Taking Bitstreams Seriously: Digital Forensics and the BitCurator Environment

Cal Lee

School of Information and Library Science University of North Carolina, Chapel Hill

Digital Preservation 2013 July 23-25, 2013 Alexandria, Virginia



Two Main Acquisition Paths

Systematic Transfer	Dealing with Whatever you Get
Close pre-coordination between	Little pre-coordination between
Producer and Archive*	Producer and Archive
Archive has (at least some) say	Archive has relatively little say in
in how materials are produced,	how materials are produced,
packaged and transferred	packaged and transferred
Relatively little need to engage in guess work or ad hoc description after the transfer	Substantial need to engage in guess work or ad hoc description after the transfer

*Archive in the OAIS sense – can be library, archives, museum, data center, ...

Examples of Systematic Transfer Developments

- Protocols and tools for transfer (e.g. BagIt)¹
- Systematic and predictable ingest workflows²
- Substantial, well-documented Producer-Archive interactions (e.g. PAIMAS)³

1. "BagIt: Transferring Content for Digital Preservation" (Video). Library of Congress, August 6, 2009. http://www.youtube.com/watch?v=l3p3ao_JSfo

2. Glick, Kevin, and Eliot Wilczek. "Ingest Guide." Tufts University and Yale University, 2006. http://dca.lib.tufts.edu/features/nhprc/reports/3 1 draftpublic3.pdf

3. Producer-Archive Interface Methodology Abstract Standard (PAIMAS). CCSDS 651.0-M-1. Consultative Committee for Space Data Systems. May 2004.

Sometimes things are a little messier



Applying Digital Forensics to Digital Collections – Previous Work*

- Ross and Gow (1999) potential relevance of advances in data recovery and digital forensics to collecting institutions
- More recently active stream of literature related to use of forensic tools and methods for digital collections, including activities at the British Library, National Library of Australia and Indiana University
- PERPOS (Georgia Tech) has applied data capture and extraction to US presidential materials
- "Computer Forensics and Born-Digital Content in Cultural Heritage Collections" symposium and report (2010)
- Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS) framework for the stewardship of born-digital materials, including digital forensics methods
- Digital Records Forensics project has articulated connections between the concepts of digital forensics and archival science
- Two Open Planets Foundation (OPF) Hackathons this year one in Copenhagen, one in Chapel Hill

*See citations in: Lee, Christopher A. "<u>Archival Application of Digital Forensics Methods for Authenticity</u>, <u>Description and Access Provision</u>." In *Proceedings of the International Council on Archives Congress, Brisbane, Australia, August 20-24, 2012*.

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)

Core BitCurator Team • Cal Lee, Pl

- Matt Kirschenbaum, Co-PI
- Kam Woods, Technical Lead
- Alex Chassonoff, Project Manager (UNC)
- Sunitha Misra, GA (UNC)
- Porter Olsen, GA (MITH)











Two Groups of Advisors

Professional Experts Panel	Development Advisory Group		
 Bradley Daigle, University of Virginia Library Erika Farr, Emory University Jennie Levine Knies, University of Maryland Jeremy Leighton John, British Library Leslie Johnston, Library of Congress Naomi Nelson, Duke University Erin O'Meara, Gates Archive Michael Olson, Stanford University Libraries Gabriela Redwine, Harry Ransom Center, University of Texas Susan Thomas, Bodleian Library, University of Oxford 	 Barbara Guttman, National Institute of Standards and Technology Jerome McDonough, University of Illinois Mark Matienzo, Yale University Courtney Mumma, Artefactual Systems David Pearson, National Library of Australia Doug Reside, New York Public Library Seth Shaw, University Archives, Duke University William Underwood, Georgia Tech 		





BitCurator Goals

- Develop a system for collecting professionals that incorporates the functionality of opensource 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 (primarily data capture and reporting) of open source software: fiwalk, bulk extractor, Guymager, The Sleuth Kit, sdhash and others
- Can be run as:
 - Self-contained environment (based on Ubuntu Linux) running directly on a computer (download installation ISO)
 - Self-contained Linux environment in a virtual machine using VirtualBox
 - As individual components run directly in your own Linux environment or (whenever possible) Windows environment

Acknowledgement to Simson Garfinkel

- Digital forensics scholar at Naval Postgraduate School
- Responsible for:
 - fiwalk
 - Bulk Extractor
 - Digital Forensics XML (DFXML) metadata conventions
 - forensicswiki.org
 - digitalcorpora.org



BitCurator-Supported Workflow



See: http://bitcurator.net

Metadata Conventions of the BitCurator Tools: Digital Forensics XML (DFXML)

000			BitCu	rator-Test-0.1.7 [Running]	
LibreOffi	ce Ca	alc		🖂 🖾 Ťį 4)) 1:03 AM 👤 BitCurator 🔱
۲	80	 Documentation and Help DFXML tag library v3.xls 	x - LibreOffice Calc		
	<u>F</u> ile	<u>E</u> dit <u>V</u> iew <u>I</u> nsert F <u>o</u> rmat	<u>T</u> ools <u>D</u> ata <u>W</u> indow	<u>H</u> elp	×
		š • 📇 🔮 🚖 🗾 🔝		🗊 🗊 🔹 🍰 👘 🕐 🕐 🔊 🚳 💑 🍏 🖉	
	-	Arial 💌 11			E · 💁 · 🗐
	A62	$r f(x) \Sigma =$	<compiler></compiler>		
		A	В	С	D
	1	Tag name	Element name	Description	May contain
	2	< <u>dfxm</u> l>	DFXML	Root element, marks the beginning and end of the DFXML metadata file. The <dfxml> element contains the primary elements reported in fiwalk's xml structure: <metadata>, <creator>, <source/>, <volume>, and <runstats>.</runstats></volume></creator></metadata></dfxml>	< <u>metadata</u> >, <creator>, <source/>, <volume>, <<u>runstats</u>>, <<u>sectorsize</u>>,<<u>pagesize</u>>,<</volume></creator>
	3				acquisition seconds>
	3	< <u>metadata</u> >	Metadata	The < <u>metadata</u> > tag provides header information that defines the <u>metadata</u> in the DFXML document. Includes <u>namespace</u> declaration, <u>namespace</u> schema location, and other information that is used to define the elements used in the XML file.	<dc:type>, <dc:creator>, <dc:title>, <dc:description>; for more information on Dublin Core element set,</dc:description></dc:title></dc:creator></dc:type>
U	4			These declarations provide information on the types of standardization schemes used to convey information in the DFXML document. The <metadata> tag may also contain high</metadata>	see (21).
				rendered in Dublin Core (dc), in order to increase interoperability.	
<u>^_</u>	-	(araatan)	Oraștar	The Creater element provides decumentation shout the	
			Creator	program and computing environment in which the disk analysis (or capture) take place. <creator> includes tags documenting the program that initiated the capture creating the DFXML file, and other contextual information about the system on which</creator>	<program>, <version>, <build_environment>, <execution_environment></execution_environment></build_environment></version></program>
- data	HI	Fiwalk / bulkextractor / attributes /	color code /Sheet2 / 🐥 /		
	She	et 1 / 5 P	ageStyle_fiwalk	STD 🗅 Sum=0	

일 💿 🖉 🗗 🛄 🛄 🚫 퇲 Left ೫

http://www.bitcurator.net/2013/02/06/dfxml-tag-library/

Metadata Generation and Reporting



See: Woods, Kam, Christopher Lee, and Sunitha Misra. "Automated Analysis and Visualization of Disk Images and File Systems for Preservation." In *Proceedings of Archiving 2013* (Springfield, VA: Society for Imaging Science and Technology, 2013), 239-244.

-		

Ubuntu Desktop







home



Imaging Tools



Forensics Tools



Additional Tools

🔜 🖬 💼 🛊 🕕) 12:57 AM L BitCurator 以





BitCurat



🔞 💿 🖉 🖃 🛄 💟 🔇 💽 Left 🕷

Acquiring Disk Images with Guymager

00		BitCurator-0.3.0 [Running]	
Guymager		📖 🐱 💷 📬 🗐 1:15 AM 👤 BitCurator	ψ
٢			
	× – O GUYMAG	Acquire image of /dev/sdb	
	Devices Misc Heip	File format	
		C Linux dd raw image (file extension .dd or .xxx)	
	Serial nr.	© Expert Witness Format, sub-format Guymager (file extension .Exx) C Advanced forensic image (file extension .aff) C Advanced forensic image	
	VB20bfff24-3f8cb5fa	Case number 1	
		Evidence number 1	
		Evaniner BitCurator test user 0 2%	
		Description A sample floppy disk image	
		Notes Additional notes go here	
		Destination	
		Image directory /home/bcadmin/Desktop/SampleData/	
		Image filename (without extension) sampleimage	
	Size :	Info filename (without extension) sampleimage	
	Image file /	Hash calculation / verification	
>_	Current speed	Calculate MD5 Calculate SHA-1 Calculate SHA-256	
	Hash calculation	☐ Re-read source after acquisition for verification (takes twice as long)	
	Image verification	Verify image after acquisition (takes twice as long)	
		Cancel Duplicate image Start	
	Trash		

Exporting Filesystem Content Using fiwalk

000	BitCurator-0.3.0 [Running]
Ubuntu Desktop	📖 💌 🗊 👣 🖛 👤 BitCurator 🔱
Computer	Fiwalk XML Annotated Features Reports
home	Fiwalk produces a DFXML file showing the volumes, directories, and files Image File Image File Image File
Imaging Tools	(home/bcadmin/Desktop/SampleData/sampleimage.E01) Output XML File
Forensics Tools	(home/bcadmin/Desktop/SampleData/sampleimage.xml)
Additional Tools	
	Cancel Close OK tCurator
9 9	

Viewing the Command Line Output



DFXML Output from fiwalk

000	BitCurator-0.3.0 [Running]		
Mozilla F	irefox 📖 🖾 💼 👣 🖬 1:46 AM 👤 E	SitCurat	or 🔱
	🗍 file:///home/bmpleimage.xml 🛟		
	$ = \frac{1}{2} $		۱ 🔥
		<u> </u>	2 💷
			â
	This XML file does not appear to have any style information associated with it. The document tree is shown below.		Ú
	$-< df_{ym}$ vorsion $-$ "1.0"		
	- <metadata></metadata>		
	<pre><dc:type>Disk Image</dc:type></pre>		
	- <creator version="1.0"></creator>		
	<program>fiwalk</program>		
	<version>4.1.0</version>		
	- build_environment>		
	<compiler>GCC 4.6</compiler>		
	<pre>library name="affild" version="3./.1"/></pre>		
242	<pre></pre> /build environment>		
	- <execution environment=""></execution>		
	- <command line=""/>		
	fiwalk -f -X /home/bcadmin/Desktop/SampleData/sampleimage.xml /home/bcadmin/Desktop/SampleData		
	/sampleimage.E01		
Ĺ	<start_time>2013-07-20T05:34:37Z</start_time>		
	<image_filename>/home/bcadmin/Deskton/SampleData/sampleimage_E01</image_filename>		
	fs start: 0		
	- <volume offset="0"></volume>		
	<pre><partition_offset>0</partition_offset></pre>		
0	<sector_size>512</sector_size>		
	<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>		
		ما 🗐 🔕	 ft ¥£

DFXML for a Specific File

```
<fileobject>
      <filename>Documents and Settings/All Users/Documents/
                 My Pictures/Sample Pictures/Blue hills.jpg
      </filename>
      <filesize>28521</filesize>
      <alloc>1</alloc>
      <used>1</used>
      <inode>6245</inode>
      . . .
      <uid>0</uid>
      <qid>0</qid>
      <mtime>1208174400</mtime>
      <ctime>1257729636</ctime>
      <atime>1257729636</atime>
      <crtime>1257729636</crtime>
      <seq>2</seq>
      <libmagic>JPEG image data, JFIF standard 1.02</libmagic>
      <byte runs>
       <run file_offset='0' fs_offset='0' img_offset='363200512'
         len='0'/>
      </byte runs>
      <hashdigest type='MD5'>
          6fb2a38dc107eacb41cf1656e899cf70
      </hashdigest>
      <hashdigest type='SHA1'>
          4eee44b18576e84de7b163142b537d2fe6231845
      </hashdigest>
</fileobject>
```

Identifying "Features" of Interest in Disk Images or Directories

Bulk Extractor

000		BitCurator-0.3.0 [Running]
Bulk Extr	actor Viewer	📖 🖾 🗊 👣 🜒 2:00 AM 👤 BitCurator 🔱
	File Edit View Tools He	8 Run bulk_extractor
	x 🔒 🎭 🖷 📁	Required Parameters Scanners
	🗙 Highlight:	Scan: Image File Raw Device Directory of Files Directory of Files
	Reports Feature Filter	Image file :op/SampleData/sampleimage.E01 wordlist
		Output Feature Directory ampleData/bulk-extractor-output
	Feature File None	General Options
		Use Banner File
		Use Alert List File
		Use Stop List File
		Use Find Regex Text File
		Use Find Regex Text
222		Noning Faranceers
	-	
7 2	Deferred Feature	Use Margin Size
	Referenced Featu	Use Min Word Size
>_		Use Max Word Size
		Use Block Size 512
		Use Number of Threads
		Scapper Controls
		Use Plugin Directory
		Use Scan Option Name
		Restore Defaults Start bulk_extractor Cancel
		😂 😳 🖉 🗗 🛄 🔯 Left #

Scanner	Description	
scan-accts	Looks for phone numbers, credit card numbers, etc	
scan_base64	Decodes BASE64 text	Scanners Discontinuity Scanners
scan_kml	Detects KML (Keyhole Markup Language) files – used to identify geographic locations	□ wordlist
scan_gps	Detects XML from Garmin GPS devices	 ✓ base16 ✓ base64
scan_aes	Detects in-memory AES (Advanced Encryption Standard) keys from the key schedules	Image: Second secon
scan_json	Detects JavaScript Object Notation files	gps 🕑 gzip
scan_exif	Detects EXIF structures from JPEG files	
scan_zip	Detects and decompresses ZIP files and zlib streams	i john i kml i net i pdf
scan_gzip	Detects and decompresses GZIP files and gzip streams	 ✓ vcard ✓ windirs ✓ winpe
scan_pdf	Extracts text from some kinds of PDF files	i winpr≜fetch
scan_hiber	Detects and decompresses Windows hibernation file fragments	
scan_winprefetch	Detects and extracts fields from windows prefetch fields from Windows prefetch files and file fragments	Cancel

000	BitCurator-0.3.0 [Running]	
Bulk Extr	ractor Viewer 📃 💌 💷 📬 📣 🖄 2:02 AM 👤 BitCurator 🗄	¥.
	File Edit View Tools Help	
	× 🖴 🎭 📲 🗊 📥	
	X Highlight:	
	Reports Feature Filter 🗌 Match case Navigation	
	× bulk_extractor Scan	T
	Image File sampleimage.E01 Feature Directory bulk-extractor-output	
	Progress Done bulk extractor scan completed. See Status, below, for details,	
	Options Departie Departs	
	bulk_extractor'	
	//home/bcadmin/Desk i bulk_extractor has completed. //home/bcadmin/Desk i Report bulk-extractor-output has been opened and is ready for viewing.	
500	Status	
	R Elapsed time: 0.4985 sec.	
	Total email features found: 0	
<u>}-</u>	Done.	
R.M.		
	Close	
	🖲 Text 🔿 Hex 🛛 🖷 🗭	
	9 💿 🖉 🗗 🛄 🔇 💽 Left 🕱	11.

Histogram of Email Addresses (Specific Instances in Context on Right)



Matching Bulk Extractor Output (Based on Byte Offsets) to fiwalk Output (Based on Filesystem Location)



Generating BitCurator Reports

BitCurator-0.3.0 [Running] 📖 💌 🗊 🚛 🕪) 2:32 AM 👤 BitCurator 😃 **BitCurator Forensics GUI** \odot Bitcurator Reports Computer Fiwalk XML Annotated Features Reports Help Produces Office Open XML and PDF reports to assist in image analysis Fiwalk XML File home e/bcadmin/Desktop/SampleData/sampleimage.xml Network Servers Annotated Feature Files Directory /bcadmin/Desktop/SampleData/annotated-features Imaging Tools Output Directory For Reports: /home/bcadmin/Desktop/SampleData/bc-reports Config File (optional): /Path/To/file Forensics Tools Command Line Output: reports/FiwalkReport.pdf System Settings 4) /home/bcadmin/Desktop/SampleData/bcreports/FiwalkDeletedFiles.pdf 5) /home/bcadmin/Desktop/SampleData/bc-Additional Tools reports/BeReport.pdf Generating Excel report /home/bcadmin/Desktop/SampleData/bcreports/sampleimage.xml.xlsx Curator Cancel Close OK °D Trash



P Droviour - Nove 4		. 🕞				
Previous 🧄 Next 🔤	(Tor 1) Fit Page width	* –				
iumbnails 🔻 🕷	Domosti Dulli Estract					RitCurator
		or reatures				
					Note:	
					FIUF:Total f	eatures unallocated to files
					FIUF:Total f	eatures unallocated to files
1	Bulk Extractor Depart Files	Footure Instances	FLTE	FILIE	FICR:Total f	eatures in compressed regions
	annotated, telephone tyt	r cature instances	4	1	2	
	annotated rfc822.txt	258	39	219	110	
	annotated_zip.txt	127	8	119	3	
	annotated_windirs.txt	466	13	453	180	
	annotated_domain.txt	653	48	605	317	
	annotated_exif.txt	2	2	0	0	
	annotated_winpe.txt	1	1	0	0	
	annotated_email.txt	500	42	458	224	
						9 0 Ø P 🗂 💟



😫 🕢 🖉 🖃 🛄 🚺 🔇 💽 Left 🕱

Nautilus Scripts

- Scripts that can be run using the GNOME file manager called Nautilus (Linux analog to Windows Explorer or Mac OS X Finder)
- Can be used in the BitCurator environment or your own Linux environment

MD5 Hashes of Files (Nautilus Script)



BitCurator Tools and Further Information

🛔 Log in

BitCurator

Software Document Navigation Nain page Recent ch Toolbox What links Related ch Special pa Printable x

	Page Decusion	View View source History Search	Go Search		
•	An Introduction to BitCurator	Software			
ation	The BICurator project brings open source cipits formatics ² tools and techniques to collecting institutors preserving toon-dipited materials. BiCurator packages these technologies in an environment where saves can create thereadling-ackaged data mapse, perform sophistoteal to trage tasks on torn-dipital materials, extract and transform metidata, and redict sensitive information. A more detailed description of the Caratorie environment can be found on addressification page.	Download our current release using the links below. Looking for previous or upcoming releases? Check our Software page.			
anges	Please test bicruiter read for sease, contacts, publications of and updates on research activities. We also invite you to pin the BRCunter community by following us at @bicruiter of an Twitter and participation of the provide comparison of the community of the second se	BitCurator Virtual Machine (Release 0.2.7) /F ((MD5)#		
	Getting Started				
here langes ges	BiCurator uses a customized Linux environment based on Ubuntu Linux 12.04.075. BiCurator is currently available as effer an ISO import P or as a pre-configured Virtual Machine (P (VM)). Users can use or install BiCurator in one of three ways:	Source Code and Packages			
ersion	1. Use as a virtual machine using VirtualBox (as demonstrated in the video below).	~~~			
l Brik	Download the ISO image and install by Moleying the installation procedure for Usurlu Linux () The STURATE on a hostifield USE disk for devinations or Coll times and Use Installation Electrical USE installation	Community and Docs a			
	The default username and password for the BitCurator environment are both "boadmin".	Quickstart Guide E Get up and running with our	quickstart guide.		
	The lutural betw steps you through the process of instaining striculator as a virtual mechine using Virtuatiox. Find more luturate on our You table of channel, or check out our Quick Start Guide on the right.	Commentation Installation guides, project require documentation.	ements, and developer		
	Installing BHCurator using Virtual Box YouTube	Join our Google Group () to post questions and directly.	comments. Or, email us		
	A of a constant on experiment of the darks to be	Frequently Asked Questions Run into a proble	m? Check our FAQs first		
		Twitter®			
		Please be patient as we transition our feed widget to Twitter	1.1 API		

BitCurator

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

Archives

June 2013

May 2013

February 2013November 2012

October 2012September 2012

August 2012

Home About People Software FAQ Publications Presentations Related Resources
WELCOME TO THE BITCURATOR PROJECT. @BITCURATOR Software
The BitCurator Project is an effort to build, test, and analyze systems and

software for incorporating digital forensics methods into the workflows of a variety of collecting institutions. Read more about BitCurator here...

Posted on July 10, 2012 by Kam Looking for the test release of the BitCurator virtual environment? You can find the latest release on the front page of our wiki, in the Downloads section. We've also posted instructions to help you get started. Posted in Uncategorized (Leave a comment

Marty Gengenbach selected for National Digital Stewardship Alliance Innovation Award Get the software Documentation and technical specifications Screencasts Google Group http://wiki.bitcurator.net/

People Project overview Publications News http://www.bitcurator.net/

Twitter: @bitcurator