

Arbeitsgruppe Sacherschließung (AG SE)

Protokoll der 13. Sitzung am 21.02.2022

- Formalia
 - Teilnehmende (inkl. Gäste)
 - Zeit & Ort
- Tagesordnung
 - TOP 1: Begrüßung, Formalia, Protokollführung, Protokoll der letzten Sitzung, Tagesordnung
 - TOP 2: Migration der RVK-Notationen von ASEQ 700g nach ASEQ 701
 - TOP 3: Anreicherung von Sacherschließungsdaten per Culturegraph
 - TOP 4: Datenfeld für freie Schlagwörter / AG Metadaten für Forschungsdaten
 - TOP 5: B3Kat-Aktionen
 - TOP 6: Migration der sonstigen Altdaten von ASEQ 700# nach ASEQ 701#
 - TOP 7: Migration der Altdaten von ASEQ 7xx# nach ASEQ 951#
 - TOP 8 und 9: Berichte
 - TOP 10: Verschiedenes

Formalia

Teilnehmende (inkl. Gäste)

- Alter, Dirk (Gast)
- Franke-Maier, Michael (Vorsitz)
- Hertel, Sabine (Gast, Protokoll)
- Dr. Kobold, Sonja
- Koglin, Lydia (Gast, bis 9:15)
- Dr. Magen, Antonie
- Dr. Schweikl, Gabriele
- Thiessen, Peter (Ständiger Gast)
- Dr. Völkl, Martin
- Wolf-Dahm, Barbara

- Dr. Ziegler-Stryczek, Birgit

entschuldigt:

- Rüter, Christian
- Merker, Kathrin (Gast)

Zeit & Ort

Datum: 21.02.2022

Zeit: 8:30-12:30

Ort: Videokonferenz

Tagesordnung

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

Das Protokoll der letzten Sitzung wird offiziell bestätigt. Sabine Hertel übernimmt dankenswerterweise die Protokollführung.

1.1. Jahresplanung 2022, kurzes Brainstorming

Die AG SE beginnt die erste Sitzung im Jahr 2022 mit einem kurzen Brainstorming zu den Prioritäten des letzten Jahres der laufenden Amtsperiode. Dabei werden die folgenden Punkte genannt:

- Fortführung der Culturegraph-Anreicherungen (Phase 2, 3 und ggf. 4)
- Bereinigung der RVK-Notationen
- Bereinigung der Felder mit DDC-Notationen

Weiterhin werden während des Brainstormings folgende Fragen gestellt:

Martin Völkl fragt, ob die AG SE auch zuständig sei für eine Optimierung des Recall im Gateway Bayern. Dort werden derzeit die abweichenden Benennungen in der Recherche nicht berücksichtigt, was man in Katalogschulungen schlecht erklären kann. Peter Thiessen (BSB) verweist auf die Kommission für Service und Information (KSI) und wird eine:n Ansprechpartner:in ermitteln.

Dirk Alter fragt nach der Bereinigung der Forms Schlagwörter in den ZDB-Daten. Peter Thiessen (BSB) antwortet, dass der Konverter von der Verbundzentrale schon seit ca. 2 Jahren daraufhin eingestellt wurde, die Forms Schlagwörter resp. Formangaben nur noch korrekt zu importieren. Diese Lösung greift allerdings nur bei Datensätzen, die seither aktualisiert wurden oder noch werden. Seit Jahren ruhende unveränderte Datensätze und neue Übernahmen aus der BVB03 haben

noch den veralteten Stand. Die noch nicht aktualisierten Datensätze werden in der nächsten Zeit sukzessive maschinell am Wochenende in Paket aktualisiert, so dass perspektivisch immer weniger veraltete Formschlagwörter in den abnehmenden Systemen ankommen sollten.

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

701i / ohne ID-Nummer:

Notationen aus 700g, die nicht in der Positivliste, aber in der Normdatei enthalten sind, werden bei der Bereinigungsaktion in 701i ohne ID-Anreicherung übernommen. Bei diesen Notationen muss intellektuell überprüft werden, ob die inhaltliche Zuordnung noch stimmt (nicht in der Positivliste enthalten -> inhaltlich geändert).

701i / mit ID-Nummer:

Notationen aus 701g, die bei einem RVK-Update ungültig werden (gelöscht/inhaltlich verändert), werden mit der ID-Nummer nach 701i verschoben. Auch diese Notationen müssen intellektuell korrigiert werden.

Es erfolgt ein wöchentlicher Abgleich an der Positivliste, diese muss quartalsmäßig nach einem RVK-Update aktualisiert werden.

701y / ohne ID-Nummer:

Hierhin werden bei der Bereinigungsaktion alle Notationen aus 700g übernommen, die weder in der Positivliste noch in der Normdatei enthalten sind. Diese Notationen sind Sonderwege, Schreibfehler, sehr alte bereits gelöschte Notationen etc.

Es sind bereits 7 Mio Altdaten umgesetzt worden. In die Lokalsysteme wird nur das Feld 701g ausgeliefert.

Peter Thiessen (BSB) wird Zahlen für den wöchentlichen Datenverkehr ermitteln, da es aus Performancegründen nicht möglich ist, bei einer Online-Übernahme automatisiert Fremdnotationen an der Normdatei und der Positivliste vorbeizuziehen. Deswegen sollen RVK-Notationen bei Online-Übernahmen nach 701i geschrieben werden und ggf., abhängig von der Größe des Datenverkehrs wöchentlich prozessiert werden.

Michael Franke-Maier und Gabriele Schweikl werden den Entwurf für die BVB-Info-Mail dahingehend noch einmal anpassen (Info zum Datentransfer aufnehmen). Die Mail soll erst zeitnah zum Sperrtermin des Feldes 700g verschickt werden.

TOP 3: Anreicherung von Sacherschließungsdaten per Culturegraph

Peter Thiessen (BSB) teilt mit, dass die DNB jetzt monatlich Culturegraph-Bündel bereitstellt und somit ein aktuellerer Datenbestand vorliegt. Die erste Anreicherungsphase könnte damit wiederholt und dabei ca. 210.000 Datensätze für Anreicherungen berücksichtigt werden. Prinzipiell ist die AG SE dafür, allerdings wird dafür plädiert, den Sondertermin mit der DNB (Herr Grund und Frau Vorndran, s. TOP 3.1.) abzuwarten, damit die Problematik der falschen Bündelung bei einer erneuten Prozessierung berücksichtigt werden kann.

3.1. Bereinigungsaktionen zu Phase 1

Stand der Dinge bzgl. falscher Werkcluster

Es wird angenommen, dass hier Fehler bei der Erstellung der Werkbündel zugrunde liegen. Michael Franke-Maier wird einen Sondertermin zu der Thematik mit der DNB vereinbaren. Sonja Kobold und Martin Völkl werden für dieses Gespräch eine kurze Präsentation der aufgefundenen Fehler vorbereiten.

Erst nach diesem Austausch mit der DNB wird die Phase 2 angegangen.

Überschreibung der Provenienzkennzeichnungen in 904a \$aDE-188 und Qualität

Michael Franke-Maier wird eine Liste der Datensätze anfordern, bei welchen das Provenienzfeld 904a/909a ... \$ DE-188 gelöscht wurde. Die Verbundzentrale kann dieses Set ggf. benutzen, um intern die Culturegraph-Routine zu optimieren, da hier z.T. Fehler produziert wurden (Anreicherung von s-codierten Formangaben und Überschreiben von korrekten Formangaben). Die Routine wird entsprechend angepasst.

Die Liste mit BV-Nummern wurde bei Erstellung des Protokolls bereits an Peter Thiessen (BSB) verschickt.

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

verschoben auf Termin nach dem Sondergespräch mit der DNB.

3.3. Phase 3 & 4: Verbundübergreifende Anreicherungen

verschoben auf Termin nach dem Sondergespräch mit der DNB.

TOP 4: Datenfeld für freie Schlagwörter / AG Metadaten für Forschungsdaten

Die AG Metadaten für Forschungsdaten hat Optionen für das Format erstellt, um in der Regel frei von Autor:innen vergebene Schlagwörter unterzubringen. Es bleibt die Wahl zwischen MAB 711

und MAB 951. Die AGSE spricht sich für MAB-951-Felder aus, da dort künftig alle nicht-gnd-konformen Schlagwörter gehalten werden und diese Felder künftig alle 700er Felder ersetzen sollen.

951 kann bereits genutzt werden.

Christian Rüter (HUB) wird gebeten, das Datenformat insgesamt genau vorzustellen. Michael Franke-Maier wird einen Vorschlag für alle benötigten Unterfelder erarbeiten (siehe auch Wiki der AGSE: AB31 - Datenformat für Forschungsdaten):

951

Indikator = _

Unterfelder:

\$a = Schlagwort - z.B. „Nahe Osten“

\$l = Sprache - z.B. „ger“

\$0 = Identifier - sofern vorhanden, z.B. Wikidata-ID

\$2 = Quelle -

z.B. Wikidata (wird vermutlich nicht immer gehen, da feste Liste), aus der LoC-Liste der "Subject Heading and Term Source Codes"

\$P @@q = Herkunft - Bibliotheks-ISIL

\$C = Herkunft - re3data.org: ID des Repositoriums, für FU z.B. OpenDOAR:4732

Zusätzlich

064a \$a Forschungsdaten

\$9 (DE-588)1098579690

Offene Fragen:

- Persistent Identifier der Repositorien in \$P ablegbar oder nur ISIL?
- Gibt es bereits Repositorien, die eine Verknüpfung der SWW mit Normdaten vorsehen?
- Umgang mit der ORCID, wird an die AG FE verwiesen
- Versorgung der SISIS-Projekte

TOP 5: B3Kat-Aktionen

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

Michael Franke-Maier erarbeitet einen Vorschlag zur Verschiebung der noch nicht in der GND vorhandenen Sachbegriffe (derzeit mit "f" in den 902ff-Feldern codiert) in die 711er, bzw. 951er-Felder.

5.2. Bereinigung von falsch s-codierten Formangaben

Die Bereinigung ist erfolgt, aber noch nicht in die Versorgungssysteme gegeben.

5.3. Bereinigung von falsch s-codierten Zeitschlagwörtern

Die Bereinigung ist noch nicht erfolgt, Top 5.1 wird aus Synergie-Effekten abgewartet.

5.4. Importkonverter für Sacherschließungsdaten aus SWB/K10plus

Die Frage von Peter Thiessen (BSB) , ob der Import auch gleich in 951 erfolgen solle, wurde von der AG SE positiv beschieden, da man damit saubere Daten von weniger sauberen formattechnisch getrennt hält.

Die Versorgung der Lokalsysteme ist noch ungeklärt.

5.5. Mapping MeSH - GND

- Status: Michael Franke-Maier hat Kontakt mit DNB aufgenommen, Daten liegen derzeit nur intern vor, Datenlieferung erfolgt in Q1/2022. Eine Benachrichtigung wurde angefordert.

5.6. Routine Anreicherung eBook-Pakete - BVB-Infomail Mitte / Ende Januar

- Status: Die Mail wurde abgestimmt und am 26.01.2022 über die Mailingliste versendet.
- Weitere Fragen, die im Zuge der Kommunikation der Anreicherungsaktion aufkamen, siehe unten (TOP 10 Verschiedenes).

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

Christian Rüter (HUB) hatte die Bereinigung der falsch abgelegten RVK übernommen und erledigt.

Martin Völkl wird das Institut für Zeitgeschichte anschreiben und um Bereinigung der IfZ-Notationen bitten.

Martin Völkl hat die Daten im Feld 700_ ohne UF2 geprüft und bittet um Unterstützung bezüglich ihm unbekannter Klassifikationen.

Michael Franke-Maier hat das Feld 700b geprüft und mit 705 begonnen. In den Feldern finden sich vielfältige falsche Daten, angefangen mit Schmierzeichen, DDC-Notationen in falschen Unterfeldern, DDC-Editionsangaben in \$a (statt \$s), zwei und mehr DDC-Notationen in einem Feld usw. Es müssen Einzelentscheidungen getroffen werden, s. Wiki der AG SE: AB27.1 - Bereinigung und Transfer von DDC-Notationen.

Die AG SE spricht sich dafür aus, auch den Öffentlichen Bibliotheken im Verbund Rechnung zu tragen und Informationen, die für diese hilfreich sind, entsprechend aufzubereiten. In einem ersten Zuge konnte der Umgang mit Informationen der DNB zu Schulbüchern und Kinder- und

Jugendliteratur geregelt werden. Michael Franke-Maier arbeitet zusammen mit Martin Völkl konkrete Vorgaben für die Verbundzentrale aus.

Die Verbundzentrale liefert eine Datei mit Systemnummern aller Datensätze, in denen in 700b der *Großbuchstabe B* vorkommt. Die DNB nutzt diesen zur Einordnung eines Titels zur Belletristik, die LoC hingegen zur Einordnung als Biografie. Weitere Attribute müssen gesucht werden, damit eine korrekte Umsetzung in Formangaben erfolgen kann.

Michael Franke-Maier wird sich beim GBV bzw. beim Projekt coli-conc darüber informieren, wie viel Informationsgehalt der Slash in DDC-Notationen beinhaltet. Die Antwort liegt bereits vor:

Slash (/) und Apostrop (') werden im DDC-Feld 045F als Trennzeichen in Notationen verwendet, das werten wir aber bislang nicht aus, sondern normalisieren die Notationen zur Weiterverarbeitung.

> Wie bewertet ihr a) den Informationsgehalt dieser Angabe und b) für wie wichtig haltet ihr den Erhalt des Slashes?

Damit haben wir uns noch nicht weiter beschäftigt, vielleicht können die Zeichen später noch hilfreich sein, vielleicht auch nicht. Ich würde die Sonderzeichen im K10plus lassen, wo sie sind. Überall wo wir durch Anreicherung selber DDC-Notationen eintragen, geschieht dies aber in normalisierter Form.

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

siehe Wiki der AG SE: AB14 - Migration der sonstigen Altdaten von ASEQ 7xx# nach ASEQ 951#

Michael Franke-Maier hat die Patenschaft für die 740 Felder übernommen.

Christian Rüter (HUB) hat die Patenschaft für die 711 Felder übernommen.

Die Felder 710 und 720 sind noch offen, um Beteiligung wird gebeten.

TOP 8 und 9: Berichte

- RVK, Gabriele Schweikl: Es gibt keine Neuigkeiten. Das nächste Anwendertreffen ist für den 09.11.2022 geplant.
- Redaktionsteam Formangaben, Michael Franke-Maier: Neue Formangaben sind inzwischen mit der Codierung "saf" in die GND eingebracht worden und könnten bereits

genutzt werden, wenn die Pickliste in MAB 064a es zulassen würde (noch nicht der Fall). Das liegt daran, dass die Kommunikation zur Veröffentlichung der neuen Begriffe noch in Abstimmung mit der FG E und der DNB ist. Die Verbundzentrale sollte nach offizieller Veröffentlichung die Begriffe möglichst zeitnah in die Positivlisten für MAB 064a übernehmen.

- Kommission für Erschließung und Metadaten, Michael Franke-Maier: Das Protokoll der letzten Sitzung ist in der Abstimmungsphase und noch nicht veröffentlicht. Themen waren z.B. die Löschung von Provenienzkennungen in $\$P$ bei intellektueller Bearbeitung, die Nutzung von Culturegraph für die Formalerschließung (z.B. Sprachencode), Berichte aus anderen internen und externen Gremien sowie die Aktionenliste der Verbundzentrale.
- FG Erschließung, Barbara Wolf-Dahm: Die Erstellung des neuen Erschließungshandbuchs, basierend auf den bisherigen D-A-CH-Regeln ist bis Ende 2022 geplant. Es soll in Wikibase veröffentlicht werden mit einer Verlinkung zu allen Arbeitsunterlagen. Das alte RDA-Toolkit soll noch so lange erhalten bleiben, bis das Erschließungshandbuch veröffentlicht werden kann. Inzwischen wurde eine UAG zur Erstellung der Schulungsmaterialien gegründet. Das Expertenteam "Qualität in der GND" arbeitet weiter an der Dublettenzusammenführung innerhalb der GND.
- ET RAVI, Martin Völkl : Zukünftig soll die Inhaltserschließung heterogener werden. Es wurden die drei Dimensionen "Präzision", "Themenbildung" und "Vollständigkeit der Wiedergabe des Themas" diskutiert. Weiterhin wurde die Tiefe der Erschließung diskutiert und ob es eine Wahl zwischen "Einzelschlagwörtern", "Thematischen Haufen" oder "Schlagwortfolgen" geben sollte. Das neue modulare Regelwerk soll ebenfalls künftig in Wikibase eingestellt werden.

TOP 10: Verschiedenes

10.1. Weitere Fragen zur Anreicherungsroutine von eBook-Paketen (und einzelner eBooks) im B3Kat

- Wäre eine kürzere Frequenz künftiger Anreicherungsaktionen von eBook-Paketen denk- und umsetzbar?
(→ Statt einmal im Jahr vielleicht einmal pro Halbjahr oder einmal pro Quartal?)
Eine vierteljährliche Anreicherung erscheint sinnvoll und wäre auch umsetzbar.
- Könnten eBook-Pakete früherer Jahrgänge bei künftigen Anreicherungsaktionen nochmals mit einbezogen werden?
(→ Eine später erfolgte, intellektuelle Inhaltserschließung von parallelen Printexemplaren könnte somit auch auf die betreffenden eBooks übertragen werden.)
Es wäre möglich, regelmäßig alle eBook-Pakete anzureichern, also auch die alten Pakete wiederholt einzubeziehen.
- Unabhängig von der Anreicherung von eBook-Paketen: Wäre eine einfachere manuelle, weniger zeitintensive Übernahme der inhaltserschließenden Metadaten von einer (Print-)Ausgabe auf eine parallele (e-Book-)Ausgabe in Aleph

denkbar?

(→ Momentan müssen Formangabe, RVK-Notationen und SWF umständlich aus dem Titelsatz des Printexemplars in denjenigen des eBooks kopiert werden.)

Peter Thiessen (BSB) prüft, was Aleph in dieser Beziehung hergibt.

- Könnte beim Einspielen von eBooks in den B3Kat automatisiert überprüft werden, ob es eine parallele Printausgabe mit inhaltserschließenden Metadaten gibt, die dann sofort übernommen würden?
Die Thematik muss an die AG FE weitergegeben werden.

10.2. Öffentlichkeitsarbeit für die GND-Kooperative: Anfrage der Arbeitsstelle für Standardisierung

- Die AfS intendiert ein Forum, in dem die GND-Verbundpartner Projekte rund um die GND vorstellen können. Aus den Verbänden können ggf. Ansprechpartner:innen für Projekte benannt werden; diese sollten zunächst an die Vertreter:innen im GND-Ausschuss gemeldet werden.