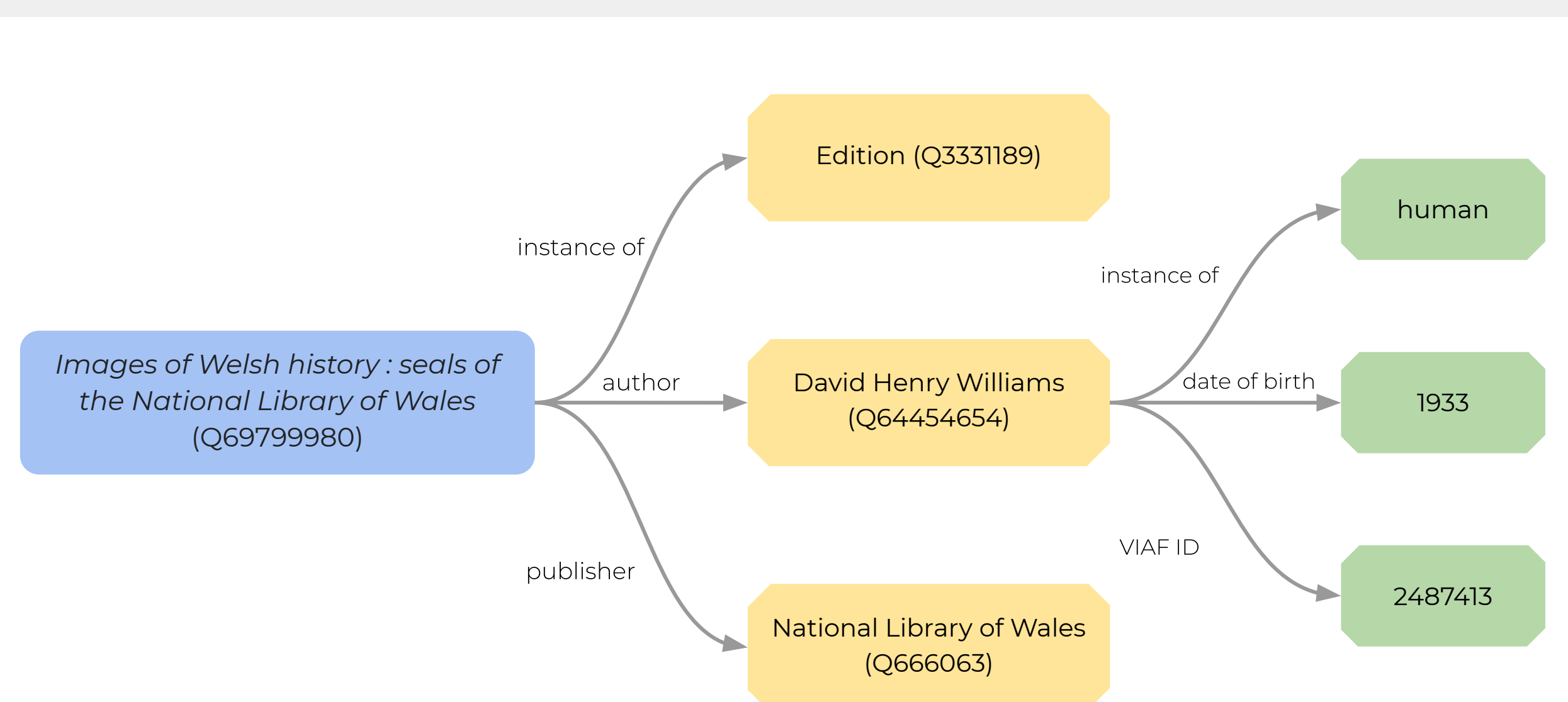


Wikidata: a Linked Open Data Platform open to everyone

Wikidata (wikidata.org) is a platform for Linked Open Data. It's one of the sister projects of Wikipedia. Just like Wikipedia, Wikidata can be edited by anyone, and the content is publicly accessible under an open license (CC0) – meaning it can be freely re-used! Wikidata editors are volunteers, and have imported many open data datasets.



Wikidata contains over **86 million data items**. Thanks to its flexible structure, it encompasses all domains of knowledge: people, geographical objects, historical events, books and much more, can be described in Wikidata. The volunteer community works together to develop modelling standards and best practices for each domain. A dedicated community project for books (wikidata.org/wiki/Wikidata:WikiProject_Books) has developed a data model based on FRBR and its work/edition distinction.



The data on Wikidata is structured and linked, making it an excellent platform to store information about books and their authors.

Open data at the National Library of Sweden

Wikimedia Sverige is a non-profit association based in Sweden, working towards making knowledge freely accessible to as many people as possible. We have an ongoing collaboration with the **National Library of Sweden**, where we work on the Wikimedia platforms (Wikidata and Wikipedia) with the open data released by the library. We also support the library in its endeavors to become an active and knowledgeable participant in the open knowledge movement.

In 2018, the National Library of Sweden became the first national library in the world to transition to linked data in its catalog. The new **Libris** – a national union catalog containing data from over 500 libraries in Sweden – is based on **Bibframe 2.0**, making the data easier to understand, analyze and re-use. Most importantly, it makes Swedish bibliographic data more compatible with the rest of the web, including free and open platforms such as Wikipedia and Wikidata.

The Swedish National Bibliography contains metadata of over 700,000 books, periodicals and other materials published in Sweden. As part of its open data strategy, the National Library of Sweden has released this data under an open license, **CC0**. The same applies to the Libris Authority File. The CC0 marking is the most permissive Creative Commons license: it means the data can be re-used by anyone, without having to credit the source. Removing the attribution requirement is important in order to be able to combine and visualize multiple datasets when crediting every single source becomes an impossibility. This is the data we have been working with.

Wikidata and bibliographic metadata around the world

Libraries and other cultural heritage institutions around the world are implementing Wikidata in their work.

- **Wikibase**, the software behind Wikidata, has been evaluated by the **German National Library** as a possible platform for the Integrated Authority File (GND).¹
- The **Library of Congress** has added Wikidata identifiers to over one million entries in its authority file.²
- The **Association of Research Libraries'** white paper on Wikidata gives more examples of how Wikidata and Linked Open Data can benefit cultural heritage organizations and scholarly communication.³
- **WikiCite** (<https://meta.wikimedia.org/wiki/WikiCite>) is an international initiative to develop open citations and linked bibliographic data to serve free knowledge, using Wikidata and Wikibase as its platform.

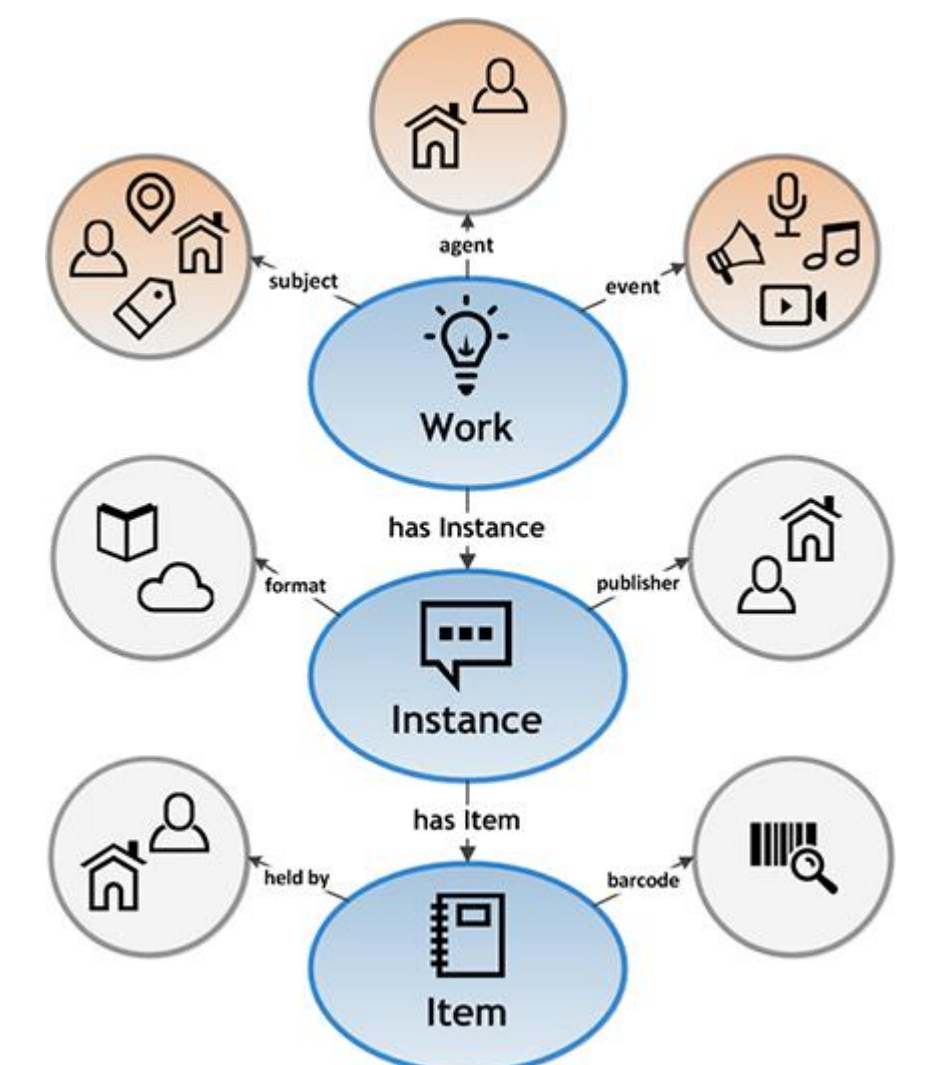
Our work with the National Library of Sweden is not unique; it is part of a global movement to make bibliographic data more robust, accessible and useful.

Working with Swedish bibliographic data on the Wikimedia platforms

We have been working on developing tools and workflows to import **bibliographic metadata** from the National Library's catalog to Wikidata. As a pilot project, we created ca. 500 items to represent book editions. We selected the books that are most often used as references on Swedish Wikipedia to make our work more relevant and interesting to the community. We also added Libris authority identifiers to over 60,000 items of authors on Wikidata.

One of the biggest **challenges** of mass-importing bibliographic data to Wikidata is that items representing authors, publishers etc. need to be uniquely identified in order to be linked from the book items. A prerequisite for doing this automatically is that the author's identity in the source dataset is stated using an authority ID and that a Wikidata item containing the same ID can be found – or created, if we can be sure it does not yet exist. We have found that many author statements in Libris are still stated as strings, making it impossible to automatically and correctly match them to Wikidata items. This problem is even more acute for publishers, where no authority IDs are used at all.

We are also exploring how bibliographic data on Wikidata can be continuously **improved** and kept up-to-date. Wikidata is a living platform, and when community members interact with the data we imported, it proves that our work is engaging and relevant. In particular, we are planning to build tools to enable Wikidata contributors to create new items based on Libris data when they are needed. We see this as an important contribution to the community which will enable the Swedish bibliographic commons to grow even after the conclusion of our project.



The Bibframe 2.0 data model, used in the National Library's new catalog, uses three levels of abstraction to organize bibliographic data.

Since data from Wikidata can be displayed on **Wikipedia**, we are investigating how our work can benefit Wikipedia editors by helping them identify and cite reliable sources. We are also examining which books are most often used as sources on Wikipedia. In the future, this will make it possible to link directly to the digitized versions of those books to make fact checking easier, as well as give an indication to the National Library which books could be prioritized in their digitization work.

An integral part of our collaboration is providing **support** and training about the Wikimedia platforms, giving Swedish librarians tools not only to understand the open knowledge ecosystem, but also to contribute with sources and other improvements themselves. That's why we have organized a hackathon for the National Library staff who are directly involved in Libris development, giving them an opportunity to explore and experiment with the broader context in which their work exists.

References

1. Fischer, Barbara and Jens Ohlig. (2020). *Could you wikify an authority file? Wikibase has been evaluated for the Integrated Authority File (GND)*. <https://blog.wikimedia.de/2020/03/04/wikibase-and-gnd/>
2. Ferriter, Meghan. (2019). *Integrating Wikidata at the Library of Congress*. <https://blogs.loc.gov/thesignal/2019/05/integrating-wikidata-at-the-library-of-congress/>
3. Association of Research Libraries Task Force. (2019). *ARL White Paper on Wikidata: Opportunities and Recommendations*. <https://www.arl.org/resources/arl-whitepaper-on-wikidata/>