

Grundlagen Metadaten: Lieferung Nicht-EDM

Gregor Pirgie, Team Kulturpool

09.07.2024

Inhalt

1. Glossar
2. Gemeinsame Ziele
3. Anforderungen an die Lieferung von Nicht-EDM
4. Kurze Einführung in EDM
5. Ablauf des Mappings
6. Mögliche Probleme erkennen
7. Mapping Tabelle(n)
8. Was passiert nach der Anbindung?

Glossar

- Datenformat vs. Metadatenformat
- Pipeline
- Kardinalität
- Restrictions
- Pflichtfelder
- Values, und Value-Types
- Kontrollierte Vokabulare
- Identifier ausliefern = URI
- Validierung
- Syntax im Kontext eines Metadatenformats
- Mapping
- Literals und Refs
- Enrichement
- Strukturierte Daten
- Domäne
- Harvest
- Harvesting Zyklen
- API / Schnittstelle
- OAI-PMH
- REST
- GraphQL
- Semantik
- Record

Gemeinsame Ziele

Warum machen wir das alles?

Nutzen

- Vernetzung
- Deduplizierung
- Übersetzung
- Semantik
- Durchsuchbarkeit / Auffindbarkeit
- Enrichement
- Austausch von Wissen
- Archivierung / Sicherung

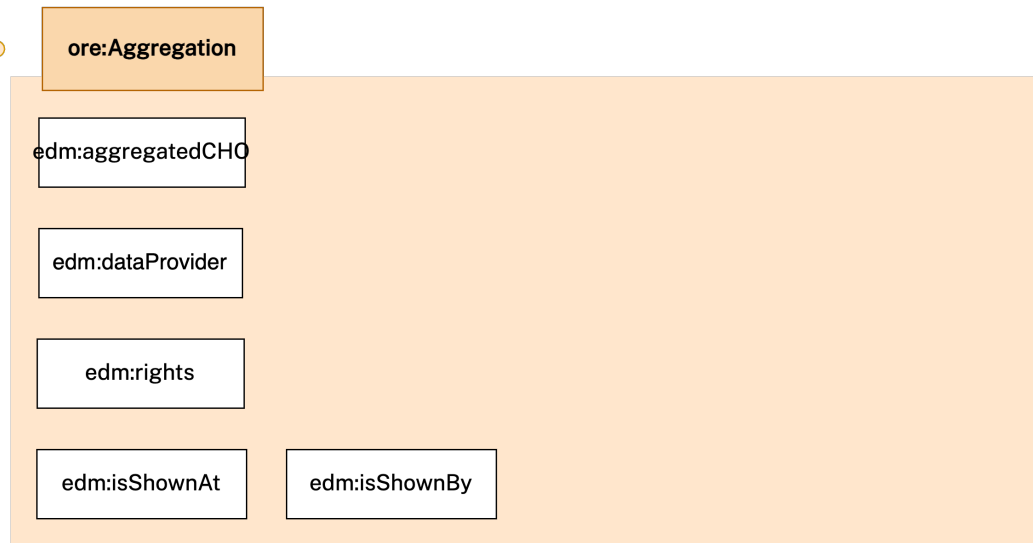
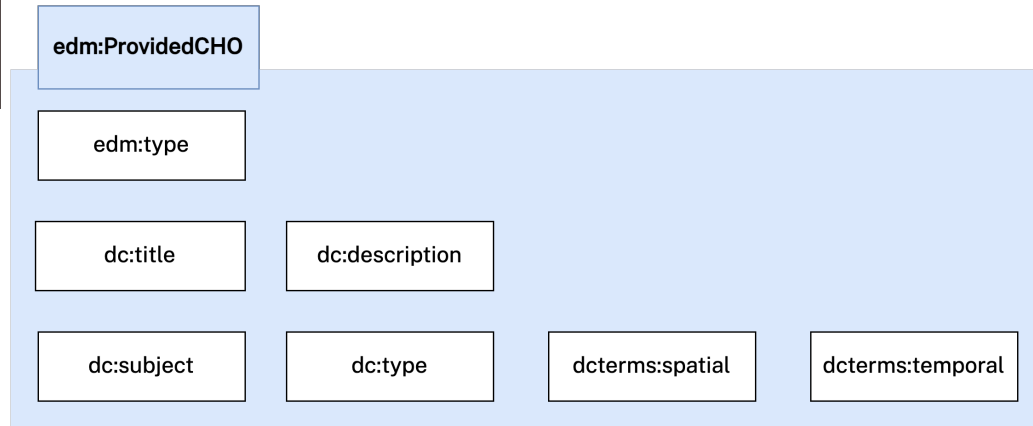
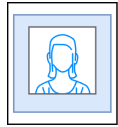
Nachhaltigkeit

- Automatisierung
- Einheitliche Daten
- Semantische Genauigkeit
- Wiederverwendbar
- Ressourcenschonend

EDM: Kurze Einführung

- Europeana Datamodel
- RDF-basiert, aber wird nur im RDF/XML-Format angenommen
- nutzt XML vor allem, um constraints zu definieren
- wir müssen gegen diese constraints validieren
 - Pflichtfelder
 - Optionale Pflichtfelder
 - Value-Types
 - Cardinality-Constraints

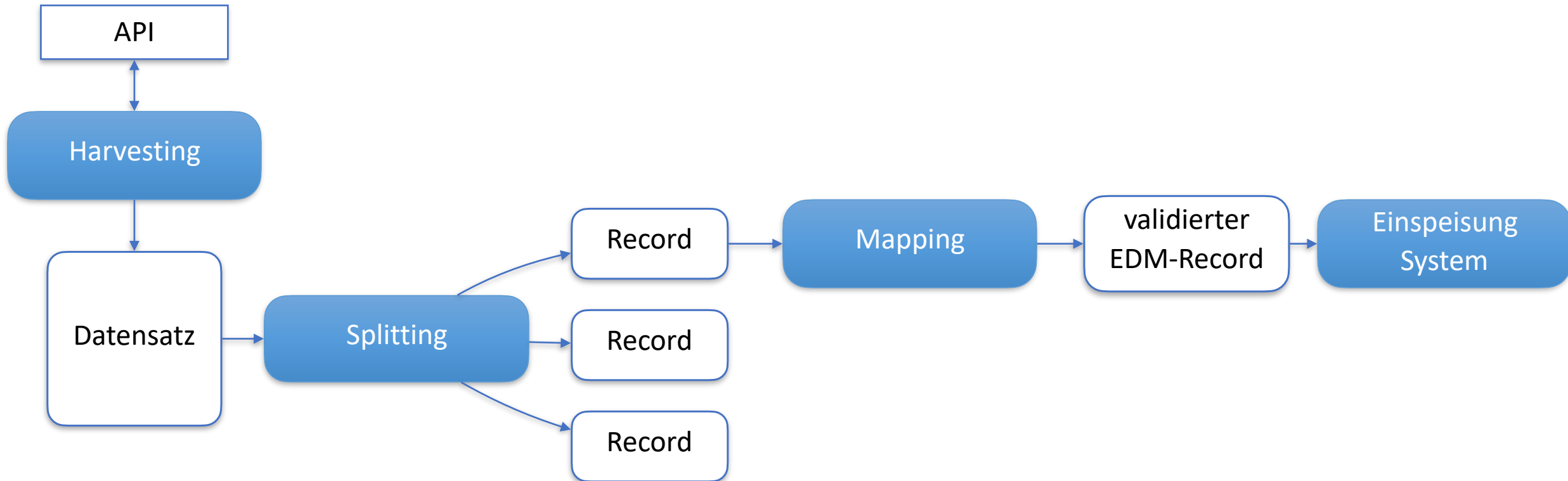
EDM Pflichtfelder



- edm:type: Medientyp IMAGE, TEXT, VIDEO, AUDIO, 3D
- dc:title: Titel des Objekts
- dc:description: Beschreibung des Objekts
- dc:subject: Thema/Inhalt des Objekts
- dc:type: Typ/Art des Objekts
- dcterms:spatial: Räumliches Thema des Objekts
- dcterms:temporal: Zeitliches Thema des Objekts

- edm:aggregatedCHO: Referenz auf die edm:ProvidedCHO Instanz
- edm:dataProvider: Bereitstellende Institution
- edm:rights: Referenz per URL auf Lizenz des Digitalisats, zB <http://creativecommons.org/licenses/by/4.0/>
- edm:isShownAt: Referenz auf Objekt im Kontext (Institutionseigener Online Katalog)
- edm:isShownBy: Referenz auf Digitalisat (zB: Bild-URL)

Einblick: Pipelines



Anforderungen an gelieferte Datensätze

EDM-Äquivalenz

- Pflichtfelder
- Optionale Pflichtfelder
- Cardinality restrictions
- URIs
- Value-Constraints

Für selektives Harvesting

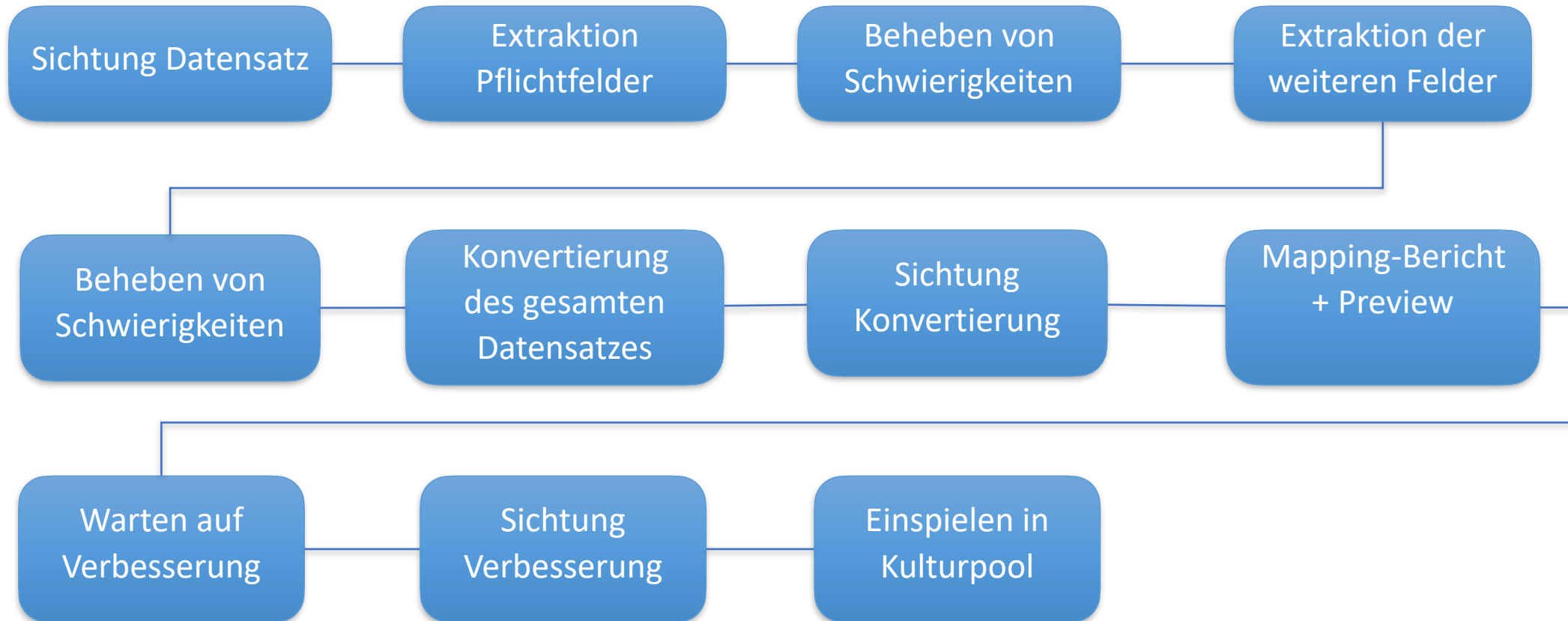
- Datestamps (letzte Veränderung am Record)
- Gelöschte Records markieren

Anbindungs-Workflow

Mapping-Prozess



Mapping-Prozess



Typische Schwierigkeiten in den Input-Daten

1. Ausgangsformat validiert nicht
2. Im Ausgangsformat werden nicht die Standardfelder verwendet, sondern Ausweichmöglichkeiten
3. Fehlende Bilder für edm:isShownBy
4. Strukturelle Fehler im Input-Format
5. Semantische Fehler im Input-Format
6. Divergenz der Records unterschätzt
7. HTML-Tags oder Markup in Texten
8. Leerzeichen in URIs
9. Vokabulare nicht verwendet
10. Vokabulare verwendet, aber nicht die URI eines Terms ausgeliefert, sondern den "String"
11. Fehlende Sprach-Tags
12. Rechts-Statements müssen URIs sein, nicht "Strings"

Mapping: Generelle Herausforderungen

- Grenze zu Interpretation ist fließend beim Mapping
- Interpretation sollte von der Institution gestützt werden (Sie haben die Expertise)
- Konzeptuell denkt man oft nur an einen Record, aber wir mappen zig-tausende Records, die sich mitunter stark unterscheiden können (innerhalb einer Institution)
- Wir können nur das mappen, was wir kennen. Wir können nicht ein generelles Mapping von LIDO auf EDM schreiben
- Kommen neue Felder hinzu, müssen wir das Mapping anpassen
- Ändern sich existierende Felder, kann das mitunter nicht auffallen – wir müssen das Mapping auch anpassen

Mapping-Tabelle

- Sollte von Ihnen geliefert werden, bevor wir mit dem Mapping beginnen
- Mappt jedes konvertierbare Element im Ausgangsformat auf das entsprechende Konstrukt im EDM
- Je nach Ausgangsformat mehr oder weniger kompliziert
 - Beispiel LIDO: wir brauchen den Pfad, den zu beachtenden Kontext, ob der Wert eines Feldes aus Attributen oder Text entnommen werden soll und alle zu berücksichtigenden Bedingungen
- Siehe: <https://wissen.kulturpool.at/books/aggregation/page/mapping-tabelle>

Nach der Anbindung

- Festlegen des Harvesting-Intervalls
- Infrastruktur und Mapping muss gepflegt werden
- Updates in den Ausgangsdaten sollten kommuniziert werden, damit wir das Mapping anpassen können

Überblick Arbeitsschritte

- Ausfüllen technisches Anbindungsformular
- Bereitstellen der Schnittstelle und der genauen Abfrage
- Auslieferung mit Dateistamps und Deletions sicherstellen
- Kooperationsvertrag unterschreiben
- Bereitstellung Mapping-Tabelle
- Abnahme Preview und Einarbeiten von notwendigen Verbesserungen
- Endabnahme
- Nachsorge: Wartung Schnittstelle, Dateistamps aktuell halten, Veränderungen an den Ausgangsdaten kommunizieren

Gibt es noch Fragen?

- Knowledgebase: <https://wissen.kulturpool.at/>
- E-Mail: info@kulturpool.at
- Online Café 15.07., 14:00
- Forum: <https://forum.kulturpool.at/>