LEAP – Technical Feature Overview

Overview of relevant Data Formats

LEAP is designed to be agnostic and complimentary to existing 3rd party extraction toolkits. Our solution is designed to be flexible and customisable, and our aim is to continually expand support for different data formats as and when appropriate.

Supported Data Input Formats

LEAP currently supports the following import formats:

Device Analysis

- Cellebrite UFDR
- MSAB Extended XML
- Oxygen Forensics XML
- MobilEdit XML & UFDR
- Grayshift .ZIP file system (currently media & document analysis only)

Media Analysis

LEAP media analysis supports all common video and image formats in **raw or zip format**. For a complete list please refer to the following:

- https://pillow.readthedocs.io/en/5.1.x/handbook/image-file-formats.html (fully supported and read only format sections)
- https://www.ffmpeg.org/general.html#Supported-File-Formats_002c-Codecs-or-Features

Forensic Images – image, video, document, and audio analysis only

- E01 forensic image
- .DD disk image

Supported File systems:

- NTFS
- FAT
- FAT32
- exFAT
- EXT3
- EXT4
- HFS+

Forensic images containing encrypt partitions are not supported. Recovery of data from unallocated space or within compound files not currently supported.

Audio Analysis (optional)

- Keyword spotting with audio watchlist – supports German, English, Arabic, French, Dutch
- The following audio file types are supported - https://www.ffmpeg.org/general.html#Audio-Codecs

Document Analysis

- Searching using watchlists and OCR
- Embedded image file extraction and object recognition

Supported document types:

- Microsoft Office (doc, docx, xls, xlsx, ppt. pptx)
- Open Office (odt, odp)
- PDF
- RTF
- TXT

Watchlists

LEAP supports the import of watchlists in the following formats:

- .txt
- LEAP custom JSON format

LEAP Analysis Profiles

LEAP supports the import of profiles in the following formats:

- ZIP
- LEAP custom JSON format

Hash Database Analysis

- LEAP supports the import of Project VIC JSON files v1.3 and v2.0

Supported Data Output Formats

LEAP currently supports the following output formats:

Automated Report

- HTML
- PDF
- PDF/A
- XML (optional feature enabled on request with custom content, not including media)

Hash Report

- CSV

Watchlists

- LEAP custom JSON format

LEAP Analysis Profiles

- LEAP custom JSON format
- ZIP

LEAP.Legal Communication Export

Included within the LEAP.Legal profile is the ability to export call logs, address books, emails, SMS and chat messages found within the extraction data in JSON or CSV format. This includes all fields allocated to each message, example field categories can include, but are not limited to:

- Recipient & Receiver Information
- Time & Date Information
- Attachment name
- App type (Gmail, WhatsApp, Viber etc)
- Message type (sent, received etc)
- Message Body

Technical Breakdown of Analysis Types

Depending on the options chosen and the data obtained by the 3rd party extraction toolkit, LEAP performs a variety of different analysis steps, which can be summarized as follows:

User Account Information

User account information present within the extraction data is aggregated and presented in tabular form. This information can be invaluable in establishing a person's identity. Typical aggregated data can include:

- Social Media Account information (Facebook, Instagram, YouTube etc)
- Messaging Applications (Whatsapp, Viber, Telegram, Skype, Facebook Messenger etc)
- Email Account Information (Gmail, Microsoft Outlook, Yahoo etc)
- Online Account information (cached credentials from website login accounts e.g. travel, retail etc)

Country Analysis

To determine a person's descent, several sources are used to check in- and outgoing communication for their countries of origin. The following categories are distinguished:

- Calls (phone and messenger)
- Text messages (SMS and messenger)
- Contacts in the address book
- Browser history, favourites and cookies
- E-mails

Language Analysis

Written texts found and extracted from the phone (i.e. messenger, notes) are analysed regarding the used languages. More than 170 languages are currently identified and in addition to this, 18 Arabic dialects are also detected.

Language identification is also supported during audio analysis for specific languages – please see the Audio Analysis section

Image & Video Analysis

Images and videos extracted from the device, forensic image or uploaded from external sources are matched against predefined classes:

- Identification and travel documents (passports, credit cards, driver's licences, etc.)
- Vehicle licence plates
- Screenshots of maps
- Terrorist propaganda and symbolism
- Weapons
- Military uniforms and masked individuals
- Documents (including OCR)
- Child Sexual Abuse Material (CSAM)

- Age & Gender detection (default 0-15 years)
- Endangered Wildlife

In addition to predefined classes, image content with dynamic nature can also be detected:

- Faces – a reference image is uploaded to LEAP and chosen as part of the analysis options. LEAP automatically finds similar faces within image files.

Audio Analysis (Optional)*

Using a 3rd party integration, LEAP supports **keyword searching** from audio files contained within a device extraction: At present the following languages are technically supported for keyword spotting but not all are available by default in LEAP:

5th generation models:
Arabic (Gulf), Arabic (Levantine), Croatian, Czech, Dutch, English (US), French, Polish, Russian, Slovak, Spanish, Swedish
4th and older generation models:
Arabic, Chinese, Croatian, Czech, Dutch, English (US), Farsi, French, German, Hungarian, Italian, Pashto, Polish, Russian, Slovak, Spanish, Turkish

**Additional licence required.*

Geo-Location Analysis

Locations saved on the phone are screened, additionally all decoded location data received with the device is displayed separately. Location data is generally extracted from the following data categories:

- Pictures
- Apps (navigation, messenger, others)
- WiFi connections
- Cell towers
- System data

Watchlists

Data from extractions is checked against known factors from the following categories and potential findings are displayed. The following table describes the different types of watchlists available:

Watchlist Type	Checks	Method	Example (*your input* vs. *found data*)
General	App names, Contents of Messages (eg chats, SMS etc), Calls, Contents of Emails, Address Book, Browser Data and File Names When OCR is enabled also checks text within media files When Document Analysis is enabled also checks for text within Document Files	Case insensitive exact match anywhere in any of the categories.	*today* matches *What are you doing today?* *today* does not match *What are you doing tdoay?* For further examples refer to Apps, Contact and URL
General (Language defined)	Audio Files	Case insensitive exact match anywhere within audio files.	*today* matches *instances where the word 'today' is spoken/present inside audio files*
App	Installed apps	For apps with less than 4 characters LEAP matches the exact phrase given. For longer app names the match is not limited to the exact phrase.	*facebook* matches *Facebook Messenger* *fa* does not match *Facebook Messenger*
Contact	SMS, messages, contacts, calls	Either the phone number or the contact name is matched with the information from the extraction. Please make sure to use the correct format for telephone numbers (use national code).	*436601234567* matches *06601234567* *436601234567* does not match *+01 01234567* *Max Mustermann* matches *max mustermann* *Max Mustermann* does not match *maxi*

Country	Own location data	The country code found in the location data is compared with the countries of the watchlist.	*China* matches *CN* (People's Republic of China) *China* does not match *TW* (Republic of China, Taiwan) For this match, choose Taiwan instead.
Hash	File hashes – media,document, audio	Case insensitive exact match of given hash and file hash in full length	*9f090a3f986a1e2d4d246e3efec87bacc791afc* matches *9f090a3f986a1e2d4d246e3efec87bacc791afc* but not *9f090a3f986a1e2d4d246e3efec87bacc791000*.
Location	Own location data	Match if given location is in given area range (radius) of location data.	*16.3875308, 48.1905363, 500* matches *16.38615,48.1904225* *16.3875308, 48.1905363, 500* does not match *16.3485455, 48.1814928*
PhotoDNA	Images	The photoDNA hashes of images are calculated and compared with given hashes. The hash is either an array of numbers or base 64 representation of the array. The distance between the given hash and calculated hash is calculated and compared	Array: 5,69,1,77,27,17,8,...,67,87,14,183,21 Base64 Hash: BUUBTRsRCDoOFxA3Dhs.....0pQOiniUT6CY4AqVUNXDrcV Shortened versions, since photoDNA hashes have the length of 144.
URL	Browser history, cookies, SMS, messages	Given URL anywhere in URLS from browser or in SMS/message texts	*t3k forensics.com* matches *http://www.t3k-forensics.com/analytics-en/* *t3k-forensics.com* does not match *https://www.linkedin.com/company/t3k-forensics/*

Optical Character Recognition

LEAP has an inbuilt multilingual OCR functionality to allow text to be displayed from images, this also enables watchlists to locate keyword hits from within image, video and document files.

Supported languages:

- German
- English
- Arabic
- French
- Other languages using standard Latin and Cyrillic alphabetic characters

By default, OCR text is extracted from the following media object classes (when detected):

- Screenshot
- Document
- Passport

- Identity card
- Text (e.g. text within a picture or video)
- Licence plates

When enabled, OCR can also be performed across documents which do not have a text layer, eg PDF.

Activity Analysis

Activity data considers all data which indicate with high probability their creation by the device itself and indication that the device is in use. This analysis type can include:

- Call data
- SMS
- Browser data
- System data

Examples include establishing the most contacted people on the device, the longest call duration, first calls and messages to the device etc

Interlink Correlation Analysis

Several extractions are compared with each other for a correlation analysis, checking the respective extractions for similarities or hints of contact between the users. Extractions of any supported data format can be compared, e.g. UFDR against XRY XML. The following categories are used for a correlation analysis:

- Direct contacts between extractions (incl. period of contact)
- Common contacts (incl. period of contact)
- Common languages and dialects
- Common locations (time-referenced and general)
- Common WIFI's
- Common websites
- Common apps (incl. order according to anomaly factor)
- File hash correlation (MD5/SHA1)
- General Watchlist correlation – multiple devices compared to the same watchlist simultaneously

Biometric Analysis

LEAP includes the following automated biometric detection options:

- Face detection – LEAP AI matches a sample face, uploaded from within the GUI, to similar faces found within images and videos within the mobile device extraction data or in raw media files from an external source

Hash Database Analysis

LEAP allows Project VIC JSON files to be uploaded via the configuration menu in the LEAP GUI.

Allows the user to choose one or more of the following Hash comparison options:

- MD5
- SHA1
- PhotoDNA

Results are displayed in a separate section of both automated PDF and HTML reports.

Pattern Recognition*

LEAP includes the ability to detect distinct 2D patterns within media content using custom computer vision technology.

Example uses include but are not limited to detecting:

- Logos/symbols (e.g. extremist insignias and symbols, corporate logos)
- Wallpaper designs (could be used to detect backgrounds and or locations in illicit images, e.g. CSAM)
- Tattoos
- Currency (bank notes, bank cheques with a unique design)
- Document header and footer designs (corporate or technical documentation)
- Works of art

Pattern recognition requires a minimum of one distinct example of the pattern or logo to be uploaded to LEAP and then selected as part of the analysis options or as part of a profile. Results are displayed in a separate section in both PDF and HTML reports.

*Feature is in beta testing but can be activated in LEAP for evaluation on request.