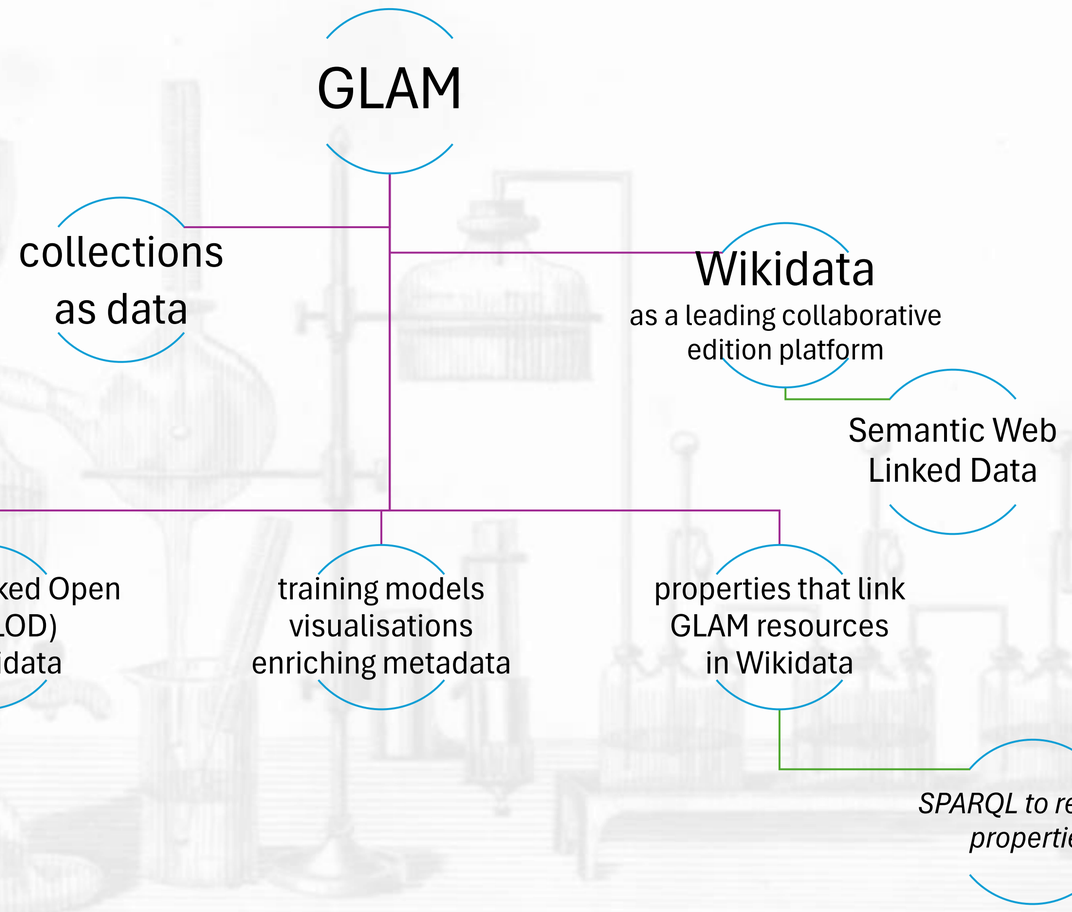
TPDL 2024

A SYSTEMATIC REVIEW OF WIKIDATA IN GLAM INSTITUTIONS: A LABS APPROACH

GALLERIES, LIBRARIES, ARCHIVES AND MUSEUMS LABS | GLAMLABS.IO

Objectives

- To provide a systematic review of Wikidata use in GLAM institutions
- To assess the benefits and limitations of using Wikidata in GLAM projects



Examples of GLAM Labs initiatives related to Wikidata

- Data Foundry at the National Library of Scotland | data.nls.uk
- Rijksmuseum in the Netherlands | data.rijksmuseum.nl
- LOD Infrastructure for Digital Humanities in Finland | seco.cs.aalto.fi/projects/lodi4dh
- National Library of Spain | datos.bne.es
- Wikidata projects of National Library of the Netherlands | www.wikidata.org/wiki/Wikidata:GLAM/Koninklijke_Bibliotheek_Nederland/CurrentProjects

Research questions and findings

1 How is Wikidata described in the current GLAM literature?

Group 1	Group 2	Group 3	Group 4
properties schema	free and structured database	collaborative knowledge base	tool for new working methods
large ontology	structured data source	knowledge base	collaborative and global metadata source
semantic model	central data management platform	knowledge graph	environment for data storage, curation & extraction
ontological model	freely available hosted platform	Web-based knowledge graph	tool of bibliographic interest in the Semantic Web

2 What types of experimentation with Wikidata are emerging in the GLAM domain?

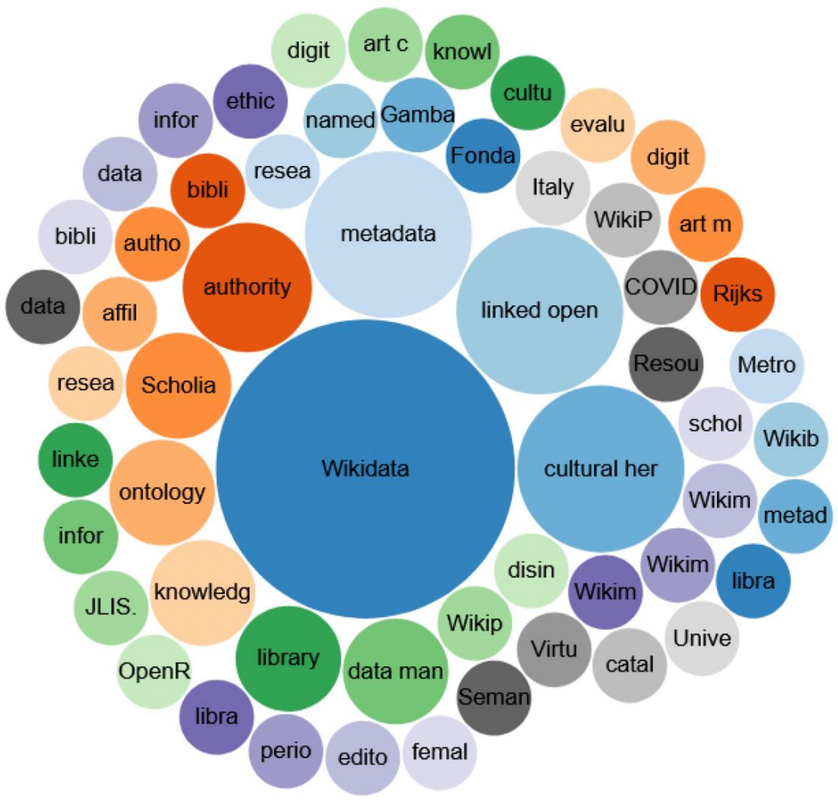
Each of the topics have been categorised using the **TaDiRAH taxonomy** https://tadirah.info/

- Data curation and publication
- Data extraction and visualisation
- Data enrichment
- Data analysis
- Data mapping
- Data quality

3 What are the challenges associated with Wikidata in GLAM?

- Imbalanced data coverage and quality in Wikidata
- Editing exercise is limited due to the lack of reliable citations
- Lack of information in Wikidata may affect reconciliation services
- Historical injustices and imbalances in information representation
- Quality of use requires improvement regarding authority control and the adoption of advanced metadata vocabularies, including provenance information
- Since Wikidata is not originally designed as a bibliographic tool:
  i) quality and completeness of bibliographic data are usually high, but not certain and
  ii) current ontology models for cataloguing rely on metadata structures involving three or more tiers that are not in Wikidata

Conclusions



Bubble chart obtained from the SPARQL endpoint of Wikidata describing the main subjects covered by the projects. The size of the bubbles indicates the relevancy of the keywords in the selection of works.

- Summary of 38 research articles on Wikidata projects published 2020-2023
  (Source: Web of Science and Scopus online research databases)
- Publicly available dataset available in a GitHub repository
- Datasets extracted from Wikidata and a README file describing the project and extraction process
- Results show Wikidata has been used for different purposes in GLAMs
- Wikidata reuse, an important FAIR principle, can be better integrated into the practices of GLAMs

Next steps

- Expanding the search strategy and integrating additional projects and approaches which can be found on Wikimedia pages
- Incorporating projects provided by other repositories and conferences
- Increasing the number of articles by adjusting the time period

github.com/hibernator11/wikidata-review

Gustavo Candela[1], Mirjam Cuper[2], Olga Holownia[3], Nele Gabriëls[4], Milena Dobreva[5], and Mahendra Mahey[6]
[1]University of Alicante, [2]KB, National Library of the Netherlands, [3]International Internet Preservation Consortium, [4]KU Leuven Libraries, [5]University of Strathclyde, [6]Tallinn University

✉ Gustavo Candela | gcandela@ua.es