Method assignment - Categorizing Dekoloniale biographies in Wikidata

Characters: 6860

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

The methodological approach of converting the biographical text into structured data had a multiple steps process. We started by analyzing the Dekoloniale biography of Joseph Bilé to identify central data categories. We found important data in the categories of family relations, occupations, political beliefs and military service. We then analyzed the Wikidata profile of Bilé (Q105336549) to identify what data, from what we had found, was missing from the profile.

Now, we had to add the missing information. For every new piece of information that we had to add, we had to find the correct "property", which is what Wikidata uses to categorize the data on the profiles. An example is that "father" (P22) was used for the data about his father, "occupation" (P106) for his work, and "member of political party" (P102) for his political beliefs. Then, at the end of adding these properties, we added a source to make sure it can be verified. The property for adding a source was "stated in" (P248), where we referred to the Dekoloniale project (Q136450525), using a subproperty of "reference url" (P854) to link directly to the specific biography.

***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 methodical 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.
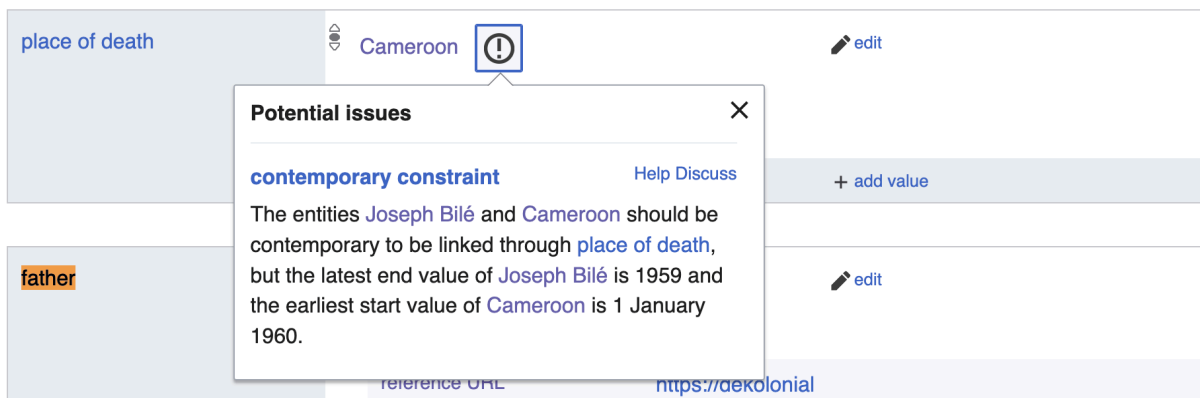

Figure 2: Screenshot from Joseph Bilé´s Wikidata page, showing the issue with adding Cameroon.

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

The practical challenges can be analyzed using concepts from Nadim (2021) and Loukissas (2019). The Cameroon problem illustrates Loukassa's idea about "All Data Are Local", because that even though Wikidata appears as a global, universal platform, it's data is really local (Loukissas, 2019) since it's formed by the institutionally defined framework (Loukissas, 2019) that the data is created inside of. Our problem was a "data artifact" (Loukissas, 2019), a kind of inconsistency, that maybe doesn't really reflect the historical reality, but instead the database's specific, local reflection of the world.

The process can moreover be seen as datafication (Nadim, 2021), since we're dealing with the translation of Bilé's life into a dataset. This process can create new "connections" but also disconnections (Nadim, 2021). On the one hand, we did create new connections, by linking Bilé to "Communist Party of Germany" (Q152541) we essentially changed the scale of his history,

and he is now a more elaborate part of this digital network, which in turn enhances his discoverability (Nadim, 2021). On the other hand, we also experienced disconnections. It was hard to represent nuances such as the discrimination, he experienced. We could register that he was an actor, but not how his career was affected by racism.

Conclusion

Trying to translate a Dekoloniale biography into Wikidata was an interesting excercise, 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 cuticle 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