IF A CHECKSUM FALLS IN A FOREST...

Elizabeth England
@elizabeengland
Digital Archivist | Johns Hopkins University
KEY TERMS

• Fixity – object is unchanged since an earlier point in time [1]
• Integrity – object is complete, unaltered in all essential respects [2]
• Authenticity – “trustworthiness”

Hi Elizabeth,

As we discussed yesterday, I was wondering if these bags are the same since they have the same byte count. Now that I’ve got them staged back, I calculated their checksums and they do not match. So I now know they are not the same – or at least not exactly the same. Looking closer, each bag contains the same number of files. Comparing files, all the checksums match except for the files listed below. At least some of these are just metadata files and they don’t match by design (e.g. they contain different filenames). Maybe it’s the same data, just processed two different ways?
SIMILARITIES

• 558 GB
• Expert Witness Format (.EWF) disk images, Guymager sub-format
• Partitioned into 2047 MB chunks
• 274 matching checksums

DIFFERENCES

• 7 checksums
• 3 BagIt files
• 4 disk image files
FILES WITH DISCREPANCIES

```
bag/
|-- data
| |-- 001.E01
| |-- 001.EGA
| |-- 001.EGY
| |-- 001.info
|-- manifest-md5.txt
|-- bag-info.txt
\-- tagmanifest-md5.txt
```

Disk image files

BagIt files
INFO FILES

bag/
  |-- data
  |   |-- 001.E01
  |   |-- 001.EGA
  |   |-- 001.EGY
  |   |-- 001.info
  |-- manifest-md5.txt
  |-- bag-info.txt
  \-- tagmanifest-md5.txt
MANIFEST FILES

bag/
|-- data
  |-- 001.E01
  |-- 001.EGA
  |-- 001.EGY
  \-- 001.info
|-- manifest-md5.txt MD5 values for data directory files
|-- bag-info.txt
\-- tagmanifest-md5.txt MD5 values for BagIt generated files
bag/
|   ├── data
|   |   ├── 001.E01
|   |   ├── 001.EGA
|   |   ├── 001.EGY
|   |   └── 001.info
|   └── manifest-md5.txt
|   └── bag-info.txt
|   └── tagmanifest-md5.txt

EWF DISK IMAGE FILES
EWF HEADER FILE

bag/
|-- data
| |-- 001.E01 header
| |-- 001.EGA
| |-- 001.EGY
| |-- 001.info
|-- manifest-md5.txt
|-- bag-info.txt
\-- tagmanifest-md5.txt
bag/
|-- data
| |-- 001.E01
| | |-- 001.EGA segment
| | |-- 001.EGY
| |-- 001.info
|-- manifest-md5.txt
|-- bag-info.txt
|-- tagmanifest-md5.txt
EWF FOOTER FILE

dir://
|-- data
| `-- 001.E01   header
| `-- 001.EGA   segment
| `-- 001.EGY   footer – MD5 value for the segments
| `-- 001.info
|-- manifest-md5.txt
|-- bag-info.txt
`-- tagmanifest-md5.txt
EWF SEGMENT FILE

bag/
|-- data
| |-- 001.E01
| |-- 001.EGA
| |-- 001.EGY
| |-- 001.info
|-- manifest-md5.txt
|-- bag-info.txt
\-- tagmanifest-md5.txt
.EGA SEGMENT FILE

EGA version 1

EGA version 2

EWF file signature
THANK YOU!

Elizabeth England
twitter: @elizabeengland
email: eengland@jhu.edu