

Preservation Planning, Preservation Actions

1. Definition der Archivformate

Die Definition der Archivformate geschieht in der Konfigurationsdatei „migration-config.xml“ von docuteam feeder. In dieser umfangreichen Datei wird für jedes Ausgangsformat ein Archivformat hinterlegt. Diese Konfiguration kommt zur Anwendung unabhängig davon, ob es sich um einen initialen Ingest oder um die Migration von Dateien handelt, die bereits im Archiv liegen. Die folgende Abbildung zeigt einen Ausschnitt aus dieser Konfigurationsdatei. Für jedes Quellformat wird sowohl ein Zielformat definiert, wie auch die für die Migration verantwortliche Applikation hinterlegt.

```

<!-- WORD PROCESSING -->

<!-- Word for Windows -->
<puid name="fmt/37"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/38"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/39"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/40"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/412"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/523"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/597"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="fmt/599"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />

<!-- Word for MS-DOS -->
<puid name="x-fmt/273"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="x-fmt/274"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="x-fmt/275"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="x-fmt/276"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />

<!-- Word for Mac -->
<puid name="x-fmt/2"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />
<puid name="x-fmt/129"> applicationID="6" targetExtension="pdf" targetPronom="fmt/354" />

<!-- Wordperfect -->
<puid name="x-fmt/42"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="x-fmt/43"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="x-fmt/44"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="x-fmt/203"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="x-fmt/393"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="x-fmt/394"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="x-fmt/609"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />

<!-- Rich Text -->
<puid name="fmt/45"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
<puid name="fmt/46"> applicationID="1" targetExtension="pdf" targetPronom="fmt/95" />
```

2. Grundlagen für das Preservation Planning ermitteln

Die Grundlage für das Preservation Planning im Cockpit von docuteam feeder über den Bericht „Dateiformate“ ermittelt. Der Bericht listet – geordnet nach Häufigkeit – alle im Archiv vorhandenen Dateiformate auf und zeigt auch ihren Pronom-Identifier (PUID) an. **Pronom** ist die Fileregistry des britischen Nationalarchivs und erlaubt es, Dateiformate und Unterverisionen davon eindeutig zu identifizieren und zu referenzieren. Aufgrund dieses Berichts wird der Entscheid gefällt, ob und welche Dateiformate in neue Zielformate migriert werden sollen.

The screenshot shows a web browser window with the URL localhost:3000/repositories/demo/file_formats. The page title is "Demo: Dateiformate". The top navigation bar includes links for "Ausführungen", "Workflows", "Schrifte", "Benutzer", and "Cockpit". Below the title, there are two tabs: "Total" and "test", with "test" being the active tab. A table displays the following data:

Name	PUID	#	%
Microsoft Word Document	fmt/40	884	49,72%
Acrobat PDF 1.3 - Portable Document Format	fmt/17	147	8,27%
Acrobat PDF 1.4 - Portable Document Format	fmt/18	147	8,27%
JPEG File Interchange Format	fmt/43	82	4,61%
Graphics Interchange Format	fmt/4	66	3,71%
Microsoft Excel 97 Workbook (xls)	fmt/61	62	3,49%
Tagged Image File Format	fmt/353	54	3,04%
Acrobat PDF 1.2 - Portable Document Format	fmt/16	51	2,87%
Rich Text Format	fmt/50	27	1,52%
HyperText Markup Language	fmt/96	27	1,52%
Acrobat PDF 1.5 - Portable Document Format	fmt/19	25	1,41%
Graphics Interchange Format	fmt/3	23	1,29%
Microsoft Word Document	fmt/39	22	1,24%
Plain Text File	x-fmt/111	16	0,90%
Acrobat PDF/A - Portable Document Format	fmt/481	13	0,73%
JPEG File Interchange Format	fmt/44	11	0,62%
Microsoft Excel 5.0/95 Workbook (xls)	fmt/59	10	0,56%
HyperText Markup Language	fmt/100	9	0,51%
Thumbs DB file	fmt/682	9	0,51%

3. Extraktion der zu migrierenden Dateien

Preservation Actions werden immer pro Dateiformat durchgeführt. Beispiel: Im Archiv liegt verschiedene JPEG-Dateien, die im Rahmen einer Preservation Action nach Tiff migriert werden sollen. In migration.config.xml ist Tiff als Archivformat von JPEG hinterlegt.

JPEG (Pronom fmt/43) -> Tiff (Pronom fmt/353)

In docuteam feeder werden mit Hilfe des Workflows „Preservation: Deliver DIP by PUID“ alle JPEG-Dateien extrahiert. Der Workflow wird gestartet und als Parameter wird im Feld „Manuelle Eingabe“ der PUID des zu migrierenden Dateiformats angegeben. **Wichtig:** der PUID muss in Anführungszeichen gesetzt werden, also „**fmt/43**“ und nicht nur **fmt/43**.

docuteam feeder

Ausführungen Workflows Schritte Benutzer Cockpit

Neue Ausführung für Preservation:
Deliver DIP by PUID

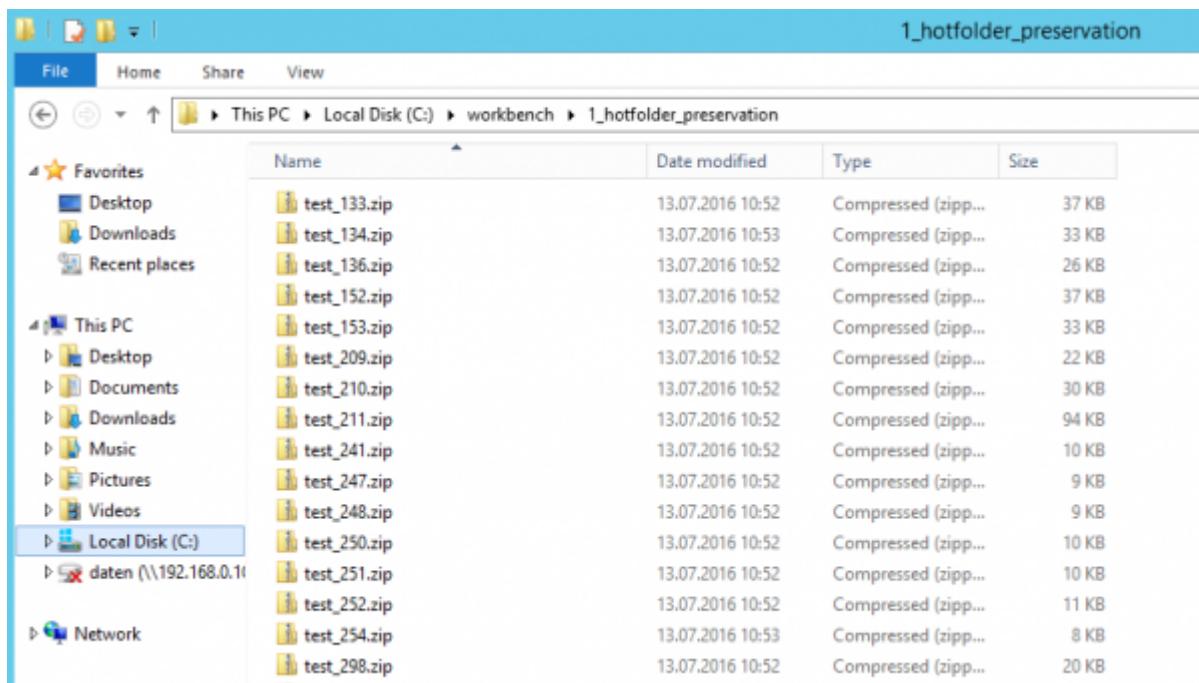
Objekt-ID

SIP

oder

Manuelle Eingabe

Die JPEG-Dateien werden je als eigene Informationspaket im Ordner **1_hotfolder_preservation** auf der Workbench gespeichert:

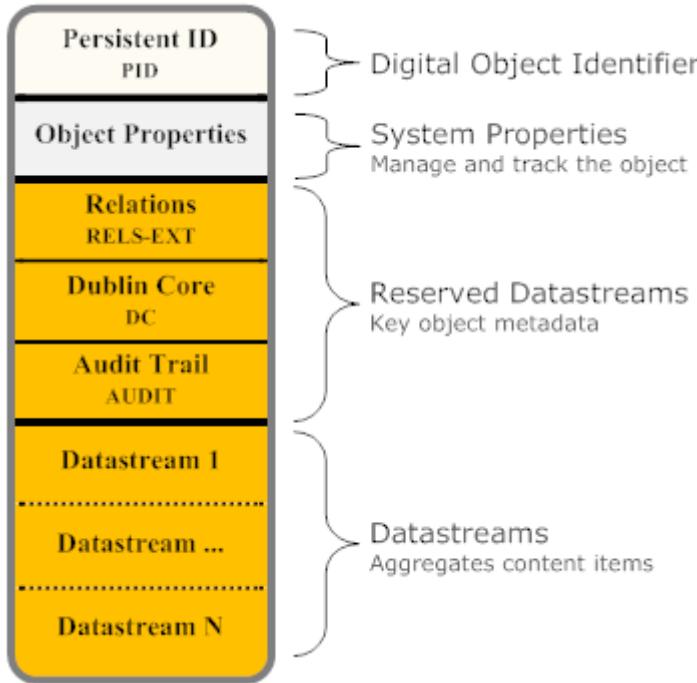


3. Durchführung der Migration

Die Preservation Action kann für Einzeldateien manuell über den Workflow „Update Fedora Objects“ durchgeführt werden. In der manuellen Eingabe wird dem Workflow der Name des zu migrierenden Pakets angegeben. Die manuell ausgeführte Preservation Action dient primär dem Testbetrieb, bei einer grösseren Menge von zu migrierenden Dateien wird der Workspace-Ordner „**1_hotfolder_preservation**“ durch einen observer überwacht und sobald dort neue Pakete eintreffen werden sie automatisiert in das neue Zielformat migriert.

The screenshot shows the 'docuteam feeder' interface. At the top, there is a navigation bar with links: 'Ausführungen', 'Workflows', 'Schritte', 'Benutzer', and 'Cockpit'. Below the navigation bar, the main title is 'Neue Ausführung für Preservation: Update Fedora Objects'. The search form is divided into two sections: 'Objekt-ID' and 'Manuelle Eingabe'. In the 'Objekt-ID' section, there is a dropdown menu labeled 'SIP' with a downward arrow. In the 'Manuelle Eingabe' section, there is a text input field containing 'test_133'. Below these fields is a button labeled 'Ausführung erstellen'. At the bottom right of the page, there is a footer with links: 'Angemeldet als admin | Logout | Dokumentation | Hilfe'.

Das [Objektmodell](#) von Fedora Commons sieht vor, dass Dateien in sogenannten „Datastreams“ gespeichert werden und dass diese Datastreams versioniert werden können. Docuteam legt für jede Datei ein eigenes Fedora-Objekt an und verknüpft die Dateien und Ordner über Relationen (beschrieben im RELS-EXT Datastream). Wenn nun eine Datei in ein neues Format migriert wird, dann wird der entsprechende Datastream nicht einfach überschrieben, sondern es wird eine neue Version davon angelegt. Dadurch bleibt der PID der Datei weiterhin gültig. Die Ausgangsdatei wird bei einer Preservation Action nicht gelöscht. Falls also im Migrationsprozess ein Fehler auftritt, kann weiterhin auf die Ursprungsdatei zurückgegriffen werden – auch wenn der Fehler erst zu einem späteren Zeitpunkt entdeckt wird. Während der Preservation Action werden zur Vorgangsdokumentation auch neue PREMIS-Metadaten geschrieben. Die PREMIS-Metadaten liegen in Fedora ebenfalls in einem eigenen Datastream, auch dieser wird in der Preservation Action versioniert.



From:
<https://wiki.docuteam.ch/> - **docuteam wiki**



Permanent link:
https://wiki.docuteam.ch/doku.php?id=docuteam:preservation_330&rev=1468475839

Last update: **2019/01/07 11:16**