

Match-&-Merge: Änderungen im Bereich der Geografika

TelKo zum Themenkomplex Match-und-Merge in der GND, 12.12.2011, 14:00 Uhr

TOP 3: Änderung der fremdsprachigen Geografika in ehemaligen GKD-Sätzen in die deutsche Form, Vorwegnahme von strukturellen Änderungen in Vorzugsbezeichnungen für Geografika

Problemskizze

- Ansetzungsänderungen im Bereich der Gebietskörperschaften (Tg-Sätze) haben Folgewirkung auf eine größere Menge anderer GND-Sätze:
 - ⇒ **Anzahl der geänderten Datensätze steigt; Anzahl auszuliefernder Sätze steigt**
 - ⇒ Ursache:
 - Datensätze werden beim M-&-M-Verfahren mitunter mehrfach geändert (Austausch von IDNs in Tb/Tf-Sätzen in 551-Feldern bei M&M von Tg-Sätzen; anschließend M&M für Tb-/Tf-Sätze, d. h. teilw. werden Sätze erneut bearbeitet)
 - zusätzlicher Änderungsbedarf durch notwendigen Austausch von Feldinhalten in die GND-Form (Datensätze werden aber vom M-&-M-Verfahren nicht erfasst, da keine Zusammenführung nötig)
 - ⇒ Übersicht zu nötigen Änderungen in Tb-/Tf- und teilw. Tp-Sätzen:
 - zahlreiche Verknüpfungen in Feld 551 zu Gebietskörperschaften in allen GND-Sätzen => werden bei Änderung der Ansetzungsform aktualisiert; Satz muss ausgeliefert werden
 - Gebietskörperschaften in 110/410 \$g; 111/411 \$c, 110 \$a/410\$a (wenn \$a = Name einer Gebietskörperschaft; kio-Codierung) werden i. d. R. in Ansetzungsform zitiert
=> d. h. Feldinhalte müssen bei Änderung der Ansetzung der Gebietskörperschaft ebenfalls korrigiert werden; Satz muss ausgeliefert werden

Fallgruppen

a) GKD: fremdsprachige Gebietskörperschaften

- GKD-Ansetzung: offizielle (i.d.R. originalsprachlicher) Name
- GND-Ansetzung: die im Deutschen gebräuchliche Form (i.d.R. ehemalige SWD-Ansetzungsform).
- BEISPIEL: GKD: Genève = SWD: Genf = GND: Genf
- nötig: 5XX: IDN-Austausch; 1XX/4XX: Austausch von Feldinhalten
- ⇒ für Fallgruppe wird maschinelle Bereinigung im Vorfeld empfohlen

b) GKD: deutschsprachige Gebietskörperschaften

- weniger Unterschiede zur SWD zu erwarten
- bei Unterschieden nötig: 5XX: IDN-Austausch; 1XX/4XX: Austausch von Feldinhalten

- ⇒ Fallgruppe wird aufgrund geringen Umfangs für maschinelle Bereinigung im Vorfeld vernachlässigt

c) GKD/SWD/PND: künftige GND-Ansetzungsform entspricht weder GKD- noch SWD-Form (strukturelle Änderungen)

- bedingt durch GND-ÜR
 - z. B.: Behandlung von erläuternden Bestandteilen bei Ortsnamen (bisher kein Namensbestandteil, GND: Namensbestandteil)
- BEISPIEL: GKD/SWD/PND: Frankfurt <Main> = GND: Frankfurt am Main
- nötig: 5XX: IDN-Austausch (GKD/SWD/PND); 1XX/4XX: Austausch von Feldinhalten (GKD/SWD)
- ⇒ für Fallgruppe wird maschinelle Bereinigung im Vorfeld empfohlen

Lösung

- Austausch **häufig** verwendeter Gebietskörperschaften in endgültige GND-Form **vor** GND-Start
- Zeitpunkt: in der Nachbereinigung (Relationierung erfolgt noch in getrennten Beständen); vor GND-Abzug
- Verfahren (Ablauf):
 - Schritt 1: maschineller Austausch auf Basis einer Konkordanz
 - Koko: alte GKD-Form + -IDN; alte SWD-Form + -IDN; GND-Form + -IDN
[sinnvoller Umfang wird noch durch Datenanalyse ermittelt; FA erstellt Koko]
 - IDN-Austausch: erfolgt nur im GKD-Bestand (alte GKD-IDN wird gegen GND-IDN getauscht); in SWD und PND bleibt die IDN erhalten, die expandierte GND-Form wird korrekt ausgeliefert (Voraussetzung: manuelle Korrektur der Ansetzungsformen, s. Schritt 2)
 - Austausch von Feldinhalten: Der komplette Inhalt festgelegter Unterfelder wird in die GND-Form geändert.
 - Schritt 2: manueller Austausch der Ansetzungsformen in GND-Gewinnersätzen von FA in neue Form
- Verfahren (System/Auslieferung):
 - programmtechnische Entwicklung im internen GND-Entwicklungssystem, d. h. Test nur in DNB möglich
 - Auslieferung: nur in der produktiver Datenbank am 19.4.2012; keine Abbildung im GND-Testsystem
- Bereiche mit weniger häufig verwendete Gebietskörperschaften werden der späteren manuellen Korrektur überlassen; ggf. mit Skriptunterstützung

Auswirkungen

- Match-&-Merge-Verfahren: keine?
- Auslieferung: Reduzierung der auszuliefernden Datenmenge auf tatsächliche Umlenkungen durch M&M
(Vermeidung unnötiger Mehrfachauslieferung eines Satzes durch Nebeneffekte des M&M im Bereich IDN-Austausch; Austausch nicht vom M&M erfasster Sätze aufgrund von Korrekturen von Feldinhalten wird vermieden, da bereits in Grundlieferung nötige Korrekturen enthalten)
- Anwender:
 - Für wichtige Gebietskörperschaften sind GND-ÜR korrekt umgesetzt.
 - Manueller Korrekturaufwand sinkt. Redaktionsverfahren wird weniger durch Folgekorrekturen belastet.
 - Akzeptanzverbesserung
 - Schulungen müssen auf den Stand ausgerichtet werden (Liste betroffener Geografika wird zu Verfügung gestellt)

Alternativen

- IDN-Austausch in 551: keine
(falls Problem nicht vor GND-Start bearbeitet wird, erfolgt maschineller Austausch bei M&M; Auslieferung betroffener Sätze nötig, da neue Nummern und Expansionen)
- Austausch von Feldinhalten in 110/410 \$g; 111/411 \$c, 110 \$a/410\$a:
 - nach GND-Start per WinIBW-Skript möglich
 - Problem: Auslieferung => nur vertragliche Mengen pro Woche möglich; Bereinigung zieht sich dann lange hin und muss betreut werden
 - Problem für Anwender: finden nicht korrekte Form vor; Schulungsaufwand, Korrekturaufwand