

Tags: Archiv Big Data, Digitalisierung, Erschliessung, Informationskompetenz, Sammlung, Zugang,

Normdaten im Archiv: Ein Dreistufenmodell

Längst sind es nicht mehr nur Bibliotheken, die das Potenzial von Normdaten für ihre Erschliessungsarbeiten nutzen. Auch Archive sind auf den Geschmack gekommen. Der folgende Artikel präsentiert ein dreistufiges Modell, das mögliche Anwendungen von Normdaten in Archiven beschreibt.

Normdaten liegen im Trend. Längst sind es nicht mehr nur Bibliotheken, die ihr Potenzial für ihre Erschliessungsarbeiten nutzen.¹ In den vergangenen Jahren wurden mögliche Nutzungs- und Weiterentwicklungsszenarien breit diskutiert, besonders dort, wo ein Bedürfnis nach eindeutig identifizier- und referenzierbaren Entitäten besteht. Heute sind die Ansprüche, die an Normdaten gestellt werden, vielfältig. So sind sie etwa ein wesentliches Element der Weiterentwicklung von Archivinformationssystemen in Richtung Linked-Data-Anwendungen, die mit dem neuen Standard Records in Contexts (RiC) verstärkt vorangetrieben werden soll. Gleichzeitig hat der Austausch über Nutzungsmöglichkeiten einen anwendungsorientierten Dialog zwischen verschiedenen Interessensgruppen aus dem GLAM-Bereich und der (geisteswissenschaftlichen) Forschung in Gang gesetzt.¹

Im Rahmen dieses Artikels sollen mögliche Anwendungsszenarien von Normdaten in Archiven skizziert werden. Wir gehen dabei davon aus, dass der Einsatz von Normdaten in ganz unterschiedlicher Form stattfinden kann und immer auch von den Schwerpunkten und Ressourcen der jeweiligen Archive abhängt. Deswegen schlagen wir ein dreistufiges Modell vor, das die Rolle von Normdaten mit einem Fokus auf die Bereiche Redaktion, Anwendung und Nachnutzung beschreibt. Die erste Stufe umfasst Einstiegsmöglichkeiten in die Nutzung von Normdaten, die zweite Stufe beinhaltet erweiterte Anwendungsmöglichkeiten zur Nutzung des Vernetzungspotenzials von Normdaten, auf der dritten Stufe wird die Erstellung und Vernetzung von Normdaten in aktuelle technologische Weiterentwicklungen eingebettet. Im Mittelpunkt steht der Einsatz von Normdaten, um Entitäten wie Personen oder Orte eindeutig zu referenzieren. Rahmenkonzepte, die beispielsweise im Kontext von RiC relevant sind, werden nicht diskutiert. Für die angeführten Beispiele greifen wir auf Aktivitäten innerhalb der Sektion Sammlungen und Archive der ETH-Bibliothek zurück. Einen Schwerpunkt bildet dabei der Einsatz der Gemeinsamen Normdatei (GND).

Der Einstieg in die Nutzung von Normdaten kann einfach sein: Es genügt – auch in einer Erschliessung, die auf dem ISAD(G)-Standard basiert – eine klar definierte Form, beispielsweise ein Feld oder eine Auszeichnung, die eine bestimmte Entität einem eindeutigen Identifikator aus einem Normdatensatz zuweist. Auf diese Weise kann unabhängig von veränderbaren Eigenschaften, beispielsweise der Schreibweise eines Namens, zweierlei erreicht werden: Eine Verknüpfung mit einer Informationsquelle, die Informationen zu bekannten Schreibvarianten enthält und eine numerische Referenz, die völlig unabhängig von der Schreibweise eine Entität eindeutig referenziert. Bereits hier kommen wichtige Vorteile der Nutzung von Normdaten zum Tragen. Zum einen liegt die Verantwortung für die Pflege der eigentlichen Normdaten bei einer externen Stelle, die auf die Redaktion und Dokumentation solcher Daten spezialisiert ist. Zum anderen ist bereits mit dieser einfachen Massnahme die Vernetzung mit weiteren Datenquellen möglich, wenn diese ebenfalls auf die gleichen Identifikatoren verweisen (z. B.: Person x = GND-ID y = Person z). Eine einfache, in der Praxis etablierte Nachnutzungsmöglichkeit für diese Form der Vernetzung sind sogenannte BEACON-Dateien. Wikipedia, Projektseite Wikipedia:BEACON, online: <https://de.wikipedia.org/wiki/...>¹ Der in der Schweiz etablierte Vernetzungsservice metagrid.ch funktioniert zwar grundsätzlich auch für Datensätze, für die keine GND-ID existiert, aber auch hier kann die GND genutzt werden, um die Verknüpfung zwischen den verschiedenen Partnerprojekten zu vereinfachen.

Ausgehend von diesen Basis-Szenarien sind die Übergänge zu weiterführenden Nutzungsformen fließend. Mit der Einführung des GND-Webformulars ist es für kleinere Projekte einfacher geworden, sich punktuell an der GND zu beteiligen und Daten beizusteuern. [2](#) Das gleiche gilt für die Anwendung auf den eigenen Daten auch für die Vernetzung mit weiteren externen Datenquellen. Neben zusätzlichen Normdaten, für Orte beispielsweise Geonames, kann insbesondere Wikidata eine sinnvolle Ergänzung sein, auch wenn es sich dabei um einen kollaborativ gepflegten Datenpool und nicht um eine zentral redigierte Normdatei handelt. Diese zusätzlichen Auszeichnungen vereinfachen und erweitern auch die Möglichkeiten für die Datennachnutzung und -integration. So nutzt die ETH-Bibliothek beispielsweise sowohl GND-IDs als auch die Q-IDs aus Wikidata als Identifikatoren, um externe Daten in ihre Swisscovery-View zu integrieren und den Nutzenden so zusätzliche Informationen anbieten zu können.

Um die Nutzung von Normdaten schliesslich langfristig auszubauen, kann es für Institutionen sinnvoll sein, sich noch stärker in die Weiterentwicklung der genutzten Normdatei einzubringen. Für die GND kann dies bedeuten, dass eine Institution eine eigene GND-Redaktion aufbaut und so weitergehende Möglichkeiten zur Mitwirkung an der GND erhält. [3.1](#) Je umfangreicher Normdaten innerhalb eines Projekts oder einer Institution eingesetzt werden, umso wichtiger wird oft auch die Frage der Automatisierung von Prozessen. Für die Nutzung von Normdaten gibt es hierfür zahlreiche Möglichkeiten. Einzelne Entitäten können im Rahmen der Erschliessung mit Lookup-Schnittstellen überprüft werden. Für grössere Datenmengen bietet sich als Einsteigertool OpenRefine mit seinen integrierten Reconciling Services an. [4.1](#)

Einen Schritt weiter geht das sogenannte Named Entity Linking, bei dem mit Hilfe von Machine-Learning-Verfahren versucht wird, eine automatisierte Verknüpfung von Entitäten mit externen (Norm-)Datenquellen zu erreichen. Spätestens hier lässt sich dann auch nur noch schwer zwischen Anwendung und Nachnutzung von Normdaten unterscheiden. Die normdatenbasierte Verknüpfung unterschiedlicher Datenquellen, beispielsweise zwischen GND und Wikidata, kann genutzt werden, um zu den jeweiligen Entitäten einen gemeinsamen Pool an Informationen zusammenzutragen, der wiederum als Ausgangsmaterial für Natural-Language-Processing-Verfahren dienen kann. Aufbauend auf das Projekt E-Periodica Next Level Access, bei dem Named Entity Recognition und Linking Verfahren auf den gesamten Bestand von E-Periodica (ca. 10 Millionen Seiten) angewendet wurden, wird aktuell in der Sektion Sammlungen und Archive der ETH-Bibliothek das Potenzial dieser NLP-Verfahren getestet und ausgelotet.

Das hier präsentierte dreistufige Modell macht aber hoffentlich eins deutlich: Vom unbestrittenen und wachsenden Nutzen von Normdaten können Archive mit einer schrittweisen Einführung profitieren, die in der Komplexität je nach Bedarf und Möglichkeiten gesteigert werden kann. Als Einstieg genügt schon ein Feld.

Weiterführende Hinweise & Links zum Thema

Für die Rolle von Normdaten in der bibliothekarischen Bestandserschließung, die in ihren Grundsätzen auch hohe Relevanz für andere Anwendungsmöglichkeiten haben kann: Klaus Haller und Claudia Fabian: Bestandserschließung, in: Die moderne Bibliothek. Ein Kompendium der Bibliotheksverwaltung, hrsg. Von Rudolf Frankenberger und Klaus Haller, München 2004, S. 222?261, online: <https://doi.org/10.1515/978311...> Zum Thema Normdaten insbesondere S. 246?253.

Wikipedia, Projektseite Wikipedia:BEACON, online: <https://de.wikipedia.org/wiki/...>

Deutsche Nationalbibliothek, GND-Webformular, online: <https://www.dnb.de/DE/Professi...>

Bernd Uttenweiler, Discovery and more: Wo sind die Personen zwischen all den Dokumenten?, Präsentationsfolien 111. BiblioCon, Hannover, 23.?26. Mai 2023, online: <https://doi.org/10.3929/ethz-b...>

Gemeinsame Normdatei, Über die GND, online: <https://gnd.network/Webs/gnd/D...>

OpenRefine, online: <https://openrefine.org/>.

1

2 Deutsche Nationalbibliothek, GND-Webformular, online: <https://www.dnb.de/DE/Professi...>

3 Gemeinsame Normdatei, Über die GND, online:

https://gnd.network/Webs/gnd/DE/UeberGND/Organisation/organisation_node.html

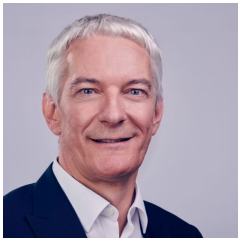
4 OpenRefine, online: <https://openrefine.org/>.



Christiane Sibille

Dr. Christiane Sibille hat Geschichte und Musikwissenschaft in Heidelberg studiert und arbeitet heute an den (digitalen) Schnittstellen zwischen Forschung, Archiven und Bibliotheken. Zu ihren beruflichen Stationen zählen die Universitäten Heidelberg und Basel, die Zürcher Hochschule der Künste sowie die Forschungsstelle Diplomatische Dokumente der Schweiz (Dodis) und das Projekt Metagrid.ch.

Seit Juni 2021 ist Christiane Sibille in der Sektion Sammlungen und Archive der ETH-Bibliothek Verantwortliche für die Fachstelle «Digital Scholarship Services». Ausserdem unterrichtet sie regelmässig zu digitalen Themen an der Universität Basel und ist Präsidentin des Vereins «Geschichte und Informatik».



Michael Gasser

Michael Gasser (*1968) studierte Geschichte und Anglistik an den Universitäten Basel und Edinburgh. Der Einstieg ins Archiv- und Bibliothekswesen erfolgte im Anschluss an mehrere Jahre in der Privatwirtschaft. Seit 2006 arbeitet Michael Gasser an der ETH-Bibliothek. Dort leitet er seit 2020 die Sektion Sammlungen und Archive. Die fortschreitende Digitalisierung, Öffnung und Vernetzung von Inhalten und Daten der unterschiedlichen Archiv- und Sammlungsbestände ist ihm ein zentrales Anliegen.

Abstract

Deutsch

An der VSA-Fachtagung 2023 «Archives in Contexts» in Sarnen hielten Dr. Christiane Sibille und Michael Gasser einen viel beachteten Einführungsvortrag zur Nutzung von Normdaten. Für den arbedo-Beitrag legen sie ihren Fokus auf ein Drei-Stufen-Modell für Archive.

Français

Lors de la journée professionnelle 2023 de l'AAS "Archives in Contexts" à Sarnen, Dr Christiane Sibille et Michael Gasser ont tenu une conférence d'introduction très remarquée sur l'utilisation des données normatives. Pour la contribution arbedo, ils mettent l'accent sur un modèle à trois niveaux pour les archives.