Wikidata and Wikibase

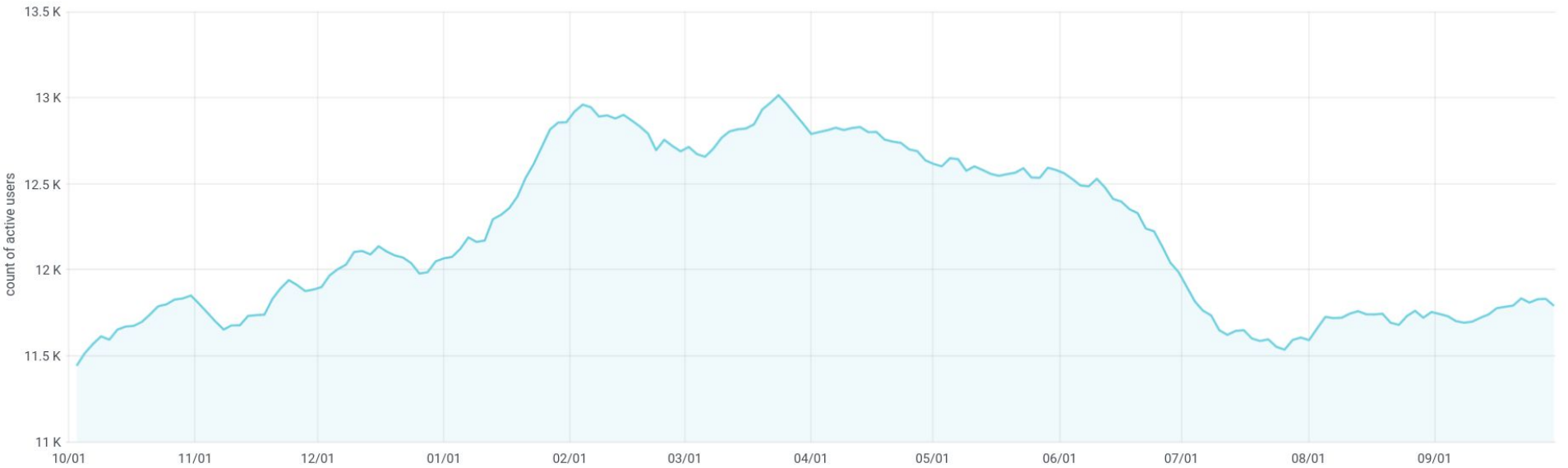
What happened? Where are we going?

Lydia Pintscher
Product Manager for
Wikidata

Sam Alipio
Product Manager for
Wikibase Ecosystem

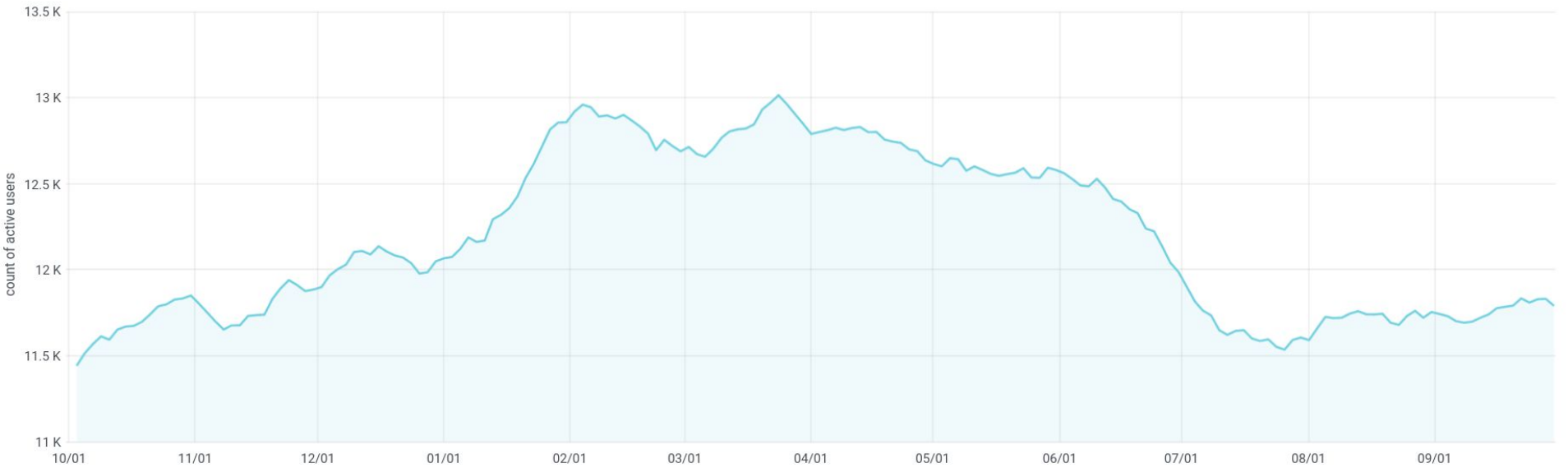
Where are we?

11.8k active editors

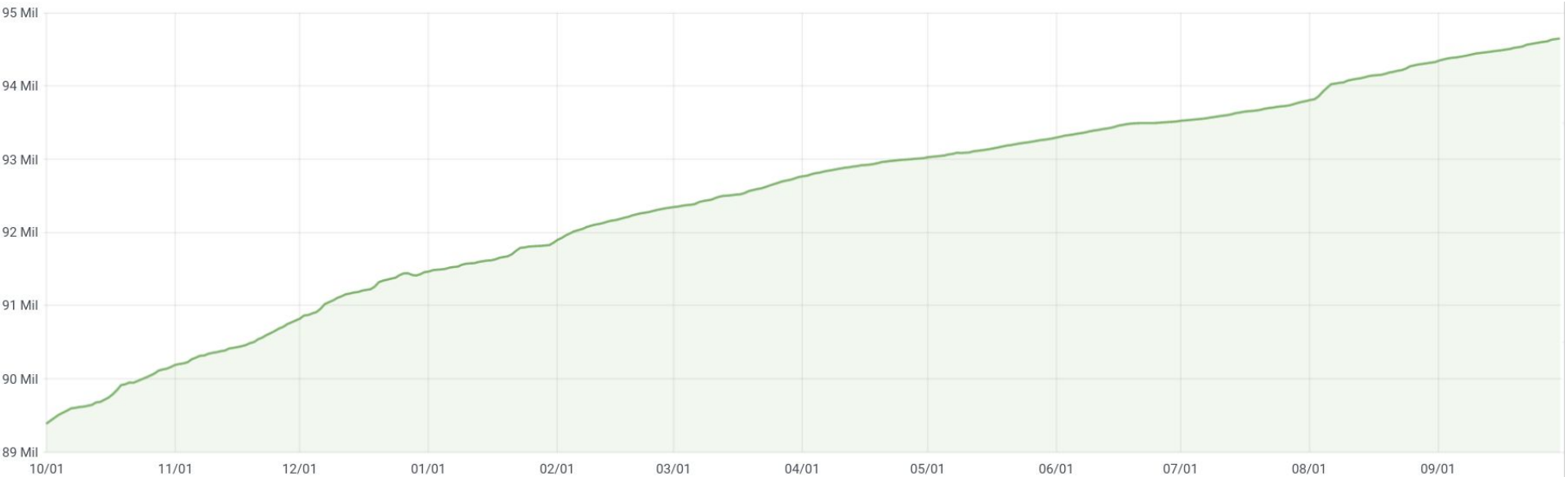


11.8k active editors

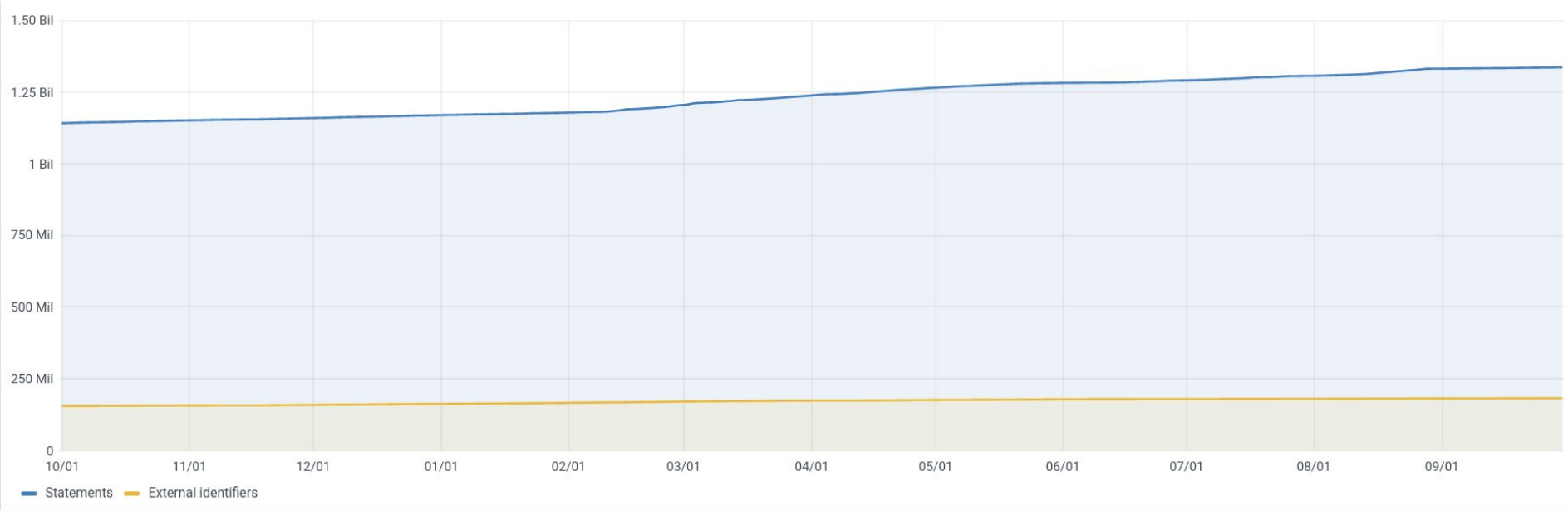
1.5 Billion edits



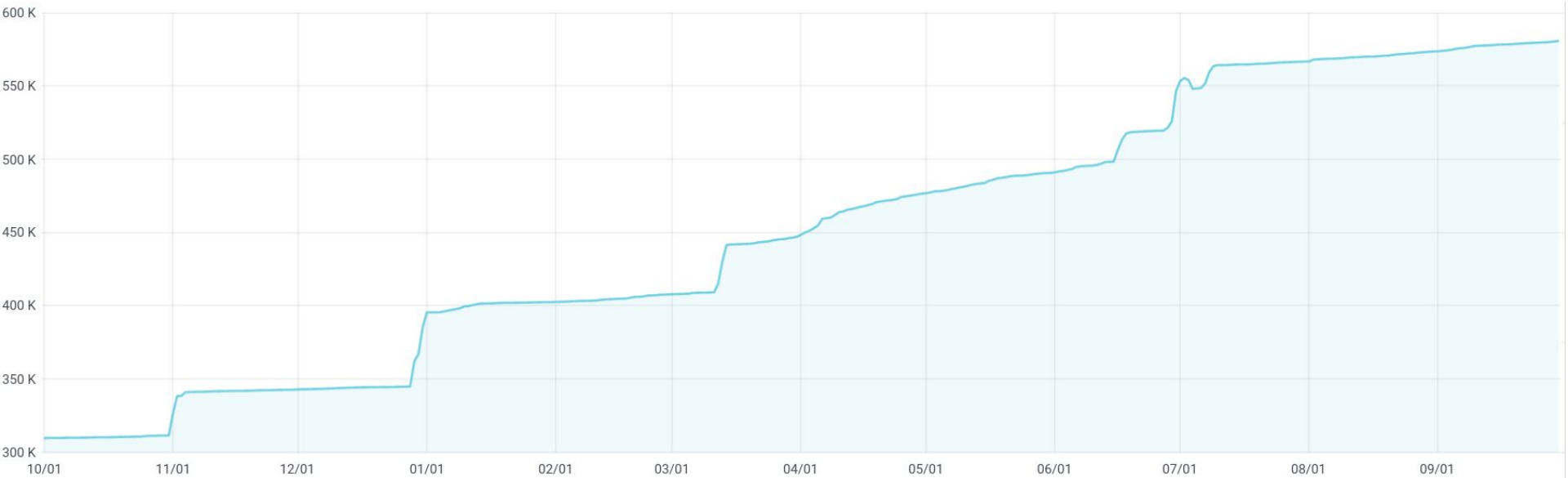
94 Million Items



1.34 Billion Statements

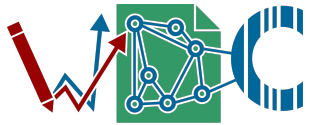


580k Lexemes



71% of Wikipedia articles

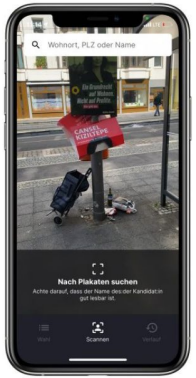
8 Million SPARQL queries/day



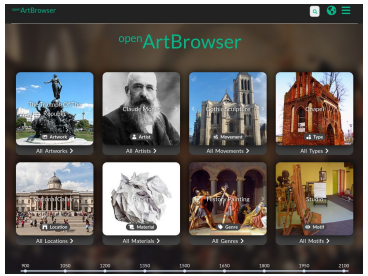
2021 - a sustainable future for Wikidata

OPEN LIBRARY

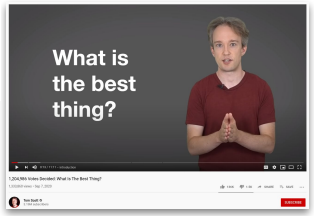
mySociety



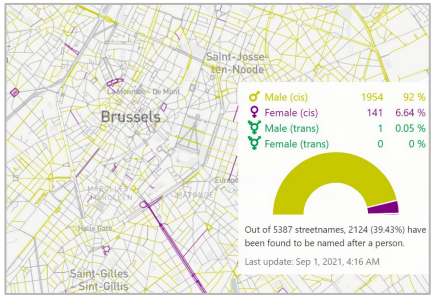
OCCRP



Quora



reddit



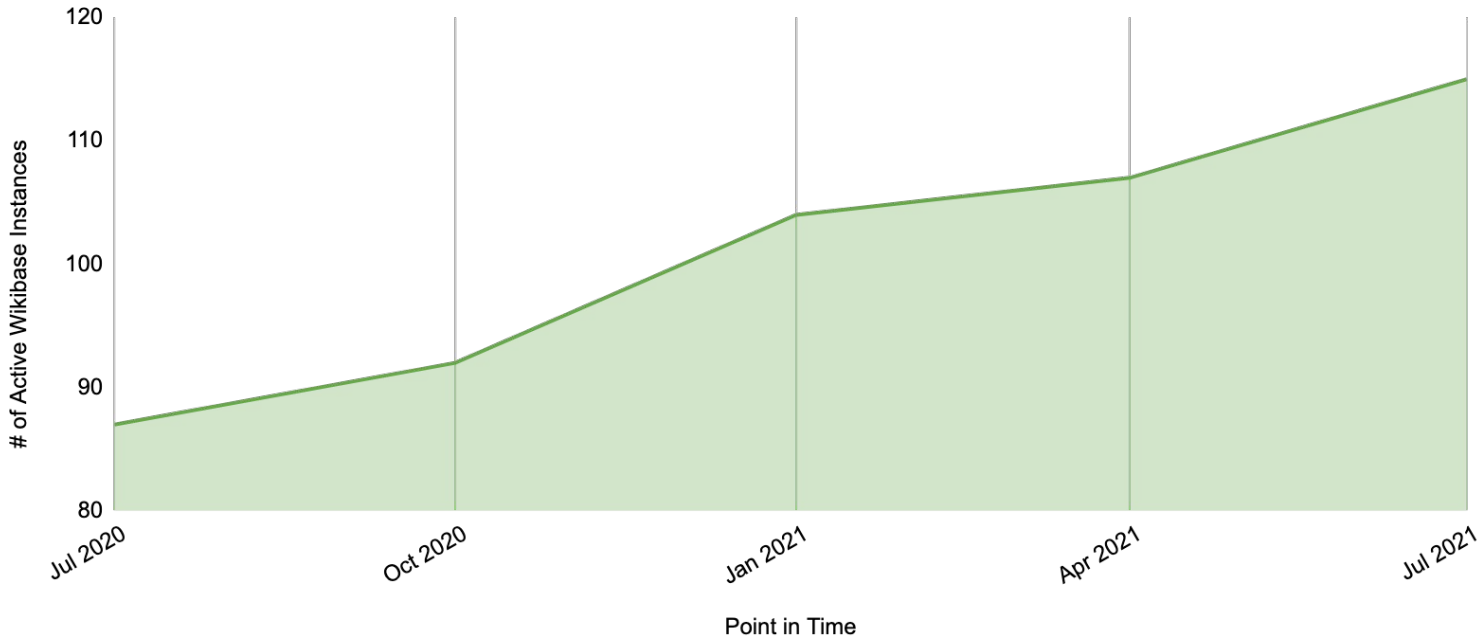
WolframAlpha™

WIKI DATA CON

2021 - a sustainable future for Wikidata

The Wikibase Ecosystem continues to grow:

There are 32% more active instances vs. a year ago.



WIKI DATA CON

2021 - a sustainable future for Wikidata

Examples of Wikibases in the ecosystem

Lingua Libre

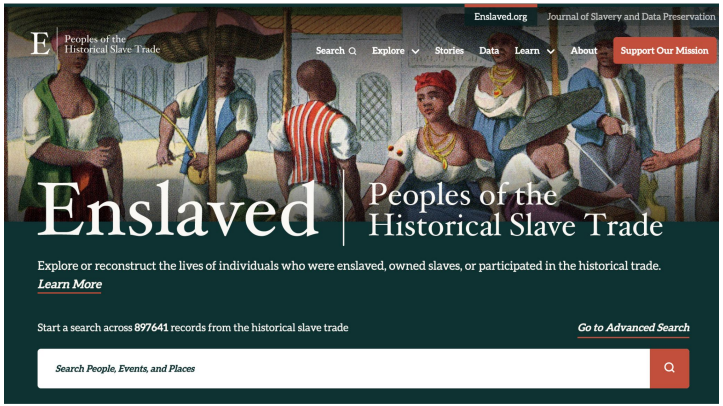
Piloting

DEUTSCHE NATIONAL BIBLIOTHEK

{BnF

BLACK BIBLIOGRAPHY PROJECT

RHIZOME



European Commission | Kohesio Open Projects
Project Information Portal for Regional Policy

New communities of practice have formed around Wikibase

Wiki Library Manifesto

Goal: *The creation and implementation of a single Linked Open Data network for art, culture and science.*

43 signatories, including national libraries, museums, university libraries

WIKIBASE STAKEHOLDER GROUP

Goal: *To commission production and maintenance of open source extensions to Wikibase.*

13 institutional members (including Wikimedia Germany), and additional individual members

WIKI DATA CON
2021 - a sustainable future for Wikidata

What happened over the last year?

Data quality

- Understand
- Improve
- Secure

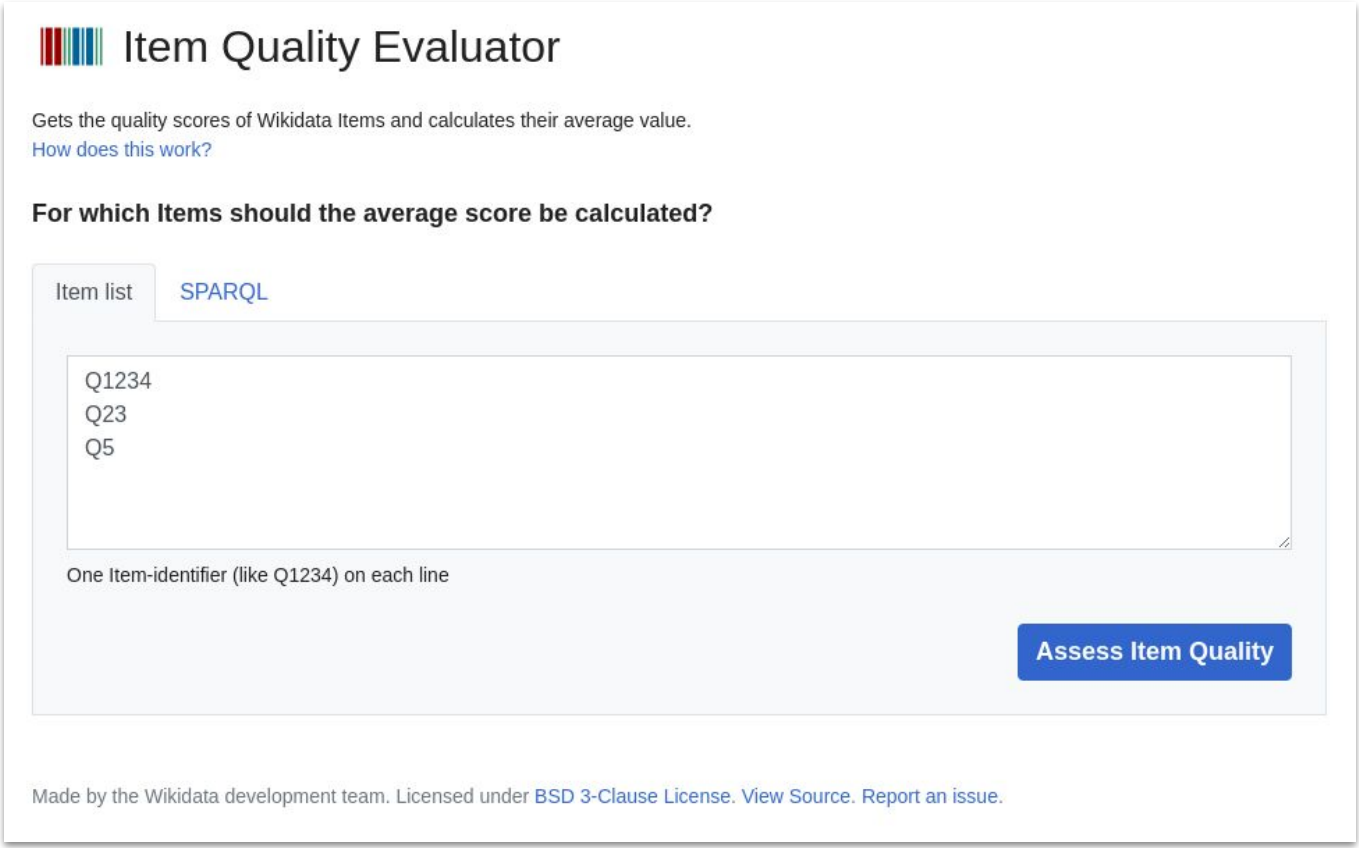
Data quality

- Understand
  - Improved ORES quality scoring for Items
  - Measuring constraint violations
  - Ontology issues research
  - Data Quality Days
- Improve
- Secure

WIKI DATA CON
2021 - a sustainable future for Wikidata

Data quality

- Understand
- Improve
  - Item Quality Evaluator
  - Curious Facts
  - Better watchlist and recent changes integration on Wikipedia and co
- Secure
  - Properties and highly-used Items are now semi-protected
  - More comparison categories



https://item-quality-evaluator.toolforge.org

WIKI DATA CON
2021 - a sustainable future for Wikidata

Item Quality Evaluator

← Refine Item selection

Average quality score based on the 50 Items selected:

2.32 Scores can go from 1 (worst quality) to 5 (best quality)

Scores for 50 Items

Download as CSV

Label	Score
Saint Patrick's Day (Q181817)	4.06
Woodstock Festival (Q164815)	3.82
Saturnalia (Q207213)	3.35
Feast of the Immaculate Conception (Q3538509)	3.23
Oktoberfest (Q102400)	3.08
San Fermín (Q829919)	3.02

WIKI DATA CON
2021 - a sustainable future for Wikidata

FOR REAL, WIKIDATA..?

Curious Facts

The world is weird. Wikidata describes that weird world. Wikidata also sometimes makes mistakes. Let's find out which is which together! Here are a bunch of statements from Wikidata that are potentially weird. Have a look. Is it true? Cool, you learned something new about the world! Is it false? Can you go and fix Wikidata? <3 (If you don't know how, let us know on the [Project chat].)

Talal bin Abdulaziz Al Saud has a Statement "spouse: Mona Al Solh" but Mona Al Solh is not found in the class "human".

This fact was established on: **2021-06-10 05:05:11**, and is based on the **2021-05-31** snapshot in hdfs of the Wikidata JSON dump. Edits made after that date are not taken into account.

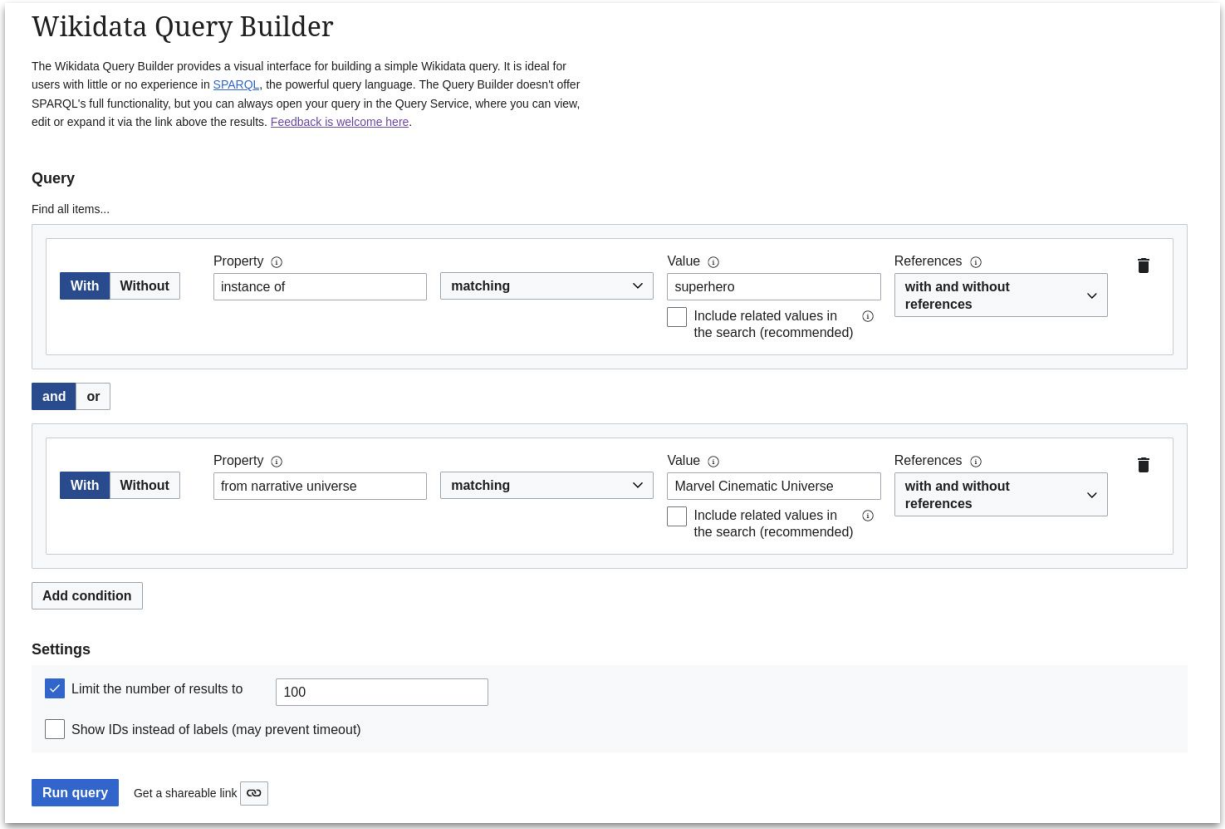
Give me some new weird stuff!

https://wikidata-analytics.wmcloud.org/app/CuriousFacts

WIKI DATA CON
2021 - a sustainable future for Wikidata

Query Service

- Scaling
  - New updater
  - Research to understand content composition better
  - Survey to understand your priorities
- Smaller improvements
  - Titles for queries!

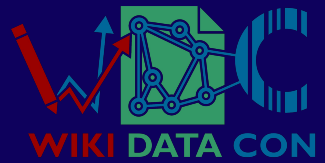


https://query.wikidata.org/querybuilder

WIKI DATA CON
2021 - a sustainable future for Wikidata

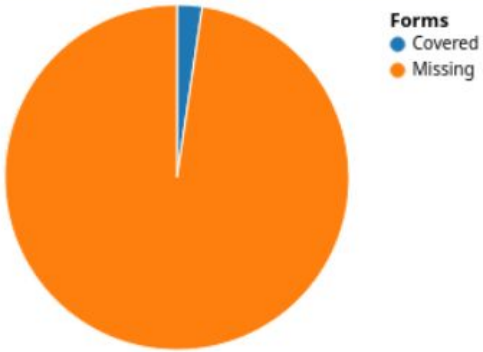
Lexicographical data

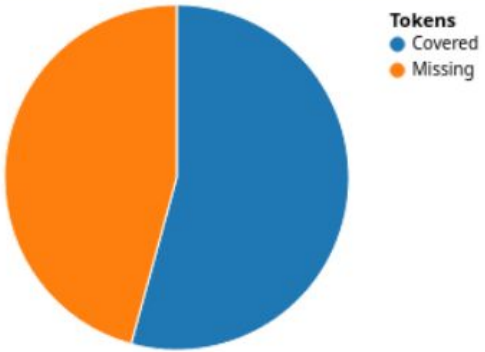
- Abstract Wikipedia - the new kid on the block
- 30 Lexic-o-days
- Focus languages
  - Hausa, Igbo, Dagbani, Malayalam, Bengali



es [edit | edit source]

- Forms in Wikidata: 10,020
- Forms in Wikipedia: 372,589
- Tokens: 405,914,020
- Covered forms: 8,767 (2.4%)
- Missing forms: 363,822 (97.6%)
- Covered tokens: 220,303,435 (54.3%)
- Missing tokens: 185,610,585 (45.7%)
- Most frequent missing forms





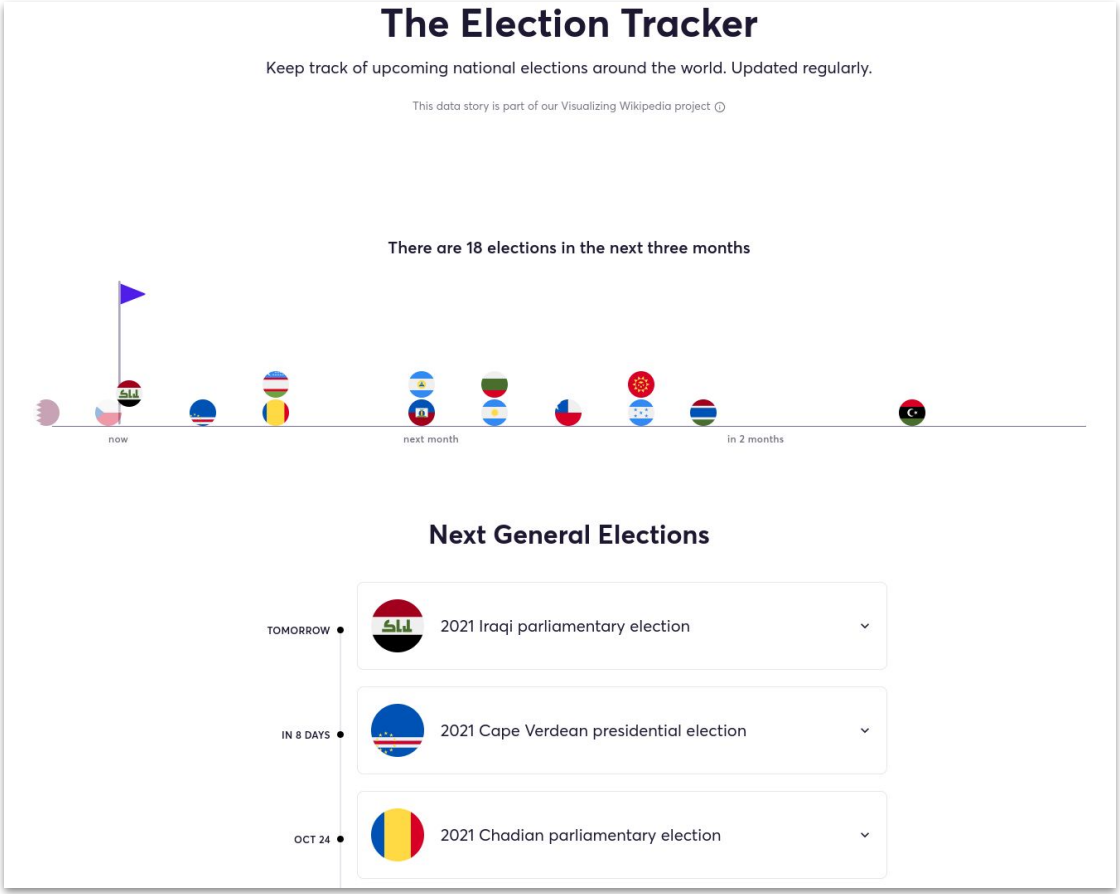
Wikidata:Lexicographical_coverage

WIKI DATA CON
2021 - a sustainable future for Wikidata

What are people building?

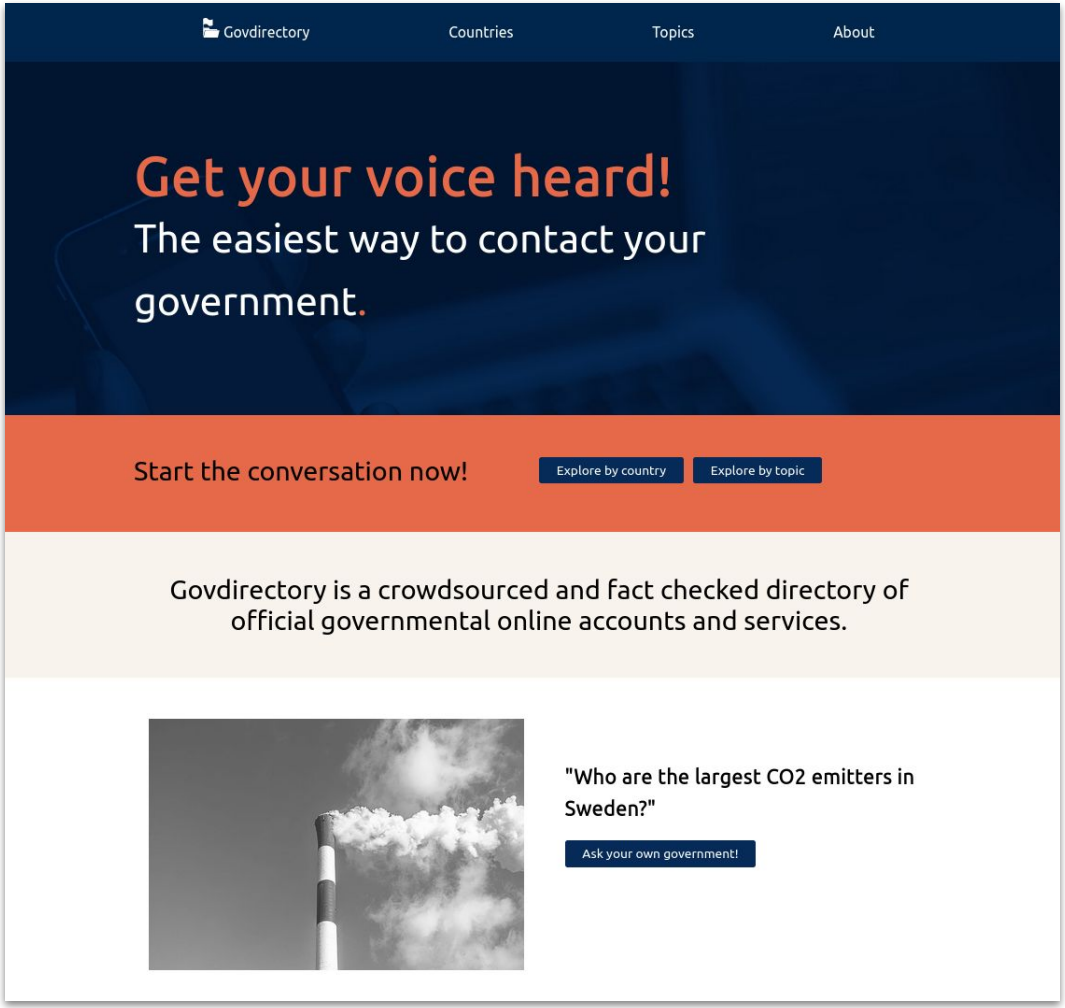
https://facethefacts.app

WIKI DATA CON
2021 - a sustainable future for Wikidata



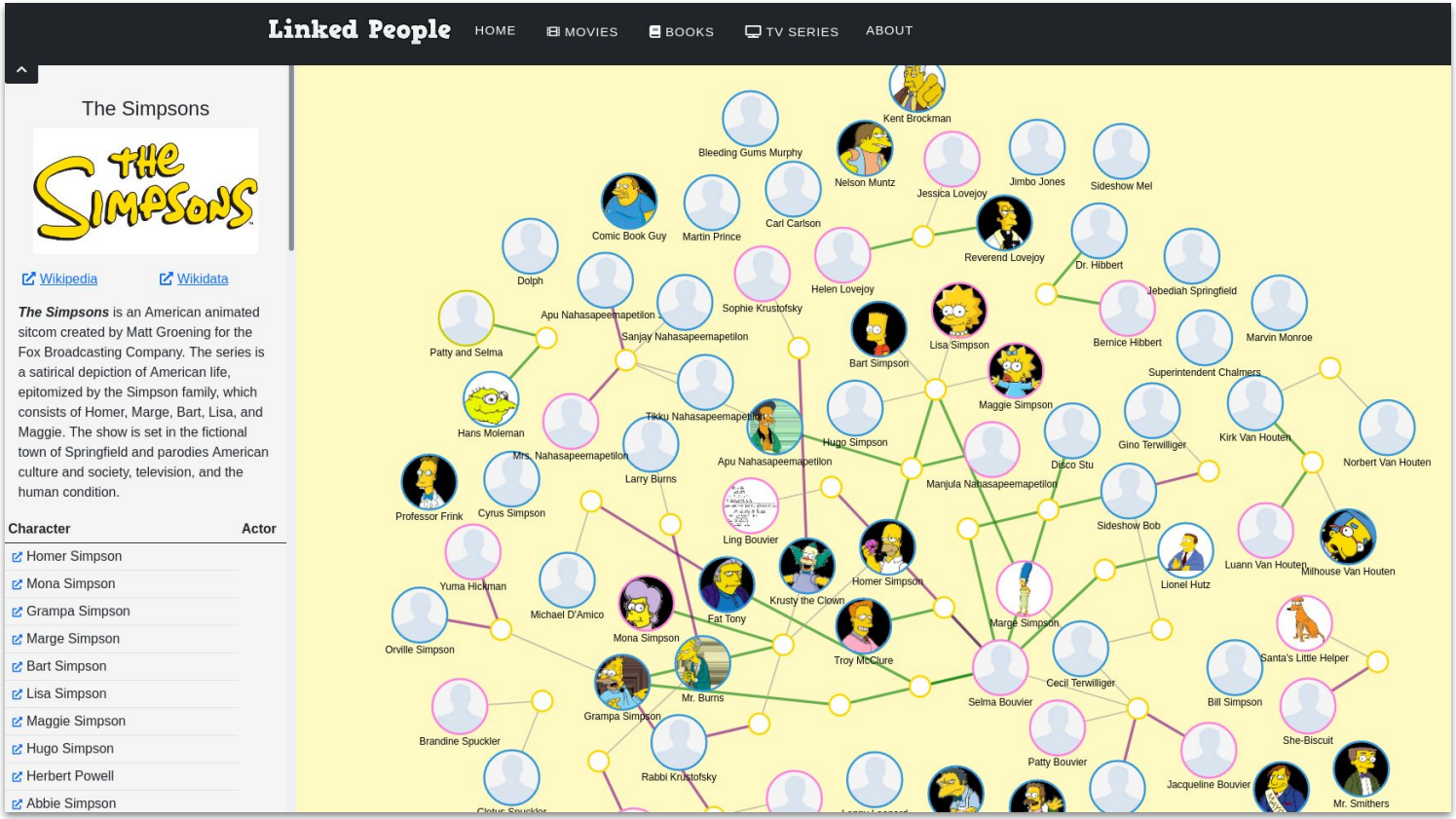
https://www.datastory.org/services/election-tracker

WIKI DATA CON
2021 - a sustainable future for Wikidata



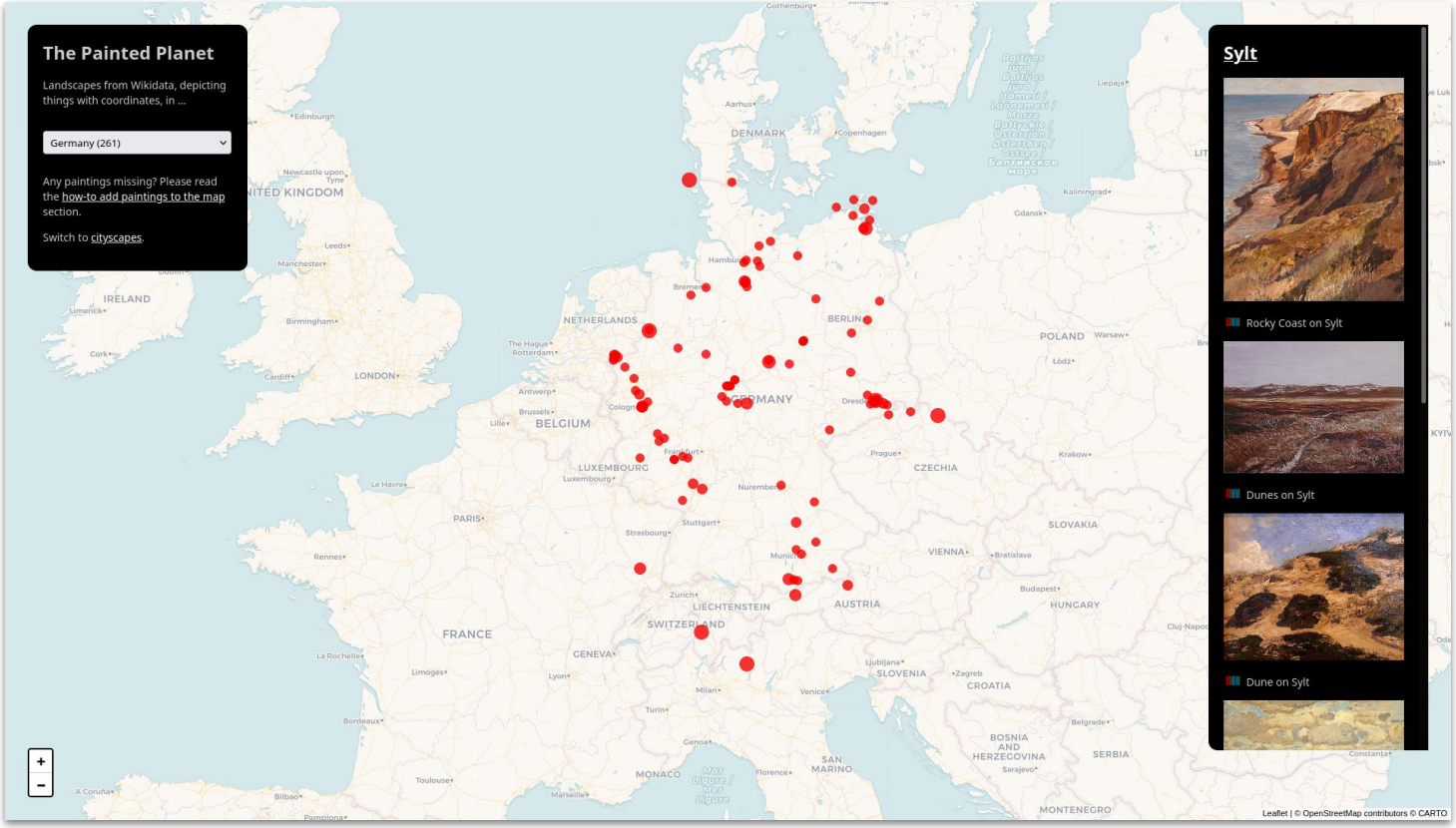
https://govdirectory.org

WIKI DATA CON
2021 - a sustainable future for Wikidata



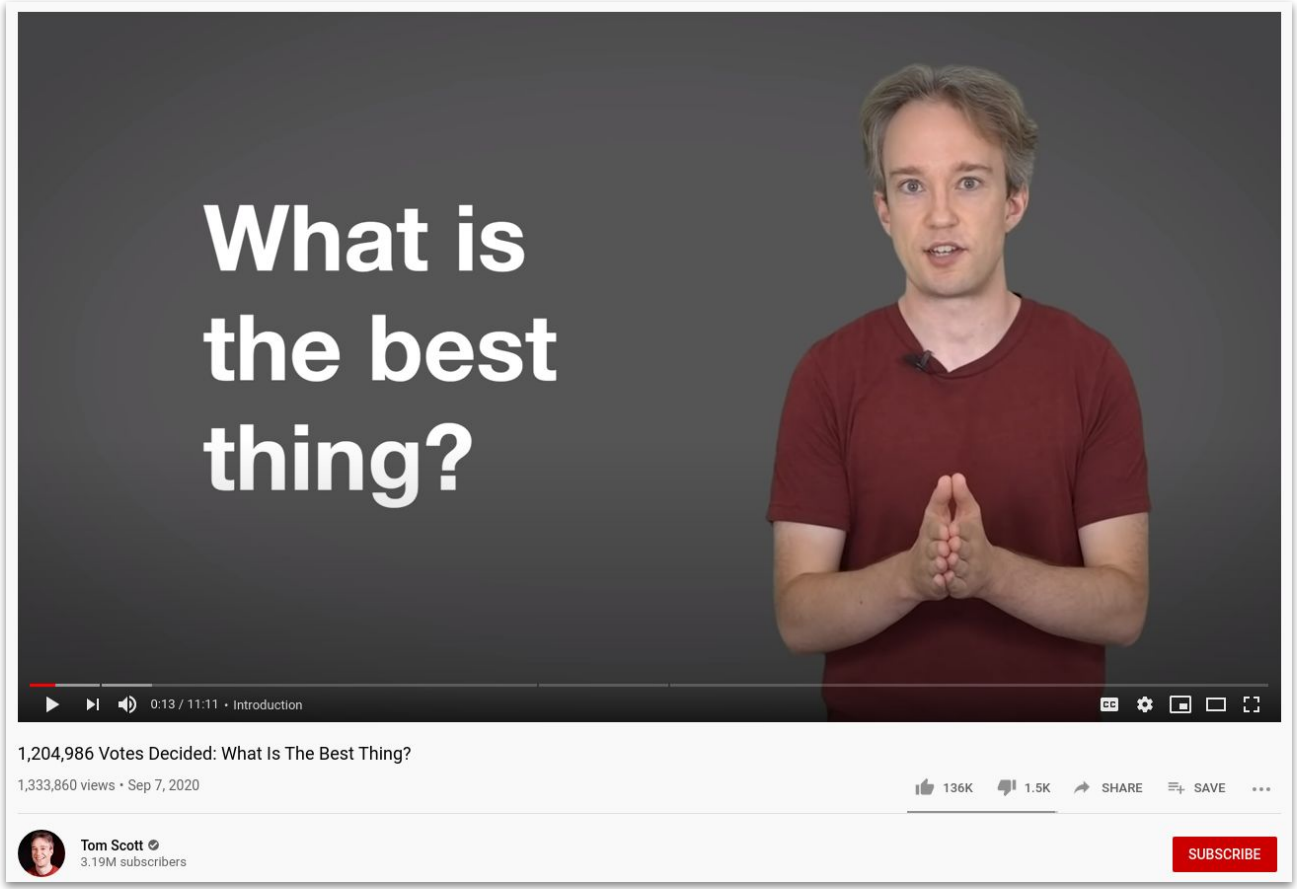
https://linkedpeople.net

WIKI DATA CON
2021 - a sustainable future for Wikidata



https://hicsuntleones.nl

WIKI DATA CON
2021 - a sustainable future for Wikidata



WIKI DATA CON

2021 - a sustainable future for Wikidata

Release Strategy & Infrastucture

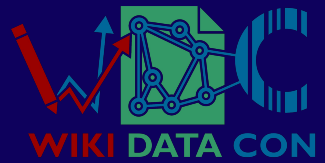
- Users wanted predictability in releases & clearer instructions for upgrading
- Established 6-month release cycle
- Built deployment pipeline & processes for releasing Wikibase Suite

Improved Documentation

- To support releases going forward, we published:
  - New installation instructions
  - New updating instructions (for Docker release)
- Each release accompanied by detailed release notes

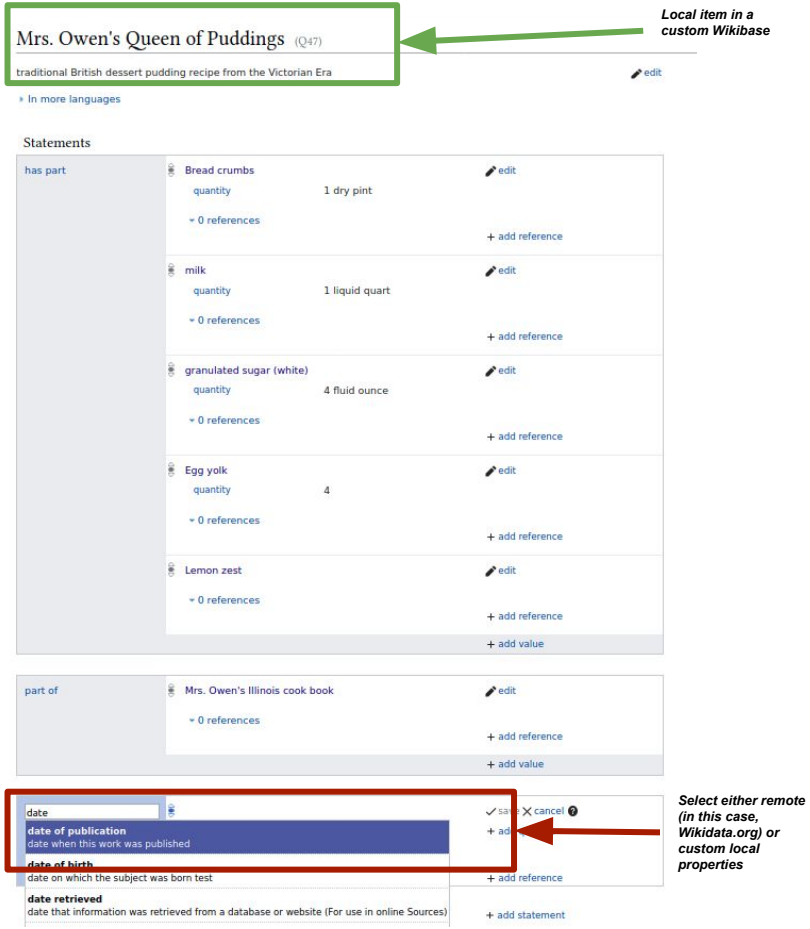
Wikibase Spring Release

- May release included 3 new features:
  - Federated Properties, WikibaseManifest, and Wikibase Pingback
- New Docker bundle included even more helpful extensions and tools



Federated Properties

- We are working on enabling deeper connections between Wikibases so they can share content (e.g. entities) and ontologies
- Wikidata plays an especially important role here



We continued working on the **Federated Properties (v2)** feature which will make it possible for a Wikibase project to enhance its local ontology with remote properties from another Wikibase (especially Wikidata).

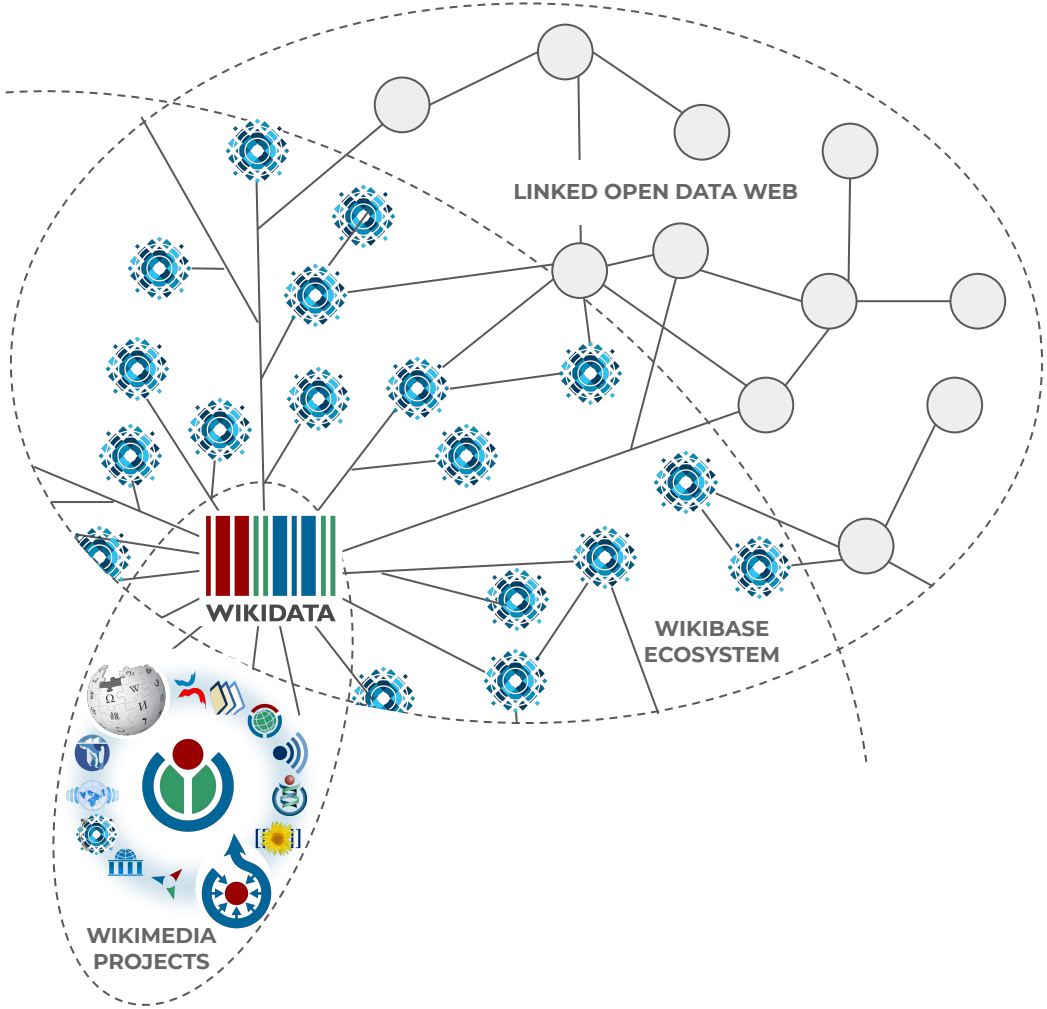
WIKI DATA CON
2021 - a sustainable future for Wikidata

What’s going to be important?

What do we work towards?

- A thriving Linked Open Data Web
- Wikidata and network of Wikibase where communities can co-create a high-quality and diverse global knowledge graph
- We together create the standards and define how the world's knowledge is organized
- People, companies and institutions generate new insights and build new apps and services that change the world for the better

WIKI DATA CON
2021 - a sustainable future for Wikidata



WIKI DATA CON

2021 - a sustainable future for Wikidata

What is important for Wikidata?

Wikidata community

creates and maintains content

Re-users

give back feedback and awareness

WIKI DATA CON

2021 - a sustainable future for Wikidata

What is changing for Wikidata?

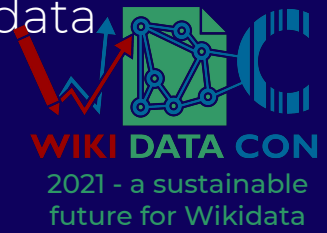
- Data quality is becoming more crucial and the pressure continues to rise
- More contributions will come from outside wikidata.org
- The Wikibase Ecosystem is growing

What should we focus on now?

- Empowering editors to increase data quality
- Equity in decision making
- Increasing re-use for impact
- Knowledge of and about underrepresented languages

Empowering editors to increase data quality

- We want to ensure that the content on Wikidata is of high quality for anyone who re-uses our data.
- We must ensure that the socio-technical system we have is set up in a way that helps editors increase the quality of Wikidata's existing data and contribute high-quality new data.



Equity in decision making

- We want to ensure that a diverse set of voices are heard when fundamental decisions are made on and about Wikidata to give more people a voice.

Increasing re-use for impact

- We want more people to benefit from the data Wikidata provides.
- Our data is available for anyone to re-use. We want to especially support projects that are aligned with our mission and values and/or that give back to Wikidata.

Knowledge of and about underrepresented languages

- More people should have access to knowledge and technology in their language. Data in their language is a fundamental building block for that.

What's important for Wikibase?

- To grow and thrive, Wikibase should be supported by a holistic ecosystem that includes diverse and interconnected knowledge hubs, a community of extension and tool developers, and more.

What should we focus on now?

- Improving onboarding into the Wikibase Ecosystem
- Enabling more extension and tool building
- Interconnections between Wikibases

Improving onboarding into the Ecosystem

- We want to ensure that more projects can independently onboard themselves into the Wikibase Ecosystem by reducing the complexity of software setup, data importing, & maintenance.
- This should make it possible for less-resourced projects to engage with Wikibase.

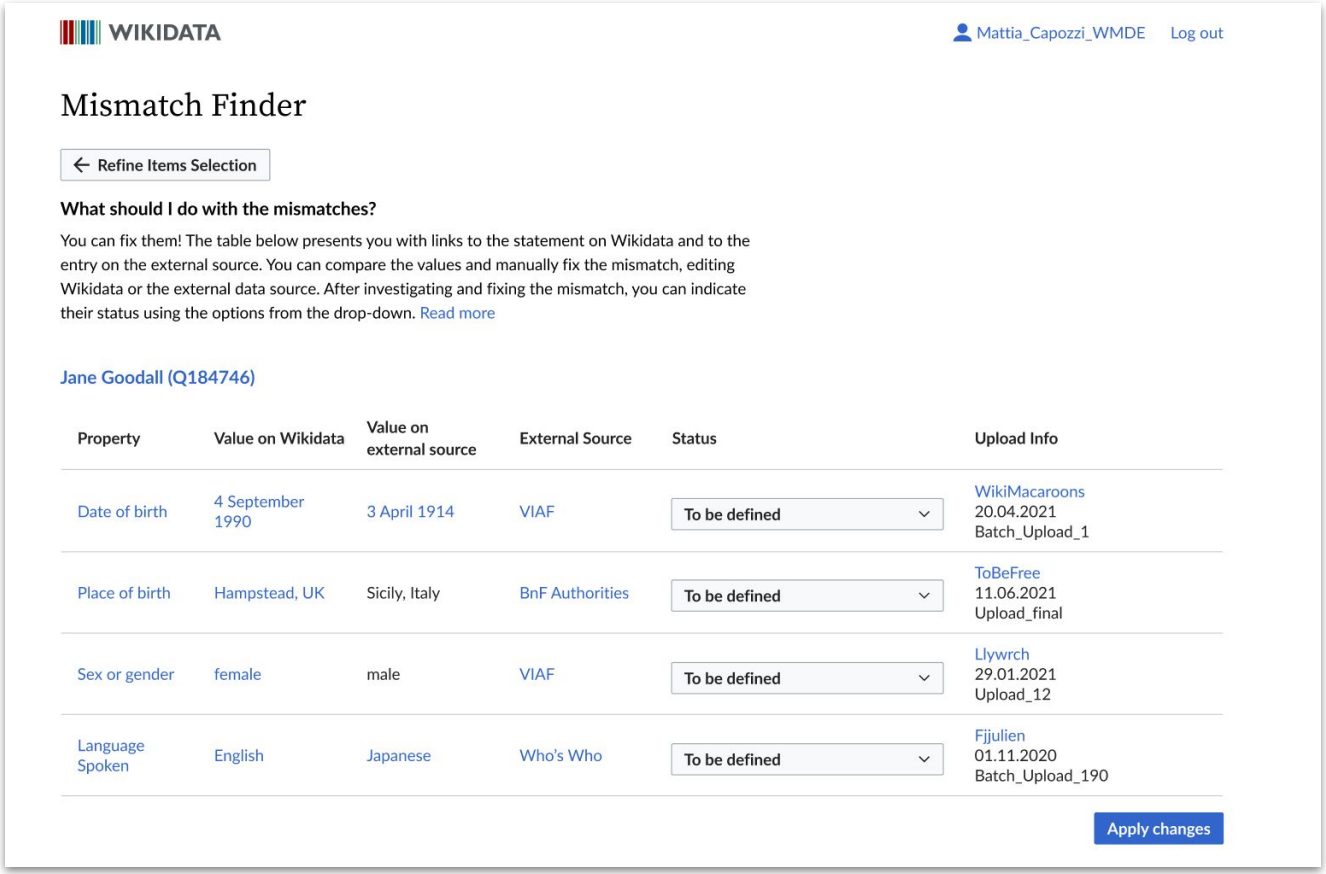
Enabling more extension and tool building

We need to ensure a healthy ecosystem of extensions as well as tools/custom interfaces based on WB APIs can develop around Wikibase, extending the functionality of the software for more use cases.

Interconnections between Wikibases

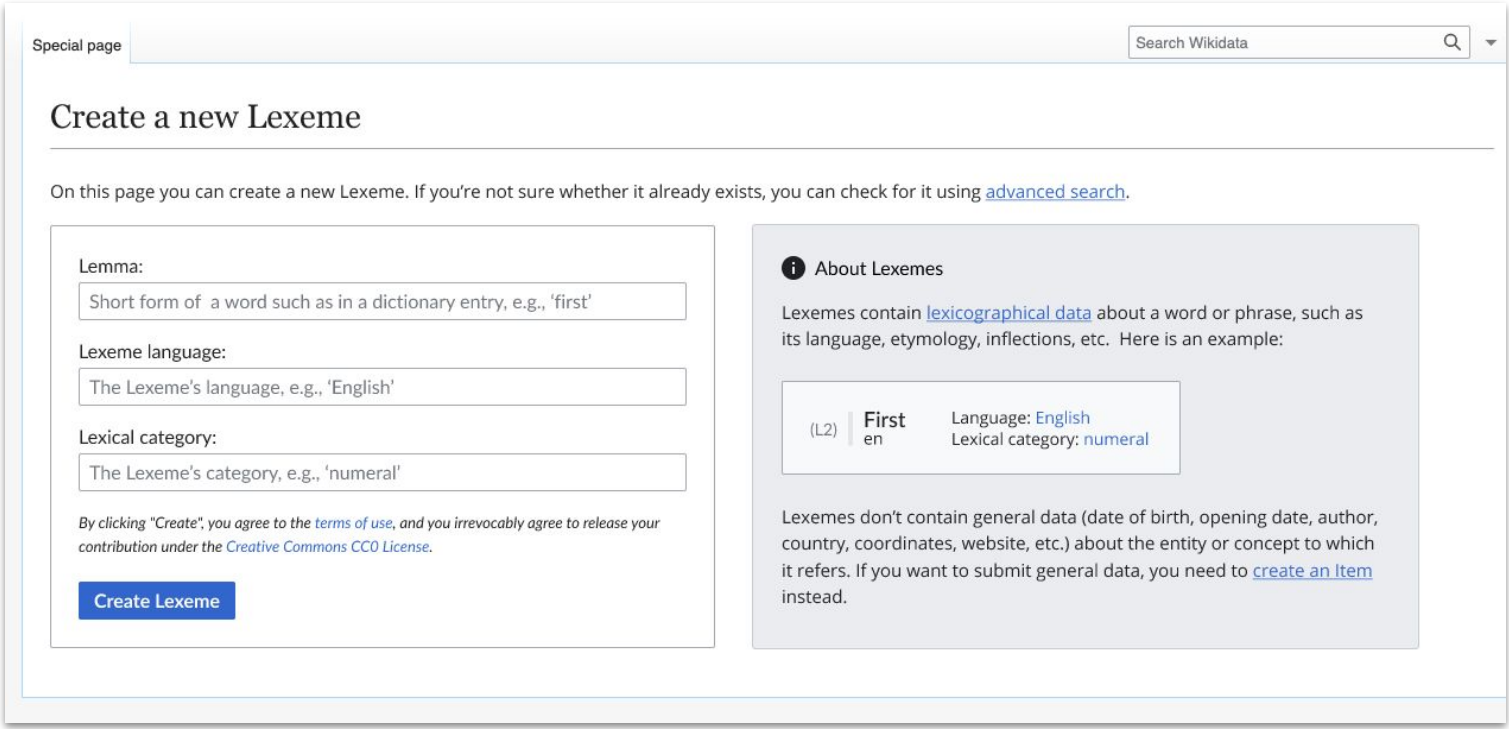
- We want to ensure Wikibases can connect with one another and Wikidata to form a linked open data web more powerful than the sum of its parts.

What’s next?



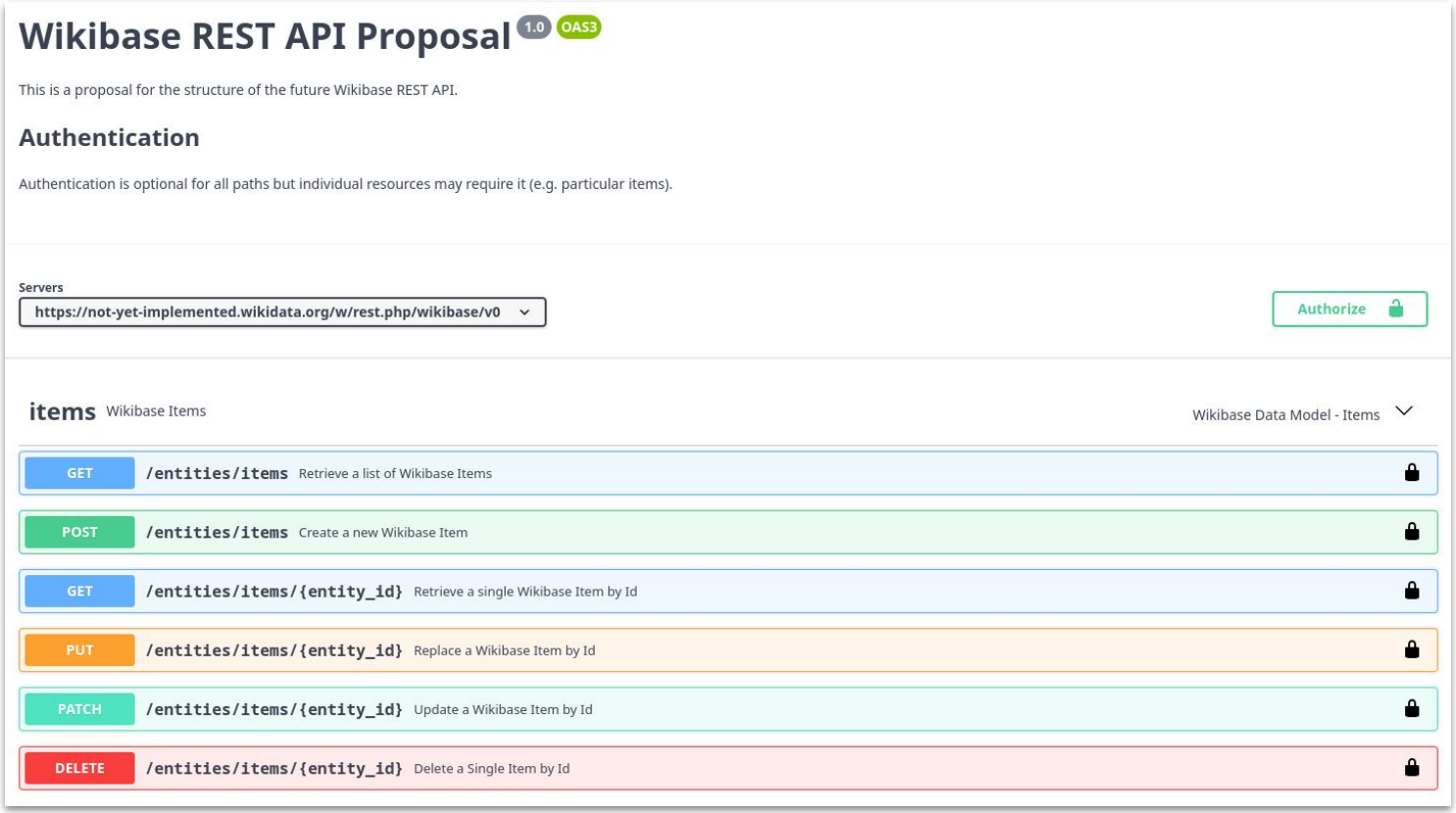
[Wikidata: Mismatch Finder]

WIKI DATA CON
2021 - a sustainable future for Wikidata

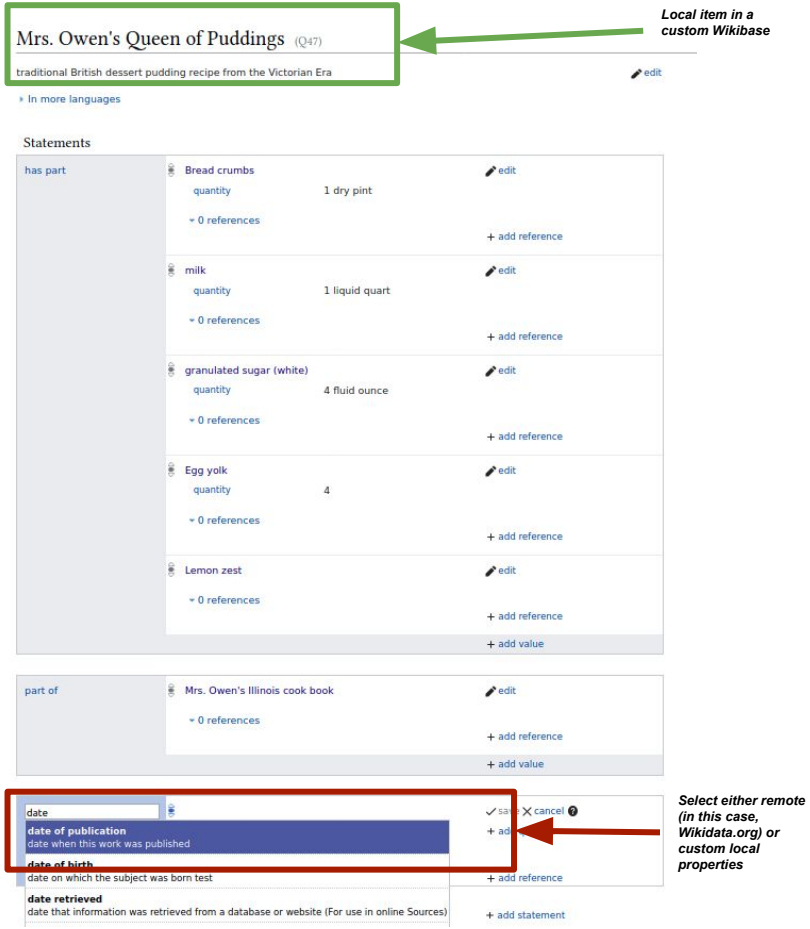


Special:NewLexeme improvements

WIKI DATA CON
2021 - a sustainable future for Wikidata



WIKI DATA CON
2021 - a sustainable future for Wikidata



We will be conducting user testing of an experimental version of Federated Properties v2 to gather feedback and ensure proper user experience and functionality with core extensions/tools.

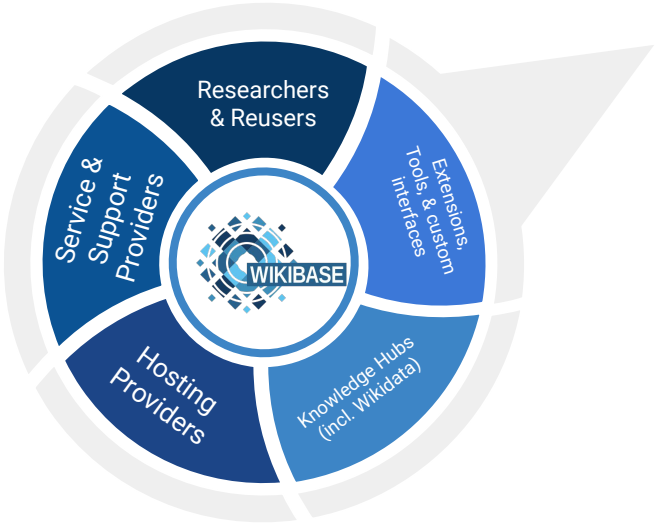
WIKI DATA CON
2021 - a sustainable future for Wikidata

Wikibase.Cloud

(Beta)

WIKI DATA CON

2021 - a sustainable future for Wikidata



A Holistic Wikibase Ecosystem

- We will engage directly with institutions, users, and developers in the ecosystem to find out what their needs, ideas, and plans are for extensions and tools built on Wikibase
- Our first requirements gathering session will happen during the WikidataCon Wikibase track

“Our admiration for Wikidata and for the people who work with it knows no bounds.

Having a single source of well modelled, massively interlinked, well managed data that anyone can query at the press of a button is a real thing of wonder.”

-- UK Parliament data team

"A lot of [research] projects get built, and often they disappear. There's all that work and data, and it just goes away. So that's also what we asked ourselves: how can we do this in a stable way that can go forward?

And that is – for us – one of the most exciting things about working with Wikibase. Wikibase has a sustainable community with a lot of people working together."

-- Dean Rehberger, Enslaved.org

WIKI DATA CON
2021 - a sustainable future for Wikidata

Thank you!

Get in touch with us:

Lydia Pintscher
lydia.pintscher@wikimedia.de

Sam Alipio
samantha.alipio@wikimedia.de