Using simple(ish) tools for appraising electronic records

Presented by Eira Tansey, Digital Archivist/Records Manager
University of Cincinnati

UNIVERSITY OF
Cincinnati

Electronic Records at UC

- Born digital vs. digitized
- Records retention schedules
- Transfer of records
- Legal vs ethical exclusion of data

UNIVERSITY OF
Cincinnati

Sensitive data

- Looking at file metadata versus looking inside files
- To accession or not to accession
- Start with these two questions…
  1. What stuff do you have?
  2. What's lurking in the stuff you have?

Beware the U. Oregon situation

F 🐦 8+ ✉ Comments (6)

Archivists Suspended in Records-Leak Case Won't Return to U. of Oregon

Two archivists at the University of Oregon who had been placed on leave after a cache of confidential records was released to a professor will not be returning to their jobs, *The Oregonian* reports.

University officials had described the release of the records as unlawful. A lawyer for William T. Harbaugh, an economics professor who had received the documents, returned the records to the university in late January. Mr. Harbaugh runs the blog UOMatters and has been a vocal advocate for transparency by Oregon's administration.

An investigation into the incident found that the records had been released before customary processing had been carried out and that they included information protected under federal and state privacy laws.

UNIVERSITY OF
Cincinnati

A brief word on appraisal…

Image via https://www.flickr.com/photos/denverjeffrey/1950409800/in/photostream/

1. What stuff do you have?

- Imagine files in nests upon nests upon nests… that's too much clicking!

- Use a combination of the command line and simple spreadsheets to get a birds' eye view

- List of file names, extensions, dates, size

- Bertram Lyons' excellent tutorial
  - http://bit.ly/1ctsgpF
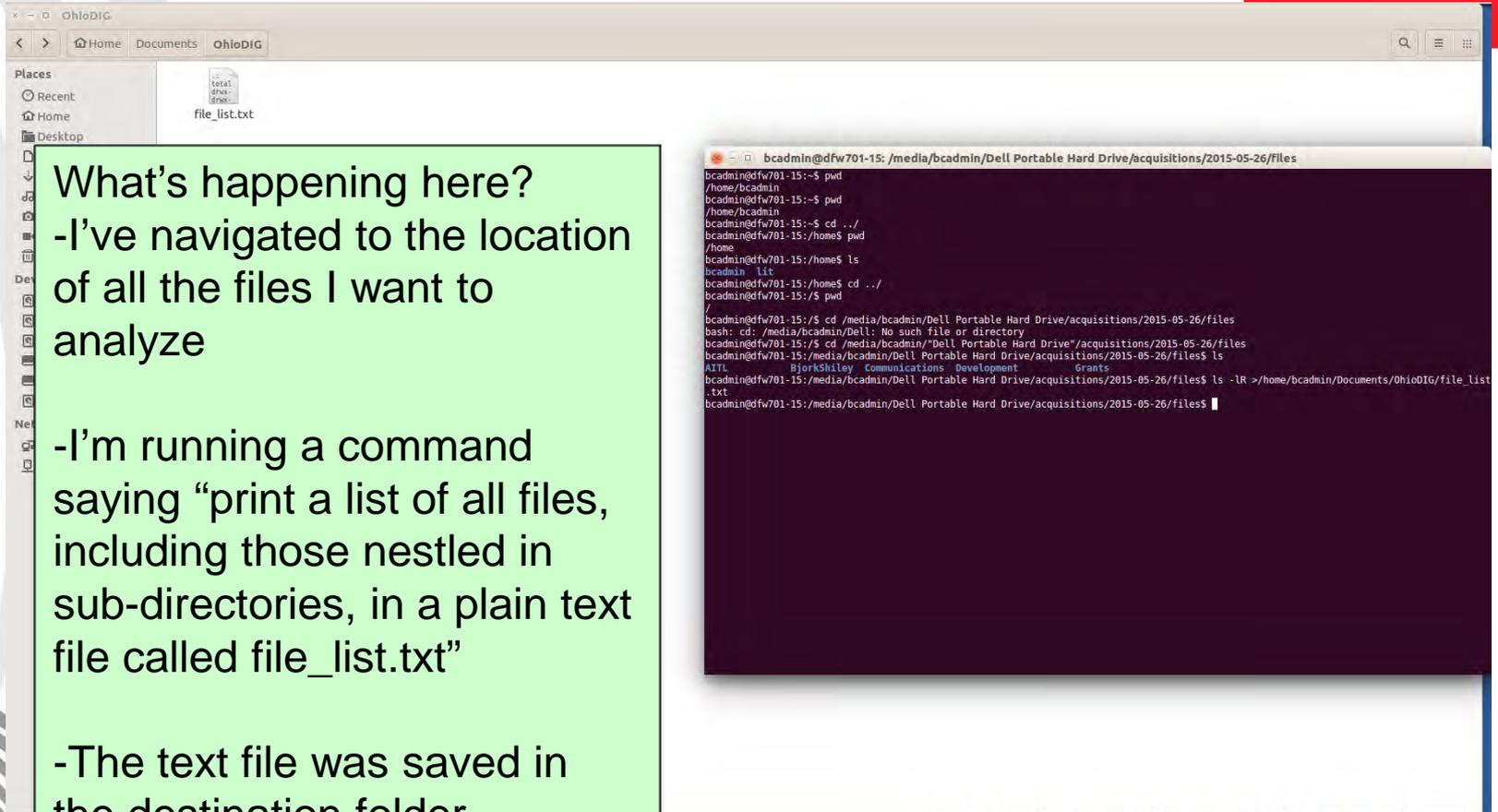
UNIVERSITY OF
Cincinnati

Fire up the terminal!

What's happening here?
-I've navigated to the location of all the files I want to analyze

-I'm running a command saying "print a list of all files, including those nestled in sub-directories, in a plain text file called file_list.txt"

-The text file was saved in the destination folder

UNIVERSITY OF
Cincinnati

Uh……. Why is this useful?

Wait, there's more!

If you don't love spreadsheets yet, now's the time to start. If you're already a true believer, do not allow others to spreadsheet-shame you.

We can take the plain text file, and transform it into something that is far more useful for appraisal.

UNIVERSITY OF
Cincinnati

Cleaning up the file

What's happening here?

If you look at the columns on the right, you see that spaces in the file names (BOO! HISS!) broke apart into several columns

After cleaning…

The file names have been rejoined using the Excel concatenate function

Cincinnati

And eventually…

The column with the numbers represents bytes, so you can determine the GB or MB of your files.

In addition, you can run a function to count the number of unique file extensions

2. What's lurking in the stuff you have?

- BitCurator is a package of open-source digital forensics tools

- Digital forensics originated in the law enforcement community to deliver legally-admissible digital evidence

- May not want to forensically image everything, but forensics tools can be helpful for other tasks

UNIVERSITY OF
Cincinnati

Diving Inside with Bulk Extractor
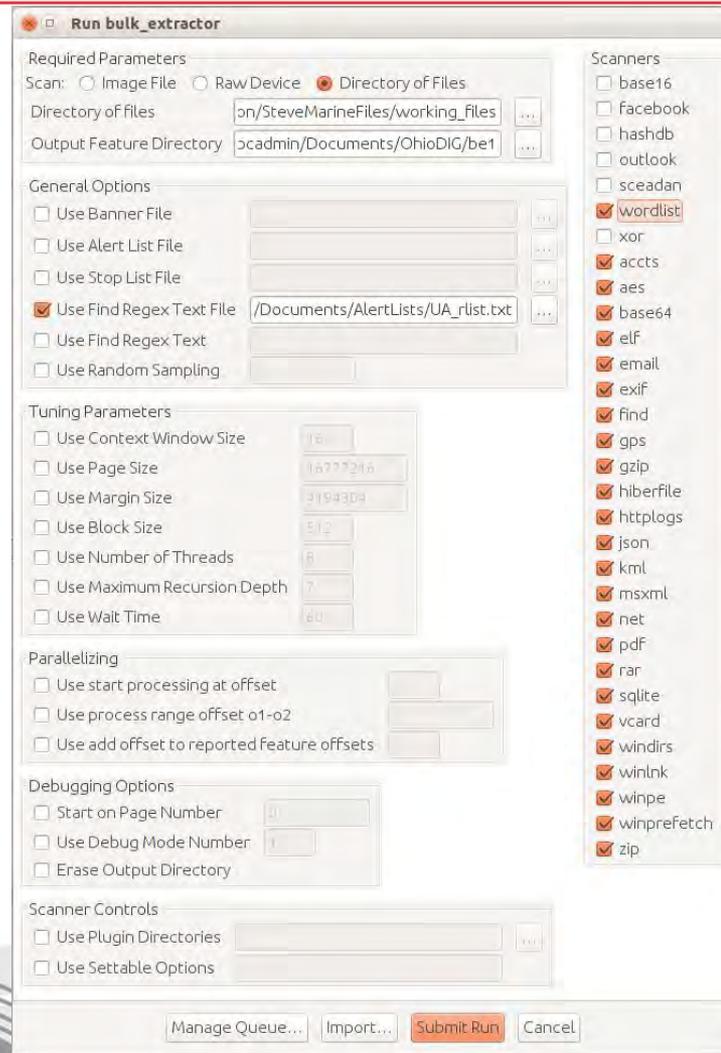
What's happening here?

-The top box is where you point BE to the origin set of files (or forensic disk image) and designate an output destination

-The middle box is where you can select a list of keywords you want to flag

-The right column allows you to select pre-programmed "scanners" (e.g., credit card numbers, emails)

UNIVERSITY LIBRARIES www.libraries.uc.edu

Cincinnati

Running Bulk Extractor
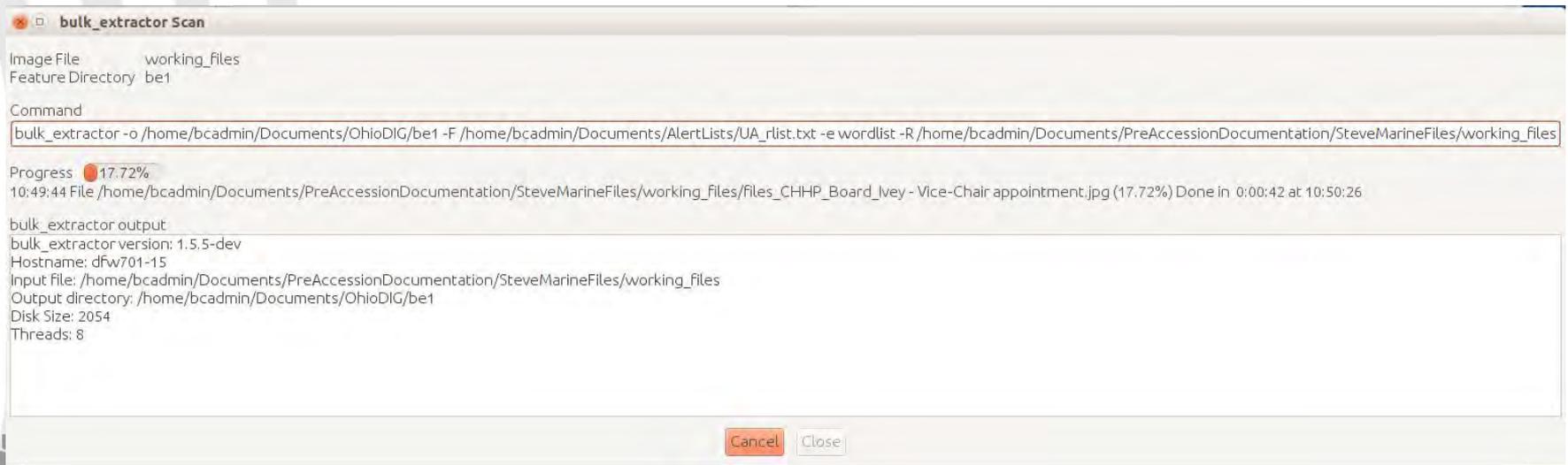
Let's Explore These Files
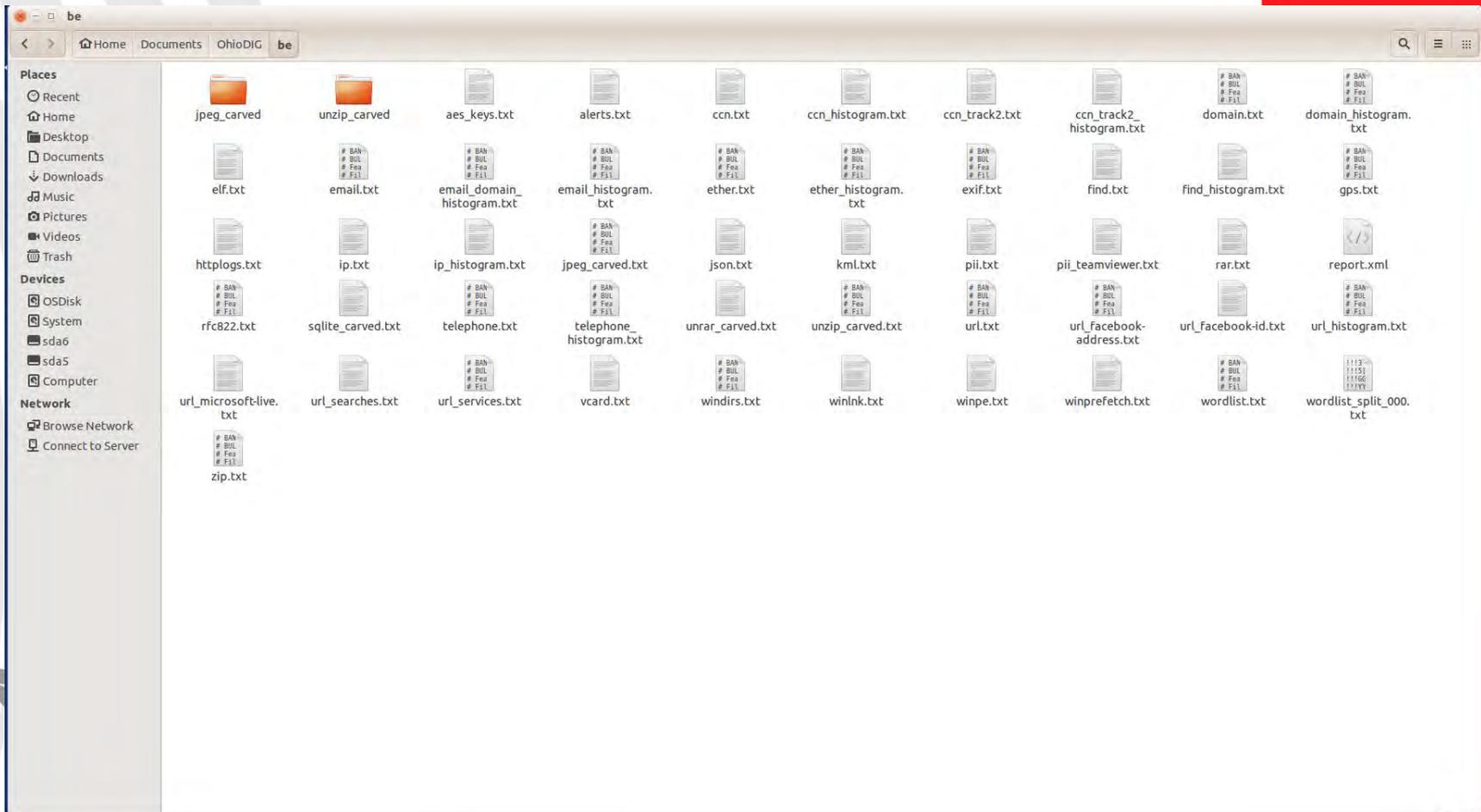
Looking Inside Flagged Files

What's happening here?

The items in bold in the left column represent scanners that found something.

The histogram shows how many files match an alert. This helps you establish rough idea of review time needed.

The screen on right shows which file has the matching word so you can determine whether to retain.

Questions or Comments?

http://www.bitcurator.net

http://coptr.digipres.org

http://digitalpowrr.niu.edu/

https://groups.google.com/forum/#!forum/digital-curation

eira.tansey@uc.edu

UNIVERSITY OF
Cincinnati