



Welcome to the Data Modelling Days 2023!

Lydia Pintscher

Portfolio Lead for Wikidata,
Wikimedia Deutschland

Léa Lacroix

Community Engagement &
events consultant,
Wikimedia Deutschland

Daria Ammalainen

UX researcher for Wikidata,
Wikimedia Deutschland

Arian Bozorg

Product Manager for Wikidata,
Wikimedia Deutschland

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



Agenda

- Who's in the room?
- Introduction & how to attend
- What is the event all about?
- Q&A
- Results of the research on ontologies issues
- Q&A
- Coming up next



Hello there!

While Lydia, Daria and Arian are quickly introducing themselves, can you tell us in the chat:

- Where are you joining us from?
- Is this your first time at a Wikidata event?
- What are you particularly interested in?



Introduction & how to attend

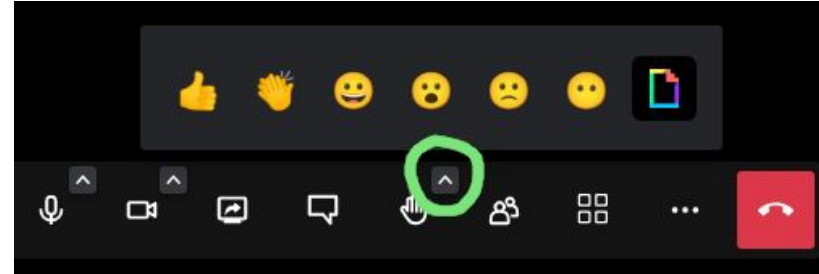
About the event

- Taking place on **November 30 - December 2**
+ additional self-organized sessions on December 3
- Most sessions take place **in this Jitsi room**
- **Program:** [[Wikidata:Events/Data Modelling Days 2023]]
[w.wiki/7iBe](https://www.wikidata.org/wiki/7iBe)
- **Open and free** without registration
...but you can sign up on the participants list: [w.wiki/89vq](https://www.wikidata.org/wiki/89vq)
- Please check the **Code of Conduct:** [w.wiki/AK\\$](https://www.wikidata.org/wiki/AK$)
and make sure you contribute to a safe and welcoming environment
- **Question, issue** during the event?
Feel free to contact me at lea.lacroix_ext@wikimedia.de
or @auregann on Telegram



Jitsi in short

- Menu bar with several options
- Make sure to **mute** your microphone and camera
- Feel free to **unmute** when you want to speak (eg during Q&A)
- Feel free to discuss and **ask questions in the chat**
- Raise hand vs reactions
- **Bugs can happen** (especially if we're many people)
- Works better on Chrome(ium) than Firefox
- In doubt: refresh the page < switch browser < reboot
- There's also a Jitsi **mobile app**



Documentation & outcomes

- Sound and video of most sessions **will be recorded** by a wonderful human *and* we have **collaborative notes** (links in the schedule)
- Please help us **taking notes** during the sessions!
- Slides and videos will be published shortly after the event, links will be added to the schedule (it will probably take a few days, please be patient)
- At the end of each session, we will take a few minutes to note down **outcomes and next steps**



Special notes to the speakers

- Thanks for your participation <3
- Please double-check the **time of your session** in the schedule, and join the Jitsi room at least **10min in advance**
- You can use the **slides template** to prepare your presentation shorturl.at/nBV24
- Please **upload your slides** and other materials to Wikimedia Commons in `[[Category:Data Modelling Days 2023 presentations]]`
- If you don't want your session to be recorded, please let me know asap at lea.lacroix_ext@wikimedia.de



Join the discussions!

- Event talk page: [w.wiki/7ibM](https://www.wikidata.org/wiki/7ibM)
- Jitsi chat during sessions (not logged)
- Hashtag: **#DataModellingDays**
- Telegram: main Wikidata group
- Join the **Wikidata community channels**: [w.wiki/4wAo](https://www.wikidata.org/wiki/4wAo)
 - Project Chat on Wikidata
 - Mailing-list
 - Weekly newsletter
 - Telegram group
 - Mastodon, Twitter, Facebook





What is this event all about?

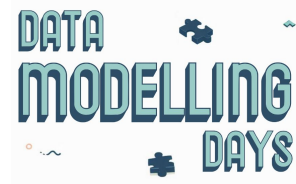
Data modelling is at the core of what we do on Wikidata.

It can be hard - it can be messy.

It can be confusing. Where to even start asking questions?

Let's get together to:

- welcome people in and help them find their way around
- discuss the nitty-gritty details of particular modeling issues
- talk about how to improve data modelling in Wikidata overall
- learn about existing and upcoming tools together



Some big picture things to keep in mind

Wikidata's data is there to be used.

Well-modelled data makes it sooo much easier to build applications on top of Wikidata's data.

But “well-modelled” for what exactly?

Some big picture things to keep in mind

Wikibase gives a lot of freedom for how to model data. That is powerful. It allows us to much more faithfully model the world. It also allows us to make things more complex than they meaningfully should be.

With great power comes great responsibility!

Some big picture things to keep in mind

The map is not the territory!

Let's find the right balance.

Let's geek out on data modelling for the next few days together! ✨





Let's have a look at the program!

[[Wikidata:Events/Data Modelling Days 2023]]

[w.wiki/7iBe](https://www.wiki/7iBe)





Any questions about the event so far?

Feel free to unmute your mic or write your question in the chat.



Results of the research on ontologies issues

Ontology issues are a big issue for our reusers

- 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

What do we even mean when we say ontology issues?

Conflicting
real-world models

Conceptual
ambiguity

Semantic drift

Upper level
ontology is
messy

Structural bugs (cycles,
redundancy, ...)

Inconsistent
modeling

Overgeneralisation

We asked reusers

- Which types of ontology issue are the worst for you?
- Which ones do you encounter most often?

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

embassy (Q3917681)

permanent diplomatic mission of higher level, representing its operator in the country the emb
ambassadorial delegation | diplomatic representation | de jure embassy

~ In more languages

Language	Label	Description
English	embassy	permanent diplomatic mission of higher level, representing its operator in the country the embassy is in
German	Botschaft	ständige diplomatische Auslandsvertretung eines Staates am Regierungssitz eines anderen Staates

All entered languages

Statements

subclass of	diplomatic mission ↳ 0 references
	location ↳ 0 references

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

Want to dive deeper into the results?

[Wikidata:Ontology issues prioritization](#)

Want to take part in our research projects?

[Sign up to Wikimedia Deutschland research participant pool](#)

What do we do about it?

- Make EntitySchemas more useful
- Make constraint violations queryable
- Make Wiki Projects mode visible
- Make showcase Items and documentation better and more discoverable
- Create automated lists for Wikipedia and Co
- Treat ontology-related statements differently in the UI
- Show the bigger picture
- Fix common mistakes automatically



Any questions?

Feel free to unmute your mic or write your question in the chat.

Coming up later today:

- 16:00 UTC (in a few minutes)

Data Modelling Clinic focused on biodiversity topics



- 17:00 UTC (in 1 hour)

Introduction to Entity Schemas

- 18:00 UTC (in 2 hours)

Let's write a Schema together

- 19:00 UTC (in 3 hours)

Modelling challenges when reusing data on Wikimedia projects

- 20:00 UTC (in 4 hours)

Social time & games

Thanks for being here <3
See you around!

Questions, issues during the event?
Feel free to write to Léa Lacroix
lea.lacroix_ext@wikimedia.de

