

**DIE ÖSTERREICHISCHE BIBLIOTHEKENVERBUND  
UND SERVICE GMBH**

**obv sg**

# **SPRINGER E-BOOK WORKFLOW IM OBV**

**DACHELA, 4. OKTOBER 2023, ONLINE**

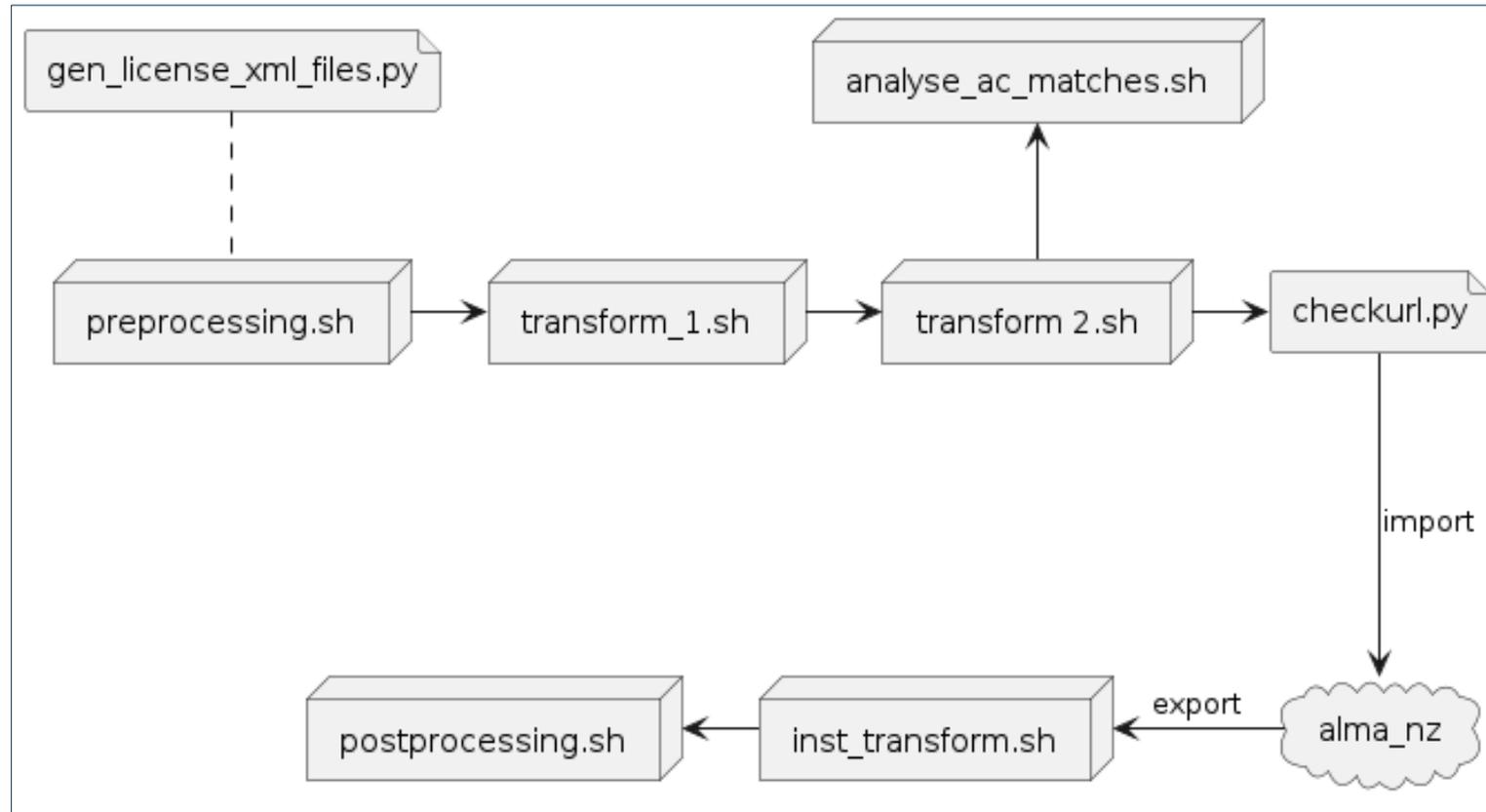
**PHILIPP KOZJAK**

## AGENDA

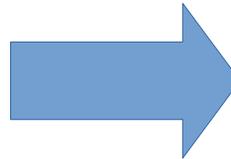
- Workflow monatlich
- Lückenfindung
- Probleme/ Ausblick

# WORKFLOW MONATLICH

- Der Workflow wird in bash ausgeführt und verwendet folgende Tools:
  - marc2xml
  - xslt/ xquery
  - python3
  - xmllint
  - alma-js
  - sftp



Paket	Kürzel hbz	OBV	FHO	BOK	UBG
ZDB-2-BHS	ebs	2013			
ZDB-2-BSP	-	2016-2023			
ZDB-2-BUM	-	2016-2023	2016-2023		
ZDB-2-CMS	ecms	1893-2004; 2011-2012; 2016-2023		2011-2012	
ZDB-2-CWD	epcwd	2013-2023	2013-2023		
ZDB-2-ECF	-	2016-2023	2016-2019		



```

<licenses>
  <license>
    <sigel>ZDB-2-CMS</sigel>
    <year>2011</year>
    <year>2012</year>
  </license>
  <license>
    <sigel>ZDB-2-SBL</sigel>
    <year>2011</year>
    <year>2012</year>
  </license>
  <license>
    <sigel>ZDB-2-SNA</sigel>
    <year>2007</year>
    <year>2008</year>
  </license>
</licenses>
    
```

- Marc File Download von Springer SFTP
- Umwandlung in Marc XML
- Deduplizierung anhand der DOI in Marc 024

- Filterung auf OBV Lizenzen
- `<xsl:if/>` mit mit einem Attribut `test="`  
    `( key ('licenseBySigel', $sigel, $licensesDoc)`  
        `intersect`  
        `key ('licenseByYear', $year, $licensesDoc)`  
    `) or $zdb-2-sob"`

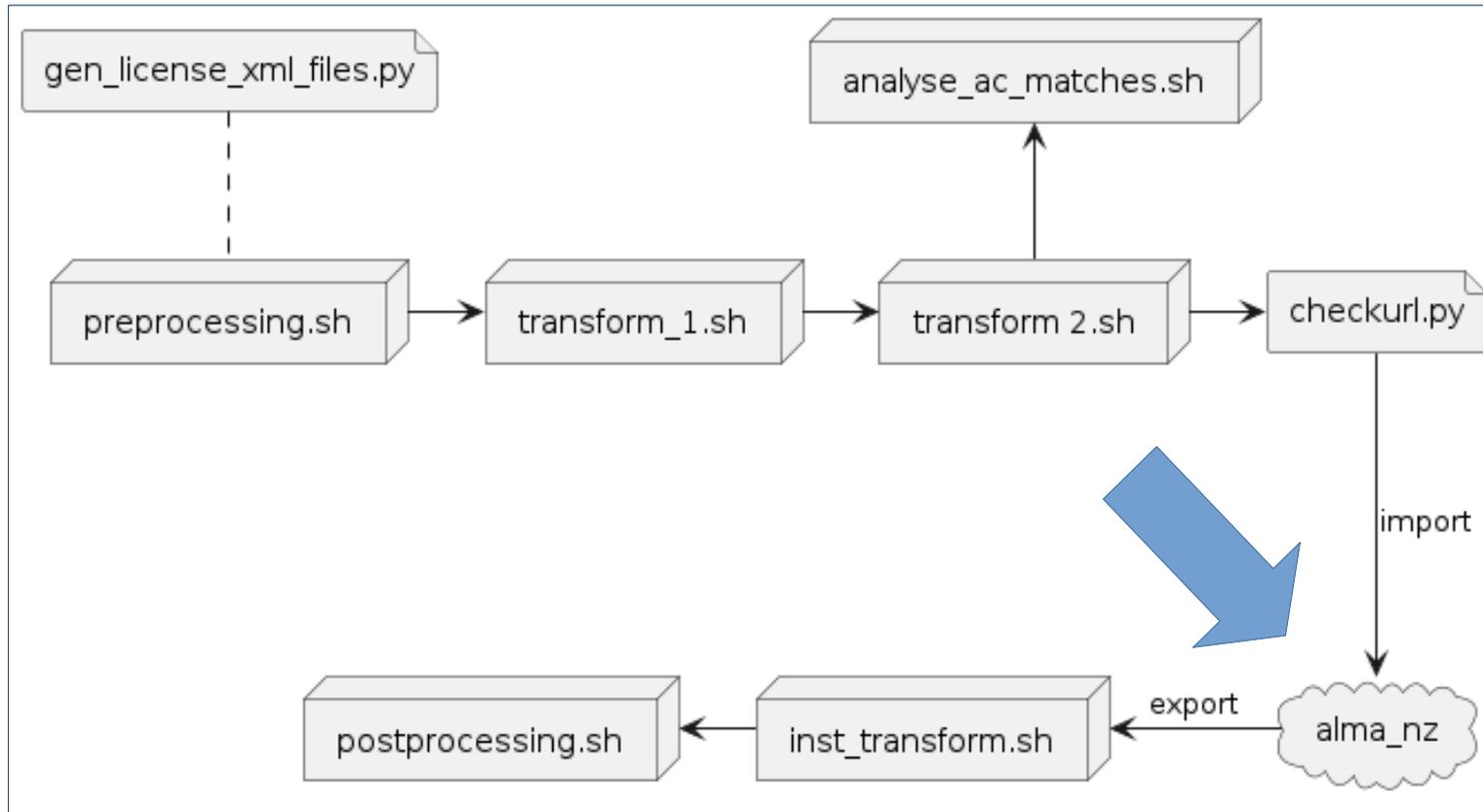
- Manipulation der Marc XML Records
  - Fixtexte setzen (etwa LDR, 040, 506 open access, 856)
  - Felder oder Subfelder entfernen (etwa 001, 003, 020 \$9, 830)
  - Felder manipulieren
    - Entferne ISBD Interpunktion in 100, 110, 245, 250, 264, 490, 505, 520, 650, 700
    - Setze Nichtsortierzeichen in 245
    - Normalisierung etwa 008, 300, 776
  - Felder hinzufügen
    - Etwa 040, 041, 044
  - Funktionen:
    - `local:entferne_Deskriptionszeichen`
    - `local:entferne_Deskriptionszeichen_personen`
    - `local:replace_008`
    - `local:Nichtsortierzeichen`

- Validierung der transformierten Daten durch RNG (sehr granular)
- SRU queries nach der DOI in der OBV Networkzone
  - Datensätze mit AC Nummer (OBV Sätze) gematched?
    - Weitere Analyse, siehe nächste Folie
  - Datensätze aus der CZ gematched?
    - Dublett einspielen

- DOI Matches von AC Sätzen
  - Datensatz ist ein manuell erfasstes Springer E-Book
  - Datensatz ist eine Print Aufnahme mit Springer DOI in Marc 024
  - Datensatz ist ein Aggregator E-Book mit Springer DOI in Marc 024
  - (Datensatz wurde in der Vergangenheit von Springer bereits geliefert)

- HTTP Response Status Codes für URLs prüfen
- `follow_redirects=True`

```
(venv) pk2@ada:~/springer_workflow$ checkurl data/august/urls.txt
200 https://doi.org/10.1007/978-3-030-85040-1 https://link.springer.com/referencework/10.1007/978-3-030-85040-1
200 https://doi.org/10.1007/978-3-662-47493-8 https://link.springer.com/referencework/10.1007/978-3-662-47493-8
200 https://doi.org/10.37307/b.978-3-503-21114-2 https://chooser.crossref.org/?doi=10.37307%2Fb.978-3-503-21114-2
200 https://doi.org/10.1007/978-1-0716-2466-1 https://link.springer.com/referencework/10.1007/978-1-0716-2466-1
200 https://doi.org/10.1007/978-1-0716-2388-6 https://link.springer.com/referencework/10.1007/978-1-0716-2388-6
200 https://doi.org/10.1007/978-3-030-93582-5 https://link.springer.com/referencework/10.1007/978-3-030-93582-5
200 https://doi.org/10.1007/978-3-662-65093-6 https://link.springer.com/referencework/10.1007/978-3-662-65093-6
```



- Filterung aus der Gesamtmenge der in den Verbundkatalog eingespielten Daten
- Nach Jahr und Sigel
- Für jede teilnehmende Institution
- Gleiche XML Lookup Technik wie bei der Filterung der OBV relevanten Daten

- Generierung einer CSV Tabelle für Statistik der NZ-Monatseinspielung nach Sigel und Jahr
- Upload institutionsspezifischer Marc XML Files auf OBVSG FTP für die jeweiligen User
  - Daten werden lokal eingespielt
  - Import Profile: Use Network = True, Match auf ID in Marc 035, Prefix (AT-OBV)
- OBV Kataloganreicherung mit Inhaltsverzeichnissen
  - Nutzung der DOI oder ISBN um die URL zu generieren

# LÜCKENFINDUNG

- Lücken zwischen NZ und IZ systematisch identifizieren
- Für jede Institution wird
  - Pro Sigel und Jahrgang
  - NZ und IZ via SRU befragt
  - Eine Übersichtstabelle generiert

Institution	Sigel	Year	IZ Results	NZ Results	IZ Results == NZ Results ?	Diff	Query
43ACC_UBBW	ZDB-2-CMS	2011	179	180	False	1	alma.mms_resource_type="Book - Electronic" AND alma.other_system_number all (AT-OBV) AND alma.local_field_912 all ZDB-2-CMS AND alma.main_pub_date=2011

- Neue SRU Abfrage für jede Zeile mit Abweichung in der Übersichtstabelle
- File mit den AC Nummern fehlender Datensätze schreiben
  - pro Institution, Sigel und Jahrgang

# PROBLEME/ AUSBLICK

- DOIs mit multiple resolution
  - Betroffene Datensätze können nicht zum Zeitpunkt der Bereitstellung ermittelt werden
- ORCID
  - Wurde bisher bei der Transformation ignoriert
  - Wird ab jetzt in Marc 100/ 700 \$9 geschrieben
  - Anreicherung der ORCIDs für bestehende Datensätze? (beträfe 2804 ORCIDs in Marc 100 sowie 42228 ORCIDs in Marc 700)
- ORCID/ GND
  - SRU Lookup der GND-ID durch die ORCID; automatische Verlinkung des Normsatzes in 100/ 700?
- Refaktorisierung des Quellcodes

**VIELEN DANK FÜR DIE AUFMERKSAMKEIT!**

Kontakt: [philipp.kozjak@obvsg.at](mailto:philipp.kozjak@obvsg.at)