

WHAT I WISH I KNEW...

Kieran O'Leary - Irish Film Institute

DPC/DRI - Dublin 2019-02-08

@kieranjol



Irish Film Institute

Format Identification via command line

```
kieran110s@kieran110s-Lenovo-ideapad-110S-11IBR: ~/ifigit/ifiscripts/rawcooked_examp
amples$ sf ~/Pictures/Screenshot\ from\ 2019-02-07\ 21-48-45.png
---
siegfried      : 1.7.8
scandate       : 2019-02-07T21:57:44Z
signature      : default.sig
created        : 2017-12-02T14:49:15+11:00
identifiers    :
  - name       : 'pronom'
    details    : 'DR0ID_SignatureFile_V93.xml; container-signature-20171130.xml'
---
filename       : '/home/kieran110s/Pictures/Screenshot from 2019-02-07 21-48-45.png'
filesize       : 69142
modified       : 2019-02-07T21:48:48Z
errors         :
matches        :
  - ns         : 'pronom'
    id         : 'fmt/11'
    format     : 'Portable Network Graphics'
    version    : '1.0'
    mime       : 'image/png'
    basis      : 'extension match png; byte match at [[[0 16]] [[69130 12]]]'
    warning    :
kieran110s@kieran110s-Lenovo-ideapad-110S-11IBR:~/ifigit/ifiscripts/rawcooked_ex
amples$
```



- [UW Audio Preservation and Restoration Guide](#): Large aggregation of Audio resources with descriptions.

Scripts

- [Carnegie Hall](#): Post-digitization quality control workflow documentation and scripts; scripts and overview documentation for generating linked data from event records; and scripts and documentation detailing workflow for batch linking metadata records from disparate CSVs to files.
- [CUNY mediamicroservices](#): Suite of bash scripts for A/V archiving.
- [FFmpeg Cookbook for Archivists](#)
- [ffmprovisr](#): Repository of useful FFmpeg command lines for archivists!
- [IFIs scripts](#): Scripts for video/DCP/image sequences/fixity for use in the IFI Irish Film Archive.
- [National Library of New Zealand \(NLNZ\)](#): Scripts for generating METS records, Rosetta-compliant SIPs, etc.
- [New York Times video microservices](#): Documentation and links to services scripts.
- [NYPL ami-tools](#): Python3 scripts and classes to help with managing bags of NYPL AMI files.
- [Puget Sound and Vision Audio Scripts](#): Bash scripts/documentation used at University of Washington for moving from a digitized WAV file to a package including BWF, TIFFs and derivatives.
- [Rockefeller Archive Center Transfer Workflow](#): Digital media inventory and transfer workflow documentation.
- [ucsb-src-microservices](#): Post-processing scripts we use at UCSB Special Research Collections AVLab.
- [U.S. National Archives](#)
- [WGBH preservation-workflows](#): Useful scripts to help automate common WGBH Media Library and Archive digital preservation workflows into a single preserve.rb Ruby script.
- [Yale University Libraries](#): Manuscripts and Archives GitHub.

dave rice

archivist | technologist

HOME

ABOUT

IMAGES

PAPERS

Reconsidering the Checksum for Audiovisual Preservation

This article was authored by Dave Rice and initially published in the [IASA Journal number 29](#) under a [CC-BY-ND](#) license.

Reconsidering the Checksum for Audiovisual Preservation:

Detecting digital change in audiovisual data with decoders and checksums

Firstly, what are checksums for?

A checksum is small data value computed from a given amount of data, such as a file or bit-stream, for the purpose of facilitating the future ability to detect changes in that given data. The generation and verification of checksums for digital archival holdings is a central principle of digital preservation and enable archivists to trust that data held within an archive is the same data that was received by the archive. Although checksum wrangling is typically a

RECENT

[Innovation and
#amia15: The N
Digitization Soft
published in the](#)



Notifi



Kieran O'Leary

@kieranjol



@dericed Great article on framemd5. Is there a way to verify/compare md5 on a per frame basis not just by eye? Is there a way to flag errors?

4:39 PM - 5 May 2015

5 Likes



1



5



Add another Tweet



Kieran O'Leary @kieranjol · 5 May 2015



Replying to @kieranjol

@dericed Can ffmpeg tell me if the md5 is not matching or must I scan the outputs by eye? Thanks, Kieran.



1



Dave Rice @dericed · 5 May 2015



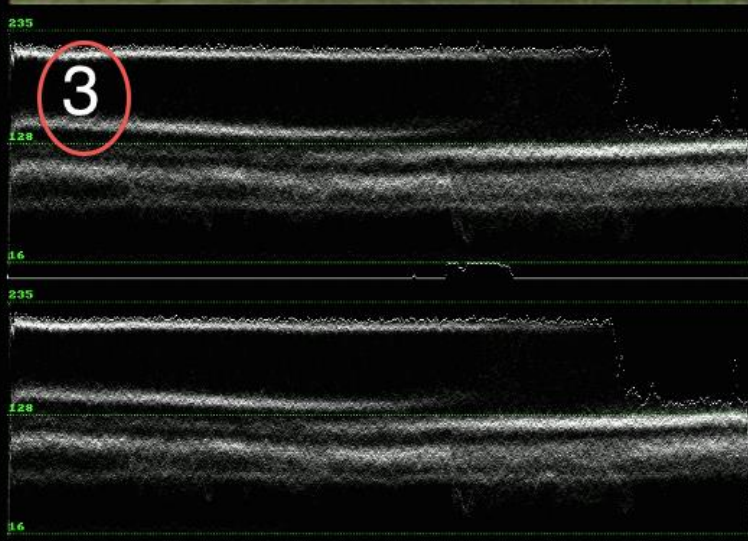
@kieranjol do you have an existing framemd5 for the file? Usually I store framemd5, then can later diff against a new framemd5 to find probs



1



```
92     if [ $(grep -v "^#" "${LOGDIR}/${SOURCEFILENAME%.*}.framemd5" | md5 -q) = $(grep -v "^#" "${LOGDIR}/${SOURCEFILENAME%.*}.framemd5" | md5 -q)
93         _report -dt "Everything looks safe. Going to delete the original."
94         _run_critical mediainfo -f --language=raw --output=XML "${SOURCEFILE}" > "${LOGDIR}/${SOURCEFILENAME%.*}_mediainfo.xml"
95         _run_critical rm -f -v "${SOURCEFILE}"
96     else
97         _report -wt "Not looking safe. Going to keep the original."
98     fi
```




```
macminis-mini:~ macmini$ tree /Volumes/pegasus2/kieran_tests/aaa0996
/Volumes/pegasus2/kieran_tests/aaa0996
├── 9d21569e-afde-401b-a57e-e3f2b32bcef1
│   ├── logs
│   │   ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_framemd5.log
│   │   ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_normalise.log
│   │   └── 9d21569e-afde-401b-a57e-e3f2b32bcef1_sip_log.log
│   ├── metadata
│   │   ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1.framemd5
│   │   ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1.mkv_medainfo.xml
│   │   ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1.mkv_mediatrace.xml
│   │   ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_dfxml.xml
│   │   ├── AD607_filmographic.csv
│   │   ├── aaa0996_pbcore.csv
│   │   ├── bf6929ee-018a-46bc-beed-12d5330aec17.mov_source.framemd5
│   │   └── supplemental
│   │       ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_source_dfxml.xml
│   │       ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_source_medainfo.xml
│   │       ├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_source_mediatrace.xml
│   │       └── bf6929ee-018a-46bc-beed-12d5330aec17.mov.qctools.xml.gz
│   └── objects
│       └── 9d21569e-afde-401b-a57e-e3f2b32bcef1.mkv
├── 9d21569e-afde-401b-a57e-e3f2b32bcef1_manifest-sha512.txt
└── 9d21569e-afde-401b-a57e-e3f2b32bcef1_manifest.md5
```

5 directories, 17 files

Some resources

- COPTR - http://coptr.digipres.org/Main_Page
- Open workflows - <https://github.com/amiaopensource/open-workflows>
- CRALS - <https://dd388.github.io/crals/>
- Ashley Blewer - <https://bits.ashleyblewer.com/blog/2014/11/04/non-technical-persons-guide-to-becoming-an-open-source-software-contributor-via-github/>
- IFIscripts - <https://ifiscripts.readthedocs.io/en/latest/index.html>
- My blog: <https://kieranjol.wordpress.com/>
- Dave Rice - Framemd5 blog: <http://dericed.com/papers/reconsidering-the-checksum-for-audiovisual-preservation/>