

Barbara K. Fischer, Jürgen Kett

# Leipziger Einsichten zur Öffnung der GND

## Bericht vom GND4C-Forum in der Deutschen Nationalbibliothek

Am 6. Juni 2019 fand in der Deutschen Nationalbibliothek in Leipzig das erste Forum zum Projekt »Die Gemeinsame Normdatei für Kulturdaten« – kurz GND4C – statt. Gut sechzig Expertinnen und Experten aus Forschung, Museum, Archiv und Bibliothek waren der Einladung der sechs Partner<sup>1</sup> des DFG geförderten Projektes gefolgt, erste Arbeitsergebnisse in Form von Postern in kleinen Gruppen nach Weltcafé-Manier zu diskutieren. Das Projekt GND4C erarbeitet seit Mai 2018 die konzeptionellen Grundlagen zur Öffnung der Gemeinsamen Normdatei (GND) über Bibliotheken hinaus und adressiert hierzu folgende Ziele:

- eine nachhaltige sparten- und fächerübergreifende Organisation aufzubauen
- Datenmodell und Regeln auf nicht-bibliothekarische Anwendungskontexte hinaus weiterzuentwickeln
- Schnittstellen und Werkzeuge zur Unterstützung nicht-bibliothekarischer Anwendungen bereitzustellen
- Kommunikation mit den verschiedenen Interessengruppen über diverse Kommunikationskanäle zu stärken und das GND-Netzwerk sichtbar zu machen

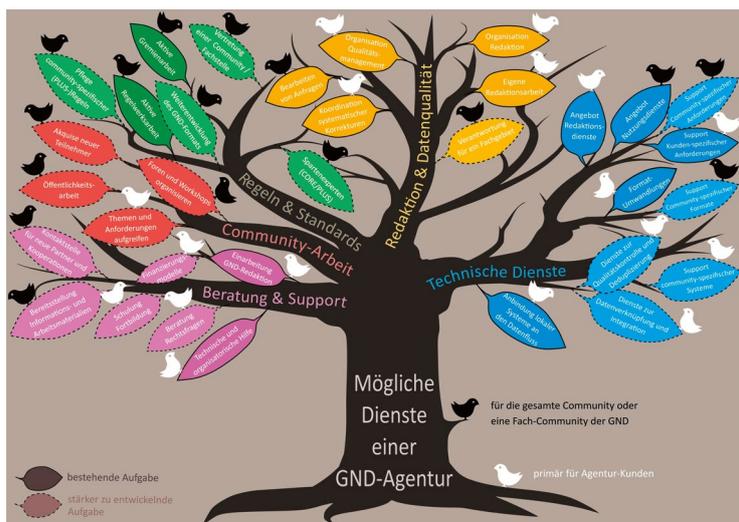
Nach einem Jahr sollten erste Zwischenergebnisse präsentiert und hinterfragt werden. Gemeinsam mit den Partnern nahmen sich die Teilnehmenden einen ganzen Tag Zeit, die vorgelegten Ergebnisse zu reflektieren: Was überzeugt? Was fehlt? Was wird kritisch gesehen? Unterstützt wurden sie dabei engagiert von Mitarbeitenden, Auszubildenden im ersten Lehrjahr sowie den Bundesfreiwilligen der Deutschen Nationalbibliothek.

## Ein Semantic Web der Kultur und Wissenschaft ist nicht kostenfrei

Die erste Erkenntnis des Forums war mehr eine Forderung: Die Anwesenden hoben hervor, dass die digitale Transformation im Kulturbereich eine strukturelle Aufgabe sei. Um Aufgaben wie die Beschreibung von Digitalisaten mit Metadaten und deren Verknüpfung über Normdaten personell leisten zu können, bedarf es einer dauerhaften Finanzierung. Ohne den politischen Willen der fördernden Ministerien in Bund und Ländern ist die spartenübergreifende und tiefgehende Vernetzung der Kulturdaten im deutschsprachigen Raum nicht möglich. Weder kann die Bereitstellung und Pflege der Normdaten allein zentral erfolgen, noch zusätzlich und wie nebenbei von den teilnehmenden Einrichtungen geleistet werden. Befristete Erschließungsprojekte können zwar den Aufbau eines kulturellen Datennetzes entscheidend voranbringen, aber nicht für eine dauerhafte Verlässlichkeit und Qualität bürgen. Hierfür werden nachhaltige Strukturen benötigt. Auf der Wunschliste der Teilnehmer und Teilnehmerinnen stand daher die Forderung, eine auf Dauer angelegte Normdatenarbeit in Förderprogrammen wie den Fachinformationsdiensten für die Wissenschaft (FID)<sup>2</sup> und der Nationalen Forschungsdateninfrastruktur (NFDI) zu verankern.<sup>3</sup> Die zweite übergreifende Einsicht war für manch Teilnehmende ernüchternd: Der Öffnungsprozess der GND ist - insbesondere aufgrund des erwähnten strukturellen Defizits - nicht von heute auf morgen umzusetzen.

Sicher ist, dass die bestehenden Organisationsstrukturen der Gemeinsamen Normdatei erweitert werden müssen. Geplant ist der Aufbau von neuen GND-Agenturen, welche die Rolle von Koordinationsschnittstellen zwischen der GND und den verschiedenen Kultursparten übernehmen sowie eine Mitwirkung der Kultureinrichtungen auf allen erforderlichen Ebenen ermöglichen. Es gibt bereits heute Agenturen in den Zentralen der Bibliotheks-

verbünde. Sie haben ein breites Aufgabenportfolio. Doch neue Bedarfe aus anderen Sparten bringen neue Aufgaben mit sich. Ein im Forum vorgestelltes Poster<sup>4</sup> illustriert »Mögliche Dienstleistungen der GND-Agenturen« als Baum. Die Dienste lassen sich grob in fünf Kategorien aufteilen: Beratung und Support (pink), Community-Arbeit (rot), Regeln und Standards (grün), Redaktion und Datenqualität (gelb) und technische Dienste (blau). Dieser Baum ist nur eine idealtypische Darstellung von vielen Ausformungen, die GND-Agenturen ihren Sparten und Communities anbieten können. Eine reale Agentur wird ihr Portfolio und die Ausprägung der Dienste auf ihre spezifische Zielgruppe ausrichten. Einige Gruppen benötigen beispielsweise eine starke Zentralredaktion, weil sie selbst kaum redaktionelle Erfahrungen und Möglichkeiten haben. Andere Agenturen haben eher die Rolle einer Koordinierungsstelle, die die redaktionelle Arbeit verteilt.



Die möglichen Dienste der GND-Agenturen als Blätter eines Baumes.  
Abbildung: Deutsche Nationalbibliothek, Jürgen Kett

Auf weiteren Postern wurden anhand zweier fiktiver Beispiele die Abstimmungsprozesse und Beteiligungswege innerhalb der künftigen GND-Kooperative dargestellt. Hierzu ist eine flexible Ausweitung der etablierten Gremien- und Arbeitsgruppenstruktur geplant, um neue Interessengruppen an der Strategie- und Produktentwicklung unmittelbar zu beteiligen. Ein mehrfach geäußerter Rat aus dem

Forum war, sich hinsichtlich der Beteiligungsmöglichkeiten an der Offenheit anderer Community-Projekte wie etwa der Wikipedia zu orientieren.

## Die offene GND ist immer eine Normdatei

Wie aus Daten Normdaten werden, welche Eigenschaften, Klassen und Beziehungen für alle gelten und wo spartenspezifisch ausdifferenziert werden sollte, diskutierten die Gäste im Weltcafé »Datenmodell«. Die folgenden Voraussetzungen sollte jeder Normdatensatz erfüllen:

- Er wird tatsächlich als Element zur Beschreibung von Kulturgütern benötigt.
- Er enthält ausreichend viele Informationen aus verlässlichen Quellen, um die durch ihn repräsentierte Entität eindeutig zu beschreiben.
- Eine Einrichtung übernimmt die Verantwortung für die nachhaltige Qualitätssicherung des Datensatzes.

Eine für unterschiedliche Anwendergemeinschaften offene Normdatei kann nur funktionieren, wenn Eigenschaften und Regeln definiert werden, die übergreifend für jede Anwendergemeinschaft verbindlich sind (CORE-Regeln). Im Zentrum steht das Basismodell, welches die grundsätzlichen Eigenschaften der GND auf abstrakter Modell-Ebene beschreibt. Die Eigenschaften, Relationen und Klassen gruppieren sich mit zunehmender Differenziertheit (Vererbungsprinzip) um diesen Kern. So ist es möglich, die GND-Regeln dergestalt weiterzuentwickeln, dass sehr allgemein gehaltene Grundregeln je nach Sachverhalt in mehr oder weniger komplexe Spezialregeln ausdifferenziert werden können. Eine Anwendergemeinschaft kann sich auf dieser Grundlage ein Set an Eigenschaften, Relationen, Klassen und Regeln definieren (PLUS-Modul). Die Analyse der Fallbeispiele im Projekt GND4C bestätigt die These, dass in den meisten Fällen der zusätzliche Bedarf lediglich darin besteht, präzisierende Regeln zu formulieren, um die allgemein zur Verfügung stehenden Klassen, Relationen und Eigenschaften anwenden zu können<sup>5</sup>.

## Fortschritte im automatisierten Import

Neu hinzukommende Sparten benötigen zu Beginn einen größeren Grundbestand von Normsätzen, die in der GND bislang fehlen und daher neu eingespielt werden müssen. Die Datensätze liegen oft schon als strukturierte Quelldaten vor. Jedoch müssen die Datenlieferanten vor einem Massenimport durch eine Analyse sicherstellen, dass die Daten eine ausreichende Qualität aufweisen und keine Dubletten in die GND gespült werden. Auf dem Forum stellte das GND4C-Team Werkzeuge vor, welche die neuen Partner bei dieser Aufgabe unterstützen sollen. Die technische Infrastruktur von GND4C, das Projectlab, besteht aus mehreren Modulen: Der Matching-Service durchsucht unterschiedliche Quell- und Zieldaten auf übereinstimmende oder ähnliche Zeichenketten. Die im Matching-Service erstellten Treffer sind Kandidaten für den Mapping-Prozess. Dabei werden im Mapping-Service semantisch übereinstimmende Begriffe beziehungsweise Entitäten unter Verwendung der Mapping-Properties nach dem «Simple Knowledge Organization System» (SKOS)<sup>6</sup> festgelegt. Bei Übereinstimmungen werden GND-Datensätze mit zusätzlichen Informationen aus den Quelldaten angereichert oder korrigiert. Am Ende des Prozesses teilt sich der Quelldatenbestand in drei Kategorien auf: neue GND-Datensätze, Ergänzungen zu bestehenden und nicht zu importierende Dubletten. Erst ab der zweiten Projektphase ist geplant, neue GND-relevante Datensätze in die GND aufzunehmen.



Heitere Stimmung bei der Diskussion des Schaubildes zur Evolution der GND-Normdatenbank.  
Foto: Deutsche Nationalbibliothek, Julia Krüger

## Eine starke Community

Auf dem Forum wurde auch das Organigramm einer offenen GND-Kooperative vorgestellt. Dabei wurden die Faktoren diskutiert, die zur Qualitätssicherung beitragen können: personelle Ressourcen, Schulungen für Redakteurinnen und Redakteure, die Dokumentation der Datenherkunft, eine nachvollziehbare Entwicklungsgeschichte des einzelnen Datensatzes und ebenso praxisorientierte wie verbindliche Regeln. Die auf dem Forum vorgestellte Beteiligungsmatrix ergänzte, wer sich wann, wo und wie einbringen kann. Unterschiedliche Positionen gab es zum Beispiel bei der Frage, wie stark die Öffnung der GND gesteuert werden sollte. Einige setzen auf die selbstregulierenden Kräfte einer möglichst großen Community, andere auf ein geregeltes Wachstum der Community mit gesteuerten Qualitätsmaßnahmen durch zentrale Einrichtungen. Einigkeit bestand allerdings in dem Punkt, dass die Bereitstellung von Ressourcen für die zuverlässige Pflege und das nachhaltige Anlegen von Normdatensätzen institutionell gefördert werden muss.

Dabei wurde deutlich, dass sich diese Normdatendarbeit nicht allein auf die Pflege der GND beschränkt, sondern auch andere Normdaten- und Informationssysteme und deren Verknüpfung mit der GND beinhalten sollte. So wäre es vorstellbar, anstelle des Imports jeglicher Thesauri, eher ein Netzwerk, vergleichbar einem Ökosystem, von miteinander interagierenden Thesauri und Normdatenbanken zu fördern. So könnte sowohl die gewünschte Vielfalt der Ontologien als auch die Stabilität der Gesamtheit der vernetzten Daten nachhaltiger sichergestellt werden.

## Zwei gute Nachrichten zum Schluss

Das GND-Datenmodell muss nicht neu erfunden werden. Die Fallanalysen im Projekt GND4C haben ergeben, dass die GND strukturell für die Modellierung nicht-bibliothekarischer Normdaten bereits geeignet ist.

Zudem bestehen bezüglich der Anforderungen der verschiedenen Interessengruppen an die GND große Schnittmengen. Die Digitalisierung und digitale

## Zeitpunkte

Vernetzung sorgen für eine zunehmende Konvergenz der Interessen. Das war auf dem Forum deutlich spürbar. Auch nach einem langen Tag so intensiver wie inspirierender Arbeit war der Wunsch

groß, das gemeinsame Entwickeln von Ideen und Lösungen fortzusetzen. Gelegenheit dazu wird es in kommenden Symposien und im Rahmen von Arbeitsgruppen geben.

## Anmerkungen

- 1 Das GND4C-Forum war eine Gemeinschaftsveranstaltung der Projektpartner des DFG-geförderten Projekts »GND für Kulturdaten« (GND4C): Bibliotheksservice-Zentrum Baden-Württemberg, Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg, Deutsche Nationalbibliothek, Deutsche Digitale Bibliothek
- 2 <[https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis\\_foerderangebote/fachinfodienste\\_wissenschaft/](https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/fachinfodienste_wissenschaft/)>
- 3 <<https://www.dfg.de/foerderung/programme/nfdi/>>
- 4 Die vollständige Dokumentation des Forums mit allen Postern, Vorträgen und Diskussionen sowie mehr zum Project Lab ist im öffentlichen Wiki nachzulesen: <<https://wiki.dnb.de/display/GND/GND4C-Forum+Dokumentation%3A+Startseite>>
- 5 siehe Fußnote 1.
- 6 SKOS ist eine Empfehlung des World Wide Web Consortiums für die Modellierung von Thesauri, Ontologien, Taxonomien, Schlagwortkatalogen und anderen kontrollierten Vokabularen; vergleiche <<https://www.w3.org/TR/skos-reference/#mapping>>