

File-Level Lab

Preserving Digital Culture

Spring 2018

Some basic file management considerations

- For creators, using a limited set of known, open file formats is recommended
 - For archivists, this can be out of our control
- Standardize filenames
- Make sure your filenames are unique even when removed from their original folders
- Keep track of the file formats in your collection and their respective risks and requirements
- Manage copies/backups

File profiling can help us:

- Assess the contents of large collections of digital files
- Identify file formats in a collection, along with their potential risks
- Assure that your technical environment can support files
- Audit information management policy compliance
- Detect duplicates

Using exiftool for file profiling

First, to view files in a directory, open your command prompt or terminal window.

Type: `$ cd directory`

To list directory contents, type: `$ ls`

For PC users, navigate to your exiftool directory

To view file information, type:

`$ exiftool filename`

Export file information to csv with exiftool

To export metadata for a single file:

```
$ exiftool -csv filename >  
itemprofile.csv
```

To export metadata for all items in a collection:

```
$ exiftool -csv . >  
collectionprofile.csv
```

Try profiling a directory with many different files, such as your downloads folder. What do you notice?

Export file
information to
csv with
exiftool

To export only file names and sizes to a csv, type:

```
$ exiftool -csv -filesize -filetype  
directory > collectionprofile2.csv
```

PRONOM

PRONOM is an online directory of file formats, intended to support to the long-term preservation of electronic records.

It contains information about software products and the file formats that they can read and write.

Originally created by the UK National Archives to support digital file formats in their collection, PRONOM is now an open registry which users can contribute new system and file format information to via an online submission form.

The PRONOM directory can be searched by file format, software, vendor, and PRONOM unique identifiers.