



Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Interfakultären Fachbereich für GeoInformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

„Vergangenheit hat Zukunft“ Das Historische Häuserbuch von Mainz im Semantic Web

vorgelegt von

Sebastian Bayer
104908, UNIGIS MSc Jahrgang 2017

Betreuer/in:
Prof. Josef Strobl

Zur Erlangung des Grades
„Master of Science – MSc“

Mainz, 22.02.2022

Zusammenfassung

Das erfolgreiche Vorhaben Linked Open Data (LOD) hat einen wichtigen Beitrag zur Verwirklichung der Vision vom Semantic Web geleistet und die nächste Evolutionsstufe des Webs eingeläutet. Die LOD-Initiative bezweckt die umfassende Informationsintegration durch die Kreierung einer gigantischen globalen „Graphdatenbank“, die als LOD-Cloud oder einfach als das Web (of Data) bezeichnet wird. Die LOD-Cloud umfasst Informationen – Daten, denen eine Bedeutung (Semantik) zugeordnet wurde. Es bedarf einer Handlungsvorschrift, damit unterschiedliche Informationen aus verschiedenen Quellen in einem flexiblen, graphenbasierten Modell einheitlich repräsentiert und zusammengeführt werden können. Die LOD-Methode ist eine Handlungsvorschrift, welche auf vier einfachen Regeln – den LOD-Prinzipien – beruht. Sie zielt auf die Integration strukturierter Informationen ins Web durch deren Verlinkung ab. Dadurch können die Informationen in unvorherzusehenden Anwendungsszenarien wiederverwendet werden und es lässt sich ein Informationsmehrwert erzielen – „The value of your own information is very much a function of what it links to [...]“ (Berners-Lee, 2006).

Die vorliegende Arbeit konzentriert sich auf die Integration von *Geoinformationen* in die LOD-Cloud. Diese spielen im Web eine zentrale Rolle, da die meisten verfügbaren Informationen einen Raumbezug aufweisen. Es wird eine bestimmte Art von vektorialen *Geodaten* fokussiert, die in einem proprietären Format gespeichert wurden. Anhand von MapInfo-Relationen aus dem historischen Häuserbuch von Mainz wird untersucht, ob die LOD-Methode auch angewendet werden kann, um den Informationsumfang von Modellen zu erhöhen, die keine konzeptuelle Grundlage aufweisen, die nur aus vektorialen Geodaten bestehen, welche ausschließlich in Form von nichtnormalisierten Relationen vorliegen (Leitfrage). Eigentlich zielt die LOD-Methode auf die Integration *strukturierter* Informationen ins Web durch deren Verlinkung ab. In dieser Arbeit besteht die Herausforderung darin, dies auf Basis der Sachdaten (thematische Eigenschaften der Geoobjekte) zu realisieren, die aus einer Sammlung *unstrukturierter* Zeichenketten (mehrwertige und zusammengesetzte Attribute) bestehen.

Um die Leitfrage zu beantworten, werden die (LOD-)Anforderungen herausgearbeitet und definiert, denen der resultierende Graph genügen muss oder soll (erstes Teilziel). Das zweite Teilziel besteht in der Konzeption und in der praktischen Umsetzung eines Workflows zwecks Generierung des LOD-konformen Graphen aus den oben charakterisierten Relationen in mehreren aufeinanderfolgenden Schritten (Transformation, Konzeptualisierung, Konsolidierung, Anreicherung bzw. Verlinkung). Im Rahmen des Transformations- und Konzeptualisierungsprozesses wird eine anwendungsfallsspezifische Lösung erarbeitet, welche auf die Identifikation von Entitäten via Uniform Resource Identifiers (URIs) sowie auf die Beschreibung von Ressourcen unter Einsatz des Resource Description Frameworks (RDF) und mithilfe darauf aufbauender Informations-Spezifikations-sprachen abzielt. Nach der Konsolidierung der RDF-Tripel in Gestalt benannter RDF-Graphen, wird der Informationsumfang durch die Erzeugung von RDF-Links auf der Instanzebene erhöht. Hierfür wird eine übertragbare Verfahrensweise vorgeschlagen, die der Ermittlung und Verknüpfung von URI-Aliasen oder URIs, welche auf ein ähnliches Individuum referieren, dient. Die Praktikabilität des Verfahrens wird im Rahmen des Verlinkungsprozesses unter Beweis gestellt, indem ausgewählte Ressourcen aus dem historischen Häuserbuch mit Ressourcen der Deutschen Nationalbibliothek (DNB) verknüpft werden (Titeldaten, Gemeinsame Normdatei). Ob und in welchem Grad der resultierende RDF-Graph den definierten (LOD-)Anforderungen genügt, wird aus der Sicht einer Datenbereiterin überprüft. Aus dieser Perspektive wird auch die Effizienz der vorgenannten Verfahrensweise kritisch beurteilt. Außerdem wird ein fiktives Anwendungsszenario konstruiert und es wird ein Informationssystem aufgebaut, damit der erhöhte Informationsumfang bzw. das erhöhte (Nach-)Nutzungspotenzial aus der Sicht einer Anwenderin demonstriert werden kann (drittes Teilziel). In diesem Zuge wird auch klargestellt, was unter dem unspezifischen Begriff „Informationsumfang“ zu verstehen ist.

Generell wird besonderer Wert auf die Verständlichkeit gelegt. Neben der Leitfrage zieht sich ein „Schlüsselbeispiel“ wie ein roter Faden durch gesamte Arbeit. Sie richtet sich an Datenbereiterinnen und Anwenderinnen aus der Praxis, die ihre Ressourcen in die LOD-Cloud integrieren oder LOD nachnutzen möchten.

Inhaltsverzeichnis

Zusammenfassung	II
Inhaltsverzeichnis	III
Abkürzungsverzeichnis	VI
Abbildungsverzeichnis	XII
Tabellenverzeichnis	XV
Listingverzeichnis	XVI
1 Einleitung	1
1.1 Motivation, Problemstellung	2
1.2 Zielsetzung, Forschungsfrage(n)	3
1.3 Hintergrund, Zielgruppe	5
1.4 Inhalte, Strukturierung	5
2 Theoretische Grundlagen	8
2.1 Web of Documents	8
2.1.1 Entwicklung, Standards	8
2.1.2 Terminologie	10
2.1.3 Web-Architekturprinzipien	15
2.2 Web of Data	18
2.2.1 Intention, Abgrenzung	19
2.2.2 Wissensrepräsentation	22
2.2.3 RDF – Grundlagen	27
2.2.4 Erweiterte Terminologie	32
2.2.5 RDF – Web-Architekturprinzipien	36
2.3 Geo Web	38
2.3.1 Standards, Geodateninfrastrukturen	38
2.3.2 Downloaddienst: Basic WFS	39
2.3.3 Grundlagen der Datenmodellierung	41
2.3.4 Moderne Systeme, OGC-Standards	44
2.4 Semantic Geospatial Web	45
2.4.1 Semantic Enablement, Uplift, Downlift	46
2.4.2 Empfehlungen, Standards	47
2.5 Zusammenfassung	51

3	Aktueller Stand der Forschung	56
3.1	Verfahrensweisen	58
3.1.1	Generierung von LGCHD	58
3.1.2	URI-Allokation	60
3.1.3	Konzeptualisierung	61
3.1.4	Verlinkung auf der Instanzebene	63
3.2	Hilfsmittel.....	65
3.2.1	RDF-izer.....	65
3.2.2	WFS-Adapter	69
3.2.3	Frameworks	70
4	Kurzportrait: Historisches Häuserbuch	74
4.1	Hintergrundinformationen und Kartengrundlagen	74
4.2	Rasterdarstellungen und Modelleigenschaften	76
4.3	Verwendungsbeispiel aus der Praxis	80
5	Datenuntersuchung, -auswahl, -anpassung	81
5.1	Historisches Häuserbuch	81
5.1.1	Vektormodell	81
5.1.2	Auswahl der Individuen zwecks Verlinkung.....	94
5.2	LOD-Cloud	94
6	Anforderungsanalyse und -definition	97
6.1	Diskussion der SDWWG-Empfehlung	98
6.1.1	URI-Kollision.....	98
6.1.2	Feature	101
6.2	LOD-Anforderungen.....	103
6.2.1	Allgemeine Anforderungen	103
6.2.2	Individuelle Anforderungen	105
6.3	Anwendungsfallspezifische Anforderungen	106
6.4	Anwendungsszenario	107
7	Lösungsansatz	110
7.1	Entwurf des Systems	110
7.1.1	Infrage kommende Applikationen.....	113
7.1.2	Entscheidungen	116

7.2	Konzeption des Workflows	120
7.2.1	Transformation.....	120
7.2.2	Konzeptualisierung	122
7.2.3	Konsolidierung.....	125
7.2.4	Anreicherung.....	125
7.2.5	Gesamter Workflow.....	131
8	Praktische Umsetzung	132
8.1	Aufbau des Systems	132
8.2	Generierung des Graphen.....	133
8.2.1	Transformation.....	133
8.2.2	Konzeptualisierung	149
8.2.3	Konsolidierung.....	159
8.2.4	Anreicherung.....	160
9	Diskussion und Demonstration	184
9.1	Anforderungen.....	184
9.2	Anwendungsszenario	187
9.3	Beantwortung der Leitfrage und Fazit	195
9.4	Übertragbarkeit	197
10	Zusammenfassung und Ausblick	198
10.1	Zusammenfassung.....	198
10.2	Ausblick	200
	Literaturverzeichnis	202
	Anhang	212

Abkürzungsverzeichnis

Themenspezifisch

AgRelOn	Ontology of agent relationships
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
AVX	Advanced Vector Extensions
AWWW	Architecture of the World Wide Web
BIBO	Bibliographic Ontology
BLOBs	Binary Large Objects
CC	Creative Commons
CDV	Kultur-Hackathon Coding da Vinci
CLI	Command-line interface
CMS	Content-Management-System
CPU	Central processing unit
CQL	OGC Common query language
CRM	Conceptual Reference Model
CRS	Coordinate reference system
CSW	Catalogue Service for the Web / Web Catalogue Service
CURIEs	Compact URIs
cURL	Client for URLs
CWA	Closed World Assumption
DB	Database
DCMI	Dublin Core Metadata Initiative
DDL	Data Definition Language
DHDN90	Deutsches Hauptdreiecksnetz 1990
DL	Description logics
DNB	Deutsche Nationalbibliothek
DNS	Domain Name System
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
DQ	Datenqualität
ESRI	Environmental Systems Research Institute
ET	Entitätstyp
ETL	Extract, Transform, Load
ETRS89	European Terrestrial Reference System 1989
FCOs	First Class Objects
FD	False Duplicate

FE	Filter Encoding
FME	Feature Manipulation Engine
FOAF	Friend of a Friend
FS	Filesystem
FU	False Unique
GDAL	Geospatial Data Abstraction Library
GDI	Geodateninfrastruktur
GFM	General Feature Model
GI	Geoinformatik
GIS	Geo-Informationssystem
GIScience	Geographic Information Science
GK3	Gauß-Krüger-Abbildung im 3°-Streifensystem
GML	Geography Markup Language
GND	Gemeinsame Normdatei
GSW.IE	Geospatial Semantic Web Interoperability Experiment
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IDE	Integrated development environment
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
INSPIRE	INfrastructure for Spatial Information in Europe
IoT	Internet of Things
IPS	Implementierungsspezifikation
IQ	Informationsqualität
IR	Inferenzregel
IRE	Identity of Resources and Entities on the web
IRW	Identity and Reference on the Web
IS	Informationssystem
ISA	Interoperability Solutions for European Public Administrations
ISO	International Organization for Standardization
IT	Informationstechnik
JSON	JavaScript Object Notation
JTS	Java Topology Suite
KS	Klassenüberschneidung

KV	Klassenvereinigung
KVP	Key-Value-Paare
LBauO	Landesbauordnung Rheinland-Pfalz
LD	Linked Data
LGCHD	Linked Geospatial Cultural Heritage Data
LGD	Linked Geospatial (Open) Data
LOD	Linked Open Data
LSDI	Linked Spatial Data Infrastructure
MIF/MID	MapInfo Data Interchange-Format
MUR	Minimal umgebendes Rechteck
MZ	Mainz
N3	Notation 3
NIC	Network interface card
NIR	Nicht-Informationsressource / Non-Information Resource
NUNA	Non-Unique Name Assumption
OAI	Open Archives Initiative
OAI-PMH	Open Archives Protocol for Metadata Harvesting
OAPI	OGC Application Programming Interface
OGC	Open Geospatial Consortium, früher Open GIS Consortium
OGR	OGR Simple Features Library
OO	Objektorientierung
ORDBMS	Object-Relational Database Management System
OS	Operating system
OWA	Open World Assumption
OWL	Web Ontology Language
OWS	OpenGIS Web Service / OGC Webservices
PDF	Portable Document Format
PK	Primary key
R2RML	RDB to RDF Mapping Language
RAID	Redundant array of independent disks
RAM	Random-Access Memory
RDB	Relational database
RDBMS	Relational Database Management System
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RDIMM	Registered Dual Inline Memory Module

REST	Representational State Transfer
RFC	Request for Comments
RLP	Rheinland-Pfalz
ROA	Resource-oriented architecture
SDWWG	Spatial Data on the Web Working Group
SFA	Simple Feature Access
SGW	Semantic Geospatial Web
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organisation System
SLES	SUSE Linux Enterprise Server
SMB	Server Message Block
SO	Similarity Ontology
SOA	Service-oriented architecture
SOAP	„Simple Object Access Protocol“; SOAP ist kein Akronym mehr.
SP	Service Pack
SPARQL	SPARQL Protocol And RDF Query Language
SQL	„Structured Query Language“; eigentlich ist SQL ein eigenständiger Name.
SRS	Spatial reference system
SSD	Solid-State-Drive
SSW	Semantic Sensor Web
Sta	Stadtarchiv
SVG	Scalable Vector Graphics
SWEET	Semantic Web for Earth and Environmental Terminology
SWS	Semantic Web Services
Sync.	Synchronisation
TAG	Technical Architecture Group
TC	Technical Committee
UCS	Universal Coded Character Set
UKB	Unterklassenbeziehung
UML	Unified Modeling Language
UNC	Uniform Naming Convention
UoD	Universe of Discourse
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
UTC	Coordinated Universal Time

UTF-8	8-Bit UCS Transformation Format
UTM	Universal Transverse Mercator
V.	Version
VAD	Virtuoso Application Distribution
VGI	Volunteered Geographic Information
VM	Virtuelle Maschine
voiD	Vocabulary of Interlinked Datasets
VOS	OpenLink Virtuoso, Open-Source Edition
W3C	World Wide Web Consortium
WE	Wurzelement
WFS	Web Feature Service
WFS-T	Transaction WFS
WGS	World Geodetic System
WKT	Well-known text
WMS	Web Map Service
WWW	World Wide Web
XLink	XML Linking Language
XML	Extensible Markup Language
XPath	XML Path Language
XQuery	XML Query Language
XSD	XML Schema Definition
XSLT	XSL Transformation

Allgemein

Abb.	Abbildung
bzw.	beziehungsweise
d.h.	das heißt
ED	Eigene Darstellung
EK	Eigene Konfiguration
engl.	englisch
et al.	et aliae
etc.	et cetera
ggf.	gegebenenfalls
GP	Gliederungspunkt
i.d.R.	in der Regel
i.V.m.	in Verbindung mit
Lis.	Listing
o. Ä.	oder Ähnliche
o.g.	oben genannt
OA	Originalabbildung
p.	page
resp.	respektive
s.	siehe
s.o.	siehe oben
s.u.	siehe unten
sog.	sogenannt
Tab.	Tabelle
u. a.	unter anderem
usw.	und so weiter
uvm.	und viele mehr
vgl.	vergleiche
z.B.	zum Beispiel

Hinweise

- In dieser Arbeit wird das generische Femininum verwendet. Ausnahmen bestätigen die Regel.
- Daraus folgt: die URL (statt der URL) sowie die URI (statt der URI).
- Links werden wie folgt dargestellt: interner [Link](#), externer [Link](#).
- (Semantic Web) URIs werden im Fließtext mithilfe von **ConsoLas** repräsentiert.
- Schlüsselwörter werden mithilfe von KAPITÄLCHEN hervorgehoben.

Abbildungsverzeichnis

Abb. 1.1: Struktur der Thesis, ED	7
Abb. 2.1: Standards des Web of Documents, OA von Allemang et al. (2020, p. 86)	8
Abb. 2.2: Spaltung des URI-Konzepts – Differenzierung zwischen Zugang und Referenz, ED	9
Abb. 2.3: Information, Kodierung, Inhalt, OA von Halpin (2013, p. 25), angepasste Darstellung	11
Abb. 2.4: URI-Syntax, ED	12
Abb. 2.5: „Inhaltsvereinbarung“, OA von Halpin (2013, p. 31), angepasste Darstellung.....	14
Abb. 2.6: Zustand des Web of Documents aus menschlicher Sicht, OA von Berners-Lee (1994b).....	18
Abb. 2.7: Zustand des Web of Documents aus der Sicht eines Computers, OA von Berners-Lee (1994b)	18
Abb. 2.8: Aufgabenerledigung im Semantic Web, OA von Berners-Lee (1994b)	20
Abb. 2.9: 5 Sterne für LOD, OA von Berners-Lee (2006).....	21
Abb. 2.10: Standards im Web of Documents / Web of Data, OA von Allemang et al. (2020, p. 86)	21
Abb. 2.11: Wissenspyramide, OA von Bill (2016, p. 363), angepasste Darstellung.....	23
Abb. 2.12: Ausgewählte Wissensrepräsentationssprachen, ED	25
Abb. 2.13: Ressourcen & Beziehungen, OA von Halpin (2013, p. 51), angepasste Darstellung.....	27
Abb. 2.14: Netzwerk-Effekt in der LOD-Cloud, Graphiken von lod-cloud.net	28
Abb. 2.15: Ein RDF-Tripel, OA von Hitzler et al. (2008, p. 36), angepasste Darstellung.....	28
Abb. 2.16: RDF-Links in RDF-Tripeln, ED	29
Abb. 2.17: Ein Literal an Objektstelle, ED	29
Abb. 2.18: Ausgewählte RDFS-Plus-Sprachmittel, ED	30
Abb. 2.19: Auszugsweise, graphische Darstellung der IRW, OA von Halpin and Presutti (2009, p. 4)	34
Abb. 2.20: 303-Weiterleitung, OA von Halpin and Presutti (2009, p. 3), angepasste Darstellung	34
Abb. 2.21: „IVB“ nach HTTP 303, OA von Halpin and Presutti (2009, p. 33), angepasste Darstellung	35
Abb. 2.22: Kommunikation mit einem WFS, OA von Vretanos (2002, p. 9)	39
Abb. 2.23: Operationsfolgen beim Basic WFS, OA von Staub (2009, p. 57), angepasste Darstellung.....	40
Abb. 2.24: Klassenhierarchie des SFA-Geometriemodells, OA von Ryden (2005, p. 11)	42
Abb. 2.25: GeoSPARQL-Konzepte, OA von Tschirner (2012, p. 43), angepasste Darstellung	49
Abb. 2.26: Eine Domänenontologie mit GeoSPARQL-Konzepten, OA von Hiebel et al. (2015, p. 5)	50
Abb. 3.1: Linked Spatial Data Infrastructure, OA von López-Pellicer et al. (2015, p. 10)	56
Abb. 3.2: LOD-Lifecycle, OA von Auer and Lange (2014, p. 8), angepasste Darstellung	57
Abb. 3.3: Workflow zwecks Generierung von LGCHD, OA von Nishanbaev et al. (2020, p. 885)	59
Abb. 3.4: Hilfsmittel für die Generierung von LGCHD, OA von Nishanbaev et al. (2019, p. 1482)	59
Abb. 3.5: Klassenhierarchie der Historicmapshen-Ontologie, OA von Scheider et al. (2014, p. 9)	62
Abb. 3.6: Graphischer Auszug aus der SO, OA von Halpin et al. (2010, p. 312)	64

Abb. 3.7: Umformung von Geometriedaten mit Geometry2RDF, OA unverändert.....	65
Abb. 3.8: Extraktion, Transformation, Laden mit TripleGeo, OA von Patroumpas et. al. (2014, p. 275)	66
Abb. 3.9: Architektur von GeoTriples, OA unverändert	67
Abb. 3.10: Funktionsweise des modularen LOD4WFS-Adapters, OA von Jones et al. (2014, p. 9)	69
Abb. 3.11: Generierung und Publikation von LGD, OA von Vilches-Blázquez and Saavedra (2019, p. 22) ...	71
Abb. 3.12: Beispiel zur Evaluation in der Silk Workbench, OA von Isele (2013, p. 179)	73
Abb. 3.13: Der typische Workflow für die Erstellung einer Link-Spezifikation, OA unverändert.....	73
Abb. 4.1: Mainz um 1620, erstellt von Falck (1972)	74
Abb. 4.2: „Schwedenplan“ (Sta Mainz, BPSP / 55 D).....	75
Abb. 4.3: Auszug aus dem Stadtplan von 1871 (Sta Mainz, BPSP / 489 D)	75
Abb. 4.4: Mainz um 1450 (gerendert von MapServer V. 7, dargestellt von Mapbender V. 3).....	76
Abb. 4.5: Mainz um 1620 (gerendert von MapServer V. 7, dargestellt von Mapbender V. 3)	77
Abb. 4.7: Mainz um 1871 (gerendert von MapServer V. 5, dargestellt von Mapbender V. 2)	78
Abb. 4.8: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts, OA unverändert	79
Abb. 4.9: Topographische Objekte um 1450, OA von Münkel (2019, p. 8).....	80
Abb. 5.1: Übersicht zur Datenbasis, ED	84
Abb. 5.2: Beispiel zur Datenerfassung in ArcGIS Pro, ED	88
Abb. 5.3: Abgleich von Objekten in der Relation mit der Karte von Falck (1978)	89
Abb. 5.4: Aggregation als Kompromiss, Kartenausschnitt aus Falck (1978)	89
Abb. 5.5: SVG-Export des ArcGIS Pro-Layouts (1450)	91
Abb. 5.6: SVG-Export des ArcGIS Pro-Layouts (1620)	92
Abb. 5.7: SVG-Darstellung, generiert über FME (1871)	93
Abb. 6.1: Exemplarische Dereferenzierung, dargestellt von PuTTY	99
Abb. 7.1: Entwurf des Systems (Teil 1), ED	110
Abb. 7.2: Entwurf des Systems (Teil 2), ED	111
Abb. 7.3: Entwurf des Systems (Teil 3), ED	112
Abb. 7.4: Entwurf des Systems (letzter Teil), ED	117
Abb. 7.5: Sync. via OAI-PMH, OA von Diederichs and Wuttke (2018, p. 7), angepasste Darstellung	119
Abb. 7.6: Verfahrensweise zwecks Beantwortung der Leitfrage, ED	120
Abb. 7.7: Verfahrensweise für die Generierung von RDF-Links auf der Instanzebene, ED	128
Abb. 7.8: Konzeption des Workflows zwecks Generierung des RDF-Graphen, ED	131
Abb. 8.1: FME-Workbench (Teil 1), EK	136
Abb. 8.2: FME-Workbench (Teil 2), EK	137

Abb. 8.3: Parameter der beiden Reprojector-Transformer, EK	138
Abb. 8.4: Parameter der beiden GeometryValidator-Transformer, EK.....	138
Abb. 8.5: Protokollierung geometrischer Probleme, EK.....	139
Abb. 8.6: Fehlerhafte Orientierung der Koordinaten, Feature Kontrolle.....	139
Abb. 8.7: Parameter der drei Tester-Transformer, EK.....	140
Abb. 8.8: Parameter der drei AttributeSplitter-Transformer, EK.....	140
Abb. 8.9: Verarbeitung unterschiedlicher Listenstrukturen, EK	141
Abb. 8.10: Vorbereitung der Personentabelle, EK	142
Abb. 8.11: Literaturtabelle, EK.....	142
Abb. 8.12: Beispiele für FD/FU, EK.....	143
Abb. 8.13. Finale Attribut-Manipulation, EK	144
Abb. 8.14: Konsolidierung mittels komplexer Listen, EK	144
Abb. 8.15: WKT-Export, EK	146
Abb. 8.16: Klassenhierarchie, EK, dargestellt von OntoGraf.....	150
Abb. 8.17: Workbench für die Verlinkung der Literaturangaben, EK	165
Abb. 8.18: Kandidatenpaare für die Verlinkung der Literatur, mit Details zu S19, EK.....	166
Abb. 8.19: Beispiel für ein mehrteiliges Werk, EK	167
Abb. 8.20: Workbench für die Verlinkung der Menschen, EK.....	171
Abb. 8.21: Beispiel für Namensvarianten, EK	173
Abb. 8.22: Kombination von Jaccard und Soft-Jaccard, EK.....	173
Abb. 8.23: Beispiel für einen fehlenden Namenszusatz, EK	174
Abb. 8.24: Workbench für die Verlinkung der Organisationen, EK	175
Abb. 8.25: Beispiel für ähnliche Namen von Organisationen, EK	175
Abb. 8.26: Kandidatenpaare für die Verlinkung der Menschen, mit Details zu A331, EK.....	176
Abb. 8.27: Kandidatenpaare für die Verlinkung der Organisationen, mit Details zu A281, EK	177
Abb. 8.28: Darstellung der Ausgangsbasis und der Ergebnisse (Verlinkung der Agenten), ED	180

Tabellenverzeichnis

Tab. 2.1: Die Semantik ausgewählter HTTP-Statuscodes im Semantic Web	33
Tab. 5.1: Eigenschaften der „Häuser-Relationen“	85
Tab. 5.2: Gemeinsame Struktur und individuelle Eigenschaften der übrigen Relationen für 1450	86
Tab. 5.3: Relationsstruktur und Eigenschaften der übrigen Objekte für 1620	86
Tab. 5.4: Gemeinsame Struktur und individuelle Eigenschaften der angepassten Relationen für 1450	90
Tab. 5.5: Relationsstruktur und Eigenschaften der übrigen Objekte für 1620	90
Tab. 5.6: Infrage kommende Wissensbasen für die Verlinkung von Individuen	95
Tab. 5.7: Infrage kommende Vokabulare für den Konzeptualisierungs- und Transformationsprozess.....	96
Tab. 8.1: Relevante Spezifikationen (VM2).....	132
Tab. 8.2: Relevante Applikationen auf VM1 und VM2.....	132
Tab. 8.3: Übersicht zu den verwendeten FME-Elementen.....	135
Tab. 8.4: Vergleich der Listenstrukturen	141
Tab. 8.5: Verknüpfung von Gebäuden mit Eigentümer, Bewohner und Literatur	145
Tab. 8.6: Attribute der Objekte, die Gebäude repräsentieren.....	146
Tab. 8.7: Attribute der restlichen Objekte.....	146
Tab. 8.8: Überlegungen zur URI-Allokation	149
Tab. 8.9: Das „Nutzungsproblem“	151
Tab. 8.10: Gebäudetypen	152
Tab. 8.11: Details zum verwendeten Grundbestand	159
Tab. 8.12: Informationen zu den Prädikaten rdau:P60327 und rdau:P60493, Quelle: RDA Registry	162
Tab. 9.1: Informationen zu den Bursen (Antworten auf die Fragen aus Gruppe 1).....	189
Tab. 9.2: Informationen für biografischen Skizzen der Bewohner (zu den Fragen 2.1, 2.2, 2.4 - 2.6.)	190

Listingverzeichnis

Lis. 8.1: WE mit Namensraum-Deklaration und Prozessierungsanweisungen (XML/XQuery), EK	147
Lis. 8.2: Sub-Template BUILDING (XML/XQuery), EK.....	147
Lis. 8.3: Property mz:hadOwner (RDF/XML), EK.....	156
Lis. 8.4: Property mz:hadOccupant (RDF/XML), EK.....	157
Lis. 8.5: Property mz:wasUsedBy (RDF/XML), EK	157
Lis. 8.6: Property mz:where (RDF/XML), angelehnt an Scheider et al. (2014)	158
Lis. 8.7: Property mz:when (RDF/XML), EK.....	158
Lis. 8.8: SPARQL-Abfrage zwecks Selektion und Erläuterung der Geoobjekte, EK.....	181
Lis. 9.1: Festlegungen für SPARQL (Select/Construct), EK	187
Lis. 9.2: Exemplarische SPARQL-Abfragen zwecks Vorbereitung, EK	188
Lis. 9.3: SPARQL-Abfrage (Fragen aus Gruppe 1), EK	189
Lis. 9.4: SPARQL-Abfrage (Fragen aus Gruppe 2), EK	190
Lis. 9.5: SPARQL-Konstruktion eines Graphen (Fragen 3.1 und 3.4), EK.....	191
Lis. 9.6: Antworten auf die Fragen 3.1 und 3.4 (RDF/XML), generiert von VOS	191
Lis. 9.7: SPARQL-Abfrage zwecks Ergänzung (Fragen aus Gruppe 3), EK.....	192
Lis. 9.8: Weitere Antworten auf die Fragen aus Gruppe 3 (JSON), generiert von VOS.....	193

1 Einleitung

Als die Stadtverwaltung Mainz Anfang der Neunzigerjahre beschloss, Geo-Informationssysteme (GIS) einzusetzen, eröffneten sich dem Mainzer Stadtarchiv neue Möglichkeiten, den reichhaltigen Schatz an raumzeitlichen Informationen zu digitalisieren und zukunftsorientiert zu publizieren. Dieses Potenzial sollte auch im Projekt „Historisches Häuserbuch von Mainz“ ausgeschöpft werden, um die bewegte Geschichte der Stadt modellhaft darzustellen. In diesem Zuge wurden digitale Vektormodelle entwickelt, die Informationen zur Topographie, über Personen und Ereignisse verschiedener Zeitabschnitte enthielten. Als Grundlagen dienten historische Quellen, z.B. analoge Karten, die mithilfe von GIS digitalisiert und georeferenziert wurden. Das Mainzer Stadtarchiv verfolgte das Ziel, allen Interessierten einen einfachen und freien Informationszugang zu ermöglichen. Die Charakteristika monolithischer GIS stellten eine Hürde für die Verwirklichung dieses Ziels dar.

Mitte der Neunzigerjahre konstituierte sich das heutige Open Geospatial Consortium (OGC) unter dem Namen Open GIS Consortium, dessen Standardisierungsbestrebungen zunächst die **syntaktische Interoperabilität** von GIS adressierten. Infolgedessen evolvierten viele GIS sukzessive zu modularen Applikationen mit offenen Schnittstellen. Diese Entwicklung wurde auch von den Standards der International Organization for Standardization (ISO) vorangetrieben, die essenzielle Konzepte normierte, insbesondere zwecks einheitlicher Modellierung räumlicher Informationen. Der effiziente Datenaustausch zwischen GIS konnte auf Basis etablierter Internet- und Webtechnologien realisiert werden, die primär von der Internet Engineering Task Force (IETF) und dem World Wide Web Consortium (W3C) standardisiert wurden. Das Rahmenwerk aus De-facto- und De-jure-Standards begünstigte die erfolgreiche Implementierung webbasierter GIS, das rasante Wachstum öffentlich verfügbarer Geoinformationen und -dienste sowie die Verwirklichung von dezentralen Geodateninfrastrukturen (GDI), die in ihrer Gesamtheit im Folgenden unter dem Begriff „**Geo Web**“ subsumiert werden.

Von dieser Entwicklung profitierte auch das Projekt „Historisches Häuserbuch von Mainz“. Der Datenaustausch mittels proprietärer Formate auf Sekundärspeichern wurde ab 2003 zunehmend überflüssig, da eine Webanwendung unter Nutzung des OGC-Standards Web Map Service (WMS) realisiert werden konnte. Das historische Häuserbuch stand fortan allen Interessierten als Rastermodell zur Verfügung. Zwecks Interaktion wurde eine grafische Benutzerschnittstelle (GUI) implementiert und der WMS wurde ein Bestandteil der wachsenden GDI der Stadt Mainz. Die Applikation wurde zwar weiterentwickelt und an die Anforderungen moderner Klienten angepasst; der webbasierte Zugriff auf die Vektormodelle ist allerdings immer noch nicht möglich. Da jene Modelle immer wieder im Fokus wissenschaftlicher Arbeiten und kreativer Projekte stehen, ist es naheliegend die GDI der Stadt Mainz um einen weiteren OGC-konformen Web Feature Service (WFS) zu bereichern. Dieser bewährte Downloaddienst bietet als OpenGIS Web Service (OWS) standardisierte Operationen für einen interoperablen Zugriff auf die Vektormodelle – zumindest in struktureller und syntaktischer Hinsicht.

Im Rahmen der interdisziplinären Projektarbeit, rückten jedoch auch **semantische Aspekte der Interoperabilität** in den Fokus und das Vorhaben erhielt neue Impulse. Der Austausch mit verschiedenen Akteurinnen im geo- bzw. kulturwissenschaftlichen Bereich, die ihre Informationen bereits ins World Wide Web (WWW) integrieren, führte zu wichtigen Erkenntnissen. Einerseits besteht Konsens hinsichtlich der grundlegenden Intention, der Öffentlichkeit einen webbasierten Zugang zu frei lizenzierten Informationen bieten zu wollen, andererseits werden unterschiedliche Methoden angewandt, um dieses Ziel zu erreichen. Neben OWS und Modellierungsansätzen des Geo Web spielen auch Methoden, Prinzipien und Technologien des **Semantic Web** eine zentrale Rolle. Sie dienen dazu, die Vision von der nächsten Evolutionsstufe des WWW zu verwirklichen, wozu insbesondere Linked Open Data bereits einen wichtigen Beitrag geleistet hat. Die LOD-Methode, die auf vier Prinzipien beruht, zielt auf die Integration strukturierter Informationen ins WWW ab, um dieses zu *einer* globalen, dezentralen Datenbank weiterzuentwickeln, die als LOD-Cloud oder Web (of Data) bezeichnet wird.

1.1 Motivation, Problemstellung

Die LOD-Methode könnte angewendet werden, um das historische Häuserbuch *ins* Web zu integrieren. Die Fakten über konkrete Individuen (sog. Instanzen) und die Konzepte aus dem historischen Häuserbuch könnten mit weiteren Ressourcen in Relation gesetzt werden. Die LOD-Cloud enthält Aussagen¹, die sich u. a. auf historische Individuen beziehen. Im historischen Häuserbuch werden ebenfalls Aussagen über historische Individuen gemacht. Diese Aussagen könnten sich gegenseitig ergänzen, indem explizit gesagt wird, welche Art von Beziehung zwischen den Referenzobjekten besteht. Dadurch könnte die Informationsmenge über ein Individuum vergrößert werden. Außerdem finden sich in der LOD-Cloud Vokabulare, die Termini enthalten, mit denen die Eigenschaften von Individuen und ihre Beziehungen untereinander beschrieben werden können. Es ist auch möglich, existierende Termini individuell zu spezialisieren. Die konzeptuelle Grundlage des historischen Häuserbuchs kann offengelegt, unmissverständlich beschrieben und mit Konzepten aus anderen Quellen verknüpft werden. Unter Anwendung der LOD-Methode lässt sich die umfassende Informationsintegration bewerkstelligen – es können nicht nur Instanzdaten, sondern auch Konzeptualisierungen kombiniert werden.

Diese Chance wurde im Rahmen der Projekte „Kultur-Hackathon Coding da Vinci Rhein-Main“² (CDV) und „Linked Pasts IV: Views From Inside The LOD-cloud“³ erkannt, die 2018 in Mainz stattfanden. Die eindrucksvollen Ergebnisse aus LOD-Projekten, die auf freiwilliger und gemeinnütziger Basis durchgeführt werden, inspirieren ebenso wie die erfolgreichen LOD-Unternehmungen öffentlicher Institutionen. Die Aussicht auf Verlinkung des historischen Häuserbuchs mit Ressourcen, die z.B. von der Deutschen Nationalbibliothek (DNB)⁴ veröffentlicht werden oder im Projekt DBpedia⁵ aus der digitalen Enzyklopädie Wikipedia extrahiert werden, motiviert zur konsequenten Beachtung der LOD-Prinzipien. Retrospektiv betrachtet, wurde der Grundstein für die vorliegende Arbeit im Rahmen der o.g. Projekte gelegt. Außerdem wurde eine Erkenntnis gewonnen:

Wenn die Daten aus dem historische Häuserbuch in einem proprietären Geodatenformat *über* das Web bereitgestellt werden, besteht die Chance auf Informationsintegration nicht. In diesem Fall manifestiert sich das historische Häuserbuch als isolierte Datensammlung. Der syntaktischen Interoperabilität kann zwar Rechnung getragen werden, indem die Daten mithilfe der universellen Auszeichnungssprache des Geo Web – der Geography Markup Language (GML) – kodiert und über einen standardkonformen WFS zum Download angeboten werden. Die automatische Konvertierung der Daten in GML-Repräsentationen führt allerdings nicht zu einer Wertsteigerung. Es wird nur ein weiteres Geodatensilo geschaffen. Für das Projekt CDV Rhein-Main 2018 wurden Teile des historischen Häuserbuchs in Form automatisch generierter GML-Instanzen mit bedeutungslosen GML-Schemata bereitgestellt. Diese konnten von Projektteilnehmerinnen um zusätzliche Daten aus anderen Quellen angereichert werden. Es wurde eine Webapplikation für einen interaktiven Stadtrundgang entwickelt⁶. Um für ähnliche Anwendungsfälle künftig eine bessere Ausgangsbasis zu schaffen, soll der Informationsumfang des historischen Häuserbuchs vergrößert werden, indem es in die LOD-Cloud integriert wird.

Um dieses Ziel zu erreichen, muss ein allgemeines Problem gelöst werden:

Aus einer Menge von vektoriellen Geoobjekten mit assoziierten Sachdaten, die in einem proprietären Geodatenformat vorliegen, muss ein gerichteter Graph geformt werden, der den LOD-Anforderungen entspricht.

¹ Der Terminus Aussage wird im Sinne eines Hyperonyms verwendet, das vor allem Behauptungen einschließen soll.

² Projektseite: <https://codingdavinci.de/de/events/rhein-main-2018>

³ Beschreibung der Hochschule Mainz: <https://www.hs-mainz.de/news/news/linked-pasts-iv/>

⁴ Linked-Data-Service der DNB:

https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/LDS/lds_node.html

⁵ Projektseite: <https://wiki.dbpedia.org/>

⁶ Siehe: https://codingdavinci.de/projects/2018_rm/Gensfleisch%20ist%20kein%20Gemuese.html#project-name

Dafür müssen alle Ressourcen (Instanzen, Konzepte) über Uniform Resource Identifiers (URIs) identifiziert werden. Die Ressourcen müssen mithilfe des flexiblen Datenmodells namens Resource Description Framework (RDF) und darauf aufbauenden Informations-Spezifikationssprachen beschrieben werden. Der RDF-Graph muss ins Web integriert werden, indem die Konzepte mit Vokabularen, die sich in der LOD-Cloud etabliert haben, beschrieben oder mit diesen in Beziehung gesetzt werden. Auf Instanzebene müssen ausgehende RDF-Links eingefügt werden. Dadurch soll die Informationsmenge über Individuen vergrößert werden.

Eigentlich zielt die LOD-Methode auf die Integration strukturierter Informationen ins WWW ab. In dieser Arbeit besteht die Herausforderung darin, dies primär auf Basis von Sachdaten – den thematischen Eigenschaften der Geoobjekte – zu realisieren, die aus einer Menge von unstrukturierten Zeichenketten (mehrwertige und zusammengesetzte Attribute) bestehen. Es muss eine Lösung gefunden werden, wie aus einer größtenteils ungeordneten Datensammlung, welche von Inkonsistenz und beträchtlicher Redundanz geprägt ist, ein LOD-konformer RDF-Graph mit verbesserter (zweckabhängiger) Informationsqualität (IQ) geformt werden kann.

1.2 Zielsetzung, Forschungsfrage(n)

In dieser Arbeit wird die Anwendbarkeit einer bekannten Methode in einem spezifischen Kontext untersucht. Aus der oben geschilderten Problematik resultiert die Leitfrage, ob die LOD-Methode auch angewendet werden kann, um den Informationsumfang von Modellen zu erhöhen, die keine konzeptuelle Grundlage aufweisen, die nur aus vektorialen Geodaten bestehen, welche ausschließlich in Form von nichtnormalisierten Relationen vorliegen. Um diese Frage beantworten zu können, müssen drei Teilziele erreicht werden:

- (1) Die Kriterien, die zur *Bewertung* des Arbeitsergebnisses / des resultierenden RDF-Graphen erforderlich sind, müssen ermittelt bzw. spezifiziert werden. Aus den grundlegenden Prinzipien, auf denen die LOD-Methode beruht, müssen die LOD-Anforderungen abgeleitet und definiert werden. Diese werden um die anwendungsfallsspezifischen Anforderungen ergänzt, die anhand ausgewählter MapInfo-Relationen aus der Sicht einer Datenbereitstellerin bestimmt werden.
- (2) Die vektorialen Geodaten aus den proprietären MapInfo-Relationen müssen zu einem RDF-Graph umgeformt werden, *der den definierten Anforderungen entspricht* und der im Vergleich zu den Ausgangsdaten *einen erhöhten Informationsumfang aufweist*. Hierzu ist ein Workflow zu konzipieren und praktisch umzusetzen. Im Rahmen des Workflows gilt es, die Ausgangsdaten zu transformieren, um die Fakten über die Individuen, ihre Eigenschaften sowie ihre Beziehungen untereinander in Tripelform beschreiben und die korrespondierenden Konzepte formalisieren zu können. Während dafür zwangsläufig die individuellen Charakteristika der Ausgangsdaten bzw. die anwendungsfallsspezifischen Anforderungen maßgeblich sind, wird zur Erhöhung des Informationsumfangs eine übertragbare Verfahrensweise entwickelt, die auf dem vierten LOD-Prinzip fußt und auf die Anreicherung des RDF-Graphen durch die Generierung von RDF-Links auf der Instanzebene abzielt. Die Praktikabilität des Verfahrens wird bewiesen, indem URI-Aliase oder URIs, welche auf ein ähnliches Individuum referieren, in den Informationsbeständen der DNB (Titeldatei, Gemeinsame Normdatei) ermittelt und mit ausgewählten Ressourcen aus dem historischen Häuserbuch verknüpft werden. Der Zielerreichungsgrad wird quantifiziert, indem auf die Anzahl der erzeugten RDF-Links auf der Instanzebene abgestellt wird.
- (3) Der erhöhte Informationsumfang muss aus Anwenderinnensicht mithilfe eines (Geo-)Informationssystems *demonstriert* werden. Es müssen nutzbare Antworten auf konkrete Fragen gegeben werden, die sich anhand der ursprünglichen MapInfo-Relationen nicht beantworten lassen. Zwecks Demonstration muss ein Anwendungsszenario konstruiert werden, da für die Präzisierung des unspezifischen Begriffs „Informationsumfang“ die Definition einer zweckabhängigen IQ-Dimension („angemessener Umfang“) herangezogen wird, die nur unter Berücksichtigung der Nutzung eruiert werden kann.

Der Informationsumfang des Modells soll durch die Erreichung des zweiten Teilziels erhöht werden. Der Umfang dieser Arbeit muss hingegen begrenzt werden. Deshalb werden die folgenden Nicht-Ziele definiert:

- Die ausgewählten Relationen werden nicht normalisiert und im Rahmen der Transformation wird keine Datenerfassung mittels zusätzlicher Primär- oder Sekundärquellen durchgeführt.
- Es wird kein Workflow geplant, der auf das ganzheitliche Datenmanagement von LOD in produktiven Umgebungen abzielt („LOD-Lifecycle“), sondern es werden jene Teilschritte konzipiert und praktisch umgesetzt, die für die Beantwortung der Leitfrage zwingend erforderlich sind. Da die Normalisierung nun schon mehrfach erwähnt wurde, sei explizit darauf hingewiesen, dass in der vorliegenden Arbeit sämtliche Aktualisierungsprozesse und die möglicherweise daraus resultierenden Anomalien ausgeklammert werden.
- Das Informationssystem (IS), das zur Demonstration erforderlich ist, wird nicht produktiv gestellt, d.h. es ist eine virtuelle Maschine (VM), die außerhalb des lokalen Netzes nicht erreichbar ist.
- Die Syntax semantischer Standards und Basistechnologien (z.B. der verschiedenen RDF-Serialisierungen oder von SPARQL) wird im theoretischen Teil nicht erläutert. Ebenso wird davon abgesehen, eine bestimmte (Auszeichnungs-)Sprache zu erörtern; nämlich: die Extensible Markup Language (XML) inklusive XSL Transformation (XSLT), XML Path Language (XPath), XML Query Language (XQuery), XML Linking Language (XLink). Selbiges gilt für die „Structured Query Language“ (SQL).
- In dieser Arbeit wird keine Applikation mithilfe einer objektorientierten Programmiersprache entwickelt oder weiterentwickelt.

Das letztgenannte Nicht-Ziel musste definiert werden, nachdem das Zeitbudget für die vorliegende Arbeit erschöpft war. An dieser Stelle soll wenigstens noch erwähnt werden, dass der angedeutete Plural in der Überschrift dieses Abschnitts darauf hinweisen soll, dass ursprünglich noch eine Hypothese verifiziert werden sollte:

Wenn die o.g. Geodaten ins Web integriert werden, dann profitieren auch GIS-Anwenderinnen davon, weil die vernetzten Informationen über einen standardkonformen Basic-WFS bereitgestellt und mithilfe eines WFS-Clients prozessiert werden können.

Diese Hypothese drängte sich im Zuge der Literaturarbeit geradezu auf, da sowohl GML als auch RDF auf ähnlichen Modellierungsansätzen (Objekt-Property-Muster, gerichteter Graph) basieren und die einleitend erwähnten Standardisierungsgremien (mittlerweile) kooperieren, um semantische und räumliche Konzepte zusammen zu bringen. Trotzdem haben die semantischen Basistechnologien (mind. RDF, SPARQL) im GIS-Bereich noch nicht vollständig Einzug erhalten. Rein pragmatisch betrachtet, sieht man sich mit dem Problem konfrontiert, dass ArcGIS for Desktop, ArcGIS Pro, MapInfo Professional und QGIS in den jeweils aktuellen Versionen⁷ WFS konsumieren können, jedoch keines der vorgenannten Desktop-GIS native RDF- oder SPARQL-Unterstützung bietet. Für manche Desktop-GIS existieren inzwischen Erweiterungen, die sich als Resultate aus Forschungsprojekten allerdings häufig noch im Alpha- oder Beta-Stadium befinden. Abgesehen davon, wird der Einsatz solcher Erweiterungen in vielen Organisationen restriktiv gehandhabt, weshalb einige Anwenderinnen diese überhaupt nicht verwenden können. Um den GIS-Anwenderinnen einen einfachen und freien Zugang zu den vektorialen Geodaten des historischen Häuserbuchs bieten zu können, sollte deshalb auch ein standardkonformer Basic-WFS eingerichtet werden. Obwohl das OGC jüngst mit der OGC API Familie vom „Remote-Procedure-Call-over-HTTP architectural style using XML for any payloads“ (Portele et al., 2019, p. 17) abgekommen ist, nimmt der WFS mit der gefilterten Download-Möglichkeit in vielen bestehenden – „legacy“ – Systemen und Infrastrukturen nach wie vor eine wichtige Funktion ein. Das o.g. Problem besteht aber auch serverseitig. Momentan operieren weder verbreitete Open-Source-Applikationen wie GeoServer oder MapServer noch proprietäre Lösungen wie ArcGIS (for) Server auf Triple- bzw. Quadstores.

⁷ Stand: Dezember 2020

Dafür kann jedoch auf „Middleware-Lösungen“ zurückgegriffen werden, die im Rahmen von Forschungsprojekten entwickelt wurden (Proxys, Wrapper, Adapter etc.). Von der Adaption der vorgenannten Applikationen (Anpassung und Erweiterung des Quellcodes) musste in der vorliegenden Arbeit letztlich abgesehen werden. Die o.g. Applikationen haben zwar noch eine Relevanz hinsichtlich des dritten Teilziels, da die vergleichende Untersuchung der vernetzten Informationen mittels SPARQL- und WFS-Clients im Zuge der Demonstration sinnvoll wäre. Dies setzt jedoch voraus, dass die infrage kommenden Applikationen zumindest in funktionaler Hinsicht dafür geeignet sind und nicht erst (weiter-)entwickelt werden müssen. Sofern die Geeignetheit für den konkreten Anwendungsfall nicht festgestellt werden kann, wird das Thema im Zuge der Produktivstellung des Systems erneut aufgegriffen.

1.3 Hintergrund, Zielgruppe

In dieser Arbeit wird eine praxisnahe Untersuchung durchgeführt. Die Anwendbarkeit der LOD-Methode und die positiven Aspekte vernetzter Informationen werden nicht anhand eines idealtypischen Beispiels eruiert. Stattdessen wird ein Experiment mit Daten durchgeführt, welche die o.g. Charakteristika aufweisen und somit keine ideale Ausgangsbasis für die Generierung eines LOD-konformen RDF-Graphen bieten. Dem liegt die Überlegung zugrunde, dass sich Kolleginnen aus der Praxis mit ähnlichen Problemstellungen konfrontiert sehen könnten, wenn sie ihre Ressourcen in die LOD-Cloud integrieren möchten. Dies könnten ebenfalls nicht-normalisierte Relationen oder z.B. semi-strukturierte Daten sein, die vielleicht das Ergebnis einer Datenmigration sind. In dieser Arbeit werden Lösungen konzipiert und praktisch realisiert, damit solchen (Daten-)Sammlungen eine Bedeutung zugewiesen werden kann, damit sie unter Beachtung der LOD-Prinzipien umgeformt werden können, sodass sie durch Verlinkung ins Web integriert werden können, um vom Netzwerk-Effekt profitieren zu können. Ob und in welchem Grad der resultierende RDF-Graph den LOD-Anforderungen genügt, wird zuerst aus der Sicht einer Datenbereiterin überprüft. Ob die Informationsqualität verbessert bzw. der Informationsumfang erhöht werden konnte, wird danach aus der Sicht einer potenziellen Anwenderin demonstriert. Dementsprechend werden in dieser Arbeit auch die vorgenannten Kolleginnen aus der Praxis adressiert, die Informationen ins Web of Data integrieren oder vernetzte Informationen verwenden möchten.

Es wird davon ausgegangen, dass die o.g. Adressatinnen ein Interesse an semantischen Basistechnologien haben und nicht notwendigerweise über genauere Kenntnisse auf diesem Gebiet verfügen. Diese mussten auch im Rahmen dieser Arbeit von Grund auf erarbeitet werden.

1.4 Inhalte, Strukturierung

In diesem Zuge entstand das zweite Kapitel, das etwas umfangreicher ausgefallen ist. Damit möchte die vorliegende Arbeit den o.g. Adressatinnen einen einfacheren Zugang zu den o.g. Technologien ermöglichen, sodass sie zur Expansion des Web of Data beitragen können, ohne zuvor terminologische Probleme klären zu müssen. In der ersten Hälfte des theoretischen Teils (Abschnitt 2.1 und 2.2) wird klargestellt, was unter dem semantischen Netz eigentlich zu verstehen ist, welche Rolle Linked (Open) Data dabei spielt und auf welchen Standards sowie Architekturprinzipien das Web basiert. Das Semantic Web soll verständlich erklärt und möglichst scharf umrissen werden. Deshalb erfolgt eine vergleichende Gegenüberstellung des Web of Data und des Web of Documents. Da nähere Kenntnisse zum Letztgenannten sehr hilfreich für das Verständnis sind, werden essentielle Standards des Web of Documents und dessen Entwicklung in Abschnitt 2.1.1 thematisiert, um anschließend die wichtigsten Begriffe (GP 2.1.2) und die Web-Architekturprinzipien (GP 2.1.3) anschaulich anhand eines Schlüsselbeispiels erklären zu können. Darauf wird in Abschnitt 2.2 immer wieder Bezug genommen, um zuerst die Charakteristika und die Intention des Web of Data hervorheben zu können. In diesem Zuge

werden auch die LOD-Prinzipien eingeführt und der Netzwerk-Effekt in der LOD-Cloud wird erklärt (GP 2.2.1). Aufgrund der Gegenüberstellung mit dem Web of Documents werden die Unterschiede und Gemeinsamkeiten besonders deutlich. Danach können die wesentlichen Grundlagen zur Wissensrepräsentation (mittels RDF) erörtert werden (GP 2.2.2, GP 2.2.3). Die wichtigsten Begriffe dazu werden unter GP 2.2.4 definiert. Schließlich wird erklärt, wie die Web-Architekturprinzipien mittels RDF umgesetzt werden können (GP 2.2.5).

In der zweiten Hälfte des theoretischen Teils (Abschnitt 2.3 und 2.4) wird der Raumbezug in den Mittelpunkt gestellt. Zunächst werden ausgewählte Standards bzw. Empfehlungen vorgestellt, mit denen die syntaktische Interoperabilität im Geo Web sichergestellt wurde. In der vorliegenden Arbeit müssen lediglich der Basic WFS als Downloaddienst (GP 2.3.2) und GML als „Auszeichnungssprache“ für vektorielle Geodaten (GP 2.3.3) erläutert werden. Die vorgenannten OGC-Standards können für die Demonstration (3. Teilziel) verwendet werden, da der gerichtete Graph nicht nur mittels RDF-Serialisierungen, sondern auch in GML-Strukturen abgebildet werden kann. Deshalb werden nicht alle Details zu GML besprochen, sondern die Parallelen zwischen RDF und GML werden in den Fokus gerückt. Der Paradigmenwechsel im Geo Web wird unter GP 2.3.4 angerissen. Anschließend werden relevante Punkte aus dem weiten (Forschungs-)Feld des Semantic Geospatial Web erläutert, nämlich grundlegende Ansätze zwecks „Semantic Enablement“ (GP 2.4.1) und Empfehlungen bzw. Standards, vor allem „GeoSPARQL“ sowie raumzeitliche Modelle und Konzeptualisierungen (GP 2.4.2).

In Abschnitt 2.5 werden die Inhalte des theoretischen Teils stichpunktartig zusammengefasst. Abgesehen von der bündigen Rekapitulation, versteht sich diese Passage auch als Brücke zum praktischen Teil. Es wird dargelegt, warum die jeweiligen Punkte für die Zielerreichung bzw. für die Beantwortung der Leitfrage wichtig sind.

In Kapitel 3 wird der aktuelle Forschungsstand überblicksmäßig dargestellt. Zuerst werden die speziellen Verfahrensweisen, auf die sich die vorliegende Arbeit methodisch bezieht, neutral vorgestellt. Selbiges gilt für die Hilfsmittel – (prototypische) Implementierungen respektive Forschungsergebnisse. Die Vor- und Nachteile, die sich bei der Anwendung bestimmter Verfahren ergeben, werden erst in Kapitel 7 (Lösungsansatz) diskutiert.

In Kapitel 4 wird das historische Häuserbuch der Stadt Mainz aus der Sicht einer Anwenderin vorgestellt. Es handelt sich lediglich um ein Kurzportrait, das der Veranschaulichung des Diskursuniversums dient. In Kapitel 5 wird die Datenuntersuchung, -auswahl und -anpassung in den Mittelpunkt gestellt. Zunächst wird das historische Häuserbuch behandelt. Die Vektormodelle werden charakterisiert und im Detail dokumentiert. Die Auswahl bestimmter MapInfo-Relationen wird begründet. Außerdem wird dargelegt, welche Relationen im Vorfeld angepasst wurden und welche Individuen für die Verlinkung ausgesucht wurden. Danach werden die Wissensbasen aus der LOD-Cloud vorgestellt, die für die Generierung des RDF-Graphen ausgewählt wurden.

Kapitel 6 ist der Anforderungsanalyse und -definition gewidmet. Es werden die Quellen genannt, die zur Definition der LOD-Anforderungen herangezogen werden. Vor der Anforderungsdefinition werden essentielle Punkte aus einer einschlägigen W3C- / OGC-Empfehlung diskutiert. Es wird begründet, warum bestimmten Quellen der Vorzug gegeben wird. Das Kapitel schließt mit der Vorstellung des fiktiven Anwendungsszenarios.

In Kapitel 7 wird zunächst das System entworfen, das für die Generierung des Graphen und für die Demonstration erforderlich ist. Nach der Diskussion der potenziellen Systemarchitekturen (GP 7.1), werden die infrage kommenden Applikationen besprochen (GP 7.1.1). Die Replikation muss kurz angerissen werden (GP 7.1.2).

Abschnitt 7.2 ist dem zweiten Teilziel dediziert: Der Workflow zwecks Generierung des RDF-Graphen wird konzipiert, die infrage kommenden Verfahren, Vokabulare, Werkzeuge etc. werden diskutiert und die Verfahrensweise für die Anreicherung bzw. für die Generierung von RDF-Links auf der Instanzebene wird vorgestellt.

Die praktische Umsetzung des Workflows wird ausführlich in Abschnitt 8.2 erörtert; es wird aufgezeigt, welche Entscheidungen getroffen wurden und welche Probleme sich ergaben. Der Aufbau des Systems nimmt dagegen eine untergeordnete Rolle ein (GP 8.1). Der empirische Teil konzentriert sich auf die Generierung des Graphen.

In Kapitel 9 wird das Arbeitsergebnis evaluiert: Zuerst wird aus der Sicht einer Datenbereitstellerin geprüft, ob bzw. inwieweit der generierte RDF-Graph die definierten Anforderungen erfüllt (GP 9.1). Nachdem der erhöhte Informationsumfang aus Anwenderinnensicht und bezugnehmend auf das fiktive Anwendungsszenario demonstriert wurde (GP 9.2), wird die Leitfrage beantwortet (GP 9.3). Schließlich wird die Übertragbarkeit bzw. die Wiederverwendbarkeit der erzielten Ergebnisse im Kontext anderer Projekte besprochen (GP 9.4).

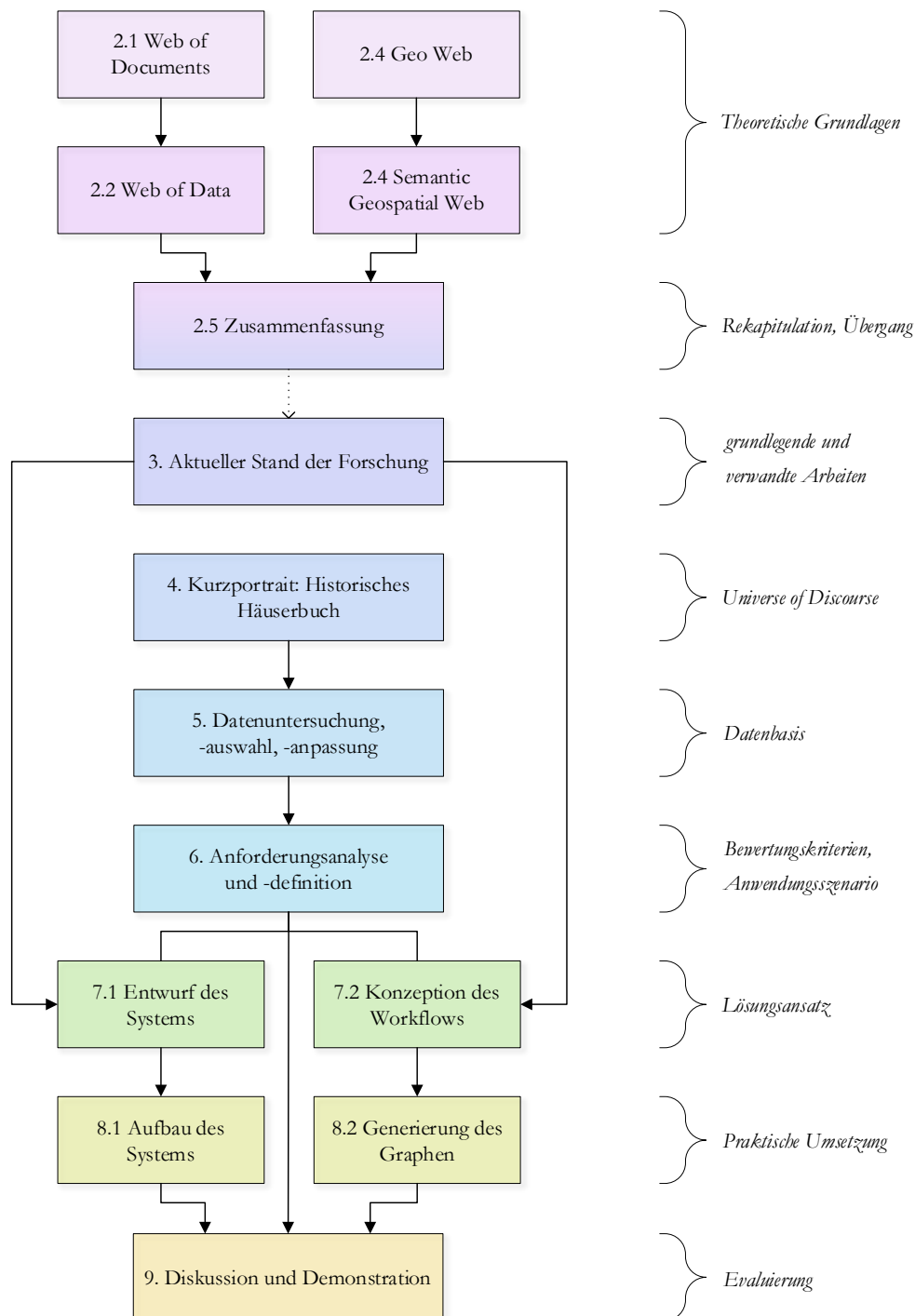


Abb. 1.1: Struktur der Thesis, ED

2 Theoretische Grundlagen

2.1 Web of Documents

Der PC und das Internet bildeten laut Halpin (2013) die technologische Basis für die erfolgreiche Realisierung des WWW und dessen Weiterentwicklung bzw. Expansion. Berners-Lee (1989) konzipierte das Web ursprünglich als „universal information space“ und forderte: „We should work toward a universal linked information system, in which generality and portability are more important than fancy graphics and complex extra facilities“.

2.1.1 Entwicklung, Standards

Um dieses Ziel zu erreichen, forcierten er und seine Mitstreiterinnen die (Weiter-)Entwicklung bzw. Standardisierung der Hypertext Markup Language (HTML), des Hypertext Transfer Protocol (HTTP) und der URIs durch die IETF (Halpin, 2013). **HTML** sollte als universelle Auszeichnungssprache der Strukturierung digitaler, plattformunabhängiger **Dokumente** und deren **Verknüpfung** (über sog. Hyperlinks) dienen (Bizer et al., 2009). Laut Steinmetz (1999) können HTML-Dokumente neben Hypertext auch die verschiedensten multimedialen Inhalte umfassen (sog. Hypermedia-Konzept). **HTTP** wurde zwar speziell für Hypermedia-Informationssysteme konzipiert, ist als generisches Protokoll aber nicht auf die **Übertragung** von Hypertext/Hypermedia beschränkt (Berners-Lee et al., 1996). **URIs** sollen an dieser Stelle als persistente, globale, universelle Identifikatoren für Ressourcen eingeführt werden (Berners-Lee, 1996). Es wird zunächst keinen Wert auf eine formal korrekte oder vollständige Definition gelegt. Wichtiger ist die Erkenntnis, dass URIs eigentlich einen universellen Charakter haben, da **jede Ressource** (eine Entität, ein Ding oder Objekt) per URI identifiziert werden kann (Berners-Lee, 1996). Das Web wird dann zu einem „virtual space for naming information“ (Halpin 2013, p. 28). Ressourcen werden zu „First Class Objects (FCOs)“ im Web (Berners-Lee, 1996). Diese abstrakte Darstellung wird im Rahmen der weiteren Erörterung sukzessive konkretisiert. Zunächst soll dies für das Verständnis der folgenden Aussagen von Berners-Lee and Fischetti (1999, p. 48) genügen: „What was often difficult for people to understand about the design was that there was nothing else beyond URIs, HTTP, and HTML. There was no central computer ‘controlling’ the Web, no single network on which these protocols worked, not even an organization anywhere that ‘ran’ the Web. The Web was not a physical ‘thing’ that existed in a certain ‘place.’ It was a ‘space’ in which information could exist“. Allemang et al. (2020) illustrierten die drei vorgenannten Standards des Web of Documents mittels der folgenden Graphik:

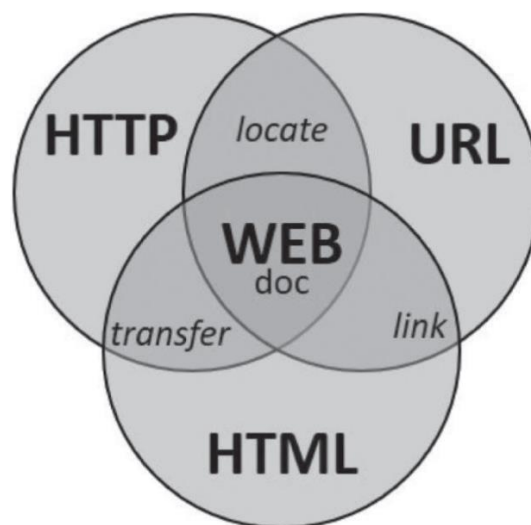


Abb. 2.1: Standards des Web of Documents, OA von Allemang et al. (2020, p. 86)

Im Web of Documents wurde eine spezifische Unterart von URIs verwendet: die Uniform Resource Locators (URLs). Während der Erarbeitung des ersten URI-Standards fanden Berners-Lees Ansichten in der IETF zunächst wenig Zustimmung (Berners-Lee and Fischetti, 1999), weshalb sein ursprüngliches URI-Konzept in zwei Standards gespalten wurde (Halpin, 2013). Die folgende Abbildung zeigt die Ergebnisse des Standardisierungsprozesses in der IETF und präzisiert URLs und Uniform Resource Names (URNs):

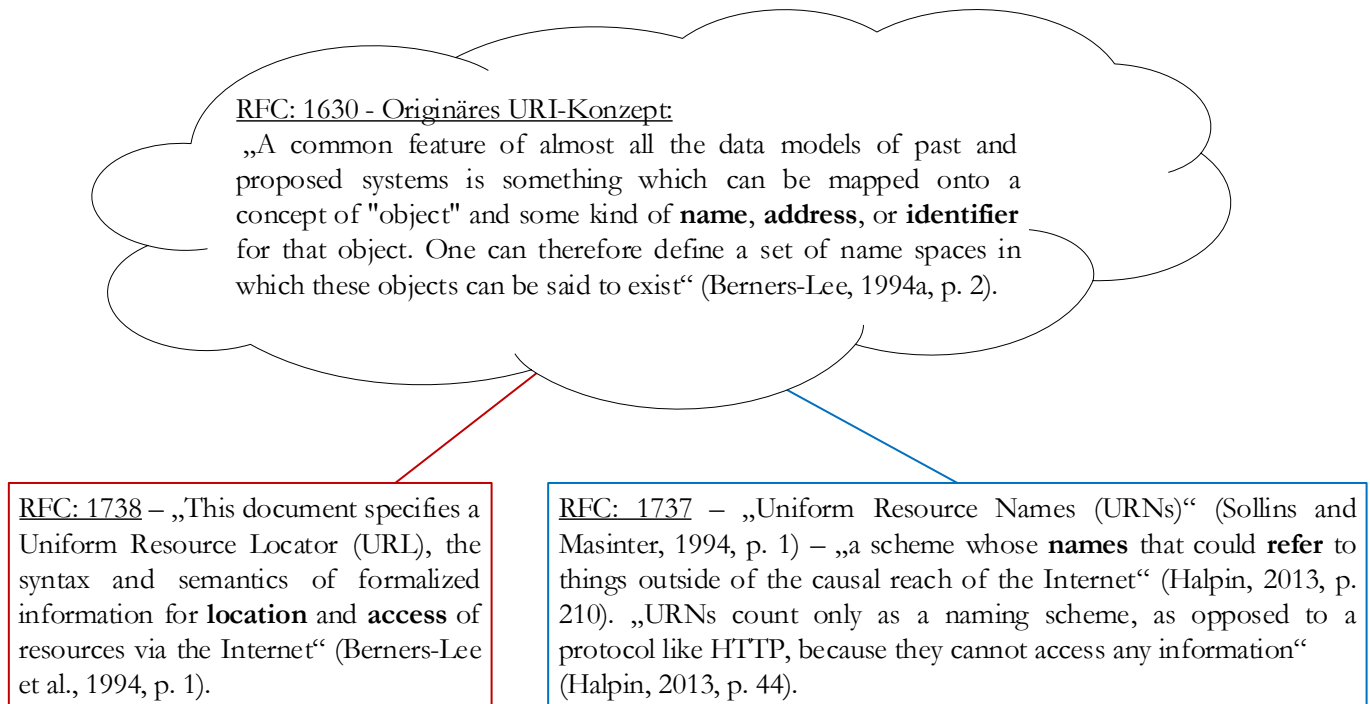


Abb. 2.2: Spaltung des URI-Konzepts – Differenzierung zwischen Zugang und Referenz, ED

Alle fett markierten Begriffe sind besonders relevant für diese Arbeit und werden später noch definiert. Wichtig ist an dieser Stelle, dass im Web of Documents URIs nicht im originären Sinne, sondern als **URLs** verwendet wurden. URLs identifizieren laut Halpin (2013) digitale Ressourcen und dienen deren **Lokalisierung**, d.h. der Ermittlung einer Adresse. Diese Adresse wird verwendet, um z.B. über **HTTP** Zugang zu den Informationen zu erhalten, die hauptsächlich in **HTML** strukturiert sind, d.h., die **HTML-Dokumente** (hier synonym: Webseite) werden per HTTP über ein Netzwerk **transferiert** (Berners-Lee et al., 1994). URLs dienen auch als **Links** um HTML-Dokumente bzw. die darin enthaltenen Informationseinheiten zu verknüpfen (Allemang et al., 2020). Dies ist der Grund, warum URLs der De-Facto-Standard im Web of Documents wurden; jede Person konnte z.B. einen Browser nutzen, um Webseiten aufzurufen bzw. um zwischen diesen zu navigieren. URNs wurden allerdings weniger häufig verwendet, da sie i.d.R. für die direkte Lokalisierung und damit für den Informationszugang via HTTP nicht geeignet sind (Halpin, 2013).

Die o.g. **Requests for Comments** (RFCs) waren die Ergebnisse eines frühen Standardisierungsprozesses in der IETF (Halpin, 2013). Generell qualifizieren sich RFCs häufig als De-Facto-Standards mit normativer Wirkung für das Internet und/oder für das Web (Halpin, 2013). Berners-Lee gab seine Vorstellungen von Universalität zunächst auf, um für das Web eine normative Grundlage zu schaffen, sodass es auf dieser Basis wachsen konnte (Berners-Lee and Fischetti, 1999). Nach der Limitierung des URI-Konzepts sah er jedoch auch den universellen Charakter von HTML durch die sog. „Browserkriege“ gefährdet (Halpin, 2013). Im Jahr 1994 gründete er das W3C, um die langfristige, strategische Weiterentwicklung des WWW durch die Ausarbeitung/Standardisierung von adäquaten Technologien/Techniken und Richtlinien sicherzustellen (Berners-Lee and Fischetti, 1999).

Für die vorliegende Arbeit sind neben den RFCs, besonders die Empfehlungen des W3C (**W3C Recommendations**) relevant. Ähnlich wie die RFCs haben die Empfehlungen des W3C normative Wirkung für das Web (Halpin, 2013). Eine frühe W3C-Empfehlung widmete sich der Standardisierung von HTML, als herstellerneutrale, universelle Auszeichnungssprache (Raggett et al., 1999), um die (ersten) „Browserkriege“ zu beenden (Halpin, 2013). Die Akzeptanz bzw. die Umsetzung der Empfehlungen führte dazu, dass HTML in verschiedenen Applikationen, vor allem in Browsern, interoperabel verwendet werden konnte, was laut Heath and Bizer (2011) als eine wesentliche Voraussetzung für die Expansion des Web of Documents anzusehen ist.

2.1.2 Terminologie

Für adäquate Begriffsdefinitionen sind weder die RFCs noch die Empfehlungen des W3C ausreichend, da essentielle Termini häufig unscharf oder widersprüchlich definiert wurden (Halpin, 2013). Deshalb erarbeiteten Presutti and Gangemi (2008) die IRE-Ontologie⁸. IRE steht für „Identity of Resources and Entities on the web“ (Presutti and Gangemi, 2008, p. 50). Halpin and Presutti (2009) entwickelten die IRE zur IRW-Ontologie weiter⁹. IRW steht entweder für „Identity and Reference on the Web“ (Halpin and Presutti, 2009, p. 521) oder für „Identity of Resources on the Web“ (Halpin and Presutti, 2009, p. 522). Halpin and Presutti (2009) formalisierten Konzepte der Web-Architektur. Diese wurde von Jacobs and Walsh (2004) im Dokument namens „Architecture of the World Wide Web, Volume One“ spezifiziert¹⁰, das dazu dient, die Web-Architekturprinzipien darzulegen und adäquate Entscheidungen für die Entwicklung neuer Web-Technologien oder Web-Applikationen aufzuzeigen. Es ist die detaillierte Exegese der informellen und ungeordneten „Design Issues: Architectural and philosophical points“ von Berners-Lee¹¹ und Fieldings Dissertation „Architectural Styles and the Design of Network-based Software Architectures“ (Fielding, 2000, p. 1). Die AWWW wurde von der Technical Architecture Group (TAG) ausgearbeitet und liegt als W3C-Empfehlung vor (Jacobs and Walsh, 2004).

Netzwerkinteraktionen werden nachfolgend anhand eines Schlüsselbeispiels erläutert: Ein **Agent** greift auf verschiedene **Ressourcen** zu, die Informationen *über* das historische Gebäude „Zum kleinen Schlüssel“ enthalten, indem er **URLs dereferenziert**, um über **HTTP Zugang** zu **Web-Repräsentationen** in verschiedenen **Kodierungen** zu erhalten. Zuerst fordert der Agent textuelle Informationen über das Gebäude an, die in HTML strukturiert sind und später Abbildungen des Gebäudes, die als Raster- oder Vektorgraphiken verfügbar sind.

Dafür sind **Protokolle** erforderlich. „A protocol is a convention for transmitting information between two or more **agents**“ (Halpin, 2013, p. 18). Hier ist HTTP das wichtigste Protokoll. „An agent is any thing capable of interacting via a protocol“ (Halpin, 2013, p. 18). Exemplarisch können hier Crawler und Browser genannt werden (Heath and Bizer, 2011). Allerdings kann man auch einen Menschen, der einen Browser verwendet, als Agent bezeichnen (Halpin, 2013). Laut Fielding et al. (1999) transferiert HTTP Informationen als Nutzlast (engl. **payload**). „A payload is the information transmitted by a protocol“ (Halpin, 2013, p. 18). Die Nutzlast wird im Rahmen einer Anfrage (engl. **request**) oder im Zuge einer Antwort (engl. **response**) via HTTP übertragen (Fielding et al., 1999). Bengel (2014, p. 23) bezeichnete dies als „Interaktion zwischen den **Clients** und dem **Server**“. Das WWW basiert auf der **Client-Server-Architektur**, die Halpin (2013, p. 18) wie folgt beschrieb: „meaning that protocols take the form of a request for information and a response with information. The client is defined as the agent that is requesting information and the server is defined as the agent that is responding to the request“. „In a protocol, an **endpoint** is any process that either requests or responds to a protocol, and so includes both client and servers“ (Halpin, 2013, p. 18). Fielding et al. (1999) erstellten eine Übersicht zu den verfügbaren **HTTP-Methoden**, die Tilkov et al. (2015, p. 53) alternativ als „**Verben**“ oder

⁸ Semantic Web Dokument (OWL-Version): <http://www.ontologydesignpatterns.org/cpont/ire.owl>

⁹ Semantic Web Dokument (OWL-Version): <http://ontologydesignpatterns.org/ont/web/irw.owl>

¹⁰ Im Folgenden wird die Abkürzung „AWWW“ verwendet.

¹¹ Siehe: <https://www.w3.org/DesignIssues/>

„Operationen“ bezeichneten. Davon ist in dieser Arbeit nur die idempotente Methode GET relevant. Die Antwort eines HTTP-Servers auf eine GET-Anfrage beinhaltet eine sog. „Status-Line“ mit einem **Statuscode** (Fielding et al., 1999, p. 39), den Halpin (2013, p. 19) als „one of a finite number of codes which gives the user-agent information about the server’s HTTP response itself“ definierte. Fielding et al. (1999) spezifizierten die **Bedeutung der Statuscodes**, die im Semantic Web besonders relevant ist (Tab. 2.1). Die Antwort eines HTTP-Servers auf eine GET-Anfrage kann weitere Informationen enthalten, wobei die „entity“ wichtig ist, die Fielding et al. (1999, p. 8) definierten und weiter differenzierten: „The information transferred as the payload of a request or response. An entity consists of meta-information in the form of **entity-header fields** and content in the form of an **entity-body**“. Für diese Arbeit sind die Header-Felder Content-Language und Content-Type wichtig, die in RFC 2616 detailliert erörtert wurden. Das Feld Content-Type enthält z.B. die Information, dass die gelieferte **Web-Repräsentation** der angefragten **Ressource** in der **Sprache** HTML **kodiert** ist. Das Feld Content-Language enthält ebenfalls eine Information zur Sprache (z.B. de = deutsch). Im Gegensatz zu den Informationen im Feld Content-Type wird hier allerdings keine **formale**, sondern eine **natürliche** Sprache spezifiziert. Beide Sprachangaben sind für die **Interpretation** der **Nachricht** wichtig. Content-Typen in HTTP wurden zu Internet **Media Types** generalisiert (Halpin, 2013). „A media type consists of a two-part scheme that separates the type and a subtype of an **encoding**“ (Halpin, 2013, p. 20). „An encoding is a set of precise regularities that can be realized by the **message**“ (Halpin, 2013, p. 22). „The message is the physical thing that realizes the regularities of the **information** due to its local characteristics. Information is whatever regularities held in common between a source and a receiver“ (Halpin, 2013, p. 21). Halpin (2013, p. 23) schlussfolgerte: „the **content** of any information-bearing message is whatever is held in common between the source and the receiver as a result of the conveyance of a particular message“. Die anschließende Definition verknüpft zwei Begriffe: „The relationship of an encoding to its content, is an **interpretation**“ (Halpin, 2013, p. 23). Die Zusammenhänge werden anhand des Schlüsselbeispiels veranschaulicht: Der Sender (engl. source) schickt dem Empfänger (engl. receiver) eine Nachricht (engl. message), die Informationen in einer bestimmten Kodierung (engl. encoding) enthält. Es kann interpretiert werden, dass sich der Inhalt (engl. content) der Nachricht auf das historische Gebäude „Zum kleinen Schlüssel“ bezieht.

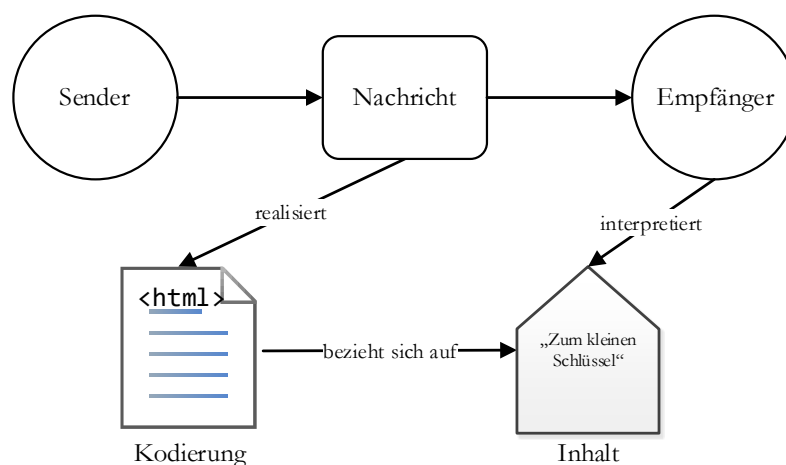


Abb. 2.3: Information, Kodierung, Inhalt, OA von Halpin (2013, p. 25), angepasste Darstellung

Dieses einfache Beispiel kann nicht ohne Weiteres auf das WWW übertragen werden, das auf die Kommunikationsinfrastruktur des Internets aufsetzt (Beißwenger and Pappert, 2018). In verschiedenen Schichten werden unterschiedliche Protokolle genutzt, die Inhalte abstrahieren bzw. Kodierungen verschachteln (Schwenk, 2010).

Laut Halpin (2013) ist für die Kodierung (mindestens) eine Sprache (engl. language) erforderlich. Halpin (2013, p. 25) definierte den Begriff wie folgt: „A **language** can be defined as a system in which information is related to other information systematically“. Er stellte klar: „In a language, this is a relationship between how the encoding of some information can change the interpretation of other encodings“ (Halpin, 2013, p. 25). Anschließend definierte er die **Syntax einer Sprache** als „a system in which the encoding of information is related to each other systematically“ und die **Semantik einer Sprache** als „a system in which the content of information is related to each other systematically“ (Halpin, 2013, p. 25). Da es im empirischen Teil dieser Arbeit häufig darum geht, Aussagen (engl. statements) über Ressourcen in einer formalen Sprache zu machen, sind noch die folgenden Termini relevant: „Any combination of terms that is valid according to the language’s syntax is a **sentence** (sometimes an ‘expression’) in the language, and any combination of terms that has an interpretation to content according to the language’s semantics is a **statement** in the language“ (Halpin, 2013, p. 25).

Jacobs and Walsh (2004) definierten das **WWW** als „an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI)“. Fielding et al. (1999, p. 8) definierten den Begriff Ressource wie folgt: „A network data object or service that can be identified by a URI“. Fielding et al. (1999) widersprachen damit den früheren Aussagen von Berners-Lee et al. (1998, p. 2): „Not all resources are network "retrievable"; e.g., human beings“. Eine Definition lieferten sie aber auch nicht; Berners-Lee war der Meinung, das Beste an Ressourcen sei u. a., dass er den Begriff nie definieren musste (Berners-Lee and Fischetti, 1999). RFC 3986 enthält auch keine Definition, sondern die folgende Aussage: „the term "resource" is used in a general sense for whatever might be identified by a URI“ (Berners-Lee et al., 2005, p. 5). Halpin (2013, p. 30) leitete daraus folgende Definition ab: „A **resource** is any thing capable of being content, or in other words, an ‘identity’ in a language“. Halpin (2013, p. 30) verknüpfte diese Definition mit den o.g. Definitionen: „a resource is typically not a particular encoding of some content but some content that can be given by many encodings“. Eine URI identifiziert also den Inhalt einer Ressource – quasi deren Sinngehalt – auf einer abstrakten Ebene, nicht jedoch den Inhalt der Ressource in einer konkreten Kodierung (Halpin, 2013). Dies kann anhand des Schlüsselbeispiels verdeutlicht werden. Vorher muss genauer erklärt werden, was eine **URI** eigentlich ist. URIs sind das zentrale Charakteristikum des Webs (Jacobs and Walsh, 2004). Das Web wird erst zu einem „universal information space“ (Halpin 2013, p. 28) wenn Ressourcen über URIs identifiziert werden. Andernfalls sind keine Referenzen möglich und die Informationen, die eine Ressource beinhalten kann, sind z.B. über HTTP nicht zugänglich (Berners-Lee, 1996). URIs bieten *einen* Namensraum mit globalem Gültigkeitsbereich und realisieren damit einen eindeutigen, weltweiten Identifikationsmechanismus für Ressourcen (Berners-Lee et al., 1998). „The Web is then first and foremost a naming initiative“ (Halpin, 2013, p. 26). RFC 3986 definiert die generische URI-Syntax (Berners-Lee et al., 2005). Der eingangs erwähnte universelle Charakter von URIs kann anhand der folgende Zeichenkette illustriert werden, die eine konkrete HTTP-URI/HTTP-URL ist. Andere Schemata werden nicht thematisiert, da im LOD-Kontext nur HTTP-URIs verwendet werden (Heath and Bizer, 2011).

scheme authority path fragment
 http://history.mainz.de/def#hadOccupant

Abb. 2.4: URI-Syntax, ED

Die dargestellten Komponenten definierte Halpin (2013, p. 27) wie folgt:

- „a **scheme** is a name of the protocol or other naming convention used in the URI“,
- „the **authority** is a name that is usually a domain name, naming authority, or a raw IP address, and so is often the name of the server“.

Im RFC 3986 finden sich folgende Definitionen:

- „The **path** component contains data, usually organized in hierarchical form, that, along with data in the non-hierarchical query component serves to identify a resource within the scope of the URI's scheme and naming authority“ (Berners-Lee et al., 2005, p. 22).
- „The **fragment** identifier component of a URI allows indirect identification of a secondary resource by reference to a primary resource and additional identifying information. The identified secondary resource may be some portion or subset of the primary resource, some view on representations of the primary resource, or some other resource defined or described by those representations“ (Berners-Lee, et al. 2005, p. 24).

Einige Komponenten von URIs wurden ausgespart. Für Details wird auf RFC 3986 verwiesen. Grundlegende Informationen zur Namensauflösung über das Domain Name System (DNS) enthalten die RFCs 1034 und 1035 (Mockapetris, 1987). Die obige URI könnte per HTTP (indirekt) dereferenziert werden, da alle protokoll-spezifischen Angaben in der URI per se enthalten sind. Das Schema (engl. scheme), hier mit „http“ angegeben, spezifiziert nur das bevorzugte Protokoll; es legt aber nicht fest, mit welchem Protokoll ein Agent die Ressource tatsächlich anfragen muss (Halpin, 2013). Generell ist eine URI nicht von einem bestimmten Protokoll abhängig (Berners-Lee, 1994a). Im weiteren Verlauf wird die Kombination aus Schema und Autorität (engl. authority) informell als **Base-URI** bezeichnet, die immer ident ist (<http://www.mainz.de>). Der Pfad (engl. path) kann variieren. Bezüglich der „fragment identifier“ ist noch erwähnenswert, dass das Format und die Interpretation, vom „media type“ abhängig ist, was clientseitig behandelt wird (Berners-Lee et al., 2005). Falls in der Anfrage eine URI mit „fragment identifier“ verwendet wird, ist dies für den Server grundsätzlich nicht relevant, denn der „fragment identifier“ wird vor der Beantwortung der Anfrage „abgeschnitten“ (Berners-Lee, 1996).

Jacobs and Walsh (2004) definierten etliche **Anforderungen**, die bei der Konstruktion von URIs beachtet werden müssen. **URI-Kollisionen** entstehen, wenn dieselbe URI *direkt* unterschiedliche Ressourcen identifiziert (Jacobs and Walsh 2004). Es ist wichtig URI-Kollisionen zu vermeiden, denn diese führen zu Mehrdeutigkeiten (Jacobs and Walsh, 2004). **URI-Allokation** bedeutet, dass die Eigentümerin einer Ressource diese per URI identifiziert (Jacobs and Walsh, 2004). URIs sind verschieden, wenn die Zeichenfolge insgesamt nicht exakt identisch ist (Jacobs and Walsh, 2004). Grundsätzlich geht man davon aus, dass eine Ressource über mehrere unterschiedliche URIs identifiziert werden kann (Jacobs and Walsh, 2004), was als „**Non-Unique Name Assumption**“ (NUNA), also die „Annahme nicht-eindeutiger Namen“ bezeichnet wird (Hitzler et al., 2008, p. 65). Diese URIs nennt man „**URI aliases**“ (Heath and Bizer, 2011) oder „**co-referential URIs**“ (Halpin, 2013, p. 30). Diese sind im Web üblich bzw. überaus sinnvoll (Heath and Bizer, 2011). Agenten sollen von einer URI keine Eigenschaften ableiten, die sich auf die identifizierte Ressource per se beziehen (Jacobs and Walsh, 2004). Das heißt, dass ein Agent, der die URI <http://www.mainz.de/ZumKleinenSchluesel.html> dereferenziert, nicht erwarten kann, dass er tatsächlich eine **Web-Repräsentation** mit Informationen *über* ein historisches Gebäude erhält, die in HTML kodiert sind. Jacobs and Walsh (2004) bezeichneten dies als **URI-Undurchsichtigkeit**.

Der Begriff „Repräsentation“ wird in normativen Dokumenten nicht einheitlich gebraucht oder definiert. Nachfolgend wird die Definition von Halpin (2013) herangezogen, um anhand des Mechanismus zur „Inhaltsvereinbarung“ (engl. **content negotiation**) kontextbezogen zu erläutern, welchem Zweck **Web-Repräsentationen** im Web of Documents dienen. Dies ist nun möglich, da alle relevanten Termini in der folgenden Definition bereits eingeführt wurden. „A Web representation is the encoding of the content given by a resource given in response to a request that is subject to content negotiation, which must then include any headers that specify an interpretation, such as character encoding and media type“ (Halpin, 2013, p. 32). Eine Ressource kann eine, keine oder mehrere Web-Repräsentation(en) haben, die ggf. in unterschiedlichen Sprachen kodiert sein können (Jacobs and Walsh, 2004). Wenn jedoch derselbe Inhalt bzw. Sinngehalt kodiert wird, handelt es sich um verschiedene Web-Repräsentationen derselben Ressource, die dementsprechend über *eine* URI identifiziert wird (Halpin, 2013). Ein Agent kann die URI verwenden, um z.B. über HTTP Zugang zur referenzierten Ressource zu erhalten, was als **Dereferenzierung der URI** bezeichnet wird (Jacobs and Walsh, 2004). Dies ist die Definition aus dem normativen Dokument, das die Architektur des WWW definiert. Für diese Arbeit ist die Definition weder richtig noch falsch. Manchmal ist der „Zugang“ jedoch unmöglich. Darin liegt der Grund für die folgende Änderung: Unter der Phrase „**Dereferenzierung einer HTTP-URI**“, ist im Rahmen dieser Arbeit der Mechanismus zu verstehen, der es ermöglicht, Informationen über die identifizierte Ressource in Form von Web-Repräsentationen über ein Netzwerk zu transferieren. Welche Web-Repräsentation einer Ressource auf eine Anfrage eines Clients vom HTTP-Server ausgeliefert wird, kann von unterschiedlichen Kriterien abhängen. Im Rahmen der „Inhaltsvereinbarung“¹² werden die erwähnten HTTP-Header verwendet, um die adäquate Web-Repräsentation zu spezifizieren bzw. zwischen Client und Server auszuhandeln (Heath and Bizer, 2011). Die Definition lautet: „the process of selecting the best representation for a given response when there are multiple representations available“ (Fielding et al., 1999, p. 70). Das Beispiel illustriert diesen Prozess:

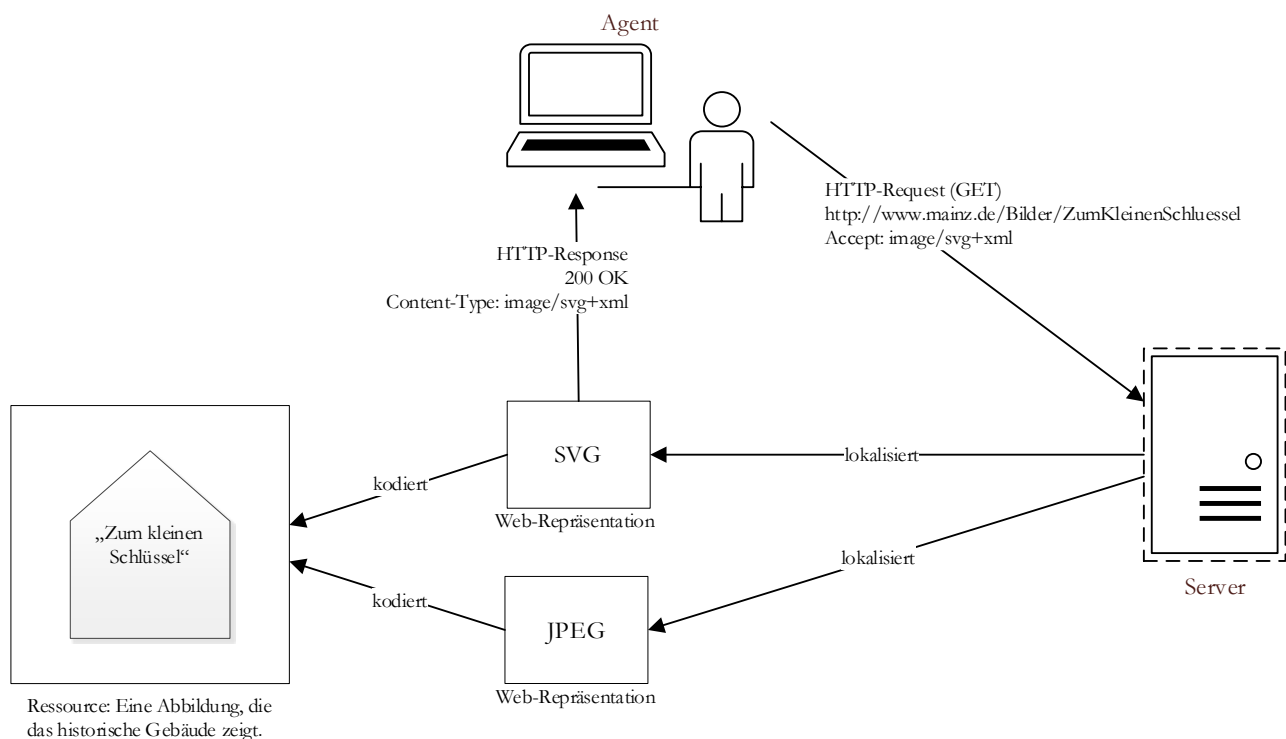


Abb. 2.5: „Inhaltsvereinbarung“, OA von Halpin (2013, p. 31), angepasste Darstellung

¹² Der Grund, warum dieser Begriff immer in Anführungszeichen steht, wird in Abschnitt 2.2.4 genannt. Dort wird die Terminologie erweitert.

Im Zuständigkeitsbereich wird die URL `http://www.mainz.de/Bilder/ZumKleinenSchluessel` konstruiert, um eine Ressource zu identifizieren, die Informationen *über* das historische Gebäude in Form einer Abbildung enthält. Im Dateisystem existiert das Bild in unterschiedlichen Formaten. Außerdem kann der Server zusätzliche Bildformate auf Anfrage generieren. Allerdings wird stets derselbe Inhalt kodiert. Die verschiedenen Formate oder Sprachen kodieren verschiedene Web-Repräsentationen derselben Ressource, die dementsprechend nur über eine URL identifiziert wird. Diese HTTP-URL wird vom Client dereferenziert, da er Zugang zu den Abbildungen über die HTTP-GET-Methode erhalten möchte. Allerdings bevorzugt der Client nicht die Rasterdarstellungen (z.B. im JPEG-Format), sondern die skalierbare Vektorgrafik (SVG). Er teilt dem Server diese Präferenz durch die Angabe „image/svg+xml“ im Request-Header-Feld „Accept“ mit. Der Server antwortet entsprechend und sendet dem Client die verfügbare SVG-Repräsentation der angefragten Ressource als Nutzlast (entity-body). Zusätzlich schickt der Server die erforderlichen Metainformationen (hier die Angabe „image/svg+xml“ im entity-header field „Content-Type“), welche der Client für die Interpretation benötigt.

Erwähnenswert ist außerdem, dass die „Inhaltsvereinbarung“ nicht auf formale Sprachen beschränkt ist. Angenommen es wurde die HTTP-URL `http://www.mainz.de/ZumKleinenSchluessel.php` konstruiert, um eine Ressource zu identifizieren, die textuelle Informationen *über* das historische Gebäude aus einer Datenbank extrahiert und bei entsprechenden Anfragen in HTML kodiert. Der Webserver wurde so konfiguriert, dass er die Informationen in deutscher und englischer Sprache ausliefern kann. Der Client macht davon Gebrauch, indem er seine bevorzugte natürliche Sprache im „Accept-Language-Feld“ der Anfrage festlegt. Die Antwort des Servers verwendet dann immer die Auszeichnungssprache HTML, trotzdem handelt es sich bei der englisch- / deutschsprachigen Webseite um verschiedene Web-Repräsentationen derselben Ressource.

2.1.3 Web-Architekturprinzipien

Berners-Lee (1996) erklärte: „By Universal I mean that the web is declared to be able to contain in principle every bit of information accessible by networks“. Dieses Zitat wird verwendet, um das erste Axiom der Web-Architektur (Berners-Lee, 1996) – das **Prinzip der Universalität** – einzuführen. Für das Web of Documents ist die (irrig) Annahme sinnvoll, dass Berners-Lee die obige Aussage auf digitale Ressourcen beschränkte, deren Inhalte in HTML-Repräsentationen kodiert werden können. Diese können als Nachricht über Netzwerkprotokolle versandt werden. Das Prinzip der Universalität besagt, dass grundsätzlich alle möglichen Ressourcen per URI identifiziert werden können und dass jede „signifikante Ressource“ (siehe GP 6.2.1) eine URI erhalten sollte (Berners-Lee, 1996). Im Web of Documents wurde das Prinzip umgesetzt, indem HTML-Dokumente per URL lokalisiert wurden, um Agenten den Zugang zu den enthaltenen Informationen per HTTP zu ermöglichen. Die Agenten konnten herausfinden, welche Inhalte sich hinter den undurchsichtigen URLs verbargen.

Das **Prinzip der Verlinkung** definierte Halpin (2013, p. 45) wie folgt: „The Principle of Linking states that any resource can be linked to another resource identified by a URI. No resource is an island, and the relationships between resources are captured by the linking, transforming lone resources into a Web. A **link** is a connection between resources“. Ein **Hyperlink** in HTML wird im Rahmen dieser Arbeit als spezielle Ausprägung eines Links angesehen, um diesen später von RDF-Links (und XLinks) mit anderen Eigenschaften abzugrenzen. Im Web of Documents wurde das Prinzip umgesetzt, indem HTML-Dokumente bzw. die darin enthaltenen Informationseinheiten über Hyperlinks verknüpft wurden (Baggio and Antonioli Corigliano, 2009). HTML-Dokumente können Hyperlinks enthalten, die weitere HTML-Dokumente **referenzieren**, welche auf den gleichen oder verschiedenen Rechnern existieren oder auch nicht (Heath and Bizer, 2011). Wenn das referenzierte HTML-Dokument existiert, dient der Hyperlink als **URL** vor allem dem **Zugang** zu den enthaltenen Informationen (Halpin, 2013). Ein Mensch könnte z.B. einen Webbrowser verwenden, um von einem HTML-

Dokument zum nächsten HTML-Dokument zu navigieren, indem er sich dieses anzeigen lässt (Bizer et al., 2009). Man bezeichnet diesen Vorgang als **Verfolgung eines Hyperlinks** (Halpin, 2013). Allgemeiner gesprochen können alle Agenten, also z.B. auch Programme wie Crawler, Hyperlinks verfolgen (Brin and Page, 1998). Wenn das referenzierte HTML-Dokument (in diesem Moment) nicht existiert bzw. zugänglich ist, spricht man von einem „**broken link**“ (Halpin, 2013, p. 45). Hyperlinks sind als wesentliche Voraussetzung für die Entstehung des Web of Documents anzusehen (Baggio and Antonioli Corigliano, 2009). Im Web gibt es keine Organisation o. Ä., bei der Links vor ihrer Veröffentlichung in einem zentralen Index registriert werden müssen (Heath and Bizer, 2011). Es gelten nur die o.g. Einschränkungen (bzgl. URI-Allokation). Da **kein zentraler Link-Index** existiert, wird aber auch nicht überprüft, ob die URL tatsächlich dereferenziert werden kann oder ob ein HTTP-Server letztlich einen Statuscode 404 sendet. Da die Konsistenz von Links nicht sichergestellt wird, qualifiziert sich das Web of Documents nach dem „Dexter Hypertext Reference Model“ (Halasz et al., 1994, p. 30) eigentlich nicht als vollwertiges Hypertextsystem (Halpin, 2013). Die Architektinnen des WWW räumten der **Skalierbarkeit** allerdings eine höhere Priorität ein als der **Link-Konsistenz** (Connolly, 2002a), was dazu führte, dass alle Agenten zum rapiden Wachstum des Web of Documents durch Verlinkung beitragen konnten, unabhängig davon, ob die Links tatsächlich physische Adressen zwecks Informationszugang lieferten.

Der Mensch muss häufig noch selbst entscheiden, ob er einem Hyperlink folgen möchte oder nicht. Diese Entscheidung ist nicht immer einfach, da Hyperlinks i.d.R. **nicht getypt** sind, d.h., in HTML ist nicht explizit spezifiziert, welche Art von Relation zwischen den beiden verknüpften Dokumenten bzw. zwischen den darin enthaltenen Informationseinheiten besteht (Bizer et al., 2009). Im Web of Documents wurden primär unidirektionale Hyperlinks verwendet (Halpin, 2013; Connolly, 2002a). Das Prinzip der Verlinkung beschränkt sich jedoch nicht auf Dokumente, die in HTML strukturiert sind, sondern es bezieht sich auf alle möglichen Ressourcen (Halpin, 2013). Damit werden z.B. auch die XML-basierten GML-Instanzen oder -Applikationsschemata erfasst, die im Geo-Web transferiert werden. Auch diese Ressourcen können auf unterschiedliche Art und Weise (bidirektionale, multidirektionale Verbindungen mit oder ohne spezifizierter Semantik) über XLinks verknüpft werden. Häufig wird von dieser Möglichkeit allerdings kein Gebrauch gemacht. In diesem Fall werden Geodatensilos geschaffen, die nicht ins Web integriert sind, da das Prinzip der Verlinkung nicht beachtet wurde.

Links sind auch für das **Prinzip der Selbstbeschreibung** essenziell, das Folgendes besagt: „individual documents become self-describing, in the sense that only widely available information is necessary for understanding them“ (Mendelsohn, 2007). Halpin (2013) leitete davon die Anforderung ab, dass alle Informationen, die ein Agent für die Interpretation einer Web-Repräsentation benötigt, über die Web-Repräsentation per se zugänglich sein müssen. Das bedeutet konkret für die Dereferenzierung einer HTTP-URL, dass die Antwort eines Webservers entweder alle erforderlichen Informationen und Metainformationen enthält oder dass diese über enthaltene Links zugänglich sind. In beiden Fällen müssen die Informationen und Metainformationen für die Interpretation ausreichend sein, d.h., der Agent kann diese verwenden, um den Sinngehalt der identifizieren Ressource zu ermitteln (Halpin, 2013). Wenn der Agent hierfür Links verfolgt, bezeichnet man den Vorgang als „follow your nose“ (Connolly, 2002b). Der Interpretationsprozess basiert auf Konzepten, die bereits vorgestellt wurden. Wenn in der Anfrage ein „fragment identifier“ verwendet wurde, interpretiert der Client diesen in Abhängigkeit vom „media type“ (Jacobs and Walsh, 2004), der normalerweise als Metainformation im „entity-header field Content-Type“ mitgeliefert wird (Fielding et al., 1999). Falls dieser fehlt, dürfen heuristische Lösungsverfahren verwendet werden, um einen wahrscheinlichen „media type“ zu ermitteln (Fielding et al., 1999). Man bezeichnet diesen Vorgang informell als „sniffing“ (Halpin, 2013, p. 20). Bemerkenswert ist, dass Fielding et al. (1999, p. 42) damit eine Ausnahme vom Prinzip der URI-Undurchsichtigkeit definierten: „the recipient MAY attempt to guess the media type via inspection of its content and/or the name extension(s) of the URI used to identify the resource“.

Das **Prinzip der offenen Welt** ist die Konsequenz aus dem Prinzip der Universalität und dem Prinzip der Verlinkung. Laut Halpin (2013) besagt es, dass die Zahl der Ressourcen im Web zunehmen kann, wenn neue Ressourcen hinzugefügt werden, damit sie mit anderen Ressourcen verlinkt werden können. Allemang et al. (2020, p. 31) charakterisierten das Web deshalb als offene Welt, in der die „Open World Assumption“ gilt. Hitzler et al. (2008, p. 145) übersetzten dies als sogenannte „Offene-Welt-Annahme (OWA)“, die besagt, „dass eine Wissensbasis immer potentiell unvollständig ist“. Deshalb gilt: Wenn nicht bewiesen werden kann, dass eine Aussage wahr ist, kann nicht angenommen werden, dass die Aussage falsch ist (Halpin, 2013). Da es keinen zentralen Index für alle URLs/Links gibt, können Agenten wie z.B. Suchmaschinen den Status des WWW zu einem bestimmten Zeitpunkt nicht exakt ermitteln, da das WWW vielleicht genau in diesem Moment, um neue Informationen erweitert wird (Halpin, 2013; Allemang et al., 2020). Insofern ist die OWA hinsichtlich des WWW durchaus nachvollziehbar. Es muss aber auch möglich sein, spezifischen Anwendungsfällen die „Closed World Assumption“ zugrunde zu legen, die z.B. aus dem Umfeld relationaler Datenbanken bekannt ist (Hitzler et al., 2008). Das W3C erarbeitete hierfür in jüngster Zeit entsprechende Empfehlungen (Allemang et al., 2020). Hitzler et al. (2008, p. 145) führten aus, dass nach der „Geschlossenen-Welt-Annahme“ davon auszugehen ist, „dass sämtliche relevanten Fakten in der Wissensbasis vorhanden sind“. Dann kann davon ausgegangen werden, dass eine Aussage falsch ist, wenn nicht bewiesen werden kann, dass die Aussage wahr ist (Halpin, 2013).

Das letzte Architekturprinzip definierte Halpin (2013, p. 48) wie folgt: „The **Principle of Least Power** states that a Web representation given by a resource should be described in the least powerful but adequate language“. Berners-Lee and Mendelsohn (2006) führten aus, dass diesem Prinzip im Rahmen der Standardisierung von HTML Rechnung getragen wurde. Der Grund ist, dass möglichst allen Agenten die Interpretation der bereitgestellten Information ermöglicht werden soll (Berners-Lee and Mendelsohn, 2006). Das Web ist ein offenes System, in dem unterschiedliche Agenten mit unterschiedlichen Fähigkeiten agieren (Allemang et al., 2020). Nicht alle Agenten können ausdrucksstarke Sprachen interpretieren (Halpin, 2013). Da die Eigentümerin einer Ressource nicht wissen kann, in welchen Anwendungsszenarien bzw. von welchen Agenten die bereitgestellten Informationen irgendwann verwendet werden, soll sie dafür Sorge tragen, die Informationen möglichst interoperabel bereitzustellen (Berners-Lee and Mendelsohn, 2006). Potenzielle Anwenderinnen sollen nicht von vornherein ausgeschlossen werden. Für die vorliegende Arbeit ist die Umsetzung in der LOD-Cloud wichtig (Ausdrucksmächtigkeit von Wissensrepräsentationssprachen), weshalb auf Details zur HTML verzichtet wird.

2.2 Web of Data

Aufgrund der erläuterten Web-Architekturprinzipien konnte das Web of Documents rasant expandieren und sich zu einem weltweiten Netz aus verknüpften Hypertext-Dokumenten entwickeln (Berners-Lee, 1994b). Der Mensch findet dort Daten, aus denen er Informationen gewinnen kann, was Berners-Lee (1994b) darstellte:

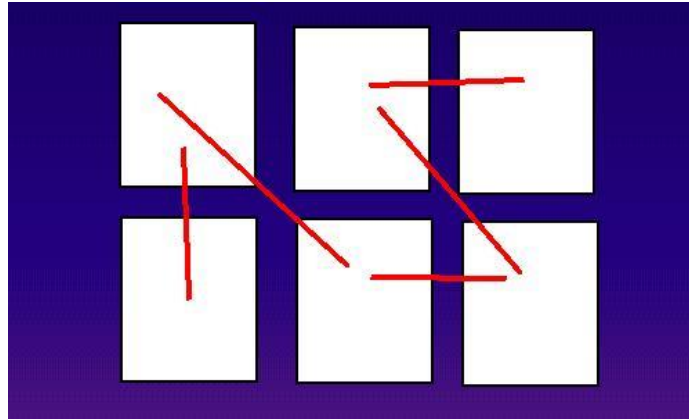


Abb. 2.6: Zustand des Web of Documents aus menschlicher Sicht, OA von Berners-Lee (1994b)

Berners-Lee betonte jedoch auch, dass das Potenzial damit noch lange nicht ausgeschöpft war und er hob einen wichtigen Aspekt des Web of Documents hervor: „To a computer, then, the web is a flat, boring world devoid of meaning“ (Berners-Lee, 1994b). Der Grund hierfür war, dass HTML verwendet wurde, um Daten zu strukturieren und zu verknüpfen. Einerseits war dies ein wesentlicher Faktor, dass dieses Entwicklungsstadium überhaupt erreicht werden konnte (Heath and Bizer, 2011); andererseits wurde die Weiterentwicklung des Web dadurch limitiert (Berners-Lee, 1994b). Die textuellen sowie multimedialen Inhalte in HTML-Dokumenten bzw. Webseiten waren auf den Menschen als Nutzerin ausgerichtet (Hitzler et al., 2008) und nicht auf die Interpretation durch Programme (Allemang et al., 2020). In syntaktischer Hinsicht konnten standardkonforme Webseiten von entsprechenden Applikationen (Browser, Parser, Crawler etc.) verarbeitet werden – die Semantik blieb bei der maschinellen Verarbeitung jedoch unberücksichtigt (Bizer et al., 2009). Der Sinngehalt konnte nur durch menschliche Interpretation ermittelt werden (Bizer et al., 2009). Deshalb charakterisierte Berners-Lee (1994b) das Web of Documents als „sinnlos“ für einen Computer. Er illustrierte diesen Zustand:



Abb. 2.7: Zustand des Web of Documents aus der Sicht eines Computers, OA von Berners-Lee (1994b)

Außerdem stellte er klar, dass die Essenz des Web eigentlich immer in den Informationen über alle möglichen Ressourcen lag, die untereinander in bestimmten Beziehungen stehen (Berners-Lee, 1994b); allerdings nicht in der Strukturierung von Daten mittels HTML und deren Verlinkung über ungetypte Hyperlinks (Halpin, 2013).

Webseiten können Informationen enthalten, die sich auf diese konkrete HTML-Repräsentation der digitalen Ressource beziehen, z.B. Angaben zur Autorin oder zum Erstellungszeitpunkt. Wesentlich häufiger finden sich in Webseiten jedoch Aussagen, Beschreibungen, Abbildungen, welche sich auf realweltliche, fiktive oder konzeptuelle Ressourcen – also **alle möglichen Ressourcen** – beziehen. Deshalb wurde in den bisherigen Beispielen immer betont, dass die jeweiligen Web-Repräsentationen Informationen *über* das historische Gebäude enthalten. Das Problem ist, dass dies in HTML-Dokumenten nicht formal ausgedrückt wird, ebenso wenig wie die Art der Beziehung zwischen Dokumenten oder Informationseinheiten formalisiert wird. Angenommen in der Webseite über das historische Gebäude findet sich die Aussage, dass dieses eine Mieterin hatte. Diese Aussage in einer natürlichen Sprache kann von einem Menschen interpretiert werden, da dieser *begreift*, was unter dem *Begriff* Mieterin zu *verstehen* ist und in welcher *Beziehung* eine Mieterin zu einem Haus steht, welche Rechte und Pflichten eine Mieterin hat etc. Eine Maschine kann diese Hintergrund- / Kontextinformationen und Zusammenhänge anhand eines „klassischen“ HTML-Dokuments, d.h. ohne semantische Erweiterungen, nicht ermitteln (Hitzler et al. 2008; Heath and Bizer, 2011). Es wurden Konzepte und Techniken (z.B. Microformate) entwickelt, die vorsahen, HTML-Dokumente semantisch zu annotieren, um z.B. die automatische Extraktion der enthaltenen Informationseinheiten über Programme zu realisieren (Heath and Bizer, 2011). Diese haben sich wegen ihrer zu geringen Ausdrucksmächtigkeit jedoch nicht wirklich etabliert (Heath and Bizer, 2011).

2.2.1 Intention, Abgrenzung

Das RDF hat sich etabliert und bildet das Herz des semantischen Netzes (Allemang et al., 2020). Berners-Lee erklärte: „Adding semantics to the web involves *two things*: allowing documents which have information in *machine-readable forms*, and allowing *links to be created with relationship values*. Only when we have this extra level of semantics will we be able to use computer power to help us exploit the information to a greater extent than our own reading“ (Berners-Lee, 1994b). RDF-basierte Informations-Spezifikations-sprachen sind für beide (o.g.) *Dinge* geeignet. Es handelt sich um formale Sprachen (Hitzler et al., 2008). Damit kann die Semantik spezifiziert werden (Hitzler et al., 2008). Es ist möglich RDF-Beschreibungen durch **maschinenverarbeitbare** Zeichenketten darzustellen, deren Syntax standardisiert wurde (Hitzler et al. 2008). Anwendungen können RDF-Beschreibungen austauschen (Hitzler et al. 2008). Die Semantik geht dabei nicht verloren (Hitzler et al. 2008). Im RDF werden URIs verwendet, um **alle möglichen Ressourcen** zu identifizieren (Allemang et al., 2020). Dadurch können **alle Agenten alle möglichen Aussagen über alle möglichen Ressourcen** treffen (**AAA-Slogan**) und dabei auch die **Beziehung zwischen Ressourcen explizit angeben** (Allemang et al., 2020). Trotz der prinzipiellen URI-Undurchsichtigkeit können Agenten die verwendeten **HTTP-URIs** (indirekt) **dereferenzieren**, um zu ermitteln, ob in einer digitalen RDF-Repräsentation Aussagen über eine digitale Ressource oder eine nicht digitale Ressource (das wäre dann z.B. das Haus „Zum kleinen Schlüssel“) getroffen werden. Im Gegensatz zum Web of Documents werden URIs im Web of Data nicht nur für die Lokalisierung von (HTML-)Dokumenten bzw. den Informationszugang via HTTP verwendet, sondern sie dienen auch der **Referenz** auf alle denkbaren Bezugsobjekte, die nicht immer über Netzwerke transferiert werden können (z.B. das o.g. Haus). „The difference between the use of a URI to **access** a hypertext web-page or other sort of information-based resource and the use of a URI to **refer** to some non-Web accessible entity or concept ends up being quite important, as this ability to representationally use URIs as ‘stands-in’ for referents forms the basis of the distinction between the hypertextWeb and the Semantic Web“ (Halpin, 2013, p. 32).

Berners-Lee (1994b) zeigte anhand der folgenden Abbildung, dass Maschinen im Web of Data zur eigenständigen Aufgabenerledigung befähigt werden sollen:

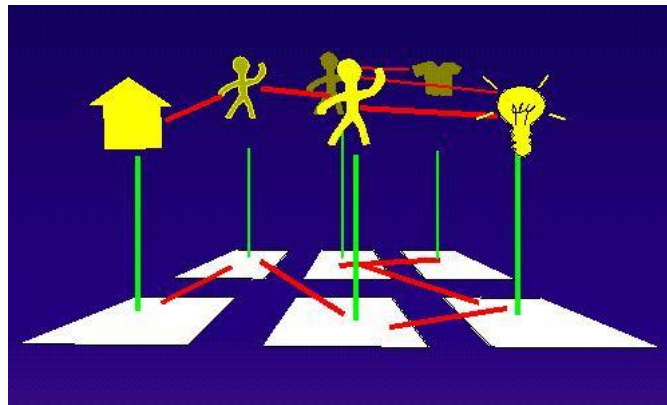


Abb. 2.8: Aufgabenerledigung im Semantic Web, OA von Berners-Lee (1994b)

Hier soll ein Programm *automatisch* ein „house“ (deutsch: Haus) ermitteln und den „transfer of ownership“ (deutsch: Eigentumsübergang) *selbstständig* „aushandeln“ (Berners-Lee, 1994b). Wie dargestellt, bleiben auch HTML-Dokumente Bestandteil des Semantic Web (vgl. GP 2.2.5), das auf den Web-Architekturprinzipien und Web-Standards aufsetzt (Bizer et al., 2009). Die folgende Beschreibung des semantischen Netzes wurde sehr häufig zitiert: „The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to **work in cooperation**“ (Berners-Lee et al., 2001, p. 3). Hitzler et al. (2008, p. 12) beschrieben die Intention hinter dem Semantic Web wie folgt: „Finde Wege und **Methoden**, Informationen so zu repräsentieren, dass Maschinen damit in einer Art und Weise umgehen können, die **aus menschlicher Sicht nützlich** und sinnvoll erscheint“. Hitzler et al. (2008) fokussierten das Nützliche (vgl. GP 2.2.2) und sie entschärften vor allem die Äußerungen der Architektinnen des (Semantic) Web. Diese führten häufig aus, dass das Semantic Web „*machine-understandable*“ sein soll (Berners-Lee, 1998). Hitzler et al. (2008, p. 48) hoben diesbezüglich hervor: „Eines der größten Missverständnisse im Umgang mit dem Semantic Web ist es vielleicht zu glauben, dass Computer [...] tatsächlich in die Lage versetzt würden, komplexe Konzepte [...] zu *verstehen*“. Wenn Maschinen die Konzepte wirklich verstehen sollen (vgl. GP 2.2.2), sieht man sich mit einer ganzen Reihe von Problemstellungen konfrontiert, die im Forschungsfeld der künstlichen Intelligenz fokussiert werden (Hitzler et al. 2008). Einige Erkenntnisse aus diesem Forschungsfeld beeinflussten die Entwicklung semantischer Basistechnologien (Halpin, 2013). Selbiges gilt auch hinsichtlich des Artikels, in dem die Idee bzw. die Vision vom Semantic Web vorgestellt wurde (Berners-Lee et al., 2001; Pohl and Danowski 2013). Dies wird auch anhand des obigen Beispiels deutlich (Abb. 2.8), welches auf die *vollautomatische programmgestützte* Aufgabenerledigung ausgerichtet ist. Mit dem subtilen Hinweis auf das deutsche (Grundstücks-)Recht wird betont, dass der beschriebene Prozess / diese (konkrete) Vision zumindest in Deutschland niemals zur Realität werden kann. Utopische Vorstellungen wichen jedoch der Erkenntnis, dass die Maschine im Semantic Web nicht zum *kognitiven* Interpretations-, Verknüpfungs- und Transformationsprozess (s.u.) befähigt werden muss (Hitzler et al., 2008; Allemang et al., 2020). Trotzdem können die u.g. Aufgaben im Semantic Web aufgrund maschineller **Unterstützung** effizienter erledigt werden (Hitzler et al., 2008).

Eine **Methode** ist eine planmäßig angewandte, begründete Vorgehensweise zur Erreichung eines Ziels unter Beachtung festgelegter Prinzipien (Hesse et al., 2020). Deshalb kann im Kontext dieser Arbeit unter Bezugnahme auf die Architektinnen des (Semantic) Web von der **LOD-Methode** gesprochen werden: „The first step is putting data on the Web in a form that machines can *naturally understand*, or converting it to that form. This creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines“ (Berners-Lee and Fischetti, 1999, p. 177). „Therefore, while the Semantic Web, or Web of Data, is the goal or

the end result of this process, Linked Data provides the means to reach that goal“ (Bizer et al., 2009, p. 18). Die LOD-Methode beruht auf den folgenden LOD-Prinzipien, die Berners Lee (2006) als Regeln bezeichnete:

- „1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things“.

Die LOD-Prinzipien wurden durch Heath and Bizer (2011) anhand konkreter Einsatzszenarien anschaulich erklärt und im Detail begründet. Die richtungsweisenden Ausführungen von Heath and Bizer (2011) fanden in vielen erfolgreichen LOD-Projekten Beachtung (Allemang et al., 2020). Berners Lee (2006) schlug ein simples 5-Sterne-Schema vor, das die wichtigsten Kriterien für die Generierung und die Publikation von LOD enthält:

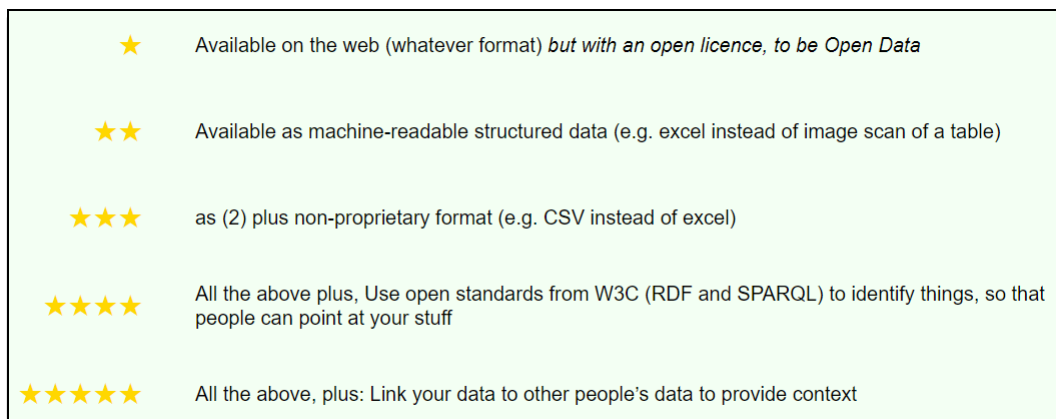


Abb. 2.9: 5 Sterne für LOD, OA von Berners-Lee (2006)

Für die Generierung und die Publikation von LOD werden Web-Standards eingesetzt, die bereits im Rahmen der Erörterung des Web of Documents vorgestellt wurden (s. Abschnitt 2.1). Diese Standards werden allerdings noch erweitert, was anhand der vollständigen Darstellung von Allemang et al. (2020) besonders deutlich wird:

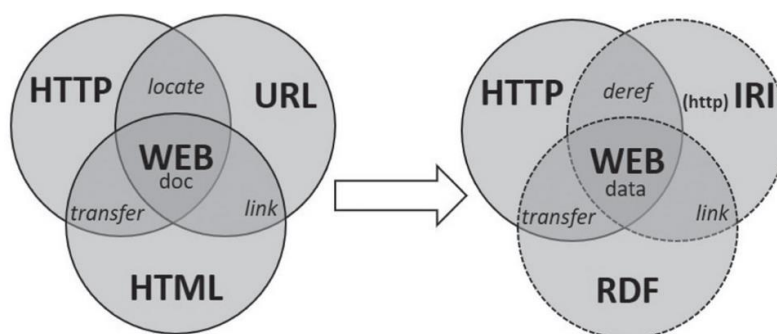


Abb. 2.10: Standards im Web of Documents / Web of Data, OA von Allemang et al. (2020, p. 86)

Es werden nicht nur HTML-Dokumente erstellt, per URL referenziert bzw. zugänglich gemacht und über Hyperlinks mit anderen Webseiten verknüpft, sondern RDF-Dokumente enthalten **dereferenzierbare HTTP-URIs** respektive IRIs, auf die alle Agenten im Web of Data **referieren** können (Allemang et al., 2020).

Die URI <http://www.mainz.de/id/743> identifiziert das „realweltliche“, historische Gebäude aus dem Beispiel. Das „eigene“ *RDF-Dokument* enthält Beschreibungen dieser Ressource, z.B. wird ausgesagt, dass dieses Gebäude im Jahr 1594 existierte. Ein anderer Agent im Web of Data könnte auf die o.g. URI referieren, um eine andere oder ergänzende Aussage über dieses Gebäude zu treffen. Er erstellt hierfür einen **RDF-Link** (Heath and Bizer, 2011). Dadurch werden die Bezugsobjekte in **Relation** gesetzt, wobei die Art dieser Beziehung explizit ausgedrückt wird (Bizer et al., 2009). Die Aussagen stammen dann zwar aus verschiedenen Quellen und liegen ggf. in unterschiedlichen Tripelstrukturen vor (s.u.), die RDF-Links verknüpfen aber die enthaltenen Ressourcen, die per URI (ggf. in unterschiedlichen Namensräumen) identifiziert wurden (Heath and Bizer, 2011). Somit kann ein globaler Informationsbestand aus verknüpften Ressourcen entstehen – das Web of Data bzw. die LOD-Cloud, in die sämtliche RDF-Beschreibungen integriert werden können (Heath and Bizer, 2011). Die LOD-Methode dient damit der umfassenden **Informationsintegration**. Je mehr Agenten die LOD-Prinzipien berücksichtigen, desto stärker ist der **Netzwerk-Effekt**, der zu einer Expansion des Web of Data führt (Allemang et al., 2020). Natürlich können (und sollen) auch im „eigenen“ *RDF-Dokument* bzw. -Graphen RDF-Links zu anderen Ressourcen eingefügt werden (Heath and Bizer, 2011). Allemang et al. (2020) hoben hervor, dass Agenten den RDF-Links folgen können, um Ressourcen zu entdecken, ohne überhaupt danach suchen zu müssen („follow your nose“). Für die **effiziente Suche** nach relevanten Informationen kann SPARQL verwendet werden (Harris and Seaborne, 2013). In dieser Arbeit sind einige SPARQL-Statements enthalten; die Syntax wird jedoch nicht erläutert, sondern es wird auf die o.g. W3C-Empfehlung verwiesen. Semantic Webstandards (s.u.) können für die Ableitung von **implizitem Wissen** aus einer formalisierten Wissensbasis i.V.m. ontologischen Schlussfolgerungsmethoden eingesetzt werden (Hitzler et al., 2008).

2.2.2 Wissensrepräsentation

Dieser Abschnitt dient der bündigen Darstellung – **nicht** der philosophischen Diskussion – der relevanten Punkte aus dem komplexen Themen- und Forschungsfeld „Wissensrepräsentation“. Dafür werden zunächst die abstrakten Definitionen (vgl. GP 2.1.2) der Termini Syntax und Semantik kontextbezogen¹³ konkretisiert.

[a] Die Syntax einer Sprache wurde als *System* eingeführt, das die Art und Weise festlegt, wie Informationen zu kodieren sind (Halpin, 2013). Für die *systematische* Kodierung sind „Regeln zur Strukturierung von Zeichen und Zeichenketten“ erforderlich (Hitzler et al., 2008, p. 13). In der vorliegenden Arbeit steht der Begriff Syntax primär¹⁴ „für eine Menge von Regeln, um [RDF-]Dokumente mit bestimmten Eigenschaften“ erzeugen zu können (Hitzler et al., 2008, p. 13). Die Syntax steht auf „der Codierungsebene [...] resp. Zeichenebene“ (Bill, 2016, p. 363). In der sog. Wissenspyramide ist dies die Basis bzw. die unterste Ebene (vgl. Abb. 2.11). Eine Menge von Regeln wird in dieser Arbeit als *strenges Regelwerk* bezeichnet. Die essentiellen Passagen aus den erörterten W3C-Empfehlungen und RFCs, die als „normativ“ charakterisiert wurden, sind strenge Regelwerke. Zeichen resp. Zeichenketten müssen in der **Interpretation** und **Informationsgewinnung** strengen Regeln folgen (Bill, 2016). Dies ist die Grundvoraussetzung, damit die vorgeannten **Prozesse**, die zwangsläufig das „Wissen eines Wissensträgers“ erfordern (Bill, 2016, p. 3), erfolgreich verlaufen können. Das Wissen bildet allerdings die Spitze der sog. Wissenspyramide (vgl. Abb. 2.11).

¹³ Diese Aussagen erheben keinesfalls Anspruch auf Allgemeingültigkeit, sondern sie werden hinsichtlich des LOD-Paradigmas getroffen. Zu Gunsten einer anschaulichen und verständlichen Erörterung wird der Umfang und der Komplexitätsgrad auf das notwendige Minimum reduziert.

¹⁴ Das Wort „primär“ bezieht sich auf das zweite operative Teilziel: Da dieses der Generierung des RDF-Graphen dediziert ist, welcher in dieser Arbeit hauptsächlich mithilfe der RDF/XML-Serialisierung (vgl. GP 2.2.3) abgebildet wird, stehen RDF-Dokumente im Fokus. Die Syntax der SPARQL-Ausdrücke muss ebenfalls korrekt sein, da die Selektion/Konstruktion anderenfalls scheitert. Selbiges gilt bzgl. XQuery etc. pp. Aufgrund der Wichtigkeit von XML für diese Arbeit, wird XML für die Erklärung verwendet.

Ein Beispiel für eine essentielle Passage aus einem strengen Regelwerk sind die Abschnitte 1-6 i.V.m. den Annexen A und B der W3C-Empfehlung für die (Auszeichnungs-)Sprache XML (Bray et al., 2008). Dort werden alle Regeln festgelegt, die ein XML-Dokument (insgesamt) erfüllen muss, um als wohlgeformt zu gelten (Bray et al., 2008). Die Wohlgeformtheit (Übereinstimmung mit der definierten Syntax) ist die Grundvoraussetzung, damit ein XML-Dokument von Programmen (z.B. von Parsern) – also von Maschinen – korrekt verarbeitet/abgearbeitet, aber nicht zwangsläufig *interpretiert* (Definition: S. 11), werden kann.

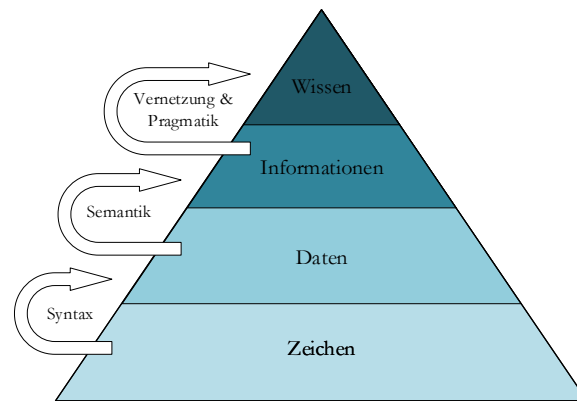


Abb. 2.11: Wissenspyramide, OA von Bill (2016, p. 363), angepasste Darstellung

[b] Die Semantik einer Sprache wurde als System eingeführt, in dem der „*content of information*“ (Definition: S. 11) systematisch miteinander verbunden ist (Halpin, 2013, p. 89). Im Bereich des Semantic Web „versteht man unter Semantik die *Bedeutung* von *Worten* bzw. Zeichen(-ketten) und ihre Beziehungen untereinander“ (Hitzler et al., 2008, p. 13). Zur Informationsgewinnung müssen Daten, die aus Zeichen(-ketten) bestehen (vgl. Abb. 2.11), in einem Kontext auf der *Bedeutungsebene* interpretiert werden (Bill, 2016). Tschirner (2012, p. 13) bezeichnete die *Worte* als „Konzeptwörter“ und den *Kontext* als „Bedeutungskontext“. Halpin (2013, p. 89) bezog sich auf Gottlob Frege und schlussfolgerte: „A sense is objective insofar as it is a shared part of an *inherently public* language, since a sense is the common property of many people, and so is **not** a part of a mode of the **individual** mind“. Diesbezüglich muss Folgendes pointiert werden: „It is precisely this notion that sense – and therefore **meaning as whole** – is ‘objective’ that is crucial for [the] meaning on the Web. The Fregean notion of sense is identical with [the] notion of *informational content*“ (Halpin, 2013, p. 89). Darauf aufbauend wird der Kontext – speziell für diese Arbeit – wie folgt unterteilt:

- *Allgemeiner Bedeutungskontext*: Hier soll die allgemein anerkannte wohldefinierte Bedeutung der Konzeptwörter, welche in universellen Vokabularen (s.u.) festgelegt ist, *die von Natur aus öffentlich sind*, **nicht** von einem **speziellen** Kontext oder von der **individuellen** Interpretation einzelner Agenten abhängen. Ein Beispiel hierfür ist das Prädikat `owl:sameAs`, dessen Semantik im OWL-Vokabular spezifiziert wurde. Wenn es eingesetzt wird, um auszudrücken, dass zwei URI-Referenzen auf denselben Referenten referieren, dann kann eine Maschine automatisch *ableiten*, dass der Referent dieselben Eigenschaften hat.
- *Spezifischer Anwendungskontext*: Hier kann die individuelle Bedeutung der Konzeptwörter in domänen- oder problemspezifischen – nicht unbedingt öffentlichen – Vokabularen (s.u.) festgelegt werden, um Unklarheiten zu vermeiden bzw. um die Interpretation zu ermöglichen (Hitzler et al., 2008; Tschirner, 2012). Als Beispiel hierfür dient das folgende leere Element, das in einem XML-Dokument verwendet werden kann: `<mine/>`. Ein Mensch assoziiert damit evtl. eine Waffe; ein anderer Mensch denkt evtl. an ein Bauteil von Schreibgeräten. Schade and Cox (2010) erörterten die Äquivalenz bzw. den Isomorphismus zwischen RDF und GML anhand von XML-Serialisierungen (vgl. GP 2.3.3). Schade and Cox (2010) zeigten (u. a. am Beispiel einer Mine), wie Zeichenketten/Wörtern eine Semantik zugeordnet

werden kann, indem sie mit den korrespondierenden Konzepten verknüpft werden. Hitzler et al. (2008) betonten, dass die Maschine dadurch nicht befähigt wird, ein komplexes Konzept wie „Mine“ zu *verstehen* – die (inhaltliche) Interpretation der kodierten Aussagen obliegt immer noch dem Menschen. Für die Interpretation können die terminologischen Zusammenhänge mithilfe von universellen Vokabularen bzw. formalen Modellierungssprachen spezifiziert und ins Web integriert werden, das auf dem Prinzip der *Selbstbeschreibung* (GP 2.2.5) basiert. Anhand von Lis. 2, Lis. 3 in der Arbeit von Schade and Cox (2010) kann ein Mensch *verstehen*, dass die Mine im spezifischen Kontext der Rohstoffgewinnung steht.

Der Grund für diese Differenzierung liegt darin, dass den universellen Vokabularen die höchste Priorität zukommt. Ein gemeinsames, objektives Verständnis der darin spezifizierten Konzeptwörter ist die *Grundvoraussetzung* für die Erstellung domänen- und/oder problemspezifischer Vokabulare. Anderenfalls ist es unmöglich das Chaos, welches aus der Natur des WWW per se und konkret aus dem AAA-Slogan hinsichtlich des Semantic Web resultiert, zu organisieren (Allemang et al., 2020). Die Notwendigkeit dafür wurde schon vor dem Aufkommen des Semantic Web erkannt (Brachman and Smith, 1980). Ist die *Grundvoraussetzung* nicht erfüllt, können Agenten in Netzwerken nicht sinnvoll kommunizieren (Allemang et al., 2020). Dies wird am Beispiel früher Ausprägungen semantischer Netzwerke deutlich: „When researchers attempted to communicate or combine their knowledge representation schemes, no-one really knew what the natural language description ‘meant’ except the author, even when semantic networks were used as a formal language“ (Halpin, 2013, p. 55). „The crucial context for **usage** of a particular semantic network was always **lost in transfer**“ (Halpin, 2013, p. 60).

Aufgrund der Leitfrage dieser Arbeit ist der Begriff Information essentiell (GP 1.2). Es wurde klargestellt, dass im Zuge der Demonstration (drittes Teilziel) nutzbare Antworten auf konkrete Fragen gegeben werden müssen. „Information ist nutzbare Antwort auf eine konkrete Fragestellung“ (Zehnder, 1998, p. 14). Es existieren unzählige Definitionen des Begriffs Information. Häufig wird aber auf den Zweck und/oder den **Nutzen** in einem bestimmten Kontext abgestellt. Bill (2016, p. 3) bezeichnete Information als „zweckbezogenes *Wissen*“.

„Das *Wissen* eines Wissensträgers“ begrenzte Bill (2016, p. 3) auf „die Menge aller von ihm als wahr angenommenen Aussagen über die repräsentierte Welt, die tatsächlich wahr sind“. Davon wird in der vorliegenden Arbeit abgewichen. Zwecks Begründung wird auf GP 2.1.3 i.V.m GP 2.2.5 (OWA) verwiesen. Außerdem wird hervorgehoben, dass von vornherein klargestellt wurde, dass der Terminus Aussage im Sinne eines Hyperonyms verwendet wird, das vor allem Behauptungen einschließen soll. Der Grund wird unter GP 3.1.4 genannt. Auf die (zwangsläufig philosophische/erkenntnistheoretische) Diskussion des Wissensbegriffs wird verzichtet.

Die o.g. **Prozesse** (siehe [a]) zielen auf die Gewinnung von Informationen ab (Bill, 2016). Daten können von einem Wissensträger durch die Anwendung von Regeln und (Transformations-)Anweisungen zu sinnvollen und verwertbaren Informationen gewandelt werden, die in einem spezifischen Anwendungskontext **genutzt** werden können (Bill, 2016). Der Wissenspyramide (Abb. 2.11) wurde der Begriff Pragmatik hinzugefügt. Damit wird (allgemein) die Orientierung auf das **Nützliche** betont. Außerdem wurde der Darstellung der Begriff **Vernetzung** hinzugefügt, womit (allgemein) betont wird, dass der kognitive Interpretations-, Verknüpfungs- und Transformationsprozess vom Wissen eines Wissensträgers und (ggf.) von dessen Erfahrungen abhängig ist. Speziell für die vorliegende Arbeit wird – in Anbetracht der Leitfrage – mit dem Begriff Vernetzung zusätzlich die kollaborative Informationsintegration und Wissensvernetzung in der **LOD-Cloud** hervorgehoben.

Information ist laut Bill (2016, p. 3) an ein **Informationsmittel** gebunden. Bill (2016) führte die **Sprache** als Beispiel für ein Informationsmittel an. Im Semantic Web spielen „Informations-Spezifikationsprachen“, die auch als „Wissensrepräsentationssprachen“ bezeichnet werden, eine *bedeutende* Rolle (Hitzler et al., 2008, p. 12). Die beiden o.g. Begriffe bezeichnen formale Sprachen, die in Form von universellen, standardisierten Vokabularen resp. W3C-Empfehlungen vorliegen (vgl. Abb. 2.12). Diese werden für die Spezifikation der Semantik

der Konzeptwörter/Fachwörter (die allgemein als *Termini* bezeichnet werden) und der *terminologischen* Zusammenhänge eingesetzt (Hitzler et al., 2008). Bezugnehmend auf die obige Differenzierung (siehe [b]) wird klargestellt, dass die W3C-Empfehlungen keine fachspezifischen Vokabulare sind, sondern sie stellen universelle Ausdrucksmittel bereit, welche es ermöglichen, „innerhalb eines [Semantic Web] Dokuments Aussagen über die semantischen Beziehungen der Termini eines beliebigen nutzerdefinierten Vokabulars zu machen. Damit kann ein neues Vokabular definiert und zumindest ein Teil seiner „Bedeutung“ im [Semantic Web] Dokument spezifiziert werden [...]“ (Hitzler et al., 2008, p. 67). Trotz der Wichtigkeit formaler Sprachen wird klargestellt, dass natürlichsprachliche Beschreibungen in Englisch (Lingua franca), welche für die Interpretation durch einen Menschen bestimmt sind, in Semantic Web Dokumenten weiterhin ihre Daseinsberechtigung haben (vgl. [b])¹⁵.

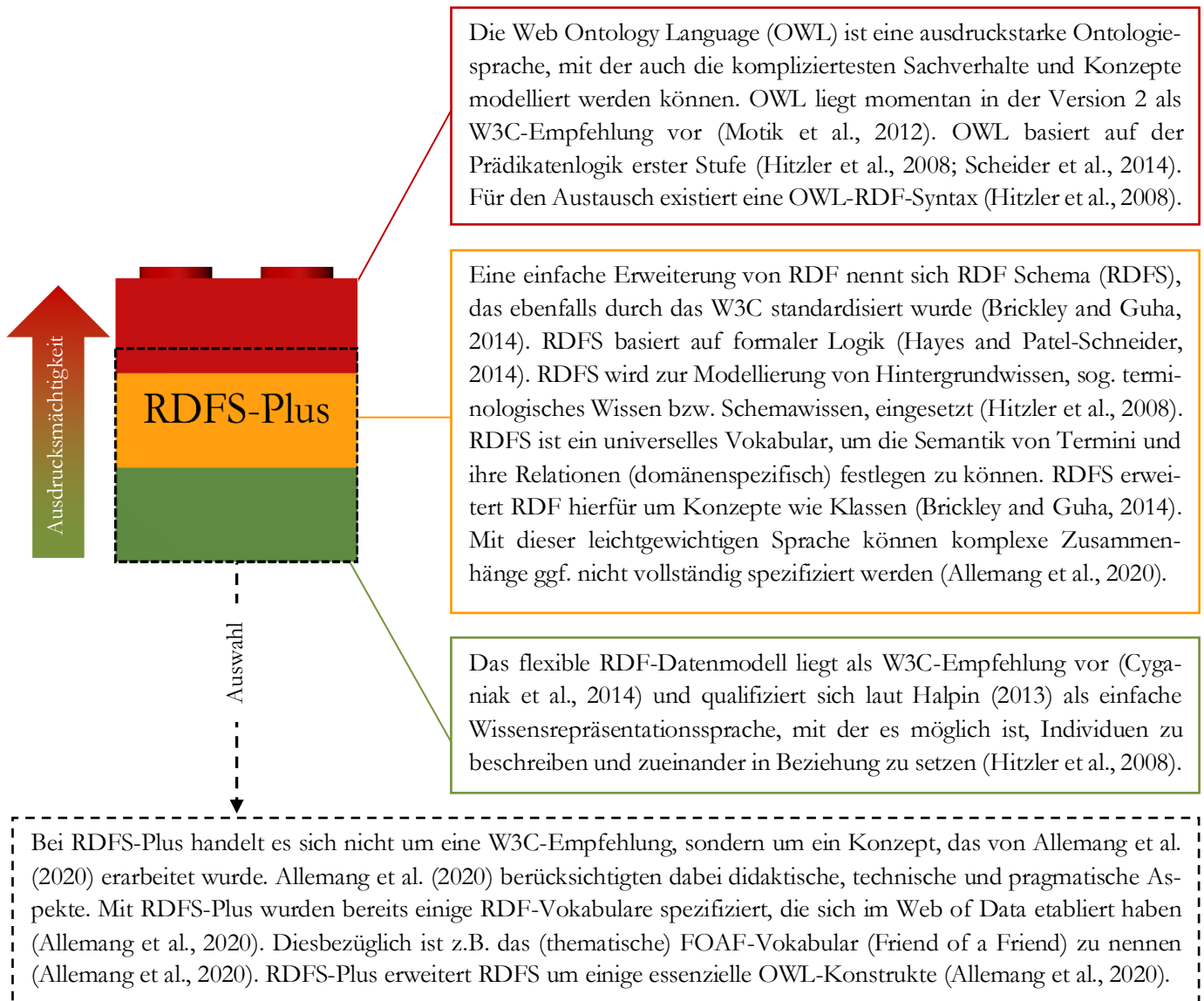


Abb. 2.12: Ausgewählte Wissensrepräsentationssprachen, ED

Mit dieser Übersicht wird auf eine detaillierte Differenzierung von OWL und auf die Auflistung aller einschlägigen W3C-Empfehlungen, z.B. zum Simple Knowledge Organization System (SKOS), verzichtet. SKOS wird ausgeklammert, da im Zuge der Anreicherung RDF-Links¹⁶ auf der Instanzebene fokussiert werden (s. GP 1.2). Im Übrigen wird auf die Ausführungen zu Vocabulary Links (s. GP 7.2.4) / zur TBox (s. GP 7.2.2) verwiesen.

¹⁵ Ein Semantic Web Dokument kann hier als RDF-Dokument angesehen werden. Die Details dazu folgen unter GP 2.2.4.

¹⁶ RDF-Links werden unter GP 2.2.3 erklärt; sie werden unter GP 6.2.1 exakt definiert.

Zusätzlich zu den genannten W3C-Empfehlungen hat das Konsortium eine Reihe von Dokumenten (sog. Primer) erarbeitet, welche die jeweiligen Konzepte anschaulich darstellen (Schreiber and Raimond, 2014). Nachfolgend werden nur ausgewählte Sprachmittel von RDFS-Plus anhand des Schlüsselbeispiels erklärt, da diese für den praktischen Teil wichtig sind. Hierfür muss noch auf den Begriff Ontologie eingegangen werden.

Eine allgemeine Definition lieferte Gruber (1995, p. 908): „An ontology is an explicit specification of a conceptualization“. Für diese Arbeit ist die speziellere Definition von Guarino et al. (2009, p. 2) relevant: „Computational ontologies are a means to formally model the structure of a system, i.e., the relevant entities and relations that emerge from its observation, and which are useful to our purposes“. Konkret für diese Arbeit – ohne Anspruch auf Allgemeingültigkeit – kann unter einer *simplem Ontologie* eine (maschinenverarbeitbare) Beschreibung der *Konzepte* (TBox) verstanden werden, die mithilfe einer leichtgewichtigen Wissensrepräsentationssprache erstellt wird, um den Austausch bzw. das Teilen von Informationen zu ermöglichen (Primärzweck) und um implizite Aussagen über einfache Schlussfolgerungsmethoden ableiten zu können (Sekundärzweck).

Da im Rahmen dieser Arbeit historische Zusammenhänge modelliert werden, können unter dem Begriff *Konzept* alle Klassen, Prädikate und Datentypen subsumiert werden, die erforderlich sind, um individuelle, historische Entitäten und ihre Beziehungen untereinander auf der Instanzebene beschreiben zu können, sodass sie maschinell verarbeitet und von menschlichen Nutzerinnen verstanden werden können. Die Instanzebene wird in ihrer Gesamtheit als „assertional box (ABox)“ bezeichnet (Nys et al., 2018, p. 39), was frei als Menge an „feststellende[n] Behauptung[en]“ übersetzt werden kann (Tschirner, 2012, p. 8). Die korrespondierenden Konzepte sind in der „TBox“ enthalten, welche die Gesamtmenge an terminologischem Schemawissen umfasst (Nys et al., 2018, p. 38). Die TBox kann auch als Vokabular bezeichnet werden und, da im konkreten Fall stets RDF für die Beschreibung der Konzepte verwendet wird, als RDF-Vokabular (Hitzler et al., 2008). Dieses wird als Wissensmodell bezeichnet, wenn der Organisations- / Kommunikationsaspekt betont werden soll (Allemang et al., 2020). Die Bezeichnung Wissensbasis schließt sowohl die ABox als auch die TBox ein (Nys et al., 2018).

Der Vorgang zwecks Formalisierung von Konzepten wurde schon als „Konzeptualisierungsprozess“ bezeichnet. Im Sinne eines Oberbegriffs soll die o.g. Bezeichnung einige Tätigkeiten zwecks Erstellung einer Wissensbasis umfassen, insbesondere die Nachnutzung von Konzepten aus existierenden, etablierten RDF-Vokabularen, das „Ontologie-Alignment“ (s.u.) und das Mapping mithilfe adäquater Domänenontologien (s. GP 2.4.2).

Diesbezüglich ist Folgendes erwähnenswert:

- (1) Es ist sinnvoll Konzepte aus RDF-Vokabularen nachzunutzen, die sich im Web of Data etabliert haben (Heath and Bizer, 2011). Dies ist sowohl in der ABox als auch in der TBox möglich (Tschirner, 2012).
- (2) Guarino (1998) unterschied vier Arten von Ontologien, von denen im konkreten Fall drei wichtig sind:
 - a. **Upper-Ontologien**, die globale, domänenübergreifende Konzepte formalisieren, um möglichst viel Wissen der Welt einzubeziehen bzw. zu ordnen (Damova et al., 2010). Mit diesen übergeordneten, abstrahierten Konzeptualisierungen lassen sich die Wissensbasen unterschiedlicher Fachbereiche überbrücken bzw. verknüpfen (Lieberman, 2006). So ist z.B. die bereits vorgestellte IRW-Ontologie mit Konzepten aus der „Descriptive Ontology for Linguistic and Cognitive Engineering“ (DOLCE) verknüpft¹⁷ (Halpin and Presutti, 2009), die sich als Upper-Ontologie qualifiziert (Borgo and Masolo, 2010, p. 1).

¹⁷ Der Korrektheit halber aber nicht des allgemeinen Verständnisses halber: Es handelt sich um die leichtgewichtige Version davon, welche „DOLCE Ultra Lite“ genannt wurde (Halpin and Presutti, 2009, p. 1). Genauer gesagt handelt es sich um eine Erweiterung der o.g. Upper-Ontologie, die als „Information Object Lite“ bezeichnet wurde (Halpin and Presutti, 2009, p. 1).

Beispiel für ein ähnliches Semantic Web Dokument: <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl>

- b. **Domänenontologien**, zu denen Lieberman (2006) alle Ontologien zählte, die Spezialwissen eines Fachbereichs umfassen. Die enthaltenen Konzepte sind im betreffenden Fachbereich gültig und bekannt (Tschirner, 2012). Domänenontologien können sich auf die fachbereichsspezifischen Anforderungen konzentrieren und müssen nicht zwangsläufig mit Konzepten anderer Domänen in Relation gesetzt werden. Letzteres ist aber vor allem im (interdisziplinären) geowissenschaftlichen Bereich häufig ein erstrebenswertes Ziel (Hbeich and Roxin, 2020). Die Semantic Web for Earth and Environmental Terminology (SWEET)¹⁸ ist ein Beispiel für eine „geospatial domain ontology“ (Stock and Guesgen, 2016, p. 173).
- c. **Applikationsontologien**, die stets für eine individuelle Problemstellung entwickelt werden, bezeichnet man auch als Problemontologien (Lieberman, 2006; Tschirner, 2012). Eine Problemontologie ist eine spezielle Ausprägung einer Domänenontologie (Lieberman, 2006).

Es ist möglich Ontologien zu verknüpfen (Jain et al., 2011). Man bezeichnet dies als **Ontologie-Alignment** (Jain et al., 2011). RDFS-Plus bietet hierfür entsprechende Konstrukte respektive Sprachmittel, mit denen z.B. Äquivalenzen von Konzepten in unterschiedlichen Vokabularen ausgedrückt oder Spezialisierungen vorgenommen werden können (Allemang et al., 2020).

2.2.3 RDF – Grundlagen

Mithilfe von Wissensrepräsentationssprachen können Ontologien erstellt werden. Halpin (2013, p. 51) definierte diese Sprachen kontextbezogen und hob einen wichtigen Aspekt hervor: „a language whose primary purpose is the representation of non-digital content in a digital encoding“. Benötigt wird also eine Sprache, mit der z.B. ausgedrückt werden kann, dass das historische Gebäude (nicht digital) aus dem Schlüsselbeispiel einen Bewohner (nicht digital) hatte und zwar in einer digitalen Form, die von Maschinen verarbeitet werden kann. Die Grundzüge des RDF-Datenmodells werden nachfolgend anschaulich anhand des Schlüsselbeispiels erläutert, das zunächst abstrakt dargestellt wird:

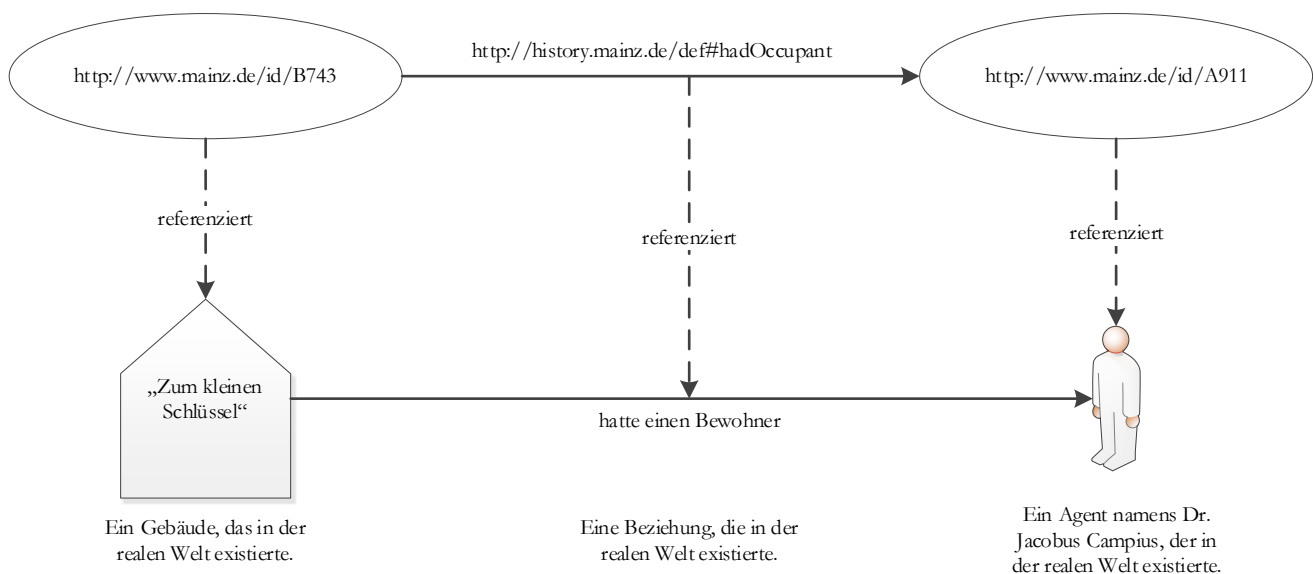


Abb. 2.13: Ressourcen & Beziehungen, OA von Halpin (2013, p. 51), angepasste Darstellung

¹⁸ Siehe: <https://github.com/ESIPFed/sweet>

Das RDF hat sich als *Lingua franca* etabliert (Heath and Bizer, 2011). Es ist der kleinste gemeinsame Nenner, auf den man sich verständigt hat, um die Informationsintegration in der LOD-Cloud zu ermöglichen. Dies führte zu einem rapiden Wachstum, was die folgende Abbildung eindrucksvoll belegt. Der Zustand der LOD-Cloud im Jahr 2007 (links) wird dem (relativ) aktuellen Status (rechts) gegenübergestellt. Die LOD-Cloud ist mittlerweile derart umfassend, weshalb die Details in diesem Dokument nicht mehr dargestellt werden können.

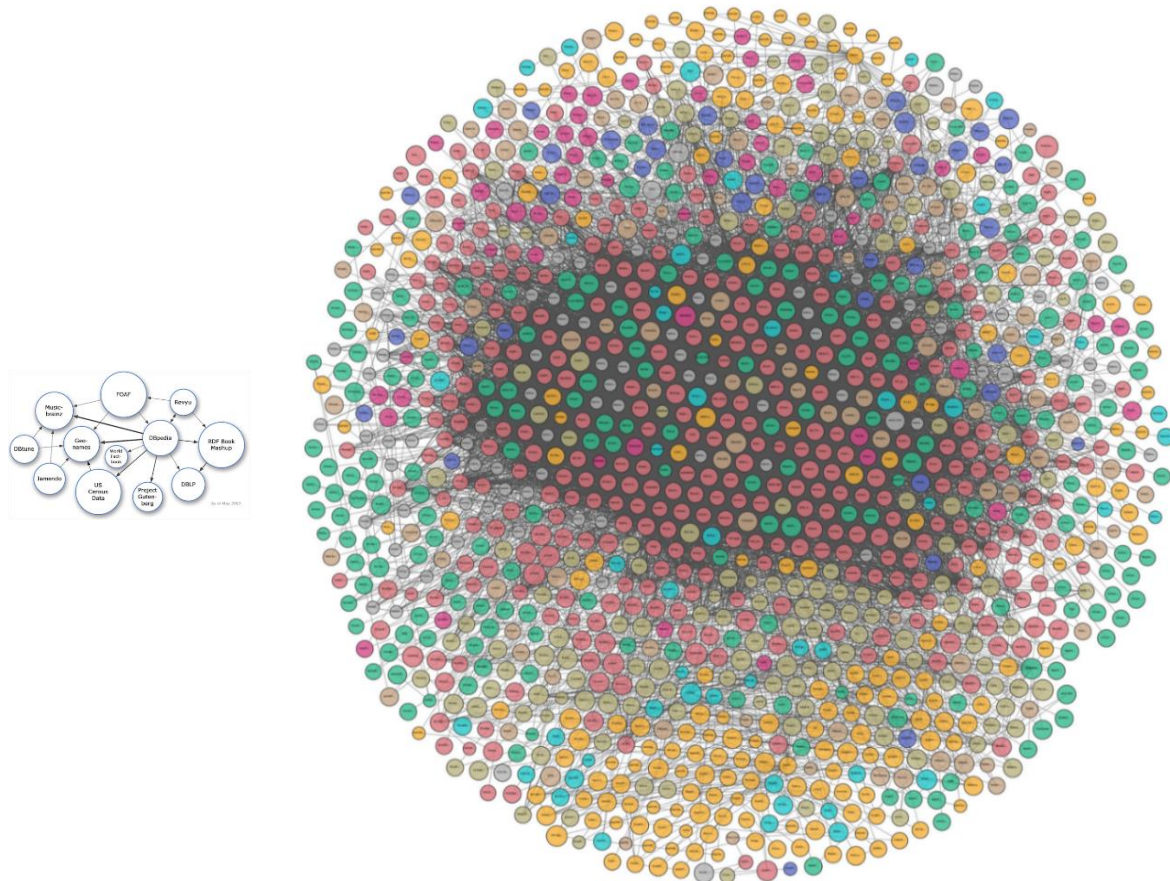


Abb. 2.14: Netzwerk-Effekt in der LOD-Cloud, Graphiken von lod-cloud.net

Das RDF dient primär der Informationsintegration (Halpin, 2013; Heath and Bizer, 2011; Allemang et al., 2020). Anhand einer Aussage im RDF kann anschaulich erklärt werden, warum Informationsbestände, die dezentral erstellt und vorgehalten werden, trotzdem sehr einfach verknüpft werden können. Im RDF werden **Graphen** beschrieben, die aus Knoten und gerichteten Kanten bestehen (Hitzler et al., 2008). Die folgende Abbildung stellt die o.g. Aussage graphisch dar (vgl. Abb. 2.13). Knoten werden durch Ovale repräsentiert und die Kante ist als Pfeil dargestellt (Hitzler et al., 2008).

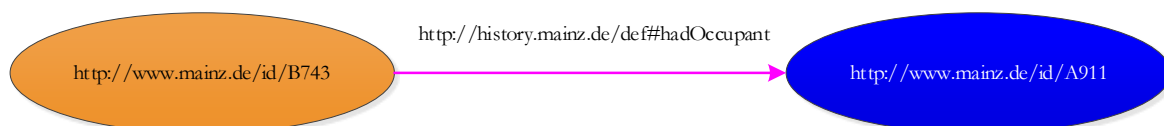


Abb. 2.15: Ein RDF-Tripel, OA von Hitzler et al. (2008, p. 36), angepasste Darstellung

Knoten und Kanten sind Ressourcen (Halpin, 2013), die im RDF *grundsätzlich* über URIs identifiziert werden (Hitzler et al., 2008). Gemäß den LOD-Prinzipien wurden die Knoten und die Kante in der obigen Graphik mit dereferenzierbaren HTTP-URIs beschriftet (Heath and Bizer, 2011). Insgesamt bezeichnet man die beiden

Knoten und die verbindende Kante als **Tripel** (Hitzler et al., 2008). Ein Tripel ist **eine Aussage** im RDF (Halpin, 2013). Den Knoten, der mit der URI <http://www.mainz.de/id/B743> beschriftet und orange eingefärbt ist, bezeichnet man als **Subjekt**; die Kante ist das sog. **Prädikat** (pink) und den blauen Knoten mit der URI <http://www.mainz.de/id/A911> nennt man **Objekt** (Hitzler et al., 2008). Das RDF ist eine formale Sprache (Halpin, 2013). Ein Tripel weist jedoch Ähnlichkeiten zu einfachen, natürlichsprachlichen Aussagen auf (Heath and Bizer, 2011). Das Prädikat – der **RDF-Link** – gibt die Beziehung zwischen Subjekt und Objekt an. Ein RDF-Link ist, wie die meisten HTML-Hyperlinks, unidirektional (Halpin, 2013). Wie HTML-Hyperlinks können RDF-Links Ressourcen referenzieren, die (ggf.) von unterschiedlichen Agenten auf verschiedenen Servern bereitgestellt werden (Heath and Bizer, 2011). Dies ist nachfolgend exemplarisch dargestellt:



Abb. 2.16: RDF-Links in RDF-Tripeln, ED

Hier wird eine Ressource im eigenen RDF-Dokument mit einer Ressource verknüpft, die in der Gemeinsamen Normdatei (GND) geführt und als LOD ins Web integriert wurde. Das Prädikat `owl:sameAs` ist ein Beispiel für einen RDF-Link, der zwei „Datenbestände“ verknüpft. Genauer gesagt werden dadurch zwei Graphen vereinigt (Hitzler et al., 2008). In der LOD-Cloud werden auf diese Weise sehr viele Graphen vereinigt; sie manifestiert sich dementsprechend als gigantischer, globaler Graph (Heath and Bizer, 2011). Anhand des Beispiels wird aber auch deutlich, dass Ressourcen im RDF sehr flexibel eingesetzt werden können (Hitzler et al., 2008). Hier ist die Ressource mit der URI <http://www.mainz.de/id/A911> sowohl Objekt als auch Subjekt. Ein gerichteter Graph ist nicht hierarchisch aufgebaut; er verfügt nicht über einen Ursprungsknoten (Hitzler et al., 2008; Tschirner, 2012). Das RDF ist für die Komposition dezentraler Informationen prädestiniert (Hitzler et al., 2008). Dies wurde hier betont. Der Plural soll klarstellen, dass **alle Prädikate RDF-Links** sind.

Das RDF hat den Status einer W3C-Empfehlung (Cyganiak et al., 2014). Das RDF basiert auf etablierten Web-Standards wie z.B. URIs (Halpin, 2013). Die einzigen relevanten Ausnahmen vom Grundsatz, dass alle Ressourcen im RDF per URI identifiziert werden, sind die leeren Knoten und die sog. Literale. Letztere dienen der Darstellung von Datenwerten im RDF (Hitzler et al., 2008). Dies ist nachfolgend exemplarisch dargestellt:

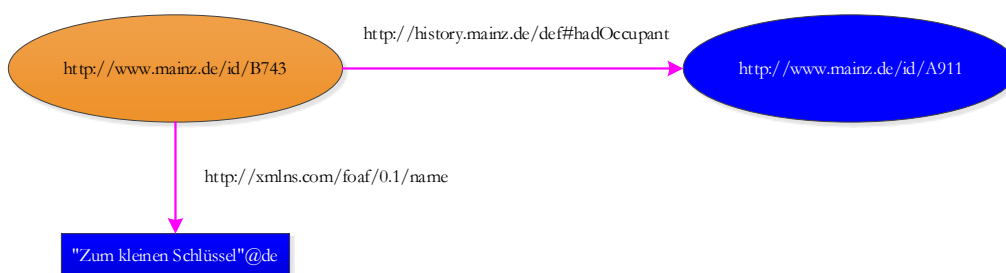


Abb. 2.17: Ein Literal an Objektstelle, ED

Das Literal ist das Objekt im Tripel, das durch ein Rechteck repräsentiert wird, um es von Objekten abzugrenzen, die per URI identifiziert werden (ovale Knotennotation). Literale dürfen nur die Rolle eines Objekts einnehmen, weshalb man keine direkten Aussagen über Literale machen kann (Hitzler et al., 2008). Im Beispiel wurde ein Literal verwendet, aber kein Datentyp angegeben, weshalb es einfach als Zeichenkette interpretiert werden soll (Hitzler et al., 2008). Man bezeichnet diese Literale als ungetypte Literale, die eine Information zur

verwendeten (natürlichen) Sprache erhalten können (hier deutsch = "@de"), wobei diese Angabe Teil des Datenwerts wird (Hitzler et al., 2008). Vereinfachend wird angenommen, dass der Datentyp für ein Literal im RDF durch die Referenz auf XML-Schema-Datentypen per URI (vgl. GP 2.4.2) festgelegt werden kann (Hitzler et al., 2008). Die folgende Abbildung zeigt u. a. den Wertebezeichner des Literals (steht in Anführungszeichen) und die XML-Schema-Datentyp-URI (steht nach der Zeichenfolge "^^"). Allerdings zeigt die Graphik auch weitere, ausgewählte RDFS-Plus-Konstrukte respektive Sprachmittel. Diese wurden ausgewählt, um die wesentlichen Aspekte hervorzuheben, die unter dem vorherigen GP aufgezählt wurden. Hierfür wurden den Farben andere Bedeutungen zugewiesen, als in den vorherigen Graphiken. Dies wird auf der nächsten Seite erklärt.

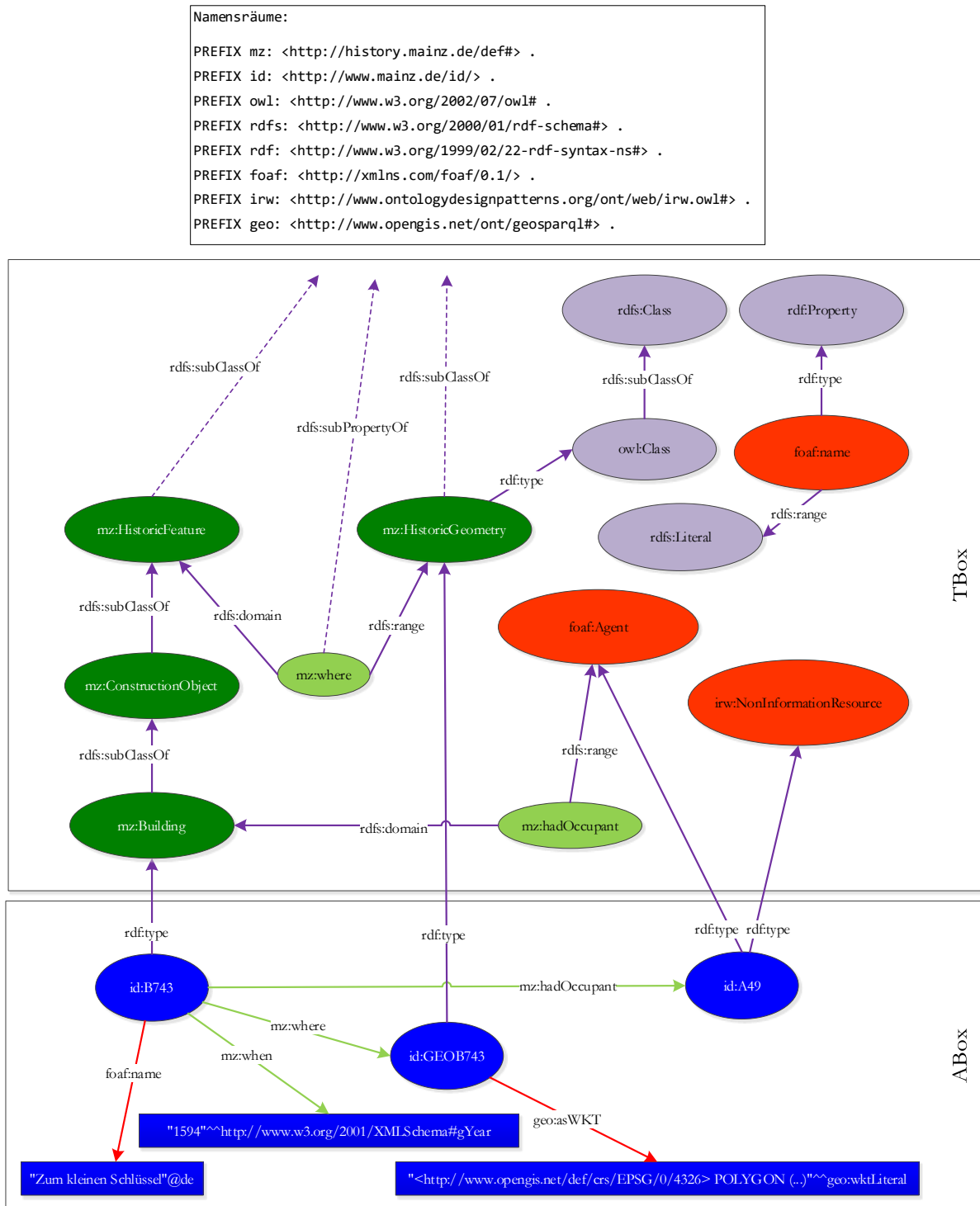


Abb. 2.18: Ausgewählte RDFS-Plus-Sprachmittel, ED

Die ovalen und rechteckigen Knotennotationen haben, ebenso wie die Pfeile, noch dieselben Bedeutungen. Die Wissensbasis ist in ABox und TBox unterteilt. Knoten in Dunkelblau mit weißer Schrift zeigen die Informationen aus dem historischen Häuserbuch auf der Instanzebene. Grüne Knoten zeigen selbst erstellte Klassen (dunkelgrün, weiße Schrift) und Prädikate (hellgrün, schwarze Schrift), die in der TBox definiert und in der ABox verwendet werden. Dort wird die Verwendung dieser Prädikate durch hellgrüne Pfeile dargestellt. Die ABox enthält auch rote Pfeile. Damit wird ausgedrückt, dass in diesem Fall Prädikate aus anderen (externen) Vokabularen nachgenutzt wurden. Die Abbildung zeigt dies exemplarisch anhand der Prädikate **foaf:name** und **geo:asWKT**. Abgesehen von Prädikaten können auch Klassen aus externen Vokabularen in der ABox referenziert werden, um Individuen als Instanzen dieser Klassen auszuweisen. Diese Klassen (als Beispiel: **foaf:Agent**, **irw:NonInformationResource**) sind ebenfalls in Rot dargestellt. Ressourcen können „getypt“ werden (Hitzler et al., 2008, p. 101). Hierfür wird das vordefinierte Prädikat **rdf:type** verwendet (Hitzler et al., 2008). Vordefiniert bedeutet, dass diese Konzepte in den offiziellen W3C-Empfehlungen definiert wurden, d.h. im konkreten Fall in der RDF-, RDFS-, oder OWL-Spezifikation. Alle ausgewählten, vordefinierten Konstrukte sind in Violett dargestellt. Wenn vordefinierte Prädikate verwendet werden, wird dies sowohl in der TBox als auch der ABox durch einen violetten Pfeil angezeigt. Vordefinierte Klassen sind violett eingefärbt. In der TBox werden auch die eigenen Klassen über das vordefinierte Prädikat **rdf:type** getypt, d.h. es wird spezifiziert, dass z.B. die Klasse, die über die URI **mz:HistoricGeometry** bezeichnet ist, eine OWL-Klasse ist, die gemäß OWL-Spezifikation selbst zur Klasse aller Klassen (**rdfs:class**) gehört (Motik et al., 2012). Genauer gesagt ist diese Klasse eine Unterklasse von **rdfs:class**, was über das vordefinierte Prädikat **rdfs:subClassOf** ausgedrückt wird. Damit lassen sich Klassenhierarchien modellieren, wie in der TBox anhand der Klasse **mz:Building** skizziert, die Unterklasse von **mz:ConstructionObject** ist usw. Dies ist auch für Prädikate möglich. RDFS definiert hierfür das Prädikat **rdfs:subPropertyOf**. Das sind Beispiele für das angesprochene Ontologie-Alignment, was über die violetten, gestrichelten Pfeile angedeutet wird. In beiden Fällen handelt es sich um Spezialisierungen von Konzepten, die – einfach ausgedrückt – im GeoSPARQL-Vokabular definiert wurden. Zum einen ist mit **mz:HistoricFeature** die Spezialisierung einer Klasse, zum anderen mit **mz:where** die Spezialisierung eines Properties dargestellt. Diese alternative Bezeichnung für Prädikate kann ebenfalls anhand der obigen Abbildung erläutert werden. Hier ist exemplarisch das Prädikat **foaf:name** dargestellt. Dieses gehört zur vordefinierten Klasse mit dem Bezeichner: **rdf:Property**, die alle Ressourcen umfasst, die Beziehungen ausdrücken (Hitzler et al., 2008). Hinsichtlich der Properties sind noch die vordefinierten Prädikate **rdfs:range** und **rdfs:domain** relevant. Diese dienen der Spezifikation von Verwendungseinschränkungen für Prädikate. Mittels **rdfs:domain** kann der Definitionsbereich eingeschränkt werden, mittels **rdfs:range** der Wertebereich (Hitzler et al., 2008). Die Graphik zeigt eine solche Spezifikation anhand des Prädikats **mz:hadOccupant**. Wenn dieses Prädikat in einem Tripel verwendet wird, kann gefolgert werden, dass das Subjekt vom Typ **mz:Building** ist und das Objekt vom Typ **foaf:Agent**. In diesem Fall werden beide Ressourcen über URIs identifiziert. Allerdings ist es z.B. auch möglich den Wertebereich mit einem Literal anzugeben, was exemplarisch anhand des Prädikats **foaf:name** dargestellt ist.

Die obige Graphik dient nur der Erläuterung. Für die maschinelle Verarbeitung ist sie ungeeignet. Graphen werden hierfür serialisiert, d.h. sie werden zerlegt und die komplexen Datenobjekte werden in lineare Zeichenketten transformiert (Hitzler et al., 2008). Zu diesem Zweck wurden zahlreiche Vorschläge und Standards erarbeitet, z.B. Notation 3 (N3), N-Tripels, Turtle, RDF/XML. Letztere RDF-Serialisierung hat den Status einer W3C-Empfehlung (Gandon and Schreiber, 2014). RDF/XML wird in dieser Arbeit hauptsächlich verwendet. Bezüglich der Syntax wird auf die o.g. W3C-Empfehlung verwiesen. Im Übrigen wird auf die Ausführungen zu Graphen und Bäumen verwiesen (s. GP 2.3.3).

2.2.4 Erweiterte Terminologie

Das URI-Konzept wurde zusammengeführt; die Spaltung in die Standards URN und URL wurde revidiert. Abb. 2.2 zeigt, dass die Standards URL und URN im Endeffekt der Differenzierung zwischen **Zugang** und **Referenz / Adressen** und **Namen** dienen, die Berners-Lee (1991) wie folgt beschrieb: „a **"name"** has tended to mean a logical way of **referring** to an object in some **abstract** name space, while the term **"address"** has been used for something which specifies the **physical location**“. Halpin (2013, p. 43) definierte diese Termini wie folgt: „a **location** is a term that can be used to **access** the thing, while a **name** is a term that can be used to **refer** to a thing. Unlike access, **reference** is the use of an **identifier** for a thing to which one is immediately causally disconnected. Generally, the relationship of a thing to another thing to which one is immediately causally disconnected is a relationship of reference to a **referent or referents**, the distal thing or things referred to by a representation. The thing which refers to the referent(s) we call the **'representation,'** and take this to be equivalent to being a symbol“. Halpin (2013, p. 43) bezog sich auf Hayes and Halpin (2008): „**Access** is the use of an identifier to create immediately a causal connection to the thing identified“.

Die o.g. Differenzierung wurde durch RFC 2396 aufgehoben „in order to define a single, generic syntax for all URI“ (Berners-Lee et al., 1998, p. 2). RFC 2396 wurde durch RFC 3986 aktualisiert (Berners-Lee et al., 2005), welcher bereits für die Erörterung der URI-Syntax herangezogen wurde. Da das ursprüngliche Konzept letztlich doch umgesetzt wurde, können URIs verwendet werden, um digitale Ressourcen zu identifizieren, deren unterschiedliche Web-Repräsentationen über das Internet transferiert werden können (Berners-Lee et al., 2005). URIs können jedoch ebenso verwendet werden, um auf alle möglichen Bezugsobjekte/Referenten (realweltliche, fiktive oder konzeptuelle Ressourcen) zu referieren (Heath and Bizer, 2011; Halpin, 2013). Bei repräsentativer Verwendung von URIs qualifizieren sie sich als „‘stands-in’ for referents“ (Halpin, 2013, p. 32). Die Referenten können aber keine Web-Repräsentationen haben (Allemang et al., 2020). Selbst wenn das historische Gebäude aus dem Beispiel noch existieren würde, kann kein Internetprotokoll das Gebäude in seine Einzelbestandteile zerlegen, verpacken und über ein Netzwerk versenden. Die digitale Ressource enthält zwar Informationen über die realweltliche Ressource, ist jedoch keinesfalls mit dieser gleichzusetzen (Lewis, 2007).

Zuvor wurde erklärt, wie mithilfe von RDF-Tripeln Aussagen über analoge Ressourcen getroffen werden können. Der Begriff Ressource wurde nicht in den Fokus gerückt. Dies ist nun zwingend erforderlich. Eine serialisierte RDF-Beschreibung (z.B. in RDF/XML) ist ein **Semantic Web Dokument** (Halpin, 2013), also eine konkrete Web-Repräsentation einer digitalen Ressource. In Semantic Web Dokumenten werden regelmäßig URIs verwendet, um nicht digitale Ressourcen zu bezeichnen (s.o.). Man nennt diese Bezeichner in diesem Fall **Semantic Web URIs** (Halpin, 2013). Diese URIs können keinen Zugang zu Web-Repräsentationen der bezeichneten (nicht digitalen) Ressourcen bieten. Nun stellt sich aber die Frage, wie diese HTTP-URIs dereferenziert werden können. Dies wird im LOD-Kontext explizit gefordert (Heath and Bizer, 2011). Außerdem ist fraglich, wie ein Agent ermitteln kann, ob eine (prinzipiell undurchsichtige) URI repräsentativ für eine analoge Ressource verwendet wurde. Diese Fragen („httpRange-14“) wurden von der TAG behandelt (Lewis, 2007).

Jacobs and Walsh (2004) hoben einen wesentlichen Aspekt von **Informationsressourcen** hervor: „all of their essential characteristics can be conveyed in a message“. Halpin (2013, p. 64) schlussfolgerte daraus: „In particular, this means that an information resource is a resource that can be realized as an information-bearing message, even with multiple encodings.“ Außerdem muss seine Definition der **Nicht-Informationsressource** herangezogen werden, da dieser Begriff in normativen Dokumenten nicht definiert wurde. Nicht-Informationsressourcen sind Ressourcen „that cannot be realized digitally by a message, but only described or depicted by digital information“ (Halpin, 2013, p. 64). „A non-information resource is a thing that is **analogue** and so resists direct digital encoding, but can only be **indirectly** encoded via representations of the thing in a suitable

language“ (Halpin, 2013, p. 65). Ob sich eine Ressource als Informationsressource oder als Nicht-Informationsressource qualifiziert, ist aber nicht davon abhängig, ob diese tatsächlich über Web-Repräsentationen verfügt. Die Frage ist, ob es theoretisch möglich ist, den Sinngehalt der Ressource in einer bestimmten Web-Repräsentation zu kodieren und diese als Nachricht über ein Netzwerk zu senden (Halpin, 2013). Deshalb ist z.B. der Inhalt/Text eines analogen Buches eine Informationsressource auch wenn dieser (gerade) nicht über das Web abgerufen werden kann, denn theoretisch wäre dies möglich (Halpin, 2013). Allerdings widersprach Halpin (2013) damit Jacobs and Walsh (2004). Halpin (2013) führte aber eine Unterklasse von Informationsressourcen ein: „a **Web resource**, an information resource that has **accessible** Web representations that realize its information. A Web resource can then be thought of as a mapping from time of request to a series of Web representation responses, where the information realized by those Web representations are the Web resource“ (Halpin, 2013, p. 66). Diese Definition ist fast identisch mit der Definition im RFC 2616; allerdings definierten Fielding et al. (1999) die „Superklasse“ (Ressource). Im Semantic Web muss weiter differenziert werden: „A **Semantic Web resource** is a resource that allows **access** to Semantic Web documents“ (Halpin, 2013, p. 66). Außerdem benötigen Nicht-Informationsressourcen im Semantic Web sog. „**associated descriptions**, resources that have as their primary purpose the representation, however incomplete, of some non-information resource“ (Halpin, 2013, p. 66). Zugehörige Beschreibungen sind also ein klassischer Fall von Metadaten (Halpin, 2013). Wenn sich die Web-Repräsentationen von digitalen Informationsressourcen auf analoge Nicht-Informationsressourcen beziehen, diese aber nicht gleichgesetzt werden dürfen (Fielding et al., 1999), dann muss im Zuge der Dereferenzierung ausgedrückt werden können, um welche Art von Ressourcen es sich handelt (Lewis, 2007). Die TAG schlug dafür die Verwendung der HTTP-Statuscodes vor („httpRange-14“):

Code / Text	ursprüngliche Bedeutung	Bedeutung im Semantic Web
„200 OK“ (Fielding et al., 1999, p. 57).	„The request has succeeded“ (Fielding et al., 1999, p. 57).	„The resource is an information resource and a representation of it has been returned“ (Lewis, 2007).
„303 SEE OTHER“ (Fielding et al., 1999, p. 62).	„The response to the request can be found under a <i>different</i> URI and SHOULD be retrieved using a GET method on that resource. This method exists primarily to allow the output of a POST-activated script to redirect the user agent to a selected resource. The new URI is <i>not a substitute reference</i> for the originally requested resource. The different URI SHOULD be given by the Location field in the response“. (Fielding et al., 1999, p. 62).	„The resource could be any resource. There is an associated resource whose URI has been returned. The associated resource might or might not be an information resource“ (Lewis, 2007).
„404 NOT FOUND“ (Fielding et al., 1999, p. 65).	„The server has not found anything matching the Request-URI. No indication is given of whether the condition is temporary or permanent“ (Fielding et al., 1999, p. 65).	„Nothing can be inferred about the nature of the resource“ (Lewis, 2007).

Tab. 2.1: Die Semantik ausgewählter HTTP-Statuscodes im Semantic Web

Wichtig ist, dass die URI, die der Server in seiner Antwort im „Location-Feld“ übermittelt, nicht dieselbe URI sein darf, die ursprünglich angefragt wurde, falls diese eine Nicht-Informationsressource identifiziert. Anderenfalls wäre dies ein Fall der URI-Kollision. Insofern kann konstatiert werden, dass die zugehörige Beschreibung eine eigene URI erhalten muss. In dieser Arbeit wird diese URI als **303-URI** bezeichnet. Eine URI, die eine Nicht-Informationsressource identifiziert, wird in der vorliegenden Arbeit als **NIR-URI** bezeichnet. Die IRW-Ontologie formalisiert diese Konzepte:

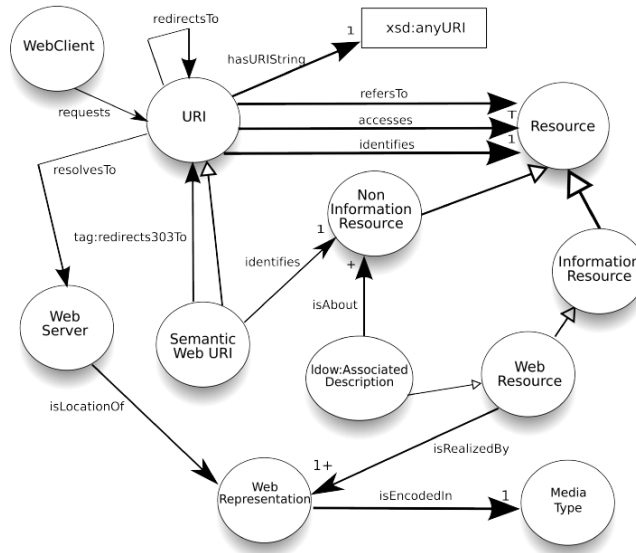


Abb. 2.19: Auszugsweise, graphische Darstellung der IRW, OA von Halpin and Presutti (2009, p. 4)

Die folgende Abbildung visualisiert die beschriebene Netzwerkinteraktion anhand des Schlüsselbeispiels:

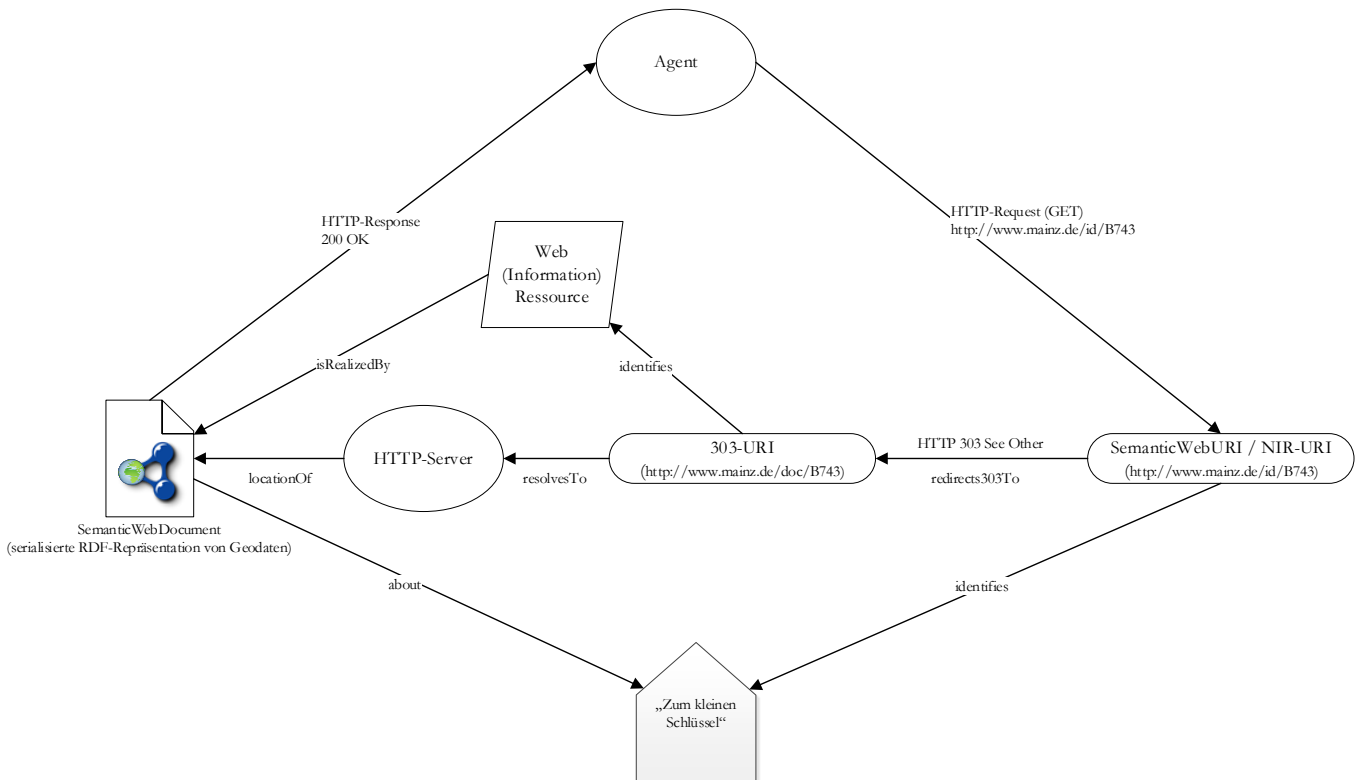


Abb. 2.20: 303-Weiterleitung, OA von Halpin and Presutti (2009, p. 3), angepasste Darstellung

Anhand der obigen Abbildung wird deutlich, dass eine NIR-URI nur indirekt dereferenziert werden kann. Dies steht nicht im Widerspruch zur sorgsam formulierten Phrase „**Dereferenzierung einer HTTP -URI**“ (s. GP 2.1.2). In diesem Prozess spielt aber auch noch die „Inhaltsvereinbarung“ (kurz: „IVB“) / „Content Negotiation“ eine Rolle. Es geht hier nicht darum, den Inhalt zwischen Client und Server auszuhandeln, denn dieser soll (!) immer ident sein. Es geht hier darum, in welcher Sprache / welchen Sprachen der Inhalt am besten kodiert wird. HTML-Darstellungen/Webseiten eignen sich vor allem für Menschen. RDF-Dokumente sind für die maschinelle Verarbeitung geeignet (Allemang et al., 2020). Deshalb wird empfohlen (zumindest) diese beiden Web-Repräsentationen anzubieten (Heath and Bizer, 2011) Die „IVB“ ist nachfolgend vereinfacht dargestellt, wobei wieder davon ausgegangen wird, dass ursprünglich eine NIR-URI dereferenziert werden sollte:

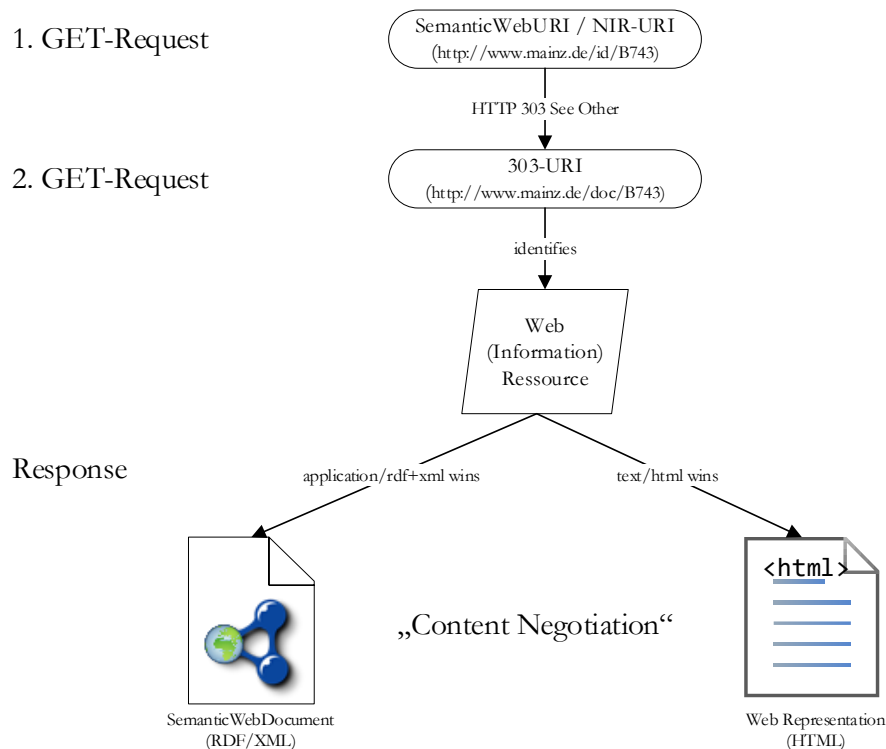


Abb. 2.21: „IVB“ nach HTTP 303, OA von Halpin and Presutti (2009, p. 33), angepasste Darstellung

Die obige Darstellung zeigt einen einfachen Fall. Mehrere Weiterleitungen (z.B. auch über HTTP 302) sind je nach Anwendungsszenario möglich bzw. sinnvoll. 303-URIs sind *eine* spezifische Realisierung der „Cool URIs for the Semantic Web“ (Sauer mann and Cyganiak, 2008). Für diese Arbeit sind noch die sog. **Hash-URIs** relevant. Hash-URIs eignen sich vor allem dann, wenn eine Gruppe zugehöriger Beschreibungen identifiziert werden soll, die sich auf verwandte Nicht-Informationsressourcen beziehen (Sauer mann and Cyganiak, 2008). Ein klassischer Fall im Semantic Web sind z.B. Ontologien (Heath and Bizer, 2011). Ein Semantic Web Dokument soll per HTTP-URL zugänglich sein (<http://history.mainz.de/def>). Es soll eine statische Textdatei lokalisiert werden, die Beschreibungen der konzeptuellen Nicht-Informationsressourcen enthält. Es ist sinnvoller dem Client dieses Semantic Web Dokument en bloc zu liefern, statt jedes Konzept separat über NIR-URIs zu identifizieren und auf 303-URIs weiterzuleiten (Heath and Bizer, 2011). Wenn der Client z.B. die NIR-URI <http://history.mainz.de/def#hadOccupant> dereferenziert, um zu erfahren, welche Bedeutung das Prädikat `hadOccupant` hat, enthält die Antwort des HTTP-Servers das komplette Semantic Web Dokument. Wie erwähnt, verwendet der Server den Teil der URI nach dem Doppelkreuz („Hash“) für die Beantwortung der Anfrage nicht (Berners-Lee et al., 2005). Dieser Teil wird vorher „abgeschnitten“ (Berners-Lee et al., 2005).

Pragmatisch betrachtet gibt es die URI <http://history.mainz.de/def#hadOccupant> aus der Sicht des Servers also nicht, sondern nur die URI <http://history.mainz.de/def>. Da die erstgenannte URI die Nicht-Informationsressource identifiziert und die zweite URI die Informationsressource, ist dies kein Fall der URI-Kollision, denn offensichtlich handelt es sich um zwei verschiedene URIs. Der Client kann den „fragment identifizier“ für die Interpretation verwenden (Berners-Lee et al., 2005). Vereinfacht ausgedrückt, verweist das Doppelkreuz (“#”) den Client auf die relevante Stelle im gelieferten Dokument. Somit kann der Client im Beispiel direkt an die Stelle des Semantic Web Dokuments „springen“, wo die Bedeutung von **hadOccupant** beschrieben wird. Wenn er sich anschließend z.B. für die Bedeutung des Prädikats **hadOwner** interessiert, braucht er keine neue Anfrage zu stellen. Hash- und 303-URIs können auch kombiniert werden (Lewis, 2007).

Die obigen Ausführungen fokussieren lediglich pragmatische Lösungsmöglichkeiten, die es Agenten im Semantic Web gestatten, die „Bedeutung“ von URIs ermitteln und sinnvolle Aussagen über alle möglichen Ressourcen in RDF-Tripelform treffen zu können. Die Identitätskrise im Semantic Web resultierte – rein pragmatisch und nicht „philosophisch“ gesehen – aus einer unklaren Verwendung von URIs. Es besteht die Gefahr, dass Aussagen ihre Bedeutung verlieren, verfälscht oder wertlos werden, wenn die URI, die eine Nicht-Informationsressource identifiziert, mit der URI verwechselt wird, welche die zugehörige Beschreibung identifiziert.

2.2.5 RDF – Web-Architekturprinzipien

Das RDF ist für die Umsetzung des AAA-Slogans (Allemang et al., 2020) geeignet – unabhängig davon, ob Aussagen über Informationsressourcen oder Nicht-Informationsressourcen gemacht werden sollen. Das **Prinzip der Universalität** wird im LOD-Kontext berücksichtigt, da auf leere Knoten verzichtet werden muss (Heath and Bizer, 2011). Die einzigen Ressourcen, die nicht per URI identifiziert werden, sind Literale. Die LOD-Cloud entstand wegen des **Prinzips der Verlinkung** (Bizer et al., 2009), was im Rahmen der Einführung der RDF-Links anschaulich dargestellt wurde. Auch die Web-Repräsentationen aus den Beispielen zum Web of Documents können ins Web of Data integriert werden. Dafür werden jedoch keine ungetypten Hyperlinks verwendet, sondern Konzepte, deren Bedeutungen in RDF-Vokabularen beschrieben wurden. So könnte man z.B. die Nicht-Informationsressource (<http://www.mainz.de/id/743>) über ein Prädikat aus dem etablierten FOAF-Vokabular (**foaf:depiction**) mit den Abbildungen aus dem Beispiel verknüpfen (z.B. <http://www.mainz.de/Bilder/ZumKleinenSchluessel.jpg>). Dies gilt auch für die zugehörige Beschreibung (<http://www.mainz.de/doc/B743>), die ebenfalls über adäquate Prädikate (z.B. **irw:isAbout**) verlinkt werden kann. Die Verknüpfung ist nicht nur in Semantic Web Dokumenten möglich, sondern auch in der Webseite per se, indem RDFa verwendet wird (Heath and Bizer, 2011). Insofern kann die Forderung von Berners-Lee (1994b) – „allowing links to be created with relationship values“ – im Web of Data in mannigfaltiger Weise umgesetzt werden.

Das LOD-Paradigma fußt auch auf dem **Prinzip der Selbstbeschreibung**, da die URIs in einer RDF-Beschreibung direkt oder indirekt dereferenzierbar sein müssen (Heath and Bizer, 2011). Tripel können zwischen verschiedenen Agenten transferiert werden, ohne dass ihre ursprüngliche Bedeutung dabei verloren geht (Hitzler et al., 2008), denn die notwendigen Informationen für die Interpretation können per HTTP abgerufen werden („follow your nose“). Diese Informationen sind zwar maschinenverarbeitbar, in einigen Fällen sind diese aber weiterhin für die Interpretation durch einen Menschen bestimmt. So kann man z.B. mit dem Prädikat **rdfs:comment** menschenlesbare, natürlichsprachliche Kommentare zu einer Ressource machen (Hitzler et al., 2008). Grundsätzlich hat der Mensch weiterhin großen Einfluss darauf, in welchem Grad die Ressourcen selbstbeschreibend sind (Heath and Bizer, 2011). Die Eigentümerin einer Ressource kann entscheiden, ob sie einer RDF-Beschreibung Tripel hinzufügt, die Metainformationen zur Ressource enthalten. Abgesehen von den o.g. Prädikaten aus Vokabularen, die Termini für eine spezielle Domäne definieren, sind auch die universellen Ausdrucksmittel von RDF/RDFS selbst in Semantic Web Dokumenten beschrieben. Wenn ein Agent z.B. erfahren

möchte, welche Bedeutung der Terminus „Prädikat“ im RDF hat, kann er die Hash-URI <http://www.w3.org/1999/02/22-rdf-syntax-ns#predicate> dereferenzieren, um eine RDF-Beschreibung zu erhalten. Wie bereits klargelegt, soll die allgemein anerkannte wohldefinierte Bedeutung der Termini, welche in universellen Vokabularen (RDF, RDFS, OWL etc.) festgelegt ist, die von Natur aus öffentlich sind, **nicht** von einem **speziellen** Kontext oder von der **individuellen** Interpretation einzelner Agenten abhängen.

Die formale Semantik von RDFS und OWL berücksichtigt die **Offene-Welt-Annahme** (Berners-Lee, 1998; Hayes, 2002). Der Grund dafür ist, dass das RDF im Web of Data primär der Informationsintegration dient (Halpin, 2013); weil Graphen aus unterschiedlichen Quellen vereinigt werden (Hitzler et al., 2008). Dabei soll es jedoch niemandem möglich sein, die ursprüngliche Bedeutung eines anderen Graphen durch das Hinzufügen von weiteren Tripeln zu verändern, also z.B. Aussagen zu negieren (Halpin, 2013). Anderenfalls müssten Agenten beim Schlussfolgern von implizitem Wissen die Graphen nach jeder Konsolidierung erneut interpretieren (Halpin, 2013). Die TBox kann nicht für die Validierung herangezogen werden. Wenn dies erforderlich ist, können mithilfe der Shapes Constraint Language (SHACL) Kriterien zwecks Validierung formuliert werden (Knublauch and Kontokostas, 2017). Im Gegensatz zu RDFS/OWL setzt SHACL auf die Geschlossene-Welt-Annahme (Allemang et al., 2020).

Abb. 2.1 verdeutlicht, dass dem **Prinzip der geringsten Ausdrucksmächtigkeit** im Rahmen der Entwicklung der einschlägigen W3C-Empfehlungen (z.B. RDFS, OWL) Rechnung getragen wurde. Die beiden vorgenannten Ontologiesprachen basieren auf dem RDF (Hitzler et al., 2008), das als einfache Wissensrepräsentationsprache (Halpin, 2013) die Basis für das Semantic Web bildet (Allemang et al., 2020). Die Einstiegshürde wurde somit bewusst niedrig gehalten, um möglichst vielen Agenten die Partizipation im Semantic Web zu ermöglichen, da neue Informationen sehr einfach ins Web of Data integriert werden können und die Nachnutzung der bereits existierenden Informationen in eigenen Anwendungen nicht zwangsläufig die Implementierung hochkomplexer Konzepte bedingt (Heath and Bizer, 2011). RDFS und OWL qualifizieren sich als Erweiterungen des RDF (Halpin, 2013). In Abhängigkeit zu den jeweiligen Anforderungen bzw. hinsichtlich der zu modellierenden Informationen kann für jedes Anwendungsszenario die angemessene Sprache ausgewählt werden, die für eine geeignete Informations-Spezifikation ausreichend ist. Sollten die Modellierungsprimitive von RDFS für die Spezifikation der terminologischen Zusammenhänge bzw. für die beabsichtigten Inferenzen nicht ausreichend sein, können komplexere Konzepte auch über die ausdrucksstärkere Ontologiesprache OWL formalisiert werden. Allerdings wurde das Prinzip der geringsten Ausdrucksmächtigkeit auch bei der Standardisierung von OWL besonders berücksichtigt, da OWL in verschiedene Teilsprachen unterteilt wurde (Motik et al., 2012). Deshalb ist es auch beim Einsatz von OWL möglich, ein adäquates Verhältnis zwischen Ausdrucksstärke und Skalierbarkeit herzustellen (Hitzler et al., 2008). RDFS-Plus sieht lediglich die Verwendung einfacher OWL-Konzepte vor, mit denen sich z.B. keine Verwendungseinschränkungen für Prädikate spezifizieren lassen, die auf der Konstruktion anonymer Klassen basieren (Allemang et al., 2020).

2.3 Geo Web

2.3.1 Standards, Geodateninfrastrukturen

Das Geo Web kann auf eine Vielzahl gemeinschaftlicher Standardisierungsbestrebungen zurückblicken, die neben universellen Austausch- und Speicherformaten auch Geodienste mit offenen Web-Schnittstellen für die unterschiedlichsten Anwendungsszenarien hervorbrachten. Mittlerweile wurden viele dieser Schnittstellen sowohl in freien als auch proprietären GIS implementiert, welche u. a. die Verarbeitung von Geoinformationen ermöglichen, die mithilfe von Standardformaten strukturiert bzw. kodiert wurden. Davon profitieren die Akteure im Geo Web, da unterschiedliche GIS von verschiedenen Herstellern zusammenarbeiten können. Man spricht in diesem Zusammenhang von Interoperabilität und meint ganz einfach, „dass Softwarekomponenten über Software-, System- und Institutionsgrenzen hinweg miteinander so kommunizieren und zusammenarbeiten können, dass der Nutzer nichts davon merkt“ (Andrae, 2009, p. 23). Dies ist eine wesentliche Voraussetzung, um diverse (Web-)GIS in dezentrale GDI¹⁹ integrieren zu können. Jene sind heute (noch) omnipräsent, denn sie ermöglichen den netzwerkbasieren und grenzübergreifenden Austausch digitaler, räumlicher Informationen zwischen den unterschiedlichsten Akteuren (Staub, 2009). Die europäische GDI „namens INSPIRE (INfrastructure for Spatial Information in Europe)“ (Bill, 2016, p. 234) belegt dies auf supranationaler Ebene²⁰.

Im Rahmen solcher Vorhaben werden, neben den **Standards** bzw. Empfehlungen der bereits vorgestellten Organisationen respektive Standardisierungsgremien, auch die De-jure- / De-facto-Standards der ISO und des OGC verwendet. Die syntaktische Interoperabilität kann über OGC Geodienste²¹ sichergestellt werden (Klien, 2007). „Syntaktische Interoperabilität mittels OGC Webservices erlaubt die interoperable Nutzung verfügbarer Daten über eine Standardschnittstelle“ (Staub, 2009, p. 25). Um dies zu erreichen, arbeiten das OGC, das W3C und die ISO zusammen (Bill, 2016). Innerhalb der ISO konstituierte sich das ISO Technical Committee (TC) 211, das die Normenreihe 19100 speziell für Geoinformationen in enger Abstimmung mit dem OGC erarbeitete (Andrae, 2009). ISO-Standards und OGC-Spezifikationen²², können differieren oder identisch sein (Andrae, 2009). Eine genaue Beschreibung der Beziehungen zwischen sämtlichen OGC-Spezifikationen und ISO-Standards bzw. eine Übersicht zu sämtlichen Dokumenten aller Standardisierungsgremien ist in dieser Arbeit wenig zielführend. Deshalb werden in den folgenden Abschnitten nur wenige konzeptuelle Grundlagen aus den einschlägigen Themenbänden (Topics) der abstrakten, plattformunabhängigen Spezifikationen des OGC (OpenGIS Abstract Specifications) behandelt. Selbiges gilt für die konkretisierenden Implementierungsspezifikationen des OGC (OpenGIS Implementation Specifications), die auf eine Auswahl rund um den Basic WFS als Downloaddienst und GML als „Datenaustauschformat“ für vektorielle Geodaten eingegrenzt werden.

GDI können aus verschiedenen Blickwinkeln betrachtet, charakterisiert und definiert werden (Bernard et al., 2005), indem z.B. organisatorische Aspekte oder die technische Infrastruktur fokussiert wird (Würriehausen, 2015). Für letzteren Fall definierte Staub (2009, p. 66) die GDI als „Summe von Geodaten, Netzen, Services und Standards“. Services manifestieren sich in GDI als standardisierte Geoinformationsdienste (GI-Dienste), die häufig mittels OWS realisiert wurden (Staub, 2009). Williamson et al. (2003) nannten zahlreiche Vorteile von GDI, wie z.B. die Vermeidung von Redundanz und verbesserte Zugriffsmöglichkeiten. Diesbezüglich hob Bill (2016) hervor, dass die Benutzerin einer GDI deren Leistungsfähigkeit regelmäßig auch anhand nicht-funktionaler Anforderungen beurteilt, wobei er insbesondere effiziente Zugriffsmöglichkeiten auf relevante Informationen betonte. Seifert (2008) betrachtete GDI auch aus der Sicht einer Anwenderin und nannte

¹⁹ GDI steht in dieser Arbeit sowohl für die Einzahl (Geodateninfrastruktur) als auch für die Mehrzahl (Geodateninfrastrukturen).

²⁰ INSPIRE-Geoportal: <https://inspire-geoportal.ec.europa.eu/>

²¹ Alternative Bezeichnung: OGC Webservices (OWS)

²² Alternative Bezeichnung: OpenGIS Spezifikationen

viele verschiedene Funktionen und Webservices, wie z.B. Register/Registry-Services, die realisiert werden sollten. Staub (2009) griff die anwenderorientierte Sicht auf und kam zu dem Schluss, dass erst mit der Realisierung aller von Seifert (2008) genannten Komponenten, von einer GDI gesprochen werden kann, die auf einer serviceorientierten Architektur (SOA) basiert. Bengel (2014, p. 252) identifizierte die Services als Grundelemente eines SOA-Systems, die „nicht auf einem Server konzentriert, sondern auf die unterschiedlichen Rechner im Netz möglichst gleichmäßig verteilt“ sind. Die **technische Infrastruktur** kann in dieser Arbeit auf die wesentlichen Komponenten reduziert werden: „Spatial Data Infrastructures usually consist of at least three fundamental components: „One or many geospatial-aware **databases** [...], a **webservice infrastructure** exposing machine-readable views [...] of geospatial data and a **web interface** to make these geospatial views accessible to end users“ (Homburg, 2020, p. 3). **GDI prägen das Geo Web**, da sie auf verschiedenen Ebenen (z.B. kommunale/nationale GDI) bzw. in unterschiedlichen Organisationseinheiten (z.B. privat/öffentlich) für mannigfaltige Zwecke realisiert wurden (Würriehausen, 2015). Unabhängig davon, ob sich eine spezifische Ausprägung einer GDI oder ein Zusammenschluss mehrerer GDI als vollwertige SOA-Implementierung qualifiziert, kann deshalb konstatiert werden, dass sowohl das Geo Web als auch SOA auf dem **Grundgedanken von verteilten Diensten** basieren. Die ressourcen-zentrische Sicht des Semantic Web bzw. Ressourcen-orientierte Architekturen (ROA), das REST-Paradigma und die Web-Architekturprinzipien erhalten im Geo Web mehr und mehr Einzug. Die Thematik wird unter GP 2.3.4 aufgegriffen, um diesen Paradigmenwechsel anzureisen.

2.3.2 Downloaddienst: Basic WFS

Der WFS bietet Zugriffsmöglichkeiten auf vektorielle Geoobjekte (Vretanos, 2002). Der WFS ist ein wichtiger Dienst in einer traditionellen GDI (= „Legacy-System“). Wenn lediglich ein lesender Zugriff möglich ist, bezeichnet man den WFS als „**Basic WFS**“ (Vretanos, 2002, p. 3). Ein „Transaction WFS“ (WFS-T) bietet darüber hinaus noch Operationen für die Transaktion von vektorialen Geoobjekten an, wobei optional eine Konkurrenzkontrolle implementiert werden kann (Vretanos, 2002, p. 3). In beiden Fällen basiert der WFS auf dem bereits definierten Client-Server-Modell, was die folgende Abbildung illustriert. Sie ist in der konkreten **WFS-Implementierungsspezifikation** des OGC (Version 1.0.0) enthalten, die Vretanos (2002) verfasste.

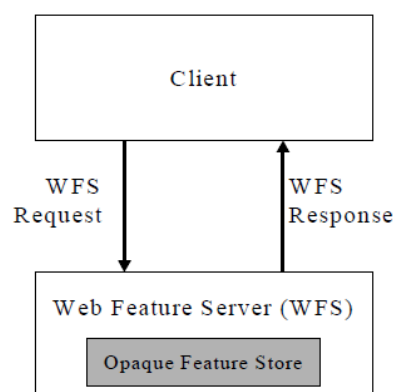


Abb. 2.22: Kommunikation mit einem WFS, OA von Vretanos (2002, p. 9)

Für diese Arbeit ist nur der Basic WFS in der Version 1.0.0 relevant. Ein Basic WFS bietet drei Operationen an (Vretanos, 2002). Die folgende Abbildung von Staub (2009, p. 57) zeigt die aufeinanderfolgenden Operationen in Form eines Sequenzdiagramms.

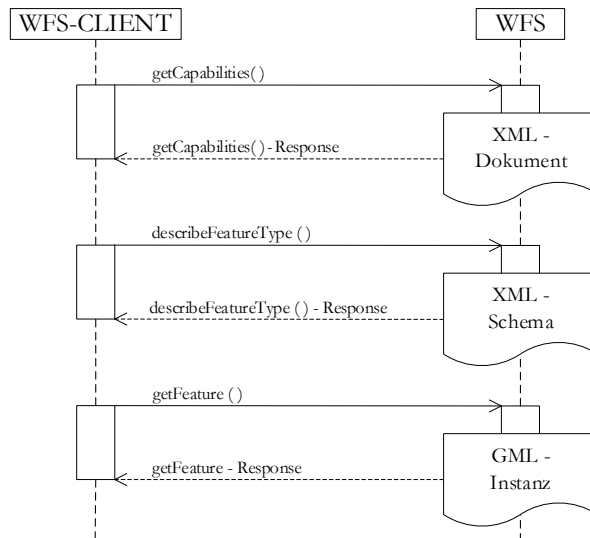


Abb. 2.23: Operationsfolgen beim Basic WFS, OA von Staub (2009, p. 57), angepasste Darstellung

Normalerweise stellt ein Client eine GetCapabilities-Anfrage vor der DescribeFeatureType-Anfrage, um schließlich eine geeignete Instanz-Anfrage über die GetFeature-Operation formulieren zu können (Vretanos, 2002). Die Details zur Parametrisierung können der o.g. WFS-Implementierungsspezifikation entnommen werden. Ein WFS muss eine GetCapabilities-Anfrage mit geeigneten Metadaten in Form eines XML-Dokuments beantworten (Vretanos, 2002). Die Implementierung der **GetCapabilities-Operation** ist eine obligatorische Anforderung (Vretanos, 2002). Die Anforderungen an das auszuliefernde XML-Dokument definierte Vretanos (2002) über ein XML Schema (Abkürzung XSD, steht für XML Schema Definition). Dieses spezifiziert **vier Sektionen**, in denen z.B. die Namen der Elemente, deren Reihenfolge etc. festgelegt werden (Vretanos, 2002).

Die **Service-Sektion** liefert allgemeine Informationen zum WFS, wobei mindestens ein Titel und eine Online-Ressource angegeben werden muss (Vretanos, 2002). Optional darf das XML-Dokument z.B. auch eine kurze Zusammenfassung und Schlüsselwörter enthalten (Vretanos, 2002). Danach folgt eine Sektion, welche die unterstützten Operationen angibt (Vretanos, 2002). Damit wird der Client über die eigentlichen **Fähigkeiten** des WFS informiert (Vretanos, 2002). Für den Basic WFS werden die drei Operationen definiert, die das Sequenzdiagramm zeigt (Vretanos, 2002). Deren Parametrisierung muss über die HTTP-Verben GET und POST möglich sein²³ (Vretanos, 2002). In letzterem Fall können die Anfrage-Parameter als Key-Value-Paare (KVP) oder als XML-Strukturen übergeben werden (Vretanos, 2002). Im Fall von GET sind die Anfrage-Parameter als KVP Teil der URI. In der Fähigkeiten-Sektion müssen wenigstens zwei Ausgabeformate angegeben werden (Vretanos, 2002). Für die DescribeFeatureType-Operation muss mindestens XSD und für die GetFeature-Operation mindestens GML2 angeboten werden (Vretanos, 2002). Letzteres wird im Sequenzdiagramm als Instanz bezeichnet, um es vom Schema abzugrenzen. Im konkreten Fall kann das Schema alternativ als GML2-Applikationsschema bezeichnet werden (Vretanos, 2002). Die Antwort auf eine GetCapabilities-Anfrage enthält noch eine weitere Sektion, welche die verfügbaren **Featuretypen** listet. Mindestens ein Featuretyp muss gelistet werden. Zu jedem Featuretyp können weitere Angaben enthalten sein, wie z.B. die vorgenannten optionalen Informationen (Vretanos, 2002). Die Angabe eines Namens und eines CRS²⁴ ist jedoch obligatorisch (Vretanos, 2002). Den Namen des Featuretyps benötigt der Client, um eine differenzierte DescribeFeatureType-Anfrage stellen zu können (Vretanos, 2002). Wird in einer DescribeFeatureType-Anfrage kein Name angegeben, liefert der WFS Beschreibungen zu sämtlichen verfügbaren Featuretypen aus (Vretanos, 2002).

²³ SOAP Bindings wurden erst für spätere WFS-Versionen definiert.

²⁴ In dieser Version des WFS wird statt CRS, das Akronym SRS (spatial reference system) verwendet.

Die letzte Sektion spezifiziert die **Filter-Möglichkeiten** (Vretanos, 2002). Diese Sektion wird nicht direkt in der WFS-Implementierungsspezifikation definiert (Vretanos, 2002). Stattdessen wird die korrespondierende Version (1.0.0) der Filter Encoding (FE) Implementierungsspezifikation referenziert, die ebenfalls von Vretanos (2001) verfasst wurde und wiederum eine konkrete XML-Realisierung der Filtersprache OGC Common query language (CQL) darstellt. Dort ist das XSD enthalten, welches die Filter-Sektion definiert (Vretanos, 2001). Die Referenz hierauf erfolgt über das XSD namens WFS-capabilities. Die WFS-Implementierungsspezifikation ist hier widersprüchlich. Einerseits wird die Filter-Sektion in den Erläuterungen als optional bezeichnet, andererseits legt das vorgenannte XSD fest, dass das Element `ogc:Filter_Capabilities` immer enthalten sein muss (Vretanos, 2001). Da gegen das WFS-capabilities.xsd validiert wird, hat es Vorrang. Die Filter-Sektion ist nicht optional. Darüber kann der Client ermitteln, welche räumlichen, vergleichenden, identifizierenden und logischen Operatoren er in einer GetFeature-Anfrage verwenden kann (Vretanos, 2001). Als Antwort auf eine **DescribeFeatureType-Anfrage** erhält der Client ein XSD/GML2-Applikationsschema, das die Struktur für einen oder mehrere Featuretypen definiert, d.h., die verfügbaren Geoobjekte werden näher beschrieben, z.B. über die Angabe der Attribute und ihrer Optionalitäten, Datentypen etc. (Vretanos, 2002). Der Client kann das Schema letztlich zur (syntaktischen) Validierung verwenden (Vretanos, 2002). Featuretypen kann man alternativ auch als Objektklassen bezeichnen (Bill, 2016) und die Basiseinheiten der Geoinformationsverarbeitung – die Geoobjekte – als Features (Andrae, 2009). Die Beschreibungen der Features werden vom WFS in Form von GML2-Instanzen auf eine **GetFeature-Anfrage** geliefert. Die GetFeature-Operation dient also dem Download der Feature-Beschreibungen, weshalb der WFS sehr häufig als Downloaddienst bezeichnet wird (Patroumpas et al., 2015). Im konkreten Fall erhält der Client als Antwort auf eine Download-Anfrage immer eine Menge von Feature-Beschreibungen, die als Feature Collection bezeichnet wird. Der Zugriff auf eine einzelne Feature-Beschreibung (über die Operation GetGmlObject) ist in dieser Version des WFS nicht möglich, da sie in einer jüngeren WFS-Implementierungsspezifikation definiert wurde (Vretanos, 2002; Vretanos, 2005).

2.3.3 Grundlagen der Datenmodellierung

Der ISO-Standard 19109 enthält grundlegende Konzepte zur Modellierung räumlicher Informationen und er definiert das Meta-Modell namens General Feature Model (ISO/TC 211, 2005). Es werden Regeln festgelegt, wie Featuretypen (engl. feature types) in einem Anwendungsschema (engl. application schema) beschrieben werden können (ISO/TC 211, 2005). Das OGC erarbeitete mit der abstrakten OpenGIS-Spezifikation, Themenband 5 (Features), eine ähnliche konzeptionelle Grundlage, die mit ISO 19109 allerdings nicht deckungsgleich ist (Kottman and Reed, 2009). Beide Standards beziehen sich jedoch auf ISO 19107 – Spatial Schema (Kottman and Reed, 2009; ISO/TC 211, 2005), wo Regeln für die Geometriemodellierung festgelegt werden (ISO/TC 211, 2003). Kottman and Reed (2009, p. 1) verwiesen für die folgende Definition des Begriffs **Feature** auf ISO 19101 – Reference model: „A feature is an abstraction of a real world phenomenon“ (ISO/TC 211, 2002). Das historische Gebäude „Zum kleinen Schlüssel“ (aus dem Beispiel) ist ein Feature im historischen Häuserbuch, da dieses abstrahierte bzw. generalisierte Objekt in diesem Projekt / im UoD von Interesse ist. Wie ausführlich dargelegt, sollen in Semantic Web Dokumenten Aussagen über dieses Feature gemacht werden oder Behauptungen aufgestellt werden. Der Terminus Feature wird für die zu generierenden RDF-Repräsentationen eine zentrale Rolle spielen. Es wird deshalb an dieser Stelle einfach gesagt, dass das Feature in dieser Arbeit als **Nicht-Informationsressource** angesehen wird. Dies wird an einer passenderen Stelle präzisiert, diskutiert und begründet (s. GP 6.1.2). Es wird die Auffassung vertreten, dass auch eine GML-Instanz als (Geo-)Web-Repräsentation der Informationsressource nur Beschreibungen des Features und nicht das Feature per se enthält. Das GML-Schema legt die Regeln fest, wie das historische Gebäude „Zum kleinen Schlüssel“ beschrieben werden kann. Hierfür wird im GML-Schema eine **Objektklasse** (bzw. ein Featuretyp/feature type, s.o.) definiert. Im Zuge der Klassendefinition werden die möglichen Eigenschaften (Attribute, Reihenfolge, Wertebereiche etc.) festgelegt, die alle Gebäude aufweisen können. Ein Individuum, z.B. das vorgenannte

historische Gebäude, ist eine konkrete Ausprägung – eine Instanz – der Gebäudeklasse. Der Name „Zum kleinen Schlüssel“ ist ein nicht-räumlicher Attributwert, der speziell für dieses Gebäude erfasst wurde, ebenso wie der räumliche Attributwert in Form von Koordinaten. Diese geben den Grundriss des historischen Gebäudes in generalisierter Form an. Es handelt sich um eine geometrische Eigenschaft. Um Geometrien modellieren zu können, werden geometrische Grundelemente benötigt, die im ISO-Standard 19107 – Spatial Schema definiert wurden (ISO/TC 211, 2003). Für diese Arbeit ist nur das **Simple-Feature-Modell** in der Version 1.1.0 relevant, das auf der konzeptionellen Ebene als ISO-Standard 19125-1 – Simple Feature Access (SFA) Part 1: Common Architecture definiert wurde (ISO/TC 211, 2004), welcher der OGC-Spezifikation mit der Nummer 05-126 entspricht (Ryden, 2005). Die vorgenannten Dokumente beschränken sich auf eine Auswahl von Geometriekonzepten aus dem ISO-Standard 19107 – Spatial Schema (ISO/TC 211, 2004; Ryden, 2005). Andrae (2013, p. 35) erklärte, dass die Simple-Feature-Standards so genannt werden, da „sie die verwendeten Geobjekte (Features) auf solche mit nicht komplexen, nichttopologischen, linear interpolierten Geometrien oder um daraus gebildete unstrukturierte Aggregate beschränken“. Dies gilt auch für die **GML-Implementierungsspezifikation** des OGC (Version 2.1.1). Diese spezifiziert das Standardausgabeformat des Basic WFS in der Version 1.0.0. Zwar wurde die GML-Implementierungsspezifikation (Cox et al., 2002) vor den o.g. Simple-Feature-Standards verabschiedet (ISO/TC 211, 2004; Ryden, 2005). Allerdings ist die folgende Übersicht zum SFA-Geometriemodell bzw. das Klassendiagramm, welches mithilfe der Unified Modeling Language (UML) erstellt wurde, ebenfalls in (fast) identischer Form in der o.g. GML-Implementierungsspezifikation enthalten (Cox et al., 2002).

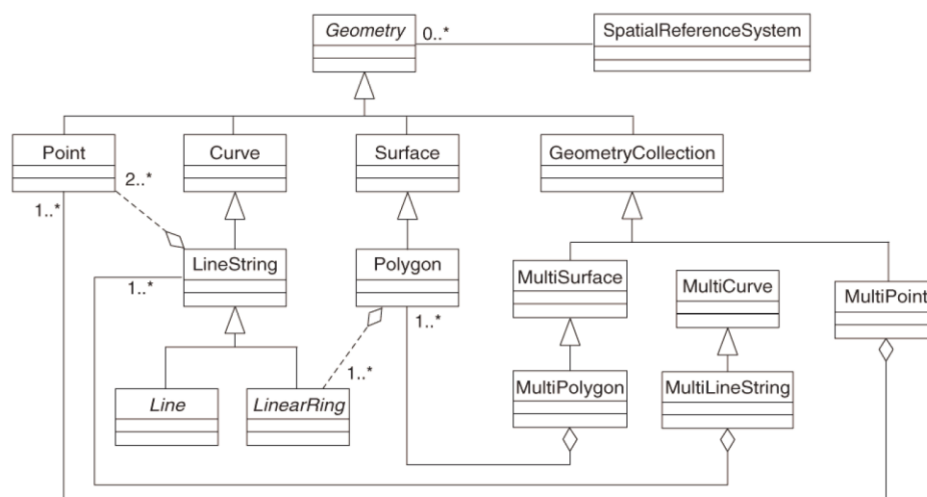


Abb. 2.24: Klassenhierarchie des SFA-Geometriemodells, OA von Ryden (2005, p. 11)

Cox et al. (2002) wiesen jedoch darauf hin, dass die Umsetzung in GML ohne die Klassen Curve, Surface, MultiSurface und MultiCurve auskommt. Im Übrigen gelten in GML2 aber die o.g. Einschränkungen bzgl. der geometrisch-topologischen Eigenschaften, weshalb nur einfache Geometrie-Beschreibungen nach dem SFA-Geometriemodell gespeichert bzw. transferiert werden können (Cox et al., 2002). Für Details zu GML2 wird auf Cox et al. (2002) verwiesen. An dieser Stelle ist der Hinweis *sinnvoller*, dass die Datenmodelle GML und RDF große Ähnlichkeiten aufweisen. Schade and Cox (2010, p. 1) erklärten in ihrer Arbeit mit dem Titel „Linked Data in SDI or How **GML is not about Trees**“, dass sowohl RDF wie auch **GML als gerichteter Graph** konzipiert wurden. Sie betonten, dass die Ausarbeitung einer frühen GML-Spezifikation (Version 1.0) – genauer gesagt ein OGC „Recommendation Paper“ (Lake and Cuthbert, 2000) – stark vom RDF-Datenmodell beeinflusst wurde (Schade and Cox, 2010). Das Dokument, das von Lake and Cuthbert (2000, p. 9) verfasst wurde, enthält ein Profil „for those *who are prepared to make use of RDF and RDF Schema*“ und eine RDFS-Spezifikation der GML 1.0.

Damit sollten Konsumentinnen in die Lage versetzt werden, dass sie die Bedeutung der Geoinformationen verstehen können, die auch beim system- bzw. anwendungsübergreifenden Informationsaustausch erhalten bleiben soll. Die Parallele zur Intention, die unter GP 2.2.1 und unter GP 2.2.2 erklärt wurde, ist offensichtlich. Es wäre – im *wahrsten Sinne* des Wortes – *unsinnig*, wenn Agent A und Agent B z.B. nicht auf die allgemein anerkannte Bedeutung des Konzeptworts „Polygon“ abstellen würden, sondern die Semantik von ihrer individuellen Interpretation und/oder ihrem spezifischen (Anwendungs-)Kontext abhängig machen würden. Möglicherweise *verwendeten zu wenige Personen* RDF(S) für die Kodierung/Beschreibung von Geoinformationen. Das RDFS-Profil und die RDFS-Spezifikation der GML sind jedenfalls in den darauffolgenden GML-Implementierungsspezifikationen nicht mehr enthalten (Cox et al., 2002; Portele 2007; Portele 2012). Aus praktischer Sicht nimmt auch die Konzeption als gerichteter Graph eine untergeordnete Rolle ein, denn es dominieren hierarchische (XML-)Baumstrukturen in Gestalt von Anwendungsschemata und Instanzdokumenten mit einfacher Elementverschachtelung (Tschirner, 2012). Aus theoretischer Sicht ist dies frappierend, denn auch GML2/GML3 fußt auf dem „**object-property pattern** and include[s] **XML linking** [...] in a manner that is **isomorphic to RDF**“ (Schade and Cox, 2010, p. 2; DeRose et al., 2010). Das bedeutet, dass GML-Objekte, die in RDF-Tripeln Subjekt und Objekt sein können, über eine GML-Property, die im RDF-Tripel an Prädikatstelle steht, miteinander verknüpft werden (Tschirner, 2012) und dass sich (Quer-)Beziehungen zwischen GML-Objekten über XLinks nicht nur abbilden lassen, sondern dass (auch formal) angegeben werden kann, welche Rolle der Link spielt bzw. welche Bedeutung er hat (Harold and Means, 2004; DeRose et al., 2010). Mit XLinks lässt sich ein gerichteter Graph beschreiben (DeRose et al., 2010). In der Theorie sind XLinks sogar noch flexibler als RDF-Links, da die Erstgenannten sogar bi- und multidirektionale Verbindungen zulassen. In der Praxis ist zu prüfen, inwieweit XLinks in einer Applikation zweckgerichtet prozessiert werden können, was aber grundsätzlich auch für RDF-Links gilt. Schade and Cox (2010) hoben die o.g. Parallelen zwischen der XML-basierten GML und RDF anhand der XML-Serialisierung von RDF hervor. Da syntaktische Grundlagen in der vorliegenden Arbeit nicht behandelt werden, wird bzgl. der Details auf Schade and Cox (2010) verwiesen.

Die Quintessenz ist, dass Linked Geospatial (Open) Data²⁵ nicht nur mittels RDF-Serialisierungen, sondern auch in GML-Strukturen kodiert werden können. Da die GML-Syntax standardisiert wurde, können GML-Strukturen maschinell verarbeitet werden. GML erlaubt die semantische Beschreibung von Ressourcen sowie deren Verknüpfung. Dieses Potenzial von GML bleibt ungenutzt, wenn Geodaten in flachen XML-Strukturen ohne Querverbindungen kodiert werden. Ein wichtiges Charakteristikum von GML und RDF ist **nicht** die Kodierung als hierarchische Baumstruktur²⁶, sondern das grundlegende Datenmodell – ein gerichteter Graph.

²⁵ Nachfolgend mit „LGD“ abgekürzt.

²⁶ RDF/XML ist ohnehin nur eine von vielen RDF-Serialisierungen.

2.3.4 Moderne Systeme, OGC-Standards

In den vorherigen Abschnitten wurden OGC-Standards resp. OGC-Implementierungsspezifikationen behandelt, die einen hohen Praxisbezug haben. Dies liegt daran, dass die entsprechenden, damals existierenden, Implementierungsspezifikationen verwendet wurden, um traditionelle GDI zu realisieren. Wie dargelegt waren dafür u. a. der WFS und die GML essentiell. Bereits der Draft „OGC Web Feature Service 3.0: Part 1 – Core“ zeigte jedoch, dass sich künftige OGC-Standards an modernen Ansätzen orientieren werden, die den Web-Architekturprinzipien entsprechen (Portele and Vretanos, 2019, p. 1). Im o.g. Draft wurde noch vom WFS gesprochen, mittlerweile nennt sich der Standard²⁷, genauer gesagt die Implementierungsspezifikation, jedoch „OGC API – Features - Part 1: Core“ (Portele et al., 2019, p. 1). Dort werden die minimalen Fähigkeiten definiert, die eine API-Features-Implementierung haben muss, um als standardkonform zu gelten (Portele et al., 2019). Vom WFS wird eigentlich nur noch an einer Stelle gesprochen, die erwähnenswert ist: „[...] the fine-grained access to features over the Web has been supported by the OGC Web Feature Service (WFS) standard (in ISO: ISO 19142) and **many** implementations of that standard for **many** years. WFS uses a Remote-Procedure-Call-over-HTTP architectural style using XML for any payloads. When the WFS standard was originally designed in the late 1990s and early 2000s this was the state-of-the-art“ (Portele et al., 2019, p. 17). Die OGC API Familie wendet sich von diesem Architekturstil ab. Für die Praxis ist der WFS aber noch relevant (Ivánová et al., 2020). In vielen GDI (z.B. im Kontext von INSPIRE) bildet der WFS (momentan) einen Grundpfeiler²⁸.

²⁷ Die Standards der OGC API Familie werden anhand des Typs der Ressource organisiert. Da sich diese Arbeit auf den WFS beschränkt, soll nur das API – Features angerissen werden. Weitere Informationen enthält die OGC API (OAPI) Common Specification: <https://github.com/engeospatial/ogcapi-common>

²⁸ Der Einsatz von OGC API – Features wird jedoch bereits empfohlen: <https://inspire.ec.europa.eu/good-practice/ogc-api—features-inspire-download-service>

2.4 Semantic Geospatial Web

Mit einer Aussage, die in diversen Publikationen im GI-Bereich bzw. in der GIScience häufig einleitend verwendet wird, lässt sich an die vorherigen Abschnitte anknüpfen. Auch Vilches-Blázquez and Saavedra (2019) erwähnten anfangs, dass ungefähr 80 % aller verfügbaren Informationen einen Raumbezug aufweisen. Dass dies auch hinsichtlich des Web of Data zutrifft, belegt nicht nur die Graphik, welche den (relativ) aktuellen Zustand der LOD-Cloud darstellt²⁹, sondern auch die Tatsache, dass etablierte Vokabulare³⁰, die räumliche Konzepte formalisieren, im Semantic Web sehr häufig (nach-)genutzt werden. Die o.g. Graphik lässt auch erkennen, dass räumliche Informationen in der LOD-Cloud oft verwendet werden, um Informationen verschiedener Domänen zu verknüpfen (Atemezing, 2015). Kurzum: Es ist sicher, dass Geoinformationen im WWW überaus wertvolle und wichtige Ressourcen sind, die interdisziplinär verwendet werden (Zhang et al., 2015).

Aus heutiger Sicht ist es deshalb besonders einleuchtend, dass Egenhofer (2002), kurze Zeit nachdem Berners-Lee et al. (2001) ihre Vision vom Semantic Web vorgestellt hatten, schon darauf hinwies, dass sich der Einsatz semantischer Standards/Technologien positiv auf die webbasierte Suche nach relevanten Geoinformationen und deren zweckgerichtete Verwendung auswirken könnte. Egenhofer (2002) stellte allerdings auch klar, dass hierfür **Ontologien** erstellt werden müssen, um die Bedeutung essentieller GI-Konzepte und -Termini adäquat formalisieren zu können, sodass diese von Maschinen prozessiert und von Menschen verstanden werden können. Egenhofer (2002, p.1) betonte, dass eine **kanonische Form** zwecks räumlicher Abfragen und für die Evaluation der spezifizierten Semantik die Basis für seine Vision vom „**Semantic Geospatial Web**“ bildet – ein Web, in dem die semantische Suche und Prozessierung von Geoinformationen möglich ist. Egenhofer (2002, p. 3) bezeichnete das „Semantic Geospatial Web“ als „stepping-stone toward the Semantic Spatial Web“.

In der vorliegenden Arbeit wird stets vom Semantic Geospatial Web (SGW) gesprochen, da sowohl die webbasierte semantische Suche und Prozessierung von Geoinformationen, wie auch viele weitere Problemstellungen, die Egenhofer nicht explizit benannte bzw. vorhersehen konnte, immer noch an der Forschungsfront in diesem Teilbereich der GIScience angesiedelt sind (Kokla and Guilbert, 2020). Diesbezüglich können z.B. Semantic Web Services (SWS), das Internet of Things (IoT) sowie Volunteered Geographic Information (VGI) und das Semantic Sensor Web (SSW) genannt werden (Kokla and Guilbert, 2020; Vilches-Blázquez and Saavedra, 2019; Guo et al., 2017; Salk et al., 2016). „Linked Spatial Data Infrastructures“ (López-Pellicer et al., 2012, p. 10) und „Linked Data Spatial Data Infrastructure[s]“ (Homburg, 2020, p. 7) sind ebenfalls Teil des Geospatial Ecosystems. Tandy et al. (2017) schrieben sogar: „spatial data infrastructures are a key component of the broader spatial data ecosystem“. Ob dies (noch) zutrifft, kann dahinstehen. Fakt ist jedenfalls, dass viele Wissenschaftlerinnen im Rahmen des **Semantic Enablements** eruierten, wie OGC-Standards, welche die Basis von (traditionellen) GDI bilden, mit **semantischen Standards und Empfehlungen** zusammengebracht werden können – „[to] bridge the gap to the semantic web community“ (Schade and Cox, 2010, p. 2). Aus der komplementären Verwendung von OGC- und Semantic Web-Standards können sich einige Vorteile ergeben, z.B. eine verbesserte Auffindbarkeit und ein vereinfachter Zugang zu Geoinformationen und Geodiensten – insbesondere für Nicht-Experten, denen das Anwendungsmuster Publish/Find/Bind/[Chain] nicht geläufig ist (Bernard et al., 2005; van den Brink et al., 2014; Vilches-Blázquez and Saavedra, 2019; Ivánová et al. 2020).

²⁹ Siehe Graphik bzw. Legende („Geography“): <https://lod-cloud.net/>

³⁰ Für ein einfaches Beispiel siehe: <https://lov.linkeddata.es/dataset/lov/vocabs/geo>

Nachfolgend werden lediglich die Resultate aus dem weiten (Forschungs-)Feld des GSW behandelt, die oben schon **hervorgehoben** wurden, weil sie für die Zielerreichung relevant sind, und zwar:

- Ansätze und Applikationen als Ergebnisse des Semantic Enablements, sofern sie sich auf den Basic WFS beziehen.
- Die o.g. kanonische Form, die laut Battle and Kolas (2011) mit dem OGC-Standard namens GeoSPARQL vorliegt.
- Weitere Standards, z.B. eine Domänenontologie, die als ISO-Standard vorliegt sowie Empfehlungen zwecks Publikation von LGD respektive „Linked Geospatial Cultural Heritage Data“ (Nishanbaev et al., 2020, p. 885) – eine treffende Bezeichnung³¹ für die Ausgangsdaten (s. Kapitel 4 und 5).

Warum bzw. für welches Teilziel die o.g. Punkte wichtig sind, wird in Abschnitt 2.5 hervorgehoben, der eine bündige Zusammenfassung enthält und der als Brücke zum praktischen Teil der Arbeit dient.

2.4.1 Semantic Enablement, Uplift, Downlift

Unter Gliederungspunkt 2.3.1 wurde ausgeführt, dass die ISO, das OGC und das W3C zusammenarbeiten, um die syntaktische Interoperabilität im Geo Web über De-facto- und De-Jure-Standards sicherzustellen. Allerdings stellten López-Pellicer et al. (2012) fest, dass dies hinsichtlich semantischer Aspekte der Interoperabilität, die Lieberman (2006) aufzählte³², sehr lange nicht der Fall war. Dies ist frappierend, da die ISO/TC 211 bereits 2006 eine Gruppe einsetzte, um zu untersuchen, wie die Standards des Geo Web von Technologien und Methodiken des Semantic Web profitieren könnten (López-Pellicer et al., 2012). Im selben Jahr wurde vom OGC das „Geospatial Semantic Web Interoperability Experiment (GSW.IE)“ durchgeführt, um herauszufinden, ob sich OWS mittels semantischer Konzepte und Technologien augmentieren³³ lassen, was auch als „**Semantic Enablement**“ bezeichnet wird (Tschirner, 2012, p. 42). Allerdings erwuchs daraus vorerst kein „liaison agreement“ zwischen der ISO/TC 211 und dem W3C bzw. zwischen dem OGC und dem W3C, die lediglich zwecks Standardisierung der Scalable Vector Graphics (SVG) kooperierten (López-Pellicer et al., 2012, p. 7).

Das GSW.IE schlug u. a. eine Referenzarchitektur und das Semantic Enablement auf Implementierungsebene (für den WFS und den CSW) vor (Lieberman, 2006). Die Semantik wurde dabei mittels der OWL spezifiziert; hinsichtlich des WFS z.B. auf Schema- und Instanzebene (Lieberman, 2006). Eine wichtige Erkenntnis des GSW.IE war diesbezüglich, dass das Semantic Enablement **nicht** notwendigerweise die Integration semantischer Konzepte in die jeweiligen (Implementierungs-)Spezifikationen bzw. die Adaption produktiver Dienste in traditionellen GDI erfordert (Lieberman, 2006). Stattdessen könnte das Semantic Enablement auch über eine **zusätzliche Komponente** realisiert werden, die in einer bestimmten Weise mit dem bestehenden OWS interagiert, z.B. indem sie einem WFS oder CSW vorgeschaltet wird (Lieberman, 2006). Je nach Intention, Funktionsweise/Implementierung kann eine solche Zusatzkomponente als Proxy, Translations-, Übersetzungs- oder Abstraktionsschicht (Lieberman, 2006); aber auch als Fassade (Schade and Cox, 2010), Wrapper (van den Brink et al., 2014), Crawler (Florczyk et al., 2010), Adapter (Jones et al., 2014; Homburg, 2020) usw. bezeichnet werden³⁴. Vilches-Blázquez and Saavedra (2019, p. 29) verwendeten in diesem Kontext den Oberbegriff „middleware“. In dieser Arbeit – d.h. ohne Anspruch auf Allgemeingültigkeit – wird das Konstrukt

³¹ Nachfolgend mit „LGCHD“ abgekürzt.

³² Zur Definitionsproblematik bzgl. der „semantischen Interoperabilität“ siehe Lieberman (2006), Seite 10, Gliederungspunkt 3.2. Für eine Definition im Kontext von OWS siehe Hobona and Brackin (2015), Seite 6, Gliederungspunkt 3.6.

³³ Es ist das Augmentieren von Datenstrukturen, Diensten und deren Spezifikationen gemeint, d.h., diese werden basierend auf semantischen Standards erweitert.

³⁴ Mitunter werden die genannten Begriffe auch einfach synonym verwendet.

„**LOD-Wrapper**“ verwendet. Damit wird eine Zusatzkomponente bezeichnet, die einem Basic WFS vorgeschaltet ist, um SPARQL-Anfragen in WFS-Anfragen zu übersetzen und WFS-Antworten (je nach Operation: XML-Dokument, XSD, GML-Instanzen) mittels RDF(S)/OWL und einer oder mehrerer Serialisierungen (Turtle, RDF/XML etc.) als Semantic Web Dokument(e) auszuliefern. Dabei handelt es sich um eine vereinfachte Sichtweise. Komplexere Ansätze z.B. die gleichzeitige Anfrage mehrerer WFS können ebenfalls über einen LOD-Wrapper realisiert werden, wie Tschirner (2012) bewies. Unabhängig vom Komplexitätsgrad kann der grundlegende Ansatz als „**Semantic Uplift**“ bezeichnet werden. „Semantic uplift is a process of converting existing structured or semi-structured data into LOD based upon semantic-web technologies“ (Abgaz et al., 2018, p. 13). „The Semantic Uplift describes the conversion from a not semantically annotated dataset to a semantically annotated dataset“ (Homburg, 2020). In dieser Arbeit – d.h. wieder ohne Anspruch auf Allgemeingültigkeit – wird noch das Konstrukt „**WFS-Adapter**“ verwendet. Damit wird eine Basic WFS-Implementierung bezeichnet, die WFS-Anfragen zuerst in SPARQL-Anfragen übersetzt, um diese an einen (oder mehrere) SPARQL-Endpunkte weiterzuleiten, die das SPARQL-Resultat anschließend entgegennimmt und es transformiert, um es (je nach Operation) als XML-Dokument, XSD, GML-Instanz auszuliefern. Solche WFS-Implementierungen können ebenfalls sehr komplex sein, wenn alle Operationen des letzten WFS-Standards (WFS-Version: 2.0.2) realisiert werden. Entscheidend ist, dass der o.g. Ansatz umgekehrt wird. Dementsprechend kann vom „**Semantic Downlift**“ gesprochen werden. „The process of converting ontological content to heterogeneous datasets is called Semantic downlift. A semantic downlift can be achieved by converting a SPARQL query result to a format common in the GIS world such as [...] GML [...]“ (Homburg, 2020, p. 22).

2.4.2 Empfehlungen, Standards

López-Pellicer et al. (2012) forderten vom ISO/TC 211 die Entwicklung von Richtlinien und Empfehlungen, die aufzeigen, wie Geodaten bzw. -informationen in die LOD-Cloud integriert werden können. Außerdem forderten López-Pellicer et al. (2012) die kooperative Erstellung von RDF-Vokabularen. Heute existiert eine **Empfehlung**, die vom **OGC und W3C** in der Spatial Data on the Web Working Group (SDWWG) gemeinsam erarbeitet wurde (Tandy et al., 2017). Die Empfehlung richtet sich auch an Datenbereitstellerinnen, die ihre Geoinformationen ins Web (of Data) integrieren möchten (Tandy et al., 2017). Sie ist sehr umfassend und berücksichtigt verschiedene Infrastrukturen, Anwendungen etc. (Tandy et al., 2017). Außerdem behandelten Tandy et al. (2017) auch viele weitere Prozesse des LOD-Lifecycles (s. Abb. 3.2), etwa die RDF-basierte Beschreibung räumlicher Informationen mithilfe des Geo-Vokabulars aus dem GeoSPARQL-Standard. In der vorliegenden Arbeit wird bei der Anforderungsdefinition und der späteren Umsetzung darauf geachtet, die Vorschläge von Tandy et al. (2017) mit den richtungsweisenden, domänenübergreifenden LOD-Empfehlungen von Heath and Bizer (2011) in Einklang zu bringen, wenn es möglich ist (s. GP 6.1 u. GP 7.2).

GeoSPARQL wurde vom OGC **standardisiert** (Perry and Herring, 2012). Im Gegensatz zum einleitend referenzierten Beispiel – das war das „W3C Basic Geo (WGS84 lat/long) Vokabular“ (Brickley, 2003) – ist GeoSPARQL³⁵ weder auf das CRS WGS84 noch auf die Beschreibung von Punktgeometrien beschränkt (Perry and Herring, 2012). Als *Vokabular* kann GeoSPARQL u. a. für die standardisierte, RDF-basierte Beschreibung von (komplexen) Geometrien verwendet werden (Perry and Herring, 2012). Da GeoSPARQL die W3C-Empfehlung(en) zu SPARQL (Harris and Seaborne, 2013; Feigenbaum et al., 2013) um räumliche Konzepte ergänzt, dient der OGC-Standard der Abfrage, der Suche und der Filterung von Geoinformationen (Perry and Herring, 2012) und deren Prozessierung³⁶ im Web (Battle and Kolas, 2011).

³⁵ Im OGC-Standard wird anschaulich erklärt, dass sich die meisten Geometrien, die mithilfe des W3C Basic Geo Vokabulars beschrieben wurden, recht einfach in eine GeoSPARQL-basierte Darstellungsform überführen lassen.

³⁶ Etwa bei der Verkettung mehrerer GeoSPARQL-Anfragen, die an jeweils an unterschiedliche Endpunkte gerichtet sind.

Für detaillierte, strukturierte Erklärungen hinsichtlich GeoSPARQL, insbesondere zu den topologischen und geometrischen **Filterfunktionen** und zum **Query Rewriting** räumlicher Anfragen wird auf Tschirner (2012) verwiesen, der sich bereits in einem frühen Stadium mit dem GeoSPARQL-Standard auseinandersetzt und diesen zusätzlich kommentiert. In Kapitel 5 wird aufgezeigt, wieso für die Kodierung der Geometrien aus dem historischen Häuserbuch lediglich die Prädikate **geo:hasGeometry** und **geo:asWKT** aus dem GeoSPARQL-Vokabular und das ASCII-basierte Format **Well-known text** (WKT) erforderlich sind. Unter GP 2.3.3 wurde SFA Part 1 (Version 1.1.0) eingeführt. WKT wird auf der konzeptionellen Ebene spezifiziert und ist für die Implementierungsebene³⁷ (OGC SFA-SQL / ISO 19125-2 SFA Part 2: SQL-Option) wichtig (Andrae, 2013). WKT³⁸ ist i.V.m. GeoSPARQL bzw. in Kombination mit dem **Geo-Vokabular** in der vorliegenden Arbeit nur für die standardisierte, RDF-basierte Kodierung von einfachen, zweidimensionalen Geometrien relevant.

Deshalb sollen nur die wesentlichen Aspekte des Geo-Vokabulars skizziert werden, statt den GeoSPARQL-Standard in seiner Gesamtheit zu thematisieren. Bezugnehmend auf Perry and Herring (2012) kann zunächst Folgendes festgehalten werden:

- Die Klassen **geo:Feature** und **geo:Geometry** haben dieselbe Superklasse (**geo:SpatialObject**). Tschirner (2012) ergänzte, dass diese Hierarchie konsistent zum GFM (ISO/TC 211, 2005) ist.
- **geo:Feature** und **geo:Geometry** sind disjunkte Klassen.
- **geo:Feature** ist die übergeordnete Klasse aller Feature-Typen und äquivalent zur Klasse **GFI_Feature** in ISO 19156.
- **geo:Geometry** ist die übergeordnete Klasse sämtlicher Geometrie-Typen und äquivalent zur Klasse **GM_Object** in ISO 19107.
- Für das Prädikat (oder Property) **geo:hasGeometry** gilt, dass der Definitionsbereich mit **geo:Feature** und der Wertebereich mit **geo:Geometry** angegeben ist. Dieses Prädikat ist als **owl:ObjectProperty** spezifiziert.
- Im Gegensatz dazu ist das Prädikat **geo:asWKT** als **owl:DatatypeProperty** spezifiziert. Der Definitionsbereich ist mit **geo:Geometry** angegeben und der Wertebereich mit **geo:wktLiteral**. Die letzte NIR-URI identifiziert einen Datentyp.

Diese Konzepte lassen sich anschaulich mithilfe der folgenden Abbildung visualisieren³⁹, die von Tschirner (2012, p. 43) übernommen wurde. Tschirner (2012) stellte die Konzepte in einem Graphen dar, dessen Notation der eines UML-Klassendiagramms entspricht. Das Klassendiagramm wurde auf die o.g. Punkte reduziert.

³⁷ Dies ist in dieser Arbeit bei der Auswahl der Datenhaltungskomponente zu beachten.

³⁸ Da sich die Ausführungen in dieser Arbeit immer auf SFA Version 1.1.0 beziehen, sei gesagt, dass die Geometrikodierung im WKT-Format den neueren SFA-Spezifikationen nicht zuwider läuft.

³⁹ Für eine alternative Darstellungsform der Klassenhierarchie, die mithilfe eines Ontologie-Editors erstellt wurde und deshalb keine Prädikate zeigt, siehe: https://www.w3.org/2015/spatial/wiki/Further_development_of_GeoSPARQL

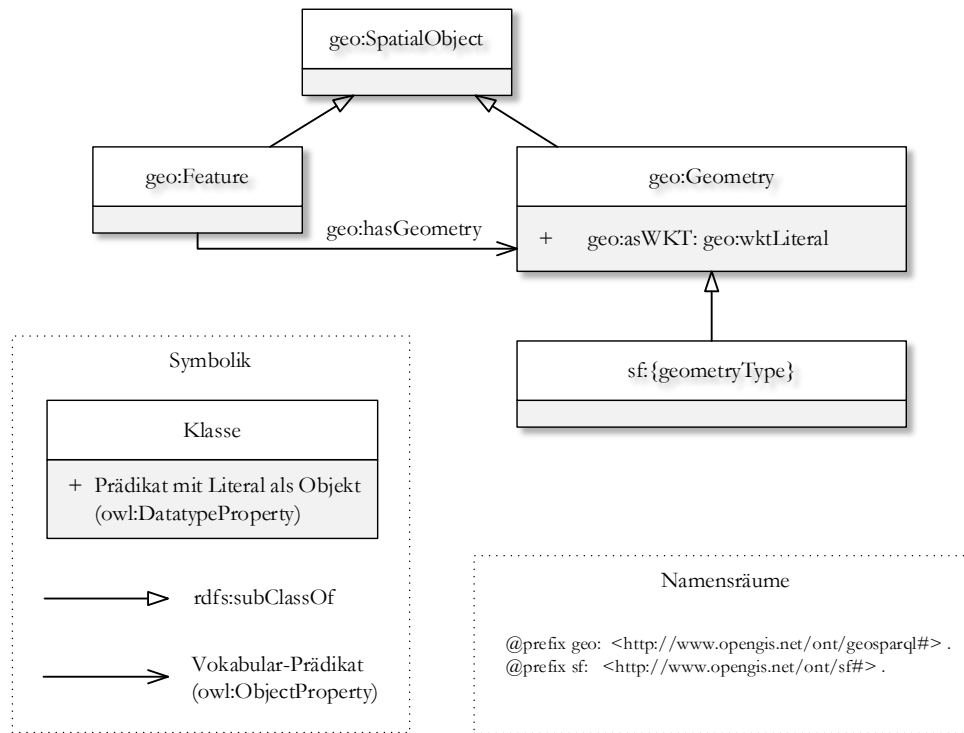


Abb. 2.25: GeoSPARQL-Konzepte, OA von Tschirner (2012, p. 43), angepasste Darstellung

Hinsichtlich der zu generierenden ABox sind die folgenden Punkte relevant:

- Es ist nicht unbedingt erforderlich die (instanziierte) Geometrie explizit zu typen. Dann muss man sich allerdings darauf verlassen, dass GeoSPARQL-konforme Implementierungen auf den geometrischen Serialisierungen (hier nur `geo:wktLiteral`) operieren, um den Geometrietyp direkt aus der Zeichenkette zu ermitteln (Tschirner, 2012). Diese Ableitung ist jedoch nicht mit den Inferenzen (konkrete Beispiele mit Inferenzregeln: GP 8.2.2) gleichzusetzen, die mit RDFS- oder OWL-Konstrukten möglich sind, denn für die Interpretation der geometrischen Serialisierung ist die Kenntnis eines domänenspezifischen Formats (WKT) notwendig, das in SFA CA spezifiziert wurde. Hitzler et al. (2008) erwähnten in diesem Zusammenhang, dass im RDF die Verwendung der Datentypen von XML Schema empfohlen wird, da diese etabliert sind und somit von vielen Applikationen erkannt und unterstützt werden. Prinzipiell kann eine RDF-basierte Implementierung bzw. eine LOD-Applikation jedoch nicht feststellen, ob eine bestimmte URI einen Datentyp identifiziert (Hitzler et al., 2008). Das Problem ist in einer (quasi geschlossenen) Umgebung mit versierten GI-Anwenderinnen, Entwicklerinnen etc. weniger relevant. Bei der (beabsichtigten) *domänenübergreifenden* Bereitstellung⁴⁰ von LGCHD ist es aber auch möglich, die Geometrie explizit zu typen, indem eine NIR-URI referenziert wird, welche dem Muster in der obigen Abbildung entspricht. Wie erwähnt, ist es hier ausreichend, dass diese NIR-URIs Simple Feature-Konzepte identifizieren.
- Hinsichtlich des CRS sollten ebenfalls die grundlegenden Ausführungen aus den Abschnitten 2.1.3 und 2.2.5 einbezogen werden. Prinzipiell ist es nicht notwendig, das CRS explizit anzugeben (Perry and Herring, 2012). In diesem Fall schreibt der GeoSPARQL-Standard Folgendes vor: „The URI

⁴⁰ Es wird *betont*, dass es durchaus sinnvoll sein kann, einige Fakten explizit auszudrücken. Gerade im Kontext von LGCHD, könnte der Zugang für potenzielle Anwenderinnen erleichtert werden. Wie bereits erwähnt, muss prinzipiell nicht immer alles formalisiert werden. Es wird die Auffassung vertreten, dass auch Aussagen sinnvoll sind, welche die Wissensbasis nicht um logische Aussagen zwecks automatischer Ableitungen erweitern, sondern die lediglich von Menschen interpretiert werden können.

Nys et al. (2018) brachten zusätzlich die „Time Ontology in OWL“ (Cox and Little, 2020) ein, weil der GeoSPARQL-Standard momentan⁴⁵ keine Konzepte enthält, mit denen zeitliche Informationen modelliert, abgefragt bzw. prozessiert werden können. Raumzeitliche Abfragen und Modelle wurden auch von Koubarakis and Kyzirakos (2010, p. 426) adressiert, die das RDF-Modell (durch „stRDF“) und SPARQL (durch „stSPARQL“) erweiterten. Letztere Erweiterung hat große Ähnlichkeiten mit „SPARQL-ST“ (Perry et al., 2011, p. 61). Im GI-Bereich / in der GIScience sind allerdings noch viele weitere Konzepte wichtig, welche im GeoSPARQL-Standard nicht formalisiert wurden. Das ISO/TC 211 kam den o.g. Forderungen von López-Pellicer et al. (2012) jedoch nach und formalisierte viele Konzepte aus den ISO 191XX-Standards. Die Ontologien stehen auf GitHub⁴⁶ bereit; die Arbeit von Hbeich and Roxin (2020) enthält eine aktuelle Übersicht dazu.

2.5 Zusammenfassung

In diesem Kapitel wurde das theoretische Fundament für den empirischen Teil der Arbeit gelegt, indem die wichtigsten Begriffe definiert und anhand von Beispielen erklärt wurden. Dabei wurden die Gemeinsamkeiten und die wesentlichen Unterschiede zwischen dem Web of Documents und dem Web of Data herausgearbeitet, um letzteres verständlich charakterisieren zu können. Nachdem klargestellt wurde, welches Ziel das LOD-Vorhaben verfolgt, wurden anschließend die Grundlagen des Geo Web sowie die Ansätze, Standards und Empfehlungen aus dem weiten (Forschungs-)Feld des GSW erläutert, die für die vorliegende Arbeit bedeutsam sind.

Die wesentlichen Inhalte werden noch einmal stichpunktartig zusammengefasst. Dieser Abschnitt dient allerdings nicht nur der bündigen Rekapitulation, sondern er versteht sich als Brücke zum praktischen Teil der Thesis, weshalb nun hervorgehoben wird, wieso bestimmte *Punkte* für die Zielerreichung relevant sind (*s. Pfeile*).

- Das **WWW** wurde als Web of Documents, Web of Data, Geo Web usw. bezeichnet. Dies erfolgte zwecks Betonung bestimmter Aspekte, zur vergleichenden Untersuchung bzw. zur Fokussierung verschiedener Entwicklungsstadien des Web. Es wurde aber klargestellt, dass die LOD-Cloud kein separater Teil des Web ist und dass es dementsprechend **nur ein Web** gibt, das auf wenigen, simplen – aber grundlegenden – Standards (URI, HTTP, RDF/HTML) und derselben Architektur bzw. denselben Architekturprinzipien (Universalität, Verlinkung usw.) basiert. Dezentralität und Offenheit sind essentiell für die Expansion des Web bzw. für das Linked Data Vorhaben, das einen möglichst starken Netzwerk-Effekt bezweckt. Die LOD-Cloud manifestiert sich als globales „universal linked information system“ und damit als das WWW, das Berners-Lee (1989) ursprünglich bereits vorgesehen hatte.
- ➔ Die Generierung eines LOD-konformen Graphen bedingt zumindest den Einsatz von drei der o.g. Standards; die zu definierenden LOD-Anforderungen entspringen den Web-Architekturprinzipien.
- Die **LOD-Methode** ist eine Handlungsvorschrift, die auf vier einfachen Regeln, den LOD-Prinzipien, basiert. Sie zielt auf die Integration strukturierter Informationen ins Web durch deren Verlinkung ab. Es soll ein semantisches Netz aus verknüpften, maschinenverarbeitbaren Informationen geschaffen werden, um die menschliche Entscheidungsfindung durch maschinelle Prozessierungen unterstützen zu können. Die ursprüngliche Vision vom semantischen Netz wurde von Ansätzen aus dem Forschungsfeld der künstlichen Intelligenz beeinflusst und zeichnete sich durch eine starke Fokussierung auf die vollautomatische programmgestützte Aufgabenerledigung etwa mittels ontologischer Schluss-

⁴⁵ Stand: Dezember 2020;

GeoSPARQL wird aktiv weiterentwickelt: <https://github.com/opengeospatial/ogc-geosparql/tree/master/1.1>

⁴⁶ Siehe: <https://github.com/ISO-TC211/ontologies>

folgerungsmethoden aus. Mit der sukzessiven Komplexitätsreduktion rückten auch die Vorteile verlinkter Informationen für menschliche Akteure⁴⁷ ins Zentrum der Aufmerksamkeit: Verknüpfte Informationen (in einer standardisierten Syntax) mit (formal) spezifizierter Bedeutung verbergen sich nicht in Silos, sondern können von potenziellen Anwenderinnen verstanden und damit in unvorherzusehenden Anwendungsszenarien nachgenutzt werden. Neben Anwenderinnen können auch Datenbereiterinnen vom Netzwerk-Effekt profitieren: Durch Verlinkung der eigenen Informationen mit Ressourcen aus der LOD-Cloud kann die IQ verbessert werden (z.B. das Wertschöpfungspotenzial).

➔ Der obige Punkt ist vor allem für die praxisnahe Leitfrage dieser Arbeit relevant: Am Beispiel eines „Geodaten-Silos“, in dem sich vektorielle Geoobjekte mit verknüpften Sachdaten verbergen, wird untersucht, ob die LOD-Methode auch angewendet werden kann, um diese Art von Daten umzuformen, um ihnen eine Bedeutung zuzuordnen und um sie durch Verlinkung ins Web zu integrieren – bildlich gesprochen – „um das Silo zu öffnen, weil dadurch der Wert des Inhalts gesteigert werden kann“. Aus der Sicht einer Datenbereiterin wird überprüft, ob und in welchem Grad die verlinkten Informationen den LOD-Anforderungen genügen. Aus der Sicht einer potenziellen Anwenderin soll demonstriert werden, dass die IQ durch die Verlinkung verbessert werden kann, wobei eine zweckabhängige IQ-Dimension („Informationsumfang“) fokussiert wird. Insofern werden die positiven Aspekte des Netzwerk-Effekts aus menschlicher Sicht untersucht; die programmgestützte Ableitung impliziten Wissens nimmt eine untergeordnete Rolle ein.

- Die breite Akzeptanz und die weitreichende Umsetzung des herstellerneutralen HTML-Standards war eine Grundvoraussetzung für die rasante Expansion des WWW. Diese Auszeichnungssprache ist quasi die **Lingua franca** des Web of Documents. Selbiges gilt hinsichtlich des RDF, das die Basis des Semantic Web bildet. RDFa ist eine „Mischform“, die für die vorliegende Arbeit keine Relevanz hat. Die XML-basierte GML war eine lange Zeit die wichtigste Auszeichnungssprache für die interoperable Übertragung vektorieller Geodaten (GML) und der korrespondierenden Metadaten (XML) im Geo Web. GML und RDF basieren auf ähnlichen Modellierungsansätzen (Graph, Object-Property-Muster).
- ➔ Die LOD-Prinzipien fordern die Verwendung von RDF und darauf aufbauenden Informationsspezifikationssprachen zwecks einheitlicher Repräsentation und Komposition unterschiedlicher Informationen aus diversen Beständen, die dezentral verwaltet werden. Einfach ausgedrückt: das Arbeitsergebnis / der resultierende RDF-Graph muss alle Kriterien aus dem simplen 5-Sterne-Schema erfüllen. Die vernetzten Informationen können in GML-Strukturen kodiert werden, die von vielen GIS importiert werden können. Im Rahmen der Demonstration könnten die Analyse- und Kombinationsmöglichkeiten mittels SPARQL- und WFS-Clients vergleichend dargestellt werden.
- **Dezentralisierung & Verlinkung:** HTML-Dokumente werden über Hyperlinks verknüpft; alle Arten von Ressourcen werden über RDF-Links verknüpft. Beide Link-Typen sind unidirektional. Im Gegensatz zu Hyperlinks sind RDF-Links jedoch getypt – es wird explizit angegeben, welche Relation zwischen den Ressourcen besteht. In GML-Schemata und -Instanzen können bidirektionale und multidirektionale XLinks für die Kodierung von gerichteten Graphen bzw. für die Verbindung zwischen Objekten und für die Verknüpfung von (XML-)Dokumenten verwendet werden. Die Link-Semantik kann spezifiziert werden. Links können gezielt und nach dem Motto „follow your nose“ verfolgt werden, sofern der Link nicht tot ist. Letzteres ist aufgrund der dezentralen Konzeption des Web möglich: da

⁴⁷ Diese Aussagen beziehen sich auf den Interpretations-, Verknüpfungs- und Transformationsprozess, der von einem Menschen durchgeführt werden muss.

die Skalierbarkeit prioritär berücksichtigt wurde, gibt es keinen Mechanismus zwecks Sicherstellung der Link-Konsistenz, etwa einen zentralen Link-Index.

➔ Der Informationsumfang des RDF-Graphen soll durch die Erzeugung von RDF-Links auf der Instanzebene erhöht werden. Die Relation zwischen den Ressourcen, die im Tripel an Subjekt- und Objektstelle stehen, lässt sich als Querverbindung zwischen GML-Objekten über XLinks abbilden.

- **Zugang & Referenz:** Die HTML war/ist wichtig (s.o.). Sie war aber nie das zentrale Charakteristikum des Web. Das hervorstechende Merkmal des WWW sind URIs. Mit diesen persistenten, universellen Identifikatoren für Ressourcen wurde ein virtueller, globaler Namens- und Informationsraum – das Web – geschaffen. Das Web ist eine Initiative, welche die „Benennung“, ggf. die „Adressierung“ von Ressourcen über URIs und damit die Kreierung von FCOs bezweckt. Mit „Benennung“ ist *hier* die Bezugnahme, die **Referenz** auf die Ressource über einen Namen in einem virtuellen, abstrakten Namensraum gemeint, während die „Adressierung“ auf die Ermittlung einer physischen Lokation abzielt und damit dem „Informationszugang“, d.h., dem **Zugang** zu Informationen, die in Web-Repräsentationen kodiert sind, dient.
- **URLs & URIs:** Der Ausdruck „Web of Documents“ referiert auf das Netz aus verknüpften Hypertext-Dokumenten, welche durch die Dereferenzierung von **URLs** lokalisiert werden, um sie via HTTP zugänglich zu machen, wobei die Bedeutung der übertragenen Daten nur durch menschliche Interpretation ermittelt werden kann. Der Ausdruck „Web of Data“ referiert auf das semantische Netz, in dem (ggf. indirekt) dereferenzierbare **HTTP-URIs** entweder für den Zugang zu Web-Repräsentationen von Informationsressourcen oder als Name, d.h. für die Referenz auf Nicht-Informationsressourcen – das sind z.B. realweltliche Objekte, abstrakte oder fiktive Konzepte, die nicht über das Netz verschickt werden können – oder für beides verwendet werden.
- Darin liegt **der wesentliche Unterschied** zwischen dem Web of Documents und dem Web of Data bzw. dem Semantic Web. Letzteres ist ein RDF-Graph, der – abgesehen von leeren Knoten und Literalen – FCOs enthält. Auf HTML-Dokumente kann im semantischen Netz weiterhin referiert werden; die URL könnte in einem RDF-Tripel z.B. an Objektstelle stehen bzw. sinnvoll verlinkt werden (etwa unter Verwendung des Prädikats **foaf:page**). Im Gegensatz zum „klassischen“ HTML, d.h. ohne semantische Erweiterungen, ist RDF jedoch eine einfache formale Sprache, für die es ausdrucksstärkere Erweiterungen (RDFS, OWL) gibt, weshalb die Semantik der Ressourcen eindeutig spezifiziert werden kann. RDF-Beschreibungen (Semantic Web Dokumente) können dank standardisierter Serialisierungen wie z.B. RDF/XML unter Berücksichtigung der festgelegten Bedeutung maschinell prozessiert werden, wobei die Semantik auch beim anwendungs- oder systemübergreifenden Austausch erhalten bleibt.
- **Semantic Web Dokumente** enthalten regelmäßig Semantic Web URIs / NIR-URIs wie z.B. Hash-URIs als spezielle Ausprägung davon. Obwohl diese keinen Zugang zu Web-Repräsentationen der referierten (nicht digitalen) Ressource bieten können und trotz der prinzipiellen URI-Undurchsichtigkeit, kann ein Agent im Zuge der indirekten Dereferenzierung Informationen zum Typ der Ressource erhalten. Hierfür können die HTTP-Statuscodes interpretiert werden, deren Bedeutung die TAG speziell für das Semantic Web festgelegt hat. Selbst wenn der Referent eine Nicht-Informationsressource ist, können Informationen über das analoge Bezugsobjekt in Form digitaler **zugehöriger Beschreibungen** geliefert werden, die jedoch zwecks Vermeidung von URI-Kollisionen und mehrdeutigen Aussagen nicht über dieselbe (NIR-)URI identifiziert werden dürfen. Die 303-URI kann im Rahmen der Weiterleitung (engl. redirect) bekannt gemacht werden. Die adäquate Sprache mit welcher der Inhalt am besten kodiert wird, kann zwischen Client und Server ausgehandelt werden – die Bezeichnung

„Inhaltsvereinbarung“ ist in diesem Kontext jedoch missverständlich, da es hier nicht darum geht, den Inhalt zwischen Client und Server auszuhandeln, denn dieser soll (!) immer ident sein.

- Im Gegensatz zu natürlichsprachlichen Referenzen (z.B. über einen Namen) oder zu Referenzen mittels URNs, liegt **der entscheidende Vorteil von URI-Referenzen** darin, dass dank der zuvor beschriebenen Mechanismen **Informationen über das Bezugsobjekt** bereitgestellt bzw. über HTTP transferiert werden können, statt die NIR-URI in RDF-Tripeln ausschließlich repräsentativ verwenden zu können. Im dezentralen Informationsraum mit undurchsichtigen URIs, die sich als nicht eindeutige Namen für alle möglichen Arten von Ressourcen manifestieren, ist dies essenziell für die Kohärenz.
- ➔ Das Problem, das zwecks Erhöhung des Informationsumfangs (im Rahmen der Anreicherung) gelöst werden muss, ist *referentieller* Natur, was anhand des folgenden fiktiven Beispiels klar wird: Vorhanden sei die natürlichsprachliche Aussage: Das Haus X hatte eine Bewohnerin namens „Jane Doe“. Es sei das Subjekt und das Prädikat einmal außen vorgelassen, um das (*Bezugs-*)Objekt in den Mittelpunkt stellen zu können, auf das in der natürlichsprachlichen Aussage über eine Zeichenkette (anscheinend ein Name) referiert wird. Auch wenn auf die Nicht-Informationsressource in RDF-Tripeln über eine NIR-URI referiert wird, kann nicht näher bestimmt werden, wer Jane Doe eigentlich ist/war, denn die zugehörigen Beschreibungen enthalten auch nur die Informationen, dass sie mit dem Haus in einer (formal) bestimmten Beziehung stand und dass sie einen Namen (Literal als Objekt) hatte. Ziel ist es, auf Basis des Namens und der vorhandenen Kontextinformationen (z.B. die Beziehungen zwischen Individuen und der Zeitpunkt ihres Bestehens) Jane Doe zu identifizieren, um zu untersuchen, ob in einem anderen Informationsbestand (GND) eine NIR-URI existiert, die auf eine „ähnliche“ Jane Doe (z.B. eine jüngere) oder „dieselbe“ Jane Doe referiert. In letzterem Fall ist die NIR-URI ein URI-Alias (wegen der NUNA). Wenn im Zuge der indirekten Dereferenzierung mehr Informationen über Jane Doe (z.B. Aussagen, Ansichten etc.) in Form der zugehörigen Beschreibungen geliefert werden können, lohnt es sich die URI-Referenzen in eine (formal) bestimmte Beziehung zu setzen, da die Menge an Informationen über Jane Doe erhöht werden kann. Potenzielle Anwenderinnen können davon profitieren.
- Dies setzt aber auch voraus, dass das „Geodaten-Silo“ geöffnet wird (s.o.). Dafür muss die Semantik der darin verborgenen Objekte allgemeinverständlich spezifiziert werden, sodass sich die Bedeutung nicht nur der Datenbereiterin, sondern z.B. auch potenziellen Anwenderinnen erschließt. Dafür bedarf es zumindest einer **simplen Ontologie**. Ihre Zwecke und Eigenschaften wurden (speziell für die vorliegende Arbeit) wie folgt beschrieben: Es handelt sich um eine (maschinenverarbeitbare) Beschreibung der Konzepte (TBox), die mithilfe einer leichtgewichtigen Wissensrepräsentationssprache erstellt wird, um den Austausch bzw. das Teilen von Informationen, d.h. von Daten mit einer zugeordneten Bedeutung, zu ermöglichen (Primärzweck) und um implizites Wissen über einfache Schlussfolgerungsmethoden ableiten zu können (Sekundärzweck).
- Mit der OWL wurde eine ausdrucksstarke RDF-basierte Ontologiesprache standardisiert. Terminologisches Wissen lässt sich auch mithilfe der einfachen RDF-Erweiterung namens RDFS spezifizieren, die ebenfalls als W3C-Empfehlung vorliegt. Für die Modellierung komplexerer Zusammenhänge ist sie allerdings nicht ausreichend. RDFS-Plus ist ein Konzept von Allemang et al. (2020), das die Erweiterung von RDFS um einige, simple – aber essenzielle – OWL-Konstrukte vorsieht. RDFS-Plus qualifiziert sich damit als **leichtgewichtige Wissensrepräsentationssprache**.
- Die erläuterten Konzepte aus dem Geo-Vokabular des GeoSPARQL-Standards sind für die RDF-basierte Beschreibung von komplexen Geoinformationen geeignet. Sie sind auch wichtig für **Applikations- und Domänenontologien**, die räumliche Konzepte umfassen. Für die Generierung von

LGCHD ist die standardisierte Domänenontologie namens CIDOC CRM prädestiniert, die mit CRMgeo um GeoSPARQL-Konzepte erweitert wurde. Für RDF und SPARQL gibt es Erweiterungen, mit denen sich raumzeitliche Informationen modellieren und abfragen lassen.

- ➔ GeoSPARQL ist für die **Transformation** der Ausgangsdaten, genauer gesagt, für die RDF-basierte Kodierung der einfachen, zweidimensionalen Geometrien, relevant. Dies gilt analog für die **Konzeptualisierung**, die mithilfe der o.g. Domänenontologie (Mapping) und durch Spezialisierung der GeoSPARQL-Konzepte (Modellierung mit RDFS-Plus-Sprachmitteln) bewerkstelligt werden kann. Die **Demonstration** kann mittels GeoSPARQL-Abfragen erfolgen, wenn der OGC-Standard vom IS (insbesondere datenbankseitig) unterstützt wird.
- Im Rahmen des **Semantic Enablements** wurde untersucht, ob (und wie) OGC- und Semantic Web-Standards komplementär eingesetzt werden können. Eine wichtige Erkenntnis ist, dass dies auch ohne die Veränderung der ursprünglichen (Implementierungs-)Spezifikationen bzw. ohne die Adaption produktiver OWS in existierenden GDI möglich ist, indem z.B. eine **Zusatzkomponente** verwendet wird. Der grundlegende Ansatz zwecks sinnhafter Überführung von *einheitlich repräsentierten Informationen*⁴⁸ in eine *spezifische Form*⁴⁹ wurde als **Semantic Downlift** bezeichnet. Eine Zusatzkomponente, mit der dies erreicht werden kann, wurde **WFS-Adapter** genannt. Es wurde klargestellt, dass damit eine Basic WFS-Implementierung gemeint ist, die WFS-Anfragen zuerst in SPARQL-Anfragen übersetzt, um diese an einen (oder mehrere) SPARQL-Endpunkte weiterzuleiten, die das SPARQL-Resultat anschließend entgegennehmen und es transformiert, um es (je nach Operation) als XML-Dokument, XSD, GML-Instanz auszuliefern.
- ➔ Dieser Ansatz und der WFS-Adapter sind für die Demonstration relevant. Hierfür können auch GIS resp. WFS-Clients eingesetzt werden, die mit standardkonformen WFS interagieren und die gelieferten XML- / GML-Strukturen einlesen können.

⁴⁸ Vereinfachend wird angenommen, dass Daten zu einheitlich repräsentierten Informationen werden, wenn die Semantik mithilfe einer RDF-basierten Ontologiesprache spezifiziert wird und wenn sie für die domänenübergreifende Verwendung mithilfe des *generischen* RDF-Datenmodells anwendungsneutral dargestellt werden.

⁴⁹ Genauer gesagt, ist eine *domänenspezifische* Darstellungsform (z.B. eine GML-Repräsentation) gemeint, die auf bestimmte IS (z.B. GIS) ausgerichtet ist.

3 Aktueller Stand der Forschung

Im Folgenden werden die aktuellen Ansätze aus dem Forschungsfeld des (Geospatial) Semantic Web überblicksmäßig dargestellt, welche auf die Generierung und Publikation von LGD/LGCHD abzielen. Abgesehen von diesen speziellen Verfahrensweisen, auf die sich die vorliegende Arbeit methodisch bezieht, werden auch (prototypische) Implementierungen (Werkzeuge) vorgestellt, die als Forschungsergebnisse aus wissenschaftlichen Arbeiten zur effizienten Erreichung des vorgenannten Ziels eingesetzt werden könnten – „ohne das Rad wieder neu erfinden zu müssen“. Die Diskussion darüber, welche Vor- oder Nachteile sich bei der Anwendung bestimmter Verfahren und beim Einsatz konkreter Werkzeuge zwecks praktischer Realisierung der operativen Teilziele ergeben, erfolgt im Rahmen der Entwicklung des eigenen Lösungsansatzes (s. Abschnitt 7.2). Zunächst erfolgt eine bündige, neutrale Vorstellung einschließlich einer Kategorisierung, die aufgrund des Komplexitätsgrades des gesamten Workflows wesentlich zur Verständlichkeit der jeweiligen Teilschritte beiträgt.

Der Grund hierfür ist darin zu sehen, dass zur Verwirklichung des o.g. Ziels (immer noch) auf zahlreiche Standards, Techniken, Werkzeuge etc. und alternative oder sich komplementierende Ansätze zurückgegriffen werden kann, die López-Pellicer et al. (2012) mithilfe der folgenden Abbildung in eine gewisse Ordnung brachten:

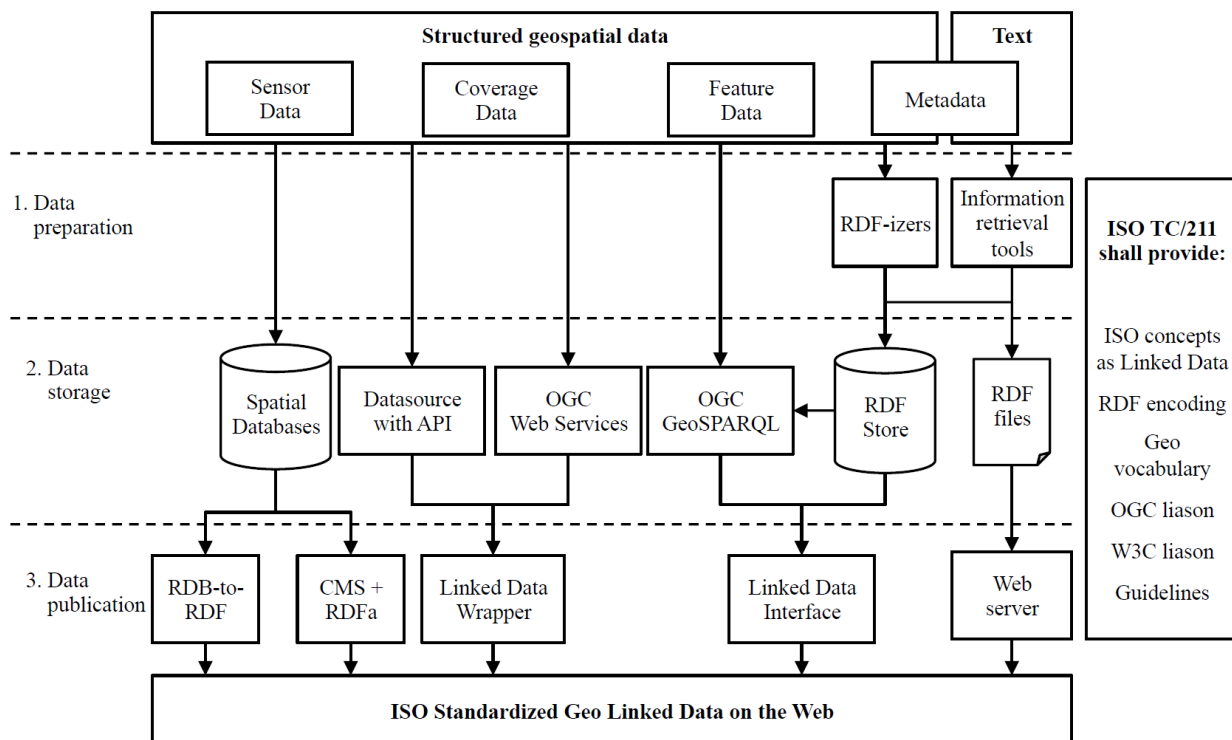


Abb. 3.1: *Linked Spatial Data Infrastructure*, OA von López-Pellicer et al. (2015, p. 10)

Die obige umfassende Abbildung wird zur Komplexitätsreduktion herangezogen. Es wird zunächst klargestellt, dass die **Verfahrensweise**, welche auf die Integration strukturierter, vektorieller Geodaten (oben als „Feature Data“ bezeichnet), inklusive der korrespondierenden Metadaten („Metadata“), unter Beachtung der LOD-Prinzipien, ins Web (of Data) abzielt, der Verständlichkeit halber in die folgenden Schritte aufgeteilt werden kann:

- (1) Analyse, Vorbereitung und Umformung der Daten (oben als „Data preparation“ bezeichnet)
- (2) Verwaltung und Speicherung der Daten/Informationen (oben als „Data storage“ bezeichnet)
- (3) Integration bzw. Publikation der Informationen (oben als „Data publication“ bezeichnet)

Dabei handelt es sich um eine vereinfachte Sichtweise. Die o.g. Teilschritte werden später weiter differenziert.

Die obigen Ausführungen suggerieren, dass die Integration von LGD/LGCHD ins Web durch die einmalige Ausführung der Teilschritte bzw. Teilprozesse erreicht werden kann. Tatsächlich ist dies nicht der Fall. Der gesamte Prozess zwecks Datenmanagement von LOD manifestiert sich als Zyklus. Die Teilprozesse werden in der folgenden Graphik durch verschiedenfarbige Kreise repräsentiert. Die grundsätzliche (aber nicht bindende) Ablaufreihenfolge wird durch verschiedenfarbige Pfeile dargestellt. Der sog. „LOD Lifecycle“ dient hier der vereinfachten Übersicht zum ganzheitlichen Datenmanagement von LOD in produktiven Umgebungen. Für Details zu den einzelnen Schritten wird auf die Arbeit von Auer and Lange (2014) verwiesen.

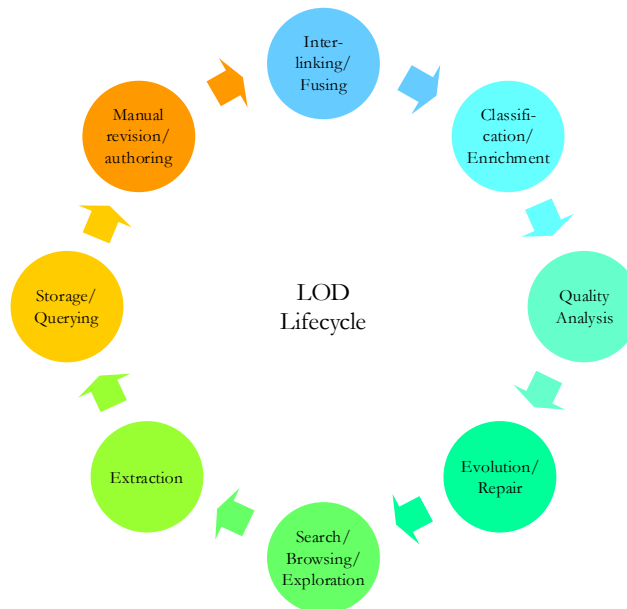


Abb. 3.2: LOD-Lifecycle, OA von Auer and Lange (2014, p. 8), angepasste Darstellung

In diesem Abschnitt soll die **Verfahrensweise** auf die Schritte reduziert werden, die für die Beantwortung der Leitfrage zwingend notwendig sind (s.u.). Da das IS, das auch zu Demonstrationszwecken erforderlich ist, nicht produktiv gestellt wird, werden einige der o.g. Abläufe im Rahmen dieser Arbeit nicht thematisiert bzw. umgesetzt (z.B. Aktualisierungsprozesse).

Zwecks Komplexitätsreduktion und Kategorisierung werden nun auch die relevanten Standards, Werkzeuge usw. fokussiert, die zwar insgesamt unter dem allgemeinen Begriff „**Hilfsmittel**“ subsumiert, jedoch auch den jeweiligen Fachwörtern untergeordnet werden können, die größtenteils in Abb. 3.1 vorhanden sind:

- (1) RDF-izer – worunter alle Werkzeuge verstanden werden, mit deren Hilfe RDF-Repräsentationen aus Daten oder Informationen generiert werden können, die zunächst nicht als Tripel/Quads vorliegen (z.B. Geoinformationen in proprietären Formaten).
- (2) Tripel- bzw. Quadstores – idealerweise mit GeoSPARQL-Unterstützung (s.o.) – sowie relationale/objektrelationale Datenbankmanagementsysteme (RDBMS/ORDBMS) mit räumlicher Erweiterung („Spatial Databases“), die u. a. zur Persistierung und Verwaltung von Geoinformationen eingesetzt werden.
- (3) Applikationen zwecks servicebasierter Bereitstellung von LGD/LGCHD, d.h. via OWS, Linked Data Wrapper, Linked Data Interfaces (darunter ist hier das GUI als Webinterface im engen Sinne aber auch der SPARQL-Endpunkt per se zu verstehen) sowie WFS-Adapter, die in der obigen Abbildung nicht explizit genannt werden. In der vorliegenden Arbeit werden diese (wegen der grundsätzlichen Parallelen zwischen RDF und GML) explizit in diese Kategorie eingegliedert.
- (4) Frameworks, die mehrere – aber nicht unbedingt alle – Hilfsmittel zusammenfassen.

Zuerst werden die wissenschaftlichen Arbeiten vorgestellt, die sich der o.g. **Verfahrensweise** widmen. Danach werden die o.g. **Hilfsmittel** vorgestellt, die – abgesehen von den Datenbanken – allesamt Resultate aus Forschungsprojekten sind. Vorab wird noch klargestellt, dass dabei lediglich Fachbeiträge berücksichtigt werden, die auf den bereits behandelten Grundlagen zum Geo Web (Eingrenzung auf den Basic-WFS in der Version 1.0 unter Verwendung von GML2), den vorgestellten Standards (GeoSPARQL als OGC-Standard, CIDOC-CRM als ISO-Standard usw.) und den skizzierten Ansätzen (Semantic Enablement, Uplift, Downlift) aufbauen. Dementsprechend bleiben z.B. Lösungsmöglichkeiten über Content-Management-Systeme (CMS) i.V.m. RDFa im weiteren Verlauf unberücksichtigt (vgl. Abb. 3.1). Selbiges gilt für die Bereitstellung (statischer) Textdateien (hier: TBox) über einen Webserver, was in der obigen Abbildung zwar visualisiert wurde, in wissenschaftlicher Hinsicht allerdings keine Relevanz hat.

3.1 Verfahrensweisen

Für die vorliegende Arbeit sind die folgenden Ansätze wichtig:

- (1) Es bedarf einer systematischen Handlungsvorschrift, die über eine festgelegte Schrittfolge vorgibt, wie aus vektorialen Geodaten LGCHD bzw. ein RDF-Graph generiert werden kann.
- (2) Es bedarf einer Strategie zwecks URI-Allokation, da prinzipiell alle Ressourcen in diesem RDF-Graphen über (indirekt) dereferenzierbare HTTP-URIs identifiziert werden müssen.
- (3) Zusätzlich ist ein Modellierungsansatz erforderlich, der auf die Konzeptualisierung von Inhalten aus historischen Karten abzielt.
- (4) Des Weiteren sind Vorschläge hilfreich, wie mit der „owl:sameAs-Problematik“ zwecks Generierung von sinnvollen RDF-Links auf der Instanzebene umgegangen werden kann.

Wie erwähnt, werden die jeweiligen Ansätze in diesem Abschnitt noch nicht im Detail diskutiert. Nach der bündigen Vorstellung wird aber kurz auf den Nutzen für die vorliegende Arbeit hingewiesen.

3.1.1 Generierung von LGCHD

Nishanbaev et al. (2020, p. 876) erarbeiteten eine **Methode**, die speziell auf die Generierung von LGCHD mithilfe von „current state-of-the-art geospatial RDF data generation tools and interlinking frameworks“ abzielt, die sie zuvor im Rahmen einer Evaluierung mit festgelegten Kriterien einer Geeignetheitsprüfung unterzogen. Die Methode kann angewendet werden, um aus räumlichen Daten, die historische Entitäten nicht mittels Tripelstrukturen und ohne Verknüpfungen zu anderen Datensätzen beschreiben, einen LOD-konformen RDF-Graphen zu erzeugen. Hierfür zerlegten Nishanbaev et al. (2020) den o.g. Workflow (vgl. GP 3) nicht in drei, sondern in fünf Teilschritte. Nishanbaev et al. (2020) visualisierten den Ablauf mithilfe der folgenden Graphik.

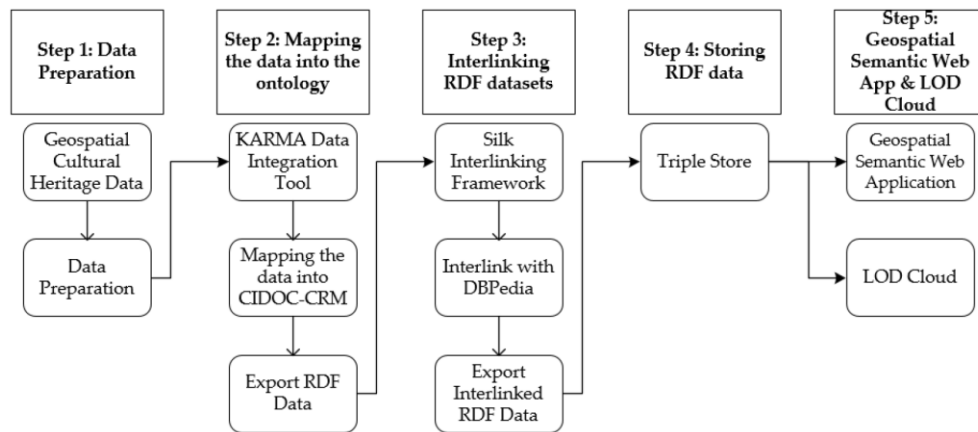


Abb. 3.3: Workflow zur Generierung von LGCHD, OA von Nishanbaev et al. (2020, p. 885)

Beim Vergleich der obigen Graphik mit der Darstellung von López-Pellicer et al. (2015) wird deutlich, dass die Teilschritte 2 und 3 (vgl. Abb. 3.1) im Prinzip unverändert sind. Auch die „Data Preparation“ ist weiterhin erforderlich; Nishanbaev et al. (2020) betonten jedoch das Mapping unter Verwendung des CIDOC-CRM und die Verlinkung der Ressourcen (auf der Instanzebene). Wenn man den Bezug zwischen den LOD-Prinzipien und den obigen Teilschritten herstellen möchte, lässt sich vereinfachend feststellen, dass dem ersten und zweiten LOD-Prinzip (URIs/HTTP-URIs) vor allem in Schritt 2, dem dritten LOD-Prinzip (RDF, SPARQL) in Schritt 4 und dem vierten LOD-Prinzip primär in Schritt 3 Rechnung getragen wird. Eigentlich gibt es aber keine klaren Trennlinien, weil im LOD-Kontext z.B. nicht nur die Verwendung von RDF-Links auf der Instanzebene (u. a. *Identity Links*), sondern auch auf der konzeptuellen Ebene (*Vocabulary Links*) propagiert wird (Heath and Bizer, 2011). Wichtiger als die Zuordnung der LOD-Prinzipien zu den jeweiligen Teilschritten ist die Tatsache, dass die obige Methode zwar angewendet werden kann, um einen RDF-Graphen zu generieren, der LOD-Anforderungen entspricht; jedoch muss der **Nutzen für die vorliegende Arbeit** erklärt werden. Zu Beginn dieses Abschnitts wurde schon erwähnt, dass Nishanbaev et al. (2020) die Werkzeuge für die Generierung von LGCHD fokussierten resp. evaluierten. Sie hoben Folgendes hervor: „this article presents a methodology demonstrating *how the evaluated tools and frameworks can be applied* to generate geospatial Linked Data“ (Nishanbaev et al., 2020, p. 876). In der vorliegenden Arbeit muss zwischen der Methode und den Hilfsmitteln differenziert werden. Dementsprechend muss die obige Verfahrensweise abstrahiert werden. Der Workflow soll nicht von vornherein auf bestimmte Werkzeuge ausgerichtet werden, sondern es soll später diskutiert werden, mit welchen Hilfsmitteln das Ziel am effizientesten erreicht werden kann. Zwecks Abstraktion kann auf die Graphik zurückgegriffen werden, die Nishanbaev et al. (2019, p. 1482) in einer früheren Arbeit verwendeten und als „Workflow: from cultural heritage data to the geospatial semantic web“ betitelten:

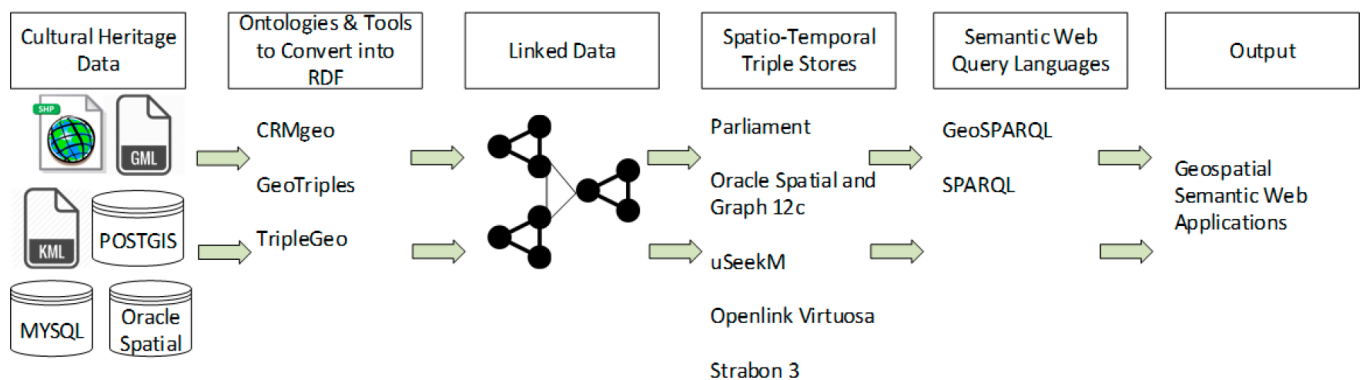


Abb. 3.4: Hilfsmittel für die Generierung von LGCHD, OA von Nishanbaev et al. (2019, p. 1482)

Beim Vergleich der beiden vorherigen Abbildungen fällt auf, dass z.B. „Karma“ ein RDF-izer ist, der zur Generierung von LGCHD eingesetzt werden kann. Allerdings existieren noch einige Alternativen wie z.B. „GeoTriples“. Dies gilt auch hinsichtlich des Datenmappings unter Verwendung einer Domänenontologie (CIDOC-CRM, ggf. i.V.m. CRMgeo). Stattdessen könnte auch die Modellierung einer Problemontologie zielführend sein. Kurzum: Wenn man für jeden Schritt alle sinnvollen Optionen berücksichtigt und dabei auch von bestimmten Modellen (z.B. RDF) sowie Informationsbeständen (z.B. Ressourcen aus DBpedia) abstrahiert, kann der Workflow in allgemeinere Teilschritte zerlegt werden:

- Schritt 1: **Vorbereitung der Daten:** Neben der obligatorischen Datenanalyse und der anschließenden Auswahl von geeigneten Daten, werden die Daten in diesem Schritt ggf. angepasst, bereinigt etc.
- Schritt 2: **Festlegung eines Datenmodells:** Prüfung, ob die ausgewählten Daten z.B. mithilfe des Relationenmodells oder „nativ im RDF“ gespeichert werden sollen.
- Schritt 3: **Generierung des Graphen:** Transformation/Umformung oder Mapping der Relationen, URI-Allokation, Modellierung einer Problemontologie oder Mapping mithilfe einer geeigneten Domänenontologie, Verknüpfung von Konzepten.
- Schritt 4: **Speicherung/Verwaltung:** Je nach Datenmodell muss z.B. ein RDF-Store mit GeoSPARQL-Unterstützung oder ein ORDBMS/RDBMS mit räumlicher Erweiterung zwecks physischer Persistierung/Konsolidierung der Daten- oder Wissensbasis ausgewählt werden.
- Schritt 5: **Verlinkung (auf der Instanzebene):** Ermittlung von URI-Aliasen oder URIs, die auf ein ähnliches Individuum referieren, um eine Aussage treffen zu können, welche Art von Beziehung zwischen den Referenzobjekten besteht, sodass die Informationsmenge über ein Individuum vergrößert werden kann.
- Schritt 6: **Publikation:** Bereitstellung der Informationen über GSW-Anwendungen und/oder LOD-Applikationen.

3.1.2 URI-Allokation

Insbesondere für die Umsetzung von Schritt 3 muss festgelegt werden, wie die notwendigen HTTP-URIs konstruiert werden sollen. van den Brink et al. (2014, p. 71) verwendeten zwecks Generierung von LGD die „Dutch URI strategy“. Die URI-**Strategie** wurde von Overbeek and van den Brink (2013) und Overbeek and Brentjens (2013) erarbeitet, die neben den einschlägigen W3C-Empfehlungen auch die Erkenntnisse aus Studien berücksichtigten, welche im Rahmen des EU-Programms namens „Interoperability Solutions for European Public Administrations“ (ISA) durchgeführt wurden. Neben „scalability, intelligibility, trust, machine-readability and human-readability“ wird vor allem die URI-Persistenz und die (indirekte) Dereferenzierbarkeit der verschiedenen URIs gewährleistet (van den Brink et al., 2014, p. 71). Die URI-Strategie sieht die Vermeidung von URI-Kollisionen durch die Verwendung von NIR-URIs/Hash-URIs und 303-URIs vor (van den Brink et al., 2014). Zwecks URI-Allokation griffen van den Brink et al. (2014, p. 71) auf das folgende „URI pattern“ zurück: `http://{domain}/{type}/{concept}/{reference}`

Zusammenfassend ist zu den URI-Bestandteilen Folgendes festzuhalten:

- {domain}** Dient neben der eindeutigen Identifikation von Ressourcen auch der Seriosität von URIs, wenn ein Agent erkennen kann, dass zwischen der URI und einem vertrauenerweckenden sozialen Gebilde (z.B. juristische Personen des öffentlichen Rechts) eine Relation besteht, wobei van den Brink et al. (2014) auch darauf hinwiesen, dass der Name einer Organisation kein Bestandteil der URI sein sollte. van den Brink et al. (2014) waren der Auffassung, dass dies die URI-Persistenz gefährdet, da sich Namen von Organisationen mit der Zeit verändern können.
- {type}** Diese Variable kann laut van den Brink et al. (2014) wie folgt belegt werden, um den URI-Typ anzugeben:
- id** für NIR-URIs
 - doc** für 303-URIs
 - def** für spezielle NIR-URIs – die Hash-URIs
- {concept}** van den Brink et al. (2014) erklärten, dass dieser URI-Bestandteil primär für menschliche Agenten vorgesehen ist. Für diese soll erkennbar sein, welches Konzept von der URI identifiziert wird, wobei auf die persistente Bezeichnung ebenfalls besonderen Wert gelegt werden sollte (van den Brink et al., 2014).
- {reference}** Diese Variable wird mit einer ID belegt (van den Brink et al., 2014).

Für die **vorliegende Arbeit** ist die skizzierte URI-Strategie von hoher Bedeutung. Der **Nutzen** liegt auf der Hand: Bei konsequenter Anwendung dieser Strategie wird dem zweiten LOD-Prinzip Rechnung getragen. Der resultierende Graph wird dann nicht nur die allgemeinen LOD-Anforderungen an URIs erfüllen, sondern es kann auch etwaigen Missverständnissen vorgebeugt werden, indem explizit ausgedrückt wird, welche Art von Ressource die jeweilige URI identifiziert. In vielen erfolgreichen LOD-Projekten wird ähnlich vorgegangen (DBpedia, GeoNames uvm.); die vorgestellte URI-Strategie speziell für LGD widerspricht den richtungsweisenden, domänenübergreifenden LOD-Empfehlungen von Heath and Bizer (2011) nicht.

3.1.3 Konzeptualisierung

Wie erwähnt, kommt im Rahmen von Schritt 3 statt des Mappings auf das CIDOC-CRM/CRMgeo auch die Modellierung einer dedizierten TBox in Frage. Genauer gesagt, sind für die vorliegende Arbeit **Modellierungsansätze** interessant, welche auf die Formalisierung historischer Konzepte in Form von (möglichst einfachen) Problemontologien abzielen, dabei aber berücksichtigen, dass nicht notwendigerweise alle vorhandenen Informationen auch spezifiziert werden können. Der Grund hierfür liegt darin, dass die Vektormodelle (s. Kapitel 5) auf Basis der (noch) vorzustellenden Primärquellen (s. Kapitel 4) erzeugt wurden. Da in diesem Zuge nicht alle Karteninhalte digitalisiert wurden, muss bei der Generierung des RDF-Graphen bedacht werden, dass dieser nicht alle Informationen enthalten wird. Trotzdem sollen damit nutzbare Antworten auf konkrete Fragen gegeben werden, die sich anhand der Ausgangsdaten nicht beantworten lassen.

Scheider et al. (2014) erarbeiteten modulare Problemontologien, um Aussagen über die Inhalte historischer Karten via RDF treffen zu können. Man könnte auch sagen, mit den Konzeptualisierungen von Scheider et al. (2014) lassen sich die Inhalte historischer Karten anders kodieren; dafür werden formale Sprachen (RDFS etc.) statt Kartenelemente (kartographische Signaturen etc.) verwendet. Das Wort „kodieren“ soll nicht suggerieren, dass Scheider et al. (2014) mithilfe von Wissensspezifikationsprachen tatsächlich alle Inhalte historischer Karten exakt und explizit beschreiben wollten. Scheider et al. (2014, p. 1) betonten: „We suggest ways to formally encode historic map content in an approximate intensional manner which still allows useful queries“. Scheider et al. (2014, p. 8) präsentierten auch einen konkreten Umsetzungsvorschlag: „A useful way of encoding such content graph patterns is to write them down as graphs with blank nodes instead of variables. Blank nodes [...] can be interpreted as existentially quantified variables, because they can be interpreted into any other node satisfying a query“. Dafür spezifizierten Scheider et al. (2014) einige Prädikate mittels OWL-Restriktionen. Abgesehen davon spezialisierten Scheider et al. (2014) auch Konzepte aus dem GeoSPARQL-Vokabular. Der **Nutzen für die vorliegende Arbeit** besteht darin, dass der letztgenannte Ansatz im Rahmen der Konzeptualisierung einfach übernommen werden kann. Auch die Verwendung von leeren Knoten ist eine Lösungsmöglichkeit für das Problem der fehlenden Informationen. Allerdings muss noch diskutiert werden, welche Vor- und Nachteile sich bei der Verwendung von anonymen Klassen im Vergleich zur Erzeugung von leeren Knoten auf der Instanzebene ergeben. Wegen des modularen Konzepts von Scheider et al. (2014, p. 9), kann auf das Semantic Web Dokument namens „historicmapsphen“ zurückgegriffen werden⁵⁰. Die nachfolgend dargestellten Klassen können gemäß individueller Anforderungen angepasst werden, wie Scheider et al. (2014) erwähnten. Die adaptierbare Klassenhierarchie ist für die Modellierung einer simplen TBox hilfreich, die sich im einfachsten Fall als eine Taxonomie von Konzepten manifestiert.

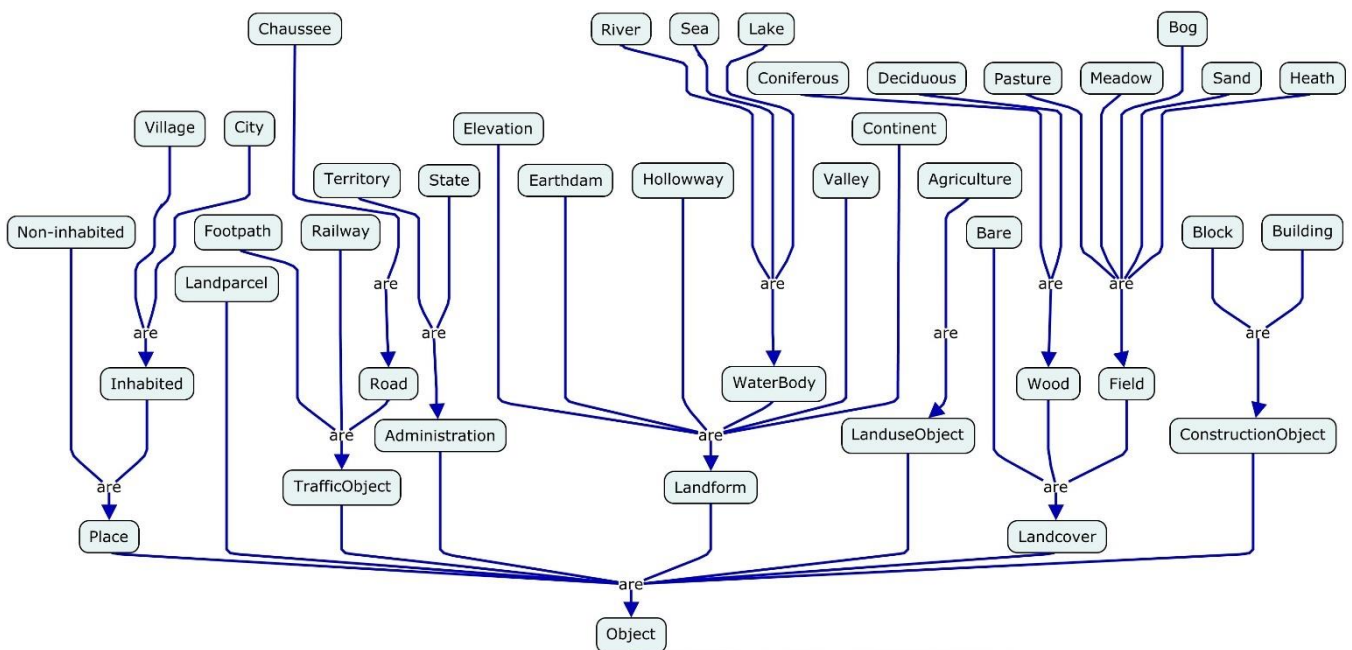


Abb. 3.5: Klassenhierarchie der Historicmapsphen-Ontologie, OA von Scheider et al. (2014, p. 9)

⁵⁰ <http://www.geographicknowledge.de/vocab/historicmapsphen#>

3.1.4 Verlinkung auf der Instanzebene

Es existieren Werkzeuge, mit deren Hilfe URI-Aliase ermittelt werden können, um sie für die Generierung von RDF-Links (genauer: *Identity Links*) zwischen Individuen auf der Instanzebene verwenden zu können. Es ist nicht unproblematisch, dass sich die ermittelten URIs häufig (offensichtlich) nicht als URI-Aliase qualifizieren, aber die jeweiligen Ressourcen trotzdem massenhaft unter Verwendung des Prädikats **owl:sameAs** verknüpft werden. Es ist fraglich, ob ein solches Vorgehen mit dem beabsichtigten Verwendungszweck bzw. mit der Semantik des vorgenannten Prädikats vereinbar ist, die Halpin et al. (2010, p. 307) unter Bezugnahme auf die OWL-Spezifikation wie folgt erklärten:

- „the built-in OWL property **owl:sameAs** links an individual to an individual“
- „Such an **owl:sameAs** statement indicates that two URI references actually refer to the same thing: the individuals have the same *identity*“

Zwecks Definition von *Identität* bezogen sich Halpin et al. (2010, p. 307) auf das Identitätsprinzip (Leibniz-Gesetz) und invertierten dieses: „if x and y share all *properties* (i.e. if they are indiscernable) then they are identical“. Im Kontext des Semantic Web ist die mögliche Inferenz relevant: „When someone says two things are the same, the two things share all the *same properties* and so every property of one thing can be inferred to be a property of the other“ (Halpin et al., 2010, p. 307). Hinsichtlich des LOD-Paradigmas stellten Halpin et al. (2010) und Halpin et al. (2015) Folgendes fest:

- In der LOD-Cloud wird das Prädikat **owl:sameAs** häufig nicht im Sinne der OWL-Spezifikation verwendet.
- Die formale Semantik des Prädikats bleibt oftmals unbeachtet, da Inferenzen im LOD-Kontext eine untergeordnete Rolle einnehmen und weil viele LOD-Applikationen nicht auf die Ableitung impliziten Wissens ausgerichtet sind.

Halpin et al. (2010) hoben hervor, dass die Problematik in der LOD-Cloud regelmäßig umgangen wird, indem RDF-Tripel, in denen das Subjekt und das Objekt über das Prädikat **owl:sameAs** verknüpft sind, nicht als objektive Aussagen zur Identität, sondern als Behauptungen, Meinungen etc. der jeweiligen Agenten angesehen werden, was Heath and Bizer (2011) und Halpin et al. (2015) bestätigten. Deshalb kann angenommen werden: „As different agents may have different sets of claims they accept, different agents may accept different identity statements and so have different inferences“ oder anders ausgedrückt: „it is one thing for me to link to your URI, but its another thing for me to believe what you say about it as though you were talking about my URI“ (Halpin et al., 2010, p. 309). Heath and Bizer (2011) empfahlen: „use owl:sameAs to express identity links, but always [...] keep in mind that the Web is a social system and that all its content needs to be treated as claims by different parties rather than as facts [...]“. Dies steht im Widerspruch zu den Ausführungen unter GP 2.2.2.

Es mutet allerdings auch seltsam an, dass die Generierung von Identity Links⁵¹ im LOD-Kontext propagiert – um nicht zu sagen explizit gefordert wird – und die W3C-Empfehlungen bzw. die Informationsspezifikationssprachen (RDFS, OWL) dafür nur wenige Prädikate bereitstellen. Ob das Prädikat **rdfs:seeAlso**, das (faktisch) keine formale Semantik besitzt, in diesem Zuge überhaupt genannt werden sollte, ist diskussionswürdig. Halpin et al. (2010, p. 311) erarbeiteten eine Ontologie namens „Similarity Ontology (SO)“. Sie boten weitere Prädikate an, mit denen nicht nur ausgedrückt werden kann, dass zwei URI-Referenzen auf denselben Referenten referieren (Halpin et al., 2010). Mithilfe der SO lassen sich auch unterschiedliche Arten von Ähnlichkeiten zwischen Individuen angeben (Halpin et al., 2010). Die folgende Graphik

⁵¹ Dies gilt insbesondere wegen des SKOS nicht für Vocabulary Links.

zeigt die Hierarchie der Properties in der SO und ihre Relationen zu den Prädikaten in den o.g. W3C-Empfehlungen (Halpin et al., 2010):

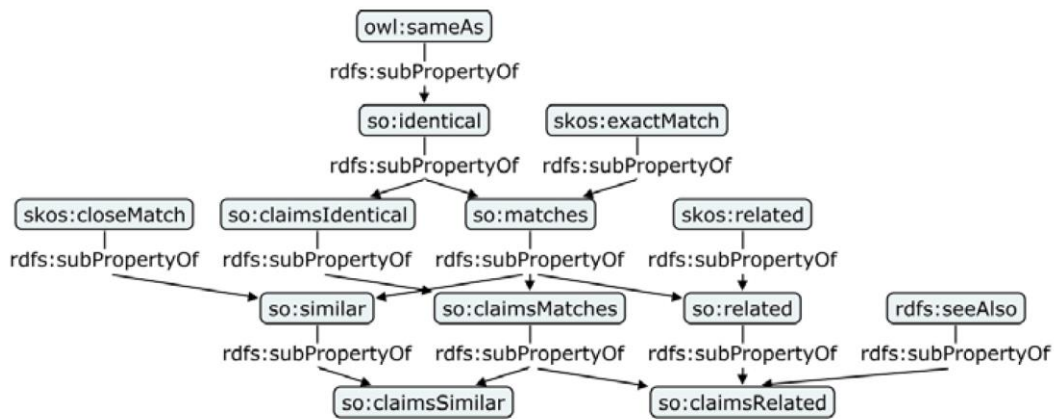


Abb. 3.6: Graphischer Auszug aus der SO, OA von Halpin et al. (2010, p. 312)

Für die **vorliegende Arbeit** ist die skizzierte **Problematik** von hoher Bedeutung, da die Informationsmenge über Individuen durch die Generierung von RDF-Links auf der Instanzebene vergrößert werden soll. Es muss festgelegt werden, ob dabei ausschließlich das Prädikat **owl:sameAs** eingesetzt und die Problematik damit einfach umgangen werden soll oder ob besser auf andere Prädikate zurückgegriffen werden soll, wenn es offensichtlich ist, dass zwei URIs nicht auf dasselbe historische Individuum referieren, die Bezugsobjekte jedoch bestimmte Ähnlichkeiten aufweisen. In letzterem Fall könnten die Prädikate aus der SO (nach-)genutzt werden. Halpin et al. (2010, p. 311) stellten den **Nutzen dieser Ontologie** im Rahmen eines Experiments unter Beweis; sie stellten abschließend allerdings fest: „However, the Similarity Ontology explores too large of a design space to be reliably deployed. A simple similarity property would be quite useful to add to RDF, such as **sub-property of rdfs:seeAlso**“. Diese Spezialisierung ist nicht erfolgt und hinsichtlich des Prädikats **rdfs:seeAlso** stellt sich die Frage, ob dieses überhaupt auf NIR-URIs an Objektstelle referenzieren sollte⁵². Die Frage, ob sich die SO in der LOD-Cloud etabliert hat, kann schon wegen der OWA nicht abschließend beantwortet werden. Eine eingeschränkte Momentaufnahme führte allerdings zu der Erkenntnis, dass die Prädikate aus der SO in erfolgreichen LOD-Projekten (z.B. DBpedia = Nukleus der LOD-Cloud) nicht verwendet werden⁵³. Dies wird im Rahmen der Entwicklung des eigenen Lösungsansatzes diskutiert (ab GP 7.2).

⁵² Siehe: <https://www.w3.org/wiki/UsingSeeAlso>

⁵³ Stand: Dezember 2020

3.2 Hilfsmittel

3.2.1 RDF-izer

Für die Transformation von **räumlichen Informationen** ins RDF wurden einige RDF-izer entwickelt, die verschiedenen funktionalen und nicht funktionalen Anforderungen gerecht werden. Beispiele dafür sind „Geometry2RDF“ (Vilches-Blázquez et al., 2010; Vilches-Blázquez and Saavedra, 2019, p. 28), „TripleGeo“ (Patroumpas et al., 2014, p. 1) und „GeoTriples“ (Kyzirakos et al., 2014; Kyzirakos et al., 2018, p. 16). Die Arbeit von Nishanbaev et al. (2020, p. 880) enthält eine aktuelle Tabelle mit wichtigen Details zu den jeweiligen RDF-izern (dort als „RDF generation tools“ bezeichnet) und die Links zu GitHub, wo der Quellcode zur Verfügung steht.

Nachfolgend werden die Funktionen, Gemeinsamkeiten und Unterschiede der o.g. RDF-izer überblicksmäßig dargestellt, statt sämtliche Details zu eruieren. Dafür werden die RDF-izer in chronologischer Reihenfolge vorgestellt. Dadurch können vor allem die Unterschiede klar herausgearbeitet werden. Zunächst ist aber festzuhalten, dass alle o.g. Applikationen treffend als „purpose-built geospatial RDF data generation tools“ bezeichnet werden können, da sie speziell für die Generierung von LGD entwickelt wurden (Nishanbaev et al., 2020). Allen o.g. RDF-izer ist also gemeinsam, dass sie die Erzeugung von RDF-Graphen aus Geoinformationen ermöglichen, die in RDBMS/ORDBMS mit räumlicher Erweiterung (z.B. PostgreSQL i.V.m. PostGIS) oder in dateibasierten Formaten (z.B. ESRI Shapefiles) gespeichert bzw. verwaltet werden (Nishanbaev et al., 2020)⁵⁴.

Die Java-Bibliothek namens **Geometry2RDF**, war das erste Werkzeug, das zu diesem Zweck entwickelt wurde und zwar bevor GeoSPARQL vom OGC standardisiert wurde (Vilches-Blázquez et al., 2010; Kyzirakos et al., 2014). Für die Verarbeitung respektive Transformation von räumlichen Informationen (GML oder WKT) wurde das Java-Framework namens Apache Jena⁵⁵, die Java-Bibliothek GeoTools⁵⁶, die Java Topology Suite (JTS)⁵⁷, die Geospatial Data Abstraction Library (GDAL) / OGR Simple Features Library (OGR)⁵⁸ sowie das Vokabular Basic Geo (WGS84 lat/long)⁵⁹ eingesetzt (Vilches-Blázquez et al., 2010; Patroumpas et al., 2014).

Die folgende Graphik zeigt die o.g. Komponenten und den grundsätzlichen Ablauf zur Erstellung eines RDF-Graphen aus Geometriedaten, die serialisiert als Zeichenkette vorliegen:

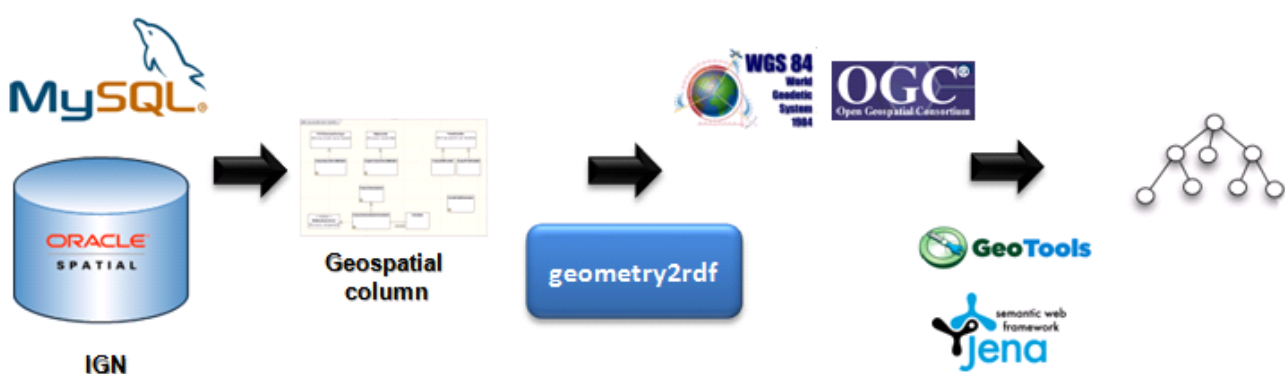


Abb. 3.7: Umformung von Geometriedaten mit Geometry2RDF, OA unverändert
 übernommen von: <https://oeg.fi.upm.es/index.php/en/technologies/151-geometry2rdf/index.html>

⁵⁴ Wie bereits erwähnt, haben Rasterdaten für die vorliegende Arbeit keine Relevanz.

⁵⁵ <https://jena.apache.org/>

⁵⁶ <https://www.geotools.org/>

⁵⁷ <https://www.osgeo.org/projects/jts/>

⁵⁸ <https://www.osgeo.org/projects/gdal/>

⁵⁹ <https://www.w3.org/2003/01/geo/>

Die Transformation der Geometriedaten wird über eine Konfigurationsdatei definiert (ggf. Koordinatentransformation, URI-Konstruktion und Verknüpfung). In diesem Zuge können neben dem Basic Geo Vokabular weitere Ontologien festgelegt werden, sodass auch RDF-Repräsentationen von Linienzügen und Polygonen generiert werden können. Der Prozess wird über die Kommandozeile (engl. command-line interface, kurz: CLI) angestoßen. Standardmäßig enthält der resultierende RDF-Graph aber keine Konzepte aus dem GeoSPARQL-Vokabular oder die angerissenen raumzeitlichen Erweiterungen stRDF/stSPARQL (Kyzirakos et al., 2014). Der Graph kann in Form einer RDF/XML-Serialisierung ausgegeben werden (Vilches-Blázquez et al., 2010).

Deshalb entwickelten Patroumpas et al. (2014) aufbauend auf Geometry2RDF (und damit basierend auf allen o.g. Java-Bibliotheken) die Java-Applikation **TripleGeo**, welche auf die Generierung von LGD (primär) unter Nachnutzung der Konzepte aus dem GeoSPARQL-Vokabular ausgerichtet ist. TripleGeo wurde allerdings auch für die Transformation von nicht-räumlichen Attributen konzipiert und nicht nur für die Umformung der Geometriedaten (Patroumpas et al., 2014). Es können vollständige RDF-Beschreibungen von Features generiert und in vielen gängigen Notationen (z.B. Turtle) serialisiert werden (Patroumpas et al., 2014). Abgesehen von diesen funktionalen Erweiterungen, ist der Umformungsprozess jedoch mit dem oben skizzierten Ablauf vergleichbar:

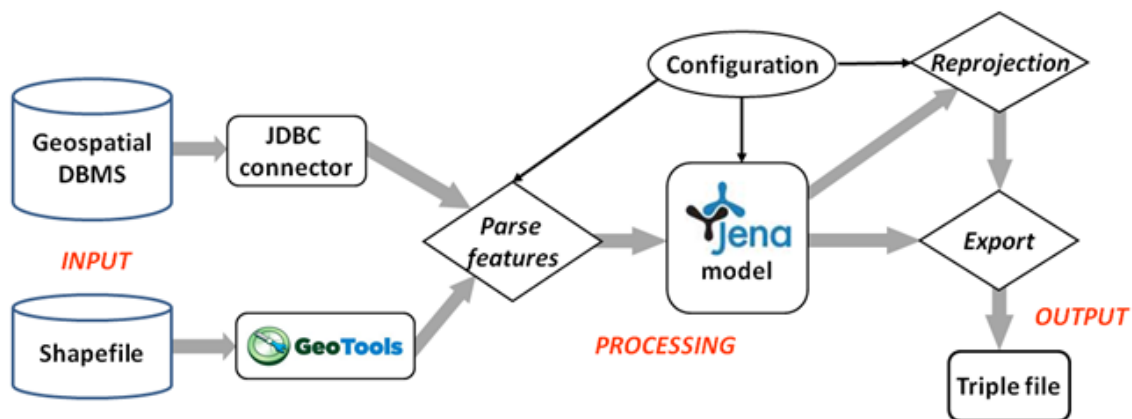


Abb. 3.8: Extraktion, Transformation, Laden mit TripleGeo, OA von Patroumpas et. al. (2014, p. 275)

Die Parameter für die Umformung werden in einer Konfigurationsdatei festgelegt (URI-Allokation, CRS etc.). Der Prozess wird über das CLI initialisiert. Patroumpas et al. (2014, p. 275) bezeichneten TripleGeo als „spatial ETL tool“⁶⁰ und betonten, dass statt GeoSPARQL auch andere Vokabulare verwendet werden können. Allerdings muss festgehalten werden, dass dann viele der „hartkodierten“ Vorgaben manuell angepasst werden müssen (z.B. in den vorhandenen XSLT-Dokumenten). Der Transformationsprozess basiert auf statischen Regeln; TripleGeo ist in mehrfacher Hinsicht nicht wirklich mit umfassenderen Spatial-ETL-Werkzeugen wie z.B. der Feature Manipulation Engine (FME) vergleichbar, da diese zwar einerseits „absolutely no RDF support“ bieten (Patroumpas et al., 2014, p. 276), andererseits aber ein flexibleres Datenmapping über eine GUI ermöglichen.

⁶⁰ Die Abkürzung „ETL“ steht *allgemein* für **E**xtraktion, **T**ransformation, **L**aden – also für drei grundlegende Prozess-Schritte zwecks Informationsintegration. Im *konkreten Fall* beschränkt sich die Extraktion zwar auf das Auslesen von räumlichen und thematischen Informationen aus vektorialen Geodatenmodellen oder -formaten; der Transformationsprozess umfasst jedoch neben der syntaktischen Umformung (etwa die reine Konvertierung bzw. Serialisierung von Geometriedaten) auch die semantische Transformation der Sachdaten (z.B. Bereinigung, Aggregation) um diese letztlich in einem RDF-Graph zusammenzuführen, indem die vorhandenen Kanten (RDF-Tripel) entweder direkt über eine entsprechende Schnittstelle in die Datenbank importiert oder zunächst als Sammlung linearer Zeichenketten mithilfe gängiger RDF-Serialisierungen exportiert werden (s.o.: „Triple file“). Dieser Workflow wird im weiteren Verlauf als „ETL-Prozess“ bezeichnet.

Flexibilität und Dynamik waren wesentliche Faktoren für die Entwicklung des RDF-izers namens **GeoTriples** (Kyzirakos et al., 2014). Ebenso wie TripleGeo bietet auch GeoTriples die Möglichkeit aus Geodaten (nachfolgend „Quelldaten“ genannt), welche in gängigen Geodatenformaten (z.B. dBase- oder textbasiert) vorliegen oder mittels Geodatenbanken (etwa PostgreSQL/PostGIS) verwaltet werden, einen RDF-Graphen unter Verwendung des GeoSPARQL-Vokabulars zu generieren (Kyzirakos et al., 2014). Allerdings stellt GeoTriples hierfür eine flexible Komponente („the mapping generator“) zur Verfügung, die via GUI oder CLI angesprochen werden kann, um basierend auf den Quelldaten automatisch Anweisungen für das Datenmapping (kurz: Mapping) zu generieren (Kyzirakos et al., 2014, p. 6). Laut Kyzirakos et al. (2014) eignet sich der Mapping-Generator sowohl für das direkte Mapping⁶¹ – ein Prozess, der vom W3C standardisiert wurde (Arenas et al., 2012) – als auch für das individuelle Mapping mithilfe der RDB to RDF Mapping Language (R2RML), die ebenfalls als W3C-Empfehlung vorliegt (Das et al., 2012). In letzterem Fall kann auch die RDF Mapping Language (RML) verwendet werden und statt GeoSPARQL kann z.B. auch stRDF/stSPARQL, CIDOC-CRM/CRMgeo speziell für die Generierung von LGCHD eingesetzt werden (Kyzirakos et al., 2014; Nishanbaev et al., 2020). Dies bedingt aber die manuelle Anpassung der automatisch generierten Mapping-Anweisungen, was Kyzirakos et al. (2014) als semi-automatischen Prozess bezeichneten. Aus den (ggf. adaptierten) Anweisungen erzeugt der R2RML- / RML-Prozessor den RDF-Graph (Kyzirakos et al., 2014). Im Gegensatz zu TripleGeo stellt GeoTriples also keine vordefinierten Transformationsanweisungen bereit, sondern erzeugt diese dynamisch. Dadurch ist auch die Persistierung per se flexibler, da die Quelldaten nicht notwendigerweise mithilfe eines Tripel- / Quadstores gespeichert werden müssen, sondern die R2RML- / RML-Mappings können auch verwendet werden, um GeoSPARQL-Anfragen zur Laufzeit in SQL-Statements zu übersetzen, wenn das RDBMS/ORDBMS mit der räumlichen Erweiterung noch mit einem entsprechenden Aufsatz ausgestattet wird (Kyzirakos et al., 2014). Die folgende Graphik stellt die angerissenen Komponenten und Funktionen von GeoTriples überblicksmäßig dar; für Details zum verwendeten RDB2RDF-Framework wird auf die Arbeit von Kyzirakos et al. (2014) verwiesen.

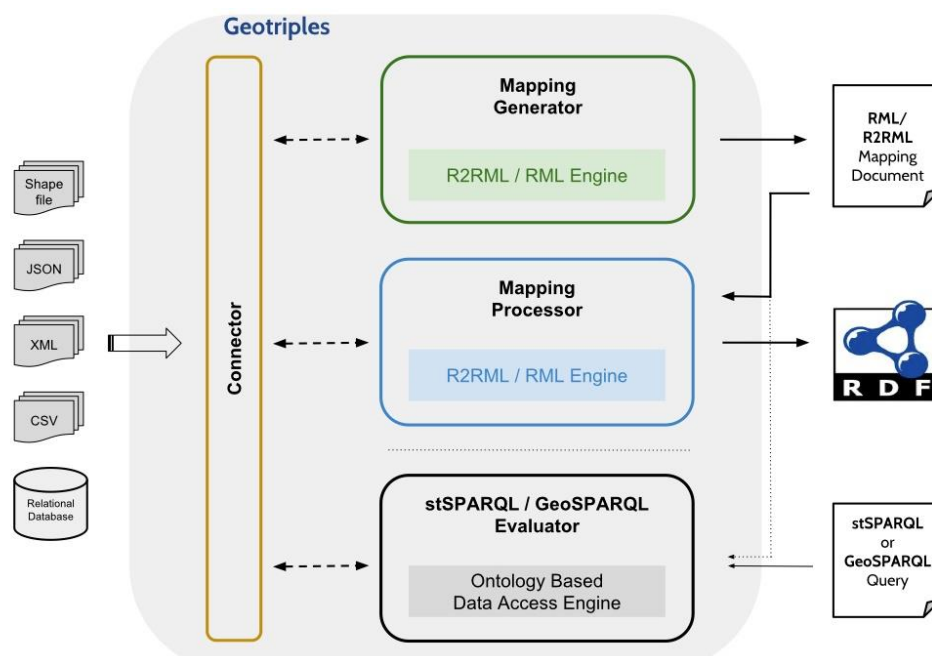


Abb. 3.9: Architektur von GeoTriples, OA unverändert
übernommen von: <http://geotriples.di.uoa.gr/>

⁶¹ In diesem Fall werden die Vorgaben für die Generierung der RDF-Repräsentationen direkt vom relationalen Datenbankschema abgeleitet.

Neben den o.g. RDF-izern, die – wie bereits erwähnt – speziell auf die Generierung von LGD/LGCHD unter Verwendung des RDF-Datenmodells ausgerichtet sind, könnten zur Erreichung des zweiten operativen Teilziels der vorliegenden Arbeit auch **Spatial-ETL-Werkzeuge** eingesetzt werden, die Patroumpas et al. (2014) erwähnten: die Applikation FME des kanadischen Unternehmens Safe Software⁶² und die freie Software GeoKettle⁶³. Diese Tools sind zwar nicht dem o.g. Zweck dediziert (d.h. es fehlt an „nativer RDF-Unterstützung“); sie können aber (ggf. unter Verwendung geeigneter Skript- / Programmiersprachen wie Java, Python, XSLT etc.) aus den o.g. Quelldaten ebenfalls RDF-Repräsentationen erzeugen. Vorteilhaft ist dabei, dass die semantische Transformation der Quelldaten (Überprüfung, Bereinigung etc.) dank vordefinierter Funktionen effizient umgesetzt werden kann. Der ETL-Prozess kann auf die Charakteristika der jeweiligen Quelldaten (z.B. uralte dBASE-Datenstrukturen) und alle weiteren individuellen Anforderungen, etwa hinsichtlich der Zieldatenstruktur (Serialisierung, Vokabulare etc.) zugeschnitten werden. Aus diesem Grund wurde FME für die RDF-Generierung bereits in einem Experiment des OGC verwendet (Hobona and Brackin, 2015). Es ist auch möglich Plugins, Bibliotheken usw. – z.B. Geometry2RDF – in Kombination mit den o.g. ETL-Tools zu verwenden⁶⁴.

Auch **skriptbasierte Lösungen** zwecks syntaktischer Transformation könnten in der vorliegenden Arbeit aufgegriffen werden. van den Brink et al. (2014) realisierten u. a. die automatische Transformation wohlgeformter GML-Instanzen in RDF-Repräsentationen mittels XSLT. Wegen vergleichbarer Modellierungsansätze (Objekt-Property-Muster, gerichteter Graph) kann die Umformung syntaktisch korrekter GML-Instanzen in RDF-Repräsentationen mit der o.g. Programmiersprache recht einfach bewerkstelligt werden (van den Brink et al., 2014). Die grundlegenden Ausführungen von Schade and Cox (2010) hierzu wurden unter GP 2.3.3 zusammengefasst. van den Brink et al. (2014) bezeichneten ihr Skript als RDF-izer und bezogen sich dabei auf die Abbildung von López-Pellicer et al. (2012) – s. Abb. 3.1. van den Brink et al. (2014) fokussierten jedoch nicht nur die Instanzebene, sondern präsentierten auch eine Methode zwecks semi-automatischer Generierung eines Semantic Web Dokuments (TBox in der OWL) auf Basis eines Modells, das mithilfe der UML erstellt wurde.

⁶² Webpräsenz des Unternehmens: <https://www.safe.com/>

⁶³ Dokumentation: https://live.osgeo.org/archive/10.5/de/overview/geokettle_overview.html

⁶⁴ Beispiele für GeoKettle: <https://github.com/marcelocaj/rdf4geokettle>

3.2.2 WFS-Adapter

Jones et al. (2014, p. 5) entwickelten die Applikation namens „LOD4WFS“ um LGD via WFS für die Nutzung in GIS bereitzustellen. Der o.g. WFS-Adapter dient dem Zweck, der unter GP 2.4.1 beschrieben wurde.

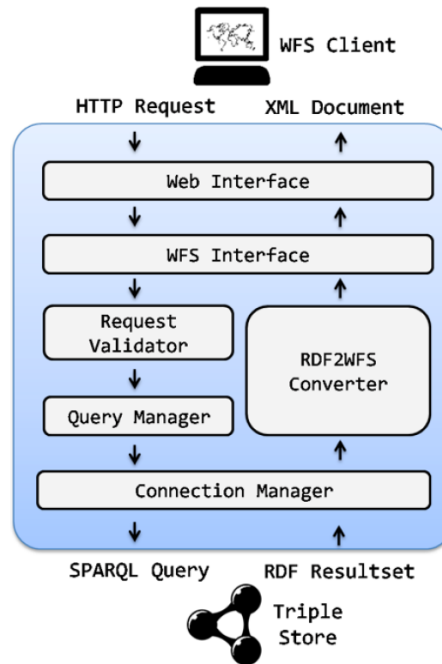


Abb. 3.10: Funktionsweise des modularen LOD4WFS-Adapters, OA von Jones et al. (2014, p. 9)

Die obige Darstellung zeigt nicht nur die Funktionsweise des WFS-Adapters, sondern auch den modularen Aufbau der Applikation (Jones et al., 2014). LOD4WFS wurde in Java implementiert (Jones et al., 2014). Es handelt sich um eine Basic WFS-Implementierung, die laut Jones et al. (2014) der WFS-Implementierungsspezifikation des OGC in der Version 1.0.0 entspricht. LOD4WFS kann GetCapabilities-, DescribeFeatureType- und GetFeature-Requests beantworten (Jones et al., 2014). Die Operationen wurden unter GP 2.3.2 beziehungsweise auf den vorgenannten OGC-Standard ausführlich besprochen. Dabei wurde auf eine Widersprüchlichkeit bzgl. FE/CQL hingewiesen. Jones et al. (2014) verschwiegen nicht, dass die CQL-Unterstützung im Zuge der Weiterentwicklung realisiert werden sollte. In der vorliegenden Arbeit wird die fehlende CQL-Unterstützung nicht als Hinderungsgrund für den Einsatz von LOD4WFS angesehen, da Jones et al. (2014, p. 8) unterschiedliche Zugriffsmethoden implementierten, die sie als „Standard Data Access“ (SDA) und als „Federated Data Access“ (FDA) bezeichneten. FDA erlaubt die individuelle Formulierung von SPARQL-Statements, die gespeichert werden können (Jones et al., 2014). Damit lassen sich sinnvolle Feature Collections generieren; z.B. kann ein FILTER-Element in einer SELECT-Abfrage verwendet werden, um bestimmte Instanzen zu selektieren, für die ein bestimmter Zeitraum angegeben ist usw. Für die Nutzung in GIS lassen sich also geeignete „Layer“ zusammenstellen. Jones et al. (2014) demonstrierten, dass via FDA auch Anfragen an mehrere SPARQL-Endpunkte gestellt werden können (Schlüsselwort SERVICE). Dadurch lassen sich Informationen aus unterschiedlichen Quellen kombinieren und per WFS bereitstellen (Jones et al., 2014). Dies kann mithilfe von LOD4WFS einfach nachvollzogen werden, da Jones et al. (2014) einige SPARQL-Statements zur Veranschaulichung beifügten, die allerdings nicht die Schlüsselwörter DESCRIBE oder CONSTRUCT enthalten. FDA und SDA setzen auf SPARQL-SELECT-Anfragen. Die Antwort eines RDF-Stores auf eine solche Anfrage erfolgt in tabellarischer Form; die spezifizierten Variablen werden entsprechend belegt. Auf dieser Grundlage generiert der o.g. WFS-Adapter (je nach Anfrage) entweder ein XML-Dokument oder ein XSD oder eine GML-Instanz.

Da SDA in der vorliegenden Arbeit eine untergeordnete Rolle einnimmt, wird für weitere Details dazu auf die Arbeit von Jones et al. (2014) verwiesen. Hinsichtlich FDA soll jedoch erwähnt werden, dass Jones et al. (2014) auch eine GUI entwickelten, weshalb die o.g. SPARQL-Statements einfach formuliert und getestet bzw. validiert werden können – insbesondere, weil auch ein HTTP-Server (Jetty) mitgeliefert wird. Abgesehen davon zeichnet sich LOD4WFS durch eine ausführliche Dokumentation⁶⁵ und durch hohe Performanz aus.

Mit LOD4WFS steht eine „ready-to-use application“ zur Verfügung, die in der vorliegenden Arbeit nachge-nutzt werden könnte. Das Ziel dieser Thesis besteht darin, aus isolierten Geodaten (ohne Links, ohne Semantik) einen RDF-Graphen zu formen, der den LOD-Anforderungen entspricht, wobei das vierte LOD-Prinzip (Ver-linkung) prioritär berücksichtigt wird. Das Resultat soll einen Mehrwert für potenzielle Anwenderinnen bieten. Aufgrund der Parallelen zwischen RDF und GML, könnten die verknüpften bzw. verlinkten Informationen mithilfe von LOD4WFS in verknüpfte GML-Strukturen mit Querverbindungen (Links) abgebildet werden. Anhand eines konkreten Anwendungsfalls könnte aufgezeigt werden, mit welchen SPARQL-Statements dieses Ergebnis erreicht werden kann und es könnte demonstriert werden, dass LGD auch für GIS-Anwenderinnen einen Mehrwert bietet. Deshalb soll der WFS-Adapter beim Entwurf des Systems (Logikschicht) berücksichtigt werden.

3.2.3 Frameworks

A framework for connecting two interoperability universes: WFS and LD

Ehrlicherweise muss gesagt werden, dass dies keine neue Idee ist. Vilches-Blázquez and Saavedra (2019, p. 22) verwendeten LOD4WFS in ihrem „framework for generating, enriching, and exploiting geospatial Linked Data“, wofür der Quellcode von LOD4WFS nicht angepasst werden musste⁶⁶. Die folgende Abbildung zeigt, dass LOD4WFS von Vilches-Blázquez and Saavedra (2019, p. 40) zwecks „Exploitation“ eingesetzt wurde.

Vilches-Blázquez and Saavedra (2019, p. 40) bezweckten damit Folgendes: „The final goal of connecting and opening these data is that users can use semantics tools and SDI technologies in a coordinated way to search, analyze, visualize, or absorb programmatically the whole data available“. Dafür entwickelten Vilches-Blázquez and Saavedra (2019, p. 31) eine Web-Applikation („GeoSPARQL Viewer“) mithilfe von Leaflet⁶⁷ und Apache Jena⁶⁸. Dieser Client soll beim Entwurf des Systems (Präsentationsschicht) auch berücksichtigt werden, da potenziellen Anwenderinnen genau die o.g. Möglichkeiten eröffnet werden sollen. Mit der Web-Applikation von Vilches-Blázquez and Saavedra (2019 p. 41) lassen sich GeoSPARQL-Abfragen stellen, welche die Generierung von dynamischen Karten ermöglichen, „since each interaction with data (query) displays an updated map related to the obtained results of each query“. Aufbauend auf LOD4WFS konnten Vilches-Blázquez and Saavedra (2019, p. 41) Folgendes erreichen: „an enhancement in the query processing, since [the] framework allows setting GeoSPARQL queries on an RDF triple store and showing an updating of the results on WFS automatically“.

⁶⁵ <https://github.com/jimjonesbr/lod4wfs>

⁶⁶ Version vom 28.06.2018, 00:42, die am 05.12.2020, 20:20, von <https://github.com/jasaavedra/GeoLOD> heruntergeladen wurde. LOD4WFS hat denselben Funktionsumfang.

⁶⁷ <https://leafletjs.com/>

⁶⁸ <https://jena.apache.org/>

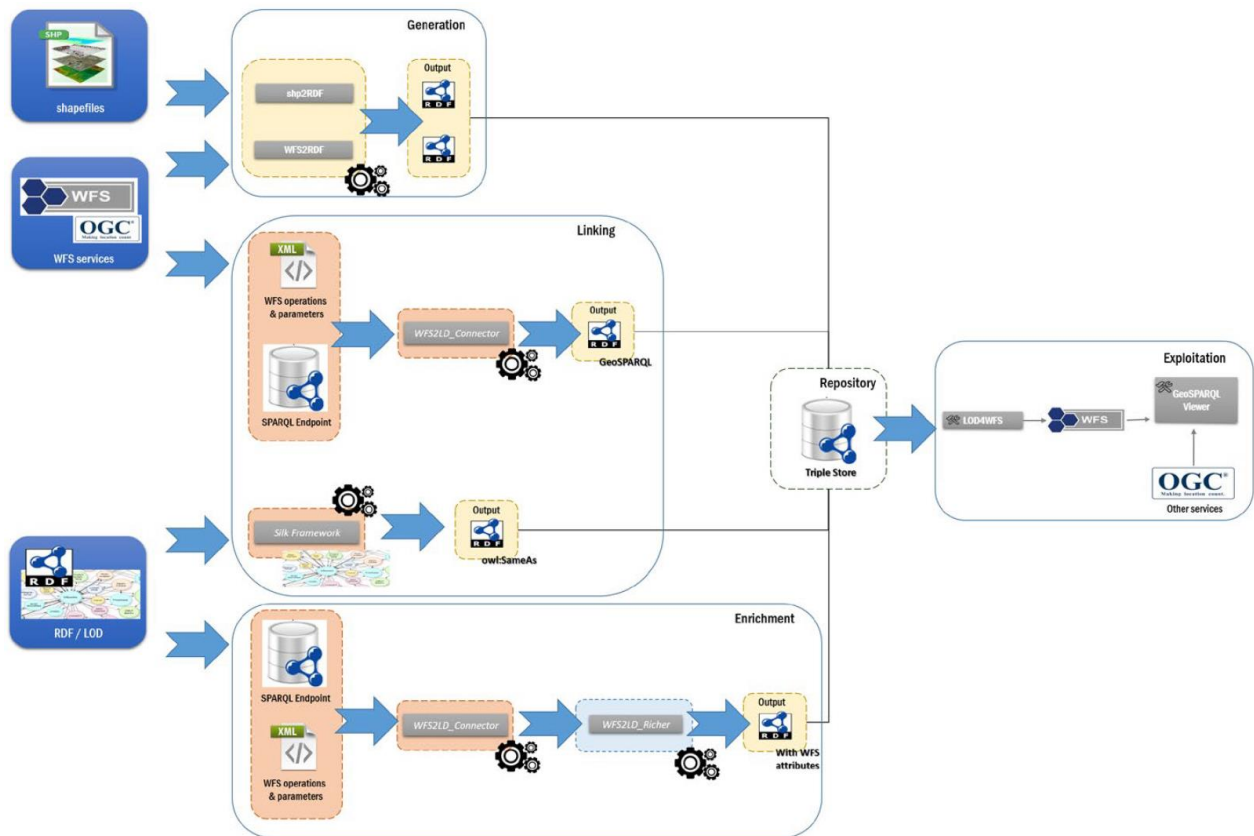


Abb. 3.11: Generierung und Publikation von LGD, OA von Vilches-Blázquez and Saavedra (2019, p. 22)

Silk Link Discovery Framework

Die obige Darstellung zeigt, dass Vilches-Blázquez and Saavedra (2019) auch „Silk“ für die Verlinkung einsetzen. Das „Silk Link Discovery Framework“ (Isele 2013, p. 173), das auch The Linked Data Integration Framework genannt wird, ist ein Open Source Framework, welches unabhängig vom oben gezeigten Framework ist⁶⁹.

Silk wurde von Wissenschaftlerinnen entwickelt, um die Verlinkung in der LOD-Cloud voranzutreiben (Volz et al., 2009). Die Wissenschaftlerinnen erkannten, dass Datenbereiterinnen auf einige Anwendungen für die Publikation von Linked Data zurückgreifen konnten, aber dass die Verlinkung mit anderen Ressourcen noch nicht hinreichend unterstützt wurde (Volz et al., 2009).

Mithilfe von Silk können Datenbereiterinnen ihre Ressourcen über RDF-Links mit anderen Ressourcen in der LOD-Cloud verknüpfen (Volz et al., 2009). In der vorliegenden Arbeit kann Silk für die Generierung von RDF-Links auf der Instanzebene eingesetzt werden. Dadurch soll die Menge der verfügbaren Information zu einem Individuum angemessen erhöht werden⁷⁰. Das zentrale Problem „ist das Finden von Paaren von Entitäten in Datensets, welche das gleiche Realwelt-Objekt beschreiben“ (Isele 2013, p. II). Das Problem kann mithilfe von Silk gelöst werden, indem zweckmäßige „link specifications“ erstellt werden (Isele 2013, p. 174).

⁶⁹ <http://silkframework.org/>

⁷⁰ Der „angemessene Umfang“ wird unter GP 6.4 behandelt. Die Verlinkung wurde anschaulich in Abschnitt 2.5 erklärt.

Isele (2013, p. 174) erklärte, dass eine „link specification“ sämtliche Informationen für die Verlinkung enthält (z.B. den Typ des zu generierenden Links, die zu erfüllenden Bedingungen bzw. Verlinkungsregeln, Datenzugriffsinformationen, Ausgabeoptionen). Sie kann mithilfe der deklarativen „Silk Link Specification Language“ dargestellt werden (Isele 2013, p. 174). Die folgenden Punkte sind für die vorliegende Arbeit besonders relevant:

- „Data Sources“, „Outputs“: Silk kann Requests an einen (oder mehrere) SPARQL-Endpunkt(e) stellen, um auf die (benannten) Graphen bzw. die Informationen zuzugreifen, die in einem RDF-Store verwaltet werden (Isele 2013, p. 175). Außerdem kann Silk die RDF-Links via SPARQL (Update) in den RDF-Store bzw. in den entsprechenden Graphen einfügen (Isele, 2013).
- „Link Type“: Silk kann Identity Links generieren, aber statt `owl:sameAs` kann auch ein anderes Property eingesetzt werden (Isele 2013, p. 176).
- „Data Sets“: Isele (2013, p. 176) erklärte, dass Silk die Formulierung von SPARQL-Abfragen erlaubt, damit aus einer Menge von Ressourcen eine Untermenge für die Verlinkung selektiert werden kann (z.B. alle Instanzen eines bestimmten Feature Types).
- „Linkage Tasks“, „Linkage Rule“: Darüber wird festgelegt wie bestimmte Typen von Entitäten, die in verschiedenen Datensätzen beschrieben werden, verlinkt werden sollen (Isele 2013, p. 176). Laut Isele (2013) können Bedingungen für die Verlinkung definiert werden. Volz et al. (2009, p. 3) bezeichneten diese Bedingungen als zentralen Teil einer „link specification“, weil hier spezifiziert wird „how similarity metrics are combined in order to calculate a total similarity value for an entity pair.“ Für Details zu den verfügbaren Operatoren und zu den erlaubten Kombinationen wird auf Abschnitt 2.5 in der Dissertation von Isele (2013) verwiesen.

Eine „link specification“ kann über die „Silk Link Discovery Engine“ ausgeführt werden (Isele 2013, p. 177); sie ist in allen Silk-Applikationen („Silk Single Machine“, „Silk MapReduce“, „Silk Server“) enthalten (Isele 2013, p. 173). Für die vorliegende Arbeit soll der Fokus auf die anwenderinnenfreundliche Web-Applikation⁷¹ namens „Silk Workbench“ gelenkt werden, die aus verschiedenen Komponenten besteht (Isele 2013, p. 177):

- Der „Workspace Browser“ über den die Graphen (Quelle und Ziel), die Präfixe, die Transformations- und Verlinkungsprozesse etc. verwaltet werden können (Isele 2013, p. 177).
- Der „Linkage Rule Editor“ für die Spezifikation der Bedingungen/Regeln (Isele 2013, p. 177). Isele (2013) hob hervor, dass sich der graphische Editor auch zum Experimentieren eignet. Die Operatoren, deren Parameter, die Kombinationen etc. können einfach variiert, getestet und verglichen werden (Versuch und Irrtum).
- Eine Komponente („Execution“), mit der die spezifizierten Verlinkungsregeln ausgeführt und die generierten Links untersucht werden können (Isele 2013, p. 177).
- Der „Reference Links Manager“, mit dem die Referenz-Links verwaltet werden (Isele 2013, p. 177).
- Der „Linkage Rule Learner“ über den Silk Verlinkungsregeln lernen kann (Isele 2013, p. 50).

⁷¹ Diese Anwendung bzw. das GUI zeichnet sich nicht nur durch ergonomische Gestaltung aus. Fehler werden ebenfalls vermieden, da die Web-Applikation sinnvolle Hinweise ausgibt; z.B. wenn ein Prozess noch nicht abgeschlossen ist. Außerdem ist sie zuverlässig. Gespeicherte Konfigurationen können problemlos wiederverwendet werden. Da es sich hier um eine Open Source Web-Anwendung handelt, ist dies bemerkenswert – vor allem, weil hierauf noch einmal Bezug genommen werden muss.

Die Silk-Workbench unterstützt die Generierung einer Link-Spezifikation. Man könnte auch sagen, dass das zeitaufwendige Suchverfahren vereinfacht wird. Die Informationskosten⁷² können reduziert werden. Dafür wurde der „Linkage Rule Learner“ entwickelt, der auf Algorithmen basiert, welche eine Menge von Referenz-Links nutzen, um Verlinkungsregeln zu lernen (Isele 2013, p. 50). „A set of reference links consist of positive reference links that identify pairs of entities that are known to match and negative reference links that identify pairs that that are known non-matches“ (Isele 2013, p. 5). Wenn es keine Referenz-Links gibt, können diese von einer Expertin erstellt werden, damit Silk die geeignetste Verlinkungsregel vorschlagen kann (Isele 2013).

Da dies in der vorliegenden Arbeit eine Rolle spielen wird, soll anhand des Beispiels von Isele (2013) illustriert werden, wie die Referenz-Links im Rahmen des Evaluierungsprozesses über das GUI erzeugt werden können:

Source: DBpedia	Target: linkedmdb	Score	Correct?
▼ Topaz_%281969_film%29	mdb.org/resource/film/230	-4.1%	☑ ? ✖
<pre> http://dbpedia.org/resource/Topaz_%281969_film%29 ?a<http://xmlns.com/foaf/0.1/name> Topaz ?a<http://dbpedia.org/ontology/releaseDate> 1969-12-19 ?a<http://dbpedia.org/property/name> Topaz http://data.linkedmdb.org/resource/film/230 ?b<http://purl.org/dc/terms/title> Topaz ?b<http://data.linkedmdb.org/resource/movie/initial_release_date> 1945 ?b<http://www.w3.org/2000/01/rdf-schema#label> Topaz </pre>			
▶ erbolt_%281910_film%29	ndb.org/resource/film/350	-14.6%	☑ ? ✖
▶ from_Brazil_%28film%29	ndb.org/resource/film/353	14.6%	☑ ? ✖
▶ kness_%282002_film%29	db.org/resource/film/2320	14.6%	☑ ? ✖
▶ a:Castaway_%28film%29	db.org/resource/film/2051	-15.8%	☑ ? ✖

Abb. 3.12: Beispiel zur Evaluation in der Silk Workbench, OA von Isele (2013, p. 179)

Isele (2013, p. 179) bezeichnete diesen Prozess als „active learning workflow“, weil in diesem Fall der „ActiveGenLink algorithm“ verwendet wird. Die Detailansicht zeigt, dass die farbig markierten URIs keine URI-Aliase sind. Hier sollen Identity Links generiert werden. Die URIs identifizieren nicht denselben Film (Isele, 2013). Daher fügt die Anwenderin einen negativen Referenz-Link hinzu (Isele, 2013). Glücklicherweise stehen ihr neben den „Namen“ der Filme, die ident sind, auch Daten zur Verfügung, die ihr die Entscheidung leicht machen. Anschließend wird sie von Silk bei der Generierung der Identity Links unterstützt. Da sie einen negativen Referenz-Link hinzugefügt hat, reagiert die intelligente Applikation entsprechend, indem sie eine Verlinkungsregel vorschlägt, die nicht ausschließlich auf einen „Namensvergleich“ setzt (Isele, 2013). Mithilfe der folgenden Darstellung soll abschließend pointiert werden, dass der Lernvorgang i.d.R. ein iterativer Prozess ist:

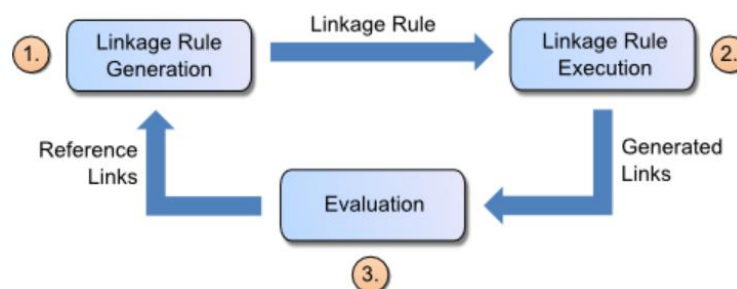


Abb. 3.13: Der typische Workflow für die Erstellung einer Link-Spezifikation, OA unverändert übernommen von: https://app.assembla.com/spaces/silk/wiki/Silk_Workbench

⁷² Es ist der Aufwand gemeint, der z.B. im Zuge der Kenntniserneuerung (Stichwort: Operatoren), bei der Selektion etc. entsteht.

4 Kurzportrait: Historisches Häuserbuch

Da die Entwicklungsgeschichte des historischen Häuserbuchs der Stadt Mainz bereits einleitend skizziert wurde, widmet sich dieser Abschnitt nun der Veranschaulichung des Diskursuniversums, wofür im IT-Bereich häufiger der englische Ausdruck Universe of Discourse (UoD) verwendet wird. Diese historische „Mini-Welt“ wird anhand geschichtlicher **Hintergrundinformationen und Kartengrundlagen** dargestellt. Für die Visualisierung der Modelle wird auf die **Rasterdarstellungen** zurückgegriffen, die den Anwenderinnen momentan über den WMS zur Verfügung stehen; man kann insofern vom externen Modell sprechen. Anhand ausgewählter Geoobjekte werden wichtige **Modelleigenschaften** (z.B. die Klassifikation der Gebäude) hervorgehoben.

Die Strukturen des zugrunde liegenden Vektormodells bleiben den Nutzerinnen der Web-Map-Applikation verborgen. Die Datenhaltungsschicht wird ausschließlich von Applikationen auf der Logikschicht angesteuert. Die Anwenderin kann die gerenderten Rasterdaten über eine graphische Benutzeroberfläche (GUI) nutzen. Ausgehend von dieser Perspektive können die Modelle zunächst grob dargestellt und charakterisiert werden. Dieses Kurzportrait schließt mit einem **Verwendungsbeispiel aus der Praxis**. Es wird aufgezeigt, wieso der WFS-gestützte Zugriff auf die Vektormodelle sinnvoll ist. Diese werden in Kapitel 5 im Detail analysiert und exakt dokumentiert, da diese Datenbasis zu einem LOD-konformen RDF-Graph transformiert werden muss.

4.1 Hintergrundinformationen und Kartengrundlagen

Die wichtigste Ausgangsbasis für die Erstellung des historischen Häuserbuchs bildete der von Falck (1972) erstellte Plan „Mainz um 1620“. Der Plan enthält neben einer kartographischen Darstellung des Stadtgebiets auch Verzeichnisse, die z.B. Auskunft zu den bebauten Grundstücken und den Namen der Gebäude geben.

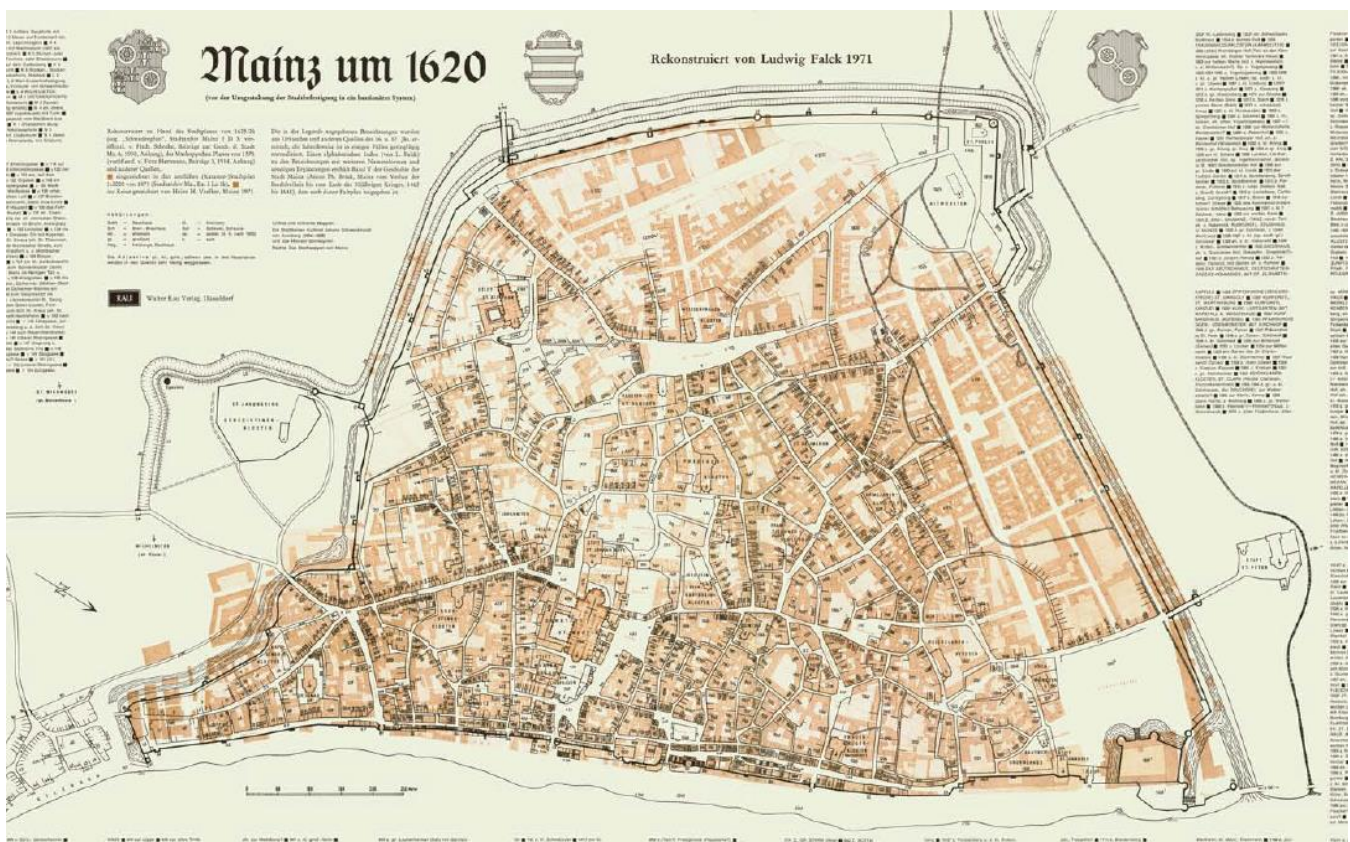


Abb. 4.1: Mainz um 1620, erstellt von Falck (1972)

Mit den ältesten Mainzer Stadtplänen standen geeignete Quellen für die Rekonstruktion des Stadtgebiets im frühen 17. Jahrhundert zur Verfügung (Falck, 1978). Eine dieser Kartengrundlagen entstand in den Jahren 1625/26 und wird gelegentlich als „Waldenburgplan“ bezeichnet, da man den damaligen kurfürstlichen Stadtkommandanten als Auftraggeber vermutet (Falck, 1978). Sicher ist allerdings, dass dieser Plan während des Dreißigjährigen Krieges von den Schweden unter Führung von König Gustav Adolf konfisziert und im schwedischen Kriegsarchiv aufbewahrt wurde, weshalb er häufig „Schwedenplan“ genannt wird (Falck, 1978).



Abb. 4.2: „Schwedenplan“ (Sta Mainz, BPSP / 55 D)

Der „Schwedenplan“ zeigt für die Jahre 1625/26 bereits die geplante bastionäre Umwallung als Konsequenz renaissancezeitlicher Artilleriekriegsführung und die neue Befestigungsanlage um den Jakobsberg – die Schweickhardsburg – während Falck (1972) in seinem Plan (Abb. 4.1) noch den spätmittelalterlichen Mauer ring mit seinen Tor- und Wehranlagen im Stichjahr 1620 fokussierte. Die Informationen für die Rekonstruktion entnahm Falck (1978) auch dem sog. „Vogelschauplan“ von Gottfried Maskopp aus dem Jahr 1575. Diesen kombinierte er mit weiteren Dokumenten und Ausgrabungsskizzen sowie mit dem ältesten maßstabrichtigen Mainzer Stadtplan aus dem Jahr 1871, welchen er im Mainzer Stadtarchiv finden konnte (Falck, 1978).



Abb. 4.3: Auszug aus dem Stadtplan von 1871 (Sta Mainz, BPSP / 489 D)

4.2 Rasterdarstellungen und Modelleigenschaften

Der „Falck-Plan“ wurde zunächst eingescannt und anschließend mithilfe des Desktop-GIS „MapInfo Professional“ georeferenziert. In Kombination mit weiteren Quellen, vor allem den relevanten Urkunden des Mainzer Stadtgerichts zu Grundstücksgeschäften und einschlägiger Literatur, wurden zahlreiche MapInfo-Relationen erzeugt. Diese repräsentieren, als separate Vektorlayer, die verschiedenen topographischen Objekte für die Zeit um 1450. Die folgende Rasterdarstellung wurde auf dieser Basis gerendert.

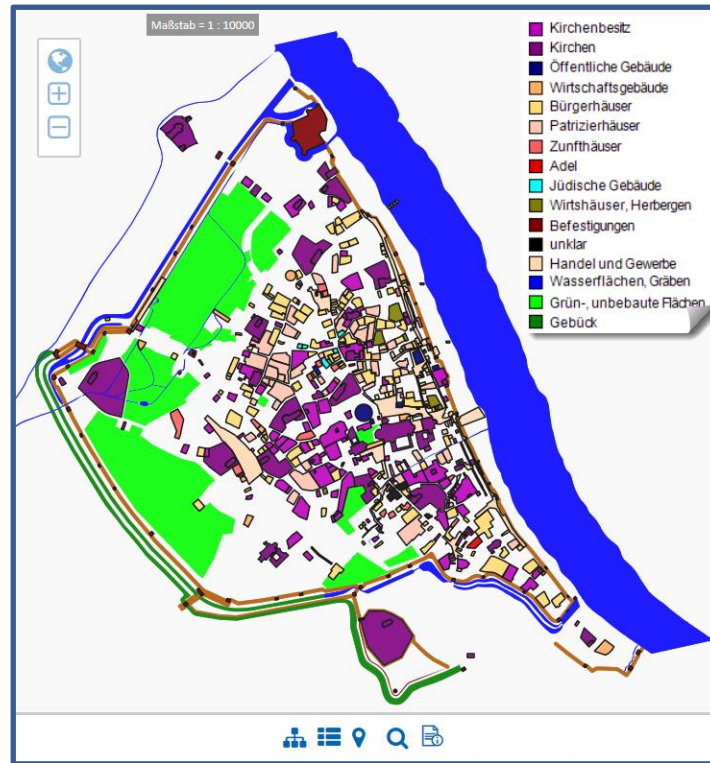


Abb. 4.4: Mainz um 1450 (gerendert von MapServer V. 7, dargestellt von Mapbender V. 3)

Nicht alle Informationen beziehen sich exakt auf das Jahr 1450, weshalb bewusst unscharf von Mainz „um 1450“ gesprochen wird. Ziel dieses Modells ist es, die mittelalterliche Stadtentwicklung der freien Stadt Mainz darzustellen. Falck (1974) erklärte die Wahl des Jahres damit, dass diese Phase um 1450 noch andauerte und erst am 28. Oktober 1462 gewaltsam beendet wurde. Jedes abgebildete Objekt trägt eine entsprechende Zeitangabe als Sachdateninformation. Die Klassifikation der 602 Gebäude basiert auf sozialtopografischen Kriterien oder Nutzungsinformationen, die zusammen mit Bewohner-, Eigentümer- und Bauinformationen ebenfalls als Attributwerte erfasst wurden. Erwähnenswert ist auch die Stadtmauer, die gemäß ihrer mittelalterlichen Ausprägung generalisiert wurde. Dies ist ein wichtiger Unterschied zum Falck-Plan, der ein ähnliches Bild für die Zeit um 1620 zeichnet.

Für das historische Häuserbuch wurde allerdings die gegenteilige Entscheidung getroffen. Das Modell für die Zeit um 1620 ist eindeutig von den bastionären Verteidigungswerken geprägt. Auch dieses Modell bezieht sich nicht auf ein Stichjahr, sondern auf einen Teilabschnitt, in dem Mainz kurfürstliche Residenzstadt war. Falck (1974) datierte den Beginn der gesamten Phase auf das Jahr 1462, als Mainz die Stadtfreiheit aufgrund der Eroberung durch Adolf II. von Nassau verlor und Buschbaum (2002) datierte das Ende der gesamten Phase auf das Jahr 1792 (Einnahme der Stadt durch das französische Revolutionsheer).

Um das Modell nicht zu überladen, wurden die Informationen auf einen Teilabschnitt beschränkt, der mit dem Beginn im Jahr 1590 und dem Ende des Dreißigjährigen Krieges (1648) grob umrissen werden kann. Die Zerstörungen, die aus dem o.g. Konflikt resultierten und ganze Quartiere veränderten (Buschbaum, 2002), wurden in diesem Modell bewusst ausgespart.

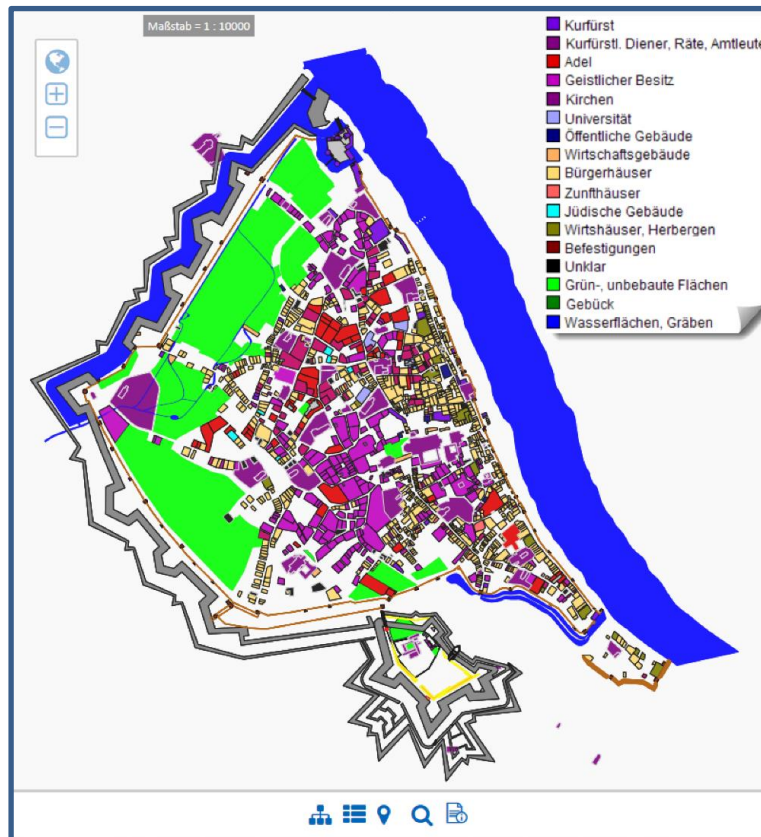


Abb. 4.5: Mainz um 1620 (gerendert von MapServer V. 7, dargestellt von Mapbender V. 3)

Die Klassifikation und die Attribute sind für beide Modelle (um 1450 / um 1620) fast identisch. Jedoch konnte die Anzahl der erfassten Gebäude (1424) und die Genauigkeit der Grundrisse erhöht werden, da eine größere Anzahl an hochwertigen Quellen zur Verfügung stand. Zusätzlich zum georeferenzierten Falck-Plan wurden für die Datenerfassung auch Urkunden und Protokolle des Mainzer Stadtgerichts, die Steuerliste von 1614 und Grundrisspläne einzelner Gebäude verwendet. Wichtige Sekundärquellen waren außerdem die Stadtaufnahmen von 1594, 1644, 1657 und 1687, die von Schrohe (1929, 1930) editiert wurden. Den Wert dieser Quellen für das historische Häuserbuch lassen die Ausführungen von Buschbaum (2002) erkennen, die eine Stadtaufnahme als systematische Datenerhebung durch eine spezielle Kommission beschrieb, welche die gesamte Stadt Mainz abschnitt und den Zustand der Gebäude und die Eigentumsverhältnisse der Grundstücke dokumentierte. Für die Rekonstruktion der Schweickhardtsburg mit dem darin integrierten Benediktinerkloster St. Jakob waren jedoch noch weitere Pläne erforderlich, die in den Stadtarchiven von Würzburg und Mainz aufbewahrt werden.

Wie aufwendig die Georeferenzierung der historischen Pläne war, lässt sich exemplarisch anhand der folgenden schematischen Abbildung erläutern, welche den Prozess zur Digitalisierung der Bundesfestung (1866) illustriert.

Das Häuserbuch für die Zeit um 1871 soll nicht verschweigen, dass Mainz allmählich zu einer mächtigen Festung ausgebaut wurde. Allerdings bietet der, von Falck verwendete, älteste maßstabrichtige Mainzer Stadtplan aus dem Jahr 1871 keine Grundlage für die Digitalisierung der Festung, da diese vollkommen ausgespart wurde. Deshalb mussten 25 einzelne Kartenblätter aus einer Mappe, die während des zweiten deutschen Einigungskriegs angelegt wurde, georeferenziert und mit einem Übersichtsplan der Reichsfestung aus dem Jahr 1873 kombiniert werden, um eine verlässliche Grundlage für die Digitalisierung der Bundesfestung zu schaffen. Ähnlich aufwendig gestaltete sich die Georeferenzierung des (Kataster-)Stadtplans von 1871, da sich herausstellte, dass der Plan falsch zusammengesetzt wurde und außerdem Lücken aufwies. Angesichts dieser Mängel mussten weitere Pläne aus dem Mainzer Stadtarchiv herangezogen werden, um eine adäquate Ausgangsbasis für die Erstellung des Häuserbuches für die Zeit um 1871 zu schaffen. Diesbezüglich ergeben sich, im Vergleich zu den bereits erläuterten Modellen, die folgenden Besonderheiten.

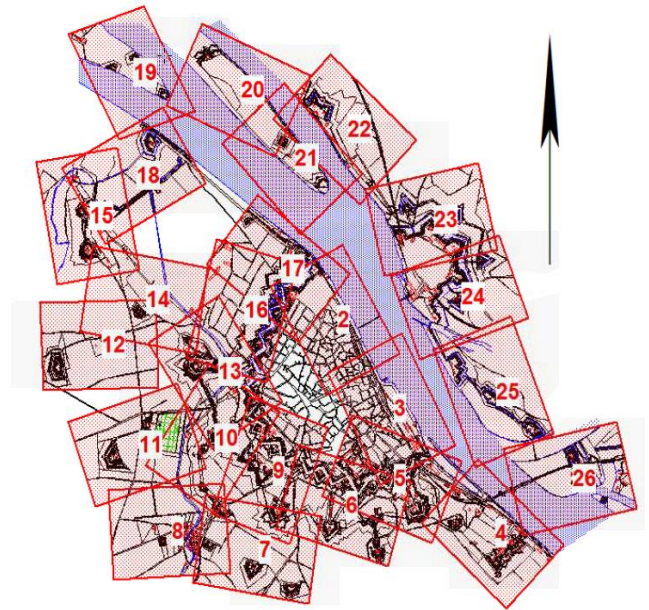


Abb. 4.6: Systematische Georeferenzierung

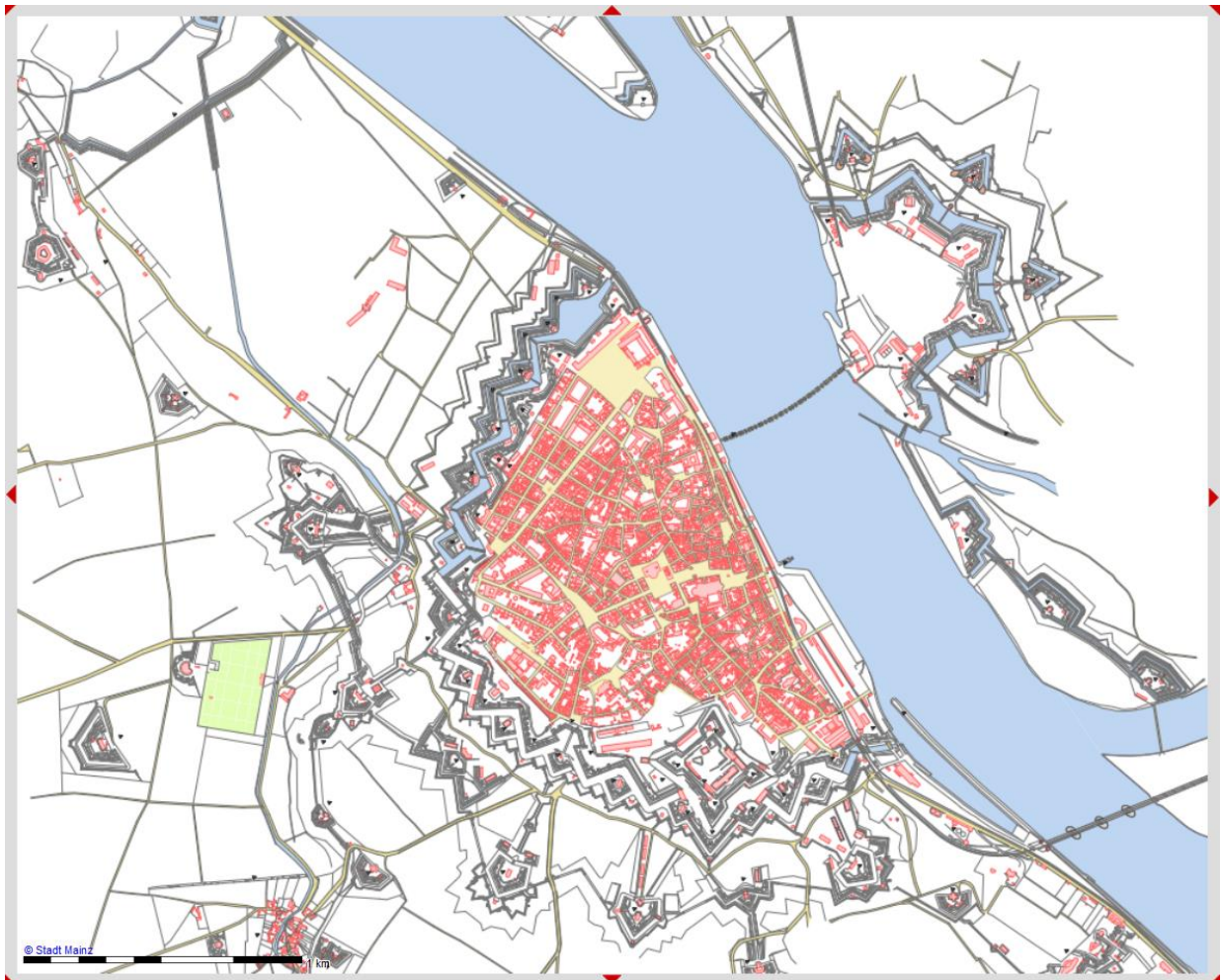


Abb. 4.6: Mainz um 1871 (gerendert von MapServer V. 5, dargestellt von Mapbender V. 2)

Die 5705 Häuser sind zwar mit einer Zeitangabe verknüpft, alle weiteren o.g. Sachdateninformationen fehlen jedoch. Im Gegensatz zu den Modellen für die Zeit um 1450 und 1620 ist deshalb keine Klassifikation anhand der Nutzung oder aufgrund sozialtopografischer Faktoren möglich. Eigentümer- und Bewohnerangaben existieren zwar, allerdings nicht medienbruchfrei, da diese in einem separaten Verzeichnis in Form gescannter Dokumente vorliegen. Der indirekte Raumbezug wird für jedes Haus zusätzlich über die erfassten Attributwerte in den Feldern „Strasse“, „Hausnummer“ hergestellt. Die übrigen topographischen Objekte (z.B. die Straßen) tragen jedoch nicht genügend Attribute, weshalb eine genauere Klassifizierung momentan nicht möglich ist.

Ergänzend wird auf die Karte von Falck hingewiesen, die ausgewählte Aspekte der Stadtentwicklung vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts zusammenfasst. Sie war für vorbereitende Maßnahmen sehr nützlich, die im folgenden Kapitel angerissen werden.

MAINZ vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts



Abb. 4.7: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts, OA unverändert
 übernommen von: <https://www.lagis-hessen.de/de/subjects/idrec/sn/ga/id/65>

4.3 Verwendungsbeispiel aus der Praxis

Münkel (2019, p. 1) nutzte für seine Masterthesis mit dem Titel „THEATRUM MOGONTIACENSIMUM“ das historische Häuserbuch von Mainz als Planungsgrundlage. Münkel (2019) entwickelte u. a. ein Besucherzentrum für das Römische Theater von Mogontiacum (der lateinische Name der heutigen Stadt Mainz). Die Rasterdarstellungen des historischen Häuserbuchs waren dafür nicht ausreichend, weshalb die Geodaten in vektorieller Form zur Verfügung gestellt wurden. Münkel (2019) beschrieb damit z.B. die städtebauliche Situation für die Zeit um 1450:



Abb. 4.8: Topographische Objekte um 1450, OA von Münkel (2019, p. 8)

Für ähnliche Vorhaben (z.B. für das eingangs erwähnte CDV-Projekt oder für Festungsforschungen) ist es sinnvoll, dass der (webbasierte) Zugriff auf die vektoriellen Geoobjekte ohne Hürden (z.B. die manuelle Konvertierung oder die Installation clientseitiger Erweiterungen) ermöglicht wird. Dafür bietet sich (momentan) der WFS an – insbesondere, da über diesen Dienst auch LGCHD bereitgestellt werden kann.

5 Datenuntersuchung, -auswahl, -anpassung

Dieses Kapitel enthält wichtige Details zu den Daten und Informationen, die zwecks Generierung und Anreicherung des Graphen verwendet werden. Neben ausgewählten MapInfo-Relationen aus dem historischen Häuserbuch, die zuerst thematisiert werden, zählen auch die Informationen der DNB und Vokabulare aus der LOD-Cloud dazu, die anschließend in Abschnitt 5.2 vorgestellt werden.

5.1 Historisches Häuserbuch

Nachdem ein überblicksmäßiger Eindruck von den Modellen des historischen Häuserbuchs gewonnen wurde, erfolgt nun die exakte Beschreibung der Datenbasis. Idealerweise sollte dies zunächst anhand eines Anwendungsschemas auf der konzeptuellen Ebene erfolgen. Ein solches Modell existiert jedoch nicht. Selbiges gilt für die Dokumentation der Anforderungen in Form eines Pflichtenhefts. Es widerspricht der Konzeption dieser Arbeit, den gesamten Modellierungsprozess – vom Anforderungskatalog über ein konzeptuelles zum logischen Datenmodell zwecks (erneuter) physischer Implementierung – zu durchlaufen, da eine möglichst direkte Transformation der Ausgangsdaten in graphenbasierte Strukturen angestrebt wird. Deshalb werden lediglich die Charakteristika der vektoriellen Geodaten (MapInfo-Relationen) untersucht bzw. dokumentiert.

5.1.1 Vektormodell

Datei- und Datenstruktur

Sämtliche relevante Daten liegen im nativen MapInfo-Datenformat vor, das alternativ als MapInfo-TAB-Format oder MapInfo-Relation bezeichnet wird⁷³. In dieser Arbeit wird nur die letztgenannte Bezeichnung verwendet. Es handelt sich um ein proprietäres Format, das im Desktop-GIS „MapInfo Professional“ standardmäßig für die Verarbeitung, Verwaltung und die Persistierung von Geoobjekten zum Einsatz kommt. Der Hersteller des o.g. GIS, früher „Pitney Bowes Software & Data“, dann „Syncsort“ und mittlerweile⁷⁴ „Precisely“ veröffentlicht(e) keine Dokumentation des Formats⁷⁵. Erwähnenswert ist die Open-Source-Bibliothek namens MITAB⁷⁶, die jedoch vom Hersteller des GIS unabhängig ist. Eine MapInfo-Relation ist dateibasiert; sie besteht i.d.R. nicht aus einer einzigen Datei, sondern aus mehreren Dateien, die Daten separat abspeichern. Die Dateien gehören aber zusammen, weshalb sie den gleichen Dateinamen aufweisen. Die Dateierweiterung differiert jedoch:

- 1) Dateiname.TAB: ASCII, beschreibt die Relationsstruktur.
- 2) Dateiname.DAT: ASCII und binär (Header, Tupel-Informationen in Form von BLOBs o. Ä.), enthält die Sachdaten. Die Dateierweiterung kann variieren, je nachdem, ob dBase oder Access eingesetzt wird.
- 3) Dateiname.IND: binär, dient der Indexierung der thematischen Daten.
- 4) Dateiname.MAP: binär, enthält geographische Informationen der Relation; dient der Visualisierung.
- 5) Dateiname.ID: binär, enthält den Index für die MAP-Datei zur Verknüpfung mit der DAT-Datei.

Die Relation kann durch weitere Dateien ergänzt werden, z.B. durch XML-Strukturen für die Speicherung von Parametern für Anfragen gegen OWS, temporäre Dateien für die Prozessierung, etc. Für spezielle Anwendungsfälle kann die obligatorische TAB-Datei ggf. ausreichend sein. I.d.R. wird aber mindestens die TAB- und die DAT-Datei verwendet (sofern keine graphische Darstellung / Visualisierung erforderlich ist). Für Vektormodelle (mit Kartendarstellung) werden i.d.R. die fünf o.g. Dateien verwendet.

⁷³ Das MapInfo Data Interchange-Format (MIF/MID) spielt für diese Thesis keine Rolle.

⁷⁴ Stand November 2020, siehe: <https://www.precisely.com/about-us/pitney-bowes-software-and-data>

⁷⁵ Eine Übersicht kann hier eingesehen werden: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000300.shtml>

⁷⁶ Projektseite: <http://mitab.maptools.org/>

Gemeinsame Objekteigenschaften

Dies trifft auf alle Relationen zu, aus denen das historische Häuserbuch besteht. Die fünf o.g. Dateien sind also immer vorhanden. Es handelt sich um Vektormodelle, die aus einer Menge von Geoobjekten bestehen, deren *gemeinsame* Eigenschaften nachfolgend erläutert werden.

Prinzipiell bietet eine MapInfo-Relation die Möglichkeit, **geometrische Eigenschaften** vektorieller Geoobjekte über verschiedene Geometrietypen zu beschreiben und gemeinsam zu speichern. Davon wurde bei der Erstellung des historischen Häuserbuchs kein Gebrauch gemacht. Jede Relation enthält lediglich eine Menge homogener Geometrien. Dies sind entweder Polygone oder Linien. Polygone können keinen/einen inneren Ring oder mehrere innere Ringe aufweisen. Linien können als MapInfo-Line oder -Polyline vorliegen. Erstere hat nur zwei Stützpunkte. Letztere hat mehrere Stützpunkte. In beiden Fällen sind die Stützpunkte geradlinig verbunden, was auch für Polygone gilt. Splines, Kreisbögen, o. Ä. wurden dementsprechend nicht verwendet. Selbiges gilt für Geometrieaggregate, -sammlungen oder -komplexe. Deshalb ist die Beurteilung, ob disjunkte Linien oder Linienzüge eine realweltliche Entität oder verschiedene Objekte repräsentieren sollen, mitunter schwierig. Häufig können qualifizierte Entscheidungen bzw. logische Zuordnungen nur aufgrund räumlicher Nähe, gleicher attributiver Information und mithilfe von historischen Quellen getroffen werden (s.u.). Es existieren ausschließlich zweidimensionale Geometrien, da keine z-Koordinaten vorhanden sind und Höhenangaben auch nicht als Attributwerte erfasst wurden. Jedem Objekt ist nur ein Geometriedatensatz zugeordnet.

Topologische Aspekte konnten bei der Datenerfassung nicht besonders berücksichtigt werden, da MapInfo keine Funktionen für die Definition von Topologieregeln bereitstellt. Die Konsistenz der Geometrien wurde teilweise über andere Erfassungsfunktionen sichergestellt (z.B. Fangfunktionen oder das automatische Vervollständigen / Schließen von Polygonen). Allerdings enthalten einige Relationen lediglich „Spaghetti-Daten“, die in topologischer Hinsicht keine Struktur aufweisen. Die Geometrie wurde ursprünglich noch unter Verwendung des Koordinatenreferenzsystems Deutsches Hauptdreiecksnetz 1990 (DHDN90) und der Gauß-Krüger-Abbildung im 3°-Streifensystem (GK3) erfasst. Alle notwendigen Angaben zum ursprünglichen CRS können bei Bedarf unter [EPSG:31467](#) eingesehen werden. Der Bezugssystemwechsel auf ETRS89/UTM ([EPSG:25832](#)) wurde vollzogen.

Die Kombination aus verwendetem Desktop-GIS (MapInfo-Version: 5.5) und dBase III limitierte die Möglichkeiten der Konsistenzsicherung auch hinsichtlich **thematischer Informationen** erheblich. Es gibt, im Gegensatz zu anderen Desktop-GIS, (abgesehen vom Datentyp) keine äquivalenten Funktionen, um Domänen oder ähnliche Integritätseinschränkungen definieren zu können. Wenn nachfolgend von Domänen gesprochen wird, bezieht sich der Begriff dementsprechend auf Erfassungsdisziplin, statt auf einen technischen Automatismus zwecks Vermeidung von Fehlerfassungen. Das Feld „Nummer“ wird in mehreren Relationen verwendet, allerdings wurden keine Schlüsseleinschränkungen definiert und die Werte erfüllen die Anforderungen an einen Primärschlüssel nicht. Es existieren (beliebige) Kombinationen von alphanumerischen Zeichen. Die Relationen per se stehen in keiner spezifizierten Abhängigkeit; Bedingungen zur Sicherung der referenziellen Integrität existieren nicht. Werte in Feldern vom Typ „Char“ dürfen lediglich eine maximale Anzahl von Zeichen nicht überschreiten.

Jedes Geoobjekt in jeder Relation verfügt über assoziierte Sachdaten. Wenn einem Objekt explizit Attributwerte in Form einer Zeichenkette zugeordnet wurden, beschreibt diese stets nominale Merkmale. Quantitative Eigenschaften sind entweder implizit vorhanden, da z.B. die Grundfläche eines Hauses aus der Geometrie abgeleitet werden kann oder sie wurden in Form numerischer Werte erfasst. Dazu zählen z.B. die Zeitangaben. Die Werte des Feldes „Nummer“ werden aufgrund ihrer Bedeutungslosigkeit hier nicht weiter berücksichtigt.

Die vorgestellten Rastermodelle (vgl. Abb. 4.4 und Abb. 4.5) für die Zeitabschnitte um 1450 und 1620 sind vor allem durch die Klassifikation der Häuser geprägt, die in quantitativer Hinsicht dominieren. Grundlage der Klassifikation bilden die Werte des Feldes „Nutzung“. Die Werte geben entweder tatsächlich Auskunft zur primären Nutzung des Gebäudes / zum „Typ“ des Gebäudes oder sie ordnen *das Gebäude* einer historischen, sozialen Gruppe zu. Anhand der Rasterdarstellungen ist erkennbar, dass z.B. die türkis eingefärbten Polygone jüdische Gebäude repräsentieren sollen. Denecke (1987, p. 199) wies im Rahmen seiner sozialtopographischen Analyse der mittelalterlichen Stadt Göttingen darauf hin, dass die sozialen Verhältnisse in einer Stadt nicht nur am sozialen Status einzelner Personen oder Gruppen greifbar werden, sondern „auch in spezifischen sozial-räumlichen – beziehungsweise Haus für Haus für das gesamte Stadtgebiet betrachtet – in sozialtopographischen Anordnungsmustern“. Allerdings erarbeitet er dieses sozialtopographische Anordnungsmuster anhand von Kriterien, die *nur für Menschen* zutreffen konnten, wie z.B. Religions- oder Berufszugehörigkeit (Denecke, 1978). Buschbaum (2002) erarbeitete vergleichbare Anordnungsmuster auf Basis des oben gezeigten Modells für die Zeit um 1620, indem sie zusätzliche Informationen aus den zahlreichen Mainzer Stadtaufnahmen analysierte. Jedoch bezog auch Buschbaum (2002) diese Informationen auf die Menschen, die mit dem jeweiligen Gebäude in einer bestimmten Beziehung standen (z.B. Besitzerinnen).

Viele Attributwerte können in mehrere Komponenten zerlegt werden, z.B. könnte der Wert im Feld „Literatur“ „Frank: Das Totenbuch des Mainzer Dominikanerklosters (1993)“ in Autor, Titel, Erscheinungsdatum aufgeteilt werden. Zusammengesetzte und mehrwertige Attribute bilden die Regel, nicht die Ausnahme.

Alle Sachdaten eines Objekts gelten für einen bestimmten Zeitpunkt, der mit dem numerischen Attributwert im Feld „Jahr“ festgelegt ist. Wie bereits erwähnt, kann diese Angabe für Objekte innerhalb *eines* Modells differieren. Insofern kann der „Wertebereich“ des Attributs „Jahr“ z.B. im zweiten Modell mit 1590 – 1648 angegeben werden, weshalb nicht exakt das Stichjahr 1620 repräsentiert wird. Die übrigen Felder können noch weitere Zeitangaben enthalten, z.B. das Feld „Bauinformationen“. Da diese jedoch als Bestandteil einer Zeichenkette ohne feste Systematik erfasst wurden, bietet sich keine strukturierte Zugriffsmöglichkeit auf diese Daten. Für die Beurteilung der **Dynamik** der Geoobjekte wird also rein auf die Werte im Feld „Jahr“ abgestellt. Diesbezüglich ist festzustellen, dass es sich um statische Modelle handelt. Die Dynamik einzelner Objekte kann anhand eines Modells nicht beschrieben werden. Dies wäre ggf. durch die Kombination mehrerer Vektorlayer (aus unterschiedlichen Modellen) und mithilfe geeigneter Operatoren möglich, sofern es sich tatsächlich um Ebenen/Layer im eigentlichen Sinne handelt (s. Abb. 5.2).

Jede Relation enthält nur wenige **Metadaten**, die am Anfang der TAB-Datei stehen, z.B.:

```
!table
!version 300
!charset WindowsLatin1
```

Am Ende der TAB-Datei, finden sich ergänzende Angaben je Relation, die für diese Arbeit jedoch wenig Relevanz aufweisen und daher nachfolgend nicht separat aufgeführt werden:

```
begin_metadata
"\IsReadOnly" = "FALSE"
"\MapInfo" = ""
"\MapInfo\TableID" = "db85db4a-6513-43bc-af24-5e61e7572422"
end_metadata
```

MapInfo-Relationsstruktur

Die nächste Abbildung dient vor allem als Übersicht zu den **ausgewählten** MapInfo-Relationen; die Begründung, warum einige MapInfo-Relationen angepasst oder ausgeklammert wurden, folgt.

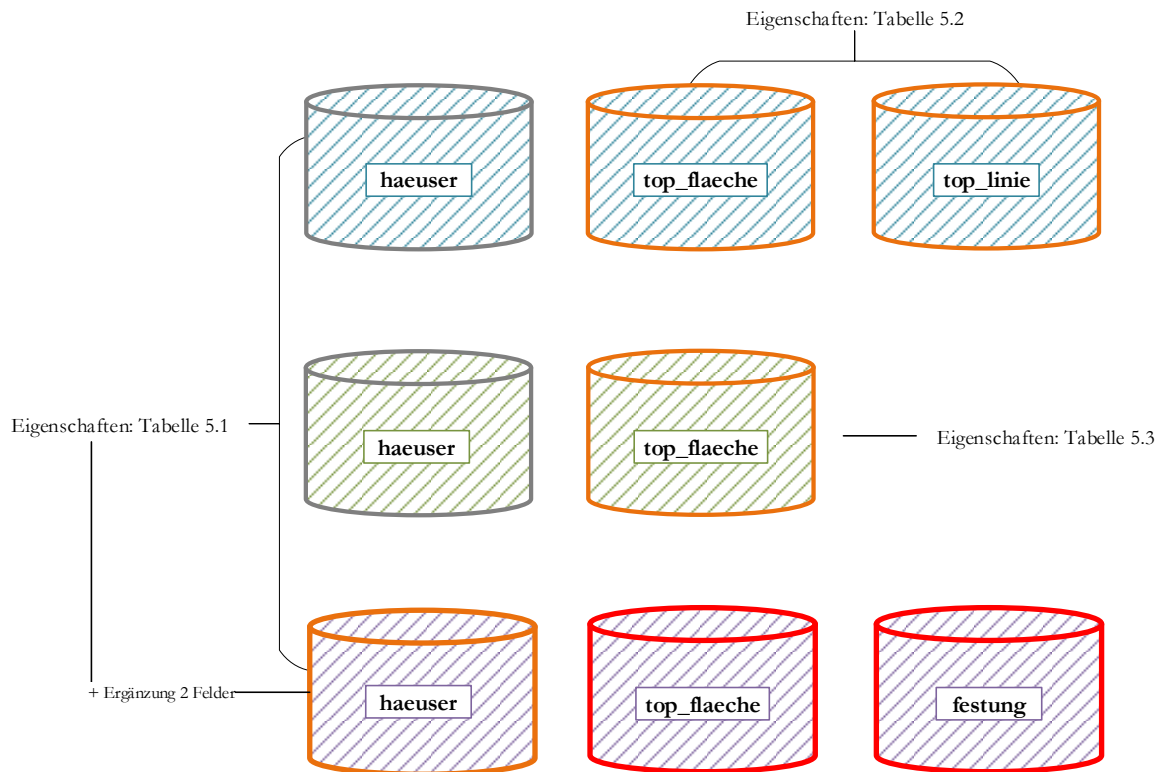


Abb. 5.1: Übersicht zur Datenbasis, ED

Das dosenähnliche Symbol („Datenbank“) repräsentiert die MapInfo-Relation insgesamt, d.h. alle erwähnten Dateien, aus denen sie sich zusammensetzt. Die Schraffur zeigt an, dass die Modelle für die verschiedenen Zeitabschnitte aus mehreren Relationen bestehen (1450 = hellblau, 1620 = grün, 1871 = violett). Die Eigenschaften der Relationen, welche nur Tupel enthalten, die Häuser repräsentieren, werden gemeinsam in Tab. 5.1 beschrieben. Für die Zeit um 1871 folgt eine Ergänzung, da die Relation zwei weitere Felder enthält. Wenn das Datenbank-Symbol eine **graue Konturlinie** aufweist, werden die Relationen **ohne Änderung** in den Transformationsprozess eingebunden. Eine **orangefarbene Konturlinie** zeigt an, dass die ursprünglichen Daten im Vorfeld **angepasst** werden mussten und eine **rote Konturlinie**, dass die Relationen im Rahmen dieser Arbeit **nicht verwendet** werden.

Die folgenden Tabellen konkretisieren die allgemeinen Informationen des vorherigen Abschnitts. Falls möglich, werden Domänen vollständig aufgeführt. Anderenfalls ersetzt ein repräsentatives Beispiel (vorerst) die Auflistung aller Werte, um die Übersichtlichkeit zu wahren.

Relationsname	1450_haeuser	1620_haeuser	1871_haeuser
Anzahl der Objekte	602	1424	5705
Größe im Dateisystem	2469 KB	4576 KB	17479 KB
Geometrie	Polygone		

Relationsdefinition (TAB-Datei)	Beispiel	Domäne/Erläuterung
Nummer Integer ;	1202	s.o.: Stichwort „Schlüsseinschränkungen“
Hausname Char (254) ;	Zum Gutenberg	variable Zeichenketten, ohne distinktives Muster, die immer vorhanden sind
Bauinformationen Char (254) ;	Protonotariat. Eh. Juristenkolleg, Universitätsburse	nicht immer vorhanden; s.o.: Stichwort „Bauinformationen“
Nutzung Char (254) ;	UNI	NULL, Adel, Befestig, Buerger, Herberge, juedisch, K_Dienst, K_Fuerst, Kir_bes, Kirche, Turm, oeffentl, Patriz, UNI, unklar, Wirt_geb, Zunft
Eigentuemmer Char (254) ;	Universität Mainz	Zwischen den Entitätstypen (ET) Haus und Person (in ihrer Rolle als Eigentümerin/Bewohnerin) besteht eine m:n-Beziehung. Es können NULL-Werte, ein- oder mehrwertige, zusammengesetzte Attributwerte in Form von variablen Zeichenketten vorkommen. Diese entsprechen keinem distinktiven Muster und qualifizieren sich nicht als eindeutige Referenzen auf die jeweiligen Bezugsobjekte (Personen). Wenn mehrere Eigentümerinnen/Bewohnerinnen existieren, trennt die Zeichenfolge Komma + Whitespace die einzelnen Personen.
Bewohner Char (254) ;	Dr. Jacobus Campius	
Literatur Char (254) ;	Schrohe: Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts / Teil 1 (1930)	Die vorstehenden Ausführungen gelten grundsätzlich analog. Eine Ausnahme ist darin zu sehen, dass sich die Zeichenketten aus „Informationseinheiten“ (z.B. die Zeichenfolge „Schrohe“) zusammensetzen, die in ihrer Gesamtheit ausreichend sind, um die jeweiligen Bezugsobjekte (Publikationen) zu identifizieren. Abgesehen davon, weisen die Zeichenketten ein distinktives Muster auf, weshalb sie automatisch in Autorin, Haupttitel, ggf. Titelnzusatz und Erscheinungsjahr zerlegt werden können.
Bild1 Char (254) ;	NULL	Wiederholungsgruppe; NULL oder syntaktisch korrekte URIs respektive URLs
Bild2 Char (254) ;		
Bild3 Char (254) ;		
Bild4 Char (254) ;		
Bild5 Char (254) ;		
Jahr Integer ;	1594	vierstellige Ziffernfolge (immer vorhanden)

Tab. 5.1: Eigenschaften der „Häuser-Relationen“

Die Relation für 1871 verfügt zusätzlich über die Spalten „Strasse“ und „Hausnummer“, die überwiegend mit Werten gefüllt wurden. Selbiges gilt hinsichtlich der Bauinformationen. Eine Jahresangabe und ein Hausname wurden für jedes Objekt erfasst. Alle anderen o.g. Felder enthielten ursprünglich jedoch nur NULL-Werte.

Die übrigen Relationen für die Zeit um 1450 haben folgende Eigenschaften:

Relationsname	1450_top_flaeche	1450_top_linie
Anzahl der Objekte	61	66
Größe im Dateisystem	48 KB	29 KB
Geometrie	Polygone	Linien oder Linienzüge
Felder (TAB-Datei)	Domäne	
Nutzung Char (8) ;	NULL, Befestig, Gebueck, Gruenanl, Wasser	NULL, Mauer
Nummer Char (10) ;	s.o.: Stichwort „Schlüsseleinschränkungen“	
Beschreibung Char (249) ;	NULL, variable Zeichenketten	
Jahr Integer ;	vierstellige Ziffernfolge (immer vorhanden)	

Tab. 5.2: Gemeinsame Struktur und individuelle Eigenschaften der übrigen Relationen für 1450

Das Modell für die Zeit um 1620 wird durch die folgende Relation komplettiert:

Relationsname	1620_top_flaeche
Anzahl der Objekte	246
Größe im Dateisystem	136 KB
Geometrie	Polygone
Felder (TAB-Datei)	Domäne
Nutzung Char (8) ;	NULL, Befestig, Bruecke, Brunnen, Festung, Gruenanl, Mauer, Wasser
Nummer Char (10) ;	s.o.: Stichwort „Schlüsseleinschränkungen“
Beschreibung Char (249) ;	NULL, variable Zeichenketten
Jahr Integer ;	vierstellige Ziffernfolge (immer vorhanden)

Tab. 5.3: Relationsstruktur und Eigenschaften der übrigen Objekte für 1620

Da für die Zeit um 1871 lediglich die Häuser-Relation verwendet wird, erfolgt keine exakte Dokumentation der restlichen Relationen. Eine kurze Charakterisierung ist im nächsten Abschnitt enthalten, der auch die Begründungen für den Ausschluss und die Anpassungen enthält.

Anpassung und Ausschluss von Relationen

Die erwähnten geometrischen, topologischen und thematischen Eigenschaften erschweren die angestrebte Generierung adäquater RDF-Repräsentationen. Deshalb mussten einige Änderungen/Ergänzungen an bestimmten Relationen vorgenommen werden. Diese können nicht (semi-)automatisch umgeformt werden. Kein Automatismus kann die manuelle Überprüfung anhand historischer Quellen ersetzen. Viele Objekte mussten mithilfe der Karte „Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts“ von Falck analysiert werden (vgl. Abb. 4.7). In Kombination mit den Erklärungen, die Falck zu den verwendeten kartographischen Signaturen lieferte, konnten einige Probleme relativiert werden (s.u.). Die Maßnahmen sind dem Tätigkeitsfeld „Datenerfassung“ zuzuordnen; sie sind nicht als integraler Bestandteil dieser Arbeit anzusehen. Dementsprechend erfolgt keine detaillierte Dokumentation, wie dies mit verschiedenen Desktop-GIS und ETL-Tools realisiert wurde. Falls erforderlich, werden die Resultate jedoch vollständig – sowohl in vektorieller als auch in tabellarischer Form – dokumentiert. Schließlich bilden sie die Ausgangsbasis für die Transformation. Diesbezüglich muss klargestellt werden, dass das Ziel dieser Arbeit nicht darin besteht, alle Probleme im Vorfeld durch Datenerfassungen zu lösen. Dafür muss der gesamte Modellierungsprozess durchlaufen werden (vgl. GP 5.1).

Die Datenerfassungen, die nachfolgend anhand repräsentativer Beispiele erläutert werden, sind erforderlich, um einige Zusammenhänge in den zu generierenden RDF-Repräsentationen *wahrheitsgemäß* angeben zu können. Im Zuge der Datenuntersuchung wurde nicht nur die Geometrie und die Topologie in Frage gestellt, sondern auch der Sinn- und Wahrheitsgehalt der vorhandenen Aussagen. Abgesehen von den Aussagen bzgl. der sozialtopographischen Anordnungsmuster (z.B. die Aussage, dass es jüdische Häuser gab), betrifft dies z.B. die Aussage, dass ein Haus bzw. ein Gebäude⁷⁷ eine Eigentümerin hatte. In dieser Arbeit wird angenommen, dass die erste Aussage *falsch* ist. Im Gegensatz dazu wird angenommen, dass die zweite Aussage *wahr* ist, weil diese Arbeit nicht auf die Erfassung von Polygonen abzielt, um sämtliche Grundstücke (oder Flurstücke) für alle Zeitabschnitte beschreiben zu können. Stattdessen wurden die Eigentümerinnen- und Bewohnerinnen-Informationen mithilfe des GIS MapInfo Professional (V. 17) in der Relation „1871_haeuser“ ergänzt, sofern dies auf Basis des historischen Adressbuchs von Mainz⁷⁸ möglich war. Die Relationsstruktur wurde nicht verändert; die erforderlichen Felder waren bereits vorhanden (s.o.). Die Systematik wurde ebenfalls beibehalten, d.h., wenn mehrere Eigentümerinnen/Bewohnerinnen eines Hauses existierten, wurden die Daten in einem Feld erfasst und durch die Zeichenfolge Komma + Whitespace getrennt. Für wenige Gebäude (111) konnte auch der Typ erfasst werden. Dabei wurden nur die Werte aus Tab. 5.1 verwendet, die tatsächlich den Gebäudetyp angeben (Befestig, Herberge, Kirche, Turm, oeffentl, Wirt_geb). Da lediglich 111 Gebäude getypt wurden (~ 2 %), konnten die Gebäude nicht sinnvoll klassifiziert werden. Die Relationen „1450_haeuser“ und „1620_haeuser“ blieben unverändert. Geometrische Eigenschaften in der Relation „1871_haeuser“ wurden nicht verändert. Eine erneute Dokumentation in tabellarischer Form ist insofern nicht erforderlich.

Die vorhandenen Geometrien und die Sachdaten in den Relationen „1450_top_flaeche“, „1450_top_linie“ und „1620_top_flaeche“ wurden in ArcGIS Pro (V. 2.6.0) verändert. Für jedes Objekt wurde ein Name erfasst. Die Informationen aus den vorgenannten MapInfo-Relationen wurden über die ArcGIS Data Interoperability Extension in Feature-Klassen importiert, die anschließend editiert wurden. Exemplarisch lässt sich der Prozess anhand einiger Objekte des Feature-Typen „Wasser“ darstellen.

⁷⁷ Zur Klarstellung des Problems wird angenommen, dass das Haus ein Gebäude ist (vgl. Legaldefinition unter GP 8.2.2 und https://www.gesetze-im-internet.de/bgb/_94.html). Die einschlägigen Regelungen sind alt; die Auflassung ist sehr alt.

⁷⁸ <https://www.dilibri.de/stbmz/periodical/structure/1225472>

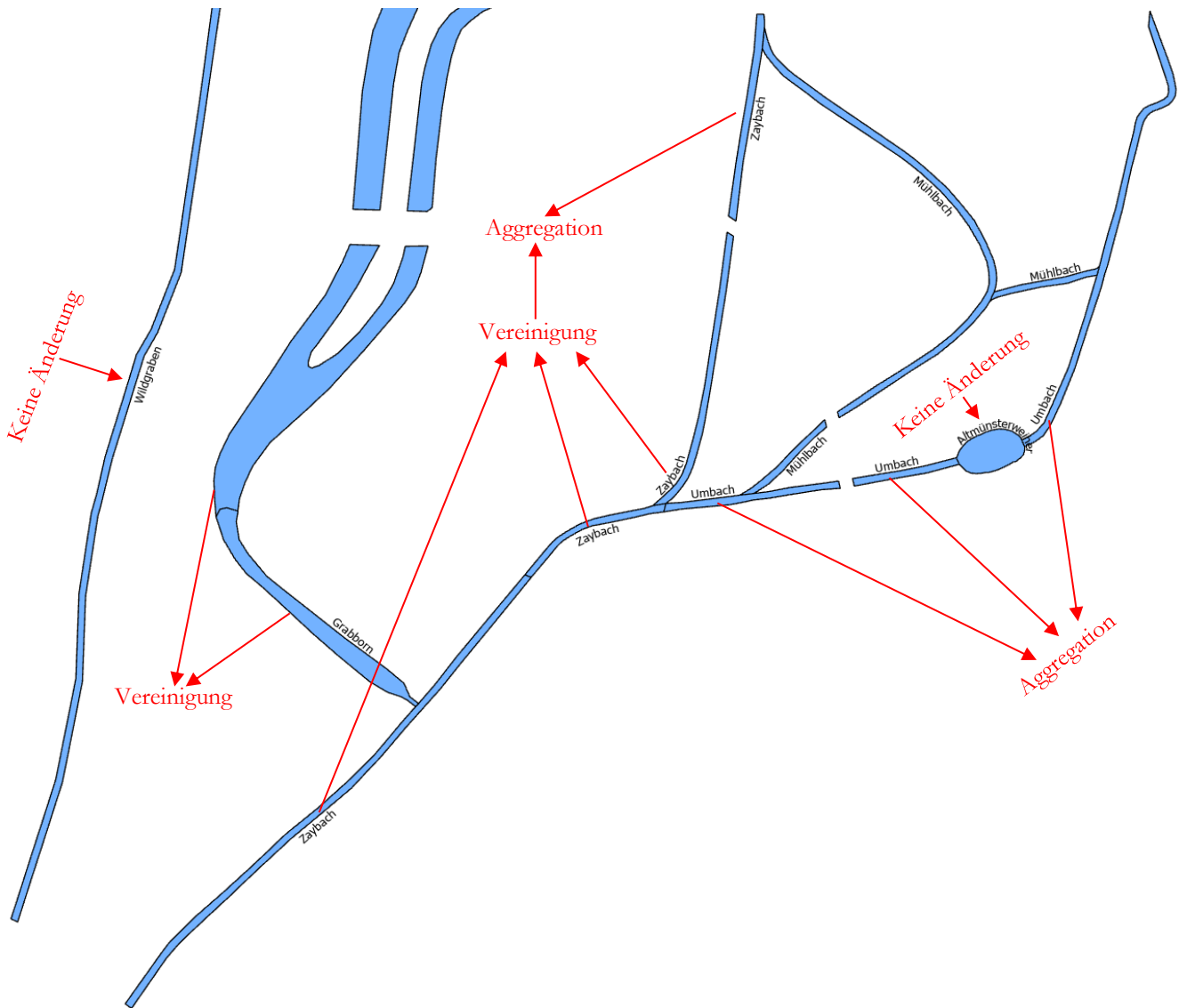


Abb. 5.2: Beispiel zur Datenerfassung in ArcGIS Pro, ED

Features wurden nur dann vereinigt, d.h., die Geometrien wurden zusammengeführt, wenn die Sachdaten zweifelsfrei auf das gesamte vereinigte Feature bezogen werden konnten. Dies wurde anhand der o.g. Quelle überprüft, die auch dazu diente, fehlende Attribute zu ergänzen. Die Aggregation von Features kann in vielen Fällen anhand historischer Quellen begründet werden. Gerade die Stadtmauer kann nicht blindlings vereinigt werden, da der Verlauf z.B. für die Zeit um 1450 an vielen Stellen ungewiss ist (Falck, 1978). Es ist also durchaus sinnvoll die disjunkten Linien(-segmente) beizubehalten, allerdings in Form einer Geometriesammlung. Die Sachdaten gelten gleichermaßen für die gesamte Sammlung. Es entsteht kein Informationsverlust. Die Datensätze, welche die Stadtmauer repräsentieren, enthalten z.B. nie eine Beschreibung (NULL) und immer dieselbe Zeitangabe.

Wenn dies nicht der Fall war, wurden die Geometrien nicht in die Sammlung aufgenommen, sondern es wurde eine separate Sammlung gebildet. Beispielsweise wird hier deutlich, dass der **Vilzbacher Burggraben** (B 57) von der **Stadtmauer** separiert werden kann:



Abb. 5.3: Abgleich von Objekten in der Relation mit der Karte von Falck (1978)

In anderen Fällen war die Aggregation ein Kompromiss, weil keine Ebenen/Layer im eigentlichen Sinne vorhanden sind (vgl. Abb. 5.2). Der folgende Kartenausschnitt z.B. lässt vermuten, dass die Altmünstergasse (315) den Umbach (412) oder den Mühlbach (410) wohl nicht wirklich unterbrochen hat:

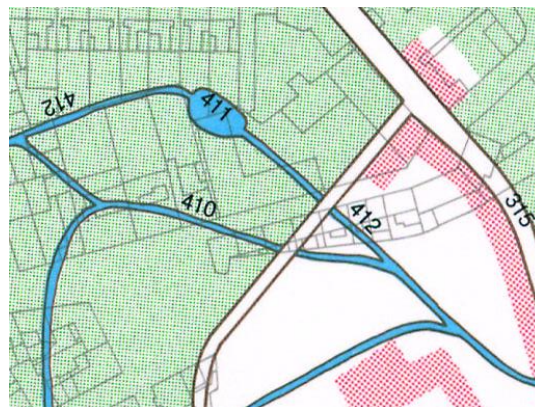


Abb. 5.4: Aggregation als Kompromiss, Kartenausschnitt aus Falck (1978)

Die spezifischen Charakteristika der verwendeten File-Geodatabase, deren Feature-Klassen, Domänen, Subtypes etc. werden nicht thematisiert. Die Resultate wurden nach Abschluss der Datenerfassung über die ArcGIS Data Interoperability Extension als MapInfo-Relationen exportiert. Deren Eigenschaften werden nachfolgend dokumentiert.

Relationsname	1450_top_flaechen	1450_top_linien
Anzahl der Objekte	25	2
Größe im Dateisystem	36 KB	19 KB
Geometrie	Polygon/MultiPolygon	MultiLineString
Felder (TAB-Datei)	Domäne	
Type Char (100)	NOT NULL, WaterBody, Pond, Creek, River, OxbowLake, Garden, GreenArea, Graveyard, Earthwall, Hedgelaying, WolfPit	NOT NULL, CityWall, Moat
Name Char (100)	NOT NULL, Zeichenkette	
Beschreibung Char (249) ;	NOT NULL, Zeichenkette	
Jahr Integer ;	NOT NULL, vierstellige Ganzzahl	

Tab. 5.4: Gemeinsame Struktur und individuelle Eigenschaften der angepassten Relationen für 1450

Relationsname	1620_top_flaechen
Anzahl der Objekte	44
Größe im Dateisystem	76 KB
Geometrie	Polygon/MultiPolygon
Felder (TAB-Datei)	Domäne
Type (100) ;	NOT NULL, Bridge, Well, Earthwall, GreenArea, Garden, PleasureGarden, Graveyard, HerbGarden, Vineyard, CityWall, Moat, WaterBody, Pond, Creek, River, OxbowLake, Gate, BastionWall
Name Char (100)	NOT NULL, Zeichenkette
Beschreibung Char (249) ;	NOT NULL, Zeichenkette
Jahr Integer ;	NOT NULL, vierstellige Ganzzahl

Tab. 5.5: Relationsstruktur und Eigenschaften der übrigen Objekte für 1620

Die angepassten Datenbestände sind nachfolgend in vektorieller Form dargestellt. Die Klassifikation wurde im Vergleich zum externen Modell geringfügig angepasst. Sie orientiert sich immer noch an den Ausgangsdaten.

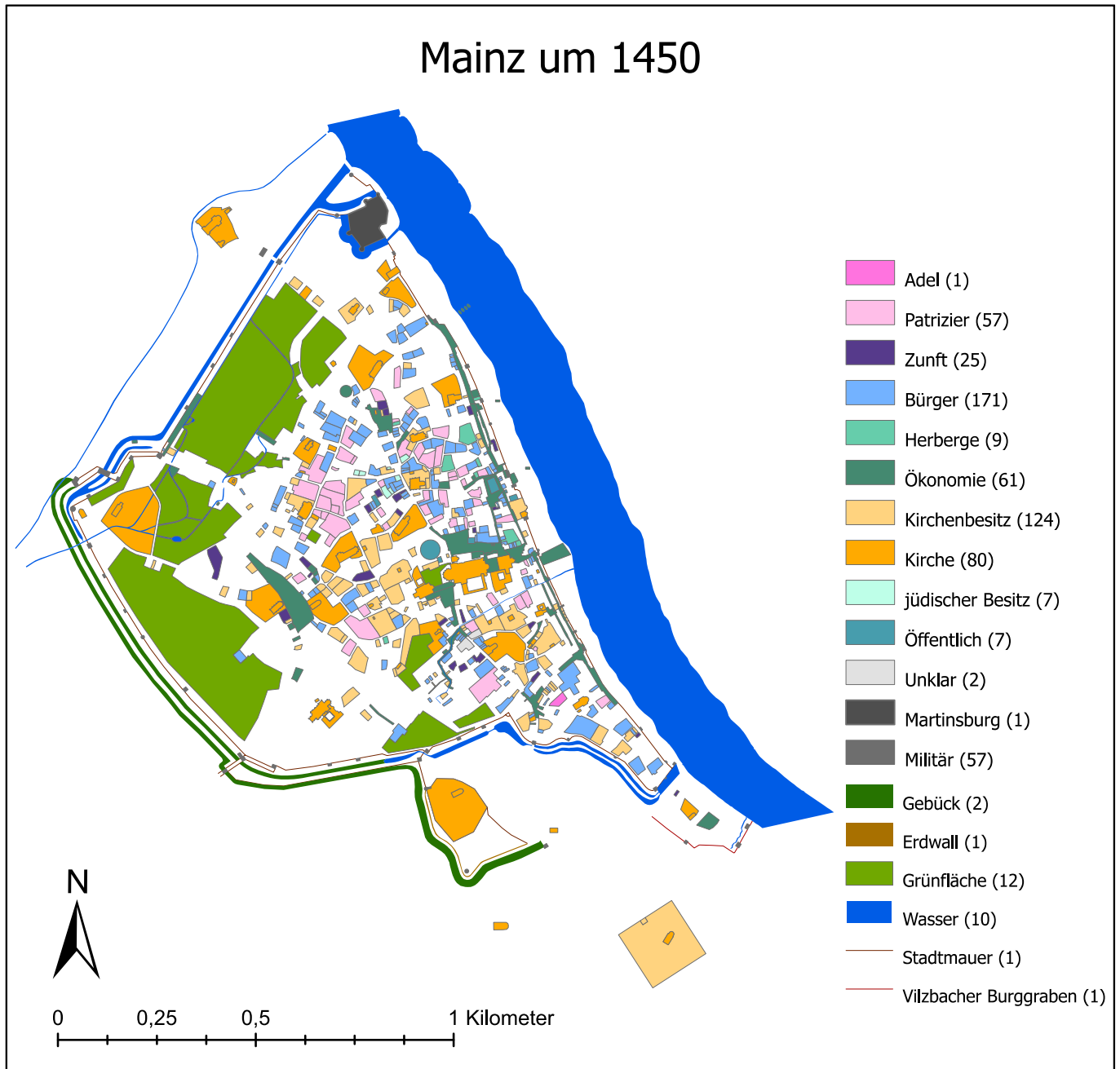


Abb. 5.5: SVG-Export des ArcGIS Pro-Layouts (1450)

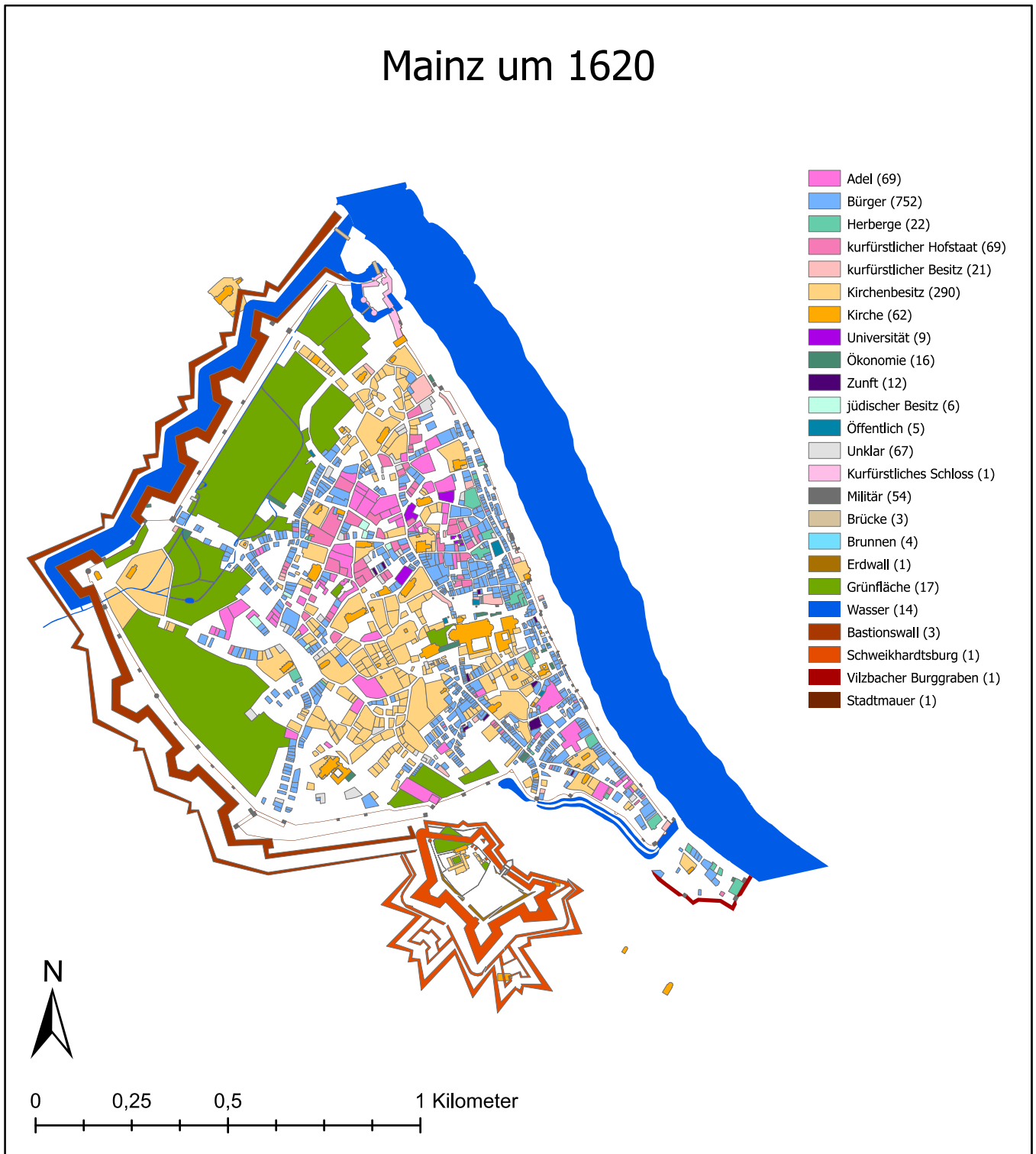


Abb. 5.6: SVG-Export des ArcGIS Pro-Layouts (1620)

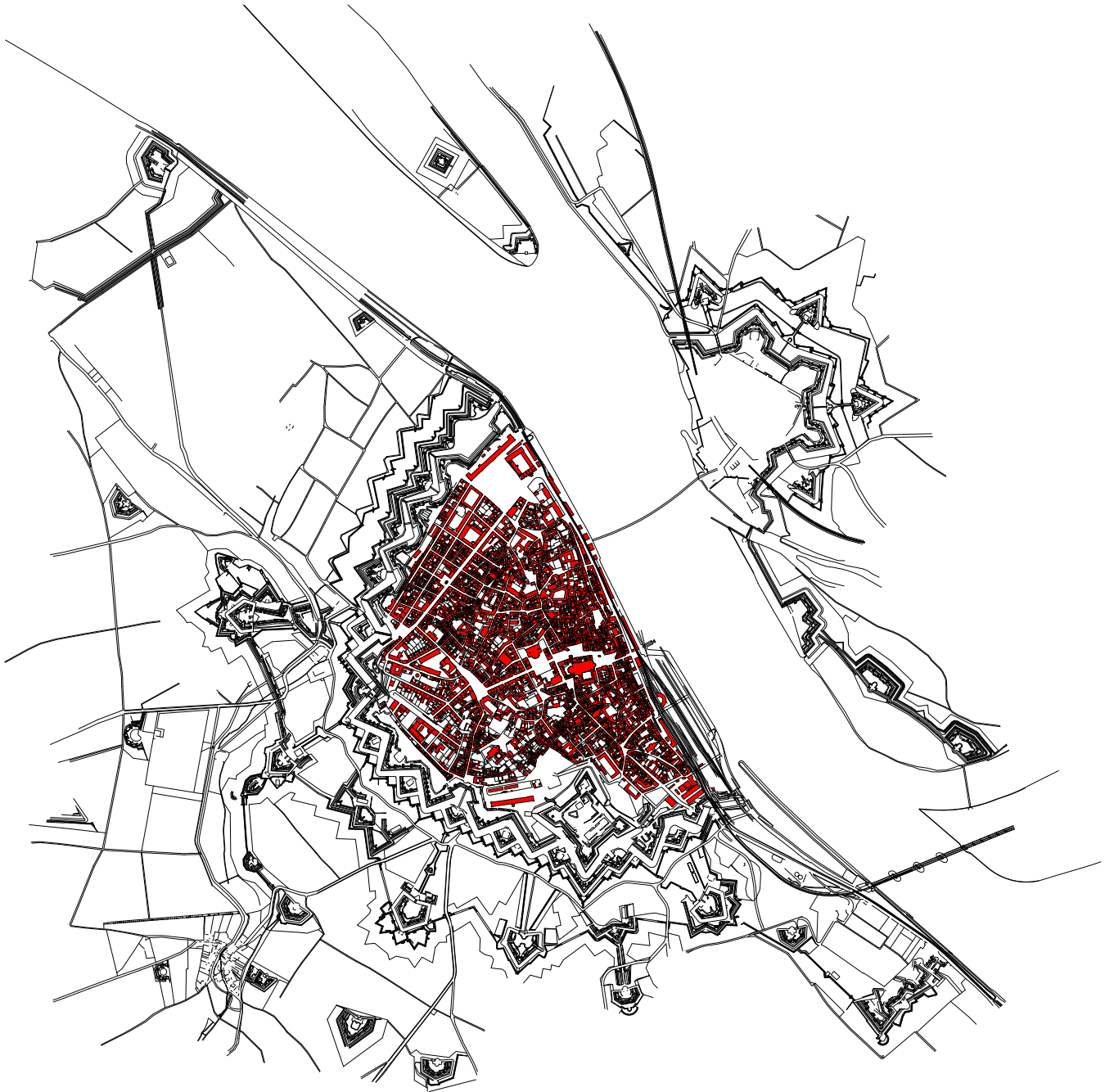


Abb. 5.7: SVG-Darstellung, generiert über FME (1871)

Für das Vektormodell der Stadt Mainz für die Zeit um 1871 wurde kein Layout in ArcGIS Pro erstellt, sondern die obige SVG-Darstellung wurde direkt aus den MapInfo-Relationen über FME Desktop generiert. Dabei wurden ausgewählte Häuser rot eingefärbt, um sie von den übrigen Objekten abgrenzen zu können. Diese können noch nicht berücksichtigt werden. Allein für die Festung müssen aus 15375 unzusammenhängenden Liniensegmenten Geobjekte resp. Informationsobjekte konstruiert werden. Dies kann in jedem beliebigen GIS nachvollzogen werden, das den Import von SVG erlaubt, da die obige Vektorgrafik einfach aus diesem PDF extrahiert werden kann (z.B. mithilfe von [Inkscape](#)).

5.1.2 Auswahl der Individuen zwecks Verlinkung

Der Informationsumfang des Modells soll durch die Generierung von RDF-Links auf der Instanzebene⁷⁹ erhöht werden, was anschaulich in Abschnitt 2.5 erklärt wurde. Für die Verlinkung mit „externen Ressourcen“ werden die folgenden Individuen (Instanzen) aus den MapInfo-Relationen ausgewählt:

- (1) Sämtliche topographische Objekte, deren Typen in Tab. 5.1, in Tab. 5.4 sowie in Tab. 5.5 angegeben wurden. Diese Geoobjekte verfügen in den ausgewählten bzw. angepassten Relationen über assoziierte Koordinaten. Es sind Gewässer wie Seen, Weiher, Bäche, Flüsse, Altarme; Grünflächen wie Friedhöfe, Gärten, Wein-, Kräuter- und Lustgärten; Infrastrukturobjekte wie Brücken, Brunnen; Befestigungen wie Barrieren (Gräben, Wälle, Tore) und Türme, die auch Gebäude sein können, ebenso wie Kirchen, Herbergen, Burgen, Paläste. Insgesamt sind es 7802 Geoobjekte. Im Anhang sind viele Beispiele enthalten.
- (2) Alle natürlichen und juristischen Personen, die im historischen Häuserbuch als Eigentümerinnen oder Bewohnerinnen genannt werden. Personen sind prinzipiell auch Geoobjekte. Allerdings sind sie nicht über Koordinaten lokalisiert, sondern in den Ausgangsdaten wird über Zeichenketten (Namen) auf die Personen referiert. Die Namen bzw. Zeichenketten sind teilweise doppelt und/oder in unterschiedlicher Form vorhanden (Schreibweise, Namenszusätze usw.), z.B. „Dr. Jacobus Campius“ und „Dr. Campius“. Die Zeichenketten entsprechen in ihrer Gesamtheit keinem distinktiven Muster. Das soll heißen, dass ein Muster erkennbar ist, das z.B. mithilfe eines regulären Ausdrucks beschrieben werden kann. Wirklich zielführend ist dies aber nicht, da lediglich ausgedrückt werden kann, dass alle Klein- und Großbuchstaben in Kombination mit Ziffern etc. vorkommen können. Es kann aber nicht ausgedrückt werden, welche Zeichenfolge den Vornamen, Nachnamen usw. darstellen soll, weshalb die Zeichenkette auch nicht automatisch in sinnvolle Komponenten zerlegt werden kann. Die Zeichenketten qualifizieren sich nicht als eindeutige Referenzen auf die jeweiligen Personen. Bestenfalls kann die Person aufgrund der vorhandenen Kontextinformationen (die Beziehung zu Häusern, der Zeitpunkt, weitere Literale wie die Bauinformationen etc.) identifiziert werden. Schlimmstenfalls müssen dafür weitergehende Recherchen angestellt werden. Insgesamt sind es 10.596 Personen. Im Anhang sind viele Beispiele enthalten.
- (3) Alle Publikationen, für welche die vorstehenden Ausführungen bedingt gelten. Auch auf die Publikationen (die prinzipiell ebenfalls Geoobjekte sind), wird in den Ausgangsdaten über Zeichenketten referiert, welche jedoch die Identifikation der jeweiligen Publikation ermöglichen und die ein distinktives Muster aufweisen, d.h. sie können automatisch in sinnvolle Komponenten zerlegt werden. Insgesamt sind es 32 Publikationen, die vollständig im Anhang gelistet sind.

5.2 LOD-Cloud

Der Zielerreichungsgrad soll möglichst hoch sein; es sollen möglichst viele der o.g. Individuen mit Ressourcen verknüpft werden, die in der LOD-Cloud vorhanden sind bzw. beschrieben werden (= „externe Ressourcen“). Über die Jahrhunderte haben in Mainz viele wichtige Personen gelebt oder gewirkt und es sind bedeutende Bauwerke entstanden. Informationen zu diesen Entitäten finden sich nicht nur im Mainzer Stadtarchiv, sondern in zahlreichen Wissensbasen (= ABox + TBox), die über das Web (of Data) frei zugänglich sind. Die folgende Tabelle listet ausgewählte LOD-Projekte, die adäquate Wissensbasen hervorbrachten.

⁷⁹ Der Begriff wird im folgenden Kapitel erklärt und definiert.

Projekt	Weitere Informationen
GeoNames	http://www.geonames.org/
LinkedGeoData	http://linkedgeodata.org/
DBpedia	http://www.dbpedia.org/
Linked-Data-Service der DNB	https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/LDS/lds_node.html

Tab. 5.6: Infrage kommende Wissensbasen für die Verlinkung von Individuen

Statt sämtliche der o.g. Projekte zu besprechen, wird dargelegt, warum die Informationen der DNB (Titeldaten, GND) für die Generierung von RDF-Links auf der Instanzebene ausgewählt werden. Diese Informationen sind dafür prädestiniert weil:

- die Linked-Data-Dienste der DNB ebenso gut dokumentiert sind, wie deren ABox und TBox⁸⁰,
- die angebotenen Informationen qualitativ hochwertig und umfassend sind (s. GP 8.2.4 und Anhang),
- auch Informationen zu älteren Publikationen sowie zu historischen Personen, Körperschaften, Geografika etc. enthalten sind (s. GP 8.2.4 und Anhang),
- die DNB ihren Beschreibungen RDF-Links hinzugefügt hat (z.B. zu Wikidata und GeoNames),
- Konzepte aus etablierten, kontrollierten Vokabularen verwendet werden (auch GeoSPARQL)⁸¹,
- die Differenzierung zwischen NIR- / 303-URIs bzw. die „Inhaltsvereinbarung“⁸² anwenderinnenfreundlich erklärt wird – ein Punkt, der in Abschnitt 6.1 aufgegriffen wird,
- strukturierte Informationen angeboten werden, die heruntergeladen werden können⁸²,
- die Nutzerinnen über Änderungen informiert werden⁸³,
- verschiedene Schnittstellen verfügbar sind, die nach Anmeldung kostenfrei genutzt werden können.

Erwähnenswert ist, dass die DNB kein SPARQL-Endpunkt anbietet⁸⁴. Zwecks Verlinkung muss der aktuelle Grundbestand (s. Tab. 8.11) heruntergeladen und in einer eigenen Datenbank verwaltet werden, sodass ein „Mehrwert-Service (value added service)“ (Diederichs and Wuttke, 2018, p. 16) realisiert werden kann. Man könnte argumentieren, dass dies im LOD-Konzept nicht vorgesehen ist (Stichwort: Dezentralität). Es gibt aber die Möglichkeit einen regelmäßigen Abgleich vorzunehmen, was unter GP 7.1.2 angesprochen wird. Der fehlende SPARQL-Endpunkt wird nicht als Ausschlusskriterium oder als Nachteil angesehen; die qualitativ hochwertigen Informationen der DNB werden verwendet. Genauer gesagt werden nur diese Informationen für die Generierung von RDF-Links auf der Instanzebene verwendet. Wenn zusätzlich auf die o.g. Wissensbasen aus den anderen Projekten zurückgegriffen wird, erhöht sich die Wahrscheinlichkeit, dass mehr Ressourcen aus dem historischen Häuserbuch verlinkt werden können. Davon wird abgesehen, weil die Analyse der ABox und der TBox ein extrem zeitaufwendiger Prozess ist. Ebenso wird in diesem Kapitel davon abgesehen, die nächsten Seiten mit Erläuterungen zur Datenanalyse zu füllen. Dies wird unter GP 8.2.4 erörtert. Dort wird auch erklärt, wieso die Bestimmung der Kandidatenmenge und der Ergebnismenge nicht trivial ist. Außerdem findet sich dort eine SPARQL-Abfrage, die so formuliert wurde, dass sie neben der Selektion auch für die Erläuterung

⁸⁰ s.: <https://d-nb.info/standards/elementset/gnd>

⁸¹ s.: https://www.dnb.de/DE/Professionell/Metadatendienste/Exportformate/RDF-Vokabulare/rdf_node.html

⁸² s.: <https://data.dnb.de/opendata/>

⁸³ s.: <https://lists.dnb.de/mailman/listinfo/lds>

⁸⁴ Stand: November 2020

verwendet werden kann. Es wird detailliert erklärt, was alles untersucht wurde (Optionalitäten, Muster und Strukturen relevanter Zeichenketten, Datentypen, Bedeutung von Termini uvm.). Für weitere Informationen zum Linked-Data-Service der DNB wird auf die o.g. URIs verwiesen. Die Grundlagen von LOD in der Bibliothekswelt wurden von Pohl and Danowski (2013) anschaulich dargelegt.

Die obigen Ausführungen beziehen sich lediglich auf den letzten Teilschritt im Workflow zwecks Generierung des RDF-Graphen (s. Abschnitt 7.2). Daneben ist ein Konzeptualisierungs- und Transformationsprozess vorgesehen. Zwecks Beschreibung der Beziehungen zwischen Individuen, für die Spezialisierung von Properties uvm. können Konzepte aus existierenden Vokabularen nachgenutzt werden. Die folgende Tabelle enthält die zweckmäßige Auswahl:

Vokabular / Ontologie / TBox...	Weitere Informationen
Vocabulary of Interlinked Datasets (voID)	https://www.w3.org/TR/void/
Dublin Core Metadata Initiative (DCMI) Metadata Terms	http://purl.org/dc/terms/
Friend of a Friend (FOAF)	http://xmlns.com/foaf/spec/
Ontology of agent relationships (AgRelOn)	https://d-nb.info/standards/elementset/agrelon
Bibliographic Ontology (BIBO)	http://purl.org/ontology/bibo/
GeoSPARQL	http://www.opengis.net/ont/geosparql
DBpedia Ontologie	https://www.dbpedia.org/resources/ontology/
IRW Ontology	http://www.ontologydesignpatterns.org/ont/web/irw.owl
GND Ontology	https://d-nb.info/standards/elementset/gnd
Creative Commons (CC)	http://creativecommons.org/ns
CIDOC CRM i.V. CRMgeo	https://www.cidoc-crm.org/

Tab. 5.7: Infrage kommende Vokabulare für den Konzeptualisierungs- und Transformationsprozess

Die Rekapitulation der Inhalte aus den Semantic Web Dokumenten bringt keinen Mehrwert. Im Anhang sind alle Semantic Web URIs enthalten, die nachgenutzt wurden. Man kann an die entsprechende Stelle im Dokument „springen“ (#Doppelkreuz). Abgesehen davon wird bei der Konzeption des Workflows (s. GP 7.2.3) auch noch begründet, wieso bestimmte Konzepte verwendet werden. Deshalb wird hier darauf verzichtet.

6 Anforderungsanalyse und -definition

In diesem Kapitel werden zuerst die LOD-Anforderungen herausgearbeitet und definiert, denen der zu generierende RDF-Graph genügen muss oder soll. Bezüglich der LOD-Anforderungen wird differenziert in **allgemeine Anforderungen** (a) und **individuelle Anforderungen** (b).

- (a) Erstere werden aus einer möglichst *objektiven Perspektive* definiert, d.h., die Anforderungen werden von Prinzipien und Empfehlungen abgeleitet, die domänenübergreifend anerkannt sind. Als Hauptquelle wird die grundlegende Arbeit von Heath and Bizer (2011) herangezogen. Die vorgenannten Autoren analysierten die „Linked Data Design Issues“ (Berners-Lee, 2006), entwickelten die Ideen weiter und sprachen verständliche und richtungweisende Empfehlungen aus, die in vielen erfolgreichen LOD-Projekten Beachtung fanden (Allemang et al., 2020). Insofern ist diese Arbeit als primäre Quelle für die Festlegung der allgemeinen LOD-Anforderungen bestens geeignet. Außerdem wird die URI-Strategie von van den Brink et al. (2014) berücksichtigt, da diese – wie bereits unter GP 3.1.2 erwähnt – den Empfehlungen von Heath and Bizer (2011) nicht widerspricht. Dies gilt nicht für die Vorschläge von Tandy et al. (2017), die nachfolgend diskutiert werden. Im Zuge der Anforderungsanalyse muss auch dargelegt werden, wieso die Definition des (essenziellen) Terminus „Feature“ aus der Arbeit von Tandy et al. (2017) nicht berücksichtigt wird und warum den Empfehlungen von Heath and Bizer (2011) zur URI-Allokation bzw. zwecks Vermeidung von URI-Kollisionen der Vorzug gegeben wird.
- (b) Die Spezifizierung der individuellen Anforderungen basiert auf den *subjektiven Erkenntnissen*, die während der Erarbeitung der theoretischen Grundlagen gewonnen wurden. In Abschnitt 2.2 wurde der „Identitätskrise des Semantic Web“ besondere Aufmerksamkeit geschenkt. Diese manifestiert sich als Hürde für die Expansion der LOD-Cloud, die wegen des terminologischen Chaos und einigen Paradoxa aufwendig genommen werden muss. Zwar hoben auch Heath und Bizer (2011) mehrfach hervor, dass die Verwendung von NIR-URIs und 303-URI essentiell für die Kohärenz des Web of Data ist. Allerdings finden sich Lösungsmöglichkeiten für die Vermeidung von Missverständnissen bzw. Mehrdeutigkeiten auch in anderen Arbeiten. Diesbezüglich sind die bereits mehrfach zitierten Arbeiten von Presutti and Gangemi (2008) und Halpin and Presutti (2009) zu nennen, die zur Formulierung der individuellen Anforderungen ebenfalls herangezogen werden. Die individuellen Anforderungen werden definiert, um sowohl die Aussagen als auch die Serialisierung möglichst einfach und verständlich zu halten. Dadurch erhöht sich die Wahrscheinlichkeit, dass das Arbeitsergebnis von möglichst vielen Anwenderinnen nachgenutzt werden kann. Möglicherweise kann die IQ in unvorherzusehenden Projekten weiter verbessert werden.

Die LOD-Anforderungen werden durch **anwendungsfallspezifische Anforderungen** ergänzt. Als Grundlage hierfür werden die Erkenntnisse verwendet, welche unter GP 5.1.1 zusammengetragen wurden.

Sämtliche Anforderungen werden aus der Sicht einer Datenbereitstellerin definiert. Aus dieser Perspektive erfolgt auch die Überprüfung. Die jeweiligen Verbindlichkeiten können den textuellen Beschreibungen entnommen werden:

- Optionale Anforderungen werden mithilfe des Schlüsselworts „*soll*“ definiert.
- Für die Spezifikation obligatorischer Anforderungen wird das Schlüsselwort „*muss*“ verwendet.

Das Kapitel schließt mit der Vorstellung eines **fiktiven Anwendungsszenarios**. Dieses wird aus Anwenderinnensicht formuliert. Es wird in Abschnitt 9.2 für die Demonstration verwendet.

6.1 Diskussion der SDWWG-Empfehlung

In der Arbeit von Heath and Bizer (2011) namens „Linked Data: Evolving the Web into a Global Data Space“ und in der von Tandy et al. (2017) verfassten SDWWG-Empfehlung namens „Spatial Data on the Web Best Practices“ wurden konträre Positionen zu grundsätzlichen Themen vertreten⁸⁵. Unterschiedliche Standpunkte und deren Diskussion sind wichtig für die Weiterentwicklung des Webs, das auch ein soziales System ist. In der vorliegenden Arbeit können nicht alle Punkte ausdiskutiert werden. Für die Anforderungsdefinition und die spätere Umsetzung muss aber zumindest entschieden werden, welchen Richtlinien zwecks URI-Allokation und welchen Empfehlungen zur Vermeidung von URI-Kollisionen gefolgt werden soll. Abgesehen davon muss noch der Begriff „Feature“ erörtert werden, da dieser für die Anforderungsdefinition ebenfalls erforderlich ist.

6.1.1 URI-Kollision

Um Klarheit zu schaffen und um Mehrdeutigkeiten zu vermeiden, wurden in der ersten Hälfte des theoretischen Teils wichtige Begriffe erklärt und Mechanismen sowie Strategien vorgestellt, die es einem Agenten ermöglichen, trotz der prinzipiellen URI-Undurchsichtigkeit, im Zuge der (indirekten) Dereferenzierung einer HTTP-URI, Informationen zum Typ der identifizierten Ressource zu beziehen. In der vorliegenden Arbeit können die Vorschläge von Tandy et al. (2017) damit nicht in Einklang gebracht werden. Außerdem wird die Auffassung vertreten, dass die spezielle SDWWG-Empfehlung für die Integration von räumlichen Informationen ins Web nicht nur sich selbst widerspricht, sondern auch der generellen W3C-Empfehlung namens „Data on the Web Best Practices“ (Lóscio et al., 2017) und den generellen Empfehlungen von Heath and Bizer (2011).

Dies kann hier zwar nicht vollumfänglich eruiert, aber zumindest exemplarisch dargelegt werden: „Following the best practices presented below, we use a HTTP URI to *unambiguously* identify Eddystone Lighthouse: <https://www.trinityhouse.co.uk/lighthouses-and-lightvessels/eddy-stone-lighthouse>“ (Tandy et al., 2017).

Tatsächlich identifiziert die o.g. URI **nicht** das realweltliche Objekt – den Leuchtturm – sondern ein digitales Dokument (eine zugehörige Beschreibung), welche(s) Informationen darüber enthält. Wenn man die Lösungsvorschläge der TAG (httpRange-14) annehmen möchte, kann man dies auch nachvollziehen, denn die obige URI kann direkt dereferenziert werden und die Antwort des Servers enthält u. a. den HTTP-Statuscode 200.

Statt darauf hinzuweisen, präsentierten Tandy et al. (2017) danach weitere URIs. „In fact, there are many URIs in use **for** Eddystone Lighthouse. The one above is provided by the owners/operators of the lighthouse; others include <https://www.wikidata.org/wiki/Q546122> from Wikidata, http://dbpedia.org/resource/Eddystone_Lighthouse from DBpedia and <http://d-nb.info/gnd/1067162240> from Deutsche Nationalbibliothek“.

Zwei der drei o.g. URIs, nämlich die der DNB und von DBpedia, qualifizieren sich jedoch als **Semantic Web URIs**, die **wirklich** der Referenz auf den (nicht digitalen) Leuchtturm sowie dem indirekten Zugang zu (digitalen) Informationen über den Leuchtturm dienen. Trotzdem kommt es dabei nie zu einer URI-Kollision, denn die o.g. URIs der DNB und von DBpedia können nur indirekt dereferenziert werden; der entsprechende Server liefert dem anfragenden Client im Rahmen der Weiterleitung / mehrerer Weiterleitungen eine andere URI, die manchmal eine relevante (generische) Informationsressource, manchmal auch eine (konkrete) Web-Repräsentation (z.B. ein RDF- oder HTML-Dokument) derselbigen identifiziert. Dies soll mit der folgenden Abbildung nochmals klargestellt werden, wofür der Korrektheit halber dieselbe URI von DBpedia⁸⁶ verwendet wurde:

⁸⁵ Anhand des Literaturverzeichnisses kann nachvollzogen werden, dass die HTML-Repräsentationen der o.g. Arbeiten untersucht wurden, die keine Seitenzahlen aufweisen.

⁸⁶ D.h. http statt https.

```

Anfrage 1 - # wget -S --header="Accept: text/html" http://dbpedia.org/resource/Eddystone_Lighthouse 2>&l | egrep '^ *HTTP|^ *Location'
HTTP/1.1 303 See Other
Location: https://dbpedia.org/resource/Eddystone_Lighthouse
Location: https://dbpedia.org/resource/Eddystone_Lighthouse [following]
HTTP/1.1 303 See Other
Location: http://dbpedia.org/page/Eddystone_Lighthouse
Location: http://dbpedia.org/page/Eddystone_Lighthouse [following]
HTTP/1.1 303 See Other
Location: https://dbpedia.org/page/Eddystone_Lighthouse
Location: https://dbpedia.org/page/Eddystone_Lighthouse [following]
HTTP/1.1 200 OK
Anfrage 2 - # wget -S --header="Accept: application/rdf+xml" http://dbpedia.org/resource/Eddystone_Lighthouse 2>&l | egrep '^ *HTTP|^ *Location'
HTTP/1.1 303 See Other
Location: https://dbpedia.org/resource/Eddystone_Lighthouse
Location: https://dbpedia.org/resource/Eddystone_Lighthouse [following]
HTTP/1.1 303 See Other
Location: http://dbpedia.org/data/Eddystone_Lighthouse.xml
Location: http://dbpedia.org/data/Eddystone_Lighthouse.xml [following]
HTTP/1.1 303 See Other
Location: https://dbpedia.org/data/Eddystone_Lighthouse.xml
Location: https://dbpedia.org/data/Eddystone_Lighthouse.xml [following]
HTTP/1.1 200 OK

```

Abb. 6.1: Exemplarische Dereferenzierung, dargestellt von PuTTY

In der ersten Anfrage wird die menschenlesbare HTML-Repräsentation präferiert; die 303-URI, die vom „DBpedia-Server“ mitgeteilt wird, enthält die Zeichenfolge „page“. In der zweiten Anfrage wird die maschinenverarbeitbare RDF-Repräsentation in der RDF/XML-Serialisierung präferiert (s. „data“). Die Semantic Web URI (s. „resource“) ist ein Bezeichner der analogen Nicht-Informationsressource – mit den Worten von Tandy et al. (2017) ausgedrückt: „for Eddystone Lighthouse“. Für die o.g. URI von Wikidata gilt dies nicht. Diese kann aber durch <https://www.wikidata.org/entity/Q546122> ersetzt werden, wenn man ausdrücken möchte, dass es eine NIR-URI für den Leuchtturm gibt. Die Systematik von Wikidata dieselbe. Die o.g. URI von Wikidata ist nicht dieselbe (s. „wiki“); der Client bekommt sie im Rahmen der Weiterleitung mitgeteilt.

Eigentlich können die obigen Aussagen nur von den Personen getroffen werden, welche die Ressourcen im Zuge der URI-Allokation (in ihrem Namensraum) identifizierten. Allerdings haben sämtliche der o.g. Akteure erklärt, welchen Sinn und Zweck ihre URIs haben⁸⁷. Die o.g. NIR-URIs der DNB und von DBpedia referieren wirklich auf den Leuchtturm. Im Gegensatz dazu kann durchaus in Frage gestellt werden, ob die URI, die Tandy et al. (2017) verwendeten, den Leuchtturm eindeutig identifiziert.

Ebenso ist fraglich, was aus einer Aussage gefolgert werden kann, in der diese URI verwendet wird. „The Spatial Thing, Eddystone Lighthouse (<https://www.trinityhouse.co.uk/lighthouses-and-lightvessels/eddytone-lighthouse>), is the “real world phenomenon” about which we want to state facts (such as the height of its light is 41 meters above sea level) and **link to other real world phenomena** (for example, that it is located at Eddystone Rocks, Cornwall; another Spatial Thing identified as <http://sws.geonames.org/2650253/> by GeoNames)“ (Tandy et al., 2017). Wenn diese Aussagen in (RDF-)Tripelform gemacht werden, wenn die obige URI an Subjektstelle steht, dann sind die Aussagen zumindest mehrdeutig: Es kann auch gefolgert werden, dass es ein Dokument gibt, dessen Licht 41 Meter über dem Meeresspiegel liegt und dass das Dokument auf den Eddystone Rocks liegt. Die o.g. URI aus dem GeoNames-Projekt ist eine **Semantic Web URI**, die tatsächlich auf die Eddystone Rocks referiert.

In der generellen W3C-Empfehlung namens „Data on the Web Best Practices“ (Lóscio et al., 2017), genauer gesagt zum „Best Practice 10“ findet sich die folgende Aussage: „Dereferencing URIs for non-information resources should result in an HTTP 303 redirect to a similar URL such as <http://data.mycity.example.com/transport/route/bus/doc/AB> that describes, i.e. gives information about, the AB bus route [...]“

⁸⁷ s.: Beschreibung des Linked-Data-Services der DNB, Version 1.1

Heath and Bizer (2011) empfehlen: „Where URIs identify real-world objects, it is **essential** to not confuse the objects themselves with the Web documents that describe them. It is, therefore, **common practice** to use different URIs to identify the real-world object and the document that describes it, in order to be **unambiguous**. This practice allows **separate statements** to be made about an object and about a document that describes that object. For example, the creation date of a person may be rather different to the creation date of a document that describes this person. Being able to distinguish the two through use of different URIs is **critical to the coherence** of the Web of Data“.

Obwohl Tandy et al. (2017) die grundlegenden Richtlinien aus der o.g. W3C-Empfehlung nicht verändern, sondern lediglich erweitern wollten, um die Besonderheiten von Geoinformationen berücksichtigen zu können, widersprachen sie der W3C-Empfehlung explizit: „However, in most cases using a single URI for both Spatial Thing and the page/document is simpler to implement and meets the expectations of most end-users. As stated in [WEBARCH] section 2.2.3 Indirect Identification, identifiers are commonly used in this way. There is no obligation to distinguish between the Spatial Thing and the page/document unless your application requires this. While there is a cost to this conflation, problems can be mitigated by avoiding making statements that confuse Spatial Thing and the page/document [...]“.

In dieser Arbeit wird die Auffassung vertreten, dass die obige Auslegung von §2.2.3 AWWW (= „WEBARCH“) in diesem Kontext fragwürdig ist. Außerdem wird der große Vorteil von HTTP-URIs in einem dezentralen System auch darin gesehen, dass es nicht nur undurchsichtige Zeichenketten sind, nicht nur nicht eindeutige Namen, sondern dass ein Agent im Zuge der (indirekten) Dereferenzierung – einfach gesagt – „ermitteln kann, was sich hinter der URI verbirgt“. Es wurde anschaulich dargelegt, wie schnell unbeabsichtigte Mehrdeutigkeiten entstehen können, welche die Kohärenz gefährden. Des Weiteren wird die Auffassung vertreten, dass dies in vielen Fällen gar nicht vermieden werden kann, wenn auf ein realweltliches Bezugsobjekt und die zugehörige Beschreibung über dieselbe URI referiert wird. Im LOD-Kontext sollen Konzepte (z.B. Prädikate) aus etablierten Vokabularen nachgenutzt werden. Wenn man den Vorschlägen von Tandy et al. (2017) folgen möchte und Mehrdeutigkeiten wirklich vermeiden will, dann werden die Verwendungsmöglichkeiten stark eingeschränkt; der AAA-Slogan gilt nicht mehr, was anhand eines fiktiven Beispiels deutlich wird:

- Das realweltliche Objekt (ein Haus) und die zugehörige Beschreibung (ein HTML-Dokument) werden über eine URI identifiziert, die im RDF-Tripel an Subjektstelle stehen soll.
- Mittels des etablierten „Foo-Vokabulars“ bzw. mithilfe des Prädikats `foo:architect` (Definitionsbereich: realweltliches Geografikum, Wertebereich: Person) soll gesagt werden, dass das Haus von einem Architekten entworfen/gestaltet wurde; mithilfe des Prädikats `foo:webDesigner` (Definitionsbereich: Web-Repräsentation, Wertebereich: Person) soll gesagt werden, dass die zugehörige Beschreibung von einem Webdesigner entworfen/gestaltet wurde – beides sind sinnvolle Aussagen.
- Dies führt aber zwangsläufig zur Mehrdeutigkeit, denn das Subjekt wird – einfach gesagt – implizit als realweltliches Geografikum und als Web-Repräsentation getypt – keine sinnvollen Aussagen.

Wenn das Haus und die zugehörige Beschreibung über unterschiedliche URIs identifiziert werden, können die sinnvollen Aussagen problemlos getroffen werden, statt auf eine verzichten zu müssen. Deshalb wird Heath and Bizer (2011) gefolgt.

6.1.2 Feature

In Abschnitt 2.3.3 sollte eigentlich der Versuch unternommen werden, mithilfe der SDWWG-Empfehlung Klarheit hinsichtlich des Begriffs „Feature“ zu schaffen, statt lediglich die knappe Definition aus ISO 19101 – „A feature is an *abstraction* of a *real world phenomenon*“ (ISO/TC 211, 2002) – zu bemühen. Wie Tandy et al. (2017) zu Recht konstatierten, ist die o.g. **Definition des Terminus „Feature“** verwirrend. Sie klärten zuerst, dass unter dem Begriff „*realweltliches Phänomen*“ greifbare, reale Objekte wie z.B. eine Straße sowie abstrakte Konzepte mit räumlicher Ausdehnung wie z.B. ein Postleitzahlbezirk und fiktionale Lokationen, die also eine mehr oder weniger genaue Georeferenz aufweisen (z.B. Dickens' London) aber auch Entitäten, für die fragliche Lokalisierungshypothesen aufgestellt wurden (z.B. zu Atlantis), subsumiert werden können (Tandy et al., 2017). Alle vorgenannten Entitäten existieren jedenfalls (ggf. aufgrund menschlicher Vorstellungskraft in einer „Ideenwelt“) außerhalb des Webs und erfüllen die Kriterien der Definition der **Nicht-Informationsressource**. Danach fokussierten Tandy et al. (2017) die *Abstraktion*. Wie Bill (2016) hervorhob, ist die exakte Abbildung der Realität bei der Konstruktion eines Geodatenmodells regelmäßig kein erstrebenswertes Ziel, weshalb die Komplexität realweltlicher Phänomene im Zuge der Abstraktion adäquat reduziert wird. Ein Feature wird dementsprechend i.d.R. Attribute aufweisen, die im UoD von Interesse sind, aber nicht alle Eigenschaften des realweltlichen Objekts. Damit könnte man argumentieren, dass das Feature auch die Kriterien der Definition der zugehörigen Beschreibung (engl. **associated description**) erfüllt. Statt Klarheit zu schaffen, verschmolzen Tandy et al. (2017) das realweltliche Phänomen und die korrespondierende Abstraktion, indem sie vorerst nicht mehr den Terminus Feature verwendeten, sondern das Konzept `w3cgeo:SpatialThing`⁸⁸ – eine Klasse die im Basic Geo (WGS84 lat/long) Vokabular spezifiziert wurde (Brickley, 2003). Wirklich geeignet sei die vorgenannte formale Definition der Klasse laut Tandy et al. (2017) aber auch nicht, denn im Gegensatz zur Klassendefinition im Geo-Vokabular des GeoSPARQL-Standards, wird nicht explizit ausgedrückt, dass das Feature und die Geometrie disjunkte Klassen sind, sondern die Klasse `w3cgeo:Point` ist eine Unterklasse von `w3cgeo:SpatialThing`. Das „Spatial Thing“ von Tandy et al. (2017) soll deshalb nicht als Äquivalent zu `w3cgeo:SpatialThing` angesehen werden, sondern es soll der Definition von `geo:Feature`⁸⁹ entsprechen. Im Semantic Web Dokument (GeoSPARQL Version 1.0.1) findet sich an der Stelle, wo die Klasse `geo:Feature` definiert wurde, der natürlichsprachliche Hinweis, dass die vorgenannte Klasse ein Äquivalent zu `GFI_Feature` sein soll, das in ISO 19156:2011 definiert wurde (Perry and Herring, 2012), wo zwecks Definition des Terminus „Feature“ wieder auf die o.g. Definition aus ISO 19101 verwiesen wird (Cox, 2010). Damit schließt sich der Kreis. Klarheit wurde allerdings nicht geschaffen, was die nachfolgenden Bedeutungserklärungen von Tandy et al. (2017) belegen:

- „Feature: Abstraction of real world phenomena. A **digital representation** of a real world entity or an **abstraction** of the real world. **Examples of features** include almost anything that can be placed in time and space, including desks, buildings, [...] and so on“.
- „Spatial thing: Anything with spatial extent, (i.e. size, shape, or position) and is a **combination** of the **real-world phenomenon** and its **abstraction (the feature)**. Examples are: people, places, or bowling balls“.
- „a Spatial Thing is actually a real or conceptual phenomenon - a lake is made from water **not** information“.

⁸⁸ `w3cgeo` = "http://www.w3.org/2003/01/geo/wgs84_pos#"

⁸⁹ `geo` = "<http://www.opengis.net/ont/geosparql#>"

In der vorliegenden Arbeit konnte nicht festgestellt werden, dass sich die Vereinigung von „realweltlichem Phänomen“ mit der korrespondierenden „Abstraktion“ eindeutig aus dem Semantic Web Dokument ergibt, das Beschreibungen der Konzepte des Basic Geo (WGS84 lat/long) Vokabulars enthält (siehe `w3cgeo:SpatialThing`). Es wurde aber festgestellt, dass der Definitionsbereich des Prädikats `geo:hasGeometry` mit `geo:Feature` und der Wertebereich mit `geo:Geometry` angegeben ist (s. GP 2.4.2). Die einfache Aussage⁹⁰, dass das „Feature“ in dieser Arbeit als **Nicht-Informationsressource** angesehen wird, hat einen pragmatischen Grund: In der zu generierenden RDF-Repräsentation sollen widerspruchsfreie Aussagen über Nicht-Informationsressourcen/Features gemacht werden. Wenn in Tripelform ausgedrückt wird, dass das Subjekt (z.B. ein Haus), das über eine NIR-URI identifiziert und in der zugehörigen Beschreibung explizit als Nicht-Informationsressource ausgewiesen wird, über das (geeignete) Prädikat `geo:hasGeometry` mit einem Objekt (eine Instanz der Superklasse `geo:Geometry`) zwecks Angabe der geometrischen Eigenschaften (Koordinaten) verlinkt wird, dann wird das Subjekt auch implizit als Instanz der Klasse `geo:Feature` ausgewiesen. Wenn die Instanzen der Klasse `geo:Feature` konkrete aber unvollständige **digitale Repräsentationen** einer Informationsressource wären, die sich auf die Nicht-Informationsressource beziehen, dann wären es zugehörige Beschreibungen und die Aussage wäre widersprüchlich. Abgesehen davon kann nicht davon ausgegangen werden, dass das Prädikat `geo:hasGeometry` für die Verlinkung einer digitalen (Web-)Repräsentation mit einem abstrakten Konzept, nämlich eine (bestimmte) Geometrie als Nicht-Informationsressource, vorgesehen ist. Das Konzept der Geometrie hat in diesem Kontext Ähnlichkeiten mit dem Konzept des Meters. Letzteres Konzept zog die TAG heran, um die Wesensmerkmale einer (Nicht-)Informationsressource zu erklären: „Even if the Web could deliver a platinum-iridium bar with two marks a meter apart at zero degrees celsius, or 1,650,763.73 wavelengths of the orange-red emission line in the electromagnetic spectrum of the krypton-86 atom in a vacuum [...], or even two marks, a meter apart on a screen, such representations are probably less than completely useful in the context of an information space. The representations [...] are not representations of the meter itself. Instead, they are representations of information resources related to the meter“ (Lewis, 2007). Im Übrigen zieht das Argument, dass sich möglicherweise nur wenige Agenten für die Spezifikationen aus der speziellen IRW-Ontologie interessieren, nicht. In sehr erfolgreichen LOD-Projekten wie DBPedia und GeoNames wird mithilfe von Konzepten aus dem etablierten FOAF-Vokabular ebenfalls ausgedrückt, dass es (Web-)Repräsentationen von Informationsressourcen gibt, die relevante Informationen enthalten, die sich auf Nicht-Informationsressourcen beziehen (z.B. mit dem Prädikat `foaf:primaryTopic`).

Die einfach formulierte Quintessenz ist, dass das Feature in der vorliegenden Arbeit als Nicht-Informationsressource angesehen werden muss, wenn widerspruchsfreie RDF-Tripel unter Nachnutzung von Konzepten aus etablierten Vokabularen generiert werden sollen. Bezugnehmend auf eine Formulierung von Tschirner (2012, p. 9) wird konstatiert, dass das Feature eine Abstraktion eines realweltlichen Phänomens ist, „das ein Konzept aus der Ideenwelt instanziiert“. Ein Feature kann in einer digitalen, zugehörigen Beschreibung lediglich beschrieben/abgebildet werden, aber das Feature per se kann nicht über ein Netzwerk transferiert werden.

Die vorstehenden Ausführungen mögen auf den ersten Blick übertrieben erscheinen. Bei genauerem Hinsehen ist die Klarstellung aber essenziell für die Kohärenz, da z.B. die Aussage, dass ein Feature einen Namen hat, zumindest mehrdeutig wird, wenn es auch der Name des Dokuments sein kann, in dem sich die Aussage findet.

⁹⁰ Richtig wäre, dass Instanzen der Klasse <http://www.opengis.net/ont/geosparql#Feature> Instanzen der Klasse <http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource> sind.

6.2 LOD-Anforderungen

6.2.1 Allgemeine Anforderungen

Bezeichnung (ID):	URI-Allokation (LOD-A1)	Typ: <u>obligatorische</u> Anforderung
Definition:	Bei der Generierung des Graphen muss das Prinzip der Universalität berücksichtigt werden. Jede <i>signifikante Ressource</i> muss über eine dereferenzierbare HTTP-URI identifiziert werden. URI-Kollisionen müssen vermieden werden. Es müssen NIR-URIs/Hash-URIs und 303-URIs erzeugt werden. Die URI-Generierung muss sich an den Empfehlungen von van den Brink et al. (2014) orientieren. Sämtliche URIs müssen in XML-Strukturen als kompakte URIs (engl. compact URIs, kurz: CURIEs) eingesetzt werden können.	
Begriffserläuterung:	<i>Signifikante Ressourcen</i> sind alle Ressourcen, die eine dereferenzierbare HTTP-URI „verdienen“. Dies ist der Fall, wenn Informationen über die identifizierte Ressource in Form von Web-Repräsentationen geliefert werden können. Dementsprechend sind Hilfsressourcen, denen lediglich eine strukturelle Funktion zukommt (z.B. um Modellierungsfehler zu vermeiden) keine <i>signifikanten Ressourcen</i> .	
Bezeichnung (ID):	Leere Knoten (LOD-A2)	Typ: optionale Anforderung
Definition:	Die Erzeugung leerer Knoten (s. GP 2.2.3) konfligiert mit Anforderung LOD-A1, weshalb darauf sowohl in der ABox als auch in der TBox verzichtet werden soll.	
Bezeichnung (ID):	Einfache SPARQL-Abfragen (LOD-A3)	Typ: <u>obligatorische</u> Anforderung
Definition:	Der Graph muss über SPARQL-Statements möglichst einfach abgefragt und umgeformt werden können. Laut Heath and Bizer (2011) wird dies erschwert, wenn Aussagen über Aussagen (Reifikation) gemacht werden oder RDF-Kollektionen oder RDF-Container verwendet werden. Der Einsatz dieser Konstrukte muss daher unterbleiben.	
Bezeichnung (ID):	Nachnutzung (LOD-A4)	Typ: optionale Anforderung
Definition:	Es soll den Empfehlungen von Heath and Bizer (2011) gefolgt werden, indem Konzepte aus Vokabularen nachgenutzt werden, die sich in der LOD-Cloud etabliert haben. Die infrage kommenden Vokabulare wurden im vorherigen Kapitel aufgezählt.	
Bezeichnung (ID):	Generierung von RDF-Links (LOD-A5)	Typ: <u>obligatorische</u> Anforderung
Definition:	Der Informationsumfang des Modells muss durch die Erzeugung von <i>RDF-Links auf der Instanzebene</i> erhöht werden. Die <i>Features</i> aus dem Häuserbuch müssen mit korrespondierenden Ressourcen verknüpft werden, welche in den Informationsbeständen der DNB (Titeldata, GND) beschrieben werden. Die Erfüllung dieser zentralen Anforderung hat eine erhöhte Priorität; der Zielerreichungsgrad wird quantifiziert, indem auf die Anzahl der o.g. Links abgestellt wird.	

- Abgrenzungen: Für die Beantwortung der Frage, ob der Informationsumfang des Modells dadurch gesteigert werden kann, wird nicht auf *interne RDF-Links*, nicht auf *Relationship Links* und nicht auf *Vocabulary Links* abgestellt. Der resultierende Graph wird diese Arten von Links (s.u.) bei der Umsetzung der übrigen Anforderungen zwar zwangsläufig enthalten; allerdings wird die Erhöhung des Informationsumfangs nur anhand der *externen RDF-Links auf Instanzebene* beurteilt, die nicht notwendigerweise *Identity Links* sind.
- Begriffserläuterungen:
- Die *Features* sind Beschreibungen ausgewählter Individuen (Instanzen), auf die in den Ausgangsdaten über Zeichenketten (Namens- und Titelinformationen) referiert wird. Die Auswahl beschränkt sich auf die Literaturangaben, die *Agenten* und die *Geoobjekte*. Prinzipiell können alle der vorgenannten Entitäten räumlich verortet werden; insofern sind sie alle *Geoobjekte*. Dieser Begriff bezieht sich im konkreten Fall nur auf die Objekte, die in den Ausgangsdaten über assoziierte Koordinaten verfügen (Gebäude, Grünflächen, Gewässer etc.). Unter dem Begriff *Agenten* werden alle natürlichen und juristischen Personen subsumiert.
 - *RDF-Links* beschreiben die Relation zweier Ressourcen. Sie bestehen i.d.R. aus drei URI-Referenzen, wobei die URIs an Subjekt- und Objektstelle die Ressourcen identifizieren, die in einer bestimmten Beziehung stehen, welche vom sie verknüpfenden Prädikat angegeben wird (Heath and Bizer, 2011).
 - *Externe RDF-Links* sind (unidirektionale) RDF-Links, die – im Gegensatz zu *internen RDF-Links* – URIs in Relation setzen, welche in unterschiedlichen Namensräumen erzeugt wurden (Heath and Bizer, 2011). Die NIR-URIs referieren also auf Bezugsobjekte, die in verschiedenen Informationsbeständen beschrieben werden (Heath and Bizer, 2011).
 - Wegen der o.g. Unidirektionalität ist zwischen *eingehenden* und *ausgehenden RDF-Links* zu differenzieren. Wenn in einem Tripel eine Aussage über die Relation der Nicht-Informationsressourcen A und B gemacht wird und die zugehörige Beschreibung von B dieses Tripel enthält, ist dieses ein *eingehender RDF-Link* zu B (Heath and Bizer, 2011). Beim *ausgehenden RDF-Link* ist die Richtung umgekehrt.
 - *Identity Links* bestehen aus *URI-Aliasen*, die im Tripel an Subjekt- und Objektstelle stehen und über das Prädikat **owl:sameAs** verknüpft sind (Heath and Bizer, 2011). Wenn der formalen Semantik des vorgenannten Prädikats Rechnung getragen wird, referieren die URIs auf denselben Referenten (Halpin et al., 2010).
 - *Relationship Links* verweisen auf Ressourcen, die Hintergrundinformationen zum Subjekt liefern können, etwa bibliographischen Angaben (Heath and Bizer, 2011).
 - *Vocabulary Links* verweisen von den beschriebenen Individuen in der ABox auf die korrespondierenden Konzepte in der TBox (z.B. durch Instanziierung) oder setzen Konzepte aus denselben oder verschiedenen Vokabularen in Relation, etwa durch Spezialisierung (Heath and Bizer, 2011).

- *RDF-Links auf der Instanzebene* sind in dieser Arbeit externe, ausgehende RDF-Links, welche die Bezugsobjekte verknüpfen, die in der ABox beschrieben werden und zwar unabhängig davon, ob die Semantik des dafür eingesetzten Prädikats formal oder natürlichsprachlich spezifiziert wurde. Die Menge an Aussagen (oder Behauptungen) über ein bestimmtes Individuum kann auch unter Verwendung anderer Prädikate als `owl:sameAs` erhöht werden. Die NIR-URIs an Subjekt- und Objektstelle müssen also nicht unbedingt *URI-Aliase* sein und das Tripel qualifiziert sich deshalb auch nicht immer als *Identity Link*.

6.2.2 Individuelle Anforderungen

Bezeichnung (ID):	Explizite Aussagen (LOD-I1)	Typ: optionale Anforderung
Erläuterung:	Die Wahrscheinlichkeit, dass die Wissensbasis in unvorherzusehenden Anwendungsszenarien zweckgerichtet verwendet werden kann, erhöht sich, wenn etwaigen Missverständnissen vorgebeugt wird. Diese sollen vermieden werden, indem explizite Aussagen gemacht werden. Der „Overhead“, der dadurch produziert wird, ist – mengenmäßig betrachtet – noch akzeptabel; der Verständlichkeit wird eine höhere Priorität eingeräumt. Dadurch soll den Empfehlungen von Presutti and Gangemi (2008) und Halpin and Presutti (2009) gefolgt werden, indem komplexe Formalisierungen in Kombination mit Konzepten aus einfachen Ontologien nachgenutzt werden.	
Begriffserläuterung:	Ein Prädikat <i>ohne formale Semantik</i> ist ein spezifiziertes Property, das keine Inferenzen erlaubt – „so in that sense it has <i>no semantics</i> “ (Allemang et al., 2020, p. 229).	
Definition:	Es soll formal aber verständlich ausgedrückt werden, ob es sich um eine Nicht-Informationsressource oder eine zugehörige Beschreibung handelt. Außerdem soll das CRS und der Typ der Geometrie angegeben werden. Neben Konzepten, die maschinelle Schlussfolgerungen zulassen, sollen auch Prädikate <i>ohne formale Semantik</i> eingesetzt werden, die Objekte referenzieren, welche für die Interpretation durch einen Menschen bestimmt sind.	

Bezeichnung (ID):	Serialisierung / Kodierung (LOD-I2)	Typ: <u>obligatorische</u> Anforderung
Definition:	Es muss ein gerichteter Graph erzeugt werden. Zur Serialisierung muss RDF/XML verwendet werden. Die Kodierung der Tripel muss dementsprechend hierarchisch erfolgen. Die resultierende Baumstruktur muss ein wohlgeformtes XML-Dokument sein. Als Zeichensatz muss Unicode verwendet werden. Die Zeichenkodierung muss mittels UTF-8 erfolgen. Das XML-Dokument muss möglichst einfach und verständlich sein.	
Erläuterung:	Mit RDF/XML lässt sich derselbe Graph unterschiedlich kodieren. Technisch betrachtet ist es egal, wie die Serialisierung genau realisiert wird, solange die Syntax stimmt. In dieser Arbeit wird eine einfache Option gewählt. Die Tripelstruktur muss erkennbar bleiben; insbesondere müssen die Subjekte anhand des XML-Elements „rdf:Description“ identifiziert werden können.	

6.3 Anwendungsfallsspezifische Anforderungen

Bezeichnung (ID): Instanzen (AS1) Typ: obligatorische Anforderung

Definition: Die Fakten über die Individuen, ihre Eigenschaften und ihre Beziehungen untereinander müssen im RDF-Graph beschrieben werden. Dafür muss grundsätzlich jeder Datensatz (Tupel) aus den ausgewählten MapInfo-Relationen (die in Tab. 5.1, Tab. 5.4, Tab. 5.5 dokumentiert wurden) in Tripelform beschrieben werden (vollständige Übernahme). Ausnahmen gelten für jene Datensätze, welche die o.g. *Features* (vgl. Begriffserläuterungen zu LOD-A5) beschreiben. Diese dürfen nicht vollständig übernommen werden, wenn es Duplikate sind. Duplikate müssen erkannt und entfernt werden. Im Zuge der Datenbereinigung müssen Inkonsistenzen beseitigt werden und sämtliche Geometrien validiert und ggf. korrigiert werden (Verbesserung der Datenqualität, kurz DQ). Die Koordinaten müssen umprojiziert werden, wofür das im LOD-Kontext empfohlene CRS (Parameter: [ESPG:4326](#)) verwendet werden muss.

Abgrenzung: Es erfolgt keine Normalisierung und keine Synthese von Relationen. Es erfolgen keine weiteren Datenerfassungen mittels zusätzlicher Primär- oder Sekundärquellen.

Bezeichnung (ID): Konzepte (AS2) Typ: obligatorische Anforderung

Definition: Im Rahmen der Konzeptualisierung müssen die *erforderlichen* Klassen und Properties mithilfe einer simplen Domänenontologie (Mapping) und/oder in Form einer einfachen, flexiblen Problemontologie spezifiziert werden (Modellierung). *Erforderlich* sind nur Klassen und Properties, die nicht unverändert aus etablierten Vokabularen übernommen werden können (s. LOD-A4) und die einen Bezug zu den ausgewählten MapInfo-Relationen aufweisen, welche in den o.g. Tabellen dokumentiert wurden (s. AS1). Der Graph muss dadurch komplettiert werden. Der Einfachheit halber und unter Berücksichtigung des Prinzips der geringsten Ausdrucksmächtigkeit muss die Komplexität auf das notwendige Minimum reduziert werden. Wenn auf Konzepte aus Domänenontologien oder sonstigen Vokabularen zwecks Spezialisierung, Mapping etc. zurückgegriffen wird, müssen diese ohne komplexe OWL-Konstrukte auskommen. Die Sprachmittel müssen auf RDFS-Plus-Konzepte beschränkt werden. Die Flexibilität muss bei der taxonomischen Einordnung berücksichtigt werden; vor allem die Klassenhierarchie muss unter Beibehaltung ihrer Grundstruktur erweiterbar oder adaptierbar sein.

Erläuterung: Die Komplettierung des Graphen fußt auf der einleitend erwähnten Intention: Die konzeptuelle Grundlage des historischen Häuserbuchs soll offengelegt, unmissverständlich beschrieben und mit Konzepten aus anderen Quellen verknüpft werden. Dadurch soll das Teilen von Informationen ermöglicht werden (Primärzweck). Die Ableitung impliziten Wissens über einfache Schlussfolgerungsmethoden nimmt eine untergeordnete Rolle ein (Sekundärzweck).

6.4 Anwendungsszenario

Eine Schülerin der gymnasialen Oberstufe (Leistungskurse: Deutsch, Englisch, Geschichte; Wahlpflichtfach: Informatik) plant ihre Facharbeit. Sie interessiert sich insbesondere für Literatur und Kunst sowie für den Dreißigjährigen Krieg, weshalb ihr der Raub der Pfälzischen Landbibliothek (Bibliotheca Palatina) durch die Katholische Liga in besonderer Erinnerung geblieben ist. In ihrer Facharbeit möchte sie sich anhand von Primär- oder Sekundärliteratur intensiv mit dem Thema „Kunstraub im Dreißigjährigen Krieg“ auseinandersetzen.

Sie möchte sich auf den Mainzer Stadtteil konzentrieren, in dem sie wohnt – die Altstadt (geographischer Rahmen). Den Kunstbegriff schränkt sie auf Bücher ein, die sich im Besitz von Privatpersonen befanden (Untersuchungsgegenstand). Im Rahmen ihrer Facharbeit kann sie zwar nicht sämtliche Informationen über die Herkunft eines Buches recherchieren, trotzdem möchte sie mit Methoden der Provenienzerschließung ermitteln, ob in der Mainzer Altstadt Buchbesitzer gewohnt haben, deren Bücher von den Schweden im Dreißigjährigen Krieg beschlagnahmt respektive geraubt wurden (Leitfrage, Nicht-Ziel, Methoden). Ihren zeitlichen Rahmen legt sie deshalb mit dem Ende des Zweiten Markgrafenkrieges (Ende 1554) – ein Konflikt in dem Kunstraub ebenfalls praktiziert wurde (König, 2018) – und mit dem endgültigen Abzug der schwedischen Truppen aus Mainz fest (Anfang 1636). Einen Ausgangspunkt für ihre Recherche findet sie schnell. Aufgrund ihres Referats zur Wissenschaftlichen Stadtbibliothek Mainz, für das die Arbeit von König (2018) sehr hilfreich war, weiß sie schon, dass es eine Verbindung zwischen der vorgenannten Stadtbibliothek und der alten Universität Mainz (Alma Mater Moguntina) bzw. der dazugehörigen Bibliothek gibt, die als gemeinsame, räumlich vereinigte Universitätsbibliothek (communis liberaria Universitatis) eingerichtet wurde. Es ist ihr bekannt, dass die Universitätsbibliothek „in einem der von Lehrkörper und Studenten als gemeinschaftliche Wohn und Arbeitsstätten genutzten Anwesen aufgestellt gewesen sein muss“, die man „Bursen“ nannte (König, 2018, p. 17).

Sie vermutet, dass die o.g. Buchbesitzer in diesen Bursen gewohnt haben könnten. Sie beschließt im historischen Häuserbuch von Mainz anhand der o.g. Kriterien zu suchen, da dieses als LOD ins Web integriert wurde und besonders viele Links zu den Ressourcen der DNB enthält. Da formale Sprachen und SPARQL im Inhaltsfeld 3 ihres Wahlpflichtfachs bereits behandelt wurden, kann sie eine effiziente Recherche via SPARQL durchführen. Sie erwartet nicht, dass sie mithilfe des RDF-Graphen ihre Leitfrage beantworten kann, aber sie erwartet, dass die verfügbaren Informationen von angemessenem Umfang sind. Ihre Anforderungen sieht sie als erfüllt an, wenn sie nutzbare Antworten auf die Fragen erhält, welche nachfolgend gruppiert aufgeführt sind:

Gruppe 1: Informationen zu den Bursen

- 1.1 Gab es in der Mainzer Altstadt im Zeitraum vom 01.01.1555 bis zum 31.12.1635 Universitätsbursen?
- 1.2 Wie lauteten die Namen der Häuser?
- 1.3 Wie lauteten die Namen der Eigentümer und der Bewohner?
- 1.4 Können RDF-Links für genauere Nachforschungen zu den Bewohnern verfolgt werden?

Gruppe 2: Informationen für die Erstellung biografischer Skizzen der Bewohner

- 2.1 Wie lauteten alle infrage kommenden Namen?
- 2.2 Welche akademischen Grade haben sie erworben und welche Berufe haben sie ausgeübt?
- 2.3 Wann wurden sie geboren?
- 2.4 Wo wurden sie geboren? Wann sind sie gestorben? In welchem Zeitraum haben sie gewirkt und an welchen Orten? Haben sie in Mainz gewirkt?

- 2.5 Standen sie mit Personen, die in Mainz gewirkt haben, in einer Freundschafts- oder Verwandtschaftsbeziehung?
- 2.6 Falls dies zutrifft, sollen alle vorgenannten Fragen auch für diese Personen beantwortet werden.

Gruppe 3: Informationen zu den Beziehungen und für die weiterführende Recherche

- 3.1 Kann die Beziehung zwischen den Personen genauer spezifiziert werden?
- 3.2 Standen die Personen in einer direkten Beziehung zu den Eigentümern der Häuser, die sie bewohnten?
- 3.3 Gibt es weitere Informationen zu diesen Eigentümern oder zu damit verknüpften Entitäten, die in thematischer Hinsicht relevant sein könnten?
- 3.4 Gibt es themenspezifische Fachliteratur, die sich genau auf die relevanten Agenten⁹¹ bezieht?

Die meisten Fachwörter, die unter GP 1.2 zwecks Spezifizierung der Leitfrage und der (Teil-)Ziele verwendet wurden, stammen aus einschlägiger Grundlagenliteratur und Lehrbüchern – z.B. wurde das Adjektiv „nicht-normalisiert“ von Bill (2016, p. 362) verwendet. Für den unspezifischen Begriff „Informationsumfang“ gilt dies nicht⁹². Wie eingangs erwähnt, soll dieser Begriff präzisiert werden. Dafür ist die Arbeit von Rohweder et al. (2021) besonders gut geeignet. Aufbauend auf das anwenderfokussierte (fit for use) Konzept von Wang and Strong (1996) führten die vorgenannten Autoren „ein Katalog von IQ-Dimensionen und Definitionen für Merkmale von Datenprodukten“ auf, der „das Ergebnis der Arbeit einer Projektgruppe in der DGIQ (Deutsche Gesellschaft für Informations- und Datenqualität)“ ist, um „die Kommunikation über Themen des IQ-Managements durch eindeutige und einheitliche Begrifflichkeiten zu vereinfachen und zu verbessern“ (Rohweder et al., 2021, p. 26). Der Katalog enthält 15 Begriffe (sog. IQ-Dimensionen) anhand derer die IQ definiert wird (Rohweder et al., 2021). Die Arbeit von Rohweder et al. (2021) zeichnet sich auch durch Verständlichkeit aus, da die Definitionen der IQ-Dimensionen anhand von Beispielen beschrieben wurden. Für die vorliegende Arbeit ist vor allem ein Beispiel zu Stammdaten relevant; hiermit lässt sich der Grad der Erfüllung der Anforderungen an das Datenprodukt bestimmen (Rohweder et al., 2021). In dieser Thesis lässt sich der vorgenannte Grad etwas näher bestimmen. Das Beispiel für Stammdaten ist dem Beispiel für Bewegungsdaten vorzuziehen, da die Daten aus dem historischen Häuserbuch die Charakteristika von Bewegungsdaten nicht aufweisen⁹³. Aus dem o.g. Katalog wird eine zweckabhängige IQ-Dimension – der „angemessene Umfang“ (Rohweder et al., 2021, p. 33) – herausgegriffen. Der Untersuchungsgegenstand der zweckabhängigen IQ-Dimensionen ist die Nutzung (Rohweder et al., 2021), weshalb das obige fiktive Anwendungsszenario konstruiert wurde. Rohweder et al. (2021, p. 33) definierten den angemessenen Umfang wie folgt: „Informationen sind von angemessenem Umfang, wenn die Menge der verfügbaren Information den gestellten Anforderungen genügt“. Rohweder et al. (2021, p. 33) illustrierten dies u.a. anhand des folgenden Beispiels zu Stammdaten: „In der Abteilung für die Vertragserstellung benötigt der Sachbearbeiter umfassende Informationen zum Kunden.

- (positiv) Im [...] Kontaktmanagementsystem stehen alle Informationen zur Verfügung.
- (negativ) Im Kontaktmanagementsystem fehlen für diesen Prozess wesentliche Informationen zum Kunden“.

⁹¹ Der Begriff wurde in Abschnitt 6.2 erläutert. In manchen Fällen ist es möglich, dass nur ein infrage kommendes Individuum beschrieben wird. Zwecks besserer Lesbarkeit wurde bei der Formulierung der Fragen auf den Singular verzichtet.

⁹² Alle anderen Eigenkreationen wurden bereits erläutert z.B. der Begriff „Konzeptualisierungsprozess“ (s. GP 2.2.2).

⁹³ Im Übrigen war es deshalb auch legitim, die Aktualisierungsprozesse vorerst, d.h. in dieser Arbeit, auszuklammern.

Für die Demonstration, dass die verknüpften Informationen (im Vergleich zu den Ausgangsdaten) einen Mehrwert für potenzielle Anwenderinnen bieten, wird das obige Beispiel wie folgt umformuliert: Eine Schülerin der gymnasialen Oberstufe benötigt für ihre Facharbeit umfassende Informationen zu den Eigentümern von Universitätsbursen, sofern in diesen Häusern Buchbesitzer gewohnt haben. Sie möchte die Bursen selektieren, um Informationen für die Erstellung biografischer Skizzen der Bewohner zu beziehen. Außerdem möchte sie Informationen zu den Beziehungen der jeweiligen Agenten und Informationen für die weiterführende Recherche erhalten. Dafür stellt sie Anfragen an das historische Häuserbuch von Mainz, das ursprünglich als schnell recherchierbares Nachschlagewerk konzipiert wurde.

- (positiv) Im Nachschlagewerk stehen alle Informationen zur Verfügung.
- (negativ) Im Nachschlagewerk fehlen für die o.g. Prozesse wesentliche Informationen zu den Eigentümern, zu den Bewohnern oder zu den Häusern.

Positiv wäre es, wenn die Schülerin alle Fragen aus den Gruppen 1-3 anhand des historischen Häuserbuchs bzw. mithilfe der Informationen der DNB beantworten kann. Dann sieht sie ihre Anforderungen als erfüllt an.

Zwecks Begriffsbestimmung wurde die o.g. IQ-Dimension aus dem Katalog herausgegriffen. Durch ein solches Vorgehen kann die DQ/IQ nicht adäquat beurteilt werden. Rohweder et al. (2021, p. 29) hoben hervor, dass die Funktionsfähigkeit eines IS erst dann gewährleistet ist, „wenn alle IQ-Dimensionen eine hohe oder zumindest ausreichende Qualität aufweisen“. Dafür müssten die systemunterstützten, dateninhärenten, zweckabhängigen und die darstellungsbezogenen IQ-Dimensionen ganzheitlich betrachtet werden (Rohweder et al., 2021). Zwecks effizienter Zielerreichung soll auf die Forschungsergebnisse aus wissenschaftlichen Arbeiten zurückgegriffen werden (vgl. GP 3.2). Wie einleitend erwähnt, setzt dies die Geeignetheit der infrage kommenden Applikationen voraus (vgl. GP 1.2). Wenn diese für den konkreten Anwendungsfall nicht gegeben ist, könnte sich am Ende herausstellen, dass das IS, welches zur Demonstration benötigt wird, lediglich ein simples „tell-and-ask system“ (Allemand et al., 2020, p. 120) mit einfachem Webinterface ist. Das obige Anwendungsszenario wurde so konstruiert, dass für die Demonstration schon ein rudimentärer SPARQL-Client ausreicht. Ergonomische Aspekte oder die einfache Zugänglichkeit spielen dann keine Rolle mehr. Informationen, die lediglich durch die Formulierung von SPARQL-Statements bezogen werden können, sind nicht für alle potenziellen Anwenderinnen einfach zugänglich; aus CONSTRUCT-Statements generierte RDF-Graphen, welche z.B. mit Hilfe von RDF/XML serialisiert werden, tragen ggf. keine Anweisungen/Festlegungen zur Darstellung etc. pp.

7 Lösungsansatz

7.1 Entwurf des Systems

Das Ziel dieser Arbeit besteht darin, Geodaten zu einem RDF-Graph umzuformen, der den LOD-Anforderungen entspricht. In diesem Zuge werden die Ausgangsdaten zu Informationen aufgewertet. Die Menge der verfügbaren Information zu Individuen wird angemessen erhöht, damit sie den gestellten Anforderungen genügt. Wie in Abschnitt 5.1 dargelegt, erfolgt dies am Beispiel vektorieller Geodaten, die bestimmte Charakteristika aufweisen. Eine Sammlung solcher Geodaten kann zweifelsohne auch als Geodaten-Silo bezeichnet werden, um das Ziel dieser Thesis noch einmal bildlich beschreiben zu können: „Durch Anwendung der LOD-Methode soll ein Geodaten-Silo geöffnet werden, weil dadurch der Wert des Inhalts gesteigert werden kann“.

Dafür wurden drei operative Teilziele definiert, nämlich die Anforderungsanalyse und -definition zwecks Bewertung des resultierenden RDF-Graphen aus der Sicht einer Datenbereitstellerin (1), die Konzeption und praktische Umsetzung eines Workflows zwecks Generierung des LOD-konformen RDF-Graphen mit einem erhöhten Informationsumfang aufgrund zahlreicher RDF-Links auf der Instanzebene (2), die Demonstration des erhöhten Informationsumfangs mithilfe eines IS oder GIS aus der Sicht einer potenziellen Anwenderin (3).

Jedes dieser Teilziele kann weiter unterteilt werden, was schon aus der Spezifikation unter GP 1.2 hervorgeht und exemplarisch anhand des zweiten Teilziels aufgezeigt wird: „Ausgangsdaten transformieren“, „Konzepte formalisieren“, „Graph anreichern“ – dies sind für sich genommen schon Teilziele, die schlichtweg der Kompaktheit halber dem zweiten Teilziel untergeordnet wurden. Die folgende Abbildung zeigt dies nur ansatzweise:

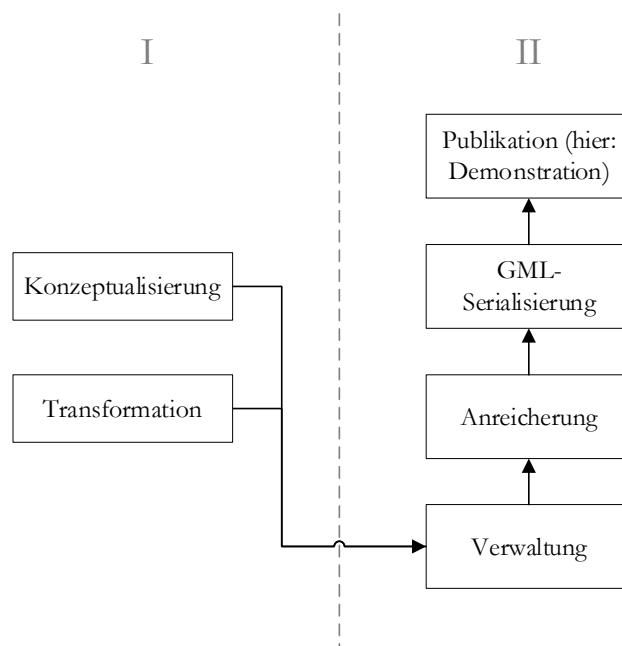


Abb. 7.1: Entwurf des Systems (Teil 1), ED

Einiges wurde bewusst ausgespart, da diese Abbildung natürlich nicht der Konzeption dieser Arbeit dient, sondern der Konzeption dieses Kapitels, in dem Lösungen entwickelt werden, die auf die Generierung und Publikation von LGCHD abzielen⁹⁴ und zwar unter der Prämisse, dass die vernetzten Informationen sowohl von LOD-Applikationen als auch von WFS-Clients konsumiert werden können (daher „GML-Serialisierung“).

⁹⁴ Dies ist lediglich eine Alternative zur bildlichen Beschreibung.

Um dieses Ziel zu erreichen, kann auf die Ergebnisse aus wissenschaftlichen Arbeiten zurückgegriffen werden. Abgesehen von den besprochenen Verfahrensweisen können auch Applikationen aus Forschungsprojekten wiederverwendet werden, um ein System für die Generierung und Publikation von LGCHD aufzubauen. In Kapitel 3 wurde – anhand von Abb. 3.1 – aufgezeigt, dass dafür viele Lösungsmöglichkeiten in Frage kommen.

Diese sollen in einer sinnvollen Reihenfolge diskutiert werden. Dafür kann das o.g. System unterteilt werden:

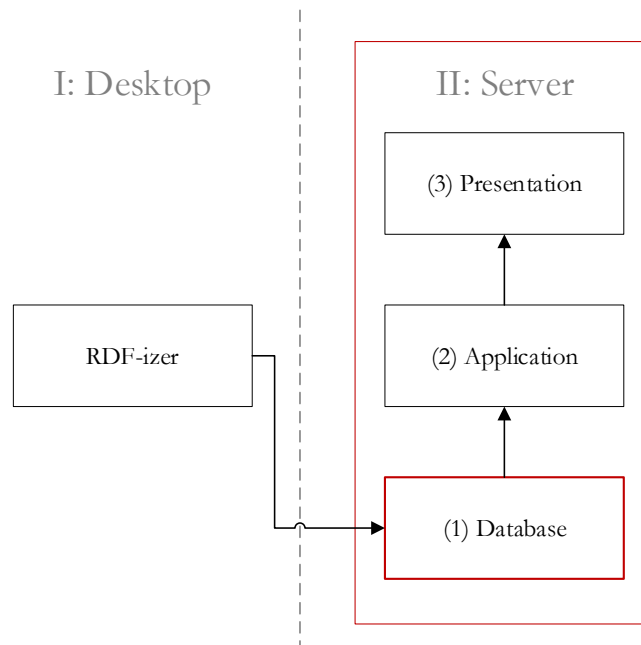


Abb. 7.2: Entwurf des Systems (Teil 2), ED

Abb. 7.2 sollte zusammen mit Abb. 7.1 betrachtet werden: Sowohl die Transformation als auch die Konzeptualisierung kann (allein) mit Desktop-Applikationen bewerkstelligt werden. Für die Prozesse, die in Abb. 7.2 auf der rechten Seite (II) aufgeführt sind, gilt dies nicht immer – zumindest für die Publikation⁹⁵ ist ein Server erforderlich. Mit der Trennlinie wird betont, dass die (virtuellen) Maschinen, auf denen die Applikationen laufen können, prinzipiell unabhängig voneinander sind und dementsprechend auch nacheinander besprochen werden können. Die rote Markierung zeigt, dass **idealerweise zuerst der Server** ausgehend von der Datenhaltungsschicht besprochen werden sollte. Die Entscheidung, ob zwecks Persistierung auf ein graphbasiertes Datenmodell oder das Relationenmodell gesetzt werden soll, beeinflusst alles Weitere:

(A) Bei der Entscheidung für ein RDBMS/ORDBMS mit räumlicher Erweiterung müssen die Geodaten aus den MapInfo-Relationen in verschiedene Datenbanktabellen eingefügt werden, wobei die Redundanzfreiheit und die Datenkonsistenz im Rahmen der Normalisierung sicherzustellen ist. Wie bereits erwähnt, gibt es mehrere Möglichkeiten, um diese Daten unter Beachtung der LOD-Prinzipien bereitzustellen ohne sie mithilfe einer Graphdatenbank speichern zu müssen:

- Um LOD-Clients zu bedienen, können RML- / R2RML-Mappings zusammen mit Applikationen verwendet werden, welche die Generierung von virtuellen RDF-Sichten und die Umsetzung von SPARQL- in SQL-Anfragen zur Laufzeit ermöglichen.

⁹⁵ In dieser Arbeit dient die zweite VM (auch) der Demonstration. Perspektivisch sollen die vernetzten Informationen über den Server für die Nachnutzung bereitgestellt werden, weshalb hier von „Publikation“ gesprochen wird.

- Anfragen von LOD-Clients können auch von LOD-Wrappern beantwortet werden, die OGC-konformen WFS vorgeschaltet werden. Bei Wahrung der Standardkonformität können Web- oder Desktop-GIS WFS-Anfragen stellen und die Resultate einlesen.

In Abb. 7.3 ist diese Option links dargestellt; die o.g. Möglichkeiten wurden in einem Block vereinfacht dargestellt bzw. zusammengefasst, der mit „*Semantic Uplift*“ beschriftet wurde, da die Details hier irrelevant sind (s.u.).

- (B) Wenn eine graphenbasierte Datenbank (hier: Triple- bzw. Quadstore oder RDF-Store) mit GeoSPARQL-Unterstützung eingesetzt wird, müssen die MapInfo-Relationen mithilfe eines RDF-izers umgeformt werden, um die Tripel entweder direkt in den RDF-Store einzufügen oder zunächst in einer gängigen Serialisierung zu exportieren. Die Datenbereinigung (z.B. Duplikaterkennung und -eliminierung) kann im Rahmen der (semantischen) Transformation vorgenommen werden. Die Publikationsmöglichkeiten unterscheiden sich ebenfalls im Vergleich zu Option A:

- Der Query-Prozessor des RDF-Stores kann via SPARQL (und über einen HTTP-Server) als Webservice agieren, um Anfragen von LOD-Clients zu beantworten, die programmgesteuert von (LOD-)Applikationen oder über das Web- bzw. Query-Interface gestellt werden können.
- Anfragen von Web- oder Desktop-GIS können von einem WFS-Adapter beantwortet werden (Semantic Downlift).

Diese Option ist rechts dargestellt:

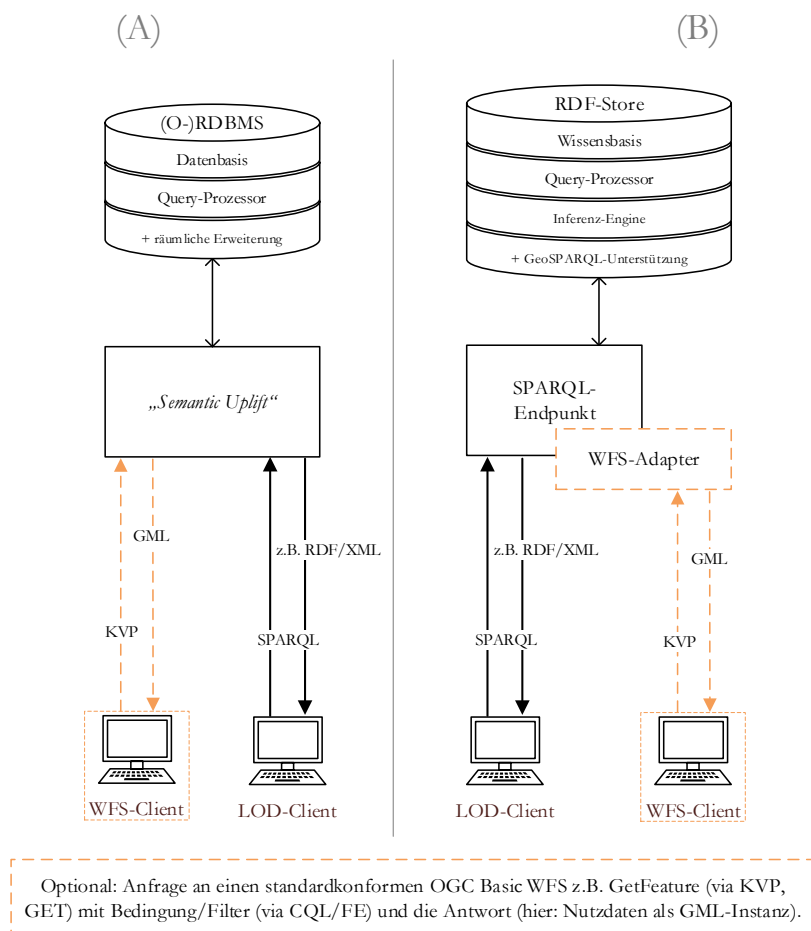


Abb. 7.3: Entwurf des Systems (Teil 3), ED

Um das Ziel zu erreichen, das zu Beginn dieses Kapitels noch einmal wiederholt wurde, müssen die Daten aus dem historischen Häuserbuch mit Ressourcen aus der LOD-Cloud kombiniert bzw. zu diesen in Relation gesetzt werden. Wenn dafür auf (vollständige) SQL-Dumps⁹⁶ und ein adäquates (O-)RDBMS zurückgegriffen wird, gestaltet sich die Integration neuer/anderer Daten recht aufwendig. Das relationale Datenbankschema muss zumindest erweitert, meist aber auch angepasst werden. Abgesehen davon, dass die Spezifika der jeweiligen (O-)RDBMS beachtet werden müssen, kann es auch schwierig sein, die eigenen Daten zu Daten aus einer externen Quelle in eine sinnvolle Beziehung zu setzen. Die Relationenschemata und Integritätsbeschränkungen müssen interpretiert werden, um auf die Semantik zu schließen, da sie nicht formal beschrieben wird. Im Gegensatz dazu ist die Daten- bzw. Informationsintegration einfach möglich, wenn für die Persistierung das flexible RDF-Datenmodell verwendet wird. Graphen können einfach vereinigt werden. Die Bedeutung oder die Struktur muss dabei nicht verändert werden. Deshalb werden die Informationen der DNB in Form mehrerer RDF-Dumps bezogen⁹⁷, die einfach in einen geeigneten RDF-Store importiert werden können, ohne überhaupt ein Schema anlegen zu müssen. Der Kontext kann über benannte Graphen angegeben werden. Aus den vorgenannten Gründen wird der Option (B), die in Abb. 7.3 vereinfacht dargestellt wurde, der Vorzug gegeben.

7.1.1 Infrage kommende Applikationen

Jones et al. (2014, p. 2) erwähnten: „making Linked Open Data available in a geo service standard will turn *all* geo-service compatible GIS tools, whether they consist of simple desktop clients or distributed service implementations, into *powerful* geographic analysis tools of the LOD cloud“.

Hier soll nicht diskutiert werden, ob LOD bzw. LGD in *allen* GIS bzw. WFS-Clients über *mächtige* Analyse- oder Schlussfolgerungsmethoden prozessiert werden kann. Es wird die Auffassung vertreten, dass LGD wegen der getypten Links, die Informationen aus verschiedenen Quellen in eine sinnvolle Relation setzen, einen Mehrwert gegenüber einer isolierten Datensammlung bietet, sofern die Links clientseitig in einer Art und Weise verarbeitet werden können, die für potenzielle Anwenderinnen nützlich und sinnvoll ist. Die Motivation GIS-Anwenderinnen nicht von vornherein von der Nutzung der o.g. Informationen auszuschließen, wurde unter GP 4.3 anhand eines Verwendungsbeispiels aus der Praxis aufgezeigt. Deshalb wurde nach einer Lösung gesucht, wie LGD über einen Basic WFS bereitgestellt werden kann. Im Rahmen der Demonstration sollte gezeigt werden, dass die Kombination von Informationen aus unterschiedlichen Quellen zu einem Mehrwert führt, von dem auch GIS-Anwenderinnen profitieren können – insbesondere, da in GIS Verschneidungen mit weiteren Geoinformationen durchgeführt werden können, um neue Erkenntnisse zu gewinnen. Anders ausgedrückt: „Um Fragen zu beantworten, die sich anhand der ursprünglichen Relationen nicht beantworten lassen“. Diese Fragen müssen nicht immer so kompliziert sein, dass die Antwort nur unter Anwendung ausgefeilter Methoden gefunden werden kann. Einige Fragen lassen sich beantworten, wenn Individuen in eine spezifizierte Beziehung zueinander gesetzt werden (was unter GP 9.2 aufgezeigt wird). Für die GML-Serialisierung ist es entscheidend, dass die Fakten über die Individuen und ihre Beziehungen untereinander nicht verloren gehen.

Trotz aller positiven Aspekte von LOD4WFS, die unter GP 3.2.2 hervorgehoben wurden, kann in der vorliegenden Arbeit keine Lösung präsentiert werden, wie mit dem o.g. WFS-Adapter vernetzte Informationen in verschachtelten GML-Strukturen mit Querverbindungen abgebildet werden können, um sie über einen OGC-konformen Basic WFS für die Nutzung in GIS bereitzustellen. Dies ist im konkreten Fall nur durch die Anpassung des Quellcodes zu erreichen, was nicht mehr realisiert werden konnte.

⁹⁶ Der Grund, warum in diesem Zusammenhang nur noch von Dumps und nicht von Diensten gesprochen wird, findet sich in Abschnitt 5.2.

⁹⁷ Tab. 8.11 listet die Details.

Wie bereits erwähnt, ergibt sich dieses Erfordernis nicht aus der fehlenden FE- / CQL-Unterstützung. Grundsätzlich können via FDA bzw. über die Formulierung geeigneter SPARQL-Statements zweckmäßige Feature Collections für die layerbasierte Verwendung in GIS zusammengestellt werden. Wenn man keinen Wert auf die Validierung gegen das XSD legt, welches unter GP 2.3.2 besprochen wurde, kann man die Filter-Sektion aus der XML-Datei entfernen, die quasi als Template für die Generierung des Capabilities-Dokuments dient. Zumindest in den Desktop-GIS QGIS 3.16.4-Hannover, ArcGIS Pro 2.6.7, ArcGIS Desktop 10.6.1 und MapInfo Professional 17.03 konnten die von LOD4WFS bereitgestellten Instanzen letztlich trotzdem eingelesen werden. Außerdem bieten alle der o.g. Desktop-GIS adäquate Funktionen für die Selektion von Geoobjekten. Im konkreten Fall bedarf es deshalb nicht zwangsläufig der gefilterten Downloadmöglichkeit. Prinzipiell könnte man die Anwenderinnen auch über einen Kommentar darauf hinweisen, dass die o.g. Möglichkeit nicht besteht, falls die Filter-Sektion im XML-Dokument verbleiben soll. In einigen Anwendungsfällen wird es wichtiger sein, dass die Informationen überhaupt bezogen werden können (als „predefined dataset“). Abb. 7.3 dient der vereinfachten Darstellung; der WFS-Client ist bei Option (B) nur einmal aufgeführt. Dies können nicht nur Desktop-GIS, sondern auch WebGIS-Clients sein, die auf der Präsentationsschicht eingesetzt werden – z.B. der Viewer, der von Vilches-Blázquez and Saavedra (2019) u. a. mithilfe von Leaflet entwickelt wurde. In diesem Fall kann individuell geregelt werden, wie der Client mit dem WFS-Adapter kommuniziert, weshalb die fehlende CQL-Unterstützung in dieser Arbeit nicht als Hinderungsgrund für den Einsatz des WFS-Adapters angesehen wird.

Problematischer ist es, dass das Applikationsschema, welches von LOD4WFS als Antwort auf eine Describe-FeatureType-Anfrage generiert wird, keine Semantik innehat. Dies gilt auch für das Capabilities-Dokument. Der Client kann die XSD zur syntaktischen Validierung verwenden; die Bedeutung kann im konkreten Fall nur abgeleitet werden, da die Elemente „sprechende“ Namen haben. Der WFS-Adapter operiert auf einem RDF-Store, in dem die ABox und die TBox verwaltet werden. Unter GP 6.3 wurde der Primärzweck der zu erstellenden TBox festgelegt. Da mit einer simplen TBox primär das Teilen von Informationen ermöglicht werden soll und die Ableitung impliziter Aussagen eine untergeordnete Rolle einnimmt, kann die semantische Interoperabilität teilweise gewährleistet werden. Dies kann im konkreten Anwendungsfall manuell über semantische Annotationen auf Service- und Modellebene erfolgen, die Maué (2009) vorschlug. Zwecks Reduktion des Konfigurations- bzw. Administrationsaufwands können die einzelnen XML-Templates zusammengeführt werden.

Das größte Problem, das im Rahmen dieser Arbeit nicht gelöst werden konnte, betrifft die Instanzen. Die Charakteristika der Ausgangsdaten wurden unter GP 5.1.1 beschrieben. Zur Veranschaulichung des Problems soll noch einmal erwähnt werden, dass eine Person X und eine Person Y in einem Haus A gewohnt haben könnten; Person Y wohnte auch in Haus B und Person X war der Eigentümer der Häuser C und D, die in der Quelle Z beschrieben wurden, welche sich aber auch auf Haus A bezieht usw. Neben den Individuen per se sind deren Beziehungen untereinander von besonderem Interesse, weshalb das Vorstehende mithilfe des RDF-Datenmodells ausgedrückt werden soll. Der resultierende RDF-Graph, der in einem Triplestore verwaltet wird, kann via SPARQL-SELECT abgefragt werden. Der Triplestore liefert das Resultat dieser Abfrage in tabellarischer Form aus. Ohne die Anpassung des Quellcodes von LOD4WFS, lassen sich GML-Instanzen mit diesem WFS-Adapter nur auf der Basis eines Resultats generieren, das vom Triplestore auf eine SPARQL-SELECT-Anfrage geliefert wird. Die Tabellenstruktur ist im konkreten Fall sehr problematisch, da sich die Informationen damit nicht sinnvoll abbilden lassen. Der WFS-Adapter verarbeitet das vorgenannte Resultat sequentiell, weshalb die vollständige Eliminierung von Redundanz aus der Tabelle notwendig ist. „Doppelte Tabellenzeilen“ treten dann auf, wenn die Belegung von Variablen (?) relevant ist, die prinzipiell unabhängig voneinander sind; z.B. wenn nach der ?ID-A von Haus A und den ?IDs-B der Bewohner von Haus A gefragt wird. Es gibt keine doppelte Tabellenzeile, aber ?ID-A ist immer gleich belegt. Durch die Verkettung von Zeichenketten lässt sich die redundante Information entfernen (GROUP_CONCAT). Durch die Kombination (fast) aller Möglichkeiten, die SPARQL 1.1 für SELECT-Statements anbietet, lassen sich sozusagen „Tabellen in die Tabelle schachteln“.

Im Quellcode von LOD4WFS muss definiert werden, welche Zeichen dafür verwendet wurden. Auf diese Weise lässt sich eine tiefer verschachtelte GML-Instanz erzeugen. LOD4WFS funktioniert dann nur noch mit einer SPARQL-Abfrage. Dies ist im konkreten Fall trotzdem nicht zielführend, weil die Beziehungen zwischen den Individuen verloren gehen. Es können keine XLinks erzeugt werden, die für die Abbildung der Beziehungen erforderlich sind. Aufgrund der oben erläuterten Charakteristika gibt es nur eine infrage kommende Darstellungsform, welche sich für die Erzeugung der GML-Instanzen mit Querverbindungen eignet und zwar ein RDF-Dokument. Zur Veranschaulichung wird auf die Fragen 3.1 und 3.4 verwiesen, die unter GP 6.4 formuliert wurden. Diese lassen sich mit einem einfachen CONSTRUCT-Statement bzw. mit einem simplen Template beantworten (Lis. 9.5). Der Graph (Lis. 9.6) ist die einzig sinnvolle Darstellung (also die RDF-Repräsentation per se und nicht die gewählte Serialisierung). Eine Tabelle, die ebenfalls von Redundanz geprägt wäre, ist nicht geeignet, da auch in diesem Fall die Relationen zwischen den Entitäten von Interesse sind. Aus diesen Gründen muss der WFS-Adapter zunächst weiterentwickelt werden, bevor er im konkreten Fall eingesetzt werden kann.

Der Quellcode von LOD4WFS bietet ein hervorragendes Grundgerüst dafür. Zuvor wurde immer wieder betont, dass sich die Ausführungen auf die vorliegende Arbeit bzw. den konkreten Anwendungsfall beziehen. Damit sollte klargestellt werden, dass hier keine pauschale Aussage zur Geeignetheit von LOD4WFS getroffen wurde. Die Daten und die SPARQL-Statements, die zusammen mit LOD4WFS bereitgestellt wurden, unterscheiden sich von den Daten, die im Rahmen dieser Arbeit verwendet werden und von den SPARQL-Statements (s. GP 9.2). Die Weiterentwicklung des WFS-Adapters ist keine Grundvoraussetzung für dessen Einsatz.

Dies belegt die Arbeit von Vilches-Blázquez and Saavedra (2019), die LOD4WFS unverändert einsetzen. Abb. 3.11 zeigt die Verbindung des WFS-Adapters zum Viewer / zum webbasierten WFS-Client, den Vilches-Blázquez and Saavedra (2019) entwickelten. Diese Webapplikation ist im Zusammenhang mit dem zuvor geschilderten Problem erwähnenswert, weil die Erzeugung von GML-Strukturen mit XLinks alleine nicht ausreichend ist. Um die Vorteile dieses Resultats anhand eines Anwendungsfalls unter Beweis zu stellen, kann nach dem WFS-Client, welcher die XLinks in einer Art und Weise prozessieren kann, die für potenzielle Anwenderinnen nützlich und sinnvoll ist, gesucht werden oder es kann ein (webbasierter) WFS-Client entwickelt werden.

Diesbezüglich wurde unter GP 2.3.3 konstatiert, dass die Vorteile von GML-Strukturen mit XLinks aus theoretischer Sicht auf der Hand liegen, aber dass praktisch geprüft werden muss, inwieweit XLinks in einer Applikation zweckgerichtet prozessiert werden können. Dies kann dahingehend präzisiert werden, dass diese Prüfung vor allem dann notwendig ist, wenn die Attribute xlink:role oder xlink:arcrole eingesetzt werden. Im Rahmen der vorliegenden Arbeit konnte der WFS-Client, welcher diese adäquat prozessieren kann, jedenfalls nicht gefunden werden. Für die Demonstration hätte noch eine Applikation (weiter-)entwickelt werden müssen, was ebenfalls nicht mehr erreicht werden konnte. Es bleibt festzuhalten, dass ein wichtiges Problem gelöst wird, indem LGD via WFS bereitgestellt wird, weil (viele) GIS standardkonforme WFS konsumieren können. Das letzte Verb kann alles und nichts bedeuten. Die Fähigkeiten der WFS-Clients müssen zusätzlich zum Dienst bzw. zur GML-Serialisierung betrachtet werden. Ein adäquater Client ist erforderlich, um anhand eines konkreten Anwendungsbeispiels demonstrieren zu können, dass die oben charakterisierten GML-Strukturen für potenzielle GIS-Anwenderinnen tatsächlich einen Mehrwert bieten.

Zu enthusiastisch (und in Ermangelung der erforderlichen Kenntnisse) wurde bei der ursprünglichen Konzeption davon ausgegangen, dass auch solche Probleme im Rahmen dieser Thesis gelöst werden können.

7.1.2 Entscheidungen

Grundsätzliche Entscheidungen

Die letztgenannte Einsicht ist eine (subjektive) Erkenntnis, die im Rahmen dieser Arbeit gewonnen wurde. Wegen der ursprünglichen Konzeption dieser Arbeit wäre ein (objektiver) Erkenntnisgewinn beim Entwurf und beim Aufbau des Systems möglich gewesen. Diese Möglichkeit besteht nicht mehr, da der Fokus neu ausgerichtet wurde: Der Workflow zwecks Generierung des LOD-konformen RDF-Graphen wurde in den Mittelpunkt gerückt; der Entwurf und der Aufbau des Systems nimmt dagegen eine untergeordnete Rolle ein (vgl. GP 1.2 und GP 1.4). Der Grund liegt darin, dass das System realisiert wird, indem Software installiert, konfiguriert und kompiliert wird, die in einschlägiger Literatur bereits hinreichend erörtert wurde (s.u.). Das Gesamtsystem (s. Abb. 7.4) ist letztendlich nur ein Mittel zum Zweck – zur Generierung des Graphen (Transformation, Modellierung, Konsolidierung, Anreicherung). Außerdem wird es für die Demonstration benötigt.

Die Begriffe, die oben eingeklammert sind, bezeichnen Teilschritte im Workflow zwecks Generierung des Graphen. Dieser Workflow wird im folgenden Abschnitt konzipiert (s. GP 7.2). Bezugnehmend auf die vorgestellten Ansätze, Verfahrensweisen und Strategien (vgl. Kapitel 3) werden **eigene** Lösungen entwickelt, die in Abschnitt 8.2 praktisch umgesetzt werden. Es wird *ausführlich* aufgezeigt, welche Entscheidungen getroffen wurden und welche Probleme sich ergaben. Im Gegensatz dazu werden die Ausführungen zum System auf das notwendige Minimum *reduziert*. Die systembezogenen Entscheidungen werden lediglich dokumentiert. In Kapitel 8 werden die relevanten Spezifikationen dokumentiert. Dies dient der Gewährleistung der Nachvollziehbarkeit bzw. Reproduzierbarkeit: Wenn dieselben Applikationen mit identischer Konfiguration verwendet werden, dann werden dieselben Ergebnisse erzielt. Hierfür muss klar sein, welche Versionen verwendet wurden.

Systembezogene Entscheidungen

Abb. 7.3 zeigt lediglich einen Teil des Servers. Die folgende finale Darstellung dient als vereinfachte Übersicht zum Gesamtsystem. Anschließend werden die systembezogenen Entscheidungen bündig dokumentiert.

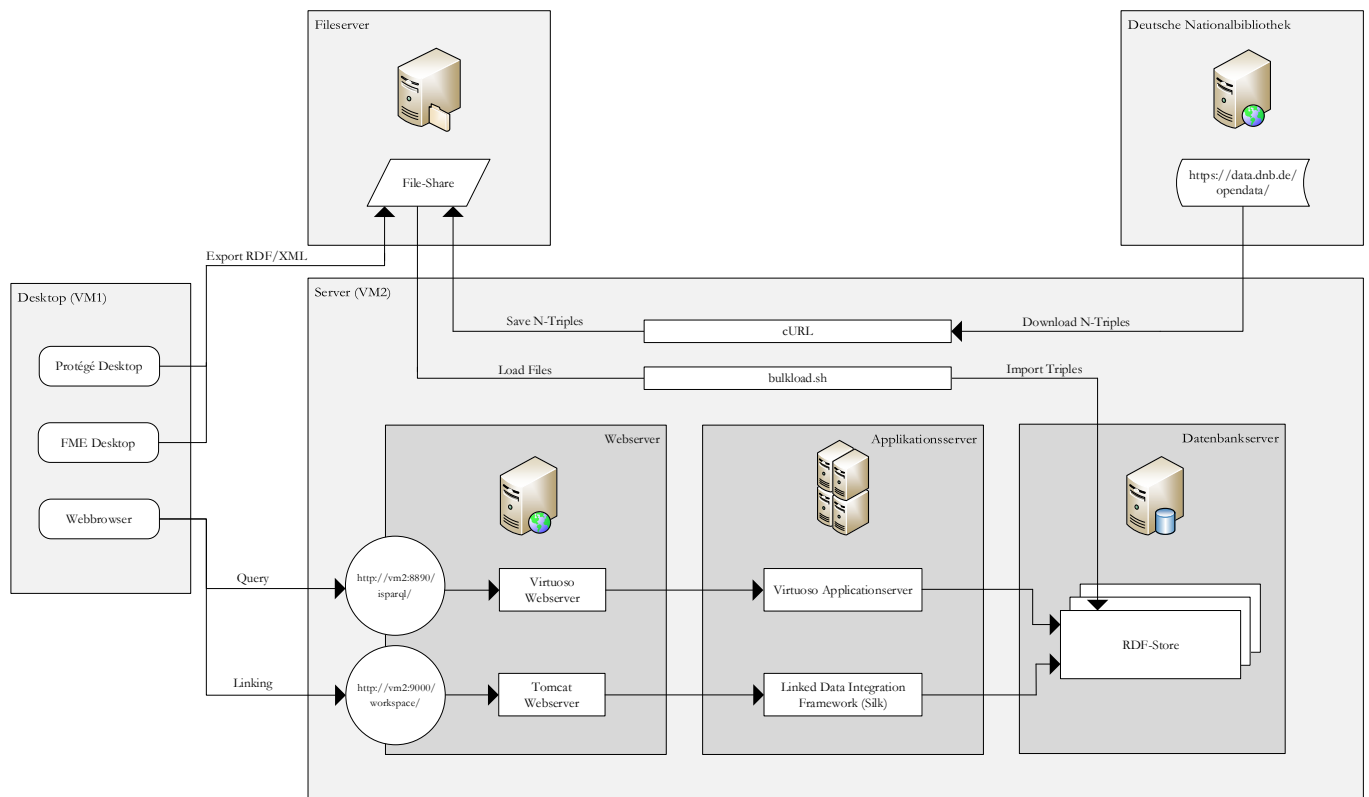


Abb. 7.4: Entwurf des Systems (letzter Teil), ED

- (1) Es werden zwei VMs installiert. In VM1 werden Desktop-Anwendungen eingesetzt; VM2 fungiert als Server.

Entscheidungen zu VM2:

- (2.1.1) Es ist sinnvoll VM2 auf weitere VMs aufzuteilen. Dies kann zu einem späteren Zeitpunkt erfolgen.
- (2.1.2) Als Betriebssystem wird SUSE Linux Enterprise Server (SLES)⁹⁸ verwendet.
- (2.1.3) Für die Virtualisierung wird VMware vSphere⁹⁹ eingesetzt.
- (2.2.1) Für die Speicherung, Verwaltung und Konsolidierung aller Tripel/Quads wird der RDF-Store von Virtuoso Universal Server (Open-Source Edition – OpenLink Virtuoso, kurz: VOS)¹⁰⁰ eingesetzt. Neben einem RDF-Store bietet VOS viele weitere Komponenten (RDBMS, ORDBMS, Fileserver, HTTP-Server uvm.). Der RDF-Store bietet GeoSPARQL-Unterstützung.

⁹⁸ <https://www.suse.com/de-de/products/server/>

⁹⁹ <https://www.vmware.com/de/products/vsphere.html>

¹⁰⁰ <http://vos.openlinksw.com/owiki/wiki/VOS>

Die Besonderheiten, die es dabei zu beachten gilt, wurden von Jovanovik et al. (2021) herausgearbeitet. Informationen zu möglichen Alternativen finden sich sowohl in der Arbeit von Jovanovik et al. (2021) als auch in der Arbeit von Nishanbaev et al. (2020).

- (2.2.2) Für die Formulierung der SPARQL-Abfragen wird die Web-Applikation iSPARQL¹⁰¹ als Virtuoso Application Distribution (VAD) in Kombination mit dem HTTP-Server eingesetzt, der in VOS integriert ist. Das HTML-Interface kann (muss aber nicht) verwendet werden. Vor dem VOS-Webserver wird der Apache HTTP Server¹⁰² als Reverse-Proxy eingesetzt (nicht abgebildet). An den SPARQL-Endpoint können Anfragen über beliebige SPARQL-Clients gestellt werden; z.B. via Client for URLs¹⁰³ (cURL).
- (2.2.3) Abb. 7.3 enthält orangefarbige gestrichelte Linien, die optionale Komponenten repräsentieren. Auf deren Einsatz wird verzichtet. Die Demonstration erfolgt mithilfe von iSPARQL.
- (2.3) Für die Anreicherung wird Silk in Kombination mit Apache Tomcat¹⁰⁴ eingesetzt. Mithilfe der Silk Workbench werden URI-Aliase oder URIs, welche auf ein ähnliches Individuum referieren, in den Informationsbeständen der DNB (Titeldaten, Gemeinsame Normdatei) ermittelt und mit ausgewählten Ressourcen aus dem historischen Häuserbuch verknüpft. Für Details zu Silk wird auf die Arbeiten verwiesen, die unter GP 3.2.3 genannt wurden.

Entscheidungen zu VM1:

- (3.1.2) Als Betriebssystem wird Windows¹⁰⁵ verwendet.
- (3.1.3) Für die Virtualisierung wird Oracle VM VirtualBox¹⁰⁶ eingesetzt.
- (3.2.1) Die o.g. Webanwendungen werden über einen beliebigen Webbrowser genutzt, der in VM1 installiert wird.
- (3.2.2) In VM1 werden Protégé Desktop und FME Desktop eingesetzt. Dies wird ausführlich im folgenden Abschnitt behandelt.

Die vorgenannten Desktop-Anwendungen sind abgebildet, damit die Entscheidungen bezüglich des Daten- bzw. Informationsaustauschs dokumentiert werden können:

- (3.3.1) Dafür wird ein Austauschverzeichnis genutzt, welches von einem virtuellen Fileserver bereitgestellt wird (SMB-Share). Dort werden nicht nur die Resultate temporär gespeichert, die mithilfe der o.g. Desktop-Anwendungen erzeugt werden, sondern hier werden auch die Informationen (RDF-Dumps) zwischengespeichert, die von der DNB zum Download angeboten werden. Der Download erfolgt via cURL. Alle Serialisierungen werden mithilfe eines Shell-Skripts und den Bulk-Loader-Funktionen von VOS in den RDF-Store importiert¹⁰⁷.
- (3.3.2) Aufgrund der Replikation muss ein regelmäßiger Abgleich erfolgen (nicht abgebildet).

¹⁰¹ <https://github.com/openlink/iSPARQL>
<https://demo.openlinksw.com/isparql/>

¹⁰² <https://httpd.apache.org/>

¹⁰³ <https://curl.se/>

¹⁰⁴ <https://tomcat.apache.org/>

¹⁰⁵ <https://www.microsoft.com/de-de/windows/windows-10-specifications>

¹⁰⁶ <https://www.virtualbox.org/>

¹⁰⁷ <http://vos.openlinksw.com/owiki/wiki/VOS/VirtBulkRDFLoader>

Unter GP 5.2 wurde erwähnt, dass die DNB kein SPARQL-Endpoint anbietet¹⁰⁸. Dort wurde auch der Grund genannt, warum die Replikation (kurz) zu thematisieren ist – „One might argue that this is against the idea of linked data, in which data should be stored decentralized and rather not stored as a local copy“ (Homburg, 2020, p. 9). Die DNB empfiehlt einen aktuellen Grundbestand (vgl. Tab. 8.11) in die eigene DB zu importieren und weist darauf hin, dass die Synchronisation serverseitig über die OAI- (Open Archives Initiative) Schnittstelle realisiert werden kann¹⁰⁹. Die folgende Darstellung zeigt, dass hierzu ein OAI-Harvester¹¹⁰ eingesetzt werden kann, der in den Grundzügen mit einem CSW-Harvester (Catalogue Service for the Web / Web Catalogue Service) vergleichbar ist.

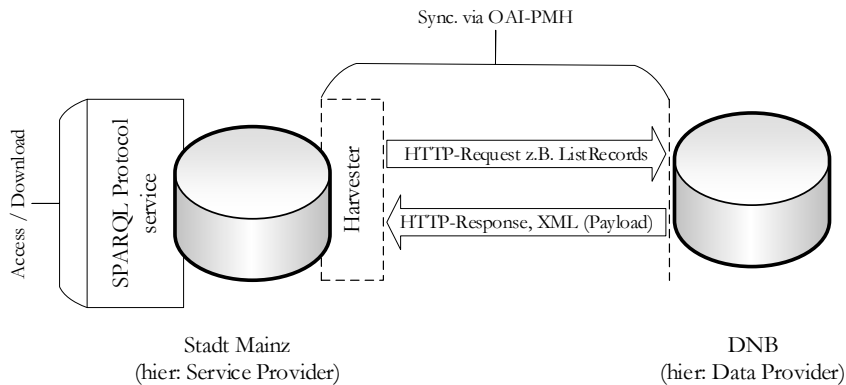


Abb. 7.5: Sync. via OAI-PMH, OA von Diederichs and Wuttke (2018, p. 7), angepasste Darstellung

Die Synchronisation wird im Zuge der Produktivstellung eingerichtet (vgl. GP 10.2). Für die Beantwortung der Leitfrage ist dies nicht erforderlich. Da die (LOD-)Grundlagen im zweiten Kapitel jedoch ausführlich besprochen wurden, ist der kurze Hinweis, dass die Synchronisation via OAI-PMH beim Entwurf des Systems berücksichtigt wurde, statthaft.

¹⁰⁸ Stand: November 2020

¹⁰⁹ Siehe: https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html

¹¹⁰ Technische Details: https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html

7.2 Konzeption des Workflows

Die folgende Graphik stellt die operativen Teilziele und damit die gesamte Verfahrensweise zwecks Beantwortung der Leitfrage überblicksmäßig dar. Da dieser Abschnitt dem zweiten Teilziel dediziert ist, wurde der Workflow zwecks Generierung des Graphen hervorgehoben. Der Workflow wird nachfolgend konzipiert und abschließend detailliert visualisiert (vgl. Abb. 7.8).

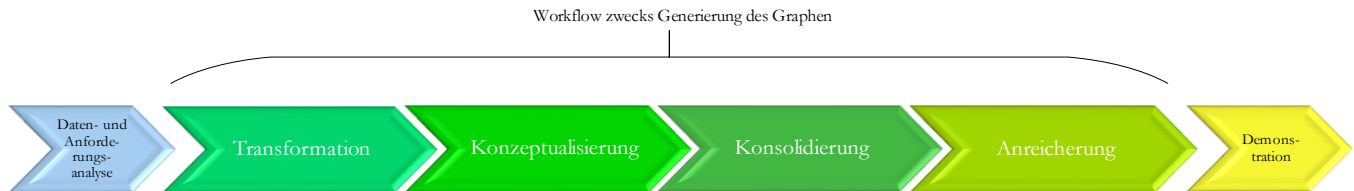


Abb. 7.6: Verfahrensweise zwecks Beantwortung der Leitfrage, ED

7.2.1 Transformation

Entscheidungen zur Umsetzung

Basierend auf den Erkenntnissen, die im Rahmen der Datenanalyse gewonnen wurden und beziehend auf die definierten Anforderungen, werden folgende Entscheidungen getroffen:

- Im Rahmen der Transformation muss jeder Datensatz (Tupel) aus den MapInfo-Relationen (die in Tab. 5.1, Tab. 5.4, Tab. 5.5 dokumentiert wurden) in Tripelform beschrieben werden. Die Werte (Attributwerte) aus den folgenden Spalten (Attribute) können dabei **unverändert** übernommen werden:
 - Jahr (vierstellige Ganzzahl)
 - Bild 1-5 (variable Zeichenketten, die syntaktisch korrekte URIs sind)
 - Beschreibung, Bauinformationen, Hausname, Name (unstrukturierte, variable Zeichenketten)
- Sämtliche Geometrien (Polygone, MultiPolygone, MultiLineStrings) müssen **validiert** und ggf. **modifiziert** werden. Die Koordinatentransformation ist ebenfalls erforderlich.
- Die übrigen Attribute müssen **manipuliert** werden (Duplikaterkennung und -eliminierung, Umbenennung, Beseitigung von Inkonsistenzen):
 - Literatur (variable Zeichenketten, die ein distinktives Muster aufweisen)
 - Bewohner, Eigentüemer (variable Zeichenketten, die kein distinktives Muster aufweisen)
 - Nutzung, Type (Zeichenketten, die insgesamt eine Menge von festen Ausprägungen bilden)
- Die jeweiligen Werte müssen zur Transformation in virtuelle Tabellen ausgelagert werden. Dabei müssen Surrogatschlüssel erzeugt werden, weil die ursprünglichen Relationen keine *geeigneten* Schlüsselkandidaten enthalten. Referenzen müssen über „Fremdschlüssel“ erfolgen, sodass die Zuordnung gewährleistet bleibt.
- Die Reihenfolge (der Zeilen und Spalten in der Relation), bleibt unberücksichtigt, weil diese für den Graphen keine Relevanz hat. Die MapInfo-Metadaten werden nicht übernommen, da diese für die RDF-Repräsentationen keine Bedeutung haben.
- Es wird die URI-Strategie von van den Brink et al. (2014) umgesetzt, die unter GP 3.1.2 vorgestellt wurde. Aufgrund des Umfangs der vorliegenden Arbeit wird die Webserver-Konfiguration nicht mehr behandelt; die 303-URIs werden im Rahmen des ETL-Prozesses erzeugt.

- Der Graph wird zunächst serialisiert (via RDF/XML) und im Filesystem (FS) gespeichert; der Import in den RDF-Store erfolgt im Zuge der Konsolidierung (Schritt 3). Für die „Validierung“ wird der „W3C RDF Validation Service“ verwendet¹¹¹. Damit wird primär die Wohlgeformtheit überprüft; die Sinnhaftigkeit wird stichprobenartig anhand repräsentativer RDF-Fragmente untersucht, die vom o.g. Dienst graphisch dargestellt werden können.
- Auf der Instanzebene fehlen notwendige Angaben häufig. Diese werden nicht manuell nacherfasst, sondern ggf. über leere Knoten dargestellt. Die vorhandenen Relationen werden semi-automatisch zu einem RDF-Graphen umgeformt, wobei der manuelle Teil auf das notwendige Minimum reduziert wird. Die händische Arbeit ist z.B. im Rahmen der Duplikatentfernung unvermeidlich bzw. wegen der angestrebten Anreicherung zwingend notwendig (GP 7.2.4).

RDF-izer

Vor diesem Hintergrund kann die Geeignetheit der vorgestellten RDF-izer diskutiert werden. Zunächst bleibt festzuhalten, dass weder TripleGeo noch GeoTriples die vorliegenden MapInfo-Relationen (in der dokumentierten Version) einlesen können. Dies ist allerdings auch beim Verzicht auf den Einsatz eines (O-)RDBMS kein Hinderungsgrund für die Verwendung dieser Applikationen, da jede Relation lediglich eine Menge homogener Geometrien enthält. Daher können die Relationen automatisch in Shapefiles konvertiert werden, welche von beiden o.g. RDF-izern gelesen werden können. Außerdem können beide Tools vollständige RDF-Repräsentationen aus den Vektormodellen generieren – es können nicht nur Geometrien umprojiziert, serialisiert und mithilfe des GeoSPARQL-Vokabulars beschrieben werden. Prinzipiell können auch alle thematischen Attribute unter Verwendung weiterer Vokabulare in Tripelstrukturen überführt werden. Wie dargelegt, soll hierfür auf Konzepte aus einigen Vokabularen zurückgegriffen werden, welche sich in der LOD-Cloud etabliert haben. Deshalb *wäre* der Applikation GeoTriples der Vorzug zu geben, weil die Adaption der individuell generierten RML- / R2RML-Mappings *effizienter* ist, als die Anpassung der Konfigurationsdateien und XSLT-Dokumente, die TripleGeo für die Umformung benötigt. Die Software GeoTriples ist ein sehr mächtiges und hilfreiches Werkzeug, um aus Geodaten, die in sinnvollen (Relations-)Strukturen vorliegen, einen RDF-Graphen zu erzeugen (ggf. sogar ohne die Verwaltung mithilfe eines ORDBMS aufzugeben). In der vorliegenden Arbeit wird jedoch der Frage nachgegangen, ob auch aus nichtnormalisierten Relationen ein LOD-konformer RDF-Graph geformt werden kann, der einen erkennbaren Mehrwert gegenüber den Ausgangsrelationen bietet. Dementsprechend wird auf die vorherige Normalisierung verzichtet. Absolut notwendige Datenbereinigungen werden im Rahmen der (semantischen) Transformation durchgeführt. Dafür sind Spatial-ETL-Werkzeuge wie FME oder GeoKettle prädestiniert. Die Erzeugung eines LOD-konformen RDF-Graphen durch Anpassung von RML- / R2RML-Mappings, die auf Basis von Relationen generiert wurden, welche überwiegend mehrwertige, zusammengesetzte Attribute und Wiederholungsgruppen enthalten, ist zwar ebenfalls möglich aber *ineffizient*.

Wichtig ist, dass der resultierende RDF-Graph anschließend sinnvoll via SPARQL abgefragt werden kann, um demonstrieren zu können, dass nun nutzbare Antworten auf konkrete Fragen gegeben werden können, die sich anhand der ursprünglichen MapInfo-Relationen nicht beantworten lassen. Dafür muss zunächst die DQ im Rahmen des ETL-Prozesses verbessert werden. Es ist zu erwarten, dass einige Prozessschritte mehrfach durchlaufen werden müssen, um die Parameter iterativ anpassen zu können. Dabei nimmt die Fehlerbehandlung und -untersuchung eine entscheidende Rolle ein. Eine Software, die eine GUI mit geeigneten Funktionen, z.B. zwecks visueller Analyse, Debugging, Logging etc. bereitstellt, ist für die praktische Umsetzung besonders gut geeignet, da sich die Datenbereinigung damit effizient zur Laufzeit bewerkstelligen lässt. FME bietet, wie unter

¹¹¹ <https://www.w3.org/RDF/Validator/documentation>

GP 8.2.1 detailliert dargelegt, alle erforderlichen Funktionen. Damit ist die Verwendung von Geometry2RDF nicht erforderlich, da sich die Geometrien ebenso mittels Transformern und Zeichenkettenfunktionen serialisieren lassen, die in FME standardmäßig zur Verfügung stehen. Der ETL-Prozess kann mit FME gemäß den individuellen Anforderungen entwickelt werden, ohne dass dafür vorhandene Codestrukturen angepasst werden müssen. Die Zieldatenstruktur kann flexibel gestaltet werden. Deshalb wird FME ausgewählt – der erste Teilschritt kann mit verhältnismäßig geringem Aufwand mithilfe einer individuell zugeschnittenen FME-Workbench umgesetzt werden.

Dies gilt im konkreten Fall nicht für die Umformung mittels XSLT. Abgesehen davon, dass die skriptbasierte Fehlerbehandlung wesentlich zeitaufwendiger ist, bringt der Zwischenschritt über GML in semantischer Hinsicht keinen Mehrwert, weil Ausgangsrelationen für die (automatische) Generierung einer sinnvollen XSD nicht geeignet sind. Die XSD bietet damit auch keine Grundlage für die automatische Ableitung einer Applikation-ontologie über Transformationsregeln. Insofern kann sich diese Arbeit zwecks Konzeptualisierung nicht auf den Ansatz von van den Brink et al. (2014) stützen. Der andere Ansatz von van den Brink et al. (2014) zur Generierung von Tripeln auf der Instanzebene kann jedoch auch ohne die vorherige, automatische Konvertierung der MapInfo-Relationen in GML-Repräsentationen aufgegriffen werden. Vor dem Hintergrund der vorstehenden Ausführungen muss das Verfahren aber etwas anders umgesetzt werden:

- Da die syntaktische Umformung nicht ausreichend ist (s.o.), wird keine automatische Transformationsroutine, sondern ein semi-automatischer ETL-Prozess entwickelt (z.B. werden Haltepunkte zwecks manueller Kontrolle verwendet).
- Die Definition der Zieldatenstruktur (hierarchische Baumstruktur) in FME wird ähnlich zur XSL Transformation umgesetzt (XML-Template / Serialisierung mittels RDF/XML); die Kontrollstrukturen werden mithilfe von XQuery definiert.

7.2.2 Konzeptualisierung

Ursprünglich war lediglich die Nachnutzung existierender Konzeptualisierungen vorgesehen. Das XML-Template, welches unter GP 8.2.1 besprochen wird, zeugt von diesem Bestreben, da viele Konzepte wiederverwendet werden. Wie erwähnt, werden ausschließlich Vokabulare nachgenutzt, die sich im Web of Data etabliert haben. Dies ist prinzipiell sinnvoller, als die Konzepte selbst zu spezifizieren; den Hauptgrund nannten Heath and Bizer (2011): „[...] it maximises the probability that data can be consumed by applications that may be tuned to well-known vocabularies [...]“. Die Absicht vorzugsweise einfache Ontologien zu verwenden, gründet auf dem Prinzip der geringsten Ausdrucksmächtigkeit. Exemplarisch kann das FOAF-Vokabular genannt werden (kommt mit RDFS-Plus aus); die nachgenutzten Konzepte aus dem Geo-Vokabular des GeoSPARQL-Standards basieren auch nicht auf komplexen OWL-Konstrukten¹¹². Die Komplexität des Graphen hätte sich erhöht, wenn nur Vokabulare zum Einsatz gekommen wären, die in der LOD-Cloud gefunden werden können. Dies soll nicht suggerieren, dass im Web keine Vokabulare vorhanden sind, mit denen sich alle Konzepte aus dem historischen Häuserbuch mit einfachen Sprachmitteln beschreiben lassen. Es soll lediglich klargestellt werden, dass nicht das gesamte Web nach geeigneten Vokabularen durchsucht werden konnte. Drei Beispiele für Vokabulare, die ermittelt werden konnten, sind die GND Ontologie, die DBpedia Ontologie und das CIDCOC CRM mit den relevanten Erweiterungen (vgl. Tab. 5.7). Die GND Ontologie oder die DBpedia Ontologie hätte als Basis für die Konzeptualisierung dienen können. Die benötigten Klassen und Prädikate können teils direkt aus dem entsprechenden Vokabular entnommen werden.

¹¹² owl:disjointWith ist eine Ausnahme, die den Kern der Aussage nicht ändert, weil die verwendeten Konzepte ansonsten ohne komplizierte Sprachmittel auskommen.

In einigen Fällen müssen sie aber in Form von Erweiterungen bzw. Spezialisierungen definiert werden. RDFS-Plus ist dafür nicht ausreichend. Die unbeabsichtigten Schlussfolgerungen müssen mithilfe von OWL-Konzepten vermieden werden, z.B. mittels OWL-Restriktionen. Diese basieren auf Beschreibungslogiken als Einschränkung der Prädikatenlogik erster Stufe (Scheider et al., 2014). Ein solches Vorgehen widerspricht der o.g. Intention, weil ein derart komplexes Semantic Web Dokument für die Beschreibung der vorhandenen Ressourcen momentan nicht erforderlich ist. Anders ausgedrückt: „The selected ontology (if any) is reused, and then an adaptation process is started in order to cope with the implicit requirements from an ontology project. This scenario is costly in many cases [...]“ (Gangemi and Presutti, 2009, p. 223). Die Verknüpfung bzw. das *direkte* Mapping unter Verwendung des CIDCO CRM ist ebenfalls nicht zielführend. Es fehlt an einer einfachen konzeptuellen Grundlage. Der zweite Teilschritt im Workflow ist deshalb nur für die Modellierung einer simplen TBox vorgesehen. Perspektivisch ist das Mapping mit den Konzepten aus dem CIDOC CRM sicherlich ebenso sinnvoll, wie die Nachnutzung der Konzepte aus dem SKOS. Vorher sollte aber klar sein, welche ET dabei relevant sind und in welcher Beziehung diese stehen. Die „Spaghetti-Daten“ (vgl. GP 5.1.1) werden allerdings nicht in den ETL-Prozess eingebunden. Aus diesen Daten müssen *zuerst* sinnvolle Informationsobjekte gebildet werden, sodass sie *wenigstens taxonomisch* geordnet werden können. Erst nach der Vervollständigung und/oder der Adaption der Klassenhierarchie sollte die Verknüpfung mit den Konzepten aus der o.g. Domänenontologie erfolgen.

Ziel, Werkzeug

Um die zuvor definierten Anforderungen trotzdem erfüllen zu können, wird zunächst das folgende Grobziel definiert: Es muss eine einfache Ontologie erstellt werden, die für zukünftige Erweiterungen als Basis dienen soll. Der Flexibilität wird eine höhere Priorität eingeräumt, als der Komplexität, die durch die folgenden Entscheidungen auf das momentan notwendige Minimum reduziert wird:

- Es sollen nur Konzepte aus den eingangs erwähnten Vokabularen spezialisiert werden. Gangemi and Presutti (2009, p. 224) betonten: „[...] the success of very simple and small ontologies like FOAF [...] shows the potential of really portable, or “sustainable” ontologies“. Weil das Geo-Vokabular aus dem GeoSPARQL-Standard ebenfalls simpel, kompakt und wegen den zahlreichen darauf aufbauenden Domänenontologien offensichtlich erfolgreich ist, wurde es neben FOAF zu Beginn erwähnt.
- Es sollen nur erforderliche Konzepte beschrieben werden. Der Begriff Konzept wurde unter GP 2.2.2 für diese Arbeit wie folgt definiert: „alle Klassen, Prädikate und Datentypen, die erforderlich sind, um individuelle, historische Entitäten und ihre Beziehungen untereinander auf der Instanzebene beschreiben zu können, sodass sie maschinell verarbeitet und von menschlichen Nutzerinnen verstanden werden können“. Erforderlich sind nur Konzepte,
 - die nicht unverändert aus etablierten Vokabularen übernommen werden können und
 - die einen Bezug zu den MapInfo-Relationen aufweisen, die in den ETL-Prozess eingebunden werden¹¹³.
- Die Klassenhierarchie soll sich an den Vorschlägen von Scheider et al. (2014) orientieren, aber auf die momentan notwendigen Klassen reduziert werden. Die Vorschläge von Scheider et al. (2014) zwecks Spezifikation der Prädikate sollen übernommen werden, sofern diese ohne OWL-Restriktionen auskommen, weil der Sprachumfang auf RDFS-Plus-Konstrukte beschränkt werden soll.
- Klassenüberschneidungen (KS) und Klassenvereinigungen (KV) können also nicht exakt modelliert, sondern nur approximiert werden. Es sollen keine anonymen Klassen konstruiert werden. Deshalb

¹¹³ Die TBox wird daher keine Konzepte wie „Kurtine“ oder „Kontergarde“ enthalten, weil die entsprechenden MapInfo-Relationen noch nicht transformiert werden können.

muss auf die OWL-Konstrukte `owl:onProperty`, `owl:allValuesFrom`, `owl:someValuesFrom` und `owl:hasValue` verzichtet werden. Sämtliche Aussagen mit dem Inhalt, dass ein bestimmtes historisches Gebäude von einer bestimmten historischen sozialen Gruppe genutzt wurde, finden sich in der ABox. Die Generierung von leeren Knoten kann auf die Instanzebene beschränkt werden (Details unter GP 8.2.1).

- Wegen der o.g. Einschränkungen lassen sich einige Aussagen nicht maschinell schlussfolgern. Deshalb sollen auch Prädikate ohne formale Semantik eingesetzt werden. Diese dienen nicht dazu, die Wissensbasis um logische Aussagen anzureichern, sondern die Literale (welche von den vorgenannten Prädikaten referenziert werden) sind für die Interpretation durch einen Menschen bestimmt.

Bezugnehmend darauf, kann das o.g. Grobziel wie folgt konkretisiert werden. Im Rahmen des zweiten Teilschritts soll der erzeugte Graph durch eine einfache TBox ergänzt werden. Die TBox muss alle erforderlichen Konzepte enthalten, die in Tab. 5.1 - Tab. 5.5 gelistet oder beschrieben wurden. Dies sind:

- Alle **Klassen**, welche in eine Hierarchie eingeordnet werden müssen, damit die Typen von historischen Geoobjekten in der ABox angegeben werden können. Die TBox muss alle Klassen enthalten, die wegen der Werte in der Spalte „Nutzung“ (originäre MapInfo-Relationen, siehe Tab. 5.1) und wegen der Werte in der Spalte „Type“ (adaptierte MapInfo-Relationen, siehe Tab. 5.4 und Tab. 5.5) erforderlich sind.
- Alle **Properties**, die zwar ebenfalls in eine Hierarchie eingeordnet werden können, aber für die (im Gegensatz zu den Klassen) Einschränkungen definiert werden müssen. Dies erfolgt durch die Festlegung der Definitions- und Wertebereiche, sodass die terminologischen Zusammenhänge sowie die Beziehungen zwischen Objekten in der ABox angegeben werden können. Außerdem soll angegeben werden, ob das Objekt, welches von einem Prädikat referenziert wird, auch Subjekt sein kann oder ob es ein `getyptes/ungetyptes` Literal ist. Die TBox muss alle Prädikate enthalten, die nicht aus etablierten Vokabularen übernommen werden können.

Für die Realisierung des zweiten Teilschritts werden die folgenden Programme eingesetzt:

- Die Modellierung der einfachen TBox erfolgt mithilfe der Funktionen von Protégé Desktop, die über eine GUI ausgeführt werden können. Tab. 8.2 enthält die Details zur Version.
- Für die Überprüfung der Schlussfolgerungen wird die Inferenzmaschine (alternative Bezeichnung: Reasoner) HermiT verwendet. Dieses Programm ist in Protégé Desktop bereits vorinstalliert.
- Die visuelle Dokumentation erfolgt mithilfe des Plugins OntoGraf. Weil sich damit nicht alle Informationen (sinnvoll) abbilden lassen, erfolgt die Dokumentation zusätzlich in textueller Form. Die einfache TBox wird über die OWL API als RDF/XML-Serialisierung exportiert (s. Anhang). Die Exportfunktion ist ebenfalls Bestandteil von Protégé Desktop.

7.2.3 Konsolidierung

Die Zusammenführung der RDF-Dumps der DNB mit den Resultaten aus Teilschritt 1 und 2 (RDF/XML-Serialisierungen, die ins Austauschverzeichnis exportiert und „validiert“ werden) ist der einfachste Prozess in diesem Workflow. Mithilfe des RDF-Stores werden benannte RDF-Graphen für das historische **Häuserbuch**, die **Titeldaten** und die **Normdaten** angelegt, um diese in SPARQL-Abfragen referenzieren zu können. Die Konzeptualisierungen werden ebenfalls im RDF-Store verwaltet, um implizite Aussagen ableiten zu können.

7.2.4 Anreicherung

Priorität des Teilschritts

Der letzte Teilschritt im Workflow (Anreicherung) ist der Wichtigste, um die Leitfrage der vorliegenden Arbeit beantworten zu können. Es wird dokumentiert und demonstriert, dass der Informationsumfang des Modells durch die Erzeugung von *RDF-Links auf der Instanzebene* erhöht werden kann. Prinzipiell wäre dies schon nach dem Import der RDF/XML-Serialisierungen in den RDF-Store möglich. Durch die Transformation und die Konzeptualisierung wird der Wert des historischen Häuserbuchs gesteigert. Problematisch ist aber, dass die Menge an konkreten Informationen überschaubar bleibt. So könnte z.B. ermittelt werden, welcher Klasse ein bestimmtes Individuum angehört, auf das in den Ausgangsdaten lediglich über eine unstrukturierte Zeichenkette referiert wird (diese Information gibt es dort nicht). Allerdings sind diese Individuen häufig Instanzen unspezifischer Klassen (z.B. **foaf:Agent**), da die Ausgangsdaten oft keine exakte Differenzierung zulassen. Außerdem eignet sich die simple TBox trotz der Verknüpfungen mit anderen Vokabularen (über *Vocabulary Links*) nur für einfache Schlussfolgerungen. Daher soll der Mehrwert nach der Generierung von RDF-Links (u. a. *Identity Links*) auf der Instanzebene demonstriert werden. Dann lassen sich tatsächlich nutzbare Antworten auf konkrete Fragen geben, die sich anhand der ursprünglichen MapInfo-Relationen nicht beantworten lassen.

Mehrwert durch Verlinkung

Im Rahmen des ETL-Prozesses wird jede Nicht-Informationsressource, die nicht über einen leeren Knoten repräsentiert wird, über eine NIR-URI identifiziert (siehe Anforderungen zur URI-Allokation). Da im Zuge der Transformation jedoch nur die notwendigsten Datenbereinigungen durchgeführt werden (s.o.), enthalten die zugehörigen Beschreibungen meist noch die ursprünglichen Daten (z.B. zusammengesetzte Attribute, die etwa den Titel, den Vor- und Nachname einer historischen Persönlichkeit angeben). Diese liegen nach der Umformung jedoch in Tripelform vor. Dadurch besteht nicht nur die Möglichkeit, die Menge an Aussagen über ein bestimmtes Individuum zu vergrößern bzw. zu ergänzen, sondern durch die Verlinkung mit den qualitativ hochwertigen Informationen der DNB könnten die vorgenannten strukturellen Probleme zumindest relativiert werden. Falls dies gelingt, erübrigt sich z.B. die Formulierung komplexer SPARQL-Abfragen, um bestimmte Teilzeichenketten aus dem Häuserbuch zu extrahieren. Stattdessen könnte der RDF-Link verfolgt werden. Auf dieser Basis könnten z.B. kreative Mashup- / LOD-Applikationen operieren, die vom Alleinstellungsmerkmal des historischen Häuserbuchs (umfangreiche geometrische und thematische Informationen über unterschiedliche Zeitabschnitte der Mainzer Stadtgeschichte) profitieren und dank des LOD-Angebots der DNB noch weitere strukturierte Informationen zu bestimmten Entitäten liefern könnten. Die Herausforderung besteht darin, im Rahmen der Anreicherung möglichst viele sinnvolle RDF-Links auf Basis der o.g. Zeichenketten zu erzeugen. Dafür muss festgelegt werden, welche Prädikate verwendet werden sollen, um die jeweiligen NIR-URIs in eine sinnvolle Relation zu setzen – vor allem dann, wenn sie nicht auf das identische Objekt referieren.

Ziel

Definition:

- Der Wert der generierten RDF-Repräsentationen wird in quantitativer Hinsicht (Erhöhung des Informationsumfangs) gesteigert, indem für ausgewählte Features (Literatur, Agenten, Geoobjekte) aus dem Häuserbuch, **semi-automatisch**, d.h. mittels **Algorithmen** (Zeichenkettenähnlichkeitsfunktionen) und durch manuelle Analyse bzw. Interpretation, jeweils **ein sinnvoller RDF-Link** generiert wird, der eine Ressource in den Informationsbeständen (Titeldaten, GND) referenziert, welche von der DNB gemäß der LOD-Prinzipien publiziert wurden. Die Zielerreichung wird quantifiziert, indem auf die Anzahl der erzeugten RDF-Links abgestellt wird. Die Zielerreichung wird dokumentiert, indem alle RDF-Links für alle Instanzen einer bestimmten Klasse beigefügt werden. In Kombination mit der exakten Dokumentation des Verknüpfungsprozesses wird die Reproduzierbarkeit gewährleistet.

Erläuterungen und Entscheidungen:

- **ein sinnvoller RDF-Link:** Jedes der o.g. RDF-Fragmente soll lediglich um ein zusätzliches Tripel erweitert werden. Jedes Feature wird nur mit einer weiteren Ressource verknüpft. Es werden nur ausgehende RDF-Links erzeugt. Die Semantik der infrage kommenden **Prädikate** soll nicht unberücksichtigt bleiben (s.u.).
- **semi-automatisch:** Der Verknüpfungsprozess besteht aus mehreren Teilschritten, die nicht automatisch durchgeführt werden können. Die potenziellen Kandidaten für die Verlinkung sollen zwar mithilfe von SPARQL-Abfragen und Algorithmen ermittelt werden, sie müssen anschließend aber manuell analysiert bzw. interpretiert werden, um eine qualifizierte Entscheidung treffen zu können, ob die Verknüpfung überhaupt sinnvoll ist. In diesem Zuge muss ggf. ein geeignetes Prädikat ausgewählt werden. Im konkreten Fall ist die manuelle Analyse für die Generierung zweckmäßiger RDF-Links zwingend notwendig.
- **Algorithmen:** Problematisch ist, dass die Referenz auf die Agenten und Publikationen in den Ausgangsdaten über Namens- / Titelinformationen erfolgt (s. Tab. 5.1 und GP 5.1.2). Im Gegensatz zu URIs sind diese Zeichenketten für die eindeutige Identifikation der Individuen größtenteils ungeeignet. Die Geoobjekte sind in den Ausgangsdaten zwar direkt über Koordinaten verortet, was die Identifikation wesentlich vereinfacht; allerdings fehlen diese räumlichen Angaben (im konkreten Fall) in der GND fast immer (s. GP 8.2.4). Die Verknüpfung ist nur möglich, indem die Ähnlichkeit von Zeichenketten (primär Namen) verglichen wird, welche auf die jeweiligen Bezugsobjekte referieren. Dafür müssen Zeichenkettenähnlichkeitsfunktionen verwendet werden (z.B. Token-basierte Ähnlichkeitsalgorithmen).
- **Prädikate:** Zur Beurteilung der Identität wird nicht auf das invertierte Identitätsprinzip abgestellt (s. GP 3.1.4). Anderenfalls müsste auf die Verwendung des Prädikats **owl:sameAs** bzw. die Generierung von *Identity Links* vollständig verzichtet werden. In Anbetracht des o.g. Ziels wird es niemals der Fall sein, dass die zu verknüpfenden Ressourcen exakt die dieselben Eigenschaften aufweisen. Das Prädikat **owl:sameAs** wird nicht im Sinne der spezifizierten Semantik verwendet. Stattdessen wird Heath and Bizer (2011) gefolgt; die *Identity Links* werden als Behauptungen/Meinungen aufgefasst und nicht als objektive Aussagen zur Identität. Trotzdem wird die Identität der jeweiligen Bezugsobjekte sorgsam überprüft, indem die zugehörigen Beschreibungen untersucht und ggf. weitere Links verfolgt werden. Wenn dabei festgestellt wird, dass die Referenten „ähnlich“ aber nicht „identisch“ sind, werden

sie über das Prädikat `rdfs:seeAlso` verknüpft. Die Entscheidung nicht auf zugehörige Beschreibungen zu verweisen, orientiert sich an den ausgewählten Informationen. Die DNB verwendet `rdfs:seeAlso` ebenfalls in Tripeln, wo Subjekt und Objekt über NIR-URIs identifiziert werden¹¹⁴. Die Properties aus der SO werden nicht verwendet, weil sie sich in der LOD-Cloud nicht etabliert haben. Heath and Bizer (2011) schlugen vor, diesen Aspekt vor dem Einsatz der jeweiligen Prädikate zu prüfen. Die Untersuchung wurde für die vorliegende Arbeit auf die GND, GeoNames, LinkedGeoData und DBpedia¹¹⁵ begrenzt. Dort werden die Properties aus der SO nicht verwendet. Für die Verlinkung werden nur die beiden o.g. Prädikate (`owl:sameAs` und `rdfs:seeAlso`) eingesetzt.

Verfahrensweise

Um das Ziel möglichst effizient erreichen zu können, wird methodisch vorgegangen. Das Diagramm¹¹⁶ auf der nächsten Seite zeigt die Verfahrensweise, die nachfolgend erläutert wird.

Der Verlinkungsprozess beginnt mit der Einrichtung der Verbindung für den Zugriff auf die benannten Graphen, die allesamt im RDF-Store verwaltet werden. Zwecks Ermittlung der potenziellen Kandidaten werden die Inhalte und die Struktur der „Zielgraphen“ (GND, Titeldaten) auf der Instanzebene (ABox) via SPARQL (SELECT, CONSTRUCT, DESCRIBE) grob analysiert. Im Rahmen von Teilprozess 1.1 werden auch die relevanten Ontologien untersucht. Neben der GND-Ontologie wird die „Ontology of agent relationships“ (AgRelOn) einbezogen, die ebenfalls von der DNB erstellt wurde. Es ist zu klären, wie die o.g. Ontologien aufgebaut sind und welche Bedeutung die Termini haben. Die Kenntnis der spezifizierten Semantik (TBox) und der Tripelstruktur (ABox) ist die Grundvoraussetzung, für die Durchführung aller weiteren (Teil-)Prozesse (s.u.).

Anschließend wird die Vorselektion mittels SPARQL durchgeführt, die dazu dient, den Zielgraphen auf Instanzen der relevanten Klassen einzuschränken (Teilprozess 1.2). Ein Beispiel für eine SPARQL-Abfrage findet sich in Abschnitt 8.2.4, wo u. a. die vorgenannten Aspekte (Semantik, Tripelstruktur) detailliert erklärt werden.

Die selektierte Untermenge wird sukzessive reduziert (Teilprozess 1.3). Die dafür notwendigen Literale werden definiert, die erforderlichen Operatoren und Transformatoren werden verknüpft, parametrisiert usw. Die Routine wird so lange adaptiert, bis eine akzeptable Ausgabe erzielt wird. Bei der Ausgabe handelt es sich nicht um eine Ergebnismenge, sondern lediglich um eine sogenannte Kandidatenmenge. In dieser befinden sich die potentiell geeigneten Paare von Ressourcen (Kandidatenpaare), welche für die Verknüpfung in Frage kommen könnten. Für die iterative Entwicklung des Workflows werden keine theoretischen Berechnungen angestellt. Basierend auf den mathematischen Grundlagen wird die prinzipielle Geeignetheit der betreffenden Ähnlichkeitsfunktion geprüft. Anschließend wird die Funktion praktisch getestet, indem die jeweiligen Parameter (ggf. mehrfach) variiert werden (Versuch und Irrtum). Der größte Vorteil der Versuch-und-Irrtum-Methode ist darin zu sehen, dass nicht untersucht werden muss, wie ein bestimmter Algorithmus implementiert wurde. Ein erheblicher Nachteil besteht allerdings darin, dass beim Vergleich längerer oder von mehreren Zeichenketten nicht alle Parameter getestet werden können. Die Spezifikationen von VM2 können Tab. 8.1 entnommen werden; anhand der Links, die unter GP 8.2.3 hinzugefügt wurden, kann nachvollzogen werden, dass der Zielgraph aus insgesamt 684.159.111 Tripeln besteht. Obwohl der Hauptspeicher keinesfalls knapp bemessen ist und ein geeigneter Sekundärspeicher (SSDs) etc. verwendet wird, terminieren bestimmte Algorithmen erst nach vielen Stunden (s. GP 8.2.4), was die möglichen Variationen praktisch limitiert.

¹¹⁴ Beispiel: <https://d-nb.info/standards/vocab/gnd/geographic-area-code#XY>

¹¹⁵ Zwecks Überprüfung kann der SPARQL-Endpunkt verwendet werden: <https://dbpedia.org/sparql/>

¹¹⁶ In Anbetracht des Grads der Automatisierung ist dies die adäquate Darstellungsform.

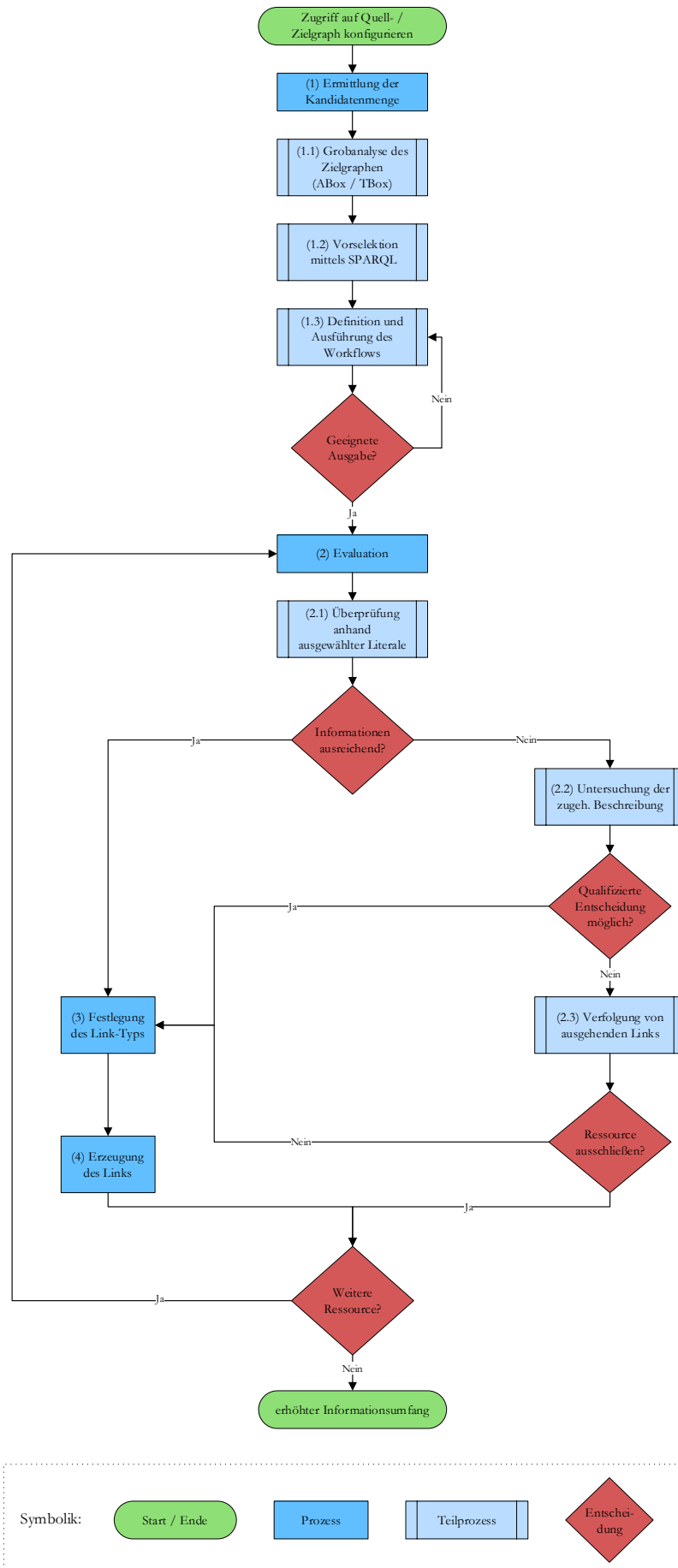


Abb. 7.7: Verfahrensweise für die Generierung von RDF-Links auf der Instanzebene, ED

Die Evaluation (Prozess 2) dient der Bestimmung der tatsächlichen Ergebnismenge. Eine stichprobenartige Überprüfung der potenziellen Kandidaten wird a priori ausgeschlossen, weil die Referenz über die o.g. Literale in den wenigsten Fällen eindeutig ist (z.B. können unterschiedliche Personen den gleichen Namen haben). Stattdessen werden die ermittelten Ressourcen einzeln untersucht, um einerseits das geeignete Prädikat auswählen zu können und andererseits jene Ressourcen aussortieren zu können, die sich fälschlicherweise in der Kandidatenmenge befinden. Die Evaluation wird als mehrstufiger Prozess durchgeführt, um den Aufwand trotzdem möglichst gering zu halten. Zunächst werden die potenziellen Kandidatenpaare (Subjekt und Objekt im Tripel) basierend auf den ausgewählten Literalen analysiert (Teilprozess 2.1). In wenigen Fällen kann auf dieser Grundlage eine qualifizierte Entscheidung getroffen werden, ob es „sinnvoll“ ist, dem Subjekt (im Häuserbuch) und dem Objekt (in den Informationen der DNB) die „Identität“ (über das Prädikat `owl:sameAs`) oder die „Ähnlichkeit“ (über `rdfs:seeAlso`) zu attestieren. Anschließend wird der „Link-Typ“ (nachfolgend entweder als „owl:sameAs-Link“ oder „rdfs:seeAlso-Link“ bezeichnet) festgelegt (Prozess 3), um das jeweilige Tripel schließlich zu erzeugen (Prozess 4). Die RDF-Links werden manuell über Aktualisierungsoperationen eingefügt (SPARQL Update), statt sie zunächst im Dateisystem zwischenspeichern. Bezugnehmend auf die Erkenntnisse, die im Rahmen der Datenanalyse gewonnen wurden, muss an dieser Stelle jedoch klargestellt werden, dass dieser Ablauf primär für die wenigen zusammengesetzten Attribute in Frage kommt, deren einfache Komponentenattribute die eindeutige Identifikation des jeweiligen Bezugsobjekts zulassen. Dies sind vor allem die Titelinformationen, die nach einer bestimmten Systematik erfasst wurden (vgl. Tab. 5.1, GP 5.1.2 oder Anhang). Die Zeichenketten (Referenzen), die als Kombination aus Erscheinungsjahr, Autorin etc. beschrieben wurden, können einfach zerlegt werden. Es ist davon auszugehen, dass sie für die eindeutige Ermittlung des jeweiligen Pendants (in den Informationen der DNB) meist ausreichend sein werden.

Da dies für die übrigen Referenzen i.d.R. nicht anzunehmen ist, sieht der alternative Prozesszweig die manuelle Untersuchung und ggf. die Beschaffung detaillierterer Informationen vor. Dies beschränkt sich zunächst auf die zugehörigen Beschreibungen der DNB (Teilprozess 2.2). Es ist davon auszugehen, dass dies regelmäßig genügt, da die Informationen der DNB nicht nur hochwertig, sondern auch sehr umfangreich sind. Aufgrund des letztgenannten Aspekts muss jedoch eingegrenzt werden, welche Ressourcen händisch überprüft werden sollen und in welcher Reihenfolge eine effiziente Untersuchung durchzuführen ist. Diese Festlegung basiert auf den Erkenntnissen, die im Rahmen von Teilprozess 1.1 gewonnen werden und auf der folgenden Orientierungshilfe, welche der möglichst schnellen sowie eindeutigen Identifikation der jeweiligen Objekte dienen soll:

- (1) Untersuchung raumzeitlicher Aspekte (z.B. Wirkungszeitpunkt/Wirkungszeitraum, geografischer Bezug etc.)
- (2) Analyse von Informationen, welche das jeweilige Individuum näher charakterisieren (z.B. biographische Beschreibungen, Definitionen etc.)
- (3) Interpretation der Beziehungen zwischen Entitäten (etwa Verwandtschaftsbeziehungen, Ordnungsprinzipien von Organisationen usw.)

Sollten die zugehörigen Beschreibungen der DNB für die Beurteilung, ob es sich um dasselbe, ein ähnliches oder ein völlig anderes Objekt handelt, nicht ausreichend sein, werden noch ausgehende Links verfolgt (Teilprozess 2.3). Falls der Link-Typ anhand der zusätzlichen Informationen aus externen Quellen immer noch nicht festgelegt werden kann, wird die Ressource nicht verknüpft. Anderenfalls wird der Link manuell erzeugt. Die letzte „Schleife“ wird so lange durchlaufen, bis alle Ressourcen aus dem Häuserbuch (und die jeweiligen Pendants in den Informationsbeständen der DNB) einmal untersucht und bestenfalls verknüpft wurden.

Hilfsmittel

Es werden einige Applikationen eingesetzt, z.B. werden die o.g. Ontologien mithilfe von Protégé Desktop untersucht, um z.B. einen schnellen Überblick über die Klassenhierarchien gewinnen zu können. Für den Zugriff auf die Graphen bzw. für deren Aktualisierung ist auch ein RDF-Store erforderlich. Wie erwähnt wird VOS dafür eingesetzt. Außerdem wird Silk bzw. die Silk-Workbench verwendet (s. GP [3.2.3](#)).

7.2.5 Gesamter Workflow

Bezugnehmend auf das zweite operative Teilziel dieser Arbeit wird der konzipierte Workflow noch einmal zusammengefasst und in den gesamten Ablauf eingeordnet (vgl. Abb. 7.6). Neben den beschriebenen Teilschritten, den notwendigen und optionalen Prozessen, den Daten, Dokumenten und der Reihenfolge ist auch der geschätzte Aufwand angedeutet (s. Abb. unten – Größe der farbigen Polygone):

- (1) Wie erwähnt, ist die semi-automatische Verlinkung auf der Instanzebene der aufwendigste Teilschritt in diesem Workflow, da (fast) jede Ressource manuell untersucht werden muss.
- (2) Die semi-automatische Umformung ist wegen der händischen Datenbereinigung im Zuge der semantischen Transformation etwas aufwendiger; einige Attribute können aber unverändert übernommen werden.
- (3) Für die manuelle Modellierung der simplen TBox werden nur einfache Vokabulare/Sprachmittel verwendet, weshalb dieser Teilschritt vergleichsweise schnell und einfach umgesetzt werden kann.
- (4) Die Tätigkeiten zwecks Konsolidierung der Informationen, sind nicht aufwendig, da die Organisation der Tripel über benannte Graphen nicht prioritär behandelt werden muss. Explizite Aussagen zum Kontext wären zwingend erforderlich gewesen, um aus dem Resultat von SPARQL-Abfragen Applikationsschemata oder Capabilities-Dokumente generieren zu können. Da die Informationen jedoch nicht über einen WFS verfügbar gemacht werden, kann darauf vorerst verzichtet werden.

Der Übersichtlichkeit halber wurde die Darstellung vereinfacht. Hierzu wurde der Informationsbezug (aus der LOD-Cloud), die „Validierung“ etc. ausgespart. Als Grundlage wurde der Workflow von Nishanbaev et al. (2020) herangezogen und an die individuellen Anforderungen angepasst (vgl. Abb. 3.3).

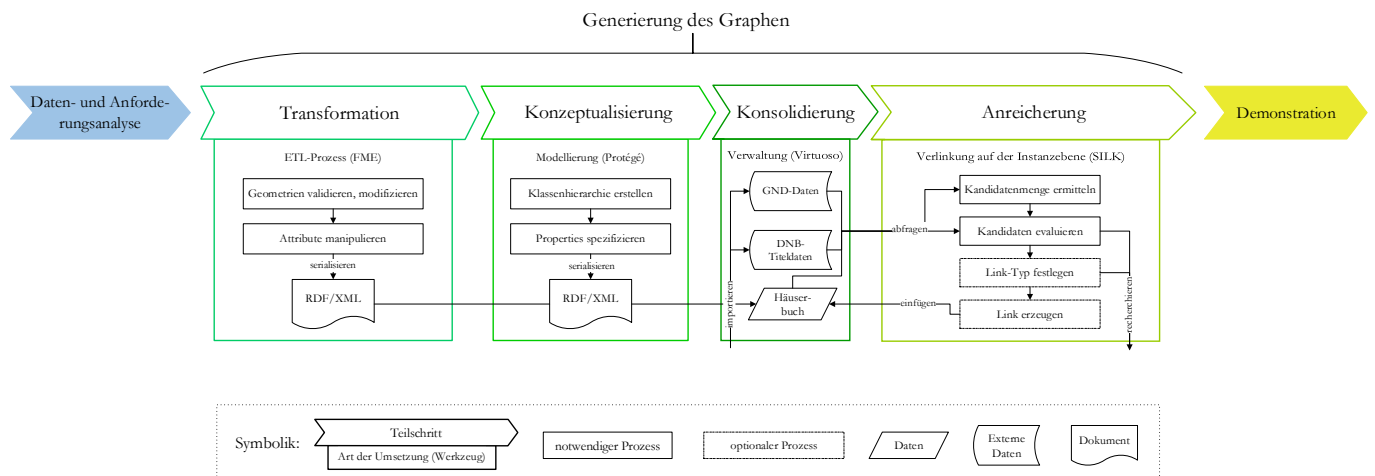


Abb. 7.8: Konzeption des Workflows zwecks Generierung des RDF-Graphen, ED

8 Praktische Umsetzung

8.1 Aufbau des Systems

Die Ausführungen zum System auf das notwendige Minimum reduziert (vgl. 7.1.2). Der empirische Teil der Arbeit konzentriert sich auf die Generierung des Graphen (s. Abschnitt 8.2). Das System, welches für die Demonstration und perspektivisch – siehe Ausblick (GP 10.2) – für die Publikation erforderlich ist, wird nicht produktiv gestellt (Nicht-Ziele, GP 1.2). Es manifestiert sich als einfaches IS. Der Aufbau des Systems beschränkt sich auf die Kompilierung, Installation und Konfiguration von existierenden Softwarepaketen innerhalb der VMs; der Einsatz eines WFS-Clients und -Adapters unterbleibt (vgl. GP 7.1.1). Wie erwähnt, zielt dieser Abschnitt auf die Gewährleistung der Nachvollziehbarkeit resp. Reproduzierbarkeit ab. Hierzu werden die Versionen der eingesetzten Applikationen und die technischen Spezifikationen angegeben. Letzteres gilt nicht für VM1 (dient der Transformation und der Konzeptualisierung), sondern nur für VM2, die neben der Demonstration für die Konsolidierung und die Anreicherung (vgl. Abb. 7.8) benötigt wird. Für die Anreicherung ist es wichtig, wann eine Routine terminiert (s GP 8.2.4). Von daher sind die folgenden Angaben relevant:

CPU	Intel(R) Xeon(R), E5-2623 v4 @ 2.6 GHz, Quad Core, AVX2
RAM	32 GB RDIMM, 2400MT/s, x8 Data Width
Disk	SSD, 500 GB, kein RAID
NIC	Broadcom 5720 QP, 1GB/s

Tab. 8.1: Relevante Spezifikationen (VM2)

Als OS für VM2 wurde SLES 12, SP 4 eingesetzt; als FS wurde ext4 gewählt. Die Datenmenge, die in VM1 erzeugt bzw. verwaltet wird, ist wesentlich geringer als jene in VM2 (vgl. GP 8.2.3 und Tab. 8.11). Die Performance von VM1 ist irrelevant. Die Tabelle listet die erforderlichen Angaben zu den eingesetzten Applikationen:

Applikation	Version	Zweck	VM	Weitere Informationen
VirtualBox	6.1.14	Virtualisierung	1	https://www.virtualbox.org/
Windows 10 Pro	21H1	OS	1	https://docs.microsoft.com/de-de/windows/
FME Desktop	2019.1.3.1	ETL	1	https://www.safe.com/fme/fme-desktop/
Visual Studio Code	1.51.1	IDE	1	https://code.visualstudio.com/
Protégé Desktop	5.5.0	Modellierung	1	https://protege.stanford.edu/
VMware vSphere	VS7	Virtualisierung	2	https://www.vmware.com/
Silk	3.3.3	Verlinkung	2	https://github.com/silk-framework/
VOS	7.2.5.1	Verwaltung...	2	https://virtuoso.openlinksw.com/

Tab. 8.2: Relevante Applikationen auf VM1 und VM2

Die Applikationen wurden nach Anleitung (vgl. o.g. URIs) installiert/konfiguriert. Dies gilt auch für VOS¹¹⁷.

¹¹⁷ Zur Kompilierung und Installation: <https://github.com/openlink/virtuoso-opensource>

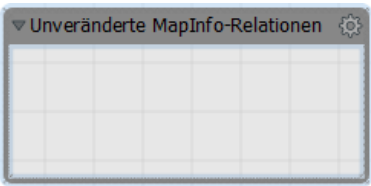
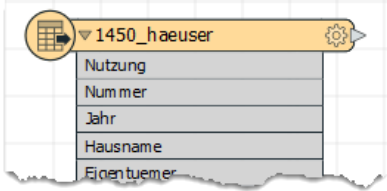
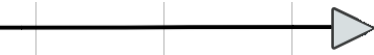
8.2 Generierung des Graphen

8.2.1 Transformation

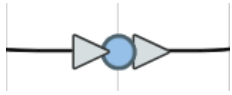
Nachfolgend wird dargelegt, wie der erste Teilschritt im Workflow zwecks Generierung des RDF-Graphen umgesetzt wurde (s. Abb. 7.8). Es wird aufgezeigt, welche Entscheidungen getroffen wurden und welche Probleme sich ergaben. Die Zielerreichung wird unter GP 9.1 diskutiert und anhand der Anforderungen beurteilt.

FME-Workbench

Prinzipiell kann der Transformationsprozess strukturiert und anschaulich mithilfe der FME-Workbench erläutert werden. Diese enthält eine graphische Übersicht in Form eines Prozessablaufdiagramms inklusive aller verwendeten Elemente und deren Beziehungen. Die Parametrisierung kann ebenfalls visualisiert werden. Es bedarf jedoch einer adäquaten Designstrategie, um die Übersichtlichkeit zu wahren. FME bietet hierfür viele geeignete Funktionen. Allerdings erschwert die erforderliche Aufteilung der Workbench auf mehrere Seiten eine verständliche Erläuterung¹¹⁸. Um dies etwas zu entschärfen, wird zunächst eine legendenähnliche Übersicht zu den verwendeten Signaturen, deren Bezeichnung und Bedeutung geliefert. Außerdem enthält die folgende Tabelle die wichtigsten Erklärungen zur Strukturierung. Selbsterklärende Komponenten werden nicht erörtert.

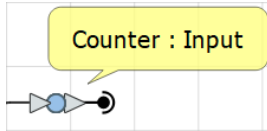
Signatur	Elementbezeichnung	Bedeutung
	Bookmark	Fungiert in der folgenden Workbench als Container, um Elemente und Prozessschritte logisch zu gruppieren. Hierfür wird stets eine aussagekräftige Überschrift und eine individuelle Farbkombination verwendet. Diese soll in der gesamten Arbeit möglichst einheitlich sein, weshalb z.B. die unveränderten MapInfo-Relationen von einer grauen Konturlinie umschlossen werden (analog zur Farbwahl unter GP 5.1.1 bzw. in Abb. 5.1).
	Reader	Liest die MapInfo-Relation ein. Für jede Relation existiert ein separater Reader. Die standardmäßige Darstellung von FME wurde nicht verändert. Im Beispiel wird angedeutet, dass immer die jeweiligen Attribute dargestellt werden. Diese werden häufig manipuliert, weshalb die Darstellung sinnvoller ist, als der ständige Verweis auf die Angaben in den vorherigen Abschnitten. Der Transformationsprozess benötigt keinen korrespondierenden Writer.
	Verbinder	Verknüpft mindestens zwei Elemente bzw. zeigt die Datenflussrichtung an (Pfeilspitze). Häufig werden Verzweigungen verwendet.

¹¹⁸ Es konnte keine Funktion gefunden werden, mit welcher die Workbench als Vektorgrafik exportiert werden kann.



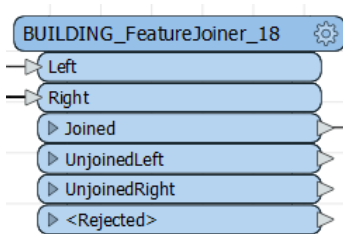
Junction

Dient der Gestaltung, da mehrere Verbinder verknüpft werden können, um die Übersichtlichkeit zu wahren.



Tunnel

Dies ist ebenfalls ein Designelement, das die Verbindung zwischen zwei Transformatoren teilweise verbirgt. Um die Zuordnung zum folgenden Transformator zu gewährleisten, wurde in der Textblase dessen Name (hier Counter) und der Port (hier Input des Counters) angegeben. Der Counter-Transformator trägt das korrespondierende Element, das stets in der entsprechenden Farbe dargestellt wurde. Textblasen zwischen verschiedenen Transformatoren wurden unterschiedlich eingefärbt.




Transformer

Über Transformer werden die Eigenschaften der Features manipuliert. Ein Transformator kann über ein oder mehrere Eingabeports und keinen oder mehrere Ausgabeports verfügen, die über Verbinder mit anderen Ports verknüpft werden können. Grundsätzlich kann zwischen Transformatoren differenziert werden, die standardmäßig installiert werden und sogenannten Custom-Transformatoren, die entweder selbst entwickelt oder von der FME-Community über einen Hub bereitgestellt werden. Transformer wurden i.d.R. nach folgendem Schema benannt: [Objektnamen]_[Standardname]_[Reihenfolge]. Der Objektnamen gibt an, auf welche Entitäten der Transformator primär wirkt. Danach folgt der Standardname des Transformators (wie von FME festgelegt). Diesbezüglich erfolgt keine Änderung, sodass bei Bedarf die detaillierte FME-Dokumentation zu jedem Transformator eingesehen werden kann. Die Ziffer bezieht sich nicht auf den Prozessablauf, sondern auf die Reihenfolge der Erläuterung. Transformer mit identischer Ziffer werden gemeinsam erörtert.

Inspektor

Separate Applikation, welche die Analyse der prozessierten Daten für jeden Prozessschritt durch Visualisierung des aktuellen Status über den internen FME Feature Store ermöglicht. Hierdurch wird eine effiziente Überprüfung während der Prozessierung ermöglicht; allerdings erfüllt der Inspektor nicht alle Anforderungen an ein vollwertiges GIS.

	Logger	Dient prinzipiell demselben Zweck wie der Inspektor, allerdings ohne graphische Komponente. Informationen werden zur Laufzeit über ein Umsetzungsprotokoll ausgegeben.
	Haltepunkt	Unterbricht den Prozessablauf, wofür optional eine Bedingung definiert werden kann.

Tab. 8.3: Übersicht zu den verwendeten FME-Elementen

Zwecks Übersichtlichkeit wurde die Parametrisierung aller Transformer nicht über die Workbench visualisiert. Da die gewählten Parameter für die Nachvollziehbarkeit bzw. Reproduzierbarkeit (häufig) wichtig sind, werden sie nachfolgend sehr oft mithilfe eines Screenshots dargestellt. Code-Blöcke werden für die Erläuterung zwar auch mittels Screenshots dargestellt; allerdings wurden sie zusätzlich in textueller Form beigefügt (im Anhang).

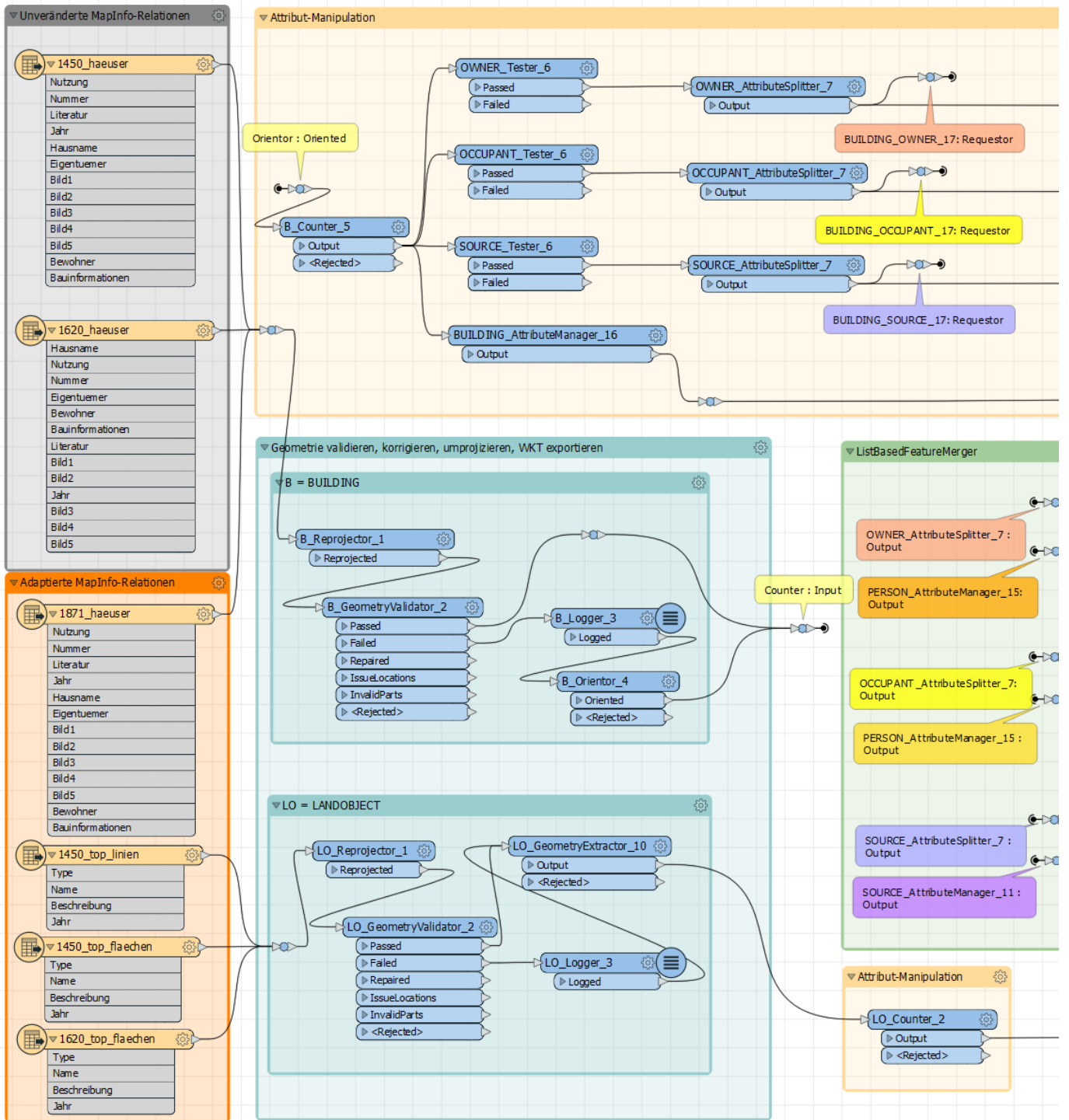


Abb. 8.1: FME-Workbench (Teil 1), EK

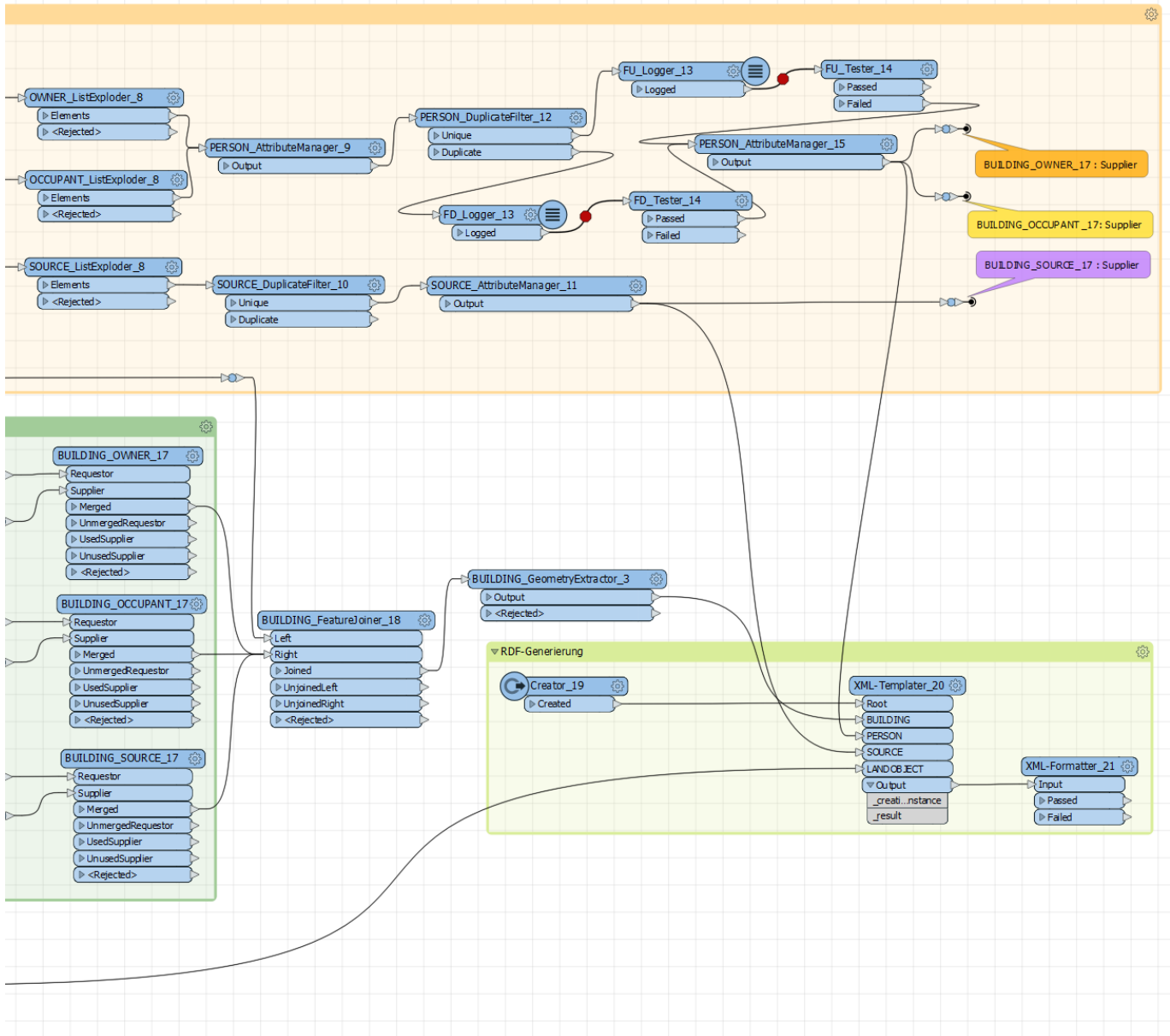


Abb. 8.2: FME-Workbench (Teil 2), EK

Nachdem die MapInfo-Relationen über die entsprechenden Reader eingelesen wurden, erfolgte zunächst die Koordinatentransformation für alle Geoobjekte mittels der beiden Reprojector-Transformer (1), die identisch parametrisiert wurden:

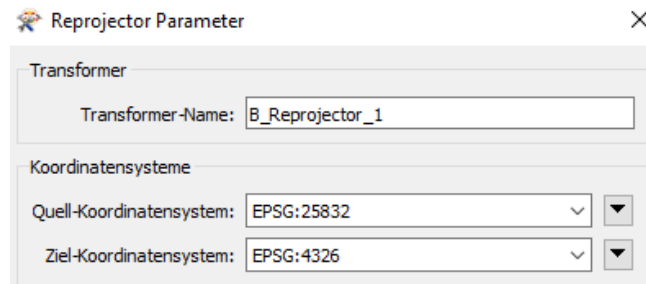


Abb. 8.3: Parameter der beiden Reprojector-Transformer, EK

Das Ziel-CRS, dessen vollständige Parameter unter [ESPG:4326](https://epsg.org/entry/4326.html) eingesehen werden können, wurde vom OGC und W3C für die Publikation räumlicher Informationen als LOD explizit empfohlen (Tandy et al., 2017). Die Geometrie wurde anschließend mittels eines GeometryValidator-Transformers (2) überprüft und danach ggf. modifiziert. Die folgenden Parameter wurden für beide verwendeten GeometryValidator-Transformer gewählt:

B_GeometryValidator_2 [GeometryValidator]	
Validierung: Zu ermittelnde Probleme festlegen: Benutzerdefiniert	
Zu ermittelnde Probleme	Parameter
Flächenausrichtung	Modus: Nur Donuts prüfen
Fehlende Stützpunkt-Normalen	
Enthält -0, NaN oder Unendlichkeit	Bemaßung verarbeiten: Bemaßung nicht verarbeiten
Enthält Null-Geometrieteile	
OGC Simple-konform	
OGC Valid-konform	OGC-Version: OGC-Version 1.2.0
Degenerierte oder korrupte Geometrien	Parameter: Z-Verbindungsmodus: Ignorieren Modus für Genauigkeit von Segmentendpunkten: Auto
Selbst-Überschneidungen in 2D	Parameter: Selbst-berührende Polygone prüfen: Ja Z-Verbindungsmodus: Ignorieren Toleranz: Automatic
Ausgabe: Listenname für ermittelte Probleme: _issues Reparaturversuch: Nein Modus für Zusammenfassung: Detailliert	

Abb. 8.4: Parameter der beiden GeometryValidator-Transformer, EK

Geometrien, die diese Testkriterien nicht erfüllten, wurden immer über den Failed-Port ausgegeben, da kein automatischer Reparaturversuch unternommen werden sollte. Das Ergebnis wäre anderenfalls von der Auswahl der Testkriterien und deren Reihenfolge abhängig gewesen¹¹⁹. In der vorliegenden Arbeit wurden die Fehler manuell analysiert und durch den Einsatz zusätzlicher Transformer beseitigt. Hierfür wurde zunächst ein Logger (3) zwecks Protokollierung verwendet. Zusätzlich wurde ein Haltepunkt mit folgender Bedingung eingefügt:

¹¹⁹ Eine detaillierte Dokumentation kann unter http://docs.safe.com/fme/html/FME_Desktop_Documentation/FME_Transformers/Transformers/geometryvalidator.htm eingesehen werden.

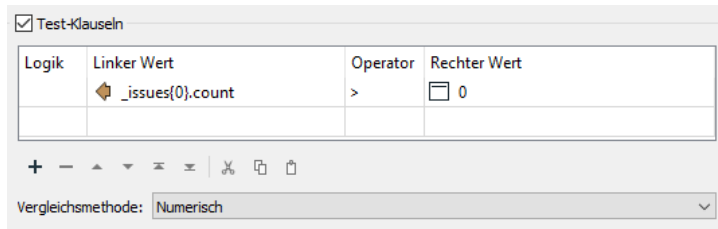


Abb. 8.5: Protokollierung geometrischer Probleme, EK

Die Bedingung basierte auf der resultierenden Liste des Failed-Ports. Es handelte sich um eine komplexe, strukturierte Liste, die mehrere Elemente mit verschiedenen Attributen pro Geobjekt enthalten konnte. Der Prozess wurde unterbrochen, sobald das Zählerattribut des ersten Listenelements (Indexposition 0) einen Wert größer 0 erhielt. Im Rahmen der anschließenden Fehleranalyse mithilfe der Feature-Kontrolle konnten die Probleme identifiziert werden. Die Stützpunkte aller Polygone wurden im Uhrzeigersinn erfasst und gespeichert. Das traf auch für alle Polygone mit mehreren Ringen zu; d.h., die Stützpunkte des inneren Rings / der inneren Ringe wurden ebenfalls immer im Uhrzeigersinn erfasst und gespeichert. Da die geometrischen Eigenschaften der Häuser-Relationen im Vorfeld nicht adaptiert wurden, machte sich der Fehler erst an dieser Stelle bemerkbar. Das repräsentative Beispiel zeigt das Fehlerprotokoll bzw. die Fehlerliste für den Mainzer Dom:

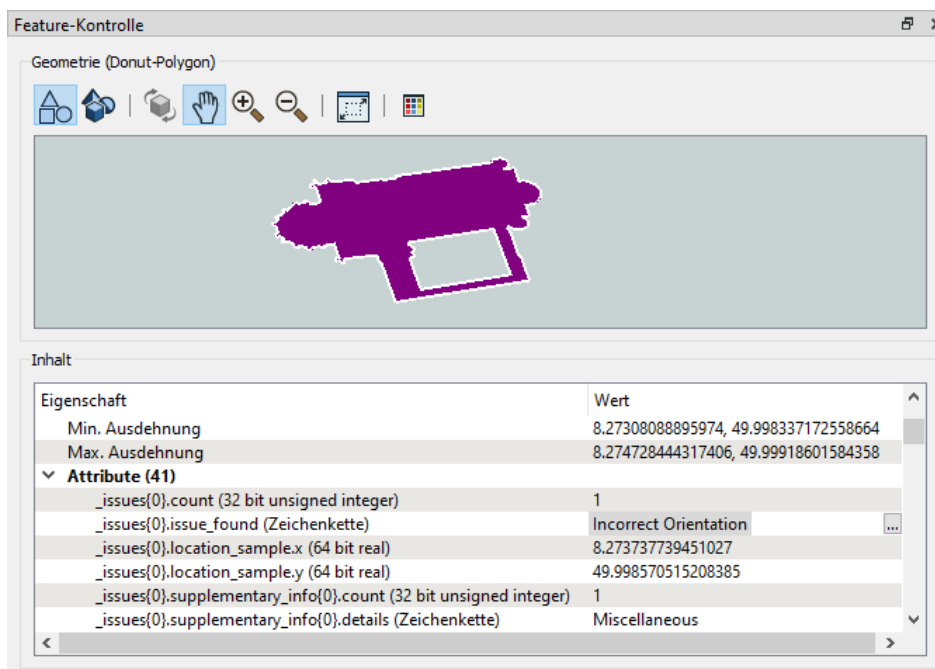


Abb. 8.6: Fehlerhafte Orientierung der Koordinaten, Feature Kontrolle

Die Fehler wurden ausgegeben, weil die Orientierung der Koordinaten des äußeren Rings und der Inselepolygone stets identisch war. Die Stützpunkte des Inselepolygons müssen jedoch in entgegengesetzter Richtung zur äußeren Grenze orientiert werden. Die Stützpunkte des äußeren Rings wurden durch den nachträglich eingefügten Orientor-Transformer (4) entgegen des Uhrzeigersinns und die Stützpunkte der Inselepolygone im Uhrzeigersinn angeordnet. Dadurch wurden auch die Stützpunkte aller Polygone ohne Inseln gegen den Uhrzeigersinn gespeichert. Der einzige Parameter des vorgenannten Transformers legte den Orientierungstyp mit der „Linken-Hand-Regel“ fest. Dies führte hier zum gewünschten Ergebnis (siehe Koordinaten im Anhang, die wegen der durchgeführten Koordinatentransformation in der Reihenfolge latitude, longitude gelistet sind).

Die übrigen Geobjekte bzw. MapInfo-Relationen, hier unspezifisch als „LANDOBJECT“ bezeichnet, wurden im Vorfeld angepasst (vgl. GP 5.1.1). Sicherheitshalber wurde die Geometrie trotzdem validiert. Die Bedingung des Haltepunkts wurde analog spezifiziert. In diesem Fall wurden jedoch keine Fehler ausgegeben. Die Extraktion der Geometrie (Transformer: LO_GeometryExtractor_10) wird mithilfe von Abb. 8.15 dargestellt.

Vorher wird die Manipulation der Attribute beschrieben. Dieser Teilprozess begann mit der Erzeugung einer numerischen ID für jedes Geobjekt mittels der beiden Counter-Transformer (5), die jeweils nur für ihren lokalen Anwendungsbereich zählten (Surrogatschlüssel mit Auto-Inkrement-Funktion, Zählbeginn 1). Der Name der Zählvariablen für die Gebäude lautete „Nummer“ und für die übrigen Objekte „loID“. Damit wurden geeignete Schlüsselkandidaten erzeugt. Für die Objekte, die keine Häuser repräsentieren, mussten keine weiteren Attribute geändert/hinzugefügt werden. Sie wurden anschließend an den XML-Templater weitergeleitet. Die unter GP 5.1.1 beschriebenen Eigenschaften der Häuser-Relationen bedingten jedoch den Einsatz zusätzlicher Transformer. Im entsprechenden Prozesszweig wurden drei Tester-Transformer (6) eingesetzt, die alle Features ohne Eigentümer-, Bewohner- oder Literaturangabe von der weiteren Verarbeitung ausschlossen.

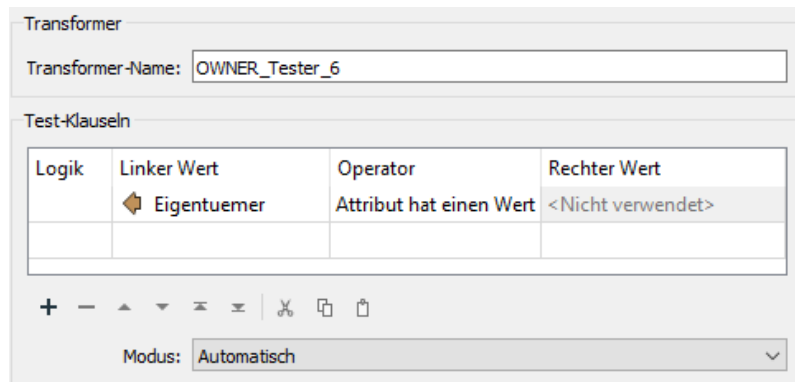


Abb. 8.7: Parameter der drei Tester-Transformer, EK

Diese zusammengesetzten Attribute dienten den drei AttributeSplitter-Transformern (7) anschließend zur Listen-Generierung. Aufgrund der erwähnten Erfassungssystematik wurden folgende Parameter gewählt, die für die weiteren AttributeSplitter-Transformer entsprechend definiert wurden:

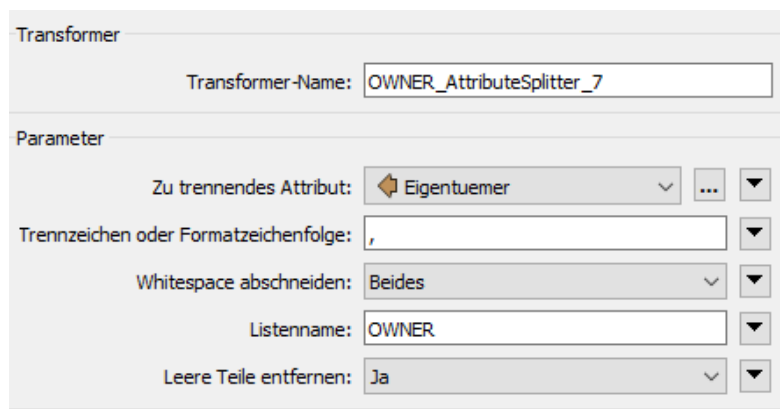


Abb. 8.8: Parameter der drei AttributeSplitter-Transformer, EK

Augenscheinlich bietet der Custom-Transformer MultiAttributeSplitter eine Möglichkeit den Prozessschritt zu konsolidieren, indem die drei Datenströme in diesem Transformer zusammengeführt werden. Es können mehrere Attribute zur Listen-Generierung angegeben werden.

Das folgende Beispiel zeigt die unterschiedlichen Listenstrukturen (MultiAttributSplitter, links; Attribute-Splitter, rechts):

Eine komplexe Liste: MULTI		Drei einfache Listen: L1, L2, L3	
Hausname	Zum kleinen Schlüssel	In jeder Liste identisch	
Nutzung	Kir_bes		
Jahr	1594		
...		...	
MULTI{0}.Bewohner	Dr. Jacobus Campius	L1: OCCUPANT{0}	Dr. Jacobus Campius
MULTI{0}.Eigentuemmer	Pfarrei St. Peter- St. Emmeran Mainz	L2: OWNER{0}	Pfarrei St. Peter- St. Emmeran Mainz
MULTI{0}.Literatur	Schrohe: Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhun- derts / Teil 1 (1930)	L3: SOURCE{0}	Schrohe: Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts / Teil 1 (1930)

Tab. 8.4: Vergleich der Listenstrukturen

Die komplexe Liste des MultiAttributeSplitters wird zwar von den folgenden ListExploder-Transformern (8) als Eingabe akzeptiert, jedoch bieten diese keine adäquaten Funktionen, um die Listenelemente anschließend manipulieren zu können. Hierfür müssen alle Listenelemente durch die Erzeugung neuer Objekte offengelegt werden, d.h., es müssen mehrere Tabellenspalten angelegt werden, welche die Informationen zu den Eigentümern, Bewohnern und die Literaturangaben aufnehmen. Dies war nur mit den drei einfachen Listen zu erreichen, deren Elemente von den ListExploder-Transformern separat offengelegt wurden. Die Parameter sind nachfolgend exemplarisch für die Bewohner-Liste (rechts) abgebildet, da die übrigen ListExploder-Transformer analog konfiguriert wurden. Auf der linken Seite ist die einzig mögliche Verwendung der komplexen Liste im ListExploder-Transformer dargestellt.

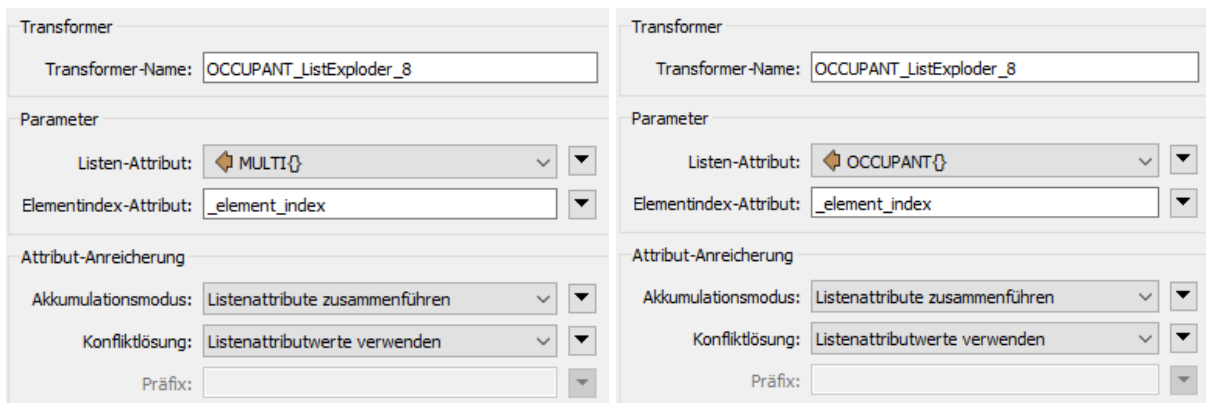


Abb. 8.9: Verarbeitung unterschiedlicher Listenstrukturen, EK

Der ListExploder-Transformer kann die komplexe Liste lediglich für die Erzeugung einer neuen Tabellenspalte nutzen, welche die Informationen zu den Indexpositionen der Listenelemente aufnimmt. Die Information, ob diese auf einen Eigentümer, Bewohner oder eine Literaturangabe verweisen, geht zwangsläufig verloren. Der ListExploder-Transformer bietet keine Möglichkeit, um die einzelnen Attributwerte der komplexen Liste offenzulegen. Die drei (separaten) einfachen Listen enthalten die notwendigen Informationen. Wenn ein Haus mit der ID 1 z.B. die Bewohner X, Y, Z hat, enthält die resultierende Tabelle nun drei Zeilen für das Haus mit der ID 1. Dies gilt entsprechend für die Tabellen mit den Einträgen für Eigentümer- / Literaturinformationen.

Der Einsatz der ListExploder-Transformer zwecks Erweiterung der Tabellen um neue Spalten und Zeilen, diente dem Ziel, die Personen- / Literaturinformationen zu trennen, um diese anschließend separat strukturieren zu können. Die Informationen mussten – ähnlich wie in einem RDBMS – verwaltet, manipuliert und verknüpft werden. Ein AttributeManager-Transformer führte deshalb im nächsten Schritt (9) die Eigentümer und Bewohner in einer Personentabelle zusammen, wobei alle überflüssigen Attribute entfernt wurden. Das Attribut „Nutzung“ wurde zunächst beibehalten, obwohl es in dieser Tabelle nicht geführt werden sollte. Es diente der Beantwortung der Frage, ob die Klassifizierung wenigstens für jene Gebäude verbessert werden kann, die über eine Information zum Eigentümer und/oder zum Bewohner verfügen. Letztlich wurde es entfernt (s.u.).

Eingabe-Attribut	Ausgabe-Attribut	Attributwert	Aktion
Hausname			Entfernen
OCCUPANT	pName		Umbenennen
OWNER	pName		Umbenennen
Jahr	Jahr		Keine Aktion
Nutzung	Nutzung		Entfernen
Nummer	Nummer		Entfernen
Bauinformationen	Bauinformationen		Entfernen
Bild1	Bild1		Entfernen
Bild2	Bild2		Entfernen

Abb. 8.10: Vorbereitung der Personentabelle, EK

Ein DuplicateFilter-Transformer (10) entfernte alle doppelten Zeilen (Schlüsselattribut-Parameter: SOURCE) in der Literaturtabelle. Während der Überprüfung im Inspektor konnten in der Ergebnistabelle keine weiteren Duplikate identifiziert werden, was auf die (im Abschnitt 5.1 beschriebene) Attributstruktur zurückzuführen ist. Bevor die endgültige Tabelle an das XML-Template weitergeleitet werden konnte, erhielt jede Quelle noch einen Schlüssel durch einen AttributeManager-Transformer (11), der auch alle überflüssigen Attribute verwarf.

Eingabe-Attribut	Ausgabe-Attribut	Attributwert	Aktion
Hausname			Entfernen
SOURCE	sName		Umbenennen
	sID	@Count(5,1)	Wert festlegen
Jahr	Jahr		Entfernen
Nutzung	Nutzung		Entfernen
Nummer	Nummer		Entfernen
Bauinformationen	Bauinformationen		Entfernen

Abb. 8.11: Literaturtabelle, EK

Bezüglich der Personentabelle erzeugte der verwendete DuplicateFilter-Transformer (12) in qualitativer Hinsicht kein vergleichbares Resultat. Die Ergebnistabelle konnte ohne manuelle Eingriffe nicht weiterverarbeitet werden. Deshalb wurden alle Informationen mithilfe der beiden Logger (13) protokolliert. Der Prozess wurde anschließend per Haltepunkt gestoppt. Sowohl die vermeintlichen Duplikate (FD = False Duplicate) als auch die scheinbar eindeutigen Attributwerte (FU = False Unique) wurden händisch überprüft. Nachfolgend werden einige Kriterien und Aspekte der Überprüfung exemplarisch anhand der Parametrisierung der folgenden Tester-Transformer (14) erläutert (Abb. 8.12). Ein Tester-Transformer (links) verarbeitete *vermeintliche* Duplikate und inkludierte die Personen wieder in den Prozess, da es faktisch keine Duplikate sind. Ein weiterer Tester-Transformer (rechts) filterte Personen, die *tatsächlich* immer noch doppelt in der Tabelle enthalten waren.

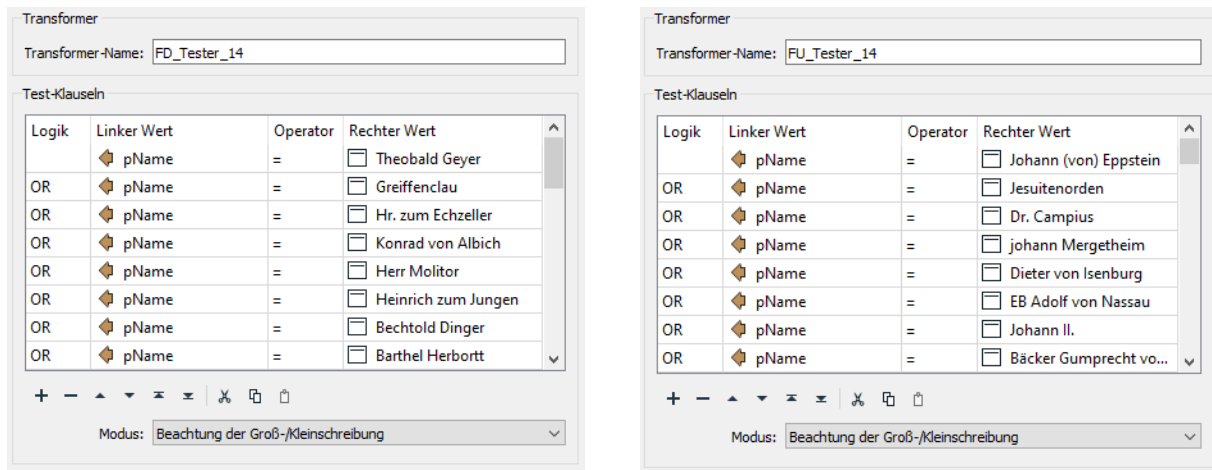


Abb. 8.12: Beispiele für FD/FU, EK

Die Entscheidungen wurden im Hinblick auf die angestrebte semi-automatische Verlinkung mit anderen Ressourcen getroffen. Vor diesem Hintergrund war es zielführend, eine Person z.B. als „Erzbischof Dieter von Isenburg“ zu bezeichnen, statt mit „EB von Isenburg“. Andernfalls besteht die Gefahr, dass im Rahmen der Anreicherung keine passenden Ressourcen für die Verlinkung ermittelt werden können. Mit dem Bekanntheitsgrad einer Person steigt die Wahrscheinlichkeit, dass die Entscheidung richtig ist, da zahlreiche Quellen im Web adäquate Informationen enthalten. Manchmal ist die Entscheidung auch äußerst schwierig oder gar nicht zu treffen, z.B. bei ähnlichen Namensschreibweisen unbekannter Personen (Fyscher, Fischer, Fysches son). In solchen Fällen wurde angenommen, dass es sich um unterschiedliche Personen handelte. Der Attribute-Manager-Transformer (15) verwendete die Ausgabe des FalseUnique-Testers am Failed-Port und die Ausgabe des False-Duplicate-Testers am Passed-Port, um jede Person in der finalen Ergebnistabelle mit einer eindeutigen ID im neu angelegten Feld „pID“ zu versehen und schließlich an den XML-Templater weiterzuleiten.

Der manuelle Eingriff bzw. die Überprüfung aller Personen zielte auf die Verbesserung der DQ ab. Hierzu wurden noch die Werte in den Feldern Eigentümer und Bewohner mit den Werten in den Feldern Nutzung und Jahr abgeglichen. Es bestand eine theoretische Chance: Geprüft wurde, ob genügend widerspruchsfreie Daten zusammengestellt werden können, um die jeweiligen, historischen (sozialen) Gruppen bedeutungsvoll bzw. substantiell instanzieren zu können. Praktisch konnte dies mit den vorhandenen Daten nicht realisiert werden. Es konnten keine substantiellen Aussagen über *bestimmte* soziale Gruppen getroffen werden – z.B., dass es eine Gruppe von Christen gab, die als Dominikaner bezeichnet werden oder dass *bestimmte* Personen Mitglieder dieser Gruppe waren. Dafür wären eingehende Untersuchungen mittels Primärquellen/Sekundärquellen erforderlich gewesen. Wegen der obligatorischen Anforderungen, die definiert wurden (vgl. AS1), war das Weglassen oder die widerspruchsbefahtene Repräsentation keine Option. Deshalb wurden die LOD-Prinzipien verletzt, indem leere Knoten erzeugt wurden. Präziser ausgedrückt: Über das XML-Template wurden anonyme Knoten generiert, da diese keine ID erhielten (vgl. Lis. 8.2: `rdf:parseType="Resource"`).

Auf alle Objekte, die Gebäude repräsentieren, musste ebenfalls ein AttributeManager-Transformer (16) angewendet werden. Überflüssige Attribute wurden entfernt. Die Spalten Eigentümer, Bewohner und Literatur waren nicht mehr erforderlich. Stattdessen sollte später eine Referenz auf die neu angelegten Tabellen mit den entsprechenden Einträgen erfolgen. Die IDs in der Personen- und Literaturtabelle, sollten quasi als „Fremdschlüssel“ dienen. Die vorbereitenden Operationen im AttributeManager-Transformer sind links dargestellt; der rechte Screenshot zeigt die Klassifizierung.

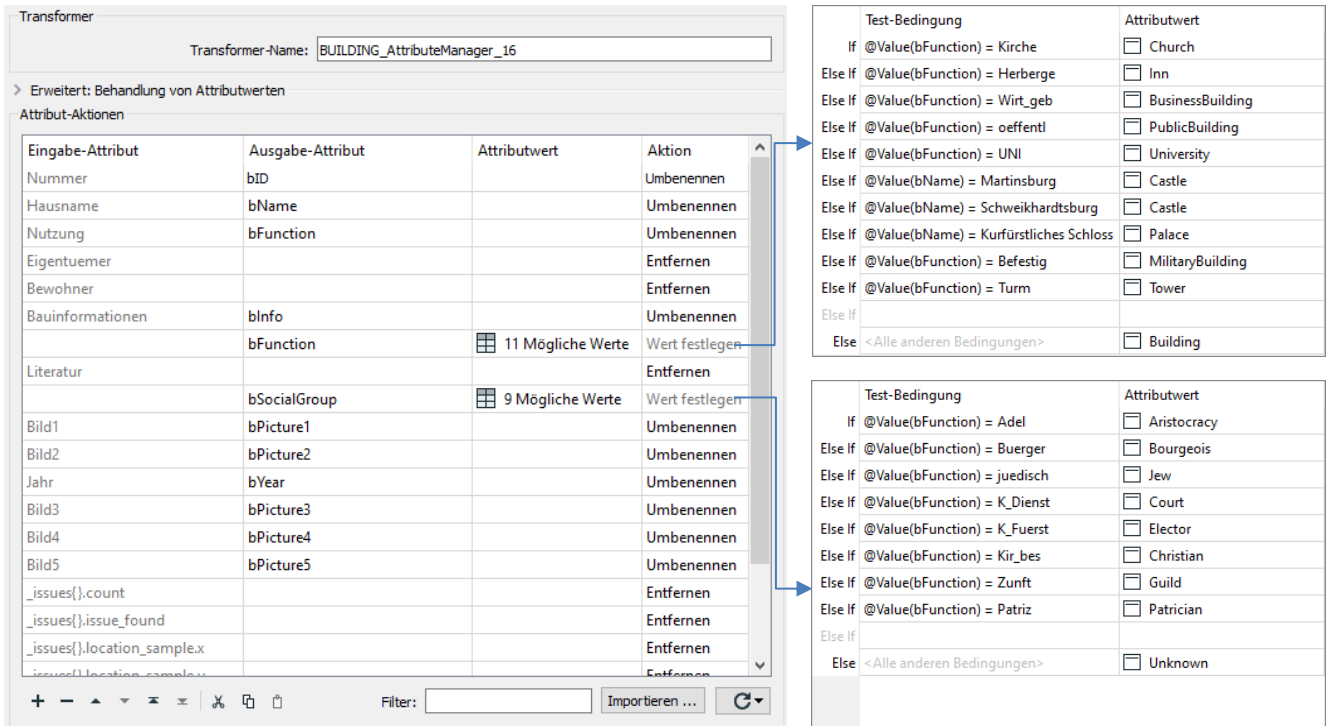


Abb. 8.13. Finale Attribut-Manipulation, EK

Die drei ListBasedFeatureMerger-Transformer (17), wurden als BUILDING_OWNER_17, BUILDING_OCCUPANT_17, BUILDING_SOURCE_17 bezeichnet und ähnlich konfiguriert, weshalb exemplarisch nur die Parameter des erstgenannten Transformers dargestellt sind:

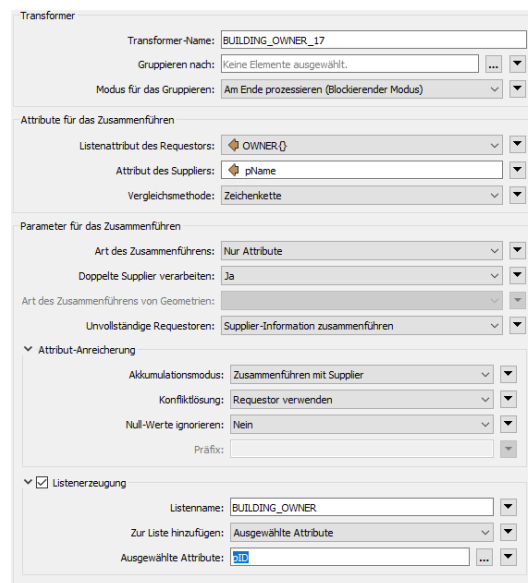


Abb. 8.14: Konsolidierung mittels komplexer Listen, EK

Die ListBasedFeatureMerger-Transformer erhielten als Requestor die einfachen Listen der Splitter-Transformer, deren Struktur bereits erläutert wurde. Als Supplier wurde die Personentabelle / die Literaturtabelle angegeben, wobei die IDs in den jeweiligen Tabellen (oben exemplarisch markiert) dann als „Fremdschlüssel“ fungierten. Das Verknüpfungsattribut war die Zeichenkette in den Feldern pName/sName der vorgenannten (geprüften bzw. bereinigten) Tabellen und die Zeichenkette in den Spalten Eigentümer, Bewohner, Literatur in der Attributtabelle der Geobjekte (Häuser). Wie erwähnt, legten die Splitter-Transformer neue Zeilen und Spalten hierfür an. Letztlich entstanden drei separate, komplexe Listen an jedem Merged-Port der ListBasedFeatureMerger-Transformer. Das folgende Beispiel zeigt Auszüge aus diesen Listen und deutet die Verknüpfung an:

BUILDING_OWNER	
Nummer	743
BUILDING_OWNER{0}.pID	37
Nummer	801
BUILDING_OWNER{0}.pID	67
...	
BUILDING_OCCUPANT	
Nummer	743
BUILDING_OCCUPANT{0}.pID	911
Nummer	801
BUILDING_OCCUPANT{0}.pID	911
...	
BUILDING_SOURCE	
Nummer	743
BUILDING_SOURCE{0}.sID	9
BUILDING_SOURCE{1}.sID	19
Nummer	801
BUILDING_SOURCE{0}.sID	9
...	

PERSON	
pID	pName
...	
37	Pfarrei St. Peter-St. Emmeran Mainz
...	
67	Universität Mainz
...	
911	Dr. Jacobus Campius

SOURCE	
sID	sName
...	
9	Schrohe: Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts / Teil 1 (1930)
...	
19	Falck: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)

Tab. 8.5: Verknüpfung von Gebäuden mit Eigentümer, Bewohner und Literatur

Die Attributwerte im Feld „Nummer“ wurden durch den Counter-Transformer (5) generiert; es waren nicht die ursprünglichen Werte des Felds „Nummer“ der Ausgangsrelationen, die keine Schlüsseigenschaften aufweisen (vgl. GP 5.1.1). Die o.g. IDs sollten nicht in separaten Listen geführt werden, sondern an die jeweiligen Geoobjekte angehängt werden. Hierfür konnten die Schlüsselattribute „bID“ und „Nummer“ vom Feature-JoinerTransformer (18) für einen Left-Join verwendet werden. Im Ergebnis waren alle Gebäude-Objekte enthalten. Diese hatten die folgenden Eigenschaften:

Attribut	Typ/Charset/Encoding	Optionalität ¹²⁰
bID	Integer, 32 Bit	[1..1]
bFunction	Varchar, Unicode/UTF-8	[1..1]
bSocialGroup	Varchar, Unicode/UTF-8	[1..1]
bName	Varchar, Windows-1252	[1..1]
bInfo	Varchar, ISO/IEC 8859-1	[0..1]
bYear	Integer, 16 Bit	[1..1]
BUILDING_OWNER{0-n}.pID	Integer, 32 Bit	[0..*]
BUILDING_OCCUPANT{0-n}.pID	Integer, 32 Bit	[0..*]
BUILDING_SOURCE{0-n}.sID	Integer, 32 Bit	[0..*]
bPicture (1-5)	Varchar, Unicode/UTF-8	[0..1]
bGeometry	Varchar, Unicode/UTF-8	[1..1]

Tab. 8.6: Attribute der Objekte, die Gebäude repräsentieren

Die übrigen Objekte hatten die folgenden Eigenschaften:

Attribut	Typ/Charset/Encoding	Optionalität
loID	Integer, 32 Bit	[1..1]
Name	Varchar, Unicode/UTF-8	[1..1]
Type	Varchar, Unicode/UTF-8	[1..1]
Beschreibung	Varchar, Unicode/UTF-8	[1..1]
Jahr	Integer, 16 Bit	[1..1]
loGeometry	Varchar, Unicode/UTF-8	[1..1]

Tab. 8.7: Attribute der restlichen Objekte

Objekte ohne Geometrie existierten nicht. Die geometrischen Eigenschaften wurden bereits erörtert. Die Geometrie wurde – analog zu den übrigen Objekten (10) – durch einen GeometryExtractor (3) als WKT offengelegt:

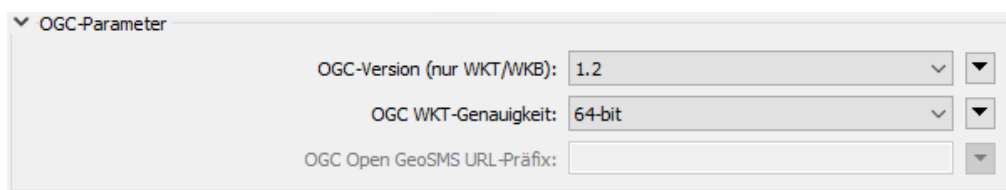
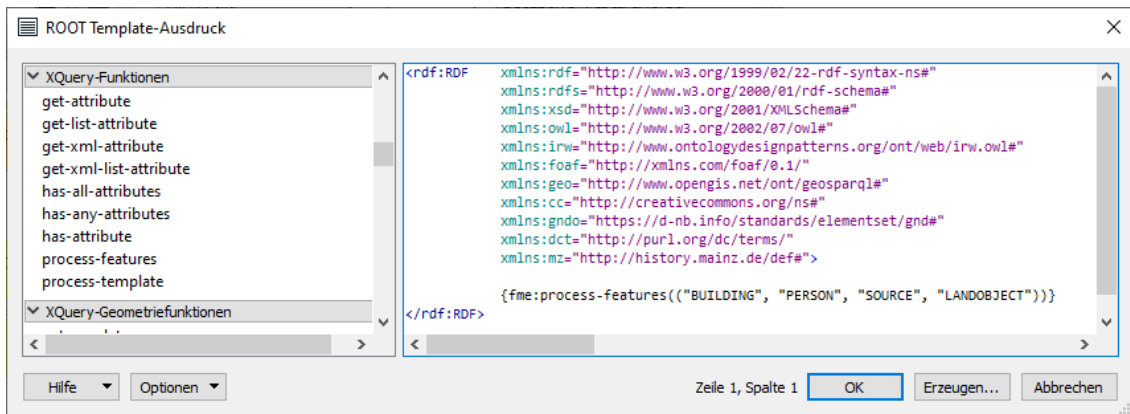


Abb. 8.15: WKT-Export, EK

Damit waren die vorbereitenden Manipulationen zwecks RDF-Generierung durch den XML-Templater abgeschlossen. Dieser verwendete verschiedene, vordefinierte XML-Fragmente, die anschließend den entstehenden Graphen kodierten. In ihrer Gesamtheit können alle notwendigen XML-Fragmente als Vorlage für die Zielstruktur angesehen werden. Diese Vorlage ist prinzipiell mit einem XSL-Template vergleichbar. Da das XML-Template mit den jeweiligen Attributwerten der manipulierten Geoobjekte gefüllt werden musste, bedurfte es zunächst eines Root-Templates zwecks Definition des Wurzelements (WE).

¹²⁰ [1..1] = kommt genau einmal vor
 [0..1] = kommt keinmal oder einmal vor
 [0..*] = kommt keinmal oder beliebig oft vor



Lis. 8.1: WE mit Namensraum-Deklaration und Prozessierungsanweisungen (XML/XQuery), EK

Die jeweiligen Sub-Templates wurden durch die XQuery-Funktion (**process-features**) aufgerufen, sobald der Creator-Transformer (19) ein Objekt erzeugte, das als Trigger für das Root-Element fungierte. Der Prolog wurde vom XML-Formater (21) geschrieben, der neben der Formatierung („Pretty-Print“) auch den Zeichensatz / die Zeichenkodierung (Unicode/UTF-8) sowie den UNC-Pfad für die Ausgabe festlegte. Die Funktionsweise kann exemplarisch anhand der folgenden Abbildung erläutert werden, die das Sub-Template BUILDING zeigt. Weil die einfacheren Sub-Templates PERSON, SOURCE, LANDOBJECT danach selbsterklärend sind, werden diese nicht separat thematisiert. Da die Syntax-Hervorhebung von FME ungewöhnlich ist, wurde das gesamte XML-Template in textueller Form dem Anhang hinzugefügt.

```
<rdf:Description rdf:about="http://www.mainz.de/id/B{fme:get-attribute("bID")}">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/B{fme:get-attribute("bID")}"/>
  <rdf:type rdf:resource="http://history.mainz.de/def#{fme:get-attribute("bFunction")}"/>
  <foaf:name xml:lang="de">{fme:get-attribute("bName")}</foaf:name>
  { if (fme:has-attribute("bInfo") and
    (string(fme:get-attribute("bInfo")) != ''))
    then <rdfs:comment xml:lang="de">{(normalize-space(fme:get-attribute("bInfo")))}</rdfs:comment>
    else () }
  <mz:when rdf:datatype="http://www.w3.org/2001/XMLSchema#Year">{fme:get-attribute("bYear")}</mz:when>
  { for $ow_ID in fme:get-list-attribute("BUILDING_OWNER{}.pID")
    return <mz:hadOwner rdf:resource="http://www.mainz.de/id/A{$ow_ID}"/> }
  { for $oc_ID in fme:get-list-attribute("BUILDING_OCCUPANT{}.pID")
    return <mz:hadOccupant rdf:resource="http://www.mainz.de/id/A{$oc_ID}"/> }
  { if (fme:get-attribute("bSocialGroup") ne "Unknown")
    then <mz:wasUsedBy rdf:parseType="Resource"><rdf:type rdf:resource="http://history.mainz.de/def#{fme:get-attribute("bSocialGroup")}"/></mz:wasUsedBy>
    else () }
  { for $i in 1 to 5
    let $P := (concat('bPicture', $i))
    return
      if (fme:has-attribute($P))
        then <foaf:depiction rdf:resource="{fme:get-attribute($P)}"/>
        else () }
  { for $s_ID in fme:get-list-attribute("BUILDING_SOURCE{}.sID")
    return <gndo:relatedWork rdf:resource="http://www.mainz.de/id/S{$s_ID}"/> }
  <mz:where rdf:resource="http://www.mainz.de/id/GEOB{fme:get-attribute("bID")}"/>
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/doc/B{fme:get-attribute("bID")}">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/B{fme:get-attribute("bID")}"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/B{fme:get-attribute("bID")}"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/">
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/id/GEOB{fme:get-attribute("bID")}">
  <rdf:type rdf:resource="http://www.opengis.net/ont/sf#Polygon"/>
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/GEOB{fme:get-attribute("bID")}"/>
  <geo:asWKT rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">&it;http://www.opengis.net/def/crs/EPSSG/0/4326&gt; {fme:get-attribute("bGeometry")}</geo:asWKT>
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/doc/GEOB{fme:get-attribute("bID")}">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/GEOB{fme:get-attribute("bID")}"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/GEOB{fme:get-attribute("bID")}"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/">
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description>
```

Lis. 8.2: Sub-Template BUILDING (XML/XQuery), EK

Die bedingten Anweisungen waren erforderlich (vgl. Tab. 8.6). Ein Objekt, das z.B. keinen Wert für „bInfo“ aufwies, sollte in der resultierenden RDF/XML-Serialisierung kein leeres XML-Element namens `rdfs:comment` erhalten. Falls Attributwerte vorhanden waren, schrieb die XQuery-Funktion `get-attribut` die entsprechenden Werte ins XML-Dokument. Die Schleifen mussten nicht verschachtelt werden; sie konnten unmöglich doppelte Geoobjekte erzeugen, da die Informationen (ggf. mittels Listenstrukturen) mit dem jeweiligen Geoobjekt verbunden wurden. Auch im XML-Templater wurden nochmals Attributwerte manipuliert. Die `normalize-space`-Funktion veränderte ggf. die übergebene Zeichenkette, indem die führenden und folgenden Whitespaces entfernt und mehrere aufeinanderfolgende Whitespaces durch ein einzelnes Leerzeichen ersetzt wurden. Dies waren lediglich kosmetische Änderungen, da die unstrukturierten Bauinformationen keine Zeichen enthielten, die nicht vom Parser analysiert werden durften.

Bezugnehmend auf die Ausführungen im Grundlagenteil dieser Arbeit wird die URI-Generierung abschließend bündig thematisiert:

- Nach der Base-URI (<http://www.mainz.de>) wurde `/id/` eingefügt, wenn die URI eine Nicht-Informationsressource identifiziert (NIR-URI). Nach der Base-URI wurde `/doc/` eingefügt, wenn die 303-URI eine Informationsressource identifiziert, deren Web-Repräsentationen zugehörige Beschreibungen sind (vgl. GP 10.2). Damit wurde dem Vorschlag von van den Brink et al. (2014) gefolgt.
- Zuletzt wurde der erzeugte Surrogatschlüssel eingefügt, der als Primärschlüssel (PK) verwendet wurde. Allemang et al. (2020) wiesen darauf hin, dass bei der Generierung von RDF-Repräsentationen auf Basis tabellenähnlicher Ausgangsstrukturen berücksichtigt werden sollte, dass numerische PK zum Problem bei der verkürzten Darstellung mithilfe des Namespace-Mechanismus werden können. Beispielsweise könnte die Definition eines Präfixes (`id`) zweckmäßig sein (s.u.), sodass <http://www.mainz.de/id/> abgekürzt und das Haus „Zum kleinen Schlüssel“ anschließend über `id:743` referenziert werden kann. Diese „compact URIs“ (CURIEs) sind im Umfeld des RDF häufig notwendig oder nützlich. Zwar wäre <http://www.mainz.de/id/743> gemäß RFC 3986 als syntaktisch korrekte URI anzusehen, jedoch besagt die W3C Recommendation für XML, dass CURIEs nach dem Präfix nicht mit einer Ziffer beginnen dürfen (Bray et al., 2008). Deshalb schlugen Allemang et al. (2020) vor, dem numerischen PK alphabetische Zeichen voranzustellen. Dies wurde bei der Erstellung des XML-Templates berücksichtigt, weshalb CURIEs verwendet werden können (z.B. `mz:B743`).
- In der ABox wurden die Nicht-Informationsressourcen explizit als Instanzen der Klasse `NonInformationResource` ausgewiesen, die in der IRW-Ontologie spezifiziert ist (vgl. Abb. 2.19). Für die zugehörigen Beschreibungen wurde analog verfahren; sie wurden als Instanzen der Klasse <http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription> ausgewiesen. Damit wurde dem Vorschlag von Halpin and Presutti (2009) gefolgt.
- Im Sub-Template LANDOBJECT wurde die URI zwecks Instanziierung der Geometrie aus der WKT-Darstellung mittels Ersetzungs- und Verkettungsfunktionen erzeugt (s. Anhang). Dies war im Sub-Template BUILDING nicht erforderlich, da Gebäude immer von Polygonen repräsentiert werden.

8.2.2 Konzeptualisierung

URI-Allokation

Van den Brink et al. (2014) folgend wird das u.g. URI-Muster zwecks Modellierung der simplen TBox verwendet:

`http://history.mainz.de/def#`

Die manuelle Erstellung der NIR-URIs beruhte auf der „URI Strategy“, die van den Brink et al. (2014, p. 70) präsentierten:

Variable	Belegung	Begründung
{concept}	history	Ist wie bei van den Brink et al. (2014, p. 72) „a generic, conceptual name“.
{type}	def	Die Nicht-Informationsressourcen werden über Hash-URIs identifiziert (Brink et al., 2014).
{domain}	mainz.de	Widerspricht zwar den Empfehlungen von Brink et al. (2014), allerdings ist es unrealistisch, dass sich der Name der Stadt in naher Zukunft ändert und abgesehen davon, wird im konkreten Fall die Geschichte der Stadt Mainz beschrieben ¹²¹ .

Tab. 8.8: Überlegungen zur URI-Allokation

Einfachheitshalber wird in den folgenden Erläuterungen auf jegliche Kurzschreibweisen verzichtet, die Turtle anbietet, mit Ausnahme der Namensräume, die zwecks Übersichtlichkeit wie folgt deklariert werden:

```
PREFIX mz: <http://history.mainz.de/def#> .
PREFIX id: <http://www.mainz.de/id/> .
PREFIX geo: <http://www.opengis.net/ont/geosparql#> .
PREFIX owl: <http://www.w3.org/2002/07/owl#> .
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

Klassen

Unter GP 6.3 wurde definiert, dass die TBox alle erforderlichen Konzepte enthalten muss, die in Tab. 5.1 - Tab. 5.5 gelistet oder beschrieben wurden. Die erforderlichen Konzepte wurden in zwei Gruppen (Klassen, Prädikate) eingeteilt (vgl. GP 7.2.2). Zunächst werden die Klassen erläutert. Der folgende Screenshot zeigt die Klassenhierarchie, die mithilfe der Software Protégé Desktop erstellt und über das Plugin OntoGraf visualisiert wurde. Die spezifizierten Properties sind nicht dargestellt; sie werden separat behandelt, denn Properties werden in RDFS nicht für Klassen erstellt. Die Featuretypen aus den o.g. Tabellen werden den Klassen in der TBox gegenübergestellt. Außerdem wird die Semantik erläutert.

¹²¹ Dienste/Applikationen sollen perspektivisch unter `https://history.mainz.de/application/...` angesprochen werden.

Historische soziale Gruppen

Eine Klasse für historische soziale Gruppen existiert in den ursprünglichen MapInfo-Relationen nicht. In der TBox ist die Oberklasse erforderlich, um den Definitions- und den Wertebereich des Prädikats `mz:wasUsedBy` in einer einfachen Form angeben zu können, da die spezifischeren Gruppen allesamt als Unterklassen von `mz:HistoricSocialGroup` ausgewiesen werden können. Alle Klassen, die unter diesem GP genannt werden, sind zunächst nur für Aussagen vorgesehen, die zum Inhalt haben, dass ein historisches Feature von einer historischen sozialen Gruppe genutzt wurde. Perspektivisch kann die Klasse `mz:HistoricSocialGroup` z.B. auch in Aussagen verwendet werden, die zum Inhalt haben, dass eine Person Mitglied einer Gruppe war (vgl. Flexibilität/ Erweiterungsmöglichkeit – [AS2/Ziel](#)). Ein Prädikat (das analog zu `foaf:member` gestaltet werden könnte) ist momentan nicht erforderlich. Zunächst müssen weitere Informationen zur historischen sozialen Gruppe erfasst werden. Da diese Informationen aktuell nicht existieren, kann in der ABox lediglich ausgedrückt werden, dass es eine/mehrere Instanz(en) von `mz:HistoricSocialGroup` (oder einer/mehrerer Unterklassen davon) gab. In Ermangelung sonstiger Informationen sind alle Instanzen von `mz:HistoricSocialGroup` jedoch anonym – sie werden von leeren Knoten repräsentiert.

Die Klassen, welche unter diesem GP genannt werden, sind nicht für direkte Aussagen über Gebäude vorgesehen, z.B.:

```
id:B404    rdf:type    mz:Jew .
```

Deshalb gibt es im resultierenden Graphen z.B. keine jüdischen Gebäude, sondern nur Gebäude, die von einer Gruppe von Juden genutzt wurden. Darin liegt ein wesentlicher Unterschied zum vorgestellten WMS bzw. den Ausgangsdaten. Die korrespondierende Legende enthält z.B. den Eintrag „jüdische Gebäude“ (vgl. [Abb. 4.5](#)).

Wichtig ist, dass Instanzen von `mz:HistoricSocialGroup` (und von den Unterklassen) eine Gruppe von Personen repräsentieren, die Gemeinsamkeiten hatten. Dies können z.B. die Dominikaner (eine Instanz von `mz:Christian`), die Grafen von Nassau-Wiesbaden (eine Instanz von `mz:Aristocracy`), eine bestimmte Strömung innerhalb des Judentums (eine Instanz von `mz:Jew`), die Scherer- und Schreinerzunft (eine Instanz von `mz:Guild`), die Diener des Kurfürsten (eine Instanz von `mz:Court`) usw. sein.

Sämtliche Werte, die in der folgenden Tabelle in der Mitte gelistet werden, sind in den originären MapInfo-Relationen enthalten, wo sie in der Spalte „Nutzung“ zu finden sind (vgl. [Tab. 5.1](#)).

Klassenname in der TBox	Wert in der Spalte „Nutzung“	In der ABox eine bestimmte Gruppe...
Aristocracy	Adel	... mit gesellschaftlichem Vorrang.
Bourgeois	Buerger	... von Bürgerlichen.
Christian	Kir_bes	... von Christen.
Court	K_Dienst	... am Fürstenhof.
Guild	Zunft	... von Gewerbetreibenden.
Jew	juedisch	... von Juden.
Patrician	Patriz	... von Privilegierten – die patrizischen Familien.

Tab. 8.9: Das „Nutzungsproblem“

Hinsichtlich der Klasse `mz:HistoricSocialGroup` bestehen noch Unwägbarkeiten, die sich mithilfe der MapInfo-Relationen nicht ausräumen lassen. Die o.g. Unterklassen wurden heterarchisch arrangiert. Die Oberklasse `mz:HistoricSocialGroup` ist eine Spezialisierung von `foaf:Group` und steht mit `geo:Feature` in keiner Relation. Die Klasse `foaf:Group` ist – im Gegensatz zu `foaf:Person` – keine Unterklasse von http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing. Es kann noch nicht abgeschätzt werden, ob die Gruppen (sinnvoll) verortet werden können.

Historische Geobjekte

Es wurden nur historische Features mit räumlichen Eigenschaften (historische Geobjekte) als Spezialisierungen von `geo:Feature` ausgewiesen. Die Oberklassen `mz:HistoricFeature`, `mz:ConstructionObject` und `mz:LanduseObject` existieren in den MapInfo-Relationen nicht. Die beiden letztgenannten Klassen wurden von Scheider et al. (2014) übernommen. Es wurden keine Äquivalenzen zur Historicmapsphen-Ontologie angegeben, weil die Unterklassen differieren. Anders als in der Historicmapsphen-Ontologie gibt es keine Klasse mit dem Bezeichner (mz:) `Landform`. Stattdessen wurde entschieden, die Klasse `mz:WaterBody` als Unterklasse von `mz:LanduseObject` auszuweisen (Begründung s.u.). Da alle Unterklassen von `mz:LanduseObject` mit dem gleichen „Namen“ bezeichnet wurden (vgl. Tab. 5.4, Tab. 5.5), ist eine Gegenüberstellung nicht erforderlich. In den angepassten MapInfo-Relationen stehen die Klassenbezeichner in der Spalte „Type“. Tab. 5.4 und Tab. 5.5 kann mit Abb. 8.16 verglichen werden, welche die Klassenhierarchie zeigt. Dies verdeutlicht, dass alle erforderlichen Klassen in der TBox abgebildet wurden.

Die Klassen, die in den Häuser-Relationen enthalten sind (vgl. Tab. 5.1), müssen jedoch thematisiert werden, da sie sich darin verbergen. Dies liegt u. a. an den teils kryptischen Bezeichnungen. Deshalb erfolgt die Gegenüberstellung mithilfe der folgenden Tabelle:

Klassenname in der TBox	Wert in der Spalte „Nutzung“	Beispiel aus der ABox
Building	unklar	Das Haus „Zum Friedberg“, das 1594 Adam Conradi (Präsenzmeister im Dom) gehörte
MilitaryBuilding	Befestig	Die „Alexander-Kaserne“ (1871)
BusinessBuilding	Wirt_geb	Das „Fladenbackhaus“, das 1433 Henne Cronenberger gehörte
Inn	Herberge	Die Herberge „Zum goldenen Adler“, die 1614 dem Wirt Johann Dieter Kaub gehörte
PublicBuilding	oeffentl	Weltliches Gerichtshaus (Neubau von 1611).
University	UNI	Die „Domus Universitatis“ (1620)
Tower	Turm	Der „Neidhartsturm“ (1450)
Church	Kirche	Der „Dom St. Martin“ (1450)

Tab. 8.10: Gebäudetypen

Abb. 8.16 zeigt, dass die o.g. Klassen – im Gegensatz zu den sozialen Gruppen – hierarchisch arrangiert wurden. Unter diesem und dem vorherigen GP wurde zunächst belegt, dass die TBox alle notwendigen Klassen enthält. Die Details dazu werden anhand möglicher Inferenzen erläutert.

Schlussfolgerungen

Dabei wird auf die ABox Bezug genommen, um einige abgeleitete Tripel darstellen zu können. Da für die Klassenhierarchie vor allem das Prädikat `rdfs:subClassOf` relevant ist, wird die korrespondierende Inferenzregel (IR) mit dem folgenden SPARQL CONSTRUCT Statement dargestellt:

```
CONSTRUCT {?id rdf:type ?Y .}
WHERE {?X rdfs:subClassOf ?Y .
      ?id rdf:type ?X .}
```

Das Beispiel aus der ABox zeigt die transitive Unterklassenbeziehung und die daraus folgenden Ableitungen.

Gegeben ist:

```
id:L037    rdf:type    mz:Bridge .
```

Daraus kann gefolgert werden:

```
id:L037    rdf:type    owl:Thing .
id:L037    rdf:type    geo:SpatialObject .
id:L037    rdf:type    geo:Feature .
id:L037    rdf:type    mz:HistoricFeature .
id:L037    rdf:type    mz:ConstructionObject .
id:L037    rdf:type    mz:Infrastructure .
```

Erwähnenswert ist diesbezüglich, dass `mz:HistoricFeature` als Unterklasse von `geo:Feature` ausgewiesen wurde, um die historische Sichtweise hervorzuheben. Insofern wurde auch `mz:HistoricGeometry` als Unterklasse von `geo:Geometry` spezifiziert. Die Entscheidung basiert auf der Tatsache, dass historische Aspekte fokussiert werden. Es ist irrelevant, ob ein historisches Feature noch existiert. Fakt ist, dass die Geometrie aus historischen Karten abgeleitet wurde. So stimmt z.B. die Geometrie des Mainzer Doms im generierten Graphen nicht mit der Geometrie des Mainzer Doms in aktuellen amtlichen Kartenwerken überein. Dies wurde für die (menschliche) Interpretation mithilfe des Prädikats `rdfs:comment` ausgedrückt, das (faktisch) keine formale Semantik besitzt, weshalb auch keine IR beigefügt wird. Der Klassenbezeichner erhält den Zusatz **Historic** an der Stelle, wo die Verknüpfung mit den externen Vokabularen über das Prädikat `rdfs:subClassOf` stattfindet. Dementsprechend gibt es noch einen Klassenbezeichner namens `mz:HistoricSocialGroup`.

KS können mit RDFS-PLUS-Konzepten approximiert werden. Bezugnehmend auf die o.g. IR ist erwähnenswert, dass nicht ausgedrückt werden kann, dass $A \cap B \subseteq C$ gelten soll, aber es kann spezifiziert werden, dass $C \subseteq A \cap B$ gelten soll. Dies wurde für die Klasse `mz:MilitaryBuilding` festgelegt:

```
mz:MilitaryBuilding    rdfs:subClassOf    mz:Building .
mz:MilitaryBuilding    rdfs:subClassOf    mz:Fortification .
```


Dies war sinnvoll, da für **mz:Building** die Legaldefinition des Begriffs Gebäude aus der Landesbauordnung Rheinland-Pfalz (LBauO) angenommen wurde (vgl. §2 II 1 LBauO bzw. Inhalt des XML-Elements namens **rdfs:comment**). In der ABox qualifizieren sich Instanzen der vorgenannten Klasse immer als Gebäude, da sie die Voraussetzungen erfüllen. Die o.g. Alexander-Kaserne ist ein Beispiel dafür¹²². Es wurden nur Objekte als **mz:MilitaryBuilding** getypt, die auch eine Wehrfunktion hatten. Deshalb sind sie nicht nur Gebäude, sondern auch Befestigungsanlagen, also Instanzen der Klasse **mz:Fortification**, was aus der Aussage:

```
id:B81      rdf:type      mz:MilitaryBuilding .
```

abgeleitet werden kann. Dies gilt insbesondere hinsichtlich der Instanzen der speziellen Unterklasse **mz:Castle**, da für die beiden Burgen bekannt ist¹²³, dass sie Räume bzw. Dächer hatten, allerdings auch der Abwehr von Angreifern dienten. Insofern wird ihnen eine Wohn- und Wehrfunktion attestiert. Diesbezüglich ist erwähnenswert, dass der letztgenannte Aspekt für Schlösser (Instanzen der Klasse **mz:Palace**) nicht angenommen wurde. Die konkreten Ausprägungen dieser Klasse haben alle repräsentativen Charakter und sind dementsprechend nicht als Befestigungsanlagen anzusehen. Trotzdem kann die Entwicklungsgeschichte dargestellt werden; z.B. wird die Martinsburg (Typ: **mz:Castle**) für die Zeit um 1450 als Vorgängerbau des Kurfürstlichen Schlosses ausgewiesen, wobei die Wehrfunktion betont wird. Später ist das Kurfürstliche Schloss als Wohn- und Repräsentativbau über eine eigene URI identifiziert und als **mz:Palace** getypt, da die Wehranlagen zerstört bzw. abgetragen wurden. Hinsichtlich der Befestigungsanlagen sollte noch erläutert werden, warum der Turm (**mz:Tower**) nur Unterklasse von **mz:Fortification** ist. Der Grund liegt in den Ausgangsdaten, die keine weitere Differenzierung zulassen. Im Rahmen der Recherche bzw. beim Abgleich mit den historischen Karten wurde klar, dass Türme auch Gebäude sein können. Aufgrund der thematischen Ausrichtung dieser Arbeit wurden die entsprechenden Attribute jedoch nicht erfasst, also vor allem die Angaben, ob ein Turm ein Dach oder Räume hatte. Deshalb wurde zunächst auf den Primärzweck eines Turms abgestellt. Hier sind alle Türme immer Befestigungsanlagen, aber nicht immer Gebäude. Wenn die o.g. Informationen nacherfasst werden, könnte man eine differenzierte Klassifizierung über den ETL-Prozess realisieren und die TBox entsprechend erweitern, ohne dass die aktuellen Beschreibungen dadurch inkonsistent werden. Man könnte die allgemeine Klasse **mz:Tower** adäquat spezialisieren und KV/KS mit OWL-Konzepten modellieren.

Bezüglich KV mittels RDFS wird auf die o.g. IR verwiesen. Über **rdfs:subClassOf** kann ausgedrückt werden, dass $C \supseteq A \cup B$ gilt.

Dies wurde z.B. für bestimmte Arten von Gewässern angegeben:

```
mz:River      rdfs:subClassOf      mz:WaterBody .
mz:OxbowLake  rdfs:subClassOf      mz:WaterBody .
```

Deshalb kann auf der Instanzebene z.B. abgeleitet werden:

Wenn

```
?id      rdf:type      mz:River .
```

Oder

```
?id      rdf:type      mz:OxbowLake .
```

¹²² Für eine Darstellung dieses Gebäudes siehe:

<http://www.festung-mainz.de/festung/kasernen/alexanderkaserne.html>

¹²³ Momentan gibt es lediglich zwei Instanzen und zwar die Schweikhardsburg und die Martinsburg.

Dann

```
?id    rdf:type    mz:WaterBody .
```

Und

```
?id    rdf:type    mz:LanduseObject .
```

usw.

Auf der Instanzebene wurden Ressourcen mitunter auch einfach der Oberklasse **mz:WaterBody** zugeordnet, nämlich dann, wenn die verfügbaren Attribute keine differenzierte Klassifizierung zuließen. Für dieses Modell ist die Gewässerordnung irrelevant, weshalb für **mz:River**, **mz:Creek**, **mz:Pond**, **mz:OxbowLake** keine Hierarchie gebildet wurde. Bezüglich der Instanzen der Klasse **mz:OxbowLake** kann anhand der Ausgangsdaten nicht ermittelt werden, ob die Altarme natürlich entstanden oder durch menschliche Einwirkungen. Letzteres ist wahrscheinlicher aber nicht sicher. Insofern werden die toten Arme genauso behandelt wie die übrigen Arten von Gewässern natürlichen Ursprungs. Die Oberklasse **mz:LanduseObject** wurde spezifiziert, um hervorzuheben, dass die Gewässer von Menschen auf eine bestimmte Art genutzt wurden; z.B. wurden im Rhein Mühlen installiert, die Umbach wurde von den Tuchwebern gewerblich genutzt, usw. Dies gilt für die Grünflächen und deren spezifische Unterklassen analog. Erwähnenswert ist diesbezüglich, dass die Instanzen der Klasse **mz:Vineyard** tatsächlich als Weingärten in der Nähe von Gebäuden anzusehen sind und nicht als Weinberge. Gerade bei Gärten wird betont, dass die Menschen diese anlegten, um die Früchte zu ziehen; beim Wein- und Kräutergarten in physischer und beim Lustgarten eher in psychischer Hinsicht. Da die Instanzen der Klasse **mz:Hedgelaying** keine Hecken waren, die (verwertbare) Früchte trugen, sondern höchstens Feinde, die sich im **Gebück** verfangen, ist diese Klasse nicht als Unterklasse von **mz:LanduseObject**, sondern als Unterklasse von **mz:ConstructionObject** ausgewiesen.

Prädikate

Da einige Prädikate aus etablierten Vokabularen übernommen werden konnten, mussten nur fünf Prädikate definiert werden. Damit können alle notwendigen Aussagen getroffen werden. Diese beantworten die Fragen:

- Wann? (indem der Zeitraum aus der Spalte „Jahr“ referenziert wird)
- Wer? (indem die Gruppe aus der Spalte „Nutzung“ sowie die Personen aus den Spalten „Eigentuemmer“ und „Bewohner“ referenziert werden)
- Wo? (indem die Geometrie im WKT-Format referenziert wird)

Damit konnten alle Angaben aus den originären und den angepassten MapInfo-Relationen (vgl. Tab. 5.1 - Tab. 5.5) in RDF-Tripelstrukturen übertragen werden. Das Prädikat, welches die Geometrie referenziert, ist dabei nicht unbedingt erforderlich, da es auch direkt aus dem GeoSPARQL-Vokabular „entnommen“ werden kann. Die Spezialisierung (als **where**) wurde von Scheider et al. (2014) übernommen.

Nachfolgend werden die Entscheidungen hinsichtlich der spezifizierten Prädikate dargelegt. Zunächst wird dabei die Verwendung der RDFS-Konzepte **rdfs:domain** und **rdfs:range** fokussiert. Vorab wird die IR für **rdfs:domain** angegeben, die für **rdfs:range** entsprechend gilt:

```
CONSTRUCT {?id rdf:type ?D .}
WHERE {?p rdfs:domain ?D .
      ?id ?p ?x .}
```

Aus der Kombination der vorstehenden IR mit der o.g. IR (`rdfs:subClassOf` i.V.m. `rdfs:domain`) folgt:

```
CONSTRUCT {?p rdfs:domain ?Y .}
WHERE {?p rdfs:domain ?D .
       ?D rdfs:subClassOf ?Y .}
```

Die Auswirkungen werden zunächst anhand des Prädikats `mz:hadOwner` dargelegt, das wie folgt definiert ist:

```
<!-- http://history.mainz.de/def#hadOwner -->

<owl:ObjectProperty rdf:about="&mz;hadOwner">
  <rdfs:label xml:lang="en">hadOwner</rdfs:label>
  <rdfs:domain rdf:resource="&mz;Building"/>
  <rdfs:range rdf:resource="&foaf;Agent"/>
  <rdfs:comment xml:lang="en">An owner is a single person, a family, a corporate body or any
  other group of persons who had the exclusive rights and the control over the entire
  property or parts of the property.</rdfs:comment>
</owl:ObjectProperty>
```

Lis. 8.3: Property `mz:hadOwner` (RDF/XML), EK

Es wurde klargestellt, dass Properties in RDFS nicht für Klassen festgelegt werden, auch wenn die obige Verschachtelung in der RDF/XML-Serialisierung dies suggeriert. Properties in RDFS sind unabhängig von jeglichen Klassen. Das bedeutet, dass `mz:hadOwner` prinzipiell für jede denkbare Ressource in der ABox verwendet werden kann. Diese RDF-Statements können nicht gegen ein RDF-Schema „validiert“ werden. Es kommt darauf an, was aus der Verwendung des o.g. Prädikats abgeleitet werden kann.

Informell formuliert: Das o.g. Property „bildet“ die Relation zwischen den (Instanzen der) Klassen `mz:Building` und `foaf:Agent` ab. Es „fordert“, dass das Subjekt in einem Tripel vom Typ `mz:Building` sein soll (spezifiziert über `rdfs:domain`) und das Objekt vom Typ `foaf:Agent` (spezifiziert über `rdfs:range`). Wenn das Prädikat `mz:hadOwner` in der ABox vorhanden ist, z.B.:

```
id:B81      mz:hadOwner      id:A511 .
```

kann u. a. Folgendes abgeleitet werden:

```
id:A511     rdf:type       foaf:Agent .
```

```
id:A511     rdf:type       owl:Thing .
```

Es ist offensichtlich, dass die o.g. kombinierte IR das Konzept der Vererbung des OO-Paradigmas umkehrt. Auch wenn es ein solches Konzept in RDFS nicht gibt, kann konstatiert werden, dass die Verwendung des Prädikats `mz:hadOwner` aufgrund der dargelegten Inferenzen dazu führt, dass das Objekt des Tripels implizit getypt wird. Dies erfolgt, bezogen auf die Klassenhierarchie, von unten nach oben und nicht umgekehrt. Die Ausgangsdaten lassen keine genauere Klassifizierung zu (z.B. ob die o.g. Person eine natürliche oder juristische Person ist). Dementsprechend sind alle Eigentümer Instanzen der o.g. (unspezifischen) Klasse. Für das Gebäude ergibt sich aus der Verwendung des Prädikats `mz:hadOwner` keine neue Information, denn die Aussage:

id:B81 rdf:type mz:Building

kann schon aus dem erzeugten (und o.g.) Tripel:

id:B81 rdf:type mz:MilitaryBuilding .

abgeleitet werden. Das Prädikat `mz:hadOccupant` wurde ähnlich definiert wie `mz:hadOwner`:

```
<!-- http://history.mainz.de/def#hadOccupant -->

<owl:ObjectProperty rdf:about="&mz;hadOccupant">
  <rdfs:label xml:lang="en">hadOccupant</rdfs:label>
  <rdfs:domain rdf:resource="&mz;Building"/>
  <rdfs:range rdf:resource="&foaf;Agent"/>
  <rdfs:comment xml:lang="en">An occupant is a single person or a family who had the right
    to live in the whole building or parts of the building.</rdfs:comment>
</owl:ObjectProperty>
```

Lis. 8.4: Property `mz:hadOccupant` (RDF/XML), EK

Es wurde erwähnt, warum sämtliche Personen lediglich Instanzen der unspezifischen Klasse `foaf:Agent` sind. Als Wertebereich der Prädikate `mz:hadOwner` und `mz:hadOccupant` wurde ebenfalls die Klasse `foaf:Agent` definiert. Der Definitionsbereich der o.g. Prädikate ist momentan mit `mz:Building` angegeben, weil das Häuserbuch nur Eigentümer- und Bewohnerinformationen zu Gebäuden enthält. Selbstverständlich ist es denkbar, dass zukünftig z.B. ausgesagt werden soll, dass ein bestimmtes **Grundstück** einen Eigentümer hatte. Die TBox kann dann adaptiert werden (vgl. Flexibilität/ Erweiterungsmöglichkeit – [AS2/Ziel](#)). Hierfür können die OWL-Konstrukte `owl:unionOf`, `owl:intersectionOf` oder `owl:complementOf` eingesetzt werden. Ebenso könnte man Hierarchien für die Prädikate definieren (s.u.), da man z.B. im Mittelalter die Verfügungsgewalt über viele Geoobjekte – auch Menschen – haben konnte (Grundherrschaft). Momentan ist dies nicht erforderlich, weil die Angaben noch fehlen. Dies gilt nicht bzgl. des Prädikats `mz:wasUsedBy`. Solche Informationen existieren bereits. Oben wurde z.B. erwähnt, dass bestimmte Gewässer von Handwerkern genutzt wurden. Es ist nicht unwahrscheinlich, dass die Tuchweber zünftige¹²⁴ Handwerker waren. Allerdings muss dies noch verifiziert werden, bevor die Tuchweber als Instanz von `mz:Guild` ausgewiesen werden können (vgl. [GP 8.2.1](#)). Wegen solcher (realistischer) Fälle wurde der Definitions- und der Wertebereich von `mz:wasUsedBy` mit jenen Klassen angegeben¹²⁵, welche mit Sicherheit alle Ressourcen enthalten, die in der o.g. Relation stehen können:

```
<!-- http://history.mainz.de/def#wasUsedBy -->

<owl:ObjectProperty rdf:about="&mz;wasUsedBy">
  <rdfs:label xml:lang="en">wasUsedBy</rdfs:label>
  <rdfs:domain rdf:resource="&mz;HistoricFeature"/>
  <rdfs:range rdf:resource="&mz;HistoricSocialGroup"/>
</owl:ObjectProperty>
```

Lis. 8.5: Property `mz:wasUsedBy` (RDF/XML), EK

¹²⁴ Zu einer Zunft gehörend.

¹²⁵ Dadurch geht keine Information verloren, da die entsprechende Unterklasse im Rahmen des ETL-Prozesses angegeben wurde.

RDFS-Konzepte können auch eingesetzt werden, um Hierarchien für Properties festzulegen. Dies erfolgt mittels des Prädikats `rdfs:subPropertyOf`. Die IR lautet:

```
CONSTRUCT {?x ?p ?y .}
WHERE {?x ?q ?y .
       ?q rdfs:subPropertyOf ?p .}
```

Diese gilt für das Prädikat `mz:where`. Die Idee für die Spezialisierung stammt von Scheider et al. (2014).

```
<!-- http://history.mainz.de/def#where -->

<owl:ObjectProperty rdf:about="&mz;where">
  <rdfs:label xml:lang="en">where</rdfs:label>
  <rdfs:domain rdf:resource="&mz;HistoricFeature"/>
  <rdfs:range rdf:resource="&geo;HistoricGeometry"/>
  <rdfs:subPropertyOf rdf:resource="&geo;hasGeometry"/>
  <rdfs:comment xml:lang="en">Note that this property is not functional [...]</rdfs:comment>
</owl:ObjectProperty>
```

Lis. 8.6: Property `mz:where` (RDF/XML), angelehnt an Scheider et al. (2014)

Da die Konstrukte `owl:FunctionalProperty` / `owl:InverseFunctionalProperty` nicht verwendet wurden, lassen sich keine Aussagen zur Identität (`owl:sameAs`) ableiten. Solche Fragen lassen sich nur individuell mittels GeoSPARQL-Abfragen klären, weil jede Ressource über eine eigene URI identifiziert wurde. In einigen Fällen ist dies offensichtlich realitätsfremd, etwa hinsichtlich eines bekannten Kenotaphs, das zwar einmal als „Drususstein“ und einmal als „Eichelstein“ bezeichnet wurde, jedoch immer dieselbe Geometrie aufweist.

Die vier beschriebenen Prädikate wurden als `owl:ObjectProperty` definiert. Über die Objekte, welche von den vier beschriebenen Prädikaten referenziert werden, können direkte Aussagen getroffen werden. Die Objekte können also Subjekte sein. Das gilt nicht für (getypte) Literale, die vom Prädikat `mz:when` referenziert werden. Das Prädikat wurde als `owl:DatatypeProperty` definiert:

```
<!-- http://history.mainz.de/def#when -->

<owl:DatatypeProperty rdf:about="&mz;when">
  <rdfs:label xml:lang="en">when</rdfs:label>
  <rdfs:domain rdf:resource="&mz;HistoricFeature"/>
  <rdfs:range rdf:resource="&xsd;gYear"/>
</owl:DatatypeProperty>
```

Lis. 8.7: Property `mz:when` (RDF/XML), EK

8.2.3 Konsolidierung

Der generierte RDF-Graph bzw. die RDF/XML-Serialisierungen wurden im Austauschverzeichnis zwischengespeichert und mit dem „W3C RDF Validation Service“ überprüft. Der aktuelle Grundbestand¹²⁶ der DNB wurde über cURL heruntergeladen und im Austauschverzeichnis zwischengespeichert. Die folgende Tabelle enthält die Details zu den Informationen, aus denen sich der vorgenannte Grundbestand zusammensetzt.

Bezeichnung	Stand	Format	Datenmenge (kB)
DNB-Titeldaten	15.06.2020 12:20 UTC ¹²⁷	N-Triples	4.927.611 (gezippt)
GND-Daten	13.10.2020 12:20 UTC ¹²⁸	N-Triples	1.692.431 (gezippt)
GND Ontology	15.10.2019 ¹²⁹	RDF/XML	322
Ontology of agent relationships	15.10.2019 ¹³⁰	RDF/XML	135

Tab. 8.11: Details zum verwendeten Grundbestand

Die TBox, welche im Rahmen des vorherigen Teilschritts modelliert wurde, wies eine geringe Datenmenge (ca. 18 kB) auf¹³¹. Das Resultat, welches vom XMLTemplater/XMLFormatter im letzten Schritt des ETL-Prozesses ins Austauschverzeichnis geschrieben wurde, umfasste 4.398.345 Tripel. Die Datei wurde gezippt (32.104 kB). Über das Webinterface von VOS wurden benannte Graphen angelegt (s. URIs – GP 9.2). Alle Serialisierungen wurden mithilfe eines Shell-Skripts und den Bulk-Loader-Funktionen von VOS in den RDF-Store importiert.

¹²⁶ Zum initialen Datenimport eines aktuellen Grundbestandes:

https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/OAI/oai_node.html

¹²⁷ Download: https://data.dnb.de/opendata/dnb-all_lds_20200615.nt.gz

¹²⁸ Download: https://data.dnb.de/opendata/authorities_lds_20201013.nt.gz

¹²⁹ Download: / Beschreibung: <https://d-nb.info/standards/elementset/gnd#>

¹³⁰ Download: / Beschreibung: <https://d-nb.info/standards/elementset/agrelon>

¹³¹ Sie ist deshalb auch vollständig im Anhang enthalten.

8.2.4 Anreicherung

Die Vorgehensweise (A) zur Erreichung des Ziels und zur Erfüllung der zentralen Anforderung LOD-A5 unterscheidet sich vom typischen Workflow (Z) für die Erstellung einer Link-Spezifikation (vgl. Abb. 7.7 und Abb. 3.13). (A) und (Z) setzen auf dasselbe Werkzeug (Silk-Workbench) zur automatisierten Unterstützung des Verfahrens (vgl. GP 3.2.3 und GP 7.2.4). Bezugnehmend auf die Erläuterungen unter den vorgenannten GP wird noch einmal klargestellt, dass sich der Grad der Automatisierung erheblich unterscheidet: (A) zielt auf die direkte Generierung der Links ab und (Z) auf die Erzeugung von Referenz-Links oder die Verwendung von existierenden Referenz-Links zwecks iterativer Optimierung der Verlinkungsregeln.

(Z) ist ein intelligentes Verfahren, das die Informationskosten senken kann. (A) ist ein Verfahren, das zwar zum Ziel aber auch zu hohen Informationskosten führen wird. Für die vorliegende Arbeit kommt nur (A) in Frage: Es ist entweder nicht erforderlich die Kosten zu senken (Fall 1 und 2) oder (Z) kann die Kosten hier nicht senken (Fall 3). Dies wird nachfolgend begründet, wobei auf die Instanzen Bezug genommen wird¹³², die unter GP 5.1.2 für die Verlinkung ausgewählt wurden:

Fall 1 (Publikationen):

- Im Häuserbuch existieren 32 Zeichenketten, die ein distinktives Muster aufweisen.
- Eine zweckmäßige Verlinkungsregel kann einfach spezifiziert werden; es ist kein Algorithmus für die iterative Optimierung notwendig.
- Das Resultat ist eine (kompakte) Kandidatenmenge. Die Kandidatenpaare, welche fälschlicherweise enthalten sind, können einfach und mit geringem Zeitaufwand entfernt werden.
- ➔ Deshalb ist die direkte Erzeugung der Links in diesem Fall am effizientesten.

Fall 2 (Geoobjekte)

- Im Häuserbuch gibt es 7802 Objekte, die über Koordinaten lokalisiert wurden. In der GND sind nicht alle Geographika über Koordinaten lokalisiert.
- Die Selektion mithilfe der Koordinaten ist nicht zwingend notwendig, weil die potenziellen Kandidaten in der GND trotzdem über eine einfache SPARQL-Abfrage mit einer räumlichen Bedingung selektiert werden können.
- Für die Verlinkung kommen nur 75 Ressourcen aus der GND in Frage.
- Die Pendants, welche im Häuserbuch beschrieben werden, können einfach mithilfe einer Zeichenketten-ähnlichkeitsfunktion ermittelt werden, die auf den Namen der Geoobjekte angewendet wird. Der Name ist in der GND und im Häuserbuch immer vorhanden.
- Auch in diesem Fall ist kein Algorithmus für die iterative Optimierung der Verlinkungsregel notwendig.
- Stattdessen ist eine Funktion erforderlich, mit welcher der Link-Typ bzw. das Prädikat individuell für jede Ressource festgelegt werden kann. Gerade weil nur wenige Links erzeugt werden können, sollen die Ressourcen nicht immer über dasselbe Prädikat verlinkt werden; z.B. ist es nicht sinnvoll immer `owl:sameAs` zu verwenden, da es sich manchmal (offensichtlich) nicht um URI-Aliase handelt. Die Ressourcen können aber Ähnlichkeiten aufweisen und dann sollen sie ebenfalls verknüpft werden. Da es eine solche Funktion in der Silk-Workbench nicht gibt, müssten die automatisch erzeugten Links im Nachgang angepasst werden oder es müsste nach einem Plugin gesucht werden.
- ➔ Davon wird abgesehen, weil auch in diesem Fall die direkte Erzeugung der Links am effizientesten ist.

¹³² Es wird auch auf die Erklärungen zu den Eigenkreationen Bezug genommen (z.B. zum „distinktiven Muster“).

Fall 3 (Agenten)

- Im Häuserbuch werden 10.596 Personen referenziert. Es wurden aufwendige Datenerfassungen durchgeführt (s. GP 5.1.1), was anschließend – aufgrund der thematischen Ausrichtung dieser Arbeit – (mehrfach) explizit ausgeschlossen wurde (vgl. AS1 und GP 7.2.1). Im Rahmen der Umformung wurden zwar Recherchen zwecks Bereinigung der Personentabelle angestellt (s. GP 8.2.1), aber es wurden keine zusätzlichen Daten erfasst. Deshalb kann die Verlinkung nur basierend auf dem Namen einer Person durchgeführt werden. Die Literale, welche im ETL-Prozess über das Prädikat `rdfs:comment` referenziert wurden, sind dafür grundsätzlich ungeeignet. Alle anderen Kriterien scheiden auch aus. Dies wird nachfolgend erklärt. Die Erkenntnis, ob es sich um dasselbe, ein ähnliches oder ein ganz anderes Bezugsobjekt handelt, kann im konkreten Fall nicht so einfach gewonnen werden, wie im Beispiel, das anhand von Abb. 3.12 illustriert wurde. Da die Daten fehlen, sind Recherchen erforderlich.
 - Geeignete Referenz-Links existieren zwar nicht; sie können jedoch erstellt werden. Die Erstellung von Referenz-Links ist mit geringerem Aufwand verbunden, als die direkte Generierung aller Links. Referenz-Links sind im konkreten Fall allerdings nicht zielführend, da ausschließlich „Namensvergleiche“ durchgeführt werden können: Selbst wenn die „geeignetste“ Verlinkungsregel verwendet wird, erübrigt sich die manuelle Überprüfung aller automatisch erzeugten Links nicht, weil diese Menge keine Ergebnismenge sein wird, sondern lediglich eine (suboptimale) Kandidatenmenge. Diese ist nicht mit den o.g. Resultaten (Fall 1 und 2) vergleichbar, weil die Zahl der Kandidatenpaare, welche fälschlicherweise enthalten sind, sehr groß ist.
- ➔ Da die Erstellung von Referenz-Links (im konkreten Fall) keinen Vorteil bringt, werden die Links direkt generiert.

Mithilfe der Silk-Workbench lässt sich (A) effizienter umsetzen, obwohl die Anwendung dafür eigentlich nicht entwickelt wurde. Dies ist sowohl auf die Funktionen der Silk-Workbench (s.u.) als auch auf nicht-funktionale Aspekte zurückzuführen, die unter GP 3.2.3 hervorgehoben wurden.

Bestimmte Komponenten der Silk-Workbench müssen für die direkte Erzeugung der Links zweckentfremdet werden. Dadurch soll jedoch nicht suggeriert werden, dass Silk den Verlinkungsprozess nicht umfassend unterstützen kann. Deshalb werden die Bezeichnungen, die unter GP 3.2.3 wörtlich zitiert wurden, in diesem Abschnitt nur einmal verwendet, um auf eine begriffliche Trennung hinzuweisen. Die Visualisierungen des Linkage Rule Editors werden „Workbench“ genannt. Damit wird die Kandidatenmenge ermittelt. Die Tabelle, welche Silk für die Evaluierung erstellt, wird „Ausgabe“ genannt. Darüber wird festgelegt, ob ein Link erzeugt werden soll. Falls dies sinnvoll ist, wird der Link anschließend manuell via SPARQL eingefügt. Silk erhält lesenden Zugriff auf den RDF-Store. Referenz-Links werden nicht hinzugefügt. Alle Anpassungen an der Workbench werden manuell vorgenommen. Der Linkage Rule Learner wird nicht genutzt. Die Standardeinstellungen der Caching- und Indexierungsfunktionen werden nicht verändert.

Für die Erläuterung wird die Workbench und die Ausgabe verwendet (analog zum ETL-Prozess). Die Workbench wird so angeordnet, dass sie abgebildet werden kann. Die Workbench ist immer ein Baum.

Zunächst wird die Verlinkung der Publikationen behandelt. Dabei wird Wert auf Kompaktheit gelegt, denn anschließend wird der Verlinkungsprozess mithilfe von (A) ausführlich besprochen.

Workbenches

Für die Verlinkung der Publikationen (nachfolgend Literatur oder Literaturangabe[n] genannt), der Agenten und der Geoobjekte wurden separate Workbenches angelegt. Der Verlinkungsprozess kann von vornherein auf bestimmte Ressourcen eingeschränkt werden. Silk bietet hierfür unter „Linking Tasks“ eine Drop-Down-Funktion, die alle Klassen listet (Feld „Target Type“). Weitere Bedingungen zwecks Selektion spezifischer Ressourcen können in den Feldern „Source Restriction“ und „Target Restriction“ via SPARQL angegeben werden.

Literatur

Für die Verlinkung der Literaturangaben wurden die Instanzen der Klasse `http://purl.org/ontology/bibo/Document` ausgewählt. Die RDF-Beschreibungen des historischen Häuserbuchs weisen die Literaturangaben immer als Instanzen dieser Klasse aus. Für deren Verlinkung kommen nur Ressourcen aus den DNB-Titeldaten in Betracht, die ebenfalls von diesem Typ sind. Weitere Einschränkungen sind nicht sinnvoll. Es ist z.B. nicht zielführend die Titeldaten der DNB via SPARQL auf die Sprachen Deutsch und Latein einzugrenzen, da das Prädikat `dcterms:language` dort nicht immer verwendet wird. Räumliche Einschränkungen o. ä. sind nicht möglich. Die folgende Workbench (s. Abb. 8.17) zeigt die Konfiguration zwecks Verlinkung der Publikationen. Aus den Quelldaten wurde das Prädikat `foaf:name` herangezogen, da dieses auf die relevanten Literale verweist. Diese Zeichenketten (engl. strings) repräsentieren alle Literaturangaben, die im historischen Häuserbuch enthalten sind. Wichtige Informationen zur Struktur dieser Zeichenketten können den Ausführungen unter GP 5.1.1 und GP 5.1.2 oder dem Anhang entnommen werden. Für die Generierung der Links wurden „Zeichenkettenähnlichkeitsfunktionen“ (Kurek, 2011, p. 10) eingesetzt. Dafür wurden im Zieldatenbestand die Objekte ausgewählt, die von den Prädikaten `dcterms:issued` und `dc:title` referenziert werden. Auf diese Objekte wurde strukturiert zugegriffen. Dafür war die spezifizierte Semantik in der TBox hilfreich. Das Prädikat `dcterms:issued` ist eine Spezialisierung des Prädikats `dcterms:date`, das ebenfalls im Vokabular namens DCMI Metadata Terms¹³³ spezifiziert wurde. Es gibt mehrere syntaktisch valide Formen, mit denen Daten angegeben werden können. Für den systematischen Zugriff auf das Objekt war daher die Erkenntnis wichtig, dass sich das Objekt in der ABox (in den relevanten Fällen) immer als vierstellige Ziffernfolge manifestiert, obwohl die Datentyp-URI `xsd:string` verwendet wird. Zusätzlich mussten die Objekte verwendet werden, die im Zieldatenbestand von den Prädikaten `rdau:P60327` und `rdau:P60493` referenziert werden. Da diese Prädikate nicht über eine „sprechende“ URI identifiziert werden, sind die wesentlichen Informationen aus der RDA Registry¹³⁴ beigefügt. Die Definitionen sind für die folgenden Erläuterungen relevant.

CURIE	Label	Definition
rdau:P60493	"has other title information"	"Relates a resource to a word, character, or group of words or characters that appears in conjunction with, and is subordinate to, a title proper of a resource."
rdau:P60327	"has statement of responsibility relating to title proper"	"Relates a resource to a statement associated with a title proper of a resource that relates to an identification or function of agents responsible for the creation of, or who contributes to a realization of, the intellectual or artistic content of a resource."

Tab. 8.12: Informationen zu den Prädikaten `rdau:P60327` und `rdau:P60493`, Quelle: RDA Registry

¹³³ Siehe: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

¹³⁴ Siehe: <https://www.rdaregistry.info/>

Die o.g. Zeichenketten wurden folgendermaßen verwendet:

- (1) **foaf:name** (sourcePath1) wurde als Eingabe definiert. Über den Transformer (regexExtract1) wurde der Nachname des Autors mithilfe eines regulären Ausdrucks entfernt. Dann wurde ein Transformer (untilCharacter1) verwendet, um das Veröffentlichungsdatum des Werks abzutrennen. Der nächste Transformer (lowerCase1) wurde eingesetzt, um alle Großbuchstaben der Eingabezeichenkette für die Ausgabezeichenkette in die entsprechenden Kleinbuchstaben umzuwandeln. Ein weiterer Transformer (removeSpecialChars1) wurde für die Entfernung aller speziellen Zeichen (**carriage return**, **ampersand**, **whitespace** etc.) eingesetzt. Dies diente der Feststellung der Ähnlichkeit zwischen dem transformierten String und dem Literal, das vom Prädikat **dc:title** (targetPath1) referenziert wird. Bevor diese Ähnlichkeit über die Jaro Distanz (jaro) ermittelt und ausgegeben wurde, musste das Literal aus dem Zieldatenbestand ebenfalls über die vorgenannten Transformer (lowerCase2 und removeSpecialChars2) verändert werden. Für die spätere Aggregation (average) wurde das Gewicht der Jaro Distanz mit 4 festgelegt. Die Ermittlung der Jaro-Distanz wurde als obligatorischer Prozessschritt definiert.
- (2) **foaf:name** (sourcePath2) wurde als Eingabe definiert. Ein Transformer (regexExtract2) wurde eingesetzt, um das Veröffentlichungsdatum des Werks mithilfe eines regulären Ausdrucks zu extrahieren. Der Datentyp des Objekts, das vom Prädikat **dcterms:issued** (targetPath2) referenziert wird, ist **xsd:string**. Es wurde geprüft, ob die beiden Zeichenketten identisch sind (equality). Das Resultat wurde im Rahmen der Aggregation nicht so stark gewichtet wie die Jaro Distanz. Das Gewicht wurde mit 1 festgelegt. Der exakte Zeichenkettenvergleich wurde als obligatorischer Prozessschritt definiert.
- (3) **foaf:name** (sourcePath3) wurde als Eingabe definiert. Ein Transformer (regexExtract3) wurde verwendet, um erst den Nachnamen des Autors mithilfe eines regulären Ausdrucks zu entfernen, bevor ein weiterer Transformer (removeSpecialChars3) eingesetzt wurde. Dessen Funktion wurde bereits beschrieben. Als weitere Eingabe wurde **rdau:P60493** (targetPath3) definiert. Hier wurde zunächst ein Transformer (removeSpecialChars4) verwendet, bevor die Ähnlichkeit der beiden Zeichenketten mithilfe der Jaro-Winkler Distanz (jaroWinkler) ermittelt und ausgegeben wurde. Für die spätere Aggregation (average) wurde das Gewicht der Jaro-Winkler Distanz mit 2 festgelegt. Dieser Prozess war für die Ermittlung **aller** potenziellen Kandidaten erforderlich. Das Literal aus dem historischen Häuserbuch musste auch mit den weiteren Titelinformationen (vgl. Tab. 8.12) abgeglichen werden, die von der DNB mithilfe des Prädikats **rdau:P60493** angegeben werden. Da dieses Prädikat im Zieldatenbestand allerdings nicht immer vorhanden ist, musste dieser Prozessschritt als optional definiert werden.
- (4) **foaf:name** (sourcePath4) wurde als Eingabe definiert. Ein Transformer (untilCharacter2) wurde verwendet, um den Nachnamen des Autors zu extrahieren. Dieser wurde mit dem Literal verglichen, welches das Prädikat **rdau:P60327** (targetPath4) referenziert. Genauer gesagt, wurde dieses Literal, das weitere Informationen zur Publikation enthält (vgl. Tab. 8.12), über einen Token-basierten Ähnlichkeitsalgorithmus (tokenize1) in seine einzelnen Bestandteile zerlegt. Die Ähnlichkeitsprüfung (relaxedEquality) war nicht so streng wie bei (2). Für eine Ähnlichkeit von 100 % genügte es, wenn ein Token mit dem Nachnamen übereinstimmte, wobei sogar korrespondierende Groß- / Kleinbuchstaben als identisch angesehen wurden, ebenso wie z.B. die Zeichen "o" und "ö". Im Zuge der Aggregation wurde das Ergebnis mit einem Gewicht von 1 berücksichtigt. Da das Prädikat **rdau:P60327** im Zieldatenbestand nicht immer vorhanden ist, musste dieser Prozessschritt als optional definiert werden.

Für weiterführende Informationen zu den verwendeten Zeichenkettenähnlichkeitsfunktionen, wird auf die Arbeit von Kurek (2011) verwiesen (s. Edit- und Token-basierte Ähnlichkeitsfunktionen). Bündige Erläuterungen zu den Ergebnissen, die mithilfe der o.g. Transformer und Operatoren erzielt wurden, finden sich nach der folgenden Darstellung, welche die Workbench zeigt.

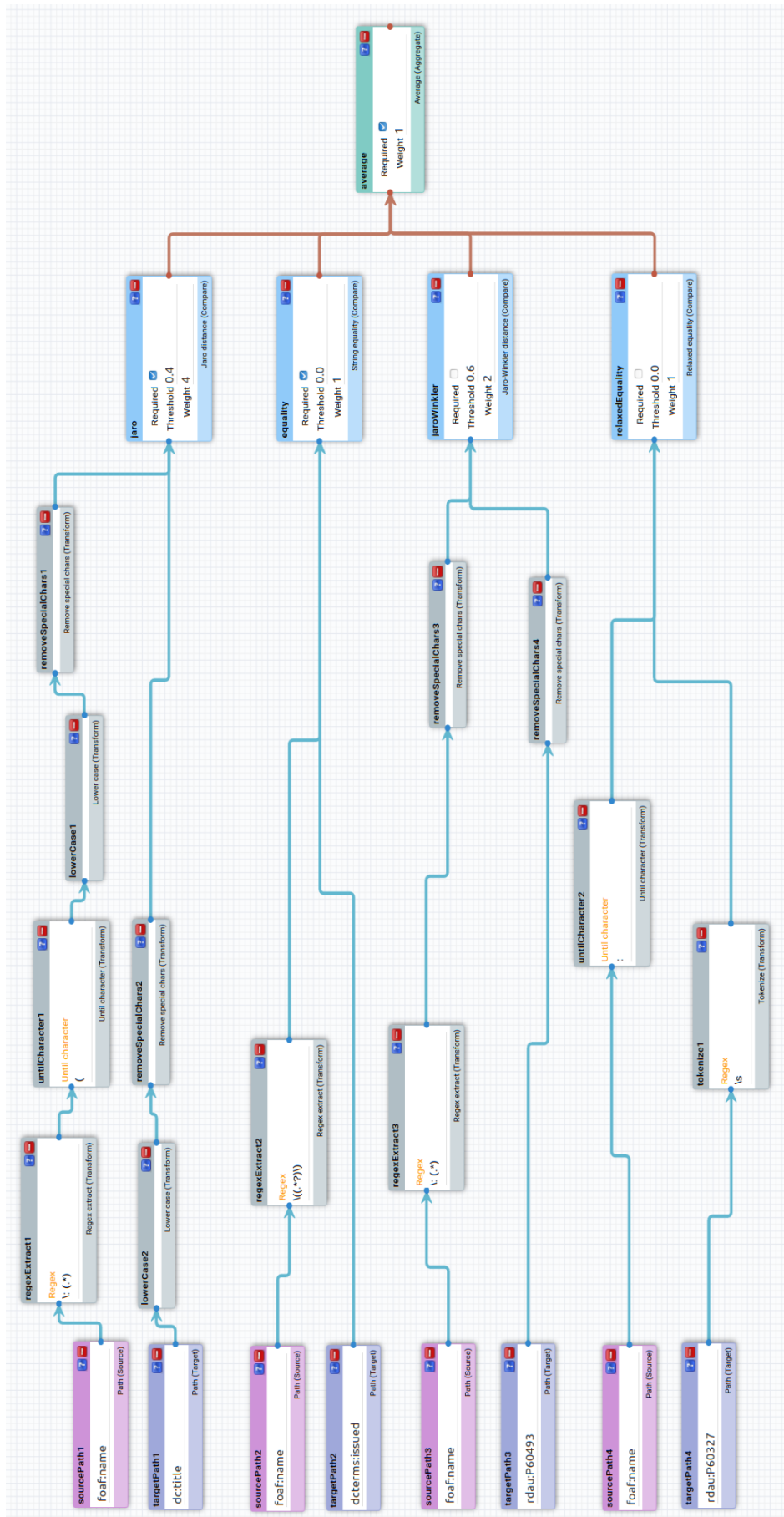


Abb. 8.17: Workbench für die Verlinkung der Literaturangaben, EK

Mit dieser Workbench wurde ein solides Ergebnis für die anschließende Evaluierung erzielt. Die Ausgabe zeigt potentielle Kandidatenpaare. Diese ist auszugsweise dargestellt. Der Anhang ist vollständig. Es wurde nach der Wahrscheinlichkeit absteigend sortiert. Die Funktionen mussten jedoch zweckentfremdet werden (vgl. GP 8.2.4). Die Ressourcen wurden untersucht, indem die jeweiligen Details über den kleinen Pfeil (links) angezeigt wurden. Danach wurde über die Buttons in der rechten Spalte festgelegt, ob ein Link eingefügt werden sollte. Die Links wurden anschließend manuell via SPARQL eingefügt.

▶ http://www.mainz.de/id/S31	https://d-nb.info/452291402	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/S12	https://d-nb.info/362677409	97.9%	✓ ? ✕
▶ http://www.mainz.de/id/S22	https://d-nb.info/362677425	93.3%	✓ ? ✕
▶ http://www.mainz.de/id/S21	https://d-nb.info/740554085	85.7%	✓ ? ✕
▶ http://www.mainz.de/id/S13	https://d-nb.info/362677387	84.9%	✓ ? ✕
▶ http://www.mainz.de/id/S1	https://d-nb.info/580166953	82.7%	✓ ? ✕
▶ http://www.mainz.de/id/S9	https://d-nb.info/363167978	81.6%	✓ ? ✕
▶ http://www.mainz.de/id/S9	https://d-nb.info/363167986	80.5%	✓ ? ✕
▶ http://www.mainz.de/id/S30	https://d-nb.info/363167994	80.2%	✓ ? ✕
▶ http://www.mainz.de/id/S16	https://d-nb.info/363167986	79.2%	✓ ? ✕
▶ http://www.mainz.de/id/S14	https://d-nb.info/930437829	78.9%	✓ ? ✕
▶ http://www.mainz.de/id/S16	https://d-nb.info/363167978	78.6%	✓ ? ✕
▶ http://www.mainz.de/id/S15	https://d-nb.info/1059715686	78.1%	✓ ? ✕
▶ http://www.mainz.de/id/S26	https://d-nb.info/362677417	77.8%	✓ ? ✕
▶ http://www.mainz.de/id/S27	https://d-nb.info/458493538	77.3%	✓ ? ✕
▶ http://www.mainz.de/id/S23	https://d-nb.info/578517760	71.3%	✓ ? ✕
▶ http://www.mainz.de/id/S2	https://d-nb.info/458493481	64.2%	✓ ? ✕
▶ http://www.mainz.de/id/S11	https://d-nb.info/458493503	63.3%	✓ ? ✕
▶ http://www.mainz.de/id/S24	https://d-nb.info/458493511	63.3%	✓ ? ✕
▶ http://www.mainz.de/id/S28	https://d-nb.info/890620601	58.3%	✓ ? ✕
▶ http://www.mainz.de/id/S4	https://d-nb.info/921274386	56.9%	✓ ? ✕
▶ http://www.mainz.de/id/S17	https://d-nb.info/363154574	50.7%	✓ ? ✕
▶ http://www.mainz.de/id/S18	https://d-nb.info/952642328	47.4%	✓ ? ✕
▶ http://www.mainz.de/id/S3	https://d-nb.info/730124045	26.3%	✓ ? ✕
▼ http://www.mainz.de/id/S19	https://d-nb.info/800741714	5.9%	✓ ? ✕

Aggregation: average (average) 5.9%

- Comparison: equality (equality) 100.0%
 - Transform: regexExtract (regexExtract2) 1978
 - Input: foaf:name (sourcePath2) Falck: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)
 - Input: dct:terms:issued (targetPath2) 1978
- Comparison: jaroWinkler (jaroWinkler) 56.6%
 - Transform: removeSpecialChars (removeSpecialChars3) MainzvomfrühenMittelalterbiszumAnfangdes17Jahrhunderts1978
 - Transform: regexExtract (regexExtract3) Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)
 - Input: foaf:name (sourcePath3) Falck: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)
 - Transform: removeSpecialChars (removeSpecialChars4) 34CKtMainzvomfrühenMittelalterbiszumAnfangdes17JahrhundertsFarbendr15000
 - Input: rdau:P60493 (targetPath3) 34C. ; [Kt.]. Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts : [Farbendr.]; 1: 5000
- Comparison: jaro (jaro) -1.17%
 - Transform: removeSpecialChars (removeSpecialChars1) mainzvomfrühenmittelalterbiszumAnfangdes17jahrhunderts
 - Transform: lowerCase (lowerCase1) mainz vom frühen mittelalter bis zum anfang des 17. jahrhunderts
 - Transform: untilCharacter (untilCharacter1) Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts
 - Transform: regexExtract (regexExtract1) Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)
 - Input: foaf:name (sourcePath1) Falck: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)
 - Transform: removeSpecialChars (removeSpecialChars2) geschichtlicheratlasvonhessen
 - Transform: lowerCase (lowerCase2) geschichtlicher atlas von hessen
 - Input: dc:title (targetPath1) Geschichtlicher Atlas von Hessen
 - Comparison: relaxedEquality (relaxedEquality) -100.0%
 - Transform: untilCharacter (untilCharacter2) Falck
 - Input: foaf:name (sourcePath4) Falck: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)
 - Transform: tokenize (tokenize1) Hess. Landesamt für Geschichtl. Landeskunde Marburg/Lahn. Begr. u. vorbereitet von Edmund E. Stengel. Bearb. von Friedrich Uhlhorn
 - Input: rdau:P60327 (targetPath4) Hess. Landesamt für Geschichtl. Landeskunde Marburg/Lahn. Begr. u. vorbereitet von Edmund E. Stengel. Bearb. von Friedrich Uhlhorn

Abb. 8.18: Kandidatenpaare für die Verlinkung der Literatur, mit Details zu S19, EK

Die URI <http://www.mainz.de/id/S9> und die URI <https://d-nb.info/800741714> identifizieren dieselbe Ressource. Es wäre nicht akzeptabel gewesen, wenn die Karte, die im Vorfeld für die Anpassung der Ausgangsdaten verwendet wurde (s. Kapitel 4 und 5), im Rahmen der Verlinkung unberücksichtigt geblieben wäre. Das Problem ist, dass das Literal, welches im historischen Häuserbuch vom Prädikat `foaf:name` referenziert wird, im konkreten Fall eine geringe Ähnlichkeit mit dem Titel hat, den die DNB mithilfe des Prädikats `dc:title` angibt. Es handelt sich also um ein Beispiel, das aufzeigt, dass der Vergleich mit der Zeichenkette notwendig war, die vom Prädikat `rdau:P60493` referenziert wird. Aufgrund der dargestellten Parameter wurde die Ähnlichkeit der beiden Zeichenketten über die Jaro-Winkler Distanz mit 56,6 % angegeben. Da dieses Ergebnis im Rahmen der Aggregation nur mit einem Gewicht von 2 berücksichtigt wurde und der Vergleich über den Token-basierten Algorithmus in Kombination mit der `relaxedEquality`-Funktion (Gewicht = 1) überhaupt keine Ähnlichkeit feststellen konnte, ergab sich insgesamt eine niedrige Wahrscheinlichkeit (5,9 %). Der Vergleich der Zeichenketten, die jeweils das Veröffentlichungsjahr des Werks angeben, änderte daran recht wenig. Hier wurde zwar eine hundertprozentige Übereinstimmung ermittelt, allerdings wurde dieses Ergebnis im Rahmen der Aggregation ebenfalls nur mit einem Gewicht von 1 berücksichtigt. Dies wurde als sinnvoll erachtet, da in einem Jahr sehr viele Werke publiziert werden. Insofern sollte das Veröffentlichungsjahr natürlich ins Gesamtergebnis eingehen, durfte es jedoch nicht entscheidend beeinflussen. Maßgeblich war hingegen die festgestellte Ähnlichkeit, welche über die Jaro Distanz ausgegeben wurde, da hierfür ein Gewicht von 4 definiert wurde. Dies war i.d.R. zielführend, denn häufig waren die Titelinformationen ausschlaggebend.

Das nächste Beispiel zeigt eine Ausnahme von dieser Regel. Es offenbart, dass die ermittelten Wahrscheinlichkeiten nur als Indikatoren anzusehen sind. Die genauere Untersuchung der Kandidaten war stets erforderlich.

Abb. 8.19: Beispiel für ein mehrteiliges Werk, EK

In diesem Fall führten alle beschriebenen Vergleiche zu akzeptablen Resultaten. Die Detailansicht zeigt aber, dass im konkreten Fall kein Link eingefügt werden durfte. Das Problem ist hier, dass die Ressource, welche über die URI <http://www.mainz.de/id/S9> identifiziert wird, eine relativ hohe Ähnlichkeit zur Ressource aufweist, welche über die URI <https://d-nb.info/363167978> identifiziert wird (81,6 %) und zusätzlich zur Ressource, welche über die URI <https://d-nb.info/363167986> identifiziert wird (80,5 %). Nur im ersten Fall identifizieren die URIs tatsächlich dieselbe Ressource (Teil 1 der Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts). Im zweiten Fall identifiziert die URI <https://d-nb.info/363167986> Teil 2 des vorgeannten Werks und korrespondiert dementsprechend mit <http://www.mainz.de/id/S16>. Da sich die beiden Zeichenketten nur in Nuancen unterscheiden, war die ermittelte Wahrscheinlichkeit jedoch ähnlich hoch.

Wenn Literaturangaben aus dem historischen Häuserbuch nicht verlinkt wurden, liegt dies daran, dass in den DNB-Titeldaten¹³⁵ kein Pendant existiert. Das historische Häuserbuch enthält, im Gegensatz zu den Agenten und Geoobjekten, eine überschaubare Anzahl an Literaturangaben (32). Deshalb konnten deren Beschreibungen mit allen erzeugten Links (24) als RDF/XML-Serialisierung dem Anhang hinzugefügt werden.

Agenten

Im Vergleich zu den Literaturangaben gestaltete sich die Verlinkung der Agenten deutlich aufwendiger. Das lag nicht nur an der Anzahl der vorhandenen Agenten (10.596), sondern auch an den Informationen und deren Struktur im Quelldaten- und im Zieldatenbestand (Normdaten der Gemeinsamen Normdatei). Die Details werden nachfolgend systematisch erörtert. Die erforderlichen (Teil-)Prozesse, Entscheidungen etc. wurden unter GP 7.2.4 thematisiert und visualisiert (vgl. Abb. 7.7).

Zwecks Ermittlung der Kandidatenmenge (Prozess 1) wurden die erwähnten Ontologien und die Fakten über konkrete Instanzen untersucht, welche in der GND vorhanden sind (Teilprozess 1.1). Basierend auf den gewonnenen Erkenntnissen wurde anschließend geprüft, ob sich die Kandidatenmenge über eine SPARQL-Abfrage bestimmen lässt. Es wäre z.B. denkbar gewesen, den GND-Datenbestand nach dem Namen der Person und dem Wirkungsjahr¹³⁶ zu durchsuchen. Idealerweise hätte die Abfrage zusätzlich durch geometrische bzw. topologische Kriterien¹³⁷ eingeschränkt werden können. Es wäre unerheblich gewesen, dass im Häuserbuch nur der Name der Person (das Literal als Objekt) unmittelbar mit dem betreffenden Subjekt (die Person) über das Prädikat `foaf:name` in Beziehung steht¹³⁸. Anhand der exemplarisch genannten Kriterien (Wirkungsjahr bzw. -zeitraum, geometrische/topologische Bedingungen, Name) und einer Analogie lassen sich einige Probleme herausstellen.

¹³⁵ Die verwendeten Informationen der DNB (Version etc.) wurden in Tab. 8.11 dokumentiert.

¹³⁶ Die Semantik des Begriffs wird hier spezifiziert: <https://d-nb.info/standards/elementset/gnd#periodOfActivity> „Exakt“ ist dabei relativ, da das Prädikat in der ABox häufig Zeiträume (z.B. ganze oder mehrere Jahre) referenziert. Nachfolgend wird das Präfix „gndo“ für <https://d-nb.info/standards/elementset/gnd#> verwendet.

¹³⁷ Gemeint ist z.B. die Prüfung, ob eine Entität (=Agent) zu einem bestimmten Zeitpunkt innerhalb eines minimal umgebenden Rechtecks (MUR) verortet werden kann.

¹³⁸ Die beiden anderen Kriterien sind einfach zu ermitteln, indem dem Pfad vom Haus (ggf. den Pfaden der Häuser) zur Person gefolgt wird. D.h. im Klartext, dass man vom Jahr des Hauses (ggf. von den Jahren der Häuser) auf das Wirkungsdatum der Person schließt und damit zumindest einen Indikator erhält. Das MUR wird aus allen Geometrien eines Modells des historischen Häuserbuchs berechnet und ggf. um einen adäquaten Toleranzbereich erweitert.

(1) Im Umfeld relationaler Datenbanken ist das Feld, welches im Rahmen einer Abfrage für die Formulierung einer einschränkenden Bedingung herangezogen werden kann, i.d.R. bekannt. Falls dies nicht zutrifft, können die notwendigen Informationen über die entsprechende Datendefinition (Statements in der Data Definition Language) ermittelt werden (z.B. der Datentyp). Generell ist dies im Kontext von LOD nicht unmöglich (vgl. GP 2.2.5: OWA oder CWA – SHACL). Im konkreten Fall konnte jedoch keine qualifizierte Entscheidung getroffen werden, welche Objekte für die Filterung via SPARQL verwendet werden sollten, um den GND-Datenbestand auf Personen einzuschränken, die in einem **bestimmten Jahr bzw. Zeitraum** agierten. Für natürliche und juristische Personen kommen u. a. die Objekte in Betracht, die mittels der folgenden Prädikate referenziert werden:

- `gndo:periodOfActivity`
- `gndo:dateOfEstablishment`
- `gndo:dateOfTermination`
- `gndo:dateOfEstablishmentAndTermination`
- `gndo:dateOfBirth`
- `gndo:dateOfDeath`
- `gndo:biographicalOrHistoricalInformation`
- ...

Im Rahmen von Teilprozess 1.1 wurde sowohl die Optionalität¹³⁹ der o.g. Prädikate in der ABox untersucht, als auch die Struktur der ggf. vorhandenen Zeichenketten. Es wurde festgestellt, dass die o.g. Prädikate und damit die für die Selektion erforderlichen Objekte häufig fehlen. Außerdem existieren viele Variationen von Zeichenketten. Es folgt ein Beispiel:

```
<rdf:Description rdf:about="https://d-nb.info/gnd/3072603-7">
  [...]
  <gndo:dateOfTermination rdf:datatype="http://www.w3.org/2001/XMLSchema#string">1577, 1612-1797
</gndo:dateOfTermination>
</rdf:Description>
```

Die Selektion über dieses Kriterium (Wirkungsjahr bzw. -zeitraum) schied aus. Es konnte keine SPARQL-Abfrage formuliert werden, die alle möglichen Arten der Datenerfassung berücksichtigt.

¹³⁹ Vgl. Erklärungen zu Tab. 8.6.

- (2) Im Umfeld objektrelationaler Datenbanken (speziell für räumliche Informationen) ist es möglich, **Koordinaten** anzugeben, um z.B. ein **MUR** mit vordefinierten Funktionen zu verwenden (etwa in PostgreSQL/PostGIS: `ST_within`). Damit kann z.B. ermittelt werden, ob das betreffende Objekt innerhalb des MURs liegt. Der Raumbezug wird im konkreten Fall allerdings nicht über primäre Metriken hergestellt (etwa mittels GeoSPARQL, WKT). Abgesehen davon, dass die entsprechenden Prädikate in der ABox ebenfalls fehlen können, sind z.B. die von `gndo:spatialAreaOfActivity`, `gndo:placeOfBusiness`, `gndo:geographicAreaCode` referenzierten Codes nicht immer konkret genug. Wenn sich lediglich die Information findet, dass ein Agent in Deutschland oder im Deutschen Reich gewirkt hat, ist dies bei der Vorauswahl nicht hilfreich. Die Zahl der irrelevanten Ressourcen ist viel zu hoch.
- (3) Je nach UoD wird man Festlegungen treffen, wie Daten erfasst werden sollen, deren Form nicht über Integritätseinschränkungen geregelt wird. Häufig ist ein einheitlicher Datenbestand erstrebenswert. Bei bestimmten Attributwerten ist hier das Gegenteil der Fall. Häufig finden sich unterschiedliche (aber trotzdem sinnvolle) **Schreibweisen für Namen** historischer Personen. Deshalb ist z.B. die Wahrscheinlichkeit, dass zwei ähnliche (aber unterschiedliche) Zeichenketten im Quell- oder Zieldatenbestand auf dasselbe Bezugsobjekt referieren (im Vergleich zu den Literaturangaben) signifikant erhöht.

Anhand der obigen Beispiele wird deutlich, warum die Kandidatenmenge nicht bereits im Rahmen von Teilprozess 1.2 via SPARQL bestimmt werden konnte. Eine SPARQL-Abfrage kam nur für die Vorselektion in Frage. Die Untermenge musste anschließend reduziert werden, indem Zeichenkettenähnlichkeitsfunktionen auf den Namen der Personen angewendet wurden. Da zur Vorselektion keine komplexe SPARQL-Abfrage formuliert werden musste bzw. konnte, wurde dies über die Drop-Down-Funktion vorgenommen. Wie bereits erwähnt, kann diese Funktion verwendet werden, um den Zieldatenbestand auf Instanzen bestimmter Klassen einzuschränken. Aufgrund der Überprüfung der Personentabelle im Rahmen des ETL-Prozesses und wegen der durchgeführten Datenerfassung waren sämtliche Typen von Agenten bekannt. Deshalb konnten bestimmte Arten von Ressourcen in der GND (z.B. Geister, Götter, musikalische Körperschaften etc.) ausgeschlossen werden, da diese im historischen Häuserbuch nicht beschrieben werden. Der Zieldatenbestand konnte von vornherein auf Instanzen der folgenden Klassen bzw. auf Instanzen der relevanten Unterklassen eingeschränkt werden:

- `gndo:CorporateBody`
 - `gndo:ReligiousCorporateBody`
 - `gndo:Company`
 - `gndo:OrganOfCorporateBody`
 - `gndo:ReligiousAdministrativeUnit`
- `gndo:Person`
 - `gndo:DifferentiatedPerson`
 - `gndo:RoyalOrMemberOfARoyalHouse`

Die Untermenge wurde sukzessive eingegrenzt (Teilprozess 1.3). Wenn bestimmte Operatoren oder Transformatoren nicht zu einer geeigneten Ausgabe führten, wurde die Workbench angepasst und erneut ausgeführt. Bevor die Umsetzung detailliert beschrieben wird, müssen die Literale thematisiert werden, welche für den Namensvergleich in Frage kamen. Für die weitere Erörterung wird zwischen juristischen und natürlichen Personen differenziert¹⁴⁰.

¹⁴⁰ Im Sinne einer Unterscheidung zwischen Menschen und Organisationen, völlig unabhängig davon, ob dies im juristischen Sinn oder aus historischer Sicht korrekt ist oder nicht. Hier ist diese Unterscheidung zielführend.

Aufgrund dieser Unterscheidung können die jeweiligen Charakteristika der Menschen oder der Organisationen und deren Auswirkungen besser hervorgehoben werden. Deshalb werden die Workbench, die resultierende Ausgabe und der Evaluierungsprozess für die Menschen und die Organisationen nacheinander behandelt.

Im konkreten Fall kommen zwei Lösungsmöglichkeiten in Frage, um die Verlinkung der Menschen durch einen Namensvergleich zu realisieren:

1. Die Ähnlichkeitsprüfung berücksichtigt alle möglichen Namensvarianten¹⁴¹ und Namensschreibweisen, die vorhanden sind.
2. Die Ähnlichkeitsprüfung wird mittels des bevorzugten Namens¹⁴² durchgeführt.

Die Umsetzung der ersten Option führt zu einer sehr großen Menge möglicher Ressourcen. Der größte Teil ist irrelevant. Dies liegt daran, dass es viele Namensvarianten für eine Person¹⁴³ geben kann und zwar hinsichtlich Vor-, Nachname, diverser Namenszusätze etc. Die manuelle Überprüfung gestaltet sich deshalb sehr aufwendig. Die Ermittlung der optimalen Konfiguration hätte auch zu viel Zeit in Anspruch genommen. Dies ist auf die Datenmenge, die gewählte Methode (Versuch und Irrtum) und auf die „Hardware“ (s. Tab. 8.1 – VM2) zurückzuführen.

Die zweite Lösung hat den grundsätzlichen Nachteil, dass zu viele Kandidaten nicht gefunden werden. Es existieren aber Ähnlichkeitsfunktionen, um unterschiedliche Namensvarianten entsprechend berücksichtigen zu können. Deshalb wurde die zweite Option umgesetzt. Hierfür wurde die folgende Workbench zwecks Verlinkung der Menschen konfiguriert:

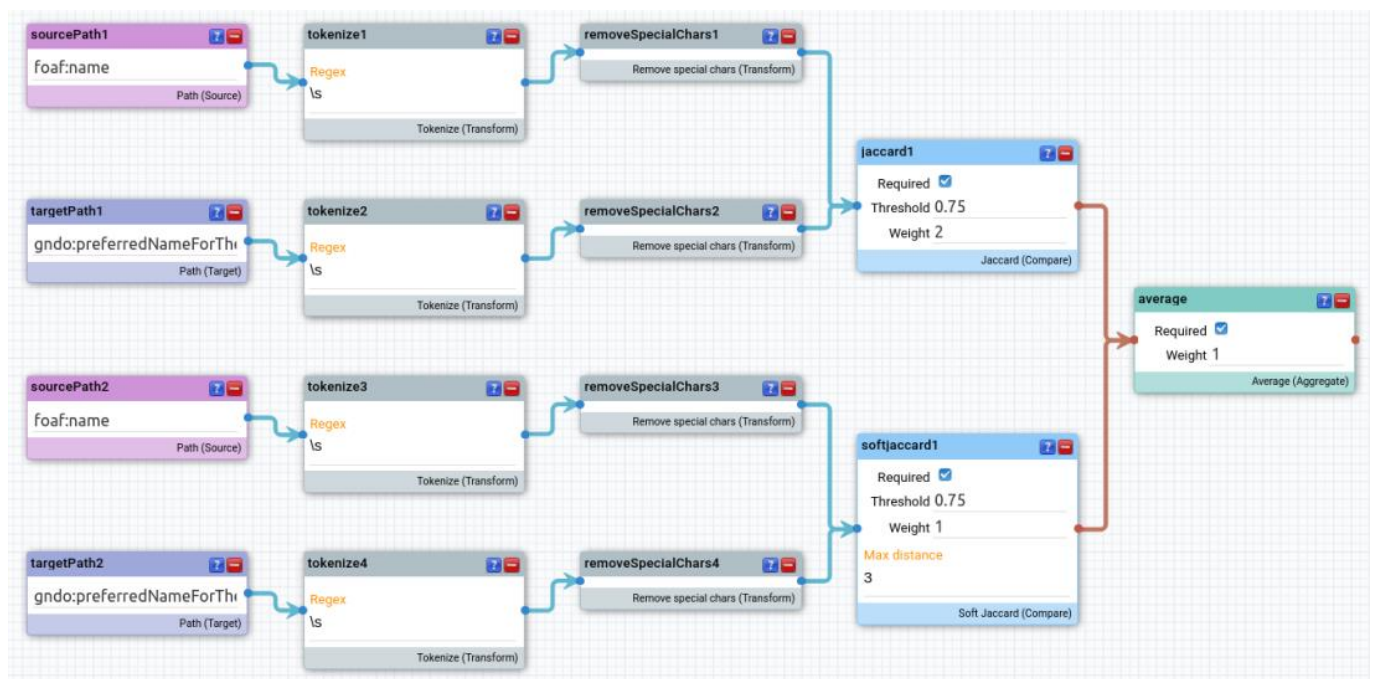


Abb. 8.20: Workbench für die Verlinkung der Menschen, EK

¹⁴¹ Siehe: gndo:preferredNameForThePerson

¹⁴² Siehe: gndo:variantNameForThePerson

¹⁴³ Beispiel: <https://d-nb.info/gnd/118543768>

Es ist erkennbar, dass die Zeichenketten verglichen wurden, welche die Prädikate `gndo:preferredName-ForThePerson` (Ziel) und `foaf:name` (Quelle) referenzieren. Dabei durften spezielle Zeichen und die Reihenfolge der Namensbestandteile, -zusätze, etc. keinen Einfluss nehmen. Deshalb wurden die Transformer (`tokenize1-4`, `removeSpecialChars1-4`) eingesetzt, welche bereits erörtert wurden. Für den Vergleich kamen Token-basierte Metriken (`jaccard1` und `softjaccard1`) zum Einsatz.

Die Berechnung der Jaccard Ähnlichkeit bot sich in diesem Fall an. Kurek (2011) führte dazu aus, dass die Wortreihenfolge bei diesem Vergleich keinen Einfluss auf den Wert der berechneten Ähnlichkeit hat. Für die Verlinkung der Menschen, war dies eine Grundvoraussetzung, da die Reihenfolge der Namensbestandteile, -zusätze etc. (im historischen Häuserbuch und in der GND) variiert. Kurek (2011) erklärte außerdem, dass schon kleinste Variationen in der Schreibweise einzelner Wörter die Ähnlichkeit stark verringern. Dies war zielführend, da der bevorzugte Name der Person restriktiv verwendet werden musste. Dadurch konnte die Kandidatenmenge zunächst begrenzt werden.

Andererseits sollten ähnliche Namensvarianten ebenfalls berücksichtigt werden. Die Kandidatenmenge musste wieder vergrößert werden. Aus diesem Grund wurde zusätzlich die Soft Jaccard Ähnlichkeit berechnet. Die Implementierung des vorgenannten Ähnlichkeitsalgorithmus verlangt die Festlegung einer maximalen Distanz (Max distance) bzw. die Angabe der Levenshtein Distanz. Die Distanz zwischen zwei Zeichenketten wird laut Kurek (2011) ermittelt, indem die minimalen Kosten berechnet werden, die bei der Überführung der einen Zeichenkette in die andere Zeichenkette entstehen. Dabei wird auf die zur Überführung nötigen Operationen (Einfügen, Löschen, Ersetzen) abgestellt, denen die entsprechenden Kosten zugeordnet werden (Kurek, 2011). Die jeweiligen Kosten der notwendigen Operationen werden addiert (Kurek, 2011). Die Levenshtein Distanz zwischen zwei Zeichenketten legt die Kosten für die vorgenannten Operationen mit 1 fest und gibt daher die minimale Anzahl der o.g. Operationen an, die erforderlich sind, um die eine Zeichenkette in die andere Zeichenkette zu überführen (Kurek, 2011). Aufgrund der spezifizierten Levenshtein Distanz (s.o.) gibt die Anwendung die Soft Jaccard Ähnlichkeit mit 100 % an, wenn es wahr ist, dass die minimale Anzahl der o.g. Operationen ≤ 3 ist. Die adäquate Levenshtein Distanz (= 3) wurde über die Versuch-und-Irrtum-Methode ermittelt. Die folgenden Beispiele belegen die Geeignetheit des o.g. Werts und zeigen ebenfalls, warum nur die Kombination der Ähnlichkeitsalgorithmen (Jaccard Ähnlichkeit und Soft Jaccard Ähnlichkeit) bzw. die Aggregation der ermittelten Ähnlichkeiten zu einem akzeptablen Ergebnis führte. Die Hintergründe zu den Namensvarianten von „Bernhard von Breidenbach“ können der zugehörigen Beschreibung entnommen werden, indem die URI dereferenziert wird. Alternativ kann der Wikipedia-Artikel¹⁴⁴ verwendet werden. Die vorgenannte Person wird im historischen Häuserbuch für das Jahr 1485 als Bewohner des Hauses „Zum Steckelberg“ (siehe Anhang, A207 und B222) ausgewiesen.

¹⁴⁴ Siehe: https://de.wikipedia.org/wiki/Bernhard_von_Breidenbach



Abb. 8.21: Beispiel für Namensvarianten, EK

Es ist erkennbar, dass spezielle Zeichen und die Wortreihenfolge keinen Einfluss auf den Wert der berechneten Ähnlichkeit hatten. Im Rahmen der Aggregation wurden die beiden berechneten Jaccard Ähnlichkeiten in einem angemessenen Verhältnis (2:1) berücksichtigt. Die Einschränkung auf eine „geeignete“ Untermenge (über Jaccard) hatte die höhere Priorität, da die potenziellen Kandidaten untersucht werden mussten (s.u.). Die berechnete Soft-Jaccard Ähnlichkeit führte im konkreten Fall und in ähnlichen Fällen dazu, dass die Wahrscheinlichkeit auf einen „angemessenen“ Wert erhöht wurde (55,6 %). Hier identifizieren die URIs dieselbe Ressource.

Das folgende Beispiel zeigt, warum die Entscheidung für die o.g. Kombination getroffen wurde. Anderenfalls wäre die Begrenzung der Kandidatenmenge (überhaupt) nicht möglich gewesen, denn die Anwendung würde in diesem Fall und in ähnlichen Fällen eine hundertprozentige Übereinstimmung ausgeben:

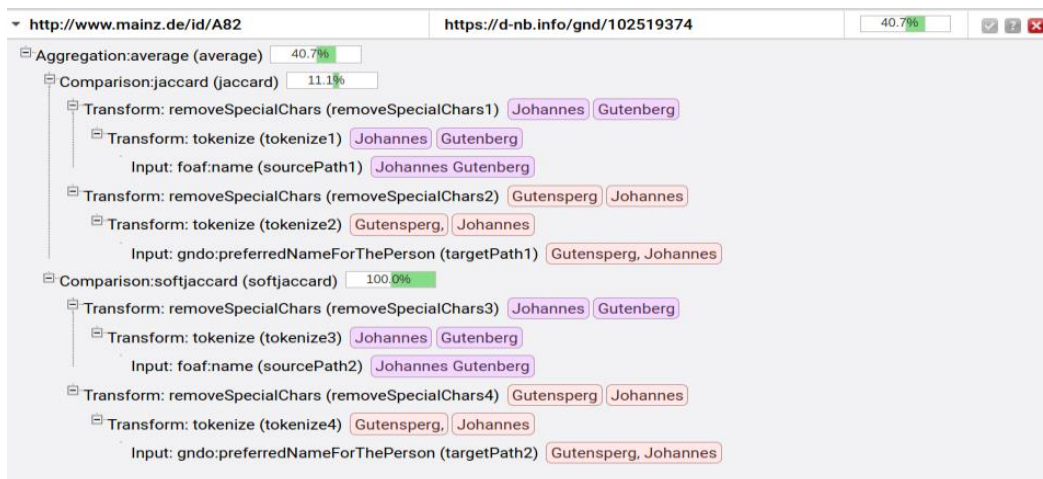


Abb. 8.22: Kombination von Jaccard und Soft-Jaccard, EK

Mithilfe der Person aus dem Schlüsselbeispiel lässt sich die Festlegung des Schwellenwerts (0,75) begründen:



Abb. 8.23: Beispiel für einen fehlenden Namenszusatz, EK

Es ist häufig der Fall, dass Adels-, Berufstitel bzw. akademische Grade und sonstige Namenszusätze, welche im Quelldatenbestand vorhanden sind, nicht als Teilzeichenfolge im Zieldatenbestand vorkommen bzw. im Literal, das von `gndo:preferredNameForThePerson` referenziert wird. Umgekehrt ist dies natürlich auch möglich. Die Jaccard Ähnlichkeit sollte dadurch nicht unverhältnismäßig stark verringert werden. Durch die Verringerung des Schwellenwerts tritt dieser Effekt ein; z.B. reduziert ein Schwellenwert von 0,5 die Ähnlichkeit im obigen Beispiel auf 33,3 %. In diesem Beispiel identifizieren die URIs dieselbe Ressource.

Die vorstehenden Ausführungen enthalten die wesentlichen Gründe, warum der umfangreiche Zieldatenbestand nicht auf die relevante Untermenge eingeschränkt werden konnte. Über die skizzierte Vorselektion via SPARQL und die Ähnlichkeitsprüfung mittels des bevorzugten Namens konnte die Kandidatenmenge lediglich näherungsweise bestimmt werden. Da die Ausgabe (s.u.) auch (viele) Kandidatenpaare enthielt, für die keine Links erzeugt werden durften, schied eine stichprobenartige Überprüfung der potenziellen Kandidatenpaare aus. Der größte Teil der Ressourcen wurde manuell untersucht.

Die relevanten Organisationen haben, im Gegensatz zu den Menschen, häufig einen eindeutigen bevorzugten Namen¹⁴⁵. Dazu kommt, dass die Mehrzahl der gesuchten Organisationen im Quell- und Zieldatenbestand die Teilzeichenfolge „Mainz“ als Bestandteil des bevorzugten Namens aufweisen. Dies gilt jedoch nicht für alle Organisationen im Mainzer Stadtgebiet (z.B. „Kloster Maria Dalheim“) und auch nicht für „auswärtige“ Organisationen, zu denen das historische Häuserbuch ebenfalls Informationen enthält (z.B. „Stift Ravengiersburg“). Auch wenn „Mainz“ einen Namensbestandteil darstellt, erübrigte sich die manuelle Überprüfung der Kandidaten meist nicht.

¹⁴⁵ Siehe: `gndo:preferredNameForTheCorporateBody`

Im Gegensatz zu den Menschen konnte der Link-Typ ohne genauere Untersuchungen der betreffenden Ressourcen häufig nicht festgelegt werden. Die o.g. Teilzeichenfolge war lediglich für die Ermittlung einer oder mehrerer infrage kommender URIs hilfreich. Dafür musste diese Teilzeichenkette nicht über einen regulären Ausdruck extrahiert werden, da eine höhere Gewichtung derselbigen nicht zielführend gewesen wäre (s.u.).

Von der Möglichkeit im Rahmen der Ähnlichkeitsprüfung alle vorhandenen Namensvarianten¹⁴⁶ und Namensschreibweisen zu berücksichtigen, wurde ebenfalls kein Gebrauch gemacht. Die Gründe dafür wurden bereits genannt. Die Zeichenkettenähnlichkeitsfunktionen wurden auf die bevorzugten Namen der Organisationen angewendet. Die folgende Workbench zeigt die Verknüpfungen der Operatoren und der Transformatoren, deren Funktion bereits erklärt wurde.

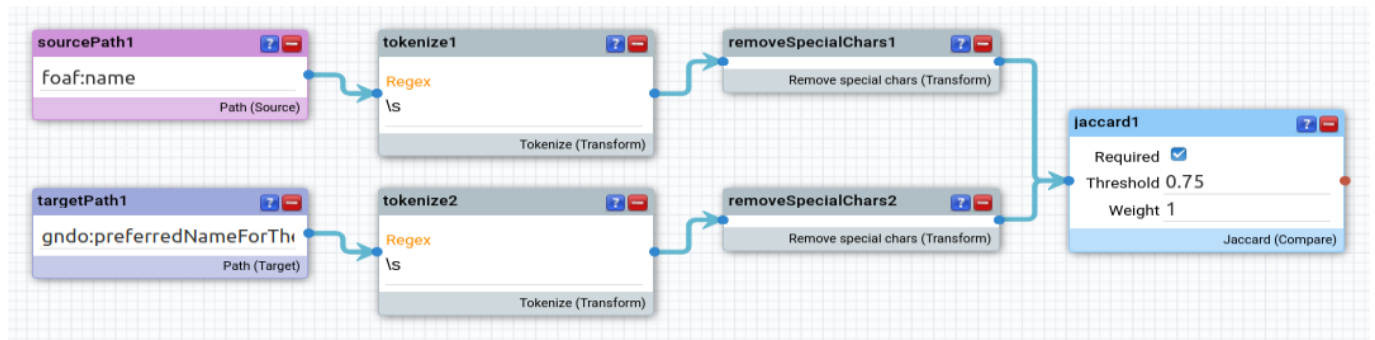


Abb. 8.24: Workbench für die Verlinkung der Organisationen, EK

Für die Verlinkung der Organisationen durften spezielle Zeichen und die Wortreihenfolge keinen Einfluss auf den Wert der berechneten Ähnlichkeit haben. Dieser Aspekt ist anhand der folgenden Beispiele erkennbar. Diese zeigen auch, warum der o.g. Schwellenwert (0,75) für die Organisationen beibehalten wurde. Häufig waren mehrere Tokens identisch. Das differierende Token oder die differierenden Tokens mussten in diesen Fällen den Ausschlag geben. Insofern wäre eine höhere Gewichtung der Teilzeichenkette „Mainz“ im Rahmen der Aggregation kontraproduktiv gewesen.

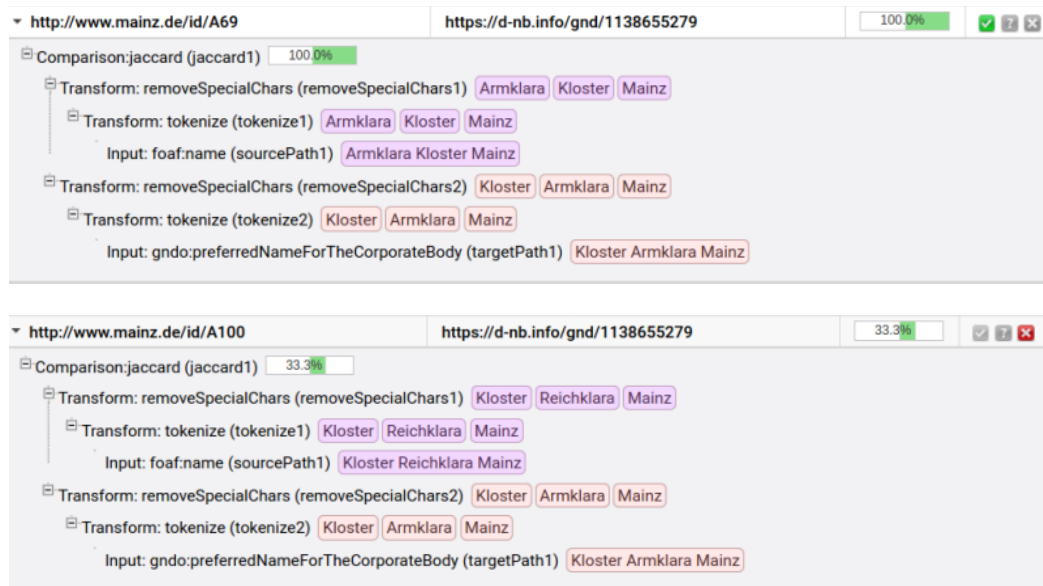


Abb. 8.25: Beispiel für ähnliche Namen von Organisationen, EK

¹⁴⁶ Siehe: gndo:variantNameForTheCorporateBody

Die Evaluation der potenziellen Kandidatenpaare wurde im Rahmen des zweiten Prozesses durchgeführt (Abb. 7.7). Zunächst wurde die tabellarische Ausgabe verwendet, welche die ausgewählten Zeichenketten listet (Teilprozess 2.1). Im Gegensatz zu den Literaturangaben, war die Festlegung des Link-Typs (Prozess 3) und die anschließende Generierung des Links selten direkt möglich (Prozess 4). Nur in wenigen Fällen konnte der Quellgraph bereits aktualisiert bzw. der Link eingefügt werden, ohne dass weitere Informationen geprüft werden mussten.

Die folgenden Abbildungen zeigen die Ausgaben (nach Wahrscheinlichkeit absteigend sortiert) für die Menschen und die Organisationen sowie ausgewählte Details. Anhand dieser Beispiele wird nachfolgend erläutert, wieso häufig die Entscheidung für genauere Untersuchungen getroffen wurde (Teilprozess 2.2 und 2.3). Auf Basis der ausgewählten Zeichenketten konnte meist noch nicht beurteilt werden, auf welche Bezugsobjekte die ermittelten URIs tatsächlich referieren, geschweige denn, ob die Verlinkung sinnvoll ist. Dafür mussten die zugehörigen Beschreibungen überprüft werden.

▶ http://www.mainz.de/id/A113	https://d-nb.info/gnd/137804393	100.0%	✓	⊖	⊕	✕
▶ http://www.mainz.de/id/A314	https://d-nb.info/gnd/1070494380	100.0%	✓	⊖	⊕	✕
▶ http://www.mainz.de/id/A64	https://d-nb.info/gnd/118656724	100.0%	✓	⊖	⊕	✕
▶ http://www.mainz.de/id/A302	https://d-nb.info/gnd/13573441X	100.0%	⊖	⊕	✕	✕
▶ http://www.mainz.de/id/A302	https://d-nb.info/gnd/135734452	100.0%	✓	⊖	⊕	✕
▶ http://www.mainz.de/id/A137	https://d-nb.info/gnd/124209807	100.0%	✓	⊖	⊕	✕
▶ http://www.mainz.de/id/A82	https://d-nb.info/gnd/118543768	100.0%	✓	⊖	⊕	✕
▶ http://www.mainz.de/id/A489	https://d-nb.info/gnd/1202574203	100.0%	✓	⊖	⊕	✕
▼ http://www.mainz.de/id/A331	https://d-nb.info/gnd/118052632	100.0%	✓	⊖	⊕	✕

Aggregation:average (average) 100.0%

Comparison:jaccard (jaccard1) 100.0%

Transform: removeSpecialChars (removeSpecialChars1) Johann von Nassau

Transform: tokenize (tokenize1) Johann, von Nassau

Input: foaf:name (sourcePath1) Johann, von Nassau

Transform: removeSpecialChars (removeSpecialChars2) Johann von Nassau

Transform: tokenize (tokenize2) Johann, von Nassau

Input: gndo:preferredNameForThePerson (targetPath1) Johann, von Nassau

Comparison:softjaccard (softjaccard1) 100.0%

Transform: removeSpecialChars (removeSpecialChars3) Johann von Nassau

Transform: tokenize (tokenize3) Johann, von Nassau

Input: foaf:name (sourcePath2) Johann, von Nassau

Transform: removeSpecialChars (removeSpecialChars4) Johann von Nassau

Transform: tokenize (tokenize4) Johann, von Nassau

Input: gndo:preferredNameForThePerson (targetPath2) Johann, von Nassau

▶ http://www.mainz.de/id/A331	https://d-nb.info/gnd/13630236X	100.0%	⊖	⊕	✕
---	---	--------	---	---	---

Aggregation:average (average) 100.0%

Comparison:jaccard (jaccard1) 100.0%

Transform: removeSpecialChars (removeSpecialChars1) Johann von Nassau

Transform: tokenize (tokenize1) Johann, von Nassau

Input: foaf:name (sourcePath1) Johann, von Nassau

Transform: removeSpecialChars (removeSpecialChars2) Johann von Nassau

Transform: tokenize (tokenize2) Johann, von Nassau

Input: gndo:preferredNameForThePerson (targetPath1) Johann, von Nassau

Comparison:softjaccard (softjaccard1) 100.0%

Transform: removeSpecialChars (removeSpecialChars3) Johann von Nassau

Transform: tokenize (tokenize3) Johann, von Nassau

Input: foaf:name (sourcePath2) Johann, von Nassau

Transform: removeSpecialChars (removeSpecialChars4) Johann von Nassau

Transform: tokenize (tokenize4) Johann, von Nassau

Input: gndo:preferredNameForThePerson (targetPath2) Johann, von Nassau

▶ http://www.mainz.de/id/A444	https://d-nb.info/gnd/137950500	100.0%	✓	⊖	⊕	✕
---	---	--------	---	---	---	---

Abb. 8.26: Kandidatenpaare für die Verlinkung der Menschen, mit Details zu A331, EK

▶ http://www.mainz.de/id/A51	https://d-nb.info/gnd/4627784-5	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A9	https://d-nb.info/gnd/5222085-0	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A37	https://d-nb.info/gnd/10075260-3	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A67	https://d-nb.info/gnd/1136480633	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A20	https://d-nb.info/gnd/2127653-5	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A177	https://d-nb.info/gnd/3072620-7	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A13	https://d-nb.info/gnd/4319017-0	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A46	https://d-nb.info/gnd/3059710-9	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A16	https://d-nb.info/gnd/3053046-5	100.0%	✓ ? ✕
▼ http://www.mainz.de/id/A281	https://d-nb.info/gnd/80092-2	100.0%	✓ ? ✕
<div style="border: 1px solid #ccc; padding: 5px;"> Comparison:jaccard (jaccard1) 100.0% <ul style="list-style-type: none"> Transform: removeSpecialChars (removeSpecialChars1) Jesuiten Transform: tokenize (tokenize1) Jesuiten <ul style="list-style-type: none"> Input: foaf:name (sourcePath1) Jesuiten Transform: removeSpecialChars (removeSpecialChars2) Jesuiten <ul style="list-style-type: none"> Transform: tokenize (tokenize2) Jesuiten Input: gndo:preferredNameForTheCorporateBody (targetPath1) Jesuiten </div>			
▶ http://www.mainz.de/id/A140	https://d-nb.info/gnd/121352718X	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A143	https://d-nb.info/gnd/3072603-7	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A123	https://d-nb.info/gnd/1213254728	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A17	https://d-nb.info/gnd/1043606785	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A100	https://d-nb.info/gnd/1138655031	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A28	https://d-nb.info/gnd/3072591-4	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A32	https://d-nb.info/gnd/4829887-6	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A367	https://d-nb.info/gnd/2041999-5	100.0%	✓ ? ✕
▶ http://www.mainz.de/id/A607	https://d-nb.info/gnd/4322161-0	100.0%	✓ ? ✕
▼ http://www.mainz.de/id/A281	https://d-nb.info/gnd/1189237067	11.1%	✓ ? ✕
<div style="border: 1px solid #ccc; padding: 5px;"> Comparison:jaccard (jaccard1) 11.1% <ul style="list-style-type: none"> Transform: removeSpecialChars (removeSpecialChars1) Jesuiten Transform: tokenize (tokenize1) Jesuiten <ul style="list-style-type: none"> Input: foaf:name (sourcePath1) Jesuiten Transform: removeSpecialChars (removeSpecialChars2) Jesuiten Rheinische Provinz <ul style="list-style-type: none"> Transform: tokenize (tokenize2) Jesuiten Rheinische Provinz Input: gndo:preferredNameForTheCorporateBody (targetPath1) Jesuiten Rheinische Provinz </div>			

Abb. 8.27: Kandidatenpaare für die Verlinkung der Organisationen, mit Details zu A281, EK

Oben sind Fälle abgebildet, in denen weitere Informationen analysiert werden mussten. Anhand des Menschen namens „Johann von Nassau“ kann das Vorgehen exemplarisch für einen „owl:sameAs-Link“ und mittels der Organisation namens „Jesuiten“ für einen „rdfs:seeAlso-Link“ dargelegt werden.

Unterschiedliche Menschen hatten mitunter den gleichen Namen (z.B. „Johann von Nassau“). Wenn also eine hundertprozentige Übereinstimmung ausgegeben wurde, erübrigte sich die manuelle Überprüfung nicht. Bezüglich der Menschen musste bzw. konnte jedoch nicht eruiert werden, welches Prädikat eingesetzt werden soll, sondern es musste nur festgestellt werden, ob die URIs in der GND tatsächlich die gesuchten Personen identifizieren. Falls dies zutraf, wurden die Menschen ausschließlich über das Prädikat **owl:sameAs** verlinkt.

Die zugehörigen Beschreibungen der DNB waren für die Beantwortung der Frage, ob es sich um „denselben“ oder einen (ganz) anderen Menschen handelt i.d.R. ausreichend (Teilprozess 2.2). Der „Zustand“ / die „Ähnlichkeit“ eines Menschen konnte anhand der zugehörigen Beschreibungen nicht beurteilt werden (Alter o. ä.).

Im ersten Schritt wurde anhand jener Kriterien manuell geprüft, die bereits erwähnt wurden. Dazu zählten z.B. der Wirkungszeitpunkt und der Wirkungszeitraum. Es wurde untersucht, ob die Zeitangaben in der GND mit den Zeiträumen im historischen Häuserbuch übereinstimmen bzw. ob es für etwaige Abweichungen plausible Gründe gibt. Zu diesen Kriterien zählte auch der geografische Bezug. Für beide Merkmale wurden bereits einige Prädikate aufgezählt, die im Zieldatenbestand eingesetzt werden.

Wenn im ersten Schritt noch keine Entscheidung getroffen werden konnte, ob es sich um die gesuchte Ressource handelte, wurden anschließend die Objekte geprüft, welche das jeweilige Individuum näher charakterisieren. Diese werden in der GND u. a. von den folgenden Prädikaten¹⁴⁷ referenziert:

- `gndo:professionOrOccupation`
- `gndo:biographicalOrHistoricalInformation`
- `gndo:definition`
- `gndo:gndSubjectCategory`
- `gndo:publication`
- ...

Im letzten Schritt dieses Teilprozesses wurden die Beziehungen der Ressourcen mithilfe der Objekte geprüft, die z.B. über die folgenden Prädikate¹⁴⁸ referenziert werden:

- `agrelon:hasSibling`
- `gndo:founder`
- `gndo:familialRelationship`
- `gndo:acquaintanceshipOrFriendship`
- `gndo:hierarchicalSuperiorOfTheCorporateBody`
- ...

Anhand des obigen Beispiels kann die Notwendigkeit für diese Prüfung dargelegt werden. So identifizieren die beiden dargestellten URIs in der GND¹⁴⁹ nicht nur Menschen, die den gleichen Namen hatten („Johann von Nassau“), sondern auch denselben Beruf (Bischof). Beide agierten u. a. in Deutschland (XA-DE). Die Lebensdaten liegen recht nah beieinander und sind teilweise unscharf angegeben. Die Person, die von letztgenannter URI identifiziert wird, agierte jedoch auch in den Niederlanden (XA-NL). Dies ist ein Hinweis, dass diese Ressource im konkreten Fall nicht als Linkziel verwendet werden sollte.

In einigen Fällen mussten noch Links verfolgt werden, d.h., es wurden URIs dereferenziert, welche die DNB ihren Informationen hinzugefügt hat (z.B. „owl:sameAs-Links“ zu Wikidata). Im Rahmen von Teilprozess 2.3 wurden nur noch HTML-Repräsentationen untersucht. Weitere Recherchen wurden nicht durchgeführt, d.h., es wurde die Entscheidung getroffen, die Ressource von der Verlinkung auszuschließen, wenn immer noch keine Festlegung möglich war. In diesen Fällen wurden also keine Links in den Quellgraphen eingefügt.

Hinsichtlich der Organisationen war es zwar einfacher, die potenziellen Kandidaten im Zieldatenbestand zu ermitteln (s.o.), dafür war es im Vergleich zu den Menschen schwieriger den Link-Typ festzulegen. Viele Organisationen konnten nicht über das Prädikat `owl:sameAs` verlinkt werden. Häufig referieren die URIs im Häuserbuch und die URIs in der GND zwar auf ähnliche aber nicht auf dieselben Bezugsobjekte. Im Häuserbuch werden historische Organisationen beschrieben. Die Beschreibungen beziehen sich mitunter konkret auf

¹⁴⁷ Die Bedeutung der Prädikate kann dem Semantic Web Dokument entnommen werden.

¹⁴⁸ Das Präfix „agrelon“ steht für <https://d-nb.info/standards/elementset/agrelon#>.

¹⁴⁹ <https://d-nb.info/gnd/118052632> und <https://d-nb.info/gnd/13630236X>

bestimmte Zeitabschnitte, Entwicklungsstadien etc. von Organisationen oder Teilen davon, die ggf. (in dieser Form) heute nicht mehr existieren.

Die oben skizzierte Untersuchung mithilfe der zugehörigen Beschreibungen der DNB musste auch für die Organisationen durchgeführt werden. Es war zu klären, ob die URI in der GND wirklich die gesuchte historische Organisation identifiziert. Der Anhang enthält einige repräsentative Beispiele, wie die (alte) Universität Mainz, für die in der GND die Zeitspanne 1477-1823 genannt wird oder das Kloster St. Michael Mainz, das laut GND im Jahr 1781 aufgehoben wurde. In diesen und ähnlichen Fällen ist die Entscheidung für den Einsatz des Prädikats **owl:sameAs** noch vertretbar. Die obige Ausgabe zeigt mit den „Jesuiten“ bzw. mit der zusätzlich hinzugefügten „Rheinischen Provinz“ ein Gegenbeispiel. Die entsprechenden URIs können dereferenziert werden. Dies verdeutlicht, dass ohne eingehende Recherchen nicht eindeutig festgelegt werden kann, ob der Link bei Verwendung des Prädikats **owl:sameAs** wirklich zur „Rheinischen Provinz“ zeigen soll (u. a. wegen der „Niederdeutschen Provinz“). Es wurde die Entscheidung getroffen, in diesen Fällen die Verlinkung (vorerst – s. GP 10.1) über das Prädikat **rdfs:seeAlso** zu realisieren. Es wurde dabei immer die Organisation (oder der Teil der Organisation) in der GND ausgewählt, die hierarchisch gesehen, d.h. bezogen auf die Organisationsstruktur, auf der höchsten Ebene angesiedelt ist. Von dort aus können ggf. weitere Links verfolgt werden, um spezifischere Informationen zu erhalten. Dieses Vorgehen wurde auch umgesetzt, wenn in der GND keine historischen Informationen zur Organisation vorhanden sind, aber dafür Links zu aktuellen Websites der Organisationen (s. RDF/XML im Anhang).

Ergebnisse

Die folgenden Ergebnisse konnten erzielt werden:

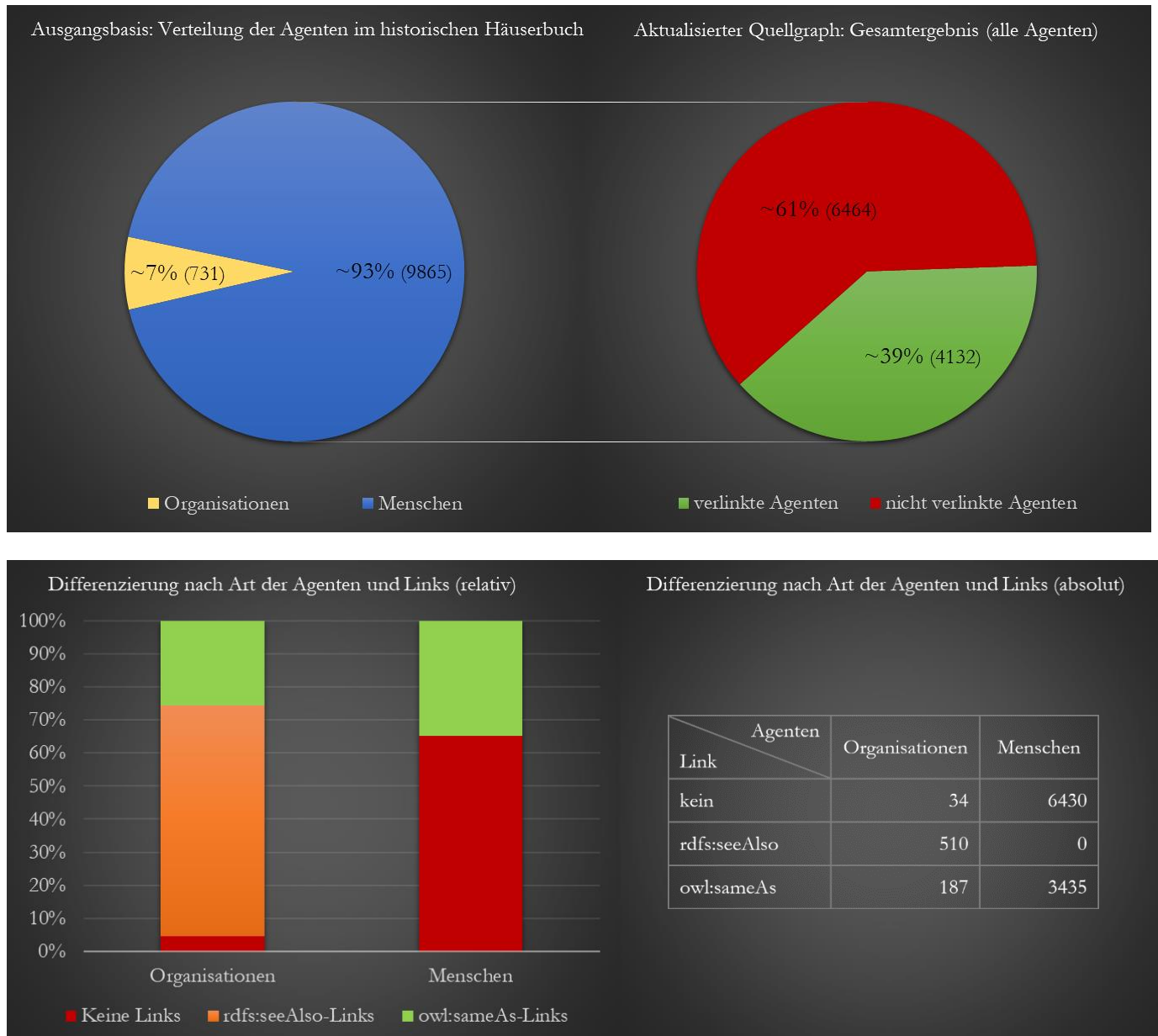


Abb. 8.28: Darstellung der Ausgangsbasis und der Ergebnisse (Verlinkung der Agenten), ED

Geoobjekte

Der Verlinkungsprozess wurde unter dem vorherigen GP ausführlich anhand der vorgestellten Verfahrensweise erläutert. Darauf wird nun verzichtet, da nur eine sehr geringe Zahl von Geoobjekten verlinkt werden konnte. Das immanente Merkmal aller Geoobjekte wurde zwar zur Ermittlung der Kandidatenmenge verwendet. Allerdings wurde festgestellt, dass dafür keine Koordinaten verwendet werden können, die den Raumbezug direkt herstellen und die bei vollständiger Angabe (z.B. CRS), im Gegensatz zu den o.g. Zeichenkettenähnlichkeitsfunktionen, eindeutige Ergebnisse liefern könnten. Im Zieldatenbestand gibt es zu wenig (relevante) Koordinaten. Dies wird anhand der folgenden SPARQL-Abfrage belegt. Diese erfüllt(e) zwei Funktionen:

- (1) sie wird nachfolgend für die Erläuterung herangezogen
- (2) sie wurde für die Selektion der potenziellen Kandidaten verwendet

```

PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gndo: <https://d-nb.info/standards/elementset/gnd#>

SELECT ?PNAME (COUNT (?VNAME) AS ?cVNAME)
WHERE
{ GRAPH <http://localhost:8890/GND>
  {
    ?FEATURE gndo:place <https://d-nb.info/gnd/4037124-4> ;
      gndo:preferredNameForThePlaceOrGeographicName ?PNAME .
    OPTIONAL { ?FEATURE gndo:variantNameForThePlaceOrGeographicName ?VNAME . }
    {
      SELECT DISTINCT ?FEATURE ?PNAME ?WKT ?GEOM ?EST ?TERM ?PROD
      WHERE
      { GRAPH <http://localhost:8890/GND>
        {
          ?FEATURE gndo:place <https://d-nb.info/gnd/4037124-4> ;
            rdf:type/rdfs:subClassOf gndo:PlaceOrGeographicName ;
            gndo:preferredNameForThePlaceOrGeographicName ?PNAME .
          OPTIONAL { ?FEATURE rdf:type ?Type .
            FILTER (?Type IN (gndo:BuildingOrMemorial,
              gndo:NaturalGeographicUnit,
              gndo:NameOfSmallGeographicUnitLyingWithinAnotherGeographicUnit,
              gndo:WayBorderOrLine)
            )
          }
          OPTIONAL { { ?FEATURE gndo:dateOfEstablishment ?EST . }
            UNION { ?FEATURE gndo:dateOfProduction ?PROD . }
          }
          OPTIONAL { ?FEATURE gndo:dateOfTermination ?TERM . }
          OPTIONAL { ?FEATURE geo:hasGeometry [
            geo:asWKT ?WKT ;
            rdf:type ?GEOM ] .
          }
        }
      }
    }
  }
}
GROUP BY ?PNAME ORDER BY ?PNAME

```

Lis. 8.8: SPARQL-Abfrage zwecks Selektion und Erläuterung der Geoobjekte, EK

Vorab wird explizit klargestellt, dass die folgenden Erläuterungen ausschließlich für Ressourcen in der GND gelten, die einen (Raum-)Bezug zur **Stadt Mainz** aufweisen:

- (1) Die Variablen `?WKT`, `?GEOM`, `?EST`, `?TERM`, `?PROD` nach den Schlüsselwörtern `SELECT DISTINCT` und alle optionalen Blöcke in der Unterabfrage belegen u. a., warum für die Verlinkung weder primäre Metriken (direkter Raumbezug) noch Daten (Zeitangaben) verwendet werden konnten. Wenn lediglich die Unterabfrage mit einem geeigneten¹⁵⁰ Tripel- / Quad-Store gegen den Zieldatenbestand (GND) gestellt wird, zeigt die resultierende Tabelle, dass lediglich zwei Geoobjekte über assoziierte Koordinaten verfügen, die mithilfe von Konzepten aus dem Geo-Vokabular des GeoSPARQL-Standards beschrieben wurden. Abgesehen davon, dass die Prädikate in der ABox – `gndo:dateOfEstablishment` (Optionalität¹⁵¹: `[0..1]`, unabhängig vom folgenden), `gndo:dateOfTermination`, (Optionalität: `[0..1]`, nur wenn das vorherige existiert), `gndo:dateOfProduction` (Optionalität: `[0..1]`, nur wenn die beiden anderen nicht existieren) – welche Objekte mit Daten referenzieren, häufig nicht vorhanden sind, gelten die obigen Ausführungen zur Datenstruktur entsprechend. Diese Zeichenketten sind für die systematische Selektion ungeeignet (vgl. Resultat der [Unter-]Abfrage). Dies gilt auch für die übrigen Objekte mit Daten, für die in der obigen SPARQL-Abfrage der Übersichtlichkeit halber keine Variablen definiert wurden; z.B. für jene, die von den Prädikaten `gndo:dateOfEstablishmentAndTermination` und `gndo:biographicalOrHistoricalInformation` referenziert werden. Die Aggregation (`COUNT`) in der äußeren Abfrage belegt die Optionalität des Objekts (`[0..*]`), das in der ABox mittels des Prädikats `gndo:variantNameForThePlaceOrGeographicName` referenziert wird. Die alternativen Namen wurden deshalb nicht berücksichtigt. Das `FILTER`-Element listet der Verständlichkeit halber die relevanten Unterklassen der einzigen infrage kommenden „Superklasse“¹⁵² (`gndo:PlaceOrGeographicName`) explizit. Diese Bedingung ist im konkreten Fall allerdings implizit im Graphen-Muster `?FEATURE gndo:place <https://d-nb.info/gnd/4037124-4>` enthalten (s.u.).
- (2) Die obige SPARQL-Abfrage wurde zur Selektion der relevanten Linkziel-Ressourcen verwendet. Die vorgenannten Variablen und die optionalen Blöcke in der inneren Abfrage – denen hier lediglich eine erläuternde Funktion zukommt – waren dafür nicht erforderlich. Das Graphen-Muster `?FEATURE gndo:place <https://d-nb.info/gnd/4037124-4>` dient der Selektion der potenziellen Kandidaten über eine räumliche Bedingung. Die vorstehende URI (<ohne Präfix>) identifiziert das Geografikum mit dem bevorzugten Namen „Mainz“ in der GND, das wegen der weiteren Informationen und den zahlreichen „owl:sameAs-Links“ als Stadt Mainz bezeichnet werden kann. Der Raumbezug wird also indirekt hergestellt. Diese räumliche Information ist etwas ungenauer, als die beiden Koordinatenangaben. Wenn das o.g. Graphen-Muster in der ABox vorhanden ist, wird das jeweilige Subjekt (`?FEATURE`) als Instanz einer/mehrerer Unterklasse(n) ausgewiesen, die im obigen `FILTER`-Element gelistet ist/sind. Die übrigen Unterklassen von `gndo:PlaceOrGeographicName` sind hier irrelevant, da diese Entitätstypen (z.B. Gebietskörperschaften) im historischen Häuserbuch entweder nicht spezifiziert sind oder entsprechende Entitäten (konkrete Ausprägungen, z.B. Wanderwege) nicht beschrieben werden. Der Vollständigkeit halber ist erwähnenswert, dass durch die Selektion über die sekundäre Metrik vier Kandidaten nicht gefunden werden. Ob das Prädikat `gndo:complexSeeReferenceSubject`, das bei den betreffenden Subjekten stets mehrfach existiert, ein Hinweis darauf ist, dass diese Ressourcen das Resultat einer Kombination aus anderen Ressourcen sind, mag dahinstehen, denn die vier betreffenden URIs¹⁵³ identifizieren dasselbe Geoobjekt (Mainzer Gutenberg-Denkmal von 1837),

¹⁵⁰ Dieser sollte Aggregationen, Property Paths etc. unterstützen, wie in der SPARQL-Spezifikation 1.1 festgelegt. In Anbetracht des zu erwartenden Ergebnisses empfiehlt sich die adäquate Konfiguration; z.B. wäre beim Einsatz von GraphDB RDFS-Plus als Ruleset ausreichend und die nicht IR-basierte Implementierung von owl:sameAs zielführend.

¹⁵¹ Vgl. Erklärungen zu Tab. 8.6.

¹⁵² Der Korrektheit halber wird darauf hingewiesen, dass dies nicht die höchste definierte Klasse in der Hierarchie ist.

¹⁵³ <https://d-nb.info/gnd/7513902-9>
<https://d-nb.info/gnd/7513905-4>

das im Modell für die Zeit um 1871 nicht beschrieben wird. Dementsprechend können diese Ressourcen auch nicht als Linkziel verwendet werden. Dies gilt ebenso für Ressourcen, die laut GND in der **Region** Mainz liegen. Im konkreten Fall werden diese Ressourcen wegen des o.g. Graphen-Musters nicht selektiert.

Das war allerdings beabsichtigt, da diese Geografika (z.B.: <https://d-nb.info/gnd/4449064-1>) viel zu weit vom ehemaligen Stadtzentrum entfernt liegen, weshalb das historische Häuserbuch an diesen Stellen keine Geoobjekte verzeichnet.

Mit der obigen SPARQL-Abfrage konnte der Zieldatenbestand auf die relevanten Kandidaten eingeschränkt werden (75 Geoobjekte in der GND). Im Zieldatenbestand¹⁵⁴ gibt es – von den irrelevanten Ausnahmen abgesehen – keine weiteren Ressourcen, welche für die Verknüpfung mit den Geoobjekten aus dem historischen Häuserbuch (grundsätzlich) in Frage kommen. Um dies zu verifizieren, wurden weitere SPARQL-Abfragen gestellt¹⁵⁵, die stets zum selben Ergebnis führten. In der Workbench wurden die Zeichenketten transformiert und in Tokens zerlegt. Es wurden keine Vergleichsoperatoren kombiniert bzw. Ergebnisse aggregiert. Als Eingabe wurden nur die beiden Objekte (Name im historischen Häuserbuch, bevorzugter Name in der GND) definiert, die über die Prädikate `gndo:preferredNameForThePlaceOrGeographicName` (GND) und `foaf:name` (Häuserbuch) referenziert werden. Die äußere Abfrage in Lis. 8.8 listet deshalb die bevorzugten Namen (?PNAME). Die Ausführungen zu den bevorzugten Namen der Organisationen gelten entsprechend (Festlegung des Link-Typs, Vorkommen der Teilzeichenkette „Mainz“ im bevorzugten Namen etc.). Die Ähnlichkeit der Namen wurde mittels des Soft Jaccard Algorithmus überprüft. Die Links konnten nach der obligatorischen Überprüfung generiert werden. Die o.g. Zeichenkettenähnlichkeitsfunktion bzw. der gesamte Prozess wurde bereits detailliert erörtert und anschaulich dargestellt.

Einige RDF-Fragmente, welche die verlinkten Geoobjekte beschreiben, konnten dem Anhang als RDF/XML-Serialisierung hinzugefügt werden. Da die vollständige Beigabe des RDF-Graphen aus Platzgründen (Koordinatenangaben im Häuserbuch) nicht möglich ist, wurde der aktualisierte Quellgraph via SPARQL abgefragt. Die resultierende Tabelle wurde dem Anhang zusätzlich hinzugefügt. Diese enthält alle selektierten Ressourcen, deren (bevorzugten) Namen und die (indirekt) dereferenzierbare URI aus der GND sowie einige der korrespondierenden Informationen aus dem historischen Häuserbuch (Name, URI, Zeit). Der Link-Typ wurde ebenfalls abgefragt und der „Match“ wurde zusätzlich markiert. Die Ressourcen und deren Verknüpfungen wurden erläutert. Von den 75 potenziellen Kandidaten in der GND wurden 30 tatsächlich als Linkziele verwendet, die vollständig im Anhang gelistet sind. 45 Ressourcen, die dafür grundsätzlich in Frage kommen, wurden nicht verwendet. Der Anhang listet die Gründe hierfür und zwar für jede einzelne Ressource, da diese im Detail überprüft wurden. Von den 7802 Geoobjekten des historischen Häuserbuchs, die im Rahmen des ETL-Prozesses umgeformt wurden, konnten 53 mit Ressourcen in der GND verknüpft werden. Es wurde **weniger als 1%** der transformierten Geoobjekte verlinkt. Ein höherer Zielerreichungsgrad ist – unter Berücksichtigung der Zieldefinition – unmöglich. Dafür wären zusätzliche Ressourcen aus der LOD-Cloud erforderlich gewesen.

<https://d-nb.info/gnd/7513904-2>

<https://d-nb.info/gnd/7513903-0>

¹⁵⁴ In Tab. 8.11 sind alle relevanten Details zu dieser Version aufgeführt.

¹⁵⁵ Es wurden z.B. alle Geografika in Rheinland-Pfalz ausgewählt, die dann über die Teilzeichenkette „Mainz“ in Kombination mit der Einschränkung auf die Instanzen der relevanten Unterklassen weiter eingegrenzt wurden. Es wurde auch überprüft, ob Subjekte lediglich mit Mainzer Stadtteilen in Bezug gesetzt wurden. Dies trifft zwar zu, jedoch liegen diese Geografika ebenfalls zu weit entfernt vom historischen Stadtkern; z.B.: <https://d-nb.info/gnd/4421824-2>

9 Diskussion und Demonstration

Das Arbeitsergebnis wird zunächst aus der Sicht einer Datenbereitstellerin und dann aus Anwenderinnensicht beurteilt. In Abschnitt 9.1 werden die Festlegungen aus Kapitel 6 aufgegriffen. Es wird geprüft, ob bzw. inwieweit der generierte RDF-Graph die definierten Anforderungen erfüllt. In Abschnitt 9.2 erfolgt die Demonstration mithilfe des Informationssystems (Abschnitt 7.1) und beziehungsweise auf das vorgestellte Anwendungsszenario (Abschnitt 6.4). Anhand konkreter Fragestellungen wird dargelegt, dass die verknüpften Informationen für potenzielle Anwenderinnen einen Mehrwert bieten. Es wird gezeigt, dass das Arbeitsergebnis im Vergleich zu den Ausgangsdaten einen erhöhten Informationsumfang aufweist. Dann wird die Leitfrage beantwortet (Abschnitt 9.3). Schließlich wird die Übertragbarkeit der gefundenen Lösungen besprochen (Abschnitt 9.4).

9.1 Anforderungen

Zur Erfüllung der Anforderung **LOD-A1** wurde jede signifikante Ressource per HTTP-URI identifiziert. Es wurde dargelegt, wie die CURIE-Empfehlungen von Allemang et al. (2020) und die URI-Strategie von van den Brink et al. (2014) umgesetzt wurden (s. GP 8.2.1). Zwecks Vermeidung von URI-Kollisionen wurden NIR-URIs und 303-URIs im Rahmen des ETL-Prozesses konstruiert, statt letztere serverseitig dynamisch zu generieren (s. Abschnitt 10.2). Die Hash-URIs wurden im Zuge der Konzeptualisierung erzeugt (s. GP 8.2.2). Die Anforderung **LOD-A2** konnte nicht vollständig erfüllt werden. Jede instanziierte Geometrie wurde über eine NIR-URI identifiziert. Diesbezüglich war der Verzicht auf leere Knoten sinnvoll, weil dadurch Aussagen über die Geometrien des historischen Häuserbuchs gemacht werden können. Es ist häufiger der Fall, dass die nicht-räumlichen Angaben zu einem Feature als gesichert gelten (z.B. die Besitzverhältnisse), die Geometrie jedoch approximiert wurde (z.B. der Grundriss). Es wäre ein erstrebenswertes Ziel z.B. die Genauigkeit, etwaige Ungewissheiten etc. in RDF-Tripelform auszudrücken. Anknüpfungspunkt dafür sollte im o.g. Fall die Geometrie sein und nicht immer das „ganze“ Feature. Abgesehen davon, bietet GeoSPARQL topologische Prädikate, mit denen in der ABox Aussagen zur „Nachbarschaft“ gemacht werden können. Diese können sich ebenfalls direkt auf die Geometrie beziehen und sind in (unvorherzusehenden) Anwendungsszenarien hilfreich, wo kein GeoSPARQL-Prozessor zur Verfügung steht. Alle Instanzen von Geometrien sind deshalb Ressourcen, die eine URI „verdienen“ (vgl. GP 6.2.1). Für die Instanzen der historischen sozialen Gruppen gilt dies nicht, da substantielle Informationen dazu fehlen. Man könnte auch sagen, dass diese Ressourcen momentan anonyme Instanzen sind, die von einem leeren Knoten repräsentiert werden. Diese sind im konkreten Fall nur Hilfsressourcen mit struktureller Funktion, die noch nicht einmal eine Knoten-ID benötigen und schon gar keine URI. Es wurden weder Aussagen über Aussagen gemacht, noch RDF-Kollektionen oder RDF-Container verwendet. Die Anforderung **LOD-A3** konnte erfüllt werden, indem mehrere Tripel mit demselben Prädikat erzeugt wurden (z.B. zur Angabe der Bewohnerinnen, wo die Reihenfolge der Aussagen keine Rolle spielt). Die in **LOD-A4** geforderte Wiederverwendung von Konzepten konnte nicht erfüllt werden. Das XML-Template und die RDF/XML-Serialisierung im Anhang zeigen, dass nicht nur Konzepte aus Vokabularen nachgenutzt wurden, die sich in der LOD-Cloud etabliert haben. Einige Konzepte wurden individuell spezifiziert. Diese Entscheidung wurde unter GP 7.2.2 begründet. Neben den vordefinierten Konstrukten aus dem RDF- / RDFS- und OWL-Vokabular konnte auf Konzepte aus den u.g. etablierten Vokabularen zurückgegriffen werden:

- Dublin Core Metadata Initiative (DCMI) Metadata Terms
- Friend of a Friend (FOAF)
- Bibliographic Ontology (BIBO)
- GeoSPARQL
- IRW Ontology

- GND Ontology
- Creative Commons (CC)

In diesem Zusammenhang bleibt festzuhalten, dass:

- die Metadaten nicht ausreichend sind, weil wichtige Angaben (z.B. zur Genauigkeit) fehlen,
- wichtige Angaben zur Organisation der Tripel / der Graphen fehlen, weil das ganze Thema in dieser Arbeit nicht prioritär behandelt wurde und deshalb keine substantiellen Aussagen mithilfe von Konzepten aus dem void getroffen werden können (zur Begründung s. GP 7.2.5),
- hauptsächlich unspezifische Konzepte aus den o.g. Vokabularen nachgenutzt wurden; Beispiel für eine Klasse: **foaf:Agent**; Beispiel für ein Prädikat **gndo:relatedWork**.

Bezüglich des letzten Punktes kann positiv hervorgehoben werden, dass alle nachgenutzten Konzepte vorher genau untersucht wurden. Deshalb können keine widersprüchlichen Aussagen gefolgert werden. Ein Beispiel ist die Verwendung des Prädikats **gndo:relatedWork**. Wegen des Definitions- und Wertebereichs kann das Subjekt problemlos eine Instanz von **gndo:BuildingOrMemorial** sein. Es ist auch unkritisch, dass gefolgert werden kann, dass das Objekt eine Instanz von **gndo:Work** ist. Die Schlussfolgerungen stehen nicht im Widerspruch zu den Aussagen, die wegen der Konzepte gefolgert werden können, die in der eigenen TBox definiert wurden. Negativ ist allerdings, dass häufig keine konkreten Aussagen abgeleitet werden können. Hierfür fehlt es an den notwendigen Angaben in den Ausgangsdaten. Es kann z.B. nicht mit Sicherheit gesagt werden, ob ein Agent eine natürliche oder juristische Person war oder ist. Die FME-Workbench wurde nicht konfiguriert, um die semantische Reichhaltigkeit immer weiter durch Datenerfassungen zu steigern. Deshalb und wegen der Verwendung des Prädikats mit dem Bezeichner **rdfs:comment**, das keine automatischen Schlussfolgerungen zulässt, kann die Bezeichnung „Wissensbasis“ in diesem Stadium durchaus in Frage gestellt werden (s.u.).

Die zentrale Anforderung **LOD-A5** konnte bis zu einem gewissen Grad erfüllt werden. Der Informationsumfang des historischen Häuserbuchs wurde erhöht. Genauer gesagt bzw. beziehend auf die Anforderungsdefinition, gilt dies allerdings nur für die Beschreibungen jener Features, die um einen RDF-Link auf der Instanzebene ergänzt werden konnten. Wie dargelegt, wurden die entsprechenden Prädikate, welche geeignete Ressourcen in den Informationsbeständen der DNB referenzieren, erst nach sorgsamer Prüfung eingesetzt. Dies kann mittels des Anhangs verifiziert werden. Rekapitulierend bleibt festzuhalten, dass der Grad der Zielerreichung (der anhand der Anzahl der erzeugten Verknüpfungen quantifiziert wird) für die Literaturangaben bei 75 % (24 Links), für die Agenten bei ca. 39 % (4132 Links) und für die Geoobjekte (53 Links) bei unter 1 % liegt. Die Wertsteigerung ist vor allem in den verlinkten (Sach-)Informationen zu sehen, die im historischen Häuserbuch mit den Geoobjekten verknüpft sind. Für LOD-Applikationen, welche die Links automatisch via SPARQL verfolgen können und für Anwenderinnen, welche die Informationen über die iSPARQL-Anwendung abfragen oder die URIs über einen Browser dereferenzieren können, bieten die verlinkten Informationen einen Mehrwert. Beispielsweise kann in einigen Fällen nun wirklich ermittelt werden, welche Person Bewohnerin eines Gebäudes war, statt nur einen Namen (in „irgendeiner“ Schreibweise) abzufragen (vgl. Abschnitt 9.2).

Die individuelle Anforderung **LOD-I1** wurde vollständig umgesetzt. Um etwaigen Missverständnissen vorzubeugen, wurde explizit angegeben, ob es sich um eine Nicht-Informationsressource oder eine zugehörige Beschreibung handelt. Dafür wurden Klassen aus der IRW-Ontologie referenziert. Da dies eine ausdrucksstarke Ontologie mit komplexen OWL-Konzepten ist, wurden die o.g. Aussagen nochmals mit Konzepten aus dem einfachen FOAF-Vokabular gemacht – genau wie im DBpedia-Projekt. Wegen des dokumentierten Wertebereichs des Prädikats **geo:hasGeometry** und den Ausführungen zum domänenspezifischen WKT-Format (vgl. GP 2.4.2), wurde die Entscheidung getroffen, jede Instanz einer Geometrie explizit zu typen, wobei die NIR-

URI des Objekts (URI-Muster aus Abb. 2.25) das adäquate Simple-Feature-Konzept auf der untersten Hierarchieebene identifiziert (vgl. Abb. 2.24). Wenn man für Schlussfolgerungen ausschließlich auf Inferenzen und keine sonstigen Interpretationen setzt, wären anderenfalls tatsächlich Informationen verloren gegangen. Es wäre dann nur klar gewesen, dass die Geometrie eine Instanz von **geo:Geometry** ist (und aller übergeordneten Klassen, weil es keine Einschränkungen gibt). Es wurde nicht das CRS verwendet, das GeoSPARQL standardmäßig vorsieht. Stattdessen wurde das CRS (Parameter: [ESPG:4326](#)) verwendet, das vom W3C und OGC im LOD-Kontext empfohlen wird. Deshalb musste die URI hinzugefügt werden, die das Dokument lokalisiert, welches das verwendete CRS beschreibt. Somit kann nachvollzogen werden, in welcher Reihenfolge die Koordinaten angegeben wurden. Im Zuge der Konzeptualisierung wurden die jeweiligen Klassen und Prädikate mithilfe von Properties ohne formale Semantik natürlichsprachlich kommentiert. Dadurch wurde der Graph zwar nicht durch logische Aussagen angereichert; die Texte sind jedoch hilfreich für die menschliche Interpretation.

Zur Erfüllung der individuellen Anforderung **LOD-I2** wurde ein gerichteter Graph erzeugt. Dieser wurde mithilfe von RDF/XML serialisiert. Das wohlgeformte XML-Dokument wurde verständlich gestaltet. Als Zeichensatz wurde Unicode verwendet. Die Zeichenkodierung erfolgte mittels UTF-8. Da es im XML-Templater von FME Desktop nicht möglich ist, Attributwerte aus Variablen in einen Character Data-Abschnitt einzufügen, wurden die reservierten Zeichen in den Elementen zur Geometrikodierung (**geo:asWKT**) gemäß der XML-Regeln umgeschrieben. Die vorstehenden Ausführungen können anhand des Anhangs verifiziert werden.

Die in **AS1** geforderte (vollständige) Übernahme und die Verbesserung der DQ konnte erreicht werden. Im Rahmen des ETL-Prozesses wurden alle Werte aus den definierten Spalten der ausgewählten MapInfo-Relationen übernommen. Dabei erfolgte:

- die Koordinatentransformation unter Berücksichtigung des empfohlenen CRS,
- die Korrektur fehlerhafter Geometrien,
- die Datenbereinigung (Duplikaterkennung und -eliminierung, Überprüfung, Korrektur), wodurch hochwertigere Tabellen (für die Literaturangaben und die Agenten) erstellt werden konnten.

Da einige Zeichenketten von Prädikaten referenziert wurden, die keine formale Semantik besitzen (vor allem **rdfs:comment**), muss klargestellt werden, dass dies nicht zu einer Wertsteigerung führte. Insbesondere die „Bauinformationen“ manifestieren sich lediglich als isolierte Datenfragmente, die bestenfalls von einem Menschen interpretiert werden können. Andererseits ist die semi-automatische Generierung von sinnhaften Informationsobjekten aus variablen, unstrukturierten Zeichenfolgen, die teils historische Bezeichnungen bilden, mithilfe einer intelligenten Software ein eigenständiges Arbeitsthema. Einige der strukturellen Probleme konnten im Zuge des Verknüpfungsprozesses relativiert werden. Die durchgeführten Datenbereinigungen waren die Grundvoraussetzung dafür.

In Abschnitt 8.2.2 wurde beschrieben, wie die Anforderung **AS2** umgesetzt werden konnte. Es wurde dargelegt, dass die erstellte TBox alle erforderlichen Klassen enthält. In der ABox wurden damit sämtliche Typen von historischen Geoobjekten angegeben, die in den ausgewählten MapInfo-Relationen vorhanden sind. Alle existierenden Werte aus der Spalte „Nutzung“ (originäre MapInfo-Relationen, vgl. Tab. 5.1) und alle vorhandenen Werte aus der Spalte „Type“ (adaptierte MapInfo-Relationen, vgl. Tab. 5.4 und Tab. 5.5) wurden bei der Klassendefinition berücksichtigt. Historische Kenntnisse wurden – soweit vorhanden – zwecks hierarchischer Strukturierung eingebracht. Es wurde klargestellt, dass die TBox für zukünftige Erweiterungen als Basis dienen soll. Dies wurde bei der Modellierung berücksichtigt. Die Überlegungen hierzu wurden durch Beispiele veranschaulicht. Dabei wurde stets das UoD fokussiert. Die Komplexität wurde auf das notwendige Minimum reduziert. Außerdem wurde dargelegt, dass die erstellte TBox alle erforderlichen Prädikate enthält. Die Festlegung

der Definitions- und Wertebereiche wurde begründet. Zusätzlich wurden die Auswirkungen theoretisch mithilfe der IR verdeutlicht und anhand von Beispielen aus der ABox demonstriert. Dadurch konnte aufgezeigt werden, dass die spezifizierten Definitions- und Wertebereiche die semantische Verbindung zwischen den erstellten Klassen und den definierten Prädikaten darstellen. Damit konnten die Beziehungen der Individuen untereinander beschrieben werden, die in den ausgewählten MapInfo-Relationen vorhanden sind. Der RDF-Graph konnte durch die einfache TBox komplettiert werden, die ihrem Primär- / Sekundärzweck gerecht wird.

9.2 Anwendungsszenario

Bezugnehmend auf das beschriebene Anwendungsszenario wird nachfolgend demonstriert, dass nun nutzbare Antworten auf konkrete Fragen gegeben werden können, die sich anhand der ursprünglichen MapInfo-Relationen nicht beantworten lassen. Hierfür werden die unter GP 6.4 formulierten Fragen via SPARQL beantwortet. Wie erwähnt, hat die SPARQL-Abfrage, die unter GP 8.2.4 zu finden ist, zwei Funktionen. Es war wichtig, dass die Selektion funktioniert und dass sich die Abfrage für die Erläuterung eignet. Wegen Letzterem wurde auf einiges verzichtet z.B. auf die explizite Klammerung der verwendeten Operatoren. Die folgenden SPARQL-Abfragen bzw. -Statements dienen nur noch der Selektion oder der Konstruktion. Die Ausführungen werden kompakt gehalten. Deshalb werden die folgenden Festlegungen auch nur einmal erwähnt:

```
# -----
# ?MeaningfulName = query variable | $_[a-z] = intermediate variable | _:[0-9] = blank node
# TD = Titeldaten | GND = Gemeinsame Normdatei | HHB = Historisches Häuserbuch
# namespace declaration:
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gndo: <https://d-nb.info/standards/elementset/gnd#>
PREFIX mz: <http://history.mainz.de/def#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdau: <http://rdaregistry.info/Elements/u/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX irw: <http://www.ontologydesignpatterns.org/ont/web/irw.owl#>
PREFIX gnd: <https://d-nb.info/gnd/>
PREFIX hhb: <http://www.mainz.de/id/>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX agrelon: <https://d-nb.info/standards/elementset/agrelon#>
PREFIX gndv: <https://d-nb.info/standards/vocab/gnd/gnd-sc#>
# -----
```

Lis. 9.1: Festlegungen für SPARQL (Select/Construct), EK

Die Formulierung geeigneter Abfragen zwecks Beantwortung der eigentlichen Fragen setzt die Beschaffung von Grundlageninformationen voraus. Die folgende Auflistung soll nur zeigen, was damit gemeint ist. Sie ist nicht abschließend und die jeweiligen Resultate werden aus den vorgenannten Gründen auch nicht beigelegt.

```

# ask for possible SubjectHeadings
SELECT *
WHERE {GRAPH <http://localhost:8890/GND>
  {?URI a gndo:SubjectHeadingSensoStricto;
    gndo:preferredNameForTheSubjectHeading|gndo:variantNameForTheSubjectHeading ?Label.
    #Try: BIND(DATATYPE(?Label) AS ?DatatypeURILabel) or BIND(LANG(?Label) AS ?HasLangTag)
    FILTER (REGEX(?Label, "buch", "i"))}}

# describe my SubjectHeading ('Buchbesitz')
DESCRIBE gnd:4681622-7

# give me a class overview
SELECT *
FROM <http://localhost:8890/HHB>
FROM <http://localhost:8890/GND>
WHERE
  {{?Class a owl:Class;
    rdfs:label ?Label.
  } OPTIONAL {?Class rdfs:comment ?Comment.}}

# show me the properties used to describe NonInformationResources / show me the metadata
SELECT DISTINCT ?Property
FROM <http://localhost:8890/HHB>
WHERE
  {$_a a irw:NonInformationResource;
    ?Property $_b.}

# give me some individuals with values
SELECT ?S ?P ?O
WHERE {GRAPH <http://localhost:8890/HHB>
  { ?S ?P ?O;
    a $_a.
    FILTER ($_a IN (foaf:Agent, mz:Building, bibo:Document))}}
LIMIT 1000

# describe a building
DESCRIBE hhb:B743

# show me some instances of mz:Building and their inferred types
SELECT DISTINCT *
FROM <http://localhost:8890/HHB>
WHERE {?I rdf:type/rdfs:subClassOf* mz:Building;
  a ?T.}
LIMIT 1000

# some data type information please
SELECT ?DTypeURI_Info ?DTypeURI_DeathDate ?DTypeURI_PeriodOfActivity
FROM <http://localhost:8890/GND>
WHERE
  {
    gnd:119636107 gndo:biographicalOrHistoricalInformation $_a;
    gndo:dateOfDeath $_b;
    gndo:periodOfActivity $_c.
    BIND(IF(isLITERAL($_a),DATATYPE($_a),"bNode or URI") AS ?DTypeURI_Info) #'rdf:langString'
    BIND(IF(isLITERAL($_b),DATATYPE($_b),"bNode or URI") AS ?DTypeURI_DeathDate) #xsd:date
    BIND(IF(isLITERAL($_c),DATATYPE($_c),"bNode or URI") AS ?DTypeURI_PeriodOfActivity) #xsd:string
    #etc.
  }
  #etc.

```

Lis. 9.2: Exemplarische SPARQL-Abfragen zwecks Vorbereitung, EK

Die Tabelle, welche auf die nächste Abfrage folgt, beantwortet alle Fragen aus Gruppe 1.

```

SELECT ?BuildingName ?Year ?BuildingInfo ?Owner ?Occupant ?Inquiry
FROM <http://localhost:8890/HHB>
FROM <http://localhost:8890/GND>
WHERE
{
  {
    {$_a a mz:University;
      foaf:name ?BuildingName;
      mz:when ?Year;
      mz:where/geo:asWKT ?fWKT;
      rdfs:comment ?BuildingInfo.
      FILTER (?Year > "1554"^^xsd:gYear &&
        ?Year < "1636"^^xsd:gYear &&
        REGEX(STR(?BuildingInfo), "burse", "i") &&
        geof:sfWithin(?fWKT, ''
          <http://www.opengis.net/def/crs/EPSSG/0/4326>
          POLYGON ((50.010833 8.271583,
            49.999969 8.260583,
            49.994699 8.267043,
            49.987569 8.296206,
            49.988899 8.298845,
            49.999258 8.283808,
            50.010833 8.271583))
            ''^^ geo:wktLiteral))
        }OPTIONAL{$_a mz:hadOwner/foaf:name $_b;
          mz:hadOwner[rdfs:seeAlso|owl:sameAs _:0].
          _:0 $_c $_d.
          $_c rdfs:subPropertyOf gndo:preferredName.
          BIND(IF(STR($_b) != $_d, CONCAT($_b, " / ", $_d), $_d) AS ?Owner)
        }
      }OPTIONAL{$_a mz:hadOccupant/foaf:name $_e;
        mz:hadOccupant/owl:sameAs ?Inquiry. # primary interest
        ?Inquiry gndo:preferredNameForThePerson $_f.
        BIND(IF(STR($_e) != $_f, CONCAT($_e, " / ", $_f), $_f) AS ?Occupant)
      }
    }
  }
}

```

Lis. 9.3: SPARQL-Abfrage (Fragen aus Gruppe 1), EK

	BuildingName	Year	BuildingInfo	Owner	Occupant	Inquiry
1	"Zum Gutenberg"@de	"1594"^^xsd:gYear	"Protonotariat. Eh. Juristenkolleg, Universitätsburse"@de	"Universität Mainz"	"Dr. Jacobus Campius / Campius, Jacobus"	https://d-nb.info/gnd/119636107
2	"Zum Algesheimer"@de	"1594"^^xsd:gYear	"Jesuitenburse und Universitätskolleg"@de	"Jesuiten"		
3	"Zum Schenkenberg"@de	"1620"^^xsd:gYear	"Universitätsburse und Kolleg"@de	"Johann Schweikhard von Cronberg / Johann Schweikhard, Mainz, Erzbischof"		

Tab. 9.1: Informationen zu den Bursen (Antworten auf die Fragen aus Gruppe 1)

Die Tabelle, welche (nicht wirklich) auf die nächste Abfrage folgt, beantwortet die Fragen 2.1, 2.2, 2.4 - 2.6.

```

SELECT ?ID (GROUP_CONCAT (DISTINCT $ _c; separator=" / ") AS ?Names) (GROUP_CONCAT(DISTINCT $ _a; separator=" / ") AS
?AcademicDegrees) ?BirthPlace ?PeriodOfActivity ?DateOfBirth ?DateOfDeath (GROUP_CONCAT(DISTINCT $ _b; separator=" / ")
AS ?PlacesOfActivity) (GROUP_CONCAT(DISTINCT $ _d; separator=" / ") AS ?Professions)
WHERE
{ GRAPH <http://localhost:8890/GND>
{
{ gnd:119636107  gndo:gndIdentifier ?ID;
  gndo:placeOfBirth/gndo:preferredNameForThePlaceOrGeographicName ?BirthPlace;
  gndo:dateOfDeath ?DateOfDeath;
  gndo:academicDegree $ _a;
  gndo:placeOfActivity/gndo:preferredNameForThePlaceOrGeographicName $ _b;
  gndo:preferredNameForThePerson|gndo:variantNameForThePerson $ _c;
  gndo:professionOrOccupation/gndo:preferredNameForTheSubjectHeading $ _d;
  gndo:periodOfActivity ?PeriodOfActivity.
} UNION { { gnd:119636107  ((gndo:familialRelationship|^gndo:familialRelationship)
|(gndo:acquaintanceshipOrFriendship|^gndo:acquaintanceshipOrFriendship)) $ _e.
  FILTER EXISTS { $ _e
    gndo:placeOfActivity/gndo:preferredNameForThePlaceOrGeographicName "Mainz"^^xsd:string. }
    $ _e gndo:preferredNameForThePerson|gndo:variantNameForThePerson $ _c;
    gndo:gndIdentifier ?ID.
  } OPTIONAL { $ _e gndo:placeOfBirth/gndo:preferredNameForThePlaceOrGeographicName ?BirthPlace.
  }
  OPTIONAL { $ _e gndo:dateOfBirth ?DateOfBirth. }
  OPTIONAL { $ _e gndo:dateOfDeath ?DateOfDeath. }
  OPTIONAL { $ _e gndo:placeOfActivity/gndo:preferredNameForThePlaceOrGeographicName $ _b. }
  OPTIONAL { $ _e gndo:academicDegree $ _a. }
  OPTIONAL { $ _e gndo:professionOrOccupation/gndo:preferredNameForTheSubjectHeading $ _d. }
  OPTIONAL { $ _e gndo:periodOfActivity ?PeriodOfActivity. }
}
}
}
GROUP BY ?ID ?BirthPlace ?PeriodOfActivity ?DateOfBirth ?DateOfDeath

```

Lis. 9.4: SPARQL-Abfrage (Fragen aus Gruppe 2), EK

	ID	Names	AcademicDegrees	BirthPlace	PeriodOfActivity
1	"119636107"	"Campius, Jacobus / Campius, Jacob / Campius, Jakob"	"Prof. / Dr. iuris utriusque"	"Eilendorf (Aachen)"	"1587-1596"
2	"138599041"	"Campius, Dionysius / D C D / D.C.D."	"Lic. iur., Dr. iur. / Dr. iuris utriusque"	"Eilendorf (Aachen)"	
	DateOfBirth	DateOfDeath	PlacesOfActivity	Professions	
1		"1609-06-03"^^xsd:date	"Bonn / Köln / Mainz"	"Jurist / Hochschullehrer / Kanzler / Stadtratsmitglied / Dekan <Kirche> / Kanoniker"	
2		"1641-08-25"^^xsd:date	"Mainz"	"Jurist / Hochschullehrer"	

Tab. 9.2: Informationen für biografischen Skizzen der Bewohner (zu den Fragen 2.1, 2.2, 2.4 - 2.6.)

Tatsächlich folgt diese Tabelle natürlich nicht auf die obige Abfrage; die Tabelle musste so formatiert werden.

Die folgenden Statements dienen der Beantwortung der Fragen 3.1 und 3.4 für die natürlichen Personen (vgl. „ID“). Der resultierende Graph wurde mithilfe von RDF/XML serialisiert. Die Namensräume gehören dazu.

```

CONSTRUCT
{
    ?PUB1 ?r3 ?P1;
    dc:title ?T1;
    rdau:P60327 ?A1.
    ?PUB2 ?r3 ?P2;
    dc:title ?T2;
    rdau:P60327 ?A2.
    ?P1 ?r1 ?P2.
    ?P2 ?r2 ?P1.
}
FROM <http://localhost:8890/TD>
FROM <http://localhost:8890/GND>
WHERE
{
    ?P1 gndo:gndIdentifizier $_1.
    ?P2 gndo:gndIdentifizier $_2.

    { { { ?P1 ?r1 ?P2. } UNION { ?P2 ?r2 ?P1. }
      } UNION { ?PUB1 ?r3 ?P1, gnd:4681622-7; dc:title ?T1;
              dcterms:creator [ gndo:preferredNameForThePerson ?A1 ]. }
    } UNION { ?PUB2 ?r3 ?P2, gnd:4681622-7; dc:title ?T2;
              dcterms:creator [ gndo:preferredNameForThePerson ?A2 ]. }

    FILTER ( $_1 = "119636107"^^xsd:string && $_2 = "138599041"^^xsd:string )
}

```

Lis. 9.5: SPARQL-Konstruktion eines Graphen (Fragen 3.1 und 3.4), EK

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:gndo="https://d-nb.info/standards/elementset/gnd#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:gnd="https://d-nb.info/gnd/"
  xmlns:agrelon="https://d-nb.info/standards/elementset/agrelon#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdau="http://rdaregistry.info/Elements/u/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">

  <rdf:Description rdf:about="https://d-nb.info/gnd/119636107">
    <gndo:familialRelationship rdf:resource="https://d-nb.info/gnd/138599041"/>
    <agrelon:hasNieceNephew rdf:resource="https://d-nb.info/gnd/138599041"/>
  </rdf:Description>

  <rdf:Description rdf:about="https://d-nb.info/gnd/138599041">
    <gndo:familialRelationship rdf:resource="https://d-nb.info/gnd/119636107"/>
    <agrelon:hasAuntUncle rdf:resource="https://d-nb.info/gnd/119636107"/>
  </rdf:Description>

  <rdf:Description rdf:about="https://d-nb.info/1059708752">
    <dcterms:subject rdf:resource="https://d-nb.info/gnd/119636107"/>
    <dc:title>Dionysius und Jacobus Campius als Buchbesitzer - eine Momentaufnahme (nicht nur) mit Blick auf die
    Mainzer Karmeliten</dc:title>
    <rdau:P60327>Ottermann, Annelen</rdau:P60327>
    <dcterms:subject rdf:resource="https://d-nb.info/gnd/138599041"/>
  </rdf:Description>

</rdf:RDF>

```

Lis. 9.6: Antworten auf die Fragen 3.1 und 3.4 (RDF/XML), generiert von VOS

Tab. 9.2 zeigt, dass diese Darstellungsform auch für die Ergebnisse von SELECT-Abfragen hier nur bedingt geeignet ist, weshalb das Resultat der letzten Abfrage, welche die o.g. Fragen für die relevanten Organisationen beantwortet, in der schlanken JavaScript Object Notation (JSON) wiedergegeben wird. Damit wird auch Frage 3.3 beantwortet.

```

SELECT ?URI ?prefName ?Term (GROUP_CONCAT (DISTINCT $_a; separator=" / ") AS ?varNames) ?Info (GROUP_CONCAT (DISTINCT
$_b; separator=" / ") AS ?assignedDirectly) (GROUP_CONCAT (DISTINCT $_f; separator=" / ") AS ?Publications)
FROM <http://localhost:8890/TD>
FROM <http://localhost:8890/GND>
WHERE
{
  { { { { ?URI a gndo:CorporateBody;
          gndo:preferredNameForTheCorporateBody ?prefName;
          gndo:dateOfTermination ?Term;
          gndo:placeOfBusiness gnd:4037124-4;
          gndo:gndSubjectCategory gndv:6.6.
          FILTER (!isBLANK(?Term) && REGEX(STR(?Term), "^(?:17|18)\\d{2}$")
                && REGEX(?prefName, "univers", "i"))
        } OPTIONAL { ?URI gndo:variantNameForTheCorporateBody $_a
          FILTER(REGEX($_a, "univers", "i")) }
        } OPTIONAL { ?URI gndo:biographicalOrHistoricalInformation ?Info. }
      } OPTIONAL { ?URI ^gndo:affiliation/gndo:gndIdentifier $_b.
        FILTER ( $_b = "119636107"^^xsd:string || $_b = "138599041"^^xsd:string ) }
      } OPTIONAL { ?PUB dcterms:subject ?URI;
        dc:title $_c;
        dcterms:creator/gndo:preferredNameForThePerson $_d. }
      } OPTIONAL { ?PUB rdau:P60493 $_e.
        BIND(IF(BOUND($_e), CONCAT($_c, " ", $_e, " (" , $_d, ") "),
              CONCAT($_c, " (" , $_d, ") ") ) AS $_f) }
    }
}
GROUP BY ?URI ?prefName ?Term ?Info
ORDER BY DESC (?Term)

```

Lis. 9.7: SPARQL-Abfrage zwecks Ergänzung (Fragen aus Gruppe 3), EK

```

{
  "head" : {
    "vars" : [
      "URI",
      "prefName",
      "Term",
      "varNames",
      "Info",
      "assignedDirectly",
      "Publications"
    ]
  },
  "results" : {
    "bindings" : [
      {
        "URI" : {
          "type" : "uri",
          "value" : "https://d-nb.info/gnd/1136480633"
        },
        "prefName" : {
          "type" : "literal",
          "value" : "Universität Mainz"
        },
        "Term" : {
          "type" : "literal",
          "value" : "1823"
        },
        "varNames" : {
          "type" : "literal",
          "value" : "Universität zu Mainz / Alma Universitas Moguntinensis / Universitas Litterarum (Mainz) / Universitas Moguntina / Universitas Electoralis Moguntina / Universidad de Maguncia / Alte Universität Mainz"
        },
        "assignedDirectly" : {
          "type" : "literal",
          "value" : "138599041"
        },
        "Publications" : {
          "type" : "literal",
          "value" : ""
        }
      }
    ],
    {
      "URI" : {
        "type" : "uri",
        "value" : "https://d-nb.info/gnd/4586520-6"
      },
      "prefName" : {
        "type" : "literal",
        "value" : "Universitätsbibliothek Mainz (1477-1798)"
      },
      "Term" : {
        "type" : "literal",
        "value" : "1798"
      },
      "varNames" : {
        "type" : "literal",
        "value" : "Bibliotheca Universitatis Moguntinae / Bibliot. Universit. Elect. Mogunt."
      },
      "Info" : {
        "xml:lang" : "de",
        "type" : "literal",
        "value" : "Im Zuge der Revolutionskriege 1798 durch die französische Regierung aufgehoben, 1803/1805 ging der Bestand in den Besitz der Stadt Mainz über und befindet sich heute in der Wissenschaftlichen Stadtbibliothek"
      },
      "assignedDirectly" : {
        "type" : "literal",
        "value" : ""
      },
      "Publications" : {
        "type" : "literal",
        "value" : "Mit 23 Bänden fing es an ... von der alten Universitätsbibliothek Mainz, heute Stadtbibliothek, zwischen Humanismus und Aufklärung (Mathy, Helmut) "
      }
    }
  ]
}

```

Lis. 9.8: Weitere Antworten auf die Fragen aus Gruppe 3 (JSON), generiert von VOS

Mithilfe der o.g. Arbeiten kann die Schülerin nicht nur die Verbindung von Jacobus Campius zur Alma Mater Moguntina herstellen, sondern sie erfährt auch, dass in der Mainzer Altstadt Buchbesitzer gewohnt haben, deren Bücher von den Schweden im Dreißigjährigen Krieg beschlagnahmt respektive geraubt wurden (Ottermann, 2014). Da sie die offenen Fragen, welche die vorgenannte Autorin identifizierte, sehr interessant findet, beschließt sie Kontakt zur Autorin aufzunehmen, um sich mit ihr abzustimmen. Dafür stellt sie noch eine letzte SPARQL-Abfrage: DESCRIBE <https://d-nb.info/gnd/112004857>

Anhand der Ausgangsdaten können nur die Fragen aus Gruppe 1 beantwortet werden. Dies ist nachvollziehbar, da mit dem Beispiel um die o.g. historische Persönlichkeit schon im theoretischen Teil dieser Arbeit begonnen wurde. Die entsprechenden Tabellen, Abbildungen und Listings enthalten alle relevanten Informationen dazu.

9.3 Beantwortung der Leitfrage und Fazit

In dieser Arbeit wurde die Anwendbarkeit einer bekannten Methode in einem spezifischen Kontext untersucht. Es wurde der Leitfrage nachgegangen, ob die LOD-Methode auch angewendet werden kann, um den Informationsumfang von Modellen zu erhöhen, die keine konzeptuelle Grundlage aufweisen, die nur aus vektoriellen Geodaten bestehen, welche ausschließlich in Form von nichtnormalisierten Relationen vorliegen. Bezugnehmend auf die definierten und größtenteils – bzw. bis zu einem bestimmten Grad – erreichten Teilziele kann diese Frage *grundsätzlich* bejaht werden, was wie folgt präzisiert wird:

In Abschnitt 5.1 wurde dargelegt, dass die ausgewählten Relationen die o.g. Eigenschaften aufweisen. Wenn solche Modelle unter konsequenter Beachtung der LOD-Prinzipien umgeformt werden, entsteht zwangsläufig ein RDF-Graph. Dieser muss oder soll bestimmten Anforderungen genügen, auf die im vorherigen Abschnitt zurückgegriffen wurde, um das Arbeitsergebnis von der theoretischen Seite bewerten und diskutieren zu können. Die negativen Aspekte bzw. die nicht vollständig erfüllten, teils optionalen, Anforderungen tangieren die positive Beantwortung der Leitfrage jedoch nicht. Sicherlich ist es suboptimal, dass während der Umformung einige Datenfragmente unter Verwendung von Prädikaten ohne formale Semantik einfach als Objekte eingefügt wurden. Dadurch kann der Wert der Daten nicht gesteigert bzw. die DQ nicht verbessert werden. Die TBox ist ebenfalls nicht optimal; an einigen Stellen ist sie verbesserungsbedürftig, was nicht zuletzt auf fehlende Informationen zurückzuführen ist. Für die Beurteilung, ob der erzeugte RDF-Graph aus der Sicht einer potenziellen Anwenderin einen angemessenen Informationsumfang bietet, wurde jedoch allein auf die RDF-Links auf der Instanzebene abgestellt (vgl. Begriffserläuterungen zur Anforderung **LOD-A5**). Der letzte Teilschritt im Workflow (Anreicherung) wurde deshalb prioritär behandelt (vgl. GP 7.2.4). Im Rahmen des ETL-Prozesses wurde die Grundvoraussetzung für die Anreicherung geschaffen (Stichwort: Datenbereinigung). Durch die Konzeptualisierung konnte der Graph komplettiert werden, was neben der Zusammenführung der verschiedenen Informationsbestände (Konsolidierung) ebenfalls eine Voraussetzung war, um ihn sinnvoll via SPARQL abfragen zu können. Mithilfe solcher Abfragen wurde in diesem Kapitel praktisch demonstriert, dass nun nutzbare Antworten auf konkrete Fragen gegeben werden können, die sich anhand der ursprünglichen MapInfo-Relationen nicht beantworten lassen. Es existieren jedoch nach wie vor Datenfragmente. Dazu zählen die bereits erörterten „Bauinformationen“. Im Rahmen dieser Arbeit konnte die umfassende Informationsintegration nicht bewerkstelligt werden. Dennoch konnte bewiesen werden, dass durch die Anwendung der LOD-Methode der Informationsumfang von Modellen *bis zu einem gewissen Grad* gesteigert werden kann, selbst wenn diese nur aus vektoriellen Geodaten bestehen, welche ausschließlich in Form von nichtnormalisierten Relationen vorliegen. Der Grad der Zielerreichung wurde anhand der Anzahl der erzeugten RDF-Links quantifiziert.

Es ist durchaus statthaft die unter GP 8.2.4 genannte Summe zu reduzieren, indem die Links abgezogen werden, welche die Organisationen über das Prädikat `rdfs:seeAlso` verknüpfen (510 Links). Dieses Prädikat hat (faktisch) keine formale Semantik. Als Begründung könnte also angeführt werden, dass sich die Bedeutung dieser Links primär der Datenbereitstellerin erschließt, aber nicht (zwangsläufig) der potenziellen Anwenderin oder der Entwicklerin einer LOD-Applikation. Allerdings sind die generierten Identity Links – wie in der LOD-Cloud üblich – mit der formal spezifizierten Semantik des Prädikats `owl:sameAs` in den allerwenigsten Fällen in Einklang zu bringen. Insofern ist die Prozessierung dieser Links ebenfalls von individuellen Interpretationen oder Implementierungen abhängig und nicht von den eigentlich beabsichtigten Inferenzen, was die Debatte – aus subjektiver Sicht – ad absurdum führt. Statt automatisch/programmgesteuert Identity Links zu generieren, wurden die zugehörigen Beschreibungen in dieser Arbeit genauestens untersucht. Statt ausschließlich Identity Links zu generieren, wurde das Prädikat `rdfs:seeAlso` eingesetzt, wenn die ermittelten NIR-URIs nicht als URI-Aliase angesehen werden konnten. In diesen Fällen wird nicht suggeriert, dass es sich um Identity Links handelt.

Während der praktischen Umsetzung des Workflows wurden weitere Erkenntnisse gewonnen:

- Zwecks Verlinkung konnte auf die Arbeitsergebnisse bzw. die Informationsbestände anderer zurückgegriffen werden, die im Netz bereits vorhanden sind. Dies ist positiv hervorzuheben, da die Erhöhung des Informationsumfangs nicht komplett in Eigenleistung durchgeführt werden musste. Dies setzt aber voraus, dass ein **geeigneter Informationsbestand** für die Verlinkung **existiert** bzw. **ermittelt** werden kann. Im konkreten Fall war ein qualitativ hochwertiger und umfangreicher Informationsbestand erforderlich. Nachdem im Zuge der Anreicherung tausende Ressourcen in den Informationsbeständen der DNB untersucht wurden, bleibt festzuhalten, dass die kostenlosen Informationen (und Dienste) der DNB besonders gut geeignet waren, da sie die genannten Kriterien erfüllen (Struktur, Korrektheit, Dokumentation, Anzahl der Aussagen über ein Individuum, Gesamtzahl der beschriebenen Individuen, Vorhandensein ausgehender RDF-Links). Die erforderliche **Geeignetheitsprüfung** war *arbeits- und zeitintensiv*, da neben den Instanzdaten auch die relevanten Konzepte aus den Ontologien untersucht werden mussten. Wenn dabei – im Gegensatz zur vorliegenden Arbeit – Informationen aus verschiedenen Quellen/LOD-Projekten nachgenutzt werden, die unterschiedliche Charakteristika aufweisen (verwendete Vokabulare, Tripelstruktur etc.), muss das *Zeitbudget* weiter erhöht werden.
- Die **Bestimmung der Kandidatenmenge** bzw. der **Ergebnismenge** zwecks Verknüpfung der Ressourcen war herausfordernd. Wenn keine eindeutige Referenz auf die Bezugsobjekte über Zeichenketten vorhanden war und diese auch kein bestimmtes Muster aufwiesen, konnte die Ergebnismenge nicht automatisch ermittelt werden. Die Kombination bestimmter Literale und Zeichenkettenähnlichkeitsfunktionen war faktisch unmöglich. Beide der vorgenannten Aspekte können im konkreten Fall auch durch die Normalisierung der Ausgangsrelationen nicht positiv beeinflusst werden. Dies gilt nicht für die Zeichenketten („Literaturangaben“), die zwecks Ermittlung des jeweiligen Pendants über reguläre Ausdrücke zerlegt werden konnten.
- Wenn eine Datenbereitstellerin die definierten Anforderungen an die zu generierenden RDF-Links auf der Instanzebene stellt und die Rahmenbedingungen/Ausgangsdaten vergleichbar sind, ist die **Evaluation der potenziellen Kandidatenpaare** ein *arbeits- und zeitintensiver Prozess*.

Die Erhöhung des Informationsumfangs durch Anwendung der LOD-Methode, insbesondere unter Beachtung des vierten LOD-Prinzips, war zwar *effizienter* als die manuelle Erfassung aller Informationen, die in den relevanten Beschreibungen der DNB vorhanden sind. Trotzdem gestaltete sich die *Anreicherung sehr aufwendig*. Die vorherige Normalisierung der Ausgangsrelationen mithilfe einer RDB hat darauf nahezu keinen Einfluss.

Fazit: Die LOD-Methode kann angewendet werden, um vektorielle Geodaten, die in nichtnormalisierten Relationen vorliegen, zu einem RDF-Graph umzuformen, der den LOD-Anforderungen entspricht. In diesem Zuge kann eine konzeptuelle Grundlage geschaffen werden, die den RDF-Graph sinnvoll ergänzt. Der Aufwand für die Transformation ist verhältnismäßig gering, wenn lediglich die notwendigsten Datenbereinigungen durchgeführt werden, statt die Daten vollständig zu zerlegen. Selbiges gilt hinsichtlich der Konzeptualisierung, sofern der Komplexitätsgrad der TBox auf das notwendige Minimum reduziert wird. Es ist möglich den Informationsumfang durch die Generierung von RDF-Links auf der Instanzebene zu erhöhen. Dieser Prozess gestaltet sich hingegen sehr aufwendig, selbst wenn auf einen adäquaten Informationsbestand aus der LOD-Cloud zurückgegriffen werden kann. Trotz hoher Informationskosten, resultierend aus der Suche, Beschaffung, Analyse, Konsolidierung und Vereinigung, ist die Verlinkung von Ressourcen effizienter als die manuelle Datenerfassung. Im Vergleich zu den Ausgangsdaten, bieten die verknüpften Informationen aus der Sicht einer Anwenderin einen erkennbaren Mehrwert, auch wenn der RDF-Graph noch isolierte Datenfragmente enthält.

9.4 Übertragbarkeit

Der konzipierte und umgesetzte Workflow zwecks Generierung des RDF-Graphen erweitert die Verfahrensweise von Nishanbaev et al. (2020), welche auf die Integration von LGCHD ins Web of Data abzielt (vgl. Abb. 3.3 und Abb. 7.8). Für ähnliche Vorhaben kann der gesamte Workflow als Orientierungshilfe dienen.

Die konkrete Umsetzung ist jedoch stark anwendungsfallspezifisch. Es bedarf daher einer individuellen Prüfung im Einzelfall, ob der ELT-Prozess bzw. die FME-Workbench und/oder die erstellte TBox wiederverwendet werden können. Um dies hervorzuheben, wurde von vornherein klargestellt, dass für die Konzeption und die praktische Realisierung der beiden ersten Teilschritte (Transformations- und Konzeptualisierungsprozess) „zwangsläufig die individuellen Charakteristika der Ausgangsdaten bzw. die anwendungsfallspezifischen Anforderungen maßgeblich sind“. Dementsprechend wurde keine generelle Transformationsmethode und kein generisches Modellierungsmuster für Ontologien entwickelt. Für ähnliche Vorhaben wird empfohlen, basierend auf den Erkenntnissen, die im Rahmen der Daten- / Anforderungsanalyse gewonnen werden, den Konzeptualisierungs- und Transformationsprozess individuell auszugestalten. Für Geeignetheitsprüfungen infrage kommender RDF-izer kann die überblicksmäßige Darstellung und die anknüpfende Diskussion hilfreich sein.

Nishanbaev et al. (2020, p. 887) sahen in ihrer „Methodology for Producing Linked Geospatial Cultural Heritage Data“ einen semi-automatischen Teilschritt zwecks Erzeugung von RDF-Links vor und wiesen in diesem Zusammenhang auf Folgendes hin: „This step usually involves verification of the correctness of the computed links, which is performed by people“. In der vorliegenden Arbeit wurde der Teilschritt detailliert ausgearbeitet. Es wurde eine Verfahrensweise entwickelt und angewendet, die sich auf ähnliche Anwendungsfälle übertragen lässt. Diese Verfahrensweise ist von manuellen Tätigkeiten geprägt, die zu hohen Informationskosten führen.

Der erstellte RDF-Graph kann für andere LOD-Projekte über ein Netzwerk übertragen werden.

10 Zusammenfassung und Ausblick

10.1 Zusammenfassung

Die erfolgreiche LOD-Initiative hat einen wichtigen Beitrag zur Verwirklichung der Vision vom Semantic Web geleistet und die nächste Evolutionsstufe des WWW eingeläutet. Die LOD-Cloud ist eine gigantische Graphdatenbank. Sie entstand durch die stringente Anwendung der LOD-Methode, welche auf vier einfachen Regeln – den LOD-Prinzipien – beruht. Die LOD-Cloud umfasst Daten, denen eine Bedeutung (Semantik) zugeordnet wurde. Die verschiedenen Informationen werden dezentral vorgehalten; sie stammen aus unterschiedlichen Quellen. Trotzdem bleibt die formal spezifizierte Semantik anwendungsübergreifend erhalten und die Informationen stehen auch nicht isoliert. Sie werden mithilfe des flexiblen, graphenbasierten RDF-Datenmodells und darauf aufbauenden Informations-Spezifikationssprachen einheitlich repräsentiert. Dank standardisierter Syntax respektive RDF-Serialisierungen können sie maschinell prozessiert werden. Da der RDF-Graph bei konsequenter Beachtung der LOD-Prinzipien (Vermeidung leerer Knoten etc.) vor allem FCOs – Ressourcen, die per URI (ggf. in unterschiedlichen Namensräumen) identifiziert wurden – enthält, lässt sich ein globales Netz aus verlinkten Informationen mittels RDF-Links realisieren. Wenn Informationen durch Verlinkung ins Web (of Data) integriert werden, können sie in unvorherzusehenden Anwendungsszenarien wiederverwandelt werden und es lässt sich ein Informationsmehrwert erzielen. Die LOD-Cloud „lebt“ wegen des Netzwerk-Effekts – für ihre weitere Expansion ist sie auf die Partizipation möglichst vieler Anwenderinnen angewiesen.

Die vorliegende Arbeit liefert einen Beitrag dazu. Im theoretischen Teil wurde Ordnung ins terminologische Chaos gebracht, indem wichtige Fachbegriffe anschaulich erklärt und möglichst präzise definiert wurden. Außerdem wurden essentielle Mechanismen anhand eines Schlüsselbeispiels erläutert und die Wesensmerkmale des Semantic Web wurden im Rahmen einer vergleichenden Gegenüberstellung des Web of Data und des Web of Documents verständlich herausgearbeitet – getrieben von der Absicht potenziellen Interessentinnen einen einfacheren Zugang zu semantischen Konzepten und (Basis-)Technologien zu ermöglichen. Auch der zweiten Hälfte des theoretischen Teils, in welcher der Raumbezug in den Mittelpunkt gerückt wurde, liegt diese Intention zugrunde. Räumliche Konzepte wurden möglichst einfach beschrieben, weil Geoinformationen omnipräsent sind. Im Web werden sie interdisziplinär eingesetzt; sie verknüpfen Wissensbasen verschiedener Domänen.

Bei der Formulierung der praxisnahen Leitfrage spielte eine bestimmte Art von Geodaten eine zentrale Rolle. In der vorliegenden Arbeit wurde kein Experiment mit idealen Daten durchgeführt. Es wurde untersucht, ob die LOD-Methode auch angewendet werden kann, um den Informationsumfang von Modellen zu erhöhen, die keine konzeptuelle Grundlage aufweisen, die nur aus vektorialen Geodaten bestehen, welche ausschließlich in Form von nichtnormalisierten Relationen vorliegen. Diese Art von Geodaten bieten keine ideale Ausgangsbasis für die Generierung eines LOD-konformen RDF-Graphen. Es wurde davon ausgegangen, dass sich Kolleginnen aus der Praxis mit ähnlichen Problemstellungen konfrontiert sehen könnten, wenn sie ihre Ressourcen in die LOD-Cloud integrieren möchten. Es wurden Lösungen erarbeitet, um solche „Geodaten-Silos“ unter Anwendung der LOD-Methode öffnen zu können, weil dadurch der Wert des Inhalts gesteigert werden kann.

Dafür wurden drei operative Teilziele definiert, nämlich die Anforderungsanalyse und -definition zwecks Bewertung des resultierenden RDF-Graphen aus der Sicht einer Datenbereiterin (1), die Konzeption und praktische Umsetzung eines Workflows zwecks Generierung des LOD-konformen RDF-Graphen mit einem erhöhten Informationsumfang aufgrund zahlreicher RDF-Links auf der Instanzebene (2), die Demonstration des erhöhten Informationsumfangs mithilfe eines IS oder GIS aus der Sicht einer potenziellen Anwenderin (3).

Im Rahmen der Anforderungsanalyse und -definition wurde zunächst festgestellt, dass die spezifische SDWWG-Empfehlung der entsprechenden, allgemeineren W3C-Empfehlung in einer zentralen Frage widerspricht. Erstere Empfehlung zielt auf die Integration von Geodaten bzw. -informationen ins Web ab. Die letztgenannte Empfehlung bezweckt genau dasselbe; sie geht jedoch nicht im Detail auf die speziellen Charakteristika räumlicher Informationen ein. Die zentrale Frage lautet: Ist es zwingend notwendig URI-Kollisionen zu vermeiden? Im Web, das auch ein soziales System ist, sind unterschiedliche Standpunkte – auch zu essentiellen Fragen – wichtig für die Weiterentwicklung. Generell kommt der Ambiguitätstoleranz in sozialen Systemen eine zentrale Bedeutung zu. Die unmissverständliche Kommunikation in einem sozialen System ist auch sicherlich der Idealfall und nicht die Regel. Aber es bedarf einer (Verkehrs-)Sprache und klarer (Kommunikations-)Regeln oder wenigstens eindeutiger Empfehlungen, dass dieser Idealfall theoretisch eintreten könnte. Die richtungweisenden, domänenübergreifenden LOD-Empfehlungen von Heath and Bizer (2011) und die Arbeiten von Presutti and Gangemi (2008) und Halpin and Presutti (2009) enthalten klare Richtlinien zwecks Vermeidung von mehrdeutigen oder widersprüchlichen Aussagen bzw. zur Bewahrung der Kohärenz der LOD-Cloud. Deshalb wurden diese Arbeiten für die Definition der LOD-Anforderungen herangezogen.

Der RDF-Graph, welcher im Rahmen des Workflows generiert wurde, genügt den obligatorischen LOD-Anforderungen. Die zentrale Anforderung LOD-A5 erfüllt er bis zu einem bestimmten Grad. Nach der Transformation der Ausgangsdaten wurde der RDF-Graph im Zuge der Konzeptualisierung ergänzt und schließlich angereichert. Im letzten Teilschritt des Workflows, der prioritär behandelt wurde, konnten insgesamt 4.209 RDF-Links auf der Instanzebene erzeugt werden. Allerdings enthält der RDF-Graph noch isolierte Datenfragmente und entspricht nicht allen optionalen Anforderungen. Dafür konnte die semi-automatische Transformation bzw. der ETL-Prozess und die manuelle Modellierung einfach und vor allem mit geringem Zeitaufwand bewerkstelligt werden. Die vorgestellte Verfahrensweise für die Generierung von RDF-Links auf der Instanzebene war zielführend. Im konkreten Fall gab es keine andere Möglichkeit. Falls möglich, sollte eine andere Verfahrensweise gewählt werden. Im Vergleich zu den vorherigen Teilschritten des Workflows, gestaltete sich der Verlinkungsprozess sehr aufwendig. Trotz hoher Informationskosten war die Verlinkung von Ressourcen effizienter als die manuelle Erfassung sämtlicher Informationen, die in den relevanten Beschreibungen der DNB bereits vorhanden sind. Schon vor der eigentlichen Verlinkung, bei der überblicksmäßigen Darstellung des aktuellen Forschungsstands, wurde die „*owl:sameAs-Problematik*“ angesprochen. Im LOD-Kontext wird die Generierung von Identity Links empfohlen. Sie werden häufig verwendet, um Individuen in Relation zu setzen, welche in unterschiedlichen Informationsbeständen beschrieben werden. Datenbereitstellerinnen können die zugehörigen Beschreibungen genau untersuchen, um diese Links zu generieren. Dies ist jedoch nicht zwingend notwendig, da Identity Links nicht als objektive Aussagen zur Identität, sondern als Behauptungen, Meinungen etc. der jeweiligen Agenten angesehen werden können. Das Problem ist, dass eine potenzielle Anwenderin oder eine Entwicklerin einer LOD-Applikation (häufig) nicht wissen kann, was sich die Datenbereitstellerin bei der Erstellung der RDF-Links gedacht hat. Im Web of Data soll die Bedeutung der *eingesetzten* Prädikate eigentlich nicht von einem speziellen Kontext oder von der individuellen Interpretation einzelner Agenten abhängen. Deshalb muss auch klargestellt werden, dass der Einsatz des Prädikats *rdfs:seeAlso* keine dauerhafte Lösung ist. Es muss weiter eruiert werden, ob sich die Nachnutzung der Konzepte aus der SO lohnt oder ob sich vergleichbare Vokabulare – *sozusagen* „*SKOS-Konstrukte auf der Instanzebene*“ – in der LOD-Cloud etabliert haben.

Es konnte demonstriert werden, dass der generierte RDF-Graph aus der Sicht einer potenziellen Anwenderin einen Mehrwert bietet. Der Informationsumfang konnte durch die Generierung von RDF-Links auf der Instanzebene angemessen erhöht werden. Via SPARQL konnten Antworten auf konkrete Fragen gegeben werden, die sich anhand der ursprünglichen Relationen nicht beantworten lassen. Der RDF-Graph konnte nicht in verschachtelten GML-Strukturen mit Querverbindungen bzw. XLinks abgebildet werden. Ein WFS-Client, der diese adäquat prozessieren kann, konnte nicht ermittelt werden. Für die Bereitstellung der vernetzten Informationen über einen WFS für die Nutzung in GIS müssen zwei Applikationen (weiter-)entwickelt werden.

10.2 Ausblick

Die Bereitstellung der vernetzten Informationen für die praktische Verwendung in anderen Projekten ist erstrebenswert. Die weitere Anreicherung bzw. Verlinkung und die Verbesserung der konzeptuellen Grundlage könnte kooperativ realisiert werden (z.B. im Austausch mit Historikerinnen). Hierzu muss das System produktiv gestellt werden. Vorher müssen die folgenden Aufgaben erledigt werden:

- Die Reorganisation der Tripel über benannte Graphen und deren Beschreibung in Semantic Sitemaps sowie mithilfe von Konzepten aus dem void. Hierzu müssen jedoch weitere Metadaten erfasst werden.
- Die Generierung von HTML-Repräsentationen sowie die Konfiguration des HTTP-Servers. Im Produktivsystem ist der serverseitigen, dynamischen Generierung der 303-URIs der Vorzug zu geben. Hierzu sind die entsprechenden Passagen aus dem XML-Template anzupassen (vgl. Anhang: überall dort, wo im XML-Template Kommata zu finden sind, die im resultierenden XML-Dokument fehlen). Der ETL-Prozess muss erneut durchlaufen werden.
- Der Ausbau des „Mehrwert-Dienstes“, weil die Informationen der DNB in Form einer lokalen Replikation vorliegen. Die Prädikate `mz:hadOwner` und `mz:hadOccupant` könnten invertiert werden. Dann könnten sie im Verlinkungsprozess eingesetzt werden, um von den Ressourcen der DNB auf das historische Häuserbuch zu verlinken. Anschließend können einfachere Recherchen durchgeführt werden (z.B. mit einem beliebigen Namen einer Person).
- Die Berücksichtigung der internen und externen Aktualisierungsprozesse. Intern ist zu prüfen, welche Datenbank und welches Web- oder Desktop-GIS für weitere Datenerfassungen geeignet ist. Die Synchronisation mit den externen Ressourcen via OAI-PMH muss eingerichtet werden.
- Die Weiterentwicklung des WFS-Adapters und des WFS-Clients.
- Die Vereinfachung von Abfragen durch den Einsatz eines geeigneten Dienstes und/oder die Einrichtung einer Webapplikation zwecks RDF-Browsing.

If the semantic web needed a symbol, a good one to use would be a Navaho dream-catcher: a small web, lovingly hand-crafted, [easy] to look at, and rumored to catch dreams; but really more of a symbol than a reality.

Pat Hayes, *Catching the Dreams*, 2003

Literaturverzeichnis

- ABGAZ, Y., DORN, A., PIRINGER, B., WANDL-VOGT, E. & WAY, A. 2018. Semantic Modelling and Publishing of Traditional Data Collection Questionnaires and Answers. *Information*, 9, 297.
- ALLEMANG, D., HENDLER, J. & GANDON, F. 2020. *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*, New York, Association for Computing Machinery.
- ANDRAE, C. 2009. Spatial Schema. ISO 19107 und ISO 19137 vorgestellt und erklärt. In: ANDRAE, C., FITZKE, J. & ZIPF, A. (eds.) *OpenGIS essentials. Die Geo-Standards von OGC und ISO im Überblick*. Heidelberg: Herbert Wichmann Verlag.
- ANDRAE, C. 2013. Simple Features. Praxisnahe Standards für einfache Geoobjekte in Datenbanken und GIS. In: ANDRAE, C., FITZKE, J. & ZIPF, A. (eds.) *OpenGIS essentials. Die Geo-Standards von OGC und ISO im Überblick*. Berlin, Offenbach: Wichmann/VDE Verlag.
- ARENAS, M., BERTAILS, A., PRUD'HOMMEAUX, E. & SEQUEDA, J. 2012. *A Direct Mapping of Relational Data to RDF* [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/rdb-direct-mapping/> [Accessed 01.01.2022].
- ATEMEZING, G. A. 2015. *Publishing and Consuming Geo-spatial and Government Data on the Semantic Web*. Doctoral Thesis, TELECOM ParisTech.
- BAGGIO, R. & ANTONIOLI CORIGLIANO, M. On the Importance of Hyperlinks: A Network Science Approach. 2009 Vienna. Springer Vienna, 309-318.
- BATTLE, R. & KOLAS, D. 2011. Linking Geospatial Data With GeoSPARQL. *Semantic Web journal - Interoperability, Usability, Applicability*, 24, 1-11.
- BATTLE, R. & KOLAS, D. 2012. Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web journal - Interoperability, Usability, Applicability*, 3, 355-370.
- BEIßWENGER, M. & PAPPERT, S. 2018. Internetbasierte Kommunikation. In: LIEDTKE, F. & TUCHEN, A. (eds.) *Handbuch Pragmatik*. Stuttgart: J.B. Metzler.
- BENGEL, G. 2014. *Grundkurs Verteilte Systeme: Grundlagen und Praxis des Client-Server und Distributed Computing*, Wiesbaden, Springer.
- BERNARD, L., FITZKE, J. & WAGNER, R. 2005. *Geodateninfrastrukturen*, Heidelberg, Herbert Wichmann Verlag.
- BERNERS-LEE, T. 1989. *Information Management: A Proposal*. [Online]. Available: <https://www.w3.org/History/1989/proposal.html> [Accessed 01.01.2022].
- BERNERS-LEE, T. 1991. *Design Issues. Naming*. [Online]. Available: <https://www.w3.org/DesignIssues/Naming> [Accessed 01.01.2022].
- BERNERS-LEE, T. 1994a. *Universal Resource Identifiers in WWW. A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web*. [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 1630. Available: <https://www.rfc-editor.org/rfc/rfc1630.txt> [Accessed 01.01.2022].
- BERNERS-LEE, T. 1994b. *World wide web future directions. Plenary Talk*. [Online]. Available: <https://www.w3.org/Talks/WWW94Tim/> [Accessed 01.01.2022].

- BERNERS-LEE, T. 1996. *Design Issues. Architectural and philosophical points*. [Online]. Available: <https://www.w3.org/DesignIssues/> [Accessed 01.01.2022].
- BERNERS-LEE, T. 1998. *What the Semantic Web can represent*. [Online]. Informal Draft. Available: <https://www.w3.org/DesignIssues/RDFnot.html> [Accessed 01.01.2022].
- BERNERS-LEE, T. 2006. *Design Issues. Linked Data*. [Online]. Available: <https://www.w3.org/DesignIssues/LinkedData.html> [Accessed 01.01.2022].
- BERNERS-LEE, T., FIELDING, R. & FRYSTYK, H. 1996. *Hypertext Transfer Protocol -- HTTP/1.0* [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 1945. Available: <https://rfc-editor.org/rfc/rfc1945.txt> [Accessed 01.01.2022].
- BERNERS-LEE, T., FIELDING, R. & MASINTER, L. 1998. *Uniform Resource Identifiers (URI): Generic Syntax*. [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 2396. Available: <https://www.rfc-editor.org/rfc/rfc2396.txt> [Accessed 01.01.2022].
- BERNERS-LEE, T., FIELDING, R. & MASINTER, L. 2005. *Uniform Resource Identifier (URI): Generic Syntax*. [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 3986. Available: <https://www.rfc-editor.org/rfc/rfc3986.txt> [Accessed 01.01.2022].
- BERNERS-LEE, T. & FISCHETTI, M. 1999. *Weaving the Web : the original design and ultimate destiny of the World Wide Web by its inventor*, San Francisco, HarperSanFrancisco.
- BERNERS-LEE, T., HENDLER, J. & LASSILA, O. 2001. The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284.
- BERNERS-LEE, T., MASINTER, L. & MCCAHERILL, M. 1994. *Uniform Resource Locators (URL)*. [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 1738. Available: <https://www.rfc-editor.org/rfc/rfc1738.txt> [Accessed 01.01.2022].
- BERNERS-LEE, T. & MENDELSON, N. 2006. *The Rule of Least Power*. [Online]. World Wide Web Consortium (W3C) - W3C Technical Architecture Group (TAG). TAG Finding. Available: <https://www.w3.org/2001/tag/doc/leastPower.html> [Accessed 01.01.2022].
- BILL, R. 2016. *Grundlagen der Geo-Informationssysteme*, Berlin, Offenbach, Wichmann/VDE Verlag.
- BIZER, C., HEATH, T. & BERNERS-LEE, T. 2009. Linked Data: The Story so Far. *International Journal on Semantic Web and Information Systems*, 5, 1-22.
- BORGO, S. & MASOLO, C. 2010. Ontological foundations of DOLCE. *Theory and applications of ontology: Computer applications*. Springer.
- BRACHMAN, R. J. & SMITH, B. C. 1980. Special issue on knowledge representation. *SIGART Bull.*, 1-138.
- BRAY, T., PAOLI, J., SPERBERG-MCQUEEN, C. M., MALER, E. & YERGEAU, F. 2008. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/xml/> [Accessed 01.01.2022].

- BRICKLEY, D. 2003. *Basic Geo (WGS84 lat/long) Vocabulary*. [Online]. World Wide Web Consortium (W3C). W3C Semantic Web Interest Group. Available: <https://www.w3.org/2003/01/geo/> [Accessed 01.01.2022].
- BRICKLEY, D. & GUHA, R. V. 2014. *RDF Schema 1.1*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/rdf-schema/> [Accessed 01.01.2022].
- BRIN, S. & PAGE, L. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Seventh International World-Wide Web Conference (WWW 1998)*. Brisbane, Australia.
- BUSCHBAUM, C. 2002. *Mainz auf dem Weg zur kurfürstlichen Residenzstadt im Spiegel der Mainzer Stadtaufnahmen*. In: Matheus, M. & Rödel W. G.: Bausteine zur Mainzer Stadtgeschichte. Mainzer Kolloquium 2000. Geschichtliche Landeskunde (Bd. 55). Stuttgart.
- CONNOLLY, D. W. 2002a. *An Evaluation of the World Wide Web with respect to Engelbart's Requirements*. [Online]. Informal Draft. Available: <https://www.w3.org/Architecture/NOTE-ioh-arch> [Accessed 01.01.2022].
- CONNOLLY, D. W. 2002b. *on media types for OWL (5.13)*. [Online]. Public mailing list. Available: <https://lists.w3.org/Archives/Public/www-webont-wg/2002Oct/0162.html> [Accessed 01.01.2022].
- COX, S., CUTHBERT, A., LAKE, R. & MARTELL, R. 2002. *OpenGIS® Geography Markup Language (GML) Implementation Specification. Version 2.1.1*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 02-009. Available: https://portal.ogc.org/files/?artifact_id=1108 [Accessed 01.01.2022].
- COX, S. & LITTLE, C. 2020. *Time Ontology in OWL*. [Online]. World Wide Web Consortium (W3C). W3C Candidate Recommendation. Available: <https://www.w3.org/TR/owl-time/> [Accessed 01.01.2022].
- CYGANIAK, R., WOOD, D. & LANTHALER, M. 2014. *RDF 1.1 Concepts and Abstract Syntax*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/rdf11-concepts/> [Accessed 01.01.2022].
- DAMOVA, M., KIRYAKOV, A., SIMOV, K. & PETROV, S. 2010. Mapping the central LOD ontologies to PROTON upper-level ontology. *Ontology Matching*, 61.
- DAS, S., SUNDARA, S. & CYGANIAK, R. 2012. *R2RML: RDB to RDF Mapping Language* [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/r2rml/> [Accessed 01.01.2022].
- DENECKE, D. 1987. *Sozialtopographie der mittelalterlichen Stadt Göttingen*. Denecke, D. & Kühn, H.: Göttingen. Geschichte einer Universitätsstadt. Von den Anfängen bis zum Ende des Dreißigjährigen Krieges (Bd.1), Göttingen, Vandenhoeck & Ruprecht. .
- DEROSE, S., MALER, E., ORCHARD, D. & WALSH, N. 2010. *XML Linking Language (XLink) Version 1.1* [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/xlink11/> [Accessed 01.01.2022].
- DIEDERICHS, K. & WUTTKE, U. 2018. *Das Open Archives Initiative Protocol for Metadata Harvesting: Zielsetzung, Funktionalität, Einsatzgebiete*. DOI: 10.5281/zenodo.1253735
- EGENHOFER, M. J. 2002. Toward the semantic geospatial web. *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*. McLean, Virginia, USA: Association for Computing Machinery.

- FALCK, L. 1972. *Mainz um 1620. Historischer Stadtplan mit alphabetischem Index*. In: Brück, A. P. & Falck, L.: Mainz vom Verlust der Stadtfreiheit bis zum Ende des Dreißigjährigen Krieges (1462–1648). Geschichte der Stadt Mainz (Bd. 5). Düsseldorf.
- FALCK, L. 1978. *Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts*. In: Geschichtlicher Atlas von Hessen / Hess. Landesamt für Geschichtl. Landeskunde. Begr. und vorbereitet von Edmund E. Stengel. Bearb. von Friedrich Uhlhorn. Karte 34 C. Marburg/Lahn.
- FEIGENBAUM, L., WILLIAMS, G. T., CLARK, K. G. & TORRES, E. 2013. *SPARQL 1.1 Protocol*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/sparql11-protocol/> [Accessed 01.01.2022].
- FIELDING, R. 2000. *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral dissertation, University of California, Irvine.
- FIELDING, R., GETTYS, J., MOGUL, J., FRYSTYK, H., MASINTER, L., LEACH, P. & BERNERS-LEE, T. 1999. *Hypertext Transfer Protocol -- HTTP/1.1* [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 2616. Available: <https://www.rfc-editor.org/rfc/rfc2616.txt> [Accessed 01.01.2022].
- FLORCZYK, A. J., MAUÉ, P., LÓPEZ-PELLICER, F. J. & NOGUERAS-ISO, J. Finding OGC Web Services in the Digital Earth. Proceedings of the Workshop Towards Digital Earth: Search, Discover and Share Geospatial Data at Future Internet Symposium, Berlin, Germany, 2010.
- GANDON, F. & SCHREIBER, G. 2014. *RDF 1.1 XML Syntax*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/rdf-syntax-grammar/> [Accessed 01.01.2022].
- GANGEMI, A. & PRESUTTI, V. 2009. Ontology Design Patterns. In: STAAB, S. & STUDER, R. (eds.) *Handbook on Ontologies*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- GRUBER, T. R. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5, 199-220.
- GUARINO, N. 1998. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, IOS press.
- GUARINO, N., OBERLE, D. & STAAB, S. 2009. What is an ontology? *Handbook on ontologies*. Springer.
- GUO, H., LIU, Z., JIANG, H., WANG, C., LIU, J. & LIANG, D. 2017. Big Earth Data: a new challenge and opportunity for Digital Earth's development. *International Journal of Digital Earth*, 10, 1-12.
- HALASZ, F., SCHWARTZ, M., GRØNBÆK, K. & TRIGG, R. H. 1994. The Dexter hypertext reference model. *Commun. ACM*, 37, 30–39.
- HALPIN, H. 2013. *Social Semantics - The Search for Meaning on the Web*. Boston, MA: Springer US.
- HALPIN, H., HAYES, P. J., MCCUSKER, J. P., MCGUINNESS, D. L. & THOMPSON, H. S. When owl:sameAs Isn't the Same: An Analysis of Identity in Linked Data. 2010 Berlin, Heidelberg. Springer Berlin Heidelberg, 305-320.
- HALPIN, H., HAYES, P. J. & THOMPSON, H. S. 2015. When owl:sameAs isn't the Same Redux: Towards a Theory of Identity, Context, and Inference on the Semantic Web. Cham: Cham: Springer International Publishing.

- HALPIN, H. & PRESUTTI, V. An Ontology of Resources: Solving the Identity Crisis. 2009 Berlin, Heidelberg. Springer Berlin Heidelberg, 521-534.
- HAROLD, E. R. & MEANS, W. S. 2004. *XML in a Nutshell: A Desktop Quick Reference*, O'Reilly Media.
- HARRIS, S. & SEABORNE, A. 2013. *SPARQL 1.1 Query Language*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/sparql11-query/> [Accessed 01.01.2022].
- HAYES, P. J. & HALPIN, H. 2008. In Defense of Ambiguity. *International Journal on Semantic Web and Information Systems*, 4, 1-18.
- HAYES, P. J. & PATEL-SCHNEIDER, P. F. 2014. *RDF 1.1 Semantics*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/2014/REC-rdf11-mt-20140225/> [Accessed 01.01.2022].
- HBEICH, E. & ROXIN, A. 2020. Linking BIM and GIS Standard Ontologies with Linked Data. *Proceedings of the Linked Data in Architecture and Construction, Dublin, Ireland*, 17-19.
- HEATH, T. & BIZER, C. 2011. *Linked Data: Evolving the Web into a Global Data Space*. [Online]. Available: <http://linkeddatabook.com/editions/1.0/> [Accessed 01.01.2022].
- HESSE, W., MERBETH, G. & FRÖHLICH, R. 2020. *Software-Entwicklung: Vorgehensmodelle, Projektführung, Produktverwaltung*. Berlin, Boston: De Gruyter Oldenbourg.
- HIEBEL, G., DOERR, M., EIDE, Ø. & THEODORIDOU, M. 2015. CRMgeo: a Spatiotemporal Model An Extension of CIDOC-CRM to link the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement. *Unpublished Report (Version 1.2)*.
- HITZLER, P., KRÖTZSCH, M., RUDOLPH, S. & SURE, Y. 2008. *Semantic Web: Grundlagen*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- HOBONA, G. & BRACKIN, R. 2015. *Testbed-11 Use of Semantic Linked Data with RDF for National Map NHD and Gazetteer Data Engineering Report*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 15-066r1. Available: <http://www.opengis.net/doc/PER/t11-linked-data-nhd> [Accessed 01.01.2022].
- HOMBURG, T. 2020. Towards a linked-data powered, INSPIRE-based national data infrastructure – The case of the Federal Agency for Cartography Germany. *International Journal of Spatial Data Infrastructures Research*, 15.
- ISELE, R. 2013. *Learning Expressive Linkage Rules for Entity Matching using Genetic Programming*. Doctoral Thesis, Universität Mannheim.
- ISO/TC 211 2002. *ISO 19101:2002 Geographic Information – Reference model*. Genf: International Organization for Standardization (ISO).
- ISO/TC 211 2003. *ISO 19107:2003 Geographic Information – Spatial Schema*. Genf: International Organization for Standardization (ISO).
- ISO/TC 211 2004. *ISO 19125-1:2004 Geographic information – Simple feature access – Part 1: Common architecture*. Genf: International Organization for Standardization (ISO).
- ISO/TC 211 2005. *ISO 19109:2005 Geographic information – Rules for application schema*. Genf: International Organization for Standardization (ISO).

- IVÁNOVÁ, I., SIAO HIM FA, J., MCMEEKIN, D. A., ARNOLD, L. M., DEAKIN, R. & WILSON, M. 2020. From spatial data to spatial knowledge infrastructure: A proposed architecture. *Transactions in GIS*, 24, 1526-1558.
- JACOBS, I. & WALSH, N. 2004. *Architecture of the World Wide Web, Volume One*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/webarch/> [Accessed 01.01.2022].
- JAIN, P., HITZLER, P., SHETH, A. P., VERMA, K. & YEH, P. Z. Ontology Alignment for Linked Open Data. 2010 Berlin, Heidelberg. Springer Berlin Heidelberg, 402-417.
- JAIN, P., YEH, P. Z., VERMA, K., VASQUEZ, R. G., DAMOVA, M., HITZLER, P. & SHETH, A. P. Contextual ontology alignment of lod with an upper ontology: A case study with proton. Extended Semantic Web Conference, 2011. Berlin, Heidelberg: Springer, 80-92.
- JONES, J., KUHN, W., KEßLER, C. & SCHEIDER, S. 2014. Making the Web of Data Available Via Web Feature Services. In: HUERTA, J., SCHADE, S. & GRANELL, C. (eds.) *Connecting a Digital Europe Through Location and Place*. Cham: Springer International Publishing.
- JOVANOVIK, M., HOMBURG, T. & SPASIĆ, M. 2021. A GeoSPARQL Compliance Benchmark. *ISPRS International Journal of Geo-Information*, 10, 487.
- KLIEN, E. 2007. A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transactions in GIS*, 11, 437-452.
- KNUBLAUCH, H. & KONTOKOSTAS, D. 2017. *Shapes Constraint Language (SHACL)*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/shacl/> [Accessed 01.01.2022].
- KOKLA, M. & GUILBERT, E. 2020. A Review of Geospatial Semantic Information Modeling and Elicitation Approaches. *ISPRS International Journal of Geo-Information*, 9, 146.
- KÖNIG, C. 2018. Mit einem Bücherhaufen fing es an... Geschichte der Universitätsbibliothek Mainz. In: Forschungsverbund Universitätsgeschichte der Johannes Gutenberg-Universität Mainz (ed.) *Beiträge zur Geschichte der Universität Mainz: Neue Folge. Band 13*. Göttingen: Vandenhoeck & Ruprecht (V&R) unipress GmbH, Mainz University Press.
- KOTTMAN, C. & REED, C. 2009. *The OpenGIS® Abstract Specification. Topic 5: Features*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 08-126. Available: https://portal.ogc.org/files/?artifact_id=29536 [Accessed 01.01.2022].
- KOUBARAKIS, M. & KYZIRAKOS, K. Modeling and Querying Metadata in the Semantic Sensor Web: The Model stRDF and the Query Language stSPARQL. 2010 Berlin, Heidelberg. Springer Berlin Heidelberg, 425-439.
- KUREK, C. 2011. *Ein Vergleich der Methoden automatisierter Dublettenerkennung*. Diplomarbeit, Fachbereich Informatik, Arbeitsbereich Softwaretechnik, Universität Hamburg.
- KYZIRAKOS, K., SAVVA, D., VLACHOPOULOS, I., VASILEIOU, A., KARALIS, N., KOUBARAKIS, M. & MANEGOLD, S. 2018. GeoTriples: Transforming geospatial data into RDF graphs using R2RML and RML mappings. *Journal of Web Semantics*, 52-53, 16-32.
- KYZIRAKOS, K., VLACHOPOULOS, I., SAVVA, D., MANEGOLD, S. & KOUBARAKIS, M. GeoTriples: a Tool for Publishing Geospatial Data as RDF Graphs Using R2RML Mappings. TC/SSN@ ISWC, 2014. 33-44.

- LAKE, R. & CUTHBERT, A. 2000. *Geography Markup Language (GML). Version 1.0*. [Online]. OpenGIS® Consortium (exOGC). exOGC Recommendation Paper. Document Number: 00-029. Available: https://portal.ogc.org/files/?artifact_id=7197 [Accessed 01.01.2022].
- LANGE, C. & AUER, S. 2014. Linking Data and Knowledge in Enterprises, Research and Society. *Proceedings of the 11th International Baltic Conference on Databases and Information Systems (Baltic DB&IS)*. Tallinn: Tallinn University of Technology Press.
- LEWIS, R. 2007. *Dereferencing HTTP URIs*. [Online]. World Wide Web Consortium (W3C) - W3C Technical Architecture Group (TAG). Draft TAG Finding. Available: <https://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14> [Accessed 01.01.2022].
- LIEBERMAN, J. 2006. *Geospatial Semantic Web Interoperability Experiment Report*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 06-002R1. Available: https://portal.opengeospatial.org/files/?artifact_id=15198 [Accessed 01.01.2022].
- LÓPEZ-PELLICER, F. J., VILCHES-BLÁZQUEZ, L. M., ZARAZAGA-SORIA, F. J., MURO-MEDRANO, P. R. & CORCHO, O. 2012. The Delft Report: Linked Data and the challenges for geographic information standardization. *Revista Catalana de Geografia vol. XVII, n° 44*.
- LÓSCIO, B. F., BURLE, C. & CALEGARI, N. 2017. *Data on the Web Best Practices* [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/dwbp/> [Accessed 01.01.2022].
- MAUÉ, P. 2009. *Semantic annotations in OGC standards*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 08-167r1. Available: https://portal.ogc.org/files/?artifact_id=34916 [Accessed 01.01.2022].
- MENDELSON, N. 2007. *The Self-Describing Web*. [Online]. World Wide Web Consortium (W3C) - W3C Technical Architecture Group (TAG). Draft TAG Finding. Available: <https://www.w3.org/2001/tag/doc/selfDescribingDocuments-2007-05-24.html> [Accessed 01.01.2022].
- MOCKAPETRIS, P. 1987. *DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION*. [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 1035. Available: <https://www.rfc-editor.org/rfc/rfc1035.txt> [Accessed 01.01.2022].
- MOTIK, B., PATEL-SCHNEIDER, P. F. & PARSIA, B. 2012. *OWL 2 Web Ontology Language. Structural Specification and Functional-Style Syntax (Second Edition)*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/owl2-syntax/> [Accessed 01.01.2022].
- MÜNKEL, J. 2019. *THEATRUM MOGONTIACENSIVM*. Master Thesis, Fakultät für Architektur, RWTH Aachen University.
- NISHANBAEV, I., CHAMPION, E. & MCMEEKIN, D. A. 2019. A Survey of Geospatial Semantic Web for Cultural Heritage. *Heritage*, 2, 1471-1498.
- NISHANBAEV, I., CHAMPION, E. & MCMEEKIN, D. A. 2020. A Comparative Evaluation of Geospatial Semantic Web Frameworks for Cultural Heritage. *Heritage*, 3, 875-890.
- NYS, G.-A., VAN RUYMBEKE, M. & BILLEN, R. Spatio-temporal reasoning in CIDOC CRM: an hybrid ontology with GeoSPARQL and OWL-Time. CEUR Workshop Proceedings, 2018. RWTH Aachen University.

- OTTERMANN, A. 2014. Dionysius und Jacobus Campius als Buchbesitzer – eine Momentaufnahme (nicht nur) mit Blick auf die Mainzer Karmeliten. In: HAUG, C. & THIELE, R. (eds.) *Sonderdruck aus: Buch – Bibliothek – Region*. Wiesbaden: Harrassowitz Verlag.
- PATROUMPAS, K., ALEXAKIS, M., GIANNOPOULOS, G. & ATHANASIOU, S. TripleGeo: an ETL Tool for Transforming Geospatial Data into RDF Triples. LWD, 2014. 275-278.
- PATROUMPAS, K., GEORGOMANOLIS, N., STRATIOTIS, T., ALEXAKIS, M. & ATHANASIOU, S. 2015. Exposing INSPIRE on the Semantic Web. *Journal of Web Semantics*, 35, 53-62.
- PERRY, M. 2008. *A Framework to Support Spatial, Temporal and Thematic Analytics over Semantic Web Data*. Doctoral Thesis, Department of Computer Science and Engineering, Wright State University.
- PERRY, M. & HERRING, J. 2012. *OGC GeoSPARQL - A Geographic Query Language for RDF Data. Version: 1.0*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 11-052r4. Available: https://portal.ogc.org/files/?artifact_id=47664 [Accessed 01.01.2022].
- PERRY, M., JAIN, P. & SHETH, A. P. 2011. SPARQL-ST: Extending SPARQL to Support Spatiotemporal Queries. In: ASHISH, N. & SHETH, A. P. (eds.) *Geospatial Semantics and the Semantic Web: Foundations, Algorithms, and Applications*. Boston, MA: Springer US.
- PIERCE, J. R. 1980. *An Introduction to Information Theory: Symbols, Signals & Noise*, New York, Dover Publications, Inc.
- POHL, A. & DANOWSKI, P. 2013. Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick. In: PATRICK, D. & ADRIAN, P. (eds.) *(Open) Linked Data in Bibliotheken*. De Gruyter Saur.
- PORTELE, C. 2007. *OpenGIS® Geography Markup Language (GML) Encoding Standard. Version: 3.2.1*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 07-036, aka International Standard ISO 19136:2007. Available: https://portal.ogc.org/files/?artifact_id=20509 [Accessed 01.01.2022].
- PORTELE, C. 2012. *OGC® Geography Markup Language (GML) - Extended schemas and encoding rules. Version: 3.3.0*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 10-129r1. Available: https://portal.ogc.org/files/?artifact_id=46568 [Accessed 01.01.2022].
- PORTELE, C. & VRETANOS, P. 2019. *OGC Web Feature Service 3.0 - Part 1: Core. Version: 3.0.0-SNAPSHOT (Editor's draft)*. [Online]. Open Geospatial Consortium Inc. (OGC). Internal OGC Ref.-No. 17-069. Available: <http://docs.opengeospatial.org/DRAFTS/17-069.html> [Accessed 01.01.2022].
- PORTELE, C., VRETANOS, P. & HEAZEL, C. 2019. *OGC API - Features - Part 1: Core. Version: 1.0*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 17-069r3. Available: <https://docs.opengeospatial.org/is/17-069r3/17-069r3.html> [Accessed 01.01.2022].
- PRESUTTI, V. & GANGEMI, A. 2008. Identity of Resources and Entities on the Web. *Int. J. Semantic Web Inf. Syst.*, 4, 49-72.
- RAGGETT, D., HORS, A. L. & JACOBS, I. 1999. *HTML 4.01 Specification*. [Online]. World Wide Web Consortium (W3C). W3C Superseded Recommendation. Available: <https://www.w3.org/TR/REC-html40/> [Accessed 01.01.2022].
- ROHWEDER, J. P., KASTEN, G., MALZAHN, D., PIRO, A. & SCHMID, J. 2021. Informationsqualität – Definitionen, Dimensionen und Begriffe. In: HILDEBRAND, K., GEBAUER, M. & MIELKE, M. (eds.) *Daten- und Informationsqualität: Die Grundlage der Digitalisierung*. Wiesbaden: Springer Fachmedien Wiesbaden.

- RYDEN, K. 2005. *OpenGIS® Implementation Specification for Geographic information - Simple feature access - Part 1: Common architecture*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 05-126, aka International Standard ISO 19125-1:2004. Available: https://portal.ogc.org/files/?artifact_id=13227 [Accessed 01.01.2022].
- SALK, C. F., STURN, T., SEE, L., FRITZ, S. & PERGER, C. 2016. Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game. *International Journal of Digital Earth*, 9, 410-426.
- SAUERMANN, L. & CYGANIAK, R. 2008. *Cool URIs for the Semantic Web*. [Online]. World Wide Web Consortium (W3C). W3C Interest Group Note. Available: <https://www.w3.org/TR/cooluris/> [Accessed 01.01.2022].
- SCHADE, S. & COX, S. Linked Data in SDI or How GML is not about Trees. 13th AGILE Conference on Geographic Information Science, 2010 Guimarães, Portugal.
- SCHEIDER, S., JONES, J., SÁNCHEZ, A. & KEBLER, C. 2014. Encoding and Querying Historic Map Content. In: HUERTA, J., SCHADE, S. & GRANELL, C. (eds.) *Connecting a Digital Europe Through Location and Place*. Cham: Springer International Publishing.
- SCHREIBER, G. & RAIMOND, Y. 2014. *RDF 1.1 Primer*. [Online]. World Wide Web Consortium (W3C). W3C Recommendation. Available: <https://www.w3.org/TR/rdf11-primer/> [Accessed 01.01.2022].
- SCHROHE, H. (ed.) 1930. *Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts*, Mainz: Wilckens.
- SCHWENK, J. 2010. Kryptographie im Internet. *Sicherheit und Kryptographie im Internet: Von sicherer E-Mail bis zu IP-Verschlüsselung*. Wiesbaden: Vieweg+Teubner.
- SEIFERT, M. 2008. *Wissenschaftlicher Beitrag für den Aufbau einer Geodateninfrastruktur zur Lösung von Aufgaben des E-Government*. Doctoral Thesis, ETH Zürich.
- SOLLINS, K. & MASINTER, L. 1994. *Functional Requirements for Uniform Resource Names*. [Online]. Internet Engineering Task Force (IETF) - Network Working Group. RFC 1737. Available: <https://www.rfc-editor.org/rfc/rfc1737.txt> [Accessed 01.01.2022].
- STAUB, P. 2009. *Über das Potenzial und die Grenzen der semantischen Interoperabilität von Geodaten: ein operationelles Verfahren zur Nutzung verteilter Systeme in Geodaten-Infrastrukturen*. Doctoral Thesis, ETH Zürich.
- STEINMETZ, R. 1999. Dokumente, Hypertext und Hypermedia. *Multimedia-Technologie: Grundlagen, Komponenten und Systeme*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- STOCK, K. & GUESGEN, H. 2016. Chapter 10 - Geospatial Reasoning With Open Data. In: LAYTON, R. & WATTERS, P. A. (eds.) *Automating Open Source Intelligence*. Boston: Syngress.
- TANDY, J., VAN DEN BRINK, L. & BARNAGHI, P. 2017. *Spatial Data on the Web Best Practices*. [Online]. World Wide Web Consortium (W3C). Open Geospatial Consortium Inc. (OGC). W3C Working Group Note. OGC Ref.-No. 15-107. Available: <https://www.w3.org/TR/sdw-bp/> [Accessed 01.01.2022].
- TILKOV, S., EIGENBRODT, M., SCHREIER, S. & WOLF, O. 2015. *REST und HTTP: Entwicklung und Integration nach dem Architekturstil des Web*, Heidelberg, dpunkt.
- TSCHIRNER, S. 2012. *Semantischer Zugriff auf INSPIRE. Verteilte Suche und Publikation von GML-Daten im Semantic Web am Beispiel INSPIRE*. Master Thesis, Zentrum für GeoInformatik (Z_GIS), Paris Lodron-Universität Salzburg.

- VAN DEN BRINK, L., JANSSEN, P., QUAK, W. & STOTER, J. E. 2014. Linking spatial data: Automated conversion of geo-information models and GML data to RDF. *International journal of spatial data infrastructures research*, 9.
- VILCHES-BLÁZQUEZ, L. M. & SAAVEDRA, J. 2019. A framework for connecting two interoperability universes: OGC Web Feature Services and Linked Data. *Transactions in GIS*, 23, 22-47.
- VILCHES-BLÁZQUEZ, L. M., VILLAZÓN-TERRAZAS, B., SAQUICELA, V., LEÓN, A. D., CORCHO, O. & GÓMEZ-PÉREZ, A. 2010. GeoLinked data and INSPIRE through an application case. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. San Jose, California: Association for Computing Machinery.
- VOLZ, J., BIZER, C., KOBILAROV, G. & GAEDKE, M. Silk – A Link Discovery Framework for the Web of Data. 18th International World Wide Web Conference, 2009.
- VRETANOS, P. 2001. *Filter Encoding Implementation Specification. Version: 1.0.0*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 02-059. Available: https://portal.ogc.org/files/?artifact_id=1171 [Accessed 01.01.2022].
- VRETANOS, P. 2002. *Web Feature Service Implementation Specification. Version: 1.0.0*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 02-058. Available: https://portal.ogc.org/files/?artifact_id=7176 [Accessed 01.01.2022].
- VRETANOS, P. 2005. *Web Feature Service Implementation Specification. Version: 1.1.0*. [Online]. Open Geospatial Consortium Inc. (OGC). OGC Ref.-No. 04-094. Available: https://portal.ogc.org/files/?artifact_id=8339 [Accessed 01.01.2022].
- WANG, R. Y. & STRONG, D. M. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12, 5-33.
- WILLIAMSON, I. P., RAJABIFARD, A. & FEENEY, M.-E. F. 2003. *Developing Spatial Data Infrastructures. From Concept to Reality*, London, CRC Press.
- WÜRRIEHAUSEN, F. 2015. *Semantische und organisatorische Interoperabilität kommunaler Geodaten im Kontext von INSPIRE*. Doctoral Thesis, Agrar-und Umweltwissenschaftliche Fakultät, Universität Rostock.
- ZEHNDER, C. A. 1998. *Leben mit Information. Informationssysteme und Datenbanken*. Wiesbaden: Vieweg+Teubner Verlag.
- ZHANG, C., ZHAO, T. & LI, W. 2015. Geospatial Data Interoperability, Geography Markup Language (GML), Scalable Vector Graphics (SVG), and Geospatial Web Services. *Geospatial Semantic Web*. Cham: Springer International Publishing.


```

<!--Quellen-->
<rdf:Description rdf:about="http://www.mainz.de/id/S1">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S1"/>
  <foaf:name xml:lang="de">Herrmann: Quellen zur Topographie und Statistik der Stadt Mainz (1914)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/580166953"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S1">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S1"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S1"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S2">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S2"/>
  <foaf:name xml:lang="de">Dertsch: Die Urkunden des Stadtarchivs Mainz Teil 1 (1962)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/458493481"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S2">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S2"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S2"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S3">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S3"/>
  <foaf:name xml:lang="de">Falck: Mainz vom Verlust der Stadtfreiheit bis zum Ende des Dreißigjährigen Krieges (1972)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/730124045"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S3">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S3"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S3"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S4">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S4"/>
  <foaf:name xml:lang="de">Mayer/Steffens: Die spätmittelalterlichen Urbare des Heiliggeistspitals in Mainz (1992)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/921274386"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S4">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S4"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S4"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S5">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S5"/>
  <foaf:name xml:lang="de">Kissel: Alt-Mainzer Häuser (1909)</foaf:name>
</rdf:Description>

```

```

<rdf:Description rdf:about="http://www.mainz.de/doc/S5">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S5"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S5"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S6">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S6"/>
  <foaf:name xml:lang="de">Schenk zu Schweinsberg; Das Geschlecht der Gensfleisch. Festschrift zum fünfhundertjährigen Geburtstag von
    Johann Gutenberg. (1900)</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S6">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S6"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S6"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S7">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S7"/>
  <foaf:name xml:lang="de">Schneider; Magenza. Topographie der jüdischen Siedlung während des Mittelalters. Masterarbeit
    Universität Trier (2016)</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S7">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S7"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S7"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S8">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S8"/>
  <foaf:name xml:lang="de">Neeb; Verzeichnis der Kunstdenkmäler der Stadt Mainz. Bd. 1: Privatbesitz (1905)</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S8">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S8"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S8"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S9">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S9"/>
  <foaf:name xml:lang="de">Schrohe; Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts / Teil 1 (1930)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/363167978" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S9">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S9"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S9"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S10">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S10"/>
  <foaf:name xml:lang="de">Falk; Zur alten Topographie von Mainz (1899)</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S10">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S10"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S10"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S11">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S11"/>
  <foaf:name xml:lang="de">Dentsch; Die Urkunden des Stadtarchivs Mainz Teil 2 (1963)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/458493503" />
</rdf:Description>

```

```

<rdf:Description rdf:about="http://www.mainz.de/doc/S11">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S11" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S11" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S12">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S12" />
  <foaf:name xml:lang="de">Schrohe: Mainz in seinen Beziehungen zu den deutschen Königen und den Erzbischöfen der Stadt bis zum
    Untergang der Stadtfreiheit (1915)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/362677409" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S12">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S12" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S12" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S13">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S13" />
  <foaf:name xml:lang="de">Schrohe: Aufsätze und Nachweise zur Mainzer Kunstgeschichte (1912)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/362677387" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S13">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S13" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S13" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S14">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S14" />
  <foaf:name xml:lang="de">Frank: Das Totenbuch des Mainzer Dominikanerklosters (1993)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/930437829" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S14">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S14" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S14" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S15">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S15" />
  <foaf:name xml:lang="de">Ochs: Gutenberg und seine frunde (2014)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/1059715686" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S15">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S15" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S15" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S16">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S16" />
  <foaf:name xml:lang="de">Schrohe: Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts / Teil 2 (1930)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/363167986" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S16">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S16" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S16" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>

```

```

<rdf:Description rdf:about="http://www.mainz.de/id/S17">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S17" />
  <foaf:name xml:lang="de">Zülch: Frankfurter Künstler (1935)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/363154574" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S17">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S17" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S17" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S18">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S18" />
  <foaf:name xml:lang="de">Schmid: Die Abtei St. Alban vor Mainz im hohen und späten Mittelalter (1996)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/952642328" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S18">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S18" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S18" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S19">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S19" />
  <foaf:name xml:lang="de">Falck: Mainz vom frühen Mittelalter bis zum Anfang des 17. Jahrhunderts (1978)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/800741714" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S19">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S19" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S19" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S20">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S20" />
  <foaf:name xml:lang="de">Wünderwein: Nova subsidia diplomatia ad selecta juris ecclesiastici Germaniae et historiarum capita elucidanda ex originalibus et authenticis documentis (1784)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/1185936173" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S20">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S20" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S20" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S21">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S21" />
  <foaf:name xml:lang="de">Stephan: Das Bürgerhaus in Mainz (1974)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/740554085" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S21">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S21" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S21" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S22">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S22" />
  <foaf:name xml:lang="de">Schrohe: Das Mainzer Geschlecht zum Jungen in Diensten des deutschen Königums und der Stadt Mainz (1933)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/362677425" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S22">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S22" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S22" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>

```



```
<rdf:Description rdf:about="http://www.mainz.de/id/S23">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S23" />
  <foaf:name xml:lang="de">Zülch: Frankfurter Urkundenbuch zur Frühgeschichte des Buchdrucks (1920)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/578517760" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S23">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S23" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S23" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S24">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S24" />
  <foaf:name xml:lang="de">Dentsch: Die Urkunden des Stadtarchivs Mainz Teil 3 (1965)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/458493511" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S24">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S24" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S24" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S25">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S25" />
  <foaf:name xml:lang="de">Dobras: Münzerhausgenossen und andere Geschlechter. Mainzer Zeitschrift 94/95 (1999/2000)</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S25">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S25" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S25" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S26">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S26" />
  <foaf:name xml:lang="de">Schrohe: Die Stadt Mainz unter kurfürstlicher Verwaltung (1920)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/362677417" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S26">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S26" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S26" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S27">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S27" />
  <foaf:name xml:lang="de">Dentsch: Die Urkunden des Stadtarchivs Mainz Teil 4 (1967)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/458493538" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S27">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S27" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S27" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S28">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S28" />
  <foaf:name xml:lang="de">Kläger: Die Mainzer Stadt- und Festungserweiterung (1988)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/890620601" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S28">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S28" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S28" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
```

```

<rdf:Description rdf:about="http://www.mainz.de/id/S29">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S29" />
  <foaf:name xml:lang="de">Dael: Antiquarische Wanderungen im Innern der Stadt Mainz (Ost/Jahr unbekannt)</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S29">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S29" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S29" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S30">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S30" />
  <foaf:name xml:lang="de">Schrohe: Die Mainzer Stadtaufnahmen des 16. bis 18. Jahrhunderts / Teil 3 (1931)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/363167994" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S30">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S30" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S30" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S31">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S31" />
  <foaf:name xml:lang="de">Kahlenberg: Kurmainzische Verteidigungseinrichtungen und Baugeschichte der Festung Mainz im 17. und 18.
  Jahrhundert (1963)</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/452291402" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S31">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S31" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S31" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/S32">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S32" />
  <foaf:name xml:lang="de">Arens: Ein gotisches Patrizierhaus in Mainz. Kultur und Wirtschaft im rheinischen Raum. (1949) </foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S32">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/S32" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S32" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<!--Agenten-->
<rdf:Description rdf:about="http://www.mainz.de/id/A207">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A207" />
  <foaf:name xml:lang="de">Bernhard von Breidenbach</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/118660047" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A207">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A207" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A207" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A100">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A100" />
  <foaf:name xml:lang="de">Kloster Reichklara Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/1138655031" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A100">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A100" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A100" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>

```



```
<rdf:Description rdf:about="http://www.mainz.de/id/A69">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A69" />
  <foaf:name xml:lang="de">Armkilara Kloster Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/1138655279" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A69">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A69" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A69" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A331">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A331" />
  <foaf:name xml:lang="de">Johann, von Nassau</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/118052632" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A331">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A331" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A331" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A281">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A281" />
  <foaf:name xml:lang="de">Jesuiten</foaf:name>
  < rdfs:seeAlso rdf:resource="https://d-nb.info/gnd/80092-2" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A281">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A281" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A281" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A67">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A67" />
  <foaf:name xml:lang="de">Universität Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/1136480633" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A67">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A67" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A67" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A911">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A911" />
  <foaf:name xml:lang="de">Dr. Jacobus Campius</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/119636107" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A911">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A911" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A911" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A177">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A177" />
  <foaf:name xml:lang="de">Kloster St. Michael Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/3072620-7" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A177">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A177" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A177" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
```

```
<rdf:Description rdf:about="http://www.mainz.de/id/A16">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A16"/>
  <foaf:name xml:lang="de">Pfarrei Sankt Ignaz Mainz</foaf:name>
  <rdfs:seeAlso rdf:resource="https://d-nb.info/gnd/3053046-5"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A16">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A16"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A16"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A37">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A37"/>
  <foaf:name xml:lang="de">Pfarrei St. Peter- St. Emmeran Mainz</foaf:name>
  <rdfs:seeAlso rdf:resource="https://d-nb.info/gnd/10075260-3"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A37">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A37"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A37"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A140">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A140"/>
  <foaf:name xml:lang="de">Kloster Maria Dalheim</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/121352718X"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A140">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A140"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A140"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A51">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A51"/>
  <foaf:name xml:lang="de">Stift Ravengiersburg</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/4627784-5"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A51">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A51"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A51"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A46">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A46"/>
  <foaf:name xml:lang="de">Pfarrei Liebfrauen Mainz</foaf:name>
  <rdfs:seeAlso rdf:resource="https://d-nb.info/gnd/3059710-9"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A46">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A46"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A46"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A314">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A314"/>
  <foaf:name xml:lang="de">Erhard von Redwitz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/1070494380"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A314">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription"/>
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document"/>
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A314"/>
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A314"/>
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  < dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dct:modified>
</rdf:Description>
```

```
<rdf:Description rdf:about="http://www.mainz.de/id/A26">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A26" />
  <foaf:name xml:lang="de">Heintze Wyd der bender</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A26">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A26" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A26" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A27">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A27" />
  <foaf:name xml:lang="de">Jeckel, Jeckel Fysches son</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A27">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A27" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A27" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A302">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A302" />
  <foaf:name xml:lang="de">Eberhard von Eppstein</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/135734452" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A302">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A302" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A302" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A9">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A9" />
  <foaf:name xml:lang="de">Stift St. Stephan Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/5222085-0" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A9">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A9" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A9" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A32">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A32" />
  <foaf:name xml:lang="de">Altmünster Kloster Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/4829887-6" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A32">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A32" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A32" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/A123">
  <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
  <foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A123" />
  <foaf:name xml:lang="de">Antoniterkloster Mainz</foaf:name>
  <owl:sameAs rdf:resource="https://d-nb.info/gnd/1213254728" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/A123">
  <rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
  <irw:isAbout rdf:resource="http://www.mainz.de/id/A123" />
  <foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A123" />
  <cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
  <cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
  <dc:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">2021-01-17</dc:modified>
</rdf:Description>
</rdf:RDF>
```

XML-Template

```

<!-- ROOT -->
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:irw="http://www.ontologydesignpatterns.org/ont/web/irw.owl#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:geo="http://www.opengis.net/ont/geosparql#"
  xmlns:cc="http://creativecommons.org/ns#"
  xmlns:gndo="https://d-nb.info/standards/elementset/gnd#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:mz="http://history.mainz.de/def#"
  {fme:process-features(("BUILDING", "PERSON", "SOURCE", "LANDOBJECT"))}
</rdf:RDF>
<!-- BUILDING -->
<rdf:Description rdf:about="http://www.mainz.de/id/B{fme:get-attribute("bID")}">
<rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
<foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/B{fme:get-attribute("bID")}"/>
<rdf:type rdf:resource="http://history.mainz.de/def#{fme:get-attribute("bFunction")}" />
<foaf:name xml:lang="de">{fme:get-attribute("bName")}</foaf:name>
  { if (fme:has-attribute("bInfo") and
    (string(fme:get-attribute("bInfo")) != ''))
    then <rdfs:comment xml:lang="de">{(normalize-space(fme:get-attribute("bInfo")))}</rdfs:comment>
    else () }
<mz:when rdf:datatype="http://www.w3.org/2001/XMLSchema#Year">{fme:get-attribute("bYear")}</mz:when>
  { for $ow_ID in fme:get-list-attribute("BUILDING_Owner").pID"
    return <mz:hadOwner rdf:resource="http://www.mainz.de/id/A{$ow_ID}"/> }
  { for $oc_ID in fme:get-list-attribute("BUILDING_OCCUPANT").pID"
    return <mz:hadOccupant rdf:resource="http://www.mainz.de/id/A{$oc_ID}"/> }
  { if (fme:get-attribute("bSocialGroup") ne "Unknown")
    then <mz:wasUsedBy rdf:parseType="Resource"><rdf:type rdf:resource="http://history.mainz.de/def#{fme:get-attribute("bSocialGroup")}" />
    </mz:wasUsedBy>
    else () }
  { for $i in 1 to 5
    let $P := (concat('bPicture', $i))
    return
      if (fme:has-attribute($P))
      then <foaf:depiction rdf:resource="{fme:get-attribute($P)}"/>
      else () }
  { for $s_ID in fme:get-list-attribute("BUILDING_SOURCE").sID"
    return <gndo:relatedWork rdf:resource="http://www.mainz.de/id/S{$s_ID}"/> }
<mz:where rdf:resource="http://www.mainz.de/id/GEOb{fme:get-attribute("bID")}" />
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/doc/B{fme:get-attribute("bID")}">
<rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
<irw:isAbout rdf:resource="http://www.mainz.de/id/B{fme:get-attribute("bID")}" />
<foaf:primaryTopic rdf:resource="http://www.mainz.de/id/B{fme:get-attribute("bID")}" />
<cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
<cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
<dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/id/GEOb{fme:get-attribute("bID")}">
<rdf:type rdf:resource="http://www.opengis.net/ont/sf#Polygon" />
<rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
<foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/GEOb{fme:get-attribute("bID")}" />
<geo:asWKT rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">&lt;http://www.opengis.net/def/crs/EPSSG/0/4326&gt;
{fme:get-attribute("bGeometry")}</geo:asWKT>
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/doc/GEOb{fme:get-attribute("bID")}">
<rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
<irw:isAbout rdf:resource="http://www.mainz.de/id/GEOb{fme:get-attribute("bID")}" />
<foaf:primaryTopic rdf:resource="http://www.mainz.de/id/GEOb{fme:get-attribute("bID")}" />
<cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
<cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
<dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description>
<!-- PERSON -->
<rdf:Description rdf:about="http://www.mainz.de/id/A{fme:get-attribute("pID")}">
<rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
<foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/A{fme:get-attribute("pID")}" />
<foaf:name xml:lang="de">{fme:get-attribute("pName")}</foaf:name>
</rdf:Description> ,
<rdf:Description rdf:about="http://www.mainz.de/doc/A{fme:get-attribute("pID")}">
<rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
<irw:isAbout rdf:resource="http://www.mainz.de/id/A{fme:get-attribute("pID")}" />
<foaf:primaryTopic rdf:resource="http://www.mainz.de/id/A{fme:get-attribute("pID")}" />
<cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
<cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
<dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description>

```



```

<!-- SOURCE -->
<rdf:Description rdf:about="http://www.mainz.de/id/S{fme:get-attribute("sID")}">
<rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
<rdf:type rdf:resource="http://purl.org/ontology/bibo/Document" />
<foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/S{fme:get-attribute("sID")}" />
<foaf:name xml:lang="de">{fme:get-attribute("sName")}</foaf:name>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/S{fme:get-attribute("sID")}">
<rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
<irw:isAbout rdf:resource="http://www.mainz.de/id/S{fme:get-attribute("sID")}" />
<foaf:primaryTopic rdf:resource="http://www.mainz.de/id/S{fme:get-attribute("sID")}" />
<cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
<cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
<dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description>
<!-- LANDOBJECT -->
<rdf:Description rdf:about="http://www.mainz.de/id/LO{fme:get-attribute("loID")}">
<rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
<foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/LO{fme:get-attribute("loID")}" />
<rdf:type rdf:resource="http://history.mainz.de/def#{fme:get-attribute("Type")}" />
<foaf:name xml:lang="de">{fme:get-attribute("Name")}</foaf:name>
<rdfs:comment xml:lang="de">{fme:get-attribute("Beschreibung")}</rdfs:comment>
<mz:when rdf:datatype="https://www.w3.org/2001/XMLSchema#Year">{fme:get-attribute("Jahr")}</mz:when>
<mz:where rdf:resource="http://www.mainz.de/id/GEOL0{fme:get-attribute("loID")}" />
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/LO{fme:get-attribute("loID")}">
<rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
<irw:isAbout rdf:resource="http://www.mainz.de/id/LO{fme:get-attribute("loID")}" />
<foaf:primaryTopic rdf:resource="http://www.mainz.de/id/LO{fme:get-attribute("loID")}" />
<cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
<cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
<dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/id/GEOL0{fme:get-attribute("loID")}">
<rdf:type rdf:resource="{concat('http://www.opengis.net/ont/sf#', (replace(replace(replace(substring-before({fme:get-attribute("loGeometry)'), ' '), 'POLYGON', 'Polygon'), 'MULTI', 'Multi'), 'LINESTRING', 'LineString'))}" />
<rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/web/irw.owl#NonInformationResource" />
<foaf:isPrimaryTopicOf rdf:resource="http://www.mainz.de/doc/GEOL0{fme:get-attribute("loID")}" />
<geo:asWKT rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">&lt;http://www.opengis.net/def/crs/EPSSG/0/4326&gt;
{fme:get-attribute("loGeometry")}</geo:asWKT>
</rdf:Description>
<rdf:Description rdf:about="http://www.mainz.de/doc/GEOL0{fme:get-attribute("loID")}">
<rdf:type rdf:resource="http://ontologydesignpatterns.org/ont/web/ldow2irw.owl#AssociatedDescription" />
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Document" />
<irw:isAbout rdf:resource="http://www.mainz.de/id/GEOL0{fme:get-attribute("loID")}" />
<foaf:primaryTopic rdf:resource="http://www.mainz.de/id/GEOL0{fme:get-attribute("loID")}" />
<cc:license rdf:resource="http://creativecommons.org/licenses/by/4.0/" />
<cc:attributionName rdf:datatype="https://www.w3.org/2001/XMLSchema#string">Stadtarchiv Mainz</cc:attributionName>
<dct:modified rdf:datatype="https://www.w3.org/2001/XMLSchema#date">{current-date()}</dct:modified>
</rdf:Description>

```

Verlinkte Geoobjekte

Die folgende Tabelle listet die Ergebnisse der SPARQL-Abfrage, die gegen den aktualisierten Quellgraphen gestellt wurde. Die Resultate wurden noch kommentiert, markiert und zwecks Lesbarkeit geeignet formatiert.

Historisches Häuserbuch (HHB)			Link-Typ	Gemeinsame Normdatei (GND)		Bemerkung zur Ressource und Begründung zum eingesetzten „Link-Prädikat“ (siehe 4. Spalte)
URI ¹⁵⁶	Name ¹⁵⁷	Modell		URI	bevorzugter Name	
mz:B120	Eichelstein	um 1450	owl:sameAs	https://d-nb.info/gnd/4753898-3	Eigelstein Mainz	Beide URIs identifizieren das Ehrenmal für General Drusus.
mz:B188	St. Gangolf	um 1450	owl:sameAs	https://d-nb.info/gnd/1147995249	Sankt Gangolf (Mainz)	Aufgrund des Bezugs zur kurfürstlichen Schloßkirche in der GND und im HHB auch für 1450. 1793 beschädigt, später zerstört.
mz:B211	Zum (großen) Stein	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/1067250816	Haus zum Stein (Mainz)	Die GND skizziert die gesamte Geschichte.
mz:B213	Pfarrkirche St. Ignaz mit Michaelskapelle	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4440449-9	Sankt Ignaz (Mainz)	Das HHB beschreibt den mittelalterlichen Vorgängerbau. Keine Angaben in der GND dazu.
mz:B368	Pfarrkirche St. Quirin	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4368131-1	Sankt Quirin (Mainz)	Mangels Details und ausgehenden Links in der GND.
mz:B370	Augustinerkirche zur Hl. Dreifaltigkeit	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4521920-5	Augustinerkirche Mainz	Die GND bezieht sich, im Gegensatz zum HHB, nicht auf einen bestimmten Zustand dieser Kirche.
mz:B374	Pfarrkirche St. Christoph	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4458963-3	Sankt Christoph (Mainz)	Die GND beschreibt den Zustand der Kirche im 20. Jh.; das HHB vor den Zerstörung im 2. Weltkrieg.
mz:B376	Stiftskirche St. Stephan	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4224871-1	Sankt Stephan (Mainz)	Das HHB enthält spezifische Bauinformationen und beschreibt ein bestimmtes Entwicklungsstadium.
mz:B448	Martinsburg	um 1450	owl:sameAs	https://d-nb.info/gnd/7745065-6	Martinsburg Mainz	Die GND und das HHB beschreiben dasselbe Geoobjekt.
mz:B498	Dom St. Martin	um 1450	owl:sameAs	https://d-nb.info/gnd/4122000-6	Dom Mainz (Mainz)	Das Vorstehende gilt entsprechend.
mz:B501	Kaufhaus	um 1450	owl:sameAs	https://d-nb.info/gnd/1050035720	Kaufhaus am Brand (Mainz)	Die Informationen im HHB und der GND sind fast deckungsgleich.
mz:B513	Kirche Mariengreden (St. Maria ad gradus, Liebfrauen)	um 1450	owl:sameAs	https://d-nb.info/gnd/4236039-0	Liebfrauenkirche Mainz	Beide URIs identifizieren die ehemalige Kirche nach dem Brand (1285).
mz:B528	Stiftskirche St. Johannes Baptist	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4369072-5	Sankt Johannes (Mainz)	Mangels Details und ausgehenden Links in der GND.
mz:B560	Karmeliterkirche St. Maria de Assumptione	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4452879-6	Karmeliterkirche Mainz	Das Vorstehende gilt entsprechend.

¹⁵⁶ Das Präfix „mz:“ steht hier für: <http://www.mainz.de/id/>

Da das System im Rahmen dieser Arbeit lediglich (teilweise) prototypisch implementiert und noch nicht produktiv gestellt wurde, sind die URIs, im Gegensatz zu jenen der DNB, noch nicht dereferenzierbar. Vollständige Beispiele für RDF-Beschreibungen wurden dem vorherigen Gliederungspunkt hinzugefügt.

¹⁵⁷ Die graue Markierung zeigt den primären Match (Workbench/Ausgabe).

mz:B579	Benediktinerkloster St. Jakob, auch St. Jakobsberg	um 1450	rdfs:seeAlso	https://d-nb.info/gnd/4670819-4	Sankt Jakob (Mainz)	Mangels Details und ausgehenden Links in der GND. 1793 fast komplett zerstört.
mz:B791	Königsteiner Hof, seit 1598 Dalberger Hof, eh. z. roten Haus	um 1620	owl:sameAs	https://d-nb.info/gnd/4741462-5	Alterer Dalberger Hof (Mainz)	Die GND und das HHB (nur 1620) beschreiben dasselbe Geoobjekt.
mz:B793	Kirche des Altmünsterklosters	um 1620	owl:sameAs	https://d-nb.info/gnd/1220381268	Altmünsterkirche (Mainz)	Sowohl die GND als auch das HHB (nur 1620) beziehen sich speziell auf die Kirche, welche ab 1662 neu errichtet wurde.
mz:B808	Stiftskirche (Schlosskirche) St. Gangolf	um 1620	owl:sameAs	https://d-nb.info/gnd/1147995249	Sankt Gangolf (Mainz)	Aufgrund des Bezugs zur kurfürstlichen Schlosskirche in der GND und im HHB. 1793 beschädigt, danach zerstört.
mz:B949	Zum Stein	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/1067250816	Haus zum Stein (Mainz)	Die GND skizziert die gesamte Geschichte
mz:B1008	Armklara Kloster (Arme Klarissen, Franziskanerinnen) seit 1619	um 1620	owl:sameAs	https://d-nb.info/gnd/4367519-0	Armklarenkirche Mainz	Die URIs beziehen sich beide auf die Kirche, welche von den Antonitern erbaut wurde und ab 1619 von den Armen Klarissen übernommen wurde
mz:B1021	Kurfürstliches Kaufhaus	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/1050035720	Kaufhaus am Brand (Mainz)	Wegen der Informationen im HHB, die auf den Eigentümerwechsel hinweisen
mz:B1131	Drususstein	um 1620	owl:sameAs	https://d-nb.info/gnd/4753898-3	Eigelstein Mainz	Beide URIs identifizieren das Ehrenmal für General Drusus
mz:B1197	Pfarrkirche St. Christoph(ell)	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4458963-3	Sankt Christoph (Mainz)	Die GND beschreibt den Zustand der Kirche im 20. Jh.; das HHB vor den Zerstörung im zweiten Weltkrieg
mz:B1299	Stift Mariengreden (St. Maria ad gradus, zu den Stufen, Staffeln), zu Unser Lieben Frauen	um 1620	owl:sameAs	https://d-nb.info/gnd/4236039-0	Liebfrauenkirche Mainz	Beide URIs identifizieren die ehemalige Kirche nach dem Brand (1285).
mz:B1305	Stift St. Johann Baptist	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4369072-5	Sankt Johannes (Mainz)	Mangels Details und ausgehenden Links in der GND.
mz:B1312	Augustiner(kloster)kirche	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4521920-5	Augustinerkirche Mainz	Die GND bezieht sich, im Gegensatz zum HHB, nicht auf einen bestimmten Zustand dieser Kirche.
mz:B1448	St. Jakob	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4670819-4	Sankt Jakob (Mainz)	Mangels Details und ausgehenden Links in der GND. 1793 fast komplett zerstört.
mz:B1461	Domstift St. Martin	um 1620	owl:sameAs	https://d-nb.info/gnd/4122000-6	Dom Mainz (Mainz)	Die GND und das HHB beschreiben dasselbe Geoobjekt..
mz:B1503	Kurfürstliches Schloss	um 1620	owl:sameAs	https://d-nb.info/gnd/4206936-1	Kurfürstliches Schloss (Mainz)	Das Vorstehende gilt entsprechend.
mz:B1591	Schweikhardtsburg (später Zitadelle)	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/1129127311	Zitadelle Mainz (Mainz)	Die GND beschreibt den Nachfolgebau der Schweikhardtsburg (nach dem Dreißigjährigen Krieg).
mz:B1602	Pfarrkirche St. Ignaz	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4440449-9	Sankt Ignaz (Mainz)	Mangels Details und ausgehenden Links in der GND.

mz:B1607	Karmeliterkirche	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4452879-6	Karmeliterkirche Mainz	Das Vorstehende gilt entsprechend.
mz:B1651	St. Stephan	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4224871-1	Sankt Stephan (Mainz)	Das HHB enthält spezifische Bauinformationen und beschreibt ein bestimmtes Entwicklungsstadium. Die GND jedoch nicht.
mz:B1719	Pfarrkirche St. Quintin	um 1620	rdfs:seeAlso	https://d-nb.info/gnd/4368131-1	Sankt Quintin (Mainz)	Mangels Details und ausgehenden Links in der GND.
mz:LO14	Marktbrunnen	um 1620	owl:sameAs	https://d-nb.info/gnd/4798065-5	Marktbrunnen Mainz (Mainz)	Die GND und das HHB beschreiben beide den ältesten und wohl schönsten Renaissancebrunnen Deutschlands.
mz:B2238	Festungskommandantur (Stadioner Hof)	um 1871	owl:sameAs	https://d-nb.info/gnd/4534174-6	Stadioner Hof (Mainz)	Die GND und das HHB beschreiben dasselbe Geoobjekt.
mz:B3216	Neubrunnen	um 1871	owl:sameAs	https://d-nb.info/gnd/7550783-3	Neuer Brunnen (Mainz)	Die GND und das HHB beschreiben denselben Brunnen, auch wenn dieser im HHB fälschlicherweise der Haus-Relation enthalten ist.
mz:B3666	St. Quintin	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4368131-1	Sankt Quintin (Mainz)	Mangels Details und ausgehenden Links in der GND.
mz:B4142	St. Christoph	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4458963-3	Sankt Christoph (Mainz)	Die GND beschreibt den Zustand der Kirche im 20. Jh.; das HHB vor den Zerstörung im 2. Weltkrieg.
mz:B4312	Zum Stein	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/1067250816	Haus zum Stein (Mainz)	Die GND skizziert die gesamte Geschichte.
mz:B4384	St. Johannis	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4369072-5	Sankt Johannis (Mainz)	Mangels Details und ausgehenden Links in der GND.
mz:B4882	Stadtheater	um 1871	owl:sameAs	https://d-nb.info/gnd/1123091404	Theaterbau des Staatstheaters Mainz (Mainz)	Die Daten und historischen Informationen, lassen darauf schließen, dass im HHB und in der GND dasselbe Geoobjekt beschrieben wird.
mz:B4888	St. Josephskapelle	um 1871	owl:sameAs	https://d-nb.info/gnd/7731500-5	Sankt Josephskapelle (Mainz)	Das Vorstehende gilt entsprechend.
mz:B4923	Augustinerkirche	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4521920-5	Augustinerkirche Mainz	Die GND bezieht sich, im Gegensatz zum HHB, nicht auf einen bestimmten Zustand dieser Kirche.
mz:B5003	St. Peter	um 1871	owl:sameAs	https://d-nb.info/gnd/4281560-5	Sankt Peter (Mainz)	Die GND und das HHB beschreiben dasselbe Geoobjekt
mz:B5224	Dom St. Martin	um 1871	owl:sameAs	https://d-nb.info/gnd/4122000-6	Dom Mainz (Mainz)	Das Vorstehende gilt entsprechend.
mz:B5346	Justizpalast, eh. Jüngerer Dalberger Hof	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4132095-5	Dalberger Hof (Mainz)	Die GND beschreibt den Vorgängerbau, an dem beim Umbau zum Justizpalast Änderungen vorgenommen wurden.
mz:B5747	St. Stephan	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4224871-1	Sankt Stephan (Mainz)	Mangels Angaben in der GND und im HHB.
mz:B6018	Großherzogliches Palais (Deutschhaus)	um 1871	owl:sameAs	https://d-nb.info/gnd/4237044-9	Deutschhaus Mainz (Mainz)	Die GND und das HHB beschreiben dasselbe Geoobjekt.

mz:B6201	Erthaler Hof (großherzogl. Provinzialdirektion Rheinhessen und Kreisamt)	um 1871	owl:sameAs	https://d-nb.info/gnd/4297193-7	Erthaler Hof (Mainz)	Das Vorstehende gilt entsprechend.
mz:B6297	St. Ignaz	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4440449-9	Sankt Ignaz (Mainz)	Mangels Angaben und Links in der GND.
mz:B6308	Karmeliterkirche	um 1871	rdfs:seeAlso	https://d-nb.info/gnd/4452879-6	Karmeliterkirche Mainz	Das Vorstehende gilt entsprechend.

Geografika in Mainz (GND)

Die folgende Tabelle listet die Kandidaten in der GND, die nicht mit dem historischen Häuserbuch verknüpft werden konnten und enthält die Gründe dafür:

URI	bevorzugter Name	Bemerkung
https://d-nb.info/gnd/4827171-8	Schloss Favorite (Mainz)	Ein – im wahrsten Sinne des Wortes – repräsentatives Beispiel einer barocken Anlage, die detailliert im Modell für die Zeit um 1750 beschrieben wird (1793 zerstört). Die Datenerfassung ist noch nicht abgeschlossen.
https://d-nb.info/gnd/7575398-4	Sankt Michael (Mainz)	Laut GND im Jahr 1330 erbaut und 1790/92 abgerissen. Die Klosterkirche wird im HHB jedoch nicht beschrieben.
https://d-nb.info/gnd/4731107-1	Isistempel Mainz	Der Isistempel wurde erst Anfang des 21. Jahrhunderts wiederentdeckt.
https://d-nb.info/gnd/7602760-0	Alter Jüdischer Friedhof (Mainz)	Wird im HHB nicht beschrieben.
https://d-nb.info/gnd/7583211-2	Ingelheimer Aue (Mainz)	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/4368799-4	Auferstehungskirche Mainz	Die Grundsteinlegung erfolgte erst im Jahr 1961.
https://d-nb.info/gnd/4298330-7	Bundesfestung Mainz (Mainz)	Wegen topologischer Probleme nicht ins RDF transformiert.
https://d-nb.info/gnd/4716717-8	Christuskirche Mainz	Wurde erst im Jahr 1903 fertiggestellt.
https://d-nb.info/gnd/4833784-5	Sankt Bernhard (Mainz)	Wurde erst im Jahr 1992 erbaut.
https://d-nb.info/gnd/4368801-9	Sankt Georg (Mainz)	Ein historischer Bezug besteht zwar; die Kirche liegt jedoch zu weit vom ehemaligen Stadtzentrum entfernt (Bretzenheim), weshalb das HHB keine Beschreibungen enthält.
https://d-nb.info/gnd/4671478-9	Maria Himmelfahrt (Weisenau)	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/1140503154	Kreuzkirche (Mainz)	Wurde erst im Jahr 1998 fertiggestellt.
https://d-nb.info/gnd/1075921007	Bahnhof Mainz Römisches Theater (Mainz)	Ist in der Häuser-Relation für die Zeit um 1871 nicht enthalten.
https://d-nb.info/gnd/1104396149	Ernst-Ludwig-Platz (Mainz)	Das Vorstehende gilt grds. entsprechend; zusätzlich wurde der Platz erst ab 1899 so genannt.
https://d-nb.info/gnd/7711900-9	Ausstellungsraum Kästrich (Mainz)	Kein Bezug zum historischen Häuserbuch.
https://d-nb.info/gnd/107359601X	Evangelisches Gemeinde- zentrum Draï (Mainz)	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/4743938-5	Große Jupitersäule (Mainz)	Diese wurde Anfang des 20. Jahrhunderts wiederentdeckt. Sie ist in der Relation, welche die Häuser für die Zeit um 1871 enthält, nicht beschrieben.
https://d-nb.info/gnd/4198373-7	Sankt Franziskus (Mainz)	Laut GND von 1982 bis 1984 erbaut.
https://d-nb.info/gnd/4621300-4	Gutenberg-Pavillon Mainz	GND: Wurde für die Dauer des Gutenberg-Jubiläums im Jahr 2000 am Rheinufer der Stadt Mainz errichtet.
https://d-nb.info/gnd/4524681-6	Staatstheater Mainz (Mainz) / Kleines Haus	Das „kleine Haus“ wurde erst Ende des 20. Jahrhunderts errichtet.
https://d-nb.info/gnd/107100932X	Maria-Hilf-Kapelle (Mainz)	Fällt in die Zeit um 1871, liegt jedoch zu weit weg vom historischen Stadtzentrum. Das Häuserbuch hat keine Einträge für Mainz-Draï.
https://d-nb.info/gnd/4832185-0	Nagelsäule Mainz	Wurde erst im Jahr 1916 erstellt. Ist im Modell für die Zeit um 1871 nicht beschrieben.
https://d-nb.info/gnd/4363567-2	Fastnachtsbrunnen Mainz	Wurde erst im Jahr 1967 fertiggestellt.
https://d-nb.info/gnd/4394374-3	Zoll- und Binnenhafen (Mainz)	Ist in der Relation, welche die Häuser für die Zeit um 1871 enthält, nicht beschrieben.
https://d-nb.info/gnd/1138999687	Sankt Achatus (Mainz)	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/7556205-4	Hauptfriedhof Mainz (Mainz)	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/4616902-7	Friedrichstraße (Mainz)	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/4394953-8	Kaiserstraße Mainz	Das Vorstehende gilt entsprechend.
https://d-nb.info/gnd/4412822-8	Löhrstraße Mainz	Das Vorstehende gilt entsprechend.

https://d-nb.info/gnd/4508107-4	Platz der Köpfe (Mainz)	Wurde erst im Jahr 1980 erstellt.
https://d-nb.info/gnd/4380953-4	Theodor-Heuss-Brücke Mainz	Aus dem Modell für die Zeit um 1871 wurden lediglich die Häuser im Stadtzentrum ins RDF transformiert
https://d-nb.info/gnd/4716995-3	Sankt Alban (Mainz)	Laut GND von 1951 bis 1952 erbaut.
https://d-nb.info/gnd/4831697-0	Hauptbahnhof (Mainz)	Die URI in der GND identifiziert den heutigen Hauptbahnhof, der im Modell für die Zeit um 1871 nicht beschrieben wird.
https://d-nb.info/gnd/7858980-0	Neue Synagoge Mainz (Mainz)	Wurde erst im Jahr 2010 fertiggestellt.
https://d-nb.info/gnd/1073594726	Evangelische Emmauskirche (Mainz)	Laut GND erfolgte die Grundsteinlegung erst am 22.04.1989
https://d-nb.info/gnd/4788769-2	Tritonplatz (Mainz)	Der Platz (oder Vorläufer davon) sind im HHB nicht beschrieben
https://d-nb.info/gnd/4849150-0	Neuer Jüdischer Friedhof (Mainz)	Fällt in die Zeit um 1871. Aus diesem Modell wurden nur Häuser ins RDF transformiert.
https://d-nb.info/gnd/4621765-4	ZDF-Medienpark Mainz	Kein Bezug zum historischen Häuserbuch.
https://d-nb.info/gnd/4659668-9	Abgeordneten- und Ministerialdienstgebäude des Landes Rheinland-Pfalz (Mainz)	Laut GND im Jahr 1999 errichtet.
https://d-nb.info/gnd/1050915100	Hauptsynagoge Mainz (Mainz)	Wurde erst 1911/12 erbaut. Ist im Modell für die Zeit um 1871 nicht beschrieben.
https://d-nb.info/gnd/119268396X	Georg Forster-Haus (Mainz)	Kein Bezug zum historischen Häuserbuch.
https://d-nb.info/gnd/7629840-1	Heilig-Kreuz-Kirche (Mainz)	Laut GND von 1953 bis 1954 erbaut.
https://d-nb.info/gnd/1076294715	Rathaus Mainz (Mainz)	Wurde von 1970 bis 1974 erbaut. Keine Grundlage für die Verlinkung zum spätmittelalterlichen Rathaus
https://d-nb.info/gnd/4731109-5	Römerpassage Mainz	Modernes Einkaufszentrum.
https://d-nb.info/gnd/116110623	Hofgarten (Mainz)	Dieser wird im Modell für die Zeit um 1871 nicht gesondert ausgewiesen.

TBox

```

<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE rdf:RDF [
  <!ENTITY mz "http://history.mainz.de/def#" >
  <!ENTITY foaf "http://xmlns.com/foaf/0.1/" >
  <!ENTITY geo "http://www.opengis.net/ont/geosparql#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
]>

<rdf:RDF xmlns="http://history.mainz.de/def#"
  xml:base="http://history.mainz.de/def"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:geo="http://www.opengis.net/ont/geosparql#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <owl:Ontology rdf:about="http://history.mainz.de/def#">
    <rdfs:comment xml:lang="en">This vocabulary describes historic objects which were
      derived from historic maps and documents. They are used to build models which show
      the changing topography and urban development of the city of Mainz from the Middle
      Ages to the present.</rdfs:comment>
    <owl:imports rdf:resource="http://www.opengis.net/ont/geosparql"/>
    <owl:imports rdf:resource="http://xmlns.com/foaf/0.1"/>
  </owl:Ontology>

  <!--
  ///////////////////////////////////////////////////////////////////////////////////////////////////////////////////
  //
  // Classes
  //
  ///////////////////////////////////////////////////////////////////////////////////////////////////////////////////
  -->

  <!-- http://history.mainz.de/def#HistoricSocialGroup -->

  <owl:Class rdf:about="&mz;HistoricSocialGroup">
    <rdfs:label xml:lang="en">HistoricSocialGroup</rdfs:label>
    <rdfs:subClassOf rdf:resource="&foaf;Group"/>
    <rdfs:comment xml:lang="en">Social stratification of the urban population based on
      profession, religious affiliation, etc.</rdfs:comment>
  </owl:Class>

```



```
<!-- http://history.mainz.de/def#Aristocracy -->

<owl:Class rdf:about="&mz;Aristocracy">
  <rdfs:label xml:lang="en">Aristocracy</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#Bourgeois -->

<owl:Class rdf:about="&mz;Bourgeois">
  <rdfs:label xml:lang="en">Bourgeois</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#Christian -->

<owl:Class rdf:about="&mz;Christian">
  <rdfs:label xml:lang="en">Christian</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#Court -->

<owl:Class rdf:about="&mz;Court">
  <rdfs:label xml:lang="en">Court</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#Elector -->

<owl:Class rdf:about="&mz;Elector">
  <rdfs:label xml:lang="en">Elector</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#Guild -->

<owl:Class rdf:about="&mz;Guild">
  <rdfs:label xml:lang="en">Guild</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#Jew -->

<owl:Class rdf:about="&mz;Jew">
  <rdfs:label xml:lang="en">Jew</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>
```

```
<!-- http://history.mainz.de/def#Patrician -->

<owl:Class rdf:about="&mz;Patrician">
  <rdfs:label xml:lang="en">Patrician</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricSocialGroup"/>
</owl:Class>

<!-- http://history.mainz.de/def#HistoricFeature -->

<owl:Class rdf:about="&mz;HistoricFeature">
  <rdfs:label xml:lang="en">HistoricFeature</rdfs:label>
  <rdfs:subClassOf rdf:resource="&geo;Feature"/>
  <owl:disjointWith rdf:resource="&mz;HistoricGeometry"/>
  <rdfs:comment xml:lang="en">ISO-19101 defines a feature as an: &quot;abstraction
of real world phenomena&quot;. These features were derived from historic
maps and documents - as mentioned above.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#HistoricGeometry -->

<owl:Class rdf:about="&mz;HistoricGeometry">
  <rdfs:label xml:lang="en">HistoricGeometry</rdfs:label>
  <rdfs:subClassOf rdf:resource="&geo;Geometry"/>
  <owl:disjointWith rdf:resource="&mz;HistoricFeature"/>
  <rdfs:comment xml:lang="en">The geometries were derived from historic
maps and documents.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#LanduseObject -->

<owl:Class rdf:about="&mz;LanduseObject">
  <rdfs:label xml:lang="en">LanduseObject</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricFeature"/>
  <rdfs:comment xml:lang="en">This is intended to be the superclass of all land
use objects which represent the human use of land and water.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#GreenArea -->

<owl:Class rdf:about="&mz;GreenArea">
  <rdfs:label xml:lang="en">GreenArea</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;LanduseObject"/>
</owl:Class>

<!-- http://history.mainz.de/def#Garden -->

<owl:Class rdf:about="&mz;Garden">
  <rdfs:label xml:lang="en">Garden</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;GreenArea"/>
</owl:Class>
```

```
<!-- http://history.mainz.de/def#Graveyard -->

<owl:Class rdf:about="&mz;Graveyard">
  <rdfs:label xml:lang="en">Graveyard</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;GreenArea"/>
</owl:Class>

<!-- http://history.mainz.de/def#HerbGarden -->

<owl:Class rdf:about="&mz;HerbGarden">
  <rdfs:label xml:lang="en">HerbGarden</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Garden"/>
</owl:Class>

<!-- http://history.mainz.de/def#PleasureGarden -->

<owl:Class rdf:about="&mz;PleasureGarden">
  <rdfs:label xml:lang="en">PleasureGarden</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Garden"/>
</owl:Class>

<!-- http://history.mainz.de/def#Vineyard -->

<owl:Class rdf:about="&mz;Vineyard">
  <rdfs:label xml:lang="en">Vineyard</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Garden"/>
  <rdfs:comment xml:lang="en">These are moderately steep vineyards located
  near the buildings.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#WaterBody -->

<owl:Class rdf:about="&mz;WaterBody">
  <rdfs:label xml:lang="en">WaterBody</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;LanduseObject"/>
</owl:Class>

<!-- http://history.mainz.de/def#Creek -->

<owl:Class rdf:about="&mz;Creek">
  <rdfs:label xml:lang="en">Creek</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;WaterBody"/>
</owl:Class>

<!-- http://history.mainz.de/def#Pond -->

<owl:Class rdf:about="&mz;Pond">
  <rdfs:label xml:lang="en">Pond</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;WaterBody"/>
</owl:Class>
```

```
<!-- http://history.mainz.de/def#OxbowLake -->

<owl:Class rdf:about="&mz;OxbowLake">
  <rdfs:label xml:lang="en">OxbowLake</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;WaterBody" />
  <rdfs:comment xml:lang="en">It is not clear if these phenomena were of natural origin or
  resulted due to human intervention. At least they were used by the people.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#River -->

<owl:Class rdf:about="&mz;River">
  <rdfs:label xml:lang="en">River</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;WaterBody" />
</owl:Class>

<!-- http://history.mainz.de/def#ConstructionObject -->

<owl:Class rdf:about="&mz;ConstructionObject">
  <rdfs:label xml:lang="en">ConstructionObject</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;HistoricFeature" />
  <rdfs:comment xml:lang="en">Superclass of all man-made objects.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#Infrastructure -->

<owl:Class rdf:about="&mz;Infrastructure">
  <rdfs:label xml:lang="en">Infrastructure</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;ConstructionObject" />
</owl:Class>

<!-- http://history.mainz.de/def#Bridge -->

<owl:Class rdf:about="&mz;Bridge">
  <rdfs:label xml:lang="en">Bridge</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Infrastructure" />
</owl:Class>

<!-- http://history.mainz.de/def#Well -->

<owl:Class rdf:about="&mz;Well">
  <rdfs:label xml:lang="en">Well</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Infrastructure" />
</owl:Class>
```

```
<!-- http://history.mainz.de/def#Fortification -->

<owl:Class rdf:about="&#x26;mz;Fortification">
  <rdfs:label xml:lang="en">Fortification</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;ConstructionObject"/>
</owl:Class>

<!-- http://history.mainz.de/def#Barrier -->

<owl:Class rdf:about="&#x26;mz;Barrier">
  <rdfs:label xml:lang="en">Barrier</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;Fortification"/>
</owl:Class>

<!-- http://history.mainz.de/def#WolfPit -->

<owl:Class rdf:about="&#x26;mz;WolfPit">
  <rdfs:label xml:lang="en">WolfPit</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;Barrier"/>
  <rdfs:comment xml:lang="en">A constructed barrier consisting of a pit with one or
  more sharpened wooden stakes.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#StoneWall -->

<owl:Class rdf:about="&#x26;mz;StoneWall">
  <rdfs:label xml:lang="en">StoneWall</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;Barrier"/>
</owl:Class>

<!-- http://history.mainz.de/def#CityWall -->

<owl:Class rdf:about="&#x26;mz;CityWall">
  <rdfs:label xml:lang="en">CityWall</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;StoneWall"/>
</owl:Class>

<!-- http://history.mainz.de/def#BastionWall -->

<owl:Class rdf:about="&#x26;mz;BastionWall">
  <rdfs:label xml:lang="en">BastionWall</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;StoneWall"/>
</owl:Class>

<!-- http://history.mainz.de/def#Earthwall -->

<owl:Class rdf:about="&#x26;mz;Earthwall">
  <rdfs:label xml:lang="en">Earthwall</rdfs:label>
  <rdfs:subClassOf rdf:resource="&#x26;mz;Barrier"/>
</owl:Class>
```

```
<!-- http://history.mainz.de/def#Gate -->

<owl:Class rdf:about="&mz;Gate">
  <rdfs:label xml:lang="en">Gate</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Barrier"/>
</owl:Class>

<!-- http://history.mainz.de/def#Hedgelaying -->

<owl:Class rdf:about="&mz;Hedgelaying">
  <rdfs:label xml:lang="en">Hedgelaying</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Barrier"/>
  <rdfs:comment xml:lang="en">Constructed by &quot;cutting branches partway through,
  laying them horizontally, and pegging them in position in order to create a strong
  thick hedge&quot; as defined by Collins Dictionary.</rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#Moat -->

<owl:Class rdf:about="&mz;Moat">
  <rdfs:label xml:lang="en">Moat</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Barrier"/>
  <rdfs:comment xml:lang="en">It is not clear if all moats were filled with water.
  </rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#Tower -->

<owl:Class rdf:about="&mz;Tower">
  <rdfs:label xml:lang="en">Tower</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Fortification"/>
  <rdfs:comment xml:lang="en">It is not clear if all towers had a roof or rooms.
  </rdfs:comment>
</owl:Class>

<!-- http://history.mainz.de/def#Building -->

<owl:Class rdf:about="&mz;Building">
  <rdfs:label xml:lang="en">Building</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;ConstructionObject"/>
  <rdfs:comment xml:lang="en">A building is a man-made structure with a roof. A building can
  be independently used and entered by humans and is primarily designed for the purpose of
  protecting humans, animals or property.</rdfs:comment>
</owl:Class>
```



```
<!-- http://history.mainz.de/def#MilitaryBuilding -->

<owl:Class rdf:about="&mz;MilitaryBuilding">
  <rdfs:label xml:lang="en">MilitaryBuilding</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Fortification"/>
  <rdfs:subClassOf rdf:resource="&mz;Building"/>
</owl:Class>

<!-- http://history.mainz.de/def#BusinessBuilding -->

<owl:Class rdf:about="&mz;BusinessBuilding">
  <rdfs:label xml:lang="en">BusinessBuilding</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Building"/>
</owl:Class>

<!-- http://history.mainz.de/def#Inn -->

<owl:Class rdf:about="&mz;Inn">
  <rdfs:label xml:lang="en">Inn</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;BusinessBuilding"/>
</owl:Class>

<!-- http://history.mainz.de/def#PublicBuilding -->

<owl:Class rdf:about="&mz;PublicBuilding">
  <rdfs:label xml:lang="en">PublicBuilding</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Building"/>
</owl:Class>

<!-- http://history.mainz.de/def#University -->

<owl:Class rdf:about="&mz;University">
  <rdfs:label xml:lang="en">University</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;PublicBuilding"/>
</owl:Class>

<!-- http://history.mainz.de/def#Church -->

<owl:Class rdf:about="&mz;Church">
  <rdfs:label xml:lang="en">Church</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;PublicBuilding"/>
</owl:Class>

<!-- http://history.mainz.de/def#Castle -->

<owl:Class rdf:about="&mz;Castle">
  <rdfs:label xml:lang="en">Castle</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;MilitaryBuilding"/>
  <rdfs:comment xml:lang="en">Class of private fortified residences.</rdfs:comment>
</owl:Class>
```

```
<!-- http://history.mainz.de/def#Palace -->

<owl:Class rdf:about="&mz;Palace">
  <rdfs:label xml:lang="en">Palace</rdfs:label>
  <rdfs:subClassOf rdf:resource="&mz;Building"/>
  <rdfs:comment xml:lang="en">A palace is not fortified.</rdfs:comment>
</owl:Class>

<!--
////////////////////////////////////
//
// Properties
//
////////////////////////////////////
-->

<!-- http://history.mainz.de/def#hadOwner -->

<owl:ObjectProperty rdf:about="&mz;hadOwner">
  <rdfs:label xml:lang="en">hadOwner</rdfs:label>
  <rdfs:domain rdf:resource="&mz;Building"/>
  <rdfs:range rdf:resource="&foaf;Agent"/>
  <rdfs:comment xml:lang="en">An owner is a single person, a family, a corporate body or any
  other group of persons who had the exclusive rights and control over the of the entire
  property or of parts of the property.</rdfs:comment>
</owl:ObjectProperty>

<!-- http://history.mainz.de/def#hadOccupant -->

<owl:ObjectProperty rdf:about="&mz;hadOccupant">
  <rdfs:label xml:lang="en">hadOccupant</rdfs:label>
  <rdfs:domain rdf:resource="&mz;Building"/>
  <rdfs:range rdf:resource="&foaf;Agent"/>
  <rdfs:comment xml:lang="en">An occupant is a single person or a family who had the right to
  live in the whole building or parts of the building.</rdfs:comment>
</owl:ObjectProperty>
```

```
<!-- http://history.mainz.de/def#wasUsedBy -->

<owl:ObjectProperty rdf:about="&mz;wasUsedBy">
  <rdfs:label xml:lang="en">wasUsedBy</rdfs:label>
  <rdfs:domain rdf:resource="&mz;HistoricFeature"/>
  <rdfs:range rdf:resource="&mz;HistoricSocialGroup"/>
</owl:ObjectProperty>

<!-- http://history.mainz.de/def#where -->

<owl:ObjectProperty rdf:about="&mz;where">
  <rdfs:label xml:lang="en">where</rdfs:label>
  <rdfs:domain rdf:resource="&mz;HistoricFeature"/>
  <rdfs:range rdf:resource="&mz;HistoricGeometry"/>
  <rdfs:subPropertyOf rdf:resource="&geo;hasGeometry"/>
  <rdfs:comment xml:lang="en">Note that this property is not functional. GeoSPARQL queries
  shall be used to determine if geometries are equal.</rdfs:comment>
</owl:ObjectProperty>

<!-- http://history.mainz.de/def#when -->

<owl:DatatypeProperty rdf:about="&mz;when">
  <rdfs:label xml:lang="en">when</rdfs:label>
  <rdfs:domain rdf:resource="&mz;HistoricFeature"/>
  <rdfs:range rdf:resource="&xsd;gYear"/>
</owl:DatatypeProperty>

</rdf:RDF>

<!-- Generated by the OWL API (version 4.5.9.2019-02-01T07:24:44Z) https://github.com/owlcs/owlapi
-->
```