



Normdatensätze (M 4.3.1)

Version 28.08.2014

Cluster 4

Verantwortlicher Partner SUB, IEG, DAI, BBAW

DARIAH-DE **Aufbau von Forschungsinfrastrukturen** **für die e-Humanities**

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: Final

Verfügbarkeit: öffentlich

Autoren:

Jenny Oltersdorf, Stefan Funk, Anna Aurast, Marcus Held, Michael Piotrowski, Stefan Schmunk, Christoph Plutte

Revisionsverlauf:

Datum	Autor	Kommentare
15.07.2014	Jenny Oltersdorf	Gliederung, erster Textentwurf
15.08.2014	Jenny Oltersdorf	Inhaltliche Ergänzung Normdaten in DARIAH
21.08.2014	Stefan E. Funk	Inhaltliche Ergänzung Geo-Browser, PND, TGN
26.08.2014	Anna Aurast, Marcus Held, Michael Piotrowski	Inhaltliche Ergänzung Labeling System
27.08.2014	Stefan Schmunk	Inhaltliche Ergänzungen CERL Thesaurus
28.08.2014	Jenny Oltersdorf	Finale Redaktion

Inhaltsverzeichnis:

1. Einleitung	3
2. Normdaten in DARIAH-DE	4
3. Bisher durchgeführte Arbeiten	6
4. Bisherige Services	10
5. Geplante Arbeiten	10

1. Einleitung

Die Arbeiten im Bereich Normdaten gehen von Erkenntnissen der DARIAH-DE Projektphase I aus. Hier konnte bereits gezeigt werden, dass die gegenwärtig existierenden Normdatensätze und Vokabulare den individuellen Ansprüchen geistes- und kulturwissenschaftlich Forschender häufig nicht gerecht werden. So sind beispielsweise in der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek (DNB) Personengruppen nicht verzeichnet, deren Publikationen während des Nationalsozialismus aus dem Bibliotheksbestand in Deutschland entfernt wurden. Sie konnten bei der Erstellung der GND nicht berücksichtigt werden. Daher sind beispielsweise die zentralen Protagonisten der jüdischen Aufklärung im 18. Jahrhundert nur fragmentarisch erfasst. Eine weitere Schwierigkeit stellt die Arbeit mit historischen Orts- und Raumtypen dar, da diese im Getty Thesaurus of Geographical Names (TGN) und von GeoNames nur sehr bedingt durch die Ortstypenklassen bedient werden.

Eine zentrale Aufgabe des DARIAH-DE Cluster 4 „Wissenschaftliche Sammlungen“ ist es darum, prototypische Nutzungsszenarien, insbesondere im Bereich Orte und Personen, zusammenzustellen, den aktuellen Bestand relevanter Vokabulare für geisteswissenschaftliche Forschung zu erfassen und ein konkretes Vokabular zu historischen Ortstypen zu entwickeln. Zusätzlich werden die bestehenden Normdatendienste hinsichtlich ihrer Anwendung und Einbindung in die DARIAH-DE Forschungsinfrastruktur untersucht und ggf. entsprechend angepasst.

In die Arbeit sind neben der *Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB)*, das *Deutsche Archäologische Institut* in Berlin (DAI), das *Leibniz-*

Institut für Europäische Geschichte in Mainz (IEG) und die *Berlin Brandenburgische Akademie der Wissenschaften* (BBAW) involviert.

Die Ergebnisse sind im vorliegenden Milestone *M 4.3.1 – Normdatensätze* zusammengefasst. Der Bericht stellt kurz die Anwendungsbereiche für Normdaten inklusive der vorhandenen Services in DARIAH-DE dar und schildert die bereits durchgeführten sowie die geplanten Arbeiten. Es ist vorgesehen, im November 2014 einen zusätzlichen Bericht zu publizieren, der über den aktuellen Implementierungsstand Auskunft gibt.

2. Normdaten in DARIAH-DE

Bisher werden Normdaten für unterschiedliche Funktionalitäten im *Geo-Browser* und der vom Salomon Ludwig Steinheim Institut für deutsch-jüdische Studien (STI) entwickelten Web-App *Orte jüdischer Geschichte* benötigt.¹ Zusätzlich nutzt das Projekt *TextGrid* die *Personennamendatei* (PND) zum Zweck der Autovervollständigung von Personennamen. Ein wünschenswertes weiteres Einsatzgebiet für Normdaten ist die Schlagwortvergabe bei der Neuanlegung von Sammlungen in der DARIAH-DE Collection Registry und im DARIAH-DE Repository.² Dies betrifft vor allem die Normdaten der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek. Die GND enthält Datensätze für Personen, Körperschaften, Kongresse, Geografika, Sachschlagwörter und Werktitel, die bis April 2012 in den getrennten Normdateien Gemeinsame Körperschaftsdatei (GKD), Personennamendatei (PND) und Schlagwortnormdatei (SWD) sowie der Einheitssachtitel-Datei des Deutschen Musikarchivs (DMA-EST-Datei) erfasst wurden.³ Im Datasheet-Editor des Geo-Browsers⁴ werden neben Daten des Getty-Thesaurus zusätzlich Normdaten aus OpenGeoNames⁵ und OpenStreetMap⁶ für Ortsnamen genutzt. Die GND wird in diesem konkreten Fall nicht berücksichtigt, weil sie keine Georeferenzen enthält und Orte lediglich als Gebietskörperschaften verzeichnet werden.

¹ <https://de.dariah.eu/web-app-orte-juedischer-geschichte>

² <http://colreg.de.dariah.eu/colreg>

³ Vgl. http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html

⁴ <http://geobrowser.de.dariah.eu/edit/>

⁵ <http://www.geonames.org/>

⁶ http://www.openstreetmap.de/faq.html#wie_daten_nutzen

Einen weiteren Schwerpunkt im Bereich der Nutzung von Normdaten stellen die Arbeiten des IEG im Rahmen von DARIAH-DE dar. Insbesondere für die historische Forschung erweist sich die Nutzung von Normdaten und kontrollierten Vokabularen sowohl als ein wichtiges Ziel als auch als eine Herausforderung, da die gegenwärtig existierenden Normdatensätze und Vokabulare häufig nicht den individuellen Ansprüchen der historisch arbeitenden Wissenschaften genügen. Das IEG geht diese Problematik auf drei Ebenen an. Erstens untersucht es anhand vorhandener Bestände, welche Anforderungen historisch arbeitende Wissenschaften an Normdateien und kontrollierte Vokabulare stellen. Zweitens erstellt es prototypische Vokabulare für unterschiedliche Bereiche historischer Forschung wie Sozial-, Militär- oder Wirtschaftsgeschichte. Drittens unterstützt das IEG die kollaborative Erstellung, Aufbereitung und Nutzung von Vokabularen und Normdaten, indem es Methoden und Werkzeuge zur Erstellung von kontrollierten Vokabularen und Normdaten entwickelt und der wissenschaftlichen Community zur Verfügung stellt.

Im Zuge einer interdisziplinären Auseinandersetzung mit Vokabularen für historische Ortstypen im Rahmen der Freien Fachsäule in der ersten Projektphase von DARIAH-DE wurden die Anforderungen der historisch arbeitenden Wissenschaften an bestehende Normdaten evaluiert und gleichzeitig die Möglichkeiten und Anforderungen an den Aufbau eines kontrollierten Vokabulars für historische Ortstypen erforscht. Dabei wurde in Kooperation mit dem *i3mainz* das sog. *Labeling System* entwickelt. Das Labeling System ermöglicht die einfache Erstellung von anwendungsspezifischen SKOS-Vokabularen, die dennoch interoperabel sind, indem diese mit anderen Referenzthesauri und Normdaten in Relation gesetzt werden. Die Vokabulare können in den DARIAH-DE Services genutzt werden und leisten so einen Beitrag zum Aufbau und der Nutzung des DARIAH-DE Repository. Dieser Ansatz ist motiviert durch die in der ersten Phase von DARIAH-DE gewonnenen Erkenntnis, dass die Entwicklung eines *einzigsten* Vokabulars nicht praktikabel ist, da die Anforderungen unterschiedlicher Forschungsfragen an die gleichen Datensammlungen zu stark voneinander abweichen. Der mit dem Labeling System verfolgte Ansatz erlaubt es hingegen, die Vorteile projektspezifischer Vokabulare (schnelle Erstellung, bekannte Terminologie) mit denen standardisierter Vokabulare und Normdaten (Konsensbildung und Interoperabilität) zu verbinden.

Einen dritten Schwerpunkt im Bereich der Normdatennutzung bilden die Arbeiten zum Personendaten-Repository (PDR). Das PDR wird an der BBAW entwickelt, um prosopografische Forschungs- und sekundär Daten aus unterschiedlichen Forschungskontexten in einer gemeinsamen Infrastruktur zu vereinen. Im Repository werden die Daten zentral gespeichert und über das Anwenderprogramm *Archiv-Editor* bearbeitbar gemacht. Für die Integration der oftmals sehr heterogenen Daten spielen Normdaten und Authority Files wie GND, und VIAF (Virtual International Authority File) eine zentrale Rolle. Personendaten aus unterschiedlichen Forschungskontexten können mit Normdaten angereichert und anschließend zusammengeführt werden. So können Daten zu Personen, die aus unterschiedlichen Perspektiven erstellt wurden, zu gemeinsamen biografischen Beschreibungen kombiniert werden.

3. Bisher durchgeführte Arbeiten

Die Daten der PND und die für den Geo-Browser relevanten Normdaten werden derzeit in einer eXist-Datenbank verwaltet. Bisher erfolgt die Nutzung der Normdaten auf der Basis eines Dumps. D.h. die Daten wurden vollständig heruntergeladen und entsprechend des Zielsystems angepasst und eingebunden. Der Nachteil dieser Verfahrensweise liegt vor allem in dem hohen datenpflegerischen Aufwand für Aktualisierung und Betreuung. Aus diesem Grund wird in DARIAH-DE II ein performanter Webservice entwickeln, der die Nutzung von Normdaten erleichtert. Zu diesem Zweck bedarf es einer Schnittstelle, die gezielte Abfragen erlaubt. Im Falle der von der DNB erzeugten GND ist dies durch eine SRU Schnittstelle gewährleistet. SRU (Search / Retrieve via URL)⁷ ist ein standardisiertes Webservice-Protokoll. Die Ergebnisse der Suche werden in einem definierten XML-Format zur Verfügung gestellt. Das SRU-Protokoll ermöglicht die gezielte Suche mittels Suchindizes und Suchbegriffen und die Übernahme der entsprechenden Treffer in die eigene Umgebung.⁸

⁷ http://de.wikipedia.org/wiki/Search/Retrieve_via_URL

⁸ Vgl. http://www.dnb.de/DE/Service/DigitaleDienste/SRU/sru_node.html

Im Juni 2014 fand an der SUB Göttingen ein internes Arbeitstreffen statt, an dem eine Reihe weiterer SUB-KollegInnen, beispielsweise aus der Abteilung Metadaten (MET), teilnahmen. Ziel des Treffens war es, ein gemeinsames Vorgehen bei der Nachnutzung von Normdatendiensten zu koordinieren. Die KollegInnen aus der Abteilung MET nutzen derzeit einen Abzug der GND-Daten für ihre Arbeit (die Aktualisierungen erfolgen wöchentlich) und sind daher an einer engen Kooperation interessiert. Neben den Vor- und Nachteilen der Nutzung eines bereits existierenden GND-Service wurde die Möglichkeit diskutiert, einen eigenen Service aufzusetzen. Zur Nachnutzung bereits bestehender Services stehen derzeit eine SRU-Schnittstelle des Bibliotheksservice-Zentrums Baden-Württemberg (BSZ)⁹ sowie eine SRU Schnittstelle des Gemeinsamen Bibliotheksverbunds (GBV) zur Verfügung. Da die Herzog August Bibliothek Wolfenbüttel (HAB) seit Juni 2014 über einen Zugang zur SRU-Schnittstelle der DNB verfügt, wurde von den Kollegen ein erster Prototyp zur Übernahme von Metadaten entwickelt, der in den kommenden Wochen an die Anforderungen in den Projekten TextGrid und DARIAH-DE angepasst wird. Anschließend soll dieser Prototyp als Service auch weiteren Projekten zur Verfügung stehen. Seit August 2014 verfügt auch die SUB über einen SRU-Zugang zu den Normdaten der DNB. Der bestehende Prototyp wird daher an der SUB in Kooperation mit der HAB weiter entwickelt.

Die Entwicklung des Labeling Systems wurde, wie bereits oben angesprochen, im Zuge eines Forschungsvorhabens initiiert, das sich mit kontrollierten Vokabularen für historische Ortstypen befasste. Dabei wurde versucht, die Ortstypen über ihre Funktionen zu beschreiben, statt über ein spezifisches, kulturabhängiges Verständnis eines Begriffs. Durch dieses Vorgehen wurde es möglich, eine Umschreibung in natürlicher Sprache als eine im jeweiligen Forschungsgebiet gängige, aber selten eindeutige Bezeichnung mit Funktionen aus einem Referenzthesaurus zu verknüpfen, um so den damit bezeichneten Begriff formal zu definieren. Darüber hinaus kann das Labeling System auch in anderen Bereichen eingesetzt werden. Um beispielsweise bestimmte Personengruppen identifizieren zu können, braucht man ein kontrolliertes Vokabular für Tätigkeitsfelder, um von traditionellen Berufsbezeichnungen, wie sie etwa in Biografien verwendet

⁹ <https://wiki.bsz-bw.de/doku.php?id=v-team:daten:sru> und <http://swb.bsz-bw.de/sru/DB=2.104/username=/password=/>

werden, abstrahieren zu können. In Biografien werden Personen beispielsweise als »Minister«, »Abgeordneter«, »Jurist«, »Publizist«, »Staatsmann« oder »Revolutionär« bezeichnet; für ein konkretes Projekt (z. B. das Projekt *Cosmopolitanities*, das den Hintergrund für den Use Case 2 in Cluster 5 darstellt) kann es jedoch sinnvoll sein, alle diese Personen einem Tätigkeitsfeld »Politik« zuzuordnen zu können. Hier bietet sich der im Rahmen der Freien Fachsäule entwickelte Ansatz an, der im Labeling System implementiert wurde. Ein Beruf wird dabei als ein Etikett (»label«) für ein Bündel von atomaren Tätigkeiten aufgefasst, die in einem Referenzthesaurus definiert sind. Auf diese Weise lassen sich auch Gemeinsamkeiten zwischen traditionell unterschiedlichen, tatsächlich sich aber teilweise überlappenden Kategorien finden.

Durch den im Labeling System implementierten Ansatz wird sowohl die Kollaboration der Forschenden als auch die semantische Austauschbarkeit und Verankerung von Datensammlungen und -sätzen (Interoperabilität) vorangetrieben. Das Labeling System nutzt grundlegende Ideen des semantischen Webs für den Aufbau von kontrollierten Vokabularen, indem es dem Anwender erlaubt, Daten sowohl in menschenverständlicher als auch in maschinell verarbeitbarer Form zu repräsentieren. Dabei werden SKOS, RDF-Schemata und OWL-Relationen eingesetzt. Das Labeling System liegt als Softwareprototyp samt Dokumentation vor. Der Prototyp und die Dokumentation werden zurzeit für den Produktivbetrieb innerhalb der DARIAH-DE Infrastruktur vorbereitet.

Für die Arbeit mit Normdaten im Personendaten-Repository wurde eine Reihe von Diensten und Werkzeugen entwickelt. Für die Abfrage von Datensätzen des PDR anhand ihrer Normdaten wie GND oder VIAF wurde das *ID-Interface* als Webservice erstellt.¹⁰ Es bietet statische Links auf Datensätze mittels GND an und liefert ein Datenblatt zur referenzierten Person im Format XML mit Einträgen zu allen vorhandenen Kategorien wie z.B. Lebensdaten, Geschlecht, Werdegang, Beziehungen etc. Die zurückgegebenen Daten zu einer Person können für die weitere Verarbeitung und Anzeige verwendet werden.

¹⁰ <https://pdrprod.bbaw.de/wiki/doku.php?id=de:doc:connect>

Die Anreicherung von Personendaten mit Normdaten spielt eine wichtige Rolle, da oftmals prosopografische Daten nicht oder nur zum Teil mit Identifikatoren ausgezeichnet sind, wenn sie in das Repositorium eingepflegt werden. Hierfür wurden im PDR zwei Werkzeuge entwickelt, die unterschiedliche Strategien verbinden. Zum einen wird mit dem Werkzeug *Concordancer*¹¹ eine automatische Anreicherung mit Normdaten verfolgt, zum anderen wird mit dem grafischen Anwenderprogramm *Investigator*¹² ein spielerischer Weg zur manuellen Auszeichnung von Datensätzen mit Normdaten untersucht. Der Concordancer ist als Webservice realisiert und erlaubt es, Anfragen nach Normdaten zu übergebenen Personendaten zu stellen. Als Parameter können Namen, Lebensdaten, Lebensorte und allgemeine Berufsbezeichnungen genutzt werden. Mit diesen Daten wird eine Datenbank mit Personendaten durchsucht, die Datensätze, die eine bestimmte Übereinstimmung mit den übergebenen Parameter aufweisen, werden mit ihren Normdaten als Antwort zurückgegeben. Als Datengrundlage für den Datenabgleich dient derzeit die *DBpedia*.¹³

Da jedoch nicht alle Datensätze automatisch mit Normdaten angereichert werden können, weil entweder keine Normdaten zu bestimmten Personen vorliegen oder aber keine ausreichend trennscharfe Identifikation der Person vorgenommen werden kann und daher nicht mit genügender Sicherheit die Übereinstimmung der Person mit einem vorhandenen Normdatensatz bestimmt werden kann, ist eine manuelle Anreicherung häufig unverzichtbar. Beim manuellen Anreichern handelt es sich um vergleichsweise repetitive Tätigkeiten. Um diese für ForscherInnen interessanter zu gestalten, wurde vom PDR ein Experiment mit einem spielerischen Ansatz begonnen. Das Webprogramm Investigator bietet einen spielerischen Ansatz zur Anreicherung von Personendaten mit GNDs und unterstützt eine kompetitive Herausforderung bei der Arbeit. Das Programm wurde entwickelt und erste interne Experimente wurden durchgeführt.

¹¹ <https://pdrprod.bbaw.de/wiki/doku.php?id=de:doc:webservices>

¹² <http://pdrprod.bbaw.de/gnd/>

¹³ <http://de.dbpedia.org/>

4. Bisherige Services

Die beiden aufgesetzten Services, die momentan die Daten der PND sowie des Getty Thesaurus zur Verfügung stellen, basieren auf einer Stand-Alone-Applikation zweier eXist-Datenbanken, die jeweils mit Abzügen der jeweiligen Daten gefüttert wurden. Sie stehen beide auf einer virtuellen Maschine von DARIAH-DE zur Verfügung (ref.dariah.eu). Für eine Aktualisierung müssen diese Dumps neu eingespielt und evtl. nachbearbeitet (Konvertierung von RDF nach XML) werden. Beide Services können per XQuery abgefragt werden¹⁴ Die Ergebnisse werden zurzeit in der Metadaten-Eingabemaske des TextGridLab¹⁵ (PND), dem Web-App Orte jüdischer Geschichte (TGN), sowie im Geo-Browser zur Ortsbestimmung genutzt (TGN).

5. Geplante Arbeiten

Es ist geplant, den von den Mitarbeitern der HAB entwickelten Prototyp zu testen und an die DARIAH-DE Infrastruktur anzupassen. Die Implementierung wird in einem zusätzlichen Bericht (geplant für November 2014) dokumentiert. Weitere geplante Arbeiten umfassen sowohl die Nutzung dieses Prototyps als produktive Instanz einer GND-Abfrage für die Nutzung in TextGrid und DARIAH-DE, als auch die weitere Nutzung des TGN für beide Projekte. Da in der GND keine Ortskoordinaten hinterlegt sind, wird von DARIAH-DE weiterhin eine Service-Instanz des TGN gehostet, die vorerst auch weiterhin aktualisiert wird. Die Nutzung der von der *The J. Paul Getty Trust*¹⁶ seit kurzem zur Verfügung gestellten SPARQL-Schnittstelle zum TGN¹⁷ wird geprüft und evaluiert, so dass zukünftig evtl. die Bereitstellung des TGN seitens DARIAH-DE entfallen kann. Für die Ortsnamen-Ergänzung im Geo-Browser¹⁸, bzw. im dazugehörigen Datasheet Editor¹⁹, ist geplant, beispielsweise in OpenGeoNames noch nicht existierende historische Ortsnamen von den NutzernInnen zunächst in einer eigenen Instanz zu sammeln und sie dann in den Datenbestand des OGN zu

¹⁴ Beispiele für TGN- und PND-Abfragen: <http://ref.dariah.eu/tgnsearch/tgnquery.xql?ac=Bam> (TGN), und <http://ref.dariah.eu/pndsearch/pndquery.xql?ln=Ringeln> (PND)

¹⁵ Vgl. TextGridLab. Download und Installation. <http://textgrid.de/registrierungdownload/download-und-installation/>

¹⁶ Vgl. The Getty. <http://www.getty.edu/>

¹⁷ Vgl. The Getty. SPARQL Query. <http://vocab.getty.edu/sparql>

¹⁸ Vgl. Geo-Browser. <http://geobrowser.de.dariah.eu/>

¹⁹ Vgl. Datasheet Editor. <http://geobrowser.de.dariah.eu/edit/>

überführen. Weiterhin soll der Geo-Browser zukünftig auch die Daten von OpenStreetMap und OpenGeoNames direkt abfragen und übernehmen können, so wie es für den TGN bereits implementiert ist.

In Zusammenarbeit mit den Kollegen der Abteilung MET der SUB ist angedacht, Daten (Personen und Körperschaften) in den CERL-Thesaurus zu überführen, sofern sie der inhaltlichen Ausrichtung des Thesaurus entsprechen. Dies betrifft folglich vor allem Datensätze, die thematisch und inhaltlich in der Frühen Neuzeit zu verorten sind.²⁰ Die SUB ist Vollmitglied im CERL-Konsortium und wird dies mit ihren entsprechenden Fachabteilungen unterstützen.²¹

Am IEG werden neben der Dokumentation von Einsatzszenarien für kontrollierte Vokabulare in der historischen Forschung in bestimmten Bereichen exemplarisch kontrollierte Vokabulare aufgebaut. Weiterführende Arbeiten am Prototyp des Labeling System und die Erweiterung der Dokumentation werden in nächster Zeit abgeschlossen sein. Geplant ist, das Labeling System als Webapplikation in den Produktivbetrieb innerhalb der DARIAH Infrastruktur zu integrieren, neue Funktionen zu entwickeln und bereitzustellen sowie exemplarische Anwendungsszenarien zu erarbeiten, um sie als generische Lösungsansätze für den gesamten Bereich der geistes- und kulturwissenschaftlichen Forschung nutzbar zu machen.

Das Ineinandergreifen von technischer Innovation und geisteswissenschaftlicher Forschung wird durch die Einbeziehung des vom IEG in Zusammenarbeit mit MInf-BA bearbeiteten Use Case 2 aus Cluster 5 gewährleistet, der sich mit der Mobilität bestimmter Personengruppen im 19. Jahrhundert beschäftigt. Eine Zielstellung ist daher, die Entwicklung eines kontrollierten Vokabulars für Tätigkeitsfelder mit Hilfe des Labeling System voranzutreiben. Der Fokus liegt dabei auf Personen in deutschen und ausgewählten europäischen Nationalbiografien; zeitlich wird das Vorhaben auf das 19. Jahrhundert beschränkt. In einem weiteren Schritt soll ein kontrolliertes Vokabular für historische Ortstypen erstellt werden, das auf die Anforderungen des Forschungsprojekts zugeschnitten und mit dem Vokabular für Tätigkeitsfelder inhaltlich verknüpft ist. Das Labeling System ist nicht auf die hier zu bearbeitenden Anwendungen beschränkt, sondern prinzipiell auf beliebige andere

²⁰ <http://cerl.sub.uni-goettingen.de/cerl/>

²¹ http://www.cerl.org/resources/links_to_other_resources/main

Szenarien übertragbar, in denen projektspezifische kontrollierte Vokabulare benötigt werden, die gleichzeitig interoperabel sind.

Für die Verbesserung der Integration von Personendaten und Normdaten im PDR werden die Werkzeuge Investigator und Concordancer weiter ausgebaut. Dabei soll insbesondere der Matching-Algorithmus des Concordancer verbessert werden. Dieser Algorithmus dient der Messung des Grades der Übereinstimmung von angefragter Person mit vorhandenen Datensätzen und steht im Zentrum der Berechnung von Konkordanzen und Identifizierungen von Personen mit Normdaten. Die Experimente mit dem Investigator sollen weiterverfolgt und das Programm entsprechend den Nutzererfahrungen verbessert werden. Neben der Arbeit mit Personennormdaten im Archiv-Editor soll im Rahmen der DARIAH-Arbeiten untersucht werden, wie der Archiv-Editor künftig auch mit geografischen Normdaten arbeiten kann. Ziel ist es, Werkzeuge für die Anreicherung von Personendaten mit geografischen Normdaten zu entwickeln, die als Vorbereitung gedacht sind um Personendaten und GeoBrowser zu verbinden.