<table>
<thead>
<tr>
<th><strong>Grant Agreement Number</strong></th>
<th>740558</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Project Title</strong></td>
<td>Tools for the Investigation of Transactions in Underground Markets</td>
</tr>
<tr>
<td><strong>Project Acronym</strong></td>
<td>TITANIUM</td>
</tr>
<tr>
<td><strong>Title of Deliverable</strong></td>
<td>Report on shared data models</td>
</tr>
<tr>
<td><strong>Deliverable Number</strong></td>
<td>D.3.4</td>
</tr>
<tr>
<td><strong>Work-package</strong></td>
<td>WP3</td>
</tr>
<tr>
<td><strong>Dissemination Level</strong></td>
<td>PU = Public</td>
</tr>
<tr>
<td><strong>Deliverable Nature</strong></td>
<td>R = Report</td>
</tr>
<tr>
<td><strong>Contractual Delivery Date</strong></td>
<td>2018-10-31</td>
</tr>
<tr>
<td><strong>Actual Delivery Date</strong></td>
<td>2018-10-31</td>
</tr>
<tr>
<td><strong>Author(s)</strong></td>
<td>Michael Fröwis, Bernhard Haslhoifer, Christian Rückert, Paulina Pesch, Thilo Gottschalk</td>
</tr>
</tbody>
</table>

**Keywords:** cryptocurrency forensics, data sharing, standards
Abstract

Cryptocurrency analytics has become a key forensic method in law enforcement and is nowadays used to investigate a wide spectrum of criminal activities. However, despite its wide adoption, the evidential value of obtained findings in court is still unclear. In this deliverable, we focus on the key ingredients of modern cryptocurrency analytic techniques, which are clustering heuristics and attribution tags, and discuss them in the light of internationally accepted legal standards and rules for substantiating suspicions and providing evidence in court. We translate derived key requirements into a data sharing framework that builds on existing standards and strengthens the evidential value of forensic cryptocurrency investigations. We believe that the implementation of our proposed framework could increase the efficiency and effectiveness of forensic cryptocurrency investigations, while safeguarding their evidential value.
## Authors

<table>
<thead>
<tr>
<th>Person</th>
<th>Role</th>
<th>Partner</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernhard Haslhofer</td>
<td>Author</td>
<td>AIT</td>
<td>Technical aspects</td>
</tr>
<tr>
<td>Michael Fröwis</td>
<td>Author</td>
<td>UIBK</td>
<td>Technical and empirical aspects</td>
</tr>
<tr>
<td>Christian Rückert</td>
<td>Author</td>
<td>KIT</td>
<td>Legal aspects</td>
</tr>
<tr>
<td>Thilo Gottschalök</td>
<td>Author</td>
<td>KIT</td>
<td>Legal aspects</td>
</tr>
<tr>
<td>Paulina Pesch</td>
<td>Author</td>
<td>KIT</td>
<td>Legal aspects</td>
</tr>
</tbody>
</table>

## Document Approval

<table>
<thead>
<tr>
<th>Person</th>
<th>Role</th>
<th>Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vincent Danjean</td>
<td>Reviewer</td>
<td>Interpol</td>
</tr>
</tbody>
</table>

## Distribution

<table>
<thead>
<tr>
<th>Person</th>
<th>Role</th>
<th>Partner</th>
</tr>
</thead>
</table>

## Revision History

<table>
<thead>
<tr>
<th>Version</th>
<th>Status</th>
<th>Author</th>
<th>Date</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>Draft</td>
<td>See above</td>
<td>2018-10-12</td>
<td>Initial Version</td>
</tr>
<tr>
<td>0.9</td>
<td>Validated</td>
<td>Bernhard Haslhofer</td>
<td>2018-10-29</td>
<td>Review revisions</td>
</tr>
<tr>
<td>1.0</td>
<td>Submitted</td>
<td>Ross King</td>
<td>2018-10-31</td>
<td>Submitted Version</td>
</tr>
</tbody>
</table>

## References

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Document</th>
<th>Date</th>
<th>Details and Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Executive Summary

Cryptocurrency analytics has become a key forensic method in law enforcement and is nowadays used to investigate a wide spectrum of criminal activities. However, despite its wide adoption, the evidential value of obtained findings in court is still unclear. In this deliverable, we focus on the key ingredients of modern cryptocurrency analytic techniques, which are clustering heuristics and attribution tags, and discuss them in the light of internationally accepted legal standards and rules for substantiating suspicions and providing evidence in court. We translate derived key requirements into a data sharing framework that builds on existing standards and strengthens the evidential value of forensic cryptocurrency investigations. We believe that the implementation of our proposed framework could increase the efficiency and effectiveness of forensic cryptocurrency investigations, while safeguarding their evidential value.
Table of Contents

Deliverable ................................................................................................................................. i

Executive Summary ................................................................................................................... iv

1 Introduction ............................................................................................................................... 1

2 Background and Related Work ............................................................................................... 2
  2.1 Address Clustering Heuristics ............................................................................................ 2
  2.2 Attribution Tags .................................................................................................................. 3
  2.3 Provenance and Digital Evidence ...................................................................................... 4

3 Empirical Analysis ................................................................................................................... 5

4 The Legal Perspective .............................................................................................................. 6
  4.1 Lawfulness of Data Processing ......................................................................................... 6
  4.2 Authenticity and integrity of data and attributed information/chain of custody .......... 7
  4.3 Reliability ........................................................................................................................ 8
    4.3.1 Clustering Reliability .................................................................................................. 9
    4.3.2 Attribution Tag Reliability ....................................................................................... 9
    4.3.3 Correct Use of Tools ................................................................................................ 9
  4.4 Qualification ....................................................................................................................... 9
  4.5 Verifiability ........................................................................................................................ 10
  4.6 Chain of Evidence ............................................................................................................. 10
  4.7 Right to Inspect the Records/Disclosure of Evidence ....................................................... 11
  4.8 Summary of Key Requirements ....................................................................................... 12

5 Data Sharing Framework ......................................................................................................... 13
  5.1 Attribution Tag Sharing ..................................................................................................... 13
  5.2 Categorization Schemes .................................................................................................... 16
  5.3 Implementation Considerations ......................................................................................... 16
  5.4 Address Cluster Sharing .................................................................................................. 17

6 Conclusions ............................................................................................................................. 19

7 References ............................................................................................................................... 19
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 [SC15], over ransomware attacks [HMA18, PCHD18], to extortion and money laundering [For15]. 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) and non-commercial analysis tools (GraphSense [HKF16], BlockSci [KGC17]) that exploit the openness of cryptocurrency transaction ledgers (blockchains). They build on a long history of research that has shown that pseudonymous addresses do not provide sufficient anonymity, neither in Bitcoin [MPJ13, AKR13, Mös13, Mon15] nor in post-Bitcoin currencies with stronger privacy-enhancing techniques such as Zcash [Que18, KYMM18] or Monero [MMLN17, KFTS17]. Investigation tools mainly rely on two complementary techniques: clustering heuristics, which supports grouping of multiple addresses into maximal subsets that can be likely assigned to the same real-world actor, and attribution tags, which is any form of context information that can be attributed to an address, 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., [KFTS17]).

However, despite the specious benefits of cryptocurrency analytics techniques in criminal investigations, the evidential value of those techniques is still largely unclear: first, certain types of transactions (e.g., CoinJoin [MB16]) could distort clustering results, unify entities that have no association in the real-world, or lead to the formation of super-clusters [HF16]. 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. Those issues amplify when attribution tags are shared among law enforcement in global knowledge bases1.

In this deliverable, we propose measures for safeguarding the evidential value of forensic cryptocurrency investigation results. After introducing the necessary background, we make a number of contributions that can be summarized as follows:

- We empirically quantify the effects of CoinJoin transactions on cluster formation and dispersion.
- We systematically investigate internationally accepted legal standards and rules for providing court-proof evidence and derive key requirements for forensic cryptocurrency investigations
- Finally, we translate those requirements into a data sharing framework for law enforcement agencies that provides safe guards for increasing the evidential value of forensic cryptocurrency investigations, while ensuring compliance with existing regulations.

---

1 Law enforcement agencies have recognized the value of information sharing to maximize investigative resources and avoid duplicate efforts [Int18].
The present deliverable tackles 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 at complying with existing regulations.

2 Background and Related Work

In this section, we briefly introduce central notions used throughout this deliverable. 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 [Nak08, BMC15, TS16, JSKW17].

2.1 Address Clustering Heuristics

There are several address clustering heuristic in use today, the most effective one being the multi-input heuristic [MPJ13, Nic15]. Its underlying intuition, which is illustrated in Figure 1: Multiple Input Clustering Heuristics, 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.

![Figure 1: Multiple Input Clustering Heuristics](source_not_found)

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 [MB16] where \( n \) parties produce a special joined transaction, as shown in Figure 2: Error! Reference source not found.. 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 this transactions schemes pose a problem because the clustering algorithm would combine all \( n \) input addresses and their respective clusters into one entity. Meaning we potentially merged \( n \) potentially independent parties into one.
A number of studies attempted to quantify the effectiveness of clustering techniques: Nick [Nic15] measured the accuracy of different clustering algorithms using a ground-truth dataset consisting of 37,585 user wallets, which he obtained via a weakness in the BitcoinJ light client implementation. His results showed that on average 69.34% of the addresses could be linked using only the multi-input heuristics. Harrigan et al. [HF16] 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. Quantifying the reliability of clustering results is not easy since no ground truth data is available. Furthermore, most clustering heuristics make assumptions about the behavior of participants. Obviously, user behavior can change.

2.2 Attribution Tags

Tagging is a collaborative process in which a user adds (mostly textual) labels (tags) to shared content. It does not rely on static, predefined taxonomic structures but on dynamic, community-driven linguistic terms and conceptions [GH06]. 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 Figure 3, a tag could for example attribute a given Bitcoin address to some real-world actor (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 [GH06]. If, in the context of cryptocurrencies, a
Semantic ambiguity of tags can also be exploited to find irregularities in address clustering: Ermilov et al. [EPY17] 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 i.e. a cluster is unlikely to be.

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.

### 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 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 [GCG10].

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. [CW03, BSHW06, BKWC01, Kar09]). Data provenance is also discussed in the field of scientific data sharing and processing to support data protection (e.g., GDPR), data ethics [HFPL] and tracking the lineage (origin and subsequent processing history) of scientific data sets [BF05]. For a more general overview on provenance management for computational tasks in various domains we refer to the survey by Freire et al. [FSKS08].

In forensic investigations, provenance information is recorded in order to ensure that collected evidence can be accepted as truthful by 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 [Gio11]. Modern forensic software can automate much of the manual work needed to produce audit trails and provide stronger guarantees by using existing digital infrastructure and techniques such as user management, digital signatures, or even blockchains [Sto17] for creating 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 [Pol07]. More recent research by Cosic and Miroslav [CB10] 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). Those controls can be implemented through a database to record 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 uttermost 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 Empirical Analysis

After having introduced clustering heuristics, attribution tags as well as provenance tracking as the key ingredients to cryptocurrency investigations, we now move on and empirically analyze how certain types of transactions could distort clustering results, unify entities that have no association in the real-world, or lead to the formation of super-clusters. In the following we only consider the multi-input heuristic as well as coinjoins.

Coinjoin transactions pose a serious false positive risk to naïve multi-input address clustering in Bitcoin and other cryptocurrencies. Coinjoins provide additional anonymity to the participants by making transactions unlinkable. Fortunately, coinjoins are in their standard form detectable by simple heuristics. Goldfeder et al.[GKRN18] describe and evaluate a coinjoin identification heuristic. BlockSci [KGC17] implements two versions of the heuristic. The full version as describe by Goldfeder et al. is NP-Complete and therefore slow on larger inputs and a solely structural approach avoiding much of the computational effort.

We use the coinjoin heuristics to evaluate how coinjoins influence the results of naïve multi-input address clustering. We investigate the 100 largest clusters found in Bitcoin produced by the naïve multi-input address clustering heuristic. We compare how many of the addresses of one cluster were also part of a potential coinjoin transaction. The resulting ratio \( \frac{\text{#AddressesCoinjoin}}{\text{#AddressesCluster}} \) gives indication of the false positive rate of a cluster.

We looked at all Bitcoin transaction from its inception until April 30th, 2018. Our dataset contains 40,049,947 clusters created using the naïve multi-input address clustering algorithm as described in [MPJ13]. We found around 2% or 724863 clusters have at least one address that was involved in a transaction that was flagged as coinjoin by one of the two coinjoin heuristics.

This result shows that naïve multi-input address clustering is significantly biased by coinjoin transactions. Clustering for law enforcement purposes should always consider coinjoins as a special case employing heuristics described in [GKRN18] either to proactively exclude potential coinjoin transactions from clustering or by at least marking clusters that contain coinjoins for manual human intervention.

Obviously this is not a complete picture, there are other heuristics that are used to cluster addresses that are used in practice such as the change heuristic [MPJ13] that also suffers from false positives. Clustering heuristics are rule of thumbs, that rely on assumptions that where reasonable when they were created. To evaluate the reliability of such heuristics we need to get data and need to reiterate whenever behavior changes.
4 The Legal Perspective

Usage of address clustering and attribution tags raises numerous yet unresolved legal questions that we address in this section. Questions mainly arise from the law of substantiation of evidence and suspicion and fall into two sub-categories:

First, the main goal of any Law Enforcement Agency that will use address clustering and cryptocurrency attribution tags is to create relevant court-proof evidence or at least reasonable suspicion as a basis for further investigations. Evidence is basically relevant if it makes the matter that requires proof more or less probable [73, AC 729]. Therefore, the data used in a criminal trial or as a basis for substantiation of a suspicion must be admissible and have an evidential value that is as high as possible. The admissibility of evidence requires at least that the gathering, processing and usage of the evidence was/is lawful under the respective domestic and/or international criminal procedure and data protection law. The evidential value of data is determined through the contained information and the quality of the analytical methods to process the data, as well as the authenticity and integrity of both. Hence, our model must meet the legal rules and standards for gaining, processing and presenting digital evidence. To date, there are no internationally valid rules for the tread 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 minimum standards proposed below were obtained from a summary of the following sources: a) written law for evidence in criminal procedures (e.g., the federal rules of evidence of the USA) [USCoA23, Cou93, fDoM07], b) case law of several supreme courts of several countries (e.g. USA and Germany), c) recommendations of international groups of forensic experts (like the SWGDE).

Second, most legal systems give the accused a right to inspect the records as a part of the guarantee of fair trial or oblige the public prosecutor’s office to disclose their evidence. Hence, the documentation of the process of clustering and tagging must meet the requirements of criminal trial record keeping.

In addition, with a general increase information exchange between LEAs the sharing of clustering data and attribution tags poses new questions in the area of data protection which are also addressed in this section.

4.1 Lawfulness of Data Processing

Data processing must be in compliance with the legal framework it takes place in. For criminal investigations the respective criminal procedure law and data protection laws set up 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 extent/scope of the processing.

Clustering and attribution techniques are specifically used to identify natural persons/suspects. The used data consequently relates to identifiable natural persons and is hence deemed personal data (c.f. [EC16a, EC16b, CSS18]) that is protected on international [EP00, oE50, EC16a, EC16b] and national [itSNYotRoI18, Par18, GP18] level as well as in subject specific laws [oST16, tUSC14, tUSC99]. As the scope of protection differs, the data protection principles in this paper are derived from the European data protection framework [EC16a, EC16b] which arguably acts as a role model internationally (c.f.[itSNYotRoI18, CSS18, Alb]) and that contains relatively high standards which help fulfilling the evidence requirements described in the following Sections.
The processing can be roughly split into (1) the *collection* of data and (2) the *subsequent processing*. Transaction data is gathered from publicly available Blockchains, while attribution data can derive from public and non-public sources. While legal implications of processing publicly available data for law enforcement purposes are still subject to ongoing discussion, both steps can arguably be based on general clauses to a certain extent. In addition, data obtained through existing law enforcement communication channels (e.g. [SIS, Error! Reference source not found., Error! Reference source not found., SIE]) can usually be seen as lawfully due to the legal framework and the safeguards included in these systems (c.f. [Ass16a, otEU16a, Ass16b, otEU06]).

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. The transparency requirement may require the investigator/prosecutor to explain the processing to a certain extent either to a data subject or the data protection authorities. Having said that, most legislations limit the transparency requirement to ensure an effective law enforcement.

While automatic decisions are generally prohibited, decisions such as ordering further investigative measures, can still be based on results of the automated clustering of cryptocurrency addresses and other analytics techniques if they not merely formalistic and all relevant aspects of the individual case are carefully taken into account. (c.f. Recital 38, Art. 11 [EC16b]). 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. Moreover, the data has to be accurate (c.f Art. 4 (1) c [EC16b], Art. 28 (1) d [otEU16b], [oMCoE87]). Since only facts can be inaccurate, data protection law does not prohibit the processing of data based on estimations or probabilistic measures. However, the principle of accuracy requires the clear distinction of facts and probabilistic or estimated results such as address clusters (Art. 7 [EC16b], Art. 29 [otEU16b], [oMCoE87]). To assess the nature and reliability of the data it is hence necessary to have sufficient meta-information 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 before. Data protection law usually requires the data controller to rectify, or in some cases, erase false data. Having said that, erasure bears the risk of automatic reproduction of false positives in some cases. Where this is the case, data should be marked to restrict it from further processing (c.f. Art. 16 (3) [EC16b]). If data has been shared, receivers of the false/outdated data have to be informed.

### 4.2 Authenticity and integrity of data and attributed information/chain of custody

The authenticity of data correlates with the probative value of information in criminal proceedings but is also often required for data protection purposes. Authenticity must be ensured in all forensic steps and procedures that involved processing of electronically stored information [ftDoM07, CCoA02, Con14], which makes comprehensive and precise documentation of data sources, tools, and applied techniques necessary. To allow an assessment of the evidential value in trials, the analysis procedure outcomes and limitations have to be explainable in a comprehensible manner. It must be ensured that data has not been altered, which can be achieved by using 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 [ftDoM07, at 546]). In this regard, it might be helpful to attach “certainly-values” to data as proposed by Casey [Cas04, p.176]. Similarly, data protection law, particularly the LED (Articles 4 para 4, 19 Article 25 para 1, recital (57), requires the data processor to be able to demonstrate compliance with data protection law to ensure accountability. 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 source data. This means at least the name of the source; time of access and liability of the source must be recorded as provenance information.

Forensic tools typically operate on-top of a specific cryptocurrency blockchain that already offers a number of features that can be recorded in order to provide authenticity and integrity: by recording the hash of the most recent block considered by the tool, one can directly relate the result of an investigation to a data structure that provides authenticity (signed transactions) and integrity proofs (hashes). As a blockchain can practically not be altered, it is sufficient to record the currency code (e.g., BTC, ETH) and the most recent blockhash as identifiers within the chain of custody.

Ensuring the authenticity and integrity of clustering techniques 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 blockchain state does is not necessarily equal to a cluster generated at a later blockchain 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 can be ensured by relating a tag to its source, its creator and generation procedures and computing a digital signature over the tag and all contextual data points. Tag integrity can be achieved by computing a hash and including that hash in the signature.

4.3 Reliability

The reliability strongly correlates with the weight of evidence in a criminal procedure. [Neu, Goo50] Hence, evidence must be based on precise and scientifically proven methods without any mistakes when collecting data, clustering addresses and attributing addresses by tags [oPPD13, Con14]. The reliability hence has to be proven for the scientific methods as well as the correct use of them. Proof can be brought forward by describing results of testing the process, the logic of the process, or by having an expert testimony [Con14, Cou93, oAftEC99]. A combination of these approaches generally increases the reliability of evidence in court. 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.

4.3.1 Clustering Reliability

Although proving the reliability of algorithms is subject to an ongoing discussion [Keh17], 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).
Assessing the overall effectiveness of a formalized heuristic could be achieved by testing it against some collected and verified ground truth. In the case of clustering heuristics, a ground truth could be a set of known cryptocurrency wallets, each carrying a set of addresses belonging to the same real-world user. However, ground-truths are currently constructed ad-hoc, often for scientific purposes. Making available a general, authoritative standard ground-truth dataset would allow the quantification of clustering effectiveness and provide specific reliability measures and probabilities, as in other forensic methods (e.g., DNA testing).

The reliability of a clustering algorithm within the scope of a particular cryptocurrency can be tested by following standard functional testing procedures, which are now widely known in software development. A function implementing the clustering heuristics, can be tested by feeding them a set of example transactions and examining the output without considering the internal structure and source code (black-box testing).

4.3.2 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 forensics tool. If, for instance, an investigator identifies a cryptocurrency donation address on a known Web site and assigns a tag to that address (e.g., “Internet Archive”) 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).

4.3.3 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. 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 [otEU16c, oMCoE87]. The degree of reliability correlates with the evidential value of information and is hence of relevance for the (free) consideration of evidence in trials and in pre-trial stages to establish/justify necessary degrees of suspicion.

4.4 Qualification

Investigators and experts who obtain and analyze electronic and digital evidence in criminal proceedings must be qualified to use the respective methods of obtaining and analyzing evidence [MAC03, Hei15] [Cas04, p. 170]. Lack of qualification in dealing with IT forensic methods can have considerable influence on the investigations and the 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 experts increases the probability that wrong conclusions will be drawn from the available evidence and, in the worst case, that the public prosecutor's office and/or the court will make their decisions on the basis of false facts [Hei15]. 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 [Hei15].

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 their authenticity and reliability.

4.5 Verifiability
The method of collecting data and gaining information must be repeatable and reproducible [Mar15, Hei15, Cas04]. This is the only way to ensure that the lawyers involved are able to follow up the acquisition of information in subsequent legal proceedings [Hei15]. With regard to limitations to disclosure and right to inspect the records, verifiability becomes even more important and must be ensured for the individual case.

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” [oSN01, Mar15]. To account for individual errors of the investigators, the formal review of a process may accompany a peer-review with different tools [Cas04, p.184].

Verifiability of cryptocurrency investigation results obtained from forensic tools can be achieved by applying the same method that already provide 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, which can be identified by respective block hash.

4.6 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. 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. The necessary quality of the facts and the necessary probability of committing the offence increase with the intensity of the investigation measures applied on the basis of the suspicion (e.g. search and seizure, detention on remand). For a conviction, the factual basis must be of the highest quality and the probability of perpetration must be highest.

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. What the systems have in common, however, is that the various stages of suspicion are described with normative terms. The descriptions of the suspicion reflect a certain necessary quality of the facts and
a necessary level of probability of the offence being committed that can be derived from the facts. Examples (from the German and US Codes of Criminal Procedure) include the "simple" suspicion to start an investigation, the reasonable suspicion for special investigation measures (e.g. seizure), sufficient suspicion to bring charges against the suspect, probable cause for arrests and warrants, and, last but not least, the "beyond all reasonable doubts" standard to convict the suspect.

It is difficult to harmonize the normative standard of evidence in criminal proceedings with statistical results of data analysis procedures. The criminal lawyers involved in the proceedings must be enabled to subsume the results of technical investigations under the normative concepts. Therefore, the relevant information must be presented to the criminal lawyers in a form and language that they can understand. 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, and alternative hypotheses must be presented and, if necessary, comprehensibly excluded. This requires a thorough understanding of the data analysis method used (see subsection "Qualification").

Finally, it is very important that the sources of information and data are reliable and traceable, when using address clustering and annotation tagging (see subsection "Reliability"). 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.

4.7 Right to Inspect the Records/Disclosure of Evidence

The right of the accused or his defense counsel to inspect the evidence gathered by the police and the public prosecutor’s office is a key element of constitutional criminal proceedings. In inquisitorial criminal procedure systems such as the German criminal procedure, this right is designed as the right to inspect records. In contradictory criminal procedure systems, such as the US-American one, on the other hand, the right to disclosure of evidence by the public prosecutor’s office exists and can include source code (e.g. UK: [96, Section 3] [Off13] [Ser18], US: [Con14, Rule 705] [Con44, Rule 16] [otUS63a, otUS63b] [Bro17, p.148] [Wil11, p.127] [Sho10]). 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 him is made transparent in order to verify that the requirements laid down in the previous Sections are met. On the other hand, law enforcement agencies have an interest in ensuring that the precise functioning of data analysis tools does not become widely known in criminal communities. This could make their use more problematic or impossible. And finally, the proprietary rights of the companies that produce and distribute the software tools used must be taken into account. [Cas04] [JRWP, CC07] In addition, a complete disclosure of the code would make it much easier to reproduce the tools. Moreover, the validation of the source code alone does not account individual errors of the investigator. [Cas04, p.184] The highly complex question of balancing arises in both systems and cannot yet be conclusively answered. A first idea could be not to include the source code of the software used in the files, but a function and usage description as exact as possible. With regard to the clustering software, this mainly concerns the heuristics, their accuracy or the degree of probability of the results but 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. [Ser18, p. 87 ff.]

4.8 Summary of Key Requirements

- **Address Clustering and Annotation Tagging** 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.
- To guarantee **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.
- 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. sets of addresses from known 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.
  - 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.
- 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, on the possible effects of adding an attribution tag to a certain address and have developed an understanding of the attached provenance information.
- **Repeatability and reproducibility** of address clustering and annotation tagging can be achieved by the same measures that guarantee the authenticity and integrity of data.
- 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.
- To meet the requirements of the *right to inspect the records/the principle of 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.

5 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 and effectiveness needs of law enforcement agencies and, on the other hand, provides court-proof evidential value to forensic investigations and adheres to current data protection legislation. In the following, we focus on sharing of attribution tags among law enforcement agencies, in a way that can, according to our analysis, be accepted as truthful by courts.

Previously, we already emphasized the key role clustering heuristics and attribution tags in cryptocurrency investigations: a single tag can deanonymize a cryptocurrency addresses 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, we first propose a lightweight data model for sharing attribution tags that should effectively balance those opposing goals. Then, we suggest a model for sharing address clusters.

5.1 Attribution Tag Sharing

Our proposed data model combines two lines of related work: first, it builds on the generic PROV standard, which was released by the W3C Provenance Working Group in April 2013 [BBC12]. That model synthesizes a long line of research and has been designed as an abstraction of domain or application specific provenance models. On a generic level it expresses the \textit{process of generating entities with activities by involved agents} and defines the three core concepts \textit{Entity}, \textit{Activity}, and \textit{Agent} as well as the relations between them. Those concepts can be refined to the purpose of an attribution tag sharing model as follows:

- **Tag** (specialization of \textit{PROV Entity}): a tag is the core concept one wants to describe provenance about. A tag can refer to some digital, physical, or purely conceptual thing and carries a unique identifier (e.g., http://exampler tool.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 name (e.g., “Internet Archive”) and can be categorized along several dimensions (see below).
- **Activity** (specializations of \textit{PROV Activity}): each tag has been generated by some activity, which started and ended at some point in time. An activity should either carry a human-readable name (e.g., "Manual Entry") fall into one or more predefined categories (see below).
- **Agent**: represents a person or machine who is responsible for an activity which is, in this case, creation or modification of tags. An agent carries a human-readable label (e.g., John Doe).

Second, the data model also considers the key legal requirements, which were defined in the previous Section, by translating them into corresponding data model fields:
• **Hashcode**: 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?).

• **Timestamps**: automated and trusted time stamping routines should record when an activity was performed on a tag.

• **Source**: each tag has been extracted from some digital or non-Digital source. Each source carries a human-readable label (e.g., Internet Archive Website) and possible some URI referring to that source (e.g., https://archive.org).

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 namespaces expressed as Internationalized Resource Identifiers (IRI). All previously introduced concepts, attributes and relations could, for instance, carry the namespace (http://titanium-project.eu/vocab/), which, for convenience reasons, can be mapped to a prefix such as titanium. Specific examples are http://titanium-project.eu/vocab/Tag and titanium:Tag, which both refer to the same concept definition expressed above. Figure Error! Reference source not found. shows the conceptual entities and relations of the proposed attribution tag data sharing model. The following listing shows an example attribution tag expressed in JSON-LD.

```json
{"@context": {
  "@vocab": "http://titanium-project.eu/vocab/",
  "category": "http://titanium-project.eu/categories/"
}
```

**Figure 4: Attribution Tag Sharing Data Model**: a data model expressing main conceptual entities for describing attribution tags in a court-proof fashion.
5.2 Categorization Schemes

Assigning uniquely identifiable categories to the main conceptual entities of the tag sharing model is key for automated data processing and algorithmic decision making. Some cryptocurrency analytics tool 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 within the law enforcement domain. However, as a starting point, we suggest considering at least consider 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. Example services are: Exchange, Wallet Provider, Miner, Marketplace, etc.

**Agent categories**: distinguishing between Person and Organization is a common refinement (cf., 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, a Device, etc.

**Activity categories**: provide information on the type of activity a tag was generated by. Common activity types are ManualEntry, Crawl, etc.

### 5.3 Implementation Considerations

Vocabularies and categorization schemes could be published on the Web my making sure that all terms (e.g., http://titanium-project.eu/vocab/address) and concepts (e.g., http://titanium-project.eu/categories/ManualEntry) carry dereferencable IRIs. This allow clients to browse available terms and categories online and to automatically verify attribution tag categories before exchanging them with others. A simple, straightforward way is to follow the implementation of schema.org, which is a generic schema for structured data on the Internet. More advanced implementations could follow the Linked Data principles [BHB09].

Our proposed data sharing model suggests that authenticity and integrity of attribution tags can be achieved by computing hashes and digital signatures over a tag and all its contextually relevant attributes. However, for a data exchange purpose, this would require a precise and agreed-upon definition of the hash computation and digital signature procedures, which could be defined as part of a larger standardization effort. 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 believe that it could also be used for sharing JSON-LD serializations of attribution tags.

### 5.4 Address Cluster Sharing

Our proposed model for sharing address clusters, which is shown in Figure 5, builds on the previously introduced attribution tag sharing model and introduces the following conceptual entities, attributes, and relationships:
• **Cluster**: a cluster 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.

• **Agent**: denotes the creator of a cluster. 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.

• **Activity**: describes the activity 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.

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

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

```json
{"@context": {
    "@vocab": "http://titanium-project.eu/vocab/",
    "category": "http://titanium-project.eu/categories/"
}
```
6 Conclusions

This deliverable focuses on cryptocurrency analytics has become a key forensic method in law enforcement and is nowadays used to investigate a wide spectrum of criminal activities. Despite its wide adoption, the evidential value of obtained findings in court is still unclear. In this deliverable, we
focused on the key ingredients of modern cryptocurrency analytic techniques, which are clustering heuristics and attribution tags, and discussed them in the light of internationally accepted legal standards and rules for substantiating suspicions and providing evidence in court. We translated derived key requirements into a data sharing framework that builds on existing standards and strengthens the evidential value of forensic cryptocurrency investigations. We believe that the implementation of our proposed framework could increase the efficiency and effectiveness of forensic cryptocurrency investigations, while safeguarding their evidential value.

7 References


[Alb] Jan Philipp Albrecht. How the gdpr will change the world.


[CCoA02] Div. V. Colorado Court of Appeals. The people of the state of colorado v. daniel huehn, 2002.
[HF16] Martin Harrigan and Christoph Fretter. The unreasonable effectiveness of address clustering. In Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and

[HFLP] Ensar Hadziselimovic, Kaniz Fatema, Harshvardhan Pandit, and David Lewis. Linked data contracts to support data protection and data ethics in the sharing of scientific data.


[JRWMP] Court of Appeals Jdg. Randolph W. Peterson. State v. underdahl (consolidated opinion). [...]"district court abused its discretion in ordering discovery because respondents failed to show that the source code is relevant and in the state’s possession or control.".


[Neu] David; Jackson Graham; Reyna Valerie; Ranadive Anjali Neumann, Cedric; Kaye. Presenting quantitative and qualitative information on forensic science evidence in the courtroom.


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.


Siena - secure information exchange network application (europol).


