Let’s keep in sync

An intro to Mismatch Finder

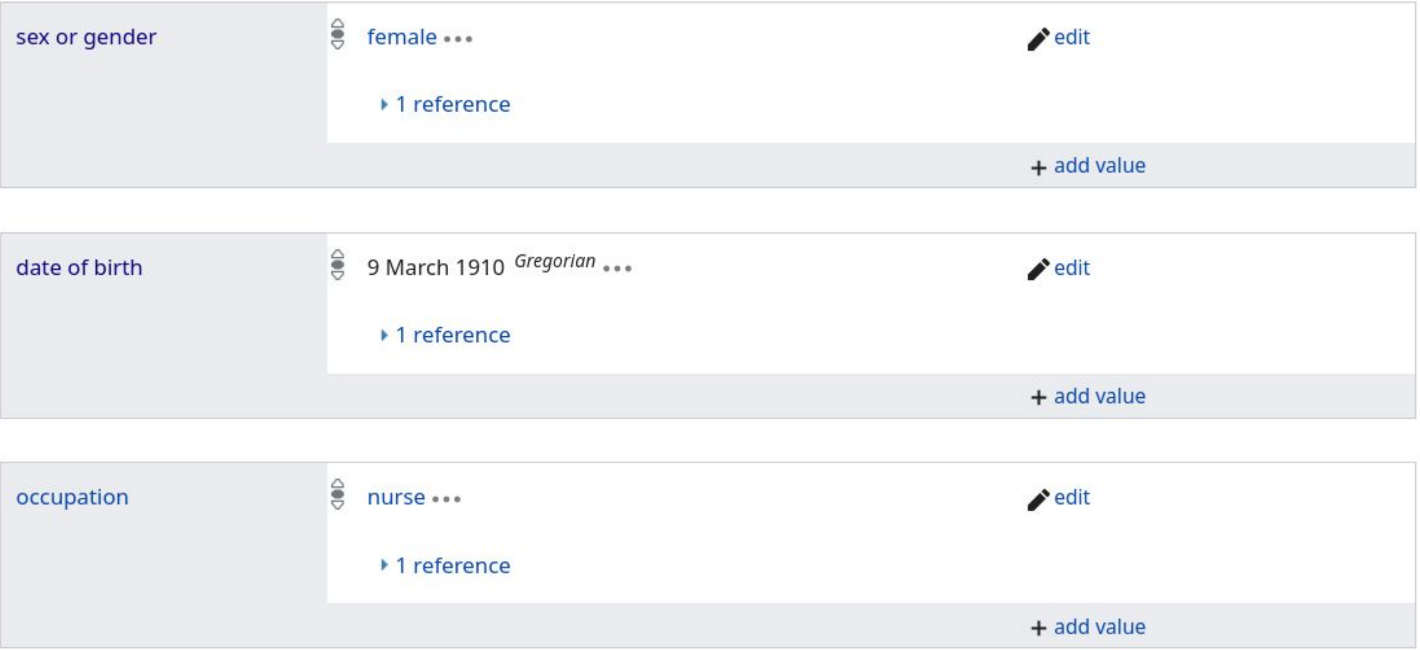
Lydia Pintscher
Product Manager for Wikidata
@nightrose

Mattia Capozzi
Wikidata UX Researcher

WIKIDATA

DATA RE–
14 - 24 March 2022
USE DAYS

It’s a cruel world out there…



sex or gender: female

1 reference

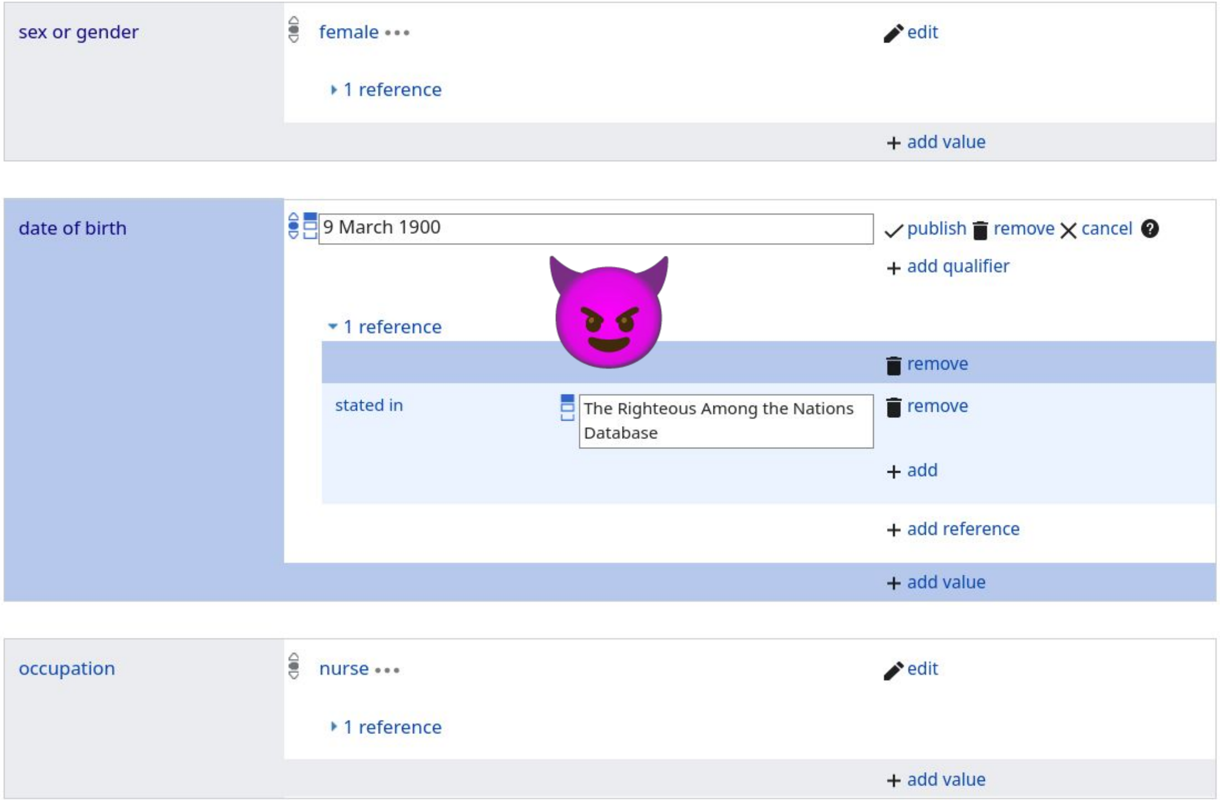
date of birth: 9 March 1910 Gregorian

1 reference

occupation: nurse

1 reference

Vandals are screwing up our data in subtle but evil ways



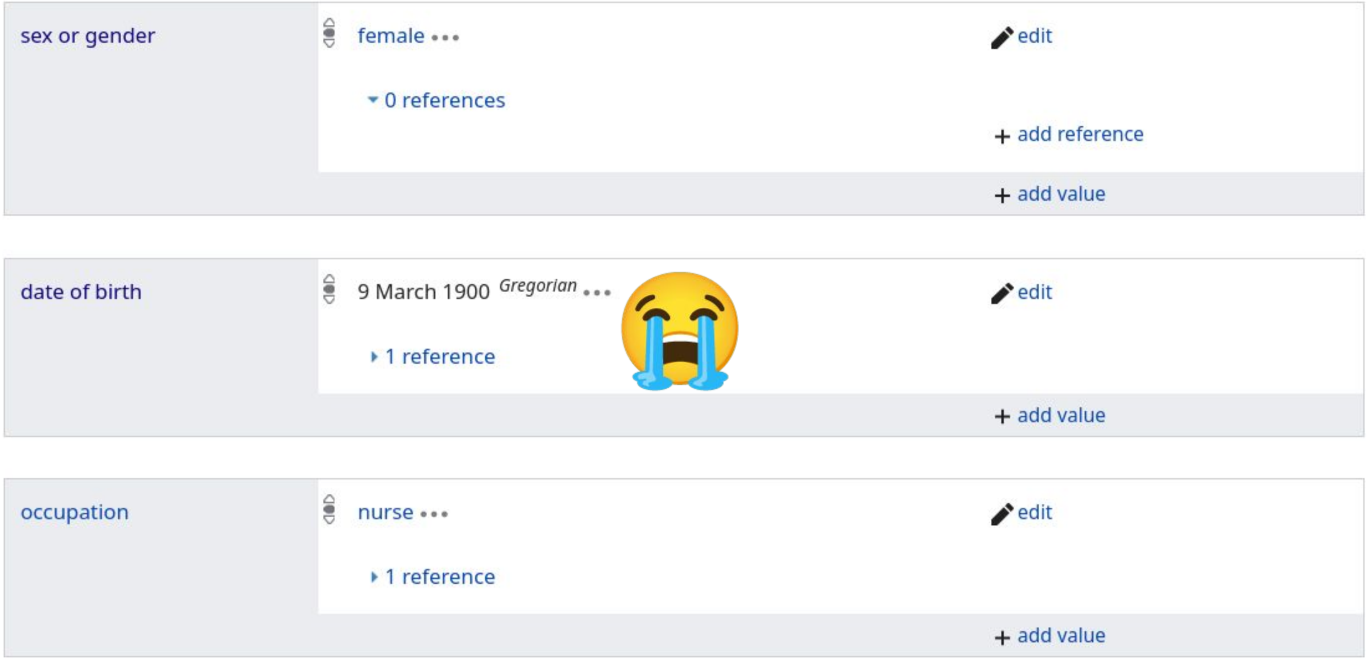
sex or gender: female

date of birth: 9 March 1900

stated in: The Righteous Among the Nations Database

occupation: nurse

Vandals are screwing up our data in subtle but evil ways



sex or gender

female

0 references

edit

add reference

add value

date of birth

9 March 1900 *Gregorian*

1 reference

edit

add value

occupation

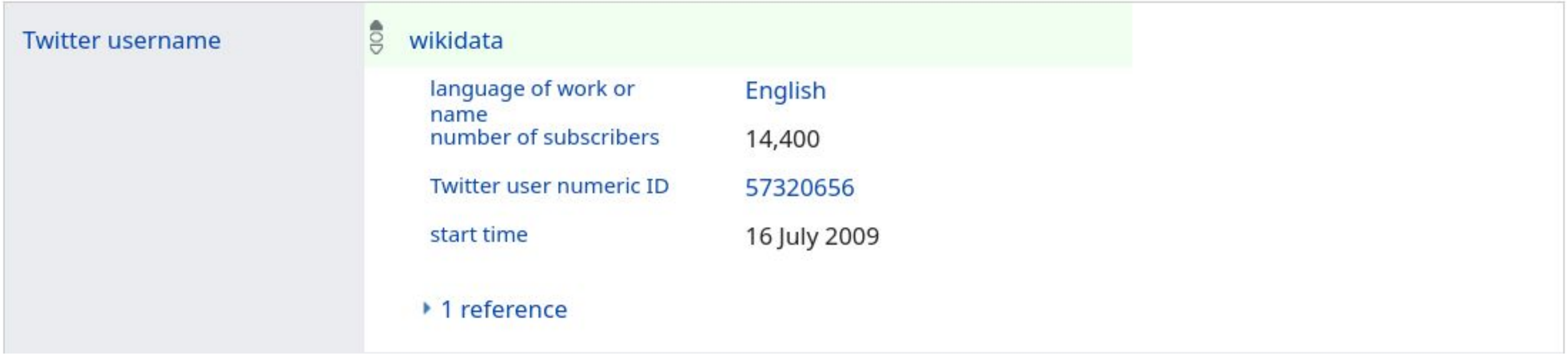
nurse

1 reference

edit

add value

Vandals are screwing up our data in subtle but evil ways



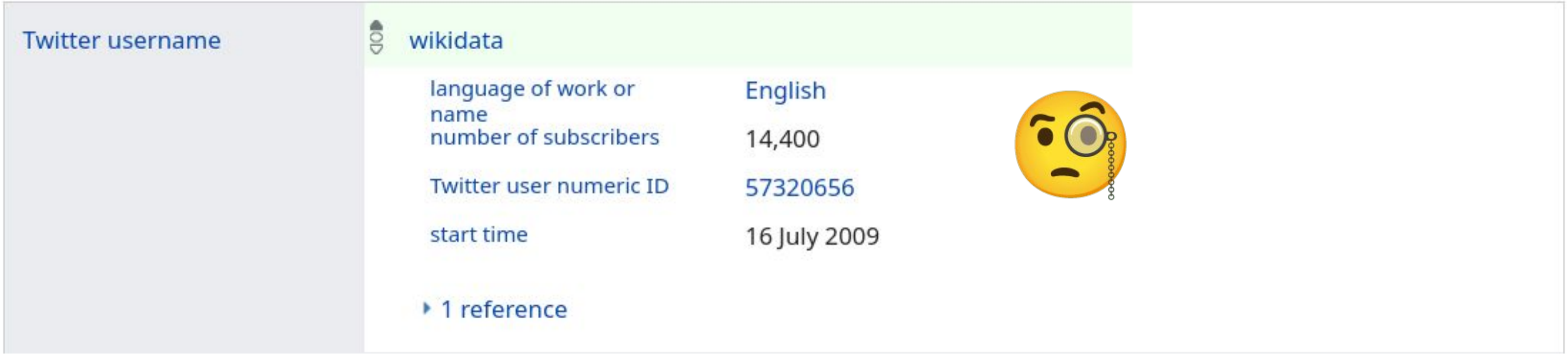
Twitter username

wikidata

language of work or name	English
number of subscribers	14,400
Twitter user numeric ID	57320656
start time	16 July 2009

▸ 1 reference

The world is changing around us
How dare it!



Twitter username

wikidata

language of work or name	English
number of subscribers	14,400
Twitter user numeric ID	57320656
start time	16 July 2009

▸ 1 reference

The world is changing around us
How dare it!

““I really wonder whether there is a way of simplifying the whole work of detecting mismatch, because it is tedious and energy consuming.”

“People are really frustrated spending time looking for mismatches in a non-structured way.”

“The work I do finding mismatches is rarely reusable, so I cannot always teach it to other people.”

Reusers wanna help but don’t really know how

Mismatch Finder to the rescue! 🦸

How Mismatch Finder works

1. Someone has a way to automatically and at scale compare Wikidata’s data against another database/website/…
2. They prepare a CSV file with these mismatches
3. They upload the file to Mismatch Finder
4. Others can review these mismatches in Mismatch Finder and figure out weather the issue is in Wikidata or the other data source and make edits accordingly or report the error to the other data source

Examples

More conventional:

- Date of birth statements for German authors between Wikidata and German National Library
- Band member names between Wikidata and MusicBrainz

Less conventional but still <3:

- Local infobox data from English Wikipedia and Wikidata
- User reported errors from a website using data from Wikidata

💪

Better data quality for Wikidata, clearer ways for re-users to give back and a more robust Linked Open Data web for everyone

Demo

Mismatch finder website, gadget and API

How you can help

Wanna review?

- Enable the user script
- Use the Mismatch Finder website

Got a way to find issues in Wikidata’s data?

- Work with us to get them into the Mismatch Finder

Find documentation for everything at
Wikidata:Mismatch Finder

Thanks for your attention!

To stay in touch:

wikidata.org/wiki/Wikidata:Mismatch_Finder
mismatch-finder.toolforge.org

Lydia Pintscher
lydia.pintscher@wikimedia.de
@nightrose

Mattia Capozzi
mattia.capozzi@wikimedia.de

WIKIDATA

DATA RE–
14 - 24 March 2022
USE DAYS