Digital Preservation and Digital Forensics: A Marriage Made in Bitstreams

Christopher (Cal) Lee
School of Information and Library Science
University of North Carolina at Chapel Hill

Digital forensics and digital preservation: Investigating good practice
Digital Preservation Coalition
26 February 2024

Some Goals When Acquiring Born-Digital Materials

- Ensure integrity of materials
- Allow users to make sense of materials and understand their context
- Prevent inadvertent disclosure of sensitive data

Fundamental Archival Principles

Provenance	• Reflect “life history” of records • Records from a common origin or source should be managed together as an aggregate unit
Original Order	Organize and manage records in ways that reflect their arrangement within the creation/use environment
Chain of Custody	• “Succession of offices or persons who have held materials from the moment they were created”[1] • Ideal recordkeeping system would provide “an unblemished line of responsible custody”[2]

1. Pearce-Moses, Richard. *A Glossary of Archival and Records Terminology*. Chicago, IL: Society of American Archivists, 2005.
2. Hilary Jenkinson, *A Manual of Archive Administration: Including the Problems of War Archives and Archive Making* (Oxford: Clarendon Press, 1922), 11.

Dangers of a “Screen Essentialist”* Perspective

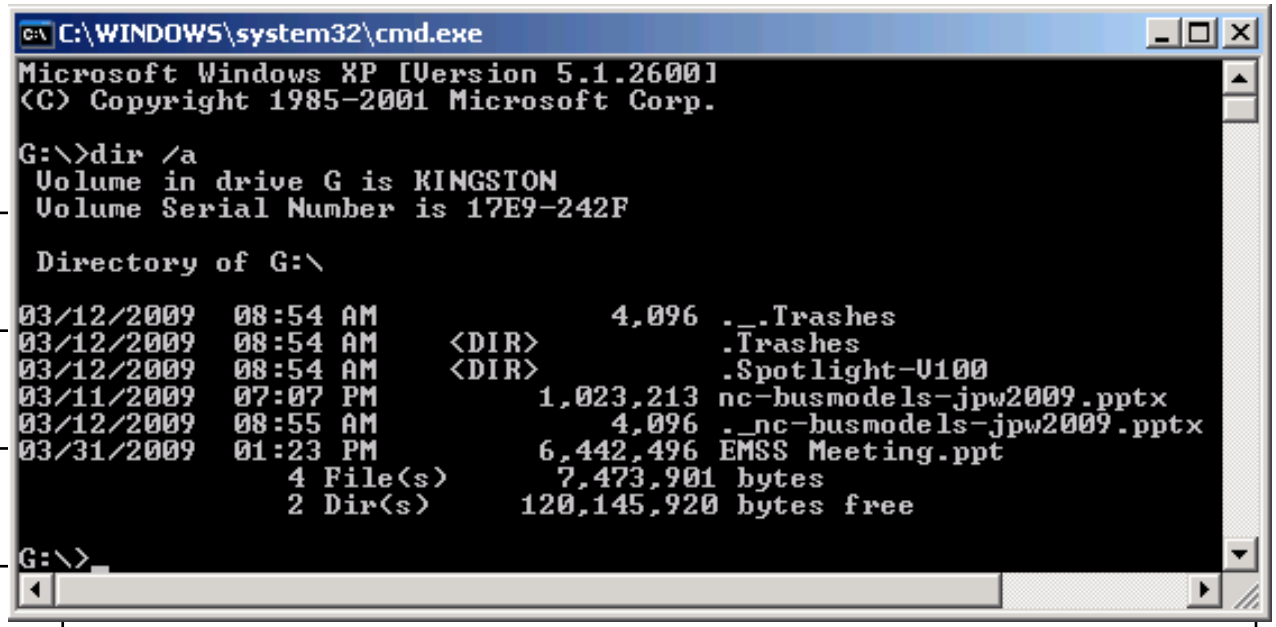
- We encounter digital objects through the activation of various computing resources - processing, storage, and input/output (I/O)
- Digital preservation involves strategies for enabling such activation over time
- One can rarely identify all essential properties of a digital object simply by looking at what happens to be on the screen during one particular encounter

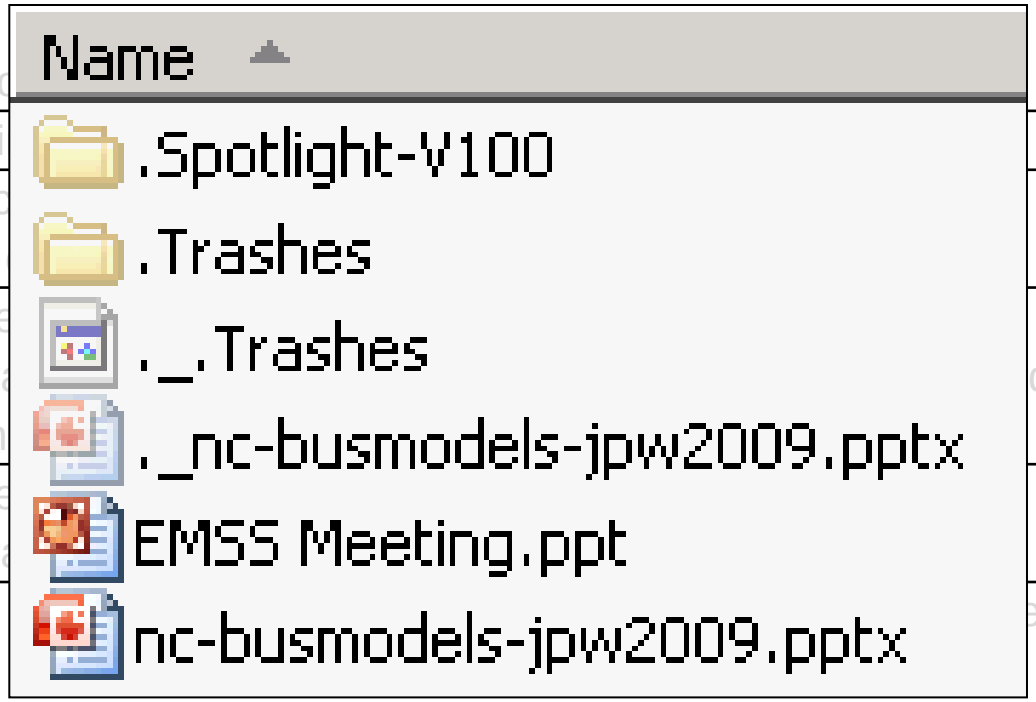
*Montfort, Nick. "The Early Materiality and Workings of Electronic Language." Modern Language Association Convention, Philadelphia, PA, December 28, 2004.

Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
6	In-application rendering	As rendered and encountered within a specific application
5	File through filesystem	Files encountered as discrete set of items with associate paths and file names
4	File as “raw” bitstream	Bitstream encountered as a continuous series of binary values
3	Sub-file data structure	Discrete “chunk” of data that is part of a larger file
2	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

Level	
Aggregation of objects	
Object or package	
In-application rendering	
File through filesystem	"ls" at directo
File as "raw" bitstream	Openi
Sub-file data structure	Extrac value
Bitstream through I/O equipment	Conne genera comm d
Raw signal stream through I/O equipment	Conne genera
Bitstream on physical medium	Using e of the drive or pits and lands on an optical disk





Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

File as “raw” bitstream

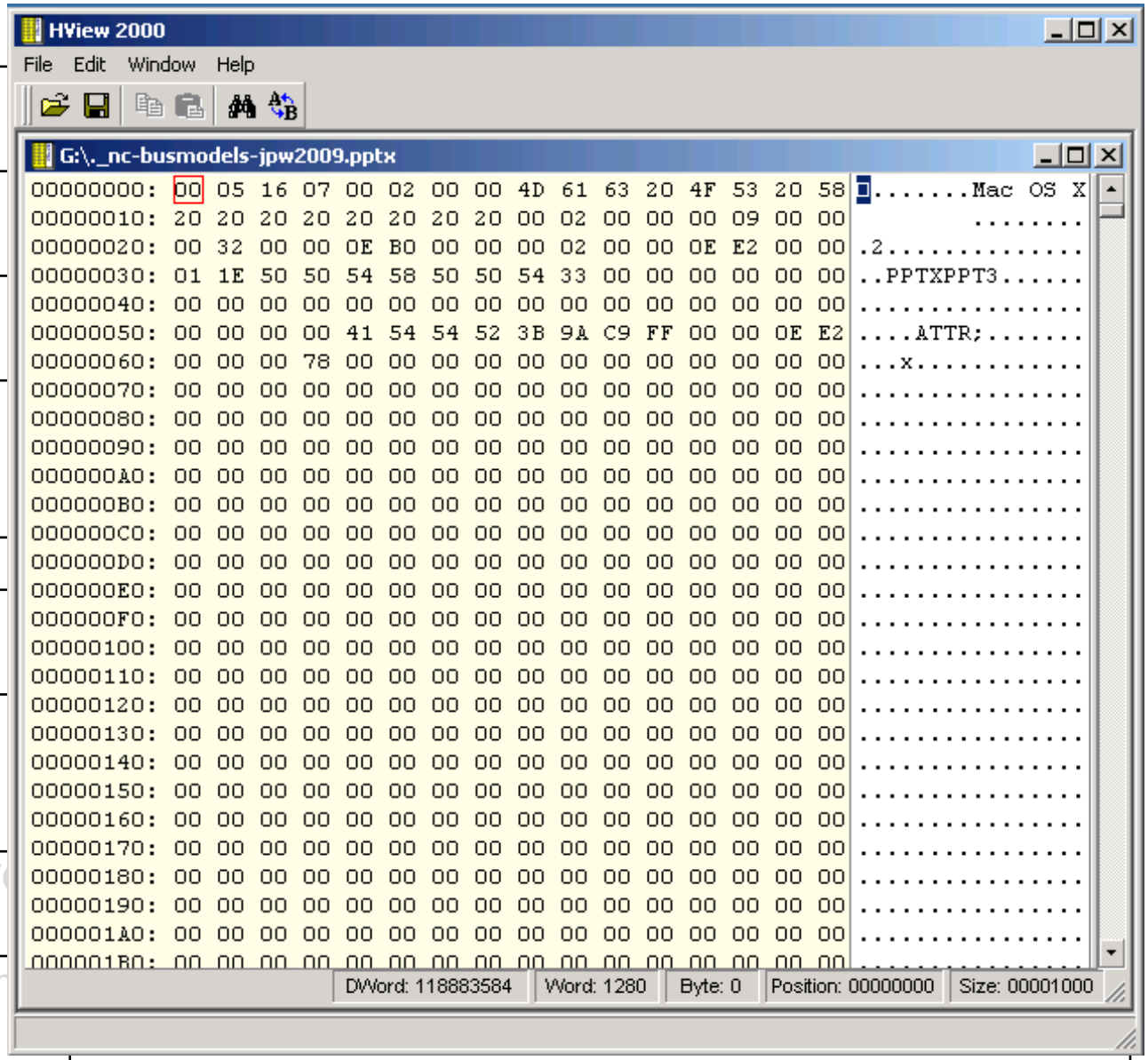
Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through I/O equipment

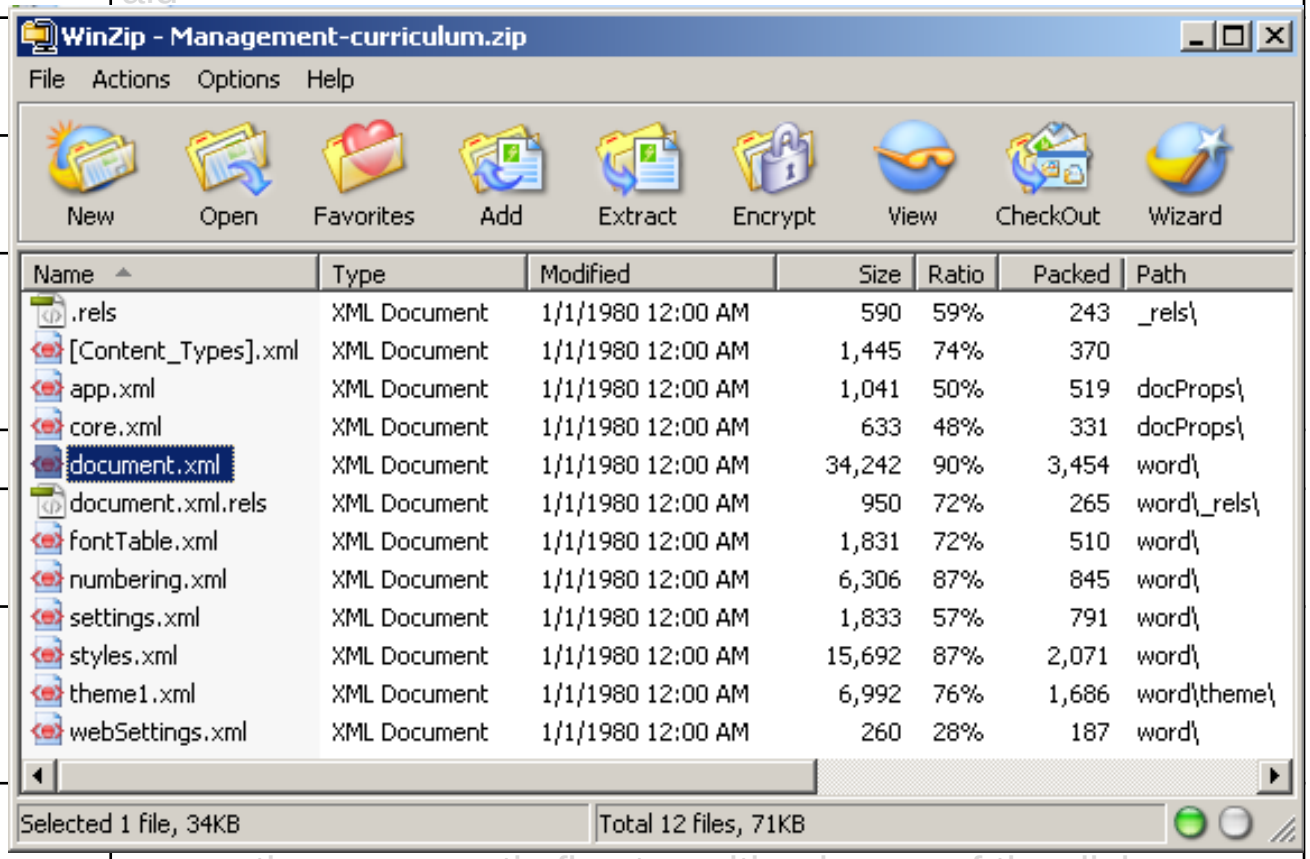
Bitstream on physical medium

drive or pits and lands on an optical disk



Interaction Examples

Level	Examples
Aggregation of objects	Browsing the contents of an archival collection using a finding aid
Object or package	
In-application rendering	
File through filesystem	
File as “raw” bitstream	
Sub-file data structure	
Bitstream through I/O equipment	
Raw signal stream through equipment	generating a magnetic flux transition image of the disk
Bitstream on physical medium	Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk



Example of EXIF Metadata from a JPEG File (Generated Using exiftool*)

```
---- ExifTool ----
ExifTool Version Number         : 9.38
---- System ----
File Name                       : IMG_20130823_151811.jpg
Directory                       : C:/Users/callee/Documents/images/digital-forensics-lab
File Size                       : 1785 kB
File Modification Date/Time     : 2013:08:23 16:36:44-04:00
File Access Date/Time           : 2013:10:14 17:13:02-04:00
File Creation Date/Time         : 2013:08:23 16:36:44-04:00
File Permissions                : rw-rw-rw-
---- File ----
File Type                       : JPEG
MIME Type                       : image/jpeg
Exif Byte Order                 : Big-endian (Motorola, MM)
Image Width                     : 2592
Image Height                    : 1944
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
---- GPS ----
GPS Img Direction               : 83
GPS Img Direction Ref           : Magnetic North
GPS Latitude Ref                : North
GPS Latitude                    : 35 deg 55' 2.24"
GPS Longitude Ref               : West
GPS Longitude                   : 79 deg 2' 57.55"
GPS Altitude Ref                : Above Sea Level
GPS Altitude                    : 0 m
GPS Time Stamp                  : 19:18:06
GPS Processing Method           : NETWORK
GPS Date Stamp                  : 2013:08:23
---- IFD0 ----
Orientation                     : Unknown (0)
Camera Model Name               : Galaxy Nexus
Modify Date                     : 2013:08:23 15:18:11
Y Cb Cr Positioning             : Centered
Y Resolution                    : 72
Resolution Unit                 : inches
X Resolution                    : 72
Make                            : Samsung
---- ExifIFD ----
Create Date                     : 2013:08:23 15:18:11
Date/Time Original              : 2013:08:23 15:18:11
Exif Version                    : 0220
Flash Energy                    : 0
Image Unique ID                 : OAEL01
Exposure Time                   : 1/17
ISO                             : 125, 0, 0
Scene Type                      : Directly photographed
Exposure Index                  : undef
Components Configuration        : Y, Cb, Cr, -
F Number                        : 2.8
Compressed Bits Per Pixel       : 0
Sensing Method                  : One-chip color area
Exposure Program                : Aperture-priority AE
Aperture Value                  : 2.6
Brightness Value                : 0
Subject Distance Range          : Unknown
Shutter Speed Value             : 1/15
Subject Distance                : 0 m
Saturation                      : Normal
Color Space                     : sRGB
Contrast                        : Normal
Metering Mode                   : Multi-spot
Flashpix Version                :
Exposure Compensation           : 0
Exif Image Height               : 1944
Max Aperture Value              : 2.6
Sharpness                       : Normal
Exif Image Width                : 2592
Focal Length                    : 3.4 mm
Digital Zoom Ratio              : 1
Light Source                    : Fluorescent
Scene Capture Type              : Standard
Flash                           : Off, Did not fire
Custom Rendered                 : Custom
White Balance                   : Auto
Exposure Mode                   : Auto
---- IFD1 ----
Compression                     : JPEG (old-style)
Image Width                     : 160
Image Height                    : 120
Thumbnail Offset                : 1239
Thumbnail Length                : 7164
---- Composite ----
Aperture                        : 2.8
GPS Altitude                    : 0 m Above Sea Level
GPS Date/Time                   : 2013:08:23 19:18:06Z
GPS Latitude                    : 35 deg 55' 2.24" N
GPS Longitude                   : 79 deg 2' 57.55" W
GPS Position                    : 35 deg 55' 2.24" N, 79 deg 2' 57.55" W
Image Size                      : 2592x1944
Shutter Speed                   : 1/17
Thumbnail Image                 : (Binary data 7164 bytes, use -b option to extract)
Focal Length                    : 3.4 mm
Light Value                     : 6.7
```

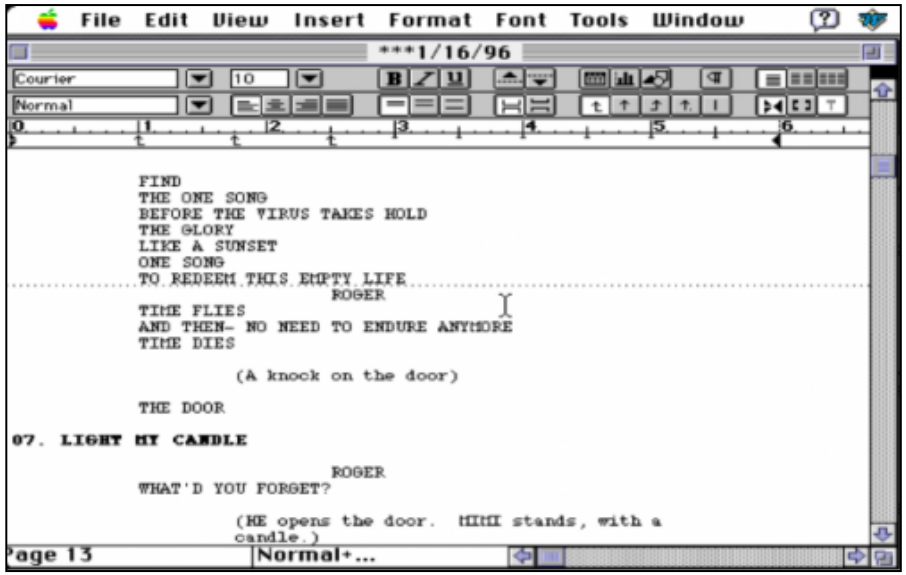
*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)

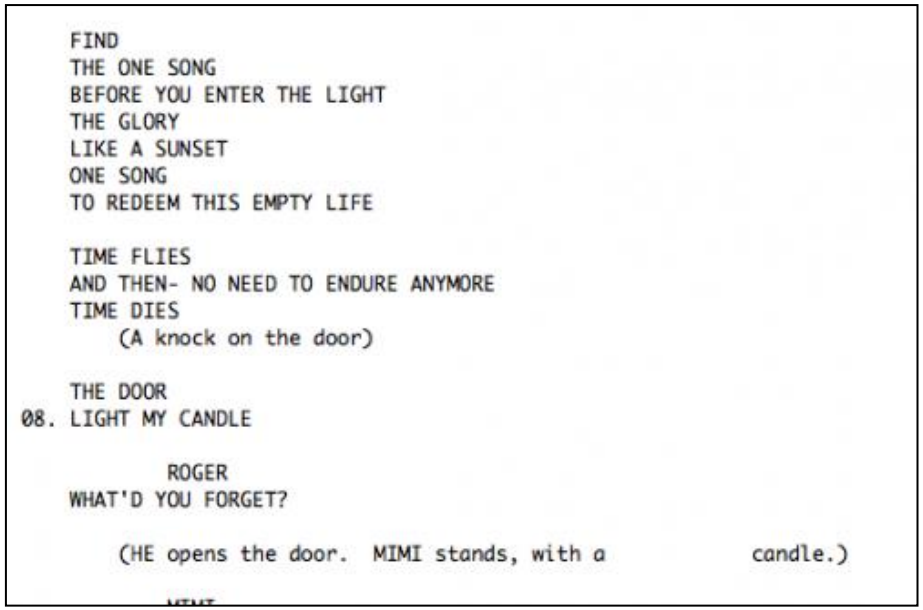
Stripping of Metadata from Images

Social Media site/system	Summary	Displays correctly?		Displays 4Cs?	Save As embedded?			Download embedded?		
		Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
500px - www.500px.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not the rights-relevant 4C fields. Metadata preserved in SaveAs file. Compared to 2013: SaveAs preserves metadata now = improvement	Green	Green	Red	Green	Green	Green	Grey	Grey	Grey
BEHANCE - www.behance.net Tested in late 2015	All rights-relevant fields and more are shown, all corectly. Embedded metadata is preserved in the SaveAs and the downloaded image file. Compared to 2013: not tested then	Green	Green	Green	Green	Green	Green	Green	Green	Green
Dropbox - www.dropbox.com Tested in late 2015	No embedded metadata shown. Embedded metadata only preserved in the downloaded image file but not in the SaveAs. Compared to 2013: also SaveAs files preserved metadata then = decline	Grey	Grey	Red	Red	Red	Red	Green	Green	Green
EyeEm - www.eyeem.com Tested in late 2015	No embedded metadata shown. SaveAs file was downscaled and all metadata was stripped off. Compared to 2013: not tested then	Grey	Grey	Red	Red	Red	Red	Grey	Grey	Grey
Facebook - www.facebook.com Tested in late 2015	No embedded metadata shown. SaveAs file preserved Copyright Notice and Creator in IIM, anything else is stripped off. Surprise: 2 IIM fields contain data generated by Facebook. Compared to 2013: at least 2 fields in IIM survive now = slight improvement	Grey	Grey	Red	Red	Orange	Red	Grey	Grey	Grey
Flickr FREE account- www.flickr.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not all rights-relevant 4Cs. Embedded metadata is stripped off SaveAs files but preserved in downloaded files. Compared to 2013: plus = any downloded file preserves metadata now; minus = even high resolution SaveAs file does not preserve it now.	Green	Green	Orange	Red	Red	Red	Green	Green	Green
Google Photo - photos.google.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not all rights-relevant 4Cs. SaveAs works only for downscaled files - only Exif metadata is preserved. Downloaded files preserved all metadata. Compared to 2013/Google+ photos: SaveAs file gets IIM and XMP metadata stripped off now = decline	Green	Green	Orange	Green	Red	Red	Green	Green	Green
Img.ly - www.img.ly Tested in late 2015	No embedded metadata shown. Embedded metadata is preserved in the high resolution/original size SaveAs image file but stripped off in a downscaled file. Compared to 2013: the loss of metadata in downscaled images was not tested in 2013.	Grey	Grey	Red	Green Red	Green Red	Green Red	Grey	Grey	Grey
Instagram - instagram.com Tested in late 2015	Tested using the Instagram iOS app v 6.4.1: No embedded metadata fields are shown. No retrieval of image files possible. Compared to 2013: then SaveAs was possible - with stripped off metadata.	Grey	Grey	Red	Grey	Grey	Grey	Grey	Grey	Grey
Joomeo - www.joomeo.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not the rights-relevant 4Cs. Embedded metadata preserved in the downloaded image files. Compared to 2013: more embedded metadata were shown then, including 4Cs = slight decline	Green	Grey	Red	Grey	Grey	Grey	Green	Green	Green
LINKED IN 2015 - www.linkedin.com Tested in late 2015	No embedded metadata shown. Only embedded Exif fields are preserved in SaveAs files. Compared to 2013: not tested then.	Grey	Grey	Red	Orange	Red	Red	Grey	Grey	Grey
Pictify - www.pictify.com Tested in late 2015	No embedded metadata shown. No retrieval of image files possible. Compared to 2013: then SaveAs was possible - with stripped off metadata.	Grey	Grey	Red	Grey	Grey	Grey	Grey	Grey	Grey
Pinterest - www.pinterest.com	No embedded metadata shown. Embedded metadata preserved in high resolution/original size images, but IIM and XMP metadata is	Grey	Grey	Red	Green	Green	Green	Grey	Grey	Grey

http://www.embeddedmetadata.org/social-media-test-results.php

Jonathan Larson Fast Save Example



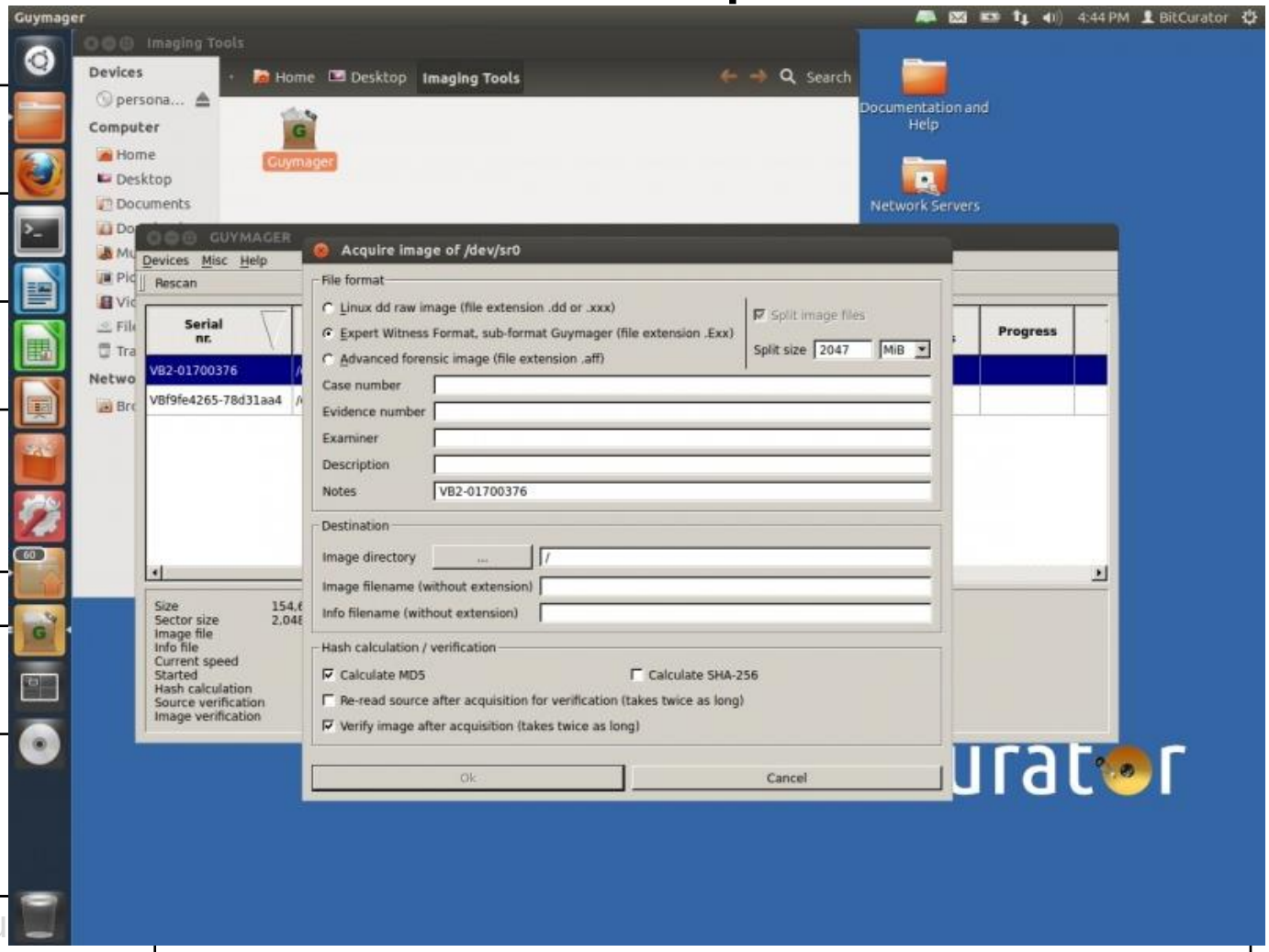


```
    FIND
    THE ONE SONG
    BEFORE YOU ENTER THE LIGHT
    THE GLORY
    LIKE A SUNSET
    ONE SONG
    TO REDEEM THIS EMPTY LIFE

    TIME FLIES
    AND THEN- NO NEED TO ENDURE ANYMORE
    TIME DIES
        (A knock on the door)

    THE DOOR
08. LIGHT MY CANDLE

            ROGER
    WHAT'D YOU FORGET?

        (HE opens the door.  MIMI stands, with a          candle.)
```

```
00028b60  09 09 09 2a 2a 2a 31 2f  31 36 2f 39 36 4f 55 52  |...***1/16/96OUR|
00028b70  20 57 45 44 44 49 4e 47  4f 4e 20 54 48 45 20 53  | WEDDINGON THE S|
00028b80  4f 46 41 53 4f 46 41 54  48 45 20 56 49 52 55 53  |OFASOFATHE VIRUS|
00028b90  20 54 41 4b 45 53 20 48  4f 4c 44 4d 45 45 54 20  | TAKES HOLDMEET |
00028ba0  59 4f 55 20 41 54 20 54  48 45 20 53 48 4f 57 49  |YOU AT THE SHOWI|
00028bb0  27 4c 4c 20 54 52 59 20  41 4e 44 20 43 4f 4e 56  |'LL TRY AND CONV|
00028bc0  49 4e 43 45 20 52 4f 47  45 52 20 54 4f 20 47 4f  |INCE ROGER TO GO|
00028bd0  43 4c 4f 53 45 20 4f 4e  43 41 4e 20 49 20 48 45  |CLOSE ONCAN I HE|
00028be0  4c 50 4d 69 73 73 20 50  6f 72 74 65 72 27 73 46  |LPMiss Porter'sF|
00028bf0  4f 52 47 45 54 20 49 54  50 41 55 4c 2a 2a 2a 2a  |ORGET ITPAUL****|
```

Douglas Reside, "'No Day But Today': A look at Jonathan Larson's Word Files," April 22, 2011, https://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files

Interaction Examples



Level	
Aggregation of objects	
Object or package	
In-application rendering	
File through filesystem	
File as “raw” bitstream	
Sub-file data structure	
Bitstream through I/O equipment	
Raw signal stream through equipment	generating a magnetic flux transition image of the disk
Bitstream on physical medium	Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

When software encounters data, it likes to change it

- Bitstreams of files (including embedded metadata)
- Filesystem information (e.g. timestamps, access permissions)
- Bitstreams residing on disks (e.g. hidden system files, content of unallocated sectors)

Digital Forensics

- "The process of identifying, preserving, analyzing and presenting digital evidence in a manner that is legally acceptable."*
- "Involves multiple methods of
  - **Discovering digital data (computer system, mobiles)**
  - **Recovering deleted, encrypted, or damaged file information**
  - Monitoring live activity
  - Detecting violations of corporate policy"**

*McKemmish, R. "What is Forensic Computing?" *Trends and Issues in Crime and Criminal Justice* 118 (1999).
**Brad Glisson, Introduction to Computer Forensics & E-discovery, University of Glasgow, Week 1 Lecture, September 2008.

Common Digital Forensics Scenarios

- Evidence seized from home/office of “person of interest” in a criminal investigation (dead forensics)
- Response to system security breach, to determine what was done, by whom and how (live forensics)

Benefits of Borrowing Digital Forensics Concepts, Tools and Methods in Digital Preservation

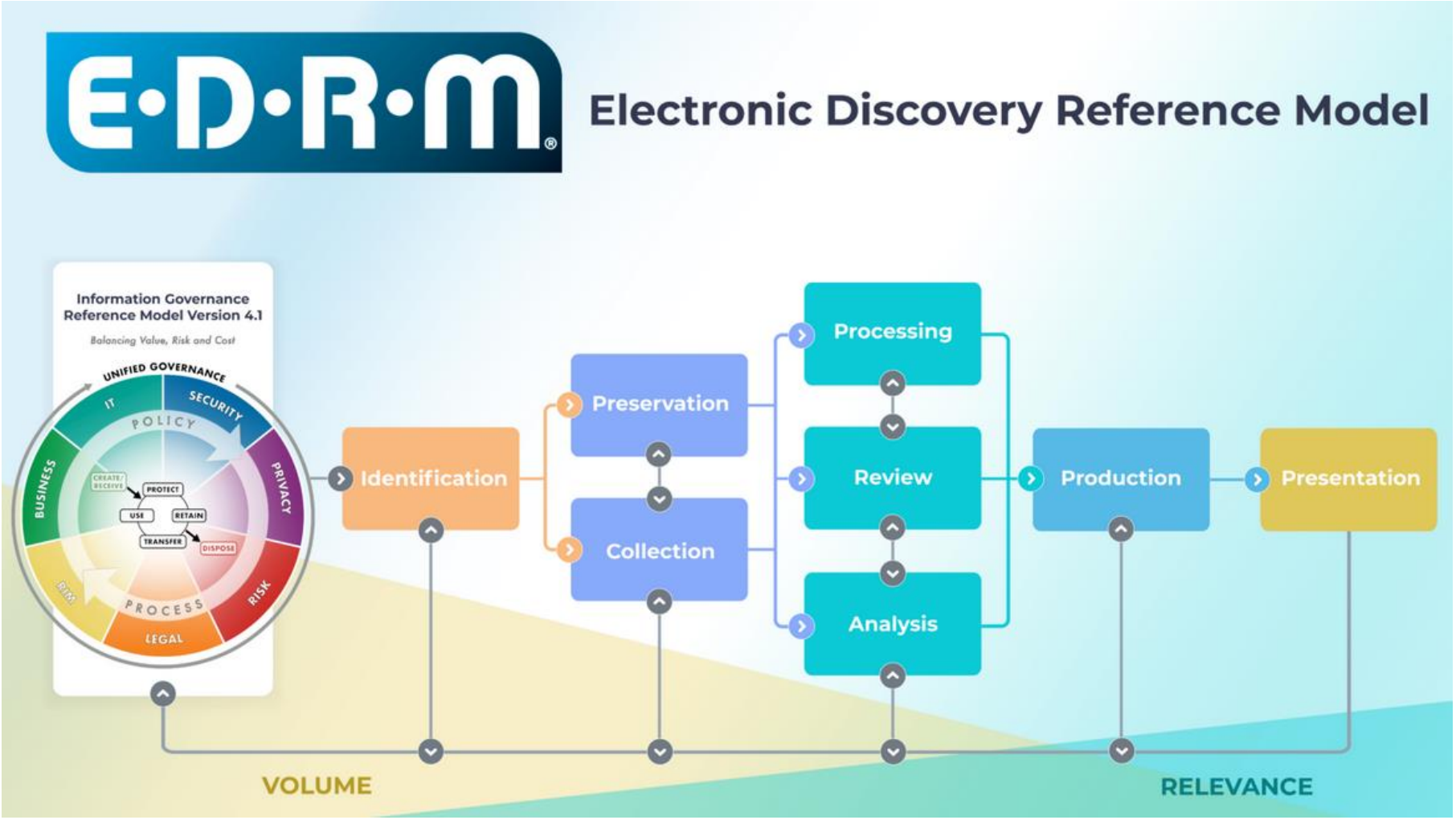
- **Not** because you're expected to solve crimes or catch malicious users
- Recognition of how data can be recovered when **layers** of technology fail or are no longer available
- **Capturing evidence** from places that are not always immediately visible
- Ensuring that actions taken on files **don't make irreversible changes** to essential characteristics (e.g. timestamps)
- Attending to the **order of volatility** – some types of data change much more quickly and often than others
- Learning about wide array of **tools and techniques** already available to deal with born-digital materials
- Established practices for **documenting** what we do, so others will know what we might have changed
- Considerable **overlap** between **technical knowledge** required to do digital forensics and ad hoc acquisition of digital materials by libraries/archives

Practices Contributing to Digital Curation Goals

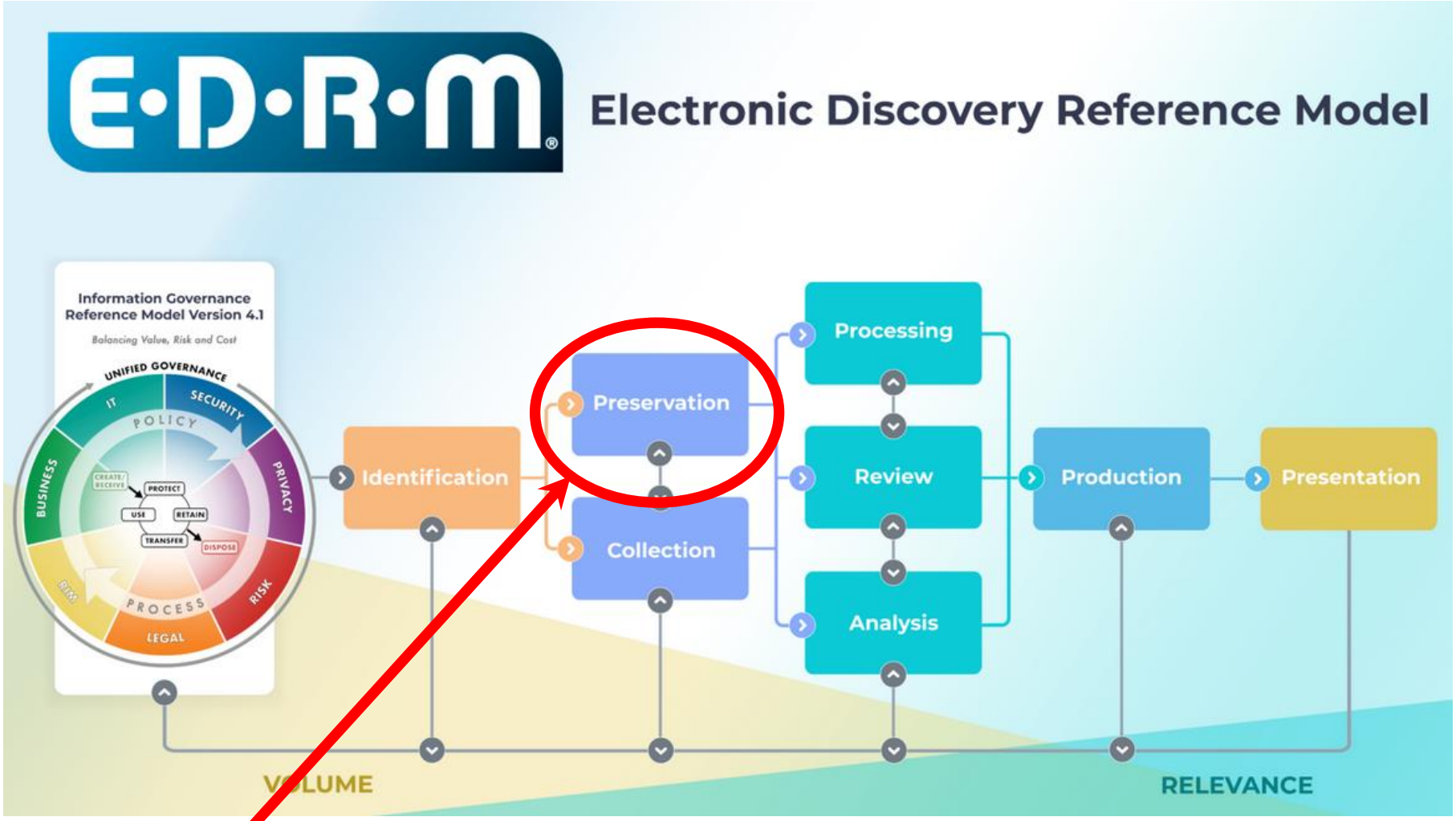
- Use of write blockers
- Generation of disk images
- Applying cryptographic hashes to bitstreams
- Capture of contextual metadata, including Digital Forensics XML (DFXML)
- Scanning bitstreams for potentially sensitive information

Need for Adaptation of Digital Forensics Tools and Tasks for DP

- Existing digital forensics tools provide valuable functionality, but they don’t always fit well into primary DP workflows
- For example, DP professionals are particularly concerned with:
  - structure and persistence of metadata
  - provisions for providing public access to data
  - support for older technologies (e.g. floppy disks, HFS)

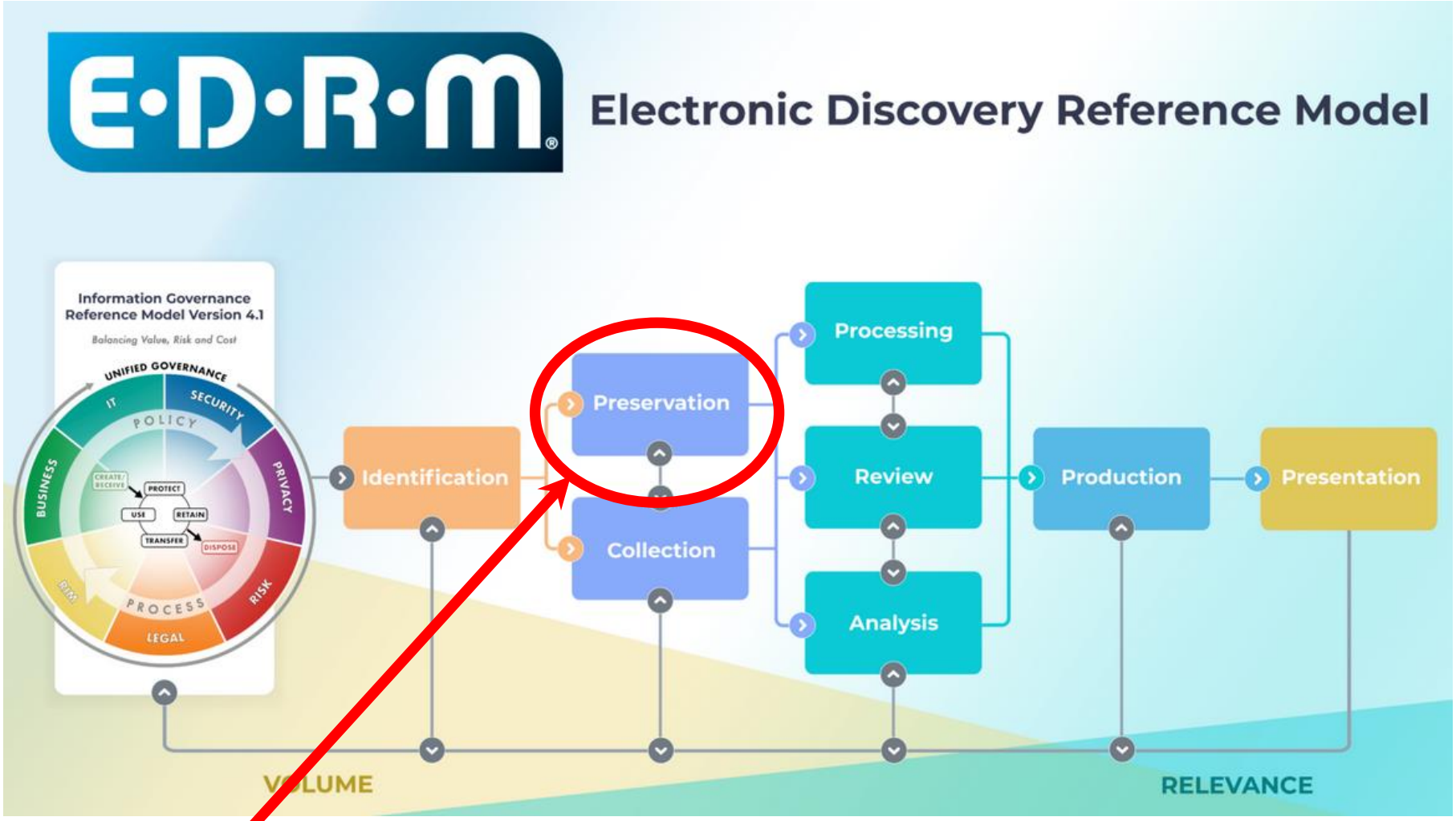


https://edrm.net/edrm-model/current/



"Ensuring the ESI [electronically stored information] is protected against inappropriate alteration or destruction"

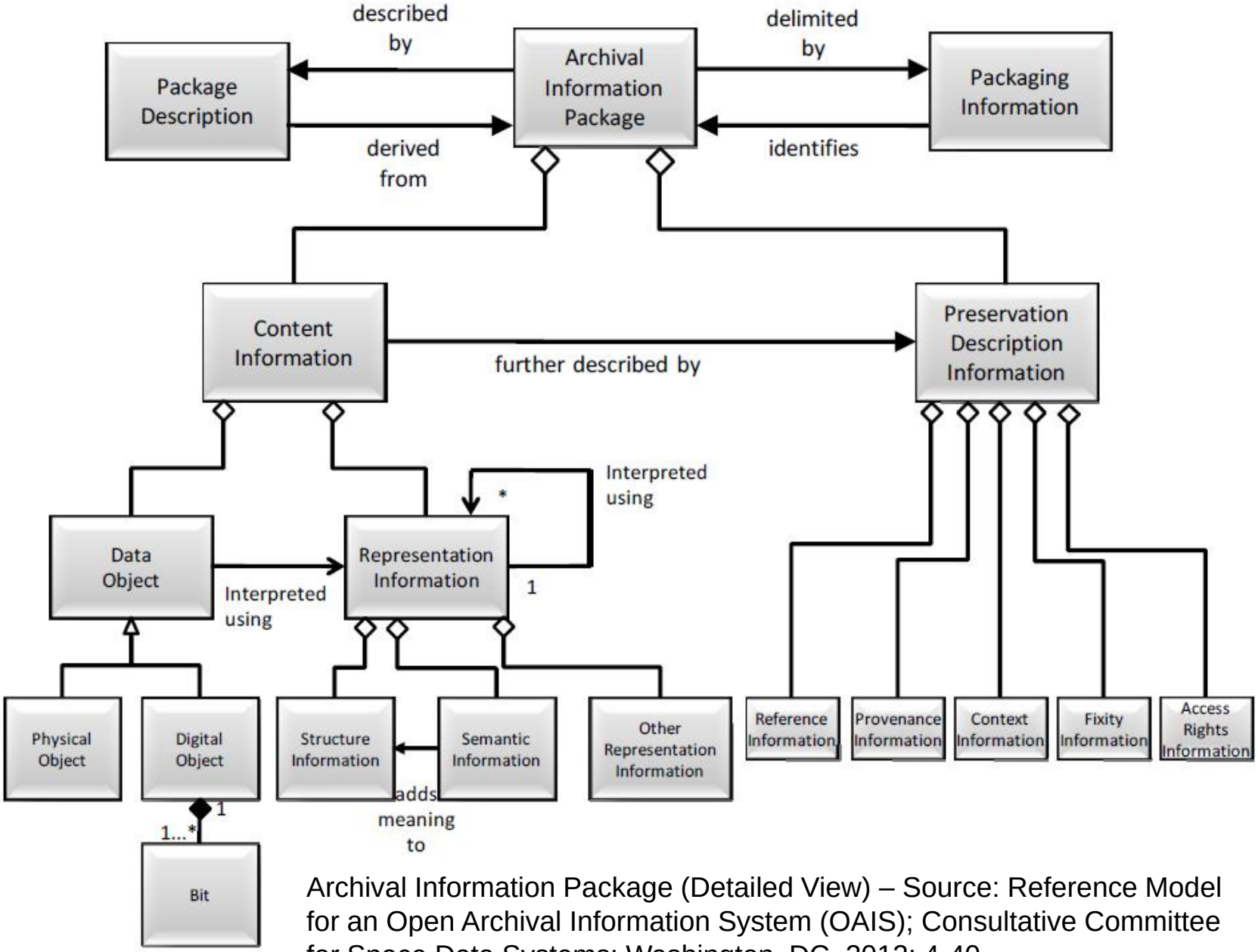
https://edrm.net/edrm-model/current/



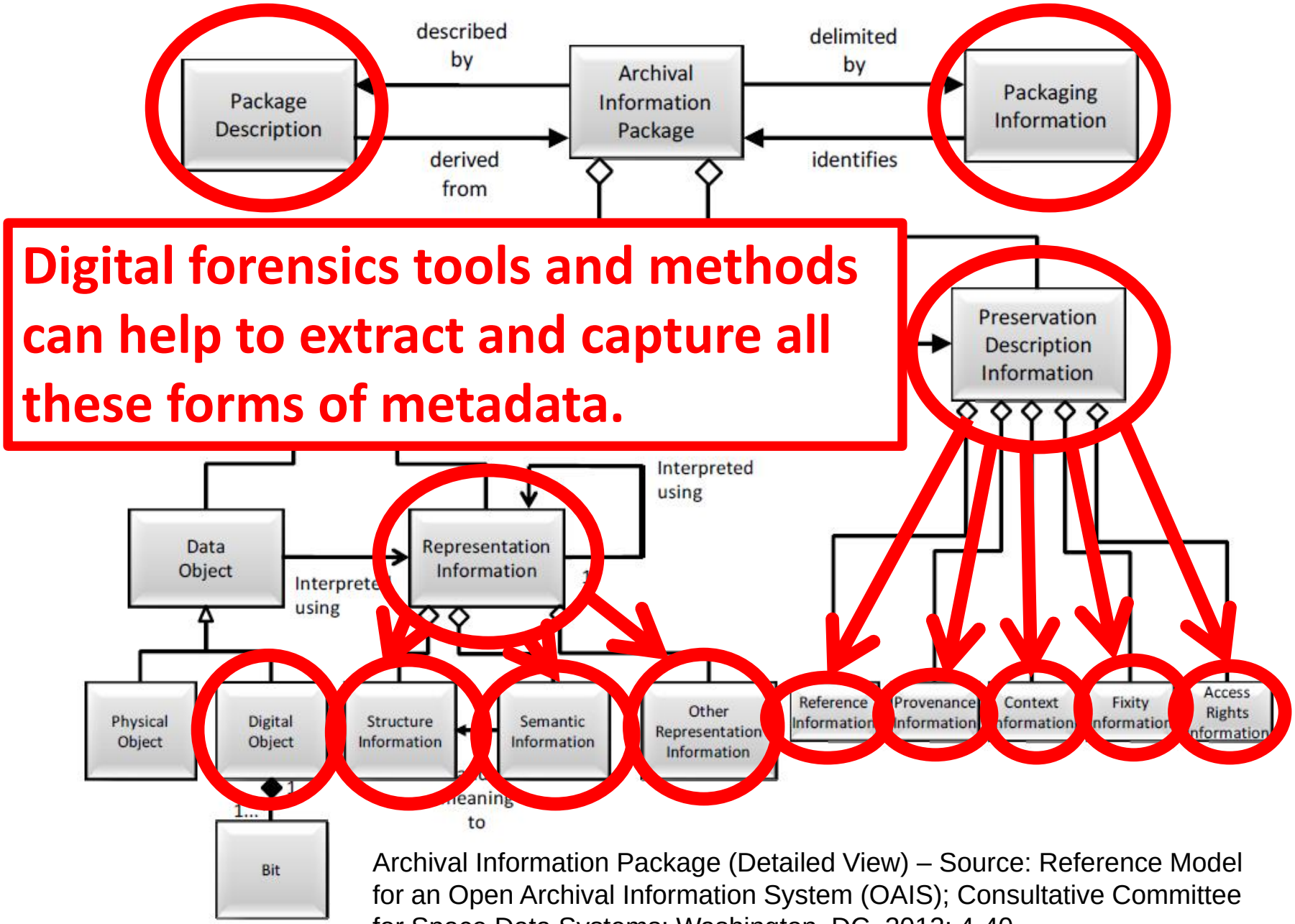
"Ensuring the ESI [electronically stored information] is protected against inappropriate alteration or destruction"

As a field DP has a much broader mandate than this.

https://edrm.net/edrm-model/current/



Archival Information Package (Detailed View) – Source: Reference Model for an Open Archival Information System (OAIS); Consultative Committee for Space Data Systems: Washington, DC, 2012; 4-40.



Archival Information Package (Detailed View) – Source: Reference Model for an Open Archival Information System (OAIS); Consultative Committee for Space Data Systems: Washington, DC, 2012; 4-40.

BitCurator

- Funded by Andrew W. Mellon Foundation
  - Phase 1: October 1, 2011 – September 30, 2013
  - Phase 2 – October 1, 2013 – September 30, 2014
- Partners: School of Information and Library Science (SILS) at UNC and Maryland Institute for Technology in the Humanities (MITH)

BitCurator Goals

- Develop a system for collecting professionals that incorporates the functionality of open-source digital forensics tools
- Address two fundamental needs not usually addressed by the digital forensics industry:
  - Incorporation into the workflow of archives/library ingest and collection management environments
  - Provision of public access to the data

BitCurator Environment*

- Bundles, integrates and extends functionality of open source software
- Can be run as:
  - Self-contained environment (based on Ubuntu Linux) running directly on a computer (download installation ISO)
  - Using “bootstrapping” installation scripts to turn any Ubuntu Linux machine into a BitCurator Environment
  - Self-contained Linux environment in a virtual machine using e.g. Virtual Box or VMWare
  - As individual components run directly in your own Linux environment or (whenever possible) Windows environment

*To read about and download the environment, see: http://wiki.bitcurator.net/

BitCurator

Installation, Configuration, and Usage Guide for Release(s): 4.x.x

https://github.com/BitCurator/bitcurator-distro/wiki/BitCurator-Quick-Start-Guide

Jammy-4.2.0 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Applications Sep 28 17:21

Home

Shared Folders and Media

BitCurator

Right Ctrl

Jammy-4.2.0 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Applications Sep 28 17:23

- Favorites
- Accessories
- Additional Tools
- Documentation and Help
- Forensics and Reporting
- Graphics
- Imaging and Recovery
- Internet
- Office
- Packaging and Transfer
- Programming
- Sound & Video
- System Tools
- Utilities

- Brasero
- cdrdao (command line)
- Clonezilla (command line)
- dcfldd (command line)
- dd (command line)
- ddrescue (command line)
- dumpfloppy (command line)
- ewfacquire (command line)
- Guymager

BitCurator

Right Ctrl

BitCurator Consortium

- Continuing home for hosting, stewardship and support of BitCurator tools and associated user engagement
- Administrative home: Educopia Institute
- Funding based on membership dues
- Software and documentation are free and open source, but membership provides benefits (e.g. support, training, development priority)

BitCurator
CONSORTIUM

About | Forum | Projects | Resources | News & Events | Get Involved

BitCurator Support Q&A

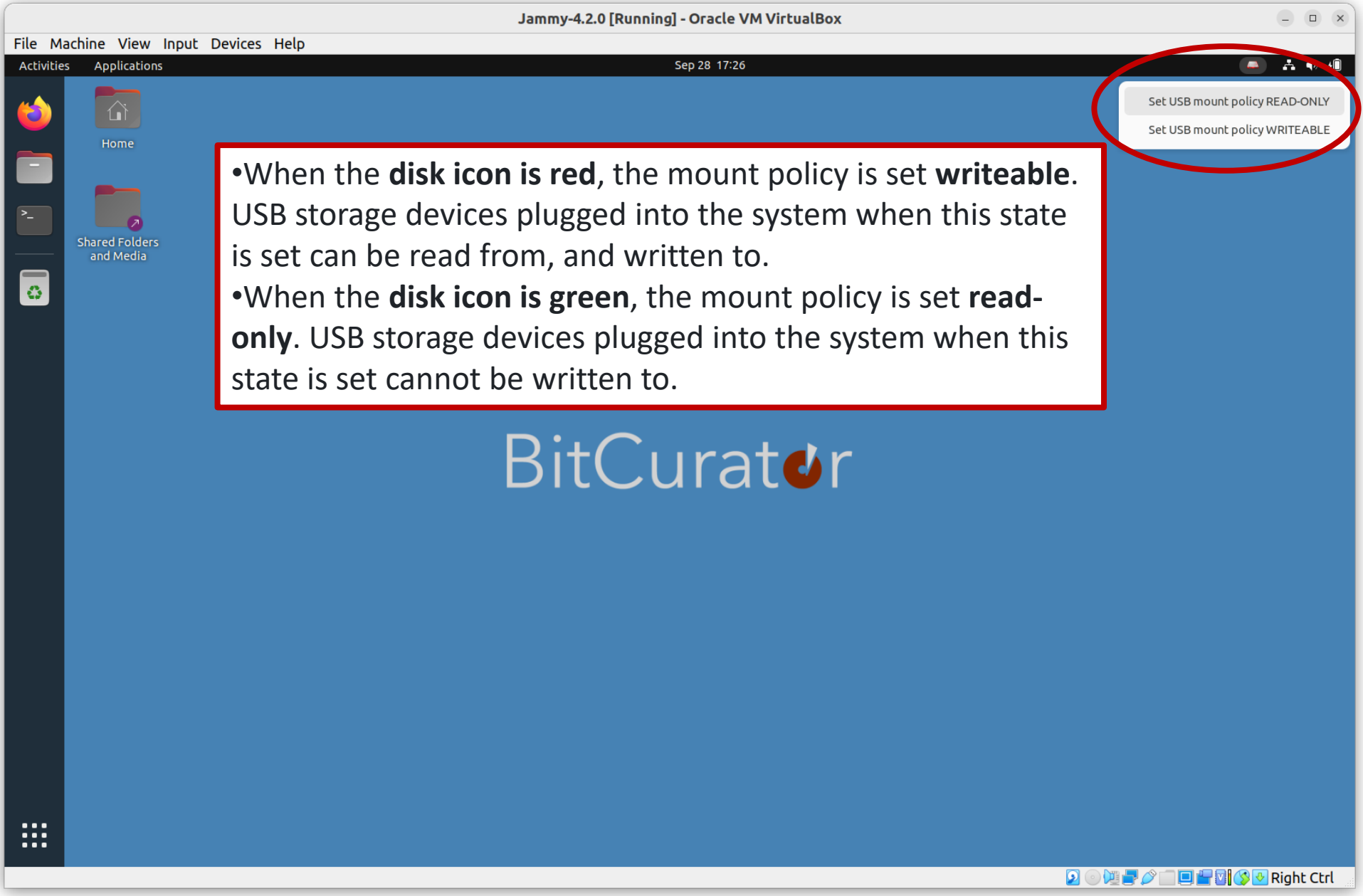
Looking for support for the BitCurator environment? Start here!

READ MORE

The BitCurator Consortium (BCC) is an independent, community-led membership association. The purpose of the BitCurator Consortium is to build a community of organizations that support practitioners responsible for the curation of born-digital materials, especially through the application of free and open-source tools.

Our organizational vision is to address the articulated needs of the BCC community—training, collaboration, research, software development, documentation, integration, code—while advocating for the expansion of digital forensics practice worldwide.

https://bitcuratorconsortium.org/


Jammy-4.2.0 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Applications
Sep 28 17:26
Set USB mount policy READ-ONLY
Set USB mount policy WRITEABLE
Home
Shared Folders and Media
•When the **disk icon is red**, the mount policy is set **writeable**. USB storage devices plugged into the system when this state is set can be read from, and written to.
•When the **disk icon is green**, the mount policy is set **read-only**. USB storage devices plugged into the system when this state is set cannot be written to.
BitCurator
Right Ctrl

Creating a Disk Image in Guymager

Jammy-4.2.1 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Applications Guymager Oct 22 02:08

Home

Devices Misc Help

Rescan

Serial nr.	Bad sectors	Progr
AA010220140505361941		
VB1b9dad02-edc1fb9c		

Size
Sector size
Image file
Info file
Current speed
Started
Hash calculation
Source verification
Image verification
Overall speed (all acquisi

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 2 EiB

Batch number 1

Item number 1

Contact name BitCurator

Description A Sample Image

Notes AA010220140505361941

Destination

Image directory ... /home/bcadmin/

Image filename (without extension) SampleData

Info filename (without extension) SampleData

Hash calculation / verification

☑ Calculate MD5 ☐ Calculate SHA-1 ☐ Calculate SHA-256

☐ Re-read source after acquisition for verification (takes twice as long)

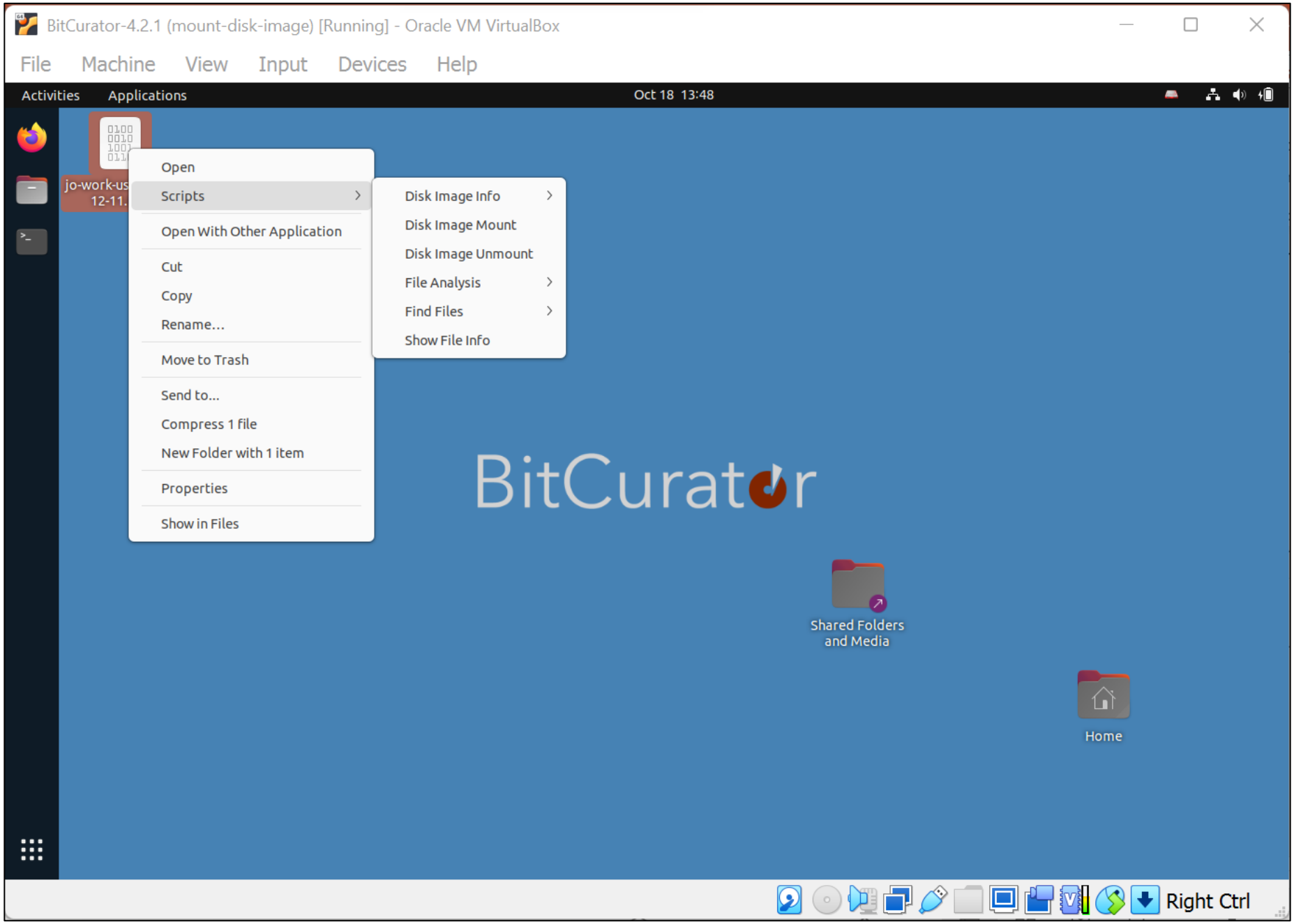
☑ Verify image after acquisition (takes twice as long)

Cancel Duplicate image... Start

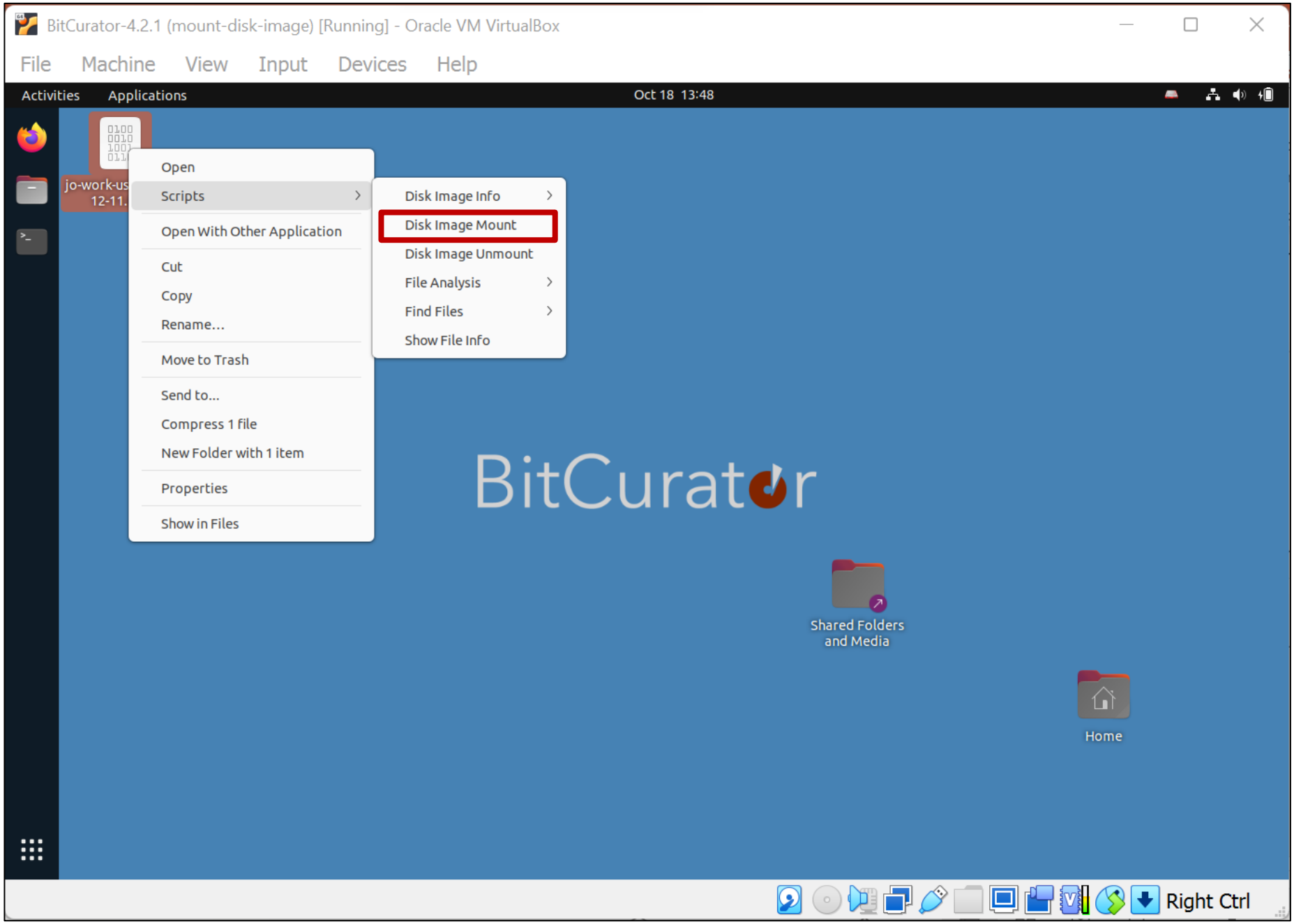
Shared Folders and Media

Right Ctrl

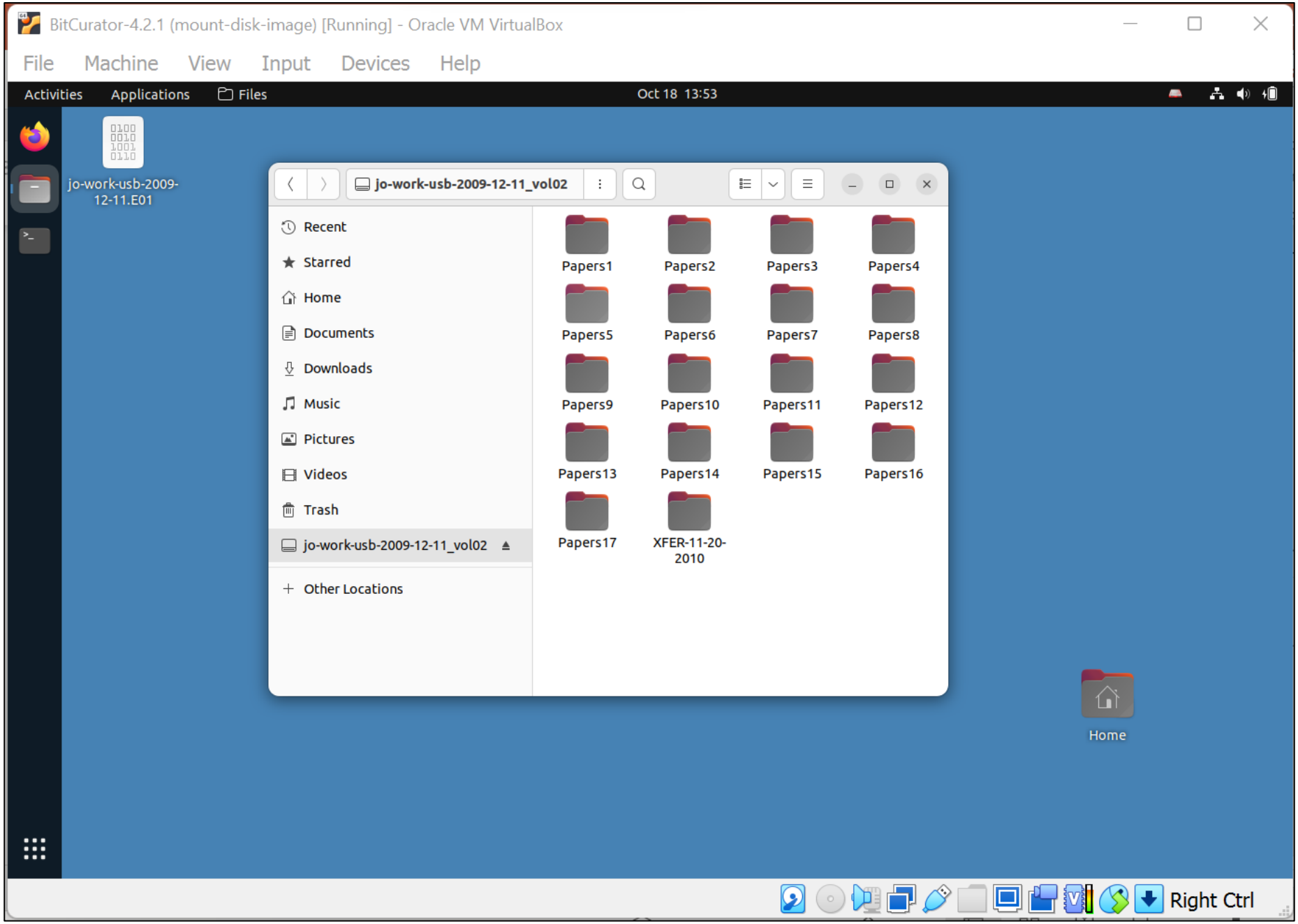
Mounting a Disk Image to Browse the Contents



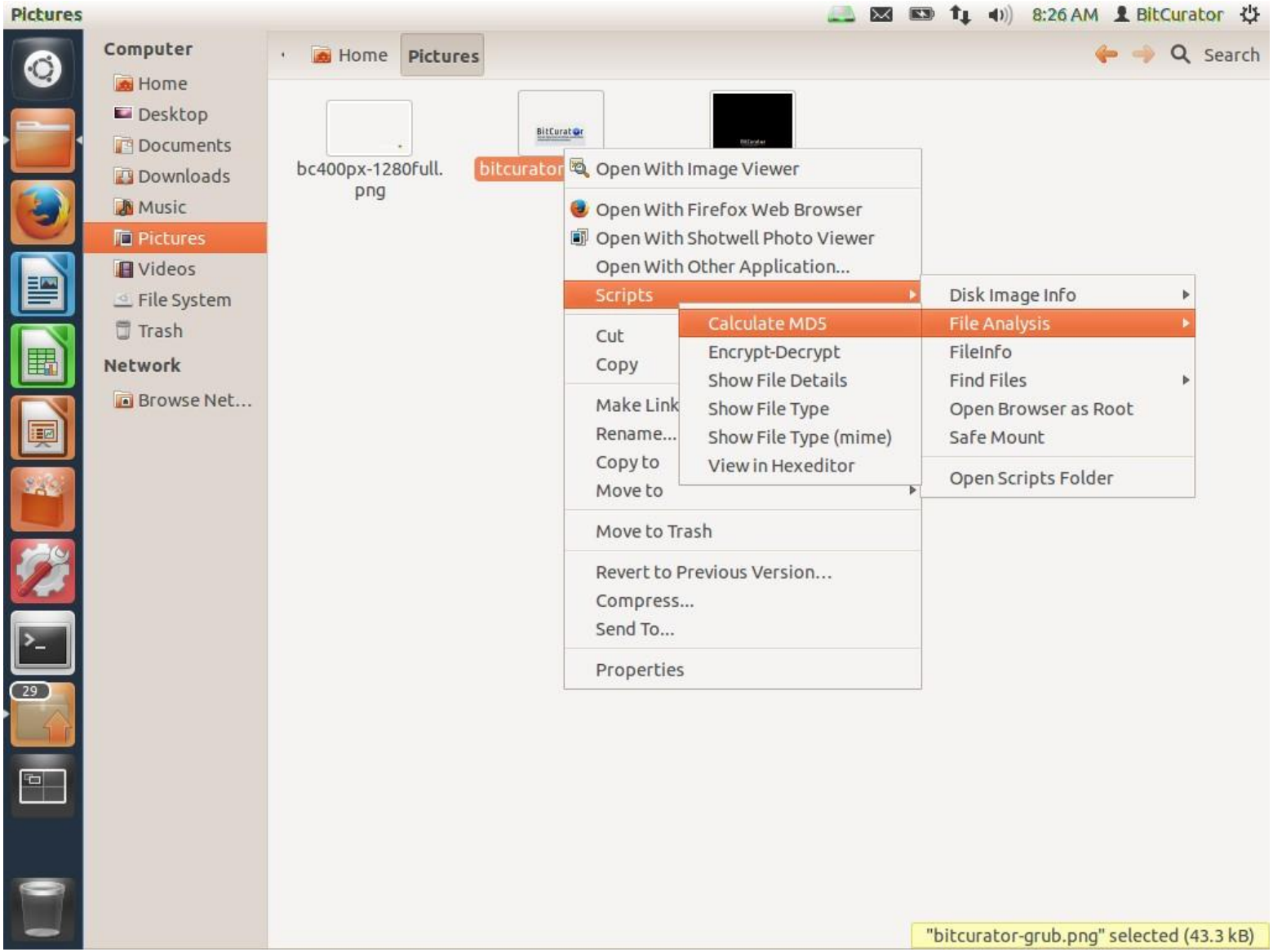
Mounting a Disk Image to Browse the Contents



Mounting a Disk Image to Browse the Contents



In BitCurator environment: Right Click on File or Directory and Calculate MD5



Calculate MD5 (Files and Directories)

8:31 AM BitCurator

Computer
- Home
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- File System
- Trash

Network
- Browse Net...

Home Pictures Search

bc400px-1280full.png

bitcurator-grub.png

bitcurator-grub-new.png

Calculate MD5 (Files and Directories)

Please choose the way you want the MD5 hash to be presented:
(1 file(s) selected)

Handling
(•) Display on screen
( ) Save to file (the selected filename + .md5 extenion)

Cancel OK

"bitcurator-grub.png" selected (43.3 kB)

Calculate MD5 (Files and Directories)

8:30 AM BitCurator

Computer
- Home
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- File System
- Trash

Network
- Browse Net...

Home Pictures Search

bc400px-1280full.png

bitcurator-grub.png

bitcurator-grub-new.png

Calculate MD5 (Files and Directories)

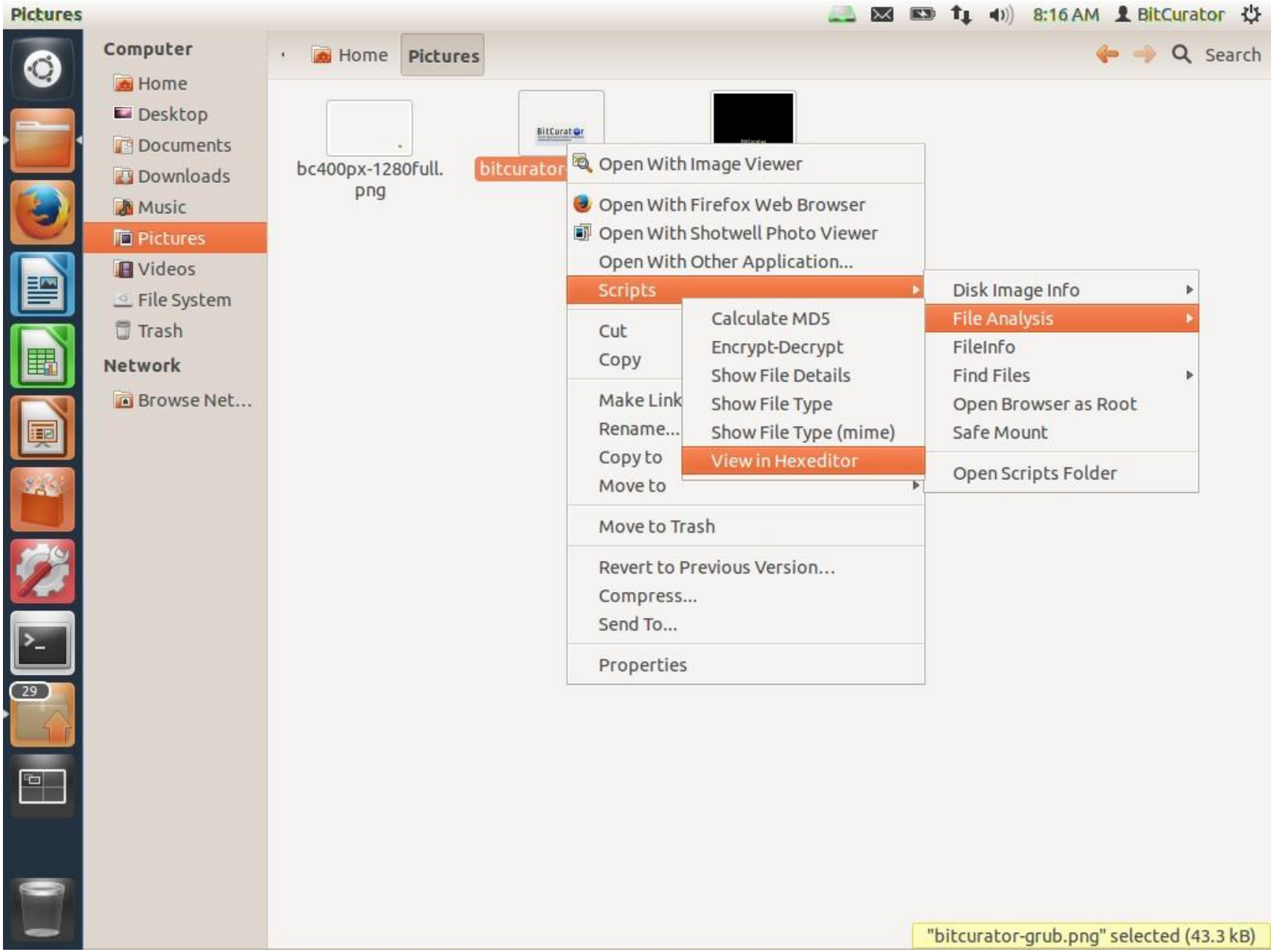
The MD5 hash of the selected file:

```
ceb2622125be1231b0fc9babee27942d  /home/bcadmin/Pictures/bitcurator-grub.png
```

Cancel OK

"bitcurator-grub.png" selected (43.3 kB)

In the BitCurator environment:



BitCurator-Basic-720px-2016.png - GHex

```
00000000 89 50 4E 47 0D 0A 1A 0A 00 00 00 0D 49 48 44 52 .PNG........IHDR
00000010 00 00 0C D4 00 00 02 18 08 06 00 00 00 1F 71 50 ..............qP
00000020 39 00 00 00 06 62 4B 47 44 00 FF 00 FF 00 FF A0 9....bKGD.......
00000030 BD A7 93 00 00 00 09 70 48 59 73 00 00 0B 13 00 .......pHYs.....
00000040 00 0B 13 01 00 9A 9C 18 00 00 00 07 74 49 4D 45 ............tIME
00000050 07 E0 0A 0B 00 2B 29 73 E3 4B 87 00 00 00 1D 69 .....+)s.K.....i
00000060 54 58 74 43 6F 6D 6D 65 6E 74 00 00 00 00 00 43 TXtComment.....C
00000070 72 65 61 74 65 64 20 77 69 74 68 20 47 49 4D 50 reated with GIMP
00000080 64 2E 65 07 00 00 20 00 49 44 41 54 78 DA EC DD d.e... .IDATx...
00000090 79 B8 65 69 59 DF FD EF A9 1E 69 66 BA 19 44 99 y.eiY.....if..D.
000000A0 07 47 40 50 C0 17 C4 29 12 0D 38 60 44 45 11 51 .G@P...)..8`DE.Q
000000B0 63 D4 28 1A 34 BE E6 D2 98 38 26 8E E0 04 51 8C c.(.4....8&...Q.
000000C0 13 06 51 79 9D 20 A8 20 A0 41 14 64 92 19 94 A6 ..Qy. . .A.d....
000000D0 69 81 16 BA 1B 19 9A 9E 9B EE AA F7 8F 7D FA A2 i...........}..
000000E0 2C BB CA AA AE 7D CE BD 87 CF E7 BA EE 6B 9F AA ,....}.......k..
000000F0 86 53 BF BD 9E F5 AC BD D6 5E FB DE CF 4E 00 00 .S.......^...N..
00000100 00 00 EB ED D4 EA EC EA B6 D5 39 D5 AD AB 5B 54 ..........9...[T
00000110 B7 DC AD C3 7F BE 69 75 66 75 93 1B 78 3C AD 3A ......iufu..x<.:
00000120 E5 28 75 A8 BA AE BA 76 F7 F1 F0 9F AF AE AE 38 .(u....v.......8
00000130 4A 5D 52 7D A0 FA E0 EE E3 E1 F5 DE EA A2 EA 9A J]R}...........
00000140 E9 0D 08 00 00 00 00 00 00 00 B0 6D 76 A6 03 00 ...........mv...
00000150 00 00 00 1C C5 2D AA 8F AE EE B8 FB 78 F8 CF 77 .....-......x..w
00000160 68 D1 3C 73 DB 16 8D 32 EB FC 1E C7 07 AA 0B 5B h.<s...2.......[
00000170 34 D7 5C B8 5B EF AA DE 59 BD 63 B7 2E 9E 0E 09 4.\.[...Y.c.....
00000180 00 00 00 00 00 00 00 B0 49 D6 F9 C3 26 00 00 00 ........I...&...
00000190 C0 7A 3B BD BA 47 75 B7 A3 D4 AD A6 03 AE 90 AB .z;..Gu.........
000001A0 FA 48 83 CD B9 BB F5 D6 DD C7 F3 5B AC 96 03 00 .H.........[....
000001B0 00 00 00 00 00 00 C0 71 D2 50 03 00 00 00 EC B5 .......q.P......
000001C0 DB 55 1F 5F 7D EC 11 75 B7 EA 94 E9 70 1B E0 C3 .U._}..u....p...
000001D0 2D 9A 6A DE 5A BD B1 7A C3 6E FD 5D 75 CD 74 38 -.j.Z..z.n.]u.t8
000001E0 00 00 00 00 00 00 00 80 55 A4 A1 06 00 00 00 58 ........U......X
```

Signed 8 bit:	-119	Signed 32 bit:	1196314761	Hexadecimal:	89
Unsigned 8 bit:	137	Unsigned 32 bit:	1196314761	Octal:	211
Signed 16 bit:	20617	Signed 64 bit:	1196314761	Binary:	10001001
Unsigned 16 bit:	20617	Unsigned 64 bit:	1196314761	Stream Length:	8 − +
Float 32 bit:	5.281654e+04	Float 64 bit:	5.292398e-260		

☑ Show little endian decoding ☐ Show unsigned and float as hexadecimal

Offset: 0x0

Scanning Disk Images and Directories for Potentially Sensitive Information with Bulk Reviewer

BitCurator-4.3.0 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Applications bulk-reviewer Oct 25 21:02

Home

Bulk Reviewer

Application Edit

Bulk Reviewer New Load

Identify, review, and remove sensitive files

Scan new directory or disk image | Load from JSON file

Shared Folders and Media

Right Ctrl

BitCurator-4.3.0 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Applications bulk-reviewer Oct 25 21:19

Home

Bulk Reviewer

Application Edit

Bulk Reviewer New Load

New session

Type

Disk Image

Name

A Sample Disk Image

Browse /home/bcadmin/SampleData.E01

Options

Use existing bulk_extractor reports

Choose directory None selected.

Social Security Number identification mode

Medium: xxx-xx-xxxx with dashes (ssn_mode=1)

Regular expressions file

Choose file None selected.

Stoplist directory

Choose directory None selected.

Include full EXIF metadata in results

Show Applications

Shared Folders and Media

Right Ctrl

BitCurator-4.3.0 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Applications bulk-reviewer Oct 25 22:00

Home

Bulk Reviewer

Application Edit

Bulk Reviewer New Load

Session: A Sample Disk Image

Source: /home/bcadmin/SampleData.E01

Save | Export files | Download CSV

+ Show file selector

Features (28170)

Showing results from: All files

Feature type: All (28170) | Show details | x Dismiss all | Undo all

Feature	Type	Note	Dismiss
utmp_carved/000/9401548800utmp	utmp_carved.txt	n/a	x Dismiss
utmp_carved/000/9401549184utmp	utmp_carved.txt	n/a	x Dismiss
utmp_carved/000/11544825856utmp	utmp_carved.txt	n/a	x Dismiss
utmp_carved/000/11544826240utmp	utmp_carved.txt	n/a	x Dismiss
214-69-9247	Social Security Number (USA)	n/a	x Dismiss
509-22-0354	Social Security Number (USA)	n/a	x Dismiss
509-23-1641	Social Security Number (USA)	n/a	x Dismiss
509-46-2701	Social Security Number (USA)	n/a	x Dismiss
476-32-6410	Social Security Number (USA)	n/a	x Dismiss
476-32-2012	Social Security Number (USA)	n/a	x Dismiss

Show Applications

Shared Folders and Media

Right Ctrl

XML Schema for Digital Forensics XML

43 commits | 1 branch | 9 releases | 1 contributor

branch: master | dfxml_schema / +

Document an XML validation step ...

ajnelson authored on Dec 4, 2014 latest commit 4c8aab566e

ref	Allow offline validation with local XSD cache	2 years ago
LICENSE.txt	Add public domain license text	2 years ago
README.md	Document an XML validation step	6 months ago
dfxml.xsd	Document an XML validation step	6 months ago

Code
Issues 8
Pull requests 0
Pulse
Graphs

HTTPS clone URL
https://github.com/c

You can clone with HTTPS or Subversion.

Clone in Desktop
Download ZIP

README.md

This is the schema repository for Digital Forensics XML, version 1.1.1.

If you intend to use the dfxml.xsd file as a DFXML document validator, note that you will also need to download two accompanying .xsd files under the "ref" directory. The easiest way to do this is by downloading the repository as a Git clone, or by downloading the zip archive from the Github page.

To report issues, questions, or feature requests, please either:

- File a Github issue, seeing first if it is already filed, here.
- Email the dfxml@nist.gov mailing list. If you wish to join the mailing list, send an email to dfxml-subscribe@nist.gov (no subject or message body is necessary), and a moderator will grant access.

https://github.com/dfxml-working-group/dfxml_schema

Operationalizing Original Order - Filesystem Metadata Output from fiwalk*


```
−<fileobject>
  −<parent_object>
      <inode>102</inode>
    </parent_object>
    <filename>Papers8/37638.BrannyPhyle.Joseph+Moore.pdf</filename>
    <partition>1</partition>
    <id>901</id>
    <name_type>r</name_type>
    <filesize>100857</filesize>
    <alloc>1</alloc>
    <used>1</used>
    <inode>6783</inode>
    <meta_type>1</meta_type>
    <mode>511</mode>
    <nlink>1</nlink>
    <uid>0</uid>
    <gid>0</gid>
    <mtime prec="2">2009-11-17T19:35:10</mtime>
    <atime prec="86400">2009-12-10T05:00:00</atime>
    <crtime prec="2">2009-12-10T19:34:11</crtime>
    <libmagic>PDF document, version 1.4 </libmagic>
  −<byte_runs>
      <byte_run file_offset="0" fs_offset="56621568" img_offset="56653824" len="100857"/>
    </byte_runs>
    <hashdigest type="md5">eb60256dabffa67cef7211bcba659815</hashdigest>
    <hashdigest type="sha1">e56f606877f10daf91dc0304ea120b35452bd36e</hashdigest>
  </fileobject>
```


*Developed by Simson Garfinkel

PREMIS (Preservation) Metadata Generated from Running BitCurator Tools – Recorded as PREMIS Events

premis.xml (~/Desktop/demo1/demo1reports/reports) - gedit


```xml
<?xml version="1.0" encoding="UTF-8"?>
<premis xmlns="info:lc/xmlns/premis-v2" version="2.0" xsi="http://www.w3c.org/2001/XMLSchema-instance">
  <object>
    <objectIdentifier>
      <objectIdentifierType>0d4e30d6-b8dc-11e3-a80f-080027f8dfea</objectIdentifierType>
      <objectIdentifierValue>/home/bcadmin/Desktop/terry-work-usb-2009-12-11.E01</objectIdentifierValue>
    </objectIdentifier>
  </object>
  <event>
    <eventIdentifier>
      <eventIdentifierType>0d4ea1ce-b8dc-11e3-a80f-080027f8dfea</eventIdentifierType>
      <eventIdentifierValue>E01/home/bcadmin/Desktop/terry-work-usb-2009-12-11.E01</
eventIdentifierValue>
    </eventIdentifier>
    <eventType>Capture</eventType>
    <eventDateTime>     Wed Jan 19 12</eventDateTime>
    <eventOutcomeInformation>
      <eventOutcome>E01</eventOutcome>
      <eventOutcomeDetail>Version:      20100226
, Image size: 512</eventOutcomeDetail>
    </eventOutcomeInformation>
  </event>
  <event>
    <eventIdentifier>
      <eventIdentifierType>19882604-b8dc-11e3-93f0-080027f8dfea</eventIdentifierType>
      <eventIdentifierValue>bulk_extractor -o /home/bcadmin/Desktop/demo1 /home/bcadmin/Desktop/terry-
work-usb-2009-12-11.E01</eventIdentifierValue>
    </eventIdentifier>
    <eventType>Feature Stream Analysis</eventType>
    <eventDateTime>2014-03-31T13:49:59Z</eventDateTime>
    <eventOutcomeInformation>
      <eventOutcome>Bulk Extractor Output</eventOutcome>
      <eventOutcomeDetail>version: 1.4.4</eventOutcomeDetail>
    </eventOutcomeInformation>
  </event>
</premis>
```

Provenance – DFXML Output

BitCurator-0.2.0 [Running]

Mozilla Firefox 8:08 PM BitCurator

file:///home/b...mpleimage.xml

file:///home/bcadmin/Desktop/SampleData/sampleimage.xml

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
-<dfxml version="1.0">
 -<metadata>
    <dc:type>Disk Image</dc:type>
  </metadata>
 -<creator version="1.0">
    <program>fiwalk</program>
    <version>4.0.2</version>
  -<build_environment>
     <compiler>GCC 4.6</compiler>
     <library name="afflib" version="3.7.1"/>
     <library name="libewf" version="20130303"/>
   </build_environment>
  -<execution_environment>
   -<command_line>
      fiwalk -f -X /home/bcadmin/Desktop/SampleData/sampleimage.xml /home/bcadmin/Desktop/SampleData
      /sampleimage.E01
     </command_line>
     <start_time>2013-03-12T00:08:28Z</start_time>
   </execution_environment>
  </creator>
 -<source>
    <image_filename>/home/bcadmin/Desktop/SampleData/sampleimage.E01</image_filename>
  </source>
  <!-- fs start: 0 -->
 -<volume offset="0">
    <partition_offset>0</partition_offset>
    <block_size>2048</block_size>
    <ftype>2048</ftype>
    <ftype_str>iso9660</ftype_str>
    <block_count>36839</block_count>
```

Batch and Case vs. Triage

- Batch and case model
  - Forensic Toolkit (FTK)
  - Autopsy
  - OpenText EnCase Forensic
- Triage
  - BitCurator
  - SIFT Workstation (SANS) – specialized OS much like BitCurator but tools focus on forensics tasks (e.g. RAM analysis)
  - Kroll Artifact Parser (KAPE) – based on model of quick results but focuses on security incident response rather than curation of materials

Challenges

- Incorporation into DP workflows, e.g. metadata conventions, connections to collection management systems
- Obsolete storage media and filesystems
- Dealing with large, internally complex data Files (including disk images)
- Provision of public access
- Defining and implementing ethical commitments

New Workflows

- Core digital curation functions involve numerous decisions based on various patterns – commonalities, differences, contextual relationships
- When patterns can be identified algorithmically, software can assist the process
- Compared to analog materials, functions are often more iterative and rely on data sources/streams shared across functions

OSSArcFlow

OSSArcFlow
MODELING AND SYNCHRONIZING OSS WORKFLOWS

Contact:
Katherine Skinner

Additional Documents:
OSSArcFlow proposal

Investigating, Synchronizing, and Modeling a Range of Archival Workflows for Born-Digital Content

Project Abstract

The Educopia Institute, in collaboration with the University of North Carolina at Chapel Hill School of Information and Library Science (UNC SILS), LYRASIS, and Artefactual, Inc., are investigating, synchronizing, and modeling a range of workflows to increase the capacity of libraries and archives to curate born digital content. These archival workflows will incorporate three leading open source software (OSS) platforms—BitCurator, Archivematica, and ArchivesSpace—and the project will be designed to generate findings that can be generalizable to settings that are using other platforms and applications.

This project will significantly impact curation practices by increasing our understanding of how institutions of different sizes and types may engage in OSS tool integration and workflow development. Our findings will be used to support a broad range of libraries and archives actively collecting and curating digital content. The knowledge gained by working with multiple institutions of different types and sizes will also broaden field-wide understanding of curation approaches and priorities, and how those impact the use of tools and capabilities in Archivematica, ArchivesSpace, and BitCurator. We expect the empirical findings about institutional needs, as well as formal workflow models, to contribute to digital curation research literature.

This project has been generously funded by the Institute of Museum and Library Services.

Project Outputs

Digital Dossiers (January 2018)

Ahead of the partner meeting on December 4-5, 2017, project partners created digital dossiers outlining the form, function, and future of digital curation at their home institutions.

1. Atlanta University Center, Robert W. Woodruff Library
2. District of Columbia Public Library
3. Duke University

https://educopia.org/research/ossarcflow

Artifacts:
Workflow Representations

Rep #1 Procedural Narrative

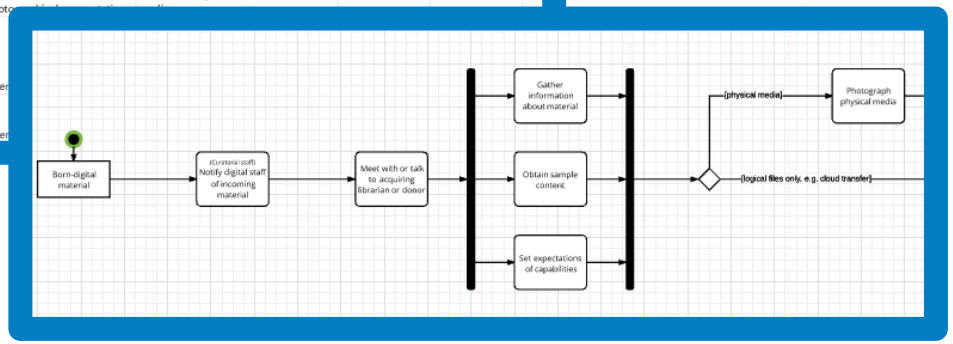
- **TRANSFER/ACCESSIONING**
  - Digital content is acquired:
    - [Digital Archivist or Archives Staff] IF on physical storage media THEN
      - [Archivist] takes picture of the media
      - IF filesystem is known THEN:
        - Triaged from physical media using either Guymager or Forensic Toolkit (FTK) Imager
      - IF filesystem is unknown THEN:
        - Triaged from physical media using Kryoflux
    - I
  - Data is m

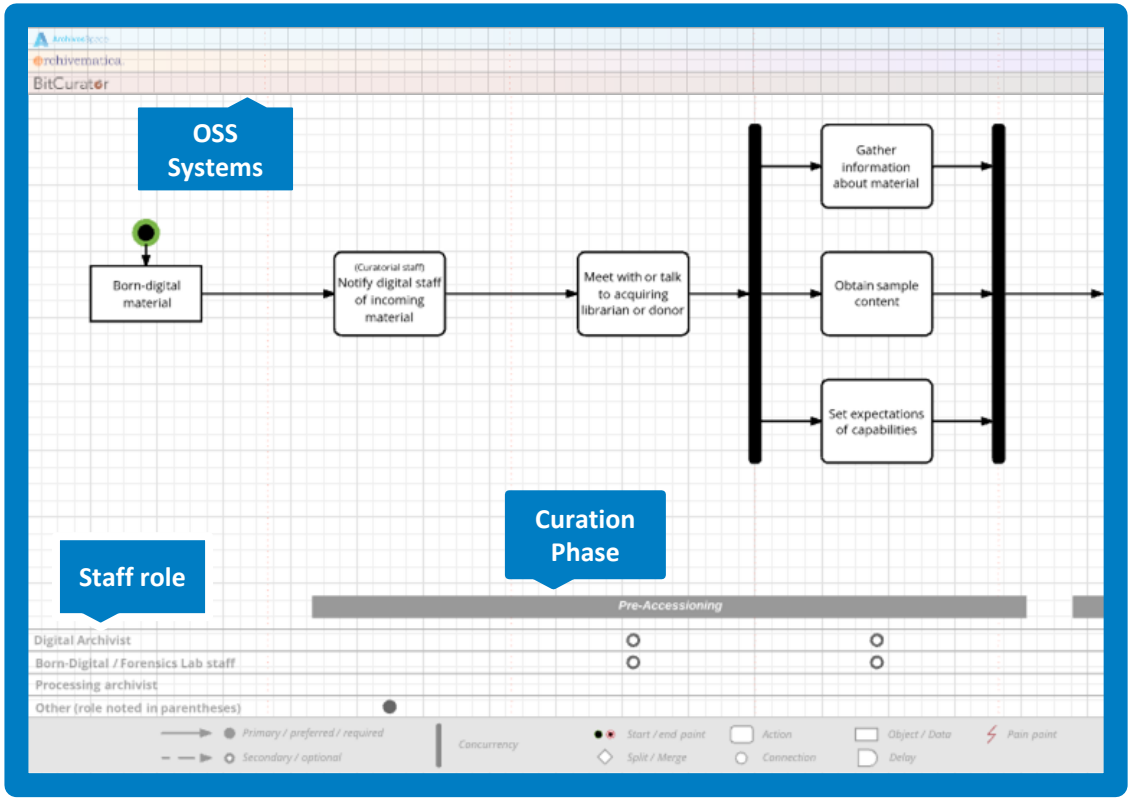
Rep #2 Tabular Steps

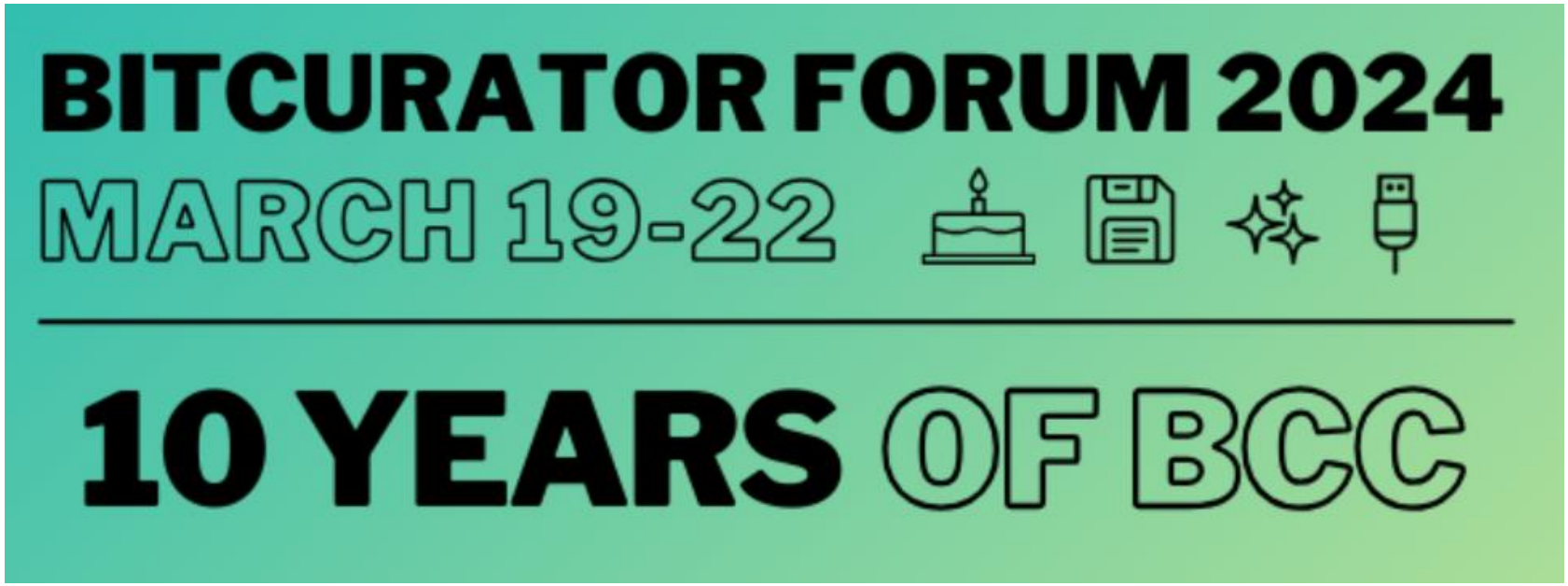
phase

	A	B	C	D	E
	phase	step	description	hardware	software
	pre-accessioning	high-level content analysis, donor negociation	[Michael and/or Curator] Appraisal meeting/donor negotiation using a survey instrument that gathers information about what the collection is and what it contains.		survey instrument (Google Sheet? Excel spreadsheet?)
	transfer/accessioning	photo	IF materials are on physical storage media, THEN [Digital Archivist or Archives Staff] takes a picture of the		
	transfer/accessioning	forer			
	transfer/accessioning	forer			

Rep #3: Visual Diagram




archivematica
BitCurator
OSS Systems
Born-digital material
(Curatorial staff)
Notify digital staff of incoming material
Meet with or talk to acquiring librarian or donor
Gather information about material
Obtain sample content
Set expectations of capabilities
Curation Phase
Staff role
Pre-Accessioning
Digital Archivist
Born-Digital / Forensics Lab staff
Processing archivist
Other (role noted in parentheses)
Primary / preferred / required
Secondary / optional
Concurrency
Start / end point
Split / Merge
Action
Connection
Object / Data
Delay
Pain point



https://bitcuratorconsortium.org/forum/