Digital forensics tools and methodologies in archival repositories

Creighton Barrett, Dalhousie University Archives

Faculty of Computer Science Seminar Series
May 16, 2017 – Jacob Slonim Room (430)
Goldberg Computer Science Building

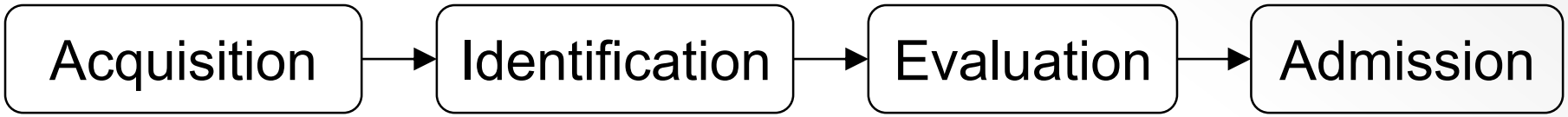
DALHOUSIE UNIVERSITY 1818 2018

Overview

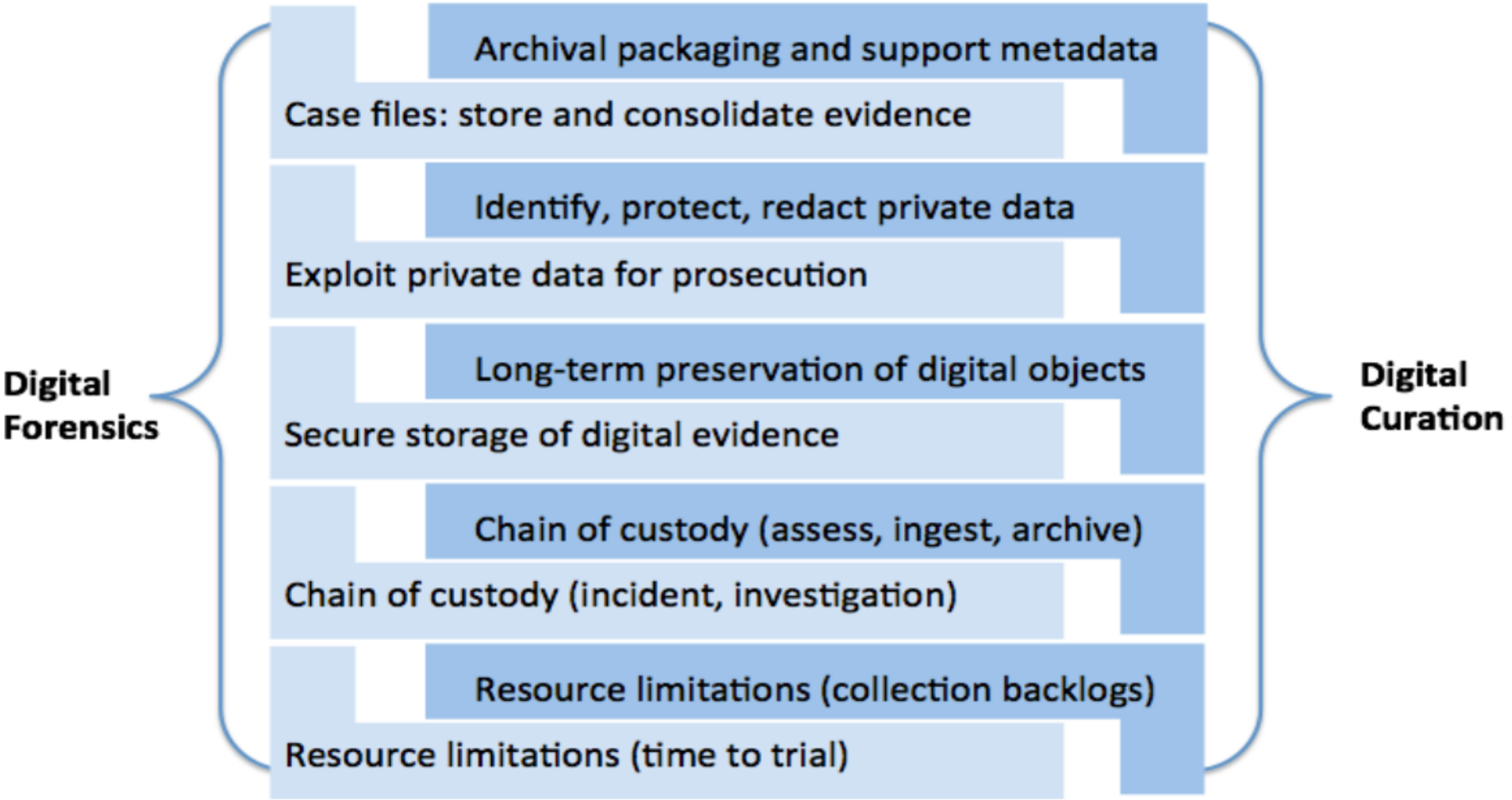
- Introduction to digital forensics in archival repositories
- Development of Dalhousie's digital forensics lab
- Forensic images
- Digital forensics tools and workflows
- Case study: Bill Freedman fonds
- Research challenges and next steps

What is digital forensics?

- Forensic science – recovery and investigation of data found in digital storage devices
- Primarily used in criminal investigations, corporate investigations (by specially trained professionals)
- Archives are adopting digital forensics techniques to support acquisition, accessioning, preservation, and access



Source: Infosec Institute, Digital Forensic Models (January 25, 2016): http://resources.infosecinstitute.com/digital-forensics-models/

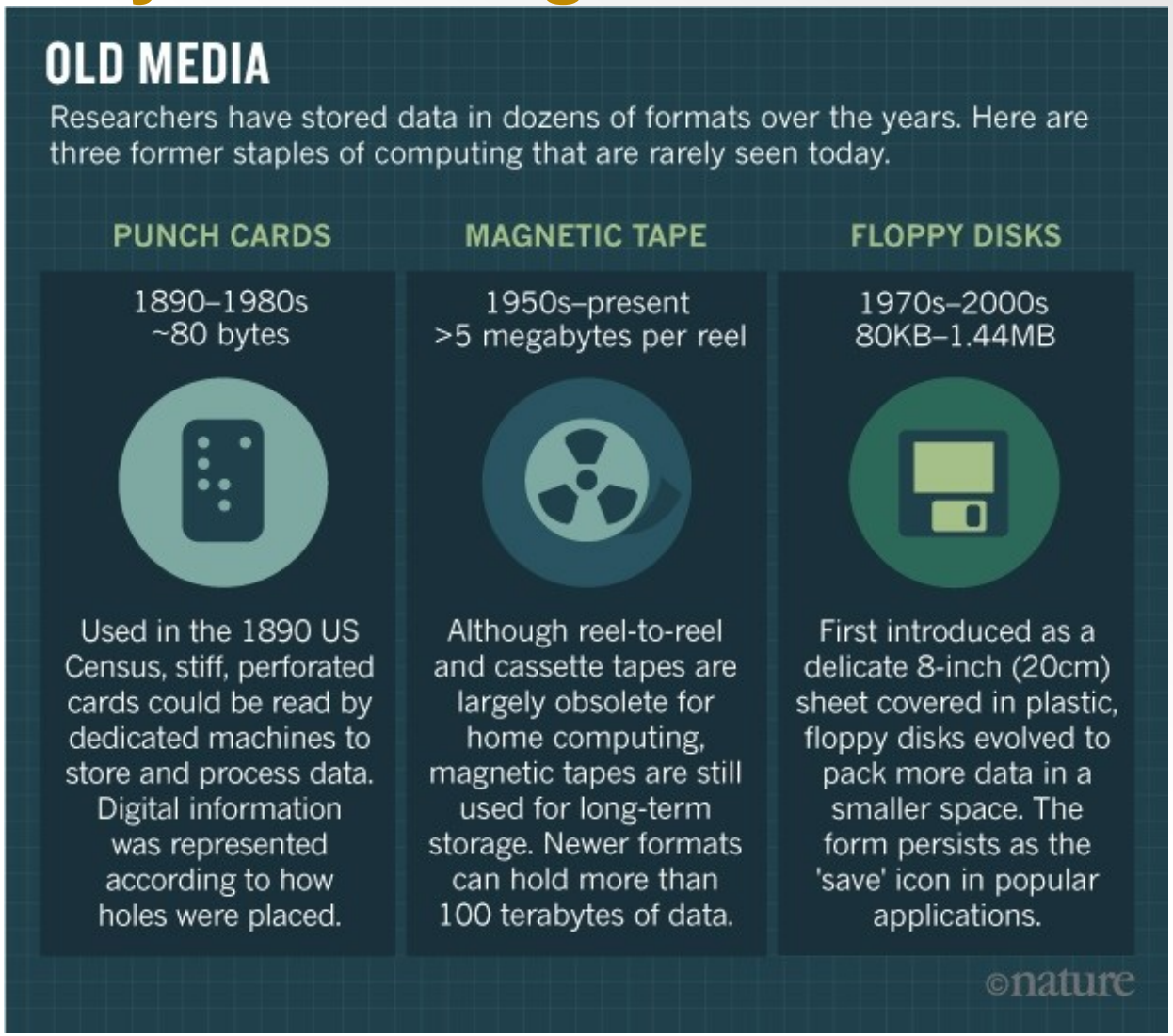


Source: Kam Woods, Preservation, Privacy, and Access: Enhancing Digital Curation Workflows with Forensic Analysis (March 21, 2017): http://wiki.bitcurator.net/downloads/kwoods-unc-digpres-v12.pdf

Unfamiliar territory for archivists

- Windows registry analysis
- Deleted files (slack space / unallocated space)
- Cached data
- Encrypted data
- Passwords
- Filesystem permissions
- Distributed systems and data

Why have a digital forensics lab?



Source: Baker, M. (2017, May 2). Disks back from the dead. *Nature, 545* (7652), 117–118. https://doi.org/10.1038/545117a

Why have a digital forensics lab?

- Archivists are now working with a wide variety of:
  - Digital storage devices
  - Computer file systems, operating systems, and software
  - File formats
- Digital storage devices are unstable and data is at risk
- Supports archival mission to preserve authenticity and integrity of records

How are archives doing digital forensics work?

- Use write-blockers to create forensic images
- Adopt forensic software (BitCurator or FTK or EnCase)
- Incorporate digital forensics tools and techniques into core archival functions
- New policy decisions (e.g., preserve forensic image or extract files?)
- Archival functions become blurred (e.g., files can be arranged before they are accessioned)

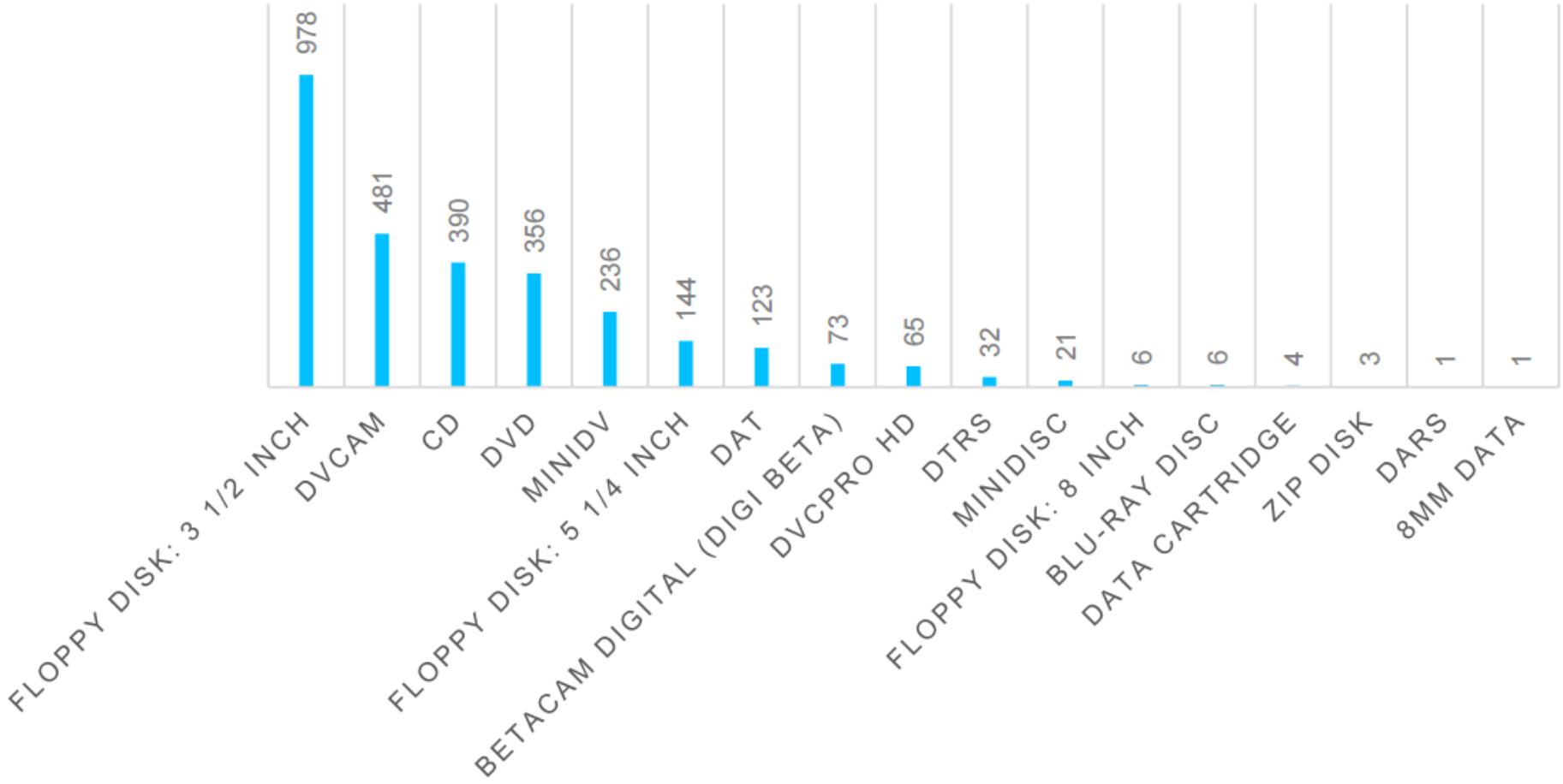
Components of forensic workstation

- Dual Intel Xeon processor
- 64 GB RAM
- NVIDIA GeForce GT740 graphics card
- Tableau T35689iu write-blocker
- AFT EX-S3 forensic card reader
- M-DISC triple burner (CD, DVD, Blu-Ray)
- 4-bay 2.5 inch RAID cage (4 x 500 GB SATA III SSDs (2 x SSDs configured in RAID 0 for database storage)
- 5-bay RAID cage (4 x 2 TB SAS HDDs configured in RAID 5 and 1 x 250 GB SATA III for OS/apps)
- Forensic Toolkit (FTK) and BitCurator software

Timeline at Dalhousie

- February 2016 – Acquire forensic workstation
- May – November 2016 – Digital archives collection assessment project: http://hdl.handle.net/10222/72663
- January 2017 – Install BitCurator and Forensic Toolkit (FTK) software
- February 2017 – Advanced computer forensics training
- May 2017 – Launch digital forensics lab
- April 2017 – Dal's first time at BitCurator Users Forum

DIGITAL MEDIA CARRIER FORMATS



Forensic images

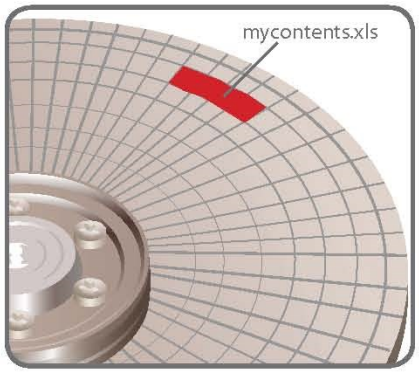
What is a forensic image?

- Complete (i.e., bit-level) copy of a hard drive or other digital storage media
- Includes unallocated space and slack space
- Includes operating system and file system
- Includes computer registry files, browser history, and other contextual information about how the computer was used
- Includes all files on the hard drive

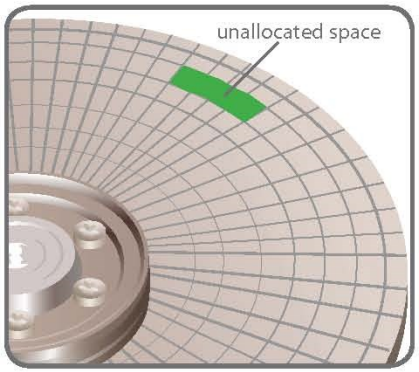
How are Deleted Files and Data Recovered?

Computers Don't Immediately Remove Data that is Deleted

Original Data

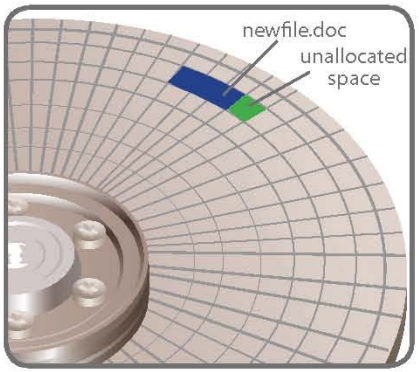


Deleted Data



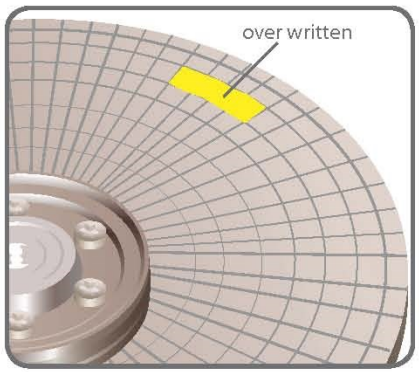
The original data is still present, but marked as unallocated space.

Partially Overwritten Data



Over time, some or all of the data can be over written. The remaining data can still be "carved" and reviewed.

Data Wiped Clean or Shreaded



The data can be wiped clean or shreaded using privacy software.

What is unallocated space?

Unallocated Space is available disk space that is not allocated to any volume. The type of volume that you can create on unallocated space depends on the disk type. On basic disks, you can use unallocated space to create primary or extended partitions. On dynamic disks, you can use unallocated space to create dynamic volumes.

PINPOINT LABORATORIES

www.pinpointlabs.com

Preserve data in slack space / unallocated space

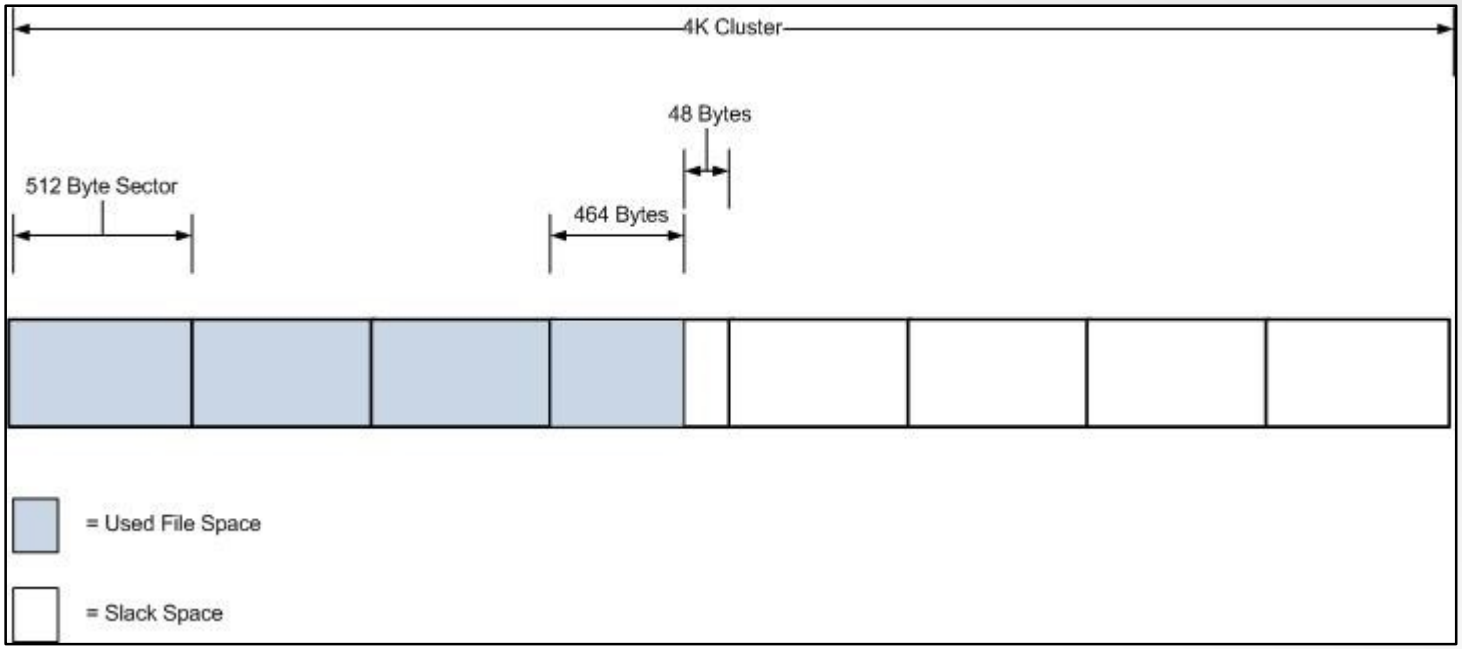


Image source: Priveon Labs Security Blog: http://blog.priveonlabs.com/sec_blog.php?title=forensic-basics-slack-space&more=1&c=1&tb=1&pb=1

Preserve information about the operating system and file system

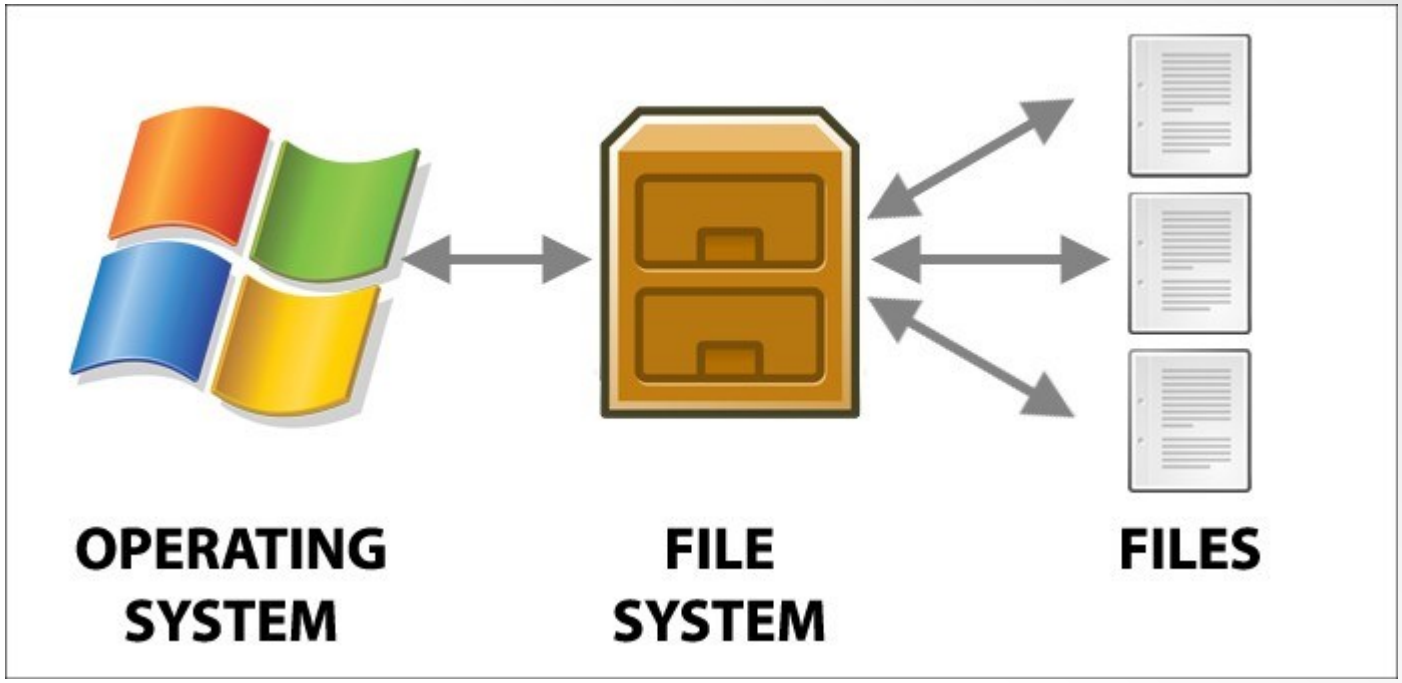


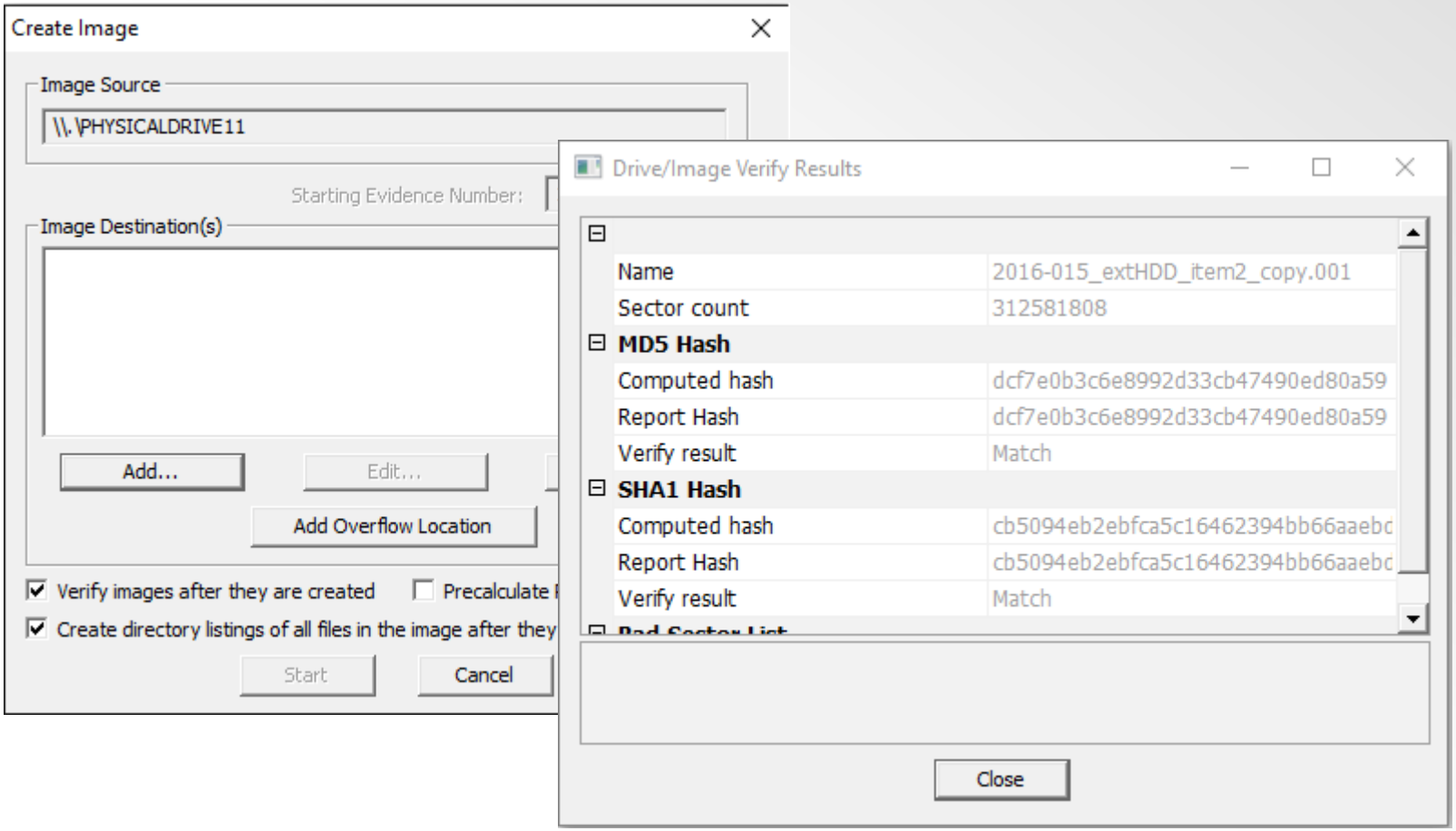
Image source: Power Data Recovery: https://www.powerdatarecovery.com/hard-drive-recovery/volume-not-contain-recognized-file-system.html

Some differences between forensic and logical images

Forensic image	Logical image
Bit-level copy of entire disk	Copy of logical partition or files in a directory
Enables registry analysis, password recovery, decryption, etc.	Not usually suitable for robust forensic analysis
Disk image is size of entire hard drive, regardless of how many files are stored	Disk image is size of files copied from logical partition or directory
Suitable for law enforcement + digital archives backlog	Suitable for packaging known files in a directory

11S45J9439ZVJ4YV036DCR REV:0B
lenovo
intel
CHINA
30V Serial ATA
ASAP

FTK Imager – Create and verify image



Digital forensics tools


FORENSIC COMPUTERS, INC.
Tableau Forensic Bridge
FORENSIC COMPUTERS
READ ONLY
M-DISC
LG
LIAN LI


Tableau Forensic Bridge
T356789iu
GUIDANCE SOFTWARE
Write Block
Read/Write
Power
Host
Device
Activity
IDE
SATA/SAS
SATA Gen3
Drive Power
PCIe
FireWire
USB 3.0


TABLEAU
2.5" IDE Hard Drive Adapter
Model TDA5-25
Pin 1


Forensic Computers
forensic-computers.com
Forensic Computers
T35689iu Forensic Bridge
Powered by TABLEAU
Pwr
Dev
Host
WrtBlk
Act
SAS
FireWire
USB 3.0
IDE
Insert SUBJECT Drive With Pins Outward
Connect Appropriate Cabling To T35689iu
HC/SDXC
MS/DUO
Forensic Series
EX-53
READ WRITE
M-DISC
LG

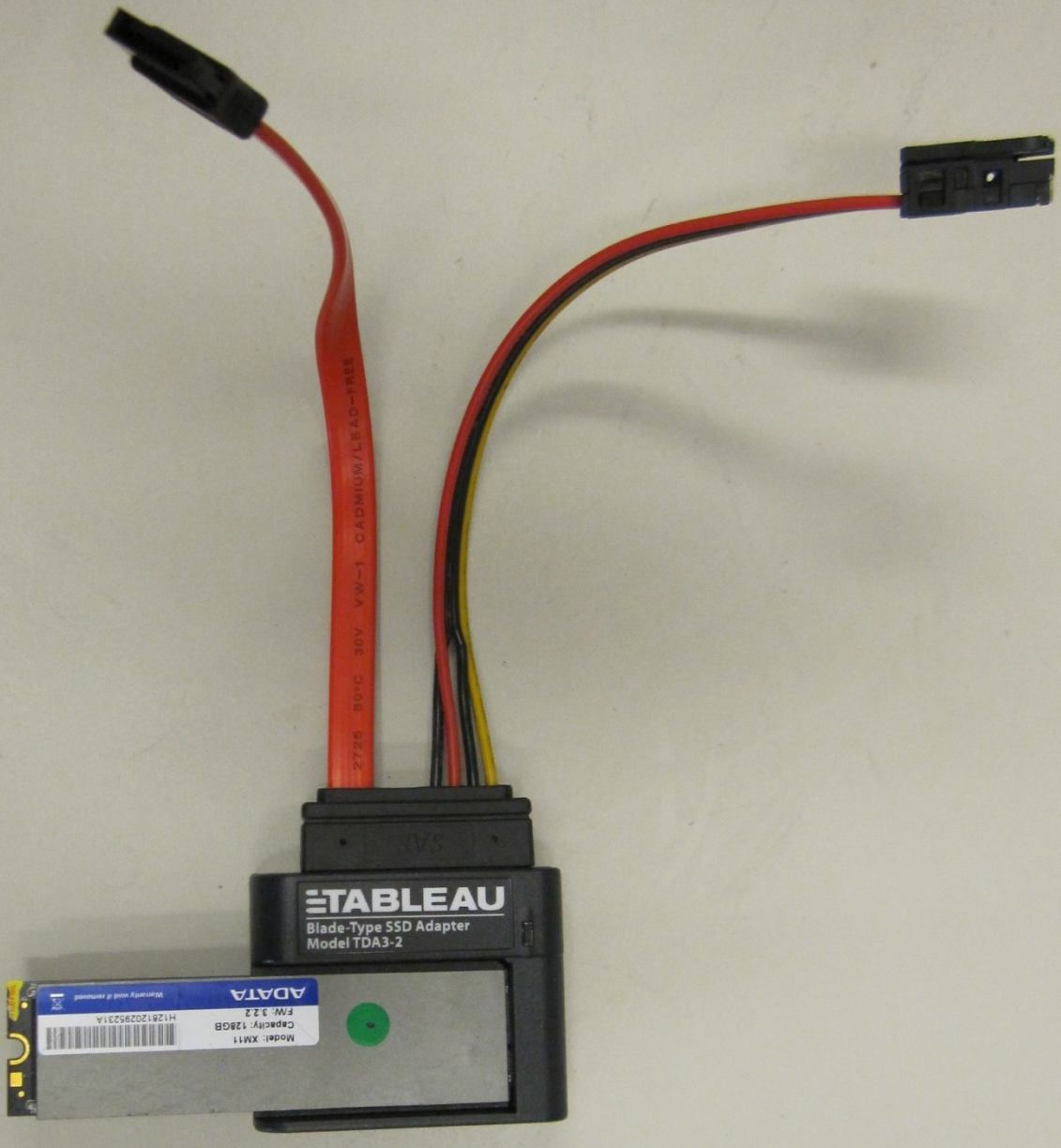
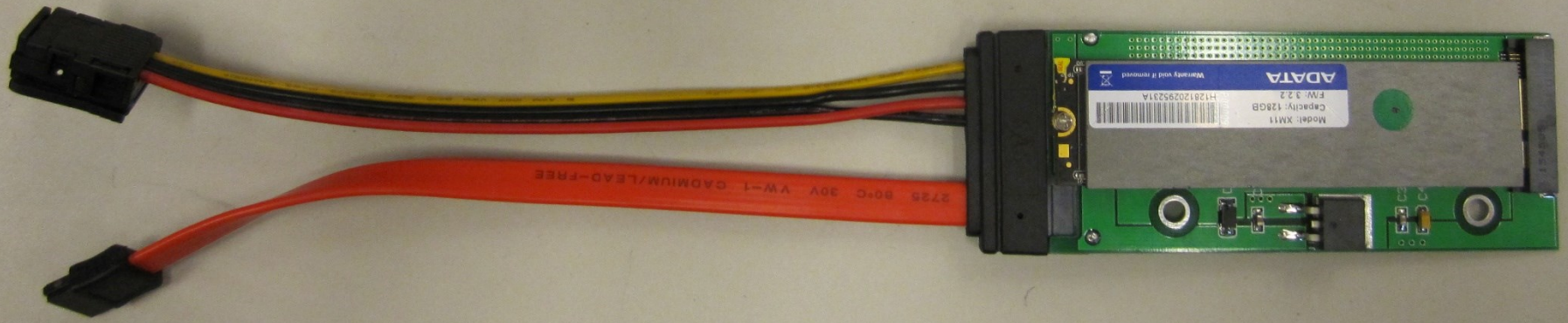
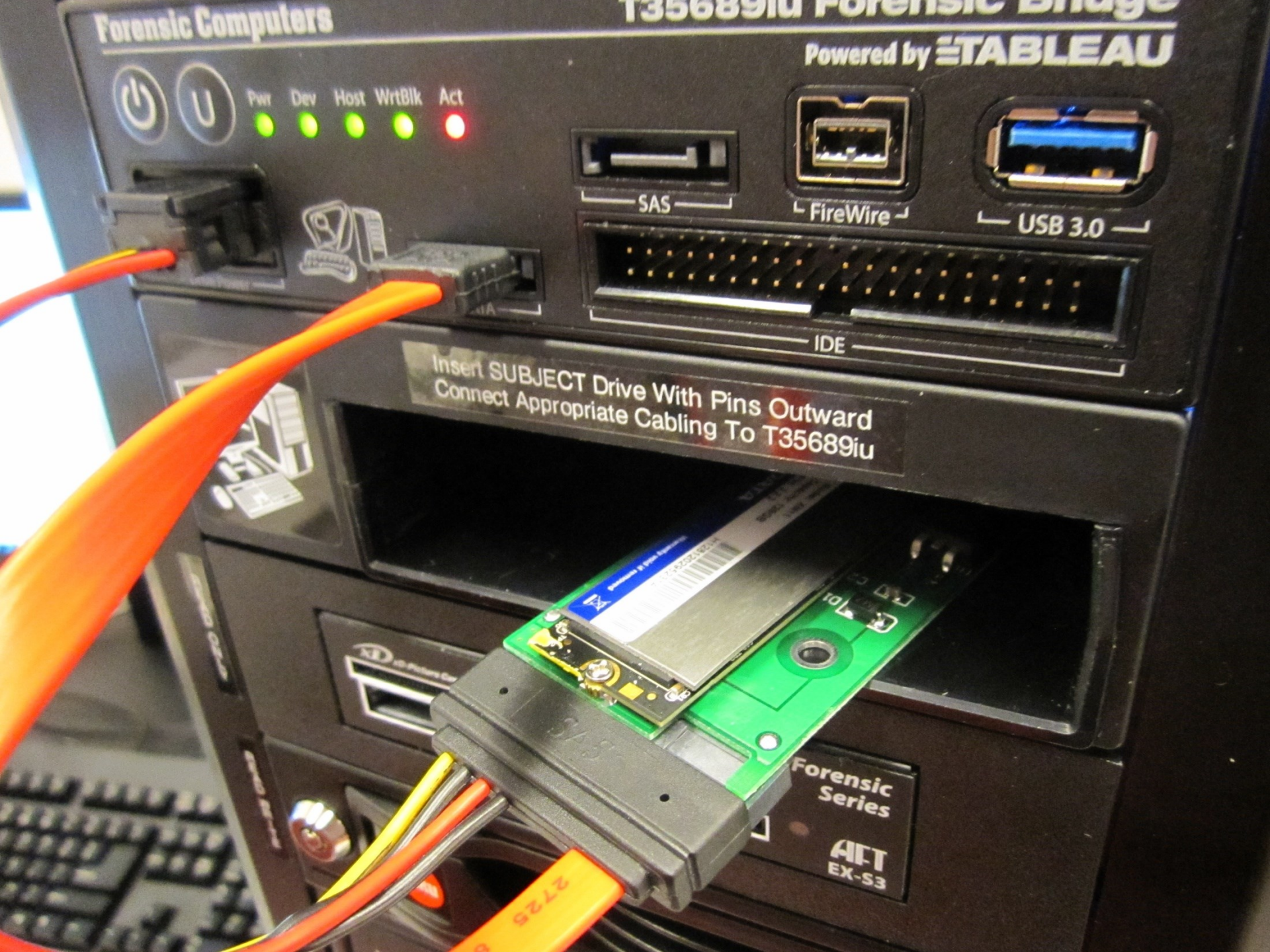


TABLEAU
Blade-Type SSD Adapter
Model TDA3-2
ADATA
Model: XM11
Capacity: 128GB
FW: 3.2.2
H12812029523 1A
Warranty void if removed
CADMIUM/LEAD-FREE
VW-1
30V
80°C
2725


ADATA
Warranty void if removed
Model: XM11
Capacity: 128GB
FW: 3.2.2
H1281202952314
2725 80°C 30V VW-1 CADMIUM/LEAD-FREE


Forensic Computers
T35689iu Forensic Bridge
Powered by TABLEAU
Pwr
Dev
Host
WrtBlk
Act
SAS
FireWire
USB 3.0
IDE
Insert SUBJECT Drive With Pins Outward
Connect Appropriate Cabling To T35689iu
Forensic Series
AFT EX-S3

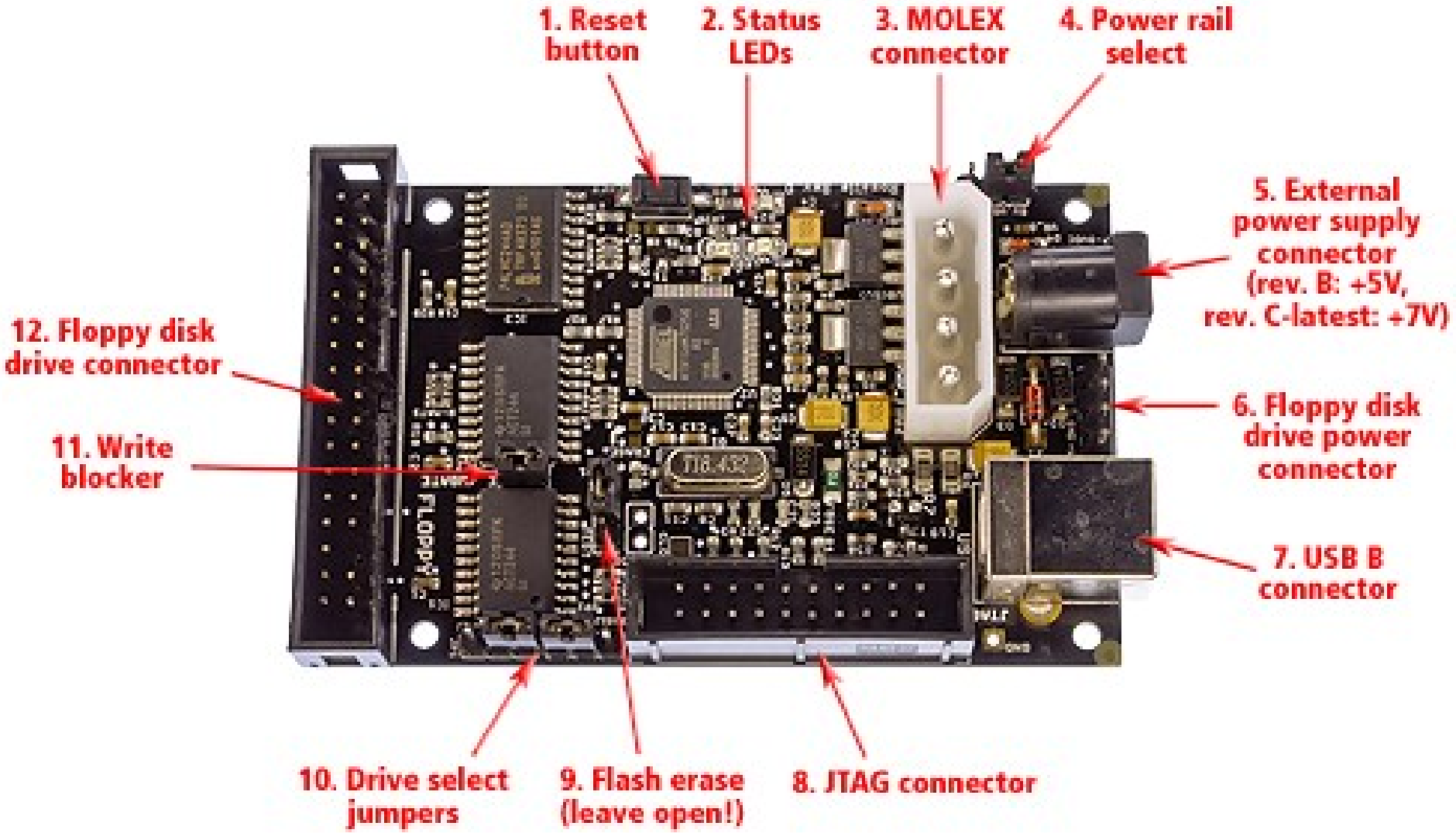


Image source: KryoFlux: https://kryoflux.com/?page=kf_tech

BitCurator

EnCase®

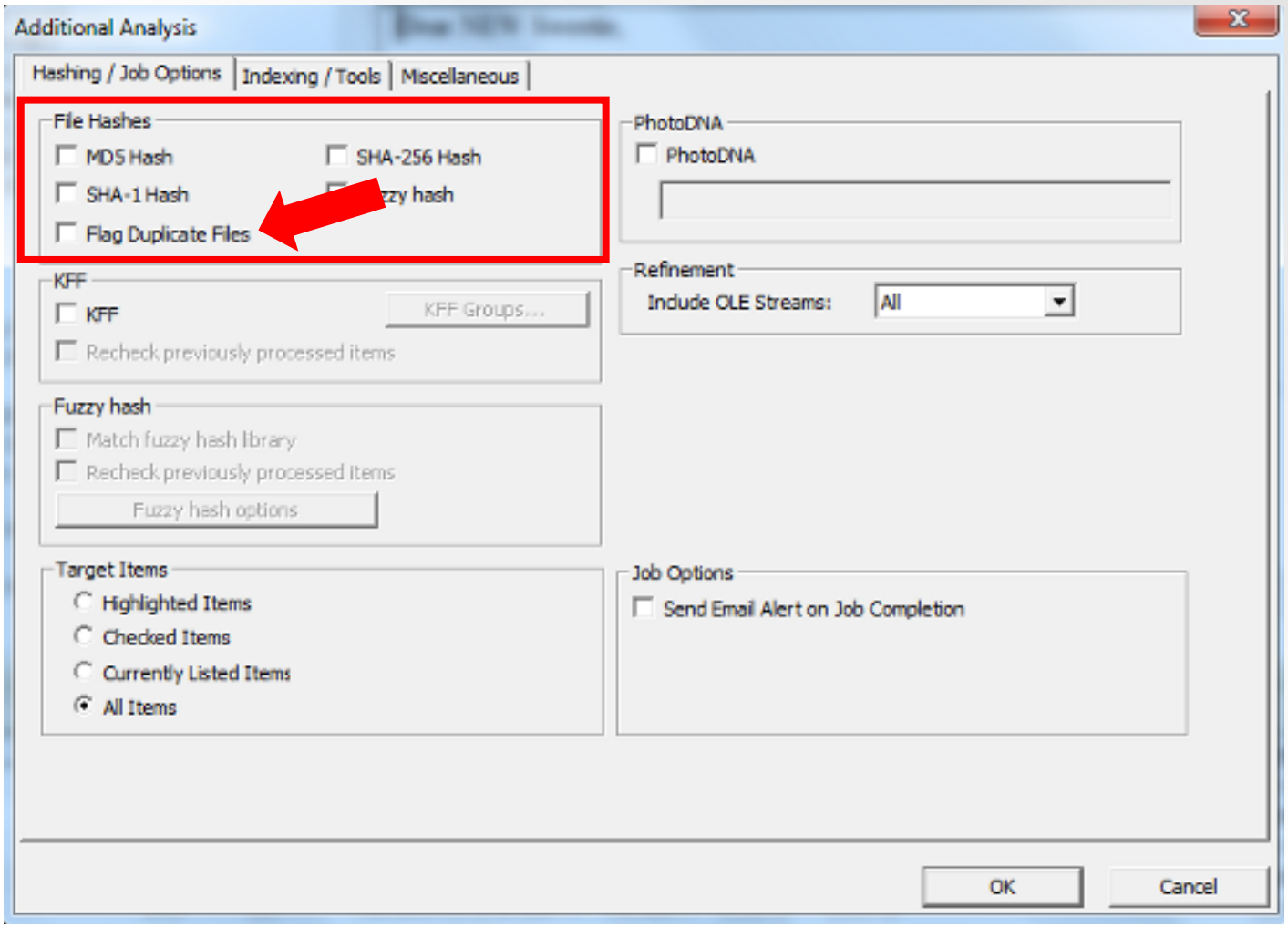
Forensic Toolkit (FTK)

- Three components
  - Database (Oracle, PostgreSQL, Microsoft SQL)
  - Graphical user interface (GUI)
  - Known file filter server (contains datasets with hash values for known file types)
- Indexing, live search, regular expression
- Oracle “Outside In” technology for previewing most file types
- Integration with other AccessData products (Registry Viewer and Password Recovery Toolkit)

FTK – Flag Duplicates

- Simple process, very time consuming, still a powerful feature
  - Scans entire file and generates MD5
  - Assigns primary status to first instance of each MD5
  - Assigns secondary status to subsequent instances of each MD5

FTK – Flag Duplicates



FSlint (finds file system “lint”)

- Duplicates
- Installed packages
- Bad names
- Name clashes
- Temp files
- Bad symlinks
- Bad IDs
- Empty directories
- Non stripped binaries
- Redundant whitespace

FSlint – Flag Duplicates

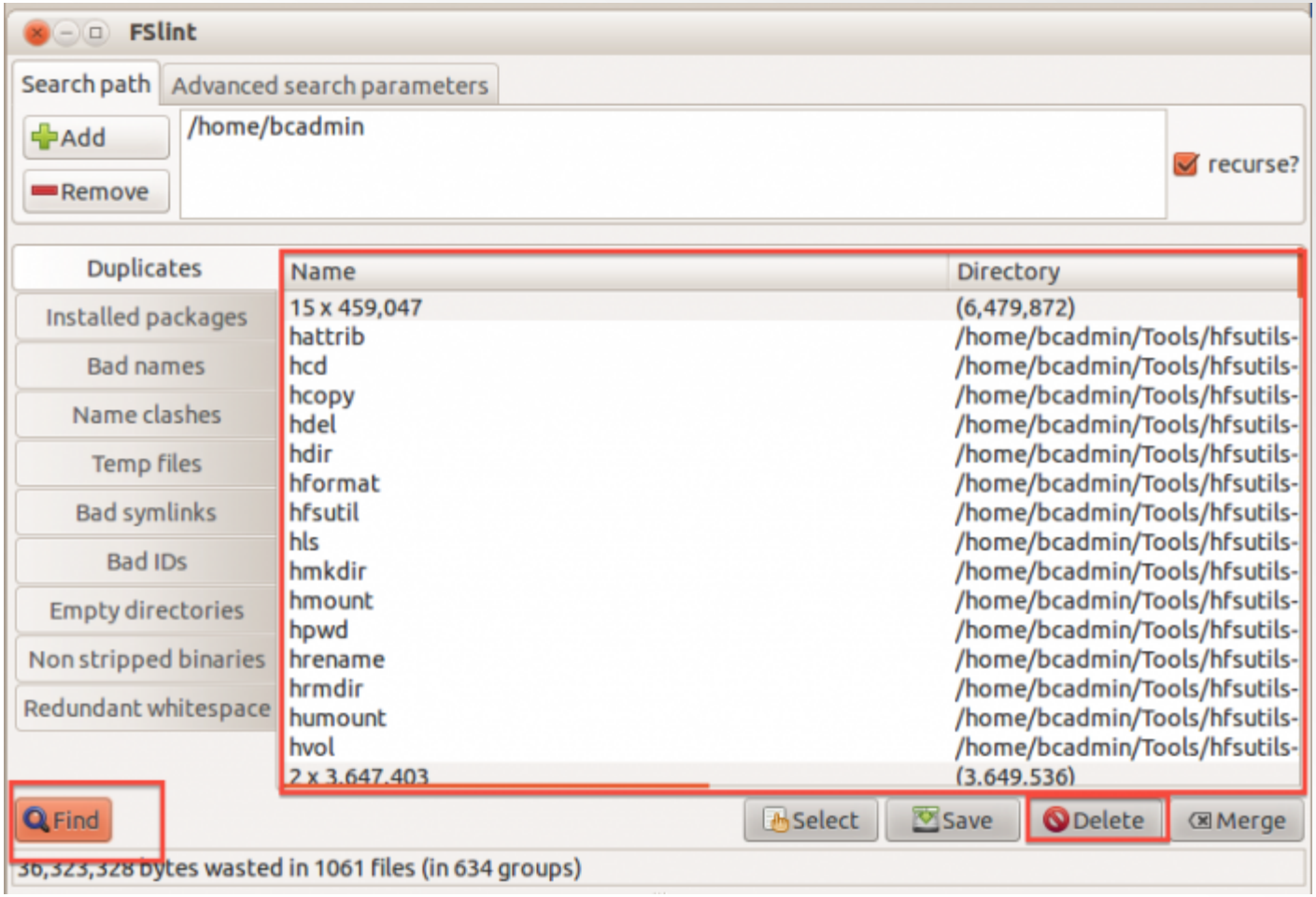


Image source (BitCurator wiki):
https://wiki.bitcurator.net/index.php?title=Identify_and_delete_duplicate_files

NSRL Reference Data Set (RDS)

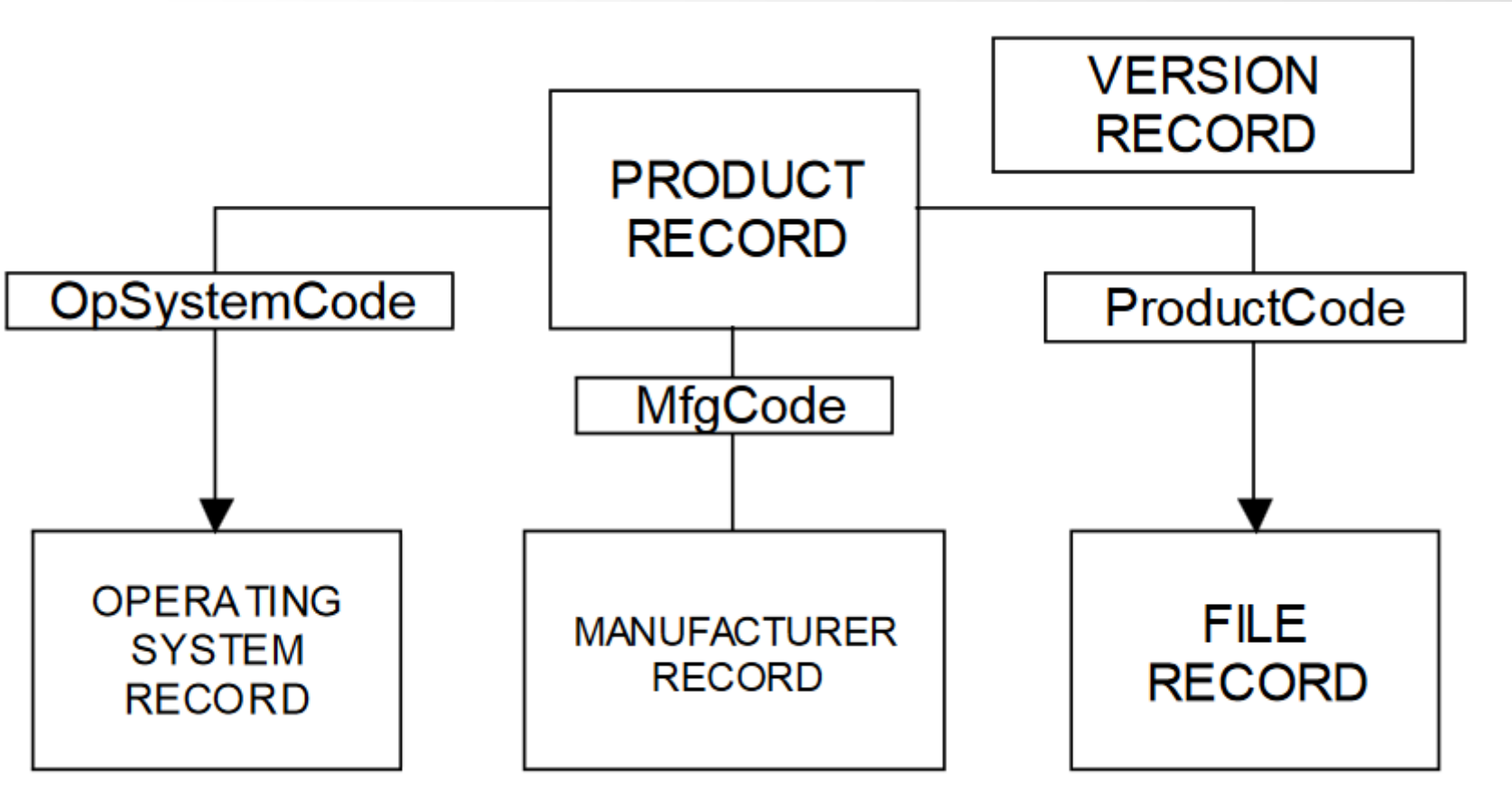


Image source (NSRL): https://www.nsrl.nist.gov/Documents/Data-Formats-of-the-NSRL-Reference-Data-Set-16.pdf

NSRL Reference Data Set (RDS)

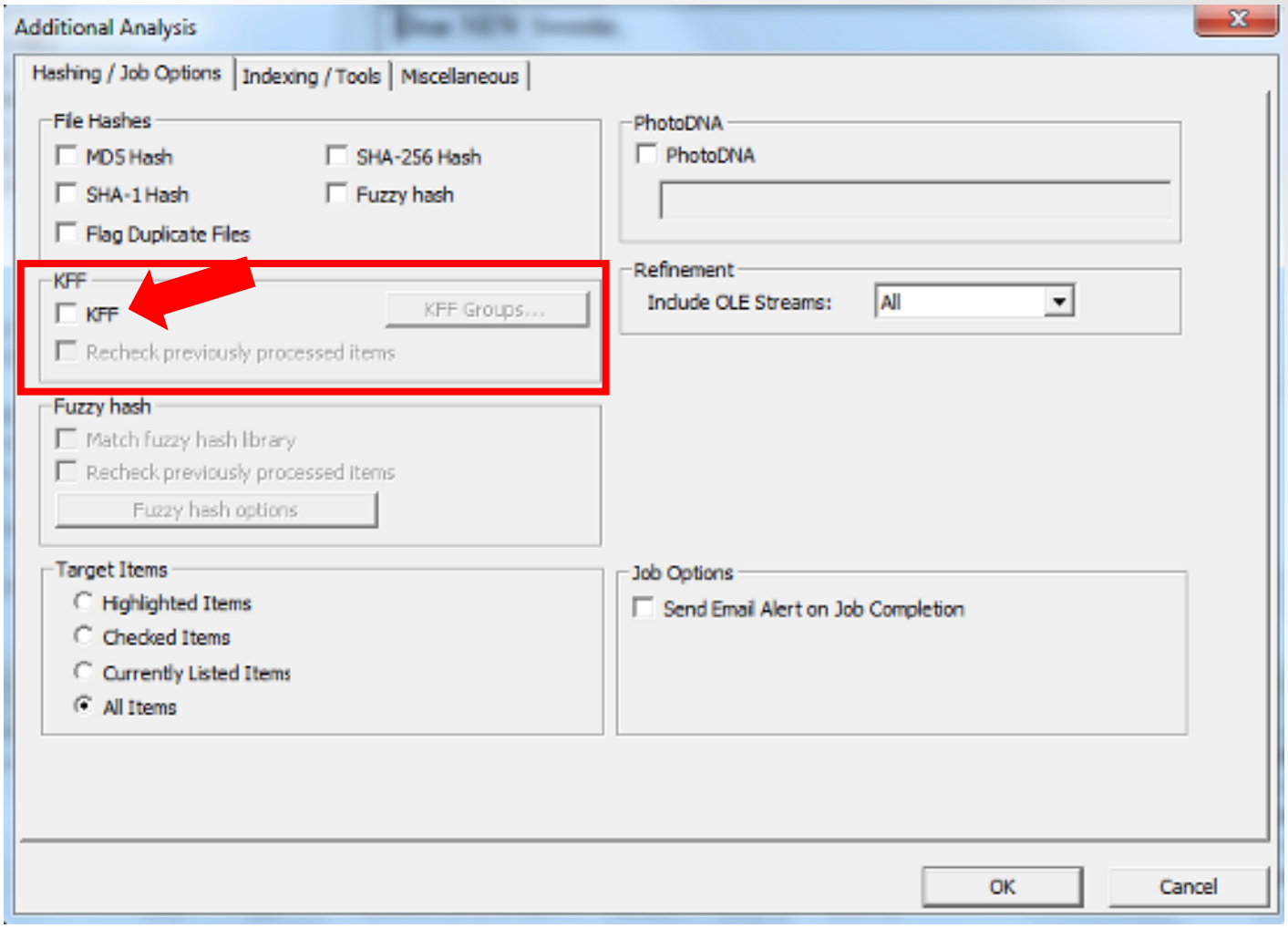
- Hashsets and metadata used in file identification
- Data can be used in third-party digital forensics tools

- RDS is updated four times each year
- As of v2.55, RDS is partitioned into four divisions:
  - Modern – applications created in or after 2000
  - Legacy – applications created in or before 1999
  - Android – Mobile apps for the Android OS
  - iOS – Mobile apps for iOS

FTK – Known File Filter (KFF)

- KFF data – hash values of known files that are compared against files in an FTK case
- KFF data can come from pre-configured libraries (e.g., NSRL RDS, DHS, ICE, etc.) or custom libraries
- FTK ships with version of NSRL RDS bifurcated into “Ignore” and “Alert” libraries
- KFF Server – used to process KFF data against evidence in an FTK case
- KFF Import Utility – used to import and index KFF data

FTK – Known File Filter (KFF)



Other tools to work with NSRL RDS

- nsrlsvr - https://github.com/rjhansen/nsrlsvr/
  - Keeps track of 40+ million hash values in an in-memory dataset to facilitate fast user queries
  - Supports custom libraries (“local corpus”)

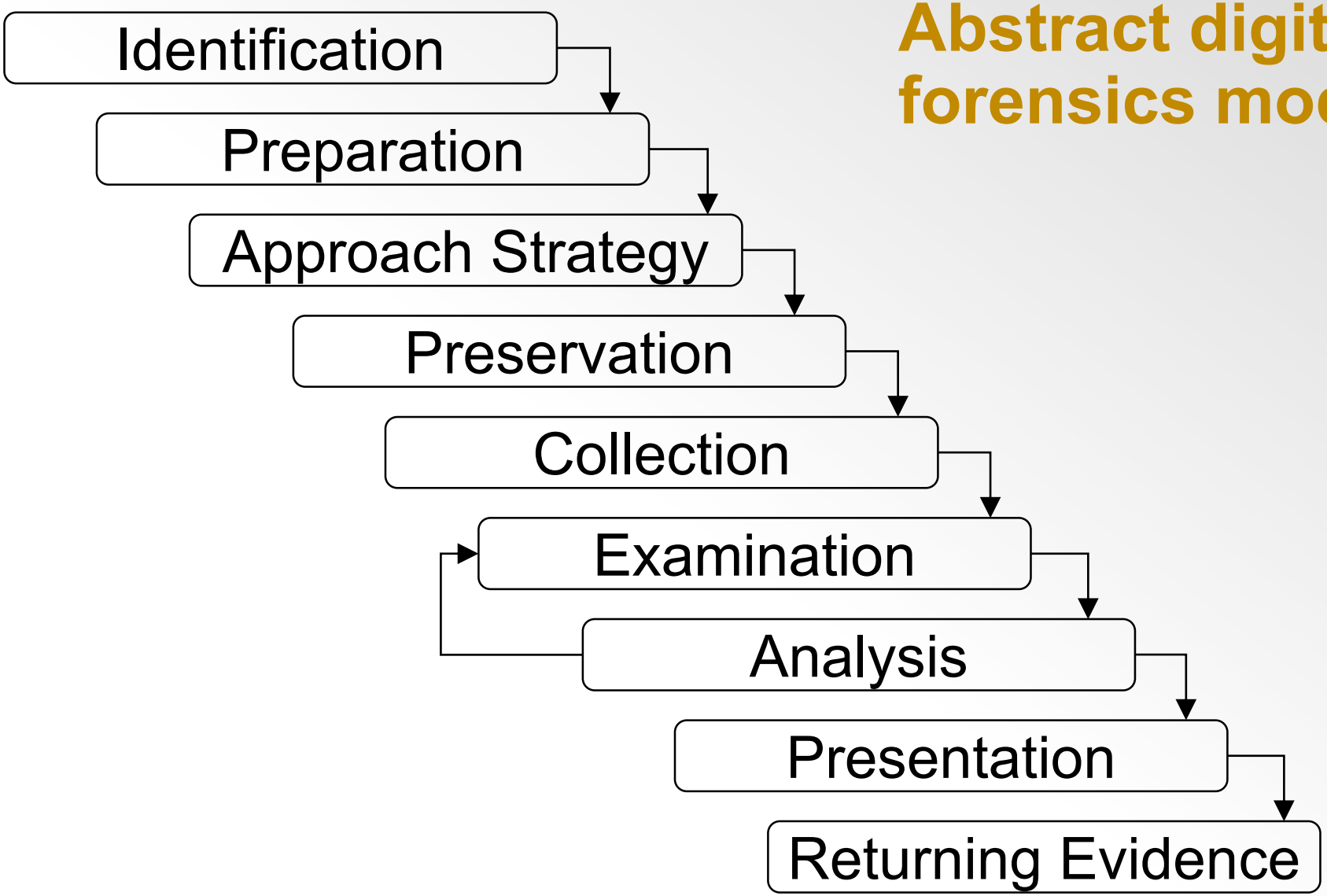
- nsrllookup - https://rjhansen.github.io/nsrllookup/
  - Command-line application
  - Works with tools like hashdeep: http://md5deep.sourceforge.net/

- National Software Reference Library - MD5/SHA1/File Name search - http://nsrl.hashsets.com

Standard	Description	Comments
FTK XML	Generated by Forensic Toolkit software	Primarily used to store templates, settings, etc.
DFXML	Digital forensics XML schema	Used in BitCurator and SleuthKit but not in FTK
PREMIS	Data Dictionary for Preservation Metadata	International metadata standard to support the preservation of digital objects
RAD	Rules for Archival Description	Canadian standard for archival description
EAD	Encoded Archival Description	XML schema for encoding and exchanging archival descriptions
ISAAR (CPF)	International Standard for Archival Authority Records	International standard for describing corporate bodies, persons, and families
EAC	Encoded Archival Context	XML schema for encoding ISAAR(CPF) records

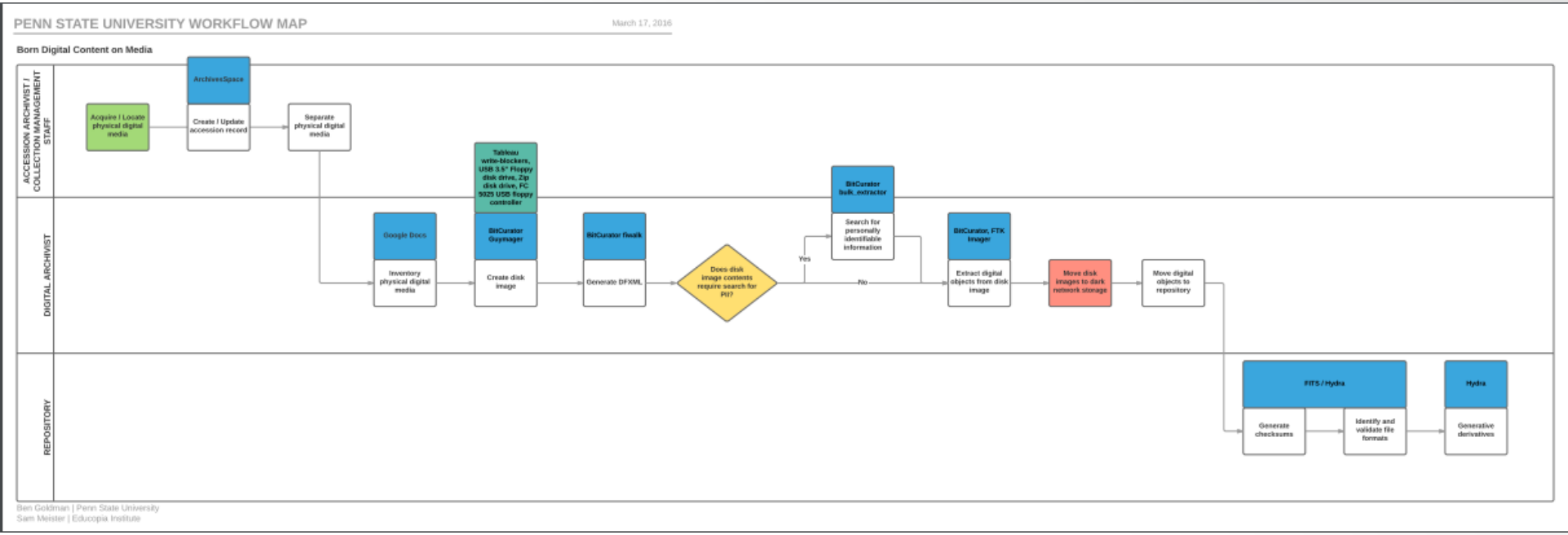
Digital forensics workflows

Abstract digital forensics model

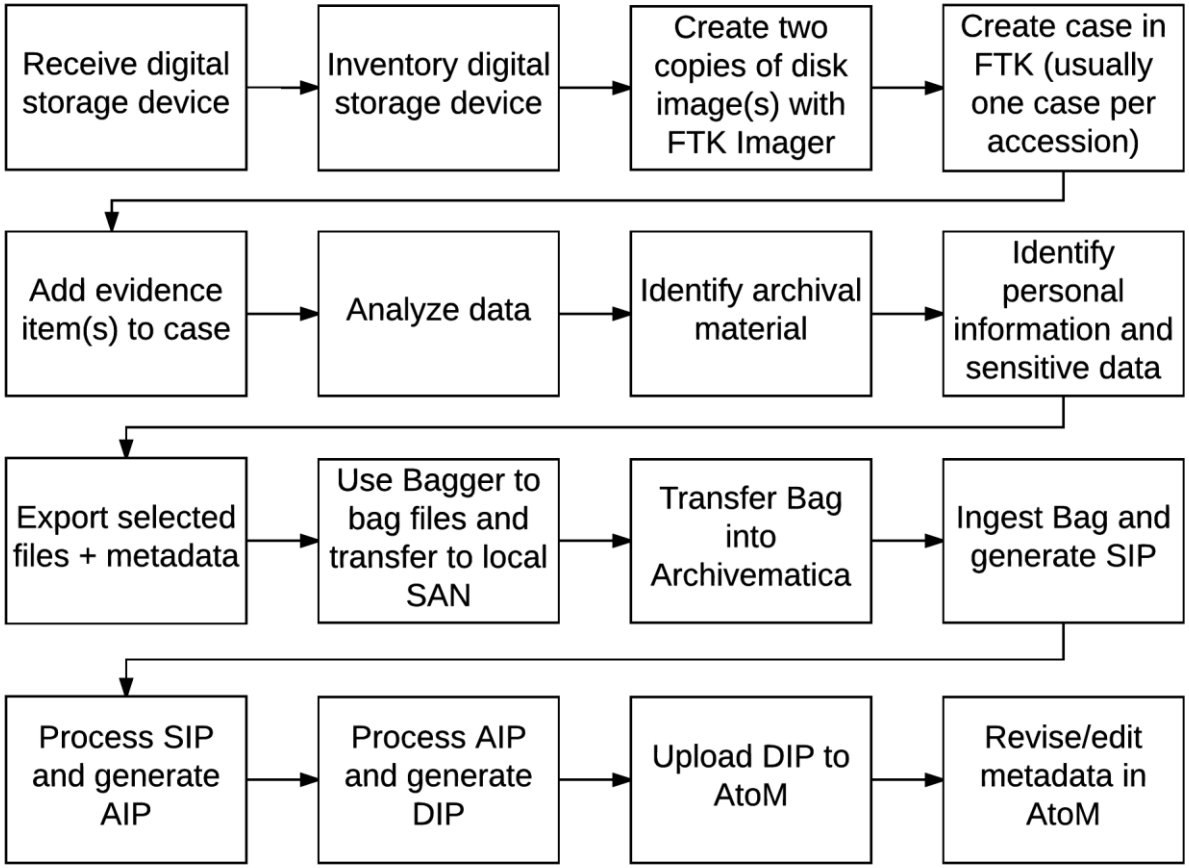


Source: Infosec Institute, Digital Forensic Models (January 25, 2016): http://resources.infosecinstitute.com/digital-forensics-models/

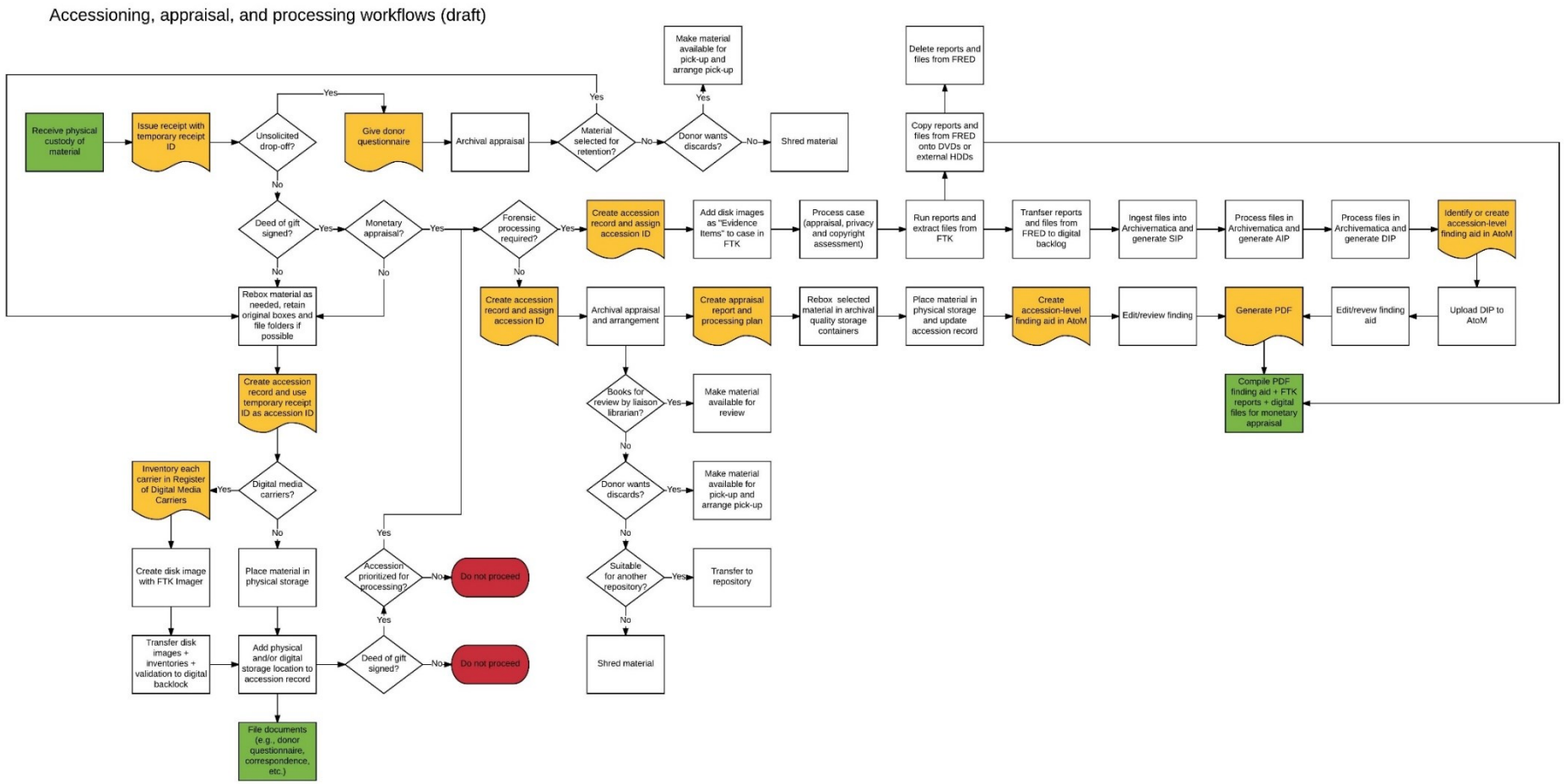
Penn State University workflow



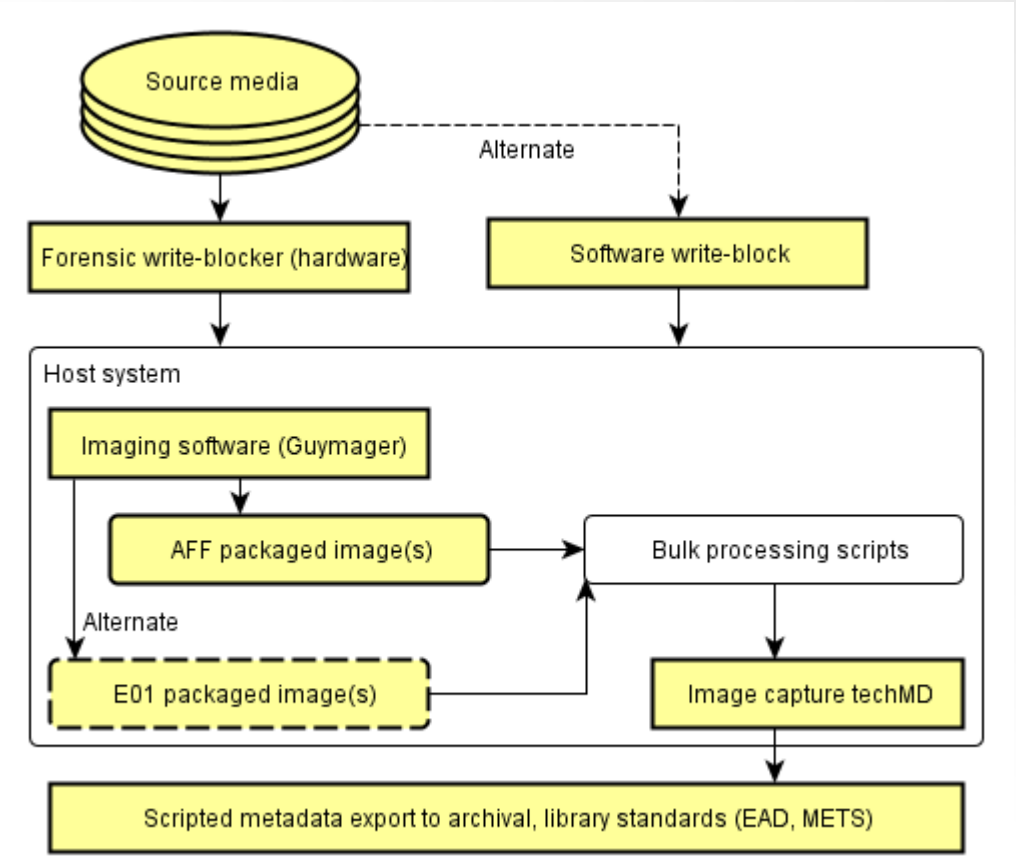
Dalhousie University workflow (draft)



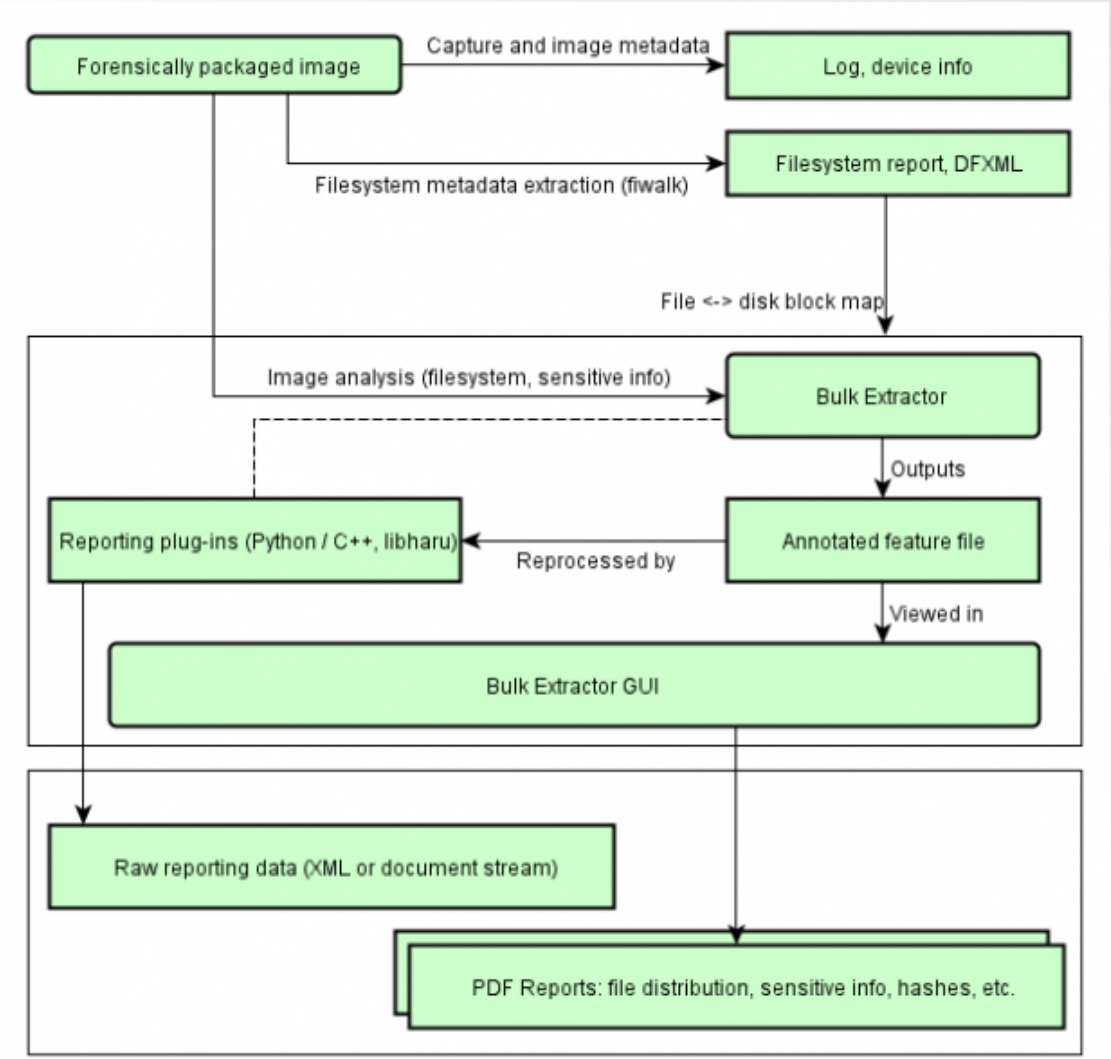
Accessioning, appraisal, and processing workflow (draft)



Forensic imaging workflow (BitCurator)



Identify privacy concerns (BitCurator)



Case study: Bill Freedman fonds

DALHOUSIE UNIVERSITY 1818 2018

Case study: Bill Freedman fonds

- Overview of the collection
- Digital forensics
- Policy, ethics, and legal context
- Next steps

Bill Freedman fonds filtered in FTK

Filter	Description	# of files	Size
Unfiltered	All files in case	26,651,084	3,568 GB
Primary status	Duplicate File indicator IS “Primary”	731,417	83.48 GB
Secondary status	Duplicate File indicator IS “Secondary”	16,569,218	271.5 GB
KFF Ignore	Match all files where KFF status IS “Ignore”	2,548,119	44.29 GB
No KFF Ignore	Match all files where KFF status IS NOT “Ignore” + KFF status IS “Not checked”	24,102,965	3524 GB
Primary status + No KFF Ignore	Match all files where duplicate file indicator IS “Primary” + KFF status IS NOT “Ignore”	626,351	71.95 GB
Actual files + Primary status + No KFF Ignore	Match all disk-bound files where duplicate file indicator IS “Primary” + KFF status IS NOT “Ignore”	103,412	61.81 GB

Policy, ethics, and legal context

- Recovery of deleted files if they appear to be archival?
- Decryption of EFS files? Other encryption methods?
- Use of Password Recovery Toolkit?
- Use of registry information, browser history, etc. to support archival appraisal?
- Modifications to standard deed of gift template?
- Monetary appraisal of born-digital archival material?

Research challenges and next steps

- Forensics in a networked environment
- Development of digital forensics workflow
- Development of lab manual
- Data transfer / storage protocols
- Finish processing Bill Freedman fonds
- Create forensic images of media identified during Digital Archives Collection Assessment