

**Kommission für Erschließung und Metadaten
des Bibliotheksverbundes Bayern (KEM)
Arbeitsgruppe Sacherschließung (AGSE)**

**4. Sitzung der Amtsperiode 2011-2013 am 24.09.2012,
UB der Humboldt-Universität Berlin, 9.30 – 15.30 Uhr**

Protokoll

Teilnehmer:

Frau Braune-Egloff, Bibliothek der FU Berlin
Frau Dilber, BVB-Verbundzentrale (Ständiger Gast)
Frau Hertel, Bibliothek der HU Berlin
Herr Holbach, BSB (Protokoll)
Frau Dr. Kobold, Bibliotheca Hertziana, Rom
Herr Rüter, Bibliothek der HU Berlin
Herr Scholz, BVB-Verbundzentrale (Gast, 9.30-11.30 Uhr)
Frau Dr. Schweikl, UB Regensburg
Herr Dr. Stumpf, UB Augsburg (Vorsitz)

Entschuldigt:

Frau Dr. Boll, UB Würzburg
Frau Gulder, BSB
Frau Merda, Bibliothek der Ohm-Hochschule, Nürnberg
Frau Reitzle, UB Augsburg

1. Formalia, Endgültige Verabschiedung des [Protokolls der 3. Sitzung](#)

Frau Hendrix, Stellvertretende Direktorin und Leiterin der Abteilung für Medienerwerbung und –erschließung, begrüßt die Teilnehmer.

Zum Protokoll der letzten Sitzung gibt es folgende Präzisierung zu TOP 6 „Sacherschließung in der ZDB, Nacharbeiten für RVK-Notationen“: Für die inhaltliche Recherche braucht es nicht die Angabe, dass es sich um eine Zeitschrift handelt; es reicht vielmehr aus, die korrekten Notationen für den jeweiligen Inhalt anzugeben. Die RVK-Anwender werden gebeten, beim nächsten Anwendertreffen zu entscheiden, ob eine vereinfachte Notation ausreichend ist, oder ob die Notation für einen Zeitschriftenbereich vergeben werden soll. Auch Bereiche für andere formale Literaturgattungen wie z.B. Serien sollten analog behandelt werden. Die für August angekündigte Rücklieferung der Titeldaten der ZDB mit den Sacherschließungsdaten ist seitens der ZDB nicht erfolgt, da die gelieferten Sacherschließungsdaten noch gar nicht eingespielt worden sind. Vgl. TOP 10.

Das Protokoll der 3. Sitzung wird verabschiedet.

2. Seiten der [AGSE](#) auf der BVB-Kommunikationsplattform

Die Verbundkonventionen für die verbale Sacherschließung im B3Kat (Schlagwort-Verbundredaktion) liegen nach wie vor auf dem Server in Augsburg. Auf der BVB-Kommunikationsplattform sind gegenwärtig neben der Beschreibung der Grundzügen der Sacherschließungsarbeit insbesondere Informationen zur AG Sacherschließung (Mitglieder, Protokolle, Arbeitspapiere ...) sowie Links zu interessanten Websites und Internetressourcen zu finden. Herr Stumpf bittet die Teilnehmer, Vorschläge für die Erweiterung des Angebots (z.B. interner Bereich, Wiki, ...) zu überlegen und über die Mailingliste zu diskutieren.

3. Datenanreicherung

a. Pfeffer-Verfahren

i. Ziele und Ablauf des Projekts

Die Prämisse ist, dass Inhaltsanalyse und –beschreibung in aller Regel auf der Werkebene ansetzen. Das Ziel ist, Datensätze im B3Kat automatisch durch Sacherschließungselemente anzureichern, die aus Titelaufnahmen zu anderen Ausgaben des gleichen Werkes (manifestations und expressions im Sinne der FRBR) im B3Kat oder in anderen Verbundkatalogen vorhanden sind. Näheres kann der Präsentation von Prof. Pfeffer (<http://de.slideshare.net/MagnusPfeffer/2012-gfkl>) entnommen werden.

Prof. Pfeffer hat der AGSE bereits Ergebnislisten, getrennt für die Bereiche Schlagwörter und RVK-Notationen, zukommen lassen, die für alle in Frage kommenden „Werke“ die entsprechenden Cluster mit den ID-Nummern enthalten. Demnach könnten angereichert werden: 2.969.381 Titel mit RVK und 2.765.967 Titel mit Schlagwortfolgen. Nur die Clusterbildung wird von Prof. Pfeffer übernommen. Für die Anreicherung im B3Kat, die die Verbundzentrale durchführen muss, ist vorgesehen, nur solche Titelsätze anzureichern, die keine RSWK- bzw. RVK-Daten haben.

ii. Optionen für die Auswahl der zu berücksichtigenden Felder, Titelsätze, Provenienzen (Verbünde)

Die von Herrn Pfeffer gelieferten Daten entstanden durch Abgleich der Titel aus B3Kat, HeBis, SWB und HBZ. Aufgrund der MARC-Struktur der zugrundeliegenden B3Kat-Daten sind auch die Bände von mehrbändigen Werken mit in die Clusterbildung einbezogen worden (was die Zahl der anzureichernden Titel zwar erheblich erhöhen würde, aber inhaltlich unerwünscht ist).

Es wird beschlossen, die Qualität der Clusterbildung zu prüfen (Gehören alle Titelsätze zu einem Werk? Gibt es noch Einzelprobleme, die zu berücksichtigen sind?), auf jeden Fall aber die Bandsätze auszunehmen. Datenanreicherung soll nur zwischen h-Sätzen erfolgen. Nach Möglichkeit sollen die von Herrn Pfeffer ermittelten Daten (Clusterbildung) verwendet werden; ein „Nachbauen“ des Verfahrens sollte vermieden werden.

iii. Vorgaben für einen Algorithmus zur Auswahl der „Quellensätze“

Für die Anreicherung mit Schlagwörtern soll folgende Reihenfolge für die Übernahme von SW-Folgen beachtet werden:

1. B3Kat - 2. SWB - 3. HBZ - 4. HeBis

Die AG wird noch prüfen, welche SWW übernommen werden sollen, wenn es mehrere Titel im Cluster aus einem Verbund gibt. Denkbar wären z. B. die Optionen:

1. kleinste ID-Nummer (BV-Nummer) für B3Kat-Aufnahmen
2. Die längsten SW-Folgen
3. Zufallsprinzip

Ein Vergleich der SW-Folgen mit einer inhaltlichen Zusammenführung (innerhalb B3Kat) wäre zwar interessant, wird aber aus Aufwandsgründen zunächst nicht weiter verfolgt und ggf. zu einem späteren Zeitpunkt wieder aufgegriffen.

Form-Schlagwörter für die physische Form (z.B.: „CD-ROM“) werden bei der Anreicherung weggelassen (eine Liste wird geliefert). Zeitschlagwörter werden übernommen, obwohl es in Einzelfällen zu einer gewissen Unschärfe bei den verschiedenen Ausgaben/Auflagen kommen kann.

Eine Kennzeichnung, dass die Schlagwortfolgen aus einer maschinellen Anreicherung stammen, erfolgt nicht.

Die Schlagwörter für Belletristik-Titel werden nicht ausgenommen, obwohl manche Bibliotheken diese Titel nicht beschlagworten, um die Treffermengen für wissenschaftliche Literatursuche nicht zu verwässern.

Vor der Anreicherung mit RVK-Notationen soll erst eine Bereinigung im B3Kat vorgenommen werden (siehe TOP 11). Dies betrifft:

1. Schreibfehler
2. nicht mehr gültige (abgeschaffte) Notationen
3. Bereichsangaben
4. „Sonderwege“ einzelner Bibliotheken (noch zu prüfen)

Zumindest die Punkte 1 und 2 sollen gleich bearbeitet/bereinigt werden. Da durch die Anreicherung wieder falsche Notationen aus den anderen Verbänden hereinkommen werden, soll die Aktion danach wiederholt werden.

Die Reihenfolge für die Verbände ist:

1. B3Kat - 2. SWB - 3. HeBis - 4. HBZ

Notationen aus Titelaufnahmen innerhalb des B3Kat werden alle additiv angereichert (d. h. alle aus allen BV-Titelaufnahmen des Clusters), also aufsummiert und dedupliziert. Aus Aufnahmen anderer Verbände werden nachrangig nach dem Zufallsprinzip die Notationen einer Titelaufnahme genommen.

Notationen aus dem Bereich Musik werden ausgenommen, weil es ansonsten zu groben Fehlern kommt, da Noten, Tonträger und Bücher, die sich oft in einem Cluster finden, drei ganz verschiedene Notationsbereiche haben.

iv. Perspektiven für die „Institutionalisierung“ des Pfefferschen Ansatzes

Das Pfeffersche Verfahren zielt auf eine retrospektive Verbesserung des bestehenden Katalogs. Für den Neuzugang sollte das Verfahren gemäß der kooperativen Katalogisierung den Austausch der Daten gewährleisten. Dieses greift allerdings nur bei Aufnahmen für dieselbe Manifestation. Mit dem Projekt Culturegraph zeichnet sich aber mittelfristig eine weitergehende Möglichkeit der Datenanreicherung, auch retrospektiv, ab.

Zu überlegen ist, ob das Pfeffer-Verfahren auf andere Informationen ausgeweitet werden sollte. In Frage kämen:

1. Weitere Sacherschließungselemente, z.B. DDC-Notationen, DNB-Sachgruppen, sonstige verbale und klassifikatorische Sacherschließungselemente
 2. Verbesserungen im Bereich der Normdatenverknüpfung, z.B. maschinelle Umverknüpfungen von Namensätzen (Tn) zu Personensätzen (Tp)
- Die Entscheidung hierüber wird zunächst zurückgestellt. Das Verfahren wird zunächst auf Schlagwörter und RVK beschränkt.

b. Sonstige / nachfolgende Arbeiten

Es liegen keine konkreten Pläne für maschinelle Aktionen vor. Die nach der Pfeffer-Anreicherung verbleibenden Defizite müssen später analysiert werden.

4. Inhaltliche Suche im Gateway Bayern

a. Beurteilung der jüngsten Verbesserungen

Im Gateway Bayern wurden vor kurzem einige neuen Funktionalitäten im Bereich der sachlichen Recherche freigeschaltet.

- Suchfilter/Fachfacette DDC-Sachgruppe im Drilldown
Eine Treffermenge kann nun nach DDC (DNB-Sachgruppen) gefiltert werden. Als Datenbasis wurden alle MAB-Kategorien, die DDC-Notationen beliebiger Provenienz und Granularität enthalten, herangezogen (700_, 700b, 705_, 705a, 705b).
- Weitersuche („Lateralsuche“) mit RVK-Notationen
Mittels MouseOver-Technik werden den Benutzern die verbalen Benennungen (inklusive aller Oberklassen) angezeigt. Sie sind durch Anklicken direkt für eine Weitersuche nutzbar, die zu Titeln mit der gleichen Notation führt.

Die AGSE begrüßt prinzipiell diese Verbesserungen für die sachliche Suche. Es handelt sich erst einmal um einen Zwischenstand; weitere Verbesserungen sind notwendig. Die Verbundzentrale des BVB wird gebeten, diese mit der AGSE abzustimmen. Insbesondere soll eine Freischaltung zukünftig erst dann erfolgen, wenn die letztendliche Realisierung von der AG begutachtet wurde.

Die verbale Anzeige des ganzen Hierarchiebaums wird als sinnvoll und notwendig erachtet, obwohl sie manchmal etwas unübersichtlich ist. Auch die Einarbeitung von Schlüsselungen ist sinnvoll, damit die Weitersuche nicht ins Leere läuft. Für die Notationen ohne Benennung in RVK-online wird ein Platzhalter-Text vorgeschlagen.

b. Weitere Desiderate

Die Suche nach Nichtdeskriptoren (abweichenden Benennungen) funktioniert nicht bei der Suche mit Schlagwörtern (weder in der freien Suche noch mit der Option „Schlagwort“; auch bei der Suche nach Personen). Dagegen sind die abweichenden Namen von Personen bei der Personensuche, die auf MAB 100 ff. aufsetzt, zu finden. Dieser schon länger festgestellte Fehler muss jetzt endlich behoben werden.

c. Nutzung des Zeitcodes für die Suchverfeinerung nach „thematischer Zeit“

Ein Konzept hierzu liegt seit längerem vor,

vgl. http://waldkauz.bibliothek.uni-augsburg.de/kfe/mat/kes_p_06.pdf (Anlage 1).

Die Teilnehmer halten die Realisierung des Konzepts nach wie vor für sinnvoll. Neben den Angaben aus den Zeit-Schlagwörtern ist allerdings auch die Heranziehung der zeitlichen Angaben aus den GND-

Normdatensätzen notwendig. Eventuell könnten auch die analytischen BSB-DDC-Notationen ausgewertet werden, da diese auch den zeitlichen Aspekt erschließen.

Die Umsetzung wird aber aufgrund wichtigerer anderer Themen mit weniger hoher Priorität versehen.

5. Anreicherung der Katalogsätze in der Datenbank vs. Mappings und Anreicherungen im Index der Suchmaschine bzw. im Suchvorgang: Prioritäten am Beispiel von RVK und DDC-Sachgruppen

Die AGSE und die Vertreter der Verbundzentrale befürworten die Anreicherung um die DNB-Sachgruppen im Aleph-Feld 700_ mit einem noch festzulegenden Suffix (analog zu DNB und ZDB). Eine vollständige Konkordanz von RVK-Grobstellen auf die DNB-Sachgruppen existiert; Frau Schweikl wird diese den Sitzungsteilnehmern zukommen lassen. Zusätzlich hat die Verbundzentrale den Versuch einer Konkordanz von RVK-Notationen auf die DNB-Sachgruppen unternommen. Basis für das maschinelle Verfahren ist das häufige gleichzeitige Vorhandensein bestimmter Notationen aus beiden Klassifikationen in den Titeldatensätzen (Kookkurrenzanalyse).

An der UB Regensburg sind Dissertationen auf einem speziellen Standort einem Fach zugeordnet, indem die Signatur den ersten Buchstaben einer RVK-Notation enthält. Damit ließe sich eine zuverlässige Zuordnung zu einer DNB-Sachgruppe herstellen. Die UB Regensburg wird der Verbundzentrale eine Liste der BV-Nummern (> 200.000) liefern. Die AGSE befürwortet die Anreicherung in den Titelaufnahmen (Feld 700_) und bittet die Bibliotheken im BVB und KOBV zu prüfen, ob weitere Bestände mit entsprechenden lokalen Signaturen für dieses Verfahren in Frage kommen.

Die AGSE wird des Weiteren prüfen, ob sich aus den Metadaten für bestimmte E-Book-Pakete (Produktsigel, auch aus Nationallizenzen) ebenfalls Möglichkeiten für ein Grob-Mapping auf die DNB-Sachgruppen ergeben.

Zu diesem Themenkomplex wird am 8.10.2012 ein Workshop mit Mitarbeitern der Verbundzentrale, der UB Regensburg und der AGSE in Regensburg stattfinden.

6. Erfahrungen und Desiderate aus der Arbeit mit der GND

a. Stand der GND-Bereinigung (Match & Merge)

Die Bereinigungen bei den Körperschaften sind mittlerweile erfolgt; als nächstes folgen die Geographika. Die Verbundzentrale prüft, ob seitens BVB/KOBV auch zwei Pakete pro Woche verarbeitet werden könnten.

Bei manuellen Umlenkungen ist bis auf Weiteres Zurückhaltung geboten. Nach Abschluss der maschinellen Läufe wird es Listen für die manuellen (verteilten) Bereinigungen geben.

Die AGSE fände es gut, wenn die DNB bis spätestens Jahresende eine Zusammenstellung der erfolgten und der nicht maschinell möglichen Korrekturen bereitstellen würde.

Die DNB plant für den 8.10.2012 eine Telefonkonferenz mit den Verbänden, um das weitere Vorgehen zu besprechen.

b. Redaktionsverfahren im B3Kat

Die Redaktionsanleitung findet sich im GND-Handbuch („Anhang für Redakteure“).

Bei Anhebungen von Datensätzen auf Level 1 soll das Selektionskennzeichen herausgenommen werden, da die „1“ die Aufforderung an die Verbundredaktion darstellt, den Satz zu ergänzen und

zu autorisieren.

c. Fehlverknüpfungen

- i. Das Problem der unzulässigen Verknüpfungen bei den Hinweissätzen ist überschaubar (ca. 300 Fälle); sie werden manuell korrigiert. Die anderen Verbände werden noch einmal auf das Problem hingewiesen.
- ii. Das Problem, dass bei der Z39.50-Übernahme aus anderen (PICA-)Verbänden bei der GND-Nummer das Präfix „/gnd“ statt nur der Nummer geliefert wird, ist bekannt. Die Verbundvertreter bringen das in die AG KVA ein.
- iii. Die Vorgaben für die Indexierung führen dazu, dass der Index in der GND oft unübersichtlich ist. Beispiel: „Drama“ als Sachschlagwort und als Ort in Griechenland stehen nebeneinander, was ein sehr genaues Hinschauen bei der Erschließung erfordert. Bei Klöstern und Kirchen ist durch den Wegfall der Ortsgebundenheit der Index aufgrund der fehlenden Unterscheidungsmerkmale ebenfalls unübersichtlich geworden. Dies beklagen namentlich die Bibliotheken des Kunstverbundes, es behindert aber generell eine präzise und schnelle Erschließung. Hier sollte nach den Match&Merge-Läufen eine manuelle Nachbearbeitung durch entsprechender Hinzufügung bzw. Indexeinträge für die Ortsnamen erfolgen.
Über BVB-INFO sollen die Sachverhalte etwas ausführlicher vermittelt werden.

d. Unverknüpfte Schlagwörter

In den Titelsätzen war bis vor kurzem in den Feldern 902ff. das Abspeichern unverknüpfter Texte nicht nur mit den Unterfeld-Indikatoren f und z möglich, sondern auch mit fast allen anderen Indikatoren. Die Verbundzentrale hat dies mittlerweile stark entschärft, da nun auch bei dem Indikator \$a eine rote Fehlermeldung kommt. Dies betrifft immerhin ca. 99% der Fälle.

e. Probleme mit Fremddaten: siehe TOP 6c, ii

7. Sonstige Fragen der Sacherschließung im B3Kat

- a. Wenn das Unterfeld \$1 belegt ist kommt es zu Mehrfacheinträgen im Schlagwortindex. Das durch DNB-Importe belegte UF soll regelmäßig gelöscht werden.
- b. Korrekturvereinbarung / Urheberkennungen**
Herr Stumpf wird eine kurzgefasste schriftliche Regelung der gegenseitigen Korrekturberechtigung im B3Kat auf die SW-Webseiten stellen. Es wird klargestellt, dass die nicht mit einer Urheberkennung versehenen SW-Folgen von allen nach bestem Wissen und Gewissen korrigiert werden dürfen, ja sollen, wenn sie Fehler enthalten. Werden auf Grund abweichender Inhaltsanalyse oder spezifischer Erschließung andere Sucheinstiege gewünscht, sollen diese i. d. R. durch Hinzufügung weiterer SW-Folgen eingebracht werden. SW-Folgen der FU Berlin mit ihrer Urheberkennung DE-188 sollen von anderen Bibliotheken nicht korrigiert werden. Die Urheberkennungen der DNB im Feld 904_ werden (außer in Sätzen aus der ZDB) nicht importiert.

8. MAB-Feld 740u

Das gesamte Feld 740 soll den Subject headings der LoC vorbehalten bleiben. Bzgl. der fehlerhaften Einträge in Feld 740u wird folgendes beschlossen:

1. Alle 3-stelligen Zahlen (= DNB-DDC-Sachgruppen) werden nach Feld 700_ (ohne Suffix) transferiert.
2. Alle anderen Einträge werden gelöscht. Ausnahme: wenn Feld 025L (bedeutet Herkunft LoC) belegt ist, denn dann ist davon auszugehen, dass das Feld 740u einen LCSH-affinen Inhalt hat.
3. Die Belegung des Feldes 740u mit VLB-Daten via DNB-Neuerscheinungsdienst sollte nach Abschaltung des Z39.50-Zugriffs gestoppt sein. Aus dem Fremddatenpool können diese Daten lt. Frau Dilber nicht stammen.

9. Regelwerksfragen RSWK; durch die EG Sacherschließung zu klärende Fragen

entfällt

10. Sacherschließung in der ZDB, Nacharbeiten für RVK-Notationen

Die Erfassung von Schlagwörtern und Notationen der im B3Kat Verwendung findenden Klassifikationen/ Fachsystematiken erfolgt in der ZDB. Die bereits Ende letzten Jahres an die ZDB/DNB gelieferten Daten für die retrospektive Ergänzung wurden entgegen anders lautender Versprechungen bisher noch nicht in die ZDB eingespielt. Dies führt dazu, dass im B3Kat der entsprechende Überschreibschutz für diese Felder nicht aufgehoben werden und eine Deduplizierung der Feldinhalte nicht erfolgen kann.

Die BSB wird noch einmal die ZDB nachdrücklich zu einer schnellen Einspeisung der Daten auffordern *[ist inzwischen geschehen; die Notationen sind eingespielt worden.]*

11. Bereinigung der fehlerhaften Einträge in 700g (RVK) im B3Kat

Herr Stumpf hat den Anfang der nach Häufigkeit geordneten Liste bereits kommentiert vorgelegt. Die Teilnehmer werden sich per E-Mail abstimmen, welche obsoleten Notationen gelöscht werden können. Diese Liste wird dann an die RVK-Redaktion übermittelt.

Die RVK-Anwendergruppe wird gebeten, darauf hinzuwirken, dass lokale Sondernotationen, auch wenn sie als Anhänge zu RVK-online registriert sind, nicht im Verbundkatalog, sondern nur im Lokalsystem erfasst werden.