Categorization – Referencing Knowledge Resources in Linked Open Data

Alexander Juby-Rasmussen (AJR), AU694667, 202208686@post.au.dk
5th Semester, Cognitive Science BSc, Arts Faculty, Aarhus University
Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark
Critical Data Studies, Elective
Curating Data
Characters: 6.274

Introduction & Context

This project contributed to Wikidata by expanding the biography of Mpondo Akwa, an overlooked figure in Cameroon's early anti-colonial movement. Due to limited documentation and colonial marginalisation, his story like many others remains underrepresented. The project aimed to make such marginalised biographies more accessible by transforming textual sources into structured linked open data. This shared contribution to building a network of contributions rather than a fixed narrative is representative of Linked Open Data and allows for greater variability and confidence in the reported work. (Ford & Iliadis, 2023)

Through a hands-on workshop, we learned how to convert narrative texts into Linked Open Data on Wikidata. This process not only connected Dekoloniale's historical materials to an open, semantic database but also highlighted how data modelling and categorisation shape historical visibility and representation. The exercise emphasised the political dimensions of knowledge creation and encouraged critical reflection on how colonial legacies continue to influence modern understandings of history.

Personal Role & Responsibilities

The assignment began with a group agreement on which biography to work with. After selecting: *'The colonial critic Mpondo Akwa [c. 1871 - c. 1914] – Germany | Cameroon'*, we collaboratively curated data from his biography into a shared database, compiling and organising information from the text into a structured format. This collaborative document provided transparency and coordination across our individual contributions, ensuring that everyone participated equally throughout the process.

We divided the biography from Dekoloniale into sections, enabling each member to transform their portion into structured and simplified data. This systematic approach allowed us to cover more aspects of Mpondo Akwa's life and expand the overall dataset.

Methodological Approach

To transform our data from the shared spreadsheet into Wikidata, we first merged our individual datasets into a single, structured spreadsheet, as we had all used the same column format. After reviewing the combined data, we analysed the existing Wikidata entry for Mpondo Akwa to assess what information was already available and to identify missing elements. This analysis quickly revealed significant gaps in sourcing and detail, as much of the existing content lacked credible references and thorough explanation. To address this, our group began adding and citing reliable sources to improve the quality and completeness of the Wikidata entry.

In line with Kitchin's discussion of "data infrastructures," this process highlighted how Wikidata's ontology structures knowledge into discrete, standardised units, shaping not only how information is stored but also how historical narratives like that of Mpondo Akwa are represented and understood within digital knowledge systems. (Kitchin, 2014)

Challenges & Considerations

A key challenge in this project was determining how to categorise and simplify complex textual information within the spreadsheet and later in Wikidata. Some details had to be omitted or modified to fit predefined columns, raising ethical concerns about whether the data might become overly subjective or misrepresent significant events. This issue was compounded by Wikidata's data model and property constraints, which limited how fully Mpondo Akwa's life could be represented.

Working within Wikidata also meant engaging with its taxonomy-based system, which organises knowledge through hierarchies and relationships. While this structure allows for: coherence, reusability and interconnected data. It also introduces challenges, if a value does not already exist in the database, linking data can lead to errors or misleading associations. Therefore, the process requires balancing accuracy and structure while staying aware of these systems both enable and constrain the representation of complex historical narratives.

Critical Reflection

Working on this project revealed that curating and modeling data are not objective acts but processes that actively shape knowledge. As Bowker and Star (Bowker & Star, 2008) note, classifications both reveal and obscure realities. Transforming complex biographical information into structured data required deliberate choices about what to include and how to represent it. In line with Ford and Illiadis (Ford & Iliadis, 2023) our work with Wikidata demonstrated that data infrastructures carry social and political implications what becomes visible depends on who contributes and how. Similarly, Clausen et al. (Clausen et al., 2023) emphasise that structuring data functions as a curatorial practice, transforming text into new, mediated forms of knowledge.

Through the workshop, it became clear that Wikidata is not a neutral database but a curatorial and semantic infrastructure, a system where both the platform and its

community shape what counts as valid knowledge. Group collaboration further highlighted this dynamic, as we continually negotiated how to represent information and define relationships within shared ontological frameworks.

Conclusion

The process of transforming decolonial information into structured data revealed how every decision shapes how a story is ultimately understood. While Wikidata supports the preservation and accessibility of historical knowledge, it also simplifies complex narratives and risks omitting essential context.

Working on Mpondo Akwa's biography showed how data modelling can make marginalised histories more visible, while also exposing the limitations of the systems that represent them. This experience has deepened my understanding of data curation as a critical, ethical, and responsibility. Although Wikidata's categories are constrained, contributing to it felt meaningful as it helps safeguard stories that might otherwise disappear.

In conclusion, the project demonstrated that curating data is not merely a technical task but a creative and ethical act of deciding which stories to preserve and how they will be remembered. Open linked data, when approached critically, can serve as a tool for: representation, connecting, sharing, and sustaining overlooked narratives within broader networks of knowledge.

Bibliography

Bowker, G. C., & Star, S. L. (2008). *Sorting things out: Classification and its consequences* (1. paperback ed., 8. print). MIT Press.

Clausen, B., Engel, D., & Wharton, G. (2023). Capturing Narrative and Data in Performance Art: The Joan Jonas Knowledge Base. In *Reconstructing Performance Art*. Routledge.

Ford, H., & Iliadis, A. (2023). Wikidata as Semantic Infrastructure: Knowledge Representation, Data Labor, and Truth in a More-Than-Technical Project. *Social Media + Society*, *9*(3), 20563051231195552. https://doi.org/10.1177/20563051231195552

Kitchin, R. (2025). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences , by Rob Kitchin. 2014. Thousand Oaks, California: Sage Publications. 222+xvii. ISBN: 978-1446287484, $100. *ResearchGate*. https://doi.org/10.1111/jors.12293