# Standard Representation for Digital Forensic Processing

Stoykova, Radina; Franke, Katrin

*Published in:*
13th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE)

*DOI:*
10.1109/SADFE51007.2020.00014

*Publication date:*
2020

1 2

# Standard Representation for Digital Forensic Processing

[[au1]]Radina Stoykova, [[au2]]Katrin Franke

*Abstract*—**bstractbstractA— This paper discusses the lack of reliability and reproducibility validation in digital forensics for a criminal trial. It is argued that this challenge can be addressed with standard data-representation for digital evidence. The representation must include reproducibility documentation on processing operations including automation, human interaction, and investigation steps. Analyzed are two blueprint articles – the CASE specification language for cyber-investigations [1] and the WANDA data standard for the documenting semi-automated hand-writing examination [2]. These two generic frameworks are studied for their granularity to support reproducibility testing by representing: (i) artefact characteristics, forensic – tool parameters and input – output logic; (ii) human and tool data interpretation; and (iii) parallel-running forensic tasks or chains of processes. Proposed is the integration of WANDA-based schema as CASE expression. The utility of such integration is demonstrated as a new module in CASE designed to meet the high standard of proof and scientific validation typically required in criminal investigations and trials. The expression ensures compliance without overburdening digital forensic practitioners.**

Index Terms— automation, digital forensics, machine- generated annotations, reliability, reproducibility, standard representation

## I. I. DIGITAL EVIDENCE VALIDITY UNDER PRESSURE?

digital evidence is the result of a controlled forensic procedure, which ensured that the data "authenticity and integrity can be validated"[3] and if "the method (or tool) used to gather and/or analyses digital evidence does change the original data set, the changes are identifiable"[4]. Validation procedures ensure that the digital forensic technique is following scientific methodology, produces accurate results, and therefore is reliable and reproducible. However, currently digital forensics are lacking formal validation procedures that can undermine their scientific credibility and render digital evidence inadmissible or low-probative in courts.

Ten years ago the NAS report [5] made the concerning statement that "no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source". The report was even more critical to the legal regimes ability to evaluate incriminating expert evidence concluding that it is "inadequate to the task of curing the documented ills of the forensic science disciplines". In 2016, another review confirmed that the issue with scientific validity is not addressed in most forensic fields, including digital evidence, calling it "a critical gap"[6]. In Europe same lack of "any [internationally] recognized quality standards for digital forensic processes and systems, and the lack of transparency" was pointed out as a major legislative challenge in the field [7]. These reports emphasize the missing effective measures to implement and enforce compliance, not the lack of regulation.

On the other hand, digital forensics, criminal investigations, and legal regulation of digital evidence are developing simultaneously and rapidly due to new advancements in technology. This is related also to more automation understood as computer-assistance in cognitively cumbersome digital forensic activities. This is expanding the experts‘ work to validation of the processing and evaluation of the legal effects of such processing in the investigation.

The rapid scientific advancements in computer-assisted forensic science render a lot of existing validation schemas outdated [8], sidetrack reproducibility studies [1], disturb accuracy testing in digital forensics [9], and in the subsequent court evaluation. This calls for new more general reproducibility framework, which can be expedited, distributed, and swiftly applied in practice.

### *A. Post-Daubert: Stringent Reliability Requirements in the Absence of Implementation Solutions*

The scientific community concerns were amplified by rather heavy evidence reliability discussions in the legal domain. Back in the days, the US Supreme Court formulated the Daubert‘s rule [2] that requires the forensic theory or technique to be tested, peer-reviewed, generally accepted in the scientific community, and accounting error rates. Moreover, the attention drawn to wrongful convictions based on unreliable expert evidence [10], required the legislators to elaborate further on the Daubert‘s criteria. In the last years detailed reliability requirements were introduced in academia, forensics, and standardization bodies [11].

*Edmond* is discussing the issues with "effectively contesting expert opinion" and difficulties for practical verification of forensic results [12]. Some of the factors legally recognized to render the expert evidence unreliable are*: (i) if* the opinion is based on a hypothesis which has not been subjected to sufficient scrutiny (including appropriate experimental or other testing), or which has failed to stand up to scrutiny; *(ii)* if the opinion is based on an unjustifiable assumption or on flawed

[[af0]]Radina Stoykova is a PhD candidate in University of Groningen, Faculty of Law, The Netherlands and the Norwegian University of Science and Technology, Department of Information Security and Communication Technology, (adi.stoykova@gmail.com).

[[af1]]Katrin Franke is a Professor at Norwegian University of Science and Technology, Department of Information Security and Communication Technology, 22 Teknologivegen, 2815 Gjøvik, Norway, (katrin.franke@ntnu.no).

data; *(iii)* if it relies on an examination, technique, method or process that was not properly carried out or applied, or was not appropriate for use in the particular case [12]. Consequently, future legislative and scientific challenge is the lack of standard validation procedures during all stages of the digital forensic process. In addition, implementation solutions to document the scientific approach with its objective measurements as well as the assumptions and interpretations made are missing.

In addition, as part of the criminal investigation digital forensics must implement effective guarantees for procedural accuracy including cross-examining and verifying the reliability of expert evidence. Interpol stated as guiding principle that a "record of all actions taken when handling electronic evidence must be created and preserved so that they can be audited. An independent third party should be able to repeat those actions and achieve the same results"[13]. This is not just a requirement in digital forensics scientific research, but in the everyday work.

Since criminal investigations require following criminal procedure and set a higher standard and burden of proof for the prosecution, not all digital forensics methods, algorithms, and advancements in machine learning are suitable for criminal procedures. There is the need to identify efficient and rigorous automated forensic methods, which however ensure data integrity, reproducibility and reliability of both the process and the result. In addition, all parties in the proceedings must have access to sufficient documentation on the forensic examination and knowledge to question digital forensic findings on relevant grounds. Other examiners must be able to replicate experiments when challenged in court or for research purposes.

This paper suggests that the representation in standard format of data necessary for validation of digital forensic processing will directly implement the legal reliability requirements, while it will not cost significant effort or time for practitioners.

### B. New Validation Problems in Automation

An additional complication to quality assurance is the shift towards automated processing in digital forensics. More automation is needed to deal with the complexity and data-volume explosion [14] that cannot rely on "time consuming ad- hoc verification" but requires formal validation procedures [15]. Currently, the scientific validation in digital forensics is predominantly focused on tool testing [16]. Often objective measurements for the forensic processing, reproducibility documentation, and quality of the decisions taken by the examiner are missing [17]. The current situation will be further complicated due to advancement in new research focused on machine-intelligence approaches to utilize computer power and reduce the information overhead in digital forensics [18]. In order to develop scientific validation methods for computational forensics and data analytics, first step is to represent such methods in standard, auditable way.

Data examination are based on combination of assumptions, expert knowledge, context data, and tools and algorithms to assist the examiner in identifying relevant artefacts. It depends on the available data for analysis, the adequacy of the computational method selected, but also on the skills and knowledge of the forensic examiner. Ergo two analysts can come to different conclusions about the same set of data. Moreover, skillful digital forensics specialists know the limitations of each forensic tool and often employ different tools for the same task to overcome these limitations. Tools retrieve different data in response to a given input parameters. Examiners have learned to take into account system-specific behaviors in order to maximize the success of e.g. search algorithms. This locks the specialists to learn complex tools, to adapt them when inadequate (e.g. additional scripts), and often purposely to leave out information knowing that this will degrade the processing quality – but this reasoning is not recorded or represented [19]. It is regretful that examiners miss minimum agreed standards for the documentation of the forensic examination to abolish tool dependencies. Moreover, a lot of forensic tools are built with extensibility in mind, which allows specialists to fine tune them, to add scripts, batch files, and plug-ins according to the case specifics and the data to be analyzed [3]. This favors automation however it also requires transparency and accuracy evaluation.

Current and future (semi-) automated forensic analysis will benefit from a standard representation for quality assurance and validation that is based on data model and tracing back certain analytical steps. Considerably, the representation of the automated tasks must ensure reproducibility and reliability where the same input and set up will generate always the same results, while the human interaction and judgement in the loop is also represented.

## II. II. RELATED WORK

Further, examined are the legal and scientific requirements for validation procedures in the digital forensic process in order to derive criteria for data-representation model.

### A. Features and Algorithms Validation

Each examination task first requires extraction and selection of reliable data characteristics. This will bring transparency in the data selection, reduction or recovery and can assist in algorithm verification and reproducibility of the method. It will serve proving that data is processed correctly, no data is omitted randomly, and the procedure is controlled by the examiner. Forensic feature representation will allow the examiner to develop new tools and document the investigation task and context. This will enable not only a well-known tools or procedures to be accepted in courts and in the forensic community, but also new methods to be testable and prove their utility over the old ones.

For example, *Garfinkel* describes forensic features extraction as "searching files and unallocated sectors for strings that match user-specified regular expressions", including compressed and encoded data [20]. He further defines pseudo-unique identifier properties of different data structures but argues that the relevance of features depends on the concreate investigation question we need to answer [21]. A rich source for feature selection is also meta data like file name, path, size, time stamps, fragmentation, status flags and hash codes [22]. However, often the forensic value of such features cannot be predetermined and often enquires enrichment with case specific forensic features.

Further, the reliability of such enrichment is an important pre- processing step for applying machine learning models in digital forensics. Issues in computational forensics are related to inappropriate selection method e.g. insufficient detection of statistical properties representative for the data set and heuristic feature search strategies [23]. In digital forensics for criminal investigations the quality standards are higher, which requires documentation of the feature selection method, processing algorithm, assumptions made, and full false negatives accuracy due to the danger of missing relevant for the case data. Further, the forensic method and results must be evaluated by digital forensics specialist before training the forensic algorithm. Standard representation of the automated processing parameters and feature selection are at the core of meeting the legal requirements for risk assessment of the computational design related to accuracy, reliability, objectivity, explainability, and accountability [4].

In addition, evaluation of the suitability of the algorithm for forensic purposes and court proceedings is needed. In criminal investigation the algorithm logic needs to be fully traceable and human understandable, even if this require full source code disclosure. In contrast to heuristic algorithms, only deterministic algorithms can be validated and reproduced for forensics. Given the same input only deterministic algorithm produce exactly the same result every time. Since no random variables in the processing must be introduced, new approaches, for example, in neural networks with random initialization [24] do not meet the reliability standard required in court. Another limitation of computational methods, that should be considered in accuracy evaluation, is that some algorithms work on linear and other on non-linear feature correlations [23]. It becomes apparent that automation adds more complexity in respect to the reliability of the digital evidence as a final output of the process if not properly documented and explained [15]. Thus, new standards for digital evidence and its admissibility in court must also include documentation on algorithms transparency and reproducibility testing. However, full source code disclosure will result in overcomplexity problem, therefore a data model for validation purposes is preferable.

### B. Validations Designs and Standard-Expression Criteria

Further compared are one legal and one scientific design for forensic process validation, to exemplify that requirements for reliability testing from both domains could be satisfied with standard data-representation.

In legal taxonomies reliability of expert evidence is used as synonym of trustworthiness and reproducibility – a "test that produces the same results on successive applications is said to be reliable, while a test that produces accurate results is said to be valid" [24]. Risinger elaborates on the requirements for legal validation of core forensic processes [25], which are adapted and extended in Table 1.

Table 1. Legal validation design

Those general requirements for legal validation of the forensic processes are elaborated further in a scientific validation design to derive data-representation prerequisites. The FRED model

[25] provides a framework for validation and evaluation in forensic interpretation of digital artefacts. The model closely examines the validation of digital forensic practices according to ISO and ENFSI requirements [26], [27]. It is extended in Table 2 with data-representation model to ensure practical implementation, which is not burdensome for the examiner.

Table 2. Description of each stage of the FRED design [25] and the derived requirement for data-representation model *in italic*

Both the legal and the scientific validation designs can be better satisfied with additional data generation, integrated in the process to enable validation of the forensic actions. Thus, this study considers that a data model for digital evidence representation is suitable to meet reproducibility and reliability requirements in court.

In summary, a standard representation of the forensic examination process must be further elaborated to support:

•**Automated processing set up** (e.g. feature and algorithm selection method, control parameters, mechanism to detect error rates, parallel processing, and accuracy);

•**Human intervention** (e.g. hypothesis, assumptions, parameter input, and output interpretation);

•**Evaluation/ Testing set up** (e.g. event reconstruction, trace back representation of dynamic recomputing, retesting and chains of processes).

It is insufficient to only require compliance with the *Daubert* or similar criteria for reliability in digital forensics. To ensure effectiveness and procedural economy, the validation must be practical and expedited. Therefore, *Daubert* must be implemented in the forensic process itself by generating data necessary for such evaluation.

Standard data-representation can ensure reproducibility and transparency in every step of the digital forensic examination and thereby protect defendants‘ rights by guaranteeing that digital evidence presented in court is properly tested, documented and traceable. It will allow prosecutors to be transparent about reasonable lines of inquiry. In addition, the defense can contribute to the investigation by providing alternative hypotheses for testing also in an early stage.

## III. III. SEMANTIC INTEROPERABILITY: CASE AND WANDA

The goal for digital-evidence standardization is a flexible standard, which can ensure security policies, information exchange, cooperation, and storage (*physical evidence preservation*). New methods, algorithms, parameters, and functions must be traceable in provenance and chain of custody records (*logical evidence preservation*).

In order to achieve broader consensus and international standardization of digital evidence, building on preservation, integration, and harmonization of existing models is preferable and does not exclude innovation. In this paper only two frameworks are examined – WANDA and CASE. They both are building on previous standard representations in forensics to achieve sustainability and harmonization [5]. The objective of both frameworks is working towards reliability in digital forensic processes by aligning terminology and semantics. They also account for the fact, that evidence is context dependent data – the storage media has physical context, the digital data has logical context, and combined they produce information in legal context [28]. Both frameworks will enable computer-assisted decisions in the investigation of crimes as it can be understood by human and machines alike.

The CASE framework and its compatibility with the underlying *Unified Cyber Ontology* (UCO) [29] are based on years long effort for enabling open ICT ecosystems in different security domains and the development of knowledge-based tools [6].

CASE is artefact-oriented framework, which lacks expressivity on process level. WANDA framework has the advantage that it is designed to represent forensic science methodology and processes for the demands of court proceedings.

While this paper supports CASE as generic framework for evidence with a broader application level, it is further argued that the framework will greatly benefit from integration of a WANDA-based expressions. CASE scope is cyber investigations, not limited to court cases, while WANDA is specifically developed to validate the investigative actions of forensic examination. WANDA allows the documentation of forensic methodology for validation purposes irrespective if it is performed by human, machine (automated), or combination of expert input and computations (semi-automated). Unlike intelligence or security investigations, forensics performed for criminal investigation must meet the highest burden and standard of proof and comply with fair trial standards. This requires full documentation of the entire processes and accurate fact finding based on objective measurements.

## IV. IV. CASE LIMITATIONS: VALIDATION, AUTOMATION, REPRODUCIBILITY

CASE specification is a generic framework to support any type of investigation on a case level including related cases [7]. It aims to standardize the whole forensic process from the first seizure of an item to the relevant for the investigation data, including all actors, objects, actions, and tools with their relationships and characteristics. It aims also to develop complete provenance documentation and chain of custody. CASE may achieve completeness and transparency in the digital investigation process as a whole. The full CASE provenance record may enable cross-border exchange or joined investigations and corroborations of traces from different origins, including features from non-forensic tools, and aligning concepts with related domains in cyber security. CASE community is discussing gradual development of glossary, thesaurus and ontology.

Until today representing automated processing is not mentioned in any CASE documentation or related workshop. There is still no clear differentiation between human and machine actions. Opinions within the CASE community were expressed, that full replay of examiner actions is not required. They argue it is sufficient detailed specification of the tool (version and configuration) and the representation of results. The argument is that complex tools like *Encase* [30] cannot be easily reversed, and such validation so far is not required by courts. On the contrary, in the current version CASE does not meet the legal requirements for cross-validation of automated and semi- automated forensic tasks. Currently, even if the courts do not examine the automated processing of data with the fast advancement in this field this will be changed. Legal attention is drawn on protecting individuals from arbitrary automated decisions, validation of algorithms and results, as well as their effect on human rights. Inferential analytics and their impact on decision making are already provoking debates with respect to the data protection legislation [31]. Considering the growing digitalization of society and the impact of digital evidence on investigations, it is hereby argued that a data-representation model, which does not meet legal requirements for court proceedings will be of limited value.

Currently, CASE ensures full documentation of the acquisition phase and will assist further forensic analysis in many ways e.g. tools output comparison, dual tool validation, elastic search and cross-device analysis. However, the current version does not include representation of data during digital forensic experiments and automated investigation tasks, nor separates machine and human operations which will be addressed further by the proposed integration of WANDA validation standard.

textitA. Trace and OriginTrace is an *Object* in CASE that can represent physical device, data structure, context data, or data embedded in data. The *duck-typing model* does not impose strict inheritance but allows the object to be defined by its characteristics (property bundles). To what extend the *duck-typing model* is suitable for the expressing exponential number of digital artefacts and characteristics, and how it will fit the hierarchical structure of ontology remains to be proven by the CASE community.

The upside of the data model is that new types of traces and their properties could be described easily, the link between the physical device and the data is preserved. It builds a tree structure where explicitly one can track the modifications from the source data to the relevant for the case information. Further, native for the trace metadata is represented as reference, while additional metadata from origin or location is added as a relationship property. This can be problematic for classification and inheritance structures in ontology reasoners. As observed in II.A features depends on the application domain and their relevance is determined by the context and the investigative questions to answer. Another challenge is the interoperability with tools that already have data models with standard expressions for more common traces. Arguably, in order the standard expression to be consistent, it should be focused on procedural and not on artefact level. Further we argue the validation documentation should map how the examiners decide according to their task how to normalize and pre- process, identify traces or analyze them further. In CASE there is no standard expression how the pre-processing for trace selection is done, which cannot help with reliability evaluation, while also the examiner interaction is not mapped.

### A. Investigation Actions and Provenance Records

Currently, CASE makes a strict separation of representing investigation action and traces. The investigation action is an activity with properties: UUID, location, performer, tool, forensic environment and object (e.g. mobile phone). The investigation action could take as an input a provenance record about device or output of another tool. It is stated that "*Forensic Action* can output other *Forensic Actions*, such as when an automated tool launches module to process *Objects*" The example is extracted mobile data, from which the *mmssms* repository is further parsed [1, Expl. 17].This schema provides provenance records of inputs and outputs, but does not support reproducibility, because does not keep track of the parameters for the processing.

Moreover, it is suitable for data examination from phone or computer, but does not fit triage, incident response or network forensics objectives. In network forensics required is the representation of the logic of processes and not necessarily representing in full all the observed nods. For example, if incident-response analyst conducts malware search, what is important to be represented is the total network topology (input) with all the machines that has been searched and the search functionality (parameters). And as result (output) the infected machines. We argue that full representation of all machines is not required. Not always representing all the results from processing in full is necessary, but if the logic from the processing is transparent, can be reconstructed. In other words, one represents only relevant for the investigation results, where irrelevant results are implicit, yet can be reconstructed on demand. The chosen approach for representation of investigation action and provenance record introduces redundancy that can challenge the maintaining of consistency. This leads to exponential growth in complexity which is error prone. In order to address those limitations, we propose further work on representing the logic of the processes and explicit representation only of relevant data. This is of utmost importance for representing automated and semi-automated investigation actions. It should be emphasized again, that data representation of investigation actions must include additional standard expression of the human and automated scientific methodology, which is necessary for reliability and reproducibility testing.

Even though parallel processing could be represented in CASE, there is no option to document chaining of automated processes. This prevents reuse of certain modules (e.g. a batch file template) or modifying them slightly according to new tasks. Based on the limited documentation about CASE, it could be argued that it does not sufficiently represent the forensic methodology for court proceedings. The framework does not differentiate between human interactions and machine input – output, lacks expressivity for documenting automated tasks, is at times redundant and error prone, and unsuitable for representing automated processes in forensic investigations. In addition, CASE is focused on artifacts representation, while it has limited expressivity on processes level.

Without negating the significant contribution of CASE standard in its broader scope, it could benefit from other more custom- made for a specific task standard. Therefore, we further argue for extending the investigation action in CASE with WANDA as a validation data expression for forensic methodology. First considered are the derived requirements for reliability and standard-expression objectives from the *FRED* and *Risinger's* validation designs in *Section II.B*. This will allow to generalize WANDA and create module for CASE that documents forensic experiments conducted in digital forensics labs. The WANDA module achieves standard-representation for reliability and reproducibility validation of forensic examination.

## V. V. THE WANDA VALIDATION SCHEMA

WANDA is a generic framework that has been developed between 2001- 2004 in response to the criticism in the Daubert case [8]. In European context, it is aiming at establishing a scientific base in the forensic examination.

Although WANDA data standard served initially for forensic analysis of handwritten samples, the framework is developed around a number of concepts with extensibility and flexibility in mind that can benefit other forensic domains. The focus of the research was to discover objective measurements for deterministic algorithms and to develop a model which ensures reproducibility for automated processing [32], in order to overcome the predominant subjective and opinion-based evaluations of forensic samples. Therefore, WANDA turned into data standard for annotation, processing, and storage of digitized handwritten documents. Most importantly this schema allows proof and verification that a tool is treating all the input data in the same way, does not omit any data and process everything according to the forensic objectives, and does not serve to personal or corporate interests.

One of the greatest advantages of WANDA over other XML schemas, is that it has process-oriented design to trace back and represent forensic data processing by separately representing human and machine-generated annotations. While both CASE and WANDA support standard data format and reproducibility of tool results, WANDA has additional expressions for investigative actions: objective measurements and forensic feature extraction; representation of computer-supported forensic processing and reproducible results [33]; modularization and extensibility in system concept representation [9]; separation between data, GUIs and processing modules; quality control in the analysis work for both examiner and tool; data repositories and working sets for fast remote data retrieval. This shows the major advantage that WANDA is process-oriented representation, while CASE is artefact-level oriented. By focusing on processing, WANDA allows to fulfill the high reliability and reproducibility requirements for court proceedings.

## VI. VI. CASE AND WANDA INTEGRATION

Having expressed the need of CASE and WANDA integration to meet the reliability and reproducibility requirements in court proceedings, we further elaborate the WANDA framework and its generalization as CASE module.

(Semi-) automated data extraction or data searches could negatively influence the decision-making process, since humans could introduce bias in the algorithm selection or influencing the processing with control parameters, even though the algorithms are functioning [34].

Therefore, it is required that we can trace back the actions performed on the data, including the logic behind automated or human based assumptions, input – output relationships, and judgements. *Forward mapping* also referred to as task planning of investigative actions will assist the forensic expert to track and correct steps in the analysis. *Backward mapping* also referred as audit trail is necessary for reproducibility and court evaluation. WANDA schema can be used for recording of the analysis operations performed on the data and reconstruction of the examination phase, even when multiple analyzing tasks are automated. Human interaction with the system is taken into account in WANDA as algorithm configuration or expert annotations. It is expressive enough to detect errors and assumptions done by the examiner, or to correct errors by repeating certain steps in the investigative process.

In CASE annotations can be keywords or free text, which can be added by both a person and a tool in addition to provenance records about investigation actions. WANDA introduces separate specification only for machine-generated annotations called *Filters*. *Filters* are defined as computer programs that process a data set and returns either process data or a set of trace characteristics. The term *Filter* could be substituted by *Tool*, used in CASE, because it could bring wrong association, that data will be always filtered out. On the contrary, WANDA *Filters* is generic term for automated processing that can represent enhancement, restoration or decryption operations as well as data-reduction tasks. Hereafter for clarity we refer to *Filters/Tool* representation schema .

If integrated in CASE the WANDA *Filters/Tool* concept could provide better and more detailed specification for automated forensic tasks. This will bring clarity of which tasks are automated and allow their representation e.g. pre-processing settings, algorithm and control parameters, extracted forensic features, and further filtering. This will omit current needs in CASE of verbosely repeating the same information, since the provenance record will represent only the logic of the processing and relevant results. This is of high importance for the reliability and reproducibility validation of the forensic examination.

WANDA links human annotations to the processing task, but does not mix them together in order to avoid any ambiguities. This allows also to identify on which phases the automated processing needs to be validated by human expert, which is a crucial safeguard required by law, when automation is presented in decision-making process.

In cases where no suspicious digital data is found or the analysis give insufficient or inconsistent results, WANDA schema could represent the performed searches and the absence of evidence, while in CASE no such representation is suggested.

Further, output of forensic actions in CASE can be taken as an input for other forensic actions. To some extend this represents how automated tool launches modules to process *Objects*, but does not describe the modules and no apparent schema exists for parallel or sequence of automated processing within CASE. WANDA, on the other hand is representing these in specifics.

Fig. 1. Keyword search in updated WANDA with controls.

### A. Implementation examples

To exemplify the utility of WANDA over current CASE investigation action representation, we developed a simple keyword search expression in Figure 1.

This structure fully supports reliability validation because it represents the input, the search method and algorithm parameters. The input defines the search area of a disk image and the keywords. Additional control parameters are set by the examiner and the type of executable module is elastic search v.7.6. The output features (match location, path, offset) could be also expressed in detail but here for simplicity are represented as .csv file. This representation is crucial for the court of law since all parties in the proceedings can verify the type of search algorithm selected, any errors or bias introduced with the key words or control parameters, and the exact version of the search module. Since the court proceedings could take years and the technology could be updated, such specification is required. The keyword search in this auditable form is fully reproducible and could be used for ranking algorithms for forensic search hit relevancy as stressed in [35].

WANDA structure is suitable for representing any type of automated or semi-automated forensic methodology. The elements $< \text{input}/ >< \text{output}/ >$ are generic and due to the simplicity of the structure new types of *Filters/Tool* inputs and outputs could be specified. Input could be user-defined or the output of another *Filter/Tool*. The *Filter/Tool* schema allows play back of the data examination that is contained in filter cached results. This allows the examiner to save certain steps of the analysis and repeat the experiments from a certain point. The CASE *Provenance Record* for output of forensic actions serves as output unique identifier, which could be aligned with WANDA *Filters/Tool* identifier to support dynamic recomputing the results of data filtering.

WANDA creates a formal standard representation for any type of (semi-) automated processing or parallel processing in order to enable testing and validation. Further, this standard expression reveals the logic of forensic actions, implicitly in a simple tree structure, that can be partially or fully committed to repository or used in further processing. Potential shortcomings with *Filter/Tool* processes must be identified and if known, additional measures to mitigate them could be recorded at any stage. The *Filter/Tool* representation allows examiners to record testing of different hypothesis and correct errors. Moreover, they record explicitly the assumptions and restrictions made before or after every (semi-) automated processing step. This is of utmost importance for algorithm auditing, as proposed by Mittelstadt [36]. He defines the need of transparent explanation why a new input was assigned a particular classification, and concludes that in most cases "reporting only the features of data relevant to the classification may be sufficient”. This significant for law requirement is fulfilled by the WANDA framework.

CASE-WANDA integrated module can be used for statistics, standardization of machine processing, and forensic processing templates.

### *B. Control Parameters Integration*

The original WANDA schema is aiming at maximizing simplicity in the play back sequence of elements. Therefore, the $<$ input $>$ includes the set of data for processing and any control parameters added by the examiner. The list of input parameters can be extended according to examination task. For reproducibility of the automated processing, it is indeed sufficient to document the control parameters in order to make them verifiable, but this does not represent the logic behind choosing specific control parameters. If more complex controls are involved in data reduction, data mining or triage models – it could be beneficial to have clear representation of the control parameters separately from the input data (compare Fig.1). For example, Quick and Choo are defining data reduction by selective imaging based on predefined parameters [37]. The same complexity in choosing control parameters is followed from data enhancement or enrichment techniques. Moreover, by choosing control parameters there is a high risk of omitting relevant or exculpatory evidence. Therefore, one suggested update in WANDA is to represent control parameters separately, as they might highly impact the investigation action, the quality of the output data [2, p. 52]. Behind this argument there is a long-overlooked theoretical challenge, which has major practical implications for fair trial guarantees. Is it necessary to represent the control parameters as part of the working hypothesis, or the hypothesis testing should be done on a higher level as part of the investigation process and not represented in the design of the forensic action? There is high dependency between the automated processing accuracy and the original algorithm configuration – the assumptions and judgements before, during or after the processing stages could introduce errors or bias. However, being able to represent the control parameters gives one a chance to detect examiners errors during cross-examination if one can ensure that the results will receive sufficient scrutiny afterwards, which is not always the case in practice. There is a certain trade-off between striving for exact documentation, while ensuring effective processing, since the system must be agnostic enough for different forensic tasks and simplicity is essential.

### *C. Validation of Plug-ins, Scripts, and Filters*

One CASE example represents the forensic path for a decompressed with *Bulk-extractor* e-mail [10]. What CASE represents is the decompression with a tool (only by name and version) the email found (offset location, hash etc.), but it does not represent how the tool was configured and thus which algorithms has been used. What *Bulk-extractor* did is started one scanner for decompression with certain parameters for input, took the output of the first scanner and a second carved the email. This can easily be represented with WANDA *Filter/Tool* module and its control parameters.

WANDA is designed for accumulating sufficient knowledge about forensic features and processing paths enables the development of custom or task-specific plug-ins and scripts for assisting data interpretation. Moreover, they can run in parallel or call automatically other filters. For example, Bulk extractor is dividing the disk image into chunks and runs multiple filters (plug-ins called scanners) each searching for different features. Some filters (zip, rar, jpeg) are recursive - they output (expand/find) data for other filters to process [20]. WANDA representation for chain of filters fully supports that. Forensic examiners could write scripts with sequence of processes to automize certain investigation tasks, and record them in WANDA standard expression. Further, the sequence of processing could be used as batch files for similar cases or exchanged among analysts[2, p. 28].

The WANDA *Filter/Tool* module allows forensic practitioners and researchers to refine the sequence of routine working stages in order to obtain consistent and further advance forensic processing. The defined protocol will then be cast into an analysis wizard that could guide the forensic expert step by step. Further, the filters can be used as pre-configured machine template to assist in efficient processing. As the template is fully documented it can be inspected or customized on a later stage.

At this moment the proposed standard representation is focused on (semi-) automated processes, so it does not represent the strength of the hypothesis tested. However, it makes transparent the configuration in (semi-) automated processing, the input and output data sets, intermediate results and error mitigation, fulfilling the international requirements for reproducibility and reliability [36],[37]. Further, ISO 17025:2017 was criticized for imposing burdensome and unsuitable for digital forensics labs requirements (Sommer, 2018). Therefore, this representation model can speed up and automate compliance, while optimizing the work of the practitioners. Moreover, the *Filter/Tool* schema fully documents tool limitations, allowing the examiner to develop own scripts, batch files, and record interactive processes. This will allow the examiner to exercise full control over the tool.

### *D. Annotations*

It is beneficial to separation between human and machine-generated annotations to document objective or possibly-biased data processing, track back, and reproduce the results of the forensic actions. WANDA is using the term *Annotations* only for human interactions and is proposing sets of qualitative categories or quantitative measures [11]. Examiner annotations are recorded with the machine processing but separated and under controlled vocabulary. The controlled vocabulary aims to avoid terminology and interpretation differences in descriptions among experts. Contrary, CASE deals with great variety of traces and actions and introduces rather free labelling, grouping, adding general notes, comments, and bookmarks to an *Object* or group of *Objects*. It is unclear from the documentation, if CASE will introduce fixed or semi-controlled vocabulary for certain annotations in order to avoid lexical ambiguity. Human annotations support database queries, case documentation, reporting and statistics [2, p. 3.2.3] and too verbose annotation schema, could defeat the purpose. Vice versa too little or missing annotation will hamper transparency and reproducibility.

Another suggestion for CASE is to adapt the annotations for legal evaluation. *Bodington et. al.* developed a digital evidence validation model in order to align the inferences made by the forensic examiner based on the data integrity and their logical expression for legal arguments [39]. . *Bodington et. al.* paper is proposing a standard logical structure to represent human annotations with the weight assigned to it. The terms "*confirmed by*" and "*negated by*" must not be understood binary. What this paper suggests is that the standard must have confidence expression which supports different frameworks for probability evaluation [40], but does not define the certainty level.

On the same *Trace* one can have both "*confirmed by*" and "*negated by*" *Annotation*, followed by a link to the process in support of the statement. Linking the sources which confirm or negate an existence of a fact, will allow the examiner on a later stage to apply different methods to estimate probabilities based on the links and with respect to the whole data analyzed. This will be a significant improvement to the proposed in CASE confidence level schema. It is purely user-defined by choosing from control list from 1 *"confirmed by other sources"*, to 2 or 3 *"probably or possibly true"* to 6 *"truth cannot be justified"*. Representing only corroborating links which support or negate the existence or nonexistence of a fact, that allows later customizing the probability assessment according to the task and features at issue and is preferable then a fixed schema with certainty levels without clear definitions.

The integration of CASE and WANDA will give sufficient representation of human and machine-generated annotations, that can contribute to evidence platforms development and improving of case management systems. This will ensure quality control, reduce subjectivity in the scientific work and improve accountability.

## VII. VII. CONCLUSION AND FURTHER WORK

This paper argues that meeting higher legal and scientific reliability requirements in digital forensic examination for court proceedings will greatly depend on standard representation and play back recording of automated and semi-automated processing. The examined legal and scientific validation designs allowed to derive standard-representation criteria to satisfy both. Further, to build up on previous standardization efforts, we propose the generalization and integration of WANDA schema into CASE expression, breaching the missing representation elements. While CASE is artefact-oriented framework, WANDA is process-oriented. The simplicity of WANDA model allows representing of any (semi-) automated processing or parallel processing in order to enable testing and validation for court proceedings. It provides quality control in forensic examination by documenting objective measurements, input – output relationships, algorithm configuration and sequence of processes which can be recorded, reproduced and reused for interactive batch plug-ins. This support database queries, case documentation, reporting, and statistics. WANDA module ensures also the separation of human and machine-generated annotations. In addition, proposed is an advanced confidence-level schema for CASE. The CASE- WANDA integration ensures compliance with international reliability and reproducibility standards without extensive burden in the practitioner day-to-day work. It will guarantee protection of defendants‘ rights by documenting that any digital evidence presented in court is sufficiently tested.

Further work to advance the specification language in digital forensic analysis and meeting the reliability requirement for criminal proceedings is related to refining the data models, representing more complex forensic methodology including computations and machine learning in digital forensics, and information exchange policy specifications. Evidence platforms and storage architecture should benefit and build on existing standard expressions and the CASE generic framework. Future efforts on achieving semantic interoperability are of great importance for developing knowledge-based tools, reliable computer-assisted decision making and reproducible forensic computations in the investigation of crimes with digital elements.

$^{S}$ubmission date 10.03.2020. This work was supported by Maria Sklodowska- Curie grand as part of Project ID: 722482 – ESSENTIAL – Evolving Security Science through Networked Technologies, Information policy and Law: https://www.essentialresearch.eu/

[1] Summary of the discussion: https://en.wikipedia.org/wiki/Replication_crisis.

[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 Bulk extractor customized plug-ins and development as described in http://downloads.digitalcorpora.org/downloads/bulk_extractor/BEUsersMa l.pdf.

[4] For example, NIST (2019). U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools, p.8. http://www.fatml.org/.

[5] CASE is based on Cybox and DFAX and elaborates on other standard expressions, see Casey E, Back G, Barnum S, Leveraging CybOX to standardize representation and exchange of digital forensic information." Proceedings of the 2nd annual DFRWS EU Conference, Digital Investigation, Volume 12, Supplement1, Elsevier, 2015, while WANDA was extension of FISH, see M. Philipp. Fakten zu FISH, Das Forensische Informations-System Handschriften des Bundeskriminalamtes - Eine Analyse nach über 5 JaWirkbetrieb. Technical report, Kriminaltechnisches Institut 53, Bundeskriminalamt, Thaerstrasse 11, 65173 Wiesbaden, Germany, 1996. In German.

[6] See https://caseontology.org/ontology/intro.html.

[7] See the example for a whole investigation case file at: https://github.com/casework/CASE/blob/master/examples/Oresteia.json.

[8] See footnote 2.

[9] For more information see also Wanda presentation at NIST:

https://www.nist.gov/system/files/documents/oles/MSSFAH-Franke-WANDA-Conference-Presentation.pdf

[10] See https://github.com/casework/CASE/blob/master/examples/bulk_extractor_forensic_path.json.

[11] WANDA *Filter/Tool* schema provides data that can be considered machine- generated annotations.

## References

[1] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Beek, and A. Nelson, 'Advancing coordinated cyber- investigations and tool interoperability using a community developed specification language', *Digital Investigation*, vol. 22, pp. 14–45, Sep. 2017, doi: 10.1016/j.diin.2017.08.002.

[2] K. Y. Franke *et al.*, 'WANDA: A generic framework applied in forensic handwriting analysis and writer identification', *http://www.clopinet.com/isabelle/Papers/his_paper_final.pdf*, 2003, Accessed: 16-Jan-2020. [Online]. Available: https://repository.ubn.ru.nl/handle/2066/63528.

[3] E. Casey, 'What does "forensically sound" really mean?', *Digital Investigation*, vol. 4, pp. 49–50, Jun. 2007, doi: 10/bh3dcs.

[4] S. Mocas, 'Building theoretical underpinnings for digital forensics research', *Digital Investigation*, vol. 1, no. 1, pp. 61–68, Feb. 2004, doi: 10.1016/j.diin.2003.12.004.

[5] National Research Council (U.S.), Ed., *Strengthening forensic science in the United States: a path forward*. Washington, D.C: National Academies Press, 2009.

[6] 'PCAST Releases Report on Forensic Science in Criminal Courts | whitehouse.gov'. https://obamawhitehouse.archives.gov/blog/2016/09/20/ pcast-releases-report-forensic-science-criminal-courts (accessed Mar. 06, 2020).

[7] Council of the European Union, 'Draft Council Conclusions on the way forward in view of the creation of an European Forensic Science Area'. Council of the European Union, 18-Feb-2016, Accessed: 27-Mar-2018. [Online]. Available: http://data.consilium.europa.eu/doc/document/ST-6078- 2016-INIT/en/pdf.

[8] A. Kloosterman *et al.*, '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.*, vol. 370, no. 1674, Aug. 2015, doi: 10.1098/rstb.2014.0264.

[9] S. Garfinkel, P. Farrell, V. Roussev, and G. Dinolt, 'Bringing science to digital forensics with standardized forensic corpora', *Digital Investigation*, vol. 6, pp. S2– S11, Sep. 2009, doi: 10.1016/j.diin.2009.06.016.

[10] 'Innocence Project', *Innocence Project*. https://www.innocenceproject.org/ (accessed Jan. 29, 2020).

[11] P. Sommer, 'Forensic science standards in fast-changing environments', *Science & justice : journal of the Forensic Science Society*, vol. 50, pp. 12–7, Mar. 2010, doi: 10.1016/j.scijus.2009.11.006.

[12] G. Edmond, 'Is Reliability Sufficient? The Law Commission and Expert Evidence in International and Interdisciplinary Perspective (Part 1)', *The International Journal of Evidence & Proof*, vol. 16, pp. 30–65, Jan. 2012, doi: 10.1350/ijep.2012.16.1.391.

[13] INTERPOL, 'Global guidelines for digital forensics laboratories, 2019: 2.3.e – Principle of electronic evidence, available at: https://www.interpol.int/en/content/download/13501/file /INTERPOL_DFL_GlobalGuidelinesDigitalForensicsL aboratory.pdf .' .

[14] L. Caviglione, S. Wendzel, and W. Mazurczyk, 'The Future of Digital Forensics: Challenges and the Road Ahead', *IEEE Security Privacy*, vol. 15, no. 6, pp. 12– 17, Nov. 2017, doi: 10.1109/MSP.2017.4251117.

[15] J. James and P. Gladyshev, 'Challenges with Automation in Digital Forensic Investigations', *undefined*, 2013. /paper/Challenges-with-Automation- in-Digital-Forensic-James-Gladyshev/6a3602e7181f9e828406b208e26b9248f0469 637 (accessed Sep. 01, 2018).

[16] B. Guttman, J. R. Lyle, and R. Ayers, 'Ten years of computer forensic tool testing', *DEESLR*, vol. 8, no. 0, Jan. 2014, doi: 10.14296/deeslr.v8i0.1963.

[17] H. Page, G. Horsman, A. Sarna, and J. Foster, 'A review of quality procedures in the UK forensic sciences: What can the field of digital forensics learn?', *Science & Justice*, vol. 59, no. 1, pp. 83–92, Jan. 2019, doi: 10.1016/j.scijus.2018.09.006.

[18] N. Beebe, 'Digital Forensic Research: The Good, the Bad and the Unaddressed', in *Advances in Digital Forensics V*, 2009, pp. 17–36, doi: 10.1007/978-3-642- 04155-6_2.

[19] M. Pollitt, M. Caloyannides, J. Novotny, and S. Shenoi, 'Digital Forensics: Operational, Legal and Research Issues', in *Data and Applications Security XVII: Status and Prospects*, S. De Capitani di Vimercati, I. Ray, and I. Ray, Eds. Boston, MA: Springer US, 2004, pp. 393– 403.

[20] S. L. Garfinkel, 'Digital media triage with bulk data analysis and bulk_extractor', *Computers & Security*, vol. 32, pp. 56–72, Feb. 2013, doi: 10.1016/j.cose.2012.09.011.

[21] S. L. Garfinkel, 'Forensic feature extraction and cross- drive analysis', *Digital Investigation*, vol. 3, pp. 71–81, Sep. 2006, doi: 10.1016/j.diin.2006.06.007.

[22] N. C. Rowe and S. L. Garfinkel, 'Finding Anomalous and Suspicious Files from Directory Metadata on a Large Corpus', in *Digital Forensics and Cyber Crime*, Berlin, Heidelberg, 2012, pp. 115–130, doi: 10.1007/978-3-642-35515-8_10.

[23] H. T. Nguyen, K. Franke, and S. Petrovic, 'Towards a Generic Feature-Selection Measure for Intrusion Detection', *2010 20th International Conference on Pattern Recognition*, pp. 1529–1532, 2010, doi: 10.1109/ICPR.2010.378.

[24] J. Bergstra and Y. Bengio, 'Random search for hyper- parameter optimization', *J. Mach. Learn. Res.*, vol. 13, no. null, pp. 281–305, Feb. 2012.

[25] G. Horsman, 'Framework for Reliable Experimental Design (FRED): A research framework to ensure the dependable interpretation of digital data for digital forensics', *Computers & Security*, vol. 73, pp. 294–306, Mar. 2018, doi: 10/gcx6dd.

[26] ISO/IEC, 'ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories', *ISO*, 2017. http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/69/66912.html (accessed Jan. 16, 2020).

[27] ENFSI, 'European Network of Forensic Science Institutes (ENFSI 2015). Forensic examination of digital technology', 2015. http://enfsi.eu/wp- content/uploads/2016/09/1._forensic_examination_of_d igital_technology_0.pdf (accessed Mar. 30, 2020).

[28] M. Pollitt, 'Computer Forensics: an Approach to Evidence in Cyberspace', in *Proceedings of the National Information Systems Security Conference*, Baltimore, Maryland, 1995, vol. 2, pp. 487–491, Accessed: 23-Apr-2018. [Online]. Available: http://www.digitalevidencepro.com/Resources/Approac h.pdf.

[29] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Beek, and A. J. Nelson, 'The Evolution of Expressing and Exchanging Cyber-investigation Information in a Standardized Form', vol. 39, pp. 43–58, Jul. 2018.

[30] 'EnCase Forensic Software - Top Digital Forensics & Investigations Solution'. https://www.guidancesoftware.com/encase-forensic (accessed Mar. 25, 2020).

[31] S. Wachter, B. Mittelstadt, and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', *International Data Privacy Law*, vol. 7, no. 2, pp. 76– 99, May 2017, doi: 10.1093/idpl/ipx005.

[32] K. Franke and S. Rose, 'Ink-deposition model: the relation of writing and ink deposition processes', in *Ninth International Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 173–178, doi:10.1109/IWFHR.2004.59.

[33] K. Franke, L. Schomaker, L. Vuurpijl, M. van Erp, and I. Guyon, 'A common ground for forensic handwriting examination and writer identification', p. 24.

[34] E. Bozdag, 'Bias in algorithmic filtering and personalization', *Ethics Inf Technol*, vol. 15, no. 3, pp. 209–227, Sep. 2013, doi: 10.1007/s10676-013-9321-6.

[35] N. L. Beebe and L. Liu, 'Ranking algorithms for digital forensic string search hits', *Digital Investigation*, vol. 11, pp. S124–S132, 2014, doi: https://doi.org/10.1016/j.diin.2014.05.007.

[36] B. Mittelstadt, 'Automation, Algorithms, and Politics| Auditing for Transparency in Content Personalization Systems', *International Journal of Communication*, vol. 10, no. 0, p. 12, Oct. 2016.

[37] D. Quick and K.-K. R. Choo, 'Big forensic data reduction: Digital forensic images and electronic evidence', *Cluster Computing*, vol. 19, Mar. 2016, doi: 10.1007/s10586-016-0553-1.

[38] ISO/IEC, 'ISO/IEC 27042:2015 Information technology — Security techniques — Guidelines for the analysis and interpretation of digital evidence', *ISO/IEC 27042:2015*, 2015. https://www.iso.org/obp/ui/#iso:std:iso-iec:27042:ed- 1:v1:en (accessed Apr. 04, 2018).

[39] R. Boddington, V. Hobbs, and G. Mann, 'Validating digital evidence for legal argument', 2008, doi: 10.4225/75/57b269e240cb7.

[40] Gladyshev , Pavel, 'Probabilistic Reasoning In Digital Forensics - DFRWS EU 2019 workshop', 2019.

https://dfrws.org/presentation/probabilistic-reasoning-in-digital-forensics/ (accessed Mar. 30, 2020).