



## ACCESSIONING FILES FROM CDS AND DVDS

### CONTENTS

I. Basic Workflow .....	1
II. Imaging vs File Extraction .....	2
III. Unique Identifier .....	3
III. Virus Scans .....	4
IV. Create Disc Image .....	5
A. IsoBuster .....	5
B. FTK Imager .....	8
C. Examine disc image .....	11
V. File Extraction .....	12
A. Which files to extract .....	12
B. Software .....	13
FTK Imager .....	13
Forensic Toolkit .....	15
Bagger .....	20
C. Post file extraction .....	21
VI. Documentation .....	22
VII. Handling of Discs Post-Digital Capture .....	22
VIII. Appendix: Understanding Common File Systems on Optical Discs .....	22

Note: If you are dealing with event recordings on optical discs, the below guidelines apply but you will also need to consult the events manual for additional steps.

### I. BASIC WORKFLOW

The following is a basic workflow for accessioning CDs and DVDs.

1. Assign a unique identifier to disc.
2. Make sure autoplay function is turned off on computer before inserting disc.
3. Scan disc for viruses. Record virus scan results and actions as Virus Check event in ASpace.
4. Create disc image or extract files from disc and store in “[accession#]\_original” folder on ira\_locked.



5. If image was created and time permits, export files from image and save in “[accession#]\_original” folder.
6. Maintain extracted files in bag.
7. Record actions in Digital File Management Note of accession record in ASpace.
8. Place disc in Alt Media box for shredding or retain disc in collection.

## II. IMAGING VS FILE EXTRACTION

To accession content on a CD or DVD, you will need to decide whether to a) create a disc image (iso file) of the disc or b) extract files from the disc. Keep in mind that even if you decide to create a disc image, files will eventually need to be exported for preservation in Rosetta.

The following are some factors to consider in making your decision. The below should not be considered all-encompassing and will need to be updated as we encounter more scenarios.

**Content:** Content refers not only to the content of the files, but also in a more general sense, as in the format type of the files and the nature/purpose of the disc.

The following are instances where we do want to create a disc image:

- You want to maintain the full functionality of the CD or DVD. For example, for discs that are programmed to automatically run, it is usually difficult for a non-technical individual to determine which file to click to manually execute a program or play a video.
  - Disc contains software files
  - Disc has an interactive menu, usually a DVD containing a Video\_TS folder with .vob files
- Disc appears to have been created for distribution, rather than for data backup. These discs usually have a custom label or insert.

The following are instances where file extraction is sufficient or makes more sense than imaging a disc:

- The disc appears to have been created for computer backup or data transfer (as opposed to distribution) and consists entirely of text/image/audio files.
- You've appraised the files and have decided you only want a portion of the files on the disc.

**Time/efficiency:** Appraisal adds significant time to accessioning discs. If you have a large number of discs and little time to examine their contents, you may want to image all the discs so that they can be preserved in Rosetta more quickly.

The computer processing time to create an image versus extracting files is about the same when using FTK Imager (based on a very small, unscientific test), so theoretically you could also choose to extract files from all the discs. As explained in the Content section, there are some circumstances where you should create a disc image, and unless you're confident that they do not apply to any of your discs, creating disc images is the safer strategy. While in many cases imaging a disc creates an extra step in the workflow since files will eventually need to be extracted from the images for preservation, it is one worth taking if it means getting the content off discs and into Rosetta sooner than later.



**Technical issues:** You will likely encounter discs that are unreadable or have bad sectors. The disc might not have been burned properly or it may be damaged due to improper handling or storage or poor manufacture quality.

If you tried unsuccessfully to create an image of a disc using FTK Imager and IsoBuster, try to extract files to salvage as much of the content on the disc as possible. If there are particularly important files that you are unable to extract, you may want to try special recovery software such as Ontrack EasyRecovery Professional.

### III. UNIQUE IDENTIFIER

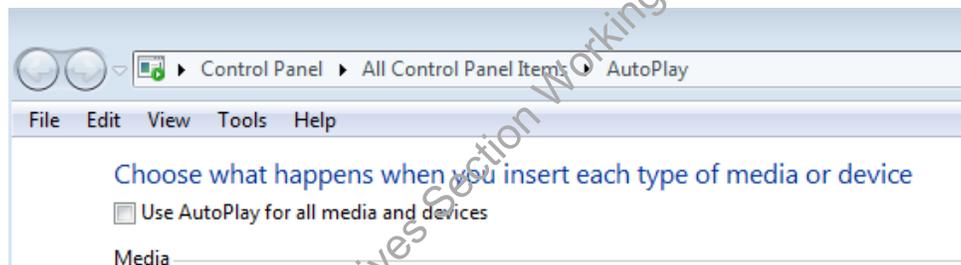
Assign each disc a unique identifier following the format that applies to your scenario. You may need to use a mix of formats within a single accession. Make sure you do not duplicate identifiers used for other digital storage media in the accession. Use the unique identifier as the filename for disc images and folder name for exported files.

Scenario	Unique Identifier
Accession consists entirely of a single disc.	[accession #]  Example: 2016ia38
Disc does not need to be retained and it is not necessary to record its original physical context.  This applies to:  -hybrid collections with <b>one or multiple</b> discs or multiple types of digital storage media  -purely digital accessions with <b>multiple</b> discs or multiple types of digital storage media.	[accession #]_i[item number]  Example: 2016ia38_i02
Disc is to be retained.  <i>or</i>  Disc is not to be retained but the original physical context needs to be recorded.	[accession #]_b[box #]_i[item #]  Example: 2016ia38_b02_i01  Note: Item numbering should restart from 01 with each box.
Event recordings  Note: We were instructed to use this format in the past, but this may no longer be necessary. We may change this in the future to align with our naming conventions for other optical discs.	gia_[accession # with underscores]_[item number]  Example: gia_2016_ia_48_01  Note: Include box number if appropriate.

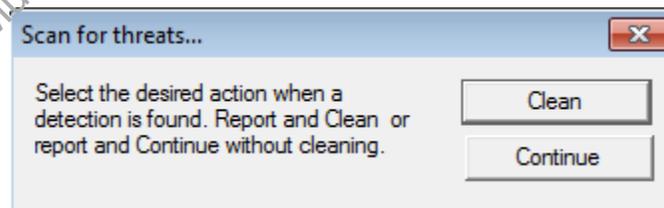
### III. VIRUS SCANS

Run a virus scan on all discs before imaging, extracting files, or examining the content of files. For certain discs, such as Getty event recordings, virus scans may not be necessary if we know the discs were created on virus-free computers. Keep in mind that although Getty computers are regularly scanned for viruses, we cannot always assume that discs transferred to us by staff were created on Getty computers. In addition, we have come across malware on staff computers that were not recognized by the antivirus program.

1. **Before you insert a disc, turn off the autoplay function on your computer:** A disc with a virus could potentially infect a computer if the disc is programmed to run automatically when inserted into the computer drive. To turn off the autoplay function, navigate to the control panel and select **AutoPlay**. In the window that appears uncheck the box next to “Use AutoPlay for all media and devices.” Click **Save** at the bottom of the window.



2. Insert disc and open Windows Explorer. Right-click the optical disc drive and select **Scan for threats. Do not open any of the files before you have verified that the disc is virus free.**
3. A message box will appear. Select **Continue** and the virus scan will begin.



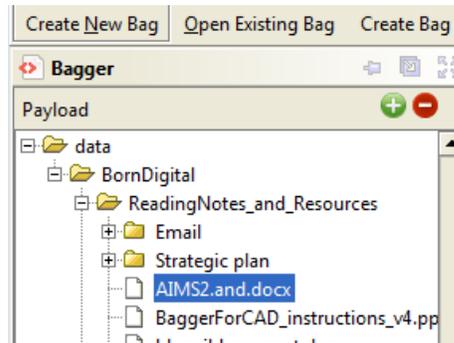
4. If one or more viruses are found, save a log of the infected files in the accession's documentation folder on ira\_locked according to the following format: [unique identifier]\_viruslog.

There are four options for dealing with infected files:

(1) Working on Fluffy (remember to turn off autoplay), extract files from the disc (see [section V.](#)) to an empty external drive and have the antivirus program clean the infected file. Make sure the antivirus software is up-to-date and that the Ethernet cord is disconnected.



(2) Use Bagger to exclude the infected file(s) from transfer. To do this simply click on the file in the payload and click  to remove from the list.



Note: You can also use FTK Imager if files on the disc are not organized in folders.

(3) If cleaning the infected file is not an option and it needs to be retained, create an image of the disc and only access the infected file through Forensic Toolkit.

(4) Do not accession the disc.

Before proceeding, consult with the Head of Institutional Archives.

5. Record virus scan results (even if no virus was found) and actions as a Virus Check event in the accessions record of ASpace. See [section VI](#).

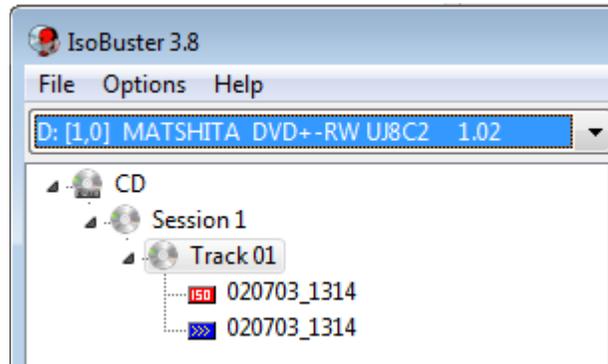
#### IV. CREATE DISC IMAGE

Creating a disc image of a CD or DVD will produce iso and cue files (or bin/cue files for multisession discs).

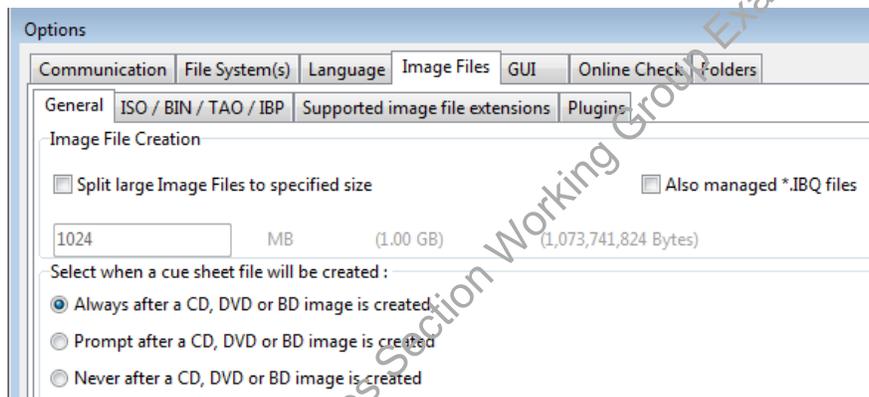
IsoBuster and FTK Imager are the two primary tools we use for creating disc images. FTK Imager has a higher success rate but breaks an image into multiple files if over 1024 MB. (Multi-part iso files are only a problem if we want to mount an image, as described in [section IV.C.](#), or burn the image to another disc.) As a general rule use IsoBuster first. If IsoBuster generates errors, use FTK Imager.

##### A. IsoBuster

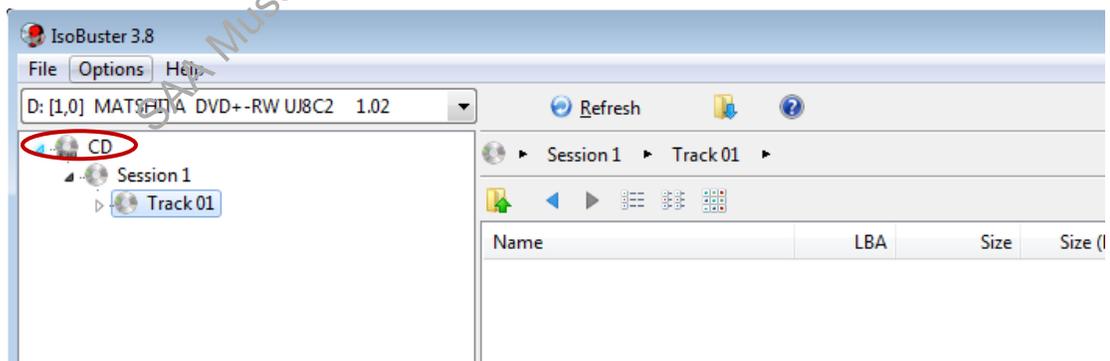
1. Load your source disc into your drive and startup IsoBuster. It should read the contents of the disc and display a tree view on the left pane. In case you do not get any such display, make sure that you have selected the correct optical drive in the drop-down box below the menu.



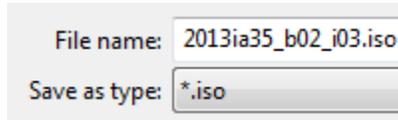
2. Now go to the menu – **Options > Image Files**. Under “Select when a cuesheet file will be created”, make sure that either **Always after a CD or DVD image is created** or **Prompt after a CD or DVD image is created** is selected. It is better to select **Always**.



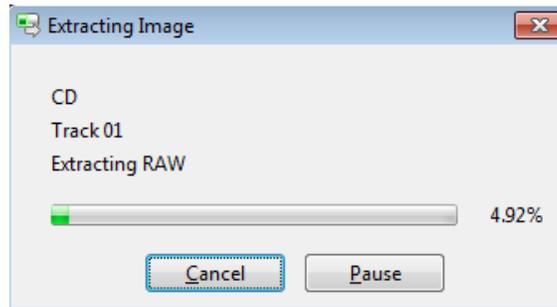
3. Click **OK** to exit dialog.
4. At the top of the tree in the left-hand pane is a small disc icon and CD or DVD-R or something similar based on the disc inserted into the drive. Right-click on this.



5. You will see a menu popup. If the disc is a CD, select **Extract CD (Image) > RAW (\*.bin, \*.iso)**. If it is a DVD, select **Extract DVD (Image) > User Data (\*.tao, \*.iso)**.
6. In the next window navigate to “[accession#]\_original” folder on ira\_locked as the save location. Enter the unique identifier with the extension .iso as the filename and click **Save**.

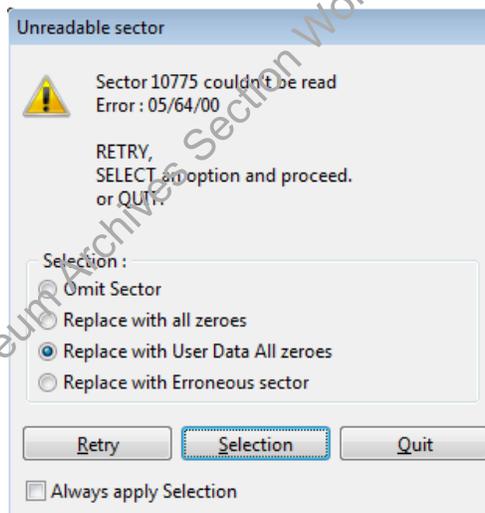


7. A progress bar for image extraction should appear.

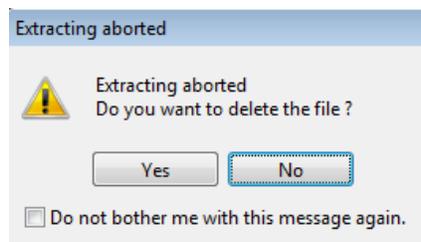


Once completed, the progress bar will disappear and a window will appear to save the \*.cue file. Use the unique identifier as the filename and click save.

If unsuccessful, a message will appear that there is an unreadable sector.



Click **Quit**. Click **Yes** in the next window to delete the file.

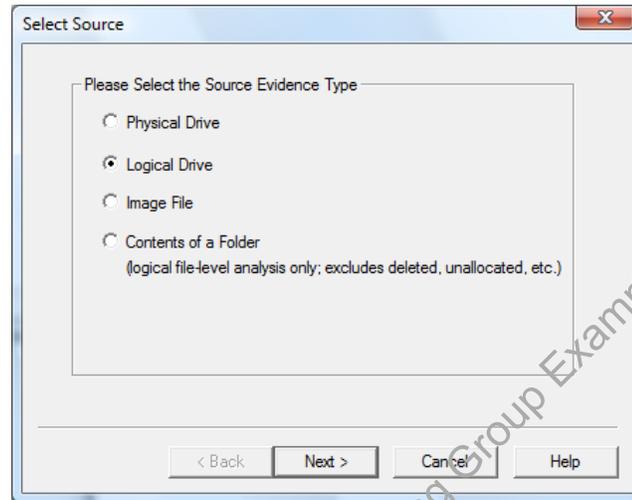


Try to image disc using FTK Imager.



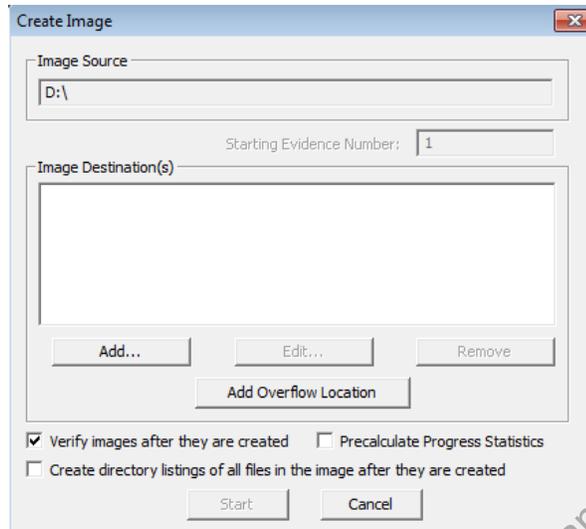
## B. FTK Imager

1. Click on Create Image in the file drop down menu.
2. You will be presented with the following source options:

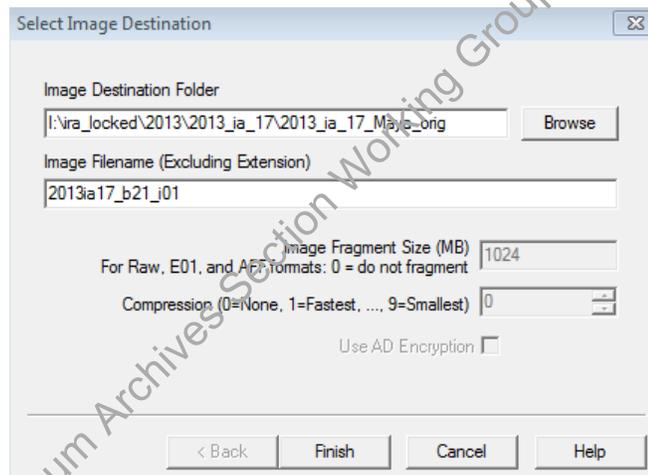


Select **Logical Drive** and click **Next**.

3. Select a drive to image. If imaging multiple discs in a single accession that have unique identifiers that are consecutively numbered, you may check the box next to Automate Multiple Removable Media. FTK Imager will automatically increment the evidence number with each image by adding “-000[#]” to the end of the filename. Note that while you cannot alter the number of leading zeros, you can adjust the starting evidence number. This is helpful if you are unable to complete imaging of all the discs in one sitting.
4. The Create Image screen will appear. If Automate Multiple Removable Media was selected in the previous window, change the starting evidence number if needed. Although you can check off **Verify images after they are created**, this feature does not work for optical discs.



5. Click **Next** and the Select Image Destination screen will come up.



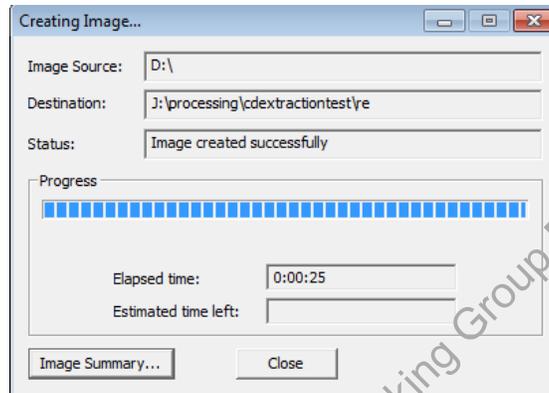
- a. Select the save location for the image file. If working on FRED or Fluffy, save the files on an external hard drive to facilitate file transfer to a networked computer. Make sure there is sufficient room on the external hard drive for the image. If imaging discs on a networked computer, save the files in “[accession #]\_original” folder on ira\_locked.
- b. Enter the unique identifier as the filename for the image. Do not include the extension.
  - If Automate Multiple Removable Media was selected earlier, include “i” at the end of the filename but **do not enter the item number**. FTK Imager will automatically add “-000[#]” to the end of the filename. We can use Renamer to remove “-” and the extra leading zeros from the filenames later. Note that if you decide to continue imaging discs within a set in a separate session, a message will appear that you may potentially overwrite an existing file. You can ignore this message if you’ve made the appropriate adjustment to the starting



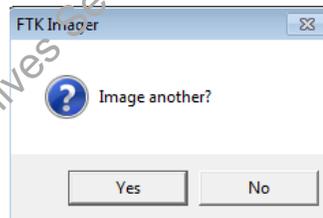
evidence number and confirmed that the file with that filename does not already exist, or if it does already exist, can be overwritten.

c. Click **Finish**.

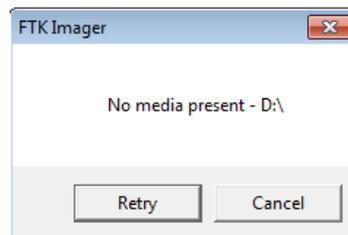
- You will return to the Create Image screen. There should now be a location listed under Image Destination(s). Click Start to begin imaging and a window with a progress bar will appear. The status message will change to “Image created successfully” once imaging has completed. You may click **Close**.



- If you checked off Automate Multiple Removable Media, a window will appear giving you the option to image another disc. Click **Yes**.

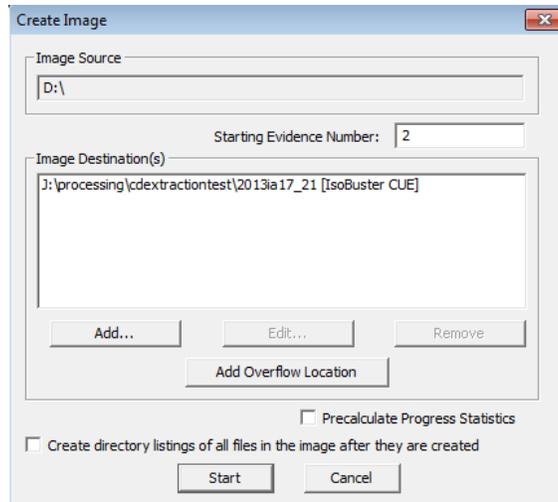


The drive will automatically pop open and a window will appear stating that no media is present.



Replace the disc with the next disc and click **Retry**.

The Create Image window will appear again. This time the starting evidence number should have gone up by one. The previous save location should also appear under Image Destination.



### C. Examine disc image

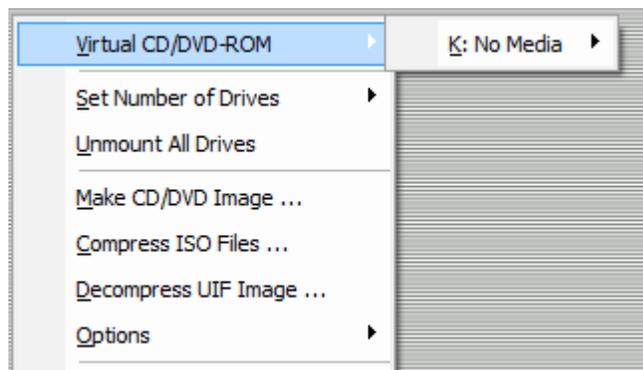
You will need special software to view the contents of your disc image. You may use FTK Imager or Forensic Toolkit, but the most user friendly method is to mount the iso file using software such as [MagicISO](#). (FTK Imager has a mount image feature but is unable to mount iso files.) MagicISO creates a virtual disc drive and allows you to run the iso file like an actual disc and view the contents through file explorer. Note: This will not work for iso files that are broken into multiple parts.

How to mount disc image using MagicISO

1. If Magiciso is installed on your computer, you should see a small icon for the software on your taskbar.



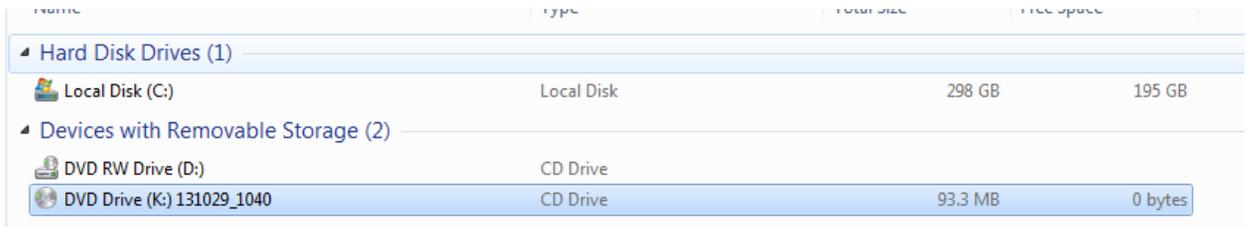
Right-click the icon and click on Virtual CD/DVD-ROM.



Select the available drive. The drive letter will vary from computer to computer. Click Mount in the next menu



2. Browse and select the iso file you want to mount and click OK.
3. Navigate to file explorer.

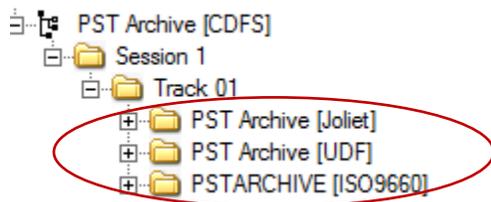


The image file should now appear mounted on the virtual drive. Double-click the drive and you should be able to see and open files within the disc image. You can also use QuickView Plus on the virtual drive.

## V. FILE EXTRACTION

### A. Which files to extract

When viewing the contents of a CD or DVD using software such as FTK Imager, Forensic Toolkit, or IsoBuster, you may see something like this:



As you click through each tree, you'll notice that the folder structure and files are the same, although the filenames may look slightly different. While it may appear that there are multiple sets of files on the disc, this is not the case. There is actually only one set of files but multiple file systems used to store information on the disc. The above square brackets indicate the file system names. While certain file systems can be only read by specific operating systems, they should all be readable using IsoBuster, FTK Imager, and Forensic Toolkit. (See the [appendix](#) if you would like more information about file systems on optical discs.)

Do not extract files at the Track 01 level or higher as that will result in multiple sets of the same files. Because filenames may be truncated in the earlier file systems, select the file system that has the least restrictive filenameing rules.

Use the following chart to determine which file system to extract:

Extract	File system
1 <sup>st</sup> choice	UDF
2 <sup>nd</sup> choice	Joliet
3 <sup>rd</sup> choice	ISO 9660
+ when present*	HFS



\*Extract HFS when present along with one of the other three file systems if you want to retain compatibility to MacOS.

In this example, you would extract folder PST Archive [UDF]:



If your disc contains file systems that are not covered in the above chart, examine and compare the file and folder names of each file system folder. Select the file system folder with the most complete names.

### B. Software

There are three different tools that we use to extract files from an optical disc or disc image:

**FTK Imager:** Fastest method to extract files from disc, but there is no option to verify that the files extracted from the disc match the original files.

**Forensic Toolkit:** You will primarily use this software when you are processing the files (i.e. searching for sensitive information and weeding non-archival files). Forensic Toolkit is only available on FRED and Fluffy.

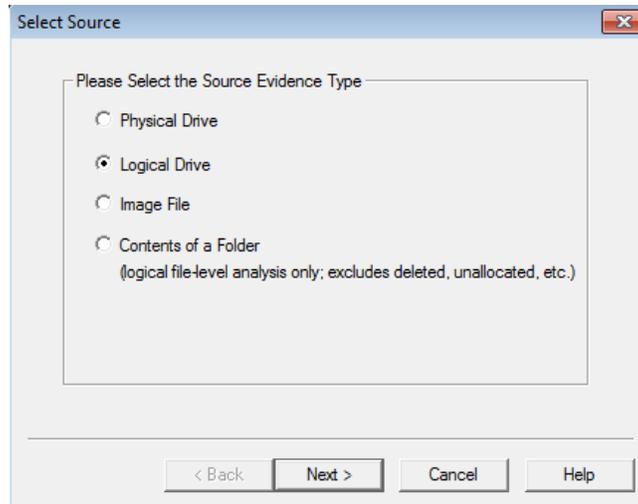
**Bagger:** Requires significantly more processing time than FTK Imager and Forensic Toolkit to copy files. This is due to Bagger calculating checksums of the original files and the copied files. While Bagger takes much longer, it ensures that exact copies of the files are made and generates documentation. The other software do not provide that kind of verification, or at least are not transparent about it. Images need to be mounted in order to use Bagger.

**Note:** It is possible to use IsoBuster to extract files, but we currently use the free version which does not allow extraction of UDF files.

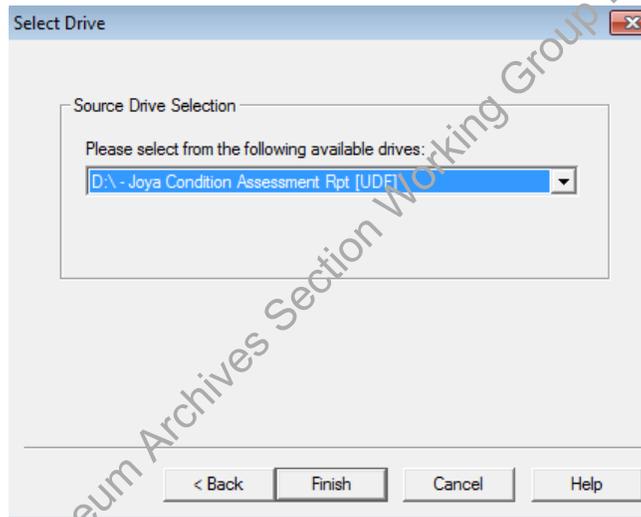
In most cases you will use FTK Imager or Bagger to export files. Use your best judgement in deciding which to use. While Bagger provides an added layer of verification, it may not be practical if you are dealing with a large number of discs.

### FTK Imager

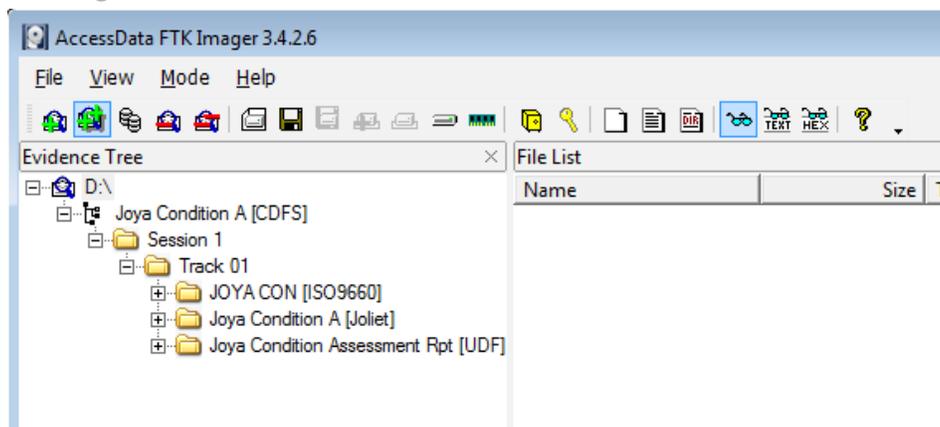
1. Click **File**→**Add Evidence Item** or click  in the toolbar.
2. Select **Logical Drive** if you want to extract files from a disc. Select **Image File** if you want to extract files from a disc image (ISO file).



3. Select the appropriate disc drive or image file and click **Finish**.

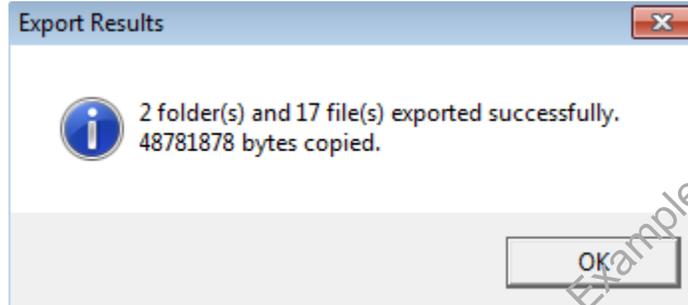


4. The contents of the optical disc or disc image should now appear in the Evidence Tree pane.





- Right-click the folder below Track 01 that you want to extract. (See [here](#) if you do not know which folder to extract.) Select Export Files and in the next window select the location to save the files. If working on FRED or Fluffy, save the files on an external hard drive and use Bagger to transfer to a networked computer. If working on a networked computer, save the files in “[accession #]\_original” folder on ira\_locked.
- Once export is successfully completed, you should see something like this:



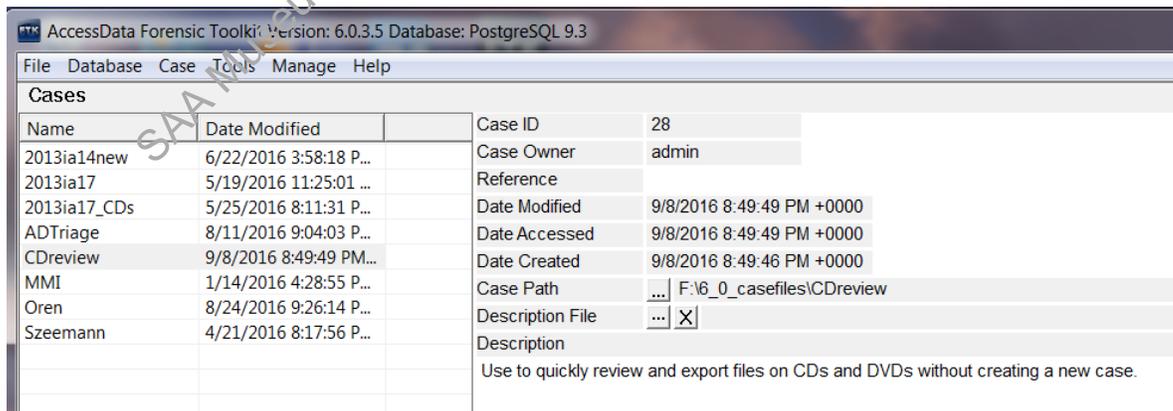
- Maintain exported files on ira\_locked in bag.

### Forensic Toolkit

(See section III.B. of [IA Electronic record accessions.pdf](#) for more thorough instructions on using Forensic Toolkit.)

- Create a new case to examine files more thoroughly and flag non-archival files and files with sensitive information. (For quick review and file extraction select CDreview. In the next window navigate to **Evidence** in the toolbar and select **Add/Remove** and skip to step 3.)

Navigate to toolbar and click on **Case**.



- Select **New** and the New Case Options window will appear.

Fill the following fields:

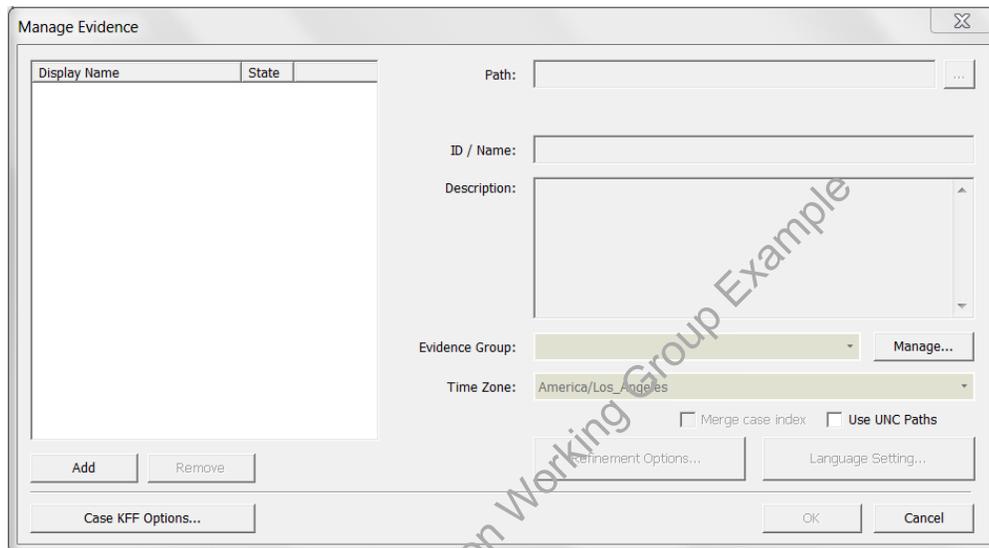
**Case Name:** Accession number



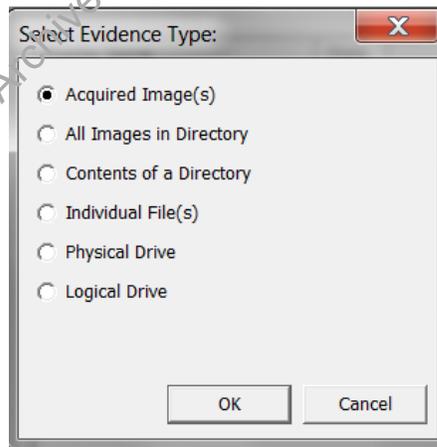
**Processing Profile:** IA default

Click **OK**.

3. In the Manage Evidence window click **Add**.

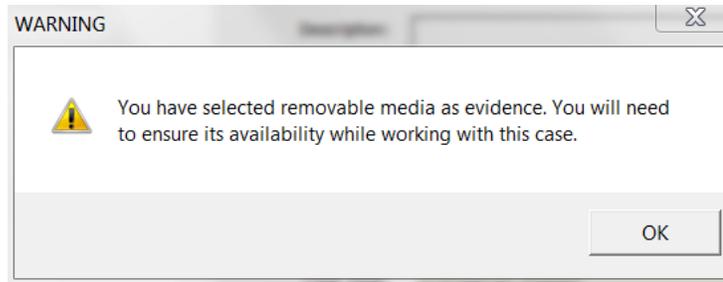


Select from the following evidence types and click **OK** to select the appropriate folder/drive.



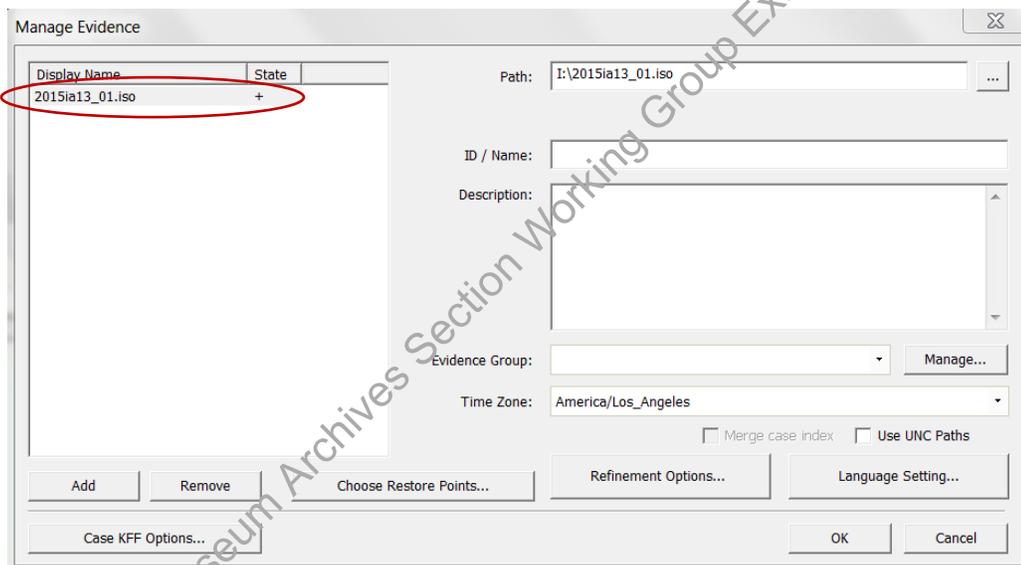
**Acquired Image:** Select this to view the contents of a disc image that you created or obtained.

If you select an image file on a USB device, click **OK** in the following window.

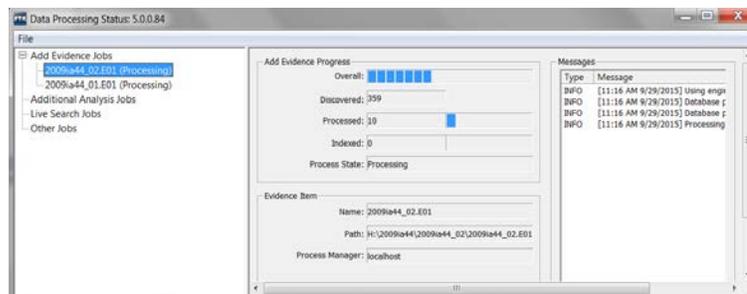


**All Images in a Directory:** Select this if there are multiple disc images in a single accession/collection that you want to view together. (See step 8. for note regarding exporting files from multiple images.)

4. The image file(s) should now appear in the Manage Evidence window. Click **OK**.



5. A Data Processing Status window will appear.



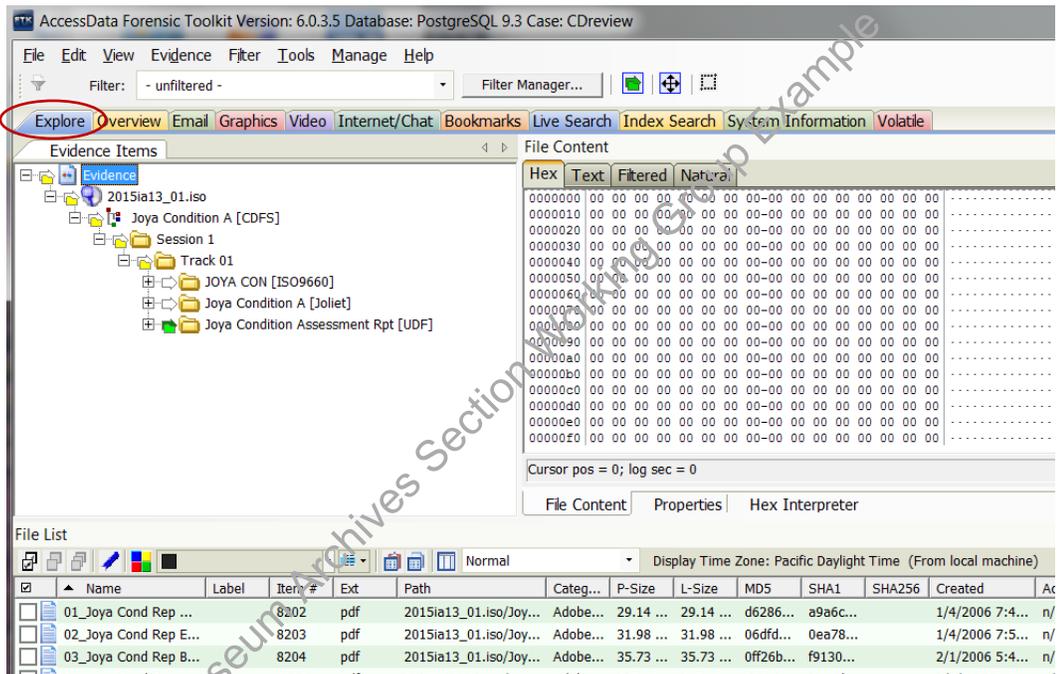
Processing time will vary depending on the size of the image and may possibly take over an hour. It should only take a few minutes at most, however, for files to load in FTK. While you will



be able to examine the contents of files before processing is completed, do not conduct index or live searches or export files until then.

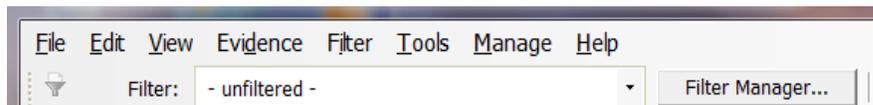
Once the processing job is done, you can search for documents with sensitive information or non-archival files to weed (exclude from file export).

- 6. See section III.B. of [IA Electronic record accessioning.pdf](#) for guidance on reviewing, marking, and filtering files. Keep in mind that if there are multiple file systems, duplicates of the files will appear. Work with files from only one file system. (See [here](#) to determine which file system to work with.) Make sure only the icon next to that file system folder is green (🟢). (See “Explorer tab” in section III.B. of [IA Electronic record accessioning.pdf](#))

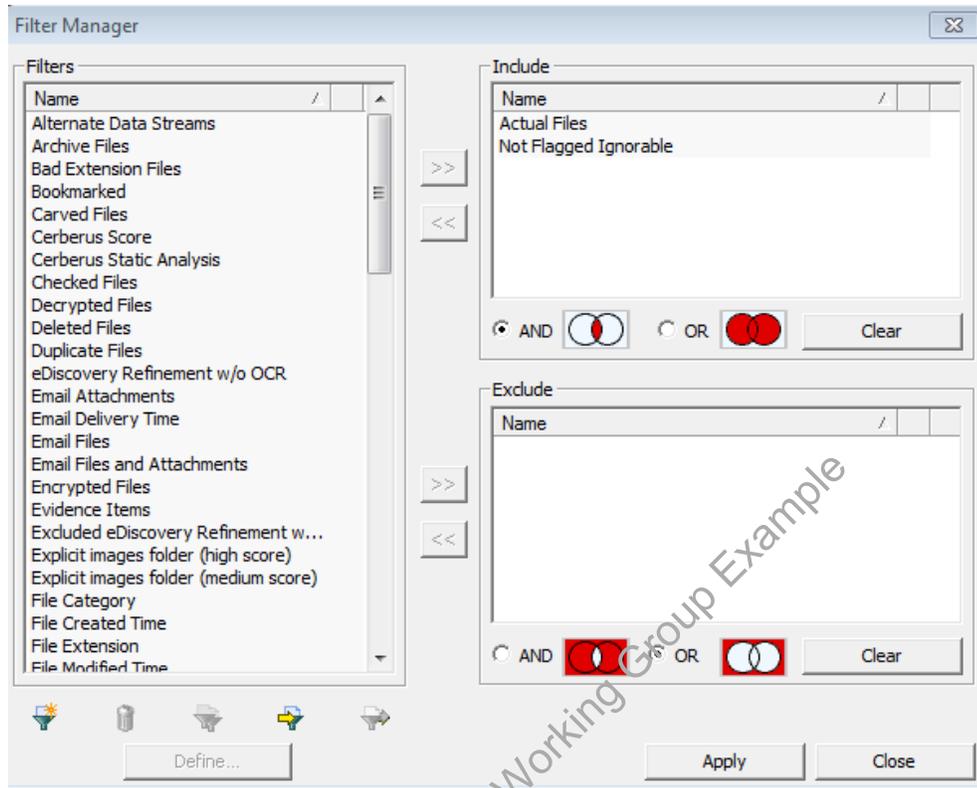


- 7. To export files, navigate to the Explorer tab. Make sure that all files that you want to export are currently displayed in the File List pane.

If you did not mark any files to ignore, select **Actual Files** in the filter drop-down menu.



If you did mark files to ignore, use the filter manager to include **Not Flagged Ignorable, Actual Files**, and any other filters.

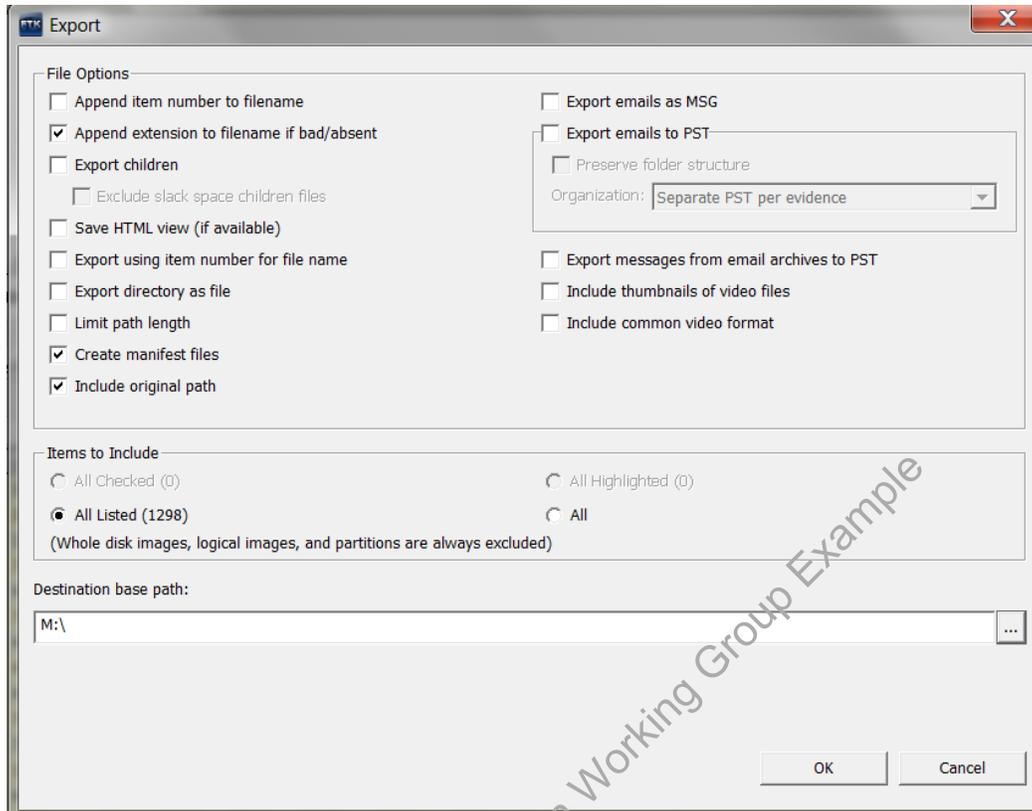


You can also deselect the  icon next to a file or folder to exclude from export.

8. Once you have finished adjusting the filters so that only the files to be exported are displayed in the File List pane of the Explore view, go to **File** in the menu bar and click **Export**.

Note: Disc image names are not exported with files. Associating disc images with exported files is therefore difficult when extracting files across multiple disc images within a case in a single export. If you need to maintain the connection between the disc image and exported files, either export files from each image, one by one, or add each image as a separate case. In most cases we do want to maintain the connection between the two.

9. In the Export Window, check off boxes as in image below.

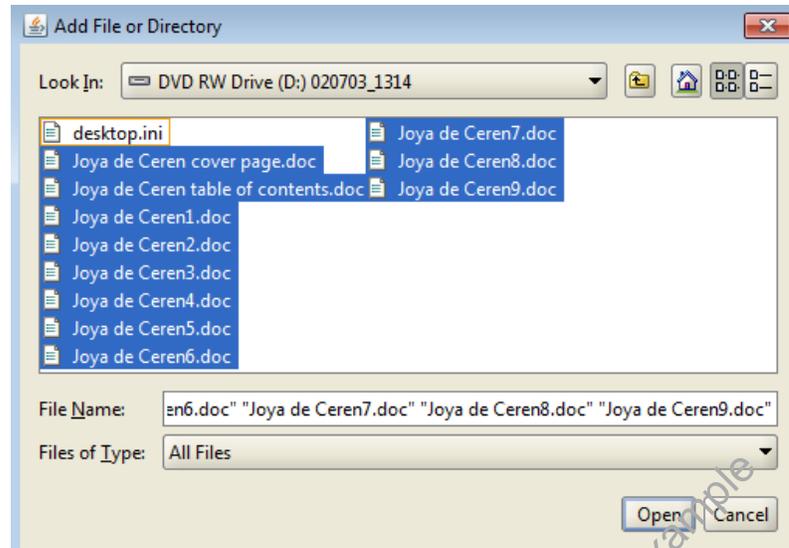


Under Destination base path select location to save the exported files, ideally to an external drive. Create a new folder, using the disc's unique identifier as the folder name. If you need to save to FRED, save on Data Drive. Click **OK**.

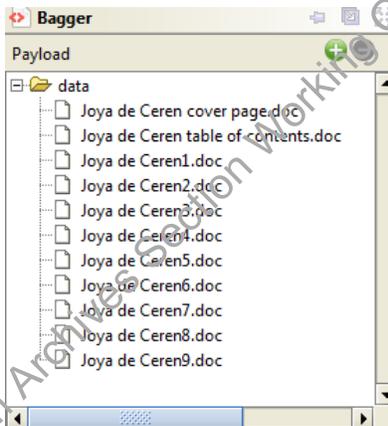
10. Once you have exported the files, you will need to transfer them from the external drive to "[accession #]\_original" folder on ira\_locked using Bagger. Validate the bag and maintain files in the bag.

## Bagger

1. Follow the instructions outlined in steps 2.a. through 2.d. of [Bagger User Guide.pdf](#) to create a new bag.
2. In the file browser dialog box navigate to and select the disc drive and click **Open**.
3. Select the file(s) and/or folder(s) to copy. To select multiple files hold the Ctrl key and click the files. To select all press Ctrl+A. Make sure to exclude desktop.ini if the file is present. Click **OK**.



4. The files or folders should now appear under “data” in the **Payload** pane.



5. Follow the rest of the instructions in [Bagger User Guide.pdf](#) to save the bag. Use the unique identifier for the bag name.
6. Validate the bag to verify files were properly transferred.
7. Maintain files in bag.

### C. Post file extraction

All exported files should be maintained in a bag in “[accession #]\_original” folder on ira\_locked. Maintain files in bags at the level that is most practical. This will vary from bags at the disc level to a single bag for the entire set of discs.



## VI. DOCUMENTATION

In ASpace use the Digital Files Management Notes field under the User Defined section of the accession record to document the work you’ve completed, work that needs to be done, and any known issues or problems. Your notes should be clear enough for someone to be able to pick up from where you left off.

Create a spreadsheet if you need to transcribe labels or document other information about each disk that might be confusing to track in ASpace. You may use [Imagingsummary\\_sample.xls](#) on [ARJ1\PRD-ARJ\ira\\_locked\BornDigital](#) as a model. Add or remove columns as necessary. Save the file under the accession’s documentation folder and make sure to reference the spreadsheet in the Digital Files Management Notes field.

Record virus scan results (even if no viruses were found) and actions as a Virus Check event in the accessions record of ASpace. Select the appropriate outcome in the drop-down menu and fill out the Outcome Note if any viruses were found. Include a link in the External Documents section to the infected files logs in the “[accession #]\_documentation” folder on ira\_locked. If you run a virus check on only some of the discs within the accession, specify in the Outcome note so it is clear which discs have been scanned.

**New Event** Event

**Basic Information**

Type \*

Outcome

Outcome Note

## VII. HANDLING OF DISCS POST-DIGITAL CAPTURE

Institutional Archives views discs as physical carriers that, in most cases, do not hold artefactual value. Once we have captured and preserved the digital content off of the discs, the discs can be placed in an Alt Media shred box for destruction. There are, however, certain circumstances in which we may decide to retain discs. Discs with custom labeling and inserts, for example, may warrant retention. Consult the Head of Institutional Archives as necessary.

## VIII. APPENDIX: UNDERSTANDING COMMON FILE SYSTEMS ON OPTICAL DISCS



**ISO 9660** is the original file system standard for CD data discs and was first published in 1988. It is sometimes referred to as CDFS (Compact Disc File System), not to be confused with the virtual Linux file system (CDFs). The standard was developed to enable multiple computer operating systems to read files on a disc. Three “levels of interchange” are described in the standard. Level 1 provides compatibility with the largest number of operating systems but is also the most restrictive in terms of file and directory name rules. Directory and filenames are limited to eight characters, with a three-character extension for filenames, which is in accordance with MS-DOS’s restrictive file naming rules. Original file and directory names that exceed eight characters will appear abbreviated with a “~”. Characters for filenames can also only contain uppercase letters, digits, or an underscore. Levels 2 and 3 have the same limitations of character type for filenames but allow for 30 characters in directory and filenames.

**Joliet** is an extension to the original ISO 9660 standard and was developed in 1995 by Microsoft for Windows 95 and later Windows operating systems. Joliet supports longer filenames, up to 64 characters, as well as spaces and Unicode characters (this includes diacritics and non-Latin script). Joliet is backwards compatible because filenames are saved in a supplementary volume descriptor that is ignored by ISO 9660. Thus you will see both ISO 9660 and Joliet directory and filenames listed in separate trees on the disc.

**UDF (Universal Disk Format)** is a file system standard that was introduced in 1995 and has since replaced ISO 9660. It is widely used on DVDs. A major advantage of UDF is that it can be used for packet writing technology – basically files can be created, modified, or deleted on a disc like on your local computer hard drive without burning an entire disc. Files can be dragged and dropped or copy and pasted to the CD using UDF-compatible applications. UDF can also support filenames up to 255 characters. Older optical drives and operating systems are unable to read UDF formatted discs and may display filenames in the Joliet format. Discs with UDF often contain ISO 9660 to allow for backwards compatibility.

**HFS (Hierarchical File System)** is a Macintosh file system released in 1985. The character limit for filenames is 31. PCs are unable to read HFS formatted discs, although it is possible using Isobuster or FTK Imager. You will sometimes see hybrid discs with both HFS and ISO 9660/Joliet so that the discs are readable on PCs and Macs. HFS should be exported if it is important to maintain compatibility with MacOS.

## References

<http://www.avpreserve.com/wp-content/uploads/2014/04/OpticalMediaPreservation.pdf>

<https://books.google.com/books?id=iw7yCQAAQBAJ&printsec=frontcover#v=onepage&q&f=false>