siegfried

roy, brunnhilde

2017-06-29, Kiel, Marco Klindt

siegfried

- Siegfried is a signature-based file format identification tool.
- It implements:
- the National Archives UK's PRONOM file format signatures
- freedesktop.org's MIME-info file format signatures
- the Library of Congress's FDD file format signatures (**beta**).

roy

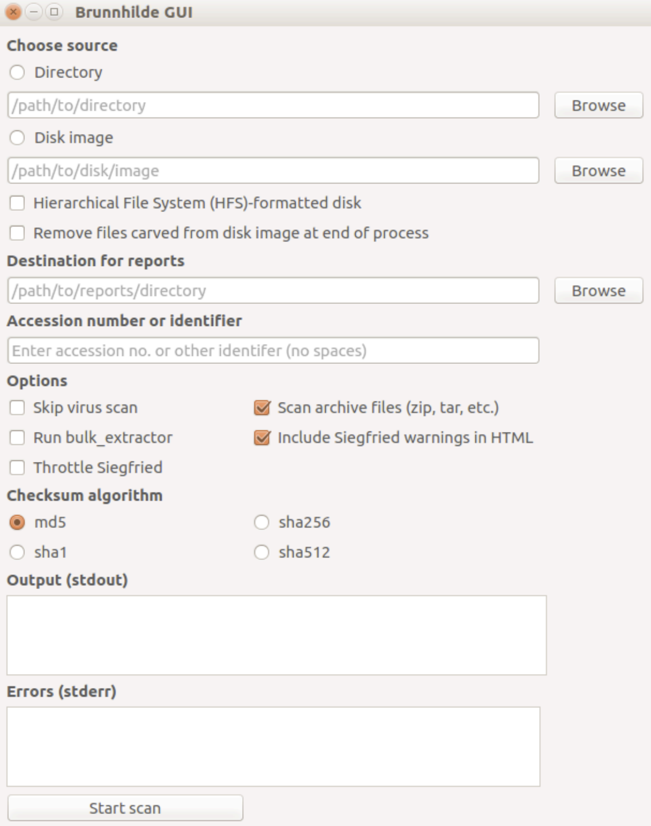
- roy bearbeitet siegfried signature-Dateien
- Kann Signaturen hinzufügen und ausklammern.
- Eigene Sets von Dateiformaten anlegen.

```
"pdfcore":[ "fmt/14 (Acrobat PDF 1.0 - Portable Document Format 1.0)",
 "fmt/15 (Acrobat PDF 1.1 - Portable Document Format 1.1)",
"fmt/16 (Acrobat PDF 1.2 - Portable Document Format 1.2)",
 "fmt/17 (Acrobat PDF 1.3 - Portable Document Format 1.3)",
 "fmt/18 (Acrobat PDF 1.4 - Portable Document Format 1.4)",
"fmt/19 (Acrobat PDF 1.5 - Portable Document Format 1.5)", "fmt/20 (Acrobat PDF 1.6 - Portable Document
Format 1.6)", "fmt/276 (Acrobat PDF 1.7 - Portable Document Format 1.7)" ],

"pdfa":[ "fmt/95", "fmt/354", "fmt/476", "fmt/477", "fmt/478", "fmt/479", "fmt/480", "fmt/481" ], ...
```

brunnhilde

- Report Generator für siegfried



Brunnhilde GUI

Choose source

○ Directory

/path/to/directory

○ Disk image

/path/to/disk/image

□ Hierarchical File System (HFS)-formatted disk

□ Remove files carved from disk image at end of process

Destination for reports

/path/to/reports/directory

Accession number or identifier

Enter accession no. or other identifer (no spaces)

Options

□ Skip virus scan ☑ Scan archive files (zip, tar, etc.)

□ Run bulk_extractor ☑ Include Siegfried warnings in HTML

□ Throttle Siegfried

Checksum algorithm

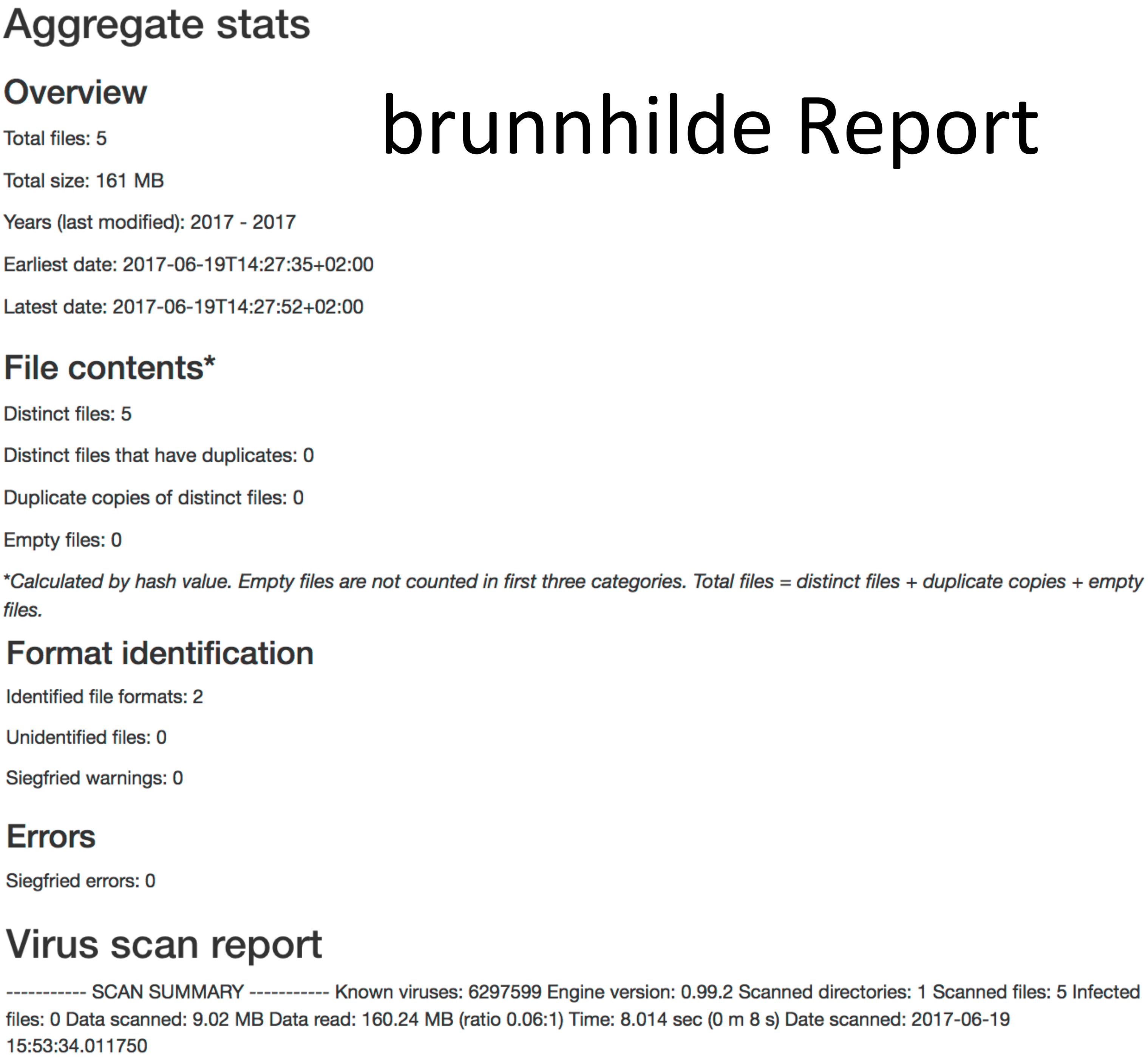
● md5 ○ sha256

○ sha1 ○ sha512

Output (stdout)

Errors (stderr)

brunnhilde Report



Aggregate stats

Overview

Total files: 5

Total size: 161 MB

Years (last modified): 2017 - 2017

Earliest date: 2017-06-19T14:27:35+02:00

Latest date: 2017-06-19T14:27:52+02:00

File contents*

Distinct files: 5

Distinct files that have duplicates: 0

Duplicate copies of distinct files: 0

Empty files: 0

**Calculated by hash value. Empty files are not counted in first three categories. Total files = distinct files + duplicate copies + empty files.*

Format identification

Identified file formats: 2

Unidentified files: 0

Siegfried warnings: 0

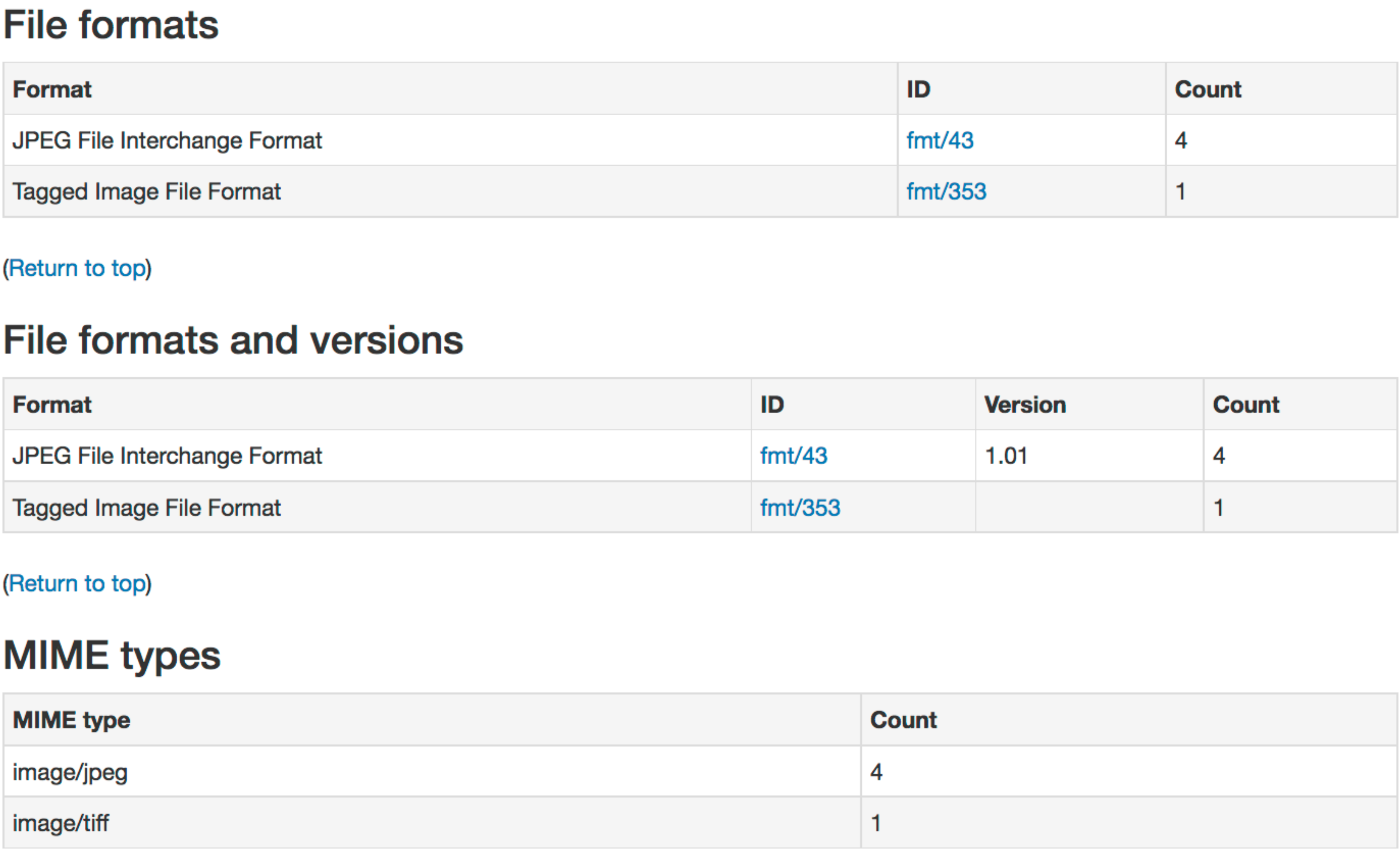
Errors

Siegfried errors: 0

Virus scan report

```
----------- SCAN SUMMARY ----------- Known viruses: 6297599 Engine version: 0.99.2 Scanned directories: 1 Scanned files: 5 Infected files: 0 Data scanned: 9.02 MB Data read: 160.24 MB (ratio 0.06:1) Time: 8.014 sec (0 m 8 s) Date scanned: 2017-06-19 15:53:34.011750
```

brunnhilde Report



File formats

Format	ID	Count
JPEG File Interchange Format	fmt/43	4
Tagged Image File Format	fmt/353	1

(Return to top)

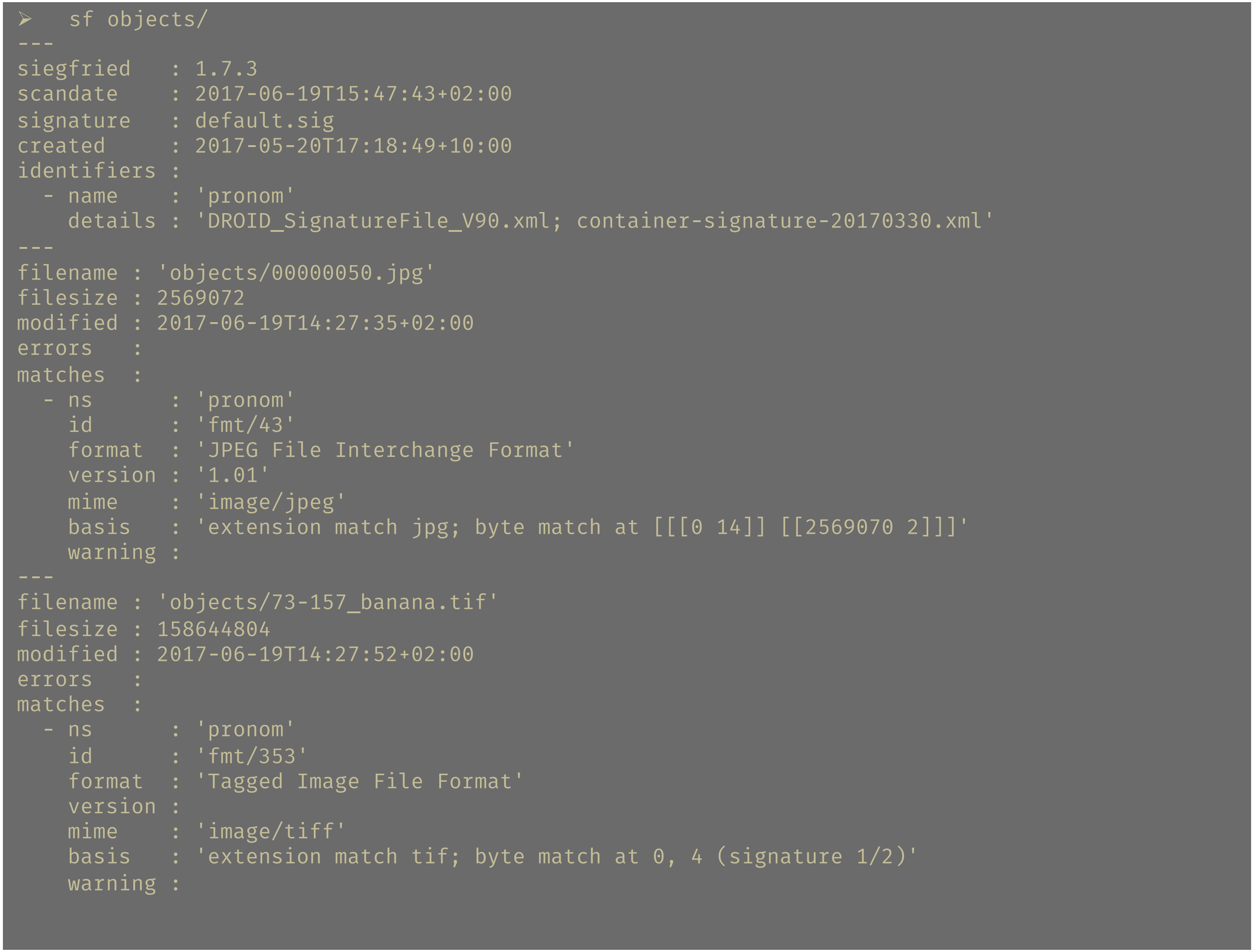
File formats and versions

Format	ID	Version	Count
JPEG File Interchange Format	fmt/43	1.01	4
Tagged Image File Format	fmt/353		1

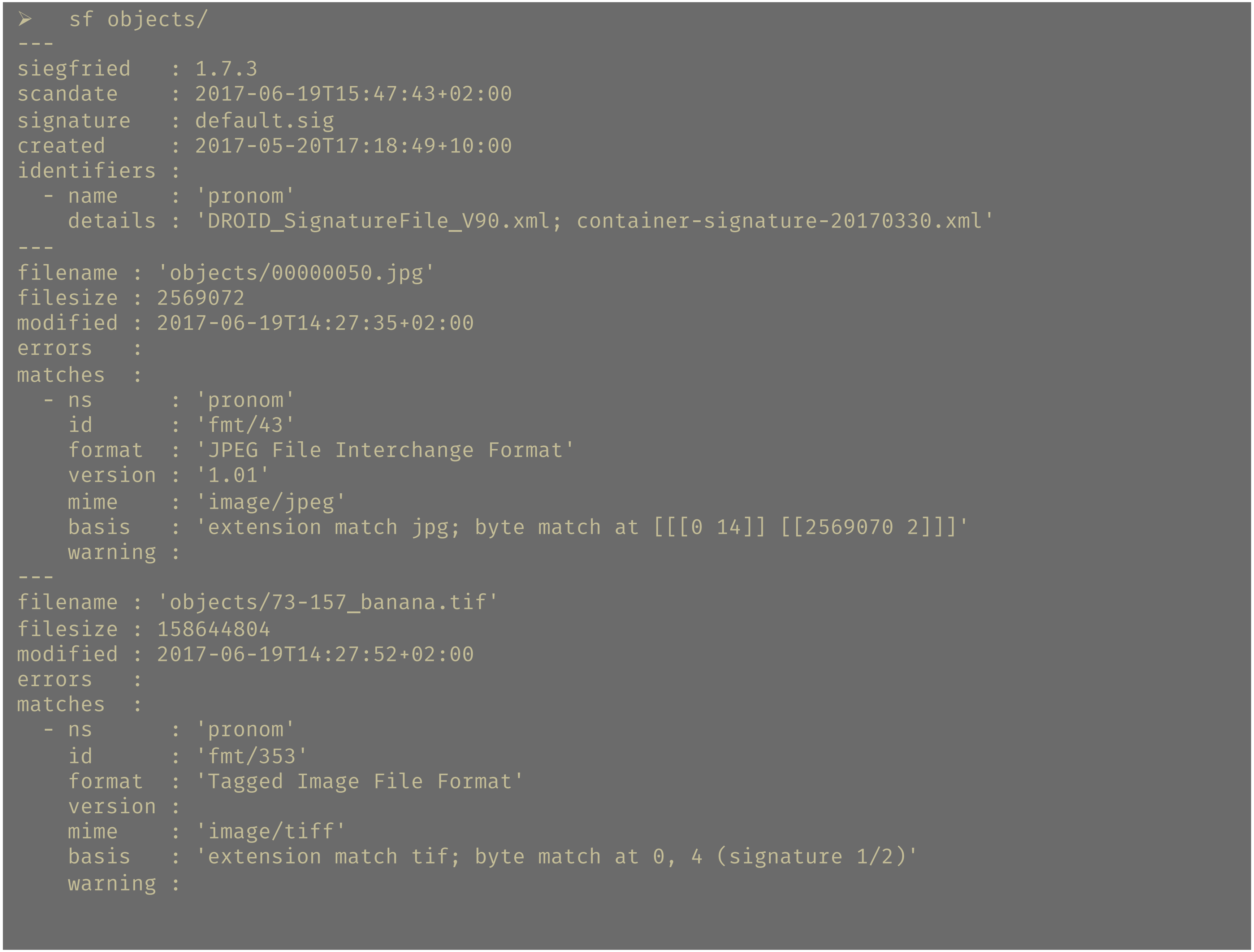
(Return to top)

MIME types

MIME type	Count
image/jpeg	4
image/tiff	1



```
➤  sf objects/
---
siegfried   : 1.7.3
scandate    : 2017-06-19T15:47:43+02:00
signature   : default.sig
created     : 2017-05-20T17:18:49+10:00
identifiers :
  - name    : 'pronom'
    details : 'DROID_SignatureFile_V90.xml; container-signature-20170330.xml'
---
filename : 'objects/00000050.jpg'
filesize : 2569072
modified : 2017-06-19T14:27:35+02:00
errors   :
matches  :
  - ns      : 'pronom'
    id      : 'fmt/43'
    format  : 'JPEG File Interchange Format'
    version : '1.01'
    mime    : 'image/jpeg'
    basis   : 'extension match jpg; byte match at [[[0 14]] [[2569070 2]]]'
    warning :
---
filename : 'objects/73-157_banana.tif'
filesize : 158644804
modified : 2017-06-19T14:27:52+02:00
errors   :
matches  :
  - ns      : 'pronom'
    id      : 'fmt/353'
    format  : 'Tagged Image File Format'
    version :
    mime    : 'image/tiff'
    basis   : 'extension match tif; byte match at 0, 4 (signature 1/2)'
    warning :
```



```
➤   sf objects/
---
siegfried   : 1.7.3
scandate    : 2017-06-19T15:47:43+02:00
signature   : default.sig
created     : 2017-05-20T17:18:49+10:00
identifiers :
  - name    : 'pronom'
    details : 'DROID_SignatureFile_V90.xml; container-signature-20170330.xml'
---
filename : 'objects/00000050.jpg'
filesize : 2569072
modified : 2017-06-19T14:27:35+02:00
errors   :
matches  :
  - ns      : 'pronom'
    id      : 'fmt/43'
    format  : 'JPEG File Interchange Format'
    version : '1.01'
    mime    : 'image/jpeg'
    basis   : 'extension match jpg; byte match at [[[0 14]] [[2569070 2]]]'
    warning :
---
filename : 'objects/73-157_banana.tif'
filesize : 158644804
modified : 2017-06-19T14:27:52+02:00
errors   :
matches  :
  - ns      : 'pronom'
    id      : 'fmt/353'
    format  : 'Tagged Image File Format'
    version :
    mime    : 'image/tiff'
    basis   : 'extension match tif; byte match at 0, 4 (signature 1/2)'
    warning :
```

Siegfried-Ausgabeformate

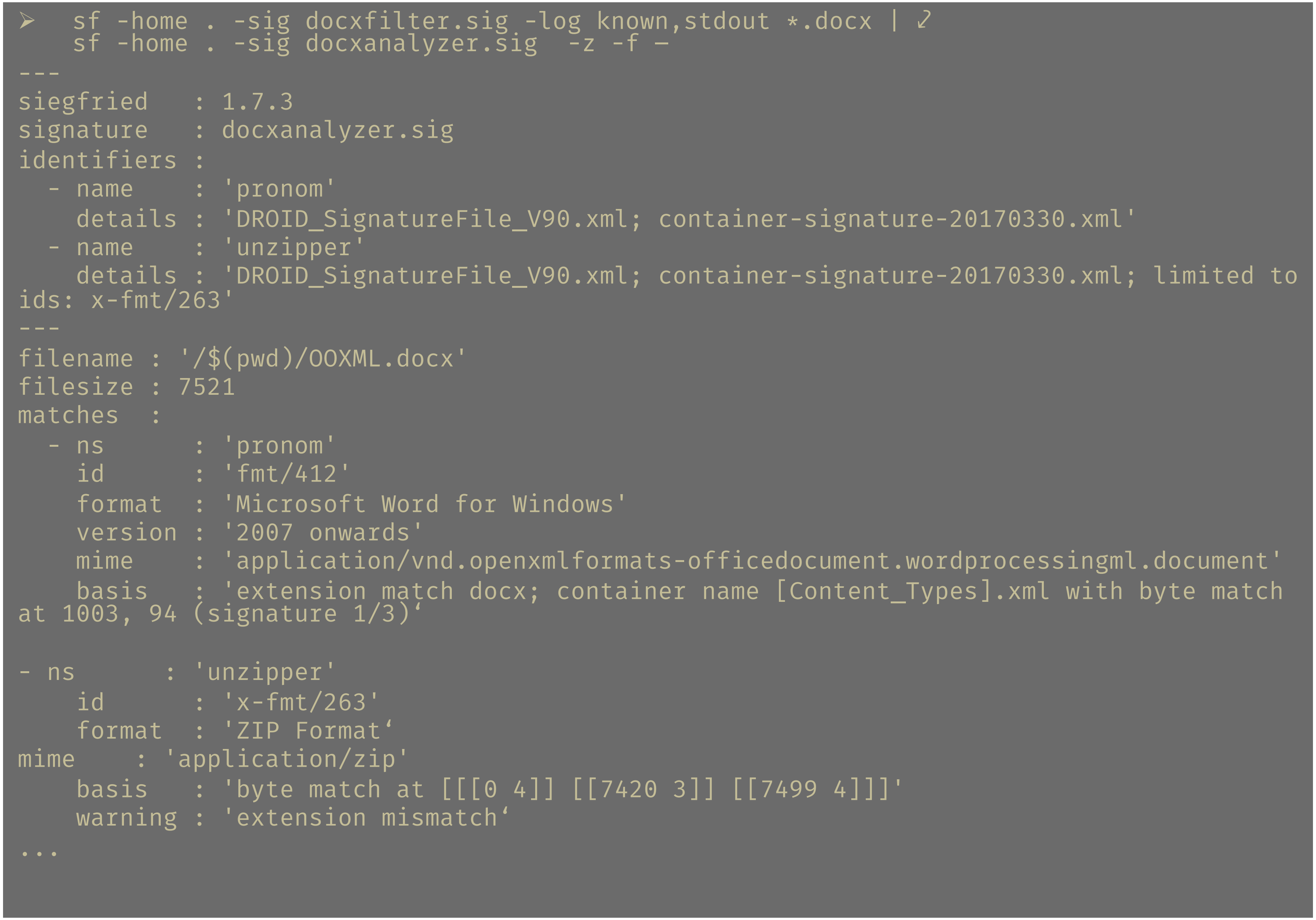
- YAML
- CSV
- JSON
- DROID-CSV

- Kann auch Dateilisten (-dateien) verarbeiten.

DOCX Analyse mit Siegfried und Roy

```
➢  roy harvest -home .
➢  roy build -limit fmt/412,fmt/494 -home . –name ↲
   docxfilter docxfilter.sig
➢  roy build -home docxanalyzer.sig
➢  roy inspect -home . docxanalyzer.sig
Identifiers
Name: pronom
Details: DROID_SignatureFile_V90.xml; container-
signature-20170330.xml
Number of filename signatures: 1780
Number of MIME signatures: 566
Number of container signatures: 146
Number of XML signatures: 0
Number of byte signatures: 1223
Number of RIFF signatures: 0
Number of text signatures: 1
```

```
➢  roy add -home . -limit x-fmt/263 -name unzipper $(pwd)/
   docxanalyzer.sig
➢  roy inspect -home . docxanalyzer.sig
Identifiers
Name: pronom
Details: DROID_SignatureFile_V90.xml; container-signature-20170330.xml
Number of filename signatures: 1780
Number of MIME signatures: 566
Number of container signatures: 146
Number of XML signatures: 0
Number of byte signatures: 1223
Number of RIFF signatures: 0
Number of text signatures: 1
Name: unzipper
Details: DROID_SignatureFile_V90.xml; containersignature-20170330.xml;
limited to ids: x-fmt/263
Number of filename signatures: 1
Number of MIME signatures: 1
Number of container signatures: 0
Number of XML signatures: 0
Number of byte signatures: 1
Number of RIFF signatures: 0
Number of text signatures: 0
```



```
➤  sf -home . -sig docxfilter.sig -log known,stdout *.docx | ⤦
   sf -home . -sig docxanalyzer.sig  -z -f –
---
siegfried   : 1.7.3
signature   : docxanalyzer.sig
identifiers :
  - name    : 'pronom'
    details : 'DROID_SignatureFile_V90.xml; container-signature-20170330.xml'
  - name    : 'unzipper'
    details : 'DROID_SignatureFile_V90.xml; container-signature-20170330.xml; limited to
ids: x-fmt/263'
---
filename : '/$(pwd)/OOXML.docx'
filesize : 7521
matches  :
  - ns      : 'pronom'
    id      : 'fmt/412'
    format  : 'Microsoft Word for Windows'
    version : '2007 onwards'
    mime    : 'application/vnd.openxmlformats-officedocument.wordprocessingml.document'
    basis   : 'extension match docx; container name [Content_Types].xml with byte match
at 1003, 94 (signature 1/3)‘

- ns      : 'unzipper'
    id      : 'x-fmt/263'
    format  : 'ZIP Format‘
mime    : 'application/zip'
    basis   : 'byte match at [[[0 4]] [[7420 3]] [[7499 4]]]'
    warning : 'extension mismatch‘
...
```

```
...

---
filename : '$(pwd)/OOXML.docx#[Content_Types].xml'
filesize : 1489
- ns      : 'pronom'
    id      : 'fmt/101'
    format  : 'Extensible Markup Language'
    version : '1.0'
    mime    : 'application/xml'
    basis   : 'extension match xml; byte match at 0, 19’
- ns      : 'unzipper'
    id      : 'UNKNOWN’
warning : 'no match’
---
filename : '$(pwd)/OOXML.docx#docProps/thumbnail.jpeg'
filesize : 251
- ns      : 'pronom'
    id      : 'fmt/41'
    format  : 'Raw JPEG Stream'
    version :
    mime    : 'image/jpeg'
    basis   : 'extension match jpeg; byte match at [[[0 3]] [[249 2]]] (signature 1/2)'
    warning :
  - ns      : 'unzipper'
    id      : 'UNKNOWN’
warning : 'no match’
```

Links

https://www.itforarchivists.com/siegfried

https://github.com/timothyryanwalsh/brunnhilde

https://github.com/timothyryanwalsh/brunnhilde-gui