



Vorlesung Open Data

Wikidata und SPARQL

Termin 11, 9. Mai 2019

Benedikt Hitz-Gamper

Forschungsstelle Digitale Nachhaltigkeit
Institut für Wirtschaftsinformatik
Universität Bern

Terminübersicht Vorlesung

21. Februar 2019	Informationen zur Vorlesung, Einführung Open Data
28. Februar 2019	Open Government Data und das Impact Measuring Framework
7. März 2019	Digitale Nachhaltigkeit und Open Data
14. März 2019	Entstehung und Anwendung des Öffentlichkeitsgesetz
21. März 2019	Geistiges Eigentum und die Open-Bewegung
28. März 2019	Open Data Journalism
4. April 2019	Open Finance und Participatory Budgeting
11. April 2019	Open Transport Data
18. April 2019	Open Geodata
25. April 2019	Semesterferien
2. Mai 2019	Aid Transparency und Open Aid
9. Mai 2019	Linked Data und Semantic Web
16. Mai 2019	Open Corporate Data
23. Mai 2019	keine Vorlesung, am Nachmittag Abschlusspräsentationen Open Data Apps
30. Mai 2019	Auffahrt



Heutiger Ablauf

1. Wikidata
2. SPARQL und Linked Data
3. Federated SPARQL Queries



Introduction to Wikidata

- > Wikidata is a **free and open** knowledge base that can be read and edited by both humans and machines
- > Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, [...]
- > Wikidata also provides support to many other sites and services beyond just Wikipedia **e.g. Linked Data**
- > The content of Wikidata is available under a **free license**, exported using **standard formats**, and can be **interlinked** to other open data sets on the linked data web



Wikidata – Hands On

> Forschungsstelle Digitale Nachhaltigkeit

University of Bern Research Center for Digital Sustainability (Q63558709)

A research center of the institute of computer science at the university of Bern

 [edit](#)

▼ In more languages [Configure](#)

Language	Label	Description	Also known as
English	University of Bern Research Center for Digital Sustainability	A research center of the institute of computer science at the university of Bern	
British English	No label defined	No description defined	
German	Forschungsstelle Digitale Nachhaltigkeit der Universität Bern	Eine Forschungsstelle des Instituts für Informatik der Universität Bern	
Swiss German	No label defined	No description defined	

Wikidata – Erster eigener Beitrag

- > Wer leitet diese Forschungsstelle?
- > Kein Wikidata Account nötig, um Einträge zu ändern/erstellen

director/manager

Matthias Stürmer

✓ publish ✕ cancel ?

+ add qualifier

▼ 1 reference

reference URL

http://www.digitale-nachhaltigkeit.unibe.ch/about_us/team/index_eng.html

remove

remove

+ add

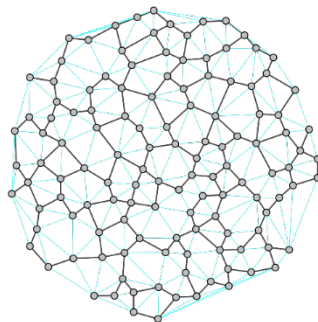
+ add reference

Wichtig: Wikidata ist eine sekundäre Quelle

Linked Data Graphs

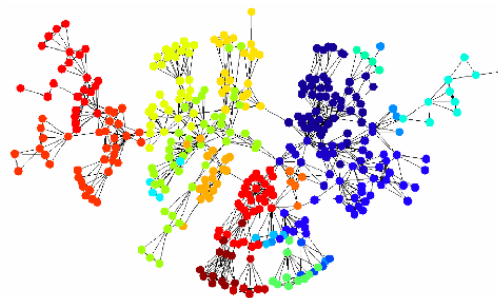
Herausforderung

- > Grosse Mengen von Linked Data sind auf den ersten Blick nicht einfach zu erfassen



Abhilfe

1. Zeige nur ein Subjekt mit allen Objekten
2. Benutze SPARQL um einen Sub-Graphen zu erzeugen



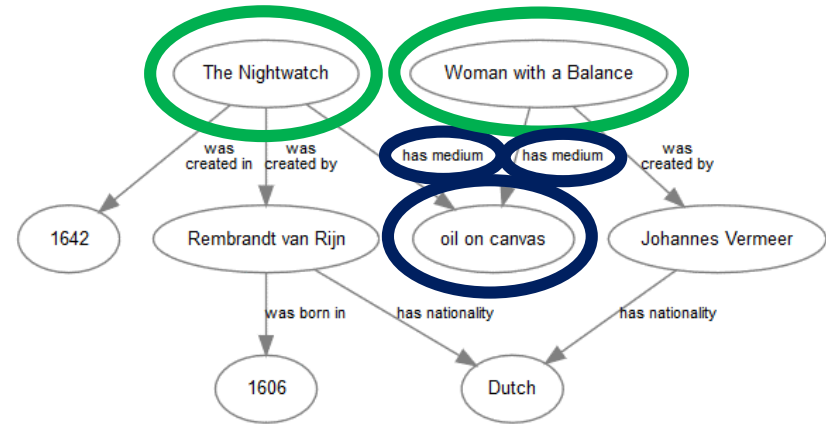
Was macht Linked Data so interessant?

- > RDF/Linked Data beinhaltet eine mächtige Abfragesprache
 - SPARQL: **S**PARQL **P**rotocol **A**nd **R**DF **Q**uery **L**anguage
- > Linked Data aus unterschiedlichen Quellen lassen sich vergleichsweise einfach kombinieren
 - sog. “Federated Queries”



SPARQL in a Nutshell

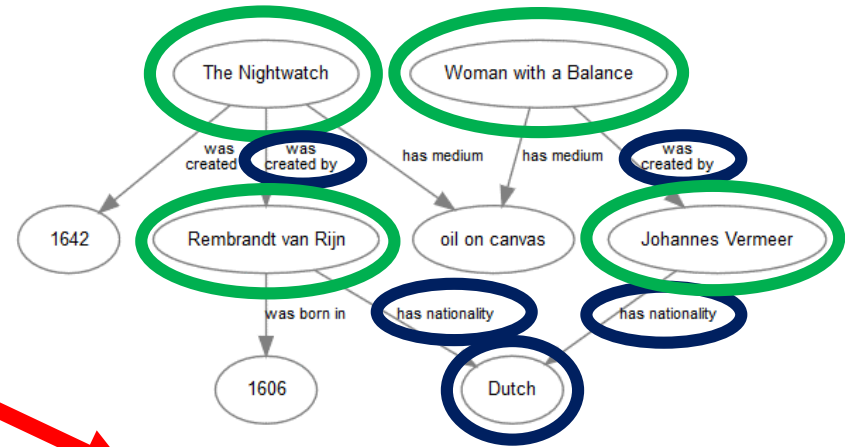
- > Pattern-Matching
- > E. g. search paintings made with «oil on canvas» technique



```
SELECT ?painting WHERE {  
  ?painting <has medium> <oil on canvas>.  
}
```

SPARQL in a Nutshell

- > Multiple Pattern-Matching
- > E. g. search paintings made by a Dutch artist



Variables (?xxx) with the same name
need to have the same value

```

SELECT ?painting WHERE {
  ?painting <was created by> ?artist.
  ?artist <has nationality> <Dutch>.
}
    
```

Wikidata – Query Service

- > Basic pattern of SPARQL queries:

```
SELECT ?subject WHERE {  
  ?subject ?predicate ?object.  
}
```



- > Query for research institutes on Wikidata:

```
SELECT ?institute WHERE {  
  ?institute wdt:P31 wd:Q31855.  
}
```

↑
instance of

↑
research institute

Wikidata – Query Service

> A little bit more interesting:

```
SELECT ?institute ?coord WHERE
{
  ?institute wdt:P31 wd:Q31855.
  ?institute wdt:P625 ?coord.
}
```

↑
coordinate location

<http://tinyurl.com/y7j8xbyi>

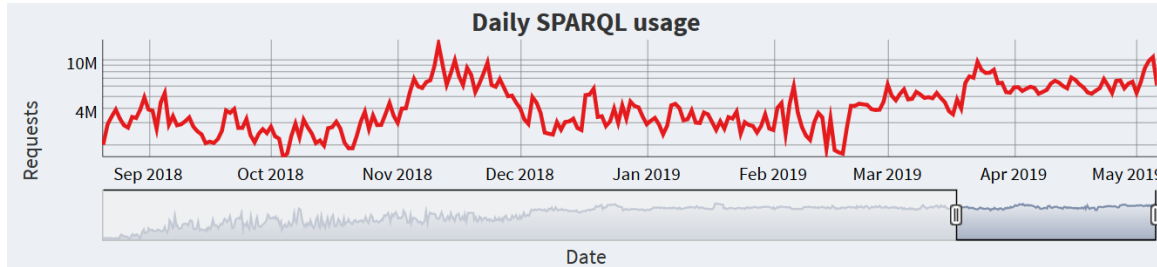


Wikidata – Multilingual at Heart

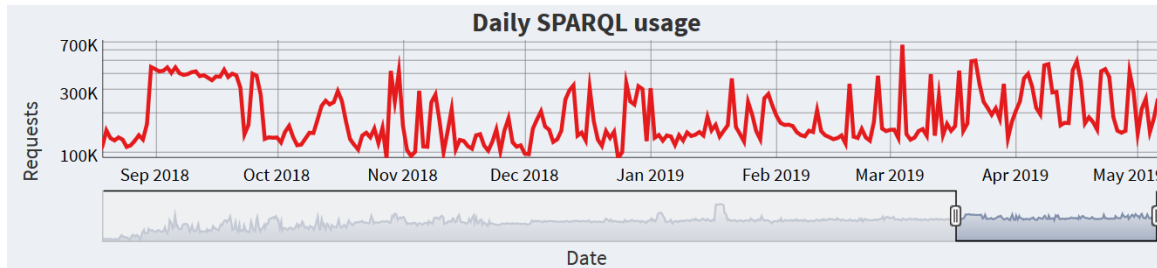
- > Why these whacky wdt:P... and wd:Q... numbers?
- > **Wikidata is multilingual at heart and does not favour any language, so no English identifiers but only numbers**
- > wdt:P... for predicates (properties)
- > wd:Q... for subjects and objects
- > How to get these numbers? → [Wikidata Query Service](#)

Wikidata und SPARQL

> Braucht das überhaupt irgendwer in der «realen» Welt?



Mit Bots



Ohne Bots

SPARQL Praxisbeispiel 1

- > Gesucht: Alle Berge in der Schweiz, die höher als 4000m sind



SPARQL Praxisbeispiel 1

> Grundgerüst für alle Wikidata SPARQL Queries

```
SELECT
```

```
WHERE {
```

```
    SERVICE wikibase:label { bd:serviceParam wikibase:language "de" }  
}
```


SPARQL Praxisbeispiel 1

> Komplette SPARQL Query

```
#defaultView:Map
```

```
SELECT ?mountain ?mountainLabel ?height ?location
```

```
WHERE {
```

```
    ?mountain wdt:P31 wd:Q8502.
```

```
    ?mountain wdt:P17 wd:Q39.
```

```
    ?mountain wdt:P2044 ?height filter (?height>4000).
```

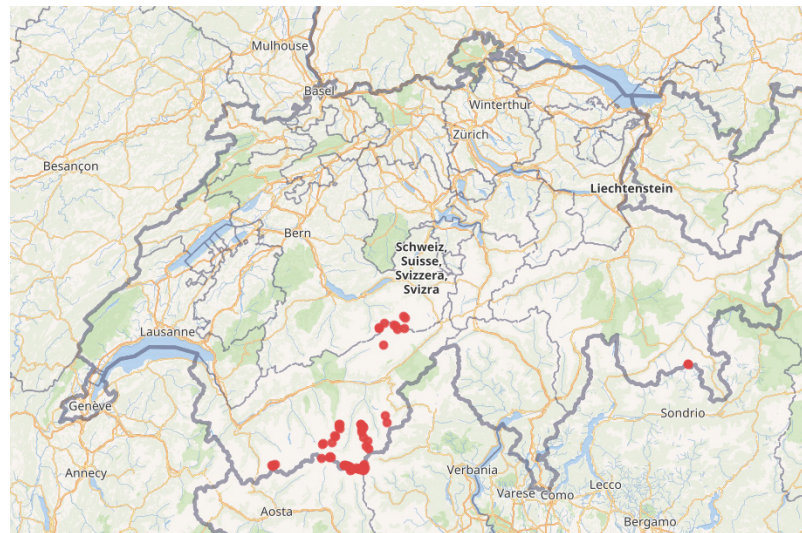
```
    ?mountain wdt:P625 ?location.
```

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "de" }
```

```
}
```

SPARQL Praxisbeispiel 1

- > Gesucht: Alle Berge in der Schweiz, die höher als 4000m sind
- > <http://tinyurl.com/hhwfgd9>
- > Hätte man auch mit tabellarischen Daten einfach lösen können...



SPARQL Praxisbeispiel 2

- > Which rivers flow into the river «Aare»?
- > <http://tinyurl.com/yaduwdfg>



```
#defaultView:Map
SELECT ?river ?riverLabel ?location WHERE {
  ?river wdt:P403 wd:Q1675;
    wdt:P625 ?location.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de". }
}
```

SPARQL Praxisbeispiel 2

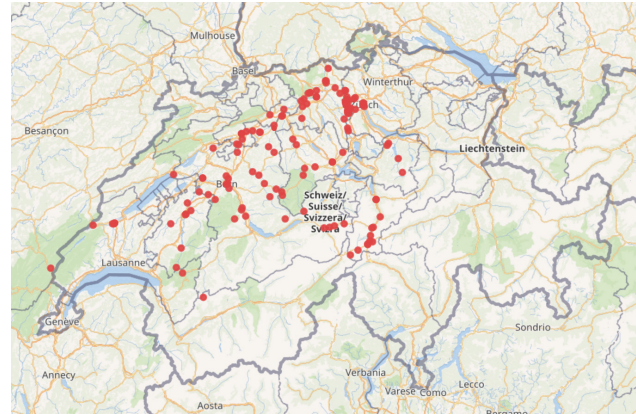
- > Which rivers flow into a river that flows into the «Aare»?
- > <http://tinyurl.com/yc9y4uub>



```
#defaultView:Map
SELECT ?river ?riverLabel ?location WHERE {
  ?river wdt:P403/wdt:P403 wd:Q1675;
  wdt:P625 ?location.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de". }
}
```

SPARQL Praxisbeispiel 2

- > Which rivers flow finally into the «Aare»?
- > <http://tinyurl.com/y8b856tu>

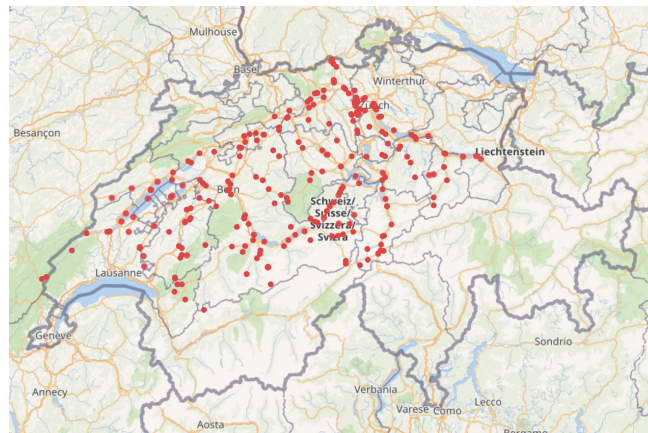


```
#defaultView:Map
SELECT ?river ?riverLabel ?location WHERE {
  ?river wdt:P403+ wd:Q1675;
    wdt:P625 ?location.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de". }
}
```

SPARQL Praxisbeispiel 2

- > What about Schils?
 - **Schils** flows (P403) into
 - **Seez** flows (P403) into
 - **Wahlensee** drains (P201) into
 - **Linth** flows (P403) into
 - **Lake Zurich** drains (P201) into
 - **Limmat** flows (P403) into
 - **Aare**

> <http://tinyurl.com/yayrkbxs>



```
#defaultView:Map
SELECT ?river ?riverLabel ?location WHERE {
  ?river wdt:P403*/wdt:P201*/wdt:P403*/wdt:P201*/wdt:P403* wd:Q1675;
  wdt:P625 ?location.
SERVICE wikibase:label { bd:serviceParam wikibase:language "de". }
```

Fazit

- > Zum Datenformat Linked Data gehört mit SPARQL eine mächtige Abfragesprache, mit der die Daten umfangreich gefiltert, aggregiert und kombiniert werden können (bspw. CSV/XML haben keine solchen Sprachen)
- > Linked Data ist gut geeignet, wenn die Daten schon inhärent netzwerkartig aufgebaut sind (Flüsse, die in Flüsse fließen, die in Flüsse fließen)

Federated SPARQL Queries

- > Daten liefern häufig spannende Einsichten, wenn diese aus verschiedenen Quellen kombiniert werden
- > Das Problem dabei: Verknüpfung aufwändige Handarbeit

Gemeinde	Einwohner
Neuenburg	34'000
Aeschlen	305
Davos	11'060

Gemeinde	Höhe [m ü. M.]
Neuchâtel	434
Oberdiessbach	605
Davos-Platz	1560

- > Sprache, Eingemeindungen, Teilgemeinden, etc...

Federated SPARQL Queries

- > Linked Data / RDF liefert die Möglichkeiten, um Daten einfach zu kombinieren
 - URI, sameAs, seeAlso, etc
- > Data Provider muss diese allerdings auch verwenden
 - Ein Aspekt von Data Quality
- > Gutes Beispiel: Linked Data Dienst des Bundesamts für Landestopografie
 - <https://ld.geo.admin.ch/boundaries/municipality/351>

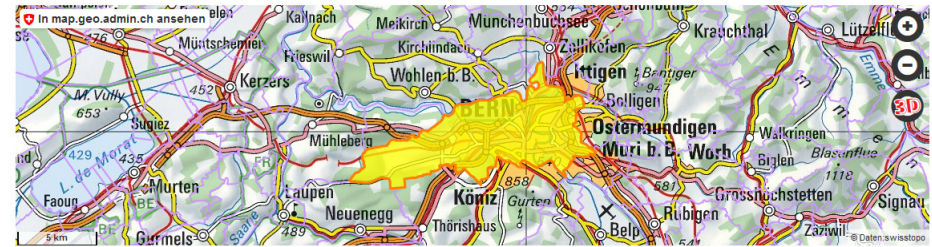
Federated SPARQL Queries

- > Verweis auf weitere Infos zur Gemeinde Bern bei Wikidata (Q70)

Bern

<https://ld.geo.admin.ch/boundaries/municipality/351>

a <http://www.opengis.net/ont/geosparql#Feature>, <http://www.geonames.org/ontology#ADM3>, <http://schema.org/AdministrativeArea>



type	Feature
type	A.ADM3
type	AdministrativeArea
hasVersion	351:2017
hasVersion	351:2016
hasVersion	351:2018
hasMap	?ch.swisstopo.swissboundaries3d-gemeinde-flaeche.fill=351
name	Bern
sameAs	351?geometryFormat=geojson
seeAlso	351
seeAlso	Q70
bfsNumber	351

Federated SPARQL Queries Hands-On

- > Gesucht:
 - Schweizer Gemeinden mit ihren geografischen Grenzen,
 - politischem Oberhaupt, dessen Parteizugehörigkeit und
 - das entsprechende Gemeindewappen
- > Quellen:
 - <https://ld.geo.admin.ch/sparql/>
 - <https://query.wikidata.org/>
- > Resultat: <https://tinyurl.com/y9o7bf2v>

Federated SPARQL Queries Hands-On

> Ausschnitt aus SPARQL Query auf <https://ld.geo.admin.ch/sparql>

```
SELECT ?wikiURI WHERE {  
  ?muni rdfs:seeAlso ?wikiURI;  
  {  
    SELECT ?wikiURI WHERE {  
      SERVICE <https://query.wikidata.org/bigdata/namespace/wdq/sparql> {  
        ?wikiURI wdt:P31 wd:Q70208;  
        wdt:P6 ?mayor;  
        wdt:P94 ?flagLink.  
      }  
    }  
  }  
}
```



Weitere (Federated) SPARQL Sample Queries

🔗 eCH0071 sample SPARQL queries

If you need help with additional queries please open an issue in this Github repository.

- [All \(not versioned\) municipality IDs in Switzerland](#)
- [All versioned municipalities](#)
- [Versioned municipalities that are still active](#)
- [Political municipalities that are still active](#)
- [Political municipalities that are still active, including admission date](#)
- [All events on 01.01.2015](#)
- [Details of all events on 01.01.2015](#)
- [How many municipalities in Switzerland changed canton?](#)
- [Federated Query: Get shapes of municipalities that stopped existing 2016](#)
- [Federated query: Show all municipality changes in 2016](#)

SPARQL lernen

- > https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial
- > DuCharme, Bob. Learning SPARQL: Querying and Updating with SPARQL 1.1, O'Reilly Media, Incorporated, 2013

