MAPPING DIGITAL FORENSICS METADATA TO PRESERVATION EVENTS USING BITCURATOR

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

MITH MARYLAND INSTITUTE FOR TECHNOLOGY IN THE HUMANITIES

Alexandra Chassanoff, Kam Woods, Christopher Lee
School of Information and Library Science, University of North Carolina at Chapel Hill

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

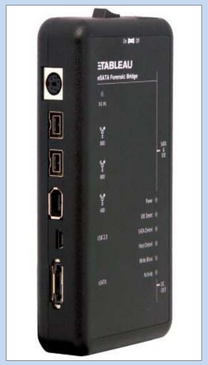
The BitCurator software environment, a suite of open-source digital forensic tools, can support the preservation goals of archivists throughout their institutional workflows.

Below are four points in an archival workflow in which preservation events can be recorded during the creation and ingest of a disk image.
We show how the output of digital forensics tools incorporated into BitCurator can be mapped to PREMIS encoded preservation events.

WORKFLOW STAGES

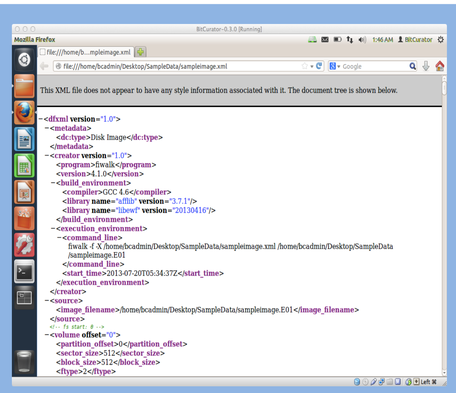
DISK IMAGING

Using Guymager, physical media is acquired as a raw bitstream packaged in an compressed forensic container



FILE ANALYSIS

Fiwalk produces a DFXML report of the contents of the file system(s).



FEATURE ANALYSIS

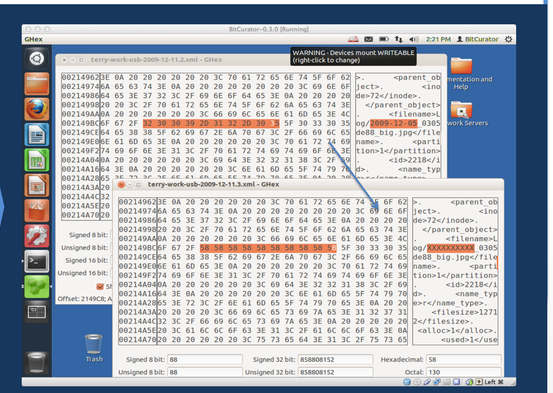
Bulk extractor performs forensic analysis of the raw bitstream and identifies features of interest.

Bulk Extractor Scanning Options

Scanner	Description
scan-accts	Looks for phone numbers, credit card numbers, etc
scan_base64	Decodes BASE64 text
scan_kml	Detects KML (Keyhole Markup Language) files – used to identify geographic locations
scan_gps	Detects XML from Garmin GPS devices
scan_aes	Detects in-memory AES (Advanced Encryption Standard) keys from the key schedules
scan_json	Detects JavaScript Object Notation files
scan_exif	Detects EXIF structures from JPEG files
scan_zip	Detects and decompresses ZIP files and zlib streams
scan_gzip	Detects and decompresses GZIP files and gzip streams
scan_pdf	Extracts text from some kinds of PDF files
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

REDACTION

The redaction tool iredact.py creates a new redacted version of a disk image



PREMIS ENCODED EVENTS

PREMIS Event 1: Image Capture

Record information about acquisition process including time, success/failure of creation, device specifics, and imaging format.

Semantic unit	Sample value(s)
eventidentifier	UUID
eventidentifier	8jb50321-6d7b-4291-89ag-a8b0fhc1f276
eventType	capture
eventDateTime	2013-03-29T 15:00:18Z
eventDetail	version="Guymager 0.4.2-2" Compilation timestamp = "2010-02-08-14.45.08" Compiled with="gcc 4.4.3" libewf version="20100226" libguytools version="1.1.1"
eventOutcomeInformation	Disk image created; Image creation failed
eventOutcomeInformation	.e01, .AFF
linkingAgentIdentifier	Forensic environment
linkingAgentIdentifier	BitCurator
linkingObjectIdentifier	UUID

PREMIS Event 2: File System Analysis

Captures technical metadata about each volume, partition, and associated file system on the disk image.

Semantic unit	Sample value(s)
eventidentifier	UUID
eventidentifier	69d02-93d1-dkj987-j308a0-k30x7g-3kd97
eventType	file system analysis
eventDateTime	2013-03-29T15:10:23Z
eventDetail	version="fiwalk 4.0.2" "2010-02-08-14.45.08" Compiled with="gcc 4.4.3" libewf version="20100226" libguytools version="1.1.1"
eventOutcomeInformation	File system analyzed; File system not analyzed
linkingAgentIdentifier	forensics tool
linkingAgentIdentifier	BitCurator
linkingObjectIdentifier	UUID

PREMIS Event 3: Feature Analysis

Documents the production of timestamped Bulk Extractor reports on specific features of interest.

Semantic unit	Sample value(s)
eventidentifier	UUID
Eventidentifier	39ag-032lj2-83-a3098gad-odua-308b
eventType	Feature stream analysis
eventDateTime	2013-03-29T15:12:36Z
eventDetail	version="bulk extractor 1.3.1"<os_sysname>Linux </os_sysname><os_release> 3.5.0-26 generic </generic </os_release><os_version># 42-Ubuntu SMP Fri Mar 8 23:18:20 UTC 2013</os_version><host>FW S306</host><arch>x86_64 </arch>
eventOutcomeInformation	X # Reports were produced; No reports were produced
linkingAgentIdentifier	Forensic environment
linkingAgentIdentifier	BitCurator
linkingObjectIdentifier	UUID

PREMIS Event 4: Redaction

Records metadata about the redaction of potentially sensitive or private information from the disk image.

Semantic unit	Sample value(s)
eventidentifier	UUID
eventidentifier	8jb50321-6d7b-4291-89ag-a8b0fhc1f276
eventType	redaction
eventDateTime	2013-03-29T16:46:13Z
eventDetail	version = "iredact.py"
eventOutcomeInformation	Redaction completed; Redaction not completed
eventOutcomeInformation	Original name="Van0C1.aff" Redacted name= "Van0C1_redacted.aff"
linkingAgentIdentifier	forensics tool
linkingAgentIdentifier	BiCurator
linkingAgentIdentifier	kamwoods
linkingObjectIdentifier	UUID
linkingObjectIdentifier	8a8215bd-3068-45f2-929d-7042dc46fb14

SAA Research Forum, August 13, 2013, New Orleans, LA Andrew W. Mellon Foundation

Visit http://wiki.bitcurator.net/

@bitcurator