### **BitCurator at the Rubenstein**



Image: General Library Exterior, before 1948. Source: Duke University Archives. License: cc-by-nc-sa.

### Outline

- Institutional context
- Digital forensics: the (oh so) brief version
- BitCurator overview
- BitCurator at Duke



Image: "Chalk" outline, 2009. Photographer, Erica Cherup. Source <u>Flickr</u>. License: <u>cc-by-nd</u>



#### Who we are and what we collect



Image: General Library Exterior, before 1948. Source: Duke University Archives. License: cc-by-nc-sa.

## **Digital Forensics, inadequately**



Image: Basic Forensic Imaging Kit. John Crel, photographer. Source: Flickr. License: cc-by-nd.



### Boring, definition part I

- Academic definition of *digital forensics*:
  - "The use of scientifically derived methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence for the purpose of reconstructing events found to be criminal."
- Highlights
  - Criminal and legal audiences
  - Many digital preservation concepts

Image: Image from Page 41 of DEC Terminal Manual, 1982. Source: Internet Archive, <u>Flickr Commons</u>.



### Boring definition part II

- Disk Image
- Filesystem v.
   Operating System
- Bit stream
- Hex code
- Fixity

Image: Image from Page 41 of DEC Terminal Manual, 1982. Source: Internet Archive, <u>Flickr Commons</u>.

#### What I talk about when I talk about digital forensics



Mapping forensic methods to archival practice

Image by Jorgen Stamp. Source: Digital Bevaring. License: cc-by-2.5-dk.

## Acquisition: physical control

- Capture materials at lowest possible level
- Original order of files on media
- Establish fingerprint, monitor over time
- Document physical context
- Create access and preservation copies

Images: (top) Porter Olsen, photographer. Source: <u>BitCurator Facebook Group</u>. (bottom) SA175 Port Lincoln Library digitisation project, 2013. Source: <u>Flickr</u>. License: <u>cc-by-nc-nd-</u> <u>2.0</u>.





## **Acquisition: intellectual control**

- Be aware of time bombs: scan for SEI, PII, viruses at scale
- Generate lists of contents, file extensions
- Extract filesystemlevel metadata
- Document internal context





## Appraisal

- www.digitalbevaring.dk
- Mount, view contents from obsolete media, systems
  - Fuzzy hashing techniques to help address duplication

Image by Jorgen Stamp. Source: <u>Digital Bevaring</u>. License: <u>cc-by-2.5-dk</u>.



## Arrangement, Description

- Abstract groups of digital files for description
- Supports MPLP-like
   arrangement
- Export contents for more complex arrangements
- Granular filesystem metadata for creation, modified dates

Image: IBM/Tabulating Machine Co. Organization Chart. Photographer, Marcin Wichary. Source: <u>Flickr</u>. License: <u>cc-by-2.0</u>.



#### Access

- Mounting disk images in the reading room
- Document creator's computer environment
- Experiments with browsing disk image via browser
- Facilitates experiments with emulation

Image: Places of Invention, National Museum of American History. Photographer, Blake Patterson. Source: <u>Flickr</u>. License: <u>cc-by-2.0</u>.

### Where to learn more about DF

- SAA Digital Forensics workshops
  - Fundamentals
  - Advanced
- Gengenbach, Martin J. (2012) The way we do it here: Mapping digital forensics workflows in collecting institutions. Masters paper, UNC-Chapel Hill. <u>Link</u>
- CLIR report
  - <u>Digital Forensics and Born-Digital Content in Cultural Heritage</u> <u>Repositories</u> (2010)
  - Born-Digital: Guidelines for Donors, Dealers, and Archival <u>Repositories</u> (2013)
- BitCurator white papers
  - From Bitstreams to Heritage (2013)
  - From Code to Community (2014)

#### **BitCurator Overview**



Image: General Library Exterior, before 1948. Source: Duke University Archives. License: <u>cc-by-nc-sa</u>.

#### **Desktop Overview**



### **Imaging Tools**

😸 – 🗉 Imaging Tools						
< > A Home Deskt	op Imaging Tools				Q ≡ :::	
Places Ø Recent	۶	۶	۶	۶		
A Home	cdrdao (command line)	dcfldd (command line)	dd (command line)	ddrescue (command line)		
☐ Documents ❖ Downloads ✔ Music	Guymager					
☑ Pictures 目 Videos 圓 Trash						
Devices						
🖻 Electronic Mat 🔺 🖿 sdað						
Sda5 Computer						
Bookmarks						
Network Browse Network Connect to Server						
fmount work Ele	ectronic Materials					



Electronic Mater



**Forensics Tools** 





C Advanced foren	sic image (file extension .aff)	Split size 20 GiB						
Case number	2012-0057 Accession number							
Evidence number	RL10152-OP-0001 Media Identifier							
Examiner	Matthew Farrell Your name Label: "Personal documents" Label transcription if available							
Description								
Notes	Acquired 2015.06.24 09:22 SN: Z2A9TNHD							
Click to b Image directory	rowse to the accession folder //media/bcadmin/							
Image filename (w	ithout extension) Media identifier without h	yphens						
Image filename (w Info filename (with	nithout extension) Media identifier without h	yphens						
Image filename (w Info filename (with Hash calculation / 1	verification	yphens						
Image filename (w Info filename (with Hash calculation / 1 IT Calculate MD5	verification	yphens ✓ Calculate SHA-256						
Image filename (w Info filename (with Hash calculation / 1 Calculate MD5	ithout extension) Media identifier without h nout extension) verification Calculate SHA-1	yphens ✓ Calculate SHA-256						

## Guymager example I

Create image dialogue

#### **Guymager Example II**

Serial nr.	Linux device	Model	State	Size	Hidden areas	Bad sectors	Progress	Average speed [MB/s]	Time remaining	FIFO qu usag [%]
CN085KRY7363915M53TC	/dev/sr0	PLDS PLDS DVD+/-RW DH-16ABS	Finished	27.5MB	unknown	0	100%	1.38		
Z2A9TNHD	/dev/sda	ATA ST3500413AS	() Idle	500.1GB	none					
	/dev/loop0	Linux Loop: RL01262_strickla	() Idle	31.4MB	unknown					
	/dev/loop1	Linux Loop: RL01262_strickla	() Idle	31.4MB	unknown					
	/dev/loop2	Linux Loop: sda1	() Idle	322.1GB	unknown					
	/dev/loop3	Linux Loop: sr0	() Idle	27.5MB	unknown					
WD-WCATR7940902	/dev/sdb	ATA WDC WD10EALX-759BA1	() Idle	1.0TB	none					

 Size
 27,498,496 bytes (26.2MiB / 27.5MB)

 Sector size
 2,048

 Image file
 /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262OP0001.dd

 Info file
 /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262OP0001.info

 Current speed
 15. May 13:56:51 (00:00:19)

 Hash calculation
 SHA-256

 Source verification
 off

#### View attached devices

#### **Progress updates**

Þ

#### **Forensics Tools**





## bulk\_extractor and Viewer

Required Parameters		Scanners
Scan: 🖲 Image File 🔿 Raw Device	e 🔿 Directory of Files	Dase16
Image file Edit these /media/	bcadmin/Electronic Materia	🗌 Facebook
Output Feature Directory ons/201	1-1040/D04515505P beout	🗌 hashdb
		🗌 outlook
General Options		🗌 sceadan
Use Banner File		wordlist
🗌 Use Alert List File		I XOF
🗌 Use Stop List File		accts
Use Find Recey Text File		aes Absoci
		Dase04
		email
		exif
Tuning Parameters		Find
🗌 Use Context Window Size		gps
🔲 Use Page Size	167-216	🛃 gzip
🗌 Use Margin Size	4194304	M hiberfile
Use Block Size		M httplogs
Use Number of Threads		🗹 json
		🗹 kml
		msxml
		M net
Parallelizing		
Use start processing at offset		Solite
Use process range offset 01-02		vcard
Use add offset to reported feature	ure offsets	windirs
		🗹 winlnk
		🗹 winpe
Start on Page Number		👿 winprefetch
Use Debug Mode Number		📝 zip
Erase Output Directory		Do not change the
Scanner Controls		
Мараде Оцеце	Import Submit Run C	ancel

caption

#### bulk\_extractor's output

File Edit View Bookmarks Tools Help

redeate ricer introduce	Image File RE0119005B0001.E01
beout	Feature File contxt
	Forensic Path 2491174257
togram.txt Histogram.File ccn histogram.txt	Feature 4305872400574147
.txt n=4	Inner
histogram.tx	dge  2/d0127040_1e##############################
- <t< td=""><td>2491171904 #.##0\).""\$"#.##0.00 ):\("\$"#.##0.00\)"."."."."*"#.##0.0</td></t<>	2491171904 #.##0\).""\$"#.##0.00 ):\("\$"#.##0.00\)"."."."."*"#.##0.0
main histog	2491171968 0 ); [Red] \("\$"#,##0.00\)7.*.2 ("\$"* #,##0 ); ("\$"* \(#,##0\)
togram.txt	2491172032 ;_("\$"* "-"_);_(@_),,}.),(* #,##0_);_(* \(\\\\\\\\\\(#,##0\));_(* "-"_);
	2491172096 _(@_)?,,,("\$"* #,##0.00_);("\$"* \(#,##0.00\);_("\$"* "-"??
cogram tyt	2491172160 _); _(0_)6.+.1[* #,##0.00_); _(* \(#,##0.00\); _(* "-"??_); _(0
ogramitere	2491172224
adityt	2491172200
EditAt	2491172416
	2491172480
Fort	2491172544
	2491172608
nistogram	2491172672
	249117/30
amitxt	2491172864 "@
s.txt	2491172928, b. @
	2491172992
Referenced Feature File con.txt	2491173056
Referenced Feature	2491173120
And the second se	24911/3184 CheckingCheckingCredit CardBreakdown
	2491173210 ) = *
	2491173376
	2491173440T
	2491173504 HardwareFerguson's AutomotiveLaValley'sMaynard Auto Sup
	2491173568 plyKibby EquipmentMerriam-Graves CorporationJoe's Equip
	2491173632 mentLF ProttierU.S. Postal ServiceRichard's AutoPami
	2491173090 (j bottarCumbertanu FarmasBJ sRume DeputWadMartSau 2401173760 (j & Sone dutozone llokonym Huherts Sonetsho Center Pi
	2491173824 ke Industries. Keith Jones, Carquest, The Bental Center, Tow
	2491173888 n Line EquipmentNorthern NurseriesSt. PierreLambert Sup
	2491173952 plyWoodstock Home & HardwareFleury's SalesMiscellaneous
	2491174016 ExpensesEquipment RepairsBarn/Garage ExpensesEquipment
	2491174080 MaintenanceRepairsPostage and DeliveryEquipment Purcha
	2491174144 sesEquipment SuppliesInspectionsExcavating ExpensesE
	2491174200 NUTPHENE RELATIONSCAPING EXPENSES, CHases,
	2491174272 Theorem Control and the second se

```
BitCurator-0.1.5 [Run
```

irefox

file:///media/...s/mythumb.xml

Interpretation of the state of the state

#### </hleobject>

-<fileobject>

<filename>Russell/CoetzeeDisgracePaper.doc</filename>

<partition>1</partition>

<id>5</id>

<name\_type>r</name\_type>

<filesize>39936</filesize>

<alloc>1</alloc>

<used>1</used>

<inode>2463754</inode>

<meta type>1</meta type>

<mode>511</mode>

<nlink>1</nlink>

<uid>0</uid>

```
<gid>0</gid>
```

<mtime>2003-12-05T21:42:26Z</mtime> <atime>2012-10-21T04:00:00Z</atime>

<crtime>2009-10-07T13:51:49Z</crtime>

-<byte\_runs>

<br/>

<hashdigest type="md5">05cb2205ae4b8318199878b4828 <hashdigest type="sha1">c7cbbe0d846f1210e11c94523b7; </fileobject>

-<fileobject>

#### fiwalk

- Filesystem metadata
- Previously command-line only
- Default outputs to DFXML

#### fiwalk converted

Fiwalk-output.xml.xlsx - LibreOffice Calc

A2:AMJ2  $\forall f(x) \Sigma = 1$ 

	A	<b>B</b> 1	C	D	E Contraction		G	i ii	and the second se	and the second sec	ĸ
1	Partition	Filename	Extension	Filesize	File format	Chang	e tim Access time	Create time	Modification time	MD5 Hash	SHA1 Hash
2	1	AnaDirector/anago	dcr	2730913	data	None	2010-01-12T05:00:00	2010-01-12T20:41:05	2003-02-07T00:27:38	6b66822b3	ebc336f5a0dd24
3	1	AnaDirector/anago	dir	14154794	data	None	2010-01-12T05:00:00	2010-01-12T20:41:08	2003-02-07T00:27:32	f322478ce2	b47ab99b19320
4	1	AnaDirector/anago	htm	1192	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:41:23	2003-02-07T00:27:42	c28cca6ebo	84770288d1275
5	1	AnaDirector/anago	dcr	2753813	data	None	2010-01-12T05:00:00	2010-01-12T20:41:23	2003-02-07T00:29:24	be3c487bat	28d7398382818
6	1	AnaDirector/anago	dir	14178208	data	None	2010-01-12T05:00:00	2010-01-12T20:41:26	2003-02-07T00:45:00	a8ff3b9c85	717d887f286ac1
7	1	AnaDirector/anago	htm	1196	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:41:41	2003-02-07T00:29:26	2793140df	19d565b84c3b9
8	1	AnaDirector/anaop	dir	9619377	data	None	2010-01-12T05:00:00	2010-01-12T20:41:41	2002-10-18T02:06:16	3b39ef8fb2	b7529e340bf1ec
9	1	AnaDirector/anarp	dir	10260380	data	None	2010-01-12T05:00:00	2010-01-12T20:41:51	2002-10-19T01:27:04	72199262e	76ff02f6c344053
10	1	AnaDirector/anaru	dcr	1706080	data	None	2010-01-12T05:00:00	2010-01-12T20:42:01	2002-10-17T14:39:04	2afa3bd9e8	fb9f06c16eb7ebf
11	1	AnaDirector/anaru	dir	9619357	data	None	2010-01-12T05:00:00	2010-01-12T20:42:04	2002-10-18T02:06:02	1c4381b98	92568b7fd50da2
12	1	AnaDirector/anaru	htm	1086	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:42:14	2002-10-17T14:39:06	84d74f7734	+b4e95905e9de3
13	1	AnaDirector/chalk.	dcr	2718297	data	None	2010-01-12T05:00:00	2010-01-12T20:42:14	2002-10-17T20:11:02	6540ac063	a4d3211c906f7d
14	1	AnaDirector/chalk.	dir	13734436	data	None	2010-01-12T05:00:00	2010-01-12T20:42:17	2002-10-17T20:31:40	c758c1518	eb74f008acf1ac5
15	1	AnaDirector/chalk.	htm	1082	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:42:31	2002-10-17T20:16:38	836ff48e2fc	4423e782e0e329
16	1	AnaDirector/chalk	dcr	2803463	data	None	2010-01-12T05:00:00	2010-01-12T20:42:31	2003-02-07T00:33:30	129a5d707	8b72986b1aa89
17	1	AnaDirector/chalk	dir	13822160	data	None	2010-01-12T05:00:00	2010-01-12T20:42:35	2003-02-07T00:33:22	7958e1dea	d9c44fc6635e11
18	1	AnaDirector/chalk	htm	1176	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:42:49	2003-02-07T00:33:32	bda098af08	fe2513a9dadd13
19	1	AnaDirector/chalk	dir	13546361	data	None	2010-01-12T05:00:00	2010-01-12T20:42:49	2002-10-17T20:10:22	5784744a0	da630823eb1bcc
20	1	AnaDirector/colort	dir	1104746	data	None	2010-01-12T05:00:00	2010-01-12T20:43:03	2002-10-17T20:41:22	7f41b125fa	1f0e6dc0b17fa0e
21	1	AnaDirector/playw	dir	14198190	data	None	2010-01-12T05:00:00	2010-01-12T20:43:05	2003-02-07T01:18:26	3e574552a	e5979f4a8257ad
22	1	AnaDirector/playw	dir	14203138	data	None	2010-01-12T05:00:00	2010-01-12T20:43:20	2003-02-07T02:28:44	00b626213	ea80f53223287b
23	1	AnaDirector/Thum	db	5120	Composite l	None	2010-01-12T05:00:00	2010-01-12T20:43:35	2007-02-19T19:37:18	03290331d	8243db60e5979
24	1	AnaDirector/tryone	dcr	2108793	data	None	2010-01-12T05:00:00	2010-01-12T20:43:35	2002-10-14T20:07:52	59ccdb5b8	591700a17e02a
25	1	AnaDirector/tryone	dir	12157240	data	None	2010-01-12T05:00:00	2010-01-12T20:43:38	2002-10-16T19:24:38	3ce003764	18e28581efe541
26	1	AnaDirector/tryone	htm	1086	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:43:50	2002-10-14T20:07:54	fec04b3a16	9772f6de630aa9
27	1	AnaDirector/tryone	dir	12175350	data	None	2010-01-12T05:00:00	2010-01-12T20:43:50	2002-10-15T11:26:58	aef4bb50a9	75e999f56c8963
28	1	AnaDirector/tryone	dir	14260646	data	None	2010-01-12T05:00:00	2010-01-12T20:44:03	2002-10-15T12:16:12	5bd002b4a	65f87d21f438b6
29	1	AnaDirector/tryone	dir	14155018	data	None	2010-01-12T05:00:00	2010-01-12T20:44:18	2002-10-15T21:02:48	845001c5ct	fdb8d2916571a7
30	1	AnaDirector/tryone	dir	14024540	data	None	2010-01-12T05:00:00	2010-01-12T20:44:32	2002-10-18T00:07:28	2a3e37354	6cf66e734ae580
31	1	AnaDirector/xmba	ipg	1579	JPEG image	None	2010-01-12T05:00:00	2010-01-12T20:44:47	2002-10-15T14:38:18	42f74aecf4	a2145ccc3fe392a
32	1	AnaDirector/ana/1	GIF	6297	GIF image o	None	2010-01-12T05:00:00	2010-01-12T20:44:47	2001-08-08T19:16:40	741974bcb	085cb88028a8d
33	1	AnaDirector/ana/1	ipg	18127	JPEG image	None	2010-01-12T05:00:00	2010-01-12T20:44:47	2002-10-21T13:47:32	6328447aa	7ec088bc044855
34	1 File Object Info	AnaDirector/ana/a	htm	2071	HTML docup	None	2010-01-12T05:00:00	2010-01-12T20:44:47	2001-05-05T12:32:16	db0cdb491	b4cd95cea88fffd
Shee	t1/1		Page	Style File Object	oformation				Sum=0		0 + 1009

#### Spreadsheet transformation

### **Reporting Tool and outputs I**

Run All Fiwalk XML Annotated Features Reports

#### Run fiwalk, annotate the bulk\_extractor output, and generate Office / PDF reports.

If you haven't run bulk\_extractor yet, use the button to the right to launch and run it first.

#### Image File

 $/media/bcadmin/Electronic\ Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262OP0001.dd$ 

#### **Bulk Extractor Feature Directory**

/media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_beout

#### Output Directory (fiwalk output, annotated features, and reports will appear in here)

/media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_rep

#### Config File (Optional)

/Path/To/File

#### Command Line Output

>> Using the default config file: /etc/bitcurator/bc\_report\_config.txt

>> PREMIS Capture Event not generated: Raw or unknown image type

>> Generating XML File for the image /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262OP0001.dd >> Command Executed for Fiwalk: ['fiwalk', '-f', '-X', '/media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_rep/fiwalk-output.xml', '/media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_rep/fiwalk-output.xml', '/media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262OP0001.dd']

#### >> Image File Selected:

>> Image File Selected: /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262OP0001.dd >> Annotate: BE Features Directory Selected: /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_beout

>> Success!!! Fiwalk created the following file:

o /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_rep/fiwalk-output.xml >> Generating Premis event for Fiwalk in: /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_rep/reports

>>> Generating bulk\_extractor Premis Event

>> Creating annotated Features

>> Success!!! Annotated feature files created in the directory:

o /media/bcadmin/Electronic Materials/processing/RL01262\_strickland/writings\_series/projects/errand\_upon\_which\_we\_came/RL01262-OP-0001\_rep/annotated-features

>> Generating BitCurator Reports:

Close

Cancel

Run

\*\*\*

....

...

...

Launch BEViewer

### **Reporting Tool and outputs II**

annotated\_ccn.txt × annotated\_rfc822.txt × 1 # Position Feature Context Filename MD5 2 # /home/bcadmin/Tools/bulk\_extractor/python/identify\_filenames.py --all --image\_filename /media/bcadmin/Electronic Materials/processing/ RL01190 simmons/RL01190USB0001.E01 /media/bcadmin/Electronic Materials/processing/RL01190 simmons/USB0001 beout /media/bcadmin/Electronic Materials/ processing/RL01190 simmons/USB0001 reports/annotated-features 3 # BANNER FILE NOT PROVIDED (-b option) 4 # BULK EXTRACTOR-Version: 1.5.5-dev (\$Rev: 10844 \$) 5 # Feature-Recorder: ccn 6 # Filename: /media/bcadmin/Electronic Materials/processing/RL01190 simmons/RL01190USB0001.E01 7 # Feature-File-Version: 1.1 8 1440854999 SOPOOMIE>:9:950/ '\x1D\x12\x0D\x0D\x0E\x0E\x0D\x0C\x0E\x0F\x12\x12\x10\x0D '\x1D\x12\x0D\x0E\x0E\x0E\x0E\x0E\x0E\x0E\x12\x12\x10\x0D 9 1449268982 SOP00MTE>:9:950/ enses\x05\x00\x00Chase\x10\x00\x00 0 2491174257 x08\x00\x00Best Buy\x07\x00\x00LL SOPOOMIE>:9:950/ \x1D\x12\x0D\x0D\x0E\x0E\x0E\x0E\x0E\x0F\x0F\x12\x12\x10\x0D 286/065815 2875479798 SOPOOMIE>:9:950/ \x1D\x12\x0D\x0D\x0E\x0E\x0D\x0C\x0E\x0F\x12\x12\x10\x0D 13 # Total features input: 5 14 # Total features located to files: 0 15 # Total features in unallocated space: 5 16 # Total features in encoded regions: 0 17 # Total processing time: 0.00024 seconds

- Maps features ID'd by bulk\_extractor to specific files, where possible
- Reports which features apply to files and which do not

#### BitCurator Disk Image Access tool

#### **Reporting Tool and outputs III**

< >

1				
N	lame		Size	Туре
	W	Chapter 1 Saraband for a Saint 1-26.doc	10.6 MB	Docum
ļ	W	Chapter 2 The Gypsy Condesa 27-42.doc	1.4 MB	Docum
	W	Chapter 3 Meeting Isabel 43-60.doc	105.0 kB	Docum
	W	Chapter 4 Charleston 61-76.doc	99.8 kB	Docum
	W	Chapter 5 Saint Teresa 77-99.doc	135.7 kB	Docum
	W	Chapter 6 Chronology 100-112.doc	247.8 kB	Docum
	W	Chapter 7 The Sackvilles & Other Connections 113-152	1.4 MB	Docum
	W	Chapter 8 Orlando 153-182.doc	178.7 kB	Docum
	W	Chapter 9 Ballad of the Sad Cafe Where it Came From	704.5 kB	Docum
	W	Chapter 10 Ballad The Town The Time The People 219-2	53.8 kB	Docum
	W	Chapter 11 Ballad The Cast 228-243.doc	471.6 kB	Docum
	W	Chapter 12 Ballad A Theory of Love 244-250.doc	47.6 kB	Docum
	W	Chapter 13 Ballad The Tale 251-275.doc	429.1 kB	Docum
	W	Chapter 15 Ballad Afterwards 276-282.doc	46.6 kB	Docum
	W	Chapter 16 Ballad Sources 283-299.doc	86.0 kB	Docum
	W	Chapter 17 Ballad The Film 300-321.doc	114.2 kB	Docum
	W	Chapter 18 The Transsexual Phenomenon 322-378.doc	2.8 MB	Docum
	W	Chapter 19 The Last Book The Last Mentor 379-416.doc	427.0 kB	Docum
	W	Chapter 20 Me Papoose Sitter Revisited 417-443.doc	1.3 MB	Docum
	and the second se			

#### **Report: File System Statistics and Files**

#### **Deleted Files**

#### Disk Image: RL01190USB0001.E01

Parti	tior	Deleted File				
1	SOrphanFiles					
1	\$OrphanFiles/_01.JPG					
i	SOrphanFiles/_02.JPG					
1	5OrphanFiles/_03.IPG					
1	SOrphanFiles/_04.JPG					
ī	\$OrphanFiles/_05.JPG					
Ţ.	SOrphanFiles/_06.JPG					
1	\$OrphanFiles/_07.JPG					
i	SOrphanFiles/_08.JPG					
1	5OrphanFiles/_09.IPG					
1	SOrphanFiles/_10,JPG					
1	SOrphanFiles/_11.JPG					
I.	SOrphanFiles/_12.JPG					
1	\$OrphanFiles/_13.JPG					
i	SOrphanFiles/_14.JPG					
1	SOrphanFiles/_15.JPG					
1	SOrphanFiles/_16.JPG					
ī	\$OrphanFiles/_17.JPG					
T.	SOrphanFiles/_18.JPG					
1	\$OrphanFiles/_19.JPG					
i.	SOrphanFiles/ 20.JPG					

Q

#### **Disk Image Access tools I**

Open disk image Close disk image Select All DeSelect All Export selections Cancel export

e System: ntries in bold are directories	i Image Info
ntries in red are unallocated/deleted files	Password: N/A
Chapter 2 The Gypsy Condesa 27-42.doc	Scaples Relay UFD Serial number: 000ECC110008D227
Chapter 3 Meeting Isabel 43-60.doc	
Chapter 4 Charleston 61-76.doc	EWF information File format: EnCase 6
Chapter 5 Saint Teresa 77-99.doc	Sectors per chunk: 64
Chapter 6 Chronology 100-112.doc	Error granularity: 1
Chapter 7 The Sackvilles & Other Connections 113-152.doc	Compression level: good (Fast) compression
Chapter 8 Orlando 153-182.doc	and the second
Chapter 9 Ballad of the Sad Cafe Where it Came From 183-218.doc	Media information Media type: removable disk
Chapter 10 Ballad The Town The Time The People 219-227.doc	Is physical: yes
Chapter 11 Ballad The Cast 228-243.doc	Bytes per sector: 512 Number of sectors: 7821312
Chapter 12 Ballad A Theory of Love 244-250.doc	Media size: 3.7 GiB (4004511744 bytes)
Chapter 13 Ballad The Tale 251-275.doc	
Chapter 15 Ballad Afterwards 276-282.doc	
Chapter 16 Ballad Sources 283-299.doc	
Chapter 17 Ballad The Film 300-321.doc	Messages
Chapter 18 The Transsexual Phenomenon 322-378.doc	>> Image file selected: /media/bradmin/Electronic
Chapter 19 The Last Book The Last Mentor 379-416.doc	Materials/processing/RL01190_simmons/RL01190USB0001.E01
Chapter 20 Me Papoose Sitter Revisited 417-443.doc	Constitue DEVAN Eller
Chapter 21 Transsexual Notes 444-542.doc	/home/bcadmin/.bcfa/RL01190USB0001.E01_dfxml.xml
Chapter 22 The Negro World 543-646.doc	
Chapter 23 Meeting John Paul 647-681.doc	>> Success!!! Fiwalk created DFXML file
Chapter 24 Portrait of a Marriage 682-693.doc	>> Generating directory tree
Chapter 25 Mentors 694-735.doc	
Chapter 26 Religion 736-800.doc	>> Done
SMBR	
SFAT1	
SFAT2	
▼ \$OrphanFiles	

#### **BitCurator Disk Image Access tool**

#### **Disk Image Access tools II**

File Tools Help				
Up Extract	Info			
Path: /				🔁 Go
<ul> <li>Intergrams 2</li> </ul>	Name	Size	Туре	Date Modified
Trash	Desktop Desktop Folder igh Part2 Stuffit 1.5.1 Trash	0 393 Ki 0 0	B File B File B File B Folder	5/13/90 12:16 AM 5/31/93 1:36 AM 5/13/90 12:08 AM 5/9/90 1:34 AM 5/31/93 1:38 AM
	HFS	Explorer	•	

### **De-duplication**, file similarity

bcadmin@bcadmin-rub:/media/bcadmin/Electronic Materials/appraisal/JHFNC/JHFNC similarity gen/CD182 CD190 CD 216\$ sdhash -t 90separator csv -c RL10	.59CD182.
Jame / Jones / Joset an Joset and La Contract and La Contract / Jame / Jame / Jones / Joset and Joset and Contract and Con	ticlo b
/home/bcadmin/besktop/image_mount/dd_PL10156CP102/Welden/bave Bloom in Iceroll in /bene/bcadmin/besktop/image_mount/dd_PL10156CP102/Welden/bave Bloom in Iceroll in /bene/bcadmin/besktop/image_mount/dd_PL10156CP102/Welden/bave Bloom in Iceroll	in Teres
/long/bcadmin/besktop/image_mount/du_ktio159cD102/weldon/bave bloom in 151aet.jpg//nome/bcadmin/besktop/image_mount/du_ktio159cD216/weldon/bave bloom	
/home/bcadmin/besktop/image_mount/dd_kilds/Weldon/bave_Bloom photo.bmp,/home/bcadmin/besktop/image_mount/dd_kilds/Delba/Weldon/bave_Bloom photo.bmp,/home/bcadmin/besktop/image_mount/dd_kilds/Delba/Weldon/bave	.o. omp, 10
/nome/bcadmin/besktop/image_mount/dd_kl/0159CD182/weldon/bave & Rose Bloom.jpg,/nome/bcadmin/besktop/image_mount/dd_kl/0159CD216/weldon/bave	.oom.jpg,
/nome/bcadmin/besktop/image_mount/dd_kLi0159CD182/weldon/Eugene & Betty Bloom.jpg//nome/bcadmin/besktop/image_mount/dd_kLi0159CD216/weldon/Eugene & B	TTY BLOO
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 004.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 004.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 005.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 005.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 006.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 006.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 016a.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 016a.j	og,100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 022.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 022.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 025.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 025.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 026.jpg,/home/bcadmin/Desktop/image_mount/dd RL10159CD216/Weldon-Farber/01 026.jpg	100
<pre>/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 032.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01 032.jpg</pre>	100
/home/bcadmin/Desktop/image mount/dd RL10159CD182/Weldon-Farber/01 035.jpg,/home/bcadmin/Desktop/image mount/dd RL10159CD216/Weldon-Farber/01 035.jpg	100
/home/bcadmin/Desktop/image mount/dd RL10159CD182/Weldon-Farber/01 037.jpg,/home/bcadmin/Desktop/image mount/dd RL10159CD216/Weldon-Farber/01 037.jpg	100
/home/bcadmin/Desktop/image mount/dd RL10159CD182/Weldon-Farber/01 041.jpg,/home/bcadmin/Desktop/image mount/dd RL10159CD216/Weldon-Farber/01 041.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_042.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_042.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_043.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_043.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_044.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_044.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01 054.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/Rabbi AmyS	heinerma
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_056.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_056.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_058.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_058.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_060.jpg./home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_060.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_061.jpg./home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_061.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_063.jpg./home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_063.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01_064.jpg./home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01_064.jpg	100
/home/bcadmin/Desktop/image_mount/dd_Bl10159CD182/Weldon-Earber/01_065.jpg/home/bcadmin/Desktop/image_mount/dd_Bl10159CD216/Weldon-Earber/01_065.jpg	100
/home/bcadmin/Desktop/image_mount/dd_RL10159CD182/Weldon-Farber/01.jpg,/home/bcadmin/Desktop/image_mount/dd_RL10159CD216/Weldon-Farber/01.jpg,100	

- Hashing, fuzzy hashing techniques
- FSlint
- sdhash, ssdeep

### **Additional tools**

😸 - 🗉 Accession Tools		× - D Additional Tools						
< > Accession T	rools · Q ≡ :::	< > A Home Deskto	Additional Tools				9, ≡ :	
Places © Recent A Home Desktop Documents Documents Downloads Downloads R Husic Pictures Videos D Videos Trash Devices S 322 GB Volume S 32 GB Volume	Bagger Degit Library (BIL)	Places ② Recent ♣ Home Desktop Documents ⇒ Downloads G Music Pictures ¥ Videos ∰ Videos ∰ Trash Devices S 322 GB Volume S 32 GB Volume	Command-Line Tool Scripts FIDO Format Identification HFS Explorer	DFXML Scripts DFXML Scripts FITS File Information Read Outlook PST File	Antiword (legacy Word to TXT and PDF) CHex Recoll	ClamTk CtkHash VLC media player		

fmount work

**Electronic Materials** 



Forensics Tools





## **Decision points: VM or Native OS?**

## VM

#### Pros

- Update with each new release quickly
- Distribute amongst staff
- Cons
  - VirtualBox, VMware support issues
  - Resource sharing
  - Virtualized storage

## Native OS

- Pros
  - Run forensic processes more efficiently
  - Less virtualized storage
- Cons
  - Linux learning curve
  - Support at your institution
  - Likely that BC can't do everything you want

#### **Decision points: Which tools to use?**



### **Specific Uses at Duke**



Image: General Library Exterior, before 1948. Source: Duke University Archives. License: cc-by-nc-sa.



## Acquisitions: Imaging

- Decide per collection, Windows or BC for imaging
- Question: who is doing the imaging?
- Floppies—Windows
- Most everything else—BC

Image: So Many Thumb Drives, 2013. Photographer, Mark Wilson. Source <u>Flickr</u>. License: <u>cc-by-2.0</u>.

🕘 💷 emacs24@bcadmin-rub

Edit Options Buffers Tools Sh-Script Help

🗁 📃 🗶 🔚 I 🥱 I 🐰 🖬 🛅 I 🖎

#### //bin/<mark>bash</mark>

```
: Title : SimGenVirusScan
: Author : "Matthew Farrell" <matthew.j.farrell@duke.edu>
: Date : 4/1/2015, 5/12/2015
: Version : 0.3
: Descript : mount a directory of disk image files, print a simple contents
, run ClamTK, generate sdhash digests its contents, and unmount
: Options :
: Depends : fuseiso, xmount, sdhash
: License : GPLv3
or file in *.{iso,E01}; do
```

f [[ \$file =~ .\*[.][eE]0[1234] ]]

#### hen

```
DIR="${file%.*}"
```

```
mkdir -p /home/bcadmin/Desktop/image_mount/"$DIR"
sudo xmount --in ewf $file /home/bcadmin/Desktop/image_mount/"$DIR"
mkdir /home/bcadmin/Desktop/image_mount/dd_"$DIR"
sudo mount -t iso9660 -o loop /home/bcadmin/Desktop/image mount/"$DIR"/"$
```

.dd /home/bcadmin/Desktop/image\_mount/dd\_"\$DIR"

```
ls -R /home/bcadmin/Desktop/image_mount/dd_"$DIR"/
```

```
find /home/bcadmin/Desktop/image_mount/dd_"$DIR"/ -maxdepth 15 -iname "*.
printf "%h,%f,%CY-%Cm-%Cd,%s\n" > /home/bcadmin/Desktop/"$DIR"_contents.csv
clamacap l /home/bcadmin/Desktop/"$DIR"_contents.csv
```

```
clamscan -l /home/bcadmin/Desktop/"$DIR".txt -r /home/bcadmin/Desktop/ima
ount/dd_"$DIR"/
```

```
sdhash -r -o /home/bcadmin/Desktop/"$file" /home/bcadmin/Desktop/image_mo
dd_"$DIR"
```

```
sudo umount /home/bcadmin/Desktop/image_mount/dd_"$DIR"
```

```
sudo umount /home/bcadmin/Desktop/image_mount/"$DIR"
```

```
rmdir /home/bcadmin/Desktop/image_mount/dd_"$DIR"
```

```
rmdir /home/bcadmin/Desktop/image_mount/"$DIR"
```

```
lif [[ $file =~ .*[.](iso|ISO)\d? ]]
```

#### hen

```
DIR="${file%.*}"
```

mkdir -p /home/bcadmin/Desktop/image\_mount/"\$DIR"

fuseiso -p \$file /home/bcadmin/Desktop/image\_mount/"\$DIR"

ls -R /home/bcadmin/Desktop/image\_mount/"\$DIR"

```
clamscan -l /home/bcadmin/Desktop/"$DIR".txt -r /home/bcadmin/Desktop/ima
ount/"$DIR"/
```

```
sdhash -r -o /home/bcadmin/Desktop/"$file" /home/bcadmin/Desktop/image_mo
"$DIR"
```

-- **iso\_e01\_simgen\_v0.5** Top L15 (Shell-script[bash]) eginning of buffer

### Acquisitions: Basic Reporting

 Script to mount disk image, create directory printout, run virus scan, create fuzzy hashes

• bulk\_extractor for sensitive content

## Analysis for Arrangement and Description

- Fiwalk reports
- FITS
- Disk image mounting

tput.xml.xlsx - Lib	reOffice Calc									
	B B ABC	RBC 🔏 🗐		🧙 • 🚈	- 🔊 🕯	Z Z M H				
▼ 11		AEE		1 %						
$\star$ $f_{(2)} \Sigma = 1$										
8	l c	Ď	i i	F		G				
Filename	Extension	Filesize	File format	Change tim	Access t	ime				
AnaDirector/anac	lo dcr	2730913	data	None	2010-03	I-12T05				
AnaDirector/anac	lo dir	14154794	data	None	2010-01	L-12T05				
AnaDirector/anac	ohtm	1192	HTML docup	None	2010-0	L-12T05				
AnaDirector/anag	odcr	2753813	data	None	2010-0:	I-12T05				
AnaDirector/anac	odir	14178208	data	None	2010-01	L-12T05				
AnaDirector/anac	ohtm	1196	HTML docup	None	2010-03	L-12T05				
AnaDirector/anac	pdir	9619377	data	None	2010-03	L-12T05				
AnaDirector/anar	p∙dir	10260380	data	None	2010-03	L-12T05				
AnaDirector/anar	u∗dcr	1706080	data	None	2010-03	L-12T05				
AnaDirector/anar	u⊧dir	9619357	data	None	2010-03	L-12T05				
AnaDirector/anar	u∙htm	1086	HTML docup	None	2010-03	L-12T05				
AnaDirector/chall	⇔dcr	2718297	data	None	2010-03	L-12T05				
AnaDirector/chall	⇔dir	13734436	data	None	2010-03	L-12T05				
AnaDirector/chall	⇔htm	1082	HTML docup	None	2010-03	L-12T05				
AnaDirector/chall	¢)dcr	2803463	data	None	2010-03	L-12T05				
AnaDirector/chall	c≱dir	13822160	data	None	2010-03	L-12T05				
AnaDirector/chall	¢ htm	1176	HTML docup	None	2010-03	L-12T05				
AnaDirector/chall	k⊳dir	13546361	data	None	2010-03	L-12T05				
AnaDirector/color	t∙dir	1104746	data	None	2010-03	L-12T05				
AnaDirector/play	w∙dir	14198190	data	None	2010-03	L-12T05				
AnaDirector/play	w∍dir	14203138	data	None	2010-03	L-12T05				
AnaDirector/Thur	n∙db	5120	Composite I	None	2010-03	L-12T05				
AnaDirector/tryor	nødcr	2108793	data	None	2010-03	L-12T05				
AnaDirector/tryor	nødir	12157240	data	None	2010-03	L-12T05				
AnaDirector/tryor	nøhtm	1086	HTML docup	None	2010-03	L-12T05				
AnaDirector/tryor	nødir	12175350	data	None	2010-03	L-12T05				
AnaDirector/tryor	nødir	14260646	data	None	2010-03	L-12T05				
AnaDirector/tryor	nedir	14155018	data	None	2010-03	L-12T05				
AnaDirector/tryor	ne dir	14024540	data	None	2010-03	L-12T05				
AnaDirector/xmb	a∗ipg	1579	JPEG image	None	2010-03	L-12T05				
AnaDirector/ana/	1+GIF	6297	GIF image o	None	2010-03	L-12T05				
AnaDirector/ana/	1 ipg	18127	JPEG image	None	2010-03	L-12T05				
AnaDirector/ana/	a∙htm	2071	HTML docup	None	2010-03	L-12T05				
mation / +			and the second		1 AC					

PageStyle\_File Object Information

## Creation of Copies

- Disk image access tools to create access sets
- Export subsets for more complex arrangements

Image: Day 301: Right to the Point, 2013. Photographer, Quinn Dombrowski. Source: <u>Flickr</u>. License: <u>cc-by-sa-2.0</u>.

DO NOT COP, I ME,



#### **Future Steps**

- Further work on similarity automation
- Leverage DFXML for other systems
- Expand use to other members of TS

Image: Metadata is a love note to the future, 2012. Source: <u>Flickr</u>. License: <u>cc-by-2.0</u>.

# BitCurater consortium

#### **Current Membership**

- British Library
- Duke University
- Emory University
- Harvard University
- Maryland Institute for Technology in the Humanities
- Massachusetts Institute for Technology
- New York University
- School of Information & Library Science, UNC-CH
- Princeton University
- Stanford University
- University of Maryland Libraries
- University of Virginia
- McMaster University
- Northwestern University
- Texas State Libraries and Archives Commission
- University of Colorado Boulder
- University of Manchester
- Yale University
- Penn State University

#### **Recent News and Plans**

- August 2015: Consortium
   Manager will join
- FY2016: Formalize the BCC Help Desk
- Fall 2015: launch BCC website
  - Workflow exchange
  - Script repository
- Fall-Winter 2015: compiling, documenting workflows
- January 2016: 2<sup>nd</sup> Annual BitCurator User Forum, UNC-Chapel Hill

Matthew Farrell <u>matthew.j.farrell@duke.edu</u> @laissezfarrell



#### Thanks!