Insights for researchers and current challenges

Lydia Pintscher
lydia.pintscher@wikimedia.de - @nightrose
Wikidata Workshop 2022
Giving more people more access to more knowledge
How it all started

- 6 months of initial development and then very minimal launch and then expansion
- Sitelinks to bootstrap content and community - task forces developed to organize initial content work
- Attracted existing editors from Wikimedia projects and new people who found a home in Wikimedia for the first time
Wikidata today

- Project at the center of the Wikimedia Projects and the Wikibase Ecosystem
- Most-edited Wikimedia project
- Large corpus of data: 100 M Items with 1.4 B statements and 700K Lexemes
- Healthy community that builds out the project: 12K active editors
- Central to technology we use every day
How has Wikidata changed?

Wikidata

General purpose data

Structured Data on Commons

Lex. data

Wikipedia

Other Wikimedia projects

External reuse

Wikibase Ecosystem
How has Wikidata changed the Wikimedia movement?

- Brought us closer together
- Lifted up the other projects
- Core piece of making the movement strategy of Knowledge as a Service and Knowledge Equity reality
How has Wikidata changed the world?

- More people have more access to more knowledge (but we’re still far from where we need to be)
- Technology we use every day is built on a more open and influenceable system
- Made the ideas of the semantic web more approachable for more people
What was crucial for Wikidata to get to where it is today?

- Community
  - Not just editing but outreach, tool building and so much more
  - Trust the community to make it work and make the right decisions
  - Provide people the tools to help themselves instead of doing everything for them
- Commitment to leave content and ontology to community
- Connection to Wikipedia and the other Wikimedia projects
- Human + machines working together

Photo by Mārtiņš Brucenieks (cc-by-sa 4.0)
What was crucial for Wikidata to get to where it is today?

- Focus on verifiability instead of truth and everything that follows from it (conflicting statements, references, qualifiers, ranks, ...)
- Multilinguality and a focus on IDs instead of one natural language
- CC-0 (reusability, less ambiguity, leveling the playing field)
- Openness and flexibility at (almost) every level -> allow people to do things you have never even thought of or dreamed about
How can you think about the Wikidata community?

- There is not one community but many subcommunities and groups taking care of different content and processes
- International and multilingual
- Some contribute to other Wikimedia projects (e.g. Wikipedia), some come from other open data projects, some contribute only to Wikidata
- For many it is a purely volunteer activity, for some it is part of their job, others even made it their job!
How can you think about the Wikidata community?

- Enjoy organizing, structuring and connecting data - bringing order to the chaos of the unstructured world
- Understand that data powers much of our day-to-day life
- Generally tech-savvy and open to new technology
- Tickled by someone using their data for something new and unexpected
- Sometimes just want to make their Wikimedia project better
- Bots are just tools we use
Community processes and venues

Ultimately work and decisions happen on Wikidata!

- Project chat
- Wiki projects
- Admin noticeboard
- Property creation
- Requests for comments

- Weekly summary
- Chats on Telegram
- Social networks
- Mailing list
- Office hours
- Events for community building around topics or events
How can you work together well with the Wikidata community?

- Reach out! Talk to people! (Wiki project, Telegram, Project chat, events)
- Your research can benefit a lot from getting input from the people that make Wikidata run every day -> fact checking and impact of your research
- Understand that for many editors this is their free time and a project they care deeply about
- Make it clear what's in it for everyone involved. Why should they help you?
- If you're not sure, reach out to me
Strategic challenges

- Empower the community to increase data quality
- Facilitate equity in decision making
- Increase reuse for increased impact
- Strengthen underrepresented languages
- Enable Wikimedia projects to share their workload
What are we doing?
What’s next?

• Building out lex. data
• Better encoding of modeling decisions with Entity Schemas
• Easier data access for developers with new REST API
• Tighten the integration between Wikidata and the other Wikimedia projects
Current research questions

- Coping with growth
  - What else can we do to give the core community the tools they need to keep the data in good shape, especially around consistent modelling and vandalism fighting?
  - How can we continue to make the power of querying a large knowledge graph available and accessible?

- A world in which most contributors to Wikidata are not on Wikidata.org
  - How can we give more people the ability to contribute without necessarily being a part of the core community and still help them contribute productively?
  - How can we keep the community together in this scenario?
Current research questions

- Reuse
  - How can we better understand how and where Wikidata’s data is used within the constraints we have?

- Equity in decision making
  - Which significant hurdles exist in our decision making processes and how can we reduce them?
Join us to celebrate one decade of Wikidata!

[[Wikidata:Tenth Birthday]]

#WikidataBirthday
Thank you

See you on Wikidata!

Email:
lydia.pintscher@wikimedia.de

Twitter:
@nightrose

Wikidata:
Q18016466