

Vorbemerkung

In der folgenden Darstellung werden Norm- und Metadaten unterschieden. Metadaten werden wiederum unterteilt in: Metadaten zu Primärquellen (Kalliope/EAD) und Sekundärquellen (ZDB- und DNB-titel/MARC21). Während die Regeln zu Sekundärquellen erschöpfend sind, gilt für die Primärquellen, dass ein umfassendes, aber potenziell auch noch erweiterbares Regelset vorliegt.

In den Ausgangsdaten für SoNAR (IDH) sind soziale Beziehungen nur in den Normdaten (auch: Referenz-/ Stammdaten) wie GND oder Wikidata explizit kodiert. Im Austauschformat der GND (MARC21) erfolgt dies für Akteure (Personen und Körperschaften) in der Form¹:

035\$0{GND-ID}	(!Entität A!)
100 110\$a{Name}	(!Entität A!)
500 510\$0{GND-ID}\$a{Name}\$4{Code f. Art d. Relation}	(!Entität B!)

Aus einer „Graph-Perspektive“ ergibt sich daraus eine direkte Beziehung zwischen zwei Akteuren:
A ---- B wobei das Merkmal für die Kante aus dem Code in 500|510\$4 übernommen wird

In den Metadaten, die die Ressourcen von Archiven und Bibliotheken beschreiben, sind Beziehungen zwischen Akteuren, die mit den Ressourcen identifiziert werden, nur implizit kodiert. Ziel ist es, diese Beziehungen regelbasiert abzuleiten und explizit als Daten zur Verfügung zu stellen. Voraussetzung ist, dass die Akteure eindeutig mit einer Normdaten-ID identifiziert sind. Um die Art und Richtungen sozialer Beziehungen (ungerichtet, gerichtet) zu berücksichtigen, ist die Auszeichnung der Rollen der Entitäten im Kontext der Ressource (Verfasser, Herausgeber, Adressat, etc.) eine weitere Voraussetzung. In Metadaten für Sekundärquellen (Publikationen) von ZDB und DNB erfolgt dies in folgender Form:

100 110\$0{GND-ID}\$a{Name}\$4{Code f. Art d. Relation}
700 710\$0{GND-ID}\$a{Name}\$4{Code f. Art d. Relation}

Aus einer „Graph-Perspektive“ besteht eine direkte Relation zwischen der Ressource und dem Akteur / den Akteuren, die die Ressource erstellt haben. Eine Relation zwischen > 1 Akteur, z.B. zwei Verfassern, kann nur abgeleitet werden (s. unten)².

Metadaten zu Primärquellen (Kalliope) sind ähnlich, aber auf Basis von EAD formal kodiert:

//persname corpname/	
@authfilenumber	{GND-ID}
@normal	{Name}
@role	{Art der Relation}

¹ s. auch Präsentation v. 17.07.2019, Einführung Datenformate

² dies erfolgt durch Tests von AP2 und AP3 durch Abfrage von n-Graden

Auch hier besteht eine direkte Relation zwischen der Ressource und dem Akteur / den Akteuren, die die Ressource erstellt haben. Eine Relation zwischen > 1 Akteur, z.B. zwischen Verfasser und Adressat, kann somit abgeleitet werden. Zusätzlich tritt bei Primärquellen, speziell von so genannten Archivbeständen, anhand des zugrundeliegenden Erschließungsprinzips (Provenienzprinzip) eine weitere Annahme über soziale Beziehungen hinzu: die zwischen Bestandsbildner (//origination/persname|corpname) und den Akteuren, die mit den Quellen eines Archivbestandes identifiziert werden (//controlaccess/persname|...).

Sowohl ZDB und DNB als auch Kalliope identifizieren in Meta- und Normdaten die Normdatei, in der die sozialen Akteure näher beschrieben sind. Dies ist zunächst eine konstruktive Feststellung, da innerhalb des deutschsprachigen Raums die GND i.d.R. als Referenzdatei gesetzt ist, aber mit Blick auf Daten, die etwa gegen Wikidata referenzieren, ist die Berücksichtigung dieser Information im Konzept zwingend:

In den MARC21 Daten wird in 035, 1XX\$0, 5XX\$0, 7XX\$0 den Identnummern das ISIL der Referenzdatei vorangestellt. Für die GND: DE-588 (z.B. 100\$0(DE-588){ID}. Kalliope benennt die Normdatei mit dem Attribut @source, z.B. //persname|corpname/@source="GND".

Mit den nachstehenden Regeln zur Ableitung sozialer Beziehungen ist es zielführend, das im LD-Format aufbereitete Datenangebot des HBZ (lobid.org³) für SoNAR (IDH) zu prüfen. Es enthält: GND, ISIL und Metadaten der Sekundärquellen der Bibliotheken des HBZ einschließlich der ZDB. Vierteljährlich werden Referenzen von GND-Entitäten zu VIAF, ISNI, Wikidata, ... von Entity Facts übernommen. Zu prüfen ist, ob die Art, in der die Daten aufbereitet und bereitgestellt sind, Prozesse zur Übernahme und Transformation oder auch Deduktion von sozialen Beziehungen im Vergleich zu Datendumps in konventionellen XML-Formaten eher geeignet sind für den Betrieb einer Forschungstechnologie SoNAR (IDH).

Regel 1: „Folgerung“ sozialer Beziehungen

Folgende allgemeine Annahme soll geprüft werden: Für alle Akteure – Personen, Körperschaften und Familien –, die gemeinsam in einem Metadatensatz genannt sind, besteht eine soziale Beziehung. Diese Beziehung heißt „ist assoziiert mit“ und ist ungerichtet. Die in den Ausgangsdaten implizit kodierten Beziehungen sollen explizit in Form von Kanten zwischen den Akteuren in SoNAR (IDH) kodiert sein:

Für Sekundärquellen (MARC21):

100|110\$0{GND-ID}\$a{Name}\$4{Code f. Art d. Relation}

700|710\$0{GND-ID}\$a{Name}\$4{Code f. Art d. Relation}

Für Sekundärquellen (lobid.org):

Key „agent“ {„id“, „gndidvertiser“, „label“}

Key „role“ {„id“, „label“}

³ <https://lobid.org/>

Für Primärquellen (Kalliope/EAD):

```
//controlaccess/persname|corpname/  
    @authfilenumber      {GND-ID}  
    @normal               {Name}  
    @source               {GND}  
    @role                 {Art der Relation}
```

Von der Regel ausgenommen sind Akteure, die in Metadaten zu a) Sekundärquellen als Index-Begriff (653) oder b) Primärquellen (@role) als „dokumentiert“ | „behandelt“ | „nicht-definiert“ kodiert sind.

Regel 2: Co-Autoren

Die Beziehungen zwischen Akteuren > 1 = Verfasser (\$4/role: „cre“) in einem Metadatensatz für Sekundärquellen heißt „sind Co-Autoren“ und ist ungerichtet.

Regel 3: Co-Herausgeber

Die Beziehungen zwischen Akteuren > 1 = Verfasser (\$4/role: „isb“) in einem Metadatensatz für Sekundärquellen heißt „sind Co-Herausgeber“ und ist ungerichtet.

Regel 4: Affiliationen

Beziehungen zwischen Personen und Körperschaften (Affiliationen) können angenommen werden:

- wenn in einem Metadatensatz für Sekundärquellen mind. 1 Verfasser (Person) und 1 Herausgeber (Körperschaft) ausgezeichnet sind
- wenn in einem Metadatensatz für Primärquellen 1 Verfasser Person und 1 Verfasser Körperschaft oder 1 Adressat Person und 1 Adressat Körperschaft ausgezeichnet sind

Die Annahme für Primärquellen beruht auf dem Prinzip, dass im Fall von Korrespondenzen die Nennung von Körperschaft und Person für Verfasser deutlich wird, dass die Person im Auftrag der Körperschaft einen Brief geschrieben resp. empfangen hat.

Die Beziehung heißt „ist affiliert mit“ und ist ungerichtet.

Regel 5: Korrespondenzen

Für alle Primärquellen, in denen die Werte `//controlaccess/genreform/@authfilenumber 4008240-4` (Briefe) | `4146609-3` (Briefwechsel) ausgezeichnet sind, soll eine Korrespondenzbeziehung zwischen Verfasser und Adressat explizit gemacht werden. Die Beziehung heißt „hat korrespondiert mit“ und ist gerichtet (unidirektional bei Briefe und `//controlaccess/persname|corpname/@role= „Verfasser : Adressat“` oder bidirektional bei Briefwechsel und „Verfasser : Adressat“ | „Korrespondenzpartner“).

Weiterhin: Es ist eine soziale Beziehung anzunehmen, wenn in einem Metadatensatz für Primärquellen mit `//controlaccess/genreform/@role 4008240-4` | `4146609-3` Verfasser in Verbindung mit einem oder mehreren Akteuren genannt sind, die mit der Rolle „erwähnte Person“ oder „behandelte Person“ (resp. Körperschaft) ausgezeichnet sind. Die Beziehung zwischen Verfasser und „erwähnte | behandelte Person“ (resp. Körperschaft) heißt „kennt“ und ist gerichtet (unidirektional). Ob dies für die Adressaten eines Briefes / Briefwechsels gilt, ist „unsicher“ und ist speziell von AP2 zu prüfen.

Regel 6: Tagebücher

Für Tagebücher (`//controlaccess/genreform/@role= 4058900-6`) soll eine soziale Beziehung zwischen Verfasser zu erwähnte / behandelte Person / Körperschaft angenommen werden (s. Korrespondenzen). Die Beziehung heißt „kennt“ und ist gerichtet (unidirektional).

Regel 7: Stammbücher / Stammbucheintrag, Alben

Für Stammbucheintragungen (`//controlaccess/genreform 4508344-7`) soll angenommen werden, dass eine soziale Beziehung zwischen dem Verfasser / Schreiber des Eintrags zum Adressaten (= Eigentümer des Stammbuchs) besteht. Die Beziehung heißt „kennt“ und ist ungerichtet.

Regel 8: Protokolle

Für Protokolle (`//controlaccess/genreform 4047552-9`) soll angenommen werden, dass eine Beziehung zwischen den Personen (`//controlaccess/persname/@role=“dokumentiert“`) (= Teilnehmer) besteht. Die Beziehung heißt „kennt“ und ist ungerichtet.

Regel 9: Archivbestände / Findbücher

Für alle `//controlaccess/persname|corpname/@role=Verfasser|Schreiber|Adressat` von Metadatensätzen eines EAD-Dokuments und die nicht gleich „Bestandsbildner“ (`//origination/persname|corpname`) sind,

soll eine Beziehung zum Bestandsbildner angenommen werden. Voraussetzung ist, dass die Metadaten im Findbuch nicht Teil einer Sammlung (`//controlaccess/genreform/@authfilenumber="4128844-0"`) sind. Die Beziehung heißt „ist assoziiert mit“ und ist ungerichtet.

Für alle Bestandsbildner eines Findbuchs (> 1 Anzahl von `//origination/persname|corpname`) soll ebenfalls eine soziale Beziehung angenommen werden. Die Beziehung zwischen mehreren Bildnern heißt „kennt“ und ist ungerichtet.