

# AG K10plus Katalogisierung

## Protokoll der Webkonferenz am 05. März 2021

### Teilnehmer:

Ute Bauer	BLB Karlsruhe
Petra Bengtson	TIB Hannover
Christina Billand	IAI Berlin
Christoph Boveland	HAB Wolfenbüttel
Angelika Englert	UB Heidelberg
Christoph Hermann	UB Freiburg
Silke Horny	BSZ Konstanz
Beate Hultschig	SB Berlin
Ursula Jacob	SULB Saarbrücken
Birgit Libelt	ZBW Kiel/Hamburg
Julia Neumann	VZG Göttingen
Katharina Schmidt	SUB Hamburg

### Gäste:

Andrea Diedrich	VZG Göttingen
Annabel Feuerstein	BSZ Konstanz
Lena Hassel	BSZ Konstanz

### Entschuldigt:

André Hohmann	SLUB Dresden
Claudia Liebl	KIM Konstanz

### Protokoll:

Julia Neumann	VZG Göttingen
---------------	---------------

### Tagesordnung:

TOP 1	Umlenkungen	2
1.1	Maschinelle Dublettenbereinigung (Match & Merge)	2
1.2	Bearbeitung manuell eingetragener Umlenkungen („1698 Umlenkung nach: !PPN!“)	3
TOP 2	Sonstiges	5
2.1	Belegung des Feldes 4233	5

## TOP 1 Umlenkungen

Frau Horny stellt die Verfahren zur Dublettenbearbeitung im K10plus vor.

### 1.1 Maschinelle Dublettenbereinigung (Match & Merge)

Die maschinelle Dublettenbereinigung erfolgt in zwei Schritten:

- **Match: Prüfung, ob eine Dublette vorliegt**

Dazu werden bestimmte Felder verglichen, z. B. Nummern, Titel, Beziehungen zu Personen, Erscheinungsdatum, Ausgabevermerk und Umfangsangabe. Die einzelnen Komponenten lassen sich dabei gewichten. Datensätze, die eine sehr hohe Ähnlichkeit aufweisen, werden als identisch betrachtet.<sup>1</sup>

- **Merge: Zusammenführung der Dubletten**

Anhand des Status in Feld 0500 sowie ggf. dem Vorhandensein von Originalschrift und dem zugrundeliegenden Regelwerk (RDA/Nicht-RDA) wird der höherwertige Datensatz zum „Qualitätsgewinner“ bestimmt.

Da nicht alle Lokalsysteme Umlenkungen automatisch verarbeiten können, wird neben dem „Qualitätsgewinner“ auch der „Bestandsgewinner“ festgelegt, d. h. der Datensatz, der technisch erhalten bleibt (PPNs in Feld 0100 und ggf. 0110).

Handelt es sich beim „Qualitätsgewinner“ nicht um denselben Datensatz wie den „Bestandsgewinner“, werden die Informationen des „Qualitätsgewinners“ in den Datensatz des „Bestandsgewinners“ übertragen. Sie bleiben bei der Zusammenführung somit vollständig erhalten. Zusätzlich werden bestimmte Felder aus dem Verlierersatz übernommen, sofern sie im „Qualitätsgewinner“ nicht vorhanden sind. Das betrifft im Wesentlichen Nummernfelder, Kataloganreicherungsdaten und Sacherschließung.

Problematisch bei der maschinellen Dublettenbereinigung sind Inhalte, die in verschiedenen Feldern erfasst sein können (z. B. bei Beziehungen zu Personen in Feld 3000 oder 3010). Möglichen Informationsverlusten durch Inhalte, die nicht übernommen werden, stehen Redundanzen durch unnötig übernommene Inhalte gegenüber. Für bestimmte Sacherschließungsfelder und Kataloganreicherungs-URLs sind Bereinigungsprogramme geplant, die dublette Inhalte löschen.

Seit dem Start von K10plus wurden maschinelle Dublettenbereinigungen für E-Books sowie für gedruckte Monografien ohne Bestand durchgeführt (166.500 Titelsätze aus den Jahren 1970-2019). Bereinigungen sind darüber hinaus in folgenden Teilbereichen geplant:

- Datensätze mit identischer EKI (Identnummer der erstkatalogisierenden Institution, Feld 2240)
- Datensätze mit identischer OCLC-Nummer (Felder 2097/2098)
- Teile monografischer Reihen, die in Feld 418X mit derselben Gesamtaufnahme verknüpft sind

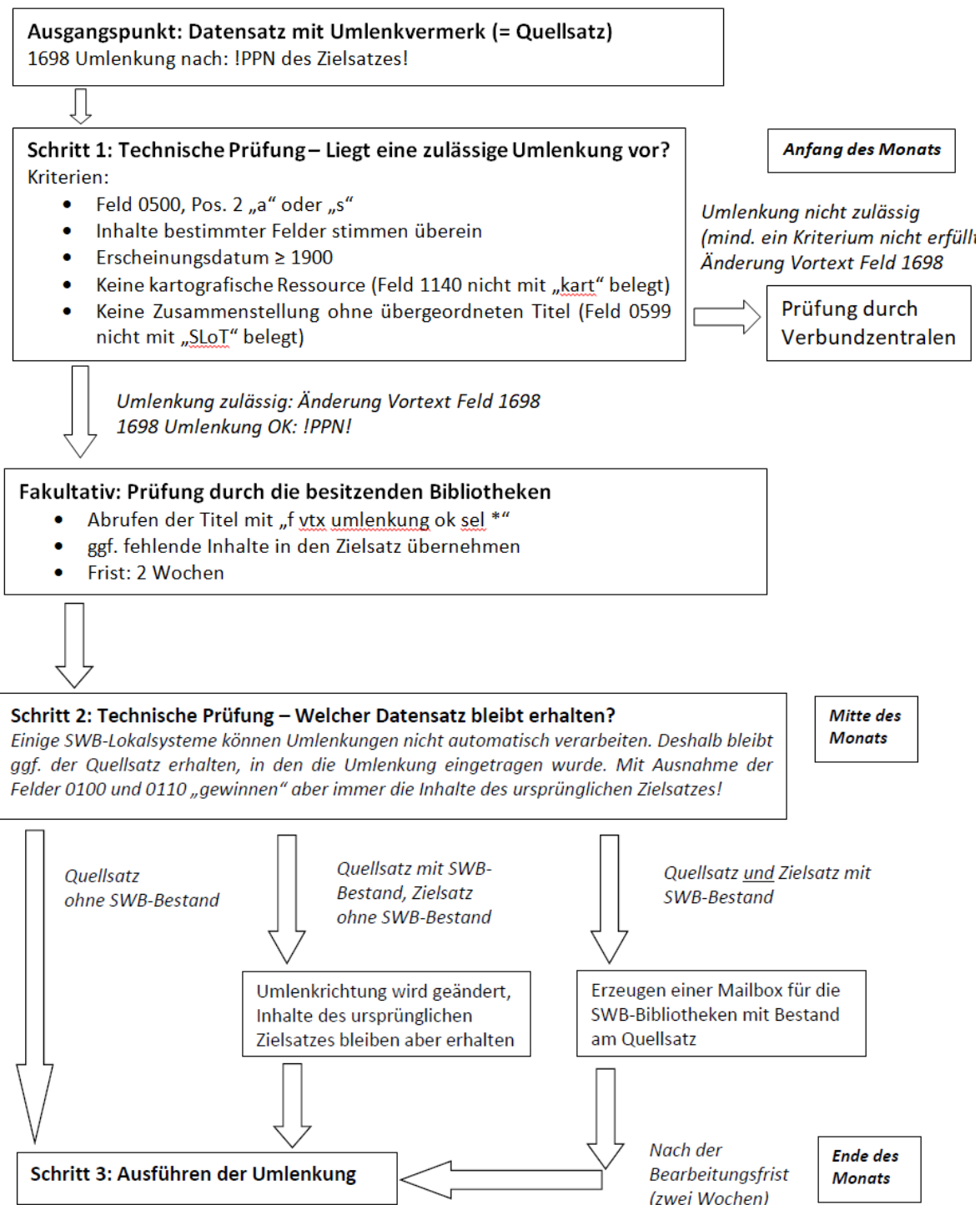
---

<sup>1</sup> Probleme bereiten dabei z. B. unterschiedliche Schreibweisen oder Namensvarianten. Sie können dazu führen, dass Datensätze maschinell nicht als Dubletten erkannt werden.

## 1.2 Bearbeitung manuell eingetragener Umlenkungen („1698 Umlenkung nach: !PPN!“)

Ausgangspunkt sind die Umlenkungen in Feld 1698, die von Katalogisierenden eingetragen wurden. Für die Bearbeitung ist seitens der Verbundzentralen folgendes Verfahren geplant:

### Ablaufplan Umlenkungen für \*a- und \*s-Sätze (Entwurf, Stand: 17.02.2021)



Für die Datensätze, in die im K10plus ein Umlenkvermerk in Feld 1698 eingetragen wurde, wurde Schritt 1 im Februar testweise durchgeführt. 60% der Datensätze wurden vom Programm als zulässige Umlenkungen erkannt und in Feld 1698 mit „Umlenkung OK“ gekennzeichnet. Die übrigen Datensätze waren Prüffälle für die Verbundzentralen oder wurden z. B. wegen des Erscheinungsdatums in der Bearbeitung zurückgestellt und erhielten einen anderen Vortext.

Bei dem weitaus größten Teil der Datensätze, die mit „Umlenkung OK“ gekennzeichnet wurden, handelt es sich um Altdatensätze, die im GBV bzw. SWB angelegt worden waren.

Die Mitglieder der AG K10plus Katalogisierung hatten als Vorbereitung für die Webkonferenz die „Umlenkung OK“-Fälle geprüft, an denen ihre Einrichtung Bestand nachgewiesen hat. Der Vortext „Umlenkung OK“ wurde den Stichproben nach korrekt vergeben, d. h. nur in Fällen, die auch bei intellektueller Prüfung als Dubletten angesehen wurden. Ausnahme: Bei einem Datensatz wurde „Umlenkung OK“ eingetragen, obwohl der Zielsatz ebenfalls ein Feld 1698 enthält. Hier ist eine Anpassung des Programms erforderlich.

Bei der Zusammenführung der Dubletten bleibt der Kern der bibliografischen Beschreibung im Zielsatz unverändert. Es werden – ähnlich wie bei der maschinellen Dublettenbereinigung (Match & Merge) – nur ausgewählte Felder aus dem Quellsatz in den Zielsatz übernommen. Das betrifft im Wesentlichen Nummernfelder, Kataloganreicherungsdaten und Sacherschließung. Eine Liste dieser Felder soll als Anhang zum Handbuch [Nutzung von K10plus-Datensätzen](#) veröffentlicht werden.

Die Prüfung ergab, dass in einigen Fällen Informationen aus dem Quellsatz verlorengehen würden, sofern die besitzenden Bibliotheken die Umlenkungen nicht prüfen und fehlende Inhalte manuell übernehmen. Das betrifft z. B.

- differenziertere Ländercodes in Feld 1700 wie „XA-DE-HH“ statt „XA-DE“ (wichtig für Pflichtexemplarbibliotheken)
- Beziehungen zu Personen und Körperschaften in den Feldern 30X0/31X0
- zusätzliche Erscheinungsorte/Verlage in Feld 4030 (wichtig für Pflichtexemplarbibliotheken)
- Verknüpfungen zu fortlaufenden Ressourcen in Feld 418X

Aus den Katalogisierungsunterlagen muss deutlich hervorgehen, dass vor dem Eintragen einer Umlenkung insbesondere die Felder zu überprüfen sind, die dem Korrekturrecht unterliegen (vgl. Handbuch [Korrekturverfahren](#), Kapitel 3.1). Stimmen z. B. die bibliografische Erscheinungsweise in Feld 0500, Pos. 2 oder das Erscheinungsdatum nicht überein, muss vorab eine Mailbox an die besitzenden Bibliotheken geschrieben werden.

Außerdem liegt es in der Verantwortung der Person, die die Umlenkung in Feld 1698 einträgt, alle Informationen aus dem Quellsatz in den Zielsatz zu übernehmen, die dort nicht vorhanden sind und nicht per Programm übernommen werden.

Wenn dies gewährleistet ist, gestaltet sich die Überprüfung der umzulenkenden Datensätze, an denen die jeweilige Einrichtung Bestand hat, deutlich einfacher. Generell halten die Mitglieder der AG K10plus Katalogisierung eine Prüfphase als Angebot für die Bibliotheken für sinnvoll. Gleichzeitig fehlen die Personalkapazitäten, um den großen Anteil der eingetragenen Umlenkungen in Altdaten zu prüfen und ggf. entsprechende Mailboxen zu bearbeiten.

In diesem Zusammenhang wird darauf hingewiesen, dass gemäß den Bestimmungen im Handbuch [Nutzung von K10plus-Datensätzen](#) keine Umlenkvermerke in Altdaten eingetragen werden dürfen (Ausnahmen: E-Books, Gesamtaufnahmen mehrteiliger Monografien und fortlaufender Ressour-

cen). Diese Regelung soll weiterhin gelten. Die AG K10plus Katalogisierung bittet die Verbundzentralen darum, dies erneut an die K10plus-Bibliotheken zu kommunizieren.

Das Verfahren zur Bearbeitung manuell eingetragener Umlenkungen wird angepasst, sodass Umlenkungen in Altdaten bei der Analyse in Schritt 1 per Programm künftig direkt mit einem anderen Vortext in Feld 1698 versehen und bei der weiteren Analyse nicht berücksichtigt werden.

Anschließend werden Detailfragen zu einzelnen „Umlenkung OK“-Fällen besprochen.

## TOP 2 Sonstiges

### 2.1 Belegung des Feldes 4233

In der TIB Hannover werden in Absprache mit dem Urheber Nutzungskopien für Inhalte von USB-Sticks oder zugesandten Dateien erstellt. Die Inhalte selbst werden in Rosetta langzeitarchiviert, die Katalogisierung erfolgt nur für die Nutzungskopie. Feld 4233 kann in diesen Fällen mit dem Code „aa“ für „Archivierung/Langzeitarchivierung gewährleistet“ im Unterfeld \$a belegt werden, da sich der Sachverhalt „Archivierung“ auf die Inhalte und nicht auf den Datenträger bezieht. Im Unterfeld \$z kann eine Anmerkung zum Sachverhalt eingetragen werden, z. B. „digitale Langzeitarchivierung“.

Julia Neumann

19.03.2021