Periodic editathons as a way to improve data quality: an experiment in Italy

Camillo Carlo Pellizzari di San Girolamo
13/09/2021

Data Quality Days 2021

Main points

- The Italian situation: library catalogues and Wikidata
- Italian monthly editathons: history, scope, structure
- Achievements
- To-do list

Italian library catalogues and authority files

SBN – National Library Service [**P396**]

- Catalogue: https://opac.sbn.it/opacsbn/opac/iccu/free.jsp
- Authority file: https://opac.sbn.it/opacsbn/opac/iccu/authority.jsp

- 6521 libraries participating, out of 12606 existing libraries (51.7%); here a detailed report
  - public libraries (6632) are mostly in SBN
  - university libraries (1373) and ecclesiastical libraries (1146) are very often outside SBN

- SBN is a VIAF member (ICCU), so it sends its authority data to VIAF, but …

Issues in SBN

- … SBN shows to the public and sends to VIAF only the authority records considered to reach a high quality standard
  - out of about 4M (estimated) personal authority records, only 0.2M are accessible to the public and sent to VIAF

- There is an ongoing effort to obtain access to a higher number of authority records
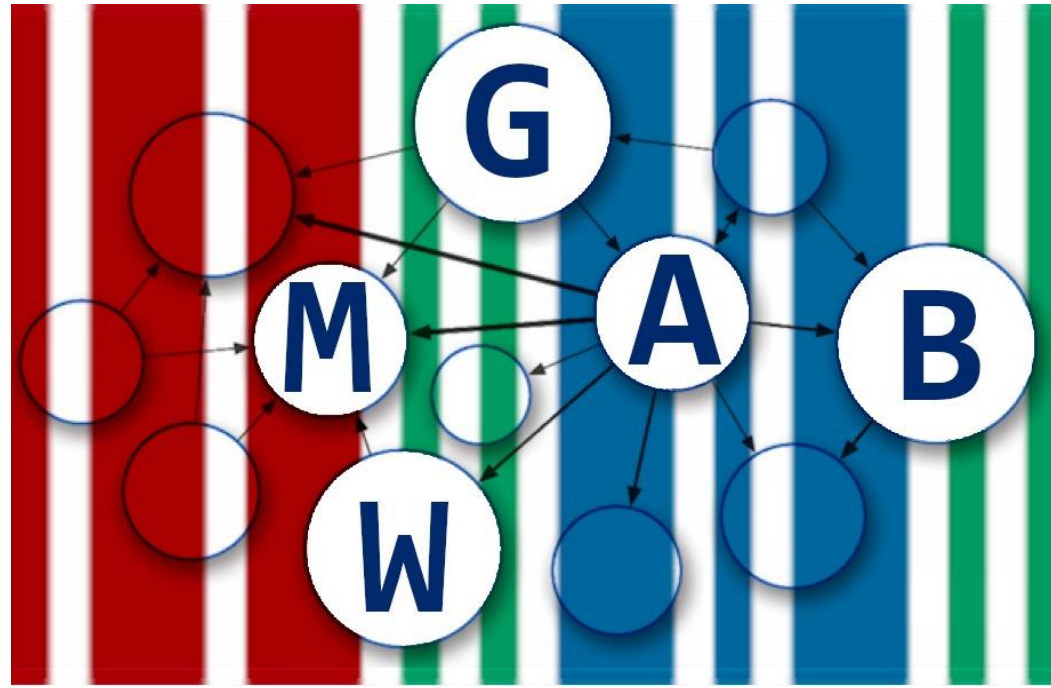
BAV – Vatican Apostolic Library [**P8034**]

- Catalogue: https://opac.vatlib.it/

- Authority file: https://opac.vatlib.it/auth/

- BAV is a VIAF member (BAV) and sends its authority data to VIAF, about 0.3M personal authority records

Some recent collective catalogues …

- BeWeB (https://beweb.chiesacattolica.it/) [**P7796**]: a collective catalogue of many ecclesiastical libraries, with few thousands personal authority records

- SHARE Catalogue (https://catalogo.share-cat.unina.it/sharecat/clusters) [**P3987**]: a collective catalogue of the libraries of 9 universities in southern Italy, with 0.6M personal authority records

- CoBiS LOD Portal (https://dati.cobis.to.it/) [**P7865**]: a collective catalogue of 12 libraries from around Turin, with 0.2M author records

… and a lot of one-library catalogues

- 11 library catalogues have had their authority file connected to Wikidata in the last few years
  - these catalogues help a lot in covering authors which are absent in foreign catalogues and are present in the non-accessible part of SBN
  - some of these middle-sized libraries proved to be more interested to Wikidata than SBN, so most Wikidata activities try to involve them

- For a complete list, see the first section of https://www.wikidata.org/wiki/Template:Italy_properties

Gruppo Wikidata per Musei, Archivi e Biblioteche (GWMAB)

Wikidata Group for Museums, Archives and Libraries

Purposes

- Wikidata page: https://www.wikidata.org/wiki/WD:GWMAB

1. involve people from Wikidata and from Italian cultural institutions

2. present Wikidata to in Italian to operators of the cultural institutions, with focus on LOD

3. organise courses and events regarding Wikidata in Italian

4. persuade Italian institutions both to contribute data to Wikidata and to use data from Wikidata in their websites

Activities

- Basic (online) courses about Wikidata in Italian for people from cultural institutions
- Presentations of projects already realised in Italy on Wikidata
- Conferences and meetings regarding Wikidata and cultural institutions
- Monthly editathons on Wikidata

  - The recordings of most activities are published on YouTube
  - GWMAB has a mailing list, social accounts (Facebook, Twitter, Instagram) and a group on Telegram

GWMAB monthly editathons

https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Editathon_mensile

History and scope

- First edition in March 2021

- Seventh edition … tomorrow!

- Periodicity: each month, second Tuesday

- Expected public: Wikidata users (both expert and not-so-expert), mainly from Italy; librarians from Italy

- Edited items: persons (P31=Q5) whose works are present in the catalogue of at least one Italian library

Structure

- Example edition: https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Editathon_mensile/2021/03

- Each editathon focuses on one library catalogue

Main parts of the page:

- Short instructions, with link to longer version

- Links to places where to discuss strange issues in items

- Lists of items to be edited

Some more details – new users

- The instructions (https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Editathon_mensile#Preparazione) guide the user in the settings of Preferences and common.js and give advices about fundamental properties that persons (Q5) items should contain

- Users can ask questions: on Wikidata; on Telegram; in an online meeting at 17 CEST (this meeting has the main purpose of solving doubts of new users about some aspects of Wikidata)

Some more details – queries

- The queries, stored in a table using {{SPARQL Inline}}, can be opened directly showing their results (easier for new users)

- A few items are present in more than one query, so that the possibility of overlap are very low

- The queries focus on four aspects to be improved:
  - missing VIAF (and other identifiers)
  - missing birth/death dates
  - missing occupation
  - missing Italian label

Some more details – results

- Results are monitored in two ways:
  - through Outreach Dashboard (but many users forget to sign in; so metrics are not reliable)
  - through the decrease of the number of results for each query: this number is written down the evening before the editathon, the morning after the editathon and at the end of the month

- The queries decreased of 273 results on average (March-August) the day after editathons

Achievements – numbers

- 23 users took part in at least one of the 6 editions
  - 9 GWMAB core members
    - 7 librarians, 1 library scientist, 1 Wikidata user (me)
  - 3 Wikidata users
  - 11 librarians
    - 8 of them belonging to the libraries object of editathons
- decrease of 1640 results in the queries

Positive effects

- Focus attention on data quality as a central issue

- Involve librarians on Wikidata with the incentive of editing items which are crucial for their own library catalogue

- Write down queries which could well be used also after the editathon

To-do list

- Organise these editathons not only online, but also in person
- Involve also foreign libraries, focusing on intersections between their catalogues and Italian libraries' ones (see [tomorrow edition](#))
- Re-involve librarians, after their library's edition, for next editions, preparing them queries for intersections with their library; encourage them to continue using queries also after editathons
- Try (again) to involve more Wikidata users, and Wikipedia users
- Introduce queries for fundamental statements (e.g. dates of birth/death, pseudonyms) completely lacking references or being referenced only from Wikipedia

Final thoughts about most difficult issues

- Libraries (not only in Italy) have few librarians, overworked and often not well educated about semantic web and LOD

- Libraries often say that they would be more interested in Wikidata and LOD if they had efficient ways to display data from Wikidata in their catalogues = ILS, especially not-open-source ones, still to be improved (but browser extensions as Entity Explosion are a good starting point)