Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

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

Reliability validation enabling framework (RVEF) for digital forensics in criminal investigations

Radina Stoykova [a], [*], Katrin Franke [b]

[a] University of Groningen, the Netherlands
[b] Norwegian University of Science and Technology, Gjøvik, Norway

ARTICLE INFO

Article history:
Received 22 November 2020
Received in revised form
8 December 2022
Accepted 10 March 2023
Available online 12 May 2023

Keywords:
Digital forensics
Digital evidence
Validation
Reliability
Reproducibility
Daubert
Forensic science
Fair trial
Criminal investigation

ABSTRACT

This paper proposes a formal *reliability validation enabling framework (RVEF)* for evaluation of digital forensics in criminal investigations. The RVEF is informed by examined theoretical and conceptual gaps between law and digital forensics related to reliability and validation. Identified are validation criteria and validation testing techniques for digital forensics as well as their limitations and challenges.

The proposed RVEF aims to satisfy the objective for documenting the chain of evidence and custody as standard process. It is a generic and extensible approach to create a formal procedure for documentation of reliability information at three levels: technology, method, and application. For each level reliability criteria are compared against international digital forensic standards, guidelines, and best practices in order to elaborate concrete minimum documentation requirements necessary to enable reliability validation by law enforcement. The framework aims to increase accountability, reliability testing, and machine-human error mitigation in digital forensics. It can also serve judges and defense lawyers to cross-examine the forensic report in a formalized process, access the proportionality of the investigation measures, and potential risks from the inappropriate use of technology.

1. Introduction: Challenges with reliability in digital forensics

A reliability crisis in digital forensics (DF) for criminal investigations is emphasized by Interpol (Reedy, 2020) and in the UK National digital forensics strategy [(The UK National Police Chiefs Council, 2020), p. 21]. Academics and practitioners also stated the lack of reliability validation in digital forensics (Hughes and Karabiyik, 2020; Horsman, 2018a; Casey, 2019; Jones and Vidalis, 2019). Standardization and governmental bodies commented on the reproducibility crisis in the field (PCAST, 2016) (Council of the European Union, 2016). Legal scholars proposed expert accreditation in digital forensics (Henseler and van Loenhout, 2018), (Kwakman et al., 2011) and discussed the absence of clear legal requirements for digital evidence reliability assessment (Risinger, 2018; Edmond, 2012; Sommer, 2010; Saks and Koehler, 2005; Risinger, 2000). The fast technological advancements in computations to assist forensic sciences makes a lot of existing validation and reproducibility studies outdated (Kloosterman et al., 2015; Horsman, 2019a; Tully et al., 2020), challenges testing in digital forensics (Garfinkel et al., 2009), and the subsequent court evaluation of digital evidence (Saks and Faigman, 2008). A comprehensive *reliability challenges taxonomy* identified underdeveloped validation procedures and limitations with testing datasets, tools, methods, and examiner work (Stoykova, 2021).

The identified lack of quality assurance and accountability mechanisms in digital forensics might lead one to expect more strict evaluation of digital evidence reliability in the criminal justice system.

To the contrary, judges "seem to be enthusiastic to rapidly embrace the products of technological progress" (Edmond and Roberts, 2011) and often assume that the digital media source of evidence is "working properly".[(Mason and Seng, 2017), Para. 6.198]. Similarly, "law enforcement and prosecuting authorities are often willing to use novel science and technology" in order to secure evidence.[(Doyle, 2019), Ch. 7]. The enhanced use of automated tools to acquire and analyze digital evidence creates the false perception that technology mitigates errors and bias, and that results from tools are reliable and trustworthy. A phenomenon described as *technological protection fallacy* (Dror, 2020).

[*] Corresponding author. University of Groningen, the Netherlands.
E-mail addresses: r.stoykova@rug.nl (R. Stoykova), katrin.franke@ntnu.no (K. Franke).

https://doi.org/10.1016/j.fsidi.2023.301554

Digital forensic suffers from proliferation of standards, which are short-lived and insufficient to reach general acceptance in the domain. For example, there is clear consensus in the DF community that following a process model is an obligatory requirement in order to meet legal and scientific objectives. However, currently over 60 different process models are proposed (Årnes, 2018). By contrast, implementable solutions for reliability validation or practical validation process models are rarely discussed in digital forensics which raises concerns for dissemination of unreliable knowledge.

In order to overcome those challenges a first step is to develop a generic framework and standard process for reliability validation of digital forensics for criminal proceedings.

This paper proposes a reliability validation enabling framework (RVEF) under which tools, methods, and examiner work can be documented for cross-examination. The framework can guide and support different validation techniques by defining a formal validation process and deriving minimum documentation requirements for law enforcement purposes. RVEF aims to overcome the limitations of current testing of tools, methods, and examiners done in isolation by linking together different validation specifications relevant to the forensic task and mapping information on the validation process as a whole.

The paper is organized as follows: *Section 2* outlines that legal and forensic science concepts related to reliability often defer, overlap, or complement each other which requires their clarification in order to ensure a validation process that satisfies both domains. Those aspects of the problem define the scope and the elements of scientific validation in digital forensics and are complemented with literature review on validation, chain of custody, and chain of evidence requirements. *Section 3* discusses specific limitations with reliability validation procedures in DF. *Section 4* defines the proposed generic RVEF as a process and clarifies validation criteria and specific requirements at technology, method, and application level. *Section 5* explains how RVEF can satisfy techno-legal objectives, while its possible limitations and improvement in future work are discussed in *Section 6*.

2. Reliability validation: Techno-legal analysis

This section introduces a techno-legal understanding of key concepts related to digital evidence reliability.

First examined are the concepts of evidence admissibility, probative value, and relevance as well as proportionality of investigative measure to show that they all depend on clear reliability procedures. Then we clarify the meaning of reliability and validation in law and in science to identify what is the reliability standard for digital evidence in criminal proceedings and what is required in digital forensics to meet such standard. This will give a theoretical background for the proposed RVEF.

2.1. Admissibility and probative value

The admissibility and probative value of the evidence are evaluated by the judge at a trial. However, to strengthen the judicial evaluation of the facts, the investigation procedure must guarantee at least minimum quality of the evidence and preliminary confirmation that it is admissible, probative, and legally obtained. This requires upholding to criminal procedure and standards for quality of the investigation which depend on the jurisdiction of each country. However, demonstrating reliability of digital evidence is a factor which influences both admissibility and probative value, and does not depend on the jurisdiction. This means that it can be internationally standardized.

2.2. Relevance in law and in science

Relevance in court is a function of two tasks: “the arguments made by lawyers and judges, as well as *examining the inherent quality of a piece of information*” [(Roberts and Zuckerman, 2010), p. 104]. This second part of the examination is routinely skipped in courts since incriminating evidence relevance is taken for granted (Edmond and Roberts, 2011). Consequently, digital forensic examiners when testifying or submitting report to the judge, are challenged with the objective to advance legal arguments about the case, and rarely on the relevance and reliability of the information. Testing the reliability of the information must be done in the digital forensic process, since the courts are ill-equipped to do so and have different objectives than thoroughly validation of forensic findings.

2.3. Reliability in law and in science

Reliability in law is a different threshold than scientific reliability because the legal evaluation of expert evidence has different objectives. In law reliability is related to trustfulness and weight of the forensic evidence in the concrete case. Judges evaluate the “forensic reliability” of the evidence against its overall probative weight and relevance to the case. Digital forensics examiner is concerned with accurate fact-finding only, while judges evaluate if the evidence is reliable but also its probative value to the case including if it is legally obtained. The development and administration of standards for scientific validity is part of digital forensics and must be done before the digital evidence is presented in court. Judge's role should be only to verify and enforce upholding to the standards.

In forensic science reliability is a property of process related to a consistent intended behavior and results (ISO/IEC 27037:2012). A test that produces the same results on successive applications is said to be *reliable* (Gross and Mnookin, 2003). The primary role of a forensic scientist is to provide guidance about the reliability of different kinds of evidence, and to develop methods and devices for increasing such reliability (Risinger, 2000).

However, currently there is no clear legal standard for digital evidence reliability. Relevant digital evidence must be processed in forensically sound manner in order to be admitted as probative. The Council of EU interpreted that *sound* digital forensics procedures must reflect the “state of art of science and technology” (Council of the European Union, 2011). It is considered that a process or a method are forensically sound if they adhere to established digital forensics principles, standards, and processes [(Årnes, 2018), p. 13]. The issue is that often in digital forensics certain reliability standards are rethought and improved in short periods of time base on technology advancement. For example, the *ACPO* guidelines, that were one of the first standards for digital forensics, stated that data on the original source must never be changed by the forensic procedure (ACPO, 2012). It was later realized that in certain cases (e.g. live acquisitions, encryption on the device or remote acquisitions) complying with this principle was impossible from a technical point of view. Therefore, the standard was subjected to reevaluation and current requirements suggest that changes to the system must be limited to the minimum and accountable, while any interference with evidence data must be justifiable (ISO/IEC 27037:2012; Adams et al., 2013). Another principle, that the data must be acquired in lowest level of abstraction, was disproved in mobile forensics where encryption and security features render the physical acquisition of data unreadable for tools or examiners.

Consequently, reliability testing during the forensic examination and according to a standardized procedure is a pre-requisite for the subsequent in court verification of the reliability of the evidence resulted from this process. According to Horsman there are three types of validation — following previous case work, existing published works, or validation via testing (Horsman, 2019b). Only validation via testing, however, meets the legal requirements for a scientific rigor.

2.4. Validation in law and in science

The Forensic science regulator in UK (FSR, 2020) guideline states that when methods or tools are novel, in-house developed scripts, or documentation of a formal validation is missing, they need to comply with software engineering verification and validation testing [(FSR, 2020), Pt. 6.2.2—6.2.4].

According to IEEE Std 1012—1998 validation is the assurance that a product, service, or system meets the needs of the customer and other identified stakeholders. It often involves acceptance and suitability with external customers. Verification is the evaluation of whether or not a product, service, or system complies with a regulation, requirement, specification, or imposed condition. Notably, SWGDE emphasise that “software testing can never prove that a tool is functioning correctly [however …] testing can lead to confidence that the tool is unlikely to fail within the situations for which it has been tested.”

Hereafter, we adopt the generic definition that validation is the *scientific methodology for demonstrating the accuracy and reliability of a process* (Hughes and Karabiyik, 2020) (Gross and Mnookin, 2003). This definition is consistent with forensic sciences and avoids requirements like “fitness for purpose”, “general acceptance” or “current state of technology” (FSR, 2020; ENFSI, 2015a; Guo et al., 2009), which are lowering the threshold for testing in digital forensics and introduce a vague, unclear legal standard. Thus, the identified need is for a generic independent validation process in digital forensics.

2.5. Documentation as a legal requirement

Documenting forensic methodology is crucial given the dependencies in digital forensic actions. Both chain of custody and chain of evidence need to be preserved in digital forensics investigations in order for any party to the criminal proceedings to be able to establish the accuracy and reliability of the digital evidence.

2.5.1. Chain of custody

The chain of custody record is a document identifying the chronology of the movement and handling of digital artefacts (ISO/IEC 27037:2012). Such a record must be established for each piece of electronic evidence (Interpol, 2019). Most forensic labs standards require a record of the control and quality of forensic procedures (ISO/IEC 17025:2017), which can be examined by both judge and defense lawyers. This is a requirement for the accessibility of the chain of custody.

In the legal context, the term chain of custody is established in the US legal tradition, while in England and Wales “continuity of evidence” covers the same concept. The chain of custody “begins prior to collection and ends when evidence is released to the owner or destroyed.” (ISO/IEC 27037:2012)

Although the chain of custody includes all phases of the digital forensic process and must be preserved together with the original data, its role during evidence identification and collection is crucial: as Roger argues, if “doubt is cast on the initial collection and management of evidence, output from the other phases is moot” (Rogers et al., 2006).

Any break in the chain of custody “can lead to questions about the validity of the evidence”[(Daniel and Daniel, 2012), p. 12], but could also be accounted for in further investigation activities.

In digital forensics, the chain of custody is an overarching principle since validation and reproducibility testing are impossible without proper documentation. In law, however, the chain of custody is just one of many ways to prove due diligence, and requirements vary strongly between different jurisdictions. Therefore, in most continental jurisdictions, although not explicitly, rules on chain of custody exist, but there are hard to interpret in digital context. For example, Norwegian criminal procedure law[1] requires a description of nature and purpose of the search, where “all objects seized shall be accurately recorded and marked in such a way as to avoid confusion”. Same broad and vague requirements for expert report are open to interpretation in how to document digital searches, seizures, and examination of data. In the absence of clear requirements for digital chain of custody and formalized protocols for each type of evidence, many forensic actions are poorly or not at all documented. Arguably, in the digital evidence domain a break in the chain of custody can be potentially compensated by testimony, but a complete absence of it cannot as there is no possibility to audit the stages of the processing and therefore the reliability of digital evidence or its compliance with a fair trial. This argument is even more important in relation to chain of evidence documentation.

2.5.2. Chain of evidence

Often chain of custody and chain of evidence are used interchangeably, but it could be argued that they are the results of different procedures, have different objectives and human rights impacts, and are related to separate evaluation.

Chain of evidence documents the digital artefact interpretation of relevance to the concrete criminal case and enables classification, reconstruction, and examination of digital events. The actual digital data which has evidential value and is related to the crime is just a small piece of the data originally collected and preserved. However, before the examination and analysis phase, the potential relevance and probative value of data seized as evidence cannot be discovered. Therefore, the chain of evidence is constructed only when all the digital evidence processing steps of acquisition, examination and analysis are completed. Moreover, as the stages are interdependent and repetative, if a certain stage is not documented there is no possibility to trace the origin of digital evidence and the forensic actions performed.

The output of the acquisition, examination, and analysis should be chained in a digital forensics report. Although it is simple for a practitioner to obtain data using forensic tools, the validation that it has been correctly obtained and the interpretation of the underlying data structures is of far greater importance, because these reveal the origin of the digital artefacts and are therefore crucial for the attribution and individualization of the digital artefacts and events to concreate suspects, locations, and timelines.

The term evidence chain is not used or established as a legal requirement. However, Schum refers to “intellectual audit trials” which explain “what questions were asked at what times, what possibilities were being entertained at various stages of an investigation, and what was the existing evidential base for entertaining these possibilities at various times.” (Schum, 2001) Also, in forensics Carrier describes “standard trace detectors for efficient trace gathering that can be compiled into event chains to support different hypothesis about the case.” (Carrier, 2006) The chain of evidence is part of and based on the chain of custody. The principle of availability requires suspects and defendants to have access to the chain of evidence and chain of custody in order to be able to cross-examine and challenge the evidence on valid grounds. Full

[1] See Norwegian Criminal Procedure Act (NCPA) section 207 §1, NCPA section 197 § 3 in conjunction with NCPA section 153 § 3.

chain of evidence and chain of custody ensures that the work of the DF examiners is limited to what was authorized, that only competent personnel had access to the data, no data is randomly omitted or destroyed, and according to the principle of proportionality the investigation interference with human rights was reduced to a minimum.

Documentation and chain of custody are often referenced as an obligatory requirement in digital forensics(Montasari, 2016; Beebe and Clark, 2004; Kohn et al., 2013) but not yet prominent as a legal requirement, while chain of evidence is not so much discussed with the exception of the theoretical works cited above. Although there is a consensus on the importance of documentation, the lack of a legal requirement for chain of custody and chain of evidence in digital forensics investigations is considered as a major drawback in the development of reliability validation procedures.

2.6. Proportionality

Proportionality of the investigative measure is a principle in law which receives little attention in digital forensics. Proportionality requires that the methods used to gather the evidence must be fair and proportionate to the interests of justice: the prejudice (i.e. the level of intrusion or coercion) caused to the rights of any party should not outweigh the probative value of the evidence (Interpol, 2019). Proportionality analysis oppose the estimated relevance and probative value of the evidence against the intrusiveness and resource demand of the method. It also requires LEA to choose the least intrusive in respect to human rights investigative measure. As argued, human rights objectives “may prevail when the public interest can be attained with a less restrictive measure, but they may be curtailed when the measure seems proportional to the objective” (Alendal et al., 2021). Although legally and internationally recognized as a principle, proportionality is routinely criticized for its difficulty to be implemented and enforced in practice (Bart van der Sloot, 2016) (Tsakyrakis, 2008).

Validation documentation of digital forensics is crucial to uphold to the proportionality principle in practice. The principle requires description of specific forensic task (scope of examination) and justification of the tools and methods chosen in the investigation.

However, empirical study in the Norwegian police showed that such documentation is often missing in DF reports (Stoykova et al., 2022). The proposed here RVEF aims to meet those objectives and to map the minimum documentation needed for proportionality assessment of the digital forensics’ methodology.

3. Specific challenges with validation in digital forensics

The theoretical background identifies two needs: (i) a generic independent validation process for digital forensics and (ii) exact documentation requirements to enable validation and minimum chain of custody/chain of evidence preservation. In order to identify further requirements for validation procedure in digital forensics, this section is focused on specific challenges with validation in practice.

3.1. Lack of a reliability standard: Daubert’s limitations

De facto reliability standards in DF are not specific enough to guide the consistent development of a validation process.

The US Supreme Court developed the Daubert standard[2] with several decisions to promote court criteria for evaluating reliability of expert evidence on scientific grounds. Since most jurisdictions don’t have clear reliability standards, Daubert had an international impact and turned into a de facto standard for digital forensics.[3] The Daubert standard requires: *(1)* the forensic method to be tested, *(2)* peer-reviewed, *(3)* generally accepted in the scientific community, *(4)* with identified error rates *(5)* and within the examiner’s expertise. In previous work we identified multiple limitations for implementation of this criteria in digital forensic related to the lack of procedures to produce the information needed for Daubert evaluation. [21, Tbl. 2] Testing is limited due to insufficient resources, quality standards, and test scenarios (Horsman, 2019a) (Horsman, 2019b) (Horsman, 2018b). It is unclear what type of peer review or competence of the reviewer is acceptable (Tully et al., 2020) (Marsico, 2004). Often there is no agreement in digital forensics what is standard or accepted method (Horsman, 2019b) (Marsico, 2004) (Sremack, 2007) (Arshad et al., 2018). The requirements for DF expert skills vary among jurisdictions (Henseler and van Loenhout, 2018) (Kwakman et al., 2011).

Daubert aims to make judges attentive to reliability issues in forensics, but under no circumstances suggests that judges or defence lawyers must perform complex scientific validation in the court room. Although digital forensics validation requires a Dauber-similar standard, the criteria are vaguely formulated, jurisdiction-specific, and insufficient to guide such a dynamic discipline. They must be used to develop a more detailed formal model for validation.

3.2. Lack of verification and validation specifications for tools

In digital forensics, often development specifications are not disclosed by the vendor (Marshall and Paige, 2018), and LEAs rely on the vendors testing that a tool is functioning appropriately. Often DF methods and tools are not specifically designed to satisfy law enforcement purposes and legal constrains (Marshall and Paige, 2018) (Page et al., 2019). For commercial digital forensic tools, Marshal and Page concluded that validation requirements for law enforcement purposes are neither clearly formulated by law enforcement, nor do DF tool vendors provide validation information according to such requirements (Marshall and Paige, 2018). Moreover development and customer specifications are changing rapidly due to changes or updates in the underlying software or the DF tool itself (Horsman, 2019a) (Tully et al., 2020). New digital forensic methods are introduced more quickly than accreditation or certification can be obtained (Tully et al., 2020).

Often in digital forensics is stated that exhausting all testing scenarios is impossible (Horsman, 2019a). A lot of the big forensic suites (EnCase, X-Ways, XRY, UFED, Cellebrite, etc) routinely used by law enforcement and accepted by the courts performed poorly in test scenarios designed by NIST (NIST, 2017a). Multi-purpose or closed-source DF tools include many functionalities but validation testing is developed only for some of those functions, leaving others not tested at all (NIST, 2017b). Further, there is no European body to perform independent validation testing for law enforcement. Consequently, LEAs themselves need to test DF products and demonstrate testing results.

[2] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U S. 579, 1993. The Daubert criteria was further elaborated in General Electric Co. v. Joiner 522 U S. 136 (1997), and Kumho Tire Co. v. Carmichaet 526 U S. 137 (1999).

[3] For example, the Scientific Working Group on Digital Evidence (SWGDE) uses Daubert to define error mitigation analysis (Scientific Working Group on Digital Evidence, 2018). Montasari based the evaluation of the quality of over 30 digital forensic process models on it (Montasari, 2016). A google Scholar search for “Daubert and digital forensics” outputs more than 3,690 results. Some authors directly refer to it as a de facto standard (Jasanoff, 2005a).

Requirement: To address these challenges the proposed RVEF must aim to provide a generic and formal process for law enforcement to test, document, and trace back digital forensics processing operations in order to meet legal requirements for evidence reliability. Such generic process must not be dependent on commercial validation, upgrades in technology, or the specifics of the case. The developed in RVEF standard procedure for minimum documentation of digital forensics actions must facilitate the design, improvement, and implementation of any type of testing methods and techniques.

3.3. *Lack of reproducibility and repeatability studies for DF methodology*

Current validation procedures are focused predominantly on tool testing, but this proves insufficient in more complex cases. Identified challenges to forensic method validation are related to lack of realistic testing data sets, lack of validation scenarios, lack of reproducibility studies, and the need for multidisciplinary peer-review (Nordvik et al., 2021).

Reproducibility refer to “the ability to replicate a measurement during repeated analysis” (Hughes and Karabiyik, 2020). Repeatability requires one investigator to be able to arrive at the same conclusion as another under similar conditions (Valjarevic and Venter, 2012). Consistent literature for the past 10 years shows that digital forensics techniques are not reproducible (PCAST, 2016) (Garfinkel, 2010). In reproducibility or repeatability studies, systematic errors can cause the results to be consistently wrong, which is a problem of validity (Foster and Huber, 1999). Therefore, repeatability study is considered as “simply a confidence indicator and one which should form part of the overall process of validating a tool” (Horsman, 2019a).

A source code audit of the tool (Gerber and Leeson, 2004) is preferable but usually such information is not accessible. Exceptionally, source code access can be requested by courts under non-disclosure agreement or if the tool is challenged by the defense. Some argue that black-box testing as a methodology level validation is preferable (Khan and Khan, 2012) (Risinger, 2018), the simplest of which is dual-tool verification. Dual-tool verification is sufficient for verification of tool results only in limited situations. For example, it is not a valid method for testing tools which reuse libraries and functionalities (Friheim, 2016). Black-box testing is inefficient for algorithm testing (Khan and Khan, 2012) and algorithm implementation errors detection (Alendal et al., 2021). Therefore, more often examiners are required to perform reverse engineering which has multiple reliability limitations as it deals with partial knowledge and subjective interpretations (Nordvik et al., 2021).

An important challenge is the lack of realistic real or synthetic testing data sets in digital forensics (Hughes and Karabiyik, 2020). The best way to estimate error rates is to perform blind proficiency test on realistic datasets. This requires samples whose properties are known and agency unaffiliated with the forensic scientist's laboratory (Saks and Faigman, 2008). It was proposed, that for each forensic functionality with its specifications, there must be a “set of references with known results” in order “any tool regardless of its original design intention, can be validated against known elements” (Guo et al., 2009). The generation and maintenance of testing data sets in compliance with data protection regulation that are sufficiently large and can keep up with the technology advancement is burdensome and requires specific knowledge. Only recently, some efforts in this direction can be found (Göbel et al., 2022).

Requirement: The development of formal validation procedures and sufficient automated documentation of the forensic methods are the key needs in order to ensure reproducibility without overburdening practitioners (Stoykova and Franke, 2020). Such documentation should enable internal LEA testing of tools and methods on case level as well as by cross-examination of similar forensic tasks and methodologies in different cases.

3.4. *Lack of accountability for examiners’ errors*

Currently, most quality standards in digital forensics are not specific as to practical solutions for evaluating the examiner work and her/his interaction with the tools (Horsman, 2019b) (Page et al., 2019).

It is hard to distinguish method errors from examiner errors “in fields where the method is primarily the judgment of the examiner.” (Saks and Faigman, 2008) and such judgment is not reflected in the case documentation.

The dependencies of data interpretation on the skills and knowledge of the forensic examiner are not well studied. Examiners learn to configure complex tools, to fine tune them, or to extend them with scripts, batch files, and plugins according to the case specifics and the data set — but this reasoning is not recorded or represented.

Several authors have examined multiple biasing factors for examiners in digital investigations such as exposure to case-irrelevant information, base rate expectations from previous investigations, failure to evaluate competitive hypotheses, or digital context information indicating intent or bad character (Sunde and Dror, 2019), (Edmond, 2016) Examiner errors are related to inaccurate data examination and tool result interpretations, as well as improper interaction with the tool’s setup. Since the same tool or method for data examination can be used in multiple investigations and trials, failing to identify limitations and errors could potentially result in reopening previous cases for re-examination once the errors are detected. Currently, there is no standard in digital forensics for “calculating error rates for both tools and specific procedures.” (Carrier, 2002) Moreover, errors related to many digital forensic activities are “systematic in nature and no statistical error rate exists.” (Lyle, 2010) *McKemmish* argues that “it is impossible to test for either the inaccuracy or accuracy of computer operations, and impossible to give a statistical rate of failure, and that there is therefore no rational basis for assuming a high rate of *reliability*” (McKemmish et al., 2008).

Requirement: Validation procedures must document examiners’ subjective judgement and interpretation. Cross-verification of results based on documenting the methods and tools, with examiner interaction in the process, is a preferred method for advancing the field.

3.5. *Lack of a standardized reporting process*

There is a lack of standardized and efficient procedures for documenting processing operations which results in resource and time consuming reporting that overburdens practitioners (Casey, 2019), (Horsman, 2018b) However, poor documentation cannot serve as an established practice (Horsman, 2018b).

Requirement: New solutions to enable and automate swift chain of custody and chain of evidence as standard documentation processes in law enforcement work with digital forensics is necessary from both theoretical and practical perspective.

4. Proposed reliability validation enabling framework for digital forensics (RVEF)

4.1. *RVEF: Generic framework*

In 2005, Saks and Koehler envisaged a paradigm shift in the forensic identification sciences where “untested assumptions and semi-informed guesswork are replaced by a sound scientific foundation and justifiable protocols” (Saks and Koehler, 2005). This effort, the authors argued further, should begin with adoption of the basic research model, which allows forensic scientists to design experiments that test the core assumptions of their fields. A research model based on the scientific method (Define problem – Hypotheses – Observation – Analysis – Evaluate Hypotheses) seems to be essential also for digital forensics. However, the advancement in computations to support every stage of this scientific process additionally complicate reliability assurance. By examining the effects of increased computations in forensic sciences and investigations, Franke and Srihari argued that contemporary forensic sciences can be defined as the intersection between technology, methodology, and application (Franke and Srihari, 2008). They concluded that in order the investigation to benefit from the advancement in technology and methodology, the application level should be supported by “new work procedures and legal frameworks … that take advantage of both knowledge domains; forensic and computational sciences”.[Franke and Srihari, 2008, p. 8] This generic definition (see Fig. 1) was not further developed by the authors or conceptualized for each of the three intersections. The hereafter proposed RVEF adapts and instrumentalized this generic model to develop a new approach to improve reliability in the digital forensics' domain. This model is selected as a base concept for several reasons. First, although abstract the model provides an insight of all levels of intersection in digital forensics as a multidisciplinary field of expertise. In this sense, it overcomes Daubert's limitations which is purpose-specific, one-dimensional models in law. The generic framework also doesn't focus on results-based requirements for forensic reports, but rather on the importance of a broader process-level perspective where technology, methodology, and application aspects can have impact on the quality of the digital forensics result. Therefore, the framework supports equally *(i)* a research model in forensics and computer science, and *(ii)* a development of a new legal approach to reliability evaluation.

A reliability validation enabling framework means a conceptual framework which identifies legal and forensic requirements for digital evidence reliability and elaborates the related measurements in digital forensic processes that need to be documented in order to meet the high-level requirements. Given the rapid developments and increased data volumes and complexities in technology and digital forensics a “one-standard-fits-all” validation cannot be applied. For these reasons and in order to be practical, the validation procedures must have the following properties.

- Formal: to be swiftly implemented in practice; extensible; machine- and human-readable; suitable for different jurisdictions and the rapid advancement in DF technology
- Validation – minimum requirements for reliability testing in DF procedures
- Framework – ideal validation criteria with considerations for practical implementation
- Covering legal and scientific objectives in law enforcement work with DF
- Documentation – enabling auditing and multitude of reliability testing methods in digital investigations.

In previous work a reliability challenges taxonomy was developed in order to summarize the problems related to reliability assurance in digital forensics (Stoykova, 2021). Based on this taxonomy, the RVEF identifies four validation criteria – data set, tool, method, and examiner. They have interdependencies to be accounted for in validation (see Fig. 2). For each of the core validation criteria RVEF defines minimum documentation requirements at three abstract levels of validation – technology, method, and application level.

First, each of the three intersections is defined. Technology must be conservatively used in the sense that tools and automated processes must be verified and validated for their ability to meet development specifications and law enforcement requirements and to achieve accurate results. The forensic methodology as a sequence of predefined steps must be tested for scientific validity. At the application level the validation process must establish if the examiner correctly selected and used the method and the tool in the concrete case according to the forensic task and the data set characteristics. All three levels have overlaps in their scope and validation requirements but also specifics to be considered.

Secondly, for each of the intersections, RVEF identifies reliability requirements. The literature review in the reliability challenges taxonomy was used as a first step to identify such requirements. In addition, best practices and guidelines from NIST (Ayers et al., 2014), ISO (ISO/IEC 27037:2012; ISO/IEC 17025:2017; ISO/IEC

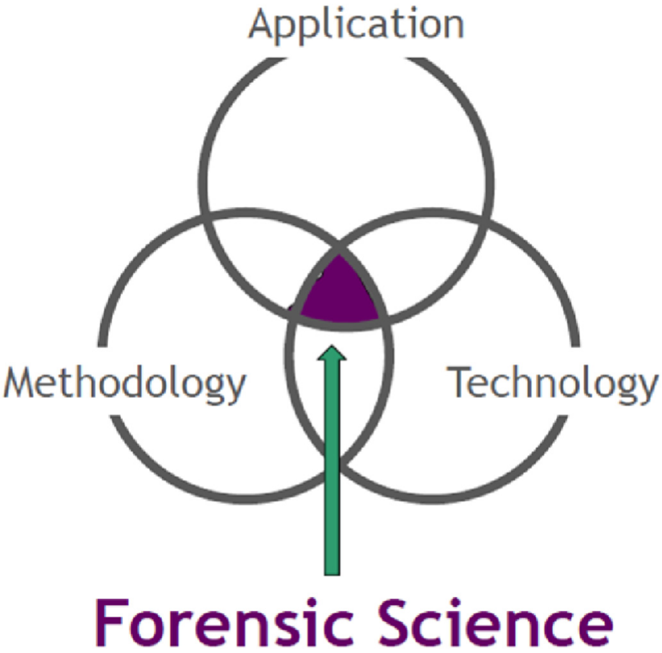

Fig. 1. Forensic science dimensions (Franke and Srihari, 2008).

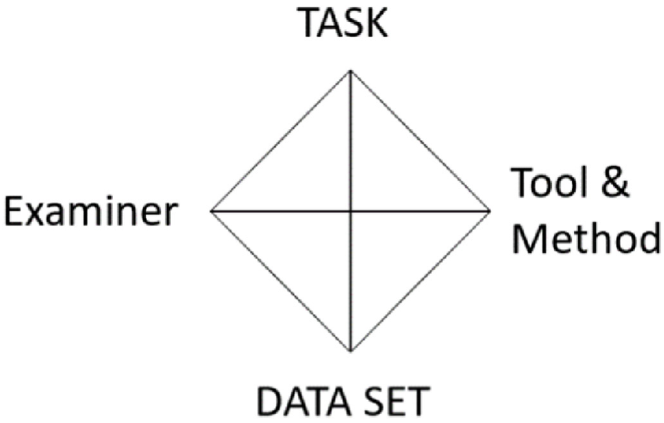

Fig. 2. Reliability criteria for digital forensics.

27041:2015; ISO/IEC 27042:2015) ENFSI (ENFSI, 2015b), and Interpol (Interpol, 2019) were examined to derive further reliability requirements for the three-levels RVEF. These standards were selected as they map fundamental digital forensic principles internationally recognized by the digital forensic community. The ENFSI best practice manual for examination of digital technology is considered suitable as it is a reliability validation standard. The NIST guide and ISO standards map general, technical, and organizational conditions, while the Interpol guidelines are focused on acquisition requirements of procedure specifically for law enforcement and define guidance notes for documentation of digital investigation. The standards are extensive, descriptive, and define sequence of steps and processes, rather than concrete criteria for validation and documentation. RVEF builds up on this by providing a consistent three-level schema for documentation, where the standards are concretized, and instrumentalized as minimum practical requirements for documentation of digital forensic work which can satisfy a validation procedure. The RVEF does not guarantee that the international standards are fulfilled but maps minimum documentation to enable reliability validation testing to make the forensic process accountable and testable. Some of the international standards suggest concrete documentation requirements.[(Interpol, 2019), p. 66], [(Ayers et al., 2014), Para. 7], [(ISO/IEC 27042:2015), Para. 9.2], [(ENFSI, 2015b), p. G], However, they are mainly on case level reporting, and require further interpretation for reliability testing purposes as provided in the RVEF. In addition, current testing of tools, methods, and examiners is done in isolation, while a practitioner needs information on the validation process as a hole. The RVEF can guide and support any type of validation technique and can link together different validation specifications relevant to the forensic task.

4.2. RVEF: Technology level

At the technology level, validation documentation must provide proof that a tool is treating all the input data in the same way, does not omit any data, processes everything according to the forensic objectives, and does not serve personal or corporate interests (Stoykova and Franke, 2020).

In this validation framework, the tool is understood as the specific functionality of the automated setup which is employed in the methodology and may include commercial software, but also in-house scripts, open-source tools, and batch files.

It is recommended that law enforcement use only validated tools and they are revalidated when updates are released.[(Interpol, 2019), Para. 3.4], [(Ayers et al., 2014), Para. 3.4], The tool documentation must include its name, version, configuration, and functions used.[(ENFSI, 2015b), Para. 4.2 and 6.6]. In commercial tools this is sufficient as the algorithm and implementation are fixed. In bigger tools specification of the concrete function used is important to detect algorithm and implementation errors. Reference to previous validation and verification testing can inform about known errors. [(ENFSI, 2015b), p. 23]

Documentation should provide reference to the tool results and tool's reported errors in output e.g. areas of the disk that were not recovered or read correctly. Technology level documentation is sufficient for dual tool verification.

4.3. RVEF: Method level

The method and tool technical requirements for validation overlap. Several forensic methods can be automated in one tool, or different tools might be necessary to construct the forensic methodology. In more complex computations it is insufficient to perform technology level validation only, as the examiner must have sufficient understanding of the use and limitations of peer-reviewed methods.

For the purpose of RVEF, the method is defined as a “concept that the work carried out by the organization is based on accepted scientific approaches, preferably consensus-based, and that any deviations from accepted scientific approaches can be substantiated in a manner considered generally acceptable by experts in that field.” (Marshall and Paige, 2018) Validation of the method is an “assessment of whether a standardized sequence of steps, often employing digital forensic tools, leads to a reliable result.” (Hughes and Karabiyik, 2020) The documentation should enable to determine if an appropriate scientific method, technique or procedure was followed(ISO/IEC 27037:2012), if the method meet the requirements of the investigation and have been appropriately tested (ISO/IEC 27042:2015). The used method for each forensic task should be validated. [ISO/IEC 27041:2015, Para. 5.5.2], [ISO/IEC 27042:2015, p. 7]

Ergo, documentation for formal validation of the method may include principles in computation and engineering, mathematical/statistical methods, or reference to peer reviewed methods, established practices, and previous work.[(ENFSI,2015), Para. 4.3] Documentation of previous work can serve as guidance for validation but must not be considered correct. In some cases, where the method is simple e.g., reference to an established practice might be sufficient. However, more complex methods or cases where peer-reviewed methods need to be modified to fit the present forensic task – an experiment or test setup should be described, including test data sets, and limitations of previous methods. Every new testing data set requires a new method level documentation.

At minimum the method for the pre-processing of the data set for input, and the feature and algorithm selection methods should be documented for reliability validation purposes. Each of these methods has certain limitations, which can impact the results and their correct interpretation. For example, it has been convincingly demonstrated that pre-processing for input has an effect on the accuracy of the results (Johnsen and Franke, 2019). Special issues with pre-processing are related to information loss due to digitalization, normalization, data reduction, data enrichment, changes to the system due to the forensic process. Changes introduced into the data set during pre-processing must be reduced to the minimum and must be documented in detail. Given the reproducibility requirements in digital forensics, only deterministic algorithms can be utilized, since they give the same results based on the same input data.

Feature selection depends on sufficient understanding of the data structure. However, in real data sets the base truth is not known and the examiner cannot be sure that the selected features are representable. Issues with features selection methods are related to insufficient detection of statistical properties representative for the data set and heuristic feature search strategies (Nguyen et al., 2010). Therefore, methods for feature extraction and selection must be fully documented and traceable.

Method level testing is resource and time consuming. Documentation about the algorithms limitations and its implementation can enable cross-validation of the method and examination of error triggering conditions.

4.4. RVEF: Application level

At the application level the DF examiner must ensure that the selected methods and tools fit the data set characteristics and the forensic task. They must be able to test if the methods and tools work correctly in the specific case and be able to document any human judgement in the setup.

The validation of the DF examiner performance is related to competence, authorization, performance and mitigation of biasing factors. Examiner's accreditation and certification are enough as entry requirements. Moreover, experts with domain or topic specific knowledge might be called in cases where an accredited digital forensic expert does not have competencies.

Documenting the investigative task is essential, as it defines the scope of the forensic examination [(Interpol, 2019), Para. 5.2.3.3] and impacts the selection of tools and method.[(ENFSI, 2015), p. 20] Defining a task also serves for proving that the scope of the investigationis proportionate and authorized, and the privacy of suspects or other human rights are not violated during the forensic actions .

Further, a description of the original data set for each forensic task must be documented. Physical preservation documentation includes authentication and identification of the initial physical carrier and digital data sources from the crime scene, digital field triage method, storage medium description, secure storage repository.[(ENFSI, 2015), p. 12] Logical preservation of the data set relates to the description of the data set structure and behavior, level of acquisition and integrity preservation methodology (e.g., unique identifiers and hash function description [(Interpol, 2019), Para. 5.1.2.3 and 5.1.3.3] (Ayers et al., 2014),), information availability assessment (errors/issues preventing access to the data and/or data source, e.g., volatility order, security locks, encryption, power fault, anti-forensics, etc.). [(Årnes,2018), p. 31 Table 2.2].

To validate the examiner work at the application level, the documentation must contain a minimum description of subjective measurements e.g., hypothesis, assumptions, decision taken based on expert knowledge (Interpol, 2019), [ISO/IEC 27042:2015, Para. 6.4] This requires justification of the selected methods and tools as to why they are suitable to solve the forensic task at hand.

The examiner's interaction with the tool must be traceable and includes parameterization of the tool or feature extraction and selection according to method specifications. Documenting the application level of the forensic examination will output the chain of evidence that can be cross-examined by other parties before or during trial.

RVEF at application level allows to perform proficiency testing (Hughes and Karabiyik, 2020) (ENFSI, 2015a). Common law jurisdictions use sometimes a practice of expert hot tubbing (Rares, 2011; Ross, 2013; Sommer, 2009) where examiners are discussing their arguments on the same facts in front of a judge. However, this practice is mainly advancing the legal arguments in the case, not necessarily improving scientific reliability. Dual investigator and random dip-sampling (Page et al., 2019) are good practices, however, it is a matter of resources since for some topic-specific knowledge there might be not two examiners with the same level of competence or sufficient cases to randomize. A more viable solution is establishing of independent, multi-disciplinary expert commissions. Such commission is best suited to perform testing of all validation criteria in all levels, but most importantly can ensure application-level reliability testing. In addition, proportionality assessment of the forensic method requires legal knowledge as well. Some formal verification methods are used also in validation for external functional testing, however risk-based testing to the best of our knowledge is not performed or published by any standardization body or in academia. For highly automated tools — such as those used in 'intelligent' triage —a validation against a human-created 'gold standard' has been proposed (James et al., 2014).

Table 2
Reliability validation techniques.

Level	Validation method
Technology	Dual-tool verification Black-box testing Reverse engineering Security testing
Method	Reproducibility/Repeatability study
Application	Proficiency testing; Dual investigator Expert hot tubbing; Expert commissions Random/Dip-sampling

Any errors or uncertainties found during the application of the method should be documented. Finally, reporting confidence levels about the traces' relevance and reliability in probabilistic terms is of key importance for advancing the investigation and consequently the trial. Output interpretation of tool results must ensure that facts and inference or opinions are kept separately.[79, Para. 13.4] The need for the fast evolution of methods and standards in digital forensics related to technology advancements, increased data volumes and complexities, means that reliability testing must be performed on the application level in each case and during daily work.

Application-level validation of computational methods in forensics is necessary given the advancements in machine-learning approaches to utilize computer power and reduce the information overhead in digital forensics. New methods need to be validated in their concrete application given the subjective nature of the pre-processing and feature selection and the importance of choosing the correct method and algorithm according to the data set structure and the forensic task.

4.5. RVEF: Summary and test scenarios

The RVEF is summarized in Table 1 above. As already stated, this is a general and formal framework for reliability validation, which can be extended and concretized for each specific digital forensic activity. The included minimum documentation is only exemplified, in order to prove that only a specific and limited number of steps and measures needs to be documented during the forensic work in order both automated and semi-automated processing operations to be clear.

In this section we discussed also that the RVEF can support the validation techniques listed in Table 2 below.

To prove practical utility the RVEF was tested in two case scenarios. The RVEF was trailed against peer-reviewed methods for file system reverse engineering in order to evaluate if they provide sufficient documentation for reliability and reproducibility studies (Nordvik et al., 2021). The second case study was conducted on investigation records from the Norwegian police and had twofold objectives: (i) to examine the records according to RVEF in order to

Table 1
RVEF overview.

Level	Minimum documentation
Technology	Tool type, name, version; Tool function used; Prior validation/verification results; Known errors reports; Tool's ability to report errors in output
Method	Algorithms and implementation; Reference to peer reviewed method; established practice; previous work; Experiment/Test setup; known limitations
Application	Forensic task; Data set; Guidelines/SOPs reference; Tool parameterization; Justification of method, algorithms, and features selection; Assessment of tool results; Confidence levels; Separation of facts and inferences

evaluate the reliability of digital evidence handled by the police; (ii) and to propose RVEF as a template for LEAs to improve their reliability assurance in their investigative work (Stoykova et al., 2022). An initial work to automate RVEF via standard expressions and to create validation testing reports efficiently was also presented (Stoykova and Franke, 2020).

5. RVEF discussion: techno-legal objectives

The proposed reliability validation enabling framework (RVEF) is motivated by the need for chain of custody and chain of evidence documentation in digital forensics investigations as a precondition to enable reliability validation. Further, validation procedures must be implemented in the design of digital forensics methodologies and processes where machine and human errors can be identified and mitigated. Generation of reliability validation documentation will assist the development of a legal reliability standard, as the digital evidence process can be audited at each processing stage for legal compliance and prejudicial effects. The RVEF ensures continuity with existing international standards for digital evidence as it adapts them to law enforcement needs and extends them with a practical framework for efficient compliance.

The generic framework does not suggest a concrete testing procedure or exhaustive testing at each stage of the digital evidence process. As Casey argued well-established technical processes might require only an audit or some level of repeatability, while evidence evaluation must include procedures standardization and testing implementation (Casey, 2016). It ensures the traceability of critical decisions in the examination, the sound cross-validation of the methodology steps and the tool parameterization by the examiner. Moreover, reliability testing can be performed on different levels or on all of them depending on the case specifics and resources.

The RVEF has the advantage that it maps the formal procedure and minimum documentation required for validation in any digital forensics task and at any stage of the evidence processing. This overcomes the limitations of high-level and abstract reliability requirements and provides a process-level validation schema as opposed to the current focus on only tool-testing or testing of tools, methods, and examiners done in isolation. The RVEF can link together different validation specifications relevant to the forensic task. Moreover, RVEF-based documentation does not require a general acceptance of methods, because they can be documented and tested with flexible and different types of validation techniques depending on the case. The RVEF allows for errors and uncertainties documentation during forensic examination. It also assists in the generation of data for developing more advanced validation approaches and test scenarios in the LEA work. It focuses on mapping the automated and semi-automated processing operations by making explicit the forensic actions in the digital forensic investigation.

The minimum documentation process can assist in administrative control and audit in routinely and easily verifiable tasks in acquisition, as well as robust lab testing during evidence examination and analysis (Casey, 2016). DF examiners and LEA can track forensic actions based on scientific methodology and separate them clearly from purely investigative actions.

The proposed framework is suitable for automation and everyday documentation of each case where digital forensic examination is performed in order to save time and guarantee expedience in practice. It can inform examiners of the minimum documentation required for validation it can be used as a template for their day-today work. This does not suggest that digital forensic actions have to be validated on a daily basis, but in case of doubt or a challenge to the digital evidence in the consequent criminal proceedings, the RVEF can ensure accountability and sufficient documentation for cross-examination. Moreover, such validation documentation can be automated and standardized to the benefit of cross-verification of results amongst examiners and amongst different labs. The main objective of the minimum documentation is to provide a guideline for practitioners. This can assist in formulating and improving development and customer specifications by law enforcement to commercial tool vendors, that can enable LEA to better test the digital forensics tools purchased. The implementation of the RVEF process by LEA can continuously ensure the tracking of the digital evidence process and the management and support optimization and error rates studies.

RVEF is flexible enough to satisfy fair trial requirements. It can serve judges and defense lawyers in cross-examining the forensic report in a formalized process, where potential dangers to the presumption of innocence and risks from the inappropriate use of technology can be mitigated. Generating the information needed for validation will assist in information-driven policy and regulation and informed decisions by the courts. Statistics and analytics can be used to evaluate the validity, proportionality, and legal compliance of different forensic methods. The RVEF also aims to ensure that if challenged in court the forensic actions can be traced back and errors are identifiable. Defense lawyers can be provided with access to such documentation in order to understand and challenge the digital forensics actions and request exculpatory data to be retrieved using different methods, features, or algorithms. Moreover, based on this three level validation documentation judges can verify that the authorization and principles of proportionality, data minimization, and fairness of processing are not violated. The application-level validation is the most complex because it evaluates the adequacy of digital forensics against the law enforcement purpose (Jasanoff, 2005b).

The propose RVEF can provide the necessary information for proportionality assessment and assist in classification of digital forensic methods and tools according to their intrusiveness. This allows more intrusive digital forensics methods to be gradually justified and authorized when less intrusive ones fail to acquire the evidence.

6. Conclusions and further work

This paper argues that digital forensics can reach a level of standardization and validation similar to the classical forensic sciences. However, we identify as major gaps the lack of clear reliability standard and the focus on quality assurance of technology, where methodology and application validation techniques are underdeveloped. As opposed to “one-standard-fits-all” lab requirements, proposed solutions should enable gradual documentation of the methods, tools, and the interaction of examiners across the process in order to enable different types of validation procedures.

To support theoretically the development of a reliability standard, we clarified concepts routinely used as a measure for quality assurance in digital evidence since often they have different nuances in the legal and forensic science domain.

The proposed reliability validation framework (RVEF) is a conceptual framework which identifies practical, legal and forensic requirements for evidence reliability and elaborates the related law enforcement requirements in digital forensic processes that needs to be documented in order to meet the high-level criteria. The framework identifies four validation criteria – data set, tool, method, and examiner.

The RVEF suggests a model for minimum documentation of three level validation requirements as a first step to address the identified reliability challenges.

A technology level documentation assures validation of tools and specific functionality of the automated setup which is employed in the investigation task.

A methodology level documentation provides proof that an accepted scientific procedure, and standardized sequence of steps is followed to provide reliable results. It includes method, algorithms, and feature selection and detailed description of dataset, experiment setup and preprocessing for input.

At the application level, it was identified that the examiner's interaction with the method and tool as well as subjective measurements must be traceable and justified according to the concrete forensic task.

The RVEF is general and needs to be elaborated and tested further. Nevertheless, the added value of RVEF is that it enables the gradual development of techno-legal standards for reliability as it facilitates any type of testing on any stage of the evidence processing. Further, it can be used by LEAs to create audit trials of digital forensic actions, which can be studied at large for reducing subjective opinions and assumptions in favour of objective measurements, formal justification of the selected methodology according to the forensic task, and large-scale reliability and error rates studies. Most importantly, the RVEF provides the minimum documentation to secure the opportunity for cross-examination and the challenging of digital evidence on valid grounds in further criminal proceedings.

Considering that digital forensics for criminal proceedings requires not only scientific validation, but also proportionality and data protection assessment, the RVEF can serve as a first step to meet these ends as well.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Adams, R., Hobbs, V., Mann, G., 2013. The advanced data acquisition model (Adam): a process model for digital forensic practice. J.Dig. Forensics, Sec. Law 8 (4). https://doi.org/10.15394/jdfsl.2013.1154.

Alendal, G., Axelsson, S., Dyrkolbotn, G.O., 2021. Chip chop — smashing the mobile phone secure chip for fun and digital forensics. Forensic Sci. Int.: Digit. Invest. 37, 301191. https://doi.org/10.1016/j.fsidi.2021.301191.

Årnes, A. (Ed.), 2018. Digital Forensics: an Academic Introduction. John Wiley & Sons Inc, Hoboken, NJ.

Arshad, H., Jantan, A.B., Abiodun, O.I., 2018. Digital forensics: review of issues in scientific validation of digital evidence. J. Inf. Proc. Syst. 14 (2), 346–376.

Ayers, R., Brothers, S., Jansen, W., 2014. Guidelines on Mobile Device Forensics. National Institute of Standards and Technology [Online]. Available: https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/cftt-technical/mobile.

Bart van der Sloot, 2016. The practical and theoretical problems with "Balancing": Delfi, coty and the redundancy of the human rights framework. Maastricht J. Eur. Comp. Law 23 (3), 439–459.

Beebe, N.L., Clark, J.G., 2004. A hierarchical, objectives-based framework for the digital investigations process. Digit. Invest. 2 (2), 147–167, 10/c74x7g.

Carrier, B., 2002. Open Source Digital Forensics Tools: the Legal Argument.

Carrier, B.D., 2006. 'A Hypothesis-Based Approach to Digital Forensic Investigations', *Theses and Dissertations Available from ProQuest*, pp. 1–169.

Casey, E., 2016. Differentiating the phases of digital investigations. Digit. Invest. 19, A1–A3. https://doi.org/10.1016/j.diin.2016.11.001.

Casey, E., 2019. The chequered past and risky future of digital forensics. Aust. J. Forensic Sci. 51 (6), 649–664. https://doi.org/10.1080/00450618.2018.1554090.

Council of the European Union, 2011. 'Council conclusions on the vision for European Forensic Science 2020 including the creation of a European Forensic Science Area and the development of forensic science infrastructure in Europe. In: 3135th JUSTICE and HOME AFFAIRS Council Meeting [Online]. Available: https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/jha/126875.pdf.

Council of the European Union, 2016. Draft Council Conclusions on the Way Forward in View of the Creation of an European Forensic Science Area. Council of the European Union [Online]. Available: http://data.consilium.europa.eu/doc/document/ST-6078-2016-INIT/en/pdf. (Accessed 27 March 2018).

Daniel, L., Daniel, L., 2012. Digital Forensics for Legal Professionals: Understanding Digital Evidence from the Warrant to the Courtroom. Syngress/Elsevier, Waltham, MA.

Doyle, S., 2019. Quality Management in Forensic Science. Elsevier: Academic Press, is an imprint of Elsevier, London, United Kingdom ; San Diego, CA, United States.

Dror, I.E., 2020. Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. Anal. Chem. 92 (12), 7998–8004. https://doi.org/10.1021/acs.analchem.0c00704.

Edmond, G., 2012. Is reliability sufficient? The law commission and expert evidence in international and interdisciplinary perspective (Part 1). Int. J. Evid. Proof 16, 30–65. https://doi.org/10.1350/ijep.2012.16.1.391.

Edmond, G., 2016. Legal versus non-legal approaches to forensic science evidence. Int. J. Evid. Proof 20 (1), 3–28. https://doi.org/10.1177/1365712715613470.

Edmond, G., Roberts, A., 2011. Procedural fairness, the criminal trial and forensic science and medicine. Syd. Law Rev. 33 (3), 36.

European Network of Forensic Science Institutes (ENFSI), 2015. Best practice manual for the forensic examination of digital technology. http://enfsi.eu/wp-content/uploads/2016/09/1._forensic_examination_of_digital_technology_0.pdf. (Accessed 30 March 2020).

European Network of Forensic Science Institutes (ENFSI), 2015. Best Practice Manual for forensic examination of digital technology [Online]. Available: https://enfsi.eu/wp-content/uploads/2016/09/1._forensic_examination_of_digital_technology_0.pdf.

Foster, K.R., Huber, P.W., 1999. In: *Judging Science: Scientific Knowledge and the Federal Courts*, 1, paperback ed. MIT Press, Cambridge, Mass.

Franke, K., Srihari, S.N., 2008. Computational Forensics: an Overview. Computational Forensics, Berlin, Heidelberg, pp. 1–10. https://doi.org/10.1007/978-3-540-85303-9_1.

Friheim, I., 2016. Practical Use of Dual Tool Verification in Computer Forensics. University College Dublin. https://doi.org/10.13140/RG.2.2.33300.81288.

Garfinkel, S.L., 2010. Digital forensics research: the next 10 years. Digit. Invest. 7, S64–S73, 10/bmnnb3.

Garfinkel, S., Farrell, P., Roussev, V., Dinolt, G., 2009. Bringing science to digital forensics with standardized forensic corpora. Digit. Invest. 6, S2–S11. https://doi.org/10.1016/j.diin.2009.06.016.

Gerber, M., Leeson, J., 2004. Formalization of computer input and output: the Hadley model. Digit. Invest. 1 (3), 214–224. https://doi.org/10.1016/j.diin.2004.07.001.

Göbel, T., Maltan, S., Türr, J., Baier, H., Mann, F., 2022. ForTrace - a holistic forensic data set synthesis framework. Forensic Sci. Int.: Digit. Invest. 40, 301344. https://doi.org/10.1016/j.fsidi.2022.301344.

Gross, S., Mnookin, J., 2003. Expert information and expert evidence: a preliminary taxonomy. Articles [Online]. Available: https://repository.law.umich.edu/articles/570.

Guo, Y., Slay, J., Beckett, J., 2009. Validation and verification of computer forensic software tools—searching Function. Digit. Invest. 6, S12–S22. https://doi.org/10.1016/j.diin.2009.06.015.

Henseler, H., van Loenhout, S., 2018. Educating judges, prosecutors and lawyers in the use of digital forensic experts. Digit. Invest. 24, S76–S82. https://doi.org/10.1016/j.diin.2018.01.010.

Horsman, G., 2018a. Framework for Reliable Experimental Design (FRED): a research framework to ensure the dependable interpretation of digital data for digital forensics. Comput. Secur. 73, 294–306. https://doi.org/10.1016/j.cose.2017.11.009.

Horsman, G., 2018b. Framework for Reliable Experimental Design (FRED): a research framework to ensure the dependable interpretation of digital data for digital forensics. Comput. Secur. 73, 294–306, 10/gcx6dd.

Horsman, G., 2019a. Tool testing and reliability issues in the field of digital forensics. Digit. Invest. 28, 163–175. https://doi.org/10.1016/j.diin.2019.01.009.

Horsman, G., 2019b. Formalising investigative decision making in digital forensics: proposing the digital evidence reporting and decision support (DERDS) framework. Digit. Invest. 28, 146–151. https://doi.org/10.1016/j.diin.2019.01.007.

Hughes, N., Karabiyik, U., 2020. Towards reliable digital forensics investigations through measurement science. WIREs Forensic Sci. n/a (n/a), e1367. https://doi.org/10.1002/wfs2.1367.

International Criminal Police Organization (Interpol), 2019. Global guidelines for digital forensics laboratories [Online]. Available: https://www.interpol.int/en/content/download/13501/file/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory.pdf.

International Organization for Standardization and International Electrotechnical Commission, 2015. ISO/IEC 27041:2015 Information technology — security techniques — guidance on assuring suitability and adequacy of incident investigative method. ISO/IEC 27041:2015. https://www.iso.org/obp/ui/#iso:std:iso-iec:27041:ed-1:v1:en. (Accessed 4 April 2018).

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 2012. ISO/IEC 27037 eForensics Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence. https://www.iso27001security.com/html/27037.html. (Accessed 3 September

2020).

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 2017. ISO/IEC 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories. ISO. http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/69/66912.html. (Accessed 16 January 2020).

International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), 2015. ISO/IEC 27042:2015 Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence. *ISO/IEC 27042:2015*. https://www.iso.org/obp/ui/#iso:std:iso-iec:27042:ed-1:v1:en. (Accessed 4 April 2018).

James, J.I., Lopez-Fernandez, A., Gladyshev, P., 2014. Measuring Accuracy of Automated Parsing and Categorization Tools and Processes in Digital Investigations, vol. 132, pp. 147–169. https://doi.org/10.1007/978-3-319-14289-0_11. *arXiv: 1502.05186 [cs]*.

Jasanoff, S., 2005a. Law's knowledge: science for justice in legal settings. Am. J. Publ. Health 95 (Suppl. 1), S49–S58. https://doi.org/10.2105/AJPH.2004.045732.

Jasanoff, S., 2005b. Law's knowledge: science for justice in legal settings. Am. J. Publ. Health 95 (Suppl. 1), S49–S58. https://doi.org/10.2105/AJPH.2004.045732.

Johnsen, J.W., Franke, K., 2019. The impact of preprocessing in natural language for open source intelligence and criminal investigation. In: 2019 IEEE International Conference on Big Data. Big Data), Los Angeles, CA, USA, pp. 4248–4254. https://doi.org/10.1109/BigData47090.2019.9006006.

Jones, A., Vidalis, S., 2019. Rethinking digital forensics. Ann.Emerg.Technol. Comput. 3, 41–53. https://doi.org/10.33166/AETiC.2019.02.005.

Khan, Mohd E., Khan, F., 2012. A comparative study of white box, black box and grey box testing techniques. Int. J. Adv. Comput. Sci. Appl. 3 (6). https://doi.org/10.14569/IJACSA.2012.030603.

Kloosterman, A., et al., 2015. The interface between forensic science and technology: how technology could cause a paradigm shift in the role of forensic institutes in the criminal justice system. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370 (1674). https://doi.org/10.1098/rstb.2014.0264.

Kohn, M.D., Eloff, M.M., Eloff, J.H.P., 2013. Integrated digital forensic process model. Comput. Secur. 38, 103–115. https://doi.org/10.1016/j.cose.2013.05.001.

Kwakman, N.J.M., Nijboer, J.A., Keulen, B.F., Elzinga, H.K., 2011. Expert Registers in Criminal Cases. Governance in Criminal Proceedings [Online]. Available: https://www.rug.nl/rechten/congressen/archief/2011/governancemeetslaw/workingpapers/papernijboerkeulen.pdf. (Accessed 25 June 2020).

Lyle, J.R., 2010. If error rate is such a simple concept, why don't I have one for my forensic tool yet? Digit. Invest. 7, S135–S139. https://doi.org/10.1016/j.diin.2010.05.017.

Marshall, A., Paige, R., 2018. Requirements in Digital Forensics Method Definition: Observations from a UK Study.

Marsico, C.V., 2004. CERIAS Tech Report: Computer Evidence V. Daubert: the Coming Conflict. Purdue University School of Technology [Online]. Available: https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2005-17.pdf.

Mason, S., Seng, D., 2017. *Electronic Evidence*, Fourth. University of London, Institute of Advanced Legal Studies. https://doi.org/10.14296/517.9781911507079.

McKemmish, R., 2008. When is digital evidence forensically sound?. In: Ray, I., Shenoi, S. (Eds.), Advances in Digital Forensics IV, vol. 285. Springer US, Boston, MA, pp. 3–15. https://doi.org/10.1007/978-0-387-84927-0_1.

Montasari, R., 2016. Review and assessment of the existing digital forensic investigation process models. Int. J. Comput. Appl. 147 (7), 41–49. https://doi.org/10.5120/ijca2016911194.

Nguyen, H.T., Franke, K., Petrovic, S., 2010. Towards a generic feature-selection measure for intrusion detection. In: 2010 20th *International Conference On Pattern Recognition*,, pp. 1529–1532. https://doi.org/10.1109/ICPR.2010.378.

Nordvik, R., Stoykova, R., Franke, K., Axelsson, S., Toolan, F., 2021. Reliability validation for file system interpretation. Forensic Sci. Int.: Digit. Invest. 37, 301174. https://doi.org/10.1016/j.fsidi.2021.301174.

Page, H., Horsman, G., Sarna, A., Foster, J., 2019. A review of quality procedures in the UK forensic sciences: what can the field of digital forensics learn? Sci. Justice 59 (1), 83–92. https://doi.org/10.1016/j.scijus.2018.09.006.

Rares, S., 2011. Using the “hot tub”: how concurrent expert evidence aids understanding of issues. Judic. Rev. 171–186.

Reedy, P., 2020. ‘Interpol Review of Digital Evidence 2016 - 2019’, *Forensic Science International: Synergy*. https://doi.org/10.1016/j.fsisyn.2020.01.015.

Risinger, D.M.. ‘Unsafe verdicts: the need for reformed standards for the trial and review of factual innocence claims’, social science research Network. Rochester, NY, SSRN Scholarly Paper ID 610665, Oct. 2004 [Online]. Available: https://papers.ssrn.com/abstract=610665. (Accessed 3 June 2019).

Risinger, D., 2018. The five functions of forensic science and the validation issues they raise: a piece to incite discussion on validation [Online]. Available: Seton Hall Law Rev. 48 (3) https://scholarship.shu.edu/shlr/vol48/iss3/6.

Risinger, D.M., 2000. Navigating Expert Reliability: are Criminal Standards of Certainty Being Left on the Dock? Social Science Research Network, Rochester, NY. SSRN Scholarly Paper ID 251033, Accessed: Jun. 25, 2020.

Roberts, P., Zuckerman, A.A.S., 2010. In: Criminal Evidence, second ed. Oxford University Press, Oxford ; New York.

Rogers, M., 2006. Dcsa: a practical approach to digital crime scene analysis. In: Krause, M., Tipton, H. (Eds.), Information Security Management Handbook, fifth ed., vol. 3. Auerbach Publications, pp. 601–614. https://doi.org/10.1201/9781420003406.ch9.4.

Ross, Andrew, 2013. Murky waters: An expert's perspective on the effectiveness of expert conclaves and “hot tubs”', *Precedent (Sydney, N.S.W.)*. No 119, pp. 30–34.

Saks, M., Faigman, D.L., 2008. Failed forensics: how forensic science lost its way and how it might yet find it. Annu. Rev. Law Soc. Sci. 4, 149–171. https://doi.org/10.1146/annurev.lawsocsci.4.110707.172303.

Saks, M.J., Koehler, J.J., 2005. The coming paradigm shift in forensic identification science. Science 309 (5736), 892–895. https://doi.org/10.1126/science.1111565.

Schum, D.A., 2001. Evidence marshaling for imaginative fact investigation. Artif. Intell. Law 9 (2/3), 165–188. https://doi.org/10.1023/A:1017941304013.

Scientific Working Group on Digital Evidence, 2018. ‘Establishing confidence in digital and multimedia evidence forensic results by error mitigation analysis v.2.0’. SWGDE, nov. 20 [Online]. Available: https://fenix.tecnico.ulisboa.pt/downloadFile/845043405513436/SWGDE_Establishing_Confidence_in_Digital_Forensic_Results_by_Error_Mitigation_Analysis.pdf. (Accessed 12 March 2022).

Sommer, P., 2009. Meetings between experts: a route to simpler, fairer trials? Digit. Invest. 5 (3–4), 146–152. https://doi.org/10.1016/j.diin.2008.11.002.

Sommer, P., 2010. Forensic science standards in fast-changing environments. Sci. Justice : J. Forensic Sci. Soc. 50, 12–17. https://doi.org/10.1016/j.scijus.2009.11.006.

Sremack, J.C., 2007. The gap between theory and practice in digital forensics [Online]. Available:. In: Presented at the Conference on Digital Forensics, Security and Law https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.823.8791&rep=rep1&type=pdf.

Stoykova, R., 2021. Digital evidence: unaddressed threats to fairness and the presumption of innocence. Comput. Law Secur. Rep. 42, 105575. https://doi.org/10.1016/j.clsr.2021.105575.

Stoykova, R., Franke, K., 2020. Standard representation for digital forensic processing. In: 2020 13th International Conference on Systematic Approaches to Digital Forensic Engineering. SADFE), pp. 46–56. https://doi.org/10.1109/SADFE51007.2020.00014.

Stoykova, R., Sep. 2021. Digital evidence: Unaddressed threats to fairness and the presumption of innocence. Computer Law Secur. Rev. 42, 105575. https://doi.org/10.1016/j.clsr.2021.105575.

Stoykova, R., Andersen, S., Franke, K., Axelsson, S., 2022. Reliability assessment of digital forensic investigations in the Norwegian police. Forensic Sci. Int.: Digit. Invest. 40, 301351. https://doi.org/10.1016/j.fsidi.2022.301351.

Sunde, N., Dror, I.E., 2019. Cognitive and human factors in digital forensics: problems, challenges, and the way forward. Digit. Invest. 29, 101–108. https://doi.org/10.1016/j.diin.2019.03.011.

The Association of Chief Police Officers in UK, 2012. ACPO Good practice guide for digital evidence [Online]. Available: https://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf.

The UK National Police Chiefs Council, 2020. Digital forensic science strategy [Online]. Available: https://www.npcc.police.uk/Digital%20Forensic%20Science%20Strategy%202020.pdf.

The United Kingdom Forensic Science Regulator. Method validation in digital forensics. FSR-G-218, issue 2'. 2020. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/921392/218_Method_Validation_in_Digital_Forensics_Issue_2_New_Base_Final.pdf.

The US President's Council of Advisors on Science and Technology (PCAST), 2016. Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods. https://obamawhitehouse.archives.gov/blog/2016/09/20/pcast-releases-report-forensic-science-criminal-courts. (Accessed 6 March 2020).

Tsakyrakis, Stavros, 2008. Proportionality: an Assault on human rights?' Jean monnet working paper 09/08 [Online]. Available: http://jeanmonnetprogram.org/wp-content/uploads/2014/12/080901.pdf. (Accessed 22 July 2020).

Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., Watson, T., 2020. Quality standards for digital forensics: learning from experience in England & Wales. Forensic Sci. Int.: Digit. Invest. 32, 200905. https://doi.org/10.1016/j.fsidi.2020.200905.

US National Institute of Standards and Technology, 2017. ‘Computer forensics tool testing program (CFTT)', *NIST*. https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt. (Accessed 29 January 2020).

US National Institute of Standards and Technology, 2017. ‘Computer forensics tool testing program (CFTT)', *NIST*. https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt. (Accessed 29 January 2020).

Valjarevic, A., Venter, H.S., 2012. Harmonised digital forensic investigation process model. In: 2012 Information Security for South Africa, pp. 1–10. https://doi.org/10.1109/ISSA.2012.6320441.