EntitySchemas

Data Modelling Days 2023

What are EntitySchemas even?

EntitySchemas allow us to encode and test data modeling decisions in a machine-readable way

- Encodes a standardised structure for a subject area of Wikidata's data
- Different standards for the syntax exist (ShEx, SHACL) - we work with ShEx

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

start = @<human>

<human> EXTRA wdt:P31 {
  wdt:P31 [wd:Q5];
  wdt:P18 . * ;                # image (portrait)
  wdt:P21 [wd:Q48270 wd:Q48279 wd:Q179294 wd:Q189125 wd:Q207959 wd:Q301702 wd:Q350374
wd:Q505371 wd:Q660882 wd:Q746411 wd:Q859614 wd:Q1052281 wd:Q1097630 wd:Q1289754 wd:Q1399232
wd:Q2449503 wd:Q3177577 wd:Q3277905 wd:Q6581072 wd:Q6581097 wd:Q7130936 wd:Q12964198
wd:Q15145778 wd:Q15145779 wd:Q18116794 wd:Q27679684 wd:Q27679766 wd:Q52261234 wd:Q93954933
wd:Q93955709 wd:Q96000630 wd:Q99485724 wd:Q99485785]?;   # gender
  wdt:P19 . ?;                    # place of birth
  wdt:P20 . ?;                    # place of death
  wdt:P569 . ? ;                  # date of birth
  wdt:P570 . ? ;                 # date of death
  wdt:P735 . * ;                  # given name
  wdt:P734 . * ;                  # family name
  wdt:P106 . * ;                  # occupation
  wdt:P1559 . ? ;               #name in native language
  wdt:P27 @<country> *;           # country of citizenship
  wdt:P22 @<human> *;             # father
  wdt:P25 @<human> *;             # mother
  wdt:P3373 @<human> *;           # sibling
  wdt:P26 @<human> *;             # spouse
  wdt:P40 @<human> *;             # children
  wdt:P1038 @<human> *;           # relatives
  wdt:P103 @<language> *;         # native language
  wdt:P1412 @<language> *;        # languages spoken, written or signed
  wdt:P6886  @<language> *;       # writing language
  rdfs:label rdf:langString+;
}

<country> EXTRA wdt:P31 {
  wdt:P31 [wd:Q6256 wd:Q3024240 wd:Q3624078] +;
}

<language> EXTRA wdt:P31 {
  wdt:P31 [wd:Q34770 wd:Q1288568] +;
}
```

Contents [hide]

geographic entity [edit | edit source]

label	description	aliases	class/Property	dependencies
adminstrative district in Turkey (E133)	schema for administrative district in Turkey		district of Turkey (Q1147395)	
Australian national parks (E30)	basic schema for national parks in Australia		national park of Australia (Q18618819), country (P17)=Australia (Q408)	
ceremonial county of England (E71)	basic schema for a ceremonial county in England		ceremonial county of England (Q180673)	
Chilean Cities (E172)	Schema to validate Chilean Cities		Chile (Q298), city (Q515)	
Chilean Volcanoes (E171)	Chilean Volcanoes		Chile (Q298), volcano (Q8072)	
city (E100)	schema of a city		city (Q515)	Imported by: Research institute (E337)
European Bathing waters (E319)	Bathing waters in Europe identified with Eionet bathingWaterIdentifier		Eionet bathingWaterIdentifier (P9616)	
health in region (E88)	basic scheme for health by region/country items		health by country or region (Q64027457)	
Lunar crater (E106)	schema for lunar craters	moon crater	lunar crater (Q1348589)	
metropolitan statistical area (E140)	schema for metropolitan statistical areas		Metropolitan Statistical Area (Q1768043)	
mountain/hill (E339)	General schema for mountains/hills, useful as a base for more specific schema		mountain (Q8502), hill (Q54050)	
municipio of Rome (E104)	schema for validating current municipi of Rome [draft]		municipio of Rome (Q525504)	
NationalFlag (E50)			national flag (Q186516)	
Nazca lines (E148)	schema for geoglyphs that make up the Nazca lines		geoglyph (Q45791), part of (P361)=Nazca Lines (Q2620036)	
sovereign state (E32)	basic schema for sovereign states		sovereign state (Q3624078)	Imported by: Research institute (E337)
Swedish beach (E305)	with P9616 number i.e. more official beach		bathing water (Q567998), country (P17)=Sweden (Q34)	
Swedish dogbeach (E304)	beskriver ett hundbad		dog beach (Q35101126), country (P17)=Sweden (Q34)	
townland (E295)	schema to describe a townland (small geographic division of Ireland)		townland (Q2151232)	
trail (E114)	schema for trails	hiking trail	trail (Q628179)	
volcano (E80)	General schema for volcanos on Wikidata	vulcano	volcano (Q8072)	

But why?!

We need EntitySchemas because...

- **Data quality:** They are an important tool for our editors to rein in the data chaos.
- **Data utility:** Consistently modelled data is easier to reuse.
- **Decision making support:** Galvanising modeling decisions around a shared artifact makes it easier to make and stick to modeling decisions.
- **Tool ecosystem:** Machine-readable encoding of Wikidata's modeling decisions makes it possible to build better tools, e.g. a form for adding new Items that automatically checks if the input is good.
- **Subsetting:** Some groups are using EntitySchemas to define which Items should be in a particular subset they are interested in.

How does it work?

- EntitySchemas are encoded in pages in Wikidata
- They have:
    - ID prefixed by the letter E:
    - Terms (label, description, aliases) in various languages
    - A section for the definition
    - Version history, discussion page, etc.

EntitySchema Discussion | Read View history More | Search Wikidata

human (E10)

language code	label	description	aliases	edit
en	human	simple schema for humans	person | human being	edit
ca	humà	schema per a éssers humans	persona | ésser humà	edit
cs	osoba	jednoduché schéma pro člověka	člověk | osoba	edit
da	menneske		person	edit
de	Mensch	einfaches Objektschema des Menschen	Person	edit
el	άνθρωπος			edit
en-gb	human	simple schema for humans	person | human being	edit
eo	homo	simpla skemo por homoj	persono	edit
es	ser humano	esquema simple para una persona	persona	edit
et	inimene	lihtne skeem inimese jaoks		edit
fi	ihminen	yksinkertainen skeema kohteelle ihminen	henkilö	edit
fr	humain	schéma simple pour un être humain	personne	edit
fy	minske		persoan	edit
gl	ser humano	esquema simple para definir unha persoa	persoa	edit
hu	ember		személy	edit
it	umano	schema per descrivere un essere umano	persona | individuo | essere umano	edit
ja	ヒト	ヒト記述用のスキーマ	人間	edit
ko	[illegible]	[illegible]	[illegible] | [illegible]	edit
lv	cilvēks		persona	edit
ms	manusia	skema penerangan manusia	orang | insan	edit
nb	menneske	enkelt skjema for mennesker	person	edit
nl	mens	simpel schema voor mensen	persoon	edit
pt	humano	esquema simples para humanos		edit
pt-br	humano	esquema para descrever seres humanos		edit
ro	om		persoană | ființă umană	edit
ru	человек	простая схема для людей	персона | персоналия	edit
sk	osoba	schéma pre dátové položky ľudí	človek	edit
sq	njeri	skema e thjeshtë për njerëzit	personi | qënie njerëzore	edit
sr	особа	проста шема за особу	човек	edit
sv	människa	ett enklare schema för människor		edit
tr	insan	insanlar için basit şema	kişi	edit
vec	èsare uman		omo | uman | persona | personajo	edit

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

start = @<human>

<human> EXTRA wdt:P31 {
  wdt:P31 [wd:Q5];
  wdt:P18 . * ;              # image (portrait)
  wdt:P21 [wd:Q48270 wd:Q48279 wd:Q179294 wd:Q189125 wd:Q207959 wd:Q301702 wd:Q350374 wd:Q505371 wd:Q660882 wd:Q746411 wd:Q859614 wd:Q1052281 wd:Q1097630
wd:Q1289754 wd:Q1399232 wd:Q2449503 wd:Q3177577 wd:Q3277905 wd:Q6581072 wd:Q6581097 wd:Q7130936 wd:Q12964198 wd:Q15145778 wd:Q15145779 wd:Q18116794
wd:Q27679684 wd:Q27679766 wd:Q52261234 wd:Q93954933 wd:Q93955709 wd:Q96000630 wd:Q99485724 wd:Q99485785]?;   # gender
  wdt:P19 . ?;                  # place of birth
  wdt:P20 . ?;                  # place of death
  wdt:P569 . ? ;                # date of birth
  wdt:P570 . ? ;               # date of death
  wdt:P735 . * ;                # given name
  wdt:P734 . * ;                # family name
  wdt:P106 . * ;                # occupation
  wdt:P1559 . ? ;             #name in native language
  wdt:P27 @<country> *;          # country of citizenship
  wdt:P22 @<human> *;            # father
  wdt:P25 @<human> *;            # mother
  wdt:P3373 @<human> *;          # sibling
  wdt:P26 @<human> *;            # spouse
  wdt:P40 @<human> *;            # children
  wdt:P1038 @<human> *;          # relatives
  wdt:P103 @<language> *;        # native language
  wdt:P1412 @<language> *;       # languages spoken, written or signed
  wdt:P6886  @<language> *;      # writing language
  rdfs:label rdf:langString+;
}

<country> EXTRA wdt:P31 {
  wdt:P31 [wd:Q6256 wd:Q3024240 wd:Q3624078] +;
}

<language> EXTRA wdt:P31 {
  wdt:P31 [wd:Q34770 wd:Q1288568] +;
}
```

check entities against this Schema | edit

human (E10)

language code	label	description	aliases	edit
en	human	simple schema for humans	person | human being	edit
ca	humà	schema per a éssers humans	persona | ésser humà	edit
cs	osoba	jednoduché schéma pro člověka	člověk | osoba	edit
da	menneske		person	edit
de	Mensch	einfaches Objektschema des Menschen	Person	edit
el	άνθρωπος			edit
en-gb	human	simple schema for humans	person | human being	edit
eo	homo	simpla skemo por homoj	persono	edit
es	ser humano	esquema simple para una persona	persona	edit
et	inimene	lihtne skeem inimese jaoks		edit
fi	ihminen	yksinkertainen skeema kohteelle ihminen	henkilö	edit
fr	humain	schéma simple pour un être humain	personne	edit
fy	minske		persoan	edit
gl	ser humano	esquema simple para definir unha persoa	persoa	edit
hu	ember		személy	edit
it	umano	schema per descrivere un essere umano	persona | individuo | essere umano	edit
ja	ヒト	ヒト記述用のスキーマ	人間	edit
ko	□□	□□□□ □□□□ □□□□ □□□□	□□□ | □□	edit
lv	cilvēks		persona	edit
ms	manusia	skema penerangan manusia	orang | insan	edit
nb	menneske	enkelt skjema for mennesker	person	edit
nl	mens	simpel schema voor mensen	persoon	edit
pt	humano	esquema simples para humanos		edit
pt-br	humano	esquema para descrever seres humanos		edit
ro	om		persoană | ființă umană	edit
ru	человек	простая схема для людей	персона | персоналия	edit

check entities against this Schema | edit

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

start = @<human>

<human> EXTRA wdt:P31 {
  wdt:P31 [wd:Q5];
  wdt:P18 . * ;                # image (portrait)
  wdt:P21 [wd:Q48270 wd:Q48279 wd:Q179294 wd:Q189125 wd:Q207959 wd:Q301702 wd:Q350374 wd:Q505371 wd:Q660882 wd:Q746411 wd:Q859614 wd:Q1052281 wd:Q1097630
wd:Q1289754 wd:Q1399232 wd:Q2449503 wd:Q3177577 wd:Q3277905 wd:Q6581072 wd:Q6581097 wd:Q7130936 wd:Q12964198 wd:Q15145778 wd:Q15145779 wd:Q18116794
wd:Q27679684 wd:Q27679766 wd:Q52261234 wd:Q93954933 wd:Q93955709 wd:Q96000630 wd:Q99485724 wd:Q99485785]?;   # gender
  wdt:P19 . ?;                    # place of birth
  wdt:P20 . ?;                    # place of death
  wdt:P569 . ? ;                  # date of birth
  wdt:P570 . ? ;                 # date of death
  wdt:P735 . * ;                  # given name
  wdt:P734 . * ;                  # family name
  wdt:P106 . * ;                  # occupation
  wdt:P1559 . ? ;               #name in native language
  wdt:P27 @<country> *;           # country of citizenship
  wdt:P22 @<human> *;             # father
  wdt:P25 @<human> *;             # mother
  wdt:P3373 @<human> *;           # sibling
  wdt:P26 @<human> *;             # spouse
  wdt:P40 @<human> *;             # children
  wdt:P1038 @<human> *;           # relatives
  wdt:P103 @<language> *;         # native language
  wdt:P1412 @<language> *;        # languages spoken, written or signed
  wdt:P6886  @<language> *;       # writing language
  rdfs:label rdf:langString+;
}

<country> EXTRA wdt:P31 {
  wdt:P31 [wd:Q6256 wd:Q3024240 wd:Q3624078] +;
}

<language> EXTRA wdt:P31 {
  wdt:P31 [wd:Q34770 wd:Q1288568] +;
}
```

```
<human> EXTRA wdt:P31 {
  wdt:P31 [wd:Q5];
  wdt:P18 . * ;                  # image (portrait)
  wdt:P21 [wd:Q48270 wd:Q48279 wd:Q179294 wd:Q189125
wd:Q207959 wd:Q301702 wd:Q350374 wd:Q505371 wd:Q660882
wd:Q746411 wd:Q859614 wd:Q1052281 wd:Q1097630 wd:Q1289754
wd:Q1399232 wd:Q2449503 wd:Q3177577 wd:Q3277905 wd:Q6581072
wd:Q6581097 wd:Q7130936 wd:Q12964198 wd:Q15145778
wd:Q15145779 wd:Q18116794 wd:Q27679684 wd:Q27679766
wd:Q52261234 wd:Q93954933 wd:Q93955709 wd:Q96000630
wd:Q25388691 wd:Q56315990]?;    # gender
```

```
wdt:P19 . ?;                    # place of birth
wdt:P20 . ?;                    # place of death
wdt:P569 . ? ;                  # date of birth
wdt:P570 . ? ;                 # date of death
wdt:P735 . * ;                  # given name
wdt:P734 . * ;                  # family name
wdt:P106 . * ;                  # occupation
```

What’s building on this?

ShEx Simple

kakapo schema (E61)

language code	label	description	aliases	edit
en	kakapo schema	schema for kakapo individuals on wikidata		edit
ja	フクロウオウム	フクロウオウムを記述するためのスキーマ		edit

check entities against this Schema | edit

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX pr: <http://www.wikidata.org/prop/reference/>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX eco: <http://purl.obolibrary.org/obo/ECO_>
PREFIX psv: <http://www.wikidata.org/prop/statement/value/>

## SELECT * WHERE {?cellline wdt:P31 wd:Q179959 .}
start = @<#kakapoSchema>

<#kakapoSchema> {
      #instance of
      p:P31 { ps:P31  IRI * };

      #sex or gender
      p:P21 { ps:P21  [
                       wd:Q44148 #male organism
                       wd:Q43445 #female organism
                      ] };

      #image
      p:P18 { ps:P18  IRI * } ?;

      #place of birth
      p:P19 { ps:P19  IRI * } ?;

      #named after
      p:P138 { ps:P138  IRI * } ?;

      #date of birth
      p:P569 { psv:P569  IRI *; ps:P569  xsd:dateTime *;  } ?;

      #Commons category
      p:P373 { ps:P373  xsd:string * } ?;

      #child
      p:P40 { ps:P40  IRI * } ?;

      #father
      p:P22 { ps:P22  IRI * } ?;

      #Twitter username
      p:P2002 { ps:P2002  xsd:string * } ?;

      #date of death
      p:P570 { ps:P570  xsd:dateTime *; psv:P570  IRI *;  } ?;

}
```

ShEx Simple


ShEx2 — Simple Online Validator
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX pr: <http://www.wikidata.org/prop/reference/>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX eco: <http://purl.obolibrary.org/obo/ECO_>
PREFIX psv: <http://www.wikidata.org/prop/statement/value/>
SELECT * WHERE {?cellline wdt:P31 wd:Q179959 .}
start = @<#kakapoSchema>
<#kakapoSchema> {
#instance of
p:P31 { ps:P31 IRI * };
#sex or gender
p:P21 { ps:P21 [
wd:Q44148 #male organism
wd:Q43445 #female organism
] };
#image
p:P18 { ps:P18 IRI * } ?;
run query to fetch entities (ctl-enter)
Query
Entities to check
SELECT ?id WHERE {
# ...
}

ShEx Simple

ShEx2 — Simple Online Validator

```
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX pr: <http://www.wikidata.org/prop/reference/>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX eco: <http://purl.obolibrary.org/obo/ECO_>
PREFIX psv: <http://www.wikidata.org/prop/statement/value/>

## SELECT * WHERE {?cellline wdt:P31 wd:Q179959 .}
start = @<#kakapoSchema>

<#kakapoSchema> {
      #instance of
      p:P31 { ps:P31  IRI * };

      #sex or gender
      p:P21 { ps:P21  [
                       wd:Q44148 #male organism
                       wd:Q43445 #female organism
                      ] };

      #image
      p:P18 { ps:P18  IRI * } ?;
```

validate (ctl-enter)

Query | Entities to check

```
    ?statement0 (ps:P10241/(wdt:P279*)) wd:Q179959.
  }
 }
}
```

✓wd:Q7530532@START

✓wd:Q63254018@START

✓wd:Q63254090@START

✓wd:Q107694640@START

✓wd:Q107695339@START

✓wd:Q107708464@START

✓wd:Q107708488@START

User script

By Teester

User script

By Teester

Denny Vrandečić (Q18618629)

ORES predicted quality: A (4.74)

Croatian computer scientist; former co-developer of Semantic MediaWiki and Wikidata; currently the lead developer of the Wikifunctions project, and an employee of the Wikimedia Foundation as a Head of Special Projects, Structured Content

Denny Vrandecic | Zdenko Vrandecic | Zdenko Vrandečić

E10 Automatically check schema

Checking against **E10:human**:

Required properties	Optional properties	Other properties
correct P31 - instance of	present P103 - native language	Not in schema P101 - field of work
	missing P1038 - relative	Not in schema P1015 - NORAF ID
	present P106 - occupation	Not in schema P1026 - academic thesis
	correct P1412 - languages spoken, written or signed	Not in schema P108 - employer
		Not in schema P1477 - birth name
	missing P1559 - name in native language	Not in schema P1581 - official blog
	present P18 - image	Not in schema P184 - doctoral advisor
	present P19 - place of birth	Not in schema P1960 - Google Scholar author ID

In more languages edit

Statements

instance of — E10: correct
- human ••• — edit — correct
- 1 reference
- + add value

image — E10: present
- edit — allowed
- 
- Dr. Denny Vrandečić-2.jpg
- 5,425 × 3,617; 5.12 MB
- 0 references
- + add reference
- + add value

sex or gender — E10: correct
- male ••• — edit — correct
- 2 references
- + add value

ca Denny Vrandečić
de Denny Vrandečić
en Denny Vrandečić
fr Denny Vrandečić
hr Zdenko Denny Vrandečić
no Denny Vrandečić
ru Врандечич, Денни
sh Zdenko Denny Vrandečić

Wikipedia (9 entries) edit

Wikibooks (0 entries) edit

Wikinews (1 entry) edit
ru Категория:Денни Врандечич

Wikiquote (0 entries) edit

Wikisource (0 entries) edit

Wikiversity (0 entries) edit

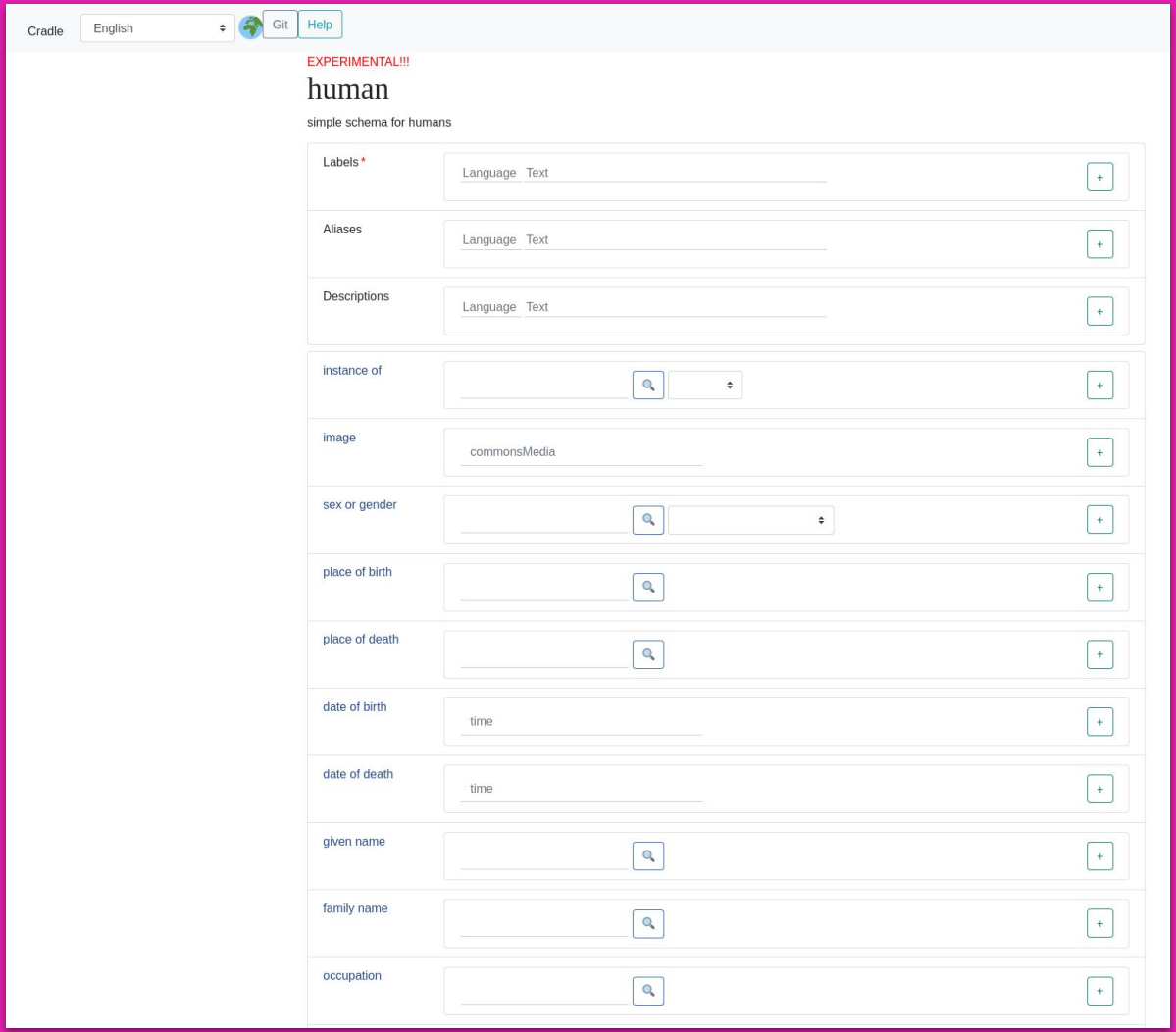
Wikivoyage (0 entries) edit

Wiktionary (0 entries) edit

Multilingual sites (1 entry) edit
commons Category:Denny Vrandečić

Cradle

By magnusmanske



How do EntitySchemas relate to Property Constraints?

EntitySchemas:

- Centered around classes
  - E.g. “Monuments should always be modelled this way.”
- Fairly powerful
- Following a semantic web standard

And they are currently...

- Not well integrated in Wikidata’s processes, tools and UI
- Comparatively hard to use and understand

Property Constraints:

- Centered around Properties
  - E.g. “Date of birth should always be used this way.”
- Limited in their expressivity
- Home-grown solution specific to Wikibase
- Very well integrated in Wikidata’s processes, tools and UI
- Comparatively easy to use and understand by now

We need both of them!

What is EntitySchema v2?

v2 is a step to bring more visibility to EntitySchemas

- EntitySchemas available in statements connecting them to classes of Items
- Increasing queryability in Query Service
- EntitySchemas will be legible for all users
  - Always displayed with their label i.e. Human (E10)
  - Translations / language fallbacks
- Alignment of EntitySchemas and Item UI e.g. termbox

Why are we doing this?

Expected impact

- More consistently modelled data
- More centralised discussions about modelling of particular classes
- Contributions made through editing tools are more in line with the modelling for the class in questions

Helpful links

- human (E10) - Wikidata
- Wikidata:Schemas
- Wikidata:WikiProject Schemas
- Wikidata:Database reports/EntitySchema directory
- Wikidata:Schema proposals
- Shape Expressions (ShEx) 2.1 Primer
- CRADLE
- User:Teester/EntityShape.js