

# Wikidata and Wikibase

What happened? Where are we going?

**Lydia Pintscher**

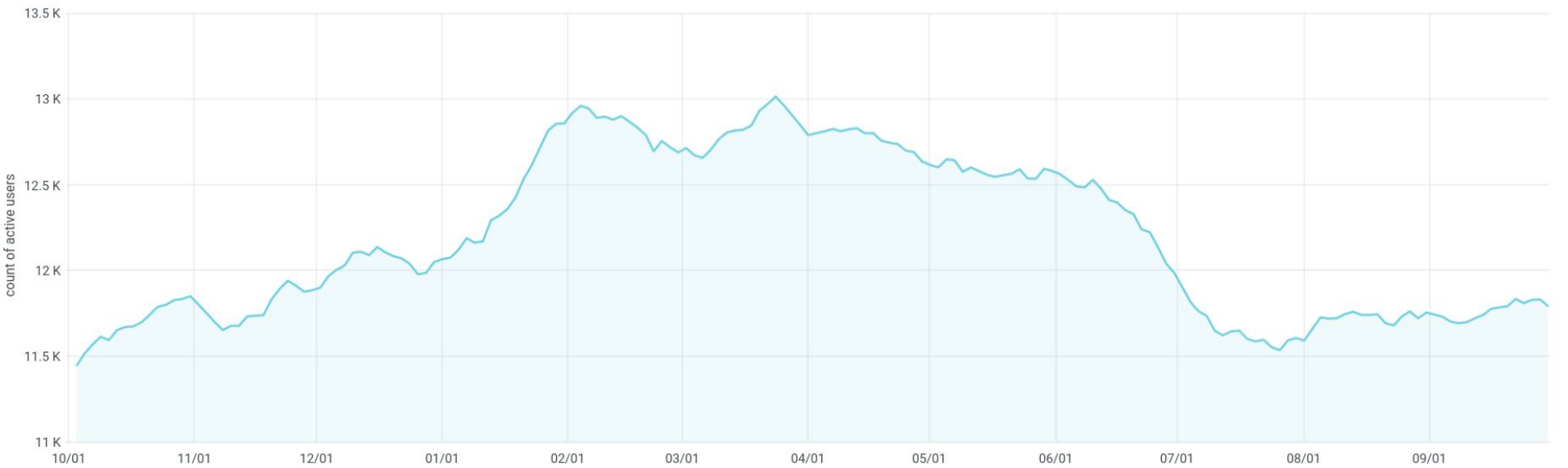
Product Manager for  
Wikidata

**Sam Alipio**

Product Manager for  
Wikibase Ecosystem

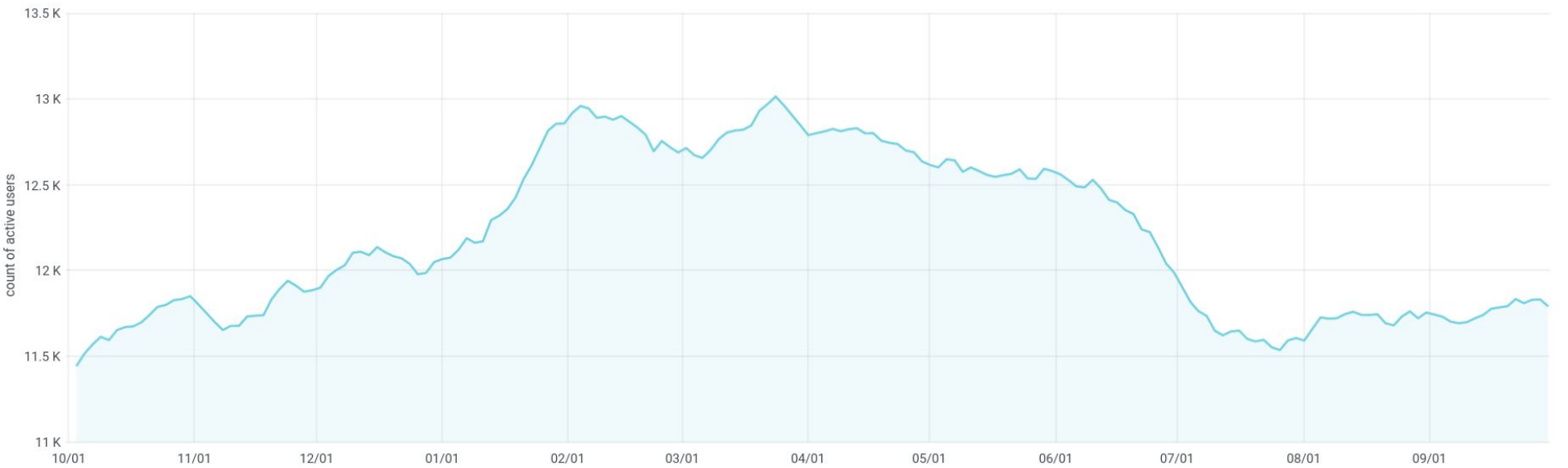
Where are we?

11.8k active  
editors

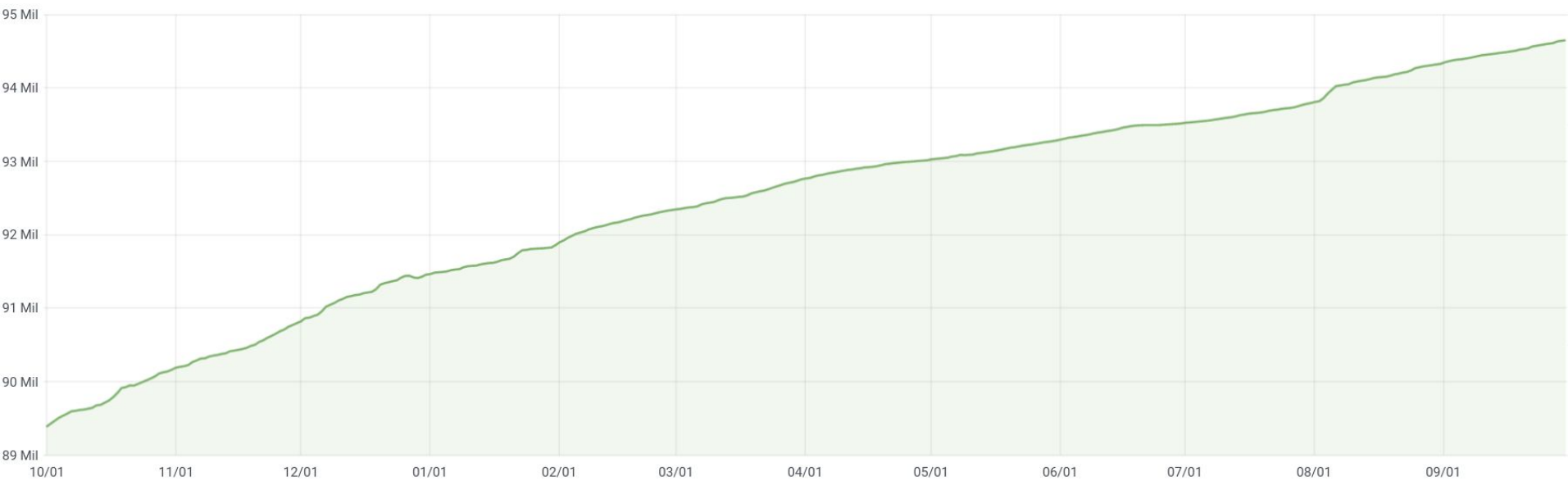


11.8k active  
editors

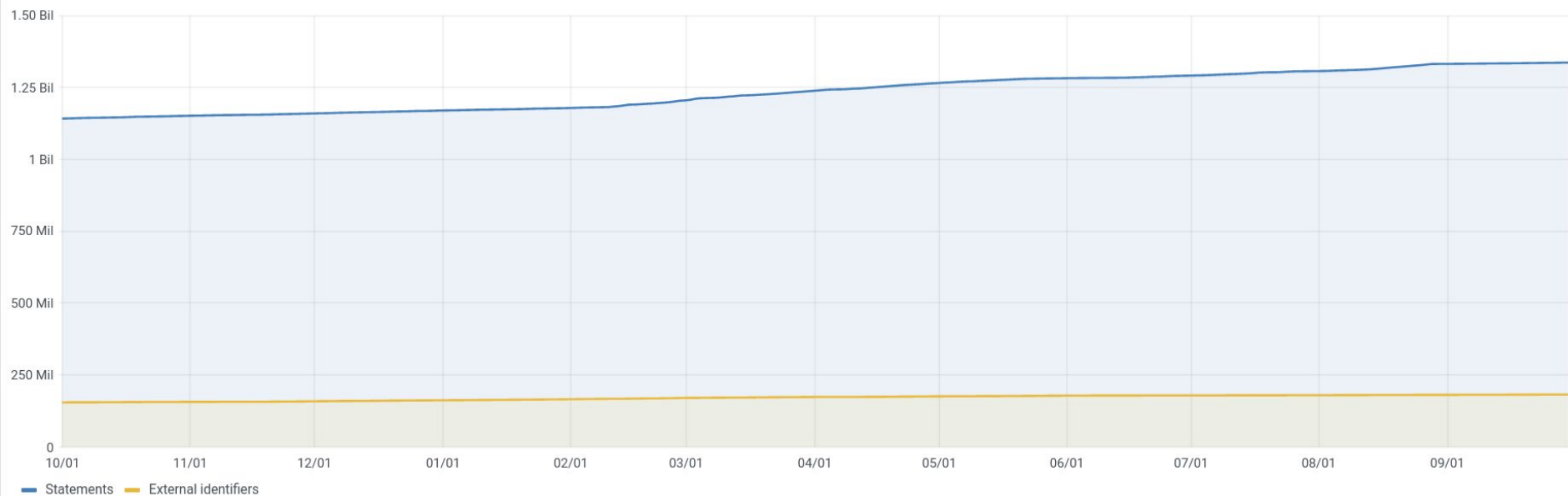
1.5 Billion  
edits



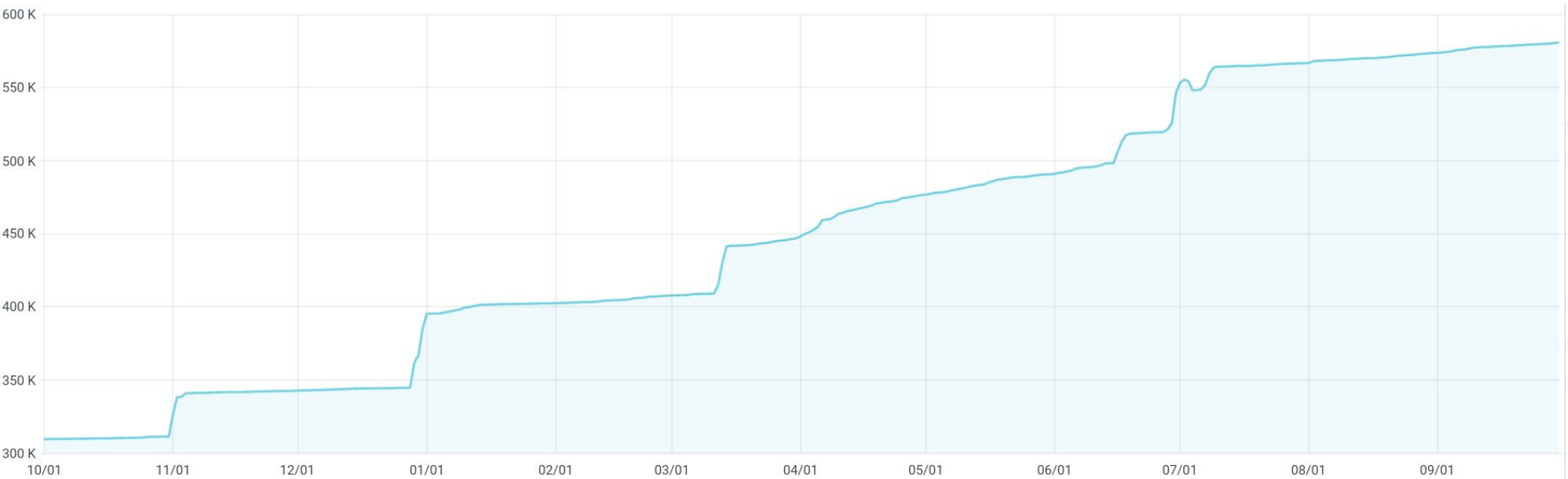
# 94 Million Items



# 1.34 Billion Statements



580k  
Lexemes



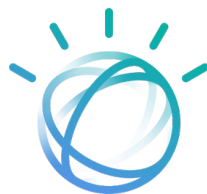
8 Million  
SPARQL  
queries/day

71% of  
Wikipedia  
articles

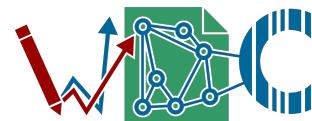
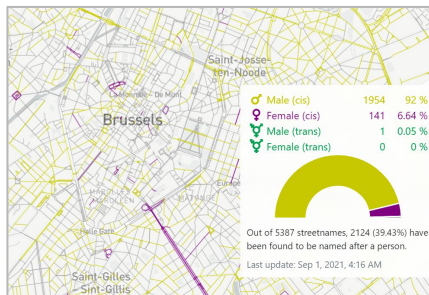
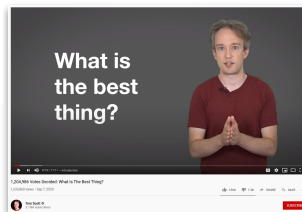
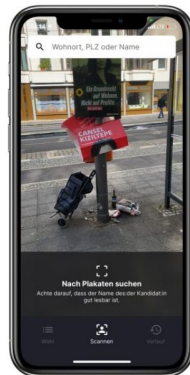
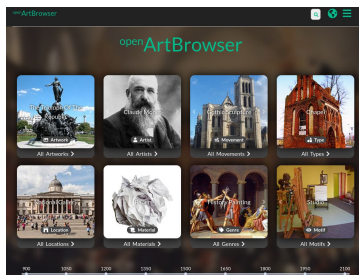




mySociety

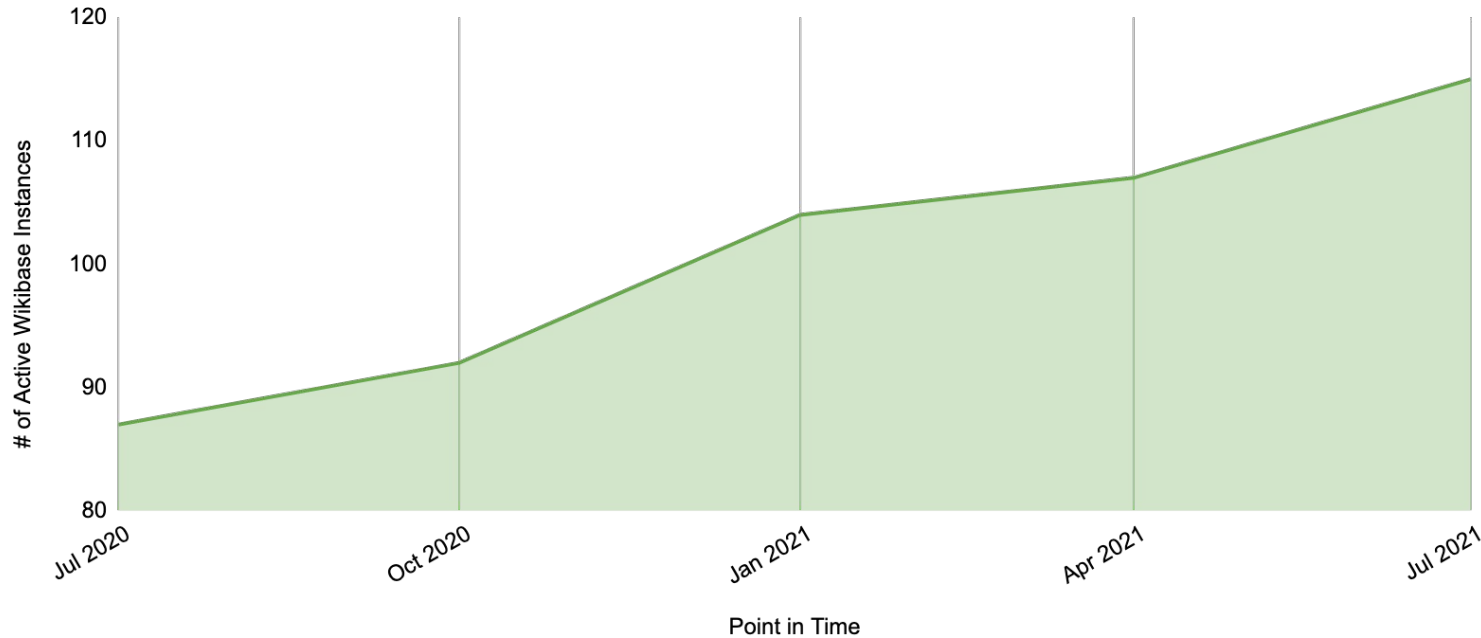


Quora



WIKI DATA CON  
2021 - a sustainable  
future for Wikidata

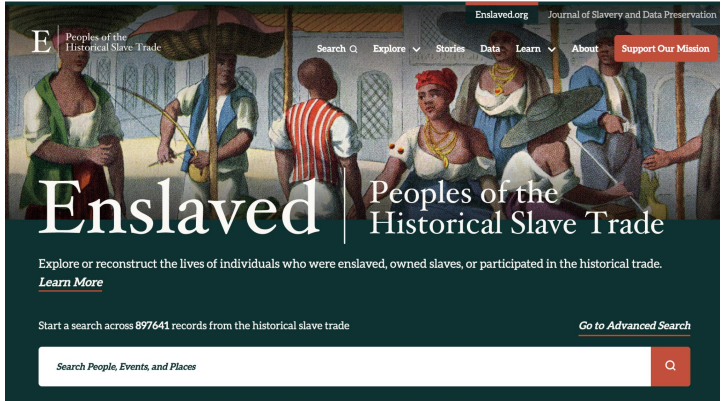
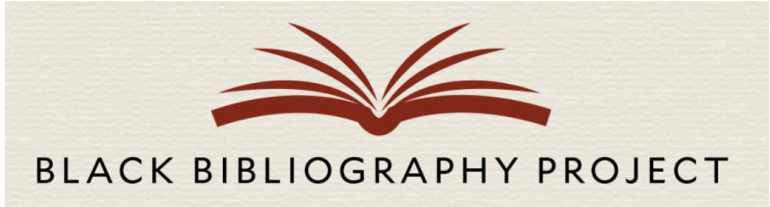
# The Wikibase Ecosystem continues to grow: There are 32% more active instances vs. a year ago.



# Examples of Wikibases in the ecosystem



Lingua Libre



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

# 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

# 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



# Item Quality Evaluator

Gets the quality scores of Wikidata Items and calculates their average value.

[How does this work?](#)

**For which Items should the average score be calculated?**

Item list

SPARQL

```
Q1234
Q23
Q5
```

One Item-identifier (like Q1234) on each line

**Assess Item Quality**

Made by the Wikidata development team. Licensed under [BSD 3-Clause License](#). [View Source](#). [Report an issue](#).

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

# 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
<a href="#">Saint Patrick's Day (Q181817)</a>	4.06
<a href="#">Woodstock Festival (Q164815)</a>	3.82
<a href="#">Saturnalia (Q207213)</a>	3.35
<a href="#">Feast of the Immaculate Conception (Q3538509)</a>	3.23
<a href="#">Oktoberfest (Q102400)</a>	3.08
<a href="#">San Fermín (Q829919)</a>	3.02

# 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>

# Query Service

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

# Wikidata Query Builder

The Wikidata Query Builder provides a visual interface for building a simple Wikidata query. It is ideal for users with little or no experience in [SPARQL](#), the powerful query language. The Query Builder doesn't offer SPARQL's full functionality, but you can always open your query in the Query Service, where you can view, edit or expand it via the link above the results. [Feedback is welcome here.](#)

## Query

Find all items...

With Without Property   Value  References

Include related values in the search (recommended)

and or

With Without Property   Value  References

Include related values in the search (recommended)

Add condition

## Settings

Limit the number of results to

Show IDs instead of labels (may prevent timeout)

Run query Get a shareable link

<https://query.wikidata.org/querybuilder>



# 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](#)



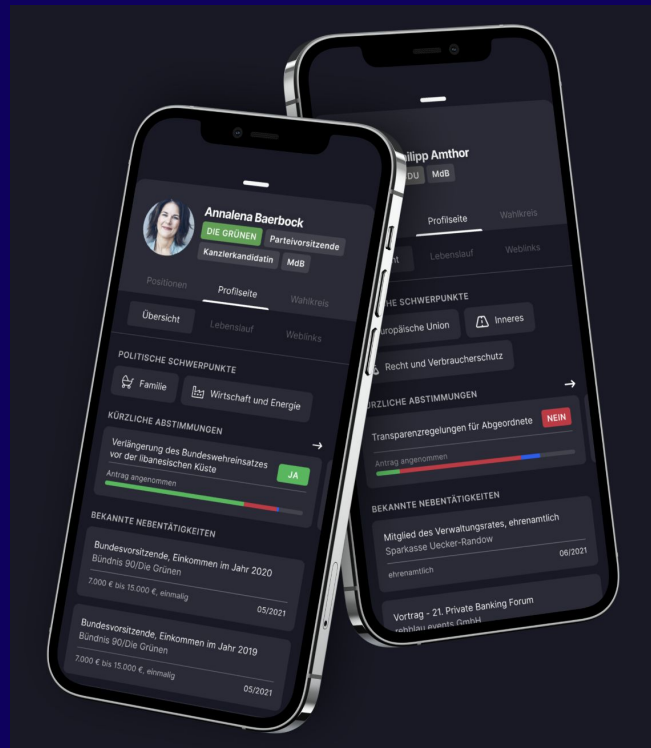
**Forms**  
● Covered  
● Missing



**Tokens**  
● Covered  
● Missing

[Wikidata:Lexicographical\\_coverage](#)

What are  
people  
building?



<https://facethefacts.app>

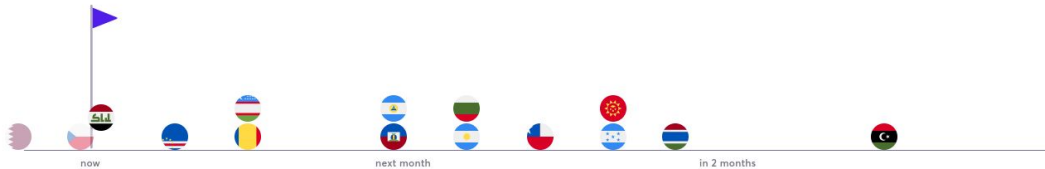


# The Election Tracker




Keep track of upcoming national elections around the world. Updated regularly.

This data story is part of our Visualizing Wikipedia project 

There are 18 elections in the next three months



## Next General Elections

TOMORROW		2021 Iraqi parliamentary election	▼
IN 8 DAYS		2021 Cape Verdean presidential election	▼
OCT 24		2021 Chadian parliamentary election	▼

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

# Get your voice heard!

The easiest way to contact your government.

Start the conversation now!

Explore by country

Explore by topic

Govdirectory is a crowdsourced and fact checked directory of official governmental online accounts and services.



"Who are the largest CO2 emitters in Sweden?"

Ask your own government!

<https://govdirectory.org>



## The Simpsons

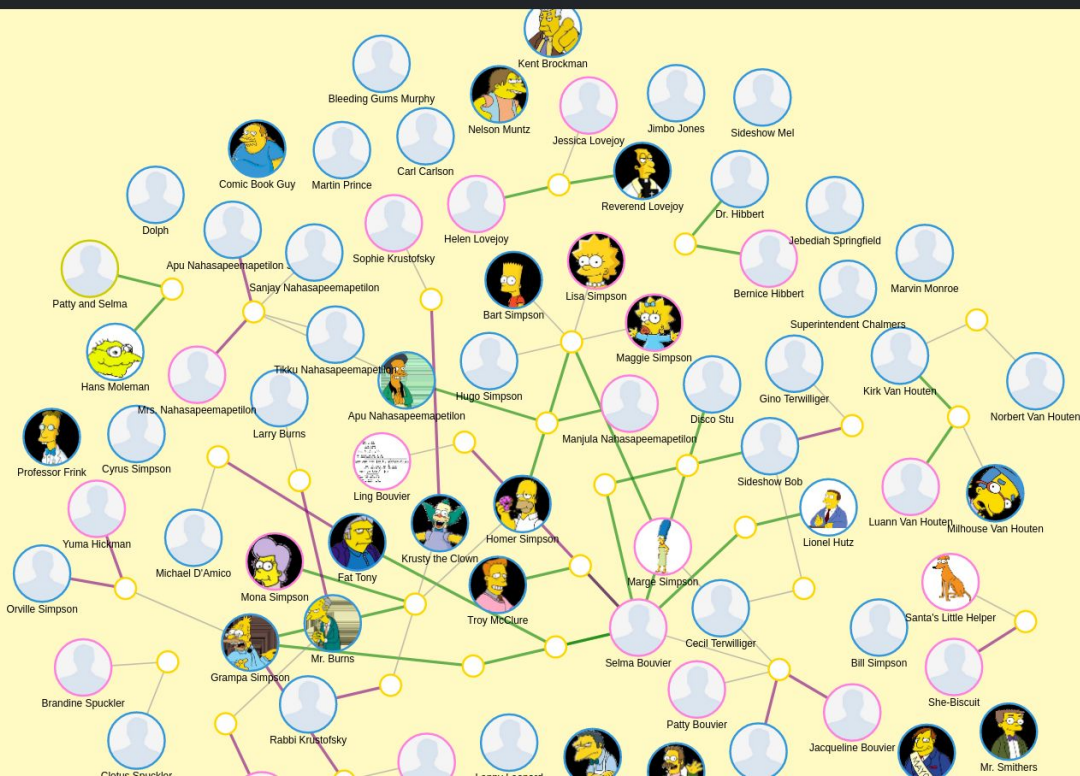


[Wikipedia](#) [Wikidata](#)

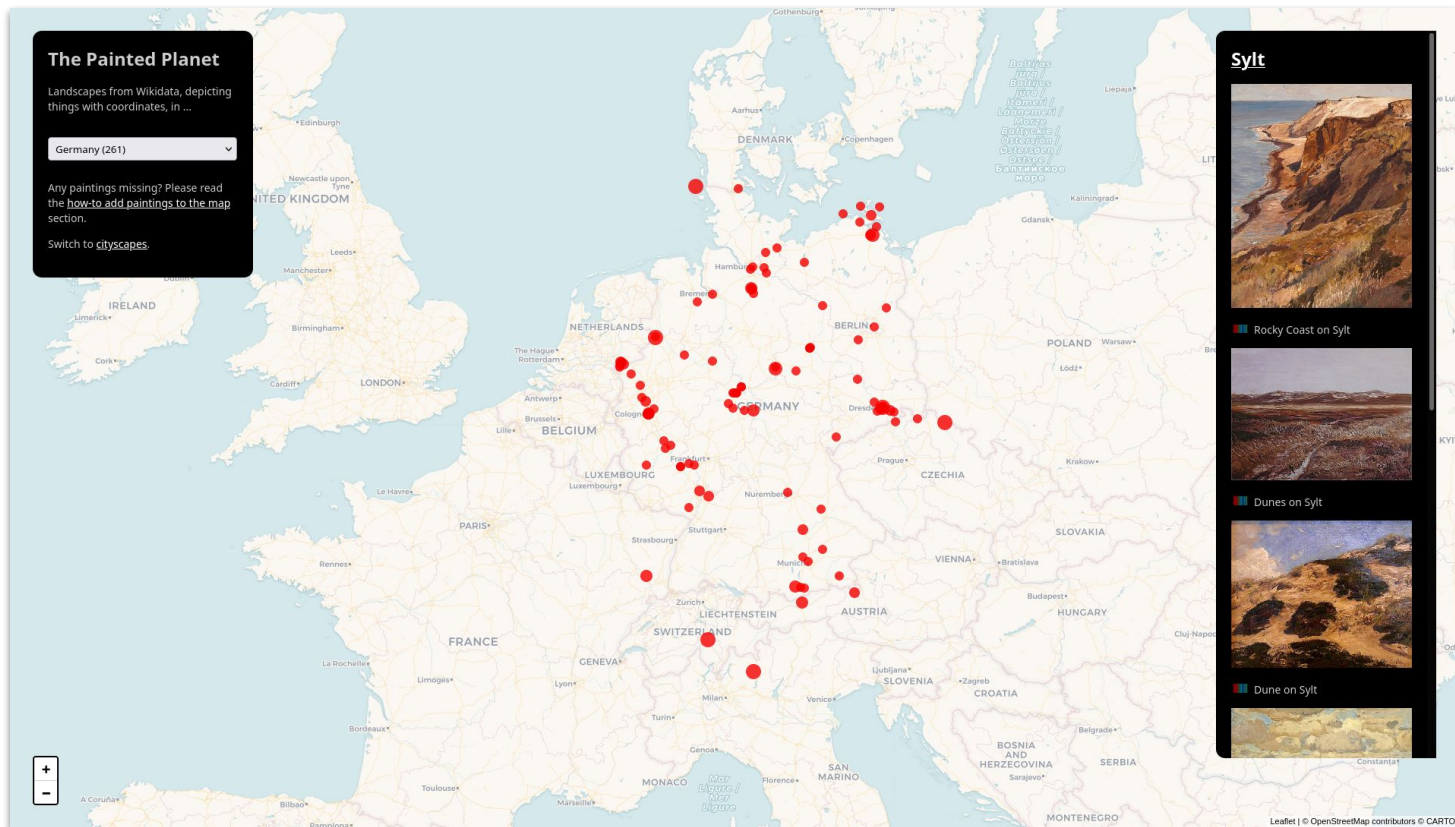
*The Simpsons* is an American animated sitcom created by Matt Groening for the Fox Broadcasting Company. The series is a satirical depiction of American life, epitomized by the Simpson family, which consists of Homer, Marge, Bart, Lisa, and Maggie. The show is set in the fictional town of Springfield and parodies American culture and society, television, and the human condition.

**Character** **Actor**

- [Homer Simpson](#)
- [Mona Simpson](#)
- [Grampa Simpson](#)
- [Marge Simpson](#)
- [Bart Simpson](#)
- [Lisa Simpson](#)
- [Maggie Simpson](#)
- [Hugo Simpson](#)
- [Herbert Powell](#)
- [Abbie Simpson](#)



<https://linkedpeople.net>



<https://hicsuntleones.nl>

# What is the best thing?



0:13 / 11:11 • Introduction



1,204,986 Votes Decided: What Is The Best Thing?

1,333,860 views • Sep 7, 2020

👍 136K    💬 1.5K    ➦ SHARE    ⌵ SAVE    ⋮



**Tom Scott** ✓  
3.19M subscribers

SUBSCRIBE



# Release Strategy & Infrastructure

- 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











Mrs. Owen's Queen of Puddings (Q47)

traditional British dessert pudding recipe from the Victorian Era

 edit

[In more languages](#)

#### Statements

has part	 Bread crumbs	 edit
	quantity	1 dry pint
	<a href="#">+ 0 references</a>	<a href="#">+ add reference</a>
	 milk	 edit
	quantity	1 liquid quart
<a href="#">+ 0 references</a>	<a href="#">+ add reference</a>	
 granulated sugar (white)	 edit	
quantity	4 fluid ounce	
<a href="#">+ 0 references</a>	<a href="#">+ add reference</a>	
 Egg yolk	 edit	
quantity	4	
<a href="#">+ 0 references</a>	<a href="#">+ add reference</a>	
 Lemon zest	 edit	
<a href="#">+ 0 references</a>	<a href="#">+ add reference</a>	
	<a href="#">+ add value</a>	

part of	 Mrs. Owen's Illinois cook book	 edit
	<a href="#">+ 0 references</a>	<a href="#">+ add reference</a>
		<a href="#">+ add value</a>

<input type="text" value="date"/>	 ✓	 X cancel
<b>date of publication</b> date when this work was published	 +	
<b>date of birth</b> date on which the subject was born test	 +	<a href="#">+ add reference</a>
<b>date retrieved</b> date that information was retrieved from a database or website (For use in online Sources)	 +	<a href="#">+ add statement</a>

Local item in a custom Wikibase

Select either remote (in this case, Wikidata.org) or custom local properties

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).

# 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



# What is important for Wikidata?

## Wikidata community

creates and maintains content



## Re-users

give back feedback and awareness



# 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?

# Mismatch Finder

[← Refine Items Selection](#)

## What should I do with the mismatches?

You can fix them! The table below presents you with links to the statement on Wikidata and to the entry on the external source. You can compare the values and manually fix the mismatch, editing Wikidata or the external data source. After investigating and fixing the mismatch, you can indicate their status using the options from the drop-down. [Read more](#)

### Jane Goodall (Q184746)

Property	Value on Wikidata	Value on external source	External Source	Status	Upload Info
Date of birth	4 September 1990	3 April 1914	VIAF	To be defined	WikiMacaroons 20.04.2021 Batch_Upload_1
Place of birth	Hampstead, UK	Sicily, Italy	BnF Authorities	To be defined	ToBeFree 11.06.2021 Upload_final
Sex or gender	female	male	VIAF	To be defined	Llywrch 29.01.2021 Upload_12
Language Spoken	English	Japanese	Who's Who	To be defined	FjJulien 01.11.2020 Batch_Upload_190

[Apply changes](#)

[Wikidata:](#)  
[Mismatch](#)  
[Finder](#)



## Create a new Lexeme

On this page you can create a new Lexeme. If you're not sure whether it already exists, you can check for it using [advanced search](#).

Lemma:

Lexeme language:

Lexical category:

*By clicking "Create", you agree to the [terms of use](#), and you irrevocably agree to release your contribution under the [Creative Commons CC0 License](#).*

Create Lexeme

### About Lexemes

Lexemes contain [lexicographical data](#) about a word or phrase, such as its language, etymology, inflections, etc. Here is an example:

(L2)	<b>First</b>	Language: <a href="#">English</a>
	en	Lexical category: <a href="#">numeral</a>

Lexemes don't contain general data (date of birth, opening date, author, country, coordinates, website, etc.) about the entity or concept to which it refers. If you want to submit general data, you need to [create an Item](#) instead.

# Wikibase REST API Proposal 1.0 OAS3

This is a proposal for the structure of the future Wikibase REST API.

## Authentication

Authentication is optional for all paths but individual resources may require it (e.g. particular items).







Servers

▼

Authorize 

**items** Wikibase Items

Wikibase Data Model - Items ▼

GET	/entities/items	Retrieve a list of Wikibase Items	
POST	/entities/items	Create a new Wikibase Item	
GET	/entities/items/{entity_id}	Retrieve a single Wikibase Item by Id	
PUT	/entities/items/{entity_id}	Replace a Wikibase Item by Id	
PATCH	/entities/items/{entity_id}	Update a Wikibase Item by Id	
DELETE	/entities/items/{entity_id}	Delete a Single Item by Id	











Mrs. Owen's Queen of Puddings (Q47)

traditional British dessert.pudding recipe from the Victorian Era

 edit

[In more languages](#)

#### Statements

has part	 Bread crumbs	 edit
	quantity 1 dry pint	
	+ 0 references	+ add reference
	 milk	 edit
	quantity 1 liquid quart	
+ 0 references	+ add reference	
 granulated sugar (white)	 edit	
quantity 4 fluid ounce		
+ 0 references	+ add reference	
 Egg yolk	 edit	
quantity 4		
+ 0 references	+ add reference	
 Lemon zest	 edit	
+ 0 references	+ add reference	
	+ add value	

part of	 Mrs. Owen's Illinois cook book	 edit
	+ 0 references	+ add reference
		+ add value

date	 X cancel
<b>date of publication</b>	 +
date when this work was published	
<b>date of birth</b>	 +
date on which the subject was born test	+ add reference
<b>date retrieved</b>	 +
date that information was retrieved from a database or website (For use in online Sources)	+ add statement

Local item in a custom Wikibase

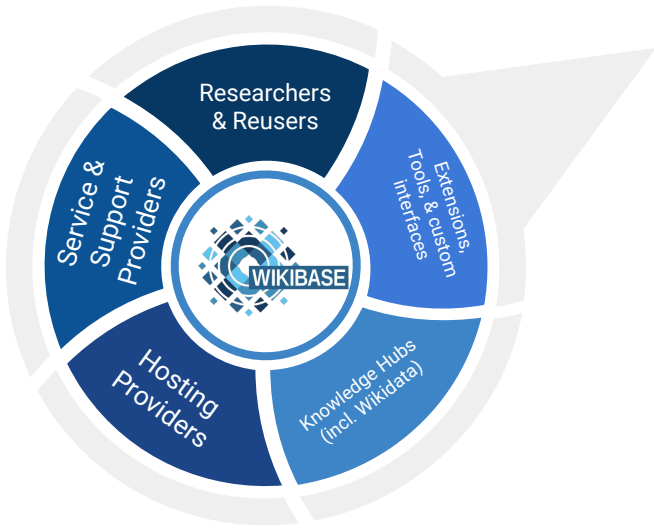
Select either remote (in this case, Wikidata.org) or custom local properties

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.

# Wikibase.Cloud

*(Beta)*





*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



# Thank you!

Get in touch with us:

**Lydia Pintscher**

[lydia.pintscher@wikimedia.de](mailto:lydia.pintscher@wikimedia.de)

**Sam Alipio**

[samantha.alipio@wikimedia.de](mailto:samantha.alipio@wikimedia.de)