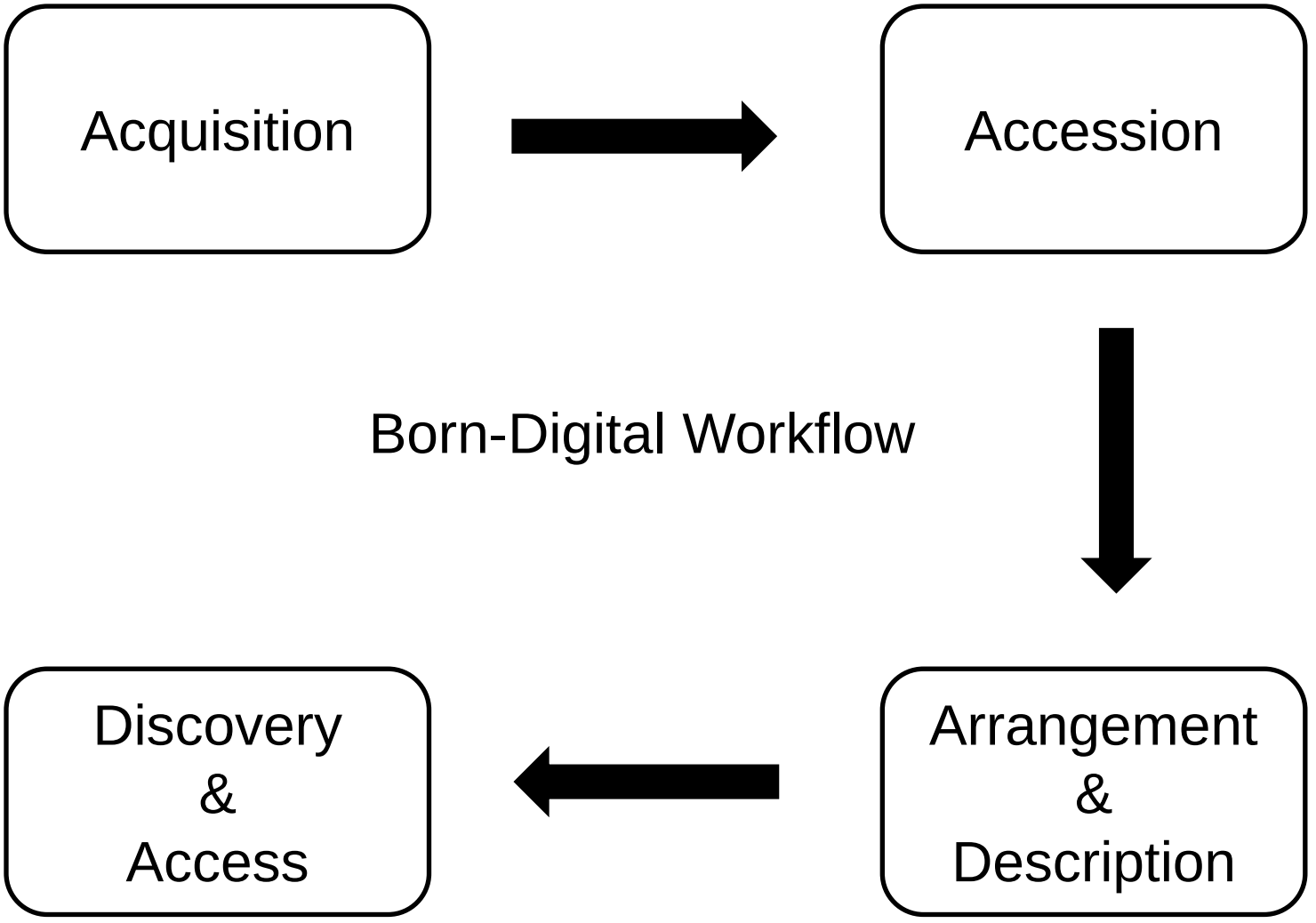
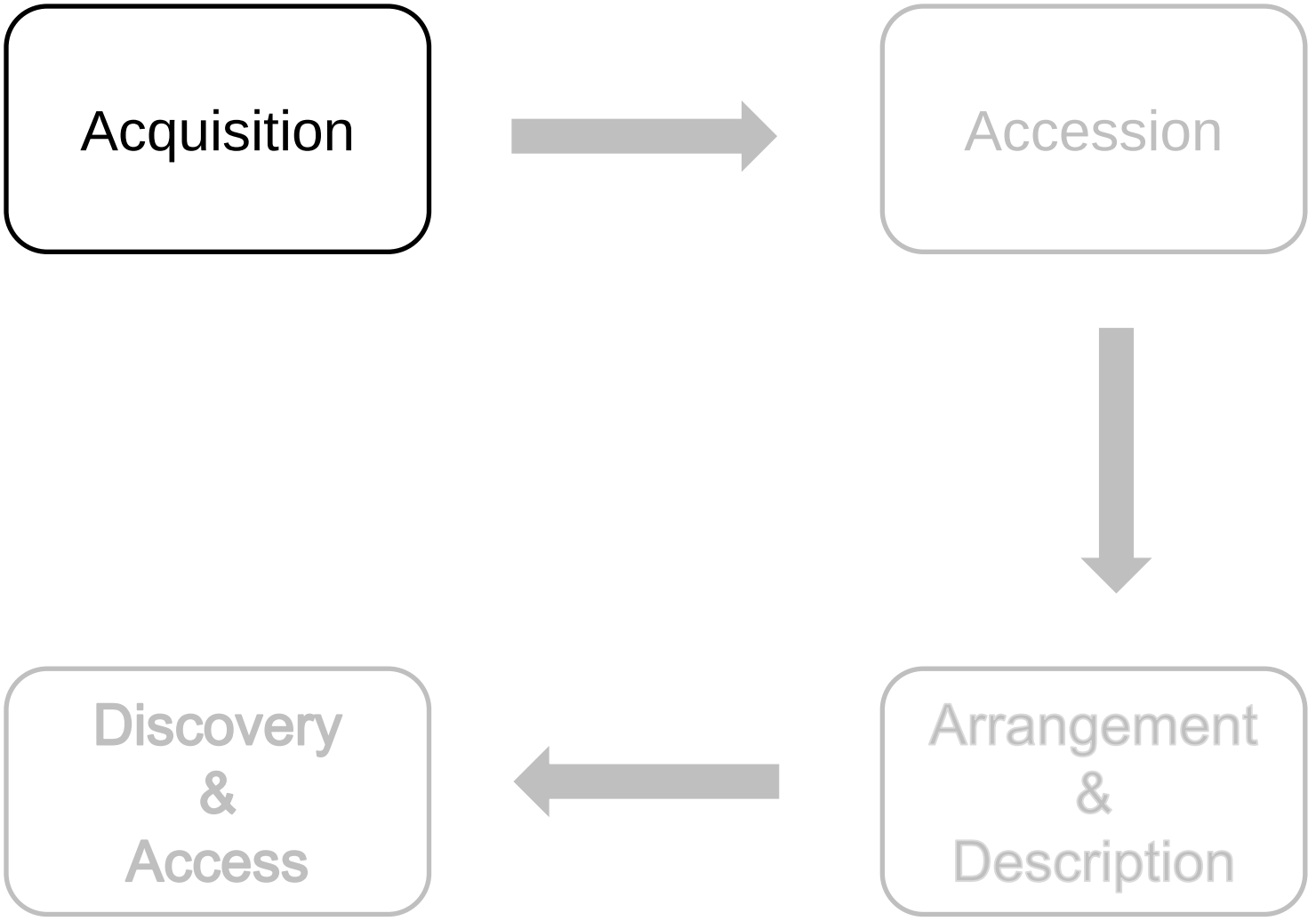
Getting Bits off Disks:

Using open source tools to stabilize and prepare born-digital materials for long-term preservation

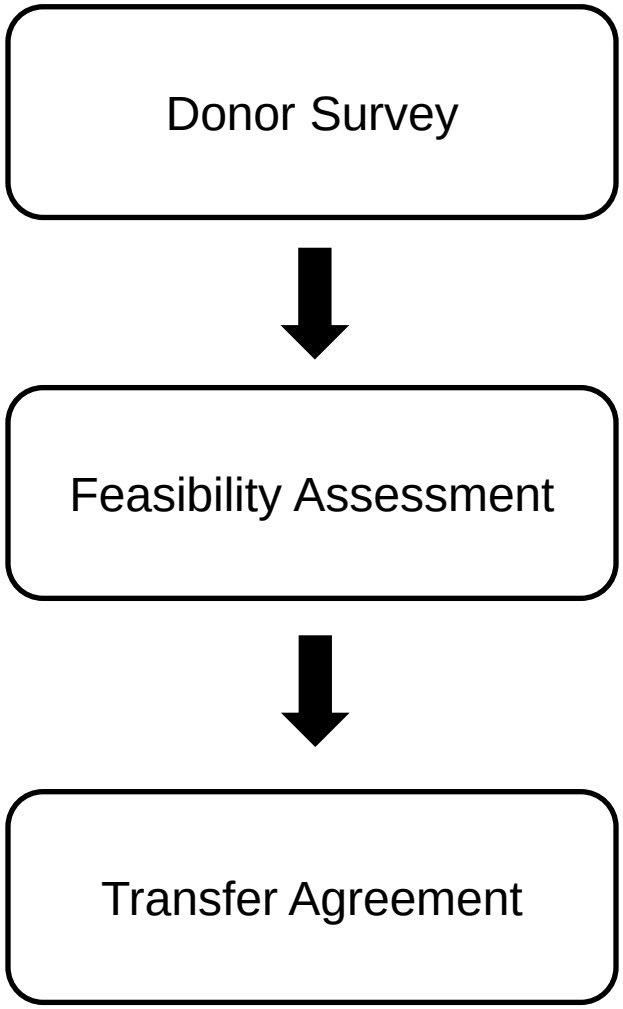
Sam Meister
University of Montana

Best Practices Exchange 2013 November 13, 2013


Acquisition
Accession
Born-Digital Workflow
Arrangement
&
Description
Discovery
&
Access


Acquisition
Accession
Arrangement
&
Description
Discovery
&
Access

Acquisition Process



Donor Survey

Creation

Context

Organization

Privacy & Security

Storage

Technical

Transfer Options

Current

Donor Survey
Sam Meister
08/16/12

Date: 09/07/12
Donor Name: Joe Crabtree – Art & Architecture Studio
Location: Donor's Residence – Missoula, MT
Accession #: 2012-061

Creation

Are you the only creator of your digital files? (e.g. does your collection contain email, documents, or other materials produced by others?)

Crabtree Leigland Arch. / Art & Architecture Studio / Crabtree Architecture

If not, who else is involved and what are their roles?

What are the earliest dates of file creation?

1996

What are the latest dates of file creation?

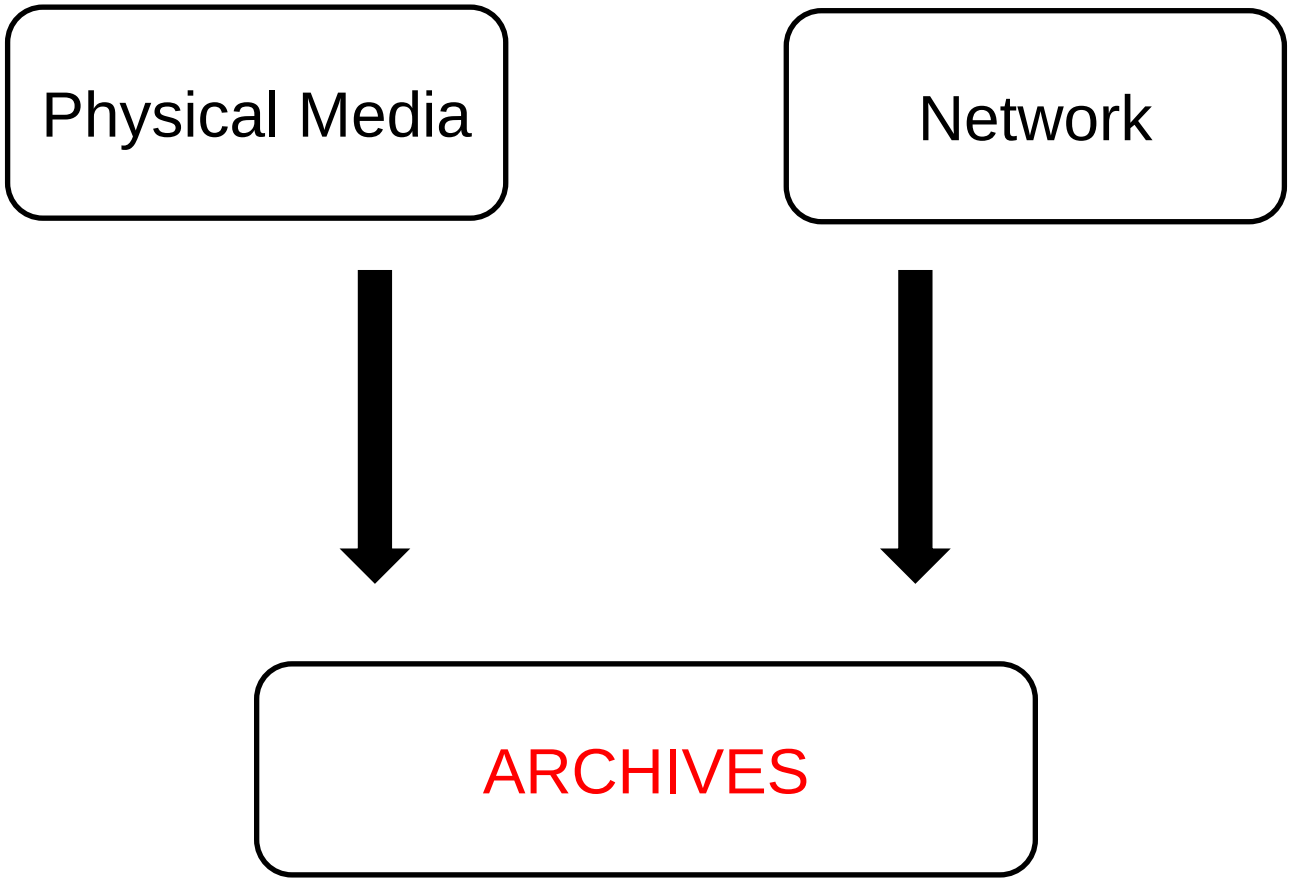
2011

Future

Feasibility Assessment

Do we have resources to feasibly acquire, preserve, and provide access to the digital materials?

Transfer



Current

Digital Materials Transfer

Date of Transfer:	2012-11-08
Transferred from:	Joe Crabtree
Transferred to:	Archives and Special Collections
Transferred by:	Sam Meister
Accession number:	2013-030

Information

Content

The digital materials include the following types of content:

Vector drawing files (AutoCAD)
3.5 inch floppy disks
CD Roms

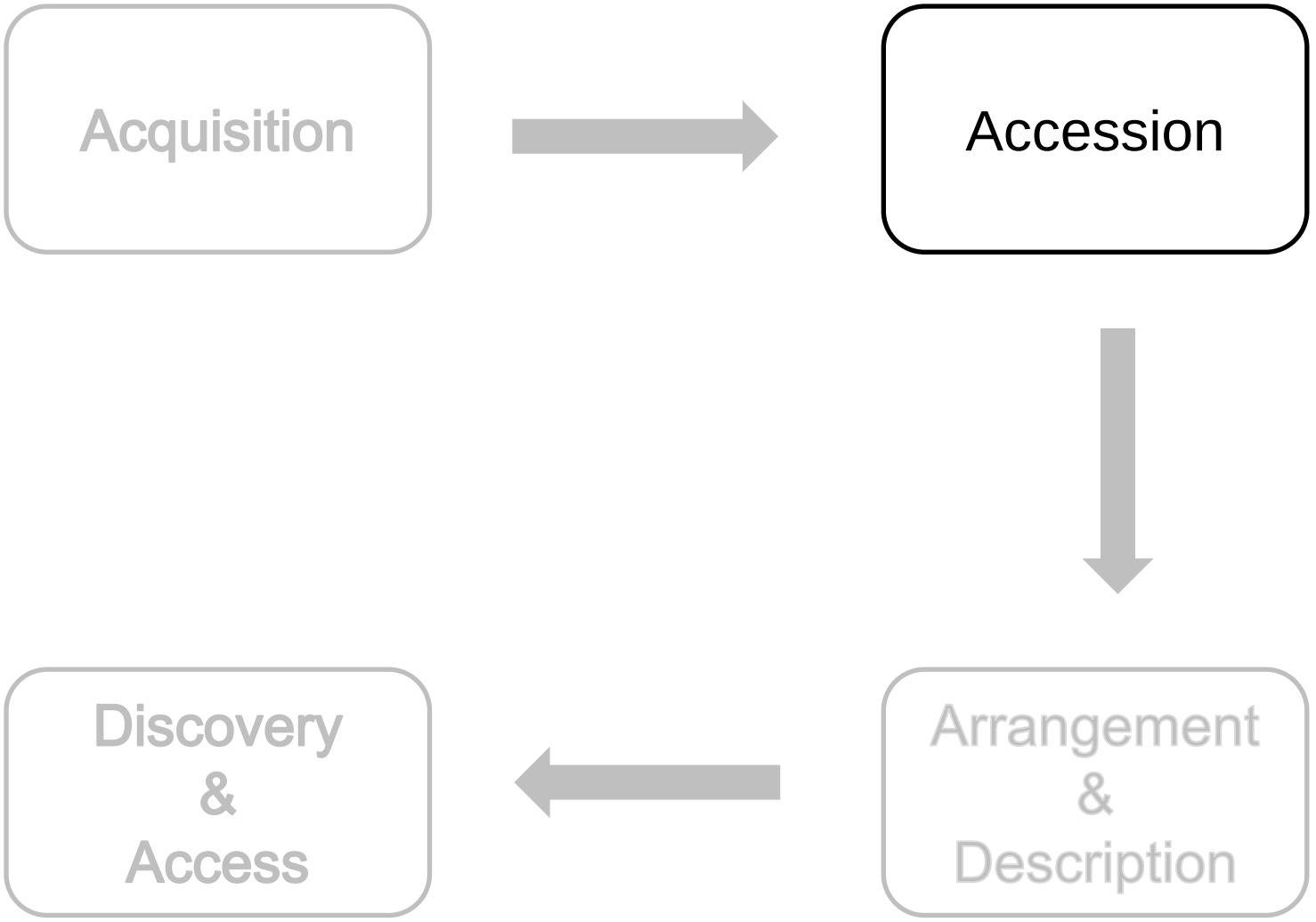
Technical

The technical environment used in the creation and management of digital materials includes:

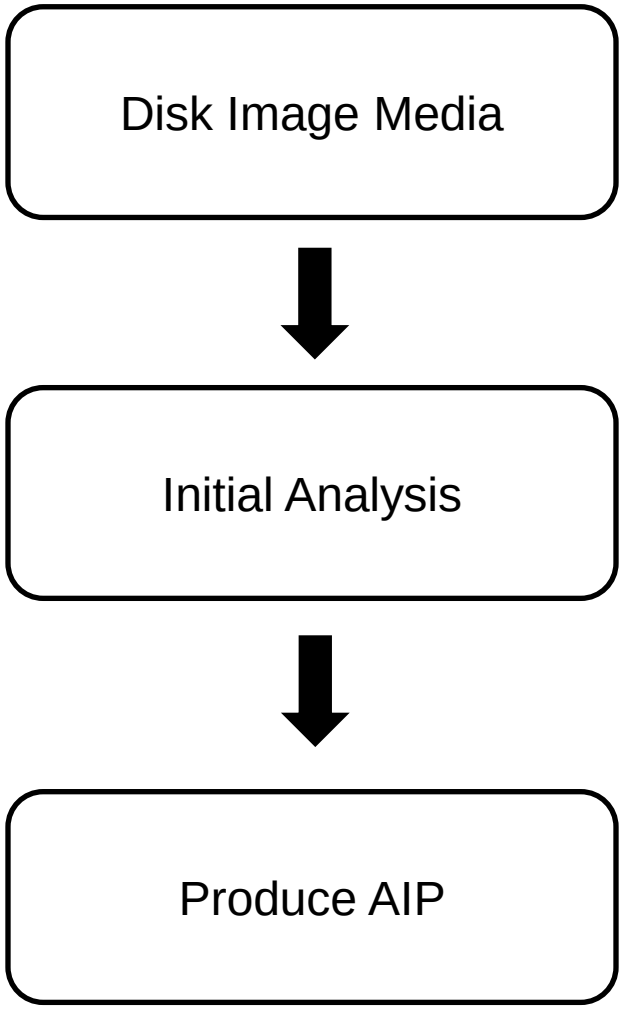
Describe details of technical infrastructure including hardware, software, operating systems, etc

Windows XP Desktop computer

Future


Acquisition
Accession
Arrangement
&
Description
Discovery
&
Access

Accession Process



Data Transfer

Hardware

3.5 Floppy Drive
5.25. Floppy Drive
Zip Drive
CD / DVD Drive
USB Write-Blocker
SATA / IDE Write-Blocker

Software

FTK Imager
Guymager
FC5205

Disk Imaging

“A single file or storage device containing the complete contents and structure representing a data storage medium or device, such as a hard drive, tape drive, floppy disk, CD/DVD/BD, or USB flash drive”

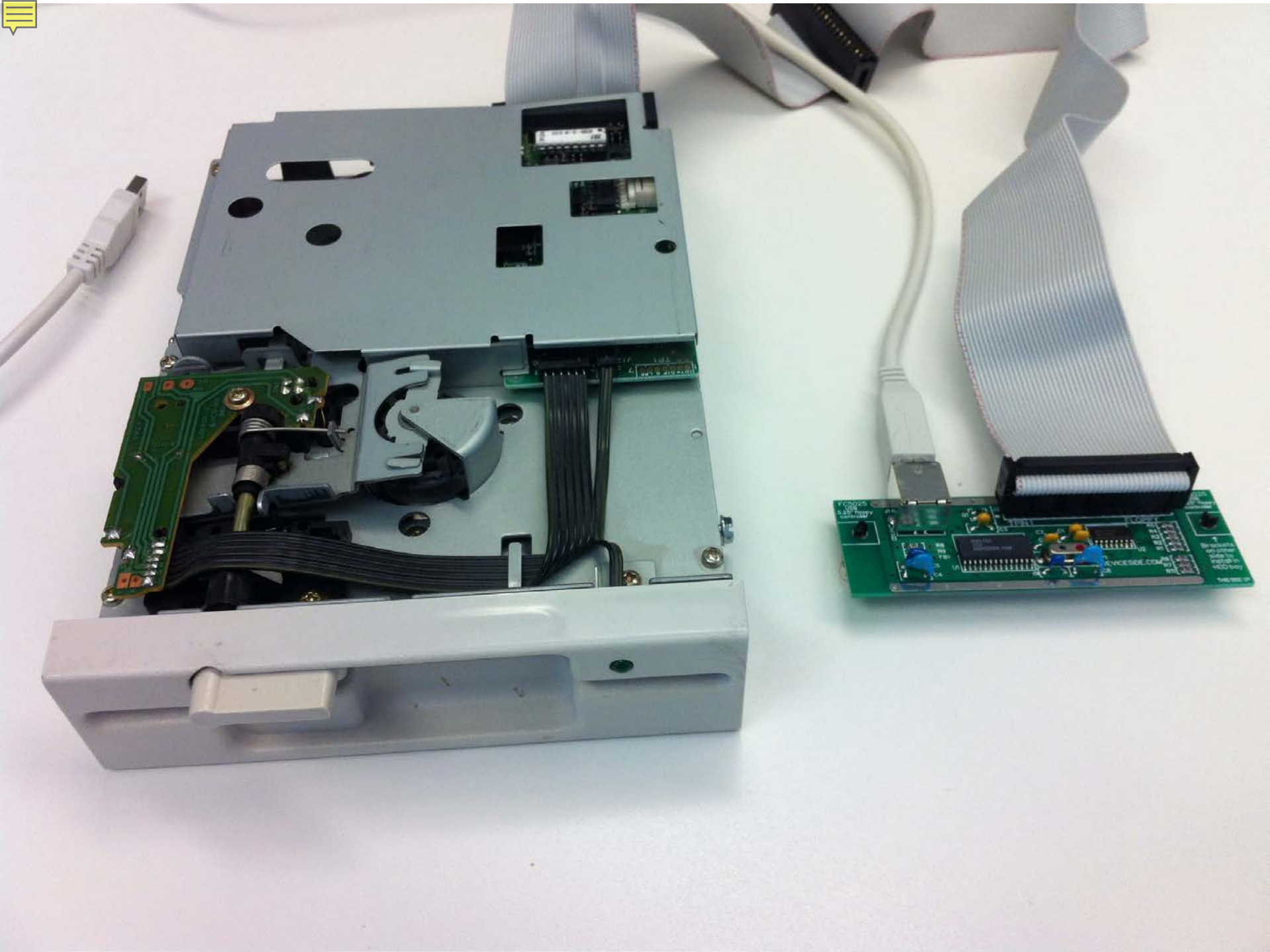
Disk Imaging

Born Digital Workstation 1.0

Disk Imaging

Born Digital Workstation 2.0

RESTRICTED
The University of Montana
1893
Mansfield Library
Missoula, MT 59812-9936
(406) 243-6866
THE UNIVERSITY OF MONTANA
3 3342 02186 1965

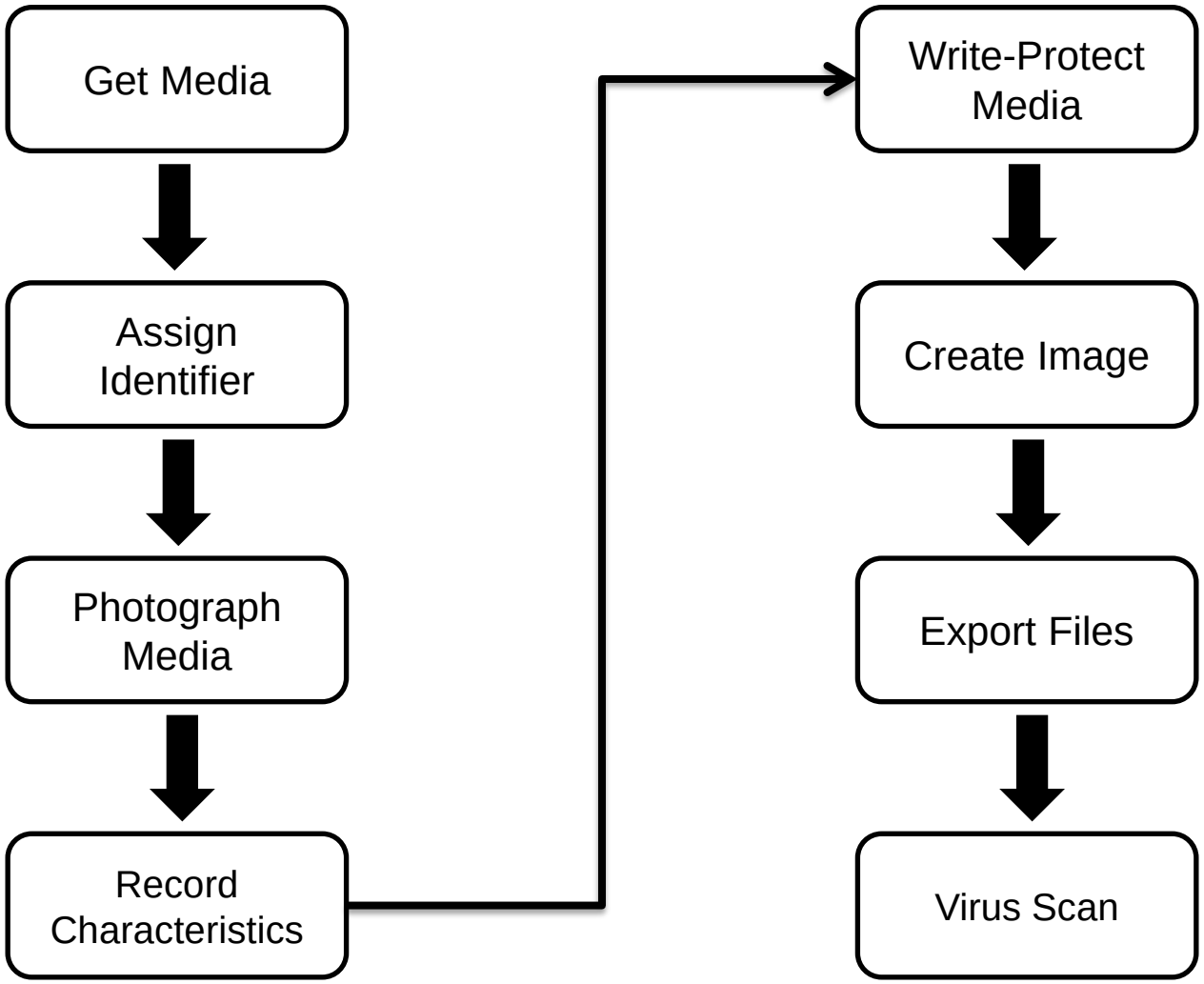


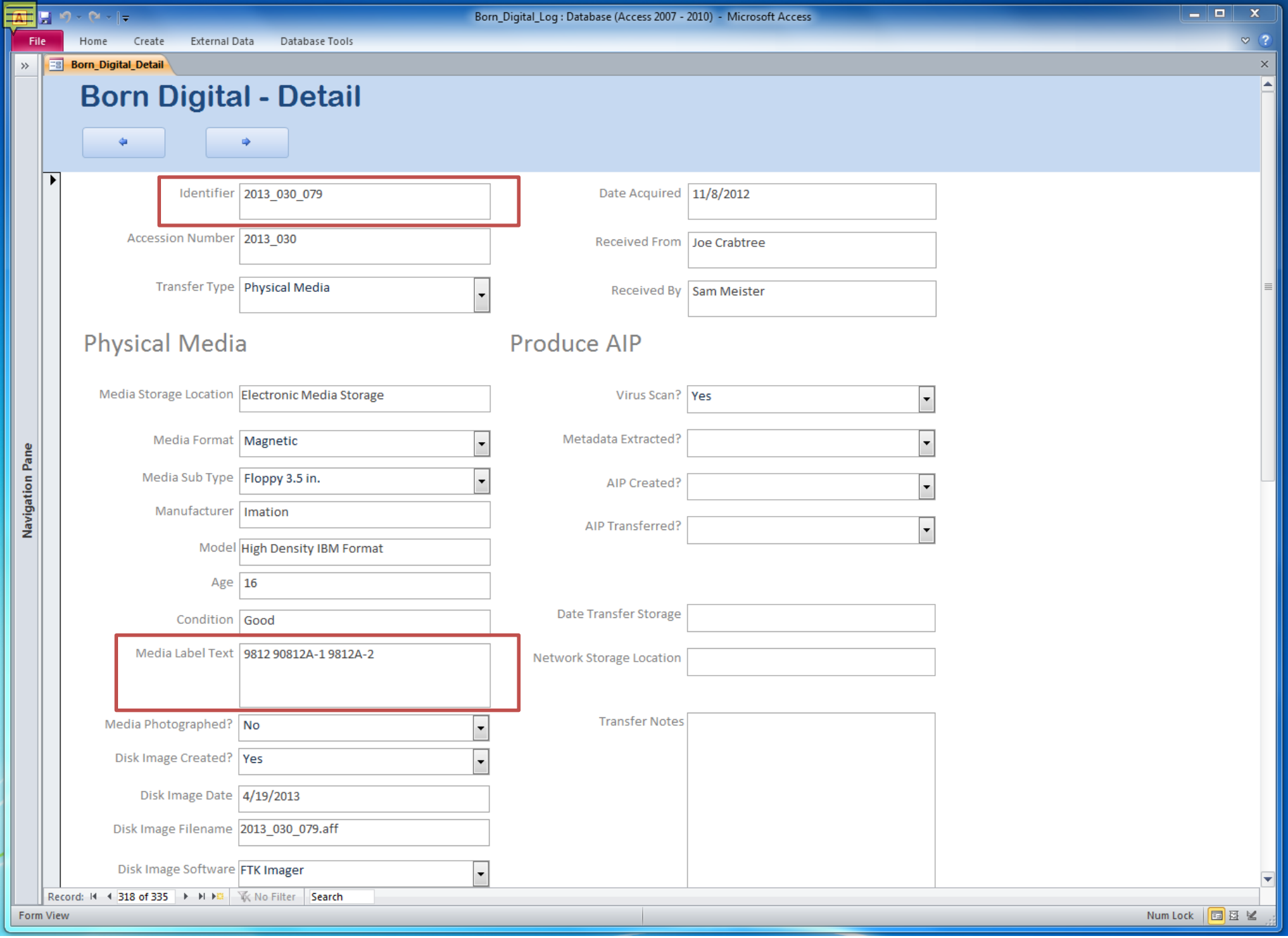
Assembled in MALAYSIA
W/O# 00000

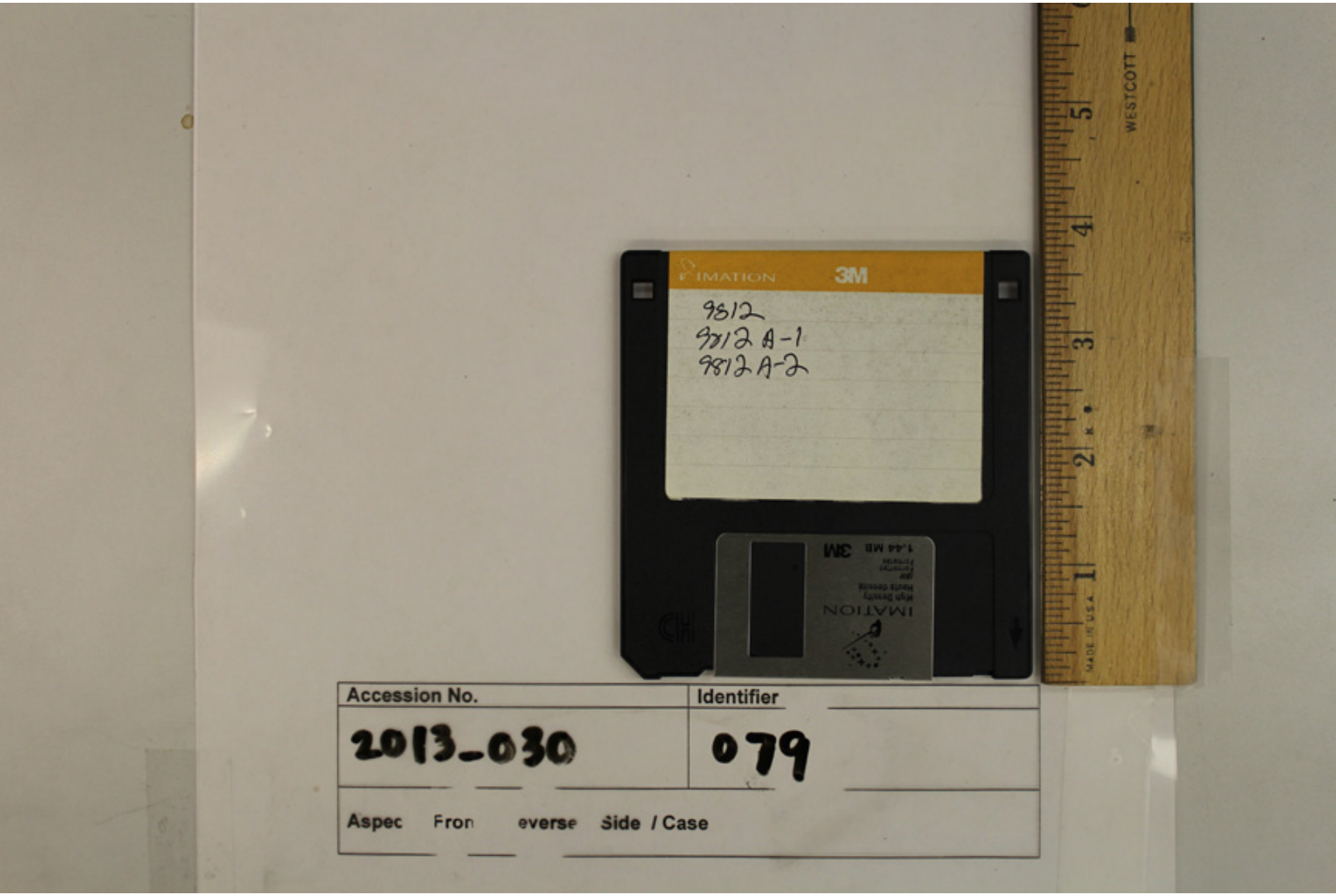
USB 2.0
Forensic USB Bridge
Power
Write Block
Activity
Device Detect
TABLEAU
DC IN
Host Detect
Menu
Enter
DC IN
FireWire — Host — USB 2.0

TABLEAU
eSATA Forensic Bridge
DC IN
800
800
400
SATA & IDE
Power
IDE Detect
SATA Detect
Host Detect
Write Block
Activity
USB 2.0
eSATA
DC OUT

Disk Imaging



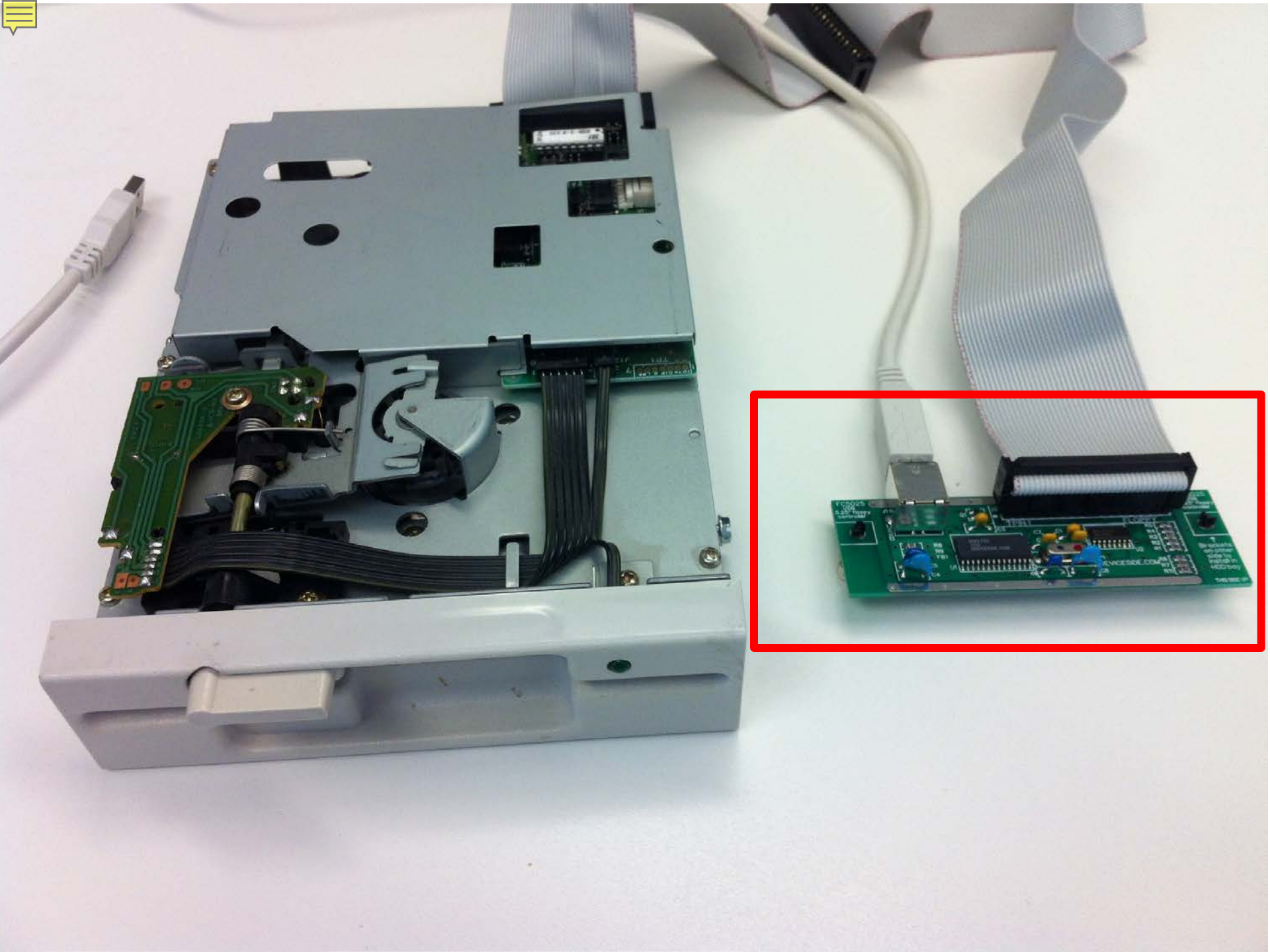

Born_Digital_Log : Database (Access 2007 - 2010) - Microsoft Access
File
Home
Create
External Data
Database Tools
Born_Digital_Detail
Born Digital - Detail
Identifier
2013_030_079
Accession Number
2013_030
Transfer Type
Physical Media
Date Acquired
11/8/2012
Received From
Joe Crabtree
Received By
Sam Meister
Physical Media
Media Storage Location
Electronic Media Storage
Media Format
Magnetic
Media Sub Type
Floppy 3.5 in.
Manufacturer
Imation
Model
High Density IBM Format
Age
16
Condition
Good
Media Label Text
9812 90812A-1 9812A-2
Media Photographed?
No
Disk Image Created?
Yes
Disk Image Date
4/19/2013
Disk Image Filename
2013_030_079.aff
Disk Image Software
FTK Imager
Produce AIP
Virus Scan?
Yes
Metadata Extracted?
AIP Created?
AIP Transferred?
Date Transfer Storage
Network Storage Location
Transfer Notes
Navigation Pane
Record: 318 of 335
No Filter
Search
Form View
Num Lock

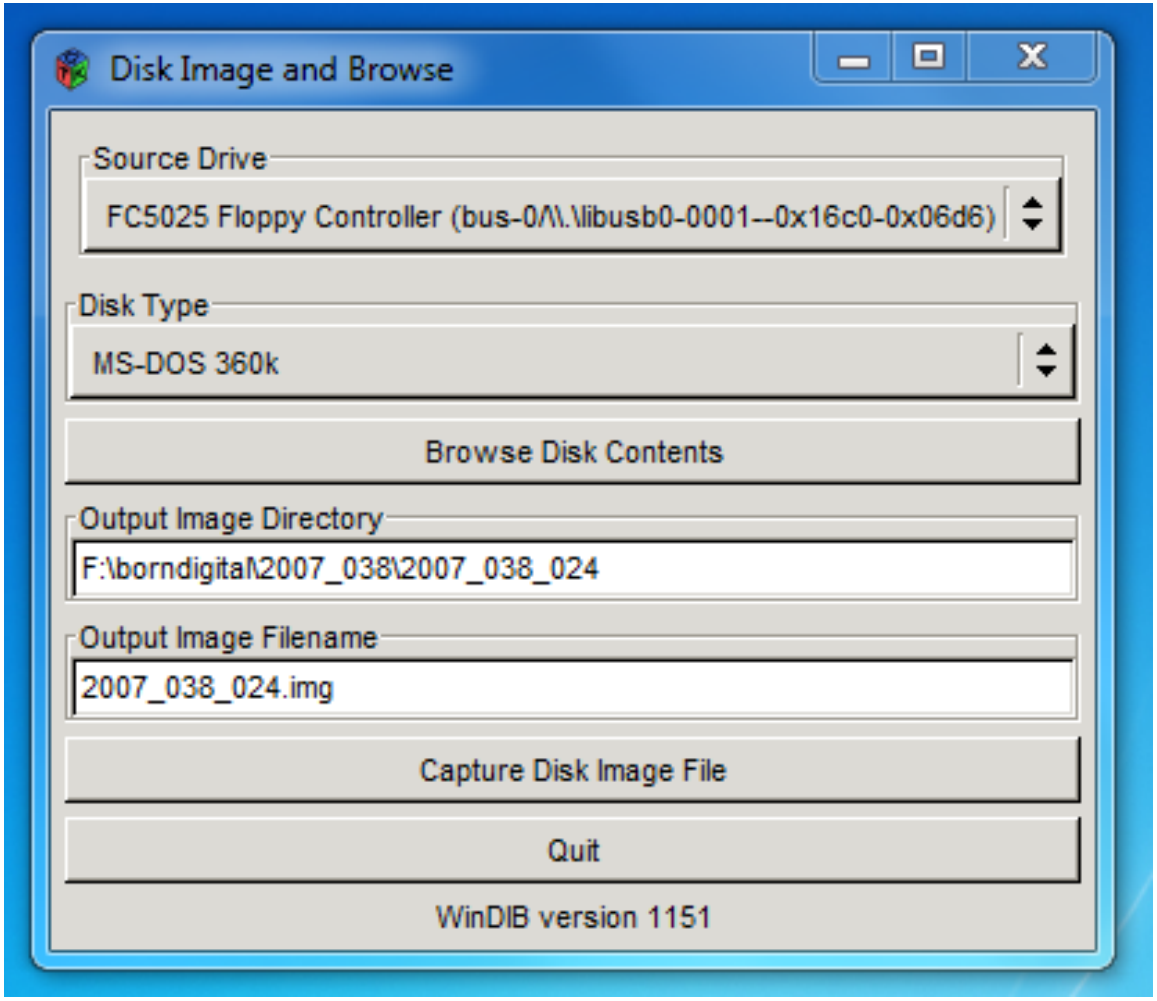


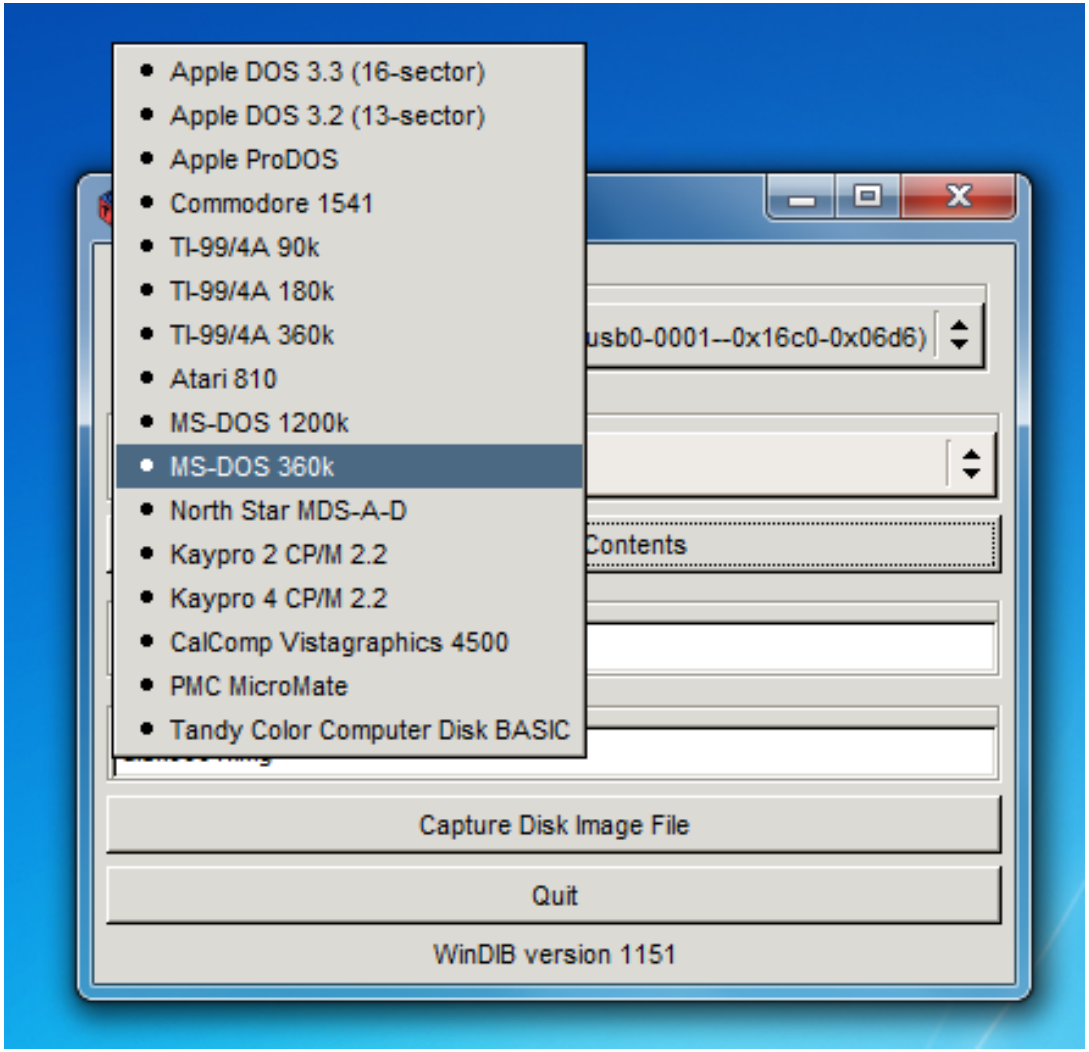
Accession No.	Identifier
2013_030	079

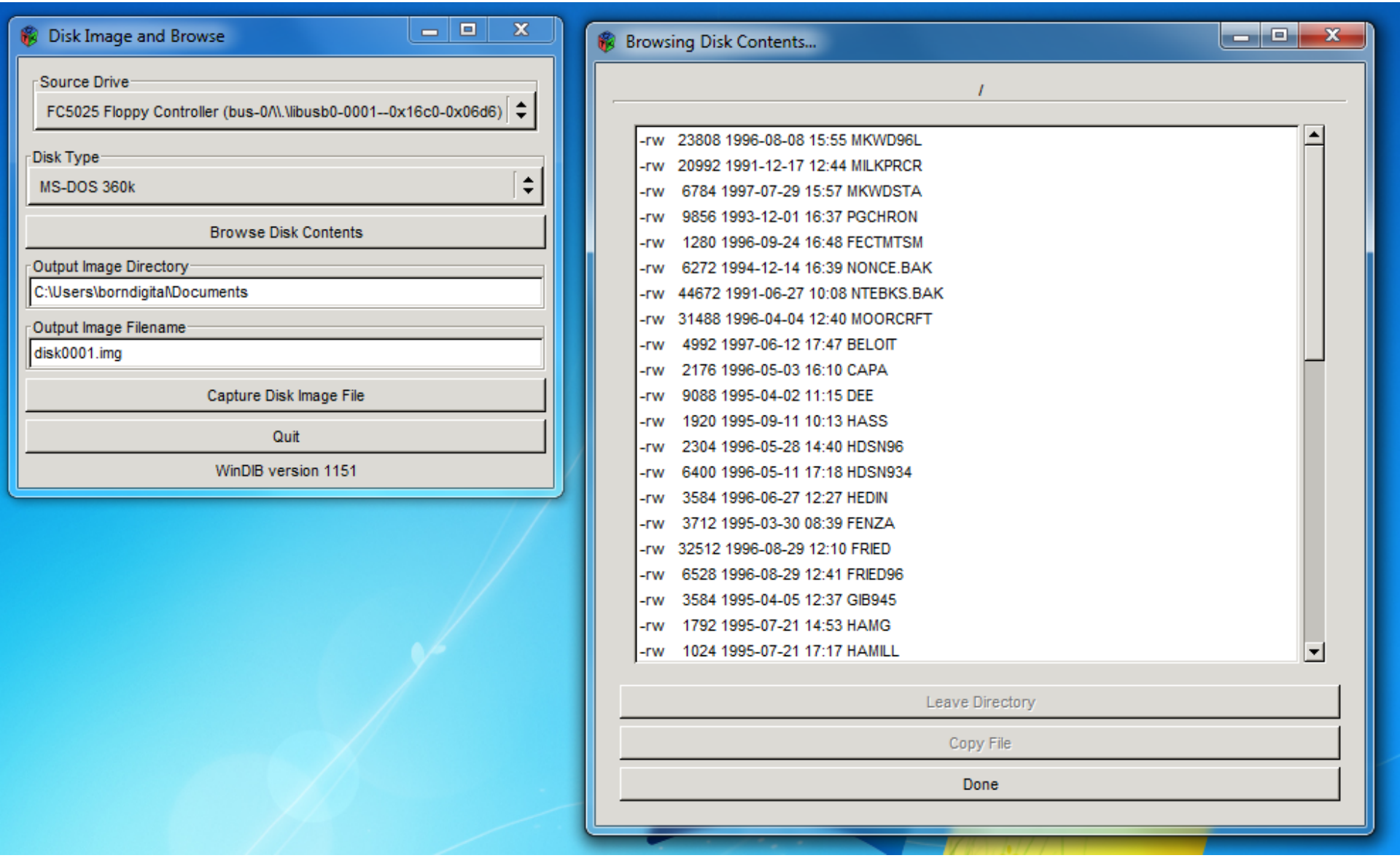
Aspec Fron everse Side / Case

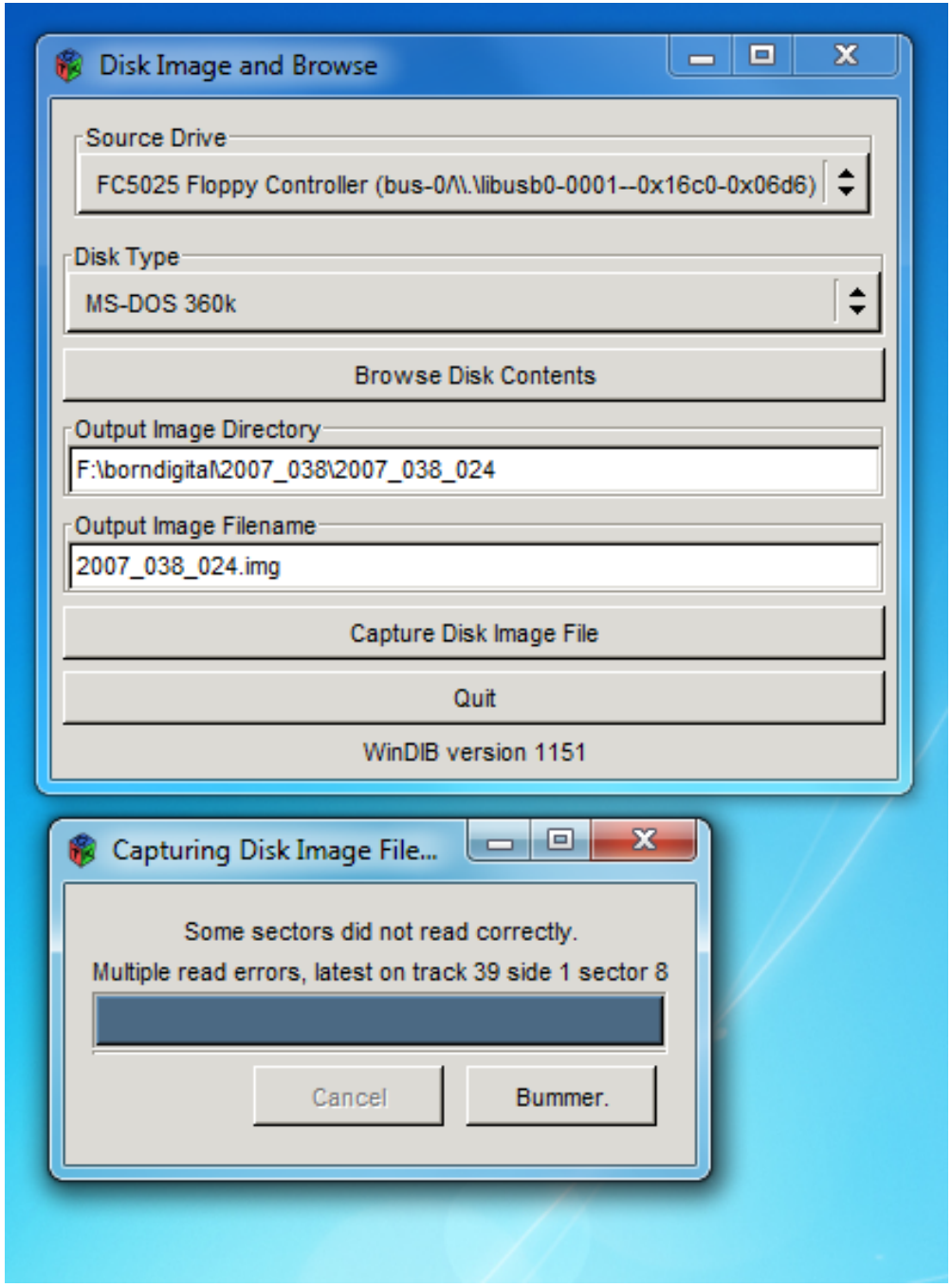
FC5205
Disk Image and Browse

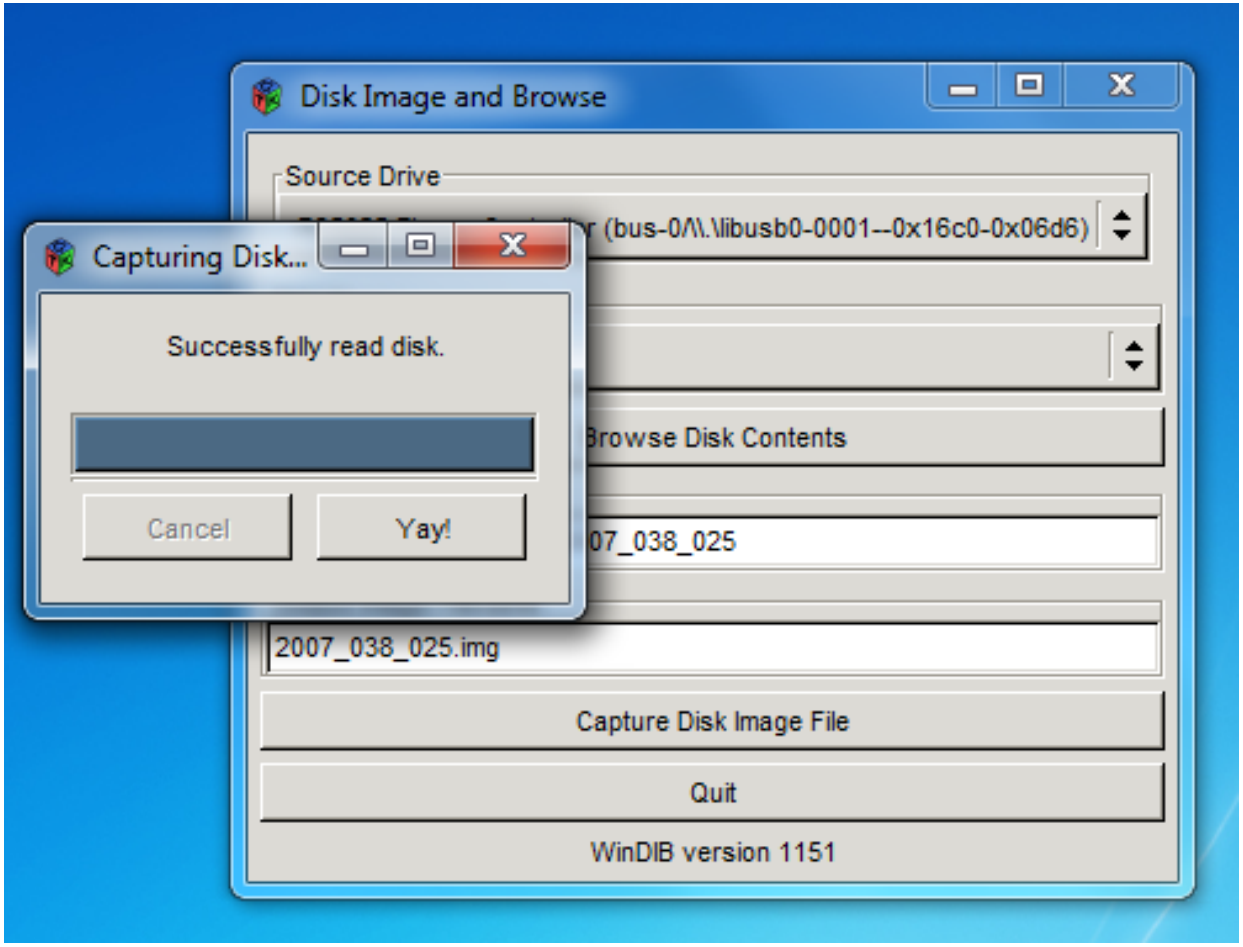



Disk Image and Browse
Source Drive
FC5025 Floppy Controller (bus-0/\\.\libusb0-0001--0x16c0-0x06d6)
Disk Type
MS-DOS 360k
Browse Disk Contents
Output Image Directory
F:\borndigital\2007_038\2007_038_024
Output Image Filename
2007_038_024.img
Capture Disk Image File
Quit
WinDIB version 1151

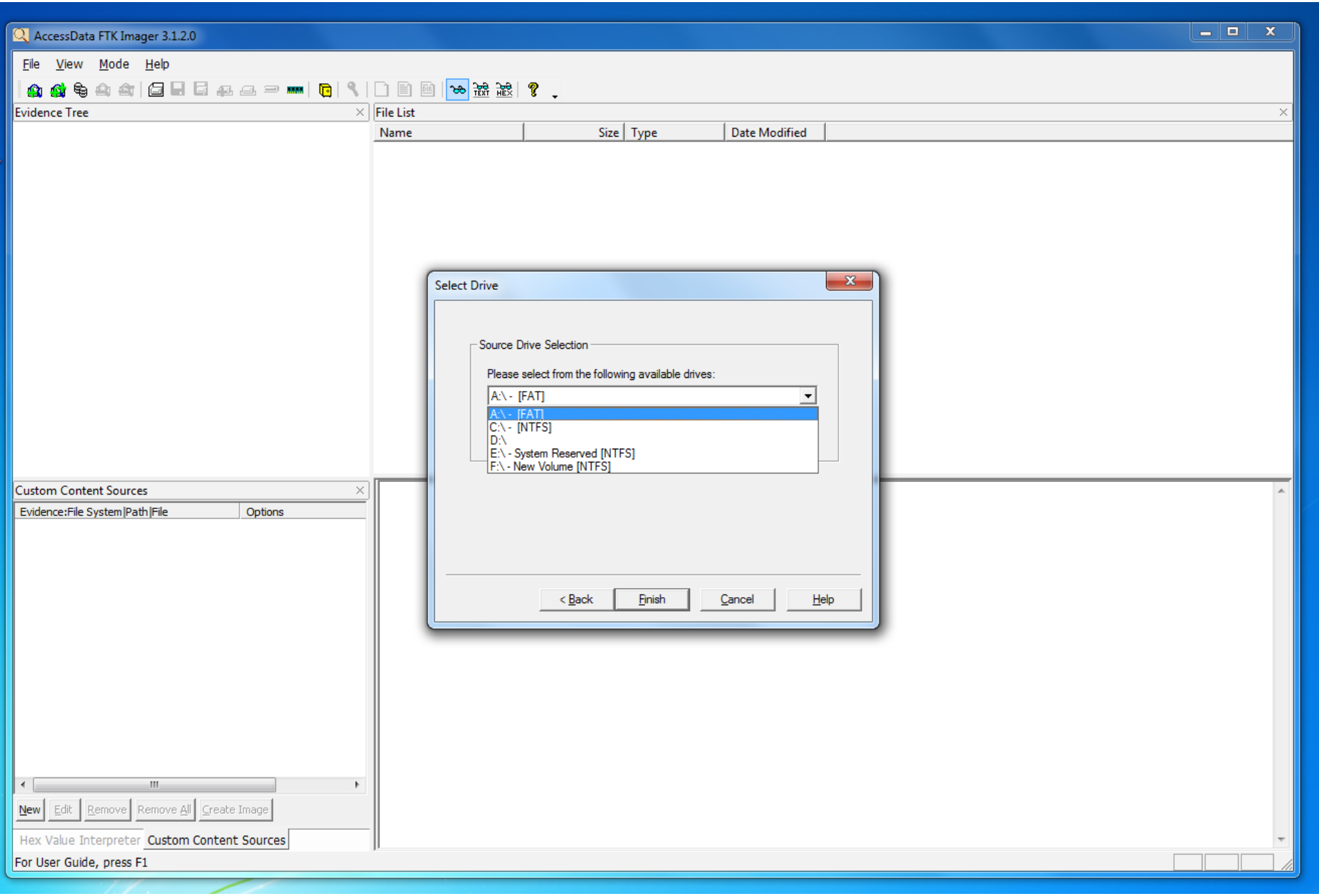

Apple DOS 3.3 (16-sector)
Apple DOS 3.2 (13-sector)
Apple ProDOS
Commodore 1541
TI-99/4A 90k
TI-99/4A 180k
TI-99/4A 360k
Atari 810
MS-DOS 1200k
MS-DOS 360k
North Star MDS-A-D
Kaypro 2 CP/M 2.2
Kaypro 4 CP/M 2.2
CalComp Vistagraphics 4500
PMC MicroMate
Tandy Color Computer Disk BASIC
usb0-0001--0x16c0-0x06d6)
Contents
Capture Disk Image File
Quit
WinDIB version 1151

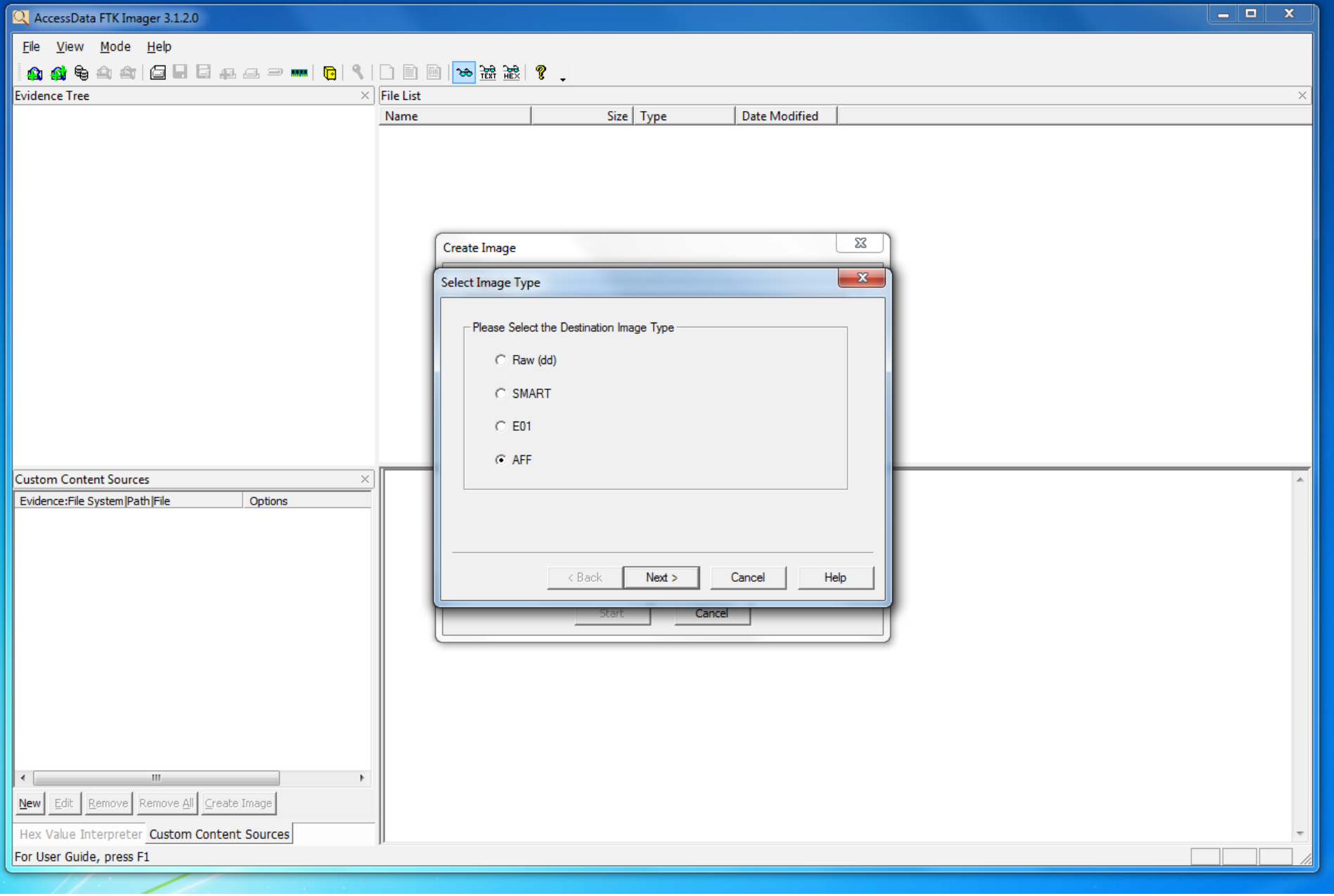

Disk Image and Browse
Source Drive
FC5025 Floppy Controller (bus-0/\\.\libusb0-0001--0x16c0-0x06d6)
Disk Type
MS-DOS 360k
Browse Disk Contents
Output Image Directory
C:\Users\borndigital\Documents
Output Image Filename
disk0001.img
Capture Disk Image File
Quit
WinDIB version 1151
Browsing Disk Contents...
/
-rw 23808 1996-08-08 15:55 MKWD96L
-rw 20992 1991-12-17 12:44 MILKPRCR
-rw 6784 1997-07-29 15:57 MKWDSTA
-rw 9856 1993-12-01 16:37 PGCHRON
-rw 1280 1996-09-24 16:48 FECTMTSM
-rw 6272 1994-12-14 16:39 NONCE.BAK
-rw 44672 1991-06-27 10:08 NTEBKS.BAK
-rw 31488 1996-04-04 12:40 MOORCRFT
-rw 4992 1997-06-12 17:47 BELOIT
-rw 2176 1996-05-03 16:10 CAPA
-rw 9088 1995-04-02 11:15 DEE
-rw 1920 1995-09-11 10:13 HASS
-rw 2304 1996-05-28 14:40 HDSN96
-rw 6400 1996-05-11 17:18 HDSN934
-rw 3584 1996-06-27 12:27 HEDIN
-rw 3712 1995-03-30 08:39 FENZA
-rw 32512 1996-08-29 12:10 FRIED
-rw 6528 1996-08-29 12:41 FRIED96
-rw 3584 1995-04-05 12:37 GIB945
-rw 1792 1995-07-21 14:53 HAMG
-rw 1024 1995-07-21 17:17 HAMILL
Leave Directory
Copy File
Done

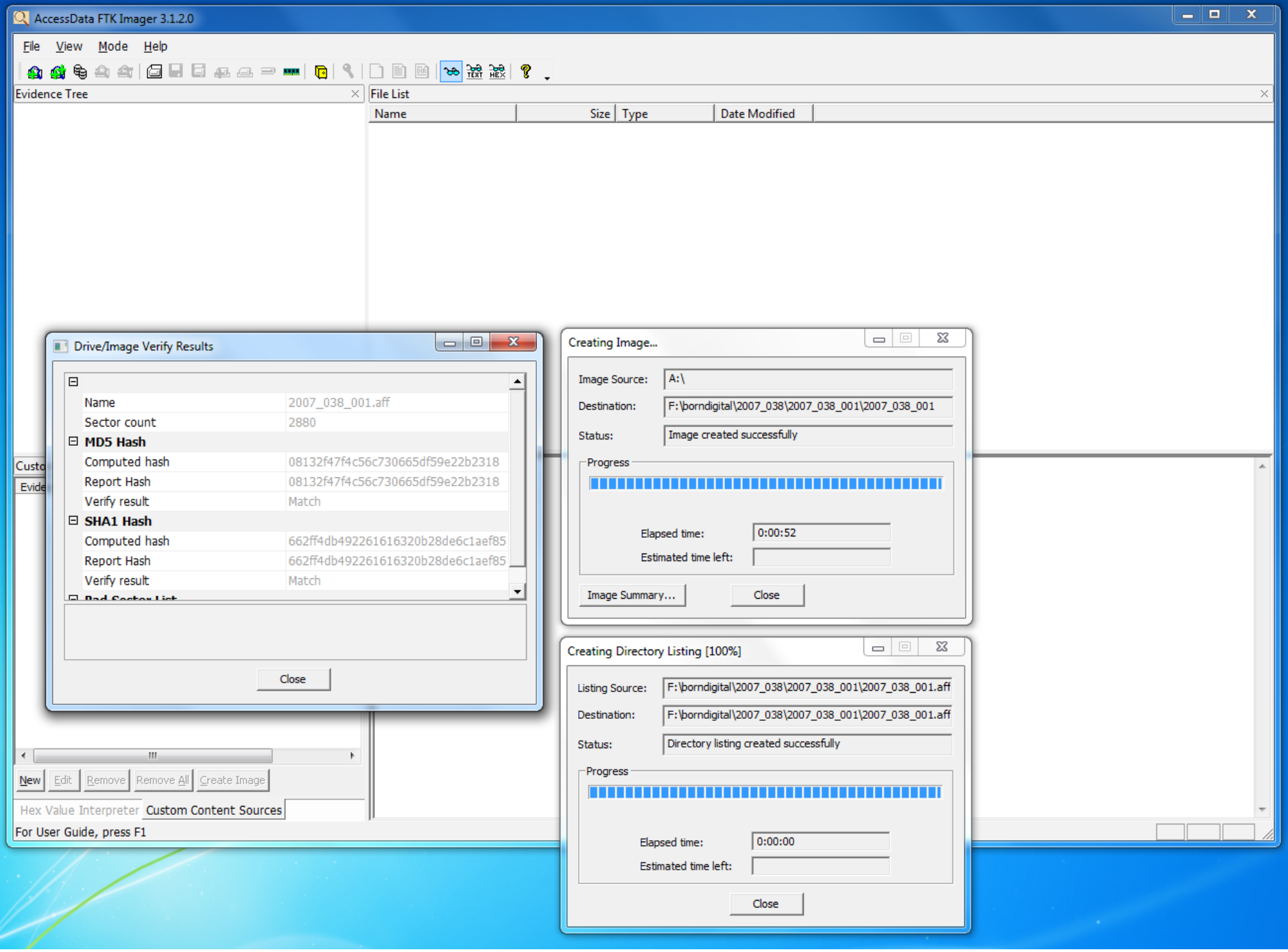

Disk Image and Browse
Source Drive
FC5025 Floppy Controller (bus-0/\\.\libusb0-0001--0x16c0-0x06d6)
Disk Type
MS-DOS 360k
Browse Disk Contents
Output Image Directory
F:\borndigital\2007_038\2007_038_024
Output Image Filename
2007_038_024.img
Capture Disk Image File
Quit
WinDIB version 1151
Capturing Disk Image File...
Some sectors did not read correctly.
Multiple read errors, latest on track 39 side 1 sector 8
Cancel
Bummer.

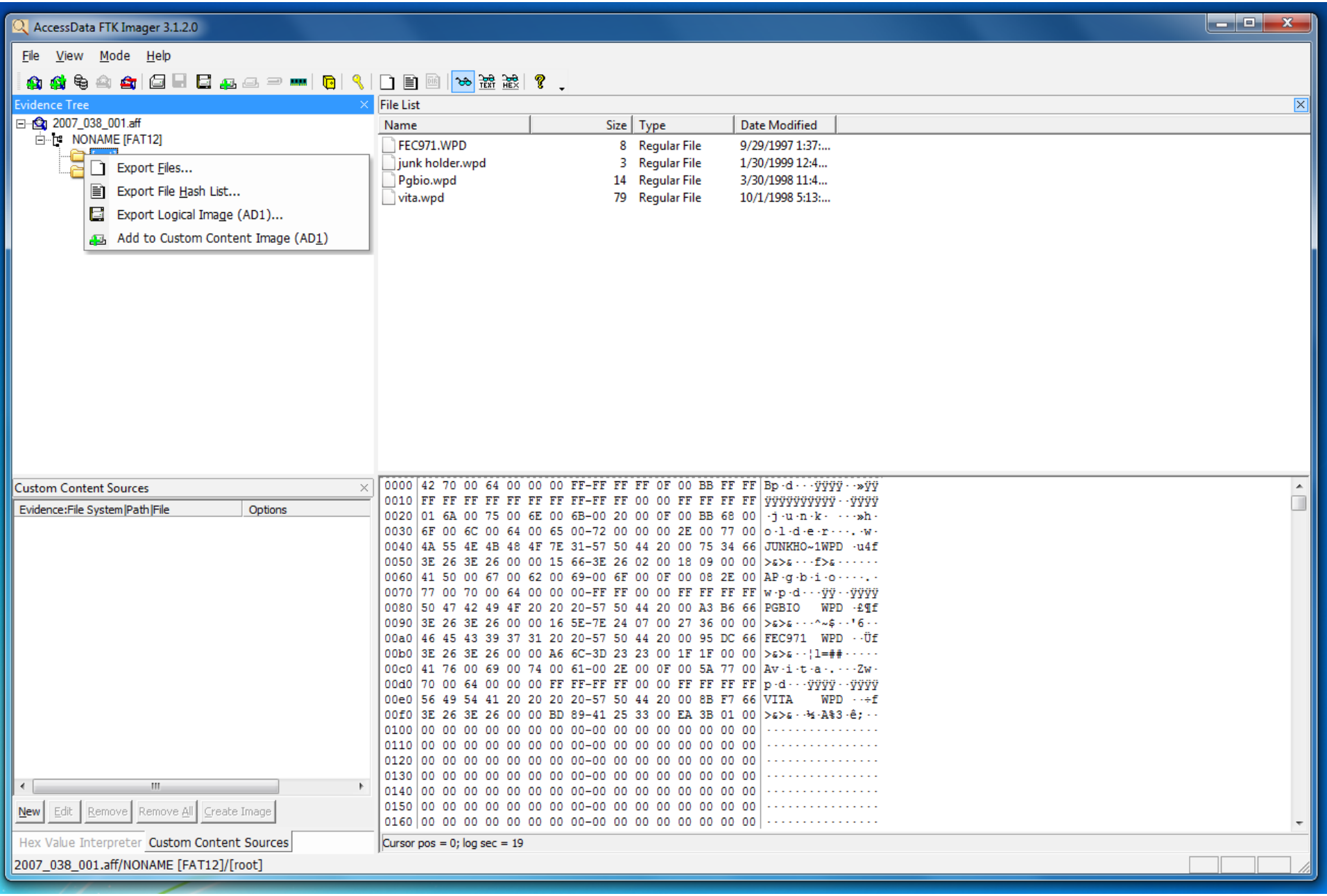

Disk Image and Browse
Source Drive
(bus-0/\\.\libusb0-0001--0x16c0-0x06d6)
Browse Disk Contents
07_038_025
2007_038_025.img
Capture Disk Image File
Quit
WinDIB version 1151
Capturing Disk...
Successfully read disk.
Cancel
Yay!

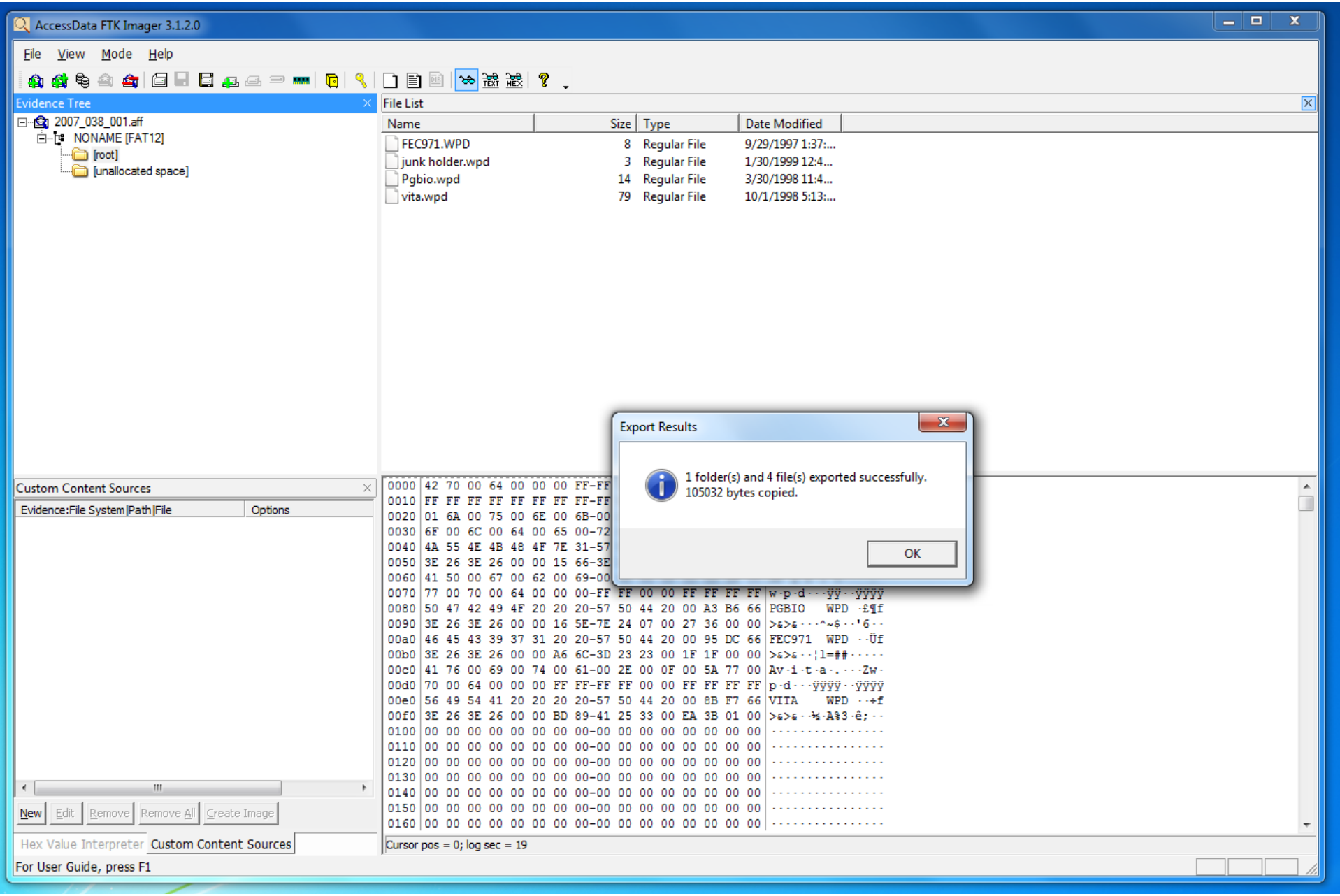
FTK Imager


AccessData FTK Imager 3.1.2.0
File View Mode Help
Evidence Tree
File List
Name Size Type Date Modified
Select Drive
Source Drive Selection
Please select from the following available drives:
A:\ - [FAT]
A:\ - [FAT]
C:\ - [NTFS]
D:\
E:\ - System Reserved [NTFS]
F:\ - New Volume [NTFS]
< Back
Finish
Cancel
Help
Custom Content Sources
Evidence:File System|Path|File
Options
New
Edit
Remove
Remove All
Create Image
Hex Value Interpreter
Custom Content Sources
For User Guide, press F1


AccessData FTK Imager 3.1.2.0
File View Mode Help
Evidence Tree
File List
Name Size Type Date Modified
Create Image
Select Image Type
Please Select the Destination Image Type
Raw (dd)
SMART
E01
AFF
< Back Next > Cancel Help
Start Cancel
Custom Content Sources
Evidence:File System|Path|File Options
New Edit Remove Remove All Create Image
Hex Value Interpreter Custom Content Sources
For User Guide, press F1


AccessData FTK Imager 3.1.2.0
File View Mode Help
Evidence Tree
File List
Name Size Type Date Modified
Drive/Image Verify Results
Name 2007_038_001.aff
Sector count 2880
MD5 Hash
Computed hash 08132f47f4c56c730665df59e22b2318
Report Hash 08132f47f4c56c730665df59e22b2318
Verify result Match
SHA1 Hash
Computed hash 662ff4db492261616320b28de6c1aef85
Report Hash 662ff4db492261616320b28de6c1aef85
Verify result Match
Close
Creating Image...
Image Source: A:\
Destination: F:\borndigital\2007_038\2007_038_001\2007_038_001
Status: Image created successfully
Progress
Elapsed time: 0:00:52
Estimated time left:
Image Summary...
Close
Creating Directory Listing [100%]
Listing Source: F:\borndigital\2007_038\2007_038_001\2007_038_001.aff
Destination: F:\borndigital\2007_038\2007_038_001\2007_038_001.aff
Status: Directory listing created successfully
Progress
Elapsed time: 0:00:00
Estimated time left:
Close
New Edit Remove Remove All Create Image
Hex Value Interpreter Custom Content Sources
For User Guide, press F1

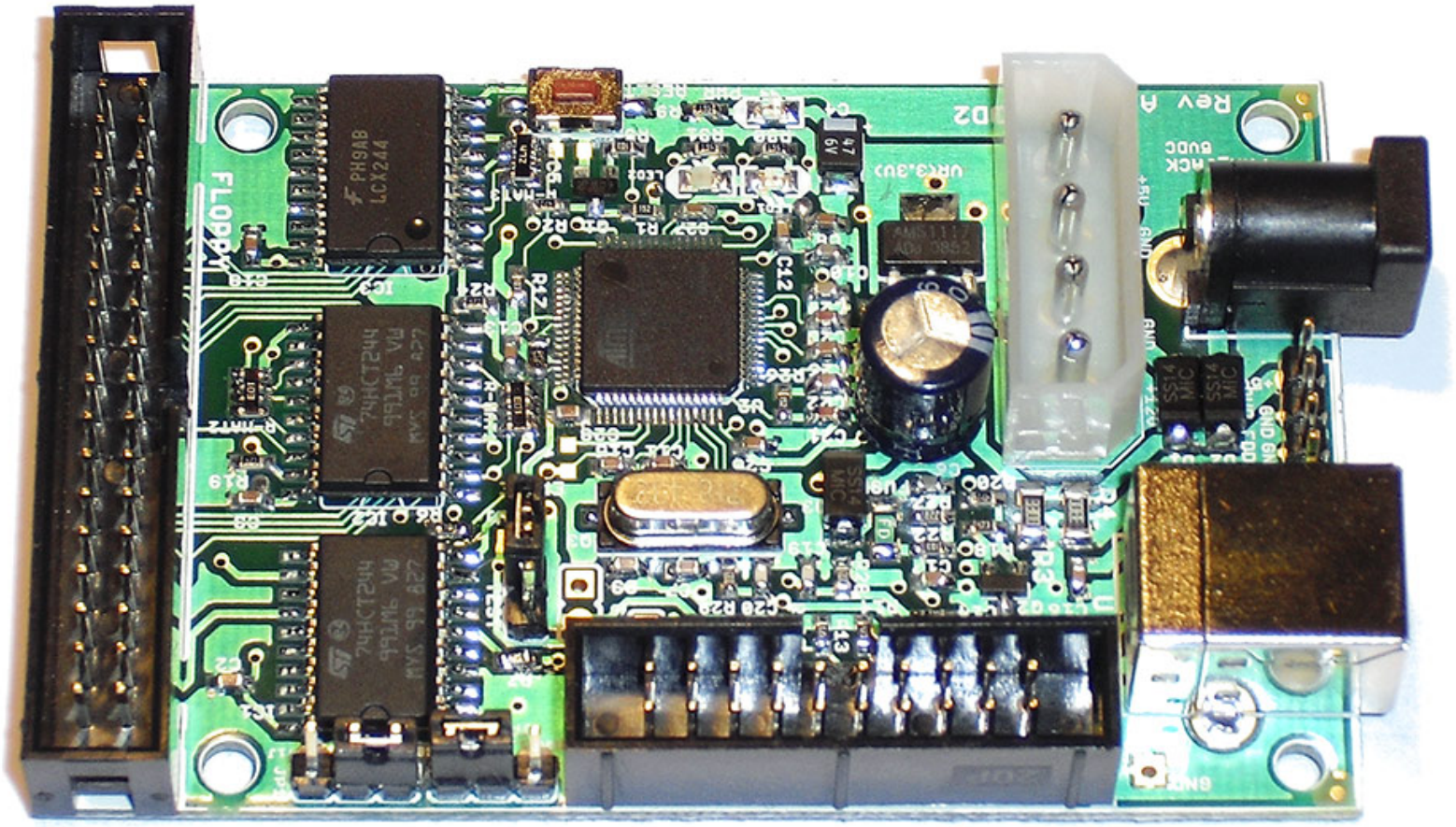

AccessData FTK Imager 3.1.2.0
File View Mode Help
Evidence Tree
2007_038_001.aff
NONAME [FAT12]
Export Files...
Export File Hash List...
Export Logical Image (AD1)...
Add to Custom Content Image (AD1)
File List
Name Size Type Date Modified
FEC971.WPD 8 Regular File 9/29/1997 1:37:...
junk holder.wpd 3 Regular File 1/30/1999 12:4...
Pgbio.wpd 14 Regular File 3/30/1998 11:4...
vita.wpd 79 Regular File 10/1/1998 5:13:...
Custom Content Sources
Evidence:File System|Path|File
Options
New Edit Remove Remove All Create Image
Hex Value Interpreter
Custom Content Sources
Cursor pos = 0; log sec = 19
2007_038_001.aff/NONAME [FAT12]/[root]


AccessData FTK Imager 3.1.2.0
File View Mode Help
Evidence Tree
2007_038_001.aff
NONAME [FAT12]
[root]
[unallocated space]
File List
Name Size Type Date Modified
FEC971.WPD 8 Regular File 9/29/1997 1:37:...
junk holder.wpd 3 Regular File 1/30/1999 12:4...
Pgbio.wpd 14 Regular File 3/30/1998 11:4...
vita.wpd 79 Regular File 10/1/1998 5:13:...
Export Results
1 folder(s) and 4 file(s) exported successfully.
105032 bytes copied.
OK
Custom Content Sources
Evidence:File System|Path|File
Options
New Edit Remove Remove All Create Image
Hex Value Interpreter
Custom Content Sources
Cursor pos = 0; log sec = 19
For User Guide, press F1

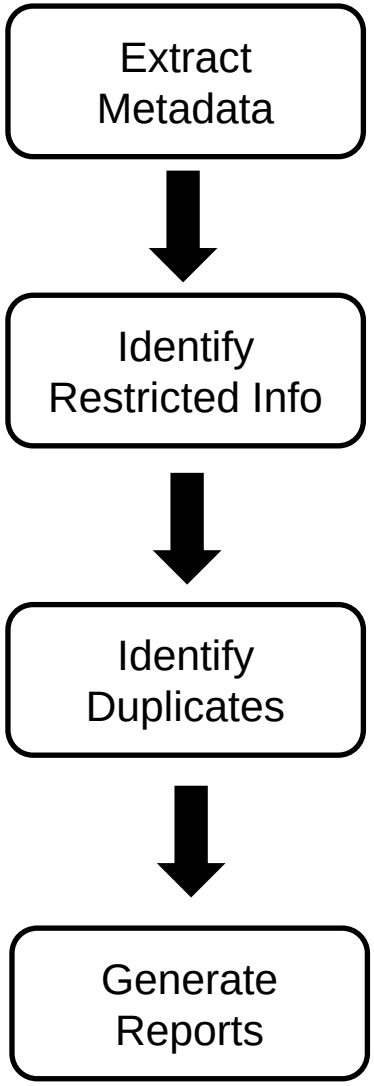
Issue:

Unknown / Unrecognized Filesystems

KryoFlux



Initial Analysis



Initial Analysis

Hardware

Software

BitCurator
fiwalk
Bulk Extractor

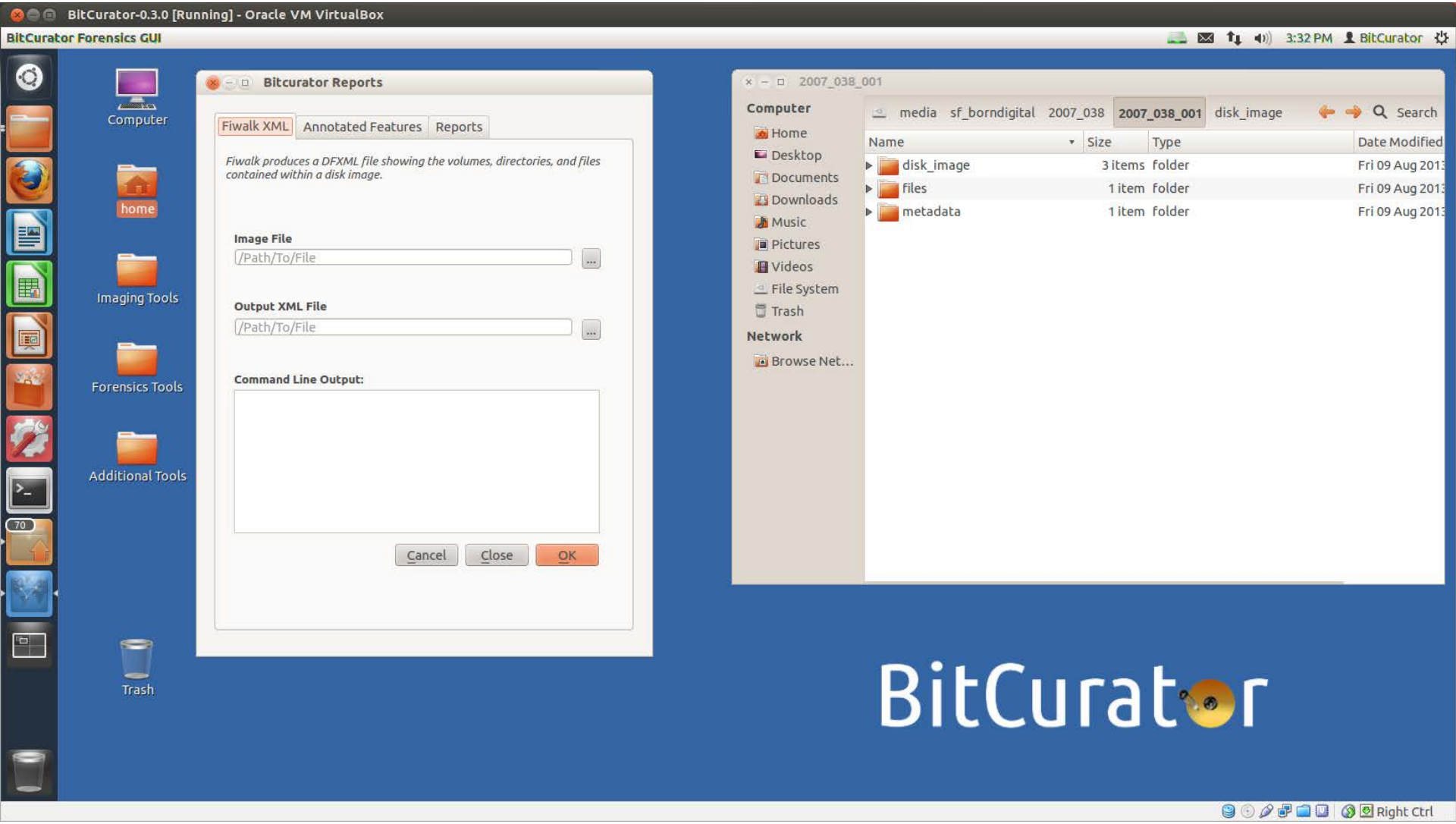
BitCurator

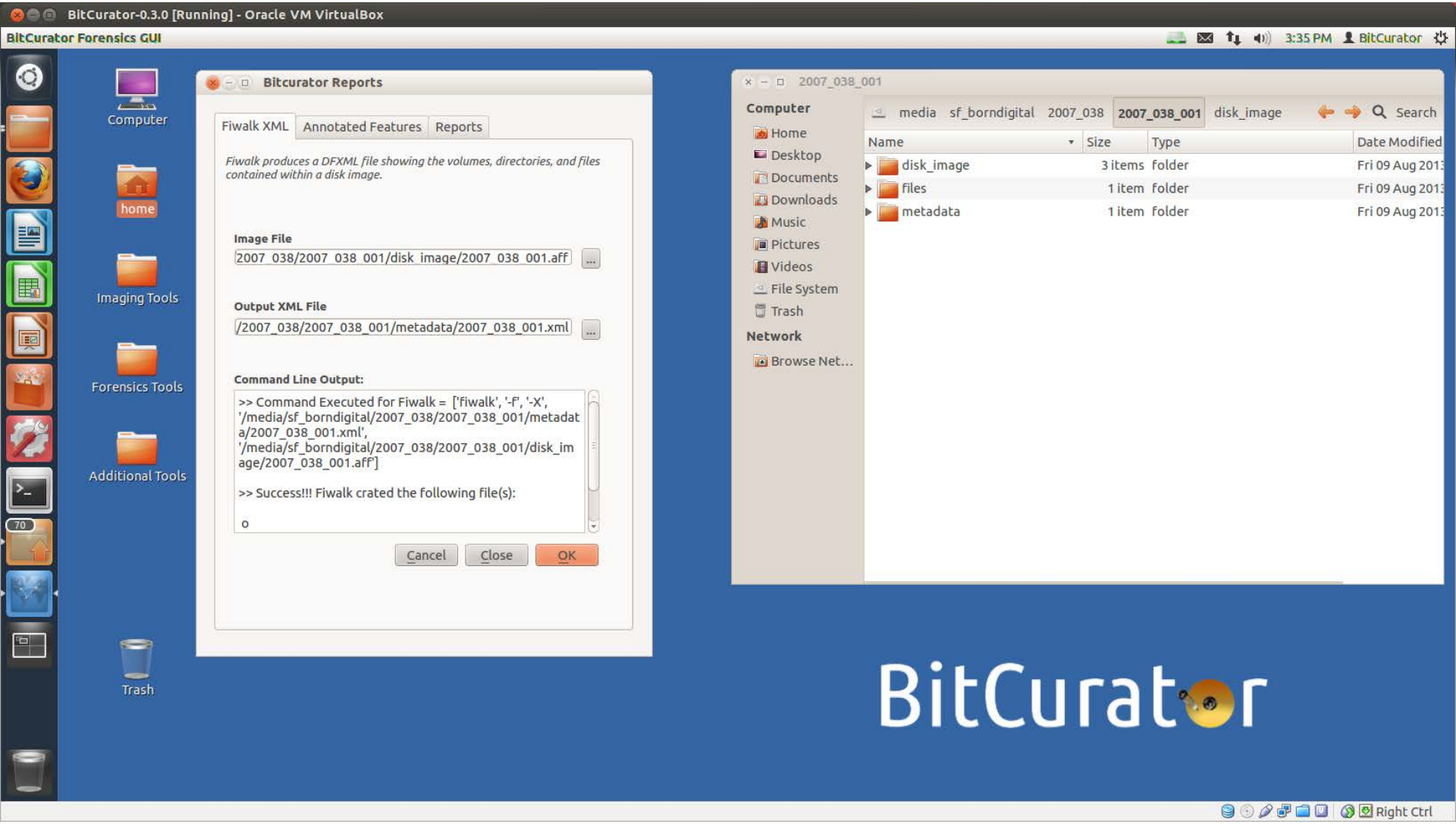
Tools for Digital Forensics Methods and Workflows
in Real-World Collecting Institutions

“an effort to build, test, and analyze systems and software
for incorporating digital forensics methods
into the workflows of a variety of collecting institutions”

BitCurator:

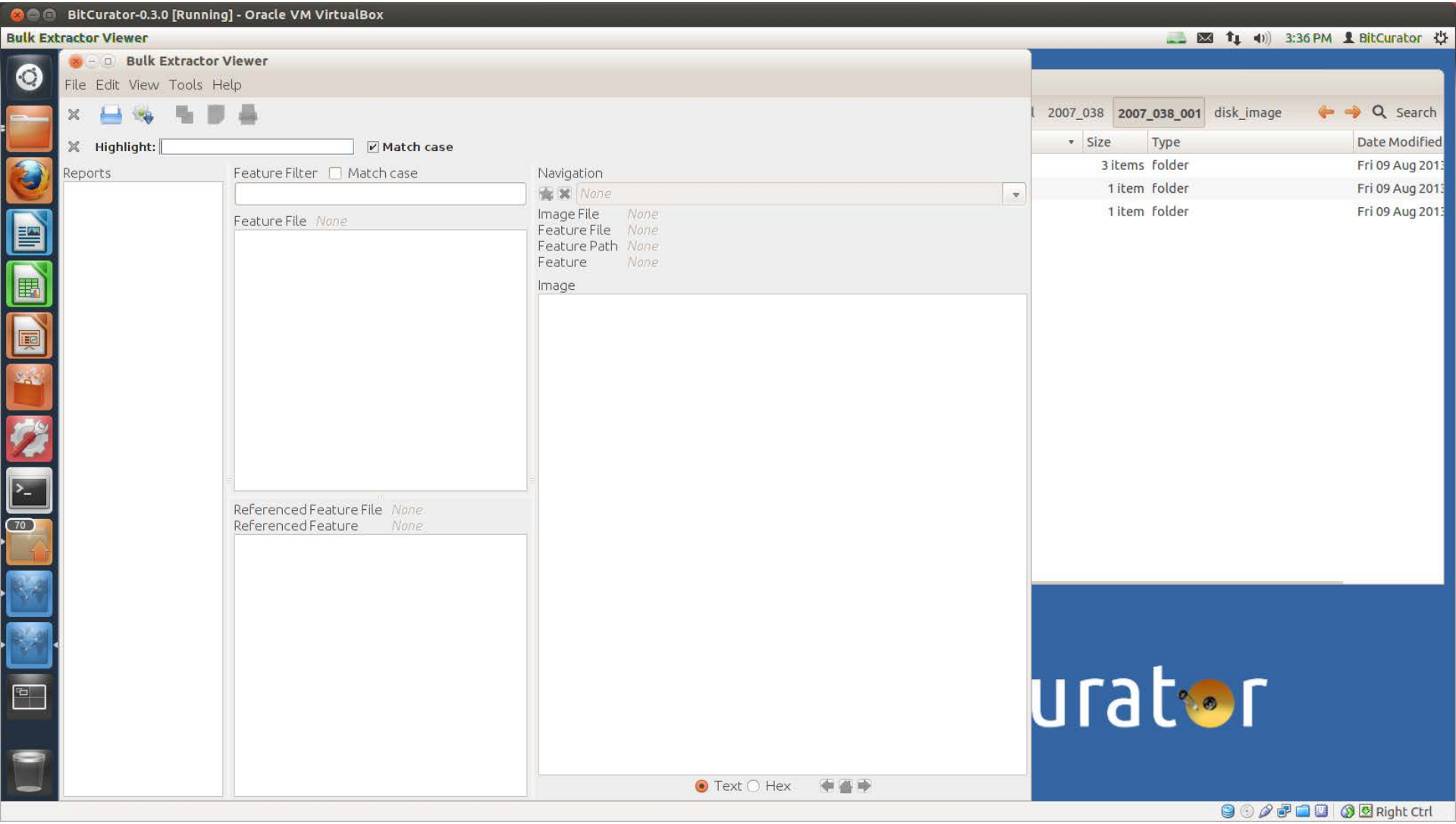
fiwalk

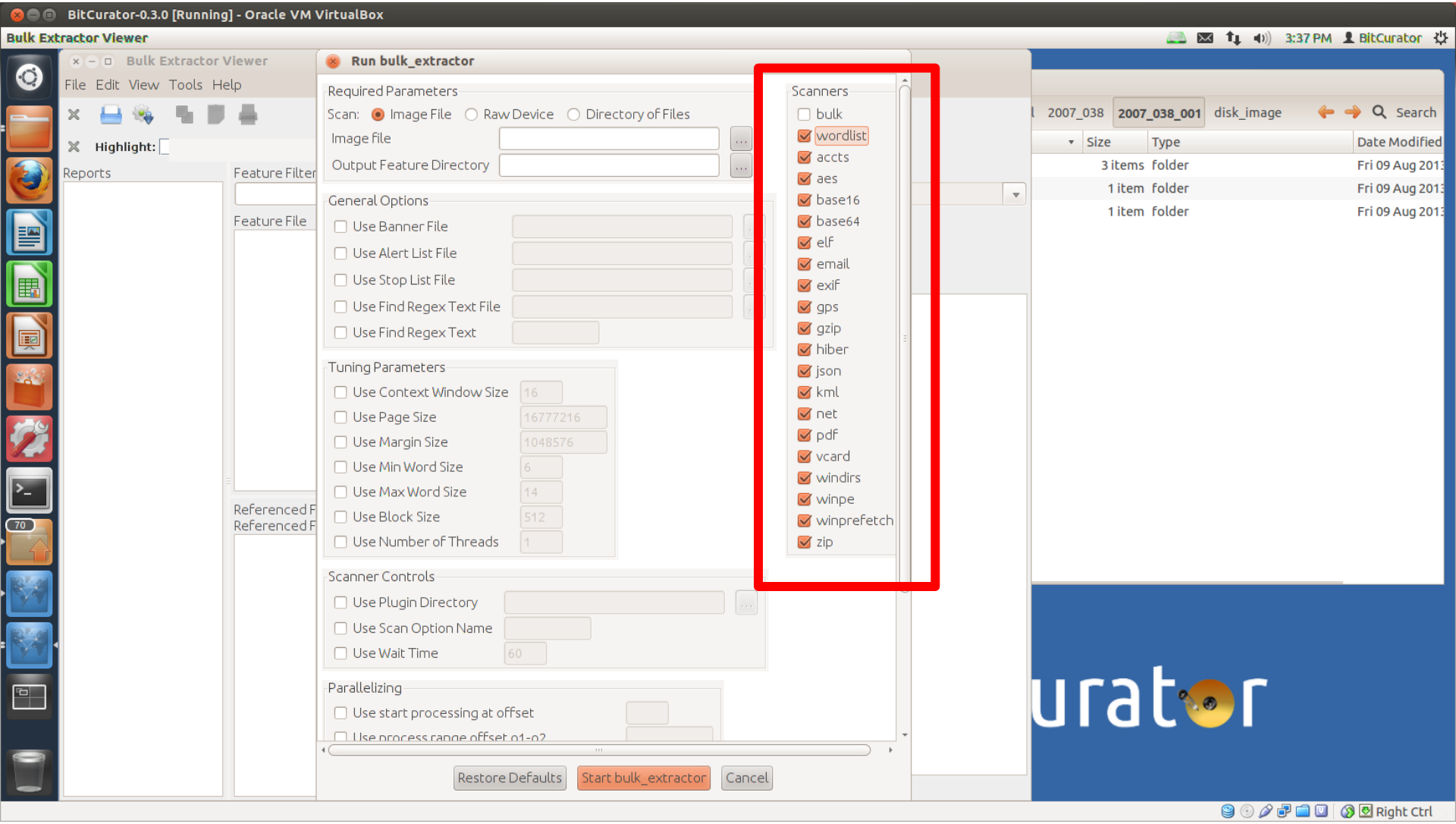

BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
BitCurator Forensics GUI
3:32 PM
BitCurator
Computer
home
Imaging Tools
Forensics Tools
Additional Tools
Trash
Bitcurator Reports
Fiwalk XML
Annotated Features
Reports
Fiwalk produces a DFXML file showing the volumes, directories, and files contained within a disk image.
Image File
/Path/To/File
Output XML File
/Path/To/File
Command Line Output:
Cancel
Close
OK
2007_038_001
Computer
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
File System
Trash
Network
Browse Net...
media
sf_borndigital
2007_038
2007_038_001
disk_image
Search
Name
Size
Type
Date Modified
disk_image
3 items
folder
Fri 09 Aug 2013
files
1 item
folder
Fri 09 Aug 2013
metadata
1 item
folder
Fri 09 Aug 2013
BitCurator
Right Ctrl

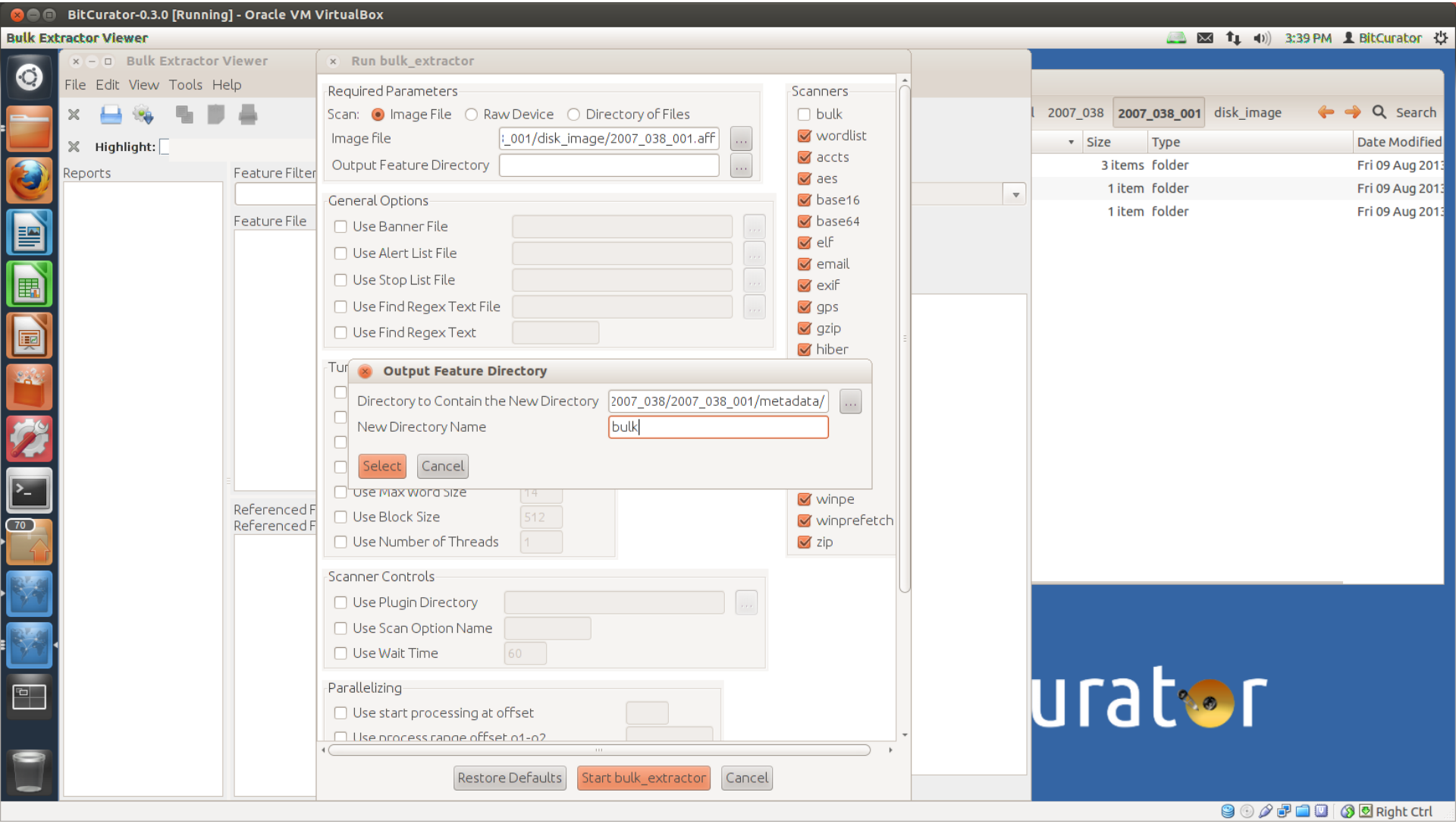

BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
BitCurator Forensics GUI
3:35 PM
BitCurator
Computer
home
Imaging Tools
Forensics Tools
Additional Tools
Trash
Bitcurator Reports
Fiwalk XML
Annotated Features
Reports
Fiwalk produces a DFXML file showing the volumes, directories, and files contained within a disk image.
Image File
2007_038/2007_038_001/disk_image/2007_038_001.aff
Output XML File
/2007_038/2007_038_001/metadata/2007_038_001.xml
Command Line Output:
>> Command Executed for Fiwalk = ['fiwalk', '-f', '-X', '/media/sf_borndigital/2007_038/2007_038_001/metadata/2007_038_001.xml', '/media/sf_borndigital/2007_038/2007_038_001/disk_image/2007_038_001.aff']
>> Success!!! Fiwalk crated the following file(s):
o
Cancel
Close
OK
2007_038_001
Computer
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
File System
Trash
Network
Browse Net...
media
sf_borndigital
2007_038
2007_038_001
disk_image
Search
Name
Size
Type
Date Modified
disk_image
3 items
folder
Fri 09 Aug 2013
files
1 item
folder
Fri 09 Aug 2013
metadata
1 item
folder
Fri 09 Aug 2013
BitCurator
Right Ctrl

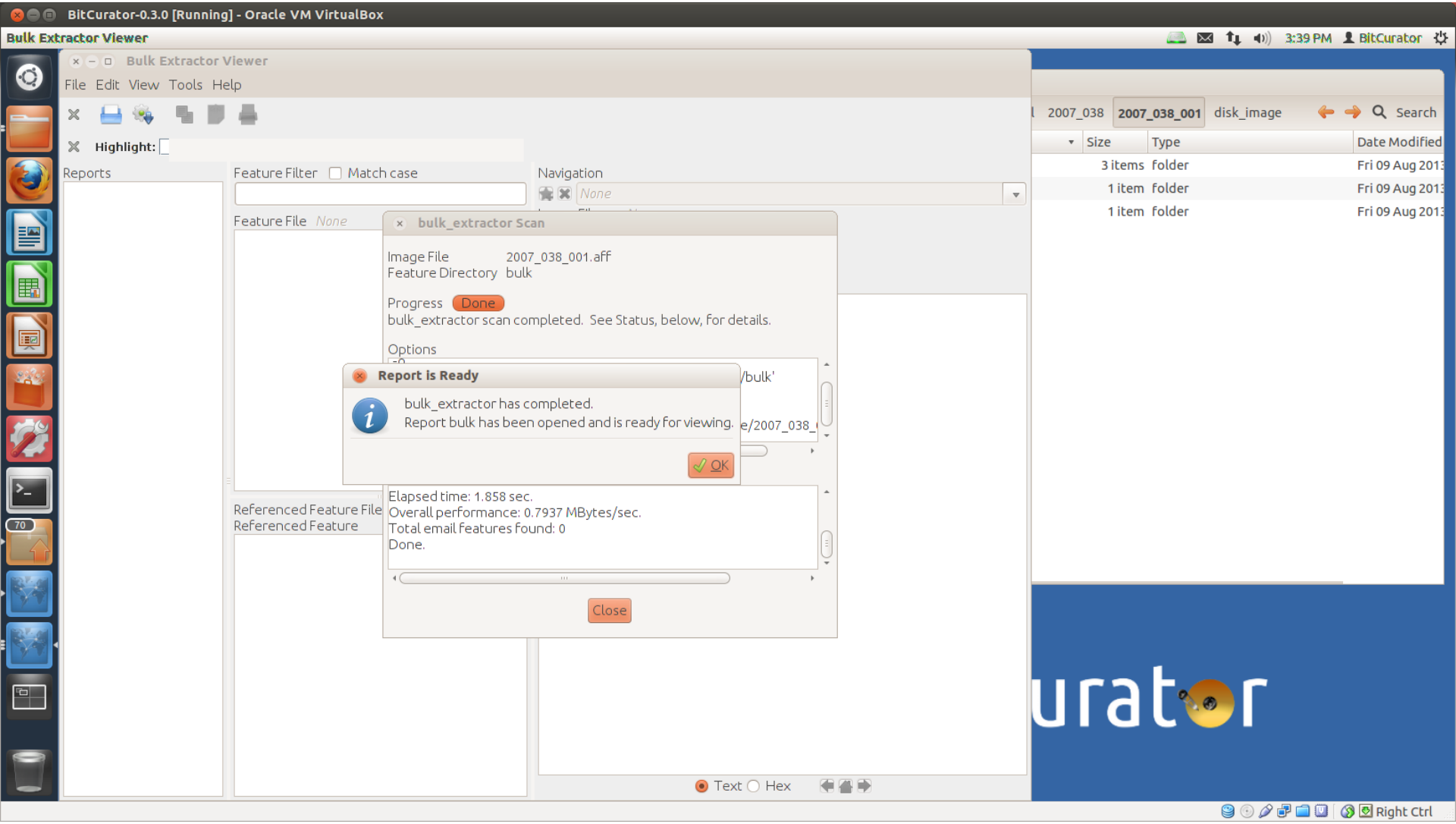
BitCurator:

bulk_extractor


BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
Bulk Extractor Viewer
3:36 PM
BitCurator
Bulk Extractor Viewer
File Edit View Tools Help
Highlight:
Match case
Reports
Feature Filter
Match case
Feature File None
Referenced Feature File None
Referenced Feature None
Navigation
None
Image File None
Feature File None
Feature Path None
Feature None
Image
Text
Hex
2007_038
2007_038_001
disk_image
Search
Size
Type
Date Modified
3 items folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
Right Ctrl

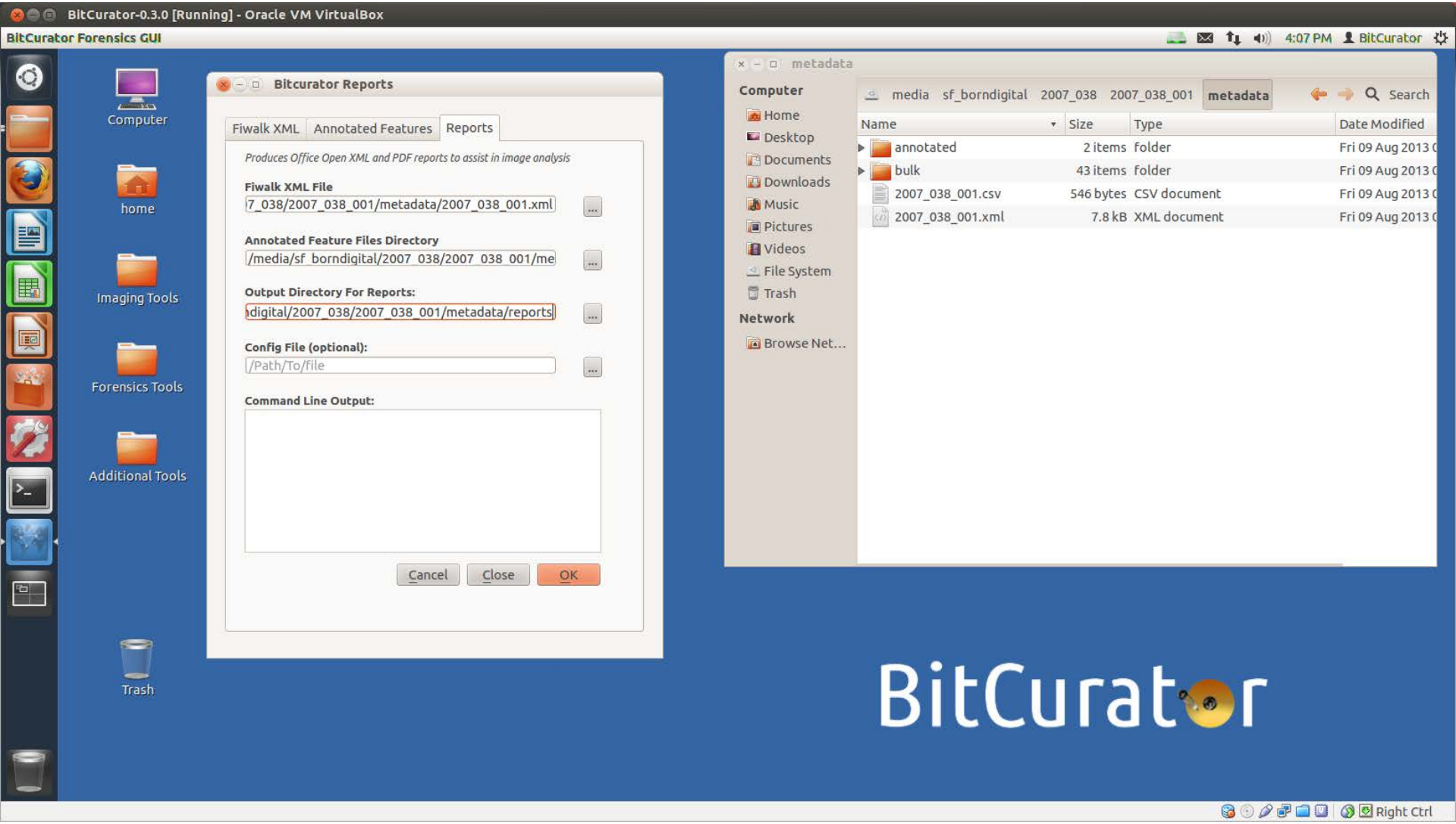

BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
Bulk Extractor Viewer
3:37 PM
BitCurator
Bulk Extractor Viewer
File Edit View Tools Help
Highlight:
Reports
Feature Filter
Feature File
Referenced F
Referenced F
Run bulk_extractor
Required Parameters
Scan: Image File Raw Device Directory of Files
Image file
Output Feature Directory
General Options
Use Banner File
Use Alert List File
Use Stop List File
Use Find Regex Text File
Use Find Regex Text
Tuning Parameters
Use Context Window Size 16
Use Page Size 16777216
Use Margin Size 1048576
Use Min Word Size 6
Use Max Word Size 14
Use Block Size 512
Use Number of Threads 1
Scanner Controls
Use Plugin Directory
Use Scan Option Name
Use Wait Time 60
Parallelizing
Use start processing at offset
Use process range offset o1-o2
Scanners
bulk
wordlist
accts
aes
base16
base64
elf
email
exif
gps
gzip
hiber
json
kml
net
pdf
vcard
windirs
winpe
winprefetch
zip
Restore Defaults
Start bulk_extractor
Cancel
2007_038
2007_038_001
disk_image
Search
Size
Type
Date Modified
3 items folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
Right Ctrl

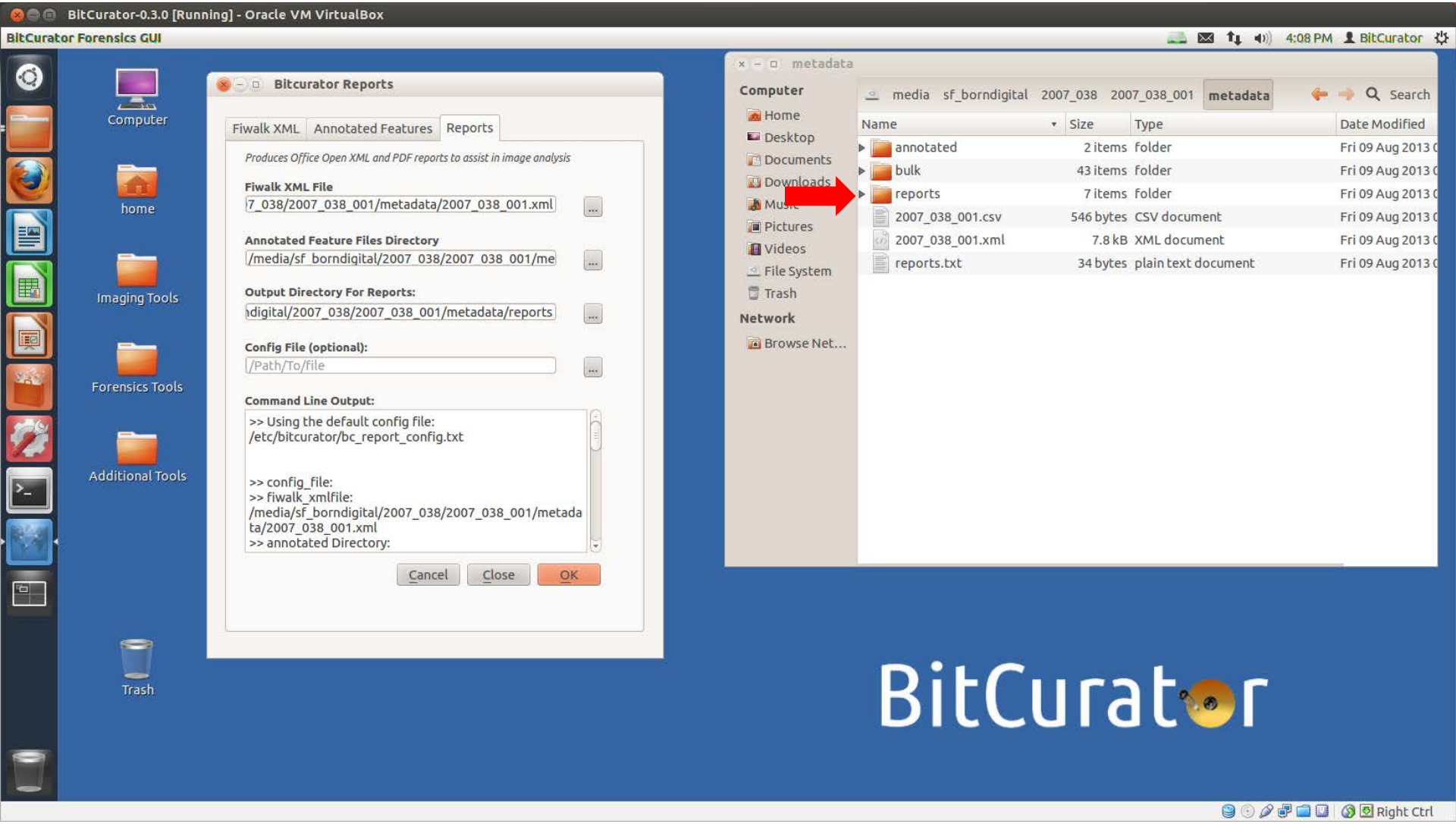

BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
Bulk Extractor Viewer
3:39 PM
BitCurator
Bulk Extractor Viewer
File Edit View Tools Help
Highlight:
Reports
Feature Filter
Feature File
Run bulk_extractor
Required Parameters
Scan: Image File Raw Device Directory of Files
Image file
:_001/disk_image/2007_038_001.aff
Output Feature Directory
General Options
Use Banner File
Use Alert List File
Use Stop List File
Use Find Regex Text File
Use Find Regex Text
Output Feature Directory
Directory to Contain the New Directory
2007_038/2007_038_001/metadata/
New Directory Name
bulk
Select
Cancel
Use Max Word Size
Use Block Size
512
Use Number of Threads
1
Scanners
bulk
wordlist
accts
aes
base16
base64
elf
email
exif
gps
gzip
hiber
winpe
winprefetch
zip
Scanner Controls
Use Plugin Directory
Use Scan Option Name
Use Wait Time
60
Parallelizing
Use start processing at offset
Restore Defaults
Start bulk_extractor
Cancel
2007_038
2007_038_001
disk_image
Search
Size
Type
Date Modified
3 items folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
Right Ctrl


BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
Bulk Extractor Viewer
3:39 PM
BitCurator
Bulk Extractor Viewer
File Edit View Tools Help
Highlight:
Reports
Feature Filter
Match case
Navigation
None
Feature File None
bulk_extractor Scan
Image File 2007_038_001.aff
Feature Directory bulk
Progress Done
bulk_extractor scan completed. See Status, below, for details.
Options
Report is Ready
bulk_extractor has completed.
Report bulk has been opened and is ready for viewing.
OK
/bulk'
e/2007_038_
Elapsed time: 1.858 sec.
Overall performance: 0.7937 MBytes/sec.
Total email features found: 0
Done.
Close
Referenced Feature File
Referenced Feature
Text Hex
2007_038
2007_038_001
disk_image
Search
Size
Type
Date Modified
3 items folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
1 item folder Fri 09 Aug 2013
urator
Right Ctrl

BitCurator:

Reports


BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
BitCurator Forensics GUI
4:07 PM
BitCurator
Computer
home
Imaging Tools
Forensics Tools
Additional Tools
Trash
Bitcurator Reports
Fiwalk XML
Annotated Features
Reports
Produces Office Open XML and PDF reports to assist in image analysis
Fiwalk XML File
7_038/2007_038_001/metadata/2007_038_001.xml
Annotated Feature Files Directory
/media/sf_borndigital/2007_038/2007_038_001/me
Output Directory For Reports:
ndigital/2007_038/2007_038_001/metadata/reports
Config File (optional):
/Path/To/file
Command Line Output:
Cancel
Close
OK
metadata
Computer
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
File System
Trash
Network
Browse Net...
media
sf_borndigital
2007_038
2007_038_001
metadata
Search
Name
Size
Type
Date Modified
annotated
2 items
folder
bulk
43 items
folder
2007_038_001.csv
546 bytes
CSV document
2007_038_001.xml
7.8 kB
XML document
BitCurator
Right Ctrl


BitCurator-0.3.0 [Running] - Oracle VM VirtualBox
BitCurator Forensics GUI
4:08 PM
BitCurator
Computer
home
Imaging Tools
Forensics Tools
Additional Tools
Trash
Bitcurator Reports
Fiwalk XML
Annotated Features
Reports
Produces Office Open XML and PDF reports to assist in image analysis
Fiwalk XML File
7_038/2007_038_001/metadata/2007_038_001.xml
Annotated Feature Files Directory
/media/sf_borndigital/2007_038/2007_038_001/me
Output Directory For Reports:
ndigital/2007_038/2007_038_001/metadata/reports
Config File (optional):
/Path/To/file
Command Line Output:
>> Using the default config file:
/etc/bitcurator/bc_report_config.txt
>> config_file:
>> fiwalk_xmlfile:
/media/sf_borndigital/2007_038/2007_038_001/metada
ta/2007_038_001.xml
>> annotated Directory:
Cancel
Close
OK
metadata
Computer
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
File System
Trash
Network
Browse Net...
media
sf_borndigital
2007_038
2007_038_001
metadata
Search
Name
Size
Type
Date Modified
annotated
2 items
folder
bulk
43 items
reports
7 items
2007_038_001.csv
546 bytes
CSV document
2007_038_001.xml
7.8 kB
XML document
reports.txt
34 bytes
plain text document
BitCurator
Right Ctrl

Produce AIP

AIP = Archival Information Package

Produce AIP

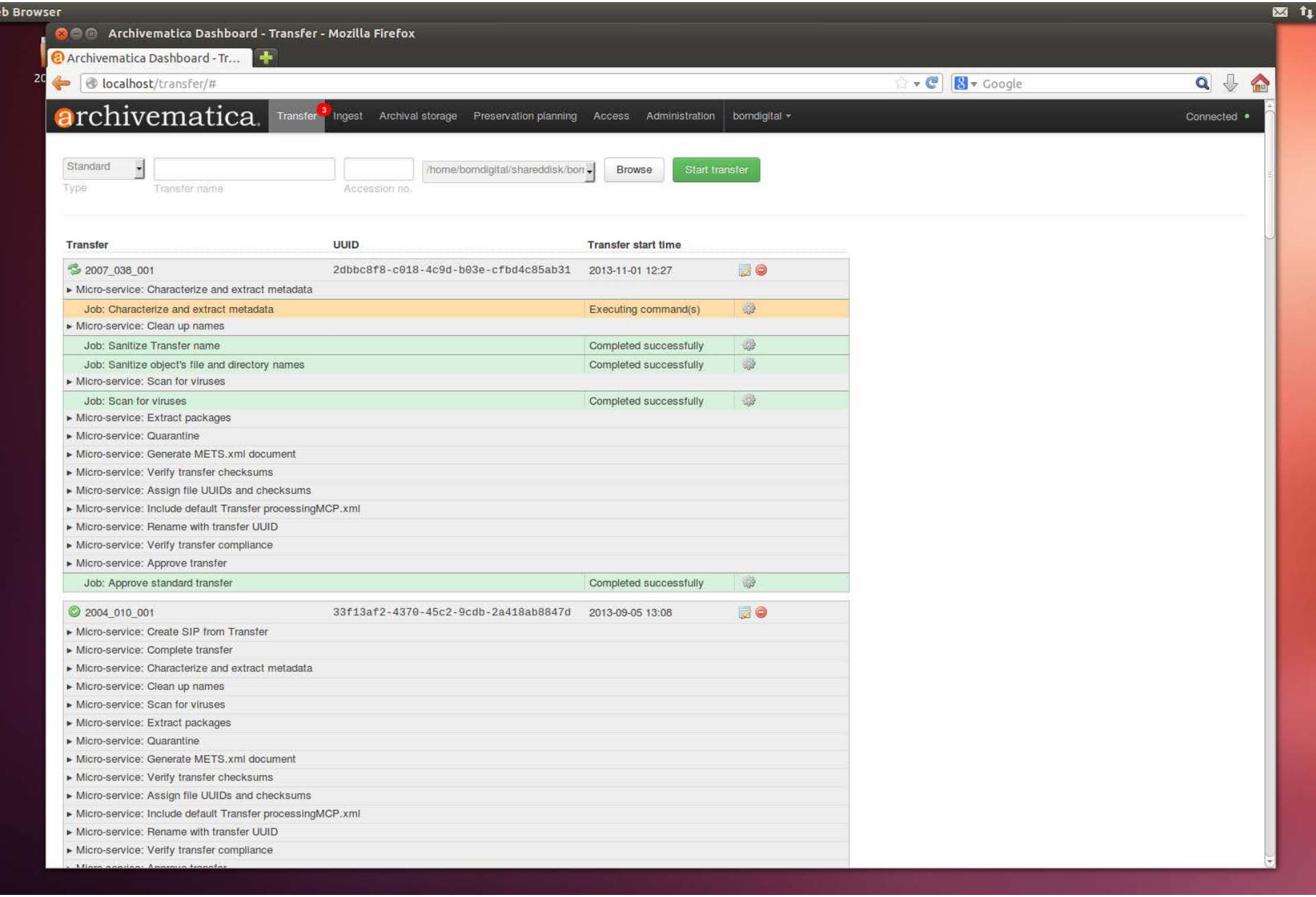
Hardware

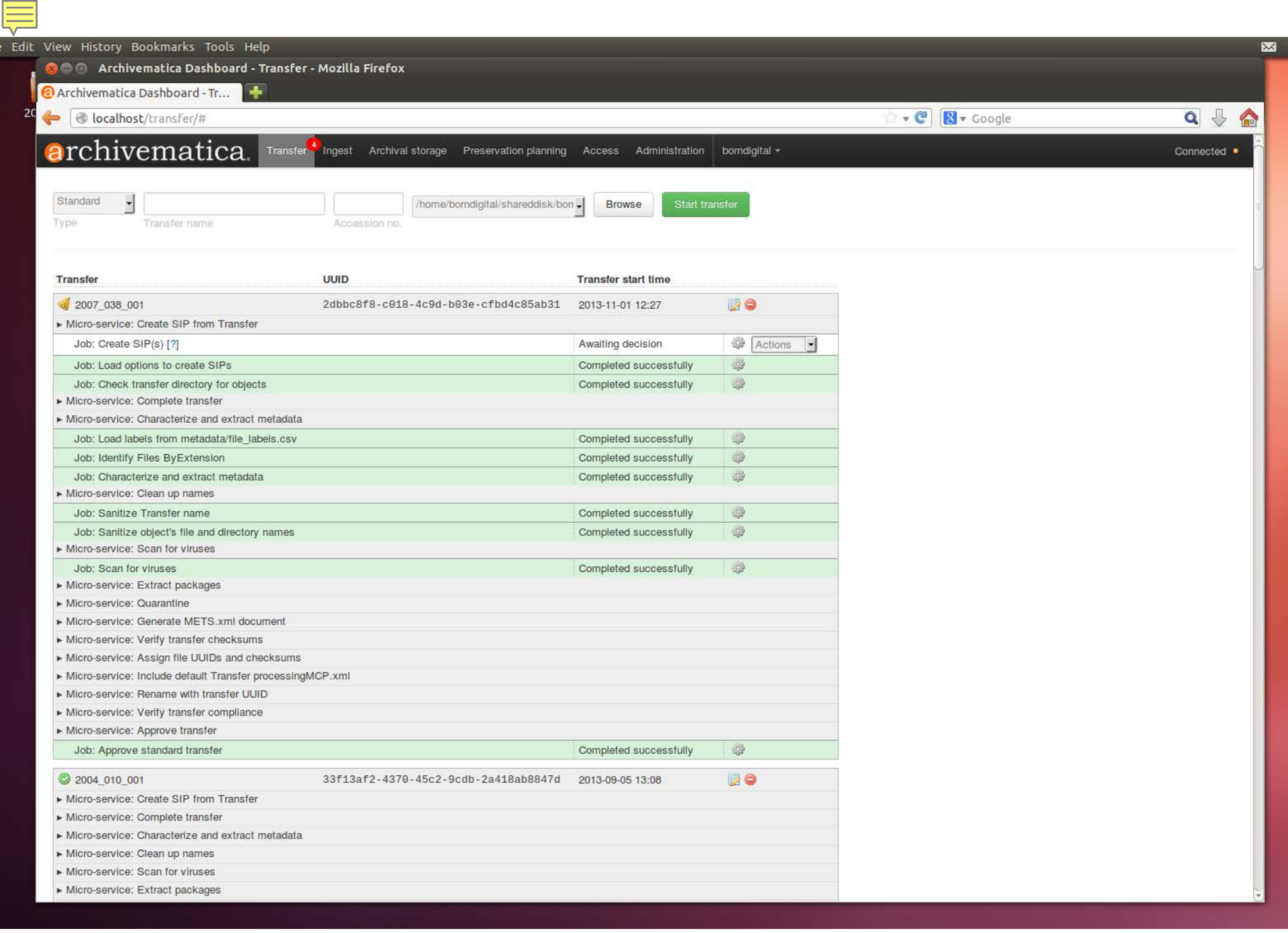
Software

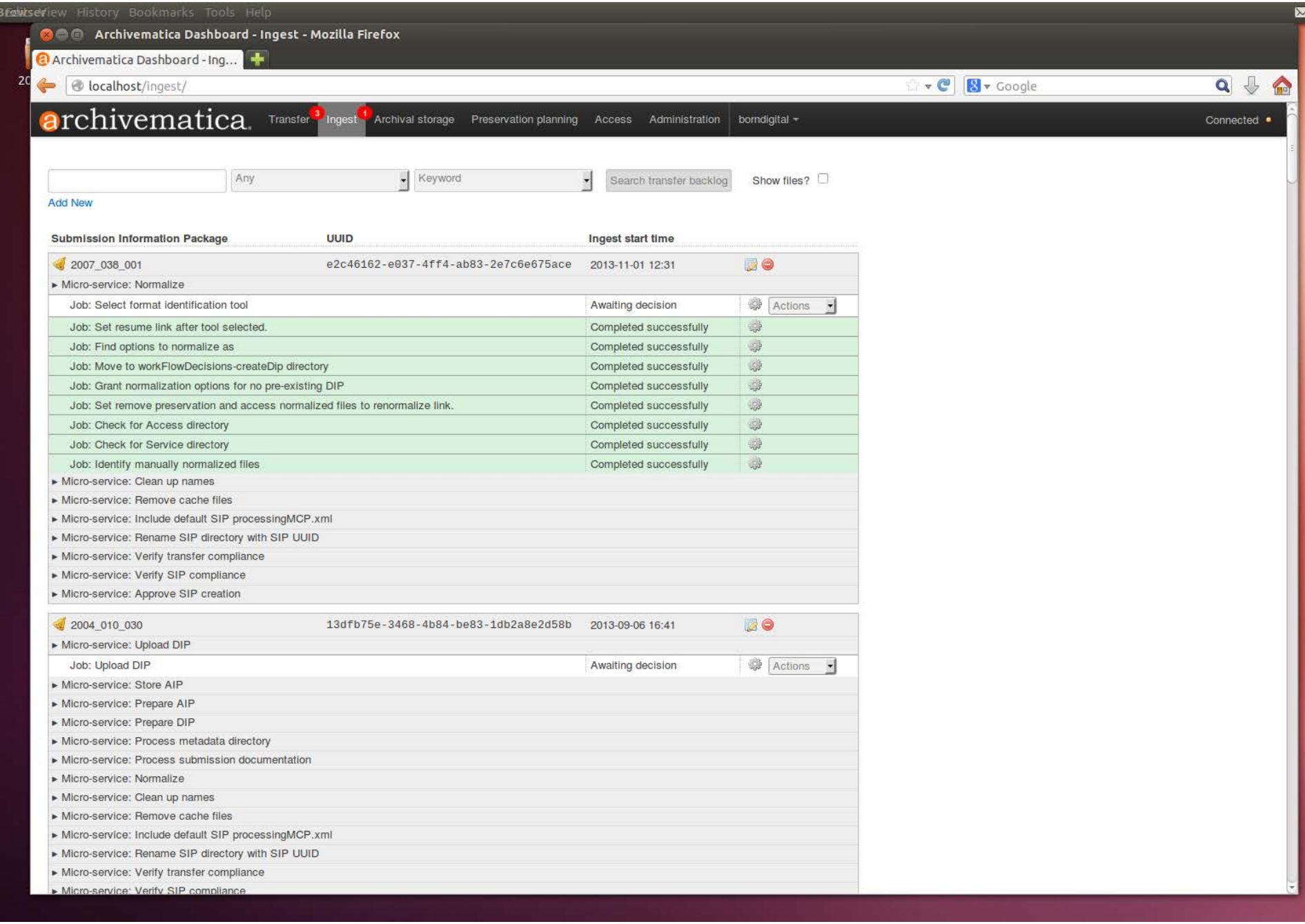
Archivematica

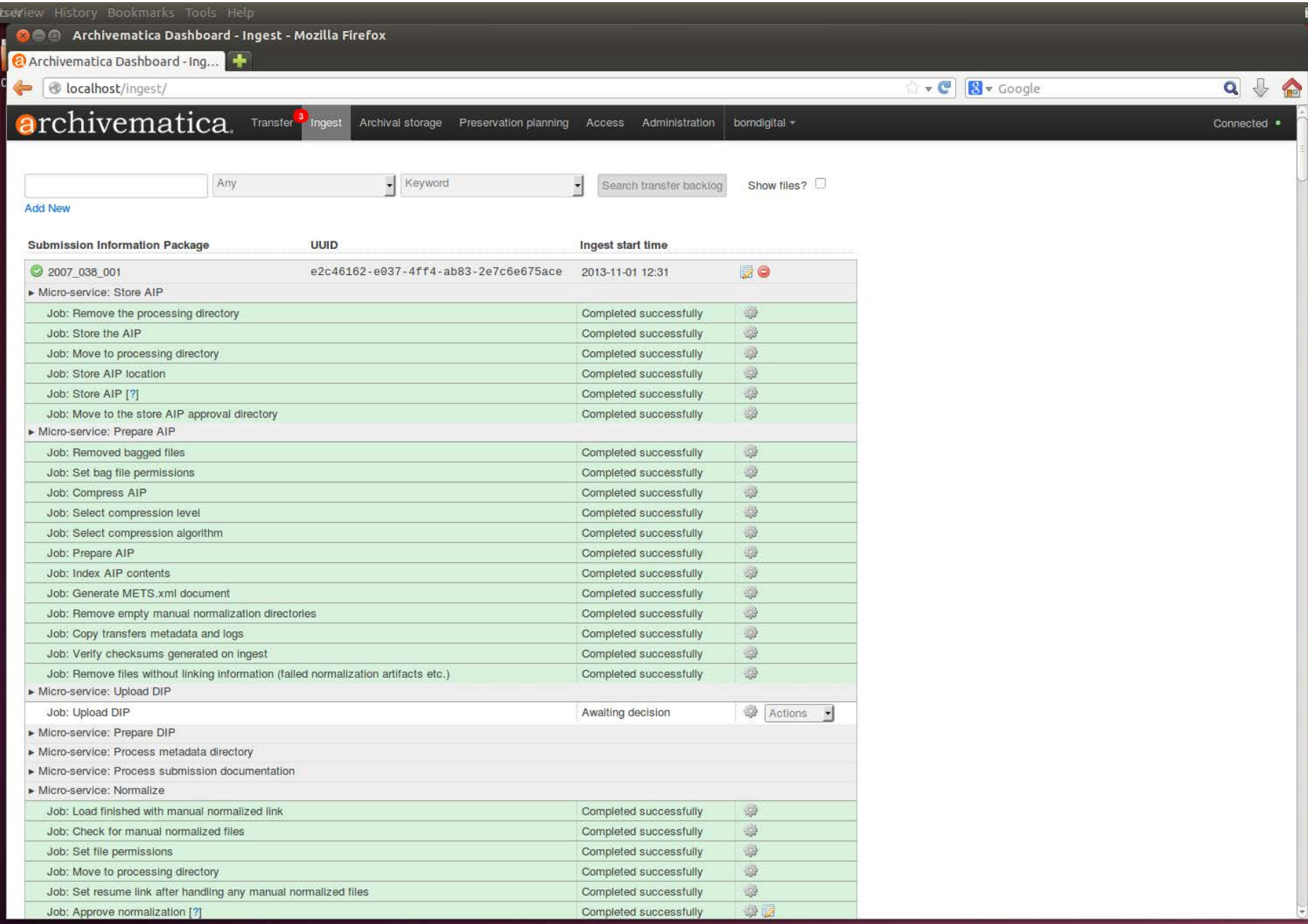
archivematica®

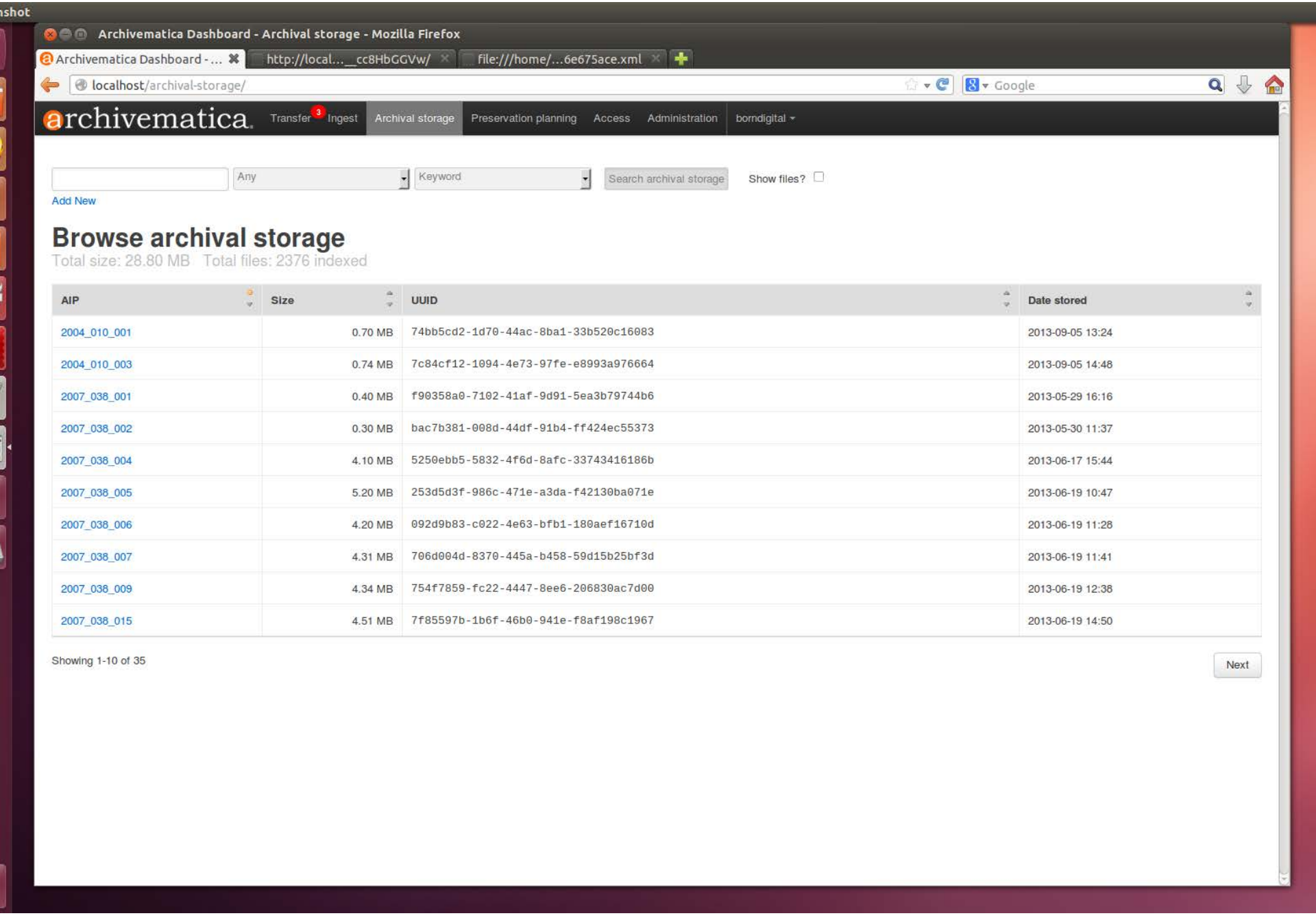
“a free and open-source digital preservation system
that is designed to maintain standards-based,
long-term access
to collections of digital objects”


Archivematica Dashboard - Transfer - Mozilla Firefox
Archivematica Dashboard - Tr...
localhost/transfer/#
Google
archivematica
Transfer
Ingest
Archival storage
Preservation planning
Access
Administration
borndigital
Connected
Standard
Type
Transfer name
Accession no.
/home/borndigital/shareddisk/born
Browse
Start transfer
Transfer
UUID
Transfer start time
2007_038_001
2dbbc8f8-c018-4c9d-b03e-cfbd4c85ab31
2013-11-01 12:27
Micro-service: Characterize and extract metadata
Job: Characterize and extract metadata
Executing command(s)
Micro-service: Clean up names
Job: Sanitize Transfer name
Completed successfully
Job: Sanitize object's file and directory names
Completed successfully
Micro-service: Scan for viruses
Job: Scan for viruses
Completed successfully
Micro-service: Extract packages
Micro-service: Quarantine
Micro-service: Generate METS.xml document
Micro-service: Verify transfer checksums
Micro-service: Assign file UUIDs and checksums
Micro-service: Include default Transfer processingMCP.xml
Micro-service: Rename with transfer UUID
Micro-service: Verify transfer compliance
Micro-service: Approve transfer
Job: Approve standard transfer
Completed successfully
2004_010_001
33f13af2-4370-45c2-9cdb-2a418ab8847d
2013-09-05 13:08
Micro-service: Create SIP from Transfer
Micro-service: Complete transfer
Micro-service: Characterize and extract metadata
Micro-service: Clean up names
Micro-service: Scan for viruses
Micro-service: Extract packages
Micro-service: Quarantine
Micro-service: Generate METS.xml document
Micro-service: Verify transfer checksums
Micro-service: Assign file UUIDs and checksums
Micro-service: Include default Transfer processingMCP.xml
Micro-service: Rename with transfer UUID
Micro-service: Verify transfer compliance


Edit View History Bookmarks Tools Help
Archivematica Dashboard - Transfer - Mozilla Firefox
Archivematica Dashboard - Tr...
localhost/transfer/#
Google
archivematica
Transfer 4
Ingest
Archival storage
Preservation planning
Access
Administration
borndigital
Connected
Standard
Type
Transfer name
Accession no.
/home/borndigital/shareddisk/born
Browse
Start transfer
Transfer
UUID
Transfer start time
2007_038_001
2dbbc8f8-c018-4c9d-b03e-cfbd4c85ab31
2013-11-01 12:27
Micro-service: Create SIP from Transfer
Job: Create SIP(s) [?]
Awaiting decision
Actions
Job: Load options to create SIPs
Completed successfully
Job: Check transfer directory for objects
Completed successfully
Micro-service: Complete transfer
Micro-service: Characterize and extract metadata
Job: Load labels from metadata/file_labels.csv
Completed successfully
Job: Identify Files ByExtension
Completed successfully
Job: Characterize and extract metadata
Completed successfully
Micro-service: Clean up names
Job: Sanitize Transfer name
Completed successfully
Job: Sanitize object's file and directory names
Completed successfully
Micro-service: Scan for viruses
Job: Scan for viruses
Completed successfully
Micro-service: Extract packages
Micro-service: Quarantine
Micro-service: Generate METS.xml document
Micro-service: Verify transfer checksums
Micro-service: Assign file UUIDs and checksums
Micro-service: Include default Transfer processingMCP.xml
Micro-service: Rename with transfer UUID
Micro-service: Verify transfer compliance
Micro-service: Approve transfer
Job: Approve standard transfer
Completed successfully
2004_010_001
33f13af2-4370-45c2-9cdb-2a418ab8847d
2013-09-05 13:08
Micro-service: Create SIP from Transfer
Micro-service: Complete transfer
Micro-service: Characterize and extract metadata
Micro-service: Clean up names
Micro-service: Scan for viruses
Micro-service: Extract packages


Archivematica Dashboard - Ingest - Mozilla Firefox
Archivematica Dashboard - Ing...
localhost/ingest/
Google
archivematica
Transfer 3
Ingest 1
Archival storage
Preservation planning
Access
Administration
borndigital
Connected
Any
Keyword
Search transfer backlog
Show files?
Add New
Submission Information Package
UUID
Ingest start time
2007_038_001
e2c46162-e037-4ff4-ab83-2e7c6e675ace
2013-11-01 12:31
Micro-service: Normalize
Job: Select format identification tool
Awaiting decision
Actions
Job: Set resume link after tool selected.
Completed successfully
Job: Find options to normalize as
Completed successfully
Job: Move to workFlowDecisions-createDip directory
Completed successfully
Job: Grant normalization options for no pre-existing DIP
Completed successfully
Job: Set remove preservation and access normalized files to renormalize link.
Completed successfully
Job: Check for Access directory
Completed successfully
Job: Check for Service directory
Completed successfully
Job: Identify manually normalized files
Completed successfully
Micro-service: Clean up names
Micro-service: Remove cache files
Micro-service: Include default SIP processingMCP.xml
Micro-service: Rename SIP directory with SIP UUID
Micro-service: Verify transfer compliance
Micro-service: Verify SIP compliance
Micro-service: Approve SIP creation
2004_010_030
13dfb75e-3468-4b84-be83-1db2a8e2d58b
2013-09-06 16:41
Micro-service: Upload DIP
Job: Upload DIP
Awaiting decision
Actions
Micro-service: Store AIP
Micro-service: Prepare AIP
Micro-service: Prepare DIP
Micro-service: Process metadata directory
Micro-service: Process submission documentation
Micro-service: Normalize
Micro-service: Clean up names
Micro-service: Remove cache files
Micro-service: Include default SIP processingMCP.xml
Micro-service: Rename SIP directory with SIP UUID
Micro-service: Verify transfer compliance
Micro-service: Verify SIP compliance


View History Bookmarks Tools Help
Archivematica Dashboard - Ingest - Mozilla Firefox
Archivematica Dashboard - Ing...
localhost/ingest/
Google
archivematica
Transfer 3
Ingest
Archival storage
Preservation planning
Access
Administration
borndigital
Connected
Any
Keyword
Search transfer backlog
Show files?
Add New
Submission Information Package
UUID
Ingest start time
2007_038_001
e2c46162-e037-4ff4-ab83-2e7c6e675ace
2013-11-01 12:31
Micro-service: Store AIP
Job: Remove the processing directory
Completed successfully
Job: Store the AIP
Completed successfully
Job: Move to processing directory
Completed successfully
Job: Store AIP location
Completed successfully
Job: Store AIP [?]
Completed successfully
Job: Move to the store AIP approval directory
Completed successfully
Micro-service: Prepare AIP
Job: Removed bagged files
Completed successfully
Job: Set bag file permissions
Completed successfully
Job: Compress AIP
Completed successfully
Job: Select compression level
Completed successfully
Job: Select compression algorithm
Completed successfully
Job: Prepare AIP
Completed successfully
Job: Index AIP contents
Completed successfully
Job: Generate METS.xml document
Completed successfully
Job: Remove empty manual normalization directories
Completed successfully
Job: Copy transfers metadata and logs
Completed successfully
Job: Verify checksums generated on ingest
Completed successfully
Job: Remove files without linking information (failed normalization artifacts etc.)
Completed successfully
Micro-service: Upload DIP
Job: Upload DIP
Awaiting decision
Actions
Micro-service: Prepare DIP
Micro-service: Process metadata directory
Micro-service: Process submission documentation
Micro-service: Normalize
Job: Load finished with manual normalized link
Completed successfully
Job: Check for manual normalized files
Completed successfully
Job: Set file permissions
Completed successfully
Job: Move to processing directory
Completed successfully
Job: Set resume link after handling any manual normalized files
Completed successfully
Job: Approve normalization [?]
Completed successfully


Archivematica Dashboard - Archival storage - Mozilla Firefox
Archivematica Dashboard - ...
http://local..._cc8HbGGVw/
file:///home/...6e675ace.xml
localhost/archival-storage/
Google
archivematica
Transfer
3
Ingest
Archival storage
Preservation planning
Access
Administration
borndigital
Any
Keyword
Search archival storage
Show files?
Add New
Browse archival storage
Total size: 28.80 MB Total files: 2376 indexed
AIP
Size
UUID
Date stored
2004_010_001 0.70 MB 74bb5cd2-1d70-44ac-8ba1-33b520c16083 2013-09-05 13:24
2004_010_003 0.74 MB 7c84cf12-1094-4e73-97fe-e8993a976664 2013-09-05 14:48
2007_038_001 0.40 MB f90358a0-7102-41af-9d91-5ea3b79744b6 2013-05-29 16:16
2007_038_002 0.30 MB bac7b381-008d-44df-91b4-ff424ec55373 2013-05-30 11:37
2007_038_004 4.10 MB 5250ebb5-5832-4f6d-8afc-33743416186b 2013-06-17 15:44
2007_038_005 5.20 MB 253d5d3f-986c-471e-a3da-f42130ba071e 2013-06-19 10:47
2007_038_006 4.20 MB 092d9b83-c022-4e63-bfb1-180aef16710d 2013-06-19 11:28
2007_038_007 4.31 MB 706d004d-8370-445a-b458-59d15b25bf3d 2013-06-19 11:41
2007_038_009 4.34 MB 754f7859-fc22-4447-8ee6-206830ac7d00 2013-06-19 12:38
2007_038_015 4.51 MB 7f85597b-1b6f-46b0-941e-f8af198c1967 2013-06-19 14:50
Showing 1-10 of 35
Next

Web Browser

Mozilla Firefox

Archivematica Dashboard - ... | http://local..._cc8HbGGVw/ | file:///home/...6e675ace.xml

file:///home/borndigital/.cache/.fr-EmMcTq/2007_038_001-e2c46162-e037-4ff4-ab83-2e7c6e675ace/data/METS.e2c46162-e037-4ff4-ab83-2e7c6

Google

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


```
−<mets xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/version18/mets.xsd">
  <metsHdr CREATEDATE="2013-11-01T18:36:46"/>
 −<amdSec ID="amdSec_1">
  −<techMD ID="techMD_1">
   −<mdWrap MDTYPE="PREMIS:OBJECT">
    −<xmlData>
     −<object xsi:type="file" xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd" version="2.2">
      −<objectIdentifier>
        <objectIdentifierType>UUID</objectIdentifierType>
        <objectIdentifierValue>86660b58-459f-47e5-a024-f06bf2d8ed82</objectIdentifierValue>
       </objectIdentifier>
      −<objectCharacteristics>
        <compositionLevel>0</compositionLevel>
       −<fixity>
          <messageDigestAlgorithm>sha256</messageDigestAlgorithm>
         −<messageDigest>
            db800ae1daeff77c250516f9c72ff4a501be1c6bb59ddaa1f7f3b3e8fdf821da
          </messageDigest>
        </fixity>
        <size>1528</size>
       −<format>
         −<formatDesignation>
            <formatName>Comma Separated Values</formatName>
            <formatVersion/>
          </formatDesignation>
         −<formatRegistry>
            <formatRegistryName>PRONOM</formatRegistryName>
            <formatRegistryKey>x-fmt/18</formatRegistryKey>
          </formatRegistry>
        </format>
       −<objectCharacteristicsExtension>
         −<fits xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version="0.6.1"
          timestamp="11/1/13 12:29 PM">
           −<identification status="CONFLICT">
             −<identity format="Plain text" mimetype="text/plain" toolname="FITS" toolversion="0.6.1">
                <tool toolname="file utility" toolversion="5.09"/>
              </identity>
             −<identity format="MPEG 1/2 Audio Layer 3" mimetype="audio/mpeg" toolname="FITS" toolversion="0.6.1">
                <tool toolname="Exiftool" toolversion="7.74"/>
              </identity>
             −<identity format="Comma Separated Values" mimetype="text/csv" toolname="FITS" toolversion="0.6.1">
                <tool toolname="Droid" toolversion="3.0"/>
```

Produce AIP

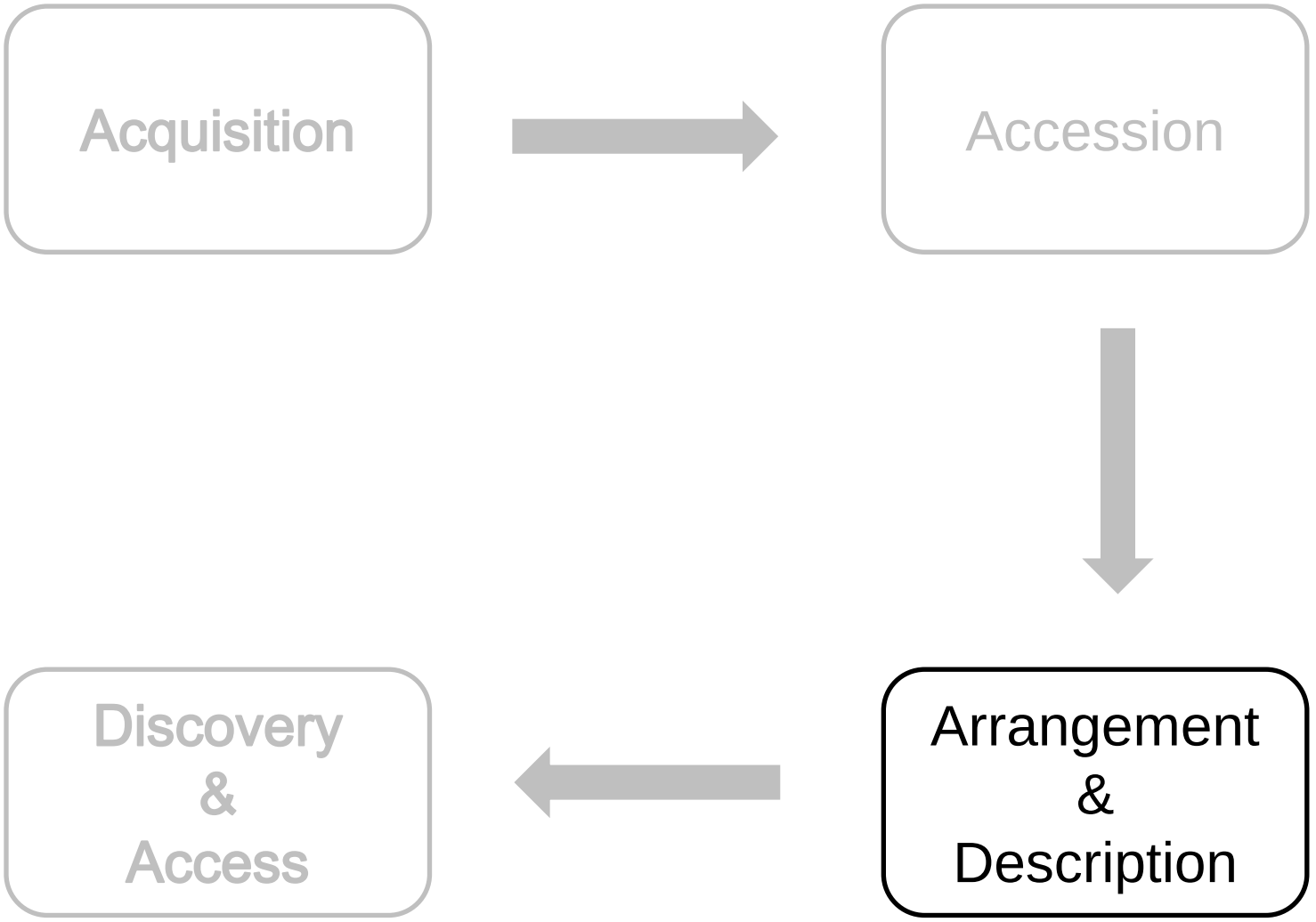
Archivematica

Current

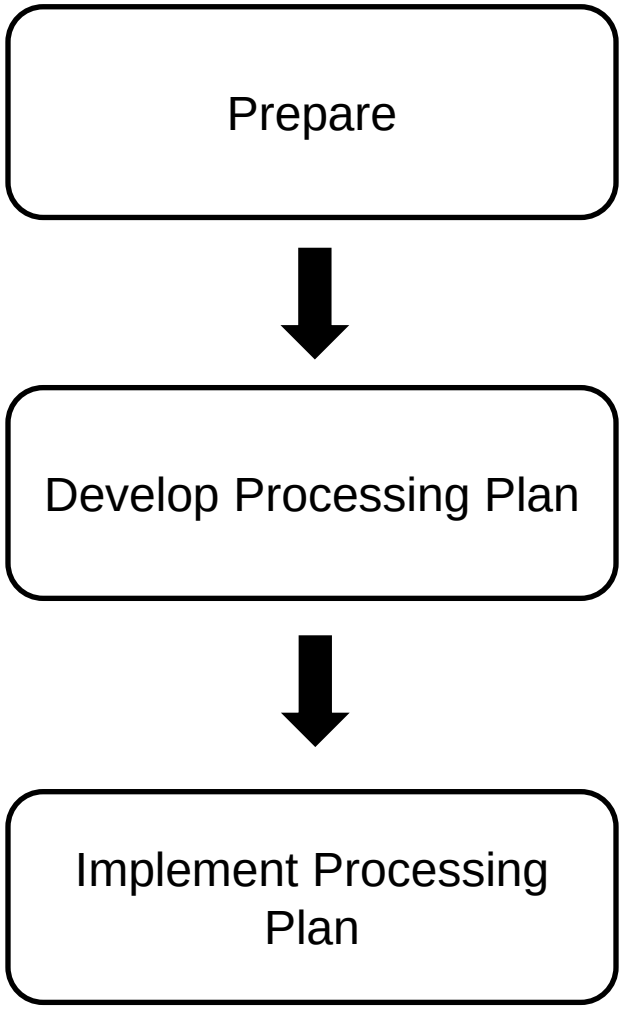
Using version 0.10 on dedicated workstation (testing as virtual server)

Future

Install version 1.0 on server with multiple client nodes (workstations)


Acquisition
Accession
Arrangement
&
Description
Discovery
&
Access

A & D


Prepare
Develop Processing Plan
Implement Processing Plan

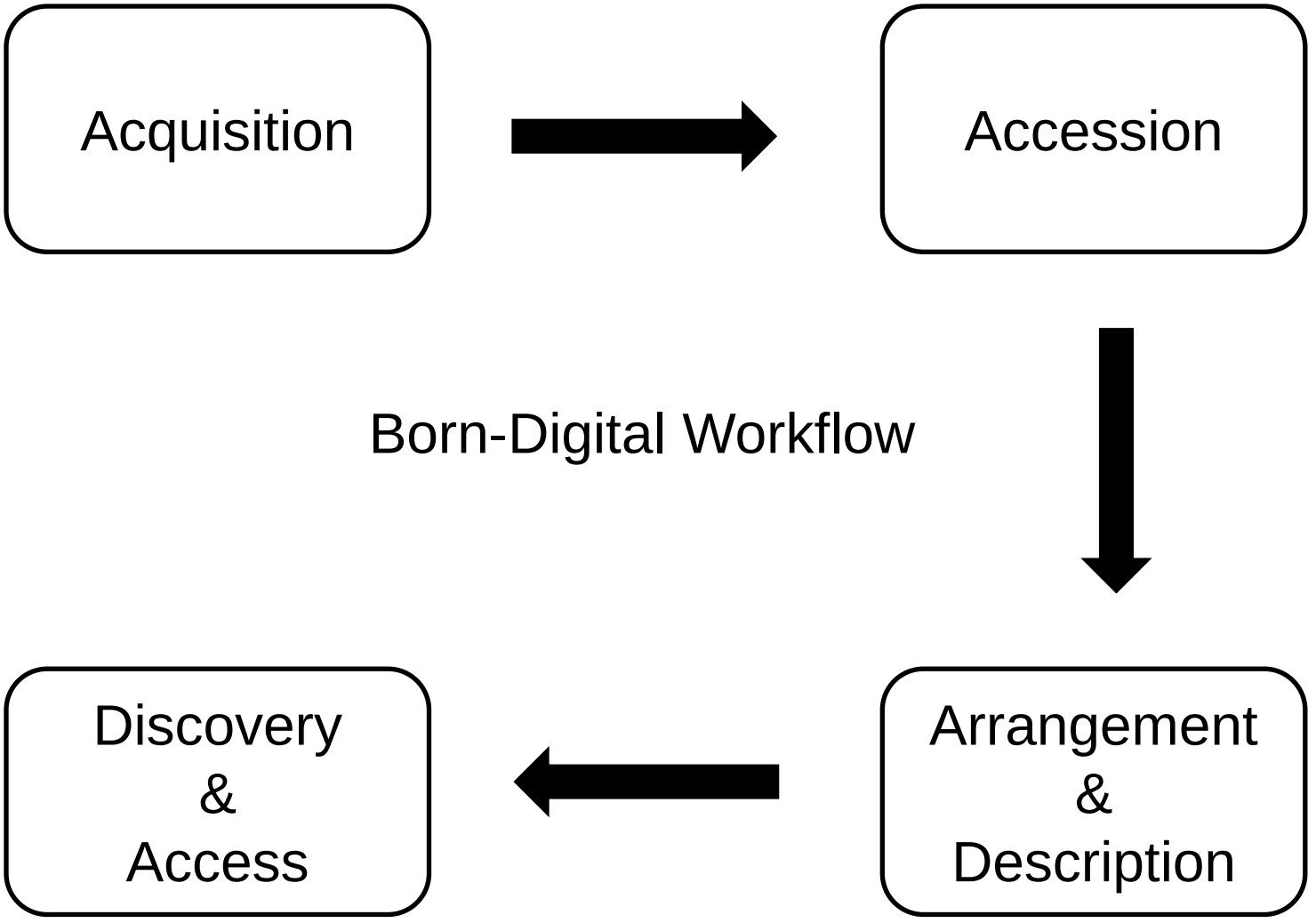
A & D

Current

- Integrate Born Digital materials into existing A&D process / tools (mix of Excel, Word, XMetal XML editor)

Future

- Determine tools needed for reviewing content (data visualization)
- Integrate Born Digital materials into collection management system


Acquisition
Accession
Born-Digital Workflow
Arrangement
&
Description
Discovery
&
Access

Lessons Learned

- Embrace iterative approach (use what you have and get what you need when you need it)
- Capture as much metadata as possible (descriptive, structural, administrative)
- Start with workflow requirements (what needs to be done) then test tools (what things will get it done)
- Build flexibility into system (may not always be ideal scenarios)

Open Source - Issues

- May require specific IT environment (Linux)
- Tools likely to change quickly
- User interfaces / experience may be simple
- Will need ongoing support from IT / Systems staff

Open Source - Benefits

- Limited initial resources needed to install and test
- Provides opportunity to engage systems / IT in new areas
- Designed and developed in collaboration with archival community
- Direct communication channels to contribute to / modify development roadmap
- Quickly build initial standards-compliant workflow

Resources

FC5205 Disk Image
http://www.deviceside.com/fc5025.html

Kryoflux
http://www.kryoflux.com/

BitCurator
http://www.bitcurator.net/

Archivematica
https://www.archivematica.org/wiki/Main_Page

Thanks!

sam.meister@mso.umt.edu
@samalanmeister