WIKIMEDIA DEUTSCHLAND

State of Wikidata usage on other wikis

RESEARCH PROGRESS REPORT – 27.06.2024

Elisha Cohen
Wikidata Integrations Team

Report contents

1.
Background

- What is "Wikidata integration"?
- What is the Wikidata Integrations Team?
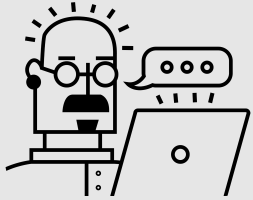 - Strategy
  - Scope

2.
Research approach

- Goals
- Methods
- Timeline

3.
Early learnings

- Three main categories of WD usage
- Additional frameworks for understanding how editors currently use WD on other projects

4.
Next steps

- First implementations
- Share out results
- Phase two of research

Questions, feedback, comments, discussion

https://docs.google.com/document/d/1JecxlCsIYx10WGWHinxDvTMfCBgWedhC8sx4IsOI6MI/edit
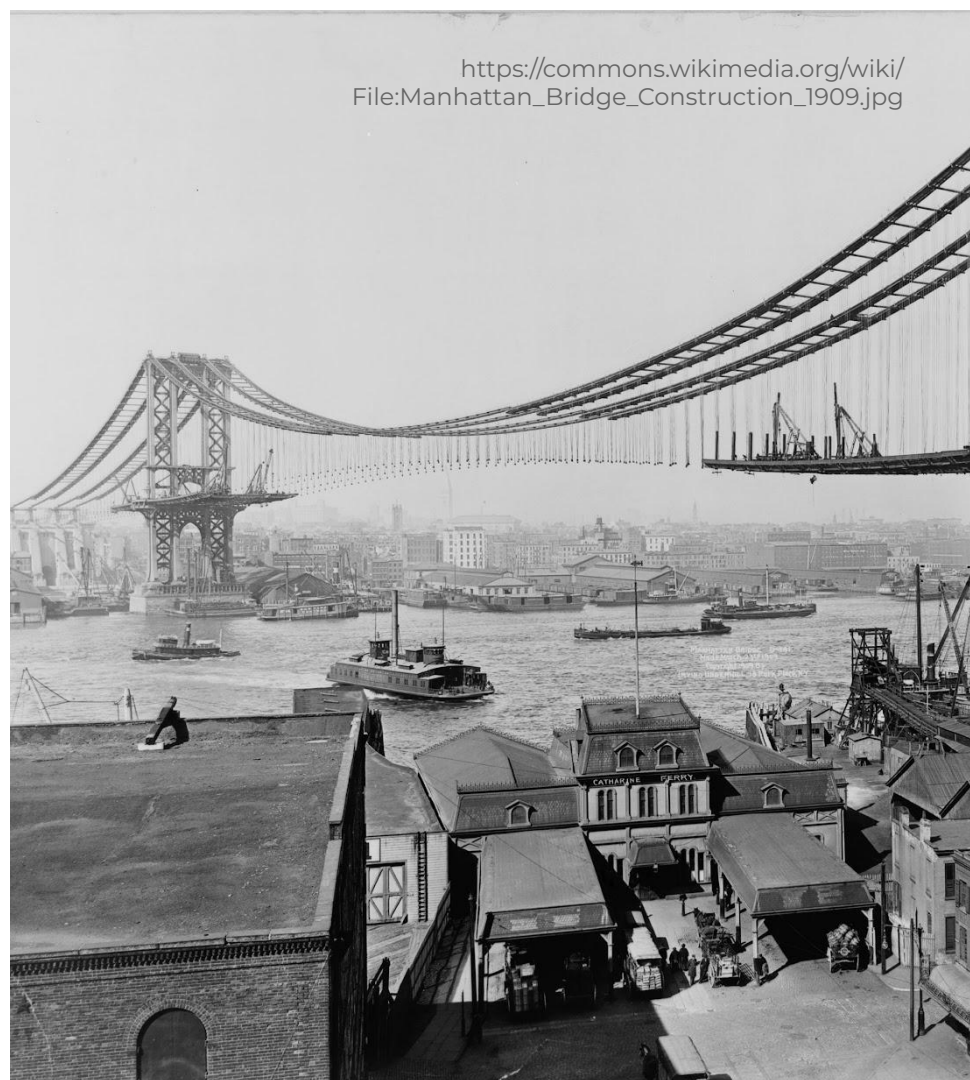
WIKIMEDIA
DEUTSCHLAND

01 Background

- What is "Wikidata integration"?
- What is the Wikidata Integrations Team?

What is "Wikidata integration"?

Any time Wikidata (WD) is used by editors to support their work on Wikipedia and other Wikimedia sibling projects

"Integration" references all the supporting features, tools, and underlying infrastructure that connect Wikidata to other projects and allow Wikidata to be used there

What is the Wikidata Integration Team (WIT)?

WIT is a Wikimedia Deutschland team dedicated to supporting editors who want to use Wikidata on other projects and improving Wikidata integration workflows

Project strategy

First steps
Support the existing core workflows and make sure those work well.

In the future
We might introduce new features and new ways that WD can be used by editors. There are many opportunities already identified in this early research, but the priority is making sure the existing, in-use functionality is usable.

Project scope

The primary focus is on supporting Wikipedia editors but ideally the features we improve also make WD usage easier on the other sibling projects.

Out of scope: supporting using Wikibase to create new knowledge graphs to serve other projects, such as was done with Structured Data on Commons (SDC).

Want to know more about the team?

Visit the project's meta page

- Background information
- Newsletter sign-up
- Team and contact details
- Resources

Primary points of contact

Product Manager
Ifeatu Nnaobi
(user page)

Community Communications Manager
Danny Benjafield
(user page, email)

02 Research approach

- Goals
- Methods
- Timeline

Primary Goals

Phase 1

Comprehensive, shared understanding and documentation of the **current state** of Wikidata integrations in the other Wikimedia projects and how data flow works

Phase 2

Learn about editor **motivations**, existing **workflows**, and **problems** encountered while using Wikidata's data on Wikipedia and the sibling projects

Supporting Goals

Set up the team for future success by:

- Creating a knowledge foundation the team can build on for the future (metric definitions and baseline numbers, shared mental models, …)

- Making project-wide scoping decisions (target user groups)

- Identifying first small projects for developer onboarding

Research plan

MIXED METHODS

PHASE 1		PHASE 2	
DESK RESEARCH	**EXPERT INTERVIEWS**	**DATA COLLECTION + ANALYSIS**	**COMMUNITY INTERVIEWS**

Review existing data
- Onwiki (help pages, disc pages)
- Phabricator (review all related past work, feature requests, component boards and bug reports)
- Existing tracking and metrics
- Past research

Collect new data
- Onwiki problem submissions

Document and inventory
- Existing features, extensions, tools, gadgets, scripts
- Differences between interfaces (skins, mobile, apps,)

Internal stakeholders
- Talk with experts within WMDE / WMF to better understand what has already happened

Community experts
- Interview editors who are known for their work in this space to learn about history, opportunities and possible roadblocks

Goals
Learn more about the frequency of use of different methods for transcluding Wikidata (embedded statements) and determine patterns across languages and projects

Establish meaningful ways to track Wikidata usage over time and collect baseline numbers

Approach
Extracting and analyzing two data sources:
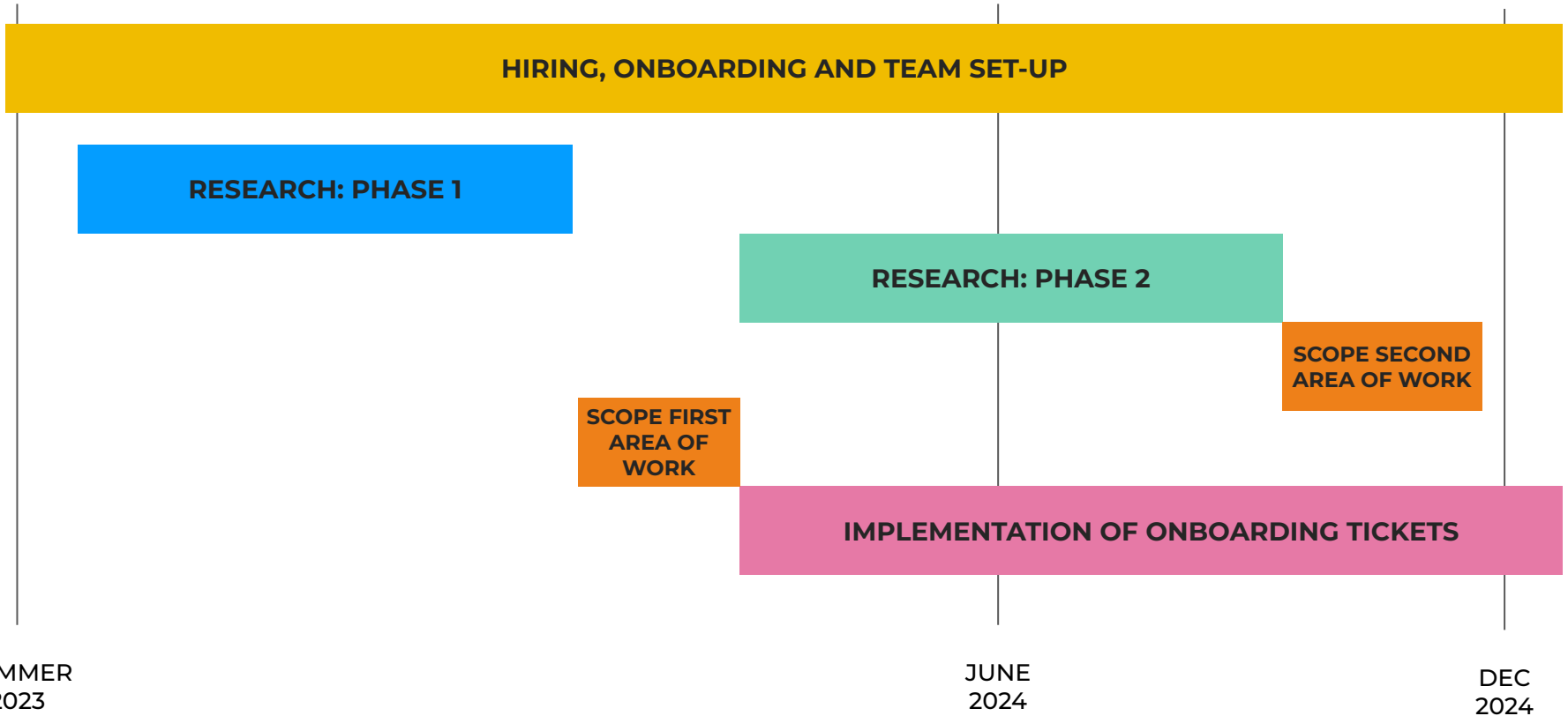- usage tracking (complete)
- wikitext (sampled)

Goal
Focused on learning about motivations, problems and troubleshooting, priorities, expectations

Recruiting for diversity factors
- Gender
- Location
- Editing experience
- Editing hrs/wk
- LTR / RTL / Non-latin scripts
- Primary wiki size
- Project(s)
- Workflows
- Frequency of integration work

What has happened so far? What is next?

HIRING, ONBOARDING AND TEAM SET-UP

RESEARCH: PHASE 1

RESEARCH: PHASE 2

SCOPE SECOND AREA OF WORK

SCOPE FIRST AREA OF WORK

IMPLEMENTATION OF ONBOARDING TICKETS

SUMMER
2023

JUNE
2024

DEC
2024

03 Early learnings

- Three main categories of WD usage
- Additional frameworks

Wikidata is used by editors in three **main** ways

https://commons.wikimedia.org/wiki/
File:Three_giraffes_01.jpg

1 Link pages

Wikidata links projects and languages together by topic

- [Sitelinks](#)
- Language switching
- Project switching

Wikipedia (80 entries) ✏️ edit

af	Ursula K. Le Guin
ar	أورسولا لي غوين
ary	ؤرسولا لگّوين
arz	اورسولا لى جوين
ast	Ursula K. Le Guin
ay	Ursula K. Le Guin
azb	اورسولا لو قویین
az	Ursula Kröber Le Quin
ba	Урсула Ле Гуин
bcl	Ursula Kroeber Le Guin
be	Урсула Кробер Ле Гуін
bg	Урсула Ле Гуин
bn	উরসুলা কে. লে গুইন
br	Ursula K. Le Guin
ca	Ursula K. Le Guin
cs	Ursula K. Le Guinová
cy	Ursula K. Le Guin
da	Ursula K. Le Guin
de	Ursula K. Le Guin
el	Ούρσουλα Λε Γκεν
en	Ursula K. Le Guin 🏅
eo	Ursula K. Le Guin
es	Ursula K. Le Guin
et	Ursula Le Guin
eu	Ursula Le Guin

Wikibooks (0

Wikiquote (21

az	Ursula Kr
be	Урсула К
bg	Урсула Л
bs	Ursula K.
cs	Ursula K.
en	Ursula K.
es	Ursula K.
et	Ursula Le
fa	ولا لو گوین
fr	Ursula K.
he	ה לה גווין
hy	Ուրսուլա
it	Ursula K.
ka	ურსულა
lt	Ursula K.
pl	Ursula K.
pt	Ursula K.
ru	Урсула К
sr	Урсула Л
tr	Ursula K.
uk	Урсула Л

2 Embed statements

Multilingual, structured data is pulled from Wikidata and displayed on article pages

- Modules and templates
  - Parser functions
  - Lua functions

- Wikifunctions (*in development*)

- Extensions
  - Kartographer (maps)
  - Graph (currently disabled)

A view of the Black Rock Mesa FD site with the doors open

Location(s)	Millard County, Utah
Coordinates	🌐 39°17′49″N 112°54′31″W ✏
Altitude	1,400 m (4,600 ft) ✏
Built	2003–2007 ✏
First light	2008 ✏
Telescope style	gamma-ray telescope ✏
Website	www.telescopearray.org ↗ ✏

2 Embed statements

Multilingual, structured data is pulled from Wikidata and displayed on article pages

- Authority control
- Infoboxes
- Coordinates
- References
- ...

Articles related to Grace Hopper	
Authority control databases	
International	FAST ↗ · ISNI ↗ · VIAF ↗ · WorldCa
National	Norway ↗ · France ↗ · BnF data ↗ · United States ↗ · Czech Republic ↗ Netherlands ↗ · Poland ↗
Academics	Association for Computing Machiner MathSciNet ↗ · Mathematics Geneal zbMATH ↗ (2 ↗)
People	Trove ↗
Other	SNAC ↗ · IdRef ↗

2 Embed statements

Required, related functions for embedded statements to work:

- Usage tracking, caching and updating
- Watchlist entries and anti-vandalism

3 Support editing

Wikidata is used by editors to support their processes and editing work on other projects

- Queries for worklists (ex. Listeria)

- Starter articles, stubs, drafts (ex. MBabel)

- Visualize gaps in knowledge (ex. WikiShootMe)

#	name	image	description	occupation	country of citizen
1	Annemarie Mooren		Dutch television actor and television director	television actor television director film director	Kingdom of the Netherlands
2	Christiane Spiero		French film director	film director screenwriter	France
3	Caroline Chu		actress	actor film director	United Kingdom
4	Anne Budgell		Canadian author	filmmaker film director writer journalist	Canada
5	Louise Johnson		Canadian filmmaker	filmmaker film director producer	Canada
6	Kita Bauchet		French film director	film director	France
7	Olga Rakhmanova		Russian and Soviet actress and director of theater and silent films, drama teacher (1871—1943)	film director stage actor drama teacher screenwriter theatrical director silent film actor film actor acting instructor	Russian Empire Soviet
8	Adriënne Solser		Dutch actress, film director and producer	film actor film director film producer	Kingdom of the Netherlands
9	Margarete Lindau-Schulz		German screenwriter	screenwriter film director	German United of America

Related: supporting tools

Many standalone tools, games, scripts, and gadgets help make specific tasks related to the three primary workflows easier by:

- Providing a way to do it without technical knowledge, often with a graphical user interface (GUI)

- Reducing platform switching between Wikidata and other wikis while adding or editing data

- Auto-checking and making suggestions for possible edits to fix missing links

Examples
- *Query Builder*

- *WE-Framework*
- *Wikivoyage Listing Editor*
- *Wikidata Bridge*

- *The Distributed Game*
- *Duplicity*

More frameworks to understand Wikidata use on other wikis

Related to basic data flow

- Data is **linked**

- Data is **displayed**

- Data is **editable**

- Data is **updated**

- Changes are **trackable** and **understandable** to editors

Supporting flow

- Make the steps related to data flow easier, more intuitive, provide additional options

- Make editing easier (get a head start with data that already exists on WD or provide work lists)

Main or content namespaces

- **Statement(s)** are inserted from a QID
  - WD-powered templates, e.g.:
    - Infoboxes
    - References
    - Coordinates
    - Authority control
  - Wikifunctions *(in development)*

- **Visualizations** are inserted from a QID or SPARQL query
  - Maps
  - Graphs

Non-content namespaces
(PROJECT, USER, HELP, ...)

- **Lists** are inserted from a SPARQL query

Content is inserted within a page

- **Statement(s)** are inserted from a QID
  - WD-powered templates, e.g.:
    - Infoboxes
    - References
    - Coordinates
    - Authority control
  - Wikifunctions *(in development)*

- **Visualizations** are inserted from a QID or SPARQL query
  - Maps
  - Graphs

- **Lists** are inserted from a SPARQL query

Content inserted into UI

- **Sitelinks**

- **Short descriptions**

No content is inserted

- Tracking **categories**

Marked / identified as coming from Wikidata

- Some templates have graphic conventions to mark if information is coming from Wikidata but it not consistent. Examples:

  - with a pencil icon

  - with a link to edit WD

- Recent changes / watchlist entries marked with a "**D**"

Not marked as Wikidata

- Sitelinks

- Short descriptions

Workflow difficulty and frequency

Beginner
Frequent

- Sitelinks

Intermediate
Sometimes

- Use WD powered infoboxes
- Re-use someone else's SPARQL query or ask someone for a list
- Anti-vandalism checks
- Use QID for citations
- Use QID for maps

Advanced
Rare

- Write/maintain WD powered template or module
- Write SPARQL queries
- Create/maintain tools, gadgets and scripts

04   Next steps

- First implementations
- Share out plan
- Phase two of research

First implementations

- Prioritizing smaller fixes in core areas where:
    - We know the change will have an widespread impact
    - Our developers need onboarding
    - Low UX effort needed focus can stay on research
    - Clearly defined scope so we can practice going from beginning to end of product development process as a new team

- First topic areas:
    - Sitelinks
        - Increase consistency and visibility of sitelinks via the language switcher and project switcher
    - Lua functions which call Wikidata
        - Bug fixing
        - Improve error handling

Share out results and build on findings

commons.wikimedia.org/wiki/File:Original_packaging_cardboard_box_for_the_Mettoy_Minor_toy_typewriter_with_instructions_for_use.jpg

Rewrite help pages

Use the information we have documented and our frameworks to update help pages to be clearer and cover all current options

commons.wikimedia.org/wiki/File:3d10_fm_de_vilafranca.jpg

Internal coordination

Make sure all teams with overlapping interests and content share research with each other and build on the information we are each learning, esp. Wikidata and Wikifunctions teams

commons.wikimedia.org/wiki/File:Post-Its.jpg

Present + collect feedback

Share reports both onwiki and with live presentations to hear from the community about if it aligns with their understanding of Wikidata integration.

Research plan

MIXED METHODS

PHASE 1		PHASE 2	
DESK RESEARCH	**EXPERT INTERVIEWS**	**DATA COLLECTION + ANALYSIS**	**COMMUNITY INTERVIEWS**

Review existing data
- Onwiki (help pages, disc pages)
- Phabricator (review all related past work, feature requests, component boards and bug reports)
- Existing tracking and metrics
- Past research

Collect new data
- Onwiki problem submissions

Document and inventory
- Existing features, extensions, tools, gadgets, scripts
- Differences between interfaces (skins, mobile, apps,)

Internal stakeholders
- Talk with experts within WMDE / WMF to better understand what has already happened

Community experts
- Interview editors who are known for their work in this space to learn about history, opportunities and possible roadblocks

Goals
Learn more about the frequency of use of different methods for transcluding Wikidata (embedded statements) and determine patterns across languages and projects

Establish meaningful ways to track Wikidata usage over time and collect baseline numbers

Approach
Extracting and analyzing two data sources:
- usage tracking (complete)
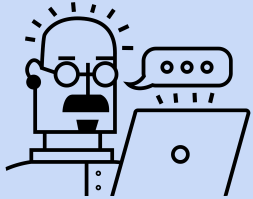- wikitext (sampled)

Goal
Focused on learning about motivations, problems and troubleshooting, priorities, expectations

Recruiting for diversity factors
- Gender
- Location
- Editing experience
- Editing hrs/wk
- LTR / RTL / Non-latin scripts
- Primary wiki size
- Project(s)
- Workflows
- Frequency of integration work

05 Questions, feedback and discussion

Questions, feedback, and discussion

https://docs.google.com/document/d/1JecxlCsIYx10WGWHinxDvTMfCBgWedhC8sx4IsOI6MI/edit

~ 5 minutes to add thoughts ~