

# **Instructions for BDPL Staff**

## **for Using the BDPLinventory Script**

by Luke Menzies

Last Updated: 10/28/2016

These Instructions can also be found at:

<https://wiki.dlib.indiana.edu/display/DIGIPRES/BDPLinventory+Script+for+Inventory+and+Checksums+of+Large+Media>

## Contents

Having a Donor Run the BDPLinventory Script.....	3
Running BDPLinventory on the BDPL Computer.....	5
Running BDPLinventory on an Item in Karst (or another IU server).....	7
Running BDPLinventory on an Item in Box (using WebDav).....	8
What the Script Actually Does.....	10

# Instructions for Having a Donor Run the BDPLinventory

## **Step 1: Have the administrator for the collection (hereafter "donor-admin.") download and run the BDPLinventory script**

### -1- Have the donor-admin. download the script

- Find out what operating system they're using, and either email the appropriate script as an attachment, or have them download it from Box.
- Also send them a copy of the "00Instructions" file (as .pdf, .docx, or .odt), located in the bdpl Box account, and on the BDPL computer:  
(BDPL computer) ~/Documents/scripts/BDPLinventory/00Instructions.pdf  
(BDPL Box) /Scripts/BDPLinventory/00Instructions.pdf
- In most cases, they will just double-click the executable file.

### -2- If the script does not work:

- One point of confusion on a Mac is that the the requests, "What folder do you want to analyze?", "In what folder will the results be placed?", etc. are NOT displayed. The user will just have to remember (as per the instructions) that the first folder is the one to be analyzed and the second folder is the destination folder.
- On Windows it is possible that the computer is missing the Windows Visual C Runtime Environment. If this is not present, you can have the donor-admin. download it and install it herself/ himself:  
<http://www.microsoft.com/en-us/download/details.aspx?id=29>  
Or, you can create an installer or package that will correct the problem, according to this site:  
<http://py2exe.org/index.cgi/Tutorial#Step5>
- Be sure to ask the donor-admin. to email you a screenshot or quote of the error message, verbatim, and to describe what happens in as much detail as possible. Also, find out exactly which version of the operating system they are using (ex. OSX 10.10.2).

## **Step 2: Get the Results from the Donor-Admin.**

### -1- Have the donor-admin. email the output file and the comparison file to you.

- the filenames will be something like this: BDPL\_Results.csv CompareMMDD\_HHMMSS.txt

### -2- Place both files into the appropriate folder in 1received.

- (i.e. ~/Documents/1received/UAC0920160003)
- Rename the .csv file:  
<BDPL accession number>-inv.csv (ex. UAC0920160003-inv.csv)
- If you have multiple Output.csv files, number them:  
<BDPL accession number>-inv<no.>.csv (ex. UAC0920160003-inv2.csv)
- Do Not rename the Comparison file.

### -3- Remember these filenames and their locations.

- When you run BDPLinventory on the BDPL computer, you will enter the name of the new .csv output file.  
(ex. UAC0920160003-inv3.csv).
- The new Comparison file will be named automatically.

### **Step 3: Run BDPLinventory on Transferred Materials and Compare Results.**

#### **-1- Transfer donor materials according to the proper BDPL protocols and run the script.**

- For Transfer instructions, see the Non-Disk Accession Procedure, "Step 5: Transfer."

<https://wiki.dlib.indiana.edu/display/DIGIPRES/Accession+Procedure+for+Non-Disk+Items>

- For running the BDPLinventory script on the BDPL computer, on Box (via WebDav), and on Karst (or other IU server), see instructions below, or on the wiki:

<https://wiki.dlib.indiana.edu/display/DIGIPRES/BDPLinventory+Script+for+Inventory+and+Checksums+of+Large+Media>

#### **-2- Open the output .csv file(s) and check its contents visually.**

- When opening the .csv, the "Character Set" should be "Unicode UTF-8", "Separator Options" should be set to "comma" only, and "Text Delimiter" should be double-quotes.
- Note how many files are listed.
- DO NOT save any changes to this file.

#### **-3- Compare the two "Comparison" .txt files**

- From the command prompt, navigate to the 1received folder and use the diff command:  
cd ~/Documents/1received/<BDPL accession no.>/  
diff <comparison file 1> <comparison file 2>  
(ex. diff Compare1028\_112904.txt Compare1028\_113244.txt)
- If the comparison files differ, see trouble-shooting tips under Step 7 of the Accession Procedure for Non-Disk Items (on the Wiki).

# Instructions for Running BDPLinventory on the BDPL Computer

## Step 1: Assign the BDPL accession number and create the appropriate folder in Ireceived.

- Accession Number Creation Guidelines:

<https://wiki.dlib.indiana.edu/display/DIGIPRES/Accession+Number+Creation+Guidelines>

### Step 1b: Mount an External Disk

-1- If you want to run the script on an external disk (flash drive, hard drive, etc.), you will need to mount the disk.

- If the disk appears in the left-hand sidebar, to mount it, just click on the icon

-2- If the disk does not appear in the left-hand sidebar, you must mount it manually.

- Mount:

`sudo mount <path to the disk> <path to the mount point>`

ex. `sudo mount /dev/sdb1 ~/Documents/MountPoint/`

- For convenience use the mount point: ~/Documents/MountPoint/

- You may need to specify the filesystem type, using the "-t" option:

ex. `sudo mount -t exfat /dev/sdb1 ~/Documents/MountPoint/`

## Step 2: Run the BDPLinventory script.

-1- Versions of the BDPLinventory script are currently in ~/Documents/scripts/BDPLinventory/, as well as on Box.

- Either double-click on BDPLinventoryLinux or use the Terminal to run BDPLinventory102016.py:

`cd ~/Documents/scripts/BDPLinventory/`

`python BDPLinventory102016.py`

-2- The script will ask for two folders:

- (a) the one to be analyzed and
- (b) the one where the results will go (~/Documents/Ireceived/<BDPL accession number>/).

\*Note: If you are running BDPLinventory, and you want to generate temporary output files, you may use ~/Desktop/Eraseme/ as the destination folder, and delete the files later.

-3- Then it asks you to enter the name of the .csv output file.

- The standard format for the output files is:

`<BDPL accession number>-inv<version number>.csv` (ex. UAC0920160004-inv2.csv)

- If the .csv file does not exist, the script will create it.

- Do not overwrite an existing .csv. When establishing the final output .csv, simply take the most recent output .csv and rename it by removing the version number.

(ex. UAC0920160004-inv2.csv => UAC0920160004-inv.csv)

\*Note: If you have any doubts about what to enter, you may enter a fake name for the .csv to avoid overwriting any existing file, and then change the name later.

- 4- When the script is finished, a message will appear that says, "All Done!".  
- Until that happens, the script is still running.

**Step 3: Check the output .csv and comparison files, according to the Accession Procedure for Non-Disk Items (Step 7)**

See => <https://wiki.dlib.indiana.edu/display/DIGIPRES/Accession+Procedure+for+Non-Disk+Items>

- This also contains trouble-shooting tips, if the comparison files differ.

## **Instructions for Running BDPLinventory on an Item in Karst (or another IU server)**

### **Step 1: Mount the Karst directory at a local mount point on the BDPL computer, using sshfs**

-1- SSHFS has already been installed on the BDPL computer.

- If not, open the command prompt and type:  
sudo apt-get install sshfs

-2- Find the correct path on Karst (or another server) by logging in with ssh:

- login to the bdpl Karst account (or other account)  
ssh <username>@karst.its.iu.edu (ex. ssh bdpl@karst.its.iu.edu)
- find out the path to the 'home' directory  
pwd (this command tells you which directory you are currently in)
- exit the Karst account  
exit

-3- Mount the Karst directory using sshfs:

- use sshfs to mount the Karst account, from a local mount point  
sshfs <user@host>:<path to home directory> <path to local mount point>  
ex. sshfs bdpl@karst.its.iu.edu:/N/u/bdpl/Karst/ /home/bcadmin/Documents/MountPoint/

### **Step 2: Run BDPLinventory**

-1- Double-click the Linux version of BDPLinventory in ~/Documents/scripts/BDPLinventory/

-2- Treat the contents of ~/Documents/MountPoint/ as though they are located on the BDPL computer.

- See Above, "Running BDPLinventory on the BDPL Computer" or the Wiki:  
<https://wiki.dlib.indiana.edu/display/DIGIPRES/Instructions+for+Running+BDPLinventory+on+the+BDPL+Computer>

### **Step 3: Check the output .csv and comparison files, according to the Accession Procedure for Non-Disk Items (Step 7)**

See => <https://wiki.dlib.indiana.edu/display/DIGIPRES/Accession+Procedure+for+Non-Disk+Items>

- This also contains trouble-shooting tips, if the comparison files differ.

## Running BDPLinventory on an Item in Box (using WebDav)

### Step 1: Donor must share folder on Box with BDPL staff.

- 1- Once donor has shared the folder with staff, it will show up in the BDPL Box account.
  - The donor should share "read" permissions with BDPL staff. "Write and "execute" permissions are not necessary.

\*Note: Donors may NOT place any sensitive data or information on Box!  
=> see: <https://kb.iu.edu/d/bfirt> for more information.

### Step 2: Lookup the External Password for the Box Account.

- 1- If you are using the BDPL Box account, ask Staff, or check the Wiki for the *external* password.
  - This password is different from the regular BDPL login password.
- 2- If you are using a personal Box account, you will need to create an external password.
  - This is so that personal IU login information will not be transmitted over the internet.
  - To do this, see: <https://kb.iu.edu/d/bccp>

### Step 3: If, for some reason, you need to install WebDav on the BDPL computer.

- 1- WebDav is already installed on the BDPL computer, if not:
  - Install WebDav:  
`sudo apt-get install davfs2`
- 2- Configure the fstab file to automatically connect to Box.
  - Find out what the current Box WebDav server is.  
See:  
<https://community.box.com/t5/Managing-Your-Content/Does-Box-support-WebDAV/ta-p/310>
  - Currently, it is: <https://dav.box.com/dav>
- 3- Create a local mount point.  
`mkdir ~/Documents/box.com`
- 4- Edit the fstab file.
  - Enter a super-user shell:  
`sudo -s`

**\*Important Note:** You should "never" run commands as root because you could accidentally botch up your whole system. However, for the sake of privileges, it is sometimes necessary. Be sure to exit the root shell by typing "exit" whenever root privileges are not absolutely necessary.

- Open the fstab file in gedit  
`gedit /etc/fstab`
- Enter the following into the fstab file, and Save:  
`https://dav.box.com/dav /home/bcadmin/Documents/box.com davfs rw,user,noauto 0 0`
- The first element is the WebDav server. The second is the mount point you created.



- Quit gedit, and exit the super-user/ root shell. (Click the X in the corner of the window, then type "exit" into the command line.)

#### **Step 4: Mount the BDPL Box account and navigate to the folder you wish to analyze.**

##### -1- Navigate to the proper Box folder.

- This is an easy way to ensure that you have properly mounted box.com, and you know the location of the folder you want to analyze.  
    mount ~/Documents/box.com  
    cd ~/Documents/box.com
- Here, ~/Documents/box.com is the current location of the box mount point. (This can be changed in the methods described in #3 above)

##### -2- Enter the user name and password.

- It will ask for the Box username, including "@indiana.edu" and the *external* password.  
    (ex. bdpl@indiana.edu)
- Note the title and location of the donor's folder.

#### **Step 6: Run the BDPLInventory script.**

##### -1- Double-click the Linux version of BDPLInventory in ~/Documents/scripts/BDPLInventory/

##### -2- Treat the contents of ~/Documents/box.com/ as though they are located on the BDPL computer.

- See Above, "Running BDPLInventory on the BDPL Computer" or the Wiki:  
    <https://wiki.dlib.indiana.edu/display/DIGIPRES/Instructions+for+Running+BDPLInventory+on+the+BDPL+Computer>

#### **Step 7: Check the output .csv and comparison files, according to the Accession Procedure for Non-Disk Items (Step 7)**

See => <https://wiki.dlib.indiana.edu/display/DIGIPRES/Accession+Procedure+for+Non-Disk+Items>

- This also contains trouble-shooting tips, if the comparison files differ.

#### **Step 8: Un-Mount the Box account.**

##### -1- Un-mount.

- sudo umount ~/Documents/box.com/  
\*Note: The command is "umount", not "unmount" as one might expect.
- When you unmount, WebDav will sync your Box files with the BDPL computer.

##### -2- If you do not properly un-mount:

- If, for some reason, you do not properly un-mount Box, you will need to re-mount it and properly unmount it.
- Otherwise, you may have an excess of temporary files, stored on the BDPL computer, that have not been sync'ed.
- If this happens, you may find a temporary file in the "lost+found" folder on box.com.
- Delete it, and allow Box to sync with the BDPL computer.

## What the Script Actually Does:

1. Using Tkinter askdirectory, it asks the user for two folder paths: the object folder and the destination folder.
2. Asks the user for the name of the Output.csv
3. Imports os and uses os.walk to "walk" the contents of the object folder
4. Uses os.stat to get statistics on each file contained in the object folder
5. Uses python-magic to find out the mime type, based on the first lines or bytes of the file
6. Generates a md5 checksum for each file, by opening the file and reading it byte-by-byte
7. Writes a comma separated value (csv) file that contains this information.
8. Creates a comparison file, which contains only filenames and md5 checksums, separated by a comma and a line break (so, it is possible to change the extension from .txt to .csv, and open this as a spreadsheet).
9. The comparison files are titled according to the time when the script was run. While the titles will be different, they should all be identical.
10. The script *should* also include hidden files (i.e. ones that begin with ".") and files that have improper file names.