

Arbeitsgruppe Sacherschließung (AG SE)

11. Sitzung am 07.10.2021

Teilnehmende (inkl. Gäste)

- Alter, Dirk (Gast, bis TOP 10.4.)
- Franke-Maier, Michael (ab 9:00 Uhr)
- Holbach, Werner (Vorsitz)
- Dr. Kobold, Sonja (ab 9:15 Uhr)
- Dr. Magen, Antonie (Protokoll)
- Merker, Kathrin (Gast, bis TOP 10.4.)
- Rüter, Christian
- Dr. Schweikl, Gabriele
- Thiessen, Peter (Ständiger Gast)
- Dr. Völkl, Martin
- Wolf-Dahm, Barbara

entschuldigt:

- Dr. Ziegler-Stryczek, Birgit

Zeit & Ort

Datum: Do., 07.10.2021

Zeit: 08.30-12.00 Uhr

Ort: Videokonferenz

TOP 1: Begrüßung, Formalia, Protokollführung, Protokoll der letzten Sitzung, Tagesordnung

Das Protokoll der letzten Sitzung, das bereits auf den Seiten der AG SE des BVB veröffentlicht ist, wird offiziell bestätigt.

TOP 2: Anreicherung von Sacherschließungsdaten per Culturegraph

Die erste Phase der Anreicherung von Sacherschließungsdaten via Culturegraph ist seit Anfang August abgeschlossen. Die Zahl der angereicherten Titeldatensätze liegt im Millionenbereich und wurde über BVB-Info veröffentlicht. Die Anreicherung wurde im Großen und Ganzen positiv aufgenommen, was eine Vielzahl einschlägiger Rückmeldungen belegt.

Von drei Fehlermeldungen abgesehen, verlief die Umsetzung von Phase 1 unproblematisch. Korrekturbedarf gab es lediglich bei der Anreicherung von Patentschriften und Inkunabeln. Bei Letzteren erwies sich aufgrund der historischen Besonderheiten dieser speziellen Buchgruppe die

Anreicherung von Ausgaben auf Werkebene als ungünstig. Beide Fälle hat Herr Thiessen bereits bereinigt. Die Rücknahme der Anreicherung war aufgrund der Provenienzzangaben (in Kombination mit den Gattungsbegriffen „Patentschrift“ bzw. „Inkunabel“) technisch unproblematisch.

Der dritte Fehler besteht darin, dass die in den Provenienzzangaben genannte URL bei der DNB noch nicht zu den gewünschten Informationen zur verwendeten Bündelung führt. Herr Thiessen hat bereits eine Mail mit der Bitte um Aktualisierung der entsprechenden Webseite an die DNB gesendet. Da er aber noch keine Rückmeldung erhalten hat, möchte er der Sache noch einmal nachgehen.

Die zweite Phase der Anreicherung von Sacherschließungsdaten via Culturegraph besteht in der Anreicherung von RVK-Notationen innerhalb des B3Kat-Bestandes. Beschlossen wird, dass hierfür zunächst die Migration der RVK-Notationen von ASEQ 700g nach ASEQ 701 (s. TOP 4) durchgeführt werden soll. Ist diese erfolgt, sollen

- a) alle Titeldatensätze, die noch über keine RVK-Notationen verfügen, und
- b) Titeldatensätze, die bereits über RVK-Notationen verfügen durch weitere additiv

angereichert werden.

TOP 3: B3Kat-Aktionen

3.1. Übernahme von Sacherschließungsdaten der Printausgabe für E-Book-Metadateneinspeicherung von Verlagsdaten

In einer ersten Testumsetzung, in der Sacherschließungsdaten der Printausgabe für E-Book-Ausgaben übernommen wurden (vgl. dazu Protokoll der 9. Sitzung der AG SE TOP 6.4. und Protokoll der 10. Sitzung TOP 5.1.), wurden insgesamt 20.221 Schlagwortfolgen angereichert. Die Zahl der angereicherten Datensätze, die vorher über keinerlei Schlagwortfolgen verfügten, beläuft sich dabei auf 10.063.

Diese erste Testumsetzung beruht auf einem Matching der Print- und elektronischen Ausgabe über ISBN. Dieses Verfahren wird auch zukünftig die Grundlage einer zu etablierenden technischen Routine sein. Die Selektion über im Vorfeld definierte Einzelpakete, die grundlegend für den Probelauf war, erwies sich hingegen als aufwändig und sollte für ein Routineverfahren nicht übernommen werden. Als Alternative schlägt Herr Thiessen eine reguläre vierteljährliche Anreicherung vor. Da die Mengen in jedem Fall überschaubar sind, ist auch voraussichtlich die Direktversorgung der Lokalsysteme garantiert, bzw. gut sicherzustellen. Dieser Vorschlag wird angenommen und mit der Absichtserklärung verbunden, dieses Procedere über eine BVB-Infomail an die Anwender*innen zu kommunizieren.

3.2. Löschung von nicht in der GND enthaltenen SWW in den Feldern 902ff.

Analog zu dem in der letzten Sitzung gefassten Beschluss, dass Formschlagwörter, die nicht in der GND enthalten sind, aber fälschlicherweise in den Feldern 902ff. abgelegt worden sind (vgl. Protokoll der 10. Sitzung, TOP 5. 2), in Feld 711 verschoben werden sollen, wird beschlossen, dass so auch mit allen übrigen nicht GND-konformen Schlagwörtern verfahren wird, die fälschlicherweise in den Feldern 902ff. abgelegt worden sind.

3.3. Bereinigung von falsch s-codierten Formangaben und Zeitschlagwörtern

Im Rahmen der Culturegraph-Anreicherung sind Herrn Völkl Erfassungsfehler (falsch s-codierte Formangaben und Zeitschlagwörter) aufgefallen, die durch die Culturegraph-Anreicherung systematisch multipliziert werden. Um die Culturegraph-Anreicherung möglichst sauber zu halten, sollten diese Fehler in einer gesonderten Bereinigungsaktion behoben werden.

Im Fall der falsch s-codierten Formangaben befinden sich diese in der Regel am Ende einer Schlagwortfolge. Für die Bereinigung dieser Fälle ist ein Abgleich mit Feld 064a erforderlich. Befindet sich hier eine Formangabe, die dem am Ende der Schlagwortfolge stehenden, gleichlautenden s-Schlagwort in Feld 902ff. entspricht, ist in aller Regel davon auszugehen, dass dieses Schlagwort eine falsch s-codierte Formangabe darstellt und gelöscht werden kann.

Eine Ausnahme bilden hier lediglich die beiden Begriffe „Kongress“ und „Ausstellung“. Befindet sich einer dieser Begriffe als s-codiertes Schlagwort am Ende einer Schlagwortfolge, muss abgeglichen werden, ob im selben Titeldatensatz das Feld 064a mit der Formangabe „Ausstellungskatalog“ bzw. „Konferenzschrift“ belegt ist. In diesen Fällen handelt es sich bei dem entsprechenden Schlagwort in den Feldern 902ff. mit hoher Wahrscheinlichkeit um ein falsch s-codiertes Forms Schlagwort und kann ebenfalls gelöscht werden.

Um Fehler bei falsch s-codierten Zeitschlagwörtern zu identifizieren, müssen alle Folgen geprüft werden, in denen sich der s-codierte Begriff „Geschichte“ befindet. Auch hier ist mit einiger Wahrscheinlichkeit davon auszugehen, dass der Begriff „Geschichte“ dann falsch s-codiert ist, wenn er sich am Ende einer Schlagwortfolge befindet (wohingegen das Auftauchen des Begriffs an anderer Stelle der Folge die richtige s-Codierung wahrscheinlich macht). Wenn sich also das s-codierte Schlagwort „Geschichte“ am Ende einer Schlagwortfolge befindet, soll diese mit den anderen Schlagwortfolgen in dem betreffenden Titeldatensatz verglichen werden. Befindet sich in einer anderen Schlagwortfolge desselben Titeldatensatzes ein z-codiertes Zeitschlagwort „Geschichte“, sollte die Schlagwortfolge, in der das Schlagwort „Geschichte“ s-codiert ist und am Ende steht, mit der falschen s-Codierung gelöscht werden.

Das Bereinigungsverfahren wird beschlossen.

3.4. Übernahme von Kataloganreicherungen und Sacherschließungsdaten aus anderen Verbänden

Ein Überblick über die augenblickliche Übernahme von Kataloganreicherungen und Sacherschließungsdaten aus anderen Verbänden zeigt, dass derzeit die Kataloganreicherung aus HEBIS, SWB, Casalini monatlich erfolgt, aus der DNB und von Amalivre wöchentlich; unser Medea reichert wöchentlich bis vierzehntägig an. Die Nachnutzung bzw. Anreicherung von Schlagwörtern erfolgt derzeit über DNB, hbz, OBV. Da der SWB seit geraumer Zeit nicht mehr MAB-, sondern MARC-Daten liefert, wird aus diesem Verbund derzeit nichts übernommen (zum Stand s. Anlage).

K10Plus-Daten werden aktuell nicht für die Schlagwort-Anreicherungen im B3Kat verwendet, da das entsprechende Verfahren noch auf der Lieferung von MAB-Daten basiert. Es besteht aber die Hoffnung, dass das bald – durch eine Erweiterung des Verfahrens zur Verarbeitung von MARC-Daten – wieder der Fall sein wird. Da allgemeiner Konsens besteht, dass die Datenanreicherung über Culturegraph (s. TOP 1) und die RVK-Übernahme (s. TOP 4) höher priorisiert sind, wird eine Routine für die Datenübernahme (Schlagwortanreicherung) aus K10Plus und ggf. anderen Verbänden derzeit noch zurückgestellt. Grundsätzlich erklärt sich Herr Thiessen aber bereit, im Namen der

Verbundzentrale zu klären, welche anderen Verbünde als Datenlieferanten in Zukunft noch in Frage kommen könnten.

TOP 4: Migration der RVK-Notationen von ASEQ 700g nach ASEQ 701

Die in der letzten Sitzung der AG SE geforderte „Positivliste“ (Notationen, deren Bedeutung sich in den letzten 10 Jahren nicht geändert hat) liegt inzwischen vor. Erstellt wurde sie dankenswerterweise von Herrn Dr. Albert Schröder (UB Regensburg), der auch das Vorgehen, das der Erstellung der Liste zugrunde liegt, dokumentiert hat. Aus dieser Liste geht hervor, dass sich im Vergleich vom ersten Quartal 2010 zum ersten Quartal 2021 von 850.000 Notationen 663.000 Notationen nicht verändert haben. Der Großteil der Änderungen fand zwischen 2018 und 2020 (hier v. a. Schlüsseländerungen in den Systematikbereichen T und Z) statt. 80% der Notationen blieben stabil.

Eine Analyse der in 700g belegten Felder im B3Kat auf der Basis der „Positivliste“ durch Herrn Thiessen hat gezeigt, dass von aktuell insgesamt 10.844.152 belegten 700g-Feldern in nur 544.622 Fällen (5,02%) die Notation nicht in der „Positivliste“ vorkam – bzw. dass sie in 94,98% der Fälle (10.299.530 belegten 700g-Feldern) in der „Positivliste“ enthalten war.

Die von Herrn Thiessen erstellte Liste der Notationen und ihrer Häufigkeiten, die die „Positivliste“ nicht passieren, zeigt, dass die häufigsten Fälle Notationen sind, die in der RVK nicht (mehr) enthalten sind, v.a. Zeitschriften-Individualnotationen, Sonderwege, Notationen mit CSN oder Verschreiber.

Mit Hilfe der „Positivliste“ können nun alle Notationen, die sich nicht geändert haben, in 701g überführt und mit der Normdaten-ID angereichert werden.

Für das weitere Vorgehen werden folgende Schritte beschlossen, die umzusetzen Herr Thiessen gebeten wird:

1. Die Notationen aus der „Positivliste“, also „saubere“ Notationen, die sich nicht geändert haben, können von 700g nach 701g (MIT ID) überführt werden
2. Notationen, die nicht in der „Positivliste“ vorkommen, werden an der Normdatei vorbeigezogen:
 - a. Sind sie in der Normdatei noch vorhanden, werden sie OHNE ID im Feld 701i abgelegt. Das Konzept dieser Notationen ist mit hoher Wahrscheinlichkeit geändert worden, hier muss intellektuell korrigiert werden.
 - b. Sind sie in der Normdatei nicht (mehr) vorhanden, werden sie im Feld 701y abgelegt und können so in den Lokalsystemen noch gefunden werden (oft handelt es sich um Sonderwege oder Zeitschriften-Individualnotationen)

Nach der Umsetzung soll Feld 700g in Zukunft gesperrt und mit einer roten Fehlermeldung versehen werden. Das Thema „Umgang mit der Einspeicherung von Fremddaten“ wurde angesprochen und soll in der nächsten Sitzung vertieft werden.

TOP 5: Manuelle Belegung von Feld 701i

Da eine rote Fehlermeldung den Workflow in dem Moment behindern könnte, in dem alte Datensätze bearbeitet werden, wird beschlossen, dass das Feld 701i zukünftig mit einer grünen Fehlermeldung versehen werden soll.

TOP 6: Migration der sonstigen Altdaten von ASEQ 700# nach ASEQ 701#

Kontrovers diskutiert wird die Frage, wie die Migration sonstiger Altdaten von ASEQ 700 nach ASEQ 701 stattfinden soll. Aufgrund der von Herrn Thiessen zur Verfügung gestellten Auswertungsdateien „mapping 700_701“ sowie „700_UF2_hauf“ wird im Plenum eine Kategorisierung von in jedem Fall zu übernehmenden (grün markiert), ohne Verlust zu löschenden (rot markiert) und nochmals zu überdenkenden und ggf. zu korrigierenden (gelb markiert) Altdaten vorgenommen. Herr Franke-Maier und Herr Völkl erklären sich bereit, in einem gesonderten Gespräch die gelb kategorisierten Fälle 700_, 700b sowie 700d genauer zu analysieren. Alle anderen gelb klassifizierten Fälle sollten ebenfalls analysiert werden. Mitarbeiter*innen werden ihnen nicht zugewiesen, es erfolgt jedoch der eindringliche Aufruf an alle, sich dieser Fälle anzunehmen.

Im Vorfeld bereits genauer analysiert hatte Herr Völkl die Daten, die in der Datei „700_ohne_UF“ enthalten sind, und dabei fünf große Blöcke (DNB-Sachgruppen, falsch abgelegte RVK-Notationen, BISAC-Code, IfZ-Notationen, AHF-Systematik) ausgemacht. Diese könnten aller Wahrscheinlichkeit nach mit Hilfe einer Syntaxanalyse in einen Algorithmus übersetzt werden, mit dessen Hilfe eine Übernahme nach ASEQ 701 erfolgen könnte.

TOP 7: Aleph-Feldhilfen für Sacherschließungsfelder

Da im Rahmen der Culturegraph-Anreicherung Feld 904b\$a dafür genutzt wird, die Provenienzangaben aus MARC-Einspeicherungen zu dokumentieren, hat sich grundsätzlicher Klärungsbedarf der Felder 904_\$a mit Feldinhalt „b“ und 904a\$a ergeben.

In Feld 904a\$a kann die ISIL der erfassenden Bibliothek verzeichnet werden.

Feld 904_\$a mit Feldinhalt „b“ diente früher als Überschreibschutz für automatisch eingespielte Schlagwortfolgen aus der DNB und wird teilweise von den Sacherschließer*innen auch noch als solcher verwendet. Da aber Schlagwortfolgen schon seit einiger Zeit nicht mehr überschrieben, sondern nur additiv ergänzt werden, ist diese Verwendung im Grunde nicht mehr notwendig.

Es wird beschlossen, die Feldhilfe anzupassen und durch eine BVB-INFO-Mail zu kommunizieren. Frau Schweikl erklärt sich bereit, nach Rücksprache mit Herrn Thiessen den Feldhilfetext entsprechend anzupassen und die BVB-INFO-Mail zu formulieren.

TOP 8: Bericht aus der RVK

Der RVK-Bericht von Frau Häusler, dessen zentraler Punkt der Hinweis auf die Positivliste (s. TOP 4) ist, wird dankend zur Kenntnis genommen.

TOP 9: Berichte aus anderen Gremien (KEM, ET RAVI, GND-Ausschuss)

Herr Holbach wird die entsprechenden Protokolle über die Mailingliste verschicken.

TOP 10: Verschiedenes

10.1. Projekt: Korrektur unklarer Inhalte als Zeitschlagwort

Wird auf die nächste Sitzung verschoben.

Manuelle Erfassung von Keywords in den Feldern 7xx

Wird auf die nächste Sitzung verschoben.

10.2. Hinweis auf den neu erschienenen Sammelband "Qualität in der Inhaltserschließung"

Bei De Gruyter Saur erschienen ist soeben der von Herrn Franke-Maier herausgegebene Sammelband [„Qualität in der Inhaltserschließung“](#), in dem sich Beiträge von ihm, Frau Wolf-Dahm, Herrn Holbach und Herrn Thiessen befinden.

10.3. Hinweis auf Workshop „Computerunterstützte Inhaltserschließung“ am 10./11. Nov. 2021

Der Workshop [„Computerunterstützte Inhaltserschließung“](#), in dem Herr Holbach ein Referat halten wird, findet am 10.-11.11.2021 online statt.

10.4. Termin der nächsten Sitzung

Die nächste Sitzung der AG SE findet am 16.12. (9:30-13:00 Uhr) wiederum als Zoom-Konferenz statt. Der ursprünglich als weiterer Termin avisierte 4.11. wird hingegen aufgrund seiner zeitlichen Nähe zur 11. Sitzung der AG SE gestrichen.

10.5. Gäste in der AG SE / Subskribenten Mailing-Liste

Alle bisherigen Gäste sollen auch zukünftig an den Sitzungen der AG SE teilnehmen. Weitere Interessierte könnten als Gäste zu Sitzungen, auf denen Spezialfragen behandelt werden, fallweise eingeladen werden.

München, 28.10.2021

Anlage zu TOP 3.4.: Übernahme von Kataloganreicherungen und Sacherschließungsdaten aus anderen Verbänden

Kataloganreicherungen Stand 29.07.21:

HEBIS - monatlich

SWB - monatlich

DNB - wöchentlich

Ama Livre - wöchentlich

Casalini - monatlich

Unser Medea - wöchentlich bis 2-wöchentlich

Was die Nachführung/Anreicherung von Schlagworten anbelangt, sind aktuell folgende Quellen im Einsatz:

DNB (in der Regel wöchentlich)

hbz - unregelmäßig/mehrwöchentlich (+RVK)

OBV - unregelmäßig/mehrwöchentlich (+RVK)

Da der SWB seit geraumer Zeit nicht mehr MAB, sondern nur noch MARC liefert, ist er leider aus dem etablierten Verfahren herausgefallen.