

Method assignment - Categorizing Dekoloniale biographies in Wikidata

Characters: 7743

Method assignment - Categorizing Dekolonial biographies in Wikidata	1
Introduction & Context	2
Personal Role & Responsibilities	2
Methodological approach	2
Challenges & Considerations	3
Critical reflection	4
Conclusion	5

Statements

instance of

human

edit

▼ 0 references

+ add reference

+ add value

image

edit

Joseph Bilé.png
194 × 275; 68 KB

► 1 reference

+ add value

Introduction & Context

This assignment analyzes the practical process that is turning a biography from the project Dekoloniale into structured data on the platform Wikidata. The purpose of this task was transferring one of these stories of marginalized people that Dekolonial has been working to put a spotlight on, into a Linked Open Data format to make the data findable, accessible, and usable. Our group's work focused on the biography of Joseph Bilé (Q105336549) who was a political activist and actor from Cameroon, who moved to Germany. Before we started working on it, his Wikidata profile consisted mostly of basic biographical facts, such as date of birth and death, and gender. It was missing important information about his political life, his family and his work, as the Dekoloniale-text was writing about.

Personal Role & Responsibilities

This assignment is individual, but we contributed to Wikidata in a group setting. The work in the group was fully done collaboratively. Instead of splitting the tasks, we decided to discuss and implement all the edits together. This approach was taken to make sure that all group members were sure of how to do it properly, and so we could learn together. From the data collection from the Dekoloniale site, the ontological research of finding the right properties, to the technical submitting of data to wikidata with source citing. The responsibility was therefore a collective one.

Methodological approach

Our methodological approach consisted of a systematic “translation” of Joseph Bilés narrative biography from Dekoloniale to Wikidata's structured triple format of subject-predicate-object. The process was split into the three phases of data identification, ontological mapping and technical implementation.

First we analyzed the biography by extracting central entities. Then we identified the data in the categories family relations, occupation and political affiliation. Afterwards we compared those findings with Bilé's existing Wikidata-item (Q105336549) to find flaws and missing information.

The central part of our work was in mapping narrative information to the correct Wikidata properties and items. The text describes Bile's political affiliation. To structure this, we chose the property “member of political party” (P102) and linked him to the item “Communist Party of Germany” (Q152541). This was a necessary conversion of the text's nuances into a concrete data-post. His occupation was registered by using the property “occupation” (P106), where we added the value “actor” (Q33999). To register his father, James Bilé M'bule, we first had to create a whole new item for the father (Q136450220) since the relation “father” (P22) required an existing target-item and not just a text string.

In the end we added the references to all the Wikidata claims by using “stated in” (P248) with a reference to the Dekoloniale project (Q136450525) and a “reference URL” (P854) to make sure the data is verifiable and lived up to the Linked Open Data standards.



Figure 1: Screenshot from Joseph Bilé’s Wikidata page, showing the column about his father.

Challenges & Considerations

The process of ‘translating’ a biography to a database gave us several methodological and ontological challenges, and highlighted the complexity between reality and this kind of database format's strict structure.

One example is the issue we faced when trying to register the location of where Bilé died. He died in what is now known as Cameroon in 1959, but Wikidata's ontology defines the entity of “Cameroon” (Q1009) as being created in 1960. The system gave us an error, when the year of death was before the entity’s “official” creation date. Our time restriction stopped us from finding out if the geographical area of Cameroon’s earlier name was defined on Wikidata, so it may be possible - but still, this does illustrate a conflict between historical geography and the databases state-based definitions.

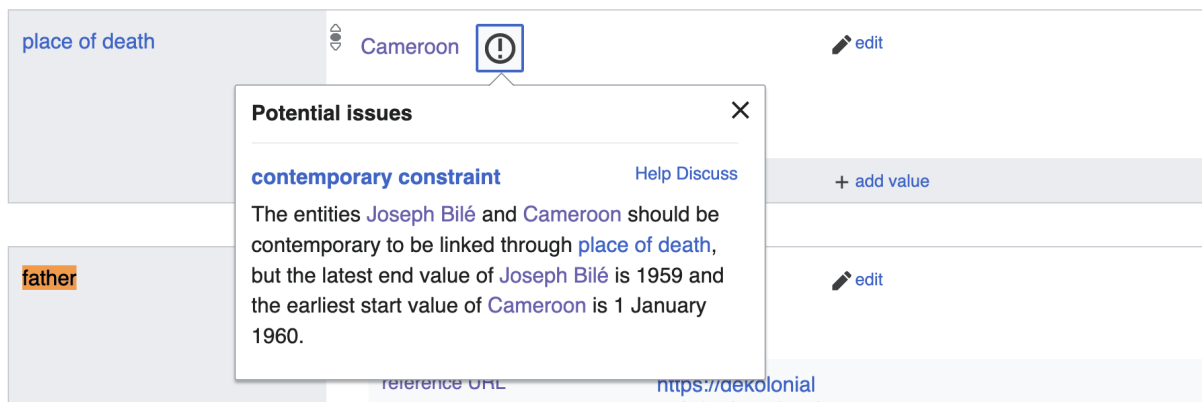


Figure 2: Screenshot from Joseph Bilé’s Wikidata page, showing the issue with adding Cameroon.

Another challenge was registering Joseph Bilé’s father, James Bilé M’Bule, as it was necessary to create a whole new item / page for him (Q136450220). First after doing that could we establish the son’s connection to his father (P22). This example illustrates how knowledge in a database such as Wikidata is relationally conditioned based on already existing data, which is what Nadim (2021) talks about, when she talks about “connections” (p. 63) and “chains of transformations” (p. 64).

Finding the fitting properties to capture the nuances in the biography was a challenge in itself. An example was that Bilé’s biography mentions his military training taking place at “The Garrison In Meiningen”, but there was no property for registering “place for military training”, which meant that this specific information was lost in translation into the structured data.

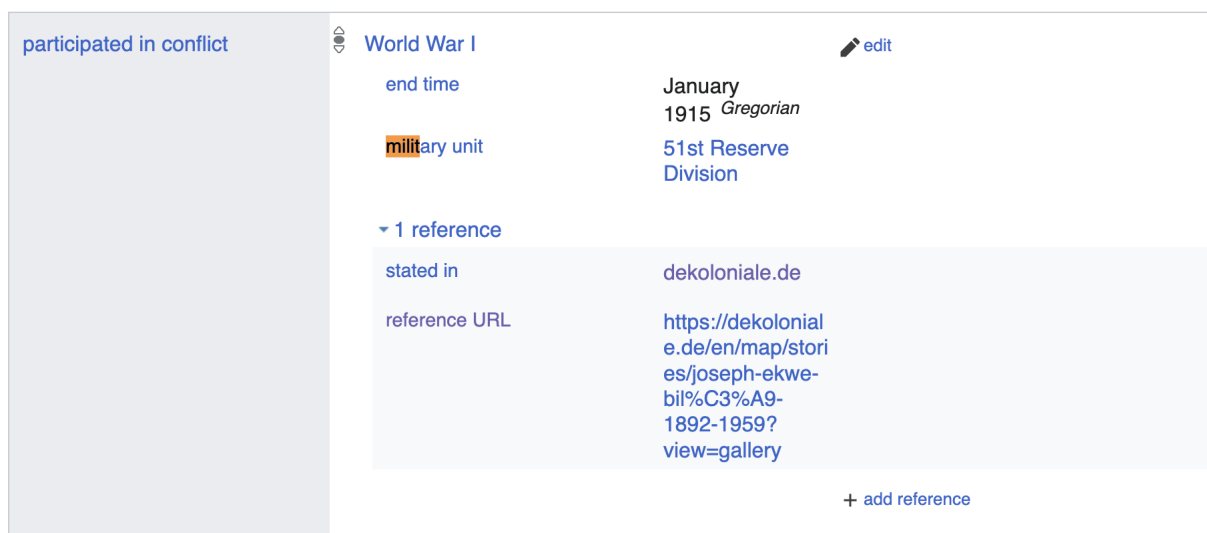


Figure 3: Screenshot from Joseph Bilé’s Wikidata page, showing the lack of military training location information.

Critical reflection

Working with converting Bilé’s biography into Wikidata made me more aware of the conflict between historical complexity and the logic of the database. This can be analyzed through Louikassas’ (2019) concept of “All Data Are Local”. Even though Wikidata appears as an universal platform, it is still formed by specific, institutional definitions.

The previously mentioned conflict regarding Cameroon’s creation date is a prime example of what Loukissas calls a “data artifact”, which is a form of database conflict where the database’s local truth has higher priority than the historical, geographical reality.

The transformation process in itself can also be seen through the eyes of Nadim's (2021) concept of datafication and "chains of transformation". When we convert a narrative of discrimination and political war into datapoints such as "member of political party" (P102), a form of abstraction happens. On one hand we create, what Nadim calls, new "connections" that change Bile's story. By linking him to the German communist party he becomes part of a globally searchable network, which enhances his "discoverability". On the other hand will this way of structuring also create disconnections. The exclusion of military training location as mentioned earlier exemplifies this, since there wasn't a fitting property for this specific context. Likewise was it hard to properly capture the nuances about the racism which affected his acting carrier and use them in the property "occupation". The datafication of Wikidata prioritizes relational connections, such as who he knew and what he was a member of, over the qualitative, human experience, which confirms Nadim's argument about data never being a neutral representation of nature, or of history, in the case of Joseph Bilé.

Conclusion

Trying to translate a Dekoloniale biography into Wikidata was an interesting exercise, where the process had our group trying to navigate between translating Bilé's rich, human history into a rigid, standardized database structure.

The task proved to us how Linked Open Data is a powerful tool that also works for creating awareness of marginalised histories, since Bilé's profile now is a bigger part of a global knowledge network. And yet, the proces isn't neutral. It leaves a bunch of curatorial choices up to the person "translating", such as: what category is most accurate? What is left out, when an accurate category is missing? What is lost in translation from biography to data points? The task shows that the contributor isn't just typing in data, but actively participating in a curricular practice that produces a new, specific form of knowledge with the limitations the format contains.

References

- Nadim, Tahani. 2021. "The Datafication of Nature: Data Formations and New Scales in Natural History." *Journal of the Royal Anthropological Institute* 27 (S1): 62–75.
<https://doi.org/10.1111/1467-9655.13480>
- Yanni Alexander Loukissas. "A Place for Plant Data", *All Data Are Local: Thinking Critically in a Data-Driven Society*, 2019
<https://direct.mit.edu/books/oa-monograph/4323/chapter-standard/181049/A-Place-for-Plant-Data>