

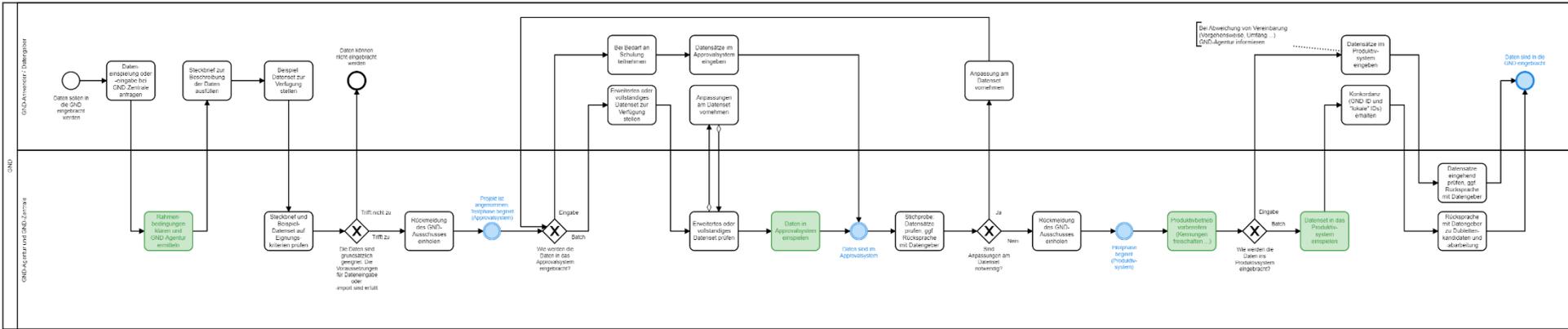
# Importe in die GND

# Lieferformate für Bulk-Uploads

- vollständiger Normdatensatz im Format MARC 21 oder MARC 21-XML oder
- „Datenmanipulation“: Ausschnitt (einzufügendes Feld + GND-ID)
- Zeichensatz: UTF-8
- [Import-Checkliste](#) (Stand: 21.08.2019)

# Import-Prozess

- [Eignungskriterien](#)
- [Workflow des Importprozesses](#)
  - technisch, Stand: 26.06.2017
  - aktuell zugriffsbeschränkt für GND-Anwender\*innen
- [Workflow aus Sicht der Datengeber](#)
  - nicht technisch, Stand: 21.07.2022
  - [„Steckbrief“ \(Template\) zum Einbringen von Datensets in die GND](#)



# Definition Match&Merge (M&M)

- Identische Datensätze sollen gefunden (Match) und zusammengeführt (Merge) werden
- Nicht identische Datensätze werden neu eingespielt, u. U. mit Verlinkung und Markierung zu einer möglichen Dublette (Kandidat)
- Dabei sollen möglichst vermieden werden:
  - dublette Datensätze
  - falsch zusammengeführte Datensätze
  - Verlust von relevanten Informationen
- [M&M Kriterien](#) (zugriffsbeschränkt)

## Vereinfachte Darstellung M&M

- gelieferte Datensätze werden mit vorhandenen Datensätzen in der GND verglichen: der Datensatz mit dem besten Vergleichswert („Matchwert“) ist der Gewinner
- Je nach Matchwert, wird der zu importierende Datensatz entweder
  - mit dem Vergleichsdatensatz gemergt (MATCH/MERGE),
  - als Kandidat zum verglichenen Datensatz neu eingespielt und markiert (CANDIDATE/POSSIBLE MATCH) oder
  - als neuer Datensatz eingespielt (NEW)

## M&M | Schwierigkeiten

- Datensätze mit wenig Informationen schwierig zu matchen
- vorh. GND-Datensatz kann andere Merkmale aufweisen als der zu importierende Datensatz
- auch wenn es nur einen passenden GND-Datensatz gibt, muss es nicht der richtige sein
- Vorgehen: „play it safe“:
  - eher NEW, als falsche Datensätze zu mergen

# Merge

- Merge für einzelne Felder der Datensätze folgt nach definierten Regeln
  - Felder, für die keine Regeln definiert werden, werden nicht hinzugefügt
  - mögliche Regeln:
    - nur Felder hinzufügen, wenn Feld oder Inhalt noch nicht vorh. ist
    - nur Felder hinzufügen, wenn sie über ein bestimmtes Unterfeld (u.U. mit bestimmtem Inhalt) verfügen
    - „bessere“ Feldinhalte (z.B. Verknüpfungen) ersetzen „schlechtere“ Felder (z.B. Strings)
    - etc.

## Kandidaten | Possible Match

- die als Kandidaten eingespielten Datensätze enthalten einen Link auf einen GND-Datensatz, der mutmaßlich ein Match (Dublette) ist (Pica Feld 169, MARC-Feld 885)
- „Entdoppeln“ über WinIBW oder zukünftig über Webcat Entdopplungsfunktion

## Kandidaten | Entdoppeln

- GLEICH: Merge in einem nächtlichen Lauf nach den gleichen Bedingungen wie beim M&M-Verfahren
  - Verliererdatensatz wird umgelenkt auf den Gewinnerdatensatz
  - Umlenkung findet regulär (Di auf Mi) statt
- UNGLEICH: im nächtlichen Lauf wird die Kandidatenmarkierung entfernt

# Vielen Dank!

Fragen?



[gnd-info@dnb.de](mailto:gnd-info@dnb.de)

<https://gnd.network/>