

Language gaps in Wikidata

Simon Cobb

Celtic Knot Conference 2018, National Library of Wales, Aberystwyth, 6th July 2018



<http://tinyurl.com/y84d9rrr>

What is Wikidata?

- Repository of structured data for Wikimedia projects
- Knowledge base
- Linked open data
- Free – data released under CC0 Public Domain Dedication
- Collaborative
- Multilingual



Structured, multilingual data

label — Douglas Adams (Q42) — **item identifier**

description — English writer and humorist
Douglas Noël Adams | Douglas Noel Adams — **aliases**

property — **educated at** — **value**

St John's College	end time	1974
	academic major	English literature
	academic degree	Bachelor of Arts
	start time	1971

qualifiers

rank — **opened references**

stated in	Encyclopædia Britannica Online
reference URL	http://www.nndb.com/people/731/000023662/
original language of work	English
retrieved	7 December 2013
publisher	NNDB
title	Douglas Adams (English)

statement group — **collapsed reference**

Brentwood School	end time	1970
	start time	1959

collapsed reference

레이블 — 더글러스 애덤스 (Q42) — **식별자 고유**

설명 — 영국의 작가
더글러스 애덤스 | 더글러스 노엘 애덤스 — **이름 다른**

속성 — **다녔던 학교** — **값**

세인트존스 칼리지	종료된 날	1974
	전공	영문학
	학위	문학사
	시작된 날	1971

한정자

등급 — **목록 출처 (상태 열린)**

다음에 서술되어 있음	브리태니커 백과사전 온라인
참고한 URL	http://www.nndb.com/people/731/000023662/
저작물의 원어	영어
확인한 날짜	7 12월 2013
발행사	NNDB
제목	Douglas Adams (영어)

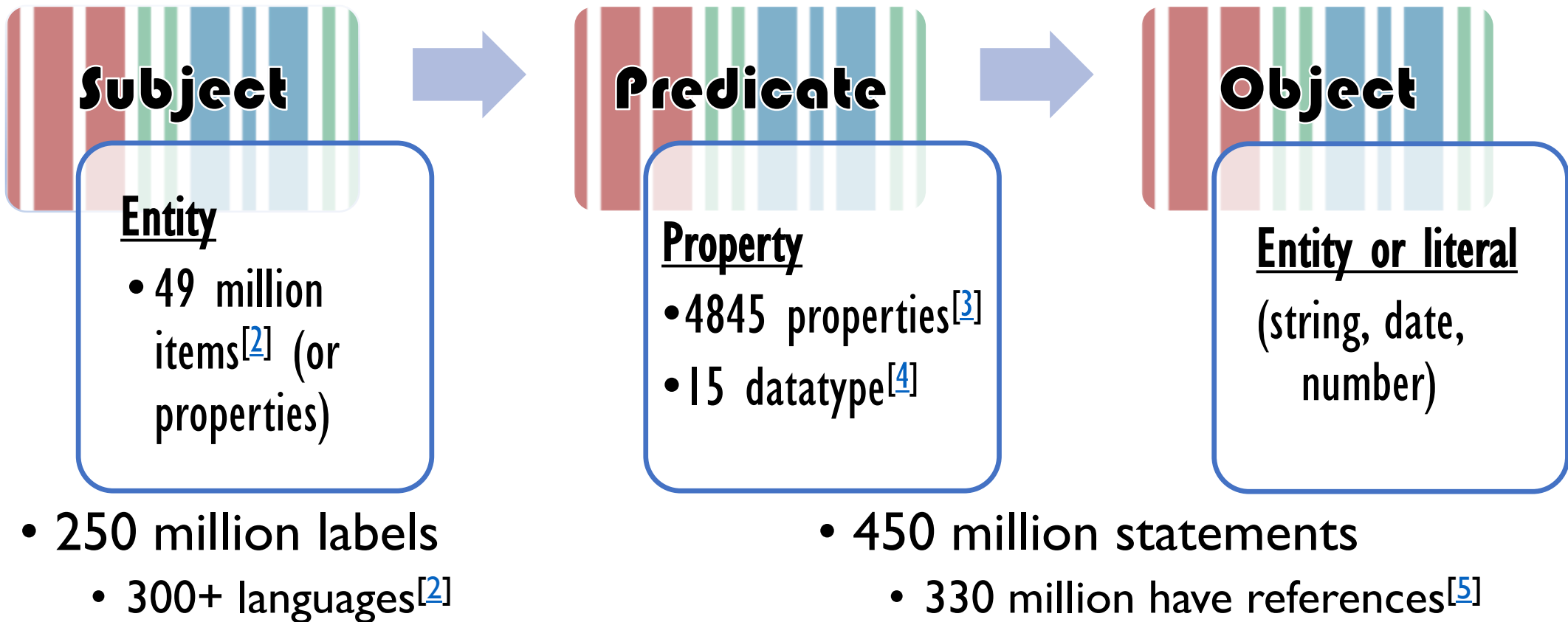
목록 출처 (상태 닫힌)

브렌트우드 스쿨	종료된 날	1970
	시작된 날	1959

서술 묶음 — **추가**

Wikidata in numbers

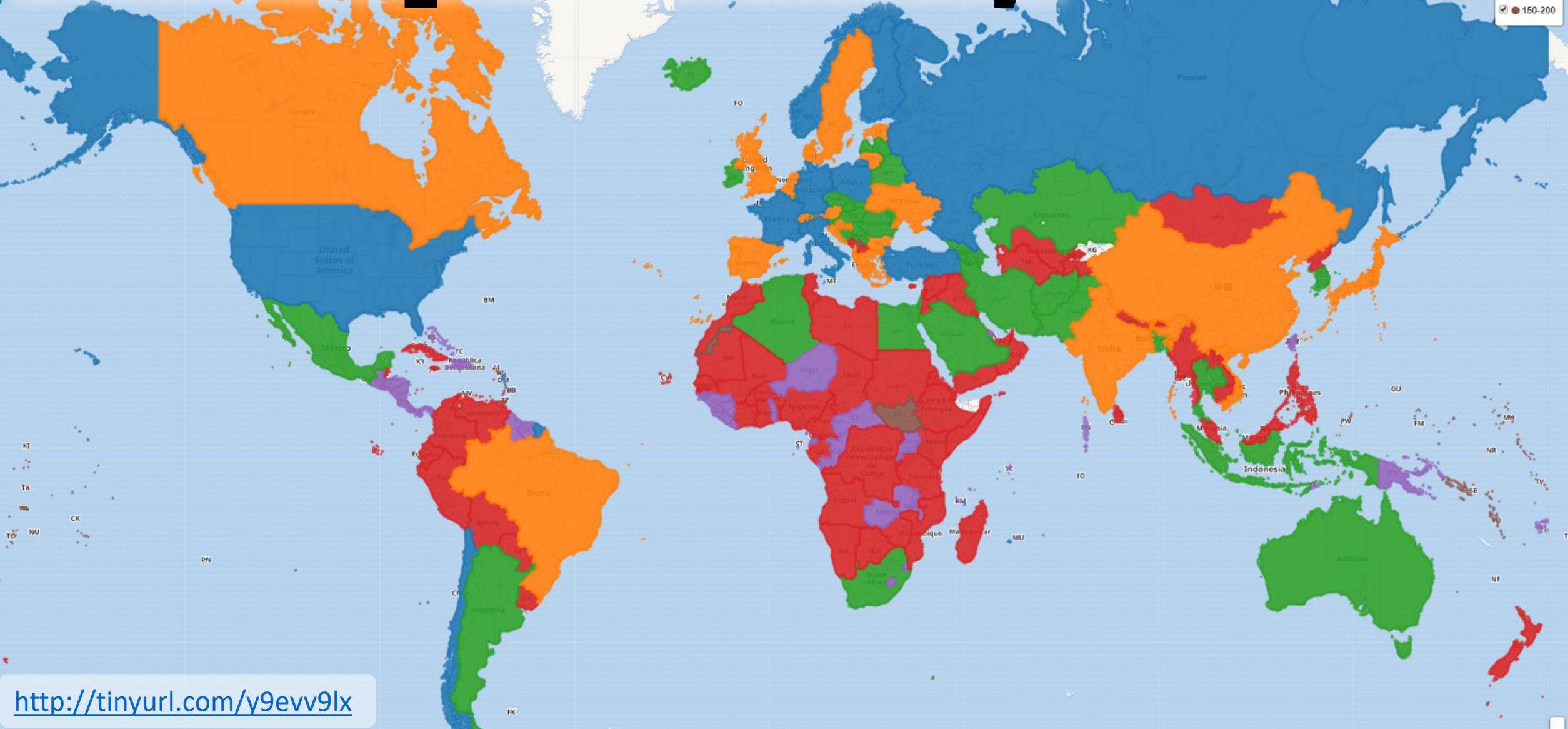
5.5 billion triples^[1]



1 - <http://tinyurl.com/ybr7gej9> ; 2 - <https://www.wikidata.org/wiki/Special:Statistics> ; 3 - <http://tinyurl.com/y7eefu4f> ; 4 - <http://tinyurl.com/y9xpjmm2> ; 5 - <https://tools.wmflabs.org/wikidata-todo/stats.php>.

Labels per country

- All layers
- 300+
- 276-300
- 251-275
- 226-250
- 201-225
- 150-200








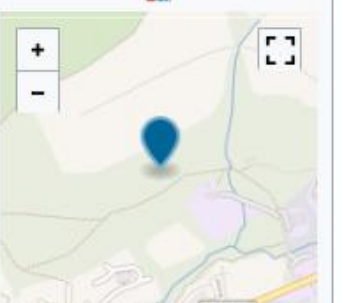


<http://tinyurl.com/y9evv9lx>

Is this sufficient?

Wikidata powered content:

- Infoboxes
- Lists
- Article placeholders
- Mapframes



Aberystwyth [Cuddier]	Aberystwyth [Hide]	Aberystwyth [Kuzhat]	Aberystwyth [Falaich]
			
tref yng Ngheredigion, Cymru	town in Ceredigion, Wales	localité britannique	town in Ceredigion, Wales
Wicipedia	Wikipedia Wikivoyage	Wikipedia	Uicipeid
Enghraifft o'r canlynol cymuned	Instance of market town, of community	Doare an elfenn Ville de marché, communauté	Instance of market town, of community
Enwyd ar ôl Afon Ystwyth	Named after River Ystwyth	Anvet eus Ystwyth	Named after River Ystwyth
Lleoliad Ceredigion, Y Deyrnas Unedig	Located in Ceredigion, United Kingdom	Lieu Ceredigion, Rouantelezh-Unanet	Làrach Ceredigion, An Rìoghachd Aonaichte
Rheolaeth awdurdod [Cuddier]	Authority control [Hide]	Autorité [Kuzhat]	Authority control [Falaich]
dynodwr VIAF: 148932214	VIAF identifier: 148932214	VIAF: 148932214	VIAF: 148932214
GND: 4079607-3	GND identifier: 4079607-3	GND: 4079607-3	GND ID: 4079607-3
			
Wikimedia maps Map data © OpenSt...	Wikimedia maps Map data © OpenSt...	Maps Données de carte © les contributeu...	Wikimedia maps Map data © OpenSt...
52° 25′ 12″ N, 4° 04′ 12″ W	52° 25′ 12″ N, 4° 04′ 12″ W	52° 25′ 12″ N, 4° 04′ 12″ W	52° 25′ 12″ N, 4° 04′ 12″ W
OpenStreetMap · Google Maps · Google Earth · Proximityrama	OpenStreetMap · Google Maps · Google Earth · Proximityrama	OpenStreetMap · Google Maps · Google Earth · Proximityrama	OpenStreetMap · Google Maps · Google Earth · Proximityrama
Reasonator · Scholia · Ystadegaeth · WikiShootMe	Reasonator · Scholia · Statistics · WikiShootMe	Reasonator · Scholia · Stadegouriezh · WikiShootMe	Reasonator · Scholia · Staitistearachd · WikiShootMe

Wikidata Query Service (SPARQL)

```
#Creative work in language or from country
SELECT DISTINCT ?x ?xLabel WHERE {
  {
    ?x wdt:P407 wd:Q9309 #language of work
  } UNION {
    ?x wdt:P495 wd:Q25 #country of origin
  } UNION {
    ?x wdt:P123 / wdt:P131* wd:Q25 #publisher is located in country
  }
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "[AUTO_LANGUAGE]".
  }
  FILTER(NOT EXISTS {
    ?x rdfs:label ?lang_label.
    FILTER(LANG(?lang_label) = "cy"). #language of missing label
  })
}
```

qid	itemLabel
wd:Q53577653	Hanes Morganwg (1874 edition)
wd:Q53576694	Hanes y Brytaniaid a'r Cymry (1873 edition)
wd:Q41515094	Windows 10 Version 1709
wd:Q33109779	Welsh Wikiquote
wd:Q33109629	Welsh Wikibooks
wd:Q33109160	Welsh Wiktionary
wd:Q29344680	Every Valley
wd:Q29167167	The County Echo
wd:Q29047341	Cambria Daily Leader
wd:Q29018974	Brecon and Radnor Express, Carmarthen and Swansea Valley Gazette and Brynmawr District Advertiser
wd:Q29018705	Barry Herald
wd:Q29018691	Barry Dock News
wd:Q29016915	Barmouth and county advertiser and district weekly news
wd:Q29005491	Amman Valley Chronicle and East Carmarthen News
wd:Q28979295	The Abergavenny Mail
wd:Q28978402	Abergavenny Chronicle
wd:Q28974212	Mid-Wales Telegraph and Observer
wd:Q28974118	Montgomeryshire Herald

Add labels with QuickStatements

QuickStatements English  [New batch](#) [Last batches](#) [Chat](#) [Git](#) [Help](#) [Sic19](#) [Your last batches](#) 

Batch on Wikidata by Sic19 [[Batches](#)] [Discuss/revert batch](#)

Status: DONE  100% (14) of 14 done

1	done Hanes Morganwg (1874 edition) ADD Label cy:"Hanes Morganwg" [Q53577653]
2	done Hanes y Brytaniaid a'r Cymry (1873 edition) ADD Label cy:"Hanes y Brytaniaid a'r Cymry" [Q53576694]
3	done Hafan [Q21167778] ADD Label cy:"Hafan"
4	done Yr Atal Genhedlaeth [Q8059638] ADD Label cy:"Yr Atal Genhedlaeth"
5	done Y Lle [Q8046153] ADD Label cy:"Y Lle"
6	done Y Clwb Rygbi [Q8046113] ADD Label cy:"Y Clwb Rygbi"
7	done Tystion vs Allfa Un [Q7861820] ADD Label cy:"Tystion vs Allfa Un"
8	done Tymor yr Heliwr [Q7860397] ADD Label cy:"Tymor yr Heliwr"
9	done NLW MS 20143A [Q21561324] ADD Label cy:"NLW MS 20143A"
10	done Heddiw [Q5697130] ADD Label cy:"Heddiw"

First Page Next Last

All errors Init

<https://tools.wmflabs.org/quickstatements/#/>

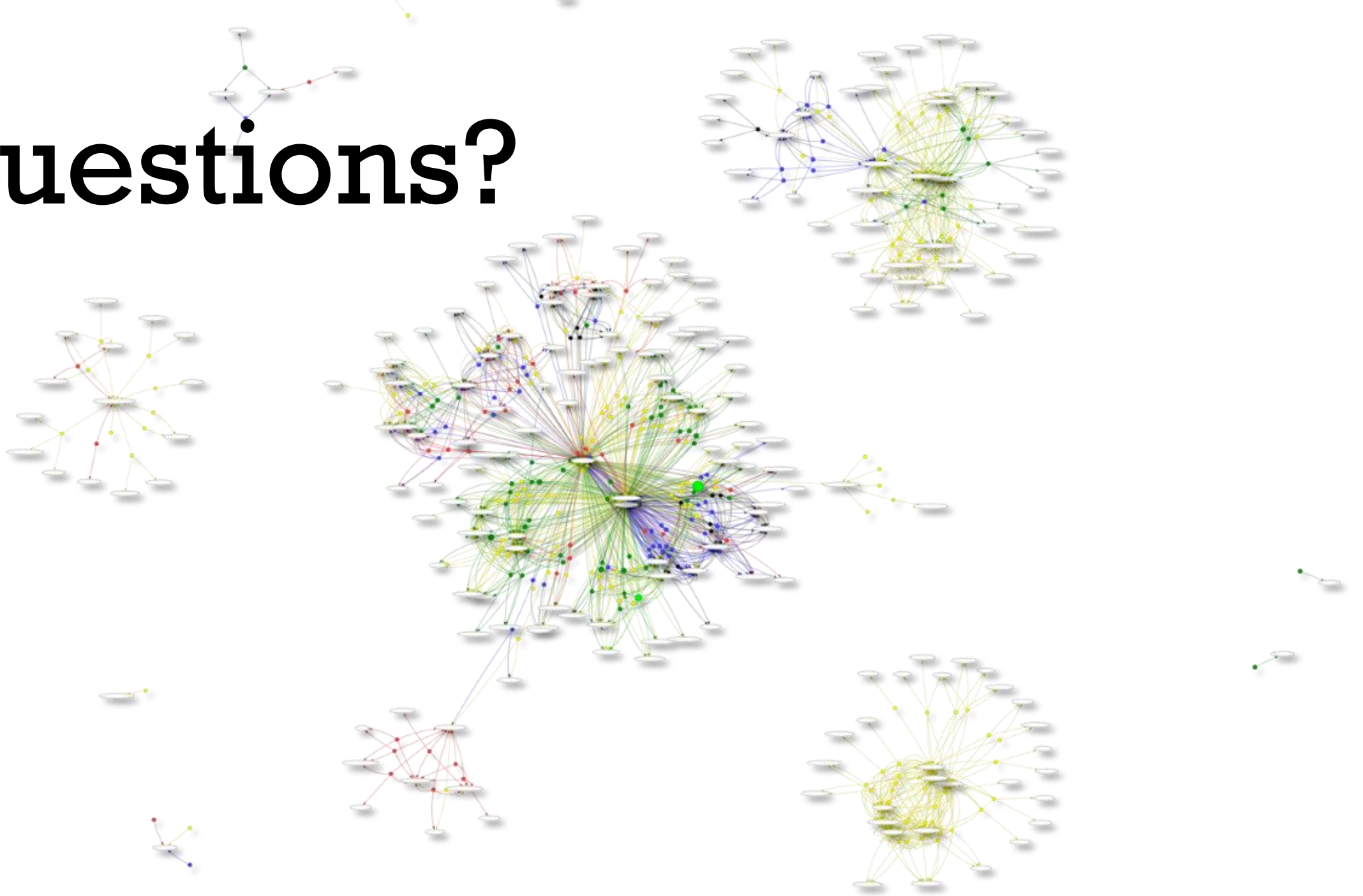


NameGuzzler / VIP Labels

The screenshot shows the Wikidata page for Gary Graham (Q53844001). The page title is "Gary Graham (Q53844001)". The main content area shows the name "Gary Graham" in various languages: British English (Gary Graham), Welsh (No label defined), English (Gary Graham), and Latin (No label defined). The "Statements" section shows "instance of" and "sex or gender". A dialog box titled "VIP's labels" is open, showing a search for "Gary Graham" and a list of 84 selected languages. The dialog box has a "Labels:" input field with "Gary Graham" and radio buttons for "Mode Label" (selected) and "Mode Aliases". The language list includes: aa, ab, ace, ady, ady-cyrl, aeb, aeb-arab, aeb-latn, af, ak, aln, am, an, ang, anp, ar, arc, am, arq, ary, arz, as, ast, atj, awa, av, avk, ay, az, azb, ba, ban, bar, bbc, bbc-latn, bcc, bcl, be, be-tarask, bg, bgn, bho, bi, bjn, bm, bn, bo, bpy, bqj, br, brh, bs, bto, bug, bxr, ca, cbk-zam, cdo, ce, ceb, ch, cho, chr, chy, ckb, co, cps, cr, crh-latn, crh-cyrl, cs, csb, cu, cv, cy, da, de, de-at, de-ch, din, diq, dsb, dtp, dty, dv, dz, ee, egl, el, eml, en, en-ca, en-gb, eo, es, et, eu, ext, fa, ff, fi, fit, fj, fo, fr, frc, frp, frr, fur, fy, ga, gag, gan, gan-hans, gan-hant, gd, gl, glk, gn, gom, gom-deva, hi, hif, hif-latn, hil, ho, hr, hrx, hsb, ht, hu, hy, hz, ia, id, ie, ig, ii, ik, ike-cans, ike-latn, ilo, inh, io, is, it, iu, ja, jam, jbo, jut, jv, ka, kaa, kab, kbd, kbd-cyrl, kbp, kea, kg, khw, ki, kiu, kj, kk, kk-arab, kk-cyrl, kk-latn, kk-cn, kk-kz, kk-tr, kl, km, kn, ko, ko-kp, koi, kr, krc, kri, krj, krl, ks, ks-arab, ks-deva, ksh, ku.

Language	Label
British English	Gary Graham
Welsh	Gary Graham
English	Gary Graham
Latin	No label defined
Afrikaans	Gary Graham
Aragonese	Gary Graham
Asturian	Gary Graham
Aymara	Gary Graham
Azerbaijani	Gary Graham
Breton	Gary Graham
Catalan	Gary Graham
Chechen	Gary Graham
Cebuano	Gary Graham
Corsican	Gary Graham
Czech	Gary Graham
Kashubian	Gary Graham
Danish	Gary Graham
German	Gary Graham
Austrian German	Gary Graham

Questions?



Identify other language gaps

Recent events

Deaths:

<http://tinyurl.com/yc2m7n2f>

Birthday today:

<http://tinyurl.com/ycdwnofs>

Events:

<http://tinyurl.com/y8qkhlf8>

Topics:

Scholia: <https://tools.wmflabs.org/scholia/topic/Q25> (example showing works related to Wales).

People

Language of name:

<http://tinyurl.com/ybos3rj3>

Speakers of:

<http://tinyurl.com/yclowz7y>

Born, died, worked in, or citizen of country:

<http://tinyurl.com/yd878pyl>

Location

Located in country:

<http://tinyurl.com/yaqfwkpi>

Located with proximity of:

<http://tinyurl.com/y83zvsvs>

Around current location:

<http://tinyurl.com/ya8njzv6>

External IDs

ODNB:

<http://tinyurl.com/yah8pyuo>

Coflein:

<http://tinyurl.com/yd835bml>

List of properties:

<http://tinyurl.com/ya33s58q>

Further information

- *Babylon: the Wikimedia translators' portal/noticeboard.* <https://meta.wikimedia.org/wiki/Meta:Babylon>.
- Kaffee, L.-A., Piscopo, A., Vougiouklis, P. et al. (2017) *A Glimpse into Babel: An Analysis of Multilinguality in Wikidata*. In: Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17). New York, NY, USA: ACM. Article 14. DOI: <https://doi.org/10.1145/3125433.3125465>.
- Kaffe, L.-A. (2017) *Languages in Wikidata*. Presentation at WikidataCon 2017, 28–29 October 2017, Berlin, Germany. https://upload.wikimedia.org/wikipedia/commons/8/81/WikidataCon_2017_-_Languages_in_Wikidata.webm.
- *WikiProject: Labels and descriptions.* https://www.wikidata.org/wiki/Wikidata:WikiProject_Labels_and_descriptions.