Dokumentation
der Bearbeitung, der Migration und des Ingests nach DIMAG der Dateien (genuin digitale Unterlagen) des Bestandes A 243 (Pressemitteilungen)

(Anleitung für die Arbeit im Landeskirchlichen Archiv Stuttgart)

Bearbeiter: Uwe Heizmann
www.uwe-heizmann.de

Stand: 18.11.2025

Inhalt

1 Vorwort ... 3
2 Verwendete Programme ... 4
3 Arbeitsordner ... 5
4 Änderung der Dateinamen ... 6
5 Analyse der Dateien ... 7
5.1 Dateiformate ... 7
5.2 Migrationstests ... 8
5.2.1 Ordnerstruktur ... 8
5.2.2 DOT-Dateien (Microsoft Word 97-2003-Vorlagen) ... 8
5.2.3 Unterschiedliche DOC-Dateien ... 9
6 Vorbereitung für die Migration ... 11
6.1 Ordnerstruktur auflösen ... 11
6.2 Dateien sortieren ... 11
7 Migration ... 12
7.1 Migrationsvorgehen ... 12
7.1.1 Mit Adobe Acrobat: DOC und DOCX nach PDF/A ... 12
7.1.2 Mit Adobe Acrobat: PDF nach PDF/A ... 12
7.1.3 Mit LibreOffice: DOC nach DOCX ... 13
7.1.4 Mit LibreOffice: DOC nach PDF/A ... 13
7.2 Migrationsschritte ... 14
7.3 Validierung ... 15
7.4 Archivformate ... 15
7.5 Migrationsbericht ... 16

8 Erkenntnisse, Empfehlungen ... 17
9 Ordnerstruktur wiederherstellen ... 18
9.1 Für Repräsentation 1 ... 18
9.2 Für die Repräsentation 2 ... 18
9.3 Für Repräsentation für die Online-Stellung ... 19
10 Ingest nach DIMAG ... 20
10.1 Projektordner ... 20
10.2 Metadaten-Datei ... 20
10.3 IngestTool: Paketierung ... 21
10.4 IngestTool: Quellelement Editor ... 22
10.5 IngestTool: Mappingwerkzeug ... 23
10.6 IngestTool: Mappinglauf starten (Ingest nach DIMAG) ... 24
11 Online-Stellung ... 26
11.1 Dateinamen ergänzen ... 26
11.2 Anbinden der Dateien an die Verzeichnungseinheiten in ACTApro Desk ... 26
11.3 Export aus ACTApro Desk ... 26

1 Vorwort

Dieses Dokument beschreibt die Bearbeitung der Pressemitteilungen und dazugehörigen Dateien bis einschließlich der Verzeichnungseinheit Nr. 343 (Dezember 2022). Sie gilt aber auch für die Bearbeitung und Migration zukünftiger Pressemitteilungen oder auch anderer Dateisammlungen, wobei zukünftige Pressemitteilungen aufgrund der potenziell kleineren Anzahl an unterschiedlichen, v.a. älteren Dateiformaten einfacher zu bearbeiten sein dürften.

Grundsätzlich werden die Original-Dateien als Repräsentation 1 übernommen. Hiervon ausgenommen sind PDF-Dateien. Diese werden nach PDF/A migriert und diese neuen PDF-Dateien als Repräsentation 1 übernommen.
Als Repräsentation 2 werden nur Archivformate von (wahrscheinlich) nicht langzeitfähigen Dateiformaten, z.B. PDF/A- von DOC-Dateien, übernommen. Nur dieses Archivformate, also das Delta zu Repräsentation 1, bilden die Repräsentation 2.
Hinsichtlich des Ingests nach DIMAG ist es wichtig zu wissen, dass die Repräsentationen gleichzeitig hochgeladen werden müssen. Eine nachträgliche Ergänzung der Repräsentation 2 an die bereits vorhandene Repräsentation 1 ist derzeit (November 2023) nicht möglich.
Neben den Repräsentation 1 und 2 wird für die Pressemitteilungen außerdem eine Repräsentation für die Online-Stellung über ACTApro Benutzung erstellt. Diese besteht hauptsächlich aus der Repräsentation 2 und den nicht zu migrierenden Dateien der Repräsentation 1. Genaueres siehe Kapitel 9.3.

Die Anleitung ist eine Mischung zwischen allgemeingültiger Anleitung und Bericht über Durchgeführtes. Deshalb wechseln auch die Zeitformen ab. Dies ist beabsichtigt, damit ersichtlich ist, was allgemeine Bearbeitungsschritte sind und welche Dinge vorkommen können.

Abhängig von der Anzahl der Ordner und der unterschiedlichen zu migrierenden Dateiformate sowie evtl. anderer Faktoren bietet sich das in dieser Anleitung beschriebene, arbeitsschrittreiche Vorgehen oder ein Vorgehen mit weniger Arbeitsschritten an.

2 Verwendete Programme

Für die Erfassung der ursprünglichen Dateinamen

- FilelistCreator/DateilistenSchreiber, Version 22.3.3

Für das Ändern der Dateinamen[1]

- Ant Renamer, Version 2.12

Für das Kürzen der Dateinamen[2]

- Advanced Renamer 3.88

Für das Löschen von Thumbs.db

- FileReaper[3]

Für die Analyse der Dateien

- IngestList, Version 6.6.10
    - erstellt eine XML-Datei, in der die analysierten Dateien einzeln aufgeführt sind (wird benötigt für das Sortieren der Dateien, vgl. Kapitel 6.2)
- TreeSize Professional, Version 9.0.3.1852
- DROID, Version 6.6.1[4]
    - reine Analyse der Dateiformate (deshalb deutlich schneller als IngestList), kann statistische Auswertungen erstellen

Für den Umbau der Ordnerstruktur

- Ant Renamer, Version 2.12
- Powershell

Für die Migration der Dateien

- Adobe Acrobat Pro, Version 2020.005.30514
- CMD
- LibreOffice Portable, Version: 7.1.0.3 (x86)

Für die Validierung der migrierten Dateien

- KOST-Val, Version 2.1.4.0
- veraPDF, Version 1.24.1

Für den Ingest nach DIMAG

- IngestTool 4.0

[1] Auch Advanced Renamer kann Dateien umbenennen. Diese müssten jedoch nach jedem Bearbeitungsschritt neu geladen werden, in Ant Renamer bleiben sie im Programm.
[2] Auch Ant Renamer kann Dateinamen kürzen, jedoch zeigt Advanced Renamer in der Vorschau Dateinamen rot an, wenn die Kürzung zu Konflikten führen würde.
[3] DIMAG-Wiki → Software für Anwender → Weitere Software, die im DIMAG-Verbund entstanden ist (DIMAG-Wiki).
[4] Download als Teil der KOST-Tools von GitHub.

3 Arbeitsordner

Für die einzelnen Bearbeitungs- und Migrationsschritte ist es zu empfehlen, lange Dateinamen (technisch: Dateipfad + Name der Datei + Dateiendung!) zu vermeiden. Deshalb sollte der Arbeitsordner „A243“ direkt auf H: liegen.
Im DIMAG-Ingest-Arbeitsordner werden die Pressemitteilung und dazugehörigen Dateien im Ordner „A-243 dauerhaft (R 0)“ pro Monat in einem Unterordner gesammelten. Der Ordner „A-243 dauerhaft (R 0)“ nach H:\A243 kopieren und in „0“ umbenennen.

4 Änderung der Dateinamen

Da die Bearbeitung der Dateien teilweise über CMD und Powershell erfolgt und um mögliche Inkompatibilitäten aufgrund der Dateinamen zu vermeiden sowie um die Online-Stellung der Dateien zu ermöglichen, werden die Dateinamen der Dateien im Ordner „0“ bearbeitet. Zuvor werden die ursprünglichen Dateinamen aber in einer Tabelle erfasst. Folgende Bearbeitung werden anschließend mit Ant Renamer vorgenommen:

- Bindestrich (-) ersetzen durch zwei Bindestriche (--), zur Unterscheidung zwischen dem ursprünglichen Bindestrich und dem Bindestrich für das Komma
- langer Bindestrich (–) ersetzen durch zwei Bindestrichen (--)
- Komma (,) ersetzen durch Bindestrich (-)
- Leerzeichen ( ) ersetzen durch Tiefstrich (_)
- Ä ersetzen durch Ae
- Ö ersetzen durch Oe
- Ü ersetzen durch Ue
- ä ersetzen durch ae
- ö ersetzen durch oe
- ü ersetzen durch ue
- ß ersetzen durch ss
- öffnende eckige Klammer [ersetzen durch öffnende runde Klammer (
- schließende eckige Klammer] ersetzen durch schließende runde Klammer)
- kaufmännisches Und (&) ersetzen durch u
- Punkt (.) löschen
- Ausrufezeichen (!) löschen
- Anführungszeichen unten („) löschen
- Anführungszeichen oben (“) löschen
- Anführungszeichen allgemein (") löschen
- Doppelkreuz (#) löschen
- Apostroph (') löschen

Anschließend werden mit Advanced Renamer die Dateinamen nach 55 Zeichen abgeschnitten (Entfernen > Anzahl: 300, ab Position: 56). Hierbei werden in der Vorschau Dateinamen rot angezeigt, wenn die Kürzung zu Konflikten führen würde. Ggf. müssen noch händische Namenskürzungen vorgenommen werden.

- in A 243 betrifft dies Dateien in der Nr. 309, 317, 324 und 338

Anschließend werden die geänderten Dateinamen in obiger Tabelle ergänzt. Für jede(s) Verzeichnungseinheit/AIP wird daraus eine Dateiliste im CSV-Format erstellt. Diese Dateilisten werden nach Wiederherstellung der Ordnerstruktur (s. Kapitel 9) in jedem Ordner pro Verzeichnungseinheit/AIP ergänzt. Darin sind die ursprünglichen und die geänderten Dateinamen aufgeführt. Der Dateiname lautet z.B. „_Dateiliste_Nr-332.csv“.

5 Analyse der Dateien

5.1 Dateiformate

Es sind 5.541 Dateien vorhanden. Diese Dateien wurden mittels TreeSize und anschließend mit IngestList analysiert und über das PRONOM file format ID (z.B. fmt/39) identifiziert.

Die Originaldateien, die nach PDF/A migriert wurden, umfassen folgende Dateiformate:

- DOC-Dateien
  - 2.214 x Microsoft Word Document V97-2003 (fmt/40)
  - 781 x OLE2 Compound Document Format (fmt/111)
  - 615 x Microsoft Word Document V6.0/95 (fmt/39)
  - 3 x Wordperfect Secondary File V5.0 (x-fmt/42)[5]
  - 4 x Microsoft Word for Windows Document V2.0 (fmt/38)
  - 1 x Rich Text Format V1.0-1.4 (fmt/45)
  - 1 x Rich Text Format V1.8 (fmt/53)
- DOCX-Dateien
  - 969 x Microsoft Office Open XML V2007 onwards (fmt/189)
- PDF-Dateien
  - 215 x Acrobat PDF 1.7 - Portable Document Format V1.7 (fmt/276)
  - 89 x Acrobat PDF 1.5 - Portable Document Format V1.5 (fmt/19)
  - 78 x Acrobat PDF 1.6 - Portable Document Format V1.6 (fmt/20)
  - 25 x Acrobat PDF 1.4 - Portable Document Format V1.4 (fmt/18)
  - 5 x Acrobat PDF 1.3 - Portable Document Format V1.3 (fmt/17)
  - 3 x Acrobat PDF/X - Portable Document Format - Exchange PDF/X-4 (fmt/488)
  - 1 x Acrobat PDF/X - Portable Document Format - Exchange 3:2002 (fmt/158)
- DOT-Dateien
  - 4 x Microsoft Word 97-2003-Vorlagen (fmt/40)

Die Originaldateien, die nicht migriert wurden (weil sie bereits Archivformate sind oder eine Migration in ein Archivformat nicht sinnvoll ist), umfassen folgende Dateiformate:

- JPG-Dateien
  - 76 x JPEG File Interchange Format V1.01 (fmt/43)
  - 73 x Raw JPEG Stream (fmt/41)
  - 12 x JPEG File Interchange Format V1.02 (fmt/44)
  - 12 x Exchangeable Image File Format (Compressed) V2.2.1 (fmt/645)
  - 3 x Exchangeable Image File Format (Compressed) V2.2 (x-fmt/391)
  - 1 x Exchangeable Image File Format (Compressed) V2.0 (x-fmt/398)

[5] Dieses Format wird gleichzeitig auch identifiziert als: Wordperfect Secondary File V5.1/5.2, Stationery for Mac OS X, Microsoft Word for MS-DOS Document V3.0, Interleaf Document, Microsoft Word (Generic) V6.0-2003, Microsoft Word Document (Password Protected) V97-2003, Compound WordPerfect for Windows Document V6 onwards bzw. folgenden PRONOM file format IDs zugeordnet: x-fmt/43, x-fmt/131, x-fmt/273, x-fmt/329, fmt/609, fmt/754, fmt/892.

- PNG-Dateien
    - 6 x Portable Network Graphics V1.0 (fmt/11)
    - 3 x Portable Network Graphics V1.2 (fmt/13)
- HTM-Dateien
    - 1 x Hypertext Markup Language V4.0 (fmt/99)
    - 1 x Vector Markup Language (fmt/583)
- Sonstige Dateien
    - 1 x PDF-Datei, Acrobat PDF/A - Portable Document Format V2b (fmt/477)
    - 1 x TXT-Datei, Microsoft Word for MS-DOS Document V4.0 (x-fmt/274)
    - 1 x BMP-Datei, Windows Bitmap V3.0 (fmt/116)

Die Migration muss in mehreren Schritten mit verschiedenen Programmen durchgeführt werden. Hierzu müssen die Dateien in verschiedene Ordner aufgeteilt werden. Genaueres siehe Kapitel 6.2.

5.2 Migrationstests

5.2.1 Ordnerstruktur

Für jede Verzeichnungseinheit/AIP gibt es einen Ordner, in dem die verschiedenen Dateien liegen. Es gibt keine Unterordner, die Struktur ist also einstufig. Es gibt 343 Verzeichnungseinheiten und damit 343 Ordner.
Das Problem hierbei ist, dass weder Adobe Acrobat noch die Stapelverarbeitung mit LibreOffice eine Migration von Dateien aus einer Struktur in eine gleiche Struktur ermöglichen. Die Migration ist nur aus einem Quellordner ohne Unterordner in ein Zielordner ohne Unterordner möglich.
Deshalb muss die Struktur aufgelöst, die Migrationen durchgeführt und anschließend die Struktur wiederhergestellt werden. Genaueres siehe Kapitel 6 bis 9.

5.2.2 DOT-Dateien (Microsoft Word 97-2003-Vorlagen)

DOT-Dateien werden im Adobe Acrobat im Auswahl-Fenster für die Dateien gar nicht angezeigt. Bei einer Migration mit LibreOffice, nach PDF/A oder DOCX, wird die Kopfzeile auf Seite 1, alles andere auf die nächste Seite verschoben. Da in A 243 jedoch nur vier DOT-Dateien vorhanden sind, werden diese händisch als DOCX-Dateien gespeichert, so dass diese Dateien zusammen mit den anderen DOCX-Dateien migriert werden können.

5.2.3 Unterschiedliche DOC-Dateien

Bei der testweisen durchgeführten Migration von DOC-Dateien konnten einige nicht von Adobe Acrobat geöffnet werden. Deshalb wurden die Dateien genauer analysiert.
Es gibt 3.619 DOC-Dateien. Diese werden von TreeSize als „Microsoft Word Document V97-2003 Dokumente“ identifiziert. Nach diesen Dateien wurde mit TreeSize gesucht, nach „Letzte Änderung“ sortiert und dann getrennt bis 1998 bzw. ab 1999 kopiert und in entsprechend benannte Ordner abgelegt. Diese Ordner wurden mit IngestList analysiert:

Ordner „DOC bis 1998“ (623 Dateien):

(1) 615 Dateien fmt/39 = Microsoft Word Document V6.0/95
(2) 4 Dateien fmt/38 = Microsoft Word for Windows Document V2.0
(3) 3 Dateien x-fmt/42 (u.a.) = Wordperfect Secondary File V5.0 (u.a.)
(4) 1 Datei fmt/45 = Rich Text Format V1.0-1.4

Ordner „DOC ab 1999“ (2.996 Dateien):

(5) 2.214 Dateien fmt/40 = Microsoft Word Document V97-2003
(6) 781 Dateien fmt/111 = OLE2 Compound Document Format
(7) 1 Datei fmt/53 = Rich Text Format V1.8

Testergebnisse

- Adobe Acrobat
  - (1) und (2) können nicht geöffnet werden
  - (3) bis (7) können problemlos geöffnet und nach PDF/A migriert werden
- LibreOffice
  - bei der Migration nach PDF/A von (1) und (3) wird die Schriftart bzw. das Layout geändert
  - bei der Migration von (1) nach DOCX wird das Layout nicht geändert
  - die Migration nach PDF/A von (2) scheint in Ordnung zu sein; mögliche Änderungen können nicht festgestellt werden, da dieses Dateiformat nicht geöffnet werden kann, da der Zugriff in den Zugriffsschutzeinstellungen im Windows Trust Center gesperrt ist
  - (4) kann problemlos geöffnet und nach PDF/A migriert werden
  - bei der Migration von (5) wird die Kopfzeile auf Seite 1, alles andere auf die nächste Seite verschoben

Genauere Untersuchung der Migration von Dateien vom Typ (1) (fmt/39)

1. bei der Migration direkt nach PDF/A mit LibreOffice wird das Layout geändert und das Textfeld „Pressemitteilungen“ geschwärzt, außerdem gehen teilweise Textteile verloren
2. wenn als Zwischenschritt die DOC-Dateien mit LibreOffice nach DOCX migriert werden, dann bekommen in den DOCX-Dateien die Buchstaben im Textfeld „Pressemitteilungen“ Schatten, außerdem wird das Textfeld, das u.a. das Datum enthält, teilweise zu klein angezeigt (oder die Schrift zu groß), so dass das Datum nicht immer zu sehen ist

a. in den aus den DOCX-Dateien mit LibreOffice migrierten PDF/A-Dateien stimmt das Layout ebenfalls nicht mehr, über dem Textfeld „Pressemitteilungen“ liegt ein kleiner Kasten und das Datum, das in DOCX-Dateien nicht zu sehen ist, ist auch hier nicht zu sehen
b. wenn diese DOCX-Dateien mit Adobe Acrobat migriert werden, dann gibt es beim Layout keine großen Unterschiede, abgesehen davon, dass die Buchstaben im Textfeld „Pressemitteilungen“ einen Schatten haben und das Datum, das in einigen DOCX-Dateien nicht zu sehen ist, auch hier nicht zu sehen ist

Da das Datum nicht immer verloren geht, teils an anderer Stelle im Dokument nochmal steht, außerdem im Enthält-Feld in ACTApro vorhanden ist und das exakte Datum evtl. auch nicht so wichtig ist, wird Methode 2b angewendet.

Fazit

- Dateien vom Typ (1) (fmt/39) werden mit LibreOffice nach DOXC migriert. Anschließend werden diese mit Adobe Acrobat nach PDF/A migriert
- Dateien vom Typ (2) und (4) (fmt/38 und fmt/45) werden mit LibreOffice nach PDF/A migriert
- Dateien vom Typ (3), (5) bis (7) (x-fmt/42, fmt/40, fmt/111 und fmt/53) werden mit Adobe Acrobat nach PDF/A migriert

Die Dateien müssen anhand der von IngestList erstellten XML-Datei (vgl. Kapitel 5.1) herausgesucht und in entsprechende Ordner verschoben werden. Genaueres siehe Kapitel 6.2

6 Vorbereitung für die Migration

6.1 Ordnerstruktur auflösen

Es ist ratsam, während des gesamten Vorgehens bis zur Wiederherstellung der Struktur die Anzahl der Dateien mittels TreeSize und einer Excel-Liste zu überwachen.

1. den Ordner „0“ kopieren und in „1“ umbenennen
2. Thumbs.db mit dem FileReaper löschen
3. Ordner „1“ in Ant Renamer laden
4. Ordnerstruktur in die Dateinamen einfügen: über „Zeichen einfügen“ → „Einzufügende Zeichenfolge“ (ab Position 0 vom Anfang): %folder1%---
5. Ordner „2“ anlegen
6. im Ordner „1“ nach allen Dateien suchen (*.*) und nach Ordner „2“ kopieren

6.2 Dateien sortieren

1. mit TreeSize im Ordner „2“ nach DOC-Dateien suchen, alle Dateien ab 1999 in den Ordner „3-1“ kopieren
2. mit TreeSize im Ordner „2“ nach DOC-Dateien suchen, alle Dateien bis einschließlich 1998 in den Ordner „3-2“ kopieren
3. in der von IngestList erstellten XML-Datei nach fmt/38 und fmt/45 suchen, die entsprechenden Dateien anhand des Dateinamens mit TreeSize im Ordner „3-2“ suchen und in den Ordner „3-3“ verschieben
4. in der von IngestList erstellten XML-Datei nach x-fmt/42 suchen, die entsprechenden Dateien anhand des Dateinamens mit TreeSize im Ordner „3-2“ suchen und in den Ordner „3-1“ verschieben
5. mit TreeSize im Ordner „2“ nach PDF-Dateien suchen und in den Ordner „3-4“ kopieren
6. mit TreeSize im Ordner „2“ nach DOCX-Dateien suchen und in den Ordner „3-5“ kopieren
7. mit TreeSize im Ordner „2“ nach DOT-Dateien suchen und in den Ordner „3-6“ kopieren
8. mit TreeSize im Ordner „2“ nach TXT-, HTM-, PNG-, JPG-, BMP- und CSV-Dateien suchen und in den Ordner „3-7“ kopieren
9. anhand der von IngestList erstellten XML-Datei die bereits vorhandene PDF/A-Datei identifizieren und diese aus dem Ordner „3-4“ in den Ordner „3-7“ verschieben

7 Migration

7.1 Migrationsvorgehen

7.1.1 Mit Adobe Acrobat: DOC und DOCX nach PDF/A

1. Werkzeuge → PDF erstellen → Mehrere Dateien → Mehrere PDFs erstellen → Weiter
2. die Dateien können aus dem Windows-Explorer direkt ins nun geöffnete Fenster gezogen werden
3. OK
4. im nächsten Fenster wird der Zielordner (siehe Kapitel 0) ausgewählt – v.a. wenn (bei vielen Dateien) mehrere Migrationsläufe nötig sind, kann hier der Quellordner der Einfachheit halber als Zielordner ausgewählt und die erstellten Dateien erst anschließend in der eigentlichen Zielordner verschoben werden
5. Start mit OK

Ungefährer Zeitbedarf:

- zwischen 15 und 27 min / 100 DOC-Dateien, vermutlich abhängig von der Netzwerkauslastung
- 16 min / 100 DOCX-Dateien

7.1.2 Mit Adobe Acrobat: PDF nach PDF/A

1. Werkzeuge → Aktionsassistent → Dokumente archivieren → Dateien hinzufügen (rechte Seite)
2. Rechtsklick auf „Dokumentbeschreibung hinzufügen“ → „Diesen Schritt überspringen“
3. Einstellung für das Speichern: „Speichern“ → dadurch wird die vorhandene PDF-Datei überschrieben, bei „Speichern unter“ müsste für jede (!) Datei ein Speicherort ausgewählt werden
4. Klick auf „Anfang“
5. bei einer Unterbrechung (z.B. wegen zu kleinen Seitenrändern) muss diese Meldung mit OK bestätigt, die Migration mit „Anfang“ fortgesetzt werden

Ungefährer Zeitbedarf: 6 min / 100 Dateien

7.1.3 Mit LibreOffice: DOC nach DOCX

1. Bedingung: LibreOffice muss unter dem unten angegeben Pfad (H:\Dokumente\...) liegen
2. Kommandozeile öffnen: Windows-Taste → „CMD“ eingeben
3. ins Verzeichnis mit den zu bearbeitenden Dateien wechseln, Eingabe im CMD z.B. „cd H:\A243\3-2“ (Kopieren und Rechtsklick ins CMD)
4. ENTER
5. exakte Eingabe der folgenden Zeilen (dito), ggf. anpassen

```
For %f in (*.doc) do
"H:\Dokumente\Programme\LibreOfficePortable\App\libreoffice\program\soffice" --
headless --convert-to docx --outdir H:\A243\ZIELORDNER %f
```

6. ENTER

Ungefährer Zeitbedarf: 16 min / 100 Dateien

7.1.4 Mit LibreOffice: DOC nach PDF/A

7. Bedingung: LibreOffice muss unter dem unten angegeben Pfad (H:\Dokumente\...) liegen
8. Kommandozeile öffnen: Windows-Taste → „CMD“ eingeben
9. ins Verzeichnis mit den zu bearbeitenden Dateien wechseln, Eingabe im CMD z.B. „cd H:\A243\3-3“ (Kopieren und Rechtsklick ins CMD)
10. ENTER
11. exakte Eingabe der folgenden Zeilen (dito), ggf. anpassen

```
For %f in (*.doc) do
"H:\Dokumente\Programme\LibreOfficePortable\App\libreoffice\program\soffice" --
headless --convert-to pdf --outdir H:\A243\ZIELORDNER %f
```

12. ENTER

Ungefährer Zeitbedarf: 16 min / 100 Dateien

7.2 Migrationsschritte

1. Folgende Zielordner anlegen:
    - 3-1a
    - 3-2a
    - 3-2b
    - 3-3a
    - 3-5a
    - 3-6a
    - 3-6b
2. er Ordner „3-4“ wird kopiert, die Kopie „3-4a“ benannt
3. die Dateien im Ordner „3-1“ werden mit Adobe Acrobat nach PDF/A migriert (Kapitel 7.1.1), Zielordner „3-1a“
4. die Dateien im Ordner „3-2“ werden mit LibreOffice nach DOCX migriert (Kapitel 0), Zielordner „3-2b“, anschließend werden diese DOCX-Dateien mit Adobe Acrobat nach PDF/A migriert (Kapitel 7.1.1), Zielordner „3-2a“
5. die Dateien im Ordner „3-3“ werden mit LibreOffice nach PDF/A migriert (Kapitel 7.1.4), Zielordner „3-3a“
6. die Dateien im Ordner „3-4a“ werden mit Adobe Acrobat nach PDF/A migriert (Kapitel 7.1.2), ersetzende Migration
7. die Dateien im Ordner „3-5“ werden mit Adobe Acrobat nach PDF/A migriert (Kapitel 7.1.1), Zielordner „3-5a“
8. die Dateien im Ordner „3-6“ werden händisch nach DOCX migriert (Datei speichern unter), Zielordner „3-6b“, anschließend werden diese DOCX-Dateien mit Adobe Acrobat nach PDF/A migriert (Kapitel 7.1.1), Zielordner „3-6a“
9. die Dateien im Ordner „3-7“ werden nicht migriert

7.3 Validierung

Die migrierten Dateien werden zuerst mit KOST-Val validiert. Nicht valide Dateien werden mit veraPDF gegengeprüft. Bei nicht validen Dateien muss entschieden und dokumentiert werden, wie damit umgegangen wird (vgl. unten).
Die Validierung mit KOST-Val benötigt ca. 30 sek / 100 Dateien.

7.4 Archivformate

Folgende Archivformate wurden bei der Migration erstellt (Analyse mit DROID):

Mit Adobe Acrobat erstellt

- Acrobat PDF/A - Portable Document Format V1b (fmt/354)
- Acrobat PDF/A - Portable Document Format V2b (fmt/477)
- hierbei ist zu erwähnen, dass zwei Dateien von Typ Acrobat PDF/X - Portable Document Format - Exchange PDF/X-4 (fmt/488) auch nach der Migration von diesem Typ sind, aber gleichzeitig je einmal
    - Acrobat PDF/A - Portable Document Format V1b (fmt/354) (Nr-281---171018_Presseeinladung_Festakte_Baden_und_Wuerttemberg_)
    - Acrobat PDF/A - Portable Document Format V2b (fmt/477) (Nr-292---180906_Handreichung_Religioese_Feiern_im_multireligioes)
- Tagged Image File Format (fmt/353)

Mit LibreOffice erstellt

- Acrobat PDF/A - Portable Document Format V1a (fmt/95)

Die unterschiedlichen PDF/A-Formate sind durch die Grundeinstellungen der Programme bedingt. Da keine PDF/A-Version eine andere ablöst, gibt es auch keinen Grund, die Einstellungen zu ändern.

7.5 Migrationsbericht

- Migration der Dateien im Ordner „3-1“ (DOC nach PDF/A)
  - es waren 22 Migrationsläufe nötig
  - bei 13 davon ist Adobe Acrobat abgestürzt
  - drei Dateien konnten zuerst nicht migriert werden, erst bei zweiten oder dritten Migrationslauf hat es funktioniert
  - alle 2.999 PDF/A-Dateien sind valide
- Migration der Dateien im Ordner „3-2“ (DOC über DOCX nach PDF/A)
  - Schritt 1: DOC nach DOCX
    - die Datei „Nr-25---TESU“ wurde nicht migriert, da sie defekt ist
    - sie konnte weder von Word noch von LibreOffice geöffnet werden
    - sie konnte aber mit Notepad++ geöffnet werden: der offensichtliche Text wurde kopiert und als neue DOCX-Datei gleichen Namens gespeichert und mit den anderen DOCX nach PDF/A migriert
  - Schritt 2: DOCX nach PDF/A
    - es waren mehrere Migrationsläufe nötig
    - bei 2 davon ist Adobe Acrobat abgestürzt
    - alle 615 PDF/A-Dateien sind valide
- Migration der Dateien im Ordner „3-3“ (DOC nach PDF/A)
  - keine Probleme, alle 5 PDF/A-Dateien sind valide
- Migration der Dateien im Ordner „3-4a“ (PDF nach PDF/A, ersetzenden Migration)
  - 411 PDF/A-Dateien sind valide, 6 sind nicht valide:

Datei	KOST-Val	veraPDF
Nr-269---161022_Flyer_OekPilgertag_Ludwbrg	invalide	invalide
Nr-273---170209_Flyer_didacta_Sonderschau_Kirche	invalide	valide
Nr-274---170302_Programmflyer_FESTIVAL_digital	invalide	invalide
Nr-279---170803_2017_Programmflyer_FESTIVAL_digital	invalide	invalide
Nr-292---180906_Handreichung_Religioese_Feiern_im_multireligioes	invalide	valide
Nr-314---20200716_Corona--EinKurzberichtzurEvangelischenBildung_	invalide	invalide

  - diese 6 Dateien können jedoch problemlos geöffnet und angezeigt werden, sie werden als Repräsentation 1 archiviert
  - auch über eine händische Migration (Datei → Speichern als) konnten keine validen Dateien erstellt werden
  - zur Sicherheit wurden aus den 6 ursprünglichen Dateien mit dem Adobe Acrobat die Seiten als TIF-Dateien[6] exportiert und im Ordner „3-4_TIF“ gespeichert, diese werden als Repräsentation 2 archiviert
- Migration der Dateien im Ordner „3-5“ (DOCX nach PDF/A)
  - es waren 7 Migrationsläufe nötig
  - bei 4 davon ist Adobe Acrobat abgestürzt
  - alle 416 PDF/A-Dateien sind valide
- Migration der Dateien im Ordner „3-6“ (DOT über DOCX nach PDF/A)
  - keine Probleme, alle 4 PDF/A-Dateien sind valide

[6] Bearbeiten → Einstellungen → Aus PDF konvertieren → TIFF: 118,11 Pixel/cm = 300 dpi, kein Farbprofil.

8 Erkenntnisse, Empfehlungen

- aufgrund von Fehlermeldungen, Warnungen (z.B. wegen zu kleinen Seitenrändern) oder Programmabstürzen während der Migration mit Adobe Acrobat kann die Migration nicht über Nacht durchgeführt werden, man muss stattdessen am PC bleiben, um eingreifen zu können
- deshalb und auch grundsätzlich empfiehlt es sich, die Anzahl der pro Migrationslauf zu verarbeitenden Dateien auf maximal 180 Dateien zu begrenzen
- da bei der Migration fehlerhafte Dateien nicht verarbeitet und somit (erst) erkannt werden, ist es zu empfehlen, die Migration durchzuführen, dann ggf. fehlerhafte Dateien auch in der Repräsentation 1 zu löschen und erst dann ins DIMAG bzw. digitale Magazin zu laden
- da aktuell (November 2024) im DIMAG mit dem IngestTool keine weitere Repräsentation an ein bestehendes AIP angebunden werden kann, sondern die verschiedenen Repräsentationen nur zusammen in einem Importvorgang importiert werden können, ist es ebenfalls zu empfehlen, zuerst die Repräsentationen zu erstellen, dann zu importieren

9 Ordnerstruktur wiederherstellen

9.1 Für Repräsentation 1

1. die Dateien aus dem Ordner „3-4a“ (PDF/A aus PDF) werden in den Ordner „2“ kopiert und ersetzen die dort vorhandenen Dateien
2. Windows-Taste, dann „Powershell“ eingeben
3. Ordnerstruktur wiederherstellen: In Powershell folgenden Code eingeben (Kopieren und Einfügen), der Vorgang sollte von selbst starten (ggf. mit ENTER starten):

```
$Files = Get-Childitem -File -Path "H:\A243\2\*" -include "*---*"
Foreach ($File in $Files) {
 $NewFile = [io.fileinfo]($File.Fullname -Replace '---','\')
 [void](md -force $NewFile.DirectoryName)
 Move-Item $File.Fullname $NewFile.Fullname
}
```

4. der Vorgang nimmt einige Zeit in Anspruch
5. der Ordner „2“ wird in „R-1“ umbenannt und nach „H:\Dokumente\DIMAG-Ingest-Arbeitsordner\A-243\input“ verschoben

9.2 Für die Repräsentation 2

1. die Dateien aus folgenden Ordnern werden in einen neuen Ordner „4“ kopiert:
   - 3-1a
   - 3-2a
   - 3-3a
   - 3-4_TIF
   - 3-5a
   - 3-6a
2. Windows-Taste, dann „Powershell“ eingeben
3. Ordnerstruktur wiederherstellen: In Powershell folgenden Code eingeben (Kopieren und Einfügen), der Vorgang sollte von selbst starten (ggf. mit ENTER starten):

```
$Files = Get-Childitem -File -Path "H:\A243\4\*" -include "*---*"
Foreach ($File in $Files) {
 $NewFile = [io.fileinfo]($File.Fullname -Replace '---','\')
 [void](md -force $NewFile.DirectoryName)
 Move-Item $File.Fullname $NewFile.Fullname
}
```

6. der Vorgang nimmt einige Zeit in Anspruch
7. der Ordner „4“ wird in „R-2“ umbenannt und nach „H:\Dokumente\DIMAG-Ingest-Arbeitsordner\A-243\input“ verschoben

9.3 Für Repräsentation für die Online-Stellung

1. die Dateien aus folgenden Ordnern werden in einen neuen Ordner „5“ kopiert:
   - 3-4a
   - 3-7
2. Windows-Taste, dann „Powershell“ eingeben
3. Ordnerstruktur wiederherstellen: In Powershell folgenden Code eingeben (Kopieren und Einfügen), der Vorgang sollte von selbst starten (ggf. mit ENTER starten):

```
$Files = Get-Childitem -File -Path "H:\A243\5\*" -include "*---*"
Foreach ($File in $Files) {
  $NewFile = [io.fileinfo]($File.Fullname -Replace '---','\')
  [void](md -force $NewFile.DirectoryName)
  Move-Item $File.Fullname $NewFile.Fullname
}
```

8. der Vorgang nimmt einige Zeit in Anspruch
9. anschließend wird die Ordnerstruktur (mit Dateien) aus dem Ordner „R-2“ in den Ordner „5“ kopiert, dabei werden gleichnamige Ordner vereint
10. aus den Ordnern „Nr-269“, „Nr-273“, „Nr-274“, „Nr-279“, „Nr-292“ und „Nr-314“ werden die TIF-Dateien (vgl. Kapitel 7.5) gelöscht
11. der Ordner „5“ wird in „A-243“ unbenannt und nach „XXXXXXXXXX\A-Bestaende“ verschoben
12. der Ordner „5“ wird in „A-243“ umbenannt und nach XXXXXXXXXX\A-Bestaende“ verschoben

10 Ingest nach DIMAG

10.1 Projektordner

Der Projektordner „A-243“ hat nebenstehende Struktur. Die Ordner „dimag“ und „dit“ werden auch bei der Anlage des Projekts „A-243“ vom IngestTool angelegt.
In den Ordner „dit“ werden von IngestTool IPAK-, IPRF- und IPRO-Dateien angelegt.
Im Ordner „input“ liegen die Ordner für die Repräsentationen, jeweils darunter die Ordner für die einzelnen AIP. Hierhin müssen die Ordner „R-1“ und „R-2“ (vgl. Kapitel 9.1 und 9.2) verschoben werden.
Ist zu einem AIP keine zweite Repräsentation vorhanden, wird auch kein Ordner benötigt.
Unterhalb des Ordners „A-243“ liegt die Datei „Metadaten.csv“ (also auf derselben Ebene wie die Ordner „dimag“ und „input“).

10.2 Metadaten-Datei

Vgl. hierzu und auch für folgenden Kapitel die entsprechenden Dateien in Doxis.

Die Datei „Metadaten.csv“ muss folgenden Metadaten (in dieser Reihenfolge) enthalten:

- Signaturanteil (Sig. 1 in ACTApro)
- Entstehungszeitraum (Laufzeit)
- Objekt-Titel (Sachakten-Titel)
- Titel R 2
- AIP-Ordner

Als „Titel R 2“ kommen abhängig vom Inhalt folgende in Frage:

- PDF/A nicht langzeitstabiler Dateiformate (Delta zu R 1)
- TIF invalider PDF/A (Delta zu R 1)
- PDF/A nicht langzeitstabiler Dateiformate, TIF invalider PDF/A (Delta zu R 1)
- …

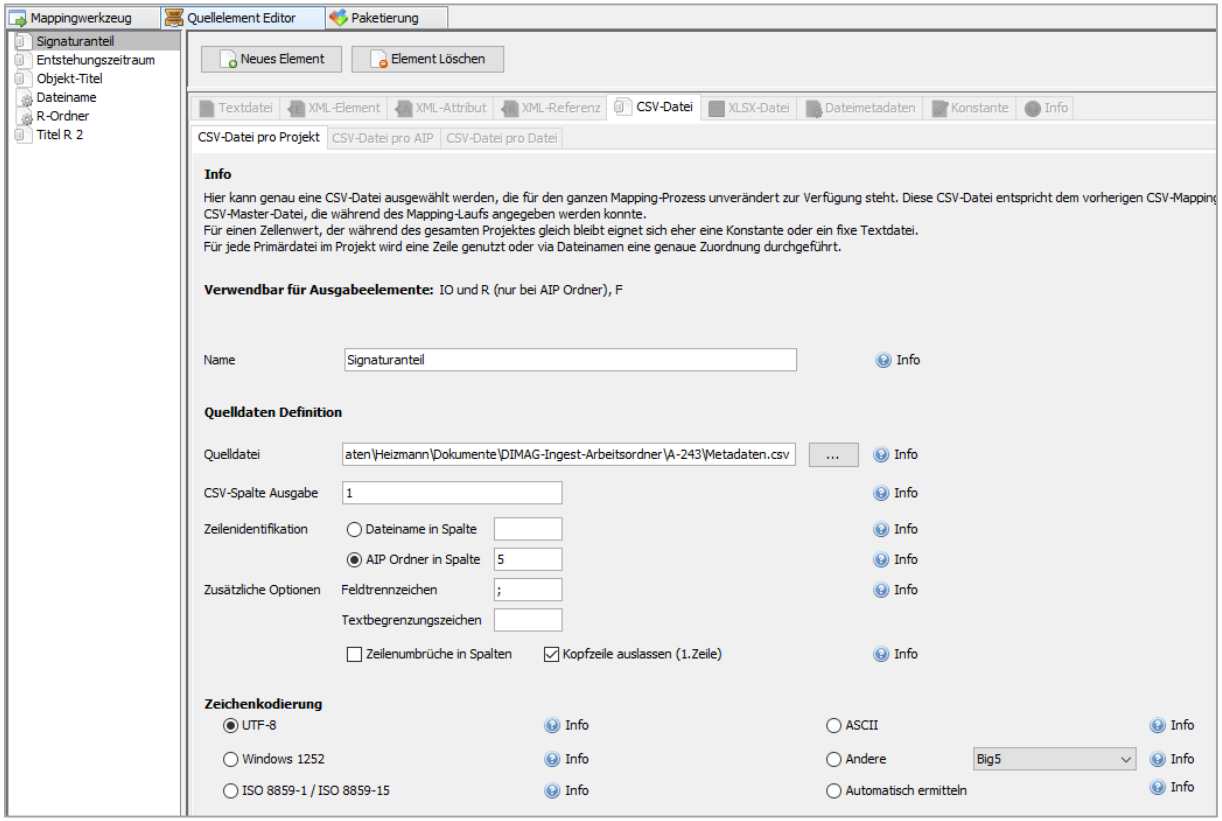
Hierzu muss überprüft werden:

- Von welchen Signaturen sind AIP für die Repräsentation 2 vorhanden?
- Welche davon enthalten etwas anderes als PDF/A (vgl. Kapitel 7.4)

10.3 IngestTool: Paketierung

- Ordnerstruktur berücksichtigen
- Ebene: 1 = Ordner unterhalb des Ordners für den Bestand
- Unterordner einbeziehen: nein
- Muster: Thumbs.db, Ausschließen (für den Fall, dass doch nicht alle Thumbs.db gelöscht wurden)

10.4 IngestTool: Quellelement Editor



Hier werden folgende Elemente angelegt:

1. Signaturanteil (CSV-Spalte Ausgabe: 1)
2. Entstehungszeitraum (CSV-Spalte Ausgabe: 2)
3. Objekt-Titel (CSV-Spalte Ausgabe: 3)
4. Dateiname
5. R-Ordner
6. Titel R 2 (CSV-Spalte Ausgabe: 4)

Für 1 – 3 und 6 gilt:

- Neues Element: CSV-Datei pro Projekt
- Quelldatei: Metadaten.csv im Projektordner (siehe Kapitel 10.1)
- Zeilenidentifikation: AIP Ordner in Spalte, 5
- Feldtrennzeichen: ;
- Textbegrenzungszeichen: {leer}
- Kopfzeile auslassen (1. Zeile): {auswählen}
- Zeichenkodierung: UTF-8

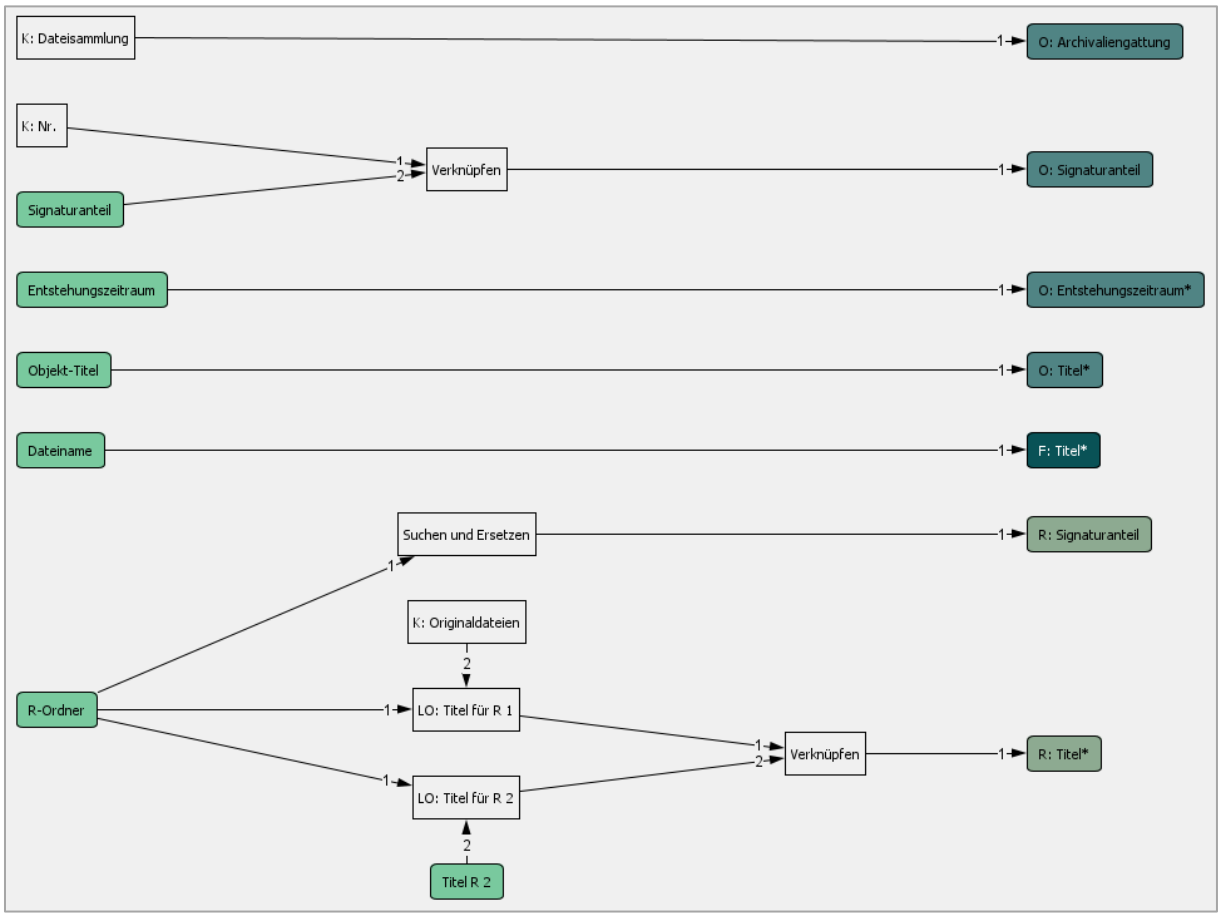
Dateiname:

- Neues Element: Dateimetadaten pro Datei
- Metadaten: Dateiname inkl. Endung

R-Ordner:

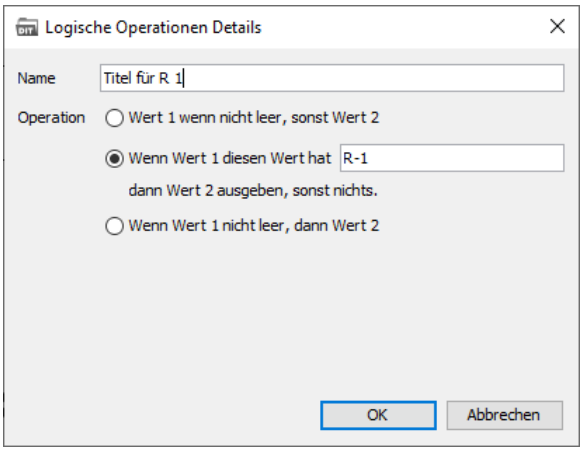
- Neues Element: Dateimetadaten pro AIP
- Metadaten: Ordnernamen auf Ebene, 2

10.5 IngestTool: Mappingwerkzeug

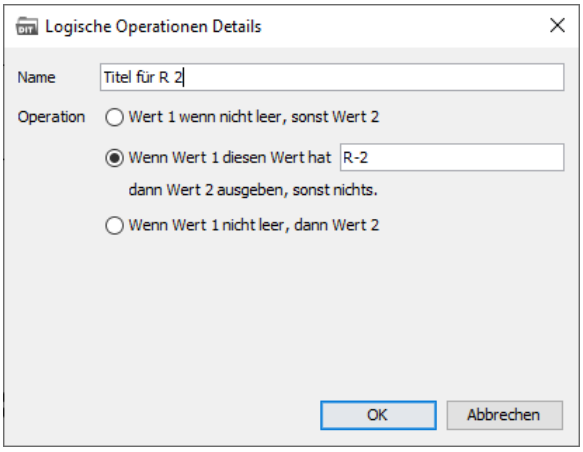


„Suchen und Ersetzen“: - durch Leerzeichen ersetzen („R-1“ → „R 1“, „R-2“ → „R 2“)

„LO: Titel für R 1“



„LO: Titel für R 2“



Die hellgrünen Felder sind die Metadaten, die aus der Metadaten.csv, den Dateinamen oder den Ordnernamen entnommen sind (siehe vorheriges Kapitel).

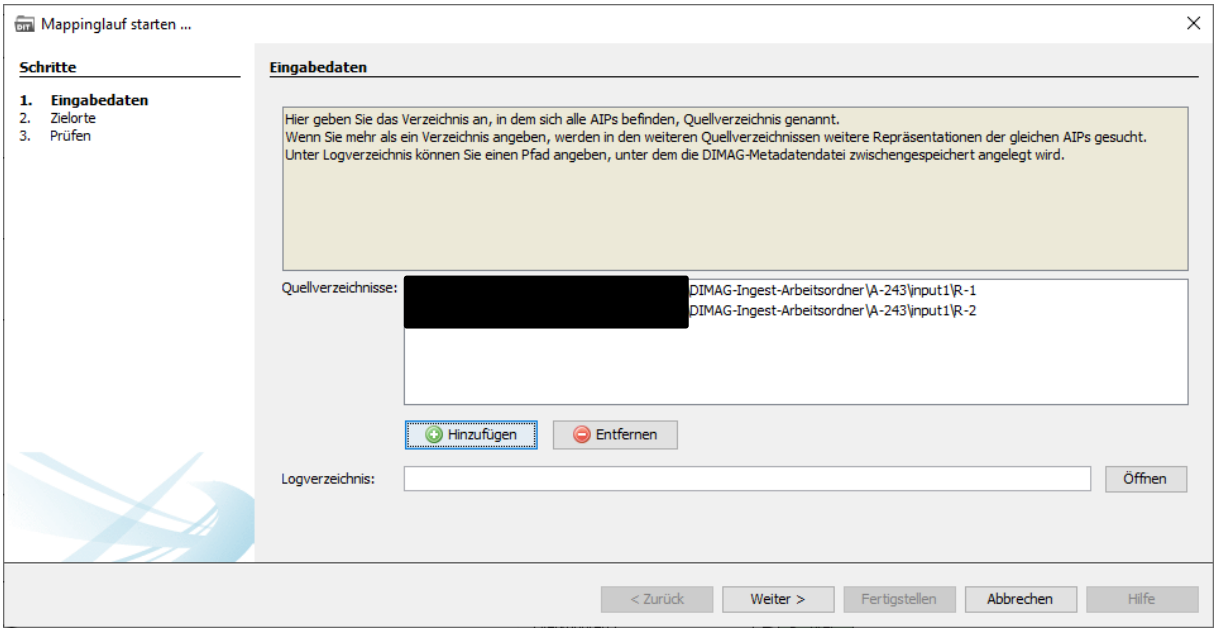
10.6 IngestTool: Mappinglauf starten (Ingest nach DIMAG)

Bei großen Dateimengen sollte der Ingest aufgeteilt werden erfolgen (Dateianzahl < 10.000). Deshalb wurden die AIP von A 243 aufgeteilt (jeweils für die Ordner „R-1“ und „R-2“):

1. input1: Nr-1 bis Nr-172
2. input2: Nr-173 bis Nr-343

Die Ordner für die zweite Lieferung wurde in einen Ordner „input2“ mit den Unterordnern „R-1“ und „R-2“ verschoben.
Die Metadaten-Datei muss nicht aufgeteilt werden. Sie kann auch Metadaten zu AIP, die im Quellverzeichnis nicht vorhanden sind, enthalten. Sie muss aber zu allen AIP im Quellverzeichnis vollständige Informationen haben.

Damit gleichzeitig die verschiedenen Repräsentationen angelegt werden, müssen die Ordner „R-1“ und „R-2“ einzeln ausgewählt werden (in dieser Reihenfolge):



Über die DIMAG-Weboberfläche wird der Bestand angelegt werden (Bestandssignatur ohne Präfix). Nach Anlage des Bestands wird die „Archival ID (AID)“ notiert. Diese ist in der DIMAG-Weboberfläche unter „Details“ des jeweiligen Bestandes zu finden.

Die Archival ID wird in IngestTool in das Feld „Anhänge ID“ eingegeben. Nach Auswahl des Verbindungsprofils und des Ziels „DIMAG“ sowie Eingabe der Passwörter wird der Ingest über „Fertigstellen“ gestartet.
Der Ingest nimmt einige Zeit in Anspruch. Nach Meldung, dass der Ingest nicht mehr länger verfolgt wird, können noch 1,5 h oder mehr (!) vergehen, bis die AIP im DIMAG zu sehen sind.

11 Online-Stellung

11.1 Dateinamen ergänzen

Vergleichbar mit der üblichen Dateibenennung für Digitalisate und andere Dateien werden die Dateinamen noch angepasst, z.B.

„PMLAKIMU.pdf" im Ordner „Nr-4" → „LKAS_A-243_Nr-4_PMLAKIMU.pdf"

11.2 Anbinden der Dateien an die Verzeichnungseinheiten in ACTApro Desk

Die Dateien in „XXXXXXXXXX\A-Bestaende\A-243" werden mit ACTApro Batch an die entsprechenden Verzeichnungseinheiten in ACTApro Desk angebunden.
Ob die Dateien zukünftiger Pressemitteilungen per ACTApro Batch oder in ACTApro Desk direkt angebunden werden, muss dann entschieden werden. Vgl. hierzu auch die Anleitung „ACTApro Batch. Anbinden von Dateien (Digitalisate u.a.) an ACTApro Desk", Kapitel 8 bzw. 7 (Version vom 09.11.2023).
Hierbei ist zu beachten: In ACTApro Benutzung muss der Dateiname angezeigt werden. Dieser muss hierfür in ACTApro Desk im Reiter „Digitale Unterlage" → Datei-Metadaten → „Titel/Bezeichnung" stehen. Beim Import über ACTApro Desk wird wegen eines Bugs der Dateiname nicht automatisch eingetragen, beim Import über ACTApro Batch wird er eingetragen.

11.3 Export aus ACTApro Desk

Für den Export – mit Vorschaubildern – wird auf die Anleitung „ACTApro Desk. Export von Beständen für ACTApro Benutzung und Archivportal-D", Kapitel 2.3 (Version vom 13.10.2023) verwiesen.
Ein (grobe) Überprüfung der Vorschaubilder ergab, dass kein falsches Vorschaubild für PDF-Dateien vorhanden ist. Bei zukünftigen Exporten muss dies ebenfalls geprüft und falsche Vorschaubilder ggf. ersetzt werden.