

DNB Leipzig

# Projektbericht „Neuimport aller Titeldaten“, Erfahrungen, Empfehlungen

**Sebastian Bertrams**

Lead Metadata Specialist

---

# WICHTIGE SCHRITTE ZUR MIGRATION

---

# Drei Phasen

1. Analyse, Auswertung und Dokumentation
2. Vereinbarung, Konfiguration und Laden
3. Überprüfung, Bewertung und Aufräumen

---

# ANALYSE, BEWERTUNG UND DOKUMENTATION

---

# Analyse, Bewertung und Dokumentation

1. Beispiele oder vollständiger Datensatzexport
2. Datenmigrationsfragen beantworten
3. Bewertungsbericht erstellt
4. Vorläufigen Translation Table erstellen

# Daten zur Analyse

1. Idealerweise ein vollständiger MARC-Export mit Daten für jedes Element
2. Die Daten müssen alles enthalten, was zur Identifizierung des Datensatzes erforderlich ist - Barcode, Standortdaten, Klassifizierung / Nummern, usw.
3. Alternative Formate sind akzeptabel, müssen jedoch besprochen und der Projektzeitraum nach Bedarf angepasst werden

# Tools um zur Datenanalyse

- Marcedit
- Ultraedit
- Ultracompare

# Marcedit

- MarcEdit ist eine Metadaten-Bearbeitungssoftware, die in erster Linie zur Erstellung und Manipulation von MARC-Datensätzen verwendet wird. Ursprünglich 1999 von Terry Reese für ein großes Projekt zur Datenbank-Bereinigung an der Oregon State University entwickelt.
- Freeware <https://marcedit.reeset.net/>



# Marcedit

- Mit Marcedit lassen sich Valdierungen auf Grundlage des Marc Regelwerks durchführen.
- Es hilft im Vorfeld schon Duplikate zu finden und diese zu entfernen.
- Es lassen sich fehlende Felder ergänzen und wenn nicht vorhanden mit Platzhaltern füllen.

# Ultraedit

- Mit Ultraedit lassen sich große Dateien öffnen und komfortabel bearbeiten.
- Es ist hier möglich gerade im Bereich der Testdatenlieferung einfache Skripte zur Transformation der Daten auszuführen
- Es ist möglich mit *regular expressions* zu arbeiten

# Ultracompare

- Mit Ultracompare ist es möglich den Effekt von Skripten auf große Datenmengen aufzuzeigen.
- Beim gegenüberstellen zweier Dateien werden übersichtlich die Veränderungen aufgezeigt.

# Frage zur Datenmigration

1. Sagen Sie uns, was Sie über Ihre Daten wissen und was Ihnen wichtig ist
2. Geben Sie an welche Daten Sie bereitstellen möchten
3. Geben Sie Besonderheiten an die der Datenanalyst bei der Auswertung Ihrer Daten beachten muss

# Bewertungsbericht

Der Datenanalyst liefert eine Auswertung der übermittelten Daten:

Liefert Statistiken und bestätigt, dass Daten vorhanden und gültig sind

1. Identifiziert, wo eine Vorverarbeitung der Daten erforderlich ist
2. Erstellt einen vorläufigen Translation Table

# Translation Table

Institution Name at OCLC	Deutsche Nationalbibliothek	Deutsche Nationalbibliothek Leipzig	Deutsche Nationalbibliothek Frankfurt am Main	Deutsche Nationalbibliothek, Deutsches Musikarchiv	Deutsche Nationalbibliothek, Deutsches Exilarchiv 1933-1945	Deutsche Nationalbibliothek, Neuerscheinungsdienst (New Release Service)
WorldCat Registry ID	195174	48919	48921	145433		195516
ISIL	DE-101	DE-101a	DE-101b	DE-101c		DE-101e
ISIL	OCLC-GW/DNB	OCLC-GW/DEB	OCLC-GW/DDB	OCLC-DE/DNM		OCLC-DE/DNE
OCLC Account Name	DEUTSCHE NATIONALBIBLIOTHEK	DEUTSCHE BUCHEREI	DEUTSCHE NATIONALBIBLIOTHEK	DEUTSCHE NATIONALBIBLIOTHEK DEUTSCHES M	DEUTSCHE NATIONALBIBLIOTHEK DEUTSCHES E	DEUTSCHE NATIONALBIBLIOTHEK NEW RELEASE SERVICE
Place/town	over-all group symbol	Leipzig	Frankfurt am Main	Leipzig	Frankfurt am Main	Frankfurt am Main
OCLC Customer Number	01EUR 20030641	01EUR 20030641	01EUR 200309109	01EUR 20165218	01EUR 20165219	01EUR 20163872
OCLC Symbol	GW/DNB	GW/DEB	GW/DDB	DE/DNM	DE/DNE	DE/DND
OCLC Institution Number	48828	36754	34508	72940	72941	73209
Information	over-all group symbol		Bis ca. Frühjahr 2018 erfolgte unter diesem Symbol der Import der DNB-Titeldaten (nur mit dem Bestandsnachweis F)			OCLC-Symbol neu seit 12/2018 DE/DND ist kein alleinstandes Symbol, sondern auch Teil der Gruppe nur mit eigener Collection (Collection for unpublished Records).
data sync collection ID	Titeldaten mit OCN: 1013423					1023921
	Titeldaten ohne OCN: 1029086					
OCLC-Scripte (Datenmanipulation)	1. MARC-Feld 689: Entfernung nur der Unterfelder mit Großbuchstaben. 2. MARC-Feld 008/06 (Leader/07): Manipulation bis Korrektur der Konversion bei DNB im Herbst 2018. 3. MARC-Feld 008/06: Manipulation 008/06 wird von "s" beim Import zu "c" 4. ?					Identisch mit GW/DNB
Sonstige	1. MARC-Feld 850: Auswertung der ISIL als Bestandsnachweis. Ist 850 nicht besetzt handelt es sich um "Kopfzeile", in diesen Fällen wird durch OCLC-Manipulation trotzdem der Bestandsnachweis für L (DE-101a) und F (DE-101b) gesetzt. 2. Entfall der DNB-Titeldatensatz mehrere OCNs erfolgt die MATCH-Prüfung mittels Titel					1. MARC-Feld 850: KEINE Auswertung (es darf kein Bestandsnachweis erzeugt werden). 2. ND-Titeldaten bleiben 365 Tage im WorldCat auch ohne Bestandsnachweis verfügbar. 3. DNB-Titeldaten dürfen ND-Titeldaten vollständig überschreiben inkl. Erzeugung Bestandsnachweis aufgrund Angaben in MARC-Feld 850.

Wir können eine Entwurfstabelle erstellen, indem wir ein Skript gegen Ihre Testdaten ausführen.

# Translation Table Mapping

1. Ihre Daten müssen sich alle in Unterfeldern des Datensatzes befinden-wir können bei Bedarf bibliographische Daten in ein entsprechendes Feld kopieren.
2. Verwenden Sie Wildcards \*, um einen beliebigen Wert darzustellen.
3. Die Tabelle wird in der Reihenfolge erstellt, von spezifischen zu allgemeinen Feldern.

# Translation Table Mapping – draft in progress

	A	B	C	D	E	F	G	H	I	J	K	L			
1	Aleph			WMS									Migration		
	930 \$b Branch Code	930 \$c Shelving Location	930 \$f Item Type	930 \$E Item Status	852 \$b Holding library branch code	Branch display name in WorldCat Discovery	852 \$c Shelving Location	008/20	008/21	New location	Do not migrate	Default			
2															
3	ARCH	ARCOL	Reference	*	YT2E	SCA Storage	Not Available [bgasref]	u	u						
4	ARCH	ARCOL	*	*	YT2E	SCA Storage	Not Available [bgas60d]	u	u						
5	ARCH	BGAS	Reference	*	YT2E	SCA Storage	Not Available [bgasref]	u	u						
6	ARCH	BGAS	*	*	YT2E	SCA Storage	Not Available [bgas60d]	u	u						
7	ARCH	GPWA	*	*	YT2B	Francis Close Hall Campus	Archives QU024 Glos Poets, Writers & Artists Collection [ref]	u	u						
8	ARCH	JOURN	60 Day	*	YT2B	Francis Close Hall Campus	Archives QU024 Journals [60d]	u	u						
9	ARCH	JOURN	Reference	*	YT2B	Francis Close Hall Campus	Archives QU024 Journals [ref]	u	u						
10	ARCH	LOC	*	*				u	u		x				
11	ARCH	MACOL	*	*				u	u		x				
12	ARCH	MAPS	*	*				u	u		x				
13	ARCH	MAREF	*	*				u	u		x				
14	ARCH	OU	*	*				u	u		x				
15	ARCH	OVER	*	*				u	u		x				
16	ARCH	REF	*	*				u	u		x				
17	ARCH	SLIDE	*	*				u	u		x				
18	ARCH	TRCPB	*	*				u	u		x				
19	ARCH	*	*	*	YT2B	Francis Close Hall Campus	Migration Exception	u	u						
20	CROSS	ENQ	Reference	*	YT2D	Central Office	Enquiry Desk [ref]	u	u						
21	FCHLC	3DAY	3 Day	*	YT2B	Francis Close Hall Campus	Main [3d]	u	u						
22	FCHLC	3DAY	3 Week	*	YT2B	Francis Close Hall Campus	Main [3wk]	u	u						
23	FCHLC	3DAY	7 Day	*	YT2B	Francis Close Hall Campus	Main [7d]	u	u			+			



# Draft Scope Statement

Eine Aufzeichnung des vorgeschlagenen Migrationsplans:  
Beschreibung der erforderlichen Vorverarbeitung

1. Eine Aufzeichnung der erforderlichen Scripte
2. Mappings für LBD Daten
3. Eine Vereinbarung zwischen uns und einem Weg, um alle Änderungen, die zu einem späteren Zeitpunkt auftreten können, zu verwalten.

---

# VEREINBARUNG, KONFIGURATION UND LADEN

---

# Vereinbarung, Konfiguration und Laden

1. Erstellen des Scope Statement:
  - a) Vorverarbeitung definiert
  - b) LBD mapping
  
2. Feld Zusammenführung
3. Laden der Testdaten
4. Finales laden

# Pre-processing der Titeldaten

1. Problematische Validierungsprobleme entfernen, z.B. unvollständige Leader-Daten
2. Aussortieren unerwünschte Datensätze, die nicht identifiziert werden können mit dem Translation Table
3. Verschieben Sie LBD-Felder auf lokale Tags, um ungewolltes Merging zu vermeiden
4. Entfernen von Validierungsfehler aus ausgewählten LBD-Feldern (689\$X, 259)
5. Aufteilen der Datensätze nach Typ, z.B. nach Datentyp, Vorhandensein von OCLC-Nummer (OCN), etc.

# LBD mapping

- LBD – Lokale Bibliographische Daten
- Verbunden mit dem Master-Datensatz, aber nur sichtbar und editierbar in Ihrem lokalen System
- Nützliche Möglichkeit, Daten in der Migration zu speichern
- Re-tag zu einem lokalen ' 9 '-Tag, um zu vermeiden, dass sie mit einem Master-Rekord zusammengeführt werden
- Validierung ist weiterhin wichtig
- Einige Felder erlauben indicators/Unterfelder, z.B. 599

# LBD mapping

LSN			System.Supplied@2014-07-15,09:32:05
OWN			UKIT2
LDR			c
001			353718899
005			20170808090518.0
004			222559278
500	<input type="checkbox"/>	<input type="checkbox"/>	Local note.
593	<input type="checkbox"/>	<input type="checkbox"/>	Custom binding in cork.
655	<input type="checkbox"/>	7	test \$2 aat

# LBD mapping

<b>Subjects:</b>	<a href="#"><u>Deakin, Roger Journeys Great Britain.</u></a> <a href="#"><u>Deakin, Roger.</u></a> <a href="#"><u>Swimming Great Britain.</u></a> <a href="#"><u>Swimming.</u></a> <a href="#"><u>Travel.</u></a> <a href="#"><u>Voyages and travels.</u></a> <a href="#"><u>Great Britain Description and travel.</u></a> <a href="#"><u>Great Britain.</u></a> <a href="#"><u>test</u></a>
<b>Genre:</b>	test
<b>Notes:</b>	Originally published: London: Chatto & Windus, 1999. Local note. Custom binding in cork.

LBD wird innerhalb  
der Master-Record-  
Anzeige angezeigt  
und nach  
Möglichkeit  
dedupliziert.

# Field merge/transfer

## Call Number Group

055 - Classification Numbers Assigned in Canada  
060/096 - Medical Call Numbers  
070 - National Agricultural Library Call Number  
080 - Universal Decimal Classification Number  
083 - Additional Dewey Decimal Classification Number  
084 - Other Classification Number  
085 - Synthesized Classification Number Components  
086 - Government Document Classification Number  
050/090 - Library of Congress Classification Numbers  
082/092 - contain Dewey Decimal Classification Numbers

## 6xx group:

600 (regardless of indicator)  
610 (regardless of indicator)  
611 (regardless of indicator)  
630 (regardless of indicator)  
648 (regardless of indicator)  
650 (regardless of indicator)  
651 (regardless of indicator)  
655 (regardless of indicator)

## Note Group

505 - Formatted Contents Note  
508 - Creation/Production Credits Note  
511 - Participant or Performer Note  
520 - Summary Note

Für verschiedene Felder gelten unterschiedliche Regeln unter Berücksichtigung von Indikatoren und Einzigartigkeit und ob der Tag bereits im Master-Rekord vorhanden ist. Wenn eines dieser Felder für Sie wichtig ist, sagen Sie dies bitte dem Datenanalysten.



# Testdaten laden

- Der Fokus liegt auf dem Mapping und Überprüfung der LBD's
- Einige Beispieldatensätze die möglichst viele verschiedene Varianten der Daten abbilden.

# Testdaten laden

- Das laden der Daten kann sich auf " staging " beschränken, sie sind nicht sichtbar in WorldCat und werden nicht indiziert oder an Matching Prozesse gesendet, um Änderungen in WorldCat zu vermeiden
- Datensätze können mit der OCLC-Nummer und speziellen Suchanfragen des Datenladers abgerufen werden.

# Finales Laden

- Scan-delete löscht alle Bibliotheksaufzeichnungen von WorldCat vor dem Laden
- Datensätze können in kleinere Dateien aufgeteilt werden und mehr als eine Sammlungskonfiguration verwenden
- Matching and merge findet statt – ‘set holds’ auf bestehenden Master-Rekord

# Finales Laden

- Wenn kein "match" gefunden wird, wird der Titel hinzugefügt
- Alle Daten werden geprüft– Datensätze mit level 2 or 3 errors bleiben 'dark' (werden nicht indexiert) und müssen nach dem Laden bearbeitet werden
- Die Bibliothek sollte nach der gleichen Methoden überprüfen ob die Daten korrekt geladen wurden wie für das Testladen.

---

# ÜBERPRÜFUNG, BEWERTUNG UND AUFRÄUMEN

---

# Überprüfung, Bewertung und Aufräumen

1. Datenlader überprüft die Validierungsfehler und macht Korrekturen, um Datensätze „light“ (sichtbar) zu machen
2. Bibliothek meldet alle Probleme, die behoben werden müssen
3. Reports werden generiert
4. Datensätze die nicht migriert sind werden gemeldet
5. Datensätze die ‘dark’ bleiben werden gemeldet

# Noch Fragen?

**Sebastian Bertrams**

Sebastian.bertrams@oclc.org

**Because  
what is  
known must  
be shared.®**