

Kommission für Erschließung und Metadaten des Bibliotheksverbundes Bayern (KEM)

Arbeitsgruppe Sacherschließung (AGSE)

**Protokoll der 3. Sitzung der Amtsperiode 2014-2016 am 23.06.2015
Bibliothek der HU Berlin, 9.00 – 15.45 Uhr**

Teilnehmer:

Herr Dr. Stefan Brecheisen, BVB- Verbundzentrale (Gast)
Herr Michael Franke-Maier, FU Berlin (Vorsitz)
Frau Eva-Maria Gulder, BSB München
Frau Sabine Hertel, FUB Berlin (Gast)
Herr Werner Holbach, BSB München
Frau Dr. Sonja Kobold, Bibliotheca Hertziana Rom
Frau Ute Ristau, HU Berlin (Gast)
Herr Christian Rüter, HUB Berlin
Frau Dr. Gabriele Schweikl, UB Regensburg (Protokoll)
Herr Eugen Underberg, UB Würzburg
Frau Monika Vogler, UB Augsburg

Entschuldigt:

Herr Dr. Gerhard Stumpf (UB Augsburg)
Frau Tatjana Merda (Bibliothek der TH Nürnberg)

1 Begrüßung, Formalia

Herr Franke-Maier übernimmt in Vertretung für Herrn Stumpf den Vorsitz. Er begrüßt die Anwesenden zur 3. Sitzung der Amtsperiode. Das Protokoll der letzten Sitzung wird verabschiedet.

2 Gesamtkonzept für Anreicherungen mit Sacherschließungsdaten im B3Kat

Grundlage der Diskussion ist das von Herrn Stumpf als Entwurf vorgelegte und von Herrn Franke-Maier ergänzte „Konzept zur Anreicherung mit Sacherschließungsdaten im B3Kat“.

Zu 3.3.:Anreicherung grundsätzlich nur, wenn das betreffende Feld im Zielsatz leer ist (keine Addition). Ausnahme wäre die Addition mit zusätzlichen Daten aus einem anderen Regelwerk, was an Hand der Belegung im Unterfeld x abgefragt werden könnte. Herr Brecheisen wird

gebeten, Machbarkeit und Aufwand für einen Abgleich an Unterfeld x mit additiver Anreicherung zu testen.

Zu Fußnote 1: Mit „Herkunftskennung“ ist hier die Regelwerkskennung gemeint.

Zu 5. Schlagwörter: Für und Wider einer Beibehaltung der Schlagwortfolgen als Textstring mit Trennzeichen „/_“ werden diskutiert:

- Intellektuell erzeugte Schlagwortfolgen, die den Inhalt oder Teilaspekte u.U. eindeutiger wiedergeben als Einzelschlagwörter
- Vergleich der Folgen und der Permutationen umständlich
- Im semantischen Netz werden Vergleiche und Vernetzungen immer mit Einzelschlagwörtern vorgenommen
- Bei Schlagwörtern in 711 ist das Regelwerk oft unbekannt
- Zu überlegen ist, ob die Folgen immer dann aufgelöst werden sollten, wenn das Regelwerk nicht bekannt ist
- Falls bei der Anreicherung eine Schlagwortfolge auf einen Zielsatz trifft, der ein Einzelschlagwort aus der Folge bereits enthält, könnte das Einzelschlagwort durch die Folge ersetzt werden

Das Konzept wird verabschiedet. Der Aspekt „Zerschlagen der Schlagwortfolgen“ wird nach lebhafter Diskussion vertagt.

Herr Franke-Maier arbeitet die Änderungen zusammen mit den unter TOP 3 beschlossenen Ergänzungen zur Erfassung der Herkunftskennzeichen ein. [Das aktualisierte Konzept wurde mittlerweile am 24.08.2015 per Mail verschickt, einzelne Punkte müssen noch geklärt werden, damit das Konzept am 30.10. in der KEM-Sitzung vorgestellt werden kann.]

Top 3 Konzept zur Herkunftskennzeichnung bei der Einspielung von SE-Daten

Dabei muss getrennt werden die Frage nach der Art und Weise, wie ein Schlagwort gebildet wurde (maschinell, intellektuell durch Autopsie oder intellektuell mit maschineller Unterstützung) von der Frage nach der Art der Einspielung. Maschinelle Einspielung wird die Regel werden.

Eine Kennzeichnung der Datensätze mit dem Hinweis darauf, dass hier schon mal eine Anreicherung vorgenommen wurde, würde bedeuten, dass diese Datensätze nicht nochmal angefasst werden müssen. Herr Brecheisen weist darauf hin, dass in Aleph beim Laden von Anreicherungen Felder auf bereits vorhandene Inhalte hin überprüft werden können.

Beschluss:

- Wenn es in Fremddaten Kennzeichnungen für eine maschinelle Indexierung gibt, sollen diese mit eingespielt und nicht gelöscht werden
- Dort, wo es die Möglichkeit gibt, soll im Feld 904 (bzw. 909, 914 ff.), 740 oder 711 eine Codierung für das Verfahren, das angewendet wurde, abgelegt werden.
- Keine Kennzeichnungen sind notwendig bei den Anreicherungen aus anderen Verbänden im Kontext des kooperativen Datenaustauschs.

- Die AGSE entscheidet von Fall zu Fall und wird dazu über jeden Prozess informiert.

Herr Brecheisen wird um eine Übersicht über laufende Batch-Läufe zu Anreicherungen gebeten.

4 Bereinigungen von angereicherten Daten in den Sacherschließungsfeldern

Herr Franke-Maier stellt eine Analyse der von Herrn Brecheisen verschickten Listen für die im B3Kat bereits vorhandenen verbalen Sacherschließungs-Daten in den Feldern 710, 711 und 740 vor:

Feld 711 b: Schreibfehler im Unterfeld x könnten durch eine Normierung der Regelwerks-Kennzeichnung eliminiert werden. Bei 234 Datensätzen handelt es sich um GND-Begriffe, die intellektuell bearbeitet werden können. DNB-Sachgruppen können nach 700_ verschoben werden.

Feld 711 blank: Liste sollte nach Häufigkeit des Vorkommens durchgesehen und entschieden werden, ob gelöscht oder umgesetzt wird. Beispiele: Forms Schlagwort electronic book in 23.620 Fällen kann gelöscht werden; \$\$x swd: Anreicherung mit SWD-Schlagwort in 900er Feldern, falls dort noch nicht vorhanden.

Feld 711 c: Herr Stumpf wird gebeten, zu klären, woher die Daten kommen.

Feld 720: Hier sollte geprüft werden, wo die Quellen der Belegung sind. Da Feld soll nicht aktiv benutzt werden.

Herr Franke-Maier bietet an, die Listen durchzusehen und Vorschläge für Bereinigungen zu machen, die mit möglichst wenig Aufwand durchgeführt werden können (TOP 50 jeder Tabelle), die Abstimmung soll per Mailverfahren erfolgen.

5 Desiderata bei der retrospektiven Anreicherung mit Sacherschließung im B3Kat

Die UB der FU Berlin plant, den alten Schlagwortkatalog, der seit der Gründung der UB bis 1989 geführt wurde, zu erfassen. Der Katalog enthält 1 Mio Kärtchen, davon 750.000 mit Schlagwortfolgen. Die Images sind nicht OCR-lesbar, Formalerschließung ist bereits erfolgt. Geplant ist eine Anreicherung in 711 mit Unterfeld x als Schlagwortfolgen (Permutationen sind Teil des Kartenbestands).

Fraglich ist noch, wann und ob über Culturegraph Anreicherungen von Sacherschließungsdaten möglich sind. Herr Franke will bei der EG SE diesbezüglich nachfragen.

6 Formangaben

a. Künftige Erfassung

In der von der DNB vorgelegten Liste „Normierte Begriffe zur Beschreibung des Inhalts“ sind sowohl Forms Schlagwörter nach RSWK enthalten als auch Begriffe aus RDA, die bei der

Beschreibung einer Ressource die Art des Inhalts angeben. Dem Vorschlag der AGFE, die grau unterlegten Begriffe obligatorisch durch die Formalerschließung und alle sonstigen, nicht ausgegrauten, Begriffe durch die Sacherschließung zu erfassen, wird zugestimmt. Damit werden mit Einführung der RDA in den RDA-Aufnahmen keine Forms Schlagwörter mehr in 9xx \$f erfasst sondern Feld 064 (a oder b) belegt, in alten RAK-Aufnahmen weiterhin Feld 9xx. Misch-Datensätze soll es nicht geben. Die bisher hinterlegte Liste der Formbegriffe soll aktualisiert, dabei die obligatorisch von der Formalerschließung zu vergebenden Begriffe an den Anfang gestellt werden.

b. Behandlung der Altdaten (Formschlagwörter) im B3Kat

Die Möglichkeit, Forms Schlagwörter, die als Sachschlagwörter (Formbegriffe mit Indikator s) an letzter Stelle einer Kette mit mindestens zwei Begriffen stehen, umzuwandeln in Forms Schlagwörter mit Indikator f, sollte geprüft werden. Die gleichzeitige Belegung von Feld 051 (Pos. 1 bis 3) muss gegeben sein.

c. Maschinelle RDA-Anpassungen

Im Freitext der Altdaten sollten Rechtschreibkorrekturen maschinell nachgezogen werden (Biographie/Biografie, Autobiographie/Autobiografie, Bibliographie/Bibliografie, Photographie/Fotografie, Geographie/Geografie). Ein Auftrag dazu ergeht an die Verbundzentrale, Frau Gulder nimmt Kontakt auf.

Da sich zum 1.10. mit der Einführung der RDA eine Anzahl von Begriffen ändert (Konferenzschrift statt Kongress), die nicht nur in 064, sondern auch in 9XX erfasst werden, erscheint eine Information über BVB-Info notwendig (Frau Gulder).

d. Anforderungen an die Suchsysteme

Facetten müssen RDA-konform werden, Forms Schlagwörter aus der thematischen Facette herausgenommen und dort untergebracht werden, wo die Formbegriffe nach RDA zu finden sind.

7 Neuausrichtung der verbalen Sacherschließung

- „Enges Schlagwort“: Präkoordinierte Begriffe (Beispiel: Autoethnografie durch Kombination von Autobiografie und Ethnomethodologie) sind den Nutzern in der Regel so bekannt. Andererseits bietet die Vergabe eines „weiten Schlagworts“ das Potenzial für die Verlinkung auch in andere Kataloge und andere Systeme. Zudem wollen kleinere Bibliotheken nicht immer das „enge Schlagwort“ vergeben. Die AGSE plädiert dafür, beide Möglichkeiten zuzulassen und das auch so zu kennzeichnen.

- Diskussion um Schlagwortfolgen:

Für den Verzicht auf Schlagwortfolgen:

- Indexierung von Einzelschlagwörtern, Vernetzung erfolgt mit Einzelschlagwörtern

- Automatische Erschließung erfolgt über Einzelschlagwörter
- Öffnung für nicht-bibliothekarische Anwender leichter möglich
- Inhaltsverzeichnisse liegen als Zusatzinformation vor
- Stellenwert für die Anzeige verloren gegangen

Für die Beibehaltung der Schlagwortfolge:

- Relevanzranking mit Hilfe der Stellung der Schlagworte innerhalb einer Folge
- Öffnung für nicht-bibliothekarische Anwender:
Die Vergabe von Schlagwörtern oder freien Deskriptoren durch Benutzer im Sinne des Social Tagging wird Zukunft haben, insbesondere wenn eine Verknüpfung zum Regelwerk bzw. der Normdatei hergestellt werden kann. Auf den Publikationsservern der Hochschulen werden durch die Benutzer freie Deskriptoren vergeben, die über eine Schnittstelle in den B3Kat übernommen werden könnten.

8 Rückblick auf den Bibliothekartag *entfällt*

9 GND

Frau Gulder berichtet, dass in der GND noch eine große Menge an Korrekturen anstehen, auch im Kontext der Rechtschreibreform. (Anm.: Die Korrektur der wichtigsten Begriffe ist mittlerweile erfolgt, wie Grafik, Autobiografie, Biografie, Bibliografie, Fotografie, Geografie). Das Match-and-merge-Verfahren für die Top-100-Begriffe wird weiter fortgeführt (z.B. bei Körperschaften Vorgänger/Nachfolger). Bei den Diözesen gibt es noch Umarbeitungsbedarf, sie sind derzeit noch als Geografika erfasst und müssen auf Körperschaftsdatensätze umgearbeitet werden. Dies soll ab Januar 2016 in Angriff genommen werden. Ebenso müssen die Datensätze für Universitäten umgearbeitet werden, da sich der Standardisierungsausschuss für eine Abschaffung der Normierung an dieser Stelle ausgesprochen hat. Dies wird in einer gemeinsamen Aktion nach dem Regionalprinzip bearbeitet, jede Universitätsbibliothek wird sich um die Korrektur der Datensätze der eigenen Universität kümmern. Die Bearbeitung wird mit Veröffentlichung des RDA-Releases 10.2015 begonnen.. Frau Gulder wird eine entsprechende Mail an die Verbundbibliotheken schicken.

Mit dem Vollumstieg auf RDA werden weitere Korrekturen notwendig. Für Spitzenorgane von Gebietskörperschaften (wie Pressestelle, Bürgermeisteramt), die derzeit im Datensatz für die Gebietskörperschaft als abweichender Name erfasst sind, sollen mit dem Vollumstieg der GND auf RDA jeweils eigene Datensätze erstellt werden. Dann müssen auch die mit diesen Datensätzen verknüpften Titel umgehängt werden. Für die Erstellung der neuen Datensätze für diese Spitzenorgane ist hier die ZDB gefragt.

Gerade bei Sachschlagwörtern fehlen noch etliche Oberbegriffe.

10 RVK

Die bisherige Redaktionsdatenbank (Masterdatenbank) wurde zum 15.03.2015 durch eine neu strukturierte Datenbank ersetzt und die Inhalte in die neue Struktur überführt. Die neue Redaktionsdatenbank zeichnet sich durch folgende Merkmale aus: Die Texte sind durchgängig in die UTF-8 Zeichenkodierung überführt, Hierarchiebeziehungen werden über Identnummern statt wie bisher über Notationen hergestellt und die Registereinträge sind über GND-Nummern mit der Gemeinsamen Normdatei (GND) verknüpft. Man trägt nun GND-IDs in das Register ein, und die Begriffe werden dann aus der GND geholt. Mehrgliedrige Schlagwörter, die als solche mit individueller GND-Nr. in der GND enthalten sind, werden auch in Zukunft aufgenommen und als solche im Register der RVK Online angezeigt.

Zur neuen Version gibt es Eingabemasken und Werkzeuge, die den Redaktionsprozess unterstützen und die Stabilität der RVK-Identnummern sicherstellen. Alle in der Analysephase erkannten Typen, nämlich Neumeldungen, Löschungen, Änderungen (Inhalt bleibt gleich) und Änderungen (auch Inhalt ändert sich), können mit der neuen Oberfläche durchgeführt werden.

Seit dem 15.03.2015 wird die inhaltliche Weiterentwicklung der RVK nur noch mit der neuen Redaktionsdatenbank betrieben.

Die Datenbank der RVK ist auf die neue Normdaten-gerechte Struktur umgestellt. Da die Programme, die aus dieser Datenbank den XML-Abzug erstellen sollen, mit Ende des zweiten Quartals 2015 nicht ganz fertig geworden sind, die laufenden Änderungen aber bereits in die neue Struktur eingepflegt werden, können die alten Programme nicht mehr verwendet werden. Der turnusmäßige update der RVK-Datenbank wurde deshalb um ein Quartal ausgesetzt, die Anwender darüber über die RVK-Mailingliste informiert.

Das Anwendertreffen wird am 30. September 2015 in Regensburg stattfinden.

11 Sonstiges

Der retrospektiven Anreicherung von MeSH- und NLM-Daten der Bibliothek der Charité Berlin wird zugestimmt.