

Altdatenkonzept:

1. Allgemeine Darstellung der Problematik:

Die Migration konnte nicht alle Erfordernisse der Übergangsregeln bedienen, so dass viele Migrationsdaten (Altdaten) nach GND-Start aufgearbeitet werden müssen.

Hierfür müssen technische Rahmenbedingungen beachtet werden, insbesondere hinsichtlich der Kapazitäten der Schnittstellen, die in bestimmten Zeitabschnitten nur einen begrenzten Umfang an Änderungen zulassen.

Redaktionelle Änderungen, die Änderungen von Titelverknüpfungen nach sich ziehen (das sind Umlenkungen oder Löschungen) oder zu einer Neuindexierung der Indices für die verknüpften Titel führen (das ist der Fall, wenn sich die Vorzugsbenennung und die abweichenden Namensformen geändert haben), können bis ein anderer Stand abgestimmt wird, nicht in größerem Umfang getätigt werden, da die korrekte Verarbeitung nicht in allen abnehmenden Systemen gewährleistet werden kann.

2. Typen der Korrekturen und Bestandsaufnahme

Es gibt für die unterschiedlichen Altdaten grundsätzlich drei Typen von Korrekturen und eine Normdatengruppe, die aufgrund ihrer umfangreichen Titelverknüpfungen besonders behandelt werden muss:

- Intellektuell vorzunehmende Änderungen; in der Regel aufgrund von nicht möglicher Umsetzung der Übergangsregeln bei der Migration (z.B.: Die Entscheidung, ob der Ort bei Körperschaften, die früher mit Ordnungshilfe erfasst wurden, Namensbestandteil ist oder nicht, konnte nicht bei der Migration erfolgen, sondern muss intellektuell vorgenommen werden, vgl. AWB-K10)
- Maschinelle Änderungen (z.B. Korrektur von gemäß Erfassungsleitfaden falsch strukturierten Zeitangaben im Feld 548)
- Aufarbeitung durch IT-Unterstützung, aber nicht ganz maschinell (z.B. per Skript den Code „ortm“ in „orta“ ändern)
- Top-500-Datensätze, die aufgrund der großen Menge von verknüpften Titeldaten besonders beachtet werden müssen. Vom Fehlertyp sind sie keine eigene Gruppe. (vgl. Wiki <https://wiki.d-nb.de/x/fIBsAw>)

Für die drei großen Normdateien GKD, PND und SWD liegen umfangreiche Bestandsanalysen der zu korrigierenden Sachverhalte vor mit entsprechenden Korrekturvorschlägen je nach Typ der Korrektur. Aus diesen Dokumenten sollen konkrete Zeitpläne erarbeitet werden.

[Bestandsaufnahme SWD](#)

[Bestandsaufnahme PND](#)

[Bestandsaufnahme GKD](#)

Bestandsaufnahme DMA: (noch in Bearbeitung)

3. Bereits getroffene Absprachen und Zeitpläne

In den Anwendungsbestimmungen und im Erfassungsleitfaden wurde die Altdaten-Problematik dargestellt und auf das Altdatenkonzept verwiesen. Aufgrund technischer Begrenzungen können die einzelnen Fallgruppen der überregionalen GND nicht sofort bearbeitet werden. Deshalb werden sie hier nicht aufgeführt, sondern es wird nur der vereinbarte grobe Rahmen dargestellt.

Seit Anfang Juli 2012 läuft das maschinelle Match-und-Merge, das sich bis ca. **Januar 2014** hinziehen wird, weil nur schnittstellenverträgliche Mengen verarbeitet werden dürfen. Das bedeutet, dass systematische Datenbereinigungen weiterhin nicht möglich sind, da die Systeme durch das Match-und-Merge-Verfahren ihre Kapazität, Änderungen zu verarbeiten, ausschöpfen. Seit März 2013 werden geografische Datensätze zusammengeführt. Es wurde beschlossen, das maschinelle Match-und-Merge-Verfahren zeitweise auszusetzen, um besonders wichtige geografische Datensätze korrigieren und zusammenführen zu können. Folgende fünf Top-500-Datensätze wurden bisher korrigiert und zusammengeführt: Italien, USA, Frankreich, Bayern, Deutschland, Deutschland <DDR>, Deutschland <Bundesrepublik>, **Großbritannien**.

3.1. Korrekturen nach Produktionsbeginn der GND (seit November 2012)

Erlaubt sind:

- maschinelles Match-und-Merge

Erlaubt sind, wenn nicht mehr als **ca. 5.000 Titel¹** mit dem GND-Datensatz verknüpft sind, folgende Aktionen:

- intellektuelle Umlenkungen im Rahmen der normalen Redaktionsarbeit
- Löschungen im Rahmen der normalen Redaktionsarbeit
- Änderungen der Vorzugsbenennung und abweichenden Namensformen im Rahmen der normalen Redaktionsarbeit
- intellektuelle Aufarbeitung der Top-500-Datensätze nur von DNB und Verbundredaktionen nach vorheriger Absprache
 - DNB erstellt Vorgaben
 - Verbundredaktionen prüfen
 - Korrektur durch DNB und Kennzeichnung der Datensätze durch DNB in Feld 667 mit "Top500 - jede Art von Änderung nur nach Absprache"

Für das maschinelle Match-und-Merge und für die Aufarbeitung von Top-500-Datensätzen sind gemeinsame Zeitpläne erstellt (Zeitplan für Match-und-Merge: <https://wiki.d-nb.de/x/-QGOAw>; Top-500-Bearbeitung: <https://wiki.dnb.de/x/VqlsB>).

Mailbox-Verkehr:

Korrekturanträge können über Mailboxfelder gestellt werden.

- Die Redaktionen können Korrekturen durchführen bis auf die 1XX- u. 4XX-Felder der Top-500-Datensätze.
 - Die Bibliotheken können im Rahmen ihrer Korrekturberechtigungen Korrekturen durchführen.
- Ausnahme: Korrekturen an Top-500-Aufnahmen sind nicht erlaubt.

Die vereinbarte Sperre für Umlenkungen und Änderungen der Vorzugsbenennungen und abweichenden Namensformen von Top-500-Datensätzen kann nicht technisch erzeugt werden; es handelt sich ausschließlich um eine redaktionelle Absprache.

¹ pro Woche sind Änderungen von bis zu 50.000 Titeln zulässig. Wenn Datensätze mit umfangreichen Titelverknüpfungen geändert werden müssen, sollte die Arbeit so geplant werden, dass es nur wenige pro Woche sind; die Änderungen sollten vorzugsweise am Freitag geschehen, um das Wochenende für die Neuindexierung nutzen zu können.

4. Kennzeichnung der Korrekturen in den Datensätzen

Da die Korrekturen der Datensätze oft aufwändig sind, sollen die Datensätze gekennzeichnet werden.

Die AG GND hat sich auf die Kennzeichnung der bearbeiteten Datensätze im wiederholbaren Feld 667 nach GND-Start gemäß des unter Punkt 3 vereinbarten Zeitplans wie folgt geeinigt:

| Nr. | Typen der Aufarbeitung | Kennzeichnung in 667 |
|-----|--|--|
| 1 | Aufarbeitung der Vorzugsbenennung, nur im Gewinnerdatensatz (Felder 1xx) <i>Für Datensätze mit bis zu 5.000² Titelverknüpfungen pro Verbund zulässig</i> | BNPe nur im Gewinnerdatensatz nur durch Redaktionen nur Level-1-Datensätze |
| 2 | Aufarbeitung des gesamten Datensatzes, nur im Gewinnerdatensatz (alle Felder) <i>Für Datensätze mit bis zu 5.000 Titelverknüpfungen pro Verbund zulässig;</i> | VPe nur im Gewinnerdatensatz nur durch Redaktionen nur Level-1-Datensätze |
| 3 | Intellektuell vorgenommene Zusammenführung einschließlich Aufarbeitung gemäß den Übergangsregeln <i>Für Datensätze mit bis zu 5.000 Titelverknüpfungen pro Verbund zulässig</i> | MMi (gilt nur für die Satzarten Tb, Tf, Tg und Tu) nur im Gewinnerdatensatz nur durch Redaktionen nur Level-1-Datensätze |
| 4 | Intellektuell vorgenommene Zusammenführung ohne Aufarbeitung gemäß den Übergangsregeln bzw. nach einem vereinfachten Verfahren <i>Für Datensätze mit bis zu 5.000 Titelverknüpfungen pro Verbund zulässig</i> | MMina (gilt nur für die Satzarten Tb, Tf) nur im Gewinnerdatensatz nur durch Redaktionen |
| 5 | Kennzeichnung der maschinell zusammengeführten Datensätze | MMm nur im Gewinnerdatensatz |
| 6 | Kennzeichnung des Verliererdatensatzes bei maschineller Zusammenführung | MMv im Verliererdatensatz (Es wird noch geprüft, ob die Kennzeichnung im Verliererdatensatz im Feld 667 erfolgen kann, oder ob aufgrund der Verkürzung des Satzes ein anderes Feld verwendet werden muss.) |

5. Beispieldatensätze in der GND mit Kennzeichnung im Feld 667

Da der Migrationsstand nicht immer den Übergangsregeln entspricht, wurde vereinbart, dass Beispieldatensätze in der GND aufgearbeitet und im Feld 667 mit „GNDBeispiel“ gekennzeichnet werden.

² pro Woche sind Änderungen von bis zu 50.000 Titeln zulässig. Wenn Datensätze mit umfangreichen Titelverknüpfungen geändert werden müssen, sollte die Arbeit so geplant werden, dass es nur wenige pro Woche sind; die Änderungen sollten vorzugsweise am Freitag geschehen, um das Wochenende für die Neuindexierung nutzen zu können.