

Universität Leipzig  
Fakultät für Mathematik und Informatik  
Wintersemester 2021/22  
Bachelorarbeit

---

# **Transformation von Daten des Amtlichen Liegenschaftskatasterinfor- mationssystem in semantische Datenformate**

---

Betreuer: Prof. Dr. Hans Gert Gräbe, Dr. Christian Zinke-Wehlmann, Norman Radke

Autor: Robin Seidel

Abgabedatum: .03.2022

# Inhaltsverzeichnis

|     |  |    |
|-----|--|----|
| 1   | Einleitung . . . . .                       | 1  |
| 1.1 | Themenbeschreibung . . . . .               | 1  |
| 1.2 | Motivation . . . . .                       | 1  |
| 2   | Grundlagen . . . . .                       | 2  |
| 2.1 | ALKIS . . . . .                            | 2  |
| 2.2 | ALKIS Daten . . . . .                      | 3  |
| 2.3 | RDF/RML . . . . .                          | 6  |
| 2.4 | RDF vs XML . . . . .                       | 6  |
| 2.5 | RDF Schema, Ontologies . . . . .           | 6  |
| 2.6 | RML/Yarrml . . . . .                       | 7  |
| 2.7 | Abkürzungen . . . . .                      | 7  |
| 3   | State of the Art . . . . .                 | 7  |
| 3.1 | Problembeschreibung . . . . .              | 7  |
| 3.2 | Möglichkeiten der Transformation . . . . . | 8  |
| 3.3 | Tripelstores . . . . .                     | 8  |
| 4   | Anforderungsanalyse . . . . .              | 8  |
| 5   | Implementierung . . . . .                  | 8  |
| 5.1 | genutzte Technologien . . . . .            | 8  |
| 5.2 | Funktionsweise . . . . .                   | 9  |
| 6   | Ergebnisse und Evaluation . . . . .        | 9  |
| 7   | Ausblick . . . . .                         | 9  |
| 8   | notizen . . . . .                          | 9  |
| 9   | Quellenverzeichnis . . . . .               | 10 |

# 1 Einleitung

## 1.1 Themenbeschreibung

## 1.2 Motivation

Mit der Einführung des World Wide Webs im Jahre 1990 durch Tim Burner Lee ist das Internet heute viel mehr als das, was es früher einmal war. Der Mensch nutzt das Internet immer häufiger als Hauptinformations- und Interaktionsmöglichkeit über diverse Unterhaltungskanäle wie Facebook, Youtube oder andere Websites. (QUELLE) Während am Anfang lediglich ein paar Dutzend Webseiten existierten sind es mittlerweile über zwei Milliarden mit 100 Milliarden Webdokumenten, die sich wiederum alle 6 Monate verdoppeln. (vgl. Meinel 2021) Aber es entstehen nicht immer nur mehr Webseiten auch die Anzahl der Internetnutzer hat sich in den letzten Jahren drastisch gesteigert. 1995 hat der Spiegel 250.000 Nutzer in Deutschland geschätzt, während 2018 durch das statistische Bundesamt ein Anstieg auf 90 wurde. (vgl. Köllinger 2003 zitiert nach: Der Tagesspiegel vom 19. August 1995, S. 23) Demnach nutzen mittlerweile 66,5 Millionen Menschen ab 10 Jahren das Internet als Informations- und Interaktionsquelle. (vgl. Statistisches Bundesamt 2018) Bei einer solchen Masse an Informationen und Daten wird es immer schwerer relevante Informationen herauszufiltern, obwohl das Hauptziel des Internets einst nicht die Unterhaltung, sondern die Informationsbeschaffung war. (QUELLE) Deshalb ist es wichtig die vorhandenen Daten mit Hilfe von RDF Graphen in einer geordneten Art und Weise zur Verfügung zu stellen. Das semantische Web war bereits eine Idee des Entwicklers/Gründers Tim Burner Lee. Er hatte die Vision, dass das Netz von Informationen nicht nur Dokumente und Daten miteinander verbindet, sondern das diesen Daten auch eine Bedeutung gegeben wird. Damit ermöglicht man einer Maschine die Anhäufung von Daten automatisiert nutzen zu können.

- beschreibung semantik web und wichtigste Prinzipien - 4 prinzipien: tim burner lee - nutzen von URI's als namen für objekte - zur eindeutigen identifikation für dinge - nutze HTTP URIs - so können Menschen die Objekte nachschauen - nutze anerkannte standarts( RDF, Sparql) - linke URIS weiter Das semantische Web ist eine Art rohe(also unearbeitete) Daten zuübertragen. (Interoperabilität) Um die Informationen die das Web liefert in einen strukturierten Weg wiederzugeben. Mit Hilfe der daraus entstehenden strukturierten Daten kann die Software diese Daten anschließend besser nutzen und verarbeiten. Das semantische Web ist damit sozusagen eine Alternative zu Webseiten die ausschließlich für die menschliche Nutzung geeignet sind. Die Vorteile die eine Semantifizierung der Daten mit sich bringt sind unter anderem das Abfragen durch Sparql Querys, eine Verbesserung der Bedeutung der Daten, das Verbinden von Daten und die Nutzbarkeit für andere Bereiche.

- überleitung auf Alkis Thematik Alkis Daten(siehe Begriffsklärungen) werden zwar in gewissen Standart und Einheitlichkeit erhoben, allerdings ist die Verarbeitung dieser

Daten hauptsächlich für GeoProgramme(näher erklären) vorgesehen. Die Art der Daten ist nicht für die optimale Verarbeitung von ComputerProgrammen und zur Nutzung für den gewöhnlichen Nutzer vorgesehen. Eine Überführung in Rdf Graphen ermöglicht es das Potenzial der Daten freizuschalten und die Nutzung für zahlreiche andere Bereiche zur Verfügung zu stellen. Quellen: (<https://www.spektrum.de/kolumne/semantic-web-wie-das-internet-inhalte-verstehen-koennte/> 1841389) paper über tim terner lees arbeit [https://www.researchgate.net/publication/307845029\\_Tim\\_Berners-Lee\\_s\\_Semantic\\_Web](https://www.researchgate.net/publication/307845029_Tim_Berners-Lee_s_Semantic_Web) paper von burnier über das semantic web 10.3.2022 <https://www.w3.org/2000/Talks/0906-xmlweb-tbl/text.htm> statistik bundesamt [https://www.destatis.de/DE/Presse/Pressemitteilungen/2018/09/PD18\\_330\\_634.html](https://www.destatis.de/DE/Presse/Pressemitteilungen/2018/09/PD18_330_634.html) Köllinger, Philipp <https://www.econstor.eu/handle/10419/151237> (Verlag und Jahr noch angeben)

## 2 Grundlagen

### 2.1 ALKIS

Das Amtliche Liegenschaftskatasterinformationssystem, kurz ALKIS, wird von der Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland, kurz ADV verwaltet und ist seit 2015 in allen Bundesländern aktiv. Kataster- und Liegenschaftsinformationen sind Geobasisdaten über sogenannte Liegenschaften, also zum Beispiel Flurstücke oder Gebäude. Diese Geobasisdaten werden durch die Vermessungs- und Katasterverwaltungen der Länder erhoben (eventuell nähere Beschreibung notwendig)

Durch die Einführung von ALKIS wurden alle Daten des Liegenschaftskatasters redundanzfrei zusammengeführt. Es wurden internationale Standards, einheitliche Objektabbildungskataloge, Dateninhalte und Datenaustauschformate geschaffen. (quelle s 10)

Bereits vor der Veröffentlichung geobasierter Daten über ALKIS wurden Daten über das automatische Liegenschaftsbuch (Textform) und die automatische Liegenschaftskarte (geographischer Teil) gesammelt. Der Nachteil liegt darin, dass teilweise technisch veraltete Konzepte, verschiedene Datenformate, getrennte und teilweise redundante Datenaufhaltung aufeinander treffen und zu dem länderspezifisch implementiert wurden. (Quelle s8 13.3) -> vorteile von alkis gegenüber vorgängersystem(seite11)

<https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/ALKIS/broker.jsp?uMen=e5670f15-8e71-3c01-e1f3-351ec0023010> <https://www.adv-online.de/AdV-Produkte/Liegenschaftskataster/ALKIS/>

11.3.22 ADV Online - ALkis Beschreibung

## 2.2 ALKIS Daten

Die Bestandsdaten aus ALKIS, können über eine sogenannte Normbasierte Austausch-schnittstelle(NAS) abgerufen werden. Diese wurde speziell für den Austausch von Daten des AAA - Models entwickelt. Das AFIS-ALKIS-ATKIS Referenzmodell, enthält neben den Grundsätzen für ALKIS auch noch Regeln für das AFIS(Amtliche Festpunktinformationssystem) und für ATKIS(Amtliche Topographisch-Kartographische Informationssystem) auf die in dieser Arbeit nicht näher eingegangen wird. (seite 8 geoinfodoc) - Daten werden im XML Format bereit gestellt

- Aufbau der XML Dokumente so einfach wie möglich und so komplex wie nötigseite 87

- Beschreibung der ALKIS Objekte
- wichtigstes Feature AX-Flurstueck: enthält folgende Daten über flurstuecke(noch ergänzen)
- alkis flurstücksdaten werden im Rahmen der Open- Data Lizenz - die Xml Dokumente werden dem Nutzer für bestimmte Schnittstellen bereitgestellt
- die Zugriffsbeschränkungen auf die Schnittstellen erfolgt länderspezifisch
- Veröffentlichung der Alkis Daten unterliegt der "Datenlizenz Deutschland - Namensnennung-Version2.0"<https://www.govdata.de/dl-de/by-2-0>
- Abruf der Daten von Schnittstellen der einzelnen Bundesländer
- Integration der Schnittstellen unterscheidet sich(siehe Tabelle)

- Abruf der Daten über 2 Arten von Geodatendiensten

Es wird generell zwischen einem WMS, Web Map Service und zwischen einem WFS, Web Feature Service unterschieden. Der erstere dient zur Visualisierung von Geodaten. Grundlage dafür sind Vektor und Rasterdaten. Ergebnis einer Anfrage auf einen WMS ist ein Kartenausschnitt. Beim WFS bleibt die volle Struktur der Daten erhalten und die Daten werden in vollem Ausmaß bereit gestellt. Und sind somit zur Weiterverarbeitung geeignet. WFS hat verschiedene Abfrageoptionen: GetCapabilities

Abruf von Fähigkeiten und Metadaten des aufgerufen Dienstes

DescribeFeatureType

abfrage von struktur eines featuretypes

GetFeature

zugriff auf daten

<https://www.adv-online.de/icc/extdeu/binarywriterservlet?imgUid=8f830072-8de8-9221-d5ad-8f138a438ad1&uBasVariant=11111111-11111111>  
geoInfoDoc

[https://www.lgln.niedersachsen.de/download/126790/Basiswissen\\_ALKIS\\_ETRS89\\_Schulungsmaterial\\_Stand\\_12.04.2010.pdf](https://www.lgln.niedersachsen.de/download/126790/Basiswissen_ALKIS_ETRS89_Schulungsmaterial_Stand_12.04.2010.pdf) bAsiswissen  
alkis 10.3

[https://www.bezreg-koeln.nrw.de/brk\\_internet/geobasis/webdienste/anleitung\\_wfs.pdf](https://www.bezreg-koeln.nrw.de/brk_internet/geobasis/webdienste/anleitung_wfs.pdf)

| Bundesland             | Zugriffsart      | Zugriffsbeschränkung | Endpoint                 |
|------------------------|------------------|----------------------|--------------------------|
| NRW                    | WFS              | keine                | siehe Quellenverzeichnis |
| Berlin                 | WFS              | keine                | siehe Quellenverzeichnis |
| Thüringen              | WFS              | auth notwendig       | siehe Quellenverzeichnis |
| Brandenburg            | WFS              | keine                | siehe Quellenverzeichnis |
| Mecklenburg-Vorpommern | WFS              | keine                | siehe Quellenverzeichnis |
| Hamburg                | WFS              | keine                | siehe Quellenverzeichnis |
| Sachsen                | Download         | keine                | siehe Quellenverzeichnis |
| Sachsen-Anhalt         | WFS              | kostenpflichtig      | siehe Quellenverzeichnis |
| Bayern                 | WFS              | auth notwendig       | siehe Quellenverzeichnis |
| Niedersachsen          | WFS              | kostenpflichtig      | siehe Quellenverzeichnis |
| Hessen                 | WFS              | keine                | siehe Quellenverzeichnis |
| Rheinland-Pfalz        | WFS              | kostenpflichtig      | siehe Quellenverzeichnis |
| Schleswig-Holstein     | WFS              | auth notwendig       | siehe Quellenverzeichnis |
| Baden-Württemberg      | WFS              | auth notwendig       | siehe Quellenverzeichnis |
| Bremen                 | nicht auffindbar |                      | siehe Quellenverzeichnis |
| Saarland               | WFS              | auth                 | siehe Quellenverzeichnis |

## 2.3 RDF/RML

- vom W3C erstellt, unterliegt Regeln

<https://www.w3.org/TR/rdf-concepts>

Das Resource Description Framework RDF ist ein universelles, maschinenlesbares Format, das für den Austausch und der Darstellung von Daten gedacht ist. Original wurde es genutzt um Metadaten zu beschreiben (Beleg fehlt noch) und später wurde es in vielen verschiedenen Gebieten und generellen Anwendungen benutzt. Die Daten im RDF Format werden als Graph angeordnet. Ein Graph besitzt Kanten und Knoten. Ein RDF Graph ist eine Ansammlung von Triples, Tripel sind Aussagen über Dinge. Ein Tripel besteht aus einem Subjekt, was konkret eine URI oder eine blank node sein kann, einem Predikat was in dem Tripel die Eigenschaft widerspiegelt. Weiterhin besteht das Tripel aus einem Objekt. Ein Beispiel wäre:

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix ex: <http://example.com/

ex:RobinSeidel a foaf:Person

ex:RobinSeidel ist das Subjekt, a das Predikat und foaf:Person das Objekt.

## 2.4 RDF vs XML

Nun könnte man sich die Frage warum sollte man RDF nutzen, wenn dies ohnehin schon auf XML beruht und die ALKIS Daten schon in diesem universellem Format vorliegen. Was ist der Vorteil von Rdf gegenüber anderen Datenaustauschformaten? <https://www.w3.org/DesignIssues/RDF-XML.html> <https://dbs.uni-leipzig.de/html/seminararbeiten/semSS99/arbeit5/Rdf.html> Eine Aussage, kann in XML durch verschiedene Arten ausgedrückt werden. Für den Menschen der Sie liest bedeuten Sie das Gleiche. Für eine Maschine/Computer sind es verschiedene XML Bäume. ..

## 2.5 RDF Schema, Ontologies

- noch nicht sicher ob notwendig rdf schema: - beschreiben vocabularies beschreibung mit rdf triples - data über data -> metadata - vocabulary, sammlung von standart formaten zur wiederverwertbarkeit (namespace) - gespeichert in owls -> !Unterschied zur Objektorientierung - wenn eine instanz eine bestimmte property einer klasse besitzt, gehört diese instanz zur klasse( umkehrung zu oo) - sehr große macht -> man nutzt standarts des semantic webs, indem man zur Metadata eine Eigenschaft zur Ressource hinzufügt,



dadurch wird diese ressource zum Member einer Klasse( was zuvor nicht der fall war)  
- subclasses existieren owl: - erweiterung zu rdf schema - ontologies beschreiben spezifische Domainen, damit sie unter Fachleuten und INteressenten besser verbreitet und genutzt werden können - .owls sind collection von triples - gibt semantic websemantic also bedeutung - man braucht nicht die ganze bedeutung abzuspeichern da bestimmte key wörter den daten bedeutung geben ontology: - an ontology is a formal, explicit specification of a shared concept - tom grupber 1992 - formale beschreibungssprache um termini einer domäne und deren beziehungen formal zu beschreiben, sodass software diese versteht

## **2.6 RML/Yarrrml**

- RDF Mapping Language
- regeln zum überführen von verschiedenen datenquellen in rdf

## **2.7 Abkürzungen**

- ALKIS: Amtliches Liegenschaftskatasterinformationssystem
- RML: RDF Mapping Language
- RDF: Resource Description Framework
- XML: eXtensible markup language
- ADV: Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland
- URL: Uniform Resource Locator
- URI: Uniform Resource Identifier
- NAS: Normbasierte Austauschchnittstelle

# **3 State of the Art**

## **3.1 Problembeschreibung**

- extract, transform, load problem
- die einzelnen Schritte beschreiben

### 3.2 Möglichkeiten der Transformation

- Alkis mapper selber programmieren
- welche Vorteile und nachteile hätte dies? - RDFMapper beschreiben - yarrml Mapper beschreiben und warum ich ihn nutze

### 3.3 Tripelstores

## 4 Anforderungsanalyse

## 5 Implementierung

- python
- konsolenanwendung
- prototyp

- hier CodeAusschnitte einfügen, um Beispiele für die Beschreibung der Implementierung zu haben

→ Daten sammeln

- HTTP Get an WFS

- Umsetzung mit HTTP Libray Requests"geschrieben in python

ODER für ALKIS Sachsen mit URLLIB, wird genutzt um beliebige URL aufzurufen und deren response zu downloaden

→ Transformation

- Auswahl welche XML Datei umgewandelt werden soll

- ausgewählte Datei wird in yarrmlRules Datei eingetragen

- es erfolgt die Erstellung von RML Regeln mithilfe yarrml Parser und selbstgeschriebener yarrml Regeln

- Aufruf RDFMapper der mit RML Regeln, die Objekte in der ausgewählten xml datei in RDF im .ttl Format umwandeln

→ Daten Speichern( )

- die produzierten rdf tripel werden in fuseki graph db gespeichert

### 5.1 genutzte Technologien

**RDF Triplestore** Was ist eine Graph Store?

Welchen store sollte ich nutzen?

neo4j vs graphDB

neo4j: graph native, index free, fast predictable preformance

- highly dynamic

<https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-triplestore/>

<https://www.youtube.com/watch?v=eDRfKAKZMQE>

<https://db-engines.com/en/system/GraphDB>

## **5.2 Funktionsweise**

- github mit howto zur selbstaussführung
- beschreibung der Transformation mit codeausschnitten

## **6 Ergebnisse und Evaluation**

- ergebnis mit beispielhaften SPARQL abfragen
- da alkis data auffindbarkeit und nutzbarkeit mangelhaft
- arbeit hat sich zum ziel gemacht dies zuverbessern

## **7 Ausblick**

- weitere Alkis Quellen
- benutzermanagment
- anreicherung der rdf daten

## **8 notizen**

- gml:id ist dokumentenweit eindeutig(durch entstehungsdatum-zeit)(quelle geoinfodoc s92)

## 9 Quellenverzeichnis

Quellen für Alkis Endpunkte der Bundesländer:

NRW: [https://www.wfs.nrw.de/geobasis/wfs\\_nw\\_alkis\\_vereinfacht?Service=WFS&Request=GetCapabilities](https://www.wfs.nrw.de/geobasis/wfs_nw_alkis_vereinfacht?Service=WFS&Request=GetCapabilities)

Berlin: [https://fbinter.stadt-berlin.de/fb/wfs/data/senstadt/s\\_wfs\\_alkis?Service=WFS&Request=GetCapabilities](https://fbinter.stadt-berlin.de/fb/wfs/data/senstadt/s_wfs_alkis?Service=WFS&Request=GetCapabilities)

Thüringen: [https://fbinter.stadt-berlin.de/fb/wfs/data/senstadt/s\\_wfs\\_alkis?Service=WFS&Request=GetCapabilities](https://fbinter.stadt-berlin.de/fb/wfs/data/senstadt/s_wfs_alkis?Service=WFS&Request=GetCapabilities)

Brandenburg: [https://isk.geobasis-bb.de/ows/alkis\\_vereinf\\_wfs?Service=WFS&Request=GetCapabilities](https://isk.geobasis-bb.de/ows/alkis_vereinf_wfs?Service=WFS&Request=GetCapabilities)

Mecklenburg-Vorpommern: [https://www.geodaten-mv.de/dienste/alkis\\_wfs\\_sf](https://www.geodaten-mv.de/dienste/alkis_wfs_sf)

Hamburg: [https://geodienste.hamburg.de/WFS\\_HH\\_ALKIS\\_vereinfacht?REQUEST=GetCapabilities&SERVICE=WFS](https://geodienste.hamburg.de/WFS_HH_ALKIS_vereinfacht?REQUEST=GetCapabilities&SERVICE=WFS)

Sachsen:

Sachsen-Anhalt:

Bayern: [https://geoservices.bayern.de/wfs/v1/ogc\\_alkis\\_ave.cgi?](https://geoservices.bayern.de/wfs/v1/ogc_alkis_ave.cgi?)

Niedersachsen: <https://numis.niedersachsen.de/www.lgl.niedersachsen.de?SERVICE=WFS&REQUEST=GetCapabilities>

Hessen: <https://www.gds.hessen.de/wfs2/aaa-suite/cgi-bin/alkis/vereinf/wfs?Service=WFS&Request=GetCapabilities>

Rheinland-Pfalz: <https://www.geoportal.rlp.de/spatial-objects/353>

Schleswig-Holstein: [https://www.schleswig-holstein.de/DE/Landesregierung/LVERMGEOSH/lvermgeosh\\_node.html?REQUEST=GetCapabilities&SERVICE=WFS](https://www.schleswig-holstein.de/DE/Landesregierung/LVERMGEOSH/lvermgeosh_node.html?REQUEST=GetCapabilities&SERVICE=WFS)

Baden-Württemberg:

Bremen:

Saarland: <https://geoportal.saarland.de/registry/wfs/325?REQUEST=GetCapabilities&VERSION=1.1.0&SERVICE=WFS>

- 4.2.22

zu einleitungserklärungen

- neo4j vs graph db: