# Embracing the D Word - Placing Archives Development in the R&D Landscape

## Cal Lee

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

Research Forum Society of American Archivists Annual Meeting August 12, 2014 Washington, DC





SCHOOL OF INFORMATION AND LIBRARY SCIENCE

# D is for Development

- Advancing the archival profession requires active research and <u>development</u>
- The archival literature includes several persuasive calls for the importance of research into issues as such user needs and the costs/benefits of archival processes.
- There has been relatively little emphasis on the role of innovative and systematic <u>development</u> in the archival enterprise.

# "Research and development (R&D) is the creation of knowledge to be used in products or processes."

Levy, David M. "Research and Development." In Concise Encyclopedia of Economics (1st ed.), edited by David R. Henderson. Library of Economics and Liberty, 2002.

# Question: How should archivists appraise and select materials on the Web?

# VidArch

- Funded initially by NSF (2005-2007)
  - "Preserving Video Objects and Context: A Demonstration Project"
  - Supported by the National Science Foundation #IIS
     455970 DigArch Program
- Second grant funded by LOC NDIIPP (2007-2009)
  - Extending archival documentation strategies
  - Partners: San Diego Supercomputer Center and Internet Archive

See: Gary Marchionini, Helen Tibbo, Cal A. Lee, Paul Jones, Robert Capra, Gary Geisler, Terrell Russell, Laura Sheble\*, Sarah Jorda, Yaxiao Song, Dawne E. Howard, Rachael Clemens, Brenn Hill (2009). *VidArch: Preserving Video Objects and Context Final Report.* http://sils.unc.edu/sites/default/files/general/research/TR-2009-01.pdf

# ContextMiner\*

(http://www.contextminer.org)

- Web-based service for building collections, through "campaigns" (i.e. sets of associated queries and parameters to harvest content over time)
- For campaign, user specifies how often to query, number of results to harvest, hosts to query
- Can collect information from various sources: blogs, Flickr, Twitter, YouTube, open Web
- Uses various site-specific APIs to collect data

\*Developed by Chirag Shah (now at Rutgers University)

#### Create a new campaign (step 1 of 2)

A campaign is like a project. Once you create a new campaign, you will be able to choose what objects you want to collect, from which sources, and how.

Enter a name for your campaign (e.g. Swine flu, Economic recession)

Create		

#### Create a new campaign (step 2 of 2)

Here you can choose what you want to collect for your campaign Economic recession.

Clicking on a source button will expand or collapse its options. Any query you add through "Common Queries" box will be added to Blogs, Twitter, YouTube, and Flickr.

Don't forget to click on "Start My Campaign" once you are done entering your queries and URLs, and configuring their options. Once your campaign is running, you can come back to this page any time to add more queries/URLs, or change any of the options.



### Parameters for a Query within a Campaign

	YouTube									
A	Add the <b>queries</b> that you want to monitor. ContextMiner will keep running them periodically on YouTube and collect the top results as per your criteria.									
	Add									
	#	Query	# of results	Criteria	What to collect	Frequency	Actions			
	1	stimulus package	100 Update	<ul> <li>Relevance</li> <li>Published</li> <li>View count</li> <li>Rating</li> </ul>	<ul> <li>Results+Basic attributes</li> <li>Results+Basic attributes</li> <li>+Periodic attributes</li> <li>Collect inlinks</li> </ul>	<ul> <li>Daily</li> <li>Weekly</li> <li>Monthly</li> </ul>	Pause Delete			

#### Three Different Campaigns for a Given User

 Select: 🗘 Action: 🗘 Go								
ID	Title	Date Created	Collection	Status	Last Export	Manage		
2	Elections 2008	2008-06-24	YouTube: 1537, In-links: 337756 Blogs: 38949 Tweets: 31616	Active	N/A	Description   Parameters   Queries		
516	Cancer	2009-06-16	Web: 500 YouTube: 532, In-links: 10757 Blogs: 707 Tweets: 3081 Flickr: 1678	Active	N/A	Campaign Options		
523	Swine flu	2009-06-17	Web: 500 YouTube: 148, In-links: 3936 Blogs: 459 Tweets: 591 Flickr: 188	Active	N/A	Campaign Options		

#### Items from YouTube within a Collecting Campaign

1	My Cam	paign	I <b>S</b> -:	> Cancer -> YouTube			Campaigr	Options
[				Search		Showing resul	ts of all the	e queries.
(	Select:	+		Action: Go [First Page] [Prev] Sho	wing pag	e 1 of 22 page	s [Next] [L	ast Page]
	- 1	+	#	Title	Query	Category	Duration	Date
	00	0	1	Prostate Cancer Drug Improves Survival	cancer survival	News	4.85 min.	2009-06-16
	00	0	2	SUNDAY NEW YORK TIMES CHEMO BRAIN AND CANCER SURVIVAL	cancer survival	News	8.75 min.	2009-06-16
	00	0	3	Treating Cancer - Dendreon's Provenge May Improve Survival Rate for Prostate Cancer	cancer survival	News	3.23 min.	2009-06-16
	00	0	4	Tips For Cancer Survival	cancer survival	News	4.15 min.	2009-06-16

### Detailed Metadata for a Video from YouTube



Crawl #	Crawl date	Views	Ratings	Avg Rating	Comments	Favorited
1	2009-06-16	2938	5	5	4	5
2	2009-06-17	2941	5	5	4	5
3	2009-06-18	2942	5	5	4	5
4	2009-06-19	2942	5	5	4	5
5	2009-06-20	2946	5	5	4	5
6	2009-06-21	2948	5	5	4	5

## Items from Blogs within a Campaign

М	y Camj	paigns	-> (	Cancer -> Blogs		Campai	gn Options
				Search	Showing	g results of all t	he queries.
	Select:	\$	4	Action: Go [First Page] [Prev] Sho	wing page 1 of 29	pages [Next]	[Last Page]
	- 1	+	#	Title	Query	Published	Collected
0	00	0	1	Cancer Research UK Researchers find clues to improve breast Snippet: Checking lymph nodes during surgery and assessing the hormone status of tumours could help improve breast cancer survival in the UK according to research published today in Annais of Oncology Author: unknown, Site: cancerresearchuk Published: 2009-06-08, 00:05:00	cancer survival	2009-06-08	2009-06-16
	00		2	Herceptin aids stomach cancer survival study   HEALTH News	cancer survival	2009-06-01	2009-06-16
	00	00	3	Campaigners blast Scots lung cancer survival rates   Scotland	cancer survival	2009-06-08	2009-06-16
	00	0	4	Predicting Breast Cancer Survival Rates   Ivanhoe's Medical	cancer survival	2009-06-01	2009-06-16

### Items from Flickr within a Campaign

My Campaign	<u>15</u> -	> Cancer -> Flickr	Ca	mpaign C	<u>Dptions</u>
		Search	Showing results of	of all the o	ueries.
 Select:		Action: Go [First Page] [Prev] S	howing page 1 of 8 pages []	Next] [Las	t Page]
-   +	<u>#</u>	Picture	Author	Query	Collected
000	1	Cancer Survival Plaza Description: Tags: chicago, skydeck, willis, tower, millenium, park, chrome, jellybean, chinese, scultpure Taken: 2009-06-15, 19:45:43 Posted: 2009-06-15, 13:03:27	Henry Lopez Real name: Henry Lopez, Location: Chicago, United States	cancer survival	2009-06-16
000	2	IMG_0699 Description: 1000 paper birds Tags: paper, birds, cancer, survival, survivor, memorial, katrina Boemig Taken: 2009-04-25, 15:36:21 Posted: 2009-04-26, 13:20:40	katrina boemig Real name: N/A, Location: N/A	cancer survival	2009-06-16
000	3	Cancer Survival Description: My husband has just been diagnosed with terminal metastatic pancreatic cancer that has invaded his spine, liver, abdomen, and lymphatic system. The doctors have sent him home to die, but that's not my truth. I believe, in every cell of my body, that he can be healed. We are throwing every compatible holistic therapy at his cancer, and I believe that we can be triumphant. I made this graphic to remind us of what it will take to survive and overcome this cancer. I'll be placing them all over the house. Tags: cancer, survivor, terminal, illness, hope, inspiration, inspirational, colorful, Holly Smith, Michael Smith, cancer survivor, overcoming cancer, holistic, naturopathy, cesium, rubidium, laetrile Taken: 2009-04-05, 14:29:35 Posted: 2009-04-05, 14:29:35	<u>TheEclecticArtisan</u> Real name: Holly Smith, Location: Atlanta, USA	cancer survival	2009-06-16
000	4	Cancer survival Description: My BIL and nephew Love this old cocoa daisy kit Tags: Taken: 2009-03-08, 22:18:37 Posted: 2009-03-08, 23:18:37	<u>Lan A</u> Real name: Lan Amphone, Location: Andover	cancer survival	2009-06-16
	_				



# Question: How should archivists process born-digital materials?

# **Overarching Goals**

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

# **Digital Forensics in Archives**

- In recent years, archivists have been applying various digital forensics methods, for example:
  - use of write blockers
  - generation of disk images
  - applying cryptographic hashes to files
  - capture of Digital Forensics XML (DFXML)
  - scanning bitstreams for personally identifying information

#### From Bitstreams to Heritage:

#### Putting Digital Forensics into Practice in Collecting Institutions



Christopher A. Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassanoff

#### http://www.bitcurator.net/docs/bitstreams-to-heritage.pdf

# Digital Forensics Lab @ UNC School of Information and Library Science



Need for Adaptation of Digital Forensics Tools and Tasks for Archivists

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

# BitCurator

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

# **BitCurator Goals**

- Develop a system for collecting professionals that incorporates the functionality of 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

# Core BitCurator Team

- Cal Lee, Pl
- Matt Kirschenbaum, Co-PI
- Kam Woods, Technical Lead
- Porter Olsen, Community Lead
- Alex Chassanoff, Project Manager
- Sunitha Misra, Software **Developer (UNC)**
- Kyle Bickoff, GA (MITH)















## Two Groups of Advisors

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





## BitCurator Environment\*

- Bundles, integrates and extends functionality 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 e.g. Virtual Box or VMWare
  - As individual components run directly in your own Linux environment or (whenever possible) Windows environment

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

#### **BitCurator-Supported Workflow Elements**



#### See: http://bitcurator.net

#### Mounted Devices set to Read-Only by Default\*



\*Not to replace hardware-based write blocking, but useful for various purposes

#### Creating a Disk Image in Guymager\*

Guymage	5 <b>r</b>	🔜 💌 🗊 🖡 🗤) 2:12 F	PM 👤 BitCu	irator 岱
	Computer	Acquire image of /dev/sdb	Documentat Help	ion and
	× - D GUYMA	File format		
	<u>D</u> evices <u>M</u> isc <u>H</u> e	C Linux dd raw image (file extension .dd or .xxx)		
	Rescan	Expert Witness Format, sub-format Guymager (file extension .Exx)		
	Sorial	C Advanced forensic image (file extension .aff)	Rod	
	nr.	Case number 1	sectors	Progre
		Evidence number 1		
		Examiner BitCurator User		
	VB45b1d326-9557	Description A sample floppy disk image		
	1045010520 5557	Notes Additional notes go here		
242		Destination		
		Image directory /home/bcadmin/Desktop/SampleData/		
69		Image filename (without extension) sampleimage		
	•	Info filename (without extension) sampleimage		•
	Size			
	Sector size			
	Info file	Calculate MDS J Calculate SHA-1 M Calculate SHA-256		
	Started	Verify image after acquisition (takes twice as long)		
F	Source verification	verify image after acquisition (takes twice as long)		
	Image vernication	Cancel Dunlicate image Start		
	9			
9	_			
	Trash			

#### \*Developed by Guy Voncken

## Mounting a Disk Image to Browse the Contents

Ubuntu	Desktop				🔜 🖾 🗊 🕇	<b>↓ 4</b> ))	4:33 PM	A L BitCurator
0	Computer	1 10 101 1001	N				Do	ocumentation and Help
	<u>ia</u>	jo-work- 12-1	Open Open With Other Application					
	home		Scripts Cut Copy		Disk Image Info File Analysis File Info Find Files		• 11 •	Network Servers
	Imaging Tools		Make Link Rename Copy to Move to	* *	Mount Disk Image Open Browser as Ro Safe Mount Unmount Disk Imag	ot		
245	Forensics Tools		Move to Trash		Open Scripts Folder	2		
1			Resize Icon Restore Icon's Original Size					
<u>&gt;_</u>	Additional Tools		Send To Compress Revert to Previous Version					
			Properties					
					BitCu	٦٢	at	٦ 😎

Trash

## Mounting a Disk Image to Browse the Contents



#### Bulk Extractor\* – Identifying Potentially Sensitive Information

File Edit View Bo	🕫 🛞 Run bulk_extractor		
🕑 × 🔒 👒 '	Required Parameters		Scanners
Reports A	e Scan: 🖲 Image File 🔿 Raw Devic	e 🔿 Directory of Files	🗆 bulk
	Image file		 🗌 wordlist
	e Output Feature Directory		
	Connect Options		✓ accts
			M base16
			 M base64
	Use Alert List File		
	🗌 Use Stop List File		🖉 email 🗉
	🗌 Use Find Regex Text File		🗹 exif
	Use Find Regex Text		🗹 find
	Use Random Sampling		🗹 gps
220	Tuning Parameters		🗹 gzip
	Use Context Window Size	16	
	e 🗌 Use Page Size	16777216	🗹 kml
	Re 🗌 Use Margin Size	4194304	🗹 net
	Use Block Size	512	🗹 pdf 🛛 🗍
<u> </u>	Use Number of Threads	4	🗹 гаг
344	Use Maximum Recursion Depth	7	Vcard
2-18			
Sole at			winpe
	Parallelizing		
	Use start processing at offset		
http://www.	forensicswiki c	ora/wiki/Rulk	xtractor

\*Developed by Simson Garfinkel

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



### **BitCurator Reporting Tool**



#### Various Specialized BitCurator Reports

000	BitCurator-Demo-0.3.4	[Running]	
Document Viewer		📖 💌 🗊 🕇 🐠	6:27 PM & BitCurator
	× − □ format_table.pd ↑ Previous ↓ Next	f       1     (1 of 1)     Fit Page Width	
Computer	S Home	Report: File System Statistics and Files File Format Table Disk Image complement Bit	BitCurator
<ul> <li>Home</li> <li>Desktop</li> <li>Documents</li> <li>Downloads</li> <li>Music</li> <li>Music</li> <li>Pictures</li> <li>Videos</li> <li>Sile Sustem</li> <li>bc_format_barg</li> </ul>	feature Traph.pdf	Format           data           news or mail, ASCII text, with CRLF line term innors           PCX ver. 2.5 image data           PDF document, vension 1.4           MS Wadows icon resource - 2 icons, 3x, 4-colors           x86 boot sector, code offset 0x52, Octors 1, dos < 4.0 BootSector (0x0)           SysBx File - GreyMatter           empty Clp archive data, at lost v1.0 to extinct)           TIFF image data, little-endun           ASCII text, with no line terminators (OpenDocument Text)           JPEG image data, JFIF standard 1.01           PE32 executable (GUI) litel 30386, f, InnoSetupself extincting archive           /PEG image data, JFIF standard 1.01	Short Form         Files           dat_sta         31           new_cors         1           PCX_sta         1           PDF_14         6           MS_ors         1           X85_ors         1           Sys_str         1           Sys_str         2           TIF_sina         2           ASC_xt-         1           PE_01         4           PE3_j ve         1           PE4_1-         2           ASC_xts         40           Com_sto         1           aven ptr         0
Previous Next   Image: state	1 (1 of 1) Fit Page Width •	Ints (by format)	

Specialized BitCurator Reports

File	Content
bc_format_bargraph.pdf	histogram of file formats found on the volume
bulk_extractor_report.pdf	high-level overview of feature locations on disk
fiwalk_deleted_files.pdf	shows paths to any deleted materials found in a given partition
fiwalk-output.xml.xlsx	Excel converted DFXML output (file system metadata)
fiwalk_report.pdf	high-level overview of file system characteristics
format_table.pdf	long-form file format names for formats shown in bar graph
premis.xml	PREMIS preservation metadata

Operationalizing Original Order - Filesystem Metadata Output from fiwalk\*

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

\*Developed by Simson Garfinkel

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



#### Exporting Selected Files from a Disk Image



# Nautilus Scripts

- Scripts that can be run using Nautilus (GNOME file manager)
- Most provide more convenient access (right click and menu selection) to functions performed by applications that could also be run directly

#### Right Click on File or Directory and Calculate MD5



>\_

29

6

Calculat	e MD5 (Files and Dire	ectories)				$\ge$	<b>⊠</b> † <sub>↓</sub>	<b>■</b> ))	8:31 AM	L BitCurat	or 🔱
0	Computer	• 🙍 Home	Pictures						4	<del>)</del> ବ୍ ବ୍	Search
	<ul> <li>Desktop</li> <li>Documents</li> <li>Downloads</li> <li>Music</li> </ul>	bc400px-128 png	Ofull.	Bitturator bitcurator-grub.png	bitcurator-grub- new.png						
	<ul> <li>Pictures</li> <li>Videos</li> <li>File System</li> </ul>										
	Trash Network		🙁 Cal	culate MD5 (Files and	Directories)						
	Browse Net		Please o (1 file(s)	choose the way you wa ) selected)	nt the MD5 hash to l	be presen	ted:				
			Har	ndling play on screen ve to file (the selected (	filename + md5 evt	enion)					
			0 34		Cancel						

Calculate MD5 (Files and Directories)

📖 🖾 📼 📬 🕕) 8:30 AM 👤 BitCurator 🔱



#### Quick Access to a Hex View:



Hex		📖 🖾 🗊 📬 🗉 🔝 🛄 🔜 🔜 🔜	<mark>tor</mark> ∜
	× – D Pictures		
9	🛞 🗆 bitcurator-grub.png - GHex		
	00000000       9       50       4E       47       0D       0A         00000012       02       80       00       00       01       E0         00000024       09       70       48       59       73       00         00000024       09       70       48       59       73       00         00000036       00       0A       4F       69       43         00000048       49       43       43       20       70       72         0000005A       54       53       E9       16       3D       F7         0000006C       20       52       42       88       80       14         0000007E       51       C1       11       45       45       04         000000042       6B       D6       BC       F7       E6       CD         000000042       6B       D6       BC       F7       E6       CD         000000042       6B       D6       BC       F7       E6       CD         000000044       08       CP       64       83       51         000000120       FF       E6       D3       00 <th>1A       0A       00       00       00       48       44       52       00       00       IHDR.       Immediation and Help         08       02       00       00       0B       13       01       00       9A       9C       18         IHDR.  </th> <th></th>	1A       0A       00       00       00       48       44       52       00       00       IHDR.       Immediation and Help         08       02       00       00       0B       13       01       00       9A       9C       18         IHDR.	
89	Signed 8 bit: -119	Signed 32 bit: 1196314761 Hexadecimal: 89	
	Unsigned 8 bit: 137	Unsigned 32 bit: 1196314761 Octal: 211	
	Signed 16 bit: 20617	Float 32 bit: 5.281654e+04 Binary: 10001001	
	Unsigned 16 bit: 20617	Float 64 bit: 5.292398e-260 Stream Length: 8 💻 🖶	

\$

Desktop

29

6



#### EnCase Disk Image Info

Computer

Network

















Computer		Mome Desktop	🔶 🄶 🔍 Search
🙍 Home			
📼 Desktop		8 EnCase Disk Image Info	1 10 101
📑 Documents		owfipfo 20120/116	1010
📴 Downloads	1	ewini 0 20130410	2009-12-11.E01
🔏 Music		Acquiry information	
Pictures		Acquisition date: Wed Jan 19 12:09:18 2011	
🖪 Videos		System date: Wed Jan 19 12:09:18 2011	
File System	· '	Operating system used: Linux	
Trash		Software version used: 20100226	
la hush		Password: N/A	
Network		FW/F information	
Browse Net		EWF INFORMATION File format: EpCase 6	
		Sectors per chunk: 64	
		Error granularity: 64	
		Compression method: deflate	
		Compression level: best compression	
		Set identifier: 4eb6701d-6cf0-2f4a-a0c6-0cb5d5e20959	
		Media information	
		Media type: fixed disk	
		Is physical: yes	
		Bytes per sector: 512	
		Number of sectors: 2068480	
		Media size: 1010 MiB (1059061760 bytes)	
		Digest hash information	
		MD5: 9c0de6c8532d7a66ddcf01861dfb6535	
		Cancel OK	

🖾 📼 🔃 🗤) 8:43 AM L BitCurator 🔱

Quick Start Guide Most recent version always available at:

http://wiki.bitcurator.net/

# BitCurator

## Quick Start Guide v0.9.16

Last updated: August 3, 2014



SCHOOL OF INFORMATION AND LIBRARY SCIENCE



MARYLAND INSTITUTE FOR TECHNOLOGY IN THE HUMANITIES

## Other Functionality to Meet Identified Needs:

Function	Tool(s)
Identify duplicate files	FSLint
Characterize files	FITS
Examine, copy and extract information from old Mac disks	HFSExplorer
Package files for storage and/or transfer	BagIt (Java) library
Scan for viruses	ClamTK
Read contents of Microsoft Outlook PST files	readpst
Examine embedded header information in images	pyExifToolGUI
Generate images of problematic disks or particular disk types	dd, dcfldd, ddrescue, cdrdao (in addition to Guymager)
Identify files that are partially similar but not identical	sdhash, ssdeep

# BitCurator Consortium

- Continuing home for hosting, stewardship and support of BitCurator tools and associated user engagement
- Administrative home: Educopia Institute
- Funding based on membership dues
- Institutions as members, with two categories of membership: Charter and General
- Software and documentation will continue to be free and open source, but membership provides further benefits (e.g. support, training, development priority)

## http://www.bitcurator.net/bitcurator-consortium/

## DIMAC (Disk Image Access for the Web)

- Developed by Sunitha Misra and Kam Woods
- To dynamically navigate and download contents of a disk image, without having to download or mount the full image
- See: <a href="https://github.com/kamwoods/dimac">https://github.com/kamwoods/dimac</a>
- Demo at:

### http://www.youtube.com/watch?v=BwiWFqxYzQ8

See: Sunitha Misra, Christopher A. Lee, and Kam Woods, "A Web Service for File-Level Access to Disk Images," *Code4Lib Journal*, <u>http://journal.code4lib.org/articles/9773</u>.

96	Disk	Image Access for the Web - DIMA	C - Mozilla	Firefox				
Di	sk Image /	Access for the We						
-	@ 127.0	.0.1:5000/image/jo-work-usb-2009-1	2-11.E01/1	☆ <b>~ で</b>	<mark>8</mark> <del>▼</del> Google	Q	$\bigcirc$	
	Disk	mage Access for the	Web					Â
	Browse o	lirectories and download files:						Ξ
			File List					
	d/ı	Filename	Size	Modified time	Deleted?			
	r	UNTITLED 1 (Volume Label Entry)	0	2009-11-23T14:58:2	8Z No			
	r	Trashes	4096	2009-11-20T18:26:1	2Z No			
	r	SC00009.JPG	334501	2009-11-11T14:16:5	6Z Yes			
	d	. <u>Trashes</u>	2048	2009-11-20T18:26:1	2Z No			
	d	.Spotlight-V100	2048	2009-11-20T18:26:1	2Z No			
	r	_SC00010.JPG	336095	2009-11-11T14:17:1	8Z Yes			
	r	_SC00011.JPG	342041	2009-11-11T14:17:2	8Z Yes			
	r	<u>Cat.mov</u>	5712046	2009-11-20T14:25:2	6Z Yes			
	r	Cat.mov	4096	2009-11-20T18:31:4	8Z Yes			
	r	MontereyKittyHQ.m4v	25123154	2009-11-20T14:00:0	0Z Yes			
	r	MontereyKittyHQ.m4v	4096	2009-11-20T18:32	2Z Yes			
	d	<u>.fseventsd</u>	2048	2009-11-23T14:58:2	8Z No			
	r	<u>IGGER~1.M4V</u>	12536015	2009-11-18T21:50:5	4Z Yes			
	r	. <u>_TiggerTheCat.m4v</u>	4096	2009-11-20T18:35:2	8Z Yes			
	r	<u>Cat.m4v</u>	1077729	2009-11-20T14:12:0	4Z Yes			
	r	. Cat.m4v	4096	2009-11-20T18:35:3	4Z Yes			•

## APPLYING FORENSICS TO PRESERVING THE PAST: CURRENT ACTIVITIES AND FUTURE POSSIBILITIES

- Organizers: Cal Lee, Jeremy Leighton John, Susan Thomas
- To be held at Digital Libraries 2014, London, September, 8-12, 2014
- One-day event, split across an afternoon and following morning
- Call for papers (one page max) deadline August 1, 2014 [but talk to me if you still want to submit something)

# Lessons and Observations

- In order to empirically test ideas in a new context, one often has to build something
- Software development is a powerful mechanism for:
  - Operationalizing archival principles
  - Testing methods, approaches and assumptions
  - Serving the archival profession
- Iterative development enables responses to "emergent needs" (often things breaking)

# Opportunities for Collaboration and Development

- Identify areas of archival practice that:
  - Are underdeveloped

Are currently done manually (but shouldn't be)

- Be realistic about what role you'll play, given your expertise and phase of your career
- Be clear about whether you're simply doing a proof concept or think the product will be used "in production" (if the latter, have a sustainability plan)

## Thank You!

Support	Page Discussion	View	View source Hi	story Search	Go Search
QA	BitCurator & is a suite of open source digital forensics & and data analy	sis tools to help collecting in	stitutions (libraries, a	archives, and museum	s) process born-
iccreencasts bout rchival workflows oftware overview ource on GitHub	ograf ar manas in unanto supports portante ograf preservation outco • Create Forensic disk images: Disk images paradaged with metadata • Analyze files and file systems: View details on file system content • Extract file system metadata: File system metadata is a critical init • Identify and reduct sensitive information: Locate private and sen • Identify and reduct sensitive information: Locate private and sen • Find more details about sensefic provinges and the fibtration on our Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fibtration on our • Find more details about sensefic provinges and the fiber one of the fiber our details about sensefic provinges and the fiber our our our • Find more details about sensefic provinges and the fiber our our our our details about sensefic provinges and the fiber our our our our our our our our details about sensefic provinges and the fiber our our our our our our our our our ou	mes using somware and pra a about devices, file systems s from a wide variety of file sy cin the chain of custody and aitive information on digital m what can be discarded. software page.	, and the creation pr ystems. in records of prover edia and prepare m	the digital forensics co rocess. hance. laterials for public acce	ss.
tCurator.net site	Documentation		Get B	litCurator	
ools					
hat links here	1. Add BitCurator to Your Archival Workflow 2. Install BitCurator			BitCurator Virtual	Machine (v0.9.16)
ecial pages	3. Prepare Your Media		<u> </u>	[Download minorar]	[mbb checksumm]
intable version	4. Data Triage 5. Using Digital Ecrepsics Tools		(2	BitCurator Installat	ion ISO (v0.9.16)&
rmanent link ige information	The following descriptions include text, and screenshot hitorials. You can	also visit our video gallery fo	watch	[Download Mirror @]	[MD5 checksum #]
	related screencasts.	and that but video ganery to		Quickstart Guide v	0.9.16  Walk through
	I. Add BitCurator to Your Archival Workflow			installation and BitC	urator use.
	How do these tools address archival concerns? An overview of	f why you'd want to use each	tool.	The BitCountry Cou	
	How does BitCurator fit into common archival workflows? Lear	n how BitCurator supports for	rensic <u>C</u>	and speak with men	bers of the team.
	Research People Support Community		Install	Using BitCurator	Community
About Blog	Research People Support Community		Install	Using BitCurator	Community
About Blog	Research People Support Community	Video Spotlig	Install	Using BitCurator	Community
About Blog ews itCurator L ugust 12, 2 ssted July 23, 20	Research People Support Community Users Gathering @ Archives*Records - 2014 14 by Porter Olsen	Video Spotlig	Install	Using BitCurator	Community
About Blog IEWS ItCurator L ugust 12, 2 osted July 23, 20 Iease join us r August 12th	Research People Support Community Users Gathering @ Archives*Records – 2014 Drater Olsen for a gathering of BitCurator users on the evening after the SAA Research Forum from 5:30PM to	Video Spotlig BitC	ht urat o	Using BitCurator	Community C
About Blog Iews itCurator L ugust 12, 2 osted July 23, 26 Iease join us (August 12th 30PM. Come	Research       People       Support       Community         Users Gathering @ Archives*Records - 2014         2014       Porter Olsen         For a gathering of BifUcrator users on the evening after the SAA Research Forum from 5:30PM to meet other BitUcurator users and share your	Video Spotlig BitC	ht urat o	Using BitCurator	Community C
About Blog Iews ItCurator L ugust 12, 2 osted July 23, 20 Iease join us f August 12th Come xperiences w	Research People Support Community Jsers Gathering @ Archives*Records – C014 114 by Porter Olsen for a gathering of BilCurator users on the evening after the SAA Research Forum from 5:30PM to e meet other BitCurator users and share your orking with digital forensics and BitCurator. In	Video Spotlig BitC	ht urat o Screenca	Using BitCurator	Community C
About Blog Iews ItCurator L ugust 12, 2 asted July 23, 260 Iease join us f August 12th :30PM. Come xperiences we ddition to colli	Research People Support Community	Video Spotlig BitC	ht urat or Screenca	Using BitCurator	Community Community thefaad wolfnes addes
About Blog lews itCurator L ugust 12, 2 osted July 23, 20 lease join us f August 12th 30PM. Come xperiences we ddition to colli itCurator Co-l	Research People Support Community Desers Gathering @ Archives*Records – 2014 114 by Porter Olsen for a gathering of BitCurator users on the evening after the SAA Research Forum from 5:30PM to meet other BitCurator users and share your orking with digital forensics and BitCurator. In aborating with other BitCurator users, meet with PI'S Cal Lee and Matt READ MORE	Video Spotlig BitC	ht uration Screenca Installing virtual mac	Using BitCurator	Community Community that and tractines atoms
About Blog lews itCurator L ugust 12, 2 stead July 23, 22 lease join us 'August 12th 30PM. Come «periences we didition to collidition to collidition to collidition to collidition to collidition to transfer and the second second second the second s	Research People Support Community	Video Spotlig BitC	ht urat Screenca Installing virtual mac	Using BitCurator	Community Control of the set of t
About Blog lews itCurator L ugust 12, 2 lease join us r August 12th 30PM. Come xperiences with ddition to coll itCurator Co-l itCurator Co-l	Research People Support Community  Assers Gathering @ Archives*Records - 2014  Assers Gathering @ Archives*Records - 2014  Assers Gathering of BitCurator users on the evening after the SAA Research Forum from 5:30PM to meet other BitCurator users and share your orking with digital forensics and BitCurator. In aborating with other BitCurator users, meet with PI's Cal Lee and Matt READ MORE  Consortium — Membership Now Open  4 by Porter Olsen	Video Spotlig BitC as a @   <u>L</u>	ht uration Screenca Installing virtual mac	Using BitCurator	the shad work flues
About Blog lews itCurator L ugust 12, 2 systed July 23, 26 4 August 12th 30PM. Come «periences wi didition to collid diffuiton to collid diffuiton to collid diffuiton to collid ticCurator Co- itCurator Co- systed July 8, 201 he BitCurator	Research People Support Community Community Desers Gathering @ Archives*Records – 2014 Different Giben for a gathering of BitCurator users on the evening after the SAA Research Forum from 5:30PM to meet other BitCurator users and share your orking with digital forensics and BitCurator. In aborating with other BitCurator users, meet with PI's Cal Lee and Matt READ MORE Consortium — Membership Now Open Wey Porter Olsen Consortium (BCC) is an independent, community	Video Spotlig BitC as a @ L	ht urat Screenca Installing virtual mac	Using BitCurator	Community Community that and toutifues ations that and toutifues takens back of the second se
About Blog itCurator L ugust 12, 2 sited July 23, 20 lease join us August 12th 30PM. Come optimiones ou didition colli tiCurator Co-l itCurator Co-l	Research       People       Support       Community         Users Gathering @ Archives*Records – 1014         UH4 by Porter Olsen         for a gathering of BitCurator users on the evening after the SAA Research Forum from 5:30PM to pomet other BitCurator users and share your orking with digital forensics and BitCurator. In aborating with other BitCurator users, meet with PI's Cal Lee and Matt         Read Matter Obsen         Consortium — Membership Now Open 4/by Parter Obsen         Consortium ((BCC) is an independent, community ip association that will serve as the host and	Video Spotlig BitC as a @ [ ]	ht urat Screenca Installing virtual mac	Using BitCurator Using BitCurator Is an analysis of the second Is trutorial: BitCurato hine in Virtua MITH Marrie	And and investments
About Blog lews itCurator L ugust 12, 2 stated July 3, 20 lease join us August 12th 30PM. Come typeriences wu ddition to colli itCurator Co-l itCurator Co-l itCurator Co-l itCurator Co-l itCurator Co-l data July 3, 20 he BitCurator ed membersh enter of admini	Research People Support Community Resear	Video Spotlig BitC as a M	ht uration Screenca Installing virtual mac	Using BitCurator	Community Community theirs and workflows theirs and workflows their and workflows become we dependence of the community become we dependence of th
About Blog About Blog SitCurator L ugust 12, 2 sitCurator L ugust 12, 2 vitases join us f August 12th :30PM. Come xperiences wu ddition to colli titCurator Co- SitCurator Co- SitCurator Co- SitCurator Co- SitCurator Co- sted July 2, 20 SitCurator Co- sted July 2, 20 SitCurator Co- SitCurator Co-	Research People Support Community      Research People Support Community      Desers Gathering @ Archives*Records -     Conta      Desers Gathering @ Archives*Records -     Deservers      Desevers      Deservers      Deservers      Deservers      Deservers	Video Spotlig BitC asa @ 5	ht uration Screenca Installing virtual mac	Using BitCurator	Community Commun
About Blog About Blog SitCurator L ugust 12, 2 sitCurator L ugust 12, 2 sitCurator L ugust 12, 2 sitCurator C 4 ugust 12th 30PM. Come speriences with ddition to coll sitCurator Co- SitCurator	Research People Support Community      Seers Gathering @ Archives*Records -     2014      Seers Gathering @ Archives*Records -     Seers Gathering @ Arc	Video Spotlig BitC as a DitC Events	ht urat Screenca Installing virtual mac	Using BitCurator	the start work free starts

Get the software Documentation and technical specifications Screencasts Google Group http://wiki.bitcurator.net/

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

Twitter: @bitcurator