Wikidata
content oriented projects
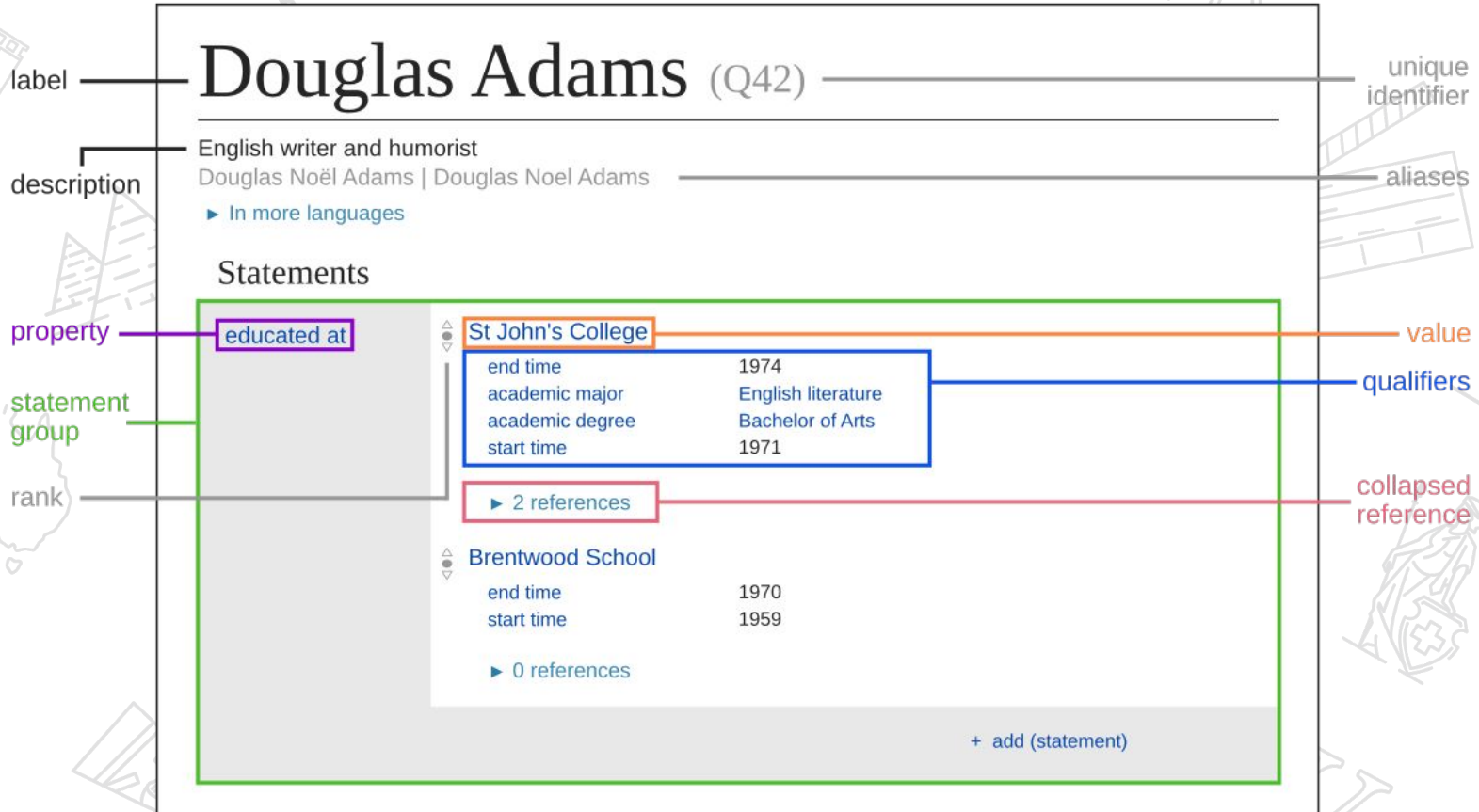
Mārtiņš Bruņenieks
CEE Meeting Belgrade 2019

You have some DATA

- It is open (as in CC0)
- It is structured (somewhat)
- It is about your favourite topic

Let's put it on Wikidata

Wikidata structure

label — Douglas Adams (Q42) — unique identifier

description — English writer and humorist
Douglas Noël Adams | Douglas Noel Adams — aliases
▶ In more languages

Statements

property — educated at

statement group

rank

St John's College — value

end time	1974
academic major	English literature
academic degree	Bachelor of Arts
start time	1971

— qualifiers

▶ 2 references — collapsed reference

Brentwood School

end time	1970
start time	1959

▶ 0 references

+ add (statement)

Starting a new Project

- Copy from others
- Describe data model — list all properties
- Invite people to join
- Use {{ping project}} to notify others

Ontology

- Think of statements that describe your data
- Map your statements to existing properties
  - [Wikidata:List of properties](#)
  - Look into similar projects
  - Using existing property with qualifier
  - [Wikidata:Property proposal](#)
  - Most likely you will be missing just a few specific properties

Workflow

↓ Obtaining useful data
↓ Ensuring licence (CC0)
↓ Preparing (cleaning data)
↓ Reconciling to avoid duplicates
↓ Running the import
↓ Maintenance queries

Preparing your data

- Suitable data formats
  - CSV, TSV, Excel
  - XML
  - Database
  - Tables in PDFs (need some work)
  - HTML pages with some structure
- Check file encoding

Preparing data — tools

- Tabula — extract data from PDFs
- OpenRefine cleanup + reconcile
- Mix'n'match — match external identifiers
- Excel/Google Sheets (addon)

Doing the import

- Quick Statements
- OpenRefine
- Writing your own scripts:
  - Python
  - Java
  - TypeScript/JavaScript
  - Wikidata API
- Do not forget the references

Tools: Wikidata Usefuls

Tools: Harvest templates

Load pages from

Wiki   en   .   wikipedia   .org

Namespace   0

Import data

Property   P   1

Template   [   ]

Parameter   [   ]

+ add alias

Filter

Category   [   ]

Depth   0

Manual list   [   ]

Already set   ☑ don't load items with the property set

Offset   0

Copy data from a Wikimedia template to Wikidata.
The tool parses templates and adds the found values to the item of the page the template is located in.

- *Wiki:* language and project name of the wiki you're working on
- *Namespace:* default value 0 corresponds to the article namespace in Wikipedia
- *Property:* destination property in Wikidata
- *Template:* source template
- *Parameter:* parameter name in the template. Use numbers for nameless parameters
- *Category:* non-mandatory option to filter pages on wiki by category
- *Manual list:* non-mandatory option to filter pages on wiki by page name or item id
- *Already set:* deactivate this option if you encounter performance issues
- *Offset:* skip a specific number of results. A negative number will start that far from the end
- *Limit:* limit the number of pages getting loaded

By clicking *get pages* a list with articles gets created. After it, when clicking *add values* the articles get parsed and claims get added to Wikidata. Alternatively, with *demo* only parsing is done without any edit action.

News

- Autoload (&load=) and autorun (&run=)
- Collaboration platform

Tools: Infobox export

Completeness dashboard

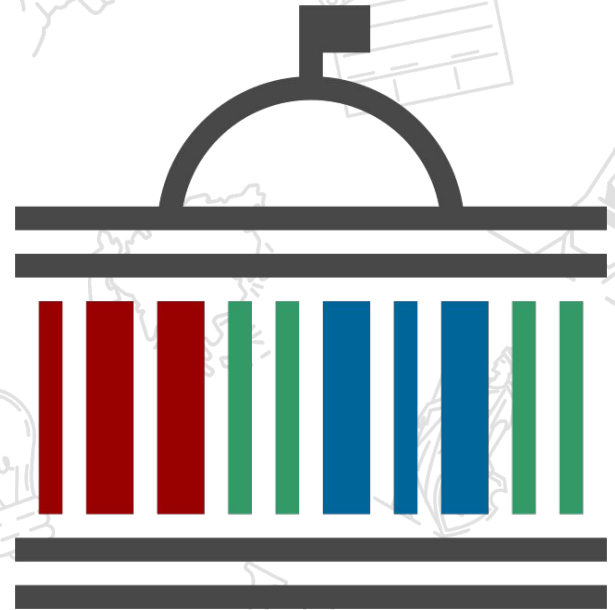| Country	Name	Count	creator	location	inception	height & width	material used	image	genre	depicts	main subject	title
	Bavarian State Painting Collections (Q812285)	17860	87.18% (15571)	44.71% (7985)	78.9% (14091)	97.69% (17447)	25.44% (4544)	49.44% (8830)	9.5% (1697)	4.69% (837)	4.95% (884)	99.09% (17698)
	Cultural Heritage Agency of the Netherlands Art Collection (Q18600731)	12795	71.4% (9135)	64.53% (8257)	93.51% (11965)	71.96% (9207)	59.68% (7636)	25.72% (3291)	15.76% (2016)	2.08% (266)	1.13% (144)	94.15% (12047)
	National Trust (Q333515)	12659	90.51% (11458)	62.37% (7896)	75.18% (9517)	92.25% (11678)	69.35% (8779)	1.55% (196)	4.15% (525)	1.93% (244)	2.05% (260)	95.03% (12030)
	Finnish National Gallery (Q2983474)	10059	99.62% (10021)	99.48% (10007)	73.28% (7371)	92.84% (9339)	31.77% (3196)	23.51% (2365)	9.07% (912)	3.32% (334)	0.52% (52)	99.89% (10048)
	Smithsonian American Art Museum (Q1192305)	8405	99.3% (8346)	99.94% (8400)	92.25% (7754)	89.28% (7504)	39.14% (3290)	37.28% (3133)	8.69% (730)	7.01% (589)	0.79% (66)	99.88% (8395)
	Metropolitan Museum of Art (Q160236)	8262	99.95% (8258)	99.7% (8237)	99.43% (8215)	76.31% (6305)	52.87% (4368)	64.37% (5318)	48.09% (3973)	32.04% (2647)	6.9% (570)	98.27% (8119)
	Nationalmuseum (Q842858)	7830	93.51% (7322)	94.11% (7369)	40.14% (3143)	90.97% (7123)	52.07% (4077)	64.01% (5012)	19.89% (1557)	37.36% (2925)	1.83% (143)	0.83% (65)
	Slovak National Gallery (Q1744024)	7564	98.84% (7476)	100.0% (7564)	96.39% (7291)	95.07% (7191)	52.37% (3961)	26.76% (2024)	32.43% (2453)	1.31% (99)	1.72% (130)	95.94% (7257)
	private collection (Q768717)	7000	99.96% (6997)	29.11% (2038)	97.1% (6797)	55.93% (3915)	86.53% (6057)	74.96% (5247)	67.51% (4726)	61.61% (4313)	8.41% (589)	5.93% (415)
	Hermitage Museum (Q132783)	6941	92.08% (6391)	99.73% (6922)	89.38% (6204)	82.74% (5743)	74.31% (5158)	37.72% (2618)	37.67% (2615)	17.42% (1209)	7.26% (504)	97.78% (6787)
	Statens Museum for Kunst (Q671384)	6811	99.74% (6793)	99.99% (6810)	92.51% (6301)	10.38% (707)	71.5% (4870)	56.91% (3876)	28.09% (1913)	18.44% (1256)	3.36% (229)	99.02% (6744)
	Art Museum of Estonia's painting collection (Q56189782)	6492	96.69% (6277)	0.0% (0)	71.58% (4647)	93.11% (6045)	99.69% (6472)	0.55% (36)	0.65% (42)	0.49% (32)	0.0% (0)	0.0% (0)
	National Gallery of Australia (Q795228)	6325	78.18% (4945)	99.98% (6324)	64.44% (4076)	57.45% (3634)	12.17% (770)	0.71% (45)	1.96% (124)	0.22% (14)	0.0% (0)	0.0% (0)
	Museo del Prado (Q160112)	6303	100.0% (6303)	99.78% (6289)	39.47% (2488)	99.33% (6261)	94.89% (5981)	45.2% (2849)	21.5% (1355)	14.07% (887)	10.65% (671)	5.73% (361)
	National Gallery of Victoria (Q1464509)	6007	79.17% (4756)	100.0% (6007)	23.79% (1429)	78.28% (4702)	28.85% (1733)	2.98% (179)	2.86% (172)	1.76% (106)	0.7% (42)	0.43% (26)
	Yale University Art Gallery (Q1568434)	5792	90.02% (5214)	99.98% (5791)	67.56% (3913)	6.54% (379)	43.32% (2509)	19.98% (1157)	6.37% (369)	2.68% (155)	1.0% (58)	82.27% (4765)
	Department of Paintings of the Louvre (Q3044768)	5621	99.75% (5607)	99.79% (5609)	98.68% (5547)	93.33% (5246)	98.45% (5534)	44.24% (2487)	56.45% (3173)	94.15% (5292)	11.53% (648)	4.25% (239)
	Royal Collection (Q1459037)	5596	91.39% (5114)	31.33% (1753)	92.3% (5165)	95.05% (5318)	70.73% (3958)	8.99% (503)	12.08% (676)	6.61% (370)	3.63% (203)	98.55% (5515)
	Rijksmuseum (Q190804)	5532	99.91% (5527)	99.71% (5516)	97.74% (5407)	92.79% (5133)	95.01% (5256)	92.97% (5143)	60.23% (3332)	38.77% (2145)	5.78% (320)	86.95% (4810)
	Staatliche Kunstsammlungen Dresden (Q653002)	5224	86.1% (4498)	99.85% (5216)	99.14% (5179)	70.29% (3672)	45.73% (2389)	9.74% (509)	9.23% (482)	5.23% (273)	3.71% (194)	87.61% (4577)
	Tate (Q430682)	5080	98.41% (4999)	99.98% (5079)	93.13% (4731)	95.18% (4835)	68.31% (3470)	8.76% (445)	6.5% (330)	3.82% (194)	1.0% (51)	95.04% (4828)
	National Gallery (Q180788)	4791	99.87% (4785)	99.87% (4785)	96.99% (4647)	99.31% (4758)	69.96% (3352)	34.63% (1659)	26.93% (1290)	14.53% (696)	9.29% (445)	99.39% (4762)
	National Museum of Art, Architecture and Design (Q1132918)	4744	99.68% (4729)	100.0% (4744)	80.67% (3827)	98.95% (4694)	51.94% (2464)	9.87% (468)	7.12% (338)	3.98% (189)	0.4% (19)	1.5% (71)
	Musée d'Orsay (Q23402)	4459	100.0% (4459)	99.96% (4457)	76.16% (3396)	95.27% (4248)	96.05% (4283)	24.98% (1114)	84.08% (3749)	49.72% (2217)	1.26% (56)	7.85% (350)
	Cleveland Museum of Art (Q657415)	4258	85.25% (3630)	99.79% (4249)	98.19% (4181)	32.1% (1367)	21.94% (934)	72.24% (3076)	10.19% (434)	5.8% (247)	2.09% (89)	99.41% (4233)
	National Gallery of Canada (Q1068063)	4172	96.91% (4043)	100.0% (4172)	91.47% (3816)	97.77% (4079)	51.49% (2148)	3.76% (157)	5.61% (234)	1.22% (51)	1.37% (57)	90.48% (3775)
	National Gallery of Art (Q214867)	4171	99.5% (4150)	99.9% (4167)	94.1% (3925)	92.86% (3873)	71.59% (2986)	64.59% (2694)	29.8% (1243)	23.06% (962)	5.42% (226)	99.66% (4157)
	Österreichische Galerie Belvedere (Q303139)	4153	98.34% (4084)	99.95% (4151)	85.82% (3564)	86.32% (3585)	61.57% (2557)	59.21% (2459)	6.57% (273)	3.11% (129)	3.06% (127)	94.41% (3921)
	Tartu Art Museum's painting collection (Q55986655)	4071	97.86% (3984)	0.0% (0)	88.04% (3584)	98.31% (4002)	99.61% (4055)	2.24% (91)	3.54% (144)	1.74% (71)	0.0% (0)	0.0% (0)
	Detroit Institute of Arts (Q1201549)	4002	88.21% (3530)	99.85% (3996)	61.07% (2444)	16.52% (661)	29.91% (1197)	52.87% (2116)	16.07% (643)	2.5% (100)	3.3% (132)	99.58% (3985)
	Art Gallery of New South Wales (Q705551)	3930	90.81% (3569)	99.97% (3929)	68.19% (2680)	5.47% (215)	5.17% (203)	5.83% (229)	6.46% (254)	3.26% (128)	0.84% (33)	0.74% (29)
	Museum of Fine Arts (Q49133)	3879	99.85% (3873)	99.66% (3866)	69.94% (2713)	88.1% (3421)	62.57% (2427)	59.94% (2325)	16.81% (652)	8.56% (332)	3.92% (152)	97.29% (3774)
	Royal Museums Greenwich (Q7374509)	3788	90.1% (3413)	99.95% (3786)	49.63% (1880)	60.53% (2293)	47.36% (1794)	18.06% (684)	3.78% (143)	2.35% (89)	0.0% (0)	99.82% (3781)
	National Galleries of Scotland (Q2051997)	3787	95.78% (3627)	100.0% (3787)	75.42% (2856)	94.22% (3568)	71.69% (2715)	8.05% (305)	8.93% (338)	4.86% (184)	2.38% (90)	96.59% (3658)
	Musée national des beaux-arts du Québec (Q2338135)	3683	99.46% (3663)	100.0% (3683)	85.07% (3133)	99.21% (3654)	48.66% (1792)	0.41% (15)	38.07% (1402)	0.3% (11)	1.41% (52)	99.1% (3650)
	Kunsthistorisches Museum (Q95569)	3346	96.65% (3234)	99.7% (3336)	80.1% (2680)	80.45% (2692)	40.23% (1346)	56.1% (1877)	16.14% (540)	7.26% (243)		

Sum of all paintings

WikiProject every politician

Goals & achievements

- Started by mySociety
- Improving information on politicians
- Covering every country in common format
- Data model is work in progress
- As of now covers ~70 countries

Data model

- The usual "human properties"
- position held (P39), with qualifiers
    - start time (P580) / end time (P582)
    - parliamentary term (P2937)
    - elected in (P2715)
    - parliamentary group (P4100)
    - electoral district (P768)
    - end cause (P1534)
    - replaces (P1365) / replaced by (P1366)

Data model

- Elections
- Election districts
- Parliamentary terms
- Cabinets of ministers
- Political positions
  - Head of government
  - Member of parliament

Data model

Maintenance lists

- Generated by Listeria/SPARQL queries
- Member list of each parliament
- Current member biographic data
- Checks
  - Missing qualifiers (no term, group, district)
  - Missing biographic data (gender, birthdate, given/first name)

Data completeness (LV)

position held (P39) qualifiers   [edit | edit source]

parliamentary term (P2937): **100%** complete

start time (P580): **100%** complete

parliamentary group (P4100): **100%** complete

electoral district (P768): **0%** complete

elected in (P2715): **0%** complete

Data completeness (LV)

Biographical data [edit | edit source]

sex or gender (P21): **100%** complete
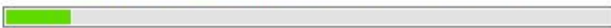
date of birth (P569): **100%** complete
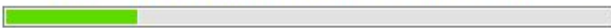
place of birth (P19): **57%** complete

given name (P735): **97%** complete

family name (P734): **11%** complete
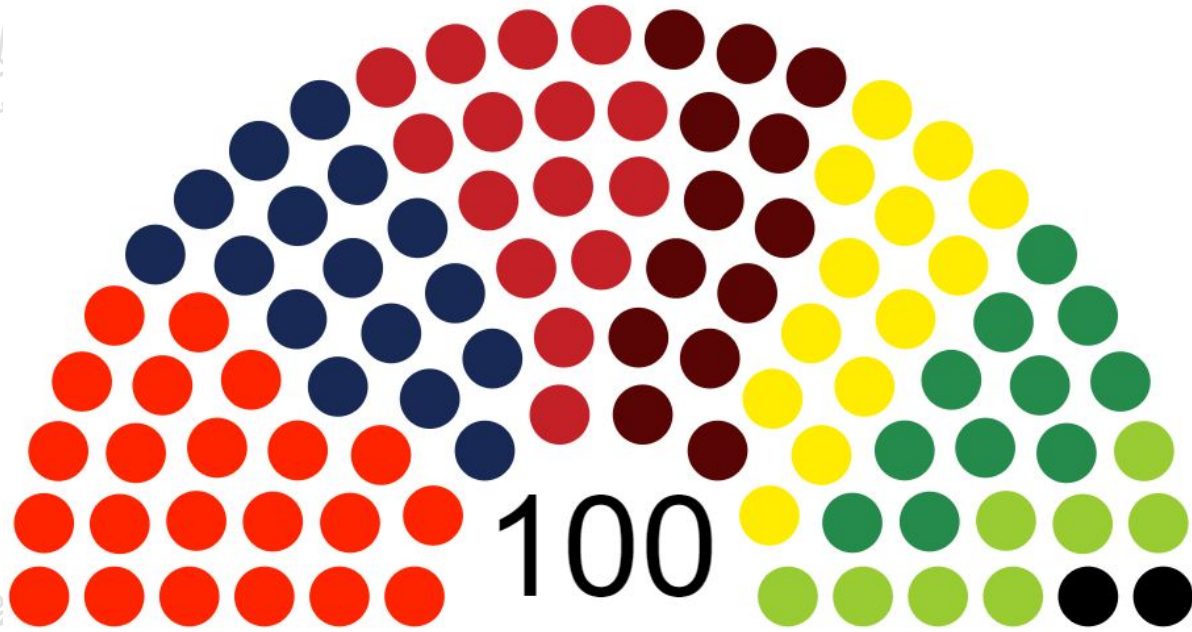
Twitter username (P2002): **22%** complete

Facebook ID (P2013): **1%** complete

Using the data

Generating parliament diagrams

100

Seeking help

- Tool manuals on Wikidata
- DO ask for help!
- Wikidata project chat (multilingual)
- [Wikidata Telegram group](#)
- [IRC](#)
- [Request a Query](#)

THANK YOU