Towards the linking of geospatial government data

Exploring the use of Semantic Web tools for integration

Gabriella Wiersma
MSc. in Geomatics

Supervisors: Linda van den Brink, Jantien Stoter
Co-reader: Hugo Ledoux
Delegate: Willem Korthals Altes

TU Delft

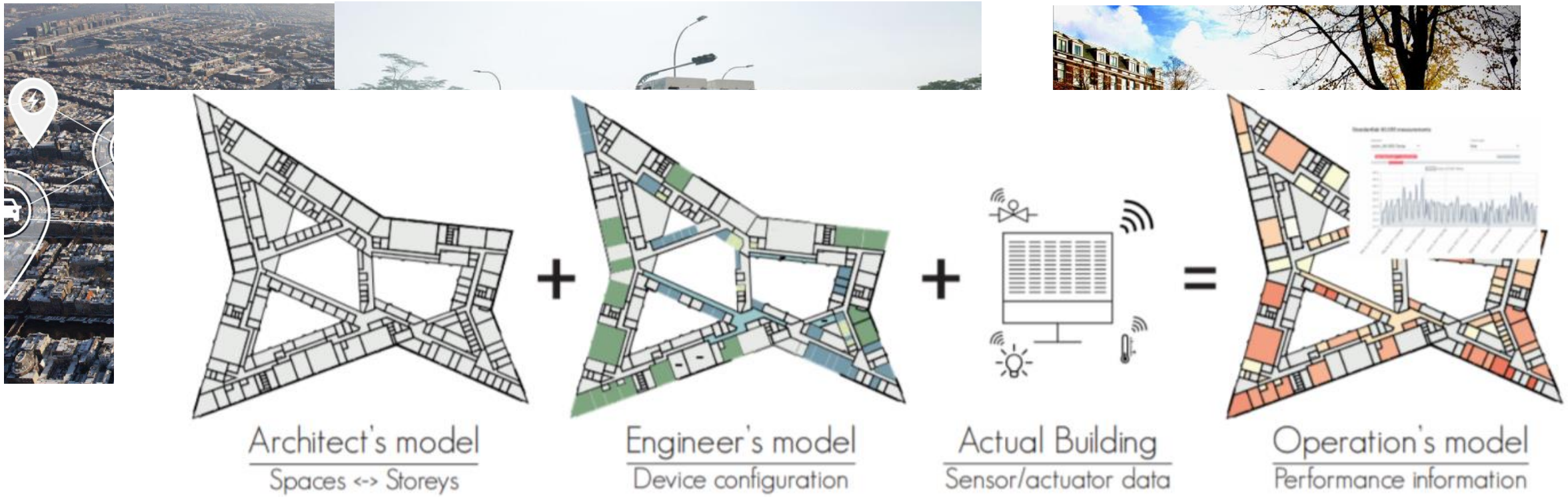
Content

1. Introduction
2. Problem statement and research questions
3. Methods:
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. Results & discussion
5. Conclusions

Content

1. **Introduction**
2. Problem statement and research questions
3. Methods:
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. Results & discussion
5. Conclusions

Introduction



http://www.semantic-web-journal.net/system/files/swj2224.pdf

Introduction

Semantic interoperability

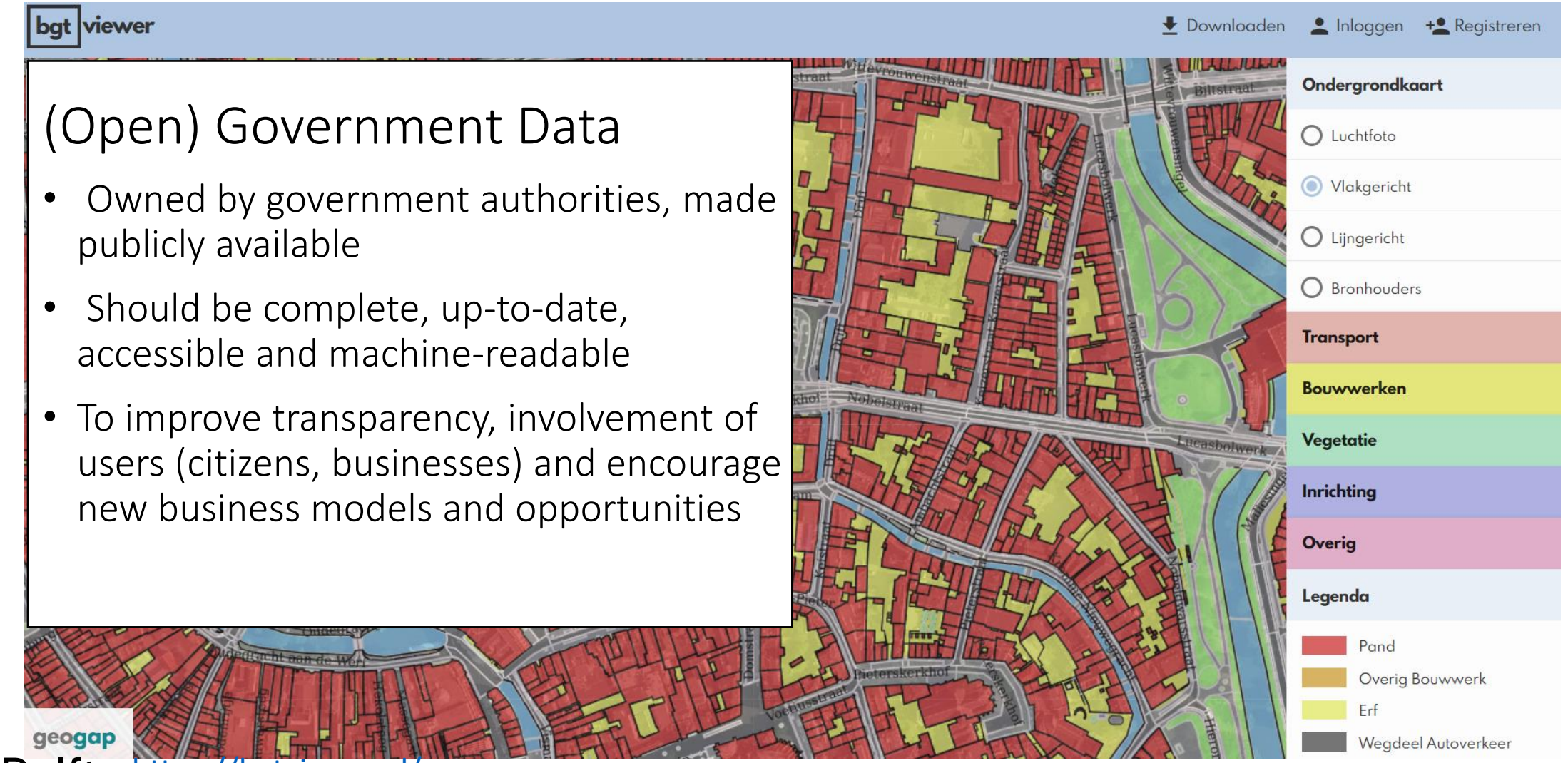
- Same domain modelled in different ways → Understanding differences that might occur due to other coverage/granularity/perspective

"Basically, we're all trying to say the same thing."

Introduction

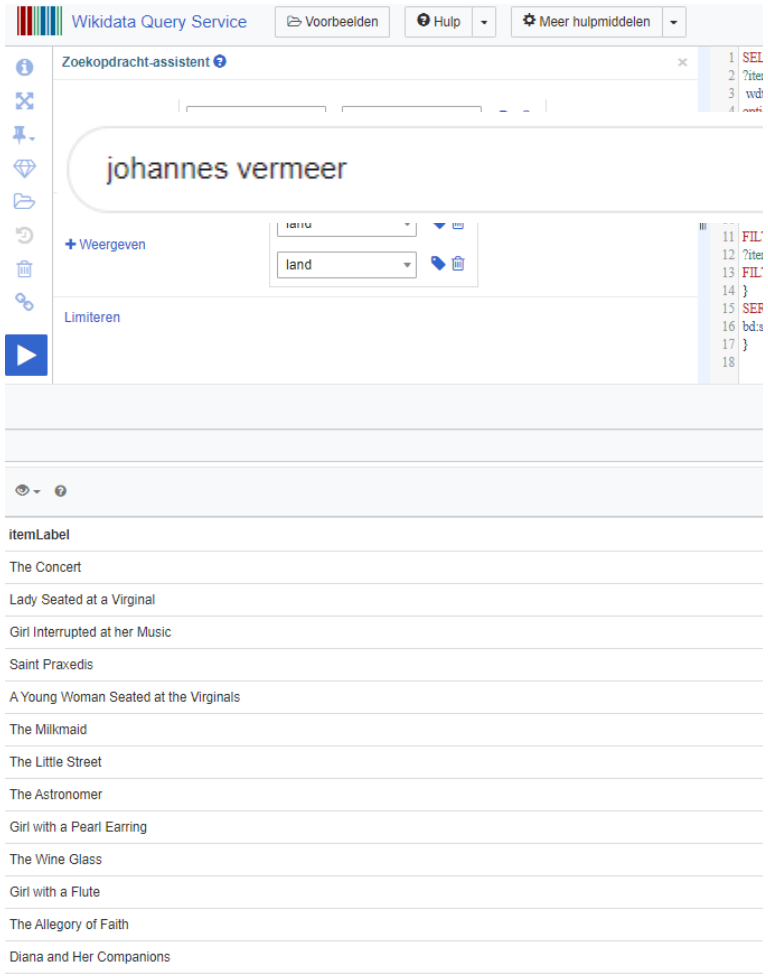
(Open) Government Data

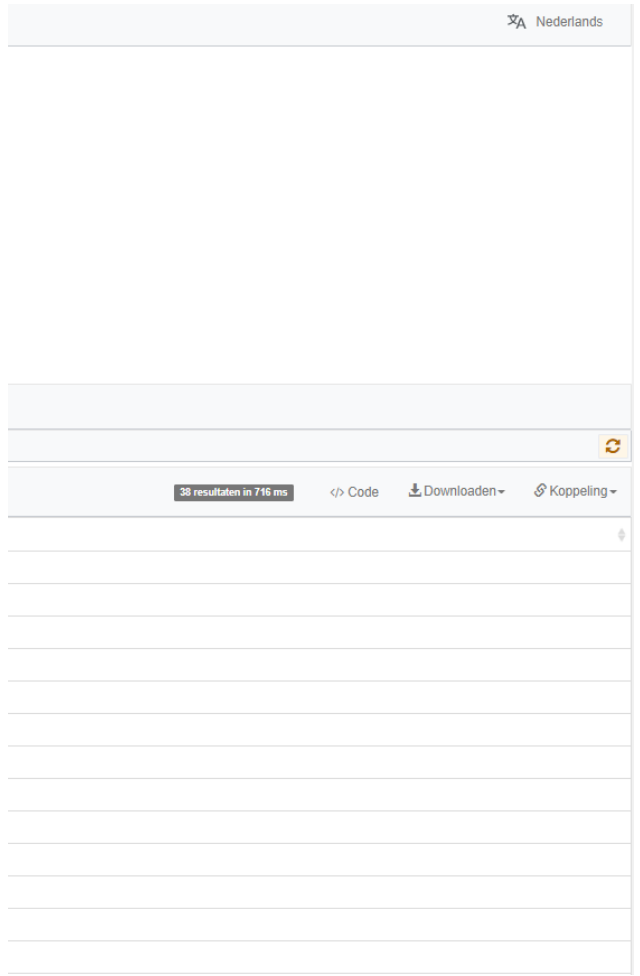
- Owned by government authorities, made publicly available
- Should be complete, up-to-date, accessible and machine-readable
- To improve transparency, involvement of users (citizens, businesses) and encourage new business models and opportunities



https://bgtviewer.nl/

Introduction



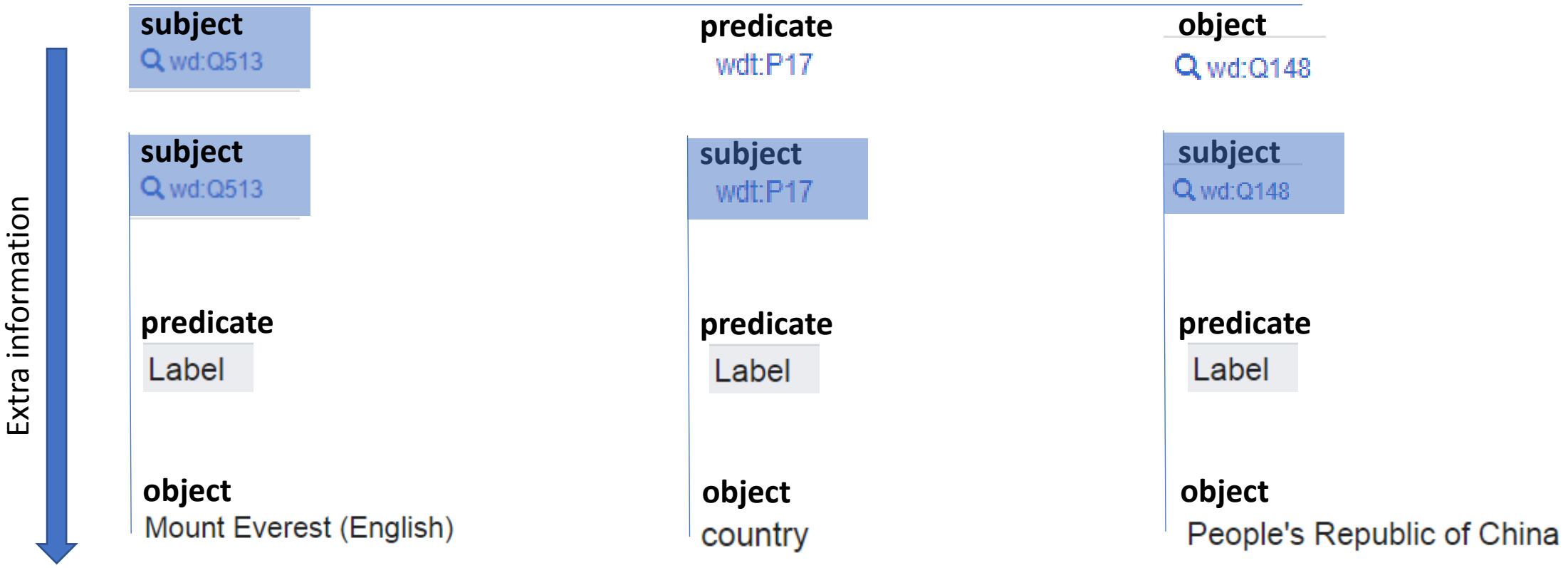


https://query.wikidata.org

Introduction

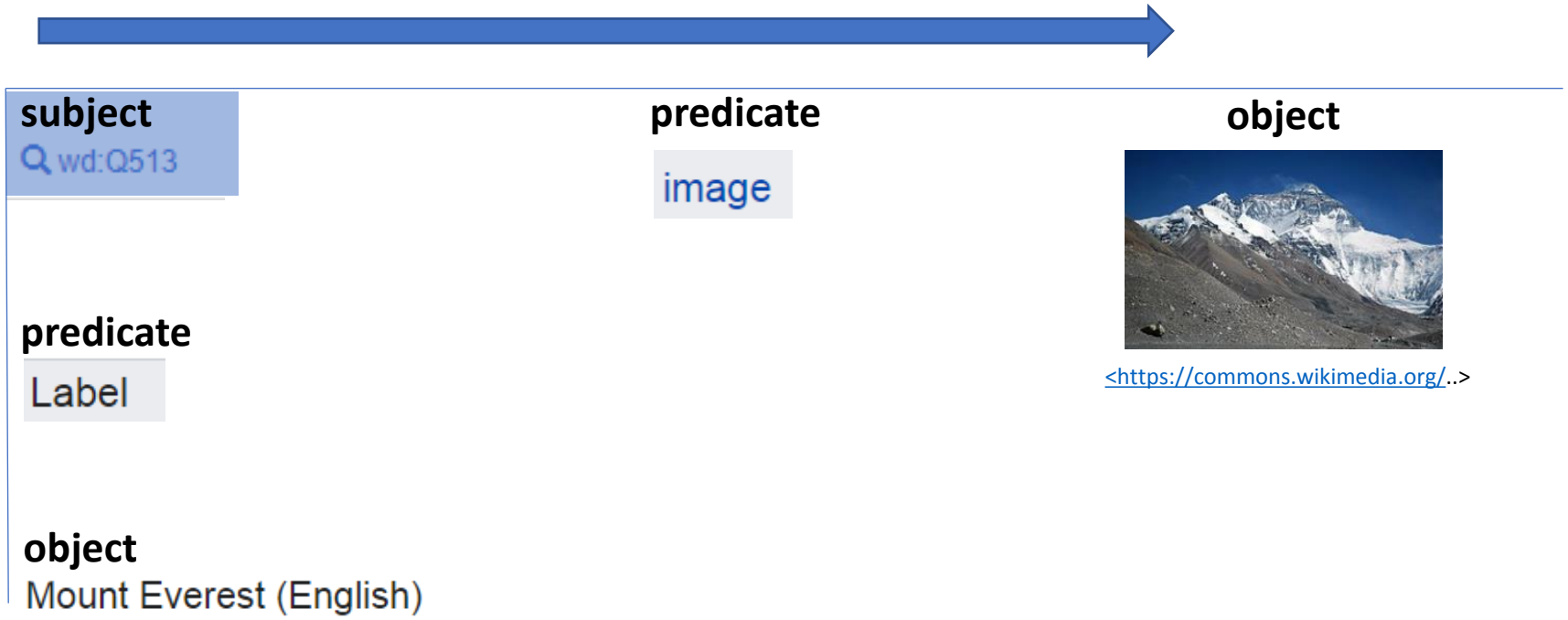
subject	predicate	object
wd:Q513	wdt:P17	wd:Q148

Introduction



Introduction

Link to external sources



Content

1. Introduction
2. **Problem statement and research questions**
3. Methods:
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. Results & discussion
5. Conclusions

Problem statement

What about geospatial data?

https://data.pdok.nl/sparql

subject

http://brk.basisregistraties.overheid.nl/id/perceel/370290171

predicate

http://brk.basisregistraties.overheid.nl/def/brk#plaatscoordinaten

object

Problem statement

Link to external sources

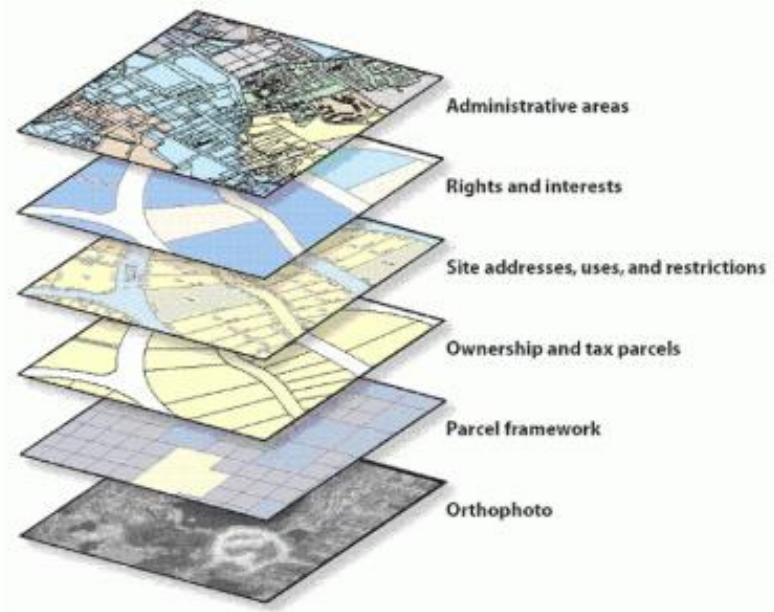
subject

http://brk.basisregistraties.overheid.nl/id/perceel/370290171

predicate

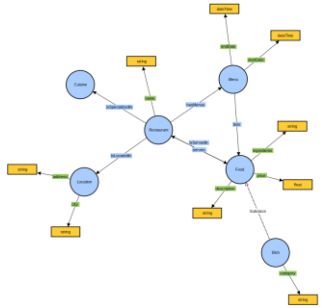
http://brk.basisregistraties.overheid.nl/def/brk#plaatscoordinaten

object

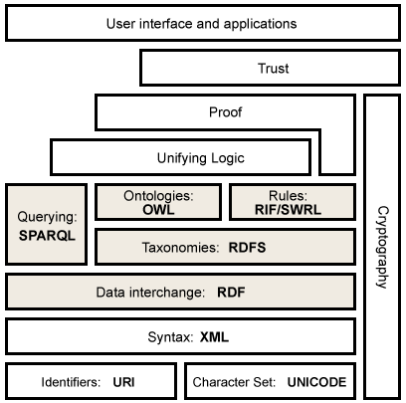


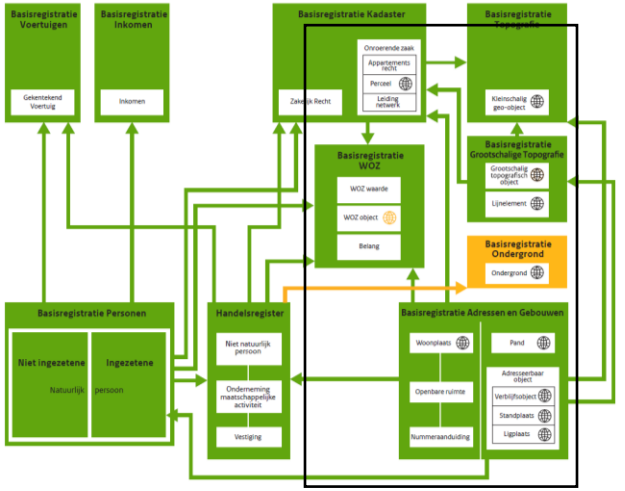
https://geospatialhistorian.wordpress.com/lessons/arcgis-lesson-5-overlay-analysis/

Research question



To what extent can *ontology-based solutions* using *semantic web technologies*, contribute to integration and use of data from *geospatial registries*?





Research question

To what extent can *ontology-based solutions* using *semantic web technologies*, contribute to integration and use of data from *geospatial registries*?

- What type of ontology-based techniques are best suited in the case of integrating data from the geo-registries?
- How does the overlap and differences between the data/models affect the correspondences between the data sources?
- What is the added value of having custom semantic relations incorporated into ontolo-gies?

Content

1. Introduction
2. Problem statement and research questions
3. **Methods:**
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. Results & discussion
5. Conclusions

Content

1. Introduction
2. Problem statement and research questions
3. **Methods:**
   - **Literature: Semantic Web and (Geo) semantics**
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. Results & discussion
5. Conclusions

Literature

Formal ontologies:

- Based on RDF data model (subject – predicate – object)

<Lassie> <rdf:type> <Dog>

- Use RDF-S (RDF Schema) to represent hierarchical relations

<Dog> <rdfs:subclassOf> <Mammal>

- Different OWL profiles for more expressive relations (to infer new knowledge)

<hasOwner> <owl:inverseOf> <isOwnerOf>
<Lassie> <hasOwner> <Jeff Miller>

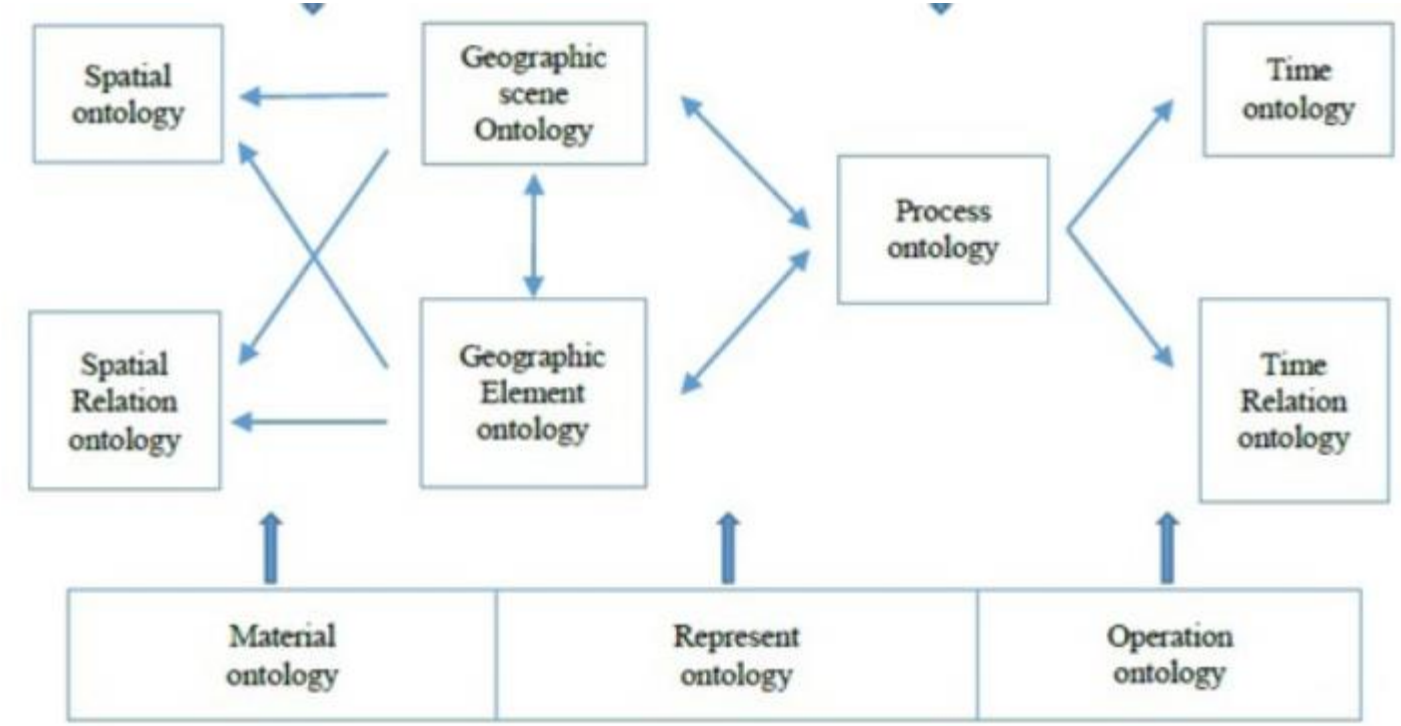
→ *Who is the owner of Lassie?*

<Jeff Miller <isOwnerOf> <Lassie>

Literature

Geosemantics + OWL:

- Formalizing concepts related to geospatial applications

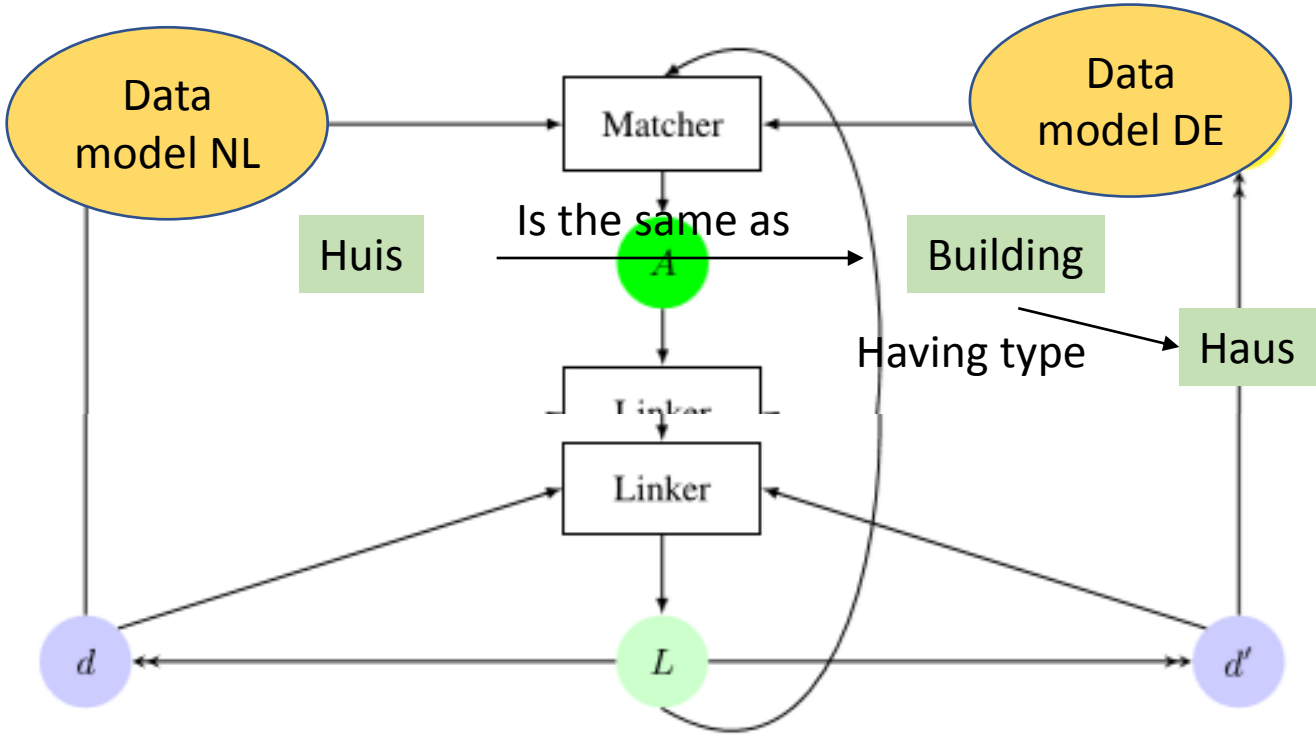


https://www.degruyter.com/view/journals/geo/10/1/article-p782.xml?language=en

Literature

Geosemantics (linked data integration):

- Ontology alignment:to understand relation between concepts from different sources
- Data interlinking: to determine how data is connected
- Geospatial data: links (L) already available through location matching



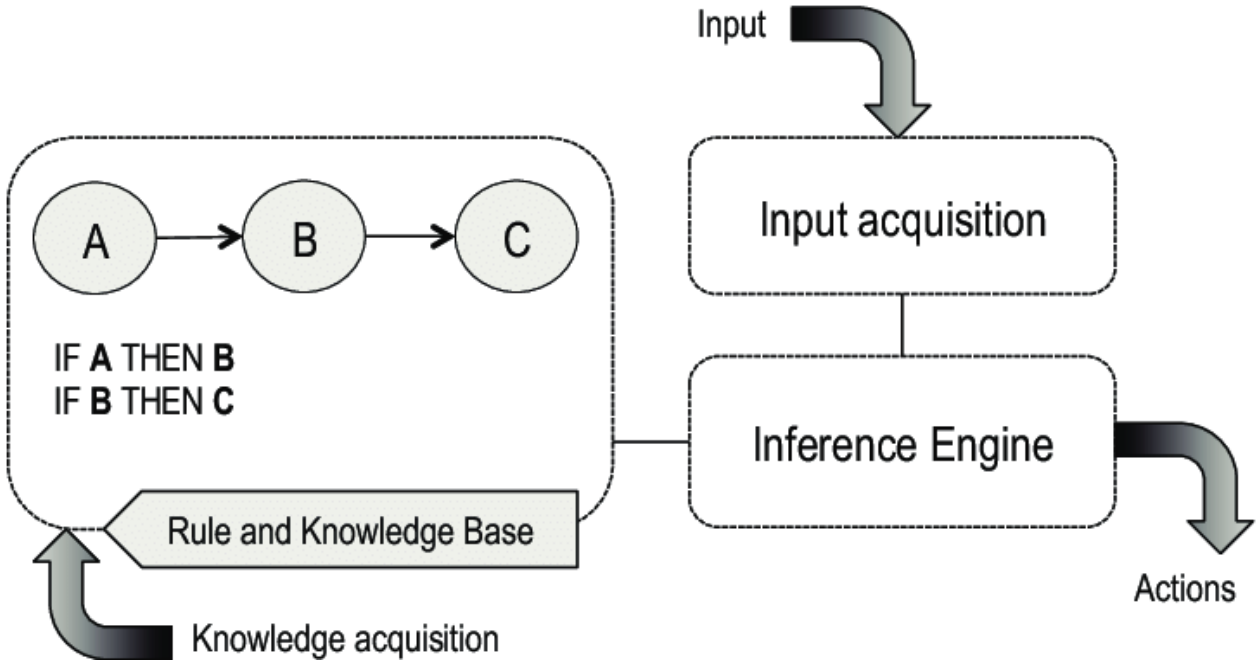
Literature

Geosemantics + rules:

- Geospatial (L) ‘overlap’ links might not suffice → **rule-based reasoning** = meaningful spatial links

- Semantic Web Rule Language (SWRL)

- Shapes Constraint Language (SHACL)
  https://www.w3.org/TR/shacl-af



https://www.researchgate.net/publication/234551921_Survey_of_Context_Provisioning_Middleware

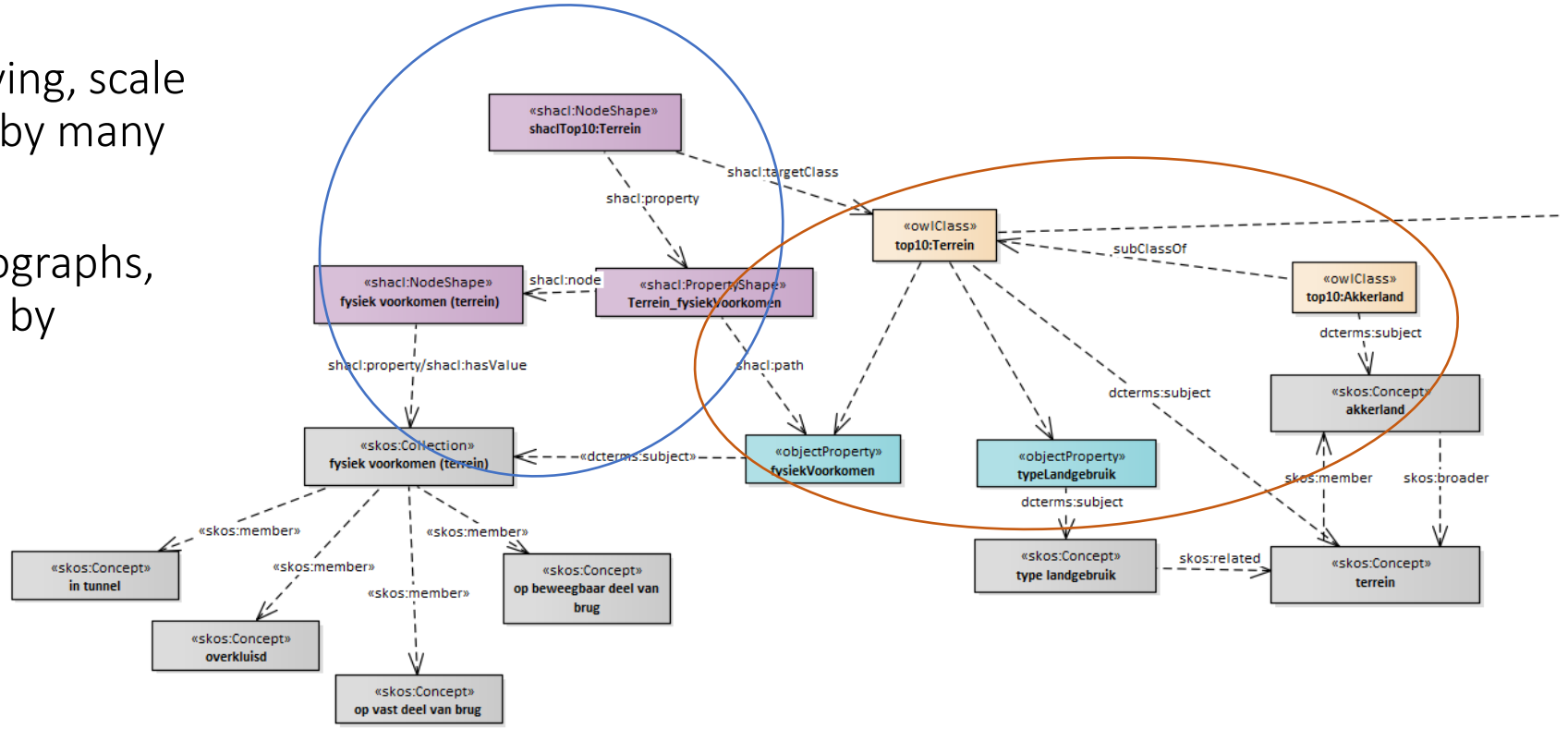
Content

1. Introduction
2. Problem statement and research questions
3. **Methods:**
   - Literature: Semantic Web and (Geo) semantics
   - **Case study: Dutch geo-registries**
   - Framework: alignment and query
4. Results & discussion
5. Conclusions

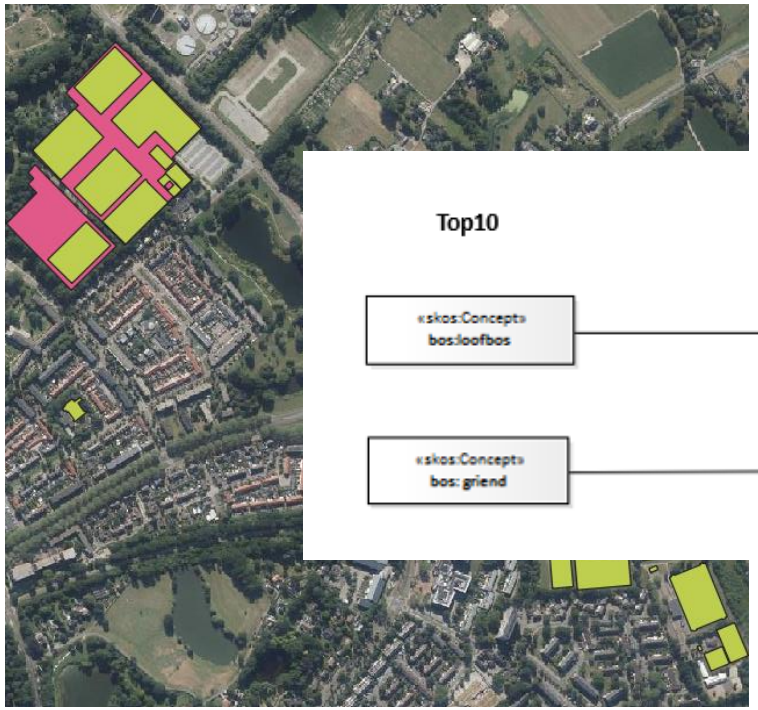
Case study

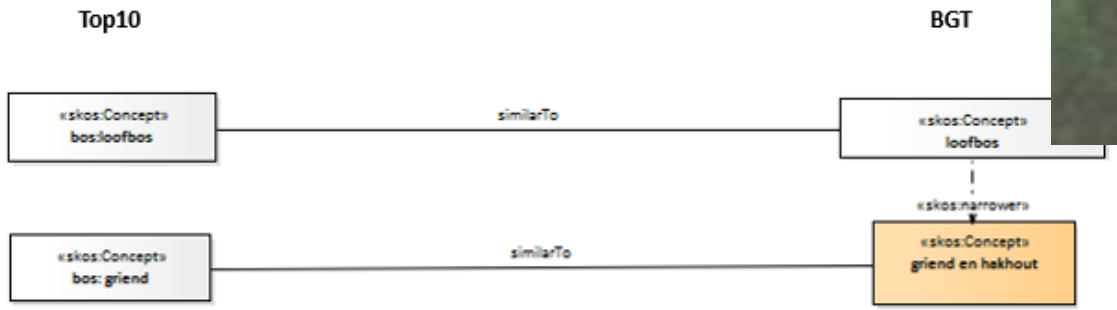
Datasets:

- BGT → terrestrial surveying, scale 1:500-1:5000, collected by many source holders
- TOP10NL → aerial photographs, scale 1:10000, managed by Kadaster



Case study





Vagueness due to incomplete representation
Optional attributes (BGT 'plus-type')

Vagueness due to dynamic aspects of phenomena:
BGT waterobjects in pink, their 'fuzzy boundaries' represented in orange

Phenomena represented by crisp concepts:
TOP10NL 'sportterrein, sportcomplex' in pink, BGT 'sportterrein' in green

Content

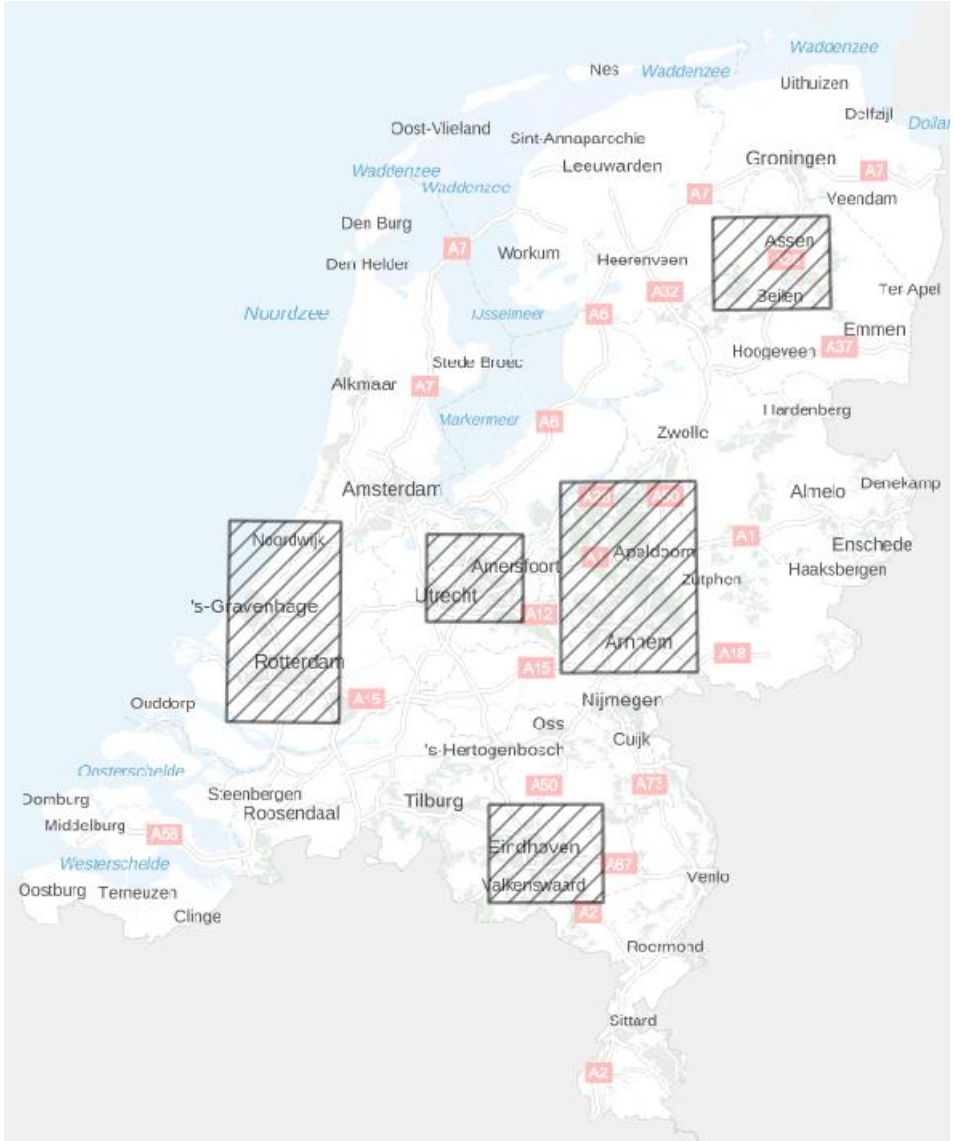
1. Introduction
2. Problem statement and research questions
3. **Methods:**
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - **Framework: alignment and query**
4. Results & discussion
5. Conclusions

Framework

Datasets/tools:

- For alignment:
  - Get data instances for selected classes in selected study areas (POSTGIS Dump → shapefiles)
  - TOP10NL ontology from PDOK
  - BGT ontology draft (adapted)
  - Link both (TOP10NL/BGT) to NEN3610-LD

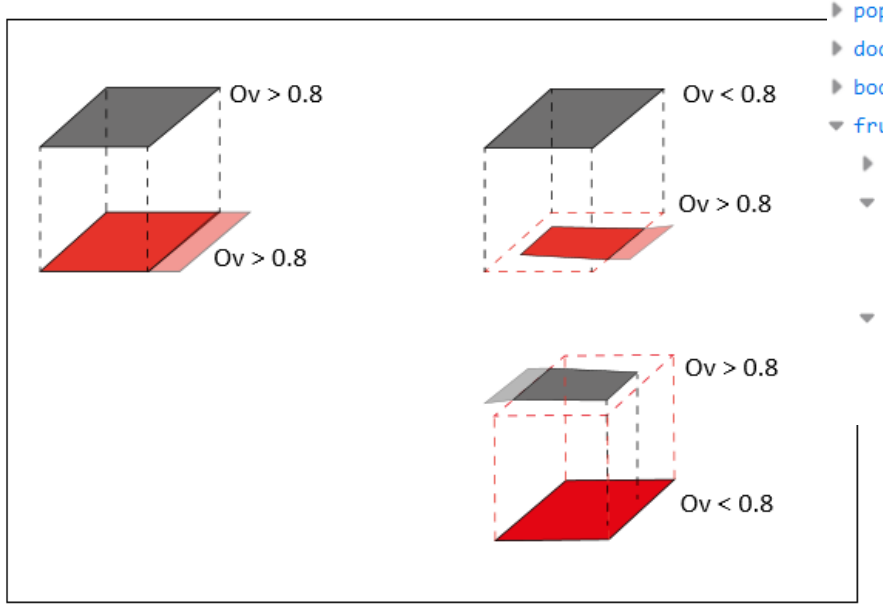
- Querying:
  - Apache Jena API
  - Invoke TOP10NL data via SPARQL <SERVICE>
  - Transform BGT GML from PDOK to RDF data, store in TDB

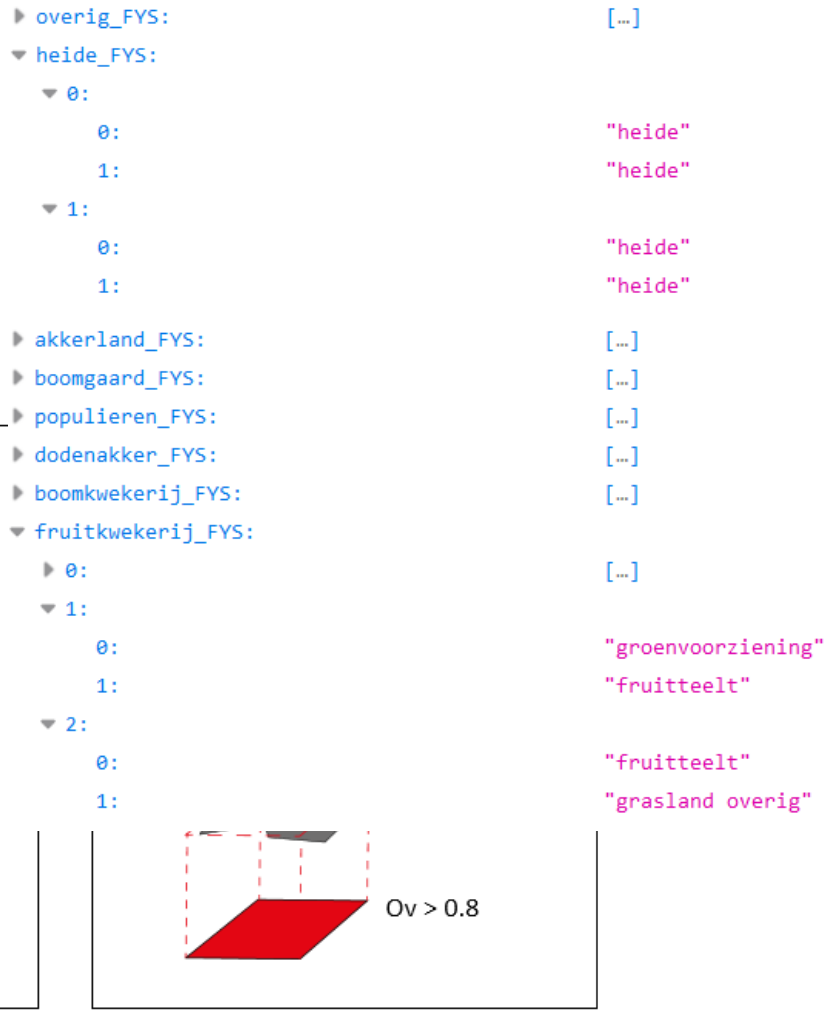


Framework

Alignment between object types:

- Select matchers: string-based and structural + extensional (leveraging the geospatial link between instances)

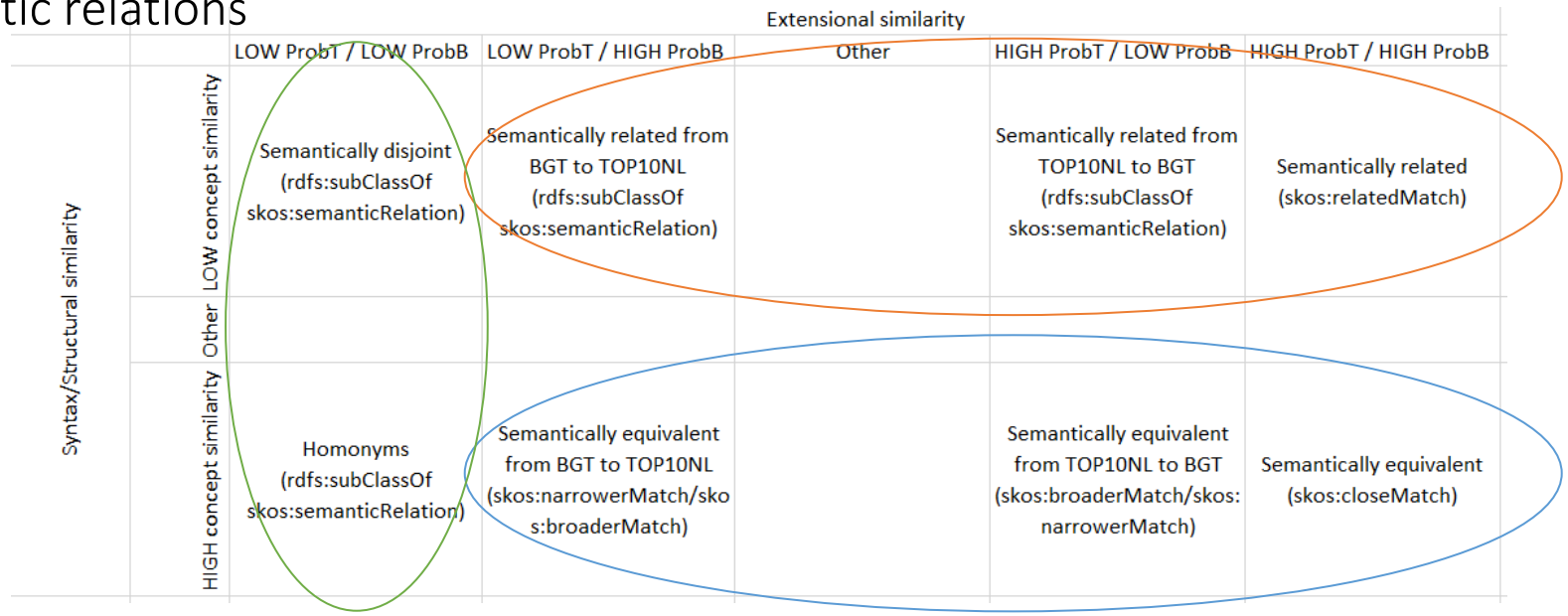




Framework

Alignment between object types:

- Select matchers: string-based and structural + extensional (leveraging the geospatial link between instances) = semantic relations

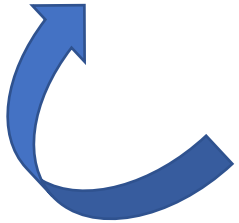


Framework

Querying

- Federated query to process geospatial links

```
maps:relation_obj1  a       maps:OverlapRelations ;
        maps:mainfeature    <http://brt.basisregistraties.overheid.nl/top10nl/id/terrein/130511636> ;
        maps:maintype       top10:Grasland , top10:Terrein ;
        maps:hasOverlap     [ a                         bgt:Zand ;
                              maps:mainRelativeArea     "9.2"^^<http://www.w3.org/2001/XMLSchema#float> ;
                              maps:secRelativeArea      "100.0"^^<http://www.w3.org/2001/XMLSchema#float> ;
                              maps:secfeature           <http://bgt.basisregistraties.overheid.nl/bgt/id/G0150.2a6ff73776bb4d5c
                            ] ;
```



```
BIND (IF((?count = 1 && ?mainarea >= 80 && ?secarea >= 80), align:equals,
IF((?count = 1 && ?mainarea >= 80 && ?secarea < 80), align:is_contained,
IF((?count = 1 && ?mainarea < 80 && ?secarea >= 80), align:contains,
IF((?count > 1 && (?sum_main > 80 || ?blue > 70)), align:aggregates,
" "
```

Framework

Querying

- Are there any possible inconsistencies in the data (compare if the concepts used in both sources are not disjoint)?

- Return all objects from dataset 1 having type X (and consider similar objects from other datasets).

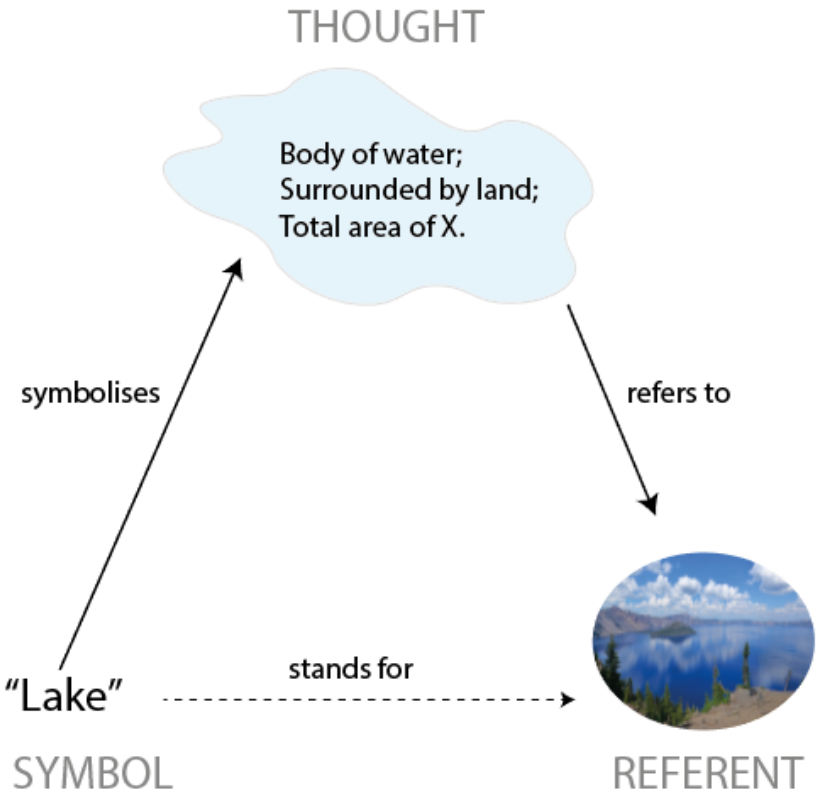
Content

1. Introduction
2. Problem statement and research questions
3. Methods:
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. **Results & discussion**
5. Conclusions

Results & Discussion

Matching

- String/Structural:
  - High similarity scores reliable; but many words found similar (during manual inspection) do not score high syntactically
  - not appropriate if there are many small syntactic variances (duin vs puin)
  - Context might be too important in some cases (hoogspannings vs laagspanningsmast)
  - Using structure: simply looking at shared NEN31610 class might be more useful than graph-based approach



Results & Discussion

Matching

- Extensional:
  - More geospatial links than expected for some objects (Terrein), less for others (Functioneel Gebied)
  - Possibly different use of concepts by source holders: Naaldbos in BGT/TOP10NL only related in three study areas
  - Focus was only on polygons, with few thresholds used determine relation types between objects – linestrings/points require other methods

Results & Discussion

Querying

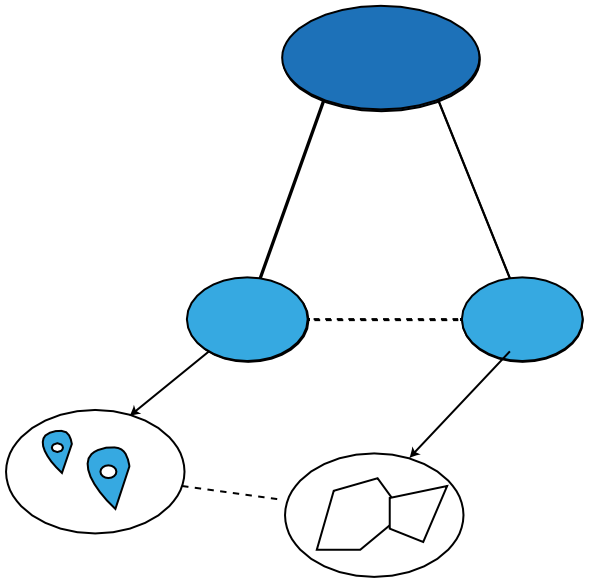
- TOP10NL Terrein -> many ‘disjoint’ relations.
- No SHACL constraint warning fired during queries, however:
  - Few disjoint relations (object types that are rarely spatially linked);

Content

1. Introduction
2. Problem statement and research questions
3. Methods:
   - Literature: Semantic Web and (Geo) semantics
   - Case study: Dutch geo-registries
   - Framework: alignment and query
4. Results & discussion
5. **Conclusions**

Conclusions

Q1: What type of ontology-based techniques are best suited in the case of integrating data from the geo-registries?



VS (SPARQL)Rules

Conclusions

Q2: How does the overlap and differences between the data/models affect the correspondences between the data sources?

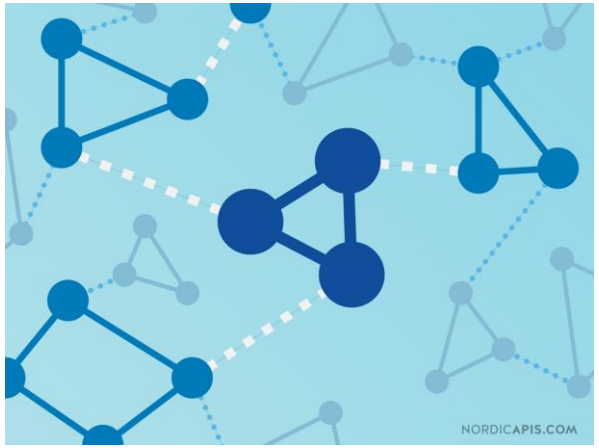
Scale ≠ Granularity

Different approaches to geometric representation

(Possibly) inconsistent interpretations

Conclusions

Q3: What is the added value of having custom semantic relations incorporated into ontologies?



Thank you for the attention