

Arbeitsgruppe Sacherschließung (AG SE)

12. Sitzung am 16.12.2021

Teilnehmende (inkl. Gäste)

- Franke-Maier, Michael
- Hertel, Sabine (Gast)
- Holbach, Werner (Vorsitz)
- Dr. Kobold, Sonja
- Dr. Magen, Antonie (Protokoll)
- Merker, Kathrin (Gast)
- Christian Rüter
- Dr. Schweikl, Gabriele
- Thiessen, Peter (Ständiger Gast)
- Dr. Völkl, Martin
- Wolf-Dahm, Barbara
- Dr. Ziegler-Stryczek, Birgit (ab 12 Uhr)

entschuldigt:

- Alter, Dirk (Gast)

Zeit & Ort

Datum: Do., 16.12.2021

Zeit: 9.30-13.00 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.

Herr Holbach, der mit dem Jahreswechsel aus der AG SE ausscheidet, übernimmt bei dieser Sitzung das letzte Mal den Vorsitz. Sein bisheriger Stellvertreter, Herr Franke-Maier, der für den Rest der Sitzungsperiode bis Ende 2022 den Vorsitz übernehmen wird, steht ihm als Co-Moderator zur Seite.

TOP 2: Migration der RVK-Notationen von ASEQ 700g nach ASEQ 701

Herr Thiessen referiert den Stand der Dinge. Der Zeitplan, den er im November zur Vorbereitung auf die 12. Sitzung der AG SE in die Runde geschickt hat, wird derzeit ausgeführt und befindet sich auf

aktuellem Stand. Die erste und zweite Tranche sind bereits umgesetzt; die dritte wird am kommenden Wochenende (18./19.12.21) in Angriff genommen. Bearbeitet sind aktuell 2,2 Millionen von insgesamt 7,5 Millionen Datensätzen.

Sobald die Umsetzung abgeschlossen ist, wird Feld 700g gesperrt. In den Feldern 701i und 701y sollen keine neuen Daten mehr erfasst werden. Die AG SE verständigt sich darauf, kurz vor dem Ende der Umsetzung (d. h. ca. eine Woche vor dem Abschluss der letzten Tranche) eine BVB-Infomail zu versenden, die über den Abschluss der Migration und über die künftige Verwendung bzw. Sperrung der betroffenen Felder informiert. Frau Schweikl erklärt sich bereit, Anfang Januar einen entsprechenden Entwurf zu formulieren und in die Runde zu schicken.

Angerissen wird in diesem Zusammenhang auch der Umgang mit Fremddaten (s. Punkt 6 des letzten Protokolls), die grundsätzlich analog zur Migration der RVK-Notationen behandelt werden können. Sofern dies technisch nicht direkt bei Online-Übernahmen während der Katalogisierung zu realisieren ist, soll der Feldinhalt zunächst automatisiert in 701i umgesetzt und dann regelmäßig offline per Batch nach Feld 701g umgesetzt werden.

TOP 3: Anreicherung von Sacherschließungsdaten per Culturegraph

Die Anreicherung von Sacherschließungsdaten per Culturegraph ist definiert, ein einschlägiger Auftrag liegt der Verbundzentrale vor.

3.1. Bereinigungsaktionen zu Phase 1

Bei der Anreicherungsaktion zu Phase 1 sind einige Fehler aufgetreten. So macht Herr Franke-Maier die Beobachtung, dass in den Feldern 902ff. teilweise noch veraltete Forms Schlagwörter wie z. B. CD-ROM vorhanden sind, die richtigerweise in Feld 064b gehören würden. Noch von Herrn Stumpf existiert eine Liste mit einschlägigen Forms Schlagwörtern, die zu einem Abgleich herangezogen werden könnte. Es wird beschlossen, dass eine Bereinigungsaktion durchgeführt werden sollte. Diese Aufgabe wird in die Agenda 2022 übernommen und soll zu einem späteren Zeitpunkt systematisch bearbeitet werden.

Frau Kobold macht auf einen anderen Fehler aufmerksam, der ihr nach der Umsetzung von Phase 1 aufgefallen ist: Es kommt nun vermehrt vor, dass gleichlautende Personen (Bsp. Francis Bacon, Philosoph, 1561-1626, GND-Nr.: 118505696 und Francis Bacon, Maler, 1909-1992, GND-Nr. 11850570X) falsch angereichert werden. Bereits im Vorfeld der Sitzung hat Frau Kobold dieses Problem mit Herrn Thiessen per Mail diskutiert. Sie verspricht, den diesbezüglichen Mailwechsel mit Herrn Thiessen in die Runde zu schicken, was zum Zeitpunkt der Protokollerstellung bereits geschehen ist.

Wenn es sich bei der falschen Personenanreicherung um einen inhaltlichen Erfassungsfehler handelt, der auf einen intellektuellen Fehler in einem Einzeldatensatz zurückgeht, ist dessen Duplikation bei der Culturegraphanreicherung nicht zu vermeiden. Behoben werden kann dieses Problem allenfalls im Nachhinein bei denjenigen Bündeln, in denen zwar der gleichlautende Name, aber unterschiedliche GND-Nummern vorhanden sind. Hierfür erforderlich wäre ein GND-Abgleich. Allerdings lassen Häufigkeit und Umfang des Fehlers darauf schließen, dass es sich bei den falschen Anreicherungen nicht immer um inhaltliche Einzelfehler handelt. Vielmehr muss angenommen werden, dass bereits der Bündelung, die durch Culturegraph erfolgt, Fehler zugrunde liegen. Es wird beschlossen, die DNB nach den Algorithmen bzw. Verfahren zu fragen, durch die die Culturegraph-Bündel generiert werden, und den DNB-Ansprechpartner in eine der nächsten Sitzungen der AG SE einzuladen (nach Mög-

lichkeit vor Beginn von Phase 3 und 4 der Culturegraphanreicherung). Herr Franke-Maier erklärt sich bereit, die entsprechenden Erkundigungen einzuziehen.

Um zumindest die Anzeige der falsch angereicherten Schlagwörter in den lokalen OPACs zu vermeiden, wird darauf hingewiesen, dass sich durch die hinterlegten Provenienzdaten die Anzeige der durch Culturegraph angereicherten Sacherschließung bei Bedarf unterdrücken lässt. Dies wird aber von der AG SE nicht empfohlen.

Darüber hinaus wird beschlossen, vor Beginn von Phase 3 und 4 eine BVB-Info-Mail zu verschicken, in der zur Rückmeldung von Fehlern und Unstimmigkeiten aufgefordert wird.

3.2. Phase 2: Anreicherung von RVK-Notationen (B3Kat-Bestand)

Herr Thiessen referiert den Stand der Dinge: Grundlagen für Phase 2 sind bereits geschaffen. Im Februar nächsten Jahres wird er Testdaten in die Runde schicken, bei denen vor allem auch Datensätze, die den von Frau Kobold beobachteten Fehler aufweisen, Berücksichtigung finden sollen.

Da die Menge der angereicherten Datensätze für die Statistik der einzelnen Häuser interessant ist, soll für Phase 2 und alle weiteren Phasen der Culturegraphanreicherung außerdem eine institutionenbezogene Anreicherungsstatistik erstellt werden.

TOP 4: B3Kat-Aktionen

4.1. Löschung von nicht in der GND enthaltenen SWW in den Feldern 902ff. mit Indikator (f-codiert)

Ein Auftrag ist durch die AG SE formuliert (s. Punkt 3.2. des letzten Protokolls). Zur Präzisierung wird Punkt 3.2. des letzten Protokolls nochmals unter dem Gesichtspunkt diskutiert, ob GND-konforme Schlagwörter einer Schlagwortfolge in 902ff. bewahrt oder ob die Folge komplett nach 711 verschoben werden soll, sobald ein nicht GND-konformes Schlagwort in ihr enthalten ist. Da auch hierfür ein GND-Abgleich notwendig ist, erklärt sich Herr Franke-Maier bereit, zu formulieren, wie ein entsprechender Abgleich aussehen muss. Anhand dieser Formulierung wird Herr Thiessen im kommenden Jahr feststellen, ob und mit welchem Aufwand ein solcher Abgleich möglich wäre.

4.2. Bereinigung von falsch s-codierten Formangaben

Auf Grundlage des Beschlusses der letzten Sitzung (s. Punkt 3.3. des letzten Protokolls) wird die Umsetzung durch die Verbundzentrale erfolgen.

4.3. Bereinigung von falsch s-codierten Zeitschlagwörtern

Auf Grundlage des Beschlusses der letzten Sitzung (s. Punkt 3.3. des letzten Protokolls) wird die Umsetzung durch die Verbundzentrale erfolgen.

4.4. Importkonverter für Sacherschließungsdaten aus SWB/K10plus

Bei Datenübernahmen aus dem SWB bzw. K10+ werden die Library of Congress Subject Headings (LCSH) nicht korrekt übernommen. Die Verbundzentrale wird den Fehler im Import-Konverter in nächster Zeit korrigieren.

4.5. Mapping MeSH - GND

Herr Franke-Maier macht darauf aufmerksam, dass die DNB eine Konkordanz [zwischen MeSH und GND](#) zur Verfügung stellt und regt an, dieses Mapping auf Verbundebene nachzunutzen. Dieser Vor-

schlag findet allgemeine Zustimmung. Herr Franke-Maier erklärt sich bereit, zu diesem Zweck das Mapping der DNB zu analysieren und hier vor allem zunächst die eindeutig identischen Konzepte zu identifizieren, bei denen eine Anreicherung weitgehend unproblematisch ist. Die per Mapping-Anreicherungen generierten Daten könnten in Feld 950 übernommen werden, das den Vorteil bietet, Provenienzangaben zu hinterlegen. Dieses Projekt wird in die Agenda für 2022 aufgenommen.

4.6. Routine Anreicherung eBook-Pakete - BVB-Infomail

Herr Thiessen wird zum Jahreswechsel die Anreicherung der eBook-Pakete mit RVK-Notationen und GND-Schlagwörtern aus parallelen Printpublikationen für 2021 vornehmen und in diesem Zusammenhang das angewendete Verfahren beschreiben. Darüber informiert werden soll in einer BVB-Info-Mail, für die Herr Thiessen einen ersten Entwurf in die Runde schicken wird.

TOP 5: Migration der sonstigen Altdaten von ASEQ 700# nach ASEQ 701#

Wie in der letzten Sitzung besprochen (s. Punkt 6 des letzten Protokolls) wurden die in den Auswertungsdateien „mapping 700_701“ sowie „700_UF2_hauf“ gelb hinterlegten Fälle im Detail analysiert. Herr Franke-Maier und Herr Völkl übernahmen die Fälle 700_, 700b sowie 700d, Frau Magen die Fälle 700r und 700s, für die ihr Herr Thiessen in der Zwischenzeit Listen mit den einschlägigen Datensätzen zusammengestellt hat. Sie verspricht, diese Listen im Nachgang der Sitzung in die Runde zu schicken, was bei Protokollerstellung bereits geschehen ist.

Eine Analyse der Liste, die den Fall 700r enthält, ergab, dass es sich hierbei überwiegend um falsch hinterlegte RVK-Notationen und um falsch hinterlegte IfZ-Notationen handelt. Da es sich um eine überschaubare Anzahl von betroffenen Datensätzen handelt, wird beschlossen, die falsch abgelegten RVK-Notationen manuell zu bereinigen. Herr Rüter erklärt sich bereit, das zu tun. Die Liste mit den falsch abgelegten IfZ-Notationen soll an das Institut für Zeitgeschichte geschickt werden, damit die Kolleg*innen dort die Datensätze entsprechend bereinigen können.

Eine Analyse der Liste, die den Fall 700s enthält, ergab, dass es sich hier um überwiegend freie Schlagwörter (z. T. auch Formangaben) handelt, in wenigen Fällen auch um Subject Headings (d. h. Feld 700s entspricht 740s) und in noch weniger Fällen Notationen der Bayerischen Bibliographie.

Eine Analyse der Liste, die den Fall 700_ enthält, ergab eine Reihe von Fehlertypen (mit Blank angehängte Großbuchstaben am Ende der Notation, vermutlich aus der DNB für die Kennzeichnung für Belletristik, Kinderbuch und Schulbuch (Bsp.: _B_K_S); mit Blank angehängte Edition am Ende der Notation, bestehend aus Blank und 2 Ziffern (Bsp.: _19_20_21); zwei Notationen in Unterfeld \$a getrennt durch einen Blank; zwei Unterfelder mit \$a und zwei unterschiedlichen DDC-Notationen; Notationen anderer Systeme in DDC-Feldern), die Herr Franke-Maier und Herr Völkl als neuen AB 27.1. im Wiki dokumentiert haben. Ferner hinterlegt ist hier eine Liste mit durchzuführenden Korrekturen und einer Transferstrategie von Feld 700 nach Feld 701. Da es sich um ein besonders komplexes Feld handelt, wollen Herr Franke-Maier und Herr Völkl in einer nächsten Sitzung die weiteren Analyseergebnisse präsentieren. Auf deren Grundlage können dann alle inhaltlichen Fragen und ein Vorgehen für die Datenmigration diskutiert bzw. erarbeitet werden.

TOP 6: Migration der Altdaten von ASEQ 7xx# nach ASEQ 951#

Die Altdaten (verbale Inhaltserschließung), die bisher in den 7xxer-Feldern (710, 711, 720, 740) abgelegt sind, sollen perspektivisch in das neue Feld 951 überführt werden. In der Vorarbeit kann der Umsetzungsprozess analog zum laufenden Umsetzungsprozess der 700er-Felder stattfinden. Erfolgen sollte er nach Beendigung der in dieser Sitzung diskutierten Aktionen (s. o. Punkt 2). Allerdings müssen vor der Datenmigration die alten 7xxer-Felder analysiert und bereinigt werden. Hierzu wird Herr Thiessen in einem ersten Schritt der AG SE Listen mit den entsprechenden Daten zur Verfügung stellen. Identifiziert werden sollten hier zunächst möglichst unproblematische Fälle, die leicht und schnell umzusetzen sind.

Es wird beschlossen, dass alle Mitglieder der AG SE in Zweierteams je ein Feld übernehmen und bis Ende des ersten Quartals 2022 analysieren. Diese Analyse sollte für eine spätere Sitzung aufbereitet und vorgestellt werden, so dass sie im Anschluss im Plenum diskutiert und eine Transferoutine erarbeitet werden kann.

Herr Rüter übernimmt Feld 711, Herr Franke-Maier Feld 740. Die Felder 710 und 720 sind noch nicht vergeben, die Zweitbearbeiter*innen für die Felder 711 und 740 stehen ebenfalls noch aus. Es wird vereinbart, dass sich die Mitglieder der AG SE nach der Sitzung per Mail die Bearbeitung der Felder aufteilen. Die Feldanalyse sollte anhand der Feldbeschreibungen, der alten Protokolle, des Wiki (bei gleichzeitiger Aktualisierung des Wiki) erfolgen und in entsprechenden Wiki-Arbeitsbereichen dokumentiert werden. Herr Rüter erklärt sich darüber hinaus bereit, einen neuen, übergreifenden Arbeitsbereich zu diesem Thema im Wiki anzulegen, was zum Zeitpunkt der Protokollerstellung bereits geschehen ist.

TOP 7: Bericht aus der RVK

Der RVK-Bericht von Frau Häusler wird dankend zur Kenntnis genommen. Herr Franke-Maier weist in diesem Zusammenhang darauf hin, dass in der Beiratssitzung u. a. auch das Thema der gendergerechten Sprache diskutiert wurde und ergänzt, dass es auch einen Aufruf der DNB für eine einschlägige Arbeitsgruppe gibt. Er schlägt vor, Frau Häusler von dieser Arbeitsgruppe in Kenntnis zu setzen. Frau Schweikl verspricht, dies zu tun.

TOP 8: Berichte aus anderen Gremien (KEM, ET RAVI, GND-Ausschuss)

8.1. Berichte aus den Gremien

Die Berichte aus der KEM, dem ET RAVI und dem GND-Ausschuss erfolgen via Protokolle der genannten Gremien, die an die Mitglieder der AG SE zu verschicken sich Herr Holbach bereit erklärt.

Frau Wolf-Dahm weist darauf hin, dass sich der GND-Ausschuss derzeit mit der Frage beschäftigt, wie sich der GND-Plus-Bereich formieren soll. Diese und andere Themen sind aber noch im Fluss. Frau Wolf-Dahm verspricht, zu berichten, sobald Ergebnisse vorliegen.

8.2. Einstellung Patenschaftsmodell für eBooks

Kontrovers diskutiert wird die Einstellung des Patenschaftsmodells für eBooks. Kritisiert wird insbesondere das [Protokoll der AG FE vom 9.9.21](#), in dem es heißt, dass auf die Hochkatalogisierung von Springer-Ebook-Paketen zukünftig verzichtet werden kann, weil entsprechende Kapazitäten fehlen und zudem „[d]urch die Culturegraph-Anreicherungen [...]eine Sacherschließung in den meisten Fäl-

len gewährleistet“ sei. Die AG SE nimmt mit Befremden zur Kenntnis, dass die Culturegraph-Anreicherung hier als Argument für den Verzicht auf eine Sacherschließung der Ebooks verwendet wird.

Herr Franke-Maier hatte bereits im Vorfeld der Sitzung eine Stellungnahme zu diesem Protokoll, insbesondere zu der zitierten Passage entworfen, mit der sich die AG SE vom Beschluss der AG FE bzw. von der genannten Argumentation distanziert. – Die Culturegraph-Anreicherung kann mitnichten eine Sacherschließung von Ebooks ersetzen. In diesem Ergebnis ist sich die AG SE einig. Es wird beschlossen, den Entwurf von Herrn Franke-Maier mit Blick auf die Zuständigkeit der AG SE zu schärfen. Herr Rüter sagt eine entsprechende Überarbeitung zu, in der vor allem dargestellt werden soll, dass Culturegraph nicht die Sacherschließung von Ebooks ersetzt oder auch nur ersetzen kann. Einen ersten Umarbeitungsentwurf hat er bereits zum Zeitpunkt der Protokollerstellung mit der Bitte um Weiterbearbeitung in die Runde geschickt.

Zudem wird in diesem Zusammenhang überlegt, dass es interessant sein könnte, zu wissen, wie viel Prozent der Ebook-Pakete in Print vorhanden und sacherschlossen sind. So könne der Bedarf für die Sacherschließung bei Ebooks ermittelt werden.

TOP 9: Verschiedenes

9.1. Projekt: Korrektur unklarer Inhalte als Zeitschlagwort

Im Vorfeld hat Herr Völkl ca. 20 einschlägige Fehlertypen identifiziert. Für jeden Fehlertyp ist zwar ein Algorithmus denkbar, mit dem der Fehler behoben werden kann. Es stellt sich aber die Frage, ob die etwaige Programmierung der Algorithmen aufgrund der geringen Anzahl von Datensätzen, die überhaupt betroffen sind, in einem akzeptablen Verhältnis steht. Nach längerer Diskussion kommt die AG SE zum Schluss, dass das nicht der Fall ist und die Fehler manuell behoben werden sollten. Dazu sollen die fehlerhaften Datensätze in 500er-Paketen im Wiki abgelegt werden. Alle Mitglieder der AG SE werden aufgefordert, sich einzelne Pakete zur Bearbeitung zu nehmen und in ihrer Bibliothek zu verteilen. Die entsprechenden Entnahmen sollten im Wiki dokumentiert werden. Herr Franke-Maier erklärt sich bereit, im Januar die 500er-Pakete sowie eine Verwaltungsdatei im Wiki zu hinterlegen.

Ferner wird beschlossen, für die 902er-Felder eine Plausiprüfung zu hinterlegen, mit deren Hilfe unzulässige Zeitschlagwörter identifiziert werden sollen.

9.2. Manuelle Erfassung von Keywords in den Feldern 7xx

Manuelle Keywords sollen zukünftig in Feld 951 erfasst werden.

9.3. Datenfeld für freie Schlagwörter

Freie Schlagwörter sollen zukünftig in Feld 951 erfasst werden.

9.4. Termin der nächsten Sitzung

Die nächste Sitzung der AG SE sollte in ca. zwei Monaten (Ende Februar) stattfinden. Herr Franke-Maier wird ein Termin-Froodle aufsetzen.

Abschließend dankt Herr Franke-Maier als stellvertretender Vorsitzender Herrn Holbach für sein langjähriges Engagement in der AG SE und wünscht ihm im Namen der AG SE alles Gute für seinen Ruhestand.

München, 17.12.21