Forensic Science International: Digital Investigation 35 (2020) 301014

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

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

Towards a conceptual model for promoting digital forensics experiments☆

Edson OliveiraJr [a, *], Avelino F. Zorzo [b], Charles Varlei Neu [c]

[a] Informatics Department, State University of Maringá (UEM), Brazil
[b] School of Technology, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil
[c] Computing Department, University of Santa Cruz do Sul (UNISC), Brazil

ARTICLE INFO

Article history:
Received 25 March 2020
Received in revised form
19 May 2020
Accepted 14 June 2020
Available online 22 August 2020

Keywords:
Concept map
Conceptual model
Digital forensics
Experimentation
Knowledge semantic-based model

ABSTRACT

Experimentation is one of the foundations for scientific evolution from the empirical point of view. Conducting experiments contributes to strengthen evidence of a given field mainly based on provided data and results, corroborated by repetitions, replications or reproducibility of an experiment, which altogether confirms or rejects pre-established hypotheses. Therefore, the proper conduction, documentation and dissemination of such experiments are essential to enable reproducibility. In this paper, we present ExperDF-CM, a conceptual model that aims to assist Digital Forensics researchers on planning, executing, analyzing and disseminating experiments. Such conceptual model was built based on almost two hundred analyzed Digital Forensics experiment papers and is mainly organized in five elements: Planning, Pre-Operation, Operation, Analysis and Interpretation, and Dissemination. We evaluated the conceptual model based on an evaluation survey and the Technology Acceptance Model (TAM) with researchers and practitioners of the Digital Forensics area. Results point out that our conceptual model is feasible for promoting reproducibility of Digital Forensics experiments, as well as it is easy to use and useful. Thus, our proposed conceptual model can contribute significantly to improve Digital Forensics experimentation and make them repeatable, replicable, and/or reproducible.

1. Introduction

Digital Forensics (DF) is the process of applying scientific methods to analyze stored information and to determine the events of a particular incident, thus making evidence usable in court (Raghavan, 2013). The DF process is composed of different phases, such as: Acquisition, Analysis and Examination, and Reporting. These phases are complex and deserve special attention on their performance, especially when handling digital data (Caviglione et al., 2017).

Different methods, processes, approaches, algorithms, tools, and several other artifacts are used in DF to allow researchers and practitioners to provide evidence legally available to be used in court (Perkel, 2018). To take the best advantages of such artifacts, they must be constantly tested and evaluated. One of the means to do it is by experimentation, which is one of the most recognized and used empirical methods to provide reliable and statistically tested evidence (Tedre and Moisseinen, 2014). Furthermore, experiments are high roads to consolidate researches by means of repetition, replication and reproduction of experiments. These activities strengthen reliability of data and provide a path to generalize results, for instance, based on meta-analysis (Hoffman and Hoffman, 2019).

By experimenting DF artifacts, researchers and practitioners evolve this research area providing experimental data and producing novelty by new findings (Casey, 2013). An increasing number of experiments have been performed in DF, which shows that it is a reliable area of investigation, as certain experimental design challenges still remain (Casey, 2013). However, most of the publicized studies are poorly reported in terms of experimentation details and its elements, such as Hypothesis, variables, threats to validity, sampling methods and experimental design. Moreover, the data used on the experiments are usually not available, lacking several important experimental elements to allow their repetition, replication or even reproduction.

To provide a way to promote reproducibility of DF experiments, we built a conceptual model to guide researchers and practitioners

☆ Preprint submitted to Forensic Science International: Digital Investigation June 16, 2020.

* Corresponding author.
E-mail addresses: edson@din.uem.br (E. OliveiraJr), avelino.zorzo@pucrs.br (A.F. Zorzo), charles1@unisc.br (C.V. Neu).

https://doi.org/10.1016/j.fsidi.2020.301014
2666-2817/

at properly planning, executing, analyzing, and disseminating DF experiments. Such conceptual model relates main concepts that should be defined in each experimentation phase to make it able to be internally or externally reproduced, thus prospectively performing meta-analysis and generalize results towards reliable and auditable evidence. To illustrate the concepts of such conceptual model we use identified excerpts from previous published DF experiments.

Therefore, this paper is driven by the following research question: **"How reproducibility of Digital Forensics experiments can be promoted?"**.

This paper is organized as follows. Section 2 discusses main literature on this work. Section 3 presents the conception and the design of ExperDF-CM. Section 4 presents a feasibility study of ExperDF-CM. Finally, Section 5 concludes this work and presents directions for future work.

2. Background and related work

This section presents a literature review on the main concepts related to experimentation in digital forensics and conceptual modeling, as well as related work.

2.1. Experimentation in Computer Science and digital forensics

Since the 80s, researchers (Mitchell and Welty, 1988) (Tichy et al., 1995) (Tichy, 1998) discuss the lack of experimentation in Computer Science due to different factors, such as, experimentation is inappropriate, too difficult, useless, and even harmful. They also report that most of the published papers did not provide evidence based on statistical methods and formal analysis. They claim a large part of Computer Science research consists of proposing new designs: systems, algorithms, and models, and they can only be objective judged on the basis of reproducible experiments.

In more recent works, e.g. Andujar et al. (2012), the authors argue that experimentation should be better understood and appreciated as a key methodology in Computer Science. According to Andujar et al., unfortunately computer scientists do not seem to agree on how experimental methods are supposed to impact their theory and practice. Despite the controversies about the scientific experimental method and its role, experiments possess some general features that are universally acknowledged and often are not even made explicit, i.e., comparison, repeatability and reproducibility, justification and explanation.

Nowadays, there are several initiatives in promoting experimentation in Computer Science. The reviews of Dieste et al. (2013) and Tedre and Moisseinen (2014) discuss how experimentation has grown in the Computer Science area and its importance. According to Dieste et al. (2013), experiments are performed for testing technologies related with quality and management of software and for analyzing outcomes related with effectiveness and effort. Besides, the major challenges faced by experimenters are to minimize the cost of running the experiment for the company and to schedule the experiment so as not to interfere with production processes.

Tedre and Moisseinen (2014) claim that experiments play a central role in science, being still unclear in Computer Science. Many questions on the relevance of experiments in computing attracted little attention. Nowadays, a variety of technically, theoretically, and empirically oriented views on experiments have emerged. As a consequence, computer related fields use experiments and experiment terminology in a variety of ways: feasibility experiment, trial experiment, field experiment, comparison experiment, and controlled experiment.

As a field of Computer Science, few Digital Forensic researchers discuss experimentation. Nance et al. (2009) propose the development of a research, education, and outreach agenda for Digital Forensics. In such agenda, experimentation is suggested as part of the future development for formalizing this research area.

Experimentation in Digital Forensics is also suggested in the work of Garfinkel (2010), which analyzes Digital Forensics research towards the next decade. They claim that "without a clear strategy for enabling research efforts that build upon one another, forensic research will fall behind the market, tools will become increasingly obsolete, and law enforcement, military and other users of computer forensics products will be unable to rely on the results of forensic analysis". They mention several Digital Forensics frameworks, however none of them provides the scale, hooks for experimentation. Furthermore, they claim that sponsors, researcher advisers and reviewers need new algorithms to be experimented with significant data sets, larger than a few dozen documents chosen from the experimenter's own system.

A study that examines the issues that are considered essential to discuss and resolve, for the proper acceptance of evidence based on scientific grounds is presented by Arshad et al. (2018). The authors also discuss challenges on the process of systematic validation of electronic evidence. Thus, they suggest open research areas, highlighting many of the issues and problems associated with the empirical evaluation of these solutions, which should be addressed by researchers and practitioners. They also review issues in the experimental validation of currently available practices.

2.2. Conceptual modeling

Due to the large amount of semantic information in Computer Science research, it is necessary to organize this knowledge in a structured and comprehensive way. Modeling information semantic is very important to effectively and efficiently manage and reuse the information generated during the whole applications development process (Wen et al., 2012; Novak and Cañas, 2006).

A conceptual model can be used as a communication tool between computer scientists and users to represent concepts and relationships between them. The goal of a conceptual model is to express the meaning of terms and concepts that are used by domain experts to discuss the problem, find the right relationships between different concepts and to communicate, to abstract and to calculate (Novak and Cañas, 2006). The conceptual model also aims to clarify the meaning of ambiguous terms, avoiding problems with different interpretations of the terms and concepts (Wen et al., 2012).

There are different types of conceptual models, which are classified according to different perspectives. The general perspective is divided into three types: Data Model, Process Model, and Behavior Model (Uthmann et al., 1999). Data model is a static conceptual model while process and behavior models are dynamic. In this paper, we focus our attention to static conceptual models due to their simplicity.

There is a bunch of conceptual data forms including (Wen et al., 2012): **Entity Relationship Model** (Chen, 1976) that provides a set of shapes and lines to deliver information; **Petri Nets** (Peterson, 1981) that graphically depicts the structure of a distributed system as a directed bipartite graph with annotations; **Unified Modeling Language (UML)** that includes class diagrams that may be annotated with expressions in a textual constraint language; and **Concept Maps** (Novak and Cañas, 2006) that are graphical tools for organizing and representing knowledge, including concepts enclosed in circled rectangles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts.

According to the functional view of conceptual models, Wen et al. (2012) classify them as: Structure-based Model, Object-

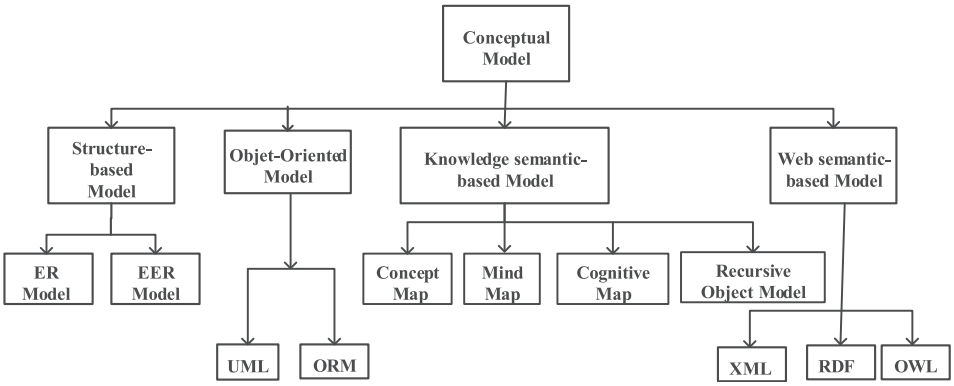

Fig. 1. Conceptual model classification from the functional view (Wen et al., 2012).

Oriented Model, Knowledge Semantic-based Model, and Web Semantic-based Model. Fig. 1 depicts such classification. An example of a conceptual map is illustrated in Fig. 2. We can observe circled rectangles representing knowledge in terms of concepts and their relationships indicated by connecting lines between two concepts. There is also a role represented as a label in such lines. In this example, Seasons "are determined by" the Amount of Sunlight.

2.3. Related work

To the best of our knowledge, there are no directly related works on conceptual modeling of DF experiments. However, in this section we discuss research towards experimentation in DF.

Casey (2013) discusses the importance of experimentation in DF. The author claims that designing good experiments is hardly a trivial undertaking. Experiments pose novel challenges towards avoiding mistakes and increasing scientific rigor in DF. Furthermore, the author also claims that flawed test results might provide incorrect decisions, potentially resulting in the loss of a person's livelihood, liberty or life. To reduce threats of incorrect conclusions, researchers must pay attention not just to the results, but also to the experimental design to assure reliability and repeatability.

Planning and conducting experiments in Digital Forensics is something that we have to accept as a possible substitute for unequivocal proof (Casey, 2013). Thus, when formulating an experiment to determine the cause or meaning of digital artifacts, thoughtful planning is required to eliminate irrelevant phenomena, and to evaluate individual causes separately. Planning includes the arrangement of the experimental environment to eliminate unwanted influences, and figuring out how to control and assess each variable separately. Specific configuration settings of a computer program or operating system can influence the findings. The version of the operating system can completely change the outcome of an experiment.

Casey (2013) states also that to obtain the clearest view of cause and effect when conducting experiments, it is also desirable to isolate each significant variable and test it individually, while holding the other variables fixed. In some circumstances, it may even be necessary to create the equivalent of a control group for a given experiment by conducting tests that help to distinguish normal usage of the computer system from the particular process that is being studied. Digital investigators survive on their reputations, and must make every effort to verify experimental findings for themselves. An error in a forensic report or expert testimony will be attributed to the digital investigator, regardless of whether the error was actually due to a mistake in someone else's work. Therefore, to be useful in Digital Forensics, experiments need to be reproducible. One must run the same experiment multiple times to verify results, and to exclude errors and extraneous variations. It is also important to keep in mind that systematic error or bias in an experiment can be difficult to detect, even between multiple runs of the same experiment. Ideally, different people will take different experimental approaches to test a theory – the more ways a Hypothesis is tested, the higher will be the confidence in its veracity.

Horsman (2018) proposes the Framework for Reliable Experimental Design (FRED) to support people engaged in the field of Digital Forensics research to contribute to reliable and robust findings. FRED focuses on the underpinning procedures involved within undertaking the reverse engineering of digital data structures and the process of extracting and interpreting digital content in a reliable way. The proposed framework is designed to be a tool for people operating within the Digital Forensic field, both in industry and academia, to support and to develop research best practices for reliable experimental design towards admissible evidence.

Both works contribute are related to this one in the sense they provide key concepts and stages during planning, conducting, analyzing, and reporting Digital Forensics experiments. Besides, they also discuss the importance of experimentation in Digital Forensics, similarly to Nance et al. (2009), Garfinkel (2010), and Arshad et al. (2018). Nance et al. (2009), for example, define a research agenda for Digital Forensics, including the need of development of scientifically rigorous repeatable experiments. Garfinkel (2010) discusses the coming Digital Forensics crisis and enforces the need of reliable and admissible evidence. For Arshad et al. (2018), to obtain actual evidence, it is essential that it can be explained and justified through systematic and rigorous experimental methods.

3. ExperDF-CM: a conceptual model for DF experiments

This section describes the **Exper**imentation in **D**igital **F**orensics **C**onceptual **M**odel (ExperDF-CM) in terms of its conception and design, as well as an application example.

ExperDF-CM is a static data model. Its conceptual data represents knowledge by means of a conceptual map with rectangles and connecting line links. We chose a knowledge semantic-based model due to characteristics of the DF experimentation domain.

The conception of ExperDF-CM is totally based on empirical results of a systematic mapping study on experimentation in DF.[1] Each concept of ExperDF-CM takes into consideration the main elements present in each of the 154 experiments performed in DF

[1] A list of 154 papers with DF experiments is available at https://doi.org/10.5281/zenodo.3516001.

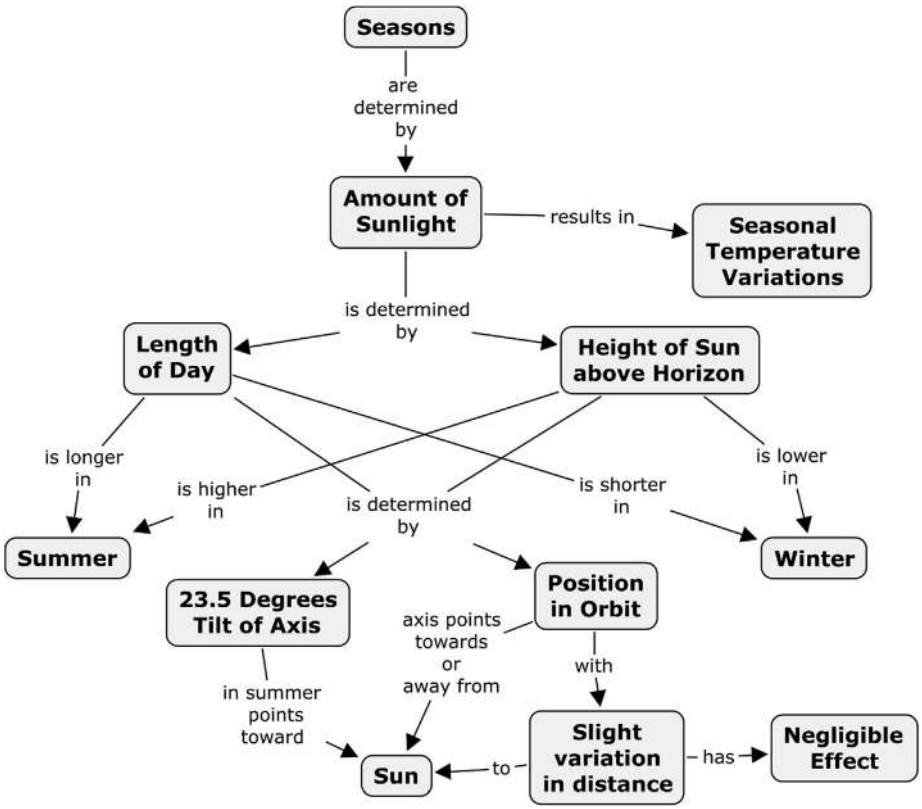

Fig. 2. An example of a conceptual map (Novak and Cañas, 2006).

research topics. As mapping studies cannot assure 100% coverage of existing studies, there might be some concepts not modeled in ExperDF-CM. In addition, note that this a first attempt to build a conceptual model for DF experiments, which we expect to evolve over the years.

DF experiments aim at establishing a cause–effect relation based on a theory and real-world observations by means of treatments and outcomes. To do so, four different stages are considered:

1. **Planning**, which is responsible for providing more control to the experiment by previously defining, for instance, goal, Hypothesis, instruments, and variables;
2. **Operation**, which executes the experiment according to its planning, mainly by collecting data with pre-defined instruments;
3. **Analysis and interpretation**, which provides data analysis mainly based on observed data (samples), such as descriptive statistics, and normality and Hypothesis testing;
4. **Reporting**, responsible for disseminating the experiment planning, operation, analysis and interpretation to interested audience, such as researchers and industry practitioners.

Thus, from such stages, we can start mapping essential concepts and relationships. Fig. 3 depicts the following DF experiment main concepts:

- **DF Experiment:** is the DF experiment itself. A DF Experiment *has* a relationship with planning and *runs* operation. It *provides* analysis and interpretation, as well as *reports* dissemination;
- **Planning:** represents the planning stage of DF experimentation;
- **Pre-Operation:** is a set of procedures ran before beginning the experiment operation itself;
- **Operation:** concept representing the operation stage of an experiment, most related with data collection;
- **Analysis and Interpretation:** concept related with statistical analysis techniques and limitation; and
- **Dissemination:** encompasses reporting experiments and provides a way to increase reproducibility of DF experiments.

The majority of DF experiments do not provide detailed descriptions on each of such stages, but we can observe that they are usually present in DF experiments descriptions along the text.

The following sections map and discuss each concept from Fig. 3. The complete ExperDF-CM view is available in Appendix 5.

3.1. Planning concept

Planning of DF experiments is very important to avoid deviations during the experiment execution, which could invalidate the data collection procedures. Fig. 4 depicts **Planning** and its related concepts.

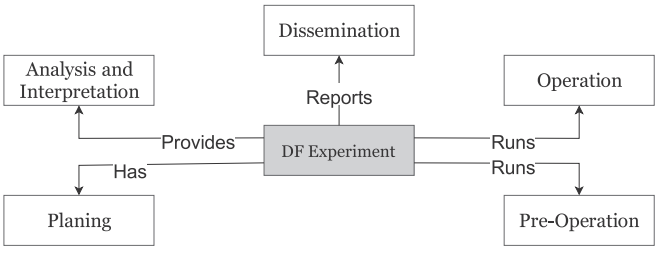

Fig. 3. ExperDF-CM main concepts.

It is important to properly define the **Variables** on every experiment, which can be **Independent** or **Dependent**. Dependent variables are the ones an experimenter wants to observe with relation to independent ones. The more independent variables values vary (cause), the more effects might be observable. On the one hand, usually, dependent variables are directly measurable, providing a clear **Dependent Variable Metric**. On the other hand, independent variables might be composed of a **Factor** assuming different values during an experiment, or a **Pre-Fixed** (constant) value. Besides, an independent variable can be a **Treatment**, which a variable can compare to, or a **Control**, which a variable is compared with.

As an example of independent and dependent variables, Kim et al. (2011) state that "*to compare the GPU-accelerated recovery by a brute-force attack with the CPU-based software approach, we measured the cracking time on a single GPU platform and a multi GPUs platform as...*", in which recovery "method" is a factor with a treatment ("GPU-accelerated method") and a control ("CPU-based software approach"). Another factor is the "GPU platform" ("single" versus "multi"). Dependent variable is the "cracking time", measured with a dependent variable metric in seconds (sec), minutes (min), or hours (hour).

Hypothesis. should be stated to provide a way to verify cause–effect relations. An experiment usually has a **Null Hypothesis** (H0), which always assumes statistical equality among samples, and at least one **Alternative Hypothesis** as the opposite of H0. An example of the definition of hypotheses is presented by Iqbal et al. (2008) when they want to "*...verify if the extracted write-print exhibits strong evidence for supporting the conclusion on authorship*". Iqbal et al. (2013) explicitly define hypothesis as "*...we perform a paired t-test on the data in* Fig. 3(a) *with the null hypothesis* H_0*:* $\mu_D = 0$ *and the alternative hypothesis* H_a*:* $\mu_D > 0$ *where* $\mu_D = \mu_{AuthorMiner2}$ μ*other method*".

Experimental Unit is the minimum element that allows the measurement of dependent variables values. Its definition is essential to understand how the dependent variables effect can be observed. Palomo et al. (2012), for instance, define experimental units as the "network packets" in a network forensics research on network visualization of traffic data as follows "*A set of network packets were captured from two different subnets from a university network with the aid of the WireShark program.*"

Design Type is a concept that directly depends on the number of factors and treatments/controls of an experiment. The higher the number of factors and treatments/controls, the more complex is the analysis of an experiment in terms of statistical tests. Wohlin et al. (2012) suggest the following Hypothesis tests to be performed to compare means of different populations according to factors and treatments (not limited to): one factor with two treatments, use T-Test or Wilcoxon (Forsyth, 2018); one factor with more than two treatments, use ANOVA or Kruskal–Wallis (Montgomery, 2008); and two factors with two treatments, use ANOVA. Taking into consideration the example of independent variables (previously

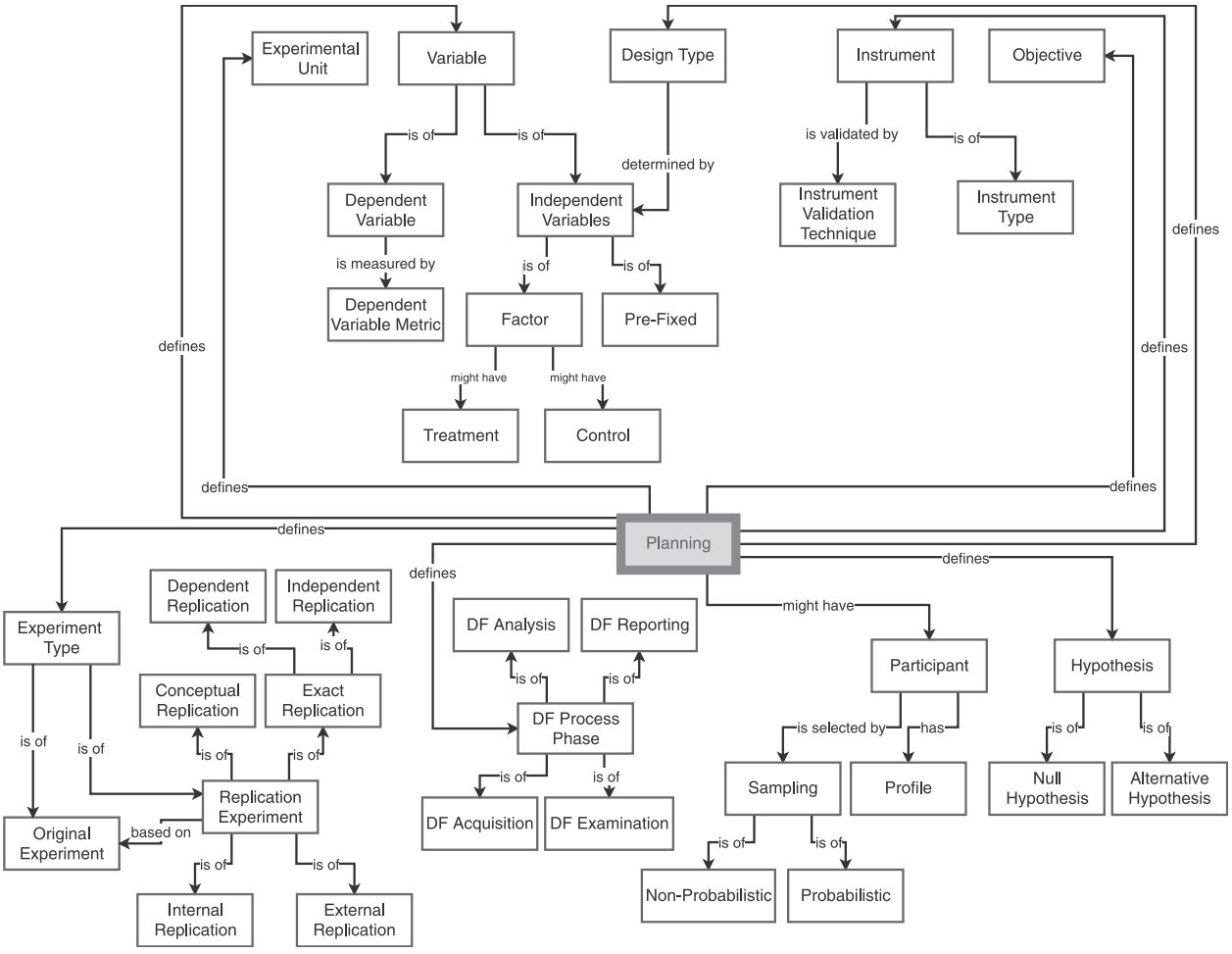

Fig. 4. Planning and related concepts.

mentioned), we could state the design type of the experiment is a 2x2 factorial as it has two factors ("method" and "GPU platform"), each of them with two treatments/controls: "GPU-accelerated method" and "CPU-based software approach" for the former and "single" and "multi" for the latter. Based on the design type of the experiment, ANOVA could be used for hypothesis testing during Analysis and Interpretation.

Instrument is a set of materials of several **Instrument Type**, e.g., source-code, questionnaires, or guidelines, that a participant handles during the execution of an experiment. They are essential to provide data from the action of participants and a way to keep the control of the experiment. Each instrument defined for an experiment should be validated by an **Instrument Validation Technique**, such as, Principal Component Analysis (PCA) (Abdi and Williams, 2010) or Cohen's Kappa Coefficient (Cohen, 1960). Hong et al. (2013) use a questionnaire as instrument for a selective search and seizure new triage model. Louis and Engelbrecht (2011) also used a questionnaire for discovering relations on textual data.

Experiment Type is a concept that can be an **Original Experiment** or a **Replication Experiment**. On the one hand, an original experiment is the one that was first conducted with no similar one previously conducted. On the other hand, a replication is conducted based on the original experiment. Replications could be performed as an (Juristo and Moreno, 2010): **Internal Replication**, by the same research group which conducted the original one; or **External Experiment**, by another research group, not involved in the original experiment. Replications can also be (Shull et al., 2008): an **Exact Replication**, following the procedures of an experiment as closely as possible; or a **Conceptual Replication**, which evaluates the same research questions with different procedures. Exact replications are divided into two sub-categories: **Dependent Replication**, with all conditions being the same or similar; and **Independent Replication**, in which one or more design elements vary.

Participant, in human-centered experiments, represents a person that is responsible for performing tasks and to generate data. Participants have **Profiles**, usually related with their experience with the experiment subject. Participants represent an interest population sample and are selected by **Sampling**, which might be (Wohlin et al., 2012): **Probabilistic**, in which the chance of each participant being selected is the same; or **Non-Probabilistic**, the opposite of probabilistic. Non-probabilistic techniques include sampling by convenience and by quota. Probabilistic sampling techniques include: simple random sampling, systematic sampling, and stratified random sampling. Baggili et al. (2013) recruited 107 students and faculty in a probabilistic manner as: "*After seeking ethical clearance, the researchers were able to disseminate a survey to 4473 e-mail addresses, which included students and faculty. The researchers gathered data for a period of two weeks, and the total number of participants in this study was (n = 107) after eliminating 188 participants with incomplete responses, 93 of which were females, and 14 of which were males*".

DF Process Phase is directly related with the DF phase (Arnes, 2017) to which an experiment is conducted. Therefore, DF experiments are conducted aiming at the following phases: **DF Acquisition**, **DF Examination**, **DF Analysis**, and **DF Reporting**. These concepts are essential to allow phases categorization, to plan replications, and to search for related/similar experiments. As examples, the following studies are performed for specific DF Phases: King and Vidas (2011) ran an experiment on analysis of solid state disk (SSD) data retention for the **DF Acquisition** phase; Breitinger et al. (2014) performed an experiment for the **DF Examination** phase on reducing the volume of digital data searching for relevant files in massive digital corpora; AlFahdi et al. (Al Fahdi et al., 2016), in their experiment, analyze suspect-oriented intelligent and automated computer forensic analysis of artifacts in the **DF Analysis** phase; and Husain et al. (2011) target all DF phases, including **DF Reporting** on an experiment on simple cost-effective framework for iPhone Forensics.

3.2. Pre-Operation concept

Setup is one of the most important concepts in DF Experiments, being well described in the majority of the analyzed papers. This concept represents the preparation of an infrastructure composed of **Software**, **Hardware** and **Algorithm**. These elements should be configured by an experimenter to establish a way to provide **Training** and to perform a **Pilot Project**. Fig. 5 depicts **Pre-Operation** and its related concepts. An example of setup description is presented by Fang et al. (2011): *"Two test hard disks are used including HD #1 with capacity of* 250GB *(total number of sectors N = 488,392,065) and speed of* 5400 rpm*, and HD #2 with capacity of* 60GB *(total number of sectors N = 117,210,240) and speed of* 4200 rpm*. The workstation running the experimental tests is configured with an Intel® CoreTM 2 CPU (E6750 at 2.*66GHz*) and 1.*97 GB RAM."

The **Software** concept is related to **Virtual Machine**, **Operating System**, and **Application**. In a virtual machine it necessary to install an operating system, which is responsible for executing applications for the virtual machine users. An example of virtual machine usage can be found in the work of Wang et al. (2016) for a computer forensic analysis model for reconstruction of the chain of evidence from volatile memory. Examples of applications are EnCase, FTK, Recuva, R-Studio, Stella Phoenix in the work of Buchanan-

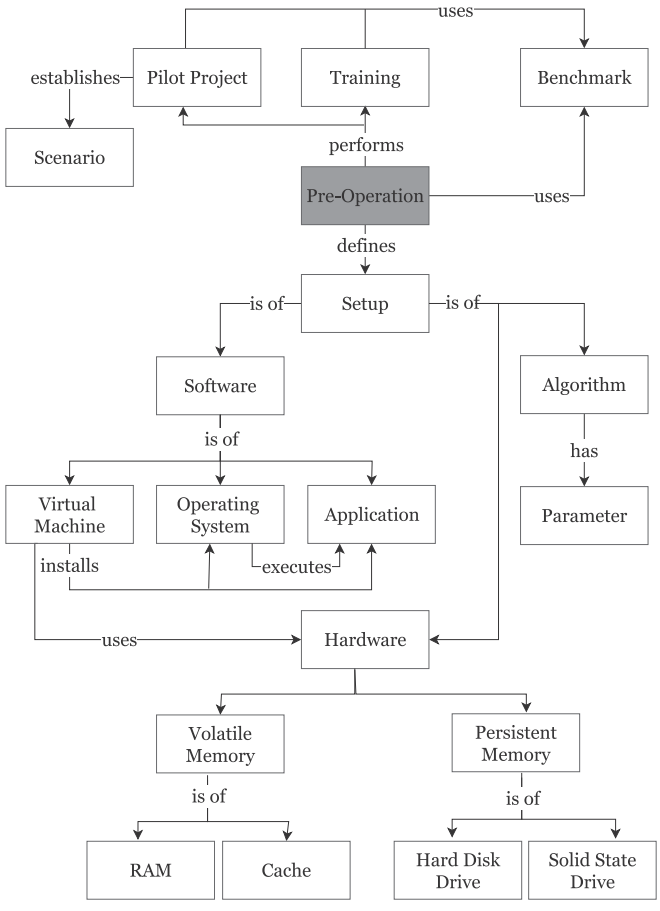

Fig. 5. Pre-Operation and related concepts.

Wollaston et al. (2013) for comparing data recovery function of forensic tools. Operating system definition is presented in the work of Cheng et al. (2017) in which a lightweight live memory forensic approach based on hardware virtualization is experimented.

Hardware setup is mainly related to **Volatile Memory** and **Persistent Memory**. Volatile Memory could be **RAM** or **Cache**, whereas Persistent Memory could be a **Hard Disk Drive** or **Solid State Drive**. Example of volatile memory used for hardware setup of an experiment is found in the work of Cheng et al. (2017) as: "*It is the first technique that is able to atomically acquire the entirety of the volatile memory, overcoming the SMM-imposed* 4GB *barrier while providing integrity guarantees and running on commodity systems*". An example of persistent memory can be found in the experiment of Hoelz et al. (2008) where a cooperative multi-agent approach to computer forensics is experimented.

The **Algorithm** concept is present in several DF experiments, and it is related with **Parameter**. Example of algorithm used is found in the study of Mohammed et al. (2018) where the Metadata Harmonisation algorithm is experimented for merging data sets through a "characterisation and harmonisation" process for Digital Forensics.

Training is usually performed aiming at refining or adjusting algorithms and their parameters (inputs/outputs), software and hardware setup. Training takes into consideration data from data sets or **Benchmark**, and instrument from the Planning concept. An example of training statement is from Ahmed et al. (2010): "*Fifty percent of the files were used for the training dataset to build representative models of file types...*". An example of benchmark is from the study of Zhou and Makris (2016), in which MiBench[2] is used. Training might also be performed to provide specialized knowledge for participants of an experiment. This kind of training is stated in the study of Shen et al. (2014) as: "*We asked subjects to complete an one-time demographic questionnaire before the collection of data began.*"

A **Pilot Project** is performed over the experiment setup infrastructure and it is used mainly for evaluating instrument previously defined. This kind of project is also known as mini experiment and it is essential for assuring a better quality of setup, training, and instrument of the experiment to be conducted. Pilot projects are usually executed under supervision of a person and it encompasses human interaction aiming to detect potential defects, bad design and/or threats to the conduction and analysis/interpretation of the experiment results. Therefore, experts on the experiment subject should be invited. Pilot projects are seen as good practices to refine and re-evaluate assumptions and hypotheses stated for an experiment. To do so, in the pilot project a scenario to be exercised by experts might be defined. Baggili et al. (2013), for example, mention that *"A pilot survey test was conducted prior to the distribution of the survey"*.

3.3. Operation concept

Operation means the execution of experimental tasks to produce data that can be analysed based on several elements from Planning and Pre-Operation. To do so, Operation follows **Operation Procedure**, collects **Data**, might use **Participant**, and performs over **Instrument**. Fig. 6 depicts **Operation** and related concepts.

Operation Procedure establishes a set of tasks necessary for the systematic execution of an experiment. Such tasks are essential, thus an experiment might produce reliable data and reduces experimental (type-I - false positive; and type-II - false negative) errors based on pre-evaluated elements during Pre-Operation. Therefore, a good description of the operation procedure contributes for reproducibility and auditability of experiments. Few studies provide such description, for example Schimid et al. (Schmid et al., 2015) mention "*a* 1 TB *external disk drive is virtually divided into regions of different sizes (*512 MB, 1 GB *and* 2 GB*) yielding approximately 2 K, 1 K and 512 regions, respectively. In separate experiments, one, two, three and four million sectors are randomly selected .... The same quantity of random sectors is individually extracted from 1 TB disk drive ... In similar lines, one million random samples from a 16 GB storage media were extracted for analysis.* "

The operation procedures might use a **Participant** to perform such tasks in human-centered experiments. Such participants produce data when performing the experimental tasks. Therefore, participants are an important source of data for the experiment. When dealing with experiments with no participants, algorithms, hardware and software take place to produce data. This encourages a well-defined setup during Pre-Operation.

Operation performs over **Instrument**, such as, models, database dumps, benchmarks, and metadata of data sets to produce data for the experiment. Whatever instrument is used, its description is essential for producing nonnoising and reliable data.

Data is the fundamental element of an experiment as it forms **Sample** mainly from independent variables to allow testing hypotheses, provide evidence, and draw conclusions. Data can be of **Original Data** or **Duplicated Data**. Original Data represent data available in any hardware/software which a Digital Forensics investigator collects from. For example, live forensics studies handle data available in random access memory (RAM), such as in (Thing et al., 2010; Maartmann-Moe et al., 2009). However, to preserve original data, several studies make Duplicated Data copies as they can get back to the original state. These kind of data are dealt in experiments such as in the work of Bakas et al. (2019).

3.4. . Analysis and interpretation concept

The **Analysis and Interpretation** concept is directly related with the performing of **Analysis Technique** and **Data Plotting**, the discussion of **Limitation** and the evaluation of **Threats to Validity** of an experiment. Fig. 7 depicts **Analysis and Interpretation** and related concepts.

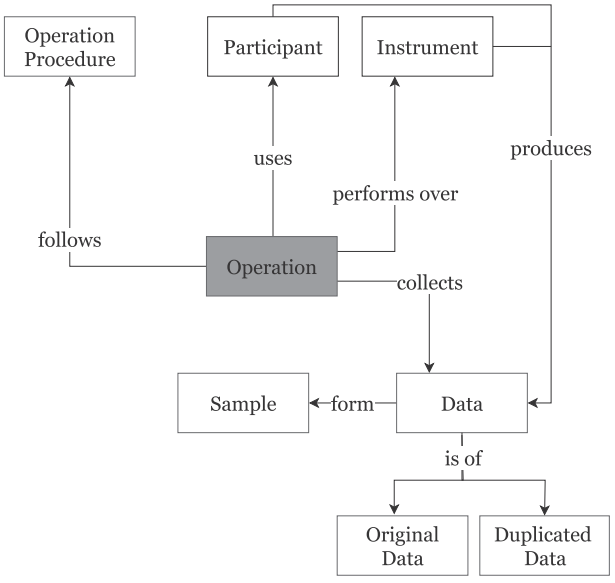

Fig. 6. The concept of operation.

[2] http://vhosts.eecs.umich.edu/mibench.

Once data are collected during the Operation of an experiment, **Data Plotting** is performed as a way to provide graphical visualization and organizations of experimental data. Data plotting might be realized by using **Table**, **Bar Chart**, **Tendency Line Chart**, **Boxplot**, **Histogram**, **Dispersion Chart**, **Decision Tree**, **Word Cloud**, and so on. Examples of studies that plot data with these techniques are (Schmid et al., 2015; Bakas et al., 2019; Maartmann-Moe et al., 2009; Shen et al., 2014; Zhou and Makris, 2016).

By plotting data, analysis can be performed in two ways: **Qualitative Analysis**, less common to DF experiments; and **Quantitative Analysis**, most usual to DF experiments. Qualitative analysis is aimed at analyzing mainly textual data, including techniques such as: Manual Verification, Focus Group, and Axial and Open Coding. Iqbal et al. (2008) use Manual Verification: "*In additional to measuring the quality of write-print using authorship identification accuracy, we also manually examined the extracted write-print and found that frequent patterns can succinctly capture combinations of features that occur frequently in the suspect's e-mails*." For quantitative analysis, it is possible to perform: **Descriptive Statistics**, **Normality Test**, **Hypothesis Test**, **Correlation**, and **Regression** upon collected data. **Descriptive Statistic** is a summary statistic that quantitatively describes or summarizes features of a collection, including: mean, median, standard deviation, and variation. The last three can be used to calculate the **Effect Size** of the test based on tested samples.

From this set of techniques, we understand that at least two of them must be conducted in every DF experiment towards reliable data analysis: **Normality Test** and Hypothesis **Test**. Normality tests provide analysis on the behavior of data samples based on a known distribution model. Experimenters usually take the normal curve for understanding sample behaviors. Normality tests that may be applied include: Kolmogorov–Smirnov (Lillefors), Quantilee-Quantile (QQ) Test, and Shapiro–Wilk. Baggili et al. (2013) perform Q–Q normality test: "*The data were analyzed using a variety of statistics. Primarily, the data were tested for normality and outliers using Q–Q plots and box plots*". Bharadwaj and Singh (2018) also performed QQ test: "*Initially, to assess the usefulness of the proposed sampling method towards the examination of the storage drive, the retrieved sector samples were evaluated using statistical Quantilee-Quantile (QQ) and null-hypothesis test*".

According to the normality of samples, one or more Hypothesis tests can be selected to test the experiment hypotheses. On the one hand, usually, a non-normal sample distribution leads to a non-parametric hypothesis test, such as: Chi Square, Mann–Whitney, and Squared Ranks. On the other hand, normal samples lead to a parametric hypotheses test, such as: TTest, ANOVA, and Paired T-Test. Example of a study that uses MannWhitney and T-Test is Bogen et al. (2010): "*Student's t-test for differences between means was applied when a data point (control/experimental pair) met the required criteria: normal distribution and uniform variance. When the data points did not meet the normal distribution and uniform variance criteria, a nonparametric Mann–Whitney test for differences between means was applied*".

Correlation might be performed between two pairs of samples to assess a possible linear association between two continuous variables. For example, Spearman Correlation Rank is used when one or both variables are skewed or ordinal, whereas Pearson Correlation Coefficient, used when both samples are normally distributed. A study that used the Pearson Correlation is Baggili et al. (2013): "*In order to test whether correlations existed between a set of measures, Pearson's correlation was used*".

Regression is a set of statistical processes for estimating the relationships among variables, i.e., the relationship between one dependent variable and one or more independent variables. Linear regression and nonlinear regression might be performed for analyzing experiment data. For example, Taha and Yoo (2018) used Logistic Regression for identifying suspects of a crime with no solid material evidences.

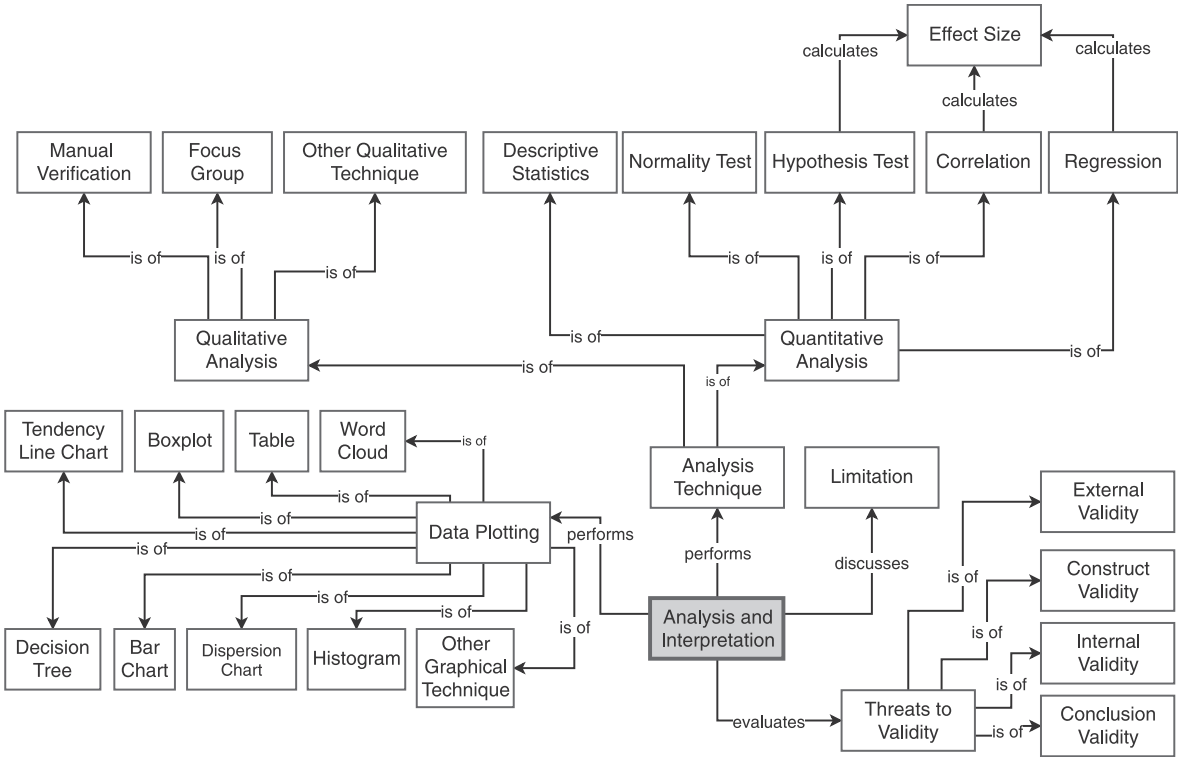

Fig. 7. The concept of analysis and interpretation.

Hypothesis, correlation and regression might calculate their **Effect Size**, which is a quantitative measure of the magnitude of a particular phenomenon. For example, Cohen's d is defined as the difference between two means divided by a standard deviation for the data.

Two concepts are of extreme importance to reduce biased experiments: **Limitation** and **Threats to Validity**. **Limitations** are discussed to provide the boundaries of the experimental study, thus a replication does not extrapolates its analysis. For example, the works of Al Sharif et al. (Al Sharif et al., 2014) and Kawaguchi et al. (2005), respectively: "*On the other hand, we must mention that none of the five tools were able to retrieve any data from the second image, which was created after a full format using windows format drive feature*" and "*In this paper, we do not consider the effect of malicious hosts, compromise of the Sign Server and network failure*". **Threats to Validity** evaluates the degree to which conclusions, constructions, internal and external relationships among variables based on the data are correct or reasonable (Wohlin et al., 2012). There are four types of validity: **Internal Validity**, **External Validity**, **Conclusion Validity**, and **Construct Validity** (Wohlin et al., 2012). Internal validity is focused on making sure a relationship is observed between the treatment and the outcome, *i.e.*, the treatment causes the outcome (effect). Threats to external validity are conditions that limit the ability to generalize the results of the experiment to industrial practice. Conclusion validity is focused on issues affecting the ability to draw the correct conclusion on relations between the treatment and the outcome of an experiment. Construct validity concerns to generalizing the result of the experiment to the concept or theory behind the experiment. Unfortunately, no analyzed study discusses Conclusion Validity.

In the work of Bogen et al. (2010), for instance, the authors discuss the following threats to validity:

- **Internal Validity:** "*The results of the experiments established a strong relationship between the case domain modeling method and an increase in effort...*";
- **Construct Validity:** "*One limitation of the study is that each evidence item is weighed equally and the quality of an examination is determined by how many evidence items are recovered...*"; and
- **External Validity:** "*The following factors should be considered before generalizing the results of these experiments: the duration of the examination activities, the size of the population, and the characteristics of the population. ...*"

3.5. Dissemination concept

The **Dissemination** concept aims at: sharing a **Data Set** of an experiment with proper **Authorship**, via a trusted **Repository** with an **Unique ID**, findable with a **Citation**; providing a **Data Management Plan** (DMP) on produced data; sharing **Diary/Annotation** on the experiment activities; and documenting **Experimental Issues**. Fig. 8 depicts **Dissemination** and related concepts.

Data Set represents all data produced and used in an experiment. Produced data might include: quantitative raw data from samples derived based on independent variables and the application of instruments with direct measurable measures, such as, Likert scale questions; and qualitative data obtained from subjective measures, such as, open questions in a questionnaire. Used data means data from an original experiment in case of a replication, data from a similar experiment, any benchmarks or even data from another research area.

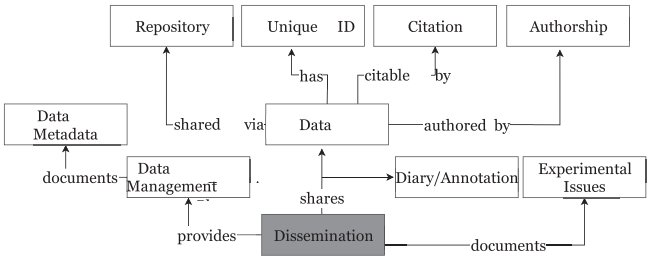

Fig. 8. Dissemination concept.

A data set must be identified with an **Unique ID**, preferable a name or a code (e.g. DOI) related to the developed research. Examples of data sets with unique ID are GovDocs[3] (Sabir et al., 2018), a publicly available data set of government documents with 986,278 files, and Enron[4] (Iqbal et al., 2008), an e-mail data set that contains 200,399 real-life e-mails from 158 employees of the Enron corporation.

Authorship is essential to identify who created such data set. Hence, authors can receive proper credits for its creation, as well as they might be consulted for specific discussions on such data set. For example, digital corpora[5] is responsible for the GovDocs data set.

Data sets usually are available in URLs of universities or government sites. However, over the years, such URLs break and data sets are no longer available. Therefore, a trusted **Repository** must be used (e.g. Zenodo, figshare.com, GitHub). Example of a study with available repositories using a DOI identifier is Uliyan et al. (2016) with a data set for image data manipulation (https://doi.org/10.1109/TIFS.2012.2218597).

All data sets should have a proper **Citation**. It should be possible to cite a data set by providing authorship, unique ID and public and permanent storage. To do so, there are several resources that facilitate the citation of a data set, such as, zenodo.org, figshare.com, and arXiv.org, or even GitHub. Lillis et al. (2018), for instance, share their HBFT algorithm using GitHub.

Diary/Annotation is a concept related to sharing all piece of data including decision making, recorded audio and/or video, pictures, draws, white-board annotations, or drafts of an experiment. These elements might contribute to understand the formulation of hypotheses, the selection of variables, the potential threats to validity of the study, the limitations, and further concerns for reproducibility.

Other important concept is **Experimental Issues**, which is related to problems and deviations during the experiment operation. For example, an inappropriate instrument used to capture data from participants. When the operation finishes, such participants should not be recruited again for the same experiment as they are, from now on, considered a threat to the study. Such problems must be shared towards avoiding facing them at a prospective experiment and as a matter of costs.

Data Forensics Management Plan (DFMP) is a Data Management Plan (DMP) for data forensics. This is related to how data are handled during and after a research project. To do so, a DFMP should consider aspects of data management, metadata generation, data preservation, and analysis for now and for preservation in the future (NSF, 2018; DCC, 2013; Wilkinson et al., 2016). DFMP usually documents the **Data Set Metadata**. The metadata describes the data set in a way one can understand data organization to either

[3] https://digitalcorpora.org/corpora/govdocs.

[4] http://www.cs.cmu.edu/~enron.

[5] https://digitalcorpora.org.

use data set in an experiment or to provide other researchers a way to reproduce conducted experiments. An example of metadata of the Enron data set is presented in (Iqbal et al., 2008).

4. ExperDF-CM feasibility evaluation survey

Empirical studies have shown that the conceptual modeling quality affects the quality of the information system (Maes and Poels, 2007). According to Moody (2005), the practice of evaluating the quality of conceptual models has more of the characteristics of an art than of an engineering discipline.

There are different techniques to evaluate conceptual models, from the user perspective to the application on actual industrial contexts. Therefore, we chose to perform an evaluation survey using the wide-knowing Technology Acceptance Model (TAM) (Davis, 1989) to perceive our conceptual model's usefulness and ease of use. We follow the guidelines of Linaker et al. (Linaker et al., 2015) for conducting surveys.

4.1. Survey objective

4.1.1. Research objective

This study aims to analyze the ExperDF-CM conceptual model's perception of usefulness and ease of use. We want to understand whether Digital Forensics researchers and practitioners find such model easy to use and useful for supporting them at performing and reporting experimental activities. To do so, we invited researchers and practitioners to take a survey by responding to a questionnaire. Information gathered up with this survey is used to provide initial evidence on the ExperDF-CM feasibility.

As far as we know, no other survey has been taken for evaluating a Digital Forensics experiments conceptual model.

All authors of this paper have discussed and agreed with the questions to be answered by respondents. The following research questions were then defined for this study:

- "RQ.1 - How ease to use ExperDF-CM is perceived from the perspective of Digital Forensics researchers and practitioners?"
- "RQ.2 - How is ExperDF-CM usefulness perceived from the perspective of Digital Forensics researchers and practitioners?"

Therefore, in this descriptive survey, we intend to measure the frequency of usefulness and ease to use factors among the investigated population.

4.1.2. Target population

Researchers and practitioners of the Digital Forensics area.

4.1.3. Sampling design

Our sampling strategy is non-probabilistic, more specifically by convenience.

4.1.4. Instrument design

We designed our main instrument (questionnaire) based on two factors of the TAM model (Davis, 1989): Perceived Usefulness (PU) and Perceived Ease of Use (EoU). We elaborated nine questions for the former, and eight questions for the latter. In addition, we defined three open questions.

4.1.5. Data analysis

We analyzed data from the survey using a Grounded Theory procedure, named Coding (Strauss and Corbin, 1990).

4.1.6. Variables

According to Davis (1989), "people tend to use or not use an application to the extent they believe it will help them perform their job better". Therefore, he is referring this as the "**perceived usefulness**" variable. In addition, even users find it useful they must believe the conceptual model is ease of use, thus, characterizing the "**perceived ease of use**".

Therefore, we can define "**perceived usefulness**" as "the degree to which a person believes that using a particular system would enhance his/her job performance". On the other hand, "**perceived ease of use**" might be defined as "the degree to which a person believes that using a particular system would be free of effort".

4.1.7. Instrumentation

We defined a set of instruments[6] to be used during the conduction of our survey:

- an Informed Consent Term (ICT) of participation, explaining that participation is voluntary and non-paid with no identification of participants;
- a Characterization Questionnaire, with few questions on the expertise, in academia or industry, on Digital Forensics, DF subareas and the experimentation process;
- a survey questionnaire with 20 questions: nine for the ExperDF-CM Perceived Usefulness, eight for Ease of Use, and three general open questions; and
- an instructional material explaining in details the ExperDF-CM conceptual model.

The survey questionnaire is composed of the following Likert-scaled questions:

- **Perceived Usefulness (PU):**
  - PU.1 - ExperDF-CM aids me to provide well-documented Digital Forensics experiments.
  - PU.2 - ExperDF-CM allows Digital Forensics experiments to be repeated, replicated and/or reproduced.
  - PU.3 - ExperDF-CM allows me to proper information retrieval regarding a Digital Forensics experiment.
  - PU.4 - ExperDF-CM contributes to a proper organization of the Digital Forensics experiment found evidence.
  - PU.5 - ExperDF-CM allows one to audit found evidence of a Digital Forensics experiment.
  - PU.6 - With ExperDF-CM I can apply meta-analysis to a set of Digital Forensics experiments towards generalization of evidence.
  - PU.7 - ExperDF-CM takes Digital Forensics experimentation process longer, but its benefits outweighs that.
  - PU.8 - ExperDF-CM improves the quality of the experiments I conduct.
  - PU.9 - Overall, I find ExperDF-CM useful in performing Digital Forensics experiments.
- **Perceived Ease of Use (EoU):**
  - EoU.1 - ExperDF-CM is objective. I often do not become confused when using it.
  - EoU.2 - ExperDF-CM is easy to use for fully document elements of a Digital Forensics experiment.
  - EoU.3 - ExperDF-CM can be used as a checklist in Digital Forensics experiments.
  - EoU.4 - ExperDF-CM can be used to provide error-prone identification in Digital Forensics experiments.
  - EoU.5 - Interacting with ExperDF-CM requires little of my mental effort.

[6] https://doi.org/10.5281/zenodo.3695857.

- EoU.6 - ExperDF-CM facilitates the support to perform and to document Digital Forensics experiments.
- EoU.7 - ExperDF-CM is easy to extend for particular cases in Digital Forensics experiments.
- EoU.8 - Overall, I find ExperDF-CM easy to use.

For each question we defined the following labels and numerical representation: “I strongly disagree” (1), “I disagree” (2), “Neutral” (3), “I do agree” (4), and “I strongly agree” (5).

Besides PU and EoU Likert scaled-questions, we also defined three open questions:

- OP.1 - What is your opinion/statement on the ExperDF-CM conceptual model for planning and conducting Digital Forensics experiments? Please, state 3–5 lines on this;
- OP.2 - Would you adopt ExperDF-CM for your prospective experiments in Digital Forensics? Please, provide details on your answer; and
- OP.3 - Would you recommend ExperDF-CM as a conceptual model to support performing Digital Forensics experiments to a colleague? Please, provide details on your answer.

The open questions were interpreted using open coding and axial coding techniques (Corbin and Strauss, 2008).

4.1.8. Selection of participants

We chose to invite as participants in this survey: (i) (co-)authors of DF experiments[7] published in the ACM, IEEE, Scopus, Springer, and Science Direct; and (ii) researchers from the National Institute of Science and Technology in Forensic Sciences (INCT Forense), a Brazilian wide project; and (iii) key distinguished researchers on DF.

4.1.9. Survey type

We decided to perform a self-administered, cross-sectional, and exploratory survey (Moll'eri et al., 2016). By being self-administered, respondents could answer in writing a set of questions; by being cross-sectional, we could gather a snapshot in time, as this survey could give us an idea on how things are for our respondents; finally, by being exploratory, we could focus on taking advantage of the respondent's experience to evaluate our conceptual model.

4.2. Execution

4.2.1. Questionnaire validation

We established the following activities to validate our questionnaire before applying it to participants:

1. **Analyze Survey Construction:** we invited five experts on survey questionnaires construction to evaluate our questions. As a result, we modified our questionnaire, discarding five questions from PU and six questions from EoU, thus resulting in the questions presented in Section 4.1.7;
2. **Simulate Taking the Survey:** we invited one expert on Digital Forensics to take the survey and providing notes on its questions to evaluate whether they capture the essence of DF experimentation. The invited evaluator is a Criminal Expert in the Brazilian Federal Police since 2003, working on seized forensic material reporting and software development for Digital Forensics;
3. **Run a Pilot Test:** we ran a pilot test with three respondents. Then, we cleaned data and analyzed towards errors and inconsistencies. None was found. We, thus, discarded such data from the final analysis.

Table 1
Survey Answers for Perceived Usefulness and ease of Use of ExperDF-CM.

Question	R.1	R.2	R.3	R.4	R.5	Mode
Perceived Usefulness						
PU.1	4	4	5	5	4	4
PU.2	4	5	4	5	5	5
PU.3	4	4	3	4	5	4
PU.4	4	4	4	5	5	4
PU.5	4	4	4	4	5	4
PU.6	4	5	3	4	5	4/5
PU.7	4	3	4	5	3	4/3
PU.8	4	4	3	5	4	4
PU.9	4	4	5	5	4	4
Perceived Ease of Use						
EoU.1	4	3	2	3	3	3
EoU.2	3	3	4	3	3	3
EoU.3	4	4	4	5	4	4
EoU.4	4	5	3	3	4	4/3
EoU.5	3	3	2	2	2	2
EoU.6	4	3	4	3	3	3
EoU.7	3	4	3	4	3	3
EoU.8	4	3	4	3	3	3

Legend: I strongly disagree = 1, I disagree = 2.
Neutral = 3, I do agree = 4, I strongly agree = 5.

4.2.2. Collecting data

We implemented the survey questionnaire using Google Forms.[8] From the invited people, five took our survey.

4.3. Analysis and interpretation

We performed the analysis in twofold: (i) tables with Likert-scale labels from each survey question; and (ii) coding for OP.1, OP.2, and OP.3 quotes from participants.

4.3.1. Perceived usefulness and ease of use analysis

We analyzed the respondents answers based on the Likert scale defined in Section 4.1.7. Table 1 presents a summary of respondents (R.x) and respective answers concerning the Perceived Usefulness (PU.y) and Ease of Use (EoU.z) of ExperDF-CM.

Perceived Usefulness. We applied Cronbach's Alpha (Cronbach, 1951) on the PU observations, which returned a value of $\alpha = 0.686$. This means that respondent answers on PU shared covariance and reliably measure the same underlying concept. Most of the answers for Perceived Usefulness of ExperDF-CM, as shown in Table 1, are concentrated in the “I do agree” label, followed by “I strongly agree”. This means, in general, that respondents found ExperDF-CM useful for supporting experimentation in Digital Forensics.

With regard to PU.1, all respondents agree ExperDF-CM aids to provide well-documented DF experiments. **PU.2** provides insights ExperDF-CM might allow DF experiments replication, repetition or reproduction. 80% of respondents agree ExperDF-CM provides a way to retrieve important information on DF experiments, whereas only 20% neither agree nor disagree based on **PU.3**. **In PU.4**, 100% of respondents agree ExperDF-CM contributes to properly organize findings of a DF experiment. The same percentage is applied to

[7] https://doi.org/10.5281/zenodo.3516001.

[8] https://bit.ly/3e0RJX8. This survey is still live in that site. We invite the reader to take that survey so we can improve ExperDF-CM.

PU.5, in which respondents agree with DF experiments might be audited. Meta-analysis might be supported using ExperDF-CM according to 80% of respondents in **PU.6**, whereas 20% neither agree nor disagree. **PU.7** provides evidence the experimentation process using ExperDF-CM might be longer than usual, however its benefits outweighs that based on 60% of the respondents. ExperDF-CM might improve the quality of experiments according to **PU.8**, in which 80% agree with it. In **PU.9**, all respondents found ExperDF-CM useful for performing DF experiments.

Perceived Ease of Use. Observing Table 1, we can see that most of respondents answers are neutral, excepting for questions EoU.3, EoU.4, and EoU.5. In **EoU.3** all respondents agreed ExperDF-CM can be used as a checklist for planning, conducting, and dissemination on evidence of DF experiments. As we expected, **EoU.4** also had most of respondents agreeing that ExperDF-CM can be used to provide error-prone identification in DF experiments. With regard to **EoU.5**, as we also expected, most respondents claim that using ExperDF-CM requires a significant mental effort. Most of the answers for the remaining questions (EoU.1, EoU.2, EoU.6, EoU.7, and EoU.8) are neutral.

4.3.2. Coding of respondents answer

By analyzing the quotes from all respondents, we came up with the following open codes: **OC.1** - Planning and Conducting DF Experiments; **OC.2** ExperDF-CM Adoption; and **OC.3** - ExperDF-CM Recommendation.

We, then, identified axial codes for each open code and respective relationships as follows: **AX.1** - Facilitate Developing DF Experiments/Practicality/Guidelines; **AX.2** - Validate/Check Forensics Data/Terminology; **AX.3** - Useful Model/Adoption; **AX.4** - Increase Students Knowledge; **AX.5** - Improves Quality and Efficiency of DF Experiments; **AX.6** - DF Phases Guide; and **AX.7** - Improvements and Constraints. Table 2 lists codes and relationships.

We can observe in Table 2 that the open codes have relationships to each other by analyzing the axial codes.

There is a strong relationship among OC.1, OC.2, and OC.3 with relation to **AX.1** ExperDF-CM facilitates the development of DF experiments, its practicality, and its use as a set of guidelines. For example, R.1 says "*it is a good model to facilitate on developing forensics experiments, validating and analyzing the forensics data*" and R.2 "*the model aids me as it is practical and relatively ease to use*".

With regard to **AX.2**, it has a relationship with OC.1 and OC.2 on validating and checking forensics data and its terminology. For instance, R.1 says "*it is a good model to... validating and analyzing the forensics data*" and R.4 claims that "*it is useful to describe experiments using a consistent and well-defined terminology*", "*it may also be used to check if any aspect of the experiment description was neglected*" and "*I used it to check my terminology and as a guideline on what to include in my experiment description*".

Ax.3 also has relationship with OC.1 and OC.2 on model usefulness and adoption. R.2 claims that "*the model is very useful to stick some lines focusing great objectives* ". R.3 claims that "*the ExperDF-CM conceptual model 'e a very useful tool for conducting experiments*" and R.4 mentions that "*it is useful to describe experiments using a consistent and well-defined terminology*". R.5 says that "*[the model] seems useful*". R.2 claims that "*certainly I would adopt ExperDF-CM for my prospective experiments in Digital Forensics*" and R.5 mentions that "*I would test [the model]*".

ExperDF-CM aids to increase the students (users) knowledge (**AX.4**) based on the opinion of R.2 "*...makes possible to increase the knowledge of students at certain aspects of different areas.*".

AX.5 has a relationship with OC.1 as R.3 states that "*with ExperDF-CM the quality and efficiency of experiments are better and provides greater practicality*".

AX.6 has a relationship with OC.3 as R.1 claims that "*as this conceptual model provides a complete set of components/stages for the experiments to be used as the based-knowledge and requirements on designing and conducting their Digital Forensics experiments*".

With regard to ExperDF-CM improvements and constraints by means of **AX.7**, such axial code has a direct relationship of improvement with OC.1 as R.1 asks for "*however, I would like to suggest on adding the information/process/procedure on the DF Phase - Collection.*" and "*[the model] looks laborious to use*". For OC.2, R.5 declares a constraint as "*depending on the effort needed I would keep using it or not [the model]*". With OC.3, R.2 states "*it is too early to recommend that.*" and R5. "*Only if the topic comes up I would present the model.*".

Table 2
Open and axial codes from the survey answers.

Codes	OC.1	OC.2	OC.3
AX.1	X	X	X
AX.2	X	X	
AX.3	X	X	
AX.4	X		
AX.5	X		
AX.6			X
AX.7	X	X	X

4.4. Discussion on results

Once we evaluated our conceptual model by analyzing the survey answers, we could draw conclusions on its feasibility based on the provided evidence.

By analyzing the evaluated perceived usefulness perspective of ExperDF-CM, we understand it is an useful tool to provide users guidance on planning, conducting and disseminating DF experiments evidence. Furthermore, based on the survey responses, ExperDF-CM can be useful for: documenting experiments; supporting replication, repetition, and reproduction of DF experiments; retrieving experiments information; organizing experiments; auditing experimental processes; supporting meta-analysis; and improving the quality of experiments.

In the perspective of the evaluated perceived ease of use of ExperDF-CM, we can draw the following feasibility-related conclusions: it is objective, avoiding confusing experimental terms and elements; it seems easy to use for documenting DF experiments; it can be used as a checklist for DF experiments conduction; it helps to find error-prone identification; it requires mental effort to be correctly used; and it seems easy to extend for a particular case of DF.

When analyzing the answers of the open questions (OP.1, OP.2 and OP.3) using coding, we could find several different open and axial codes. These codes aided us to identify the relationship among the open questions and how respondents perceived ExperDF-CM use for planning and conducting DF experiments, its possibility of adoption, and its recommendation to the DF community.

Overall, the quotes from respondents confirm their Likert-scale responses, thus providing feedback and suggestions on the ExperDF-CM feasibility. Responses went toward the following evidence on ExperDF-CM: (i) it facilitates developing DF experiments, it is practical, and it can be used as a guideline for experiments; (ii) it can be used to check and validate forensics data and experiments terminology; (iii) it is a useful model and can be easily adopted; (iv) it increases users knowledge; (v) it improves the quality and efficiency on DF experiments; (vi) it provides a guidance for DF phases during experimentation; and (vii) it has important ease of use constraints and it needs specific improvements to improve its ease of use.

4.5. Data sharing

All data used to perform this survey is available at https://doi.org/10.5281/zenodo.3695857.

4.6. Threats to validity

In this section we discuss the main threats related to the validity of this survey.

4.6.1. Face validity

We mitigate this threat by verifying that the open questions are totally related to the Likert-scale ones. Thus, for this survey, there is a strong relationship on what we measured and what we were supposed to measure, according to (Privitera, 2016).

4.6.2. Content validity

One potential threat to this survey is the fact that we did not ask respondents to answer specific DF experiments domain characteristics. As we asked general DF experiments questions, we could not reduce this threat. A further survey could explore specific characteristics to make ExperDF-CM encompassing specific use cases.

4.6.3. Internal Validity

We mitigate this threat by adopting Perceived Usefulness and Perceived Ease of Use classical measures. Besides, we asked open questions to determine the level of ExperDF-CM aids planning and conducting DF experiments, whether it could be adopted and further recommended. We, then, demonstrated a strong relationship between Likert-scale answers and open question answers.

4.6.4. External validity

To mitigate this threat, we provided respondents a document containing several examples for each ExperDF-CM element with published articles in DF experiments. Although we could not generalize results of this survey, it was possible to demonstrate evidence on the feasibility of the model.

4.6.5. Conclusion Validity

As we had a small sample of respondents, we tried to mitigate this threat by inviting experienced practitioners who work on actual DF projects and cases, thus making their responses more accurate than DF newcomers.

5. Conclusion

In this paper we presented ExperDF-CM, a conceptual model for promoting Digital Forensics experimentation that aims to assist researchers on planning, executing, analyzing and disseminating experiments evidence. This model is based on 154 DF experiment articles and is organized in five main experimental elements: “Planning”, “Pre-Operation”, “Operation”, “Analysis and Interpretation”, and “Dissemination”. Each of these elements is decomposed into several other experimental elements to support users on evaluating DF theories and technologies.

We illustrated ExperDF-CM with actual cases from a diversity of article excerpts, demonstrating how the elements of a DF experiment are spread out in a document, which makes it difficult to understand an experiment, thus compromising its repeatability, replicability or reproducibility.

We evaluated ExperDF-CM feasibility by means of a survey with DF experiment experienced respondents. They answered 20 questions about the usefulness and ease of use of ExperDF-CM, as well as its potential to improve planning, conduction and dissemination of experiments evidence, adoption and recommendation of the model. Our results show that ExperDF-CM is feasible for promoting DF experimentation.

As future work, we intend to: (i) rerun the survey with specific DF experiments situations, such as for mobile cloud forensics; (ii) develop an ontology to formalize the ExperDF-CM concepts and elements, thus providing a way for performing inferences on our dataset of DF experiments; and (iii) develop a recommendation system for DF experiments based on content-based filtering recommendation techniques for different types of users, from industry practitioners to lecturers and students.

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.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors would like to thank the Brazilian Institute for Forensics Science and Technology (Instituto Nacional de Ciência e Tecnologia Forense - INCT Forense) for providing Prof. Edson OliveiraJr and Dr. Charles V. Neu postdoctorate stipends. Avelino F. Zorzo is financed by CNPq/Brazil.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fsidi.2020.301014.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis, WIREs Comput. Stat 2, 433–459. https://doi.org/10.1002/wics.101.

Ahmed, I., suk Lhee, K., Shin, H., Hong, M., 2010. Content-based file-type identification using cosine similarity and a divide-and-conquer approach. IETE Tech. Rev. 27, 465–477. https://doi.org/10.4103/02564602.2010.10876780.

Al Fahdi, M., Clarke, N., Li, F., Furnell, S., 2016. A suspect-oriented intelligent and automated computer forensic analysis. Digit. Invest. 18, 65–76. https://doi.org/10.1016/j.diin.2016.08.001 https://doi.org/10.1016/j.diin.2016.08.001.

Al Sharif, S., Al Ali, M., Salem, N., Iqbal, F., El Barachi, M., Alfandi, O., 2014. An approach for the validation of file recovery functions in digital forensics' software tools. In: 6th International Conference on New Technologies, Mobility and Security (NTMS), 2014, pp. 1–6. https://doi.org/10.1109/NTMS.2014.6814005.

Andujar, C., Schiaffonati, V., Schreiber, F.A., Tanca, L., Tedre, M., van Hee, K., van Leeuwen, J., 2012. The role and relevance of experimentation in informatics. In: Technical Report 11/2012, Informatics Europe. http://www.informatics-europe.org/images/documents/informatics-experimentation_2013.pdf.

Arnes, A., 2017. Digital Forensics, first ed. Wiley.

Arshad, H., Jantan, A.B., Abiodun, O.I., 2018. Digital forensics: review of issues in scientific validation of digital evidence. J. Inf. Process Syst. 14 https://doi.org/10.3745/JIPS.03.0095.

Baggili, I., Al Shamlan, M., Al Jabri, B., Al Zaabi, A., 2013. Cybercrime, censorship, perception and bypassing controls: an exploratory study. In: Rogers, M., Seigfried-Spellar, K.C. (Eds.), Digital Forensics and Cyber Crime. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 91–108.

Bakas, J., Naskar, R., Dixit, R., 2019. Detection and localization of interframe video forgeries based on inconsistency in correlation distribution between haralick coded frames. Multimed. Tool. Appl. 78, 4905–4935. https://doi.org/10.1007/s11042-018-6570-8 https://doi.org/10.1007/s11042-018-6570-8.

Bharadwaj, N.K., Singh, U., 2018. Efficiently searching target data traces in storage devices with region based random sector sampling approach. Digit. Invest. 24, 128–141. https://doi.org/10.1016/j.diin.2018.02.004 http://www.sciencedirect.com/science/article/pii/S1742287617303249.

Bogen, A.C., Dampier, D.A., Vaughn, R., Reese, D.S., Allen, E.B., Carver, J.C., 2010. Structured forensics examination planning with domain modeling: a report of three experiment trials. J. Digit. Forensic Pract. 3, 23–32. https://doi.org/10.1080/15567280903376896 https://doi.org/10.1080/15567280903376896.

Breitinger, F., Winter, C., Yannikos, Y., Fink, T., Seefried, M., 2014. Using approximate matching to reduce the volume of digital data. In: Peterson, G., Shenoi, S. (Eds.), Advances in Digital Forensics X. Springer Berlin Heidelberg, Berlin, Heidelberg,

pp. 149–163.
Buchanan-Wollaston, J., Storer, T., Glisson, W., 2013. Comparison of the data recovery function of forensic tools. In: Peterson, G., Shenoi, S. (Eds.), Advances in Digital Forensics IX. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 331–347.
Casey, E., 2013. Experimental design challenges in digital forensics. Digit. Invest. 9, 167–169. https://doi.org/10.1016/j.diin. 2013.02.002.
Caviglione, L., Wendzel, S., Mazurczyk, W., 2017. The future of digital forensics: challenges and the road ahead. IEEE Secur. Priv. 15, 12–17. https://doi.org/10.1109/MSP.2017.4251117.
Chen, P.P.-S., 1976. The entity-relationship model: toward a unified view of data. ACM Trans. Database Syst. 1, 9–36. https://doi.org/10.1145/320434.320440 http://doi.org/10.1145/320434.320440.
Cheng, Y., Fu, X., Du, X., Luo, B., Guizani, M., 2017. A lightweight live memory forensic approach based on hardware virtualization. Inf. Sci. 379, 23–41. https://doi.org/10.1016/j.ins.2016.07.019 http://www.sciencedirect.com/science/article/pii/S0020025516305011.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. https://doi.org/10.1177/001316446002000104.
Corbin, J., Strauss, A., 2008. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, 3 ed. SAGE Publications, Inc., Thousand Oaks, CA https://doi.org/10.4135/9781452230153.
Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555.
Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 13, 319. https://doi.org/10.2307/249008.
DCC, 2013. Digital curation center - data management plans. http://www.dcc.ac.uk/resources/data-management-plans.
Dieste, O., Juristo, N., Martinc, M.D., 2013. Software industry experiments: a systematic literature review. In: 1st International Workshop on Conducting Empirical Studies in Industry. CESI), pp. 2–8. https://doi.org/10.1109/CESI.2013.6618462.
Fang, J., Jiang, Z., Yiu, S., Hui, L., 2011. An efficient scheme for hard disk integrity check in digital forensics by hashing with combinatorial group testing. In: International Journal of Digital Content Technology and its Applications. https://doi.org/10.4156/jdcta.vol5.issue2.35.
Forsyth, D., 2018. Probability and Statistics for Computer Science. https://doi.org/10.1007/978-3-319-64410-3.
Garfinkel, S.L., 2010. Digital forensics research: the next 10 years. Digit. Invest. 7, S64–S73. https://doi.org/10.1016/j.diin.2010.05.009 http://doi.org/10.1016/j.diin. 2010.05.009.
Hoelz, B.W.P., Ralha, C.G., Geeverghese, R., Junior, H.C., 2008. A cooperative multi-agent approach to computer forensics. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 2, pp. 477–483. https://doi.org/10.1109/WIIAT.2008.55, 2008.
Hoffman, J.I., 2019. Chapter 36 - meta-analysis. In: Hoffman, J.I. (Ed.), Basic Biostatistics for Medical and Biomedical Practitioners, second ed. Academic Press, pp. 621–629. https://doi.org/10.1016/B978-0-12-817084-7.00036-X.
Hong, I., Yu, H., Lee, S., Lee, K., 2013. A new triage model conforming to the needs of selective search and seizure of electronic evidence. Digit. Invest. 10, 175–192. https://doi.org/10.1016/j.diin.2013.01.003.
Horsman, G., 2018. 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.
Husain, M.I., Baggili, I., Sridhar, R., 2011. A simple cost-effective framework for iphone forensic analysis. In: Baggili, I. (Ed.), Digital Forensics and Cyber Crime. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 27–37.
Iqbal, F., Hadjidj, R., Fung, B.C., Debbabi, M., 2008. A novel approach of mining write-prints for authorship attribution in e-mail forensics. Digit. Invest. 5, S42–S51. https://doi.org/10.1016/j.diin.2008.05.001 (the Proceedings of the Eighth Annual DFRWS Conference).
Iqbal, F., Binsalleeh, H., Fung, B.C.M., Debbabi, M., 2013. A unified data mining solution for authorship analysis in anonymous textual communications. Inf. Sci. 231, 98–112. https://doi.org/10.1016/j.ins.2011.03.006.
Juristo, N., Moreno, A.M., 2010. Basics of Software Engineering Experimentation, first ed. Springer Publishing Company, Incorporated.
Kawaguchi, N., Obata, N., Ueda, S., Azuma, Y., Shigeno, H., Okada, K., 2005. Efficient log authentication for forensic computing. In: Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop, pp. 215–223. https://doi.org/10.1109/IAW.2005.1495955.
Kim, K., Lee, S.S., Hong, D., Ryou, J., 2011. Gpu-accelerated password cracking of PDF files. Trans. Internet Inf. Syst. 5, 2235–2253.
King, C., Vidas, T., 2011. Empirical analysis of solid state disk data retention when used with contemporary operating systems. Digit. Invest. 8, S111–S117. https://doi.org/10.1016/j.diin.2011.05.013 http://doi.org/10.1016/j.diin.2011. 05.013.
Lillis, D., Breitinger, F., Scanlon, M., 2018. Expediting mrsh-v2 approximate matching with hierarchical bloom filter trees. In: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, vol. 216, pp. 144–157. https://doi.org/10.1007/978-3-319-73697-6_11.
Linaker, J., Sulamn, S., Host, M., de Mello, R., 2015. Guidelines for Conducting Surveys in Software Engineering. Lund University, Sweden. Technical Report 1.1.
Louis, A., Engelbrecht, A., 2011. Unsupervised discovery of relations for analysis of textual data. Digit. Invest. 7, 154–171. https://doi.org/10.1016/j.diin.2010.08.004.
Maartmann-Moe, C., Thorkildsen, S.E., Arnes, A., 2009. The persistence of memory: forensic identification and extraction of cryptographic keys. http://www.sciencedirect.com/science/article/pii/S1742287609000486 Digit. Invest. 6, S132–S140. https://doi.org/10.1016/j.diin.2009.06.002 (the Proceedings of the Ninth Annual DFRWS Conference).
Maes, A., Poels, G., 2007. Evaluating quality of conceptual modelling scripts based on user perceptions. Data Knowl. Eng. 63, 701–724. https://doi.org/10.1016/j.datak.2007.04.008 http://www.sciencedirect.com/science/article/pii/S0169023X07000754. 25th International Conference on Conceptual Modeling (ER 2006).
Mitchell, J., Welty, C., 1988. Experimentation in computer science: an empirical view. Int. J. Man Mach. Stud. 29, 613–624. https://doi.org/10.1016/S0020-7373(88).80069-5 http://www.sciencedirect.com/science/article/pii/S0020737388800695.
Mohammed, H., Clark, N., Li, F., 2018. Automating the harmonisation of heterogeneous data in digital forensics. In: Jøsang, A. (Ed.), Proceedings of the 17th European Conference on Information Warfare and Security. Academic Conferences and Publishing International Limited, pp. 299–306.
Molleri, J.S., Petersen, K., Mendes, E., 2016. Survey guidelines in software engineering: an annotated review, 58:1–58:6. In: Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '16. ACM, New York, NY, USA. https://doi.org/10.1145/2961111.2962619. http://doi.org/10.1145/2961111.2962619.
Montgomery, D., 2008. Design and Analysis of Experiments. John Wiley & Sons.
Moody, D.L., 2005. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. Data Knowl. Eng. 55, 243–276. https://doi.org/10.1016/j.datak.2004.12.005 http://www.sciencedirect.com/science/article/pii/S0169023X04002307. quality in conceptual modeling.
Nance, K., Hay, B., Bishop, M., 2009. Digital forensics: defining a research agenda. In: 2009 42nd Hawaii International Conference on System Sciences, pp. 1–6. https://doi.org/10.1109/HICSS.2009.160.
Novak, J.D., Cañas, A.J., 2006. The Theory Underlying Concept Maps and How to Construct and Use Them, Research Report 2006-01 Rev 2008-01. Florida Institute for Human and Machine Cognition. http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm.
NSF, 2018. National science foundation - data management guidance for computer & information sciences & engineering proposals and awards. https://www.nsf.gov/cise/cise_dmp.jsp.
Palomo, E., North, J., Elizondo, D., Luque, R., Watson, T., 2012. Application of growing hierarchical som for visualisation of network forensics traffic data. Neural Network. 32, 275–284. https://doi.org/10.1016/j.neunet.2012.02.021 http://www.sciencedirect.com/science/article/pii/S0893608012000500. selected Papers from IJCNN 2011.
Perkel, J.M., 2018. A toolkit for data transparency. Nature 560, 513–515. https://doi.org/10.1038/d41586-018-05990-5.
Peterson, J.L., 1981. Petri Net Theory and the Modeling of Systems. Prentice Hall PTR, Upper Saddle River, NJ, USA.
Privitera, G.J., 2016. Research Methods for the Behavioral Sciences, 2 ed. SAGE Publications, Inc., Thousand Oaks, CA.
Raghavan, S., 2013. Digital forensic research: current state of the art. CSI Trans. ICT 1, 91–114. https://doi.org/10.1007/s40012-012-0008-7.
Sabir, M.F., Jones, J.H., Liu, H., Mbaziira, A.V., 2018. A non-algorithmic forensic approach for hiding data in image files. In: Proceedings of the 2Nd International Conference on Compute and Data Analysis, ICCDA 2018. ACM, New York, NY, USA, pp. 60–64. https://doi.org/10.1145/3193077.3193087 http://doi.org/10.1145/3193077.3193087.
Schmid, M.R., Iqbal, F., Fung, B.C., 2015. E-mail authorship attribution using customized associative classification. http://www.sciencedirect.com/science/article/pii/S1742287615000572 Digit. Invest. 14, S116–S126. https://doi.org/10.1016/j.diin.2015.05.012 (the Proceedings of the Fifteenth Annual DFRWS Conference).
Shen, C., Cai, Z., Maxion, R.A., Guan, X., 2014. On user interaction behavior as evidence for computer forensic analysis. In: Shi, Y.Q., Kim, H.-J., Perez-Gonzalez, F. (Eds.), Digital-Forensics and Watermarking. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 221–231.
Shull, F.J., Carver, J.C., Vegas, S., Juristo, N., 2008. The role of replications in empirical software engineering. Empir. Software Eng. 13, 211–218. https://doi.org/10.1007/s10664-008-9060-1 https://doi.org/10.1007/s10664-008-9060-1.
Strauss, A., Corbin, J., 1990. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage Publications, Newbury Park, California.
Taha, K., Yoo, P.D., 2018. A forensic system for identifying the suspects of a crime with no solid material evidences. In: IEEE Intl Conf on Dependable, Autonomic and Secure Computing, pp. 576–583. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00107.
Tedre, M., Moisseinen, N., 2014. Experiments in computing: a survey. Sci. World J. 2014, 1–12. https://doi.org/10.1155/2014/549398.
Thing, V.L., Ng, K.-Y., Chang, E.-C., 2010. Live memory forensics of mobile phones. http://www.sciencedirect.com/science/article/pii/S174228761000037X Digit. Invest. 7, S74–S82. https://doi.org/10.1016/j.diin.2010.05.010 (the Proceedings of the Tenth Annual DFRWS Conference).
Tichy, W.F., 1998. Should computer scientists experiment more? Computer 31, 32–40. https://doi.org/10.1109/2.675631.
Tichy, W.F., Lukowicz, P., Prechelt, L., Heinz, E.A., 1995. Experimental evaluation in computer science: a quantitative study. J. Syst. Software 28, 9–18. https://doi.org/10.1016/0164-1212(94)00111-Y http://www.sciencedirect.com/science/article/pii/016412129400111Y.

Uliyan, D.M., Jalab, H.A., Wahab, A.W.A., Shivakumara, P., Sadeghi, S., 2016. A novel forged blurred region detection system for image forensic applications. Expert Syst. Appl. 64, 1–10.

Uthmann, C.v., Becker, J., 1999. Guidelines of modelling (gom) for business process simulation. In: Scholz-Reiter, B., Stahlmann, H.-D., Nethe, A. (Eds.), Process Modelling. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 100–116.

Wang, F., Hu, L., Hu, J., Zhao, K., 2016. Computer forensic analysis model for the reconstruction of chain of evidence of volatile memory data. Multimed. Tool. Appl. 75, 10097–10107. https://doi.org/10.1007/s11042-015-2798-8 https://doi.org/10.1007/s11042-015-2798-8.

Wen, K., Zeng, Y., Li, R., Lin, J., 2012. Modeling semantic information in engineering applications: a review. Artif. Intell. Rev. 37, 97–117. https://doi.org/10.1007/s10462-011-9221-2 https://doi.org/10.1007/s10462-011-9221-2.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3. https://doi.org/10.1038/sdata.2016.18.

Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., Wesslen, A., 2012. Experimentation in Software Engineering: an Introduction, 2 ed. Kluwer Academic Publishers, Norwell, MA, USA.

Zhou, L., Makris, Y., 2016. Hardware-based workload forensics: process reconstruction via tlb monitoring. In: IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2016, pp. 167–172. https://doi.org/10.1109/HST.2016.7495577.