

Increasing access to free and open knowledge for speakers of underserved languages on Wikipedia



Lucie-Aimée Kaffee
@frimelle

Find the slides here:

<http://tinyurl.com/fosdem-articleplaceholder>

State of languages online now

Usage Statistic for Content Languages for Websites



Number of Articles	Number of Wikis
5M+	1
1M - 4M	11
100K - 1M	44
10K - 100K	75
1K - 10K	108
0 - 1000	48

• Statistics at 12:00, 26 January 2016 (UTC)

1 000 000+ articles

No ↕	Language ↕	Language (local) ↕	Wiki ↕	Articles ↕
1	English	English	en	5,065,136
2	Swedish	Svenska	sv	2,747,186
3	Cebuano	Sinugboanong Binisaya	ceb	1,924,975
4	German	Deutsch	de	1,901,286
5	Dutch	Nederlands	nl	1,853,041
6	French	Français	fr	1,718,191
7	Russian	Русский	ru	1,284,991
8	Waray-Waray	Winaray	war	1,259,463
9	Italian	Italiano	it	1,249,218
10	Spanish	Español	es	1,228,955
11	Polish	Polski	pl	1,154,080
12	Vietnamese	Tiếng Việt	vi	1,142,383

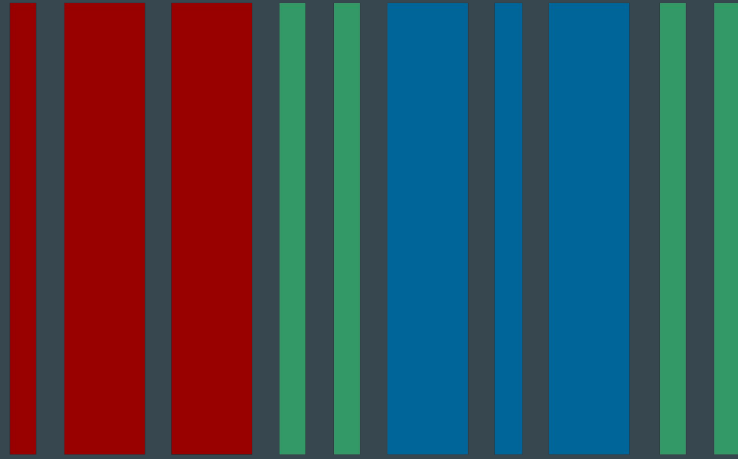
First-language speakers

Language	Speakers (millions)
Chinese	1,197
Spanish	399
English	335
Hindi	260
Arabic	242
Portuguese	203
Bengali	189
Russian	166

Language	Speakers (millions)
Japanese	128
Lahnda	88.7
Javanese	84.3
German	78.1
Korean	77.2
French	75.9
Japanese	128
Lahnda	88.7

**“this limit to the web’s
accessibility proves that it can be
just as insular and discriminative
as the modern world at large**

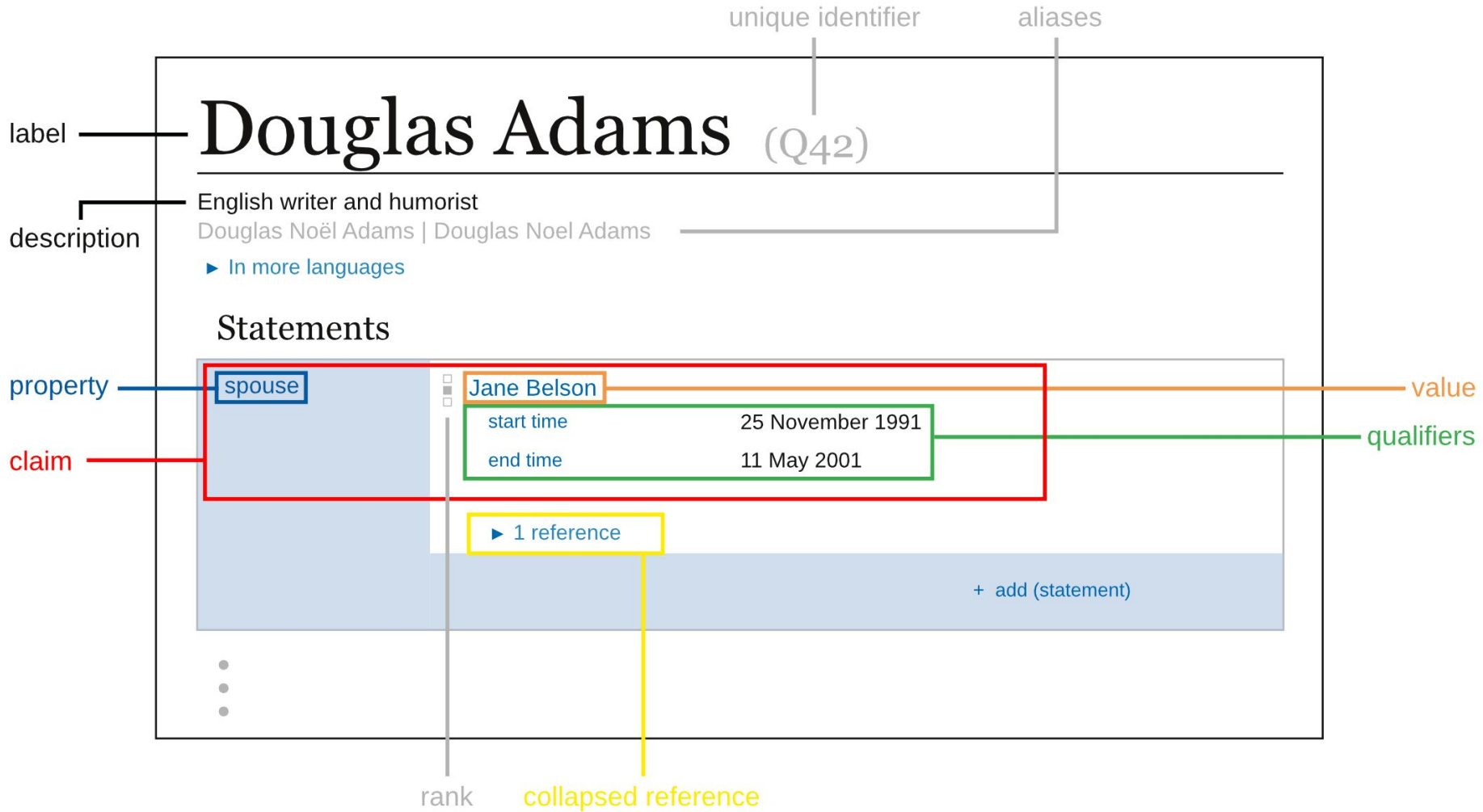
What are we going to do about this?



WIKIDATA

Basics: Wikidata

- Free knowledge base by the Wikimedia movement
- Started ~3 years ago
- Structured data
- Linked data
- Persistent identifiers
- Multiple languages
- People, places, events, and more
- Powered by Wikibase (open source MediaWiki extension)
- Data licenced under CC-0

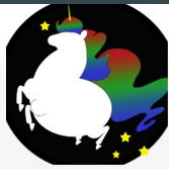




Article Placeholder

- MediaWiki extension
- In active development
- “Live” generated content pages for Wikipedia
- With multilingual data from Wikidata

<https://www.mediawiki.org/wiki/Extension:ArticlePlaceholder>



Ada Lovelace

Create an article

English mathematician, considered the first computer programmer

place of death

Marylebone^[18]

place of birth

London^[19]

cause of death

uterine cancer^[20]

occupation

mathematician^[21]

programmer

poet

computer scientist

engineer

sex or gender

female^{[22][23][24][25]}

father

Lord Byron



External Resources

National Portrait Gallery (London) person identifier	mp50509
--	---------

Example for Ada Lovelace

- User editable layout due to Lua modules
- Ordering of data happens on-wiki
- Pretty and useful default
- Encourages reader to create article
- May lower the number of bot created stub articles

How to contribute?

Besides editing and translating data on Wikidata

- Translate the extension itself on Translatewiki
 - Contribute to the code (PHP, JavaScript, CSS, Lua)
 - Work with Lua modules on-wiki
 - Help out with documentation/tests
-

Get in touch!

With Lucie

`[[User:Frimelle]]`

@frimelle

lucie.kaffee@gmail.com

<http://fuzzle.me>

With Wikidata

#wikidata on Freenode

@wikidata

wikidata@lists.wikimedia.org

www.wikidata.org/wiki/Wikidata:Project_chat

This slides are licensed under the [Creative Commons Attribution-Share Alike 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/) license.

Slide 3: Usage of content languages for websites, http://w3techs.com/technologies/overview/content_language/all (last visited Jan. 29, 2016).

Slide 4: List of Wikipedias, https://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=15272498 (last visited Jan. 29, 2016).

Slide 5: Table 3. Languages with at least 50 million first-language speakers, <https://www.ethnologue.com/statistics/size> (last visited Jan. 29, 2016).

Slide 6: The Internet Isn't Available in Most Languages, <http://www.theatlantic.com/technology/archive/2015/11/the-internet-isnt-available-in-most-languages/417393/> (last visited Jan. 29, 2016).

Slide 8: [Wikidata logo](#), [[User:Planemad]], public domain

Slide 9: [WikidataTermDescriptorDiagram](#), Charlie Kritschmar [[User:Incabell]] CC-BY-SA

Slide 10: [Magic Unicorn](#), Logo of the ArticlePlaceholder extension, Charlie Kritschmar [[User:Incabell]] CC-BY-SA