

Esther Scheven

# Qualitätssicherung in der GND

## *Inhalte*

Allgemeines zur GND

Heterogenität in der GND

Qualitätssicherung: Organisation

Qualitätssicherung: Datenformat

Ergebnisse aus dem Arbeitsteam zur Qualität

Gesprächsreihe

# Was ist die GND?

- **G**emeinsame **N**ormdatei (seit 2012)

Informationen zur GND: [Webseite zur GND](#)

[Unterseite auf der DNB-Webseite](#)

[Wiki für die Redaktionsarbeit](#)

- hervorgegangen aus vier Vorgängerdateien  
(siehe [Geschichte zur GND](#))

# Aktuelle Daten

## Datenvolumen

Körperschaften (Satzart Tb):	1.501.689
Konferenzen (Satzart Tf):	854.461
Geografika (Satzart Tg):	317.051
Personen (Satzart Tp):	5.572.400
Sachbegriffe (Satzart Ts):	203.361
Werke (Satzart Tu):	468.257

*Stand: 9.6.2021*

## Kooperative Pflege

drei Nationalbibliotheken (D-A-CH)

8 Bibliotheksverbände

(BVB – BibliotheksVerbund Bayern; BSZ – Bibliotheksservice-Zentrum Baden-Württemberg; GBV – Gemeinsamer Bibliotheksverbund der Länder Bremen, Hamburg, MecklenburgVorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen und der Stiftung Preußischer Kulturbesitz; HeBis – Hessisches BibliotheksInformationsSystem; hzb – Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen; SLSP Swiss Library Service Plattform; KOBV – Kooperativer Bibliotheksverbund Berlin-Brandenburg, Zuse Institut Berlin; OBV – OBVSG/Österreichischer Bibliothekenverbund.)

zahlreiche kleinere Anwender

# Aktuelle Daten

## GND-Dashbord: statistische Auswertungen der GND und ihrer Verknüpfungen mit den Titeldaten der DNB

### Relationen

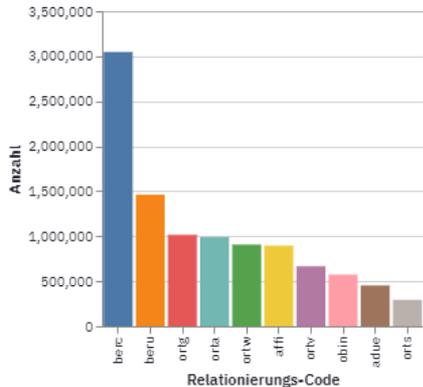
GND-Datensätze können mit anderen Datensätzen verlinkt (>relationiert<) werden. Die Art der Verlinkung wird über einen Relationierungscode beschrieben. Hier sind die am häufigsten verwendeten Relationierungscode zu sehen. Die Auflösung der wichtigsten Codes gibt es [hier](#).

Zeige Top ...

10

5

122



Relationierungscode

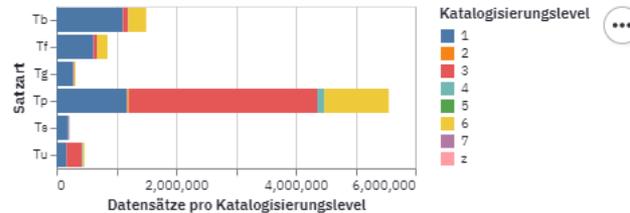
- berc
- beru
- orig
- orta
- ortw
- affi
- ortv
- obin
- adue
- orts

### GND Statistik allgemein

GND-Entitäten gesamt: 8.907.221

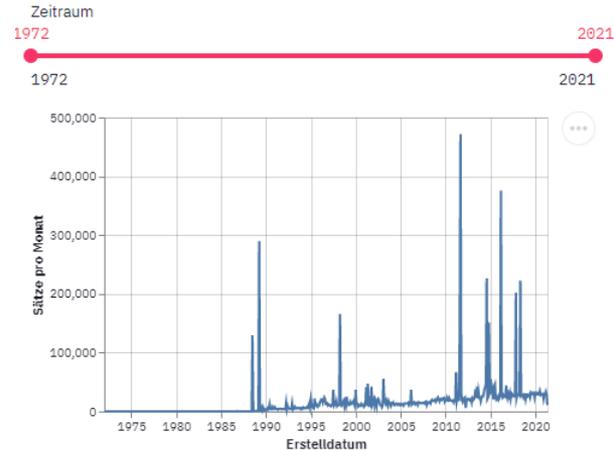
### Entitäten und Katalogisierungslevel

Alle GND-Entitäten können in verschiedenen Katalogisierungsleveln (1-7) angelegt werden. Je niedriger das Katalogisierungslevel, desto verlässlicher die Daten, weil Sie dann von qualifizierten Personen erstellt bzw. überprüft wurden.



### Zeitverlauf der GND-Datensatzerstellung

Auf einer Zeitreihe wird die Anzahl der monatlich erstellten GND-Sätze aufgetragen. Die ersten Sätze stammen aus dem Januar 1972



## Webseite zur GND (seit April 2021)

### Herzlich willkommen auf den Webseiten der Gemeinsamen Normdatei (GND)

Die GND (Gemeinsame Normdatei) ist die größte Normdatensammlung für Kultur- und Forschungsdaten im deutschsprachigen Raum. Ursprünglich ein Arbeitswerkzeug in Bibliotheken wächst im Zuge der digitalen Transformation mit ihrer Öffnung ihre Bedeutung als ein zentrales Angebot von Knotenpunkten im Netz. Die GND stellt Verbindungen innerhalb unseres gemeinsamen Wissensnetzes her. Gehen Sie auf Entdeckungsreise im GND network. Wir freuen uns über Ihren Besuch. Diese Website ist mehr als nur eine Informationsplattform. Die GND bietet Ihnen viele Möglichkeiten, aktiv zu werden, daher bietet Ihnen die Website viele Wege an sie zu erkunden:



Sie suchen Informationen über die GND, ihre Struktur auf organisatorischer und struktureller Ebene sowie eine allgemeine



Die vielen Millionen Normdatensätze stehen allen Menschen und Maschinen zur freien Nutzung zu Verfügung. Wenn Sie

GND-Blog

Neues und interessantes rund um die GND.

[MEHR](#)

**GNDCon 2.0**  
 Vom 07. bis zum 11. Juni richten die Partnereinrichtungen der GND-Kooperative die zweite Convention zur Gemeinsamen Normdatei (GNDCon 2.0) aus. Digital, divers und dezentral diskutieren an fünf Tagen Menschen mit verschiedenen Interessen und aus diversen Bereichen ganz unterschiedliche Fragen, Aufgaben und Ansätze rund um das Thema GND.

## Wiki für die Redaktionsarbeit

Seiten / ILLTIS-Seiten

### Informationsseite zur GND

Erstellt von Trunk, Daniela, zuletzt geändert von Thüncher, Sylvia am 2021-05-27

Willkommen im Wiki der Gemeinsamen Normdatei

[Allgemeine Informationen](#) | [Regeln](#) | [Erfassung](#) | [Codes](#) | [Redaktion, Umgang mit Altdateien, Rangfolge der Nachschlagewerke usw.](#) | [Formate, Indexierung und technische Informationen](#) | [RDA-Schulungsurter](#)

Die Gemeinsame Normdatei (GND) enthält Datensätze für Personen, Familien, Körperschaften, Konferenzen, Gebietskörperschaften bzw. Geografika, Sachbegriffe und Werktitel.

Die GND wird kooperativ geführt und an der Deutschen Nationalbibliothek (DNB) gehalten. Die GND-Normdaten stehen allen Interessierten in den Formaten MARC 21 Authority, MARC 21-xml und RDFxml kostenfrei unter der Lizenz CC0 1.0 zur Verfügung.

[Für Fragen und Hinweise an Frau Thüncher \(s.t.hue\)](#)

**Allgemeine Informationen**  
 Allgemeine Informationen zur GND finden Sie auf den [Webseiten der Deutschen Nationalbibliothek](#).

**Aktuelles und Wichtiges zur GND-Anwendung**  
 Aktuelle Informationen zur GND

**Regeln**  
 Normdatensätze für Personen, Familien, Körperschaften, Konferenzen und Gebietskörperschaften sowie Werke (mit Ausnahme von Sammlungen im Sinne der [EH-S-05](#)) werden nach dem Regelwerk "Ressour für den deutschsprachigen Raum durch Anwendungsregeln (AWR) und Erläuterungen (ERL) ergänzt.

**RDA-Toolkit**  
 RDA-Anwendungsrichtlinien

Normdatensätze für Sachbegriffe, Sammlungen im Sinne der [EH-S-05](#) und Geografika wie Bauwerke, Großplastiken, Wege, Naturparks, naturräumliche Einheiten usw. werden nach den "Regeln für die Schlagregeln für die Schlagwortkatalogisierung (RSWK)

**Erfassung**  
 Hilfen zur Erfassung einzelner Entitäten nach RDA und RSWK, inkl. Beispiele in PICA3 und Aleph:  
 Erfassungshilfen für allgemeine Sachverhalte  
 Erfassungshilfen für Personen und Familien  
 Erfassungshilfen für Körperschaften und Konferenzen  
 Erfassungshilfen für Gebietskörperschaften bzw. Geografika  
 Erfassungshilfen für Werke

Beschreibung des Erfassungsformates pro Feld:  
 Erfassungleitfaden PICA  
 Erfassungleitfaden Aleph (B3Kat, hzb, OBV)  
 Erfassungleitfaden Aleph (IDS - RDA-Toolkit-Login erforderlich)  
 Erfassungleitfaden Alma (kobv, OBV)

# Anspruch an eine Normdatei

- nur *ein* Normdatensatz für einen Sachverhalt
- gleiche Regeln für gleiche Sachverhalte
- dokumentierte Regeln
- alle Beteiligten arbeiten nach den gleichen Regeln
- gemeinsame / abgestimmte Redaktionsabläufe

Wie kann dies praktisch sichergestellt werden?

# Herausforderung „Heterogenität in der GND“

- Von den ca. 9 Mio Datensätze der GND stammen ca. 4,9 Mio aus den Vorgängerdateien (d.h. sie wurden in einem anderen Datenformat und nach anderen Regeln erfasst).
- Änderungen der Regelwerke seit GND-Start
- unterschiedliche Qualitätslevel bei der Erfassung
- unterschiedliche Regeln für unterschiedliche Benutzergruppen

# Herausforderung durch Öffnung der GND

- Vorgängerdateien dienten der bibliothekarischen Erschließung und folgten bibliothekarischen Regelwerken.
- Andere Anwendungsgemeinschaften haben andere und weitere Bedürfnisse.
- Es gibt ein eingespieltes System der Redaktionsabläufe innerhalb der Bibliotheksgemeinschaft; neue Anwendungsgemeinschaften müssen hier integriert werden.

# Was dient alles der Qualitätssicherung?

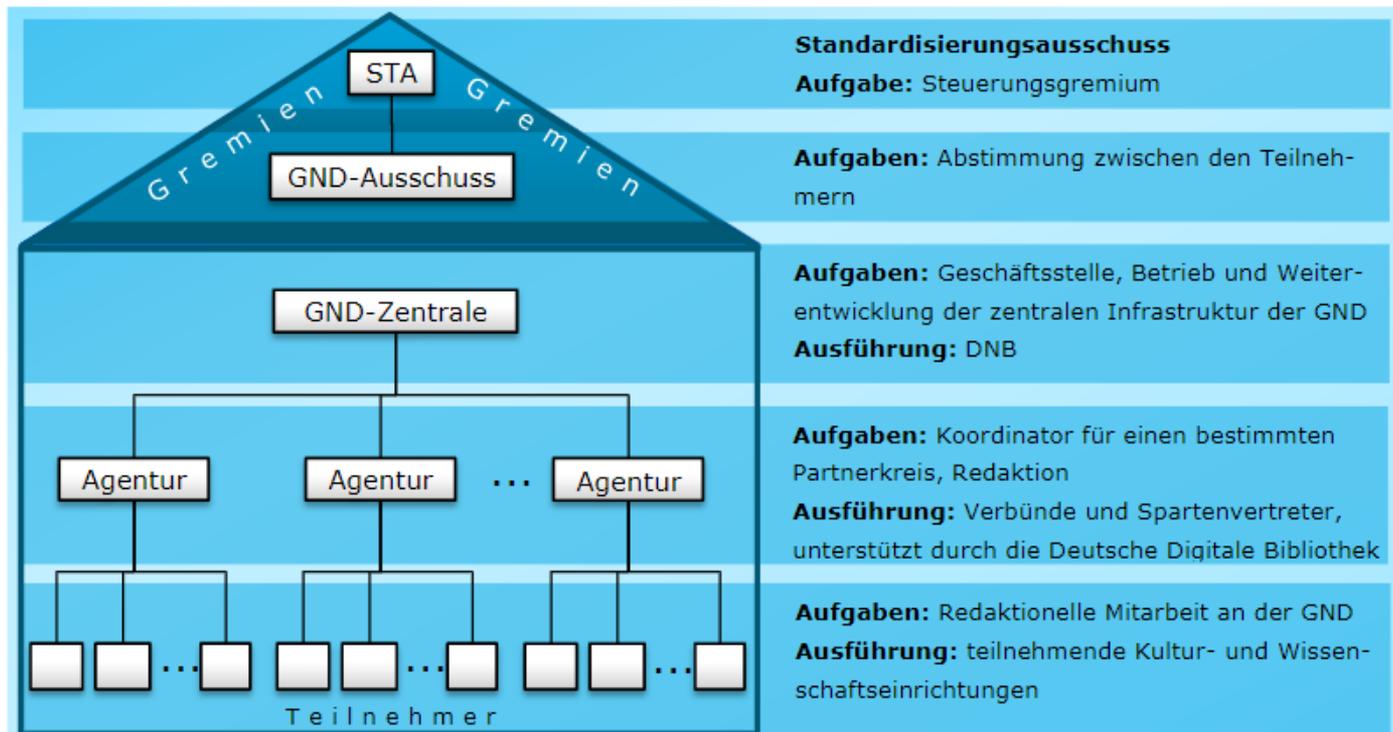
- ausgefeilter organisatorischer Rahmen:
  - GND-Ausschuss (Strategie; Pflege; Weiterentwicklung)
  - GND Kooperative (Rechte und Pflichten bei GND-Teilhabe)
    - Redaktionsstellen (Verbundzentralen u.ä.)
  - Fachgruppe Erschließung (Abstimmung von Regeln)
  - regelmäßige Telefonkonferenz zu redaktionellen Fragen
  - Arbeitsteam Qualität in der GND
  - Gesprächsreihe zur Öffnung der GND
  - regelmäßige DNB-interne Arbeitstreffen zur GND-Pflege

# Was dient alles der Qualitätssicherung?

- Datenformat (Unterfelder statt Deskriptionszeichen)
  - Dokumentation der Regeln ([öffentliche Wiki-Seite](#))
  - Unterstützung der Katalogisierenden durch Skripte und Datenmasken
  - Versionierung des Datensatzes in der WinIBW (seit 2017)
  - Archivbestand der Vorgängerdateien (GKD, SWD, EST, PND)
- (siehe auch: [Qualitätsmanagement in der GND, 2017](#))



# Organisation der GND-Kooperative



# Vereinbarungen der GND-Kooperative

- Jede kooperierende Einrichtung hat die Pflicht, Verantwortung für die Pflege der Daten zu übernehmen.
- Jede kooperierende Einrichtung informiert die anderen Einrichtungen über geplante Aktivitäten mit Bezug zur GND (Projekte, Präsentationen und Kooperationen).

(aus der [Kooperationsvereinbarung, August 2020](#))

# Arbeitsteam Qualität in der GND

*Auftrag:* Die im GND-Ausschuss gesammelte Qualitätsdefizite der GND *gemeinsam* bearbeiten und beheben.

- seit Oktober 2019; vom GND-Ausschuss einberufen
- Alle Verbände sind vertreten.
- Telefonkonferenzen ca. alle 8 Wochen
- Leitung bei der DNB

# Arbeitsteam Qualität in der GND

- Löschung von Ts6- und Ts7-Datensätzen des hbz
- Ergänzen von Ländercodes und Entitätencodes
- Reduzierung von Dubletten
- Aufarbeitung von Kandidaten-Datensätzen
- Priorisierung von Validationen

# Aufarbeitung von Kandidaten

Bei Dateneinspielungen werden Datensätze, die nicht eindeutig als Match erkannt und zusammengeführt werden können, als Kandidat gekennzeichnet. Sie können gemäß einer Anweisung in dem öffentlichen Wiki geprüft, zusammengeführt oder entkoppelt werden. Das geht rasch und elegant und wurde im Qualitätsteam für die vorhandenen Datensätze durchgeführt (ca. 24.000 Mai 2020; ca. 3000 Mai 2021).

<https://wiki.dnb.de/x/tAbeBg>

# Reduzierung von Dubletten

- Konferenzdatensätze (Umsetzung ca. Sommer / Herbst 2021)
- Körperschaftsdatensätze (Umsetzung ca. Herbst / Winter 2021)
- Personendatensätze (Umsetzung ca. Sommer / Herbst 2021)

# Personen-Dubletten und Culture-Graph

*Culture-Graph?*



Eine verbundübergreifende Plattform für Dienste und Projekte rund um die Themen Datenanalyse und Datenvernetzung.

Mittels Culture-Graph werden Werkcluster gleicher Werke erzeugt; auf diese Weise können Normdatenangaben übernommen werden, z. B. Personennormdaten in Titelaufnahmen, die keine Normdatenverknüpfung haben. Es lassen sich so auch Dublette Datensätze erkennen.

# Personen-Dubletten und Culture-Graph

- Ca. 12.000 potentielle Kandidaten

Die Datensätze, die den strengen [Match&Merge-Kriterien](#) entsprechen, werden direkt zusammengeführt (neben dem Namen müssen immer noch zwei weitere Elemente identisch sein (Geburts- oder Sterbejahr und Beruf oder Geburtsort oder Sterbeort)).

## Gesprächsreihe: Eignungskriterien

- Es besteht ein berechtigter Bedarf und die Daten dienen dem Zweck der GND.
- Die Daten stehen unter freier Lizenz.
- Die Daten sind verlässlich.
- **Die Daten entsprechen den aktuell gültigen Regelungen der GND.**
- **Die Pflege der Daten ist nachhaltig gewährleistet.**
- Die betrieblichen und technischen Rahmenbedingungen sind gegeben.

[siehe GND-Webseite](#)

# Gesprächsreihe: Dokumentationsplattform mit Wikibase als Grundlage

## Thematische Einstiege und Sichten

- Einstiegsseiten für die wichtigsten Themen und Anwendungskontexte
- Generierte Indexseiten für den Schnelleinstieg und Überblick

## Regelpyramide

- Basis: Allgemeine Regeln und Infos
- Darauf aufbauend: Regeln für spezielle Ressourcen und Anwendungen

## Vernetzung mit Erschließungsregeln („Verwendung“)

- CORE-Regeln: Anwendungs-unabhängig
- PLUS-Regeln: Anwendungs-spezifisch

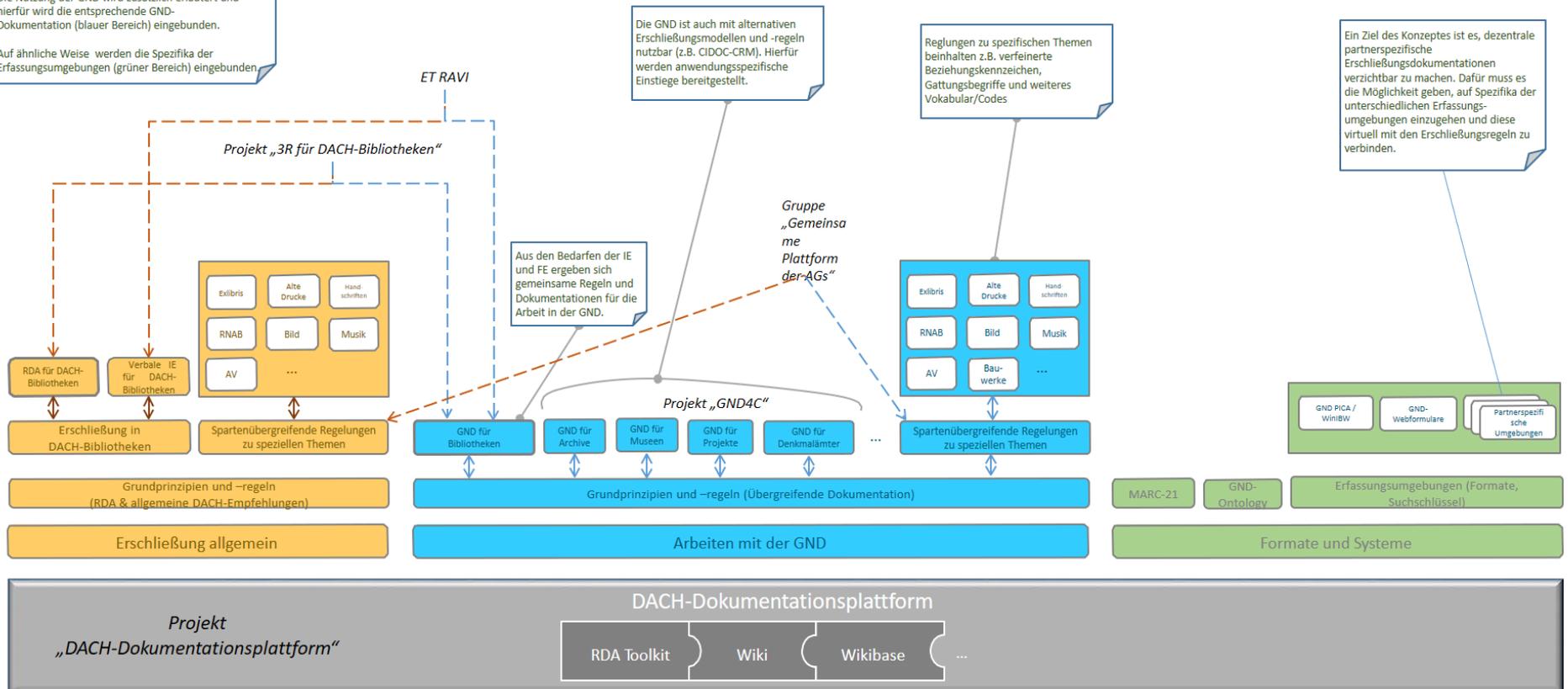
## Datenbank-basiert

- Nachnutzung von zentral verwalteten Informationen
- Leichter Zugriff auf Codes und Notationen

Regeln sind in der allgemeinsten Form **formatunabhängig** und **unabhängig von der GND**.

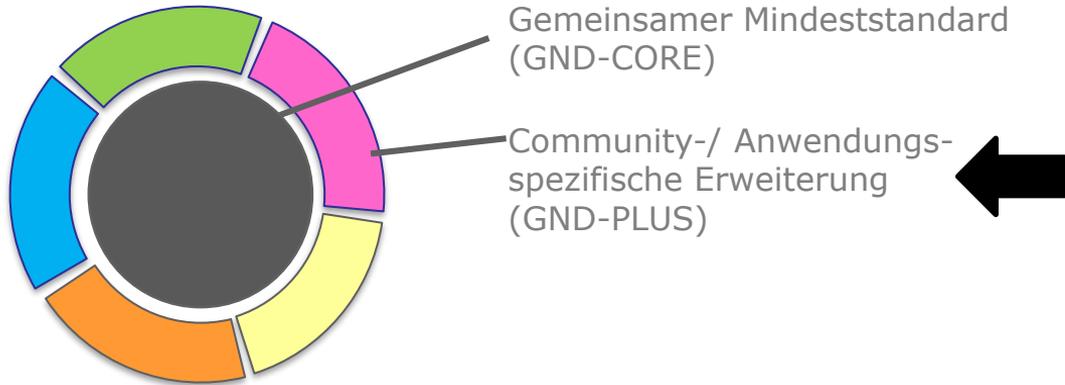
Die Nutzung der GND wird zusätzlich erläutert und hierfür wird die entsprechende GND-Dokumentation (blauer Bereich) eingebunden.

Auf ähnliche Weise werden die Spezifika der Erfassungsumgebungen (grüner Bereich) eingebunden.



# Gesprächsreihe: Plus-Markierungen in der GND

GND-Entwicklungsprogramm:  
*„Modulares Normdatensystem aufbauen“*



Nicht jeder muss alles erfassen.  
-> Kennzeichnung von  
zusätzlichen Elementen  
-> d.h. wir haben Daten über  
Daten

Hinweis:  
Für bestimmte Felder gibt  
es bereits heute Informationen  
über die Metadatenherkunft.

Qualitätssicherung durch „Metametadaten“

*Durch welche Quelle ist die Aussage bzw. der Wert belegt?*

Ganzer Datensatz: Feld 670 (Quellenangabe); feldbezogen \$v (Bemerkungsfeld) etc.

*Wie gesichert ist sie?*

Feld 670; Angabe des Levels etc.

*Wer ist für sie verantwortlich?*

Ganzer Datensatz: Feld 903 (Redaktionsstelle)

## Metametadaten sind Aussagen über Aussagen

(Es gibt bereits jetzt in der GND dafür viele Möglichkeiten.)

*Welchem Standard folgt sie?*

z.B. Feld 040 (Regelwerk) oder Unterfelder \$u/\$2 bei Koordinaten

*Wann wurde sie ergänzt?*

Seit 2017 erkennbar (s revp)

*Welcher (maschinelle) Prozess wurde verwendet?*

Einspielungen der ORCID-Ids in Feld 024 in \$v gekennzeichnet.

Die GND ist vielfältig wie die Steine eines Mosaiks.

Die Sicherung ihrer Qualität hat zahlreiche Facetten, kann aber nur ***gemeinsam*** erfolgen.

Nur so können aus den einzelnen Steinen weiterhin schöne Mosaik  
gelegt werden.



Wenn Sie weiteren Kontakt wünschen:

[e.scheven@dnb.de](mailto:e.scheven@dnb.de)

© der beiden Fotos: Esther Scheven