Managing and Consuming Completeness Information for Wikidata Using COOL-WD

KRDB Research Centre, Free University of Bozen-Bolzano

Radityo Eko Prasojo, **Fariz Darari**, Simon Razniewski, Werner Nutt

COLD 2016 @ Kobe, Japan

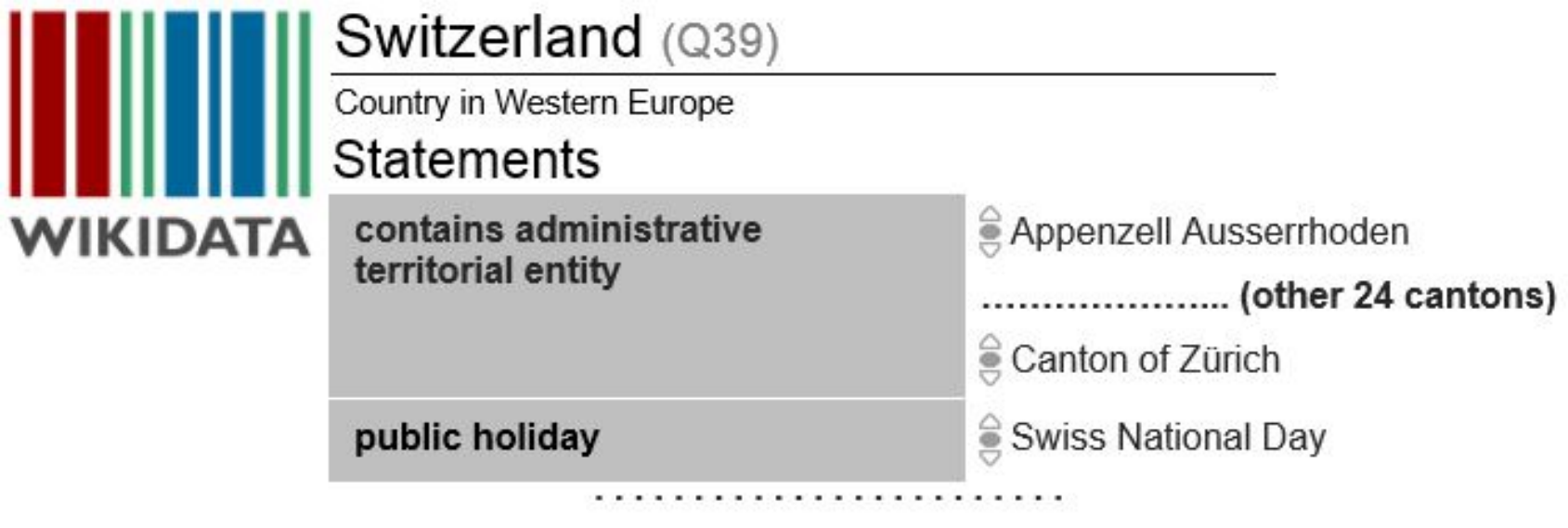
October 18, 2016

Supported by the project MAGIC, funded by the province of Bolzano

Web data is mostly incomplete

- Wikidata is missing the fact that Michael Sottile is a cast member of the movie Reservoir Dogs.
- As per YAGO, the average number of children per person is 0.02.
- DBpedia contains currently only 6 out of 35 Dijkstra Prize winners.

Cantons of Switzerland in Wikidata



All Swiss cantons by Swiss constitution

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

The federal Council
The portal of the Swiss government

Federal Constitution of the Swiss Confederation

of 18 April 1999 (Status as of 1 January 2016)

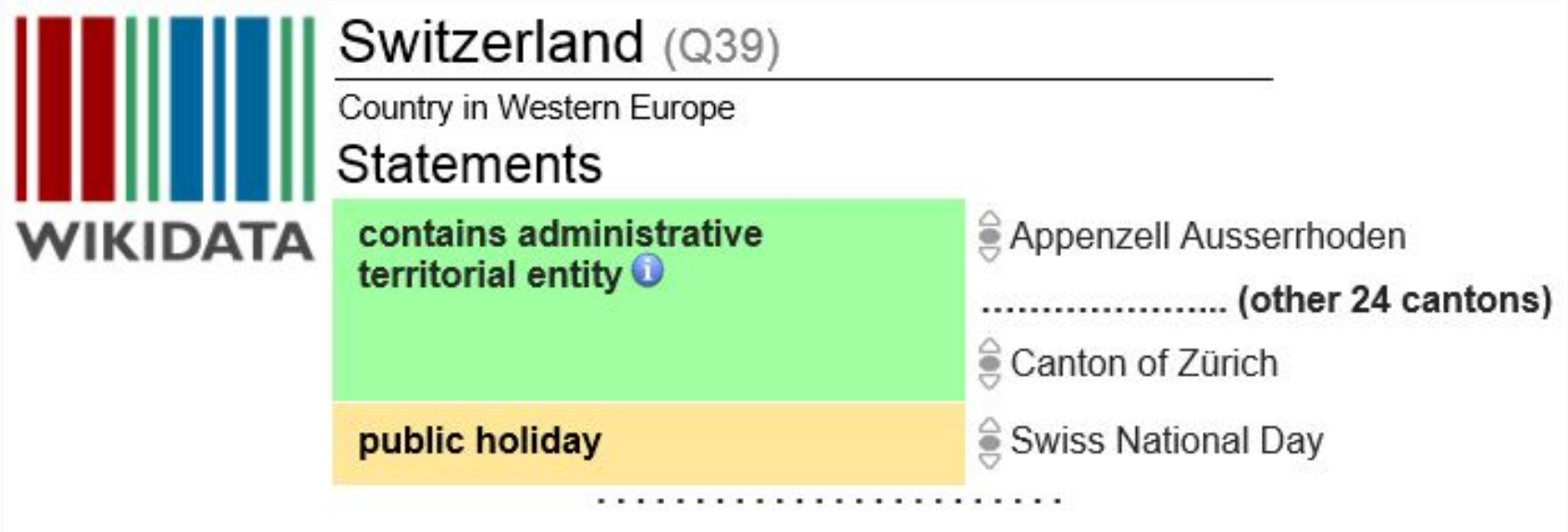
– Title 1 General Provisions

– Art. 1 The Swiss Confederation

Exact match with Wikidata!

The People and the Cantons of Zurich, Bern, Lucerne, Uri, Schwyz, Obwalden and Nidwalden, Glarus, Zug, Fribourg, Solothurn, Basel Stadt and Basel Landschaft, Schaffhausen, Appenzell Ausserrhoden and Appenzell Innerrhoden, St. Gallen, Graubünden, Aargau, Thurgau, Ticino, Vaud, Valais, Neuchâtel, Geneva, and Jura form the Swiss Confederation.

Wikidata is complete for cantons of Switzerland!



Completeness Statements[1]

Syntax:

$$(s, p)$$

Semantics:

Graph G has completeness statement (s, p)

$\downarrow$

G is complete for all p-values of s that exist in reality

Example:

Wikidata has completeness statement $(Q39, P150)$

$\downarrow$

Wikidata is complete for all
administrative territorial divisions/cantons ($=$ P150)
of Switzerland ($=$ Q39)

[1] Darari et al. Enabling Fine-Grained RDF Data Completeness Assessment. ICWE 2016.

Completeness Statement in RDF

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix spv: <http://completeness.inf.unibz.it/sp-vocab#> .
@prefix coolwd: <http://cool-wd.inf.unibz.it/resource/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

wd:Q2013 spv:hasSPStatement coolwd:statement-Q39-P150.
coolwd:statement-Q39-P150 a spv:SPStatement;
  spv:subject wd:Q39;
  spv:predicate wdt:P150;
  prov:wasAttributedTo [foaf:name "Fariz Darari";
    foaf:mbox <mailto:fariz.darari@stud-inf.unibz.it>];
  prov:generatedAtTime "2016-05-19T10:45:52"^^xsd:dateTime;
  prov:hadPrimarySource
<https://www.admin.ch/.../index.html#a1>.
```

COOL-WD

We have developed

a completeness management tool for Wikidata

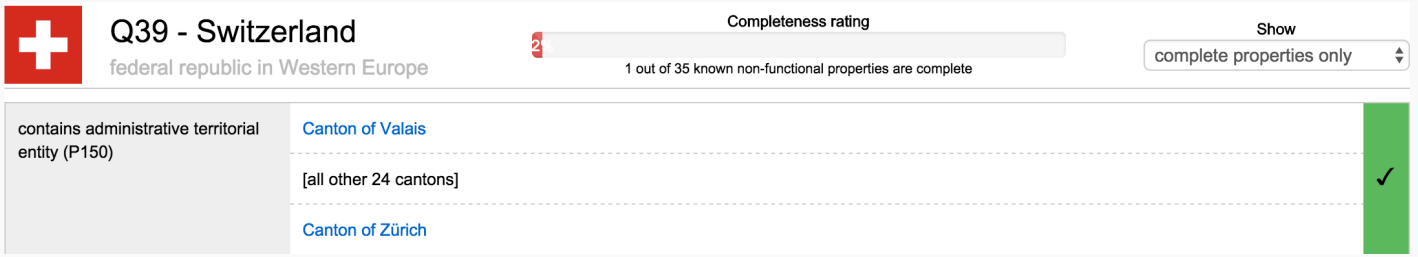
The management feature comprises:

- browsing Wikidata entities enriched with completeness statements
- adding and removing completeness statements
- updating completeness provenance

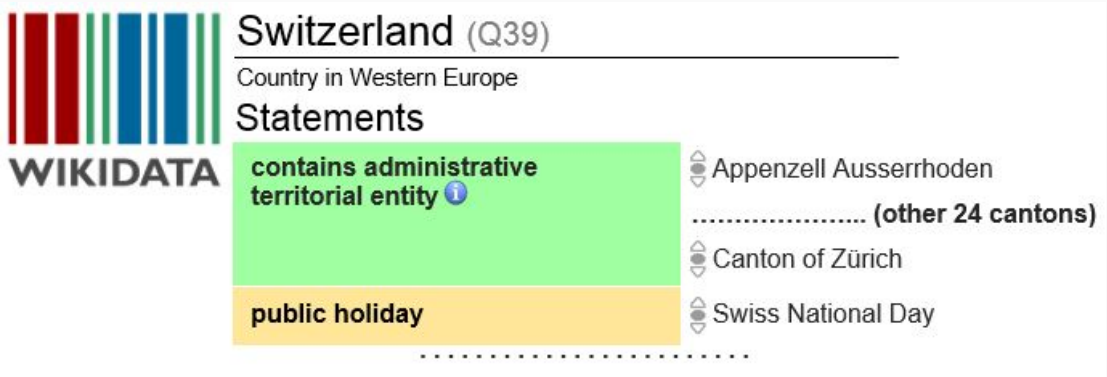
As for now, we have more than 10000 real completeness statements.

COOL-WD interfaces

1. The Web interface, accessible at `http://cool-wd.inf.unibz.it/`

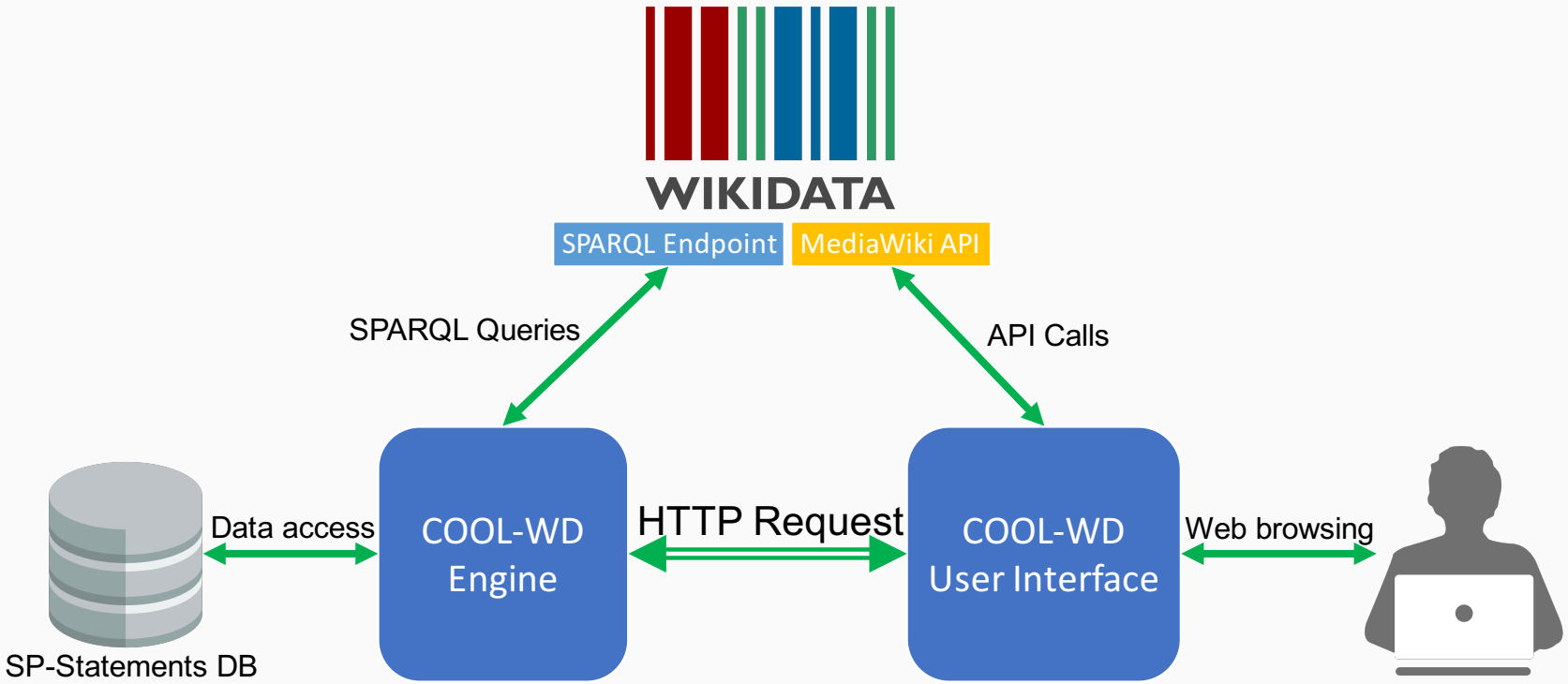


2. The COOL-WD Gadget, available for Wikidata users by importing our `cool-wd.js`[2] to their common.js page



[2] `https://www.wikidata.org/wiki/User:Fadirra/coolwd.js`

COOL-WD Web Interface: Architecture



Consuming completeness information using COOL-WD

- Completeness tracking of Wikidata entities

Q11911 - Canton of Bern

canton of Switzerland

Completeness rating

16%

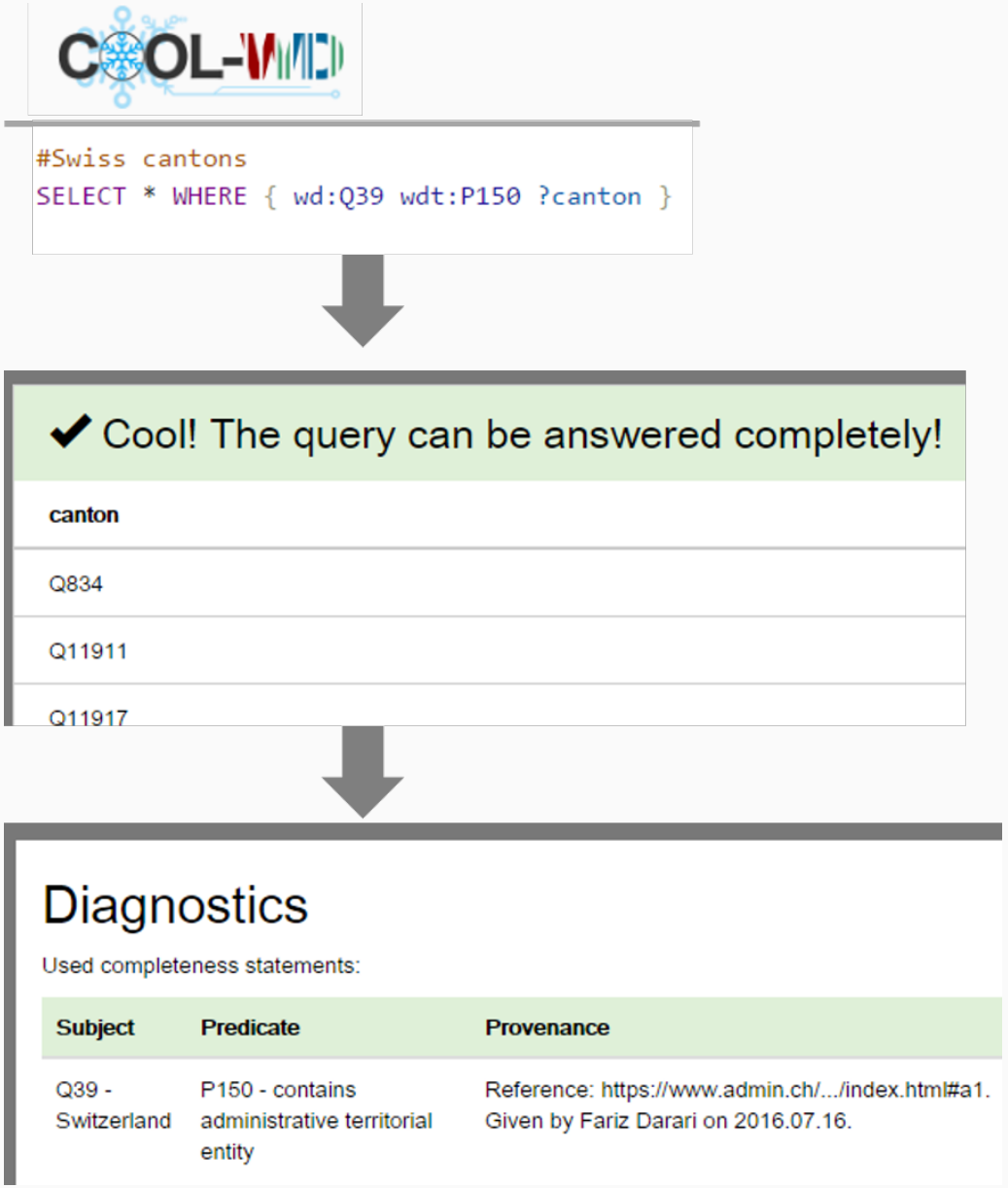
2 out of 12 known non-functional properties are complete

- Completeness analytics

Class name	#Objects	Property	Completeness percentage	Complete entities
Cantons of Switzerland	26	official language	15.38%	Canton of Geneva, Canton of Bern, Ticino, Canton of Zürich, Show less
Cantons of Switzerland	26	head of government	3.85%	Canton of Bern

Consuming completeness information using COOL-WD (2)

- Query completeness assessment



Conclusions

- Parts of information in Wikidata are complete, but so far there is no way to capture them
- COOL-WD manages and consumes completeness information of Wikidata
- Our framework can also be adopted by similar KBs like YAGO and DBpedia
- If you want more details on extracting completeness information from text: "How to Extract Cardinality Information from Text" (Wednesday evening poster session).

Thank you!

Backup slides

How to create completeness statements?

KB contributors

Paid crowd workers

Web extraction

COOL-WD, which is also pre-populated using the three approaches above.

Creating CS: KB contributors

- No-value statements
  - Stating the non-existence of information:
    Complete for all Elizabeth I's children (in reality she had none)
  - 7600 statements were imported
  - among the top 15: "member of political party", "spouse", "child", and "country of citizenship".

Creating KB: Paid crowd workers

- 900 SP-statements were crowd sourced
- Pricey
- Task is deemed too difficult for general crowd workers

Creating KB: Web extraction

- Mining cardinality information
  - Extracting information in Wikipedia like:
    Obama has two children
  - Then checking if the cardinality matches with the facts in Wikidata
  - 2200 statements were imported for the “child” relation