About Developing a Cross-Check System for Judicial Case Searching and Correlation

Gerardo Giardiello
Fabrizio Turchi
Institute of Legal Informatics and Judicial Systems,
National Research Council of Italy (CNR-IGSG)[1]

Abstract

In a recent EU publication, a report commissioned by the European Union related to the Cross-border Digital Criminal Justice environment, a set of specific business needs have been identified. Some of the most relevant ones have been: i) the interoperability across different systems needs to be ensured, ii) the stakeholders need to easily manage the data and ensure its quality, allowing them to properly make use of it (e.g. use the data as evidence in a given case) and iii) the stakeholders investigating a given case should be able to identify links between cross-border cases. Therefore, solutions are needed to allow the stakeholder to search and find relevant information they need for the case they are handling. The article presents a set of solutions to address the highlighted needs, including a 'Judicial Cases Cross-Check System'. Such a system should provide a tool being able to search for case-related information and identify links among cases that are being investigated in other EU Member States or by Justice and Home Affairs (JHA) agencies and EU bodies. To facilitate the development of the above solution, a standard representation of the metadata and data of the evidence should be adopted. In particular the Unified Cyber Ontology (UCO) and Cyber-investigation Analysis Standard Expression (CASE), dedicated to the digital forensic domain, seem the most promising one to this aim. Moreover it provides a structured specification for representing information that are analysed and exchanged during investigations involving digital evidence.

Keywords: Judicial Case Correlation, Evidence Standard, Case Ontology, Judicial Investigation

Introduction

A recent report prepared for the European Commission on Cross-border Digital Criminal Justice (Debski et al. 2020), has highlighted how the modernisation of judicial cooperation is paramount for an efficient fight against crime in view of the rapid progress of the technologies and their potential malicious or threatening use. EU Member States and Justice and Home Affairs (JHA) agencies stressed the need to be able to search for case-related information and identify correlations with cases under investigation in other EU Member

[1] Authors‘ emails: gerardo.giardiello@igsg.cnr.it, fabrizio.tuchi@igsg.cnr.it.

States/JHA agencies and EU bodies. To accomplish that goal and devise a technical solution, it is important to address all the issues raised by the involved stakeholders, of which the most important are:

- Is it acceptable, for all Member States, to maintain a central criminal cases database considering the potential data protection issues?
- Would a central database require a legal basis?
- Keeping information/evidence at national level and providing query system from abroad might be an alternative solution but would it generate a duplication of the same data being stored in multiple systems?

In this article none of the above issues will be addressed, instead the focus will be put on the standard representation of the metadata and data of a piece of evidence, based on an ontology that has been developing as an open and cost-free resource for the digital forensic community in a broad sense, including all the stakeholders involved in cross-border judicial cooperation.

Nevertheless, the adoption of the standard (see Section "The UCO/CASE standard") would facilitate both central and distributed technical solutions. In the case of a central solution, the first option, it is easy to imagine how powerful a system can be, considering that each digital trace would be represented in the same formally structured manner. In the second option, the distributed solution, the investigative information could be easily retrieved relying on the metadata representation of the pieces of evidence, also taking into account that UCO/CASE provides specific explicit ontology properties to support appropriate handling of shared information, based on the Information Exchange Policy[2] or the Traffic Light Protocol,[3] and also on enhancing data protection and intelligent analysis of digital evidence (Casey, E., Barnum, S., Griffith, R., Snyder, J., van Beek, H. & Nelson, A. (2017).

The UCO/CASE standard

UCO[4] stands for Unified Cyber Ontology, a foundation for standardized information representation across the cyber security domain; CASE[5] that stands for Cyber-Investigation Analysis Standard. UCO/CASE provides a standard language, actually a set of ontologies, for representing information collected, extracted, analysed and exchanged during investigations involving digital evidence. UCO/CASE is a community-developed ontology designed to provide a standard for interoperability and analysis of investigative information in a broad range of cyber-investigation domains, including digital forensic science, incident response, counter-terrorism, criminal justice, forensic intelligence. The UCO/CASE community is a consortium of academic, government and law enforcement, plus commercial and non-profit organisations. To perform digital investigations effectively, there is a pressing need to harmonise how information significant to cyber-investigations is represented and exchanged. UCO/CASE enables the merge of information from different data sources and forensic tool outputs to allow more comprehensive and cohesive analysis (Casey et al., 2018). The main UCO/CASE goals are:

- to foster Interoperability between digital investigation systems and tools;
- to automate normalisation and combination of differing data sources to facilitate analysis and exploration of investigative questions (who, when, where, what, etc.), maintaining provenance at all phases of digital investigation lifecycle;
- to ensure all analysis results are traceable to their sources (Chain of Evidence).

The first two points foster the development of a Judicial Cases Cross-Check system for case searching and correlation, based on the interoperability and normalisation of the data and metadata. The last point is more connected to the admissibility of a piece of evidence because it reveals which file a relevant digital trace comes from.

[2] Information Exchange Policy (IEP), https://www.first.org/iep

[3] Traffic Light Protocol Definitions and Usage, https://www.cisa.gov/tlp

[4] Unified Cyber Ontology (UCO) A foundation for standardized information representation across the cyber security domain/ecosystem, see unifiedcyberontology.org.

[5] An international standard supporting automated combination, validation, and analysis of cyber-investigation information, see caseontology.org.

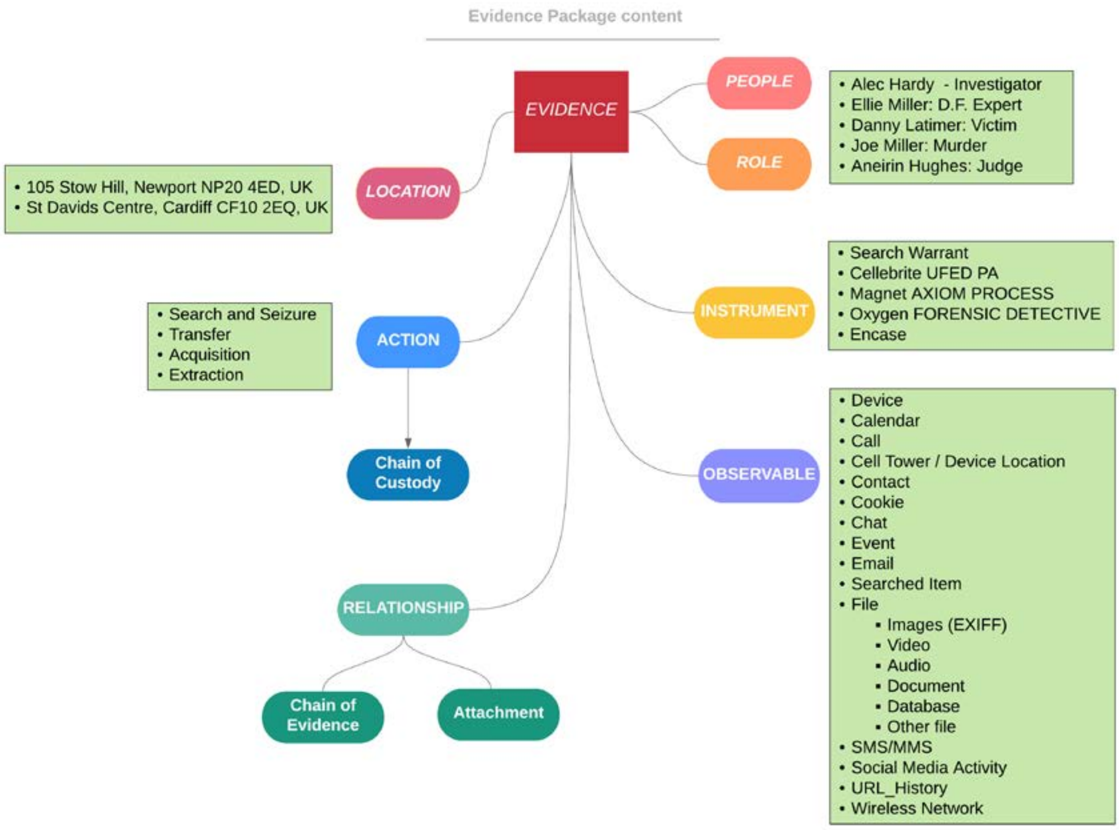
These features allow significant advantages because they comprise a wider view of the need for representing relevant cyber information and the adoption of more neutral solutions without promoting proprietary models. On the other hand, the approval of change proposals or the implementation of new cyber items to be included in the ontologies is quite slow because a broad consensus is needed/required from all the members of the community.

UCO/CASE ontology classes

The ontologies consist of the following main classes (see Figure 1):

- People involved in the evidence life-cycle, from search and seizure to the report before the Court, technical and legal (subjects, victims, authorities, examiners etc.).
- Surrounding information about Legal authorization (i.e., search warrant).
- Information about the Process/Lifecycle (i.e. seizing, acquisition, analysis etc.).
- Information about the Chain of custody by identifying Who did What, When and Where from the moment the Evidence has been gathered.
- Actions performed by people (seizing, acquisition, analysis etc.).
- Source of evidence, that is physical objects involved in the investigative case (e.g., hard disk, smartphone) but even digital source of evidence (i.e., memory dump).
- Description of the Objects inside the digital evidence and their Relationships (e.g., *Contained_Within, Extracted_From* etc.).

Figure 1: UCO/CASE ontology, main Classes (Source: by authors)



UCO/CASE to meet investigation needs

The need for a standard to represent and exchange electronic evidence has been augmented by the rising relevance of the digital evidence in a wide range of circumstances within investigative cases and the requirement upon a standard language to represent a broad range of forensics information and processing results has become an increasing need within the forensics community (Casey, 2011). The standard language is also able to meet another pressing need: processing big volumes of investigative information from different various data sources and finding correlation within them in an accurate and efficient manner. Research activities conducted in this field have been used to develop and propose many languages, but, at the moment, UCO/CASE represents the most suitable standards to representing data and metadata related to evidence for a variety of goals including the availability of a more powerful processing relying on artificial intelligence techniques. This is due to the following reasons:

- it has been developed in the cyber security environment but it also includes lots of essential elements to representing digital forensics information;
- it allows to describe technical, procedural and judicial information as well;
- it has been developed with the extensibility in mind so it is adaptable to the fast-pace development of technology, therefore permits the introduction of new elements to incorporate forensics information not envisaged yet.

It is also worth mentioning that UCO/CASE standard is able to represent the provenance of an evidence. For cyber-investigation purposes, to help establish the authenticity and reliability of information, it is important to capture where it originated or was found, as well as how it was handled after it was found. Provenance includes collection documentation, chain of custody details, audit logs from forensic acquisition tools, and integrity records, which all help to establish the trustworthiness of cyber-investigation information (Casey et al., 2017).

UCO/CASE and other standards

It is worth mentioning that existing standards for exchanging general criminal justice information, including the National Information Exchange Model (NIEM), have not kept pace with the evolution of electronic evidence. Moreover, there are there some similarities between the UCO/CASE standard and ISO/IEC 27037:2012 (Information technology — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence). ISO standards developed for Information Security (2700 series) and Forensic Science (ISO/TC 272 Forensic sciences) provide high level requirements and recommendations for specific practices/processes. Nevertheless, they do not provide a standard for representing and exchanging data. On the contrary UCO/CASE can be used to implement and strengthen certain requirements illustrated in ISO standards to fulfil the objectives of efficiency and quality.

It should be also highlighted that the UCO/CASE standard language that has become popular among many important stakeholders such as Europol, U.S. Department of Defence Cyber Crime Centre - DC3, NFI, Cellebrite, Magnet Forensic and others.

Another UCO/CASE feature worthy of attention is that the standard language has been recently moved under the Linux Foundation: a quite remarkable news that encourages widespread use of this standard in a broad range of cyber-investigation domains to foster interoperability, establish authenticity, and advance analysis.

The UCO/CASE standard has been used in many European projects, among which it is worth mentioning:

- Intelligence Network & Secure Platform for Evidence Correlation and Transfer[6] (INSPECTr, GA 833276). It aims at developing a shared intelligent platform for gathering, analysing and presenting key data to help in the prediction, detection and management of crime in support of many LEA at local, national and international level. The data will originate from the outputs of free and commercial forensic tools integrated by online resource gathering. The data

[6] INSPECTr Project, Intelligence Network & Secure Platform for Evidence Correlation and Transfer. The principal objective of INSPECTr will be to develop a shared intelligent platform and a novel process for gathering, analysing, prioritising and presenting key data to help in the prediction, detection and management of crime in support of multiple agencies at local, national and international level, see https://inspectr-project.eu.

from the tools will be represented in UCO/CASE standard using a set of parsers still under development for Cellebrite UFED_PA, Magnet Forensic AXIOM, MASB XAMN and OXYGEN Forensic Detective.

- Electronic Xchange of e-Evidences with e-CODEX[7] (EXEC II, GA INEA/CEF/ICT/A2019/2065024). Within the project's activities it has been developed a proof of concept (Evidence Exchange Standard Application, EESP Application) being able to create/prepare the Evidence Package (E-Package), safely encrypted, for facilitating its exchange through e-EDES and over e-CODEX and being able to support a standard for the representation of metadata of the Evidence, by using the UCO/CASE language/ontology to propose sensible solutions for the exchange of large file of evidence, based on a decentralised architecture

- Linking EVIDENCE into e-CODEX for EIO and MLA procedures in Europe[8] (EVIDENCE2e-Codex, GA 766468): the project provided a contribution to the exchange of digital evidence within the EIO/MLA legal instruments among Competent/Judicial Authorities in the EU Member States and beyond.

UCO/CASE main aims

One of the most common issues in dealing with the outcome of a forensic acquisition or analysis, concerns the possibility to verify findings extracted/generated by forensics tools. This need is becoming even clearer considering the ever-increasing speed of innovation involving digital devices and the consequences on forensics tools (i.e., operating system, data storage strategies, etc.). The lack of a standardised format for representing the output of forensics tools makes it difficult to compare results produced by different tools with similar features/functionalities. The use of a common standard language would offer many advantages:

- it would allow comparing results produced by different versions of the same forensics tool in order to evaluate the progress in terms of information extraction and interpretation;

- it would speed the automatic search activity avoiding analysing the same information already processed by the previous version of the tool;

- it would foster the data and information exchange between different organisations and different actors involved in the investigation.

At the moment no commercial forensic tool is able to directly generate their output in UCO/CASE standard. The UCO/CASE community is endeavouring to create a middle-layer software (parser) to convert the output from an open format (i.e., XML, CSV etc.) generated by the commercial tool into UCO/CASE standard. At the time of writing this article a parser for both Cellebrite UFED-PA and Magnet Forensic AXIOM is available in UCO/CASE repositories,[9] freely accessible to all Community members and broadly to all forensic community.

An investigation generally involves many different tools and data sources, therefore pulling together information from these various data sources and tools is time consuming, and error prone. Tools that support UCO/CASE can extract and ingest data, along with their context, in a normalized format that can be automatically combined into a unified collection to strengthen correlation and analysis.

Moreover, cyber-investigation information, to be effective, needs to be represented and shared in a form that is usable in any contexts (i.e., digital forensic science, incident response, and situational awareness etc.) and is flexible enough to accommodate evolving requirements.

The main aim of UCO/CASE is the interoperability - to enable the exchange of cyber-investigation information between tools, organizations, and countries. The power of such a standard is that it supports automated

[7] The EXEC II project (Electronic Xchange of e-Evidences) is the follow-up project of the previous EXEC and EVIDENCE2-e-CODEX projects. See https://www.e-codex.eu/EXECII.

[8] The EVIDENCE2e-CODEX project aimed at creating a legally valid instrument to exchange digital evidence related to MLA and EIO procedures over e-CODEX by providing the legal and technical communities with 'ready to use' information on EIO, digital evidence and e-CODEX and a 'true to life' example of how electronic evidence can be shared over e-CODEX in a secure and standardized way to support MLA and EIO cases, see https://evidence2e-codex.eu.

[9] The XML SAX parser for UFED/Cellebrite extracts some digital traces (Cyber items) from XML reports generated by UFED Physical Analyser (version 7.x) and convert them into UCO/CASE as JSON-LD files, see https://github.com/casework/CASE-Implementation-UFED-XML and https://github.com/casework/CASE-Implementation-AXIOM.

normalization, combination correlation, and validation of information, which means less time extracting and combining data, and more time analysing information. The interoperability is ensured not only within a single investigative case that may include many digital devices, but also throughout different investigative cases to find correlation and overcome, for instance, issues like the linkage blindness that is the failure to recognise a pattern that links one crime to another, such as crimes committed by the same perpetrator in different jurisdictions.

UCO/CASE observables

To represent cyber-investigations information, it is necessary to capture details about specific traces and their context such as manufacturers and serial numbers of storage media, network connection details, and names of files stored on a removable USB device with associated date-time stamps and cryptographic hash values. To represent this variety of information, as well as other non-trace cyber-investigation information (identities, locations, tools, etc.), UCO/CASE defines "Objects" and potentially associated "Property Bundles" containing details about the object itself.

Objects encompass any concept pertaining to cyber-investigations including traces such as a mobile device, a file extracted from a device, an email address extracted from a file, a location extracted from EXIF metadata, or non-trace concepts such as a forensic action carried out by an examiner.

UCO/CASE is able to represent certain types of information that cross the cyber domain as core entities. They consist of a set of data and metadata for describing (see Figure 6) the following items:

- Objects and their associated properties, including data sources (mobile devices, storage media, memory) and well-known digital objects such as files and folders, messages, documents, files (images, video, audio etc.) and logs (browser history, events).

- A set of data and metadata for describing all actions (i.e., tasks).

- Actors (e.g.: subjects, victims, authorities, examiners etc.).

- Tools (i.e., digital tools for carrying out different forensics processes).

- Objects relationships (e.g., Contains, Extracted From etc.), in particular for expressing the Chain of Evidence, that is which file (archive, database, etc.) a specific digital trace (Observable in term of the ontologies) has been extracted from.

- Objects inside the digital evidence and their Relationships (e.g., Contained_Within, Extracted_From etc.).

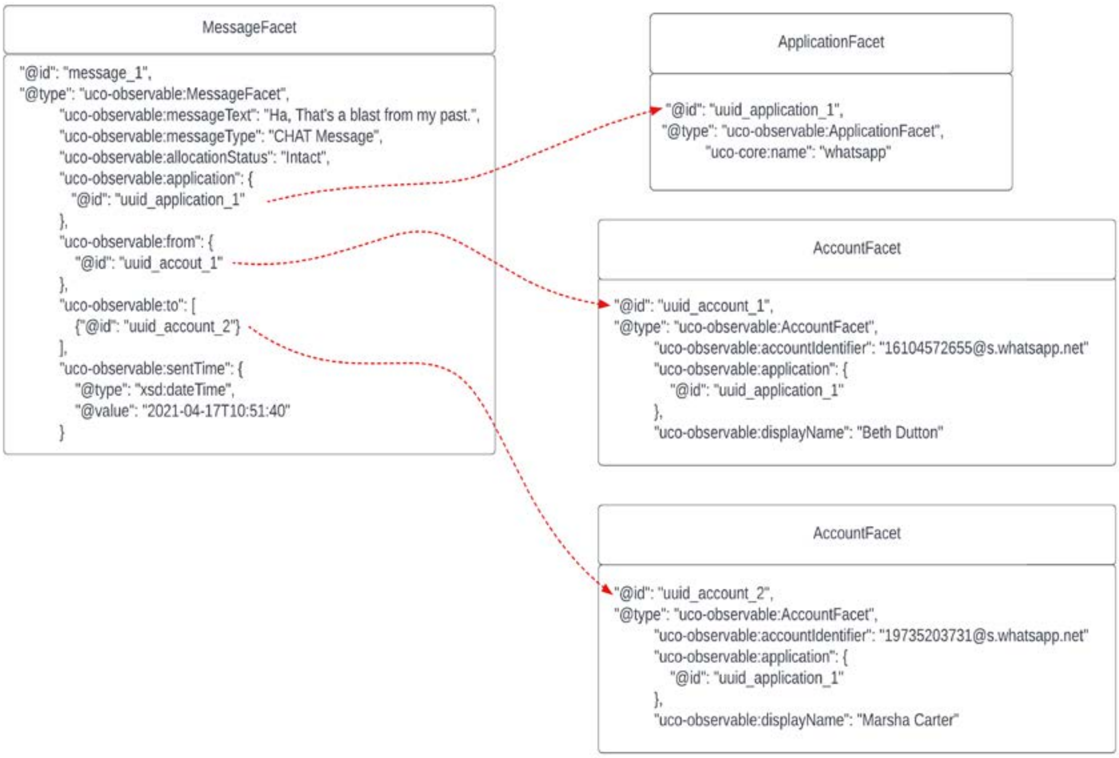
CASE supports any serialisation (default JSON-LD),[10] and can be utilised in any context, including criminal, corporate and intelligence. JSON-LD is 100% valid JSON with some specific JSON structures defined which allow full structural and semantic validation of each object, array and field in the JSON content to a relevant ontological specification for that element.

Each Object is assigned an identifier (@id) that can be used to refer to the Object that cannot be changed that points to another Object, representing a relationship to that other Object. In the proposed approach, such references are represented using an embedded property that specifies the @id of another Object.

Figures 2, 3 and 4 show a CHAT message, represented in CASE and serialised in /JSON-LD format, along with the references to an Application Account and Application Observable Objects that contain the Accounts involved in the communication (Message) and the Application in use.

[10] JSON-LD is a lightweight Linked Data format. It is easy for humans to read and write. It is based on the already successful JSON format and provides a way to help JSON data interoperate at Web-scale. JSON-LD is an ideal data format for programming environments, REST Web services, and unstructured databases such as Apache CouchDB and MongoDB, see https://json-ld.org.

Figure 2: UCO/CASE representation of a Chat Message, along with Account and Application Observables (Source: by authors)



```
"@id": "message_1",
"@type": "uco-observable:MessageFacet",
        "uco-observable:messageText": "Ha, That's a blast from my past.",
        "uco-observable:messageType": "CHAT Message",
        "uco-observable:allocationStatus": "Intact",
        "uco-observable:application": {
          "@id": "uuid_application_1"
        },
        "uco-observable:from": {
          "@id": "uuid_accout_1"
        },
        "uco-observable:to": [
          {"@id": "uuid_account_2"}
        ],
        "uco-observable:sentTime": {
          "@type": "xsd:dateTime",
          "@value": "2021-04-17T10:51:40"
        }
```

```
"@id": "uuid_application_1",
"@type": "uco-observable:ApplicationFacet",
        "uco-core:name": "whatsapp"
```

```
"@id": "uuid_account_1",
"@type": "uco-observable:AccountFacet",
        "uco-observable:accountIdentifier": "16104572655@s.whatsapp.net"
        "uco-observable:application": {
          "@id": "uuid_application_1"
        },
        "uco-observable:displayName": "Beth Dutton"
```

```
"@id": "uuid_account_2",
"@type": "uco-observable:AccountFacet",
        "uco-observable:accountIdentifier": "19735203731@s.whatsapp.net"
        "uco-observable:application": {
          "@id": "uuid_application_1"
        },
        "uco-observable:displayName": "Marsha Carter"
```

Correlations examples

An investigation generally involves many different tools and data sources, creating separate store-room of information. Manually pulling together information from these various data sources and tools is time consuming, and error prone. Tools that support CASE can extract and ingest data, along with their context, in a standard format that can be automatically combined into a unified collection to strengthen correlation and analysis. This offers new opportunities for searching, contextual analysis, pattern recognition, machine learning, and visualisation. Moreover, organisations involved in joint investigations can share information using CASE.

In addition to searching for specific keywords or characteristics within a single case or across multiple cases, having a structured representation of cyber-investigation information allows more sophisticated processing such as data mining, or NLP techniques.

A crucial aspect of information representation and exchange is being able to specify the allowed/authorised conditions for sharing and to enforce exchange policies. At this aim UCO/CASE provides for data markings that CASE can use to support proper handling of shared information: practically any marking mechanism can be employed, including Traffic Light Protocol (TLP) and Information Exchange Policy (IEP).

Overall system

The potentialities of the system, illustrated trough the below examples, and explained in a descriptive manner are underpinned by the following conditions:

- having at disposal a shared criminal cases database either based on a decentralised solution, or a central solution, including both metadata and data of the pieces of evidence. It is almost needless to say that the issues of location and jurisdiction need to be addressed, taking into account the increasingly frequency of cross-border crimes;
- a common format (UCO/CASE ontology) for data homogenisation and data discovery. Once the information is represented in a format not tied to a proprietary system where the possibilities to develop tailored tools are all open to each need that can arise.

Correlation example: ascertain if a file has been exchanged during communication between two suspects, relying on the hash file value

The investigative context is the following: two mobile devices, belonging to two suspects, have been seized and the investigative aim is to discover if a specific file, whose hash value HASH_1 is known, has been exchanged between the two devices (DEVICE_1 and DEVICE_2). The data extracted from the DEVICE_1 is not complete; the sought communication has been deleted by the SUSPECT_1 and the carved data, extracted by using a forensic tool, don't allow the potential evidence to be found because is incomplete.

To bear in mind that the example refers to the same investigative case, but the sought data could also be retrieved throughout different investigative cases, provided that the two requirements described above at the beginning of Section 3.1, are met.

Considering how the Artifact/Digital Traces are expressed in UCO/CASE the retrieval process is the following (see Figure 3):

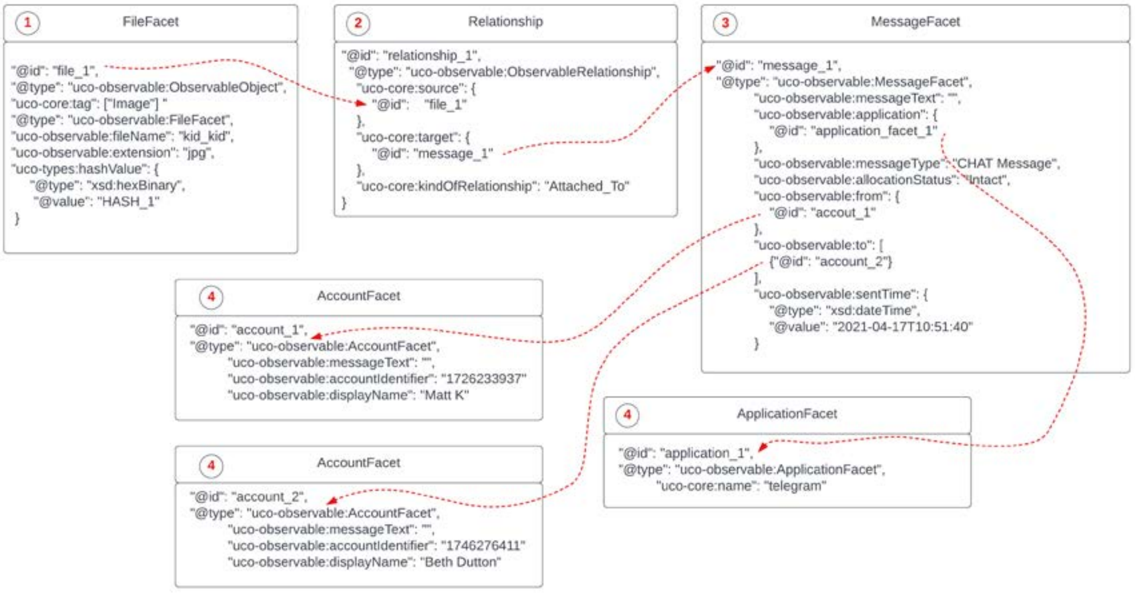
- The HASH_1 is scanned among all the File Observables of the shared database, serialised in JSON format. From this Observable the unique identifier (UUID_FILE) is taken, an identifier that is associated with each Observable.
- The Relationships Observables of kind "Attached_To" are raked to find the value UUID_X in the source property. Once that Observable has been identified its target property, the unique identifier (UUID_MESSAGE), is obtained. That Observable is a Chat/Message, that had the file as an attachment.
- By using the UUID_MESSAGE it is possible to detect the Message Observable and in turn the phone numbers involved in the communication, relying on the properties FROM and TO, always expressed as @id references.
- By retrieving the @id of the two identifiers involved in the Chat/Message Telegram it turns out that the two people who exchange the file with the hash HASH_1 are the ones identified by the following properties:

  PERSON_1: *accountIdentifier*=1726233937 and *displayName*=Matt K

  PERSON_2: *accountIdentifier*=1746276411 and *displayName*=Beth Dutton

that are the suspects under investigation.

Figure 3: Correlation example based on hash file, retrieval process based on UCO/CASE, overview (Source: by authors)



```
"@id": "file_1",
"@type": "uco-observable:ObservableObject",
"uco-core:tag": ["Image"] "
"@type": "uco-observable:FileFacet",
"uco-observable:fileName": "kid_kid",
"uco-observable:extension": "jpg",
"uco-types:hashValue": {
    "@type": "xsd:hexBinary",
     "@value": "HASH_1"
}
```

```
"@id": "relationship_1",
 "@type": "uco-observable:ObservableRelationship",
  "uco-core:source": {
    "@id":    "file_1"
  },
  "uco-core:target": {
    "@id": "message_1"
  },
  "uco-core:kindOfRelationship": "Attached_To"
}
```

```
"@id": "message_1",
"@type": "uco-observable:MessageFacet",
      "uco-observable:messageText": "",
      "uco-observable:application": {
         "@id": "application_facet_1"
      },
      "uco-observable:messageType": "CHAT Message",
      "uco-observable:allocationStatus": "Intact",
      "uco-observable:from": {
         "@id": "accout_1"
      },
      "uco-observable:to": [
         {"@id": "account_2"}
      ],
      "uco-observable:sentTime": {
         "@type": "xsd:dateTime",
         "@value": "2021-04-17T10:51:40"
      }
```

```
"@id": "account_1",
"@type": "uco-observable:AccountFacet",
      "uco-observable:messageText": "",
      "uco-observable:accountIdentifier": "1726233937"
      "uco-observable:displayName": "Matt K"
```

```
"@id": "account_2",
"@type": "uco-observable:AccountFacet",
      "uco-observable:messageText": "",
      "uco-observable:accountIdentifier": "1746276411"
      "uco-observable:displayName": "Beth Dutton"
```

```
"@id": "application_1",
"@type": "uco-observable:ApplicationFacet",
      "uco-core:name": "telegram"
```

Correlation example: to find any kind of outgoing communication originating from a given phone number

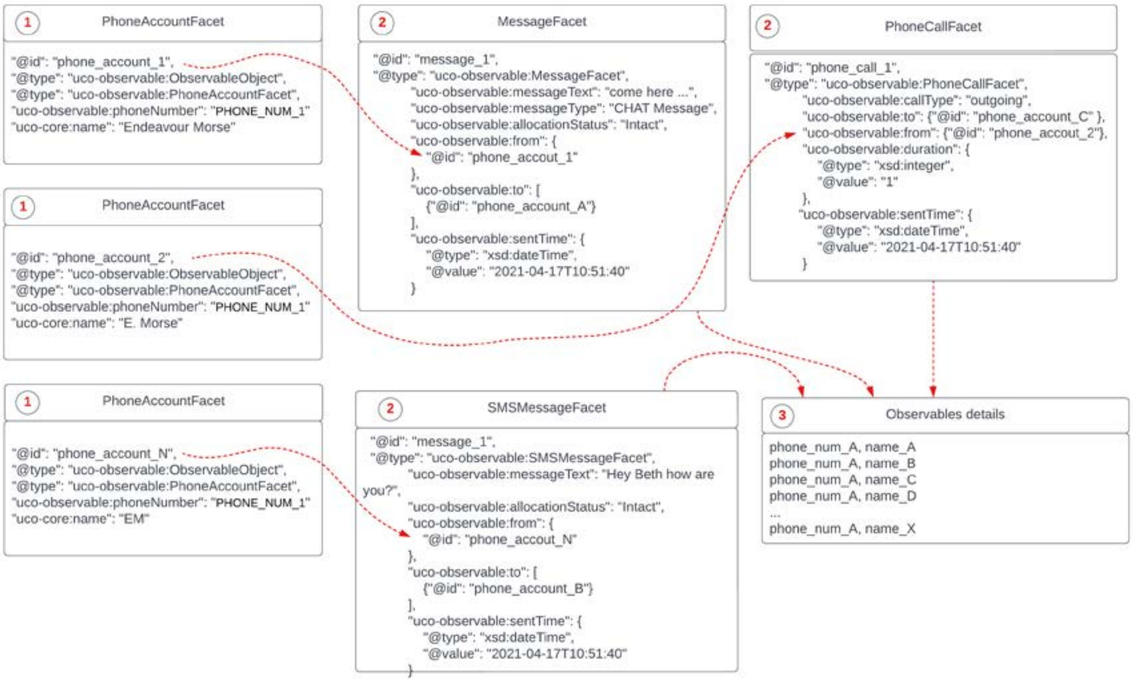
The investigative context is the following: starting from a lot of seized devices, the correlation aims to find any kind of outgoing communication originating from the phone number PHONE_NUM_1. The phone number will be searched in Call, SMS, MMS and Chat Messages selecting only the ones where the PHONE_NUM_1 plays the role of Caller/Sender property.

The retrieval process is the following (see Figure 4):

1. The phone number PHONE_NUM_1 is dug among all the Phone Account Observables of the shared criminal cases database. From the list of the retrieved Observables, all the unique identifiers (UUID_PHONE_NUM_1, UUID_PHONE_NUM2, …, UUID_PHONE_NUM_N) are taken.

2. All the UUID_PHONE_NUM_X identifiers selected in the previous step are searched in the FROM property of all Observables of kind PhoneCallFacet, SMSMessageFacet, MMSMessageFacet and MessageFacet (Chat).

3. The details of each retrieved Observable are presented below (Figure 4, frame labelled with number 3) along with the data related to the people involved in each communication item spotted.



```
"@id": "phone_account_1",
"@type": "uco-observable:ObservableObject",
"@type": "uco-observable:PhoneAccountFacet",
"uco-observable:phoneNumber": "PHONE_NUM_1"
"uco-core:name": "Endeavour Morse"
```

```
"@id": "phone_account_2",
"@type": "uco-observable:ObservableObject",
"@type": "uco-observable:PhoneAccountFacet",
"uco-observable:phoneNumber": "PHONE_NUM_1"
"uco-core:name": "E. Morse"
```

```
"@id": "phone_account_N",
"@type": "uco-observable:ObservableObject",
"@type": "uco-observable:PhoneAccountFacet",
"uco-observable:phoneNumber": "PHONE_NUM_1"
"uco-core:name": "EM"
```

```
"@id": "message_1",
"@type": "uco-observable:MessageFacet",
        "uco-observable:messageText": "come here ...",
        "uco-observable:messageType": "CHAT Message",
        "uco-observable:allocationStatus": "Intact",
        "uco-observable:from": {
           "@id": "phone_accout_1"
        },
        "uco-observable:to": [
           {"@id": "phone_account_A"}
        ],
        "uco-observable:sentTime": {
           "@type": "xsd:dateTime",
           "@value": "2021-04-17T10:51:40"
        }
```

```
"@id": "message_1",
"@type": "uco-observable:SMSMessageFacet",
        "uco-observable:messageText": "Hey Beth how are
you?",
        "uco-observable:allocationStatus": "Intact",
        "uco-observable:from": {
           "@id": "phone_accout_N"
        },
        "uco-observable:to": [
           {"@id": "phone_account_B"}
        ],
        "uco-observable:sentTime": {
           "@type": "xsd:dateTime",
           "@value": "2021-04-17T10:51:40"
        }
```

```
"@id": "phone_call_1",
"@type": "uco-observable:PhoneCallFacet",
        "uco-observable:callType": "outgoing",
        "uco-observable:to": {"@id": "phone_account_C" },
        "uco-observable:from": {"@id": "phone_accout_2"},
        "uco-observable:duration": {
           "@type": "xsd:integer",
           "@value": "1"
        },
       "uco-observable:sentTime": {
           "@type": "xsd:dateTime",
           "@value": "2021-04-17T10:51:40"
        }
```

```
phone_num_A, name_A
phone_num_A, name_B
phone_num_A, name_C
phone_num_A, name_D
...
phone_num_A, name_X
```

Figure 4: Correlation example based on phone number, retrieval process based on UCO/CASE, overview (Source: by authors)

Conclusions

To perform digital investigations effectively, there is a pressing need to harmonise how information relevant to cyber-investigations is represented and exchanged. The CASE specification language and underlying UCO support information standardization and interoperability for tools and organizations dealing with cyber-investigations. In addition to sharing cyber-investigation information on a specific case, sharing traces or patterns of particular activities in a standardized format can help others find similar traces and patterns in new cases, both in national and international judicial cooperation.

The standard is one of the most effective formalisms to represent data and metadata of an evidence and it appears particularly versatile to both a central criminal cases database and distributed databases incorporated into national systems. Moreover, it encompasses relevant aspects: it allows to indicate the grade of information sharing, preserving their privacy, and also to strengthen the admissibility of a potential evidence based on the detailed description of its Chain of Document and Chain of Evidence.

This article has illustrated a draft retrieval system based on the open and free UCO/CASE language standard highlighting the significant advantage provided by the standard, especially considering alternative proprietary systems that are closed and hinder the interoperability among different systems and various organisations. A significant example of the use of the standard has been implementing within the INSPECTr project, by using ElasticSearch[11] as storage of the data represented in UCO/CASE and Kibana as a user interface to visualize the evidence data and navigate the Elastic Stack.

Having at its disposal a standard representation would streamline the investigations and improve the effectiveness of the search for correlation within different cases both in national and cross-border scenarios. Such a system would also be beneficial for joint investigation team (JIT) scene where it is of utter importance to efficiently carry out criminal investigations in one or more of the involved States, achieving one of the main impacts of the use of the standard: to dedicate less time extracting and combining data and more time analysing info to find links and patterns.

References

- Casey, E. (2011) "Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet". 3rd edition, Elsevier, Amsterdam.
- Casey, E., Barnum, S., Griffith, R., Snyder, J., van Beek, H. & Nelson, A. (2017) Advancing coordinated cyber- investigations and tool interoperability using a community developed specification language, *Digital Investigation,* vol. 22, pp.14–45.
- Casey, E., Barnum, S., Griffith, R., Snyder, J., van Beek, H. & Nelson, A. (2018) The Evolution of Expressing and Exchanging Cyber-Investigation Information in a Standardized Form. In: Biasiotti, M., Mifsud Bonnici, J., Cannataci, J., Turchi, F. (eds) Handling and Exchanging Electronic Evidence Across Europe. Law, Governance and Technology Series, vol 39. Springer, pp. 43-58.
- Casey, E., Biasiotti. M.A., Turchi, F. (2017) "Using Standardization and Ontology to Enhance Data Protection and Intelligent Analysis of Electronic Evidence", ICAIL 17, DESI VII Workshop on "Using Advanced Data Analysis in eDiscovery & Related Disciplines to Identify and Protect Sensitive Information in Large Collections", Strand Campus, King's College London, UK, pp. 10.
  Available at: http://users.umiacs.umd.edu/~oard/desi7/papers/EC.pdf
- Debski T., Heimans D., Verheggen H., Bille W. & Kamarás E. (2020) "Cross-border Digital Criminal Justice, Final Report", by Deloitte, Luxembourg: Publication Office of the European Union.
  Available at: https://www.legalbusinessworld.com/post/report-xbordercriminaljustice

[11] Elasticsearch is a distributed, JSON-based search and analytics engine, https://www.elastic.co.