



Practical Wikibase Federation

Arthur Smith
@ArthurPSmith

This session is recorded: Please mute your microphone and camera when you're not speaking.



The need for federation

Wikidata has ~100 million items; some datasets many times larger:

- Astronomical objects. GAIA catalog has ~1.7 billion stars
- Legal entities (businesses). Over 200 million in OpenCorporates
- Scholarly articles. OpenAlex has 243 million
- Chemicals/materials. CAS registry has over 200 million
- Genealogy databases - many have 1 billion or more entries

Even if they could fit in Wikidata, some of these might not be considered to meet notability criteria.

May also be incompatible with CC0 license.



Wikicite example

Suppose we had a separate wikibase for scholarly articles. How should it federate with wikidata?

- Wikidata already has over 40 million; identical items should be linked somehow
- Item-valued properties on article items (author, main subject, language, published in, etc.) should either point to the Wikidata entries, or have local entries with their own item-valued properties that may point to Wikidata entries. (authors have employers etc.)
- Wikicite Wikibase UI should display federated items with labels and interlinking
- Query services should be federated also

Where are we on federation?

- SPARQL federation works (but may be slow) - Wikidata Query Service federates with many other SPARQL endpoints
- Wikibase has a “Federated Properties” feature where the properties for the Wikibase instance are copied from Wikidata (in testing?)
- “Two-way federation” between instances is “probably a long way off” - [mw:Wikibase/Federation](#)
- Federation is not mentioned in the 2023 development plan for Wikidata/Wikibase...

Have we lost our way here?

Other examples of federation

- World wide web: interlinking documents with URL's, resolved via DNS etc.
- Linked Data: graphs based on URL's
- Email: SMTP protocol
- OAuth: shared authentication
- Fediverse - Mastodon, Lemmy, Pixelfed etc.: based on ActivityPub protocol

Do we need something as involved as ActivityPub? Maybe simple URL-style interlinking (maybe with caching) will suffice?

Proposal 1: New Data Type

Items have datatype “wikibase-item”. Add a new “federated-item” type.

- Data entry as Q12345@<federated wikibase url>
- JSON value would add a “wikibase” attribute, the federated URL
- UI would fetch entity data from federated URL (possibly caching)
- Simplest possible change to support federation

But:

- Each item-valued property would be duplicated with an equivalent federated-item property

Proposal 2: Modify item datatype

New optional attribute for item datatype: “wikibase” = federated wikibase url

- Item-valued properties would immediately allow federated items
- Data entry as Q12345@<federated wikibase url>
- UI would fetch entity data from federated URL (possibly caching)

Challenge: every application looking at wikibase items needs to check for the new attribute for where to get item data.

Proposal 3: Modify all entity datatypes

Allow optional “wikibase” = federated wikibase url on items and properties; also lexemes and any other supported entities

- Federating properties this way may add some complexity, but avoids duplication
- More consistent treatment of wikibase entities in federated context
- Better to make this change all at once?

Some questions

- Could we switch property datatypes from item to federated item one by one?
- Do we only need wikidata.org federation (to start with)?
 - I.e. expecting other wikibases to federate only back to Wikidata, not outward federation from Wikidata to others or between others
- Should there be a whitelist and/or blacklist for federation urls?
- Does the target wikibase need to know that an item is being used in another wikibase (i.e. subscribe/follow records)?
 - And how to handle “What links here”?
- How to handle redirects?
- How should caching work?
 - “Purge” should presumably re-fetch all remote and local data on a page
- Other questions?

Summary

- Proposal 1: federated item datatype
- Proposal 2: modify item datatype to allow federation
- Proposal 3: allow federation on all (entity) datatypes
- Possible next steps:
 - RFC somewhere? How do we reach consensus?
 - Wikibase phabricator task(s)
 - Code!