

Barbara K. Fischer, Jürgen Kett, Sarah Hartmann, Mathias Manecke

Eine Stimme im Orchester der Öffnung der GND

Ein Werkstattbericht aus der Arbeitsstelle für Standardisierung an der Deutschen Nationalbibliothek

Dieser Artikel wurde bereits im BUB 02-03/2021 in einer älteren Fassung veröffentlicht.

Wenn Menschen Themen mit Verve und Ausdauer diskutieren, dann ist das in der Regel ein Zeichen für Engagement und Verbundenheit. Das Thema »Öffnung der GND« hat diese guten Qualitäten. Es betrifft und bewegt viele Menschen. Es wirft Fragen zu dem großen Thema Kooperation auf, im Detail und in ganz unterschiedlichen Kontexten. Im viel beachteten Comic »Sapiens« zur Geschichte der Menschwerdung des Historikers Yuval Noah Harari findet sich das folgende Zitat: »Was uns wirklich unterscheidet, ist die Art und Weise, wie wir in großer Zahl kooperieren.«¹ Es beschreibt auch unsere Arbeit in der Arbeitsstelle für Standardisierung (AFS) an der Deutschen Nationalbibliothek. Denn im Kern besteht unsere Arbeit im Organisieren von Kooperation. Unsere Aufgabe ist es, über nationale und disziplinäre Grenzen hinweg, die Erschließung von Wissensressourcen zu ermöglichen. Wir organisieren Zusammenarbeit durch die Förderung der Konsensfindung zu Standards, mit denen wir letztlich die Welt beschreiben und dabei für alle gleichermaßen verständlich bleiben wollen. Mittels der Standards wird in der Gemeinschaft der deutschsprachigen Bibliotheken definiert, wie Publikationen präziser als mittels der natürlichen Sprache beschrieben werden sollen, damit andere darauf eindeutig Bezug nehmen können. Dabei kommt die Gemeinsamen Normdatei (GND) zum Einsatz. Harari zielt auf das große Ganze des Menschseins und wie dieses sich beispielsweise von dem Sein der Schimpansen unterscheidet. Die Arbeit des Katalogisierens, der Beschreibung von Medien nach

den Regeln der Formal- und Inhaltserschließung, ist weit entfernt von den Herausforderungen der Homo Sapiens in der Steinzeit. Und doch ist sie gewissermaßen nur ein anderer Abschnitt auf demselben Strahlenbündel. In Folge der »kognitiven Revolution«² von damals stehen wir vor der Herausforderung der digitalen Transformation von heute. Auch diese werden wir eben durch unsere Fähigkeit zur Kooperation meistern. Das ist das, was wir können.

Im Zuge der Öffnung der GND für Gemeinschaften außerhalb des Bibliothekswesens ist in den letzten Jahren eines immer deutlicher geworden: Die GND ist weitaus mehr als nur eine Sammlung von neun Millionen Normdatensätzen zu Personen, Orten, Körperschaften, Konferenzen, Werken und Allgemeinbegriffen.³ Sie bezeichnet außerdem eine Organisationsstruktur, die die Verfasstheit ihrer derzeitigen Anwender*innen reflektiert. Sie rekurriert auf ein bestimmtes Datenmodell, das sich an den Bedarfen ihrer Anwender*innen orientiert. Sie kennt spezielle Regeln und kann als ein Spezialwerkzeug in einer abgegrenzten technischen Umgebung gesehen werden, das von den Erfordernissen einer Bibliotheksgemeinschaft geprägt wird. Aber die neuen Anwendungsgruppen sind anders organisiert. Sie haben andere Datenmodelle. Sie erschließen die Gegenstände ihres Interesses nach anderen Regeln und nutzen eine andere technische Infrastruktur. Trotzdem sind sie an der Nutzung der GND-Normdaten sehr interessiert. Sie wollen nicht nur die Identifikatoren in ihrer Erschließungsarbeit verwenden, sondern auch neue GND-Datensätze anlegen können, wenn sie dafür einen Bedarf sehen. Sie wollen ein aktiver Teil der GND-Gemeinschaft werden. Dabei müssen wir gemeinsam genau überlegen, was und wieviel wir verändern können, ohne die GND im Kern zu beschädigen. Denn ihre verlässliche Qualität wollen alle erhalten wissen. Unse-

re Aufgabe ist es auch jetzt wieder, die Zusammenarbeit im Sinne einer kollektiven Intentionalität zu organisieren.

Ein Instrument der Öffnung

Die Öffnung der GND ist ein Konzert eines ganzen Orchesters von Akteur*innen und Maßnahmen. Ein Instrument in diesem Orchester, ein Ansatzpunkt für eine behutsame Anpassung, ist die technische Umgebung, in der die GND verankert ist. Nicht der Austausch der bisherigen technischen Infrastruktur, sondern die Idee, eine parallele Struktur anzubieten, hat unsere Aufmerksamkeit auf die Datenbanksoftware Wikibase⁴ gelenkt. Wikibase ist eine Open-Source-Software der Wikimedia Foundation. Diese hat schon die Software Mediawiki entwickelt, mit der weltweit Millionen Wikis betrieben werden. Das bekannteste Wiki ist Wikipedia aus dem Hause Wikimedia. Um Wikipedia zu verbessern, ersann man vor acht Jahren das Projekt Wikidata. Eine Datenbank für strukturierte Daten, mit denen man – für Menschen wie Maschinen gleichermaßen lesbar – die Welt beschreiben kann. Die Software hinter Wikidata ist Wikibase. Wikibase bringt einige Eigenschaften mit, die Kooperationen in großer Zahl vereinfachen sollen:

- Sie bietet webbasierte Zugänge.
- Sie erlaubt das parallele kollaborative Arbeiten.
- Sie protokolliert automatisch Versionsgeschichte und Editor*in.
- Sie bietet zu jedem Datensatz eine eigene Diskussionsseite an.
- Sie ist auf multilinguale Anwendungscommunities ausgerichtet.
- Sie bietet ein einfaches und flexibles (allerdings auch limitiertes) Datenmodell.
- Sie ist einfach und intuitiv in der Bedienung und Erfassung neuer Inhalte.

Diese Eigenschaften haben wir 2019 in der Deutschen Nationalbibliothek intensiv untersucht und in unserer Evaluation⁵ gemeinsam mit Wikimedia Deutschland zusammengetragen. Auch aktuelle Schwächen des Systems sowie Entwicklungspotenziale wurden in diesem Rahmen erörtert. Noch er-

füllt das System längst nicht alle Anforderungen für den Einsatz als ideales Redaktionssystem und Drehscheibe für Kultureinrichtungen. Hierzu hat es sich noch zu wenig von seinem Ursprung als Software für Wikidata emanzipiert. Dennoch sahen wir die Grundvoraussetzungen für einen produktiven Einsatz im Kontext der AfS gegeben. Wichtig ist weniger der jetzige Stand des Produktes, sondern das Potenzial, das in seiner Weiterentwicklung und dem Aufbau einer breiten Community im Kultursektor liegt.

Im Jahr 2020 haben wir einerseits überlegt, wie wir Wikibase möglichst effektiv für die Öffnung der GND nutzbar machen können und haben andererseits die Voraussetzungen geschaffen, um unsere Pläne möglichst effizient umsetzen zu können. Wir haben entschieden, auf drei Ebenen aktiv zu werden. Wir wollen:

- der GND als Normdatei in einer Wikibase-Datenbank ein zweites Zuhause einrichten. Dort können die neuen Anwendungsgemeinschaften leichter und unabhängiger von den bestehenden Strukturen Vorschläge für neue GND-Datensätze machen und ihre Daten leichter mit der GND abgleichen, um Dubletten in der GND zu vermeiden.
- den Schulterschluss zu Wikimedia sowie anderen Institutionen suchen, die auch Wikibase nutzen wollen, um gemeinsam an der Verbesserung der Software zu arbeiten, um letztlich ein Ökosystem für Kultur- und Forschungsdaten aufzubauen.
- und drittens möchten wir die Regelwerke selbst, die die GND und unsere Katalogisierungsarbeit bestimmen, neu ordnen, zugänglicher und anpassungsfähiger für Veränderungen machen.

Der Zweitwohnsitz der GND

In der Bibliothekswelt dient die GND wie schon ihre vier Vorläufer seit langem als Werkzeug zur Referenzierung und Rationalisierung. Sie ist eingebunden in bestimmte Regelwerke und proprietäre Softwarestrukturen, die jedoch für Anwender*innen außerhalb des Bibliothekswesens relativ unzugänglich sind. Wir denken, dass wir auf Grundlage

von Wikibase einigen dieser Zielgruppen die Mitarbeit an der GND erleichtern können.

Daher werden wir 2021 alle bislang vorliegenden GND-Datensätze mit ihren entsprechenden Verknüpfungen in eine Wikibase-Instanz importieren. Das klingt nach einer einfachen Aufgabe. Doch bislang sind die Importschnittstellen von Wikibase noch sehr an den Bedarfen von Wikidata ausgerichtet. Daher haben wir uns professionelle Unterstützung von einem Wikibase-Spezialisten gesucht, der uns bei dem Transfer der Datenbankstruktur, bei dem Datenimport und bei der Erstellung von anwendungsfreundlichen Eingabemasken als Dienstleister zur Seite steht. Als nächstes werden wir dann erfahrene und neue GND-Anwender*innen einladen, die Erfassungs- und Suchprozesse in der neuen Umgebung zu testen, damit wir sie weiter verbessern können.

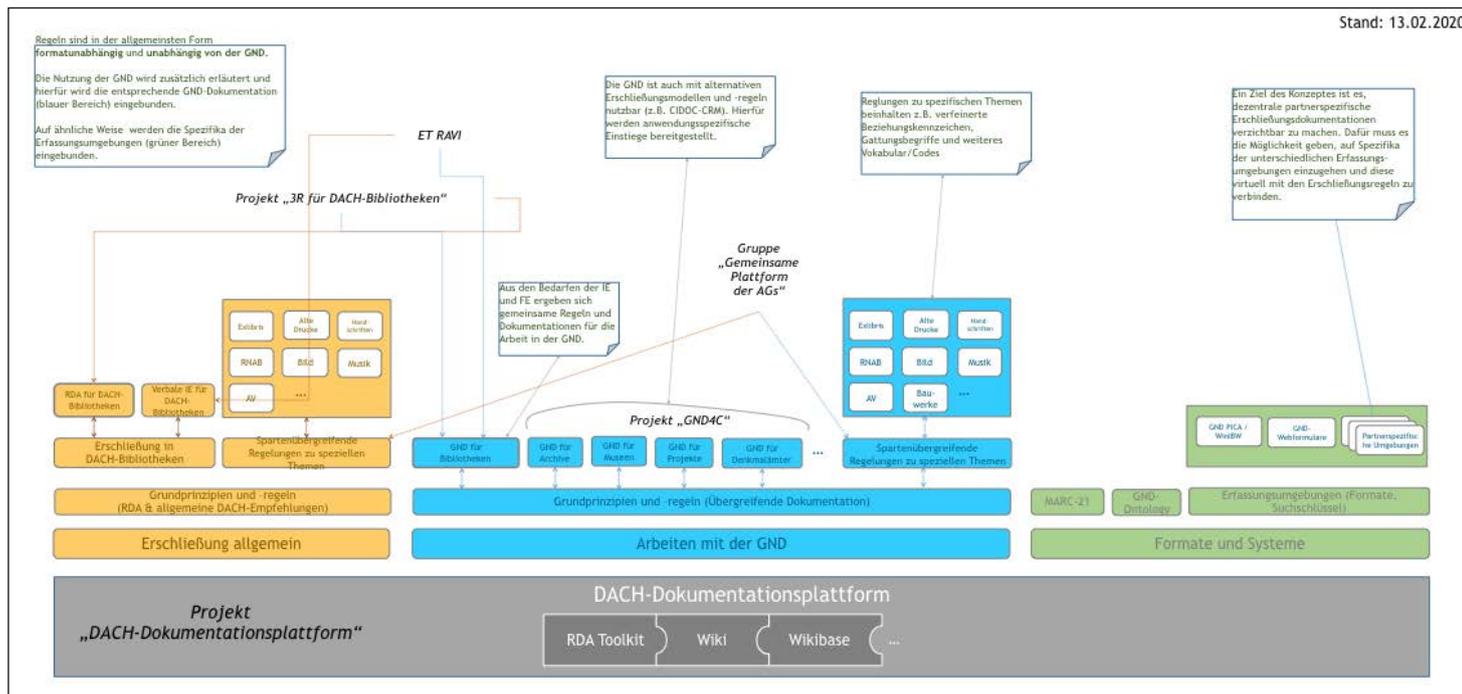
In der zweiten Jahreshälfte 2021 planen wir einen technischen Workflow zur Synchronisierung der GND-Wikibase-Instanz mit dem CBS-System.⁶ Geplant ist, dass neue oder Anwender*innen ohne WinIBW⁷-Zugang, ihre Daten als Vorschlag in die Wikibase-Instanz eingeben.

Ein langfristiges Ziel ist es, eine leicht bedienbare und unterstützende Erfassungsumgebung für die GND anzubieten. Erste Schritte in diese Richtung sind die gegenüber den bibliothekarischen Erfassungssystemen deutlich leichter zugänglichen GND-Webformulare.⁸ Über diese lassen sich bislang Personen und Körperschaften erfassen. Der Ansatz ist perspektivisch allerdings nicht flexibel genug. Neben den zwei genannten Satzarten der GND gibt es vier weitere. Diese sechs Satzarten vereinen circa 50 Entitätencodes⁹ mit jeweils spezifischen Eigenschaften, mittels derer die jeweiligen Entitäten durch GND-Datensätze erfasst werden können. Benötigt würde dafür eine dynamische Eingabemaske, die sich je nach Wahl des Entitätstyps oder Nutzungskontextes anpasst, benötigte und typische Eingabelemente anbietet, auf sinnvolle Eingaben hinweist und den Nutzenden so durch den Eingabeprozess geleitet. Ob Wikibase hierfür mittelfristig eine Grundlage bietet, muss sich noch zeigen. Aktuell bietet Wikibase solche Features noch nicht. Eine Anpassung der generischen Eingabeoberfläche ist »ab Werk« nicht vorgesehen. Auch gibt es keine Möglichkeit, das Angebot auf sinnvolle Elemente

oder Werte einzuschränken. Die Nutzer*innen sind stets mit dem vollen Umfang der Eigenschaften und Werte konfrontiert und erhalten keine Entscheidungshilfe. Ein Ziel für 2021 ist es, zu ermitteln, ob dies über die Entwicklung einer Wikibase-Erweiterung ermöglicht werden kann und welche Änderungen an Wikibase durch Wikimedia durchgeführt werden müssten, um die Erstellung anpassbarer, assistierender Eingabemasken besser zu unterstützen.

Das WikiLibrary-Manifest

Das zweite Feld, auf dem wir in 2021 aktiv sind, ist der Schulterschluss mit Wikimedia Deutschland und anderen Institutionen zur Verbesserung von Wikibase als technische Infrastruktur. In immer mehr Kontexten wird bei der Bereitstellung von Daten auf die so genannten FAIR Data Prinzipien (Findability, Accessibility, Interoperability und Reusability)¹⁰ geachtet. Daten sollen stärker miteinander vernetzt werden, um insgesamt leichter neues Wissen generieren zu können. Dies gilt insbesondere für Daten, die mit öffentlichen Mitteln erstellt wurden. Das ist für viele Einrichtungen eine große Herausforderung. Es stellt sich die Frage, ob man seine Daten in ein Sammelbecken für strukturierte Daten oder Datenportale geben soll. Ist man bereit, sich auf all die Konsequenzen einzulassen, wie zum Beispiel den Verlust von Gestaltungsmöglichkeiten des Datenmodells, der Erfassungsregeln oder der Qualitätssicherung? Oder setzt man auf Insellösungen und muss stattdessen mit einer eingeschränkten Sichtbarkeit und Nachnutzung der Daten leben? Mit der Öffnung der GND wollen wir eine Alternative schaffen. Wir setzen uns dafür ein, ein zuverlässiges, maschinenlesbares und gemeinschaftlich unterhaltenes Linked-Open-Data-Netzwerk für Kunst, Kultur und Wissenschaft als tragfähige Grundlage für FAIRes Wissen zu schaffen. Nicht die zentrale Plattform favorisieren wir, sondern ein offenes Netzwerk miteinander verknüpfter Datenbanken. Dafür ist ein gemeinsamer organisatorischer Rahmen notwendig. Diesen möchten wir in einem Netzwerk bieten. Ein Netzwerk ist immer nur so gut wie die Partner, die es knüpfen. Gemeinsam mit Wikimedia Deutschland



Schematische Darstellung des Projekts DACH-Dokumentationsplattform mit den Blöcken Erschließung allgemein, Arbeiten mit der GND und Formate und Systeme

hat die Deutsche Nationalbibliothek daher das WikiLibrary-Manifest¹¹ veröffentlicht. Knapp vierzig Einrichtungen sind bislang der Einladung gefolgt. Das Manifest lädt die unterzeichnenden Institutionen ein, sich auf der Grundlage folgender Leitlinien zu engagieren:

- Die Förderung freier Lizenzen für Daten und deren Softwareumgebung.
- Die Schaffung von Freiräumen, in denen vielfältige Communitys wachsen und gedeihen können.
- Das Angebot strukturierter Daten auf der Grundlage der FAIR-Data-Prinzipien, um Daten auf transparente Weise in Information umzuwandeln und so Wissen zu schaffen, das FAIR ist.
- Die Förderung gemeinsamer Mindeststandards.
- Die Schaffung offener Organisationsstrukturen und deren Einbettung in bestehende Systeme.
- Die Bereitstellung von Ressourcen, um für alle, die beitragen wollen und sich aktiv um Daten und Wissen bemühen, zugängliche und anwendungsfreundliche Benutzungsoberflächen zu schaffen.
- Die Förderung der Datenkompetenz im digitalen Wandel.

Mindestens ebenso bedeutungsvoll ist im Verbund mit Wikimedia Deutschland die gemeinschaftliche Umsetzung von konkreten Maßnahmen durch alle Unterzeichnenden. Es geht um die Förderung von Wikibase als eine vielversprechende technische Infrastruktur zur Speicherung, Bearbeitung und zum Austausch von Daten auf der Grundlage der FAIR-Data-Prinzipien. Wir möchten Wikibase zu einer anwendungsfreundlichen Referenzdatenbanksoftware für Datahubs formen, um so das angestrebte Datenökosystem zu fördern. Dazu laden wir weitere Institutionen aus der Bibliothekswelt, aus allen Bereichen von GLAM (galleries, libraries, archives and museums) und den Geisteswissenschaften ein, mittels Wikibase ein Ökosystem strukturierter Daten zu schaffen, das einem wirklichen Semantic Web für FAIRes Wissen näherkommt.

Die DACH-Dokumentationsplattform¹²

In den letzten Jahren sind die Resultate der Computerlinguistik mit Hilfe der künstlichen Intelligenz immer ausgereifter geworden. Schreibprogramme ziehen den Inhalt aus strukturierten Datenbanken

und bauen gewissen Vorgaben folgend mit den Elementen die Texte. Vor diesem Hintergrund stehen unsere Überlegungen, die Regelwerke für die Formal- und Inhaltserschließung¹³ sowie die Erfassungshilfen für die GND in strukturierter Form in einer Wikibase-Instanz zu erfassen.

Seit Jahrzehnten dokumentieren wir detailliert in elaborierten Texten genaue Anweisungen, wie welches Datenfeld beispielsweise in der GND erfasst werden muss. Ausgangspunkte für diese Textarbeit sind die Regelwerke der Formal- und der Inhaltserschließung, die Erfordernisse und Beschränkungen der jeweils zur Erschließung angewandten Software sowie schließlich auch die Erfordernisse des Datenaustausches. Jedes Mal, wenn an einer Stelle in diesem komplizierten Beziehungsgeflecht ein Detail geändert wird, muss in vielen Texten, die auf den Punkt Bezug nehmen, die Änderung ebenfalls umgesetzt werden. Das bedeutet jedes Mal eine intensive Recherchearbeit durch eine große Zahl von PDF-Seiten. Eine andere Folge dieser Art des Wissensmanagements ist, dass viele Detailinformationen, zum Beispiel wie man ein Datum eingibt, wie man eine Berufsbezeichnung erfasst oder welchen Code man für welches Land verwendet, an unterschiedlichen Stellen wiederholt werden müssen, um unnötiges Blättern zu vermeiden. Im Fall einer Änderung gilt es den Überblick zu haben, wann und wo genau die Änderung noch greift. Das birgt ein gewisses Risiko für Fehler und Unübersichtlichkeit. Sicher senkt es jedoch die Anwendungsfreundlichkeit der Handreichung, denn immer wieder sind Änderungen erforderlich.

Der Grundgedanke ist frappierend einfach. Bleiben wir erst einmal bei der GND selbst. Die Anzahl der Felder, mittels derer man in den Datenformaten Pica oder Marc 21 Entitäten für Normdatensätzen beschreiben kann, ist überschaubar, nämlich circa 300. Diese Datenfelder oder Elemente dienen dazu, Aussagen über Eigenschaften, Relationstypen, Teilbestände oder Entitätencodes zu den jeweiligen zu beschreibenden Entitäten zu machen. Die Datenelemente beinhalten definierende Merkmale und je nach Datenformat unterschiedliche Kodierungen. Legt man alle Elemente in einer entsprechenden Datenbank ab, können die Datenelemente wie in einem Baukasten nach den Regeln der zugrundeliegenden Regelwerke modular zusammengestellt werden.

Die bisherige Erfassungshilfe allein für Personen mit allen dazugehörigen Entitätencodes in der GND umfasst 46 Seiten.¹⁴ Die Elemente, die erfasst werden müssen, sind hingegen wenige. Neben dem Namen, sind es in erster Linie die Lebensdaten, der Beruf und eventuell noch Verknüpfungen zu weiteren Datensätzen, wie Ortsnamen als Geburts- oder Wirkungsort oder ähnliches. Für jeden der Entitätencodes in der Satzart Personen wird jedes Mal erneut beschrieben, wie beispielsweise das Element »Ort« modelliert werden muss. Wären diese Definitionen in einer Datenbank abgelegt, könnte in die Regel einfach das jeweilige Element eingefügt werden. Das heißt, ändert sich die Regel für die Merkmale zur Erfassung einer Gebietskörperschaft,¹⁵ ändert man diese zentral an einer Stelle und alle anderen Orte, an denen dieses Element zum Einsatz kommt, werden ebenfalls aktualisiert. Es ist dasselbe Prinzip wie es auch bei den Normdatensätzen in Bibliothekskatalogen angewandt wird.

Wir haben angefangen, alle Elemente, die in der GND verwendet werden, in einer strukturierten Form zu beschreiben. Dazu übernehmen wir die Vorgaben aus den Regelwerken. Nun wird es die Herausforderung sein, lesbare Fließtexte zu schreiben, in die die Elemente sinnvoll eingebettet werden können. Diese können dann einerseits deutlich konziser als bisher aktualisiert werden und gleichzeitig perspektivisch als Grundlage für die Erstellung der Eingabemasken für die Datenbank mit allen GND-Datensätzen selbst dienen.

Manchmal tut es gut, sich den Sinn und Zweck seiner Arbeit vor Augen zu führen, um motiviert zu bleiben, um den Fokus zu behalten oder um Dritten zu vermitteln, wieso diese Arbeit wichtig ist und Förderung braucht. Mit diesem Werkstattbericht möchten wir Ihnen unsere Arbeit und die Ideen dahinter näherbringen. Es liegt eine spannende Zeit der Pionierarbeit vor uns. Reizvoller wird diese Arbeit noch durch die parallel laufenden Wikibase-Projekte in den neugegründeten Konsortien der Nationalen Forschungsdaten Infrastruktur-Initiative (NFDI) sowie weiterer großer Universal- und Nationalbibliotheken in Europa und Amerika, mit denen wir in engem Austausch stehen.

Auf der digitalen Großveranstaltung GNDCon 2.0¹⁶ im Juni 2021 wird das Projekt zum ersten Mal der Öffentlichkeit präsentiert und erlaubt einen Blick in den Maschinenraum.

Anmerkungen

- 1 Yuval Noah Harari: Sapiens. Der Aufstieg. Graphic Novel. München: C.H. Beck, 2020, S.68.
- 2 Zum Konzept der »Kognitiven Revolution« vgl. Yuval Noah Harari: Ein kurze Geschichte der Menschheit. München: Dt. Verlags-Anstalt, 2013, S.11-100.
- 3 Die Satzart Konferenzen der GND lässt besonders deutlich die Verschränkung der GND mit ihren Anwender*innen der Bibliothekswelt hervortreten. Denn diese Satzart beschreibt eine besondere Art der Herausgeber von Publikationen.
- 4 Link zur Wikibase Website: <<https://wikiba.se/>>
- 5 Link zum Blogbeitrag über die Evaluation: <<https://wiki.dnb.de/pages/viewpage.action?pageId=167019461>>
- 6 CBS: Proprietäre Bibliotheks-Datenbank-Software von OCLC.
- 7 WinIBW: Lizenz gebundene Software zur Eingabe von Daten in die GND.
- 8 Das GND-Webformular bietet sich für Anwender*innen insbesondere aus Kulturinstitutionen an, wie kleineren Bibliotheken, Archive und Museen, die eine geringe Menge an Datensätzen in die GND einbringen bzw. ändern möchten. <https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_Webformular/gnd_webformular.html>
- 9 Details zur Entitätencodierung in der GND <<https://wiki.dnb.de/download/attachments/90411323/entitaetenCodes.pdf>>
- 10 Information zu den Fair-Data-Prinzipien <https://www.forschungsdaten.org/index.php/FAIR_data_principles>
- 11 Unter diesem Link können Sie das Manifest über ein einfaches Formular als Einrichtung mitzeichnen:
<<https://www.wikimedia.de/projects/wikilibrary-manifest/>>
- 12 DACH-Dokumentationsplattform: Die Plattform soll alle Regelwerke zur bibliothekarischen Erschließung und die Erfassungshilfen für die GND im deutschen Sprachraum (Deutschland, Österreich und Schweiz) vereinen.
- 13 Gemeint sind die Regelwerke RDA und RSWK.
- 14 Siehe auch <<https://wiki.dnb.de/pages/viewpage.action?pageId=90411361&preview=/90411361/94831186/EH-P-01.pdf>>
- 15 Eine Gebietskörperschaft ist ein Entitätencode aus der Gruppe der Geografika oder Orte.
- 16 Informationen zur GNDCon 2.0 und Anmeldung über diesen Link: <<https://www.dnb.de/DE/Kulturell/Veranstaltungskalender/Fachveranstaltungen/20210607GNDcon.html>>