

Stefan Hein

Metadaten für die Langzeitarchivierung

15. September 2011



Gliederung

1. Einführung

Motivation, Was sind Metadaten?

THEORIE

2. Metadatenstandards

DC, METS, PREMIS und Co.

3. Generierung von Metadaten

Metadatentools

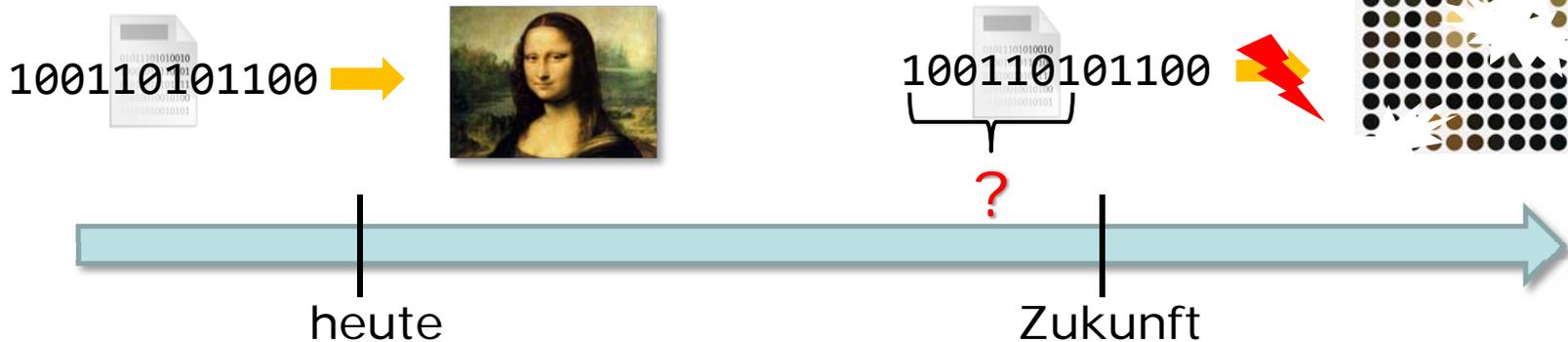
PRAXIS

4. Anwendungsszenarien

und nun ?

Einführung - Motivation

Das Problem:



➔ **Welche** Informationen brauchen wir, um den Strom aus Nullen und Einsen auch in Zukunft noch interpretieren können?

➔ **Wie** können wir diese Informationen gewinnen?

Einführung – Metadaten (1)

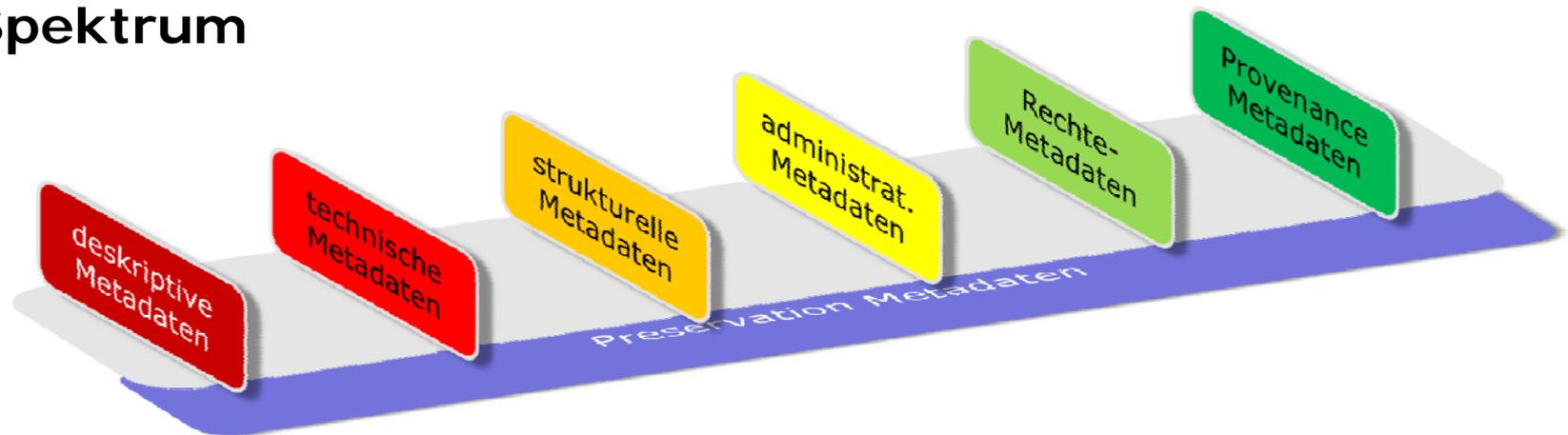
OAIS-Referenzmodell:

- Representation Information: *Informationen, um ein digitales Objekt in einer für den Nutzer verständlichen Form darzustellen.*
- Preservation Description Information (PDI): Provenance, Reference, Fixity und Context Information.

 **Metadaten** sind das Konzept, diese im OAIS-Referenzmodell definierten Informationen zu erfassen und zu speichern.

Einführung – Metadaten (2)

Spektrum



- Metadattentypen dienen unterschiedlichen Bedürfnissen
- Metadaten zur Langzeitarchivierung (Preservation Metadata) bilden einen Querschnitt durch alle Typen

Gliederung

1. Einführung

Motivation, Was sind Metadaten?

THEORIE

2. Metadatenstandards

DC, METS, PREMIS und Co.

3. Generierung von Metadaten

Metadatentools

PRAXIS

4. Anwendungsszenarien

und nun ?

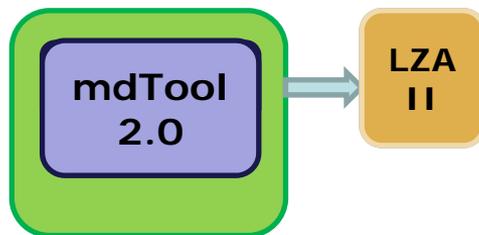
Metadatenstandards

- Standards definieren und vereinheitlichen die einzelnen Elemente, den Wertebereich und den strukturellen Aufbau von Metadaten
- *Die Verwendung von Standards verbessert die:*

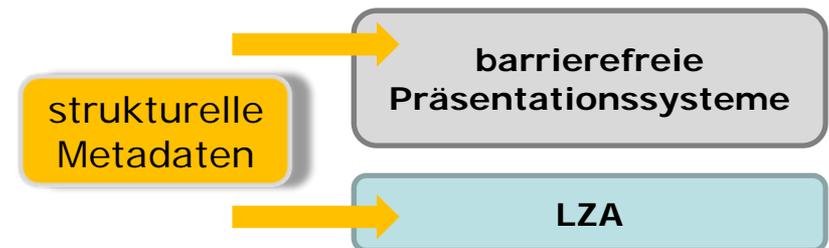
- Interoperabilität



- Austauschbarkeit

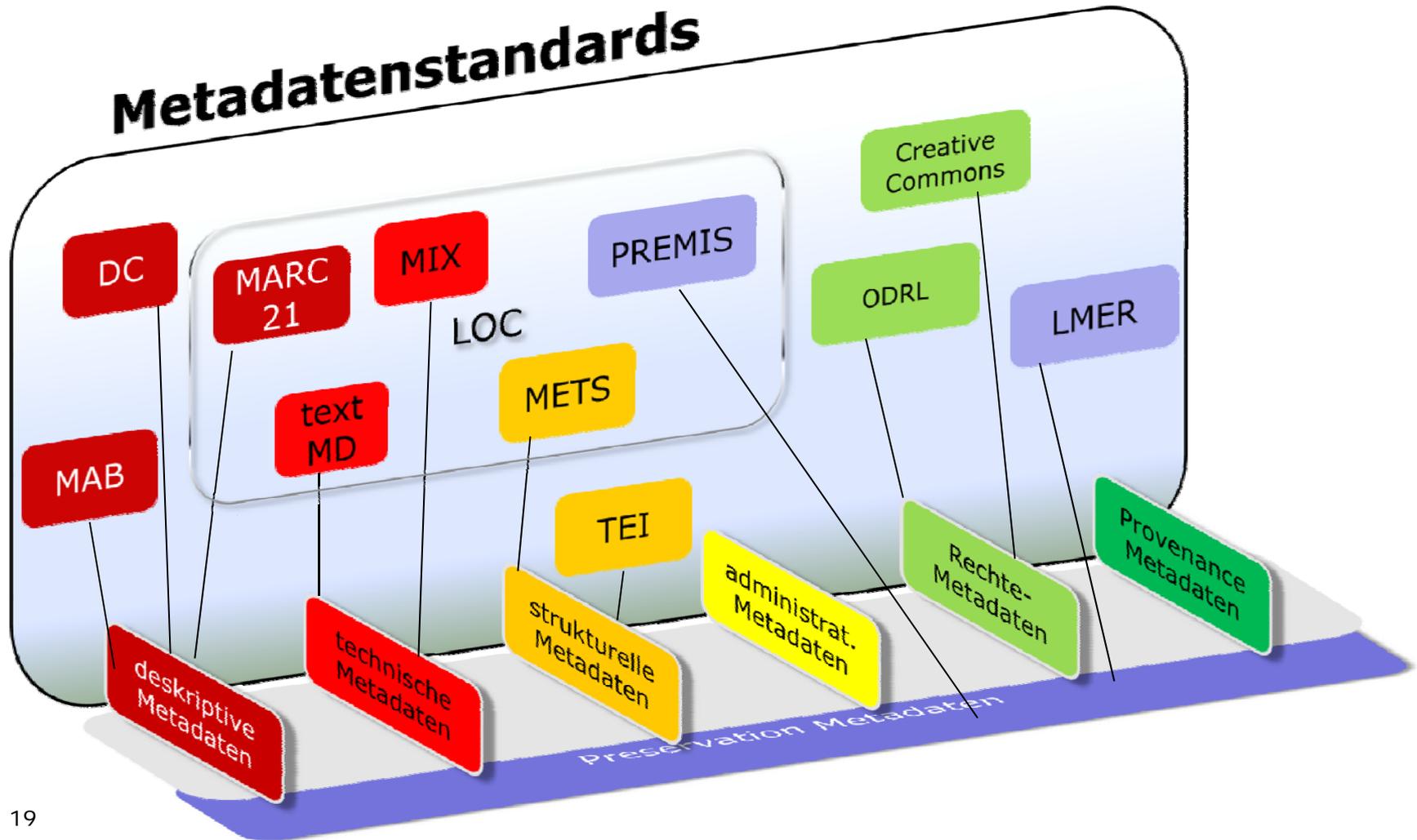


- Wiederverwendbarkeit



- Transparenz
- Integrationsfähigkeit

Metadatenstandards



Gliederung

1. Einführung

Motivation, Was sind Metadaten?

THEORIE

2. Metadatenstandards

DC, METS, PREMIS und Co.

3. Generierung von Metadaten

Metadatentools

PRAXIS

4. Anwendungsszenarien

und nun ?

Generierung von Metadaten

Was lässt sich automatisiert generieren bzw. erfassen?

- Dateiformat (Identifizierung)
- formatspezifische **technische** Metadaten
- **strukturelle** Metadaten
- Dokumentenbeschränkungen
- Dateiformat-Validität
- administrative Metadaten
- Provenance Metadaten

Generierung von Metadaten

Wie generieren und ablegen?

- Integration von **Metadatentools** in den
Verarbeitungsworkflow (Ingest, Preservation Action)
- Metadaten sollten maschinenlesbar (XML) kodiert
werden und wenn möglich einem anerkannten
Metadatenstandard entsprechen
- ggf. Normalisierung des Outputs
- Speicherung im AIP und im Data-Management des LZA

Metadatentool: FITS (1)

- Beispiel : JPEG-Datei

```
<?xml version="1.0" encoding="UTF-8" ?>
<fits xmlns="http://hul.harvard.edu/ois/xml/ns/fits/fits_output" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output
http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version="0.5.0" timestamp="24.03.11 11:32">
  <identification>
    <identity format="JPEG File Interchange Format" mimetype="image/jpeg">
      <tool toolname="Jhove" toolversion="1.5" />
      <tool toolname="file utility" toolversion="5.03" />
      <tool toolname="Exiftool" toolversion="7.74" />
      <tool toolname="Droid" toolversion="3.0" />
      <version toolname="Jhove" toolversion="1.5">1.01</version>
      <externalIdentifier toolname="Droid" toolversion="3.0" type="puid">fmt/43</externalIdentifier>
    </identity>
  </identification>
  <fileinfo>
    <size toolname="Jhove" toolversion="1.5">9071</size>
    <lastmodified toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">2011:02:11 14:54:50+01:00</lastmodified>
    <filename toolname="OIS File Information" toolversion="0.1" status="SINGLE_RESULT">bauarbeiter.jpg</filename>
    <md5checksum toolname="OIS File Information" toolversion="0.1" status="SINGLE_RESULT">e63f2ff66703d55ee29aaad329cf4d9b</md5checksum>
    <fslastmodified toolname="OIS File Information" toolversion="0.1" status="SINGLE_RESULT">1297432490221</fslastmodified>
  </fileinfo>
  <filestatus>
    <well-formed toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</well-formed>
    <valid toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</valid>
  </filestatus>
  <metadata>
    <image>
      <byteOrder toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">big endian</byteOrder>
      <compressionScheme toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">JPEG (old-style)</compressionScheme>
      <imageWidth toolname="Jhove" toolversion="1.5">100</imageWidth>
      <imageHeight toolname="Jhove" toolversion="1.5">100</imageHeight>
      <colorSpace toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">YCbCr</colorSpace>
      <YCbCrSubSampling toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">2 2</YCbCrSubSampling>
      <samplingFrequencyUnit toolname="Jhove" toolversion="1.5">in.</samplingFrequencyUnit>
      <xSamplingFrequency toolname="Jhove" toolversion="1.5">96</xSamplingFrequency>
      <ySamplingFrequency toolname="Jhove" toolversion="1.5">96</ySamplingFrequency>
      <bitsPerSample toolname="Jhove" toolversion="1.5">8 8 8</bitsPerSample>
      <samplesPerPixel toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">3</samplesPerPixel>
      <ingestLevel toolname="Didig0" toolversion="0.7" status="SINGLE_RESULT">4</ingestLevel>
    </image>
  </metadata>
</fits>
```

Metadatentool: FITS (2)

- Beispiel : JPEG-Datei

```

<identification>
  <identity format="JPEG File Interchange Format" mimetype="image/jpeg">
    <tool toolname="Jhove" toolversion="1.5" />
    <tool toolname="file utility" toolversion="5.03" />
    <tool toolname="Exiftool" toolversion="7.74" />
  </identity>
</identification>

<fileinfo>
  <size toolname="Jhove" toolversion="1.5">9071</size>
  <lastmodified toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">2011:02:11 14:54:50+01:00</lastmodified>
</fileinfo>

<metadata>
  <image>
    <byteOrder toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">big endian</byteOrder>
    <compressionScheme toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">JPEG (old-style)</compressionScheme>
    <imageWidth toolname="Jhove" toolversion="1.5">100</imageWidth>
    <imageHeight toolname="Jhove" toolversion="1.5">100</imageHeight>
    <colorSpace toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">YCbCr</colorSpace>
    <YCbCrSubSampling toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">2 2</YCbCrSubSampling>
    <samplingFrequencyUnit toolname="Jhove" toolversion="1.5">in.</samplingFrequencyUnit>
    <xSamplingFrequency toolname="Jhove" toolversion="1.5">96</xSamplingFrequency>
    <ySamplingFrequency toolname="Jhove" toolversion="1.5">96</ySamplingFrequency>
    <bitsPerSample toolname="Jhove" toolversion="1.5">8 8 8</bitsPerSample>
    <samplesPerPixel toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">3</samplesPerPixel>
    <ingestLevel toolname="Didig0" toolversion="0.7" status="SINGLE_RESULT">4</ingestLevel>
  </image>
</metadata>
  
```

Gliederung

1. Einführung

Motivation, Was sind Metadaten?

THEORIE

2. Metadatenstandards

DC, METS, PREMIS und Co.

3. Generierung von Metadaten

Metadatentools

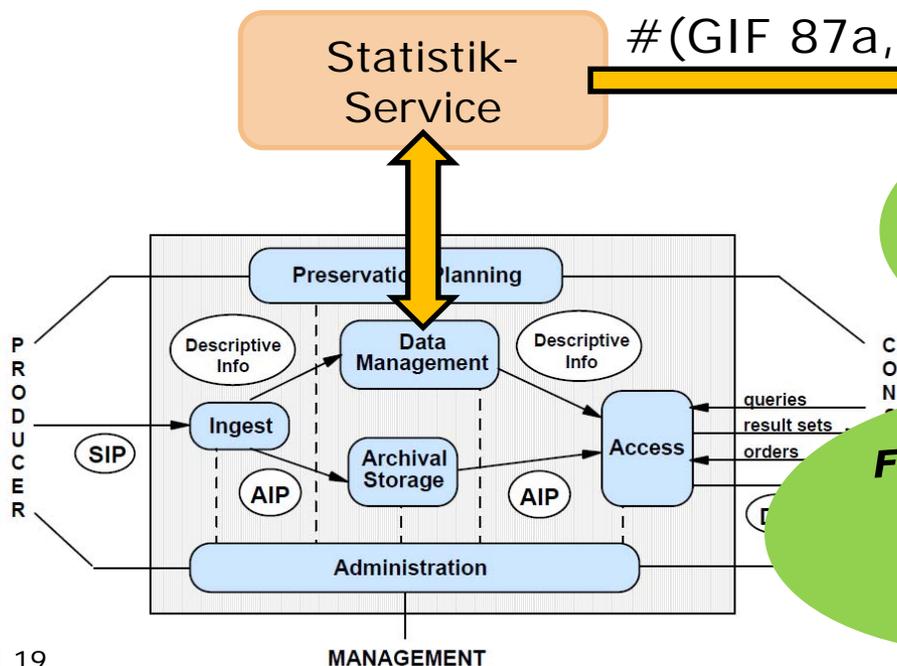
PRAXIS

4. Anwendungsszenarien und nun ?

Anwendungsszenario 1

Preservation Planning

Metadaten als wertvolle Datenquelle



Wir müssen überprüfen, ob damit einhergehend auch die Unterstützung durch Viewer und Migrationstools rückläufig ist.

Wieviel GIF-Dateien der Version 87a wurden eigentlich in den letzten 3 Jahren archiviert?

.. wenn ja, sollten wir migrieren.

Diese Formatversion scheint langsam obsolet zu werden!



Anwendungsszenario 2 - Migration

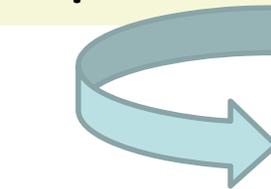
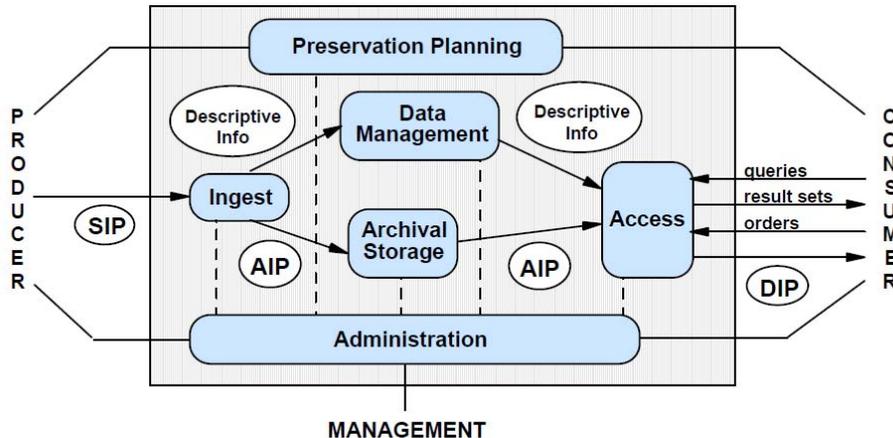
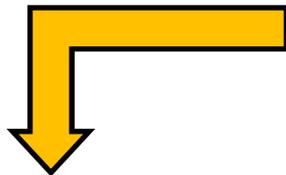
```
<imageWidth toolname="Jhove" toolversion="1.5">200 </imageWidth>
<imageHeight toolname="Jhove" toolversion="1.5">40 </imageHeight>
```

Überprüfung
der
signifikanten
Eigenschaften

Qualitätssicherung



?



SELECT *
WHERE format="GIF"
AND version="87a"

Anwendungsszenario 3

Die Generierung von technischen Metadaten als ein Baustein des Risiko Managements

beim Ingest – also noch vor der Archivierung:

- problembehaftete Objekte **frühzeitig** erkennen

```
<filestatus>  
  <well-formed toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">false</well-formed>  
  <valid toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">false</valid>  
  <message toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">No PDF trailer offset=24608685</message>  
</filestatus>
```

- **frühzeitiges** Erkennen von Sicherheitsbarrieren

```
<isRightsManaged toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">yes</isRightsManaged>  
<isProtected toolname="Exiftool" toolversion="7.74" status="SINGLE_RESULT">yes</isProtected>
```

Quellen

- JHOVE :
<http://hul.harvard.edu/jhove/index.html>
- JHOVE 2:
<https://confluence.ucop.edu/display/JHOVE2Info/Home>
- FITS:
<http://code.google.com/p/fits/>
- OAIS-Referenzmodell:
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

Vielen Dank für Ihre Aufmerksamkeit

Haben Sie noch Fragen ???

Stefan Hein
Deutsche Nationalbibliothek
Informationstechnik
Adickesallee 1
D-60322 Frankfurt am Main
Tel.: +49-69-1525-1722
<mailto:s.hein@dnb.de>
<http://www.dnb.de>