

Intelligente Datenerfassung, Haltung und Bereitstellung innerhalb der öffentlichen Verwaltung

Projektbericht

Arbeitskreis Architektur
der GDI-DE

Version	1.1
Datum	26.10.2021
Status	Final
Autoren	Falk Würriehausen, Timo Homburg, Claire Ponciano, Jean-Jacques Ponciano, Markus Schaffert
Herausgeber	AK Architektur
Ablage	GDI-DE Wiki
Haftungsausschluss	Dieser Bericht fasst die Ergebnisse der Architekturmaßnahme A1.14 zur intelligenten Datenerfassung, Haltung und Bereitstellung innerhalb der öffentlichen Verwaltung zusammen. Aussagen daraus können nicht im Rahmen gesetzlicher Verpflichtungen und deren Umsetzung geltend gemacht werden.

Änderungshistorie

Version	Datum	Änderung	Autor
0.1	10.07.2020	Initiale Version	Falk Würriehausen, Timo Homburg, Claire Prudhomme
0.2	04.09.2020	Fachliche Ergänzungen	Falk Würriehausen, Timo Homburg, Claire Prudhomme
0.3	10.09.2020	Redaktionelle Überarbeitung	Falk Würriehausen
1.0	20.10.2020	Abstimmung im AK Architektur	AK Architektur
1.1	26.10.2021	Ergänzung des Zwischenberichts zur initialen Version des Projektberichts	Falk Würriehausen, Claire Prudhomme, Jean-Jacques Ponciano, Markus Schaffert

Inhaltsverzeichnis

1. Einleitung.....	5
2. State Of The Art	5
2.1 <i>Das Geospatial Semantic Web</i>	6
2.2 <i>INSPIRE und das Geospatial Semantic Web</i>	6
2.3 <i>Mögliche Vorteile einer semantischen Datenintegration</i>	6
2.4 <i>XPlanung und INSPIRE</i>	7
2.5 <i>Weitere mögliche semantische Konvertierungsvorhaben</i>	7
2.6 <i>Technische Bereitstellung</i>	8
2.7 <i>Metadaten – DCAT, ISO 19115, ADMS.....</i>	8
3. Linked Data Architektur	8
3.1 <i>Semantic Uplift.....</i>	10
3.1.1 <i>Semantic Uplift mit dem GMLImporter.....</i>	10
3.1.2 <i>Extraktion von Ontologien aus XML Schemas.....</i>	10
3.1.3 <i>Semantic Uplift mittels Mapping Schemas</i>	10
3.1.4 <i>Schemaspeicherung.....</i>	11
3.2 <i>Datenhaltung: GraphDB.....</i>	11
3.3 <i>Datenexploration: GeoPubby</i>	11
3.4 <i>Datenexploration: Ontologiebrowser.....</i>	12
3.5 <i>Semantic Downlift: SemanticWFS</i>	13
3.5.1 <i>OGC API Features und der SemanticWFS</i>	13
3.6 <i>Verwaltung semantischer Metadaten.....</i>	15
3.6.1 <i>Semantischer Uplift von Metadaten</i>	15
3.6.2 <i>Verknüpfung von Metadaten mit Daten</i>	15
3.6.3 <i>Semantischer Downlift von Metadaten: OGC API Records</i>	15
3.7 <i>Anreicherung</i>	16
4. Ontologien, ihre Dokumentation und Verifikation	16
4.1 <i>Ontologien.....</i>	16
4.2 <i>Ontologiedokumentation</i>	17
4.3 <i>Verifikationsprozess</i>	17
4.4 <i>Ontologie für Metadaten und räumlich-zeitliche Daten</i>	18

5. Anwendungsfälle	18
5.1 Point Of Interest Daten.....	18
5.1.1 Erste Phase der Datenverarbeitung	19
5.1.2 Thematische Karte.....	20
5.2 ALKIS Testdaten	21
5.3 Anwendungsfall Schulen	22
5.4 Anwendungsfall XPlanung.....	22
5.5 Anwendungsfall UNESCO Weltkulturerbe	23
5.6 Zeitliche Landkarte	24
6. Fazit.....	25
7. Quellen.....	27
7.1 Wissenschaftliche Referenzen	27
7.2 Web-Quellen.....	27
7.3 Links zu Projektergebnissen.....	28

1. Einleitung

Die Problematik des Managements heterogener Daten sowie ihrer Verknüpfung ist ein wiederkehrendes Problem in der Geoinformatik sowie in den Geodateninfrastrukturen. Aktuelle Lösungen des Geodatenmanagements integrieren verschiedene Ressourcen wie Shapefiles bzw. GML oder KML Dateien innerhalb eines üblicherweise anwendungsfallspezifisch angelegten Datenbankschemas in Geodatenbanken wie z.B. POSTGIS. Die Bereitstellung der Geodaten erfolgt üblicherweise über OGC Webservices (z.B. WMS, WFS) oder als Downloads der jeweiligen Datenformate. Ein Datenbankbackend ermöglicht einer anwendungsfallspezifischen (Web-)Anwendung die Darstellung des für den jeweiligen Anwendungsfall benötigten Kartenmaterials, um somit userspezifische Probleme zu lösen. Hierbei erfolgt das Management der Daten auf der Ebene der Syntax. Die Daten werden grundsätzlich syntaktisch definiert (Welche Felder sollen in dem jeweiligen Datensatz oder in der jeweiligen Datenbanktabelle vorhanden sein?). Eine semantische Betrachtung und Modellierung der Daten nach Inhaltsgesichtspunkten (Was stellen die Datensets dar und welche Beziehungen haben sie zu anderen Datenquellen?) ist hierbei jedoch meistens nicht mit inbegriffen. Gerade diese Zusatzinformationen und Klassifizierungen schaffen jedoch die Möglichkeit einer einfacheren Integration von verschiedenen Datenquellen aus Bund und den Ländern, welche im Kontext verschiedener Anwendungsfälle zu Anwendung kommen könnten.

Die Maßnahme stellt sich dieser Herausforderung. Ihr Fokus liegt a) auf der Datenerfassung mit Hilfe intelligenter Plattformen, die eine Nutzer-Integration ermöglichen (z.B. Crowdsourcing, Cloudspeicherung) und b) auf der semantischen Integration von Daten in nachhaltigen Modellen (z.B. Schulstandorte in „XErleben“). Auch soll untersucht werden, welche Mehrwerte eine gemeinsam organisierte Bereitstellung für INSPIRE bietet, z.B. auf zentraler (Bundes-/Landes-) und dezentraler (Kommunal-)Ebene. Dabei ist eine semantische Modellierung für Daten anhand von Beispielen umzusetzen, um das Konzept der semantischen Integration und deren Vorteile zu erforschen.

2. State Of The Art

Das Semantic Web wurde im Jahre 2001 als das maschinenlesbare Web der Daten definiert [1] und wird in den Auszeichnungssprachen RDF (Resource Description Framework) [2] und OWL (Web Ontology Language) [3] beschrieben. Grundlage des Semantic Web ist das Konstrukt eines Triples, d.h. der Beschreibung einer Resource mittels eines Subjekts, Prädikates und eines Objektes (z.B. Berlin isCapitalOf Germany). Jedes Subjekt, Prädikat und Objekt wird hiermit über eine URI eindeutig beschrieben und kann somit als Referenz für eine Verlinkung von Datensets dienen. Ansammlungen von Triples werden in sogenannten Triple Stores (Datenbanken für Triples) gespeichert und im Internet vorgehalten. Triple Stores werden mit einer eigenen, vom W3C standardisierten Anfragesprache SPARQL [4] angefragt, welche seit einigen Jahren ebenfalls eine Geoerweiterung namens GeoSPARQL [5] für Abfragen für Georelationen bereitstellt. Schnittstellen zu Triple Stores, sogenannte SPARQL Endpoints, geben Benutzern sowie weiteren Anwendungen die Möglichkeit, Daten nicht nur im eigenen Triple Store abzufragen, sondern

auch durch Federated Queries weitere Ergebnisse von anderen Datenrepositorien einfach in das Abfrageergebnis zu integrieren [6].

2.1 Das Geospatial Semantic Web

Als Geospatial Semantic Web bezeichnet man die Teile des Semantic Web, welche mit Geodaten in Verbindung stehen. Aktuell existieren unter anderem die folgenden Triple Stores mit Geodaten:

- LinkedGeoData: Eine Ontologie über OpenStreetMap [L1]
- Wikidata: Koordinaten von Wikidatakonzepten sind oftmals annotiert [L2]
- GND: Koordinaten von Objekten der deutschen Nationalbibliothek [L3]

Qualitativ hochwertige von offiziellen Organisationen bereitgestellte Geodaten finden sich jedoch bisher kaum im Semantic Web, sodass viele verschiedene Anwendungen dieser Daten mit crowdgesourceten OpenStreetMap [L4] Daten vorgenommen werden. Hier besteht die Chance für die GDI-DE, eine Vorreiterrolle durch eine geschickte Integration von Daten einzunehmen.

2.2 INSPIRE und das Geospatial Semantic Web

Der INSPIRE Standard [L5] wurde als syntaktischer Standard zur Vereinheitlichung von europäischen Geodatenätzen von der europäischen Kommission vorgeschlagen und wurde in den einzelnen Nationalstaaten mittlerweile umgesetzt. Parallel dazu hat die europäische Kommission Möglichkeiten zur Erstellung einer INSPIRE Ontologie eruiert [L6, 7]. Dies zeigt, dass das Interesse der europäischen Kommission in die Richtung einer semantischen Repräsentation von Geodaten durchaus gegeben ist und vermutlich in Zukunft angegangen wird. Das SemanticGIS Projekt der Hochschule Mainz hat unabhängig hiervon ebenfalls eine INSPIRE Ontologie aus den offiziellen XML Schemas sowie Verbindungen zu anderen Semantic Web Ressourcen entwickelt (sog. Interlinks), welche eine einfache Integration anderer heterogener Datenquellen ermöglichen können [8]. Es liegt somit nahe, die Integration von GDI-DE und GDI-DE-nahen Daten in das Semantic Web zu erforschen, um für diese zukünftige Entwicklung vorbereitet zu sein.

2.3 Mögliche Vorteile einer semantischen Datenintegration

Eine semantische Datenintegration würde es der GDI-DE erlauben, Daten einfacher mit anderen Datenquellen mit Lizenzkennzeichnung, Lineageannotationen und Datenqualitätsangaben zu verknüpfen. Eine Verlinkung von bereits offenen Datenrepositories wie Wikidata mit GDI-DE Daten kann die externe und interne Entwicklung von Anwendungen unterstützen. Eine Möglichkeit stellen Anwendungen dar, durch die sich amtliche Daten mit externen offenen Daten (klar gekennzeichnet) anreichern lassen („enrichment“). Angereicherte amtliche Daten könnten für Kunden aus einem breiten Feld von Fachdisziplinen einen Mehrwert bieten.

Abstrakt werden diese Möglichkeiten in folgender Abbildung dargestellt: Datenquellen aus verschiedenen heterogenen Datenformaten werden in einer Wissensbasis zusammengeführt und können anschließend entweder angefragt, visualisiert oder in unterschiedlichen Kontexten wieder in die Ausgangsformate rückkonvertiert werden [9].

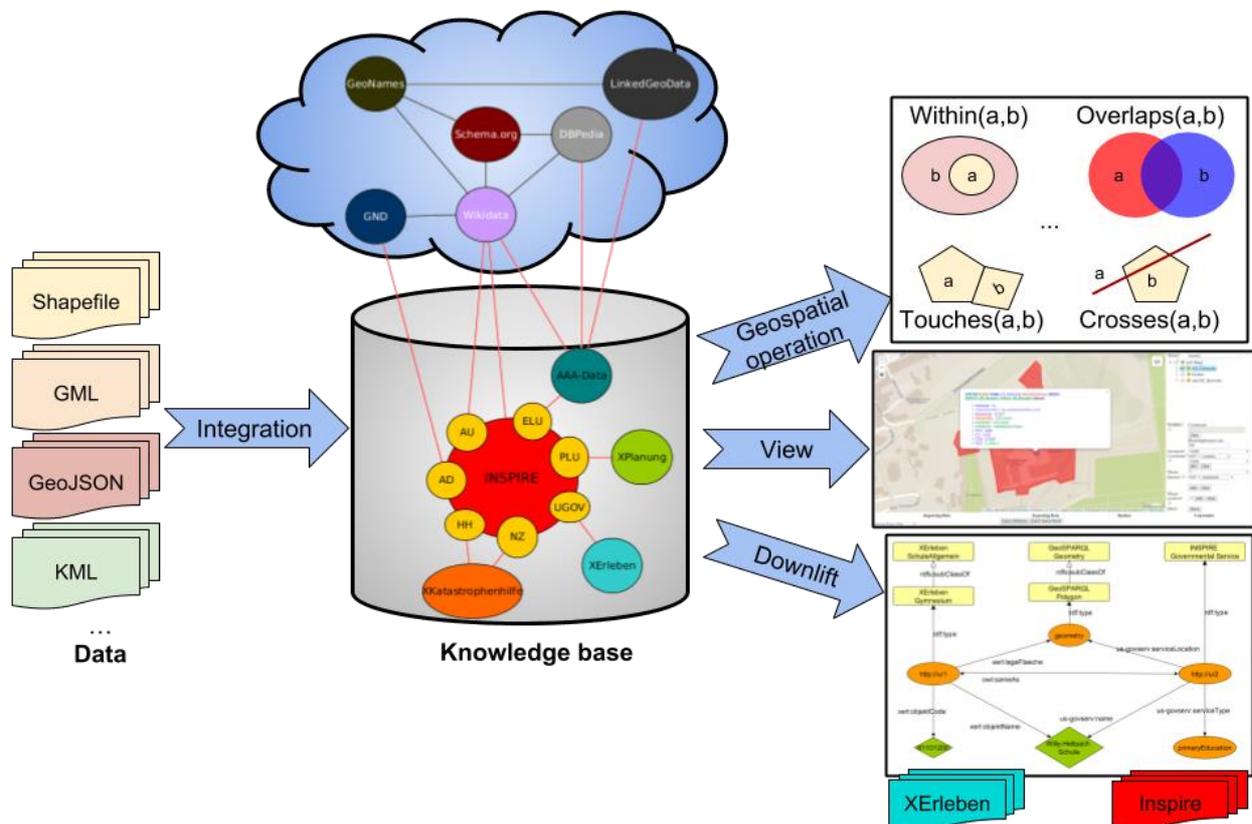


Abbildung 1 Projektübersicht: Linked Data SDI

2.4 XPlanung und INSPIRE

XPlanung [L7, 10] eignet sich aus mehreren Gründen als Experiment für eine semantische Integration. Zunächst existieren bereits konkrete Überführungsregeln von XPlanung auf INSPIRE-PLU (Planned Land Use) [11, 12], sodass die Relationen zwischen den beiden Formaten auch semantisch eindeutig definiert werden können. Desweiteren haben Bauleitpläne noch so gut wie nicht Einzug in das Semantic Web gehalten, da eine konkrete semantische Definition und eine unterstützende Community bislang fehlen. Bauleitpläne können jedoch in Kombination mit anderen Daten besonders im Zeitverlauf interessante Einblicke für potenzielle Benutzer des Semantic Webs eröffnen oder auch nur als semantisch erschlossenes Archiv für Bauleitpläne im Zeitverlauf mittels semantischer Technologien angesprochen werden.

2.5 Weitere mögliche semantische Konvertierungsvorhaben

Weitere mögliche Konvertierungen, die für die GDI DE interessant sein können, sind semantische Beziehungen zu ALKIS Datenbeständen, sowie Gebietsgrenzen, welche möglicherweise auch mit anderen Institutionen wie der deutschen Nationalbibliothek Überschneidungen aufweisen.

2.6 Technische Bereitstellung

Für die technische Bereitstellung von Semantic Web Daten mittels eines Triple Stores und eines sog. SPARQL Endpoints wurde im Vorhaben prototypisch eine Lösung beim BKG aufgesetzt und mittels Beispielanwendungen der Nutzen der Verlinkung und möglicher Konvertierungen aufgezeigt.

2.7 Metadaten – DCAT, ISO 19115, ADMS

Data Catalog Vocabulary (DCAT, <https://www.w3.org/TR/vocab-dcat-3/>). DCAT ist ein RDF-Vokabular, das entwickelt wurde, um die Interoperabilität zwischen im Web veröffentlichten Datenkatalogen zu erleichtern. DCAT ermöglicht es einem Bereitsteller, Datensätze und Datendienste in einem Katalog unter Verwendung eines Standardmodells und -vokabulars zu beschreiben, das die Nutzung und Zusammenführung von Metadaten aus mehreren Katalogen erleichtert. Dadurch lässt sich die Auffindbarkeit von Datensätzen und Datendiensten verbessern. Es ermöglicht darüber hinaus einen dezentralisierten Ansatz für die Veröffentlichung von Datenkatalogen und erlaubt eine föderierte Suche nach Datensätzen in Katalogen an mehreren Standorten unter Verwendung desselben Abfragemechanismus und derselben Struktur.

ISO19115 (<https://www.iso.org/standard/53798.html>). Die ISO 19115-1:2014 definiert das Schema, das für die Beschreibung von geografischen Informationen und Diensten durch Metadaten erforderlich ist. Es liefert Informationen über die Identifikation, den Umfang, die Qualität, die räumlichen und zeitlichen Aspekte, den Inhalt, den Raumbezug, die Darstellung, die Verteilung und andere Eigenschaften von digitalen geografischen Daten und Diensten.

Asset Description Metadata Schema (ADMS) vom W3C (<https://www.w3.org/TR/vocab-adms/>): ADMS ist ein Profil von DCAT, das zur Beschreibung von (semantischen) Assets verwendet wird, die als in hohem Maße wiederverwendbare Metadaten (z.B. xml-Schemata, generische Datenmodelle) und Referenzdaten (z.B. Codelisten, Taxonomien, Vokabulare) definiert sind, die für die Entwicklung von eGovernment-Systemen verwendet werden.

3. Linked Data Architektur

Für den im Vorhaben entwickelten Prototyp wurden Funktionen durch die Entwicklung von drei Webdiensten bereitgestellt: 1. ein Webdienst für den Datenimport (Importer), 2. ein Webdienst für die Datenverwaltung (SemanticWFS) und 3. ein Webdienst für die Verwaltung von Metadaten sowie räumlich-zeitlichen Daten (Geotime <https://github.com/Cprudhomme/GeoTimeWFS>). Der Zugang zu allen Funktionen des Prototyps erfolgt über eine Webschnittstelle, wie in der folgenden Abbildung dargestellt.

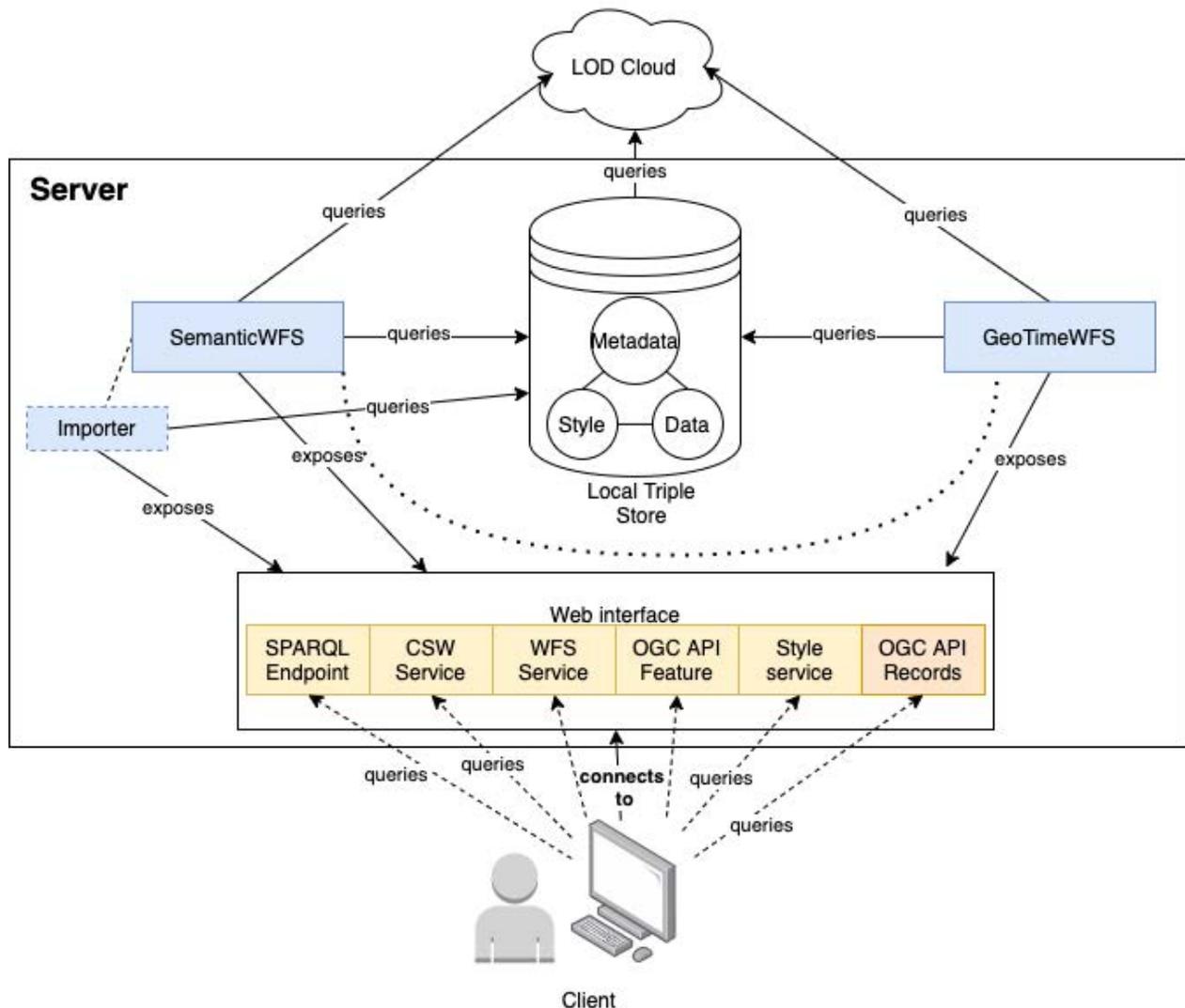


Abbildung 2 Webschnittstelle des im Vorhaben entwickelten Prototyps

Um Linked Data prototypisch im Rahmen der GDI-DE bereitzustellen, wurden verschiedene Komponenten entwickelt, welche die Bereitstellung von Linked data (SPARQL Endpoint), die Konvertierung von verschiedenen Geoformaten nach RDF (Semantic Uplift) sowie die Exploration und die Rückkonvertierung von Linked data in Geoformate für deren Bereitstellung (Semantic Downlift) ermöglichen sollten.

Infolge wurden semantikbasierte Werkzeuge für die Verwaltung von Metadaten, Datenversionen und zeitlichen Karten sowie für die Speicherung von Schemata für die gesammelten Daten entwickelt. Alle Komponenten zusammen bilden die Basis für eine Linked Data Infrastruktur in der GDI-DE und können in verschiedene Arbeitsabläufe integriert werden.

3.1 Semantic Uplift

Als Semantic Uplift bezeichnet man die Konvertierung von (Geo)daten nach RDF. Der Semantic Uplift-Prozess beinhaltet jedoch mehr als eine schlichte Konvertierung von Datenformaten. In diesem Prozess wird die Semantik, d.h. mindestens eine mögliche maschinenlesbare Interpretation der Informationen im jeweiligen Datensatz erfasst und mit abgespeichert. In der Maßnahme haben wir drei verschiedene Arten des Semantic Uplifts implementiert, welche verschiedene Voraussetzungen für ihre Anwendung aufweisen:

3.1.1 Semantic Uplift mit dem GMLImporter

Der GMLImporter (<https://github.com/i3mainz/GMLImporter> [E1]) ist ein Programm, welches ein XML-basiertes Geodatenfile (z.B. GML) nach RDF konvertieren kann. Jedes Element in einem XML File ist nach Definition in einem Namespace definiert, sodass mittels eines automatisierten Prozesses alle Informationen, die im jeweiligen XML Format vorliegen, verlustfrei nach RDF konvertiert werden können. Bei dieser Form der Konvertierung wird verlangt, dass die Interpretation der Daten, d.h. die Semantik, bereits in Form einer Datenformats-Ontologie vorliegt. Diese Datenformats-Ontologien können in einem Prozess (beschrieben in Abschnitt 3.1.2) extrahiert und anschließend manuell zu weiteren Konzepten im Semantic Web interlinkt werden.

3.1.2 Extraktion von Ontologien aus XML Schemas

Es gibt verschiedene Tools, welche die Extraktion von Ontologien aus XML Schemas ermöglichen. Viele dieser Tools basieren auf XSL Stylesheets, welche eine Konvertierung auf syntaktischer Ebene ermöglichen sollen. Aufgrund von möglichen Ambiguitäten in der Beschreibung von XML Schemas - es gibt verschiedene Möglichkeiten, ein Schema für einen Datensatz zu beschreiben - funktionieren einige dieser Vorlagen gut für einige XSD Schemas und schlechter auf anderen XSD Schemas. In jedem Fall sind jedoch noch einige manuelle Korrekturen vorzunehmen und die extrahierten Ontologien zu verifizieren - ebenfalls ein Schritt der in der Maßnahme erprobt wurde.

Eine solche automatisch extrahierte und manuell verifizierte Ontologie kann anschließend als Basis für die Integration sämtlicher Daten im beschreibenden Format über den GMLImporter verwendet werden. Zum Beispiel können nach Extraktion und Verifikation der AAA6 Ontologie aus den XML Schemas für AAA6 jegliche AAA6 formatierte Datensätze automatisch mittels des GMLImporters in den Triple Store integriert werden. Der Aufwand beschränkt sich hier also auf die einmalige Erstellung und Verifikation einer Ontologie pro Datenformat.

3.1.3 Semantic Uplift mittels Mapping Schemas

Viele Geodaten sind nicht in einem XSD Schema definiert, sondern liegen als Shapefiles, GeoJSON oder als andere Geodaten vor, deren Interpretation entweder in einer Dokumentation hinterlegt wurde (z.B. PDF) oder im schlimmsten Falle nur den Autoren der Daten selbst bekannt ist. In diesem Falle ist es notwendig, ein Mapping der relationalen Geodaten auf ein Ontologievokabular vorzunehmen.

Hierfür haben wir ein Importertool und ein XML Datenformat entwickelt, welches diese Mappings darstellen kann. Sobald ein Mapping einmal erstellt wurde, kann der Import der so formatierten Geodaten automatisiert werden.

3.1.4 Schemaspeicherung

Zur Speicherung der Schemata, die für das Uplifting von Daten verwendet werden, kann über die Weboberfläche ein Schematyp (XSD oder Mapping) ausgewählt, das zu speichernde Schema geladen und die upgelifteten Daten aus diesem Schema ausgewählt werden. Die Validierung dieses Backups ermöglicht die Speicherung des Schemas auf dem GDI-Gitlab mit Hilfe der Login- und Passwort-Informationen des Benutzers. Außerdem können das Schema und sein Speicherort über das DCAT-Vokabular mit den Daten in der Ontologie verknüpft werden.

3.2 Datenhaltung: GraphDB

Für die Maßnahme wurde ein Triple Store am BKG aufgesetzt, um die Linked data Daten zu integrieren und bereitzustellen. Nach Bewertung verschiedener Triple Stores wurde ein GraphDB Triple Store gewählt, welcher vollen Support für GeoSPARQL sowie ein einheitliches Userinterface für den Import von Linked data bietet.

Pro Datentyp wurden in der GraphDB verschiedene Repositories angelegt, welche jeweils einen SPARQL Endpoint, d.h. einen Webservice für Anfragen in der SPARQL Query Sprache bereitstellen. Desweiteren bietet die GraphDB die Möglichkeit, für Experten selbst Anfragen in einem Webinterface an die GraphDB zu formulieren und sich über die Daten, welche als Linked Open Data bereitgestellt werden, zu informieren.

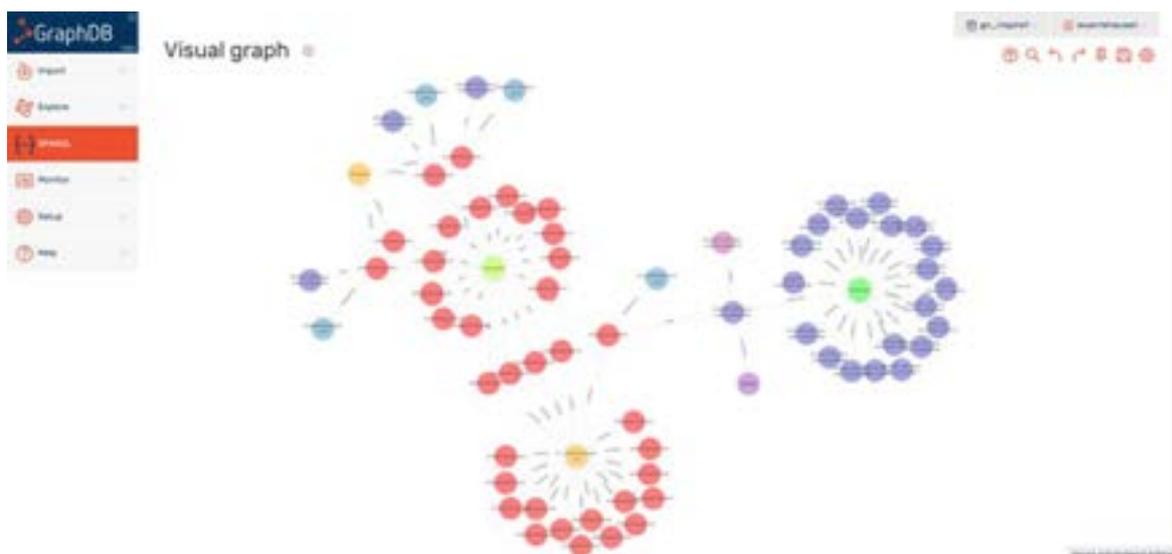


Abbildung 3: Visuelle Oberfläche von GraphDB

3.3 Datenexploration: GeoPubby

Eine Voraussetzung von Linked Open Data Anwendungen ist es, jede Ressource eines jeden Triples mit einer URI zu beschreiben. Eine Best Practice ist es, diese URIs auch im Webbrowser

Jedes Feature innerhalb einer Collection ist über eine persistente URI erreichbar.
 Alle Seiten, welche bereits erwähnt wurden, sind zudem neben ihrer HTML Repräsentierung auch in JSON und XML maschinenlesbar anfragbar.
 Schließlich spezifiziert die OGC API Features Spezifikation die Nutzung von OpenAPI als Dokumentationsmethode. OpenAPI beschreibt die API Calls und ermöglicht es Entwicklern, diese in ihre Applikationen einzubauen.

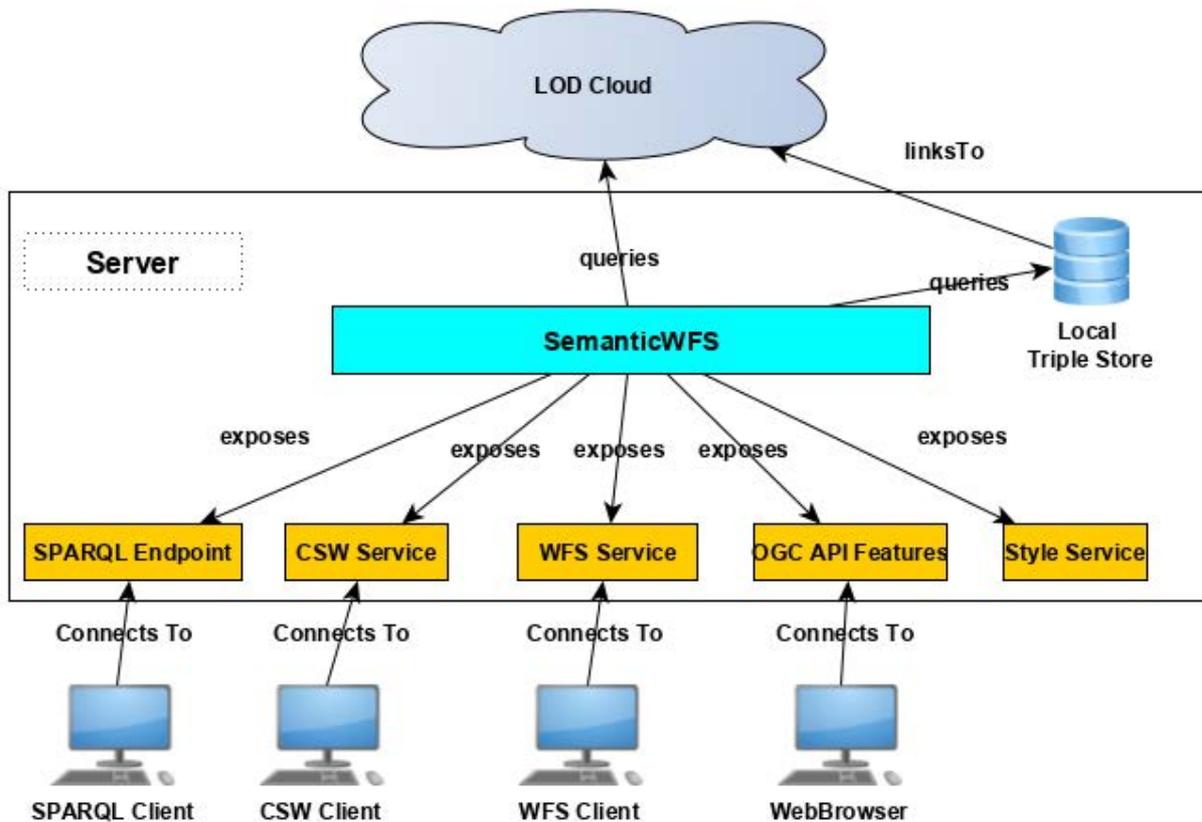


Abbildung 6: SemanticWFS Übersicht

Durch diese Architektur gewährleisten wir eine Interoperabilität zur Geocommunity, eine Publizierung von Linked Data für die Semantic Web Community und eine große Vielfalt an möglichen Formaten für die gegebenen Geodaten.

Ggf. wäre es im Rahmen des weiteren Projekts interessant, das Potenzial von schema.org für die Bereitstellung von Geodaten der GDI-DE für den Web-Mainstream zu untersuchen. Dies empfiehlt auch die OGC API Features Spezifikation in Recommendation 16.

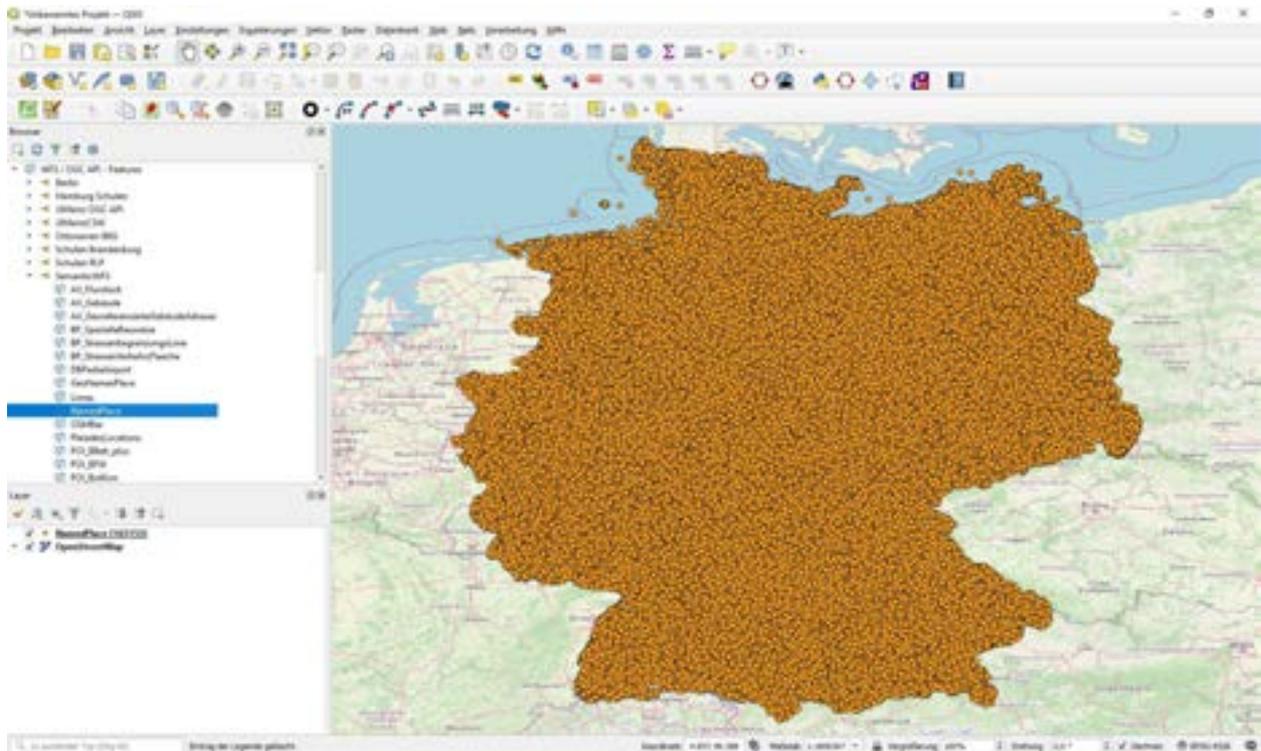


Abbildung 7: Collection Beispiel: Named Place

3.6 Verwaltung semantischer Metadaten

Die Verwaltung semantischer Metadaten erfolgt durch das Umwandeln in RDF-Tripel, die Verknüpfung der Metadaten mit den zugehörigen Daten und die Rück-Umwandlung in eine XML-Datei.

3.6.1 Semantischer Uplift von Metadaten

Die semantische Metadaten-Uplift-Funktionalität entspricht der Umwandlung einer XML-Datei nach dem ISO 19115-Format in ein RDF-Tripel.

Diese Funktion basiert auf dem Vokabular von ISO 19115. Sie ermöglicht es, den Inhalt der XML-Datei semantisch zu interpretieren.

3.6.2 Verknüpfung von Metadaten mit Daten

Nach der semantischen Integration von Daten und Metadaten werden diese mit Hilfe des DCAT-Vokabulars verknüpft. Dieses Vokabular ermöglicht die Verknüpfung eines Datensatzes und seines Standorts mit Metadaten.

3.6.3 Semantischer Downlift von Metadaten: OGC API Records

Der Prototyp bietet einen Webdienst für Metadaten. Dieser Webservice folgt den neuen Empfehlungen der OGC API Records (<https://ogcapi.org/records/overview.html>), die darauf

abzielen, den CSW (Catalog Service) Standard zu ersetzen. Dieser Webdienst ermöglicht unter anderem den Zugriff auf Kataloge und die darin gespeicherten Metadaten. Die Implementierung dieses Webdienstes liefert Ergebnisse in den Formaten JSON und HTML.

3.7 Anreicherung

Die Anreicherungsfunktion ermöglicht es dem Benutzer, Informationen aus Wikidata abzurufen, um sie der Ontologie hinzuzufügen und thematische Karten zu erstellen. Um die Nutzung dieser Funktionalität für Nicht-Experten in der Semantik zu erleichtern, kann der Nutzer eine Klasse seiner Wahl auswählen, die eine SPARQL-Abfrage generiert. Diese ermöglicht es, die Individuen dieser Klasse und ihren Standort abzufragen. Erfahrene Benutzer können ihre eigene Abfrage ändern oder erstellen. Darüber hinaus kann der Benutzer eine maximale Anzahl von Elementen definieren, die aus der Abfrage abgerufen werden sollen. Das Ergebnis der Anreicherung wird dann in einer Tabelle mit den verschiedenen Elementen, ihrer Bezeichnung und ihrem Standort sowie in einer Karte angezeigt, die den Standort der verschiedenen Elemente visualisiert. Das Ergebnis kann dann der Ontologie als Geodatensatz hinzugefügt werden, indem ein Name für den Datensatz festgelegt wird.

4. Ontologien, ihre Dokumentation und Verifikation

In der Maßnahme wurden verschiedene Ontologien für verschiedene Datenformate aus XML Schemas extrahiert und durch den Arbeitskreis Architektur verifiziert.

4.1 Ontologien

Alle Ontologierepräsentationen liegen pro Ontologie in verschiedenen Dateien vor, welche unterschiedliche Komponenten des Ontologiemodells beinhalten:

1. Konzept-Ontologie: Der Teil der Ontologie, der die Konzepte des Datenformats modelliert
2. Style-Ontologie: Der Teil der Ontologie, welcher Kartenstyles für die Konzepte der Konzeptontologie vorhält
3. Interlinking-Ontologie: Der Teil der Ontologie, welcher Links zu anderen Ontologien im Semantic Web aufweist
4. Codelisten-Ontologie: Der Teil der Ontologie, welcher Codelisten (sofern vorhanden) in OWL abbildet

Die folgenden Ontologien wurden auf diese Weise erstellt:

- AAA6, AAA7 Ontologie
- INSPIRE4 Ontologie mit allen Annexen und Codelisten
- XErleben2 Ontologie mit allen Codelisten
- XPlanung5, XPlanung 5.2 Ontologie

4.2 Ontologiedokumentation

Es existieren verschiedene Tools zur Dokumentation von Ontologien wie OWLDoc [L8] oder Widoco [L9]. Hierbei haben diese Tools unterschiedliche Zielsetzungen. OWLDoc generiert Ontologiedokumentationen, um diese lokal auf dem Rechner als HTML betrachten zu können. Widoco erstellt Ontologiedokumentationen mit dem Ziel der Auflösung von in den Ontologien enthaltenen URIs. Die Ontologiedokumentationen können somit unter der jeweiligen URL, welche eine Auflösung bewerkstelligen kann, gehostet werden. Durch einen Content-Renegotiation Mechanismus ist es mit Widoco Dokumentationen zudem möglich, auch eine RDF Repräsentation der jeweiligen Klasse durch Änderung des HTTP Anfrageheaders zu ermöglichen. Im Projekt wurde zunächst OWLDoc für die Dokumentation der Ontologien in der Alphaphase der Verifikation verwendet. Anschließend wurde Widoco für die Phase der Public Beta, in der die Ontologie bereits der Öffentlichkeit für Kommentare zur Verfügung gestellt wird, verwendet. Im Zuge der Ontologiedokumentation wurden verschiedene Unzulänglichkeiten in der Darstellung der Ontologiedokumentationen sichtbar. Im Speziellen wurde eine erschwerte Navigation durch die teilweise enorme Anzahl an Klassen bemängelt. Hierfür wurden Contributions und Issues für das Widoco Dokumentationstool erstellt, sodass ein neues Release der Widocosoftware diese Unzulänglichkeiten beheben wird.



Abbildung 8: Ontologiedokumentation Übersicht

4.3 Verifikationsprozess

In der Maßnahme haben wir 4 Phasen des Ontologieverifikationsprozesses definiert, welche wir als Best Practices für die Erstellung und Verifikation weiterer Ontologien ansehen:

1. Extraktion und manuelle Nachbearbeitung der extrahierten Ontologie -> Alpha Version
2. Korrektur der Ontologie nach interner Begutachtung durch Experten -> Beta Version
3. Begutachtung durch die breitere Öffentlichkeit mit öffentlichem Feedback
4. Release der Ontologie und ihrer Beschreibung an dem dafür vorgesehenen Ort im Web

Im Laufe der Maßnahme haben wir die Phasen 1-3 des Ontologieverifikationsprozesses durchlaufen, sodass nun eine Begutachtung der Öffentlichkeit stattfinden und nach dieser eine offizielle Standardisierung der Ontologien erfolgen kann.

Dieser letzte Schritt liegt außerhalb des Umfangs der Spezifikation der Maßnahme und wird durch die GDI-DE koordiniert werden.

4.4 Ontologie für Metadaten und räumlich-zeitliche Daten

- Die Ontologie für raum-zeitliche Metadaten und Daten besteht aus drei Hauptvokabularen, die miteinander verbunden sind:
- ISO 19115: Das Vokabular ISO 19115 wurde aufgrund seiner Standardisierung zur Beschreibung von Metadaten gewählt. Die Norm ISO 19115 wird für die Erstellung von XML-Metadaten-Dateien außerhalb der semantischen Welt empfohlen. Die Integration dieses Vokabulars soll die Aufhebung von XML-Dateien gemäß dem ISO 19115-Standard erleichtern.
- DCAT: Das DCAT-Vokabular basiert auf dem Dublin-Core-Vokabular, einem der in der semantischen Welt am häufigsten verwendeten Vokabulare zur Darstellung von Metadaten. Mit dem DCAT-Vokabular können Metadaten mit zugehörigen Datensätzen verknüpft werden. Dieses Vokabular wurde mit dem ISO 19115 Vokabular verbunden, um den Austausch und die Nutzung von Metadaten, die in die semantische Welt eingebettet sind, zu erleichtern. Dank der Verbindung zwischen dem DCAT- und dem ISO 19115-Vokabular können Metadaten sowohl als Dateien als auch innerhalb des Semantic Web leicht umgewandelt und gemeinsam genutzt werden. Das Mapping zwischen diesen beiden Vokabularen folgt den Empfehlungen des W3C (https://www.w3.org/2015/spatial/wiki/ISO_19115_-_DCAT_-_Schema.org_mapping)
- ADMS: ADMS basiert auf dem DCAT-Vokabular und ermöglicht die Darstellung von Datensätzen als Versionen. Diese Versionen sind mit einer Gültigkeitsdauer verbunden und mit anderen Versionen desselben Datensatzes verknüpft. Die Darstellung von Datensätzen in Form von Versionen ermöglicht sowohl die Verwaltung von raum-zeitlichen Daten als auch die Entwicklung und Korrektur von Datensätzen.

5. Anwendungsfälle

In der Maßnahme wurden verschiedene Anwendungsfälle für die Integration von Geodaten als Linked Data definiert. Alle Anwendungsfälle sind in jedem der vorher angesprochenen Systeme: GeoPubby und SemanticWFS verfügbar.

5.1 Point Of Interest Daten

Der Mehrwert des semantischen Webs besteht darin, dass die Daten gemeinsam genutzt werden können, so dass jeder seine eigenen Daten anreichern kann. Das Semantic Web wurde daher in diesem Projekt für den Anwendungsfall der POIs genutzt. Dieser Anwendungsfall, der auf die Anreicherung der BKG-Daten abzielt, wurde in zwei Schritten entwickelt. Der erste Schritt bestand darin, die Daten zu sammeln und sie mit den Konzepten des Semantic Web zu

verknüpfen. Der zweite Schritt bestand darin, aus den Daten des semantischen Webs thematische Karten zu erstellen.

5.1.1 Erste Phase der Datenverarbeitung

Das BKG stellt eine Reihe von Point Of Interest Daten im Shapefile Format in verschiedenen Kategorien bereit. Diese Daten werden jährlich neu erfasst und teilen sich einige Attribute in der Datenstruktur. Für alle Point Of Interest Daten wurden Mapping Schemas zu Wikidata Konzepten erstellt, um ihre Attribute zu beschreiben. Zudem wurde für jedes Point Of Interest Datenformat ein Interlinking zu Konzepten aus INSPIRE erstellt (inspire:GovernmentalService + Typ). Für ausgewählte Daten wurde außerdem ein Mapping, welches nur auf INSPIRE Konzepten beruhte, erstellt.



Abbildung 9: Integration von Point Of Interest

5.1.2 Thematische Karte

Dank der Anreicherung von verknüpften offenen Daten ist es möglich, thematische Karten von verschiedenen Points of Interest wie z. B. Kinos, Restaurants, Schulen oder Polizeistationen zu erstellen. Die Abbildung zeigt ein Beispiel für eine thematische Karte für Schulen. Diese thematischen Karten werden anhand von Daten aus dem semantischen Web entwickelt. Sie zeigen den Standort und die Eigenschaften von Individuen in einer als Thema gewählten Klasse.

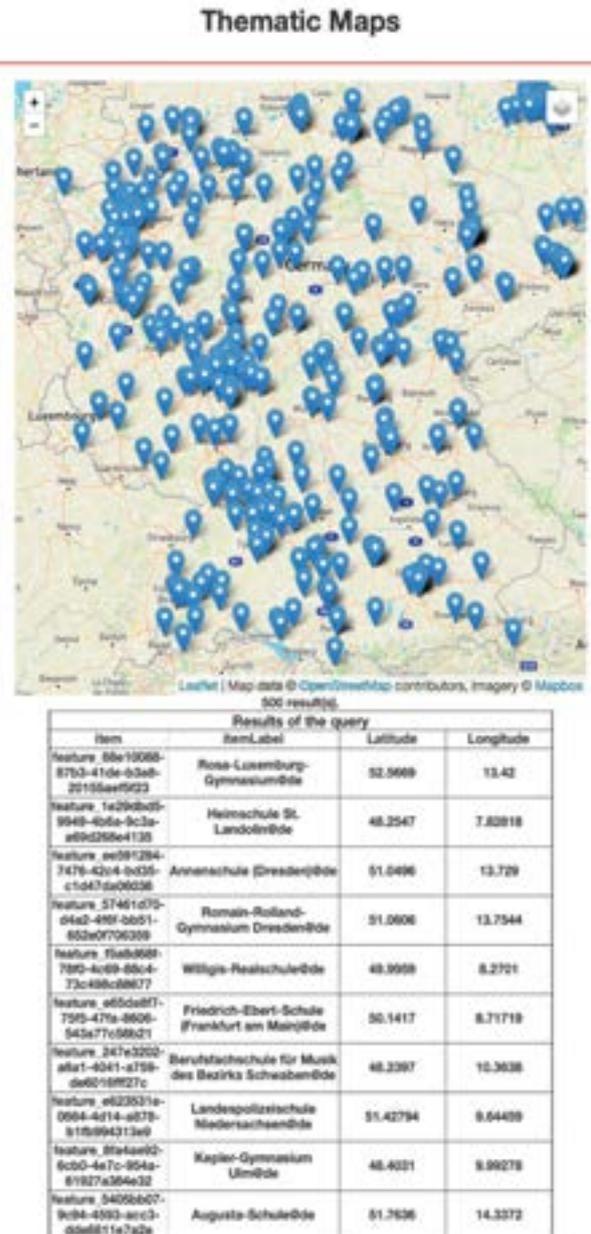


Abbildung 10: Themakarte: Point Of Interest

5.3 Anwendungsfall Schulen

Der Anwendungsfall Schulen in Deutschland erforderte die Integration von Schuldaten aus verschiedenen Bundesländern in verschiedenen Datenformaten. Hierbei wurden unterschiedliche Importmethoden verwendet, die wir im Folgenden exemplarisch erläutern:

GMLImporter:

Der GMLImporter wurde auf keinem der Datensets angewendet, weil keines der Datensets zu einer bereits im Vorfeld extrahierten Ontologie passte.

ImporterTool:

Da der GMLImporter nicht für die Schuldaten verwendet werden konnte, wurden für alle zu importierenden Schuldaten Mappingschemas erstellt. Hierbei wurde zunächst Wikidata als Ziel des Mappings gewählt, da Wikidata das Mapping von z.B. vielen verschiedenen Schultypen zuließ und die meisten zu mappenden Beziehungen bereits in Wikidata vordefiniert waren.

Alle Mappings erhielten neben eines Mappings zu einer Wikidata Schulklasse (wd:Q3914) auch Mappings zu INSPIRE:GovernmentalService mit einer entsprechenden Codelistenbeschreibung für education.

Nach Definition der Mappingschemas erfolgte die Konvertierung nach TTL mittels des Importertools anhand der Mappingschemabeschreibung. Die meisten Datensets dieses Anwendungsfalls waren in den Formaten SHP, GeoJSON und GML gegeben. In allen Fällen konnte das Importertool eine Konvertierung erreichen. Im Falle der Daten aus NRW waren die Schuldaten in XML gegeben und Codelisten, welche weitere semantische Dateninhalten beschreiben in weiteren XML Dateien beigelegt. Dies verlangte es, die gegebenen Codelisten mit in das Mappingschema zu integrieren, anschließend die XML Datei mittels QGIS in ein Geodatenformat umzuwandeln und anschließend das Importertool für die Konvertierung zu nutzen.

Schließlich wurden alle konvertierten Schuldaten in einem Repository in der GraphDB gespeichert und somit für die Verifikation z.B. mittels GeoPubby oder dem SemanticWFS verfügbar gemacht.

5.4 Anwendungsfall XPlanung

Dieser Anwendungsfall beinhaltet XPlanung Daten, welche mittels des GMLImporters importiert wurden. Der Standard XPlanung beinhaltet Äquivalenzbeziehungen von XPlanung zum äquivalenten INSPIRE Annex Planned Landuse welcher auch vor dieser Maßnahme bereits die Konvertierung von XPlanung Daten auf INSPIRE-PLU ermöglichte. Mit Linked Data ist es nun möglich, die Daten einmalig vorzuhalten und je nach Bedarf eine unterschiedliche Interpretation/Sicht auf die Daten anzunehmen. So kann beispielsweise auf denselben Daten ein FeatureType, welcher INSPIRE-GML zurückgibt, und ein weiterer FeatureType, welcher ebendiese Daten als XPlanGML zurückgibt, definiert werden. Eine doppelte Datenhaltung entfällt somit.

Für den Anwendungsfall XPlanung wurde ebenfalls ein Featuretype im SemanticWFS und ein Repository in der GraphDB erstellt. Zudem sind die Ergebnisse in GeoPubby abrufbar.

5.5 Anwendungsfall UNESCO Weltkulturerbe

Im Anwendungsfall UNESCO Weltkulturerbe konnte von Seiten des Arbeitskreises keine vollständige Abdeckung aller UNESCO Weltkulturerbestätten in Deutschland bereitgestellt werden. Alle UNESCO Welterbestätten sind jedoch in Linked Data Repositories wie Wikidata oder auch LinkedGeoData (OpenStreetMap) mit Attributen, welche zu einem Großteil nicht in den Daten des Arbeitskreises erfasst waren (aber auch mit möglicherweise schlechter zu verifizierender Qualität) gespeichert. Daraus können sich durchaus Chancen für die GDI-DE ergeben, denn die Erstellung einer Karte aller Welterbe mit Markierung der Datenherkunft ist mittels integrierten Linked data Ressourcen stark vereinfacht.

Auch für die UNESCO Welterbe Datensets wurde das Importertool für die Konvertierung verwendet. Hierfür wurden ebenfalls Mappingschemas im GMLImporter hinterlegt. Im Gegensatz zu den Schulen wurden auf verschiedene Typen innerhalb der UNESCO Datensets gemappt.

Für das UNESCO Weltkulturerbe ist neben der Darstellung im SemanticWFS und über GeoPubby eine Themenkarte erstellt worden in der neben den integrierten Daten auch UNESCO Welterbestätten aus OpenStreetMap und Wikidata dargestellt werden.

Der Anwendungsfall UNESCO Welterbe zeigt somit relativ gut die Möglichkeiten auf, nicht-vollständige Datensets durch weitere Linked Data Ressourcen anzureichern.



Abbildung 13: Anwendungsfall UNESCO Weltkulturerbestätten

5.6 Zeitliche Landkarte

Das BKG verwaltet eine große Menge an Geodaten, die jedes Jahr aktualisiert werden. Es ist daher notwendig, Versionen von mehreren Datensätzen zu verwalten. Da das semantische Web die Anreicherung dieser Daten mit verwandten offenen Daten ermöglicht, ist es für das BKG interessant, die Datensätze als RDF-Tripel zu speichern und gleichzeitig die Visualisierung und Entwicklung dieser Datensätze als Versionen zu ermöglichen. Eine solche Datendarstellung erleichtert die Anreicherung und Korrektur von Daten über das semantische Web. Es erleichtert auch die Verwaltung von Versionen dieser Datensätze, wenn sie aktualisiert, korrigiert und angereichert werden. Es wurde eine Fallstudie entwickelt, um die Nützlichkeit und Verwendung einer solchen Darstellung zu demonstrieren.

Das Ergebnis dieser Fallstudie wird in Form einer Zeitkarte präsentiert, die generiert wurde, um die Entwicklung der deutschen Grenzen von 1998 bis 2000 zu visualisieren. Die integrierten Daten für diesen Anwendungsfall stammen aus dem BKG-Datenprodukt „Verwaltungsgebiete Historisch“¹. Dieser Anwendungsfall besteht aus einer Karte, die sich entsprechend den Interaktionen des Benutzers mit einem Zeitschieber entwickelt, wie unten dargestellt.



Drag the slider to display the current version information.

Abbildung 14: Raum-zeitliche Daten: Zeitliche Landkarte

¹<https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/verwaltungsgebiete/verwaltungsgebiete-historisch-vg-hist.html>.

6. Fazit

In der Maßnahme wurden zunächst die **Grundlagen für eine Linked Data Architektur** in der GDI-DE gelegt. Für die Integration der verschiedenen Usecases war es möglich, deren Ergebnisse in einer Form zu präsentieren, die für den Arbeitskreis einfach nachvollziehbar und verständlich war. Dies wurde in erster Linie durch die Komponenten GeoPubby und SemanticWFS erreicht, wobei ersteres 1:1 das Datenmodell abbildet und zweiteres die Sicht der Datenbereitstellung durch die GDI-DE widerspiegelt. Beide Visualisierungen dienen als Prototypen und können für weitere Anforderungen in der GDI-DE angepasst werden.

In einer **zweiten Phase** wurde die Komponente Geotime entwickelt. Diese ermöglicht:

- die Verwaltung von Metadaten in semantischer Form,
- die Verwaltung und Speicherung des Schemas, das beim Uplift der Daten verwendet wird,
- die Versionsverwaltung von Datensätzen, zur Erstellung von zeitlichen Karten und,
- die Entwicklung von thematischen Karten aus semantischen Webdaten.

Die Anwendung bietet auch einen speziellen Metadaten-Webservice, der den Empfehlungen der OGC API Records folgt. Diese Funktionen werden den Nutzern über eine Web-Client-Schnittstelle zur Verfügung gestellt, die im Folgenden dargestellt wird:

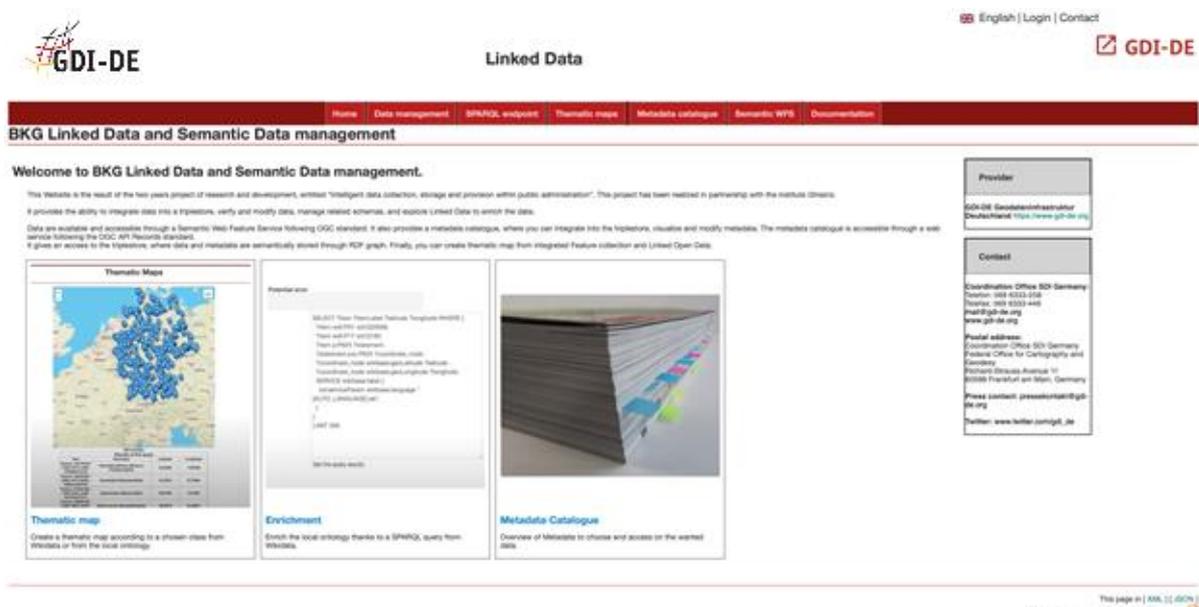


Abbildung 15: Web-Client der Linked-Data-Plattform (Prototyp)

Neben den Web-Schnittstellen wurden auch verschiedene **Dokumentationsmaterialien** erstellt. Das erste Dokumentationsmedium beschreibt die bereits extrahierten Ontologien. Diese Ontologien wurden nicht nur dokumentiert; die Dokumentation liegt nun auch in einer Form vor, welche sich mittels Content-Negotiation zur Auflösung von URIs nutzen lässt, sofern sie unter den richtigen URIs gehostet werden. Dies erlaubt sowohl das Hosten von RDF Daten als auch das Hosten von Definitionen, die mit den RDF Daten mitgeliefert werden. Schließlich wurde der Import und die Konvertierung von Geodaten nach RDF in verschiedenen Varianten erforscht und

behandelt. Es existieren nun Varianten für einen Import mittels vorextrahierten Ontologien und mittels eines Mappingschemas in verschiedenen Formaten.

Schließlich wurden Potenziale der **verschiedenen Anwendungsfälle** in Form von Karten präsentiert. Der UNESCO Welterbe-Anwendungsfall deckte auf, dass GDI-DE Daten nicht vollständig waren, aber durch Linked Data Ressourcen ergänzt werden können. Der XPlanung Usecase zeigt auf, dass Daten mittels semantischer Technologien unterschiedlich (z.B. einmal als INSPIRE und einmal als XPlanung-Datensatz) interpretiert werden können. Diese Interpretation verlangt kein dupliziertes Speichern von Daten. Weitere Konvertierungsvorhaben, wie der Anwendungsfall der Schulen, der Points Of Interests oder der ALKIS Testdaten zeigten die Übertragbarkeit des Konzeptes auf weitere Anwendungsfälle und Datensets.

Der Anwendungsfall der Zeitkarte, die die Entwicklung der deutschen Grenzen darstellt, veranschaulichte, wie die semantische Darstellung von Datensätzen als Versionen genutzt werden kann. Dank der semantischen Verwaltung von Versionen ist es möglich, die Entwicklung von Geodaten im Laufe der Zeit infolge von Aktualisierungen, Korrekturen oder Anreicherungen zu visualisieren.

Das Vorhaben ermöglichte die Entwicklung von Werkzeugen für die Darstellung von raumzeitlichen Daten und ihren Metadaten in semantischer Form sowie deren Anreicherung durch das semantische Web. Diese Werkzeuge haben ihre Effizienz in verschiedenen Anwendungsfällen bewiesen. Sie erleichtern die gemeinsame Nutzung von GDI-DE Daten als verknüpfte Daten und den Abruf von Daten aus dem semantischen Web. Es wäre jedoch weiterhin gewinnbringend, die Integration von raumzeitlichen Daten unter demselben Vokabular in einer Wissensdatenbank zu vereinheitlichen und die verschiedenen im Rahmen dieses Projekts entwickelten Funktionalitäten zukünftig in einer „**Spatial Data on the Web/Linked Data**“ Plattform zusammenzuführen, um eine robuste Architektur für verknüpfte Daten zu schaffen. Darüber hinaus wäre es künftig auch zielführend, die verschiedenen Funktionen in einem Hackathon durch externe Nutzer zu testen, und diese dabei zu neuen Anwendungsfällen zu animieren.

7. Quellen

Die Quellen dieses Dokuments bestehen aus wissenschaftlichen Referenzen, Web-Quellen und Links zu den Projektergebnissen.

7.1 Wissenschaftliche Referenzen

- [1] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american*, 284(5), 34-43.
- [2] Lassila, O., & Swick, R. R. (1998). Resource description framework (RDF) model and syntax specification.
- [3] Antoniou, G., & Van Harmelen, F. (2004). Web ontology language: Owl. In *Handbook on ontologies* (pp. 67-92). Springer, Berlin, Heidelberg.
- [4] Seaborne, A., & Prud'hommeaux, E. (2006). SPARQL query language for RDF, W3C recommendation (January 2008).
- [5] Battle, R., & Kolas, D. (2011). Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*, 3(4), 355-370.
- [6] Rakhmawati, N. A., Umbrich, J., Karnstedt, M., Hasnain, A., & Hausenblas, M. (2013). Querying over Federated SPARQL Endpoints---A State of the Art Survey. *arXiv preprint arXiv:1306.1723*.
- [7] ERDOĞDU, M. & YALÇIN, G. (2012). Geo-Ontology for INSPIRE Data Themes.
- [8] Homburg, T., Prudhomme, C., Würriehausen, F., Karmacharya, A., Boochs, F., Roxin, A., & Cruz, C. (2016, July). Interpreting heterogeneous geospatial data using semantic web technologies. In *International Conference on Computational Science and Its Applications* (pp. 240-255). Springer, Cham.
- [9] Homburg, T., Prudhomme, C., & Boochs, F. (2018). Semantische Geoinformationssysteme: Integration heterogener Geodaten am Beispiel XErleben. *Fach Austausch Geoinformation by GeoNet.MRN*.
https://www.researchgate.net/publication/329371465_Semantic_Geographic_Information_System_Integration_and_management_of_heterogeneous_geodata
- [10] Krause, K. U., & Munske, M. (2016). Geostandards XPlanung und XBau. *zfv--Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, (5), 141.
- [11] Benner, J. (2013). Von XPlanung zu INSPIRE--Automatische Erzeugung von INSPIRE Planned Land Use Daten aus XPlanGML. *Allgemeine Vermessungs-Nachrichten (avn)*, 119.
- [12] Friebe, T. (2018). From XPlanung and XPlanGML to INSPIRE PLU. In *INSPIRE Conference 2018*.

7.2 Web-Quellen

- [L1] LinkedGeoData, eine Ontologie über OpenStreetMap:
<https://wiki.openstreetmap.org/wiki/LinkedGeoData>
- [L2] Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page
- [L3] GND, die Integrierte Autoritätsdatei der deutschen Nationalbibliothek:
https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html

- [L4] OpenStreetMap: <https://www.openstreetmap.org>
- [L5] INSPIRE Standard, als syntaktischer Standard zur Vereinheitlichung von europäischen Geodatenätzen von der europäischen Kommission vorgeschlagen:
<https://inspire.ec.europa.eu>
- [L6] INSPIRE Ontology: <https://inspire.ec.europa.eu/glossary/Ontology>
- [L7] XPlanung Wiki: http://www.xplanungwiki.de/index.php?title=Xplanung_Wiki
- [L8] OWLDoc für Dokumentation von Ontologien:
<https://protegewiki.stanford.edu/wiki/OWLDoc>
- [L9] Widoco für Dokumentation von Ontologien: <https://github.com/dgarijo/Widoco>

7.3 Links zu Projektergebnissen

- [E1] GMLImporter, ein Programm welches ein XML-basiertes Geodatenfile (z.B. GML) nach RDF konvertieren kann: <https://github.com/i3mainz/GMLImporter>
- [E2] GeoPubby, eine Modifikation des Tools Pubby (<https://github.com/cygri/pubby>) welches besagte URI Auflösung vornehmen kann:
<https://github.com/i3mainz/geopubby>
- [E3] GDI-DE Instanz von GeoPubby: <http://ld.gdi-de.org/id/>
- [E4] SemanticWFS, ein WFS/OGC API Features Service mit Triple Store Backend
<https://github.com/i3mainz/semanticwfs>
- [E5] Web-Client GeoTimeWFS: <https://geotimewfs.herokuapp.com>
- [E6] GeoTimeWFS im github: <https://github.com/Cprudhomme/GeoTimeWFS/wiki>
- [E7] Benutzerdokumentation: <https://geotimewfs.herokuapp.com/documentation>