

GND in SISIS-SunRise – Nacharbeiten

*Hintergründe und Überblick zu Nacharbeiten in SISIS-SunRise aufgrund der GND-Einführung
(insbes. Datenmanipulation durch ID-Tausch bei einzelnen Normdatensätzen)*

1) GND-Konzept in SISIS-SunRise

Die **Veränderungen in SISIS-SunRise**, speziell der SIKOM-Versorgungsschnittstelle sind dokumentiert und hatten für OCLC zur Realisierung in den Versionen SISIS SunRise V4.1p11 und V3.7p12g gedient. Das Papier ist hier zu finden: http://www.bib-bvb.de/sisis/tips/papers/GND-Szenario_11-2010.pdf

Zusammenfassung der aktuellen Implementierung:

- **Normdatei-Struktur:** In SISIS-SunRise bleiben die drei "klassischen" Normdateien getrennt erhalten
- **Identifikationsnummern:** In den drei lokalen Normdateien müssen vor GND-Start nicht die neuen IDs eingebracht werden, da diese i.d.R. bijektiv zwischen den alten PND-, GKD- und SWD-IDs und den neuen GND-IDs abgebildet werden können (s. aber Kapitel "4) ID-Tausch ..."). Die SIKOM prüft das Vorhandensein lokaler Normsätze immer mit und ohne Präfix.
- **Normdaten-Updates:** Normdaten werden im Verbund an dem "GND"-Target über Z39.50 abgeholt (MARC21-Format, UTF-8-Zeichensatz) und SIKOM verarbeitet anschließend jeden GND-Normsatz indem dieser via GND-ID einer der drei lokalen Normdateien in SISIS-SunRise zugeordnet wird. Bei Titel-Updates erfolgt dies aus dem Kontext der Verknüpfungsfelder im Titel (100ff, 200ff, 902ff), bei Normdaten-Updates sind ggf. auch zwei lokale Normdateien zu aktualisieren (z.B. bei Update eines Personen-Satzes in der GND, der lokal im Kontext der PND und der SWD verwendet wird).
Hinweis: In den Titel-MAB2-Feldern 902ff kann aufgrund einer Längenbeschränkung des Feldes nur die verkürzte GND-ID (ohne Präfix) aus den Verbunddatenbanken geliefert werden. Eine Z39.50-Recherche in der GND der Verbunddatenbanken ist auch mit dieser verkürzten ID möglich.
- **Normdaten-Inhalte:** In den Lokalsystemen existieren im Rahmen dieses Szenarios mit GND-Start die ursprünglichen Normdaten der PND, GKD und SWD neben den neuen GND-Sätzen.

Bis spätestens zum **Zeitpunkt der Einführung der GND** muss eine der beiden genannten **SISIS-SunRise-Versionen mit dem entsprechenden SIKOM-Patch** im Einsatz sein. Mit Aktivierung des entsprechenden Parameters und der neuen Normdaten-Konvertertabellen kann dann die Datenversorgung über SIKOM weiterlaufen. Damit müssen zum Zeitpunkt der Einführung der GND in den lokalen SISIS-SunRise-Systemen **keine weiteren Datenbereinigungen zwingend und sofort** durchgeführt werden.

Es gibt jedoch **Nacharbeiten**, die mehr oder weniger zeitnah zur Einführung der GND durchgeführt werden sollten. Diese Arbeiten werden im folgenden motiviert und beschrieben.

2) GND-Migration in den Verbunddatenbanken des BVB und des hbz

Allgemeine Informationen

Im Rahmen der GND-Migration wird in den Aleph-Verbänden eine **neue Normlibrary** eingerichtet (**BVB18 bzw. HBZ18**), in welche die GND-Daten der DNB importiert werden. Die Daten wurden durch die DNB strukturell und inhaltlich aus den PND-, GKD- und SWD-Daten konvertiert und im MARC21-Format und UTF8-Zeichensatz ausgeliefert.

Die **DNB führt keine explizite Zusammenführung der Sätze/Entitäten** im Rahmen der GND-Migration durch (z.B. keine Zusammenführung einer Körperschaft der Formalerschließung aus der GKD mit einer Körperschaft der Sacherschließung aus der SWD). Jeder "alte" Satz hat also genau eine Entsprechung in der "neuen" Normdatei, wobei die neue ID der alten ID mit vorangestelltem fixen Präfix entspricht. Eine Ausnahme bilden hierbei die Personenschlagwörter.

Sonderfall 1) Personenschlagwörter aus der SWD

Bei den Personenschlagwörtern handelt es sich bereits jetzt (vor der GND-Umstellung) um **eine Entität bei der DNB (PND-Satz)**, die in MAB2 aber noch "übergangsweise" als zwei getrennte Sätze (als PND- und SWD-Satz) mit jeweils eigener ID von der DNB geliefert wurden. Der Import nach Aleph und die Verteilung an die Lokalsysteme erfolgte ebenfalls über zwei Libraries (BVB10/HBZ10: Personennamen und BVB12/HBZ12: Schlagwörter).

Mit Umstieg auf die GND ändert die DNB ihre Lieferpraxis: Jeder Datensatz wird nur noch einmal im MARC21-GND-Format ausgeliefert. Daher wird als **Seiteneffekt bei der GND-Migration in Aleph de facto eine Zusammenführung dieser Personenschlagwörter (SWD-Daten) auf die GND** durchgeführt, da in der Aleph-GND-Umgebung (BVB18, HBZ18) die Entität nur noch als ein "gemeinsamer" Personensatz weiter existiert.

Sonderfall 2) hbz: Zusammenführung regionaler hbz-Normdaten auf überregionale Normdaten

Im Zuge des GND-Projektes hat das hbz eine **Zusammenführung der regionalen Normdatensätze (Personen, Körperschaften und Schlagwörter) auf überregionale Normdatensätze** vorbereitet. Die maschinellen Abgleiche zur Ermittlung der regionalen Dubletten innerhalb der drei Normdateien wurden im August 2011 durch einen externen Dienstleister durchgeführt und die als dublett identifizierten Normsätze in Form von Konkordanzlisten dokumentiert.

Daneben wurden die Sätze, zu denen kein überregionales Pendant ermittelt werden konnte, entsprechend den Vorgaben der DNB aufbereitet. Das hbz hat diese Daten im September 2011 als Neuzugänge an die DNB geliefert. Lediglich ein Teil der regionalen Schlagwortsätze ist nicht an die DNB geliefert worden und wird in der hbz-Verbunddatenbank entweder ersatzlos gelöscht oder ggf. in bestimmten Fällen ein alternatives Schlagwort intellektuell vergeben.

Das hbz hat alle Arbeiten zur Dublettenbereinigung der Normdaten inklusive Lieferung der Neuzugänge an die DNB noch vor der produktiven Einführung der GND abgeschlossen, da sich bei den Planungen zur GND-Migration abzeichnete, dass eine **Kopplung der beiden Projekte** nicht nur die Aufwände für das hbz reduziert, sondern auch die Auswirkungen auf die Lokalsysteme minimiert.

Die **physikalische Zusammenführung der als dublett identifizierten Normdaten** (d. h. die Umverknüpfungen in den Titeldaten zu ausschließlich GND-Sätzen) wird daher **erst im Rahmen der Migration zur GND stattfinden**, so dass die notwendigen Titelkorrekturen (Austausch der Identifikationsnummern und Ansetzungsformen in den normdatenrelevanten Feldern) mit den zugehörigen Auswirkungen auf die Replikation in Richtung der Lokalsysteme nicht doppelt durchgeführt werden müssen. Die dubletten regionalen Normdatensätze werden in der hbz-Verbunddatenbank daher erst im Rahmen der GND-Migration physikalisch zusammengeführt (eine Migration von HBZ10-, HBZ11- und HBZ12-Daten entfällt)

Zielzustand in den Verbunddatenbanken des BVB und des hbz

Folgender **Zielzustand** existiert in den Verbunddatenbanken des BVB und des hbz nach der Migration:

- **Neue Norm-Library BVB18/HBZ18:** Die neue Normdatei BVB18/HBZ18 enthält alle GND-Daten der DNB (MARC 21), insbes. inkl. neuer ID in Feld 001 und ggf. anderer Ansetzungs- und Verweisungsformen (wg. Änderungen Format/Regelwerk).
- **Keine regionalen Normdatensätze:** Nach GND-Umstieg existieren keine regionalen Normdatensätze mehr.
- **Titeldaten:** Die Titeldaten in BVB01/HBZ01 erhalten die Anbindung/Verknüpfung an die (neuen) Normsätze (ID) und enthalten die ggf. anderen Ansetzungsformen der Normdaten in einer differenzierten Aleph-internen Unterfeldstruktur.

Erreicht wird dieses Ziel über den Weg, dass alle "alten IDs" der Normsätze vor der GND-Migration ("Verlierer-IDs") in die GND-Sätze ("Gewinner-IDs") integriert, in Aleph gespeichert und indexiert werden (z.T. unter Abschneiden des GND-Präfixes "DE-588"):

- **Sonderfall "Personenschlagwörter":** Die Personenschlagwörter (Personen-Entitäten) werden von der DNB auch mit ihren alten "Zweit-IDs" aus der ehemaligen SWD geliefert.
- **Sonderfall regionale "hbz-Verlierer-Normdatensätze":** Das hbz reichert die GND-Sätze um die alten regionalen IDs an (d.h. um die Verlierer-IDs der Dublettenzusammenführung).
- **Migration Titeldaten:** Die in den noch nicht migrierten Titeln der Verbunddatenbank verknüpften Normdaten können daher über alle "alten" IDs (Normdaten-IDs der DNB ohne GND-Präfix, IDs von DNB-Personenschlagwörtern und IDs von regionalen hbz-Sätzen) die neue Normdatenverknüpfung zur GND aufbauen.

3) Datenstand in SISIS-SunRise nach der GND-Migration: Übergangsszenario

Im Rahmen der GND-Migration der Verbunddatenbanken des BVB und des hzb werden **keine Aufträge für die Versorgungsschnittstelle** generiert (downtime).

Nach der Migration der Verbunddatenbanken existiert daher folgender Datenstand in den SISIS-SunRise-Lokalsystemen:

- **Titeldaten: Datenstand vor der GND-Migration des Verbundsystems**
- **Normdaten: Datenstand vor der GND-Migration des Verbundsystems**

Mit **Produktionsstart** der Verbunddatenbanken wird die Versorgungsschnittstelle wieder aktiviert. Alle Updates ab diesem Zeitpunkt beinhalten migrierte Titeldaten (d.h. neue GND-Ansetzungen und Verknüpfungs-IDs sowie die aktuellen GND-Sätze). Das Lokalsystem kann mit dieser neuen Situation umgehen während gleichzeitig die nicht aktualisierten Titel und Normsätze im Lokalsystem vor der GND-Migration i.d.R. unverändert bestehen bleiben (zur Heterogenität der Daten bzw. zur Dauer dieses Übergangsszenarios s. Abschnitt "5) GND im Produktionsbetrieb - weitere Aspekte").

Im Rahmen des oben beschriebenen Übergangsszenarios besitzen **die Sonderfälle "Personenschlagwörter" und "regionale hzb-Normdatensätze"** die Eigenschaft, dass die **Bijektivität der IDs (alt - neu) nicht gegeben** ist.

Hintergrund: Während bei vielen GND-Sätzen die Abbildung der alten ID zu der neuen GND-ID durch Abschneiden/Hinzufügen des Präfixes erreichbar ist und dadurch auf einen ggf. schon im Lokalsystem vorhandenen Normsatz gematcht werden kann, ist die Abbildung der IDs nicht für die o.g. Sonderfälle möglich.

Als **Konsequenz** der fehlenden Abbildungsmöglichkeit der ID ergibt sich z.T. folgende Situation:

- **1. Dubletten zu Normsätzen:** Durch Titel-Updates können Normdaten als Dubletten im Lokalsystem erzeugt werden, die ggf. schon unter anderer ID vorhanden sind
- **2. Fehlende Normdaten-Updates:** Normdaten-Updates aus der Verbunddatenbank werden nicht über Z39.50 übernommen, wenn Updates auf neue GND-IDs erfolgen, die im Lokalsystem noch nicht vorhanden sind.

4) ID-Tausch in Normdatensätzen in SISIS-SunRise-Lokalsystemen

Um im Rahmen des Übergangsszenarios die o.g. Konsequenzen von fehlender Abbildungsmöglichkeit der IDs im Lokalsystem zu minimieren, sollten **Nacharbeiten** durchgeführt werden, die mehr oder weniger **zeitnah bzw. parallel zur GND-Migration der Verbunddatenbanken** stattfinden können.

Die Nacharbeiten bestehen im Austausch von IDs zu Normdatensätzen, die die Verbundzentralen des BVB und des hzb den Lokalsystemen zum Zeitpunkt der GND-Migration in Form von Konkordanzlisten zur Verfügung stellen:

- **Personenschlagwörter:** Diese Konkordanz beinhaltet die Abbildung der Verlierer-IDs der Personenschlagwörter der DNB (SWD) zu den GND-Sätzen der entsprechenden Personen-Entität. Die endgültige Konkordanz wird aus der aktuellen GND-Lieferung der DNB erzeugt.
- **Regionale hzb-Normdatensätze:** Diese Konkordanzen beinhalten die Abbildung von "Verlierer-IDs" regionaler hzb-Normdatensätze (getrennt nach Personen, Körperschaften, Schlagwörter) zu den überregionalen "Gewinner"-Sätzen. Diese Konkordanzen liegen bereits endgültig vor.

Für den Austausch der Normdaten-IDs hat der BVB Tools entwickelt, die auch von hzb-Bibliotheken getestet wurden. Diese werden zusammen mit einer Dokumentation zur Verfügung gestellt.

Die Umsetzung der Konkordanzen gegen die drei lokalen Normdateien führt zu folgenden **Problemfällen**, die im Rahmen weiterer Schritte nachgearbeitet werden sollten:

- **1) Identische Gewinner-ID für 2-n Verlierer-IDs:** Dieser Fall tritt auf, wenn z.B. mehr als ein regionaler hzb-Normdatensatz auf den gleichen überregionalen Normdatensatz zusammengeführt wurde (d.h. die Konkordanz für eine Gewinner-ID n Zeilen mit Verlierer-IDs besitzt). In diesen Fällen sollte eine Nachversorgung der Titel der betroffenen Normdatensätze (in denen ein Austausch der ID nicht möglich ist) über den SIKOM-Testmodus erfolgen.
Hinweis: Die Anzahl der nachzuversorgenden Datensätze kann in sehr großen Bibliotheken einige 100.000 Titel betragen. Aufgrund der Menge der Titel, der Anzahl der Lokalsysteme und dem vermutlich hohem Interesse einer zeitnahen Bereinigung sind die Aktivitäten der Lokalsysteme mit den Verbunddatenbanken zu koordinieren, um zu viele parallelen Zugriffe und mögliche Auswirkungen auf die Rechnerkapazität der Verbundsysteme zu vermeiden.

Zu diesem Aspekt finden bereits Analysen zur Evaluierung optimaler Lösungsmöglichkeiten zwischen den Pilotbibliotheken ULB Münster/USB Köln und dem hbz statt. Es sind auch bereits Tests zwischen dem hbz und den Pilotbibliotheken ULB Münster und USB Köln angelaufen, um den Durchsatz von Z39.50-Recherchen zu testen. Dadurch kann auch ermittelt werden, welche begrenzenden Faktoren zentral/lokal existieren.

- **2) Gewinner-ID schon im Lokalsystem vorhanden:** In diesem Fall befinden sich im Lokalsystem bereits die Partner einer Dublettenzusammenführung, d.h. sowohl eine Verlierer-ID als auch eine Gewinner-ID, die beide mit Titeln verknüpft sind. Dadurch sind bereits im Vorfeld der GND-Einführung im Lokalsystem Dubletten vorhanden, die dann einen Austausch der Norm-IDs bei den Verlierer-IDs verhindern
- **3) Normdaten, die durch die Konkordanzen nicht getroffen wurden:** Hier handelt es sich entweder um lokale Abweichungen zur Verbunddatenbank, um Datenfehler im Verbundsystem oder um eine unvollständige Umsetzung der Dublettenbereinigung im hbz. Das hbz wird voraussichtlich vor und nach der GND-Migration Datenbereinigungen in den Titeldaten vornehmen (z.B. Löschungen von nicht mehr vorhandenen Normdaten-IDs).

5) GND im Produktionsbetrieb – weitere Aspekte

Vollständige Datenmigration/Nachversorgung

Mit Zusammenführung der drei getrennten Normdateien auf die GND wird auch eine **strukturelle und inhaltliche Migration der Normdaten von der DNB** durchgeführt (entsprechend der dafür erstellten Übergangsregeln RAK/RSWK - RDA). In diesem Zusammenhang werden daher auch **Änderungen und Erweiterungen im Bereich der Ansetzungs- und Verweisungsformen** durchgeführt.

Nach der Einführung der GND entsprechen daher viele Normdatensätze in den Lokalsystemen zunächst inhaltlich nicht den GND-Sätzen und es ergeben sich **Heterogenitäten**. Da diese Unterschiede teilweise auch Ansetzungen betreffen, ist es sinnvoll die **lokalen Normdatensätze auf den aktuellen Stand** zu bringen, damit auch lokal die GND komplett umgesetzt werden kann.

Dafür ist es denkbar in einem **parallelen SIKOM-Prozess** alle lokalen Normdatensätze, die noch keine GND-ID besitzen, nachzuversorgen: Die betreffenden Normdaten-IDs könnten dazu lokal ermittelt und eine Eingabedatei für den parallelen SIKOM-Testmodus erstellt werden. Die Abarbeitung würde dann parallel zum normalen Betrieb erfolgen und sukzessive die Normdatensätze auf den aktuellen Stand bringen.

Alternativ sind auch **Exportlieferungen aus den Verbänden** denkbar, die ggf. lokal importiert werden könnten: Z.B. MAB2-Titeldatendienste für eine Bibliothek (ausgehend von der Selektion der Bestände), um die damit verknüpften spezifischen GND-Normdaten für eine Bibliothek zu gewinnen. Oder alternativ auch ein Gesamtexport der GND, wobei aus dieser Menge dann jedes Lokalsystem die relevanten Sätze ermitteln müsste.

Gegenstand dieses Papiers ist nicht die Bewertung, ob ein Übergangsszenario im Lokalsystem für einen längeren Zeitraum oder nur für eine deutlich kürzere und absehbare Zeit akzeptiert wird. Die Entscheidungsaspekte dazu sind sicher sehr vielschichtig und können je nach Blickwinkel (Systemadministration, Katalogisierung, Benutzung) unterschiedlich bewertet werden.

Entitäten-Zusammenführung der DNB (nach GND-Migration)

Die DNB hat bereits angekündigt, ca. 10 Wochen nach Abschluss der GND-Migration (Ende Juni 2012) **Entitäten maschinell zusammenzuführen** (z.B. alte Körperschaften der GKD und der SWD).

Nach bisherigem Stand werden diese Zusammenführungen über die Replikation der beteiligten Titel (d.h. über die Updates der Umverknüpfungen der Normdaten in den Titelfeldern) an die Lokalsysteme realisiert. Die Verbundsysteme und die DNB werden hier **Absprachen** treffen, die darauf abzielen die Anzahl der maschinellen Normdaten-Umlenkungen pro Tag/Woche zu portionieren und zu kennzeichnen. Darüber könnten die Verbände, bei denen insbesondere die Anzahl der daraus resultierenden Titel-Updates relevant ist, eine **zielgerichtete Versorgung** anstoßen bzw. Lokalsysteme mit ID-Listen versorgen (z.B. für einen Parallelversorgung über den SIKOM-Testmodus oder für einen weiteren Austausch von IDs).

Auch durch intellektuelle redaktionelle Arbeiten werden mit GND-Start **manuell Entitäten zusammengeführt**, wobei hier auf Basis der Normdatensätze keine Mengenprobleme zu erwarten sind. In diesen Fällen sind jedoch die Auswirkungen auf die Anzahl der verknüpften Titeldaten von den Verbänden vermutlich nicht steuerbar (z.B. bei Umlenkungen von ehemaligen GKD-/SWD-Körperschaften mit vielen verknüpften Titeln: "Bundesrepublik Deutschland").

Normdatenstruktur in SISIS-SunRise

Hinsichtlich der Normdatenhaltung in den SISIS-SunRise-Systemen bleibt es zunächst bei den **getrennten lokalen Normdateien**.

Hinweis **von Frau Caballero-Castro (OCLC)**: Ob die SunRise Anwender die Normdaten zukünftig im Lokalsystem zusammenführen möchten liegt in der Entscheidung der Kunden. SunRise an sich kann eine Normdatei unterstützen, d.h. die SunRise Software unterstützt dies bereits. Aktuell wurde SIKOM für den GND-Support so angepasst, dass mit drei lokalen Normdaten gearbeitet wird. Dieses müsste für Anwender, die mit einer Normdatei lokal arbeiten wollen, geprüft und angepasst werden.