

Aufbau einer Infrastruktur zur Implementierung sachthematischer Zugänge im Archivportal-D am Beispiel des Themenkomplexes „Weimarer Republik“

Antrag auf Förderung eines Vorhabens im Bereich „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS), Förderprogramm „e-Research-Technologien“

Antragstellende Personen:

Prof. Dr. Robert Kretzschmar (Stuttgart), Dr. Michael Hollmann (Koblenz), Dr. Elisabeth Niggemann (Frankfurt am Main), Matthias Razum (Eggenstein-Leopoldshafen)

Beschreibung des Vorhabens

1 Ausgangslage und eigene Vorarbeiten

Sachthematische Zugänge zu Kulturgut in übergreifenden Portalen mit qualifizierten themenbezogenen Recherchemöglichkeiten sind aus wissenschaftlicher Sicht seit langem ein Desiderat. Die Präsentation von Objekten in neuem Kontext führt im Sinne des Semantic Web zu alternativen Sinnzusammenhängen und eröffnet weitere Perspektiven zu bereits bestehenden und sich durch die alternative Darstellung neu ergebenden Forschungsfragen.

Daneben existiert aber auch vonseiten der bereitstellenden Kultureinrichtungen ein Bedarf an einer thematisch strukturierten und kontextualisierten Präsentation von zu bestimmten Sachthemen vorhandenen oder anlassbezogen aufbereiteten Inhalten. Für die Archivsparte dient diese auch der Positionierung innerhalb der wissenschaftlichen Informationsinfrastruktur, in der gemeinsame Forschungsvorhaben z.B. zu historischen Jubiläen eine bedeutende Rolle spielen. Ein konkreter Anlass besteht etwa im für die deutsche Geschichte zentralen Themenkomplex „Weimarer Republik“, für den in den kommenden Jahren ein verstärktes Forschungsinteresse antizipiert wird, weshalb hierzu in mehreren Archivverwaltungen bereits Digitalisierungsvorhaben bearbeitet werden - so in den Projekten „Weimar – die erste deutsche Demokratie“ im Bundesarchiv und „Von der Monarchie zur Republik – Digitalisierung von Quellen des Landesarchivs zur Demokratiegeschichte“ im Landesarchiv Baden-Württemberg (LABW).¹

Benötigt wird daher eine Möglichkeit, die Ergebnisse der o.g. Projekte sowie künftig weiterer verwandter Vorhaben von Archiven gemeinsam themenbezogen im übergreifenden Portalkontext zugänglich und recherchierbar zu machen.² Als bereits bestehende, archivspartenübergreifend und flächendeckend angelegte deutschlandweite Plattform bietet das Archivportal-D für die Implementierung der benötigten Funktionalitäten die besten Voraussetzungen.

¹ Vgl. hierzu die untenstehenden Projektbeschreibungen.

² Ein gesteigertes wissenschaftliches Interesse, das Archive bedienen möchten, wird zudem im Zuge der 100. Jährung der NS-Zeit erwartet.

Projekt "Aufbau eines Archivportals-D" [Landesarchiv Baden-Württemberg (LABW), FIZ Karlsruhe, Deutsche Nationalbibliothek (DNB)]

Das Archivportal-D wurde 2012-2017 in zwei DFG-finanzierten Förderphasen als Teilprojekt der Deutschen Digitalen Bibliothek (DDB) entwickelt.³ Seit seinem erfolgreichen Start im Herbst 2014 hat es sich inzwischen mit über 100 beteiligten Archiveinrichtungen und ca. 11 Mio. Verzeichnungseinheiten als zentrales nationales Nachweissystem von Archivgut und Archivinformationen⁴ deutscher Archive etabliert. Sowohl seitens wissenschaftlicher Nutzer und Nutzerinnen⁵ als auch teilnehmender Archive erfährt das Portal eine hohe Akzeptanz als nachhaltige Plattform zur Präsentation von Erschließungsleistungen und digitalisiertem Archivgut, die dank ihrer kontinuierlich wachsenden Datenbasis immer wertvoller und nützlicher wird.⁶

Die Zukunft und Nachhaltigkeit des Angebots ist dabei organisatorisch abgesichert: Die Deutsche Digitale Bibliothek und das Landesarchiv Baden-Württemberg verständigten sich 2016 in einem „Memorandum of Understanding“ auf den gemeinsamen Weiterbetrieb des Archivportals-D nach Ende der DFG-Förderung. Neben dem damit gewährleisteten Betrieb des Portals soll das Angebot nach Bedarf über weitere Projektfinanzierungen funktional weiterentwickelt werden, um bestehende und neue Anforderungen insbesondere der wissenschaftlich Nutzenden umzusetzen. Hierzu zählt die Implementierung thematischer Zugänge im Portal.⁷

Deutsche Digitale Bibliothek [Deutsche Nationalbibliothek (DNB), FIZ Karlsruhe, Landesarchiv Baden-Württemberg (LABW), Bundesarchiv (BArch)]

Die Deutsche Digitale Bibliothek (DDB) verfolgt das Ziel, das Kultur- und Wissenserbe aus Deutschland in digitaler Form zusammenzuführen und über ihre Angebote an unterschiedliche Zielgruppen zu vermitteln. Dazu gehören neben dem DDB-Portal vor allem Anwendungen, die auf dem API der DDB basieren, wozu insbesondere auch das Archivportal-D zählt. Derzeit weist die DDB mehr als 22 Mio. Objekte aus ca. 350 Einrichtungen aller Kultursparten nach. Alle vier Antragsteller sind aktive Mitglieder im Kompetenznetzwerk "Deutsche Digitale Bibliothek" mit derzeit 14 Mitgliedern. Der DDB-Geschäftsbereich Technik, Entwicklung, Service mit der übergreifenden Projektkoordination und -steuerung sowie der Servicestelle ist an der DNB angesiedelt. Hier werden die konzeptionelle Weiterentwicklung und die Betreuung der Datenpartner verantwortet. FIZ Karlsruhe ist der technische Betreiber der DDB und darüber hinaus verantwortlich für die Softwareentwicklung eines Großteils der DDB-Komponenten. LABW und BArch betreiben gemeinsam die Fachstelle Archiv mit den Schwerpunkten Community-Arbeit und Betreuung der archivischen Datenpartner.

³ Ausführliche Projektbeschreibung: <https://www.landearchiv-bw.de/web/54267>.

⁴ Seit Ende 2016 sind die Adressdaten von über 2.500 Archiven im Archivportal-D recherchierbar. Damit wird zum ersten Mal ein bundesweites Verzeichnis deutscher Archive online kostenlos zur Verfügung gestellt: <https://www.archivportal-d.de/info/aktuelles/Archivverzeichnis>. Zur zentralen Rolle des Archivportals-D für Archive und Archivnutzer in Deutschland vgl. Frank M. Bischoff, Archive, in: Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, Hrsg. von Laura Busse, Wilfried Enderle, Rüdiger Hohls, Gregor Horstkemper, Thomas Meyer, Jens Prellwitz, Annette Schuhmann, Berlin 2016 (=Historisches Forum, Bd. 19), <http://www.clio-online.de/guides/sammlungen/archive/2016>.

⁵ Dies zeigten u.a. die positiven Reaktionen bei der Vorstellung des Archivportals-D auf dem Deutschen Historikertag 2016 innerhalb der Sektion "Recherche und Weiterverarbeitung. Digitale Angebote der Archive für die historische Forschung im Netz". Als Vertreter der historischen Forschung kommentierte Rüdiger Hohls die außerordentlich gut besuchte Sektion. "Er bezeichnete das Archivportal-D als sehr gelungen, die Zusammenführung von Informationen aus verschiedenen Bereichen als Quantensprung. Der archivübergreifende Zugriff auf Findmittel unterstützt eine effizient geplante Forschungsreise." (Auszug aus einer Zusammenfassung zur Sektion von Gudrun Fiedler, Leiterin des Standort Stade des Niedersächsischen Landesarchivs, abrufbar unter: <http://www.hsozkult.de/conferencereport/id/tagungsberichte-6784>).

⁶ Das Archivportal-D verzeichnet täglich im Schnitt ca. 650 Nutzungen (Stand Mai 2017). Die Nutzerzahlen sind seit dem Portal-Launch 2014 stark angestiegen. Für 2017 ist wieder mit einem deutlichen Zuwachs zu rechnen, da bereits zur Mitte des Jahres die Gesamtnutzungszahl von 2016 erreicht sein wird (109.151 Visits). Ein sehr großer Anteil der Nutzer kommt dabei über Suchmaschinen auf das Portal-Angebot, jedoch finden sich bei den weiteren Verweis-Quellen viele dezidiert wissenschaftliche Einrichtungen bzw. deren Informationssysteme und Kataloge.

⁷ Vgl. Becker/Maier/Uhde/Wolf (2016), Wolf (2016), S. 47-63 und Reisacher/Krauth (2015), S. 36 f.

Die DDB bildet das organisatorische und technische Rückgrat für das Archivportal-D: Sie stellt dafür nicht nur die zentrale Datenhaltung bereit, sondern übernimmt auch die Prozesse zur Zusammenführung und Integration der Datenbestände, die im Archivportal-D sichtbar werden (insbesondere Akquise, Datenclearing, Transformation, Review, Ingest und Aktualisierung).

Projekt „Von der Monarchie zur Republik“ (LABW)

Von 2015 bis November 2017 führt das Landesarchiv Baden-Württemberg in Kooperation mit dem Lehrstuhl für Landesgeschichte des Historischen Instituts der Universität Stuttgart ein vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg finanziertes Forschungsprojekt zur Demokratiegeschichte im deutschen Südwesten durch. Durch die wissenschaftliche Aufarbeitung und Digitalisierung unterschiedlichster Quellen sollen die gewaltigen politischen und gesellschaftlichen Transformationsprozesse ab den letzten Kriegsmonaten 1918 bis etwa 1923 veranschaulicht und ein Anstoß für die weitere Erforschung dieser Umbruchphase gegeben werden.

Nach fachlichen und wissenschaftlichen Kriterien wurden aus den Beständen des LABW projektrelevante Quellen ausgewählt, digitalisiert und mit den zugehörigen Erschließungsinformationen verknüpft. Diese sind mit Personen- und Ortsnormdaten angereichert und wurden jeweils Begriffen einer projektspezifischen Sachklassifikation zugeordnet, um die Quellen für eine themenbezogene Präsentation aufzubereiten. Die Ergebnisse werden Ende 2017 sowohl als Quellensammlung im Online-Informationssystem des Landesarchivs als auch in einem neu entwickelten Themenschwerpunkt im landeskundlichen Informationssystem LEO-BW⁸ online bereitgestellt. Angestrebt wird darüber hinaus eine spätere Vernetzung der Inhalte mit den Angeboten anderer Einrichtungen durch die Präsentation der Bestände im Archivportal-D.

Projekt „Weimar – die erste deutsche Demokratie“⁹ (BArch)

Seit November 2015 bis voraussichtlich Dezember 2020 baut das Bundesarchiv im Rahmen des Projektes „Weimar – Die erste deutsche Demokratie“ ein themenbezogenes Webangebot mit Quellen des Bundesarchivs zur Weimarer Republik auf. Anlässlich kommender Jahrestage knüpft das Bundesarchiv damit an die erfolgreiche Präsentation von Archivgut in Quellenportalen an¹⁰ und macht eine Auswahl einzigartiger Quellen zur Geschichte der Weimarer Republik für jedermann in einem neuen Internetportal zugänglich.

Die Quellen umfassen den Zeitraum 1918/19 bis 1929, wobei der inhaltliche Schwerpunkt des Portals zunächst auf den frühen Jahren bis ca. 1924 liegt. Zur Digitalisierung ausgewählt wird forschungs- und themenrelevantes Archivgut (Filme, Fotos, Schriftgut und Töne), das eine ergebnisoffene Auseinandersetzung mit der Epoche ermöglicht.¹¹

Allein aus staatlichen Akten und Unterlagen privater Herkunft werden ab Anfang des Jahres 2017 über einen Zeitraum von vier Jahren voraussichtlich mehr als 4 Millionen Digitalisate entstehen. Die Auswahl der Quellen, die Bereitstellung der Digitalisate und die Präsentation im Internet werden in einer breit aufgestellten Projektgruppe vorbereitet, in der alle beteiligten Abteilungen des Bundesarchivs vertreten sind und spezifische Fachkompetenzen sowie gewachsene Bestandskenntnisse einbringen können. Die Freischaltung des Portals ist für Anfang

⁸ <https://www.leo-bw.de/>.

⁹ Mehr Informationen zum Projekt: <https://www.bundesarchiv.de/oeffentlichkeitsarbeit/meldungen/05232/index.html.de> und <https://blogweimar.hypotheses.org/>; vgl. auch Herrmann/Zahnhausen (2016).

¹⁰ Vgl. die Angebote zum Ersten Weltkrieg (<https://www.ersterweltkrieg.bundesarchiv.de/>) sowie zum 25-jährigen Jubiläum der Wiedervereinigung (<http://wiedervereinigung.bundesarchiv.de/>).

¹¹ Daher ist es ein wichtiges Ziel des Projektes, insbesondere im Bereich Schriftgut - neben der Präsentation von ausgewählten Einzelstücken - die Quellen in ihrem Kontext zugänglich zu machen und folglich immer ganze Aktengruppen oder Teilbestände zu digitalisieren.

2018 geplant. Bis Ende 2020 werden dann sukzessive weitere Digitalisate und Angebote für die Nutzer, wie z.B. virtuelle Galerien, eingepflegt.

Am Beispiel des Themas "Weimarer Republik" soll daher auf Basis dieser Digitalisierungsprojekte eine Präsentationsmöglichkeit für thematische Quellensammlungen im Archivportal-D entwickelt werden. Hierfür gibt es jedoch bislang keine Referenzklassifikation bzw. kein gemeinsames Vokabular zur einheitlichen Verschlagwortung und Verknüpfung der Informationen.¹² Zugleich ist zu konstatieren, dass die benötigten sachthemenatischen Referenzen bzw. klassifikatorischen Erschließungsinformationen in der Regel nicht oder nur mit größerem Implementierungsaufwand in den Datenhaltungen der liefernden Einrichtungen vorgehalten und damit an Portale weitergeben werden können.¹³ Hier setzt das im Folgenden geschilderte Arbeitsprogramm an.

1.1 Projektbezogene Publikationen

1.1.1 Veröffentlichte Arbeiten aus Publikationsorganen mit wissenschaftlicher Qualitätssicherung, Buchveröffentlichungen sowie bereits zur Veröffentlichung angenommene, aber noch nicht veröffentlichte Arbeiten

- Daniel Fähle, Gerald Maier, Tobias Schröter-Karin und Christina Wolf: Archivportal-D. Funktionalität, Entwicklungsperspektiven und Beteiligungsmöglichkeiten, in: Archivar 68 (2015), S. 10-19.
- Michael Hollmann, Deutschland in zwei Nachkriegszeiten. Der Einstieg in das Online-Archiv des Bundesarchivs, in: Archivar. Zeitschrift für Archivwesen 69 (2016), S. 6-9.
- Susanne Laux und Christina Wolf: Forschungsprojekt "Von der Monarchie zur Republik" gestartet: Digitalisierung von Quellen zur Demokratiegeschichte im deutschen Südwesten 1918-1923, in: Archivnachrichten (52) 2016, S. 28-29.
- Harald Sack, Linked Data Technologien - Ein Überblick, in: Linked Enterprise Data - Management und Bewirtschaftung vernetzter Unternehmensdaten mit Semantic Web Technologien. Hg. von T. Pellegrini, H. Sack, and S. Auer. Berlin 2014, S. 21-62.
- Christian Hentschel, Timur Pratama Wiradarma and Harald Sack, If we did not have imagenet: Comparison of fisher encodings and convolutional neural networks on limited training data, in: Advances in Visual Computing, Vol. 9475 of Lecture Notes in Computer Science 2015, S. 400-409.

¹² Je nach Erschließungspraxis in den Kultur- und Wissenseinrichtungen finden sich teilweise bereits Sachschlagwörter in den Datenbeständen von Datenpartnern der DDB. Aufgrund der Heterogenität in diesem Bereich das Vokabular betreffend sowie aufgrund der fehlenden Durchgängigkeit systematischer Schlagwortvergaben erscheint dies jedoch bislang nicht als Grundlage für übergreifende systematische Themenzugänge geeignet. Vgl. hierzu die Ausführungen von Jutta Lindenthal zum Thema "Datenqualität und Retrieval. Vorschläge zur Verbesserung der Suche in der Deutschen Digitalen Bibliothek": http://jl.balilabs.de/DDB/DQ/DDB_Datenqualit%C3%A4t_Retrieval_1.0.pdf.

¹³ Eine flächendeckende sachthemenatische Erschließung wird auch in Zukunft nicht erfolgen. Sie kann allerdings anlassbezogen und in Auswahl vorgenommen werden.

2 Ziele und Arbeitsprogramm

2.1 Voraussichtliche Gesamtdauer des Projekts

Für das Projekt ist eine Gesamtlaufzeit von 36 Monaten vorgesehen. Eine Förderung wird zunächst für die Dauer von 24 Monaten beantragt. Die Förderung für eine anschließende zweite Projektphase wird vor dem Abschluss der ersten Projektphase beantragt. Gewünschter Beginn des Projekts sowie der Förderung ist der 1. Januar 2018.

2.2 Ziele

Gegenstand der ersten Projektphase ist die Erweiterung des zentralen deutschen Archivportals (Archivportal-D) zu einem Online-Informationssystem für die themenbezogene Recherche und Präsentation von archivischen Erschließungsinformationen und digitalem Archivgut sowie der Ausbau des Portals zu einer Plattform für kooperatives Arbeiten an archivischen Daten. Auf Basis der o.g. Digitalisierungsprojekte zur „Weimarer Republik“ soll im Archivportal-D beispielhaft ein sachthematischer Zugang zu digitalisierten Beständen zum Themenfeld „Weimarer Republik“ konzeptioniert und implementiert werden, der später von allen Datenpartnern des Archivportals-D nachhaltig genutzt und für andere Themenkomplexe adaptiert werden kann. Die geplante zweite Projektphase dient dann der weiteren Konsolidierung und Optimierung der entwickelten Komponenten sowie der Ausweitung auf weitere Nutzergruppen und Anwendungsszenarien.

Eine wesentliche Herausforderung des Projekts liegt in den technischen Beschränkungen der datenliefernden Einrichtungen, deren Datenhaltung bzw. Erschließungssysteme nicht oder nur eingeschränkt die notwendige Erweiterung ihres Datenmodells um thematische Bezüge bieten. Neben den technischen Limitierungen stellen in vielen Archiven Erschließungsregularien ein Hindernis dar, die die notwendige sachthematische Indizierung verhindern. Um angesichts der Probleme 1) eines fehlenden einheitlichen Referenzvokabulars, 2) bei der Datenhaltung und Erschließung in den Archivsoftwares sowie 3) bei der Datenbereitstellung und -verarbeitung Rechnung zu tragen, soll daher ein flexibles, nachgelagertes Verfahren zur Metadatenanreicherung bzw. der Vergabe und Pflege der sachthematischen Bezüge zum Einsatz kommen.¹⁴ Hierbei kann auf die bereits im Portal vorliegenden Daten zurückgegriffen werden.

Ein erster Schritt besteht darin, eine sachthematische Klassifikation mit relevanten und fachhistorisch fundierten Kategorien in Abstimmung mit einem Gremium aus Vertretern aus Wissenschaft und Archivwesen für den Themenbereich „Weimarer Republik“ zu definieren.¹⁵ Auf der Ebene der Metadaten ist es das Ziel, eine einheitliche Lösung zu finden, um eine übergreifende, thematische Zuordnung überhaupt erst zu ermöglichen.

Da das Archivportal-D auf dem Backend der Deutschen Digitalen Bibliothek (DDB) aufsetzt, müssen für die Realisierung des Vorhabens Anpassungen an der gemeinsamen Datenhaltung sowie an damit in Verbindung stehenden Workflows und Verarbeitungsprozessen vorgenommen werden. Die im Rahmen des Sonderprojekts „DDB 2017“ überarbeitete Datenhaltung der DDB ermöglicht dabei einen neuen, flexiblen Umgang mit Metadaten. Ziel des Projekts ist daher, ein Verfahren zu entwickeln, mit dessen Hilfe Erschließungsinformationen nachgelagert im Backend der beiden Online-Portale mit Verknüpfungen zu Sachbegriffen und Themenkategorien eines kontrollierten Vokabulars angereichert werden können.

Schwerpunkt des Projekts wird die Entwicklung einer oder mehrerer kooperativer Komponenten zur flexiblen Verwaltung von Themenkontexten und hierarchischen

¹⁴ Das Verfahren wird eine entscheidende Erschließungsunterstützung darstellen. Darüber hinaus wird es allerdings weiterhin essentiell sein, die Standardisierung im Erschließungsprozess und in Hinblick auf das Austauschformat EAD(DDB) voranzutreiben.

¹⁵ Zugleich wird ein entsprechendes Mapping der erstellten Klassifikation auf Linked-Data-Ressourcen (z.B. Wikidata, DBpedia) bzw. weitere geeignete Referenzvokabulare wie die GND geprüft und vorgenommen.

Sachklassifikationen sein, die zunächst die Projektbeteiligten, später auch weitere Archivmitarbeiter/innen sowie wissenschaftliche Nutzer und Nutzerinnen dazu befähigt bzw. befähigen, Objekte (Verzeichnungseinheiten) oder ganze Bestände bestimmten Kategorien selbstständig zuzuordnen. Durch dieses Vorgehen ist eine nachhaltige Befüllung und Pflege der Themenkategorien, auch über das Projektende hinaus, möglich. Eine vorgesehene Möglichkeit zur Rückspielung der angereicherten Daten zu den Datenlieferanten erhöht die grundsätzliche Datenqualität in deutschen Archiveinrichtungen und macht die erweiterten Erschließungsinformationen auch in den Ursprungssystemen für weitere Linked-Data-Anwendungen nutzbar.

In einem weiteren Arbeitspaket soll ein Verfahren zur algorithmischen Zuordnung von Archivdaten zur definierten sachthematischen Klassifikation auf der Basis semantischer Technologien erforscht, prototypisch umgesetzt und getestet werden. Eine besondere Herausforderung besteht in diesem Bereich in der großen inhaltlichen Heterogenität der Ausgangsdaten, die eine präzise Zuordnung der Daten erschweren.¹⁶

Mit dem Projektvorhaben soll den Erwartungen wissenschaftlicher Nutzer und Nutzerinnen entsprochen werden, neue Erkenntnisse durch die einrichtungsübergreifende thematische Zusammenführung von Archivalien in Online-Portalen gewinnen zu können.¹⁷ Die Präsentation von Objekten in einem weiteren Kontext eröffnet neue Perspektiven zu bereits bestehenden und sich neu ergebenden Forschungsfragen. Insbesondere im Archivportal-D schafft ein thematischer Zugang einen zusätzlichen Einstieg, mit dem sich das wissenschaftliche Arbeiten effizienter gestalten lässt und der einen schnellen, bundesweiten Überblick über vorhandenes Material bietet.

2.3 Arbeitsprogramm und Umsetzung

AP 1: Entwicklung eines Metadaten-Modells zur sachthematischen Referenzierung und Anpassung der Datenhaltung

Für die Etablierung sachthematischer Zugänge im Archivportal-D ist eine Erweiterung bzw. Modifikation des bisherigen Metadatenkonzepts notwendig. Dieses basierte auf dem Grundsatz, dass die von den liefernden Einrichtungen bereitgestellten Erschließungsinformationen zwar formal für die Portalpräsentation aufbereitet werden, aber inhaltlich unverändert bleiben. Zusätzliche Metadaten, wie beispielsweise Referenzen auf eine sachthematische Klassifikation, hätten demnach immer bereits in einer standardisierten Form in den Lieferdaten vorhanden sein müssen. Diese Voraussetzung ist aufgrund der oben beschriebenen Barrieren Limitationen nicht gegeben. Stattdessen Deshalb soll in technologisch-organisatorischer Hinsicht ein neuer Weg beschritten werden, der die Vorteile einer Unabhängigkeit von Liefersystemen und Archivsoftwaresysteme mit einer hohen Flexibilität bei Aufbau und Verwaltung von Themenzugängen im Archivportal-D (Skalierbarkeit) und später auch in der DDB kombiniert. Dazu soll auf der Grundlage der bereits vorliegenden Erschließungsinformationen eine nachgelagerte Anreicherung mit den nötigen sachthematischen Metadaten erfolgen. Die neue, überarbeitete Architektur der DDB, auf deren Backend bzw. Datenhaltung das Archivportal aufsetzt, unterstützt

¹⁶ Weiterführend dazu auch: Sack (2014), S. 21-62 und Hentschel/Wiradarma/Sack (2015), S. 400-409.

¹⁷ In seinem Empfehlungsschreiben beschreibt Patrick Sahle u.a. die unterschiedlichen Nutzerbedürfnisse: „Das Portal bietet einen guten Einstieg auf der Ebene einer allgemeinen Suche und durch die tektonische Struktur der Bestände in den Archiven. Beides sind essentielle Zugänge, sie holen aber nicht alle potentiellen Benutzer dort ab, wo ihre Informationskompetenz und bereits vorhandene Orientierung enden. Denn für eine freie Suche muss man bereits wissen, nach welchen Begriffen sinnvoll gesucht werden kann. Und für den Weg über Archivtektoniken muss man bereits ein gewisses Verständnis für Archivstrukturen haben. Dieses Problem kann vor allem durch sachthematische Zugänge gelöst werden, durch die zugleich auch das große historische Fach- und Kontextwissen in den Archiven nutzbar gemacht werden kann. Hier exemplarisch mit der Weimarer Republik anzufangen scheint mir äußerst naheliegend, weil hier mit einem sehr breiten Interesse bei den Nutzern zu rechnen ist.“

bereits grundsätzlich derartige Anwendungsfälle. Datenmodell und Datenhaltung sowie Verarbeitungsverfahren und Workflows müssen hierfür jedoch angepasst werden.

Aus dem hierarchischen Charakter der angestrebten sachthematischen Referenzierung ergibt sich eine neue Anforderung an das Backend, da die im AP 2 fachlich zu definierende Klassifikation als mehrstufige Facette umgesetzt werden muss. In einem ersten Schritt erfolgt daher die Erstellung eines Konzepts für hierarchisch strukturierte Facetten für die sachthematischen Navigationszugänge und eine dahingehende Erweiterung des globalen Datenmodells.

Die durch nachgelagerte Anreicherung gewonnenen Metadaten sollen in einer neu zu schaffenden Datenschicht vorgehalten werden. Diese separate Datenhaltung erfolgt aus rechtlich-organisatorischen und technologischen Erfordernissen, denn die erweiterten Metadaten sind von den eigentlichen Lieferdaten (Kernmetadaten) zu trennen und müssen zugleich effizient und flexibel zu verarbeiten sein. Bei der Etablierung der separaten Datenschicht muss sichergestellt werden, dass die erweiterten sachthematischen Metadaten mit den jeweiligen originären Datenobjekten im DDB-Repositorium persistent verknüpft sind, damit die Zuordnung bei Datenaktualisierungen erhalten bleibt.

Die Modifikationen des Datenmodells und der Datenhaltung sollen auch für die Auslieferung der Daten über das API¹⁸ der DDB nutzbar gemacht werden. So ist eine Erweiterung der vorhandenen API-Methoden vorgesehen, um die thematisch angereicherten Daten nachnutzen zu können, indem z.B. themenbezogene Daten-Subsets für Anwendungsszenarien der Digital Humanities abgerufen werden können. Darüber hinaus muss eine Möglichkeit zur Re-Integration angereicherter Daten in Liefersystemen implementiert werden. Hierzu ist geplant, Einrichtungen bei Bedarf die sachthematischen Bezüge zu ihren Daten in Form von Konkordanzlisten zur Verfügung zu stellen.¹⁹

Arbeitsschritte:

- Konzeption der hierarchisch strukturierten Facetten und deren Modellierung
- Erweiterung des Datenmodells von DDB und Archivportal-D für die Abbildung sachthematischer hierarchischer Facetten
- Konzeption und Umsetzung einer neuen datenhaltenden Schicht für angereicherte Metadaten bzw. User Generated Content
- Anpassung der Datenverarbeitungsprozesse und Workflows an die erweiterte Architektur
- Erweiterung des API-Umfangs der DDB zur Ausgabe der angereicherten thematischen Metadaten bzw. von thematischen Daten-Subsets
- Konzeption und Realisierung eines Daten-Rückkanals zur Re-Integration angereicherter Metadaten in Ursprungssysteme

AP 2: Erstellung einer Referenz-Sachklassifikation für den Themenbereich „Weimarer Republik“ und intellektuelle Zuordnung von Beständen

Als Basis der im Archivportal-D beispielhaft umzusetzenden thematischen Quellensammlung „Weimarer Republik“ wird eine Sachklassifikation in Form einer fachhistorisch fundierten Systematik benötigt. Die Themenstruktur muss als mehrstufige Klassifikation aufgebaut sein, da sie, von übergeordneten Klassen ausgehend, hierarchisch abgestuft sachthematische Teilaspekte beinhaltet. Als eine Grundlage dient die bestehende dreistufige Klassifikation aus dem

¹⁸ Application Programming Interface, s. <https://api.deutsche-digitale-bibliothek.de/>.

¹⁹ Eine Beschreibung des auf Linked-Data-Prinzipien basierenden Datenmodells der DDB findet sich u.a. in Rühle/Schulze/Büchner (2014).

Digitalisierungsprojekt des LABW.²⁰ Diese wird mit der geplanten thematischen Gliederung des Bundesarchiv-Projekts abgeglichen und auf Übereinstimmungen geprüft. Hierbei sollen zentrale Begrifflichkeiten sowie Strukturen im Sinne eines Anforderungskatalogs für eine übergreifende und anschlussfähige Klassifikation zur "Weimarer Republik" herausgearbeitet werden.

Mit Blick auf die Erstellung einer übergreifenden Systematik sollen bestehende einschlägige sachthematische Vokabulare und Klassifikationen (z.B. Systematik des Instituts für Zeitgeschichte, Teilhierarchien der Wikipedia-Sachsystematik) auf ihre Eignung für das Projekt hin geprüft werden. Hierzu ist u.a. die Einbeziehung der beim digiCULT-Verbund eG vorliegenden Expertise vorgesehen. Bei der Konzeption der thematischen Klassifikation wird aber auch von Beginn an eine enge Zusammenarbeit mit Expertinnen und Experten aus Wissenschaft und Archiven angestrebt. Vorgesehen sind hierzu zwei Workshops, um das zu verwendende Vokabular hinsichtlich sowohl der fachhistorischen Nutzenden als auch der bereitstellenden Archive tragfähig zu gestalten. Um die Skalierbarkeit der zu wählenden Referenz-Sachklassifikation sicherzustellen - also die Anwendbarkeit auf weitere Einrichtungen und Inhalte anderer Sparten -, ferner die Erweiterbarkeit auf zusätzliche Themen sowie auf die Umsetzung in der DDB-Oberfläche, erfolgt eine enge Abstimmung mit der DNB als technischem Koordinator der DDB.²¹

Nach der erfolgten Festlegung der Referenz-Sachklassifikation soll diese - wo es möglich ist - über Cross-Konkordanzen mit GND-Sachbegriffen verknüpft werden, um im Rückgriff auf die Normdatei eine portalübergreifende semantische Vernetzung der klassifizierten Erschließungsinformationen zu ermöglichen.

Schließlich erfolgt die intellektuelle Zuordnung der Bestände und der einzelnen Objekte des Bundesarchivs und des Landesarchivs Baden-Württemberg zu den einzelnen definierten Klassen. Hinsichtlich der LABW-Daten wird dies über die Definition von Konkordanzlisten im konzeptionellen Mapping realisiert, die auf eine Zuweisung der proprietären Projekt-Klassifikation zu der neu erstellten Referenz-Klassifikation abzielen. Um die intellektuelle Zuordnung perspektivisch effizienter zu gestalten bzw. idealerweise durch automatisierte Verfahren zu unterstützen, wird in AP 5 eine entsprechende Technologie erprobt.

Arbeitsschritte:

- Erstellung eines Fachkonzepts mit zentralen Begrifflichkeiten für die Themenstruktur der Weimarer Republik
- Evaluierung einschlägiger sachthematischer Vokabulare bzw. Klassifikationen
- Planung und Durchführung zweier Workshops zur Abstimmung der Systematik mit Vertretern aus Fachwissenschaft und Archivwesen
- Festlegung der übergreifenden Referenz-Sachklassifikation
- Ermittlung von Cross-Konkordanzen zu GND-Sachbegriffen
- konzeptionelles Mapping der LABW-Systematik auf die Referenz-Sachklassifikation
- Intellektuelle Zuordnung der Datenbestände des BArch zur Referenz-Systematik

²⁰ Diese Klassifikation ist auf der ersten Ebene territorial (Baden, Hohenzollern und Württemberg) untergliedert. Die Ebenen 2 (Politisches Leben, Wirtschaft und Soziales, Gesellschaft, Kultur und Bildung) und 3 (bspw. Friedensschluss, Revolution, Rätebewegung etc.) sind für alle Regionen identisch und bilden thematische Haupt- und Untergliederungspunkte ab. Die Klassifikation verwendet bereits Sachschlagworte, die in der GND nachgewiesen sind.

²¹ Die Klassifikation soll flexibel gestaltet werden, so dass ein zeitlich größerer Rahmen abgedeckt und eine prinzipielle inhaltliche Erweiterung vorgenommen werden kann, um z.B. die Bestände weiterer Archive ergänzen zu können.

AP 3: Konzeption und Umsetzung von Klassifikationswerkzeugen zur Pflege von hierarchischen Sachklassifikationen sowie zur Zuordnung von Daten(beständen) zu den einzelnen Klassen

Um die in AP 2 erstellte mehrstufige Referenz-Klassifikation hinterlegen und pflegen zu können, wird eine entsprechende Werkzeug-Anwendung mit Nutzeroberfläche benötigt. Diese Anwendung soll nicht nur für die Weimar-Thematik einsetzbar sein, sondern künftig für alle weiteren Themenzugänge und Klassifikationen im Archivportal-D sowie der DDB genutzt werden können. Mit der Möglichkeit zur Pflege der Klassifikationen selbst soll eine Flexibilität gegeben sein, diese später anzupassen oder auszubauen. Eine derartige Anwendung zur flexiblen Vokabularverwaltung wird möglicherweise neu zu konzipieren und umzusetzen sein, ggf. kann für die genannten Zwecke aber auch xTree eingesetzt oder nachgenutzt werden.

Ein zweites zu realisierendes Werkzeug dient der kollaborativen Zuordnung von Datenbeständen oder einzelnen Dokumenten bzw. Verzeichnungseinheiten zu den jeweiligen Klassen sowie zur Ergänzung einer Sortierung für die Anzeigereihenfolge. Dafür greift das Tool auf die in der Vokabularverwaltung hinterlegte Klassifikation zurück und speichert die intellektuell zugewiesenen Referenzen als erweiterte Metadaten in der unter AP 1 skizzierten neuen Datenschicht ab. Die Bedienung der Werkzeug-Oberfläche soll möglichst intuitiv erfolgen können, so dass diese Komponente z.B. auch für nicht IT-affines Fachpersonal in den datenliefernden Archiven sowie ggf. zukünftig für alle Nutzende zur Verwendung zur Verfügung gestellt werden kann.²² Datenpartner können so, auch nach Projektende, ihre eigenen Bestände passenden Themenkategorien zuweisen; Fachpersonal oder Forschende könnten kuratierte Themensammlungen erstellen. Angestrebt wird ein gestaffeltes Vorgehen, wonach ein Prototyp des Tools zunächst für den projektinternen Gebrauch erstellt wird, ein Ausbau des Werkzeugs für weitere Nutzergruppen (inkl. Designelementen und intuitiver Usability) hingegen erst in der zweiten Projektphase erstellt werden soll.

Parallel zur Etablierung der genannten Werkzeuge muss eine Lösung für die Fälle gefunden werden, in denen Lieferdaten bereits thematische Zuordnungen enthalten. Hierzu ist vorgesehen, die proprietären Zuordnungen ggf. in das Verknüpfungs-Tool zu laden und dort über ein hinterlegtes Mapping eine Abbildung auf die in den Portalen verwendete Klassifikation zu ermöglichen.

Arbeitsschritte:

- Anforderungsdefinition für eine Vokabularverwaltung zur Pflege flexibler Klassifikationen
- Umsetzung der Anforderungen in einem Vokabularmanagement-Werkzeug oder Nachnutzung von xTree
- funktionale und technische Konzeption einer Anwendung zur Zuordnung von Archivbeständen zur Referenz-Klassifikation
- Erarbeitung eines Nutzerrechte- und Rollenmodells zum Einsatz des genannten Software-Werkzeugs als kollaborative Komponente
- Entwicklung und Evaluierung eines Prototyps
- Toolbasierte Umsetzung der intellektuellen Zuordnungen und Sortierung der einzelnen Datensätze

AP 4: Konzeption und Realisierung des sachthematischen Recherchezugangs im Frontend des Archivportals-D

²² Die Funktionalität könnte ggf. analog von Datei- oder Ordnersystemen aufgebaut werden, die eine (Mehrfach-)Zuordnung bspw. durch Drag & Drop von Beständen zu Klassifikationspunkten ermöglicht.

Die bestehenden Zugänge zum digitalen Archivgut werden in der Nutzeroberfläche (Frontend) des Archivportals-D um sachthematische Komponenten erweitert. Dies umfasst zum einen die Etablierung eines entsprechenden Rechercheeinstiegs, zum anderen aber auch die thematische Kontextualisierung bei der Präsentation des Archivguts auf Treffer- und Merklisten sowie Detailseiten. Dabei wird im Projekt eine fachgerechte Umsetzung zu entwickeln sein, die sich in die bestehende nutzerfreundliche Gesamtkonzeption von Recherche und Präsentation im Archivportal-D einfügt. Zugleich werden für die Erweiterung Anpassungen an den bestehenden Funktionalitäten und am Layout des Archivportals-D notwendig.

Der neue sachthematische Zugang wird an zentraler Position auf der Startseite verankert. Analog zum bestehenden Archiv-Strukturbaum auf den Trefferlisten-Seiten soll ein Themen- bzw. Klassifikationsbaum realisiert werden. Dieser soll die Navigation über die einzelnen Klassen innerhalb des Themenbereichs ermöglichen (Browsing), aber auch im Sinne einer Facette fungieren – wenn der Rechercheeinstieg beispielsweise durch eine Freitextsuche erfolgt ist. Ergebnislisten einer textlichen Suche oder einer Recherche über Archive und Tektoniken können damit nachträglich auf thematische Zugehörigkeiten bzw. Referenzen eingegrenzt werden. Zusätzlich soll hierbei das über die Verknüpfung von Klassen mit GND-Sachbegriffen gewonnene kontrollierte Vokabular zur Facettierung herangezogen werden.

Vorgesehen ist ferner ein Ausbau der Merklistenfunktionalität, um die thematischen Zuordnungen auch hier abzubilden und neue Möglichkeiten zur Verwaltung privater thematischer Zusammenstellungen anzubieten. Neben der Exportfunktion im PDF-Format sollen zusätzliche Ausgabe- und Weitergabemöglichkeiten (“Sharing”) für individuelle Themensammlungen konzipiert werden. Schließlich wird auf den Objekt-Detailseiten in geeigneter Form der sachthematische Kontext visualisiert, der zugleich wiederum als explorativer Zugang in das übergeordnete Themengebiet dienen soll.

Arbeitsschritte:

- Konzeption der sachthematischen Recherche und Präsentation auf der Basis der entwickelten Klassifikation (vgl. AP 2) für die Portaloberfläche
- Überarbeitung bzw. Anpassung des Interaktionsdesigns (Layout) und Integration der sachthematischen Komponenten in das funktionale Gesamtkonzept des Archivportals-D
- Entwicklung eines funktionalen Prototyps zur projektinternen Evaluierung hinsichtlich der Konvergenz zu fachlich-wissenschaftlichen Anforderungen und Optimierung der allgemeinen Nutzerfreundlichkeit (Usability)
- Implementierung der sachthematischen Komponenten in die Nutzeroberfläche des Archivportal-D-Produktsystems unter Berücksichtigung des Anpassungsbedarfs aus dem evaluierten Prototypen

AP 5: Definition, Test und Implementierung eines Algorithmus zur Generierung automatisierter Zuordnungsempfehlungen

Ergänzend zu den in AP 3 erläuterten Klassifikations-Werkzeugen, die die Umsetzung von intellektuell definierten Zuweisungen von Beständen oder einzelnen Erschließungsinformationen zum Gegenstand haben, sollen in einem Teilprojekt auch die Möglichkeiten einer (teil-)automatisierten Zuordnung zur Referenz-Sachklassifikation konzipiert, implementiert und evaluiert werden. Die Umsetzung soll unter Nutzung semantischer Technologien in Verbindung mit automatisierter Klassifikation auf Basis maschineller Lernverfahren erfolgen. Ziel des Arbeitspakets ist es, den Umfang der sachthematisch zuordenbaren Archivbestände signifikant zu erhöhen, indem die intellektuelle Verknüpfung durch algorithmisch ermittelte Vorschläge (*recommendations*) unterstützt wird. Zusätzlich ist geplant, weitere Bestände, die zunächst nicht für eine thematische Zuordnung zum Komplex “Weimarer Republik” vorgesehen waren, auf ihre Relevanz hin zu analysieren sowie ggf. einzelne sachthematisch einschlägige Gliederungspunkte

oder Objekte bzw. Archiveinheiten zu identifizieren und mit dem entsprechenden Klassifikationspunkt der Quellensammlung "Weimarer Republik" zu verknüpfen.

Anhand hierzu festzulegender repräsentativer Testdatensätze soll zunächst ein automatisierter Metadatenabgleich gegen das Vokabular der Referenzklassifikation erfolgen. Zum vorgegebenen Referenzvokabular werden mit Hilfe externer Datenquellen, wie z.B. Wikidata und DBpedia, synonyme Begriffe und Begriffsvarianten ermittelt, die unter Einbeziehung bereits klassifizierter und annotierter Archivalien fortlaufend ergänzt und als Grundlage für eine automatisierte Klassifikation als Trainingsdaten herangezogen werden.

Zudem sollen Parameter wie Keyword-Häufigkeiten und Freitexte in definierten Erschließungsfeldern als zusätzliche Informationsquelle für eine automatische Klassifikation herangezogen werden. Ferner soll betrachtet werden, ob bei fehlenden Keywords zusätzliche Kontextinformationen genutzt werden können, um Zuordnungsvorschläge zu ermitteln. Es wird auch zu prüfen sein, inwieweit externe Datenquellen wie Wikidata im Falle unzureichender Kontextualisierung in den Erschließungsinformationen Hinweise auf relevante Keywords für eine Zuordnung erbringen können.

Der entwickelte Algorithmus dient als Basis eines konfigurierbaren Anwendungsprototyps, der zur Analyse definierter Bestände herangezogen werden kann und Zuordnungsvorschläge zur sachthemen Klassifikation generiert, die wiederum Eingang in das unter AP 3) erläuterte Zuordnungswerkzeug finden. Die Vorschläge müssen anschließend intellektuell geprüft, also bestätigt oder verworfen werden. Die im Zuge der semantischen Analyse erzeugten Metadaten werden als Linked Data aufbereitet und zur Verfügung gestellt.

Arbeitsschritte:

- Auswahl geeigneter bereits klassifizierter Archivbestände als Trainingsdaten.
- Semantische Analyse der Archivinhalte, Verknüpfung mit externen Datenquellen und Zurverfügungstellung der generierten Metadaten als Linked Data
- Ermittlung geeigneter Kenngrößen (Feature) für die automatische Klassifikation, Bestimmung eines dazu geeigneten maschinellen Lernverfahrens und anschließendes Training des Klassifikators
- Entwicklung eines konfigurierbaren Prototyps
- Etablierung einer SPARQL-Schnittstelle zum semantischen Abgleich mit externen Ressourcen
- Bewertung der Ergebnisse und Dokumentation der Vorgehensweisen

AP 6: Projektkoordination, Öffentlichkeitsarbeit, Evaluation und Kooperation mit der DDB-Servicestelle

Aus der engen Verknüpfung bzw. technischen Verzahnung von Archivportal-D und DDB ergibt sich ein hoher Abstimmungs- und Koordinationsbedarf zwischen LABW, der DDB-Servicestelle bzw. dem Technischen Koordinator der DDB an der DNB, dem FIZ Karlsruhe sowie dem Bundesarchiv. Über regelmäßige Gespräche und Treffen zwischen Vertretern der genannten Institutionen wird eine enge Abstimmung sichergestellt. Eine nachhaltige Verankerung der neuen Komponenten im bestehenden Gefüge der DDB-Plattform ist damit gewährleistet. Aufgrund der stark kooperativen Ausrichtung des Vorhabens besteht darüber hinaus grundsätzlich zwischen allen Partnern ein kontinuierlich hoher Abstimmungsbedarf.

Die beim LABW angesiedelte Projektkoordination übernimmt sämtliche Aufgaben des übergreifenden Projektmanagements und organisiert ferner die projektinterne Evaluation und die Öffentlichkeitsarbeit. Für die archivischen und wissenschaftlichen Communities sind Vorträge über die neuen sachthemen Komponenten des Archivportals-D sowie Beiträge in einschlägigen Fachzeitschriften geplant. Zudem sollen die neuen Funktionalitäten des Archivportals-D auf

wissenschaftlichen Tagungen aktiv beworben werden. Die Projektergebnisse werden entsprechend dokumentiert und frei zugänglich gemacht.

Arbeitsschritte:

- Projektmanagement (Zeit- und Ressourcenplanung, Berichtswesen, Koordination von Projektteam und -aufgaben)
- Koordination mit der DDB-Servicestelle/technischer DDB-Koordination
- Präsentation auf Archivtagen und weiteren Veranstaltungen der Archiv- und Wissenschafts-Community
- Öffentlichkeitsarbeit für verschiedene Nutzerzielgruppen: wissenschaftlich Forschende und Lehrende, Studierende sowie interessierte Laien
- Vorstellung der im Projekt entwickelten Klassifikation in geeigneten Foren der Fachöffentlichkeit
- Veröffentlichungen in fachwissenschaftlichen Medien
- Dokumentation der Ergebnisse
- Projektinterne Evaluation (vgl. 2.4)

Ausblick auf geplante Schwerpunkte der zweiten Projektphase (Monate 25-36)

- Konsolidierung der Projektergebnisse
- Schulungen zum Einsatz der im Projekt erarbeiteten Sachklassifikation und zur Anreicherung archivischer Bestände mit dem entwickelten Klassifikationstool
- instruktive Aufbereitung der Projektergebnisse in Leitfäden und E-Learning-Videos
- Verknüpfung zusätzlicher Archivbestände
- Implementierung des Vorschlags-Algorithmus im Klassifikationstool
- Erweiterung des Klassifikationstools zu einer Web-2.0-Komponente für alle Nutzer/innen
- Ausbau des Klassifikationstools zur Anwendung für weitere Nutzungsszenarien (Archivalientypen, Zielgruppenorientierte Zugänge für Genealogen etc.)
- ggf. Erweiterung der Referenz-Sachklassifikation auf das Themengebiet Nationalsozialismus
- Ausweitung des Vorschlags-Algorithmus auf *Visual Object Detection* zur Ermittlung klassifikatorischer Zuordnungen über Bildinhalte

2.4 Maßnahmen zur Erfüllung der Förderbedingungen und Umgang mit den Projektergebnissen

Nachhaltigkeit und Rückkoppelung in die Fachcommunitys

Die Projektpartner verpflichten sich zur dauerhaften Bereitstellung der im Zuge der Umsetzung des Vorhabens generierten Ergebnisse. Diese werden nachhaltig über die Bereitstellung des Themenzugangs „Weimarer Republik“ verfügbar sein, der zudem weiter gepflegt und um neue Bestände ergänzt werden kann. Über die enge Verzahnung der technischen und organisatorischen Strukturen von Archivportal-D und DDB, die in einem Betriebskonzept zwischen DDB und LABW mit Blick auf den Dauerbetrieb und Weiterentwicklungsaspekte abgestimmt und festgeschrieben wurden, ist die Nachhaltigkeit des Angebots zudem abgesichert. Die entwickelten Komponenten (neue Datenschicht, Klassifikationswerkzeuge) sind außerdem für viele weitere Anwendungsszenarien nachnutzbar. Die erstellte Referenz-Klassifikation kann ebenfalls auf weitere Themenbereiche ausgeweitet werden und ist in sich flexibel, so dass künftig neue Anforderungen an das Vokabular abgebildet werden können.

Die Verbreitung der Projektergebnisse ist über deren Dokumentation und anschließende Vermittlung über eine entsprechende Öffentlichkeitsarbeit gewährleistet. Von Beginn an wird im Projekt ein Austausch mit den Fachcommunitys hergestellt, Zwischenergebnisse sollen auf ihre Praxistauglichkeit für die historische Forschung hin von potenziellen Anwendern/innen evaluiert werden. Somit wird sichergestellt, dass sich die entwickelten Komponenten tatsächlich am Bedarf

orientieren. Insbesondere erfolgt eine Einbeziehung von Fachhistorikern/innen, Informationswissenschaftlern/innen und Vokabularexperten/innen bei der Erstellung der Klassifikation (Workshops). Archive, als weitere Ziel- /Nutzergruppe, werden ebenfalls einbezogen, indem etwa Digitalisierungsprojekte zum Thema "Weimarer Republik" aufgegriffen werden. Archivische Gremien (z.B. KLA) werden frühzeitig über das Vorhaben informiert und einbezogen. Interessierten Archiveinrichtungen werden die Werkzeuge und Möglichkeiten zur sachthemenatischen Präsentation im Archivportal-D schnellstmöglich zur Verfügung gestellt. Über die Beteiligung der DDB ist perspektivisch auch die spartenübergreifende Nutzung der thematischen Komponenten gegeben.

Wissenschaftlicher Bedarf und Anwendungsbeispiele

Seitens der historischen Fachöffentlichkeit existiert seit Längerem die Forderung an Archive, alternative Zugänge zum Archivgut neben dem einschlägigen Provenienzprinzip und über die Volltext-Recherchemöglichkeiten in Informationssystemen hinaus anzubieten. Die beigefügten Empfehlungsschreiben belegen dabei exemplarisch den Bedarf der Fachwissenschaft an qualifizierten übergreifenden Quellensammlungen im Allgemeinen, im Besonderen gerade auch bezogen auf die fokussierte Weimarer Republik als demokratiegeschichtlich zentraler Periode der deutschen Geschichte. Der renommierte Neuzeithistoriker Wolfram Pyta sieht daher hier "sowohl für die Forschung als auch für die Lehre ein wichtiges Desiderat".

Über die vom Bundesarchiv und vom LABW initial sachthematisch zugänglich gemachten Bestände hinaus werden im Archivportal-D künftig Quellen auf einen Blick erscheinen, die sowohl die reichsweite wie auch regionalgeschichtliche Dimension zu längerfristigen Entwicklungslinien (etwa den Aufbau und die Zerschlagung des demokratischen Systems, unterschiedlichste soziale und ökonomische Prozesse etc.) sichtbar machen. Auch enger gefassten Einzelthemen wie beispielsweise dem Friedensschluss, zu Reparation und Inflation, dem Kapp-Putsch oder auch den Freiwilligenverbänden des "Grenzschutz Ost" werden dargestellt.

Technologien und Interoperabilität

Für das Projekt werden bestehende Komponenten (Frontend des Archivportals-D, Backend der DDB, Ladesoftware) nachgenutzt. Der überwiegende Teil der zu erstellenden Software nutzt Werkzeuge und Komponenten, die sich bereits für das Archivportal-D im Einsatz befinden (Cassandra, Apache Spark, Solr, Groovy/Grails). Die Veröffentlichung von Software als Open Source erfolgt unter GitHub. Die Entwicklungen finden als Teil der DDB statt, die als nationale Infrastruktur und als Aggregator für die Europeana auch die internationale Anschlussfähigkeit sicherstellt. Bei den Erschließungsinformationen wird auf verbreitete XML-Profile (EAD(DDB), METS/MODS) zurückgegriffen und damit die Interoperabilität sichergestellt. Die erarbeitete Referenz-Klassifikation wird zur freien Nachnutzung öffentlich zur Verfügung gestellt und über Cross-Konkordanzen zu bestehenden Vokabularen (insbesondere GND) anschlussfähig gemacht.

Projektinterne Evaluation

Die Realisierung sämtlicher Software-Komponenten erfolgt in einem an das agile Software-Entwicklungsverfahren SCRUM angelehnten Prozess. Dieser ist geprägt von iterativen Entwicklungsschritten, die eine kontinuierliche Zwischenevaluierung ermöglichen. Substanzielle Entwicklungsergebnisse, insbesondere aber die jeweils bereitgestellten Prototypen werden intensiv sowohl seitens des Projektteams wie auch durch externe wissenschaftliche

Ansprechpartner (z.B. die Autoren der Empfehlungsschreiben) begutachtet.²³ Hinsichtlich der zu erstellenden Referenz-Klassifikation (AP 2) ist vorgesehen, den zweiten der genannten Workshops dezidiert der Diskussion bzw. Begutachtung des Vokabular-Entwurfs durch Expertinnen und Experten aus Wissenschaft und Archiven zu widmen. Die sich hieraus ergebenden anwendungsbezogenen Änderungswünsche fließen in die endgültige Modellierung der Klassifikation ein. Schließlich stellt die Qualitätssicherung einen wesentlichen Aspekt des allgemeinen Projektmanagements dar.

Anhand der beiden Projekte zur Weimarer Republik (LABW/BArch) erfolgt zudem eine Erprobung der Konzepte und der resultierenden Softwares an realen Beispielen, die auch eine Evaluierung der semiautomatischen Empfehlung von Zuordnungen (AP 5) ermöglichen, da hier die Ergebnisse der algorithmischen mit denen der intellektuellen Zuordnung verglichen werden können.

3 Literaturverzeichnis

- Irmgard Christa Becker, Gerald Maier, Karsten Uhde und Christina Wolf (Hrsg.): 19. Archivwissenschaftliches Kolloquium: Netz werken – Das Archivportal-D und andere Portale als Chance für Archive und Nutzung, Marburg 2016.
- Tobias Herrmann/Vera Zahnhausen, Auf dem Weg zum Digitalen Lesesaal: Das Projekt "Weimar – Die erste deutsche Demokratie", demnächst in: Tagungsdokumentation zum Deutschen Archivtag Bd. 21 (86. Deutscher Archivtag 2016 in Koblenz). Powerpoint-Präsentation abrufbar unter: <https://blogweimar.hypotheses.org/>.
- Martin Reisacher und Wolfgang Krauth: Vernetzen als Herausforderung - die Deutsche Digitale Bibliothek, in: Archivnachrichten (51) 2015, S. 36 f.
- Stefanie Rühle, Francesca Schulze und Michael Büchner: Applying a Linked Data Compliant Model: The Usage of the Europeana Data Model by the Deutsche Digitale Bibliothek (2014), abrufbar unter: <http://dcevents.dublincore.org/IntConf/dc-2014/paper/view/231/225>.
- Christina Wolf: Eines für alle: das Archivportal-D. Neue Zugangswege zu Archivgut, in: Neue Wege ins Archiv - Nutzer, Nutzen, Nutzung. 84. Deutscher Archivtag 2014 in Magdeburg (Tagungsdokumentationen zum Deutschen Archivtag, Band 19), S. 47-63, Fulda 2016.

²³ Die Softwares durchlaufen einen internen Code-Review durch jeweils zwei weitere Entwickler. Konzeption und Architektur der Softwares durchlaufen jeweils Reviews durch den technischen Koordinator. Die Konzeption der fachlichen Anforderungen erfolgt durch das Projektmanagement und wissenschaftliche Ansprechpartner/innen.