

# 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 -->
<quid name="fmt/37" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/38" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/39" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/40" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/41" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/523" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/597" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/599" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />

<!-- Word for MS-DOS -->
<quid name="x-fmt/273" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/274" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/275" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/276" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />

<!-- Word for Mac -->
<quid name="x-fmt/2" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/129" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />

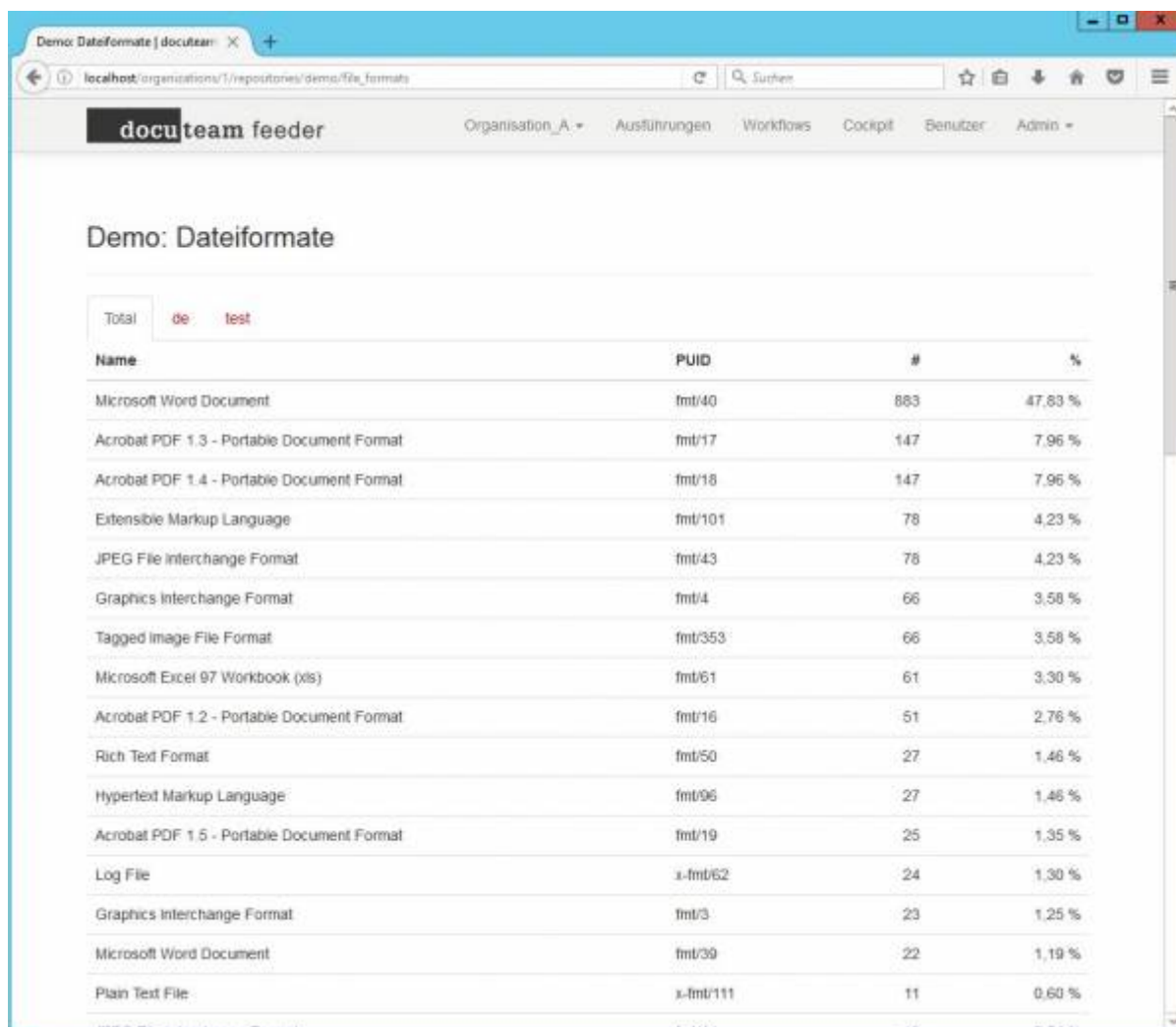
<!-- Wordperfect -->
<quid name="x-fmt/42" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/43" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/44" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/203" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/393" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/394" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="x-fmt/609" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />

<!-- Rich Text -->
<quid name="fmt/45" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/46" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/47" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/48" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/49" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/50" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/51" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/52" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="fmt/53" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />

<!-- RTF -->
<quid name="rtf" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="rtf" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
<quid name="rtf" applicationID="6" targetExtension="pdf" targetPronom="fmt/479,fmt/480,fmt/481" />
```

## 2. Grundlagen für das Preservation Planning ermitteln

Die Grundlagen für das Preservation Planning werden 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-Identifizier (PUID) an. [Pronom](#) ist die Fileregistry des britischen Nationalarchivs und erlaubt es, Dateiformate und Unterversionen 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 the 'Demo: Dateiformate' page in the Docuteam Feeder application. The interface includes a navigation bar with 'Organisation\_A', 'Ausführungen', 'Workflows', 'Cockpit', 'Benutzer', and 'Admin'. Below the navigation bar, there are tabs for 'Total', 'de', and 'test'. The main content is a table listing various file formats with their corresponding PUIDs, counts, and percentages.

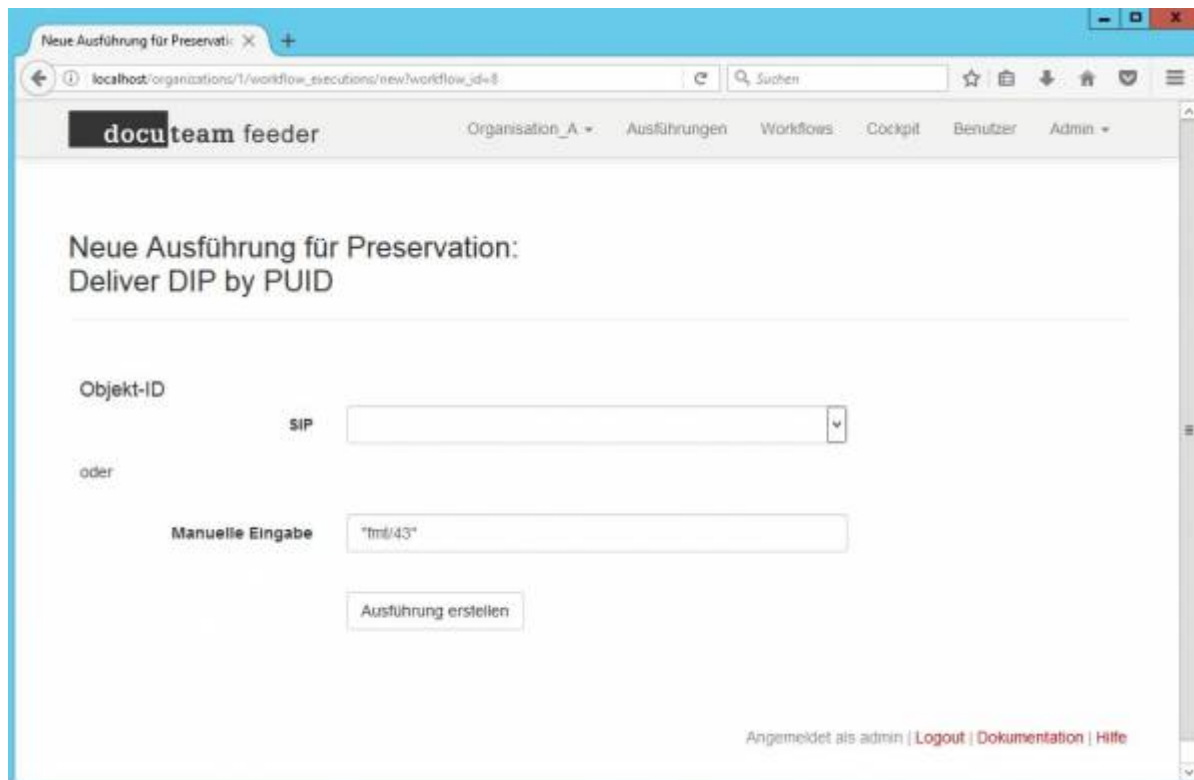
| Name                                       | PUID      | #   | %       |
|--------------------------------------------|-----------|-----|---------|
| Microsoft Word Document                    | fmt/40    | 883 | 47.83 % |
| Acrobat PDF 1.3 - Portable Document Format | fmt/17    | 147 | 7.96 %  |
| Acrobat PDF 1.4 - Portable Document Format | fmt/18    | 147 | 7.96 %  |
| Extensible Markup Language                 | fmt/101   | 78  | 4.23 %  |
| JPEG File Interchange Format               | fmt/43    | 78  | 4.23 %  |
| Graphics Interchange Format                | fmt/4     | 66  | 3.58 %  |
| Tagged Image File Format                   | fmt/353   | 66  | 3.58 %  |
| Microsoft Excel 97 Workbook (xls)          | fmt/61    | 61  | 3.30 %  |
| Acrobat PDF 1.2 - Portable Document Format | fmt/16    | 51  | 2.76 %  |
| Rich Text Format                           | fmt/50    | 27  | 1.46 %  |
| Hypertext Markup Language                  | fmt/96    | 27  | 1.46 %  |
| Acrobat PDF 1.5 - Portable Document Format | fmt/19    | 25  | 1.35 %  |
| Log File                                   | x-fmt/62  | 24  | 1.30 %  |
| Graphics Interchange Format                | fmt/3     | 23  | 1.25 %  |
| Microsoft Word Document                    | fmt/39    | 22  | 1.19 %  |
| Plain Text File                            | x-fmt/111 | 11  | 0.60 %  |

### 3. Extraktion der zu migrierenden Dateien

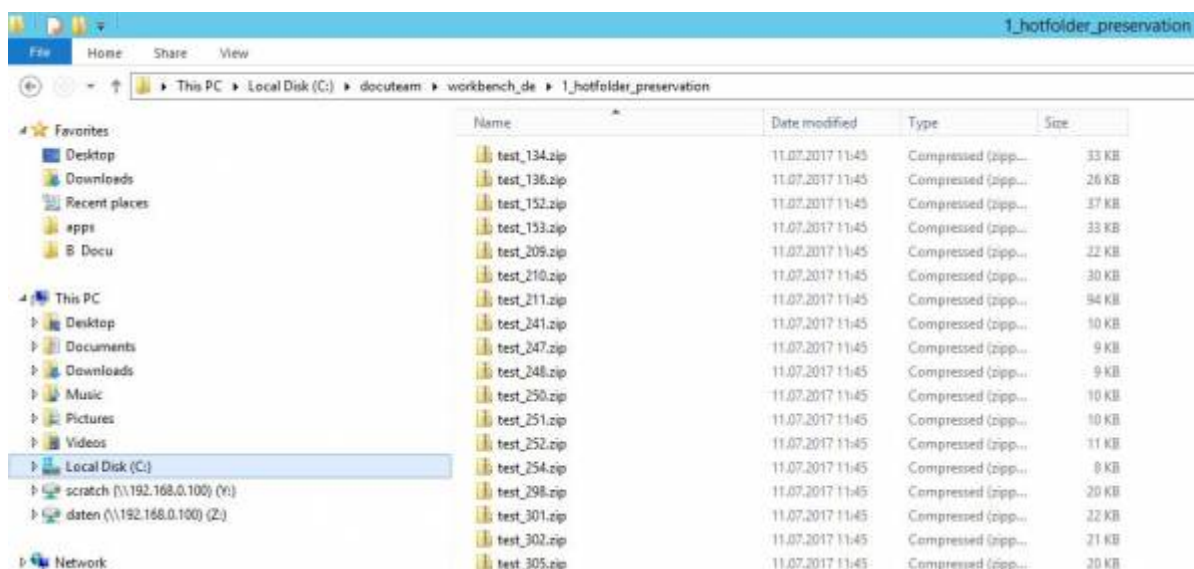
Preservation Actions werden immer pro Dateiformat durchgeführt. Beispiel: Im Archiv liegt verschiedene JPEG-Dateien, die im Rahmen einer Preservation Action nach Tiff migriert werden sollen. in migration.config.xml ist Tiff als Archivformat von JPEG hinterlegt.

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**.



Die JPEG-Dateien werden je als eigene Informationspakete im Ordner 1\_hotfolder\_preservation auf der Workbench gespeichert:

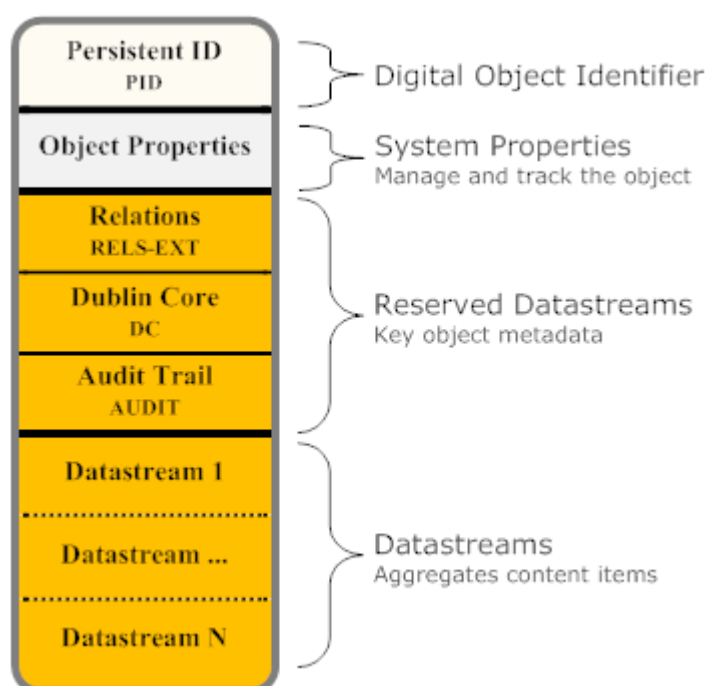


### 3. Durchführung der Preservation Action

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 a web browser window with the address bar displaying 'localhost/organizations/1/workflow\_executions/new/workflow\_id=9'. The page title is 'Neue Ausführung für Preservation: Update Fedora Objects'. The interface includes a navigation bar with 'Organisation\_A', 'Ausführungen', 'Workflows', 'Cockpit', 'Benutzer', and 'Admin'. The main form has two input fields: 'Objekt-ID' with a dropdown menu showing 'SIP', and 'Manuelle Eingabe' with the text 'test\_134'. A button labeled 'Ausführung erstellen' is at the bottom. The footer shows '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 Objekte 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 auch zukünftig jederzeit auf die Ursprungsdatei zurückgegriffen werden. Während der Preservation Action werden zur Vorgangsdokumentation auch neue PREMIS-Metadaten geschrieben. Diese liegen in Fedora ebenfalls in einem eigenen Datastream und auch dieser wird versioniert.



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

Permanent link:  
[https://wiki.docuteam.ch/doku.php?id=docuteam:preservation\\_330&rev=1546856669](https://wiki.docuteam.ch/doku.php?id=docuteam:preservation_330&rev=1546856669)

Last update: **2019/01/07 11:24**

