*Electronic Records Modules*
Electronic Records Committee
Congressional Papers Section
**Society of American Archivists**

# Accessioning Born-Digital Content with BitCurator

**John Caldwell**
***University of Delaware***
jcald@udel.edu

**Date Published: March 8, 2018**

**Module#:** ERCM015

**Updated 2017-06**
**CPS Electronic Records Committee**

**Overview and Rationale**

Digital forensics is a field of criminal and computer science encompassing the recovery and examination of data created, managed and stored on a digital device. Originally developed for law enforcement, the application of digital forensics tools and methods has expanded into the spheres of digital humanities, information science and digital preservation.

In the archival world, digital forensics is used to analyze born-digital records, maintain their integrity and ensure long-term access. One way to do this is to create and manage forensic disk images (complete copies of both the structure and content of a computer disk) to preserve the original content and generate access copies. Creating disk images and storing them in the repository's preservation system is akin to the traditional accessioning process for analog archival materials--you are securing physical and intellectual control over digital records. One of the most widely used tools for doing this in the archival world is BitCurator.

BitCurator is a software environment which includes a collection of free and open source digital forensics tools to aid archivists in accessioning, appraising and processing born digital records. BitCurator is maintained by the BitCurator Project, a joint project between the School of Information and Library Science at the University of North Carolina, Chapel Hill and the Maryland Institute for Technology in the Humanities to develop a system for information professional in collecting repositories which incorporated digital forensics tools and techniques.[1] Since its inception in 2011, BitCurator has continued to be developed, including work on creating an access model for forensic disk images.

This module will identify specific tools and walk archivists through one possible workflow for using BitCurator to accession born-digital content in their repositories.

**Procedural Assumptions**

It is assumed that the repository is running the most recent version of BitCurator (1.8.12). It is also assumed that the repository has all of the necessary peripheral drives necessary to read and interact with electronic media in your collection, including, but not limited to, a physical write blocker (e.g. Tableau Forensic USB Bridge (T8-R2)), floppy disk drives for 3.5" and 5.25" disks (either built-in or peripheral), and an optical disk drive (either built-in or peripheral).

Because this module only focuses on the accessioning and ingest processes of using BitCurator, it will not discuss in detail the PII (personally identifiable information) and other reports generated by application in BitCurator or other software components in the BitCurator Environment. More

[1] UNC School of Information and Library Science, "BitCurator--About," BitCurator.net, https://www.bitcurator.net/bitcurator/ (accessed December 13, 2017).

information on these reports can be found on the BitCurator Wiki: https://wiki.bitcurator.net/index.php?title=BitCurator_Environment.

**Hardware and Software Requirements**

BitCurator is an Ubuntu-based Linux environment which can run on either a dedicated machine or through a virtual machine.

If running on a dedicated machine, users should be familiar with using the command line for installing BitCurator from the Live ISO. The ISO and installation instructions can be downloaded from the BitCurator wiki: https://wiki.bitcurator.net/index.php?title=Main_Page.

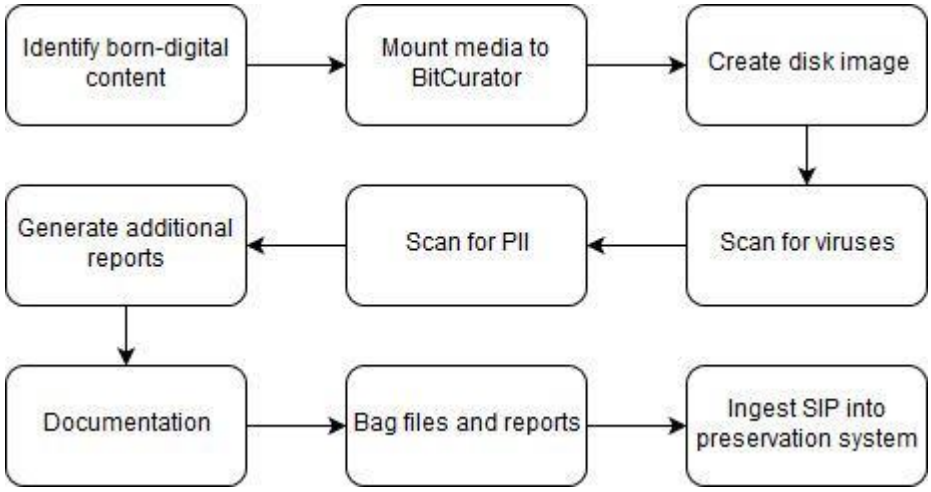
If running through a virtual machine, users will need to download the most recent version of VirtualBox, in addition to the BitCurator VM and the VirtualBox extension pack. Links to the necessary software and installation instructions can be downloaded from the BitCurator wiki: https://wiki.bitcurator.net/index.php?title=Main_Page.

All of the utilities mentioned in this module are included in BitCurator on install. These utilities are:

- Guymager
- ClamTK Virus Scanner
- Bulk Extractor Viewer
- BitCurator Reporting Tool
- Bagger
- Grsync
- A text editor

All of these utilities (or something similar), with the exception of the BitCurator Reporting Tool, can be procured and used outside of the BitCurator Suite. *Therefore, this workflow can be adapted for repositories NOT using the BitCurator Suite to accession born-digital collection material.*

**Workflow**



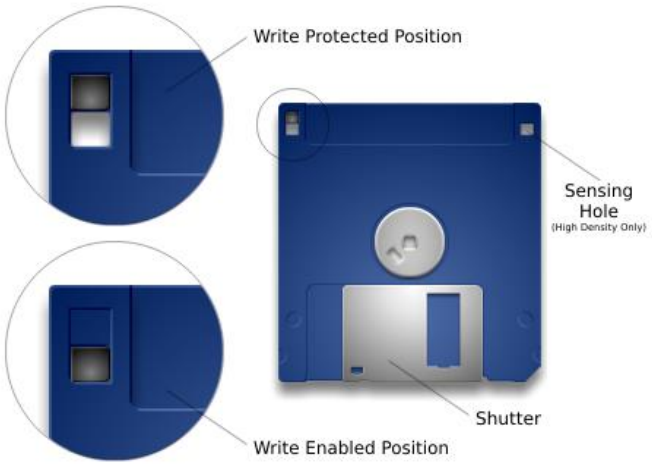

## I. Step 1: Identify born-digital content

1. Accessioning archivist checks for born-digital content.
2. If born-digital content is a part of a hybrid collection, document the removal of digital media from its place in the physical collection.

## II. Step 2: Mount media to BitCurator

1. Log into BitCurator.
2. BitCurator has a built-in software write blocker, to prevent accidentally overwriting data on removable media, which is active when BitCurator's mount policy is set to READ-ONLY. **Make sure that the mount policy is READ-ONLY. The disk icon in the top right corner should be colored green**.


3. Insert media into the appropriate drive or write blocker.
    a. Optical media **does not need to be physically write protected.** Insert media directly into the optical media drive.
    b. Magnetic media, such as floppy disks, **needs to be physically write protected**.
        i. To write block a 3.5" floppy disk, switch the write protect tab so that the space is open. You should see two "holes" (like a pair of eyes) in the floppy disk.



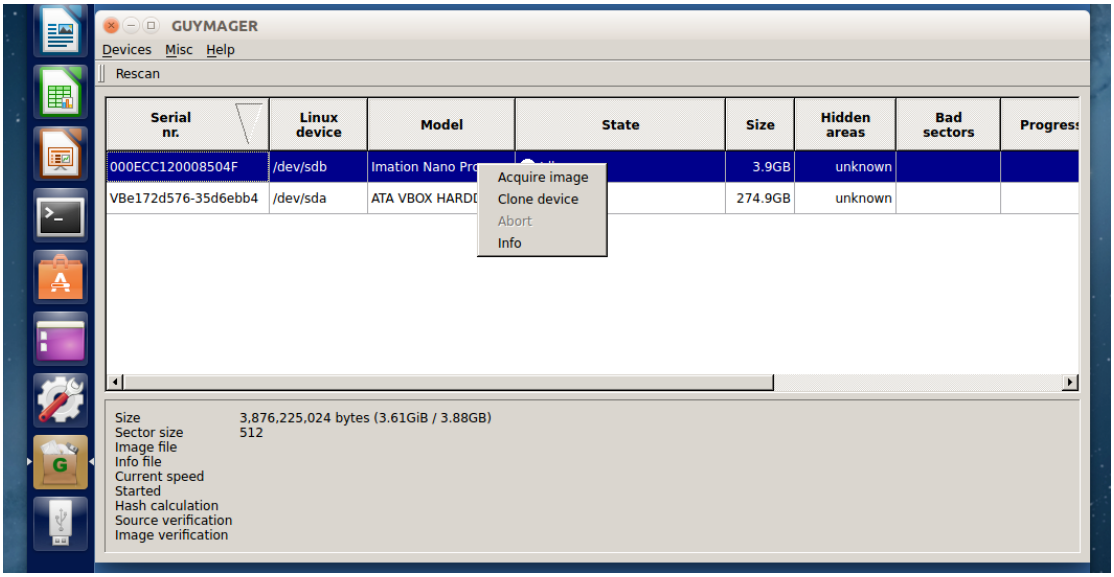

ii. The 5.25” floppy disks are write protected by the adapter connected to the external floppy drive. For additional protection, cover the Write Enable Notch with Scotch Tape.

c. USB flash drives and external hard drives **need to be physically write protected**. Recommended tool: Tableau Forensic USB Bridge (T8-R2).

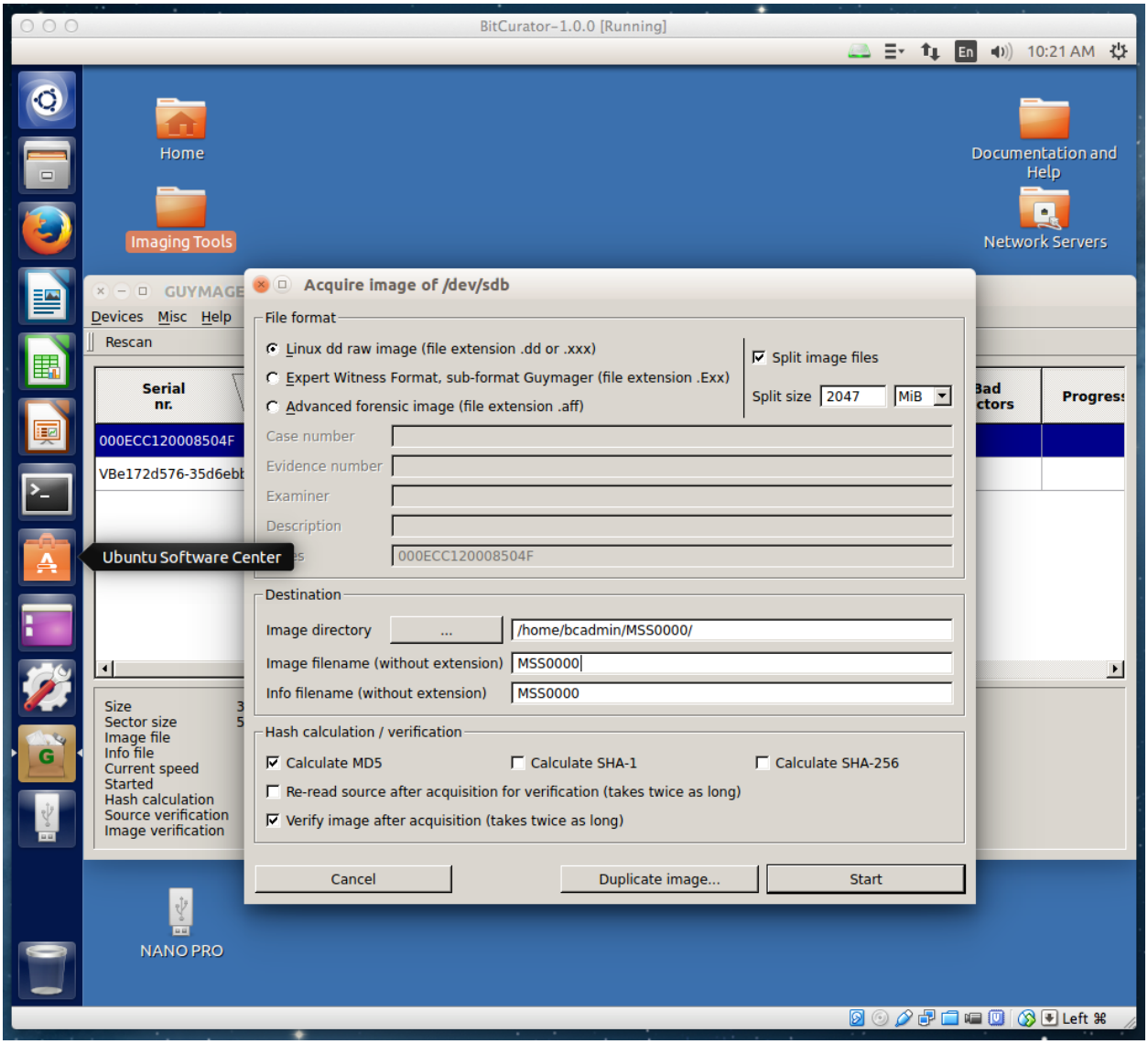
4. The media should be safely mounted in the BitCurator environment. Do not attempt to open any files from the mounted drive. Look for a media icon to appear on the side bar (see image below). This image may look like a flash drive, or it may look like a disk (each type of media displays as a different icon).


## III. Step 3: Create disk image

1. Create a directory to store your disk image.
    a. Right click anywhere on the Desktop.
    b. Select “New Folder” from the drop-down menu.
    c. Name the folder using the collection number or accession number (e.g. MSS_XXXX).
    d. Create a new folder for each new piece of media in the collection. This is the **media folder**.
    e. Within each media folder, create a folder labeled “bulk_extractor.” This is where you will store the Bulk Extractor reports later in the accessioning process.
2. Double click on the **Imaging and Recovery Tools** folder located on the desktop.
3. Double click on the **Guymager** icon.
4. Once Guymager loads, you will see a list of all disks connected to the computer. Right click on the disk you want to image and select “Acquire image.”



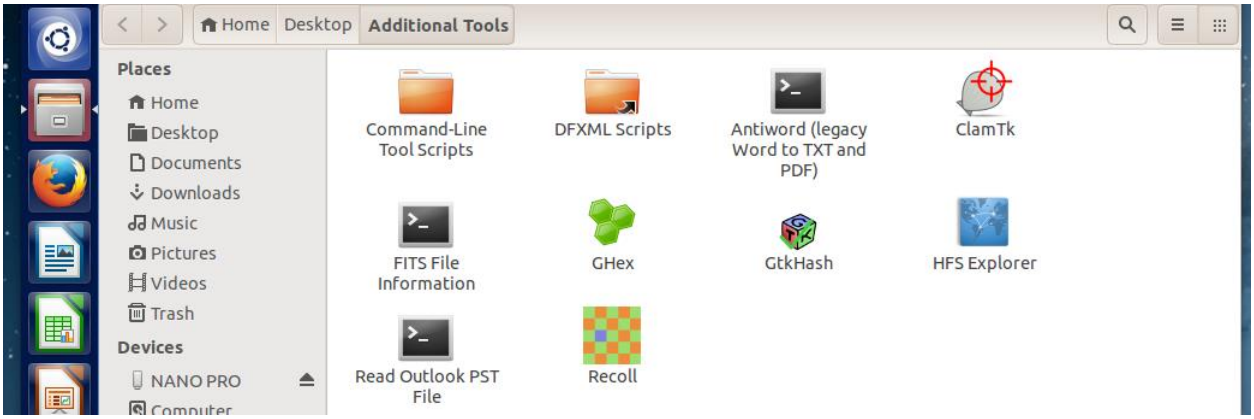

5. Select the image format “Linux dd raw image.”
6. Choose the image directory (e.g. /home/bcadmin/Desktop/MSS_XXXX/01.”
7. Enter the image filename without a file extension. This name should match the title of the media folder (e.g. MSS_XXXX_01).



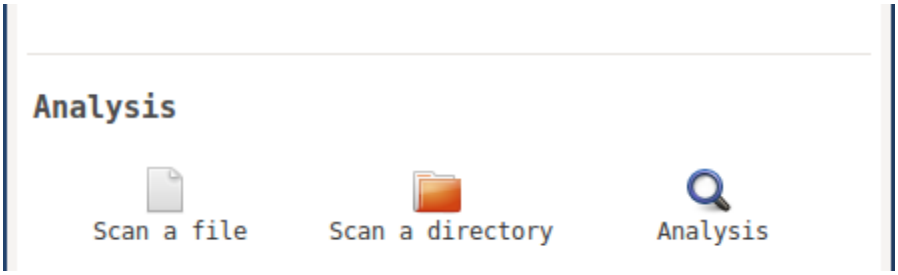

8. Click “Start.”
9. When Guymager has finished creating the disk image, exit the application.
10. Verify the image by navigating to the directory you created. There should be two files: the disk image and an info file.
    a. The info file is a log file generated by Guymager.

## IV. Step 4: Scan for viruses

1. Double click the **Additional Tools** folder located on the desktop.
2. Double click the **ClamTK** icon.



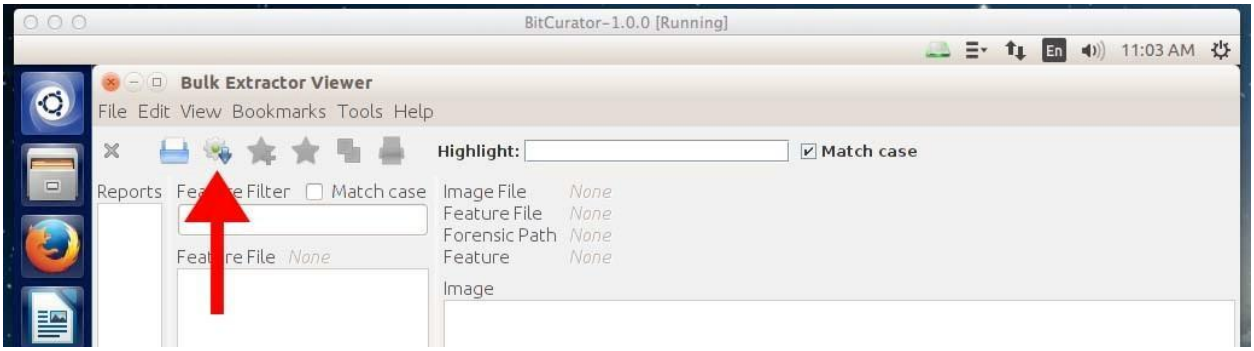

3. Under the “Analysis” section, choose “Scan a directory.”
4. Select the media folder to scan and click “OK.”



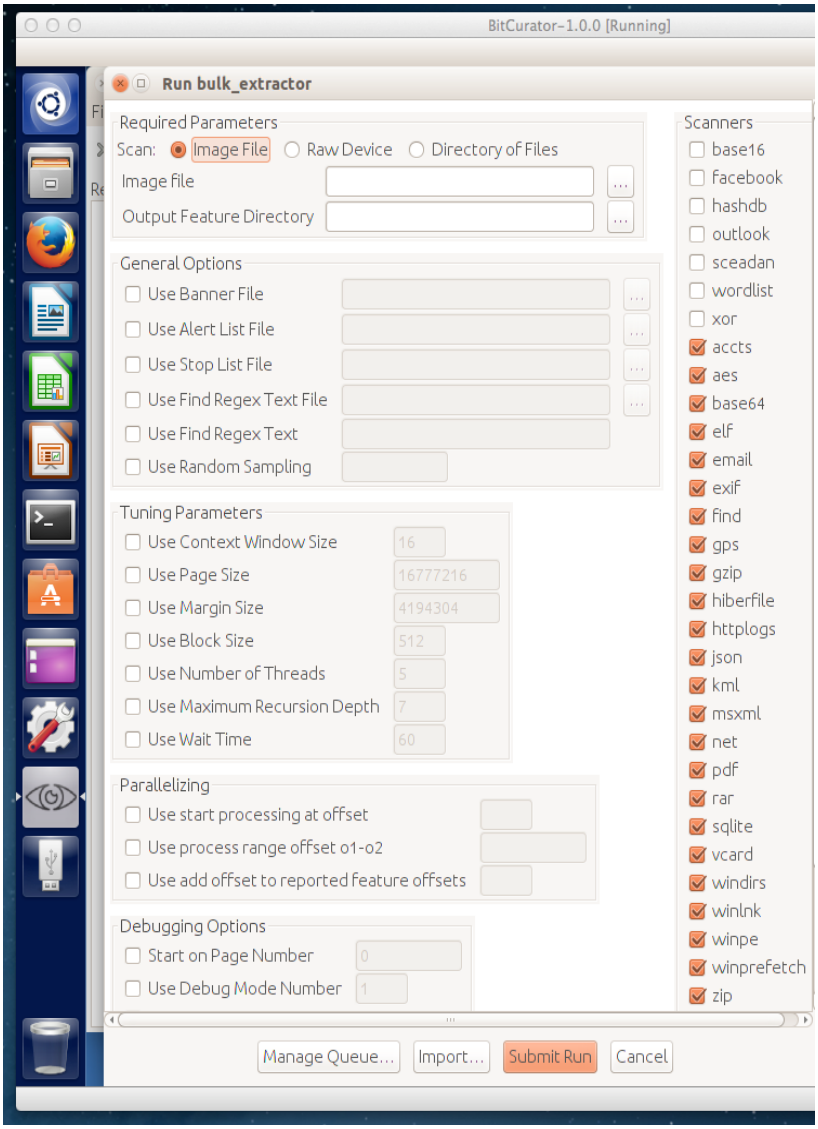

5. Quarantine files as necessary. If any files are quarantined, be sure to include this in the documentation.
6. When the virus scan is complete, click “Exit.”

## V. Step 5: Scan for PII

1. Double click on the **Forensics and Reporting Tools** folder on the Desktop.
2. Double click on the **Bulk Extractor Viewer** icon.
3. Click on the gear icon with the down arrow to open the “Run bulk_extractor” window.




4. Select the type of media you want to scan. In most cases, this will be your disk image, so select the “image file” radial button.
5. Type or navigate to the location of the disk image in the “Image file” field (e.g. /home/bcadmin/Desktop/MSS_XXXX/01/MSS_XXXX_01.000).
6. Enter the name of the directory to store the Bulk Extractor output. Choose the “bulk_extractor” folder created earlier (e.g. /home/bcadmin/Desktop/MSS_XXXX/01/bulk_extractor).
7. Check or uncheck the scanner options from the list on the right. The default is for all scanners to be checked. A list of scanners and their descriptions is available.



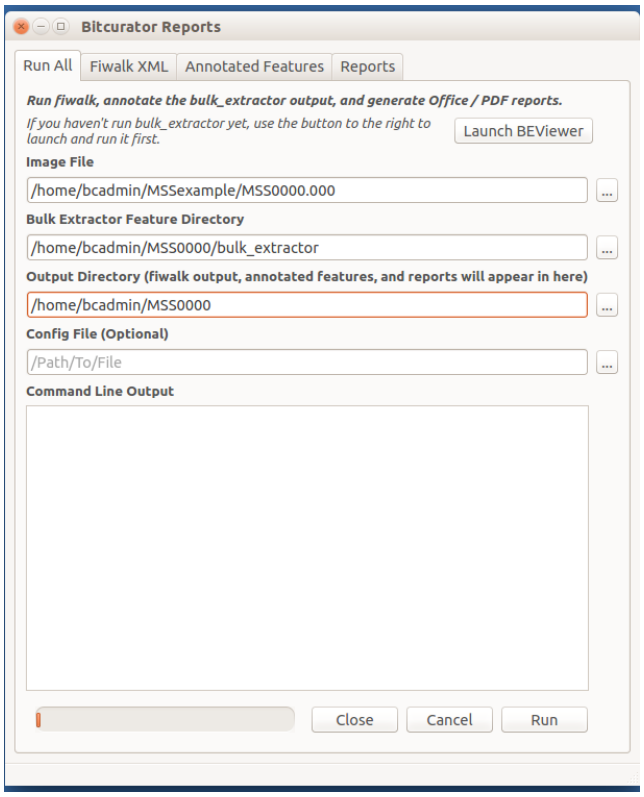

8. Click “Submit Run” at the bottom of the screen to begin the scan.
9. The reports will be automatically saved in the “bulk_extractor” folder. When Bulk Extractor has finished running, exit the program.

## VI. Step 6: Generate additional reports

1. Double click on the **Forensics and Reporting Tools** folder on the Desktop.
2. Double click on the **BitCurator Reporting Tool** icon.
3. Type or navigate to the location of the disk image in the “Image file” field (e.g.

/home/bcadmin/Desktop/MSS_XXXX/01/MSS_XXXX_01.000).

4. In the “Bulk Extractor Feature Directory,” type or navigate to the location of the feature directory (e.g. /home/bcadmin/Desktop/MSS_XXXX/01/bulk_extractor).
5. In the “Output Directory,” type or navigate to the location of the media directory (e.g. /home/bcadmin/Desktop/MSS_XXXX/01).




6. Click the “Run” button.
7. In the Command Line Output box, look for a “success” message. When this message is received, click the “Close” button.
8. The following is the list of reports generated:

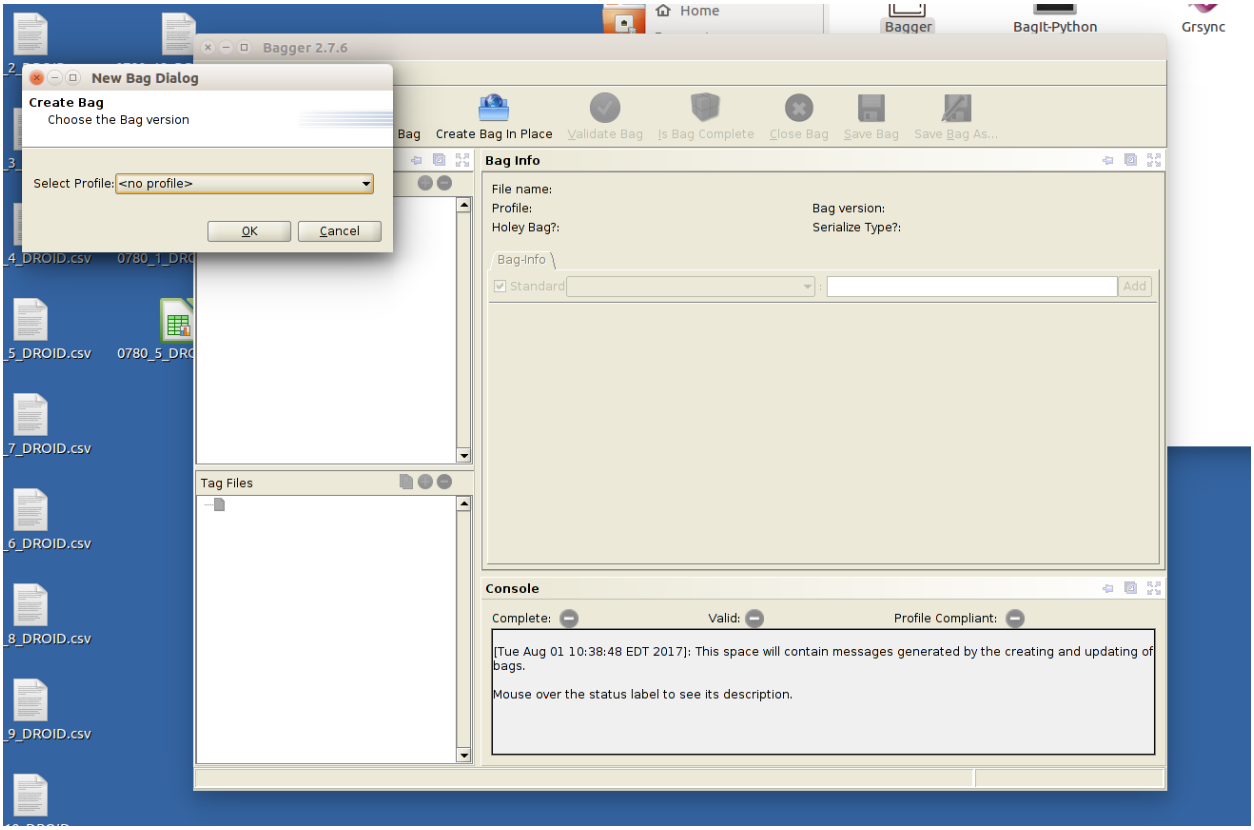
## VII. Step 7: Documentation

1. Create a “README” file.
    a. Document collection information, the type of media (e.g. CD, 3.25” floppy disk), any identifying information from the physical media (e.g. label information), size of data on media, and note any special actions taken during the accessioning process (e.g. quarantined files).

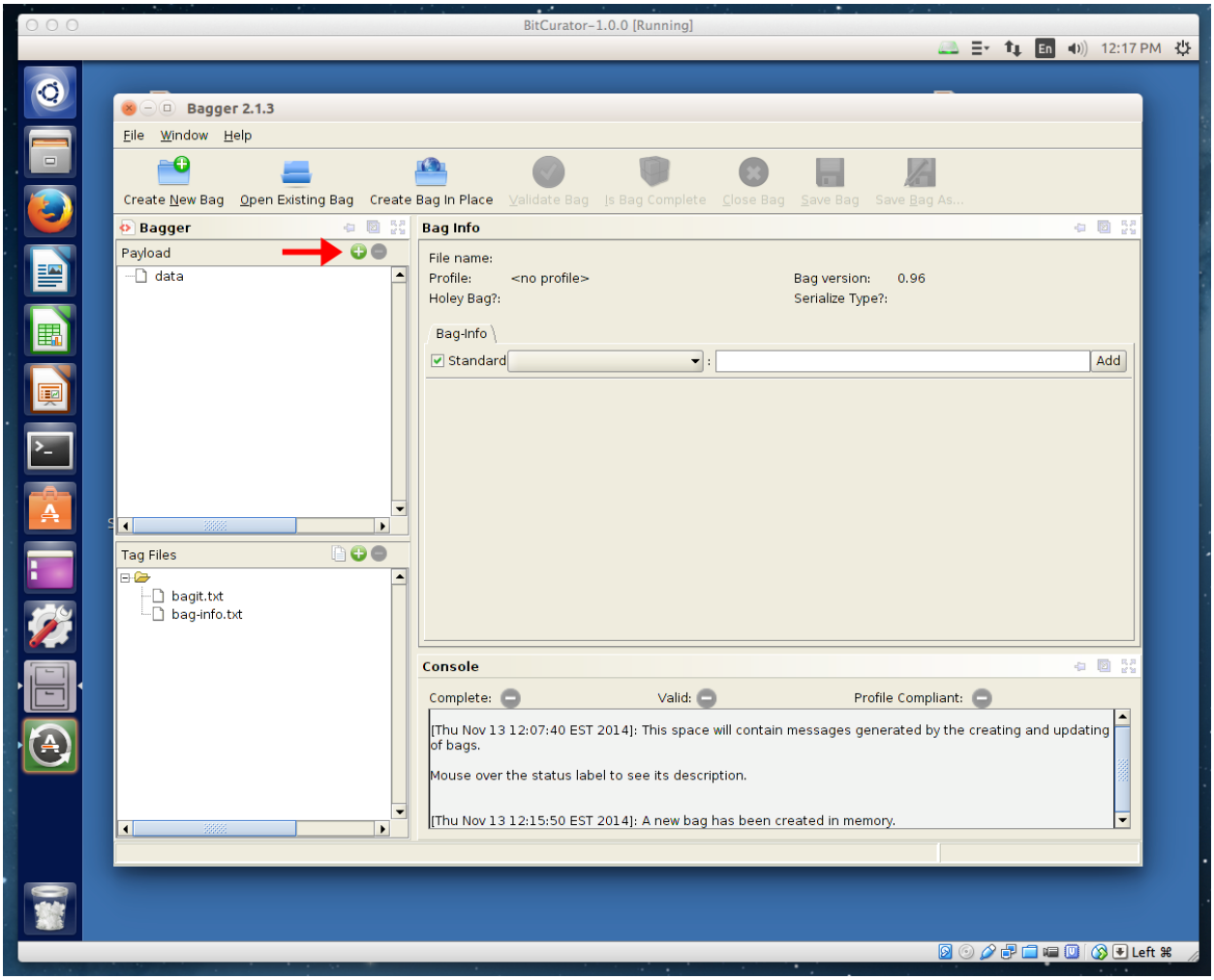
2. Save the “README” file in the media folder.
3. If your repository uses ArchivesSpace or other archival management system, document the disk imaging and ingest as part of your accession record for the collection.

## VIII. Step 8: Bag content

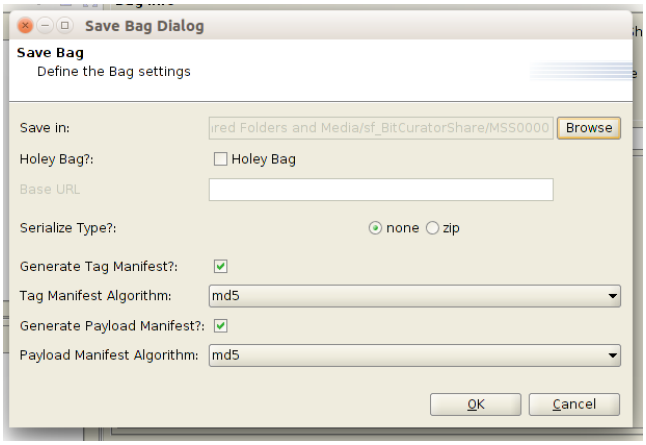
1. Double click on the **Packaging and Transfer Tools** folder on the Desktop.
2. Double click the **Bagger** icon.
3. Select the “Create New Bag” button from the top of the Bagger application.
4. Select the default profile (<no profile>) and click “OK.”




5. Select the “Add Files” button (the green plus sign in the “Payload” form).



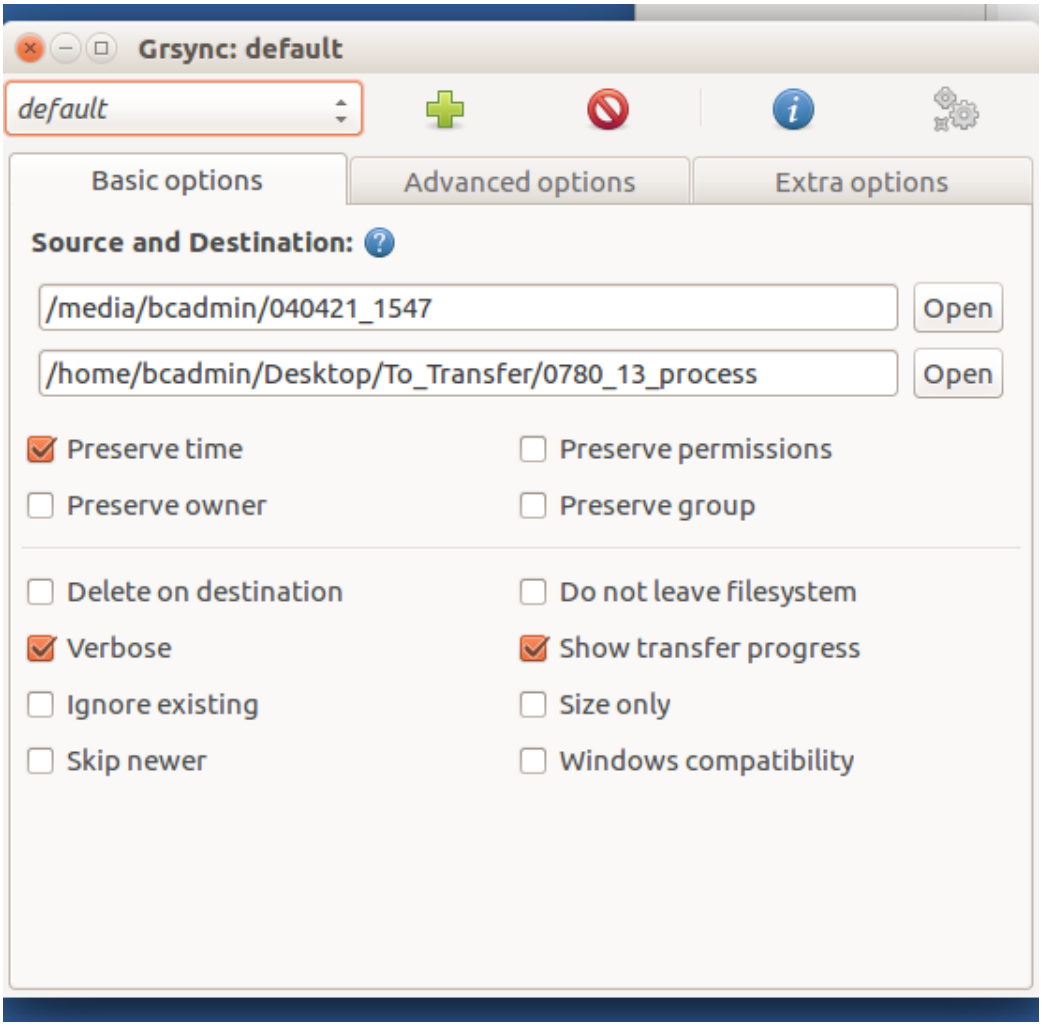

6. Navigate to the media folder and click “Open.”
7. Click the “Save Bag As” button from the top of the Bagger application.
8. Click “Browse” to choose a location to save the bag.
9. Leave all of the default options selected and press “OK.”



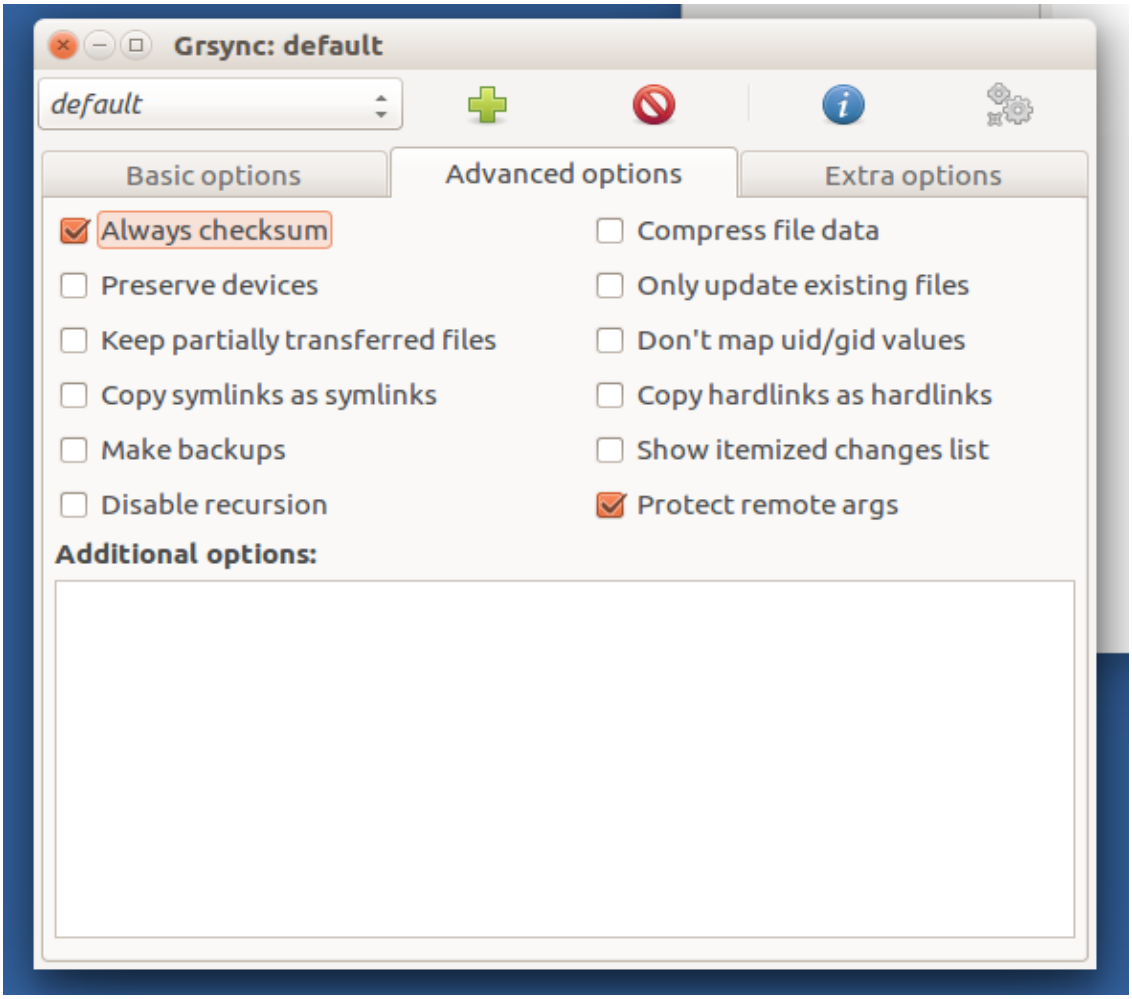

10. A dialog box will indicate that the bag was saved successfully. Click “OK” and close Bagger.

## IX. Step 9: Ingest into preservation system

1. Double click the **Packaging and Transfer Tools** folder on the desktop.
2. Double click the **Grsync** icon.
3. Under the “Basic Options” menu:
    a. Select the “Source” for the data being copied. Click on the “Open” button and navigate to the bag you just created and press “OK.”
    b. Select the “Destination” for the data. This should be where **preservation** copies of electronic files are stored. Click the “Open” button and navigate to the desired directory and press “OK.”
    c. Make sure the “Preserve time,” “Verbose,” and “Show transfer progress” checkboxes are selected.




4. Under the “Advanced options” menu, make sure the “Always checksum” and “Protect remote args” checkboxes are selected.




5. Hit the gears icon on the top right to begin the copy process. A dialog box will pop up showing the progress of the copying operation. When the copy operation is complete, click “Close.”
6. At this point, unmount the media and eject from BitCurator.
    a. Right-click on the media icon and select “Unmount media.”
    b. Once the media icon disappears from the side bar, it is safe to eject the media from the computer.
        i. When dismounting media from BitCurator, you may be prompted to input a password. The default password within the BitCurator environment is “bcadmin.”
7. Return the physical disk to storage, or dispose of the media, as per institutional practice.