TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Forst-, Geo- und Hydrowissenschaften – Institut für Kartographie

Linked Data – an MRDB at web scale?

Stefan Hahmann and Dirk Burghardt

Zürich, 12.09.2010

Outline

1. Semantic Web
2. Linked Data
3. Linked Data vs. MRDB
4. Challenges
5. Ongoing Work

Zürich, 12.09.2010

1 Semantic Web

- 1989:
  - Tim Berners-Lee – „inventor" of the World Wide Web (WWW)
  - WWW initially developed for human consumption
- Late 1990's, Vision:
  - From „machine-readable" web of documents
  - To „machine-understandable" web of data
- 1999:
  - W3C
  - Resource Description Framework (RDF)

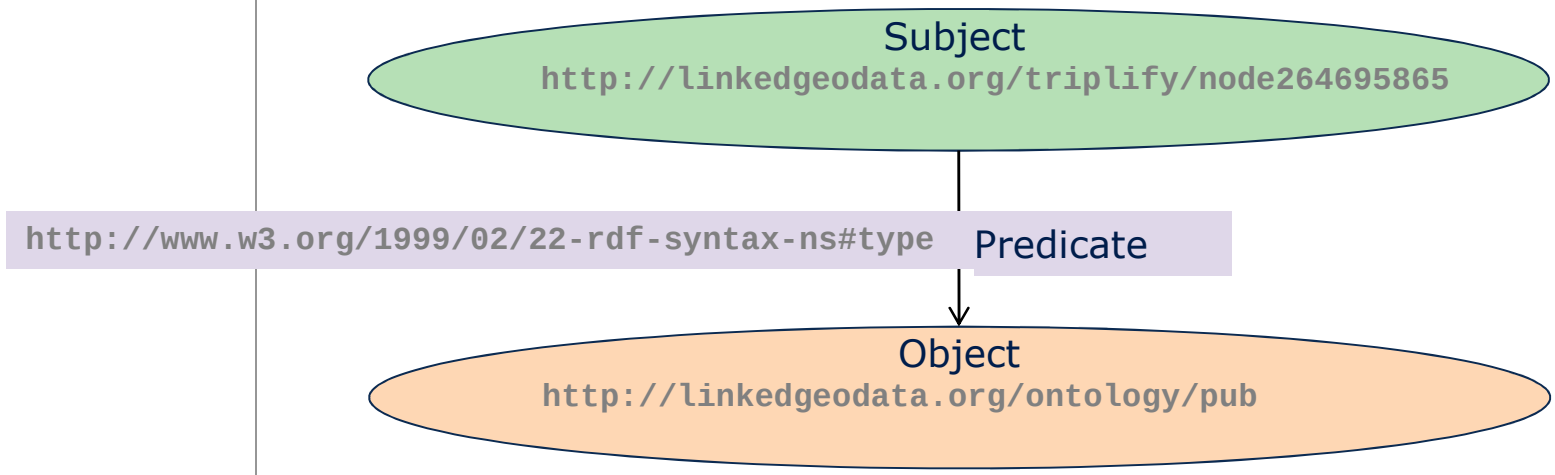
1 Semantic Web

- Resource Description Framework (RDF)
- Resource
  - arbitrary thing
  - identified by a URI
- e.g.:
  - LinkedGeoData
    - Project that publishes OpenStreetMap data in the Semantic Web using RDF
  - POI Café „B'Liebig" Dresden
  - URI: `http://linkedgeodata.org/triplify/node264695865`

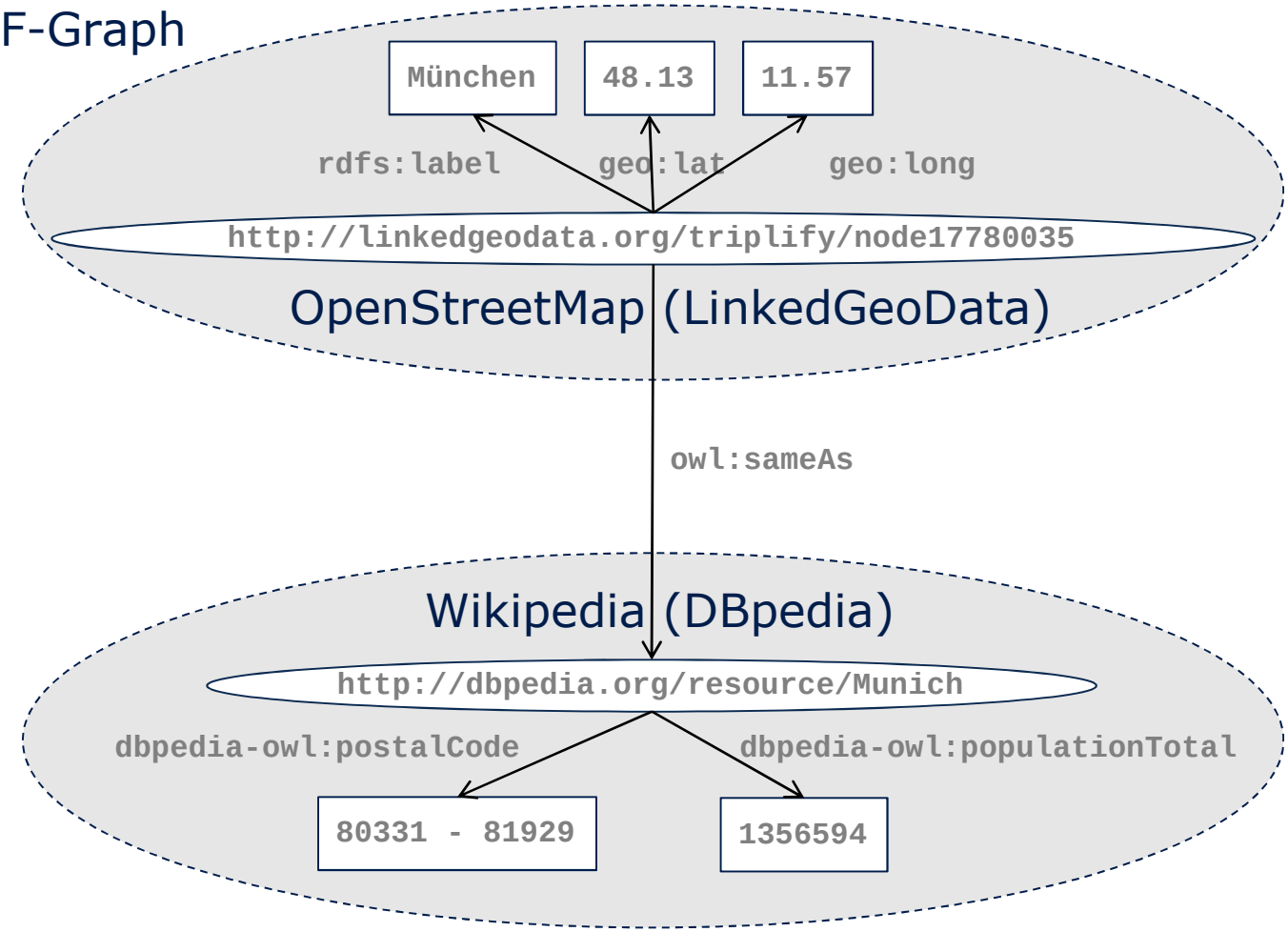
- Subject (Resource)
- Predicate (Named Property)
- Object (Resource or Literal)

RDF Triple

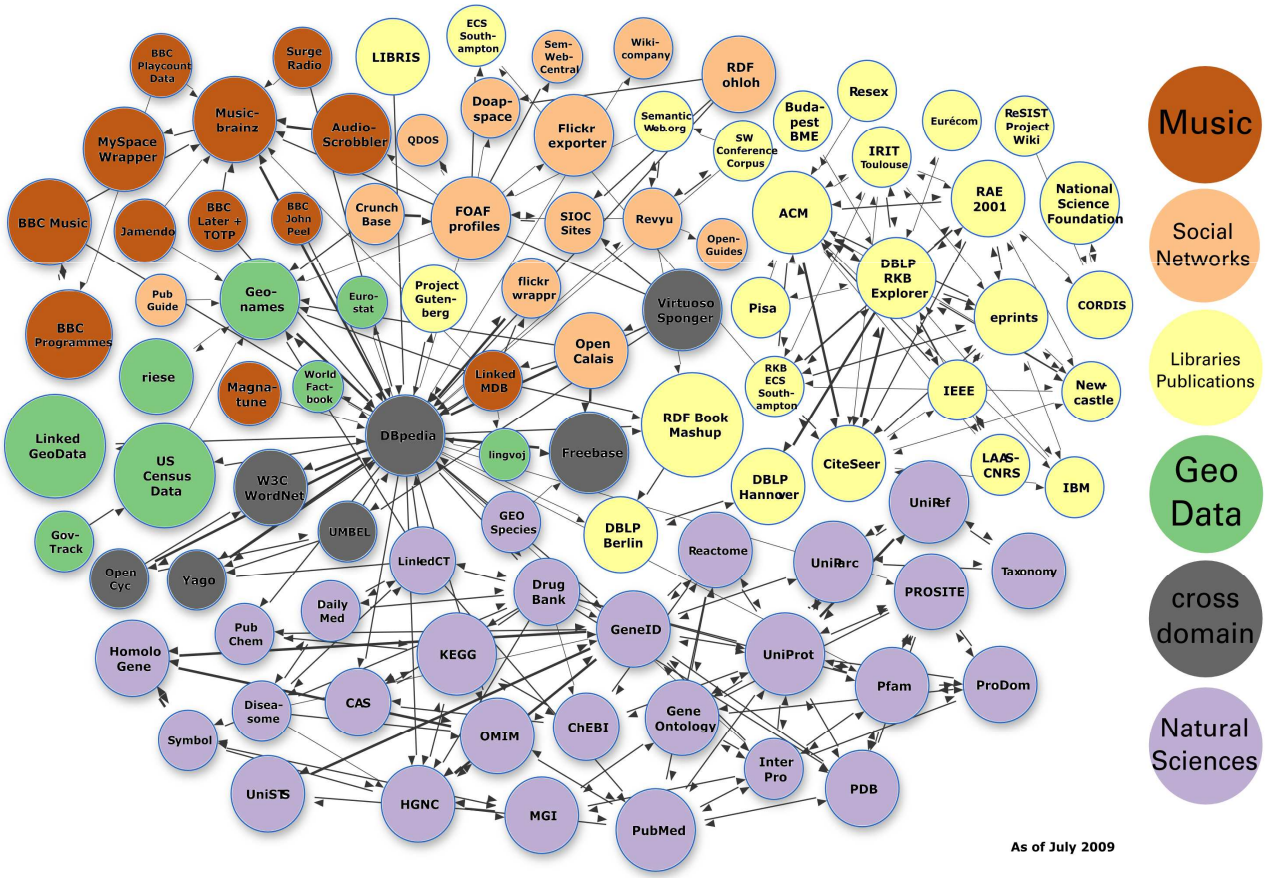
- E.g.: The LinkedGeoData Feature with ID=264695865 has the RDF type LinkedGeoData pub



- RDF-Graph



- Linked Open Data Cloud



TECHNISCHE UNIVERSITÄT DRESDEN

3 Linked Data vs. MRDB - Similarities

- a database (environment), which contains several representations of the same geographic entity (Sarjakoski 2007)
- different views of the same object are linked with each other (Sarjakoski 2007)
- geometry driven feature matching
- semantic matching: database schemas, RDF vocabularies, OWL ontologies

3 Linked Data vs. MRDB - Differences

MRDB	Linked Data
• database	• web
• focus on different geometric and semantic abstraction levels	• focus on different representations of the same entity: different (media) type and content of information
• Level of Detail strongly considered	• Level of Detail sparsely considered
• corresponding objects at different scales are explicitly linked	• marginal resolution dependent representation of geodata
• persistence and consistency can supervised by the producer	• persistence and consistency cannot be guaranteed by web links
• catalogs that contain verbal descriptions of attributes and class hierarchies, UML diagrams	• use of RDF (owl descriptions) for meta data
• corporate data	• web / distributed data
• authority-driven	• community-driven

4 Challenges

A map producer‘s point of view to Linked Data:

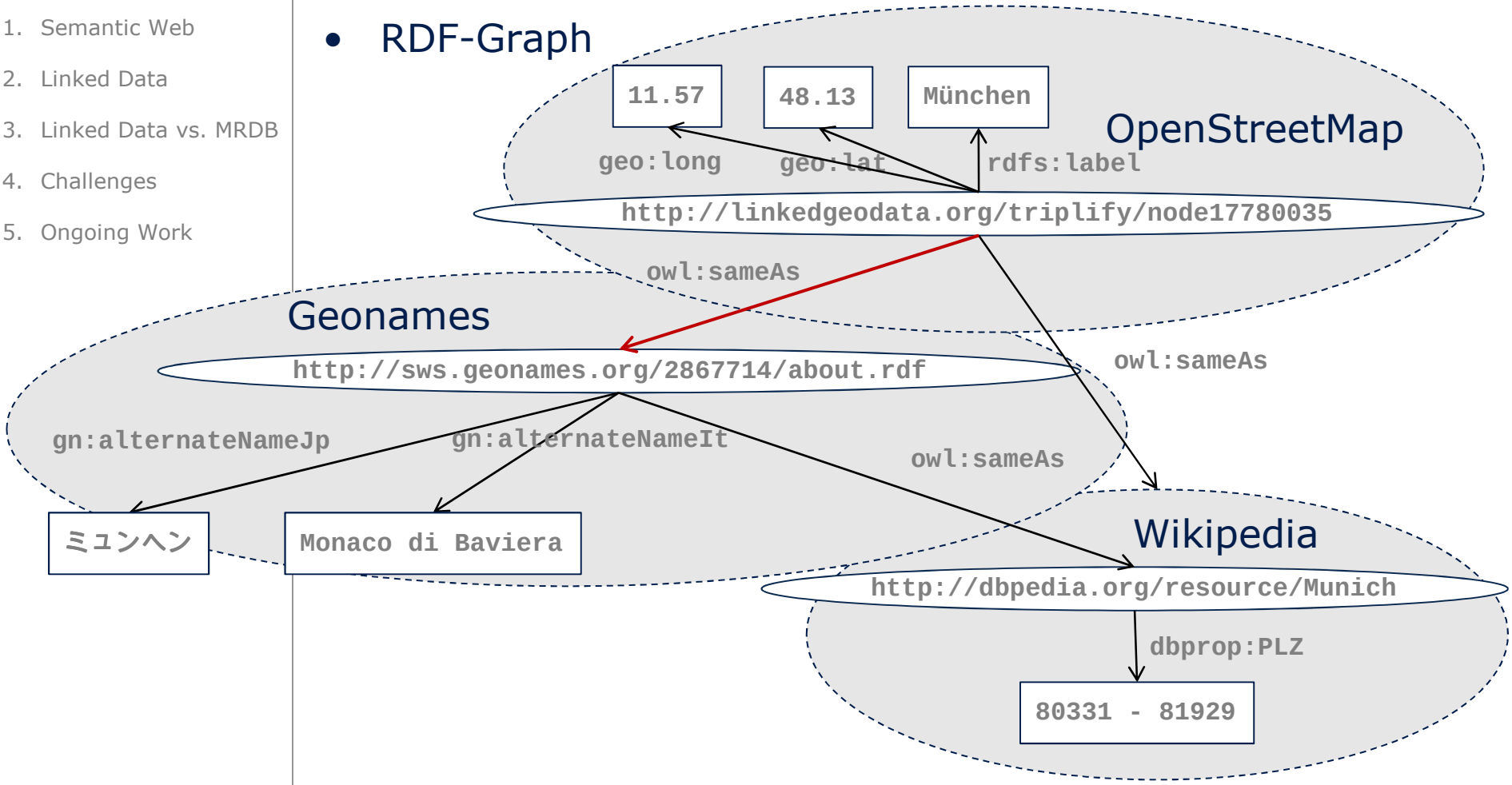
- Linked Data: a technology that can simplify access to geodata
- Advantage over Web Feature Service (WFS)?
  - RDF adds semantic descriptions
- However: Lack of standard semantic vocabularies
  - For both spatial and non-spatial terms
  - RDF is only a framework for describing semantics, not the semantics itself

4 Challenges

- Basic Geo Vocabulary (point, lat, long, alt)
  - http://www.w3.org/2003/01/geo/wgs84_pos
- W3C Geospatial Vocabulary
  - http://www.w3.org/2005/Incubator/geo/XGR-geo-20071023/
  - in discussion
  - uses a subset of gml classes and properties
- Geometric matching remains
- Semantic matching remains
  - As long as no standard (commonly accepted and used) semantic descriptions exist
  - Growing Motivation to use formalized semantic descriptions by using RDF?!

5 Ongoing Work

- RDF-Graph



„residential areas“ within 2 datasets
study area: Germany

LinkedGeoData / OpenStreetMap

Place	Sum
City	102
Town	2297
Village	36183
Suburb	7895
Hamlet	22349
Sum	**68826**

Geonames

Code	Feature description	Sum
P.PPL	populated place ... where people live and work	73990
P.PPLL	populated locality ... only with small group of buildings	2236
P.PPLX	section of populated place	2297
Sum		**78551**

5 Ongoing work

1. Semantics

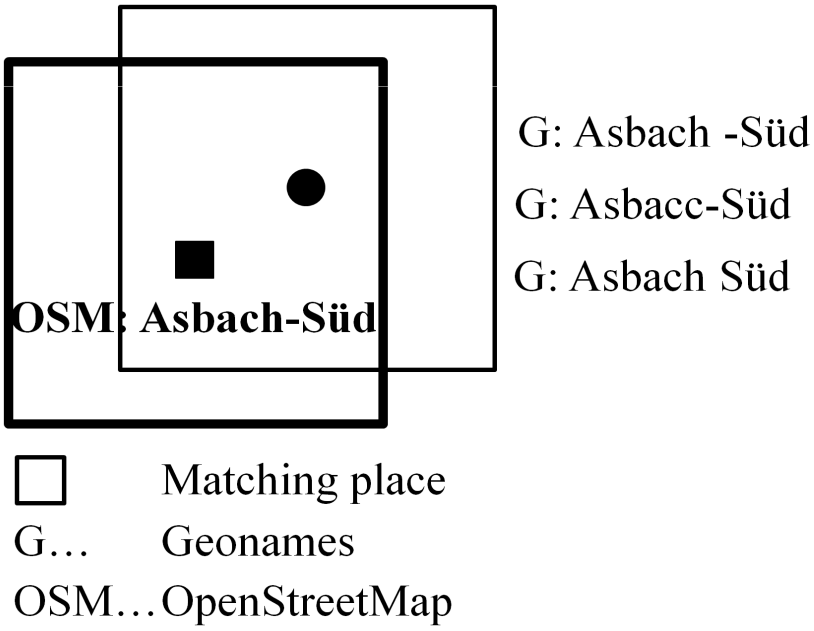
- Residential area type of both datasets
- City, Town, Village, Suburb, Hamlet (OSM Tags)
- PPL, PPLA, PPLL, PPLX (Geonames, featureCodes)

2. Geometry

- Buffer
- Bounding Box
- (0.05 x 0.05 deg.)
- Overlapping Boxes

3. Name Similarity

- Levenshtein Distance
- Threshold <= 1



5 Ongoig Work

Place	Matching percentage (%)
City	81
Town	84
Village	89
Suburb	73
Hamlet	49
Sum	74

Feature Code	Matching (%)
P.PPL	67
P.PPLA	100
P.PPLL	20
P.PPLX	53
Sum	65

5 Ongoing Work

- SPARQL (***S****PARQL* ***P****rotocol* ***a****nd* ***R****DF* ***Q****uery* ***L****anguage)*
  - Similar to SQL
  - Query RDF Graphs
  - Use results as input for thematic mapping
  - Performance - according cross domain queries?

- Validation of user generated data by comparing different data sets with each other
  - Consistency
  - Completeness
  - Syntactic correctness (toponym ambiguity)

Summary

Linked Data – an MRDB at web scale?

- Linked Data is more a webwide Geodatabase than a webwide MRDB
- Methods of MRDB are needed to produce resolution dependent Linked Data
- Our own experiments showed that
  - the task on geometric matching remains
  - there is a need for standard RDF descriptions for Cartography / Geography to enable semantic interoperability

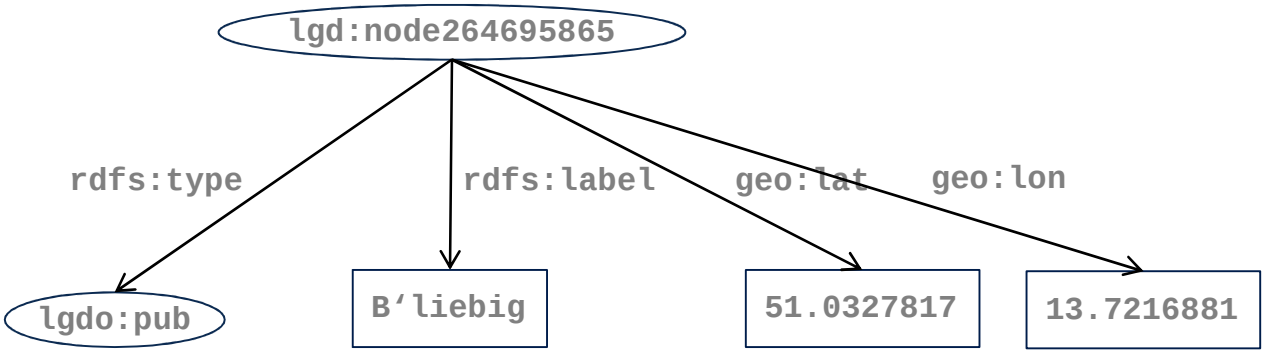
»Wissen schafft Brücken.«

Stefan.Hahmann@tu-dresden.de
Dirk.Burghardt@tu-dresden.de

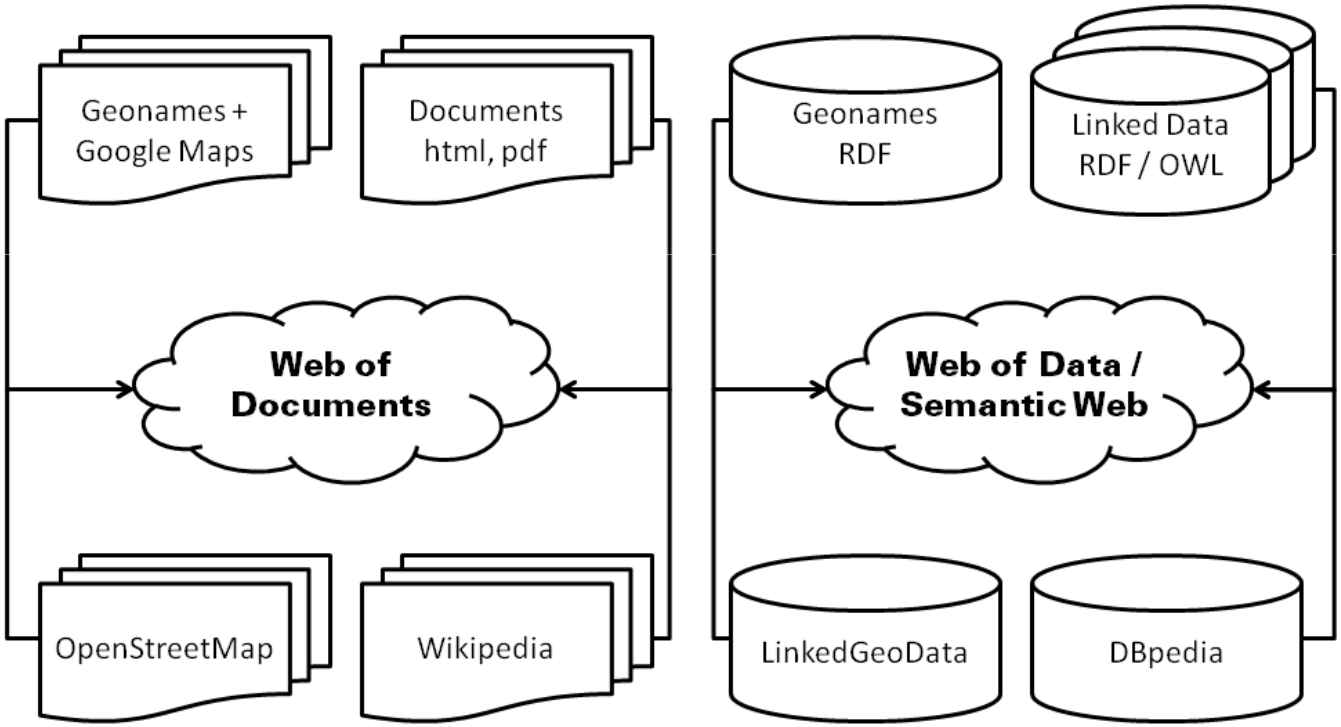
```
Prefix lgd: <http://linkedgeodata.org/triplify/>
Prefix lgdo:<http://linkedgeodata.org/ontology/>

Prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>
Prefix rdfs:<http://www.w3.org/2000/01/rdf-schema>

Prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos>
```



2 Linked Data



2 Linked Data

- Linked Data rules (according to Tim Berners-Lee):
  - (HTTP-) URI to identify things
  - Link data to other data
  - Make data accessible
  - Shared vocabularies!

5 Outlook

	MRDB	Linked Data
• Major Methods	• generalisation	• semantic web technology (RDF, OWL)
• Purpose	• high quality and effective map production • derive different type of maps from different representation levels	• access to spatial and nonspatial Information • cross domain data access (SPARQL)
• Resarch areas	• automated generalisation • updates • context modelling	• Standard semantic descriptions of geographic features (shared vocabularies) • semantic interoperability • self validating data

5 Outlook

- SPARQL (***S****PARQL* ***P****rotocol* ***a****nd* ***R****DF* ***Q****uery* ***L****anguage)*
  - Similar to SQL
  - Query RDF Graphs
  - Use results as input for thematic mapping
  - Performance - according cross domain queries?
- Validation of user generated data by comparing different data sets with each other
  - Consistency
  - Completeness
  - Syntactic correctness (toponym ambiguity)
- Validation of user generated data through logical reasoning
  - E.g. no streets with speed limit above 50km/h within cities