

WikiProvenance: Are there enough references to every (known) fact on Wikidata?

John Samuel



Aim

- ▶ Understanding how different Wikimedia properties document an item
- ▶ Understanding how references and external identifiers are used
- ▶ Provide 'real-time' [statistics](#) to Wikimedia contributors
- ▶ Compare statistics of different items

WikiProvenance

WikiProvenance: Visual interface to understanding and comparing the sources describing Wikidata items

Features:

- ▶ Get [statistics](#) on sources describing Wikidata items
- ▶ [Navigate](#) articles on Wikimedia projects and external sources
- ▶ [Compare](#) statistics between two items
- ▶ [Search](#) Wikidata items
- ▶ Use of [bookmarkable](#) links for sharing

Statistics on sources describing an item

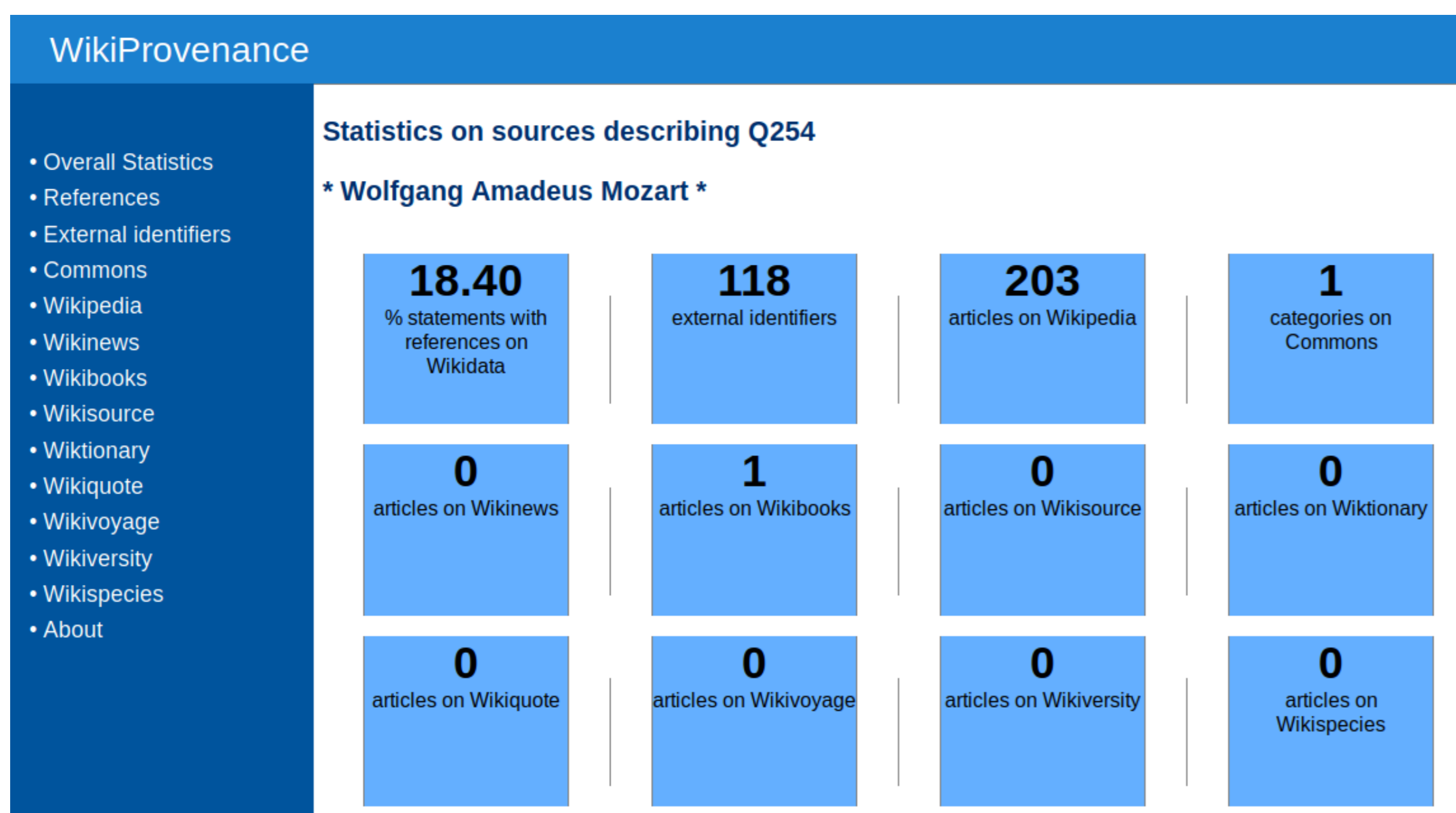


Figure: Translated label statistics

Comparison of statistics on sources describing Johann Sebastian Bach *

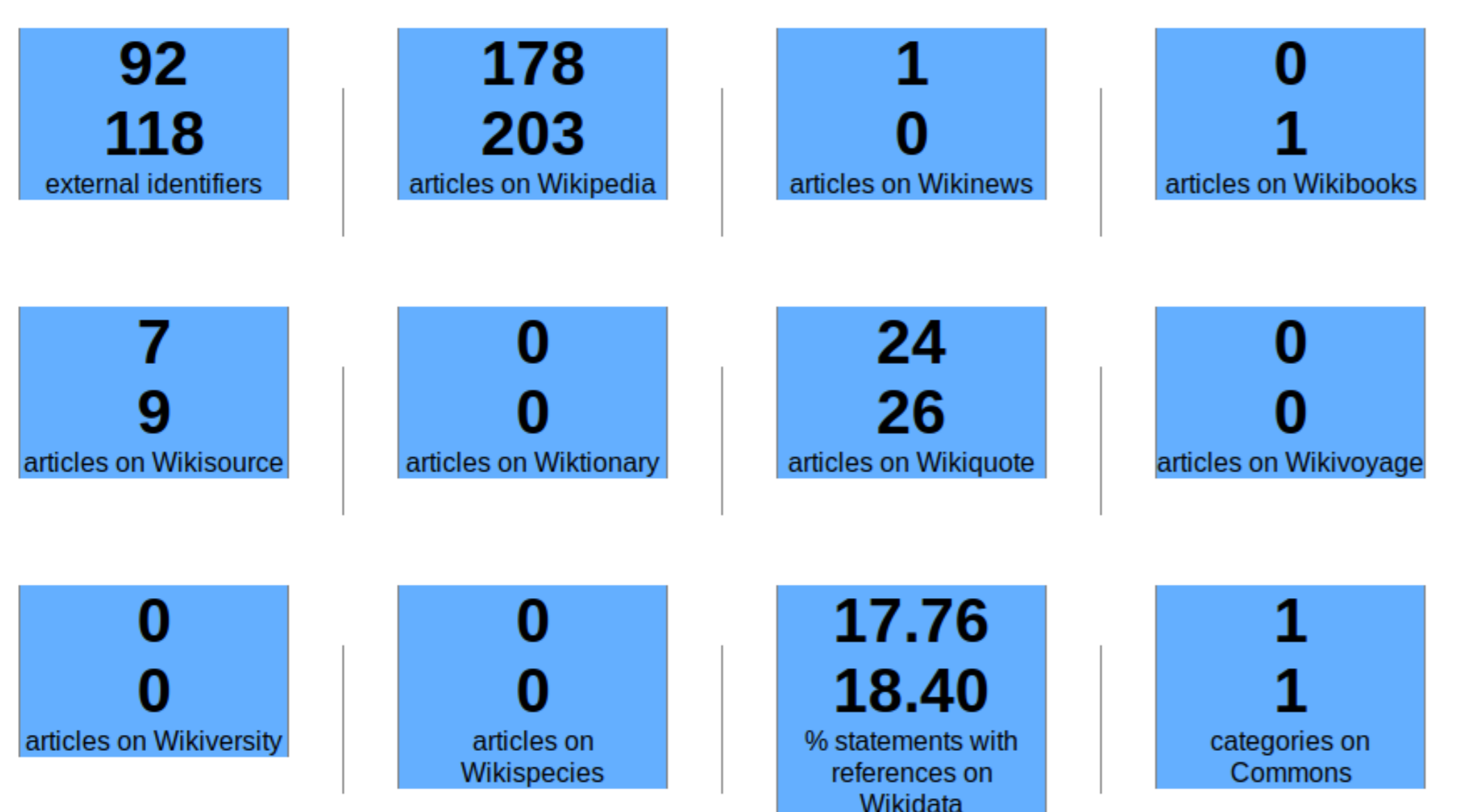


Figure: Comparison of statistics

Future Works

- ▶ Understanding types of references in Wikidata
 - ▶ Imported from Wikipedia properties (by bots etc.)
 - ▶ Reference URLs
 - ▶ Books, journals, scientific articles
- ▶ Incorporating Wikipedia references and OpenCitations to qualify an item

Questions concerning Wikidata items

Wikidata: Shift from non-structured textual content to structured data enabling the possibility of adding references to every statement (or fact)?

1. **Who** is talking about it?
 - 1.1 Languages
 - 1.2 External sources
 - 1.3 Media
2. **What** are they talking about?[Sam17, Sam18]
3. **Which** statements have **references**?
4. **How** different items belonging to same class differ in the use of references?

URL

Check <https://tools.wmflabs.org/wikiprovenance>

Wikidata: Multilingual labels



Figure: Labels in different languages

Wikidata: References and External Sources

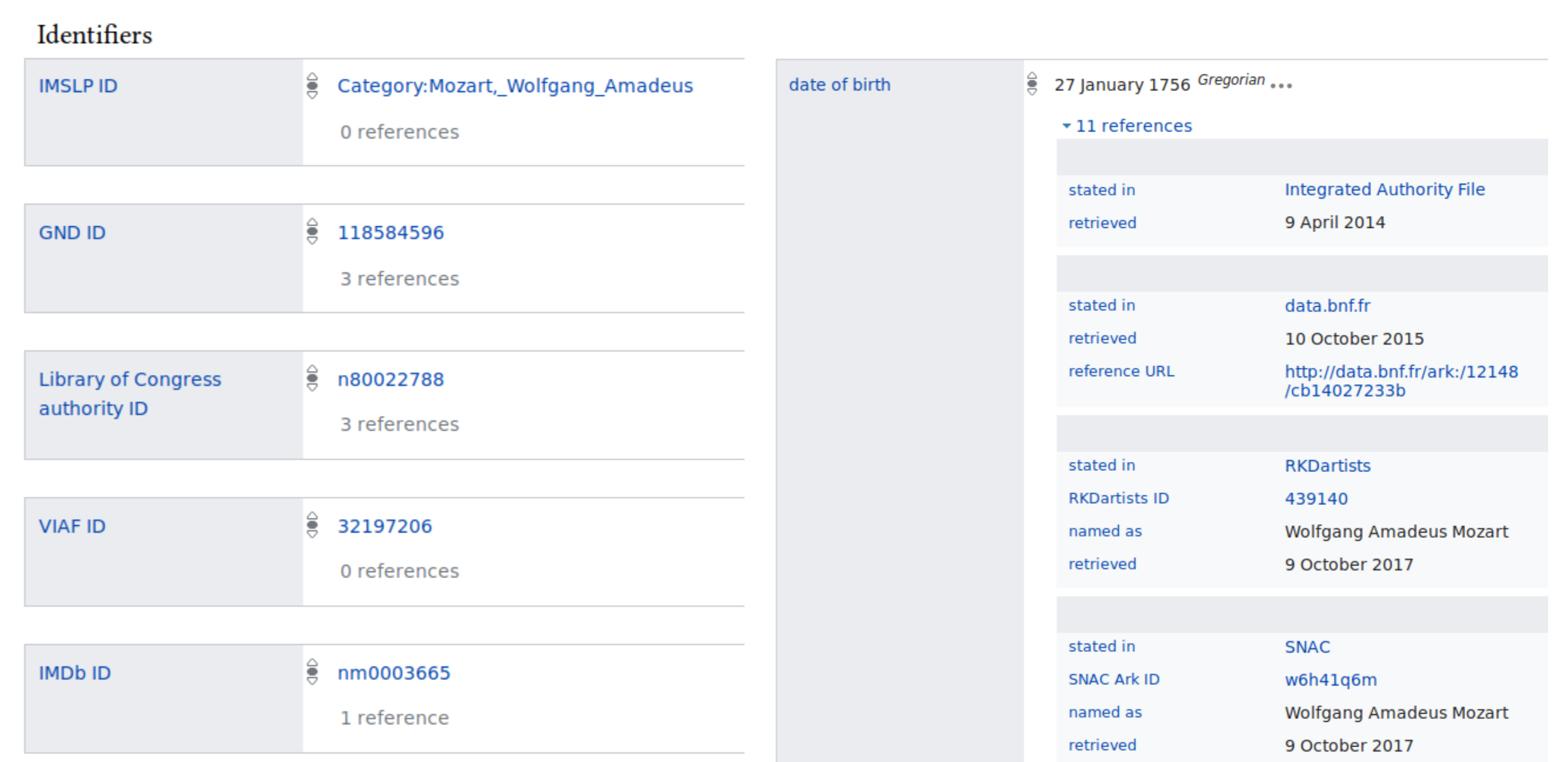


Figure: Identifiers on external sources and References

References

- ▶ John Samuel. Collaborative approach to developing a multilingual ontology: A case study of wikidata. In Emmanouel Garoufallou, Sirje Virkus, Rania Siatra, and Damiana Koutsomihia, editors, *Metadata and Semantic Research*, pages 167–172, Cham, 2017. Springer International Publishing.
- ▶ John Samuel. Analyzing and visualizing translation patterns of wikidata properties. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings*, pages 128–134, 2018.