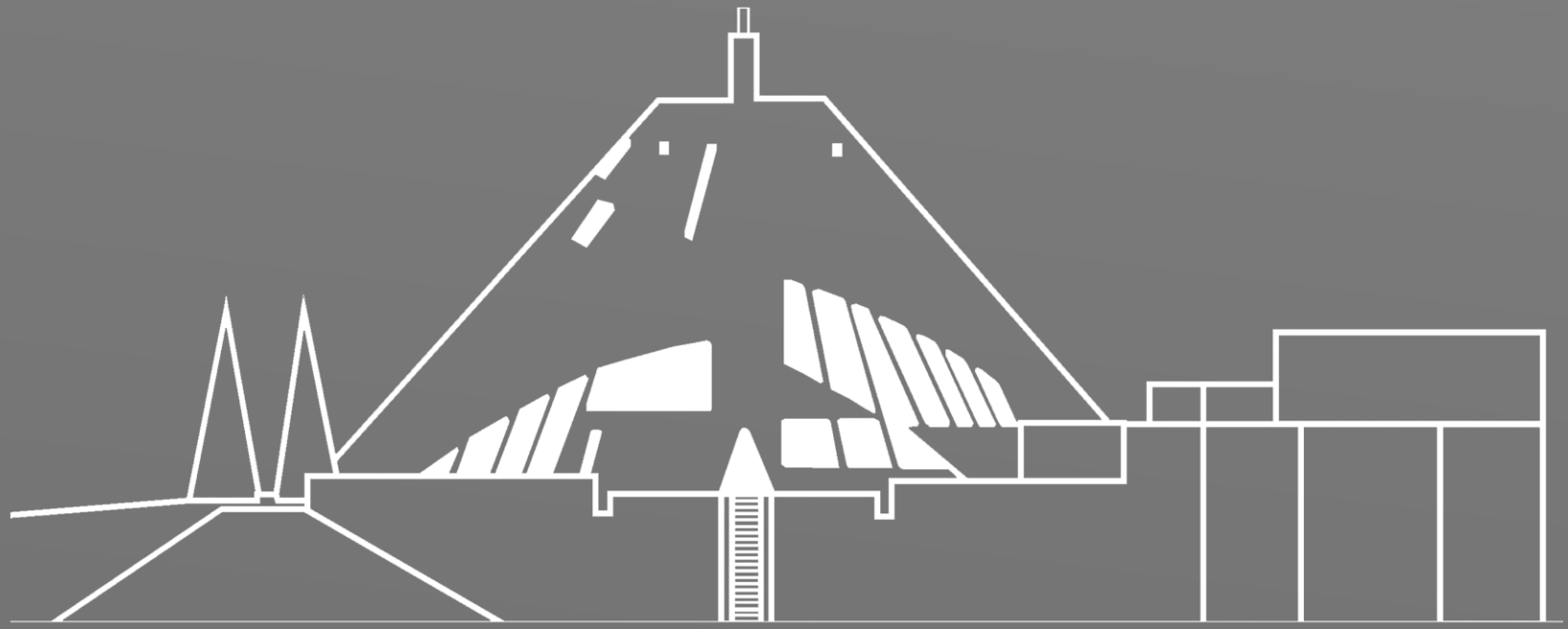
BORN-DIGITAL PRESERVATION BASIC PRINCIPLES AND PRACTICES



Classroom Building 215, University of Wyoming

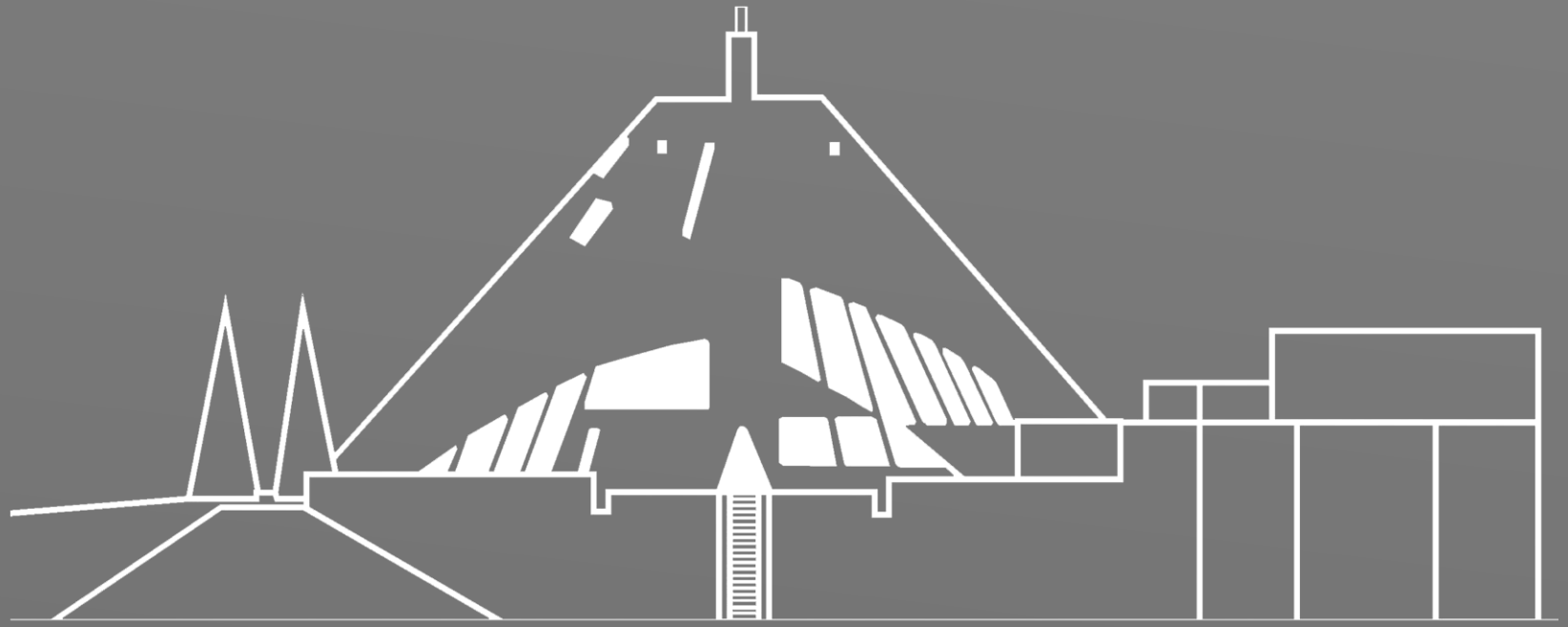
Tuesday, March 7 3:30 PM

Tyler G. Cline
Digital Archivist and Head of Digital Programs
UW American Heritage Center

tyler.cline@uwyo.edu

BASIC PRINCIPLES AND PRACTICES

- Identifying physical media
- Identifying file systems and file types
- Working with older hardware
- Using preservation tools



INTRODUCTION What are born-digital records?

DEFINING BORN-DIGITAL RECORDS

born dig·it·al (adjective)
digital information
originally created
in electronic form

Source: SAA Dictionary

- Created electronically
- Not human-readable
- Hardware and software dependent
- Non-linear (many ways to order the records)

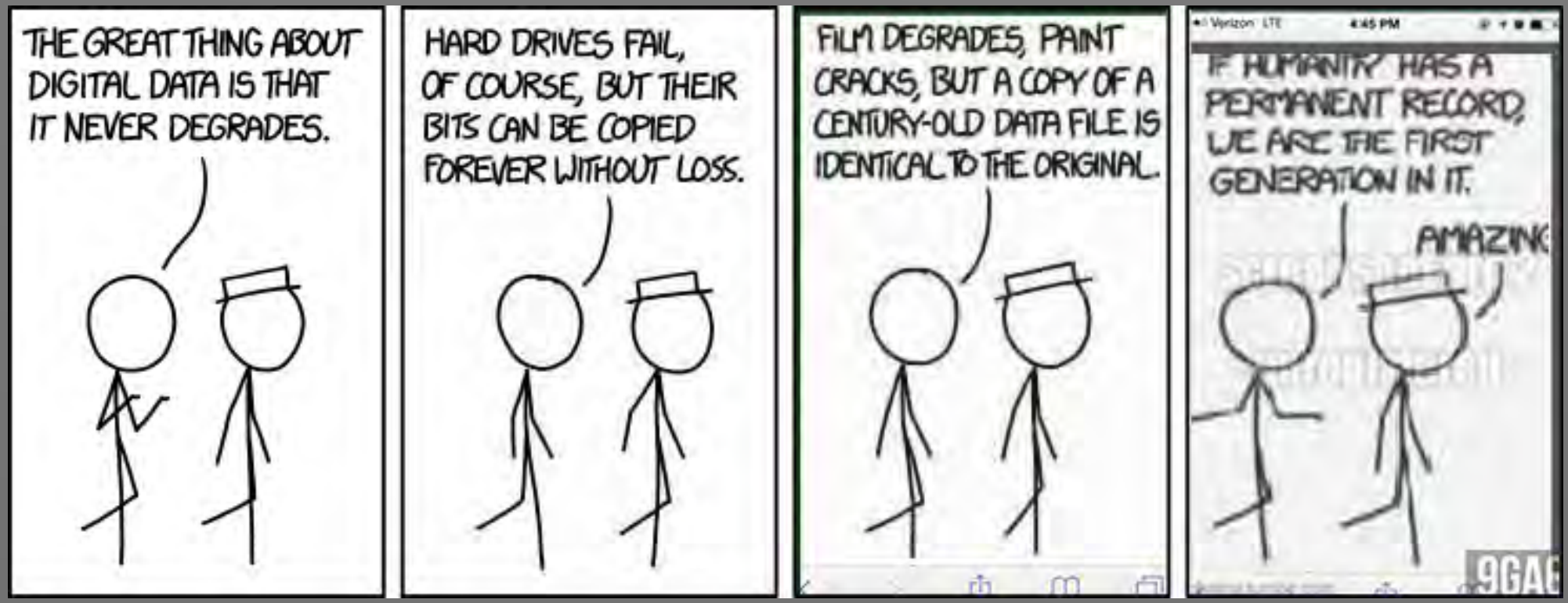
DIFFERENCES FROM PHYSICAL RECORDS

PHYSICAL:

Generally high longevity when stored in climate-controlled conditions

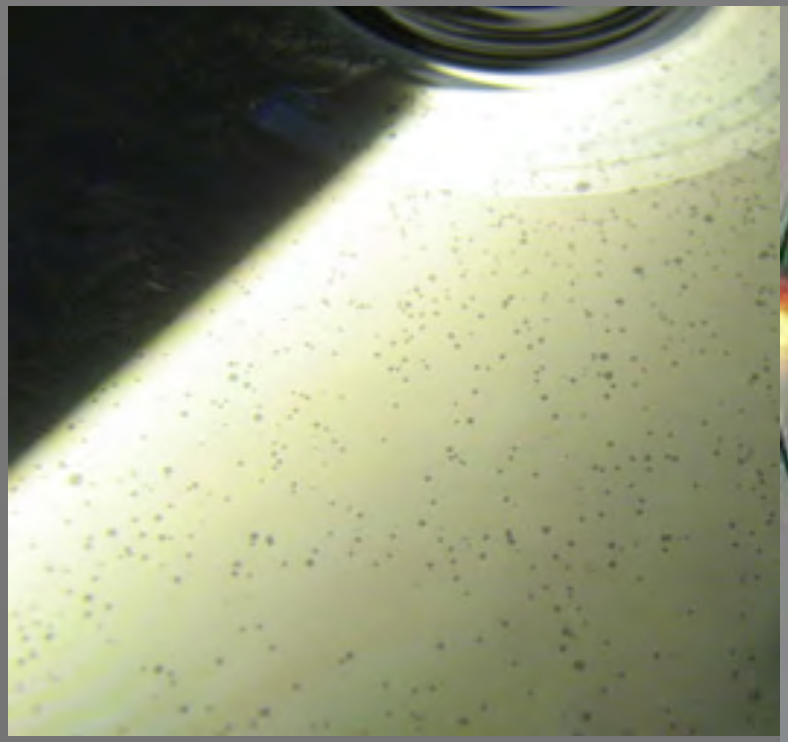
DIGITAL:

Low longevity for some formats



Source: XKCD

PHYSICAL MEDIA LIFESPAN

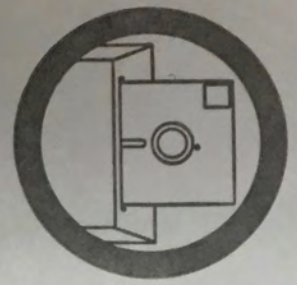


Left to Right:

- Disc rot from failure in manufacturing process
- Scratches from improper handling
- Flaking from improper storage temperature (heat)

For extended media life—
here's how to take care of your flexible disk

Precision surface.
No fingers, please!

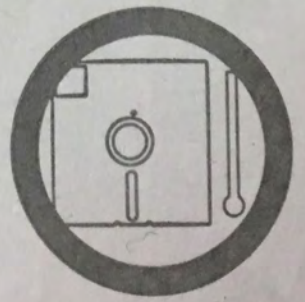


For your disk's sake
(and the system's, too)
insert disk carefully.

Magnetic fields erase.
Keep them far away.

Keep it safe—
in the jacket
when not in use.

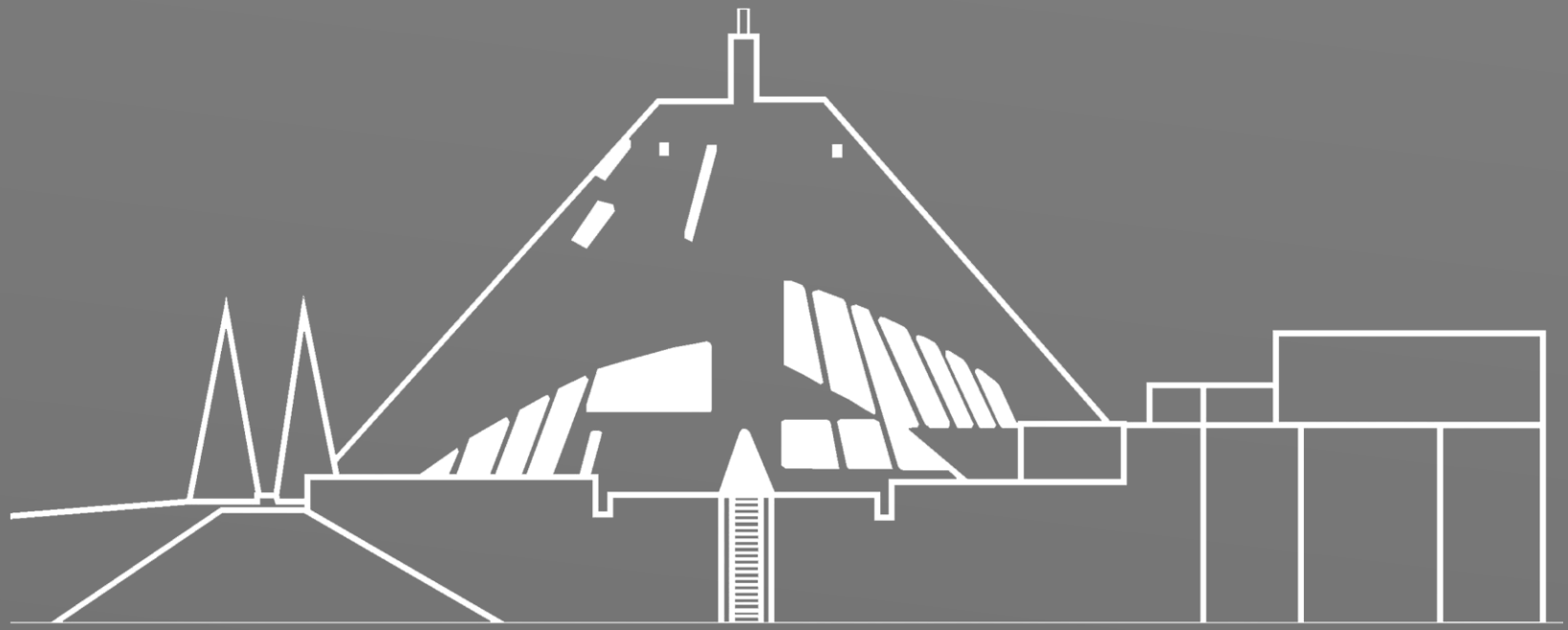
Bending and folding
may damage.
Handle with care.



Keep disks comfortable.
Store at: 10° to 52° C.
50° to 125° F.

DIFFERENCES FROM PHYSICAL RECORDS

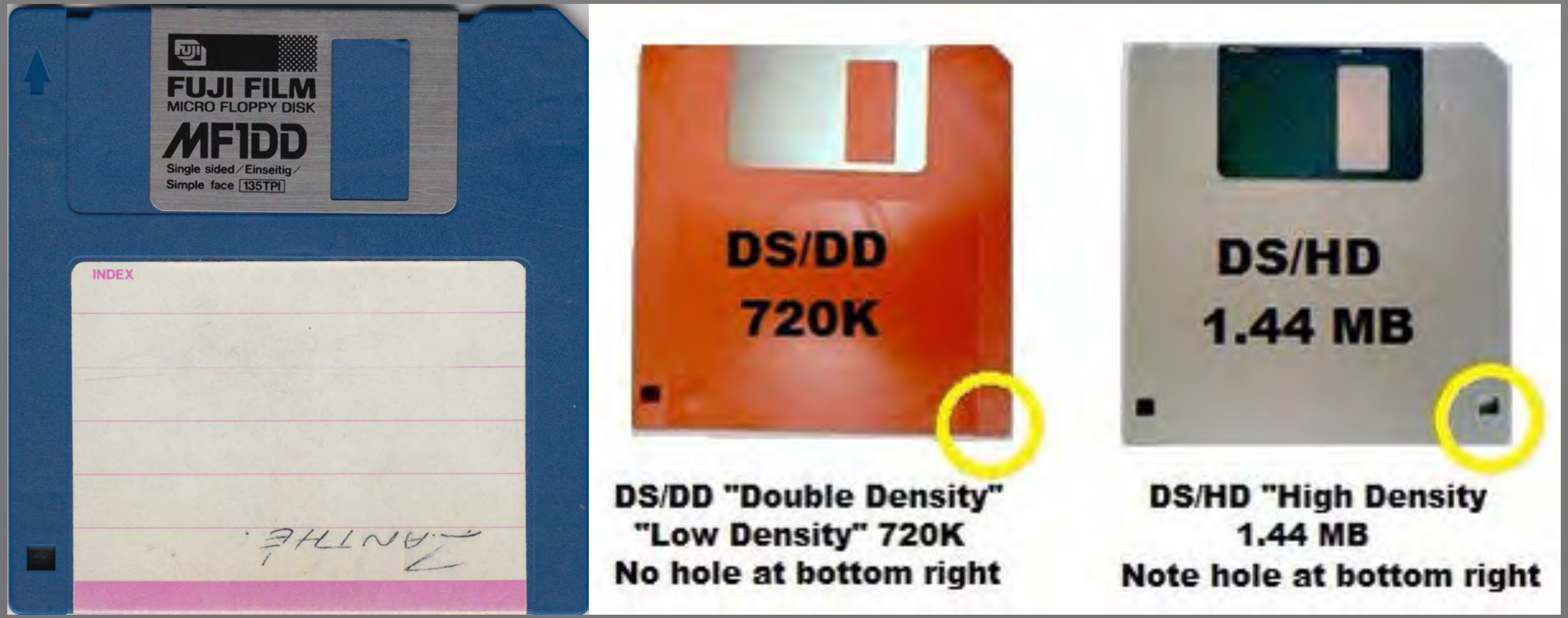
- Requires compatible software
- Requires compatible hardware
- Requires controlled storage environment
- Easily damaged by handling, defects
- Time is of the essence



IDENTIFYING PHYSICAL MEDIA

MAGNETIC MEDIA

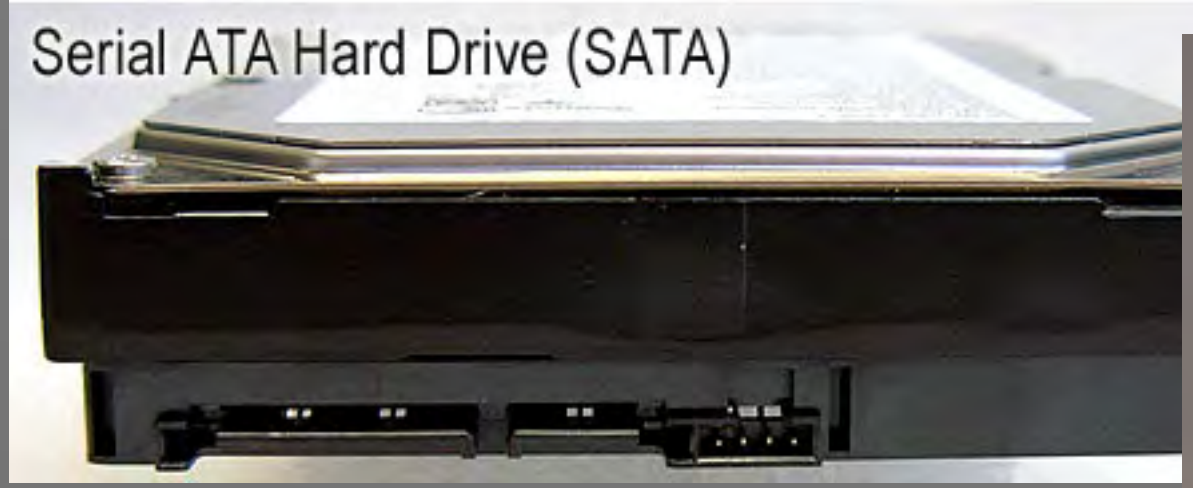
Left to right: 8” disk, 5 ¼” disk, 3 ½” disk



Left to right:

- 400 kilobyte “single density” floppy disk
- 720 kilobyte “double density” floppy disk
- 1.44 megabyte “high density” floppy disk

Left to right: Zip 100 disk, LS120 "SuperDisk"



Top left: 3.5” hard disk drives

Lower right: 2.5” hard disk drives

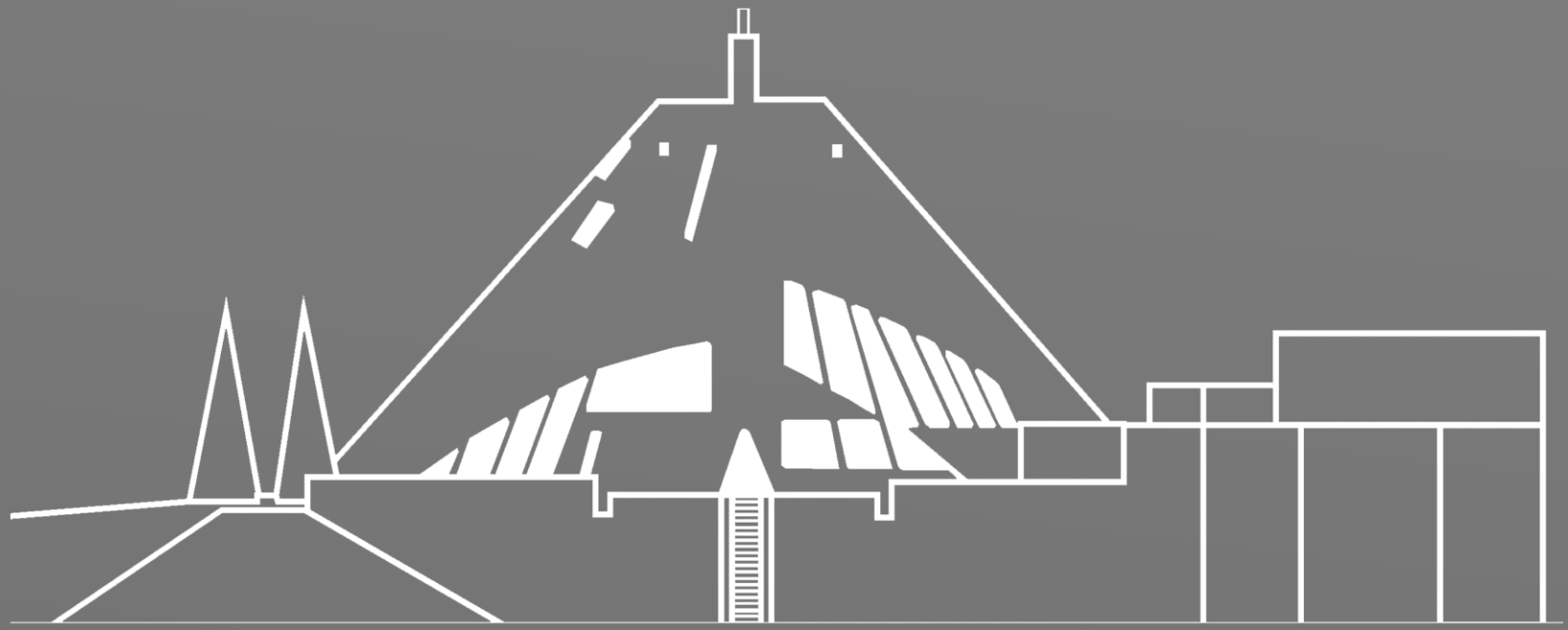
OPTICAL DISCS

Left to right:

- Mini CD, CD-R
- Dual-layer DVD
- Quad-layer Blu-Ray disc

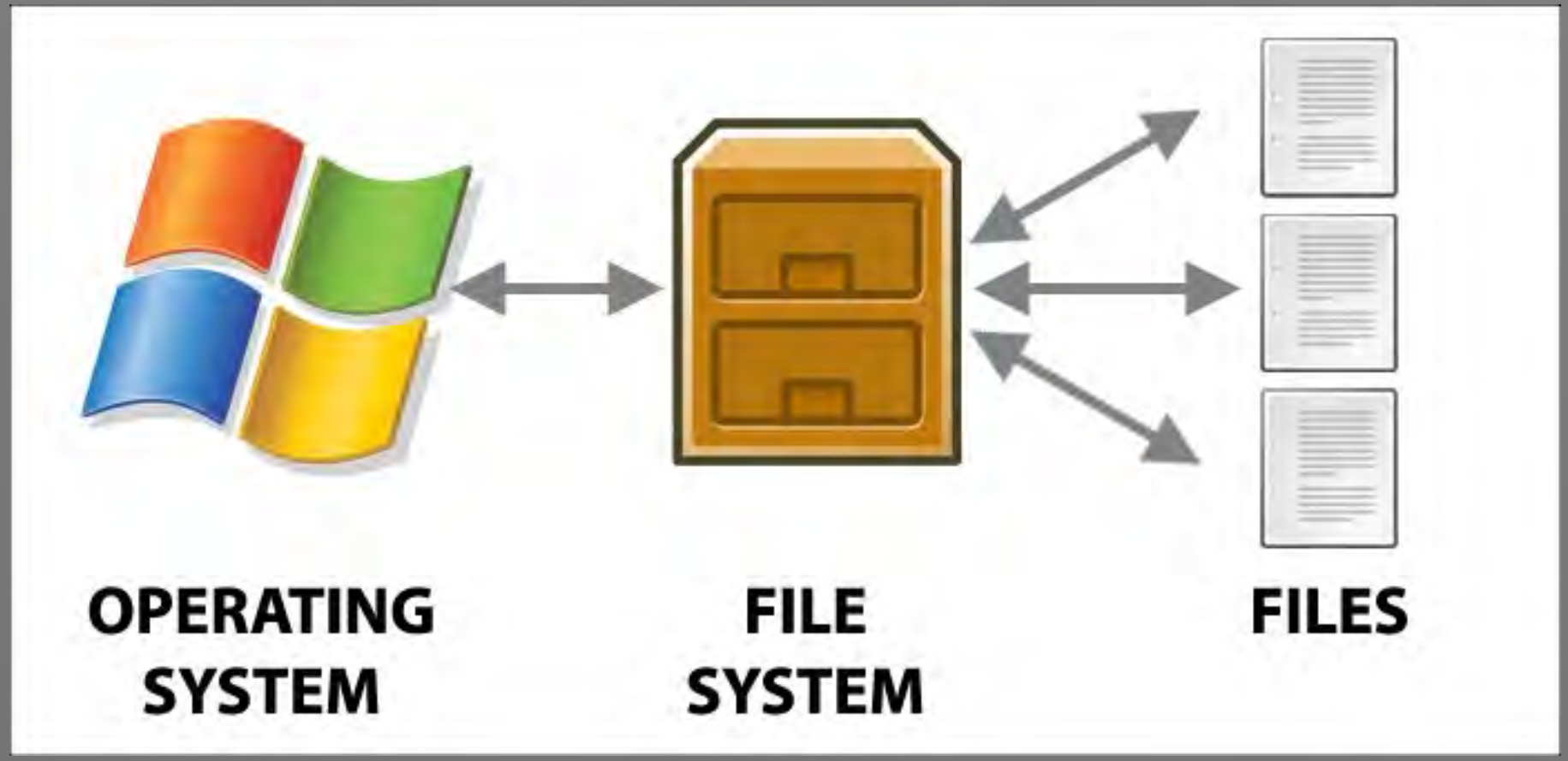
FLASH-BASED MEDIA

Clockwise: USB flash drive, Comact Flash Card, SD Card, Mini SD Card, MicroSD Card, Solid State Drive



FILE SYSTEMS How computers organize information

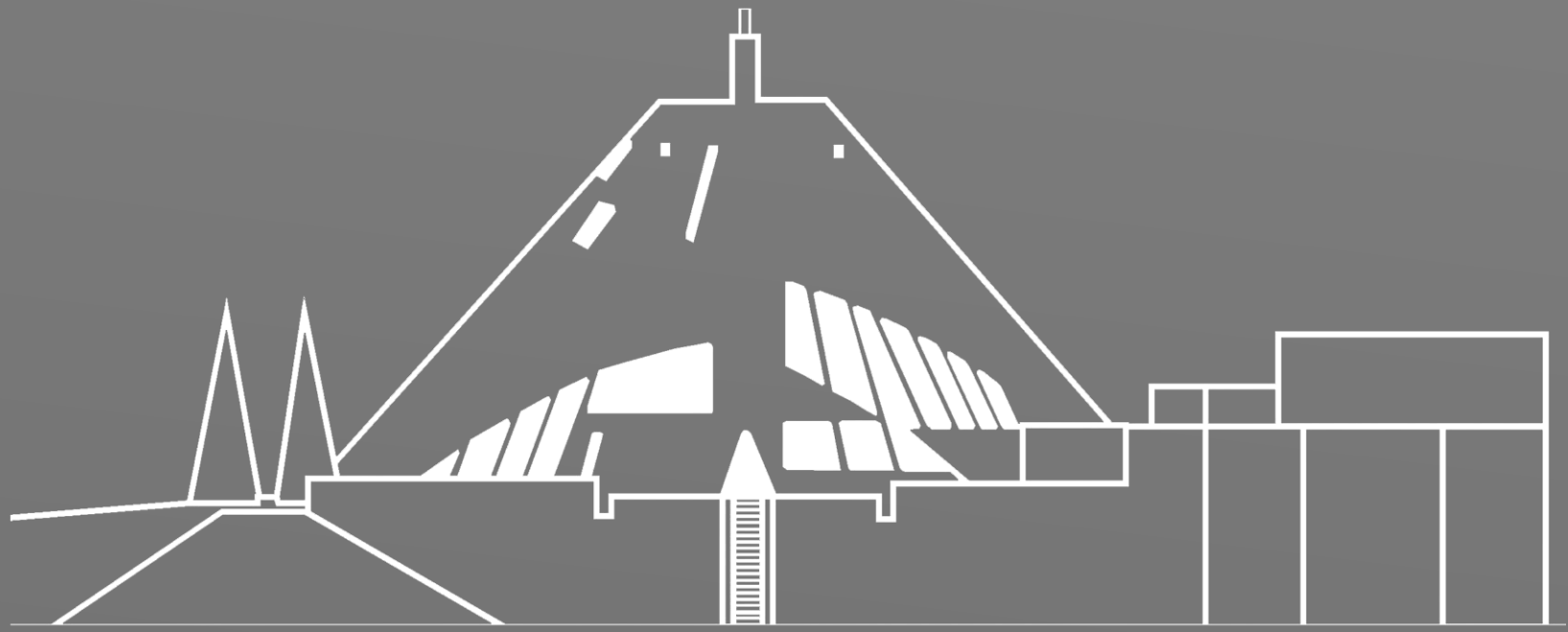
FILE SYSTEMS



The file system organizes and interprets the discrete digital information (files).

Left to right:

- Apple //e computer using DOS 3.3 (disk operating system)
- IBM PC using FAT (file allocation table) file system
- Macintosh 512k using HFS (hierarchal file system)



WORKING WITH PHYSICAL MEDIA

OBTAINING HARDWARE

eSynic

Esynic USB Floppy Drive 3.5" USB External Floppy Disk Drive Portable 1.44 MB FDD USB Drive Plug and Play for PC Windows 10 7 8 Windows XP Vista Mac Black

4.5 out of 5 stars ▾ 350 customer reviews | 21 answered questions

Price: $16.95 Prime

In Stock.

Want it Tuesday, March 7? Order within **22 hrs 56 mins** and choose **One-Day Shipping** at checkout.

Details

Sold by eSynic-us and Fulfilled by Amazon. Gift-wrap available.

Eligible for amazonsmile donation.

Color: **Black**

$17.95 Prime

$16.95 Prime

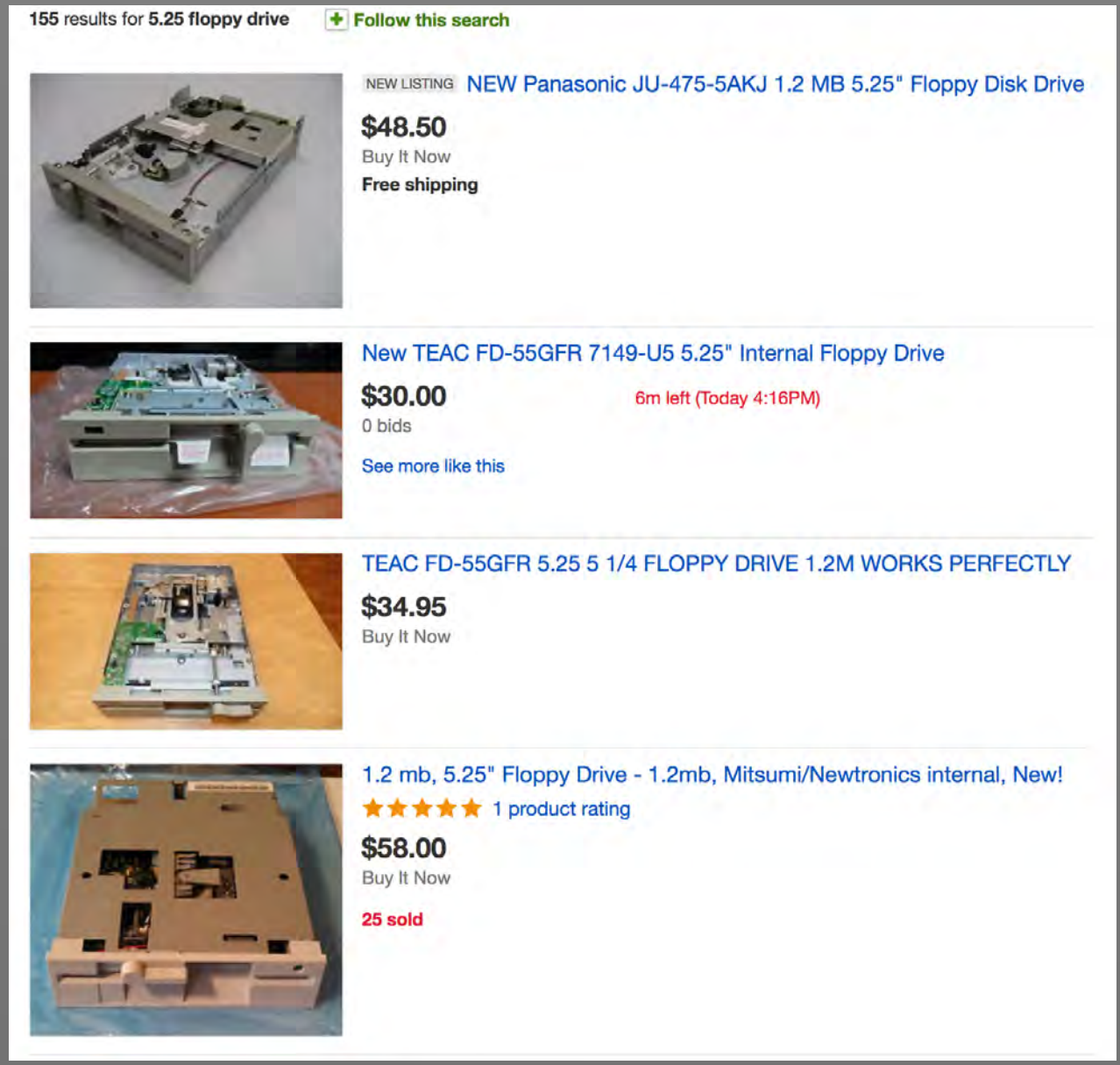
Source: amazon.com

OBTAINING HARDWARE

Source: amazon.com

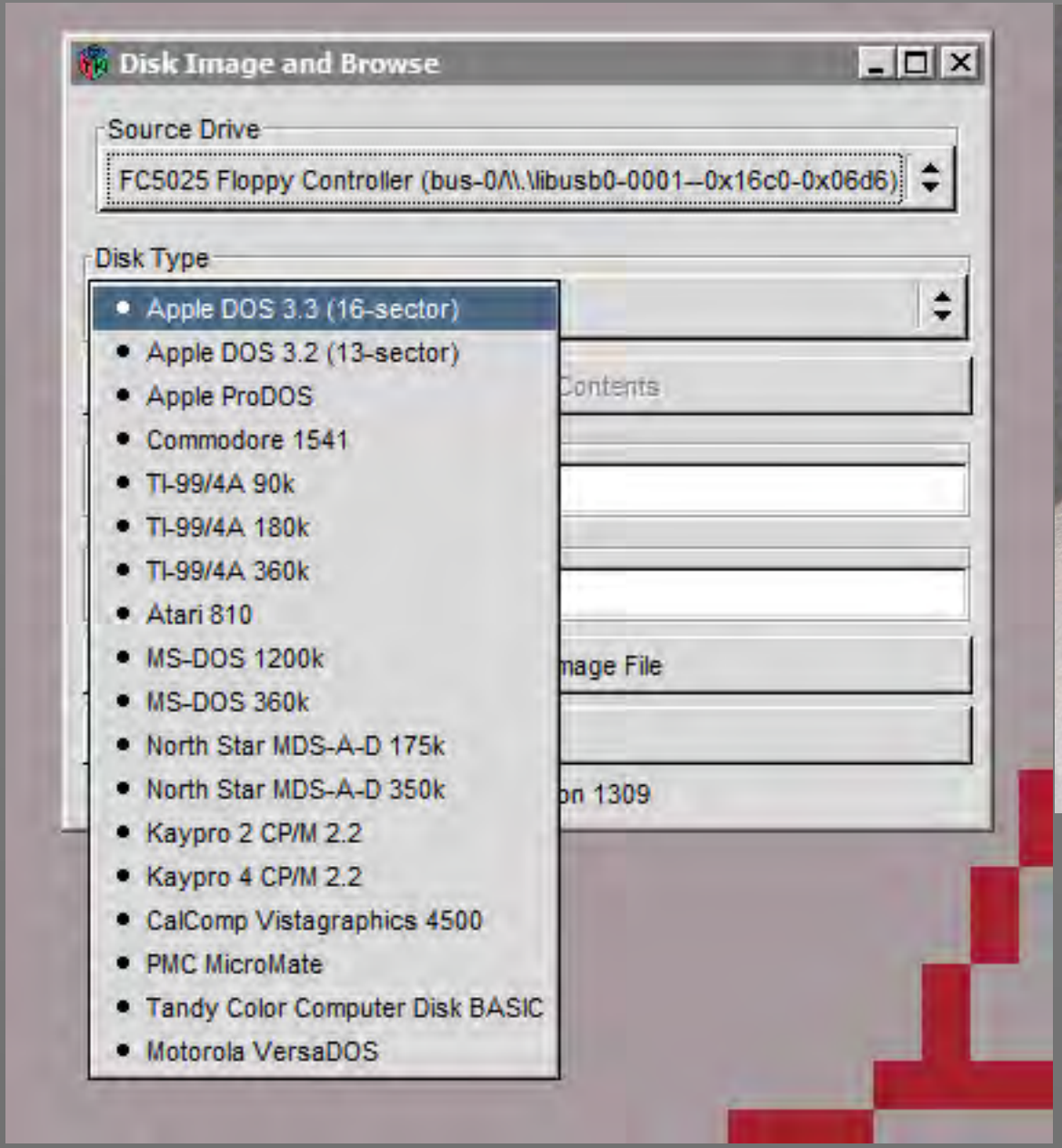
OBTAINING HARDWARE

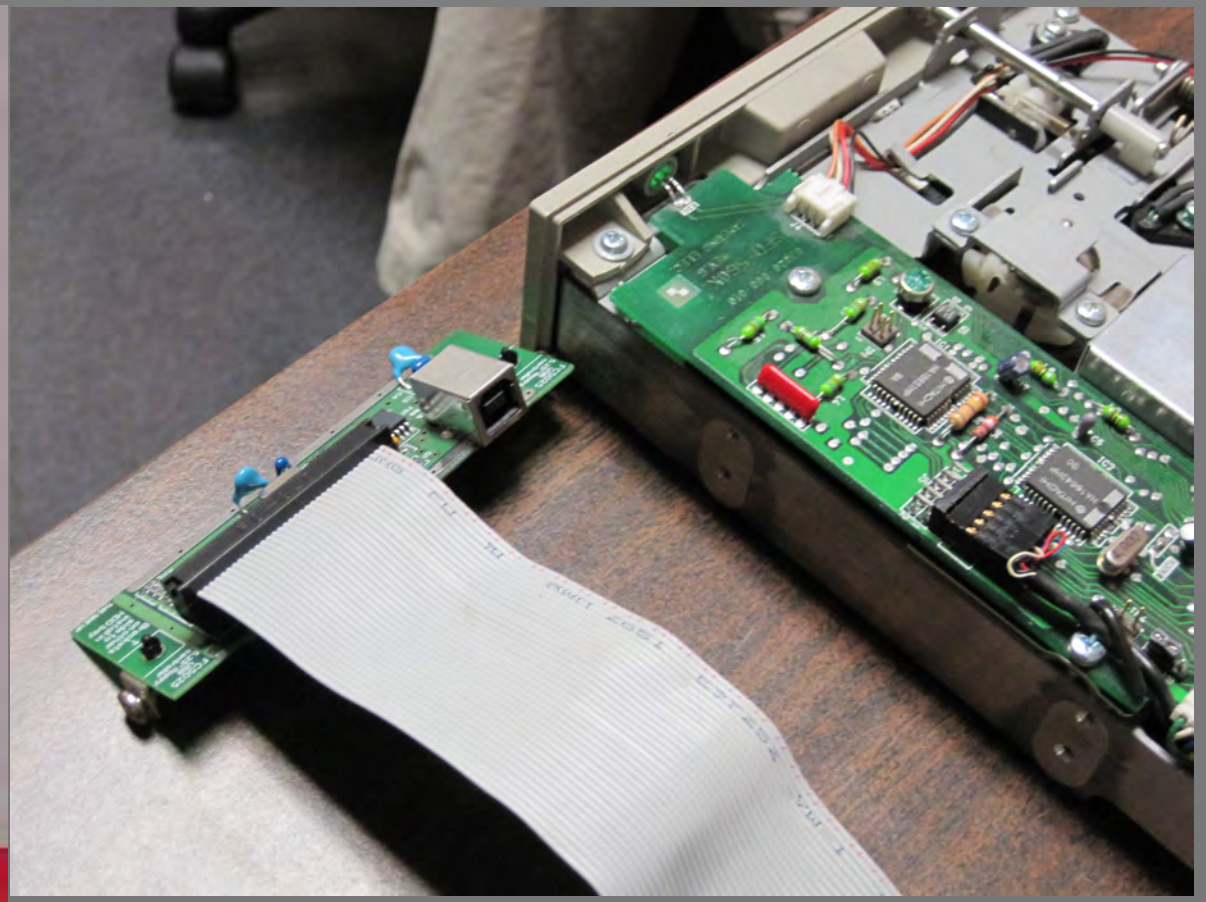
- eBay
- Amazon
- IT Departments
- Thrift Stores
- Always buy more than one!



Source: eBay search for "5.25 floppy drive"

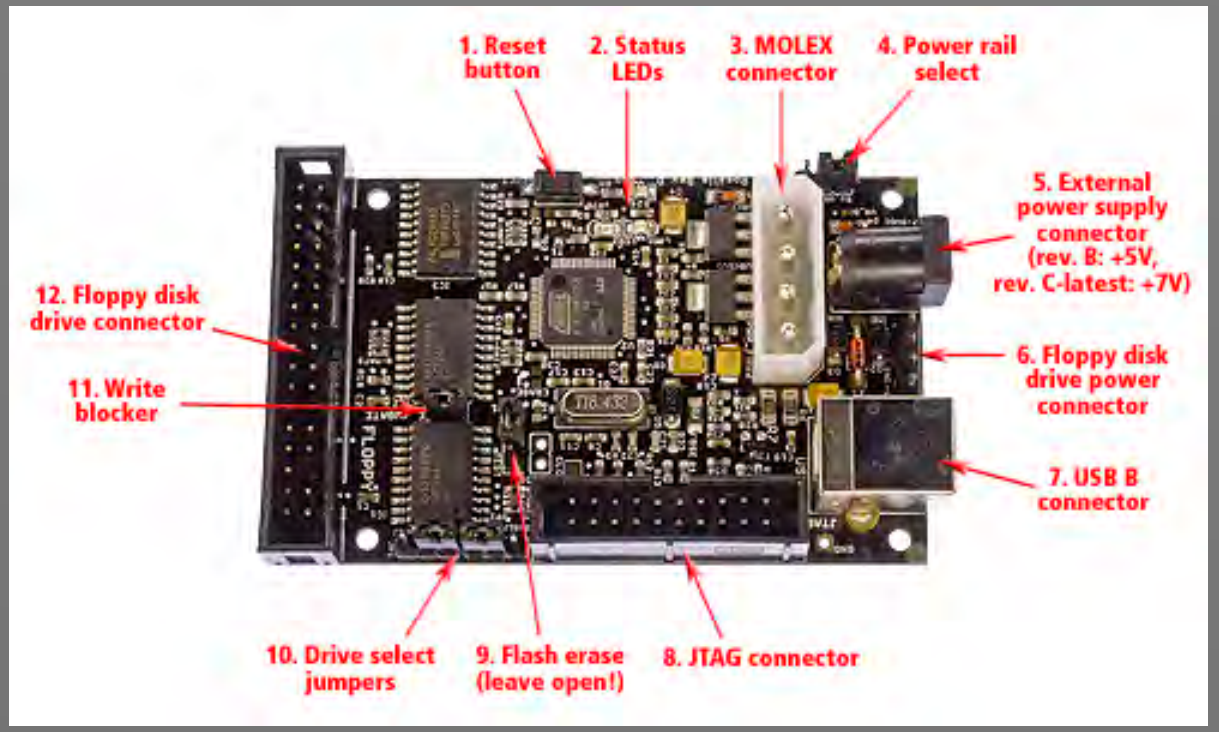
USING 5.25” FLOPPY DRIVES: FC5025

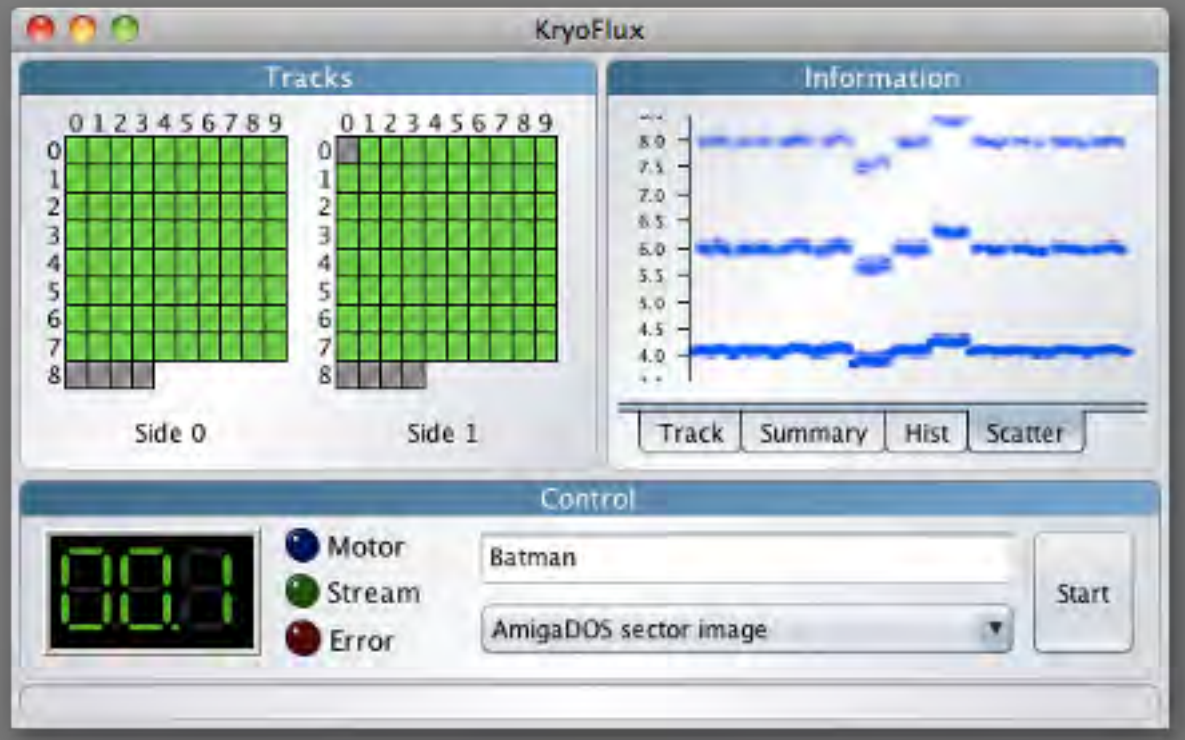




Left: FC5025 Software interface, listing compatible file systems
Right: FC5025 hardware, connected to a 5.25” disk drive

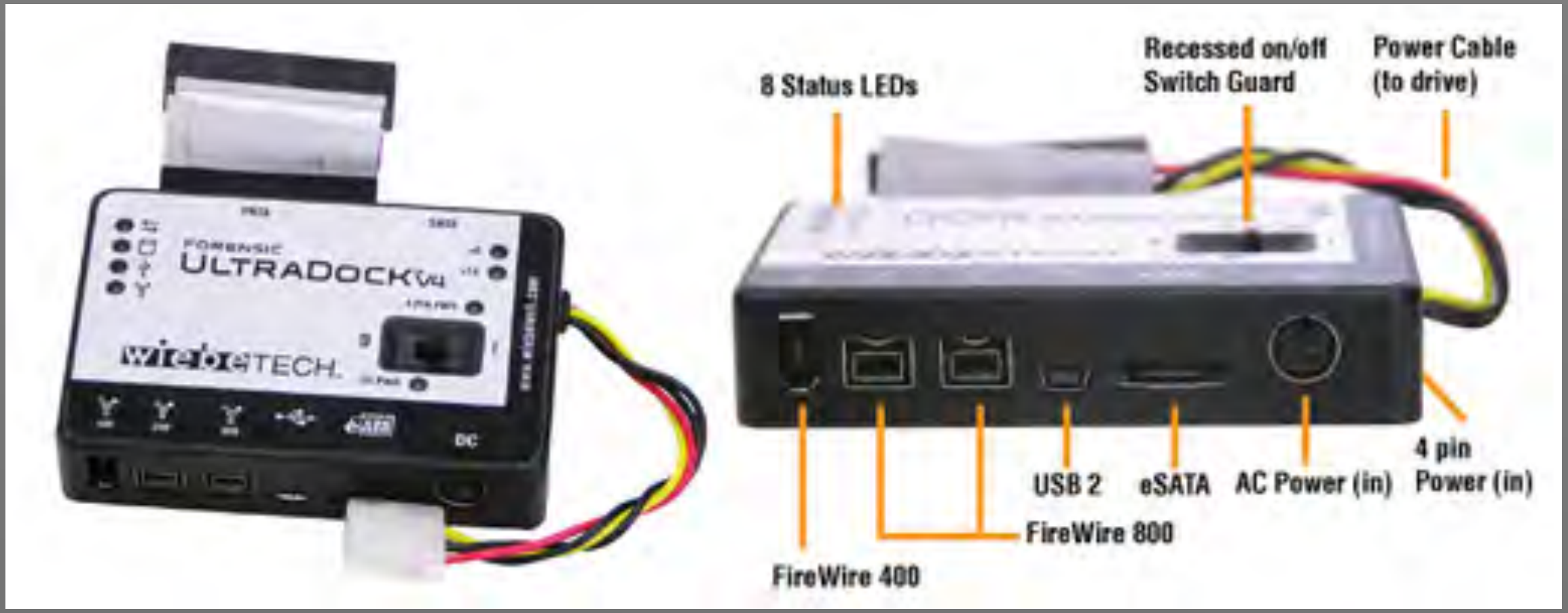
USING 5.25” FLOPPY DRIVES: KRYOFLUX



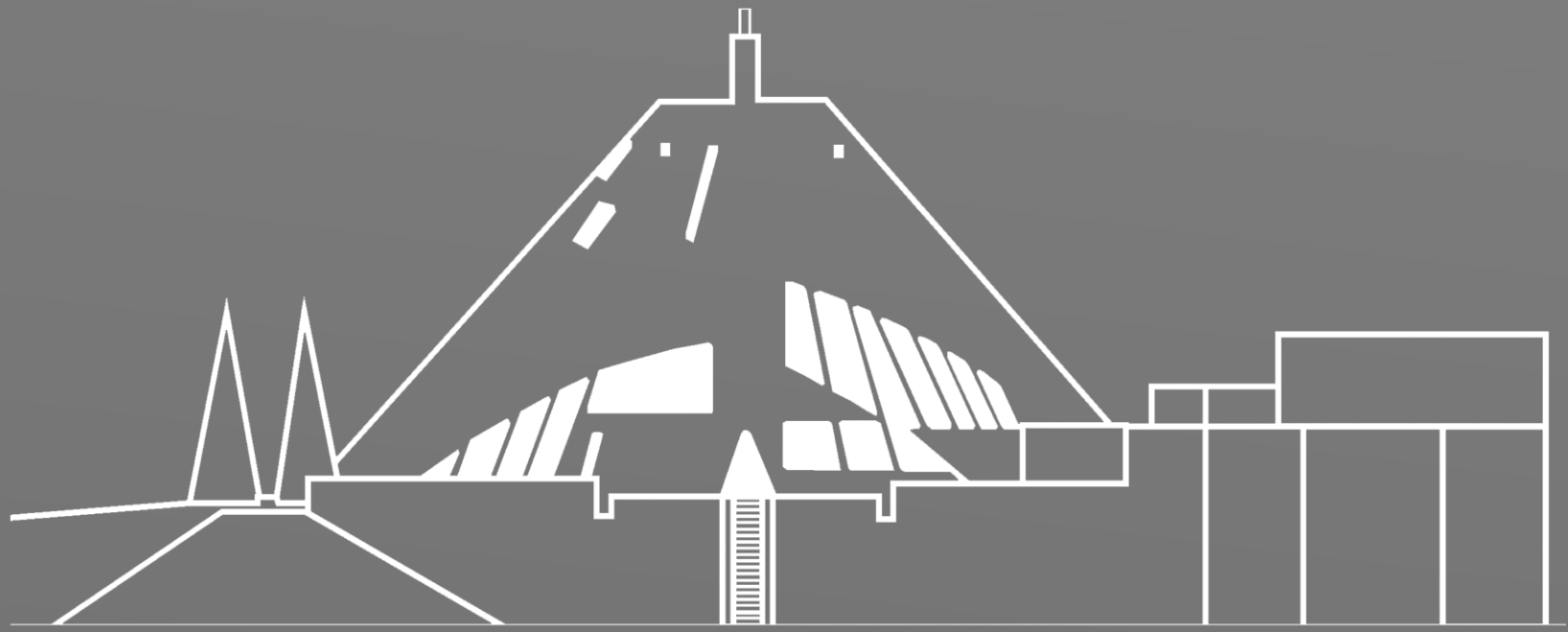


Left: KryoFlux hardware
Right: KryoFlux software running in MacOS X

WRITE-BLOCKERS FOR HARD DISKS AND FLASH DRIVES



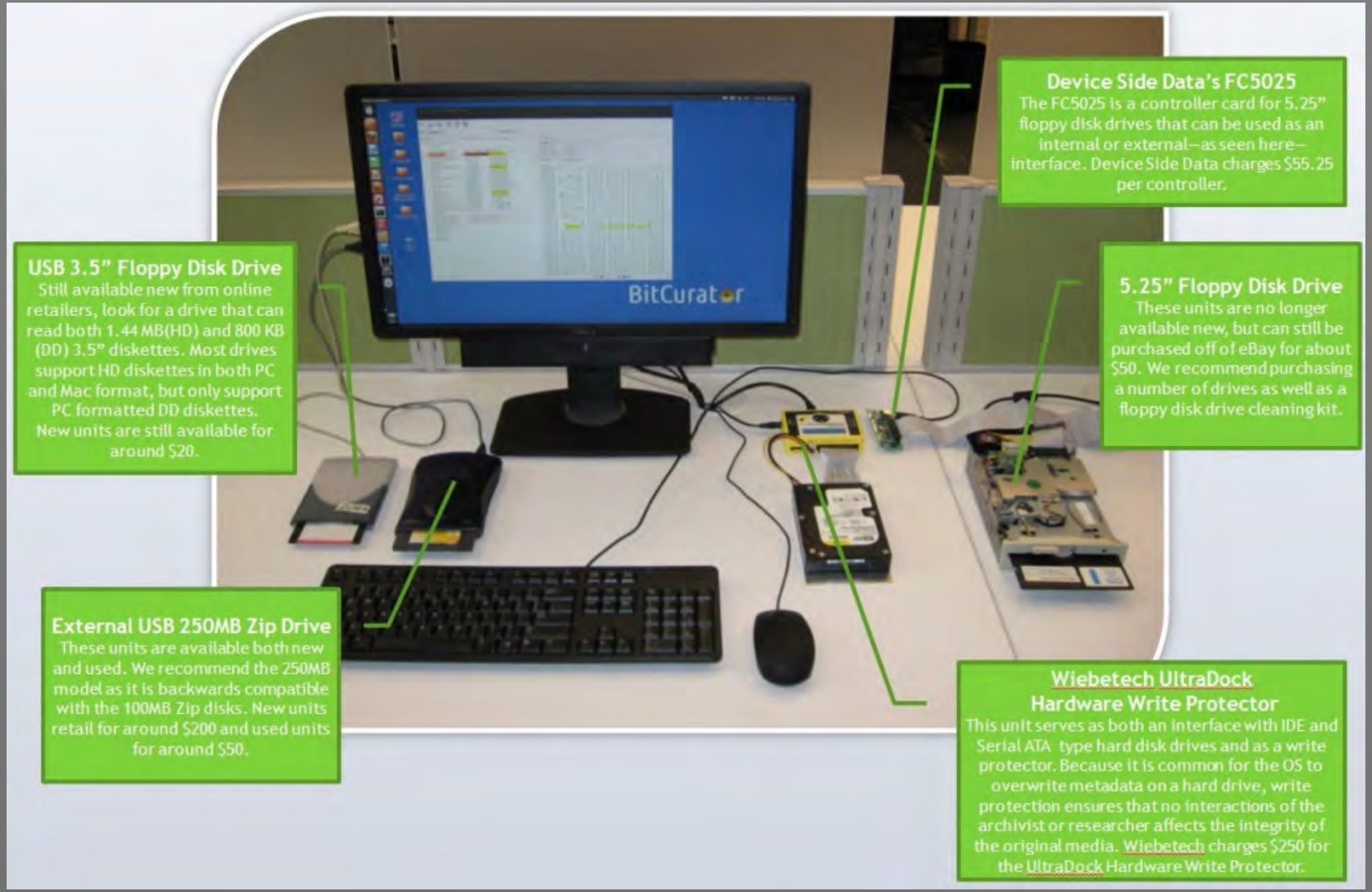
Source: UltraDock



STEPS FOR BASIC PRESERVATION

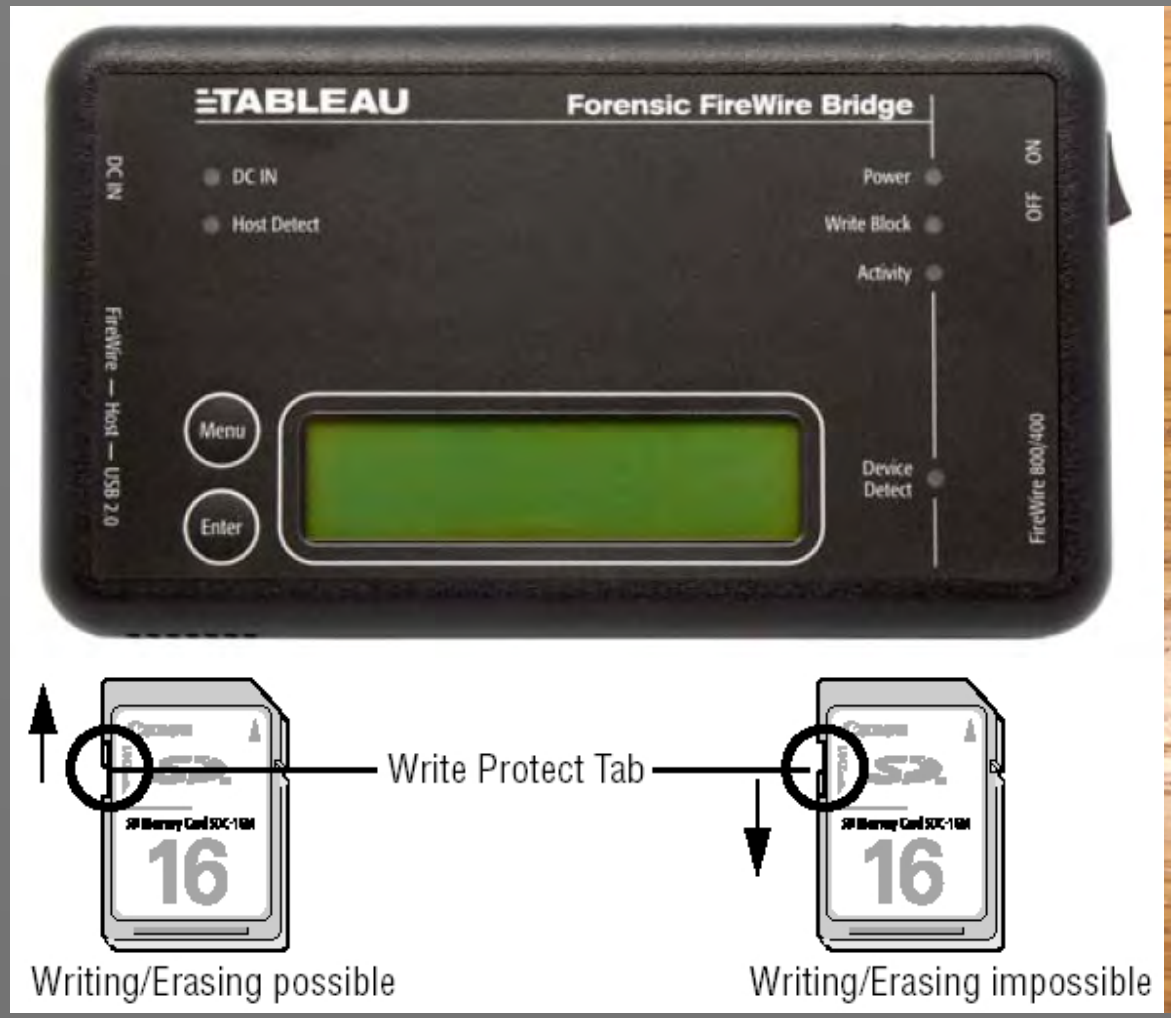
1.	Use a "clean" computer (a dedicated computer that is regularly scanned with up-to-date antivirus software and that is not used for online activities that may introduce viruses or used for other work that might be affected by viruses introduced when accessing media).
2.	Use a write blocker on the clean computer and set any write-protect tabs on the media to prevent changes to the content.
3.	Insert disk in the appropriate drive (or other medium in appropriate reader or attach other storage device to appropriate port). Do not attempt to open any files.
4.	Create a directory on the clean machine for the current project, with a subdirectory for the data files.
5.	Copy data from physical media to the subdirectory. Consider copying the data as a disk image, which is a single file that contains an exact, sector-by-sector bitstream copy of the disk's content and ensures that various forms of essential metadata and technical dependencies will be retained. Alternatively, directly copy directories and files from the original medium to the subdirectory, but note that various forms of associated data and metadata may not be transferred. Moving the data directly from the original medium into a zip-compressed archive can help to preserve some file system metadata (e.g., timestamps, directory structures, and file permissions).
6.	Generate a copy of the disk directory information (file names, sizes, extensions, and dates). Store a digital copy in the project directory and print out a copy to keep with the collection.
7.	Generate and record a checksum (a unique value based on the contents of a file) on the disk image. Alternatively, if you copied the files instead of copying a disk image, generate and record a checksum on each file in the subdirectory.
8.	Create a readme file containing pertinent information from the above steps, indicating the related analog materials and documenting each step taken. Store the file in the project directory and store a printout of the readme file with the physical collection materials.
9.	Copy the project directory to trustworthy archival storage where it will receive regular back up, with a copy stored in another location.
10.	Return the original physical media to storage. (This is optional, but it provides one more possibility for back up if something goes wrong. If you choose not to retain the media, discard them responsibly.)
11.	Create or update an associated finding aid, collection level record, or accession record with information about the steps that were taken and the location of the files.

Source: Erway, Demystifying Born Digital I



Source: BitCurator/Porter Olson

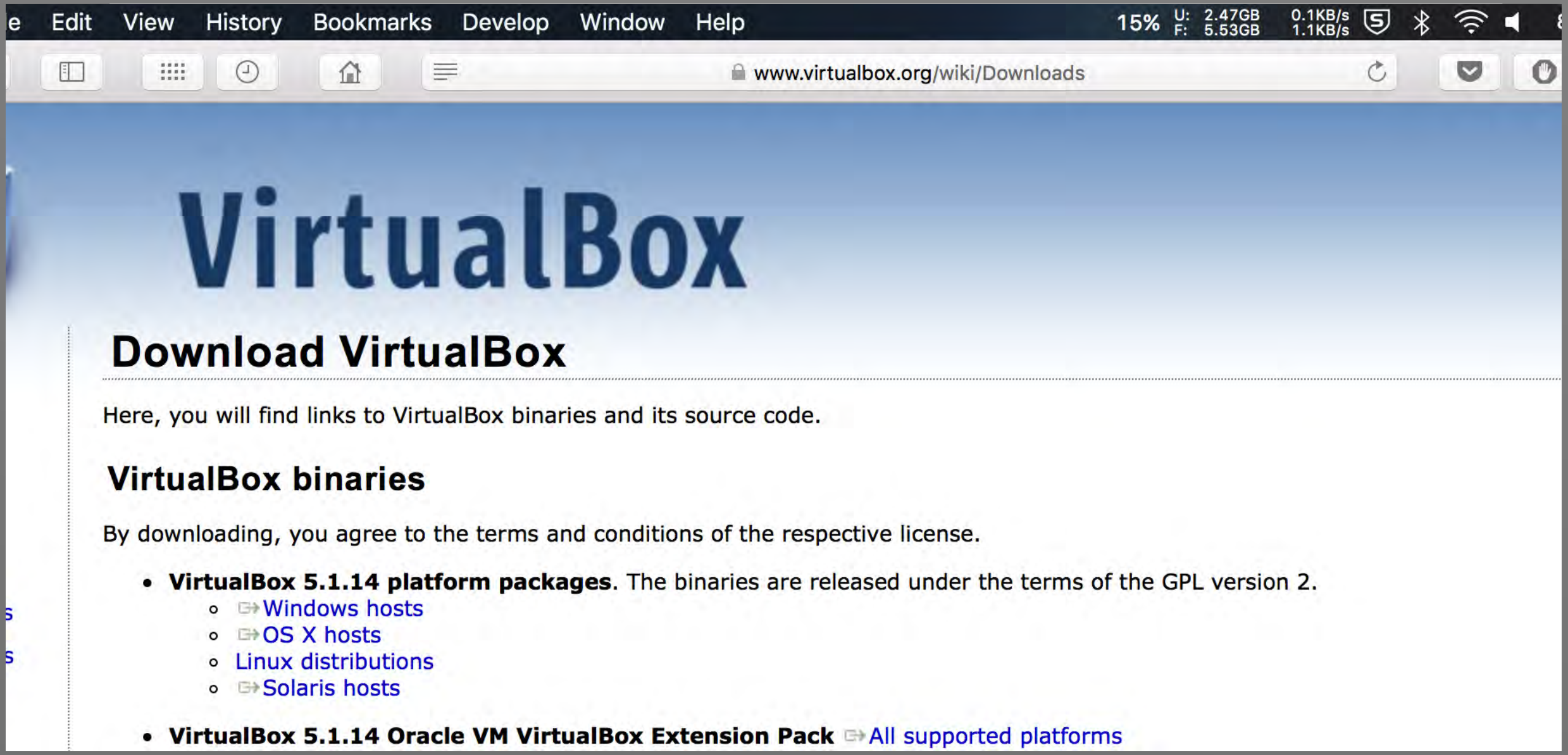
BORROWING FROM DIGITAL FORENSICS: WRITE-BLOCKERS



Clockwise: USB write-blocker, write-protect tab on a 3.5” disk, write-protect tab on an SD card

BITCURATOR

- Suite of tools for migration, metadata creation
- Based on Ubuntu Linux
- Installs easily on most PCs or in VirtualBox
- Free download, frequently updated


e Edit View History Bookmarks Develop Window Help
15% U: 2.47GB F: 5.53GB 0.1KB/s 1.1KB/s
www.virtualbox.org/wiki/Downloads
VirtualBox
Download VirtualBox
Here, you will find links to VirtualBox binaries and its source code.
VirtualBox binaries
By downloading, you agree to the terms and conditions of the respective license.
VirtualBox 5.1.14 platform packages. The binaries are released under the terms of the GPL version 2.
Windows hosts
OS X hosts
Linux distributions
Solaris hosts
VirtualBox 5.1.14 Oracle VM VirtualBox Extension Pack All supported platforms

BitCurator

Welcome to the BitCurator wiki!

This is the documentation portal for the BitCurator family of projects. Looking for the latest release of the BitCurator Environment? Check out the **Quick Links** section below. Visit the **BitCurator blog** for news and information about the Andrew W. Mellon Foundation grants that have supported these projects.

Original content on this wiki is Creative Commons-licensed. The BitCurator wiki currently includes 22 pages. View statistics on users, pages, and edits.

Contributing

The BitCurator wiki is open to the community. To participate, click on **Request account**. You'll be asked for your name, email address, bio (why we require this information). You will receive an email with temporary credentials and instructions once an admin has approved you.

Developer? Help us improve projects listed on our GitHub BitCurator organization profile.

Community support for previously grant-funded BitCurator projects is provided by the BitCurator Consortium. If you or your organization use BitCurator tools, please consider joining the Consortium!

BitCurator Projects

Source code and documentation for BitCurator projects can be found on the BitCurator GitHub IO page. Follow the links below for additional information on the individual projects.

BitCurator NLP
BitCurator Access
BitCurator

BitCurator NLP

The BitCurator NLP project is developing software for collecting institutions to extract, analyze, and produce reports on features of interest in text extracted from born-digital materials contained in collections. The software will use existing natural language processing software libraries to identify and report on those items (such as entities and topics) likely to be relevant to archival appraisal, information organization, and access activities.

Quick Links

BitCurator Environment (Current Release)

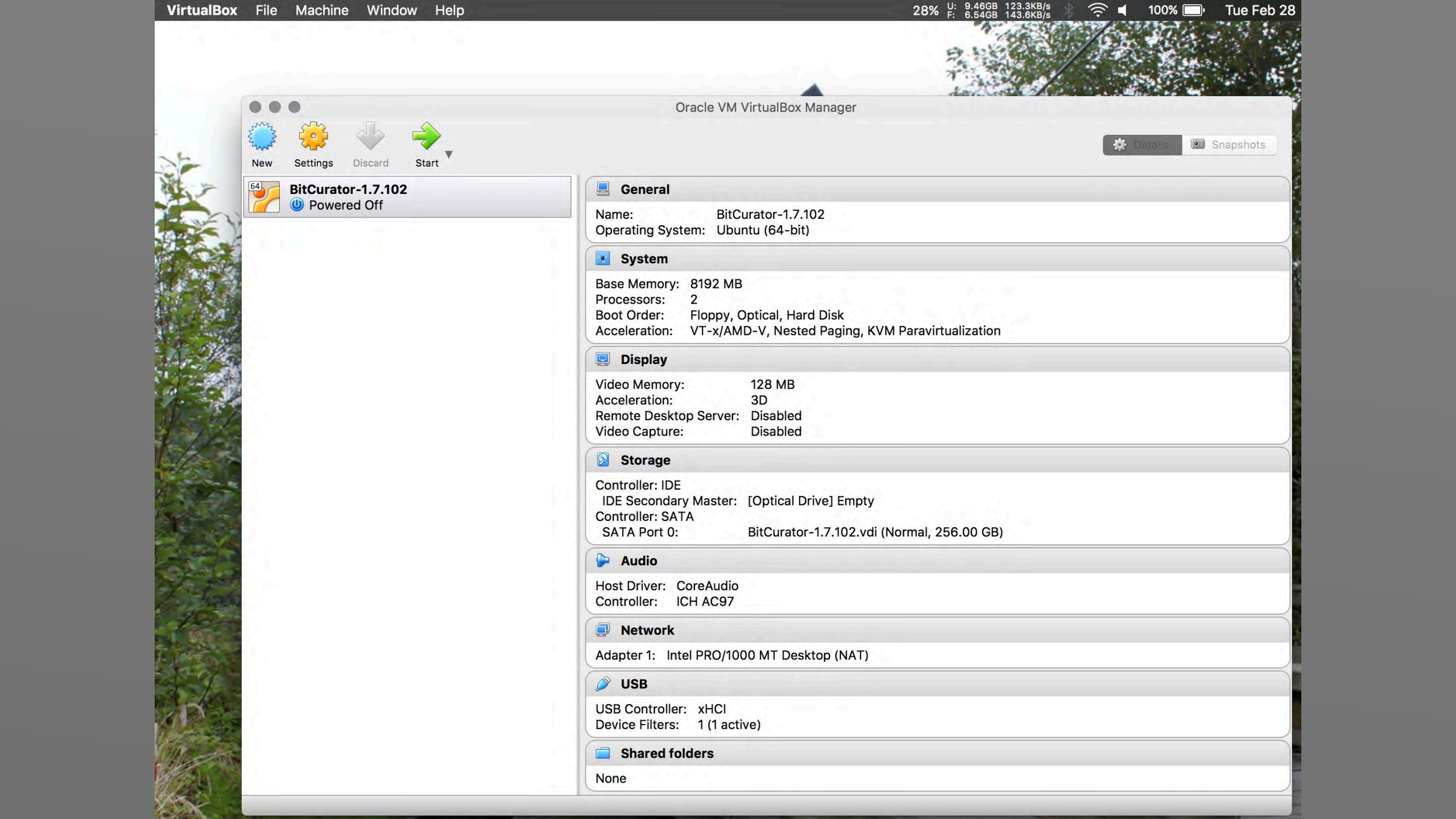
BitCurator Virtual Machine (v1.7.104)
[Torrent] [MD5]

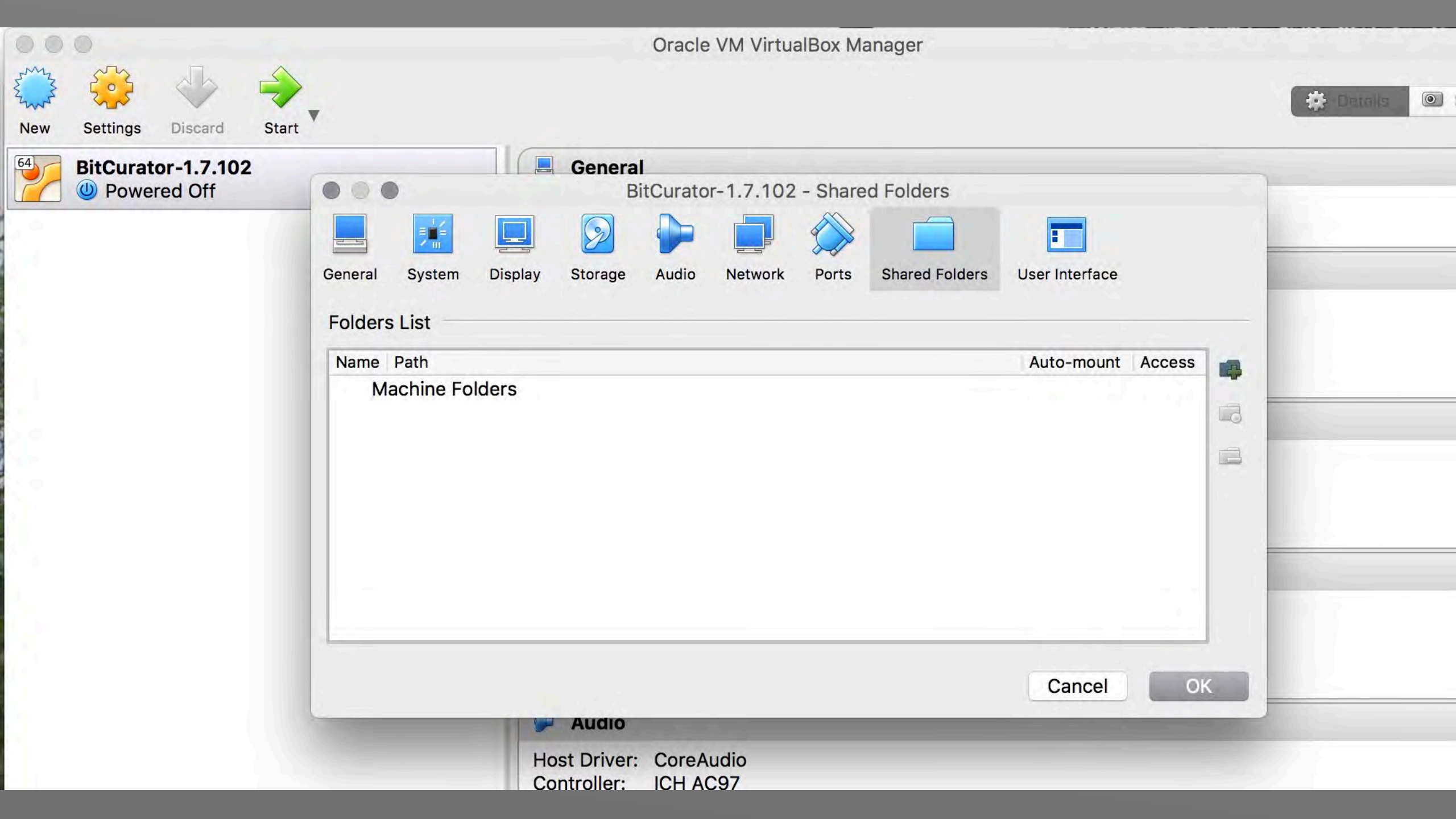
BitCurator Installation ISO (v1.7.104)
[Torrent] [MD5]

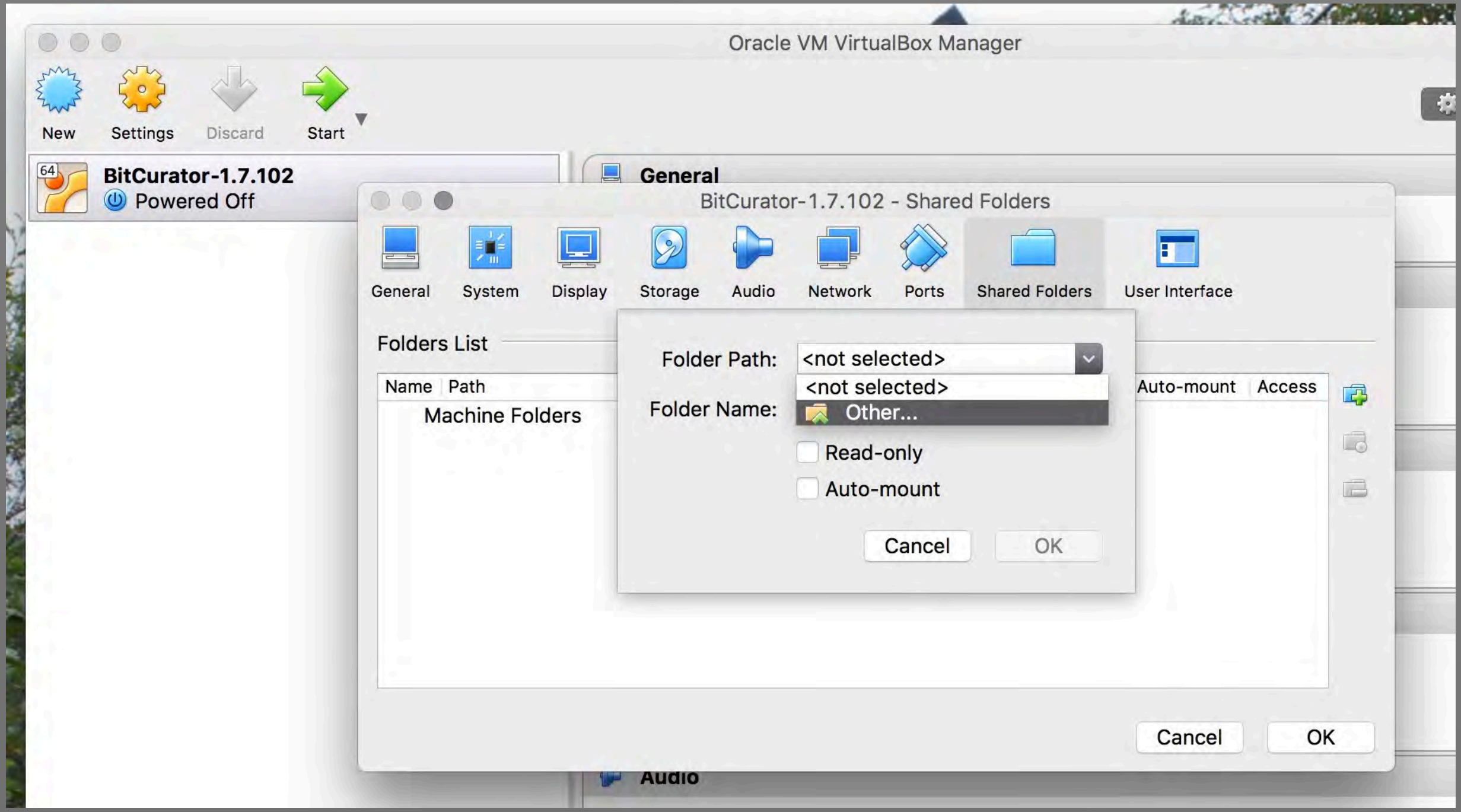
Quickstart Guide Installing and using BitCurator.

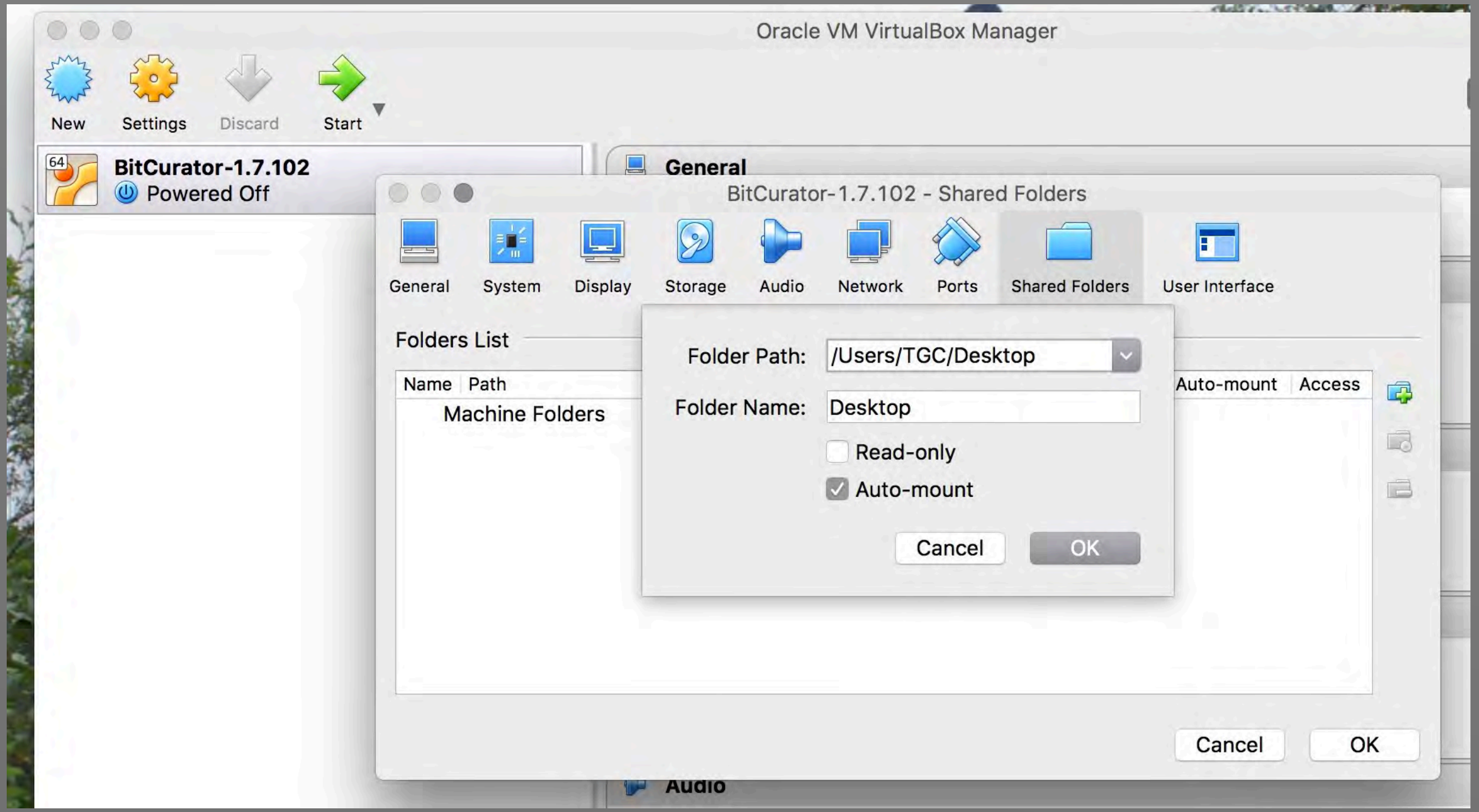
BitCurator Access Tools (Releases on GitHub)

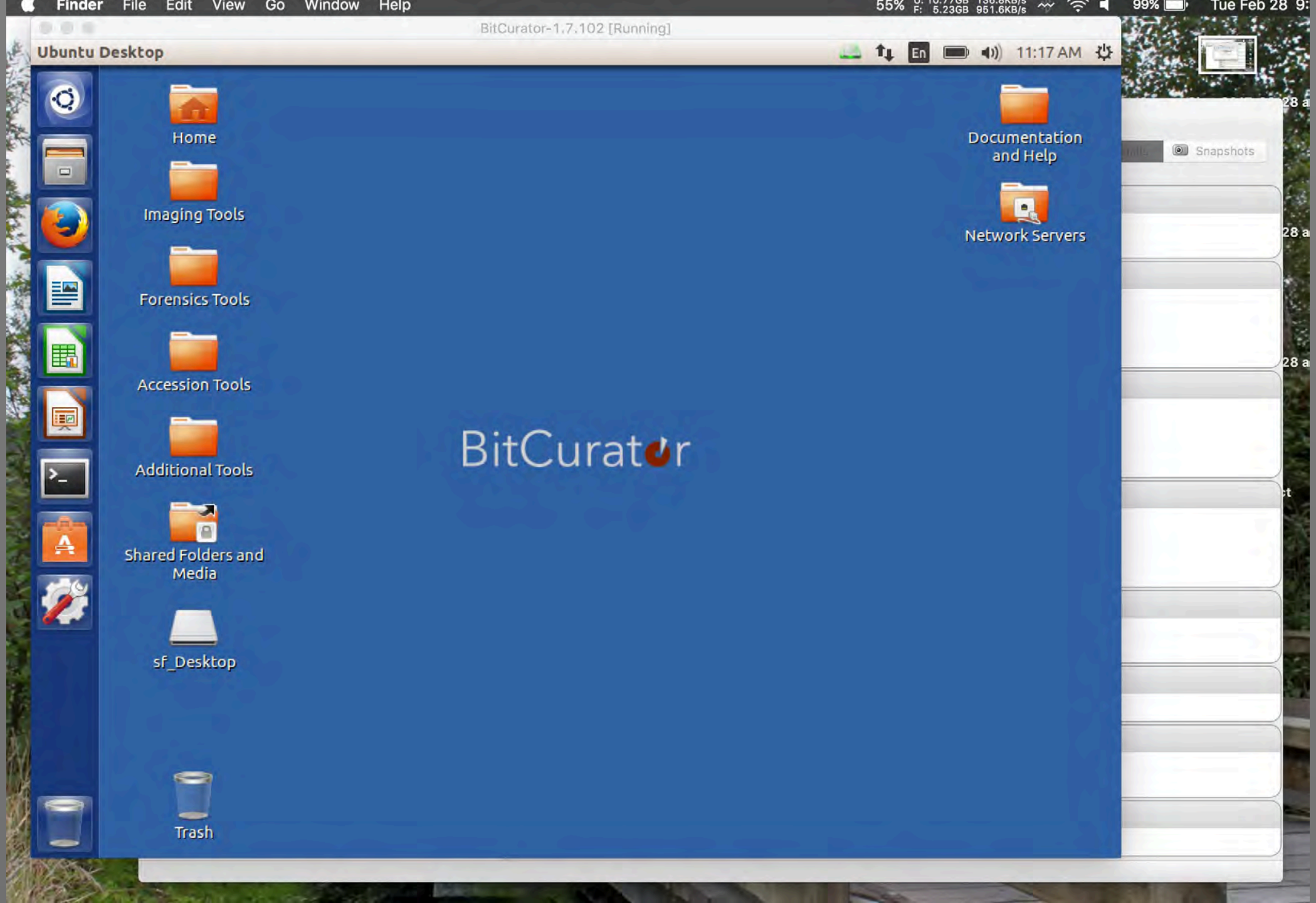
Open "https://wiki.bitcurator.net/downloads/BitCurator-1.7.104.tar.gz" in a new tab


VirtualBox File Machine Window Help
28% U: 9.46GB 123.3KB/s F: 6.54GB 143.6KB/s
100% Tue Feb 28
Oracle VM VirtualBox Manager
New
Settings
Discard
Start
Details
Snapshots
BitCurator-1.7.102
Powered Off
General
Name: BitCurator-1.7.102
Operating System: Ubuntu (64-bit)
System
Base Memory: 8192 MB
Processors: 2
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, KVM Paravirtualization
Display
Video Memory: 128 MB
Acceleration: 3D
Remote Desktop Server: Disabled
Video Capture: Disabled
Storage
Controller: IDE
IDE Secondary Master: [Optical Drive] Empty
Controller: SATA
SATA Port 0: BitCurator-1.7.102.vdi (Normal, 256.00 GB)
Audio
Host Driver: CoreAudio
Controller: ICH AC97
Network
Adapter 1: Intel PRO/1000 MT Desktop (NAT)
USB
USB Controller: xHCI
Device Filters: 1 (1 active)
Shared folders
None


Oracle VM VirtualBox Manager
New
Settings
Discard
Start
Details
BitCurator-1.7.102
Powered Off
General
BitCurator-1.7.102 - Shared Folders
General
System
Display
Storage
Audio
Network
Ports
Shared Folders
User Interface
Folders List
Name
Path
Auto-mount
Access
Machine Folders
Cancel
OK
Audio
Host Driver: CoreAudio
Controller: ICH AC97

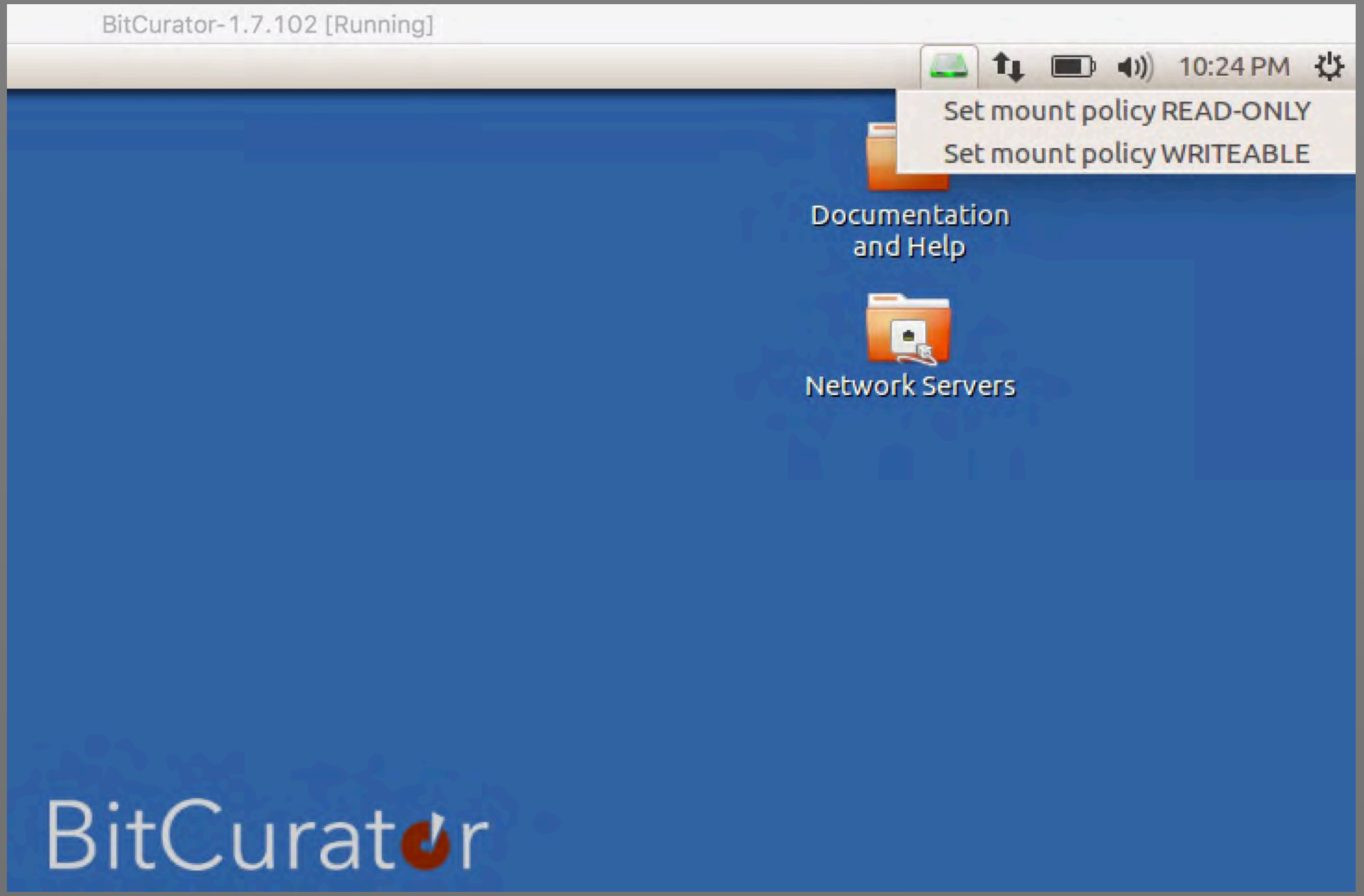

Oracle VM VirtualBox Manager
New
Settings
Discard
Start
BitCurator-1.7.102
Powered Off
General
BitCurator-1.7.102 - Shared Folders
General
System
Display
Storage
Audio
Network
Ports
Shared Folders
User Interface
Folders List
Name
Path
Machine Folders
Auto-mount
Access
Folder Path:
<not selected>
<not selected>
Other...
Folder Name:
Read-only
Auto-mount
Cancel
OK
Cancel
OK
Audio

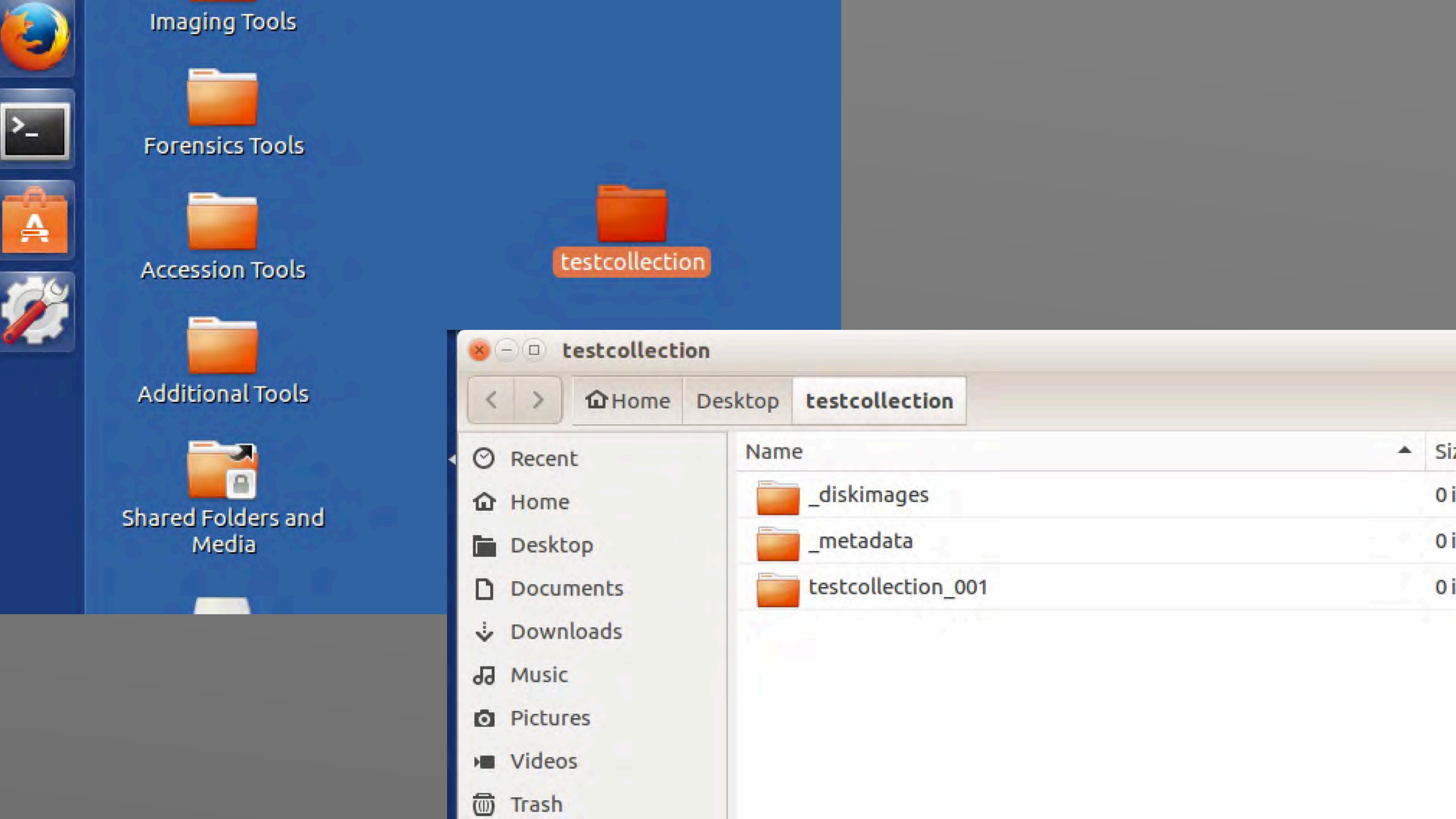

Oracle VM VirtualBox Manager
New
Settings
Discard
Start
BitCurator-1.7.102
Powered Off
General
BitCurator-1.7.102 - Shared Folders
General
System
Display
Storage
Audio
Network
Ports
Shared Folders
User Interface
Folders List
Name
Path
Machine Folders
Auto-mount
Access
Folder Path:
/Users/TGC/Desktop
Folder Name:
Desktop
Read-only
Auto-mount
Cancel
OK
Cancel
OK
Audio

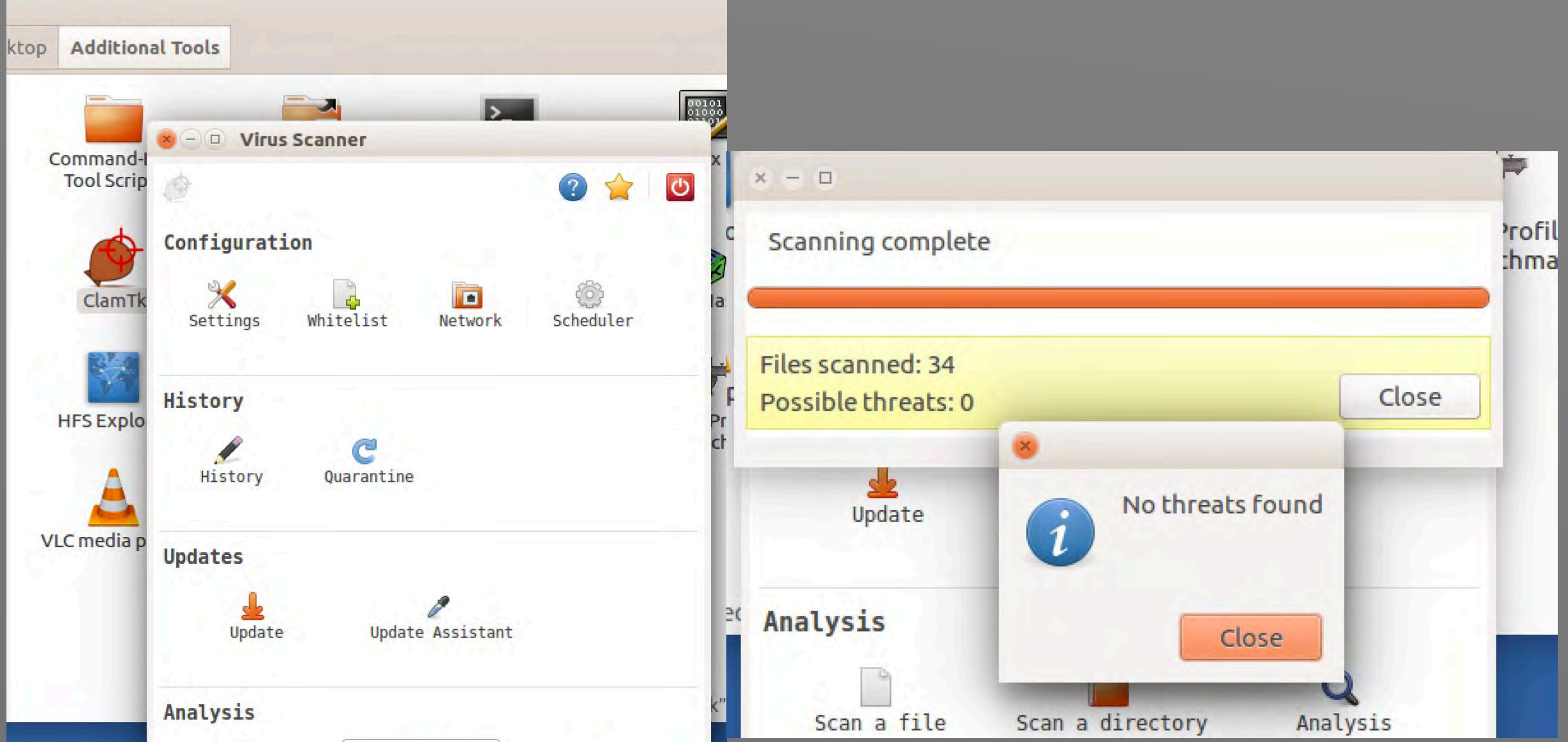

Finder File Edit View Go Window Help
55%
99%
Tue Feb 28
BitCurator-1.7.102 [Running]
Ubuntu Desktop
11:17 AM
Home
Imaging Tools
Forensics Tools
Accession Tools
Additional Tools
Shared Folders and Media
sf_Desktop
Trash
Documentation and Help
Network Servers
BitCurator
Snapshots

1.	Use a "clean" computer (a dedicated computer that is regularly scanned with up-to-date antivirus software and that is not used for online activities that may introduce viruses or used for other work that might be affected by viruses introduced when accessing media).
2.	Use a write blocker on the clean computer and set any write-protect tabs on the media to prevent changes to the content.
3.	Insert disk in the appropriate drive (or other medium in appropriate reader or attach other storage device to appropriate port). Do not attempt to open any files.
4.	Create a directory on the clean machine for the current project, with a subdirectory for the data files.
5.	Copy data from physical media to the subdirectory. Consider copying the data as a disk image, which is a single file that contains an exact, sector-by-sector bitstream copy of the disk's content and ensures that various forms of essential metadata and technical dependencies will be retained. Alternatively, directly copy directories and files from the original medium to the subdirectory, but note that various forms of associated data and metadata may not be transferred. Moving the data directly from the original medium into a zip-compressed archive can help to preserve some file system metadata (e.g., timestamps, directory structures, and file permissions).
6.	Generate a copy of the disk directory information (file names, sizes, extensions, and dates). Store a digital copy in the project directory and print out a copy to keep with the collection.
7.	Generate and record a checksum (a unique value based on the contents of a file) on the disk image. Alternatively, if you copied the files instead of copying a disk image, generate and record a checksum on each file in the subdirectory.
8.	Create a readme file containing pertinent information from the above steps, indicating the related analog materials and documenting each step taken. Store the file in the project directory and store a printout of the readme file with the physical collection materials.
9.	Copy the project directory to trustworthy archival storage where it will receive regular back up, with a copy stored in another location.
10.	Return the original physical media to storage. (This is optional, but it provides one more possibility for back up if something goes wrong. If you choose not to retain the media, discard them responsibly.)
11.	Create or update an associated finding aid, collection level record, or accession record with information about the steps that were taken and the location of the files.

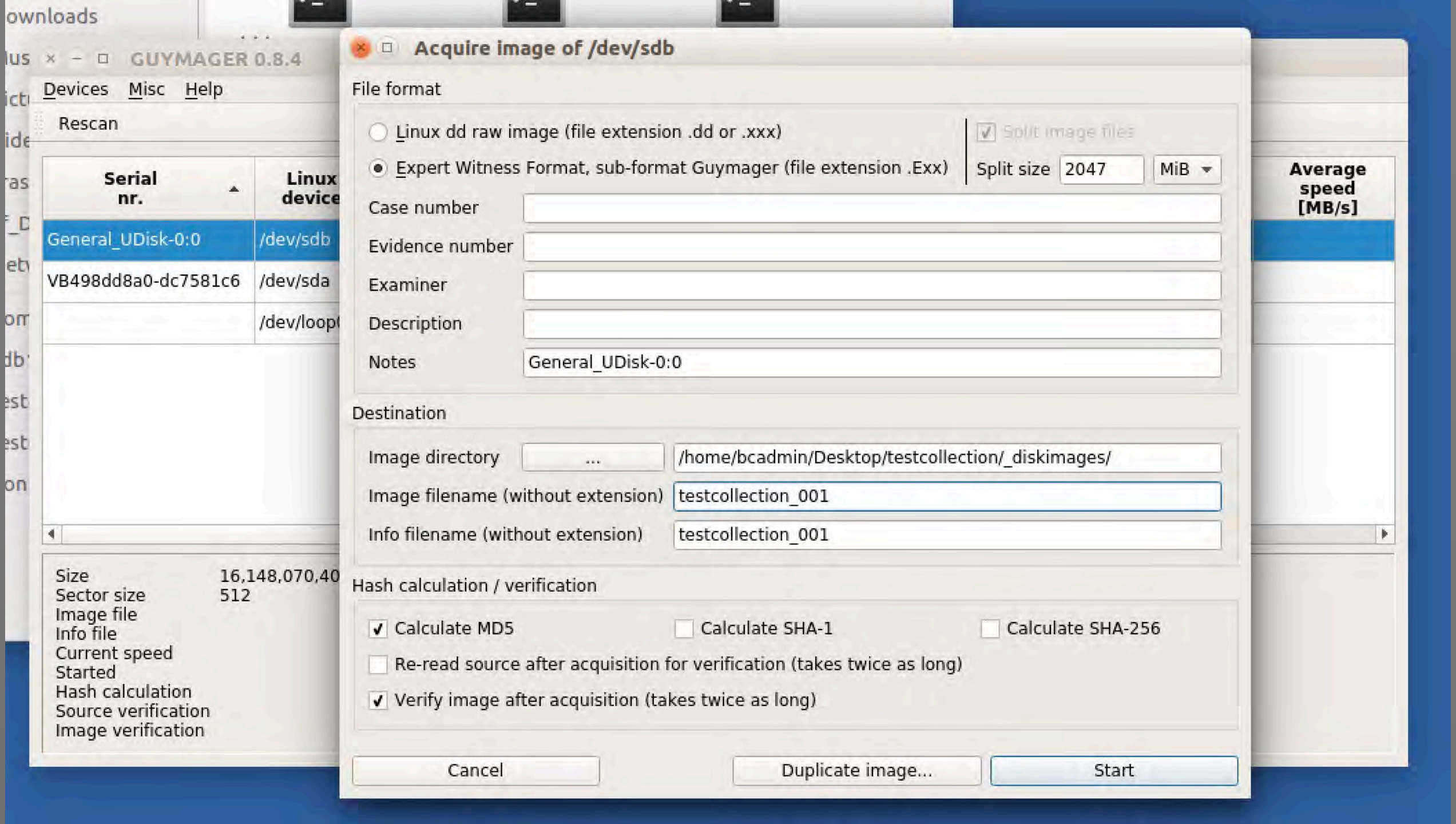
Source: Erway, Demystifying Born Digital I


BitCurator-1.7.102 [Running]
10:24 PM
Set mount policy READ-ONLY
Set mount policy WRITEABLE
Documentation and Help
Network Servers
BitCurator


Imaging Tools
Forensics Tools
Accession Tools
Additional Tools
Shared Folders and Media
testcollection
testcollection
Home
Desktop
testcollection
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Name
_diskimages
_metadata
testcollection_001


ktop
Additional Tools
Command-
Tool Scrip
ClamTk
HFS Explo
VLC media p
Virus Scanner
Configuration
Settings
Whitelist
Network
Scheduler
History
History
Quarantine
Updates
Update
Update Assistant
Analysis
Scan a file
Scan a directory
Analysis
An update is available
Scanning complete
Files scanned: 34
Possible threats: 0
Close
Update
No threats found
Close
Analysis
Scan a file
Scan a directory
Analysis

- Copy files from discs to their subfolders on the desktop
- Create a forensic disk image using Guymager
- Use the .e01 “Expert Witness” disk image format


GUYMAGER 0.8.4
Devices
Misc
Help
Rescan
Serial nr.
Linux device
Average speed [MB/s]
General_UDisk-0:0
/dev/sdb
VB498dd8a0-dc7581c6
/dev/sda
/dev/loop
Size
16,148,070,40
Sector size
512
Image file
Info file
Current speed
Started
Hash calculation
Source verification
Image verification
Acquire image of /dev/sdb
File format
Linux dd raw image (file extension .dd or .xxx)
Expert Witness Format, sub-format Guymager (file extension .Exx)
Split image files
Split size
2047
MiB
Case number
Evidence number
Examiner
Description
Notes
General_UDisk-0:0
Destination
Image directory
...
/home/bcadmin/Desktop/testcollection/_diskimages/
Image filename (without extension)
testcollection_001
Info filename (without extension)
testcollection_001
Hash calculation / verification
Calculate MD5
Calculate SHA-1
Calculate SHA-256
Re-read source after acquisition for verification (takes twice as long)
Verify image after acquisition (takes twice as long)
Cancel
Duplicate image...
Start

GUYMAGER 0.8.4

Devices Misc Help

Rescan

Serial nr.	Linux device	Model	State	Size	Hidden areas	Bad sectors	Progress	Av s [
General_UDisk-0:0	/dev/sdb	General UDisk	Finished - Verified & ok	16.1GB	unknown	0	100%	
VB498dd8a0-dc7581c6	/dev/sda	VBOX_HARDDISK	Idle	274.9GB	unknown			
	/dev/loop0	sdb2	Idle	15.8GB	unknown			

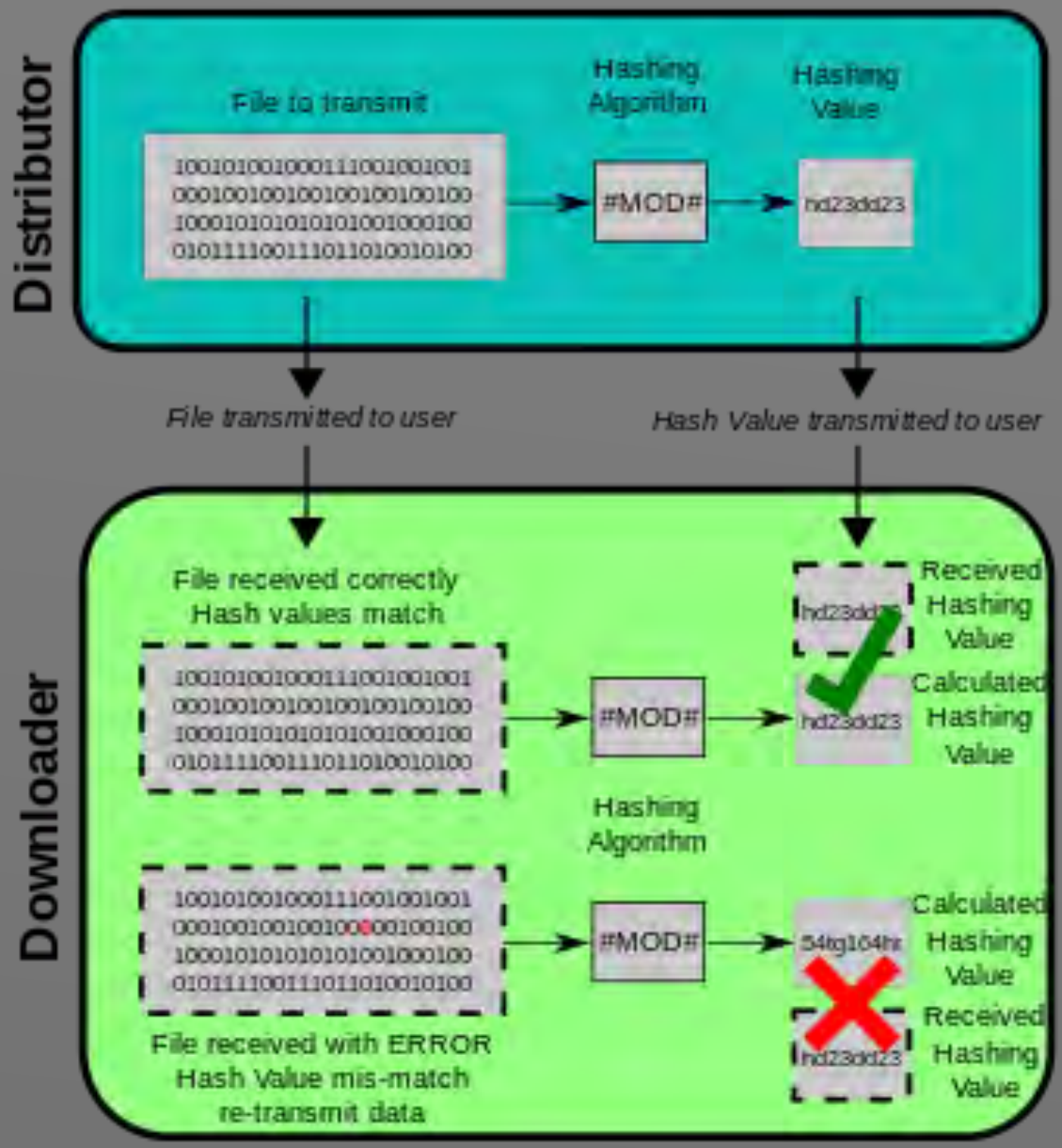
```
Image path and file name: /home/bcadmin/Desktop/testcollection/_diskimages/testcollection_001.Exx
Info  path and file name: /home/bcadmin/Desktop/testcollection/_diskimages/testcollection_001.info
Hash calculation        : MD5
Source verification     : off
Image verification      : on

No bad sectors encountered during acquisition.
State: Finished successfully

MD5 hash                : 859f70fe8ae18504481331c0cc93aa1b
MD5 hash verified source  : --
MD5 hash verified image   : 859f70fe8ae18504481331c0cc93aa1b
SHA1 hash               : --
SHA1 hash verified source : --
```

CHECKSUMS (FILE HASHES)

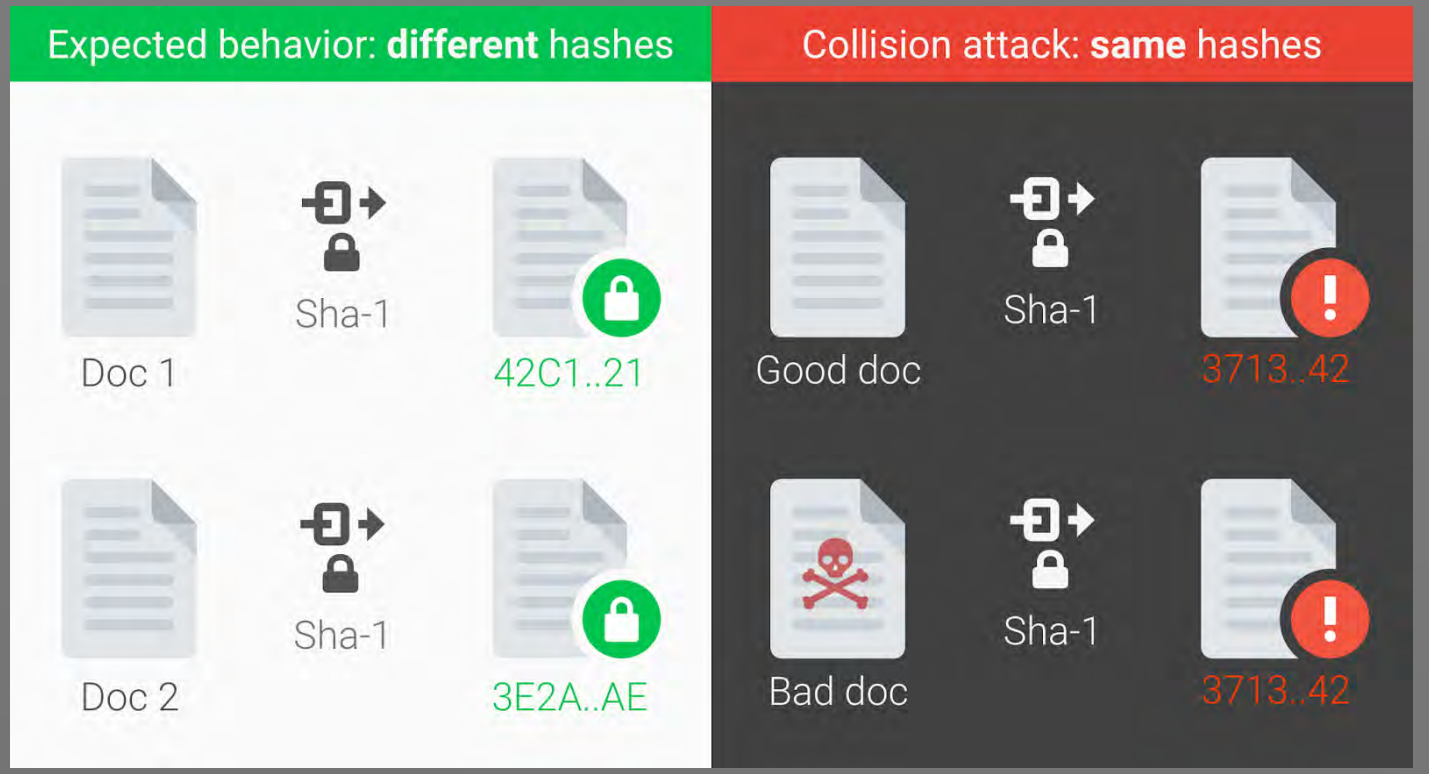
- Digital "fingerprint"
- Calculates a value for each file or disk image
- If bits in a file change, the checksum will show as invalid



Source: Wikimedia Commons

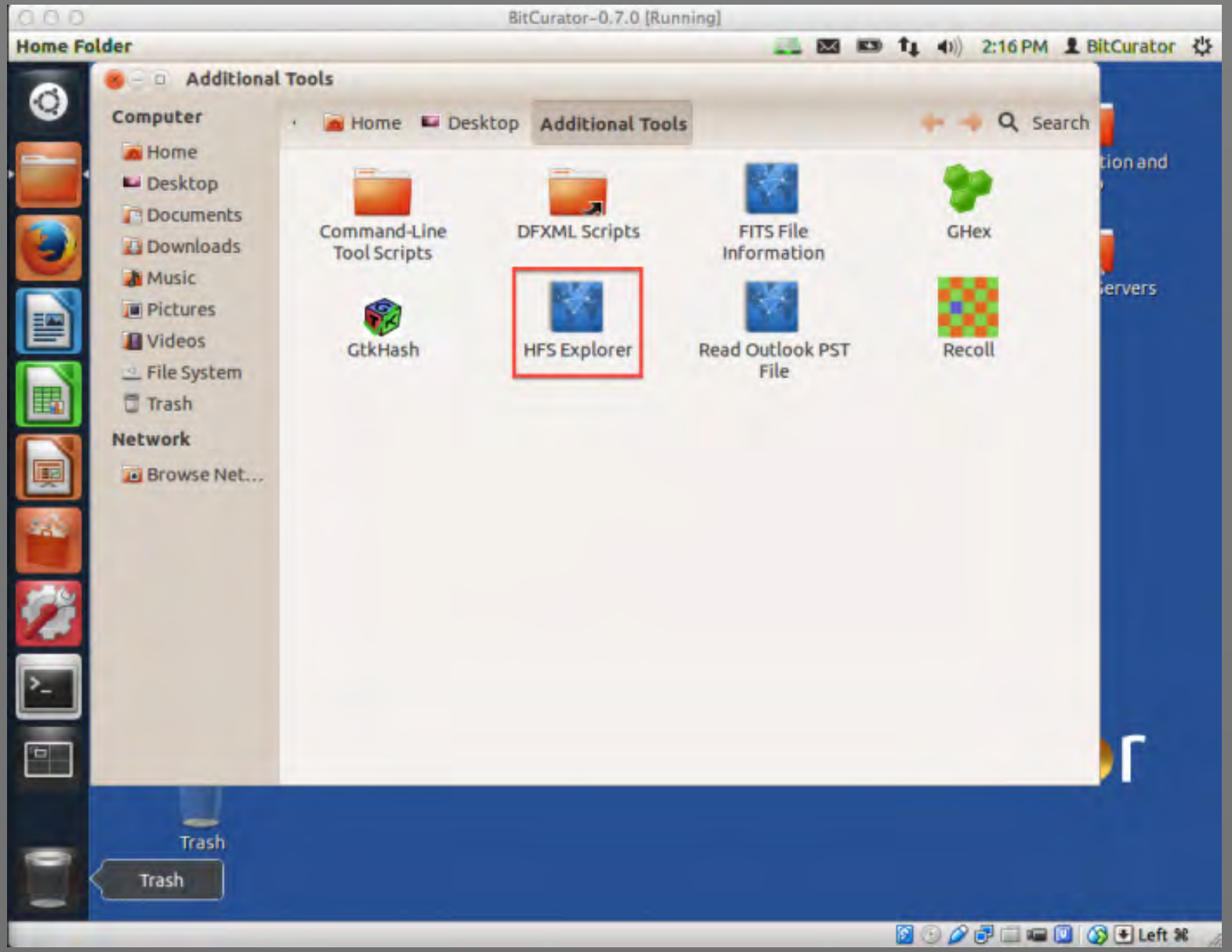
CHECKSUMS (FILE HASHES)

- Two common types of checksums
- MD5 and SHA
- Technically insecure
- Okay to use for archival work in controlled environment

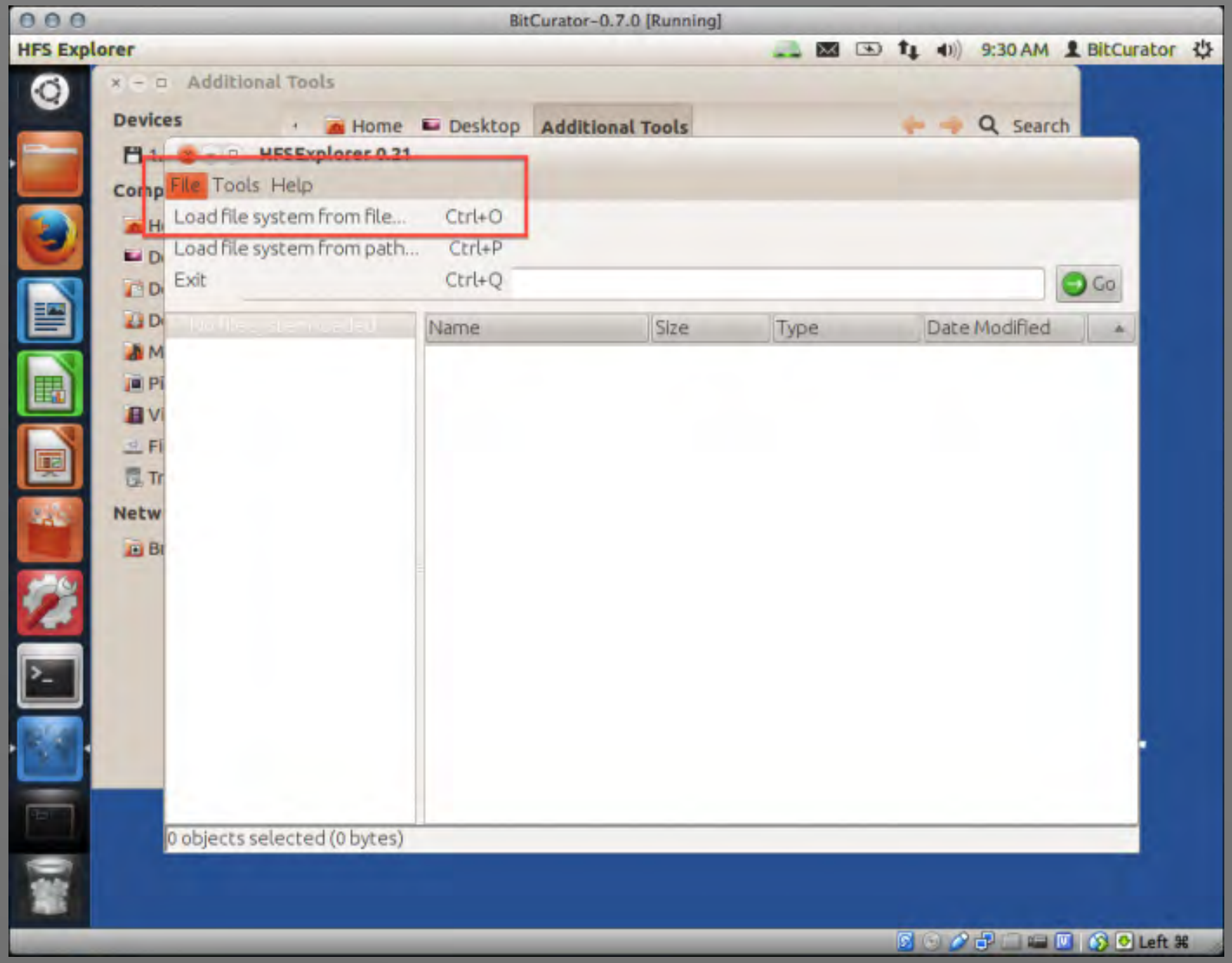


Source: Google Research & Cryptology Group Netherlands

MACINTOSH DISKS



MACINTOSH DISKS



DESCRIPTION: WHAT TO INCLUDE

- Unique Identifier
- Disk Label
- Date Ingested
- Steps Taken
- Disk image checksum

Source: XKCD

American Heritage Center
Collection Digital Management Form

Collection Title:

Accession & Accretion:

Disk Number: ahcdm_#####_###

Accession Date:

Accessioned By:

Master Location: #####

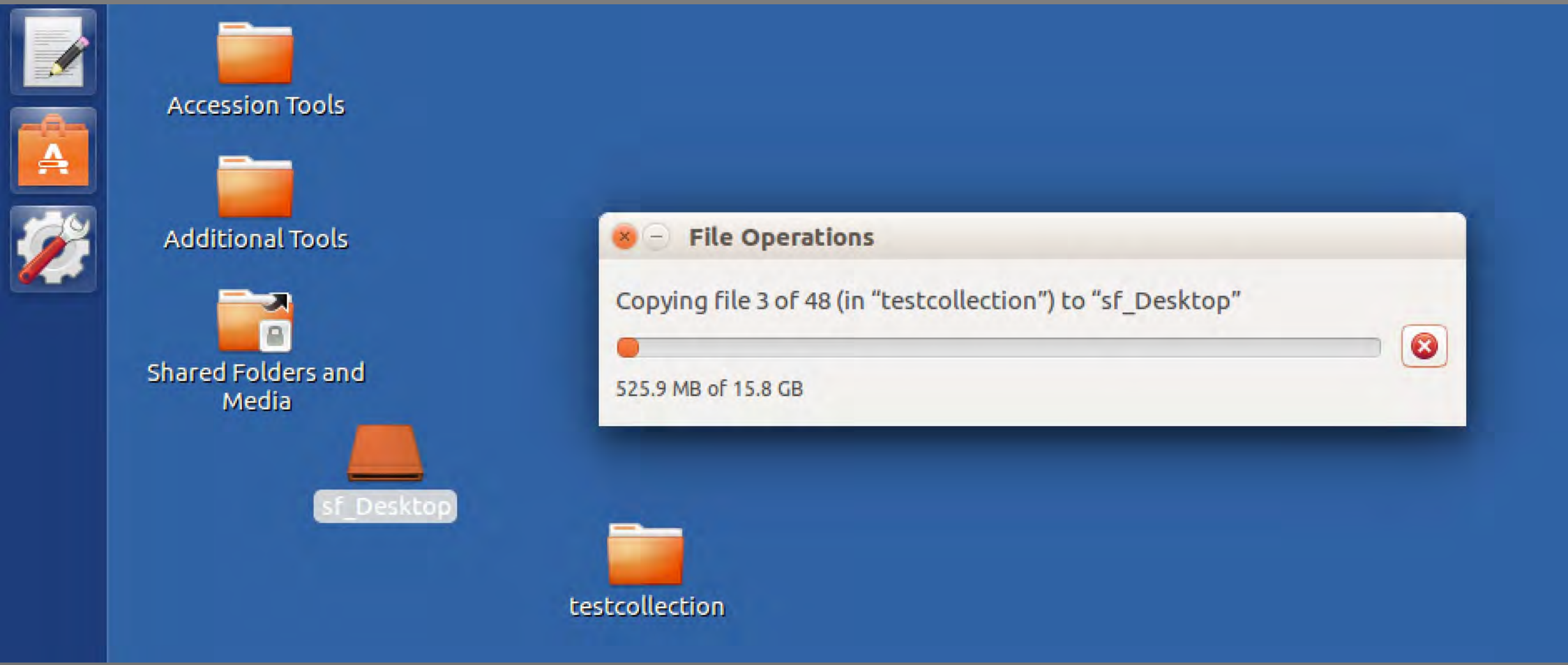
Access Location: #####

Disk Image Checksum: .md5

DISK LABEL:

OTHER NOTES:

- ☐ Virus Scan Performed
- ☐ Disk Image Created
- ☐ BitCurator Report Generated
- ☐ Data Accessioner Run
- ☐ Disk Images and Metadata Transferred to Network
- ☐ Physical Media Stored


Accession Tools
Additional Tools
Shared Folders and Media
sf_Desktop
testcollection
File Operations
Copying file 3 of 48 (in “testcollection”) to “sf_Desktop”
525.9 MB of 15.8 GB

REDUNDANCY: DIFFERING STORAGE LOCATIONS

- Two differing storage locations
- One for general use (access)
- One for limited use (dark archive)
- More backups are okay! (differing geographic areas)
- Start with what you can afford and implement

(e.g. two external drives are okay; RAID drive; networked storage w/backups

YOU HAVE NOW PRESERVED DIGITAL RECORDS!

- Identified disk types and used appropriate hardware
- Disk image captured
(from BitCurator, HFS Explorer, KryoFlux, or FC5025)
- File contents saved in unmodified form
(used hardware or software write-blocker)
- Checksums created (MD5 or SHA256)
- Disk information saved in separate file
(date, label, preservation actions)
- Two copies in different locations

QUESTION AND ANSWER PERIOD

Tyler G. Cline
Digital Archivist and Head of Digital Programs
UW American Heritage Center

USEFUL LINKS

- OCLC Demystifying Born-Digital Series
  http://www.oclc.org/research/publications/library/born-digital-reports.html
- Oracle VirtualBox
  https://virtualbox.org/wiki/Downloads
- BitCurator
  http://wiki.bitcurator.org
- Tableau Forensic Write-Blockers
  https://www2.guidancesoftware.com
- FC5025 for 5 ¼" floppy disks
  http://shop.deviceside.com
- KruoFlux for 5 ¼" and 3 ½" floppy disks
  http://webstore.kryoflux.com