



Wikidata as a Knowledge Base for People of the Greco-Roman World

MARGHERITA FANTOLI

VALERIA IRENE BOANO

EVELIEN DE GRAAF

CAMILLO CARLO PELLIZZARI DI SAN GIROLAMO

*Author affiliations can be found in the back matter of this article

DISCUSSION PAPER

ubiquity press

ABSTRACT

We explore the potential of Wikidata as a knowledge base (KB) for individuals associated with the Greco-Roman world, focusing on both the discoverability of relevant records and the current state of coverage. We develop a SPARQL query that combines multiple parameters to retrieve approximately 30,000 individuals and we propose several strategies to enhance the findability of the records. To assess coverage, we first compare the query results to a random sample of individuals described in the *Realencyclopädie der classischen Altertumswissenschaft* (Pauly–Wissowa), and then we evaluate how well Wikidata represents named individuals found in a corpus of ancient Greek and Latin texts.

CORRESPONDING AUTHOR:

Margherita Fantoli

Faculty of Arts, KU Leuven,
Leuven, Belgium

margherita.fantoli@kuleuven.be

KEYWORDS:

Greco-Roman world;
prosopography; Pauly's
Realencyclopädie
der classischen
Altertumswissenschaft; semantic
annotation; coverage

TO CITE THIS ARTICLE:

Fantoli, M., Boano, V. I.,
de Graaf, E., & Pellizzari di
San Girolamo, C. C. (2026).
Wikidata as a Knowledge
Base for People of the Greco-
Roman World. *Journal of
Open Humanities Data*, 12: 21,
pp. 1–16. DOI: <https://doi.org/10.5334/johd.457>

Following the definition of the Handbook of Semantic Web Technologies (Domingue et al., 2011, p. 79), semantic annotation establishes a bidirectional relationship between unstructured or semi-structured data (such as texts) and ontologies and knowledge bases. By enriching texts with structured contextual information, interpretation and historical reconstruction are significantly facilitated. One of the most common forms of semantic annotation, at least in the area of classical studies on which this paper focuses, is the disambiguation of named entities through the linking to a knowledge base (Named Entity Disambiguation or Linking), as seen in some notable projects such as Digital Athenaeus (Berti, 2021),¹ Digital Periegesis (Foka et al., 2021),² ToposText.³ To carry out such annotation effectively, the choice of the reference knowledge base plays a crucial role in ensuring coverage, interoperability, and annotation quality.

Focusing specifically on the annotation of ancient individuals, several databases exist that target socially, ethnically, or chronologically defined groups to varying extents (Trismegistos People,⁴ Trismegistos Authors,⁵ Lexicon of Greek Personal Names,⁶ MANTO,⁷ Prosopographia Imperii Romani):⁸ they are curated by identifiable domain experts and offer a high degree of quality and accuracy. In fact, they aim at gathering data for supporting the broader field of prosopography, intended as the “investigation of the common background characteristics of a group of actors in history by means of a collective study of their lives” (Stone, 1971), for specific time periods and geographical areas.⁹ However, none of them can function as a comprehensive prosopography of the ancient world—one that includes both prominent figures and ordinary individuals across the Greco-Roman world.

Despite efforts like the SNAP:DRGN working group to promote Linked Open Data for prosopographical resources,¹⁰ compliance with the World Wide Web Consortium (W3C) Semantic Web standards remains uneven. For example, the Lexicon of Greek Personal Names offers searchable records with unique identifiers but lacks APIs, SPARQL endpoints, or structured downloads.¹¹ The Prosopographia Imperii Romani provides an API returning partial data in JSON, while Trismegistos offers multiple APIs, including RDF. The Digital Prosopography of the Roman Empire goes further, enabling SPARQL queries on RDF data, fully embracing Semantic Web principles. Other Linked Open Data projects are discussed in Bond et al. (2021).

Against this technically heterogeneous landscape, Wikidata, an established key player in the Semantic Web (Vrandečić et al., 2023), and increasingly used within the Humanities (Zhao, 2023) and the galleries, libraries, archives, and museums (GLAM) sector (Candela et al., 2024; Van Veen, 2019), potentially offers a unified framework for semantic annotation. Wikidata is a free and open knowledge base which organizes knowledge into “items”, each with a unique identifier (Q followed by a number), representing things like people, places, or concepts. These items are described using “properties” (identified by P followed by a number), which define relationships such as “place of death” or “occupation”. Information is stored as simple

1 <https://www.digitalatheneaus.org/#page-top/>. All links in the paper have been accessed 01.11.2025. Digital Athenaeus consists of a digital semantically enriched edition of the *Deipnosophists* of Athenaeus of Naucratis.

2 <https://www.periegesis.org/en/index.php>. In this case, the focus is on the *Periegesis Hellados* by Pausanias.

3 <https://topostext.org/>. ToposText annotates Greek places and other entities in a vast corpus of English translations of ancient Greek and Latin sources. An overview of the status of semantic annotation projects for the ancient Greek and Latin works is provided in Fantoli, *forthcoming*.

4 https://www.trismegistos.org/ref/about_naw.php. Trismegistos People started as a resource focusing on personal names of non-royal individuals living in Egypt in documentary texts between 800 BCE and 800 CE, and is now being expanded to the whole Mediterranean.

5 See <https://www.trismegistos.org/authors/about.php>. The resource targets all authors who wrote between 800 BCE and 800 CE.

6 <https://www.lgpn.ox.ac.uk/>. It traces every bearer of every Greek name from the late 8th century BCE to about 600 CE. It currently contains almost 400,000 ancient Greeks.

7 <https://www.manto-myth.org/manto>. MANTO is defined as a dynamic digital portal of Greek myth.

8 <https://pir.bbaw.de/#/overview>. The resource focuses on the leadership elite of the Roman Empire during the Early and High Imperial periods, from the Battle of Actium (31 BCE) to the reign of Diocletian (284–305 CE).

9 Bradley 2025 highlights how the definition by Stone efficiently aligns with the characteristics of digital relational databases and with the notion of triple structuring the Semantic Web.

10 <https://snapdrgn.net/about.html>. SNAP:DRGN stands for Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names.

11 See https://www.w3.org/2001/sw/wiki/Main_Page for further explanation of the recommended standards.

statements linking an item (the subject) to a property and a value (the object), which can be another item or a data value like a date or number. As a collaborative resource, anyone can create new items on Wikidata, whose quality is monitored by the user community. However, this process does not provide the same level of control as expert-curated resources, and because contributions often reflect the interests of individual editors, certain areas are inevitably better represented than others. Despite this potential limitation, Wikidata is becoming an increasingly important resource also for projects related to the classical world. Several databases, including those mentioned above, link to Wikidata, and the links are frequently embedded in Wikidata through the use of specific properties: for instance, MANTO provides a link to the Wikidata item in the search interface (see [Figure 1](#)), and the MANTO ID is listed within the Wikidata corresponding item by means of the Wikidata property P9736.¹²

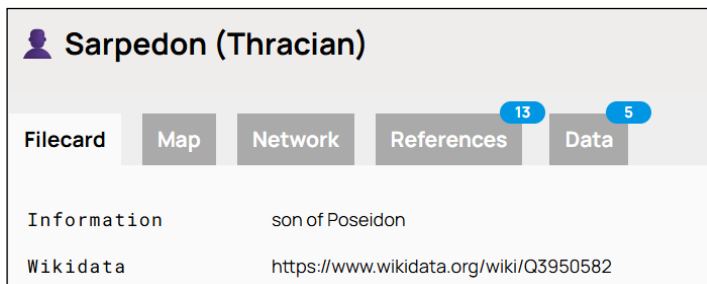


Figure 1 Screenshot of part of the MANTO page of the Thracian hero Sarpedon, including the Wikidata link.¹³

Other projects more heavily rely on Wikidata as a platform for data sharing and pooling, embracing the collaborative nature of the wiki environment. A telling example is the International (Digital) Dura-Europos Archive (IDEA),¹⁴ which integrates archaeological (meta) data related to Dura-Europos excavations in the Wikidata ecosystem and leverages Wikidata’s LOD framework for valorizing the collections ([Thornton et al., 2024a](#) and [2024b](#)). The Linked Ancient Greek and Latin (LAGL) project integrates the information on ancient authors derived from the Digital Athenaeus, the Digital Harpocraton, and the Digital Suda projects (i.e. their names, stable identifiers, and passages in which they are cited), with Wikidata-derived prosopographical information, such as the place of birth, death, occupation, ancient sources and works ([Berti, 2025](#)).

Unlike domain-specific resources, however, the usability of Wikidata as a prosopographical tool for the ancient world remains difficult to assess. We do not yet know how extensive its coverage is, how reliable its data are, or how easily ancient individuals can be retrieved. These challenges linked to the Wikidata environment are not specific to the ancient world, and have been discussed in relation to several domains (see [Farda-Sarbas & Müller-Birn, 2019](#); [Piscopo & Simperl, 2019](#) for general overviews).

This paper examines the coverage and retrievability of entries in Wikidata from the perspective of scholars of classical (ancient Greek and Latin) antiquity seeking a knowledge base for semantic annotation of ancient texts. Recent initiatives, such as the Digital Periegesis Project and the HIPE 2022 task, highlight the growing use of Wikidata for annotating persons in Greek and Latin sources ([Foka et al., 2020](#); [Ehrmann et al., 2022](#)).¹⁵ Our scope spans historical individuals mentioned in Greek or Latin texts from the 15th century BCE to the late 5th century CE,¹⁶ excluding mythological figures. This includes not only people living in regions where these languages were spoken, but also figures from neighbouring areas who interacted with the Greco-Roman world.¹⁷

¹² The same happens for instance for the Trismegistos author ID (P11252), the Digital Latin Library Catalog author ID (P8122), and the ToposText person ID (P8069).

¹³ See <https://manto.unh.edu/viewer.p/60/2616/object/6580-8189980>.

¹⁴ https://www.wikidata.org/wiki/Wikidata:WikiProject_IDEA.

¹⁵ The Digital Periegesis project also provides a browsable set of Wikidata QIDs (humans, mythological creatures, but also, for instance, epithets) “that might prove useful for annotating persons in an ancient Greek literary text”, see <https://www.periegesis.org/en/articles.php?aid=33&page=0>.

¹⁶ This working definition is not without challenges. For instance, consider a case where one individual is attested in a relevant Greco-Roman source, while a closely related figure (e.g., a relative or associate) appears only in non-Greco-Roman sources. Should we exclude the latter, and is such a distinction even feasible within Wikidata’s data model? As discussed in Section 4, we adopt an inclusive approach to address these complexities.

¹⁷ We acknowledge the difficulty of defining clear boundaries for what constitutes an “ancient Greek or Roman person”. Section 3 discusses the working criteria we adopted in constructing our dataset.

Our analysis addresses two aspects of Wikidata’s use. First, we develop and publish a query to identify records within our scope, since Wikidata lacks a category for “ancient Greek or Roman people”. We provide both the initial results and an updated dataset reflecting improvements to items and the query. Section 3 details the criteria behind the query to enable reuse in similar studies.

Second, we evaluate the dataset for accuracy and coverage (Sections 4.1–4.2). Coverage is assessed through two comparisons: (1) with the *Realencyclopädie der classischen Altertumswissenschaft* (RE, Pauly et al., 1894) and (2) with individuals mentioned in a selected corpus of Greek and Latin texts under typical annotation scenarios. The RE, published between 1893 and 1978, remains a foundational reference for classical studies. It offers one of the most comprehensive treatments of Greek and Roman cultures, drawing on textual, archaeological, and documentary sources. Its enduring relevance is reflected in later adaptations such as *Der Kleine Pauly* (1964–1975), *Der Neue Pauly* (1996–), and Brill’s online edition. Both comparisons consider presence in Wikidata and retrievability via our query, leading to refinements in both. Section 5 outlines potential applications and reuse scenarios.

(2) DATASET DESCRIPTION

REPOSITORY LOCATION

<https://doi.org/10.7910/DVN/42QMWG>

REPOSITORY NAME

JOHD Dataverse

OBJECT NAME

folders: Annotation, Query, Entities

FORMAT NAMES AND VERSIONS

CSV, TSV, TXT

CREATION DATES

01/09/2024–01/11/2025

DATASET CREATORS

Margherita Fantoli, dataset curator and data annotator; Valeria Irene Boano, dataset curator and data annotator; Evelien de Graaf, dataset curator and data annotator; Camillo Carlo Pellizzari di San Girolamo, dataset curator and data annotator; Herbert Verreth, data annotator.

LANGUAGE

English for metadata; Latin; Ancient Greek

LICENSE

CC BY-NC-SA 4.0

PUBLICATION DATE

2025-11-03

Sections 3 and 4 detail the contents of the Dataverse folders.

(3) METHOD

(3.1) QUERY DEVELOPMENT

Formulating a SPARQL query to retrieve “ancient people” from Wikidata is complex. Several criteria, such as time and place of living, nationality, language, religion, and cultural

contributions, must be considered, and these are spread across multiple properties in the Wikidata ontology. Combining them effectively requires carefully balancing precision (to avoid noise) and recall (to include relevant entries).

Because the final query is lengthy, we provide its full text in the Dataverse (“Query” folder) and summarize its components here. It relies on multiple inclusion and exclusion criteria and was iteratively refined based on result analysis. We discuss the most complete version,¹⁸ though evaluation used an earlier version with minimal differences,¹⁹ both are available in the Dataverse. Criteria added later are marked [new].

- “Instance of” criteria: items with best-rank values of P31 (instance of) that include at least one of:
 - Q5 (human)
 - Q21070568 (human whose existence is disputed)
 - Q64643615 (prosopographical phantom) [new]²⁰
- “Inclusion” criteria (see [Table 1](#)): The retrieved items need to match at least one of the following criteria:
 - I1–I3: Temporal filters based on birth, death, floruit.
 - I4–I5: Conditional inclusion based on ancient world properties or languages.
 - I6–I7: Citizenship, political entities, and time periods.
 - I8: Source-based inclusion.
- “Exclusion” criteria (see [Table 2](#)): items are excluded if they match at least one of the following criteria, relative mainly to Chinese ancient people:
 - E1: Property related to Chinese history databases
 - E2–E4: Chinese names, birthplace, citizenship
 - E5: being associated with Jainism

| CODE | CRITERION TYPE | PROPERTY | VALUE(S) | NOTES |
|----------|-----------------|---|--|--|
| I1 | Temporal | P569 (birthdate) | 1501 BCE < x < 501 CE | |
| I2 | Temporal | P570 (deathdate) | 1501 BCE < x < 551 CE | |
| I3 [new] | Temporal | P1317 (floruit) | 1501 BCE < x < 551 CE | |
| I4 | Contextual | Various: all those having P31 (instance of) -> Q56248884 (Wikidata property related to the Ancient World) | Any | Having at least one property classified as related to the Ancient World. |
| I5 | Language spoken | P1412 (languages spoken, written or signed) | Q397 (Latin), Q35497 (Ancient Greek) | |
| I6 | Geopolitical | P27 (country of citizenship) | Q1747689 (Ancient Rome), Q42834 (Western Roman Empire); any item having P31 (instance of) -> Q148837 (polis), Q3932025 (Hellenistic kingdom) | Being a citizen of Ancient Rome and/or Western Roman Empire; ²¹ and/or of an item classified as a polis and/or a Hellenistic kingdom. |

Table 1 Overview of inclusion criteria for the Wikidata query.

(Contd.)

¹⁸ Updated query: <https://qlever.dev/wikidata/H1Iu6K>.

¹⁹ Old query: <https://qlever.cs.uni-freiburg.de/wikidata/f9c57g> (now redirecting to <https://qlever.dev/wikidata/f9c57g>).

²⁰ Person whose existence has been assessed at some point by modern scholarship and then has been refuted by more recent scholarship.

²¹ The Wikidata property P27 in principle refers to proper citizenship, but it cannot be excluded that in some cases it was used more broadly.

| CODE | CRITERION TYPE | PROPERTY | VALUE(S) | NOTES |
|------|-------------------------|-----------------------------|--|--|
| I7 | Temporal/ contextual | P2348 (time period) | Any item having P361 (part of) -> Q11772 (Ancient Greece), Q1747689 (Ancient Rome), Q105747718 (Greco-Roman Egypt), Q217050 (late antiquity) | Being associated with a time period that is, directly or indirectly, part of Ancient Greece, Ancient Rome, Greco-Roman Egypt, or Late Antiquity. |
| I8 | Source-based | P1343 (described by source) | Q1138524 (Pauly-Wissowa), Q47500198 (1870 Dictionary of Greek and Roman Biography and Mythology) | Described by Pauly-Wissowa and/or the 1870 Dictionary of Greek and Roman Biography and Mythology. |

| CODE | EXCLUSION TYPE | PROPERTY | VALUE(S) | NOTES |
|------|----------------|---|--|--|
| E1 | Identifier | P497 (China Biographical Database Project), P9613 (text data entity ID) | Any | Appearing in China-related databases |
| E2 | Geopolitical | P27 (country of citizenship) | Any item having P31 (instance of) -> Q50068795 (historical Chinese state), and/or Q836688 (ancient Chinese state) [new] | Being a citizen of an item labelled as an ancient or historical Chinese state. |
| E3 | Cultural | P734 (family name) | Any item having P31 (instance of) -> Q1093580 (Chinese family name) | Having a family name classified as Chinese. |
| E4 | Geographical | P19 (place of birth) | Any item having P131 (located in the administrative territorial entity, directly or indirectly) -> Q148 (People's Republic of China) | Being born in a place which is located in the People's Republic of China. |
| E5 | Religious | P140 (religion or worldview) | Q9232 (Jainism) | |

Table 2 Overview of exclusion criteria for the Wikidata query.

Some specifications are needed to fully describe the query: first, for each property, we always use only best-rank values: deprecated-rank values are always discarded; if a property has preferred-rank values, they are the only ones included, whilst if there are no preferred-rank values, all normal-rank values are included.²²

Moreover, we applied some additional filters on some of the criteria:

- I4: if the item has best-rank values of P569 (date of birth) and/or P570 (date of death) and/or P1317 (floruit) and they are not unknown values, it must fall into the criteria I1, I2, I3; this is to exclude, for instance, medieval and modern authors receiving some of the identifiers related to the ancient world, such as a Pinakes author ID (P6831).
- I5: if the item has best-rank values of P569 and/or P570 and/or P1317 and they are not unknown values, it must fall into the criteria I1, I2, I3; this is to exclude later people speaking or writing ancient Greek and Latin, especially modern philologists.

Some criteria deserve a broader discussion, since their application impacts the overall evaluation:

- Dates: Our goal is to create a prosopographical resource for the Greco-Roman world, whose temporal boundaries are, of course, inherently uncertain. The chosen dates roughly correspond to the rise of the Mycenaean civilization and the fall of the Western Roman Empire. Naturally, this latter boundary is highly conventional and results in the inclusion or exclusion of individuals who lived relatively close in time and were undoubtedly influenced by Greco-Roman culture—not to mention those living in the Eastern Roman Empire. This criterion, therefore, should be seen as flexible and adaptable to the specific needs of each project or resource.

²² See <https://www.wikidata.org/wiki/Help:Rank>.

- Exclusion criteria: It is striking that our exclusion criteria primarily target individuals associated with Chinese civilization. This choice is motivated by two factors. First, in our initial query tests, the majority of irrelevant results fell into this category. Not excluding explicitly China-related records falling within our time-frame results in ca. 5000 additional results to query, i.e. a significant drop in precision. Second, we are reasonably confident that this exclusion would omit only a minimal number of individuals relevant to our focus (if any), as Sino-Roman contacts at the time were very shallow and mostly limited to the silk trade.²³ In contrast, for several other civilizations—whose members, as discussed below, are sometimes incorrectly retrieved by our query—it is more difficult to apply strict exclusion criteria. This is due to the fluid boundaries of Greek civilization and the Roman Empire. Such complexity affects, for instance, Egyptian, Babylonian, and Jewish civilizations, whose interactions with the Greco-Roman world evolved significantly over time and whose histories were often subjects of interest for ancient Greek and Latin authors. This aspect is further discussed in Section 4, when commenting on the results of the query.

(3.2) ANNOTATION

In order to evaluate how well Wikidata represents named individuals associated with literary sources of the Greco-Roman world, we conducted a specific case study: we compared the query results to the individuals mentioned in a selected corpus of ancient Greek and Latin texts. This section describes the annotation process of linking named individuals occurring in it to both Wikidata and the Wikisource version of the *RE*. The file containing the annotation data and a description of the format used is found in the Dataverse folder Annotation.

The corpus comprises two subsets of differing sizes, each annotated according to slightly different criteria. The first subset (Subset 1, henceforth) was developed within the context of the NIKAW project²⁴ and includes ancient Greek and Latin texts by authors from different cultural and historical contexts (de Graaf, Forthcoming). The second subset (Subset 2) is significantly smaller and consists exclusively of books II-VI of the *Naturalis Historia* by Pliny the Elder. Its annotation was performed with the aim of studying the canon of people created by Pliny, in the context of the MECANO doctoral network (Boano, Forthcoming).²⁵ Together, these two datasets contain both Greek and Latin texts spanning a time period from the 4th century BCE to the 4th century CE. Different genres and text types are included, such as letters, (philosophical) treatises, speeches, biographies, histories, dialogues, miscellanies. The diversity within the data means that, although not exhaustive, it encompasses a wide spectrum of individuals referenced in Greco-Roman literature, making it appropriate for assessing the relative coverage of knowledge bases.

The annotation process consisted of two steps: identifying named individuals and linking them to both the *RE* and Wikidata. In both corpora, the first step followed the general principle that a person is defined as “any identifiable single individual, including deities and anthropomorphic mythological figures” (Palladino et al., 2024). The second step entailed linking each mentioned individual to the corresponding *RE* article and Wikidata item. For Wikidata, the linking was performed by pointing to the item identifier (e.g. *M. Vipsanius Agrippa* was linked to the item Q48174). Linking to the *RE* required extracting relevant data from the Wikisource register and importing it into a separate database, where each *RE* entry was assigned a unique identifier.²⁶ The linking process differed slightly between the two subsets. In Subset 1, the named entities were first manually linked to *RE* identifiers and subsequently semi-automatically matched to their corresponding Wikidata items. Any *RE* entries already linked to Wikidata items in the Wikisource register were linked automatically. All remaining entities were linked manually. When a Wikisource page already exists for a *RE* article, it is also connected to the corresponding Wikidata item, improving the alignment between the two resources (Figure 2).

²³ McLaughlin & Kim, 2021, provide a complete overview of the very sparse direct contacts between the Roman and Chinese empire in the first centuries of the Roman Empire.

²⁴ <https://research.kuleuven.be/portal/en/project/3H220323>.

²⁵ <https://mecano-dn.eu/>.

²⁶ Note that the Wikisource version of the *RE* is an ongoing project, and so the articles and their relevant data are not (yet) stable. For the purpose of this project, we used the data version as scraped for the NIKAW project in spring 2024 (de Graaf et al., 2024).

| | |
|---|-----------------------------|
| Paulys Realencyclopädie der classischen Altertumswissenschaft | |
| korrigiert [Ausklappen] | |
| < Europos 10 | Europs 2 > |
| Sohn des Aigialeus, Vater des Telchin, in der Genealogie der Sikyonier | |
| Band VI,1 (1907) S. 1310 ↗ –1311 ↗ | |
| Europs (Sohn des Aigialeus) in der Wikipedia | |
| Europs in Wikidata | |
| Bildergalerie im Original ↗ | |
| Register VI,1 | Register er |
| Linkvorlage für WP [Ausklappen] | |

Figure 2 Metadata for the RE Wikisource entry of *Europos 1*,²⁷ linking to the corresponding Wikidata item *Europs* (Q654037).

Manual linking involved checking for variant spellings across different languages and was prioritised for individuals appearing in multiple works, followed by those associated with a RE identifier. For individuals mentioned in only one work and not present in the RE, only a small alphabetical sample, approximately 40% of the total, was annotated. In Subset 2, instead, the linking to the two knowledge bases was performed in parallel and entirely manually.

In both subsets, tokens were annotated individually, without merging multi-token entities. Names like *Gaius Julius Caesar* were linked per token, so a single name counts as multiple attestations of the same person—for example, three for *Gaius Julius Caesar*. With regard to nested multi-token entities, the two subsets followed different annotation criteria. In Subset 1, each token referencing a different individual was annotated with a link to the respective person: for instance, in the sequence *Africani fratris nepos* (Cic., *Tusc. Disp.* 1, 81), each token was linked to a different individual (i.e. the RE Wikisource entities *Cornelius 335*,²⁸ *Fabius 109*²⁹ and *Fabius 107*³⁰). In Subset 2, instead, the annotation was highly dependent on proper names and did not take into account nested entities: for instance, in the sequence *Daedali filio* (Plin. *Nat. Hist.* 3, 102), only *Daedali* was tagged and linked to the respective entity.³¹

As previously mentioned, the annotation guidelines we followed led to the inclusion of deities and mythological figures in the dataset. To accurately evaluate the coverage of the Wikidata query over our corpus, it was necessary to modify the dataset. The query includes only instances of “human”, “human whose existence is disputed” and “prosopographical phantom”. However, our dataset, as illustrated in [Figure 3](#), included entities from many other Wikidata categories. Retaining these additional categories would have skewed the comparison results, artificially inflating the number of unmatched entities. To avoid this bias, we removed all the entities that were not “instance of” (P31) the three categories listed above, by leveraging the established links to Wikidata. The entities that did not have a link to Wikidata were annotated manually as human or not human. After removing all the non-human entities, the dataset contained 19,262 attestations out of the original 24,807 (77.65%), corresponding to 2,279 unique entities out of

²⁷ https://de.wikisource.org/wiki/RE:Europos_1.

²⁸ https://de.wikisource.org/wiki/RE:Cornelius_335.

²⁹ https://de.wikisource.org/wiki/RE:Fabius_109.

³⁰ https://de.wikisource.org/wiki/RE:Fabius_107.

³¹ https://de.wikisource.org/wiki/RE:Daidalos_1.

(4) RESULTS AND DISCUSSION

The query was run using the QLever SPARQL engine (Bast & Buchhold, 2017; Patel-Schneider, 2025) and, as a consequence of the dynamic nature of Wikidata, the results change nearly every day, but are, at the time of writing, stabilized little above the 30,000 entries. In the Dataverse folder “Entities”, we include the json containing the full information for the 30,447 original items (full_json_retrieved_entities.json), a tsv file of the same results (reduced_dataframe_retrieved_entities.tsv), with a limited number of columns and having undergone further cleaning (removal of clearly non-relevant entities), the list of IDs of the items returned by the updated query on November 1st 2025 (wikidata_ancient_people_IDs_01_11_2025.tsv), the jupyter notebook to scrape Wikidata for the full information starting from the ID list (Wikidata_scrape.ipynb) and the two files used for evaluation (accuracy_random_sample_1000.csv, and pauly_random_sample_coverage.csv).

(4.1) ACCURACY OF THE QUERY

To evaluate the accuracy of the query results, we randomly selected 1,000 items and assessed whether they fell within the scope of our dataset. The assessment was conducted by two scholars specializing in the ancient world. Of the sample, 835 items were deemed relevant, while 165 were not (yielding an accuracy of 83.5%).

The non-relevant items fall into the following categories:

- Individuals with missing or incorrect metadata which caused the filters to malfunction.³² Most of these cases involve people outside the temporal scope of the query, but whose Wikidata items lacked proper dating.³³ Similarly, in some instances, Chinese nationality was absent or not adequately marked.³⁴
- Individuals within the temporal scope but unrelated to the Greco-Roman world, such as Vikings,³⁵ Japanese,³⁶ or Ancient Egyptians.³⁷ As discussed above, we chose not to implement strict exclusion rules in the query, with the exception of the case of Chinese entities, in order to keep the query flexible and generalizable. However, such rules could be added if needed.

When establishing which items fall within the scope of our dataset and which do not, we inevitably engage with the contested concept of ‘Western civilization’, traditionally rooted in the Greco-Roman world and early Christian Europe.³⁸ Excluding ancient eastern civilizations like China or Japan emphasizes this “western” focus, yet boundaries remain fluid.³⁹ For instance, figures from the Sasanian Empire were included due to frequent interactions with Rome.⁴⁰ Our approach is pragmatic: we target Wikidata items relevant for annotating Greek and Latin sources. Consequently, non-western individuals mentioned by these authors – such as Aktisanes, a Nubian king cited by Hecataeus⁴¹ – were also included, along with figures near the dataset’s temporal limits, even when dating is uncertain.⁴²

³² While analyzing these cases, we have solved most of the issues, so that now the items would not be retrieved by the query. We still provide the identifiers with an indication of the information added for the sake of illustration.

³³ E.g. <http://www.wikidata.org/entity/Q120963002>.

³⁴ E.g. <http://www.wikidata.org/entity/Q55697388>.

³⁵ E.g. <http://www.wikidata.org/entity/Q5887569>.

³⁶ E.g. <http://www.wikidata.org/entity/Q11391819>.

³⁷ E.g. <http://www.wikidata.org/entity/Q817116>.

³⁸ In light of the extensive debate about defining “Western civilization” in contrast to the so-called “Orient,” we refer to Federici (1995) and Goody (2006), who situate references to the Greco/Greco-Roman legacy within the construction of “Western” discourse.

³⁹ This aspect is extensively discussed in Mathisen 2003, which outlines the increasingly broader scope of the Prosopography of the Later Roman Empire.

⁴⁰ E.g. <http://www.wikidata.org/entity/Q56351145>. See Chen 2021 for an extended discussion of the intense and highly complex relations between the “Sasanian East” and the “Roman West”.

⁴¹ <http://www.wikidata.org/entity/Q279845>.

⁴² E.g. <http://www.wikidata.org/entity/Q120962912>.

Our guiding principle was inclusivity within and around the Greco-Roman context, even at the cost of some inconsistencies. Through our discussions, we defined a set of empirical inclusion criteria to assist with decisions on temporal and geographical edge cases. These criteria are not intended as necessary conditions, meaning a person does not need to meet all of them, but rather as sufficient ones for resolving ambiguous cases: if at least one criterion applies, the individual is included:

- The person is mentioned in the *RE*.
- The person is identified in Wikidata, based on primary and secondary sources, as an early Christian martyr, bishop, monk, nun, saint, or another religious category within the Christian tradition.
- The person is cited by an ancient Greek or Latin author relevant to our scope.
- The person travelled to a location within the Greek world or the Roman Empire.
- The person was involved in battles against Greek or Roman armies.

(4.2) COVERAGE OF THE QUERY AND OF WIKIDATA

(4.2.1) Comparison between *RE* and Wikidata

As mentioned in the Introduction, the *RE* is considered a reference work for classical studies. Even though its coverage, as can be expected of any scholarly work, is not entirely perfect, it can be taken as a good example of a printed resource that has supported the work of classical philologists and historians over the last decades.

We examined how many *RE* entries appear in Wikidata and what can be inferred about those missing. From the local instance of the Wikisource edition of the *RE*, we sampled 1,000 people, discarding non-human entries (e.g., mythological figures,⁴³ animals,⁴⁴ names),⁴⁵ leaving 911. Of these, 407 are in Wikidata (44%). About 250 had existing links; the rest were found manually.

Categories that are regularly missing in Wikidata include people mentioned only in a single inscription,⁴⁶ or individuals known solely from one attestation in an archaeological finding.⁴⁷ This is likely a direct consequence of the nature of Wikidata as a community-edited resource reflecting the interests of different individual editors. Also frequently missing are people related to more ‘famous’ figures (e.g. through family or friendship ties) but who are mentioned only in one or two sources.⁴⁸ Overall, even though more statistical information would be needed to reach definitive conclusions, it appears that Wikidata covers individuals for whom information is available in multiple sources. Moreover, certain categories, such as Roman consuls, ancient Greek athletes and sculptors, and literary authors, are particularly well represented. In terms of the coverage of our query, all of the items identified in Wikidata were also included in our query results.

(4.2.2) General coverage of Wikidata compared to *RE* and annotated texts

Analysis of the annotated data allows for several observations regarding Wikidata’s coverage of individuals mentioned in our texts. [Figure 4](#) illustrates how linkable attestations are distributed among entities covered by both the *RE* and Wikidata, those covered by only one of the two, and those that cannot be linked, represented by a separate bar. [Figure 5](#) presents the same distribution but at the level of unique individuals. In general, most attestations in our corpus are covered by both knowledge bases. However, the proportion of mentions not covered by

⁴³ E.g. Olenos 2, the son of Hephaistos: the corresponding Wikidata item (<https://www.wikidata.org/wiki/Q65124873>) is an instance of “Greek mythological character” (Q22988604).

⁴⁴ E.g. Dias 5, in the *RE*, referring to the horses of Amphiaraos, cf. https://de.wikisource.org/wiki/RE:Dias_5. Not yet recorded in Wikidata.

⁴⁵ E.g. Pagius, Pagurius in the *RE*, which are “alte römische Gentilnamen” (ancient Roman *nomina gentilicia*). Not yet recorded in Wikidata.

⁴⁶ E.g. Aristetas 9, Archon in the city of *Regium* (contemporary Reggio Calabria), cf. https://de.wikisource.org/wiki/RE:Aristetas_9.

⁴⁷ E.g. Dorion 5, sculptor known only from one signature on the basis of a statue, cf. https://de.wikisource.org/wiki/RE:Dorion_5.

⁴⁸ E.g. Priscus 33, father of the Pope Celestine I, cf. https://de.wikisource.org/wiki/RE:Priscus_33; “Lysitheides 2”, a friend of Themistocles, is mentioned by two sources under two different names, cf. https://de.wikisource.org/wiki/RE:Lysitheides_2.

either is noteworthy, especially when considering unique individuals, where they account for 8.8% of the total. A crucial note in correctly interpreting these results is the presence of 1,586 attestations of Jesus Christ,⁴⁹ who does not appear in the *RE*: this significantly skews the overall distribution of attestations covered only by Wikidata and explains the shift in proportions observed when comparing attestation-level data to individual-level data.

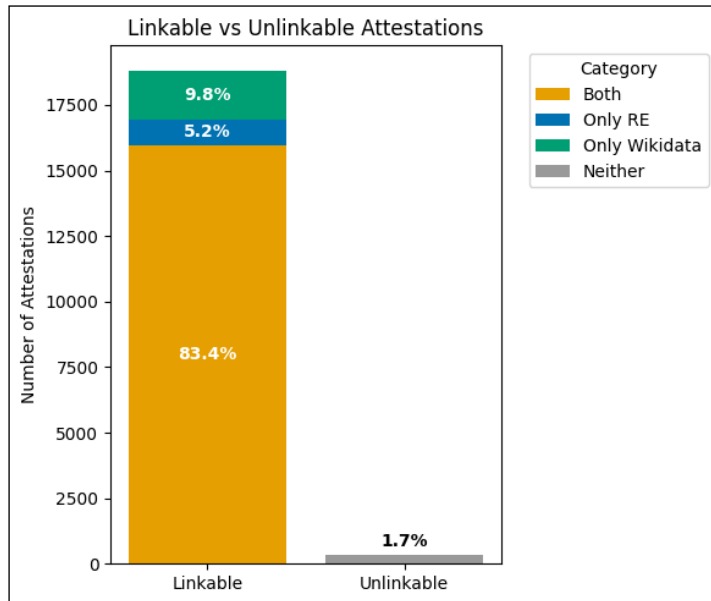


Figure 4 Bar plot that shows how the linkable attestations are distributed between the *RE* and Wikidata.

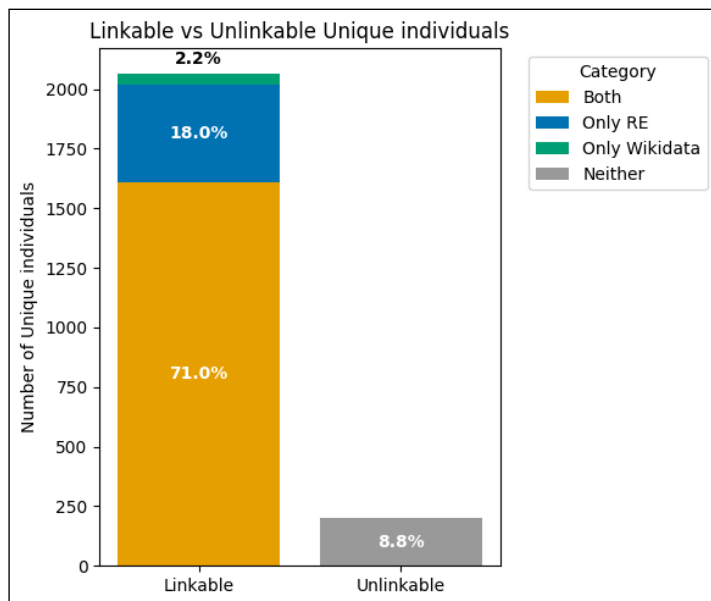


Figure 5 Bar plot that shows how the linkable unique entities are distributed between the *RE* and Wikidata.

From this perspective, we observe that, at the time of writing, the *RE* still provides better coverage of the individuals mentioned in our corpus. However, the *RE*, in terms of included entities, is a closed project, and its degree of representation of ancient individuals cannot be extended. On the other hand, Wikidata is an ongoing project, thus leaving the door open for an enhancement of its coverage in relation to the ancient world. In Section 5, we discuss how such extensions could be covered by community initiatives.

A more detailed examination of the data shows how individuals who are either represented exclusively in Wikidata or not linkable at all (i.e., absent from both Wikidata and the *RE*) are distributed across the works in our corpus (see Table 4). In Aristotle’s *Metaphysica*, Seneca’s *Epistulae Morales*, and Lactantius’s *Divinae Institutiones*, all mentioned individuals are fully represented in Wikidata. Interestingly, these texts span both ancient Greek and Latin as well as different time periods. The texts least covered by Wikidata, with a substantial number of individuals lacking representation, are Diogenes Laertius’s *Vitae Philosophorum* and Apuleius’

⁴⁹ <https://www.wikidata.org/wiki/Q302>.

Apologia. However, the two cases differ significantly. In the case of Diogenes Laertius, in fact, the number of attestations not covered by Wikidata constitutes only around 10% of the total, whereas for Apuleius' *Apologia* the percentage rises to approximately 50%. The latter can then be considered the least represented work in terms of the mentioned individuals.

| AUTHOR_WORK | ATTESTATIONS_NO_WIKIDATA | ENTITIES_NO_WIKIDATA | NOT_LINKABLE_ATTESTATIONS |
|----------------------|--------------------------|----------------------|---------------------------|
| Ael_VH | 67 | 49 | 14 |
| Amm_Marc | 26 | 16 | 2 |
| Diog_Laert | 505 | 340 | 174 |
| Strabo | 4 | 4 | 0 |
| Cic_Tusc | 15 | 10 | 4 |
| Clem_Al_Strom | 43 | 42 | 15 |
| Tac_Hist | 73 | 35 | 0 |
| Gal_PHP | 2 | 2 | 0 |
| Origen_C_Cels | 5 | 3 | 4 |
| Plut_Mor_Quaest_conv | 174 | 61 | 43 |
| Lactant_Div_Inst | 0 | 0 | 0 |
| Plato_[Epist] | 21 | 7 | 14 |
| Sen_Ep | 0 | 0 | 0 |
| Apul_Apol | 326 | 32 | 32 |
| Arist_Metaph | 0 | 0 | 0 |
| Dion_Hal_Dem | 18 | 9 | 5 |
| Tert_De_anim | 12 | 8 | 9 |
| Pliny_NH | 14 | 11 | 2 |

Table 4 The number of attestations and unique entities not covered by Wikidata or not linkable at all, distributed per work.

In Apuleius's *Apologia*, the most frequently cited individual not represented in Wikidata is his wife Pudentilla, followed by his accusers Herennius Rufus, Sicinius Aemilianus, Sicinius Pontianus, and Iunius Crassus. Interestingly, the historicity of the events narrated in this work remains unverifiable, despite numerous scholarly efforts (Hunink, 1997). The critical factor, however, is that all these individuals are known solely from this text. This pattern is common among figures not linkable to Wikidata, who frequently appear in only a single work.

However, in some cases, even people who appear in more than one text are not represented by Wikidata. This is often true for family members of well-known individuals. Table 5 lists a sample of such individuals, with their own RE label and links to their more prominent relative.

| RE LABEL | RELATIONSHIP | RELATIVE'S RE LABEL | RELATIVE'S WIKIDATA ITEM |
|---------------|--------------|---------------------|---|
| Babys 3 | father of | Pherekydes 3 | https://www.wikidata.org/wiki/Q311485 |
| Bloson 2 | father of | Herakleitos 10 | https://www.wikidata.org/wiki/Q41155 |
| Damasos 6 | brother of | Demokritos 6 | https://www.wikidata.org/wiki/Q41980 |
| Exekestides 1 | father of | Solon 1 | https://www.wikidata.org/wiki/Q133337 |
| Gryllos 2 | father of | Xenophon 6 | https://www.wikidata.org/wiki/Q129772 |
| Peirithoos 3 | father of | Alkmaion 6 | https://www.wikidata.org/wiki/Q188332 |

Table 5 A sample of individuals not appearing in Wikidata and their relationship with well-known and represented people.

In conclusion, the analysis of Wikidata's coverage within our corpus demonstrates a generally strong representation of ancient individuals, though there remains room for improvement. In particular, the investigation revealed a clear correlation between the lack of representation and the isolated or infrequent occurrence of individuals in ancient texts.

Within the annotated items that were represented in Wikidata, a small number (ca. 40) had not been retrieved by the query. Some items were missed due to incomplete data in Wikidata items, while other items were missed in the original query but were included in the improved one. This observation reveals that the quality of the entries is a key-factor for the construction of the desired knowledge base.

(5) IMPLICATIONS/APPLICATIONS

Our study highlights several conclusions about Wikidata’s usability for ancient Greek and Roman individuals. First, a query combining multiple cultural factors yields highly reliable results, as nearly all evaluated individuals were retrieved. Accuracy could improve with refined exclusion criteria.

Raising awareness within the Classics Wikidata community would enhance metadata quality—adding language, nationality, and dates improves discoverability. Our query is easily adaptable by adjusting parameters such as time span or nationality, contributing to best practices for SPARQL in Digital Classics. WikiProjects (Antiquity, Ancient Greece, Ancient Rome)⁵⁰ and initiatives like Pelagios or Linked Pasts offer platforms for coordination, documentation, and annotation sprints.⁵¹

The dataset serves as a foundation for prosopographies, data enrichment, and quantitative studies on the representation of antiquity online. Preliminary work explored Wikipedia page views to analyze the “online popularity” of ancient people, and investigated the authority sources used to support information via Wikidata properties like “described by source” (P1343), external IDs and references associated with statements.

The dataset also supports annotation of canonical texts, integration with tools like the Digital Periegesis search interface of Wikidata items (see Section 1), and practical applications – e.g., training Named Entity Linking models on Wikidata and Wikipedia tailored to classical antiquity. Recently, a quality analysis of the entries resulted in batch edits improving thousands of occupation statements.⁵²

ACKNOWLEDGEMENTS

We would like to thank the team behind the publication of the Wikisource *RE* for their invaluable work, as well as the Trismegistos+ team for their thorough integration of data with the *RE* and for their infrastructure support during the annotation phase. We are also grateful for the feedback received on the initial stages of this work during the Semantic Annotation for the Ancient World conference, organized by the TALOS Centre at the University of Crete (May 2024).

FUNDING STATEMENT

This research was supported by the NIKAW project (Interdisciplinary Network, funded by the Bijzonder Onderzoeksvonds of the KU Leuven, project code IDN/22/001) and the MECANO (The Mechanics of Canon Formation and the Transmission of Knowledge from Greco-Roman Antiquity) MSCA Doctoral Network (Grant agreement ID: 101120349).

COMPETING INTERESTS

One of the authors of the paper is an editor of the special issue Wikidata Across the Humanities: Datasets, Methodologies, Reuse, to which the paper was submitted. Moreover, the corresponding author has acted as a reviewer for this journal.

AUTHOR CONTRIBUTIONS

Margherita Fantoli: Conceptualization, Funding Acquisition, Data Curation, Methodology, Supervision, Writing – original draft

Valeria Irene Boano: Conceptualization, Data Curation, Methodology, Writing – original draft

50 https://www.wikidata.org/wiki/Wikidata:WikiProject_Antiquity, https://www.wikidata.org/wiki/Wikidata:WikiProject_Ancient_Greece, https://www.wikidata.org/wiki/Wikidata:WikiProject_Ancient_Rome. Beyond offering a platform for storing documentation, these projects (via the talk pages) can host discussions about modelling and quality issues and how to solve them.

51 See <https://pelagios.org/activities/people/> and <https://linkedpasts.hcommons.org/>.

52 See https://www.wikidata.org/w/index.php?title=Wikidata_talk:WikiProject_Ancient_Rome&oldid=2443076648#Ancient_Roman_occupations.

AUTHOR AFFILIATIONS

Margherita Fantoli  orcid.org/0000-0003-3191-4860

Faculty of Arts, KU Leuven, Leuven, Belgium

Valeria Irene Boano  orcid.org/0009-0003-2791-1365

Faculty of Arts, KU Leuven, Leuven, Belgium

Evelien de Graaf  orcid.org/0009-0006-8650-1595

Faculty of Arts, KU Leuven, Leuven, Belgium

Camillo Carlo Pellizzari di San Girolamo  orcid.org/0000-0003-2699-1693

Classe di Lettere e Filosofia, Scuola Normale Superiore, Pisa, Italy

REFERENCES

- Bast, H., & Buchhold, B.** (2017). QLever: A Query Engine for Efficient SPARQL + Text Search. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 647–656. <https://doi.org/10.1145/3132847.3132921>
- Berti, M.** (2021). *Digital Editions of Historical Fragmentary Texts*. Propylaeum. <https://doi.org/10.11588/PROPYLAEUM.898>
- Berti, M.** (2025). Linked Ancient Greek and Latin (LAGL) and Wikidata: Structuring and Reusing Data of Classical Literature. *Journal of Open Humanities Data*, 11, 72. <https://doi.org/10.5334/johd.423>
- Boano, V. I.** (Forthcoming). From the critical edition to a digital semantic annotation: implementing a digital pipeline for Pliny the Elder's *Naturalis Historia*. *Semantic Annotation for the Ancient World 2025, Conference proceedings*.
- Bond, S. E., Dilley, P., & Horne, R.** (Eds.). (2021). Linked Open Data for the Ancient Mediterranean: Structures, Practices, Prospects. *Institute for the Study of the Ancient World*. <http://hdl.handle.net/2333.1/gqnk9kz2>
- Bradley, J.** (2025). Building a Digital Prosopography. *Digital Classics Online, Bd. 10*, 77–106 Seiten. <https://doi.org/10.11588/DCO.2024.10.103975>
- Candela, G., Cuper, M., Holownia, O., Gabriëls, N., Dobрева, M., & Mahey, M.** (2024). A Systematic Review of Wikidata in GLAM Institutions: a Labs Approach. In A. Antonacopoulos, A. Hinze, B. Piwowarski, M. Coustaty, G. M. Di Nunzio, F. Gelati, & N. Vanderschantz (Eds.), *Linking Theory and Practice of Digital Libraries* (Vol. 15178, pp. 34–50). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72440-4_4
- Chen, A. H.** (2021). Late Antiquity Between Sasanian East and Roman West: Third-Century Imperial Women as Pawns in Propaganda Warfare. In W. V. Harris & A. Hunnell Chen (Eds.), *Late-Antique Studies in Memory of Alan Cameron* (pp. 168–197). BRILL. https://doi.org/10.1163/9789004452794_010
- de Graaf, E.** (Forthcoming). Challenges in Disambiguating Ancient Individuals with Paulys Realencyclopädie. *Semantic Annotation for the Ancient World 2025, Conference proceedings*.
- de Graaf, E., Depauw, M., & Fantoli, M.** (2024). “Nescio Carneades iste qui fuerit”: Evaluation of Knowledge Bases for Named Entity Linking for Latin Texts. In C. Swaelens, M. Deforche, I. D. Vos, & E. Lefever (Eds.), *Proceedings of The First Workshop on Data-driven Approaches to Ancient Languages* (pp. 1–19). Language & Translation Technology Team (LT3).
- Domingue, J., Fensel, D., & Hendler, J. A.** (2011). *Handbook of Semantic Web Technologies*. Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-92913-0>
- Ehrmann, M., Romanello, M., Najem-Meyer, S., Doucet, A., & Clematide, S.** (2022). Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents. In A. Barrón-Cedeño, G. Da San Martino, M. Degli Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, & N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (Vol. 13390, pp. 423–446). Springer International Publishing. https://doi.org/10.1007/978-3-031-13643-6_26
- Fantoli, M.** (Forthcoming). Semantic Annotation: Resources, Tools, and Methods. In E. Barker, O. Bobou, & R. Raja (Eds.), *The Oxford Handbook of Digital Classical Studies*. Oxford University Press.
- Farda-Sarbas, M., & Müller-Birn, C.** (2019). *Wikidata from a Research Perspective — A Systematic Mapping Study of Wikidata*. arXiv. <https://doi.org/10.48550/ARXIV.1908.11153>
- Federici, S.** (Ed.). (1995). *Enduring Western civilization: The construction of the concept of Western civilization and its “others.”* Praeger.

- Foka, A., Barker, E., Konstantinidou, K., Mostofian, N., Demiroglu, O. C., Kiesling, B., & Talatas, L. (2020). Semantically geo-annotating an ancient Greek “travel guide” Itineraries, Chronotopes, Networks, and Linked Data. *Proceedings of the 4th ACM SIGSPATIAL Workshop on Geospatial Humanities* (pp. 1–9). <https://doi.org/10.1145/3423337.3429433>
- Foka, A., Kostantinidou, K., Mostofian, N., Talatas, L., Brady Kiesling, J., Barker, E., Demiroglu, O. C., Palm, K., McMeekin, D. A., & Vekselius, J. (2021). Heritage metadata: A digital Periegesis. In K. Golub & Y.-H. Liu (Eds.), *Information and Knowledge Organisation in Digital Humanities: Global Perspectives*. Routledge. <https://doi.org/10.4324/9781003131816-11>
- Goody, J. (2006). *The theft of history*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511819841>
- Hunink, V. (1997). *Apuleius of Madauros: Pro Se de Magia: Apologia*. Gieben.
- Mathisen, R. W. (2003). The Prosopography of the Later Roman Empire: Yesterday, Today and Tomorrow. In *Fifty Years of Prosopography: The Later Roman Empire, Byzantium and Beyond*. Oxford University Press. <https://doi.org/10.5871/bacad/9780197262924.003.0003>
- McLaughlin, R., & Kim, H. (2021). Roman envoys and trade ambassadors in Han China. In *Rome and China Points of Contact* (1st ed.). Routledge. <https://doi.org/10.4324/9781315280738-2>
- Palladino, C., Fantoli, M., de Graaf, E., Berti, M., Romanello, M., Yousef, T., Beersmans, M., Gheldof, T., Soffiantini, L., & Litta Modignani Picozzi, E. (2024, May 28). Experience and Challenges with Named Entities – Workshop at DHBenelux 2024. Do all roads lead to {Rome: LOC}? Experiences and challenges with named entities in ancient and medieval corpora, KU Leuven. *Zenodo*. <https://doi.org/10.5281/zenodo.11366870>
- Patel-Schneider, P. F. (2025). Benchmarking SPARQL Engines on Wikidata Queries. In N. Ferranti, M. Machado, J. Gomes Junior, S. Razniewski, A. Polleres, & M. Krötzsch (Eds.), *Proceedings of the 5th Wikidata Workshop* (Wikidata 2025). Nara, Japan, November 2, 2025. CEUR Workshop Proceedings.
- Pauly, A. F. v., Wissowa, G., Kroll, W., Mittelhaus, K., Ziegler, K., Gärtner, H., & Wünsch, A. (1894). *Paulys Real-Encyclopädie der classischen Altertumswissenschaft: Neue Bearbeitung*. Druckenmüller.
- Piscopo, A., & Simperl, E. (2019). What we talk about when we talk about wikidata quality: a literature survey. *Proceedings of the 15th International Symposium on Open Collaboration*, 1–11. <https://doi.org/10.1145/3306446.3340822>
- Stone, L. (1971). Historical studies today. *Daedalus*, 100(1), 46–79.
- Thornton, K., Seals Nutt, K., & Chen, A. (2024a). 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, K., Seals-Nutt, K., & Chen, A. (2024b). Dura-Europos Stories: Developing interactive storytelling applications using knowledge graphs for cultural heritage exploration. *Semantic Web*, 1–15. <https://doi.org/10.3233/SW-243552>
- Van Veen, T. (2019). Wikidata: From “an” Identifier to “the” Identifier. *Information Technology and Libraries*, 38(2), 72–81. <https://doi.org/10.6017/ital.v38i2.10886>
- Vrandečić, D., Pintscher, L., & Krötzsch, M. (2023). Wikidata: The Making Of. *Companion Proceedings of the ACM Web Conference 2023*, 615–624. <https://doi.org/10.1145/3543873.3585579>
- Zhao, F. (2023). A systematic review of Wikidata in Digital Humanities projects. *Digital Scholarship in the Humanities*, 38(2), 852–874. <https://doi.org/10.1093/lhc/fqac083>

TO CITE THIS ARTICLE:

Fantoli, M., Boano, V. I., de Graaf, E., & Pellizzari di San Girolamo, C. C. (2026). Wikidata as a Knowledge Base for People of the Greco-Roman World. *Journal of Open Humanities Data*, 12: 21, pp. 1–16. DOI: <https://doi.org/10.5334/johd.457>

Submitted: 03 November 2025

Accepted: 22 December 2025

Published: 03 February 2026

COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.