WDProp: Web Application to Analyse Multilingual Aspects of Wikidata Properties

OpenSym 2021, 15th-17th September, 2021

John Samuel

CPE Lyon



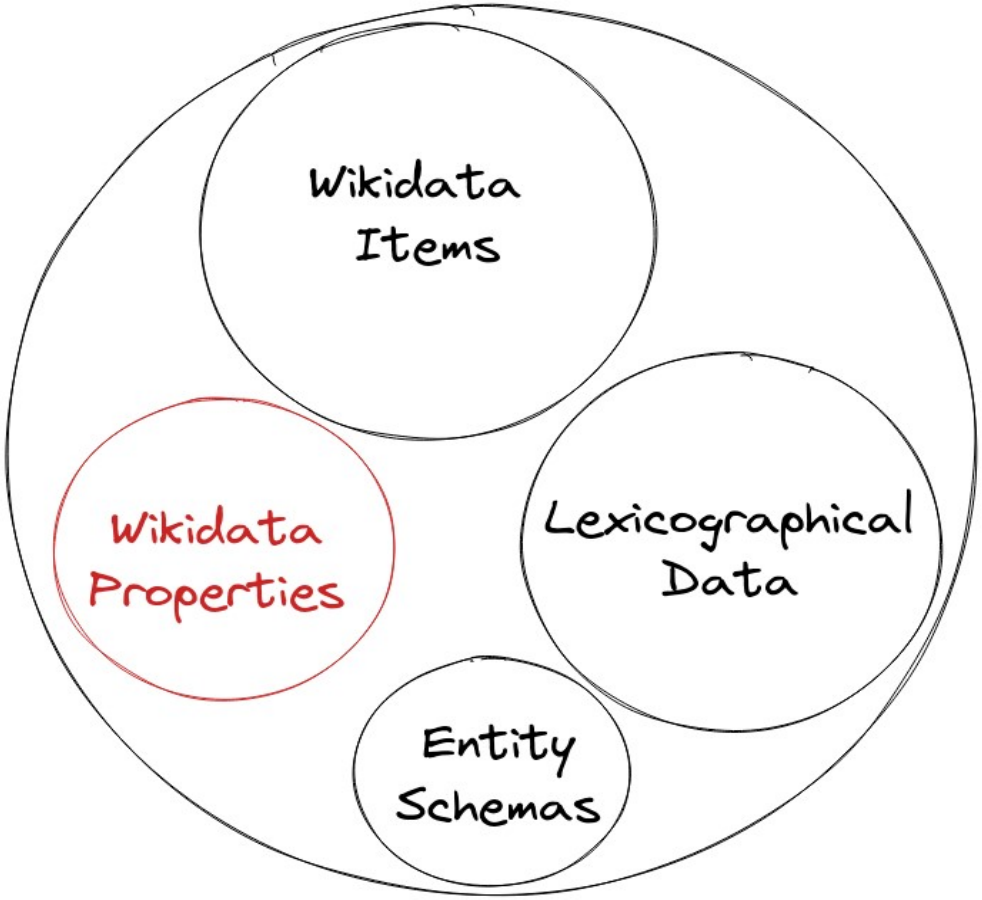
Outline

1. Introduction
2. WDProp: Wikidata Property Analysis
3. Conclusion and Future Works

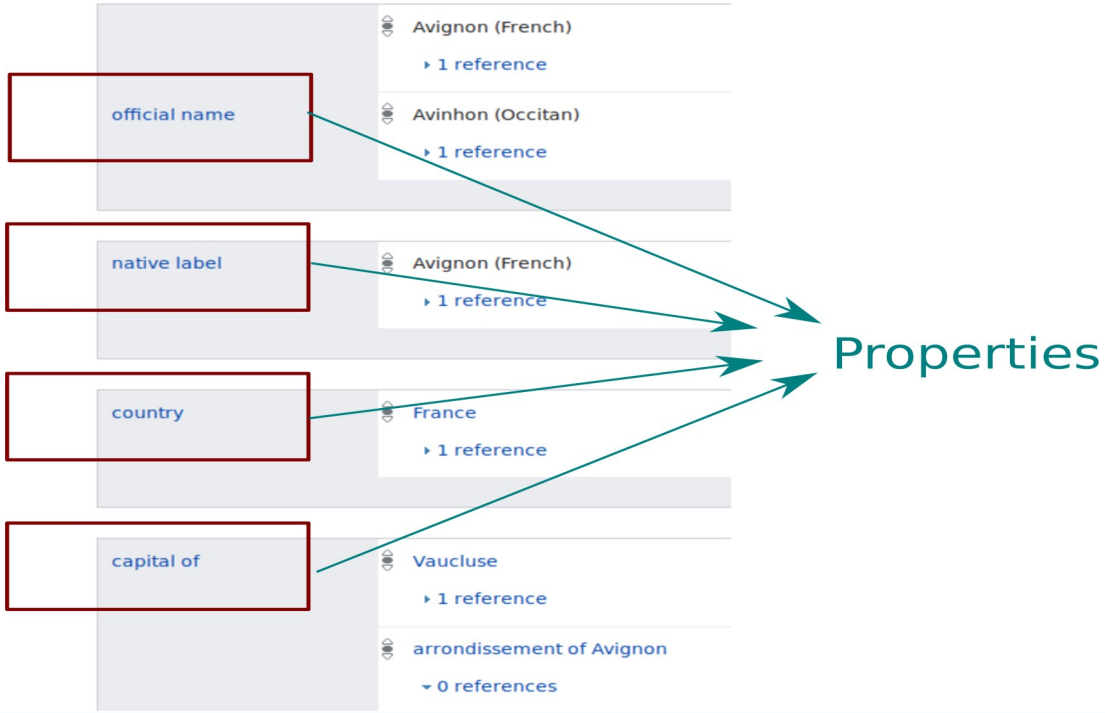
1. Introduction

- Wikidata
  - Started in 2012
  - is free, open, linked, structured, collaborative and multilingual knowledge base
  - From multi-(sub)domain multilingual Wikipedia sites to a single-domain multilingual website
- Collaborative Multilingual Multi-domain Ontology development
  - Property proposal, discussion and voting
  - Property creation and translation
  - Property usage
  - (Property deletion)

1. Introduction

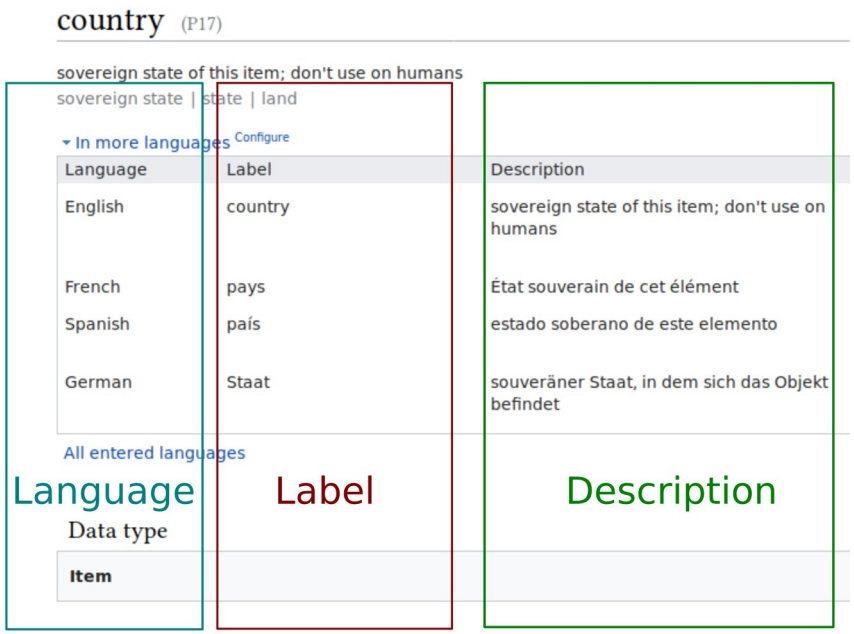


Wikidata Properties



Wikidata Properties

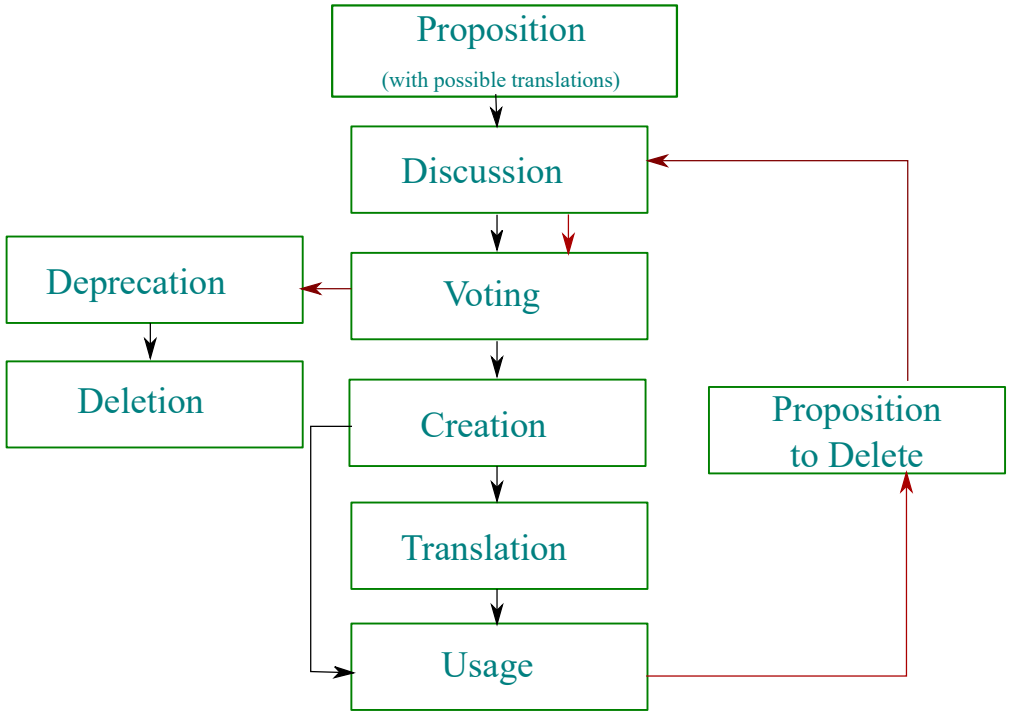
Wikidata Property: Labels and descriptions



https://www.wikidata.org/wiki/Property:P17

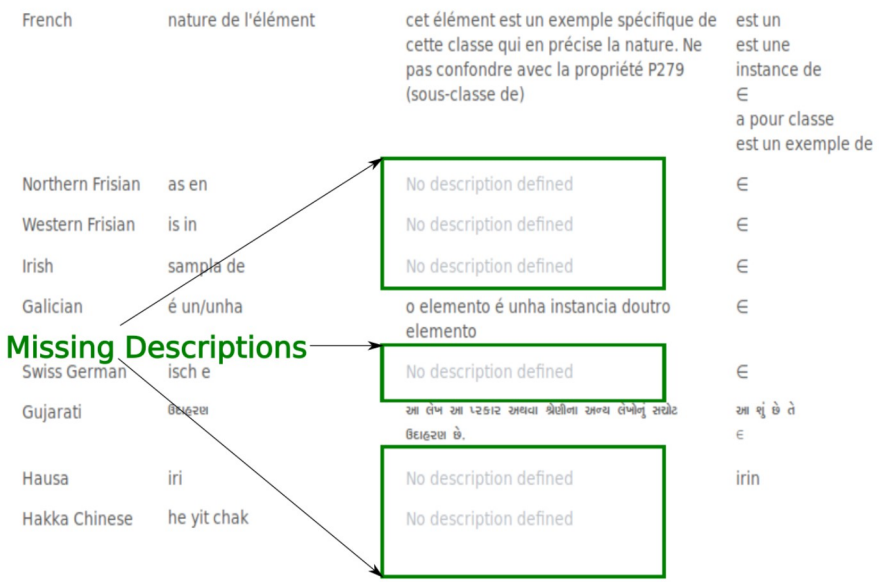
Example Wikidata Property: country (P17)

Wikidata Property: from Proposal to Usage



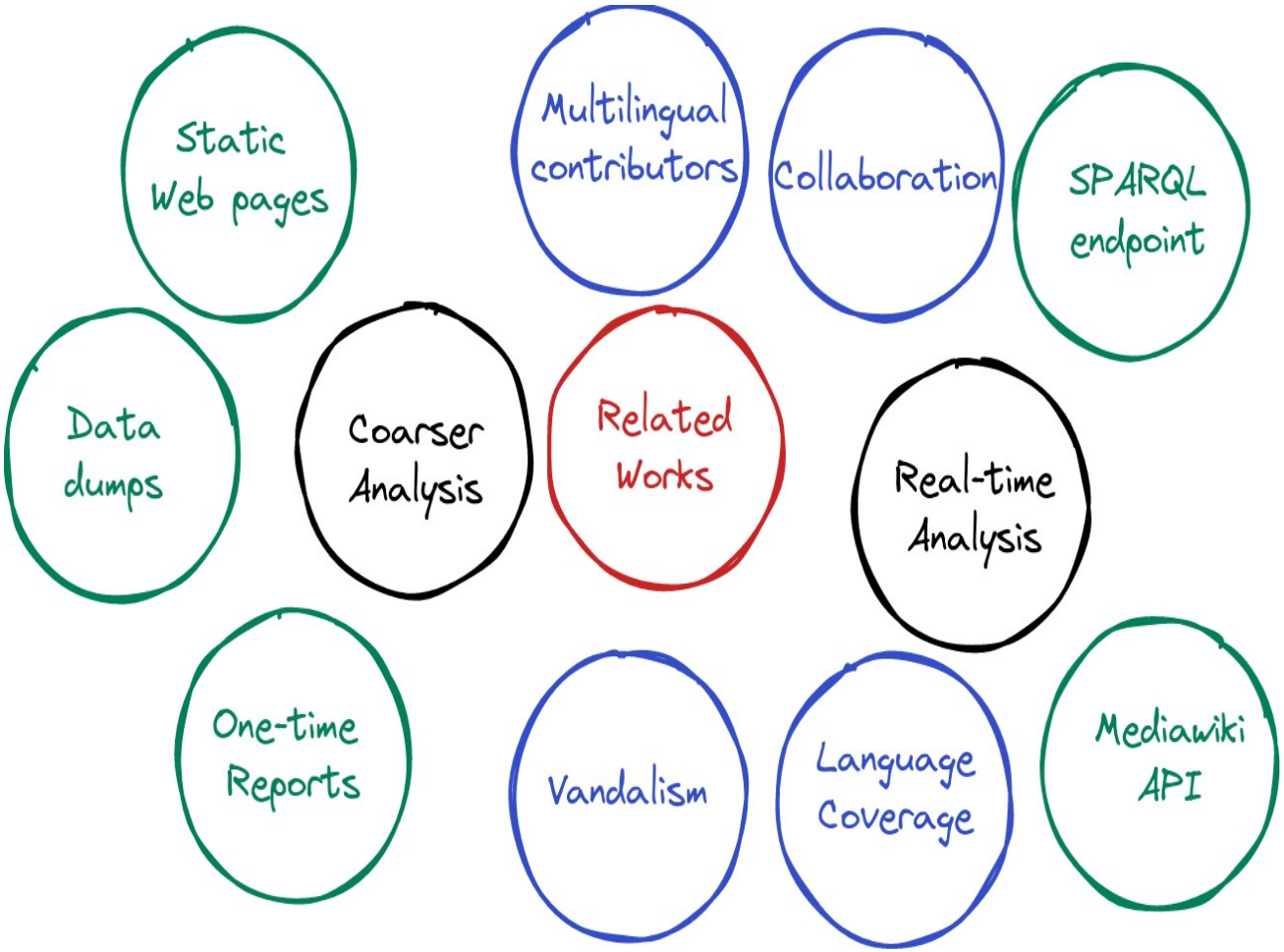
Property proposal, creation and usage on Wikidata

Wikidata Property: Missing descriptions



Missing multilingual descriptions on Property P31

1. Introduction

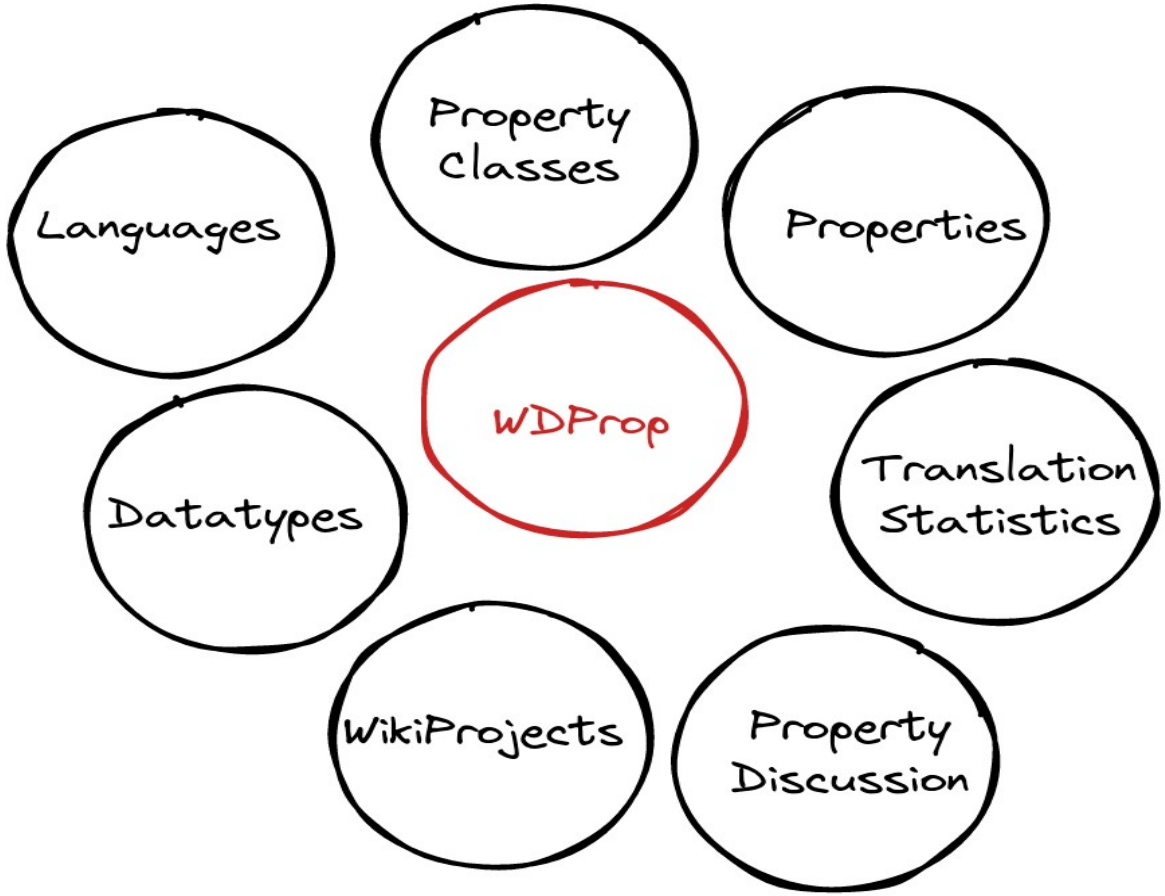


2. WDProp: Wikidata Property Analysis

WDProp

1. WDProp:
   - Collaborative Multilingual Multi-domain Ontology development: is it possible to achieve a truly multilingual experience?
2. Properties:
   - Not so much influenced by bots!
3. Goals:
   - Understanding Wikidata property proposal, creation and translation
   - Available templates and their usage
   - Providing real-time statistics to (multilingual) contributors

2. WDProp: Wikidata Property Analysis



2. WDProp: Wikidata Property Analysis

Towards 'Everything' about Wikidata properties



Information on Wikidata Properties

2. WDProp: Wikidata Property Analysis

- WDProp
  - Get real-time translation statistics
  - Navigate supported languages, properties, datatypes, classes
  - Compare translation statistics
  - Find available properties for an entity
  - Shows associated Wikidata SPARQL endpoints and Mediawiki API
  - Based on basic web technologies (HTML, Javascript, CSS)
- URL
  - https://wdprop.toolforge.org/
  - Source code: https://github.com/johnsamuelwrites/wdprop

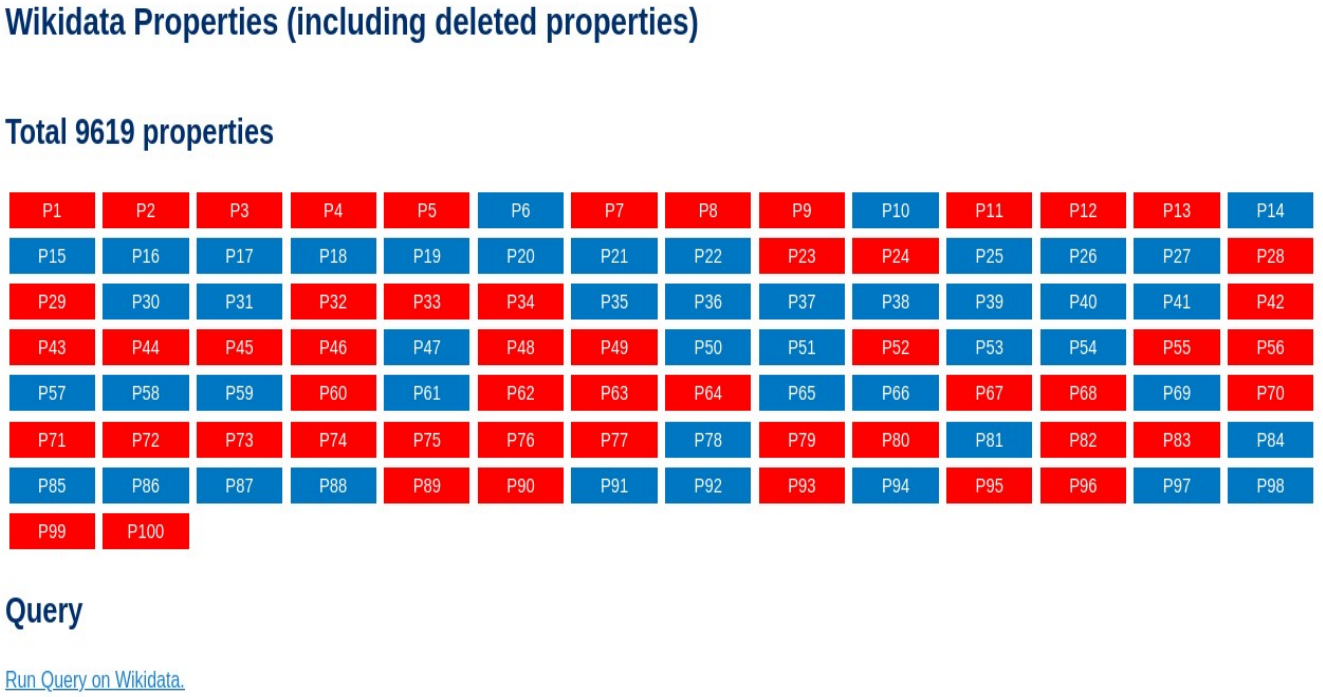
2. WDProp: Wikidata Property Analysis

Wikidata Properties



Wikidata properties

Wikidata Properties



Wikidata properties: All included deleted ones

2. WDProp: Wikidata Property: P856

Wikidata Properties

P856

- Property details
- Label missing translation in languages
- Description missing translation in languages
- Alias missing translation in languages
- Alias missing translation in languages
- Label already translated in languages
- Description already translated in languages

Details

Feature	Value
Link	https://www.wikidata.org/entity/P856
Translation Path	path.html?property=P856 (Visualization: P856)
Provenance information	provenance.html?property=P856

Wikidata Property: P856 (official website)

2. WDProp: Wikidata Property: P856

Wikidata Properties

Label missing translation in the following languages

Total 203 languages

aa	ab	ace	ady	ak	ang	arc	ary	as	av	avk	awa	ay	azb
ban	bar	bat-smg	bcl	be-x-old	bi	bjn	bm	bo	bpy	bug	bxr	cbk-zam	cdo
ceb	ch	cho	chr	chy	co	cr	crh	csb	cu	cv	din	dsb	dty
dv	dz	ee	eml	en-simple	en-x-simple	ext	ff	fiu-vro	fj	frp	fur	gag	gan
						gcr	glk	gn	gom	gor	got	gv	hak
haw	hif	ho	hz	ie	ig	ii	ik	inh	is	iu	jam	jbo	kaa
kab	kbd	kbp	kg	ki	kj	kl	km	koi	kr	krc	ks	ksh	ku
lad	lbe	lez	lfn	lg	li	lij	lld	lmo	ln	lo	lrc	ltg	lzh
mad	map-bms	mdf	mh	mhr	min	mis	mn	mni	mnw	mo	mrj	mus	mwl
myv	mzn	na	nah	new	ng	nn-hognorsk	nov	nrm	nso	nv	ny	om	pag
							pam	pap	pcd	pdc	pfl	pi	pms
pnt	rm	rmy	rn	roa-tara	ru-sib	rue	rup	rw	sa	sah	sc	sd	se
sg	shn	simple	skr	sm	smn	sn	so	srn	ss	st	stq	su	sw
szy	tay	tcy	tet	tk	tl	tlh	tn	to	tpi	trv	tum	tw	ty
tyv	udm	ug	uz	ve	vep	vls	wa	war	wo	wuu	xal	xh	xmf
yo	za	zea	zh-min-nan	zh-yue	zu								

Missing labels in languages: P856

Wikiprojects

Data models [edit]

Generic language [edit]

Title	ID	Data type	Description
image	P18	Commons media file	illustration and image: image of relevant illustration of the subject; if available, use more specific properties (sample: coat of arms image, locator map, flag image, signature image, logo image, collage image)
instance of	P31	Item	instance of: that class of which this subject is a particular example and member
programming paradigm	P3966	Item	programming paradigm: programming paradigm in which a programming language is classified
logo image	P154	Commons media file	logo: graphic mark or emblem commonly used by commercial enterprises, organizations and products
developer	P178	Item	video game developer and software developer: organisation or person that developed the item
copyright license	P275	Item	license: license under which this copyrighted work is released
designed by	P287	Item	designer: person(s) or organization which designed the object
typing discipline	P7078	Item	type system: specifies the typing discipline implemented, used, or mandated in a programming language
operating system	P306	Item	operating system: operating system (OS) on which a software works or the OS installed on hardware
software version identifier	P348	String	software version: numeric or nominal identifier of a version of a software program or file format, current or past
inception	P571	Point in time	date of establishment: date or point in time when the subject came into existence as defined
influenced by	P737	Item	contributing factor: this person, idea, etc. is informed by that other person, idea, etc., e.g. “Heidegger was influenced by Aristotle”

WikiProjects: Programming Languages

2. WDProp: Wikidata Property Analysis

Wikiprojects

WikiProject Association football (Details)
WikiProject Anime and Manga (Details)
WikiProject Patents (Details)
WikiProject Roads (Details)
WikiProject Energy (Details)
WikiProject Mathematics (Details)
WikiProject Star Trek (Details)
WikiProject Visual arts/Item structure (Details)
WikiProject Kākāpō (Details)
WikiProject Companies (Details)
WikiProject Archival Description (Details)
WikiProject Early Modern England and Wales (Details)
WikiProject Museums (Details)
WikiProject sum of all paintings (Details)
WikiProject Duplicates (Details)
WikiProject property constraints (Details)
WikiProject Templates (Details)
WikiProject India/Newsletter/2019 10 01 (Details)
WikiProject LGBT (Details)
WikiProject Greek War of Independence (Details)
WikiProject Women (Details)
WikiProject Armenia (Details)
WikiProject Historical Place (Details)

WikiProjects

Wikiprojects

Wikidata WikiProject

Total 37 properties

Property	Property label
P131	located in the administrative territorial entity
P154	logo image
P31	instance of
P17	country
P159	headquarters location
P276	location
P571	inception
P791	ISIL
P856	official website
P576	dissolved, abolished or demolished date
P669	located on street
P625	coordinate location
P968	e-mail address

WikiProjects: Properties

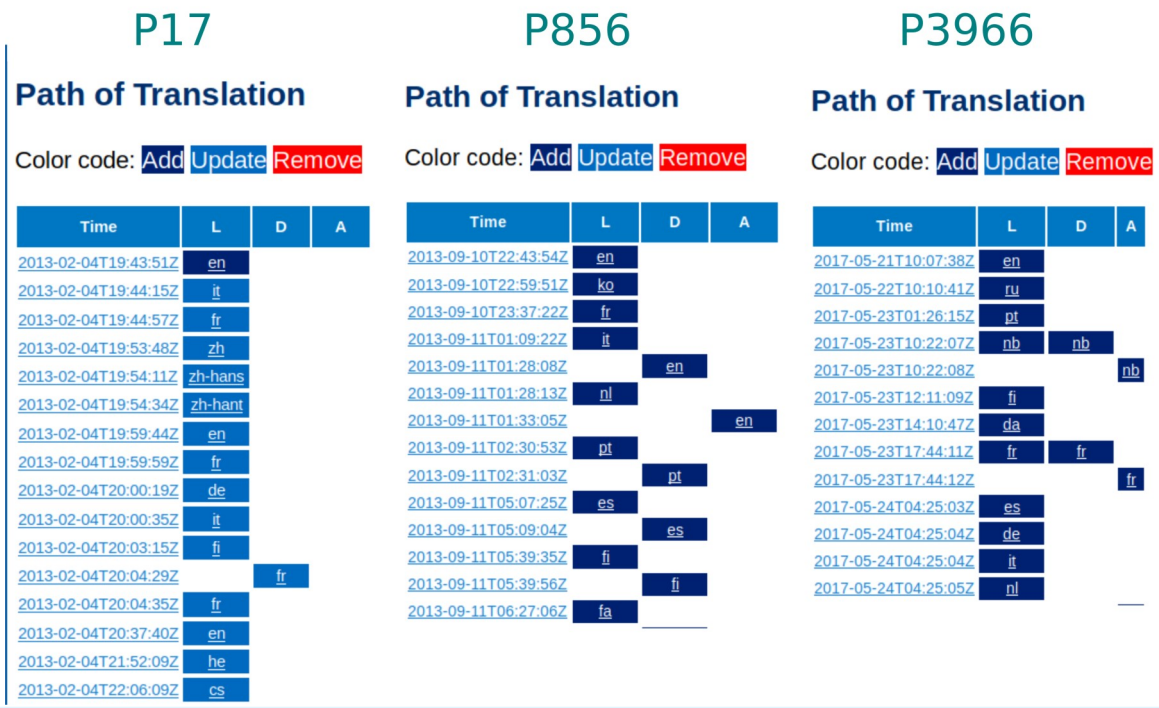
2. WDProp: Wikidata Property Analysis

Wikidata Properties: Statistcs of Multilingual Labels



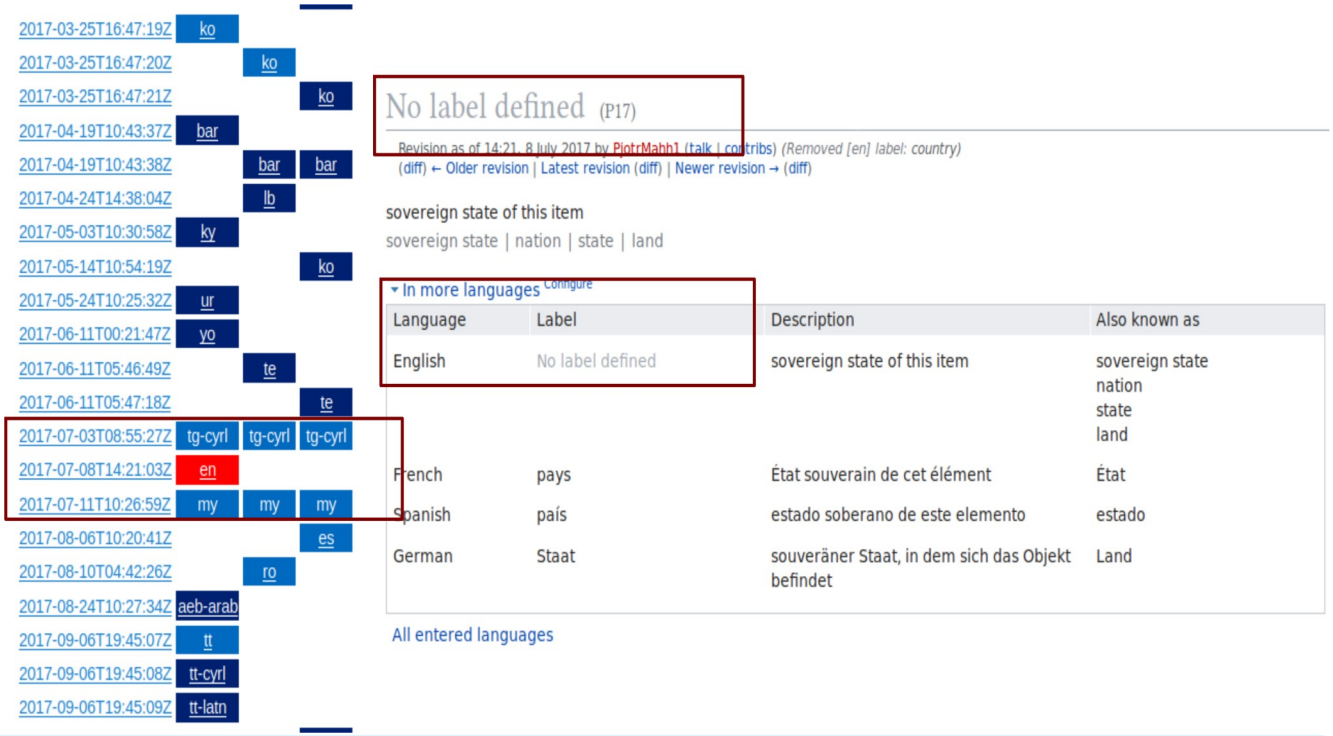
Real time information on number of multilingual labels

WDProp: Translation path



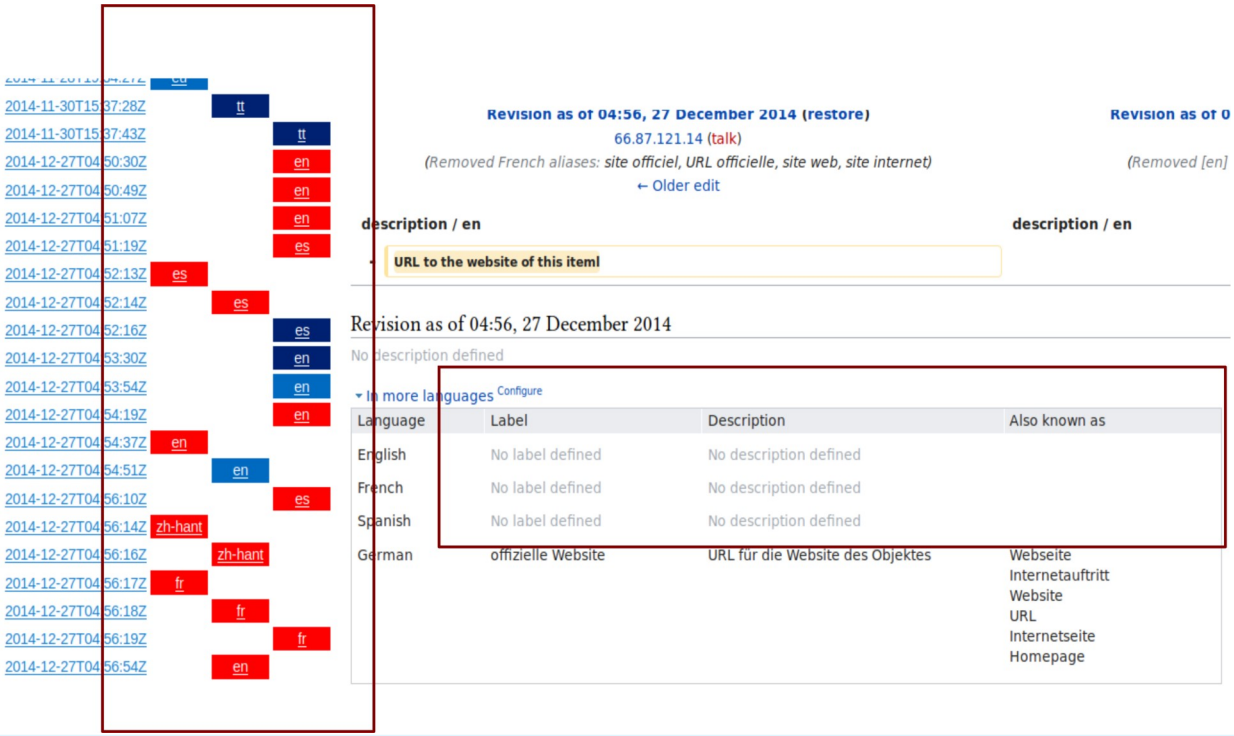
Wikidata properties: Translation path of three Wikidata Properties

WDProp: Translation path



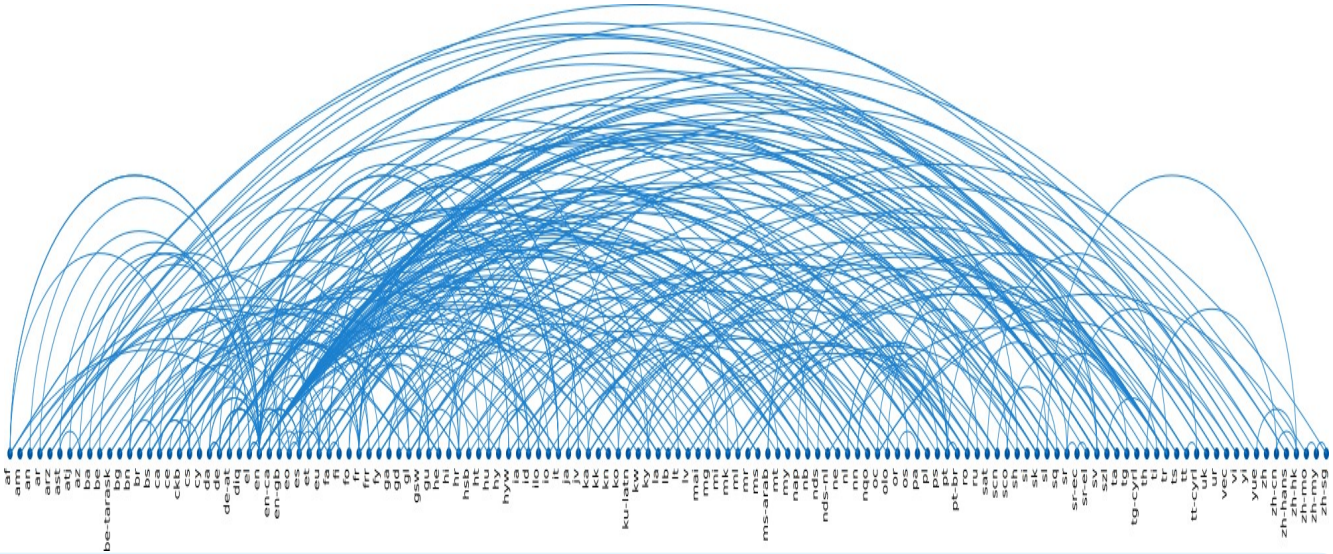
Wikidata properties: Possible vandalism (removal of labels)

WDProp: Translation path



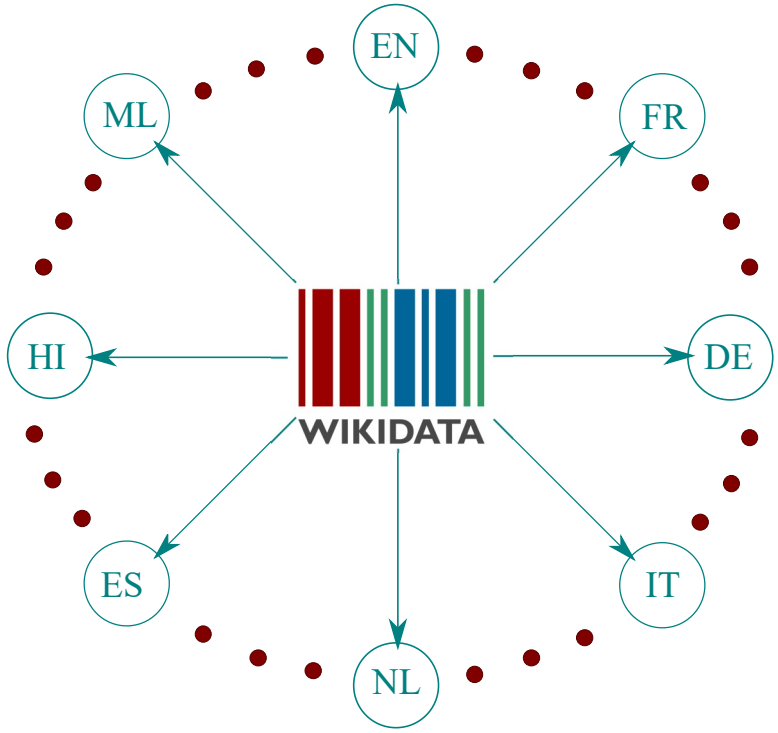
Wikidata properties: Possible vandalism (removal of labels, descriptions and aliases)

WDProp: Translation path



Wikidata properties: Visualization of Translation

Wikdiata to Wikipedia



Exporting data from Wikidata to multiple multilingual Wikipedia articles

3. Conclusion and Future Works

Future Works

1. Search:
   - Improving the search experience
2. Visualize:
   - Export as SVG
3. Download:
   - API
   - Download as JSON, CSV

References

1. Chamard, Thibaut, and John Samuel. Multilingual Wikidata Property Translation Flow Dataset. Zenodo, 8 July 2019. DOI.org (Datacite), doi:10.5281/ZENODO.3271357.
2. Kaffee, L. A., Piscopo, A., Vougiouklis, P., Simperl, E., Carr, L., & Pintscher, L. (2017, August). A glimpse into Babel: an analysis of multilinguality in Wikidata. In Proceedings of the 13th International Symposium on Open Collaboration (p. 14). ACM.
3. Müller-Birn, C., Karran, B., Lehmann, J., & Luczak-Rösch, M. (2015, August). Peer-production system or collaborative ontology engineering effort: What is Wikidata?. In Proceedings of the 11th International Symposium on Open Collaboration (p. 20). ACM.
4. Pellissier Tanon, T., & Kaffee, L. A. (2018, April). Property label stability in Wikidata: evolution and convergence of schemas in collaborative knowledge bases. In Companion of the The Web Conference 2018 on The Web Conference 2018 (pp. 1801-1803). International World Wide Web Conferences Steering Committee.

References

5. Samoilenko, A., Karimi, F., Kunegis, J., Edler, D., & Strohmaier, M. (2015, June). Linguistic influence patterns within the global network of Wikipedia language editions. In Proceedings of the ACM Web Science Conference (p. 54). ACM.
6. Samuel, J. (2017) Collaborative Approach to Developing a Multilingual Ontology: A Case Study of Wikidata. In : Research Conference on Metadata and Semantics Research. Springer, Cham, 2017. p. 167-172.
7. Samuel, J. (2018). Towards Understanding and Improving Multilingual Collaborative Ontology Development in Wikidata. In: WikiWorkshop 2018
8. Samuel, John. Johnsamuelwrites/Wdprop: V0.12. v0.12, Zenodo, 2021. DOI.org (Datacite), doi:10.5281/ZENODO.1174371.
9. Stefaner, M., Taraborelli, D., & Ciampaglia, G. L. (2011). Notabilia–Visualizing Deletion Discussions on Wikipedia.

Thank you

Questions?