Masterarbeit

zur Erlangung des akademischen Grades
Master

Technische Hochschule Wildau

Fachbereich Wirtschaft, Informatik, Recht

Studiengang Bibliotheks-informatik M. Sc.

Thema (deutsch): Ein Knowledge Graph für wissenschaftliche Sammlungen :
Generierung von Linked Open Data für heterogene museale
Sammlungen auf der Basis des ASCH-Modells

Thema (englisch): A Knowledge Graph for Academic Collections : Providing Linked
Open Data for Heterogeneous Museum Collections on the Basis of the
ASCH Model

Autor/in: Antje Niemann

Matrikel-Nr.: 50041781

Seminargruppe: BIM/16

Betreuer/in: M.A. Petra Keidel

Zweitgutachter/in: Prof. Dr. rer. nat. Janett Mohnke

Eingereicht am: 24.03.2019

ABSTRACT

Die vorliegende Arbeit liefert eine prototypische Anwendung zur besseren Sichtbarkeit und Verfügbarkeit musealer Sammlungen im Semantic Web.

Es wird ein Verfahren vorgestellt, das die LIDO-XML-Metadaten, die die Sammlungsobjekte der wissenschaftlichen Sammlungen der Georg-August-Universität Göttingen beschreiben, in Linked Open Data (LOD) umwandelt und in einem Triplestore bereitstellt. Der dargestellte Ansatz sowie die Umsetzung eignen sich zur Nachnutzung auch durch kleinere museale Einrichtungen ohne vertiefte Semantic-Web-Erfahrung.

Als Basis für die Generierung der RDF Statements dient das ASCH-Modell, ein unlängst entwickeltes RDF-Datenmodell für heterogene museale Sammlungen, dessen Proof of Concept ein weiteres Ziel dieser Arbeit darstellt.

Für die Datenanalyse werden die Quelldaten in eine XML-Datenbank importiert und mit XPath und XQuery untersucht. Die Konvertierung der Daten in das Serialisierungsformat RDF/XML erfolgt anhand eines XSLT-Stylesheets. Außerdem werden die Daten in das RDF-Format Turtle überführt und in einen mit SPARQL abfragbaren Triplestore geladen.

Die für die Umsetzung notwendigen Komponenten der Abfrage- und Programmiersprachen sowie der Tools werden anhand von Beispielen erläutert.

Die Bereitstellung des Knowledge Graph erfolgt unter Beachtung der Linked Data Principles und der vom World Wide Web Consortium (W3C) veröffentlichten Best Practices for Publishing Linked Data. Ferner finden die besonderen Anforderungen des Kulturerbesektors, insbesondere der wissenschaftlichen Sammlungen, Berücksichtigung.

This thesis provides a prototypical application for better visibility and availability of museum collections in the Semantic Web.

A method is presented that transforms the LIDO-XML metadata, which describe the objects of the academic collections of the Georg-August-University Göttingen, into Linked Open Data (LOD) and publishes it in a triplestore. The illustrated approach and the implementation are suitable for reuse also by smaller museums without deepened Semantic Web experience. The basis for the LOD generation is the ASCH model, a recently developed RDF data model for heterogeneous museum collections. Its proof of concept is another goal of this thesis.

For data analysis, the source data is imported into an XML database and examined with XPath and XQuery. The data conversion into RDF/XML is realized by an XSLT style sheet. In addition, the data is converted into Turtle and published in a triplestore enabling networking with external databases by means of SPARQL.

The components of the query and programming languages and the tools necessary for the implementation are explained by examples.

The Knowledge Graph adheres to the Linked Data Principles and the W3C Best Practices for Publishing Linked Data. Furthermore, the requirements of the cultural heritage sector, in particular the scientific collections, are considered.

INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS	VII
TABELLENVERZEICHNIS	IX
ABKÜRZUNGSVERZEICHNIS	XI
DANKSAGUNG	XV
1 EINLEITUNG.....	1
1.1 Motivation und Ziele	1
1.2 Aufbau der Arbeit	3
2 SEMANTIC WEB.....	5
2.1 Linked Data und URIs.....	5
2.2 Resource Description Framework	7
2.3 Knowledge Graph	9
2.4 Generierung und Modellierung von Wissen	9
2.5 Veröffentlichung von Linked Data	12
3 KULTURERBESEKTOR	15
3.1 Erfassung musealer Objekte	15
3.2 Wissenschaftliche Sammlungen	16
3.3 Portale und Erfassungsdatenbanken.....	18
3.4 Metadatenformat LIDO	22
4 GÖTTINGER SAMMLUNGSPORTAL.....	24
4.1 Entstehung, Ziele und Webauftritt.....	24
4.2 Ausgangsdaten	26
4.2.1 Herkunft und Sichtbarkeit.....	26
4.2.2 Syntax und Semantik.....	28
4.2.3 Tools für die Datenanalyse	31
5 ASCH-MODELL	35
5.1 Entstehung und Zweck	35
5.2 Modellierte Klassen	36
5.3 Nutzung von Standards	39
5.4 Eigene Ontologien	41
5.5 Application Profiles.....	43
6 MAPPING	47
6.1 Erarbeitung der Mapping-Vorgaben	47
6.2 Konvertierung in Linked Open Data	55
6.2.1 RDF/XML und Turtle.....	55
6.2.2 XSLT-Code	58
6.2.3 Batch-Verarbeitung.....	64
6.3 Ergebnisse und Testing.....	66

7	BEREITSTELLUNG DER DATEN	69
7.1	Triplestore.....	69
7.2	SPARQL	72
8	FAZIT UND AUSBLICK	75
8.1	Fazit.....	75
8.2	Ausblick.....	78
	LITERATURVERZEICHNIS	81
ANHANG 1:	Mapping-Tabelle lido2rdf	XVII
ANHANG 2:	XSLT-Stylesheet lido2rdfxml.xsl	XXIII
ANHANG 3:	Datensatz record_kuniweb_929301 in LIDO-XML.....	XXXVII
ANHANG 4:	Datensatz record_kuniweb_929301 in RDF/XML	XLV
ANHANG 5:	Datensatz record_kuniweb_929301 in Turtle	XLVIII
ANHANG 6:	Datensatz record_kuniweb_929301 als RDF-Graph	L
ANHANG 7:	Readme.txt der beiliegenden CD	LI

ABBILDUNGSVERZEICHNIS

Abbildung 1: Aufbau eines URIs	6
Abbildung 2: RDF-Triple.....	8
Abbildung 3: Blank Node	8
Abbildung 4: Linked Open Data Cloud	14
Abbildung 5: LIDO-Elemente	22
Abbildung 6: Ausschnitt aus einem LIDO-Datensatz.....	23
Abbildung 7: CIDOC CRM-Eintrag Beginning of Existence	23
Abbildung 8: Herkunft der Ausgangsdaten	26
Abbildung 9: Struktur des LIDO-Datensatzes im Sammlungsportal	29
Abbildung 10: Ausschnitt aus einem LIDO-Datensatz mit Schlagwort-URI	30
Abbildung 11: Skript zur Zusammenführung sämtlicher Quelldateien.....	32
Abbildung 12: XML-Datei mit zehn mittels XInclude eingebundenen Quelldateien	32
Abbildung 13: XQuery-FLOWR-Anfrage und Antwort.....	34
Abbildung 14: Das ASCH-Modell. Basisklassen	37
Abbildung 15: LIDO-Terminologie, eventType	39
Abbildung 16: Provenance Description Ontology	42
Abbildung 17: Item Description Ontology, Klasse Appearance	42
Abbildung 18: Item Description Ontology, Eigenschaft isGatheredInto.....	42
Abbildung 19: In PDO und IDO verknüpfte KOS.....	43
Abbildung 20: ASCH Application Profile, Klasse Event.....	45
Abbildung 21: ASCH Application Profile, Klasse Dimension.....	45
Abbildung 22: ASCH Application Profile, Properties description und end.....	46
Abbildung 23: Struktur der LIDO-Ausgangsdaten, eventActor	52
Abbildung 24: Struktur der LIDO-Ausgangsdaten, recordWrap	53
Abbildung 25: RDF/XML-Darstellung eines Records	56
Abbildung 26: Graph eines Records	56
Abbildung 27: Turtle-Darstellung eines Records.....	57
Abbildung 28: Arbeiten mit Oxygen	59
Abbildung 29: XSLT-Stylesheet, handle-URI.....	61

Abbildung 30: XSLT-Stylesheet, Objektgattung	61
Abbildung 31: XSLT-Stylesheet, Titel	62
Abbildung 32: XSLT-Stylesheet, Funktion lido:language()	62
Abbildung 33: XSLT-Stylesheet, Orte des Events	63
Abbildung 34: Skript batch_lido2rdfxml.sh	65
Abbildung 35: Ausschnitt aus Skript batch_rdfxml2ttl.sh	66
Abbildung 36: Fuseki Triplestore, Upload von zwei Datensätzen	70
Abbildung 37: Fuseki Triplestore, Info-Schirm	70
Abbildung 38: Fuseki SPARQL-Endpoint	71
Abbildung 39: SPARQL-Select nach den Bezeichnungen der Sammlungsobjekte	73
Abbildung 40: SPARQL-Select nach Event-Typen	73
Abbildung 41: SPARQL-Describe-Anfrage nach einem Record mit Ergebnisanzeige	74
Abbildung 42: SPARQL-Select mit Verknüpfung zur DBpedia	74

TABELLENVERZEICHNIS

Tabelle 1: Klassen des ASCH-Modells.....	38
Tabelle 2: RDF-Subjekte im Göttinger Sammlungsportal	48
Tabelle 3: Mapping-Tabelle, Objektgattung und Klassifikation	49
Tabelle 4: Mapping-Tabelle, Titel und Inschrift	50
Tabelle 5: Mapping-Tabelle, Typ und Akteur des Events.....	52
Tabelle 6: Mapping-Tabelle, Record	54

ABKÜRZUNGSVERZEICHNIS

AAT	Art and Architecture Thesaurus
ADABweb	Allgemeine Denkmaldatenbank webbasiert
AP	Application Profile
B3Kat	Gemeinsamer Verbundkatalog des Bibliotheksverbundes Bayern (BVB) und des Kooperativen Bibliotheksverbundes Berlin-Brandenburg (KOBV)
BARTOC	Basel Registry for Knowledge Organization Systems
BN	Blank Node
BVB	Bibliotheksverbund Bayern
CC	Creative Commons
CC0	Creative Commons - Public Domain
CC-BY	Creative Commons - Namensnennung
CC-SA	Creative Commons - Share Alike
CIDOC	Comité International pour la Documentation
CRM	Conceptional Reference Model
CSV	Comma-Separated Values
DANTE	Datendrehscheibe für Normdaten und Terminologien
DC	Dublin Core
DCMES	Dublin Core Metadata Element Set
DCMI	Dublin Core Metadata Initiative
DCTERMS	Dublin Core Metadata Initiative Metadata Terms
DDB	Deutsche Digitale Bibliothek
DNB	Deutsche Nationalbibliothek
DPLA	Digital Public Library of America
DSP	Description Set Profile
DWC	Darwin Core
EAD	Encoded Archival Description
EDM	Europeana Data Model
FLOWR	For, Let, Order-by, Where, Return
FOAF	Friend of a Friend
GBV	Gemeinsamer Bibliotheksverbund
GND	Gemeinsame Normdatei
GVK	Gemeinsamer Verbundkatalog
HBZ	Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICOM	International Council of Museums
ID	Identifier
IDO	Item Description Ontology

IIIF	International Image Interoperability Framework
ISIL	International Standard Identifier for Libraries and Related Organizations
ISLI	International Standard Link Identifier
ISO	International Organization for Standardization
JSON-LD	Javascript Object Notation-Linked Data
KENOM	Kooperative Erschließung und Nutzung der Objektdaten von Münzsammlungen
KOBV	Kooperativer Bibliotheksverbund Berlin-Brandenburg
KOS	Knowledge Organization System
LD	Linked Data
LIDO	Lightweight Information Describing Objects
LOD	Linked Open Data
LOV	Linked Open Vocabularies
MARC	Machine-Readable Cataloging
MB	Megabyte
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
NMN	Niedersächsischer Monographiennachweis
NMS	Nomisma Ontology
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OBG	Oberbegriffsdatei
OWL	Web Ontology Language
PDO	Provenance Description Ontology
PROV	Prov Ontology
PURL	Persistent Uniform Resource Locator
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RFC	Request for Comments
RIOT	RDF Input Output Technology
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol And RDF Query Language
SQL	Structured Query Language
TEI	Text Encoding Initiative
TEIC	Text Encoding Initiative Consortium
Turtle	Terse RDF Triple Language
UMAC	University Museums and Collections
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VZG	Verbundzentrale des GBV
W3C	World Wide Web Consortium

XML	Extensible Markup Language
XSL	Extensible Stylesheet Language
XSLT	XSL Transformation

DANKSAGUNG

Mein Dank geht zunächst an die Gruppe Metadaten der SUB Göttingen. Aus dieser Gruppe kam die Anforderung, Linked Open Data für das Göttinger Sammlungsportal zu erstellen und dabei das ASCH-Modell in die praktische Anwendung zu nehmen. Als Neueinsteigerin im Linked-Data-Kontext konnte ich mich mit allen Fragen an Stefanie Rühle, Susanne Al-Eryani, Uwe Sikora, Timo Schleier und Alex Jahnke wenden. Der XSLT-Einstieg ist mir dank der Einführung und guten Laune von Uwe Sikora ebenfalls leichtgefallen.

Es gibt weitere Kolleginnen und Kollegen in der SUB Göttingen, denen ich danken möchte. Zum einen meinem direkten Team in der Bereichsbibliothek Kulturwissenschaften, vor allem Petra Elbir, das mir stets den Rücken freigehalten hat, zum anderen Hans-Werner Hilse, mit dem ich viele inhaltliche Fragen diskutieren durfte. Dorothea Schuller danke ich für ihre hilfreichen Antworten auf meine Fragen zur Literaturverwaltung.

Außerdem gilt mein Dank der gesamten SUB Göttingen, die mir diese Art der Fortbildung ermöglicht und meine reduzierte Leistungsfähigkeit in der Zeit des Masterstudiums mit getragen hat.

Aus der Verbundzentrale des GBV möchte ich Frank Dührkohp und Tobias Helms danken, die den Server für den Triplestore gestellt haben.

Karsten Heck aus der Zentralen Kustodie der Universität Göttingen danke ich dafür, dass er mir die Zugänge in die Erfassungsportale freigeschaltet hat, sodass ich mich über die Quellen der zu konvertierenden Daten informieren konnte.

Die Informationen von Jana Hentschke von der Deutschen Nationalbibliothek zu den Themen Meta-Metadaten und Reification waren ebenso wertvoll wie die von Martin Stricker aus der Koordinierungsstelle für wissenschaftliche Sammlungen in Deutschland zum Stand von LOD in seinem Bereich. Beiden vielen Dank.

Insgesamt hat mir der Austausch mit den vielen kompetenten Kolleginnen und Kollegen, die ich im Rahmen der Masterarbeit näher kennenlernen durfte, sehr viel Spaß gemacht.

Petra Keidel von der TH Wildau gilt mein großer Dank. Sie war immer für mich ansprechbar und hat mich bestmöglich betreut.

Meine Masterarbeit hätte ich berufsbegleitend nicht schreiben können ohne meine Familie. Ich danke meinem Mann und unseren beiden Kindern von Herzen, dass sie sich selbst zurückgenommen und mich auf jede ihnen mögliche Weise unterstützt haben. Zuletzt gilt mein Dank meiner Mutter, die versucht hat, ihre Trauer vor mir zu verbergen und mich so zu entlasten.

1 EINLEITUNG

*A century ago, the resource in question was oil.
Now similar concerns are being raised by the giants that deal in data,
the oil of the digital era (The Economist 2017).*

Wir leben in einer netzgetriebenen Welt. Das Internet ist aus Wissenschaft, Kultur und Alltag nicht mehr wegzudenken. Auch museale Objekte werden zunehmend im Netz erlebbar, sei es durch hochauflösende Bilder oder detaillierte Informationen zu ihren Eigenschaften und ihrem Lebenszyklus. Die vorliegende Arbeit greift diesen Trend auf und liefert eine prototypische Anwendung zur besseren Sichtbarkeit und Verfügbarkeit musealer Sammlungen im Netz.

1.1 MOTIVATION UND ZIELE

Die Georg-August-Universität Göttingen beherbergt mehr als 40 heterogene Sammlungen, deren Bandbreite von der Algensammlung über die Schädelammlung bis hin zum Rechnermuseum oder der Humanembryologischen Sammlung reicht.¹ Die bessere Nutzbarkeit der Sammlungen für die Lehre, Forschung sowie für interessierte Bürgerinnen und Bürger der Stadt ist Teil des Zukunftskonzepts der Universität (Beisiegel et al. 2013, S. 7). Diesem Ziel folgt auch das 2017 online gestellte *Sammlungsportal*.² Es soll zukünftig sämtliche Objekte der Göttinger Sammlungen verzeichnen und anhand digitaler Reproduktionen für jedermann im Netz erlebbar machen (Beisiegel et al. 2013, S. 5).

Obwohl die im Sammlungsportal hinterlegten Informationen strukturiert und standardisiert sind, können sie gegenwärtig nicht optimal nachgenutzt werden. Sie befinden sich in einer Art Silohaltung, da sie nur innerhalb des Portals auffindbar sind. Außerdem ist eine zusätzliche Wissensgenerierung durch die automatische Vernetzung mit externen Informationen im Web nicht möglich. Es fehlen die Funktionalitäten, die auf den Technologien des Semantic Web beruhen, das die maschinelle inhaltliche Einordnung und Wiederverwendbarkeit der Informationen vorsieht.

Diese Lücke soll mit der vorliegenden Arbeit geschlossen werden. Es soll ein Pool mit RDF Statements, ein sogenannter *Knowledge Graph*, bereitgestellt werden, der die Informationen zu den Objekten des Sammlungsportals global sichtbar und verknüpfbar macht.

¹ Liste sämtlicher Sammlungen: <https://sammlungen.uni-goettingen.de/sammlungen> (15.01.2019).

² <https://sammlungen.uni-goettingen.de> (15.01.2019).

Einen weiteren Anreiz für die Beschäftigung mit dem Thema stellt das ASCH-Modell³ dar, ein von 2014 bis 2017 an der Universität Göttingen entwickeltes RDF-Datenmodell für die Beschreibung von heterogenen musealen Objekten und ihren Digitalisaten. Das ASCH-Modell bildet die Basis für die Generierung von Linked Open Data im Rahmen dieser Arbeit. Mit der Konvertierung der wesentlichen Informationen in das neue Format wird gleichzeitig der Machbarkeitsbeweis für das ASCH-Modell angetreten.

Schließlich üben die Themen *Daten* und *Semantic Web* einen besonderen Reiz aus. Der Terminus *Big Data* ist allgegenwärtig. Heath und Bizer sprechen von einem „Data Deluge“ und fühlen sich „surrounded by data“ (2011, S. 1). Vielfach werden Daten als „Öl des 21. Jahrhunderts“ (Spitz 2017) bezeichnet. Gleichzeitig sind in Deutschland nicht viele wissenschaftliche Sammlungen auszumachen, die sich mit Linked-Data-Anwendungen beschäftigen.⁴ Mit dem Datenbestand des Sammlungsportals und dem ASCH-Modell besteht die Möglichkeit, sich dem Thema auf eine sehr praxisbezogene Weise zu nähern. Es soll eine Herangehensweise und Umsetzung aufgezeigt werden, die von kleineren musealen Einrichtungen ohne weitreichende Semantic-Web-Erfahrung nachgenutzt werden kann, um die eigenen Daten als Linked Data zu veröffentlichen.

Die Entwicklung und Bereitstellung der Daten sollen den *Linked Data Principles* (Berners-Lee 2006/2010) sowie den vom *World Wide Web Consortium (W3C)*⁵ veröffentlichten *Best Practices for Publishing Linked Data* (Hyland et al. 2014) entsprechen.

Anstatt eine einzelne der über 40 Sammlungen mit sämtlichen Spezifika zu betrachten, wird ein Ansatz verfolgt, der sämtliche Datensätze und Objekte des Sammlungsportals zu berücksichtigen versucht und sich auf die gemeinsamen und sinnvoll verwertbaren Informationen konzentriert. Dieses generelle Erkenntnisinteresse ermöglicht gleichsam die Übertragung der Verfahrensweise auf eine Vielzahl von musealen Sammlungen.

Die vielen Mapping-Verfahren vorausgehende Datenbereinigung ist nicht Gegenstand dieser Arbeit. Zum einen würde die Beschäftigung mit dem Thema Datenbereinigung den Rahmen der Arbeit überschreiten, zum anderen ergab die Analyse der Ausgangsdaten ihre hinreichende Qualität.

³ <http://asch.wiki.gwdg.de> (15.01.2019).

⁴ *CodingDaVinci*, der deutsche Hackathon für Kulturdaten, weist unter <https://codingdavinci.de/daten/> (15.01.2019) keine universitätsweite wissenschaftliche Sammlung im RDF-Format nach. Laut Martin Stricker von der *Koordinierungsstelle für wissenschaftliche Sammlungen in Deutschland* entwickeln gegenwärtig einige wenige universitäre Sammlungen RDF-Daten, gleichzeitig sei das konkrete Angebot noch sehr begrenzt. Er nennt das *KOSMOS-Projekt* der Universität Bonn (<https://kosmos.uni-bonn.de>, 15.01.2019) und *Objekte im Netz* der Universität Erlangen-Nürnberg (<http://objekte-im-netz.fau.de/projekt>, 15.01.2019).

⁵ <https://www.w3.org/> (15.01.2019).

1.2 AUFBAU DER ARBEIT

Der Aufbau dieser Arbeit folgt im Wesentlichen den einzelnen Arbeitsschritten, die es bis zur Bereitstellung der RDF-Daten des Sammlungsportals der Georg-August-Universität Göttingen auszuführen gilt.

Zunächst wird in den Kapiteln 2 und 3 das theoretische Fundament der Arbeit gelegt. Es dient als Unterstützung für die konkreten Überlegungen und Vorgehensweisen in den sich anschließenden Kapiteln.

Kapitel 2 beschäftigt sich mit den grundlegenden Prinzipien und Technologien des Semantic Web. Zentrale Begriffe wie *Linked Data*, *Uniform Resource Identifier (URI)* und *Knowledge Graph* finden ebenso Berücksichtigung wie das zugrundeliegende Datenmodell *Resource Description Framework (RDF)*. Der Abschnitt *Generierung und Modellierung von Wissen* widmet sich der Wissensanreicherung im Web. Es wird beschrieben, wie die inhaltliche Einordnung der Informationen erfolgt. *Knowledge Organization Systems* und Schema-Sprachen spielen in diesem Kontext eine wichtige Rolle. Schließlich gibt der letzte Abschnitt des zweiten Kapitels einen Überblick über die Möglichkeiten der Veröffentlichung und Suche von Linked Data. In diesem Zusammenhang findet auch das Thema *Open Data* eine kurze Betrachtung.

Das dritte Kapitel nimmt den Kulturerbesektor in den Blick und legt dabei seinen Schwerpunkt auf die wissenschaftlichen Sammlungen. Dies liegt nahe, da das Göttinger Sammlungsportal die Objekte wissenschaftlicher Sammlungen verzeichnet. Des Weiteren werden die Systeme, die beispielgebend für das Sammlungsportal sind und auf deren Datenerfassung es zu großen Teilen beruht, näher betrachtet. Auf diese Weise kommen zum einen für den Kulturerbesektor typische Aspekte zur Sprache, zum anderen wird eine Grundlage für die Analyse der Ausgangsdaten in Kapitel 4 geschaffen. Zuletzt beschäftigt sich das dritte Kapitel mit dem LIDO-Datenformat. LIDO ist ein weit verbreiteter Standard im musealen Bereich (Simou 2014, S. 3; Hagedorn-Saupe 2012, S. 205). Außerdem ist es das Primärformat des Sammlungsportals.

Daran anschließend gilt es, in Kapitel 4 die Ausgangssituation und in Kapitel 5 das Zielformat kennenzulernen, um die realistischen Möglichkeiten der Zielerreichung auszuloten.

Das Göttinger Sammlungsportal stellt den Ausgangspunkt dar. Es ist wichtig, die Ziele und den Kontext des Portals zu verstehen, damit die relevanten Informationen bei der Konvertierung ausgewählt werden können. Aus demselben Grund erfährt die Benutzeroberfläche des Portals eine kurze Betrachtung. Den weitaus größten Teil des vierten Kapitels nimmt die Beschäftigung mit den Daten ein. Dabei werden zunächst ihre Herkunft und ihr Weg in das

Sammlungsportal zusammengefasst. Daraus ergeben sich Rückschlüsse auf die Beschaffenheit der Daten. Außerdem wird der Frage nachgegangen, ob es neben der Entwicklung eines eigenen Mappings alternative Möglichkeiten gibt, Linked Open Data für das Sammlungsportal zu erhalten. Der Unterabschnitt 4.2.2 nimmt Struktur (Syntax) und Inhalt (Semantik) der Ausgangsdaten in den Blick. Es wird eine kompakte Beschreibung des Ausgangsformats geliefert. Des Weiteren werden wichtige Eckpunkte und Aspekte benannt, anhand derer die Datenanalyse erfolgt. Schließlich finden die hinzugezogenen Tools Berücksichtigung.

Das Thema des fünften Kapitels ist das ASCH-Modell, das die Basis für das Zielformat darstellt. Nach Informationen zur Entstehung und dem Zweck des Modells werden die wichtigsten Klassen sowie die verwendeten Ontologien skizziert. Schließlich stehen die *Application Profiles (APs)* im Fokus, in denen das RDF-Vokabular der *Asch-Domain*, und somit auch dieser Arbeit, zusammengestellt ist. Dabei erfolgt eine Konzentration auf die APs *Event* und *Item*.

Das sechste Kapitel, das sich mit der Umwandlung der Daten beschäftigt, ist in drei Abschnitte unterteilt. Zunächst gilt es, die Vorgaben für das Mapping zu erarbeiten. Das Vorgehen wird an wenigen Beispielen dokumentiert; die gesamte Mapping-Tabelle ist in Anhang 1 enthalten. Anhand dieser Vorgaben kann anschließend im zweiten Abschnitt des Kapitels die tatsächliche Umwandlung der Daten in den Blick genommen werden. Dazu ist zunächst das RDF-Zielformat festzulegen. Hier fiel die Entscheidung auf RDF/XML und – in einem zweiten Schritt – auf Turtle. Die im gegebenen Kontext wichtigen Merkmale der Formate werden vorgestellt. Im Anschluss steht der XSLT-Code im Fokus, das Stylesheet, das die Umwandlung der Daten realisiert. Dabei wird nach einer knappen allgemeinen Einführung in XSLT das Vorgehen bei der Entwicklung des Codes beschrieben. Außerdem werden beispielhafte Passagen aus dem Code betrachtet. Der zweite Abschnitt schließt mit der Beschreibung der Batchverarbeitung, also der Anwendung des Umwandlungsprogramms auf alle Dokumente des Sammlungsportals. Der dritte Abschnitt des Kapitels wirft einen Blick auf die Ergebnisse der Konvertierung und beschreibt die Testverfahren.

Ein letzter wichtiger Schritt ist die Bereitstellung der Daten, die für das Sammlungsportal in Form eines *Fuseki Triplestores* erfolgt. Kapitel 7 beschreibt den Aufbau des Triplestores und liefert erste Beispiele für dessen Abfrage mit der *SPARQL Protocol and RDF Query Language (SPARQL)*.

Die im Rahmen der vorliegenden Arbeit abgebildete und umgesetzte Generierung von Linked Open Data wird schließlich im letzten Kapitel bewertet. Die Arbeit endet mit einem Ausblick auf mögliche Schritte, die auf den erzielten Ergebnissen aufbauen und diese weiterentwickeln.

2 SEMANTIC WEB

*I have a dream for the Web [...].
Machines become capable of analyzing all the data on the Web -
the content, links and transactions between people and computers.
A 'Semantic Web' (Berners-Lee und Fischetti 1999, S. 157).*

Das *Semantic Web* stellt eine Erweiterung des herkömmlichen Webs dar. Während Letzteres, das *Web of Documents*, ganze Dokumente über Hyperlinks verknüpft, ohne die Art der Verknüpfung zu spezifizieren, werden im Semantic Web einzelne Dinge, etwa Personen, Orte oder abstrakte Konzepte, miteinander verknüpft. Durch die Beschreibung der Verbindungen zwischen den Dingen wird dem Netz Struktur und Logik hinzugefügt. Auf diese Weise werden die Informationen maschinenlesbar (Berners-Lee et al. 2001, S. 34).

Heath und Bizer bescheinigen dem Semantic Web, das auch als *Web of Data* bezeichnet wird, „a revolutionary opportunity for deriving insight and value from data“ (2011, S. 5). Im Web of Data seien nicht nur Dokumente, sondern auch Daten „a first class citizen“ (2011, S. VI).

2.1 LINKED DATA UND URIS

Die Idee vernetzter, wiederverwertbarer und maschinenlesbarer Daten ist eng verknüpft mit dem Term *Linked Data*. Linked Data steht für eine Reihe von Technologien, die die semantische Verknüpfung von Daten im World Wide Web möglich machen (Bizer et al. 2009, S. 15).

Die *Linked Data Principles* wurden von Tim Berners-Lee, dem Begründer des World Wide Web und Vorsitzenden des W3C, eingeführt (2006/2010):

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

URIs sind eindeutige globale Identifikatoren für jedes einzelne im Web beschriebene Ding. Nicht nur Webseiten und digitale Ressourcen bekommen im Semantic Web einen URI, sondern sämtliche reale Dinge und abstrakte Konzepte. Bezogen auf den Kontext dieser Arbeit benötigt zum Beispiel jedes Sammlungsobjekt, jede Sammlung, jeder Ort und jede Person einen eigenen eindeutigen URI.

URIs folgen einem definierten Schema⁶. Außerdem ändern sie sich nicht. Berners-Lee bezeichnet persistente, stabile URIs als „cool URIs“ (1998). Das W3C reglementiert in einem eigenen Dokument das Design und den Umgang mit URIs. Sie sollen einfach, stabil und handhabbar sein (Sauermann und Cyganiak 2008).⁷

URIs dienen nicht nur zur eindeutigen Identifizierung von Dingen, sondern können, sofern sie HTTP-URIs sind, zum Auffinden und Abrufen von Web-Dokumenten dienen, die diese Dinge beschreiben (Auer et al. 2011, S. 3).

Dazu müssen sie dereferenzierbar sein, was bedeutet, dass HTTP-Clients den URI aufrufen und die Beschreibung der identifizierten Ressource anzeigen können (Auer et al. 2011, S. 6).

Grundsätzlich lassen sich im Semantic-Web-Kontext zwei Strategien für dereferenzierbare URIs unterscheiden: 303 URIs und Hash URIs.

Beim Aufruf einer 303 URI antwortet der Server mit dem HTTP Response Code 303 (see other) und der URI des Web-Dokuments, das das *Real World Object* beschreibt. Diese wird vom Client dann in einem zweiten Schritt dereferenziert. Durch diesen sogenannten *303 Redirect* kann zwischen dem *Real World Object* und seiner Beschreibung unterschieden werden. Gleichzeitig haben Clients die Möglichkeit, die Beschreibung des *Real World Objects* über seinen URI aufzurufen.

Die nachfolgende Abbildung beschreibt die einzelnen Teile eines Hash URIs, der dadurch gekennzeichnet ist, dass am Ende, durch ein Hash eingeleitet, ein spezieller Teil, das *Fragment*, enthalten ist. Durch das Fragment wird der URI zu einem eindeutigen Identifikator für ein *Real World Object*.

Abbildung 1: Aufbau eines URIs (Jahnke 2017, Folie 14)

Beim Aufruf der URI schneidet das HTTP Protokoll das Fragment ab und liefert die Beschreibung für sämtliche Objekte mit demselben Nicht-Fragment-Teil. Auf diese Weise wird ebenfalls die Eindeutigkeit der Identifikatoren für die *Real World Objects* und ihre Beschreibung gewährleistet (Sauermann und Cyganiak 2008).

⁶ Erstmals im RFC 1630 (Berners-Lee 1994) definiert und aktualisiert im RFC 3986 (Berners-Lee et al. 2005).

⁷ Neben den in diesem Abschnitt genannten Quellen finden sich weitere Informationen zu URIs bei Heath und Bizer 2011, S. 5–19. Für einen kurzen Überblick siehe Dimou 2017, S. 1.4.

Darüber hinaus können verschiedene Repräsentationen einer Beschreibung über *Content Negotiation* aufgerufen werden. Content Negotiation ist ein HTTP-Mechanismus, der es ermöglicht, zwischen verschiedenen Repräsentationen eines Dokuments auszuwählen, ohne dedizierte URIs zu vergeben. Die Idee besteht darin, dass HTTP-Clients mit jeder Anfrage HTTP-Headers verschicken, in denen das zurückzuliefernde Format vermerkt ist. Server lesen den Header und liefern entsprechend zurück (Fielding et al. 2014).

Der zweiten Forderung der *Linked Data Principles* folgend handelt es sich bei dem URI in Abbildung 1 um einen HTTP-URI, gekennzeichnet durch das Schema HTTP am Anfang.

Der dritte Punkt der *Linked Data Principles* verlangt die Bereitstellung standardisierter Informationen durch den URI. Diese Forderung ist eng verknüpft mit dem Modell, auf dem die Veröffentlichung von Linked Data im Netz beruht, dem Resource Description Framework.⁸

Das vierte Prinzip geht über den eigenen Datenpool hinaus und empfiehlt die Einbindung von externen URIs zur globalen automatisierten Wissensgenerierung und -erweiterung.

2010 ergänzt Berners-Lee seine *Linked Data Principles* um ein *Star Rating System* (2006/2010). Durch die Vergabe von bis zu fünf Sternen für die Qualität der Daten soll ein Anreiz für die Veröffentlichung von hochwertigen und offenen Daten geschaffen werden:

- ★ *Available on the web (whatever format) but with an open licence, to be Open Data*
- ★★ *Available as machine-readable structured data (e.g. excel instead of image scan of a table)*
- ★★★ *as (2), plus non-proprietary format (e.g. CSV instead of excel)*
- ★★★★ *All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff*
- ★★★★★ *All the above, plus: Link your data to other people's data to provide context*

Die *Linked Data Principles* sowie das *Star Rating System* stellen gewichtige Kriterien für die im Zuge dieser Arbeit zu generierenden Daten dar.

2.2 RESOURCE DESCRIPTION FRAMEWORK

Das Resource Description Framework (RDF)⁹ ist das Datenmodell, auf dem das Semantic Web beruht. Jedes im RDF beschriebene konkrete Ding oder abstrakte Konzept ist als *Resource* definiert. Es werden Aussagen über Resources in Form sogenannter *Triples* gemacht.

⁸ Siehe unten Kap. 2.2.

⁹ Umfassende Einführungen in das Thema liefern der RDF Primer (Manola und Miller 2004) und dessen Erweiterung und Aktualisierung (Schreiber und Raimond 2014).

Ein Triple, auch als *RDF Statement* bezeichnet, besteht aus den drei Teilen *Subjekt*, *Prädikat* und *Objekt*.

Abbildung 2: RDF-Triple

Die Elemente Subjekt und Prädikat werden stets durch einen URI repräsentiert. Das Objekt kann auch ein *Literal*, eine Zeichenfolge, sein. Literale können durch die Attribute *Sprache* und *Datentyp* spezifiziert werden.

Neben den Elementen URI und Literal kann ein Triple ein weiteres Element enthalten, einen *Blank Node*. Blank Nodes sind Knoten oder *Resources*, die (noch) keinen URI besitzen. Sie dienen als URI-Platzhalter oder Konnektor und sind nicht weltweit, sondern nur im originären Datenbestand adressierbar. Die Funktion eines Blank Nodes wird in Abbildung 3 dargestellt.

Abbildung 3: Blank Node (Gandon und Schreiber 2014, Figure 1)

Der URI *http://www.w3.org/TR/rdf-syntax-grammar* hat den Titel *RDF 1.1 XML Syntax* in Form eines Literals sowie einen Herausgeber, der in diesem Fall keinen URI hat und stattdessen von einem Blank Node repräsentiert wird. Dieser Blank Node wiederum ist das Subjekt für die beiden Triples *Blank Node -> fullName -> Dave Beckett* und *Blank Node -> homePage -> http://purl.org/net/dajobe*.

Um die RDF-Informationen in einer sequentiellen Darstellungsform abzubilden, gibt es zahlreiche RDF-Serialisierungsformate.

In dieser Arbeit ist *RDF/XML*¹⁰ das primäre Zielformat, in das die Ausgangsdaten des Sammlungsportals zu konvertieren sind.¹¹ In *RDF/XML* werden die einzelnen Bestandteile der Triples als XML-Elemente abgebildet. Außerdem findet das Serialisierungsformat *Terse RDF Triple Language (Turtle)*¹² Anwendung. Turtle ist ein sehr einfaches, für Menschen gut lesbares *RDF*-Format und eignet sich deshalb zur manuellen Qualitätskontrolle der *RDF* Statements.¹³ Zur besseren Übersichtlichkeit der Serialisierungen können URIs nach dem Prinzip der XML-Namensräume abgekürzt werden. Das Präfix und der Teil des URIs, den es repräsentiert, sind zu Beginn eines Dokuments einmalig zu definieren.¹⁴

2.3 KNOWLEDGE GRAPH

Jedes *RDF* Statement ist auch ein *Graph*, der aus zwei *Knoten* besteht, die über eine *Kante* miteinander verbunden sind. Der ausgehende Knoten ist das Subjekt, die gerichtete Kante das Prädikat und der zweite Knoten das Objekt, das von neuem das Subjekt weiterer Statements sein kann. Auf diese Weise können weit verzweigte Netze entstehen. Ist das Objekt ein Literal, ist keine weitere Verknüpfung an dieser Stelle möglich. Literale bilden stets Endpunkte. Werden *Resources* über URIs identifiziert und Vokabulare und Ontologien wiederverwendet, verbinden sich die *RDF* Statements der unterschiedlichen Quellen automatisch zu einem „giant global graph“ (Oren und Schenk 2011, S. 82).

Der Begriff *Knowledge Graph* ist vor allem in Zusammenhang mit Google bekannt. Die Bezeichnungen *Google Knowledge Graph*¹⁵ und *Knowledge Graph* werden häufig synonym verwendet.¹⁶ Gleichzeitig ist jede aus *RDF* Statements bestehende Datenbasis ein *Knowledge Graph*.

2.4 GENERIERUNG UND MODELLIERUNG VON WISSEN

In den vorangegangenen Abschnitten wurden die Technologien und Ziele des Semantic Web

¹⁰ Spezifikation *RDF/XML*: Gandon und Schreiber 2014.

¹¹ Zu *RDF/XML* und Turtle siehe unten Kap. 6.2.1.

¹² Spezifikation Turtle: Prud'hommeaux und Carothers 2014.

¹³ Als weitere *RDF*-Serialisierungsformate seien N-Triples (Beckett 2014), JSON-LD (Sporny et al. 2014), TriG (Bizer und Cyganiak 2014) und N-Quads (Carothers 2014) genannt.

¹⁴ Einen Lookup-Dienst für *RDF* Namensräume bietet <http://prefix.cc/> (19.01.2019).

¹⁵ Der 2012 eingeführte *Google Knowledge Graph* bedient sich der Semantic-Web-Technologien, um die Suchergebnisse zu Menschen, Orten oder Sachverhalten um weitere Fakten oder Bilder aus externen Quellen, zumeist aus *Wikidata*, anzureichern. Diese zusätzlichen Informationen werden in einem eigenen Kasten rechts (Desktop und Tablet) oder oberhalb (Smartphone) der eigentlichen Suchergebnisse angezeigt.

¹⁶ *Wikipedia* führt die Informationen zum *Google Knowledge Graph* zum Beispiel unter dem Eintrag *Knowledge Graph*.

beschrieben. Im Folgenden wird der Frage nachgegangen, wie die kontextuelle Einordnung der Daten erfolgt. Es geht beispielsweise darum, wie maschinenlesbar und eindeutig ausgedrückt werden kann, dass es sich bei dem Term *Golf* um den Sport und nicht um das Auto oder die Landschaft handelt.

Knowledge Organization Systems (KOS) wie Vokabulare und Ontologien liefern Antworten auf diese Fragen. Es existiert keine klare Abgrenzung zwischen den beiden Begriffen, häufig werden sie synonym verwendet (W3C 2015). Dimou grenzt die beiden Begriffe wie folgt voneinander ab: Ein Vokabular lege eine feste Bedeutung für eine Menge von zentralen Termen fest. Häufig seien Beschreibungen enthalten, in welchem Verhältnis die Terme zueinander stünden, zum Beispiel in Form von Über- und Unterordnungen. Der Begriff Ontologie werde vielfach in Zusammenhang mit präzisen Aussagen über einen Ausschnitt der Welt verwendet. Es würden die in diesem Ausschnitt der Welt existierenden Dinge, die Verbindungen zwischen ihnen sowie die logischen Schlüsse beschrieben, die sich aus diesen Beziehungen ableiten ließen (2017, S. 1.10–1.11).

Nachfolgende Definition liefert das W3C (W3C OWL Working Group 2012):

Ontologies are formalized vocabularies of terms, often covering a specific domain and shared by a community of users. They specify the definitions of terms by describing their relationships with other terms in the ontology.

Liegen diese Ontologien maschinenlesbar vor, können sie für das sogenannte *Reasoning* oder *Inferencing* hinzugezogen werden. Dabei wird neues Wissen auf Basis der bestehenden Triples und in Kombination mit den darin genutzten Ontologien generiert. Die KOS bilden in dem Fall die Grundlage für den Wissenszugewinn. Wenn beispielsweise eine *Resource* zu einer Klasse gehört, die einer Oberklasse zugeordnet ist, so gehört sie automatisch auch dieser Oberklasse an (Pohl 2013, S. 240).

Die Art und Weise, wie diese Definitionen und Spezifikationen standardisiert und maschinenlesbar in den Ontologien abzubilden sind, regeln Schema- oder Beschreibungssprachen wie *RDF Schema (RDFS)*¹⁷, *Simple Knowledge Organization System (SKOS)*¹⁸ oder *Web Ontology Language (OWL)*¹⁹.

Während SKOS eine einfache Veröffentlichung und Kombination kontrollierter, strukturierter und maschinenlesbarer Vokabulare für das Semantic Web ermöglicht, dienen OWL und RDFS vor allem der Beschreibung von *Domain*-Modellen mit Klassen und Eigenschaften und ihren Beziehungen untereinander. Heath und Bizer sehen OWL als eine Erweiterung von RDFS an,

¹⁷ Spezifikation RDFS: Brickley und Guna 2014.

¹⁸ Spezifikation SKOS: Miles und Bechhofer 2009.

¹⁹ Spezifikationen OWL: Bechhofer et al. 2004; W3C OWL Working Group 2012.

weil sie zusätzliche grundsätzliche Modellierungsterme wie `owl:equivalentClass` und `owl:equivalentProperty` einführt (2011, S. 56–60).

Die beiden Hauptklassen der RDFS-Sprache sind `rdfs:Class` und `rdfs:Property`. Jedes konkrete Ding oder abstrakte Konzept, das im RDF beschrieben wird, ist eine `rdfs:Resource`. Die Mitglieder einer Klasse werden Instanzen genannt. `rdfs:Resource` und `rdfs:Property` sind Instanzen von `rdfs:Class`. Properties werden auch als Eigenschaften bezeichnet, weil sie die Art der Verbindung zwischen Subjekt und Objekt in einem RDF Statement beschreiben und als Prädikat des Triples fungieren. Instanzen und Unterklassen erben alle Eigenschaften ihrer Hauptklasse. Dass eine *Resource* zu einer bestimmten Klasse gehört, wird mit der Property `rdf:type` ausgedrückt.²⁰

Schemasprachen machen also die Aussagen global verständlich und maschinenlesbar. Auf Basis der gemeinsamen Ansätze können unterschiedliche Quellen vernetzt und die Triples um Logik und Kontext angereichert werden.

Ein grundsätzliches Verständnis von RDFS und OWL ist nicht nur für das Design von Ontologien wichtig, sondern ebenso für ihre korrekte Anwendung. Mit Properties wie `rdfs:range` und `rdfs:domain` wird beispielsweise festgelegt, welche Daten grundsätzlich zu einem Triple zusammengesetzt werden dürfen. Diese Vorgaben muss jedes Mapping zu RDF berücksichtigen.

“Ontologies are ten a penny nowadays and the underlying concepts are as diverse as the knowledge fields they represent” (Al-Eryani und Rühle 2018, S. 3). Suominen und Hyvönen sprechen von „Wildwuchs bei den Datenmodellen“ (2017, S. 1). Dieser fortschreitenden Verbreitung von Ontologien, Konzepten und Modellen begegnet gegenwärtig eine Reihe von Vokabularsuchmaschinen wie zum Beispiel *Linked Open Vocabularies (LOV)*²¹ und *Basel Registry for Knowledge Organization Systems (BARTOC)*²². Gleichzeitig etablieren sich Dienste, die Vokabulare zentral verwalten und über öffentliche Schnittstellen bereitstellen. Als Beispiele seien *xTree*²³ von *DigiCULT*²⁴ sowie die *Datendrehscheibe für Normdaten und Terminologien (DANTE)*²⁵, ein Vokabularserver der *Verbundzentrale des GBV (VZG)*²⁶, genannt.

Trotz der Vielzahl an Ontologien kann das Design neuer Konzepte sinnvoll sein. Al-Eryani und Rühle präferieren zwar die Nachnutzung vorhandener und weit verbreiteter Ontologien, da

²⁰ Klassen sind in der Regel an einem großen, Properties an einem kleinen Anfangsbuchstaben auszumachen.

²¹ <http://lov.okfn.org/dataset/lov/> (15.01.2019).

²² <http://bartoc.org/> (15.01.2019).

²³ <http://xtree-public.digicult-verbund.de/vocnet/> (15.01.2019).

²⁴ <https://www.digicult-verbund.de/de> (15.01.2019).

²⁵ <https://dante.gbv.de/search> (15.01.2019).

²⁶ <https://www.gbv.de/> (15.01.2019).

sie die besten Möglichkeiten für die globale Vernetzung und Interoperabilität böten, gleichzeitig nennen sie Gründe für das Design eigener Ontologien:

Unfortunately, experience has shown that the available ontologies often only partly meet the requirements of particular projects ranging from minor differences to the incompatibility of relevant parts. In addition, new knowledge fields arise and the cross-domain description of knowledge fields becomes more and more important. Such shifting and changing of domains may require a new way of modelling the relations between things in a certain domain. In these cases, designing a new ontology may be inevitable (Al-Eryani und Rühle 2018, S. 3).

Der Wert einer neuen Ontologie und ihre Nachnutzung hängen von der Qualität der Dokumentation in zum Beispiel RDFS oder OWL und der maschinenlesbaren Veröffentlichung ab.

Der Generierung von Wissen im Linked-Data-Kontext geht die Modellierung des jeweiligen Wissensraums, der sogenannten *Domain*, voraus. Es wird festgelegt, welche Klassen mit welchen Eigenschaften vorkommen können und welche Verbindungen zwischen ihnen bestehen. Für diese Definitionen können unterschiedliche Ontologien hinzugezogen werden. Dieses *Mixing and Matching* verschiedener Ontologien mit dem Ziel, den jeweiligen Ausschnitt der Welt bestmöglich beschreibbar zu machen, ist wesentlich für Application Profiles (Al-Eryani und Rühle 2018, S. 6). Application Profiles stellen die für den konkreten Anwendungsfall relevanten Terme zusammen. „The term application profile has been applied to describe the tailoring of standards for specific applications“ (Nilsson et al. 2008). Damit wird zum einen die Generierung neuer Daten in einheitlicher Form erleichtert, zum anderen wird eine Dokumentation zur Verfügung gestellt, die die Nachnutzung der Daten möglich macht. Außerdem können Application Profiles zur Validierung benutzt werden, sofern sie maschinenlesbar vorliegen (Al-Eryani und Rühle 2018, S. 6). Das *Singapore Framework for Dublin Core Application Profiles* (Nilsson et al. 2008) gibt Empfehlungen für das Design von Application Profiles, um maximale Interoperabilität und die Wiederverwertbarkeit der Daten zu erzielen.

2.5 VERÖFFENTLICHUNG VON LINKED DATA

Das Hauptziel von Linked Data ist die Weiterverwendung und Verknüpfung mit externen Daten überall auf der Welt in einem globalen Netz.

Im Folgenden werden drei Möglichkeiten zur Veröffentlichung von RDF-Triples vorgestellt:

Bei der Dereferenzierung werden einzelne Beschreibungssätze in einem oder mehreren RDF-Serialisierungsformaten als Webresource vorgehalten. Ein Beispiel liefert der *Linked-Data-*

*Service der Deutschen Digitalen Bibliothek*²⁷. Jeder Datensatz liegt auch in RDF-Formaten vor, die über die Webseite aufrufbar sind.²⁸

Eine zweite Möglichkeit ist der Data Dump, der den gesamten Datenbestand in einem RDF-Serialisierungsformat zur Verfügung stellt. Aktualisierungen werden in der Regel als separate Dateien abgelegt. Als Beispiel sei der *LinkedOpenData-Service* des *B3Kat*²⁹ genannt.

Eine dritte Methode, Linked Open Data zu veröffentlichen, ist der *SPARQL Endpoint*. In diesem Fall werden beliebig viele Triples in einen sogenannten *Triplestore* geladen, der über das Web frei zugänglich ist. Die Einspielung weiterer Triples ist jederzeit möglich. Ein Triplestore ist eine auf den Umgang mit RDF-Triples optimierte Datenbanksoftware.³⁰

Die Abfragesprache, mit der in einem Triplestore gesucht wird, ist die *SPARQL Protocol and RDF Query Language (SPARQL)*³¹. Es handelt sich dabei um eine Sprache, mit der man Graphen auf das Vorhandensein bestimmter Teile und Muster hin abfragen kann. Die Anwendung von SPARQL setzt Kenntnisse ihrer komplexen Syntax sowie der Struktur der Daten voraus.

Im Unterschied zu den beiden zuerst genannten Veröffentlichungsmethoden, die die Daten entweder eines einzelnen Datensatzes oder eines ganzen Datensets zur Verfügung stellen, bietet der Triplestore zusätzlich eine Abfrageumgebung.

Im Verlauf dieser Arbeit sollen die Triples des Göttinger Sammlungsportals in einem Triplestore abgelegt werden.³²

Linked Data kann sein ganzes Potential nur als *Linked Open Data (LOD)* entfalten, wenn die Daten offen für die Allgemeinheit sind und nachgenutzt werden dürfen. Die Sichtbarkeit der eigenen Daten wird signifikant gesteigert, sobald fremde Anwendungen sie leicht und ohne rechtliche Hürden integrieren können und die Daten unter einer offenen Lizenz zur Verfügung gestellt werden.³³

²⁷ http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedd_data_node.html (15.01.2019).

²⁸ Der Link zum Datensatz, der die Universität Göttingen beschreibt, lautet beispielsweise <http://d-nb.info/gnd/2024315-7> (15.01.2019). Das Rückgabeformat bestimmt sich über HTTP Content Negotiation. Der ausdrückliche Link zur zugehörigen Turtle-Repräsentation lautet <http://d-nb.info/gnd/2024315-7/about/lds> (15.01.2019).

²⁹ <https://lod.b3kat.de/doc/de/startseite/> (15.01.2019).

³⁰ Eine Übersicht über verschiedene Softwarelösungen zum Aufbau eines Triplestores findet sich bei solid IT 2019a.

³¹ Spezifikationen SPARQL: Prud'hommeaux und Seaborne 2008; Harris und Seaborne 2013.

³² Siehe unten Kap. 7.

³³ Die Open Access Bewegung setzt sich für die leichte Zugänglichkeit und Nachnutzbarkeit der Daten ein. Weiterführende Informationen finden sich bei Dreßen 2018 und Gerdes 2018.

Die gemeinnützige Organisation *Creative Commons (CC)*³⁴ hat unterschiedliche Lizenzen entwickelt, die die Nachnutzbarkeit der Daten auf einfache Weise regeln und dem Urheber der Daten gleichzeitig Rechtssicherheit bieten. So kann der Urheber beispielsweise mit der Lizenz *CC0* auf alle Rechte verzichten oder mit *CC-BY* seine Nennung bei Weiterverwertung festlegen. Die *CC-SA*³⁵-Lizenz regelt die Weitergabe der Daten unter gleichen Bedingungen.

Als Beispiele für Datensets in der LOD-Cloud seien *DBpedia*³⁷, die ihre Daten aus strukturierten Einträgen der Wikipedia speist, und *Wikidata*³⁸, ein Projekt von *Wikimedia Deutschland*, das Daten für jedes Wikipedia Projekt liefert, genannt. In beiden Fällen handelt es sich um *Crossdomain Knowledge Graphs*, die ihre Daten unter einer CC-Lizenz zur Verfügung stellen und SPARQL-Endpoints³⁹ für die Suche anbieten (Färber et al. 2018, S. 82–83).⁴⁰

⁴⁰ Weitere Informationen zur LOD-Cloud finden sich unter <http://lodlaundromat.org/wardrobe/> (15.01.2019).

3 KULTURERBESEKTOR

*Heritage is our legacy from the past,
what we live with today, and what we pass on to future generations.
Our cultural and natural heritage are both irreplaceable sources of life and inspiration.
They are our touchstones, our points of reference, our identity (UNESCO 2008, S. 5).*

Museale Objekte sind authentische und einzigartige Zeugnisse der Vergangenheit (Dührkohp 2015). Während sie lange Zeit vor allem im Rahmen von Ausstellungen der einzelnen Kultureinrichtungen vor Ort erlebbar waren oder der Zugang aus konservatorischen Gründen nur einem begrenzten Personenkreis vorbehalten war, eröffnet der digitale Wandel neue Möglichkeiten der Zugänglichkeit. Durch die Bereitstellung von digitalen Repräsentationen der Sammlungsobjekte im Internet werden diese weltweit erlebbar (Chowdhury und Ruthven 2015b, S. XV).

Linked Open Data bietet Kulturinstitutionen zusätzliche Möglichkeiten, die eigenen Daten auffindbar, verfügbar und nachnutzbar zu machen und sie durch Informationen aus externen Quellen anzureichern (Oomen et al. 2012). Vor allem große virtuelle Bibliotheken wie die *Europeana*⁴¹, die Millionen von Kunstwerken, Artefakten, Büchern, Videos und Audios aus ganz Europa beschreibt und die zugehörigen Informationen als Linked Open Data vorhält, unterstützen diese Entwicklung und bringen sie voran.⁴² Gemäß Scholz, Partner and Operations Manager bei *Europeana*, entwickelt sich die 2008 als digitale Bibliothek Europas online gegangene *Europeana* „mehr und mehr zur Drehscheibe für die Gesamtheit des digitalen Kulturerbes Europas“ (2015, S. 125). Als weitere Beispiele seien die *Digital Public Library of America (DPLA)*⁴³, das finnische *Kulttuurisampo*⁴⁴ und die *Deutsche Digitale Bibliothek (DDB)*⁴⁵, die Plattform für Kultur und Wissen in Deutschland, genannt.

3.1 ERFASSUNG MUSEALER OBJEKTE

Im Vergleich zu den traditionellen Bibliotheksobjekten, den Büchern aus dem Buchhandel, ist die Erfassung musealer Objekte komplexer. Neben einer detaillierten Beschreibung, die beispielsweise Informationen zu den Maßen, dem Material oder den Farben eines Objektes enthält, umfasst die Erschließung die Recherche über den Ursprung des Objekts, den Samm-

⁴¹ <https://www.europeana.eu/portal/de> (15.01.2019).

⁴² Weiterführende Informationen zur *Europeana* und ihrem Linked Data Ansatz finden sich bei Haslhofer und Isaac 2011 und Gradmann 2010.

⁴³ <https://dp.la/> (15.01.2019).

⁴⁴ <http://www.kulttuurisampo.fi> (15.01.2019).

⁴⁵ <https://www.deutsche-digitale-bibliothek.de/> (15.01.2019).

ler oder die verschiedenen Ereignisse, die es im Laufe seines Daseins durchlebt hat. „Da Objekte nur zu uns sprechen, wenn sie in einen Kontext gestellt werden, sind Forschungen zur Biographie der Objekte [...] notwendig“ (Al-Eryani et al. 2018, S. 549). Die Provenienzforschung lege Informationen ganz unterschiedlicher Art offen und ermögliche die Einordnung in bestimmte Zusammenhänge sowie Aussagen über die Echtheit eines Objektes (Al-Eryani et al. 2018, S. 549).

Dieser aufwändigen Forschungsarbeit, die der Metadatenerfassung musealer Objekte häufig vorausgeht, ist es auch geschuldet, dass Einrichtungen zur Bewahrung des kulturellen Erbes in vielen Fällen zurückhaltend bezüglich der Veröffentlichung ihrer Daten sind. Ein weiterer Grund für diese Zurückhaltung liegt in der Ungewissheit, die häufig mit Provenienzaussagen einhergeht. Datierungen, Ortsangaben oder auch beteiligte Personen und Einrichtungen lassen sich nicht immer zweifelsfrei belegen. Mitunter spielen auch kommerzielle Interessen eine Rolle (Schleier 2018).

Ferner ergeben sich aus der Verschiedenartigkeit der Objekte im Kulturerbesektor besondere Herausforderungen. Die Techniken und Standards für die Beschreibung, Indexierung und Suche müssen den unterschiedlichen Objektarten, seien es Gemälde oder Musikinstrumente, Schädel oder Bauwerke, gerecht werden (Chowdhury und Ruthven 2015a, S. 4).

Außerdem verweisen Chowdhury und Ruthven (2015a, S. 4) auf die enorme Bandbreite an Nutzerinnen und Nutzern im Kulturerbesektor. Informationen richteten sich einerseits an Experten wie Historiker oder Archäologen und andererseits an die breite Öffentlichkeit, bis hin zu Schülerinnen und Schülern. Der Informationsbedarf der jeweiligen Zielgruppen sei beim Design des Zugangssystems zu berücksichtigen.

Eine weitere Besonderheit von Museumsbeständen im Vergleich zu Bibliotheksbeständen liegt in der Einzigartigkeit der musealen Objekte. Während sich in Bibliotheken aufgrund der vielen Exemplare des gleichen Typs eine kooperative Erschließung auf Basis einer gemeinsamen Datengrundlage längst als Standard etabliert hat,⁴⁶ ist diese Art der Erschließung wie auch die Dublettenkontrolle für Museen von geringem Belang. Lediglich im Bereich der Druckgrafik und bei numismatischen Beständen sind mehrere identische Exemplare anzutreffen (Dührkohp 2015, S. 138).

3.2 WISSENSCHAFTLICHE SAMMLUNGEN

Wissenschaftliche oder akademische Sammlungen nehmen eine besondere Rolle im Kultur-

⁴⁶ Als Beispiele seien der Gemeinsame Verbundkatalog des GBV, der B3Kat und der HeBIS-Verbundkatalog genannt. Im März 2019 soll ferner der K10plus, der Katalog für zehn Bundesländer, die Stiftung Preußischer Kulturbesitz und weitere Einrichtungen, in Betrieb genommen werden.

erbe- und Museumskontext ein. Sie sind Dokumente der Forschungsaktivitäten ihrer Einrichtungen und somit einzigartige Zeugnisse der Wissens- und Wissenschaftsgeschichte (Georg-August-Universität Göttingen 2012, S. 9).

Vor allem die traditionsreichen Universitäten sind vielfach im Besitz umfangreicher Sammlungen, die sie seit ihren Gründungen kontinuierlich aufgebaut haben.⁴⁷ Im letzten Jahrhundert erfuhren die Sammlungen eine abnehmende Wertschätzung. Allemeyer et al. (2012, S. 11) nennen als Gründe die Verteilung einzelner Einrichtungen auf verschiedene Fakultäten sowie einen Paradigmenwechsel, dessen Auslöser in der Entwicklung neuer Forschungsmethoden besonders in den Naturwissenschaften zu finden sei. Häufig sei die aufwändige Pflege der Sammlungsbestände aus Kostengründen eingespart worden.

Die Autoren konstatieren ein Umdenken seit der Jahrtausendwende und machen die „Renaissance der Universitätssammlungen“ an einer Reihe von Ausstellungen in den Universitäten in Berlin, Halle, Tübingen, Erlangen und Leipzig im ersten Jahrzehnt des 21. Jahrhunderts fest (Allemeyer et al. 2012, S. 11).

Die Einrichtung eines eigenen Komitees des *International Council of Museums (ICOM)* für die Belange von *University Museums and Collections (UMAC)*⁴⁸ im Jahr 2000 verdeutlicht ebenfalls das steigende Interesse an universitären Sammlungen.

Das seit 2004 aufgebaute *Informationssystem Universitätssammlungen in Deutschland* dokumentiert Bestände und Geschichte von existierenden und nicht mehr vorhandenen Sammlungen von Universitäten in Deutschland und trägt damit zur besseren Sichtbarkeit und bundesweiten Vernetzung der Einrichtungen bei (Allemeyer et al. 2012, S. 13).

Spätestens seit den vom Wissenschaftsrat (2011) herausgegebenen Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen sind die akademischen Sammlungen endgültig in das Interesse der Hochschulleitungen gerückt. Der Wissenschaftsrat spricht den Sammlungen eine wichtige Rolle als Infrastruktur für Forschung und Lehre zu; ihre Objekte stellten „für die Forschung in zahlreichen wissenschaftlichen Disziplinen eine unentbehrliche Grundlage dar“ (Wissenschaftsrat 2011, S. 7). Er empfiehlt darüber hinaus eine beratende und koordinierende Einrichtung, welche 2012 als *Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland*⁴⁹ ihre Arbeit am Hermann von Helmholtz-Zentrum für Kulturtechnik der Humboldt-Universität zu Berlin aufnimmt (Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland 2012).

⁴⁷ Vgl. <https://portal.wissenschaftliche-sammlungen.de/kennzahlen> (15.01.2019). Als Beispiele seien die Sammlungen der Universitäten Heidelberg, Tübingen, Jena, München und Berlin genannt.

⁴⁸ <http://umac.icom.museum/> (15.01.2019).

⁴⁹ <https://wissenschaftliche-sammlungen.de> (15.01.2019).

Ein wichtiges Projekt der Koordinierungsstelle ist das *Portal Wissenschaftliche Sammlungen*⁵⁰, das Informationen zu Sammlungen, Beständen und Sammlungsaktivitäten wie Tagungen, Digitalisierungsprojekten und Ausstellungen ebenso bereithält wie zu Akteuren und Einrichtungen. Das Portal löst das 2004 gegründete *Informationssystem Universitätssammlungen in Deutschland* ab und verzeichnet am 01.09.2018 deutschlandweit 1085 Sammlungen an 88 Universitäten (Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland 2018).

3.3 PORTALE UND ERFASSUNGSDATENBANKEN

Als Folge des wachsenden Interesses an musealen Objekten und des Trends, diese digital erlebbar zu machen, sind in den vergangenen Jahren zahlreiche Programme zur Digitalisierung und Katalogisierung von Kulturgut initiiert worden. Der Großteil der Ergebnisse dieser Projekte wurde über Portale oder Websites öffentlich zugänglich gemacht (Hagedorn-Saupe 2012, S. 205).

Nachfolgend erfahren die Portale *Kulturerbe Niedersachsen*⁵¹ und *KENOM*⁵² sowie die Erfassungsdatenbanken *naniweb*⁵³, *kuniweb*⁵⁴ und *kenom*⁵⁵ eine nähere Betrachtung, zum einen, weil sie als typische Beispiele für die Entwicklungen im Kulturerbesektor herangezogen werden können, und zum anderen und insbesondere, weil aus ihnen das Göttinger Sammlungsportal hervorgegangen ist.

Die beiden Portale waren mit Blick auf ihr Webinterface, ihre Funktionalität sowie die zugrundeliegende Software beispielgebend für das Sammlungsportal. Aus den drei Erfassungsdatenbanken *naniweb*, *kuniweb* und *kenom* wurde und wird zukünftig der Großteil des Datenbestands des Sammlungsportals ausgespielt.

Durch die Auseinandersetzung mit den Vorgängerportalen und den Erfassungssystemen wird das Verständnis für die im Zuge dieser Arbeit zu konvertierenden Daten geschärft und in gewisser Weise dem Kapitel 4 vorgegriffen, das dem Göttinger Sammlungsportal und seinen Daten gewidmet ist.

Es wird außerdem deutlich, worauf es im musealen Kontext ankommt. Digitale Reproduktionen von Objekten, stufenloses Zoomen der Abbildungen sowie die Beschäftigung mit Samm-

⁵⁰ <https://portal.wissenschaftliche-sammlungen.de> (15.01.2019).

⁵¹ <https://www.kulturerbe.niedersachsen.de> (15.01.2019).

⁵² <https://www.kenom.de> (15.01.2019).

⁵³ <https://naniweb.gbv.de> (15.01.2019).

⁵⁴ <https://kuniweb.gbv.de> (15.01.2019).

⁵⁵ <https://kenom.gbv.de> (15.01.2019).

lungsstätten, Fundorten und Objektgattungen sind typisch und wichtig im Kulturerbesektor. Außerdem spielt das digitale Erleben eine basale Rolle (Hagedorn-Saupe 2012, S. 196).

Darüber hinaus zeigen die nachfolgenden Beispiele, dass die Verbreitung der Daten außerhalb der Portale noch nicht realisiert ist. Der Nachholbedarf des Sektors im Bereich Digitalisierung lässt sich auch daran festmachen, dass es in Niedersachsen erst seit 2013 zentrale Erfassungsportale für museale Objekte gibt.⁵⁶

Das Portal *Kulturerbe Niedersachsen*

Das 2012 eröffnete Portal *Kulturerbe Niedersachsen* ermöglicht der interessierten Öffentlichkeit einen zentralen Zugang zu ausgewählten digital erfassten Kulturgütern aus Bibliotheken, Archiven, Museen und anderen Kultureinrichtungen des Landes Niedersachsen.

Die Digitalisate sowie sämtliche Metadaten werden von den einbringenden Einrichtungen zur Verfügung gestellt. Die Metadatenlieferungen bestehen im Wesentlichen aus den Austauschformaten METS/MODS für Bibliotheken, EAD für Archive und LIDO für Museen. Für eine einheitliche Präsentation werden die Daten in ein auf METS/MODS-basierendes Format umgewandelt (Dührkohp 2015, S. 131).

Für den Webauftritt kommt der *intra*-Viewer zum Einsatz. „Der Viewer stellt bereits alle wesentlich benötigten Funktionalitäten wie zum Beispiel stufenloses Zoomen, Vollbildanzeige, Navigationsstrukturen und Tag-Clouds zur Verfügung“ (Dührkohp 2015, S. 132).

Die Metadaten sind auf Basis eines *Solr*-Volltextindex durchsuchbar (Dührkohp 2015, S. 133). Außerdem besteht die Möglichkeit, über eine einschränkende Suche zu den Objekten zu navigieren. Der Einstieg kann über den Namen einer Sammlungsstätte, eine bestimmte Objektgattung (z. B. Malerei oder Fotografie) oder über eine zeitliche Einordnung erfolgen.

Die OAI-PMH-Schnittstelle, die die Metadaten bereitstellt, kann gemäß Aussage der VZG noch nicht freigegeben werden, weil sich nicht jede beitragende Einrichtung mit der Weitergabe einverstanden erklärt habe. Über diese OAI-Schnittstelle solle perspektivisch die Bereitstellung der Daten für die *DDB* und von dort in die *Europeana* erfolgen (Dührkohp 2018).

Die Erfassungsdatenbank *kuniweb*

Wurden in der Anfangsphase von *Kulturerbe Niedersachsen* sämtliche Daten dezentral erfasst und erst im Portal wie oben beschrieben zusammengeführt, liegt seit 2013 mit *kuniweb* eine zentrale Erfassungsdatenbank für museale Sammlungen in Niedersachsen vor. Sie ist komplett web- und browserbasiert und wurde mit Hilfe der Software *easyDB* der Firma *Pro-*

⁵⁶ Im Bibliothekssektor wurde die zentrale Erfassung in Niedersachsen bereits 1977 mit dem Niedersächsischen Monographienachweis (NMN) begonnen (Becker und Müller-Dreier 2005, S. 129).

*grammfabrik*⁵⁷ realisiert. Die Software ermöglicht das Inventarisieren, Archivieren und Recherchieren von digitalen Bild- und Multimediadaten sowie die Eingabe der zugehörigen Metadaten (Dührkohp 2015, S. 136–137).

Da der größte Teil der aktuellen Daten im Göttinger Sammlungsportal aus *kuniweb* stammt, lohnt sich ein detaillierter Blick auf das Datenerfassungsschema. Neben zahlreichen Freitextfeldern, zum Beispiel für den Titel, die Inventarnummer oder die Beschreibung des Objekts, bietet das Erfassungsportal etwa 20 verschiedene Datenfelder, die einzig die kontrollierte Eingabe über eine Auswahl vorgegebener Inhalte zulassen. Diese Inhalte, repräsentiert durch URIs, entstammen nationalen und internationalen Wissensorganisationssystemen. Beispielsweise wird die *Gemeinsame Normdatei (GND)*⁵⁸ der *Deutschen Nationalbibliothek*⁵⁹ für die Erfassung von Personen, Körperschaften und Schlagwörtern eingebunden. Für Ortsangaben werden die freie Ortsdatenbank *GeoNames*⁶⁰ und für ikonografische Angaben zu den Bildinhalten die Klassifikation *Iconclass*⁶¹ hinzugezogen. Die sogenannte *Oberbegriffsdatei (OBG)*⁶² bietet ebenfalls Normdaten für die Klassifizierung. Weitere museale Vokabulare kommen für die Beschreibung von Materialien, Technik, Stil und Epoche zum Einsatz. Über die Titel im *Gemeinsamen Verbundkatalog (GVK)*⁶³ können die Objekte mit Literaturangaben verknüpft werden. Eine weitere Verknüpfungsmöglichkeit besteht mit der Denkmaldatenbank *ADABweb*⁶⁴ des *Niedersächsischen Landesamtes für Denkmalpflege*, wodurch eine einheitliche Verzeichnung von archäologischen Fundstellen oder Baudenkmalern möglich ist.

In *kuniweb* wird somit die automatische Erfassung von zahlreichen HTTP-URIs, den grundlegenden Bestandteilen von Linked Data, sichergestellt. Weitere wichtige Vorteile dieser Erfassungsmethode liegen in der Minimierung von Eingabefehlern und in der Sicherstellung der Einheitlichkeit der Daten.

Das Internformat für die Erfassungsdatenbank ist ein proprietäres XML. *Kuniweb*-Daten können in LIDO sowie in METS/MODS exportiert werden (Dührkohp 2018).

Das Erfassungsportal *naniweb*

Während sich *kuniweb* vorrangig an Museen in Niedersachsen richtet, ist die Klientel des drei Jahre jüngeren Erfassungsportals *naniweb* vor allem im naturwissenschaftlichen Bereich verortet. Organische und mineralische Fundstücke und Präparate niedersächsischer Einrichtungen werden erfasst.

⁵⁷ <http://www.programmfabrik.de> (15.01.2019).

⁵⁸ <https://www.dnb.de/gnd> (15.01.2019).

⁵⁹ <https://dnb.de/> (20.01.2019).

⁶⁰ <https://www.geonames.org> (15.01.2019).

⁶¹ <http://www.iconclass.nl> (15.01.2019).

⁶² <http://obg.vocnet.org> (15.01.2019).

⁶³ <https://gso.gbv.de> (15.01.2019).

⁶⁴ <http://www.adabweb.info> (15.01.2019).

Naniweb basiert auf derselben Software wie *kuniweb*. Aufgrund der inhaltlichen Ausrichtung sind naturwissenschaftliche Vokabulare wie zum Beispiel die *Biological Collections Ontology*⁶⁵ oder die *Environment Ontology*⁶⁶ eingebunden (Dührkohp 2018).

Der wesentliche Unterschied und eine Verbesserung zu *kuniweb* liegt in dem eventbasierten Ansatz von *naniweb*. Zu jedem Objekt können beliebig viele Ereignisse erfasst werden. Diese Ereignisse werden standardisiert typisiert,⁶⁷ und es ist eine eindeutige Zuordnung von Personen, Orten und Daten zu genau diesem Ereignis möglich. Damit ist *naniweb* ereignis- oder eventbasierten RDF-Modellen wie *CIDOC CRM*⁶⁸ oder auch *ASCH* sehr viel zuträglicher als *kuniweb* und bietet im LIDO-Export eine bessere Datenqualität.

Zum Zeitpunkt der Erstellung dieser Arbeit ist ein Export in LIDO aus *naniweb* möglich, die METS/MODS-Schnittstelle für die Weitergabe der Daten an das Portal *Kulturerbe Niedersachsen* ist jedoch noch nicht eingerichtet (Dührkohp 2018).

Die Erfassungsdatenbank *kenom* und das *KENOM-Portal*

Kenom steht für *Kooperative Erschließung und Nutzung der Objektdaten von Münzsammlungen*. Die ebenfalls mit *easyDB*-Software betriebene Erfassungsdatenbank *kenom* – *Digitaler Münzkatalog* geht auf das Jahr 2013 zurück. Anders als *kuniweb* und *naniweb* ist das auf Münzen, Medaillen und Geldscheine spezialisierte System nicht auf niedersächsische Einrichtungen beschränkt. Projektpartner aus Thüringen, Sachsen-Anhalt und Schleswig-Holstein waren von Anfang an beteiligt. Neben den spezifischen auf numismatische Bestände ausgerichteten Erfassungsfeldern und Vokabularen besteht ein weiterer Unterschied von *kenom* zu *naniweb* und *kuniweb*: Während Museumsobjekte in den meisten Fällen Unikate sind, kommen numismatische Bestände vielfach in mehreren Exemplaren vor (Dührkohp 2015, S. 138–141).

Zur *kenom*-Erfassungsdatenbank gehört auch das 2015 eröffnete *KENOM-Portal*, *KENOM - Virtuelles Münzkabinett*, das die Bestände der breiten Öffentlichkeit digital zugänglich macht. Über eine OAI-PMH-Schnittstelle⁶⁹ werden die Daten in den Formaten Dublin Core Element Set (DCES) und LIDO bereitgestellt. Die Etablierung eines Geschäftsgangs für die Verbreitung der Daten über die *Deutsche Digitale Bibliothek* in die *Europeana* wird laut Dührkohp (2018) derzeit mit der *DDB* verhandelt.

⁶⁵ <http://purl.obolibrary.org/obo/bco.owl> (15.01.2019).

⁶⁶ <http://www.environmentontology.org> (15.01.2019).

⁶⁷ Die Typisierung erfolgt über LIDO *eventTypes*. Vgl. <http://terminology.lido-schema.org/eventType> (15.01.2019).

⁶⁸ Siehe unten Kap. 3.4.

⁶⁹ https://dokumentation.digitalebibliothek.gbv.de/kenom:Dokumentation_der_OAI-Schnittstelle_von_KENOM (15.01.2019).

3.4 METADATENFORMAT LIDO

Nach Eckert sind Metadaten „strukturierte Daten, die die Eigenschaften einer Ressource beschreiben“ (2013, S. 98). Ein Metadatenformat kann daraufhin als Dokumentation definiert werden, die das Design der strukturierten Datenbeschreibung darstellt (Al-Eryani und Rühle 2018, S. 5).

Das Metadatenformat LIDO (Lightweight Information Describing Objects) ist das gängige Austauschformat im musealen Kontext (Simou 2014, S. 3; Hagedorn-Saupe 2012, S. 205).

Es ist ein XML-Format. Somit weisen LIDO-Datensätze die für XML typische Baumstruktur auf und bestehen aus Elementen, Attributen, ihren Wertzuweisungen und den Inhalten der Elemente.

LIDO hält insgesamt 14 verschiedene Elementgruppen (*Wraps*) für die Eingabe von im musealen Kontext wichtigen Informationen bereit. Neben der RecordID enthalten lediglich drei Sektionen Pflichtfelder. Die nachfolgende Abbildung zeigt die 14 Elementgruppen; die drei Pflichtbereiche sind hellbraun gekennzeichnet.

Abbildung 5: LIDO-Elemente (ICOM/CIDOC 2010a)

Diese 14 Elementgruppen sind je einem der beiden übergeordneten Bereiche *descriptiveMetadata-Wrap* (elf Elemente) oder *administrativeMetadata-Wrap* (drei Elemente) zugeordnet. Der deskriptive Teil ist zusätzlich in die vier in der obigen Abbildung kursiv gedruckten Zwischenhierarchien *ObjectClassifications*, *ObjectIdentifications*, *Events* und *Relations* unterteilt.

Abbildung 6 zeigt einen Ausschnitt aus einem LIDO-Datensatz. Das *titleWrap* als erstes Element des *objectIdentificationWrap* wird angezeigt.


```

<lido:objectIdentificationWrap>
  <lido:titleWrap>
    <lido:titleSet>
      <lido:appellationValue lido:pref="preferred">La Primavera / Spring</lido:appellationValue>
      <lido:appellationValue lido:pref="preferred" xml:lang="en">La Primavera / Spring</lido:appellationValue>
    </lido:titleSet>
  </lido:titleWrap>

```

Abbildung 6: Ausschnitt aus einem LIDO-Datensatz

LIDO ist nicht auf die Datenhaltung in den Bereichen Erwerbung und Ausleihe ausgerichtet. Es geht um die möglichst umfassende Beschreibung der Museumsobjekte, inklusive Provenienzinformationen, in einer Vielzahl von Online-Services sowie um die gemeinsame Nutzung, den Austausch und die Verknüpfung der Daten im *World Wide Web* (ICOM/CIDOC 2010b).

The strength of LIDO lies in its ability to support the full range of descriptive information about museum objects. It can be used for all kinds of object, e.g. art, architecture, cultural history, history of technology, and natural history (Tsalapate et al. 2012).

LIDO basiert auf dem vom *International Committee on Documentation* entwickelten Conceptual Reference Model (CRM), kurz CIDOC CRM, einer Ontologie, die Klassen (z. B. E63 Beginning of Existence) und Properties (z. B. P152 hasParent) mit Unter- und Oberklassen sowie Beschreibungen für den Kulturerbesektor definiert und mit URIs maschinenlesbar vorhält (Le Boeuf et al. 2015). Die nachfolgende Abbildung zeigt beispielhaft den CIDOC CRM Eintrag der Klasse *E63 Beginning of Existence* mit Ober- und Unterklassen.

Abbildung 7: CIDOC CRM-Eintrag Beginning of Existence

Auf CIDOC CRM geht die Ereignisorientierung von LIDO zurück. Für jedes Ereignis ist darstellbar, wer, wann und wo an welcher Art von Event beteiligt war.

LIDO hält zahlreiche Felder für die Eingabe von URIs bereit. Zudem können beliebige Vokabulare eingebunden werden. Somit stellen im LIDO-Format vorliegende Daten eine gute Voraussetzung für die Generierung von Linked Open Data dar.⁷⁰

⁷⁰ Neben einer umfassenden Dokumentation des LIDO Metadatenformats (Coburn et al. 2010) stellt die *LIDO Working Group* auf der *xTree.public*-Website des *digiCULT*-Verbunds unter <http://terminology-view.lido-schema.org/> (15.01.2019) eine LIDO-Terminologie zur Verfügung, die URIs und Beschreibung zu einzelnen Begriffskategorien vorhält. Zum Beispiel sind die LIDO Event-Typen hier dokumentiert.

4 GÖTTINGER SAMMLUNGSPORTAL

Forschungsressource, didaktisches Lehrmittel oder akademisches Erbe - die Sammlungen der Universität Göttingen bilden im Einzelnen und in ihrer außergewöhnlichen Vielfalt und Bandbreite einen einzigartigen Bestand (Beisiegel et al. 2013, S. 4).

Datenmapping setzt eine gründliche Analyse und ein basales Verständnis der Ausgangsdaten und des Zielformats voraus. Dieses Kapitel beschäftigt sich mit der Ausgangslage, dem Göttinger Sammlungsportal und seinen Daten.

4.1 ENTSTEHUNG, ZIELE UND WEBAUFTRITT

Insgesamt beherbergen die 43 Sammlungen der Georg-August-Universität Göttingen Tausende von Objekten, verteilt auf über 30 Standorte auf dem gesamten Campus.

Den Empfehlungen des Wissenschaftsrats (2011) folgend sieht die Strategie der Universität die Erhaltung und bessere Sichtbarkeit der universitären Sammlungen vor (Beisiegel et al. 2013, S. 7). Zu diesem Zweck wurden die digitale Erschließung und die damit einhergehende fach- und ortsunabhängige Nutzbarkeit der Sammlungen vorangetrieben.

Seit Dezember 2017 besteht das Göttinger Sammlungsportal, das die verschiedenen Erschließungsformate zusammenfasst und webbasiertes, sammlungsübergreifendes Suchen ermöglicht (Beisiegel et al. 2013, S. 5). Es soll mittelfristig den Gesamtbestand der Göttinger Sammlungen nachweisen und „für aktuelle und möglichst auch zukünftige Fragestellungen die relevanten Informationen enthalten und ausgeben“ (Beisiegel et al. 2013, S. 6).

Gemäß der Aussage der Präsidentin der Universität „ist es der Georgia Augusta [...] ein wichtiges Anliegen, [...] [die Sammlungen] für Forschung und Lehre zu nutzen, die Öffentlichkeit über die zentrale Rolle der akademischen Sammlungen in der Wissenschaft zu informieren und die beeindruckenden Schätze auch interessierten Bürgerinnen und Bürgern sowie Schülerinnen und Schülern zugänglich zu machen“ (Beisiegel 2012, S. 9).

Mit Stand Oktober 2018 waren im Sammlungsportal 21.071 Objekte ganz unterschiedlicher Gattungen aus fast allen universitären Sammlungen verzeichnet. Neben zum Teil umfangreichen Beschreibungen sind hier außerdem IIF-konforme⁷¹ Bilder der Objekte verfügbar.

⁷¹ IIF steht für International Image Interoperability Framework. Weitere Informationen finden sich unter <https://iiif.io/> (15.01.2019).

Wie die Portale *Kulturerbe Niedersachsen* und *KENOM Münzkabinett* basiert auch die Benutzeroberfläche des Göttinger Sammlungsportals auf dem *intranda*-Viewer, der einen *Solr*-Index für die Suche in den Daten nutzt (Schleier 2018). Neben der Volltextsuche besteht die Möglichkeit des Sucheinstiegs über eine bestimmte *Objektgattung* oder *Sammlung*. Eine einschränkende Suche ist mit den Kriterien *Objekt mit Bild*, *Sammlungen*, *Orte*, *Personen* und *Bildrechte* möglich. Die zu den einzelnen Objekten angezeigten Informationen sind auf mehrere Panels verteilt. Zusätzlich zu den vier Panels *Grunddaten*, *Zitieren und Nachnutzen*, *Klassifikationen*, *Taxonomien und Schlagworte* und *Beschreibung* sind Panels für *Verknüpfungen* und jedes erfasste *Event* vorgesehen.⁷² Außerdem werden die digitalen Repräsentationen angezeigt und deren Verwertungsrechte und Urheber genannt. Die Betrachtung der Benutzeroberfläche ist insofern lohnenswert, als sie als Indikator für die Relevanz der Daten für das RDF-Mapping zu werten ist. Die hier angezeigten und zur Filterung angebotenen Felder sollten nach Möglichkeit auch als Linked Data abgebildet werden.

Hinsichtlich der Nachnutzung der Daten bekennt sich die Universität Göttingen zu den Prinzipien von Open Science und Open Access. Auf der Website des Sammlungsportals ist die freie Nachnutzung wie folgt geregelt: „Alle Informationen zu den Objekten (Metadaten) auf sammlungen.uni-goettingen.de dürfen Sie [...] ohne Nachfrage (wie die gemeinfreien Objekte) teilen, nutzen, bearbeiten (Lizenz CC0)“ (Georg-August-Universität Göttingen 2017).

Die Nachnutzung der sich aktuell auf Fotos beschränkenden digitalen Objektrepräsentationen ist zu großen Teilen auch über CC-Lizenzen⁷³ geregelt, wobei häufig die Namensnennung des Urhebers (CC-BY) und Nachnutzung unter gleicher Lizenz (CC-BY-SA) verlangt werden.

Dieser offene Umgang mit den Rechten an den Daten spiegelt das Anliegen der Universität Göttingen wider, die Nachnutzbarkeit und Sichtbarkeit der eigenen Daten zu erhöhen. Aus diesem Grund ist das Portal auch im Registry of Research Data Repositories⁷⁴ verzeichnet. Gleichzeitig bietet diese Verzeichnung keine Lösung gegen die Silohaltung der Daten.

Um für zukünftige Fragestellungen gerüstet zu sein und zur besseren Auffindbarkeit, Nachnutzbarkeit und Vernetzung der Daten, ist die Bereitstellung von Linked Open Data der logische nächste Schritt.

⁷² Beispiel: <https://hdl.handle.net/21.11107/2b88bf2b-486f-4d52-9431-5072e5934f69> (26.01.2019).

⁷³ Siehe oben Kap. 2.5.

⁷⁴ <https://www.re3data.org/> (15.01.2019).

4.2 AUSGANGSDATEN

Während der vorangehende Abschnitt allgemeine Informationen zum Göttinger Sammlungsportal enthält und somit den Kontext der Daten beschreibt, richtet sich der Fokus im Folgenden direkt auf die Quelldaten.

4.2.1 HERKUNFT UND SICHTBARKEIT

Das Primärformat des Göttinger Sammlungsportals ist LIDO-XML⁷⁵. Bevor die Daten in das Sammlungsportal integriert werden konnten, mussten sie in diesem Format vorliegen.

Zu Beginn dieser Arbeit verzeichnete das Sammlungsportal 21.071 Objekte. Somit lagen 21.071 LIDO-XML-Datensätze, jeweils in einer eigenen Datei, als Ausgangsdaten für die Konvertierung nach Linked Open Data vor.

In Kapitel 3 wurde bereits auf die Erfassungsportale eingegangen, aus denen Daten in das Sammlungsportal ausgespielt wurden und zukünftig werden. Nachfolgend ist visualisiert, aus welchen Herkunftsquellen die in dieser Arbeit zu konvertierenden Datensätze stammen.

Abbildung 8: Herkunft der Ausgangsdaten

Der Großteil der Bestände stammt aus *kuniweb* und *kenom*. Das *naniweb*-Erfassungsportal ist das jüngste, was ein Grund für die noch geringe Datenmenge ist. Die heterogenen Quellen reichen von Access-Datenbanken bis zu Excel-Listen (Schleier 2018).

Während *kuniweb*, *naniweb* und *kenom* einen ausgereiften Datenexport in LIDO implementiert haben, gingen der Integration der heterogenen Quellen individuell angepasste Datenbereinigungen und -vereinheitlichungen voraus. Letztlich haben sämtliche im Göttinger

⁷⁵ Siehe oben Kap. 3.4.

Sammlungsportal enthaltene Daten mindestens eine automatische Aufbereitung durchlaufen. Dabei lag eine wichtige Anforderung in der Bereitstellung von URIs für die transportierten Inhalte (Schleier 2018).

Die über die Erfassungsdatenbanken *kuniweb*, *naniweb* und *kenom* aufgenommenen Daten verfügen über eine einheitliche Struktur, die eine Umwandlung in LIDO erleichtert; *naniweb* ist zudem eventbasiert aufgebaut. Außerdem ist durch die Verknüpfung mit zahlreichen Wissensorganisationssystemen⁷⁶ und ihren URIs eine gute Qualität der Inhalte gewährleistet.

Des Weiteren wurden nicht sämtliche Informationen, die in den Herkunftsdatenbanken hinterlegt sind, in das Sammlungsportal übernommen. Manche Felder, wie beispielsweise der Versicherungswert, sind nicht für die Veröffentlichung freigegeben. Außerdem spielte bei der Auswahl der Felder der Zweck des Portals, die Objekte der Göttinger Sammlungen sichtbarer und vernetzbarer zu machen, eine maßgebliche Rolle (Schleier 2018).

Auch wenn die Daten des Sammlungsportals qualitativen Abstufungen unterliegen, sind sie in ihrer Gesamtheit gut aufbereitet, einheitlich strukturiert und mit URIs angereichert. Dies unterstützt den Ansatz dieser Arbeit, die Datenbereinigung nicht explizit zu behandeln.

Hinsichtlich der globalen Verfügbarkeit und Vernetzbarkeit der Ausgangsdaten lassen sich nachfolgende Aussagen treffen: Die Metadaten des Göttinger Sammlungsportals sind nur insofern verfügbar, als jeder Datensatz einzeln im LIDO-Format abrufbar ist, und zwar aus der Vollanzeige des Sammlungsportals heraus.⁷⁷ Für die Daten, die auch im *KENOM-Portal* angezeigt werden, existiert außerdem eine OAI-Schnittstelle⁷⁸, über die sich Interessierte die Daten abholen und zum Beispiel in eigene Datenbanken einspielen können. Aktuell gibt es keinen Datentransfer in die *Deutsche Digitale Bibliothek*, über welche in einem nächsten Schritt die Weitergabe an die *Europeana* möglich wäre.

Selbst wenn der automatische Datentransfer zur *DDB* und *Europeana* mittelfristig realisiert ist und damit ein Teil der Daten aufgrund des *Europeana*-RDF-Formats⁷⁹ als LOD vorliegen wird, ist die Bereitstellung von LOD im Rahmen des Sammlungsportals weiterhin relevant. Der Ansatz der *Europeana* besteht weniger darin, die Sammlungen und ihre Objekte mit allen Ereignissen abzubilden, sondern zielt vornehmlich auf die Erfassung der Kulturgüter Europas. Demzufolge finden dort weniger, respektive andere, Informationen Berücksichtigung. Zum Beispiel werden Ereignisse wie auch die mit den Ereignissen verknüpften Akteure im *Europeana*-Format nicht entsprechend abgebildet (Simou 2014, S. 12).

⁷⁶ Siehe oben Kap. 3.3.

⁷⁷ Beispiel-LIDO-Datensatz: https://sammlungen.uni-goettingen.de/lidoresolver?id=record_kuniweb_936444 (15.01.2019).

⁷⁸ <http://www.kenom.de/oai/> (15.01.2019).

⁷⁹ <https://pro.europeana.eu/resources/standardization-tools/edm-documentation> (15.01.2019).

Angesichts dieser aktuell und zukünftig eher eingeschränkten globalen Zugänglichkeit und Verwertbarkeit der Daten als Linked Data erscheint ein speziell auf das Sammlungsportal ausgerichtetes RDF-Mapping sinnvoll. Nur so sind alle relevanten Informationen abbildbar.

4.2.2 SYNTAX UND SEMANTIK

Für die Analyse der Ausgangsdaten stellt sich zunächst die Frage nach einer Dokumentation, die die Struktur der Daten, mögliche Felder sowie zugelassene Inhalte beschreibt. Für die in dieser Arbeit vorliegenden Ausgangsdaten stehen mehrere Dokumentationen zur Verfügung: Zum einen das *LIDO V 1.0 Specification Document*⁸⁰, das LIDO-Regelwerk, das jedes mögliche Element mit seinen Attributen, Ober- und Unterelementen und Informationen zur Wiederholbarkeit benennt, und zum anderen jeweils eigene Dokumentationen des *kuniweb-LIDO-Exportformats*⁸¹ und des *kenom-LIDO-Exportformats*⁸². Letztere zielen direkt auf die in dieser Arbeit vorliegenden Ausgangsdaten und erhalten somit den Vorrang gegenüber der allgemeinen LIDO-Dokumentation. Da sich *kuniweb*- und *kenom*-Exportformate mit Blick auf die benutzen LIDO-Felder nicht wesentlich unterscheiden und *kuniweb* gleichzeitig deutlich mehr Objektgattungen abdeckt als *kenom*, das vornehmlich auf Münzen ausgelegt ist, wird für die Analyse der Ausgangsdaten die *kuniweb*-Dokumentation zugrunde gelegt. Sie hat außerdem das *culture-Element* im *eventWrap*, welches *kenom* fehlt.

Im weiteren Vorgehen gilt es, anhand der *kuniweb*-Dokumentation einen groben Überblick und ein grundsätzliches Verständnis für den Aufbau der Daten, die genutzten Felder und die Art der Nutzung zu gewinnen. Die *kuniweb*-Dokumentation hat gegenüber der allgemeinen LIDO-Dokumentation außerdem den Vorteil, dass sie an zahlreichen Beispielen die möglichen Felder und Inhalte veranschaulicht. Die Information, wie häufig ein Feld vorkommen kann oder muss, ist ebenfalls enthalten. Die Felder des *kuniweb*-LIDOs entsprechen im Wesentlichen denen des Standard-LIDOs. Wenige *Wraps* aus dem Standard werden nicht verwendet.⁸³

Ein für das Mapping der Daten wichtiges Spezifikum aus der *kuniweb*- und *kenom*-Dokumentation ist die Präferenz der deutschen Sprache. Die textlichen Metadaten sind demnach grundsätzlich in Deutsch verfasst, es sei denn eine abweichende Sprache ist mit dem Attribut *xml:lang* explizit vermerkt.

Abbildung 9 zeigt die LIDO-Wrap-Elemente, die in den Datensätzen des Sammlungsportals vorkommen können. Die für diese Anwendung wichtigen Elemente lassen sich in drei Berei-

⁸⁰ <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf> (15.01.2019).

⁸¹ <http://hdl.handle.net/428894.vzg/4fd5051b-d2a4-47a6-9656-bc1c342d0df4> (15.01.2019).

⁸² <http://hdl.handle.net/428894.vzg/6711ef39-5762-46a8-b638-a69ce7473fe2> (15.01.2019).

⁸³ *State/Edition* und *Rights* aus dem Standard LIDO werden nicht verwendet.

che aufteilen. Der erste Bereich enthält zweimal das Element *objectPublishedID*. Das erste *objectPublishedID*-Element beinhaltet einen URI, der das Sammlungsobjekt identifiziert. Dieser URI beginnt stets mit der Zeichenkette *hdl.handle.net*. Der URI im anderen *objectPublishedID*-Element identifiziert den Datensatz, auch *Record* genannt, der dieses Objekt beschreibt.

Abbildung 9: Struktur des LIDO-Datensatzes im Sammlungsportal

Der zweite Bereich enthält die Metadaten, die das Sammlungsobjekt charakterisieren (*descriptiveMetadata*). Hier sind der Titel des Objektes, eventuelle Beschriftungen, seine Maße sowie inhaltliche Einordnungen über eine Klassifikation oder Schlagwörter verzeichnet. Zum Teil sind auch detaillierte Beschreibungen des Objekts enthalten und Verknüpfungen zu anderen Objekten oder Literaturangaben angegeben. Von besonderer Bedeutung ist das *eventWrap*, das die Ereignisse enthält, die ein Objekt im Laufe seines Daseins durchlebt hat. Die am häufigsten vorkommenden Event-Typen sind *Production*, *Commissioning* und

Beginning of Existence.⁸⁴ Darüber hinaus gibt es Event-Informationen außerhalb des *EventWraps*. Das *objectIdentificationWrap* enthält den Aufbewahrungsort (*Custody*) und Informationen zur Benutzung (*Use*) und zum Herstellungszweck (*Production_purpose*).

Schließlich umfasst der dritte Bereich (*administrativeMetadata*) die administrativen Metadaten sowie die URIs der digitalen Repräsentationen, also der Bilder der Objekte. Diese sind im *resourceWrap* zusammen mit den zugehörigen Lizenzinformationen abgelegt. Der *recordWrap* hält Informationen zu dem Datensatz vor; etwa, wann er erstellt wurde und wer der Rechteinhaber ist. Es werden auch die Rechte für die Verwertung der Metadaten genannt. Gemäß den Nutzungshinweisen auf der Website des Sammlungsportals⁸⁵ sollten alle Datensätze die CC0-Lizenz aufweisen. Die Datenanalyse im Rahmen dieser Arbeit ergab jedoch ein anderes Bild, sodass die individuellen Rechte eines jeden Datensatzes beim Mapping berücksichtigt werden müssen, anstatt, wie zunächst geplant, einheitlich die CC0-Lizenz zu erzeugen.⁸⁶

Bei der Betrachtung der einzelnen Elemente ist ein besonderes Augenmerk auf Felder mit URIs zu legen, denn URIs sind in einem RDF-Triple die Bestandteile, über die eine weitere Verknüpfung im Web of Data möglich ist. Für URIs hält das LIDO-Format das Feld *conceptID* in zahlreichen *Wraps* bereit. Das Beispiel in Abbildung 10 zeigt die Angabe des Schlagworts *Teich*, zum einen als URI der Iconclass-Ontologie mit type-Attribut im Feld *conceptID* und zum anderen als Literal im Feld *term*.

```
<lido:objectRelationWrap>
  <lido:subjectWrap>
    <lido:subjectSet>
      <lido:displaySubject>Teich, Tümpel</lido:displaySubject>
      <lido:subject>
        <lido:subjectConcept>
          <lido:conceptID lido:type="http://terminology.lido-schema.org/identifier_type/uri">
            http://iconclass.org/25H2143</lido:conceptID>
          <lido:term>Teich, Tümpel</lido:term>
        </lido:subjectConcept>
      </lido:subject>
    </lido:subjectSet>
  </lido:subjectWrap>
```

Abbildung 10: Ausschnitt aus einem LIDO-Datensatz mit Schlagwort-URI

Neben der Dokumentation, die gewissermaßen die theoretische Sicht auf die Daten darstellt, müssen die realen Inhalte betrachtet werden. Es ist zu überprüfen, ob die Eingaben gemäß den Vorgaben in der Dokumentation erfolgt sind. Ferner kann die Betrachtung der Inhalte der Felder Rückschlüsse auf ihre Relevanz für das Mapping liefern. Wird ein Feld beispiels-

⁸⁴ Diese Reihenfolge ist das Ergebnis einer BaseX-Anfrage. Die Abfrage ist auf der CD zur Arbeit unter *eventTypes_count.xq* enthalten.

⁸⁵ <https://sammlungen.uni-goettingen.de/nutzung/> (15.01.2019).

⁸⁶ Diese Abweichung zwischen den Vorgaben der Universität und den Informationen in den Datensätzen befindet sich aktuell in Klärung bei den Sammlungen, die für die Erfassung der Daten verantwortlich zeichnen.

weise sehr selten belegt und sind die Inhalte ohne großen Mehrwert, kann die Berücksichtigung im Mapping entfallen. Insbesondere mit Blick auf den Ansatz dieser Arbeit, sich auf das Wesentliche und allen Gemeinsame zu konzentrieren, ist dieser Aspekt von Relevanz.

Darüber hinaus sind die Daten des Sammlungsportals teilweise homogener als in der Dokumentation gefordert. So ist zum Beispiel die Objektgattung laut Dokumentation nicht zwingend als URI zu hinterlegen, ein Term ist ausreichend. Die Datensätze im Sammlungsportal enthalten jedoch allesamt einen URI. Das Mapping kann den Term somit vernachlässigen.

Im Folgenden gilt es, zwei Fragestellungen zu beantworten:

- Welche Inhalte stehen in den Feldern? Sind die Inhalte interessant für das Mapping oder enthalten sie beispielsweise Redundanzen?
- Welche Inhalte lassen sich mit dem ASCH-Modell abbilden?

Für die Beantwortung des ersten Punkts werden Tools hinzugezogen, die im nächsten Abschnitt erläutert werden.

Die zweite Frage lässt sich erst beantworten, wenn auch das ASCH-Modell, also das Zielformat, bekannt ist. Ihre Beantwortung wird somit nach der Betrachtung des ASCH-Modells im Rahmen des Mappings in Kapitel 6 erfolgen.

4.2.3 TOOLS FÜR DIE DATENANALYSE

Ein erster, einfacher Ansatz, die 21.071 Datensätze und Dateien über einen beliebigen Editor durchsuchbar zu machen, bestand darin, sämtliche Files zu einer Datei zusammenzufügen. Dies lässt sich mit dem Unix-Befehl `cat` bewerkstelligen. Das Ergebnis ist allerdings nur für sehr einfache Suchen und das Zählen von String-Häufigkeiten hilfreich. Außerdem geht die XML-Struktur der Daten verloren, und der Aufruf der 500 MB großen Datei nimmt eine nicht unerhebliche Zeit in Anspruch. Vor allem sind für die meisten Fragestellungen komplexere Anfragen mit Bedingungen notwendig, die sich nicht abbilden lassen. Hilfreich ist diese Gesamtdatei lediglich für eindeutige Fragestellungen wie zum Beispiel nach der Häufigkeit des `lido:culture-Elements`. Insgesamt empfiehlt sich dieser Ansatz allenfalls als zusätzliches Werkzeug.

Der zweite Ansatz greift auf das Kommandozeilen-Werkzeug *xsltproc*⁸⁷ zurück, mit dem mittels *XSLT-Stylesheets*⁸⁸ komplexe Anfragen auf XML-Dokumenten ausgeführt werden können. Dazu müssen zunächst die einzelnen LIDO-Dateien zu validem LIDO-XML zusammenge-

⁸⁷ Weitere Informationen finden sich unter <http://xmlsoft.org/XSLT/xsltproc.html> (15.01.2019).

⁸⁸ Siehe unten Kap. 6.2.2.

führt werden. Das Shell-Skript in Abbildung 11 erzeugt eine entsprechende XML-Datei, die sämtliche 21.071 Files via *XInclude* einbindet. Wichtig ist, dass das Shell-Skript und die LIDO-Dateien in demselben Verzeichnis liegen und die Ausgabe des Skripts in die neue Gesamtdatei umgeleitet wird.⁸⁹

```
#!/bin/sh
echo '<?xml version="1.0" encoding="UTF-8"?>'
echo '<lido:lidoWrap xmlns:lido="http://www.lido-schema.org" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xi="http://www.w3.org/2001/XInclude"
  xsd:schema="http://www.lido-schema.org http://www.lido-schema.org/schema/v1.0/lido-v1.0.xsd">'
ls record*.xml|while read f; do
  echo '<xi:include href="'$f'"/>'
done
echo '</lido:lidoWrap>'
```

Abbildung 11: Skript zur Zusammenführung sämtlicher Quelldateien

Als Ergebnis liefert das Skript eine Datei wie sie in Abbildung 12 dargestellt ist. In diesem Beispiel wurden zehn Elemente eingebunden.

```
<?xml version="1.0" encoding="UTF-8"?>
<lido:lidoWrap xmlns:lido="http://www.lido-schema.org" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xi="http://www.w3.org/2001/XInclude"
  xsd:schema="http://www.lido-schema.org http://www.lido-schema.org/schema/v1.0/lido-v1.0.xsd">
  <xi:include href="record_DE-MUS-062622_kenom_122793.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122794.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122795.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122796.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122797.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122798.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122799.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122800.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122801.xml"/>
  <xi:include href="record_DE-MUS-062622_kenom_122802.xml"/>
</lido:lidoWrap>
```

Abbildung 12: XML-Datei mit zehn mittels XInclude eingebundenen Quelldateien

Liegt diese Gesamtdatei (z. B. *allRecords.XML*) vor, kann sie mittels XSLT-Stylesheets (z. B. *search.xsl*) folgendermaßen durchsucht werden:

```
xsltproc --xinclude search.xsl allRecords.xml | uniq90
```

Bei einem umfangreichen Ausgangsdatensatz wie in diesem Fall mit 21.071 Dateien hat sich dieses Vorgehen ebenfalls als nicht ideal erwiesen. Die Antwortzeiten betrugen bisweilen über zehn Sekunden.

⁸⁹ Beispielsweise könnte ein Skript-Aufruf wie folgt aussehen: *mkAllRecordsXML.sh > allRecords.xml*.

⁹⁰ *Uniq* ist nicht nötig, verringert aber die Rückgabe von Leerzeilen für jeden Datensatz, der die abgefragten Bedingungen nicht erfüllt. Optional kann ein gepipetes *less* eine bessere Lesbarkeit der Ausgabe bewirken. Diese wird auch durch die Umleitung der Ausgabe in eine eigene Datei erreicht.

Schließlich konnte mit *BaseX*⁹¹ ein geeignetes Tool gefunden werden. *BaseX* ist ein XML-Datenbankmanagementsystem unter anderem für die Speicherung und Abfrage großer XML-Kollektionen. Es ist plattformunabhängig und wird unter einer Open-Source-Lizenz angeboten. Nach Download und Installation der Software kann eine große Zahl von XML-Dokumenten in wenigen Minuten importiert und indiziert werden. Abfragen auf den Daten sind im Anschluss mit den standardisierten Abfragesprachen *XPath* und *XQuery* zielgerichtet und mit kurzen Antwortzeiten möglich.

Mit *XPath*⁹² können XML-Elemente und Attribute gezielt adressiert werden. Sogenannte *XPath*-Achsen, die den Verlauf definieren, in dem die umgebenden Knoten abgearbeitet werden, ermöglichen die Auswertung komplexer Knotenlisten. Außerdem steht eine große Anzahl an *XPath*-Funktionen zur Verfügung. Für die Analyse der LIDO-Daten des Sammlungsportals kam vor allem die *XPath*-Node-Function *text()* zum Einsatz. Auch auf die zahlreichen *XPath*-Möglichkeiten, mit Hilfe von *child::**, *descendant::** etc. oder den entsprechenden Kurzschreibweisen, den Prädikaten, durch die XML-Dokumente zu navigieren, musste kaum zurückgegriffen werden. Die Suche über den absoluten Pfad war in der Regel die schnellere und in jedem Fall eindeutige Variante.

Auch für die Nutzung von *XQuery* sind im gegebenen Kontext rudimentäre Kenntnisse ausreichend. *XQuery* setzt auf *XPath* auf, benutzt jedoch eine völlig andere Syntax, die entfernt an die Datenbankabfragesprache SQL erinnert. Bei der Analyse der LIDO-Datensätze kam das *FLOWR-Prinzip* von *XQuery* zum Einsatz. Der Begriff *FLOWR* resultiert aus den Anfangsbuchstaben der Klauseln *For*, *Let*, *Order-by*, *Where* und *Return*, die häufig für die Formulierung der Abfrage benutzt werden.

Abbildung 13 zeigt die *FLOWR*-Abfrage nach den URIs für die unterschiedlichen Rollen, die Beteiligte innerhalb eines Events einnehmen können. Gleichzeitig wird gezählt, wie häufig die einzelnen Rollen in der Datenbank vorkommen. Die *XQuery*-Funktionen *count()*, *contains()* und *concat()* kommen zum Einsatz. Die *where*-Klausel beschränkt die Ausgabe auf *MARC Relator Terms*⁹³ über die Zeichenfolge *id.loc.gov/vocabulary/relators*. Die ersten Zeilen der Antwort werden unterhalb der Abfrage ebenfalls angezeigt. Der *MARC Relator ctb* (Contributor) kommt zum Beispiel in 9952 der abgefragten Elemente vor.

⁹¹ Eine Einführung in die Funktionalitäten von *BaseX* liefert die Getting-Started-Dokumentation unter http://docs.basex.org/wiki/Getting_Started (15.01.2019). Gleichzeitig ist die ausgiebige Beschäftigung mit der Software für die Betrachtung der LIDO-Metadaten nicht zwingend notwendig. Es wird nur ein geringer Teil der Möglichkeiten von *BaseX* für den konkreten Anwendungsfall benötigt.

⁹² Einen Einstieg in *XPath* und *XQuery* bieten die Websites <https://www.w3schools.com/> (15.01.2019) und <https://www.data2type.de/> (15.01.2019) sowie die Einführungen von Grupe (2018) und Vonhoegen (2018). Umfassendere Informationen hält das W3C bereit: Kay 2017; Clark und DeRose 2016.

⁹³ *MARC Relator Terms* definieren die Rolle eines Akteurs (z. B. Contributor oder Creator). Eine Liste sämtlicher *MARC Relator Terms* findet sich unter <http://id.loc.gov/vocabulary/relators.html> (15.01.2019).

Abbildung 13: XQuery-FLWOR-Anfrage und Antwort

Im Anschluss werden einige für die Datenanalyse relevante Fragestellungen und beispielhafte Code-Snippets aufgeführt.⁹⁴

- Was steht in einem Element?
for \$roleActor //lido:roleActor/lido:conceptID
- Wie oft kommt ein bestimmtes Element vor?
count(//lido:legalBodyName)
- Wie oft steht ein bestimmter Term in einem Element?
count(//lido:roleActor[lido:term="Münzmeister"])
- Was steht wie oft in einem Element?
for \$x in distinct-values(//lido:roleActor/lido:term)
let \$count := count(//lido:roleActor/lido:term[text() = \$x])
- Wie oft kommt ein bestimmtes Attribut vor?
count(//lido:titleSet[@lido:type="former"])
- Was steht in einem Element bei einem bestimmten Attribut
//lido:objectDescriptionSet[@lido:type="usage"]/lido:descriptiveNoteValue
- Welche Attribute kommen in einem Element wie oft vor?
for \$x in distinct-values(//lido:titleSet/@*)
let \$count := count(//lido:titleSet[@*=\$x])

Auf *BaseX* wurde im Rahmen dieser Arbeit insbesondere bei der Erstellung der Mapping-Tabelle zurückgegriffen, nachdem das Zielformat verstanden war und es ganz konkret um die Fragen ging, aus welchen LIDO-Feldern welche RDF-Aussagen zu generieren sind.

⁹⁴ Vollständige beispielhafte BaseX-Abfragen sind auf der CD zu dieser Arbeit enthalten.

5 ASCH-MODELL

Herr Bause, den ich, theuerster Gönner, Ihrem Wohlwollen inständig empfehle, wird einige Kleinigkeiten für das Museum, als ein Gläschen mit 3 Amphibien aus Astrachan, eine Schinesische Statera, ein paar gelbe türkische Papusch mit Saffianen Ueberstrümpfen, 3 kupferne rußische Münzen dort zu überreichen das Vergnügen haben (Asch 1783).

Das Zielformat ist ein RDF-Format, das einen bestimmten Ausschnitt der realen Welt in den Blick nimmt. Dieser Ausschnitt wird in einem *Domain*-Modell beschrieben. Die Basis für die im Rahmen dieser Arbeit zu generierenden Daten als Linked Open Data ist das ASCH-Modell.⁹⁵

Das ASCH-Modell erscheint aus zwei Gründen geeignet: Zum einen ist es auf RDF Statements ausgerichtet, zum anderen wurde es eigens für die Welt der akademischen Sammlungen mit ihren vielfältigen Objekten konzipiert.⁹⁶

5.1 ENTSTEHUNG UND ZWECK

Das ASCH-Modell ist ein Semantic-Web- und Linked-Open-Data-fähiges Metadatenmodell, das im Rahmen des DFG-geförderten Projekts „Entwicklung von interoperablen Standards für die Kontextualisierung heterogener Objekte am Beispiel der Provenienz Asch“ von 2014 bis 2017 an der Georg-August-Universität Göttingen entwickelt wurde (SUB Göttingen 2014a).

Auf Basis des Modells sollen vor allem museale Objekte unterschiedlicher Sammlungen institutionsübergreifend mit RDF Statements beschreibbar sein. Die Metadaten der digitalisierten Objekte sollen für das Semantic Web nutzbar gemacht werden. Ein besonderer Fokus liegt auf der Darstellung und Kontextualisierung von Provenienzen, die gerade für museale Objekte von großer Bedeutung sind (Al-Eryani et al. 2018, S. 548–549).

„Das Modell sollte die Möglichkeit zur einheitlichen strukturierten Beschreibung von Provenienzinformatoren und zur Veröffentlichung der Daten als Linked Open Data bieten, da beides sich als Desiderat bei bisher verwendeten Modellen herauskristallisiert hatte“ (Al-Eryani et al. 2018, S. 549).

⁹⁵ Sämtliche Informationen zum ASCH-Modell sind im ASCH-Wiki enthalten: SUB Göttingen 2014b.

⁹⁶ Gleichzeitig ist das ASCH-Modell nicht auf diese *Domain* begrenzt (Rühle 2018a, 2018b).

Als Namensgeber für das ASCH-Modell diente Georg Thomas von Asch (1729-1807), ein Förderer der Universität Göttingen, der ihr Schenkungen ganz unterschiedlicher Art überließ, die gegenwärtig in den verschiedenen Sammlungen der Universität beherbergt sind.⁹⁷ Die Sammlung Asch mit ihrer großen Bandbreite an Materialien und ihren heterogenen Aufbewahrungsorten bildet im Kleinen genau den Ausschnitt der realen Welt ab, auf den das neue Modell zielt. Am Beispiel einzelner Objekte der Provenienz Asch wurde das Modell entwickelt und validiert (SUB Göttingen 2014a).

Für die Entwicklung des ASCH-Modells wurden auf Basis einer Stakeholder-Befragung *Use Cases* entworfen. Klassen und Eigenschaften konnten somit anhand konkreter Bedarfe konzipiert werden (SUB Göttingen 2017).

Wie bereits im Titel des DFG-Projekts verankert, sind Interoperabilität und Heterogenität wichtige Faktoren für das ASCH-Modell. Informationen zu ganz unterschiedlichen Objekten sollen sich mit externen Informationen vernetzen können. Aus diesem Grund liegen in der Nachnutzung von Standards, der Verlinkung mit ihnen und der maschinenlesbaren Veröffentlichung der modelleigenen Ontologien weitere Schwerpunkte (Rühle 2018c).

5.2 MODELLIERTE KLASSEN

Im Mittelpunkt des ASCH-Modells steht das Sammlungsobjekt, das *Item*. Dieses Item sowie die Ereignisse, die ihm im Laufe seines Lebenszyklus widerfahren, die *Events*, gilt es ebenso zu beschreiben wie seine digitalen Repräsentationen und Evidenzen⁹⁸ zu Objektinformationen.

Abbildung 14 veranschaulicht die grundlegenden Klassen und Properties des ASCH-Modells. Die Klassen werden als Ellipsen visualisiert, ihre Bezeichnungen sind großgeschrieben. Die Properties werden als gerichtete Pfeile dargestellt. Sie sind in Kleinbuchstaben vermerkt. Das Subjekt eines RDF-Triples ist der Ausgangspunkt des Pfeils, das Objekt der Zielpunkt. Einzig die Beziehung zwischen *Item* und *Event* ist in die eine (*wasAffectedBy*) als auch in die andere (*happenedTo*) Richtung beschrieben.⁹⁹

Informationen zur Kardinalität sind in Abbildung 14 nicht enthalten. Es ist also nicht ersichtlich, wie häufig die einzelnen Klassen miteinander in Beziehung stehen können.

⁹⁷ Die Schenkungen umfassen Bücher, Karten, Zeichnungen, Münzen, Naturalien und Ethnographica (SUB Göttingen 2014b).

⁹⁸ Evidenzen, etwa Tagebücher, die den Versand eines Objekts belegen, können im Rahmen dieser Arbeit nicht behandelt werden. Außerdem nehmen sie im Göttinger Sammlungsportal einen vergleichsweise geringen Anteil ein. Weiterführende Informationen zu Evidenzen finden sich bei Al-Eryani et al. 2018, S. 553–554.

⁹⁹ Es ist nur eine Richtung notwendig, da SPAPQL sogenannte *inversePaths*, also umgekehrte Suchen, vorsieht (Harris und Seaborne 2013). Im Zuge dieser Arbeit wird *happenedTo* deshalb vernachlässigt.

Dass ein `Item` beispielsweise von vielen `Events` betroffen sein kann (`wasAffectedBy`), jedoch nur eine `Appearance` (`physicalDescription`) haben darf, wird erst unter Hinzuziehung der Application Profiles des Modells deutlich.¹⁰⁰

Abbildung 14: Das ASCH-Modell. Basisklassen (SUB Göttingen 2014b)

Es ist hingegen ersichtlich, welche Klassen und welche Properties miteinander RDF Statements bilden können. Eine Instanz der Klasse `Item` kann beispielsweise mit der Property `hasAppellation` und einer Instanz der Klasse `Appellation` ein Triple ergeben, nicht aber mit einer Instanz der Klasse `Collection`. `hasAppellation` erwartet also ein Objekt der Klasse `Appellation` oder ihrer Unterklassen.

Damit ist auch vorgegeben, dass der Zielpunkt dieser Verbindungen wiederum eine Klasse sein muss, also kein Literal sein darf. Es werden URIs, respektive Blank Nodes, erwartet. Diese letzte Aussage ist für sämtliche in Abbildung 14 dargestellten Beziehungen gültig und wird in Abschnitt 5.5 näher erläutert.

Von den Klassen `Item` und `Event` gehen die meisten Pfeile aus. Daraus lässt sich bereits ersehen, dass sie gewichtige Klassen im ASCH-Modell sind.

¹⁰⁰ Siehe unten Kap. 5.5.

Die nachfolgende Tabelle¹⁰¹ listet die grundlegenden und im Rahmen dieser Arbeit relevanten Klassen auf und liefert eine kurze Beschreibung.

Klasse	Beschreibung
Item	Sammlungsobjekt; Ding aus der realen Welt
Appearance	Physikalische Charakteristika eines Items (z. B. Maße oder Inschriften)
Appellation	Bezeichnung oder Titel eines Items
Collection	Menge an Items, die zu einem speziellen Thema zusammengetragen wurden (z. B. eine bestimmte Sammlung, der ein Item angehört)
DigitalRepresentation	Digitale Repräsentation eines Items (z. B. ein Bild)
Dimension	Dimension oder Größe eines Items; etwas Quantifizier- oder Messbares (in Zusammenhang mit Appearance)
Event	Aktivität oder Ereignis im Lebenszyklus eines Items
Agent	Person, Organisation oder Gruppe, die mit dem Item über ein Event verknüpft ist
Place	Geographischer Ort, der mit dem Item über ein Event verknüpft ist
Time	Zeitangabe, die mit dem Item über ein Event verknüpft ist
Custody	Aufbewahrungsort als Unterklasse von Event (nicht in Abbildung 14 enthalten)
Record	Datensatz, der ein Item beschreibt (nicht in Abbildung 14 enthalten)

Tabelle 1: Klassen des ASCH-Modells

In der Tabelle ist nicht berücksichtigt, dass für die Klasse `Item` die Unterklassen `Artifact` und `Naturalism` definiert sind, die wiederum Unterklassen besitzen. Ebenso unterteilt sich `Agent` in die Unterklassen `Person`, `Group` und `Corporate Body`. Diese Informationen sind in den LIDO-Ausgangsdaten zwar enthalten, können jedoch vernachlässigt werden, da den Sammlungsobjekten diese Information bereits über ihre Objektgattung bzw. den Agent-URI zugeordnet wird. Es werden also spezifischere Informationen hinzugezogen, die über Vererbung die allgemeineren Informationen automatisch enthalten.

Auch die Klasse `Event` besitzt zahlreiche Unterklassen, die die Art des Ereignisses spezifizieren. Bis auf die Klassen `Custody` (Aufbewahrungsstätte) und `Exploration` (Entdeckungsreise), die ASCH-eigene Klassen sind, wird hier das breite Spektrum der *eventTypes* der *LIDO Terminology*¹⁰² genutzt. Abbildung 15 zeigt den Einstieg in die Liste der *LIDO-eventTypes* und im rechten Bereich einen Teil der Beschreibung des Ereignisses *Beginning of existence*.

¹⁰¹ Die Tabelle wurde in Anlehnung an die Tabelle im ASCH-Wiki (SUB Göttingen 2017) erstellt.

¹⁰² <http://terminology.lido-schema.org/eventType> (15.01.2019).

Selected vocabulary: LIDO-Terminologie Info

- + actor_type (RECORDS IN PROGRESS) (en)
- eventType (RECORDS IN PROGRESS) (en)
 - + Activity (Not for indexing) (en)
 - Beginning of existence (en)
 - + Creation (en)
 - + Production (en)
 - Transformation (en)
 - + End of existence (en)
 - Unspecified event (en)
- + identifier_type (en)

Details: Beginning of existence (en)

Symbolic URI
<http://terminology.lido-schema.org/eventType/beg>

URI
<http://terminology.lido-schema.org/lido00028>

Type
 Concept

In Concept scheme
<http://terminology.lido-schema.org/eventType>

Abbildung 15: LIDO-Terminologie, eventType

Die besondere Bedeutung der Events ist nicht nur anhand der Vielzahl von Unterklassen ersichtlich, sondern zeigt sich vor allem darin, dass Beteiligte (*Agent*), Orte (*Place*) und Zeitangaben (*Time*) nicht direkt mit dem Sammlungsobjekt in Beziehung stehen.¹⁰³ Stattdessen sind sie mit einem Event verknüpft, das das Item beeinflusst hat. Beispielsweise ist ein Item von dem Event *Custody* betroffen. Mit dem Event *Custody* ist dann ein *Agent* verbunden. Im Fall des Göttinger Sammlungsportals ist dies zumeist das aufbewahrende universitäre Institut.

Diese Konzentration auf die Events, die typisch für das ASCH-Modell ist, entspricht der in den Ausgangsdaten vorhandenen und mit dem LIDO-Format eng verknüpften Ereignisorientierung.

5.3 NUTZUNG VON STANDARDS

RDF spielt seine ganze Stärke aus, wenn es nicht isoliert in einer Anwendung betrachtet wird, sondern die benutzten Begriffe etablierten Standards entsprechen. Die größtmögliche Interoperabilität wird erreicht, wenn Klassen und Eigenschaften bereits außerhalb der eigenen Anwendung bekannt und vielfach genutzt sind und somit eine anwendungsübergreifende einheitliche Beschreibung möglich ist.

Aus diesem Grund wird auch im ASCH-Modell so oft wie möglich auf vorhandene Metadatenstandards zurückgegriffen. Al-Eryani et al. (2018, S. 561–562) benennen nachfolgende Standards, die im ASCH-Modell primär Verwendung finden, zum einen, weil sie RDF-fähig sind und zum anderen, weil sie den Ausschnitt der realen Welt berücksichtigen, auf den das ASCH-Modell zielt.¹⁰⁴

¹⁰³ Siehe oben Abbildung 14.

¹⁰⁴ In den Klammern sind die Präfixe angegeben, die in den ASCH Application Profiles als Abkürzungen dienen.

- Dublin Core (dc oder dcterms)
- Europeana Data Model (edm)
- PROV Ontology (prov)
- CIDOC Conceptual Reference Model (crm)
- Darwin Core (dwc)
- Nomisma Ontology (nms)

Während es sich bei *Dublin Core*¹⁰⁵ sowie der *PROV Ontology*¹⁰⁶ um sogenannte *Cross-Domain*-Modelle mit allgemeinem Wirkungskreis handelt, sind die anderen vier Standards als *domain-spezifisch* zu bezeichnen. Die *Nomisma Ontology*¹⁰⁷ liefert einen Thesaurus für numismatische Begriffe. *Darwin Core*¹⁰⁸ wurde für die Beschreibung biologischer Konzepte entworfen. Das *Europeana Data Model (EDM)*¹⁰⁹ zielt auf die Präsentation der Kulturgüter Europas im Europeana-Portal ab. *CIDOC CRM*¹¹⁰ schließlich ist ebenfalls für die Beschreibung von Kulturgütern konzipiert, es ist jedoch deutlich komplexer als EDM. Mit Blick auf die *domain*-übergreifenden Standards lassen sich folgende zentrale Aussagen treffen: *Dublin Core* umfasst zum einen die 15 Kernelemente des *Dublin Core Metadata Element Set (DCMES)*, das sehr allgemeine Beschreibungen für alle Wissensgebiete enthält, und zum anderen die *DCMI Metadata Terms*, die weitere Spezifizierungen vorsehen (Dublin Core Metadata Initiative 2012b, 2012a). Beide Vokabulare wurden von der *Dublin Core Metadata Initiative* entwickelt. Die PROV Ontology stellt ein Vokabular für die Abbildung von Provenienzen zur Verfügung.

Darüber hinaus ist das ASCH-Modell offen für weitere Standards. So kam zum Beispiel im Verlauf dieser Arbeit der Term `foaf:primaryTopic` hinzu. Außerdem ist die Property `foaf:name` in den Application Profiles zu finden. *FOAF*¹¹¹ steht für *Friend of a Friend* und ist eine Ontologie mit 19 Kernbegriffen für die Modellierung sozialer Netzwerke.

Damit dieses *Mixing and Matching* vor allem von Eigenschaften aus unterschiedlichen Ontologien möglich ist, müssen die jeweiligen Klassen entsprechend modelliert sein. Dies realisiert das ASCH-Modell durch die Definition eigener Klassen, die gleichzeitig als Unterklassen von Standards definiert sind. Es bedarf also ASCH-eigener Ontologien. Der nächste Abschnitt widmet sich diesem Thema.

¹⁰⁵ <http://purl.org/dc/elements/1.1/> (15.01.2019) und <http://purl.org/dc/terms/> (15.01.2019).

¹⁰⁶ <https://www.w3.org/ns/prov#> (15.01.2019).

¹⁰⁷ <http://nomisma.org/ontology#> (15.01.2019).

¹⁰⁸ <http://rs.tdwg.org/dwc/terms/> (15.01.2019).

¹⁰⁹ <http://www.europeana.eu/schemas/edm/> (15.01.2019).

¹¹⁰ <http://erlangen-crm.org/170309/> (15.01.2019).

¹¹¹ <http://xmlns.com/foaf/0.1/> (15.01.2019).

5.4 EIGENE ONTOLOGIEN

Bei den in Tabelle 1 beschriebenen Klassen handelt es sich mit einer Ausnahme¹¹² um ASCH-eigene Klassen. Hier wurde also nicht auf vorhandene Klassen eines Standards zurückgegriffen. Der Grund dafür liegt vor allem in dem *Mixing-and-Matching*-Ansatz des Modells. Vorhandene Standards sind beschränkt auf die definierten Terme. Geht es aber darum, flexibel und interoperabel zu sein und aus sämtlichen Standards das Relevante nachzunutzen und Fehlendes selbst zu definieren, ist die Modellierung eigener Klassen als Unterklassen mehrerer Standards das adäquate Mittel. Indem nämlich eigene Klassen als Unterklassen externer Standards definiert werden, erben sie die Eigenschaften der Oberklassen.

Trotz des damit geschaffenen breiten Spektrums an nachnutzbaren Eigenschaften kommt das ASCH-Modell nicht ohne eigene Eigenschaften aus. Nicht alle in der *ASCH-Domain* relevanten Informationen lassen sich mit den vorhandenen Vokabularen abbilden.

Für die ASCH-eigenen Klassen (*Classes*) und Eigenschaften (*Properties*) wurden zwei Ontologien entwickelt, von denen sich die eine vor allem mit dem Sammlungsobjekt, dem Item, und die andere mit dem Event befasst.

- Item Description Ontology (IDO)¹¹³
- Provenance Description Ontology (PDO)¹¹⁴

Beide Ontologien sind über einen *Persistent Uniform Resource Locator (PURL)* erreichbar. Mit diesem URL, der auf einen Linkresolver zeigt, der den tatsächlichen URL des GitHub-Repositorys zurückgibt, wird eine konstante und persistente Information bereitgestellt.¹¹⁵

IDO und PDO sind in dem RDF-Serialisierungsformat Turtle geschrieben. Sie bestehen aus RDF-Triples und sind maschinenlesbar. Wie in Abschnitt 2.4 erwähnt, werden Ontologien vor allem mit den Schemasprachen RDFS und OWL verfasst. So ist es auch im Falle der ASCH-Ontologien. Ihre Kernelemente werden nachfolgend kurz vorgestellt.

Nach der Definition der URI-Präfixe wird zunächst beschrieben, worum es sich bei dem Dokument grundsätzlich handelt, nämlich um eine `owl:Ontology`. Titel (`rdfs:label`), eine Beschreibung (`rdfs:comment`) und die Versionsnummer (`owl:versionInfo`) werden außerdem geliefert. Abbildung 16 zeigt die Darstellung dieser Informationen in fünf Turtle-Triples.

¹¹² Die Klasse `Dimension` ist eine CIDOC CRM-Klasse.

¹¹³ <http://purl.org/asch/ido/180704#> (15.01.2019).

¹¹⁴ <http://purl.org/asch/pdo/180704#> (15.01.2019).

¹¹⁵ Vgl. <https://archive.org/services/purl/> (15.01.2019).

```
<http://purl.org/asch/pdo/180704#> a owl:Ontology ;
    rdfs:label "Provenance Description Ontology (PDO)"@en ;
    rdfs:comment "This ontology defines core classes and properties for the description of provenance."@en ;
    rdfs:comment "Created by Stefanie Rühle (SUB Göttingen)"@en ;
    owl:versionInfo "2018-07-04 (version 3.0)" .
```

Abbildung 16: Provenance Description Ontology

In Abbildung 17 wird die IDO-Klasse `ido:Appearance` definiert, und zwar als `owl:Class` und als Unterklasse von `rdfs:Resource` und `crm:E26_Physical_Feature`. Die persistente Adresse ist mit dem HTTP-URI <http://purl.org/asch/ido/180704#Appearance> ebenfalls gegeben. Der Name der Klasse (`rdfs:label`) ist auf Englisch und Deutsch vermerkt.

```
ido:Appearance a owl:Class ;
    rdfs:label "appearance"@en , "Aussehen"@de ;
    rdfs:subClassOf rdfs:Resource , crm:E26_Physical_Feature ;
    rdfs:isDefinedBy <http://purl.org/asch/ido/180704#Appearance> ;
    rdfs:comment "The physical characteristics of an item."@en .
```

Abbildung 17: Item Description Ontology, Klasse Appearance

Abbildung 18 liefert ein Beispiel für die Definition einer IDO-Eigenschaft oder -Property.

```
ido:isGatheredInto a owl:ObjectProperty ;
    rdfs:label "is gathered into"@en , "in Sammlung"@de ;
    rdfs:subPropertyOf dcterms:isPartOf , edm:isGatheredInto , crm:P46i_forms_part_of ;
    rdfs:isDefinedBy <http://purl.org/asch/ido/180704#isGatheredInto> ;
    rdfs:comment "Relates a resource with a collection of which it is a member."@en .
```

Abbildung 18: Item Description Ontology, Eigenschaft isGatheredInto

Es wird die Eigenschaft `isGatheredInto` definiert. Mit der in der Definition enthaltenen Property `rdfs:subPropertyOf` wird ein Mapping auf die Standards DCTERMS, EDM und CRM realisiert. Eine eigene Property war in dem Fall nötig, weil `edm:isGatheredInto` noch nicht hinreichend dokumentiert¹¹⁶ und `dcterms:isPartOf` zu allgemein ist (Rühle 2018c). `crm:P46i_forms_part_of` schließlich ist hinsichtlich *Range* und *Domain* auf `E18_Physical_Thing`¹¹⁷ beschränkt.

Insgesamt halten sich die ASCH-eigenen Terme in einem überschaubaren Rahmen. In IDO werden 13 Klassen und 12 Eigenschaften und in PDO 9 Klassen und 30 Eigenschaften defi-

¹¹⁶ `edm:isGatheredInto` fehlt zum Beispiel in der Definition of the Europeana Data Model (Europeana Foundation 2016). Es findet jedoch Erwähnung in den EDM mappings, refinements and extensions im Collection Profile (Europeana Foundation 2015).

¹¹⁷ Vgl. http://erlangen-crm.org/docs/ecrm/120111/objectproperties/P46iformspartof_548267046.html (15.01.2019).

niert. Von den 30 PDO-Eigenschaften beziehen sich 12 allein auf Agent-Rollen wie `pdo:hasShipper` oder `pdo:hasUser`.¹¹⁸

Dass das ASCH-Modell breit und interoperabel aufgestellt ist, sei am Ende dieses Abschnitts durch Abbildung 19 dokumentiert. Es sind die in IDO und PDO verknüpften KOS mit Präfix und URI abgebildet.¹¹⁹

```
pdo: <http://purl.org/asch/pdo/180704#> .
ido: <http://purl.org/asch/ido/180704#> .
rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
owl: <http://www.w3.org/2002/07/owl#> .
xsd: <http://www.w3.org/2001/XMLSchema#> .
skos: <http://www.w3.org/2004/02/skos/core#> .
crm: <http://erlangen-crm.org/170309/> .
foaf: <http://xmlns.com/foaf/0.1/> .
gndo: <http://d-nb.info/standards/elementset/gnd#> .
dcterms: <http://purl.org/dc/terms/> .
edm: <http://www.europeana.eu/schemas/edm/> .
prov: <http://www.w3.org/ns/prov#> .
dwc: <http://rs.tdwg.org/dwc/terms/> .
dc: <http://purl.org/dc/elements/1.1/> .
dcat: <http://www.w3.org/ns/dcat#> .
dcmitype: <http://purl.org/dc/dcmitype/> .
aat: <http://vocab.getty.edu/aat/> .
marcrel: <http://id.loc.gov/vocabulary/relators/> .
eventType: <http://terminology.lido-schema.org/eventType> .
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
gnd: <http://d-nb.info/gnd/> .
nms: <http://nomisma.org/ontology#> .
```

Abbildung 19: In PDO und IDO verknüpfte KOS

5.5 APPLICATION PROFILES

Application Profiles (APs) enthalten die für den konkreten Anwendungsfall relevanten Terme.¹²⁰ Der konkrete Anwendungsfall ist im ASCH-Kontext die *Domain* der heterogenen wissenschaftlichen Sammlungen.

In den vorherigen beiden Abschnitten wurde ersichtlich, dass die Terme des ASCH-Modells zum einen aus vorhandenen Standards und zum anderen aus eigenen Ontologien stammen. Typisch für das ASCH-Modell ist das *Mixing and Matching* dieser Terme, das in den Application Profiles dokumentiert ist.

¹¹⁸ Wie in Kap. 6 deutlich werden wird, werden diese Properties für das Sammlungsportal nicht benötigt, weil die Quelldaten weiter verbreitete Properties, die MARC Relator Terms, enthalten.

¹¹⁹ Gleichzeitig liefert der Blick in IDO und PDO keine vollständige Liste sämtlicher im ASCH-Modell genutzter Standards. Dazu müssen die Application Profiles hinzugezogen werden.

¹²⁰ Siehe oben Kap. 2.4.

Insgesamt hält das ASCH-Modell mehr als 20 Application Profiles, respektive *Description Set Profiles (DSP)*¹²¹, vor, die sich in folgende Bereiche unterteilen lassen:

- Item Description Application Profile¹²²
- Provenance Description Application Profile¹²³
- Application Profile for the Description of Metadata Provenance¹²⁴
- TEI¹²⁵ Application Profile for Evidence¹²⁶
- MODS Profile for the Description of Evidence¹²⁷

Mit Blick auf die Ziele dieser Arbeit sind vor allem die beiden ersten APs interessant. Bei der an dritter Stelle genannten AP geht es um Meta-Metadaten, also um Metadaten, die Metadaten beschreiben. Hier findet das *Record Description Set Profile*¹²⁸ Berücksichtigung, das zum Beispiel Terme bereithält, die die Rechte an den Daten beschreiben. Die beiden zuletzt genannten APs behandeln Evidenzen, die im Rahmen dieser Arbeit unberücksichtigt bleiben.¹²⁹

Das Item Description AP¹³⁰ enthält eigene Profile für die schon in Abbildung 14 genannten ASCH-Klassen *DigitalRepresentation*¹³¹, *Appearance*¹³², *Dimension*¹³³ und *Appellation*¹³⁴. Außerdem gibt es hier eigene Profile für *NumismaticItem*¹³⁵ und *Organism*¹³⁶ und *Text*¹³⁷, die jedoch angesichts des begrenzten zeitlichen Rahmens dieser Arbeit nicht betrachtet werden.

¹²¹ Im ASCH-Kontext unterteilt sich ein Application Profile häufig in mehrere Description Set Profiles (DSP). Pro Klasse existiert ein DSP, das die Properties dieser Klasse enthält.

¹²² http://asch.wiki.gwdg.de/index.php/Item_Description_AP (15.01.2019).

¹²³ http://asch.wiki.gwdg.de/index.php/Provenance_Description_AP (15.01.2019).

¹²⁴ http://asch.wiki.gwdg.de/index.php/Meta-Metadata_AP (15.01.2019).

¹²⁵ TEI steht für *Text Encoding Initiative*. Es geht unter anderem um die Strukturierung von Texten und die Kennzeichnung von in ihnen enthaltenen Informationen. Weitere Informationen finden sich auf der Website des *TEI Consortiums* unter <http://www.tei-c.org/> (15.01.2019) und im GitHub-Repository unter <https://github.com/TEIC/TEI-Simple> (15.01.2019).

¹²⁶ http://asch.wiki.gwdg.de/index.php/TEI_Application_Profile (15.01.2019).

¹²⁷ http://asch.wiki.gwdg.de/index.php/MODS_Application_Profile (15.01.2019).

¹²⁸ http://asch.wiki.gwdg.de/index.php/Record_DSP (15.01.2019).

¹²⁹ Siehe oben Kap. 5.2.

¹³⁰ http://asch.wiki.gwdg.de/index.php/Item_Description_AP (15.01.2019).

¹³¹ http://asch.wiki.gwdg.de/index.php/Digital_Representation_DSP (15.01.2019).

¹³² http://asch.wiki.gwdg.de/index.php/Appearance_DSP (15.01.2019).

¹³³ http://asch.wiki.gwdg.de/index.php/Dimension_DSP (15.01.2019).

¹³⁴ http://asch.wiki.gwdg.de/index.php/Appellation_DSP (15.01.2019).

¹³⁵ http://asch.wiki.gwdg.de/index.php/Numismatic_Item_DSP (15.01.2019) und weitere.

¹³⁶ http://asch.wiki.gwdg.de/index.php/Organism_DSP (15.01.2019) und weitere.

¹³⁷ http://asch.wiki.gwdg.de/index.php/Text_DSP (15.01.2019).

Bedeutsam sind hingegen die Provenienzen. Hier sind die Description Set Profiles *Event*¹³⁸, *Agent*¹³⁹, *Place*¹⁴⁰ und *Time*¹⁴¹ vollumfänglich relevant. Schließlich finden Terme aus dem *Holdings Description Set Profile*¹⁴² Anwendung, dessen Kontext der Aufbewahrungsort der Objekte ist.

Anhand der sich anschließenden Beispiele in den Abbildungen 20 bis 22 wird ersichtlich, wie Application Profiles zu lesen sind. Es werden Klassen und Eigenschaften betrachtet, jeweils mit einem Term aus einem externen Standard sowie einer ASCH-eigenen Ontologie.

Event

Term: pdo:Event

Type of Term: Class

URI: <http://purl.org/asch/pdo/180704#Event>

Definition: Something that happened and affected the lifecycle of the resource.

Comment: To specify an event use a class from the list of eventtypes.

Abbildung 20: ASCH Application Profile, Klasse Event

Dimension

Term: crm:E54_Dimension

Type of Term: Class

URI: http://erlangen-crm.org/160331/E54_Dimension

Definition: This class comprises quantifiable properties that can be measured by some calibrated means and can be approximated by values, i.e. points or regions in a mathematical or conceptual space, such as natural or real numbers, RGB values etc.

Abbildung 21: ASCH Application Profile, Klasse Dimension

Der ASCH-Term `pdo:Event` in Abbildung 20 und der CRM-Term `crm:E54_Dimension` in Abbildung 21 werden als Klassen (Type of Term: Class) definiert und beschrieben. Sie sind über einen HTTP-URI adressierbar. Die Definition der Klassen in den Application Profiles entspricht im Wesentlichen der in den Ontologien. Die übergeordneten Klassen sind jedoch nicht enthalten.

Unterhalb der einzelnen Klassen werden in den APs stets die zugehörigen und möglichen Eigenschaften angezeigt. Abbildung 22 zeigt zwei Eigenschaften der Klasse *Event*, `dc:description` und `pdo:end`. Dass es sich in beiden Fällen um Eigenschaften, respektive Properties, handelt, besagt der *Type of Term*.

¹³⁸ http://asch.wiki.gwdg.de/index.php/Event_DSP (15.01.2019).

¹³⁹ http://asch.wiki.gwdg.de/index.php/Agent_DSP (15.01.2019).

¹⁴⁰ http://asch.wiki.gwdg.de/index.php/Place_DSP (15.01.2019).

¹⁴¹ http://asch.wiki.gwdg.de/index.php/Time_DSP (15.01.2019).

¹⁴² http://asch.wiki.gwdg.de/index.php/Holding_DSP (15.01.2019).

description

Term: dc:description
Type of Term: Property
URI: <http://purl.org/dc/elements/1.1/description>
Definition: An account of the resource.
Occurrence: 0-n
Has Domain: pdo:Event
Has Range: rdfs:Literal

end

Term: pdo:end
Type of Term: Property
URI: <http://purl.org/asch/pdo/180704#end>
Definition: Relates to a resource representing the date of ending.
Comment: Use to describe the time the event ended.
Occurrence: 0-1
Has Domain: pdo:Event
Has Range: pdo:Time

Abbildung 22: ASCH Application Profile, Properties description und end

dc:description ist eines der 15 Dublin Core-Kernelemente. Ein Event kann in diesem Fall keine oder beliebig viele Beschreibungen haben. Dies wird durch die *Occurrence* 0-n ausgedrückt. Wichtig sind des Weiteren die Definitionen der *Domain* (Has Domain) und der *Range* (Has Range). Die *Domain* beschreibt das Subjekt eines Triples mit dieser Property. dc:description erwartet also in diesem Kontext ein Subjekt vom Typ pdo:Event. Die *Range* beschreibt das Objekt des Triples. Im Fall von dc:description wird ein rdfs:Literal, also einfacher Text und kein URI, erwartet.

Die zweite Property pdo:end kann keinmal oder genau einmal für dasselbe Subjekt vorkommen (Occurrence: 0-1). Das Subjekt muss ein pdo:Event sein (Has Domain) und das Objekt vom Typ pdo:Time (Has Range). Als Objekt ist in dem Fall kein Literal wie beispielsweise eine Zeitangabe zulässig, sondern lediglich eine Klasse oder ein Blank Node vom Typ pdo:Time.

Wird keine *Domain* oder *Range* angegeben, gibt es keine Beschränkung. Gäbe es also bei pdo:end keine *Range*, wäre als Objekt eine *Resource* oder ein Literal möglich.

Diese Vorgaben und Informationen in den Application Profiles sind bei der Generierung der Triples von fundamentaler Bedeutung. Zukünftige Anwenderinnen und Anwender der RDF-Daten des Sammlungsportals werden die Application Profiles hinzuziehen, um die Struktur der Metadaten zu verstehen.

6 MAPPING

*Linked data is essential to actually connect the semantic web.
It is quite easy to do with a little thought, and becomes second nature (Berners-Lee 1998).*

Während Tim Berners-Lee die Generierung von Linked Data als „quite easy“ und „to do with a little thought“ (1998) einschätzt, stellt Simou fest: „The mapping from one model to the other is not a trivial task“ (2014, S. 2). Simou verweist auf die Gebundenheit an *Domain*- und *Range*-Vorgaben, die mit der Benutzung von Ontologien einhergehen. Eine weitere Schwierigkeit ergebe sich dadurch, dass ein „tree-structured data model“ in ein „graph-RDF structured model“ zu transformieren sei.

Um gerade diese Transformation geht es in der vorliegenden Arbeit und insbesondere in diesem Kapitel. Nachdem in den vorangehenden beiden Kapiteln die Quelldaten und das Zielformat eingeführt worden sind, soll im Folgenden die LIDO-XML-Baumstruktur in die RDF-Struktur, die das ASCH-Modell vorgibt, umgewandelt werden.

6.1 ERARBEITUNG DER MAPPING-VORGABEN

Der erste Schritt besteht in der Erstellung einer Mapping-Tabelle mit Vorgaben, anhand derer die Inhalte aus den LIDO-Daten zusammen mit den Termen aus den ASCH APs in Triples zu transformieren sind. Die Erarbeitung der Vorgaben für die Datenkonvertierung stellt die aufwändigste Aufgabe dieser Master Thesis dar. Die sich daran anschließende Programmierung des Umwandlungscodes ist vergleichsweise einfach, wenn auf gründlich recherchierte und korrekte Vorgaben zurückgegriffen werden kann.

Die Vorgehensweise lässt sich in folgende Arbeitsschritte untergliedern:

- Die Ausgangsdaten werden entlang der *kuniweb*-Dokumentation betrachtet.
- Die Subjekte der Triples sind aus den LIDO-Feldern zu extrahieren.
- Die Prädikate der Triples sind vor allem aus den ASCH-APs zu ermitteln.
- Die Objekte der Triples sind aus den LIDO-Feldern oder den ASCH-APs zu entnehmen.
- Prädikate und Objekte sind stets mit Blick auf die *Range* und *Domain* in den Applikation Profiles auszuwählen.
- Konkrete Inhalte in den LIDO-Feldern sind mit *BaseX*-Abfragen abzusichern.
- Die Wiederholbarkeit der Felder (*Occurrence*) ist zu beachten.
- Die *Linked Data Principles* sind zu berücksichtigen.

Die *kuniweb*-Dokumentation beschreibt die möglichen LIDO-Elemente und geht dabei *Wrap* für *Wrap* vor. Dieser Reihenfolge folgen auch die Mapping-Überlegungen. Jedes *Wrap* wird mit Blick auf das Ziel der Arbeit und die Möglichkeiten des ASCH-Modells untersucht. Der XPath der verwertbaren Elemente sowie die mit ihnen zu generierenden Triples werden notiert. Darüber hinaus wird festgehalten, ob die LIDO-Felder mehrfach belegbar sind.

Auf diese Weise entsteht eine Tabelle mit Vorgaben für die Konvertierung der Daten. Die gesamte Mapping-Tabelle für die Aufgabenstellung dieser Arbeit ist in Anlage 1 enthalten. Sie kann im Folgenden nur in Ausschnitten betrachtet werden.¹⁴³ Durch die Beschreibung ausgewählter Beispiele sollen zum einen die Überlegungen beim Mapping und zum anderen der Umgang mit den APs und DSPs des ASCH-Modells dargelegt werden.

Zunächst gilt es, die Subjekte, also das erste Element der Triples, zu identifizieren. Neben dem zentralen Subjekt, dem Sammlungsobjekt (*Item*), gibt es für jedes Sammlungsobjekt genau einen Datensatz (*Record*), der die Beschreibung des Objekts enthält, und zumeist mehrere digitale Repräsentationen (*DigitalRepresentation*) der Sammlungsobjekte.

Die nachfolgende Tabelle liefert den XPath der drei aus den LIDO-Daten zu extrahierenden Subjekte. Außerdem nennt sie die Triples, die aus dem jeweiligen Subjekt, dem Prädikat *rdf:type* und der entsprechenden ASCH-Klasse als Objekt gebildet werden können.

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
Item/Sammlungs- objekt	/lido:lido/lido:objectPublishedID	Item/rdf:type/pdo:Item	Nein	Enthält Zeichenket- te <i>hdl.handle</i>
Record/Datensatz	/lido:lido/lido:objectPublishedID	Record/rdf:type/pdo:Record	Nein	Enthält Zeichen- kette <i>sammlungen.uni- goettingen.de</i>
/lido:lido/lido:administrativeMetadata/lido:resourceWrap/ (Occ.-0-1)				
Digitale Repräsen- tation des Items (Image/Jpeg)	lido:resourceSet/lido:resource Representa- tion/lido:linkResource	DigRep/rdf:type/ ido:DigitalRepresentation	Ja	Type „provided...“

Tabelle 2: RDF-Subjekte im Göttinger Sammlungsportal

Mit *rdf:type* erhält die *Resource*, identifiziert durch den URI, eine semantische Einordnung. Sämtliche Sammlungsobjekte sind als Instanzen der Klasse *pdo:Item* definiert. Die Datensätze sind Instanzen der Klasse *pdo:Record* und die digitalen Repräsentationen der Klasse *ido:DigitalRepresentation*.

¹⁴³ Die Tabellen 2-6 sind Ausschnitte aus der Mapping-Tabelle in Anhang 1.

Die Anzahl der unterschiedlichen Subjekte der Klassen `Item` und `Record` im gesamten Knowledge Graph entspricht der Anzahl der LIDO-Quelldateien, denn jede Datei beschreibt genau ein Sammlungsobjekt. Hinsichtlich der `DigitalRepresentations` ergibt eine *BaseX*-Anfrage insgesamt 34.544 verwertbare Bilder der 21.071 Sammlungsobjekte.¹⁴⁴

Außerdem kommen Blank Nodes als Subjekte vor. Auch in diesem Fall erfolgt die semantische Einordnung als Instanz einer Klasse mit `rdf:type`.

Die sich anschließende Tabelle 3 ist das Ergebnis der Analyse des *objectClassificationWraps*. Die Quelldaten enthalten einen HTTP-URI, der die Objektgattung¹⁴⁵ identifiziert. Daraus ergibt sich die Generierung eines Triples aus dem Item-URI, der Property `rdf:type` und dem URI der Objektgattung. Dieser Objektgattungs-URI ist so spezifisch, dass auf weitere in den Quelldaten enthaltene allgemeinere Informationen¹⁴⁶ verzichtet werden kann. Diese sollen zukünftig über die Ontologie der Koordinierungsstelle transportiert werden. Ist die Objektgattung dort als Unterklasse anderer Gattungen und Konzepte definiert, lassen sich diese automatisch auf die Göttinger Sammlungsobjekte übertragen.¹⁴⁷

Auch mit Blick auf die Klassifikation werden nur URIs, abgelegt im LIDO-Feld *conceptID*, berücksichtigt und ebenfalls mit der `rdf:type`-Property mit dem Item zu einem Triple verbaut.

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wiederholbar?	Kommentar
/lido:lido/lido:descriptiveMetadata/lido:objectClassificationWrap/ (Occ. 1)				
Objektgattung	lido:objectWorkTypeWrap/lido:objectWorkType/lido:conceptID	Item/rdf:type/URI	Nein	Für jedes Item ist eine Objektgattung vorhanden.
Klassifikation	lido:classificationWrap/lido:classification/lido:conceptID	Item/rdf:type/URI	Ja	Nur URIs werden berücksichtigt.

Tabelle 3: Mapping-Tabelle, Objektgattung und Klassifikation

Tabelle 4 nimmt das *objectIdentificationWrap* in den Blick, in dem die Bezeichnungen der Sammlungsobjekte und ihre Inschriften notiert sind. Die Bezeichnung ist als Literal in dem LIDO-Feld *appellationValue* hinterlegt. Laut *kuniweb*-Dokumentation wird die bevorzugte

¹⁴⁴ Es werden nur die Repräsentationen mit dem Typ http://terminology.lido-schema.org/resourceRepresentation_type/provided_representation berücksichtigt. Vgl. xQuery-Datei `resourcWrap_count_allDigRep.xq` auf der der Arbeit beiliegenden CD.

¹⁴⁵ Die URIs der Objektgattungen werden von der Koordinierungsstelle für wissenschaftliche Universitäts-sammlungen in Deutschland unter <https://portal.wissenschaftliche-sammlungen.de/ConceptScheme/33> (15.01.2019) gepflegt. Aktuell sind die URIs der Objektgattungen noch nicht maschinenlesbar.

¹⁴⁶ Die Quelldaten enthalten z. B. die sehr allgemeinen CIDOC CRM Konzepte *Man-Made Object*, *Biological Object* und *Physical Object* in dem Feld `lido:category`.

¹⁴⁷ An diesem Beispiel wird der automatische Erkenntnisgewinn ersichtlich, der durch Vererbung erreicht werden kann.

Bezeichnung in den LIDO-Daten mit dem Attribut *sortorder=1* gekennzeichnet. Das ASCH-Modell sieht für die Bezeichnungen eines Sammlungsobjekts die Property *ido:hasAppellation* vor, die als Objekt nur URIs der Klasse *ido:Appellation* zulässt. Deshalb ist ein Blank Node der Klasse *ido:Appellation* zu generieren. Dieser Blank Node stellt das Subjekt für die daraufhin zu bildenden Triples dar. In dem ASCH Appellation DSP finden sich die Properties *skos:prefLabel* und *skos:altLabel* für die Bezeichnung; diese erlauben als Objekt ein Literal. Somit wird ein Triple mit *skos:prefLabel* und dem mit *sortorder=1*-Attribut gekennzeichneten Literal generiert und für weitere Namen die Property *skos:altLabel* verwendet. Außerdem ist bei Literalen stets zu erwägen, ob ein *language-* oder *datatype*-Attribut hinzuzufügen ist.¹⁴⁸

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wiederholbar?	Kommentar
/lido:lido/lido:descriptiveMetadata/lido:objectIdentificationWrap/ (Occ. 1)				
Titel Bezeichnung	lido:titleWrap/lido:titleSet/ lido:appellationValue	Item/ido:hasAppellation/BN BN/rdf:type/ido:Appellation BN/skos:prefLabel/Literal	Nein	skos:prefLabel bei Attribut <i>sortorder = '1'</i>
		bzw. BN/skos:altLabel/Literal	Ja	Ansonsten skos:altLabel
Inschrift	lido:inscriptionsWrap/lido:inscriptions/ lido:inscriptionTranscription	Item/ido:physicalDescription/ BN BN/rdf:type/ido:Appearance BN/ido:inscription/Literal	Ja	Kein language-Attribut, da Daten nicht eindeutig; Typ der Inschrift nicht aussagekräftig; Abgleich mit objectMeasurementWrap, da es nur eine ido:Appearance geben darf

Tabelle 4: Mapping-Tabelle, Titel und Inschrift

Die Überlegungen sind bei den Inschriften ähnlich. Das LIDO-Feld enthält ein Literal. Im ASCH-Modell gibt es die Klasse *ido:Appearance* für sämtliche Informationen zum Erscheinungsbild eines Sammlungsobjekts. Jedes Item kann gemäß ASCH Item DSP eine oder keine *ido:Appearance*, verknüpft mit der Property *ido:physicalDescription*, haben. Diese Property lässt als Objekt außerdem nur Instanzen der Klasse *ido:Appearance* zu. Es muss somit ein Blank Note der Klasse *ido:Appearance* generiert werden, mit dem dann ein Triple mit *ido:inscription* und dem Inschriften-Literal gebildet werden kann. Die Property *ido:inscription* findet sich im ASCH Appearance DSP. Die Quelldaten enthalten zum Teil

¹⁴⁸ Die intensive Beschäftigung mit den Language- und Datatype-Attributen übersteigt den zeitlichen Rahmen dieser Arbeit. Bei uneinheitlichen Daten wurde auf die Verwendung der Attribute verzichtet.

mehrere Inschriften für ein Sammlungsobjekt; in dem Fall sind mehrere Triples mit demselben Blank Node als Subjekt, der Property `ido:inscription` und dem jeweiligen Literal als Objekt zu generieren. Die in der *kuniweb*-Dokumentation erwähnten unterschiedlichen Inschrift-Typen werden in dieser Anwendung ignoriert. Die Datenanalyse hat hier ergeben, dass in den Quelldaten kaum differenziert wird und der Mehrwert dieser Information gering einzuschätzen ist. Wichtig ist hingegen die ASCH-Vorgabe, dass ein Item nicht mehr als eine `ido:Appearance` haben darf. Sämtliche Informationen zum Erscheinungsbild eines Sammlungsobjekts, etwa in anderen Wraps, müssen somit mit dem einen Blank Node vom Typ `ido:Appearance` verknüpft werden. Diese Vorgabe ist in der Kommentar-Spalte der Tabelle vermerkt und kommt in Zusammenhang mit dem *objectMeasurementWrap* zum Tragen, das Informationen zu den Maßen der Sammlungsobjekte enthält.

Ein wesentlicher Teil der Informationen zu einem Sammlungsobjekt wird über seine Events transportiert. Im Folgenden werden die Überlegungen zu den Elementen *eventType* und *eventActor* des *eventWraps*¹⁴⁹ nachvollzogen. Die Ergebnisse finden sich zusammengefasst in Tabelle 5.

Gemäß der *kuniweb*-Dokumentation und Analyse der Quelldaten kann pro Sammlungsobjekt kein oder ein *eventWrap* vorliegen. Ist eines vorhanden, gibt es 1-n *eventSets*. Jedes *eventSet* enthält ein Event. Es können somit beliebig viele Events pro Sammlungsobjekt vorkommen.¹⁵⁰ Dies passt zum ASCH-Modell: Die Property `ido:wasAffectedBy` hat eine *Occurrence* von 0-n und erwartet als Objekt eine Klasse. Das Objekt ist in dem Fall immer ein Blank Note, weil es aktuell keine eigenen URIs für jede Item-Event-Kombination gibt. Die Quelldaten enthalten jedoch URIs der LIDO Terminology¹⁵¹, die den Event-Typ identifizieren. Somit kann aus dem Blank Node zusammen mit `rdf:type` und diesem Event-Typ-URI ein weiteres Triple gebildet werden.

Eine Besonderheit liegt bei den LIDO-Event-Typen *Use* und *Production* vor. Hier müssen die Triples in Abstimmung mit dem *objectIdentificationWrap* gebildet werden, das ebenfalls Informationen zum Gebrauch und zur Produktion der Items enthalten kann. Wichtig ist, dass diese Events pro Item nur einmal vorkommen. Außerdem enthält das *repositoryWrap* Informationen zum aktuellen Aufbewahrungsort des Sammlungsobjekts. Da das ASCH-Modell die Aufbewahrung als Event definiert, sind Triples mit `ido:wasAffectedBy` und *Custody*-Event anzulegen.¹⁵²

¹⁴⁹ Außerdem kann das *eventWrap* die Elemente *eventDate*, *eventPlace*, *culture* und *eventMaterialsTech* enthalten. Ihre Vorgaben zur Konvertierung sind in Anhang 1 notiert.

¹⁵⁰ Die *kuniweb*-Dokumentation nennt maximal 11 Events pro Item; unter Berücksichtigung des *naniweb*-Erfassungsportals, das die Anzahl der Events nicht beschränkt, wird hier keine Grenze angenommen.

¹⁵¹ Siehe oben Kap. 5.2.

¹⁵² Jedes der 21.071 Items hat mindestens einen Aufbewahrungsort und somit ein individuelles Event der Klasse *Custody*, repräsentiert durch einen Blank Node.

Jedes *eventSet* kann beliebig viele mit dem Event verknüpfte Akteure verzeichnen. Das Subjekt der Triples ist wiederum der im vorherigen Absatz genannte Blank Node, der die Item-Event-Verknüpfung identifiziert. Das ASCH-Modell sieht als Property `pdo:participant` vor. Diese Property kommt jedoch nur zur Anwendung, wenn den Quelldaten die spezifischere Information zur Rolle des Akteurs in Form eines URIs des MARC Relator Terms fehlt.¹⁵³

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
/lido:lido/lido:descriptiveMetadata/lido:eventWrap/lido:eventSet (Occ. 0-1/1-n)				
Typ des Events	<code>lido:event/lido:eventType/ lido:conceptID</code>	<code>Item/ido:wasAffectedBy/BN BN/rdf:type/URI</code>	Ja	Abgleich mit den Events Use und Production aus dem object-DescriptionWrap notwendig
Akteur des Events	<u>URI des eventActor:</u> <code>lido:event/lido:eventActor /lido:actorInRole/lido:actor/ lido:actorID</code> <u>Rolle des eventActor:</u> <code>lido:event/lido:eventActor/ lido:actorInRole/lido: roleActor/lido:conceptID</code>	<code>BN/marcrel:.../URI bzw. BN/pdo:participant/URI</code>	Ja	Nur URIs! Falls die Daten die marcrel-Property enthalten, soll diese genutzt werden, ansonsten <code>pdo:participant</code> . Mehrere eventActors mit je mehreren Rollen möglich

Tabelle 5: Mapping-Tabelle, Typ und Akteur des Events

Die Mapping-Tabelle kann nicht sämtliche Überlegungen abbilden. So ist Abbildung 23 ein weiteres Ergebnis der intensiven Beschäftigung mit der *kuniweb*-Dokumentation und den Quelldaten des LIDO-Elements *eventActor*. Es reicht nicht aus, das relevante Element auszumachen, sondern seine Wiederholbarkeit und Abhängigkeit von anderen Elementen müssen ebenfalls verstanden sein, um die Datenkonvertierung korrekt umzusetzen.

```

eventActor 0-n
  displayActorInRole 1 (unwichtig Literal)
  actorInRole 0-1
    actor 1
      actorID 0-1
      nameActorSet 1 (unwichtig Literal)
      roleActor 1-n (ein Akteur kann mehrere Rollen haben)
      conceptID 1-3 (marcrel, GND und AAT)

```

Abbildung 23: Struktur der LIDO-Ausgangsdaten, eventActor

¹⁵³ Siehe unten Kap. 6.2.2.

Die Quelldaten können beliebig viele Elemente *eventActor* (0-n) enthalten; diese haben maximal ein *actorInRole*-Element (0-1), das wiederum in jedem Fall einen *actor* mit *nameActorSet* und maximal einer *actorID* hat. Außerdem sind pro *actor* 1-n *roleActor*-Elemente mit 1-3 URIs im Feld *conceptID* vorhanden. Ein Akteur kann also mehrere Rollen besitzen, die jeweils mit bis zu 3 URIs identifiziert werden. Es wurde im Rahmen dieser Arbeit entschieden, von diesen drei URIs, die allesamt die Rolle des Akteurs mit einem Standard-Vokabular beschreiben und somit redundant sind, einzig die MARC Relator Terms zu berücksichtigen.¹⁵⁴ Eine weitere Entscheidung, die bei der Erstellung der Mapping-Tabelle zu treffen war, bestand darin, nur Akteure zu berücksichtigen, die eine *actorID* und somit einen URI haben. Die Abbildung gibt ebenfalls an, welche Bereiche als unwichtig eingeschätzt wurden.

Nachdem bisher ausschließlich das Item beschreibende Elemente (*descriptiveMetadata*) berücksichtigt wurden, soll dieser Abschnitt mit einem Blick auf die administrativen Daten (*administrativeMetadata*) enden. Das *recordWrap* enthält Informationen zu den Verwertungsrechten der Daten und den Rechteinhabern. Das Subjekt der RDF Statements ist in diesem Fall der URI, der das Record, also den Datensatz, identifiziert.

Abbildung 24 zeigt die relevanten Felder des *recordWraps* mit seinen *Occurrences*.

Abbildung 24: Struktur der LIDO-Ausgangsdaten, *recordWrap*

Jeder Datensatz hat als Pflichtfeld genau ein *recordWrap*. Die Rechte an den Metadaten sowie der Rechteinhaber sind dort im Element *recordRights* hinterlegt. Dieses kommt zweimal vor, wobei das zweite Element stets den Verwertungsrecht-URI im Feld *rightsType/conceptID* enthält, während das erste Element in *rightsType/term* den Begriff Verwertungsrecht aufweist.¹⁵⁵ Im ersten *recordRights*-Element wird unter *rightsHolder* der Rechteinhaber zumeist

¹⁵⁴ Die DINI-AG empfiehlt unter Berücksichtigung der weiten Verbreitung die Verwendung von MARC Relator Terms (DINI-AG KIM / Gruppe Titeldaten 2018, S. 6).

¹⁵⁵ Dies konnte mit einer *BaseX*-Abfrage verifiziert werden. Vgl. *BaseX*-Abfragen, die auf der CD zur Arbeit enthalten sind.

Mapping

mit einem URI in *conceptID* und immer mit Namen als Literal in *legalBodyName/appellationValue* angegeben.¹⁵⁶

Diese Quelldatenanalyse ergibt zusammen mit den Informationen in den ASCH Record DSP die in der sich anschließenden Tabelle notierten Mapping-Vorgaben für den Datensatz.

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Triple wiederholbar?	Kommentar
/lido:lido/lido:administrativeMetadata/lido:recordWrap/ (Occ.-1)				
Record/ Datensatz Verknüpfung Datensatz mit Item	<u>Record:</u> /lido:lido/lido:objectPublishedID	<u>Record/foaf:primaryTopic/Item</u> <u>Record/rdf:type/pdo:Record</u>	Nein	Record enthält <i>sammlungen.uni-goettingen.de</i> ; Pro Item gibt es ein Record.
Recht/Lizenz für die Ver- wertung des Datensatzes	lido:recordRights[2]/lido:rightsType/lido:conceptID	<u>Record/dc:rights/URI</u>	Nein	
Rechteinhaber des Datensatzes	<u>URI:</u> lido:recordRights[1]/lido:rightsHolder/lido:legalBodyID <u>Literal:</u> lido:recordRights[1]/lido:rightsHolder/lido:legalBodyName/lido:appellationValue	<u>URI:</u> <u>Record/dcterms:rightsHolder/URI</u> <u>Literal:</u> <u>Record/dcterms:rightsHolder/BN</u> <u>BN/rdf:type/pdo:Agent</u> <u>BN/foaf:name/Literal</u>	Nein	
Datum der Erstellung des Datensatzes	lido:recordInfoSet/lido:recordMetadataDate	<u>Record/dcterms:created</u> <u>@rdf:datatype/Literal</u>	Nein	recordInfoSet- type lido00470 and record- MetadataDate- type created

Tabelle 6: Mapping-Tabelle, Record

Die Property *foaf:primaryTopic* stellt die Verbindung zwischen Item und Record her. Das Record wird mit dem Prädikat *dc:rights* mit dem Recht verbunden. Im ersten *recordRights*-Element wird unter *rightsHolder* der Rechteinhaber zumeist mit einem URI in *conceptID* und immer mit Namen als Literal in *legalBodyName/appellationValue* angegeben.¹⁵⁷ Sollte kein URI des Rechteinhabers im Datensatz enthalten sein, wird das Literal als Objekt des Triples übernommen, Prädikat ist stets *dcterms:rightsHolder*. Da dieses Prädikat gemäß den

¹⁵⁶ Die hier beschriebene Hinterlegung der Rechte und Rechteinhaber in den LIDO-Daten ist aus der *kuniweb*-Dokumentation nur schwer ersichtlich. Zahlreiche *BaseX*-Abfragen und Analysen waren notwendig, um den Aufbau der Daten zu verstehen. Entsprechendes gilt für das *resourceWrap*.

¹⁵⁷ Dies haben ebenfalls *BaseX*-Abfragen ergeben. Die Suche nach *rightsHolder* mit *recordRights[2]* liefert keine Ergebnisse.

ASCH-APs nur Objekte vom Typ `pdo:Agent` zulässt, ist im Falle eines Literals ein Blank Node zu definieren, der über `foaf:name` mit dem Literal zu verknüpfen ist. Das Datum der Datensatz-Erstellung wird mit dem Prädikat `dcterms:created` und dem Datum als Literal mit Datentyp hinterlegt.

Die ausgewählten Beispiele liefern einen Eindruck vom Vorgehen bei der Erstellung der Mapping-Vorgaben, die in der Mapping-Tabelle zusammengefasst sind. Gleichzeitig wird ersichtlich, dass die Mapping-Tabelle die Komplexität mancher Überlegungen nicht vollständig abbildet. Zum einen wäre eine detailliertere Beschreibung zu Lasten der Übersichtlichkeit der Tabelle gegangen, zum anderen bestand die Notwendigkeit nicht, weil die Erarbeitung der Vorgaben und die Umsetzung in ein Programm in einer Hand lagen. Die Einzelheiten waren zu jedem Zeitpunkt bekannt.

6.2 KONVERTIERUNG IN LINKED OPEN DATA

Liegen mit der Mapping-Tabelle konkrete Vorgaben für die Programmierung der Datenumwandlung vor, sind als nächste Schritte folgende Entscheidungen zu treffen:

- In welches Serialisierungsformat soll konvertiert werden?
- In welcher Sprache und mit welchem Editor soll programmiert werden?

6.2.1 RDF/XML UND TURTLE

Das Zielformat ist ein Semantic-Web-fähiges Format, ein RDF-Serialisierungsformat.¹⁵⁸

Spätestens zu Beginn der Programmierarbeit an der Transformationsdatei muss das Serialisierungsformat feststehen. Die Entscheidung fiel in diesem Fall auf RDF/XML, weil eine Konvertierung von einem XML-Format (LIDO) in ein anderes (RDF) besonders gut von XML-Editoren unterstützt wird.

In RDF/XML werden die Bestandteile der Triples als XML Elemente dargestellt. Die wichtigsten Elemente sind das root-Element `RDF` und das `Description`-Element, das eine *Resource* mit dem `about`-Attribut identifiziert und das Elemente enthält, die die *Resource* beschreiben. Diese enthaltenen Elemente sind entweder die Prädikate der Triples oder ein weiteres `Description`-Element, das einen Blank Node, und somit ein neues Subjekt, repräsentiert.

In Abbildung 25 sind die Triples zu einem Record (Datensatz) in RDF/XML dargestellt. Anhand des Beispiels lassen sich die meisten für diese Arbeit wichtigen Aspekte erläutern.

¹⁵⁸ Siehe oben Kap. 2.5.


```
<rdf:Description rdf:about="http://sammlungen.uni-goettingen.de/objekt/record_kuniweb_1113859">
  <foaf:primaryTopic rdf:resource="https://hdl.handle.net/21.11107/ed388193-6b0b-4028-89dc-c8bd799b505f"/>
  <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Record"/>
  <dcterms:created rdf:datatype="http://purl.org/dc/terms/W3CDTF">2017-11-29T06:44:52+01</dcterms:created>
  <dc:rights rdf:resource="https://creativecommons.org/publicdomain/zero/1.0/deed.de"/>
  <dcterms:rightsHolder>
    <rdf:Description>
      <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Agent"/>
      <foaf:name>Rechnermuseum der GWDG</foaf:name>
    </rdf:Description>
  </dcterms:rightsHolder>
</rdf:Description>
```

Abbildung 25: RDF/XML-Darstellung eines Records

Das Attribut `rdf:about` nennt das Subjekt, in diesem Fall den URI, der den Datensatz repräsentiert. Das Prädikat des ersten Triples ist `foaf:primaryTopic`. Das Objekt ist das Sammlungsobjekt, identifiziert durch den handle-URI und transportiert in dem Attribut `rdf:resource`. Das nächste Triple beschreibt den Typ mit dem Prädikat `rdf:type` und einem URI-Objekt. Das dritte Triple enthält als Objekt ein Literal. Es gibt kein Attribut `rdf:resource`, sondern Text zwischen einem Start- und End-Tag `dcterms:created`. Außerdem enthält das Element ein `rdf:datatype`-Attribut, das dem Literal den Datentyp *date* (*W3CDTF*) zuweist. Das vierte Triple enthält die Rechte, die durch das Prädikat `dc:rights` und eine URI im `rdf:resource`-Attribut transportiert werden. Beim Rechteinhaber (`dcterms:rightsHolder`) wird mit `rdf:Description` ein Blank Node eingeleitet, der das Subjekt von zwei weiteren Triples ist. Zum einen wird der Typ des Blank Nodes mit dem `rdf:resource`-Attribut und zum anderen der Name des Rechteinhabers als Literal dargestellt.

Auch wenn die Beschriftung der Knoten und Kanten in der folgenden Abbildung 26 nicht lesbar ist, visualisiert dieser Graph, der dem RDF/XML aus Abbildung 25 entspricht, das zuvor Beschriebene. In Summe liegen sieben Triples vor, von denen ein Triple (das untere) auf einen Blank Node verweist, der wiederum das Subjekt von zwei neuen Triples ist. Insgesamt haben zwei Triples ein Literal als Objekt, was durch die Rechtecke gekennzeichnet ist.

Abbildung 26: Graph eines Records

Zusammenfassend lässt sich Folgendes zu RDF/XML festhalten: Bei der Erstellung des Transformationscodes ist vor allem zu unterscheiden, ob für das Objekt ein URI vorliegt, woraufhin das `rdf:resource`-Attribut zu benutzen ist, oder als Literal, woraufhin die Zeichenfolge

zwischen Start- und End-Tag der Property zu positionieren ist. Im Falle eines Literals ist ferner eine Spezifizierung des Datentyps mit `rdf:datatype`¹⁵⁹ oder der Sprache mit `rdf:language` zu erwägen. Bezogen auf Blank Nodes gilt, dass sich diese durch eine weitere `rdf:Description`-Ebene generieren lassen.

Im Gegensatz zu RDF/XML sind die Triples im Serialisierungsformat Turtle leicht ersichtlich. Aus diesem Grund kommt Turtle in dieser Arbeit bei der manuellen Validierung der Ergebnisse zum Einsatz.

Die für das Record in Abbildung 25 gebildeten Triples werden in Turtle wie folgt dargestellt:

```
<http://sammlungen.uni-goettingen.de/objekt/record_kuniweb_1113859>
  foaf:primaryTopic      <https://hdl.handle.net/21.11107/ed388193-6b0b-4028-89dc-c8bd799b505f> ;
  rdf:type               pdo:Record ;
  dcterms:created        "2017-11-29T06:44:52+01"^^dcterms:W3CDTF ;
  dc:rights              <https://creativecommons.org/publicdomain/zero/1.0/deed.de> ;
  dcterms:rightsHolder   _:b15 .

_:b15  rdf:type      pdo:Agent ;
       foaf:name     "Rechnermuseum der GWDG" .
```

Abbildung 27: Turtle-Darstellung eines Records

Dem Subjekt, identifiziert durch den URI in Winkelklammern, folgen untereinander die zugehörigen Prädikate mit ihren Objekten. Das Semikolon zeigt an, dass es ein weiteres Triple mit demselben Subjekt gibt. Nach einem Punkt folgt ein neues Subjekt. Literale stehen in Anführungszeichen. Der Datentyp wird mit doppeltem Zirkumflex (^^) eingeleitet.¹⁶⁰ Die Blank Nodes, gekennzeichnet durch `_:b`, werden pro Datensatz durchgezählt. Eine weitere Möglichkeit, Blank Nodes in Turtle darzustellen, besteht in der Kennzeichnung durch eckige Klammern.

In dem RDF/XML- sowie dem Turtle-Beispiel oben kommen Präfixe zur Abkürzung von URIs aus einzelnen Namensräumen (z.B. foaf) zum Einsatz. Diese Abkürzungen, die den Text leichter lesbar machen, werden in beiden Formaten auf ähnliche Weise am Anfang einer Datei definiert.¹⁶¹

¹⁵⁹ Das in dieser Arbeit erstellte Mapping enthält den Datentyp *date* für Zeitangaben und den Datentyp *float* bei der Angabe der Ortskoordinaten. Eine detaillierte Beschäftigung mit den Datentypen konnte im Rahmen dieser Arbeit aus Kapazitätsgründen nicht erfolgen.

¹⁶⁰ Außerdem könnte ein Sprachcode, eingeleitet durch das At-Zeichen (@), das Literal spezifizieren.

¹⁶¹ Definition des foaf-Namensraums in RDF/XML: `<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/">`
Definition des foaf-Namensraums in Turtle: `@prefix foaf: http://xmlns.com/foaf/0.1/.`

6.2.2 XSLT-CODE

XSL Transformation (XSLT) ist eine Programmiersprache zur Umwandlung von XML-Dokumenten und bietet sich somit für die Konvertierung der LIDO-XML-Daten in RDF/XML-Daten an. Das im Rahmen dieser Arbeit entwickelte XSLT-Stylesheet ist in Anhang 2 enthalten.

Eine allgemeine Beschreibung von XSLT liefert Vonhoegen (2018, S. 259–260):

XSLT ist eine in XML definierte Sprache, die ihrerseits den Prozess der Umwandlung eines XML-Dokuments in ein anderes XML-Dokument oder ein anderes Ausgabeformat beschreibt. Das Quelldokument und das Ergebnisdokument, das mit Hilfe des Stylesheets erzeugt wird, können dabei beliebig weit voneinander abweichen. [...] Das Stylesheet mag beispielsweise nur einen Teil der Informationen aus dem Quelldokument beachten und das Ergebnisdokument um beliebig viele Informationen erweitern, die nicht im Quelldokument vorhanden sind. [...] Die Quelldatei muss wohlgeformt sein, mehr nicht. [...] Die Transformation wird kodiert in Form von XSLT-Stylesheets, die selbst wohlgeformte XML-Dokumente sind.

XML-Dokumente lassen sich mit jedem Texteditor erstellen, gleichzeitig empfiehlt sich für die Arbeit mit XML-Dateien ein XML-Editor bzw. eine XML-Entwicklungsumgebung. Auf diese Weise kann beispielsweise das Setzen der Ende-Tags und die Überprüfung der Groß- und Kleinschreibung an ein Programm übergeben werden. Die Eingabe kann außerdem gegenüber einem vorhandenen Schema überprüft werden, und Syntax-Highlighting erleichtert die Lesbarkeit. Des Weiteren lässt sich die Hierarchie der Elemente zumeist in einer Bauman-sicht anzeigen, und der XPath liegt mit einem Mausklick vor. Komplette XML-Entwicklungsumgebungen unterstützen die Arbeit darüber hinaus durch kontextsensitive Elementlisten oder Attribut-Inspektoren. Häufig sind auch XML-Prozessoren integriert, die die Wohlgeformtheit und Gültigkeit der Dokumente schon während der Entwicklung überprüfbar machen.¹⁶²

In dieser Arbeit wurde das XSLT-Stylesheet mit dem *Oxygen XML-Editor 20.1*¹⁶³ erstellt. Das Vorgehen lässt sich wie folgt beschreiben: Grundlage bildet die Mapping-Tabelle, die es Schritt für Schritt abzuarbeiten gilt. Die dort notierten Triples sollen über das Stylesheet generiert werden. Anhand einer LIDO-Sample-Quelldatei¹⁶⁴, die im Verlauf der Arbeit an dem

¹⁶² Vonhoegen (2018, S. 38) nennt als Beispiel für einen einfachen XML-Editor die von Microsoft seit Anfang 2007 angebotene freie Version von XML Notepad und als Beispiele für XML-Entwicklungswerkzeuge XMLSpy und Stylus Studio.

¹⁶³ <https://www.oxygenxml.com/> (15.01.2019).

¹⁶⁴ Die Sample-Datei ist auf der CD zur Arbeit enthalten.

Stylesheet um fehlende, aber mögliche LIDO-Elemente erweitert wird, werden alle Programmierarbeiten unmittelbar überprüft. Der in Oxygen enthaltene XSLT-Prozessor liest Quell- und Transformationsdatei auf Anforderung ein, führt die im Stylesheet enthaltenen Anweisungen durch und zeigt die entsprechende RDF/XML-Ausgabedatei im Anschluss an.

Abbildung 28 vermittelt einen Eindruck von dem Vorgehen bei der Programmierung des Umwandlungscodes.

Abbildung 28: Arbeiten mit Oxygen

Grün umrandet sind links die Sample-Quelldatei und rechts das XSLT-Stylesheet dargestellt. Der XSLT-Prozessor *Saxon*¹⁶⁵ ist links rot markiert. Die Ausgabedatei kann zu jedem Zeitpunkt erzeugt werden. Sie wird in der obigen Abbildung im unteren Bereich angezeigt.

Die Elemente und Attribute, über die der XSLT-Prozessor seine Anweisungen erhält, müssen dem Namensraum *http://www.w3.org/1999/XSL/Transform* entstammen; als Präfix wird *xsl* verwendet. Ebenso sind die Namensräume sämtlicher im Code verwendeter URI-Abkürzungen zu definieren.

Das Wurzelement einer XSLT-Datei ist *xsl:stylesheet*, in das sämtliche Komponenten einzufügen sind. Dieses Element verlangt ein *version*-Attribut; dieses ist in der vorliegenden XSLT-Datei *Version 2.0*.

¹⁶⁵ <http://saxon.sourceforge.net/> (17.01.2019).

Von den möglichen direkten Kindern von `xsl:stylesheet`, den sogenannten Top-Level-Elementen, kommen `xsl:output`, `xsl:template` und `xsl:variable` in dieser Arbeit zur Anwendung. `xsl:output` legt die Methode fest, die bei der Erzeugung der Ausgabedatei beachtet werden soll. Im vorliegenden Fall wird mit `indent="yes"` eine Ausgabe erzeugt, die der hierarchischen Struktur des Dokuments durch Einzüge Rechnung trägt. `xsl:variable` wird zur Definition von Variablen genutzt. Variablen dienen im Rahmen dieser Arbeit vor allem der besseren Übersichtlichkeit.

Die wichtigsten Anweisungen an den XSLT-Prozessor erfolgen in einer beliebigen Abfolge von Template-Regeln, eingepackt in das Element `xsl:template`. Das Attribut `match` gibt das Suchmuster an, auf das die formulierten Regeln anzuwenden sind. Das Suchmuster „/“ findet beispielsweise den Wurzelknoten, woraufhin die Regeln in diesem Template auf sämtliche Knoten anzuwenden sind. Die innerhalb dieses Templates aufgeführte Anweisung `xsl:apply-templates` ist gewissermaßen als Generalanweisung zu verstehen, sämtliche im Stylesheet vorhandenen Templates auszuführen.

Innerhalb eines Templates können zwei Varianten vorkommen, zum einen XSLT-Elemente mit Anweisungen für den Prozessor, erkennbar an dem Präfix `xsl`, und zum anderen sogenannte literale Ergebniselemente, die der Prozessor unverändert in die Ausgabedatei übernimmt. Dies sind beispielsweise sämtliche RDF-Tags wie `rdf:Description` oder die Property-Tags, die einen wichtigen Teil der Ausgabedatei ausmachen.

Die Knoten in der Quelldatei können über XPath-Ausdrücke adressiert werden. Das vorliegende Stylesheet nutzt zumeist den absoluten Pfad. Dies erschien bei der Bearbeitung weniger fehleranfällig, zumal der Editor den XPath aus der Quelldatei per Mausklick kopiert. Ferner ist der Code auf diese Weise für eventuelle Nachnutzungen besser nachvollziehbar.

Neben den bisher erwähnten XSL-Elementen kommen in dem über 600 Zeilen umfassenden XSLT-Stylesheet dieser Arbeit nur sehr wenige weitere Elemente zur Anwendung. Dies kann als Indiz für die leichte Anwendbarkeit und Effektivität von XSLT im gegebenen Kontext und bei ähnlich gelagerten Problemstellungen gewertet werden. „Aus der beeindruckenden Vielfalt fantastischer Möglichkeiten“ (Grupe 2018, S. 370) von XSLT und XPath müssen insgesamt nur gut ein Duzend Elemente bekannt sein, um die Konvertierung zu realisieren:

Mit Hilfe von `xsl:value-of` mit `select`-Attribut können die Werte einzelner Elemente oder Attribute der Quelldatei für die Zieldatei selektiert werden. Es folgt eine Variable oder ein XPath. Mit `xsl:for-each` und `select`-Attribut können sämtliche Vorkommen eines Feldes beachtet werden. Das `xsl:if`-Element mit `test`-Attribut legt die Bedingungen fest, unter denen ein Triple zu bauen ist. Im einfachsten Fall besteht die Bedingung in dem Vorhandensein eines bestimmten LIDO-Felds, es können jedoch auch bestimmte Inhalte in den Feldern oder Attributen getestet werden. `xsl:choose` in Verbindung mit `xsl:when` und `xsl:otherwise` erlaubt eine mehrstufige Fallunterscheidung. `xsl:element` und

`xsl:attribute` ermöglichen die flexible Generierung von eigenen Element- und Attributnamen. Schließlich können mit `xsl:function` eigene Funktionen geschrieben werden, die Parameter `xsl:param` übergeben. Diese beiden letzten Elemente kommen mit dem `name`-Attribut zum Einsatz.

Im Folgenden wird der Gebrauch einzelner Elemente anhand ausgewählter Passagen aus dem Stylesheet veranschaulicht.

Abbildung 29 zeigt die Extraktion des Item-URIs, also des URIs, der das Sammlungsobjekt identifiziert. Es wird die Variable `$var_handle-uri` definiert und mit dem Inhalt des Feldes `lido:objectPublishedID` gefüllt, falls im Text `hdl.handle` enthalten ist. Die Bedingung wird durch einen XPath-Ausdruck mit der Funktion `contains()` und dem XPath-Knotentyp `text()` formuliert. Der Aufruf der Variable, die das Subjekt der Triples darstellt, erfolgt in diesem Fall im `rdf:about`-Attribut.

```
<!-- VARIABLES -->
<!-- handle-URI; every record contains a handle-URI-->
<xsl:variable name="var_handle-uri" select="lido:objectPublishedID[contains(text(), 'hdl.handle')]/text()" />

<rdf:RDF>
  <!-- handle-URI as rdf-subject -->
  <rdf:Description rdf:about="{ $var_handle-uri }">
    <xsl:apply-templates />
  </rdf:Description>
```

Abbildung 29: XSLT-Stylesheet, handle-URI

Das Beispiel in Abbildung 30 baut ein Triple mit dem Subjekt aus Abbildung 29 und dem Objektgattungs-Objekt. Das Feld, aus dem das Objekt zu extrahieren ist, wird mit dem `match`-Attribute angesprochen, der Text wird in die Variable `$var_objektgattung` geschrieben und mit der Property `rdf:type` mit dem Subjekt verknüpft. Sollten fälschlicherweise Leerstellen in dem Objekt vorkommen, werden diese zuvor mit der XPath-`replace`-Funktion gelöscht. Anstelle der `replace`-Funktion empfiehlt sich ebenfalls die Funktion `normalize-space()`, die Leerstellen zu Beginn und am Ende löscht sowie mehrere Leerstellen durch eine ersetzt.

```
<xsl:template match="lido:descriptiveMetadata[1]/lido:objectClassificationWrap[1]/lido:objectWorkTypeWrap[1]/lido:objectWorkType[1]/lido:conceptID[1]">
  <!-- A falsely catalogued space is deleted by the replace-function -->
  <xsl:variable name="var_objektgattung" select="."/text()" />
  <rdf:type rdf:resource="{replace($var_objektgattung, ' ', '')}" />
</xsl:template>
```

Abbildung 30: XSLT-Stylesheet, Objektgattung

Der Aufbau der Triples, die die Titel des Sammlungsobjekts enthalten, ist in Abbildung 31 notiert. Es wird ein Blank Node vom Type `ido:Appellation` ausgegeben und für jedes `lido:titleSet` wird nach dem Attribut `lido:sortorder='1'` gesucht. Ist dieses vorhanden, wird mit

der Property `skos:preflabel` und dem language-Attribut ein Triple mit Literal gebaut. Andernfalls wird die Property `skos:altLabel` verwendet.

```
<!-- Item's title. Every title contains a sortorder=1-attribute.
      Titleset and appellationValue can be repeated -->
<xsl:if test="@lido:titleWrap">
  <ido:hasAppellation>
    <rdf:Description>
      <rdf:type rdf:resource="http://purl.org/asch/ido/180704#Appellation" />
      <xsl:for-each select="lido:titleWrap/lido:titleSet">
        <xsl:choose>
          <xsl:when test="@lido:sortorder = '1' ">
            <xsl:for-each select="lido:appellationValue">
              <skos:prefLabel xml:lang="{lido:language(.)}">
                <xsl:value-of select="./text()" />
              </skos:prefLabel>
            </xsl:for-each>
          </xsl:when>
          <xsl:otherwise>
            <xsl:for-each select="lido:appellationValue">
              <skos:altLabel xml:lang="{lido:language(.)}">
                <xsl:value-of select="./text()" />
              </skos:altLabel>
            </xsl:for-each>
          </xsl:otherwise>
        </xsl:choose>
      </xsl:for-each>
    </rdf:Description>
  </ido:hasAppellation>
</xsl:if>
```

Abbildung 31: XSLT-Stylesheet, Titel

In diesem Beispiel kommt auch die einzige eigene Funktion des Stylesheets zum Einsatz. *lido:language()* ist in Abbildung 32 abgebildet.

```
<!-- Function which sets the language-attribute @xml:lang. It is set 'de', if no language is given,
      otherwise the value is displayed -->
<xsl:function name="lido:language">
  <xsl:param name="node" />
  <xsl:value-of select="if ( not($node/@xml:lang) ) then ('de') else ( $node/@xml:lang )" />
</xsl:function>
```

Abbildung 32: XSLT-Stylesheet, Funktion *lido:language()*

Die Funktion *lido:language()* mit dem Parameter *node* vergibt den Sprachcode *de*, falls in dem betreffenden Feld kein Code vorhanden ist. Ansonsten wird der Sprachcode aus der Quelldatei übernommen. Die Definition einer anwendungsspezifischen Funktion ist hier sinnvoll, weil sich die Frage nach dem Sprachcode stetig wiederholt.

Abbildung 33 zeigt den Ausschnitt aus dem Stylesheet, der die Informationen zu den mit dem Event verknüpften Orten aus den *lido:eventSets* in RDF-Triples umwandelt. Das Subjekt, dessen Generierung nicht in der Abbildung enthalten ist, ist der Blank Node, der die Item-Event-Verknüpfung identifiziert. Für jeden *lido:eventPlace* wird nach einem URI im Element *lido:placeID* gesucht. Ist dieser vorhanden, wird mit der Prädikat `pdo:tookPlaceAt` ein Triple

erzeugt. Fehlt der URI, wird bei Vorhandensein von *lido:gml* ebenfalls ein Triple mit *pdo:tookPlaceAt* gebildet, das als Objekt einen Blank Node der Klasse *pdo:Place* erhält. Mit diesem Blank Node als Subjekt werden anschließend zwei weitere Triples erzeugt, die den Breiten- und Längengrad des Ortes enthalten. Beide Werte sind, getrennt durch eine Leerstelle, in dem Element *gml:pos* enthalten. Der Breitengrad ist der Wert vor der Leerstelle, der Längengrad der nach der Leerstelle. Mit den Funktionen *substring-before()* und *substring-after()* lassen sich die als Glekommazahlen (float) definierten Literale extrahieren.

```
<!-- Event place -->
<!-- 0-n eventPlaces are possible -->
<xsl:for-each select="lido:event/lido:eventPlace">
  <!-- If there is a place-URI, gml:pos is not considered -->
  <xsl:choose>
    <xsl:when test="lido:place/lido:placeID[@lido:type =
      'http://terminology.lido-schema.org/identifier_type/uri']">
      <pdo:tookPlaceAt rdf:resource="{lido:place/lido:placeID}" />
    </xsl:when>
    <xsl:otherwise>
      <xsl:if test="lido:place/lido:gml">
        <pdo:tookPlaceAt>
          <!-- Creating a blank note -->
          <rdf:Description>
            <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Place" />
            <wgs84_pos:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
              <xsl:value-of select="substring-before(
                lido:place/lido:gml/gml:Point/gml:pos/text(), ' ')" />
            </wgs84_pos:lat>
            <wgs84_pos:long rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
              <xsl:value-of select="substring-after(
                lido:place/lido:gml/gml:Point/gml:pos/text(), ' ')" />
            </wgs84_pos:long>
          </rdf:Description>
        </pdo:tookPlaceAt>
      </xsl:if>
    </xsl:otherwise>
  </xsl:choose>
</xsl:for-each>
```

Abbildung 33: XSLT-Stylesheet, Orte des Events

Ein wichtiger Aspekt bei der Generierung der RDF-Daten ist schließlich die Vermeidung von redundanten Triples. Dies soll an zwei Beispielen veranschaulicht werden. Die Quelldaten enthalten bei Vorhandensein eines MARC Relator Terms für die Rolle eines Akteurs außerdem einen entsprechenden GND-URI. Dieser ansonsten sehr wertvolle URI wird in dem Fall nicht berücksichtigt, weil der MARC Relator Term als wichtiger eingeschätzt wird und nicht zweimal dasselbe ausgedrückt werden soll. Ein ähnlicher Fall liegt bei der Materialart im eventWrap vor. Hier erhält der AAT-URI den Vorrang vor dem GND-URI gleichen Inhalts. Sind beide nicht vorhanden, wird das Literal verwendet. Zur Abbildung dieser Rangfolge bei der Materialart kommen im Code verschachtelte *xsl:choose*-Elemente zum Einsatz.

Zusammenfassend lässt sich am Ende dieses Abschnitts konstatieren, dass das Stylesheet mit einer überschaubaren Menge an Elementen die zum Teil komplexen Mapping-Vorgaben

umsetzen kann. Die Schwierigkeit ergibt sich zumeist aus der Datenhaltung der Quelldatei. Dort die Muster zu finden und angesichts der tiefen Verschachtelung der LIDO-XML-Elemente die Übersicht zu behalten, ist die eigentliche Herausforderung. XSLT in Zusammenspiel mit dem Editor Oxygen kann hingegen eine leichte Erlernbarkeit und Handhabung bescheinigt werden. Die Konvertierung der eigenen Daten in Linked Data erscheint auf diese Weise auch für Neueinsteigerinnen und Neueinsteiger möglich.

6.2.3 BATCH-VERARBEITUNG

Nach der Entwicklung des XSLT-Stylesheets anhand einer LIDO-Sample-Datei, die schrittweise um denkbare Elemente zu ergänzen ist, bzw. aus der zu Testzwecken Elemente zu löschen sind, muss die Konvertierung auf sämtliche Quelldateien angewendet werden.

Dazu wird Folgendes benötigt:

- Verzeichnis mit den 21.071 LIDO-Quelldateien
- XSLT-Stylesheet zur Konvertierung
- XSLT-Prozessor
- Validierungsprogramm für die Ausgabedateien (optional)
- Ausgabeverzeichnis für die 21.071 RDF/XML-Dateien

Als XSLT-Prozessor wurde *Saxon*¹⁶⁶ für Java ausgewählt.

Das Validierungsprogramm *xmllint* stammt aus der *libxml2*-Softwaresammlung¹⁶⁷ und validiert die Ausgabedokumente gegen das Relax NG Schema für RDF/XML 1.1¹⁶⁸.

Für die Konvertierung des gesamten Dateien-Stapels wurde das Shell-Skript in Abbildung 34 geschrieben.

Es verlangt folgende drei Argumente: das XSLT-Stylesheet, das Verzeichnis mit den Quelldateien (`input_dir`) und das Verzeichnis, das die Ausgabedateien aufnimmt (`output_dir`).

Für jede mit *.xml* endende Datei im Eingangsordner (`$input_dir`) wird das JAR-File *saxon9he.jar* (XSLT-Prozessor) ausgeführt, wobei jeweils eine LIDO-Datei (*-s*) und das Stylesheet (*-xsl*) geladen und nach der Datenkonvertierung im Zielverzeichnis (`$output`) unter dem alten Namen (ohne Endung) ergänzt um *.rdf.xml* abgelegt werden. Anschließend wird diese neue Datei gegen das Relax NG Schema für RDF/XML (*rdfxml.rng*) validiert.

¹⁶⁶ Download von Saxon XSLT und XQuery Processor: <https://sourceforge.net/projects/saxon/> (15.01.2019).

¹⁶⁷ <http://xmlsoft.org/> (15.01.2019).

¹⁶⁸ <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/rdfxml.rng> (15.01.2019).


```
#!/bin/bash

# variables
xslt=$1
input_dir=$2
output_dir=$3
script_dir=`dirname "${BASH_SOURCE[0]}"`

# check if xslt-file is provided else exit
if [ ! -f "$xslt" ]; then
    echo "you need to enter a xslt-script as 1st argument!"
    exit 1
fi

# check if input_dir is set else exit
if [ ! -d "$input_dir" ]; then
    echo "you need to enter an existing input dir as 2nd argument!"
    exit 1
fi

# check if output_dir is set else exit
if [ ! $output_dir ]; then
    echo "you need to enter an output dir as 3rd argument!"
    exit 1
fi

# make output_dir if not already existing
if [ ! -d "$output_dir" ]; then
    mkdir $output_dir
fi

##### do the conversion #####
for file in $input_dir/*.xml; do
    [ -e "$file" ] || continue
    filename=$(basename -- "$file")

    # LIDO/XML to RDF/XML
    name="${filename%.*}.rdf.xml"
    java -jar $script_dir/saxon/saxon9he.jar -s:$file -xsl:$xslt > $output_dir/$name

    # validate generated RDF/XML
    xmllint --noout --relaxng $script_dir/rdfxml.rng $output_dir/$name || exit 1
done
```

Abbildung 34: Skript batch_lido2rdfxml.sh

Zur besseren manuellen Überprüfung der Mapping-Ergebnisse wurden die XML/RDF-Dateien zusätzlich in das Serialisierungsformat Turtle überführt. Diesem Zweck dient ein weiteres Skript (*batch_rdfxml2ttl.sh*), das in Ausschnitten in Abbildung 35 dargestellt ist. Die Schleife über alle Dateien entspricht der im obigen Skript. Da dieses Programm kein XSLT-Stylesheet benötigt, kommt es mit zwei Argumenten aus (*input_dir* und *output_dir*). An die Stelle des Saxon-Prozessors tritt ein Open Source Programm zur Änderung des Serialisierungsformats: Die *RDF-Input-Output-Technologie (RIOT)* des Java Frameworks *Apache Jena*.¹⁶⁹ Über das Argument *output* wird das Ziel-Serialisierungsformat, in diesem Fall Turtle, übermittelt.

¹⁶⁹ Download: <https://jena.apache.org/download/index.cgi> (10.01.2019).

Dokumentation: <https://jena.apache.org/documentation/io/> (19.01.2019).

```
##### do the conversion!! #####

for file in $input_dir/*.xml; do
    [ -e "$file" ] || continue
    filename=$(basename -- "$file")

    # RDF-XML to Turtle
    name="{filename%.*}.ttl"
    $script_dir/apache-jena-3.8.0/bin/riot --output=ttl $file > $output_dir/$name
done
```

Abbildung 35: Ausschnitt aus Skript batch_rdfxml2ttl.sh

Wichtig ist, dass die Skripte in demselben Verzeichnis liegen, in das Saxon wie auch die Apache-Jena-Distribution geladen wurden.

6.3 ERGEBNISSE UND TESTING

Als Ergebnis des Mappings liegen die 21.071 LIDO-Datensätze zusätzlich in den beiden RDF-Formaten RDF/XML und Turtle vor. Bezogen auf den Umfang rangiert das LIDO-Verzeichnis mit 497 MB weit vor den RDF-Verzeichnissen RDF/XML (146 MB) und Turtle (108 MB). Obwohl die transportierten Inhalte in den beiden RDF-Verzeichnissen die gleichen sind, ist die Summe der Turtle-Dateien kleiner, was sich auf das Fehlen der XML-Tags zurückführen lässt. Die Größe des LIDO-Verzeichnisses resultiert ebenfalls vor allem aus dem größeren Overhead der XML-Dateien.

Der überwiegende Teil der Inhalte der LIDO-Dateien konnte gleichzeitig in RDF Statements überführt werden. Soweit in den LIDO-Daten vorhanden, werden folgende Informationen über die Triples transportiert:

Informationen zum Sammlungsobjekt/Item

- Typ `pdo:Item`
- Bezeichnung/Name
- Klassifikationen
- Schlagwörter
- Signatur/Zugangsnummer
- Beschreibung
- Erscheinungsbild
 - Inschriften
 - Maßangaben
- Verknüpfte Werke
- Sammlungszugehörigkeit
- Referenzen/Literaturangaben im GVK
- Event

- Typ (z. B. Creation, Beginning of Existence, Custody)
 - Akteur
 - Datum
 - Ort
 - Ethnie
 - Material, Technik, Farbe
- Digitale Repräsentationen

Informationen zum Datensatz/Record

- Verknüpfung zum Item
- Typ `pdo:Record`
- Rechteinhaber des Datensatzes
- Lizenz für die Verwertung des Datensatzes
- Datum der Erstellung des Datensatzes

Informationen zu den Digitalen Repräsentationen

- Typ `ido:DigitalRepresentation`
- Rechteinhaber der digitalen Repräsentation
- Urheber der digitalen Repräsentation
- Lizenz für die Verwertung der digitalen Repräsentation
- Datum der Erstellung der digitalen Repräsentation

Durchschnittlich wurden pro Datensatz etwa 60 Triples erstellt, was einer Gesamtzahl von 1.183.447 Triples entspricht.

Es sind Verknüpfungen zu zahlreichen externen Quellen via HTTP-URI enthalten, aus denen zusätzliches Wissen generiert werden kann. Als Beispiele seien die Links zur Gemeinsamen Normdatei (55.000 Triples), dem Art and Architecture Thesaurus (9400 Triples), der Zeitschriftendatenbank (67.400 Triples), GeoNames (8800 Triples), zur LIDO-Terminology (41.300 Triples), Iconclass (1500 Triples) und schließlich zur Oberbegriffsdatei (7000 Triples) genannt. Außerdem verweisen mehr als 10.000 Triples auf Literaturangaben im Gemeinsamen Verbundkatalog (GVK).¹⁷⁰

Hinsichtlich der automatischen Validierung der Quell- und Ausgabedateien lassen sich folgende Aussagen treffen:

- Die LIDO-Quelldateien validieren gegen das LIDO-Schema¹⁷¹ und sind somit hinsichtlich ihrer Struktur einheitlich, was die Ausgabedateien vorhersehbarer macht.

¹⁷⁰ Die Zahlen wurden durch SPARQL-Anfragen ermittelt. Siehe unten Kap. 7.2.

¹⁷¹ <http://www.lido-schema.org/schema/v1.0/lido-v1.0.xsd> (15.01.2019).

- Das XSLT-Stylesheet validiert gegen das XSL-Schema¹⁷².
- Die RDF-Ausgabedateien werden gegen ein basales Relax NG Schema für RDF/XML¹⁷³ validiert.
- Bei der Konvertierung der Ausgabedaten zu Turtle sowie bei ihrer Einspielung in den Triplestore wird das RDF-Format mit den Apache Jena-Werkzeugen¹⁷⁴ validiert.
- Es fehlt ein Schema für das ASCH-Modell, das, basierend auf den ASCH Application Profiles und den Vorgaben in den eingebundenen Ontologien, überprüft, ob die benutzten Klassen und Properties gemäß den Vorgaben eingesetzt und ob *Domain*- und *Range*-Vorgaben eingehalten werden. Die Erstellung dieses Schemas übersteigt den Rahmen dieser Arbeit.¹⁷⁵ Stattdessen erfolgte eine stichprobenartige manuelle Überprüfung der Ergebnisse.

Es wurde ein Testpool mit 20 Turtle-Files, bestehend aus je fünf Dateien aus *kuniweb*, *naniweb*, *kenom* und den heterogenen Quellen, erstellt und die enthaltenen Daten unter Berücksichtigung folgender Kriterien sorgfältig geprüft.

- Sind die Subjekte richtig verwendet?
- Stehen die Blank Nodes in den richtigen Zusammenhängen und haben sie die korrekten Typ-Zuweisungen?
- Werden die URIs mit ihren Präfixen korrekt dargestellt?
- Stehen Literale in Anführungsstrichen und nicht abgekürzte URIs in Winkelklammern?
- Sind die Properties gültig und ohne Tippfehler?
- Gibt es leere Literal-Objekte?
- Fallen Abweichungen zu den ASCH-APs auf?

Im Rahmen dieses Testings erfolgte außerdem ein *Parsing* der Zieldateien beim *Validation Service*¹⁷⁶ des W3C . Vor allem die dort auswählbare Graph-Ansicht bietet eine zusätzliche Möglichkeit der Fehlerkontrolle.¹⁷⁷

Zusammenfassend kann den RDF-Daten eine gute Qualität bescheinigt werden.

¹⁷² <http://www.w3.org/2007/schema-for-xslt20.xsd> (15.01.2019).

¹⁷³ <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/rdfxml.rng> (15.01.2019).

¹⁷⁴ Die Tiefe der Validierung ist jedoch in den Dokumentationen von Apache Jena nicht spezifiziert. Syntaktische Fehler sowie beispielsweise Leerstellen in URIs werden angezeigt.

¹⁷⁵ Derartige Schema-Dateien konnten bei anderen Linked-Data-Anwendungen wie zum Beispiel beim Linked-Data-Service der DNB, der Bayerischen Staatsbibliothek, des HBZ oder der Universität Mannheim ebenfalls nicht ausgemacht werden.

¹⁷⁶ <https://www.w3.org/RDF/Validator/> (15.01.2019).

¹⁷⁷ Ein Beispiel für die Graph-Darstellung des Validation Services ist in Anhang 6 enthalten.

7 BEREITSTELLUNG DER DATEN

Duh — it's a graph. Relationships are first-class citizens and the links between data add huge value, as well as flexibility (Stichbury 2017).

Eine wichtige Komponente im *Linked Data Life Cycle* ist die globale Bereitstellung der Daten (Al-Mutawa und Al-Aama 2017, S. 34). Dieses Kapitel beschreibt die Veröffentlichung der Daten in einem *Fuseki*¹⁷⁸ Triplestore und führt in die SPARQL-Suche ein.

7.1 TRIPLESTORE

Datenbankmanagementsysteme, die Triples speichern und verarbeiten können, werden als RDF Stores oder Triplestores bezeichnet (solid IT 2019b).

Der *Fuseki* Triplestore ist ein Projekt von Apache Jena, dem Open Source Framework für semantische Netze, dessen RIOT-Technologie bereits bei der Umwandlung der RDF/XML-Triples in das Turtle-Serialisierungsformat zum Einsatz kam.¹⁷⁹ *Fuseki* kann als Java-Webanwendung (WAR-File) oder als eigenständiger Server ausgeführt werden (Apache Jena Fuseki 2018).

Im Rahmen dieser Arbeit wurde der Triplestore zunächst auf einem lokalen Server aufgesetzt.¹⁸⁰ Apache Jena stellt die Out-of-the-Box-Lösung auf ihrer Website zum Download¹⁸¹ zur Verfügung. Nach der Installation kann der *Fuseki* Server unter Windows durch Ausführen der Datei *fuseki-server.bat* gestartet und daraufhin in jedem Browser unter *localhost* mit Port 3030 aufgerufen werden. Neben der einfachen Installation sprechen für *Fuseki* ein komfortables Webinterface und gute Upload-¹⁸² und Antwortzeiten.

Zum Laden der Triples ist zunächst ein Datenset einzurichten, das als persistent zu definieren ist. Dies kann über die Kommandozeile oder über das *Fuseki* Webinterface erfolgen. Im Webinterface ist dazu unter *manage datasets* die Auswahl *update data* auszuwählen. Abbildung 36 zeigt die Rückmeldung des Servers bei erfolgreichem Upload von zwei Datensätzen mit 115 und 113 Triples.

¹⁷⁸ <https://jena.apache.org/documentation/fuseki2/> (21.01.2019).

¹⁷⁹ Siehe oben Kap. 6.2.3.

¹⁸⁰ Eine Anleitung findet sich bei Panchal 2017.

¹⁸¹ <https://jena.apache.org/download/> (15.01.2019).

¹⁸² Sämtliche RDF Statements des Göttinger Sammlungsportals ließen sich in weniger als einer Stunde in den Triplestore laden.

Abbildung 36: Fuseki Triplestore, Upload von zwei Datensätzen

Neben den konvertierten Daten des Sammlungsportals wurden die RDF Statements der ASCH-eigenen Ontologien IDO und PDO ebenfalls in den Triplestore eingespielt.

Unter der Registerkarte *info* zeigt der Triplestore statistische Informationen an, beispielsweise zu erfolgreichen Suchen. Außerdem ist die Gesamtzahl der Triples in dem jeweiligen Datenset abrufbar.

Abbildung 37: Fuseki Triplestore, Info-Schirm

Des Weiteren stellt *Fuseki* einen SPARQL-Endpoint für die semantische Suche zur Verfügung, dargestellt in Abbildung 38. Im mittleren Bereich des *Query*-Schirms ist die Suche unter Angabe der verwendeten URI-Präfixe zu formulieren. Oberhalb des SPARQL-Endpoint-URL des Datensets können weitere Präfixe definiert und auswählbar gemacht werden, wobei *prefix.cc* als Dienst integriert ist. Die Ergebnis-Triples werden unterhalb des Suchfelds aufgelistet. Dass das Prädikat `rdf:type` in dem Beispielergebnis unten nicht als vollständiger URI angezeigt wird, liegt in der Definition des URI-Präfixes in der Anfrage begründet.

The screenshot shows the Fuseki SPARQL query interface. At the top, there's a section for 'SPARQL query' with a text area for entering queries. Below this are 'EXAMPLE QUERIES' and 'PREFIXES' sections. The 'PREFIXES' section shows a list of prefixes: `rdf`, `rdfs`, `owl`, `xsd`, and a button to add more. The 'SPARQL ENDPOINT' section contains the URL `http://localhost:3030/sammlungsportal_goettingen/query`. The 'CONTENT TYPE (SELECT)' dropdown is set to 'JSON', and the 'CONTENT TYPE (GRAPH)' dropdown is set to 'Turtle'. The main query area contains the following SPARQL query:

```

1
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX owl: <http://www.w3.org/2002/07/owl#>
5 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
6 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
7
8
9 SELECT ?subject
10 WHERE {
11   ?subject ?predicate ?object
12 }
13 LIMIT 25

```

Below the query is the 'QUERY RESULTS' section, which shows a table of results. The table has four columns: 'subject', 'predicate', and 'object'. The results are as follows:

	subject	predicate	object
1	<code><https://hdl.handle.net/21.11107/3db5d0ba-b0d0-4a62-96b2-223b707f77bb></code>	<code>rdf:type</code>	<code><http://portal.wissenschaftliche-sammlungen.de/Object/Genre/224></code>
2	<code><https://hdl.handle.net/21.11107/3db5d0ba-b0d0-4a62-96b2-223b707f77bb></code>	<code>rdf:type</code>	<code><http://purl.org/asch/pdo/180704#item></code>

Abbildung 38: Fuseki SPARQL-Endpoint

Nach erfolgreicher Testphase auf dem lokalen Rechner wurde der *Fuseki* Triplestore mit seinen fast 1,2 Millionen Triples auf einem Webserver unter <https://fuseki.qbv.de/>¹⁸³ global zugänglich gemacht.¹⁸⁴

¹⁸³ Dabei handelt es sich um die Administratoren-Version, die mit Login erreichbar ist (Benutzername: admin; Passwort: verbundzentrale5000). Eine Version ohne Login soll laut VZG nach Implementierung des Rechtemanagements im zweiten Quartal 2019 zur Verfügung stehen (Dührkohp 2018).

¹⁸⁴ Die Installation konnte anhand folgender Anleitungen durchgeführt werden: <https://github.com/NatLibFi/Skosmos/wiki/InstallTutorial> (15.01.2019) und <https://gist.github.com/nichtich/113ecc2831d01eb4ae91508b88c0217c> (15.01.2019).

Mit der Online-Stellung der Daten sind weiterführende verteilte Tests sowie die Verknüpfung mit anderen SPARQL-Endpoints möglich. Der nächste Abschnitt liefert eine Einführung in SPARQL und präsentiert beispielhafte Anfragen innerhalb der neuen Datenbank sowie erste Experimente, die eigenen Daten mit externen Datensets zu verknüpfen.

7.2 SPARQL

Mit der *SPARQL Protocol and RDF Query Language (SPARQL)* können Triplestores durchsucht werden. „SPARQL is to Semantic Triple Stores what XQuery is to XML, JavaScript [...] is to JSON stores and SQL is to relational data” (Cagle 2013).

Die SPARQL-Suchkriterien werden in Form von *Triple Patterns* oder *Graph Patterns*, angegeben, wobei Variablen anstatt eines konkreten Subjekts, Prädikats oder Objekts benutzt werden können. Glimm (2011, S. 138) spricht in diesem Zusammenhang von *Subgraph Matching*, wobei die *Graph Patterns*, gegebenenfalls mit Variablen als *Wildcards*, mit den Triples des abgefragten Graph abgeglichen und die Übereinstimmungen ausgegeben werden.

Anhand von Beispielanfragen in den im Rahmen dieser Arbeit generierten Daten sollen nachfolgend die grundsätzlichen Bestandteile der Sprache skizziert werden.¹⁸⁵

Abbildung 39 enthält eine einfache SPARQL-Suche nach den ersten 25 englischsprachigen Bezeichnungen (prefLabel) und zeigt im unteren Teil die ersten drei Ergebnisse an.

Nach Angabe der Namensräume für die Abkürzungen der URIs beginnt die Suche mit der *Select-Klausel*. Wie bei SQL wird anschließend angegeben, welche Felder zurückgegeben werden sollen; in diesem Fall das Item und sein Name. Variablen sind durch ein Fragezeichen gekennzeichnet. Die *Where-Klausel* enthält die Bedingungen in Form von *Triple Patterns*. In diesem Beispiel sind die Prädikate feste Werte.

Mit Hilfe des booleschen FILTER-Ausdrucks ist eine weitere Eingrenzung möglich. Der SPARQL-Filter ist mit einer Reihe von Operatoren nutzbar. In dem Beispiel unten kommt er mit dem Language-Operator zum Einsatz. Als weitere Beispiele für Operatoren seien *isUri*, *isBlank*, *isLiteral*, *Str*, *Datatype* und *Regex* genannt.

Die Ergebnisanzeige kann mit Suchmodifikatoren wie LIMIT, OFFSET oder ORDER BY beeinflusst werden. In dem Beispiel in Abbildung 39 wird die Ergebnisanzeige auf 25 Triples limitiert.

¹⁸⁵ Einen Einstieg in SPARQL ermöglichen die Cheat Sheets von Feigenbaum 2008 und Beckett 2006. Weiterführende Informationen finden sich bei Prud'hommeaux und Seaborne 2008 sowie Harris und Seaborne 2013.

The screenshot shows a SPARQL query interface. The query is as follows:

```

1 PREFIX ido: <http://purl.org/asch/ido/180704#>
2 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3
4 # Return max. 25 English preLabels
5
6 SELECT ?item ?name
7 WHERE {
8   ?item ido:hasAppellation ?object.
9   ?object skos:prefLabel ?name.
10  FILTER (lang(?name) = 'en')
11 }
12 limit 25

```

Below the query, the interface shows "QUERY RESULTS" with buttons for "Table" (selected), "Raw Response", and a download icon. It indicates "Showing 1 to 25 of 25 entries". The results are displayed in a table with two columns: "item" and "name".

	item	name
1	<https://hdl.handle.net/21.11107/fedd34b-eca6-4c9c-835a-5d2c52b5e344>	"Fertility-Doll"@en
2	<https://hdl.handle.net/21.11107/6f24a1d1-0831-4e93-bd5a-92e07e8043be>	"Black-Venus"@en
3	<https://hdl.handle.net/21.11107/69209ded-e105-4414-8516-3558d3d50ed5>	"Fertility Symbol"@en

Abbildung 39: SPARQL-Select nach den Bezeichnungen der Sammlungsobjekte

In der nächsten Abbildung wird nach sämtlichen Sammlungsobjekten mit ihren Event-Typen und eventuellen Start- und Endedaten gesucht. Zusätzlich wird an diesem Beispiel die Verknüpfung von SPARQL *Graph Patterns* mit dem OPTIONAL-Konstruktor deutlich. Als Folge gibt es Ergebnisse mit und ohne Datum. Weitere Konstruktoren sind UNION und MINUS. Außerdem findet sich innerhalb des OPTIONAL-Konstruktors ein Beispiel für zwei Property Paths, mit denen in diesem Fall die Blank Nodes übersprungen werden.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX pdo: <http://purl.org/asch/pdo/180704#>
PREFIX ido: <http://purl.org/asch/ido/180704#>
# Return all events, that affected all objects in the database
# Display start and end year if present
SELECT ?subject ?eventType ?start ?end
WHERE {
  ?subject ido:wasAffectedBy ?event.
  ?event rdf:type ?eventType.
  OPTIONAL {
    ?event pdo:start/rdf:value ?start;
           pdo:end/rdf:value ?end.
  }
}

```

Abbildung 40: SPARQL-Select nach Event-Typen

Neben *Select* stehen für die Suche die Klauseln *Describe* und *Ask* und zur Bearbeitung der Daten *Construct* zur Verfügung. *Ask* gibt *true* oder *false* zurück. Ein Beispiel für eine *Describe*-Anfrage und ihr Ergebnis liefert Abbildung 41. Es werden sämtliche Informationen abgefragt, die zu einem bestimmten Datensatz vorhanden sind.

```
Describe ?record
WHERE {
  ?record foaf:primaryTopic <https://hdl.handle.net/21.11107/c0c109d0-6830-4f12-a95e-5363028d11b4>.
}

<http://sammlungen.uni-goettingen.de/objekt/record_kuniweb_1109801>
  a pdc:Record ;
  dc:rights <https://creativecommons.org/publicdomain/zero/1.0/deed.de> ;
  dcterms:created "2017-11-29T06:44:52+01"^^dcterms:W3CDTF ;
  dcterms:rightsHolder <http://ld.zdb-services.de/resource/organisations/DE-MUS-304718> ;
  foaf:primaryTopic <https://hdl.handle.net/21.11107/c0c109d0-6830-4f12-a95e-5363028d11b4> .
```

Abbildung 41: SPARQL-Describe-Anfrage nach einem Record mit Ergebnisanzeige

Das Beispiel in Abbildung 42 präsentiert ein erstes Experiment zur Verknüpfung mit einer externen Quelle. In diesem Fall wird mit dem SERVICE-Operator der SPARQL-Endpoint von *DBpedia* eingebunden. *DBpedia* verknüpft mit `owl:sameAs` zu den GND-URIs. Diese sind auch im Göttinger Knowledge Graph enthalten, zum Beispiel als *EventActors*, und können somit als Anker dienen. Daraufhin sind sämtliche in *DBpedia* enthaltene Informationen zu den *EventActors* mit GND-URI auch für die Göttinger Daten verfügbar. In dem vorliegenden Beispiel werden Geburtsdaten und Nationalitäten der Akteure abgefragt und ausgegeben.¹⁸⁶

```
10 SELECT ?subject ?person ?birthdate ?nationality
11 WHERE {
12   ?subject ido:wasAffectedBy/marcel:cre ?actor.
13
14   SERVICE <http://dbpedia.org/sparql>
15   {
16     ?person owl:sameAs ?actor.
17     OPTIONAL{
18       ?person dbo:birthDate ?birthdate;
19       dbo:nationality ?nationality.}
20   }
21 }
22 order by desc (?birthdate)
23
```

QUERY RESULTS

Table Raw Response

Showing 1 to 71 of 71 entries Search: Show 1000 entries

	subject	person	birthdate	nationality
1	<https://hdl.handle.net/21.11107/c0c109d0-6830-4f12-a95e-5363028d11b4>	<http://dbpedia.org/resource/Gustav_Heinrich_Johann_Apollon_Tammann>	"1861-5-28"^^xsd:date	<http://dbpedia.org/resource/Baltic-German>
2	<https://hdl.handle.net/21.11107/record_DE-MUS-069123_254>	<http://dbpedia.org/resource/Heinrich_Burkhardt>	"1861-10-15"^^xsd:date	<http://dbpedia.org/resource/Germany>

Abbildung 42: SPARQL-Select mit Verknüpfung zur DBpedia

Die Suche bleibt also nicht auf das eigene Datenset beschränkt, sondern bei geeigneten URIs können Informationen aus anderen Quellen eingebunden werden.

¹⁸⁶ Vgl. zum Beispiel http://dbpedia.org/page/Gustav_Heinrich_Johann_Apollon_Tammann (15.01.2019).

8 FAZIT UND AUSBLICK

There is quite a lot that one can talk about for RDF, but there is no need to talk about it all in one place (Tauberer 2008).

Abschließend gilt es zu überprüfen, inwiefern die Ziele dieser Arbeit erreicht werden konnten. Außerdem werden auf den Ergebnissen aufbauende nächste Schritte skizziert.

8.1 FAZIT

Das vornehmliche Ziel dieser Arbeit bestand in der Generierung von Linked Open Data für das Göttinger Sammlungsportal. Die RDF-Daten konnten nicht nur generiert, sondern außerdem in einem *Fuseki* Triplestore bereitgestellt werden. Als Ergebnis liegt ein Knowledge Graph mit beinahe 1,2 Millionen Triples vor.

Die Qualität der Daten soll den *Linked Data Principles*¹⁸⁷ entsprechen.

Die ersten beiden *Linked Data Principles* (HTTP-URIs für jede *Resource*) sind mehrheitlich eingehalten. Sämtliche Sammlungsobjekte sind über HTTP-handle-URIs identifiziert, die über 303 Redirect auf den jeweiligen Datensatz mit eigenem URI verweisen.¹⁸⁸ Es sind also unterschiedliche URIs für das Sammlungsobjekt (Item) und seinen Datensatz (Record) vorhanden. Ebenso liegen für digitale Repräsentationen, Orte, Personen, Klassifikationen und Schlagwörter HTTP-URIs vor. Es fehlen jedoch URIs für die einzelnen Göttinger Sammlungen, denen die Objekte angehören können. Hier sind mit `ido:isGatheredInto` und `dc:identifier` nur Literale verknüpft.¹⁸⁹ Außerdem sei die große Anzahl an Blank Nodes angemerkt. Vorzugsweise hätte zum Beispiel jede Item-Event-Kombination einen eigenen URI und wäre damit auch von außen direkt ansprechbar.

Die Einhaltung der *Linked Data Principles* 3 (standardisierte Informationen zu den URIs) und 4 (Links zu anderen URIs) kann ebenfalls bescheinigt werden. Es liegen für die Eigenschaften und Klassen durchgängig maschinenlesbare Beschreibungen in einer Schema-Sprache vor, und zwar sowohl für nachgenutzte Standardontologien wie RDF oder DC als auch für die im Rahmen des ASCH-Modells entwickelten Ontologien PDO und IDO. Dasselbe gilt für die Viel-

¹⁸⁷ Siehe oben Kap. 2.1.

¹⁸⁸ Die URIs der Sammlungsobjekte enthalten die Zeichenfolge *hdl.handle.net*, die der Datensätze *sammlungen.uni-goettingen.de*.

¹⁸⁹ Laut Rühle (2018c) befindet sich ein *International Standard Link Identifier (ISLI)* gegenwärtig in der Diskussion zusammen mit der DDB und anderen Partnern.

zahl externer URIs wie die der GND oder der Datenbank GeoNames. Ein erstes Beispiel für die Entdeckung weiterer Informationen durch externe Links zeigt die SPARQL-Anfrage in Abbildung 42, die eine Verknüpfung zur *DBpedia* und damit den Zugriff auf sämtliche dort hinterlegten Informationen enthält. Ebenso können die Daten des Göttinger Sammlungsportals aus externen SPARQL-Endpoints heraus angesprochen und weiterverwendet werden.

Das Fehlen von maschinenlesbaren Informationen wurde im Bereich der Objektgattungen ausgemacht. Daraufhin konnte im Verlauf der Arbeit Kontakt mit der *Koordinierungsstelle für wissenschaftliche Sammlungen in Deutschland* aufgenommen und eine entsprechende Aufbereitung der Daten verabredet werden.¹⁹⁰

Außerdem können dem Göttinger Knowledge Graph fünf Sterne im Sinne von Berners-Lees *Star Rating System*¹⁹¹ für offene Daten bescheinigt werden. Die beschreibenden Daten zu den Objekten wie auch ihre digitalen Repräsentationen unterliegen einer *Creative-Commons-Lizenz* und sind als Linked Open Data über den SPARQL-Endpoint zugänglich. Die Daten sind strukturiert und maschinenlesbar und enthalten zahlreiche Verknüpfungen zu externen Daten.

Auch das in dieser Arbeit gewählte Vorgehen bei der Generierung von Linked Data hält den vom W3C postulierten *Best Practices for Publishing Linked Data* (Hyland et al. 2014) stand. Im Folgenden erfahren die zehn empfohlenen Kriterien (Steps) einen kurzen Abgleich.

Step #1 Prepare stakeholders: Es wurde zunächst ein Verständnis für die Daten, ihre Ziele und den musealen Kontext aufgebaut, sodass die für die Stakeholder wichtigen Daten in die Umwandlung genommen werden konnten. Außerdem basiert das ASCH-Modell auf einer Stakeholder-Befragung.

Step #2 Select a dataset: Der Wissenschaftsrat hat wissenschaftlichen Sammlungen eine hohe Forschungsrelevanz bescheinigt. Somit wurde ein Datenbestand ausgewählt, dessen Nachnutzung für andere wertvoll ist.

Step #3 Model the data: Das Mapping wurde auf Basis des ASCH-Modells erstellt.

Step #4 Specify an appropriate licence: Die Georg-August-Universität Göttingen hat die Daten unter CC-Lizenz freigegeben. Ferner sind Lizenzinformationen für die Daten eines jeden Sammlungsobjekts und seiner Bilder enthalten.

Step #5 Good URIs for linked data: Es liegen HTTP-URIs vor (vgl. Bewertung der *Linked Data Principles* oben).

¹⁹⁰ Siehe oben Kap. 6.1.

¹⁹¹ Siehe oben Kap. 2.1.

Step #6 Use standard vocabularies: Eigenschaften und Klassen basieren auf Standards (vgl. Bewertung der *Linked Data Principles* oben).

Step #7 Convert data: Die Daten wurden automatisiert mit Hilfe eines XSLT-Stylesheets in einem Batch-Prozess konvertiert.

Step #8 Provide machine access to data: Die Bereitstellung ist über SPARQL-Endpoint realisiert. Die ASCH-Ontologien sind in RDFS beschrieben und im Internet aufrufbar.

Step #9 Announce new data sets: Die Bekanntgabe des Graph über *CodingDaVinci*¹⁹² und die Linked Open Data Cloud ist in Planung.¹⁹³

Step #10 Recognize the social contract: Die dauerhafte Verfügbarkeit und die Bereitstellung regelmäßiger Updates des Graph sind im Ausblick benannte nächste Schritte.¹⁹⁴

Angesichts der bescheinigten Datenqualität und des Best-Practice-Vorgehens bei der Umwandlung der Daten kann diese Arbeit als Grundlage für weitere Anwendungen dienen. Die Vorgehensweise bei der Generierung von Linked Data eignet sich für die Nachahmung durch andere museale Sammlungen, insbesondere, wenn die Quelldaten in LIDO vorliegen. Gleichzeitig können die Mapping-Vorgaben sowie das XSLT-Stylesheet nicht eins zu eins adoptiert werden, denn sie wurden mit Blick auf das Göttinger Sammlungsportal erstellt. Die Berücksichtigung des jeweiligen Ausgangspunkts ist für jedwede zukünftige Anwendung unerlässlich. Unterschiede bei der Belegung der Ausgangsfelder und bei der Verwendung von URIs und Vokabularen sind ebenso unvermeidlich wie unterschiedliche Schwerpunkte bei der Beurteilung der für das Mapping relevanten Felder. Die Überlegungen und Methoden dieser Arbeit sind gleichwohl gute Beispiele, anhand derer die eigenen Ausgangsdaten untersucht und letztlich in die Umwandlung genommen werden können. *BaseX* ist als Tool für die Datenanalyse ebenso zu empfehlen wie XSLT für die Konvertierung und RDF/XML als primäres Serialisierungsformat. Dasselbe gilt für das Jena-RIOT-Tool, das die Umwandlung in das Turtle-Format realisiert. Mit Blick auf den eigentlichen Umwandlungsprozess, die Batch-Verarbeitung, und die Veröffentlichung der Daten in einem Triplestore sind die in dieser Arbeit entwickelten Verfahren ebenfalls vollständig übertragbar.

Ferner liefert diese Arbeit die für ein erfolgreiches Mapping grundlegenden Informationen zu Linked Data, XPath, XQuery und XSLT. Der Fokus liegt dabei auf den für die Zielerreichung wesentlichen Aspekten. Diese sind wiederum so ausführlich gehalten, dass sie auch Neueinsteigerinnen und Neueinsteigern einen Zugang in die Linked-Data-Thematik und die verwendeten Werkzeuge und (Programmier-) Sprachen ermöglichen.

¹⁹² <https://codingdavinci.de/> (15.01.2019).

¹⁹³ Siehe unten Kap. 8.2.

¹⁹⁴ Siehe unten Kap. 8.2.

Schließlich sollte mit der Generierung des Göttinger Knowledge Graph der Machbarkeitsbeweis für das ASCH-Modell angetreten werden. Die Eignung und Nutzbarkeit des Modells lassen sich sowohl an der Qualität der erzeugten Zieldaten als auch daran festmachen, dass sich alle relevanten Informationen abbilden ließen. Das ASCH-Modell ermöglicht einen leichten Einstieg in die RDF-Welt. Es ist sehr gut dokumentiert, und die Menge der Klassen und Eigenschaften ist überschaubar. Anwenderinnen und Anwender müssen sich nicht durch eine Vielzahl von Vokabularen arbeiten, um passende Terme zu finden. Ausgangsdaten in LIDO-XML lassen sich sehr gut abbilden. Für die wichtigen Elemente gibt es Entsprechungen in den ASCH-APs. Der Ausschnitt der Welt, auf den das ASCH-Modell zielt, die *Domain* der wissenschaftlichen Sammlungen, ist sehr gut abgedeckt. Außerdem besteht durch die Offenheit des Modells für Erweiterungen und das *Mixing-and-Matching*-Prinzip die Möglichkeit, neue Elemente zu integrieren. Von Vorteil ist, dass das ASCH-Modell zweigleisig aufgestellt ist. Neben der präferierten Einbindung vorhandener Standards existieren die eigenen Ontologien, die maschinenlesbar in Turtle vorliegen und die neue Elemente bei Bedarf beschreiben können. Die Klassen und Properties des ASCH-Modells sind logisch und eindeutig aufgebaut, was eine Zuordnung und die Bildung von Triples erleichtert. Es lassen sich auch komplexe Sachverhalte abbilden. Dazu bedarf es allerdings der Blank Nodes, zum Teil sogar in mehreren Stufen.

8.2 AUSBLICK

Mit der Bereitstellung der Daten und ersten Experimenten mit dem SPARQL-Endpoint endet diese Arbeit. Es konnte die Verknüpfung mit *DBpedia* und der Zugriff auf die dort enthaltenen Informationen gezeigt werden, gleichzeitig steht die Umsetzung sinnvoller *Use Cases* noch aus. Das ASCH-Modell hat zahlreiche *Use Cases* erarbeitet. Es wird zu prüfen sein, inwiefern sich diese auf den Knowledge Graph des Sammlungsportals übertragen lassen.

Des Weiteren werden im vorangehenden Abschnitt in den *Steps 9 und 10 der Best Practices for Publishing Linked Data* bereits zwei Anforderungen benannt, die als nächste Schritte in die Umsetzung zu nehmen sind: Die Bewerbung des Services sowie die dauerhafte Bereitstellung. Mit Blick auf den ersten Punkt ist die Vorstellung des Knowledge Graph beim nächsten *Hackathon*¹⁹⁵ von *CodingDaVinci* vorgesehen. Außerdem ist der Eintrag in die *LOD Cloud*¹⁹⁶ in Vorbereitung. In das ASCH-Wiki sollen die Ergebnisse dieser Arbeit ebenfalls aufgenommen werden. Ferner wurde der *Koordinierungsstelle für wissenschaftliche Sammlun-*

¹⁹⁵ Die Veranstaltungen werden unter <https://codingdavinci.de/events/> (15.01.2019) angezeigt.

¹⁹⁶ Unter <https://lod-cloud.net/authorize?target=/add-dataset> (15.01.2019) ist die Anmeldung neuer Datensets möglich.

gen in Deutschland der Vorschlag unterbreitet, in ihrem Portal einen Bereich mit Informationen und Links zu den Metadatenformaten der Sammlungen einzurichten.

Mit der Bewerbung des Services geht die Projektierung seiner dauerhaften Bereitstellung wie auch der nutzerfreundlichen Präsentation im Internet einher. Der SPARQL-Endpoint ist in die Website der Göttinger Sammlungen einzubinden und entsprechend einzuführen. Ferner sollen die Daten als *Data Dump* ähnlich dem *B3Kat* bereitgestellt werden. Schließlich ist geplant, dem Beispiel der *DDB* folgend, zusätzlich zum LIDO-Export-Format bei jedem Objekt einen Export der Daten in RDF/XML und Turtle anzubieten.

Neben der Bereitstellung der aktuell vorhandenen Daten ist außerdem ein Workflow für die Generierung und Veröffentlichung der Updates zu etablieren.

Gemäß dem breit aufgestellten Ansatz dieser Arbeit wurden die spezifischen Daten im Sammlungsportal sowie die zugehörigen ASCH-APs noch nicht hinreichend berücksichtigt. Die Betrachtung der ASCH-Profile für *Numismatic Item*, *Organism*, *Text* und *Evidence* in Zusammenhang mit den Quelldaten ist somit ein weiteres mögliches Handlungsfeld. Hinzu kommt das Meta-Metadata-AP, bei dem es zum Beispiel um Informationen zu den RDF Statements geht.¹⁹⁷ Das Ergebnis der Beschäftigung mit den speziellen ASCH-APs wären zusätzliche Triples für den Knowledge Graph dieser Arbeit.

Es gibt einen weiteren Grund für die Beschäftigung mit den Spezifika der numismatischen und organischen Objekte. Die VZG hat ihr Interesse angemeldet, das Mapping dieser Arbeit für die Konvertierung der Daten ihrer Kulturerbe-Portale nachzunutzen. Dies wird problemlos möglich sein, da die Daten des Sammlungsportals aus diesen Systemen ausgespielt worden sind und somit dieselbe Struktur aufweisen. Mit der Nachnutzung des Mappings für die VZG-Kulturerbe-Portale kämen zu den bisherigen gut 21.000 Datensätzen weitere 85.000 Sätze im RDF-Format hinzu (Schleier 2018).

Ein weiteres wichtiges und spannendes Thema, dessen Betrachtung sich im Nachgang dieser Arbeit lohnt, ist das sogenannte *Reasoning* oder *Inferencing*.¹⁹⁸ Dabei geht es darum, aus bereits vorhandenem Wissen neues Wissen abzuleiten und die Informationen in den generierten Triples und den eingebundenen Ontologien vollumfänglich zu nutzen und weiter zu verwerten. Dabei ist auch zu prüfen, welche zusätzlichen Daten aus verknüpften Wissenssystemen in den Triplestore einzuspielen sind, zum Beispiel aus Performancegründen. Apache Jena hält außerdem eigene *Reasoner*¹⁹⁹ für seinen *Fuseki* Triplestore vor, deren Funktions-

¹⁹⁷ Ein wichtiges Thema in diesem Bereich ist die sogenannte *RDF Reification*. Eine Beschreibung findet sich bei Manola und Miller 2004.

¹⁹⁸ Siehe oben Kap. 2.4. Ferner finden sich Einführungen in das Thema *Reasoning* bei Sakr et al. 2018, S. 21–24, 109–126.

¹⁹⁹ <https://jena.apache.org/documentation/inference/> (15.01.2019).

weise und Mehrwert in den Betrieb genommen werden könnten. Die Beschäftigung mit dem *Reasoning* in Verbindungen mit weiteren Experimenten im SPARQL-Endpoint, vorzugsweise zusammen mit den Wissenschaftlerinnen und Wissenschaftlern der musealen Einrichtungen, erscheint für die Entwicklung sinnvoller *Use Cases* für die RDF-Triples ebenfalls aussichtsreich.

Abschließend lässt sich festhalten, dass im Rahmen dieser Arbeit eine Herangehensweise vorgestellt und in die Umsetzung genommen wurde, die sich zur Nachnutzung eignet. Auf Basis des ASCH-Modells konnte ein Knowledge Graph für die wissenschaftlichen Sammlungen der Georg-August-Universität Göttingen bereitgestellt werden. Diesen gilt es im Nachgang zu verstetigen, zu erweitern, zu verbessern und in eine sinnvolle Nutzung zu überführen.

LITERATURVERZEICHNIS

- Al-Eryani, Susanne; Bucher, Gudrun; Rühle, Stefanie (2018): Ein Metadatenmodell für gemischte Sammlungen. In: *Bibliotheksdienst* 52 (7), S. 548–564. DOI: 10.1515/bd-2018-0066.
- Al-Eryani, Susanne; Rühle, Stefanie (2018): Best Practice Guide for Application Profiles and Ontologies. Created in the context of the project Developing interoperable metadata standards for contextualizing heterogeneous objects from collections, exemplified by objects of the provenance von Asch (ASCH) funded by Deutsche Forschungsgemeinschaft (DFG). SUB Göttingen. Online verfügbar unter http://asch.wiki.gwdg.de/images/3/39/BPG_v6.pdf, zuletzt geprüft am 02.02.2019.
- Allemeyer, Marie Luisa; Graepler, Daniel; Ude-Koeller, Susanne (2012): Die Universität Göttingen und ihre Sammlungen. In: *Dinge des Wissens. Die Sammlungen, Museen und Gärten der Universität Göttingen*. Göttingen: Wallstein Verlag, S. 11–17.
- Al-Mutawa, Rihab Fahd; Al-Aama, Arwa Yousuf (2017): Linked Open Data Life Cycle. In: *International Journal of Computer Applications* 175 (3), S. 34–39. DOI: 10.5120/ijca2017915489.
- Apache Jena Fuseki (2018). Online verfügbar unter <https://jena.apache.org/documentation/fuseki2/>, zuletzt geprüft am 11.02.2019.
- Asch, Georg Thomas von (1783): Korrespondenz Georg Thomas von Asch mit Christian Gottlob Heyne. Cod_Ms_Asch_1_1_1783-1787_bl_020r. Online verfügbar unter http://resolver.sub.uni-goettingen.de/purl?DE_611_BF_5619_1783_1787, zuletzt geprüft am 15.01.2019.
- Auer, Sören; Lehmann, Jens; Ngonga Ngomo, Axel-Cyrille (2011): Introduction to Linked Data and its Lifecycle on the Web. In: Axel Polleres, Claudia d’Amato, Marcelo Arenas, Siegfried Handschuh, Paula Kroner, Sascha Ossowski und Peter Patel-Schneider (Hg.): *Reasoning web. Semantic technologies for the web of data*. 7th international summer school 2011, Galway, Ireland, August 23 - 27, 2011. Tutorial lectures, Bd. 6848. Berlin: Springer (Lecture Notes in Computer Science, 6848), S. 1–75.
- Bechhofer, Sean; van Harmelen, Frank; Hendler, Jim; Horrocks, Ian; McGuinness, Deborah; Patel-Schneider, Peter; Stein, Lynn Andrea (2004): *OWL Web Ontology Language Reference* (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/owl-ref/>, zuletzt geprüft am 01.02.2019.
- Becker, Hans-Jürgen; Müller-Dreier, Armin (2005): Von der automatisierten Zeitschriften- und Serienkatalogisierung zur Digitalen Bibliothek. Vierzig Jahre EDV-Einsatz in Erwerbung und Katalogisierung. In: Margo Bargheer und Klaus Ceynowa (Hg.):

- Tradition und Zukunft - die Niedersächsische Staats- und Universitätsbibliothek Göttingen. Eine Leistungsbilanz zum 65. Geburtstag von Elmar Mittler. Unter Mitarbeit von Elmar Mittler. Göttingen: Univ.-Verl. Göttingen, S. 117–142.
- Beckett, Dave (2006): SPARQL RDF Query Language Reference v1.8. Online verfügbar unter <https://www.dajobe.org/2005/04-sparql/SPARQLreference-1.8.pdf>, zuletzt geprüft am 02.02.2019.
- Beckett, David (2014): RDF 1.1 N-Triples. A line-based syntax for an RDF graph (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/n-triples/>, zuletzt geprüft am 02.02.2019.
- Beisiegel, Ulrike (2012): Vorwort. In: Dinge des Wissens. Die Sammlungen, Museen und Gärten der Universität Göttingen. Göttingen: Wallstein Verlag, S. 9–10.
- Beisiegel, Ulrike; Reich, Mike; Pietzner, Katrin (2013): Die Sammlungen, Museen und Gärten der Universität Göttingen. Göttingen: Göttingen University Press.
- Berners-Lee, T.; Fielding, R.; Masinter, L. (2005): RFC 3986. Uniform Resource Identifier (URI). Generic Syntax. IETF. Online verfügbar unter <https://www.ietf.org/rfc/rfc3986.txt>, zuletzt geprüft am 17.01.2019.
- Berners-Lee, Tim (1994): RFC 1630. Universal Resource Identifiers in WWW. IETF. Online verfügbar unter <https://tools.ietf.org/html/rfc1630>, zuletzt geprüft am 02.02.2019.
- Berners-Lee, Tim (1998): Cool URIs don't change. (W3C Style). Online verfügbar unter <https://www.w3.org/Provider/Style/URI>, zuletzt geprüft am 02.02.2019.
- Berners-Lee, Tim (2006/2010): Linked Data - Design Issues. Online verfügbar unter <https://www.w3.org/DesignIssues/LinkedData.html>, zuletzt geprüft am 02.02.2019.
- Berners-Lee, Tim; Fielding, R.; Irvine, U. C.; Masinter, L. (1998): RFC 2396. Uniform Resource Identifiers (URI). Generic Syntax. IETF. Online verfügbar unter <https://www.ietf.org/rfc/rfc2396.txt>, zuletzt geprüft am 02.02.2019.
- Berners-Lee, Tim; Fischetti, Mark (1999): Weaving the Web. The original design and ultimate destiny of the World Wide Web by its inventor. 1. ed. San Francisco, Calif.: HarperSanFrancisco.
- Berners-Lee, Tim; Hendler, James; Lassila, Ora (2001): The Semantic Web. In: *Scientific American* 284 (5), 28–37.
- Bizer, Christian; Cyganiak, Richard (2014): RDF 1.1 TriG. RDF Dataset Language (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/trig/>, zuletzt geprüft am 02.02.2019.

- Bizer, Christian; Heath, Tom; Berners-Lee, Tim (2009): Linked Data - The Story So Far. In: *International Journal on Semantic Web and Information Systems* 5 (3), S. 1–22. DOI: 10.4018/jswis.2009081901.
- Brickley, Dan; Guna, R. V. (Hg.) (2014): RDF Schema 1.1 (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/rdf-schema/>, zuletzt geprüft am 02.02.2019.
- Cagle, Kurt (2013): Semantics and Search : MarkLogic 7 Gets RDF. Online verfügbar unter <http://blogs.avalonconsult.com/blog/search/semantics-search-marklogic-7-gets-rdf/>, zuletzt geprüft am 02.02.2019.
- Carothers, Gavin (Hg.) (2014): RDF 1.1 N-Quads. A line-based syntax for RDF datasets (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/n-quads/>, zuletzt geprüft am 02.02.2019.
- Chowdhury, Gobinda; Ruthven, Ian (2015a): Managing digital cultural heritage information. In: Cultural Heritage Information. Access and management. Unter Mitarbeit von Ian Ruthven und G. G. Chowdhury. London: Facet Publishing (iResearch), S. 1–12.
- Chowdhury, Gobinda; Ruthven, Ian (2015b): Preface. In: Cultural Heritage Information. Access and management. Unter Mitarbeit von Ian Ruthven und G. G. Chowdhury. London: Facet Publishing (iResearch), S. XV–XVII.
- Clark, James; DeRose, Steve (Hg.) (2016): XML Path Language (XPath). Version 1.0 (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/xpath-10/>, zuletzt geprüft am 02.02.2019.
- Coburn, Erin; Richard Light, Gordon; Stein, Regina; Vitzthum, Axel (2010): LIDO - Lightweight Information Describing Objects Version 1.0. International Council of Museums (ICOM). Online verfügbar unter <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>, zuletzt geprüft am 02.02.2019.
- Dimou, Anastasia (2017): High Quality Linked Data Generation from Heterogeneous Data. PhD. University of Antwerp. Antwerpen. Online verfügbar unter http://sws.semanticweb.org/sites/default/files/PhDdissertation-Dimou_0.pdf, zuletzt geprüft am 02.02.2019.
- DINI-AG KIM / Gruppe Titeldaten (2018): Empfehlungen zur RDF-Repräsentation bibliografischer Daten. Deutsche Initiative für Netzwerkinformation e. V. Kompetenzzentrums Interoperable Metadaten (DINI Schriften). Online verfügbar unter <https://nbn-resolving.org/urn:nbn:de:kobv:11-110-18452/2153.3-7>, zuletzt geprüft am 08.01.2019.
- Dreßen, Angela (2018): Open Access und Open Data – Wo stehen wir? In: *Bibliothek : Forschung und Praxis* 42 (1), S. 28–35. DOI: 10.18452/18644.

- Dublin Core Metadata Initiative (2012a): DCMI: DCMI Metadata Terms. Online verfügbar unter <http://dublincore.org/documents/dcmi-terms/>, zuletzt geprüft am 02.02.2019.
- Dublin Core Metadata Initiative (2012b): DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description. Online verfügbar unter <http://dublincore.org/documents/dces/>, zuletzt geprüft am 02.02.2019.
- Dührkohp, Frank (2015): Inventarisieren im Verbund. Kulturerbe, kuniweb, KENOM. In: Caroline Y. Robertson-Von Trotha und Ralf H. Schneider (Hg.): Digitales Kulturerbe. Bewahrung und Zugänglichkeit in der wissenschaftlichen Praxis. Karlsruhe: KIT Scientific Publishing (Kulturelle Überlieferung - digital, 2), S. 129–143.
- Dührkohp, Frank (2018): Die GBV-Datenbanken im Bereich Kulturerbe. Göttingen, 17.10.2018. Expertengespräch.
- Eckert, Kai (2013): Die Provenienz von Linked Data. In: Patrick Danowski und Adrian Pohl (Hg.): (Open) Linked Data in Bibliotheken. Berlin [u.a.]: De Gruyter, S. 97–121.
- Europeana Foundation (2015): EDM mappings, refinements and extensions. Europeana Pro. Online verfügbar unter <https://pro.europeana.eu/page/edm-profiles>, zuletzt geprüft am 02.02.2019.
- Europeana Foundation (2016): Definition of the Europeana Data Model v5.2.7. Online verfügbar unter https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.7_042016.pdf, zuletzt geprüft am 15.01.2019.
- Färber, Michael; Ell, Basil; Menne, Carsten; Rettinger, Achim (2018): A Comparative Survey of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. In: *Semantic Web* 9 (1), S. 77–129. DOI: 10.3233/SW-170275.
- Feigenbaum, Lee (2008): SPARQL by Example. The Cheat Sheet, 2008. Online verfügbar unter http://www.iro.umontreal.ca/~lapalme/ift6281/sparql-1_1-cheat-sheet.pdf, zuletzt geprüft am 02.02.2019.
- Fielding; Roy; Reschke; Julian (2014): RFC 7231. Hypertext Transfer Protocol (HTTP/1.1). Semantics and Content. IETF. Online verfügbar unter <https://tools.ietf.org/html/rfc7231#section-3.4>, zuletzt geprüft am 17.01.2019.
- Gandon, Fabien; Schreiber, Guus (Hg.) (2014): RDF 1.1 XML Syntax (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/rdf-syntax-grammar/>, zuletzt geprüft am 02.02.2019.
- Georg-August-Universität Göttingen (2012): Dinge des Wissens. Die Sammlungen, Museen und Gärten der Universität Göttingen. Göttingen: Wallstein Verlag.

- Georg-August-Universität Göttingen (2017): Nutzung - Sammlungsportal der Georg-August-Universität Göttingen. Online verfügbar unter <https://sammlungen.uni-goettingen.de/nutzung/>, zuletzt geprüft am 02.02.2019.
- Gerdes, Thomas (2018): Die Open-Science-Bewegung und ihre Bedeutung für die wissenschaftlichen Bibliotheken. Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, Berlin. Online verfügbar unter <https://doi.org/10.18452/18983>, zuletzt geprüft am 02.02.2019.
- Glimm, Birte (2011): Using SPARQL with RDFS and OWL Entailment. In: Axel Polleres, Claudia d'Amato, Marcelo Arenas, Siegfried Handschuh, Paula Kroner, Sascha Ossowski und Peter Patel-Schneider (Hg.): Reasoning web. Semantic technologies for the web of data. 7th international summer school 2011, Galway, Ireland, August 23 - 27, 2011. Tutorial lectures. Berlin: Springer (Lecture Notes in Computer Science, 6848), S. 137–201.
- Gradmann, Stefan (2010): Knowledge = Information in Context. On the Importance of Semantic Contextualisation in Europeana (Europeana White Paper, 1). Online verfügbar unter https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20White%20Paper%201.pdf, zuletzt geprüft am 02.02.2019.
- Grupe, Wilfried (2018): XML. Grundlagen, Technologien, Validierung, Auswertung. 1. Aufl. Frechen: MITP.
- Hagedorn-Saupe, Monika (2012): Dokumentieren und Digitalisieren. Erschließung, Vernetzung und Access - Zugang für alle. In: Bernhard Graf und Volker Rodekamp (Hg.): Museen zwischen Qualität und Relevanz. Denkschrift zur Lage der Museen. 1. Aufl. Berlin: G+H Verl. (Berliner Schriftenreihe zur Museumsforschung, 30), S. 195–207.
- Harris, Steve; Seaborne, Andy (Hg.) (2013): SPARQL 1.1 Query Language (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/sparql11-query/>, zuletzt geprüft am 02.02.2019.
- Haslhofer, Bernhard; Isaac, Antoine (2011): data.europeana.eu. The Europeana Linked Open Data Pilot. In: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, S. 94–104. Online verfügbar unter <http://dcpapers.dublincore.org/pubs/article/view/3625>, zuletzt geprüft am 02.02.2019.
- Heath, Tom; Bizer, Christian (2011): Linked data. Evolving the web into a global data space. 1. ed. San Rafael, Calif.: Morgan & Claypool (Synthesis lectures on the semantic web, theory and technology, 1).
- Hyland, Bernadette; Atemezing, Ghislain; Villazón-Terrazas, Boris (Hg.) (2014): Best Practices for Publishing Linked Data. W3C Working Group Note 09 January 2014.

- W3C. Online verfügbar unter <https://www.w3.org/TR/ld-bp/>, zuletzt geprüft am 02.02.2019.
- ICOM/CIDOC (2010a): LIDO mandatory elements. Online verfügbar unter <http://network.icom.museum/cidoc/working-groups/lido/lido-overview/lido-mandatory-elements/>, zuletzt geprüft am 02.02.2019.
- ICOM/CIDOC (2010b): What is LIDO. Online verfügbar unter <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/>, zuletzt geprüft am 02.02.2019.
- Jahnke, Alex (2017): Linked Data und RDF. Niedersächsische Staats- und Universitätsbibliothek Göttingen. Göttingen, 28.08.2017. Online verfügbar unter https://www.eromm.org/dcgkb/lib/exe/fetch.php?media=schulungen:public:linked_open_data_1.pdf, zuletzt geprüft am 02.02.2019.
- Kay, Michael (Hg.) (2017): XPath and XQuery Functions and Operators 3.1 (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/xpath-functions-31/>, zuletzt geprüft am 02.02.2019.
- Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland (2012): Aufgaben der Koordinierungsstelle. Online verfügbar unter <https://wissenschaftliche-sammlungen.de/de/uber-uns/aufgaben/>, zuletzt geprüft am 02.02.2019.
- Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland (2018): Kennzahlen zu wissenschaftlichen Sammlungen an deutschen Universitäten. Online verfügbar unter <https://portal.wissenschaftliche-sammlungen.de/kennzahlen>, zuletzt geprüft am 02.02.2019.
- Le Boeuf, Patrick; Doerr, Martin; Ore, Christian Emil; Stead, Stephen (Hg.) (2015): Definition of the CIDOC Conceptual Reference Model. Version 6.2. CIDOC CRM Special Interest Group. Online verfügbar unter http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_6.2.pdf, zuletzt geprüft am 02.02.2019.
- Library of Congress (2017): MARC Code List for Relators. LC Linked Data Service: Authorities and Vocabularies (Library of Congress). Online verfügbar unter <http://id.loc.gov/vocabulary/relators.html>, zuletzt geprüft am 02.02.2019.
- Manola, Frank; Miller, Eric (Hg.) (2004): RDF Primer (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/rdf-primer/>, zuletzt geprüft am 02.02.2019.
- McCrae, John P.: The Linked Open Data Cloud. Online verfügbar unter <https://lod-cloud.net/>, zuletzt geprüft am 02.02.2019.

- Miles, Alistair; Bechhofer, Sean (Hg.) (2009): SKOS Simple Knowledge Organization System Reference (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/skos-reference/>, zuletzt geprüft am 02.02.2019.
- Nilsson, Mikael; Baker, Thomas; Jonston, Pete (2008): The Singapore Framework for Dublin Core Application Profiles. Online verfügbar unter <http://dublincore.org/documents/singapore-framework/>, zuletzt geprüft am 02.02.2019.
- Oomen, Johan; Baltussen, Lotte Belice; van Erp, Marieke (2012): Sharing cultural heritage the linked open data way. Why you should sign up (Annual Museums and the Web Conference, 16). Online verfügbar unter https://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data, zuletzt geprüft am 02.02.2019.
- Oren, Eyal; Schenk, Simon (2011): Semantic Web Basics. In: Raphael Troncy, Benoit Huet und Simon Schenk (Hg.): Multimedia semantics. Metadata, analysis, and interaction. Chichester, West Sussex, U.K, Hoboken, N.J: Wiley, S. 81–98.
- Panchal, Ronak (2017): Easy install and start Apache Jena Fuseki Server into windows 10 - YouTube. Online verfügbar unter <https://www.youtube.com/watch?v=3WTtKalfqnk&t=182s>, zuletzt geprüft am 02.02.2019.
- Pohl, Adrian (2013): Glossar. In: Patrick Danowski und Adrian Pohl (Hg.): (Open) Linked Data in Bibliotheken. Berlin [u.a.]: De Gruyter, S. 235–244.
- Prud'hommeaux, Eric; Carothers, Gavin (Hg.) (2014): RDF 1.1 Turtle. Terse RDF Triple Language (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/turtle/>, zuletzt geprüft am 02.02.2019.
- Prud'hommeaux, Eric; Seaborne, Andy (Hg.) (2008): SPARQL Query Language for RDF (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/rdf-sparql-query/>, zuletzt geprüft am 02.02.2019.
- Rühle, Stefanie (2018a): Item Description Ontology (IDO). Version 3.0. GitHub-Repository subugoe/metadata. Niedersächsische Staats- und Universitätsbibliothek Göttingen. Göttingen. Online verfügbar unter <http://http://purl.org/asch/ido/180704>, zuletzt geprüft am 02.02.2019.
- Rühle, Stefanie (2018b): Provenance Description Ontology (PDO). Version 3.0. GitHub-Repository subugoe/metadata. Niedersächsische Staats- und Universitätsbibliothek Göttingen. Göttingen. Online verfügbar unter <http://http://purl.org/asch/pdo/180704>, zuletzt geprüft am 02.02.2019.
- Rühle, Stefanie (2018c): Entwicklung des ASCH-Modells. Göttingen, 21.09.2018. Expertengespräch.

Sakr, Sherif; Wylot, Marcin; Mutharaju, Raghava; Le-Phuoc, Danh; Fundulaki, Irini (2018): Linked data. Storing, querying, and reasoning. Cham: Springer. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-319-73515-3>, zuletzt geprüft am 15.01.2019.

Sauermann, Leo; Cyganiak, Richard (Hg.) (2008): Cool URIs for the Semantic Web. W3C Interest Group Note 03. W3C Interest Group. Online verfügbar unter <https://www.w3.org/TR/cooluris/>, zuletzt geprüft am 02.02.2019.

Schleier, Timo (2018): Das LIDO-XML des Göttinger Sammlungsportals. Göttingen, 11.10.2018. Expertengespräch.

Scholz, Henning (2015): Europeana - Digitale Dienstleistungs-Infrastruktur für Europas Kulturerbe. In: Ellen Euler, Monika Hagedorn-Saupe, Gerald Maier, Werner Schweibenz und Jörn Sieglerschmidt (Hg.): Handbuch Kulturportale. Online-Angebote aus Kultur und Wissenschaft. Berlin: De Gruyter Saur (De Gruyter Saur Reference), S. 125–132.

Schreiber, Guus; Raimond, Yves (Hg.) (2014): RDF 1.1 Primer (W3C Working Group Note). Online verfügbar unter <https://www.w3.org/TR/rdf11-primer/>, zuletzt geprüft am 02.02.2019.

Simou, Nikolaos (2014): AthenaPlus D3.2. Description of the LIDO to EDM mapping. Online verfügbar unter <http://www.athenaplus.eu/getFile.php?id=285>, zuletzt geprüft am 15.01.2019.

solid IT (2019a): Ranking of RDF Stores - DB-Engines Enzyklopädie. Online verfügbar unter <https://db-engines.com/en/ranking/rdf+store>, zuletzt geprüft am 11.02.2019.

solid IT (2019b): RDF Stores - DB-Engines Enzyklopädie. Online verfügbar unter <https://db-engines.com/de/article/RDF+Stores>, zuletzt geprüft am 02.02.2019.

Spitz, Malte (2017): Daten - Das Öl des 21. Jahrhunderts? Nachhaltigkeit im digitalen Zeitalter. 1. Auflage. Hamburg: Hoffmann und Campe.

Sporny, Manu; Kellogg, Gregg; Lanthaler, Markus (Hg.) (2014): JSON-LD 1.0. A JSON-based Serialization for Linked Data (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/json-ld/>, zuletzt geprüft am 02.02.2019.

Stichbury, Jo (2017): WTF is a knowledge graph? – Hacker Noon. Unpicking a tangle of terminology to conclude it's semantic, smart and alive. Online verfügbar unter <https://hackernoon.com/wtf-is-a-knowledge-graph-a16603a1a25f>, zuletzt geprüft am 02.02.2019.

SUB Göttingen (2014a): ASCH - Projektdetails. Entwicklung von interoperablen Standards für die Kontextualisierung heterogener Objekte am Beispiel der Provenienz Asch. Online verfügbar unter <https://www.sub.uni-goettingen.de/projekte-forschung/projektdetails/projekt/asch/>, zuletzt geprüft am 02.02.2019.

- SUB Göttingen (2014b): ASCH-Wiki Main Page. Göttingen. Online verfügbar unter http://asch.wiki.gwdg.de/index.php/Main_Page, zuletzt geprüft am 02.02.2019.
- SUB Göttingen (2017): Use Cases - Asch. Online verfügbar unter http://asch.wiki.gwdg.de/index.php/Use_Cases, zuletzt geprüft am 02.02.2019.
- Suominen, Osmo; Hyvönen, Nina (2017): From MARC silos to Linked Data silos? In: *o-bib. Das offene Bibliotheksjournal* 4 (2), S. 1–13. DOI: 10.5282/o-bib/2017H2S1-13.
- Tauberer, Joshua (2008): What is RDF and what is it good for? Online verfügbar unter <https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md>, zuletzt geprüft am 02.02.2019.
- The Economist (Hg.) (2017): The world's most valuable resource is no longer oil, but data. Online verfügbar unter <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>, zuletzt geprüft am 02.02.2019.
- Tsalapate, Eleni; Stein, Regine; Simou, Nikolaos; Drosopoulos, Nasos (2012): Evolving LIDO based aggregations into Linked Data. Helsinki (CIDOC 2012 Helsinki). Online verfügbar unter http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/Conference_Papers/2012/simou.pdf, zuletzt geprüft am 02.02.2019.
- UNESCO (2008): World Heritage Information Kit. Online verfügbar unter <https://whc.unesco.org/uploads/activities/documents/activity-567-1.pdf>, zuletzt geprüft am 02.02.2019.
- Verbundzentrale des GBV (2018): kuniweb. Dokumentation des LIDO-Exportformats für kuniweb. Online verfügbar unter <http://hdl.handle.net/428894.vzg/4fd5051b-d2a4-47a6-9656-bc1c342d0df4>, zuletzt geprüft am 02.02.2019.
- Vonhoegen, Helmut (2018): XML. Einstieg, Praxis, Referenz. 9., aktualisierte Auflage. Bonn: Rheinwerk Verlag (Rheinwerk Computing).
- W3C: W3C RDF Validation Service. Online verfügbar unter <https://www.w3.org/RDF/Validator/>, zuletzt geprüft am 02.02.2019.
- W3C (2015): Vocabularies. Online verfügbar unter <https://www.w3.org/standards/semanticweb/ontology>, zuletzt geprüft am 02.02.2019.
- W3C OWL Working Group (Hg.) (2012): OWL 2 Web Ontology Language. Document Overview. 2. Edition (W3C Recommendation). Online verfügbar unter <https://www.w3.org/TR/owl2-overview/>, zuletzt geprüft am 02.02.2019.
- Wissenschaftsrat (2011): Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen (Drs. 10464-11). Berlin. Online verfügbar unter

<https://www.wissenschaftsrat.de/download/archiv/10464-11.pdf>, zuletzt geprüft
am 02.02.2019.

ANHANG 1: Mapping-Tabelle lido2rdf

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
/lido:lido/lido:objectPublishedID (Occ 2)				
Item/Sammlungs- objekt	/lido:lido/lido:objectPublishedID	Item/rdf:type/pdo:Item	Nein	Enthält Zeichenkette <i>hdl.handle</i>
Record/Datensatz	/lido:lido/lido:objectPublishedID	Record/rdf:type/pdo:Record	Nein	Enthält Zeichenkette <i>sammlungen.uni- goettingen.de</i>
/lido:lido/lido:descriptiveMetadata/lido:objectClassificationWrap/ (Occ. 1)				
Objektgattung	li- do:objectWorkTypeWrap/lido:objectWorkType/lido:conceptID	Item/rdf:type/URI	Nein	Für jedes Item eine Objektgattung vh.
Klassifikation	li- do:classificationWrap/lido:classification/lido:conceptID	Item/rdf:type/URI	Ja	Nur URIs werden berücksichtigt
/lido:lido/lido:descriptiveMetadata/lido:objectIdentificationWrap/ (Occ. 1)				
Titel bzw. Bezeichnung	li- do:titleWrap/lido:titleSet/lido:appellationValue	Item/ido:hasAppellation/BN BN/rdf:type/ido:Appellation BN/skos:prefLabel/Literal bzw. BN/skos:altLabel/Literal	Ja: altLabel- Nein: prefLabel	skos:prefLabel bei Attribut sortorder = '1' Ansonsten skos:altLabel
Inschrift	li- do:inscriptionsWrap/lido:inscriptions/lido:inscriptionTranscription	Item/ido:physicalDescription/BN BN/rdf:type/ido:Appearance BN/ido:inscription/Literal	Ja	Kein language- Attribut, da Daten nicht eindeutig; Typ wird nicht be- rücksichtigt, da Daten nicht aussagekräftig; Abgleich mit object- MeasurementWrap notwendig, da es nur eine ido:Appearance geben darf
lido:repositoryWrap/lido:repositorySet/ (Occ. 1/1)				
Aufbewahrungs- ort Einrichtung	lido:repositoryName/lido:legalBodyID	Item/ido:wasAffectedBy/BN BN/rdf:type/pdo:Custody BN/marcrel:cur/URI	Nein	
Aufbewahrungs- ort Koordinaten des Ortes	li- do:repositoryLocation/lido:gml/gml:Point/gml:pos	BN/pdo:tookPlaceAt/BN-P BN-P/rdf:type/pdo:Place BN-P/wgs84_pos:lat/Literal BN-P/wgs84_pos:long/Literal	Nein	Kein URI in den Daten vh. Längen- und Brei- tengrad befinden sich in einem Feld und müssen anhand der Leerstelle identifiziert werden.
Aufbewahrungs- ort Signatur bzw. Zugangsnummer	lido:workID	Item/dc:identifier/Literal	Nein	
lido:objectDescriptionWrap/lido:objectDescriptionSet (Occ. 0-1/1-n)				

Anhang 1: Mapping-Tabelle lido2rdf

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
Beschreibung	[@lido:type='description']/lido:descriptiveNoteValue	Item/dc:description @xml:lang/Literal	Ja	lido:type="description"
Gebrauch	[@lido:type='usage']/lido:descriptiveNoteValue	Item/lido:wasAffectedBy/BN BN/rdf:type/http://terminology.lido-schema.org/eventType/use BN/pdo:hasPurpose @xml:lang/Literal	Ja pdo:hasPurpose	Abgleich mit Event-Wrap notwendig, damit kein zweites Event Use entsteht.
Produktion	[@lido:type='production_purpose']/lido:descriptiveNoteValue	Item/lido:wasAffectedBy/BN BN/rdf:type/http://terminology.lido-schema.org/eventType/production BN/pdo:hasPurpose @xml:lang/Literal	Ja pdo:hasPurpose	Abgleich mit Event-Wrap notwendig, damit kein zweites Event Production entsteht.
lido:objectMeasurementWrap/lido:objectMeasurementsSet/ (Occ. 0-1/1-n)				
Maßangaben Typ	lido:objectMeasurements/lido:measurementsSet/lido:measurementType	Item/lido:physicalDescription/BN BN/rdf:type/lido:Appearance BN/lido:numberOfParts/Literal BN/dcterms:extent/BN-D BN-D/rdf:type/crm:E54_Dimension BN-D/dc:type/URI bzw. Literal Literal bei „width“ und „size“ Ansonsten bezeichnet URI den Typ: height: http://d-nb.info/gnd/4229619-5 depth: http://d-nb.info/gnd/4334905-5 length: http://d-nb.info/gnd/4224141-8 diameter: http://d-nb.info/gnd/4273067-3 circumference: http://d-nb.info/gnd/7583293-8 weight: http://d-nb.info/gnd/4157270-1 volume: http://d-nb.info/gnd/4136953-1	Ja: Dimensi- ons	Abgleich mit inscription notwendig, damit kein zweites lido:Appearance entsteht.
Maßangaben Einheit	lido:objectMeasurements/lido:measurementsSet/lido:measurementUnit	BN-D/crm:P91_has_unit @xml:lang/Literal	Dimensi- ons: Ja	
Maßangaben Wert	lido:objectMeasurements/lido:measurementsSet/lido:measurementValue	BN-D/crm:P91_has_value/ Literal	Dimensi- ons: Ja	
Maßangaben Notiz	lido:objectMeasurements/lido:extentMeasurements	BN-D/crm:P91_has_value/ Literal	Dimensi- ons: Ja	
/lido:lido/lido:descriptiveMetadata/lido:eventWrap/lido:eventSet (Occ. 0-1/1-n)				

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
Typ des Events	li- do:event/lido:eventType/lid o:conceptID	Item/ ido:wasAffectedBy /BN BN/ rdf:type /URI	Ja	Abgleich mit den Events Use und Pro- duction aus dem objectDescripti- onWrap notwendig
Pro Event kann es Actors, Dates, Places, Culture und Material geben (vgl. die nächsten sechs Zeilen)				
Akteur des Events	<u>URI des eventActors:</u> li- do:event/lido:eventActor/lid o:actorInRole/lido:actor/lido :actorID <u>Rolle des eventActors:</u> li- do:event/lido:eventActor/lid o:actorInRole/lido:roleActor/ lido:conceptID	BN/ marcrel:... /URI bzw. BN/ pdo:participant /URI	Ja	Nur URIs! Falls die Daten die marcrel-Property enthalten, soll diese genutzt werden, ansonsten pdo:participant. Mehrere eventActors mit je mehreren Rol- len möglich
Datum des Events	<u>EarliestDate:</u> li- do:event/lido:eventDate/lid o:date/lido:earliestDate <u>LatestDate:</u> li- do:event/lido:eventDate/lid o:date/lido:latestDate <u>Qualifier:</u> Im type-Attribut von ear- liestDate und latestDate	BN/ pdo:start /BN-T (earliest- Date) bzw. BN/ pdo:end /BN-T (latestDate) BN-T/ rdf:type /pdo:Time BN-T/ rdf:value @ rdf:datatype /Literal BN-T/ pdo:qualifier @ xml:lang /Literal	Nein Pro Event nur ein earliest und ein latest date	Als Qualifier sind „approximate“ und „estimate“ möglich mit Language-Attribut „en“
Ort des Events	<u>Place-URI:</u> li- do:event/lido:eventPlace/lid o:place/lido:placeID <u>Koordinaten:</u> li- do:eventPlace/lido:place/lid o:gml/gml:Point/gml:pos	<u>URI:</u> BN/ pdo:tookPlaceAt /URI <u>Koordinaten (falls keine URI vh.)</u> BN/ pdo:tookPlaceAt /BN-P BN-P/ rdf:type /pdo:Place BN-P/ wgs84_pos:lat /Literal BN-P/ wgs84_pos:long /Literal	Ja	Ist ein URI aus Geo- Names vh., werden die Koordinaten nicht berücksichtigt. Längen- und Breiten- grad befinden sich in einem Feld und müs- sen anhand der Leer- stelle identifiziert und selektiert werden.
Ethnie in Zusam- menhang mit dem Event	li- do:event/lido:culture/lido:co nceptID	BN/ pdo:culturalContext /URI	Ja	Es werden nur URIs berücksichtigt.

Anhang 1: Mapping-Tabelle lido2rdf

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
Material und Technik in Zusammenhang mit dem Event	<u>URI:</u> lido:event/lido:eventMaterial sTech/lido:materialsTech/lido: o:termMaterialsTech/lido:co nceptID <u>Literal:</u> li- do:event/lido:eventMaterial sTech/lido:materialsTech/lido: o:termMaterialsTech/lido:ter m[1]	li- do:type=".../termMaterialsTech _type/material" BN/pdo:material/URI oder Lit- eral li- do:type=".../termMaterialsTech _type/technique" BN/pdo:technique/URI oder Literal	Ja	Ist ein AAT-URI vh., wird nur dieser be- rücksichtigt; ansons- ten der GND-URI und als dritte Wahl das erste Literal mit lan- guage-Attribut
Farbe in Zusammen- hang mit Material und Technik	li- do:event/lido:eventMaterial sTech/lido:materialsTech/lido: o:termMaterialsTech/lido:ter m[1]	li- do:type="http://terminology.lid o-schema.org/lido00479" BN/ido:colour @xml:lang/Literal	Ja	Farbe in jeder Spra- che übernehmen
/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/ (Occ. 0-1)				
Schlagwort = Subject	<u>Sachbegriff:</u> li- do:subjectWrap/lido:subject Set/lido:subject/lido:subject Concept/lido:conceptID <u>Akteur als Schlagwort:</u> li- do:subjectWrap/lido:subject Set/lido:subject/lido:subject Actor/lido:actor/lido:actorID <u>Ort als Schlagwort:</u> li- do:subjectWrap/lido:subject Set/lido:subject/lido:subject Place/lido:place/lido:placeID	Item/dc:subject/URI	Ja	0-1 subjectWrap mit 1-n subjectSets Es werden nur URIs berücksichtigt.
Verknüpfte Ob- jekte = rela- tedWorks isPartOf bzw. relation	li- do:relatedWorksWrap/lido:r elated- WorkSet/lido:relatedWork/li do:object/lido:objectWebRe source	Falls "ist Teil von" in li- do:relatedWorksWrap/lido:relat ed- WorkSet/lido:relatedWorkRelTy pe/lido:term Item/dcterms:isPartOf/URI Ansonsten: Item/dc:relation/URI	Ja	
Verknüpfte Ob- jekte = rela- tedWorks GVK-URIS	li- do:relatedWorksWrap/lido:r elated- WorkSet/lido:relatedWork/li do:object/lido:objectID	Item/dcterms:isReferencedBy/ URI	Ja	@lido:type='http://te rminology.lido- sche- ma.org/identifie r_type/uri']

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
Verknüpfte Objekte = relatedWorks Collection/Sammlung	li-do:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/li-do:object/lido:objectID	Item/ido:isGatheredInto/BN BN/rdf:type/ido:Collection BN/dc:identifier/Literal	Ja	Attribute @lido:type='http://terminology.lido-sche-ma.org/identifizier_type/local_identifizier' oder 'http://terminology.lido-sche-ma.org/lido00100'
/lido:lido/lido:administrativeMetadata/lido:recordWrap/ (Occ. 1)				
Metadaten zum Datensatz/Record Verknüpfung Record mit Item	<u>Record:</u> /lido:lido/lido:objectPublishedID	Record/foaf:primaryTopic/Item	Nein	Record enthält <i>sammlungen.uni-goettingen.de</i> ; Pro Item gibt es ein Record.
Rechteinhaber des Records	<u>URI:</u> li-do:recordRights[1]/lido:rightsHolder/lido:legalBodyID <u>Literal:</u> li-do:recordRights[1]/lido:rightsHolder/lido:legalBodyName/lido:appellationValue	<u>URI:</u> Record/dcterms:rightsHolder/URI <u>Literal:</u> Record/dcterms:rightsHolder/BN BN/rdf:type/pdo:Agent BN/foaf:name/Literal	Nein	
Recht / Lizenz für die Verwertung des Records	li-do:recordRights[2]/lido:rightsType/lido:conceptID	Record/dc:rights/URI	Nein	
Datum der Erstellung des Records	lido:recordInfoSet/lido:recordMetadataDate	Record/dcterms:created @rdf:datatype/Literal	Nein	recordInfoSet-type lido00470 and recordMetadataDate-type created
/lido:lido/lido:administrativeMetadata/lido:resourceWrap/ (Occ.-0-1)				
Digitale Repräsentation des Items (image/jpeg)	li-do:resourceSet/lido:resourceRepresentation/lido:linkResource	Item/ido:hasDigitalRepresentation/URI DigRep/rdf:type/ ido:DigitalRepresentation	Ja	Type „provided...“ wird berücksichtigt. Es kann viele DigRep eines Items geben, jew. mit Rechten etc.
Rechteinhaber der digitalen Repräsentation des Items (image/jpeg)	<u>Verwertungsrecht URI:</u> li-do:resourceSet/lido:rightsResource/lido:rightsHolder/lido:legalBodyID <u>Verwertungsrecht Literal:</u> li-do:resourceSet/lido:rightsResource/lido:rightsHolder/lido:legalBodyName/lido:appellationValue	<u>URI:</u> DigRep/dcterms:rightsHolder/URI <u>Literal:</u> DigRep/dcterms:rightsHolder/BN BN/rdf:type/pdo:Agent BN/foaf:name/Literal	Nein	Pro DigRep ein Rechteinhaber. lido:rightsType/lido:term = 'Verwertungsrecht'

Anhang 1: Mapping-Tabelle lido2rdf

Beschreibung	LIDO XPath	RDF Triple Subjekt/Prädikat/Objekt	Element wieder- holbar?	Kommentar
Urheber der digitalen Repräsentation des Items (image/jpeg)	<u>Urheberrecht URI:</u> li-do:resourceSet/lido:rightsResource/lido:rightsHolder/lido:legalBodyID <u>Urheberrecht Literal:</u> li-do:resourceSet/lido:rightsResource/lido:rightsHolder/lido:legalBodyName/lido:appellationValue	<u>URI:</u> DigRep/dc:creator/URI <u>Literal:</u> DigRep/dc:creator/Literal	Nein	Pro DigRep ein Urheber lido:rightsType/lido:term = 'Urheberrecht'"
Recht/Lizenz für die Verwertung der digitalen Repräsentation	li-do:resourceSet/lido:rightsResource/lido:rightsType/lido:conceptID	DigRep /dc:rights/URI		
Datum der Erstellung der digitalen Repräsentation	li-do:resourceSet/lido:resourceDateTaken/lido:date/lido:earliestDate	DigRep/dcterms:created @rdf:datatype/Literal	Nein	Pro DigRep ein Datum

ANHANG 2: XSLT-Stylesheet lido2rdfxml.xsl

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
3      xmlns:lido="http://www.lido-schema.org"
4      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
5      xmlns:dc="http://purl.org/dc/elements/1.1/"
6      xmlns:skos="http://www.w3.org/2004/02/skos/core#"
7      xmlns:gml="http://www.opengis.net/gml"
8      xmlns:gn="http://www.geonames.org/ontology#"
9      xmlns:wgs84_pos="http://www.w3.org/2003/01/geo/wgs84_pos#"
10     xmlns:marcrel="http://id.loc.gov/vocabulary/relators/"
11     xmlns:ido="http://purl.org/asch/ido/180704#"
12     xmlns:pdo="http://purl.org/asch/pdo/180704#"
13     xmlns:owl="http://www.w3.org/2002/07/owl#"
14     xmlns:crm="http://erlangen-crm.org/170309/"
15     xmlns:aat="http://vocab.getty.edu/aat/"
16     xmlns:gnd="http://d-nb.info/gnd/"
17     xmlns:nms="http://nomisma.org/ontology#"
18     xmlns:dcmitype="http://purl.org/dc/dcmitype/"
19     xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
20     xmlns:edm="http://www.europeana.eu/schemas/edm/"
21     xmlns:dcterms="http://purl.org/dc/terms/"
22     xmlns:prov="http://www.w3.org/ns/prov#"
23     xmlns:foaf="http://xmlns.com/foaf/0.1/"
24     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
25     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
26     xsi:schemaLocation="http://www.w3.org/1999/XSL/Transform
27         http://www.w3.org/2007/schema-for-xslt20.xsd"
28     version="2.0">
29
30     <!-- Version 1_DRAFT-2018-12-23: ASCH_lido2rdfxml_20181223.xsl
31
32     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
33     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
34
35     @Author:      Antje Niemann
36     @Date:        23.12.2018
37     @Version:     1
38
39     XSIT to transform the LIDO-XML of the Sammlungsportal der
40     Georg-August-Universität Göttingen to RDF/XML
41     The underlying RDF-model is the ASCH-model
42     http://asch.wiki.gwdg.de/index.php/ASCH_Model
43
44     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
45     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
46
47     -->
48
49     <xsl:output indent="yes" />
50
51     <xsl:template match="/">
52         <xsl:apply-templates />
53     </xsl:template>
54
55     <xsl:template match="lido:lido">
56
57         <!-- VARIABLES -->
58
59         <!-- handle-URI; every record contains a handle-URI-->
60         <xsl:variable name="var_handle-uri"
61             select="lido:objectPublishedID[contains(text(), 'hdl.handle')]/text()" />
62
63         <!-- record-URI; every record contains a record-URI with exactly one right
64             and one rightsHolder and one lido creation date -->
65         <xsl:variable name="var_record-uri"
66             select="lido:objectPublishedID[contains(text(),
67                 'sammlungen.uni-goettingen.de')]/text()" />
68
69         <!-- record_rightsHolder is always given in the first
70             lido:recordRights-element (Sammlungsportal-data does not contain important
71             rightsHolder information in the second wrapper) -->
72         <xsl:variable name="var_record_rightsHolder_URI"
```

```

60     select="lido:administrativeMetadata/lido:recordWrap/lido:recordRights[1]/lido:
61     rightsHolder/lido:legalBodyID/text()" />
    <!-- record_rightsHolder as literal -->
    <xsl:variable name="var_record_rightsHolder_literal"
62     select="lido:administrativeMetadata/lido:recordWrap/lido:recordRights[1]/lido:
63     rightsHolder/lido:legalBodyName/lido:appellationValue/text()" />
    <!-- record_rights are only considered with type uri which is always given
    in the second lido:recordRights-element -->
    <xsl:variable name="var_record_rights"
64     select="lido:administrativeMetadata/lido:recordWrap/lido:recordRights[2]/lido:
    rightsType/

        lido:conceptID[@lido:type='http://terminology.lido-schema.org/identifier_t
        ype/uri']/text()" />
65     <!-- record_rightsDate is only considered with recordInfoSet-type lido00470
    and recordMetadataDate-type created -->
66     <xsl:variable name="var_record_rightsDate"
    select="lido:administrativeMetadata/lido:recordWrap/

        lido:recordInfoSet[@lido:type='http://terminology.lido-schema.org/lido0047
        0']/lido:recordMetadataDate[@lido:type='http://terminology.lido-schema.org
        /recordMetadataDate_type/created']/text()" />
67
68
69 <rdf:RDF>
70     <!-- Handle-URI as rdf-subject -->
71     <rdf:Description rdf:about="{ $var_handle-uri }">
72         <xsl:apply-templates />
73     </rdf:Description>
74
75     <!-- Record-URI as rdf-subject: MetaMetaData: The record which describes
    the item (handle_URI). The record's creation date, rights and
    rightsHolder are considered -->
76     <!-- Example for programming with variables which are defined at the top
    -->
77     <rdf:Description rdf:about="{ $var_record-uri }">
78         <foaf:primaryTopic rdf:resource="{ $var_handle-uri }" />
79         <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Record" />
80         <!-- date: There could be records without a date -->
81         <xsl:if test="$var_record_rightsDate">
82             <dcterms:created rdf:datatype="http://purl.org/dc/terms/W3CDTF">
83                 <xsl:value-of select="$var_record_rightsDate" />
84             </dcterms:created>
85         </xsl:if>
86         <!-- right is always given as URI -->
87         <dc:rights rdf:resource="{ $var_record_rights }"
        />
88         <!-- rightsHolder: URI is first choice, literal second. In any case
        a rightsHolder has to be delivered -->
89         <xsl:choose>
90             <xsl:when test="$var_record_rightsHolder_URI">
91                 <dcterms:rightsHolder
                    rdf:resource="{ $var_record_rightsHolder_URI }" />
92             </xsl:when>
93             <xsl:otherwise>
94                 <xsl:if test="$var_record_rightsHolder_literal">
95                     <dcterms:rightsHolder>
96                         <!-- Blank note pdo:Agent -->
97                         <rdf:Description>
98                             <rdf:type
                                rdf:resource="http://purl.org/asch/pdo/180704#Agen
                                t" />
99                             <foaf:name>
100                                 <xsl:value-of
                                    select="$var_record_rightsHolder_literal" />
101                                 </foaf:name>
102                             </rdf:Description>
103                         </dcterms:rightsHolder>
104                     </xsl:if>
105                 </xsl:otherwise>
106             </xsl:choose>
107         </rdf:Description>
108

```



```

109      <!-- DigitalRepresentation (image/jpeg) as rdf-subject: There could be
110      more than one DigitalRepresentation -->
111      <xsl:for-each
112      select="lido:administrativeMetadata[1]/lido:resourceWrap[1]/lido:resourceS
113      et">
114      <xsl:if
115      test="lido:resourceRepresentation[@lido:type='http://terminology.lido-
116      schema.org/resourceRepresentation_type/provided_representation']/
117      lido:linkResource[@lido:formatResource='image/jpeg']">
118      <!-- DigitalRepresentation as subject -->
119      <rdf:Description
120      rdf:about="{lido:resourceRepresentation[@lido:type='http://termino
121      logy.lido-schema.org/resourceRepresentation_type/provided_represen
122      tation']/lido:linkResource[@lido:formatResource='image/jpeg']/text
123      ()}">
124      <rdf:type
125      rdf:resource="http://purl.org/asch/ido/180704#DigitalRepresent
126      ation" />
127      <!-- Date is not always given -->
128      <xsl:if
129      test="lido:resourceDateTaken/lido:date/lido:earliestDate">
130      <dcterms:created
131      rdf:datatype="http://purl.org/dc/terms/W3CDTF">
132      <xsl:value-of
133      select="lido:resourceDateTaken/lido:date/lido:earliest
134      Date" />
135      </dcterms:created>
136      </xsl:if>
137      <!-- Rights are always given as URI in conceptID and there
138      is only one conceptID in one of the three possible
139      rightsResources -->
140      <dc:rights
141      rdf:resource="{lido:rightsResource/lido:rightsType/lido:concep
142      tID[@lido:type='http://terminology.lido-schema.org/identifier_
143      type/uri']/text()}" />
144      <!-- Verwertungsrecht and Urheberrecht -->
145      <xsl:for-each
146      select="lido:rightsResource">
147      <!-- RightHolder is always given in the rightsResource
148      with rightsType/term = "Verwertungsrecht". Normally a
149      URI is given, a literal is considered if no URI is found
150      -->
151      <xsl:if test="lido:rightsType/lido:term =
152      'Verwertungsrecht'">
153      <xsl:choose>
154      <xsl:when
155      test="lido:rightsHolder/lido:legalBodyID[@lido:typ
156      e='http://terminology.lido-schema.org/identifier_t
157      ype/uri']">
158      <dcterms:rightsHolder
159      rdf:resource="{lido:rightsHolder/lido:legalBod
160      yID[@lido:type='http://terminology.lido-schema
161      .org/identifier_type/uri']/text()}" />
162      </xsl:when>
163      <xsl:otherwise>
164      <dcterms:rightsHolder>
165      <!-- Blank note pdo:Agent -->
166      <rdf:Description>
167      <rdf:type
168      rdf:resource="http://purl.org/asch/pdo
169      /180704#Agent" />
170      <foaf:name>
171      <xsl:value-of
172      select="$var_record_rightsHolder_1
173      literal" />
174      </foaf:name>
175      </rdf:Description>
176      </dcterms:rightsHolder>
177      </xsl:otherwise>
178      </xsl:choose>
179      </xsl:if>
180      <!-- Creator of the photograph is always given in the

```

```

146         rightsResource with rightsType/term = "Urheberrecht".
147         Normally no URI is given but a literal -->
148         <xsl:if test="lido:rightsType/lido:term = 'Urheberrecht'">
149             <xsl:choose>
150                 <xsl:when
151                     test="lido:rightsHolder/lido:legalBodyID[@lido:typ
152                         e='http://terminology.lido-schema.org/identifier_t
153                         ype/uri']">
154                     <dc:creator
155                         rdf:resource="{lido:rightsHolder/lido:legalBod
156                             yID[@lido:type='http://terminology.lido-schema
157                             .org/identifier_type/uri']/text()}" />
158                 </xsl:when>
159                 <xsl:otherwise>
160                     <dc:creator>
161                         <xsl:value-of
162                             select="lido:rightsHolder/lido:legalBodyNa
163                                 me/lido:appellationValue/text()" />
164                     </dc:creator>
165                 </xsl:otherwise>
166             </xsl:choose>
167         </xsl:if>
168     </xsl:for-each>
169 </rdf:Description>
170 </xsl:if>
171 </xsl:for-each>
172 </rdf:RDF>
173 </xsl:template>
174
175 <!-- Objektgattung: each record belongs exactly to one Objektgattung; the fields
176 are not repeatable -->
177 <!-- ATTENTION: The machinereadable ontology for the Objektgattungen is still
178 missing (Todo Martin Stricker, Koordinierungsstelle) -->
179 <xsl:template
180     match="lido:descriptiveMetadata[1]/lido:objectClassificationWrap[1]/lido:objectWor
181     kTypeWrap[1]/lido:objectWorkType[1]/lido:conceptID[1]">
182     <!-- A falsely catalogued space e.g. at the beginning of the element is
183     deleted by the replace-function -->
184     <xsl:variable name="var_objektgattung" select="./text()" />
185     <rdf:type rdf:resource="{replace($var_objektgattung, ' ', '')}" />
186     <!-- Every Sammlungsobjekt is also an ASCH-Type pdo:Item -->
187     <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Item" />
188     <!--<rdf:type rdf:resource="{./text()}" />-->
189 </xsl:template>
190
191 <!--Item's classification. Only classifications with URI are taken into account
192 -->
193 <xsl:template
194     match="lido:descriptiveMetadata/lido:objectClassificationWrap/lido:classificationW
195     rap">
196     <xsl:for-each
197         select="./lido:classification/lido:conceptID[contains(@lido:type,
198             'terminology.lido-schema.org/identifier_type/uri')]">
199         <rdf:type rdf:resource="{./text()}" />
200     </xsl:for-each>
201 </xsl:template>
202
203 <!-- ObjectIdentificationWrap -->
204 <xsl:template match="lido:descriptiveMetadata/lido:objectIdentificationWrap">
205     <!-- Item's appearance (dimension and inscription) -->
206     <xsl:if test="(lido:inscriptionsWrap) or
207         (lido:objectMeasurementsWrap/lido:objectMeasurementsSet/lido:objectMeasurement
208         s/lido:measurementsSet/lido:measurementType
209         = ('count', 'height', 'width', 'depth', 'diamter', 'circumference',
210           'weight', 'size', 'volume' ))">
211         <ido:physicalDescription>
212             <!-- Blank note ido:Appearance-->
213             <rdf:Description>
214                 <rdf:type
215                     rdf:resource="http://purl.org/asch/ido/180704#Appearance" />
216                 <!-- objectMeasurementsSet is repeatable. Each

```

```

194         objectMeasurementsSet contains only one dimension -->
195         <xsl:for-each
196         select="lido:objectMeasurementsWrap/lido:objectMeasurementsSet/lido:objectMeasurements/lido:measurementsSet" >
197             <xsl:choose>
198                 <xsl:when test="lido:measurementType =
199                     'count'">
200                     <ido:numberOfParts>
201                         <xsl:value-of
202                         select="lido:measurementValue/text()" />
203                     </ido:numberOfParts>
204                 </xsl:when>
205                 <xsl:otherwise>
206                     <dcterms:extent>
207                         <!-- Blank Note crm:E54_Dimension -->
208                         <rdf:Description>
209                             <rdf:type
210                             rdf:resource="http://erlangen-crm.org/170309/E
211                                 54_Dimension" />
212                             <!-- measurement type -->
213                             <xsl:if test="lido:measurementType =
214                                 'height' " >
215                                 <dc:type
216                                 rdf:resource="http://d-nb.info/gnd/4229619
217                                     -5" />
218                             </xsl:if>
219                             <xsl:if test="lido:measurementType = 'depth'
220                                 " >
221                                 <dc:type
222                                 rdf:resource="http://d-nb.info/gnd/4334905
223                                     -5" />
224                             </xsl:if>
225                             <!-- For the following types no URI is found
226                             in d-nb; so the text is displayed as literal
227                             -->
228                             <xsl:if test="lido:measurementType =
229                                 ('width', 'size') " >
230                                 <dc:type xml:lang="en">
231                                     <xsl:value-of
232                                     select="lido:measurementType[@xml:lang
233                                         = 'en']/text()" />
234                                 </dc:type>
235                             </xsl:if>
236                             <xsl:if test="lido:measurementType =
237                                 'length' " >
238                                 <dc:type
239                                 rdf:resource="http://d-nb.info/gnd/4224141
240                                     -8" />
241                             </xsl:if>
242                             <xsl:if test="lido:measurementType =
243                                 'diameter' " >
244                                 <dc:type
245                                 rdf:resource="http://d-nb.info/gnd/4273067
246                                     -3" />
247                             </xsl:if>
248                             <xsl:if test="lido:measurementType =
249                                 'circumference' " >
250                                 <dc:type
251                                 rdf:resource="http://d-nb.info/gnd/7583293
252                                     -8" />
253                             </xsl:if>
254                             <xsl:if test="lido:measurementType =
255                                 'weight' " >
256                                 <dc:type
257                                 rdf:resource="http://d-nb.info/gnd/4157270
258                                     -1" />
259                             </xsl:if>
260                             <xsl:if test="lido:measurementType =
261                                 'volume' " >
262                                 <dc:type
263                                 rdf:resource="http://d-nb.info/gnd/4136953
264                                     -1" />

```



```

233         </xsl:if>
234         <!-- unit: kuniweb-files contain 2
235             unit-fields in German and English;
236             kenom-files contain just one without
237             given language-attribute -->
238         <xsl:choose>
239             <xsl:when
240                 test="lido:measurementUnit[@xml:lang =
241                     'en']">
242                 <crm:P91_has_unit xml:lang="en">
243                     <xsl:value-of
244                         select="lido:measurementUnit[@xml:
245                             lang = 'en']" />
246                 </crm:P91_has_unit>
247             </xsl:when>
248             <xsl:otherwise>
249                 <crm:P91_has_unit xml:lang="en">
250                     <xsl:value-of
251                         select="lido:measurementUnit" />
252                 </crm:P91_has_unit>
253             </xsl:otherwise>
254         </xsl:choose>
255
256         <!-- value -->
257         <crm:P90_has_value>
258             <xsl:value-of
259                 select="lido:measurementValue" />
260         </crm:P90_has_value>
261         <!-- note. As different languages can be
262             found in this field no language information
263             can be given -->
264         <xsl:if test="../lido:extentMeasurements">
265             <crm:P3_has_note>
266                 <xsl:value-of
267                     select="../lido:extentMeasurements" />
268             </crm:P3_has_note>
269         </xsl:if>
270     </rdf:Description>
271 </dcterms:extent>
272 </xsl:otherwise>
273 </xsl:choose>
274 </xsl:for-each>
275
276 <!-- Item's inscription -->
277 <!-- lido:inscriptions is repeatable.
278 lido:inscriptionTranscription is to be displayed -->
279 <xsl:for-each select="lido:inscriptionsWrap/lido:inscriptions">
280     <xsl:if test="lido:inscriptionTranscription">
281         <ido:inscription>
282             <xsl:value-of
283                 select="lido:inscriptionTranscription/text()" />
284         </ido:inscription>
285     </xsl:if>
286 </xsl:for-each>
287
288 </rdf:Description>
289 </ido:physicalDescription>
290 </xsl:if>
291
292 <!-- Item's title. Every title contains a sortOrder=1-attribute. Titleset
293 and appellationValue can be repeated -->
294 <xsl:if test="lido:titleWrap">
295     <ido:hasAppellation>
296         <rdf:Description>
297             <rdf:type
298                 rdf:resource="http://purl.org/asch/ido/180704#Appellation"
299             />
300             <xsl:for-each select="lido:titleWrap/lido:titleSet">
301                 <xsl:choose>
302                     <xsl:when test="@lido:sortorder = '1' ">
303                         <xsl:for-each select="lido:appellationValue">
304                             <skos:prefLabel xml:lang="{lido:language(.)}">

```



```

289         <xsl:value-of select="./text()" />
290     </skos:prefLabel>
291 </xsl:for-each>
292 </xsl:when>
293 <xsl:otherwise>
294     <xsl:for-each select="lido:appellationValue">
295         <skos:altLabel xml:lang="{lido:language(.)}">
296             <xsl:value-of select="./text()" />
297         </skos:altLabel>
298     </xsl:for-each>
299 </xsl:otherwise>
300 </xsl:choose>
301 </xsl:for-each>
302 </rdf:Description>
303 </ido:hasAppellation>
304 </xsl:if>
305
306 <!-- Item's repository -->
307 <!-- All fields are mandatory and not repeatable-->
308 <xsl:if test="lido:repositoryWrap/lido:repositorySet">
309     <dc:identifier>
310         <xsl:value-of
311             select="lido:repositoryWrap/lido:repositorySet/lido:workID/text()" />
312     </dc:identifier>
313     <ido:wasAffectedBy>
314         <!-- Creating a blank-note -->
315         <rdf:Description>
316             <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Custody" />
317             <xsl:variable name="var_repository"
318                 select="lido:repositoryWrap/lido:repositorySet/lido:repositoryName
319                     /lido:legalBodyID/text()" />
320             <marc:cur rdf:resource="{ $var_repository }" />
321             <pdo:tookPlaceAt>
322                 <!-- Creating a blank note -->
323                 <rdf:Description>
324                     <!-- In the LIDO-wrapper no place-URI is given, just
325                     lido:gml. Place-URI is not really necessary, as URI
326                     (legalBodyID) of the custody is given -->
327                     <rdf:type
328                         rdf:resource="http://purl.org/asch/pdo/180704#Place" />
329                     <wgs84_pos:lat
330                         rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
331                         <xsl:value-of
332                             select="substring-before(lido:repositoryWrap/lido:repo
333                                 sitorySet/lido:repositoryLocation/lido:gml/gml:Point/g
334                                 ml:pos/text(), ' ')" />
335                         </wgs84_pos:lat>
336                     <wgs84_pos:long
337                         rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
338                         <xsl:value-of
339                             select="substring-after(lido:repositoryWrap/lido:repo
340                                 sitorySet/lido:repositoryLocation/lido:gml/gml:Point/gm
341                                 l:pos/text(), ' ')" />
342                         </wgs84_pos:long>
343                     </rdf:Description>
344                 </pdo:tookPlaceAt>
345             </rdf:Description>
346         </ido:wasAffectedBy>
347     </xsl:if>
348
349 <!-- Item's description -->
350 <!-- objectDescriptionSet is repeatable -->
351 <xsl:if test="lido:objectDescriptionWrap">
352     <!-- Type description -->
353     <xsl:for-each
354         select="lido:objectDescriptionWrap/lido:objectDescriptionSet [@lido:type='d
355             escription']">
356         <dc:description
357             xml:lang="{lido:language(lido:objectDescriptionSet/lido:descriptiveNot
358                 eValue)}">
359             <!-- Many line feeds (\n = LF) are falsely catalogued and have

```

```

342         to be replaced by space; normalize-space-function also possible
343         -->
344         <xsl:variable name="var_descriptionNoteValue"
345         select="./lido:descriptiveNoteValue/text()" />
346         <xsl:value-of select="replace($var_descriptionNoteValue, '\n', '
347         ')" />
348         </dc:description>
349     </xsl:for-each>
350
351     <!-- Type usage -->
352     <!-- As the event "use" can only appear once, the following is only
353     considered, when there does not exist an use event in the eventWrap -->
354     <xsl:if
355     test="(lido:objectDescriptionWrap/lido:objectDescriptionSet[@lido:type='usage']
356     and not
357     (../lido:eventWrap/lido:eventSet/lido:event/lido:eventType/lido:conceptID = 'http://terminology.lido-schema.org/eventType/use'))" >
358     <ido:wasAffectedBy>
359         <rdf:Description>
360             <rdf:type
361             rdf:resource="http://terminology.lido-schema.org/eventType/use"
362             />
363             <xsl:for-each
364             select="lido:objectDescriptionWrap/lido:objectDescriptionSet[@lido:type='usage']">
365                 <pdo:hasPurpose
366                 xml:lang="{lido:language(lido:objectDescriptionSet/lido:descriptiveNoteValue)}">
367                     <xsl:value-of select="./lido:descriptiveNoteValue" />
368                 </pdo:hasPurpose>
369             </xsl:for-each>
370         </rdf:Description>
371     </ido:wasAffectedBy>
372 </xsl:if>
373
374     <!-- Type production_purpose -->
375     <!-- As the event "production" can only appear once, the following is
376     only considered, when there does not exist an production event in the
377     eventWrap -->
378     <xsl:if
379     test="(lido:objectDescriptionWrap/lido:objectDescriptionSet[@lido:type='production_purpose']
380     and not
381     (../lido:eventWrap/lido:eventSet/lido:event/lido:eventType/lido:conceptID = 'http://terminology.lido-schema.org/eventType/production'))" >
382     <ido:wasAffectedBy>
383         <rdf:Description>
384             <rdf:type
385             rdf:resource="http://terminology.lido-schema.org/eventType/production" />
386             <xsl:for-each
387             select="lido:objectDescriptionWrap/lido:objectDescriptionSet[@lido:type='production_purpose']">
388                 <pdo:hasPurpose
389                 xml:lang="{lido:language(lido:objectDescriptionSet/lido:descriptiveNoteValue)}">
390                     <xsl:value-of select="./lido:descriptiveNoteValue" />
391                 </pdo:hasPurpose>
392             </xsl:for-each>
393         </rdf:Description>
394     </ido:wasAffectedBy>
395 </xsl:if>
396 </xsl:if>
397
398 </xsl:template>
399
400 <!-- Item's events -->
401 <xsl:template match="lido:descriptiveMetadata/lido:eventWrap">
402     <xsl:for-each select="lido:eventSet">
403         <ido:wasAffectedBy>

```

```

387 <rdf:Description>
388 <!-- Event type -->
389 <rdf:type
390   rdf:resource="{lido:event/lido:eventType/lido:conceptID}" />
391
392 <!-- Event "production" and further information in
393   objectIdentificationWrap. Only one event "production" is
394   possible -->
395 <xsl:if test="(lido:event/lido:eventType/lido:conceptID =
396   'http://terminology.lido-schema.org/eventType/production'
397   and
398   /lido:lido/lido:descriptiveMetadata[1]/lido:objectIdentificati
399   onWrap[1]/lido:objectDescriptionWrap/lido:objectDescriptionSet
400   [@lido:type='production_purpose'])">
401
402   <xsl:for-each
403     select="/lido:lido/lido:descriptiveMetadata[1]/lido:objectIden
404     tificationWrap[1]/lido:objectDescriptionWrap/lido:objectDescri
405     ptionSet[@lido:type='production_purpose']">
406     <pdo:hasPurpose
407       xml:lang="{lido:language(lido:objectDescriptionSet/lido:de
408       scriptiveNoteValue)}">
409       <xsl:value-of select="."/lido:descriptiveNoteValue" />
410     </pdo:hasPurpose>
411   </xsl:for-each>
412 </xsl:if>
413
414 <!-- Event "use" and further information in
415   objectIdentificationWrap. Only one event "use" is possible -->
416 <xsl:if test="(lido:event/lido:eventType/lido:conceptID =
417   'http://terminology.lido-schema.org/eventType/use'
418   and
419   /lido:lido/lido:descriptiveMetadata[1]/lido:objectIdentificati
420   onWrap[1]/lido:objectDescriptionWrap/lido:objectDescriptionSet
421   [@lido:type='usage'])">
422
423   <xsl:for-each
424     select="/lido:lido/lido:descriptiveMetadata[1]/lido:objectIden
425     tificationWrap[1]/lido:objectDescriptionWrap/lido:objectDescri
426     ptionSet[@lido:type='usage']">
427     <pdo:hasPurpose
428       xml:lang="{lido:language(lido:objectDescriptionSet/lido:de
429       scriptiveNoteValue)}">
430       <xsl:value-of select="."/lido:descriptiveNoteValue" />
431     </pdo:hasPurpose>
432   </xsl:for-each>
433 </xsl:if>
434
435 <!-- Event actor -->
436 <xsl:for-each select="lido:event/lido:eventActor">
437   <xsl:if test="lido:actorInRole/lido:actor/lido:actorID">
438     <xsl:for-each select="lido:actorInRole/lido:roleActor">
439       <xsl:choose>
440         <xsl:when test="lido:conceptID[contains(text(),
441         'id.loc.gov/vocabulary/relators/')] ">
442           <xsl:variable name="var_marcrel_property"
443             select="lido:conceptID[contains(text(),
444             'id.loc.gov/vocabulary/relators/')] /text()" />
445           <xsl:variable name="var_elementPart_marcrel"
446             select="substring-after($var_marcrel_property,
447             'id.loc.gov/vocabulary/relators/')" />
448           <xsl:element
449             name="marcrel:{$var_elementPart_marcrel}">
450
451             <xsl:attribute name="rdf:resource">
452               <xsl:value-of
453                 select="../lido:actor/lido:actorID" />
454             </xsl:attribute>
455           </xsl:element>
456         </xsl:when>
457         <xsl:otherwise>
458           <pdo:participant

```



```

429         rdf:resource="{../lido:actor/lido:actorID}"
430     />
431     </xsl:otherwise>
432 </xsl:choose>
433 </xsl:for-each>
434 </xsl:if>
435 </xsl:for-each>
436
437 <!-- Event date -->
438 <!-- Only one start date and one end date are possible in one
439 eventSet -->
440 <xsl:if
441 test="lido:event/lido:eventDate/lido:date/lido:earliestDate">
442     <pdo:start>
443         <!-- Blank note start date -->
444         <rdf:Description>
445             <rdf:type
446                 rdf:resource="http://purl.org/asch/pdo/180704#Time" />
447             <rdf:value
448                 rdf:datatype="http://purl.org/dc/terms/W3CDTF">
449                 <xsl:value-of
450                     select="lido:event/lido:eventDate/lido:date/lido:e
451                         arliestDate" />
452             </rdf:value>
453             <xsl:if
454                 test="lido:event/lido:eventDate/lido:date/lido:earlies
455                     tDate[@lido:type]" >
456                 <pdo:qualifier xml:lang = "en">
457                     <xsl:value-of
458                         select="lido:event/lido:eventDate/lido:date/li
459                             do:earliestDate/@lido:type" />
460                 </pdo:qualifier>
461             </xsl:if>
462         </rdf:Description>
463     </pdo:start>
464 </xsl:if>
465 <xsl:if
466 test="lido:event/lido:eventDate/lido:date/lido:latestDate">
467     <pdo:end>
468         <!-- Blank note end date -->
469         <rdf:Description>
470             <rdf:type
471                 rdf:resource="http://purl.org/asch/pdo/180704#Time" />
472             <rdf:value
473                 rdf:datatype="http://purl.org/dc/terms/W3CDTF">
474                 <xsl:value-of
475                     select="lido:event/lido:eventDate/lido:date/lido:l
476                         atestDate/text()" />
477             </rdf:value>
478             <xsl:if
479                 test="lido:event/lido:eventDate/lido:date/lido:earlies
480                     tDate[@lido:type]" >
481                 <pdo:qualifier xml:lang = "en">
482                     <xsl:value-of
483                         select="lido:event/lido:eventDate/lido:date/li
484                             do:latestDate/@lido:type" />
485                 </pdo:qualifier>
486             </xsl:if>
487         </rdf:Description>
488     </pdo:end>
489 </xsl:if>
490
491 <!-- Event place -->
492 <!-- 0-n eventPlaces are possible -->
493 <xsl:for-each select="lido:event/lido:eventPlace">
494     <!-- If there is a place-URI, gml:pos is not considered -->
495     <xsl:choose>
496         <xsl:when test="lido:place/lido:placeID[@lido:type =
497             'http://terminology.lido-schema.org/identifier_type/uri']"
498             >
499             <pdo:tookPlaceAt

```

```

478         rdf:resource="{lido:place/lido:placeID}"
479     />
480 </xsl:when>
481 <xsl:otherwise>
482     <xsl:if test="lido:place/lido:gml">
483         <pdo:tookPlaceAt>
484             <!-- Creating a blank note -->
485             <rdf:Description>
486                 <rdf:type
487                     rdf:resource="http://purl.org/asch/pdo/180
488                     704#Place" />
489                 <wgs84_pos:lat
490                     rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
491                     <xsl:value-of
492                         select="substring-before(lido:place/li
493                         do:gml/gml:Point/gml:pos/text(), '
494                         ')" />
495                 </wgs84_pos:lat>
496                 <wgs84_pos:long
497                     rdf:datatype="http://www.w3.org/2001/XMLSchema#float">
498                     <xsl:value-of
499                         select="substring-after(lido:place/li
500                         do:gml/gml:Point/gml:pos/text(), '
501                         ')" />
502                 </wgs84_pos:long>
503             </rdf:Description>
504         </pdo:tookPlaceAt>
505     </xsl:if>
506 </xsl:otherwise>
507 </xsl:choose>
508 </xsl:for-each>
509
510 <!-- Event culture -->
511 <xsl:for-each select="lido:event/lido:culture">
512     <xsl:if test="lido:conceptID[@lido:type =
513         'http://terminology.lido-schema.org/identifier_type/uri']">
514         <pdo:culturalContext rdf:resource="{lido:conceptID}" />
515     </xsl:if>
516 </xsl:for-each>
517
518 <!-- Event material, technique and color -->
519 <!-- AAT-URI is first choice, second choice is gnd-URI and third
choice the first term -->
520 <xsl:for-each select="lido:event/lido:eventMaterialsTech">
521     <!-- material -->
522     <xsl:if
523         test="lido:materialsTech/lido:termMaterialsTech[@lido:type =
524         'http://terminology.lido-schema.org/termMaterialsTech_type/mat
525         erial']">
526         <xsl:choose>
527             <xsl:when
528                 test="lido:materialsTech/lido:termMaterialsTech/lido:c
529                 onceptID">
530                 <xsl:choose>
531                     <xsl:when
532                         test="lido:materialsTech/lido:termMaterialsTec
533                         h/lido:conceptID[contains(text(),
534                         'vocab.getty.edu/aat')]">
535                         <pdo:material
536                             rdf:resource="{lido:materialsTech/lido:ter
537                             mMaterialsTech/lido:conceptID[contains(tex
538                             t(), 'vocab.getty.edu/aat')]}" />
539                     </xsl:when>
540                     <xsl:otherwise>
541                         <pdo:material
542                             rdf:resource="{lido:materialsTech/lido:ter
543                             mMaterialsTech/lido:conceptID[contains(tex
544                             t(), 'd-nb.info/gnd')]}" />
545                     </xsl:otherwise>
546                 </xsl:choose>
547             </xsl:if>
548         </xsl:choose>
549     </xsl:if>
550 </xsl:for-each>

```

```

520         </xsl:when>
521         <xsl:otherwise>
522             <pdo:material
523                 xml:lang="{lido:language(lido:materialsTech/lido:termMaterialsTech/lido:term)}">
524                 <xsl:value-of
525                     select="lido:materialsTech/lido:termMaterialsTech/lido:term[1]/text()" />
526             </pdo:material>
527         </xsl:otherwise>
528     </xsl:choose>
529 </xsl:if>
530
531 <!-- technique -->
532 <xsl:if
533     test="lido:materialsTech/lido:termMaterialsTech[@lido:type =
534         'http://terminology.lido-schema.org/termMaterialsTech_type/technique']">
535     <xsl:choose>
536         <xsl:when
537             test="lido:materialsTech/lido:termMaterialsTech/lido:conceptID">
538             <xsl:choose>
539                 <xsl:when
540                     test="lido:materialsTech/lido:termMaterialsTech/lido:conceptID[contains(text(),
541                         'vocab.getty.edu/aat')]">
542                     <pdo:technique
543                         rdf:resource="{lido:materialsTech/lido:termMaterialsTech/lido:conceptID[contains(text(),
544                             'vocab.getty.edu/aat')]}" />
545                     </xsl:when>
546                     <xsl:otherwise>
547                         <pdo:technique
548                             rdf:resource="{lido:materialsTech/lido:termMaterialsTech/lido:conceptID[contains(text(),
549                                 'd-nb.info/gnd')]}" />
550                         </xsl:otherwise>
551                     </xsl:choose>
552                 </xsl:when>
553                 <xsl:otherwise>
554                     <pdo:technique
555                         xml:lang="{lido:language(lido:materialsTech/lido:termMaterialsTech/lido:term)}">
556                         <xsl:value-of
557                             select="lido:materialsTech/lido:termMaterialsTech/lido:term[1]/text()" />
558                     </pdo:technique>
559                 </xsl:otherwise>
560             </xsl:choose>
561         </xsl:if>
562
563 <!-- colour -->
564 <xsl:for-each
565     select="lido:materialsTech/lido:termMaterialsTech[@lido:type =
566         'http://terminology.lido-schema.org/lido00479']/lido:term[xml:lang='en']">
567     <ido:colour xml:lang="{lido:language(.)}">
568         <xsl:value-of select="." />
569     </ido:colour>
570 </xsl:for-each>
571 </xsl:for-each>
572
573 </rdf:Description>
574 </ido:wasAffectedBy>
575 </xsl:for-each>
576 </xsl:template>

```



```

562      <!-- End Event -->
563
564
565      <!--lido:objectRelationWrap -->
566      <xsl:template match="lido:descriptiveMetadata/lido:objectRelationWrap">
567
568          <!--Item's subject. Only subjects with URI are taken into account -->
569          <!-- URIs in lido:subjectSet is repeatable; subjectConcept, subjectActor and
570          subjectPlace are taken into account -->
571          <xsl:if test="lido:subjectWrap">
572              <xsl:for-each select="lido:subjectWrap/lido:subjectSet">
573                  <xsl:if test="lido:subject/lido:subjectConcept/lido:conceptID">
574                      <dc:subject
575                          rdf:resource="{lido:subject/lido:subjectConcept/lido:conceptID}"
576                      />
577                  </xsl:if>
578                  <xsl:if test="lido:subject/lido:subjectActor/lido:actor/lido:actorID">
579                      <dc:subject
580                          rdf:resource="{lido:subject/lido:subjectActor/lido:actor/lido:actorID}"
581                      />
582                  </xsl:if>
583                  <xsl:if test="lido:subject/lido:subjectPlace/lido:place/lido:placeID[@lido:type = 'http://terminology.lido-schema.org/identifier_type/uri']">
584                      <dc:subject
585                          rdf:resource="{lido:subject/lido:subjectPlace/lido:place/lido:placeID}"
586                      />
587                  </xsl:if>
588              </xsl:for-each>
589          </xsl:if>
590
591          <!-- Item's related works. 4 different cases are considered -->
592          <!-- lido:relatedWorkSet is repeatable -->
593          <xsl:if test="lido:relatedWorksWrap">
594              <xsl:for-each select="lido:relatedWorksWrap/lido:relatedWorkSet">
595                  <xsl:choose>
596                      <!-- First case: "ist Teil von" -->
597                      <xsl:when test="(lido:relatedWorkRelType[lido:term = 'ist Teil von'])
598                          and
599                          (lido:relatedWork/lido:object/lido:objectWebResource)"
600                          >
601                          <!-- Falsely catalogued space in the URL are deleted by the
602                          replace-function -->
603                          <xsl:variable name="var_objectWebResource"
604                              select="lido:relatedWork/lido:object/lido:objectWebResource/text()" />
605                          <dc:terms:isPartOf
606                              rdf:resource="{replace($var_objectWebResource, ' ', '')}" />
607                      </xsl:when>
608                      <xsl:otherwise>
609                          <!-- Second case: other related items and weblinks -->
610                          <xsl:if
611                              test="(lido:relatedWork/lido:object/lido:objectWebResource)
612                              and not
613                              (lido:relatedWork/lido:object/lido:objectID[@lido:type='http://terminology.lido-schema.org/identifier_type/uri'])">
614                              <!-- Falsely catalogued space in the URL are deleted by
615                              the replace-function -->
616                              <xsl:variable name="var_objectWebResource"
617                                  select="lido:relatedWork/lido:object/lido:objectWebResource/text()" />
618                              <dc:relation
619                                  rdf:resource="{replace($var_objectWebResource, ' ', '')}" />
620                          </xsl:if>
621                      </xsl:otherwise>
622                  </xsl:choose>
623                  <!-- Third case: Links to GVK-Titles -->
624                  <xsl:if
625                      test="lido:relatedWork/lido:object/lido:objectID[@lido:type='http://terminology.lido-schema.org/identifier_type/uri']">

```



```

608         <dc:terms:isReferencedBy
609             rdf:resource="{lido:relatedWork/lido:object/lido:objectID}" />
610     </xsl:if>
611     <!-- Fourth case: Link to the Collection -->
612     <xsl:if
613         test="(lido:relatedWork/lido:object/lido:objectID[@lido:type=('http://
614             terminology.lido-schema.org/identifier_type/local_identifier',
615             'http://terminology.lido-schema.org/lido00100') ])"
616         and
617         (lido:relatedWork/lido:object/lido:objectID[@lido:source='Digitale
618             Sammlung - Kulturerbe Niedersachsen'])">
619         <ido:isGatheredInto>
620             <!-- Blank node collection -->
621             <rdf:Description>
622                 <rdf:type
623                     rdf:resource="http://purl.org/asch/ido/180704#Collection"
624                 />
625                 <dc:identifier>
626                     <xsl:value-of
627                         select="lido:relatedWork/lido:object/lido:objectID" />
628                     </dc:identifier>
629             </rdf:Description>
630         </ido:isGatheredInto>
631     </xsl:if>
632 </xsl:for-each>
633 </xsl:if>
634 </xsl:template>
635
636 <!--Item's digital representation. Actually only photographs are considered in
637 the LIDO data -->
638 <!-- lido:resourceSet is repeatable; each resourceSet contains one
639 resourceRepresentation with attribute ...provided_representation (better picture
640 quality) -->
641 <xsl:template match="lido:administrativeMetadata">
642     <xsl:if test="lido:resourceWrap">
643         <xsl:for-each select="lido:resourceWrap/lido:resourceSet">
644             <xsl:if
645                 test="lido:resourceRepresentation[@lido:type='http://terminology.lido-
646                     schema.org/resourceRepresentation_type/provided_representation']/lido:
647                     linkResource[@lido:formatResource='image/jpeg']" >
648                 <ido:hasDigitalRepresentation
649                     rdf:resource="{lido:resourceRepresentation[@lido:type='http://term
650                         inology.lido-schema.org/resourceRepresentation_type/provided_repre
651                         sentation']/lido:linkResource[@lido:formatResource='image/jpeg']/t
652                         ext()}" />
653             </xsl:if>
654         </xsl:for-each>
655     </xsl:if>
656 </xsl:template>
657
658 <!-- Function which sets the language-attribute @xml:lang. It is set 'de', if no
659 language is given, otherwise the value is displayed -->
660 <xsl:function name="lido:language">
661     <xsl:param name="node" />
662     <xsl:value-of select="if ( not ($node/@xml:lang) ) then ('de') else (
663         $node/@xml:lang )" />
664 </xsl:function>
665
666 <xsl:template match="text()" />
667
668 </xsl:stylesheet>

```

ANHANG 3: Datensatz record_kuniweb_929301 in LIDO-XML

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <lido:lido xmlns:lido="http://www.lido-schema.org" xmlns:gml="
  "http://www.opengis.net/gml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.lido-schema.org
  http://www.lido-schema.org/schema/v1.0/lido-v1.0.xsd">
3   <lido:lidoRecID lido:type=
    "http://terminology.lido-schema.org/identifier_type/local_identifier">
    record_kuniweb_929301</lido:lidoRecID>
4   <lido:objectPublishedID lido:type=
    "http://terminology.lido-schema.org/identifier_type/uri">
    http://sammlungen.uni-goettingen.de/objekt/record_kuniweb_929301
    </lido:objectPublishedID>
5   <lido:objectPublishedID lido:type=
    "http://terminology.lido-schema.org/identifier_type/uri">
    https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022
    </lido:objectPublishedID>
6   <lido:category>
7     <lido:conceptID lido:type=
      "http://terminology.lido-schema.org/identifier_type/uri">
      http://www.cidoc-crm.org/crm-concepts/#E22</lido:conceptID>
8     <lido:term xml:lang="en">Man-Made Object</lido:term>
9   </lido:category>
10  <lido:descriptiveMetadata xml:lang="de">
11    <lido:objectClassificationWrap>
12      <lido:objectWorkTypeWrap>
13        <lido:objectWorkType>
14          <lido:conceptID lido:type="URI">
            http://portal.wissenschaftliche-sammlungen.de/ObjectGenre/244
          </lido:conceptID>
15          <lido:term>Kleidung und Accessoire</lido:term>
16        </lido:objectWorkType>
17      </lido:objectWorkTypeWrap>
18    <lido:classificationWrap>
19      <lido:classification lido:type="Hessische Systematik">
20        <lido:term>Kleidung &gt; Frauenkleidung &gt; Fußbekleidung
          (Frauenkleidung)</lido:term>
21      </lido:classification>
22      <lido:classification lido:type="Hessische Systematik">
23        <lido:term>Kleidung &gt; Regionaltypische Kleidung &gt; Fußbekleidung
          (Regionaltypische Kleidung)</lido:term>
24      </lido:classification>
25      <lido:classification lido:type="Oberbegriffsdatei">
26        <lido:conceptID lido:type=
          "http://terminology.lido-schema.org/identifier_type/uri">
            http://obg.vocnet.org/00002113</lido:conceptID>
27        <lido:term>Kleidung* &gt; Schuh &gt; Pantoffel</lido:term>
28      </lido:classification>
29    </lido:classificationWrap>
30  </lido:objectClassificationWrap>
31  <lido:objectIdentificationWrap>
32    <lido:titleWrap>
33      <lido:titleSet lido:sortorder="1">
34        <lido:appellationValue>Pantoffeln aus Marokko</lido:appellationValue>
35      </lido:titleSet>
36      <lido:titleSet lido:type="other">
37        <lido:appellationValue xml:lang="en">slippers</lido:appellationValue>
38      </lido:titleSet>
39    </lido:titleWrap>
40    <lido:repositoryWrap>
41      <lido:repositorySet lido:type=
        "http://terminology.lido-schema.org/repositorySet_type/current_repository_
        or_location">
42        <lido:repositoryName>
43          <lido:legalBodyID lido:type=
            "http://terminology.lido-schema.org/identifier_type/uri">
              http://ld.zdb-services.de/resource/organisations/DE-MUS-056211
            </lido:legalBodyID>
44          <lido:legalBodyName>
45            <lido:appellationValue>Georg-August-Universität Göttingen /
              Ethnologische Sammlung der Universität</lido:appellationValue>
46          </lido:legalBodyName>
47          <lido:legalBodyWeblink>
```

```

48         http://www.uni-goettingen.de/de/sammlung/28899.html
49     </lido:repositoryName>
50     <lido:workID>Af 423 a-b</lido:workID>
51     <lido:repositoryLocation>
52         <lido:namePlaceSet>
53             <lido:appellationValue>Ethnologische Sammlung der Universität
54             </lido:appellationValue>
55         </lido:namePlaceSet>
56         <lido:gml>
57             <gml:Point>
58                 <gml:pos>51.53512 9.94042</gml:pos>
59             </gml:Point>
60         </lido:gml>
61         <lido:partOfPlace>
62             <lido:placeID lido:type=
63                 "http://terminology.lido-schema.org/identifizier_type/uri">
64                 http://sws.geonames.org/2918632/</lido:placeID>
65             <lido:namePlaceSet>
66                 <lido:appellationValue>Göttingen</lido:appellationValue>
67             </lido:namePlaceSet>
68             <lido:gml>
69                 <gml:Point>
70                     <gml:pos>51.53443 9.93228</gml:pos>
71                 </gml:Point>
72             </lido:gml>
73         </lido:partOfPlace>
74     </lido:repositoryLocation>
75 </lido:repositorySet>
76 </lido:repositoryWrap>
77 <lido:objectDescriptionWrap>
78     <lido:objectDescriptionSet lido:type="usage">
79         <lido:descriptiveNoteValue>Fußbekleidung für Frauen
80         </lido:descriptiveNoteValue>
81     </lido:objectDescriptionSet>
82     <lido:objectDescriptionSet lido:type="description">
83         <lido:descriptiveNoteValue>Bei diesen Schuhen handelt es sich um ein
84         Paar offene Slipper für Frauen.
85     </lido:descriptiveNoteValue>
86     Die Pantoffeln sind aus zwei Teilen gefertigt worden. Die Schuhkappe besteht, wie
87     die Sohle, aus naturbelassenem, gelben Leder. Hier finden sich jedoch mit gold- und
88     rosafarbenem Garn bestickte Flächen als auch ein zentrales Blumenornament auf dem
89     Schaft des Schuhs. Des Weiteren wurde eine blaue Quaste auf jedem Pantoffel
90     appliziert. Die Sohle ist mit rotem Samt in der Lauffläche gepolstert.
91 </lido:descriptiveNoteValue>
92     <lido:sourceDescriptiveNote>Ethnologische Sammlung
93     </lido:sourceDescriptiveNote>
94 </lido:objectDescriptionSet>
95 </lido:objectDescriptionWrap>
96 <lido:objectMeasurementsWrap>
97     <lido:objectMeasurementsSet>
98         <lido:displayObjectMeasurements>Breite: 10 cm
99         </lido:displayObjectMeasurements>
100     <lido:objectMeasurements>
101         <lido:measurementsSet>
102             <lido:measurementType xml:lang="en">width</lido:measurementType>
103             <lido:measurementType>Breite</lido:measurementType>
104             <lido:measurementUnit xml:lang="en">cm</lido:measurementUnit>
105             <lido:measurementUnit>cm</lido:measurementUnit>
106             <lido:measurementValue>10</lido:measurementValue>
107         </lido:measurementsSet>
108     </lido:objectMeasurements>
109 </lido:objectMeasurementsSet>
110 </lido:objectMeasurementsSet>
111 <lido:objectMeasurementsSet>
112     <lido:displayObjectMeasurements>Länge: 23 cm
113     </lido:displayObjectMeasurements>
114     <lido:objectMeasurements>
115         <lido:measurementsSet>
116             <lido:measurementType xml:lang="en">length</lido:measurementType>
117             <lido:measurementType>Länge</lido:measurementType>
118             <lido:measurementUnit xml:lang="en">cm</lido:measurementUnit>

```



```

106         <lido:measurementUnit>cm</lido:measurementUnit>
107         <lido:measurementValue>23</lido:measurementValue>
108     </lido:measurementsSet>
109 </lido:objectMeasurements>
110 </lido:objectMeasurementsSet>
111 </lido:objectMeasurementsWrap>
112 </lido:objectIdentificationWrap>
113 <lido:eventWrap>
114     <lido:eventSet>
115         <lido:event>
116             <lido:eventType>
117                 <lido:conceptID lido:type=
118                     "http://terminology.lido-schema.org/identifier_type/uri">
119                     http://terminology.lido-schema.org/eventType/collecting
120                 </lido:conceptID>
121                 <lido:term xml:lang="en">Collecting</lido:term>
122                 <lido:term>Sammelereignis</lido:term>
123             </lido:eventType>
124             <lido:eventActor>
125                 <lido:displayActorInRole>Asch, Georg Thomas von (Sammler)
126                 </lido:displayActorInRole>
127                 <lido:actorInRole>
128                     <lido:actor lido:type=
129                         "http://terminology.lido-schema.org/lido00163">
130                         <lido:actorID lido:type=
131                             "http://terminology.lido-schema.org/identifier_type/uri">
132                             http://d-nb.info/gnd/118650599</lido:actorID>
133                         <lido:nameActorSet>
134                             <lido:appellationValue>Asch, Georg Thomas von
135                             </lido:appellationValue>
136                         </lido:nameActorSet>
137                     </lido:actor>
138                     <lido:roleActor>
139                         <lido:conceptID lido:type=
140                             "http://terminology.lido-schema.org/identifier_type/uri">
141                             http://vocab.getty.edu/aat/300025234</lido:conceptID>
142                         <lido:conceptID lido:type=
143                             "http://terminology.lido-schema.org/identifier_type/uri">
144                             http://d-nb.info/gnd/4179037-6</lido:conceptID>
145                         <lido:conceptID lido:type=
146                             "http://terminology.lido-schema.org/identifier_type/uri">
147                             http://id.loc.gov/vocabulary/relators/col</lido:conceptID>
148                         <lido:term>Sammler</lido:term>
149                     </lido:roleActor>
150                 </lido:actorInRole>
151             </lido:eventActor>
152         </lido:event>
153     </lido:eventSet>
154 </lido:eventSet>
155 <lido:event>
156     <lido:eventType>
157         <lido:conceptID lido:type=
158             "http://terminology.lido-schema.org/identifier_type/uri">
159             http://terminology.lido-schema.org/eventType/production
160         </lido:conceptID>
161         <lido:term xml:lang="en">Production</lido:term>
162         <lido:term>Herstellung</lido:term>
163     </lido:eventType>
164     <lido:eventDate>
165         <lido:displayDate>18. Jahrhundert</lido:displayDate>
166         <lido:date>
167             <lido:earliestDate lido:type="approximate">1701
168             </lido:earliestDate>
169             <lido:latestDate lido:type="approximate">1800</lido:latestDate>
170         </lido:date>
171     </lido:eventDate>
172     <lido:eventMaterialsTech>
173         <lido:displayMaterialsTech>Leder (gelb)</lido:displayMaterialsTech>
174         <lido:materialsTech>
175             <lido:termMaterialsTech lido:type=
176                 "http://terminology.lido-schema.org/termMaterialsTech_type/material">

```

```
158         <lido:conceptID lido:type=  
159             "http://terminology.lido-schema.org/identifier_type/uri">  
160             http://vocab.getty.edu/aat/300011845</lido:conceptID>  
161         <lido:conceptID lido:type=  
162             "http://terminology.lido-schema.org/identifier_type/uri">  
163             http://d-nb.info/gnd/4034974-3</lido:conceptID>  
164         <lido:term>Leder</lido:term>  
165     </lido:termMaterialsTech>  
166     <lido:extentMaterialsTech>gelb</lido:extentMaterialsTech>  
167 </lido:materialsTech>  
168 </lido:eventMaterialsTech>  
169 <lido:displayMaterialsTech>Textiltechnik &gt; gestickt (gold,  
170     rosa, blau)</lido:displayMaterialsTech>  
171 <lido:materialsTech>  
172     <lido:termMaterialsTech lido:type=  
173         "http://terminology.lido-schema.org/termMaterialsTech_type/techni  
174         que">  
175         <lido:term>gestickt</lido:term>  
176         <lido:term lido:addedSearchTerm="yes">Textiltechnik  
177         </lido:term>  
178     </lido:termMaterialsTech>  
179     <lido:extentMaterialsTech>gold, rosa, blau  
180     </lido:extentMaterialsTech>  
181 </lido:materialsTech>  
182 </lido:eventMaterialsTech>  
183 </lido:event>  
184 </lido:eventSet>  
185 <lido:eventSet>  
186 <lido:event>  
187     <lido:eventType>  
188         <lido:conceptID lido:type=  
189             "http://terminology.lido-schema.org/identifier_type/uri">  
190             http://terminology.lido-schema.org/eventType/use</lido:conceptID>  
191         <lido:term xml:lang="en">Use</lido:term>  
192         <lido:term>Gebrauch</lido:term>  
193     </lido:eventType>  
194     <lido:eventPlace>  
195         <lido:displayPlace>Afrika</lido:displayPlace>  
196         <lido:place>  
197             <lido:placeID lido:type=  
198                 "http://terminology.lido-schema.org/identifier_type/uri">  
199                 http://sws.geonames.org/6255146/</lido:placeID>  
200             <lido:namePlaceSet>  
201                 <lido:appellationValue xml:lang="en" lido:pref="alternate">  
202                     Africa</lido:appellationValue>  
203                 <lido:appellationValue lido:pref="preferred">Afrika  
204                 </lido:appellationValue>  
205                 <lido:sourceAppellation>Geonames</lido:sourceAppellation>  
206             </lido:namePlaceSet>  
207             <lido:gml>  
208                 <gml:Point>  
209                     <gml:pos>7.1881 21.09375</gml:pos>  
210                 </gml:Point>  
211             </lido:gml>  
212         </lido:place>  
213     </lido:eventPlace>  
214 </lido:eventPlace>  
215 <lido:displayPlace>Marokko</lido:displayPlace>  
216 <lido:place>  
217     <lido:placeID lido:type=  
218         "http://terminology.lido-schema.org/identifier_type/uri">  
219         http://sws.geonames.org/2542007/</lido:placeID>  
220     <lido:namePlaceSet>  
221         <lido:appellationValue xml:lang="en" lido:pref="alternate">  
222             Morocco</lido:appellationValue>  
223         <lido:appellationValue lido:pref="preferred">Marokko  
224         </lido:appellationValue>  
225         <lido:sourceAppellation>Geonames</lido:sourceAppellation>  
226     </lido:namePlaceSet>  
227     <lido:gml>  
228         <gml:Point>
```

```

211         <gml:pos>32 -6</gml:pos>
212     </gml:Point>
213 </lido:gml>
214 </lido:place>
215 </lido:eventPlace>
216 </lido:event>
217 </lido:eventSet>
218 </lido:eventWrap>
219 <lido:objectRelationWrap>
220 <lido:relatedWorksWrap>
221 <lido:relatedWorkSet>
222 <lido:relatedWork>
223     <lido:displayObject>„Siberia and Russian America : culture and art
        from the 1700s ; the Asch collection, Göttingen = Sibirien und
        Russisch-Amerika : Kultur und Kunst des 18. Jahrhunderts ; die
        Sammlung von Asch, Göttingen. Sibirien und Russisch-Amerika :
        Kultur und Kunst des 18. Jahrhunderts ; die Sammlung von Asch,
        Göttingen“. Prestel, München [u.a.], 2007. (S. 273)
    </lido:displayObject>
224 <lido:object>
225     <lido:objectWebResource lido:formatResource="text/html">
        http://gso.gbv.de/DB=2.1/PPNSET?PPN=532043979
    </lido:objectWebResource>
226     <lido:objectID lido:type=
        "http://terminology.lido-schema.org/identifier_type/uri">
        uri.gbv.de/document/gvk:ppn:532043979</lido:objectID>
227     <lido:objectNote lido:type="description">„Siberia and Russian
        America : culture and art from the 1700s ; the Asch collection,
        Göttingen = Sibirien und Russisch-Amerika : Kultur und Kunst
        des 18. Jahrhunderts ; die Sammlung von Asch, Göttingen.
        Sibirien und Russisch-Amerika : Kultur und Kunst des 18.
        Jahrhunderts ; die Sammlung von Asch, Göttingen“. Prestel,
        München [u.a.], 2007.</lido:objectNote>
228 </lido:object>
229 </lido:relatedWork>
230 <lido:relatedWorkRelType>
231     <lido:term>Beschrieben in</lido:term>
232 </lido:relatedWorkRelType>
233 </lido:relatedWorkSet>
234 </lido:relatedWorksWrap>
235 <lido:relatedWork>
236     <lido:displayObject>Academisches Museum</lido:displayObject>
237 <lido:object>
238     <lido:objectID lido:type=
        "http://terminology.lido-schema.org/identifier_type/local_identif
        ier" lido:source="Digitale Sammlung - Kulturerbe Niedersachsen">
        slg_1047</lido:objectID>
239     <lido:objectNote lido:type="title">Academisches Museum
    </lido:objectNote>
240     <lido:objectNote lido:type="description">Das Akademische Museum
        der Universität Göttingen geht auf den Ankauf der
        Privatsammlung des Gelehrten Christian Wilhelm Büttner
        (1716-1801) durch die Georgia Augusta im Jahr 1773 zurück.
        Bereits bestehende Planungen zum Aufbau eines
        Universitätsmuseums bekamen damit den entscheidenden Impuls für
        einen rasanten Wachstumsprozess der Sammlungen aus allen drei
        Naturreichen sowie Ethnographica, Kunstwerke und Instrumente.
        Für das Jahr 1840 kann der Gesamtbestand auf ca. 12.000 Objekte
        geschätzt werden. Die Amtszeit des ersten Inventarisierers und
        späteren Oberaufsehers des Museums, Johann Friedrich Blumenbach
        (1752-1840) gilt als wesentliche Blüte- und
        Konsolidierungsphase des Museums. Nach dessen Tod und bis zur
        Eröffnung des neuen Naturhistorischen Museums der Universität
        1878 begann bereits die Ausdifferenzierung der akademischen
        Sammlungen in vier Abteilungen: die mineralogisch-geologische
        unter Johann Friedrich Hausmann (1782-1859); die zoologische
        unter Arnold Adolf Berthold (1803-1861); die
        anthropologisch-zootomische unter Rudolf Wagner (1805-1864);
        die ethnographische unter Johann Friedrich Osiander (1787-1855).
    </lido:objectNote>
241 </lido:object>
242 </lido:relatedWork>

```



```

243         <lido:relatedWorkRelType>
244             <lido:term xml:lang="en">collection</lido:term>
245             <lido:term>Sammlung</lido:term>
246         </lido:relatedWorkRelType>
247     </lido:relatedWorkSet>
248     <lido:relatedWorkSet>
249         <lido:relatedWork>
250             <lido:displayObject>Ethnologische Sammlung</lido:displayObject>
251             <lido:object>
252                 <lido:objectID lido:type=
253                     "http://terminology.lido-schema.org/identifizier_type/local_identifizier" lido:source="Digitale Sammlung - Kulturerbe Niedersachsen">
254                     slg_1021</lido:objectID>
255                 <lido:objectNote lido:type="title">Ethnologische Sammlung</lido:objectNote>
256             </lido:object>
257         </lido:relatedWork>
258         <lido:relatedWorkRelType>
259             <lido:term xml:lang="en">collection</lido:term>
260             <lido:term>Sammlung</lido:term>
261         </lido:relatedWorkRelType>
262     </lido:relatedWorkSet>
263 </lido:descriptiveMetadata>
264 <lido:administrativeMetadata xml:lang="de">
265     <lido:recordWrap>
266         <lido:recordID lido:source="kuniweb" lido:type=
267             "http://terminology.lido-schema.org/identifizier_type/local_identifizier">929301</lido:recordID>
268         <lido:recordType>
269             <lido:conceptID lido:type=
270                 "http://terminology.lido-schema.org/identifizier_type/uri">
271                 http://terminology.lido-schema.org/recordType/item-level_record</lido:conceptID>
272             <lido:term>Einzelobjekt</lido:term>
273             <lido:term xml:lang="en">item</lido:term>
274         </lido:recordType>
275         <lido:recordSource>
276             <lido:legalBodyName>
277                 <lido:appellationValue>kuniweb</lido:appellationValue>
278             </lido:legalBodyName>
279             <lido:legalBodyWebLink>https://kuniweb.gbv.de</lido:legalBodyWebLink>
280         </lido:recordSource>
281         <lido:recordRights>
282             <lido:rightsType>
283                 <lido:term>Verwertungsrecht</lido:term>
284             </lido:rightsType>
285             <lido:rightsHolder>
286                 <lido:legalBodyID lido:type=
287                     "http://terminology.lido-schema.org/identifizier_type/uri">
288                     http://ld.zdb-services.de/resource/organisations/DE-MUS-056211</lido:legalBodyID>
289                 <lido:legalBodyName>
290                     <lido:appellationValue>Georg-August-Universität Göttingen /
291                     Ethnologische Sammlung</lido:appellationValue>
292                 </lido:legalBodyName>
293                 <lido:legalBodyWebLink>
294                     http://www.uni-goettingen.de/de/sammlung/28899.html</lido:legalBodyWebLink>
295             </lido:rightsHolder>
296         </lido:recordRights>
297         <lido:recordRights>
298             <lido:rightsType>
299                 <lido:conceptID lido:type=
300                     "http://terminology.lido-schema.org/identifizier_type/uri">
301                     https://creativecommons.org/publicdomain/zero/1.0/deed.de</lido:conceptID>
302                 <lido:term>CC0 1.0 Universell (CC0 1.0)</lido:term>
303                 <lido:term xml:lang="en">CC0 1.0 Universal (CC0 1.0)</lido:term>
304             </lido:rightsType>
305         </lido:recordRights>

```



```
297 <lido:recordInfoSet lido:type="http://terminology.lido-schema.org/lido00471">
298 <lido:recordMetadataDate lido:type=
    "http://terminology.lido-schema.org/recordMetadataDate_type/created">
299 2013-01-30T11:47:17+00</lido:recordMetadataDate>
    <lido:recordMetadataDate lido:type=
    "http://terminology.lido-schema.org/recordMetadataDate_type/modified">
    2017-10-19T10:50:26+00</lido:recordMetadataDate>
300 </lido:recordInfoSet>
301 <lido:recordInfoSet lido:type="http://terminology.lido-schema.org/lido00470">
302 <lido:recordMetadataDate lido:type=
    "http://terminology.lido-schema.org/recordMetadataDate_type/created">
    2017-11-29T06:44:52+01</lido:recordMetadataDate>
303 </lido:recordInfoSet>
304 </lido:recordWrap>
305 <lido:resourceWrap>
306 <lido:resourceSet lido:sortorder="1">
307 <lido:resourceID lido:type=
    "http://terminology.lido-schema.org/identifier_type/local_identifier"
    lido:source="kuniweb">
    record_kuniweb_929301_media/record_kuniweb_929301_391166.jpg
    </lido:resourceID>
308 <lido:resourceRepresentation lido:type=
    "http://terminology.lido-schema.org/resourceRepresentation_type/provided_r
    epresentation">
309 <lido:linkResource lido:formatResource="image/jpeg">
    http://sammlungen.uni-goettingen.de/iiif/image/record_kuniweb_929301/re
    cord_kuniweb_929301_391166.jpg/full/full/0/default.jpg
    </lido:linkResource>
310 <lido:resourceMeasurementsSet>
311 <lido:measurementType xml:lang="en">width</lido:measurementType>
312 <lido:measurementUnit xml:lang="en">pixel</lido:measurementUnit>
313 <lido:measurementValue>1772</lido:measurementValue>
314 </lido:resourceMeasurementsSet>
315 <lido:resourceMeasurementsSet>
316 <lido:measurementType xml:lang="en">height</lido:measurementType>
317 <lido:measurementUnit xml:lang="en">pixel</lido:measurementUnit>
318 <lido:measurementValue>1277</lido:measurementValue>
319 </lido:resourceMeasurementsSet>
320 </lido:resourceRepresentation>
321 <lido:resourceRepresentation lido:type=
    "http://terminology.lido-schema.org/resourceRepresentation_type/preview_re
    presentation">
322 <lido:linkResource lido:formatResource="image/jpeg">
    http://sammlungen.uni-goettingen.de/iiif/image/record_kuniweb_929301/re
    cord_kuniweb_929301_391166.jpg/full/300,/0/default.jpg
    </lido:linkResource>
323 <lido:resourceMeasurementsSet>
324 <lido:measurementType xml:lang="en">width</lido:measurementType>
325 <lido:measurementUnit xml:lang="en">pixel</lido:measurementUnit>
326 <lido:measurementValue>300</lido:measurementValue>
327 </lido:resourceMeasurementsSet>
328 <lido:resourceMeasurementsSet>
329 <lido:measurementType xml:lang="en">height</lido:measurementType>
330 <lido:measurementUnit xml:lang="en">pixel</lido:measurementUnit>
331 <lido:measurementValue>216</lido:measurementValue>
332 </lido:resourceMeasurementsSet>
333 </lido:resourceRepresentation>
334 <lido:resourceDateTaken>
335 <lido:displayDate>23. März 2009</lido:displayDate>
336 <lido:date>
337 <lido:earliestDate>2009-03-23</lido:earliestDate>
338 <lido:latestDate>2009-03-23</lido:latestDate>
339 </lido:date>
340 </lido:resourceDateTaken>
341 <lido:rightsResource>
342 <lido:rightsType>
343 <lido:term>Verwertungsrecht</lido:term>
344 </lido:rightsType>
345 <lido:rightsHolder>
346 <lido:legalBodyName>
347 <lido:appellationValue>Ethnologische Sammlung
    </lido:appellationValue>
```

```
348         </lido:legalBodyName>
349     </lido:rightsHolder>
350 </lido:rightsResource>
351 <lido:rightsResource>
352     <lido:rightsType>
353         <lido:conceptID lido:type=
            "http://terminology.lido-schema.org/identifier_type/uri">
            https://creativecommons.org/licenses/by-sa/4.0/deed.de
        </lido:conceptID>
354         <lido:term>Namensnennung - Weitergabe unter gleichen Bedingungen
            4.0 Deutschland (CC BY-SA 4.0 DE)</lido:term>
355         <lido:term xml:lang="en">Attribution-ShareAlike 4.0 Germany (CC
            BY-SA 4.0 DE)</lido:term>
356     </lido:rightsType>
357 </lido:rightsResource>
358 </lido:resourceSet>
359 </lido:resourceWrap>
360 </lido:administrativeMetadata>
361 </lido:lido>
362
```

ANHANG 4: Datensatz record_kuniweb_929301 in RDF/XML

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <rdf:RDF xmlns:lido="http://www.lido-schema.org"
3      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4      xmlns:dc="http://purl.org/dc/elements/1.1/"
5      xmlns:skos="http://www.w3.org/2004/02/skos/core#"
6      xmlns:gml="http://www.opengis.net/gml"
7      xmlns:gn="http://www.geonames.org/ontology#"
8      xmlns:wgs84_pos="http://www.w3.org/2003/01/geo/wgs84_pos#"
9      xmlns:marcrel="http://id.loc.gov/vocabulary/relators/"
10     xmlns:ido="http://purl.org/asch/ido/180704#"
11     xmlns:pdo="http://purl.org/asch/pdo/180704#"
12     xmlns:owl="http://www.w3.org/2002/07/owl#"
13     xmlns:crm="http://erlangen-crm.org/170309/"
14     xmlns:aat="http://vocab.getty.edu/aat/"
15     xmlns:gnd="http://d-nb.info/gnd/"
16     xmlns:nms="http://nomisma.org/ontology#"
17     xmlns:dcmitype="http://purl.org/dc/dcmitype/"
18     xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
19     xmlns:edm="http://www.europeana.eu/schemas/edm/"
20     xmlns:dcterms="http://purl.org/dc/terms/"
21     xmlns:prov="http://www.w3.org/ns/prov#"
22     xmlns:foaf="http://xmlns.com/foaf/0.1/"
23     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
24     xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
25     <rdf:Description rdf:about=
26         "https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022">
27         <rdf:type rdf:resource=
28             "http://portal.wissenschaftliche-sammlungen.de/ObjectGenre/244"/>
29         <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Item"/>
30         <rdf:type rdf:resource="http://obg.vocnet.org/00002113"/>
31         <ido:physicalDescription>
32             <rdf:Description>
33                 <rdf:type rdf:resource="http://purl.org/asch/ido/180704#Appearance"/>
34                 <dcterms:extent>
35                     <rdf:Description>
36                         <rdf:type rdf:resource=
37                             "http://erlangen-crm.org/170309/E54_Dimension"/>
38                         <dc:type rdf:lang="en">width</dc:type>
39                         <crm:P91_has_unit rdf:lang="en">cm</crm:P91_has_unit>
40                         <crm:P90_has_value>10</crm:P90_has_value>
41                     </rdf:Description>
42                 </dcterms:extent>
43             </rdf:Description>
44             <dcterms:extent>
45                 <rdf:Description>
46                     <rdf:type rdf:resource=
47                         "http://erlangen-crm.org/170309/E54_Dimension"/>
48                     <dc:type rdf:resource="http://d-nb.info/gnd/4224141-8"/>
49                     <crm:P91_has_unit rdf:lang="en">cm</crm:P91_has_unit>
50                     <crm:P90_has_value>23</crm:P90_has_value>
51                 </rdf:Description>
52             </dcterms:extent>
53         </ido:physicalDescription>
54         <ido:hasAppellation>
55             <rdf:Description>
56                 <rdf:type rdf:resource="http://purl.org/asch/ido/180704#Appellation"/>
57                 <skos:prefLabel rdf:lang="de">Pantoffeln aus Marokko</skos:prefLabel>
58                 <skos:altLabel rdf:lang="en">slippers</skos:altLabel>
59             </rdf:Description>
60         </ido:hasAppellation>
61         <dc:identifier>Af 423 a-b</dc:identifier>
62         <ido:wasAffectedBy>
63             <rdf:Description>
64                 <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Custody"/>
65                 <marcrel:cur rdf:resource=
66                     "http://ld.zdb-services.de/resource/organisations/DE-MUS-056211"/>
67                 <pdo:tookPlaceAt>
68                     <rdf:Description>
69                         <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Place"/>
70                         <wgs84_pos:lat rdf:datatype=
71                             "http://www.w3.org/2001/XMLSchema#float">51.53512</wgs84_pos:lat>
72                         <wgs84_pos:long rdf:datatype=
```



```

67         "http://www.w3.org/2001/XMLSchema#float">9.94042</wgs84_pos:long>
68     </rdf:Description>
69 </pdo:tookPlaceAt>
70 </rdf:Description>
71 </ido:wasAffectedBy>
72 <dc:description xml:lang="de">Bei diesen Schuhen handelt es sich um ein Paar
offene Slipper für Frauen. Die Pantoffeln sind aus zwei Teilen gefertigt
worden. Die Schuhkappe besteht, wie die Sohle, aus naturbelassenem, gelben
Leder. Hier finden sich jedoch mit gold- und rosafarbenem Garn bestickte
Flächen als auch ein zentrales Blumenornament auf dem Schaft des Schuhs. Des
Weiteren wurde eine blaue Quaste auf jedem Pantoffel appliziert. Die Sohle ist
mit rotem Samt in der Lauffläche gepolstert. </dc:description>
72 </ido:wasAffectedBy>
73 <rdf:Description>
74 <rdf:type rdf:resource=
"http://terminology.lido-schema.org/eventType/collecting"/>
75 <marcrel:col rdf:resource="http://d-nb.info/gnd/118650599"/>
76 </rdf:Description>
77 </ido:wasAffectedBy>
78 <ido:wasAffectedBy>
79 <rdf:Description>
80 <rdf:type rdf:resource=
"http://terminology.lido-schema.org/eventType/production"/>
81 <pdo:start>
82 <rdf:Description>
83 <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Time"/>
84 <rdf:value rdf:datatype="http://purl.org/dc/terms/W3CDTF">1701
</rdf:value>
85 <pdo:qualifier xml:lang="en">approximate</pdo:qualifier>
86 </rdf:Description>
87 </pdo:start>
88 <pdo:end>
89 <rdf:Description>
90 <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Time"/>
91 <rdf:value rdf:datatype="http://purl.org/dc/terms/W3CDTF">1800
</rdf:value>
92 <pdo:qualifier xml:lang="en">approximate</pdo:qualifier>
93 </rdf:Description>
94 </pdo:end>
95 <pdo:material rdf:resource="http://vocab.getty.edu/aat/300011845"/>
96 <pdo:technique xml:lang="de">gestickt</pdo:technique>
97 </rdf:Description>
98 </ido:wasAffectedBy>
99 <ido:wasAffectedBy>
100 <rdf:Description>
101 <rdf:type rdf:resource="http://terminology.lido-schema.org/eventType/use"
/>
102 <pdo:hasPurpose xml:lang="de">Fußbekleidung für Frauen</pdo:hasPurpose>
103 <pdo:tookPlaceAt rdf:resource="http://sws.geonames.org/6255146"/>
104 <pdo:tookPlaceAt rdf:resource="http://sws.geonames.org/2542007"/>
105 </rdf:Description>
106 </ido:wasAffectedBy>
107 <dcterms:isReferencedBy rdf:resource=
"http://uri.gbv.de/document/gvk:ppn:532043979"/>
108 <ido:isGatheredInto>
109 <rdf:Description>
110 <rdf:type rdf:resource="http://purl.org/asch/ido/180704#Collection"/>
111 <dc:identifier>slg_1047</dc:identifier>
112 </rdf:Description>
113 </ido:isGatheredInto>
114 <ido:isGatheredInto>
115 <rdf:Description>
116 <rdf:type rdf:resource="http://purl.org/asch/ido/180704#Collection"/>
117 <dc:identifier>slg_1021</dc:identifier>
118 </rdf:Description>
119 </ido:isGatheredInto>
120 <ido:hasDigitalRepresentation rdf:resource=
"http://sammlungen.uni-goettingen.de/rest/image/record_kuniweb_929301/record_kun
iweb_929301_391166.jpg/full/full/0/default.jpg"/>
121 </rdf:Description>
122 <rdf:Description rdf:about=
"http://sammlungen.uni-goettingen.de/objekt/record_kuniweb_929301">

```

```
123     <foaf:primaryTopic rdf:resource=  
124       "https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022" />  
125     <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Record" />  
126     <dcterms:created rdf:datatype="http://purl.org/dc/terms/W3CDTF">  
127       2017-11-29T06:44:52+01</dcterms:created>  
128     <dc:rights rdf:resource=  
129       "https://creativecommons.org/publicdomain/zero/1.0/deed.de" />  
130     <dcterms:rightsHolder rdf:resource=  
131       "http://ld.zdb-services.de/resource/organisations/DE-MUS-056211" />  
132   </rdf:Description>  
133   <rdf:Description rdf:about=  
134     "http://sammlungen.uni-goettingen.de/rest/image/record_kuniweb_929301/record_kuniwe  
135     b_929301_391166.jpg/full/full/0/default.jpg">  
136     <rdf:type rdf:resource="http://purl.org/asch/ido/180704#DigitalRepresentation" />  
137     <dcterms:created rdf:datatype="http://purl.org/dc/terms/W3CDTF">2009-03-23  
138     </dcterms:created>  
139     <dc:rights rdf:resource=  
140       "https://creativecommons.org/licenses/by-sa/4.0/deed.de" />  
141     <dcterms:rightsHolder>  
142       <rdf:Description>  
143         <rdf:type rdf:resource="http://purl.org/asch/pdo/180704#Agent" />  
144         <foaf:name>Georg-August-Universität Göttingen / Ethnologische Sammlung  
145         </foaf:name>  
146       </rdf:Description>  
147     </dcterms:rightsHolder>  
148   </rdf:Description>  
149 </rdf:RDF>
```

ANHANG 5: Datensatz record_kuniweb_929301 in Turtle

```
1 @prefix lido: <http://www.lido-schema.org> .
2 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
3 @prefix dc: <http://purl.org/dc/elements/1.1/> .
4 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
5 @prefix gml: <http://www.opengis.net/gml> .
6 @prefix gn: <http://www.geonames.org/ontology#> .
7 @prefix wgs84_pos: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
8 @prefix marcrel: <http://id.loc.gov/vocabulary/relators/> .
9 @prefix ido: <http://purl.org/asch/ido/180704#> .
10 @prefix pdo: <http://purl.org/asch/pdo/180704#> .
11 @prefix owl: <http://www.w3.org/2002/07/owl#> .
12 @prefix crm: <http://erlangen-crm.org/170309/> .
13 @prefix aat: <http://vocab.getty.edu/aat/> .
14 @prefix gnd: <http://d-nb.info/gnd/> .
15 @prefix nms: <http://nomisma.org/ontology#> .
16 @prefix dcmitype: <http://purl.org/dc/dcmitype/> .
17 @prefix dwc: <http://rs.tdwg.org/dwc/terms/> .
18 @prefix edm: <http://www.europeana.eu/schemas/edm/> .
19 @prefix dcterms: <http://purl.org/dc/terms/> .
20 @prefix prov: <http://www.w3.org/ns/prov#> .
21 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
22 @prefix xsi: <http://www.w3.org/2001/XMLSchema-instance> .
23 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
24
25 <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
26   rdf:type
27     <http://portal.wissenschaftliche-sammlungen.de/ObjectGenre/244> ;
28   rdf:type      pdo:Item ;
29   rdf:type      <http://obg.vocnet.org/00002113> ;
30   ido:physicalDescription _:b0 .
31
32   _:b0   rdf:type      ido:Appearance ;
33         dcterms:extent _:b1 .
34
35   _:b1   rdf:type      crm:E54_Dimension ;
36         dc:type        "width"@en ;
37         crm:P91_has_unit "cm"@en ;
38         crm:P90_has_value "10" .
39
40   _:b0   dcterms:extent _:b2 .
41
42   _:b2   rdf:type      crm:E54_Dimension ;
43         dc:type        gnd:4224141-8 ;
44         crm:P91_has_unit "cm"@en ;
45         crm:P90_has_value "23" .
46
47 <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
48   ido:hasAppellation _:b3 .
49
50   _:b3   rdf:type      ido:Appellation ;
51         skos:prefLabel "Pantoffeln aus Marokko"@de ;
52         skos:altLabel  "slippers"@en .
53
54 <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
55   dc:identifier      "Af 423 a-b" ;
56   ido:wasAffectedBy _:b4 .
57
58   _:b4   rdf:type      pdo:Custody ;
59         marcrel:cur
60         <http://ld.zdb-services.de/resource/organisations/DE-MUS-056211> ;
61         pdo:tookPlaceAt _:b5 .
62
63   _:b5   rdf:type      pdo:Place ;
64         wgs84_pos:lat   "51.53512"^^xsd:float ;
65         wgs84_pos:long  "9.94042"^^xsd:float .
66
67 <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
68   dc:description      "Bei diesen Schuhen handelt es sich um ein Paar offene
69   Slipper für Frauen. Die Pantoffeln sind aus zwei Teilen gefertigt worden.
70   Die Schuhkappe besteht, wie die Sohle, aus naturbelassenem, gelben Leder.
71   Hier finden sich jedoch mit gold- und rosafarbenem Garn bestickte Flächen
72   als auch ein zentrales Blumenornament auf dem Schaft des Schuhs. Des
73   Weiteren wurde eine blaue Quaste auf jedem Pantoffel appliziert. Die Sohle
```

```
        ist mit rotem Samt in der Lauffläche gepolstert. "@de ;
        ido:wasAffectedBy _:b6 .
67
68
69   _:b6   rdf:type      <http://terminology.lido-schema.org/eventType/collecting> ;
70         marcrel:col   gnd:118650599 .
71
72   <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
73     ido:wasAffectedBy _:b7 .
74
75   _:b7   rdf:type      <http://terminology.lido-schema.org/eventType/production> ;
76         pdo:start     _:b8 .
77
78   _:b8   rdf:type      pdo:Time ;
79         rdf:value      "1701"^^dcterms:W3CDTF ;
80         pdo:qualifier  "approximate"@en .
81
82   _:b7   pdo:end       _:b9 .
83
84   _:b9   rdf:type      pdo:Time ;
85         rdf:value      "1800"^^dcterms:W3CDTF ;
86         pdo:qualifier  "approximate"@en .
87
88   _:b7   pdo:material  aat:300011845 ;
89         pdo:technique  "gestickt"@de .
90
91   <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
92     ido:wasAffectedBy _:b10 .
93
94   _:b10   rdf:type      <http://terminology.lido-schema.org/eventType/use> ;
95         pdo:hasPurpose  "Fußbekleidung für Frauen"@de ;
96         pdo:tookPlaceAt <http://sws.geonames.org/6255146/> ;
97         pdo:tookPlaceAt <http://sws.geonames.org/2542007/> .
98
99   <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
100     dcterms:isReferencedBy <http://uri.gbv.de/document/gvk:ppn:532043979> ;
101     ido:isGatheredInto    _:b11 .
102
103   _:b11   rdf:type      ido:Collection ;
104         dc:identifier    "slg_1047" .
105
106   <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
107     ido:isGatheredInto    _:b12 .
108
109   _:b12   rdf:type      ido:Collection ;
110         dc:identifier    "slg_1021" .
111
112   <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022>
113     ido:hasDigitalRepresentation
114       <http://sammlungen.uni-goettingen.de/rest/image/record_kuniweb_929301/record_k
115       uniweb_929301_391166.jpg/full/full/0/default.jpg> .
116
117   <http://sammlungen.uni-goettingen.de/objekt/record_kuniweb_929301>
118     foaf:primaryTopic
119       <https://hdl.handle.net/21.11107/67db8b46-afaa-4b48-a8be-58eb18a7a022> ;
120     rdf:type      pdo:Record ;
121     dcterms:created "2017-11-29T06:44:52+01"^^dcterms:W3CDTF ;
122     dc:rights
123       <https://creativecommons.org/publicdomain/zero/1.0/deed.de> ;
124     dcterms:rightsHolder
125       <http://ld.zdb-services.de/resource/organisations/DE-MUS-056211> .
126
127   <http://sammlungen.uni-goettingen.de/rest/image/record_kuniweb_929301/record_kuniweb_9
128   29301_391166.jpg/full/full/0/default.jpg>
129     rdf:type      ido:DigitalRepresentation ;
130     dcterms:created "2009-03-23"^^dcterms:W3CDTF ;
131     dc:rights
132       <https://creativecommons.org/licenses/by-sa/4.0/deed.de> ;
133     dcterms:rightsHolder _:b13 .
134
135   _:b13   rdf:type      pdo:Agent ;
136         foaf:name      "Georg-August-Universität Göttingen / Ethnologische Sammlung" .
```


ANHANG 6: Datensatz record_kuniweb_929301 als RDF-Graph

http://asch.wiki.gwdg.de/index.php/File:Graph_Record_kuniweb_929301.png

ANHANG 7: Readme.txt der beiliegenden CD

```
1
2 #####
3
4 Readme.txt
5
6 @Author: Antje Niemann (TH Wildau, Matrikelnummer 50041781)
7 @Date: 14.02.2019
8
9 This CD belongs to the master thesis by Antje Niemann:
10
11 "Ein Knowledge Graph für wissenschaftliche Sammlungen"
12
13 The CD contains the files and programs that were developed and used in the course of
14 the thesis.
15
16 #####
17
18 The CD is divided into 7 directories with the following content:
19
20 Directory 1_text:
21 -----
22 MA_Niemann_Knowledge_Graph_2019.pdf: Text Master-Thesis
23
24 Mapping_Table_lido2rdf.pdf: Mapping-table which contains the
25 requirements and constructions to do the LIDO-XML-RDF/XML conversion
26
27 Graph_Record_kuniweb_929301_Pantoffeln_aus_Marokko.PNG Graph for
28 record_kuniweb_929391. An example.
29
30 Directory 2_data:
31 -----
32 Directory all_files_lidoxml: 21.071 LIDO/XML files
33 Directory all_files_rdfxml: 21.071 converted RDF/XML files
34 Directory all_files_turtle: 21.071 converted turtle files
35 Directory testing: 20 test-files and the sample file in each of the
36 three metadata formats
37
38 Directory 3_baseX_queries: Files with exemplary queries
39 -----
40
41 Directory 4_xsltproc:
42 -----
43 mkallrecordsxml.sh: File to integrate all LIDO-XML files via XInclude
44 allrecords.xml: Result file of mkallrecordsxml.sh
45 demoSearch.xslt: Example search in all LIDO-XML files
46
47 Directory 5_xslt:
48 -----
49 lido2rdfxml.xsl: XLST style sheet which does the conversion
50
51 Directory 6_conv:
52 -----
53 batch_lido2rdfxml.sh: Batch process for the conversion from LIDO to RDF/XML
54 batch_rdfxml2ttl.sh: Batch process for the conversion from RDF/XML to Turtle
55
56 rdfxml.rng: RELAX NG Compact Schema for RDF/XML syntax validation.
57 Downloaded at
58 http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/rdfxml.rng (19.01.2019)
59
60 Directory apache-jena-3.8.0: RIOT tool for conversion into Turtle. Downloaded
61 at https://jena.apache.org/download/ (19.01.2019)
62
63 Directory saxon: XSLT processor. Downloaded at
64 https://sourceforge.net/projects/saxon (19.01.2019)
65
66 Directory 7_sparql_queries: Files with exemplary queries
67 -----
```


SELBSTSTÄNDIGKEITSERKLÄRUNG

Ich erkläre hiermit, dass ich die vorliegende Abschlussarbeit selbstständig verfasst und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Werken wörtlich oder sinngemäß übernommenen Gedanken sind unter Angabe der Quellen gekennzeichnet.

Datum: 18. Februar 2019

Unterschrift:

A handwritten signature in black ink, appearing to read 'Justus Mann', written in a cursive style.