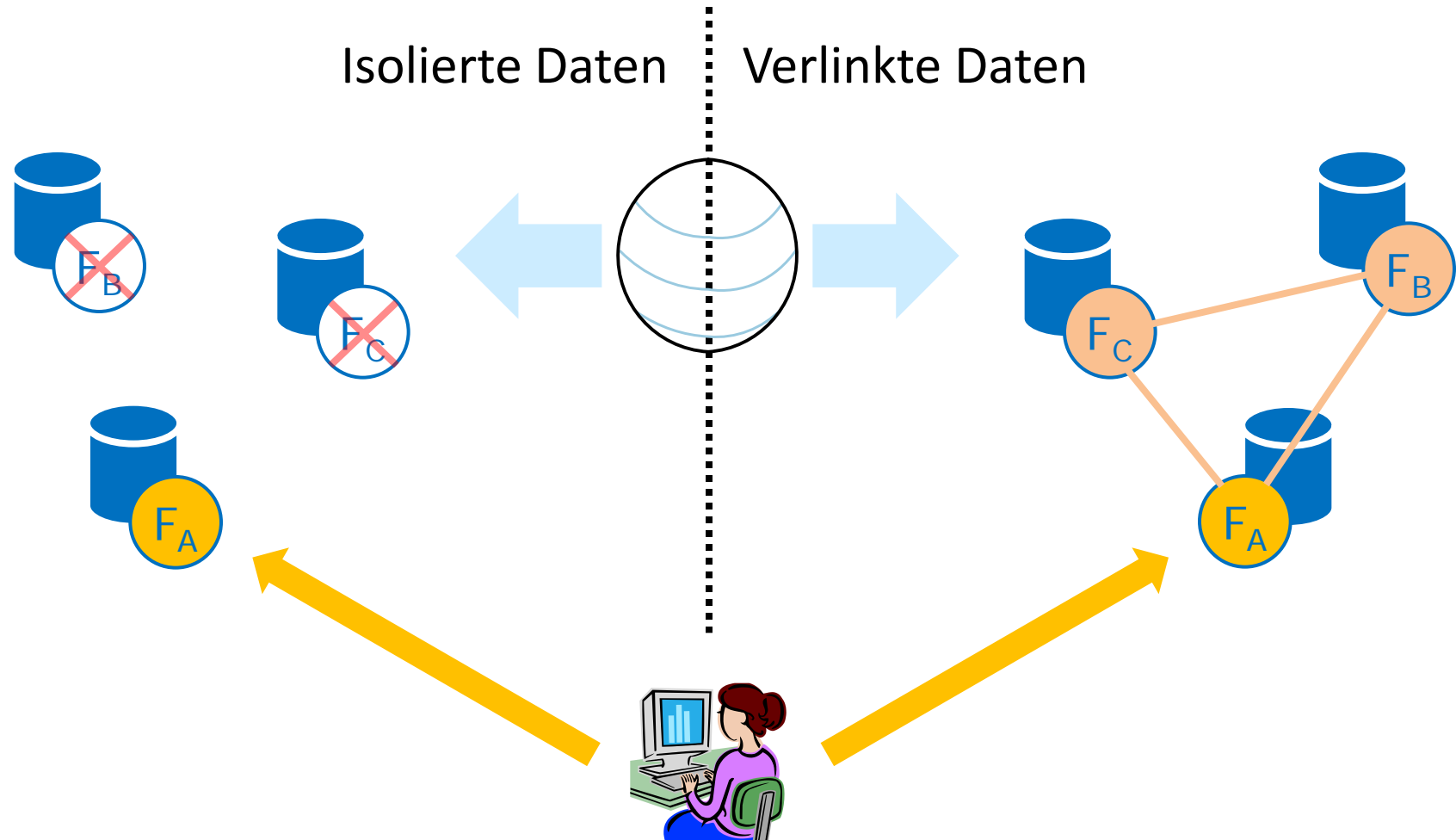


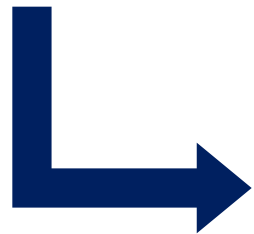
Geodatenintegration unter Nutzung von Linked Data Prinzipien

**Stefan Wiemann
Technische Universität Dresden**

Motivation



- Web of documents
 - Dokumentorientiert
 - Verlinkung über Hyperlinks



- Web of data (Semantic Web)
 - Datenorientiert
 - Gemeinsame Ontologien & Vokabulare
 - Nutzung des Linked Data Paradigma

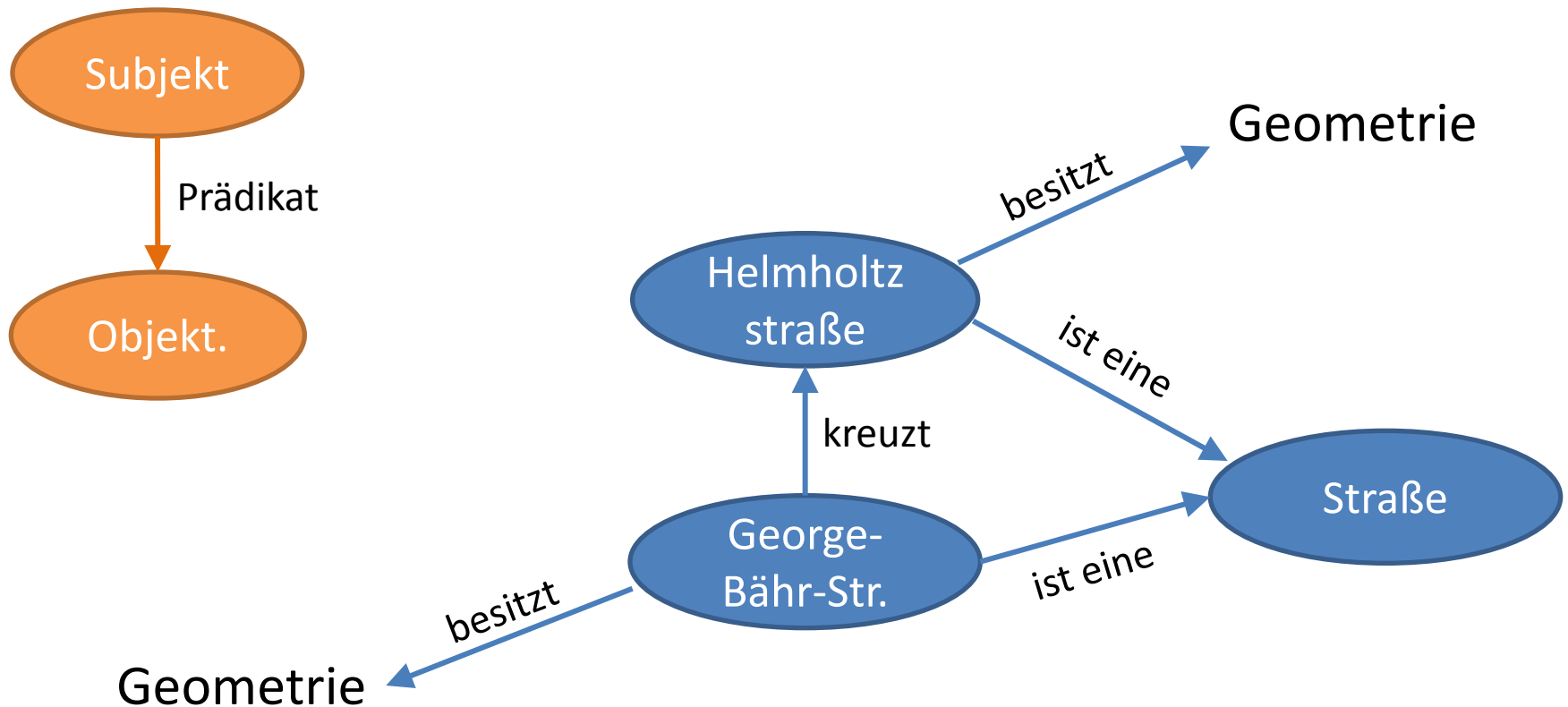
- Grundlegende Prinzipien
 - Nutzung von URIs (unique resource identifier) zur Identifikation von Objekten
 - Nutzung von HTTP URIs, um einen standardisierten Zugriff auf das Objekt zu ermöglichen
 - Bereitstellung aussagekräftiger Informationen über das Objekt (kodiert in RDF)
 - Einbettung weiterführender Links

Linked Data (3)

- ★ Daten verfügbar im Internet, möglichst lizenziert als Open Data
- ★★ maschinenlesbare Struktur der Daten
- ★★★ Daten kodiert in einem nicht-proprietären Datenformat
- ★★★★ Nutzung offener Standards des W3C (RDF & SPARQL) zur Identifikation von Objekten, damit andere darauf verweisen können
- ★★★★★ Links zu anderen Daten, um weiterführende Informationen zu erhalten

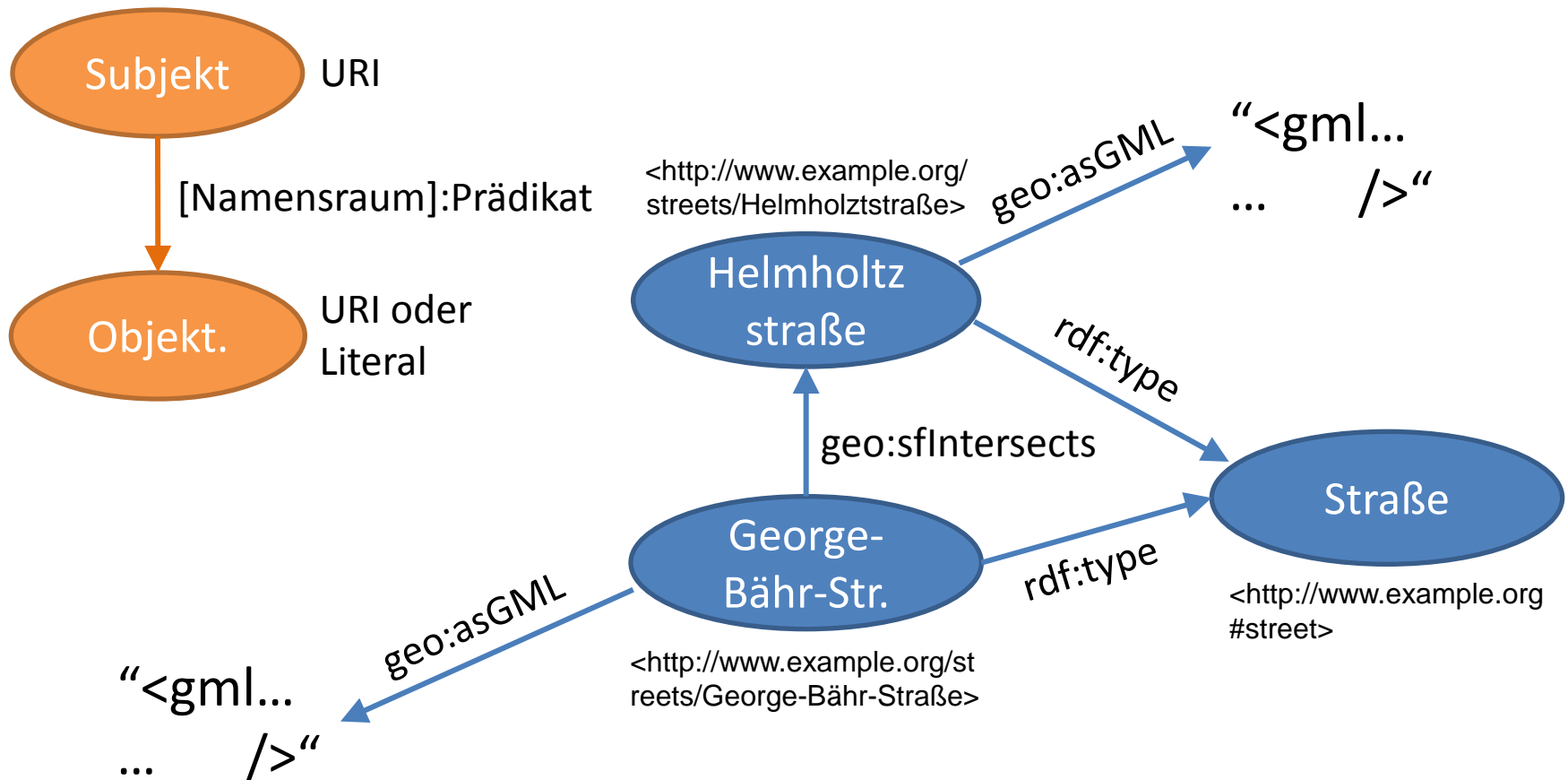
Quelle: Berners-Lee 2006, <http://www.w3.org/DesignIssues/LinkedData.html>

- Beschreibung von Daten über Tripel

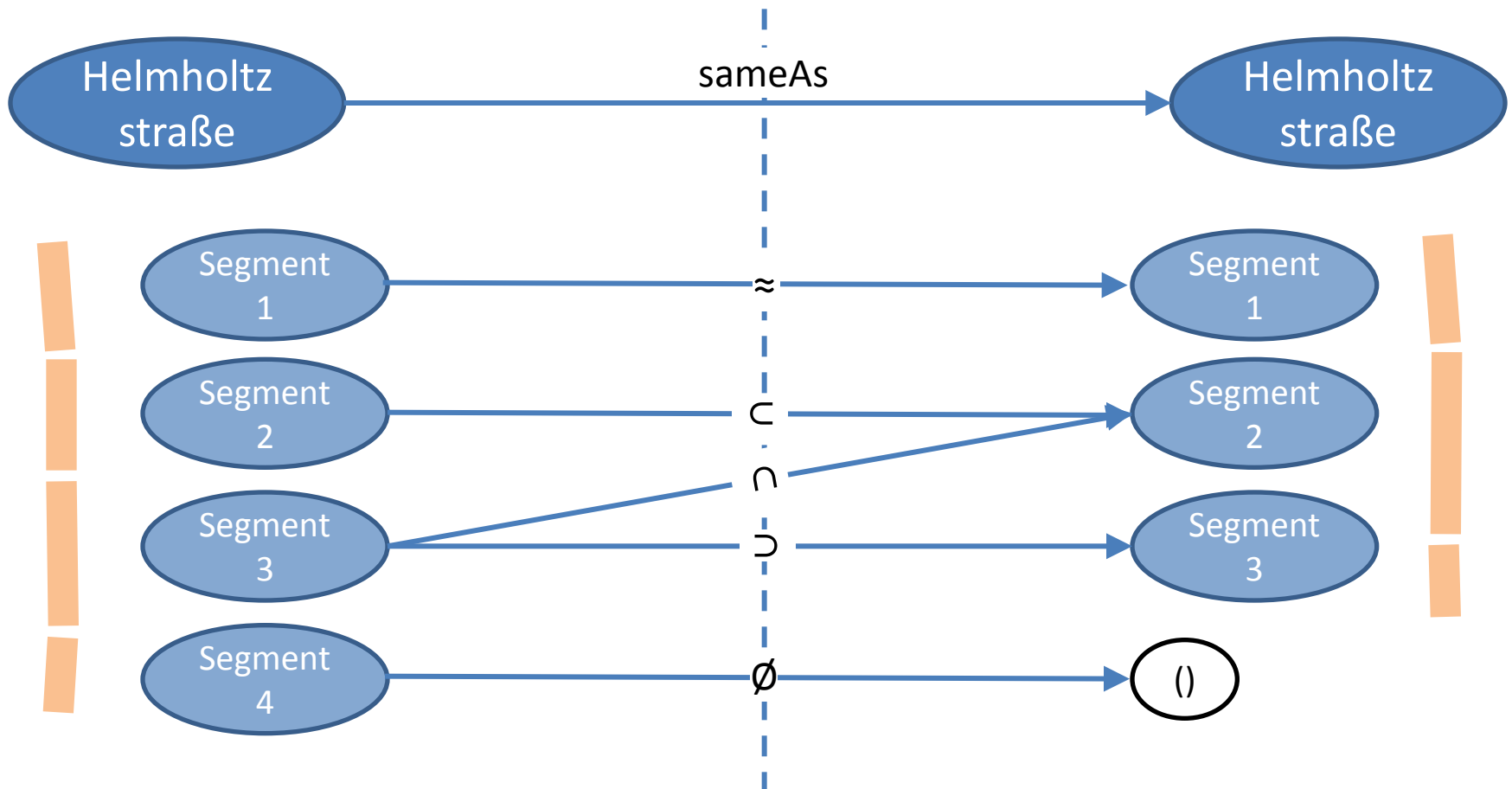


RDF (2)

- Nutzung einheitlicher Ontologien & Vokabulare



- Relationen zwischen Datensätzen



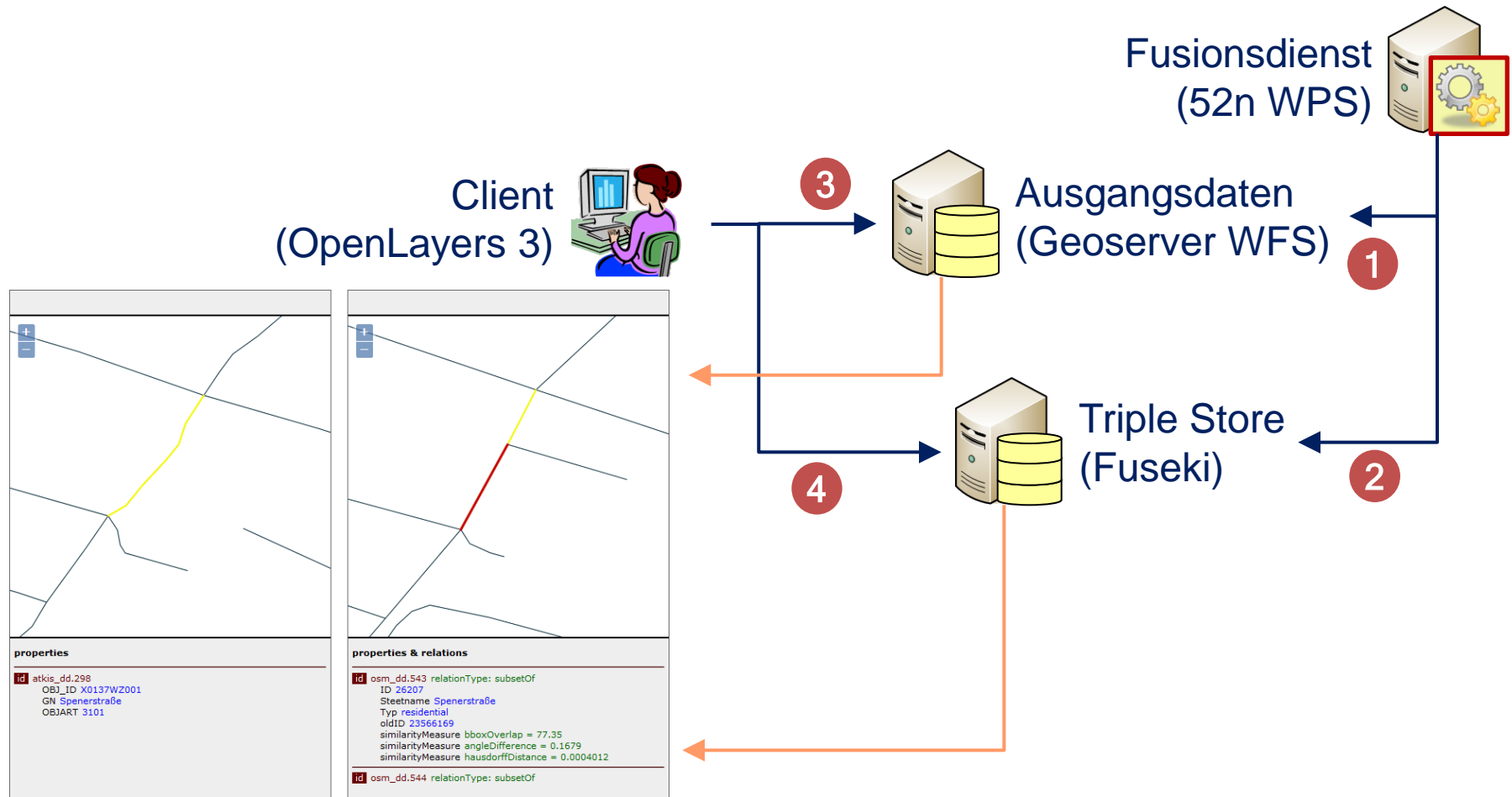
- Abfragesprache für Linked Data
 - Schnittstelle zu RDF Triple Stores
 - Vergleichbar mit SQL für relationale DB

- OGC Standard: GeoSPARQL
 - Definition geometrischer und topologischer Funktionen

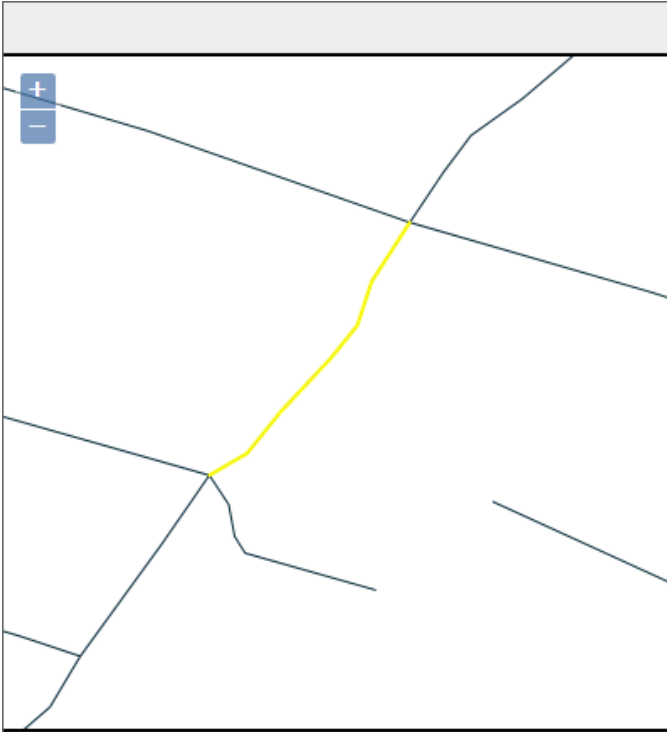
- Zielgerichtete Aktualisierung von Datensätzen
 - Fehlende Objekte (kein Gegenpart im jeweils anderen Datensatz)
 - Unvollständige oder fehlerhafte Objekte (überlappende Objekte)
 - Fehlerhafte Attribuierung (unterschiedliche thematische Attribute zugeordneter Objekte)

- Integrierte Nutzung mehrerer Datensätze
 - Ausgleich von Unsicherheiten
 - ATKIS → OSM: höhere geometrische und thematische Genauigkeit
 - OSM → ATKIS: höhere Aktualität, Vielzahl thematischer Attribute
 - Thematische Anreicherung durch weiterführende Links (z.B. Definitionen, Historie)

Implementierung (1)



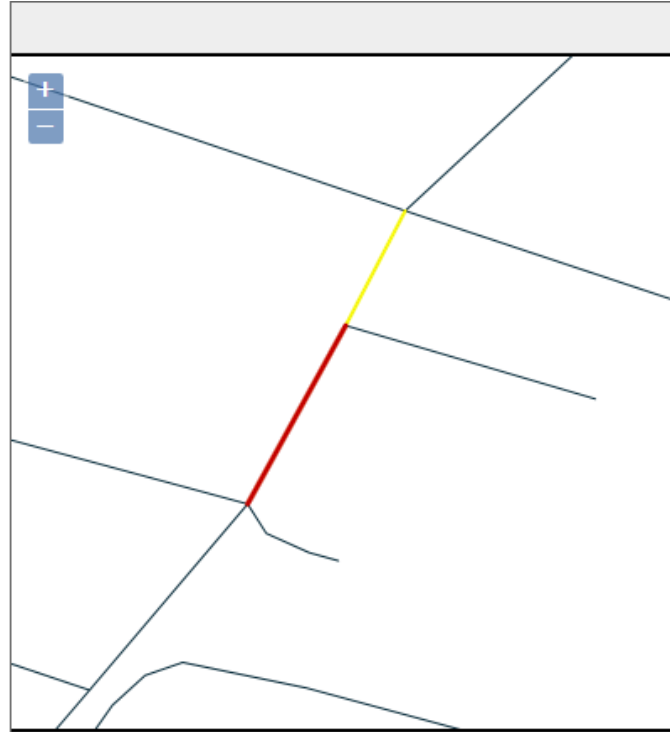
Implementierung (2)



A map view showing a network of grey lines representing streets. One segment is highlighted in yellow. In the top-left corner, there is a blue button with a '+' sign above a '-' sign, indicating zoom controls.

properties

id atkis_dd.298
OBJ_ID X0137WZ001
GN Spenerstraße
OBJART 3101



A map view showing the same network of grey lines. One segment is highlighted in red, and another segment is highlighted in yellow. The zoom controls are visible in the top-left corner.

properties & relations

id osm_dd.543 relationType: subsetOf
ID 26207
Steetname Spenerstraße
Typ residential
oldID 23566169
similarityMeasure bboxOverlap = 77.35
similarityMeasure angleDifference = 0.1679
similarityMeasure hausdorffDistance = 0.0004012

id osm_dd.544 relationType: subsetOf

Danke für die Aufmerksamkeit!

Stefan Wiemann

Technische Universität Dresden

stefan.wiemann@tu-dresden.de

- Abfrage einer Relation

[...]

```
SELECT ?targetURI ?similarityOperation ?similarityValue
WHERE {
    ?relation f:instanceOf <http://.../relation> .
    ?relation f:hasReference [variable referenceURI] .
    ?relation f:hasTarget ?targetURI .
    ?relation f:hasSimilarityMeasure ?similarityMeasure .
    ?similarityMeasure f:hasOperationURI ?similarityOperation .
    ?similarityMeasure f:hasValue ?similarityValue .
}
```

Fusionsstrategie

- Update, change detection, quality comparison, enrichment ,...

Fusion

- Auflösung Konflikte, conflation, Bestimmung Unsicherheiten

Suche & Bewertung

- Identifizierung von Kandidaten (basierend auf Metadaten)

Aspekte:
**Schema, Geometrie,
Topologie, Semantik**

Vergleich & Harmonisierung

- Transformation, Sub-settings, Generalisierung...

Konflikt Analyse

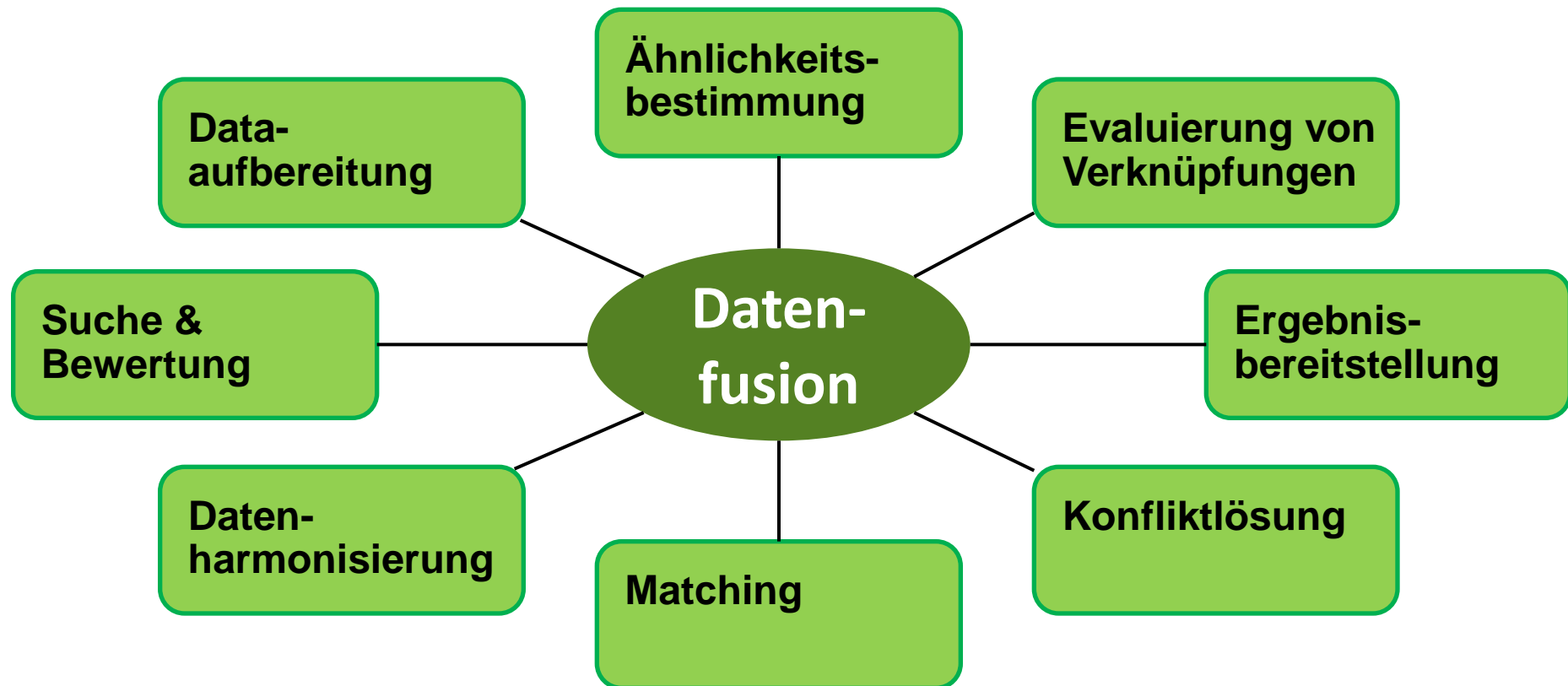
- Duplikate, change detection...

Matching

- Ähnlichkeitsbestimmung
- Identifizierung Objektverknüpfungen



Datenfusion



- **Geometrische Ähnlichkeitsmaße**
 - Hausdorff-Distanz (maximale Minimaldistanz zw. Punktmenge)
 - Fréchet-Distanz (parametrisierte Distanzmessung zw. Linien)
 - Überlappungen, Sinuosität,...
- **Topologische Ähnlichkeitsmaße**
 - Spider-Function (Betrachtung abgehender Kanten und deren Winkel)
 - Round-Trip-Walk (Verifizierung von Matches über Nachbarelemente)
- **Semantische Ähnlichkeitsmaße**
 - Logisches Schließen auf Basis von Referenzontologien (reasoning)
 - Attributvergleich
 - Numerisch: Differenzbildung
 - Zeichenketten: Damerau-Levenshtein-Distanz (Editierdistanz)

- Aufbau eines Online-Informationssystemes zur kollaborativen Erfassung (Crowdsourcing) von Umweltinformationen
- Testgebiete: UNESCO Biosphärenreservate
- Fusion mit Bestandsdaten (z.B. INSPIRE), u.a. zur Qualitätssicherung

