

Konzept zur Anreicherung mit Sacherschließungsdaten im B3Kat

1. Vorbemerkung

Anreicherungen mit inhaltserschließenden Metadaten dienen der besseren Beschreibung und Suchbarkeit von Inhalten bzw. Themen in Titeldatensätzen, nicht zuletzt der quantitativen Relevanz der deutschsprachigen Metadaten in großen heterogenen Suchräumen. Sie erfolgen weitgehend maschinell nach intellektuell überprüften Vorgaben bzw. auf der Basis plausibler Feldzuordnungen in konsistent gepflegten Austauschformaten.

Die Daten können innerhalb des B3Kat (von Titelsatz zu Titelsatz) 1:1 (als Datenfeld oder -feldgruppe) übertragen, aus anderen Feldinhalten maschinell generiert oder aus externen Datenquellen eingespielt werden, z. B. aus anderen (auch lokalen) Katalogdatenbanken oder von anderen Anbietern aus dem Publikationswesen.

Von Anreicherung spricht man dann, wenn nicht ganze Titelsätze ausgetauscht, sondern Titelsätzen Ausschnitte spezifischer inhaltlicher Metadaten hinzugefügt werden.

Inhaltserschließende Metadaten beziehen sich grundsätzlich auf die Werkebene, können aber auch einzelne Elemente enthalten, die die Manifestationsebene betreffen (z. B. Formschlagwörter, bestimmte Notationen). Sie werden in den heutigen Katalogisierungsszenarien („zusammengesetzte Beschreibung“ nach RDA) an Titeldatensätzen auf der Manifestationsebene verankert.

Es geht immer um urheberrechtlich unproblematische Metadaten, also in der Regel um bibliothekarische Klassifikationsdaten und Schlagwörter sowie urheberrechtsfreie Abstracts etc.

Die routinemäßige Nachführung von Schlagwörtern der DNB ist nicht Gegenstand dieses Konzepts.

2. Nutzen der Anreicherung

Der größte Nutzen liegt in den Anreicherungen, die

- a) noch nicht erschlossene Inhalte erstmals durch inhaltsbezogene Metadaten auffindbar machen,
- b) Titelsätze erstmals mit Sacherschließungsdaten versehen, die den im deutschsprachigen Raum verbreiteten Standards entsprechen,
- c) zu Titelsätzen Normdatenverknüpfungen liefern, die es erlauben, die Dokumentinhalte in ein Netz von Linked-Data-Entitäten einzubinden.

GND-Verknüpfungen und gültige Notationen der gängigen Klassifikationen (vor allem RVK und DDC, aber auch LCC und Basisklassifikation) haben den höchsten Wert, aber auch Anreicherungen in Form von Fachklassifikationen, Termini aus Fachthesauri und anderen deutschsprachigen Schlagwörtern sind sinnvoll.

Autorenstichwörter können mit Herkunftsbezeichnung und Kennzeichnung als „Freier Deskriptor“ ebenfalls angereichert werden.

Schlagwörter in anderen Sprachen sind dann interessant, wenn sie z.B. durch das MACS-Projekt mit der GND gematched werden können. Das trifft auf französische (Rameau) und englische Schlagwörter (LCSH) zu. Schlagwörter in anderen Sprachen stellen in unserem Kontext nur einen geringen Mehrwert dar.

Während das Pfeffer-Verfahren auf RVK und RSWK beschränkt war, könnte künftig auch in den übrigen 7xx-Feldern angereichert werden.

Redundanzen können vorübergehend vor der Option einer künftigen Bereinigung in Kauf genommen werden. Bereinigungen sollten sukzessive nach bestimmten Vorgaben, z. B. korrekte Feldbelegungen oder Deduplizierung, erfolgen.

3. Allgemeine Vorgaben

Anreicherungen sollen nur unter folgenden Bedingungen stattfinden:

1. Übernahme echter Sacherschließungsdaten, die Informationen über die in Dokumenten behandelte Thematik geben. Mit sachfremden Elementen (z. B. reine Formangaben, Zielgruppenangaben) durchsetzte Datensegmente sollen ausgeschlossen bzw. vor oder nach der Anreicherung bereinigt werden.
2. Übernahme in die offiziellen für SE-Daten bestimmten Felder des Aleph-Internformats (MAB/ASEQ-Felder 7xx und 9xx), woraus sie mit den üblichen Routinen für die Recherche bereitgestellt werden können.
3. Anreicherung grundsätzlich nur, wenn das betreffende Feld im Zielsatz ganz leer ist (Addition in Ausnahmefällen denkbar, wenn Herkunftskennzeichnung möglich ist)¹
4. Normdatenverknüpfung per ID-Nummern nur mit der GND (mit überregionaler GND-Verknüpfungsnummer) und nur in den Feldern MAB 902 ff. (Weiterentwicklung von Klassifikationen wie RVK zu Normdateien vorbehalten)
5. Strikte Trennung von Notationen und verbalen Elementen (Schlagwörtern); verbale Benennungen aus Klassifikationen ggf. nur in Schlagwortfelder, nicht zusammen mit Notationen in Klassifikationsfelder
6. Felder sollen nur mit sauberem Feldanfang befüllt werden, nicht (Beispiel GBV):
740a L \$\$œœGerman fictionœœ21st centuryœœHistory and criticism
740a L \$\$œœGerman fictionœœ20th centuryœœHistory and criticism
740b L \$\$œœRussia (Federation)œœIn literature
7. MAB/ASEQ-konforme Belegung der Unterfelder: z. B. sollten Regelwerkskennungen, die mit Slash an den Schlagwörtern hängen, bei der Anreicherung in die richtigen Unterfelder eingespielt werden.
8. Da immer mehr Anreicherungen aus heterogenen Quellen kommen, ist die Kennzeichnung ihrer Herkunft und des zu Grunde liegenden Regelwerks sehr wichtig (s. das Konzept zur Herkunftskennzeichnung).²

4. Klassifikationen

Notationen werden nur angereichert in Feld 700 (indikator-spezifisch bzw. mit blank) und in Feld 705 (DDC analytisch) unter Beachtung der Unterfelder.

Legende: 700 blank = Systematik der katalogisierenden Institution

- Sachgruppen DNB (SF2 = DNB)
- Sachgruppen ZDB (SF2 = ZDB)
- BSB-Dezimalklassifikation (G: S: Z:)
- DFG-Systematik der Kunstbibliotheken (SF2 = KUBA2)
- Notationen der Berliner Charité
- (Ob neue Inhalte in dieses Feld kommen, muss unbedingt abgestimmt werden.)

¹ In manchen Fällen kann in einem Feld additiv angereichert werden, aber nur wenn sich der alte und der neue Inhalt durch Herkunftskennungen unterscheiden lassen.

² Wenn möglich sollten die Felder darüber hinaus mit einem Anreicherungskennzeichen versehen werden.

5. Schlagwörter

Schlagwörter sollen, sofern ihre Kategorien gemäß Herkunftskatalog gekennzeichnet sind, immer in die korrekten Unterfelder und mit korrekten Indikatoren eingestellt werden. Falls sie nicht unterschieden werden können, muss der Indikator blank (nicht spezifiziert) benutzt werden.

Für Schlagwortfolgen, die als Textstring vorliegen, soll als einheitliches Trennzeichen „_ / _“ (=blank slash blank) verwendet werden.

Bei Schlagwörtern in Majuskelschrift soll - falls im Rahmen der Anreicherungsroutine möglich - beim Import die Großschreibung auf Grund der Lesbarkeit korrigiert werden: SOCIAL SCIENCES zu Social Sciences

5.1 Feld 710 Schlagwörter und Schlagwortketten

Das Feld 710 ist wegen der fehlenden Herkunftsbezeichnung nicht zu empfehlen und sollte insofern nicht aktiv belegt werden.

MAB-Indikatoren:

- blank = nicht aufgegliedert
- a = Sachschlagwort
- b = geographisch-ethnographisches Schlagwort
- c = Personenschlagwort
- d = Körperschaftsschlagwort
- f = Formschlagwort
- g = Gattungsbegriff (nur VD18, nicht MAB!)
- z = Zeitschlagwort

Das Feld soll nur mit konsistenten, i. d. R. regelbasierten Daten aus lokalen Anwendungen belegt werden, die entweder abgeschlossen sind oder von den anwendenden Bibliotheken in diesem Feld weiter gepflegt werden. Heterogene Daten oder Daten unklarer Herkunft sollen hier nicht abgelegt werden. Eine Möglichkeit der Herkunftsbezeichnung besteht in Feld 710 nicht.

Die vorhandenen Daten stammen überwiegend aus Migrationen von Spezialbibliotheken. Über Anreicherungen in Feld 710 muss auf fachlicher Ebene (AGSE) jeweils eigens entschieden werden.

5.2 Feld 711 Schlagwörter und Schlagwortketten nach anderen Regelwerken

[zur Unterscheidung von 710 ist die Feldbenennung ungeeignet; auch in 710 befindet sich SE nach anderen Regelwerken]

Indikator:

- blank = nicht spezifiziert
- a = Sachschlagwort
- b = geographisch-ethnographisches Schlagwort
- c = Personenschlagwort
- d = Körperschaftsschlagwort
- f = Formschlagwort
- z = Zeitschlagwort

Unterfelder:

- \$a = Schlagwort oder Schlagwortkette (NW)
- \$v = Verweisungsformen (V)
- \$x = Regelwerk (NW)
- \$y = Herkunft (NW)

Dieses Feld ist für alle Einzelschlagwörter und Schlagwortfolgen aus verschiedenen Sprachen vorgesehen. Regelwerk (UF x) und Herkunft (UF y) sind wenn irgend möglich anzugeben.

Subject headings (insbesondere LCSH) sollen möglichst in 740, nicht in 711 angereichert werden. Sie sind daran erkennbar, dass für die Teile der mehrteiligen Strings die Unterfelder genutzt werden.

Wenn in Schlagwortdaten Notationen, Nummern und Text vorhanden sind, soll nur der Text als Schlagwort in Feld 711 übernommen werden. Gültige Notationen sollen wo möglich in die korrekten Felder überführt werden. Das Einspielen verbaler Benennungen zu bereits vorhandenen Notationen wäre in selektiven Projekten denkbar, wobei die Verständlichkeit der Einträge und die Nützlichkeit der zusätzlichen Sucheinstiege abzuwägen wären.

5.3 Feld 740 (Subject headings)

Neben LCSH enthält Feld 740 auch englischsprachige Schlagwörter anderer Herkunft und gelegentlich auch Schlagwörter in anderen Sprachen.

Anreicherungen in diesem Feld sollen so regel- und formatgerecht wie möglich erfolgen, die Headings sollen also unter Nutzung der korrekten Unterfelder und möglichst nicht als Textstrings erfasst werden.

Indikatoren:

- p = Personal Name used as Subject
- c = Corporate Body Name used as Subject
- f = Family Name used as Subject
- n = Name and Title used as Subject
- t = Title used as Subject
- s = Topical Name used as Subject
- g = Geographical Name used as Subject
- u = Uncontrolled Subject

Unterfelder:

"Die Definition der Teilfelder wird durch das verwendete Regelwerk bestimmt" (MAB-Titel)
„Die Bedeutung der Unterfelder kann nicht angegeben werden, da sie durch das verwendete Regelwerk definiert wird“. Eine einheitliche Definition der Unterfelder für Herkunftskennzeichnung und Regelwerk ist erforderlich.³

5.4 Felder 902 ff. (RSWK mit GND-Verknüpfung)

Die RSWK-Daten haben absolute Priorität bei Schlagwort-Anreicherungen. Die GND-Nummernverknüpfung (nur unter Verwendung der überregionalen GND-Nummern) ist obligatorisch. GND-konforme Schlagwörter ohne Nummern, deren Nummern sich vor dem Import auch nicht ermitteln lassen, dürfen nur in Feld 711 angereichert werden.

Zusätzlich zu den GND-Verknüpfungen gibt es Zeitschlagwörter (Indikator z) und Forms Schlagwörter (Indikator f) als Freitext. Da die Quellsätze zum größten Teil nicht RDA-konform sind, werden auch Forms Schlagwörter (Indikator f) weiterhin angereichert, obwohl mit der Einführung von RDA die Formangaben nicht mehr in den Sacherschließungsfeldern hinzugefügt werden.

³ Prinzipiell sind in Feld 740 Unterfelder zulässig, sie könnten aber nur in Kombination mit dem Indikator beschrieben werden. Ggf. könnte man, bis in MARC auf Vorschlag der DNB generell Unterfelder für Herkunftangaben eingeführt sind, interimistisch 740 UF \$2 analog zu Unterfeld \$2 bei 700_ damit belegen.

5.5 Sonstige Felder

MAB 720 (Stichwörter) ist nicht mit Schlagwörtern zu belegen und nicht zusammen mit Schlagwörtern zu indexieren.

MAB 730 (PRECIS) wird nicht angereichert und nicht indexiert.

MAB 750 / 753 / 756 : Anreicherung mit "inhaltlichen Zusammenfassungen" (Abstracts usw.) in Textform (Feld 750 1. Zusammenfassung, Feld 753 2. Zusammenfassung, Feld 756 3. Zusammenfassung) möglichst je nach zutreffendem Indikator. Für die Indexierung und Suche werden alle Indikatoren zusammengefasst.

753 und 756 sollen nur dann belegt werden, wenn 750 schon mit einem anderen Text belegt ist. Die MAB-Felder 751, 752, 754, 755, 757, 758 sollen bei Anreicherungen nicht belegt werden.

ASEQ 064a (Angaben zur Art des Inhalts)

Die neuen Felder für RDA-konforme Formangaben bleiben bei reinen Sacherschließungs-Anreicherungen außer Betracht, soweit es um Altdaten geht. Bei der Anreicherung von RDA-Titelsätzen, die keine Forms Schlagwörter mehr in den Schlagwortfolgen enthalten, soll das neue Feld 064a für die Art des Inhalts ergänzend zur eigentlichen Sacherschließung ebenfalls einbezogen werden.

6. Konzept zur Herkunftskennzeichnung bei der Einspielung von Sacherschließungsdaten

Die Überlegungen zur Herkunftskennzeichnung beziehen sich auf die üblichen Anreicherungen, aber auch auf künftige Anreicherungen aus Culturegraph.

In folgenden für die Sacherschließung relevanten Feldern ist eine Herkunftsbezeichnung überhaupt möglich:

- Feld 700 (im Unterfeld 2, besonders wichtig in 700 blank)
- 705 (im Unterfeld A: von der LoC vergeben / von der DNB vergeben / von anderen Institutionen)
- Feld 711 (im Unterfeld y: Herkunft)
- Feld 740 ? (im Unterfeld 2: Regelwerk, Feld für die Herkunftskennzeichnung ungeklärt; vgl. 5.3)
- Feld 904, 909, 914 ff. (Erläuterungen zur 1.. 2., 3.,... Schlagwortfolge: Herkunft/Nutzung)

Herkunftskennzeichnungen außerhalb dieser Felder, also das Anhängen an inhaltliche Metadaten oder den Missbrauch anderer Felder, soll es nicht geben. In Zukunft muss darauf geachtet werden, dass Herkunftskennungen nicht in den Schlagwort- oder Notationsindex einfließen.

Von Fall zu Fall kann entschieden werden, zusätzlich zum Herkunftskennzeichen ein Anreicherungskennzeichen zu setzen, sofern damit ein Mehrwert zu erzielen ist (z.B. für statistische Auswertungen o.Ä.)

6.1 Feld 711, UF y

Die Herkunftskennzeichnungen werden bei Einspeicherungen vergeben, nicht im laufenden Betrieb.

Bei maschinellen Einspielungen sollen Herkunftskennungen immer mit übernommen werden, und zwar wenn irgend möglich in das dafür vorgesehene (Unter)Feld.

Manuelle Korrekturen in 7XX sollen nicht erfolgen.

6.2 Felder 904 ff.

Herkunftsbezeichnungen im RSWK-Feld 904 / 909 / ... haben die Form der ISIL der jeweiligen Bibliothek bzw. eine Verbundkennung; sie werden nicht allgemein gepflegt und sind nicht vor Manipulationen geschützt. Die Kennungen schützen lediglich die Schlagwortfolge durch maschinelles Überschreiben.

Bei maschinellen Einspielungen von RSWK-Daten mit GND-Verknüpfung, die intellektuell erstellt wurden, sollen Herkunftskennungen nicht generiert und vorhandene nur dann übernommen werden, wenn es dafür besondere fachliche Gründe gibt.

Eine eigene, über die ISIL bzw. Verbundkennung hinausgehende Kennzeichnung von GND-Schlagwörtern aus automatischer Erschließung im RSWK-Segment wird jedoch als zwingend angesehen.

Zu beachten: Ist das Feld 904 usw. belegt, werden im Rahmen des Aleph-Mechanismus zur Nachführung von Schlagwortfolgen die DNB-Schlagwörter additiv eingespielt.

AG Sacherschließung der Kommission für Erschließung und Metadaten

Stand: 10.02.2016