

Ontology issues in Wikidata

Everything in neat and tidy boxes? Not quite!

Lydia Pintscher

Q18016466

Portfolio Lead for Wikidata

Using Wikidata as a source of knowledge requires effort

Often too much for small re-users



Ontology issues
are a big problem
for easy re-use of
our data

- Unexpected query results and relations
- Inconsistency between similar concepts
- Even simple inferences are problematic, when connecting information
- Especially problematic for smaller and medium-size reusers

So many different types
of ontology issues!



Conflicting
real-world
models

Conceptual
ambiguity

Structural bugs (cycles,
redundancy, ...)

Upper level
ontology is
messy

Overgeneralisation

Inconsistent
modeling

Semantic drift



We asked reusers



1: Conceptual ambiguity

- not clearly defined what an Item refers to -> conceptual overloading
- example: an embassy as a location and as a diplomatic mission
- unclear what individual statements refer to + semantic drift

1: Conceptual ambiguity

- Playing into this:
 - Wikipedia connections
 - Cross-language concept mismatch
- Splitting could be even worse

2: Inconsistent modelling

- Similar kinds of data are modelled differently
- Happens within one domain as well as across domains

Runner-ups

- Complexity introduced by conflicting real-world models
- Unclassified Items
- Messy upper-level ontology

Some ideas for solutions



Making EntitySchemas more useful

- Define how certain classes should be modelled
- Will be the base for tools and integrations in the future:
 - Automated checking of Items for violations of the associated Schema
 - Forms for creating new Items
 - ...

Making constraint violations queryable

- Some constraint violations are a good indicator of an ontology issue
- Currently hard to find statements violating a constraint in a particular area of interest and work together on fixing them

Making Wiki Projects more visible

- Natural place for coming together to discuss how to model data in a certain domain
- Currently very hidden and not very active and inviting -> new and casual contributors are not aware of them or are confused



Making showcase
Items and
documentation
better and more
discoverable

- Useful tools to guide people towards consistent modeling
- Too hidden to be found by people who would need it the most

Creating automated lists for Wikipedia and co

- Would expose inconsistent modelling to more people and hopefully lead to more people working towards consistency
- Listeria is doing that but not (yet) getting to places where it would really have a huge impact on modeling consistency



Treating ontology-related statements differently in the UI

- Currently no hints that some statements are more impactful than others (e.g. instance of, subclass of, part of, ...)
- Give them separate section and dedicated help popups

Showing the bigger picture

- Often unaware that edits to ontology-related statements have potentially large impact beyond the Item being edited
- How can we help people see this impact?

Fixing common mistakes automatically

- Unintended inconsistencies in the modeling, especially where there are two equally good ways to model something
- Automatically change statements to conform with agreed-upon modeling
- `Template:Autofix` does part of this but needs improvements

Want to dive deeper into it?

-> [Wikidata:Ontology issues prioritization](#)



See you on Wikidata!

Join us at the [Data Modelling Days!](#)

Get in touch:

- lydia.pintscher@wikimedia.de
- [nightrose@mastodon.social](https://mstdn.social/@nightrose)
- [User:Lydia Pintscher \(WMDE\)](#)
- [Q18016466](#)