

**Digital Preservation Resources
Community AV Archiving Fair (AV Fair)
Portland, OR, December 1, 2018**

Contact Info:

Sandra Yates, CA, DAS
Archivist & Special Collections Librarian
Texas Medical Center Library, McGovern Historical Center
saneyates@gmail.com

Digital Preservation Suggested Readings

Preservation Storage Criteria, Version 2, May 2017

The Preservation and Archiving Special Interest Group (PASIG), <https://preservationandarchivingsig.org>
[Available here for a limited time](#)

AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship, 2012

https://dcs.library.virginia.edu/files/2013/02/AIMS_final_text.pdf

Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read
In-house

Julianna Barrera-Gomez, Ricky Erway, 2013

<https://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf>

Born Digital: Guidance for Donors, Dealers, and Archival Repositories

Gabriela Redwine, 2013

<https://www.clir.org/wp-content/uploads/sites/9/pub159.pdf>

Managing Electronic Records in Manuscript Collections: A Case Study from the Beinecke Rare
Book and Manuscript Library

Michael Forstrom, 2009

<http://americanarchivist.org/doi/pdf/10.17723/aarc.72.2.b82533tvr7713471>

From Accession to Access: A Born-Digital Materials Case Study

Cyndi Shein, 2014

<https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1036&context=westernarchives>

Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in
Manuscript Repositories

Ben Goldman, 2011

<https://rbm.acrl.org/index.php/rbm/article/viewFile/343/343>

The Open Archival Information System Reference Model: Introductory Guide

Brian F. Lavoie, 2004

<https://www.dpconline.org/docs/technology-watch-reports/91-introduction-to-oais/file>

Capturing and Processing Born-Digital Files in the STOP AIDS Project Records: A Case Study

Laura Wilsey, 2013

<https://digitalcommons.usu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1026&context=westernarchives>

Digital Preservation Workstation

Building a Digital Curation Workstation with BitCurator (update)

kamwoods, 2013

<https://bitcurator.net/2013/08/02/building-a-digital-curation-workstation-with-bitcurator-update/>

Digital Archaeology: Recovering your Digital History by Doug Reside, Curator, Billy Rose

Theatre Division, New York Public Library for the Performing Arts, Dorothy and Lewis B.

Cullman Center, July 23, 2012

<https://www.nypl.org/blog/2012/07/23/digital-archaeology-recovering-your-digital-history>

How to Build a Computer, Lesson 1: Hardware Basics

Whitson Gordon, 2011

<https://lifelife.com/5826509/how-to-build-a-computer-from-scratch-lesson-1-hardware-basics>

Digital Preservation Programs and Software

BitCurator

The BitCurator NLP project is developing software for collecting institutions to extract, analyze, and produce reports on features of interest in text extracted from born-digital materials contained in collections. We are using open source natural language processing libraries to identify items likely to be relevant to preservation, information organization, and access activities. These may include entities (e.g. persons, places, and organizations), potential relationships among entities (e.g. those entities that appear together within documents or set of documents), and topic models to provide insight into how concepts are naturally clustered within the documents. We are developing software that will allow users to create customized reports from text discovered in disk images, providing both command-line executables and a public Python API to extend the capabilities of external tools.

<https://bitcurator.net/>

<https://github.com/bitcurator/bitcurator-nlp/wiki>

Exiftool

Exiftool is an open-source program that allows you to read, write and edit meta information. It is a command-line tool, so you will need access to the Terminal program in your operating system. Exiftool produces a list of files and embedded exif metadata within media files. Exif metadata provides important information about the creation of files, especially media specifications like camera used, time, duration, etc.

<https://www.sno.phy.queensu.ca/~phil/exiftool/>

Data Accessioner

The DataAccessioner was built out of the need for a simple GUI interface to allow the Duke University Rare Book, Manuscripts, & Special Collections Library (now the David M. Rubenstein Rare Book & Manuscript Library) Technical Services staff an easy way of migrating data off disks and onto a file server for basic preservation, further appraisal, arrangement, & description. It also provides a way to integrate common metadata tools at the time of migration rather than after the fact. With a simplified interface and being written in Java it is intended to be easily adopted by smaller institutions with little or no IT staff support.

<http://dataaccessioner.org/>

File Analyzer

The NARA File Analyzer Tool walks a directory tree and performs a "File Test" on each file that is encountered. A File Test is a simple set of actions that are performed upon a single file such as filename validation, file size statistical analysis, checksum calculation, file type extraction. Depending on the action, the content of the file may or may not be read. The File Analyzer tool can be run as a GUI in which the results are displayed in a table. The File Analyzer can also be run in batch mode. In batch mode, the results will be written to a tab-separated file. The GUI version of the application allows the results of multiple executions to be merged. The merged information can be filtered to display matching values and mismatched values.

<https://github.com/usnationalarchives/File-Analyzer>

DROID

DROID is an open-source program, developed by The National Archives to perform automated batch identification of file formats. It runs in a Java environment, so it is platform-independent. It can run on Windows, Mac, and Linux machines, using the same instructions. DROID produces a list files in a source directory, providing important information about the types of files and file counts within a file system.

<http://digital-preservation.github.io/droid/>

Handbrake

HandBrake is an open-source video transcoder available for Linux, Mac, and Windows, licensed under the GNU General Public License (GPL) Version 2. HandBrake takes videos you already have and makes new ones that work on your mobile phone, tablet, TV media player, game console, computer, or web browser—nearly anything that supports modern video formats. HandBrake works with most common video files and formats, including ones created by

consumer and professional video cameras, mobile devices such as phones and tablets, game and computer screen recordings, and DVD and Blu-ray discs. HandBrake leverages tools such as FFmpeg, x264, and x265 to create new MP4 or MKV video files from these Sources.

<https://handbrake.fr/>

Digital Preservation Hardware Options

FC5025 controller card for 5.25" floppy

Device Side Data's FC5025 USB 5.25" floppy controller plugs into any computer's USB port and enables you to attach a 5.25" floppy drive. Even if your computer has no built-in floppy controller, the FC5025 lets you read those old disks. And it's not just for IBM PC disks – it also understands formats used by Apple, Atari, Commodore and TI, among others.

<http://www.deviceside.com/>

<http://www.deviceside.com/fc5025.html>

WiebeTech Forensic UltraDock v5 (Write Blocker)

The Forensic UltraDock™ is a professional drive write blocker that provides fast forensically-sound access to bare hard drives. Digital investigators, IT managers, and technicians rely on the FUD's simple and easy to use interface to study or inspect a drive.

https://www.cru-inc.com/products/wiebetech/wiebetech_forensic_ultradock_v5/

How to Create Directory List from Hard Drive

The following instructions will help you print a list of media files and embedded metadata from an external hard drive using three different open-source programs..

Exiftool

Exiftool is an open-source program that allows you to read, write and edit meta information. It is a command-line tool, so you will need access to the Terminal program in your operating system. Exiftool produces a list of files and embedded exif metadata within media files. Exif metadata provides important information about the creation of files, especially media specifications like camera used, time, duration, etc.

Directions are based on Linux-based Ubuntu system within a Virtualbox. You might need to modify according to your operating system. Windows machines cannot access external drives, so you will need to transfer the files onto your computer to run the command.

1. Download and install instructions:
<https://www.sno.phy.queensu.ca/~phil/exiftool/install.html>
2. Connect drive to USB port. Procedures will be very similar using a terminal in a Mac (skip the VirtualBox step).
3. In VirtualBox, Devices > USB to mount drive in Ubuntu
4. Open the terminal and change directory to external drive and enter exiftool command:
 - a. `cd PATH/TO/EXTERNAL/DRIVE`
 - b. `exiftool -common -createdate -duration -mimetype -recurse -csv directoryname > /path/to/project-folder/out.csv`
5. Embedded metadata from all media files on the hard drive will be exported to a CSV file

Example:

```
cd /media/sandra/  
exiftool -common -createdate -duration -mimetype -recurse -csv  
dir > /home/sandra/project/out.csv
```

How to find a directory path?

Windows: In Windows Explorer, select the directory that you want. Right-click and select Properties. Select the Location tab to see the directory path.

Mac: In Finder, select the directory that you want. Control-click and select Get Info.

Linux: In file system, select the directory that you want. Right-click and select Properties.

Copy/Paste/Write path and refer to the correct path using the command above.

DROID

DROID is an open-source program, developed by The National Archives to perform automated batch identification of file formats. It runs in a Java environment, so it is platform-independent. It can run on Windows, Mac, and Linux machines, using the same instructions. DROID produces a list files in a source directory, providing important information about the types of files and file counts within a file system.

1. Download and install instructions: <http://digital-preservation.github.io/droid/>
2. Open DROID (filename and location will be something like this, \droid-binary-6.4-bin/droid-ui-6.4.jar)
3. Add new profile, click on green + sign
4. Navigate to your source directory the directory that you would like to list files from. Click OK.
5. Click Start
6. Export Details
 - a. Once the analysis is complete, you should be able to expand the file structure within the profile window
 - b. Click Export icon at the top of the screen
 - c. Select the profile that you want to export (usually, Untitled-1), Select One Row per file, and Click Export profiles....
 - d. Select your project directory where you would like to save output files, name the file (projectname-DROID-filelist.csv), and Click Save.
7. You can generate different reports, for example:
 - a. Check profile
 - b. Select report from drop down menu
 - i. File Count and Sizes by file extension
 - c. Click Report on Files...
8. Export into your project directory as projectname_DROID_TypeReport (example: CottonVideo-1_DROID_TypeReport)

FileAnalyzer

FileAnalyzer is an open-source program, developed by the US National Archive and Records Administration (NARA). It runs in a Java environment, so it is platform-independent. It can run on Windows, Mac, and Linux machines, using the same instructions. FileAnalyzer produces a file list from a source directory including file formats and metadata.

1. Open FileAnalyzer (filename and location will be something like this,
C:\File-Analyzer\bin\fileAnalyzer.jar)
2. Complete the following fields under the Criteria tab
 - a. Root Directory to Scan: navigate to the source directory on an external drive or your computer.
 - b. Output Directory: navigate to the project directory where the reports will be saved
 - c. Action to perform: you will perform 1 action:
 - i. Count Files By Type
 - d. Max Items: 10000
 - e. Check [filenames have no periods]
3. Under the Details tab, export the Details directory list
 - a. Click on Export Table
 - b. Save to project directory as filename_Details.txt
4. Under Type tab, export the file count by type list
 - a. Click on Export Table
 - b. Save to project directory as filename_byType.txt
5. Create Excel spreadsheet from text files
 - a. Open excel or alternate spreadsheet program > Import Data Text file
 - i. Or, right-click on file > Open with Excel (or alternate program)
 - b. Confirm data is formatted into columns (might have to selected tab-separated)
 - c. Click OK

Other Exiftool Command Examples

Replace *italics/underline text* with your specific directory path or file name.

#return common embedded info in html form

```
exiftool -common -h [path to directory] > out.html
```

#example: outside directory c:\Users\Sandra\Pictures

```
exiftool -common -h \Users\Sandra\Pictures\2017AstrosParade > out.html
```

#example: within directory c:\Users\Sandra\Pictures

```
exiftool -common -h 2017AstrosParade > out.html
```

#return all embedded info, includes duration for video files

```
exiftool -ee -h 2017AstrosParade > out.html
```

#return xml format

```
exiftool -ee -x 2017AstrosParade > out.xml
```

#return common metadata in csv format

```
exiftool -common -csv [path to directory] > out.csv
```

#return ee metadata in csv format

```
exiftool -ee -csv [path to directory] > out.csv
```

#example of multi-directory and subdirectories, recursively

```
exiftool -ee -recurse -csv Videos > outee.csv
```

#return common metadata in csv format, adding meta info.

```
exiftool -T -common -createdate -duration -mimetype -recurse -csv [path to directory] > out.csv
```

#example in linux, creating file and metadata list from an external hard drive.

```
cd /media/dir
```

#find drive name using the ls command. Keeping your drive name short makes this 10x easier.

```
ls
```

```
exiftool -common -createdate -duration -mimetype -recurse -csv [path to directory] >
```

```
/home/sandra/project/out.csv
```