

Tool demo: making sense of a collection

Paul Wheatley

paul.wheatley@dpconline.org

[@prwheatley](#)

www.dpconline.org

Scale = automation = software tools

- Digital typically means working at scale
- Automation becomes a necessity
- Software tools
- Techie stuff
- But...
- Hands on experience





Digital **Preservation** Coalition

```
10101010101111010001010101000
10010100101110100101010101001
00100101010100100000101010101
01111101000101010100010010100
10111010010101010100100100101
01010010000010101010101111101
00010101010001001010010111010
01010101010010010010101010010
00001010101010110101010111101
00010101010001001010010111010
01010101010010010010101010010
00001010101010111110100010101
01000100101001011101001010101
01001001001010101001000001010
10101011111010001010101000100
1010010111010010101010100100...
```

```
SOI
APP0 JFIF
1.2
APP13 IPTC
APP2 ICC
DQT
SOF0 200x392
DRI
DHT
SOS
ECS0
RST0
ECS1
RST1
ECS2...
```





Digital**Preservation**Coalition

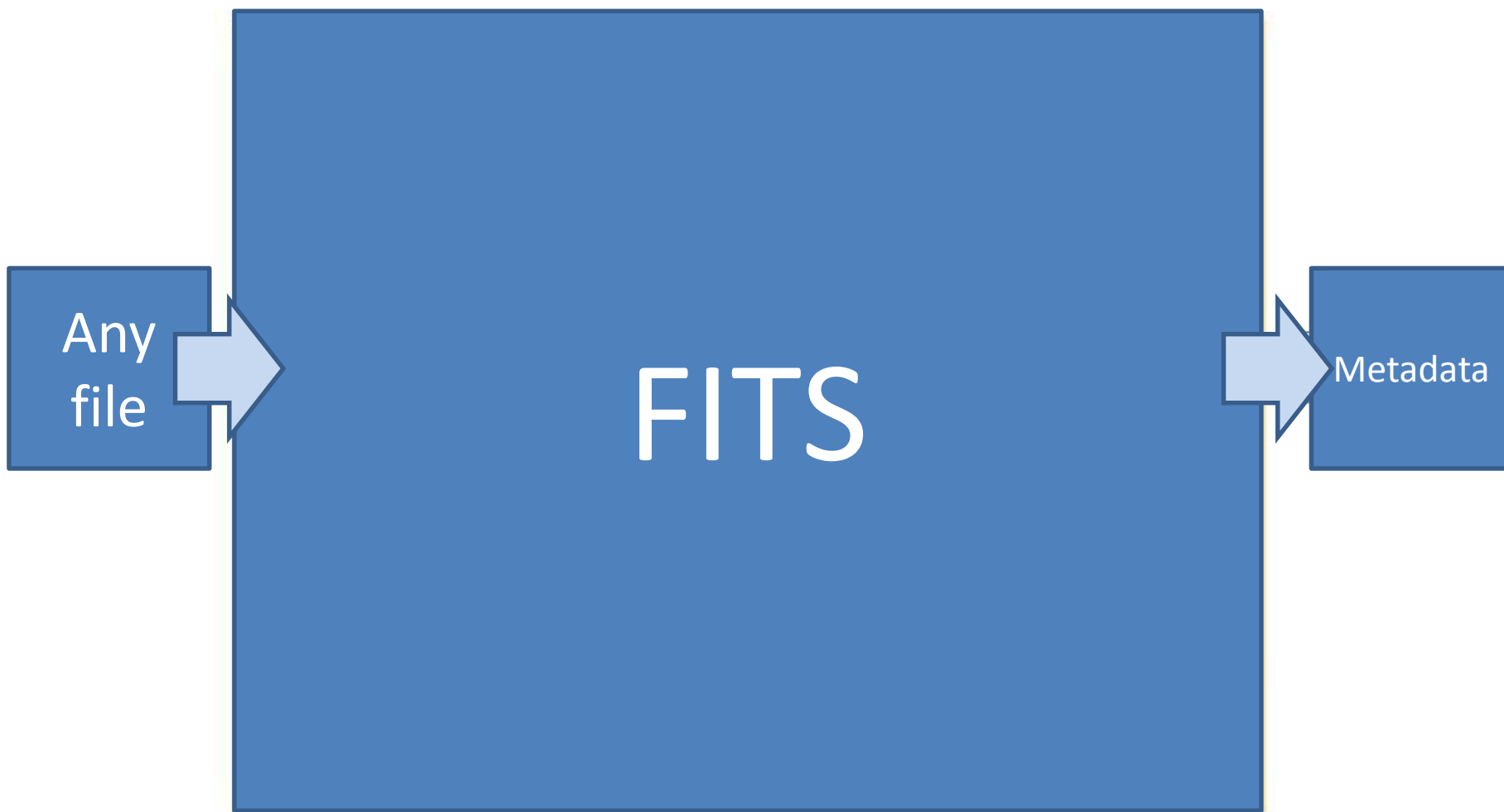
Making sense of a collection

- Significant challenge, but fundamental to DP
- Understand the data, then assess risks, plan, take action to preserve
- Characterisation:
 - How many files?
 - How big are the files?
 - What file formats?
 - Is the data dynamic or interactive?
 - Does it contain personal information?
 - Is it encrypted?

File formats: what flavour is your data?

- What are they likely to contain?
- What are the likely preservation risks associated with that file format?
- What software do I need to render, interpret or make sense of that file format?
- Caveat: Fuzzy concept, doesn't give all the answers
 - [Abusing file formats](#) by Ange Albertini (chapter 6)

FITS – File Information Tool Set



Paul Wheatley

paul.wheatley@dpconline.org

[@prwheatley](#)

www.dpconline.org

- [FITS](#): Content profiling tool
- Works with FITS to visualize FITS output metadata
- Setup can be a little more taxing



The background of the slide is a dark, atmospheric photograph. It shows a silhouette of a mountain range in the foreground, with a sky filled with wispy, light-colored clouds. The overall color palette is dark, with shades of blue, grey, and black.

TRUST NO ONE

Assume nothing, validate everything

Stuff happens

- Whenever a digital collection is moved, processed, curated or altered in any way.... things can go wrong!
 - Network dropouts at critical times
 - Disks get full, subsequent data copied there is lost
 - Software bugs lead to unexpected results
 - Human error leads to all sorts of issues
- Stuff happens a lot more at scale!