Linking Biodiversity Literature and Data

Persistent Identifier Exchange Between BHL and Wikidata

Mike Lichtenberg
Biodiversity Heritage Library
United States

LIVING DATA 2025

October 22, 2025

BHL
Biodiversity Heritage Library

The Biodiversity Heritage Library

- An open access digital library for biodiversity literature and archives
- Global consortium working to digitize library collections and make them freely available
- A key component of the core infrastructure for biodiversity research
- Services, data exports, and APIs allow researchers to download and reuse content

Biodiversity Heritage Library

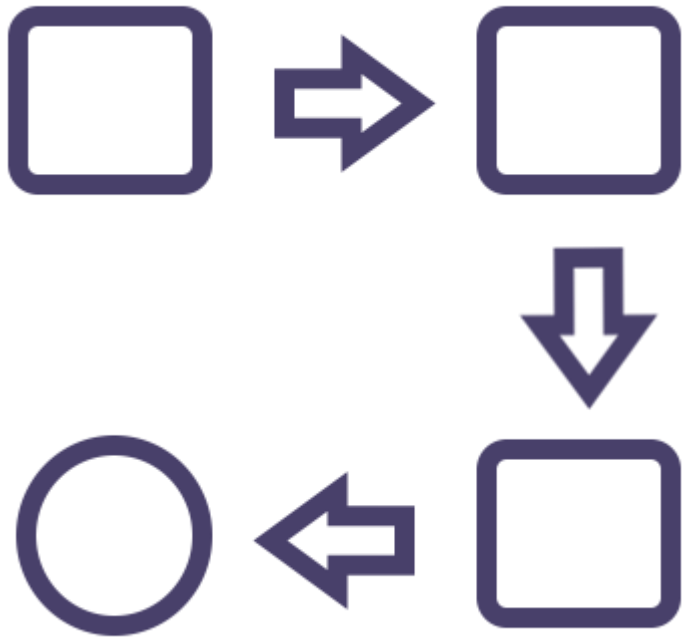
Wikidata

- Free and open knowledge base that can be read and edited by both humans and machines
- Central storage for the structured data of its Wikimedia sister projects
- Content is freely available, exportable in standard formats, and linkable to other open data sets
- A hub for persistent identifiers (PIDs) for domains such as publications, authors, and scientific names

Why Exchange Identifiers?

- Build stronger links between BHL and other biodiversity platforms
- Enhance discoverability of biodiversity literature
- Make curation and corrections easier

The Workflow



Source: Mike Lichtenberg

1. Use SPARQL queries to extract identifiers from Wikidata

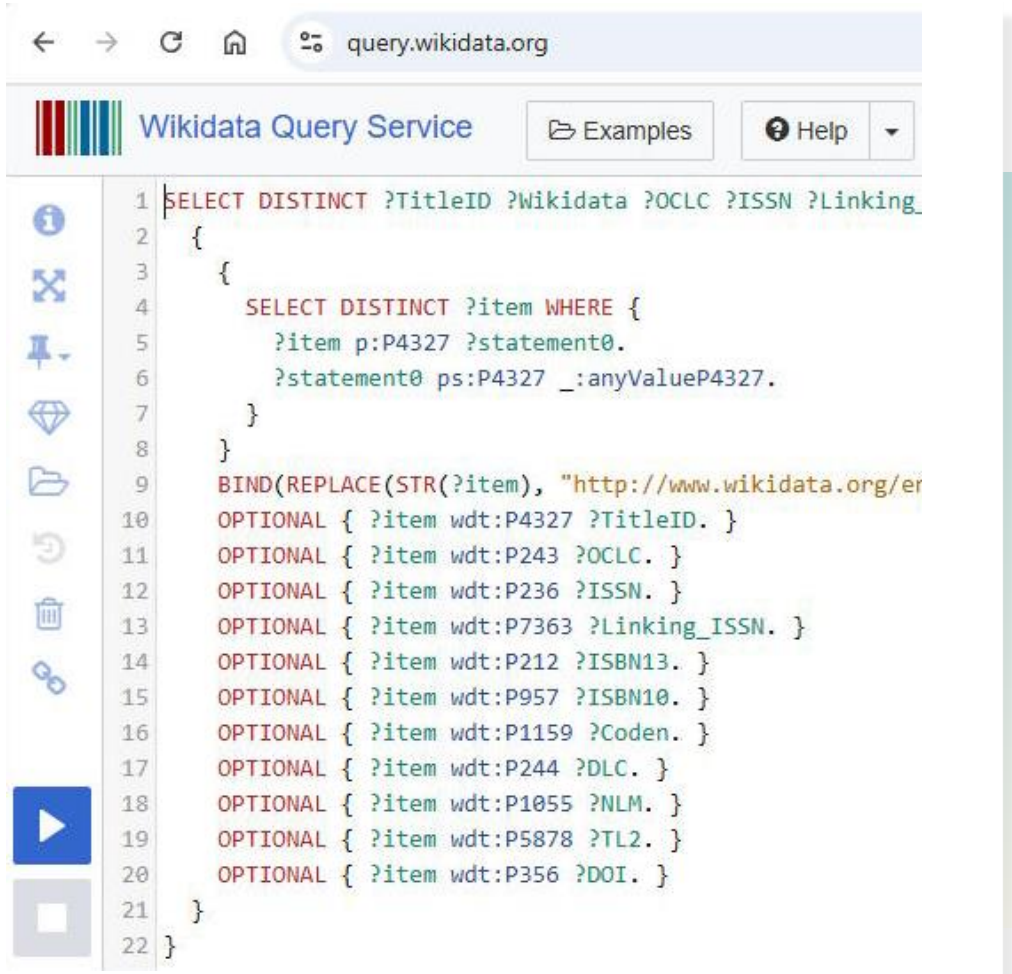
2. Import data into BHL

3. Extract from BHL lists of identifiers requiring curation
   - Identifiers missing from Wikidata
   - Identifiers with irregularities

4. Use Wikidata tools to examine data extracted from BHL and apply changes to Wikidata

Querying Wikidata With SPARQL

- SPARQL is a query language for metadata stored in the Resource Description Framework (RDF) format
- SPARQL is similar to SQL
- The Wikidata Query Service is a tool for writing and executing SPARQL queries, exporting results, and generating code
  - Includes templates and tutorials
- Use federated queries to include results from all of Wikidata



Querying Wikidata With SPARQL: An Example Query

```
SELECT DISTINCT ?TitleID ?Wikidata ?OCLC ?DOI
WHERE {
 {
 {
  SELECT DISTINCT ?item WHERE {
    ?item p:P4327 ?statement0.
    ?statement0 ps:P4327 _:anyValueP4327.
   }
 }
  BIND(REPLACE(STR(?item),
"http://www.wikidata.org/entity/", "") AS
?Wikidata )
  OPTIONAL { ?item wdt:P4327 ?TitleID. }
  OPTIONAL { ?item wdt:P243 ?OCLC. }
  OPTIONAL { ?item wdt:P356 ?DOI. }
 }
UNION
 { SERVICE wdsubgraph:scholarly_articles {
 {
  SELECT DISTINCT ?item WHERE {
    ?item p:P4327 ?statement0.
    ?statement0 ps:P4327 _:anyValueP43272.
   }
 }
  BIND(REPLACE(STR(?item),
"http://www.wikidata.org/entity/", "") AS
?Wikidata )
  OPTIONAL { ?item wdt:P4327 ?TitleID. }
  OPTIONAL { ?item wdt:P243 ?OCLC. }
  OPTIONAL { ?item wdt:P356 ?DOI. }
 }
 }
}
```

1
2
3
4

From Wikidata to BHL

TitleID	Wikidata	OCLC	DOI
29753	Q51381722	31896870	10.5962/BHL.TITLE.29753
110114	Q51386632		10.5962/BHL.TITLE.110114
169133	Q97683524	1290604141	10.1206/0003-0082(2006)3539[1:RCIBAN]2.0.CO;2
169133	Q97683524	442549156	10.1206/0003-0082(2006)3539[1:RCIBAN]2.0.CO;2

Wikidata Query Results

- BHL import process is fully automated, with no human curation of data
  1. Identifiers are compared to existing BHL data
  2. Any identifiers not already in BHL are added

- Others creating similar processes may prefer a different workflow

From BHL To Wikidata

- BHL does not just take from Wikidata; it also contributes back
- After data is added to BHL, a downloadable report is produced
- Report includes
  - BHL data missing from Wikidata
  - Data requiring curation
- Wikimedians make use of the report and Wikidata tools to curate the data and apply changes to Wikidata
- Feedback loop with BHL staff for corrections

Benefits

- Repeatable, adaptable workflow
- 65000 identifiers added to BHL and 8000 to Wikidata
  - Improved discoverability
  - Improved trust in the metadata
  - All identifiers available via Wikidata SPARQL queries and BHL APIs/Exports
- Discoverable data quality issues

Vultur gryphus
https://biodiversitylibrary.org/page/29568719

Supporting Information

Presentation Slides and Code

- **https://github.com/mlichtenberg/LivingData2025**

11-Minute Introduction to SPARQL

- https://youtu.be/FvGndkpa4K0?si=wLus8Qk1MxF5ycHQ

Wikidata Query Service

- https://query.wikidata.org

Wikidata Property Lookup

- https://www.wikidata.org/wiki/Wikidata:List_of_properties

Background Information: Wikidata Graph Split

- https://meta.wikimedia.org/wiki/WikiCite/WDQS_graph_split

BHL Wikidata Harvest Source Code

- https://github.com/gbhl/bhl-us/tree/master/WDHarvest

Thank you!

Please Support BHL!