Using Big DFIR Data in Autopsy and Other Tools

Brian Carrier

Basis Technology

SANS 2020 DFIR Summit

Copyright © Basis Technology 2020

It’s All About Efficiency

- We think about how to make examiners more efficient.
- Examples:
    - Automatically flag things previously tagged as notable
    - Prioritizing how files are displayed (not just by folder or alphabetical)
    - Giving context about an item (how and where it was used)

- All of these use data you might be “throwing away”.
- We’re going to talk about “re-using” your data.

Copyright © Basis Technology 2020

It’s All About History

Those who cannot remember the artifacts they saw before are condemned to analyze them again

Carrier - 2020

Copyright © Basis Technology 2020

It’s All About History

Those who cannot remember the artifacts they saw before are condemned to analyze them again

Carrier - 2020

Those who cannot remember the past are condemned to repeat it

George Santayana - 1905

Copyright © Basis Technology 2020

The Past is Important

- So much of what we do is based on the past.
    - We are trained to do things based on what was seen in the past.
    - Our tools parse data that we useful before in past cases
    - The NIST NSRL contains hashes of files that someone in the past processed
    - Hash sets of child exploitation material are from past cases
    - IOCs are from past cases
    - Topic-based keywords (drug terms, etc.) are based on past experience.
    - ….
- A lot of digital investigations is about applying past knowledge to the current case.

Copyright © Basis Technology 2020

Problem: Scaling

- It’s hard to remember all past notable things.
  - It’s even harder to know what your colleagues saw

- We’ve found it’s also important to remember the boring things.
  - There is A LOT of boring stuff

Solution

- Make your tools do the remembering.
- Save as much data as you can.

- Let’s look at how we’ve done this in:
    - Autopsy: Local Repository
    - Cyber Triage: Remote Global Repository

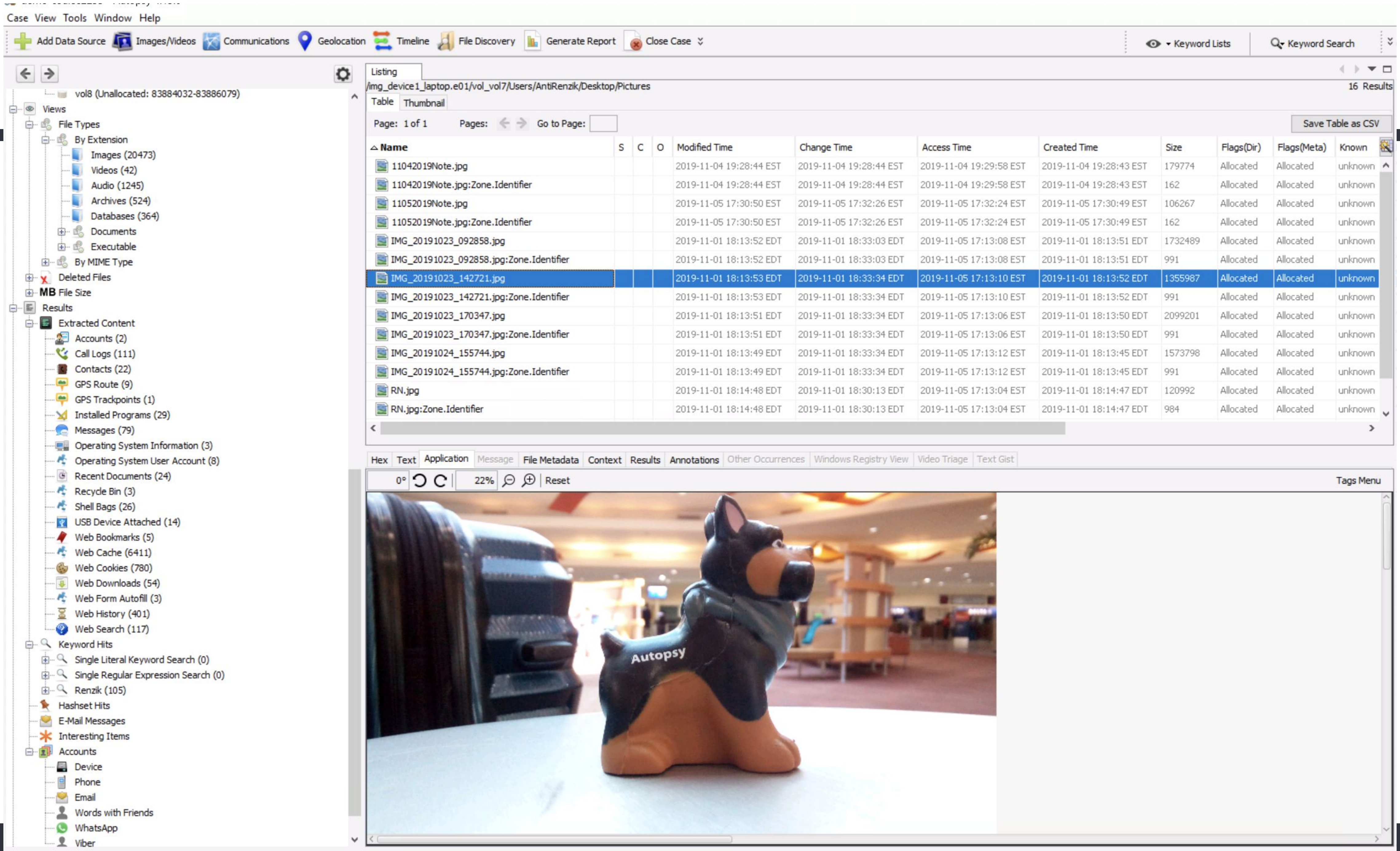
Copyright © Basis Technology 2020

Autopsy

What is Autopsy?

- Open source digital forensics platform.
- Designed to be:
  - Easy to use
  - Extensible with open plug-in frameworks
- Supports hard drives, media cards, and smart phone formats.
- Has all of the standard features, plus some unique features.

Copyright © Basis Technology 2020



vol8 (Unallocated: 83884032-83886079)
Views
File Types
By Extension
Images (20473)
Videos (42)
Audio (1245)
Archives (524)
Databases (364)
Documents
Executable
By MIME Type
Deleted Files
MB File Size
Results
Extracted Content
Accounts (2)
Call Logs (111)
Contacts (22)
GPS Route (9)
GPS Trackpoints (1)
Installed Programs (29)
Messages (79)
Operating System Information (3)
Operating System User Account (8)
Recent Documents (24)
Recycle Bin (3)
Shell Bags (26)
USB Device Attached (14)
Web Bookmarks (5)
Web Cache (6411)
Web Cookies (780)
Web Downloads (54)
Web Form Autofill (3)
Web History (401)
Web Search (117)
Keyword Hits
Single Literal Keyword Search (0)
Single Regular Expression Search (0)
Renzik (105)
Hashset Hits
E-Mail Messages
Interesting Items
Accounts
Device
Phone
Email
Words with Friends
WhatsApp
Viber

/img_device1_laptop.e01/vol_vol7/Users/AntiRenzik/Desktop/Pictures

16 Results

Page: 1 of 1 Pages: Go to Page:

Name	S	C	O	Modified Time	Change Time	Access Time	Created Time	Size	Flags(Dir)	Flags(Meta)	Known
11042019Note.jpg				2019-11-04 19:28:44 EST	2019-11-04 19:28:44 EST	2019-11-04 19:29:58 EST	2019-11-04 19:28:43 EST	179774	Allocated	Allocated	unknown
11042019Note.jpg:Zone.Identifier				2019-11-04 19:28:44 EST	2019-11-04 19:28:44 EST	2019-11-04 19:29:58 EST	2019-11-04 19:28:43 EST	162	Allocated	Allocated	unknown
11052019Note.jpg				2019-11-05 17:30:50 EST	2019-11-05 17:32:26 EST	2019-11-05 17:32:24 EST	2019-11-05 17:30:49 EST	106267	Allocated	Allocated	unknown
11052019Note.jpg:Zone.Identifier				2019-11-05 17:30:50 EST	2019-11-05 17:32:26 EST	2019-11-05 17:32:24 EST	2019-11-05 17:30:49 EST	162	Allocated	Allocated	unknown
IMG_20191023_092858.jpg				2019-11-01 18:13:52 EDT	2019-11-01 18:33:03 EDT	2019-11-05 17:13:08 EST	2019-11-01 18:13:51 EDT	1732489	Allocated	Allocated	unknown
IMG_20191023_092858.jpg:Zone.Identifier				2019-11-01 18:13:52 EDT	2019-11-01 18:33:03 EDT	2019-11-05 17:13:08 EST	2019-11-01 18:13:51 EDT	991	Allocated	Allocated	unknown
IMG_20191023_142721.jpg				2019-11-01 18:13:53 EDT	2019-11-01 18:33:34 EDT	2019-11-05 17:13:10 EST	2019-11-01 18:13:52 EDT	1355987	Allocated	Allocated	unknown
IMG_20191023_142721.jpg:Zone.Identifier				2019-11-01 18:13:53 EDT	2019-11-01 18:33:34 EDT	2019-11-05 17:13:10 EST	2019-11-01 18:13:52 EDT	991	Allocated	Allocated	unknown
IMG_20191023_170347.jpg				2019-11-01 18:13:51 EDT	2019-11-01 18:33:34 EDT	2019-11-05 17:13:06 EST	2019-11-01 18:13:50 EDT	2099201	Allocated	Allocated	unknown
IMG_20191023_170347.jpg:Zone.Identifier				2019-11-01 18:13:51 EDT	2019-11-01 18:33:34 EDT	2019-11-05 17:13:06 EST	2019-11-01 18:13:50 EDT	991	Allocated	Allocated	unknown
IMG_20191024_155744.jpg				2019-11-01 18:13:49 EDT	2019-11-01 18:33:34 EDT	2019-11-05 17:13:12 EST	2019-11-01 18:13:45 EDT	1573798	Allocated	Allocated	unknown
IMG_20191024_155744.jpg:Zone.Identifier				2019-11-01 18:13:49 EDT	2019-11-01 18:33:34 EDT	2019-11-05 17:13:12 EST	2019-11-01 18:13:45 EDT	991	Allocated	Allocated	unknown
RN.jpg				2019-11-01 18:14:48 EDT	2019-11-01 18:30:13 EDT	2019-11-05 17:13:04 EST	2019-11-01 18:14:47 EDT	120992	Allocated	Allocated	unknown
RN.jpg:Zone.Identifier				2019-11-01 18:14:48 EDT	2019-11-01 18:30:13 EDT	2019-11-05 17:13:04 EST	2019-11-01 18:14:47 EDT	984	Allocated	Allocated	unknown

0° 22%

Features

Standard

- Hash calculation and lookup
- Indexed keyword search
- Registry analysis
- Web artifacts
- Email
- Carving
- …..

Unique

- Multi-user collaborative cases
- Automatically analyze data 24x7
- Analysis-driven acquisition
- Triage
- Timeline
- Communications
- …..

Copyright © Basis Technology 2020

Autopsy’s Short-term Memory

- Data is compartmentalized by case to keep data sets small.
- Each case has its own:
  - Folder
  - Database (SQLite or PostgreSQL)
  - Solr index
  - …..
- You can (statically) import past knowledge with:
  - Hash sets
  - Keyword lists
  - Interesting item rules (file name rules)

Copyright © Basis Technology 2020

Autopsy’s Long-term Memory

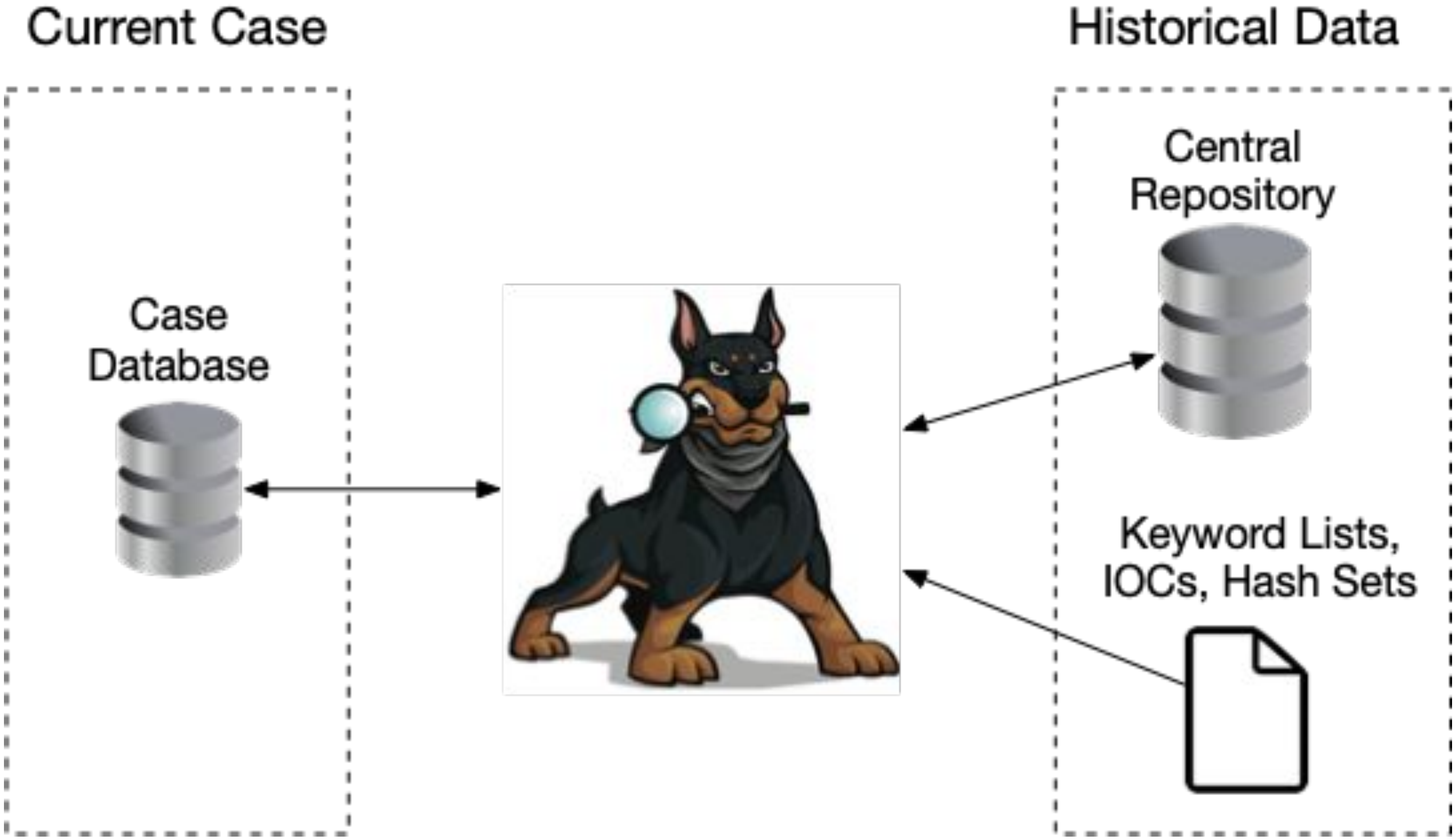
- The Central Repository spans cases and is dynamically updated.
- Can be single-user (SQLite) or multi-user (PostgreSQL)

- It stores:
  - Identifiers from past cases:
    - Hashes, Emails, USB Device IDs, Wifi SSID, ICCID, Domains, etc.
  - Comments
  - Tags

- First released in 2017 and now enabled by default

Copyright © Basis Technology 2020

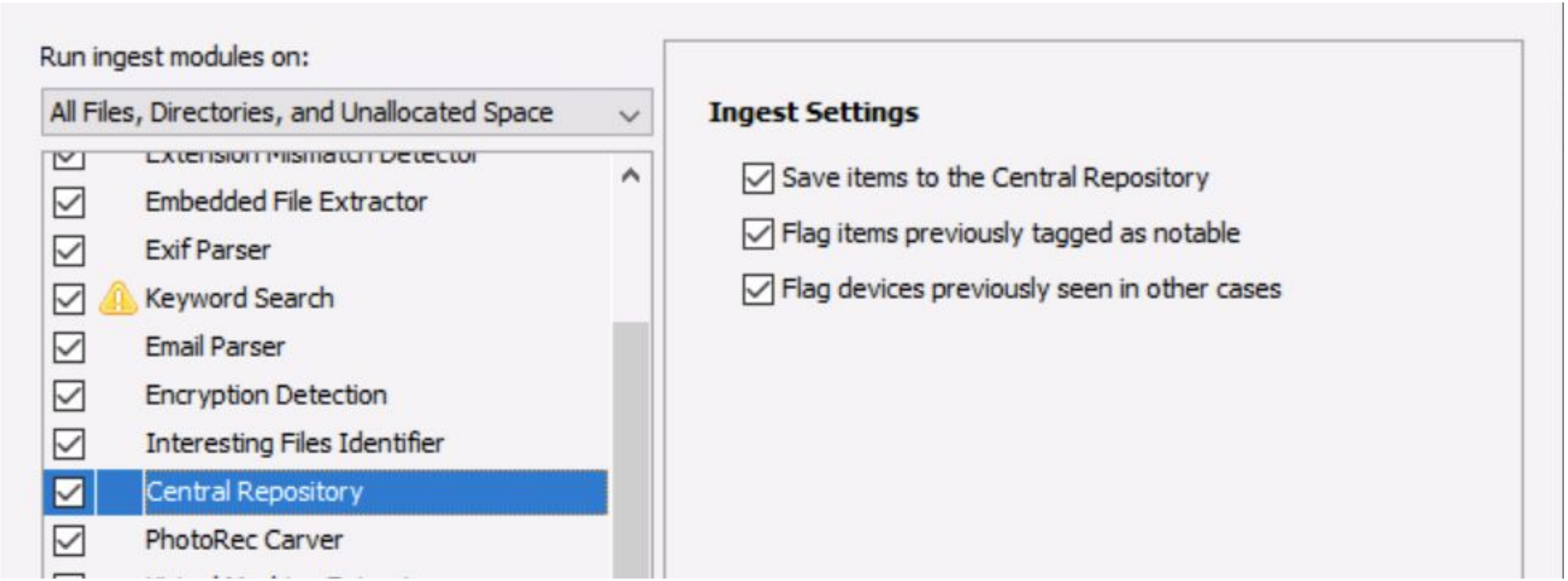
Architecture



Copyright © Basis Technology 2020

How It Is Updated

The Central Repository ingest module saves hash values and identifiers during ingest (previously called “Correlation Engine”)



Run ingest modules on:

All Files, Directories, and Unallocated Space

- ☑ Extension Mismatch Detector
- ☑ Embedded File Extractor
- ☑ Exif Parser
- ☑ ⚠ Keyword Search
- ☑ Email Parser
- ☑ Encryption Detection
- ☑ Interesting Files Identifier
- ☑ Central Repository
- ☑ PhotoRec Carver

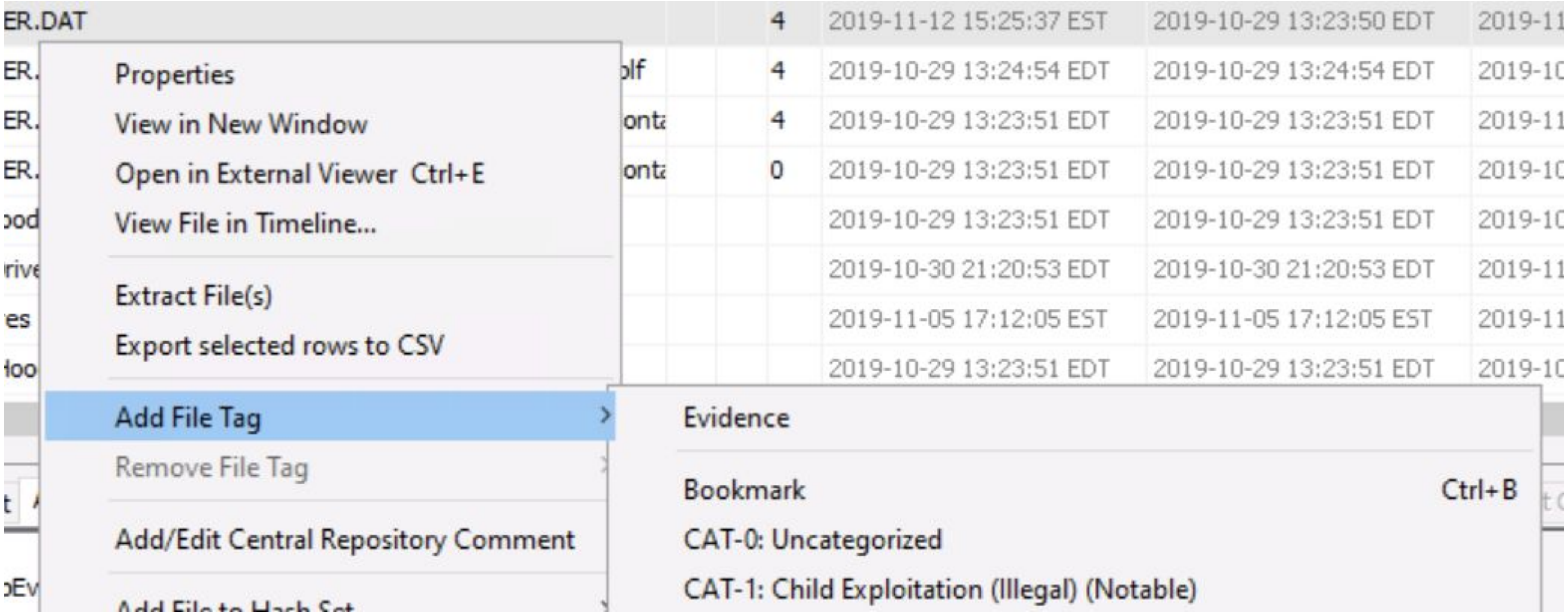
Ingest Settings

- ☑ Save items to the Central Repository
- ☑ Flag items previously tagged as notable
- ☑ Flag devices previously seen in other cases

Copyright © Basis Technology 2020

How It Is Updated (2)

When you tag an item, the entry in the Central Repository is updated.

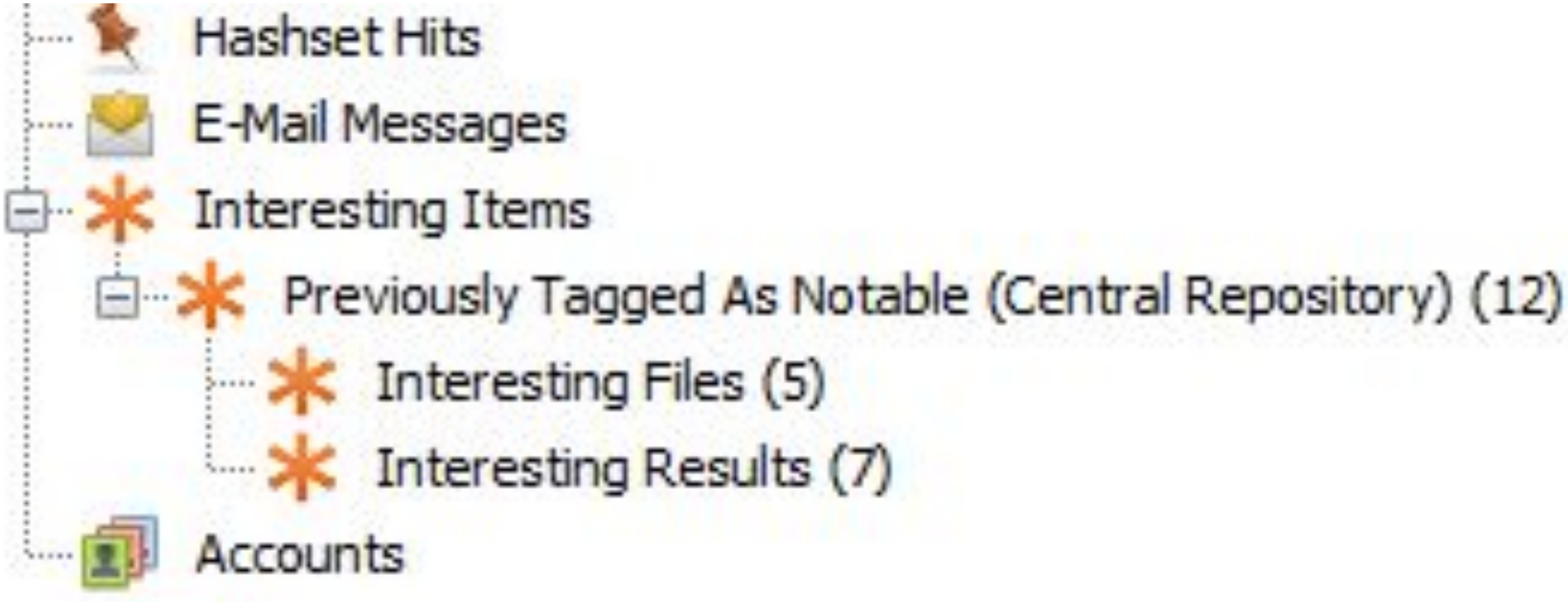


ER.DAT		4	2019-11-12 15:25:37 EST	2019-10-29 13:23:50 EDT	2019-11
ER.	olf	4	2019-10-29 13:24:54 EDT	2019-10-29 13:24:54 EDT	2019-1C
ER.	ontا	4	2019-10-29 13:23:51 EDT	2019-10-29 13:23:51 EDT	2019-11
ER.	ontا	0	2019-10-29 13:23:51 EDT	2019-10-29 13:23:51 EDT	2019-1C
ood			2019-10-29 13:23:51 EDT	2019-10-29 13:23:51 EDT	2019-1C
rive			2019-10-30 21:20:53 EDT	2019-10-30 21:20:53 EDT	2019-11
es			2019-11-05 17:12:05 EST	2019-11-05 17:12:05 EST	2019-11
ioo			2019-10-29 13:23:51 EDT	2019-10-29 13:23:51 EDT	2019-1C

Copyright © Basis Technology 2020

How It Is Used: Remembering Notability

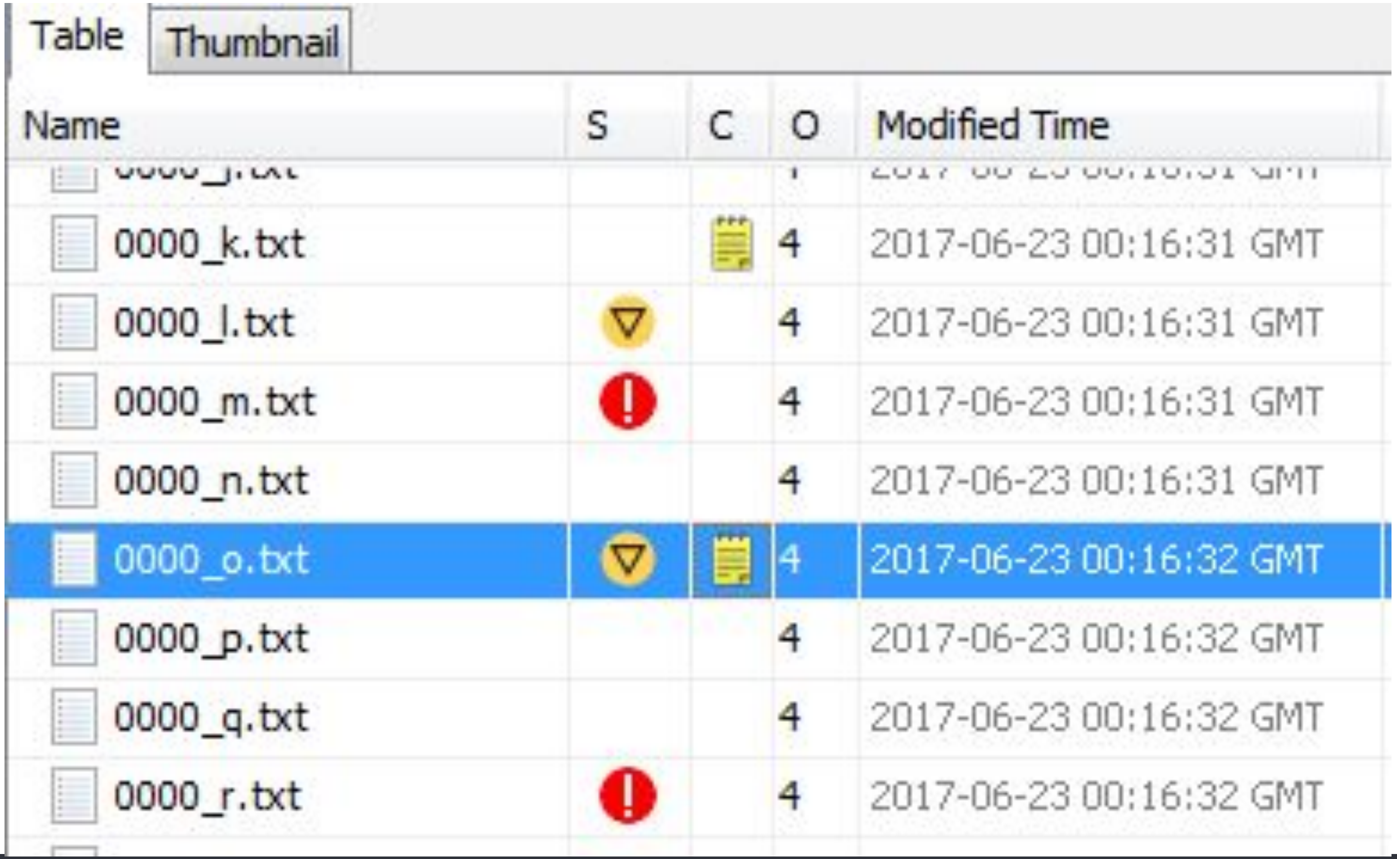
- Automatically flag files that were marked as “Notable” in a past case
  - A dynamic hash set



Copyright © Basis Technology 2020

How It Is Used: Past Occurrences (table)

- Show how often a file was seen in the past
  - The ‘O’ column is for Occurance

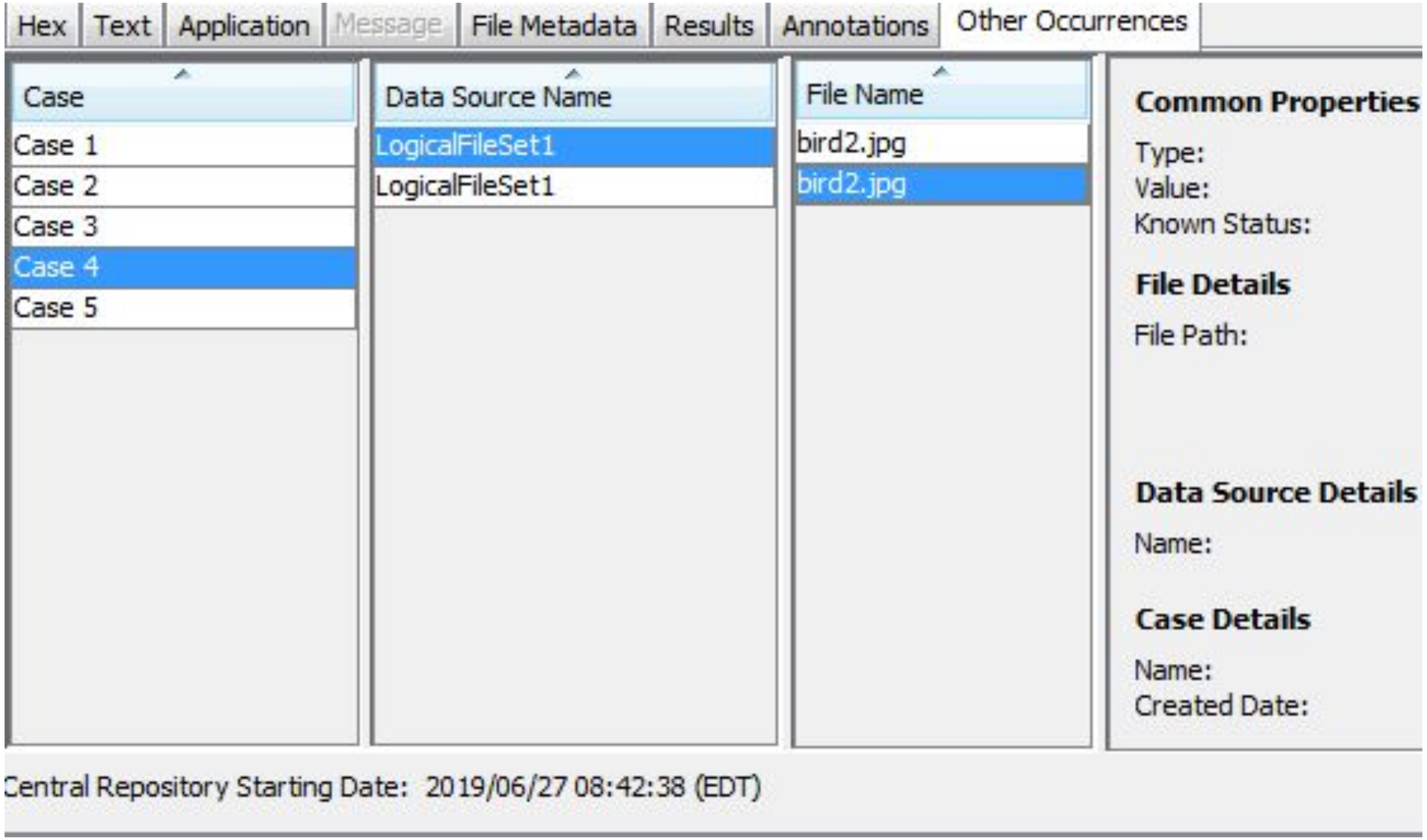


Name	S	C	O	Modified Time
0000_j.txt			4	2017-06-23 00:16:31 GMT
0000_k.txt		📝	4	2017-06-23 00:16:31 GMT
0000_l.txt	▽		4	2017-06-23 00:16:31 GMT
0000_m.txt	!		4	2017-06-23 00:16:31 GMT
0000_n.txt			4	2017-06-23 00:16:31 GMT
0000_o.txt	▽	📝	4	2017-06-23 00:16:32 GMT
0000_p.txt			4	2017-06-23 00:16:32 GMT
0000_q.txt			4	2017-06-23 00:16:32 GMT
0000_r.txt	!		4	2017-06-23 00:16:32 GMT

Copyright © Basis Technology 2020

How It Is Used: Past Occurrences (viewer)

Show how often a file was seen in the past and where.



Case
Case 1
Case 2
Case 3
Case 4
Case 5

Data Source Name
LogicalFileSet1
LogicalFileSet1

File Name
bird2.jpg
bird2.jpg

Common Properties

Type:
Value:
Known Status:

File Details

File Path:

Data Source Details

Name:

Case Details

Name:
Created Date:

Central Repository Starting Date: 2019/06/27 08:42:38 (EDT)

Copyright © Basis Technology 2020

New Use: Ranking

- We’ve been focusing on showing the most relevant files first.
- General theory:
  - If you saw a file 10 times before and didn’t think it was relevant, it’s probably not relevant the 11th time either.
  - A file you’ve seen 50 times before is less relevant than a file you’ve seen 5 times before.
  - ….
- The Central Repository data allowed us to implement this and deprioritize the boring stuff.

Copyright © Basis Technology 2020

New UI: File Discovery

- A new search UI in Autopsy.
- Goal is to allow user to define what they are looking for
  - NOTE: This is an incrementally evolving feature that changes each quarter
- User picks:
  - Features they care about
  - How they want to see the results
- The Central Repository allows the user to search or display by past occurrence.

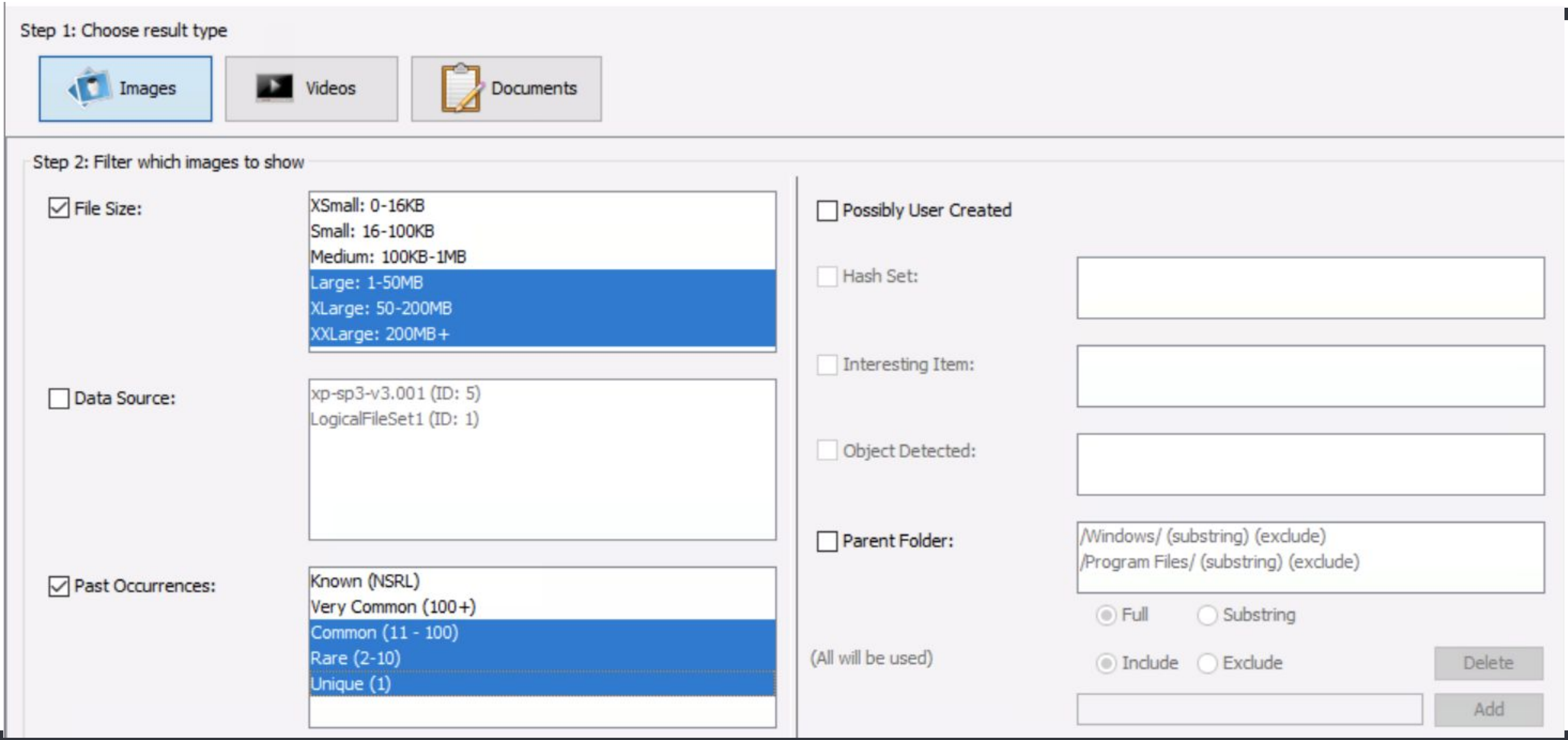
Copyright © Basis Technology 2020

File Discovery: Example Queries for Pictures

- Show all unique or rare pictures that are big. Organized by parent folder.
    - i.e. Focus on possibly user created images. Organized by how the owner organized them.
- Show all big pictures. Organize by frequency to focus on unique files first.
    - I.e. Focus on all high res pictures (including ones from past cases), but focus first on unique ones.
- ….

Copyright © Basis Technology 2020

File Discovery: Pick Type and Criteria



Step 1: Choose result type

Step 2: Filter which images to show

☑ File Size:
- XSmall: 0-16KB
- Small: 16-100KB
- Medium: 100KB-1MB
- [✓ Large: 1-50MB]
- [✓ XLarge: 50-200MB]
- [✓ XXLarge: 200MB+]

□ Data Source:
- xp-sp3-v3.001 (ID: 5)
- LogicalFileSet1 (ID: 1)

☑ Past Occurrences:
- Known (NSRL)
- Very Common (100+)
- [✓ Common (11 - 100)]
- [✓ Rare (2-10)]
- [✓ Unique (1)]

□ Possibly User Created

□ Hash Set:

□ Interesting Item:

□ Object Detected:

□ Parent Folder:
- /Windows/ (substring) (exclude)
- /Program Files/ (substring) (exclude)

● Full ○ Substring

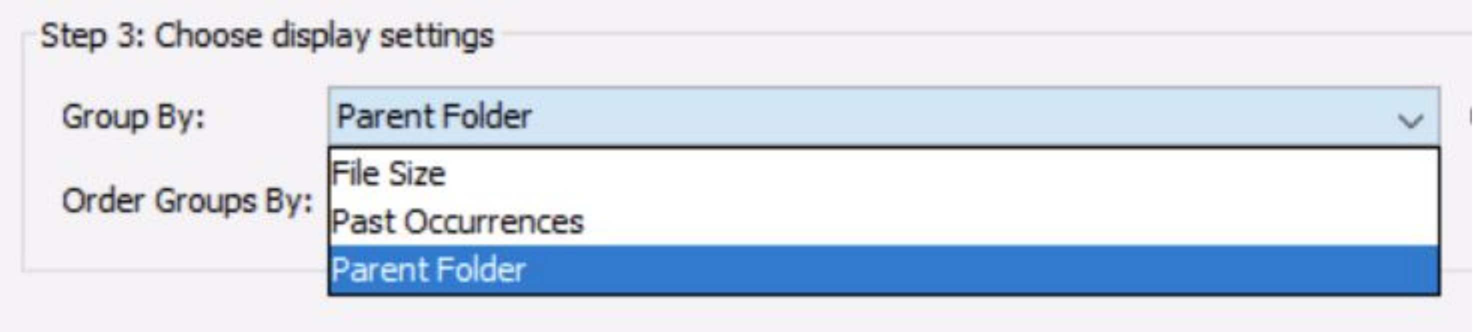
(All will be used)

● Include ○ Exclude

Copyright © Basis Technology 2020

File Discovery: Pick Display Options

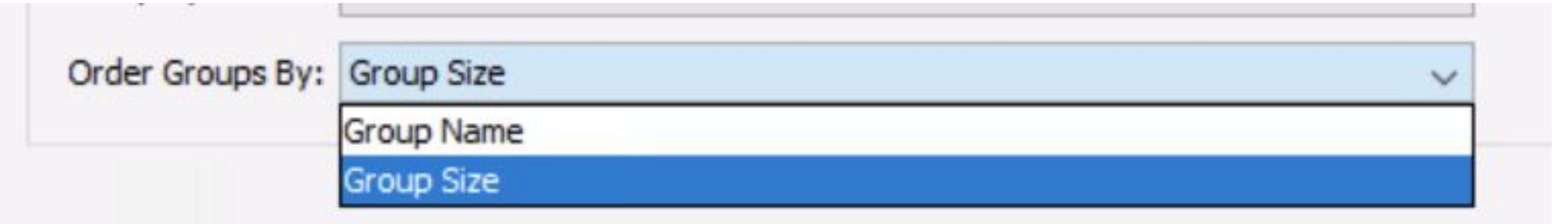
- Results are grouped to make it easier to organize:



Step 3: Choose display settings

Group By: Parent Folder

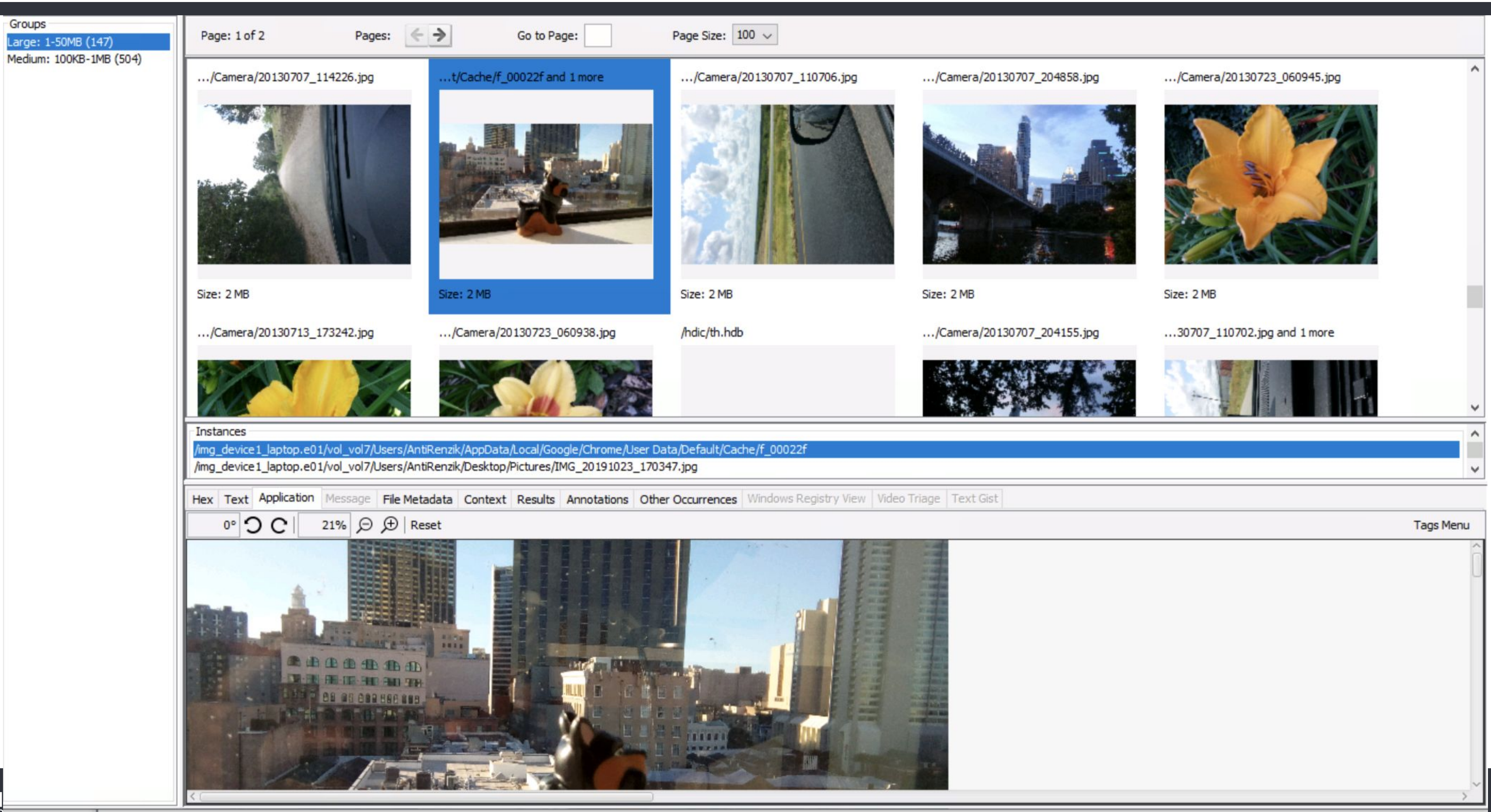
- You can pick the order the groups are displayed:



Order Groups By: Group Size

Copyright © Basis Technology 2020

Example Picture Results, Grouped By Size



/img_device1_laptop.e01/vol_vol7/Users/AntiRenzik/AppData/Local/Google/Chrome/User Data/Default/Cache/f_00022f
/img_device1_laptop.e01/vol_vol7/Users/AntiRenzik/Desktop/Pictures/IMG_20191023_170347.jpg

Cyber Triage
(Global Repository)

Central Repository is YOUR History

- The Autopsy Central Repository knows only what you've put into it.
- That's good and bad.
- Bad:
  - You haven't seen everything before
  - May take a while to build up enough data
- Good:
  - Something is better than nothing
  - Many labs are offline and can't access a global repository

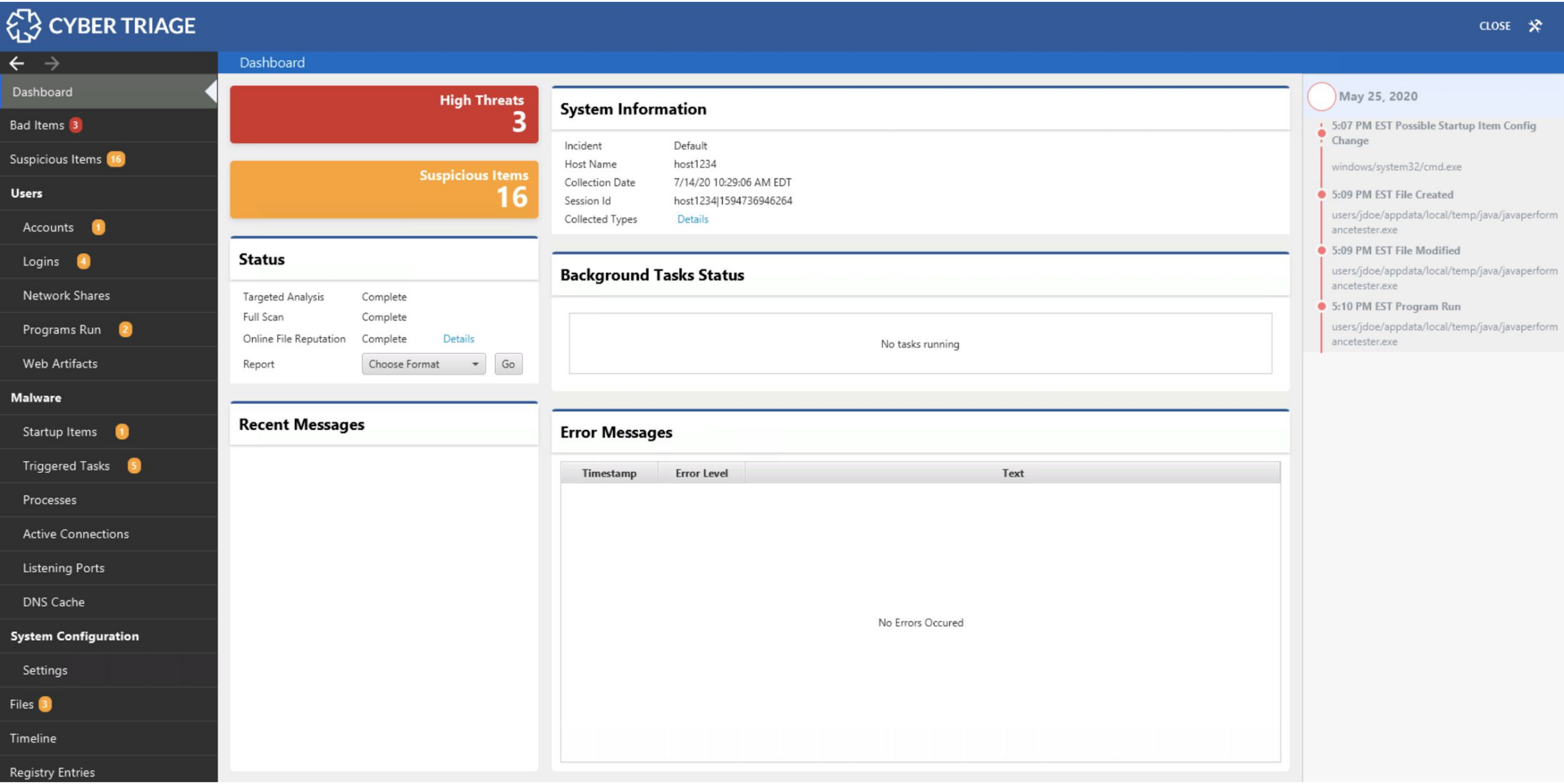
Copyright © Basis Technology 2020

What is Cyber Triage?

- Automated intrusion forensics tool.
- Hyper-focused on intrusion-related artifacts.
  - Not general purpose like Autopsy
- Collects select artifacts from a live system.
  - Start up, program run, web artifacts, WMI actions, logins, network, etc.
- Automatically scores the artifacts as bad or suspicious.
- User reviews high threat items and dives in.

CYBER TRIAGE

Copyright © Basis Technology 2020



High Threats 3

Suspicious Items 16

Status

Targeted Analysis	Complete	
Full Scan	Complete	
Online File Reputation	Complete	Details
Report	Choose Format ▾	Go

Recent Messages

System Information

Incident	Default
Host Name	host1234
Collection Date	7/14/20 10:29:06 AM EDT
Session Id	host1234|1594736946264
Collected Types	Details

Background Tasks Status

No tasks running

Error Messages

Timestamp	Error Level	Text

No Errors Occured

May 25, 2020

5:07 PM EST Possible Startup Item Config Change
windows/system32/cmd.exe

5:09 PM EST File Created
users/jdoe/appdata/local/temp/java/javaperformancetester.exe

5:09 PM EST File Modified
users/jdoe/appdata/local/temp/java/javaperformancetester.exe

5:10 PM EST Program Run
users/jdoe/appdata/local/temp/java/javaperformancetester.exe

Copyright © Basis Technology 2020

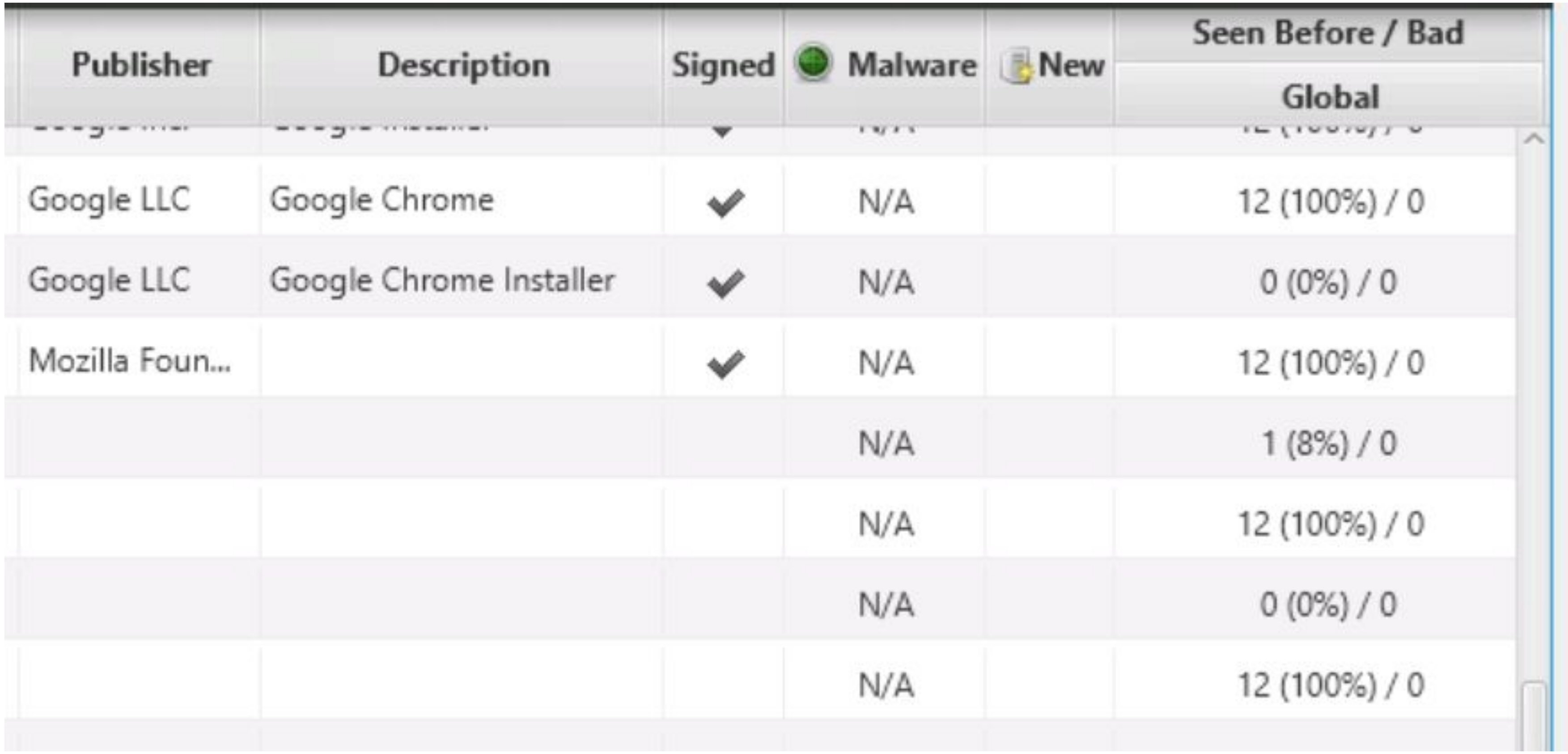
Cyber Triage's Memory

- Similar concepts as Autopsy
  - Databases for storing artifacts.
  - Remembers your past scores / tags, comments, etc.
- When you look at an artifact, it will tell you:
  - Other past cases it was seen in
  - If it was flagged as "Bad" in the past.
- Helps you to determine:
  - Is this artifact unique to this system and possibly part of the attack?
  - Other systems that could be compromised
  - …..

Copyright © Basis Technology 2020

Past Frequency

- Each row shows if it was seen before and if it was marked as bad.



Publisher	Description	Signed	Malware	New	Seen Before / Bad
					Global
Google Inc.	Google Installer	✓	N/A		12 (100%) / 0
Google LLC	Google Chrome	✓	N/A		12 (100%) / 0
Google LLC	Google Chrome Installer	✓	N/A		0 (0%) / 0
Mozilla Foun...		✓	N/A		12 (100%) / 0
			N/A		1 (8%) / 0
			N/A		12 (100%) / 0
			N/A		0 (0%) / 0
			N/A		12 (100%) / 0

Copyright © Basis Technology 2020

But, I Want More Data

- Finding outliers (unique instances) is critical in incident response.
  - Unique processes or startup items should be reviewed.

- Sometimes your past cases aren’t enough
  - You may not do many investigations
  - No one has seen everything

- Wouldn’t it be useful to know how common or rare something is amongst others in the industry?

Copyright © Basis Technology 2020

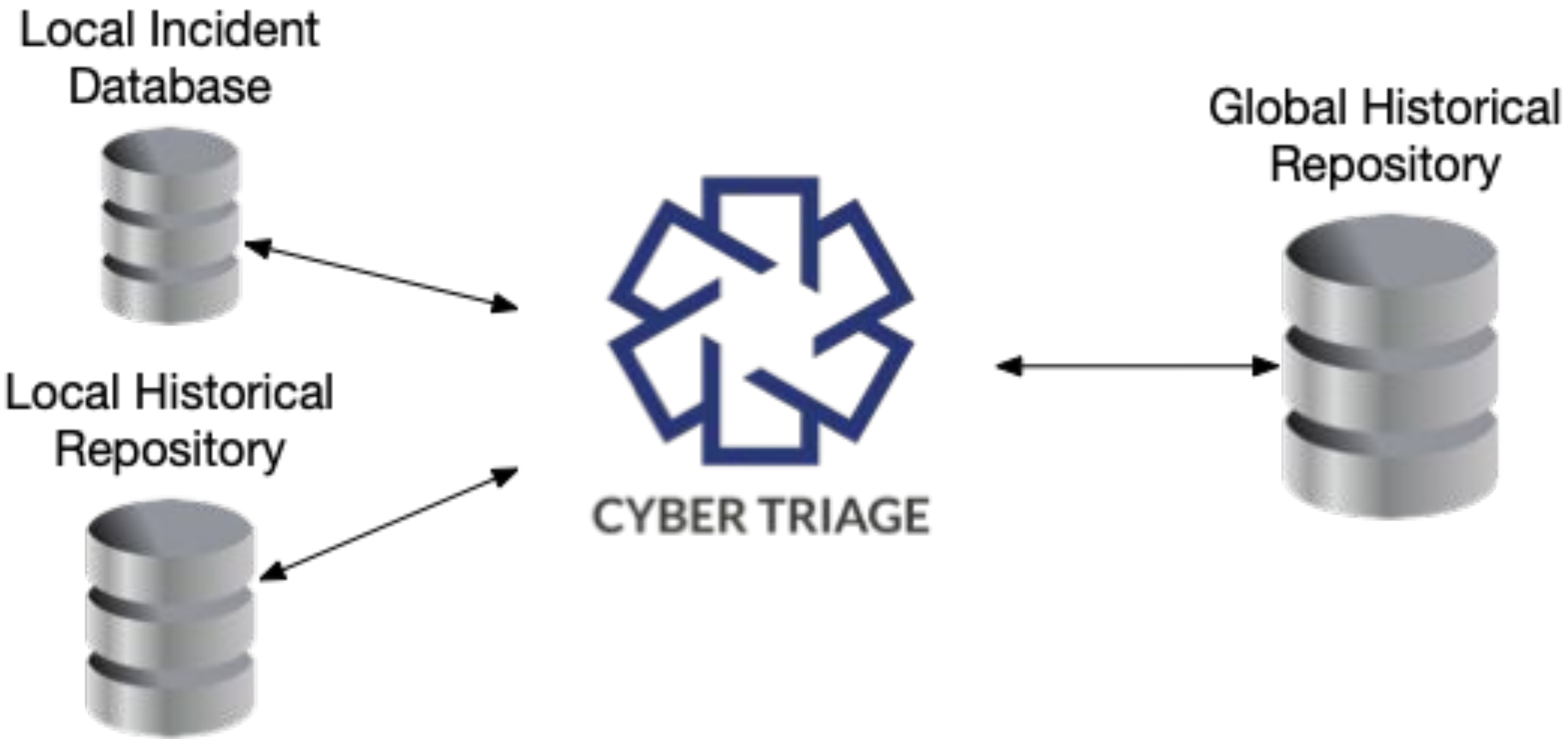
Global Repository For File Hashes

- Cyber Triage is building up a global repository for frequency analysis.

- Cyber Triage has an online file reputation service:
  - Identifies a file as good or bad
  - Backed by 40+ malware scanning engines at ReversingLabs
- It stores anonymous data about hash frequency
- It will soon provide global frequency results:
  - Unique, rare, common, etc.

Copyright © Basis Technology 2020

Architecture



Copyright © Basis Technology 2020

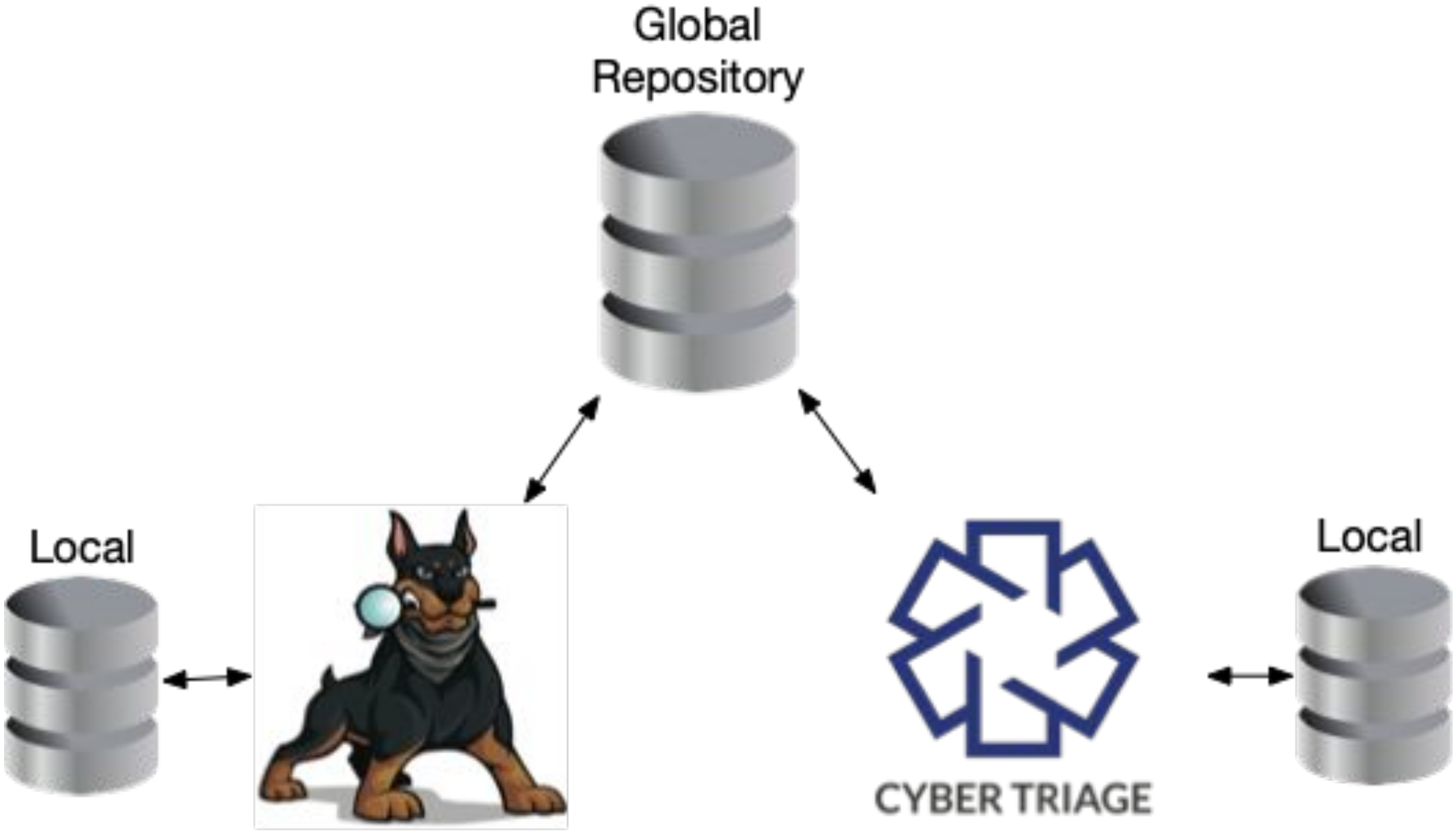
Global Repository for Other Artifacts

- It will expand beyond file hashes (start up items, processes, etc.)
- We’ll be defining anonymized “hash” functions for other data types.

- Possible Example
  - Type: Startup Item
  - Normalize the path and hash the string
  - Path: C:\Users\jdoe\AppData\Local\Temp\BLAH.EXE
  - Hash: SHA256(\users\REMOVED\appdata\local\temp\blah.exe)

Copyright © Basis Technology 2020

Coming Soon: Autopsy Can Use Global Repository



Copyright © Basis Technology 2020

Summary

- Saving your data is key to solving your future big data problems.
- Relevance and ranking are a big part of the data overload problem.

- Don’t throw away your data - reuse it.

Those who cannot remember the artifacts they saw before are condemned to analyze them again

Carrier - 2020

Copyright © Basis Technology 2020

OSDFCon

- 1-day event dedicated to open source software.
  - October 21, 2020
- It will be virtual this year
- Agenda is still being figured out
- Topics typically include incident response, memory forensics, Correlation, and more
- Free for US Government employees.

http://www.osdfcon.org/

Copyright © Basis Technology 2020

Online Training

- Autopsy: There is an 8-hour training available online.
  - http://training.autopsy.com
  - 100K people enrolled during our free COVID offering!
  - Free for US Law Enforcement

- Cyber Triage:
  - A free 3-hour “Intro to DFIR” training is coming next month.
  - An 8-hour hands-on training is coming in the Fall.

Copyright © Basis Technology 2020

Downloads and Contact

Free Download

www.autopsy.com

Free Evaluation

www.cybertriage.com

Brian Carrier

brianc <at> basistech <dot> com

Connect on LinkedIn or Twitter

Copyright © Basis Technology 2020