Forensic Science International: Digital Investigation 33 (2020) 200902

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Safeguarding the evidential value of forensic cryptocurrency investigations☆,☆☆

Michael Fröwis [a,*], Thilo Gottschalk [b], Bernhard Haslhofer [c], Christian Rückert [d], Paulina Pesch [b]

[a] University of Innsbruck, Technikerstr. 21a, 6020, Innsbruck, Austria
[b] Karlsruhe Institute of Technology, Vincenz-Prießnitz-Str. 3, 76131, Karlsruhe, Germany
[c] AIT Austrian Institute of Technology, Giefinggasse 4, 1210, Vienna, Austria
[d] Friedrich-Alexander University of Erlangen-Nuremberg, Schillerstraße 1, 91054, Erlangen, Germany

ARTICLE INFO

Article history:
Received 3 July 2019
Received in revised form
19 December 2019
Accepted 20 December 2019
Available online 6 March 2020

Keywords:
Digital forensics
Cryptocurrencies
Digital evidence
Safeguards
Legal

ABSTRACT

Analyzing cryptocurrency payment flows has become a key forensic method in law enforcement and is nowadays used to investigate a wide spectrum of criminal activities. However, despite its widespread adoption, the evidential value of obtained findings in court is still largely unclear. In this paper, we focus on the key ingredients of modern cryptocurrency analytics techniques, which are clustering heuristics and attribution tags. We identify internationally accepted standards and rules for substantiating suspicions and providing evidence in court and project them onto current cryptocurrency forensics practices. By providing an empirical analysis of CoinJoin transactions, we illustrate possible sources of misinterpretation in algorithmic clustering heuristics. Eventually, we derive a set of legal key requirements and translate them into a technical data sharing framework that fosters compliance with existing legal and technical standards in the realm of cryptocurrency forensics. Integrating the proposed framework in modern cryptocurrency analytics tools could allow more efficient and effective investigations, while safeguarding the evidential value of the analysis and the fundamental rights of affected persons.

1. Introduction

Tracking and tracing payment-flows in Cryptocurrencies by analyzing transactions in the underlying, publicly-available Blockchain, has become a key forensic method in law enforcement. It is used to investigate a wide spectrum of criminal activities relying on the pseudo-anonymous nature of cryptocurrencies, ranging from the purchase of illicit goods and services on Darknet markets (Soska and Christin, 2015), over ransomware attacks (Huang et al., 2018; Paquet-Clouston et al., 2018), to extortion and money laundering (FATF, 2015). A typical forensic investigation starts from one or more suspect addresses and traces monetary flows up to some known exit point, which is typically an exchange or a wallet provider service, where Cryptocurrencies are converted back into fiat currencies.

Cryptocurrency investigations are nowadays supported by a number of commercial (e.g. Chainalysis, Elliptic, etc.) and non-commercial analysis tools (e.g. BlockSci; Kalodner, Goldfeder, Chator, Möser and Narayanan (2017) that exploit the openness of the Cryptocurrency transaction ledger also known as *Blockchain*. They build on a long history of research that has shown that pseudonymous addresses do not provide sufficient anonymity, neither in Bitcoin (Meiklejohn et al., 2013; Androulaki et al., 2013; Möser, 2013; Monaco, 2015) nor in post-Bitcoin currencies, with stronger privacy-enhancing techniques, such as ZCash (Quesnelle, 2018; Kappos et al., 2018) or Monero (Miller et al., 2017; Kumar et al., 2017), which has shown to be traceable until early 2017.

Investigation tools mainly rely on two complementary techniques: *Address clustering heuristics* or clustering heuristics for short, which are used to group multiple addresses into maximal subsets (*clusters*) that can be likely assigned to the same real-world actor, and *attribution tags*, which are any form of context

☆ This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 740558.

☆☆ Author names in alphabetical order.

* Corresponding author.
E-mail addresses: michael.froewis@uibk.ac.at (M. Fröwis), thilo.gottschalk@kit.edu (T. Gottschalk), bernhard.haslhofer@ait.ac.at (B. Haslhofer), christian.rueckert@fau.de (C. Rückert), paulina.pesch@kit.edu (P. Pesch).

https://doi.org/10.1016/j.fsidi.2019.200902

information that can be attributed to an address, transaction or cluster, such as the name of an exchange hosting the associated wallet or some other *personally identifiable information* (PII) of the account holder. The strength lies in the combination of these techniques: a tag attributed to a single address belonging to a larger cluster can easily de-anonymize hundreds of thousands Cryptocurrency addresses (c.f. Kumar et al. (2017)).

However, despite the promising benefits of before mentioned Cryptocurrency analytics techniques in criminal investigations, the evidential value of those techniques as well as implications for digital forensics remain largely unclear: first, certain types of transactions (e.g. CoinJoins, Möser and Böhme (2016)) could distort clustering results, unifying entities that have no association in the real-world and can lead to the formation of so called super-clusters (Harrigan and Fretter, 2016). Second, false, unreliable, or intentionally misplaced attribution tags could associate unrelated actors with a given cluster and lead to suspicions against innocent people or even to false convictions.

Law enforcement agencies (LEAs) increasingly recognize the value of information sharing to maximize investigative resources and avoid duplicate efforts (Interpol, 2018). This also applies to sharing attribution tags and address clusters in cryptocurrency forensics. Both detection methods become more effective by sharing obtained information. In that regard, the lack of a standardized ontology and analytical approaches additionally amplifies the aforementioned risks of forensic Cryptocurrency analysis in particular if attribution tags are shared in a framework that does not safeguard the evidential value of the shared content (Casey and Back, 2015; Garfinkel, 2010).

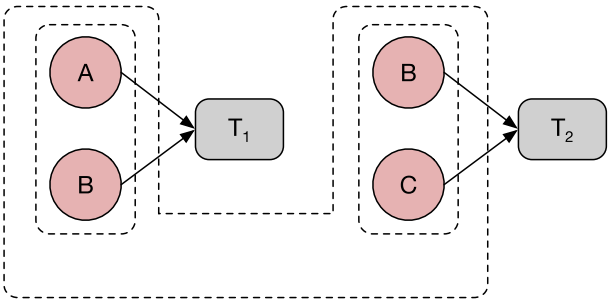
In this paper, we propose measures for safeguarding the evidential value of forensic Cryptocurrency investigation results. After introducing the necessary background in Section 2, we make two major contributions that can be summarized as follows:

- First, we systematically investigate internationally accepted legal standards and rules for providing court-proof evidence and derive key requirements for forensic Cryptocurrency investigations and discuss possible risks of clustering approaches based on an exemplary empirical analysis of CoinJoin transactions in the 100 largest Bitcoin clusters in Section 3.
- Second, we translate those requirements into a data sharing framework for law enforcement agencies that provides safeguards to maintain the evidential value of forensic Cryptocurrency investigations by ensuring compliance with existing regulations in Section 4.

To the best of our knowledge, this paper is the first to tackle Cryptocurrency forensics and analytics from a combined legal and technical perspective. We believe that it can therefore simultaneously serve as a blueprint for law enforcement investigators, prosecutors, or Cryptocurrency analytics tool providers who aim to comply with existing regulations.

2. Background and related work

In this section, we briefly introduce central notions and concepts used throughout this paper. While we do not attempt to give a complete introduction to the underlying technology of Cryptocurrencies, we direct the reader to existing literature, such as Nakamoto (2008); Bonneau et al. (2015); Tschorsch and Scheuermann (2016); Judmayer et al. (2017). In the following, we use Bitcoin as a running example but most findings can be easily translated to any UTXO (Unspent Transaction Output) based Cryptocurrency.



a **Multi-Input Clustering Heuristics**: Addresses A and B are inputs of transaction T_1 and must therefore be controlled by the owner of the corresponding private keys. The same holds for addresses B and C of T_2. Since address B occurs in the set of inputs of T_1 and T_2 one can infer that addresses A, B, C, and D are controlled by the same actor.

Fig. 1a. Multi-Input Clustering Heuristics: Addresses A and B are inputs of transaction T_1 and must therefore be controlled by the owner of the corresponding private keys. The same holds for addresses B and C of T_2. Since address B occurs in the set of inputs of T_1 and T_2 one can infer that addresses A, B, C, and D are controlled by the same actor.

2.1. Address clustering heuristics

There are several address clustering heuristics in use today, the safest, most effective and most studied one being the multi-input heuristic (MIH) (Meiklejohn et al., 2013; Nick, 2015). Its underlying intuition, which is illustrated in Fig. 1a, is that if two addresses (i.e. A and B) are used as inputs in the same transaction while one of these addresses along with another address (i.e. B and C) are used as inputs in another transaction, then the three addresses (A, B and C) must somehow be controlled by the same actor, who conducted both transactions and therefore possesses the private keys corresponding to all three addresses.

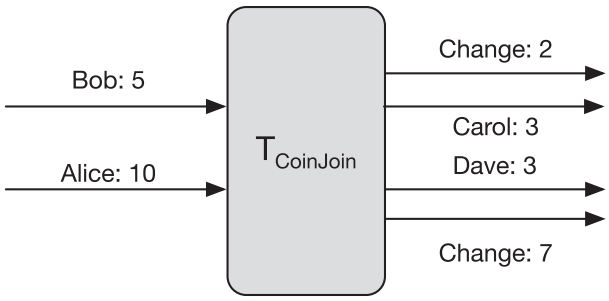
The underlying assumption of the multiple-input heuristics holds for most Bitcoin transactions although there are known obfuscation mechanisms that can violate this assumption: for example, transactions that are generated by a mixing scheme called CoinJoin (Möser and Böhme, 2016) where n parties produce a special joint transaction. Fig. 1b sketches the structure of a CoinJoin transaction. This scheme is used to conceal the relationships between inputs and outputs and in the end who of the n parties transacted with whom. For multi-input address clustering these transaction schemes pose a problem because the clustering algorithm would combine all n input addresses and their respective clusters into one entity. Meaning we merged n potentially independent parties into one (c.f. Section 3.3).

All clustering heuristics have in common that they rely on certain behavioral patterns to group addresses that likely belong to the same owner. Although different clustering heuristics depend on different behavioral assumptions,[1] they all suffer from similar problems.

To systematically evaluate the effectiveness of those heuristics, ground truth data[2] is needed, but not available and generally hard to obtain (Nick, 2015). Furthermore, the reliability of the heuristics to some extend depend on user behavior which can change and leaves room for obfuscation (nopara73, 2017). To depict the related

[1] For example: Private keys are not shared, thus nobody can create transactions with inputs not belonging to her (*multi-input*); Or nobody would include unnecessary inputs, thus we can infer change outputs (*optimal change*).

[2] Disjoint sets of addresses known to belong together (*clusters*).



b **CoinJoin Transaction** (based on Möser and Böhme (2016)): several individual payments from multiple parties are combined into a single transaction.

Fig. 1b. CoinJoin Transaction (based on Möser and Böhme (2016)): several individual payments from multiple parties are combined into a single transaction.

risks from a legal perspective we only consider the multi-input heuristic within this paper. But out of given reasons most of the results and general problems also apply to other clustering heuristics. For a taxonomy of different clustering heuristics we refer to (Meiklejohn et al., 2013; Androulaki et al., 2013).

A number of studies attempted to quantify the effectiveness of clustering techniques (Nick, 2015): measured the accuracy of different clustering algorithms using a ground-truth dataset consisting of 37 585 user wallets, which was obtained via a vulnerability in the BitcoinJ light client implementation. The results showed that on average 69.34% of the addresses could be linked using only the multi-input heuristics. Harrigan and Fretter (2016) studied reasons for the effectiveness of multi-input address clustering and came to the conclusion that address reuse and avoidable merging are the main drivers. They also measured the growth patterns of clusters and found that merges of two large clusters are rare in general and could be an indicator of wrongfully merged clusters.

The reliability of clustering results is of uttermost importance for forensic investigations. Wrong clustering results can lead to missed- or even false convictions.

2.2. Attribution tags

Tagging is a collaborative process in which a user adds (mostly textual) labels or *tags* to shared content. It does not rely on static, predefined taxonomic structures but on dynamic, community-driven linguistic terms and conceptions (Golder and Huberman, 2006). Tagging became popular with the launch of sites like Delicious and Flickr around 2005 and is now a standard feature that can be found in many social media sites. When applied in the context of Cryptocurrencies, as shown in Fig. 2, a tag could for example attribute a given Bitcoin address to some real-world actor (e.g. Internet Archive).

Despite their wide-spread adoption, tagging systems still face a number of problems: a tag can be ambiguous and have many related meanings (polysemy), multiple tags can have the same meaning (synonymy), or the semantics of a tag might range from very specific to very general because people describe resources along a continuum of specificity (Golder and Huberman, 2006). If, in the context of Cryptocurrencies, a user attributes an address with the tag "ransomware" it is not entirely clear whether this tag means that the tagged address belongs to a victim or the originator of a cybercrime attack or that the tag is just somehow related to this topic (c.f. Article 6 EU-LED). Further contextual information, such as a more detailed description, can only be determined after reconciling label-based tags with their authors.

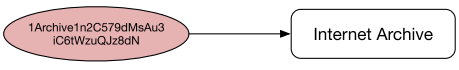


Fig. 2. Attribution Tag: an attribution tag attributes contextual information to a Cryptocurrency address.

Semantic ambiguity of tags can also be exploited to find irregularities in address clustering: Ermilov et al. Ermilov et al. (2017) devised a new clustering algorithm that uses tagging data as additional clustering criteria. They categorize tags on addresses and create so called negative pairs, if such a negative or conflicting pair would be introduced in a cluster in course of merging two clusters, cluster formation is aborted (e.g. a cluster is unlikely to be an exchange and a darknet market at the same time).

As for clustering the reliability of tag data is crucial for investigations. The reliability of tags mainly depends on the origin of the tag as well as its processing history (c.f. 3.3).

2.3. Provenance and digital evidence

In large scale data sharing efforts, data is continuously added, modified, or deleted by users having different backgrounds, technical skill, and intentions. Data integrated from several sources into another, possibly diverging context is therefore never fully clean, certain, and only as trustworthy as its source. In order to assess the *quality*, *uncertainty*, and *authority* of data, users must therefore know the sources and applied data generation and processing routines.

Provenance refers to sources of information that describe the entities and processes involved in producing, delivering, or otherwise influencing a data artifact. It provides a critical foundation for assessing quality and authenticity, as well as enabling trust and allowing reproducibility. Provenance is crucial in deciding whether information is to be trusted, how it should be integrated with other diverse information sources and how to credit originators when reusing it (Gil et al., 2010).

The importance of provenance has been recognized in a number of application areas: in the field of databases and data warehousing, provenance information represents the *lineage of data*, which is a historical record of the data and its origins. It can be used for tracing root causes of errors in data analytics processes, data-dependency analysis, as well as auditing or compliance analysis (c.f. Cui and Widom (2003); Karvounarakis (2009)). Data provenance is also discussed in the field of scientific data sharing and processing to support data protection, data ethics (Hadziselimovic et al., 2017) and tracking the lineage (origin and subsequent processing history) of scientific data sets (Bose and Frew, 2005). For a more general overview on provenance management for computational tasks in various domains we refer to the survey by Freire et al. (2008).

In forensic investigations, provenance information is recorded in order to provide sufficient information to evaluate the authenticity, integrity and reliability of evidence and thus if it can be used in court. Traditionally provenance information was mainly based on filling paper- or electronic-forms with the name of the investigators, a description of the evidence under examination and some kind of hash code (Giova, 2011). Modern forensic software can automate much of the manual work needed to produce so called audit trails and provide stronger guarantees by using existing digital infrastructure and techniques such as user management, digital signatures, or even blockchains (Stoffers, 2017) for creating and securing provenance information. Over the last two decades a number of studies investigated forensic procedures and process models. A review by Pollitt has shown that there is no consistent, generic model applicable to criminal investigations (Pollitt, 2007).

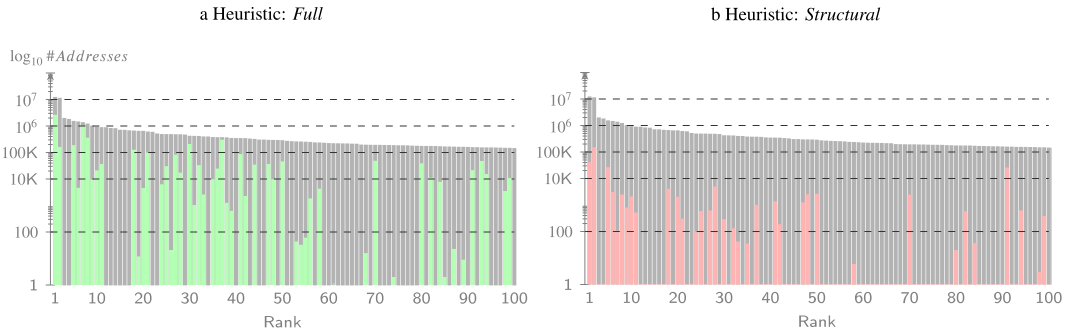


Fig. 3. #Addresses found in the 100 largest Bitcoin clusters (naïve multi-input heuristic) until April 30th, 2018. Compared to #Addresses in the cluster potentially involved in a CoinJoin transaction, note the log scale. a)*Full* b) *Structural.*

More recent research by Cosic and Miroslav (Cosic and Baca, 2010) proposes a conceptual digital evidence management framework (DEMF) to improve the chain of custody of digital evidence in all investigation phases. They suggest using hash codes for fingerprinting of evidence (*what*), hash similarity to control changes (*how*), biometric identification and authentication for digital signing (*who*), automatic and trusted time stamping (*when*), as well as GPS and RFID for geo-location (*where*). These measures can be implemented through a database which records activities done by first responders, forensic investigators, court expert witness, law enforcement personnel, and police officers.

Reliable provenance tracking of digital evidence (e.g. clustering data, attribution tags) is of high importance in digital forensics and cryptocurrency investigations. Especially the source of data and the analysis methods applied to data are relevant to determine the reliability of evidence provided in court.

3. The legal perspective

Following the technical fundamentals, we now assess the legal conditions that govern the forensic analysis of cryptocurrencies. To date, there are no internationally valid rules for the treatment of digital evidence and the legal rules for substantiating suspicions and providing evidence in court vary greatly from one country to another. Nevertheless, a number of general rules can be developed which should meet the evidence standards of most countries.

The forensic analysis of cryptocurrencies in law enforcement is governed by two main interests. First, the tools must produce relevant, court-admissible evidence or at least reasonable suspicion as a basis for further investigations. Evidence is relevant if it makes the matter that requires proof more or less probable (c.f. UK v. Kilbourne, 1973; AC 729). Second, the outcomes must comply with general legal standards and in particular preserve the fair trial rights of the accused. These rights are mainly enshrined in the *law of substantiation of evidence and suspicion*. To identify the essence of these requirements we hence analysed a) written law for evidence in criminal procedures (e.g. US-FRE; GER-CCP; NL-CCP), b) case law of several supreme courts of several countries (e.g. USA and Germany) (Eschelbach, 2018; Miebach, 2016, overview over German supreme court decisions) ((Frye v. US; Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993; Lorraine v. Markel, 2007), c) recommendations of international groups of forensic experts (e.g. the SWGDE), d) scientific publications on the evidential value of IT forensic investigations (Casey, 2011; Heinson, 2015; Mohay et al., 2003; Maras, 2015; Casey et al., 2017), e) the often overlapping legal requirements laid down in data protection law (European Commission, Parliament of the European Union, 2016b, Article 29 Working Party, 0000; CA State Senate, CA State Assembly, 2018).

Based on the analysis we identify seven key requirements for the forensic processing of cryptocurrencies. Namely, Lawfulness (Section 3.1), Authenticity (Section 3.2), Reliability (Section 3.3), Qualification (Section 3.4), Verifiability (Section 3.5), Chain of evidence (Section 3.6) and the right to inspect the records (Section 3.7).

3.1. Lawfulness of data processing

Data processing must be in compliance with the legal framework it takes place in (Casey, 2011, pp. 56 et sqq.). For criminal investigations the respective criminal procedural law and data protection law constitute the rules for the processing and in particular require a sufficient legal basis and compliance with data protection principles. The required quality of the legal basis depends on the respective level of protection as well as the scope of the processing.

Clustering and attribution techniques are specifically used to identify natural persons (i.e. suspects). The used data consequently relates to identifiable natural persons and is hence deemed *personal data* (c.f. CFR; ECHR; EU-GDPR; EU-LED) that is protected on international (e.g. CFR; ECHR; EU-GDPR; EU-LED) and national (e.g. IN-PDPB, 2018; UK-DPA, 2018; GER-BDSG, 2018; US-FISMA, 2014) level as well as in subject specific laws and regulations (e.g. NIST-800-171; US-GLBA, 1999). As the scope of protection differs, the data protection principles in this paper are derived from the European data protection framework, namely the General Data Protection Regulation (EU-GDPR) and the so called Law Enforcement Directive (EU-LED). Although the GDPR is not directly applicable in the area of law enforcement (EU-GDPR, Article 2 (2) lit. d), the general concepts are transferable and reappear in the LED as well as in national legislation. In addition, where forensic analyses are outsourced and conducted by (private) third parties (e.g. van Baar et al. (2014)), the GDPR can remain applicable. The European approach arguably acts as a role model internationally (c.f. IN-PDPB, 2018; US-CCPA; Albrecht, 2016) and also contains relatively high standards which help to fulfill the evidence requirements described in Sections 3.2 to 3.6.

Processing can be roughly split into (1) the *collection* of data and (2) the *subsequent processing*. Transaction data is gathered from a publicly available Blockchain, while attribution data can derive from public and non-public sources. Legal implications of processing publicly available data for law enforcement purposes are still subject to ongoing discussion, however, both steps can arguably be based on general clauses to a certain extent. In addition, data obtained through existing law enforcement communication channels such as SIENA, the Schengen Information System (SIS) or Interpol's I24/7 and I-Link can usually be seen as lawful due to their legal frameworks and the safeguards included in these systems (Interpol RPD, Art. 34, 37, 63), (Europol Regulation, Art. 17) (EU-SFD, 2016).

The processing (collection/analysis) of data should be limited to the extent necessary for the specific investigation (*purpose limitation*). The data volume (*data minimization*) and retention dates (*storage limitation*) of data have to be limited to what is necessary for the specific purpose and the data *integrity* has to be ensured. The *principle of transparency* requires the investigator/prosecutor to explain the processing to a certain extent either to the data subject or the data protection authorities (e.g. EU-LED, Recital 38). In practice, most legislations limit the transparency requirement to ensure effective law enforcement (c.f. 3.7). While *fully automated decision making* is generally prohibited, decisions such as ordering further investigative measures, can still be based on results of automated analytics techniques (e.g. clustering of Cryptocurrency addresses) if the decisions are not merely formalistic. That means that all relevant aspects of the individual case carefully have to be taken into account (c.f. EU-LED, Art. 11, Recital 38 (Rich, 2016, "totality-of-the-circumstances analysis"),) by a natural person. In order to mitigate the risk of misinterpretation of probabilistic results as facts and enable the decision-maker to assess its significance, the latter must be well-trained and the software they use as clear and differentiated as possible (c.f. Section 3.4).

Moreover, the data has to be *accurate* (c.f. EU-LED, Art. 4 (1) lit. c; Europol Regulation, Art. 28 (1) lit. d; Recommendation No. R (87) 15, 1987). Since only facts can be inaccurate, data protection law does not prohibit the processing of data based on estimations (heuristics) or probabilistic measures. However, the principle of accuracy requires the clear distinction of facts and probabilistic or estimated results such as address clusters (EU-LED, Art. 7; Europol Regulation, Art. 29; Recommendation No. R (87) 15, 1987) and inaccurate results need to be rectified. To assess the nature and reliability of the data it is hence necessary to have sufficient meta-information (see 4.1 and 4.4) available. In addition, the used clustering heuristics have to be reviewed steadily since changes in Cryptocurrency network protocols or user behavior can significantly limit the reliability of results or even render a heuristic obsolete.

The resulting risk of false positives raises the question how to deal with the finding that an address has been erroneously attributed to a cluster or a tag contains erroneous information. Data protection law usually requires the data controller to rectify, or in some cases, erase false data. c. f. (EU-LED, Art. 16 (3)). In both cases, measures to avoid automatic reproduction of erroneous data have to be implemented (e.g. mark the address/cluster as erroneous; exclude the address from clustering or tagging). If data has been shared, receivers of the false or outdated data must be informed.

3.2. Authenticity and integrity (chain of custody)

The authenticity and integrity of data correlates with the probative value of information in criminal proceedings (Casey, 2011, p. 59 et sqq.) but is also a data protection requirement (e.g. Art. 29 EU-LED; c. f. 3.1). Authenticity must be ensured in all forensic steps and procedures that involve the processing of electronically stored information (Lorraine v. Markel, 2007; Colorado v. Huehn, 2002; US-FRE). This makes comprehensive and precise documentation of data sources, tools, and applied techniques necessary (see Section 3.5). The preferable way of presentation is subject to an ongoing discussion (e.g. Neumann et al., 2016), however, the analysis procedure, outcomes and limitations have to be explainable in a comprehensible manner that allows an assessment of the evidential value in trials (Neumann et al., 2016; Sjerps and Berger, 2012; Lund and Iyer, 2017, p. 22). It must be ensured that data has not been altered, which can be achieved by technical means e.g. through the use of digital signatures. If data is changed, it must be clear how the alteration exactly changed the data.

Verifying authenticity of data requires that the original data/source is known and that all changes were tracked (Lorraine v. Markel, 2007; at 546). In this regard, it might be helpful to attach 'certainty-values' to data as proposed by Casey (2011, p. 70). Similarly, data protection law particularly requires the data processor to be able to demonstrate compliance with data protection law to ensure accountability ((c.f. EU-LED, Art. 4, 19, 25, Recital 57, WP29 - LED, 2017, p. 25; WP29 - Accountability, 2010; US-FISMA, 2014; NIST-800-53). The same requirements apply for data from external sources. A simple strategy for proving authenticity and integrity of data is to guarantee reproducibility of results by applying the same technique on the same data source. This means at least the name of the source, time of access and liability of the source must be recorded as provenance information. Since availability of online data often is not guaranteed, additional measures should be implemented to enable the user to prove authenticity of the data (Heinson, 2015, p. 147) (e.g. by archiving content in the WARC File Format (ISO 28500); archive.org (Kelly and Weigle, 2012).

Cryptocurrency forensic tools that operate on-top of a specific Cryptocurrency benefit from the underlying concepts of Blockchain technology which already provide strong authenticity and integrity guarantees (e.g. signed transactions) and integrity proofs (proof of work, hash linked list). Outcomes of the analysis are hence easily reproducible, given the *currency code* (e.g. BTC, ETH), the most recent *block hash* as well as the *analysis method* are recorded (including software version, git commit or other identifiers) within the chain of custody, given the analysis method is deterministic.

Ensuring the authenticity and integrity of clustering results is more challenging: tools implementing such techniques create new tool-specific data points that group known Cryptocurrency addresses into a set of clusters, which are usually identified by some tool-specific identifier. Since clustering algorithms run periodically over an evolving transaction ledger, generated clusters are volatile meaning that a cluster generated at a certain Cryptocurrency state is not necessarily equal to a cluster generated at a later Cryptocurrency state. In order to provide authenticity and integrity of clustering results, tools must implement deterministic cluster identifiers that remain stable when a cluster refers to the same set of addresses over several runs and change when the underlying set of addresses changes. This could be achieved by computing a hash (*cluster hash*) over the lexically sorted set of addresses within each cluster and also allow interoperability and comparability of clustering results across tools.

Attribution tag authenticity and integrity can be strengthened by relating a tag to its source, its creator (e.g. using digital signatures) and generation procedures.

3.3. Reliability

Reliability correlates with the weight of evidence in a criminal procedure (Neumann et al., 2016; Good and Irving, 1950) and is hence of relevance for the (free) consideration of evidence in trials and in pre-trial stages to establish necessary degrees of suspicion (e.g. US: *'reasonable suspicion'* (Terry v. Ohio, 1968), GER: *'sufficient factual indication'* (GER-CCP,§ 152 (2)), UK: *'belief that a crime may have been committed'* (UK-CCP - Code of Practice, Section 2.1), NL: *'reasonable suspicion'* (Section 27 § 1; Singelnstein, 2018). Consequently, digital/electronic evidence should be based on precise and scientifically verified methods irrespective of whether the digital investigation is conducted "in-house" or if it is outsourced (e.g. DFaaS: van Baar et al., 2014).

In addition, the methods must be applied properly when collecting and processing data (UK Code for Crown Prosecutors; US-FRE). Reliability hence has to be evaluated for the scientific methods as well as the correct use of them. A lack of scientific

verifiability reduces the evidential value of the found evidence or, in the worst case results in the complete loss of it. Proof can be brought forward by describing results of testing the process, the logic of the process, or by having an expert testimony (US-FRE; Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993; Kumho Tire Co. v. Carmichael, 1999) (Heinson, 2015). A combination of these approaches generally increases the reliability of evidence in court and the unavailability of a specific approach increases the importance of the other. In the given context, the reliability of information can be influenced on three levels: (i) the implementation of clustering heuristics, (ii) annotations, and (iii) the correct use of the tools.

Clustering Reliability: Although proving the reliability of algorithms is subject to an ongoing discussion (Kehl et al., 2017), we can assume that a reliability assessment of clustering processes must consider the underlying (usually formalized) heuristics, its implementation (algorithm), and its functioning when being applied on a particular Cryptocurrency (logic of the process).

For example, so called CoinJoin transactions can undermine the logic of naïve[3] multi-input heuristics (see 2.1) and cause false cluster merges. CoinJoins violate the assumption that all inputs of the transaction are controlled by the same entity. The logic of the process must hence acknowledge this special case. Goldfeder et al. (2018) developed and evaluated a heuristic to identify potential CoinJoin transactions. They find the heuristic to be reliable and claim that at current state of research there is no known way to create a CoinJoin transaction without the indicators that would lead to a detection by the heuristic. To show the relevance of evaluation of the underlying logic, we searched for CoinJoin transactions using BlockSci (Kalodner et al., 2017) from Bitcoins inception until April 30th, 2018 and tested them against a naïve MIH (Meiklejohn et al., 2013) cluster-dataset of 40 049 947 clusters. In the whole dataset 723 247 of the clusters or around 2% contain at least one CoinJoin with a mean of around 17 addresses involved, produced by roughly 2.4 CoinJoin transactions. This example shows that naïve multi-input address clustering can be significantly biased by CoinJoin transactions (c.f. Fig. 3), especially if we look at larger clusters. Furthermore, new privacy focused bitcoin wallets such as Wasabi[4] include CoinJoins as easy to use privacy enhancing feature (Coindesk - Alyssa Hertig, 2019). This could ultimately lead to an increased usage of CoinJoins because of the reduced technical barrier and thus more false cluster merges. Clustering for law enforcement purposes thus must consider CoinJoins as a special case by employing heuristics described in Goldfeder et al. (2018) to proactively exclude or mark potential CoinJoin transactions in clusters to ensure the correctness of the process. The underlying problem holds true for other heuristics e.g. based on other logic assumptions such a change (Meiklejohn et al., 2013) or behavior patterns (nopara73, 2017). It is hence necessary to continuously reassess the underlying assumptions and address the identified shortcomings. Failure to do the one or the other easily results in a substantially lowered reliability of the analysis and can even render it useless.

Beyond the logic of the process, the overall effectiveness of a formalized heuristic could be achieved by testing it against some collected and verified ground truth (e.g. black-box testing). In the case of clustering heuristics, ground truth could be a set of known and verified Cryptocurrency wallets, each carrying a set of addresses belonging to the same real-world user. However, as discussed in Section 2, ground-truth wallet data is hard to obtain, often constructed ad-hoc, for scientific purposes only. Creation of a general, authoritative standard ground-truth dataset would ease the quantification of clustering effectiveness and provide specific reliability measures and probabilities, as in other forensic methods (e.g. DNA testing). At EU level, the 5th Anti-Money Laundering Directive (AMLD) stipulates the creation of a central user database. In future, LEA databases may provide ground-truth data and allow better reliability evaluation of forensic tools and could also be used to generate cyber-threat intelligence (CTI) (Ribaux and Wright, 2014, p. 498, p. 498).

Furthermore, clustering algorithms usually rely on user behavior assumptions and the respective Cryptocurrency protocol. Both are subject to change, so the underlying assumption of a clustering algorithm may lose validity over time. Assumptions and algorithms hence have to be reviewed and revised regularly. In addition, if the assumptions are known, the user behavior can deliberately be changed to influence the analysis outcomes. For example, users could create crafted CoinJoin transactions to mislead the multiple-input heuristic (see 2.1) or trick the change heuristic (Meiklejohn et al., 2013) into miss-classifying the change output, effectively merging clusters of senders and receivers. As long as the effectiveness of clustering remains difficult to evaluate, the evidential value of clustering techniques is limited and the description of the logic of the process (e.g. *heuristics, user behavior, protocols*) and expert testimonies become more important for clustering techniques to maintain a certain evidential value.

Attribution Tag Reliability: The reliability of an annotation tag largely depends on its origin, its generation procedure, and how it is processed and assigned to a certain address or cluster by a forensic tool. If, for instance, an investigator identifies a Cryptocurrency donation address on a known website and assigns a tag to that address (e.g. “Internet Archive”) after thorough consideration then we can consider this as being a highly reliable attribution tag. If a tag is extracted from a dataset that has been crawled by an unknown entity and unknown technical procedures at an unknown point in time and is then assigned to a large number of Cryptocurrency addresses via a tool's clustering algorithm, then we can consider this tag as being on the other side of the reliability spectrum. Since it is hard to quantify attribution tag reliability in a universal and interoperable manner, each attribution tag should provide details about its origin (*source*) and its *generation process* (e.g. manual extraction vs. automated crawl).

Correct Use of Tools: To prove the correct use of tools the interaction with the software has to be logged extensively. Logging also helps to explain investigation steps and should include information on the technical configuration to help assessing the overall reliability. Extensive logging can also help proving compliance with other general evidence rules and data protection rules (3.2 and 3.1). Where analysis results are shared between LEAs they should contain information on the points stated above to allow an assessment of reliability of the information at all times (Europol Regulation, 2016; Recommendation No. R (87) 15, 1987).

3.4. Qualification

Investigators and experts who use cryptocurrency analytics methods to obtain and analyze electronic and digital evidence in criminal proceedings must be qualified to use those methods (Mohay et al., 2003; Heinson, 2015; Casey, 2011). Lack of qualification in dealing with IT forensic methods can have considerable influence on investigations and the resulting evidence. On the one hand, the lack of qualification of the investigators involved lowers the evidential value of the investigation results. On the other hand, the lack of or inadequate qualification of investigators and court-appointed experts increases the probability that wrong conclusions will be drawn from the available evidence and, in the worst

[3] With naïve we mean multi-input clustering that does not have special handling for CoinJoins e.g. remove potential CoinJoin before clustering.

[4] https://wasabiwallet.io/.

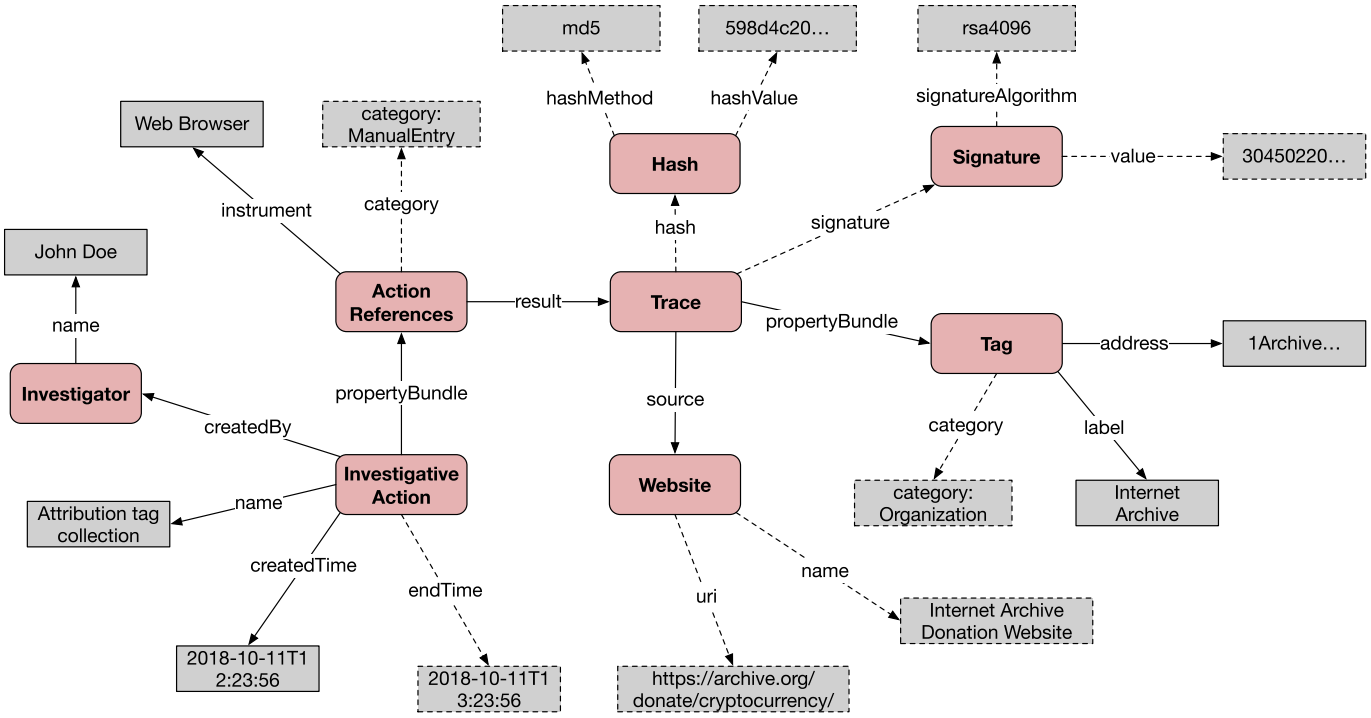


Fig. 4. Attribution Tag Sharing Model: a data model expressing main conceptual entities for describing attribution tags.

case, that the public prosecutor's office and/or the court will make their decisions on the basis of false facts or assumptions (US National Research Council, 2009; Heinson, 2015). Unfortunately, there are no international standards yet on what qualifications investigators and experts must have in handling electronic and digital evidence. To ensure acceptable minimum standards, investigators should at least have completed a certification course for the forensic software used and have basic training in IT forensics. Supervising investigators and court-appointed experts should have a university degree in IT forensics (Heinson, 2015).

Investigators who are involved in cryptocurrency investigations and use available tools should have demonstrated knowledge (e.g. certified training) on the basic architecture of cryptocurrencies, which includes the P2P communication layer as well as the blockchain that holds the transaction ledger. Specialized trainings should also cover the functionality of clustering heuristics, possible effects of adding an attribution tag to a certain address, and an understanding of attached provenance information in order to correctly assess and present (US National Research Council, 2009, p. 47) their authenticity and reliability.

3.5. Verifiability

The method of collecting data and gaining information must be repeatable and reproducible (Maras, 2015; Heinson, 2015; Casey, 2011). This ensures that involved lawyers can follow up the acquisition of information in subsequent legal proceedings (Heinson, 2015). With regard to limitations to disclosure and right to inspect the records (see 3.7), verifiability becomes even more important and must be ensured for the individual case. If, due to technical circumstances, repeatability or reproducibility cannot be achieved, the evidential value of the results obtained decreases considerably.

Following the recommendations of the US National Institute of Standards and Technology (NIST), repeatability means precision under repeatability conditions. Repeatability conditions are conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. Reproducibility describes precision under reproducibility conditions, which are defined as conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment (NIST, 2001; Maras, 2015). To account for individual errors of the investigators, the formal review of a process can be accompanied by peer-review with different tools (Casey, 2011, p.74).

Verifiability of Cryptocurrency analysis results obtained from forensic tools can be achieved by applying the same method that already provides authenticity and integrity of data: if identifiers of Cryptocurrency clusters are computed over a given Blockchain state (identified by block hash) by applying a specified hash function over the sorted set of addresses contained in a cluster, then the method becomes repeatable and reproducible when being applied on the same state of the underlying Blockchain.

3.6. Chain of evidence

In order to be used in rule of law criminal proceedings, the linking of circumstantial evidence and the conclusions drawn from it must be logical, consistent and compelling (Heinson, 2015, p. 136 et sqq.). Therefore, convictions and the establishment of a suspicion presuppose that the facts on which they are based have a certain quality, and that it can be concluded with a certain probability from the facts that the suspect/accused is indeed the offender (Schulz, 2001, p.593 et sqq.; Rich, 2016, p. 887). The degree of suspicion required for investigative measures and the standard of the court's persuasion for a conviction vary widely between the different national criminal procedural systems. However, they have in common that the various stages of suspicion are described with normative terms. The descriptions of the suspicion reflect a certain required quality of the facts and a required level of probability of a committed offense that can be derived from the facts (Schulz, 2001,

p. 593 et sqq.). Examples (US, UK, GER) include the “simple” - in German law sometimes also referred to as “reasonable” (Stehle, 2016) - suspicion to start an investigation (GERCCP, § 152 (2)), the “reasonable suspicion” for special investigation measures (UK PACE Code A; Terry v. Ohio, 1968; GER-CCP, § 100a), “probable cause”/ “urgent suspicion” for arrests and warrants (US Constitution, 4th Amdt.) (Illinois v. Gates, 1983; Stehle, 2016), and, last but not least, the “beyond a reasonable doubt” standard to convict the suspect (Victor v. Nebraska, 1993; Miles v. US, 1880; Woolmington v. DPP, 1935; BGH NJW, 1990, 1549) (GER-CCP, § 261) (Casey, 2011, p. 55). The necessary quality of the facts and the necessary probability of someone having committed the offense increases with the intensity of the investigation measures applied on the basis of the suspicion (e.g. search and seizure) (Schulz, 2001, p. 567 et sqq.)., For a conviction, the factual basis must be of the highest quality and the probability of the accused being the perpetrator must be sufficiently high to convince the court beyond reasonable doubt (Miles v. US, 1880).

It is difficult to harmonize the normative standard of evidence in criminal proceedings with statistical results of data analysis procedures (Rich, 2016; Meinicke, 2015). The criminal lawyers involved in the proceedings must be enabled to subsume the results of technical investigations under the normative concepts (Rich, 2016; Meinicke, 2015). Therefore, the relevant information must be presented to the criminal lawyers in a form and language that they can understand (Heinson, 2015, p. 130; Mohay et al., 2003, p. 132, 133; Casey, 2011, p. 78) When using data analysis techniques such as address clustering, both the software tools used and the investigating IT experts must therefore be able to provide exact information about which evidence is to be derived from the analysis and what conclusions can be drawn from it with what probability. In addition, possible sources of error must be identified (Rich, 2016), alternative hypotheses must be presented and, if necessary, comprehensibly excluded. This requires a thorough understanding of the data analysis method used (Rich, 2016) (see 3.4).

Finally, it is very important that the sources of information and data are reliable and traceable, when using address clustering and annotation tagging (see 3.3). If information that is not absolutely certain is used in the analysis, this circumstance and the resulting consequences for the result of the analysis and the suspicion or proof of the crime must be communicated to the criminal lawyers involved. A proposal for a linguistic categorization of the reliability of digital evidence can be found at (Casey, 2011, p. 69 et sqq.).

3.7. Right to inspect the records/disclosure of evidence

The right of the accused or her defence counsel to inspect the evidence gathered by the police and the public prosecutor's office is a key element of constitutional criminal proceedings Wessing (2018). In inquisitorial criminal procedure systems such as the German criminal procedure, this right is designed as the right to inspect records (GERCCP, § 147). Similarly, contradictory criminal procedure systems, such as the US-American one, constitute the right to disclosure of case-relevant evidence by the public prosecutor's office (e.g. UK (UK-CCP, Section 3) (UK Guidelines.

On Disclosure; UK Disclosure Manual), US: (US-FRE, Rule 705) (US-CCP, Rule 16), (Brady v. Mary, 1963; US v. Bagley, 1985) (Brown, 2017, p.148).). Additionally, similar rights can also arise from data protection (e.g. EULED, Recital 38, c. f. 3.1). When applying data analysis methods in either system, the question arises what information about the software tools used and the data and information processed must be disclosed to the defendant and his defenders. When answering this question, different interests have to be weighed. The defendant has a legitimate interest in ensuring that the method of gathering the evidence presented against her is made transparent in order to verify that the requirements laid down in Sections 3.1 to 3.6 are met (Wessing, 2018; Chessman, 2017; Singelnstein, 2018). One approach could be the disclosure of source code of the used software (Chessman, 2017; Meinicke, 2015). Having said that, law enforcement agencies have an interest in ensuring that the precise functioning of data analysis tools does not become widely known in criminal communities, which could make their use more problematic or impossible (Wilson, 2011, p.127; Short, 2010). (US v. Johnson, 2015). Additionally, the proprietary rights of the companies that produce and distribute the software tools used must be taken into account (Casey, 2011; Edwards and Veale, 2017; Chessman, 2017) (Minnesota v. Underdahl, 2008; NY v. Cialino, 2007; Wisconsin v. Loomis, 2016) Moreover, the validation of the source code alone does not account for individual errors of the investigator (Casey, 2011, p. 74). The highly complex question of balancing arises in both systems and cannot yet be conclusively answered. What is certain, however, is that the right to inspect the records/disclosure of evidence must be given high priority because of its enormous importance for constitutional criminal proceedings (Chessman, 2017). A first idea could be not to include the source code of the software, but a function and usage description (e.g. with regard to different description approaches (Wessing, 2018; Wachter et al., 2018).

With regard to address clustering, the right to inspect the records and disclosure of evidence mainly concerns the heuristics, their accuracy and also the usage of the tools in the specific case (e.g. searches in the database). In any case, the sources of the annotated information (e.g. other law enforcement agencies, private companies, publicly available sources) and the degree of reliability of the sources must be disclosed (UK Disclosure Manual, p.87 et sqq.) (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993; Sjerps and Berger, 2012).

3.8. Summary of key requirements

Lawfulness: Address clustering and annotation tags have to comply with the requirement for a *legal basis* and with *data protection principles*. Decisions, such as ordering further investigative measures may only be based on the results of *automated clustering* of Cryptocurrency addresses if the final decision is made by a human investigator and is not merely formalistic. Where relevant results are shown to be inaccurate, rectification is necessary.

Authenticity and Integrity: To ensure *authenticity and integrity* of the used address data, it is sufficient to record the currency code (e.g BTC, ETH) and the most recent block hash as identifiers within the chain of custody. Regarding clustering techniques, it is necessary to implement cluster identifiers that remain stable when a cluster refers to the same set of addresses over several runs and change when the underlying set of addresses changes. This can be achieved by computing a “cluster hash”. The authenticity of attribution tags should be assured by relating a tag to its source, its creator and generation procedures and computing a digital signature over the tag and all contextual information. This also increases the reliability of the tags.

Reliability: In order to achieve the highest possible level of reliability, the following measures should be taken:

- Testing the formalized heuristic against some collected and verified ground truth, ideally against a general, authoritative standard ground-truth data set (e.g. (shared) sets of addresses from known (seized) Cryptocurrency wallets).
- Testing the reliability of the clustering algorithm within the scope of particular Cryptocurrencies by using standard functional testing procedures and testing the function implementing

the clustering heuristic by feeding them a set of example transactions in a black-box test.
- Continuous review and rectification of the underlying clustering assumptions (e.g. MIH).
- Logging intensively the use of the software by investigators. When sharing analysis results: Sharing also any information that is necessary to assess the reliability of the information at all times.

Qualification: There are no international standards for the required qualification of IT-forensic investigators. The investigators involved in using address clustering and annotation tagging should at least have completed a *certified training* on the basic architecture of Cryptocurrencies, on the functionality of clustering heuristics, and on the possible effects of adding an attribution tag to a certain address and have developed a understanding of the attached provenance information.

Verifiability: Repeatability and Reproducibility of address clustering and annotation tagging can be achieved by the same measures that guarantee the authenticity and integrity of data. If the source of the tags is online data, it must be stored locally and permanently to guarantee the availability of the source.

Chain of Evidence: When using the results of address clustering techniques and annotation tagging in criminal proceedings, the criminal lawyers involved must be enabled to *subsume the results of the technical investigations under the normative concepts* of the respective criminal procedure code. Both the software tools used and the investigating IT experts have to provide exact information about which evidence is to be derived from the analysis and what conclusions can be drawn from it with what probability. The information must be presented in a language and form that is comprehensible to lawyers.

Right to Inspect the Records/Disclosure of Evidence: It is necessary to disclose a function and usage description of the used techniques as exact as possible. This means to disclose at least the heuristics used, the degree of probability of the results and the usage of the software tools in the specific case. Also, the sources, the process of generation of annotated information and the degree of reliability of both, the sources and the generation process must be disclosed.

4. Data sharing framework

After having analyzed the technical and legal factors influencing the evidential value of Cryptocurrency analytics techniques, we now proceed and propose a framework for data sharing and provenance tracking that, on the one hand, considers the efficiency of law enforcement agencies and, on the other hand, preserves evidential value of forensic investigations by adhering to legal key requirements (Section 3.8) in the context of law enforcement. In the following, we focus on sharing of attribution tags among law enforcement agencies, in a way that helps to ensure compliance with the key requirements such as authenticity, reliability or chain of evidence and helps to simplify the disclosure of evidence.

Previously, we emphasized the key role of clustering heuristics and attribution tags in Cryptocurrency investigations: a single tag can deanonymize a Cryptocurrency address and, when being combined with clustering techniques, also an entire address cluster that possibly represents some real-world actor or Cryptocurrency services like exchanges or wallet providers. Therefore, sharing attribution tags and cluster information among law enforcement agencies would certainly improve the effectiveness of forensic Cryptocurrency investigations.

The challenge in attribution tag sharing lies in finding the right trade-off between law enforcement needs, existing legal and ethical standards, as well as technical effort and practical feasibility. In the following, in Section 4.1, we first propose a lightweight data model for sharing attribution tags that should effectively balance those goals. Then, in Section 4.4, we suggest a model for sharing address clusters.

4.1. Attribution tag sharing

Our proposed attribution tag sharing model builds on the *Cyber-investigation Analysis Standard Expression (CASE)*[5] specification in order to maximize interoperability between tools and organization. Applying that model is somehow natural fit, since CASE is increasingly adopted as a standard for cyber-investigations and cryptocurrency forensics has become a standard digital forensics format. It also turned out that CASE already provides most of the semantic constructs for describing attribution tags.

Attribution tag collection and sharing can be considered as being a cyber-investigation and represented using a CASE `Investigation` object. More specifically, tag collection forms the beginning of a chain of custody and thereby an `Investigative Action`. Attribution tag collection can either be performed manually or by utilizing some associated Tool, such as a crawler. This can be expressed by utilizing the CASE Action Reference class. For representing an attribution tag, which is a specific type of `Trace`, we propose to extend CASE by a specific property bundle (`Tag`) that provides descriptive elements for attribution tags. Thus, already defined CASE concepts can easily be refined for attribution tag sharing as follows:

InvestigativeAction: each tag is the result of some `InvestigativeAction` that started and ended at some point in time and is carried out using some instrument (e.g., Web Browser) by some real-world actor. Technical details of the used instrument can be specified and named (e.g., Web browser) in an associated property bundle.

Investigator: represents a person who is responsible for an activity which is, in this case, creation or modification of tags. An agent carries a human-readable name (e.g. “John Doe”).

Tag: is a specific form of `Trace` and represents an attribution tag that attributes contextual information to some cryptocurrency identifier (e.g. an address). A tag can refer to some digital, physical, or purely conceptual thing and carries a unique identifier (e.g. http://exampletool.com/tag/1) in order to bind its meaning to a certain application context and to avoid naming collisions across contexts. A tag usually carries a human readable `label` (e.g. “Internet Archive”) and can be categorized (`category`) along several dimensions (see below).

Second, the data model also considers the authenticity and integrity requirements summarized in Section 3.8 by translating them into corresponding data model fields:

Hash: is a fingerprint of the tag description and provides integrity. It can be computed over the sorted set of attribute value pairs following an agreed-upon, standard hash function (*what*?).

Signature: an optional attribute to present the authenticity of an Agent (*who*?). It can either follow an agreed upon signature scheme or an signature abstraction such as described in RFC5126 (Pinkas, 2008).

Timestamps: automated and trusted time stamping (e.g. based on RFC3161 (Adams, 2001) or extensions) routines should record *when* an `Investigative Action` was performed. CASE create Time and `endTime` can be used for that purpose.

Source: each tag has been extracted from some digital or non-digital source. Each source carries a human-readable label (e.g.

[5] https://caseontology.org/ontology/start.html.

"Internet Archive Website") and possible some URI referring to that source (e.g. https://archive.org). To preserve the availability of the referenced document even if the original provider is not available anymore the WARC File Format (ISO 28500; ISO (2017)) could be used.

In order to avoid naming collisions, all concepts and relations, as well as their instances should carry *qualified names*. This can be achieved by assigning name spaces expressed as Internationalized Resource Identifiers (IRI). All previously introduced concepts, attributes and relations could, for instance, carry the namespace (http://case.example.org/core#), which, for convenience reasons, can be mapped to a prefix such as `case`. Specific examples are http://case.example.org/core# and `case:Tag`,[6] which both refer to the same concept definition expressed above. Fig. 4 shows the conceptual entities and relations of the proposed attribution tag data sharing model. Listing 1 shows an example attribution tag expressed in JSON-LD (Sporny et al., 2014) a JSON based linked data format specified by W3C.

```
{
  "@context": {
    "@vocab": "http://case.example.org/core#",
    "category": "http://example.com/category#"
  },
  "@graph": [
    {
      "@id": "tag-collection-uuid",
      "@type": "InvestigativeAction",
      "name": "Attribution tag collection",
      "createdBy": {"@id": "investigator1-uuid"},
      "createdTime": "2018-10-11T12:23:56",
      "endTime": "2018-10-11T13:23:56",
      "propertyBundle": [
        {
          "@type": "ActionReferences",
          "instrument": "Web Browser",
          "category": "category:ManualEntry",
          "result": [
            "http://exampletool.com/tag/1"
          ]
        }
      ]
    },
    {
      "@id": "investigator1-uuid",
      "@type": "Investigator",
      "name": "John Doe"
    },
    {
      "@id": "http://exampletool.com/tag/1",
      "@type": "Trace",
      "createdBy": "investigator1-uuid",
      "propertyBundle": {
          "@type": "Tag",
          "label": "Internet Archive",
          "address": "1Archive1n2C579dMsAu3iC6tWzuQJz8dN",
          "category": {
            "@id": "category:Organization"}
        },
      "source": {
          "@type": "Website",
          "name": "Internet Archive Donation Website",
          "uri": "https://archive.org/donate/cryptocurrency/"
        },
      "hash": {
          "@type":"Hash",
          "hashMethod": "md5",
          "hashValue": "598d4c200461b81522a3328565c25f7c"
        },
      "signature": {
          "@type":"Signature",
          "signatureAlgorithm": "rsa4096",
          "value": "304502206e21798a42fae0e854"
        }
      }
    ]
}
```

[6] CASE does not yet define a concept Tag. However, this could be introduced in future releases.

4.2. Categorization schemes

Assigning uniquely identifiable categories to the main conceptual entities of the tag sharing model is key for automated data processing. Some cryptocurrency analytics tools might, for instance, reject attribution tags that were automatically crawled from some (darknet) websites. Since making such decisions automatically based on manually entered descriptions is often error-prone (e.g. "Web Crawl" vs. "webcrawl"), data models should draw categories from pre-defined, agreed-upon vocabularies, which could be shared among stakeholders within a specific domain or application context. Such vocabularies should define categorization terms for each type of entity in the data sharing model (`Tag`, `Agent`, `Activity`, `Source`).

The definition of a full categorization schemes encompassing all relevant use cases is out of scope of this paper, but could be subject to a larger standardization effort. However, as a starting point, we suggest to consider at least the following categories for above entities:

Tag Categories: besides carrying a human-readable name (e.g. "Internet Archive") it could also be categorized by the type of real-world actor it represents. A real-world actor could be an `Organization`, an `Individual`, or an entity providing some service function in a Cryptocurrency ecosystem. For example a service might be: an `Exchange`, a `Wallet Provider`, a `Miner`, a `Marketplace`, etc.

Agent Categories: distinguishing between `Person` and `Organization` is a common refinement (c.f. FOAF vocabulary) for an `Agent` concept. Another possible use of categorization schemes could be the definition of reliability attributes (low, medium, high), which can be assigned to agents.

Source Categories: should denote the type of source tags were extracted from. A tag could be extracted from a `Website`, a `Data Dump` (e.g. from seized devices), a `Device`, a `Tor Hidden Service`, etc.

Action Categories: provide information on the type of action a tag was generated by. Common action types are `ManualEntry`, `Crawl`, etc. Action categories could be extended to provide additional details about the activity itself such as the tool and version used to create this tag. This enables better reproducibility.

4.3. Implementation considerations

Vocabularies and categorization schemes could be published on the Web by making sure that all terms (e.g. http://case.example.org/core#label) and concepts (e.g. http://example.com/category#ManualEntry) carry dereferencable IRIs. This allows searching and browsing available terms and categories online and to automatically verify attribution tag categories before exchanging them with others. Furthermore, it also provides a definition and documentation of terminology used in forensic investigations. A simple, straightforward way is to follow the implementation of `schema.org`, which is a generic schema for structured data on the Internet.

Previously we suggested that hashes and digital signatures can provide authenticity and integrity of attribution tags. However, for a data exchange purpose, this would require a precise and agreed-upon definition of the hash computation and digital signature procedures. Alternatively, one could use existing Git infrastructures for storing and publishing attribution tags. Git has its origin in distributed software development and is now the de-facto standard for publishing and tracking changes in source code files. It automatically creates hashes over each file and allows users to digitally sign their contents after each commit. Git is increasingly used for sharing smaller and even large datasets (Git LFS). Therefore, we

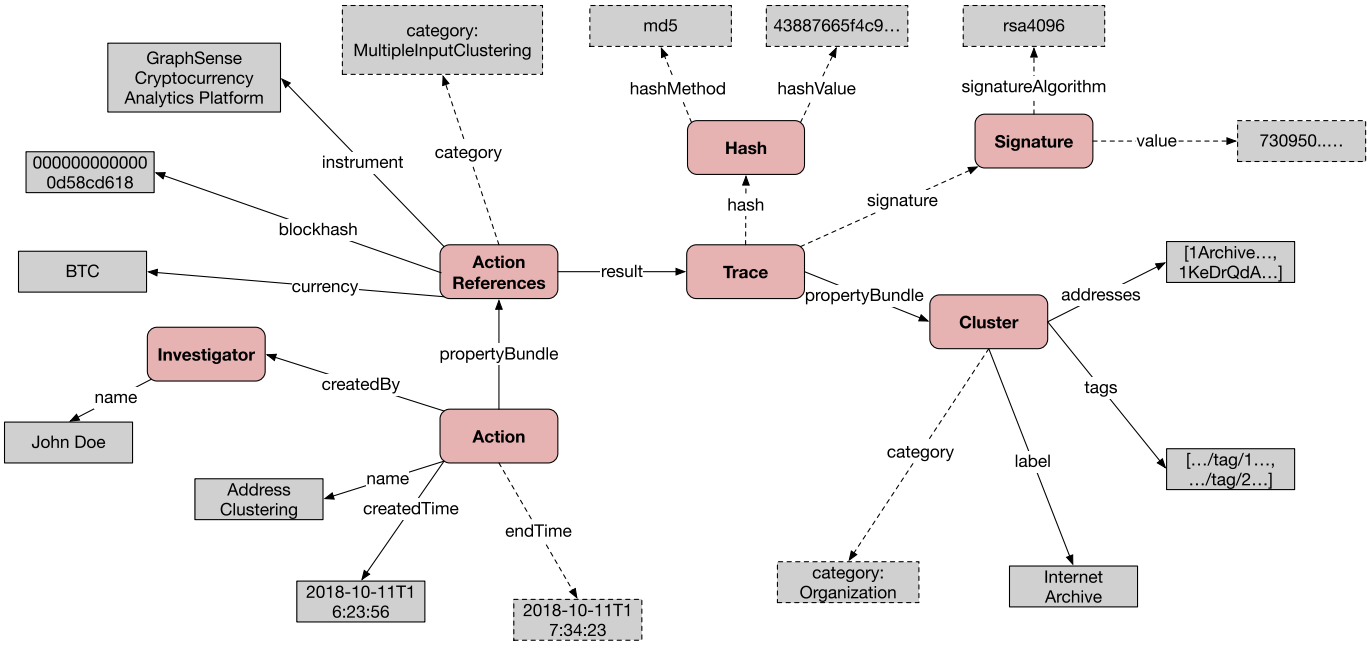


Fig. 5. Cluster Sharing Model: a data model expressing main conceptual entities for describing and sharing address clusters.

believe that it could also be used for sharing JSON-LD serializations of attribution tags. Although Git has good support for change tracking, Git has shortcomings when it comes to specifying fine grained access control policies. If tags are very sensitive and specific tags should not be shared with every participant other approaches need to be considered. Centralized services such as Europol's SIENA provide a data sharing solution with law enforcement ready data access controls.

4.4. Address cluster sharing

Sharing cluster information is a key ingredient to increase the reliability of clustering algorithms by allowing law enforcement agencies to share ground truth data (e.g. from seized wallet) to evaluate the accuracy of the clustering heuristics. Better reliability evaluations strengthen the evidential value of the findings produced using clustering algorithms (see Section 3.3). Furthermore, shared clusters combined with constant cluster identifiers, can be used to rectify (see Section 3.1) false clustering results by incorporating the shared (and rectified) clusters in the clustering process. Our proposed model for sharing address clusters, which is shown in Fig. 5, builds on the previously introduced attribution tag sharing model and introduces the following conceptual entities, attributes, and relationships:

Action: describes the process that produced a cluster. When clusters were created algorithmically the underlying procedure or heuristics must be named and, in the ideal case, be drawn from some controlled vocabulary that provides an exact definition of that procedure.

Cluster: a cluster is a specific type of `Trace` and represents a set of Cryptocurrency `addresses` and is a key entity to be exchanged within Cryptocurrency investigations. A cluster can carry a number of tags, which can be referenced by their unique (possibly dereferencable) IRIs. Authenticity can be shown by digitally signing the cluster with all its contextually relevant attributes.

Investigator: denotes the persons who controls the clustering method. Typically, clusters are created by forensic tools that implement certain heuristics or by manually identifying a set of addresses belonging to the same real-world actor.

Just as in attribution tag sharing model, all vocabulary terms and used categories should carry *qualified names*, which could be implemented as dereferencable IRIs. Listing 2 shows a JSON-LD serialization of the above cluster model example.

```
{
  "@context": {
    "@vocab": "http://case.example.org/core#",
    "category": "http://example.com/category#"
  },
  "@graph": [
    {
      "@id": "clustering-process-uuid",
      "@type": "Action",
      "name": "Address Clustering",
      "createdBy": {"@id": "investigator1-uuid"},
      "createdTime": "2018-10-11T16:23:56",
      "endTime": "2018-10-11T17:34:23",
      "propertyBundle": [
        {
          "@type": "ActionReferences",
          "instrument": "GraphSense Cryptocurrency Analytics Platform",
          "category": "category:MultipleInputClustering",
          "currency": "BTC",
          "blockhash": "0000000000000d58cd618",
          "result": [
            {"@id": "http://exampletool.com/cluster/1"},
            {"@id": "http://exampletool.com/cluster/2"}
          ]
        }
      ]
    },
    {
      "@id": "http://exampletool.com/cluster/1",
      "@type": "Trace",
      "createdBy": {"@id": "investigator1-uuid"},
      "propertyBundle": {
          "@type": "Cluster",
          "addresses": [
            "1Archive1n2C579dMsAu3iC6tWzuQJz8dN",
            "1KeDrQdATuXaZFW4CL9tfe2zpQ5SrmBFWc",
            "1Lyr44vPADgVFqqMCvykx5KxG15PKVYjW3"
          ],
          "tags": [
            {"@id":"http://exampletool.com/tag/1"},
            {"@id":"http://exampletool.com/tag/2"}
          ]
        },
      "hash": {
          "@type":"Hash",
          "hashMethod": "md5",
          "hashValue": "43887665f4c94641c357c111aa12acb1"
        },
      "signature": {
          "@type":"Signature",
          "signatureAlgorithm": "rsa4096",
          "value": "73095072206e21798a42fae0e854"
        }
      }
    ]
}
```

5. Discussion

We systematically analyzed the possibilities and shortcomings of known cryptocurrency forensics techniques from a combined legal and technical perspective. Our legal analysis has shown that so far there are no internationally binding standards for measuring, securing, or increasing the evidential value of the results of Address clustering and Annotation tagging. However, by synthesizing different criminal procedural codes and data protection regulations, minimum standards can be obtained which can claim a certain validity in any constitutional criminal procedure.

Address clustering and Annotation tagging have to comply with the requirement of having a legal basis and with data protection principles. To ensure a high evidential value, Address clustering and Annotation tagging have to meet the requirements of the chain of custody (authenticity and integrity) and reliability. Based on an analysis of CoinJoin appearances in MIH-clusters we show the necessity of continuous reassessment of underlying assumptions for clustering heuristics and the importance of proactive measures to maintain the reliability of the approach. Investigators using these techniques must be sufficiently qualified to do so (e.g. by certified courses). The results of Address clustering and Annotation tagging must be repeatable and reproducible (if possible). Moreover, the criminal lawyers involved must be enabled to subsume the results under the normative concepts of the applicable criminal procedure code by presenting the results in an language and form that is comprehensible for lawyers. To meet the requirement to inspect the records/the principle of disclosure of evidence it is necessary to disclose a exact function and usage description of the used heuristics, the used software tools and the source and generation process of annotated information. Decisions on further investigative measures (e.g. a property search) and a conviction may only be based on automated findings if the final decision is made by a person and her decision is not merely formalistic.

Our analysis showed the need for reliable ground truth data. This could be achieved by sharing seized wallets or at least the addresses belonging to a wallet to build a comprehensive ground truth dataset, possibly via a central database. Furthermore, the effectiveness of clustering heuristics often relies on assumptions about user behavior. User behavior can obviously change and thus potentially invalidate the underlying assumptions, therefore constant monitoring and reevaluation is needed.

We have used the insights into technical challenges and the legal requirements of Cryptocurrency investigations to develop a data sharing model that helps to preserve the evidential value of the gathered evidence. It has been shown that it is possible to meet (most) legal requirements for securing the evidential value and complying with the principles of data protection through easy-to-implement and practically applicable measures. Furthermore, we laid a stepping stone for future data sharing and standardization efforts in the field of Cryptocurrency investigations.

A clear technical limitation of this paper is that we only investigated the multi-input clustering heuristic and the Bitcoin system. Moreover, we did not take every existing or imaginable mixing technique into account. Nevertheless, many of the findings can be transferred to other Cryptocurrency systems, heuristics and mixing techniques or at least serve as a basis for further research in those areas. In the legal part, we tried to cover a broad spectrum of legal systems by using both sources from the Anglo-American legal systems (Common Law) and those from the continental legal systems (mostly Civil Law). Of course, the findings of the legal analysis still need to be adapted for practical application in particular countries for the criminal procedure codes applicable there. However, our model can serve as a blueprint for doing so since it establishes minimum standards for constitutional criminal procedures.

6. Conclusions

In this paper, we discussed clustering heuristics and attribution tags, which are the two key techniques implemented in forensic tools used in Cryptocurrency investigations. By empirically quantifying the effect of CoinJoin transactions we illustrated that the reliability of clustering heuristics can only be upheld if the logic of the underlying process is verified and the outcomes are correctly interpreted. We discussed those techniques in the light of internationally accepted legal standards and rules for substantiating suspicions and providing evidence in court. From that, we derived a set of legal key requirements and translated them into a data sharing framework that builds on existing legal and technical standards. We propose the implementation of this framework in tools used for Cryptocurrency investigations to safeguard the value of produced evidence.

Future research on the technical side should focus on additional metrics that can help in quantifying the reliability of clustering results. On the legal side, the possibilities and limitations of merging the results of (automated) Address clustering techniques with the normative concept of suspicion as well as the necessary persuasion of the court for a conviction need to be further examined. It is also worthwhile to consider possibilities for standardizing attribution tags and cluster sharing in the field of international law enforcement. Finally, our paper can be an anchor point for future research on other heuristics, mixing techniques, and the applicability of Address clustering and Annotation tagging in other Cryptocurrency systems to foster the evidential value of Cryptocurrency analyses.

References

Nakamoto, S., 2008. Bitcoin: a peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf. (Accessed 1 July 2015). https://bitcoin.org/bitcoin.pdf.

Adams, C., 2001. Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). Technical Report. IETF Network Working Group. RFC 3161.

Albrecht, J.P., 2016. How the GDPR will change the world. Eur. Data Prot. L. Rev. 2, 287. https://doi.org/10.21552/EDPL/2016/3/4.

Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T., Capkun, S., 2013. Evaluating user privacy in bitcoin. In: International Conference on Financial Cryptography and Data Security. Springer, pp. 34–51.

Bonneau, J., Miller, A., Clark, J., Narayanan, A., Kroll, J.A., Felten, E.W., 2015. SoK: research perspectives and challenges for bitcoin and cryptocurrencies. In: I EEE Symposium on Security and Privacy. http://www.ieee-security.org/TC/SP2015/papers-archived/6949a104.pdf.

Bose, R., Frew, J., 2005. Lineage retrieval for scientific data processing: a survey. ACM Comput. Surv. 37, 1–28. https://doi.org/10.1145/1057977.1057978, 10.1145/1057977.1057978.

Brown, D.K., 2017. Discovery in State Criminal Justice. Reforming Criminal Justice.

Casey, E., 2011. Digital Evidence and Computer Crime, 3rd. Ed. (Elsevier) Academic Press, London, UK; San Diego, CA, USA.

Casey, Eoghan, Back, G.B.S., 2015. Leveraging CybOX™ to standardize representation and exchange of digital forensic information. Digit. Invest. 12, 102–110. https://doi.org/10.1016/j.diin.2015.01.014. (Accessed 19 June 2019) [Online].

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. Digit. Invest. 22, 14–45. https://doi.org/10.1016/j.diin.2017.08.002. (Accessed 17 June 2019) [Online].

Chessman, C., 2017. A "source" of error: computer code, criminal defendants, and the constitution. Calif. Law Rev. 179, 302–333.

Coindesk - Alyssa Hertig, 2019. 100 bitcoin users perform what might Be largest 'CoinJoin' transaction ever). https://www.coindesk.com/bitcoin-users-perform-what-might-be-the-largest-coinjoin-ever?utm_source=BlockchainBites&utm_medium=Email&utm_campaign=2019-06-11. (Accessed 17 December 2020) [Online].

Colorado Court of Appeals, Div, V., 2002. The People of the State of Colorado V. Daniel Huehn. https://law.justia.com/cases/colorado/court-of-appeals/2002/00ca0505-0.html. (Accessed 17 June 2019) [Online].

Cosic, J., Baca, M., 2010. Do we have full control over integrity in digital evidence life cycle?. In: Information Technology Interfaces (ITI), 2010 32nd International Conference on, IEEE, pp. 429–434.

Cui, Y., Widom, J., 2003. Lineage tracing for general data warehouse transformations. The VLDB Journal—The International Journal on Very Large Data Bases 12, 41–58.

Edwards, L., Veale, M., 2017. Slave to the Algorithm? Why a 'Right to an Explanation'

is probably not the remedy you are looking for. Duke Law & Technology Review 18 16. (Accessed 17 June 2019) [Online].
Ermilov, D., Panov, M., Yanovich, Y., 2017. Automatic bitcoin address clustering. In: Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on, IEEE, pp. 461–466.
Eschelbach, R., 2018. Regeln für die Würdigung. In: Beck'scher Onlinekommentar StPO mit RiStBV und MiStra. Beck.
European Parliament, Council of the European Union, 2016b. Regulation (EU) 2016/794. Article 29 (1), : The reliability of the source of information originating from a Member State shall be assessed as far as possible by the providing Member State using the following source evaluation codes: A, B, C, X; (2) The accuracy of information originating from a Member State shall be assessed as far as possible by the providing Member State using the following information evaluation codes: 1, 2, 3, 4.; [Online]; accessed 16-June-2019. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0794%&from=EN.
FATF, 2015. Virtual Currencies: Guidance for a Risk-Based Approach. Technical Report. Financial Action Task Force.
Freire, J., Silva, C.T., Koop, D., Santos, E., 2008. Provenance for computational tasks: a survey. Comput. Sci. Eng. 10, 11–21. https://doi.org/10.1109/MCSE.2008.79 doi.ieeecomputersociety.org/10.1109/MCSE.2008.79.
Garfinkel, S.S., 2010. Digital forensics research: the next 10 years. Digit. Invest. 7, 64–73. https://doi.org/10.1016/j.diin.2010.05.0090. (Accessed 17 June 2019) [Online].
Gil, Y., Cheney, J., Groth, P., Hartig, O., Miles, S., Moreau, L., Da Silva, P.P., Coppens, S., Garijo, D., Gomez, J.M., Missier, P., Myers, J., Sahoo, S., Zhao, J., 2010. Provenance XG Final Report. techreport. W3C.
Giova, G., 2011. Improving chain of custody in forensic investigation of electronic digital systems. International Journal of Computer Science and Network Security 11, 1–9.
Golder, S.A., Huberman, B.A., 2006. Usage patterns of collaborative tagging systems. J. Inf. Sci. 32, 198–208. https://doi.org/10.1177/0165551506062337.
Goldfeder, S., Kalodner, H., Reisman, D., Narayanan, A., 2018. When the cookie meets the blockchain: privacy risks of web payments via cryptocurrencies. Proceedings on Privacy Enhancing Technologies 2018 179–199.
Good, Irving, J., 1950. Probability and the Weighing of Evidence (–).
Hadziselimovic, E., Fatema, K., Pandit, H., Lewis, D., 2017. Linked data contracts to support data protection and data ethics in the sharing of scientific data. In: Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci). CEUR Workshop Proceedings, pp. 55–62.
Harrigan, M., Fretter, C., 2016. The unreasonable effectiveness of address clustering. In: Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences, IEEE, pp. 368–373.
Heinson, D., 2015. IT-forensik. Mohr Siebeck, Tuebingen, Germany.
Huang, D.Y., McCoy, D., Aliapoulios, M.M., Li, V.G., Invernizzi, L., Bursztein, E., McRoberts, K., Levin, J., Levchenko, K., Snoeren, A.C., 2018. Tracking ransomware end-to-end. In: Tracking Ransomware End-To-End. IEEE, 0.
Interpol, 2018. INTERPOL holds first DarkNet and cryptocurrencies working group. https://www.interpol.int/News-and-media/News/2018/N2018-022. (Accessed 30 October 2018) [Online].
ISO, 2017. WARC File Format. https://www.iso.org/standard/68004.html. (Accessed 12 November 2018) [Online].
Judmayer, A., Stifter, N., Krombholz, K., Weippl, E.R., 2017. Blocks and Chains: Introduction to Bitcoin, Cryptocurrencies, and Their Consensus Mechanisms. Synthesis Lectures on Information Security, Privacy, and Trust. Morgan & Claypool Publishers.
Kalodner, H.A., Goldfeder, S., Chator, A., Möser, M., Narayanan, A., 2017. Blocksci: design and applications of a blockchain analysis platform. CoRR abs/1709.02489. http://arxiv.org/abs/1709.02489, arXiv:1709.02489.
Kappos, G., Yousaf, H., Maller, M., Meiklejohn, S., 2018. An empirical analysis of anonymity in Zcash. In: 27th USENIX Security Symposium (USENIX Security 18). USENIX Association, Baltimore, MD, pp. 463–477. https://www.usenix.org/conference/usenixsecurity18/presentation/kappos.
Karvounarakis, G., 2009. Provenance in Collaborative Data Sharing. University of Pennsylvania, Philadelphia, PA.
Kehl, D., Guo, P., Kessler, S., 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Technical Report. Responsive Communities Initiative. Berkman Klein Center for Internet & Society, Harvard Law School. (Accessed 16 June 2019) [Online].
Kelly, M., Weigle, M.C., 2012. WARCreate: create wayback-consumable WARC files from any webpage. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, New York, NY, USA, pp. 437–438. https://doi.org/10.1145/2232817.2232930.
Kumar, A., Fischer, C., Tople, S., Saxena, P., 2017. A traceability analysis of monero's blockchain. In: European Symposium on Research in Computer Security. Springer, pp. 153–173.
Lund, S., Iyer, H., 2017. Likelihood Ratio as Weight of Forensic Evidence: A Closer Look. Technical Report. Statistical Engineering Division, Information Technology Laboratory - National Institute of Standards and Technology, USA [Online]. (Accessed 17 June 2019).
Maras, H.M., 2015. Computer Forensics, second ed. Jones & Bartlett Learning, Burlington, MA, USA.
Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S., 2013. A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 2013 Conference on Internet Measurement Conference. ACM.
Meinicke, D., 2015. Big Data und Data-Mining: automatisierte Strafverfolgung als neue Wunderwaffe der Verbrechensbekämpfung? K&R 6, 377–384.
Miebach, K., 2016. Empirical judgement in criminal proceedings. In: Münchener Kommentar zur StPO. Beck.
Miller, A., Möser, M., Lee, K., Narayanan, A., 2017. An empirical analysis of linkability in the monero blockchain arXiv preprint arXiv:1704.04299.
Mohay, G.M., Anderson, A., Collie, B., McKemmish, R.D., Vel, O.d., 2003. Computer and Intrusion Forensics. Artech House, Inc., Norwood, MA, USA.
Monaco, J.V., 2015. Identifying bitcoin users by transaction behavior. In: SPIE Defense+ Security. International Society for Optics and Photonics, 945704–945704.
Möser, M., 2013. Anonymity of bitcoin transactions. In: Münster Bitcoin Conference, pp. 17–18.
Möser, M., Böhme, R., 2016. Join me on a market for anonymity. In: Proceedings of the Workshop on the Economics of Information Security (WEIS). University of California at Berkeley (Publication status: Accepted).
Neumann, C., Kaye, D., Jackson, G., Reyna, V., Ranadive, A., 2016. Presenting quantitative and qualitative information on forensic science evidence in the courtroom. Chance 29, 37–43.
Nick, J.D., 2015. Data-driven De-anonymization in Bitcoin. Master's thesis. ETH-Zürich.
of Standards, N.I., (NIST), T.. Recommended security controls for federal information systems and organizations. https://csrc.nist.gov/publications/detail/sp/800-92/final. (Accessed 17 June 2019) [Online].
of Standards, N.I., (NIST), T., 2001. General test methodology for computer forensic tools. https://www.nist.gov/sites/default/files/documents/2017/05/09/Test-Methodology-7.doc. (Accessed 28 June 2018) [Online].
nopara73, 2017. New bitcoin anonymity technique: the clusterfuck wallet. https://medium.com/@nopara73/new-bitcoin-anonymity-technique-the-clusterfuck-wallet-d48aa1787324. (Accessed 28 June 2018) [Online].
Paquet-Clouston, M., Haslhofer, B., Dupont, B., 2018. Ransomware Payments in the Bitcoin Ecosystem arXiv preprint arXiv:1804.04080.
Pinkas, D., 2008. CMS Advanced Electronic Signatures (CAdES). Technical Report. IETF Network Working Group. RFC 5126.
Pollitt, M.M., 2007. An ad hoc review of digital forensic models. In: Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'07). https://doi.org/10.1109/SADFE.2007.3.
Quesnelle, J., 2018. An Analysis of Anonymity in the Zcash Cryptocurrency. Master's thesis. University of Michigan-Dearborn. http://hdl.handle.net/2027.42/143130.
Ribaux, O., Wright, B.T., 2014. Expanding Forensic Science through Forensic Intelligence. https://doi.org/10.1016/j.scijus.2014.05.001. (Accessed 17 June 2019) [Online].
Rich, M.L., 2016. Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment, vol. 164. University of Pennsylvania Law Review, pp. 871–929.
Schulz, L., 2001. Normiertes Misstrauen. Vittorio Klostermann. Franfurt a.M., Germany.
Short, C., 2010. Guilt by Machine: the Problem of Source Code Discovery in Florida DUI Prosecutions. Florida Law Review [Online]. (Accessed 16 June 2019).
Singelnstein, T., 2018. Predictive Policing: algorithmenbasierte Straftatprognosen zur vorausschauenden Kriminalintervention. Neue Z. Strafr. 1–9.
Sjerps, M., Berger, C., 2012. How clear is transparent? Reporting expert reasoning in legal cases. Law Probab. Risk 11, 318–329. https://doi.org/10.1093/lpr/mgs017.
Soska, K., Christin, N., 2015. Measuring the longitudinal evolution of the online anonymous Marketplace ecosystem. In: USENIX Security Symposium, pp. 33–48.
Sporny, M., Kellogg, G., Lanthaler, M., 2014. JSON-LD 1.0: a JSON-Based Serialization for Linked Data. Technical Report. W3C.
State Senate, C.A., CA State Assembly, 2018. California consumer privacy act. (Accessed 17 June 2019) [Online]. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.
Stehle, S., 2016. Empirical judgement in criminal proceedings. In: Legal Certainty in a Contemporary Context. Springer, pp. 125–146.
Stoffers, M., 2017. Trustworthy Provenance Recording Using a Blockchain-like Database. Ph.D. thesis. Leipzig University. http://elib.dlr.de/111772/1/thesis.pdf.
Tschorsch, F., Scheuermann, B., 2016. Bitcoin and beyond: a technical survey on decentralized digital currencies. IEEE Communications Surveys Tutorials 18, 2084–2123. https://doi.org/10.1109/COMST.2016.2535718.
van Baar, R.B., van Beek, H.M.A., van Eijk, E.J., 2014. Digital Forensics as a Service: a Game Changer. https://doi.org/10.1016/j.diin.2014.03.007. (Accessed 17 June 2019) [Online].
Wachter, S., Mittelstadt, B., Russel, C., 2018. Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv. J. Law Technol. 31 https://doi.org/10.2139/ssrn.3063289. (Accessed 16 June 2019) [Online].
Wessing, R., 2018. Akteneinsichtsrecht, Besichtigungsrecht; Auskunftsrecht des Beschuldigten. In: Beck'scher Onlinekommentar StPO mit RiStBV und MiStra. Beck.
Wilson, A.J., 2011. Discovery of breathalyzer source code in DUI prosecutions. Washington journal of law, technology and arts 7. http://digital.law.washington.edu/dspace-law/handle/1773.1/1069. (Accessed 16 June 2019) [Online].
Wisconsin Supreme Court, 2016. State of Wisconsin v. Loomis. https://law.justia.com/cases/wisconsin/supreme-court/2016/2015ap000157-cr.html.S.
Lorraine v. Markel American Insurance Company, 2007, United States District Court

for the District of Maryland, https://federalevidence.com/pdf/2013/02Feb/Lorraine.v.MarkelAm.241 F.R.D. 534_D.Md.2007.pdf, [Online]; accessed 17-June-2019.
Kumho Tire Company, Ltd, et al. v. Carmichael et. al., 1999, United States Court of Appeals for the Eleventh Circuit, https://supreme.justia.com/cases/federal/us/526/137/, [Online]; accessed 17-June-2019.
Terry v. Ohio, 1968, Supreme Court of the United States, https://supreme.justia.com/cases/federal/us/392/1/, [Online]; accessed 17-June-2019.
Illinois v. Gates, 1983, Supreme Court of the United States, https://supreme.justia.com/cases/federal/us/462/213/, [Online]; accessed 17-June-2019.
Victor v. Nebraska, 1993, Supreme Court of the United States, https://supreme.justia.com/cases/federal/us/511/1/, [Online]; accessed 17-June-2019.
Brady v. State of Maryland, 1963, Supreme Court of the United States, https://supreme.justia.com/cases/federal/us/373/83/, [Online]; accessed 17-June-2019.
United States v. Bagley, 1963, Supreme Court of the United States, https://supreme.justia.com/cases/federal/us/473/667/, [Online]; accessed 17-June-2019.
Jdg. Randolph W. Peterson, Court of Appeals, State of Minnesota v. Underdahl (Consolidated Opinion), 2008, https://www.courtlistener.com/opinion/1917266/state-v-underdahl/, [Online]; accessed 17-June-2019.

Michael Frömis is a PhD student in the Department of Computer Science at the University of Innsbruck. His research focuses on privacy, smart contract analysis, and blockchain forensics. Before joining the University of Innsbruck, he has worked as a software developer in a bank.

Thilo Gottschalk is a legal research associate at the Institute for Information and Business Law (IIWR) at Karlsruhe Institute of Technology. His research is focused on legal implications of data processing for law enforcement purposes with a particular interest in novel policing methods, IT security, software development, data protection and privacy.

Bernhard Haslhofer works as a Senior Scientist at the Austrian Institute of Technology's Digital Insight Lab. At the moment, he primarily works on novel methods for analyzing the structure and dynamics of cryptocurrency ecosystems, with a special focus on Post-Bitcoin cryptocurrencies such as Monero or Zcash.

Christian Rückert is a senior legal research associate at the Friedrich-Alexander-University Erlangen-Nuremberg and at the Institute for Information and Business Law (IIWR) at Karlsruhe Institute of Technology. He holds a PhD in Criminal Law. His research is focused on legal implications of data processing in criminal proceedings.

Paulina Jo Pesch is a senior legal research associate at the Institute for Information and Business Law (IIWR) at Karlsruhe Institute of Technology. She holds a PhD in Civil Law. She has researched cryptocurrencies, blockchain technology, smart contracts, and the legal implications of the technologies since 2014. Her current research is mainly focused on data protection law.