Accession Procedure for Disk Images (DEPRECATED)

This information is not current and is only provided for historical purposes.

Receiving

Step 1: Check media out from collection owner, transport to Born Digital Preservation Lab.

Step 2: Create new Media Log file from medialog-template.xlsx for new collection if one does not exist.

- With the title consisting of <collection code>medialog.xlsx (ex. UACmedialog.xlsx)

Step 3: Note receiving information in Media Log (see here for help) for all physical media received.

- Generate accession number following creation guidelines and note in Media Log.
- Label physical object with accession number.

**Note: at this point the University Archives has taken responsibility for generating BDPL #'s and labelling media from their collections.

Step 4: Engage write-protect tabs on media, if applicable => Safe Handling of Removable Media (Disks).

Step 5: Create a new folder for each object, in ~/Documents/1received/ on the BDPL computer.

- Title with accession number (ex. ~/Documents/1received/UAC2016040002/).
- For creating multiple folders and Info CSV's at once, follow the instructions located in the ~/Documents/scripts/MakeFolders/ folder.

Step 6: Create new Info CSV document based on the latest version of the template, stored in the BDPL Box folder:

> DIGIPRES-start-here => 2-Born Digital Preservation Lab => BDPL Info-Templates

- Create the title: (Collection Code)(Accession Number)-info.csv for each object (see here for help).

  ex. ~/Documents/1received/UAC0420160002/UAC0420160002-info.csv

Step 7: Note receiving information in the Info CSV file.

**Note: Entries into the CSV file should NOT include any commas or quotation marks.

Pre-Ingest

Step 8: Plug media into computer w/ write-blocking hardware, write-protect tabs, and read-only software engaged.

- Follow all safe handling procedures.

Step 9: Open collection Media Log file and associated Info CSV document

- Note pre-ingest information in both.

**Note: When opening the Info CSV file for the first time (ex. UAC0420160002.csv) "Select All" (ctrl-A) and "Format Cells" (right-click and select). Set the "Number" to "Text" and "Alignment" to "Left". This will maintain the same formatting for every Info CSV file.

Step 10: Create a raw (.dd) disk image w/ Guymager. For CD's and DVD's use DDRescue:

> ddrescue -n -b2048 /dev/sr0 <path to output .ISO> <path to map file .map>
>
> ddrescue -d -r1 -b2048 /dev/sr0 <path to output .ISO> <path to map file .map>
>
> *mapfiles can be deleted

- Save disk images in associated folder at /home/bcadmin/Documents/1received/<accession number>
- Title with (Collection Code)(Accession Number).
- In Guymager, ensure that "Calculate MD5" and "Verify image after acquisition" are selected, and that "Split Image" is not selected.

Step 11: Make sure the disk imaging was successful, and record the appropriate information in the Media Log and .csv file.

- Enter the following information into Media Log and .csv document: Image type, filesize, filesize notes, and "successful/ unsuccessful"
- For CD and DVD media, making sure write-protection (both physical and virtual) is engaged, mount the original disk.
- Using BitCurator Disk Image Access, open the disk image, and quickly check to see that the disk image contains the same files as the original disk.

**Note: This step is especially important for CD and DVD media, because we have had issues in the past.

Step 12: Remove physical media and place back in any packaging.

- Be sure to properly eject media before removing.

- Transcribe anything written on the physical disk, entering it into the Info CSV document.
- "Physical Notes" and "Original Media Type Notes" both refer to any details about the physical disk that are printed, by the manufacturer, on the disk. Serial numbers, for example, tell future researchers where disks were manufactured and which disks were produced/ sold in the same batch.

Step 13: Create MD5 file, using Nautilus, and store in object folder.

- Right-click on the disk image, and select Scripts => File Analysis => Calculate md5. Select the "Save To File" option.

Step 14: Using FiWalk and the FiClam plugin, create a structural DFXML and run Clam Antivirus on each item in the disk image.

- Using the command prompt, navigate to Tools/ficlam (cd Tools/ficlam).
- The filename for the .dfxml is going to be the same as the disk image, but with a different extension: "(Collection code)(Accession number)".xml
- While at Tools/ficlam, enter the following into the command line:

    fiwalk -c clamconfig.txt -X <path to output .xml> <path to disk image>

 **Rather than type the command, you may:

    (a) Open ~/Documents/ClamCommand.txt and cut-and-paste the fiwalk command (use shift-ctrl-v to paste at the command prompt). Then correct the folder and file names.

    (b) In the Terminal, navigate to ~/Documents/scripts/FiClamGUI/ and run the FiClamGUI.py script:

        python FiClamGUI.py

- Using Nautilus (the Ubuntu file manager), navigate to the object folder and open the dfxml file (double-click).
- Use "Find" (Ctrl-f) => "</clam" to jump to each file listed in the report, and make sure the value for "<clamav_infected>0</clamav_infected>" is "0"

    **If the value for <clamav_infected>1</clamav_infected> is "1", note this in the Info CSV, move the folder to ~/Documents/1received/4review/, and alert managing staff.

Step 16: Access the disk image using BitCurator Image Access (if this does not work, you can also just use the mount command).

- Undertake initial analysis.
- Enter any anomalies or comments into "pre-ingest notes" in the Media Log.
- If any issues are found, note this in the Info CSV, move folder to ~/Documents/1received/4review/ and alert managing staff.

Step 17: Move object folder to ~/Documents/2readyforingest.

Ingest

Step 18: Create a tar package of object folder.

Step 19: Deposit into SDA utilizing internal deposit workflow.