

Vom gedruckten Gazetteer zum digitalen Ortsverzeichnis

Digitale geographische Hilfsmittel aus Bürgerwissenschaft und Digital Humanities im Vergleich

Bedeutung des Raumes für die Geschichtswissenschaft

Der Raum ist neben der Zeit eine elementare Kategorie der Geschichtswissenschaft. Beide Grundkategorien sollten bei der Verarbeitung und Analyse von umfangreichen Datenbeständen zum einen mehr Berücksichtigung erfahren, zum anderen in ihren strukturierenden und Interpretationen erleichternden Funktionen ernst genommen werden. Dafür erforderlich sind einfach anzuwendende Tools, die einen Schwerpunkt auf die Koppelung von Zeit und Raum legen, um dadurch eine Verortung im geographischen und im historischen Raum zu ermöglichen. Bekanntlich agiert Geschichte in einem geographischen Nebeneinander und einem chronologischen Nacheinander.¹ Somit gehört zur Geschichtswissenschaft, didaktisch gesprochen, eine „Raum-Zeit-Kompetenz“:² Geschichtliche Prozesse spielen sich in Räumen ab, jeder Raum ist geschichtlich geworden, und Räume wurden in verschiedenen Zeiten unterschiedlich bewertet. Studentinnen und Studenten der Geschichtswissenschaften werden bereits zu Beginn des Studiums mit den Formalia und der Unabdingbarkeit der Quellenkritik konfrontiert. Eines von deren Elementen ist die räumliche Verortung von Quellen: Briefe werden an einem Ort verschickt und sind an einen anderen Ort adressiert, Bücher werden in unterschiedlichen Städten publiziert, Reisende bewegen sich im geographischen Raum. Um die Dimensionen des historischen Raumes genauer abzustecken, werden Werkzeuge gebraucht, die in der Lage sind, räumliche Zuweisungen zu ermöglichen und gleichzeitig die Veränderung des Raumes in seiner zeitlichen Entwicklung abzubilden. Dass Ortsnamen einem historischen Wandel unterliegen oder im Laufe der Zeit verschwinden können, sollte nicht dazu führen, solche Räume nicht historisch zu analysieren, sondern dass die Geschichtswissenschaft sich Hilfsmittel erarbeitet, um Schwierigkeiten bei der Aufbereitung,

DOI: 10.25365/rhy-2021-14



Anne Purschwitz, Institut für Geschichte, Martin-Luther-Universität Halle-Wittenberg, Emil-Abderhalden-Str. 26–27, 06108 Halle (Saale), Deutschland, anne.purschwitz@geschichte.uni-halle.de; Jesper Zedlitz, Verein für Computergenealogie (Geschäftsstelle), Piccoloministraße 397a, 51067 Köln, Deutschland, jzedlitz@compgen.de

1 Ahasver von Brandt, *Werkzeug des Historikers*, Stuttgart 1958, 22.

2 Für das Folgende vgl. Walter Sperling, *Die Stellung der Historischen Geographie in einem modernen geographischen Curriculum*, in: *Erdkunde* 26 (1982), 79–84, 81.

Präsentation und Recherche zu überwinden. Im Verlauf der Zeit änderten sich Schreibweisen, Namen wurden überformt oder neu vergeben, die Zugehörigkeit zu Sprachräumen konnte variieren und sich ebenfalls auf die Namengebung von Wohnplätzen und Siedlungen auswirken. Die Mehrheit der aktuell verfügbaren Ortsverzeichnisse berücksichtigt diesen Wandlungsprozess nur unzureichend bis gar nicht. Doch gerade die Kombination von Zeit und Raum stellt für die Anwendung in der Geschichtswissenschaft eine zentrale Anforderung an Ortsverzeichnisse dar.

„Orte haben eine historisch-politische Dimension, die bei einer übergreifenden Registererfassung erst sichtbar zu einem Problem wird. [...] Für die Visualisierung von Briefen etwa sind historische Karten ein Desiderat; generell auch Geodaten für Flächen. Und alle mit Geodaten versehenen Einträge müssen mit einem Zeitstempel kombiniert sein, denn beispielsweise die Altstadt von Jerusalem ist eben heute nicht am selben Ort wie vor 2.000 Jahren.“³

Gazetteers im Kontext historischer Geodatenysteme

Seit einigen Jahrzehnten werden im deutschsprachigen Raum Webportale zur historischen Geographie aufgebaut, unter anderem HGIS-Germany⁴, die *Historical GIS datafiles* des Mosaic-Projekts⁵, HistoGIS⁶ und verschiedene lokale Portale aus dem Umfeld der *Time Machine Organization*. Diese bieten tendenziell zwei verschiedene Funktionalitäten. Erstens – darauf deuten schon die Selbstbenennungen als GIS oder auch als Webapplikation hin – geht es um das aktive Online-Erstellen von Karten, mit denen strukturierte historische Informationen visualisiert werden können. Dazu gehören nicht nur verschiedene graphische Funktionen, sondern auch solche der Datenabfrage und -bearbeitung. Ein alternativer Weg hin zur Karte wäre der, ein GIS im engeren Sinne des Begriffs zu nutzen. Den Industriestandard bietet ArcGIS von ESRI⁷; den in der universitären Forschung dominierenden Open-Source-Standard bietet QGIS.⁸ Die zweite Funktionalität besteht darin, Shapefiles, also Dateien mit georeferenzierten Informationen, zu erstellen und zu archivieren, unabhängig davon, ob und welche Webanwendungen geboten werden. Dabei geht es vor allem um historische Umgren-

3 Roland Kamzelak, Von der Raupe zum Schmetterling oder Wie fliegen lernen – Editionsphilologie zwischen Infrastruktur und Semantic Web, in: Ders./Timo Steyer (Hg.), *Digitale Metamorphose: Digital Humanities und Editionswissenschaft* (Sonderband der Zeitschrift für digitale Geisteswissenschaften 2), DOI: 10.17175/sb002_004.

4 <https://www.digihist.de/> (20.4.2021).

5 <https://censusmosaic.demog.berkeley.edu/data/historical-gis-files> (20.4.2021)

6 Vgl. <https://histogis.acdh-dev.oew.ac.at/> (20.4.2021). HistoGIS versteht sich als Repitorium historischer Polygondaten (einschließlich großer Teile der zuvor benannten Datenbestände) mit einheitlichen Metadaten. Schwerpunkte liegen aktuell auf der Periode zwischen dem ausgehenden 18. Jahrhundert und 1918 für Österreich-Ungarn und das Deutsche Reich. Ausführlicher dazu: Matthias Schlögl/Peter Andorfer, HistoGIS: Vom Punkt zur Fläche in Raum und Zeit, in: *Digital Humanities: multimedial und multimodal*. 6. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ (DHd 2019), Frankfurt am Main und Mainz, DOI: 10.5281/zenodo.4622271.

7 Vgl. <https://www.esri.de/de-de/home> (10.5.2021).

8 Vgl. <https://www.qgis.org/de/site/> (10.5.2021).

zungen, aber auch um Strecken (z.B. Wasserwege) und Punktdaten (z.B. Orte), jeweils mit einem definierten zeitlichen Bezug.

Mit dem Begriff „Gazetteer“ (Ortsverzeichnis, Ortslexikon) wird im vorliegenden Beitrag eine andere und wissenschaftlich ältere Funktionalität angesprochen. Auch wenn sie zum Teil Geokoordinaten enthalten, stellen Gazetteers einen eigenen Typ von Informationssammlung dar, der weder mit Systemen zur Analyse geographischer Daten verwechselt werden sollte, noch mit Repositorien von punkt-, linien- oder flächenbezogenen kartographischen Informationen. Bei Gazetteers steht die eindeutige Identifikation von Ortsbezeichnungen und die Dokumentation ihrer Zuordnung zu übergeordneten Strukturen über historische Brüche hinweg im Zentrum.

Gazetteers blicken dabei auf eine lange Entwicklungsgeschichte zurück. Erste Gazetteers sind bereits aus der Antike bekannt. Mit der zunehmenden Vermessung der Welt wurden sie zunächst ein immer wichtigerer Bestandteil von Atlanten und ähnlichen Nachschlagewerken. Seit dem 19. Jahrhundert finden sie sich dann auch als eigenständige Publikationen.⁹ Ein Teil dieser gedruckten mehrbändigen Verzeichnisse hat in den letzten Jahrzehnten seinen Weg in das Semantic Web gefunden und nutzt die Vielzahl der dadurch eröffneten digitalen Funktionsweisen. Am deutlichsten spiegelt sich die Digitalisierung von Gazetteers wahrscheinlich in Meyers Gazetteer wider. Basierend auf Meyers gedrucktem *Orts- und Verkehrslexikon des Deutschen Reiches*, dessen Ziel es war, jeden Ort im Deutschen Reich (1871–1918) aufzulisten und übergeordneten zivilen und juristischen Verwaltungseinheiten zuzuordnen, wurde eine für die Online-Nutzung aufbereitete Version erarbeitet. Sie ergänzt das klassische Lexikon um Such- und Filterfunktionen, stellt Digitalisate und Transkripte zur Verfügung und ermöglicht die Verortung der ca. 200.000 Orte auf historischen Karten des Deutschen Reiches.¹⁰

In den vergangenen Jahren sind in unterschiedlichen Disziplinen Gazetteers erstellt, jeweils für sie relevante Informationen zusammengetragen und im Semantic Web publiziert worden.¹¹ Die große Zahl der Verzeichnisse spiegelt dabei den offensichtlichen Bedarf an strukturierten geographischen Daten wider, zeigt aber auch die variierenden Herangehensweisen, Datenerhebungen, Standards und Interessen. Ein Schwerpunkt dieser digitalen (nicht digitalisierten) Gazetteers besteht mittlerweile in der Zusammenführung von Raumdaten (Koordinaten), Sachdaten (Ergänzung der Raumdaten um die übergeordneten Strukturen,

9 So z.B. Heinrich Rudolph, Vollständigstes geographisch-topographisch-statistisches Orts-Lexikon von Deutschland, so wie der unter Oesterreichs und Preussens Botmässigkeit stehenden nichtdeutschen Länder: enthaltend: alle Städte, Flecken, Pfarr- Kirch- und andere Dörfer, Ort- und Bauerschaften, Kirchspiele, Schlösser, Rittergüter, Vorwerke, Weiler, Hüttenwerke, Mühlen, Höfe, merkwürdige Ruinen, Krüge, Einschichten, Einöden u.s.w., 2 Bde., Leipzig 1870–1872.

10 Vgl. <https://www.meyersgaz.org/> (10.5.2021).

11 Aus der Fülle seien genannt: Die wohl größte, frei verfügbare Datenbank bietet GeoNames. Sie enthält über 25 Millionen geographische Bezeichnungen auf der ganzen Welt, die in jeweils eine von neun Klassen und einen von über 645 Codes kategorisiert sind. Neben Ortsnamen in verschiedenen Sprachen sind auch Längen- und Breitengrad, Höhe über Meer, Einwohnerzahl, administrative Unterteilung und Postleitzahlen in der Datenbank enthalten. Hier finden sich jedoch nahezu ausschließlich aktuelle Ortsnamen (<https://www.geonames.org/>, 1.3.2022). – Zunächst für Ortsdaten der antiken Welt entwickelt, inzwischen auch zeitlich erweitert wurde das Projekt Pelagios (<https://pelagios.org/>, 10.5.2021). Historisch-geographische Daten zur antiken Welt stellt auch Pleiades zu Verfügung (<https://pleiades.stoa.org/>, 10.5.2021). Der World-Historical Gazetteer beabsichtigt, Inhalte und Dienste bereitzustellen, die es Historiker*innen, Studierenden und der allgemeinen Öffentlichkeit ermöglichen, räumliche und zeitliche Schlussfolgerungen und Visualisierungen selbst durchzuführen (<http://whgazetteer.org/>, 10.5.2021).

z.B. Kreis, Kirche, Staat) und Metadaten (Beschreibungen der Daten). Die dafür erforderlichen Informationen können punktuell für bestimmte Zeitschnitte erhoben und aufbereitet werden. Für Historikerinnen und Historiker von besonderem Interesse ist dabei die Zusammenführung von Sachdaten zu einem Ort unter Berücksichtigung von Zeiträumen, mit anderen Worten: Zeitangaben und Zeiträume zu verzeichnen, ohne jedoch anlässlich von Diskontinuitäten jeweils ‚neue‘ Orte zu schaffen, sondern vielmehr mit dem Ziel, die historische Genese von Siedlungen im administrativen Kontext nachvollziehbar zu machen.

Für die meisten Historiker*innen ist das Sammeln und Strukturieren solcher Informationen kein Selbstzweck. Es dient vielmehr bestimmten praktischen Bedürfnissen. Eine nicht ungewöhnliche Situation besteht etwa darin, dass man in einem Quellenbestand eine größere Zahl an Ortsnamen vorfindet und diese eindeutig, effizient und zugleich auch fehlerfrei identifizieren möchte. Der folgende Beitrag fragt nach der Praxistauglichkeit solcher digitalen Tools an zwei Beispielen: dem aus der Genealogie stammenden Geschichtlichen Ortsverzeichnis (GOV) und dem im Arbeitsfeld der Digital Humanities entstandenen und in der Geschichtswissenschaft mittlerweile gut etablierten und genutzten Dariah Geo-Browser (DGB). Während das Geschichtliche Ortsverzeichnis eindeutig als digitaler Gazetteer einzustufen ist, greift der Dariah-Browser als Portal mit ausgebauter Kartenfunktionalität auf andere Datenquellen zu, darunter den Getty Thesaurus of Geographic Names, der aktuelle wie auch historische Ortsdaten enthält.

Anwendungsbeispiel Auswandererbriefe

Für einen konkreten Anwendungstest wurde ein Datensatz genutzt, der eine Vielzahl von praxisnahen Problemen in Hinsicht auf die quellenbasierte Erhebung von Ortsnamen bietet und zudem seinen Schwerpunkt nicht ausschließlich im deutschsprachigen Raum findet, nämlich die Briefe der Deutschen Auswandererbriefsammlung (DABS).¹² Auswandererbriefe, die vornehmlich von ‚einfachen‘ Menschen geschrieben wurden, sind eine durch nichts zu ersetzende Quelle für die Kultur- und Alltagsgeschichte der Migration, aber auch für die Sozial-, Mentalitäts- und Sprachgeschichte. Sie sind neben den wenigen erhaltenen Tagebüchern die einzigen zeitgenössischen und tatsächlich subjektiven sozialgeschichtlichen Zeugnisse bzw. Ego-Dokumente für die Prozesse der Auswanderungsentscheidung sowie der Orientierung und Integration in Nordamerika. Diese ganz persönlichen Zeugnisse über das Erleben des Gastlandes – Fremdes und Vertrautes, Sprachprobleme und Heimweh, Dis-

12 Die folgende Darstellung orientiert sich an einer dankenswerter Weise von Nora Isabelle Mergner zur Verfügung gestellten Selbstbeschreibung des Projekts (<http://www.auswandererbriefe.de/>, 1.3.2022). Auf der Grundlage der Deutschen Auswandererbriefsammlung sind zwei große Editionen in Print-Form erschienen: Wolfgang Helbich/Walter D. Kamphoefner/Ulrike Sommer (Hg.), Briefe aus Amerika. Deutsche Auswanderer schreiben aus der Neuen Welt 1830–1930, München 1988; Wolfgang Helbich/Walter D. Kamphoefner (Hg.), Deutsche im Amerikanischen Bürgerkrieg. Briefe von Front und Farm 1861–1865, Paderborn 2002. Für eine Auswertung einzelner Briefserien unter Fragestellungen von Erinnerungsgeschichte und dem Aspekt der Narrativität vgl. Ursula Lehmkuhl, Johann Heinrich Carl – The Revolutionary: The History and Collective Memory of a German-American Family, 1852–2004, in: *Studia Migracyjne* 40/1 (2014), 31–56; Ursula Lehmkuhl, Reading Immigrant Letters and Bridging the Micro-Macro Divide, in: *Studia Migracyjne* 40/1 (2014), 9–30. Zum Genre Auswandererbrief vgl. Ursula Lehmkuhl, Das Genre „Auswandererbrief“, in: Eve-Marie Becker u.a. (Hg.), *De Gruyter-Handbuch Brief*, Berlin 2020, 639–653.

kriminierung und wirtschaftlichen Erfolg, Zerrissenheit zwischen alter und neuer Heimat, Ängste, Hoffnungen, Stolz auf Erreichtes und auf die eigene ethnische Gruppe – machen Auswandererbriefe zu einer idealen Quelle, um Lebens- und Wissenswelten der emigrierten Deutschen zu rekonstruieren.

Aktuell umfasst die Deutsche Auswandererbriefsammlung knapp 12.000 zwischen 1800 und 2004 verfasste Briefe mit über 36.000 Briefseiten von 1.523 namentlich erfassten Briefschreiberinnen und -schreibern. Für die Anwendungsstudie stellten Ursula Lehmkuhl und Nora Isabelle Mergner eine Liste mit Auswandererbriefen des ‚langen‘ 19. Jahrhunderts (1800–1930) zur Verfügung, diese umfasst insgesamt 8.299 Briefe aus 414 Briefserien. Die zu Grunde liegende Datenbank basiert auf der auch im Internet einsehbaren Übersicht über die Bestände der Deutschen Auswandererbriefsammlung.¹³ Sie enthält Angaben zu Namen, Herkunftsort und -region der Briefschreibenden, Umfang der Korrespondenz, Transkriptionen und zusätzliches Kontextmaterial, Zeitraum der Briefkorrespondenz und Wohnorte in den USA.

Die Geolokalisation deutscher Auswander*innen und ihrer Briefpartner*innen ist für die Erforschung der deutschen Emigration im ‚langen‘ 19. Jahrhundert von besonderer Relevanz, denn sie ermöglichen es, die Mobilitätsmuster nach der Migration sichtbar zu machen. Dies ist vor allem im Hinblick auf die Frage nach dem Zusammenhang von geographischer und sozialer Mobilität wichtig. Es kann beobachtet werden, dass Ausgewanderte, die eine hohe geographische Mobilität nach der Migration zeigen, auch in wirtschaftlicher und sozialer Hinsicht eine hohe *upward mobility* aufweisen. Das Thema „Mobilität nach der Migration“ soll unter Leitung von Ursula Lehmkuhl in zwei Forschungsprojekten mit Hilfe von digitalen Auswertungsmethoden untersucht werden. Auf der anderen Seite ist die Geolokalisation für die bürgerwissenschaftliche Erforschung der deutschen Auswanderung wichtig. So können regionale Cluster sowohl für die Herkunfts- als auch die Siedlungsorte identifiziert werden. Das ermöglicht eine gezielte Einbeziehung von Ortshistoriker*innen und regional arbeitenden Genealog*innen, wodurch die weitere biographische und Sacherschließung des Briefmaterials unterstützt werden soll. Die auf lokaler und regionaler Ebene Forschenden verfügen über Kenntnisse, die von unschätzbarem Wert sind, und sollen künftig in Erschließung, Identifikation und Kontextualisierung mit einbezogen werden. Die Deutsche Auswandererbriefsammlung beabsichtigt daher, auf Grundlage der durch Geolokalisation identifizierten regionalen Cluster bürgerwissenschaftliche Forschungsprojekte zu spezifischen Briefserien zu initiieren.

Tools der Geo-Identifikation

Geschichtliches Ortsverzeichnis (GOV)

Der bürgerwissenschaftliche *Verein für Computergenealogie* (CompGen) hat in den vergangenen Jahrzehnten ein auf flächendeckende Erfassung und größtmögliche Erschließungstiefe zielendes historisches Ortsverzeichnis entwickelt. Das GOV erfasst für Deutschland und viele andere Länder Ortsnamen (in variierenden Schreibweisen), kirchliche und staatliche Zuge-

13 <http://www.auswandererbriefe.de/quellenbestand.html> (1.3.2022).

hörigkeiten im Zeitverlauf sowie geographische Koordinaten in einer Form, die weit über andere Ansätze (GND, GeoNames) hinausgeht.¹⁴ Das 1992 von Heinz Augustin als „Genealogisches Ortsverzeichnis“ initiierte Projekt wurde im Jahr 1995 an CompGen übergeben und dort unter Einbeziehung einer Vielzahl von Familienforscher*innen zum „Geschichtlichen Ortsverzeichnis“ weiterentwickelt und mit Daten gefüllt. Eine web-basierte Anwendung steht seit dem Jahr 2000 zur Verfügung. Aktuell beträgt die Abdeckung auf Ebene der Siedlungsplätze (also unterhalb der Gemeinden) für das späte Kaiserreich etwa 80 Prozent.¹⁵ Insgesamt verzeichnet das GOV ca. 1,25 Millionen Einträge (Stand Juni 2021), die beständig ergänzt und erweitert werden. Bisherige räumliche Erschließungsschwerpunkte sind Europa, die USA sowie Australien.¹⁶

Ziel des Geschichtlichen Ortsverzeichnisses ist es, Siedlungen, Wohnplätze, Orte etc. eindeutig zu identifizieren. Da dafür allein ein Name nicht ausreichend ist, erfolgt die Einordnung in administrative Zusammenhänge, wobei an dieser Stelle klar zwischen verschiedenen Formen von politischer, kirchlicher und gerichtlicher Zugehörigkeit unterschieden wird. Da historische Daten einer Vielzahl von Veränderungen unterliegen, kann auch eine solche Einbettung in übergeordnete Ebenen Änderungen und Verschiebungen erfahren. Die in den Quellen vorgefundenen Namen müssen nicht eine Siedlung bezeichnen, sondern können ebenso auch auf Verwaltungseinheiten verweisen.

Das GOV enthält daher eindeutige Identifikatoren sowohl für Siedlungen als auch für administrative Objekte, die in ihren hierarchischen Kombinationen im zeitlichen Verlauf unterschiedliche, als topologische Beziehungen explizit modellierte Verbindungen eingehen können. Die räumlichen und zeitlichen Schwerpunkte bei der Erfassung von historischen Verwaltungsobjekten im GOV liegen im 20. und späten 19. Jahrhundert in Deutschland und dem östlichen Mitteleuropa. Zeiten werden als Zeiträume, nicht Querschnitte (*time slices*) modelliert.¹⁷ Entstanden ist ein hierarchisches Modell, das zwischen Siedlung und Administration trennt, beide Ebenen aber miteinander in Beziehung setzt – wobei auch Mehrfachverknüpfungen möglich sind, da eine Siedlung zeitlich etwa zu unterschiedlichen politischen und kirchlichen Administrationen gehören kann. Im Bereich des Semantic Web sind solche Identifikatoren in Form von Uniform Resource Identifier (URI) üblich. Das GOV bietet für Wohnplätze und Verwaltungsobjekte genau solche URIs zur eindeutigen Identifizierung an.

Aktuell verzeichnet das GOV für Siedlungen und Verwaltungseinheiten 276 Typen.¹⁸ Das zentrale Element im Modell ist das GOV-Objekt. Ein GOV-Objekt ist über seine Eigenschaften und über seine Beziehungen zu anderen GOV-Objekten definiert und wird über eine GOV-Kennung eindeutig identifiziert. Alle Beziehungen zwischen GOV-Objekten können eine zeitliche Komponente aufweisen. Für die Nutzer und Nutzerinnen umgesetzt wird die ‚geographische Zeitreise‘ in Form einer Graphdatenbank.

14 Online unter: gov.genealogy.net (30.5.2021).

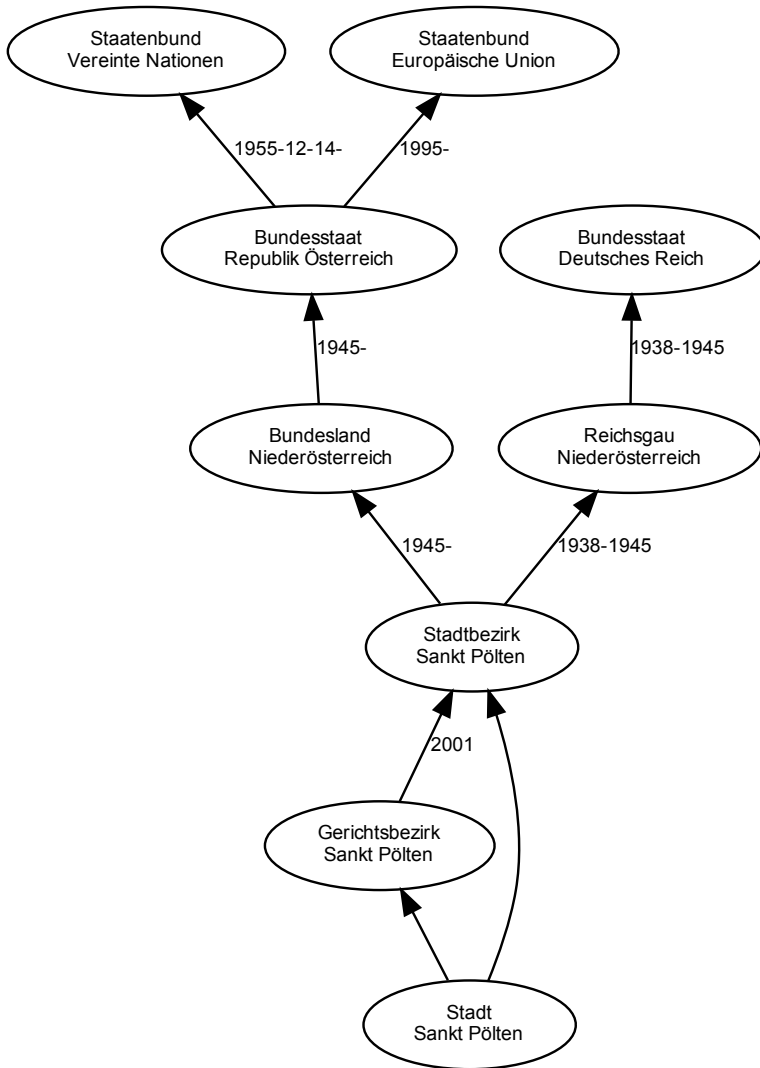
15 Genauer zur Abdeckung siehe: <http://wiki-de.genealogy.net/GOV#Umfang> (30.5.2021).

16 Die Daten für die USA und Australien stammen von der National Imagery and Mapping Agency (NIMA).

17 Jesper Zedlitz/Norbert Luttenberger, A Survey on Modelling Historical Administrative Information on the Semantic Web, in: International Journal on Advances in Internet Technology 7/3&4 (2014), 218–231.

18 Die Übersicht der Ortstypen findet sich unter: <http://gov.genealogy.net/type/list> (30.5.2021).

Abbildung 1: Visualisierung der Zugehörigkeiten und historischen Entwicklung der Stadt St. Pölten im GOV



Quelle: http://gov.genealogy.net/item/show/object_306817 (30.5.2021)

Zu den Orten selbst sind folgende Informationen im GOV erfasst: geographische Lokalisation (Koordinaten und Position auf Karte)¹⁹, Eigenschaften (z.B. Bevölkerungsentwicklung,

¹⁹ Die Angabe der Geokoordinaten im GOV orientiert sich am World Geodetic System 84 (WGS84) oder am Europäischen Terrestrischen Referenzsystem 1989 (ETRS89). Vgl. Department of Defense, World Geodetic System 1984. Its definition and relationships with local geodetic systems, Rockville, MD, 1991.

Postleitzahl), deutschsprachige, fremdsprachige und/oder frühere Namen, Zugehörigkeit zu politischen, kirchlichen und/oder gerichtlichen Verwaltungsbezirken.²⁰

Um die wissenschaftliche Nachnutzung und die Überprüfbarkeit der Informationen zu garantieren, sind für die meisten im Rahmen des GOV erhobenen Daten und Metadaten die entsprechenden Quelleninformationen hinterlegt.²¹ Zudem wurden die Orte auch in anderen Datenbanken referenziert: So finden sich Links zu GeoNames und dem Amtlichen Ortschaften-/Ortsverzeichnis Bayern²². Aktuell weist das GOV 770.000 Verbindungen zu GeoNames auf. Schwieriger gestaltet sich die Anbindung an die GND, da dort bisher nicht eindeutig zwischen Verwaltungsobjekt (z.B. Gemeinde) und Siedlungsobjekt (z.B. Dorf) unterschieden wird. In der GND werden Orte als Geographika und als Körperschaften abgebildet, deren Geokoordinaten GeoNames entstammen.²³ Bei der Erfassung von Geographika werden hier Name, Quelle, die (administrative) Zugehörigkeit ohne zeitliche Dimension und Namensänderungen berücksichtigt, durch das Fehlen von ‚Zeitstempeln‘ werden teilweise neue Entitäten mit neuen ‚Charakteristiken‘ geschaffen; gerade dieser Modellierung setzt das GOV eine belastbare Alternative entgegen. Aus diesem Grund existieren aktuell nur 9.200 Verbindungen zur GND.

Dariah Geo-Browser (DGB)

Der Dariah Geo-Browser²⁴ besteht aus zwei Komponenten: einem Browser zur Daten-Visualisierung und einem Datasheet Editor zur Dateneingabe, Abfrage und Recherche. Er wird seit 2011 basierend auf den Entwicklungen von europeana4D und GeoTemCo²⁵ weiterentwickelt. Dem Selbstverständnis des Projekts zufolge²⁶ ermöglicht er „eine vergleichende Visualisierung mehrerer Anfragen und unterstützt die Darstellung von Daten und deren Visualisierung in einer Korrelation von geographischen Raumverhältnissen zu entsprechenden Zeitpunkten und -abläufen.“ Es soll also ermöglicht werden, zeitliche und räumliche Bezüge von Daten- und Quellensammlungen herzustellen und zu analysieren.

Listen von Ortsnamen werden im DARIAH-DE Getty Thesaurus Service nachgeschlagen. Somit stellt der DGB eine Benutzeroberfläche für Recherche und Visualisierung hauptsächlich für den Getty Thesaurus of Geographic Names (TGN) dar,²⁷ ermöglicht in der Einzelabfrage aber zugleich die Nutzung von GeoNames und OpenStreetMap.

Der TGN²⁸ soll das Katalogisieren, Recherchieren und Finden von Informationen über visuelle Werke und verwandte Themen unterstützen. Als Thesaurus ist TGN ein semantisches

20 Die Minimalanforderung an einen Datensatz für die Aufnahme in das GOV besteht in der Angabe des (Orts-) Namens und seiner Koordinaten bzw. einer administrativen Zuordnung.

21 Zu jedem Eintrag im GOV können die Daten in rdf abgerufen werden. Beispiel-Link: <http://gov.genealogy.net/semanticWeb/about/KIRGTEJO31VX> (Kirchspiel Telgte), 1.3.2022.

22 <https://www.bavarikon.de/places?lang=de> (1.3.2022).

23 Vgl. dazu: <https://wiki.dnb.de/pages/viewpage.action?pageId=127172808> (14.5.2021).

24 Vgl. <https://geobrowser.de.dariah.eu/> (14.5.2021).

25 <http://www.informatik.uni-leipzig.de:8080/geotemco/> (14.5.2021).

26 Vgl. <https://geobrowser.de.dariah.eu/> (14.5.2021); <https://de.dariah.eu/geobrowser> (21.2.2022), mit dem folgenden Zitat.

27 Vgl. <https://wiki.de.dariah.eu/display/publicde/Technische+Dokumentationen> (14.5.2021).

28 Vgl. <https://www.getty.edu/research/tools/vocabularies/tgn/> (14.5.2021).

Netzwerk, das sich auf geographische Orte konzentriert, sowohl als administrative Einheiten als auch auf deren physische Merkmale. Inhaltlich fokussiert der TGN auf historische Daten und auf solche aktuellen Informationen, die das Katalogisieren und Auffinden von visuellen Werken erleichtern. Als Projekt des *Getty Research Institute* ist der TGN Teil der Getty Vocabulary Program (GVP) Ontology, ist also darauf ausgelegt, mit anderen Getty-Vokabularen wie dem Art & Architecture Thesaurus (AAT) kompatibel zu sein.²⁹ Da aus der Kunst und Architektur stammende Elemente dieser Ontologie sich auf Definitionen von Domänen und Bereichen von Objekt-Eigenschaften niederschlagen, ist die Logik des TGN-Modells nicht unmittelbar für eine eher akteursbezogene geschichtswissenschaftliche Perspektive zugänglich. Der Thesaurus selbst hat die folgenden drei Beziehungen: Äquivalenzbeziehungen (Synonyme), hierarchische Beziehungen (Ganzes/Teil) und assoziative Beziehungen (verschiedene Arten von anderen Beziehungen). Er ist polyhierarchisch und mehrsprachig, enthält multikulturelle, historische, archivische und andere Arten von Namen und Informationen zu einem Ort. Jeder Datensatz im TGN wird durch dauerhafte, eindeutige numerische Bezeichner identifiziert. Im Dezember 2020 umfasste TGN 2,5 Millionen Datensätze und 4,1 Millionen Ortsnamen.³⁰

Anwendungsspektrum und Praxistest

Bevor mit der Identifikation der Ortsnamen begonnen werden kann, ist eine vorbereitende Bearbeitung erforderlich. Hierbei wurde besonderer Wert darauf gelegt, dass ein für beide Tools gleichermaßen anwendbarer Datensatz entsteht. Die im GOV angebotene Funktionalität einer Suche nach Schreibvarianten wird im Folgenden also nicht in die vergleichende Diskussion von Trefferzahlen einbezogen. Alle Schritte der Vorverarbeitung können mithilfe einer gängigen Tabellenkalkulation erfolgen.

Die amerikanischen Ortsverzeichnungen umfassen 832 Einträge zu Briefen oder Briefserien, diese liegen in der Form „(Kürzel für den Staat)-(Kürzel für den Bundesstaat), Ortsname“ vor, also z.B. „US-MI, Madison“. Die Zuordnung der Bundesstaaten erfolgte dabei aufgrund der im Projekt aus dem Briefmaterial eruierten Informationen. Nach der Vorverarbeitung (Entfernung von Duplikaten, Vereinheitlichung unterschiedlicher Schreibweisen, Trennung der Einzelangaben, Auflösung der in der Datenbank verwendeten Kürzel für die Bundesstaaten etc.) ergab sich eine bereinigte Datenmenge von nur noch 395 Einträgen.

Die deutschen Ortsnamen weisen eine ähnliche Struktur auf. Hier verzeichnet sind ebenfalls „Staat-Kürzel Bundesland, Ortsname“, also z.B. „D-NW, Unna“. Die Bereinigung der 399 Rohdatensätze folgte den gleichen Prinzipien, hier blieben 300 Ortsnamen für die Zuordnung erhalten, von denen 76 durch die Angabe eines Kreises oder anderer Zusätze näher bestimmt werden. Sowohl für die amerikanischen wie für die deutschen Angaben gilt, dass mit dem Vorkommen unauffindbarer, weil fehlerhaft verzeichneter oder dem falschen Bundesstaat oder Bundesland zugeordneter Orte gerechnet werden muss.

29 Vgl. Getty Vocabularies: Linked Open Data – Semantic Representation, <http://vocab.getty.edu/doc/gvp-lod.pdf> (14.5.2021).

30 Vgl. <https://www.getty.edu/research/tools/vocabularies/tgn/faq.html#number> (17.5.2021).

Sowohl im GOV als auch im DGB besteht die Möglichkeit einer Einzel- und einer Listenabfrage. Die Einzelabfrage ist deutlich zeitaufwendiger. Aber gleichzeitig kann durch die individuelle Entscheidung für jeden Datenpunkt die Qualität der Ergebnisse gesichert werden, und es muss keine nachträgliche händische Zuordnung bei Mehrfachtreffern oder Unsicherheiten erfolgen. Nach der Bereinigung der Daten bestand ein zweiter Arbeitsschritt darin, alle Orte einzeln im GOV und im DGB zu suchen. Gearbeitet wurde bei der Einzelabfrage im GOV mit der erweiterten Suchoption unter Nutzung der Typgruppe „Wohnplatz“. Im DGB kann bei der Einzelabfrage zwischen TGN, GeoNames und OpenStreetMaps gewählt werden. Gearbeitet wurde zunächst mit TGN, da auch die Listenabfrage des Data-sheet Editors ausschließlich TGN nutzt. Für Datenpunkte, zu denen TGN keinen Eintrag verzeichnete, wurden erneute Abfragen bei GeoNames und OpenStreetMaps durchgeführt. Eine Vorauswahl von Wohnplätzen, Siedlungen, Administration etc. ist innerhalb der Dariah-Einzelabfrage nicht möglich, so dass in der Ergebnisübersicht unter anderem auch Flüsse, Bezirke und Berge aufgeführt werden; die Selektion aus dieser Treffermenge muss individuell erfolgen. Nach der Identifikation des Ortes wurden die Geodaten und die GOV-ID bzw. die im DGB bereitgestellte Getty-ID händisch in die Datentabelle übernommen.

Ein wesentlicher Unterschied zwischen GOV und DGB besteht darin, dass das GOV über eine automatische Ortsnamenerkennung (Toponymresolution) verfügt, die regelbasiert den Informationsgehalt der Datenbank auch über den einzelnen Ortseintrag hinaus nutzt. Sie wurde von Dennis Sen als informatische Studienabschlussarbeit entwickelt³¹ und steht zurzeit als Beta-Version zur Verfügung.³² Der Informationsgehalt des GOV wird für mehrere kombinierbare Heuristiken genutzt: hierarchische Beziehungen (Telgte, Kreis Warendorf), Unschärfe von Benennungen (Telchde), Typ der Verwaltungs- oder Siedlungseinheit (Telgte, Gogericht), geographische Nähe (Telgte bei Münster).

Für die Nutzung der Toponymresolution im GOV benötigt man eine csv-Datei im Format UTF8 unter Nutzung des Feldtrenners Tabulator. Toponyme selbst können Wildcards („\?“ „*“) und explizite Markierungen von Unschärfe („\“)³³ enthalten. Es besteht die Option, die Abfrage auf bestimmte Typklassen (z.B. Wohnplatz, Gericht, Kirche) oder beliebige Kombinationen von ihnen einzugrenzen. Ebenfalls möglich sind Sprachbeschränkungen und die Eingrenzung der Suche auf bestimmte Regionen (z.B. durch Angabe des Kreises) oder auf individuell definierte Zeiträume. Die Einbeziehung zusätzlicher Informationen zu den Ortsnamen erhöht die Trefferqualität der Toponymresolution im GOV. Gleichzeitig werden mögliche Fehler in der Datenaufnahme offensichtlich, denn stimmen Ortsangabe und Bundesland nicht überein, erfolgt keine Zuweisung. Diese fehlenden Identifikationen können dann gezielt an die Quellen zurückgegeben und einer neuerlichen Kontrolle unterzogen werden. Nach der Eingrenzung der Suchparameter und dem Hochladen der csv-Datei beginnt die Abfrage und gibt die Ergebnisse ebenfalls als csv-Datei aus. In dieser werden die Anzahl der Nennungen im Gesamtdatensatz (Frequency), Längen- und Breitengrad und die GOV-ID – jeweils unter Berücksichtigung der angegebenen Einschränkungen – verzeichnet.

31 Dennis Sen, *Toponym Resolution on Historical Serial Sources*, Master's Thesis, CAU Kiel 2016, DOI: 10.5281/zenodo.5703327 (9.12.2021).

32 <https://gov-dev.genealogy.net/toponymResolution> (9.12.2021).

33 Sen, *Toponym Resolution*, 38.

Für die Listenabfrage im Dariah-Browser von Geo-Koordinaten und die Identifizierung wird der Datasheet Editor (DGB) genutzt. Dieser ist primär für eine Darstellung der Orte im Geo-Browser erforderlich und ausgelegt.³⁴ Für die Auswahl wird im Datasheet Editor die Funktion „exact-match“ des DARIAH-DE Normdatendienstes genutzt,³⁵ der auf den TGN zugreift. Daher können an dieser Stelle die Geo-Koordinaten semi-automatisch ermittelt werden. Der Dariah-Browser unterscheidet sich von der GOV-Toponymresolution dadurch, dass die genannten Heuristiken (Hierarchie, Nähe, Schreibweisen und Ortstyp) nicht genutzt werden; Schreibvariationen müssen daher zwingend schon vor der Abfrage bereinigt werden, während die anderen Heuristiken in händische Einzelentscheidungen einfließen können. Damit Forschungsdaten im DGB summarisch identifiziert und visualisiert werden können, müssen diese im CSV-Format vorliegen. Mit dem Datasheet Editor³⁶ können die zu analysierenden Daten eingegeben oder eine bereits vorliegende csv-Datei importiert werden. Dafür ist es erforderlich, die Ortsnamen in eine Spalte mit dem Titel „Address“ einzutragen, zusätzlich ist die Anlage einer Spalte „Longitude“, „Latitude“ und „Getty-ID“ bereits in der csv-Datei sinnvoll.³⁷ Wichtig an dieser Stelle ist die Berücksichtigung der Groß- und Kleinschreibung. Für den DGB muss die csv-Datei im UTF8-Format und mit Kommata als Trennzeichen gespeichert sein.³⁸ Werden bei der TGN-Abfrage mehrere Einträge gefunden, übernimmt der Datasheet Editor jeweils die Koordinaten des ersten Treffers in die Tabelle. Falls der ermittelte Ort nicht der gewünschte Ort ist, gibt es die Option per Dropdown-Menü den ‚richtigen‘ Eintrag händisch aus der Treffermenge zu wählen. Die georeferenzierten Daten können abschließend als csv-Datei heruntergeladen werden, parallel erfolgt die Speicherung der erstellten Abfragen als csv-Datei und Datasheet über das jeweilige Nutzerkonto (Own-Storage). Für jedes Projekt vergibt der DGB zudem eine interne Projekt-ID.³⁹

Ergebnisse

Einzelabfrage

Tabelle 1 stellt die Trefferanzahl der beiden Tools bei händischen Abfragen dar. Die Ergebnisse wurden in drei Kategorien eingeteilt: Entweder wurde ein Ort richtig zugeordnet („gefunden“), oder es gab keinen passenden Treffer („nicht gefunden“), oder es wurden mehrere mögliche Orte identifiziert („uneindeutig“). In beiden Teildatenbeständen war ein knappes

34 Dafür erforderlich ist jedoch eine Anmeldung bzw. Registrierung, diese erfolgt mit einem DARIAH-Account oder mit jedem anderen eduGAIN Föderations-Account.

35 Vgl. <https://wiki.de.dariah.eu/display/publicde/DARIAH-DE+Normdatendienste> (25.5.2021); alternativ könnte auch eine Autocomplete-Funktion genutzt werden.

36 Vgl. <http://geobrowser.de.dariah.eu/edit> (25.5.2021).

37 Diese können auch nachträglich mit Hilfe des Datasheet Editors generiert werden.

38 Die Beschreibung der für die csv-Datei erforderlichen Variablen kann durch Klicken des „How to fill the table“-Buttons rechts vom Datasheet im Datasheet Editor geöffnet werden. Es müssen nicht alle Felder ausgefüllt werden, um ein Datasheet zu erstellen.

39 Die DGB Projekt-ID scheint nicht statisch zu sein, denn mit der Umstellung auf die Version 3.6 hat sich das Format der IDs der Datensätze leicht geändert. Alle ‚alten‘ Datensatzreferenzen blieben erhalten, sind aber nur noch lesbar und nicht mehr editierbar, erst nach einer Migration können sie erneut bearbeitet werden (<https://geobrowser.de.dariah.eu/beta/doc/de/datasheet.html>, 25.5.2021).

Fünftel der Orte weder mit dem einen noch dem anderen Tool eindeutig aufzufinden. Bei den übrigen gelang die Suche mit GOV besser als mit dem Dariah-Browser. Vom in GOV nicht auffindbaren Rest konnte ein Teil im Dariah-Browser gefunden werden und umgekehrt, sodass bei der händischen Arbeit sicherlich zu empfehlen ist, beide Tools nacheinander zu nutzen.

Tabelle 1: Vergleich der Suchergebnisse im GOV und im DGB bei händischer Suche, US- und deutscher Datensatz (in Prozent)

GOV	DGB			Gesamt
	gefunden	uneindeutig	nicht gefunden	
US-Daten (N=395)				
gefunden	64	11	0	75
uneindeutig	4	3	0	7
nicht gefunden	2	7	9	18
Gesamt	70	21	10	100
Deutsche Daten (N=300)				
gefunden	56	16	7	79
uneindeutig	5	6	3	14
nicht gefunden	0	1	6	7
Gesamt	61	22	16	100

Im GOV dauert eine Suchabfrage maximal 10 Millisekunden. Die eventuell daran anschließende Selektion der Treffer kann unter freier Verwendung der Filterfunktion übersichtlich und zügig erfolgen.

Von den 395 amerikanischen Ortsnamen konnten bei der händischen Einzelabfrage im GOV 296 (75 Prozent) eindeutig zugeordnet werden. 28 Datenpunkte (7 Prozent) konnten im jeweils verzeichneten Bundesstaat mehrfach nachgewiesen werden; hier bräuchte es weiterführende Informationen (z.B. Angabe des County). Datenpunkte, die ganz ohne Treffer blieben, wurden in der Kategorie „nicht gefunden“ zusammen mit solchen Einträgen gezählt, wo sich die betreffenden Orte zwar in den USA befinden, aber Bundesstaat und Ortsangabe nicht passend sind. In diesen Fällen ohne Zuordnung im angegebenen Bundesstaat erfolgt eine erneute Überprüfung anhand des Quellenmaterials.

Im DGB wurden für die amerikanischen Daten je Eintrag in der Einzelabfrage durchschnittlich 107 Treffer erzielt, was vor allem in der fehlenden Kombinationsmöglichkeit von Namen und Bundesstaat begründet ist. Dadurch erhöht sich die Treffermenge exponentiell, was sich auch in der Dauer der Suchanfrage widerspiegelt: Im Schnitt dauerte die Ergebnisausgabe je Datenpunkt etwa 30 Sekunden. Alle Treffer werden als Dropdown-Liste angezeigt, ohne dass weitere Filtermöglichkeiten eine Einschränkung der Treffermenge bzw. Begrenzung z.B. auf einzelne Staaten zulassen. Das manuelle Durchsuchen der Treffermenge gestaltet sich, etwa im Fall von San Francisco (647 Treffer), mühsam und zeitaufwändig. Erschwerend bei der händischen Selektion der Treffer wirkt sich die Tatsache aus, dass eine interne Sortierung der Treffermenge beim DGB nicht zweifelsfrei zu erkennen ist.

Abbildung 2a: Ergebnisse des GOV und des DGB in der Einzelabfrage amerikanischer Orte

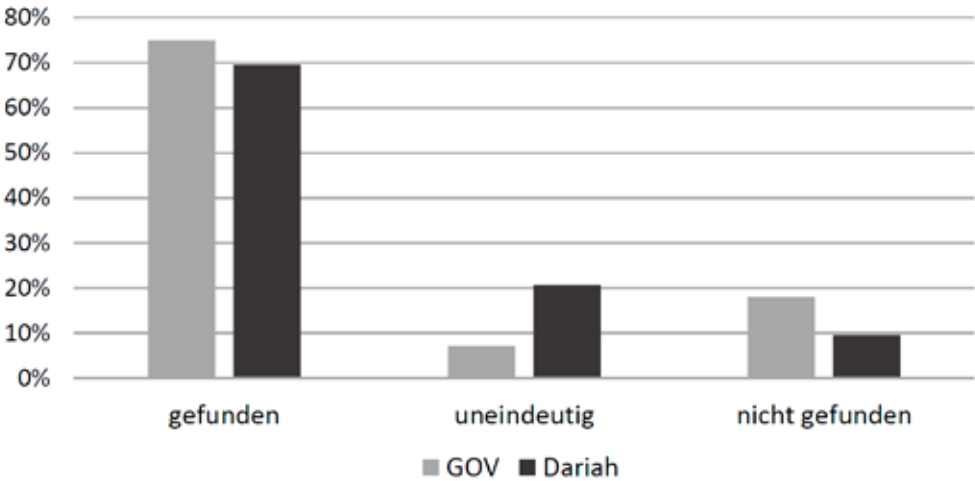
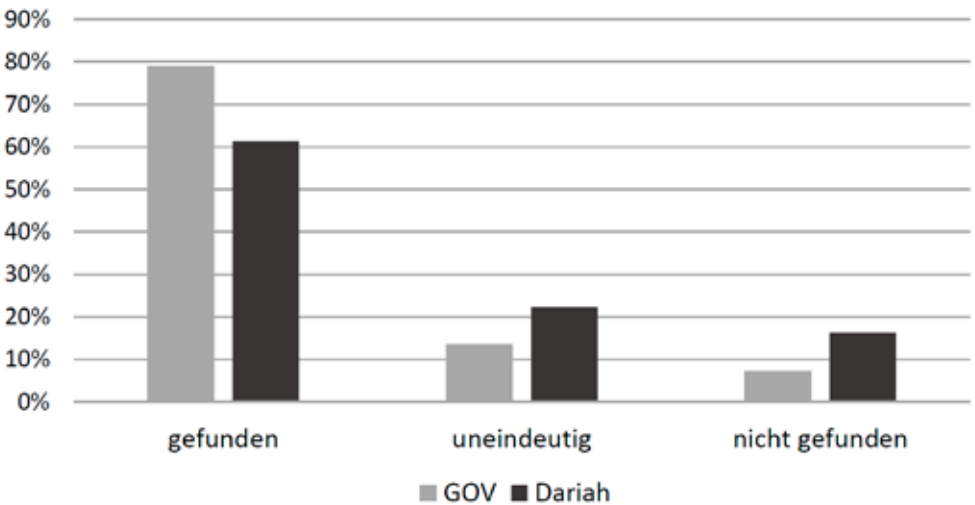


Abbildung 2b: Ergebnisse des GOV und des DGB in der Einzelabfrage deutscher Orte



Quelle: Eigene Darstellung auf Basis von Tabelle 1.

Eine eindeutige Identifizierung nach händischer Selektion der Gesamtreffermenge erreicht der DGB bei den amerikanischen Ortsnamen, ausschließlich unter Einbeziehung des TGN, für 275 Datenpunkte (70 Prozent); nur 10 Prozent (38 Datenpunkte) können keinen Eintrag vorweisen. Überprüft man die fehlenden Einträge mit einem der beiden anderen Gazetteers, verringert sich die Anzahl der fehlenden Einträge kaum; für nur vier Datenpunkte bietet OpenStreetMaps zusätzliche Treffer an, während GeoNames die Trefferanzahl nicht erhöht.

Bei den deutschen Ortsnamen erweist sich die Möglichkeit der phonetischen Suche,⁴⁰ wie sie das GOV anbietet, als besonders hilfreich, da die Schreibung der Umlaute durchaus variiert und in der häufig vorgefundenen Umschreibung (ae, oe, ue) zunächst keine Zuordnung erfolgt. Gleiches gilt etwa für Schreibungen mit „K“ oder „C“.

Von den 300 deutschen Ortsnamen können mit Hilfe eines analogen Vorgehens wie bei den amerikanischen Ortsnamen mit dem GOV 237 (79 Prozent) eindeutig identifiziert werden, lediglich bei 22 Datenpunkten (7 Prozent) gelang in der Einzelabfrage keine Zuordnung. Für einige der deutschen Ortsnamen besteht die Schwierigkeit des mehrfachen Vorhandenseins eines Ortsnamens in einem Bundesland (41 Datenpunkte, 14 Prozent). Da die administrativen Zugehörigkeiten in allen Fällen vorhanden sind, würde die Zuweisung eines Kreises diese Zahl wahrscheinlich deutlich verringern.

Bei Dariah ist in der Einzelabfrage die durchschnittliche Anzahl der Treffer bei den deutschen Orten deutlich geringer als bei den amerikanischen. Im Schnitt werden je Abfrage 20 Treffer erzielt. Auf Grund der deutlich geringeren Treffermenge je Datenpunkt bei den deutschen Orten verkürzt sich die Abfragezeit je Eintrag auf ca. vier Sekunden. Erschwerend für eine eindeutige Zuordnung im DGB ist die Tatsache, dass im DGB-Datenbestand bei den deutschen Ortsnamen nicht analog wie bei den amerikanischen Ortsnamen eine Zuordnung zu Verwaltungseinheiten vorgenommen wurde, sodass jeweils nur Land und Bundesland aufgeführt werden. Selbst eine nähere Bestimmung des Ortes durch Angabe des Kreises könnte infolgedessen keine genauere Zuordnung ermöglichen. Auf Grund der fehlenden Metadaten liegt der Anteil der eindeutig identifizierten Ortsnamen im DGB deutlich unter dem Ergebnis des GOV.

Listenabfrage

Tabelle 2 stuft analog zu Tabelle 1 die Ergebnisse der automatischen Tools ein. Dabei kommt auch die Kategorie „Fehler“ vor – hiermit ist gemeint, dass die Toponymresolution bzw. die Suche über den Datasheet Editor ein eindeutiges Ergebnis liefert, das aber vom Ergebnis der händischen Recherche abweicht. Schaut man zunächst auf die Fehlerraten, sieht man am amerikanischen Material unbefriedigende Ergebnisse sowohl für das GOV als auch für den DGB, allerdings aus unterschiedlichen Gründen. Das GOV findet automatisch über die Hälfte der Orte, allerdings bei einer Fehlerrate von einem Zehntel. Der Dariah-Browser bietet dagegen nur wenige eindeutige Ergebnisse, und von diesen ist dann auch noch ein Fünftel falsch. Wer beide automatische Tools kombiniert, hat kaum einen Vorteil gegenüber der alleinigen Nutzung der GOV-Toponymresolution: Von 163 im GOV nur uneindeutig oder gar nicht gefundenen Orten entdeckt die DGB-Automatik nur acht richtig und fünf falsch. Empfehlenswert für US-Daten dürfte es sein, in einem ersten Lauf die Toponymresolution einzusetzen, die Ergebnisse aber zu überprüfen und mit beiden Tools händisch weiterzuarbeiten.

Für Deutschland sind die Ergebnisse beim GOV anders und deutlich besser. 13 Fehler (etwa 4 Prozent) bei 300 Fällen wird man möglicherweise tolerieren können – nicht zu Ver-

40 Grundlegend basiert die phonetische Suche in GOV auf: https://kowa.hs-augsburg.de/beispiele/db/postgresql/volltext/phonet/dict_phonet/phonet/ (29.4.2021).

öffentlichungszwecken, aber doch für die Erstellung interner Findmittel. Mit 250 Treffern von 300 nimmt die Toponymresolution den Nutzern und Nutzerinnen durchaus spürbar Arbeit ab. Die automatischen Funktionen des Dariah-Browsers führen zwar auch nur noch zu wenigen Fehlern, ergeben aber deutlich weniger eindeutige Treffer. Für die Funktionsweise des Dariah-Browsers kennzeichnend sind die vielen uneindeutigen Ergebnisse: Man bekommt fast in drei Vierteln der Fälle Vorschlagslisten, aus denen man von Hand den richtigen Ort herausuchen muss. Der Dariah-Browser erweist sich damit als geeignetes Tool für ein semiautomatisches Vorgehen für den Fall, dass man sich von vornherein für ein händisches Durchkorrigieren entscheidet.

Tabelle 2: Anzahl von Suchergebnissen im GOV und im DGB bei automatisierter Suche, US- und deutscher Datensatz (in Prozent)

GOV	DGB				Gesamt
	gefunden	uneindeutig	nicht gefunden	Fehler	
US-Daten (N=395)					
gefunden	9	34	4	1	49
uneindeutig	1	13	5	1	19
nicht gefunden	1	16	4	1	22
Fehler	1	9	1	0	10
Gesamt	12	73	13	2	100
Deutsche Daten (N=300)					
gefunden	32	32	17	2	83
uneindeutig	1	1	1	0	4
nicht gefunden	0	6	2	0	8
Fehler	0	3	1	0	4
Gesamt	34	42	21	3	100

Die praktische Funktionalität der beiden Systeme unterscheidet sich also. Bei der Listenabfrage im DGB wird die vorhandene Tabelle automatisch mit Koordinaten und Getty-IDs befüllt. Unterhalb der tabellarischen Ansicht werden die Zuordnungen unter der Überschrift „Place selection“ detailliert aufgelistet. Hier offenbart sich zunächst die Schwierigkeit, dass die Sortierung der Einträge nicht der zuvor hochgeladenen tabellarischen Übersicht entspricht; interne IDs werden nicht mitgeführt und auch die zusätzliche Verzeichnung von Bundesländern oder Staaten erscheint nicht. Die Darstellung der Ergebnisse beschränkt sich somit auf den Ortsnamen (in der Spalte „Address“) und in einem Dropdown-Menü die Auflistung der potentiellen Zuordnungen.

Bei einem großen Teil der Zuweisungen von Ortsnamen im DGB kommen mehrere Möglichkeiten in Betracht. Da der DGB bei der Auswahl der vorgeschlagenen Zuordnung einem Zufallsprinzip folgt, stimmt nur ein kleinerer Teil davon mit dem gewünschten Ort überein. Diese Vorschläge sollten also nicht ohne Überprüfung genutzt werden; sie werden hier zu den uneindeutigen Treffern gezählt. Eine annähernd zweifelsfreie Identifikation innerhalb der

Abbildung 3a: Ergebnisse der Listenabfrage amerikanischer Orte im GOV und im DGB

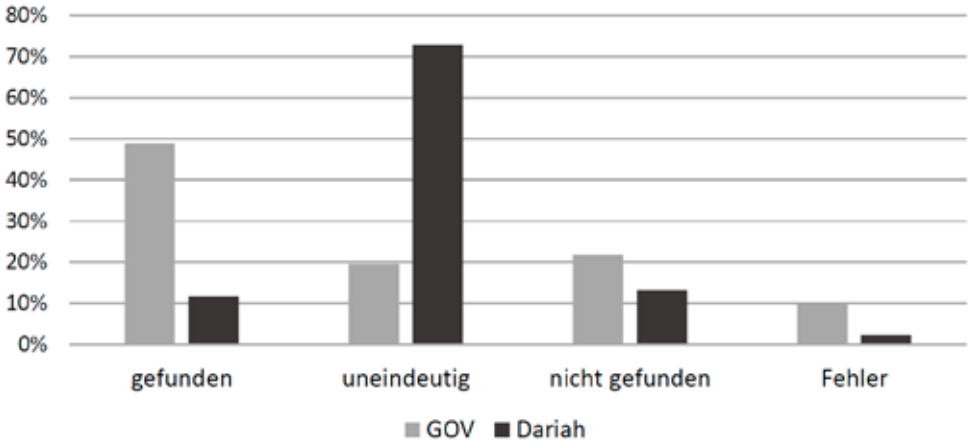
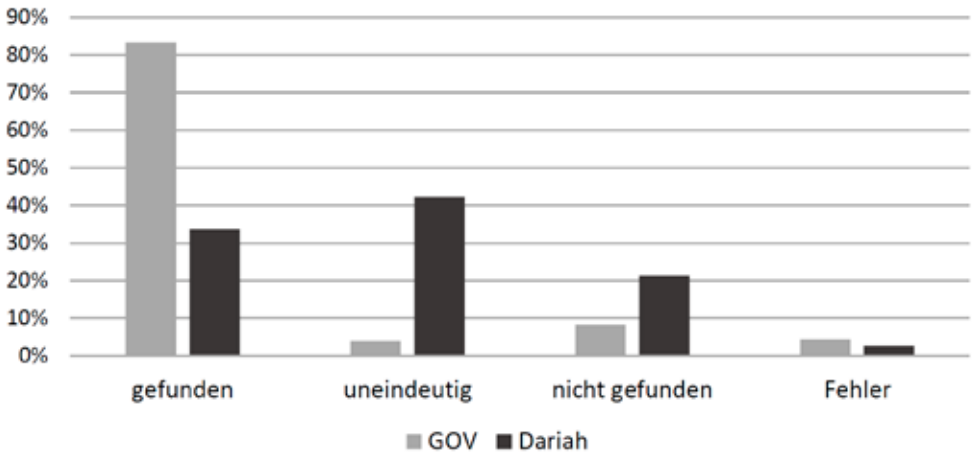


Abbildung 3b: Ergebnisse der Listenabfrage deutscher Orte im GOV und im DGB



Quelle: Eigene Darstellung auf Basis von Tabelle 2.

Listenabfrage des DGB erscheint vor diesem Hintergrund nur möglich, wenn es weltweit nur einen Ort mit entsprechendem Namen und derselben Schreibweise im gesamten TGN gibt. Auch die Anzahl der möglichen Zuordnungen ist nicht in der Übersichtsdarstellung ersichtlich. Dies wäre theoretisch von Vorteil, da Datenpunkte mit nur einem Treffer unter Umständen nicht nochmals überprüft werden müssten. Jedoch weisen jeweils etwa ein Viertel dieser Zuordnungen im DGB ebenfalls Unsicherheiten auf, da sie beispielsweise auf einen Bahnhof oder Kreis und nicht einen Wohn- oder Siedlungsplatz verweisen.

Die Listenabfrage des GOV (Toponymresolution) bietet demgegenüber deutliche Vorteile: Durch die Einbeziehung zusätzlicher Informationen (z.B. Bundesland, Kreis, „in der Nähe“,

„bei“ etc.) können sehr viel mehr eindeutige Zuweisungen erzielt werden, die keiner händischen Nachkontrolle bedürfen. Problematisch an der Listenabfrage im GOV sind die aktuell noch erheblichen Laufzeiten (mehrere Minuten bei ca. 300 Datensätzen). Betont werden muss, dass aufgrund der Datenlage die Ergebnisse für die deutschen Orte deutlich belastbarer sind als bei den amerikanischen Ortsnamen.

Visualisierung und Datenexport

Im GOV erfolgt innerhalb der Einzelabfrage zunächst eine Visualisierung der möglichen Treffer auf einer OpenStreetMap-Karte. In der Detailansicht eines selektierten Ortes können die Nutzer*innen dann zwischen unterschiedlichem Kartenmaterial wählen und dieses auch exportieren. Zur Verfügung stehen Google Earth als kml-Datei (Download) sowie die Visualisierung in GoogleMaps, OpenStreetMaps, wikimapia und MapQuest (Browser); ebenfalls erfolgen kann eine Verknüpfung mit dem Virtuellen Kartenforum 2.0, betrieben von der SLUB Dresden.⁴¹ Registrierten Nutzer*innen bietet sich zudem die Möglichkeit der Nutzung von historischen Messtischblättern. Die Registrierung ist kostenlos möglich; es genügt die Angabe einer E-Mail-Adresse. Die in der Listenabfrage generierte csv-Datei kann problemlos in unterschiedliche Formate (GIS-Systeme) übernommen werden.

Eine Visualisierung von Einzelabfragen im DGB ist nur möglich, wenn der Treffer in das DataSheet übernommen wird. Bei der summarischen Abfrage kann nach Kontrolle und Zuweisung im DGB die Gesamttabelle inklusive Koordinaten und Getty-ID als csv-Datei heruntergeladen werden, wie auch die Visualisierung im Geo-Browser selbst erfolgen. Es besteht aber nicht die Möglichkeit, in der Einzelabfrage identifizierte Ortsinformationen automatisch in das DataSheet zu überführen. Somit muss zuerst die Tabelle gefüllt sein, an der dann händisch Änderungen und Zuweisungen vorgenommen werden können. Der umgekehrte Weg einer Identifikation und daran anschließenden Aufnahme der Informationen ist nicht vorgesehen.

Für die Visualisierung der im DataSheet Editor generierten Ergebnisse steht im DGB unterschiedliches Kartenmaterial zur Verfügung. Als Grundkarte nutzt auch der DGB OpenStreetMap, daneben stehen zeitlich differenzierte Umrisskarten zur Auswahl. Diese historischen Karten wurden im Rahmen des Projekts europeana4d entwickelt bzw. als Public Domain-Material genutzt und über den GeoServer von Dariah zur Verfügung gestellt.⁴² Zusätzlich legt der DGB einen Schwerpunkt auf die Kombination von Raum und Zeit. In diesem Zusammenhang bietet die Visualisierung den Vorteil, dass die Datensätze mit Zeitpunkten bzw. Zeiträumen versehen und in der graphischen Umsetzung mit Filter- und Selektionsoptionen verknüpft werden können. Des Weiteren steht die Möglichkeit einer Animation der geographischen Einträge im Zeitverlauf zur Verfügung. Vorteilhaft an dieser Stelle ist die Möglichkeit, auch Datensätze aus mehreren Dateien gleichzeitig zu visualisieren. Ebenso können bereits bestehende Dateien (csv; kml/kmz) direkt in den Browser geladen werden. Im DGB visualisierte Daten können als kml-Dateien, die Visualisierung selbst hingegen nicht heruntergeladen werden.

41 <https://kartenforum.slub-dresden.de/vkviewer/?welcomepage> (7.6.2021).

42 <https://ref.de.dariah.eu/geoserver/web/> (25.5.2021).

Zusammenfassung

Um mit einem der beiden vorgestellten Tools arbeiten zu können, müssen die Daten eine gewisse Struktur aufweisen, die es dann aber ermöglicht, die gleichen Daten ohne großen Aufwand in unterschiedlichen Verfahren abfragen zu können. Die Datenaufnahme sollte einheitlich erfolgen – für die Nachnutzung hilfreich ist das Vermeiden von Abkürzungen und Sonderzeichen, die getrennte Aufnahme von Zusatzinformationen und die strikte Trennung von Datenpunkten (keine Mehrfachbezeichnungen in einer Zelle). Die Datenstruktur selbst orientiert sich damit (auch in einer Excel-Tabelle) an der Logik relationaler Datenbanken, das heißt sie sollte redundanzfrei und konsistent sein.

In der Listenabfrage des GOV ist es möglich, hierarchische Informationen (z.B. administrative Strukturen) zu berücksichtigen, ergänzt um die Option einer phonetischen Suche bzw. einer Suche nach ähnlichen Schreibweisen (sowohl in der Einzel- als auch der Listenabfrage). Allein diese Faktoren machen die Suche und damit auch die Abfrageergebnisse deutlich belastbarer als dies im DGB der Fall ist. Hier scheint das GOV optimal an die Bedürfnisse von Historikern und Historikerinnen angepasst, die häufig vor dem Problem unsicherer Schreibweisen, Namensvarianten oder fehlerhafter Transkriptionen stehen. Die Einbeziehung zusätzlicher Informationen zu Ortsnamen und die bereits vor der Abfrage bestehende Möglichkeit, Filterfunktionen zu nutzen, erhöht die Trefferqualität der Abfrage im GOV. Werden beispielsweise die deutschen Ortsnamen ohne die Angabe von Bundesländern abgefragt, erhält man für 63 Prozent eine richtige Zuweisung, wird hingegen die Bundeslandangabe integriert, erhöht sich der Anteil korrekter Ergebnisse auf 79 Prozent. Gleichzeitig werden mögliche Fehler in der Datenaufnahme offensichtlich, denn stimmen Ortsangabe und Bundesland nicht überein, erfolgt keine Zuweisung in anderen Bundesländern.

Der DGB erweist sich in der Listenabfrage und der Zuweisung der dort erzielten Ergebnisse als problematisch. Als besonders hemmend muss die Vorab-Zuordnung der Ergebnisse auf den zufällig ersten Treffer gesehen werden. Dasselbe gilt für die fehlende Option, die Abfrage um Metadaten zu ergänzen oder eine vorangehende Eingrenzung der Suchparameter vorzunehmen. Die rein zufällige Zuweisung von Treffern in der Listenabfrage zwingt die Nutzer*innen sämtliche Ergebnisse einzeln zu überprüfen. In diesem Fall bietet die Listenabfrage keine Vorteile gegenüber der Einzelrecherche, doch nur an dieser Stelle besteht die Möglichkeit den Datensatz zu exportieren, um ihn in GIS-Systemen weiter zu visualisieren. Somit sind die Nutzenden gezwungen, den Datasheet Editor zu verwenden. Der Mehraufwand der nachträglichen Ergebniskontrolle wiegt den Mehrwert einer mit Koordinaten und Getty-ID gefüllten csv-Datei dabei kaum auf. In Einzelfällen ist es zudem erforderlich, direkt auf den TGN zurückzugreifen, um Unsicherheiten in der Zuweisung ausschließen zu können. Auch an dieser Stelle bietet das GOV eine größere Nutzerorientierung, denn nach der Listenabfrage können über die GOV-ID direkt weiterführende Informationen zu einem Datensatz abgefragt werden.

Abschließend kann konstatiert werden, dass beide Tools in der Einzelabfrage ähnliche Resultate erzielen, was auf eine ähnliche Raumabdeckung hindeutet. Im Fall von relativ wenigen Datensätzen sind sie zudem in ihrer Anwendbarkeit vergleichbar. Forschenden, die mit größeren Datensammlungen arbeiten wollen, bietet die Listenabfrage des GOV hingegen deutlich bessere Nutzungsoptionen. Filterfunktionen, die Einbeziehung von Metadaten und die detaillierte Verzeichnung der raum-zeitlichen Entwicklung von Ortsnamen machen

das GOV zu einer geeigneten Rechercheplattform für historische Ortsnamen, während insbesondere der zeitliche Korrektur- und Kontrollaufwand bei der Listenabfrage im DGB dessen praktische Anwendbarkeit stark einschränkt. Selbst wenn im Fall einer händischen Korrektur die Ergebnisse des DGB ebenfalls belastbar erscheinen, weist er zusätzlich die Schwäche auf, dass die sich wandelnden (administrativen) Beziehungen von Siedlungen nur eingeschränkt und ausschließlich über den Umweg TGN sichtbar gemacht werden können, während die im GOV hinterlegten Metadaten deutlich über andere Normdatensätze hinausgehen und vor dem Hintergrund der intensiven Weiterentwicklung und Betreuung durch Bürgerwissenschaftler*innen das GOV zu einem attraktiven Werkzeug für Historiker*innen werden lassen.