PRONOM, Siegfried, and Brunnhilde Exercise

BitCuratorEdu

Last Updated: January 18, 2022

About This Exercise

Author

Cal Lee

Description

This hands-on exercise is meant to introduce students to tools for file format analysis, including PRONOM, Siegfried, and Brunnhilde. These slides are excerpted from Cal Lee's SAA "Advanced Digital Forensics" class. The sample data referenced in these slides is available here: https://github.com/BitCurator/bcc-dfa-sample-data/

Learning object type

Lesson plan/materials

Learning objectives

This learning object might be used in a lesson to satisfy the following learning objectives:

- Practice using tools in the BitCurator Environment.

Exercise: Using PRONOM

The PRONOM technical registry contains information about a wide variety of file formats, including versioning information. You can find it at http://www.nationalarchives.gov.uk/PRONOM/Default.aspx. PRONOM has an online search feature that can be used to view the registry.

Click on “Search PRONOM” and navigate to the “File Format” tab. Clicking on the first search button (under “1. File Formats”) will allow you to view all of the entries in the registry.

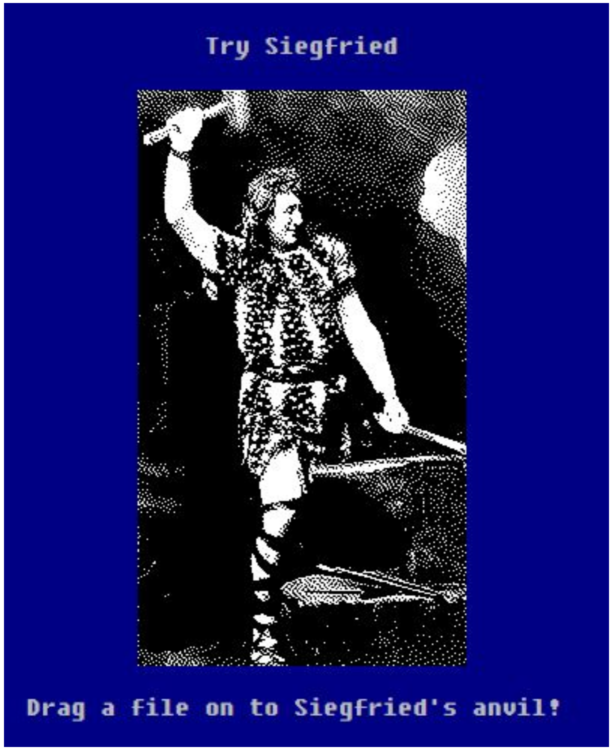
DROID incorporates information from PRONOM. It also uses file magic and file format extensions to provide a “best effort” at identifying file types. If you’d like to know more about DROID, you can find a quick demonstration video at: http://vimeo.com/24718678

Note: We’ll see DROID output in the Siegfried exercise later.

Siegfried

http://www.itforarchivists.com/siegfried/

- Signature-based file format identification tool
  - PRONOM file format signatures (National Archives of UK) (default)
  - MIME-info file format signatures (freedesktop.org)
  - FDD file format signatures (Library of Congress)
- Unlike FITS, does not have validation built in, and fewer extraction tools, but much lighter weight.
- Has a lot of customization for output
  - CSV
  - YAML (text)
  - DROID CSV
  - JSON
  - stdout



Try Siegfried

Drag a file on to Siegfried's anvil!

Brunnhilde

https://github.com/tw4l/brunnhilde

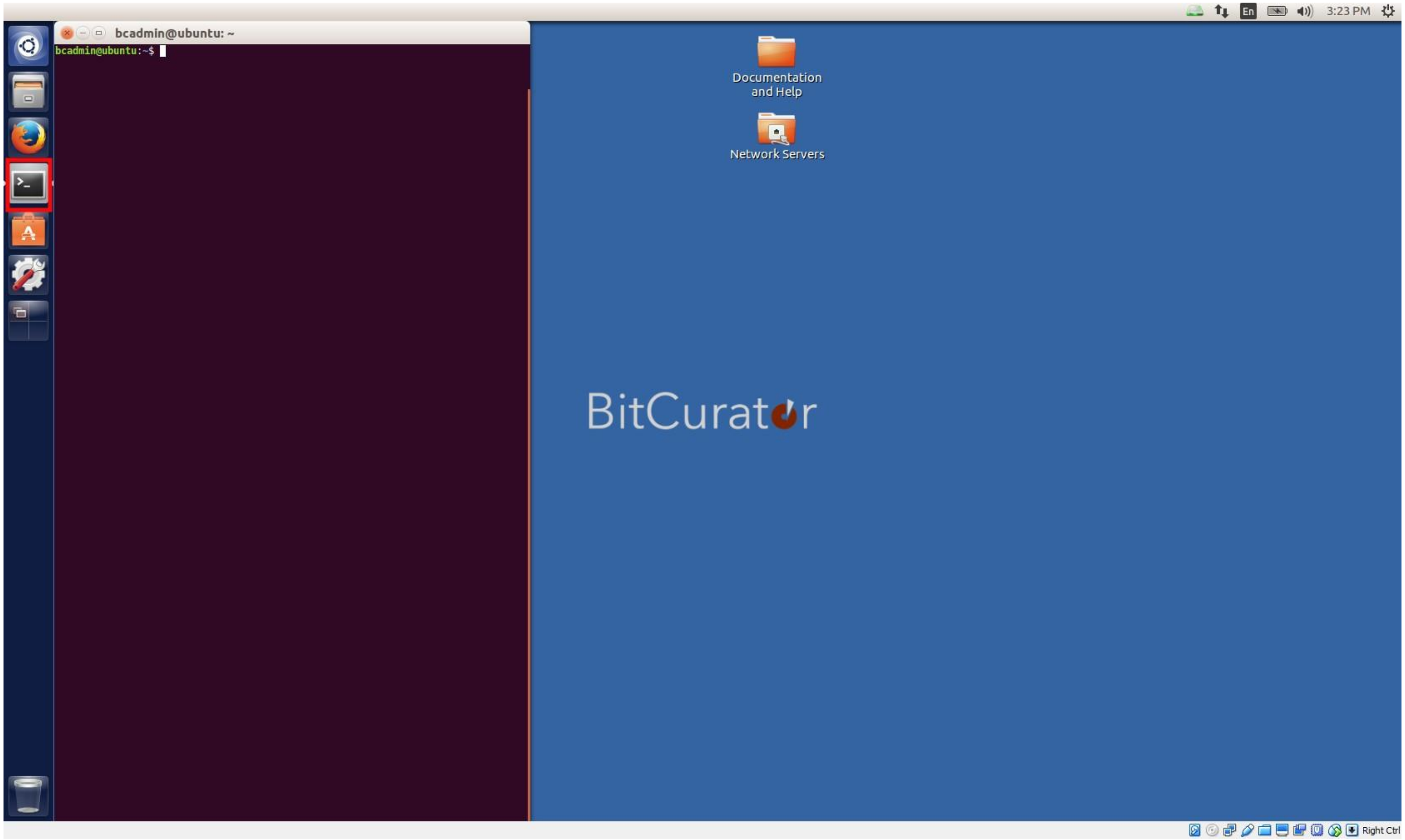
- Reporting companion for Siegfried
  - Requires Siegfried (but running Brunnhilde also runs Siegfried)
  - Command-line and GUI (we’ll be using the CLI version later)
- Reports generated
  - HTML (human readable)
  - Siegfried CSV
  - Directory tree
  - Other CSVs extracted from Siegfried logs (e.g., warnings, unidentified files)
- Can run other processes too, but not required
  - Virus scan
  - bulk_extractor
  - Disk image processing

Exercise: Siegfried and Brunnhilde

- Over the next few slides, we will run Siegfried and Brunnhilde in over the same set of files in several different ways
- Goals
  - Generate characterization and related technical metadata
  - Illustrate how the data can be configured for different uses
  - Identify decision points when data are unclear
- Source files to analyze: file_ident_ex directory in the Sample Data folder
- Note: the next two slides can be skipped if applications are already installed in your version of BitCurator (1.8.0 or later)

Installing Siegfried/Brunnhilde

- Start up the BitCurator VM (if it's not already running)
- Open a Terminal Window



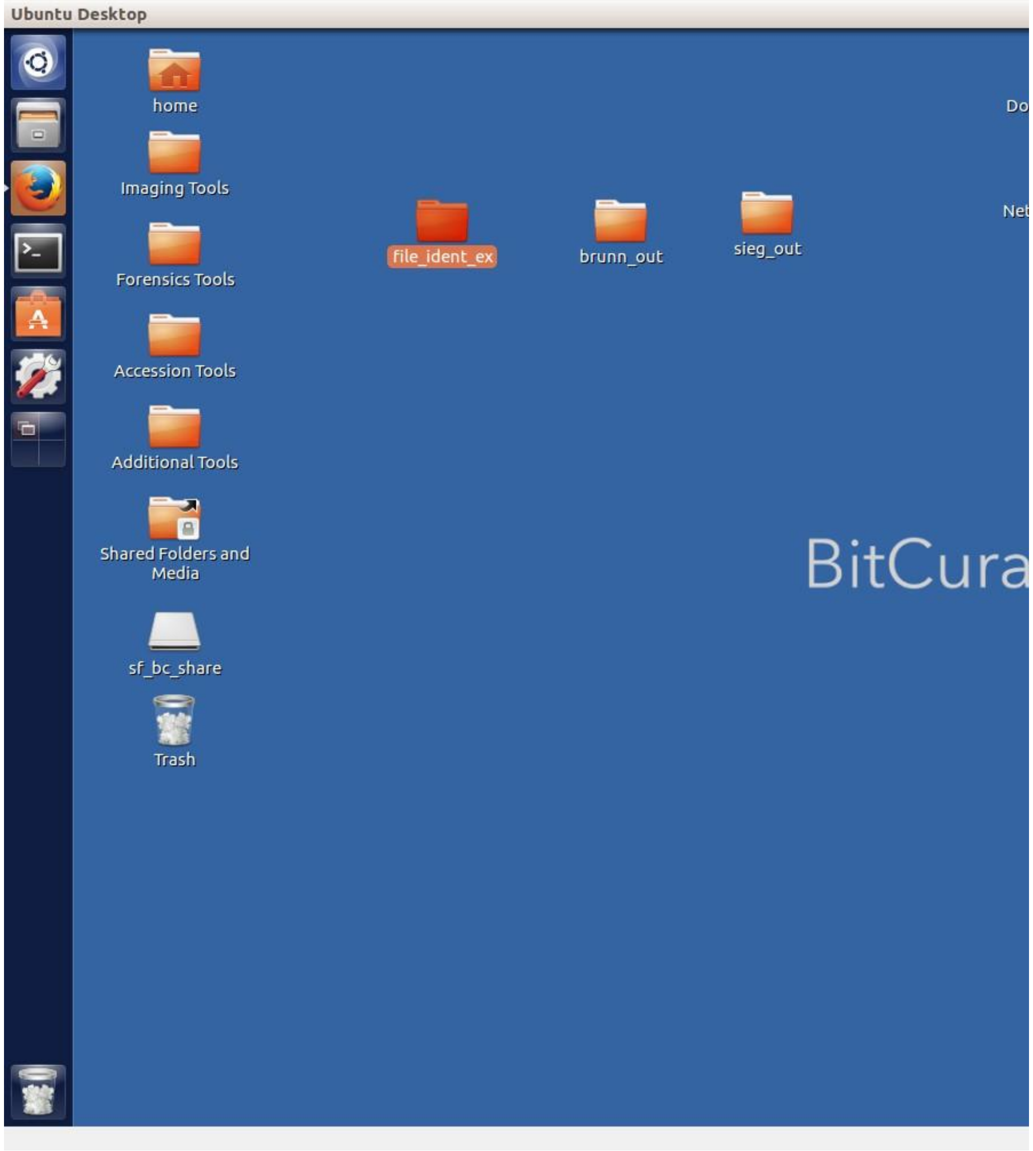
```
bcadmin@ubuntu:~$
```

Installing Siegfried/Brunnhilde

- Enter the following commands in Terminal:
  - wget -qO - https://bintray.com/user/downloadSubjectPublicKey?username=bintray | sudo apt-key add –
  - echo "deb http://dl.bintray.com/siegfried/debian wheezy main" | sudo tee -a /etc/apt/sources.list
  - sudo apt-get update && sudo apt-get install Siegfried
  - sudo pip install brunnhilde
- Note: there is a text file in the sample files that includes these commands if you want to copy/paste them

Running Siegfried and Brunnhilde I

- Start up the BitCurator VM (if it's not already running)
- Create "sieg-out" and "brunn-out" folders on the desktop
- Drag "file_ident_ex" folder from Sample Data to the Desktop



Running Siegfried and Brunnhilde II

- Open a Terminal window
- At the prompt, enter the following command:

  -
    ```
    sf ~/Desktop/file_ident_ex/ > ~/Desktop/sieg_out/sieg_out.yaml
    ```
- Open the sieg_out directory and look around
  - What did the command do?
  - What does the file tell you?
  - How would you characterize the data presented in the file?
  - Does anything strike you as odd? Particularly useful?

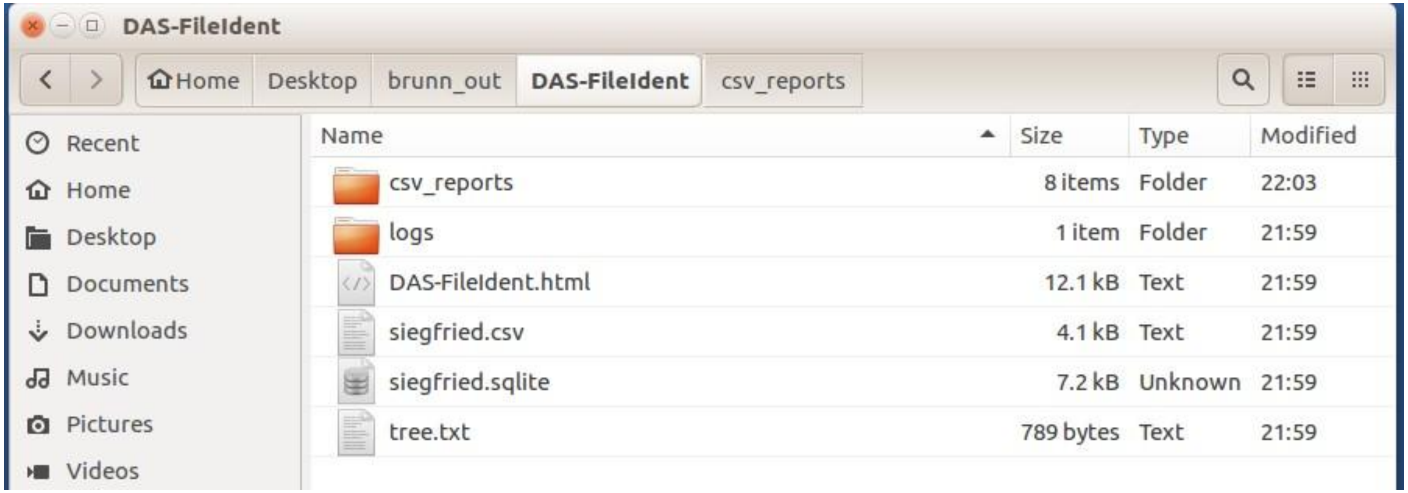
Running Siegfried and Brunnhilde III

- In the same Terminal window, enter the following commands:
  - sf -droid ~/Desktop/file_ident_ex/ > ~/Desktop/sieg_out/sieg_out-droid.csv
  - sf -json ~/Desktop/file_ident_ex/ > ~/Desktop/sieg_out/sieg_out-json.json
- Open the sieg_out directory and look around
  - What did the commands do?
  - How do these files differ from the one created on the previous slide?
  - Between the three output files, which do you think is most useful presentation of the data? (Hint: there may be more than one answer)
  - Do any of these files strike you as particularly useful? Particularly worthless?

Running Siegfried and Brunnhilde IV

- In the same Terminal window, enter the following command:

  -
    ```
    Brunnhilde.py -w ~/Desktop/file_ident_ex/ ~/Desktop/brunn_out/ DAS-FileIdent
    ```
- Open the brunn_out directory and look around
  - Did Brunnhilde perform any tasks over and above Siegfried?
  - How do the Brunnhilde output files differ from those generated by Siegfried?
  - Inspect csv_reports. How would you characterize what you see here?
  - Are the files here that Brunnhilde and Siegfried found problematic? What conclusions might you draw from them?
  - Is there information that Brunnhilde highlighted that you missed in Siegfried’s output?



Name	Size	Type	Modified
csv_reports	8 items	Folder	22:03
logs	1 item	Folder	21:59
DAS-FileIdent.html	12.1 kB	Text	21:59
siegfried.csv	4.1 kB	Text	21:59
siegfried.sqlite	7.2 kB	Unknown	21:59
tree.txt	789 bytes	Text	21:59

1.44MB
2HD IBM
imation
12/11/03
BitCuratorEdu
Advancing the adoption of digital forensics tools and methods in libraries and archives through professional education efforts
EDUCOPIA INSTITUTE
Community Cultivators
INSTITUTE of Museum and Library SERVICES

Most resources from the BitCuratorEdu project are intentionally left with basic formatting and without project branding. We encourage educators, practitioners, and students to adapt these materials as much as needed and share them widely.

The BitCuratorEdu project is a three-year effort funded by the Institute of Museum and Library Services (IMLS) to study and advance the adoption of digital forensics tools and methods in libraries and archives through professional education efforts. This project is a partnership between Educopia Institute and the School of Information and Library Science at the University of North Carolina at Chapel Hill, along with the Council of State Archivists (CoSA) and several Masters-level programs in library and information science.