

E-Papers der Archivschule Marburg

Hochschule für Archivwissenschaft

Nr. 24

Manuel Limbach

Mehrwert oder Mehrarbeit?

Voraussetzungen und Möglichkeiten des Einsatzes von Normdaten in der Sacherschließung des Bundesarchivs am Beispiel des „Themenportals Wiedergutmachung“

Transferarbeit des 55. wissenschaftlichen Lehrgangs
an der Archivschule Marburg

Betreuer der Archivschule Marburg:
Betreuer des Ausbildungsarchivs:

Dr. Robert Meier
Dr. Karsten Kühnel

Marburg/Lahn 2023

HESSEN



Die Archivschule Marburg ist eine Einrichtung des Landes Hessen.

Mehrwert oder Mehrarbeit?

Voraussetzungen und Möglichkeiten des Einsatzes von Normdaten in der Sacherschließung des Bundesarchivs am Beispiel des „Themenportals Wiedergutmachung“

von Manuel Limbach

1. Einleitung

Die Erschließung von Archivgut dient seiner Auffindbarkeit und Zugänglichmachung. Enorme Erschließungsrückstände zwingen die Archive jedoch, das Verhältnis von Erschließungsaufwand und dem Mehrwert einer Tiefenerschließung für die Nutzerinnen und Nutzer stets neu auszuloten. Im Bundesarchiv begegnet man dieser Problematik mit einer „Ressourcenkonzentration auf ausgewählte Bestände“.¹ Steht hierbei vor allem das effiziente Arbeiten im Vordergrund, wirft der digitale Wandel zusätzlich die Frage nach der Effektivität der gewählten Erschließungsansätze auf. Das auf die Maschinenlesbarkeit von Daten basierende Web 3.0 (Semantic Web) sowie veränderte Nutzererwartungen und -gewohnheiten stellen ein Archiv, das den Anspruch erhebt, zeitgemäß erschließen zu wollen, vor ganz neue Herausforderungen.² Diese liegen allen voran in der Schaffung von nutzerfreundlichen, webbasierten Zugängen sowie digitalen Auswertungsmöglichkeiten durch eine semantische Vernetzung der Erschließungsinformationen über Normdaten.

Das Bundesarchiv, in dem die Anreicherung von Erschließungsinhalten mit Normdaten unter anderem aus der Gemeinsamen Normdatei (GND) seit einiger Zeit intensiv diskutiert wird, hat den Weg hin zu einer deutlich ausgeprägteren Sacherschließung erst kürzlich mit seiner Beteiligung an der Entwicklung des „Themenportals Weimarer Republik“ eingeschlagen und möchte ihn nun durch die Realisierung eines Themenportals zur „Wiedergutmachung nationalsozialistischen Unrechts“ weitergehen. Durch den Aufbau von sachthematischen Systematiken im Archivportal-D sowie den Einsatz von normdatenreferenzierten Schlagwörtern (Deskriptoren) sollen einerseits Forschungsdesiderate und Nutzererwartungen erfüllt und andererseits (semi-)automatisierte Verfahren der Normdatenanreicherung für einen flächendeckenden Einsatz im Rahmen der Sacherschließung im Bundesarchiv erprobt werden.

Worin liegt nun die Motivation des Verfassers dieser Arbeit für die folgende Untersuchung begründet? Obwohl die nachträgliche Verschlagwortung in der Archivcommunity teilweise skeptisch gesehen wird,³ ist es das

¹ Stefanie Jost: Zugänglichmachung on demand – (nicht nur) Erschließung neu gedacht. Der Strategieprozess im Bundesarchiv, in: Erschließung 2.0: Erwartungen.- Probleme.- Lösungen. (Tagungsband zum 23. Sächsischer Archivtag in Leipzig), Chemnitz 2020, S. 35–39, hier S. 37. Es handelt sich hierbei um eine Kombination aus unterschiedlichen Ansätzen: Eine Bestandserschließung findet künftig nur noch bei so genannten Leitbeständen (u.a. B 136 Bundeskanzleramt) statt, während sich die Erschließung ansonsten in Richtung von themen- und politikfeldbezogenen Projekten verlagert. Hier werden also als relevant erachtete Unterlagen einzelner Bestände geplant und je nach Relevanz flach, mittel oder tief erschlossen. Derweil soll bei unerschlossenen und nicht derart priorisierten Unterlagen nur noch eine Erschließung on Demand stattfinden.

² Vgl. zur digitalen Transformation im Bundesarchiv in erster Linie Andrea Hänger/Michael Hollmann: Das Bundesarchiv im digitalen Wandel, in: Das Bundesarchiv im digitalen Wandel (Forum. Das Fachmagazin des Bundesarchivs, 2018), S. 4–33.

³ Bernhard Homa schließt etwa die Integration von Sachschlagworten für die Erschließung im Landesarchiv NRW ausdrücklich aus (vgl. Bernhard Homa: Voraussetzungen und Kriterien für den Einsatz von Normdaten im Landesarchiv NRW (Transferarbeit des 50. wissenschaftlichen Lehrgangs an der Archivschule Marburg), Marburg/Lahn 2019 (E-Paper zuletzt abgerufen am 7. Nov. 2022 unter: < <https://archiv.ub.uni-marburg.de/es/2020/0010/pdf/sbh.pdf> >). Mario Glauert verweist insbesondere auf die enormen und eher zunehmenden Erschließungsrückstände, weshalb er eine „Schlagwortergänzung“ als „Form der erweiterten Erschließung“ kaum

erklärte Ziel des Bundesarchivs, „den Einsatz von Normdaten [...] gerade mit Blick auf GND-Sachbegriffe“ voranzutreiben.⁴ Allerdings ist man sich auch hier des Aufwands durchaus bewusst, den die Schaffung eines „zweifelloso sehr nutzerfreundlichen Zugangs“ mit sich bringen wird: So handele es sich bei der Verschlagwortung „im Grunde [um] eine zweite Erschließung unter anderen Vorzeichen.“⁵ Ob eine flächendeckende Ausbreitung des „vielversprechenden Ansatzes“ möglich ist, werde sich zeigen.⁶

In der vorliegenden Transferarbeit soll deshalb untersucht werden, ob eine Sacherschließung durch Verschlagwortung u. a. mit GND-Sachbegriffen aktuell in Bezug auf Archivgut mit thematischem Schwerpunkt und perspektivisch für das gesamte Archivgut im Bundesarchiv in Frage kommen kann, ob der Mehrwert also den Mehraufwand rechtfertigen oder sogar unter bestimmten Voraussetzungen überwiegen kann. Nach einer eher theoretischen Darstellung zur Relevanz einer Erschließung mit Normdaten im Archivbereich sowie des diesbezüglichen Status-Quos und den Planungen im Bundesarchiv soll der Aufwand einer nachträglichen, normdatenbasierten Verschlagwortung von Archivgut beispielhaft anhand der Erstellung eines Konzepts für eine Teilsystematik für das „Themenportal Wiedergutmachung“ überprüft werden. Geht es hier bereits darum, Voraussetzung und Möglichkeiten einer Reduzierung der Anstrengungen durch ein strukturiertes Vorgehen unter den gegenwärtigen Bedingungen zu beschreiben, sollen im Anschluss heute schon in Ansätzen mögliche bzw. künftige Alternativen einer sich teilweise auf Künstliche Intelligenz stützenden Automatisierung von Datenbereinigung, Verschlagwortung und Normdatenanreicherung in den Blick genommen werden. Das abschließende Fazit wird die Ergebnisse dieses zweigeteilten Lösungsvorschlags zusammentragen und die voranstehende Leitfrage (Mehrwert oder Mehrarbeit?) beantworten.

2. Die Erschließung mit Normdaten im Archivwesen und im Bundesarchiv

„Der Archivar soll den Inhalt des Archivguts nicht interpretieren“ galt über Jahrzehnte nicht nur im Bundesarchiv als Devise.⁷ Die Erschließung von Archivgut sollte nach diesem Verständnis nur zum Inhalt der Unterlagen hinführen. Eine sachthematische Einordnung findet – wenn überhaupt – über die Klassifikation statt. Der „selbstredende Aktentitel“ und die übrigen Erschließungsinhalte verzichten zudem in der Regel auf die Verwendung eines kontrollierten Vokabulars. Mit einer immer unüberschaubarer werdenden Masse an online bereitgestellten Erschließungsinformationen hat sich in den vergangenen zwei Jahrzehnten zumindest das Wissen um die Relevanz einer auf Standardisierung und Eindeutigkeit setzenden sowie einer die Vernetzung von Daten ermöglichenden Erschließung im deutschen Archivwesen wie auch im Bundesarchiv stetig weiterentwickelt.

2.1 Standardisierung und Normierung der archivischen Erschließung

Die Geschichte der Erschließungsstandards in Deutschland ist lang. Für die alte Bundesrepublik steht der Name Johannes Papritz sinnbildlich für erste Empfehlungen einer Standardisierung, während es im Osten die „Ord-

empfehlenswert findet (Mario Glauert: Müssen wir anders verzeichnen? Erschließung zwischen analogen Archivgewohnheiten und digitalen Nutzererwartungen, in: *Archivpflege in Westfalen-Lippe* 91 (2019), S. 32–37, hier S. 36).

⁴ Mirjam Sprau/Tobias Herrmann: Das Themenportal „Wiedergutmachung nationalsozialistischen Unrechts“ im Archivportal-D, in: *Archivar. Zeitschrift für Archivwesen* 74 (2021), S. 275–277, hier S. 277.

⁵ Tobias Herrmann/Mirjam Sprau: Themenportale – Neue Formen digitaler Präsentation, in: *Archive und Erinnerungskulturen. Zwischen Bereitstellung und Geschichtspolitik* (Forum. Das Fachmagazin des Bundesarchivs, 2021), S. 71–75, hier S.73.

⁶ Die Skepsis gründet sich insbesondere auf die enormen Bearbeitungsrückstände vor allem im Schriftgutbereich. Schon eine Erschließung nach aktuellen Grundsätzen würde beispielsweise in der Abteilung B (Bundesrepublik Deutschland) einen Arbeitsaufwand von knapp 1100 Personenjahren erfordern (vgl. hierzu Tobias Herrmann: Das Bundesarchiv in Zahlen, in: *Das Bundesarchiv im digitalen Wandel* (Forum. Das Fachmagazin des Bundesarchivs, 2018), S. 36–39, hier Tabelle 4, S. 38).

⁷ Vgl. Angelika Menne-Haritz: Erschließung, in: Marcel Lepper/Ulrich Raulff (Hgg.): *Handbuch Archiv. Geschichte, Aufgaben, Perspektiven*, Stuttgart 2016, S. 207–217, insbesondere S. 207 (so nicht wörtlich bei Menne-Haritz, aber in der Praxis alltäglich in vergleichbarer Form zu vernehmen).

nungs- und Verzeichnungsgrundsätze für die staatlichen Archive der DDR“ (OVG) waren, die die Grundlage einer standardisierten Erschließung bildeten.⁸

Seit Mitte der 1990er Jahre werden durch den Internationalen Archivrat (ICA) Standards für die Verzeichnung archivischer Unterlagen herausgegeben.⁹ ISAD(G) sieht eine hierarchische Beschreibung von einzelnen Beständen, also eine Stufenerschließung nach dem Top-Down-Prinzip vor. Gemeinsam mit dem auf ISAD(G) folgenden internationalen Standard zur normierten Beschreibung von Körperschaften, Personen und Familien, ISAAR(CPF), liegen seit längerem zwei Standards für die Erschließung von Archivgut vor,¹⁰ die die Erstellung bzw. den Rückgriff auf Normdateien im Archivbereich vorsehen.¹¹ Die Akzeptanz der Standards in der Archivwelt ist dagegen bis heute unterschiedlich hoch.¹² Zudem wird die umfassende Eignung von ISAAR (CPF)-konformen Normdaten für die Sacherschließung bisweilen kritisch gesehen, da das Hauptaugenmerk des Standards auf der Normierung der Provenienzstellen liegt.¹³

Die XML-Standards Encoded Archival Description (EAD) und Encoded Archival Context (EAC(CPF)) stellen die maschinenlesbare Umsetzung der beiden genannten Standards dar. Mit EAD(DDb) liegt seit 2012 ein deutsches Profil vor, das als Austauschformat für das Archivportal-D dient. Für EAC(CPF) mangelt es dagegen an einem deutschen Pendant. Zwar wird der Standard durch das Archivportal Europa verwendet. Allerdings existiert kein zentraler Anbieter, der eine Normdatei für Körperschaften, Personen und Familien verwaltet und pflegen könnte.¹⁴

Auch deshalb arbeitet eine ICA-Expertengruppe für archivische Erschließung (EGAD) am Aufbau des Standards „Records in Contexts“ (RiC), der zum Teil eine Weiterentwicklung der bisherigen ICA-Standards darstellt

⁸ Johannes Papritz: Die archivische Titelaufnahme bei Sachakten, Marburg 1967; Ordnungs- und Verzeichnungsgrundsätze für die staatlichen Archive der Deutschen Demokratischen Republik, herausgegeben von der Staatlichen Archivverwaltung im Ministerium des Innern der DDR, siehe: < <https://www.fh-potsdam.de/sites/default/files/2022-05/landesfachstelle-archivberatung-erschlie%C3%9Fung-ordnungs-und-verzeichnisgrunds%C3%A4tze-staatliche-archiv-fhpotsdam.pdf> > (zuletzt abgerufen am 22. Nov. 2022).

⁹ Zur Entwicklung der internationalen Erschließungsstandards vgl. Silke Jagodzinski: Internationale Standards für archivische Erschließung. Von ISAD(G) zu Records in Contexts, in: *Archivar. Zeitschrift für Archivwesen* 73 (2020), S. 134–141.

¹⁰ Mit ISDF (International Standard Description of Functions) und ISDIAH (International Standard Description of Institutions with Archival Holdings) liegen zudem seit Ende der 2000er Jahre zwei weitere Standards zur Beschreibung von Funktionen und Archiven vor (vgl. zu ISDF Karsten Kühnel: Funktionen als Räume für institutionelle Episoden – Normdateien nach ISDF, in: Monika Storm (Hg.): *Transformation ins Digitale* (85. Deutscher Archivtag in Karlsruhe), Fulda 2017, S. 135–151).

¹¹ So heißt es in der deutschen Übersetzung von ISAD(G) bezüglich der Validität von Suchergebnissen: „Die Aussagekraft und Eindeutigkeit der über Zugriffspunkte ermittelbaren Informationen wird verbessert, wenn Namen und andere Elemente in Normdateien erfasst und einer besonderen Begriffskontrolle unterworfen werden“ (Rainer Brüning/Werner Heegewaldt/Nils Brübach (Hg.): *ISAD(G) - Internationale Grundsätze für die archivische Verzeichnung*, Zweite, überarbeitete Ausgabe, Marburg 2002 [Übersetzung der Veröffentlichung von 2000], S. 20.

¹² „Though ISAD(G) has significantly influenced international archival descriptive practice, ISAAR(CPF) has some use, and both ISDF and ISDIAH very little“ (ICA: *Records in Contexts. Conceptual Model. Consultation Draft v0.2 July 2021*, S. 8, abgerufen unter: < https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf > ; zuletzt am 22. Nov. 2022).

¹³ Bezüglich des Normdatenstandards ISAAR (CPF) hat sich Bernhard Homa unter anderem aus diesem Grund skeptisch im Hinblick auf seine Eignung für umfassende archivische Erschließungszwecke gezeigt. Eine Anreicherung mit ISAAR (CPF)-konformen Normdaten komme seiner Ansicht nach vor allem für eine Erschließung im Nachlassbereich (Beispiel Kalliope-Portal) in Frage (vgl. Homa: *Voraussetzungen*, S. 9 f.). Dieser Meinung ist beizupflichten. Während Homa jedoch stattdessen ausschließlich die Gemeinsame Normdatei (GND) als Normdatenrepositorium für das Landesarchiv NRW empfiehlt, betrachtet der Verfasser dieser Arbeit andere Normdateien (z. B. Wikidata) als ebenso relevant für die künftige Erschließung. Entsprechend wird das Bundesarchiv die Weiterentwicklung seines Archivinformationssystems zu BASYS 3 Akte mit einer generischen Schnittstelle versehen, um die Verknüpfung zu unterschiedlichen Normdateien zu ermöglichen (vgl. hierzu Kapitel 2.2.1).

¹⁴ Vgl. Homa: *Voraussetzungen*, S. 10.

und insgesamt zu mehr Akzeptanz innerhalb des Archivwesens führen soll.¹⁵ Er geht von einem erweiterten Provenienzprinzip aus. Anders als ISAD(G) erlaubt RiC eine multidimensionale Verzeichnung, also eine Kontextualisierung über Bestandsgrenzen hinweg, die eben nicht nur Registraturbildner, sondern unter anderem auch Sachbegriffe (RiC Subject) miteinbezieht. RiC hat den Anspruch, die archivische Verzeichnung gewissermaßen „fit“ für die „Linked Data“-Welt zu machen, weshalb eine Normierung und Disambiguierung der durch den Standard dokumentierten Entitäten von großer Bedeutung ist.

Eine Normierung der erschlossenen Namen, Orte oder von Allgemeinbegriffen fand derweil in deutschen Archiven anders als im Bibliotheksbereich lange Zeit nicht statt. Einen wichtigen Impuls zur Beschäftigung mit Normdaten im Archiv lieferte im Jahr 2014 Franz-Josef Ziwes mit seinem vielbeachteten „Leuchttürme“-Beitrag.¹⁶ Dass allerdings noch heute ein Vortrag zum Thema „Erschließung biographischer Unterlagen: Nutzung der Gemeinsamen Normdatei (GND) durch Archive“ mit den zwei rhetorischen Fragen „Wozu Normdaten?“ und „Was sind und wie funktionieren Normdaten?“ beginnt,¹⁷ ist geradezu bezeichnend für das offenbar weiterhin existierende diesbezügliche „Akzeptanzproblem“ unter den Archivarinnen und Archivaren.

Dabei wird das Potential, das „Normdaten als Erschließungsinstrument“¹⁸ in sich bergen, in den letzten Jahren durchaus verstärkt gesehen, was auch an der sich allmählich durchsetzenden Überzeugung liegt, dass eine spartenübergreifende Vernetzung der Informationsangebote von Kulturinstitutionen in einer digitalisierten Welt nur gelingen kann, wenn sich auch außerhalb der Bibliotheken, also im Museums- und Archivwesen, die Anreicherung der Erschließungsinformationen mit Normdaten etabliert. Die „Öffnung“ der in erster Linie durch die Deutsche Nationalbibliothek betriebenen GND für nicht-bibliothekarische Kultursparten und Anwendungskontexte steht sinnbildlich für dieses „Umdenken“ und treibt diese Tendenz weiter voran.¹⁹

Neben Wikidata, das seine Normdaten aus den Daten der WikiMedia (u.a. Wikipedia) speist und diese mit eindeutigen Identifikatoren zur Verfügung gestellt, ist die GND die wohl bekannteste und umfangreichste Datensammlung im Kulturbereich innerhalb des deutschsprachigen Raums,²⁰ die ursprünglich aus unterschiedlichen Normdateien hervorgegangen ist.²¹ Die Datei umfasst über 9 Millionen Datensätze unterschiedlicher Enti-

¹⁵ Vgl. zum Referenzmodell „Records in Contexts“ Gerhard Müller/Silke Jagodzinski: Die Erschließung des Kontexts. Neue Perspektiven auf ein bewährtes Prinzip, in: *Archivpflege in Westfalen-Lippe* 90 (2019), S. 10–17.

¹⁶ Franz-Josef Ziwes: *Archive als Leuchttürme. Die Erschließung mit Normdaten als Aufgabe und Chance*, in: *Archive ohne Grenzen. Erschließung und Zugang im europäischen und internationalen Kontext*. 83. Deutscher Archivtag in Saarbrücken (Tagungsdokumentation zum Deutschen Archivtag 18), Fulda 2014, S. 79–87.

¹⁷ Vgl. Livestream-Mitschnitt des Vortrags von Andreas Neuburger (LABW) unter <<https://www.youtube.com/watch?v=5XcdCoIZtew>> (zuletzt abgerufen am 7. März 2022). In der Verschriftlichung des Beitrags fühlt sich Neuburger im Hinblick auf Normdaten zu der Aussage genötigt, dass die Frage nicht mehr sei, „ob Archive hier aktiv werden, sondern sehr viel eher, wann und wie dies geschieht“ (Andreas Neuburger: *Vom Desiderat zur gelebten Praxis! Die Nutzung der Gemeinsamen Normdatei (GND) in der Erschließung*, in: *Archivpflege in Westfalen-Lippe* 95 (2021), S. 30–34, hier S. 33).

¹⁸ Nils Brübach: *Zur archivischen Erschließung von Filmen und Fotos mit Normdaten*, in: *Archivpflege in Westfalen-Lippe* 87/2017, S. 16–19, hier S. 16.

¹⁹ Die GND soll künftig als eine Art Knotenpunkt für die Vernetzung von Kulturinstitutionen dienen. Das DFG-Projekt „GND für Kulturdaten“ (GND4C) steht sinnbildlich für dieses Vorhaben. Zur Repräsentation von Sachbegriffen im GND-Datenmodell aus Museumsperspektive vgl. Jens M. Lill/Jutta Lindenthal/Detlev Balzer: *Personeninformationen und Sachbegriffe aus Museen in der GND und ihre Repräsentation im GND-Datenmodell*, in: Detlev Balzer u.a.: *Das Projekt „GND für Kulturdaten“ (GND4C)*, in: *O-bib* 2019/4, S. 59–97, hier S. 70–78. Aus Archivsicht hierzu knapper Neuburger: *Vom Desiderat*, S. 32. Womöglich wird es bei weiterem Ausbau der Kooperation mit der GND notwendig sein, auf eine Abbildung neuer archivspezifischer Entitäten hinzuwirken.

²⁰ Weitere einschlägige Normdateien und Thesauri sind etwa die Normdatei der Library of Congress (Library of Congress Control Number – LCCN), die virtuelle internationale Normdatei (Virtual International Authority File – VIAF) oder der Getty Thesaurus für geographische Namen (Thesaurus of Geographic Names – TGN).

²¹ Sie entstand 2012 aus den zuvor getrennt existierenden Personennamendatei (PND), Gemeinsame Körperschaftsdatei (GKD), Schlagwortnormdatei (SWD) sowie der Einheitssachtitel-Datei des Deutschen Musikarchivs (DMA-EST-Datei) (vgl. Renate Behrens-

tätentypen (Stand: November 2021).²² Neben Personen, Körperschaften, Konferenzen/Veranstaltungen, Geografika und Werktiteln zählen hierzu Sachschlagwörter. Zu Letzteren gehören unter anderem die Allgemeinbegriffe, die keinem anderen Entitätentyp zuzuordnen sind (Entitäten-Codierung: saz). Im Bereich der Sachschlagwörter stellt die GND ein kontrolliertes Schlagwortsystem dar, das neben der Eindeutigkeit (Angabe von Definitionen) eine vollständige Relationierung durch die Angabe von Ober- und Unterbegriffen, Synonymen und verwandten Begriffen anstrebt.

Die Vorteile einer Erschließung mit normdatenbasiertem, kontrolliertem Vokabular liegen auf der Hand und sollen nur knapp beschrieben werden.²³ Die Eindeutigkeit, Identifizierbarkeit und Auffindbarkeit von Begrifflichkeiten und Personen wird für Mensch und Maschine durch gewisse „konstitutive[...] Inhalte[...] eines Normdatensatzes“²⁴ garantiert: Durch die Angabe von Merkmalen, die der Mensch für eine Unterscheidung benötigt (z. B. bei Begriffen: Vorzugsbenennung, alternative Schreibweisen, Synonyme; bei Personen: Lebensdaten, Hinweise zum Beruf etc.) und die Vergabe eines Identifikators (ID), der auch die maschinelle Weiterverarbeitung ermöglicht. Die weiteren Vorteile der Referenz auf eine konkrete Normdaten-Entität leiten sich hieraus gewissermaßen ab: Die Vernetzung und Interoperabilität der Daten ist möglich (Linked Data); erweiterte Recherchemöglichkeiten ergeben sich (Facetten-Suche); es bieten sich neue, digitale Darstellungsmöglichkeiten (Netzwerk-Visualisierungen).²⁵ In Kombination mit einer weitgehend lizenzfreien Nachnutzbarkeit der Erschließungsdaten etwa über das Archivportal-D lässt sich auf diese Weise nicht zuletzt auch ein „FAIRer“ Zugang zu digitalisiertem Archivgut gewährleisten, was zudem immer häufiger Voraussetzung für die Vergabe von Drittmitteln ist.²⁶

Für die Anreicherung von Daten in sachthemenatischen Themenportalen sind die Sachschlagwörter der Normdateien besonders relevant. Denn eine Sachrecherche unterscheidet sich grundsätzlich von einer Personenrecherche. Während bei einer Person zwar „Namensvettern“ sowie unterschiedliche Schreibweisen eines Namens beim Suchergebnis ins Gewicht fallen können, ist die Wahrscheinlichkeit dennoch hoch, dass eine Person „recherchierbar“ ist, wenn im Rahmen einer Erschließung Namen zumindest grundsätzlich erfasst werden. Durch eine Verknüpfung mit einer Normdatei erlangt die Personenrecherche zusätzlich Eindeutigkeit. Die Suche eines Allgemeinbegriffs über die Volltextrecherche erweckt zwar den Anschein, ähnliche Ergebnisse auswerfen zu können wie eine Google-Recherche. Sie wird aber aller Wahrscheinlichkeit nach niemals mit dem auf Semantik, Linguistik und Statistik basierenden Google-Algorithmus mithalten können.²⁷ Über den Suchschlitz lässt sich nur das auswerfen, was zuvor in Aktentiteln, Enthält-Vermerken etc. erschlossen wurde. Synonyme und ver-

Neumann: Die Gemeinsame Normdatei (GND). Ein Projekt kommt zum Abschluss, in: *Dialog mit Bibliotheken* 24/1 (2012), S. 25–28, hier S. 25; zuletzt abgerufen am 7. Nov. 2022 unter < <https://d-nb.info/1026682169/34> >).

²² < https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/Gesamtabzuege/gesamtabzuege_node.html > (zuletzt abgerufen am 7. Nov. 2022).

²³ Vgl. hierzu ausführlich u.a. Gerhard Müller: Die Normdatei als ein Mittel in der Erschließung von Archivbeständen, in: *Brandenburgische Archive* 32 (2015), S. 8-16; außerdem bezogen auf die GND Daniel Fähle: Gemeinsame Normdatei (GND) und Archive – eine vielversprechende Liaison, in: *Erschließung 2.0: Erwartungen.- Probleme.- Lösungen.* (Tagungsband zum 23. Sächsischer Archivtag in Leipzig), Chemnitz 2020, S. 51–53, insbes. S. 51 f.

²⁴ Ebd., S. 51.

²⁵ Vgl. ebd., S. 51 f.

²⁶ FAIR steht als Akronym für “findable, accessible, interoperable, reusable”; im Kontext der Forschungsdatenarchivierung hierzu Gerald Maier/Daniel Fähle/Andreas Neuburger: Bereitstellung, Aufbereitung, Langzeitsicherung. Funktionen der Archive in der Forschungsdateninfrastruktur, in: *Archivar. Zeitschrift für Archivwesen* 73 (2020), S. 13–18, hier S. 15.

²⁷ Vgl. Daniel Fähle: Herausforderungen bei der Benutzerführung in Portalen, in: Irmgard Christa Becker (Hg.): *Netzwerken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung.* Beiträge zum 19. Archivwissenschaftlichen Kolloquium der Archivschule Marburg 2015 (Veröffentlichungen der Archivschule Marburg 61), S. 269–283, hier S. 280; Glauert: Müssen wir anders verzeichnen?, S. 35.

wandte Begriffe werden außer Acht gelassen, während Homonyme (gleiches Wort, andere Bedeutung) die Recherche sogar in eine falsche Richtung lenken können.

Diese Problematik einer semantischen Unschärfe kann eine Anbindung von Sachschlagwörtern an Archivgut im Rahmen der Sacherschließung auflösen. Im Gegensatz zu Stichwörtern, die in den Erschließungsinformationen vorkommen müssen, entstammen Schlagwörter üblicherweise einem Thesaurus, einem kontrollierten Vokabular, das darüber hinaus im Idealfall – wie im „Themenportal Weimarer Republik“ und künftig im „Themenportal Wiedergutmachung“ – zusätzlich auf eine Normdatei referenziert. Hierdurch werden die Voraussetzungen geschaffen, damit eine Volltextsuche perspektivisch zu einer semantischen Suche ausgebaut werden kann.²⁸

Im Gegensatz zur Indexierung und Anreicherung mit Personen- oder Ortsnormdaten²⁹ ist in deutschen Archiven eine normdatenbasierte Verschlagwortung von Sachbegriffen eher unüblich. Hierfür dürfte einerseits eine gewisse Scheu vor einem vermeintlichen Aufweichen des Provenienzprinzips durch eine ergänzende Pertinenzerschließung verantwortlich sein, andererseits aber auch das hiermit zusammenhängende Fehlen von Thesauri im Archivbereich.³⁰ Die Bedeutung von mit Normdaten angereicherten Sachbegriffen dürfte jedoch deutlich steigen – und zwar als technische Voraussetzung für semantische Verknüpfungen und damit auch zur Anwendung neuer Erschließungsstandards, wie etwa RiC. Denn erst durch die Verknüpfung mit einer Normdatei können aus „herkömmlichen“ Erschließungsdaten zeitgemäße „Linked Data“ werden und diese damit erst in neue Kontexte eingebunden und recherchierbar werden.

2.2 Gegenwart und Zukunft der Erschließung und Verschlagwortung mit Normdaten im Bundesarchiv

Für einen flächendeckenden Normdateneinsatz im Archiv müssen zwei Grundvoraussetzungen erfüllt sein: Die technische Weiterentwicklung der Archivfachinformationssysteme sowie eine weitgehend automatisierte Normdatenanreicherung.³¹ Wo steht das Bundesarchiv in dieser Hinsicht und was sehen aktuelle Pläne vor?

Eine Verknüpfung von Erschließungsinformationen mit Normdaten findet im Bundesarchiv bisher nur in wenigen Ausnahmefällen statt. Neben dem erhöhten Arbeitsaufwand, der bislang vor einem flächendeckenden Normdateneinsatz zurückschrecken ließ, sind auch die technischen Voraussetzungen hierfür nicht gegeben. So bereiten insbesondere die existierenden Archivfachinformationssysteme Schwierigkeiten. Eine Normdatenanreicherung ist zumeist nur durch aufwändige, oft manuelle Lösungen im Sinne einer maschinenlesbaren Weiterverwendung der Daten möglich.

Pläne für die Weiterentwicklung des existierenden Archivfachinformationssystems BASYS 2 Akte um Möglichkeiten einer Normdatenverknüpfung werden im Bundesarchiv aktuell vorangetrieben. Zugleich wird durch die Etablierung von weiteren digitalen Zugängen zu Archivgut der Normdateneinsatz in gewisser Weise erprobt, so insbesondere durch „Online-Plattformen“ wie das „Themenportal Weimarer Republik“.

²⁸ Vgl. Nils Meyer: Sachthematische Zugänge im Archivportal-D. Archive und ihre Bestände zusammenführen und neu entdecken, in: *Archivar. Zeitschrift für Archivwesen* 72 (2019), S. 37–39, insbes. S. 38.

²⁹ Zahlreiche – so auch viele der hier bereits genannten – Beiträge zur Integration von Normdaten beschäftigen sich mit Personen- oder Ortsnormdaten und (automatisierten) Verfahren der Anreicherung. Vgl. hierzu außerdem Daniel Fähle/Nadine Seidu: Mit „Match & Merge“ zur GND? Erprobung von Anreicherungsverfahren im Archivportal-D-Projekt, in: *Archivar. Zeitschrift für Archivwesen* 70 (2017), S. 196–198.

³⁰ Vgl. Glauert: Müssen wir anders verzeichnen?, S. 36.

³¹ So die einhellige Meinung dreier ArchivarInnen des Landesarchivs Baden-Württemberg, des Landesarchivs Hessen und des Bundesarchivs (vgl. Wolfgang Krauth/Peter Sandner/Sina Westphal: Archivische Erschließung mit Normdaten, in: *Archivar. Zeitschrift für Archivwesen* 73 (2020), S. 142–144, hier S. 143).

2.2.1 Archivfachinformationssysteme heute und morgen

Die Anbindung von Personennormdaten findet zwar in unterschiedlichen Abteilungen des Bundesarchivs statt, dabei jedoch je nach Unterlagenart auf unterschiedliche Weise. Im Bildarchiv gelang durch die vorübergehende Kooperation mit Wikimedia im Jahr 2008 innerhalb von wenigen Monaten ein für die damaligen Verhältnisse technischer Meilenstein:³² So konnte mit Hilfe eines durch Wikipedianer entwickelten Tools ein einmaliger Abgleich der im Bundesarchiv erstellten Personenliste (knapp 60.000 Einträge) mit der damaligen Personennormdatei (heute Teil der GND) sowie mit vorhandenen Wikipedia-Einträgen durchgeführt werden. Nach erfolgtem Abgleich wurde die um die PND-ID ergänzte Personenliste in die interne Bilddatenbank DC5 eingespielt und ermöglichte dort eine automatische Verlinkung auf Wikipedia-Artikel.

Allerdings blieb es bei dieser zeitweiligen teilautomatisierten Anreicherung. Eine dauerhafte Schnittstelle zur GND wurde nicht implementiert und existiert in DC5 auch heute nicht. Bei neu zu erschließenden Personen wird eine GND-ID manuell in die vorhandenen Textdateien der DC5-Personenliste eingetragen und kann anschließend über die Volltextsuche im Digitalen Bildarchiv recherchiert werden.

In BASYS 3 Film gibt es seit 2016 die Möglichkeit, GND-Personennormdaten, aber auch GND-Sachschlagworte sowie die übrigen GND-Entitäten einzubinden.³³ Mittlerweile ist die Version 2.0 von BASYS 3 Film im Wirkbetrieb. Nun ist durch Abfrage der kostenfreien SRU-Schnittstelle³⁴ der GND eine gezielte Suche und unmittelbare Übernahme von Personen- und Körperschaftsdatensätzen als Ergänzung der filmographischen Angaben sowie die einfache Indexierung mit GND-Sachschlagworten möglich. Die meisten inhaltlichen Beschreibungen finden unter den „Manifestationen“ (Versionen/Exemplare der Filmwerke) statt. Die Einbindung der GND-Sachschlagworte in einer hierarchischen Systematik, einem kontrollierten Vokabular findet nicht statt. Im Rahmen der Erschließung kann die Verschlagwortung daher relativ flexibel, aber damit theoretisch auch fehlerhaft erfolgen. Ein eigener Thesaurus wurde aufgrund der Komplexität und Themenvielfalt im Filmbereich nicht aufgebaut. Anbieten würde sich allenfalls eine Anbindung an einen „general purpose“-Thesaurus, etwa den des Getty Research Institutes. Auch ein Einspeisen von bislang nicht in der GND existierenden Personen oder Schlagwörtern erfolgt nicht. Die Daten der GND werden somit nur gespiegelt und gespeichert, was aber dennoch perspektivisch Möglichkeiten für eine Vernetzung im Semantic Web eröffnet.

Auch im Bereich der Editionen erfolgt eine Personennormdaten-Referenzierung. Während die Online-Edition der „Akten der Reichskanzlei“³⁵ ebenso wie die Nachlassdatenbank (ZDN)³⁶ über eine Verlinkung zur deutschen Biografie (www.deutsche-biografie.de) Verknüpfungen zur GND besitzen, findet sich in der Online-Edition „Die Kabinettsprotokolle der Bundesregierung“³⁷ unter dem Reiter „Biographien“ immerhin die manuell eingetragene GND-ID der jeweiligen Person allerdings ohne Link.

Die Erschließung von Schriftgut wird im Bundesarchiv seit 2018 mit BASYS 2 Akte durchgeführt. Eine Integration von GND-Normdaten ist nur durch manuelles Hinterlegen der GND-ID im Modul „Provenienzen“ möglich. Auf diese Weise kann zumindest die Voraussetzung für ein nachträgliches maschinelles Auslesen geschaf-

³² Für diesen Abschnitt zum Bereich Bildarchiv wird sich auf die Auskunft von Oliver Sander (BArch, Referat IT 2) per E-Mail vom 8. Febr. 2022 bezogen.

³³ Für diesen Absatz wird auf die Auskünfte von Georg Eckes (BArch, Referatsleiter IT 3) zurückgegriffen (Videokonferenz am 11. Febr. 2022).

³⁴ SRU steht für „Search/Retrieve via URL“ und ist ein Standard, der von der Library of Congress veröffentlicht wurde. Als maschinenlesbares Austauschformat dient bei dieser Abfrage das XML-Format Marc 21.

³⁵ < <https://www.bundesarchiv.de/aktenreichskanzlei/1919-1933/0000/index.html>> (zuletzt abgerufen am 21. Nov. 2022).

³⁶ < <https://www.bundesarchiv.de/nachlassdatenbank/>> (zuletzt abgerufen am 21. Nov. 2022).

³⁷ < <https://www.bundesarchiv.de/cocoon/barch/0000/index.html>> (zuletzt abgerufen am 21. Nov. 2022).

fen werden. Allerdings wird schon intensiv an einer Weiterentwicklung hin zu BASYS 3 Akte gearbeitet, wo die Anbindung der Erschließungsinformationen an Normdateien fest verankert sein wird.³⁸ Konkrete Überlegungen beinhalten den Aufbau einer barrierefreien, modularisierten Webanwendung. Die Bestände, Provenienzen, Serien, Subserien und Verzeichnungseinheiten sollen künftig mit GND-Normdaten für Personen, Körperschaften, Geografika und Ereignisse manuell angereichert werden können. In der Diskussion ist zudem die Verknüpfung mit GND-Sachschlagworten zu bestimmten Rechtsformen (Oberste Bundesbehörde, Bundesoberbehörde etc.). Hierdurch wäre es künftig möglich, sich sämtliche Provenienzen mit einer bestimmten Rechtsform anzeigen zu lassen. Ein Nachhalten der Fachaufgaben der Behörden in den Stammdaten könnte darüber hinaus durch GND-Sachschlagworte unterstützt werden, auch um sich die Möglichkeit einer Funktionserschließung gemäß RiC zu eröffnen.

Es wird also insgesamt eine Schnittstelle für die manuelle Verknüpfung mit existierenden Normdateien und Thesauri angestrebt. Durch die Ausweitung der Zusammenarbeit, insbesondere mit der GND, soll auf bereits vorhandenes aufgebaut werden.³⁹ Das Einspeisen von eigenen Erschließungsdaten in die GND über so genannte GND-Agenturen wird diskutiert und im Rahmen der Themenportal-Projekte im Kleinen erprobt. Doch soll es sich um eine generische Schnittstelle handeln, die eine Verbindung zu unterschiedlichen Normdateien (GND, Wikidata etc.) und die Integration der dort hinterlegten Daten im eigenen System möglich macht.⁴⁰ Die Daten sollen dabei regelmäßig nachgeladen werden. Auch die Suchfunktion in BASYS 3 Akte soll entsprechend weiterentwickelt werden und Normdaten künftig in die Suche einbeziehen.

Von den Vorteilen der Nutzung von Normdaten hinsichtlich der eindeutigen Adressierbarkeit und der archivübergreifenden Nachnutzung der Daten überzeugt, geht es dem Bundesarchiv darum, mit BASYS 3 Akte die Weichen für eine zeitgemäße Recherchierbarkeit und Vernetzung seiner Erschließungsdaten im Semantic Web zu stellen. Gleichzeitig ist man sich bewusst, dass hiermit „nur“ eine normdatengerechte Archivsoftware entsteht. Diese allein wird nicht für eine flächendeckende Verknüpfung von Erschließungsdaten und Normdaten sorgen. Ein solcher Prozess wird weitgehend automatisiert erfolgen müssen.⁴¹

Zudem ist man sich im Klaren, dass die Entwicklung von BASYS 3 Akte sowie die perspektivische Bereitstellung von Unterlagen über einen virtuellen Lesesaal noch ein paar Jahre benötigen könnten.⁴² So lässt sich das „Ausweichen“ auf „Themenportale“ – neben dem Hauptgrund, einer archivübergreifenden Zugänglichmachung von digitalisierten Unterlagen und Erschließungsinformationen – in gewisser Weise auch als Notlösung bezeichnen. Vor allem aber bieten sich hier digitale Plattformen, auf denen sich neue Wege einer archivischen Erschließung erproben lassen.

³⁸ Zum Folgenden eine Zusammenfassung der Auskünfte von Sarah Bormer, Vincent Kutz, Oliver Sander (alle BArch, Referat IT 2) vom 21. Dez. 2021 und 10. Jan. 2022.

³⁹ Damit erübrigt sich auch vorerst die Diskussion der Frage, ob es möglich bzw. notwendig ist, für eine perspektivisch flächendeckende Verschlagwortung einen einheitlichen Bundesarchiv-Thesaurus oder gar einen solchen für die gesamte deutsche Archivlandschaft anzustreben. Wichtig wird es bis auf weiteres sein, zentral verwaltete Normdateien bzw. Thesauri für Archivzwecke tauglich zu machen.

⁴⁰ Durch die generische Schnittstelle soll BASYS 3 Akte flexibel bleiben, auch im Hinblick auf künftige Weiter- bzw. Neuentwicklungen innerhalb des Semantic Web (Auskunft von Sarah Bormer, 21. Dez. 2021).

⁴¹ Die Perspektiven hierzu, werden in Kapitel 4, insbesondere 4.2 skizziert.

⁴² Telefonische Auskunft von Maurice Augel (BArch, Referat IT 1) vom 7. Febr. 2022.

2.2.2 „Themenportal Weimarer Republik“ – Systematik und Verschlagwortung

Das Bundesarchiv stellt Erschließungsinformationen zu seinem Archivgut auf gleich mehreren nationalen und internationalen Online-Plattformen zur Verfügung.⁴³ Die ersten, noch deutlich kleineren Portale dieser Art entstanden im deutschen Archivwesen schon vor etwa 25 Jahren.⁴⁴ Mit dem Aufbau eines Onlineportals für das kulturelle Erbe Europas (Europeana) in den 2000er Jahren bestand die Notwendigkeit einen nationalen Aggregator in Deutschland auf den Weg zu bringen (Deutsche Digitale Bibliothek – DDB). Innerhalb der DDB wurde 2014 das Archivportal-D als spartenspezifischer Zugang zu Unterlagen aus deutschen Archiven freigeschaltet, das wiederum als Aggregator für das Archivportal Europa (Archives Portal Europe) dient.⁴⁵

Heute gelten Archivportal-D-Projekte wie das „Themenportal Weimarer Republik“ oder das aktuell im Aufbau befindliche und hier näher zu betrachtende Themenportal zur „Wiedergutmachung“ als Errungenschaft. Sie stellen gewissermaßen Experimentierfelder dar, um mit Hilfe technischer, digitaler Tools, die Erschließung und Präsentation von Archivgut erneut den veränderten Nutzergewohnheiten und -erwartungen anzupassen und sich dem Semantic Web zu öffnen.

Zwei Ziele verbanden sich daher mit dem Pilotprojekt eines Portals zur Weimarer Republik: Ein Thesaurus zu eben jener Epoche sollte einerseits einen thematisch strukturierten Zugang zu Archivgut der Jahre 1918/19 bis 1933 innerhalb des Archivportals-D ermöglichen. Es sollten hierdurch „Perspektiven für eine zeitgemäße Bereitstellung von Erschließungsinformationen“ aufgezeigt werden.⁴⁶ Dabei scheint die Akzeptanz des als Desiderat⁴⁷ bezeichneten sachthematischen Zugangs für die Nutzerinnen und Nutzer evident.⁴⁸ Andererseits erhoffte man sich hierdurch, der archivischen Erschließung „neue[...] Impulse“ zu verleihen.⁴⁹ Beim „Themenportal Weimarer Republik“ geht es also nicht bloß um eine nutzerfreundliche Ergänzung der bisherigen Art der Zugänglichmachung von Archivgut. Es sollen grundsätzlich neue Wege einer archivischen Verzeichnung beschrritten werden, die den Erfordernissen des digitalen Wandels sowie der digitalen Vernetzung Rechnung tragen.

Da die Verschlagwortung mit Sachbegriffen und die Schaffung von sachthematischen Zugängen außerhalb von Bibliotheken eher unüblich ist, erscheinen ein paar Vorbemerkungen notwendig, um die Nützlichkeit adäquat beschreiben zu können. Eine archivische Klassifikation ergibt sich aus der Struktur eines Bestandes und orientiert sich gemäß Provenienzprinzip an den Aufgaben, Kompetenzen und Funktionen der jeweiligen Provenienzstelle. Eine systematische Ordnung nach Sachbetreffen ohne Berücksichtigung des Entstehungskontexts

⁴³ Ein Überblick findet sich hier: < <https://www.bundesarchiv.de/DE/Content/Artikel/Finden/Unterseiten-Weitere-Recherchesysteme/uebergreifende-portale.html> > (abgerufen am: 16. Nov. 2022).

⁴⁴ Den technischen Fortschritt aufgreifend wurde damals (wie heute) der Versuch unternommen, die Sichtbarmachung und Zugänglichkeit von Archivgut zu verbessern. Einige dieser Frühentwicklungen existieren heute nicht mehr (vgl. Karsten Uhde: Suchet, so werdet ihr finden – die Geschichte der Archivportale, in: Irmgard Christa Becker (Hg.): Netz werken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung. Beiträge zum 19. Archivwissenschaftlichen Kolloquium der Archivschule Marburg 2015 (Veröffentlichungen der Archivschule Marburg 61), S. 47–61).

⁴⁵ Zum Archivportal Europa vor allem Susanne Waidmann: Archivportale auf europäischer Ebene, in: ebd., S. 133-155, Unterlagenspezifisch ist zudem der „Kalliope-Verbund“ als nationales Nachweisinstrument für Nachlässe und Autographen besonders erwähnenswert, insbesondere wegen seiner Relevanz im Bereich der normdatenbasierten Erschließung (Personen, Körperschaften, Orte).

⁴⁶ Meyer: Sachthematische Zugänge, S. 38.

⁴⁷ Der „Bedarf an einem sachthematischen, niederschweligen Zugang zu archivischen Quellen“ sei insbesondere seitens der Digital Humanities und von Archiven geäußert worden (vgl. Herrmann/Sprau: Themenportale, S. 73).

⁴⁸ Technisch bedingt lassen sich keine exakten Zahlen zur Nutzung des Themenportals "Weimarer Republik" nennen, da das vorhandene Tracking Tool bisher nicht zwischen einer Suche im Themenportal und einer Suche im "normalen" Archivportal unterscheidet. Es ließen sich jedoch rund 500 Aufrufe für die Startseite des Themenportals pro Monat ermitteln, was knapp 1/8 der Startseitenauf-rufe für das Archivportal-D entspreche. Insgesamt gehe man daher, insbesondere beim Fachpublikum, von einer guten Akzeptanz des Themenportals aus (Auskunft von Nils Meyer (LABW), E-Mail vom 14. Febr. 2022).

⁴⁹ Meyer: Sachthematische Zugänge, S. 38.

(Pertinenzprinzip) findet überwiegend in Bibliotheken statt. Der Unterschied zwischen archivischer Klassifikation und bibliothekarischer Systematik besteht darin, dass die Systematik „auf einem kontrollierten Vokabular basiert und [...] in allen Bibliotheken identisch verwendet werden kann“.⁵⁰

Wenn nun aber Unterlagen mehrerer Archive und ganz unterschiedlicher Provenienz in einem Online-Portal zusammengebracht werden, kann eine navigierende Suche weiterhin nur auf Grundlage der jeweiligen „hoch proprietären Klassifikationen“⁵¹ erfolgen, wie dies im Archivportal-D allgemein der Fall ist. Ein sachthematisch strukturierter Zugang lässt sich lediglich über eine Systematik, einen Thesaurus erreichen, wie er für das „Themenportal Weimarer Republik“ entwickelt wurde und für das „Themenportal Wiedergutmachung“ angedacht ist. Diesem kontrollierten Vokabular können in letzter Stufe Schlagwörter zu bestimmten Themen zugeordnet sein, die wiederum mit dem Archivgut verknüpft werden. Am Provenienzprinzip wird festgehalten, da der Entstehungskontext und die Herkunft des Archivguts jederzeit über Angaben und Verlinkung zum „Datengeber“ (das jeweilige Archiv) nachvollziehbar bleiben.

Die inhaltliche Erstellung des Thesaurus zur Weimarer Republik stellt eine Gemeinschaftsarbeit des Bundesarchivs und des Landesarchivs Baden-Württemberg dar. Sie erfolgte über die kollaborative Vokabularverwaltung digiCULT.xTree, eine webbasierte Applikation, die mehr als 200 Vokabulare und Teilvokabulare verwaltet. Sie nutzt hierfür unter anderem die Standards RDF und JSON und ermöglicht einen Abgleich mit und ein Mapping auf Normdaten der GND, Wikidata, VIAF u. a.⁵²

Der Thesaurus ist online über das Zugangsportal des Vokabular-Management-Tools xTree.public abrufbar.⁵³ Der auf dieser Systematik basierende sachthematische Zugang im Themenportal „Weimarer Republik“ gliedert sich aktuell in 17 Oberkategorien. Die darunter liegenden Themen führen zu derzeit ca. 1000 Schlagwörtern, die allesamt mit Archivgut (Stand Februar 2022: 44.929 Objekte aus 31 Archiven) verknüpft sind. Innerhalb des xTree-Vokabulars verfügen sie zusätzlich über stabile URIs, bevorzugte und alternative Bezeichnungen (hier auch die weibliche Form von bestimmten Begriffen) sowie einen dreistelligen Entitätencode (z. B. „saz“ für Allgemeinbegriff). Die Problematik, dass die GND- oder Wikidata-Begriffe nicht immer exakt zu den herausgearbeiteten Schlagwörtern passen, wird durch das so genannte „Mapping“ aufgefangen, bei dem festgelegt wird, ob der Begriff „exact“, „broader“ oder „narrower“ mit den jeweiligen Sachbegriffen übereinstimmt.⁵⁴

Der Prozess der eigentlichen Verschlagwortung des Archivguts soll durch ein so genanntes „Schlagwort-tool“ erheblich erleichtert werden. Das Tool ist eine vom FIZ Karlsruhe, dem Landesarchiv Baden-Württemberg und dem Bundesarchiv entwickelte, überaus anwenderfreundliche Webanwendung. Verzeichnungseinheiten können durch die teilnehmenden Archive aus dem Archivportal-D unmittelbar ausgewählt und mit Deskriptoren verknüpft werden. Durch einen entsprechenden Algorithmus werden aus den Erschließungsdaten heraus zu manchen Objekten automatisch Vorschläge für eine Indexierung gemacht. Das „Schlagworttool“ wird seit dem

⁵⁰ Glauert: Müssen wir anders verzeichnen?, S. 35.

⁵¹ Ebd., S. 35.

⁵² Die Möglichkeit einer Anreicherung des xTree-Thesaurus zur „Weimarer Republik“ auf Deskriptor-Ebene besteht für weitere teilnehmende Archive. Der Zugriff erfolgt über eine REST/JSON-Schnittstelle, was auch die Nachnutzung zu anderen Forschungszwecken möglich macht (vgl. hierzu unter anderem die Informationen auf der Startseite von xTree.public unter <<http://xtree-public.digicult-verbund.de/vocnet/?action=start&lang=de>> ; zuletzt abgerufen am 15. Nov. 2022).

⁵³ Vgl. <<http://xtree-public.digicult-verbund.de/vocnet/?uriVocItem=http://vocab.archivportal-d.de/&startNode=01099&lang=de&d=n>> (zuletzt abgerufen am 15. Nov. 2022).

⁵⁴ Vgl. hierzu das Beispiel unter Anhang 2.

ersten Release laufend weiterentwickelt. Es ist beabsichtigt, das Tool eines Tages auch für interessierte Nutzerinnen und Nutzer freizuschalten.⁵⁵

Eine Nachnutzung der im Archivportal-D geschaffenen Themenportal-Infrastruktur (Systematikerstellung über xTree, „Schlagworttool“) durch andere Archive, etwa im Rahmen von übergreifenden Erschließungs-, Digitalisierungs- und Beständezusammenführungsprojekten ist ausdrücklich erwünscht.⁵⁶ Was den Arbeitsaufwand angeht, kann das Bundesarchiv also für seinen Teil der inhaltlichen Entwicklung des „Themenportals Wiedergutmachung“ auf Bestehendes zurückgreifen.

3. „Themenportal Wiedergutmachung“ – Erstellung einer Teilsystematik unter besonderer Berücksichtigung von Aufwand und Ertrag

3.1 Rahmenbedingungen und Herausforderungen des Projekts

Den Ausgangspunkt für die Überlegung, Archivgut zur Wiedergutmachung nationalsozialistischen Unrechts im Rahmen eines archivübergreifenden Themenportals zugänglich zu machen, bilden ein politischer Anlass und Auftrag: Ebenso wie die Entschädigungsleistungen der Bundesrepublik Deutschland an Betroffene in den kommenden Jahren auslaufen, fallen auch die entsprechenden Aufgaben des hiermit auf Bundesebene federführend verantwortlichen Bundesministeriums der Finanzen (BMF) weg. Das vom BMF initiierte Projekt „Transformation der Wiedergutmachung“ will sich deshalb gegen ein Vergessen dieser über 70 Jahre laufenden Wiedergutmachung wenden und insbesondere für Angehörige, Forschung und interessierte Öffentlichkeit einen niederschweligen Zugang zu digitalisierten, einschlägigen Wiedergutmachungsakten in deutschen Archiven ermöglichen.⁵⁷

Bis zum Inkrafttreten des Bundesentschädigungsgesetzes (BEG) im Jahr 1953 war die Wiedergutmachung in der Bundesrepublik nicht bundeseinheitlich geregelt und damit Ländersache. Doch auch in der Folgezeit blieb die Durchführung der Bundesgesetze im Bereich der Wiedergutmachung überwiegend Aufgabe der Länder. Daher besteht für die Zugänglichmachung von Unterlagen im „Themenportal Wiedergutmachung“ ein erheblicher Abstimmungsbedarf zwischen Bundesarchiv und den Landesarchiven.⁵⁸ Dennoch existiert eine umfangreiche Überlieferung auf Bundesebene.⁵⁹

Sachakten zum Thema „Wiedergutmachung“ sind im Bundesarchiv überwiegend erschlossen. Hier stellt sich die Frage, ob für einen sachthematischen Zugang und die Verschlagwortung im Themenportal eine „Nach-

⁵⁵ Auskunft von Nils Meyer (LABW), E-Mail vom 14. Febr. 2022. Die Problematik eines „ungesteuerten“ und wenig aussagekräftigen „Social Taggings“ wird man durch die Notwendigkeit der Verschlagwortung mit Hilfe des Tools verhindern können. Allerdings wird das Tool und sein Algorithmus hierfür noch deutlich weiterentwickelt werden müssen (vgl. zur teilweise unzufrieden stellenden Leistung des Tools Kapitel 4.2; eine eher negative Einschätzung der Nützlichkeit von „Social Tagging“ findet sich bei Glauert: *Müssen wir anders verzeichnen?*, S. 36).

⁵⁶ So die Auskunft von Nils Meyer (LABW), E-Mail vom 14. Febr. 2022.

⁵⁷ Siehe hierzu etwa den Bericht über die Kooperationsvereinbarung zwischen BMF und Bundesarchiv (< https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Themen/Oeffentliche_Finanzen/Vermögensrecht_und_Entschädigungen/themenportal-wiedergutmachung.html> ; zuletzt abgerufen am 15. Nov. 2022). Das BMF sieht im Themenportal zugleich einen Beitrag zum Kampf gegen Antisemitismus (vgl. Das Archivierungsprojekt der Wiedergutmachung und seine Bedeutung im Kampf gegen den Antisemitismus (Monatsbericht des BMF, Januar 2021), zuletzt abgerufen am 8. Nov. 2022 unter: < https://www.bundesfinanzministerium.de/Monatsberichte/2021/01/Inhalte/Kapitel-3-Analysen/3-7-archivierungsprojekt-wiedergutmachung-pdf.pdf?__blob=publicationFile&v=4>).

⁵⁸ Die Koordinierung auf Bund-Länder-Ebene setzte bereits vor über zehn Jahren ein. Eine ARK-Bund-Länder-Arbeitsgruppe „Wiedergutmachung“ veröffentlichte im Jahr 2009 einen Abschlussbericht (< https://www.bundesarchiv.de/DE/Content/Downloads/KLA/wiedergutmachung-abschlussbericht.pdf?__blob=publicationFile> ; zuletzt abgerufen am 17. Nov. 2022).

⁵⁹ Vgl. hierzu die Dokumentation der ARK-Bund-Länder-Arbeitsgruppe „Wiedergutmachung“ hinsichtlich der Überlieferungslage in den deutschen staatlichen Archiven (< https://www.bundesarchiv.de/DE/Content/Downloads/KLA/wiedergutmachung-dokumentation.pdf?__blob=publicationFile> ; zuletzt abgerufen am 17. Nov. 2022).

erschließung“ erfolgen müsste. Da hier die Erschließung auf Vorgangsebene jedoch häufig nur die Erarbeitung und Fortschreibung von Gesetzen und Verordnungen und damit in der Regel nicht mehr als formaljuristische Inhalte in den Blick nimmt, erscheint eine „Nacherschließung“ mit „Enthält-Vermerken“, an die eine Verschlagwortung anknüpfen könnte, weder nützlich noch unter den derzeitigen Rahmenbedingungen realistisch.⁶⁰

Einschlägige Einzelfallakten finden sich vor allem im Bestand B 106 (Bundesministerium des Innern). Hier existiert eine umfangreiche Personen-Kartei zu Einzelfällen, die dem „Gesetz zur Regelung der Wiedergutmachung nationalsozialistischen Unrechts für Angehörige des öffentlichen Dienstes“ (BWGöD) zugeschrieben wurden. In diesem Bereich war ausnahmsweise das BMI federführend. Die ursprüngliche Erschließung dieser Akten im Bundesarchiv war besonders flach: So blieben die der Kartei zugehörigen Einzelfallakten weitestgehend unerschlossen, während sie zugleich innerhalb der Klassifikation nur nach den Anfangsbuchstaben der Nachnamen sortiert waren. Noch vor Start des Projekts „Themenportal Wiedergutmachung“ fand zumindest eine Erfassung von Namen, Vornamen, Geburtsdatum und -ort statt.

Zwar wiesen die einzelnen Verzeichnungseinheiten und die ihnen zugeordneten Erschließungsdaten danach disambiguierende, also eindeutige Merkmale auf, die für ein Matching mit GND-Personendaten unter Umständen ausreichend sein dürften. Dennoch kam man im Hinblick auf eine spätere Einbindung im Themenportal zu dem Schluss, dass die Erschließung hierüber deutlich hinaus gehen müsste.⁶¹ Zweifellos ist für die Zuordnung von personenbezogenen Einzelfallakten zu einer sachthemen Systematik mitsamt Verschlagwortung eine tiefergehende Erschließung sinnvoll, insbesondere wenn die Zuordnung nachträglich und ausschließlich auf Grundlage der Erschließungsinformationen und eben nicht über eine erneute Aktenautopsie erfolgen soll. Auf Namens Ebene werden daher aktuell folgende Merkmale ergänzt: Ggf. Band, Name, Vorname, ggf. Akadem. Grad, ggf. Namensvorsatz, Lebensdaten, Orte, Antragsgegenstand, letzte Amtsbezeichnung, letzte Dienststelle, Antragsteller (falls mehrere Namen erfasst) sowie ggf. Mitgliedschaft in der NSDAP. Die Aufnahme der GND-ID im Feld Beschreibung ist zumindest für Personen der Zeitgeschichte vorgesehen. Diskutiert, jedoch unberücksichtigt bleibt dagegen vorerst das Verfahrensergebnis, was im Hinblick auf den sachthemen Zugang und eine Facettensuche mit Hilfe der entsprechenden Schlagwörter bedauerlich erscheint, aber offenbar nicht ohne tiefergehende Aktenautopsie möglich ist.⁶²

Im Gegensatz zum „Themenportal Weimarer Republik“ stellt der Aufbau eines „Themenportals Wiedergutmachung“ eine ungleich größere inhaltliche, aber auch rechtliche Herausforderung dar: Wir haben es in diesem Kontext mit einem hohen Anteil an personenbezogenem Schriftgut zu tun, bei dem die Schutzfristen teilweise noch nicht abgelaufen sein dürften. Hier wird es also im Vorfeld der Zugänglichmachung der Erschließungsinformationen und Digitalisate zur Implementierung einer umfassenden Rechteverwaltung innerhalb des Archivportals-D kommen müssen.⁶³

Was den Aufbau des Themenportals unter inhaltlichem Gesichtspunkt – insbesondere im Hinblick auf die Systematikerstellung – anbelangt, muss konstatiert werden, dass sich der Thesaurus für eine Epoche von jenem

⁶⁰ Telefonische Auskunft von Kerstin Oldenhage (BArch, Referatsleiterin B 1) vom 2. Febr. 2022. Wie mit Sachakten für das „Themenportal Wiedergutmachung“ insgesamt umzugehen ist, muss wohl noch abschließend geklärt werden. Auch in der KLA-Unterarbeitsgruppe „Wiedergutmachung“ wird dieses Thema eventuell noch einmal Thema werden müssen.

⁶¹ Telefonische Auskunft von Martina Scheid (BArch, Referat B 1) vom 1. Febr. 2022 sowie Videokonferenzen mit Mirjam Sprau und Kevin Dubout (BArch, Referat GW 1 und Projekt „Themenportal Wiedergutmachung“) am 14. Jan., 21. Jan. und 25. Febr. 2022.

⁶² Die Erschließung orientiert sich an einem Antragsformular, in dem der Antragsteller oder die Antragstellerin die hierfür aufgenommenen Merkmale eintragen (Auskunft von Martina Scheid, 1. Febr. 2022). Das Verfahrensergebnis müsste sich dagegen aus der Akte selbst ergeben. Im Rahmen der unten vorgestellten Systematik wird das Verfahrensergebnis dennoch aufgegriffen, weil es für den sachthemen Zugang einen Mehrwert bietet und es evtl. noch nachträglich erschlossen wird (vgl. Kapitel 3.2).

⁶³ Wie eine solche aussehen könnte, kann an dieser Stelle nicht weiter vertieft werden.

eines Politikbereichs doch deutlich unterscheiden muss. Wiedergutmachung fand in ganz unterschiedlichen Formen und auf Grundlage von diversen Gesetzen und Verordnungen statt – oder eben nicht statt.⁶⁴ Die Einzelfallakten sachthematisch zu systematisieren und zu indexieren, lässt sich tatsächlich nur auf Grundlage der oben beschriebenen tiefen Erschließung bewerkstelligen. Entsprechend vielschichtig sind die Begrifflichkeiten, die in diesem Kontext entstanden, die im Folgenden bezüglich eines Teilaspekts der „Wiedergutmachung“ in den Blick genommen werden sollen.

3.2 Konzept für eine Teilsystematik und normdatenbasierte Verschlagwortung

Ein Konzept für eine Gesamtsystematik und normdatenbasierte Verschlagwortung für sämtliche im Bundesarchiv für das „Themenportal Wiedergutmachung“ in Frage kommende Bestände würde den Rahmen dieser Transferarbeit sprengen. Es erscheint daher sinnvoll, sich zunächst den Einzelfallakten eines Bestandes zuzuwenden, deren Erschließung im Hinblick auf eine Einbindung in das „Themenportal Wiedergutmachung“ im Bundesarchiv voll im Gange ist. Die BWGöD-Einzelfallakten des Bestands B 106 (BMI) bieten sich aufgrund der oben erwähnten strukturierten, tiefen Erschließung für den Entwurf einer ersten Systematik im Themenfeld „Wiedergutmachung“ an. In den ca. 1300 Akten (Aufbewahrungseinheiten) befinden sich jeweils zwischen 10 und 50 Vorgänge. Insgesamt liegen unter dem Aktenzeichen 23* knapp 4950 Verzeichnungseinheiten, von denen ein großer Teil bereits erschlossen wurde. Diese sollen den folgenden Überlegungen als Datengrundlage dienen.

Orientieren wird sich der Vorschlag für eine sachthematische Teilsystematik mit relevanten und fachhistorisch fundierten Kategorien und normdatenbasierten Schlagwörtern zur „Wiedergutmachung nationalsozialistischen Unrechts“ vornehmlich an den im Rahmen des Projekts „Themenportal Weimarer Republik“ gemachten Erfahrungen. Diese haben ihren Niederschlag in zwei durch das Bundesarchiv und das Landesarchiv Baden-Württemberg erstellten und im Archivportal-D hinterlegten „Guidelines“ zur Systematikerstellung und Verschlagwortung gefunden, die für die künftige Gestaltung von Themenportalen hilfreich sein sollen.⁶⁵ Hiernach lauten die Leitfragen für den Aufbau einer Systematik und damit für den sachthematisch strukturierten Zugang: Welche Themen werden Nutzerinnen und Nutzer konkret interessieren? Wie kann der Einstieg in diese Themen aussehen und welche Schlagwörter müssten auf der untersten Ebene vorhanden sein, um zu diesen Themen möglichst breit und vielseitig hinzuführen und zugleich eine „Facetten-Suche“ zu ermöglichen?

Für das nachfolgende Konzept wurde eine Mischung aus deduktivem und induktivem Ansatz gewählt:⁶⁶ Einerseits ließen sich die möglichen Themen und Inhalte der Einzelfallakten auf Grundlage des Gesetzestextes zum BWGöD⁶⁷ sowie der Inhaltsverzeichnisse von einschlägiger Fachliteratur⁶⁸ eingrenzen. Hier ging es darum,

⁶⁴ Eine gute Übersicht über die unterschiedlichen Rechtsgrundlagen bietet das so genannte „Kalendarium“ des BMF (vgl. Bundesministerium der Finanzen: Kalendarium zur Wiedergutmachung von NS-Unrecht. Gesetzliche und außergesetzliche Regelungen sowie Richtlinien im Bereich der Wiedergutmachung nationalsozialistischen Unrechts, Berlin 2019).

⁶⁵ Siehe „Themenportale gestalten. Sachthematische Zugänge im Archivportal-D“ (künftig: Handbuch: Themenportale gestalten) sowie „Handbuch zur Aufbereitung von Archivgut für das Themenportal zur Weimarer Republik“ (künftig: Handbuch zur Aufbereitung von Archivgut), beides herausgegeben von Bundesarchiv und Landesarchiv Baden-Württemberg und zur Verfügung gestellt unter: < <https://www.archivportal-d.de/content/node/6025> > (zuletzt aufgerufen am 10. Nov. 2022). Auf diese beiden Guidelines wird sich der Verfasser dieser Arbeit auch im Rahmen seines Konzepts für eine Teilsystematik für das „Themenportal Wiedergutmachung“ stützen.

⁶⁶ Vgl. Handbuch: Themenportale gestalten, S. 2–4.

⁶⁷ „Gesetz zur Regelung der Wiedergutmachung nationalsozialistischen Unrechts für Angehörige des öffentlichen Dienstes“ vom 11. Mai 1951, in: BGBl. 1951 Teil I Nr. 21, S. 291.

⁶⁸ Die Forschung zur Wiedergutmachung im öffentlichen Dienst ist bisher überschaubar und eher durch regionale Studien geprägt. Vgl. Otto Gnirs: Die Wiedergutmachung im öffentlichen Dienst, in: Hugo Finke u.a.: Entschädigungsverfahren und sondergesetzliche

den gesetzgeberischen Umgang mit der Wiedergutmachung im öffentlichen Dienst um die „historische Wirklichkeit“ als Ergebnis der Forschung zu diesem Themenkomplex zu ergänzen. Auf diese Weise sollen gewisse „Leerstellen“ der deutschen Wiedergutmachungspolitik, die sich aus den Akten kaum ergeben, in der Systematik abgebildet werden.⁶⁹ So war beispielsweise der jüdische öffentliche Dienst in der ursprünglichen Fassung des Gesetzes nicht erwähnt.⁷⁰

Andererseits wurde basierend auf einer Wortgutsammlung ein Überblick über den tatsächlichen Inhalt der Einzelfallakten gewonnen. Hierfür fand ein Auslesen der in den Erschließungsinformationen eingetragenen Werte unterhalb der Namenszusatzbezeichnungen „Antrags-/Verfahrensgegenstand“, „Amtsbezeichnung/Dienstgrad“, „Letzte Dienststelle“ und „Mitgliedschaft“ statt. In einer Excel-Liste erfolgte anschließend die Bereinigung um Duplikate und leere Felder. Unter „Antrags-/Verfahrensgegenstand“ enthielt die Liste nun 109 unterschiedliche Werte, unter „Amtsbezeichnung/Dienstgrad“ 698, unter „Letzte Dienststelle“ 1137 sowie unter „Mitgliedschaft“ 69.

Anschließend wurde manuell eine Eingrenzung und Zusammenführung der tatsächlich vorhandenen Werte durch eine Bereinigung um fehlerhafte Einträge (Rechtschreibfehler, Leerzeichen oder weitere Angaben hinter/vor dem Wert etc.) sowie Synonyme vorgenommen oder Abkürzungen aufgelöst. Im Folgenden wird ein Überblick über die erstellte Systematik und die ausgewählten Deskriptoren gegeben:⁷¹

Die sechsstufige Systematik gliedert sich unterhalb des Themas „Entschädigung“ in eine Kategorie „Entschädigung nach anderen Regelungen bis 1990“ unter der wiederum die „Beschäftigten im Öffentlichen Dienst“ eingruppiert sind.⁷² Ab hier erfolgt eine mehrgliedrige Aufschlüsselung in die Kategorien „Antragsteller“, „Antrags- und Verfahrensgegenstand“ sowie „Verfahrensergebnis“. Die „Antragsteller“ spalten sich dann noch einmal in Kategorien auf, die sie näher beschreiben („Geschädigte und Versorgungsberechtigte“, „Dienst- und Arbeitsstelle“, „Mitgliedschaft“). Gleiches gilt für den „Antrags- und Verfahrensgegenstand“ („Schädigung und Benachteiligung“, „Gewährung und Neuberechnung einer Leistung“) sowie das „Verfahrensergebnis“ („Gewährung“, „Teil-Gewährung“, „Nicht-Gewährung“, „Gerichtliche Überprüfung“).

Dagegen wurde sich bewusst gegen die Aufnahme der Rechtsgrundlage in die Systematik ausgesprochen, da diese schon über die Klassifikation abgebildet ist und entsprechend angesteuert werden kann. Zudem blieben die im Rahmen der Erschließung unter „Bemerkung“ eingetragenen GND-IDs von 66 Personen der Zeitgeschichte unberücksichtigt. Eine Indexierung und Verknüpfung mit Normdaten wird für die jeweiligen Archivalien über die Personenseiten des Archivportals-D erfolgen.⁷³

Die Sacherschließung durch Verschlagwortung mit „Keywords“ soll „vermittelnd vom Thema zum Archivgut führen“ und „griffige Bezeichnungen für einzelne Phänomene, Ereignisse, Gegenstände, Institutionen etc. [lie-

Entschädigungsregelungen, München 1987, S. 265–303 sowie Christian Römmer: Entschädigung Erster Klasse? Die Wiedergutmachung im öffentlichen Dienst in Hamburg nach dem Zweiten Weltkrieg, München/ Hamburg 2003.

⁶⁹ Vgl. Sprau/Herrmann: Das Themenportal „Wiedergutmachung nationalsozialistischen Unrechts“, S. 277.

⁷⁰ In die Systematik wurde der Begriff „Jüdische Gemeinde“ als Schlagwort unter „Dienst- und Arbeitsstelle“ aufgenommen. Außerdem ließen sich durch Kombination der Schlagwörter „Krankenhaus“, „Schule“ usw. mit dem Schlagwort „Religionsgemeinschaft“ bestimmte konfessionelle, aber auch sonstige jüdische Dienstherrn verschlagworten.

⁷¹ Die folgende Systematik findet sich unter Anhang 1.

⁷² Bis zu diesem Knotenpunkt müssten weitere Themen (z. B. Rückerstattung) und Kategorien (z. B. Entschädigung nach Bundesentschädigungsgesetz, Entschädigung nach Entschädigungsrentengesetz usw.) parallel laufen, deren Aufnahme in die Systematik jedoch an dieser Stelle zu weit führen würde. Ab diesem Knotenpunkt beansprucht die folgende Systematik jedoch eine gewisse Vollständigkeit der Kategorien und Schlagwörter, wie sie sich zumindest nach Auswertung von Literatur und Erschließungsdaten darstellen.

⁷³ Vgl. Handbuch zur Aufbereitung von Archivgut, S. 20 f.

fern], die zu einem bestimmten Thema gehören“.⁷⁴ Die für die hier erstellte „BWGöD-Systematik“ entwickelten Deskriptoren finden sich entsprechend als Begriffe nur in Einzelfällen identisch in den Erschließungsdaten (vor allem unter „Antrags- und Verfahrensgegenstand“) wieder. Auf Grundlage der um Synonyme bereinigten Werte erfolgte nämlich ein Gruppieren dergleichen unter Sammel- und Oberbegriffen. Bei den „Mitgliedschaften“ sind etwa eine Vielzahl an Gliederungen oder angeschlossenen Verbänden der NSDAP feststellbar. Hier wurden häufig auftretende Werte herausgehoben (so „SS“ oder „SA“) sowie andere Körperschaften wiederum unter „Politische Organisation“ oder „Berufsverband“ (NS-Rechtswahrbund, NS-Ärztebund etc.) zusammengefasst, um die Systematik nicht mit Schlagwörtern zu überladen.⁷⁵ Auf diese Weise ließen sich in diesem Bereich sechs Schlagwörter (NSDAP, SS, SA, Politische Organisation, Berufsverband, Reichskulturkammer) ermitteln.

Im Hinblick auf die „Geschädigten und Versorgungsberechtigten“ lässt sich ein Unterschied zwischen den im BWGöD genannten Anspruchs- bzw. Versorgungsberechtigten und den in den Einzelfallakten auftauchenden Personen feststellen. Im BWGöD hatte man es versäumt, den Begriff „öffentlicher Dienst“ eindeutig zu definieren.⁷⁶ Referendare zählten dazu; Personen, die ihr Studium beendet hatten, jedoch nicht mehr in den Vorbereitungsdienst übernommen wurden (z. B. geprüfte Rechtskandidaten), dagegen nicht. Auch tauchten im Laufe der Jahre immer wieder neue Anspruchsgruppen (etwa Professoren und Privatdozenten von Hochschulen) auf, die letztlich zu Änderungen des Gesetzes und einschlägiger Rechtsprechung führten.⁷⁷ Daher ist die Verschlagwortung zwar in dieser Kategorie differenzierter als dies nach dem Gesetzestext anzunehmen wäre. Dennoch wurde auch hier ebenso wie bei den „Dienst- und Arbeitsstellen“ in vielen Fällen ein Zusammenführen für sinnvoll erachtet.⁷⁸

Dies lässt sich am ehesten anhand der Schlagwörter „Polizeibeamter“ und „Hochschullehrer“ bzw. „Polizei“ und „Hochschule“ verdeutlichen. So tauchen in den Erschließungsdaten unterschiedliche Polizei-Dienstgrade (Revieroberleutnant der Schutzpolizei, Polizei-Hauptmann usw.) ebenso auf wie diverse Formen von Professoren, Privatdozenten, Wissenschaftlichen Mitarbeitern etc. Bei den Dienststellen kommen darüber hinaus in der Regel noch Ortsangaben ins Spiel. So ist etwa von der „Deutschen Handelsschule Mähren-Ostrau/Böhmen“ oder der „Deutschen Universität Prag“ bzw. der „Gendarmerie Königsberg/Ostpreußen“ die Rede. Gleich der Vorgehensweise beim „Themenportal Weimarer Republik“ erscheint es in solchen Fällen sinnvoll, dass Geografika auch im „Themenportal Wiedergutmachung“ eine eigene, parallellaufende Systematik erhalten, wobei die Zuordnung eines normdatenreferenzierten Ort-Schlagworts zur jeweiligen Verzeichnungseinheit ebenso verpflichtend sein sollte. Daher war in diesen Fällen ein Zusammenführen zum Oberbegriff „Hochschule“ bzw. „Polizei“ die logische Konsequenz.

Die Gliederungspunkte der Kategorie „Antrags- und Verfahrensgegenstand“ orientieren sich dagegen stärker am Gesetzestext und der Literatur. Die Deskriptoren führen hier häufig verschiedene Synonyme auf den im Gesetz stehenden Schadenstatbestand bzw. die jeweilige Leistung zurück, während gleichzeitig auf ein ent-

⁷⁴ Handbuch: Themenportale gestalten, S. 4.

⁷⁵ Vgl. hierzu Handbuch zur Aufbereitung von Archivgut, S. 23. Hiernach sind „Sachbegriffe“ (saz) dem Entitäten-Typ „Körperschaft“ (kiz) in zu kleinteiligen Fällen vorzuziehen. Die Verschlagwortung der Einzelfallakte eines Arztes, der Mitglied im NS-Ärztebund war, müsste also bestenfalls durch eine Kombination der Deskriptoren „Arzt“, „Berufsverband“ und „NSDAP“ erfolgen.

⁷⁶ Vgl. Heike Anke Berger: Deutsche Historikerinnen 1920–1970. Geschichte zwischen Wissenschaft und Politik, Frankfurt/New York 2007, S. 296.

⁷⁷ Vgl. etwa die Entscheidung des Bundesverfassungsgerichts vom 12. Januar 1965, 2 BvR 454, 470/62, in: BVerfGE 18, 288. Diese bezieht sich auf das Gesetz zur Regelung der Wiedergutmachung nationalsozialistischen Unrechts für Angehörige des öffentlichen Dienstes in der Fassung des Gesetzes vom 18. August 1961 (BGBl. 1961 Teil I Nr. 68, S. 1349).

⁷⁸ Um den Rahmen der Arbeit nicht zu sprengen, sollen an dieser Stelle nur ein paar Beispiele der Deskriptor-Erstellung erwähnt werden.

sprechendes GND-Äquivalent zu achten war. Dies gilt etwa für die „Hinterbliebenenversorgung“ statt „Hinterbliebenenbezüge“ oder die „Wiedereinstellung“ statt „Wiederanstellung“.

Für sämtliche Kategorien bleibt anzumerken, dass alternative Bezeichnungen (bei „Beamter“ etwa „Beamte“, „Beamtentum“, „Beamtin“, „Berufsbeamtentum“, „Berufsbeamter“ oder „Staatsbeamter“) im xTree-Thesaurus nachzuhalten und gegebenenfalls hierfür ebenfalls ein Mapping zur GND („Beamtin“, GND-ID: 4005085-3) oder Wikidata vorzunehmen wäre. Auch hier müsste sich am Vorgehen bei der Erstellung des Thesaurus zum „Themenportal Weimarer Republik“ orientiert werden.⁷⁹

Insgesamt lässt sich für das hier beschriebene Systematikkonzept für die BWGöD-Einzelfallakten festhalten, dass es auf Grundlage von Gesetzestext, Literatur und vorhandenen Erschließungsdaten gelungen ist, einen Vorschlag zu erarbeiten, der als Basis für einen sachthematischen Zugang zu diesem Komplex der Wiedergutmachung dienlich sein kann. Alle ausgewählten Deskriptoren decken die aus den genannten Quellen ersichtlichen inhaltlichen Schwerpunkte ab und sind zudem nahezu vollständig auf Normdateien, insbesondere die GND, referenzierbar.⁸⁰

Ob die hier entworfene Systematik für diesen Teilaspekt der Wiedergutmachung wirklich tragfähig ist oder nicht doch noch im manchen Fällen nachjustiert werden muss, wird die in den kommenden Monaten anstehende Zuordnung der Verzeichnungseinheiten zu den Schlagwörtern über das „Schlagworttool“ zeigen.

3.3 Einschätzung von Aufwand und Ertrag

Wie sieht es nun mit dem Verhältnis von Aufwand und Ertrag aus? Bot das „Themenportal Weimarer Republik“ bereits einen niedrighschwelligen Einstieg in die Suche nach Unterlagen einer Epoche, so wird ein solcher Zugang nun für Unterlagen zum Thema „Wiedergutmachung“ geschaffen. Schließlich geht es bei einer intuitiven und explorativen Suche weniger um zielgerichtetes Recherchieren als vielmehr um ein unklares Suchziel. Für das eigene Thema sollen zusätzliche Quellen ausfindig gemacht werden, die durch eine Volltextsuche womöglich übersehen worden wären. Doch erst die Verknüpfung von Archivgut ganz unterschiedlicher Provenienz mit normierten Schlagwörtern, also – neben der thematischen Verortung durch die Einbindung in eine Systematik – die semantische Absicherung durch Referenz auf eine Normdatei, garantiert eine Maschinenlesbarkeit und bietet damit innerhalb des Themenportals die Möglichkeit einer Suche über Filter und Facetten sowie darüber hinaus die perspektivische Einbindung der Erschließungsinformationen und Digitalisate ins Semantic Web.

Rein praktisch schafft eine Verschlagwortung der BWGöD-Einzelfallakten im Rahmen des „Themenportals Wiedergutmachung“ – entsprechend des oben skizzierten Vorschlags – also die Voraussetzung für eine andersartige Darstellungsform und Zugänglichmachung. Archivgut lässt sich hierdurch auf unterschiedliche Art und Weise entdecken. Im Rahmen einer Facettensuche kann etwa je nach Wunsch und Forschungsvorhaben eine thematische Eingrenzung der Treffer erfolgen. Interessiert man sich für Hinterbliebene, die aufgrund der NSDAP-Mitgliedschaft des verstorbenen Angehörigen, z. B. eines Polizeibeamten, auf Probleme bei der Hinterbliebenenversorgung stießen, so könnte eine Filterung über die Schlagwörter „Hinterbliebenenversorgung“, „Polizeibeamter“ und „NSDAP“ erfolgen.

Allerdings ist dieser unzweifelhafte Mehrwert für die Nutzerinnen und Nutzer mit einem nicht unerheblichen Mehraufwand für die an der inhaltlichen Entwicklung des „Themenportals Wiedergutmachung“ beteiligten Ar-

⁷⁹ Vgl. daher Anlage 2: Mapping im xTree-Thesaurus zur Weimarer Republik.

⁸⁰ Die Deskriptoren „Wartestandsbeamter“, „Gewährung“, „Teilgewährung“ und „Nichtgewährung“ sollten entsprechend dem Vorgehen beim „Themenportal Weimarer Republik“ gesammelt und gegebenenfalls später gebündelt in Absprache mit der GND dort als Datensatz eingespeist werden.

chive verbunden.⁸¹ Denn eine tiefe, sich über Monate erstreckende Erschließung, wie im Fall der BWGÖD-Einzelfallakten, oder eine nachträgliche Datenbereinigung ohne zusätzliche technische Hilfsmittel kann nicht regelmäßig das Ziel sein. Auch die weitgehend manuelle Systematikerstellung und das Finden passender Deskriptoren, die zusätzlich in der GND oder einer anderen Normdatei vorhanden sein müssen, kann für einzelne, wenige Themen und Epochen ein gangbarer Weg sein. Für die Ausbreitung des Ansatzes der sachthematischen Zugänge auf das gesamte Bundesarchiv und seine Unterlagen müssten perspektivisch – neben der Weiterentwicklung der Archivinformationssysteme – weitere Softwaretools für automatisierte Verfahren eingesetzt werden.

Denn das Ziel der Systematikerstellung liegt letztlich darin, einen Mittelweg zwischen notwendiger und überflüssiger Komplexität zu finden. Es geht um eine Bündelung der Inhalte, die allerdings auf eine entsprechende Erschließungsqualität angewiesen ist. Die Unterstützung dieses Prozesses durch entsprechende Tools müsste folglich bei der Datenbereinigung einsetzen. Einer solchen Datenaufbereitung sollte sich im besten Fall innerhalb der gleichen Software eine Anreicherung der Erschließungsinformationen mit Normdaten bzw. mit genormten Deskriptoren anschließen.

Das „Schlagworttool“ bringt hier zwar einen technischen Fortschritt. Vermutlich wird dennoch in einigen Fällen eine erneute Aktenautopsie unumgänglich sein. Denn auch der KI-unterstützte Algorithmus des Tools stellt lediglich auf den ersten Blick eine Arbeitserleichterung dar. Die automatische, kontextuelle Verschlagwortung stößt noch zu schnell an ihre Grenzen, was teilweise auch an der unzureichenden Erschließungsdatenqualität liegt.⁸² Die Lösung des Problems eines zu großen Ressourceneinsatzes im Falle einer flächendeckenden Ausbreitung des beschriebenen Erschließungsansatzes wird daher zukünftig nur in einer weitgehend automatisierten Datenbereinigung und Normdatenanreicherung liegen können. Die diesbezüglichen Möglichkeiten und Perspektiven sollen daher abschließend genauer betrachtet werden.

4. Automatisierte Datenbereinigung, Verschlagwortung und Normdatenanreicherung – Möglichkeiten und Perspektiven

Durch das Konzept zur Erstellung einer Teilsystematik und einer Verschlagwortung für das „Themenportal Wiedergutmachung“ wurde ein Weg aufgezeigt, wie eine möglichst ressourcenschonende, thematische Sacherschließung sowie Anreicherung mit Normdaten zumindest für ausgewählte Archivbestände und Themen strukturiert und unterstützt durch wenige technische Hilfsmittel unter den gegenwärtigen Bedingungen erfolgen kann. Nach diesem praktischen Lösungsvorschlag soll nun im Rahmen eines eher theoretischen Ausblicks nach einer perspektivischen Lösung für eine womöglich flächendeckende Ausbreitung des beschriebenen Ansatzes auf das gesamte Bundesarchiv gesucht werden.

Bereiteten die gezielt für das Themenportal erarbeiteten unterschiedlichen Erschließungsinformationen aufgrund ihrer Heterogenität bereits teilweise Probleme, so darf man im Hinblick auf Bestände, die in den vergangenen Jahrzehnten im Bundesarchiv unterschiedlich tief erschlossen wurden, von einer ungleich „mäßigeren“ Erschließungsdatenqualität ausgehen. Bei besonders flach erschlossenem Archivgut müsste zunächst dafür gesorgt werden, dass zumindest die Stammdaten über aussagekräftige Beschreibungen bzw. Einleitungen ver-

⁸¹ Im Bundesarchiv band etwa die Systematikerstellung zum „Themenportal Weimarer Republik“ eine Vollzeitkraft mindestens für ein halbes Jahr. In der Projektgruppe zum „Themenportal Wiedergutmachung“ arbeiten aktuell sechs Vollzeitkräfte, von denen sich zwei ArchivarInnen um die archivischen Inhalte kümmern, die wiederum beraten werden von einem Historiker als wissenschaftlichen Mitarbeiter (Auskunft von Mirjam Sprau und Kevin Dubout (BArch, Referat GW 1 und Projekt „Themenportal Wiedergutmachung“) vom 25. Febr. 2022).

⁸² Auf die Grenzen des Schlagworttools soll im Kapitel 4.2 noch einmal genauer eingegangen werden, da dort auch die möglichen Perspektiven zur Sprache kommen sollen.

fügen, die gewissermaßen thematische Anknüpfungspunkte für eine (teil-)automatisierte Verschlagwortung bieten könnten.

Insgesamt muss das Ziel die algorithmische Indexierung, also die weitgehend automatisierte, flächendeckende Verschlagwortung sein. Entscheidend für eine solche Lösung wird aber weniger die Erschließungstiefe sein, sondern vielmehr die Datenqualität. Dieser müsste sich perspektivisch eine KI-gestützte, semantische Verschlagwortung, die über die derzeitigen Möglichkeiten des „Schlagworttools“ hinausgeht, sowie eine automatisierte Normdatenanreicherung anschließen.

4.1 Datenbereinigung und -anreicherung am Beispiel des Tools „OpenRefine“

Im Bundesarchiv wird derzeit im Rahmen des Projekts „Themenportal Wiedergutmachung“ mit „OpenRefine“ ein Open-Source-Tool getestet, mit dem auf unterschiedliche Art und Weise die Zusammenstellung des „Wiedergutmachungs-Vokabulars“ erleichtert werden könnte. Denn mit dem Programm soll es möglich sein, riesige Datenmengen zu bearbeiten.⁸³ Insbesondere eine Datenbereinigung, aber auch eine Normdatenanreicherung soll sich hiermit teilweise automatisiert vornehmen lassen. An dieser Stelle wird eine knappe Einführung in die Funktionsweise des Tools gegeben sowie seine potentielle Nützlichkeit für das Bundesarchiv und für seine Überlegungen eingeschätzt, die Sacherschließung mit Normdaten nach Möglichkeit auf das gesamte Haus auszubreiten.

Importieren lassen sich in die Webanwendung neben Excel-Dateien weitere übliche Dateiformate (CSV, JSON, XML). Die Datenbereinigung basiert bei „OpenRefine“ vor allem auf zwei Säulen: Der Korrektur von fehlerhafter sowie der Vereinheitlichung von heterogener Erschließung. Typische Fehlerquellen bei der Erschließung sind etwa Leerzeichen hinter einem Begriff, die vom Computer lediglich als „gewollter“ Zahlencode interpretiert wird. Eine Suchmaschine wird hierdurch aber überfordert sein. Bei einer Recherche nach dem entsprechenden Begriff wird dieser mit abschließendem Leerzeichen regelmäßig unberücksichtigt bleiben, wenn nicht mit Platzhaltern (etwa einem *) gesucht wird. „OpenRefine“ ermöglicht eine Disambiguierung, also eine Begriffsklärung, durch das semi-automatisierte Erkennen von (Rechtschreib-)Fehlern und Synonymen sowie das Auflösen von Homonymen. So gelingt unter anderem das Entfernen von führenden und nachstehenden Leerzeichen oder auch die Zusammenfassung von aufeinanderfolgenden Leerzeichen.

Durch eine „Cluster and Edit“-Funktion macht das Programm automatisch Vorschläge für eine Vereinheitlichung der Schreibweisen von bestimmten Einträgen, so etwa bei Wörtern, die um einen Buchstaben abweichen.⁸⁴ Dient diese Funktion einerseits noch der Datenbereinigung, kann es andererseits auch für die Erstellung einer Systematik und Verschlagwortung nützlich sein, da sich hierüber in begrenztem Umfang eine Gruppierung von ähnlichen Begriffen (z. B. Bürgermeister, Amtsbürgermeister, Oberbürgermeister) vornehmen lässt. Hier orientiert sich der Vorschlag des Tools an dem am häufigsten auftretenden Wert. Dabei muss bisweilen eine manuelle Ergänzung oder Abänderung des Vorschlags (Oberschullehrer -> Lehrer) vorgenommen oder dieser ganz abgelehnt werden. Die Auswahl lässt sich anschließend zusammenführen und neu gruppieren bzw. schließen.

⁸³ Eine ausführliche Darstellung der Möglichkeiten und Grenzen von OpenRefine findet sich in einer aktuellen informationswissenschaftlichen Abschlussarbeit (vgl. Christiane Klaes: Linked Open Data-Strategien zum Identity Management in einer Fachontologie. Prototypische Entwicklung eines Workflows zur Aufbereitung und zum Interlinking von Personennamen, Abschlussarbeit zur Erlangung des akademischen Grades Master of Arts (M.A.) im Studiengang Internationales Informationsmanagement – Informationswissenschaft, Stiftung Universität Hildesheim 2021, abgerufen am 22. Nov. 2022 unter: < https://www.researchgate.net/publication/354325553_Linked_Open_Data-Strategien_zum_Identity_Management_in_einer_Fachontologie_-_Prototypische_Entwicklung_eines_Workflows_zur_Aufbereitung_und_zum_Interlinking_von_Personennamen#pf39 >).

⁸⁴ Vgl. etwa die fehlerhafte Schreibweise des Begriffs „Hinterbliebenenversorgung“ (siehe hierzu Anhang 3: Cluster-Beispiel aus OpenRefine). Hierauf beziehen sich auch die anderen Beispiele dieses Absatzes.

Die Datenanreicherung erfolgt über die Funktion „Reconciliation“, durch die eine semi-automatisierte Verknüpfung mit einer Vielzahl von Referenzdatensets der Linked Open Data (LOD) Cloud möglich ist (z.B. GND, Wikidata, VIAF). Es findet zunächst ein „Matching“-Vorgang statt, dem sich eine Verknüpfung mit einer Normdatei anschließt. Der Abgleich gegen die GND erfolgt über den durch das Hochschulbibliothekszentrum des Landes NRW betriebenen LOD-Dienst lobid-GND, der über eine Programmierschnittstelle (API) Zugriff auf die GND hat. Konkret muss in OpenRefine über einen Link auf die Service-API des Dienstes „GND reconciliation for OpenRefine“ zugegriffen werden. Hier lässt sich auch festlegen, mit welchen Typen von Entitäten ein Abgleich stattfinden soll. Wo das „Matching“ nicht erfolgreich war, ist eine manuelle Nachbesserung erforderlich. Anschließend kann eine Anreicherung mit den im jeweiligen GND-Datensatz angelegten Werten vorgenommen werden. Die Festlegung, welcher Wert übernommen wird, muss einmalig für die jeweilige Spalte erfolgen, für die der Datenabgleich durchgeführt wurde. Die angereicherte Tabelle kann anschließend in das Archivfachinformationssystem zurückimportiert werden.

Die Erarbeitung der oben vorgestellten „BWGÖD-Teilsystematik“ erfolgte überwiegend manuell und war entsprechend zeitaufwendig. Ein Nachstellen der einzelnen Schritte mit „OpenRefine“ führt zu dem Ergebnis, dass eine solche Tool-Unterstützung durchaus eine wesentliche Aufwandsreduzierung mit sich bringt. Insbesondere die Datenbereinigung und das „Clustering“ sorgen für einen deutlichen Zeitgewinn. Natürlich befreit „OpenRefine“ – zumindest in seiner aktuellen Version⁸⁵ – nicht von einer umfangreichen Qualitätskontrolle. Umso größer jedoch die Datenmengen, desto weniger dürfte das manuelle Nachbessern gegenüber einem gänzlich manuellen Verfahren ins Gewicht fallen.

Zwar lässt sich in der semi-automatisierten Datenbereinigung und Datenanreicherung mit „OpenRefine“ ein Schritt in die richtige Richtung sehen, durch den man dem Ziel einer effizienteren Verschlagwortung und Normdatenanreicherung näherkommen kann. Ein vollautomatisiertes Verfahren wird derzeit nur durch eine hierüber hinausgehende KI-gestützte, semantische Verschlagwortung zu erreichen sein.

4.2 KI-gestützte, semantische Verschlagwortung

Kann Künstliche Intelligenz (KI) überhaupt mit menschlicher Intelligenz konkurrieren? Die Beantwortung dieser Frage kann und soll nicht Gegenstand dieser Arbeit sein.⁸⁶ Es lässt sich derzeit skizzieren, was heute schon mit KI im Bereich der Erschließung, Verschlagwortung und Anreicherung mit Normdaten für Archive möglich erscheint und bereits in Teilen der europäischen Archivwelt erprobt wird.

Seit einigen Jahren nehmen erste „Vorläufer“ von KI vor allem im Kontext der Digitalisierung von Archivgut auch Einzug ins Archivwesen. So sind insbesondere OCR- (Optical Character Recognition) und HTR-Softwarelösungen (Handwritten Text Recognition) auf dem Vormarsch. Mit dem europäischen Kooperationsprojekt „Transkribus“ ist etwa die Digitalisierung und Texterkennung von maschinen- und handschriftlichen Unterlagen möglich.⁸⁷ Auch ein Teilprojekt des Bundesarchivs zum Komplex „Wiedergutmachung nationalsozialistischen Unrechts“ wird sich in diesem Schrifterkennungs-Rahmen bewegen. Zwar soll es bei der automatischen Identifizierung von nicht gekennzeichneten Lastenausgleichs-Einzelfallakten, die sich auf Vermögen beziehen,

⁸⁵ OpenRefine 3.5 wurde am 26. Jan. 2022 veröffentlicht. Die an dieser Stelle beschriebenen Funktionen und Möglichkeiten beziehen sich auf diese Version.

⁸⁶ Einen niederschweligen Einstieg in das Thema Künstliche Intelligenz (KI) bietet ein Online-Glossar des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS (< https://www.iais.fraunhofer.de/de/forschung/kuenstliche-intelligenz.html#faq_faqitem_88176962-answer >; zuletzt abgerufen am 3. Nov. 2022). Wenn nicht anders vermerkt, greift die folgende Darstellung bei der Erklärung der einzelnen Begriffe aus dem KI-Umfeld auf dieses Glossar zurück.

⁸⁷ Vgl. zum EU-Projekt READ und Transkribus u.a. Günter Mühlberger/Tamara Terbul: Handschriftenerkennung für historische Schriften. Die Transkribus Plattform, in: b.i.t. online 21 (2018) Nr. 3, S. 218–222.

das unter Wert von jüdischen Voreigentümern erworben wurde, zum Einsatz von KI kommen. Womöglich lassen sich die einschlägigen Akten aber bereits durch „einfache“ OCR-Verfahren ausfindig machen, so dass nicht anspruchsvollere KI-Ansätze erprobt werden müssen.⁸⁸

Der Algorithmus des oben vorgestellten „Schlagworttools“ innerhalb des Archivportals-D, das die Indexierung von Archivgut im Rahmen der Themenportale vereinfachen soll, ist dagegen ein gutes Beispiel für die Möglichkeiten, aber auch Grenzen von ersten „wirklichen“ KI-Lösungen, wie sie derzeit im Archivwesen getestet werden.⁸⁹ Auffällig ist die Tatsache, dass der Algorithmus nur für einen Teil der im Themenportal befindlichen Daten eine passende Verschlagwortung anbietet. Nach Auskunft des Landesarchivs Baden-Württemberg (LABW) liegt dies daran, dass der Algorithmus sehr rechenintensiv ist und daher zunächst nur auf die Daten des LABW und des Bundesarchivs angewendet wurde. Mit dem technischen Betreiber des Algorithmus sei daher ein „bedarfsorientiertes Verfahren“ vereinbart worden.⁹⁰

Auf technischer Seite würde ein Dauereinsatz des Algorithmus das „Schlagworttool“ also derzeit offenbar zu sehr „belasten“. Hierüber ließe sich wohl zeitweilig hinwegsehen, wenn es zumindest für die zur Verfügung stehenden Unterlagen durch seine Schlagwortvorschläge eine qualitativ hochwertige Lösung bieten würde. Allerdings ergibt sich bei genauerem Hinsehen regelmäßig die Ungenauigkeit, häufig auch die Unbrauchbarkeit der vorgeschlagenen Deskriptoren. Probleme sind insbesondere bei Schlagwörtern zu erwarten, bei denen der Inhalt eigentlich interpretiert werden müsste bzw. nur im Kontext zu verstehen ist (z.B. Zuordnung einer Verzeichnungseinheit mit dem Antrags-/Verfahrensgegenstand „Gewährung von Versorgungsleistungen“ zum Schlagwort „Ruhegehalt“ oder „Hinterbliebenenversorgung“). Der Algorithmus des „Schlagworttools“ stößt hier deshalb an seine Grenzen, da er lediglich eine Ähnlichkeitssuche der Erschließungsdaten mit den Schlagwörtern des xTree-Thesaurus einschließlich der laut Wikidata und GND „verwandten Begriffe“ durchführt. Aus diesem Grund reduziert auch das „Schlagworttool“ derzeit den Arbeitsaufwand nur bedingt, da ein enormer Qualitätssicherungsbedarf bestehen bleibt.

Entsprechende Herausforderungen könnten die Algorithmen perspektivisch offensichtlich nur bewältigen, wenn die von ihnen genutzten KI-Modelle auf natürliche Sprachverarbeitung (Natural Language Processing – NLP) und – als Teil davon – beispielsweise auf Eigennamenerkennung (Named Entity Recognition – NER), Wortembedding (Word Embedding), der Schlüsselworterkennung (Keyword Spotting – KWS) bzw. der Schlüsselwortextraktion (Keyword Extraction) aufbauen würden.

Die natürliche Sprachverarbeitung (NLP) zielt insgesamt auf die Verarbeitung natürlich-sprachiger Textdaten ab. Computer, die auf strukturierte Daten angewiesen sind, sollen also die Kommunikation des Menschen, die häufig kaum strukturiert ist und Raum für Interpretationen lässt, verstehen lernen.⁹¹ Dies gelingt unter anderem durch die semantische Analyse mittels Eigennamenerkennung (NER). Sie bewirkt die Zuordnung von bestimmten Entitäten (Personen, Orten, Körperschaften) zu entsprechenden Begriffen etwa eines Satzes sowie ihre

⁸⁸ Vgl. Auskunft von Karsten Kühnel (BArch, Abt. B – Außenstelle Bayreuth, Leiter des Lastenausgleichsarchivs), E-Mail vom 21. Febr. 2022.

⁸⁹ Das „Schlagworttool“ und seine Funktionsweisen wurden für diesen Abschnitt der Transferarbeit getestet. Neben den hierbei gemachten eigenen Erfahrungen werden in den folgenden Ausführungen auch die Erkenntnisse der Projektverantwortlichen, insbesondere des LABWs, zur Sprache kommen.

⁹⁰ Auskunft von Nils Meyer (LABW), E-Mail vom 22. Febr. 2022. Archive, die ihre Unterlagen mit Hilfe des Algorithmus verschlagworten wollen, müssten also die entsprechenden Bestände zunächst mitteilen, so dass der Betreiber den Algorithmus über die entsprechenden Daten laufen lassen könnte.

⁹¹ Zu NLP siehe die Darstellung des Fraunhofer-Instituts für Naturwissenschaftlich-Technische Trendanalysen INT (Sonja Grigoleit: Natural Language Processing, in: Europäische Sicherheit und Technik (April 2019), S. 69; abgerufen am 3. Nov. 2022 unter: < <https://www.int.fraunhofer.de/content/dam/int/de/documents/EST/EST-0419-Natural-Language-Processing.pdf> >).

Extraktion. Mit NER lassen sich daher nicht nur Wörter miteinander vergleichen, sondern auch Texte zum Teil semantisch interpretieren, so dass zumindest eine Grunderfassung der Inhalte möglich ist.

In der deutschen Archivpraxis wird dem Ansatz der Eigennamenerkennung (NER) aktuell großes Potential zugeschrieben. So hat das Hessische Landesarchiv erst kürzlich gemeinsam mit dem Hessischen Landesamt für geschichtliche Landeskunde unter Verwendung von NER eine möglichst effiziente, automatisierte nachträgliche Anreicherung von Erschließungsinformationen mit Normdaten (insbesondere mit Ortsnormdaten) erprobt.⁹² Und auch das Landesarchiv Baden-Württemberg plant in den nächsten Jahren den Einsatz der Eigennamenerkennung zur Prüfung und Normalisierung von Schlagwörtern, um beispielsweise die Personenindexierung der vielen Partnerarchive im Archivportal-D/DDB zu vereinheitlichen und ggf. eine Anreicherung mit GND-Datensätzen zu ermöglichen.⁹³

Neben NER dürfte aktuell auch das Keyword Spotting Potential für die Erschließung mit normdatenbasierten Schlagwörtern bieten. Hier geht es nicht nur um das Aufspüren von bekannten Schlüsselwörtern in Dokumenten,⁹⁴ sondern auch um das Entdecken von verwandten Begriffen wie sie etwa in Normdateien hinterlegt sind. Der Ansatz wird beispielsweise im Finnischen Nationalarchiv genutzt.⁹⁵ Demgegenüber dreht Keyword Extraction den Prozess um und identifiziert automatisch Begriffe, die das Thema eines Dokuments am besten beschreiben.⁹⁶ Während Keyword Spotting somit zur Verschlagwortung bei vorhandenem Thesaurus genutzt werden könnte, wäre Keyword Extraction eine nützliche KI-Unterstützung im Hinblick auf die Schlagworterkennung und zum Aufbau eines kontrollierten Vokabulars.

In welche Richtung sich der Einsatz von KI im Archivbereich weiterentwickeln könnte und im Sinne des Semantic Web weiterentwickeln müsste, zeigt sich derzeit auf europäischer Ebene. Für das Archivportal Europa wird ebenfalls an dem Thema automatische Verschlagwortung bzw. automatische Themenerkennung gearbeitet, um eine sprachenunabhängige Suche nach bestimmten Themen bzw. eine „Ähnlichkeitssuche“ zu ermöglichen. Ein wichtiges Projekt stellt in diesem Zusammenhang die Entwicklung einer automatischen Themenerkennungs-Software („automated topic detection tool“) dar. Im Archivportal Europa wäre ein solches Tool besonders nützlich, da es semantische Verknüpfungen zwischen Dokumenten und Erschließungsinformationen in

⁹² Vgl. Stefan Aumann/Peter Sandner: Normdaten im Hessischen Landesarchiv. Erprobung von Verfahren zur Anreicherung der Erschließungsdaten in Arcinsys, in: Archivnachrichten aus Hessen Nr. 21/2 (2021), S. 104-107. Dabei nutzten die Projektpartner den zur linguistischen Analyse historischer Texte im Rahmen des Projekts „Deutsches Textarchiv“ entwickelten KI-gestützten Webservice „Cascaded Analysis Broker (CAB)“ (vgl. < <https://www.deutsches-textarchiv.de/doku/software#cab> > ; zuletzt abgerufen am 15. Nov. 2022). Als kontrolliertem Vokabular bedienten sie sich hierfür einem Webservice des „Historischen Ortslexikons“, einem Modul des Landesgeschichtliche Informationssystem Hessen (LAGIS). Da einige Ortsnamen in Hessen mehrfach auftauchen, mussten zwecks Disambiguierung weitere Webservices zusätzliche Kontextinformationen beisteuern, um letztlich die entsprechenden Normdaten (GND, Geonames) den wahrscheinlichsten Orten zuzuordnen (vgl. Aumann/Sandner: Normdaten im Hessischen Landesarchiv, S. 106).

⁹³ So die Auskunft von Nils Meyer (LABW), E-Mail vom 14.02.2022). Eine vollautomatische, kontextuelle Verschlagwortung und Normdatenanreicherung mit GND-Datensätzen sieht offenbar auch der durch das Freiburger Archiv Soziale Bewegungen mitentwickelte „Digitale Archivar“ vor. Dieser soll insbesondere für Kommunalarchive als Webservice die Arbeit einer Indexierung und Klassifizierung von Archivgut erleichtern (vgl. < <https://staccared.com/index.html> >). Leider ließen sich für diese Transferarbeit keine zusätzlichen Informationen zu diesem vielversprechend klingenden Projekt gewinnen.

⁹⁴ Eine solche Keyword-Spotting-Funktion bietet etwa Transkribus (vgl. < <https://readcoop.eu/de/glossary/keyword-spotting-kws/> > ; zuletzt abgerufen am 3. Nov. 2022).

⁹⁵ Die Funktionsweise der Schlüsselworterkennung wurde erst kürzlich von Günter Mühlberger auf dem 25. Archivwissenschaftliches Kolloquium der Archivschule Marburg vorgestellt. Der noch nicht veröffentlichte Vortrag trug den Titel „Automatisierte Handschriftenerkennung und Keyword Spotting am Beispiel des Finnischen Nationalarchivs“.

⁹⁶ Feifan Liu u. a.: Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts, in: proc. of the annual conference of the North American chapter of the Association for Computational Linguistics, Boulder 2009, S. 620–628, hier S. 621.

unterschiedlichen Sprachen herzustellen wüsste, die meist nicht auf kontrolliertem Vokabular basieren. Sein Algorithmus wird sich unter anderem der kontextabhängigen Worteinbettung (Contextual Word Embedding) bedienen. Trainiert wurde er insbesondere mit Texten aus Wikipedia sowie unterschiedlichen Sprachen. Mit dem Tool soll es möglich sein, die Bedeutung der Inhalte von archivischen Unterlagen zu erfassen, in dem (vereinfacht gesagt) ein bestimmter, mathematischer Wert (ein Vektor) ganz bestimmten Kontexten zugewiesen wird, und diese sich dadurch maschinell Themen zuordnen lassen. Gleichzeitig ist über das Tool das Identifizieren von Entitäten möglich. Dabei geht der Algorithmus über die erwähnte Eigennamenerkennung (NER) insofern hinaus, da er sogleich ein Mapping mit gleichen Werten in Wikidata und VIAF vornimmt. Auf diese Weise werden nicht nur Personen und Orte als Entitäten erfasst, sondern im Grunde all das, was auch eine Wikipedia-Seite besitzt.⁹⁷

Welchen dieser Ansätze könnte das Bundesarchiv nun intensiver erproben? Die hier genannten Arten der KI-gestützten Textanalyse erscheinen allesamt geeignet, eine eventuell flächendeckende Sacherschließung des Archivguts mit Schlagwörtern sowie die Anreicherung der Erschließungsdaten mit Normdaten nachhaltig zu vereinfachen und diese Verfahren womöglich vollständig zu automatisieren. Nun existieren bereits so genannte Textanalyse-Frameworks, die diese unterschiedlichen KI-Ansätze teilweise zusammenbringen und durch „Basisalgorithmen“ die beschriebenen Interpretationsaufgaben leisten können, so dass sich die Frage stellt, ob es nicht für das Bundesarchiv ausreicht, sich dieser Frameworks zu bedienen?

Ein Beispiel für ein deutschsprachiges Framework wäre etwa „Flair“, das als gemeinschaftliche Forschungs- und Entwicklungsarbeit der Humboldt-Universität Berlin (Institut für Informatik) sowie Zalando Research verschiedene Arten der Textanalyse beherrscht (NER, Word Embedding etc.) und das auch für das Bundesarchiv durchaus attraktiv wäre.⁹⁸ Anzumerken wäre allerdings, dass Frameworks wie „Flair“ den KI-Einstieg zwar erleichtern, ihre „Basisalgorithmen“ aber wahrscheinlich Probleme mit den archivspezifischen Textformen (bestehend aus „Behördendeutsch“, historischen Wörtern und Sprachformen) hätten, so dass sie hierauf erst trainiert werden müssten.

Diesem Vortrainieren von Sprachmodellen kommt ohnehin eine besondere Bedeutung zu. Ob dann auf historische Sprachformen trainierte Algorithmen oder die KI-unterstützte „Interpretation“ von Inhalten besser funktioniert, muss letztlich die Praxis zeigen.⁹⁹ In jedem Fall erfüllt das Bundesarchiv durch seine riesige Menge an Erschließungsdaten zumindest die für Deep-Learning-Algorithmen, also für das Trainieren neuronaler Netze, notwendigen Voraussetzungen.

5. Fazit und Ausblick

Die Ausgangsfrage dieser Transferarbeit lautete, ob es – unter Berücksichtigung von Aufwand und Ertrag – gegenwärtig und zukünftig im Bundesarchiv möglich und sinnvoll erscheint, Archivgut im Rahmen der Erschließung durch Schlagwörter und Normdaten anzureichern. Der zur Beantwortung dieser Frage entwickelte Lösungsvorschlag ist zweigeteilt:

1. Die derzeit im Bundesarchiv genutzten Archivfachinformationssysteme erfüllen nur bedingt die Voraussetzung für eine Verschlagwortung und Normdatenanreicherung. Daher ist momentan für eine Erprobung einer

⁹⁷ Vgl. zu der Funktionsweise der Themenerkennungs-Software im Archivportal Europa das Online-Paper von Kerstin Arnold: Aggregation as a service. Automatic topic detection and collaborative topic tagging in Archives Portal Europe’s multilingual environment [2021], zuletzt abgerufen am 3. Nov. 2022 unter: < https://www.digitaltreasures.eu/wp-content/uploads/2021/12/AutomaticTopicDetection_ArchivesPortalEurope-Kerstin-Arnold.pdf > .

⁹⁸ Siehe hierzu < <https://www.informatik.hu-berlin.de/de/forschung/gebiete/ml/Flair/flair> > (zuletzt abgerufen am 3. Nov. 2022).

⁹⁹ Vgl. Auskunft Nils Meyer (LABW), E-Mails vom 14. und 22. Februar 2022.

derartigen Sacherschließung ein Ausweichen auf geeignetere Systeme notwendig. Anhand des Beispiels „Themenportal Wiedergutmachung“ im Archivportal-D ließ sich zeigen, wie unter den aktuellen Voraussetzungen der Erschließung und den technischen Möglichkeiten der Themenportal-Infrastruktur eine weitgehend manuelle Systematikerstellung sowie eine durch das „Schlagworttool“ unterstützte GND-referenzierte Indexierung des einschlägigen Archivguts gelingen kann – wenn auch unter erschwerten Rahmenbedingungen aufgrund heterogener Datenqualität.

Zwar lässt sich deshalb der entstehende Mehraufwand eines solchen Erschließungsansatzes zumindest für einzelne Themen bzw. Epochen rechtfertigen. Aufwand und Ertrag würden aber vermutlich in einem Missverhältnis stehen, reduziere man das Nutzen einer Referenz auf kontrollierte Vokabulare und Normdateien allein auf die hierdurch zu erzielende Disambiguierung von Sachbegriffen, Körperschaften usw. Vielmehr macht erst die Verknüpfung der einzelnen Verzeichnungseinheiten mit Deskriptoren im Themenportal die Schaffung eines niedrigschwelligen Zugangs zu Archivgut möglich. Zugleich ergibt sich damit das nachhaltig in die Zukunft weisende Hauptnutzen: Durch ihre Vernetzung über URIs können die Erschließungsinformationen Teil des Semantic Web werden.

2. Eine flächendeckende, normdatenbasierte Indexierung, deren Mehrwert für die Forschung von heute und morgen zwar außerordentlich wäre, kann dagegen nur bei möglichst geringem Ressourceneinsatz in Betracht gezogen werden. Die notwendigen Voraussetzungen müssen hierfür im Bundesarchiv geschaffen werden, so unter anderem durch die stetige Weiterentwicklung der existierenden Archivsoftware im Sinne eines „Fitmachens“ für eine Teilnahme am Web 3.0.

Zudem wird ein solcher Weg nicht ohne weitgehend automatisierte Verfahren zu beschreiten sein. Perspektivisch wird es darum gehen, auch durch KI-gestützte Algorithmen, die auf neuronale Sprachmodelle und die semantische Textanalyse mit Ansätzen aus dem Bereich Natural Language Processing zurückgreifen, die Datenqualität im Bundesarchiv zu erhöhen und eine automatisierte Verschlagwortung des Archivguts vorzunehmen. Die auf diese Weise generierten Schlagwörter müssten zugleich über stabile Identifier mit Normdateien wie der GND verknüpft werden bzw. dort zwecks Ausbaus einer tatsächlich spartenübergreifenden Schlagwortsystematik eingespeist werden.

Vorerst wird sich das Bundesarchiv daher im Bereich der Verschlagwortung und Normdatenanreicherung mit Teillösungen zufriedengeben müssen. Neben der oben beschriebenen Entwicklung von normdatengerechter Archivsoftware, wird man auch den eigentlichen Prozess der Erschließung stetig mit entsprechenden KI-unterstützten Tool-Boxen ausrüsten müssen. Auf diese Weise wird das eigene Archivgut durch die Generierung von Linked Data Teil des Semantic Web werden können. Dem Bundesarchiv könnte sich hierdurch zugleich die Möglichkeit eröffnen, seine Erschließung mittelfristig auf zukunftsweisende internationale Konzepte wie den im Aufbau befindlichen Standard „Records in Contexts“ (RiC) umzustellen.

6. Zusammenfassung

Mit der Transferarbeit werden die Voraussetzungen und Möglichkeiten einer Anreicherung der Erschließung im Bundesarchiv mit Schlagwörtern und Normdaten untersucht. Die Arbeit verdeutlicht die Relevanz dieses Sacherschließungsansatzes für eine Vernetzung der Erschließungsdaten innerhalb des Semantic Web und beschreibt die archivfachlichen und -technischen Rahmenbedingungen, um diesen Weg aktuell und perspektivisch zu beschreiten.

Unter Heranziehung eines Systematik- bzw. Erschließungs-orientierten Ansatzes (deduktiv/induktiv) sowie unter Berücksichtigung von Aufwand und Ertrag erfolgt für Einzelfallakten des Bestands B 106 (BMI) die Entwicklung einer Teilsystematik für das im Aufbau befindliche „Themenportal Wiedergutmachung“. Zusätzlich

werden die Chancen einer effizienteren Verschlagwortung und Normdatenanreicherung durch automatisierte, KI-gestützte Verfahren erörtert.

Im Ergebnis lässt sich die Verfolgung eines solchen Sacherschließungsansatzes durch das Bundesarchiv gegenwärtig aufgrund des Mehrwerts für die Nutzerinnen und Nutzer zwar für ausgewählte Bestände rechtfertigen. Seine ressourcenschonende Ausdehnung auf sämtliche Unterlagen wird derweil nur durch KI-Lösungen zu erreichen sein.

7. Abkürzungsverzeichnis

API =	Application Programming Interface
ARK =	Archivreferentenkonferenz des Bundes und der Länder
BASYS =	BundesArchiv IT-SYStem
BArch =	Bundesarchiv
BEG =	Bundesentschädigungsgesetz
BGBl =	Bundesgesetzblatt
BMF =	Bundesministerium der Finanzen
BMI =	Bundesministerium des Innern
BSZ-BW =	Bibliotheksservice-Zentrum Baden-Württemberg
BVerfGE =	Entscheidungen des Bundesverfassungsgerichts
BWGG =	Bundesgesetz zur Wiedergutmachung im öffentlichen Dienst
CAB =	Cascaded Analysis Broker
CSV =	Comma-separated values
DC 5 =	Digital Collections 5 (Bildatenbank Bundesarchiv)
DDB =	Deutsche Digitale Bibliothek
DFG =	Deutsche Forschungsgemeinschaft
DMA-EST-Datei =	Einheitssachtitel-Datei des Deutschen Musikarchivs
DNB =	Deutsche Nationalbibliothek
EAC =	Encoded Archival Context
EAC-CPF =	Encoded Archival Context – Corporate Bodies, Persons and Families
EAD =	Encoded Archival Description
FAIR =	Findability, Accessibility, Interoperability, Reusability
FIZ =	Fachinformationszentrum
GKD =	Gemeinsame Körperschaftsdatei
GND =	Gemeinsame Normdatei
GND4C =	GND for Cultural Data
HTR =	Handwritten Text Recognition
ICA =	International Council on Archives
ID =	Identifikator

ISAAR(CPF) =	International Standard Archival Authority Record for Corporate Bodies, Persons and Families
ISAD(G) =	General International Standard Archival Description
ISDF =	International Standard for Describing Functions
ISDIAH =	International Standard for Describing Institutions with Archival Holdings
JSON =	JavaScript Object Notation
KI =	Künstliche Intelligenz
KLA =	Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder
KWS =	Keyword Spotting
LABW =	Landesarchiv Baden-Württemberg
LAGIS =	Landesgeschichtliche Informationssystem Hessen
LCCN =	Library of Congress Control Number
LOD =	Linked Open Data
MARC =	Machine-Readable Cataloging
NER =	Named Entity Recognition
NLP =	Natural Language Processing
OCR =	Optical Character Recognition
OVG =	Ordnungs- und Verzeichnungsgrundsätze für die staatlichen Archive der DDR
PND =	Personennamendatei
RDF =	Resource Description Framework
READ =	Recognition and Enrichment of Archival Documents
RiC =	Records in Contexts
SRU =	Search/Retrieve via URL
SWD =	Schlagwortnormdatei
TGN =	Thesaurus of Geographic Names
URI =	Uniform Resource Identifier
VIAF =	Virtual International Authority File
XML =	Extensible Markup Language

8. Quellen- und Literaturverzeichnis

Literatur

- Aumann, Stefan/Sandner, Peter, Normdaten im Hessischen Landesarchiv. Erprobung von Verfahren zur Anreicherung der Erschließungsdaten in Arcinsys, in: Archivnachrichten aus Hessen Nr. 21/2 (2021), S. 104–107.
- Behrens-Neumann, Renate, Die Gemeinsame Normdatei (GND). Ein Projekt kommt zum Abschluss, in: Dialog mit Bibliotheken 24/1 (2012), S. 25–28.
- Berger, Heike Anke, Deutsche Historikerinnen 1920-1970. Geschichte zwischen Wissenschaft und Politik, Frankfurt/New York 2007.

- Brübach, Nils, Austauschformate und Normierung – welche Anforderungen stellen Portale an Archive?, in: Becker, Irmgard Christa (Hg.), Netz werken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung. Beiträge zum 19. Archivwissenschaftlichen Kolloquium der Archivschule Marburg 2015 (Veröffentlichungen der Archivschule Marburg 61), S. 157–182.
- Brübach, Nils, Zur archivischen Erschließung von Filmen und Fotos mit Normdaten, in: Archivpflege in Westfalen-Lippe 87/2017, S. 16–19.
- Brüning, Rainer/Heegewaldt, Werner/Brübach, Nils (Hg.), ISAD(G) - Internationale Grundsätze für die archivische Verzeichnung, Zweite, überarbeitete Ausgabe, Marburg 2002.
- Bundesgesetzblatt, 1951 Teil I Nr. 21, S. 291; 1961 Teil I Nr. 68, S. 1349.
- Bundesministerium der Finanzen, Kalendarium zur Wiedergutmachung von NS-Unrecht. Gesetzliche und außer-gesetzliche Regelungen sowie Richtlinien im Bereich der Wiedergutmachung nationalsozialistischen Un-rechts, Berlin 2019.
- Entscheidungen des Bundesverfassungsgerichts (BVerfGE), Bd. 18, S. 288.
- Fähle, Daniel, Gemeinsame Normdatei (GND) und Archive – eine vielversprechende Liaison, in: Erschließung 2.0: Erwartungen.- Probleme.- Lösungen. (Tagungsband zum 23. Sächsischer Archivtag in Leipzig), Chem-nitz 2020, S. 51–53.
- Fähle, Daniel, Herausforderungen bei der Benutzerführung in Portalen, in: Becker, Irmgard Christa (Hg.), Netz werken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung. Beiträge zum 19. Ar-chivwissenschaftlichen Kolloquium der Archivschule Marburg 2015 (Veröffentlichungen der Archivschule Marburg 61), S. 269–283.
- Fähle, Daniel/Seidu, Nadine, Mit „Match & Merge“ zur GND? Erprobung von Anreicherungsverfahren im Archiv-portal-D-Projekt, in: Archivar. Zeitschrift für Archivwesen 70 (2017), S. 196–198.
- Glauert, Mario, Müssen wir anders verzeichnen? Erschließung zwischen analogen Archivgewohnheiten und digitalen Nutzererwartungen, in: Archivpflege in Westfalen-Lippe 91 (2019), S. 32–37.
- Gnirs, Otto, Die Wiedergutmachung im öffentlichen Dienst, in: Finke, Hugo u.a., Entschädigungsverfahren und sondergesetzliche Entschädigungsregelungen, München 1987, S. 265-303.
- Grigoleit, Sonja, Natural Language Processing, in: Europäische Sicherheit und Technik (April 2019), S. 69.
- Herrmann, Tobias, Das Bundesarchiv in Zahlen, in: Das Bundesarchiv im digitalen Wandel (Forum. Das Fachma-gazin des Bundesarchivs, 2018), S. 36–39.
- Herrmann, Tobias/Sprau, Mirjam, Themenportale – Neue Formen digitaler Präsentation, in: Archive und Erinne-rungskulturen. Zwischen Bereitstellung und Geschichtspolitik (Forum. Das Fachmagazin des Bundesar-chivs, 2021), S. 71–75.
- Hänger, Andrea/Hollmann, Michael, Das Bundesarchiv im digitalen Wandel, in: Das Bundesarchiv im digitalen Wandel (Forum. Das Fachmagazin des Bundesarchivs, 2018), S. 4–33.
- Homa, Bernhard, Voraussetzungen und Kriterien für den Einsatz von Normdaten im Landesarchiv NRW (Trans-ferarbeit des 50. wissenschaftlichen Lehrgangs an der Archivschule Marburg), Marburg/Lahn 2019 (E-Paper abrufbar unter: <https://archiv.ub.uni-marburg.de/es/2020/0010/pdf/sbh.pdf>).
- Jagodzinski, Silke, Internationale Standards für archivische Erschließung. Von ISAD(G) zu Records in Contexts, in: Archivar. Zeitschrift für Archivwesen 73 (2020), S. 134–141.

- Jost, Stefanie, Zugänglichmachung on demand – (nicht nur) Erschließung neu gedacht. Der Strategie-prozess im Bundesarchiv, in: Erschließung 2.0: Erwartungen.- Probleme.- Lösungen. (Tagungsband zum 23. Sächsischer Archivtag in Leipzig), Chemnitz 2020, S. 35–39.
- Klaes, Christiane, Linked Open Data-Strategien zum Identity Management in einer Fachontologie. Prototypische Entwicklung eines Workflows zur Aufbereitung und zum Interlinking von Personennamen, Abschlussarbeit zur Erlangung des akademischen Grades Master of Arts (M.A.) im Studiengang Internationales Informationsmanagement – Informationswissenschaft, Stiftung Universität Hildesheim 2021.
- Kühnel, Karsten, Funktionen als Räume für institutionelle Episoden – Normdateien nach ISDF, in: Storm, Monika (Hg.), Transformation ins Digitale (85. Deutscher Archivtag in Karlsruhe), Fulda 2017, S. 135–151.
- Krauth, Wolfgang/Sandner, Peter/Westphal, Sina, Archivische Erschließung mit Normdaten, in: Archivar. Zeitschrift für Archivwesen 73 (2020), S. 142–144.
- Lill, Jens M./Lindenthal, Jutta/Balzer, Detlev, Personeninformationen und Sachbegriffe aus Museen in der GND und ihre Repräsentation im GND-Datenmodell, in: Balzer, Detlef u.a., Das Projekt „GND für Kulturdaten“ (GND4C), in: O-bib 2019/4, S. 59–97.
- Liu, Feifan u. a., Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts, in: proc. of the annual conference of the North American chapter of the Association for Computational Linguistics, Boulder 2009, S. 620–628.
- Maier, Gerald/Fähle, Daniel/Neuburger, Andreas, Bereitstellung, Aufbereitung, Langzeitsicherung. Funktionen der Archive in der Forschungsdateninfrastruktur, in: Archivar. Zeitschrift für Archivwesen 73 (2020), S. 13–18.
- Menne-Haritz, Angelika, Erschließung, in: Marcel Lepper/Ulrich Raulff (Hgg.): Handbuch Archiv. Geschichte, Aufgaben, Perspektiven, Stuttgart 2016, S. 207–217.
- Meyer, Nils, Sachthematische Zugänge im Archivportal-D. Archive und ihre Bestände zusammenführen und neu entdecken, in: Archivar. Zeitschrift für Archivwesen 72 (2019), S. 37–39.
- Mühlberger, Günter/Terbul, Tamara, Handschriftenerkennung für historische Schriften. Die Transkribus Plattform, in: b.i.t. online 21 (2018) Nr. 3, S. 218–222.
- Müller, Gerhard, Die Normdatei als ein Mittel in der Erschließung von Archivbeständen, in: Brandenburgische Archive 32 (2015), S. 8–16.
- Müller, Gerhard/Jagodzinski, Silke, Die Erschließung des Kontexts. Neue Perspektiven auf ein bewährtes Prinzip, in: Archivpflege in Westfalen-Lippe 90 (2019), S. 10–17.
- Neuburger, Andreas, Vom Desiderat zur gelebten Praxis! Die Nutzung der Gemeinsamen Normdatei (GND) in der Erschließung, in: Archivpflege in Westfalen-Lippe 95 (2021), S. 30–34.
- Papritz, Johannes, Die archivische Titelaufnahme bei Sachakten, Marburg 1967.
- Römmer, Christian, Entschädigung Erster Klasse? Die Wiedergutmachung im öffentlichen Dienst in Hamburg nach dem Zweiten Weltkrieg, München/ Hamburg 2003.
- Sprau, Mirjam/Herrmann, Tobias, Das Themenportal „Wiedergutmachung nationalsozialistischen Unrechts“ im Archivportal-D, in: Archivar. Zeitschrift für Archivwesen 74 (2021), S. 275–277.
- Staatliche Archivverwaltung im Ministerium des Innern der Deutschen Demokratischen Republik (Hg.), Ordnungs- und Verzeichnungsgrundsätze für die staatlichen Archive der Deutschen Demokratischen Republik [1964], abrufbar unter: https://www.fh-potsdam.de/fileadmin/user_dateien/2_studieren-FB_Infowiss/landesfachstelle/archivberatung/erschliessung/OVG.pdf.

Uhde, Karsten, Suchet, so werdet ihr finden – die Geschichte der Archivportale, in: Irmgard Christa Becker (Hg.), Netz werken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung. Beiträge zum 19. Archivwissenschaftlichen Kolloquium der Archivschule Marburg 2015 (Veröffentlichungen der Archivschule Marburg 61), S. 47–61.

Waidmann, Susanne, Archivportale auf europäischer Ebene, in: Becker, Irmgard Christa (Hg.), Netz werken. Das Archivportal-D und andere Portale als Chance für Archive und Nutzung. Beiträge zum 19. Archivwissenschaftlichen Kolloquium der Archivschule Marburg 2015 (Veröffentlichungen der Archivschule Marburg 61), S. 132–155.

Ziwes, Franz-Josef, Archive als Leuchttürme. Die Erschließung mit Normdaten als Aufgabe und Chance, in: Archive ohne Grenzen. Erschließung und Zugang im europäischen und internationalen Kontext. 83. Deutscher Archivtag in Saarbrücken (Tagungsdokumentation zum Deutschen Archivtag 18), Fulda 2014, S. 79–87.

9. Internetressourcen [letzter Abruf]

Arnold, Kerstin, Aggregation as a service. Automatic topic detection and collaborative topic tagging in Archives Portal Europe's multilingual environment [2021], https://www.digitaltreasures.eu/wp-content/uploads/2021/12/AutomaticTopicDetection_ArchivesPortalEurope-Kerstin-Arnold.pdf [3. Nov. 2022].

Archivreferentenkonferenz des Bundes und der Länder, Abschlussbericht der ARK-Bund-Länder-Arbeitsgruppe „Wiedergutmachung“, Düsseldorf 2009, https://www.bundesarchiv.de/DE/Content/Downloads/KLA/wiedergutmachung-abschlussbericht.pdf?__blob=publicationFile [17. Nov. 2022].

Archiv Soziale Bewegungen, Der Digitale Archivar, <https://staccared.com/index.html> [15. Nov. 2022].

ARK-Bund-Länder-Arbeitsgruppe, Übersicht über die Überlieferung und Rechtsgrundlagen zur Wiedergutmachung nationalsozialistischen Unrechts in der Bundesrepublik Deutschland in den staatlichen Archiven, Düsseldorf 2010, https://www.bundesarchiv.de/DE/Content/Downloads/KLA/wiedergutmachung-dokumentation.pdf?__blob=publicationFile [17. Nov. 2022].

Bundesarchiv, Akten der Reichskanzlei (Online-Edition), <https://www.bundesarchiv.de/aktenreichskanzlei/1919-1933/0000/index.html> [21. Nov. 2022].

Bundesarchiv, Die Kabinettsprotokolle der Bundesregierung (Online-Edition), <https://www.bundesarchiv.de/cocoon/barch/0000/index.html> [21. Nov. 2022].

Bundesarchiv/Landesarchiv Baden-Württemberg, Handbuch zur Aufbereitung von Archivgut für das Themenportal zur Weimarer Republik, <https://cms.archivportal-d.de/sites/default/files/media/document/2021-06/Guideline%20zur%20Verschlagwortung.pdf> [15. Nov. 2022].

Bundesarchiv/Landesarchiv Baden-Württemberg, Themenportale gestalten. Sachthematische Zugänge im Archivportal-D, <https://cms.archivportal-d.de/sites/default/files/media/document/2021-05/Guideline%20zum%20Aufbau%20von%20Systematiken.pdf> [15. Nov. 2022].

Bundesarchiv, Nachlassdatenbank, <https://www.bundesarchiv.de/nachlassdatenbank/> [21. Nov. 2022].

Bundesarchiv, Übergreifende nationale und internationale Portale, <https://www.bundesarchiv.de/DE/Content/Artikel/Finden/Unterseiten-Weitere-Recherchesysteme/uebergreifende-portale.html> [16. Nov. 2022].

Bundesministerium der Finanzen, Das Archivierungsprojekt der Wiedergutmachung und seine Bedeutung im Kampf gegen den Antisemitismus (Monatsbericht, Januar 2021), <https://www.bundesfinanz>

ministerium.de/Monatsberichte/2021/01/Inhalte/Kapitel-3-Analysen/3-7-archivierungsprojekt-wiedergutmachung-pdf.pdf?__blob= publicationFile&v= 4 [8. Nov. 2022].

Bundesministerium der Finanzen, Themenportal Wiedergutmachung, https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Themen/Oeffentliche_Finzen/Vermoeigensrecht_und_Entschaedigungen/themenportal-wiedergutmachung.html [15. Nov. 2022].

Deutsches Textarchiv, Linguistische Analyse historischer Texte (CAB), <https://www.deutsches-textarchiv.de/doku/software#cab> [15. Nov. 2022].

Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS, Online-Glossar zum Thema Künstliche Intelligenz, https://www.iais.fraunhofer.de/de/forschung/kuenstliche-intelligenz.html#faq_faqitem_88176962-answer [3. Nov. 2022].

Humboldt-Universität zu Berlin (Institut für Informatik), Das Flair NLP Framework, <https://www.informatik.hu-berlin.de/de/forschung/gebiete/ml/Flair/flair> [3. Nov. 2022]

International Council on Archives, Records in Contexts. Conceptual Model. Consultation Draft v0.2 July 2021, https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf [22. Nov. 2022].

Neuburger, Andreas, „Erschließung biographischer Unterlagen: Nutzung der Gemeinsamen Normdatei (GND) durch Archive“ [Livestream-Mitschnitt des Vortrags vom 72. Westfälischer Archivtag 2021], <https://www.youtube.com/watch?v= 5XcdCoiZtew> [7. Nov. 2022].

READ-COOP, Transkribus Glossar: Keyword Spotting, <https://readcoop.eu/de/glossary/keyword-spotting-kws/> [3. Nov. 2022].

xTree.public, Startseite, <http://xtree-public.digicult-verbund.de/vocnet/?action= start&lang= de> [15. Nov. 2022].

xTree.public, Vokabular: Archivportal-D zur Weimarer Republik, <http://xtree-public.digicult-verbund.de/vocnet/?uriVocItem= http://vocab.archivportal-d.de/&startNode= 01099&lang= de&d= n> [15. Nov. 2022].

10. Gesprächsvermerke und sonstige Auskünfte

E-Mail von Karsten Kühnel (BArch, Abt. B – Außenstelle Bayreuth, Leiter des Lastenausgleichsarchivs) vom 21. Febr. 2022.

E-Mail von Oliver Sander (BArch, Referat IT 2) vom 8. Febr. 2022.

E-Mails von Nils Meyer (LABW) vom 14. Febr. und 22. Febr. 2022.

Telefonat mit Kerstin Oldenhage (BArch, Referatsleiterin B 1) am 2. Febr. 2022.

Telefonat mit Martina Scheid (BArch, Referat B 1) am 1. Febr. 2022.

Telefonat mit Maurice Augel (BArch, Referat IT 1) am 7. Febr. 2022.

Videokonferenz mit Georg Eckes (BArch, Referatsleiter IT 3) am 11. Febr. 2022.

Videokonferenz mit Sarah Bormer (BArch, Referat IT 2) am 21. Dez. 2021.

Videokonferenz mit Vincent Kutz, Oliver Sander (beide BArch, Referat IT 2) am 10. Jan. 2022.

Videokonferenzen mit Mirjam Sprau und Kevin Dubout (BArch, Referat GW 1 und Projekt „Themenportal Wiedergutmachung“) am 14. Jan., 21. Jan. und 25. Febr. 2022.

Anhang 1: Teilsystematik für das „Themenportal Wiedergutmachung“ unter Angabe der zu den Schlagwörtern passenden GND-IDs

[Weitere Themen]

Entschädigung

I. Entschädigung nach anderen Regelungen bis 1990

1. Beschäftigte im Öffentlichen Dienst

a. Antragsteller

i. Geschädigte und Versorgungsberechtigte

1. Beamter	4005081-6
2. Angestellter	4002025-3
3. Arbeiter	4112560-5
4. Lehrer	4035088-5
5. Hochschullehrer	4025243-7
6. Bürgermeister	4136676-1
7. Arzt	4003157-3
8. Pflegepersonal	4174087-7
9. Polizeibeamter	4135253-1
10. Berufssoldat	4144806-6
11. Ruhestandsbeamter	4225045-6
12. Wartestandsbeamter	-
13. Versorgungsempfänger	4188025-0
14. Hinterbliebener	4254622-9

ii. Dienst- und Arbeitsstelle

1. Schule	4053474-1
2. Universität	4061778-6
3. Polizei	4046595-0
4. Gemeindeverwaltung	4019990-3
5. Kreisverwaltung	4393380-4
6. Bezirksverwaltung	4145207-0
7. Provinzialverwaltung	4194366-1
8. Landesverwaltung	4393385-3
9. Ministerialverwaltung	4134363-3
10. Jüdische Gemeinde	4162805-6
11. Religionsgemeinschaft	4049401-9
12. Deutsches Reich. Wehrmacht	26159-2
13. Verkehrsbetrieb	4062911-9
14. Deutsche Reichsbahn	2008004-9
15. Bank	4004436-1
16. Theater	4059702-7
17. Gericht	4121346-4

iii. Mitgliedschaft

1. NSDAP	1012979-0
2. SS	10162246-6
3. SA	63215-6
4. Politische Organisation	4129522-5
5. Berufsverband	4005953-4
6. Reichskulturkammer	18668-5

b. Antrags- und Verfahrensgegenstand

i. Schädigung und Benachteiligung

1. Entlassung	4152363-5
2. Versetzung	4187992-2

- 3. Pensionierung 4173668-0
- 4. Beförderung 4129072-0
- 5. Beamtenverhältnis 4134677-4

ii. Gewährung und Neuberechnung einer Leistung

- 1. Rente 4049496-2
- 2. Hinterbliebenenversorgung 4159911-1
- 3. Ruhegehalt 4178676-2
- 4. Übergangsgeld 4331470-3
- 5. Rechtsstellung 4134078-4
- 6. Besoldungsdienstalter 4144908-3
- 7. Überversicherung 4186653-8
- 8. Härteausgleich 4120417-7
- 9. Wiedereinstellung 4203335-4

c. Verfahrensergebnis

- i. Gewährung -
- ii. Teilgewährung -
- iii. Nichtgewährung -
- iv. Prozess 4047577-3

*Anlage 2: Mapping im xTree-Thesaurus zur Weimarer Republik***Details: Beamter****URI**

<http://vocab.archivportal-d.de/00066>

Typ

Begriff

Bezeichnung

Beamter	Bevorzugt	de
Beamte	Alternativ	de
Beamtentum	Alternativ	de
Beamtin	Alternativ	de
Berufsbeamtentum	Alternativ	de
Berufsbeamter	Alternativ	de
Staatsbeamter	Alternativ	de

Notation

saz

Mapping

exakt <http://d-nb.info/gnd/4005081-6>
übereinstimmend

annähernd
übereinstimmend <http://d-nb.info/gnd/4005085-3>
(close)

exakt <http://www.wikidata.org/entity/Q58891819>
übereinstimmend

Quelle: <http://xtree-public.digicult-verbund.de/vocnet/?uriVocItem=http://vocab.archivportal-d.de/00700&startNode=01099&lang=de> (abgerufen am 16. März 2022).

Anhang 3: Cluster-Beispiel aus OpenRefine

5	6	<ul style="list-style-type: none"> Hinterbliebenenversorgung (2 Zeilen) Hinterbliebenenversorgung Hinterbliebenenversorgung Hinterbliebenenversorgung Hinterbliebenenversorgung 	<input checked="" type="checkbox"/>	Hinterbliebenenversorgung
5	6	<ul style="list-style-type: none"> Bürgermeister (2 Zeilen) 1. Bürgermeister 2. Bürgermeister Amtsbürgermeister Oberbürgermeister 	<input checked="" type="checkbox"/>	Bürgermeister
5	6	<ul style="list-style-type: none"> Oberschullehrer (2 Zeilen) Berufsschullehrer Oberschullehrerin Realschullehrer Volksschullehrer 	<input checked="" type="checkbox"/>	Lehrer
5	7	<ul style="list-style-type: none"> Stadt Breslau (2 Zeilen) Stadt Stettin (2 Zeilen) Stadt Berlin Stadt Elbing Stadt Wehlau 	<input type="checkbox"/>	Stadt Breslau

Quelle: OpenRefine 3.5 (zuletzt verwendet am 25. Febr. 2022).