



Publikation der Rückmeldungen und Beobachtungen zu den integrierenden Use-Cases (R 4.2.2)

Version 17. August 2017

Cluster 4

Verantwortlicher Partner SUB, DAI

DARIAH-DE Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1610A bis J, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Projekt: DARIAH-DE: Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

BMBF Förderkennzeichen: 01UG1610A bis J

Laufzeit: März 2016 bis Februar 2019

Dokumentstatus: Final

Verfügbarkeit: öffentlich

Autoren:

Matteo Romanello, DAI

Claudio Leone, SUB

Tobias Gradl, MInfBa

Stefan E. Funk, SUB

Timo Steyer, HAB

Revisionsverlauf:

Datum	Autor	Kommentare
15.06.2017	Matteo Romanello	Erstellung des Dokuments
06.07.2017	Tobias Gradl	Text und Bilder zu §2.2 hinzugefügt
14.07.2017	Claudio Leone	§2.1 hinzugefügt
14.07.2017	Matteo Romanello	§2.2 fertig gestellt
17.07.2017	Claudio Leone	Einleitung ausgeführt
17.07.2017	Stefan E. Funk	Allgemeine Überarbeitung
18.07.2017	Tobias Gradl	§3 hinzugefügt
18.07.2017	Stefan E. Funk	Formatierung
02.08.2017	Timo Steyer	Überarbeitung Use-Case MWW
17.08.2017	Stefan E. Funk	Endredaktion



Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 3.0 Deutschland zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <http://creativecommons.org/licenses/by/3.0/de/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

Inhaltsverzeichnis

1. Einleitung	5
2. Use-Cases	6
2.1. Use-Case Forschungsverbund Marbach Weimar Wolfenbüttel (MWW)	6
2.2. Use-Case Archäologie (Deutsches Archäologisches Institut – DAI)	10
2.2.1. Datenimport	11
2.2.2. Schemaregistrierung	12
2.2.3. Datenanreicherung	16
2.2.4. Erstellung des Mappings zwischen Datenstrukturen	20
2.2.5. Visualisierung der Mapping-Ergebnisse	22
3. Technische Rückmeldungen	24
4. Literaturverzeichnis	25
5. Abbildungsverzeichnis	26

1. Einleitung

Ziel des Arbeitspaketes 4.2 in Cluster 4 ist es, Geisteswissenschaftler bei der Erstellung und Nutzung von geisteswissenschaftlichen Sammlungen und Forschungsdaten zu unterstützen. AP 4.2 kollaboriert daher eng mit Institutionen und Forscherinnen und Forschern und unterstützt dabei vor allem im Hinblick auf den Aufbau, die Pflege und die deskriptive Erfassung der Sammlungen und Kollektionen von Forschungsdaten. Ziel ist es dabei, die Entwicklungsarbeit von Arbeitspaket 4.3 kontinuierlich bei der Integration der forschungs- und nutzerbasierten Anforderungen der Geistes- und Kulturwissenschaften in die DARIAH-DE Forschungsdaten-Föderationsarchitektur zu unterstützen und zugleich der Heterogenität dieser Anforderungen gerecht zu werden.

Der Forschungsverbund Marbach Weimar Wolfenbüttel (MWW) repräsentiert exemplarisch das Szenario einer institutionsübergreifenden Kooperation mit eigener Infrastruktur und der zielorientierten Nutzung einiger DARIAH-DE Dienste. Der Anwendungsfall steht auch exemplarisch für den Umgang mit Datensammlungen, die sich aus bereits vorhandenen Daten unterschiedlicher Herkunft zusammensetzen.

Anhand eines Use-Cases aus der Archäologie des Deutschen Archäologischen Instituts (DAI) wird die Nutzung der von DARIAH-DE bereitgestellten Forschungsdateninfrastruktur und angebundener Dienste in der fachwissenschaftlichen Anwendung überprüft.

Der vorliegende Report fasst die bisherigen Rückmeldungen und Beobachtungen zu den integrierenden Use-Cases zusammen und zeigt exemplarische Umsetzungen der sich ergebenden Anforderungen auf. Dabei wird auch auf einige technischen Herausforderungen und Probleme eingegangen, die bei der Nutzung von DARIAH-DE Diensten aufgetreten sind.

2. Use-Cases

2.1. Use-Case Forschungsverbund Marbach Weimar Wolfenbüttel (MWW)

Zum Forschungsverbund Marbach Weimar Wolfenbüttel (MWW) gehören mit dem Deutschen Literaturarchiv Marbach, der Klassik Stiftung Weimar und der Herzog August Bibliothek Wolfenbüttel drei wichtige geisteswissenschaftliche Zentren, die nicht nur bedeutende Sammlungen und materielle Zeugnisse aus über fünf Jahrhunderten deutscher und europäischer Kultur-, Literatur- und Ideengeschichte bewahren, sondern sich auch der Erforschung und Transformation des kulturellen Erbes widmen. In diesem Kontext wird eine gemeinsame hausübergreifende digitale Infrastruktur aufgebaut, die neue Zugänge und Arbeitsmethoden zu den Sammlungen ermöglichen wird. Die Forschungsumgebung nimmt dabei neue Entwicklungen aus dem Bereich der Digital Humanities auf, möchte aber auch selber neue Impulse liefern.¹

Zentrale Komponenten der Forschungsumgebung sind der MWW-Forschungsraum und der Verlässliche Speicher. Die bedeutenden Sammlungen der Einrichtungen sollen dabei auf eine neue Ebene der Sichtbarkeit und Nutzung gebracht werden. Eine besondere Herausforderung stellen in diesem Zusammenhang die heterogenen Datenbestände aus Archiven, Datenbanken von Museen und Forschungsprojekten dar. Die inhaltliche und technische Heterogenität bewirkte bisher, dass die vorhandenen Sammlungen noch nicht einmal innerhalb der einzelnen Einrichtungen des Verbundes in einer gemeinsamen Suche recherchiert werden konnten. Durch die Kooperation mit DARIAH-DE und dem Gebrauch der Services der Generischen Suche² von DARIAH-DE konnte für den MWW-Use-Case dem infrastrukturellen Bedarf einer bestandsübergreifenden Suche entsprochen werden: Im Forschungsraum ermöglicht jetzt eine bestandsübergreifende Suche erstmals die gemeinsame Recherche über alle Sammlungen der Verbundeinrichtungen, über Tools können die Bestände für weitere Forschungsprozesse analysiert, visualisiert und bearbeitet werden.

¹ <http://www.mww-forschung.de/>

² <https://de.dariah.eu/generische-suche>

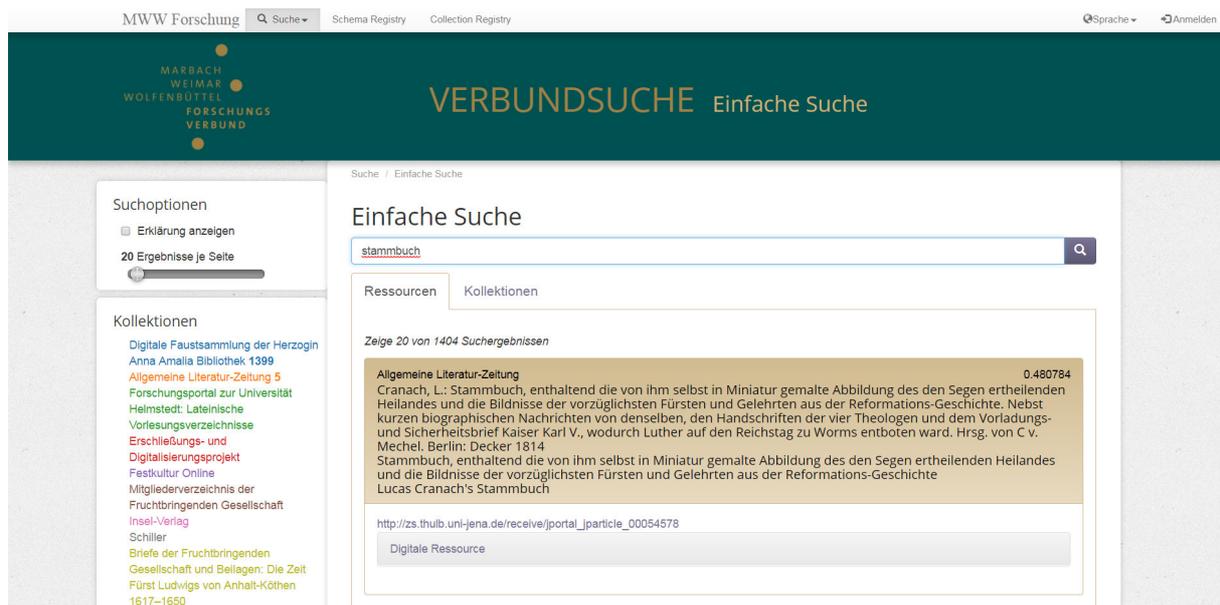


Abbildung 1: Bestandsübergreifende Suche in den Sammlungen des Forschungsverbundes MWW

Dafür wurde eigens eine eigene MWW-Instanz der Collection Registry implementiert, in die die Sammlungen eingetragen werden und mittels eines verbundübergreifenden eigenen Metadatenformates (MWW Metadata Model), einem geplanten Mapping mit dem DARIAH-DE Collection Description Data Model (DCDDM)³, als auch in ihrem ursprünglichen Datenformat in der Generischen Suche integriert werden. Über die Indizierung in der Generischen Suche wird über Verlinkungen die Zugriffsmöglichkeit auf die Daten und Dateien gewährleistet. Auf diese Weise soll zukünftig eine Verbindung mit dem virtuellen Forschungsraum des MWW-Verbundes sichergestellt werden, zum anderen soll die Integration lizenzrechtlich unbedenklicher Sammlungen in die DARIAH-DE-Instanz über die Collection Registry ermöglicht werden.

Die Sammlungen werden zurzeit von MWW-Mitarbeiterinnen und -Mitarbeitern eingetragen, die Schemata registriert und die Mappings hergestellt. Technisch werden die Sammlungen sowohl über standardisierte Schnittstellen (OAI-PMH, SRU) wie auch als statische XML-Files geliefert.

³ <https://wiki.de.dariah.eu/pages/viewpage.action?pageId=52728662>; <https://github.com/DARIAH-DE/DCDDM>

Veröffentlicht	Titel	Kategorie	Format	Datum
Veröffentlicht	A:Schiller - [Bestand, . Sammlung]	Archivbestand	XML/Text file (online); XML/Text file (online)	7. Juli 2017 14:46:46
Veröffentlicht	Briefe der Fruchtbringenden Gesellschaft und Beilagen: Die Zeit Fürst Ludwigs von Anhalt-Köthen 1617–1650 (FG1)	Digitale Edition	XML/Text file (online); XML/Text file (online)	3. Mai 2017 09:18:33
Veröffentlicht	Die Porträtsammlung der Herzog August Bibliothek Wolfenbüttel (Porträtsammlung)	Porträtsammlung	Reading Room	20. Oktober 2016 10:04:19
Veröffentlicht	Digitale Editionen an der Herzog August Bibliothek	Digitale Edition		5. Januar 2017 13:42:25
Veröffentlicht	Digitale Faustsammlung der Herzogin Anna Amalia Bibliothek	Digitale Sammlung der Herzogin Anna Amalia Bibliothek	OAI-PMH	1. Juni 2017 11:24:07
Veröffentlicht	Festkultur Online (FKO)	Alte Drucke	OAI-PMH	26. Oktober 2016 16:25:38
Veröffentlicht	Forschungsportal zur Universität Helmstedt: Lateinische Vorlesungsverzeichnisse (Lateinische Lektionskataloge)	Alte Drucke	XML/Text file (online); XML/Text file (online)	11. Mai 2017 19:28:42
Veröffentlicht	Forschungsportal zur Universität Helmstedt: Rechenschaftsberichte (Rechenschaftsberichte)	Handschriften	XML/Text file (online); XML/Text file (online)	8. März 2017 15:16:06
Veröffentlicht	HAB-Handschriftendatenbank (MSS-DB)	Handschriften	Reading Room	5. Juli 2017 16:09:21
Veröffentlicht	Insel-Verlag (Bestand 50)	Verlagsarchiv	XML/Text file (online)	12. Dezember 2016 14:17:26
Veröffentlicht	Johann Wolfgang Goethe: Repertorium sämtlicher Briefe 1764-1832	Briefrepertorium	XML/Text file (online); XML/Text file (online)	4. Mai 2017 17:44:15

Abbildung 2: Ausschnitt aus der MWW-Instanz der Collection Registry

Für die Anlage der Metadaten werden zudem dort, wo es sinnvoll ist, kontrollierte Vokabulare eingebunden, die als Auswahl zu den betreffenden Metadatenfeldern zukünftig angeboten werden sollen. Neben der Auswahl vorgegebener Termini gibt es zusätzlich auch Freitextfelder, wie z. B. bei der Angabe von räumlichen und zeitlichen Angaben, aber auch die kontrollierten Vokabulare bleiben erweiterbar.

Die geplante weitere Einbindung kontrollierter Vokabulare, wie z. B. „Getty Thesaurus of Geographical Names“ (TGN)⁴, „Gemeinsame Normdatei“ (GND)⁵, „Library of Congress Subject Headings“ (LCSH)⁶, „OpenStreetMap“ (OSM)⁷ oder „OpenGeoNames“ (OGN) steht im Zusammenhang mit einer verbesserten Recherchierbarkeit und stärkeren Nachnutzbarkeit von Sammlungsdaten und ist auch für die Collection Registry generell angedacht. Sie erlaubt zudem eine Weiterverarbeitung über angebundene DARIAH-DE Dienste, wie z. B. über den Geo-Browser⁸, wofür eine Bereitstellung räumlich-zeitlicher Daten vereinfacht wird. Hier wäre beispielsweise eine Zeit-Raum-Visualisierung von institutionsbezogenen Sammlungen und ihren Über-

⁴ <http://www.getty.edu/research/tools/vocabularies/tgn/>

⁵ http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html

⁶ <http://id.loc.gov/authorities/subjects.html>

⁷ <https://www.openstreetmap.de/>

⁸ <https://de.dariah.eu/geobrowser>

schneidungen und Zusammenhängen denkbar. Am Beispiel des MWW-Verbundes zeigt sich, wie sehr die Collection Registry, ihre Einbindung in die DARIAH-DE Data Federation Architecture (DFA)⁹ und das Zusammenspiel mit weiteren Diensten über die bloße Beschreibung und Präsentation von Sammlungsdaten hinaus einen erweiterten Nutzen bietet und die Gestaltung eines Forschungsraumes erlaubt.

Damit verbunden ist aber auch eine geplante ansprechende Präsentation der Sammlung durch Visualisierungen auf der Startseite, auf der einzelne Bilder als Eye Catcher die einzelnen Sammlungen repräsentieren sollen. Hierdurch soll eine Art Schaulaufen entstehen, das dem Nutzer einen schnellen Überblick darüber verschafft, welche Sammlungen der Verbund MWW anbietet. Diese Art der visuell unterstützten Präsentation könnte beispielsweise auch gerade für kleinere Projekte einen besonderen Mehrwert für die Sichtbarkeit ihrer Datensammlungen darstellen.

Der besondere Anwendungsfall von MWW erlaubt zudem insgesamt die Entwicklung und Definition von Workflows, die standardisiert und für Gedächtnisinstitutionen nachnutzbar gestaltet werden können. Im regelmäßigen Austausch zwischen MWW und DARIAH-DE wird ein Erweiterungsbedarf evaluiert und z. B. für die Optimierung der Generischen Suche genutzt. Bis 2018 sollen ca. 30 Sammlungen, die die inhaltliche und technische Vielfalt der Sammlungen des Verbundes MWW repräsentieren, in eine eigene DARIAH-Instanz integriert sein – davon sind zur Zeit bereits 15 integriert worden.

⁹ Für einen Überblick über die Konzepte und Komponenten der DARIAH-DE-Föderationsarchitektur siehe Gradl, T.; Henrich, A. (2016).

2.2. Use-Case Archäologie (Deutsches Archäologisches Institut – DAI)

Zu den Datenbeständen des Deutschen Archäologischen Instituts¹⁰ (DAI) gehören u.a. 2-D-Bilder, 3-D-Modelle, verschiedene Arten kontrollierter Vokabulare, geographische Daten und natürlich auch Grabungsdatenbanken. Die Frage nach dem Umgang mit Datensammlungen, die aus bereits existierenden, heterogenen Daten verschiedener Herkunft abgeleitet sind, ist deswegen hoch relevant und wird in Zukunft immer wichtiger werden. Diese Heterogenität der Daten ist für eine digitale Forschungsinfrastruktur, die auch die Bedürfnisse archäologischer Forschung bedienen können soll, eine erhebliche Herausforderung.

Hier hat sich bereits in der ersten Förderphase gezeigt, dass das Anbieten einfacher Workflows den tatsächlichen Bedürfnissen nicht gerecht wird. Ein generischer Arbeitsablauf, wie bereits im Use-Case MWW skizziert, wird dem nicht gerecht.

Das Ziel des Use-Cases ist es, zu überprüfen, ob die von DARIAH-DE bereitgestellte Forschungsdateninfrastruktur auch den Bedürfnissen der Wissenschaftlerinnen und Wissenschaftler in der Archäologie dienen kann. Insbesondere wurde untersucht, wie ein archäologischer Datensatz durch die DARIAH-DE Tools importiert, angereichert und schließlich visualisiert werden kann.

Der abgebildete Arbeitsablauf besteht aus folgenden Schritten:

- Datenimport
- Schemaregistrierung
- Erstellung von Mappings zwischen Datenstrukturen
- Mappen bzw. Überführen der Datenwerte
- Visualisierung und Analyse

Die Beispieldaten, die für diesen Use-Case benutzt werden, stammen aus einer Datenbank der DAI-Grabung in Pergamon¹¹ und beschreiben etwa 100 keramische Funde. Die Daten beinhalten u. a. wichtige Informationen in unstrukturierter Form, insbesondere die Provenienz- und Datierungsangaben. Um diese Daten analysieren und visualisieren zu können, z. B. mittels des Geo-Browsers, müssen zunächst die unstrukturierten Angaben in eine strukturierte Form überführt werden.

¹⁰ <http://www.dainst.org>

¹¹ <http://www.dainst.org/projekt/-/project-display/14186>

Bei der Überführung und Anreicherung der Daten kommen die Dienste des DARIAH-DE Data Modeling Environment (DME) zum Einsatz (bisher bekannt als Schema- und Crosswalk-Registry¹², Dokumentation und Software befinden sich derzeit im Umstellungsprozess), und insbesondere ihre Fähigkeit, auf externe Ressourcen zuzugreifen. Zwei ad-hoc-API-Wrapper wurden erstellt, damit zeit- und ortsbezogene Textangaben wie „Grobdatierung: hellenistisch-kaiserzeitlich“ oder „Provenienz: Pergamon“ mit den entsprechenden und in Zahlen ausgedrückten Werten kartiert werden können. Schließlich werden die angereicherten Daten mittels des DARIAH-DE Geo-Browsers visualisiert, damit die zeitliche und geographische Verteilung der in den Datensätzen beschriebenen Objekte visuell abgebildet werden kann.

2.2.1. Datenimport

Die archäologische Grabung in Pergamon, sowie viele andere vom DAI ausgeführte Grabungen und Projekte, werden mittels der iDAI.Field Software dokumentiert (vgl. Abbildung 3: Die Eingabemaske für Keramik-Objekte in der iDAI.Field Software). iDAI.Field ist ein modulares Dokumentationssystem für Feldforschungsprojekte, das am DAI entwickelt wurde und in ca. 50 verschiedenen Projekten eingesetzt wurde.¹³ Die durch iDAI.Field gesammelten Daten werden in einer FileMaker-Datenbank gespeichert und können später in verschiedenen Formaten (XML, Excel) exportiert werden.

In der Umsetzung des Use-Cases wurde der XML-Export von FileMaker benutzt. Da die Namen der aus FileMaker generierten XML-Elemente in einigen Fällen Zeichen beinhalten, die in dem XML-Format nicht erlaubt sind (z.B. <[Auto_Objektkennung_Index]>), war es notwendig, solche unerlaubten Zeichen aus den Dateien zu entfernen, um eine valide XML-Datei produzieren zu können.

¹² Es wurde die aktuelle Beta-Version unter <https://minfba.de.dariah.eu/schereg/> verwendet.

¹³ https://www.dainst.org/forschung/forschung-digital/idai.welt/data/-/asset_publisher/Pt831lfwO8uH/content/idai-field

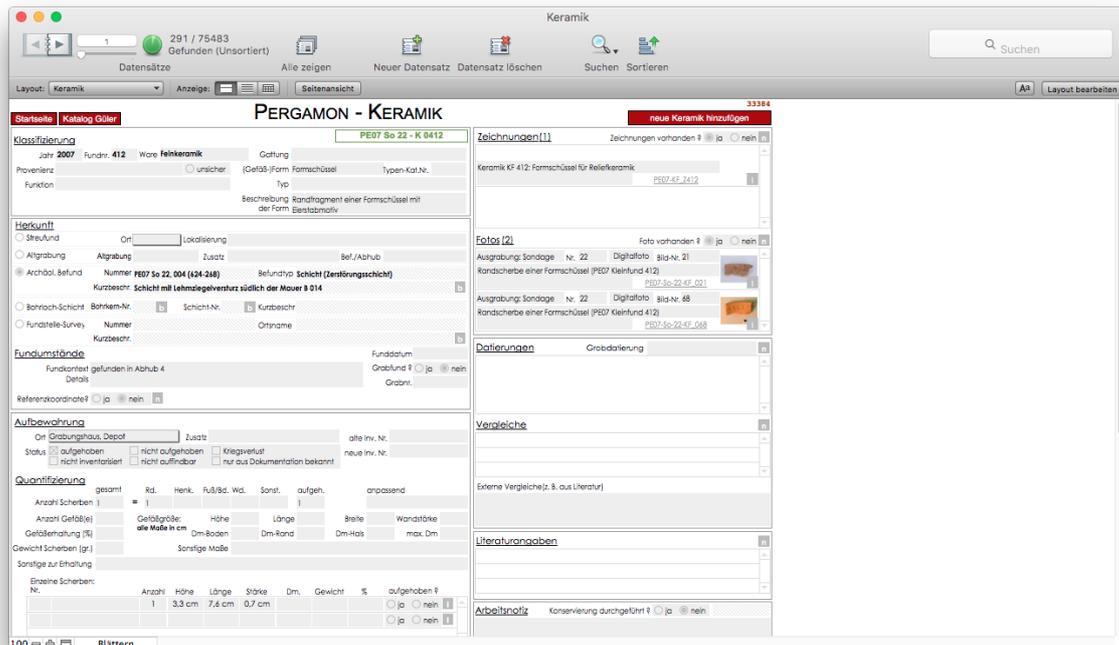


Abbildung 3: Die Eingabemaske für Keramik-Objekte in der iDAI.Field Software

2.2.2. Schemaregistrierung

Da unser Ziel die Visualisierung der Pergamon-Daten mittels des Geo-Browsers ist, müssen die Daten in ein Format umgewandelt werden, das der Geo-Browser als Eingabeformat annehmen kann, wie die Keyhole Markup Language (KML)¹⁴. In den Tools der DARIAH-DE Forschungsinfrastruktur wird eine solche Umwandlung auch „Schema Mapping“ genannt: Das Mapping wird zwischen einem Input- und einem Output-Format der Daten erstellt. Deswegen ist der erste Schritt des Arbeitsablaufs die Registrierung der zwei beteiligten Schemata im DARIAH-DE DME.

Das DME bietet zwei Möglichkeiten, um ein Schema einzutragen: Entweder manuell oder durch das Hochladen der entsprechenden XML Schema Definition (XSD)-Datei. Da kein XSD-Schema für das FileMaker-XML-Format verfügbar war, wurde es durch die Software Oxygen¹⁵ erzeugt. Im Gegensatz dazu ist das KML-Schema online ver-

¹⁴ <https://wiki.de.dariah.eu/display/publicde/Geo-Browser+Dokumentation#Geo-BrowserDokumentation-Spezifikationen%C3%BCrdieNutzung>

¹⁵ <https://www.oxygenxml.com/>

ffügbar¹⁶ und wurde – nach dem Hochladen der XSD-Datei–, automatisch in das DME importiert.

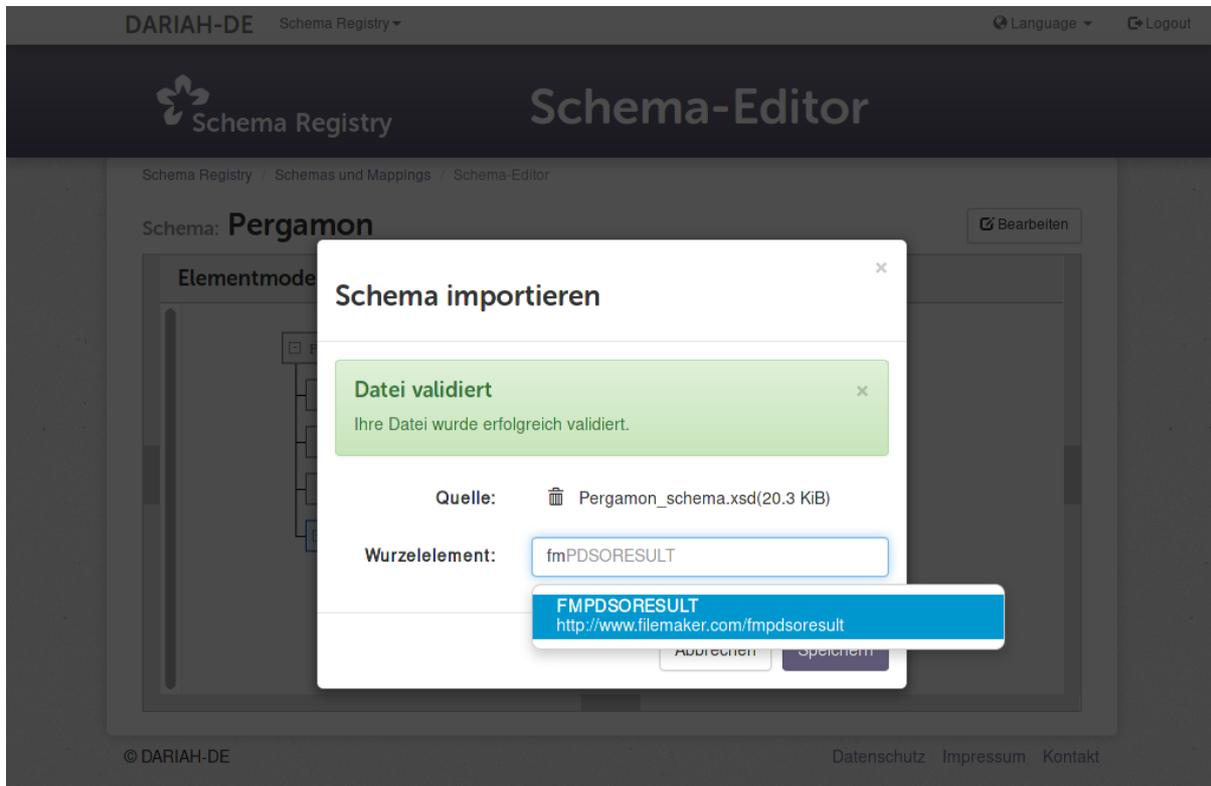


Abbildung 4: Import eines Schemas in das Data Modeling Environment

Sobald das Schema importiert ist, kann ein Wurzelement angelegt werden, das festlegt, wo in der XML-Datei ein Datensatz beginnt. Beispielsweise beinhaltet die Pergamon-XML-Datei, deren Wurzelement `<FMPDSORESULT>` ist (vgl. Abbildung 4: Import eines Schemas in das Data Modeling Environment und Abbildung 5), mehrere Datensätze, jede in einem `<ROW>` Element enthalten. Deswegen wurde das Element `<ROW>` als Wurzelement des Datensatzes angelegt.

¹⁶ <http://schemas.opengis.net/kml/2.2.0/ogckml22.xsd>

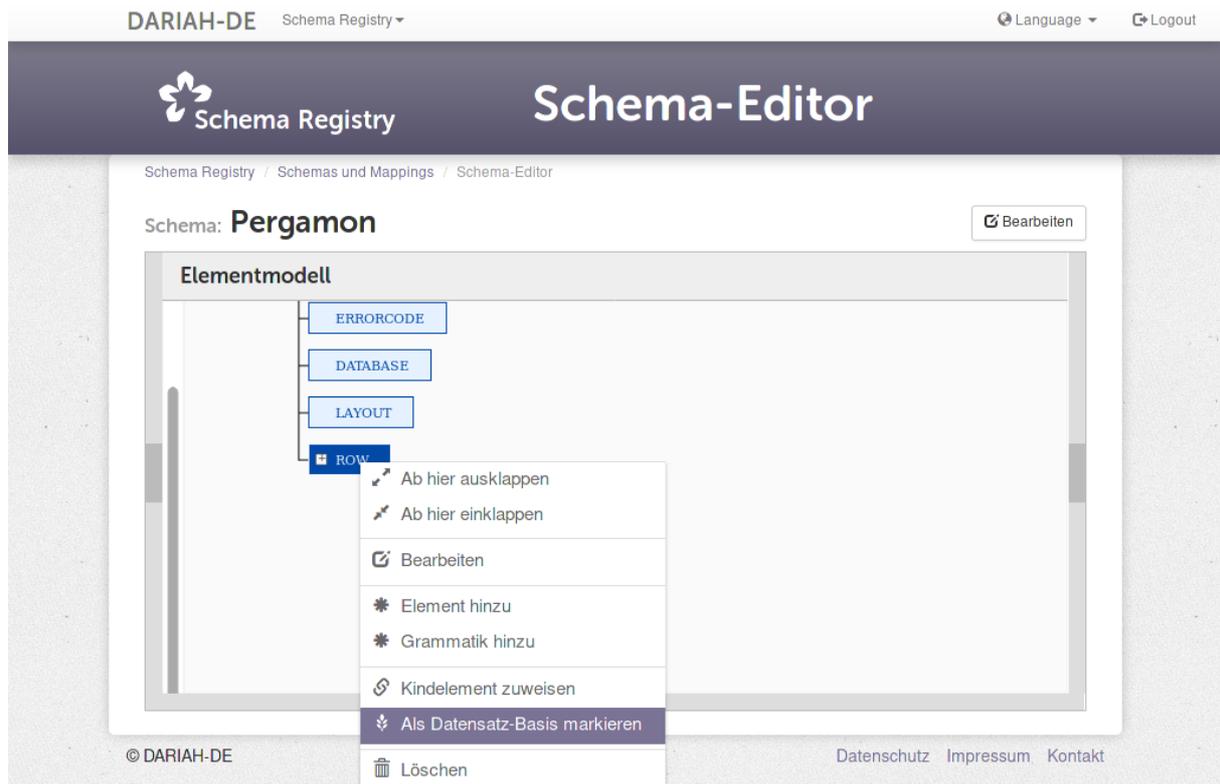


Abbildung 5: Markierung eines Elements des Schemas als Datensatz-Basis

Nach dem Import eines Schemas in das DME ist es auch möglich, einige Beispieldaten zu importieren. Dadurch kann man testen, ob die Daten aus der SR richtig prozessiert werden. Die Beispieldaten können entweder in einem Input-Feld hinzugefügt oder aus einer Datei (vgl. Abbildung 6) gelesen werden. Insbesondere bei Verwendung größerer Dokumente empfiehlt sich der Datei-Upload – insgesamt muss berücksichtigt werden, dass die Funktionalität des DME explizit kleinere Beispieldaten unterstützt und so z. B. keine Laufzeitoptimierungen wie Multithreading oder Caching anwendet. Ein Transformationsdienst für größere Datenmengen ist derzeit in Entwicklung.

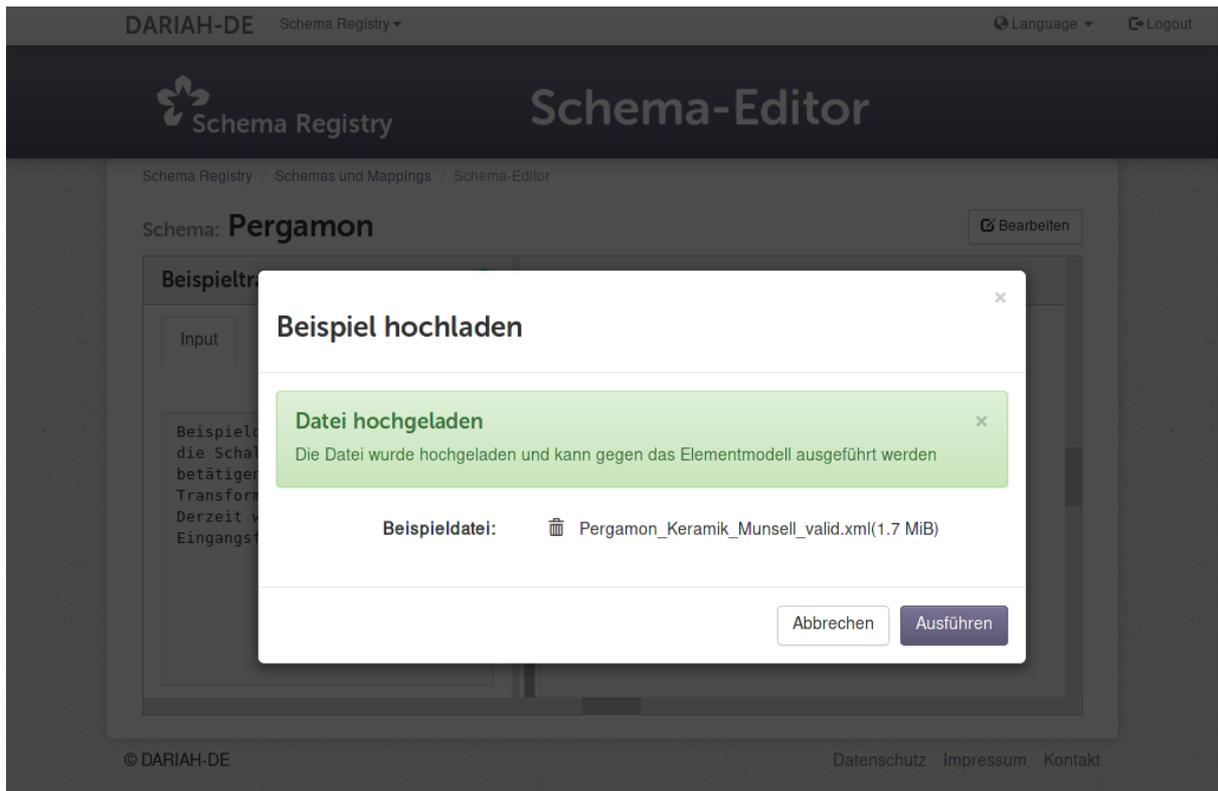


Abbildung 6: Hochladen einer Beispieldatei

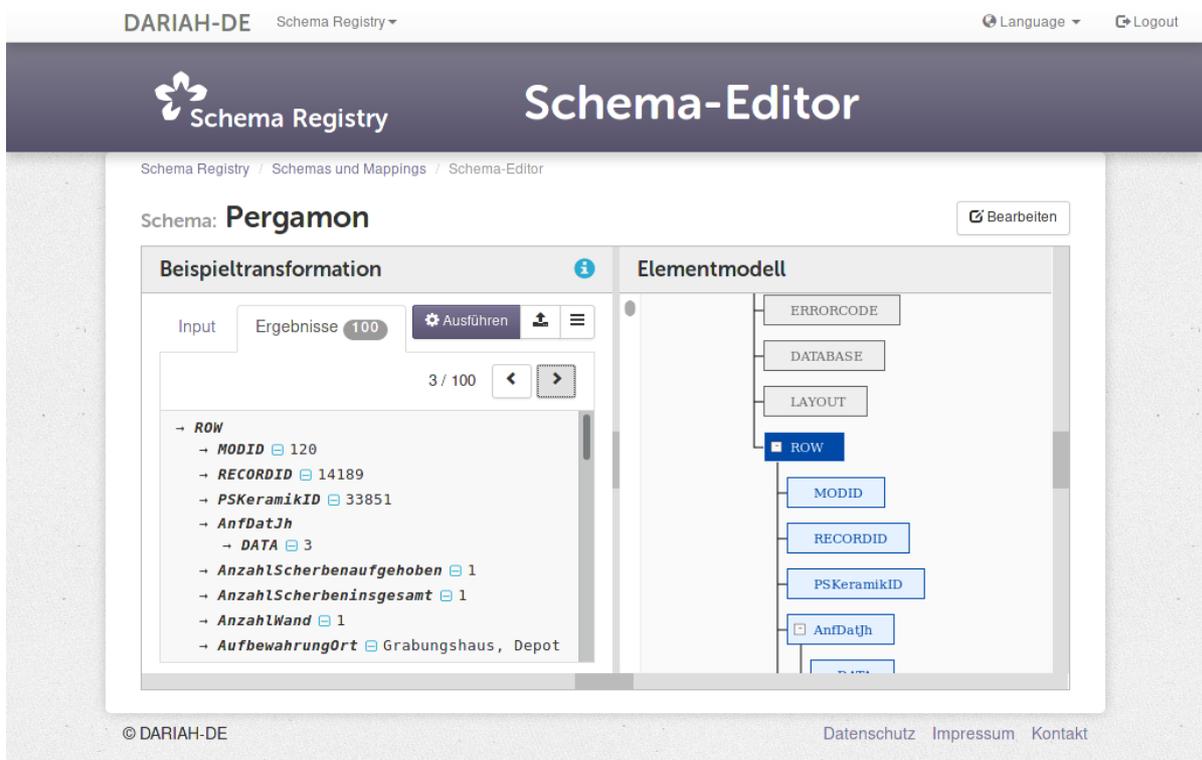


Abbildung 7: Import des Pergamon-Schemas in das Data Modeling Environment

2.2.3. Datenanreicherung

Die Funktionalitäten der SR gehen weit über die einfache Assoziation bestehender Felder hinaus. Die SR kann durch Plugins erweitert werden, damit neue Funktionen und Skripte für die Bearbeitung dieser Felder definiert werden können. Es besteht beispielsweise die Möglichkeit, einen Natural Language Processing (NLP) Dienst einzubinden, der strukturierte Informationen wie Personen- und Ortsnamen aus einem Freitextfeld automatisch extrahiert.¹⁷

Diese Datenanreicherungsfunktion des Tools passt sehr gut zum hier behandelten Use-Case. Wie bereits erwähnt, sind die Provenienz- und Datierungsangaben nur in einer unstrukturierten Form vorhanden und müssen deswegen konvertiert werden. Diese Konvertierung erfolgt in unserem Fall über zwei Normdatenvokabularien, die im Bereich Archäologie vom DAI als Basisdienste betrieben werden: Der iDAI.Gazetteer¹⁸ – der Ortsnamen mit Koordinaten verbindet – und der iDAI.Chronontology¹⁹ – der Zeitbegriffe mit Datierungen verbindet. Beide Ressourcen lassen sich über eine API-Schnittstelle, die die inhaltlichen Daten im JSON Format zurückgibt, mit dem Data Modeling Environment verbinden.

iDAI.ChronOntology

Die ChronOntologys API ermöglicht u. a. eine Suche durch Freitext. Zum Beispiel ist es durch die Suchabfrage

<<http://chronontology.dainst.org/data/period/?q=kaiserzeitlich>>

möglich, nach Zeitbegriffen zu suchen, die den String „Kaiserzeitlich“ beinhalten (vgl. Abbildung 8). Wie man an den Zeilen 27-34 sehen kann, sind die Datierungen „-27“ (begin) und „476“ (end) mit dem Begriff „kaiserzeitlich“ verbunden.

¹⁷ <https://de.dariah.eu/cosmotool>

¹⁸ <https://gazetteer.dainst.org>. Vgl. auch Cuy et al. 2014.

¹⁹ <http://chronontology.dainst.org/>

```

1  {
2  "results" : [
3  {
4    "resource" : {
5      "names" : {
6        "de" : [
7          "Römische Kaiserzeit" ,
8          "Kaiserzeitlich"
9        ],
10       "en" : [
11         "Roman Empire" ,
12         "Roman Imperial"
13       ]
14     },
15     "types" : [
16       "politisch"
17     ],
18     "provenance" : "chronontology" ,
19     "note2" : "Die Römische Kaiserzeit bildet einen Epochenabschnitt der klassischen Antike,
20 zwischen der von Expansion und Anpassungskrisen bestimmten späten Römischen
21 Republik und der Spätantike, in welcher der Mittelmeerraum einen
22 Transformationsprozess durchlief und Westrom unterging. [...]" ,
23     "note3" : "2359913" ,
24     "hasTimespan" : [
25       {
26         "timeOriginal" : "13. Januar 27 v. Chr. - 476 n. Chr" ,
27         "begin" : {
28           "at" : "-27" ,
29           "atPrecision" : "ca"
30         },
31         "end" : {
32           "at" : "476" ,
33           "atPrecision" : "ca"
34         }
35       }
36     ],
37     "relations" : {},
38     "id" : "bCTpXj4LVC2n" ,
39     "type" : "period"
40   }
41 ] ,
42 "total" : 4
43 }
44 }

```

Abbildung 8: JSON-Antwort für <http://chronontology.dainst.org/data/period/?q=kaiserzeitlich>

Abbildung 9: Erweiterung originärer Daten um Ergebnisse der Chronontology API

Die Ausführung von Funktionalität der Chronontology API ist im Namensraum DAI::CHRONONTOLOGY möglich. Im vorliegenden Fall gestaltet sich das Kommando wie folgt:

```
Chronontology = DAI::CHRONONTOLOGY::QUERY(@Chronon);
```

Um aus der Menge der zurückgegebenen Einträge ein Intervall errechnen zu können, werden zudem Kommandos aus dem MATH Funktionsraum verwendet:

```
BeginMin =
```

```
  MATH::MIN(@Response.Item.Resource.Timespan.Begin.At);
```

```
EndMax = MATH::MAX(@Response.Item.Resource.Timespan.End.At);
```

Die durch verschiedene Namensräume für den Nutzer zur Verfügung stehenden Funktionen ermöglichen es, mit Zeitintervallen umzugehen. Das Feld Grobdatierung beispielsweise beinhaltet in der Datenbank oftmals nicht einen einzelnen Zeitbegriff, sondern ein Zeitintervall (z. B. „hellenistisch-kaiserzeitlich“). In solchen Fällen können durch eine grammatikalische Definition (Regel etwa: “Bindestriche teilen mehrere Terme auf”) die Daten vorverarbeitet werden. Die Transformationsregel verarbeitet dann die übergebenen Terme, die jeweils in Chronontology durch die API durchsucht werden. Dann ermöglichen DME Funktionen wie MATH::MIN und MATH::MAX es, ein Zeitintervall zu generieren, das als Beginnndatum den Beginn der hellenistischen Zeit und als Enddatum das Ende der kaiserzeitlichen Zeit aufweist.

Schema Registry / Schemas und Mappings / Schema-Editor

Schema: Pergamon Bearbeiten

Beispieltransformation

Input Ergebnisse 100 Ausführen

7 / 100

- OberflächeÜberzugFarbeinnenMunsell 5 YR 6/6
- OberflächeÜberzugQualität matt dünn
- OberflächeÜberzugSeite innen
- Provenienz Pergamon
 - Location
 - Coordinate
 - Latitude 27.183953
 - Longitude 39.132126
 - Referenzkoordinatenvorhanden nein
 - TonFarbeAußen hellbraun

Elementmodell

- Provenienz
 - Location
 - Location
 - Coordinate
 - Latitude
 - Longitude
- Provenienzunsicher

© DARIAH-DE Datenschutz Impressum Kontakt

Abbildung 10: Erweiterung originärer Daten um Ergebnisse der Gazetteer API

Vergleichbar mit der Chronontology API können auch Funktionen der Gazetteer API auf Daten angewandt werden. Diese stehen im Namensraum DAI::GAZETTEER zur Verfügung. Im vorliegenden Beispiel wird der für einen Query zurückgegebene, erste Treffer als wahrscheinlichste Koordinate verwendet und in den Daten berücksichtigt:

```
Location = DAI::GAZETTEER::TOPCOORD(@Location);
```

2.2.4. Erstellung des Mappings zwischen Datenstrukturen

The screenshot shows the 'Mapping-Editor' interface in the DARIAH-DE Schema Registry. The source schema is 'Pergamon' and the target schema is 'KML_Geobrowser'. The 'Elementmodell' section displays two tree structures. The source schema tree includes 'Coordinate' (with sub-elements 'Latitude' and 'Longitude'), 'Provenienzunsicher', 'Referenzkoordinatavorhand...', 'SchichtNummer', 'TonBemerkung', and 'TonFarbeAußen'. The target schema tree includes 'Placemark' (with sub-elements 'Name', 'Point', 'Coordinates', 'TimeSpan', 'Begin', and 'End'). A yellow arrow points from the 'Coordinate' element in the source schema to the 'Coordinates' element in the target schema. The interface also features a 'Bearbeiten' button and a 'Löschen' button.

Abbildung 11: Erstellung eines Mappings zwischen Quellschema und Zielschema

Auch für die Definition von Mappings können Transformationsregeln spezifiziert werden. Für die Angabe von Koordinaten nach dem KML Schema wird so z. B. folgende Regel definiert:

```
Coordinates = CORE::IFTHENELSE(NULLOREMPTY(@Latitude),  
    "null",  
    CORE::CONCAT(@Latitude, ",", @Longitude)  
);
```

Koordinaten werden demnach nur angelegt, wenn @Latitude (für Daten im Quellschema) gesetzt ist. Zur Erzeugung eines Strings „Latitude, Longitude,“ wird die Konkatinationsanweisung aus dem CORE Namensraum verwendet.

Im Gegensatz zu den Koordinaten können Zeitintervalle direkt übernommen werden und erfordern lediglich eine simple Zuweisung der Elemente:

Begin = @BeginMin;

End = @EndMax;

The screenshot shows the DARIAH-DE Schema Registry Mapping-Editor interface. The top navigation bar includes 'DARIAH-DE', 'Schema Registry', 'Language', and 'Logout'. The main header features the 'Schema Registry' logo and the title 'Mapping-Editor'. Below the header, the breadcrumb trail reads 'Schema Registry / Schemas und Mappings / Mapping-Editor'. The main content area is titled 'Quellschema: Pergamon Zielschema: KML_Geobrowser' and includes 'Bearbeiten' and 'Löschen' buttons. The interface is divided into two main panels: 'Beispieltransformation' and 'Elementmodell'. The 'Beispieltransformation' panel shows a table of results with columns for 'ROW' and various fields like 'MODID', 'RECORDID', 'PSKeramikID', 'AnzahlFuß', 'AnzahlScherben', 'AnzahlScherbeninsgesamt', 'AufbewahrungOrt', 'AufbewahrungStatus', and 'AutoObjektkenung'. The 'Elementmodell' panel displays a hierarchical diagram of the KML schema, with elements like 'Placemark', 'Name', 'Point', 'Coordinates', 'TimeSpan', 'Begin', and 'End'. A purple line indicates a mapping from the 'AutoObjektkenung' field to the 'Begin' element within the 'TimeSpan' element of the 'Coordinates' element of the 'Point' element of the 'Placemark' element. The footer includes '© DARIAH-DE' and links for 'Datenschutz', 'Impressum', and 'Kontakt'.

Abbildung 12: Mappings von Feldern zwischen Pergamon- und KML Schema

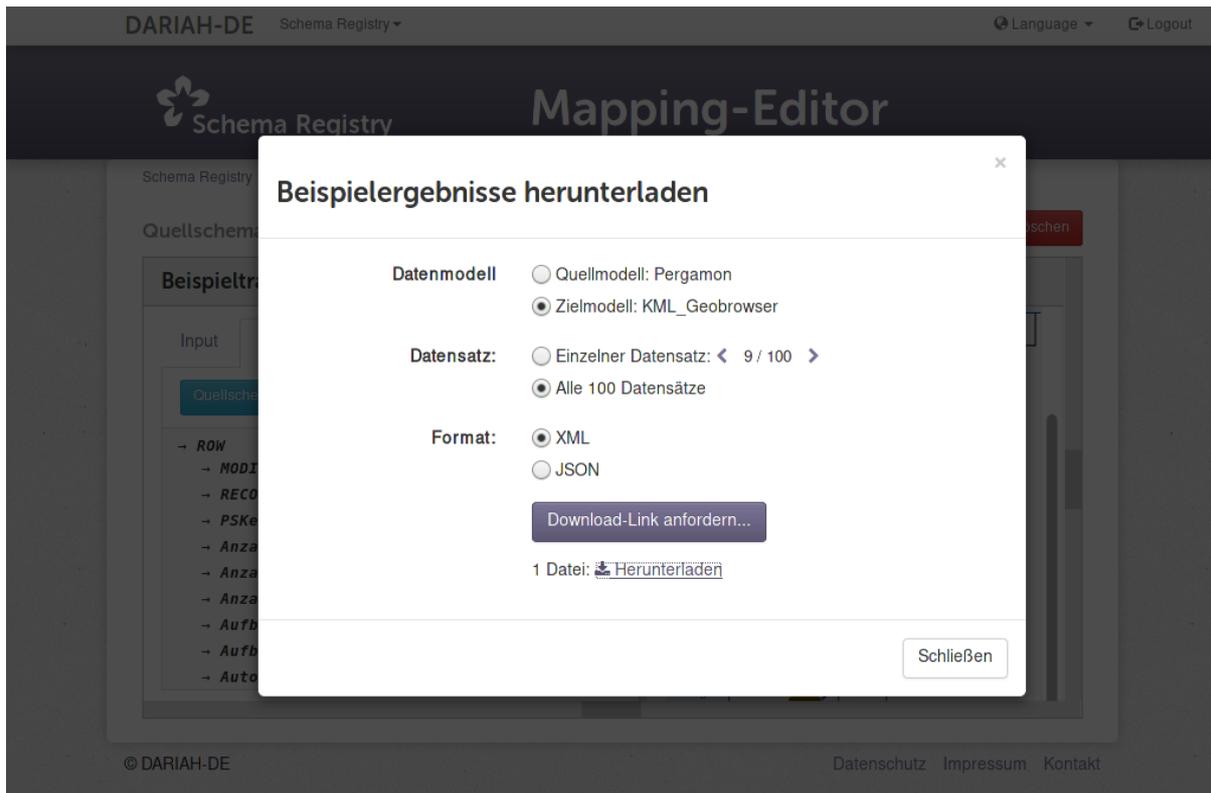


Abbildung 13: Herunterladen der Mapping-Ergebnisse

2.2.5. Visualisierung der Mapping-Ergebnisse

Sobald das Mapping der Daten ausgeführt ist, können die Daten in XML oder JSON heruntergeladen werden (vgl. Abbildung 13).²⁰ Die generierte KML Datei kann dann für den Geo-Browser hochgeladen werden. Zur Zeit ist die einzige manuelle Veränderung der Daten die Entfernung der XML Namespaces im KML. Ohne diese Änderung würden die Daten zwar hochgeladen werden, aber es würde nichts auf der Karte und auf der Timeline angezeigt werden.

Wie in Abbildung 14 abgebildet, können die 100 archäologischen Datensätze auf einer historischen Karte (the Barrington Atlas map of the Roman Empire) dargestellt werden. Nur 16 der 100 Datensätze haben Ortsangaben und können deshalb positioniert werden (Pergamon: 15, Knidos: 1), während fast alle eine Zeitangabe aufweisen. Die Möglichkeit, eine historische Karte auszuwählen, bietet einen zusätzlichen Nutzen, da sie eine bessere Kontextualisierung der Daten ermöglicht und ist sehr gelungen umgesetzt.

²⁰ Es dauert zur Zeit ungefähr eine Minute, das Mapping von 100 Datensätzen aus der Pergamon-Datenbank nach KML auszuführen.

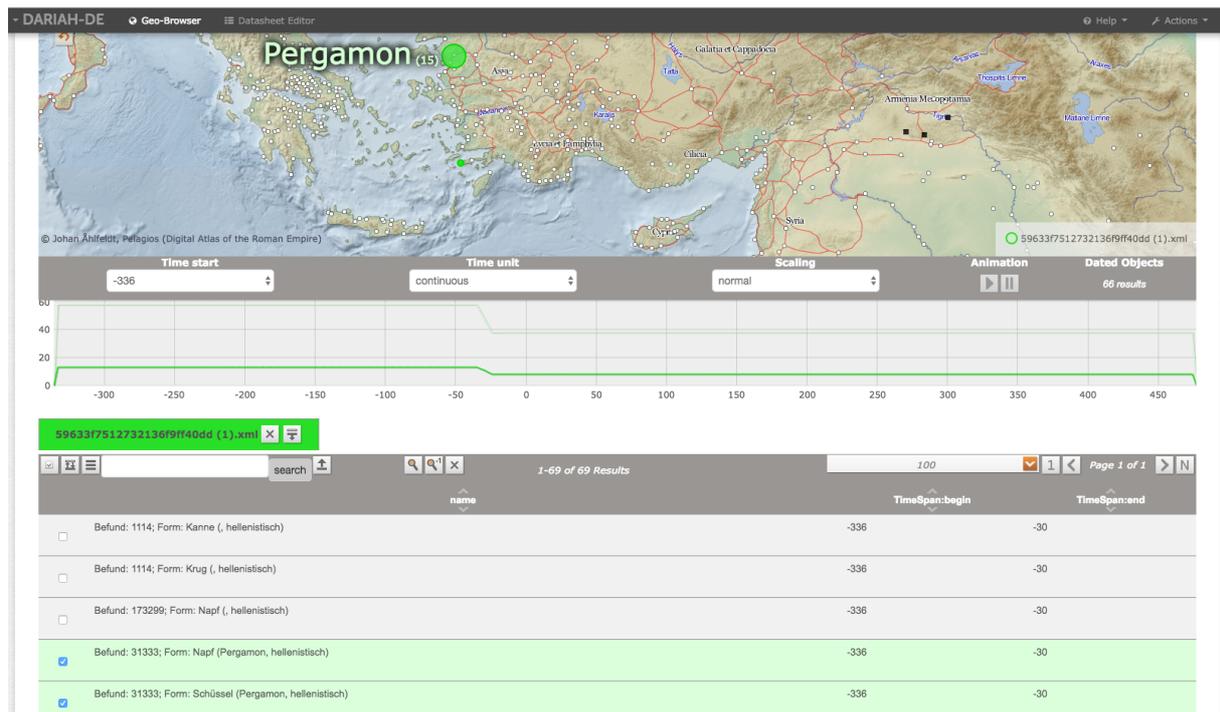


Abbildung 14: Visualisierung der Mapping-Ergebnisse mittels des DARIAH-DE Geo-Browsers

Obwohl dieser Use-Case mit einer relativ geringen Menge von Daten umgesetzt wurde, hat er deutlich gezeigt, dass das DARIAH-DE Data Modeling Environment flexibel genug ist, um mit archäologischen Daten umgehen zu können.

3. Technische Rückmeldungen

Die Umsetzung beider Anwendungsfälle werden mit Hilfe verschiedener DARIAH-DE Komponenten unterstützt. So wurde in Use-Case 2 (Archäologie) vorwiegend das DARIAH-DE DME untersucht, dieses zielt jedoch im Hinblick auf die konkrete Transformation von Daten lediglich auf Beispieldaten ab. Für eine stabile, infrastrukturelle Integration von Daten (wie diese Beispielsweise für Use-Case 1 – MWW erforderlich ist), kann auf weitere Komponenten der DARIAH-DE Föderationsarchitektur zurückgegriffen werden.

Einträge in der Collection Registry ersetzen (aus einer rein technischen Betrachtung) so beispielsweise die manuelle Eingabe von Daten. Durch einen Verweis auf die Quelle originärer Daten und die Assoziation mit einem oder mehreren Schemata im DME können Dienste wie die DARIAH-DE Generische Suche oder der derzeit entwickelte Transformationsdienst stets auf aktuelle Stände von Daten und ihrer Modelle zurückgreifen.

Eine nachhaltige und sichere Speicherung von Daten kann durch Verwendung des DARIAH-DE Repository²¹ erreicht werden, welches selbst über Schnittstellen mit weiteren Komponenten der DFA interagiert.²²

Für die spezifische Umsetzung von Use-Case 2 (Archäologie) in der DARIAH-DE DME war die Implementierung von Adaptern der DAI Schnittstellen erforderlich, um die unter 2.2.3 verwendeten API Aufrufe anbieten zu können. Die generische Umsetzung erlaubt es weiteren Use-Cases diese Adapter (wie auch die unterschiedlichen weiteren, die bereits zur Verfügung stehen) nachzunutzen. Verschiedene weitere Funktionalität der DME z. B. zum direkten Export transformierter Ergebnisse in XML wurden so implementiert, dass diese generisch für weitere Datenmodelle und Mappings nachgenutzt werden können. So wird erreicht, dass mit zunehmender Zahl von Use-Cases die Zahl der funktionalen Erweiterungen der DME zunimmt und mittelfristig die Umsetzung einer Vielzahl weiterer, komplexer Use-Cases ermöglicht werden kann.

²¹ <https://de.dariah.eu/repository>

²² DARIAH-DE Schnittstellenkonzept (M4.3.1):

https://wiki.de.dariah.eu/download/attachments/14651583/M4.3.1_final.pdf?version=1&modificationDate=1488529263513&api=v2

4. Literaturverzeichnis

Cuy, Sebastian; Gerth, Philipp; Heiden, Maximilian; Kolbmann, Wibke; Schmidle, Wolfgang (2014): iDAI.gazetteer – ein Referenzsystem für altertumswissenschaftliche Ortsinformationen als Teil einer digitalen Forschungsinfrastruktur. In *Kölner und Bonner Archaeologica* 4, S. 203-212.

Gradl, T.; Henrich, A. (2016): Die DARIAH-DE-Föderationsarchitektur: Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen, *Bibliothek Forschung und Praxis*. Band 40, Heft 2, Seiten 222-228, ISSN (Online) 1865-7648, ISSN (Print) 0341-4183, DOI: <https://doi.org/10.1515/bfp-2016-0027>, Juli 2016

Schmunk, S.; Funk, S. (2016): Das DARIAH-DE- und das TextGrid-Repositorium: Geistes- und kulturwissenschaftliche Forschungsdaten persistent und referenzierbar langzeitspeichern. *Bibliothek Forschung und Praxis*, Band 40, Heft 2, Seiten 213-221, ISSN (Online) 1865-7648, ISSN (Print) 0341-4183, DOI: <https://dx.doi.org/10.1515/bfp-2016-0020>, Juli 2016

5. Abbildungsverzeichnis

Abbildung 1: Bestandsübergreifende Suche in den Sammlungen des Forschungsverbundes MWW	7
Abbildung 2: Ausschnitt aus der MWW-Instanz der Collection Registry	8
Abbildung 3: Die Eingabemaske für Keramik-Objekte in der iDAI.Field Software	12
Abbildung 4: Import eines Schemas in das Data Modeling Environment	13
Abbildung 5: Markierung eines Elements des Schemas als Datensatz-Basis.....	14
Abbildung 6: Hochladen einer Beispieldatei	15
Abbildung 7: Import des Pergamon-Schemas in das Data Modeling Environment ...	15
Abbildung 8: JSON-Antwort für http://chronontology.dainst.org/data/period/?q=kaiserzeitlich	17
Abbildung 9: Erweiterung originärer Daten um Ergebnisse der Chronontology API .	17
Abbildung 10: Erweiterung originärer Daten um Ergebnisse der Gazetteer API	19
Abbildung 11: Erstellung eines Mappings zwischen Quellschema und Zielschema .	20
Abbildung 12: Mappings von Feldern zwischen Pergamon- und KML Schema	21
Abbildung 13: Herunterladen der Mapping-Ergebnisse	22
Abbildung 14: Visualisierung der Mapping-Ergebnisse mittels des DARIAH-DE Geo- Browsers.....	23