DFRWS EU 2024

An abstract model for digital forensic analysis tools:

A foundation for systematic error mitigation analysis

Chris Hargreaves, University of Oxford, UK

Eoghan Casey, University of Lausanne, Switzerland

Alex Nelson, NIST, US

20th March 2024

Paper: Hargreaves, C., Nelson, A. and Casey, E., 2024. An abstract model for digital forensic analysis tools-A foundation for systematic error mitigation analysis. Forensic Science International: Digital Investigation, 48, p.301679.

NIST Disclaimer

- The views and opinions expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of any agency of the U.S. government or any other organization. Any mention of a vendor or product is not an endorsement or recommendation. Logos and trademarks are copyright their respective owners.

Motivation

Need for automation

- Volume of data
- Volume of devices
- Volume of cases
- Complexity of interpretation
- Increasing use artificial intelligence

Need for demonstrably correct results

- “the courts have the expectation that the methods to produce the data that an expert bases their opinion on are valid” (UK Forensic Science Regulator 2020)
- The need for auditable automated forensic processes
- Need to stop silent failures

Solutions?

Open source tools

Dual tool verification

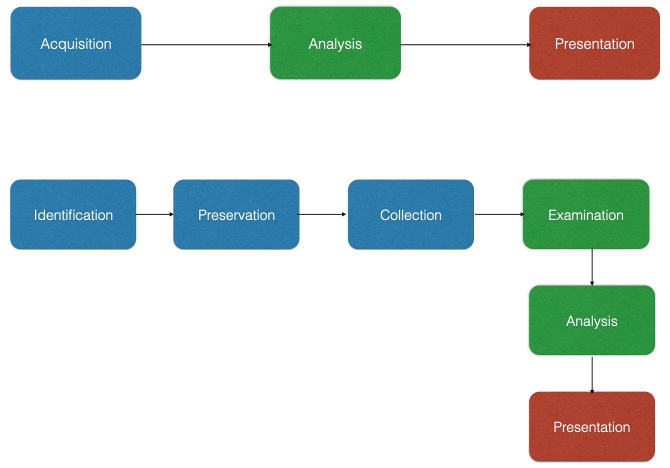
Education and Training

Validation Programmes

Lab Information Management Systems (LIMS)

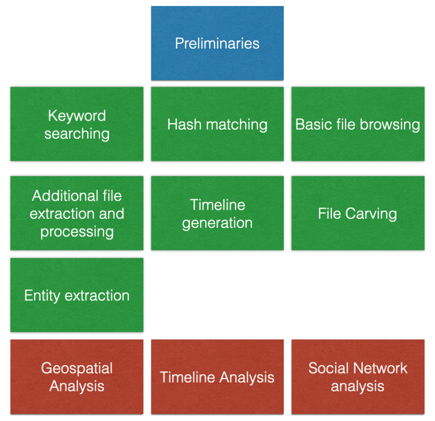
Custom In-House Mechanisms

Deconstructing forensic analysis process



Acquisition → Analysis → Presentation

Identification → Preservation → Collection → Examination → Analysis → Presentation



Preliminaries

Keyword searching

Hash matching

Basic file browsing

Additional file extraction and processing

Timeline generation

File Carving

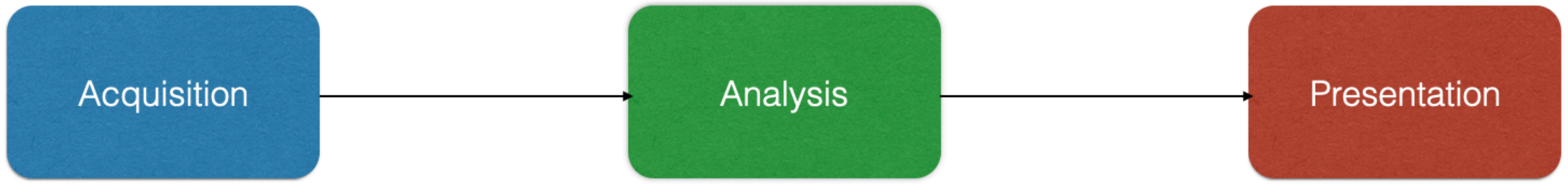
Entity extraction

Geospatial Analysis

Timeline Analysis

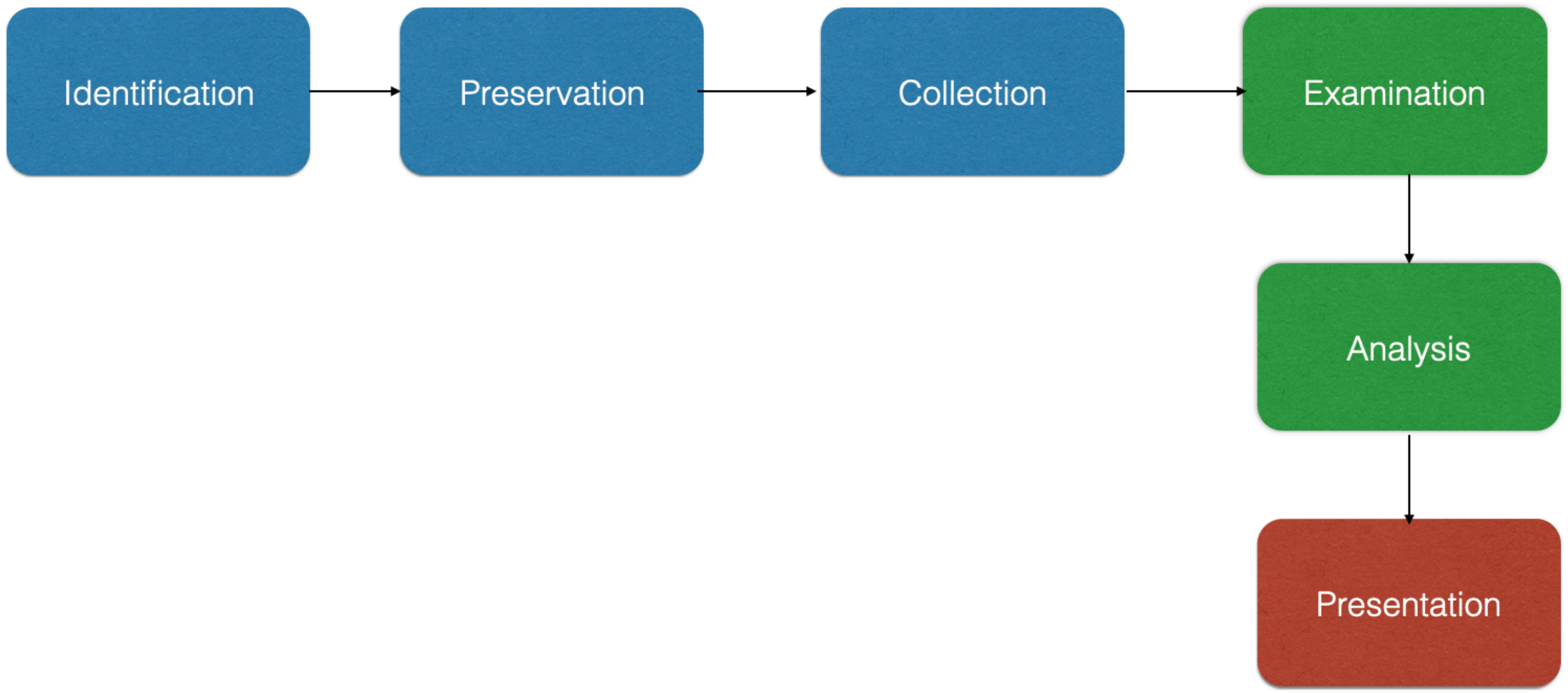
Social Network analysis

Deconstructing forensic analysis process



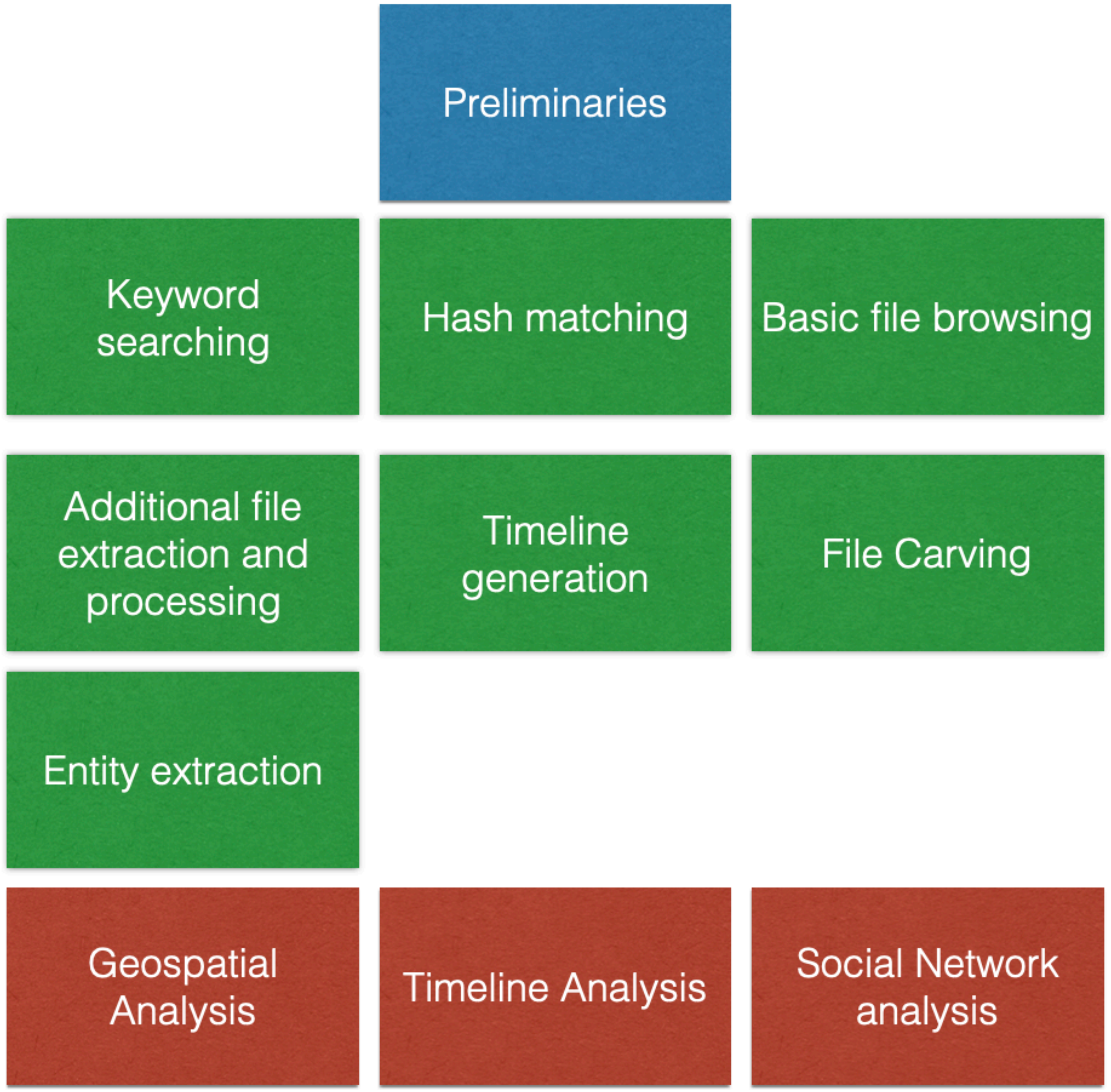
Acquisition → Analysis → Presentation

“Open Source Digital Forensics Tools: The Legal Argument” Carrier (2002)



Identification → Preservation → Collection → Examination → Analysis → Presentation

“A Road Map for Digital Forensic Research” Palmer/DFRWS (2001)



Preliminaries

Keyword searching

Hash matching

Basic file browsing

Additional file extraction and processing

Timeline generation

File Carving

Entity extraction

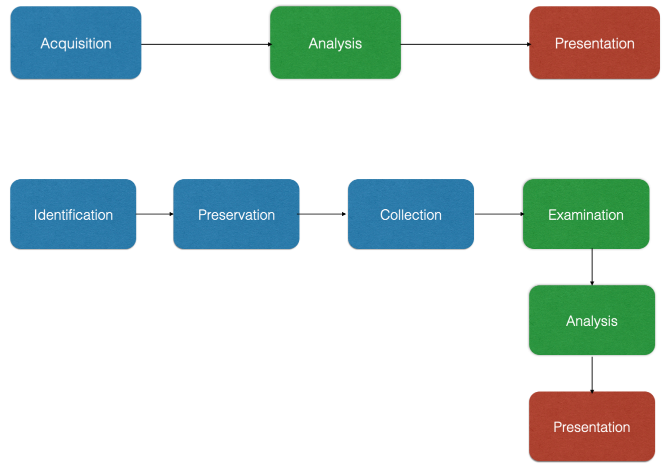
Geospatial Analysis

Timeline Analysis

Social Network analysis

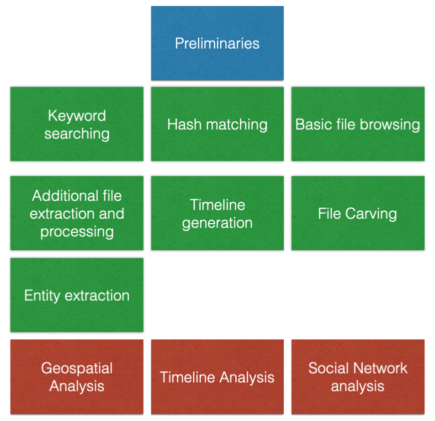
“Digital Forensic Analysis Techniques ”Hargreaves (undated)

Deconstructing forensic analysis process



Acquisition → Analysis → Presentation

Identification → Preservation → Collection → Examination → Analysis → Presentation



Preliminaries

Keyword searching

Hash matching

Basic file browsing

Additional file extraction and processing

Timeline generation

File Carving

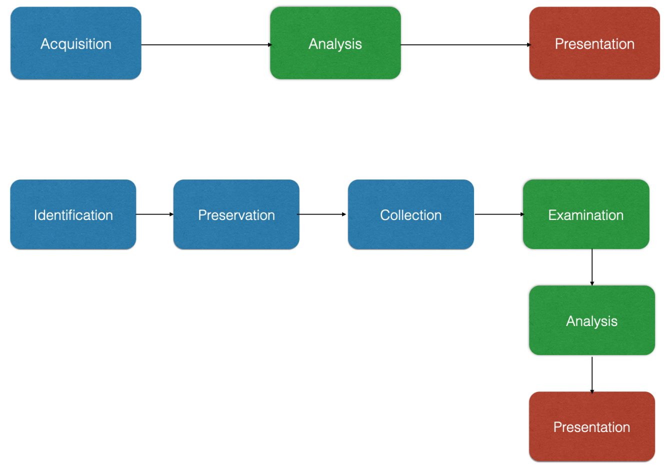
Entity extraction

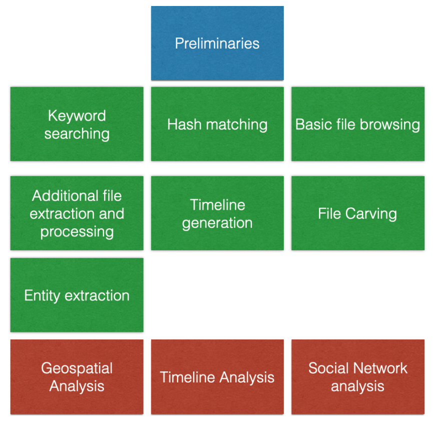
Geospatial Analysis

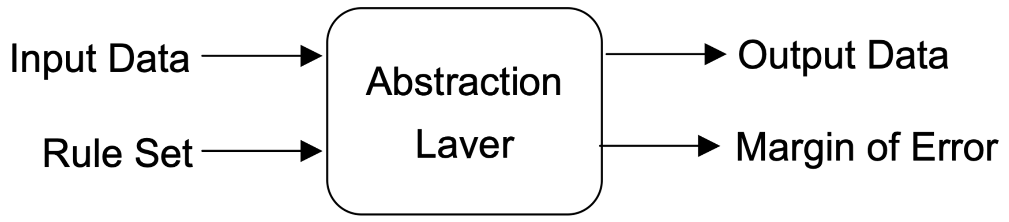
Timeline Analysis

Social Network analysis

Deconstructing forensic analysis process







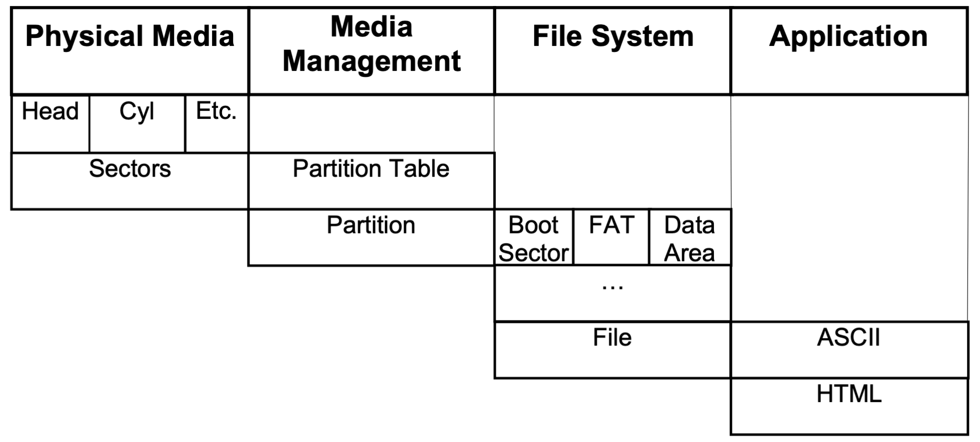
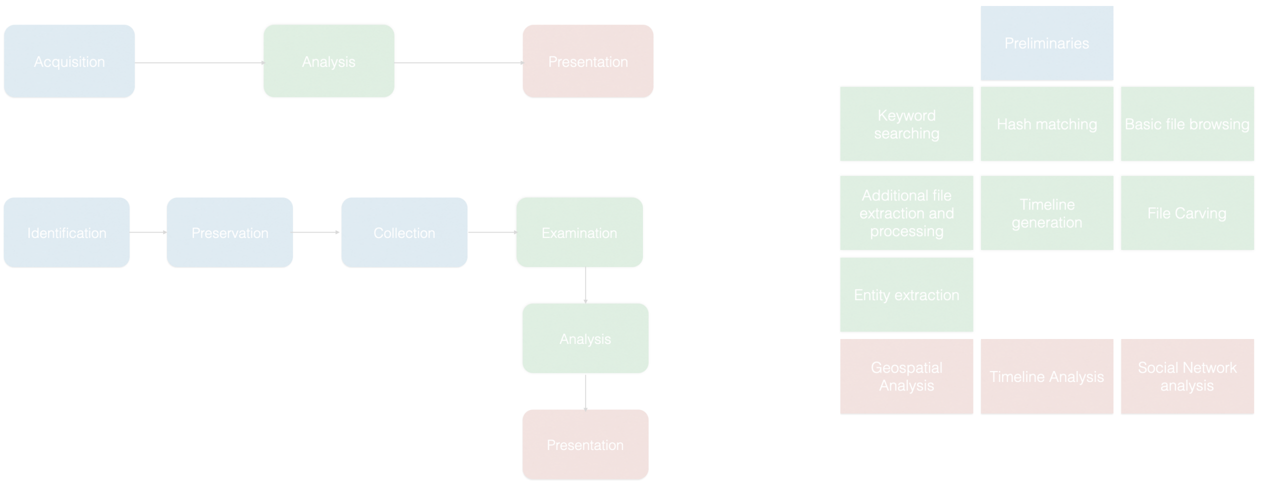


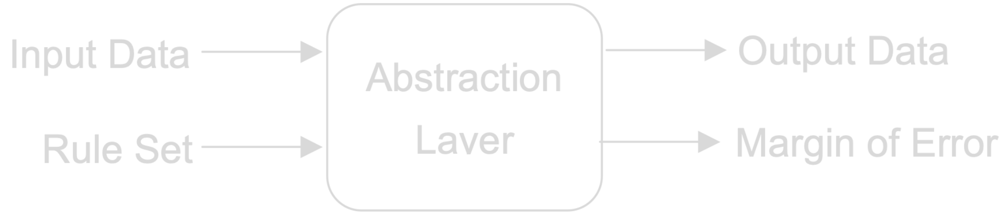
Figure 2: Abstraction Levels and Layers of an HTML File

Carrier (2003)

Abstract model of forensic analysis tools

Deconstructing forensic analysis process





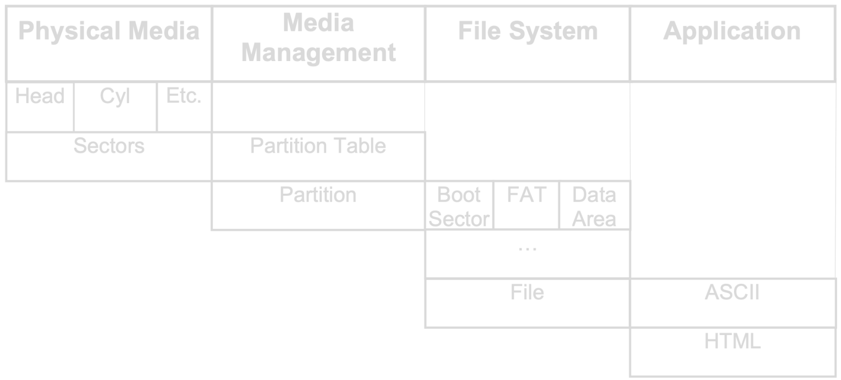


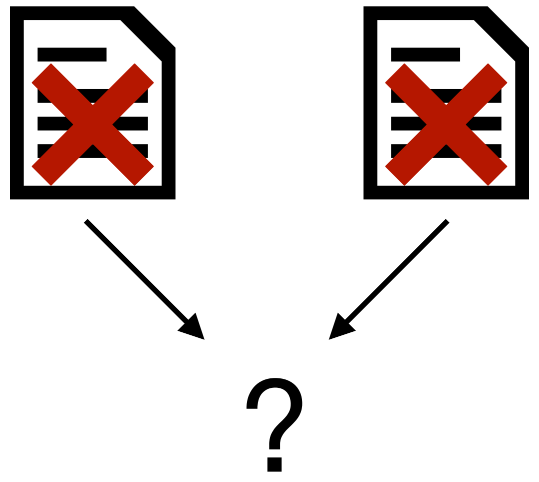
Figure 2: Abstraction Levels and Layers of an HTML File

Carrier (2003)

Carrier (2003)

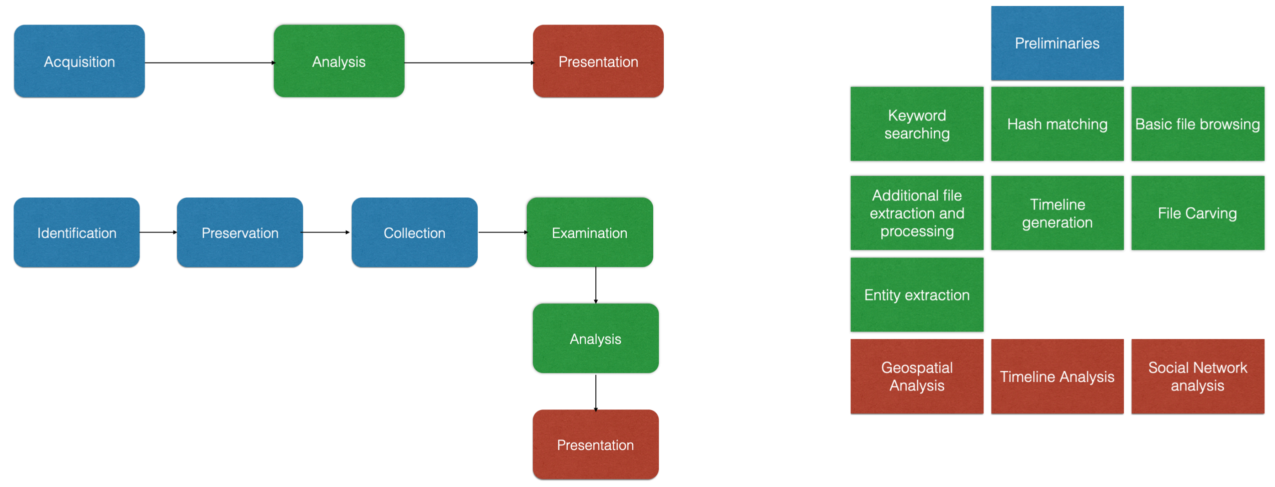
Abstract model of forensic analysis tools

Errors in digital forensic tools

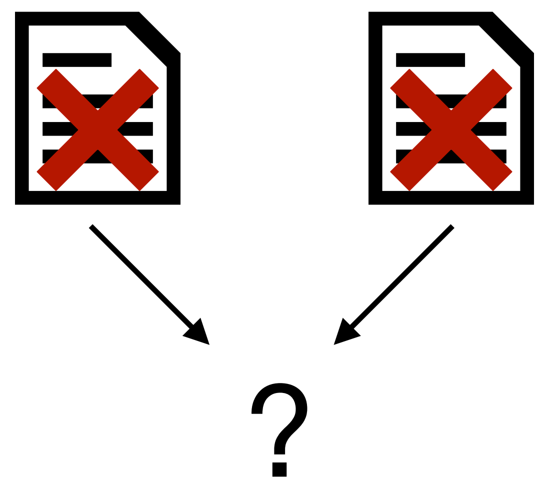


Standardised representation of digital forensic results (CASE)

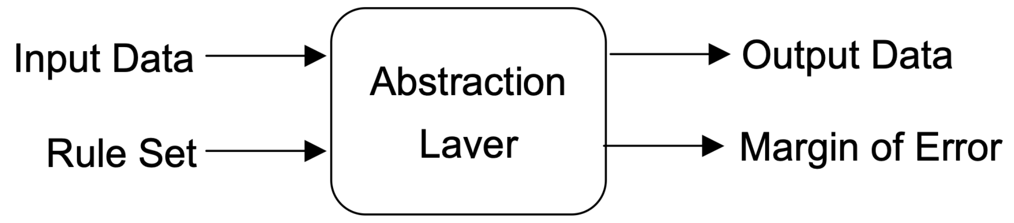
Deconstructing forensic analysis process



Errors in digital forensic tools



Standardised representation of digital forensic results (CASE)



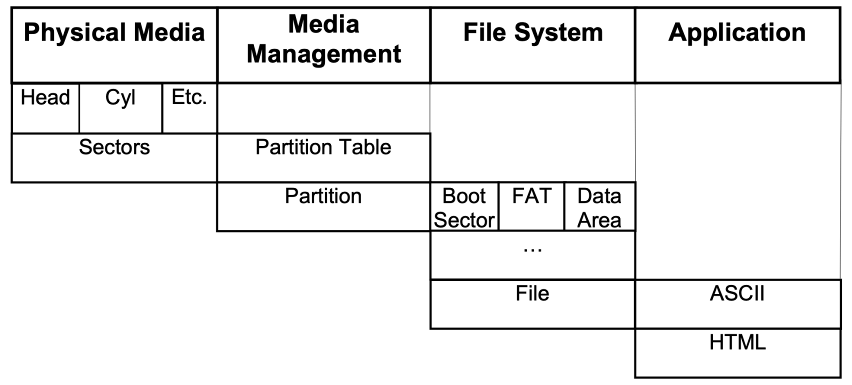


Figure 2: Abstraction Levels and Layers of an HTML File

Carrier (2003)

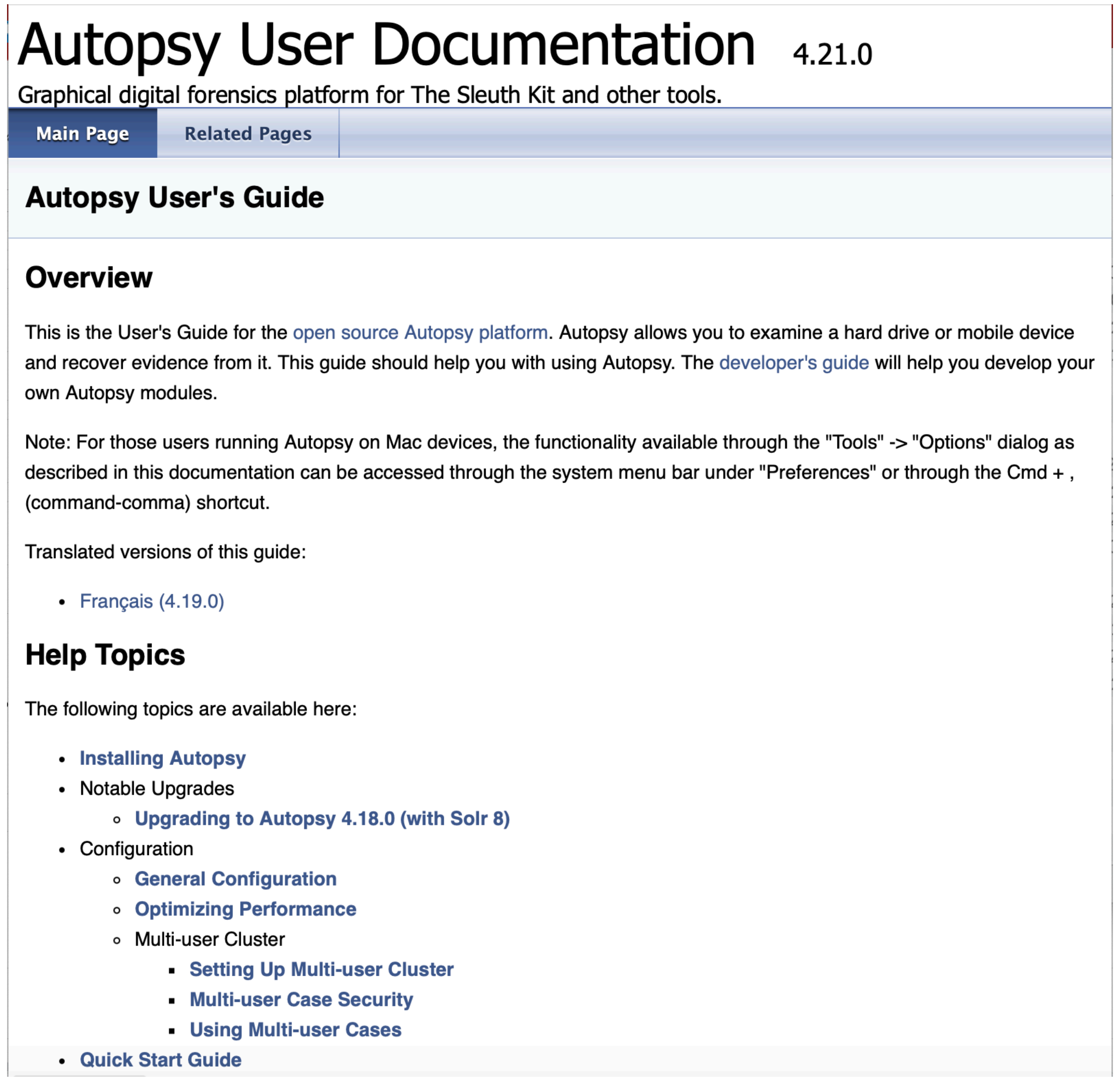
Abstract model of forensic analysis tools

A Foundation for Systematic Error Analysis

Methodology

- Deconstruction of digital forensic tool process - review of tool features
- Describing, grouping, and abstracting of these features
- Examples of tool error at each layer
- Illustration of error propagation
- Demonstration of how output at each layer allows error to be pinpointed

Review of Tool Features



Autopsy User Documentation 4.21.0

Graphical digital forensics platform for The Sleuth Kit and other tools.

Autopsy User's Guide

Overview

This is the User's Guide for the open source Autopsy platform. Autopsy allows you to examine a hard drive or mobile device and recover evidence from it. This guide should help you with using Autopsy. The developer's guide will help you develop your own Autopsy modules.

Note: For those users running Autopsy on Mac devices, the functionality available through the "Tools" -> "Options" dialog as described in this documentation can be accessed through the system menu bar under "Preferences" or through the Cmd + , (command-comma) shortcut.

Translated versions of this guide:

- Français (4.19.0)

Help Topics

The following topics are available here:

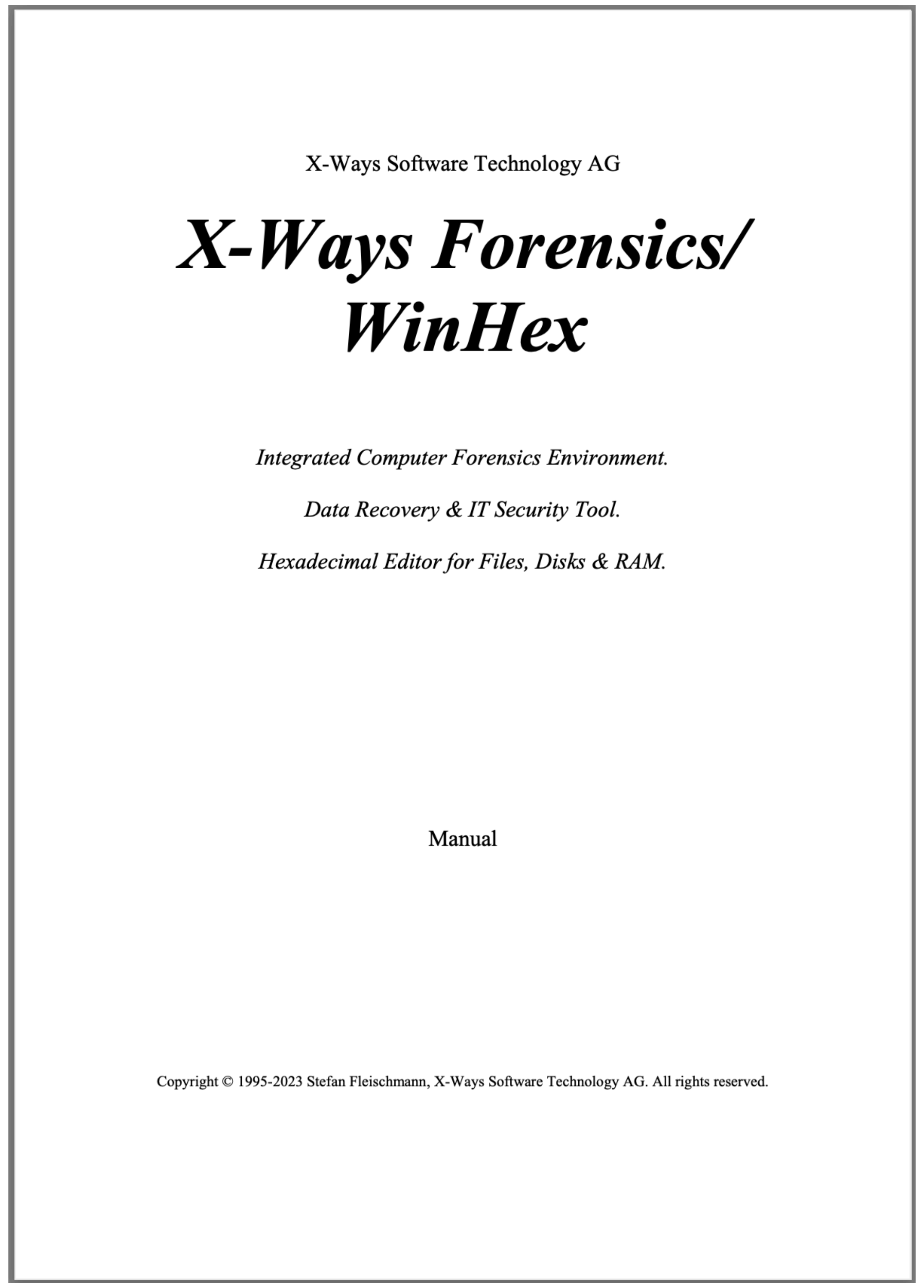
- Installing Autopsy
- Notable Upgrades
  - Upgrading to Autopsy 4.18.0 (with Solr 8)
- Configuration
  - General Configuration
  - Optimizing Performance
  - Multi-user Cluster
    - Setting Up Multi-user Cluster
    - Multi-user Case Security
    - Using Multi-user Cases
- Quick Start Guide

http://sleuthkit.org/autopsy/docs/user-docs/4.21.0/

MAGNET AXIOM™

USER GUIDE

https://www.magnetforensics.com/products/magnet-axiom/



X-Ways Software Technology AG

X-Ways Forensics/ WinHex

Integrated Computer Forensics Environment.

Data Recovery & IT Security Tool.

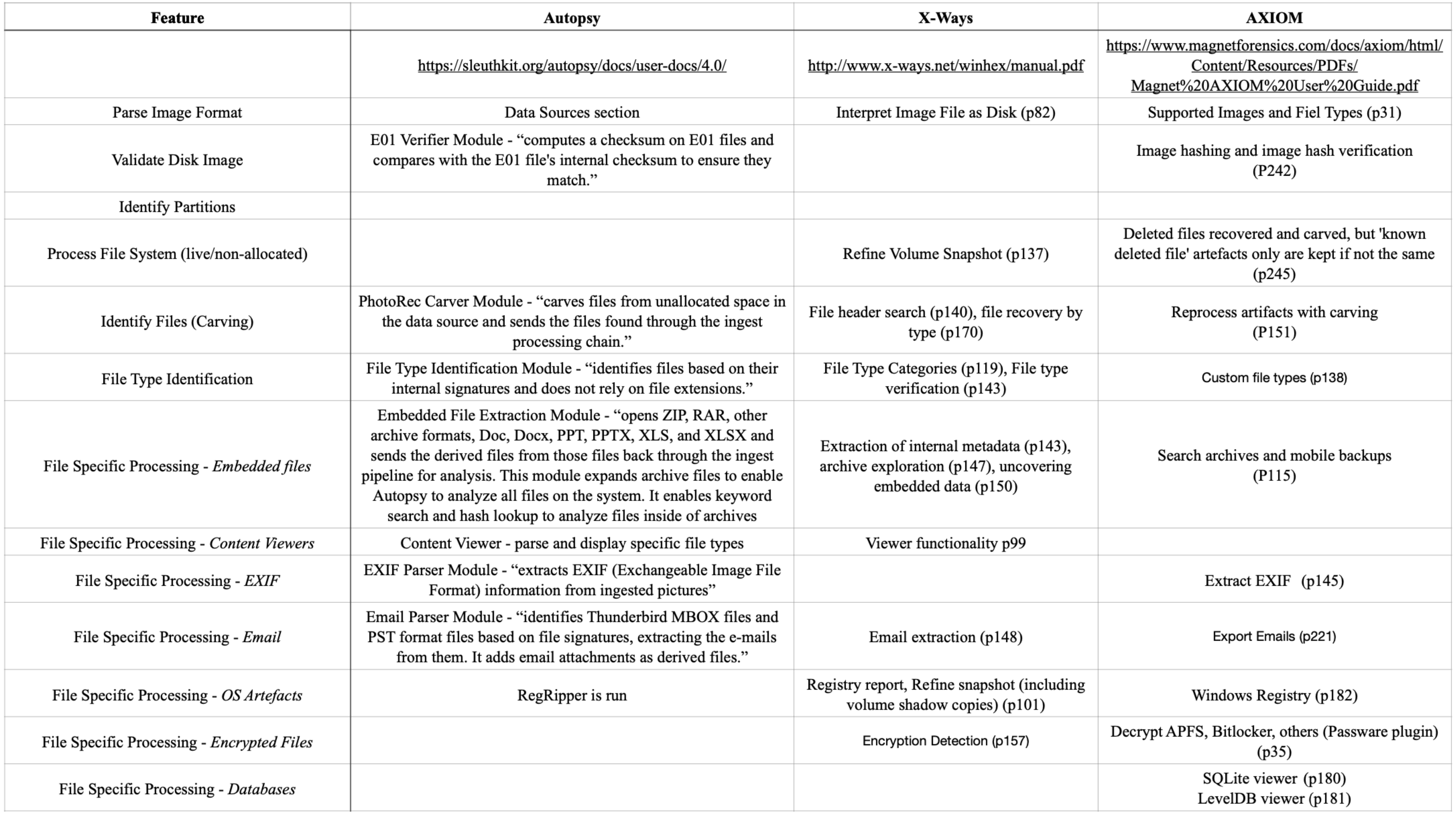
Hexadecimal Editor for Files, Disks & RAM.

Manual

Copyright © 1995-2023 Stefan Fleischmann, X-Ways Software Technology AG. All rights reserved.

http://www.x-ways.net/winhex/manual.pdf

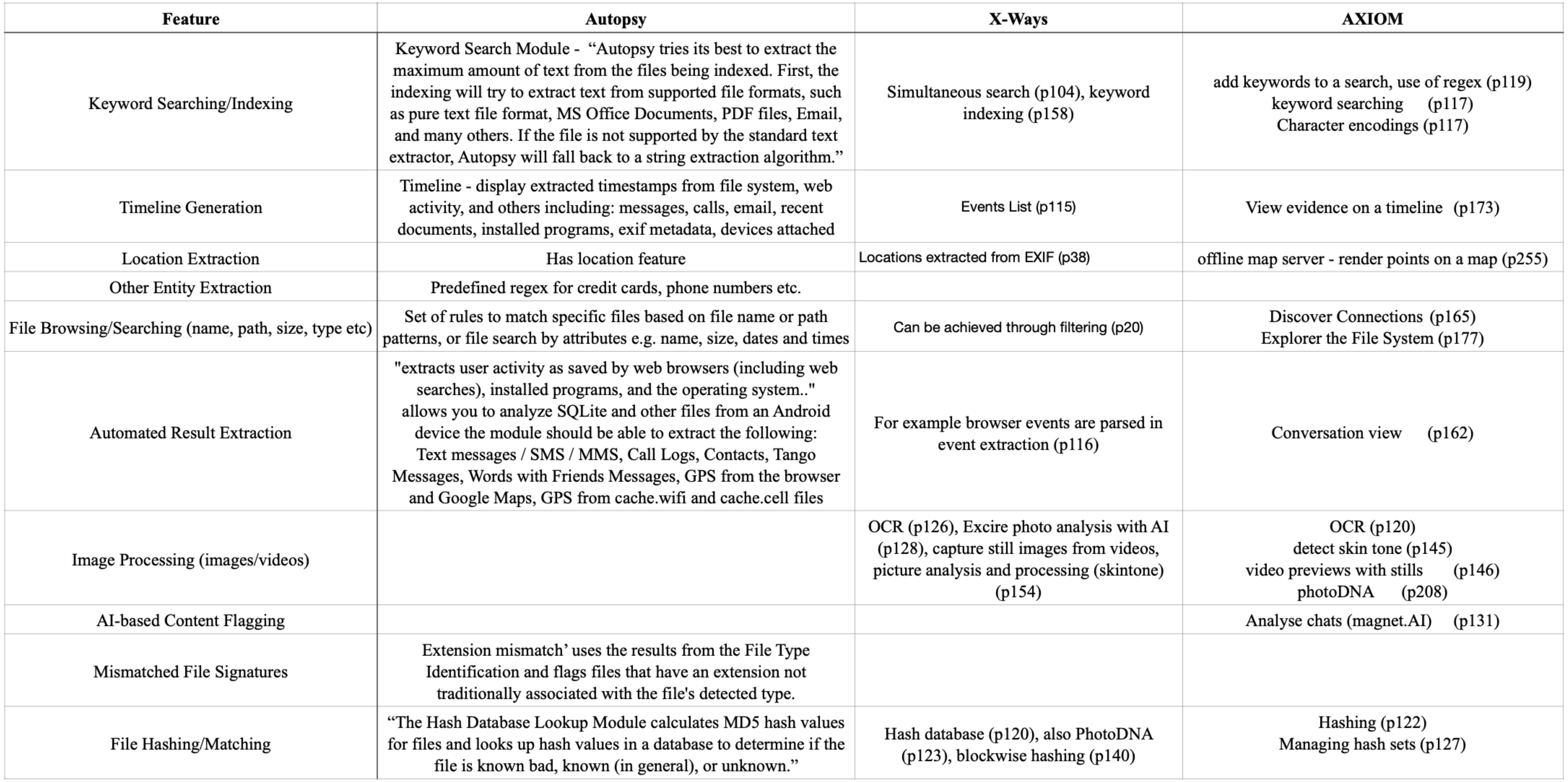
Review of Tool Features



Feature	Autopsy	X-Ways	AXIOM
	https://sleuthkit.org/autopsy/docs/user-docs/4.0/	http://www.x-ways.net/winhex/manual.pdf	https://www.magnetforensics.com/docs/axiom/html/Content/Resources/PDFs/Magnet%20AXIOM%20User%20Guide.pdf
Parse Image Format	Data Sources section	Interpret Image File as Disk (p82)	Supported Images and Fiel Types (p31)
Validate Disk Image	E01 Verifier Module - “computes a checksum on E01 files and compares with the E01 file's internal checksum to ensure they match.”		Image hashing and image hash verification (P242)
Identify Partitions			
Process File System (live/non-allocated)		Refine Volume Snapshot (p137)	Deleted files recovered and carved, but 'known deleted file' artefacts only are kept if not the same (p245)
Identify Files (Carving)	PhotoRec Carver Module - “carves files from unallocated space in the data source and sends the files found through the ingest processing chain.”	File header search (p140), file recovery by type (p170)	Reprocess artifacts with carving (P151)
File Type Identification	File Type Identification Module - “identifies files based on their internal signatures and does not rely on file extensions.”	File Type Categories (p119), File type verification (p143)	Custom file types (p138)
File Specific Processing - *Embedded files*	Embedded File Extraction Module - “opens ZIP, RAR, other archive formats, Doc, Docx, PPT, PPTX, XLS, and XLSX and sends the derived files from those files back through the ingest pipeline for analysis. This module expands archive files to enable Autopsy to analyze all files on the system. It enables keyword search and hash lookup to analyze files inside of archives	Extraction of internal metadata (p143), archive exploration (p147), uncovering embedded data (p150)	Search archives and mobile backups (P115)
File Specific Processing - *Content Viewers*	Content Viewer - parse and display specific file types	Viewer functionality p99	
File Specific Processing - *EXIF*	EXIF Parser Module - “extracts EXIF (Exchangeable Image File Format) information from ingested pictures”		Extract EXIF (p145)
File Specific Processing - *Email*	Email Parser Module - “identifies Thunderbird MBOX files and PST format files based on file signatures, extracting the e-mails from them. It adds email attachments as derived files.”	Email extraction (p148)	Export Emails (p221)
File Specific Processing - *OS Artefacts*	RegRipper is run	Registry report, Refine snapshot (including volume shadow copies) (p101)	Windows Registry (p182)
File Specific Processing - *Encrypted Files*		Encryption Detection (p157)	Decrypt APFS, Bitlocker, others (Passware plugin) (p35)
File Specific Processing - *Databases*			SQLite viewer (p180) LevelDB viewer (p181)

* Not intended as full tool feature comparison, just generation of a feature list. Blanks do not mean the tool does not have that feature.

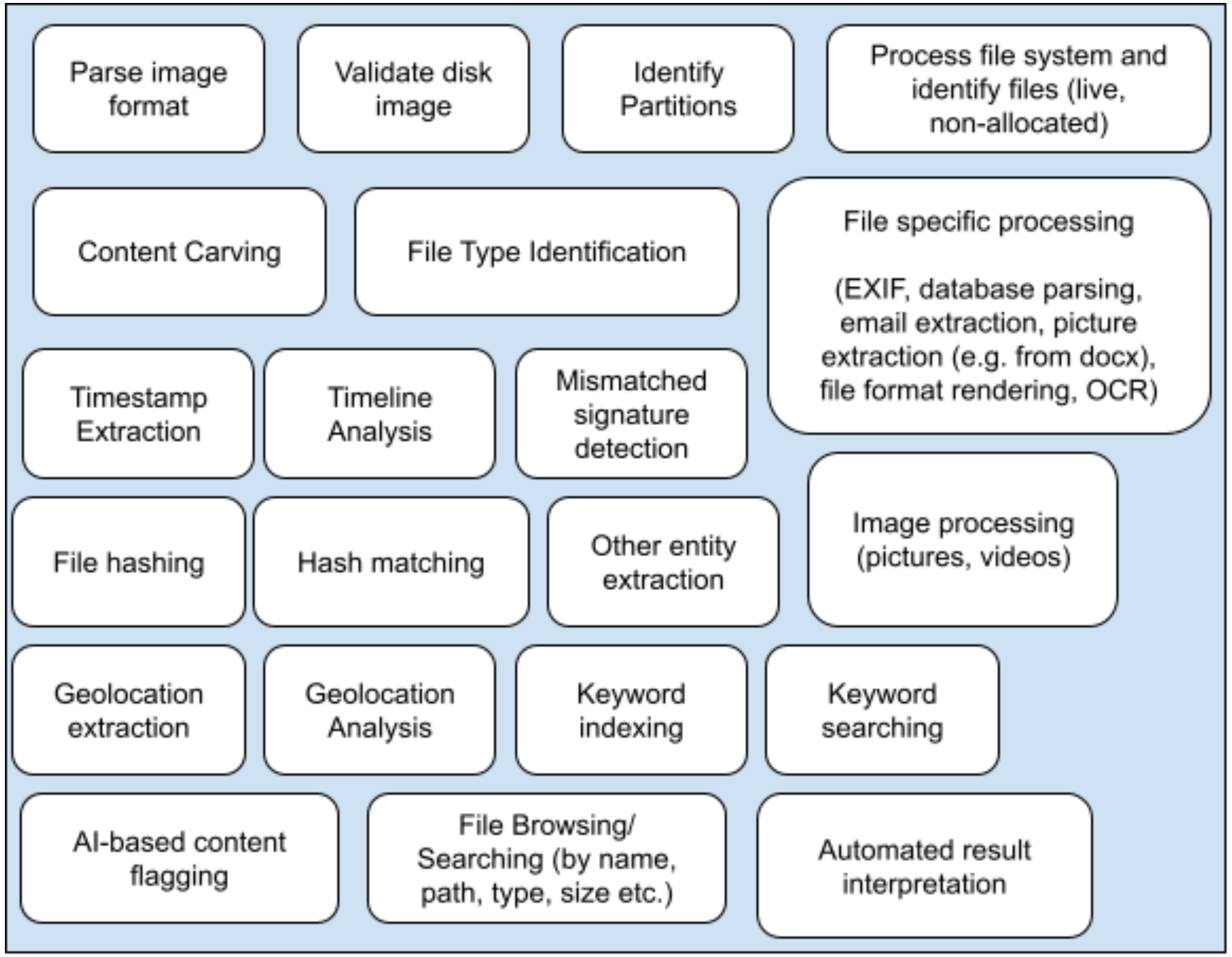
Review of Tool Features



Feature	Autopsy	X-Ways	AXIOM
Keyword Searching/Indexing	Keyword Search Module - "Autopsy tries its best to extract the maximum amount of text from the files being indexed. First, the indexing will try to extract text from supported file formats, such as pure text file format, MS Office Documents, PDF files, Email, and many others. If the file is not supported by the standard text extractor, Autopsy will fall back to a string extraction algorithm."	Simultaneous search (p104), keyword indexing (p158)	add keywords to a search, use of regex (p119) keyword searching (p117) Character encodings (p117)
Timeline Generation	Timeline - display extracted timestamps from file system, web activity, and others including: messages, calls, email, recent documents, installed programs, exif metadata, devices attached	Events List (p115)	View evidence on a timeline (p173)
Location Extraction	Has location feature	Locations extracted from EXIF (p38)	offline map server - render points on a map (p255)
Other Entity Extraction	Predefined regex for credit cards, phone numbers etc.		
File Browsing/Searching (name, path, size, type etc)	Set of rules to match specific files based on file name or path patterns, or file search by attributes e.g. name, size, dates and times	Can be achieved through filtering (p20)	Discover Connections (p165) Explorer the File System (p177)
Automated Result Extraction	"extracts user activity as saved by web browsers (including web searches), installed programs, and the operating system.." allows you to analyze SQLite and other files from an Android device the module should be able to extract the following: Text messages / SMS / MMS, Call Logs, Contacts, Tango Messages, Words with Friends Messages, GPS from the browser and Google Maps, GPS from cache.wifi and cache.cell files	For example browser events are parsed in event extraction (p116)	Conversation view (p162)
Image Processing (images/videos)		OCR (p126), Excire photo analysis with AI (p128), capture still images from videos, picture analysis and processing (skintone) (p154)	OCR (p120) detect skin tone (p145) video previews with stills (p146) photoDNA (p208)
AI-based Content Flagging			Analyse chats (magnet.AI) (p131)
Mismatched File Signatures	Extension mismatch' uses the results from the File Type Identification and flags files that have an extension not traditionally associated with the file's detected type.		
File Hashing/Matching	"The Hash Database Lookup Module calculates MD5 hash values for files and looks up hash values in a database to determine if the file is known bad, known (in general), or unknown."	Hash database (p120), also PhotoDNA (p123), blockwise hashing (p140)	Hashing (p122) Managing hash sets (p127)

* Not intended as full tool feature comparison, just generation of a feature list. Blanks do not mean the tool does not have that feature.

Review of Tool Features



Parse image format

Validate disk image

Identify Partitions

Process file system and identify files (live, non-allocated)

Content Carving

File Type Identification

File specific processing

(EXIF, database parsing, email extraction, picture extraction (e.g. from docx), file format rendering, OCR)

Timestamp Extraction

Timeline Analysis

Mismatched signature detection

File hashing

Hash matching

Other entity extraction

Image processing (pictures, videos)

Geolocation extraction

Geolocation Analysis

Keyword indexing

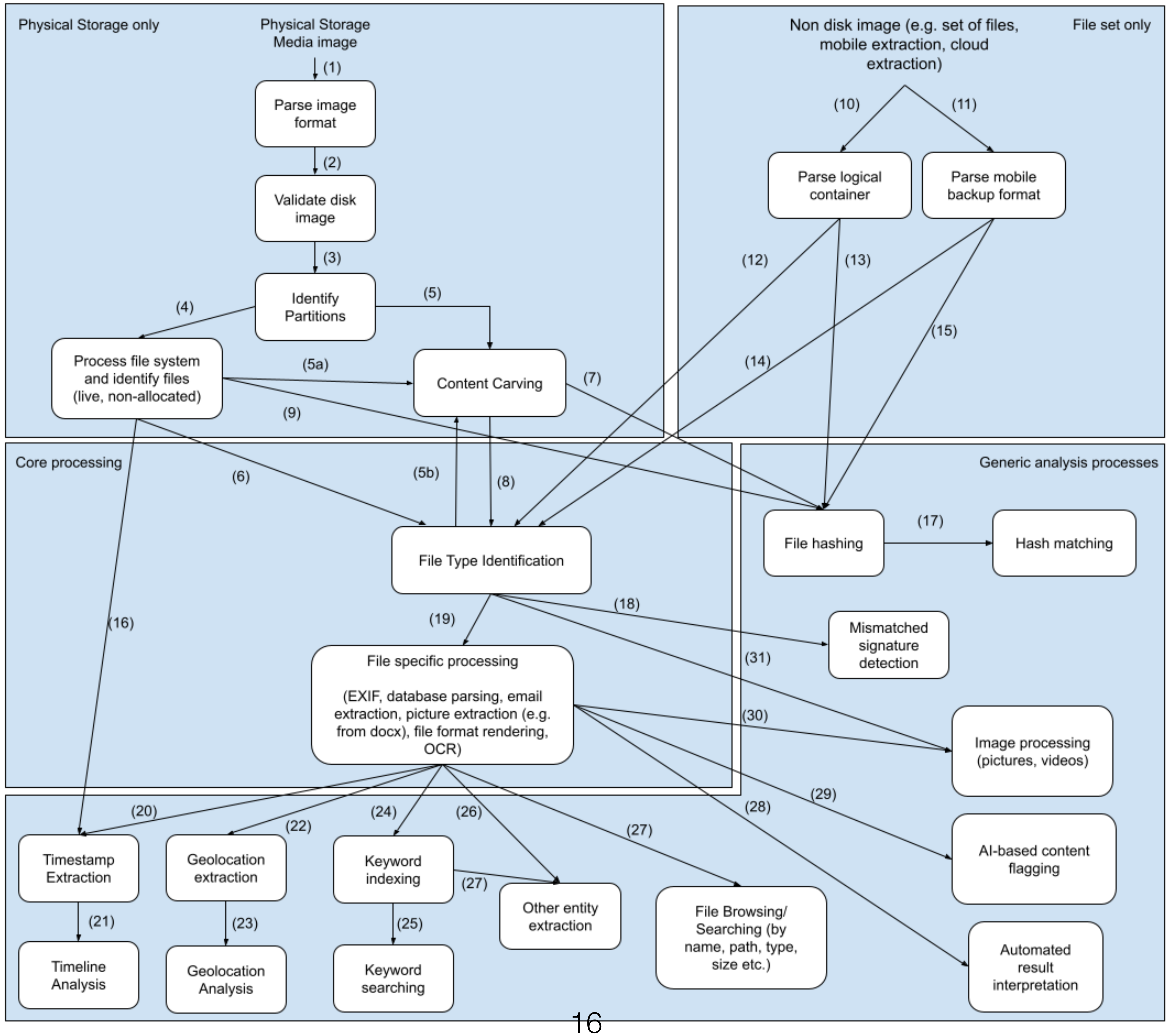
Keyword searching

AI-based content flagging

File Browsing/ Searching (by name, path, type, size etc.)

Automated result interpretation

Diagram, dependencies, and errors



Physical Storage only

Physical Storage Media image

(1)

Parse image format

(2)

Validate disk image

(3)

Identify Partitions

(4)

(5)

Process file system and identify files (live, non-allocated)

(5a)

Content Carving

(7)

(9)

Non disk image (e.g. set of files, mobile extraction, cloud extraction)

File set only

(10)

(11)

Parse logical container

Parse mobile backup format

(12)

(13)

(14)

(15)

Core processing

(6)

(5b)

(8)

File Type Identification

(16)

(19)

(18)

File specific processing

(EXIF, database parsing, email extraction, picture extraction (e.g. from docx), file format rendering, OCR)

Generic analysis processes

File hashing

(17)

Hash matching

Mismatched signature detection

(31)

(30)

Image processing (pictures, videos)

(29)

(28)

AI-based content flagging

Automated result interpretation

(20)

(22)

(24)

(26)

(27)

Timestamp Extraction

Geolocation extraction

Keyword indexing

(27)

Other entity extraction

File Browsing/ Searching (by name, path, type, size etc.)

(21)

(23)

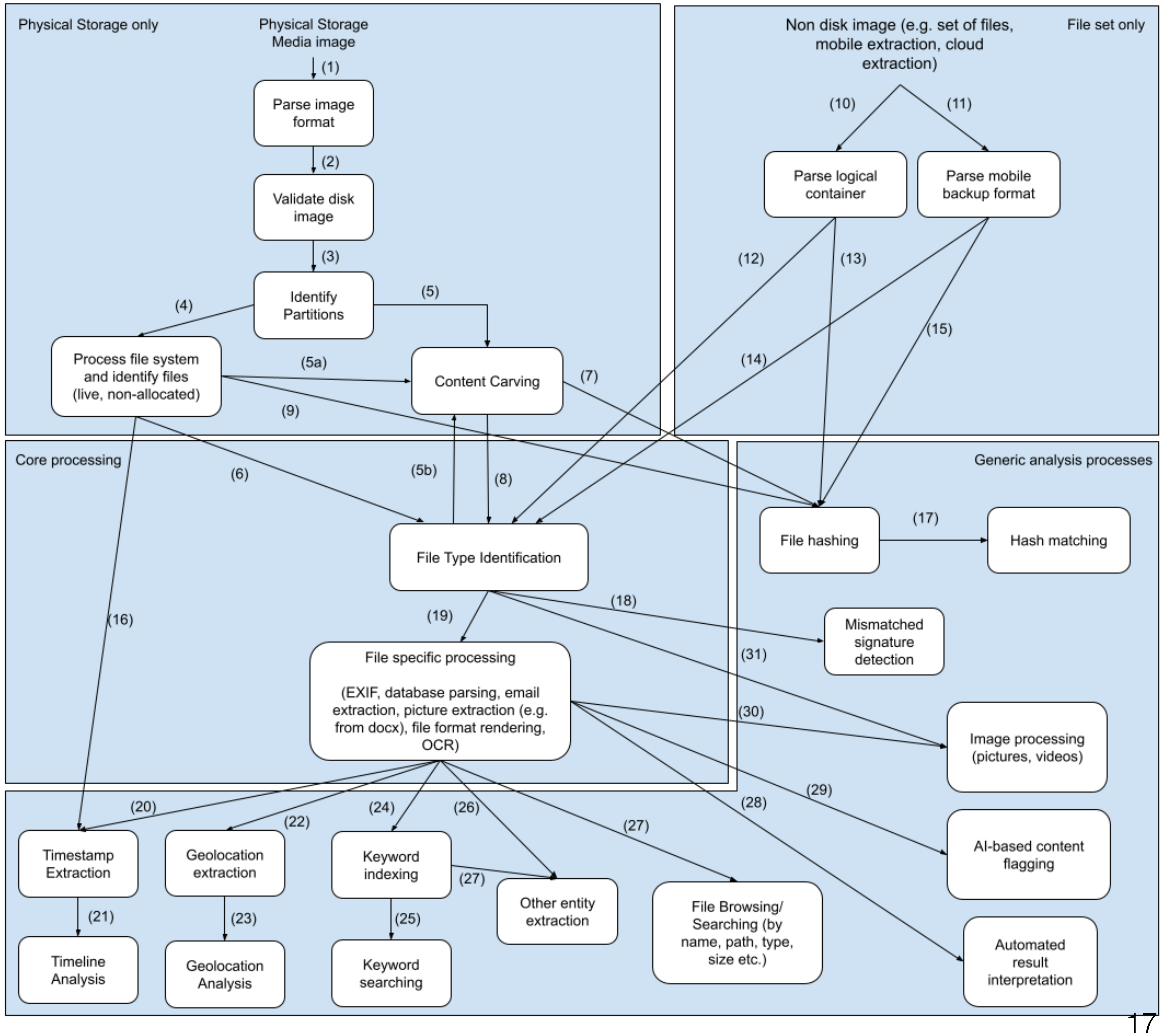
(25)

Timeline Analysis

Geolocation Analysis

Keyword searching

Diagram, dependencies, and errors

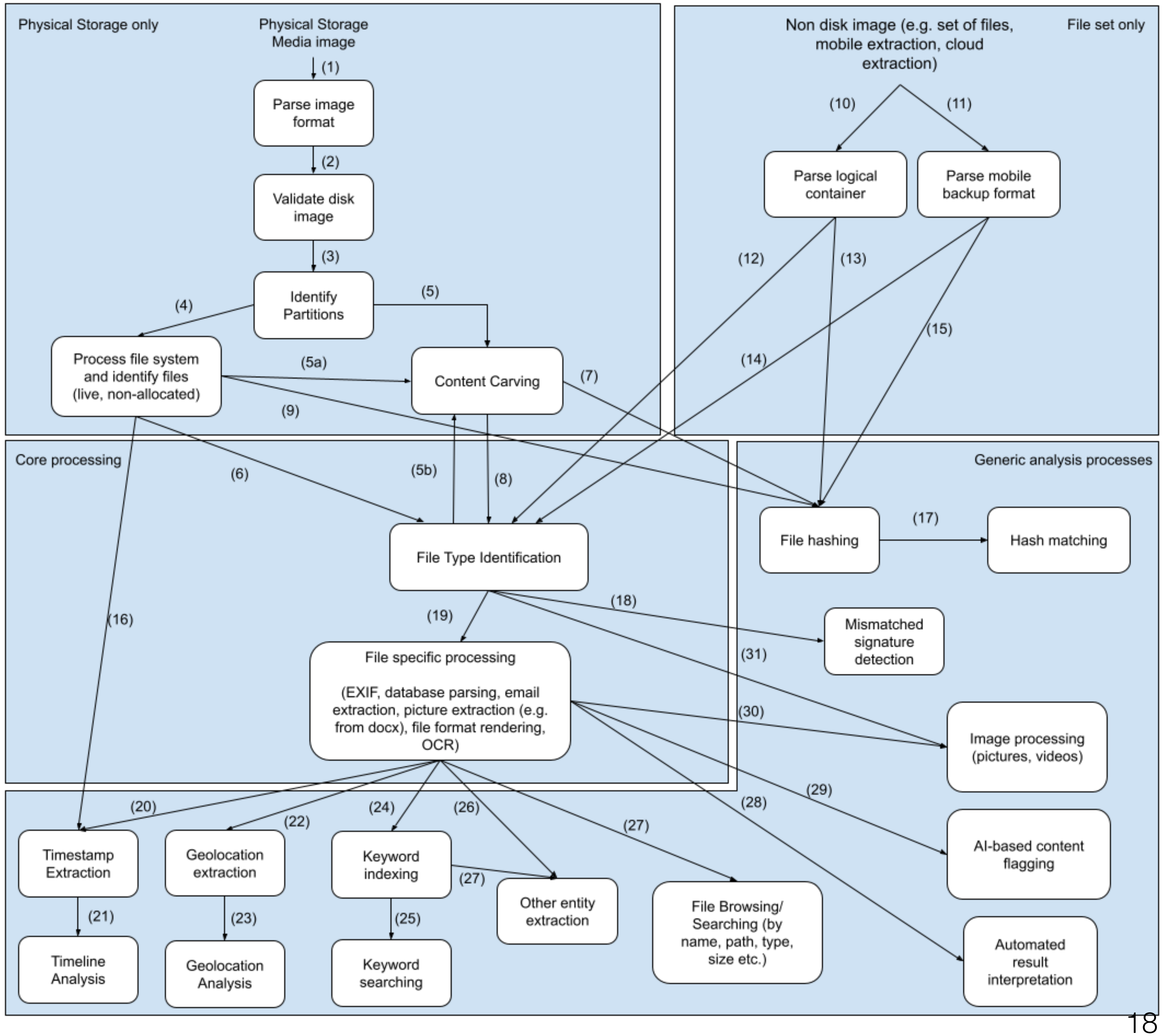

Physical Storage only
Physical Storage Media image
(1)
Parse image format
(2)
Validate disk image
(3)
Identify Partitions
(4)
(5)
Process file system and identify files (live, non-allocated)
(5a)
Content Carving
(7)
(9)
Non disk image (e.g. set of files, mobile extraction, cloud extraction)
File set only
(10)
(11)
Parse logical container
Parse mobile backup format
(12)
(13)
(14)
(15)
Core processing
(6)
(5b)
(8)
(16)
File Type Identification
Generic analysis processes
File hashing
(17)
Hash matching
(18)
(19)
Mismatched signature detection
(31)
File specific processing

(EXIF, database parsing, email extraction, picture extraction (e.g. from docx), file format rendering, OCR)
(30)
Image processing (pictures, videos)
(29)
(28)
AI-based content flagging
(20)
(22)
(24)
(26)
(27)
Timestamp Extraction
Geolocation extraction
Keyword indexing
(27)
Other entity extraction
File Browsing/ Searching (by name, path, type, size etc.)
Automated result interpretation
(21)
(23)
(25)
Timeline Analysis
Geolocation Analysis
Keyword searching

ASTM (2018) ‘E3016-18 - Standard Guide for Establishing Confidence in Digital and Multimedia Evidence Forensic Results by Error Mitigation Analysis'

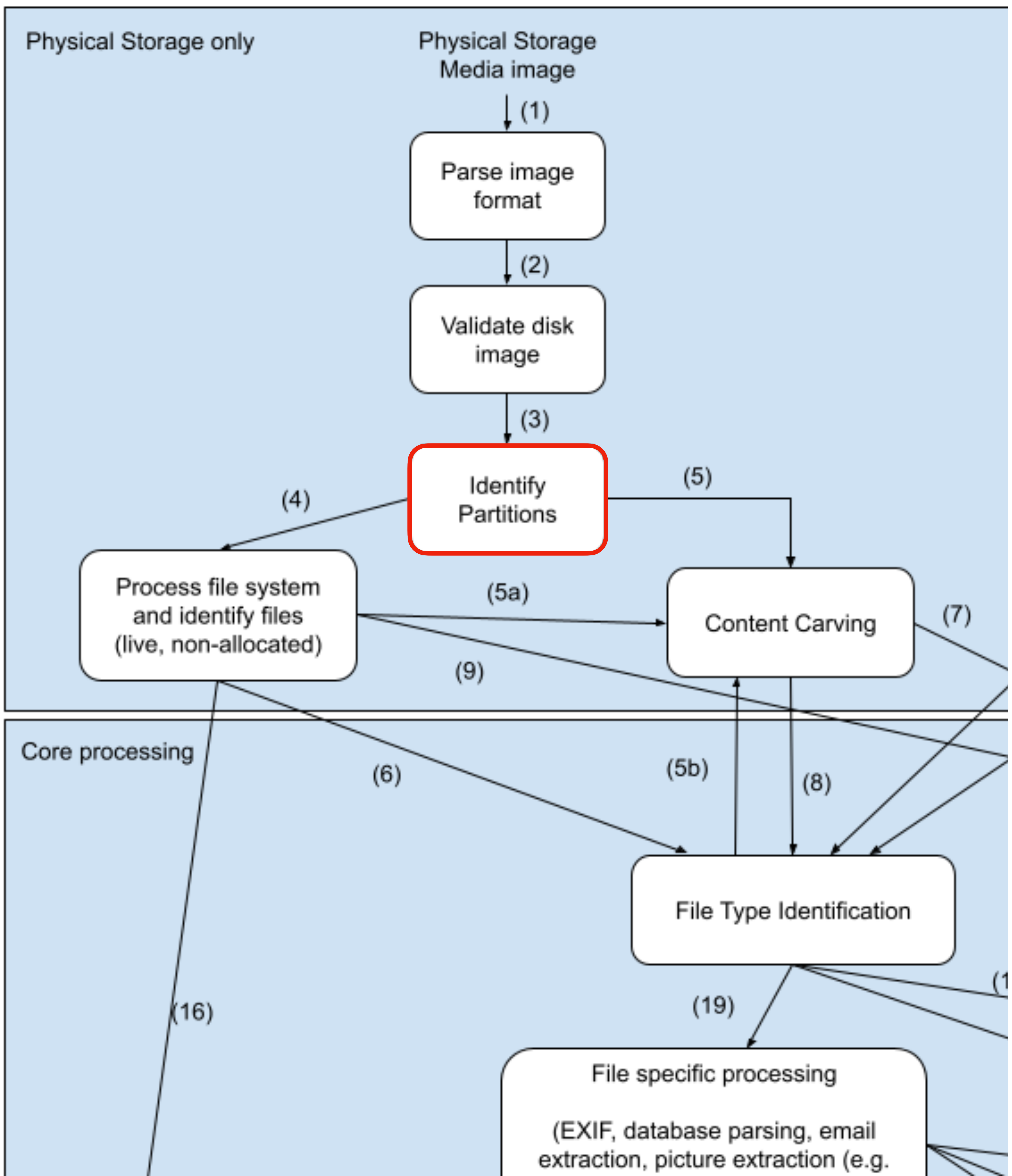
- Types of error:
    - **Incompleteness** - relevant information has not been acquired or found (INCOMP)
    - **Inaccuracy**
        - •***Existence*** - do all artefacts reported as present actually exist (INAC-EX)
        - •**Alteration** - does a tool alter data in a way that changes its meaning? (INAC-ALT)
        - •***Association*** - for every set of items identified by a given tool, is each item truly part of that set (INAC-AS)
        - •***Corruption*** - does the forensic tool detect and compensate for missing and corrupted data (INAC-COR)
    - **Misinterpretation** - The results have been incorrectly understood (MISINT)
- Applying this analysis has documented 77 possible error sources, of 6 types, over the 23 stages of processing. This is far from exhaustive.

Diagram, dependencies, and errors


Physical Storage only
Physical Storage Media image
(1)
Parse image format
(2)
Validate disk image
(3)
Identify Partitions
(4)
(5)
Process file system and identify files (live, non-allocated)
(5a)
Content Carving
(7)
(9)
Non disk image (e.g. set of files, mobile extraction, cloud extraction)
File set only
(10)
(11)
Parse logical container
Parse mobile backup format
(12)
(13)
(14)
(15)
Core processing
(6)
(5b)
(8)
File Type Identification
(16)
(19)
(18)
File specific processing
(EXIF, database parsing, email extraction, picture extraction (e.g. from docx), file format rendering, OCR)
Generic analysis processes
File hashing
(17)
Hash matching
Mismatched signature detection
(31)
(30)
Image processing (pictures, videos)
(29)
(28)
AI-based content flagging
Automated result interpretation
(20)
(22)
(24)
(26)
(27)
Timestamp Extraction
Geolocation extraction
Keyword indexing
(27)
Other entity extraction
File Browsing/ Searching (by name, path, type, size etc.)
(21)
(23)
(25)
Timeline Analysis
Geolocation Analysis
Keyword searching

ASTM (2018) ‘E3016-18 - Standard Guide for Establishing Confidence in Digital and Multimedia Evidence Forensic Results by Error Mitigation Analysis'

- Types of error:
  - **Incompleteness** - relevant information has not been acquired or found (INCOMP)
  - **Inaccuracy**
    - •***Existence*** - do all artefacts reported as present actually exist (INAC-EX)
    - •**Alteration** - does a tool alter data in a way that changes its meaning? (INAC-ALT)
    - •***Association*** - for every set of items identified by a given tool, is each item truly part of that set (INAC-AS)
    - •***Corruption*** - does the forensic tool detect and compensate for missing and corrupted data (INAC-COR)
  - **Misinterpretation** - The results are displayed in a manner that encourages, or does not prevent misinterpretation (MISINT)
- Applying this analysis has documented 77 possible error sources, of 6 types, over the 23 stages of processing. This is far from exhaustive.



For example: Identify Partitions

Description

- Parsing the partition scheme in use, GPT or MBR based, including all EPTs
- Examining unpartitioned space for deleted but recoverable partitions

Dependencies

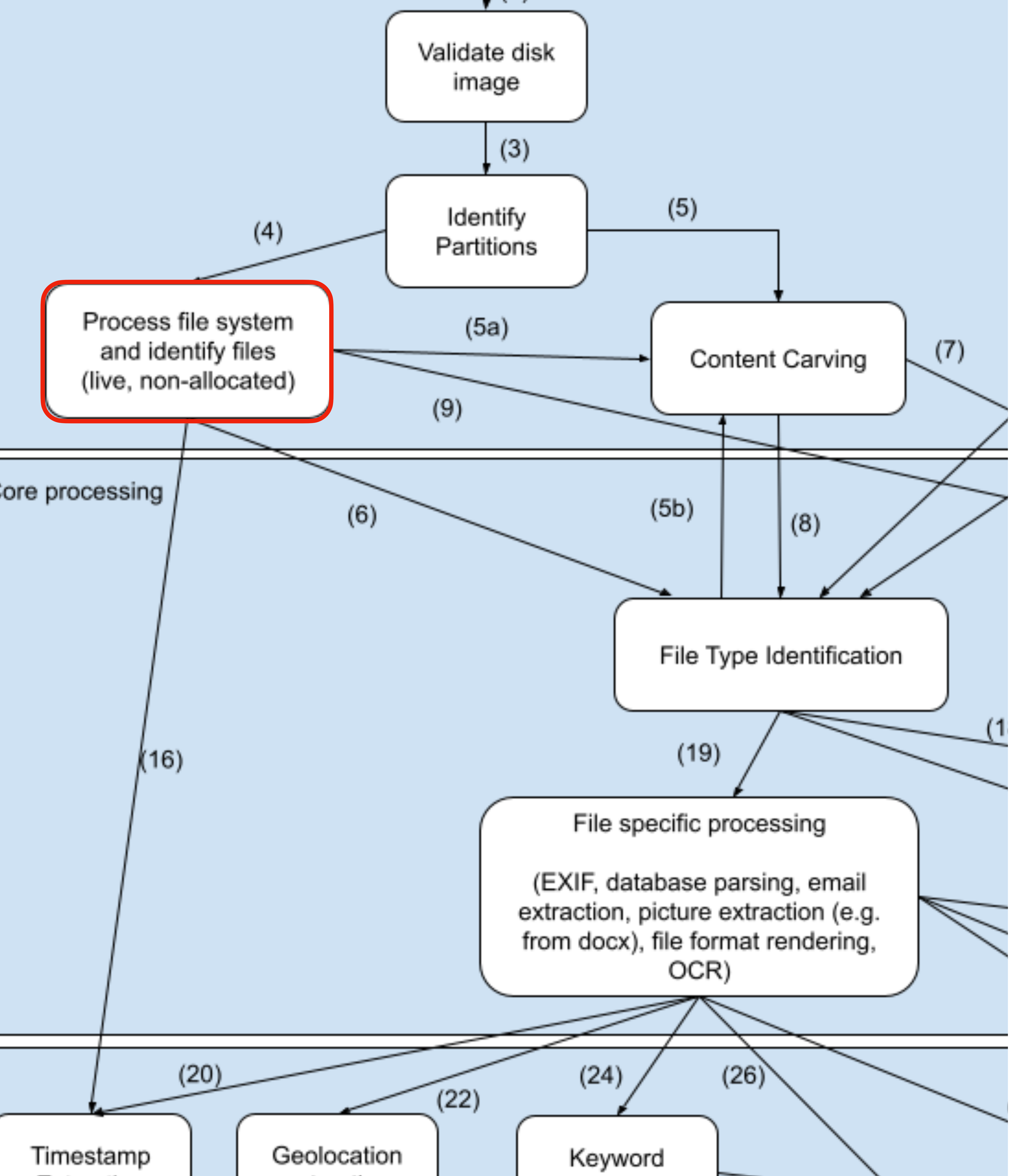
d3, if data from disk image is not validated then sectors may be missing or corrupt to prevent partition reconstruction from being performed correctly

Potential error introduced at this stage

- Incorrect parsing of partition table(s) - could result in incomplete partition list (INCOMP) or an incorrect one (INAC-EX)

- Incorrect assumptions about sector size (512 rather than 4096) or following pointers incorrectly could all miss partitions (INCOMP)

- Missing deleted but recoverable partitions - failing to search unpartitioned space for VBRs could miss entire deleted partitions (INCOMP)



For example: Process File System

Description

- Partition inspected, valid file system identified, file system parsed according to documented or reverse engineered specification.
- Usually also includes attempt to recover non-allocate files that still retain references within the file system

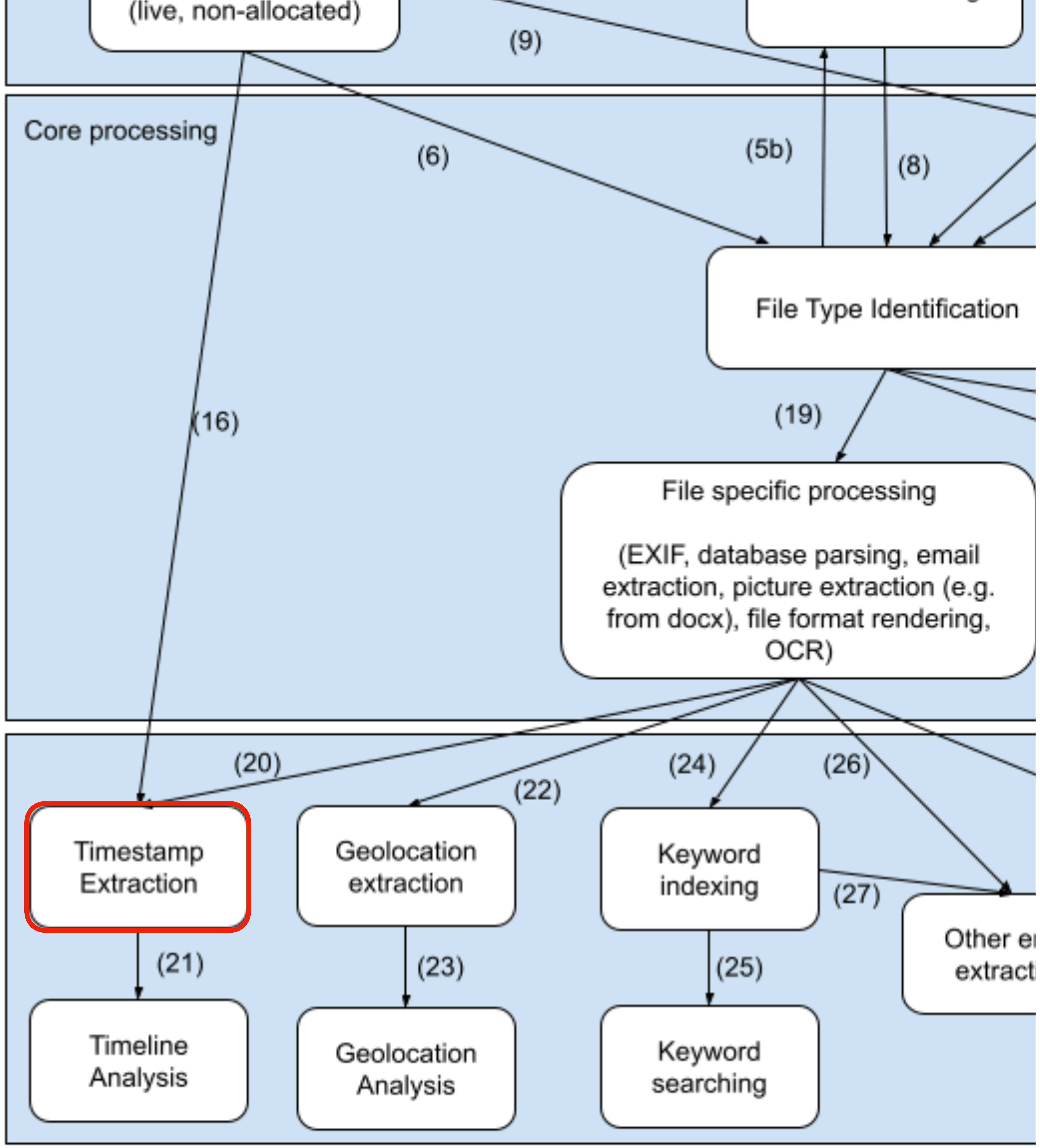
Dependencies

d4, if a partition is not identified, file system processing does cannot occur within that partition.

Potential error introduced at this stage

INCOMP	INAC-EX	INAC-AS	INAC-ALT	INAC-COR	MISINT
Failure to identify known file system type e.g. trigger missed for file system or lack of support for file system known within digital forensic science	Additional files	Attributing non-allocated file content to a different file	Inaccurate live file metadata extracted		Failure to communicate uncertainty in file recovery results e.g. where content may be partially overwritten
Missing live files: mistake in specification		Attributing metadata to the wrong file	Inaccurate live file content extracted		
Missing live files: mistake in implementation			Inaccurate non-allocated file metadata extracted		
Missing non-allocated but recoverable files			Inaccurate non-allocated file content extracted		

Many are described in more detail in Casey et al (2019)



For example: Timeline Generation

Description

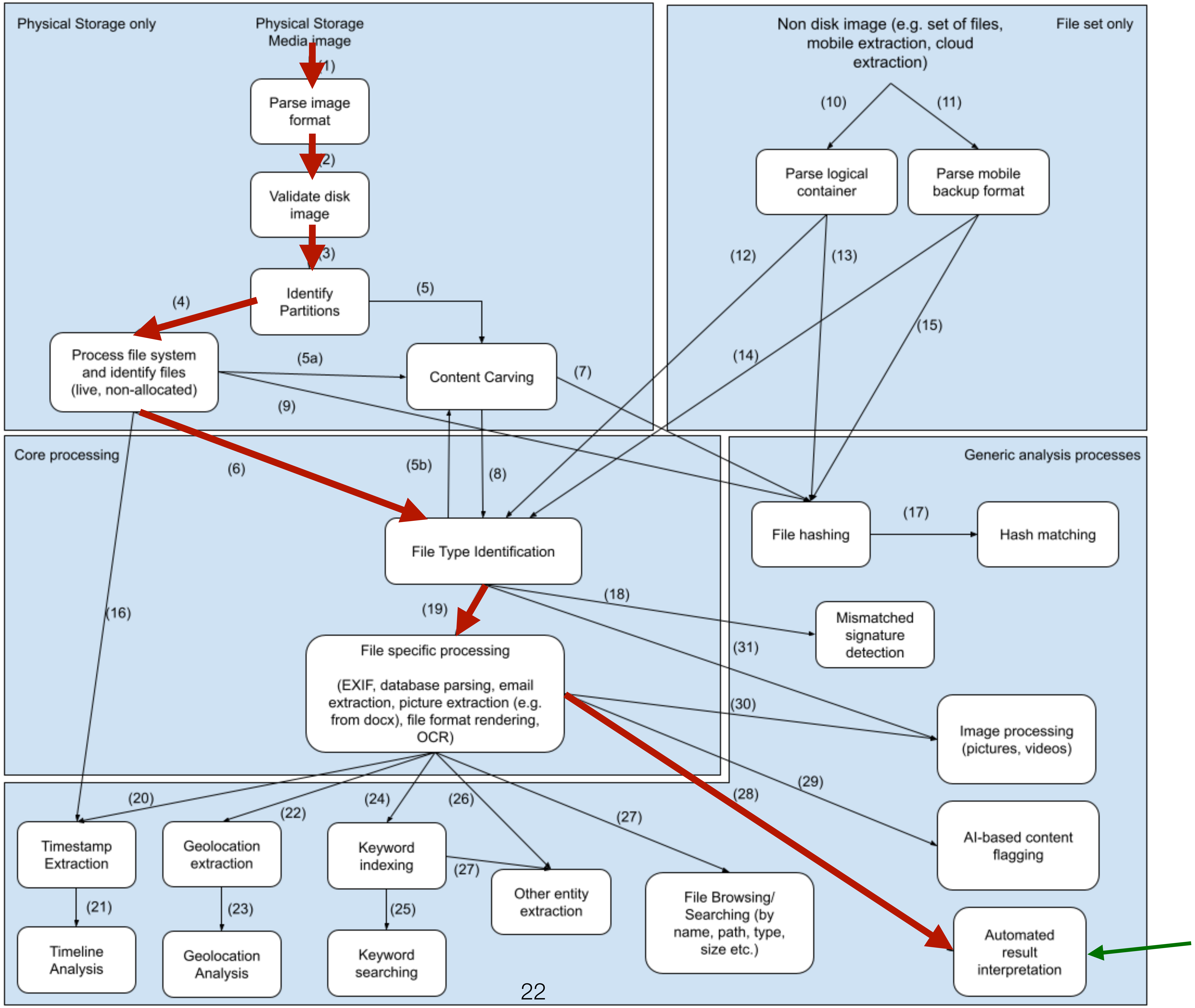
- Extracting time stamps from file system, and applying file specific processing to extract timestamps from within files eg Windows Registry

Dependencies

d16 for file system timestamp extraction, and *d20* for if file specific internal timestamps are to be extracted.

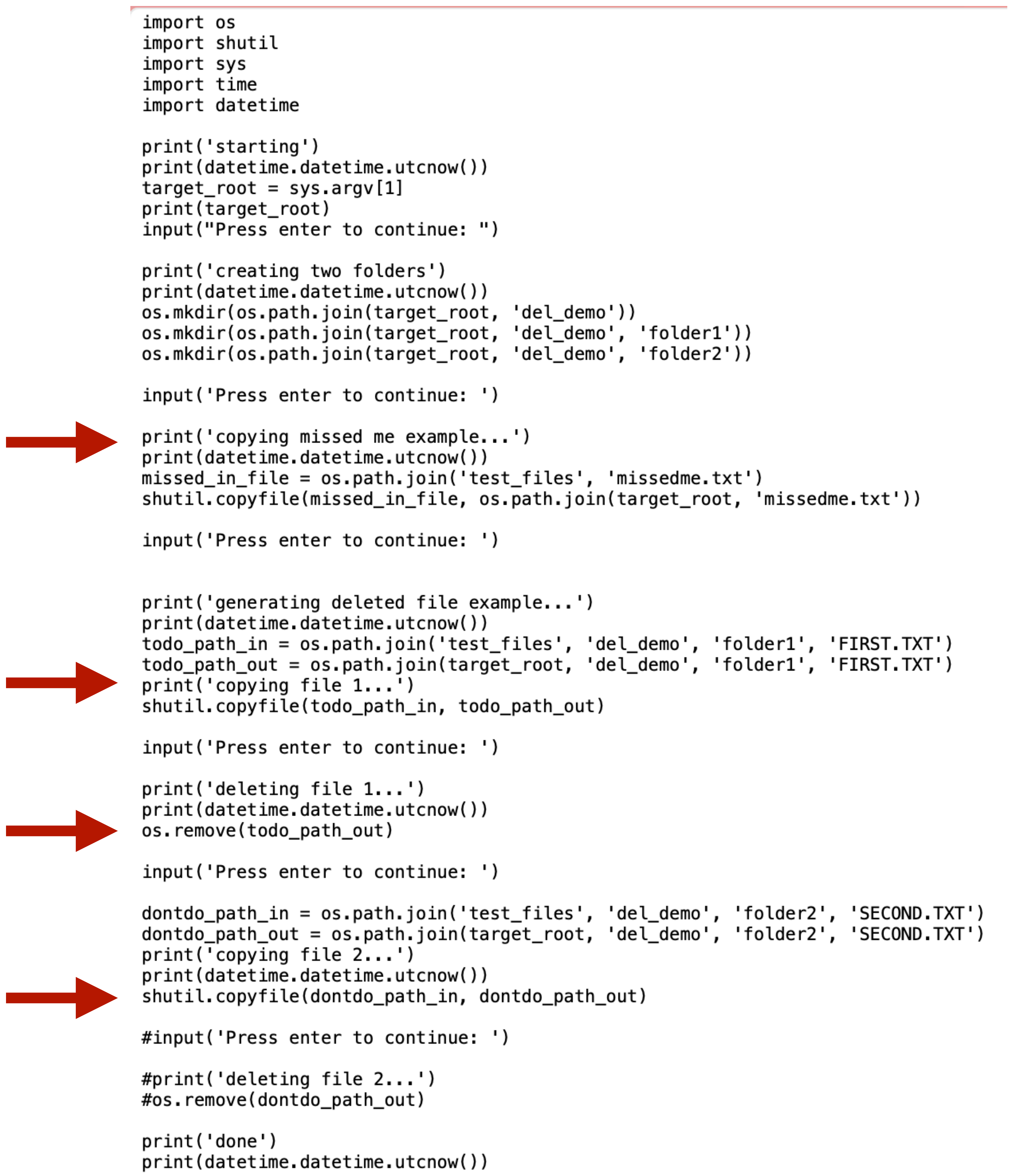
Potential error introduced at this stage

INCOMP	INAC-EX	INAC-AS	INAC-ALT	INAC-COR	MISINT
Failure to extract timestamps from a file type that contains extractable low-level events.	Creation of a fabricated timestamp entry	Incorrect interpretation applied to a timestamp e.g. message received rather than read	Incorrectly processing a file such that incorrect timestamps are added to a timeline	Failure to apply known clock offset	Overly simplified timestamp without clear communication e.g. timestamp resolution
			Incorrect timestamp normalisation (either interval, epoch, or text string parsing)	Failure to detect inaccurate clock used for timestamp generation	Failure to report the possibility of inaccurate clock
			Incorrectly applying a timezone		


Physical Storage only
Physical Storage Media image
(1)
Parse image format
(2)
Validate disk image
(3)
Identify Partitions
(4)
(5)
Process file system and identify files (live, non-allocated)
(5a)
Content Carving
(7)
(9)
Non disk image (e.g. set of files, mobile extraction, cloud extraction)
File set only
(10)
(11)
Parse logical container
Parse mobile backup format
(12)
(13)
(14)
(15)
Core processing
(6)
(5b)
(8)
File Type Identification
Generic analysis processes
File hashing
(17)
Hash matching
(16)
(18)
(19)
Mismatched signature detection
(31)
File specific processing
(EXIF, database parsing, email extraction, picture extraction (e.g. from docx), file format rendering, OCR)
(30)
Image processing (pictures, videos)
(29)
(28)
AI-based content flagging
(20)
(22)
(24)
(26)
(27)
Timestamp Extraction
Geolocation extraction
Keyword indexing
(27)
Other entity extraction
File Browsing/ Searching (by name, path, type, size etc.)
Automated result interpretation
(21)
(23)
(25)
Timeline Analysis
Geolocation Analysis
Keyword searching

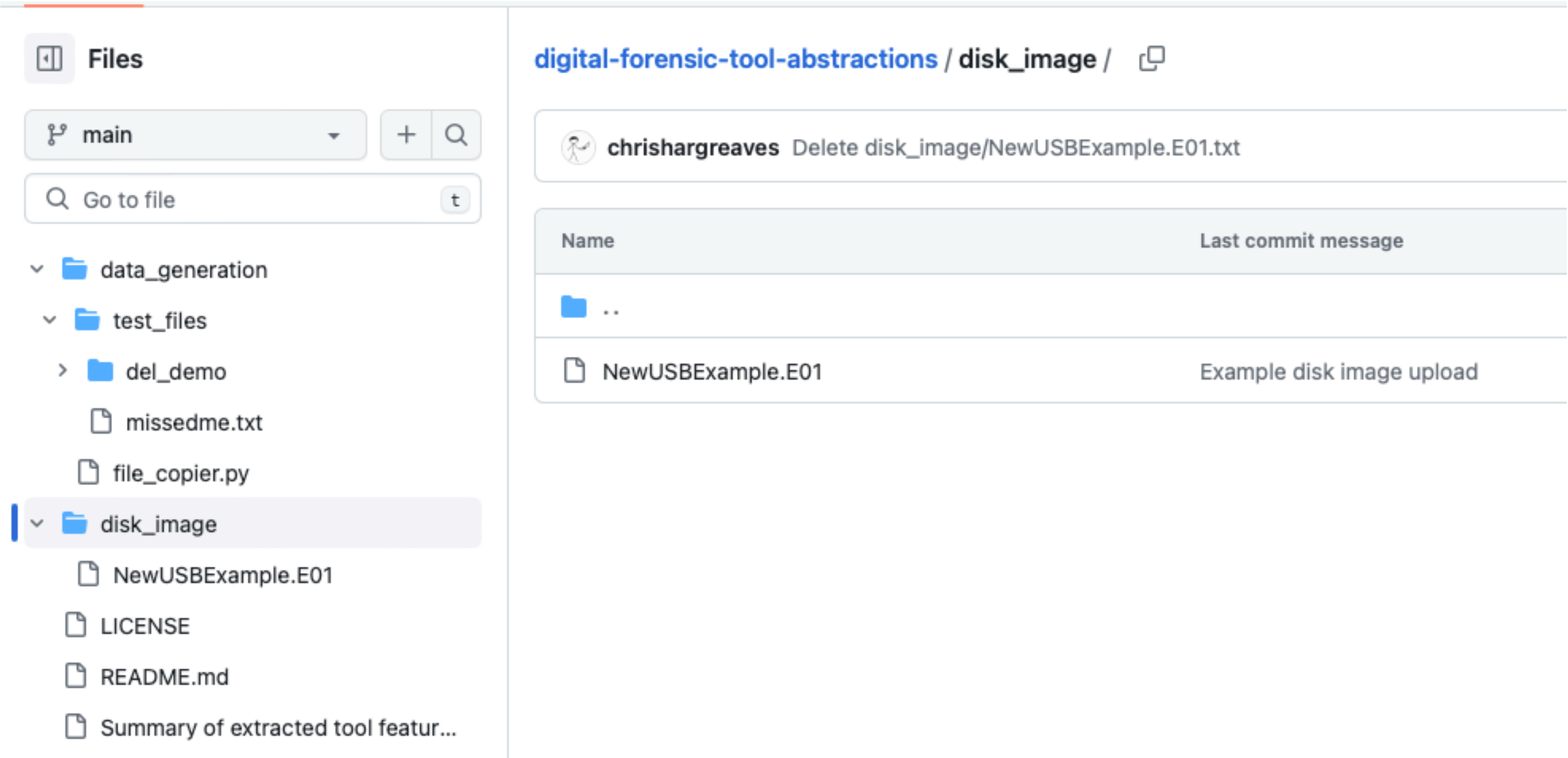
e.g. display messages

Illustration of Error Propagation



```
import os
import shutil
import sys
import time
import datetime

print('starting')
print(datetime.datetime.utcnow())
target_root = sys.argv[1]
print(target_root)
input("Press enter to continue: ")

print('creating two folders')
print(datetime.datetime.utcnow())
os.mkdir(os.path.join(target_root, 'del_demo'))
os.mkdir(os.path.join(target_root, 'del_demo', 'folder1'))
os.mkdir(os.path.join(target_root, 'del_demo', 'folder2'))

input('Press enter to continue: ')

print('copying missed me example...')
print(datetime.datetime.utcnow())
missed_in_file = os.path.join('test_files', 'missedme.txt')
shutil.copyfile(missed_in_file, os.path.join(target_root, 'missedme.txt'))

input('Press enter to continue: ')


print('generating deleted file example...')
print(datetime.datetime.utcnow())
todo_path_in = os.path.join('test_files', 'del_demo', 'folder1', 'FIRST.TXT')
todo_path_out = os.path.join(target_root, 'del_demo', 'folder1', 'FIRST.TXT')
print('copying file 1...')
shutil.copyfile(todo_path_in, todo_path_out)

input('Press enter to continue: ')

print('deleting file 1...')
print(datetime.datetime.utcnow())
os.remove(todo_path_out)

input('Press enter to continue: ')

dontdo_path_in = os.path.join('test_files', 'del_demo', 'folder2', 'SECOND.TXT')
dontdo_path_out = os.path.join(target_root, 'del_demo', 'folder2', 'SECOND.TXT')
print('copying file 2...')
print(datetime.datetime.utcnow())
shutil.copyfile(dontdo_path_in, dontdo_path_out)

#input('Press enter to continue: ')

#print('deleting file 2...')
#os.remove(dontdo_path_out)

print('done')
print(datetime.datetime.utcnow())
```

file_copier.py



chrishargreaves Delete disk_image/NewUSBExample.E01.txt

Name	Last commit message
..	
NewUSBExample.E01	Example disk image upload

https://github.com/chrishargreaves/digital-forensic-tool-abstractions

Identification of existence of, but not the specifics of the problem

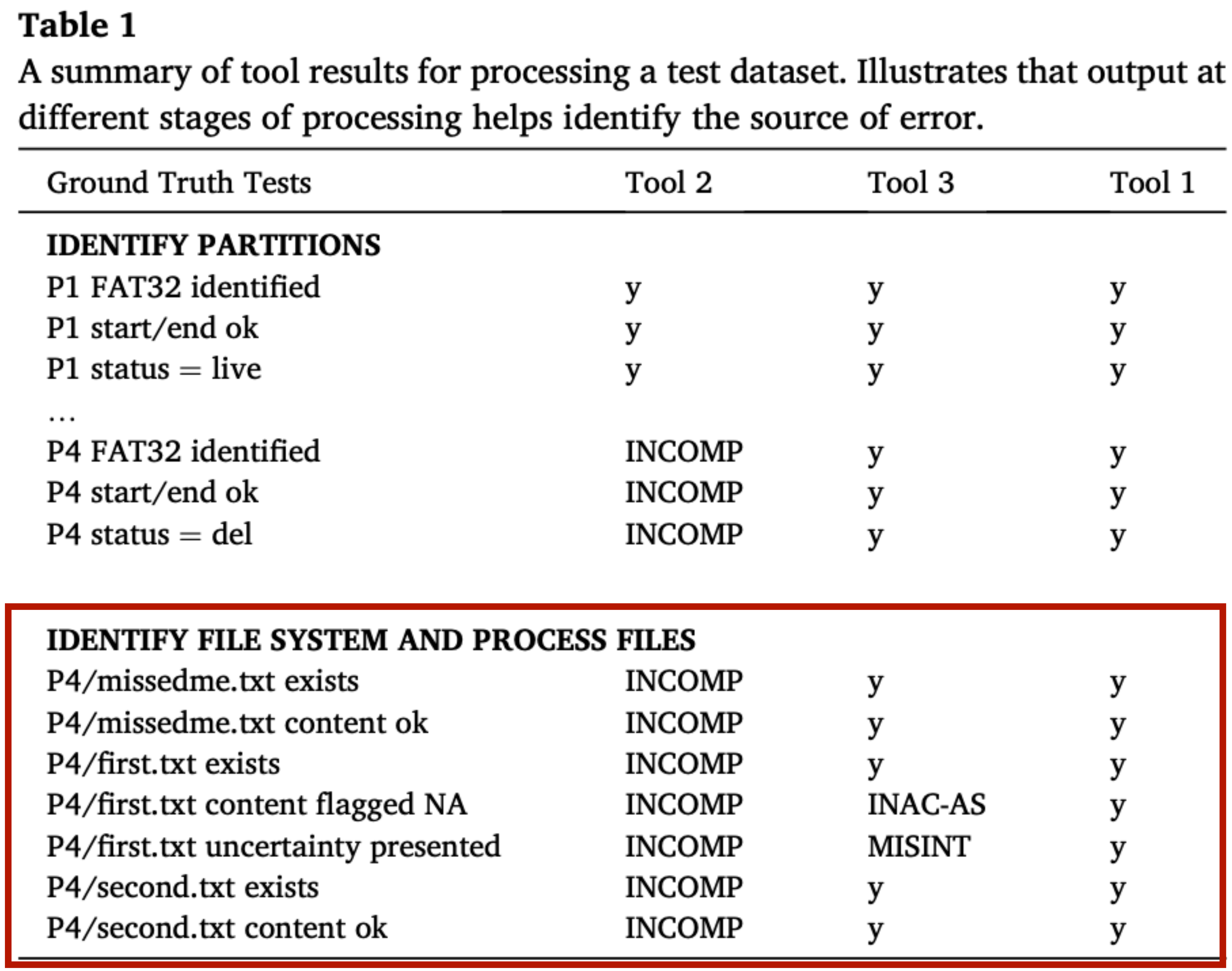


Table 1
A summary of tool results for processing a test dataset. Illustrates that output at different stages of processing helps identify the source of error.

Ground Truth Tests	Tool 2	Tool 3	Tool 1
IDENTIFY PARTITIONS			
P1 FAT32 identified	y	y	y
P1 start/end ok	y	y	y
P1 status = live	y	y	y
…			
P4 FAT32 identified	INCOMP	y	y
P4 start/end ok	INCOMP	y	y
P4 status = del	INCOMP	y	y
IDENTIFY FILE SYSTEM AND PROCESS FILES			
P4/missedme.txt exists	INCOMP	y	y
P4/missedme.txt content ok	INCOMP	y	y
P4/first.txt exists	INCOMP	y	y
P4/first.txt content flagged NA	INCOMP	INAC-AS	y
P4/first.txt uncertainty presented	INCOMP	MISINT	y
P4/second.txt exists	INCOMP	y	y
P4/second.txt content ok	INCOMP	y	y

Identification of existence of, but not the specifics of the problem

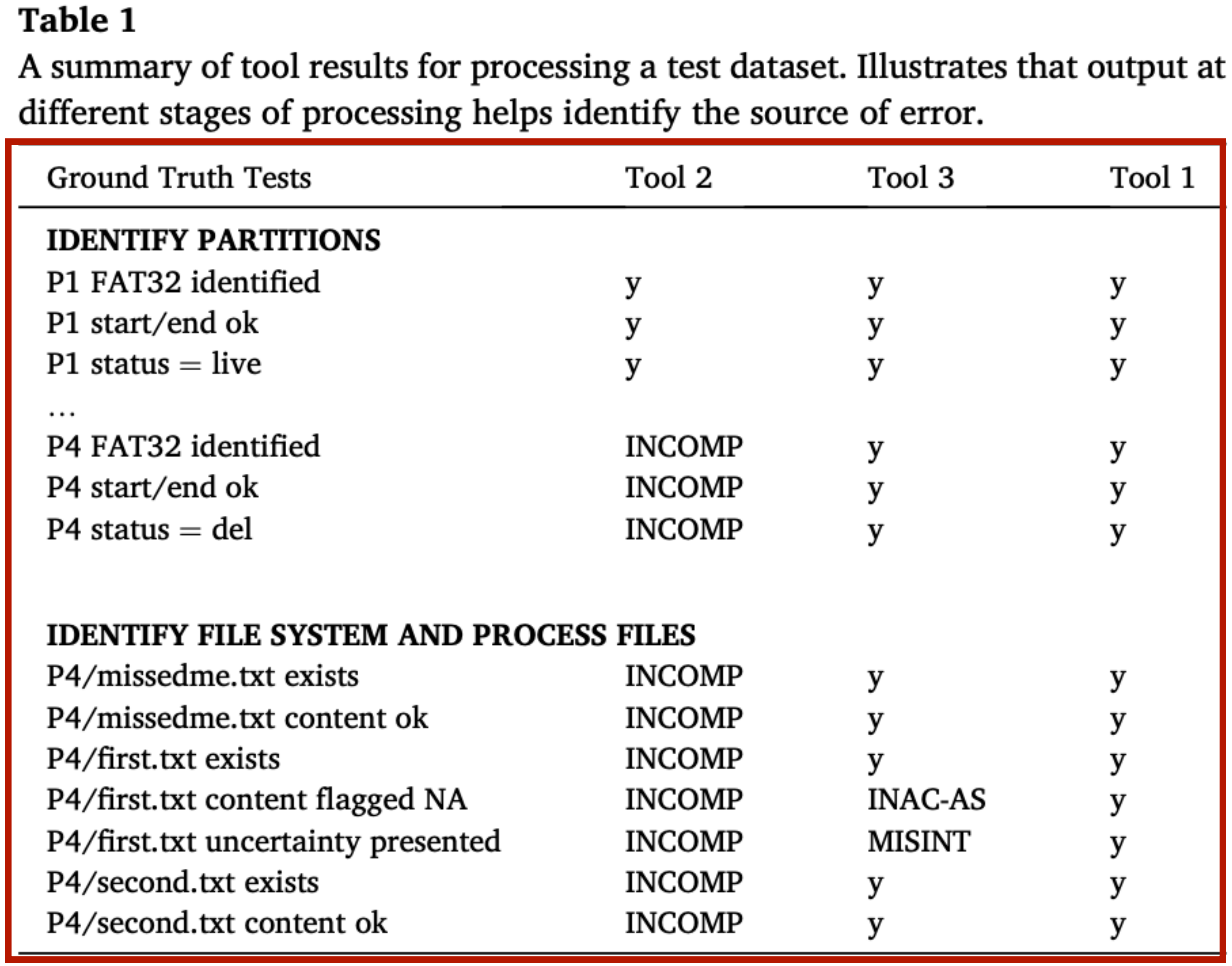


Table 1

A summary of tool results for processing a test dataset. Illustrates that output at different stages of processing helps identify the source of error.

Ground Truth Tests	Tool 2	Tool 3	Tool 1
IDENTIFY PARTITIONS			
P1 FAT32 identified	y	y	y
P1 start/end ok	y	y	y
P1 status = live	y	y	y
…			
P4 FAT32 identified	INCOMP	y	y
P4 start/end ok	INCOMP	y	y
P4 status = del	INCOMP	y	y
IDENTIFY FILE SYSTEM AND PROCESS FILES			
P4/missedme.txt exists	INCOMP	y	y
P4/missedme.txt content ok	INCOMP	y	y
P4/first.txt exists	INCOMP	y	y
P4/first.txt content flagged NA	INCOMP	INAC-AS	y
P4/first.txt uncertainty presented	INCOMP	MISINT	y
P4/second.txt exists	INCOMP	y	y
P4/second.txt content ok	INCOMP	y	y

Identification of existence of, but not the specifics of the problem

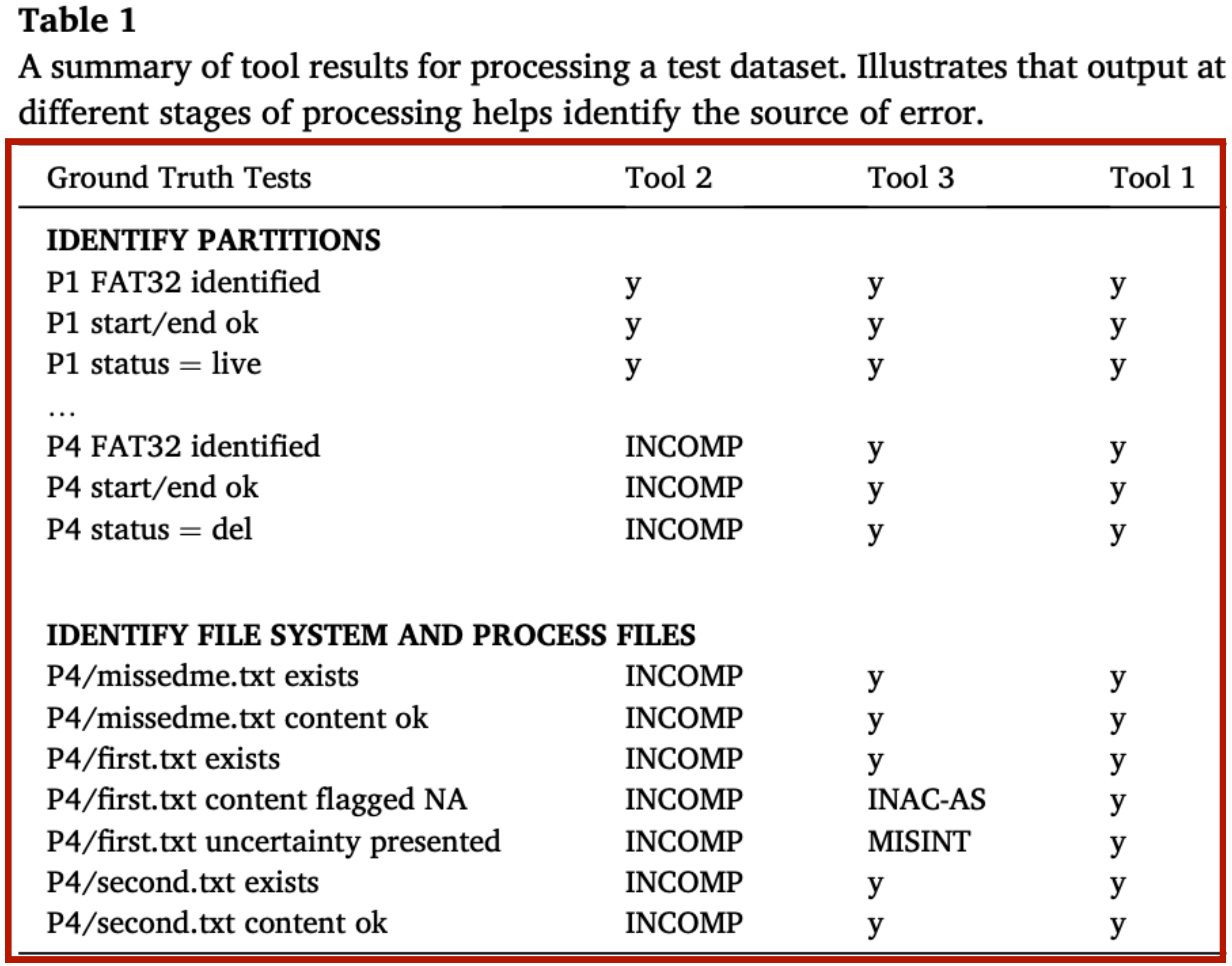


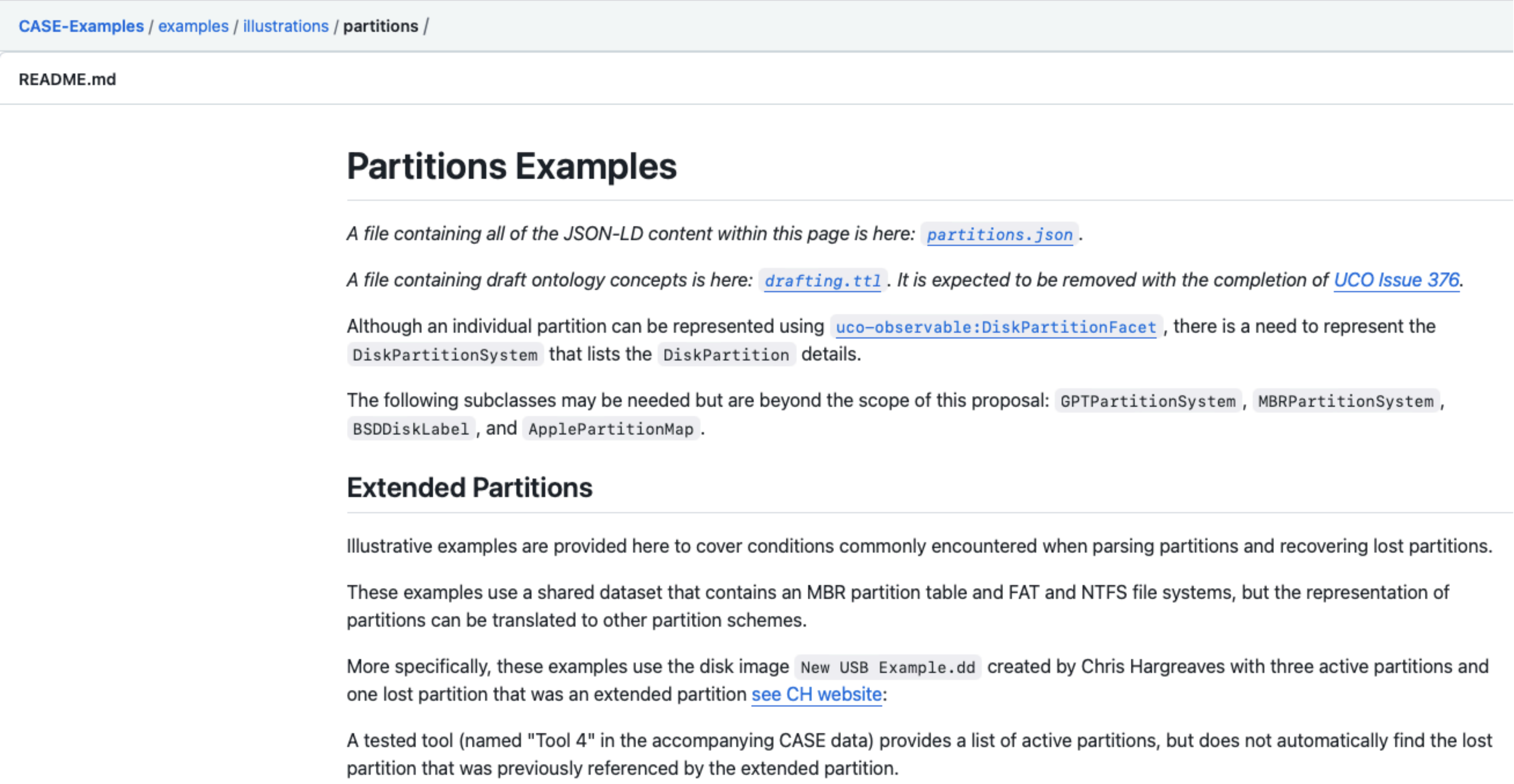
Table 1
A summary of tool results for processing a test dataset. Illustrates that output at different stages of processing helps identify the source of error.

Ground Truth Tests	Tool 2	Tool 3	Tool 1
IDENTIFY PARTITIONS			
P1 FAT32 identified	y	y	y
P1 start/end ok	y	y	y
P1 status = live	y	y	y
…			
P4 FAT32 identified	INCOMP	y	y
P4 start/end ok	INCOMP	y	y
P4 status = del	INCOMP	y	y
IDENTIFY FILE SYSTEM AND PROCESS FILES			
P4/missedme.txt exists	INCOMP	y	y
P4/missedme.txt content ok	INCOMP	y	y
P4/first.txt exists	INCOMP	y	y
P4/first.txt content flagged NA	INCOMP	INAC-AS	y
P4/first.txt uncertainty presented	INCOMP	MISINT	y
P4/second.txt exists	INCOMP	y	y
P4/second.txt content ok	INCOMP	y	y

If tools could output results at intermediate stages, the origin of errors could be more easily identified

CASE Example

Github / Casework / CASE-Examples



CASE-Examples / examples / illustrations / partitions /

README.md

Partitions Examples

A file containing all of the JSON-LD content within this page is here: `partitions.json`*.*

A file containing draft ontology concepts is here: `drafting.ttl`*. It is expected to be removed with the completion of UCO Issue 376.*

Although an individual partition can be represented using `uco-observable:DiskPartitionFacet`, there is a need to represent the `DiskPartitionSystem` that lists the `DiskPartition` details.

The following subclasses may be needed but are beyond the scope of this proposal: `GPTPartitionSystem`, `MBRPartitionSystem`, `BSDDiskLabel`, and `ApplePartitionMap`.

Extended Partitions

Illustrative examples are provided here to cover conditions commonly encountered when parsing partitions and recovering lost partitions.

These examples use a shared dataset that contains an MBR partition table and FAT and NTFS file systems, but the representation of partitions can be translated to other partition schemes.

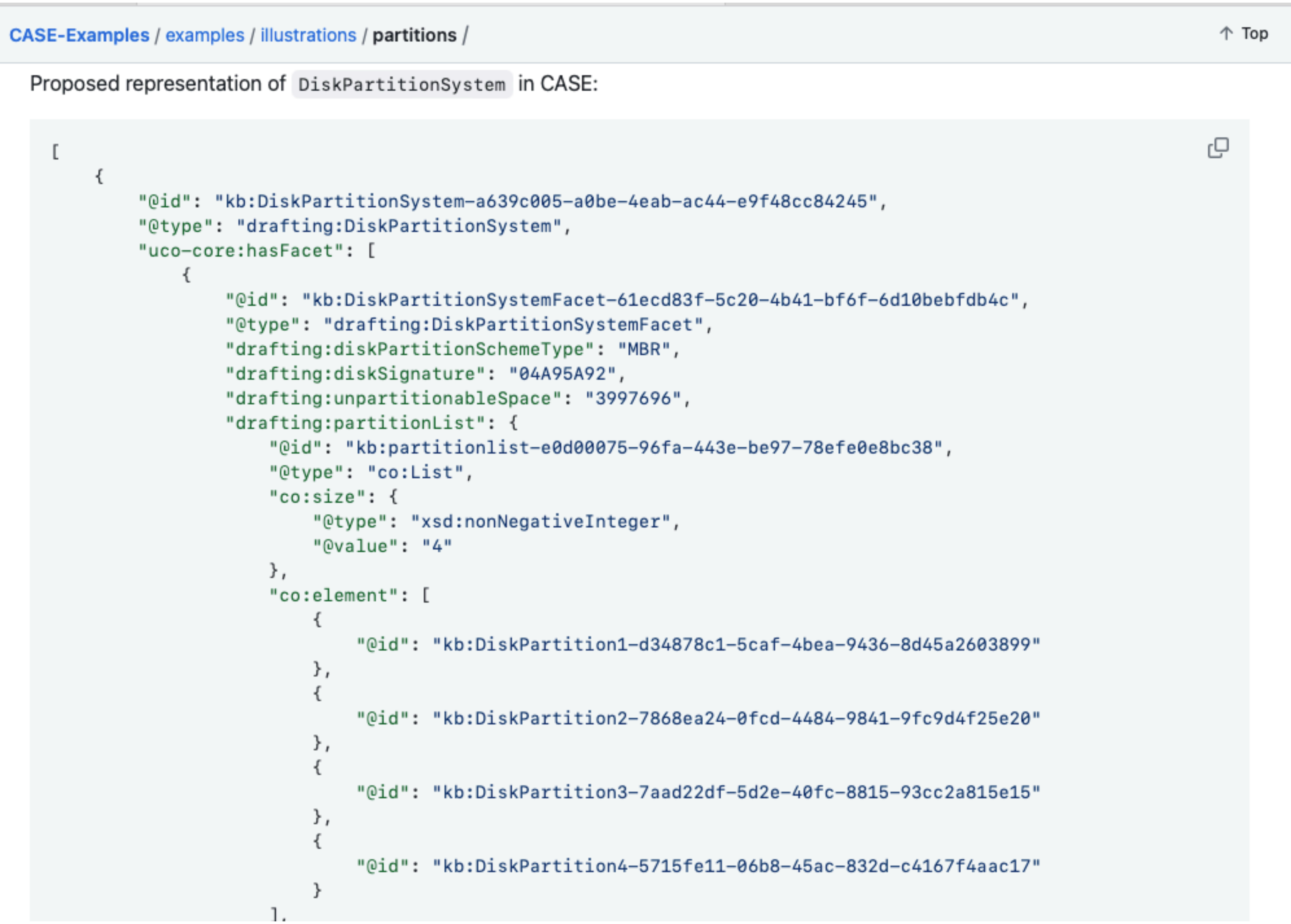
More specifically, these examples use the disk image `New USB Example.dd` created by Chris Hargreaves with three active partitions and one lost partition that was an extended partition see CH website:

A tested tool (named "Tool 4" in the accompanying CASE data) provides a list of active partitions, but does not automatically find the lost partition that was previously referenced by the extended partition.

https://github.com/casework/CASE-Examples/tree/master/examples/illustrations/partitions#partitions-examples

ObservableObjects & Facets

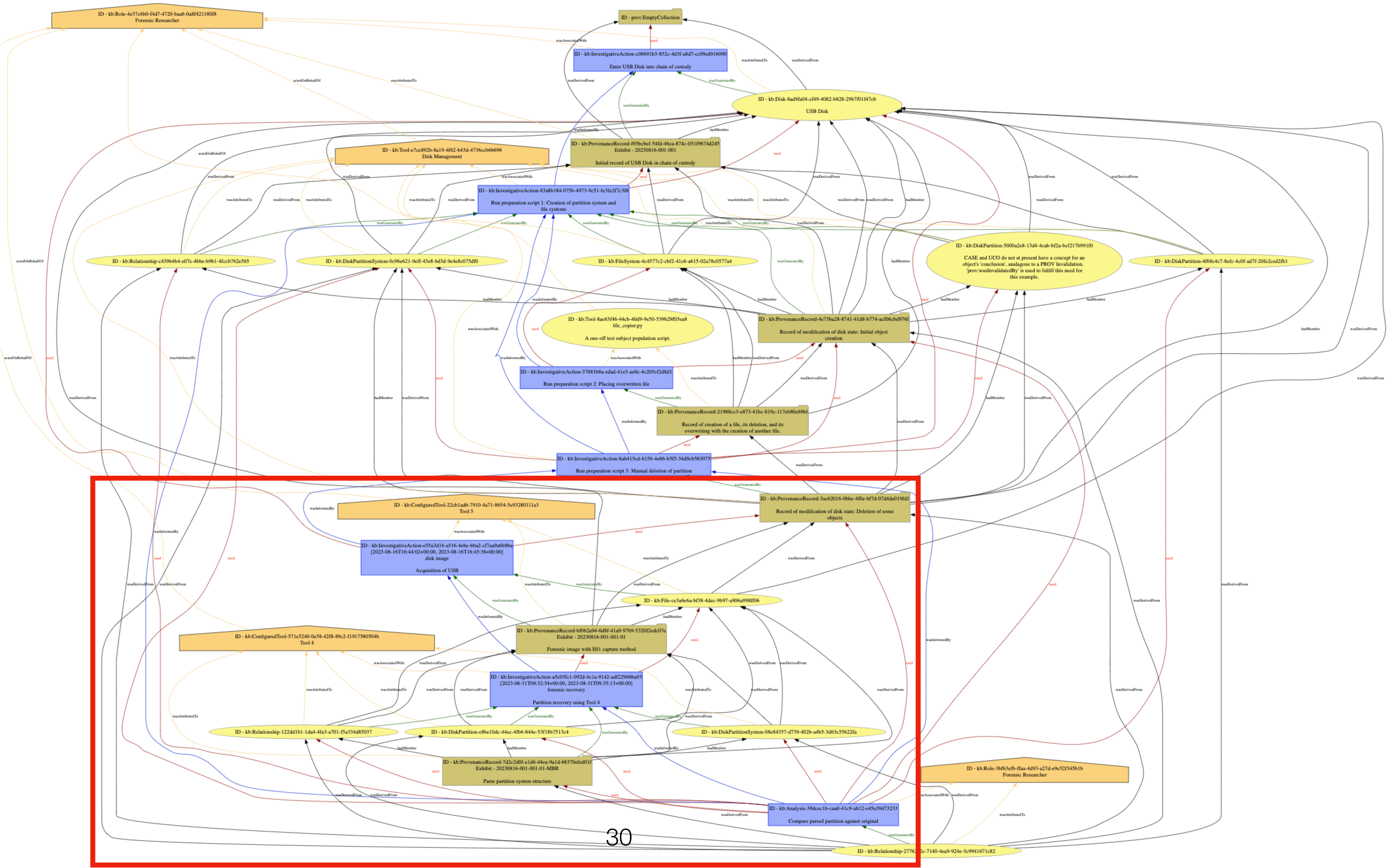
- DiskPartitionSystem
- partitionList
- DiskPartition
- RecoveredObjectFacet
- InvestigativeAction (Analysis)
- AnalyticResult



Proposed representation of `DiskPartitionSystem` in CASE:

```
[
    {
        "@id": "kb:DiskPartitionSystem-a639c005-a0be-4eab-ac44-e9f48cc84245",
        "@type": "drafting:DiskPartitionSystem",
        "uco-core:hasFacet": [
            {
                "@id": "kb:DiskPartitionSystemFacet-61ecd83f-5c20-4b41-bf6f-6d10bebfdb4c",
                "@type": "drafting:DiskPartitionSystemFacet",
                "drafting:diskPartitionSchemeType": "MBR",
                "drafting:diskSignature": "04A95A92",
                "drafting:unpartitionableSpace": "3997696",
                "drafting:partitionList": {
                    "@id": "kb:partitionlist-e0d00075-96fa-443e-be97-78efe0e8bc38",
                    "@type": "co:List",
                    "co:size": {
                        "@type": "xsd:nonNegativeInteger",
                        "@value": "4"
                    },
                    "co:element": [
                        {
                            "@id": "kb:DiskPartition1-d34878c1-5caf-4bea-9436-8d45a2603899"
                        },
                        {
                            "@id": "kb:DiskPartition2-7868ea24-0fcd-4484-9841-9fc9d4f25e20"
                        },
                        {
                            "@id": "kb:DiskPartition3-7aad22df-5d2e-40fc-8815-93cc2a815e15"
                        },
                        {
                            "@id": "kb:DiskPartition4-5715fe11-06b8-45ac-832d-c4167f4aac17"
                        }
                    ],
```

Visualisation


ID - kb:Role-4e37c6b0-f4d7-4720-baa6-0a8f421180f8
Forensic Researcher
ID - prov:EmptyCollection
ID - kb:InvestigativeAction-c08691b3-852c-4d3f-a8d7-cc09ed916090
Enter USB Disk into chain of custody
ID - kb:Disk-8ad9fa04-cf49-4082-b828-29b7f01f47cb
USB Disk
ID - kb:Tool-e7ce892b-8a19-4f62-b43d-4736ccb6b696
Disk Management
ID - kb:ProvenanceRecord-f95bcbef-54fd-46ca-874c-051f9674d245
Exhibit - 20230816-001-001
Initial record of USB Disk in chain of custody
ID - kb:InvestigativeAction-83a8b184-075b-4973-9c51-fe3fe2f7c3f8
Run preparation script 1: Creation of partition system and file systems
ID - kb:Relationship-c439b4b4-e07e-4bbe-b9b1-4fccb762e565
ID - kb:DiskPartitionSystem-0c96e621-9eff-43e8-bd3d-9e4e8c075df0
ID - kb:FileSystem-4c4577c2-cbf2-41c6-a615-02a78c0577a4
ID - kb:DiskPartition-5000a2e8-13d4-4cab-bf2a-bcf217b991f0
CASE and UCO do not at present have a concept for an object's 'conclusion', analagous to a PROV Invalidation. 'prov:wasInvalidatedBy' is used to fulfill this need for this example.
ID - kb:DiskPartition-4f68c4c7-8efc-4c0f-ad7f-20fe2ced2fb1
ID - kb:Tool-8ae63f46-44cb-40d9-9e50-539b29f05ea8
file_copier.py
A one-off test subject population script.
ID - kb:ProvenanceRecord-4e73ba28-8741-41d8-b774-acf06cbd976f
Record of modification of disk state: Initial object creation
ID - kb:InvestigativeAction-37881b8a-edad-41e3-ae0c-4c205cf2d8d1
Run preparation script 2: Placing overwritten file
ID - kb:ProvenanceRecord-21988cc3-e873-41bc-819c-117eb80e69bf
Record of creation of a file, its deletion, and its overwriting with the creation of another file.
ID - kb:InvestigativeAction-8ab415cd-6156-4e86-b5f5-34d9cb583075
Run preparation script 3: Manual deletion of partition
ID - kb:ProvenanceRecord-3ac62016-0bbe-4f0e-bf7d-07d4de019fd1
Record of modification of disk state: Deletion of some objects
ID - kb:ConfiguredTool-22cb1ad6-7910-4a71-8654-5e93280111a3
Tool 5
ID - kb:InvestigativeAction-e55a3d16-a516-4ebe-bba2-cf7aa9a6b8ba
[2023-08-16T16:44:02+00:00, 2023-08-16T16:45:38+00:00]
disk image
Acquisition of USB
ID - kb:File-ce3a6e6a-bf38-4dec-9b97-a906a998ff06
ID - kb:ConfiguredTool-571e5240-0e58-42f8-89c2-f1917580504b
Tool 4
ID - kb:ProvenanceRecord-bf062a94-6d8f-41a9-97b9-5320f2edc07e
Exhibit - 20230816-001-001-01
Forensic image with E01 capture method
ID - kb:InvestigativeAction-a5c03fc1-092d-4c1a-9142-adf22966ba93
[2023-08-31T09:32:54+00:00, 2023-08-31T09:35:13+00:00]
forensic recovery
Partition recovery using Tool 4
ID - kb:Relationship-122dd1b1-1da4-4fa3-a701-f5a334d85037
ID - kb:DiskPartition-e8be10dc-44ec-4fb6-844e-53f18b7513c4
ID - kb:DiskPartitionSystem-08e84357-d739-402b-a4b5-3d63c55622fa
ID - kb:ProvenanceRecord-7d2c2d0f-e1d6-44ea-9a1d-68370efed01f
Exhibit - 20230816-001-001-01-MBR
Parse partition system structure
ID - kb:Role-3bfb3efb-ffaa-4d93-a27d-e9e52f345b1b
Forensic Researcher
ID - kb:Analysis-39dcec1b-caa0-41c9-ab12-e45a39d73233
Compare parsed partition against original
ID - kb:Relationship-2776[illegible]2c-7140-4ea9-924e-5c9941671c82
wasAssociatedWith
used
wasDerivedFrom
wasGeneratedBy
wasAttributedTo
actedOnBehalfOf
wasInformedBy
hadMember

Visualisation

ID - kb:ConfiguredTool-22cb1ad6-7910-4a71-8654-5e93280111a3
Tool 5
ID - kb:ProvenanceRecord-3ac62016-0bbe-4f0e-bf7d-07d4de019fd1
Record of modification of disk state: Deletion of some objects
wasInformedBy
wasInformedBy
wasAssociatedWith
used
used
used
wasAttributedTo
wasDerivedFrom
ID - kb:InvestigativeAction-e55a3d16-a516-4ebe-bba2-cf7aa9a6b8ba
[2023-08-16T16:44:02+00:00, 2023-08-16T16:45:38+00:00]
disk image
Acquisition of USB
wasDerivedFrom
wasDerivedFrom
wasAttributedTo
wasGeneratedBy
wasDerivedFrom
wasGeneratedBy
ID - kb:File-ce3a6e6a-bf38-4dec-9b97-a906a998ff06
wasInformedBy
hadMember
ID - kb:ProvenanceRecord-bf062a94-6d8f-41a9-97b9-5320f2edc07e
Exhibit - 20230816-001-001-01
Forensic image with E01 capture method
used
ID - kb:ConfiguredTool-571e5240-0e58-42f8-89c2-f1917580504b
Tool 4
wasAssociatedWith
wasDerivedFrom
used
wasDerivedFrom
wasDerivedFrom
used
ID - kb:InvestigativeAction-a5c03fc1-092d-4c1a-9142-adf22966ba93
[2023-08-31T09:32:54+00:00, 2023-08-31T09:35:13+00:00]
forensic recovery
Partition recovery using Tool 4
wasAttributedTo
wasAttributedTo
wasDerivedFrom
wasDerivedFrom
wasAttributedTo
wasDerivedFrom
wasDerivedFrom
wasAttributedTo
wasGeneratedBy
wasGeneratedBy
wasGeneratedBy
wasDerivedFrom
ID - kb:Relationship-122dd1b1-1da4-4fa3-a701-f5a334d85037
ID - kb:DiskPartition-e8be10dc-44ec-4fb6-844e-53f18b7513c4
wasGeneratedBy
ID - kb:DiskPartitionSystem-08e84357-d739-402b-a4b5-3d63c55622fa
hadMember
hadMember
wasInformedBy
hadMember
ID - kb:ProvenanceRecord-7d2c2d0f-e1d6-44ea-9a1d-68370efed01f
Exhibit - 20230816-001-001-01-MBR
Parse partition system structure
used
used
used
used
wasDerivedFrom
wasDerivedFrom
ID - kb:Analysis-39dcec1b-caa0-41c9-ab12-e45a39d73233
Compare parsed partition against original
wasDerivedFrom
wasGeneratedBy
ID - kb:Relationship-27762

Suggested output at each layer of abstraction

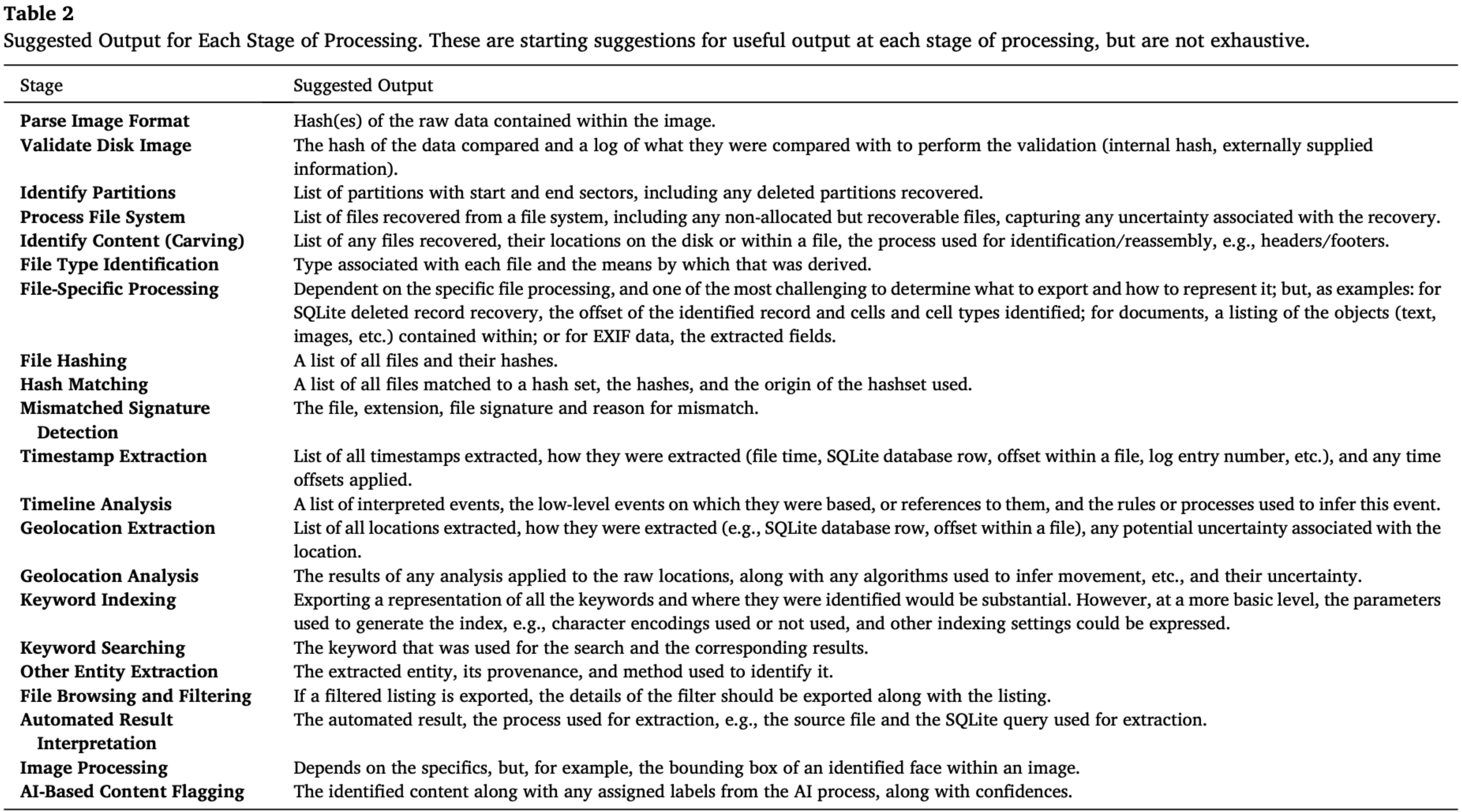


Table 2
Suggested Output for Each Stage of Processing. These are starting suggestions for useful output at each stage of processing, but are not exhaustive.

Stage	Suggested Output
Parse Image Format	Hash(es) of the raw data contained within the image.
Validate Disk Image	The hash of the data compared and a log of what they were compared with to perform the validation (internal hash, externally supplied information).
Identify Partitions	List of partitions with start and end sectors, including any deleted partitions recovered.
Process File System	List of files recovered from a file system, including any non-allocated but recoverable files, capturing any uncertainty associated with the recovery.
Identify Content (Carving)	List of any files recovered, their locations on the disk or within a file, the process used for identification/reassembly, e.g., headers/footers.
File Type Identification	Type associated with each file and the means by which that was derived.
File-Specific Processing	Dependent on the specific file processing, and one of the most challenging to determine what to export and how to represent it; but, as examples: for SQLite deleted record recovery, the offset of the identified record and cells and cell types identified; for documents, a listing of the objects (text, images, etc.) contained within; or for EXIF data, the extracted fields.
File Hashing	A list of all files and their hashes.
Hash Matching	A list of all files matched to a hash set, the hashes, and the origin of the hashset used.
Mismatched Signature Detection	The file, extension, file signature and reason for mismatch.
Timestamp Extraction	List of all timestamps extracted, how they were extracted (file time, SQLite database row, offset within a file, log entry number, etc.), and any time offsets applied.
Timeline Analysis	A list of interpreted events, the low-level events on which they were based, or references to them, and the rules or processes used to infer this event.
Geolocation Extraction	List of all locations extracted, how they were extracted (e.g., SQLite database row, offset within a file), any potential uncertainty associated with the location.
Geolocation Analysis	The results of any analysis applied to the raw locations, along with any algorithms used to infer movement, etc., and their uncertainty.
Keyword Indexing	Exporting a representation of all the keywords and where they were identified would be substantial. However, at a more basic level, the parameters used to generate the index, e.g., character encodings used or not used, and other indexing settings could be expressed.
Keyword Searching	The keyword that was used for the search and the corresponding results.
Other Entity Extraction	The extracted entity, its provenance, and method used to identify it.
File Browsing and Filtering	If a filtered listing is exported, the details of the filter should be exported along with the listing.
Automated Result Interpretation	The automated result, the process used for extraction, e.g., the source file and the SQLite query used for extraction.
Image Processing	Depends on the specifics, but, for example, the bounding box of an identified face within an image.
AI-Based Content Flagging	The identified content along with any assigned labels from the AI process, along with confidences.

Potential benefits

- Improvements to tool testing and validation process
- Help with building error-focused datasets
- Dual tool verification during a case
- Historic case re-evaluation
- Improved artefact provenance

Summary

- Tool error is, and will likely always be a problem.
- Proposed solutions are to make everything open source, do more validation and testing, rely on dual tool verification etc.
- However, this is a huge undertaking that doesn’t scale or distribute well. Or disengages with a huge section of the digital forensics community (e.g. commercial tool vendors).
- Alternatively, adding intermediate output is a pragmatic approach to cross-tool validation, or validation against carefully constructed error-focused data sets, and CASE can help facilitate that.
- This abstraction of tool features allows datasets to be constructed that aren’t aiming to “show that it works”, they actively try to target specific features and explore edge cases.

Future work

- Refinement of abstract model, including detail of ‘file specific processing’
- Further work on potential errors at each stage
- Mapping ‘suggested output’ to CASE ontology
- Potentially implement intermediate output, at least in open source tools
- Integration with Continuous Integration framework
- It is now easier to structure effort on designing and building *error focused* data sets

Questions?