

# WIKIDATA DAYS AND WIKICITE

BERN '25 **WIKIDATA  
DAYS** {} wikicite

 **wikicite**  
BERN '25

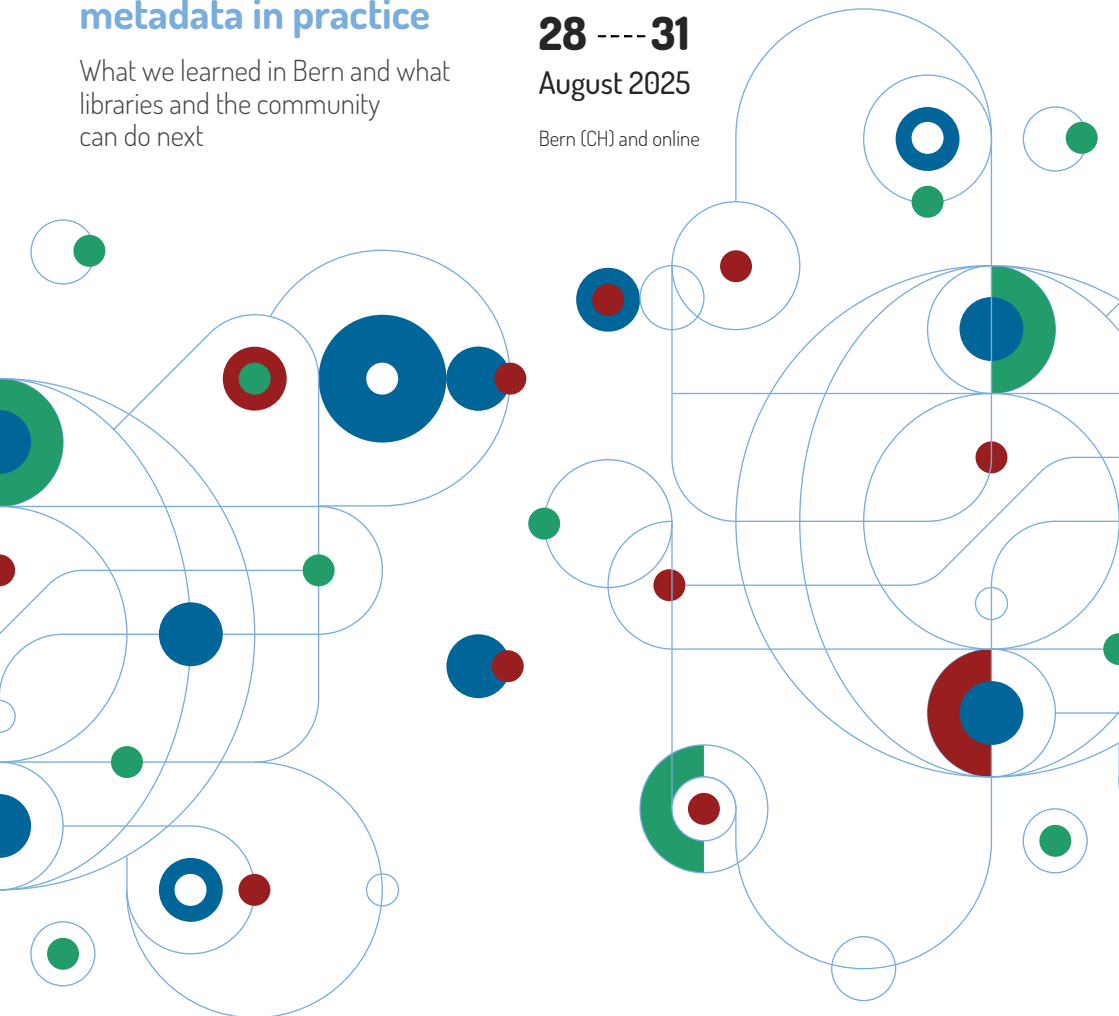
## Open bibliographic metadata in practice

What we learned in Bern and what  
libraries and the community  
can do next

**28 ---- 31**

**August 2025**

Bern (CH) and online





WIKIDATA DAYS  
AND WIKICITE

2025

Imagine a world where  
libraries, researchers and  
online communities all share  
a common language for  
citations and books.

... One language that's free,  
reliable and easy to build on.

# THE CONTEXT IN BRIEF



## WikiCite

WikiCite is a global community initiative launched in 2016 to build and maintain open, collaborative bibliographic and citation data integrated in Wikidata and supported by free technologies such as Wikibase.

Its mission is to:

- Improve the **verifiability** of knowledge across Wikimedia projects
- Enable libraries and research institutions to **reuse and enrich bibliographic metadata openly**
- Build **shared infrastructure** based on free and open technologies, including Wikibase

## Wikidata Days

Wikidata Days are community-driven events focused on practical use cases, peer learning and institutional engagement around Wikidata which emphasize:

- Applied research
- Library workflows
- Hands-on exchange between institutions and volunteers

# The event in Bern

From 28 to 31 August 2025, the communities behind Wikidata and WikiCite convened in Bern and online for three days of structured exchange, technical discussion and strategic reflection on the future of open bibliographic infrastructure.

Organized primarily with Wikimedia CH and the Swiss National Library, with partners including Swiss institutions and thematic user groups (DHWiki / DARIAH-EU, WikiClassics UG).

Discussions concentrated on six interrelated themes: Authority control, Data round-tripping, Academia and Wikimedia projects, Ontologies and controlled vocabularies, Infrastructure and scalability and the Wikibase ecosystem and federation

The hybrid event brought together Wikimedians, librarians, researchers, developers and institutional partners. **More than 100 participants** attended onsite and remotely.



At the time of the event, WikiCite-related scholarly data included approximately **45 million publications** and **312 million citation links**, underscoring both the scale and the infrastructural complexity of the ecosystem.

# A format designed for different audiences

## Day 1



### Wikidata Days Libraries and research

Sessions and case studies oriented to libraries and research, with practical entry points for University libraries

## Day 2



### WikiCite Institutions and infrastructure

Wikibase federation, governance, technical scvaling and partner perspectives

## Day 3



### WikiCite Community sessions and do-a-thons

Participatory discussions, poster sessions, tool demos and hands-on work

## Lessons Learned about the format

- Hybrid works best with a clear “spine”: one main room and strong online facilitation
- Some planned tutorials were dropped: the strongest value emerged from theme clustering and do-a-thons.



## Closing reflections

WikiCite 2025 demonstrated that open bibliographic infrastructure has reached a stage of structural maturity. The questions are no longer experimental but architectural and strategic.

Sustaining this ecosystem requires coordinated stewardship across institutions, community contributors and technical maintainers.

Open scholarly metadata is a **shared responsibility** and a **shared opportunity**.



### Looking ahead

We plan to continue with a recurring format focused on “Wikidata and Research”, building a stable bridge between Wikimedia communities, libraries and academic stakeholders.

## Where to start

[Program page](#)[View on-line](#) [Recordings and slides](#)[View on-line](#) [Community coordination](#)[View on-line](#)

# WHY THIS MATTERS



## Libraries, research, Wikimedia: one shared problem

Whenever you read a Wikipedia article, consult a University publication, or search for a book in a library catalog, you are depending on citations and metadata — the structured information that tells you who created a work, when it was published, where it appeared and how it connects to other research.

If this information is incomplete, inconsistent, or locked behind paywalls, knowledge becomes harder to verify and reuse.

Libraries, research institutions and Wikimedia projects, all depend on bibliographic metadata and citations to make knowledge findable,



verifiable and reusable. However, critical components of this infrastructure remain fragmented across catalog systems, institutional repositories, proprietary databases and heterogeneous data models.

**WikiCite and Wikidata provide a shared, open layer that enables structured bibliographic data to be connected, enriched and reused across systems.**



## The event enable

- **For Wikimedia communities**
  - Better references and structured citation metadata to support verifiability
  - Tools such as Scholia and other reuse workflows built on open data
- **For libraries and institutions**
  - A practical pathway to publish and connect authority data and bibliographic metadata as linked open data
  - Workflows for reconciliation and round-tripping (improvements can travel back to source systems), supported by a global community

# Three strategic insights from 2025



## 1. Authority control as a shared layer

Unique identifiers and systematic disambiguation are foundational across catalogs, repositories and knowledge graphs. The transition from label-based description to identifier-based modeling is now broadly recognized as essential for interoperability.



## 2. Round-tripping over one-way export

Data exchange must be bidirectional. Corrections and improvements made within Wikidata should flow back to source systems. Sustainable collaboration requires feedback loops rather than static data dumps.



## 3. Infrastructure as a strategic priority

The scale of scholarly content has driven architectural transformation (e.g., WDQS graph split). Infrastructure is no longer a background concern; it is central to sustainability and performance.

***WikiCite is sustained as much by its community as by its technology.***



# From centralization to federation

The **scholarly graph split** and the expansion of specialized Wikibase instances mark a transition from centralized growth to federated infrastructure.

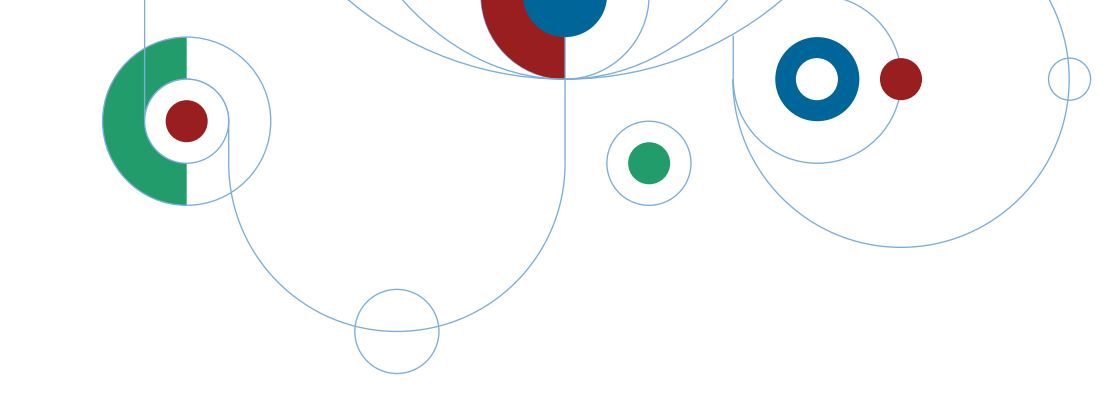
Federated architectures are increasingly recognized as the operational norm rather than a theoretical preference.



In May 2025, Wikidata split its query service into a **main graph** and a **scholarly graph** to manage data volume. The scholarly graph, accessed at a dedicated SPARQL endpoint, contains data for items identified as academic, such as papers, theses and studies. The split is based on specific properties and QIDs to separate research data.

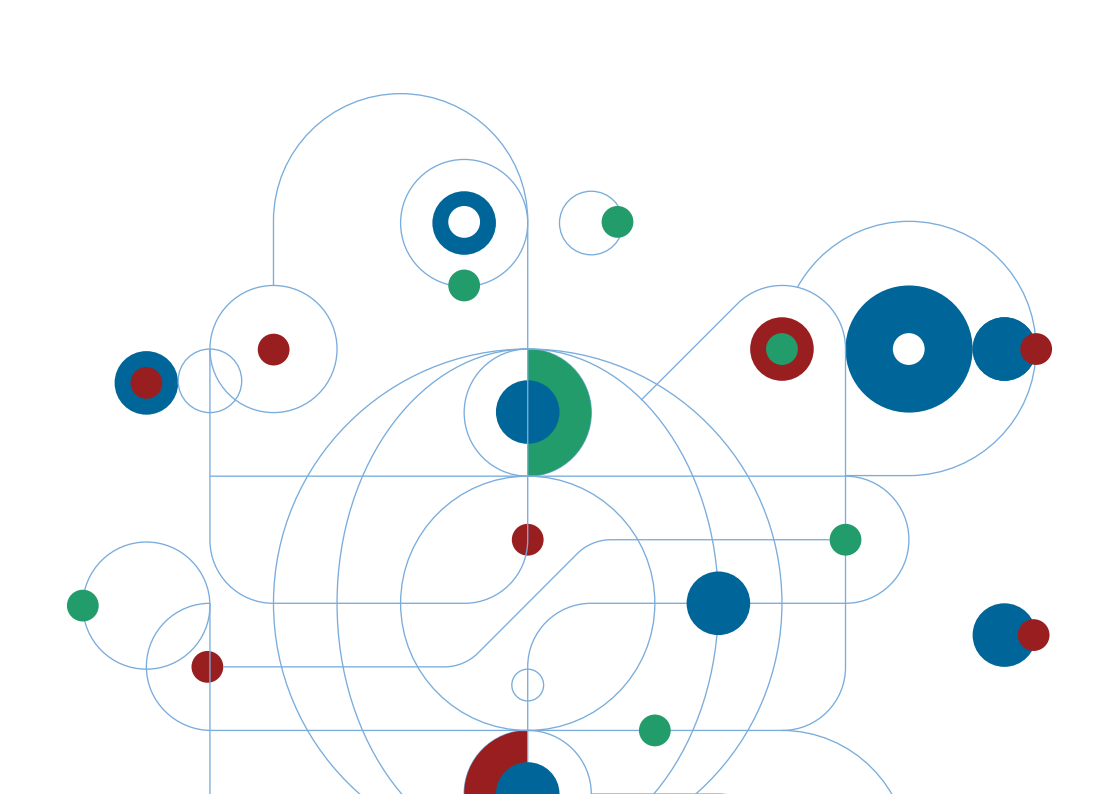
## What this means in practice

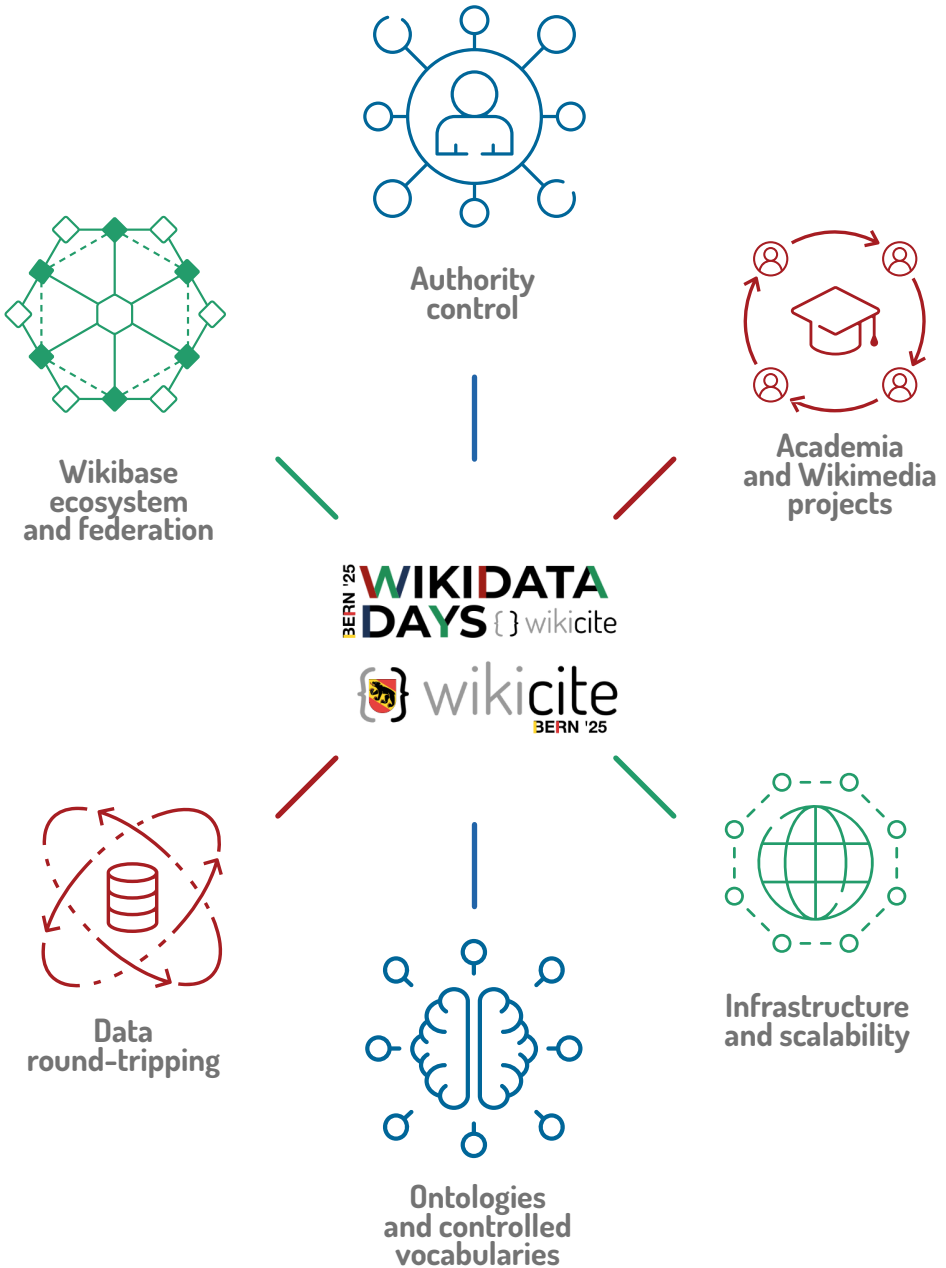
- Tooling and documentation need a “post-split default”
- Institutions can host specialized data while remaining interoperable
- Governance and sustainability become central topics, not afterthoughts



# THE 6 THEMES

Across the Wikidata Days and the WikiCite days, talks and discussions converged on recurring themes.





1

# Authority control



Authority control uses stable identifiers to link people, works, subjects, and institutions, so they can be clearly identified and shared across systems.

## •→ Key takeaways

- Authority control began in traditional library catalogs and has expanded into large linked data systems. Identifiers are now the main way different systems can work together.
- Large national authority graphs demonstrate how coordinated institutional collaboration can scale effectively
- Many libraries are now testing Wikibase to move from older cataloging systems, such as Machine-Readable Cataloging (MARC), to linked data models based on the Resource Description Framework (RDF).

## •→ Spotlight

Connecting the Integrated Authority File (GND), Wikidata, and national authority projects shows how different systems can work together by using identifiers and shared data models.

## •→ What the community can do next

- Improve author and work items with reliable identifiers; document best practices for reconciliation and disambiguation
- Use tools like OpenRefine, Mix'n'match, and bots to create and manage matching workflows, with clear checks for quality.

## •→ What institutions can do next

- Publish/align authority data using stable IDs; start with a pilot subset (e.g., one collection or one institution unit)
- Establish a feedback channel for corrections and engage with WikiCite maintainers

## •→ Featured talks

Authority from a Librarian's Perspective in Contrast with Wikidata

[View on-line](#) 

Linking GND and Wikidata

[View on-line](#) 

Czech Authority Files Meet Wikidata

[View on-line](#) 



## **2** Data round -tripping

Round-tripping refers to the continuous, bidirectional synchronization of data between institutional systems and Wikidata.

## •→ Key takeaways

Sustainable collaboration depends on structured workflows:

- Identify issue → flag in Wikidata → notify source → confirm correction
- Tools such as “deprecated rank” help mark known wrong data as problematic until the original source corrects it
- National “slow data” approaches mean improving data carefully over time through manual review, outreach, and training, in order to build trust and long-term collaboration.t

## •→ Spotlight

Without feedback loops, quality problems scale faster than volunteer capacity — especially when imports contain systemic errors

## •→ What the community can do next

- Maintain a shared “round-tripping playbook” and track cases (what was fixed, where, when).
- Create small task forces around specific source systems (repositories, national catalogs, etc.).

## •→ What institutions can do next

- Nominate a contact person for corrections; define a lightweight issue intake channel.
- Start with high-impact entities (researchers, theses, institutional publications) and scale gradually.

## •→ Featured talks

Data round-tripping for “Swiss WikiCite”

[View on-line](#) 

Homonyms clean-up for “Swiss WikiCite”

[View on-line](#) 

WikiCite in Switzerland and Italy — New Perspectives

[View on-line](#) 



# **3 Academia and Wikimedia projects**

Wikidata is increasingly used to model scholarly profiles, research networks and non-traditional research outputs, enabling new discovery and analysis workflows.

## •→ Key takeaways

- Scholarly profiling is shifting from “lists of publications” to knowledge graphs of people, outputs, affiliations and relationships
- Proposals like the Wikimedia Science Hub aim to support global science editors and connect practices across domains
- Historical sources and Wikisource-linked workflows show how structured metadata can unlock large corpora (bots, registers, automated indexing)

## •→ Spotlight

Example spotlights include large-scale thesis metadata and modeling of retracted research signals to improve responsible reuse

## •→ What the community can do next

- Improve modeling guidelines for theses, datasets, software and other outputs.
- Strengthen quality for non-English scholarship and books/editions

## •→ What institutions can do next

- Publish subsets of institutional repositories to Wikidata with identifiers and minimal, high-quality fields
- Co-design use cases (profiling, discovery, catalog enrichment) with communities

## •→ Featured talks

Planning the Wikimedia Science Hub

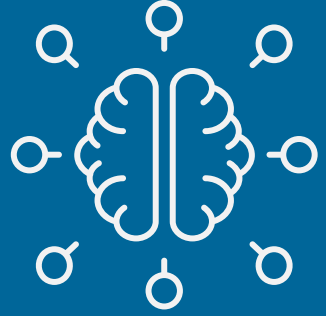
[View on-line](#) 

Linking New Zealand Theses to the World Through Wikidata

[View on-line](#) 

Retracted articles in Wikidata and how to use them

[View on-line](#) 



# 4

## **Ontologies and controlled vocabularies**

Controlled vocabularies and ontologies help use the same meanings across data, but they require careful modeling.

## •→ Key takeaways

- Thesauri and ontologies have different purposes. Wikidata works well for many kinds of entities, but it is more complex to use for broad or general concepts
- Common modeling inconsistencies (e.g., subclass vs. instance confusion) reduce reasoning capacity and reuse
- Domain initiatives show how harmonization pipelines can turn messy sources into referenced, queryable knowledge

## •→ Spotlight

The goal is pragmatic interoperability: decide what belongs in Wikidata and what stays canonical in institutional vocabularies, then maintain mappings over time.

## •→ What the community can do next

- Improve class descriptions and add missing generalization links where needed
- Document “minimal ontology patterns” for library and scholarly data

## •→ What institutions can do next

- Decide where Wikidata is the right layer (linking + multilingual access) vs where institutional vocabularies remain canonical
- Provide mappings (even partial) and maintain them over time

## •→ Featured talks

Using Wikidata for Controlled Vocabularies

[View on-line](#) 

Nuovo Soggettario and Wikidata

[View on-line](#) 

LOTUS initiative and its links to WikiCite

[View on-line](#) 



## **5** Infrastructure and scalability

Wikidata had so much scholarly data that it had to change its technical structure. Academic content was separated so searches and queries could keep working well, and tools now need to adjust.

## •→ Key takeaways

- In May 2025, approximately 45 million scholarly articles were separated into a distinct query graph to reduce timeouts and protect performance for general queries
- Queries may require federation / UNION patterns to reach both graphs.
- Some WDQS-dependent projects (e.g., Scholia) were impacted and require updates and documentation work

## •→ Spotlight

Complementary building blocks include LibraryBase (staging for bibliographic metadata) and the Wikipedia Citations Database (extracts and normalizes citations)

## •→ What the community can do next

- Maintain a shared set of “post-split query patterns” and update tutorials
- Prioritize tool fixes that unblock everyday scholarly workflows

## •→ What institutions can do next

- Allocate time to adapt dashboards and SPARQL queries
- Consider specialized Wikibases for large bibliographic corpora, connected back to Wikidata

## •→ Featured talks

WDQS Graph Split — Overview and Query-a-thon

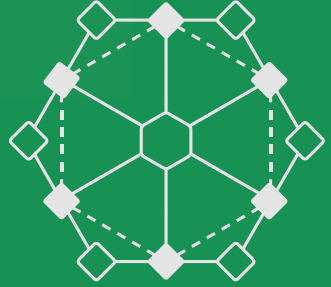
[View on-line](#) 

The State of LibraryBase

[View on-line](#) 

Wikipedia Citations Database

[View on-line](#) 



# 6 Wikibase ecosystem and federation

The future strategy is a federated Wikibase ecosystem: specialized datasets hosted on their own instances, connected back to Wikidata as a hub

## •→ Key takeaways

- More than 1,500 Wikibase instances are currently active across domains
- Federation reduces load on Wikidata and enables institutions to keep stewardship of domain datasets
- Governance models and sustainable maintenance frameworks remain ongoing challenges

## •→ Spotlight

Federation is becoming the default architecture for open scholarly infrastructure and it needs documentation, governance and practical recipes.

## •→ What the community can do next

- Document federation best practices (what works, what breaks, why)
- Build shared expectations for data governance and interoperability

## •→ What institutions can do next

- Start with a pragmatic dataset boundary (what belongs in your instance vs Wikidata)
- Plan for long-term maintenance: staff time, backups, upgrades and community touchpoints

## •→ Featured talks

Federation and data governance in Wikidata and the Wikibase

[View on-line](#) 

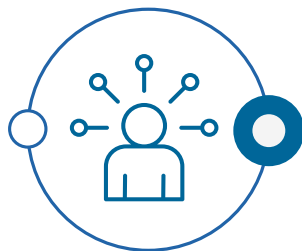
Past, Present and Future of Wikidata and Wikibase

[View on-line](#) 

The Linked Open Data strategy and what it means for your chapter

[View on-line](#) 

# CROSS-CUTTING INSIGHTS



## GLAM and collections as data

GLAM (Galleries Libraries Museums and Archives) institutions increasingly treat collections as computationally reusable data and Wikidata can be a connector layer that improves discoverability and context.

## → Key takeaways

- “Collections as Data” requires clear licensing and access (APIs / exports)
- Projects connecting journals, catalogs and legacy publications show how Wikidata can enrich context and reuse.

## → What the community can do next

- Provide reconciliation support and mapping templates
- Create plain-language documentation for GLAM staff

## → What GLAM institutions can do next

- Confirm rights/licensing on metadata
- Choose a pilot dataset and identifier strategy
- Publish + establish a correction loop

## → Featured talks

Making GLAM resources more accessible and reusable: a FAIR case study

[View on-line](#) 

Linked Open Data awareness in Polish LAM sector — survey results

[View on-line](#) 



## Highlights from community discussions

GLAM institutions increasingly treat collections as computationally reusable data and Wikidata can be a connector layer that improves discoverability and context.

- Citation Watchlist: monitoring source changes and flagging potentially problematic sources
- Modeling and surfacing retracted research to support responsible reuse
- Interest in LLM/RAG workflows with continued focus on verification, provenance and reliable pipelines

### A practical stance on automation

Automation is useful when it improves curation capacity and transparency — but it must strengthen provenance, not weaken it.



# CALL TO ACTION



## For libraries and institutions

- Choose an identifier strategy (ORCID / DOI / national IDs) and apply consistently
- Publish a pilot dataset with minimal, high-quality fields (start small, iterate)
- Create a round-tripping channel: who receives corrections, how they are tracked, when they are applied
- Explore whether a specialized Wikibase instance is appropriate for your domain
- Document licensing and access pathways so data can be reused responsibly



## For the community

- Maintain and document “gold standard” query patterns post-split
- Run focused cleanup drives (homonyms, identifiers, books/editions)
- Improve onboarding materials for institutions (plain language + templates)
- Strengthen validation via constraints, shapes and maintenance WikiProjects

# HOW TO COLLABORATE

## A collaboration loop that works

- 1 **Define a shared subset** - collection, institution unit, repository slice
- 2 **Agree on minimal fields and identifier requirements**
- 3 **Import and reconcile** - tools and mapping
- 4 **Quality pass** - constraints, sampling, issue tracking
- 5 **Round-trip corrections** - back to the source system
- 6 **Publish outcomes** - what improved, what remains open

### Roles

- **Institutions:** data stewardship, corrections, licensing clarity
- **Community:** reconciliation workflows, modeling patterns, QA and documentation
- **Tool builders:** UX, federation support, sustainable pipelines



# CREDITS

## Editorial preparation

- Ilario Valdelli (ORCID: 0009-0009-3732-8273) 

## Contributions and inputs

- Sara De Monaco (ORCID: 0009-0004-0504-5184) 
- Houcemeddine Turki (ORCID: 0000-0003-3492-2014) 
- Alessandro Marchetti (ORCID: 0000-0002-6125-3920) 
- Camillo Carlo Pellizzari di San Girolamo (ORCID: 0000-0003-2699-1693) 

## Event organization

- Wikimedia CH
- Swiss National Library
- Bern University of Applied Sciences

## License and acknowledgements

Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This booklet is a synthesis of discussions and outcomes from Wikidata Days and WikiCite 2025. It does not represent a formal scientific publication but an editorial summary prepared for dissemination purposes.



# ACADEMIC SOURCES

- **FAIR and open multilingual clinical trials in Wikidata and Wikipedia. Research Ideas and Outcomes**  
Rasberry & Mietchen (2021)
- **The invisible citation commons**  
Ayers and Klein (2021)
- **Wikidata and authority control for researchers the cases of Switzerland and Italy**  
Marchetti & Pellizzari di San Girolamo (2024)
- **Environmental impact assessment reports in Wikidata and Wikibase**  
Nielsen et al. (2023)

# OTHER SOURCES

- **WikiCite annual reports 2016-2021**  
[https://commons.wikimedia.org/wiki/Category:WikiCite\\_annual\\_reports](https://commons.wikimedia.org/wiki/Category:WikiCite_annual_reports)
- **WikiCite events and coordination page**  
[https://meta.wikimedia.org/wiki/WikiCite\\_2025](https://meta.wikimedia.org/wiki/WikiCite_2025)
- **Wikidata: Status updates**  
[https://www.wikidata.org/wiki/Wikidata:Status\\_updates](https://www.wikidata.org/wiki/Wikidata:Status_updates)
- **WikiProject Data round-tripping**  
[https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Data\\_round-tripping](https://www.wikidata.org/wiki/Wikidata:WikiProject_Data_round-tripping)

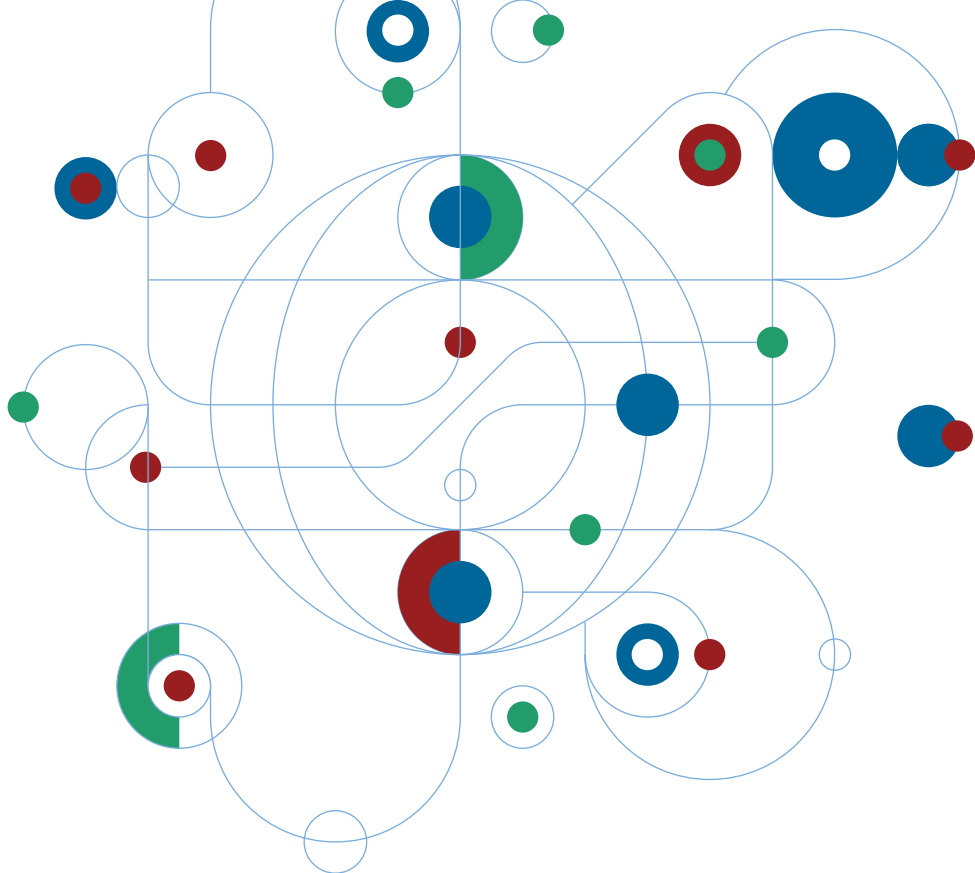
# WIKIDATA DAYS AND WIKICITE 2025

WikiCite is a shared commitment  
to verifiability, multilingual knowledge  
and open scholarly stewardship.

Join the community · Contribute data  
Start a pilot · Share a use case

Program →  
Commons →  
Meta →





BERN '25 **WIKIDATA**  
**DAYS** { } wikicite

{ } wikicite  
BERN '25

## Organizers and Partners



 Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra



**science  
et cité**  




Design: [zerogravita.com](http://zerogravita.com)