



Migration nach Fedora 6

Thomas Bernhart, Andreas Nef

Agenda

- Begrüssung
- Fedora Commons
- Migrations-Prozess
- Live-Demo
- Metadaten/Speicherstruktur
- Architektur/Default-Setup
- Zeitplanung
- Diskussion

Fedora Commons

Fedora steht für “Flexible Extensible Digital Object Repository Architecture”

- 2003 Version 1.0: XML
- 2008 Version 3.0
- 2014 Version 4.0: Linked Data, Modeshape für interne Speicherung
- 2015 Version 3.8.1
- 2018 Version 5.0
- 2021 Version 6.0: Oxford Common File Layout (OCFL) für interne Speicherung (Ablösung von Modeshape)

Fedora Commons

Einsatz bei docuteam ab Version 3.x:

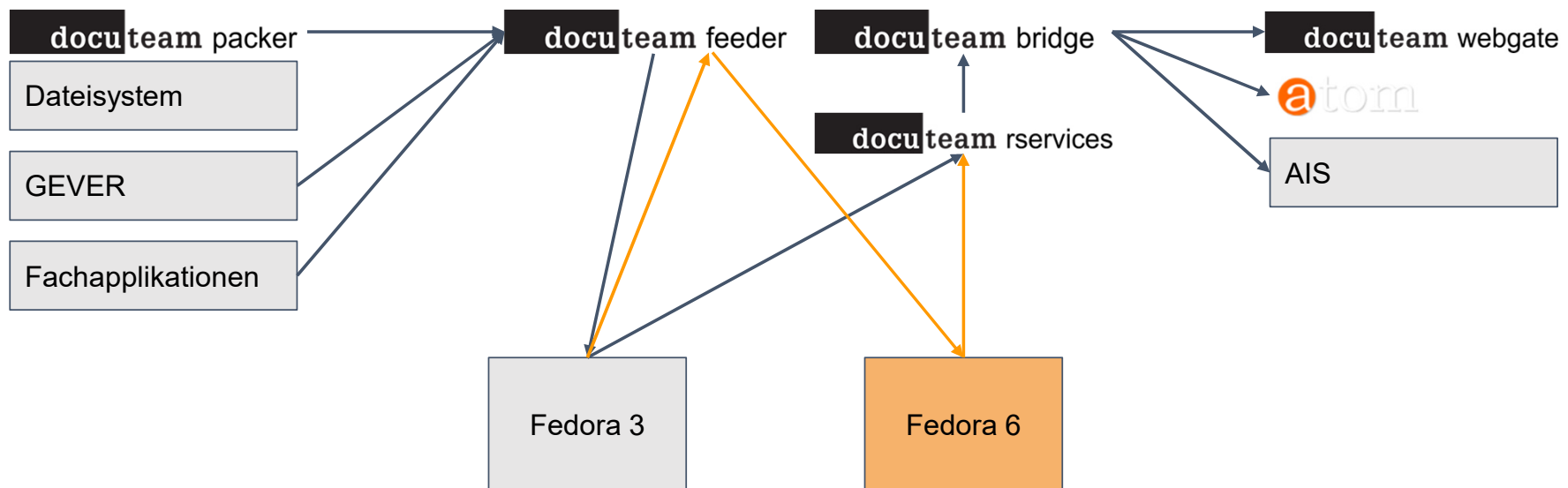
- Zuerst mit Java-Komponenten (heute: docuteam actions)
- Feeder als Workflow-Engine
- Fedora 3 als Repository
- Fedora GSearch als Suchindex
- EAD und PREMIS als zentrale Standards für Metadaten

Fedora 4 und 5 übersprungen

- Metadaten und Binaries in unterschiedlichen Systemen
 - Neues Datenmodell (Linked Data anstelle von XML)
 - Linked Data Standards für Archive
- RiC-O 0.2 Anfang 2021 veröffentlicht

Migrations-Prozess

- Keine direkte Migration auf Storage-Ebene, sondern über reguläre Fachprozesse
- Aufbau der F6-Instanz während laufendem Betrieb
- Möglichkeit für Parallelbetrieb von F3 und F6 während Migration



Migrations-Prozess

Ablauf:

- PIDs verifizieren F3 <> F6
 - Events in feeder
- Workflow pro zu verarbeitende PID
 - DIP-Erstellung ab F3
 - Bei Bedarf QA/Umwandlungen mit üblichen actions
 - Umwandlung nach Linked Data
 - Upload nach F6

In der Migrationszeit kann damit ein Parallelbetrieb stattfinden, während dessen die Pakete laufend migriert werden, neue bei Bedarf sowohl in F3 wie in F6 gespeichert werden, und nach der Verifikation einfach die F3-Instanz aufgehoben wird.

Migrations-Prozess

Matterhorn METS -> Matterhorn RDF

- Umwandlung von XML (METS) nach RDF (RiC) mittels RML ([RDF Mapping Language](#))
- RML kann auf unterschiedliche Datenquellen angewendet werden, hier wird direkt das XML verwendet.
- Für die Ausführung des Mappings benutzen wir [CARML](#).

```
...
rr:subjectMap [
  rr:template "recordresources/{../../../../../../../../@ADMID}" ;
];

rr:predicateObjectMap [
  rr:predicate rico:hasOrHadTitle ;
  rr:objectMap [
    rml:reference "EAD:did/EAD:unittitle[@label = 'main']" ;
  ];
];

rr:predicateObjectMap [
  rr:predicate rico:scopeAndContent ;
  rr:objectMap [
    rml:reference "EAD:scopecontent/EAD:p" ;
  ];
];
...
```

Demo

Metadaten

Fedora 3: XML mit EAD und PREMIS (und ...)



Fedora Digital Object

List Datastreams

Object Identifier (PID): CH-002003-9:38054

Version Date: current

Datastream ID		Datastream Label	MIME Type
RELS-EXT	view content	Relationship Metadata	application/rdf+xml
EAD	view content	Descriptive Metadata	text/xml
PREMIS	view content	Preservation Metadata	text/xml
ORIGINAL	view content	Content	application/pdf
DC	view content	Dublin Core Record for this object	text/xml

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <foxml:digitalObject VERSION="1.1" PID="CH-002003-9:38054" xmlns:foxml="info:fedora/fedora-system:def/foxm
3   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4   xsi:schemaLocation="info:fedora/fedora-system:def/foxml# http://www.fedora.info/definitions/1/0/foxml1
5 >
6   <foxml:objectProperties> [6 lines]
12  <foxml:datastream ID="AUDIT" STATE="A" CONTROL_GROUP="X" VERSIONABLE="false"> [17 lines]
30  <foxml:datastream ID="RELS-EXT" STATE="A" CONTROL_GROUP="X" VERSIONABLE="true"> [15 lines]
46  <foxml:datastream ID="EAD" STATE="A" CONTROL_GROUP="X" VERSIONABLE="true">
47    <foxml:datastreamVersion ID="EAD.0" LABEL="Descriptive Metadata" CREATED="2021-09-07T13:54:46.878Z
48      FORMAT_URI="urn:isbn:1-931666-22-9" SIZE="3001">
49      <foxml:contentDigest TYPE="MD5" DIGEST="3ecbbdbb98c7a10953ad3d1accb628f"/>
50      <foxml:xmlContent>
51        <EAD:ead xmlns:EAD="urn:isbn:1-931666-22-9" xmlns:xlink="http://www.w3.org/1999/xlink">
52          <EAD:eadheader> [7 lines]
60          <EAD:archdesc level="otherlevel">
61            <EAD:did> [2 lines]
64            <EAD:dsc>
65              <EAD:c level="otherlevel" otherlevel="Undefinient">
66                <EAD:did>
67                  <EAD:unittitle label="main">2014-09-23 Protokoll</EAD:unittitle>
68                  <EAD:physdesc>
69                    <EAD:extent unit="Datei(en)">1</EAD:extent>
70                    <EAD:physdesc>
71                      <EAD:unitid type="isVisible">true</EAD:unitid>
72                      <EAD:did>
73                        <EAD:dao xlink:role="simple" xlink:href="CH-002003-9:38054"/>
74                      </EAD:dao>
75                    </EAD:physdesc>
76                  </EAD:unitid>
77                </EAD:dsc>
78              </EAD:c>
79            </EAD:archdesc>
80          </EAD:ead>
81        </foxml:xmlContent>
82      </foxml:datastreamVersion>
83    </foxml:datastream>
84    <foxml:datastream ID="PREMIS" STATE="A" CONTROL_GROUP="X" VERSIONABLE="true">
85      <foxml:datastreamVersion ID="PREMIS.0" LABEL="Preservation Metadata" CREATED="2021-09-07T13:54:46.
86        FORMAT_URI="info:lc/xmlns/premis-v2" SIZE="10933">
87        <foxml:contentDigest TYPE="MD5" DIGEST="b5dd6995f8c642584a22245602d320bd"/>
88        <foxml:xmlContent>
89          <PREMIS:premis xmlns:PREMIS="info:lc/xmlns/premis-v2" version="2.2">
90            <PREMIS:object xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="PREMIS:
91              <PREMIS:objectIdentifier>
92                <PREMIS:objectIdentifierType>DocuTeam</PREMIS:objectIdentifierType>
93                <PREMIS:objectIdentifierValue>_20210907154936826</PREMIS:objectIdentifierValue>
94              </PREMIS:objectIdentifier>
95              <PREMIS:objectCharacteristics>
96                <PREMIS:compositionLevel>0</PREMIS:compositionLevel>
97                <PREMIS:fixity>
98                  <PREMIS:messageDigestAlgorithm>SHA-512</PREMIS:messageDigestAlgorithm>
99                  <PREMIS:messageDigest>c72f5a3f483eac85c0f0be94bb83e9bb2bdc8a2466655b1e23
100                </PREMIS:fixity>
101              </PREMIS:objectCharacteristics>
102            </PREMIS:premis>
103          </foxml:xmlContent>
104        </foxml:datastreamVersion>
105      </foxml:datastream>
106    </foxml:objectProperties>
107  </foxml:digitalObject>
```

F6: RDF mit RiC und PREMIS

```

1 PREFIX : <undefined>
2 SELECT ?record ?pid
3 WHERE {
4   ?identifier <https://www.ica.org/standards/RiC/ontology#isOrWasIdentifierOf> ?record.
5   ?identifier <https://www.ica.org/standards/RiC/ontology#_type> "PID".
6   ?identifier <https://www.ica.org/standards/RiC/ontology#name> ?pid
7 }

```

QUERY RESULTS

Table Raw Response

Showing 1 to 3 of 3 entries

	record	pid
1	<https://099repository-green-cosmos.docuteam.cloud/fcrepo/rest/recordresources/_20211019141954863>	"CH-002003-9:2171"
2	<https://099repository-green-cosmos.docuteam.cloud/fcrepo/rest/recordresources/_20211019142009579>	"CH-002003-9:2174"
3	<https://099repository-green-cosmos.docuteam.cloud/fcrepo/rest/recordresources/_20211019142028194>	"CH-002003-9:2177"

Showing 1 to 3 of 3 entries

Properties

ebucore: **hasMimeType**
application/pdf

fedora: **created**
2021-10-18T22:10:00.902975Z

fedora: **createdBy**
fedoraAdmin

fedora: **lastModified**
2021-10-18T22:10:00.902975Z

fedora: **lastModifiedBy**
fedoraAdmin

http://www.loc.gov/premis/rdf/v3/ **fixity**
https://099repository-green-cosmos.docuteam.cloud/fcrepo/rest/#fixity__20210907154936826

.../ **storedAt**
file:///496_1 Vorprojekt/2014-09-23 Protokoll.pdf/2014-09-23 Protokoll.pdf

https://www.ica.org/standards/RiC/ontology# **isOrWasComponentOf**
https://099repository-green-cosmos.docuteam.cloud/fcrepo/rest/recordresources/_20211019000934273/instances/_20211019000934318

ldp: **contains**
https://099repository-green-cosmos.docuteam.cloud/fcrepo/rest/recordresources/_20211019000934273/instances/_20211019000934318/binaries/_20211019000934318/aea737ef-3096-45f2-912e-c7add71ecd04

rdf: **type**
<http://fedora.info/definitions/v4/repository#Container>
<http://fedora.info/definitions/v4/repository#Resource>
<http://www.loc.gov/premis/rdf/v3/File>
<http://www.w3.org/ns/ldp#BasicContainer>
<http://www.w3.org/ns/ldp#Container>
<http://www.w3.org/ns/ldp#RDFSResource>
<http://www.w3.org/ns/ldp#Resource>
<https://www.ica.org/standards/RiC/ontology#Instantiation>

F3: objectStore (XML) und datastreamStore (binaries)

```
$ find . -name "*CH-002003-9%3A38054*"
./objectStore/CH-002003-9/0e/ea/info%3Afedora%2FCH-002003-9%3A38054
./datastreamStore/CH-002003-9/76/8d/info%3Afedora%2FCH-002003-9%3A38054%2FORIGINAL%2FORIGINAL.0
$
```

FOXML mit Referenz auf das binary:

```
...
<foxml:datastream ID="ORIGINAL" STATE="A" CONTROL_GROUP="M" VERSIONABLE="true">
  <foxml:datastreamVersion ID="ORIGINAL.0" LABEL="Content" CREATED="2021-09-07T13:54:46.878Z"
MIMETYPE="application/pdf" SIZE="139992">
    <foxml:contentDigest TYPE="SHA-512" DIGEST="c72f5a3f483each85c0f09...7b8aa5a67eb6cfb10715108a76a45"/>
    <foxml:contentLocation TYPE="INTERNAL_ID" REF="CH-002003-9:38054+ORIGINAL+ORIGINAL.0"/>
  </foxml:datastreamVersion>
</foxml:datastream>
...
```

Speicherstruktur

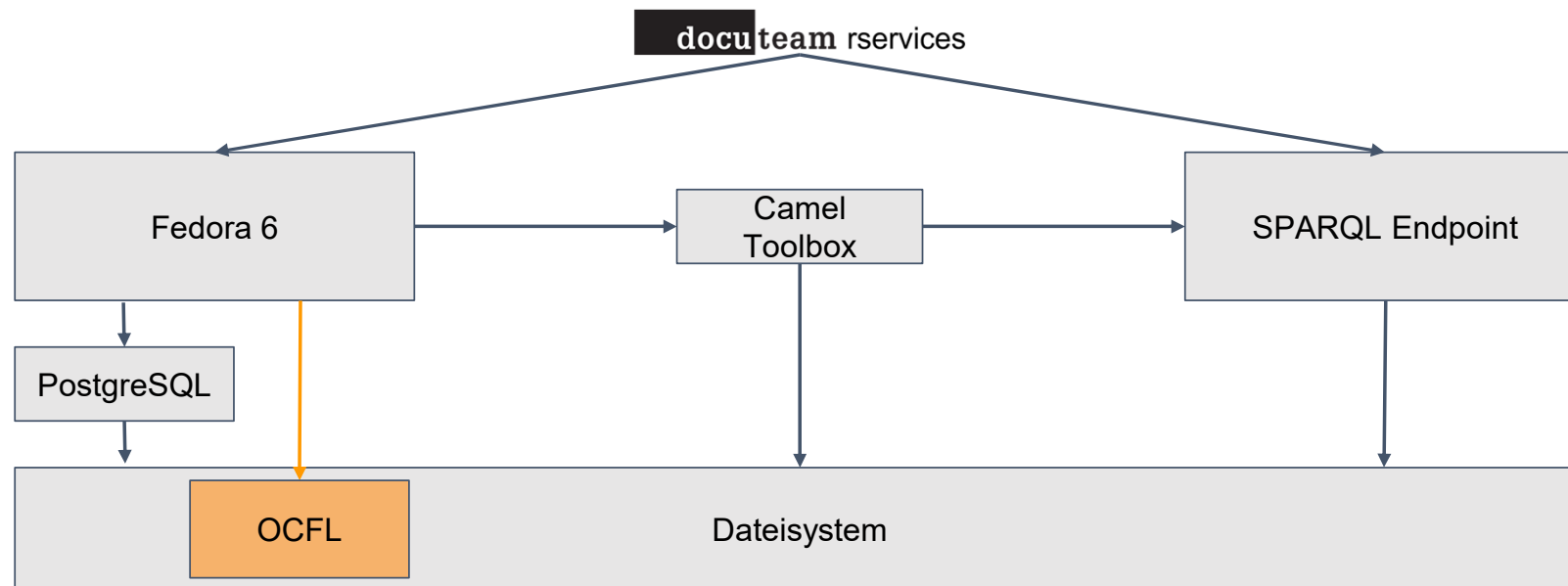
F6: OCFL

- versioned
- self-contained
- metadata together with binaries

```
[ $ tree 725/7f3/e12/7257f3e12459731972a20d47affce3f893cc5891d5368e21d91239a4d0197503
725/7f3/e12/7257f3e12459731972a20d47affce3f893cc5891d5368e21d91239a4d0197503
├── 0=ocfl_object_1.0
├── inventory.json
├── inventory.json.sha512
├── v1
│   ├── content
│   │   └── fcr-container.nt
│   ├── inventory.json
│   └── inventory.json.sha512
├── v2
│   ├── content
│   │   ├── instances
│   │   │   └── _20211019084008465
│   │   │       ├── binaries
│   │   │       │   └── _20211019084008465
│   │   │           └── fcr-container.nt
│   ├── inventory.json
│   └── inventory.json.sha512
├── v3
│   ├── content
│   │   ├── instances
│   │   │   └── _20211019084008465
│   │   │       ├── binaries
│   │   │       │   └── _20211019084008465
│   │   │           ├── 478c9ebb-4c45-4a22-9b58-3480863f9c58
│   │   │           └── 478c9ebb-4c45-4a22-9b58-3480863f9c58~fcr-desc.nt
│   ├── inventory.json
│   └── inventory.json.sha512
├── v4
│   ├── content
│   │   └── pid
│   │       └── fcr-container.nt
│   ├── inventory.json
│   └── inventory.json.sha512
17 directories, 16 files
$ █
```

Architektur Repository Server

- Relevante Daten in OCFL gespeichert
- SPARQL Endpoint als Suchindex
- rservices als API-Layer auf Fedora 6 und SPARQL Endpoint



Zeitplanung

- Q4/2021:
 - Finalisierung Mapping
 - Abschluss Default-Konfiguration/-Setup
- Q1/2022:
 - Migration cloud
 - Migration erste Kunden
- Q2/2022:
 - Ausrollen Kunden

Diskussion

Kontakt

Thomas Bernhart

Projektleiter

t.bernhart@docuteam.ch

Andreas Nef

Technical Lead

a.nef@docuteam.ch

docuteam AG

Im Langacker 16

CH-5405 Baden

T +41 56 470 03 37

<https://www.docuteam.ch/>

