

Stefan Hein

Tools zur Generierung technischer Metadaten



Der rote Faden

- 1. Leistungsspektrum**
was Metadatentools leisten können
- 2. Metadatentools**
und ihre Vielfalt
- 3. Auswahlkriterien**
die Wahl des passenden Tools
- 4. JHOVE & Co.**



Der rote Faden

- 1. Leistungsspektrum**
was Metadatentools leisten können
2. Metadatentools
und ihre Vielfalt
3. Auswahlkriterien
die Wahl des passenden Tools
4. JHOVE & Co.



Metadatentools und ihr Leistungsspektrum

I. Dateiformate identifizieren

Welches Dateiformat hat ein vorliegendes digitales Objekt?

II. Datei gegen das Format validieren

Kann dem vertraut werden, was ein digitales Objekt vorgibt zu sein?

III. techn. Metadaten extrahieren

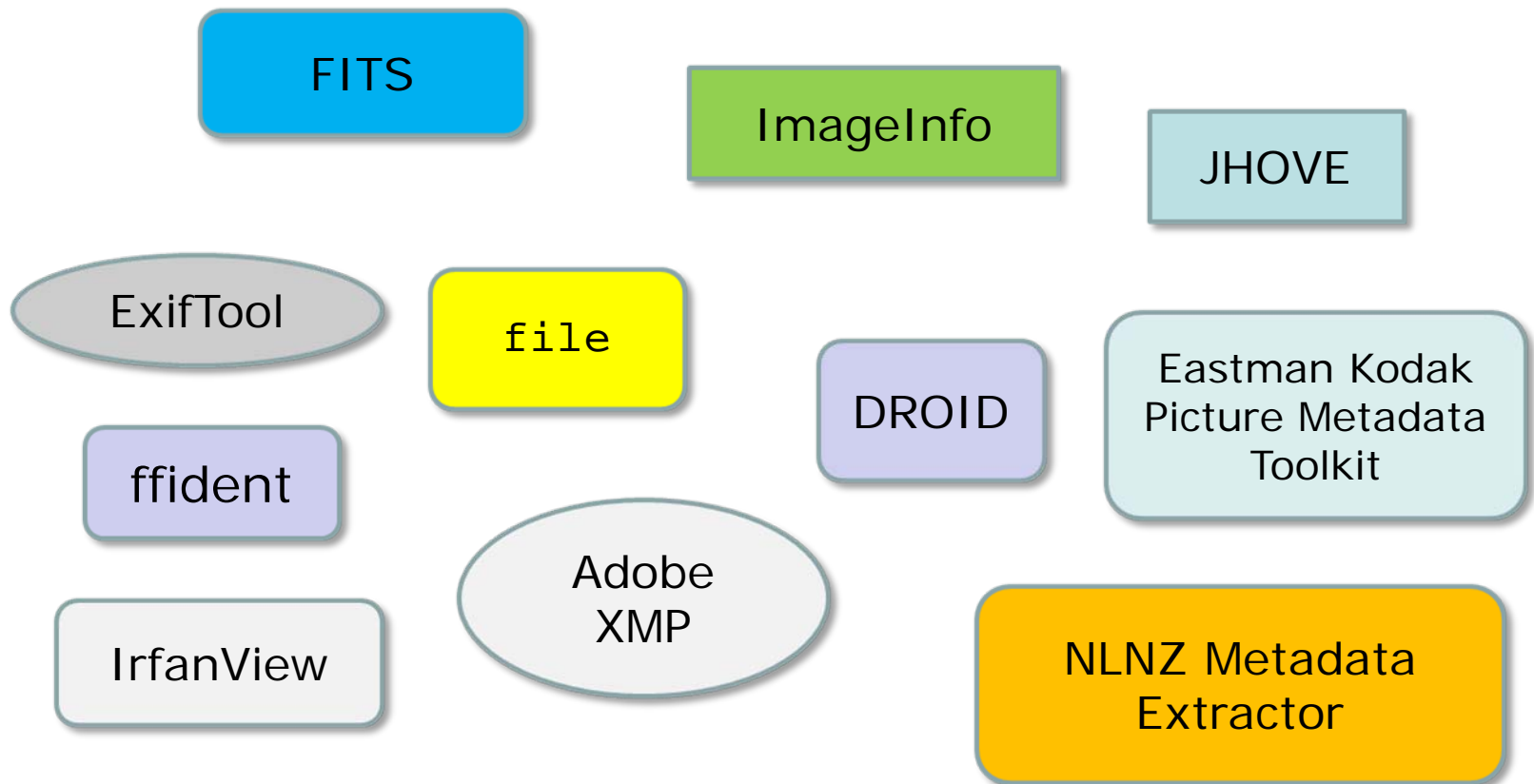
Welches sind die technischen Eigenschaften eines vorliegenden digitalen Objekts?

Der rote Faden

1. Leistungsspektrum
was Metadatentools leisten können
- 2. Metadatentools
und ihre Vielfalt**
3. Auswahlkriterien
die Wahl des passenden Tools
4. JHOVE & Co.



Metadatentools



Ordnung nach funktionalen Fähigkeiten

<p>Identifikation</p> <p>DROID, ImageInfo, fident, JHOVE, ExifTool, NLNZ-MDExt, file, IrfanView, Adobe XMP</p>	<p>Validation</p> <p>JHOVE, Adobe XMP, „XML-Validatoren“</p>
<p>Techn. MD Extraktion</p> <p>ImageInfo, JHOVE, ExifTool, NLNZ-MDExt, IrfanView, Adobe XMP</p>	

Der rote Faden

1. Leistungsspektrum
was Metadatentools leisten können
2. Metadatentools
und ihre Vielfalt
- 3. Auswahlkriterien**
die Wahl des passenden Tools
4. JHOVE & Co.



Auswahlkriterien (1)

- Anforderungen an Funktionalität und Formatunterstützung
- Anforderungen an Skalierbarkeit, Performance, Robustheit , Zuverlässigkeit (qualitative Kriterien)
- Dokumentation, Verfügbarkeit (Weiterentwicklung), Support (Integrierbarkeit, Austauschbarkeit, Wartung, Anpassbarkeit bei der Entwicklung von Erweiterungen)
- Open-Source vs. proprietär (Lizenzrecht)

Auswahlkriterien (2)

- standardisierte und dokumentierte Schnittstellen
- automatisierte Verarbeitung (Stapelverarbeitung) z. B. per Kommandozeile bzw. Aufruf über Schnittstellen
- Softwarearchitektur (z. B. modularer Aufbau)
- Systemanforderungen

Der rote Faden

1. **Leistungsspektrum**
was Metadatentools leisten können
2. **Metadatentools**
und ihre Vielfalt
3. **Auswahlkriterien**
die Wahl des passenden Tools
4. **JHOVE & Co.**



JH VE

JSTOR/Harvard Object Validation Environment

- Zusammenarbeit von *Journal STORage (JSTOR)* und *Harvard University Library* zur Format-Validation
- im Jahr 2005 veröffentlicht
- Java 1.4 (plattformunabhängig)
- Swing-basierte GUI oder Kommandozeile
- Open-Source unter LGPL



JSTOR/Harvard Object Validation Environment

Für die **Dateiformate:**

[ASCII](#) and [UTF-8](#) encoded text, [GIF](#), [JPEG2000](#), and [JPEG](#), and [TIFF](#) images, [AIFF](#) and [WAVE](#) audio, [PDF](#), [HTML](#), and [XML](#)

bietet JHOVE folgende **Funktionen:**

Identifikation	X	Validation	X
Extraktion techn. MD	X	Checksummen	X
Container-Verarbeitung	O	Erkennen von Beschränkungen	X (nur PDF)
Standardisierte Ausgabe (XML, z. T. unter Verwendung von Metadatenstandards)	X		

JHOVE im Vergleich (1)

	DROID	JHOVE	NLNZ
Identifies formats	Yes (on basis of file extension only)	Yes	Yes (on basis of file extension only)
Identifies versions	Yes	Yes	Yes
Confirms well-formed and valid	No	Yes	No
Extracts technical metadata	No	Yes	Yes (limited)
Outputs xml reports	Yes	Yes	Yes

Quelle: http://artefactual.com/wiki/index.php?title=DROID,_JHOVE,_NLNZ_Metadata_Extractor

	DROID	JHOVE	NLNZ
Office documents	DBF, DOC, Lotus formats, MS Works formats, OpenOffice formats, MDB, MPP, PDF, PPT, PST, PUB, RTF, StarOffice formats, TXT, VSD, WPD, WS and other WordStar formats, XLS (BIFF) and others	PDF, ASCII, UTF-8	DOC, MS Works formats, OpenOffice formats, PDF, PPT, WPD, XLS
Images	BMP, CDR and other Corel formats, DWG and other AutoCad formats, DXF, EPS, GeoTIFF, GIF, JPEG, JPEG2000, PageMaker documents, PCX, PNG, PS, PSD, PSP, SWF and other Macromedia formats, SVG, TIFF and others	GIF, JPEG, JPEG2000, TIFF	BMP, GIF, JPEG, TIFF
Sound / moving image	AIFF, ASF, AVI, MIDI, MOV, MP3, MPG, Real Audio (RM/A), WAV and others	AIFF, WAV	WAV, MP3
Markup languages	GML, HTML, ODF, XML, XHTML and others	HTML, XHTML, XML	HTML, XML

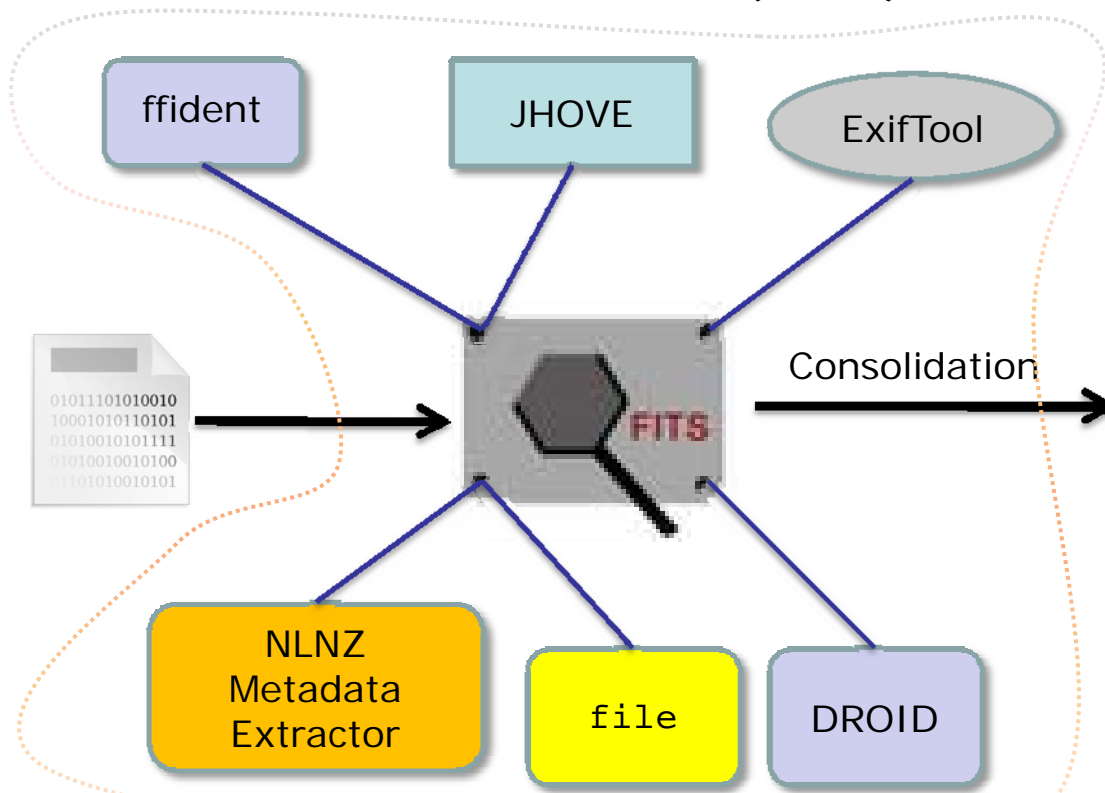
Quelle: http://artefactual.com/wiki/index.php?title=DROID,_JHOVE,_NLNZ_Metadata_Extractor

JHOVE2

- verbessertes Datenmodell:
1 object = n files = m formats
rekursive Verarbeitung von Containern
und Verzeichnissen
- **generalisiertes
Charakterisierungskonzept** (Policy-
basierte Bewertungskomponente)
- **verbesserte Identifikation**
- **Dateiformate:** ICC color profile, JPEG 2000,
PDF, SGML, Shapefile, TIFF, UTF-8, Wave,
XML, Zip
- neue Architektur -> **nicht**
kompatibel zu JHOVE 1
- Java 1.6
- aktuelle Beta-Version
unter BSD open source
license
- buffered I/O

Framework

– File Information Tool Set (FITS)



```
<?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="0.4.2"
timestamp="9/2/10 3:41 PM">
<identification>
<identity format="Tagged Image File Format" mimetype="image/tiff">
<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" />
<tool toolname="NLNZ Metadata
```

FITS-Output Object

Quellen (1)

- JHOVE :
<http://hul.harvard.edu/jhove/index.html>
- JHOVE 2:
<https://confluence.ucop.edu/display/JHOVE2Info/Home>
- Adobe XMP:
<http://www.adobe.com/products/xmp/>
- FITS:
<http://code.google.com/p/fits/>

Quellen (2)

- Unix-Manpage file:
<http://unixhelp.ed.ac.uk/CGI/man-cgi?file>
- ExifTool:
<http://www.sno.phy.queensu.ca/~phil/exiftool/>
- NLNZ Metadata Extractor:
<http://meta-extractor.sourceforge.net/>
- DROID
<http://droid.sourceforge.net/>

Quellen (3)

- ImageInfo:
<http://web.archive.org/web/20060715192746/schmidt.devlib.org/image-info/index.html>
- Eastman Kodakt Picture Metadata Toolkit:
<http://picturemetadata.sourceforge.net/>
- ffident:
<http://web.archive.org/web/20060715192855/schmidt.devlib.org/ffident/index.html>
- IrfanView: <http://www.irfanview.com/>

Vielen Dank für Ihre Aufmerksamkeit

