Framework for Reliable Experimental Design (FRED): A research framework to ensure the dependable interpretation of digital data for digital forensics.

Graeme Horsman
Faculty of Computer Science

The David Goldman
Informatics Centre
St Peter's Way
Sunderland
SR6 0DD

Email: graeme.horsman@sunderland.ac.uk
Phone: 0191 515 2381

Abstract
The establishment of fact forms the cornerstone of any forensic discipline, with digital analysis being no exception. Practitioners are under an obligation as expert witnesses to provide factual accounts of digital scenarios, which must be underpinned by robust knowledge and evidential findings. To achieve this level of reliability, investigatory research must be suitably planned, implemented and analysed in a way which instills confidence in the accuracy of any findings. This is particularly important as digital forensic organisations are now facing the impending requirement to have acquired ISO/IEC 17025 accreditation. This article proposes the Framework for Reliable Experimental Design (FRED) to support those engaged in the field of digital forensics research to contribute reliable, 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 resource for those operating within the digital forensic field, both in industry and academia, to support and develop research best practice within the discipline.

Keywords:- Digital forensics; expert evidence; admissibility; research; digital evidence.

1 Introduction

As a field, digital forensics (DF) remains in a constant state of change, driven by technological changes. In an effort to prevent the construction of another definition of DF, focus is drawn to Raghavan's (2013) following characterisation. 'Digital forensics is a branch of science that involves the application of scientific principles to the investigation of artifacts present in one or more digital devices in order to understand and reconstruct the sequence of events that must have transpired in generating the said artifacts' (Raghavan, 2013, p.91). The key components to highlight within this definition lie with the need to 'understand' and 'reconstruct' digital data, a requirement for practitioners in all investigation scenarios. To carry out these tasks requires the use of effective research strategies, underpinned by the implementation of a robust research methodologies in order to aid the accurate interpretation and understanding of digital data. Yet currently, there is limited guidance available to practitioners and academics supporting the construction of valid DF research at this level.

Atkinson (2014, p.248) suggests that a DF practitioner 'is unlikely to, and possibly incapable of, fully understanding the ever-changing nuances of the contiguous technological layers their expertise rests upon'. The vast divergence in technological devices poses a challenge, and to ensure DFs continued effectiveness, reliance is placed upon research into a number of problem areas, generated by both professional organisations, industry practitioners, and academia, and this was acknowledged in 2010 by Garfinkel. Now, in 2017 this stance remains unchanged, and in an environment which remains in a constant state of development, practitioners are now more reliant than ever on research carried out by themselves and others. Lillis et al., (2016) provide an overview of future areas of research for the DF field with numerous other proposals of a DF research agenda having been offered in previous years (see Beebe, 2009; Nance et al., 2009; Karie and Venter, 2015; Jerman-Blažič and Klobučar, 2016). Defining areas of future interest helps to provide direction for DF and supports the effective allocation of available resources. Yet it has been suggested that a disconnect exists between industry and researchers (Sremack, 2007), where standards of accuracy may diverge and the possibility of misinterpretations of digital data may exist.

Any contribution to DF knowledge from research must be valid. Christensen, et al., (2014) state that 'validity can best be thought of as the overall probability of reaching the correct conclusion, given a specific method and data. Methods that are considered "valid" give us the correct conclusion more often than chance'. The need for robust research lies with the level of reliance placed upon it. Practitioners operating in industry may depend on information and guidance published in order to analyse content found during an investigation. The need to entrust in the work of others may be due to a number of factors including a lack of time or resources available to independently carry out the level of required research themselves. To provide context, the depth of experimentation which can take place in academic environments, where the dissemination of research is often a key component of that profession (see for example, within the United Kingdom, the Research Excellence Framework (http://www.ref.ac.uk/) and its impact on universities) may be greater than time allows in industry (Sremack, 2007). The creation, publication and sharing of domain specific research is a positive act, helping to combat criticism of the field of DF having frequently operated with a 'silo mentality' (Rogers and Seigfried, 2004; Biros et al., 2008; Spyridopoulos et al., 2013). Yet, caution should be taken with regards to the validity and accuracy of research contributions being placed in circulation, particularly where research documenting the interpretive analysis of artefacts is offered with accompanying data parsing facilities. Should the accuracy of any interpretation be flawed, its application in the real world can lead to severe consequences. Further, should erroneous DF research enter mainstream viewing, it can lead to its widespread use, potentially jeopardizing multiple cases. As has been highlighted in multiple publications, there is a lack of standardisation in DF (Beebe, 2009; Karyda and Mitrou, 2007; Bulbul et al., 2013) where testing and research procedures vary, and arguably, there is also a lack of consistency and clarity in the procedures and methodologies adopted and utilised during DF research.

There is a need for standardisation and transparency in DF research methodologies to allow sufficient peer-review of practices, secondary interpretation of data and the ability to assess the reliability of results that are offered in any contribution to knowledge. This article offers the Framework for Reliable Experimental Design (FRED) in an attempt to formalise a base from which to design and carry out robust DF research providing for the accurate interpretation of digital data. FRED is not concerned with the high-level investigation

process, but is targeted for application at the research level and at those carrying out analysis of applications and digital artefacts to establish their functionality. Existing DF frameworks frequently focus on the targeted analysis of specific data types, artefacts or systems (discussed further in Section 3). However, FRED provides low level investigative support at the point of the planning and development of robust experimental design and implementation, regardless of the data under investigation. Existing DF frameworks often overlook the process of planning, constructing and undertaking a valid experiment which may jeopardise the reliability of results. The FRED framework can be applied to almost all DF investigative research and documents the stages involved in order to design and carry out effective testing to achieve valid results, supporting the DF practitioner during their investigations and research. It is designed to be a resource for industry and academia to help standardise and openly document the stages of any testing and research undertaken, to allow effective scrutiny, peer-review and evaluation of any outputs offered. FRED is a six step framework, supporting the planning, implementation, and analysis of digital data in order to establish a factually accurate outcome

2 The need for admissible evidence

It is imperative that the results of any DF investigation are reliable, based on procedures and interpretation derived from sound research. Where an investigation surrounds illegal acts, in England and Wales, the standard of proof to be obtained is 'beyond reasonable doubt', with a majority of at least 10 (of 12) jury members agreeing on a verdict needed. Under direction of the trial judge, DF expert witnesses are under a responsibility to present juries with the information needed to make an informed decision. This is often not a straightforward process, with Naughton (2007, p18) stating 'criminal trials are not a consideration of factual innocence or factual guilt in any straightforward sense. They are highly technical affairs which attempt to determine if defendants are 'guilty' or 'not guilty' of criminal offences on the basis of the reliability of the evidence before the court'. The criminal justice system is not perfect and wrongful convictions have been documented, with Naughton and Tan (2010) drawing attention to the following comments raised in *Director of Public Prosecutions v. Shannon [1974] 59 Cr.App.R.250*.

> 'The law in action is not concerned with absolute truth, but with proof before a fallible human tribunal to a requisite standard of probability in accordance with formal rules of evidence'.

The presentation of expert evidence is a powerful tool, one which has the ability to sway jury decision making (see for example the cases of Sally Clark and Angela Cannings, where expert evidence testimony regarding cot death were considered misleading (Naughton and Tan, 2010)). In addition, the reliance and trust by jurors in expert evidence can cloud judgement and decision making (Lovett and Kovera, 2008). In England and Wales, experts are under a duty following The Criminal Procedure Rules 2015 Section 19 to 'help the court to achieve the overriding objective by giving opinion which is objective and unbiased'. Evidence is admissible if it is of assistance to the court, the expert has relevant experience and is impartial, and, that the evidence is reliable (Crown Prosecution Service, 2014). Expanding on the need for reliability, the following guidance is offered.

> 'There should be a sufficiently reliable scientific basis for the expert evidence or it must be part of a body of knowledge or experience which is sufficiently

> organised or recognised to be accepted as a reliable body of knowledge or experience...The reliability of the opinion evidence will also take into account the methods used in reaching that opinion, such as validated laboratory techniques and technologies, and whether those processes are recognised as providing a sufficient scientific basis upon which the expert's conclusions can be reached. The expert must provide the court with the necessary scientific criteria against which to judge their conclusions' (Crown Prosecution Service, 2014).

Courts may take into account a number of factors for determining the reliability of expert evidence in England and Wales, with guidance provides in the *Criminal Practice Directions [2015] EWCA Crim 1567* Section 19A with focus drawn to points a,b,c,d and h.

> '(a) the extent and quality of the data on which the expert's opinion is based, and the validity of the methods by which they were obtained;
>
> (b) if the expert's opinion relies on an inference from any findings, whether the opinion properly explains how safe or unsafe the inference is (whether by reference to statistical significance or in other appropriate terms);
>
> (c) if the expert's opinion relies on the results of the use of any method (for instance, a test, measurement or survey), whether the opinion takes proper account of matters, such as the degree of precision or margin of uncertainty, affecting the accuracy or reliability of those results;
>
> (d) the extent to which any material upon which the expert's opinion is based has been reviewed by others with relevant expertise (for instance, in peer-reviewed publications), and the views of those others on that material;
>
> (h) whether the expert's methods followed established practice in the field and, if they did not, whether the reason for the divergence has been properly explained'.

In direct application to DF, the Criminal Practice Directions raise a number of concerns. A lack of defined research methodologies in DF leads to varying standards of practice, making an assessment of the validity of any utilised method difficult. A poorly constructed methodology prevents the accurate assessment of the precision or margin of uncertainty in results obtained, both by a third party, and worryingly, by an investigator. In addition, a methodology which is poorly documented cannot be established as best practice within the field or effectively peer reviewed, meaning that the validity of its outputs cannot be determined.

Ensuring the reliability of presented evidence forms the crux of an expert's role, and the need to present a factual account is crucial with Sallavaci and George (2013) stating that now, experts are facing increasing requirements to prove the validity of their work. Olivier (2016, p.49) states the importance of this, noting 'science is a quest for truth. The law, when considering disputes, often need to determine facts. Facts are claims that are true'. Mistakes during the interpretation of digital data during a forensic investigation can have significant

impacts on all parties involved in the process. Misinterpretation may lead to the prosecution of the innocent or exoneration of the guilty. The consequences wrongful conviction for the individual are severe, impacting upon all aspects of life. Scott (2010) highlights the psychological effects which can be incurred by those wrongly convicted, impacting upon health, relationships and the ability to engage in a 'normal' lifestyle. In addition, compensation for wrongly convicted individuals may be sought (see for example, Gov.uk (n.d.) in conjunction with sections 133, 133a and 133b of the Criminal Justice Act 1988 for procedural details for claims). Incorrectly interpreted evidence can also impact on the expert and their continued credibility for operating in the field, as well as the organisation they represent, drawing scrutiny upon areas such as sufficient knowledge, training and suitability of practices.

With the importance of establishing reliable DF evidence evident, the lack of documented DF research methodologies to support the development of valid research remains a concern. This omission may lead to divergences in practices and no consistent approach to establishing the reliability of work carried out. The issue of establishing reliability is not just a problem at a research-level, but also a legal challenge, with Ireland and Beaumont (2015, p.6) suggesting that in the United Kingdom, 'judges continue to be provided with no real guidance on how they should determine evidential reliability'.

There are a series of frameworks which focus on the high-level core elements of an investigation, notably acquire, examine, analyse, and report (Baryamureeba and Tushabe, 2004, Carrier and Spafford, 2004; Reith et al., 2002; von Solms et al., 2006; Kohn et al., 2006, Selamat et al., 2008). In addition, there are numerous frameworks which focus on elements of a DF investigation such as network events, forensic readiness, social networks; triage and cloud investigations (Pilli et al., 2010; Elyas, et al., 2014; Jang and Kwak, 2015; Martini,and Choo, 2012; Hitchcock, et al., 2016). Yet there is an oversight in the development of frameworks for developing robust evidential data derived from planned testing processes, the reverse engineering of evidential artifacts and the interpretation of any discovered data.

To provide an example, currently frameworks may suggests a need to investigate artefact of type 'A' during an investigation of a certain type. Such frameworks may go on to highlight the particular functions of artefact 'A'. Yet there are a lack of guidance defining the adequate procedures needed to fully test, examine, and interpret the content and function of artefact 'A' to ensure any retrieved results are accurate. In essence, this article argues the need for guidance to support practitioners and academics in DF at a 'research and testing level' construct and implement forensically sound testing methodologies to ensure the potential admissibility of their work.

2.1 Governance and validation:- ISO/IEC 17025

The requirements for valid testing and establishment of robust evidence in forensic sciences including DF have been formalised by the International Organisation for Standardization (n.d.a) leading to the development of a number of standards designed to improve organisational performance, enhance credibility and in some instances, meet legal requirements (International Organisation for Standardization, n.d.b; n.d.c). Whilst many have traditionally been pursued by DF organisations (for example, ISO 9001 Quality management; ISO/IEC 27001 Information security management), the Forensic Science

Regulator (2016) has defined a deadline of October 2017 where DF organisations should have achieved ISO/IEC 17025 certification. The International Organisation for Standardisation (n.d.) defines ISO/IEC 17025 as follows.

> 'ISO/IEC 17025:2005 specifies the general requirements for the competence to carry out tests and/or calibrations, including sampling. It covers testing and calibration performed using standard methods, non-standard methods, and laboratory-developed methods.' (International Organisation for Standardization, n.d.)

To provide an example, a non-standard method may simply involve procedures for investigating and interpreting data from an unfamiliar or undocumented system, device or file system (Bryant, 2016, p.144). Such methods are likely to be unpublished and not subject to field-wide review. Conversely, standard methods include those which have been tested, validated and documented, typically resulting in acceptance by those in the field, and are widely used within the confines of which they were developed for, such as conventional disk imaging processes (National Institute of Standards and Technology, 2016). Finally, laboratory-developed methods include those procedures constructed internally by staff to provide an effective analysis of a device where no existing documented procedures are available. The United Kingdom Accreditation Service (2016) suggest a lot of investigatory methods fall within the category of 'laboratory developed'.

A push for ISO/IEC 17025 certification in DF is seen as a method of ensuring standards of quality and organisational competence (Barbara, 2012; House of Commons Science and Technology Committee, 2016), bringing DF into line with other forensic disciplines (Beckett,and Slay, 2007). Such accreditation has also been previously called for by the Committee on Identifying the Needs of the Forensic Science Community in 2009 (p.25) who stated that 'laboratory accreditation and individual certification of forensic science professionals should be mandatory'. In addition, adherence to such quality standards is argued as ensuring the maintenance of "public confidence and to reduce the potential for miscarriages of justice" (Home Office, 2016 p.9). Whilst the imposition of this requirement upon DF has not been met with universal acceptance (see Professor Peter Sommer et al's. (2017) recent practitioner survey regarding consensus over the proposed mandatory adoption of ISO/IEC 17025), with arguments of cost and complexity being mooted, it remains an impending prerequisite for DF laboratories.

As stated above, ISO/IEC 17025 concerns establishing requirements for competence, 'for a management system for providers of laboratory-based forensic science services to demonstrate their ability to deliver consistently products and services that meet the requirements of their customers in the Criminal Justice System (CJS)' (Forensic Science Regulator, 2016. p.12). Testing and calibration form one of ISO/IEC 17025's core targets in order to regulate the development of forensically sound procedures of this type, ensuring that any techniques utilised are 'capable of meeting the requirements of the 'trier of fact'' (Guo, Slay and Beckett, 2009). Testing and validation forms a major part of a DF practitioners role to ensure the validity of results gathered during an investigation and proof of such processes must be documented in line with section 5.4 of ISO/IEC 17025. The Forensic Science Regulators (2014) Codes of Practice and Conduct for Digital Forensic Services draws reference to this need.

> ‘The provider shall ensure that, for the range of the digital forensics methods it uses, the validation requirements take account of staff competency levels, the nature and difficulty of the tasks to be carried out, and the level of acceptability of the method in the wider forensic science and criminal justice community’. Forensic Science Regulators (2014, p5 at 6.1.4)

Methods used for testing and validation must be robust as well as be designed to allow effective validation to take place. Procedural developments for testing have been offered by the National Institute of Standards and Technology’s (2015) Computer Forensics Tool Testing (CFTT), which focuses on a core set of functions including disk imaging, carving and string searching. These methods provide a foundation from which to build future testing and for validation purposes, yet it must be noted that they are non-exhaustive and do not cover all of the available techniques and software utilised by practitioners currently. In addition, they target the validation of specific vendor tools. In addition, the Scientific Working Group on Digital Evidence (2017a) provide best practice guidance regarding digital data analysis. Of particular interest is Scientific Working Group on Digital Evidence (2017b) documentation regarding error mitigation in DF tools.

Whilst both SWGDE and NIST provide guidance and support to practitioners seeking to validate procedures, they remain focused on specific areas of an investigation. In 2009, Slay et al., indicated that there was a lack of ‘verifiable, repeatable testing protocols ...and a new paradigm should be adopted that treats a tool or process independently of the mechanism used to validate it’. Arguably this situation remains, leading to the proposal of FRED offered in this article. Flandrin et al., (2014) suggest that existing methodologies for validation and testing are often either too complex to be effectively utilised or lack extensive coverage of the required aspects of an investigation. FRED is an attempt to simplify the design, implementation and testing process, whilst providing a formalised framework from which to develop forensically sound testing/validation procedures.

3 The methodology proposal

Sommer (2010, p12) stated that ‘regulatory trends in forensic science point strongly to the need for exhaustive testing of all findings and tools. At the same time a number of jurisdictions suggest a judicial test for the admissibility of novel scientific evidence. But in fields such computers and cellphones, the rate of change is faster than the normal times required for peer-reviewed publication’. As a result, establishing effective strategies for discovering the facts associated with a set of digital data is a difficult task. The rate of change in technology exacerbates this issue, as the lifespan of some analysis techniques or interpretative guidance can be short lived due to the rate of release of updated versions of software which may change the internal structure or fundamental workings of an application and its associated data (see for example Mozilla (2017) Firefox, with over 150 version updates across 15 years). With this in mind, frameworks which are artefact-specific can be of limited use. Yet frameworks providing a practitioner with the support to implement robust testing on any form of digital data are arguably a valuable tool and remain in short supply.

In the United States, following the case of *Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993)*, the ‘Daubert Test’ was established to assess the reliability of expert evidence (see Scanlon and Kechadigy (2014) for a detailed breakdown of the test). This four stage

assessment evaluates the testing carried out on a procedure, it's known error rates, publication and subsequent peer-review of the process and acceptance of the procedure within the specific domain's scientific community. Although organizations such as the National Institute of Standards and Technology (2015) and programs such as the Computer Forensics Tool Testing (CFTT) Project offer some support, largely, the burden of obtaining reliable 'Daubert Standard' forensic evidence lies with the practitioner themselves.

Where currently frameworks typically fall within one of two categories, either 'generic procedural' covering the stages of an investigation or 'artefact-specific' where a specific application is targeted for analysis, the proposed FRED methodology targets the research and testing of digital data phase (see Figure 1). Typically at this stage, practitioners are trying to establishing a chain of events by attempting to understand the function and evidential worth of any given digital data, often attempting to simulate/reconstruct an event. Erroneous decision making and testing flaws at this point could undermine any future results, jeopardizing a case.

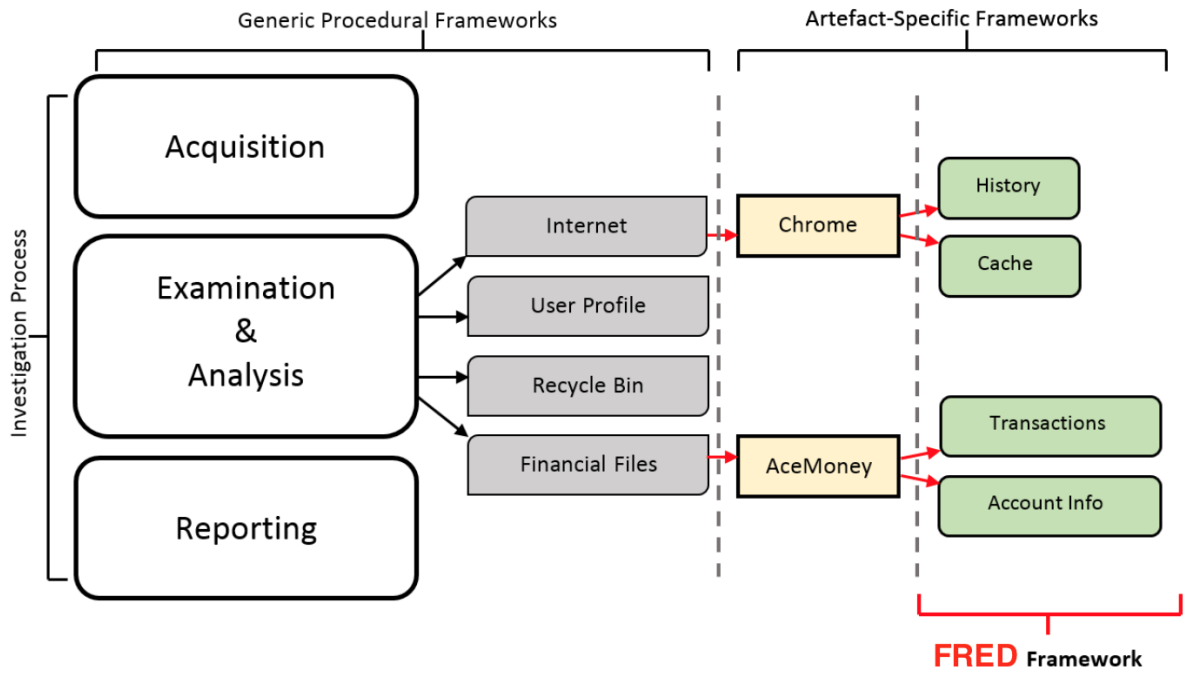

Figure 1: The application of the FRED research framework. FRED is application independent, and functions as a support mechanism during research and testing to ensure robust findings from which to build case hypotheses.

As part of ISO/IEC 17025, Section 5.4 requires organisations to identify whether methods used are acknowledged/accepted by the field (either nationally or internationally) and that relevant, up-to-date standards are utilised. Generally, all testing methods must be validated, whether standard or non-standard (Forensic Science Regulators, 2016). The problem remains that there are limited frameworks which provide support for testing and validation of results at the level which FRED is proposed. Whilst SWGDE and NIST both provide targeted testing at top level procedures such as disk imaging, FRED is aimed at validating and testing the forensic interpretation of relevant artefacts, from which thousands may exist without standards for effective parsing and evaluation of information. Currently there are limited

frameworks defining the stages of this process and the requirements involved. FRED offers a standardised procedures for developing and documenting robust testing of this type.

Figure 2 documents the structure of FRED and its six core stages, namely 'plan', 'implement', 'evaluate', 'repeat', 'analyse' and 'confirm' and each stage is discussed in depth in Section 3.

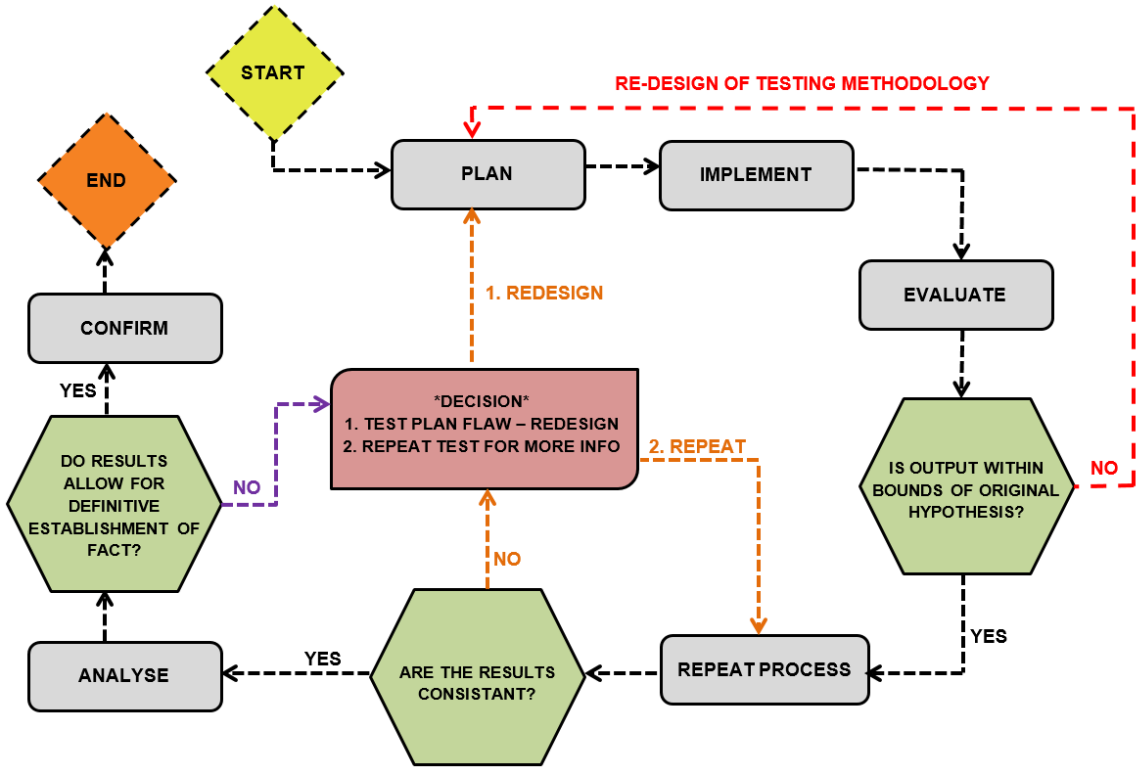

Figure 2: The FRED research framework.

3.1 Plan

The planning stage of any research provides a solid base from which to acquire a reliable set of results. Defining a suitable plan may in some instances be straight forward and in others be multifaceted, dependant upon the complexity of the digital data structures under investigation. As part of the planning phase it should be clear what any research is trying to establish, therefore it is necessary to define the goal of the research from the outset.

i. What is the goal?

What may seem an obvious question can easily be overlooked. Those undertaking DF research need to be clear on what is hoped to be achieved by the research/test which is planned. Failure to define specific goals may result in nonspecific and inapplicable findings from a given test. To provide an example, goal setting can be as simple as establishing where Internet history is stored by a particular Internet browsing application or as complex as reverse engineering the internal metadata structure of any Internet history data. As part of the process of determining a research goal, reconnaissance of any existing documented findings (published or vendor specific manuals) should not be discounted as a resource from which to acquire initial and 'potentially' accurate understanding of an artefact from which to derive a research hypothesis. Although where research into unconventional applications/artefacts is taking place, such resources are unlikely to exist.

To achieve the research goal, a plan must take into consideration the potential need to dynamically adapt to unforeseen issues, events and newly acquired knowledge from preliminary findings. A plan must facilitate testing which is comprehensive, with consideration given not only to the need to find evidential content of a particular type, but also to prove that it also does not exist in a different form elsewhere. To provide an example, consider the development of a plan to examine the basic content of the ThumbCache in Windows 7 systems. The ThumbCache maintains small thumbnail images of files viewed in 'thumbnail view', collected from across the operating system (for an indepth discussion of the ThumbCache see Parsonage (2012) and Quick et al., (2014)). A plan to establish the content and structure of the ThumbCache forms the goal of the plan, with a hypothesis of 'the ThumbCache contains thumbnail images' derived from an initial review of published documentation.

The ThumbCache can be thought of as a linear structure of images, which when approached procedurally, the user may discover that the first thumbnail stored is of type .jpg. It would at which point be easy to assume that all stored thumbnails are of type .jpg and assume that the goal of testing is achieved. Yet, the Thumbcache also stores thumbnail images in .png format (see for example the thumbcache_256.db located at C:\Users\<UserName>\AppData\Local\Microsoft\Windows\Explore and .bmp for the thumbcache_96.db). In this case (and all testing cases), a developed plan, must be capable of exhausting all possibilities before assumptions can be made on the attainment of the research goal. This need for exhaustive planning is documented in Figure 3. Here, following planned 'Route 1' would result in an erroneous interpretation of the thumbcache file structure. Testing via 'Route 2' would provide a comprehensive understanding of the artefact.

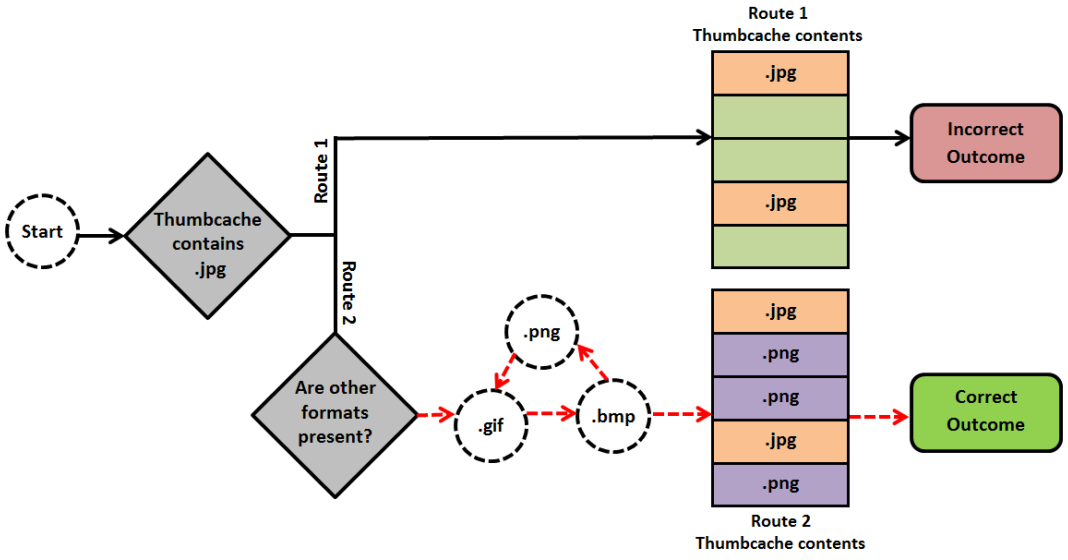

Figure 3: A testing example.

Consideration should also be given to the experimental conditions and environment used during testing.

ii. Test Conditions and Environment

Consideration of the testing environment is key to ensuring the validity of outputs. Test results must be accurate, repeatable and applicable, where all three variables are subject to the platform from which testing is taking place. To provide a basic example, research into the functionality of the Microsoft Windows Recycle Bin for a Windows 7 operating system is not applicable to legacy systems such as XP. The same can be determined of software applications and their artefacts subject to scrutiny from a case under investigation, results derived from testing need to be applicable to the exact conditions which were experienced by a defendant. Suitable setup of the testing environment is therefore crucial. The following options are available to the user.

1. *Flat analysis*: Flat analysis involves the examination of evidence captured within a standard forensic image file format (i.e. .E01, .dd). This can either be from a live investigation or after a test environment has been imaged in a forensic image format. Although this method captures the state of a system for a given point in time it has the following disadvantages:
    a. *It is static*: This approach prevents dynamic testing from taking place due to an image being a snapshot in time. Therefore a real-time examination of system-triggered events during testing is not possible. Instead, system events must be initiated by test data, then captured immediately after via the imaging process.
    b. *A partial picture*: As an image is a snapshot it may only provide a partial description of events triggered from test data (depending on the time that the image was captured and the time taken for system events to initiate and complete after test data has been used) or from a defendant's device. To enable a greater understanding of events will likely require significant repetition of the testing and imaging process. This process can be time consuming and resource intensive (Ayers, 2009).
2. *Reverse image:* The reverse image and boot is a well established process for replicating the exact function of a suspect system. The process involves writing an image file back to a separate hard disk drive, providing a bootable version of a defendant's device. The benefits of this approach is that testing can take place in an environment which is exactly the same as those experienced by a defendant and the researcher has access to all of their applications, setup using a defendant's configurations, in a usable environment. The environment is dynamic, where a user can test multiple functionality in one session. When complete, the device can be powered off and accessed using traditional write-protection equipment and examined. However the following issues must be considered.
    a. *Time and Resources*: Setting up the environment takes time to complete. This includes the time taken to image the original device then write it to a secondary device in order to boot from it.
    b. *Validity*: Once the device is booted, thousands of system changes are occurring every second. Further, after a first set of testing, the environment is contaminated with test data. At which point, the environment would either need to be wiped and re-reversed imaged again (effectively sanitized of all additional interactions) or any testing plan must be dynamically adapted to accommodate new unique test data in order to differentiate actions across all

separate test which are run (see Section 3.2 for a discussion of unique test data and its impact).

c. *No control*: Once the system is live, it is subject to change by applications installed by a defendant and potentially unknown to the researcher. Contamination of results from 3rd party sources on the device, outside the confines of the test plan is possible and have to be assessed and any impact addressed with 100% reliability.

3. *Utilising virtual machine configurations*: Virtual machine (VM) configurations offer a flexible environment implementable with limited resources. There are two approaches to take, the first through the generation of VM configuration files using an application like Virtual Forensic Computing (GetData, 2017) to virtualise a defendant's computer platform (similar to the reverse imaging techniques noted above). The second option, where replication of a defendant's setup is not necessary involves the use of a standard VM configuration (sanitized blank operating system installed from an install disk/.iso) where associated virtual hard drive files can be extracted and interpreted from within a separate digital forensic package. VMs provide an environment which can be sanitized and reconstructed with ease and relatively low cost in terms of time and resources.

The choice of platform from which to implement testing depends upon the goal of the research. If the focus is to replicate events which have occurred in an investigation currently live, (coined in this article as ‘reproductive research’) then a platform/setup capable of the recreation of a defendant's exact setup is crucial to ensuring reliable findings which can apply to that investigation. However, if the focus it simply to develop an understanding of a system artefact/log/process (coined in this article as ‘explorative research’), then a platform choice can be made with regard to factors such as compatibility, support (in terms of which platform the target application/artefact is likely to be found) and available resources.

In addition, any plan needs to be able capable of recreation (see discussion at Section 3.4) and should incorporate a setup and set of procedures which permit this. Recreation is needed both in the pursuit of reliable findings and also so that any procedure can be followed by a third party if necessary to demonstrate how results were obtained. This is a well documented principle in best practice guidelines globally (ACPO, 2012; U.S. Department of Justice, 2004).

3.2 Implement

Following a suitably designed plan, stage two of FRED involves its implementation. Implementation requires the user to carry out a series of actions to simulate user behaviour in line with the planned methodology. These actions constitute the ‘data set’ which is used during testing.

i. Data set: Data set usage is a key factor for consideration and arguably, this element should also be considered during methodology planning. Determining the test data set to be used during any research (to simulate standard user behaviour, which can then be evaluated) involves establishing a suitable set of test content which can be used to derive reliable results from the planned tests, ensuring that all outcomes have been exhausted. Test data can include a set of actions (such as saving or editing a file, or utilising an application's function) or a set of inputs (such as searching for specific terms or creating

specific content). Effective test content needs to be diverse enough to implement testing in sufficient depth to exhaust all potential outcomes and to ensure the functionality of a given application or artefact can be fully understood. Also, at later stages of FRED, a sufficiently disparate and extensive data set allows for the repetition of testing to establish consistent behaviours during the latter stages of FRED.

Contamination is also an element which must be considered, meaning that test data must be unique enough to be identifiable as the actions of the tester, with no possibility of data being derived from alternative sources. The use of a sanitized test environment will support this process (discussed in section 3.1), limiting the potential for contamination. To provide an illustration of the need for the use of an appropriate test dataset to simulate user actions, take for example the testing of an Internet browser application such as Internet Explorer (IE). On installation, IE by default maintains default favorites, bookmarks and a preconfigured homepage (www.msn.com on a default Windows 7 installation). Any test data used must be distinguishable from data derived from sources beyond the confines of the test (such as any pre-established website data create on installation of the application). Therefore in this case, any websites used to test IE's functionality must not already be contained as part of any default data acquired as part of it's installation. Consideration must also be given to the presence of any website data from other sources within a test system setup (other browser applications which may be installed or browsing data from past sessions which may be embedded within system artefacts or the unallocated regions of the system and potentially contaminate any findings). To verify that test data is viable for use in this scenario, it is best practice to first establish that any chosen test data does not exist on the system prior to testing, through preliminary keyword searching and analysis. Issues of contamination can also be limited by pre-test environment verification processes (testing to establish that chosen test data does not already exist) and the use of clean test environments.

3.3 Evaluate

Following a successful test implementation, the outcome must be evaluated. In order to do this, the effect of the implemented tests on a system or series of artefacts must be identified and collected.

i. Identifying and capturing changes brought about by testing: Once a test phase has been implemented, changes to digital data must be identified. These changes represent how a particular artefact/application behaves following system usage, and the data which is being stored documenting these actions. The process of identifying changes is not straightforward, due to the number of modifications occurring on an operating systems (OS) every second and both changes to application specific files and generic operating system structures and log data must be investigated and collected in whole (see Figure 4).

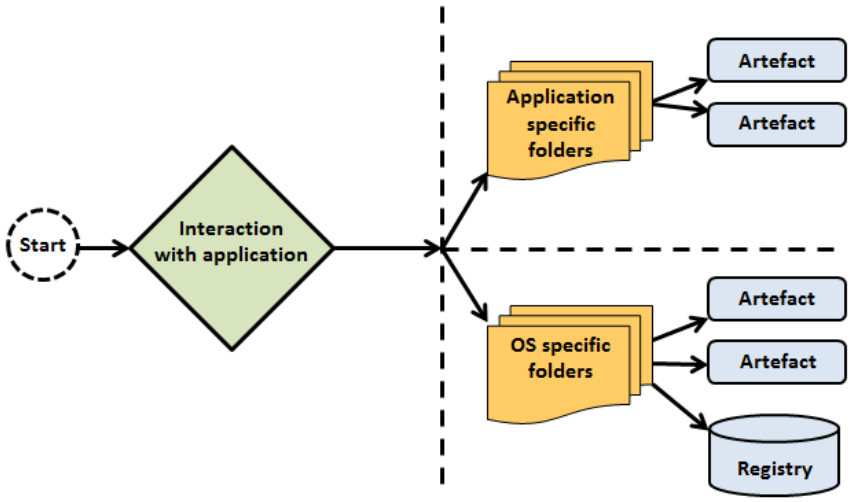

Figure 4: A representation of changes occurring following application testing.

When dealing with an isolated system artefact (for example, examining the Thumbcache as discussed in section 3.1), identifying where changes occur following testing may not prove an issue as data may be confined solely to within that artefact. However, when trying to identify changes brought about by testing of an application which is potentially triggering system wide events, the task poses a greater challenge. Keyword searching techniques can help to support the identification of changes or storage of test criteria used, linking back to the implementation stage of FRED and the need to use unique test data as discussed previously. In addition, the analysis of time stamps can help to highlight potential areas of evidential interest and reduce the amount of redundant files subject to an investigation. After a test process, if a file's modified timestamp remains unchanged, this is an indication that it has remained unaffected by any testing which has taken place (subject to any malicious tampering of file system metadata). Figure 5 provides an example of how focusing on time stamp information can help to separate redundant files from those likely to contain evidential content relating to a process under investigation.

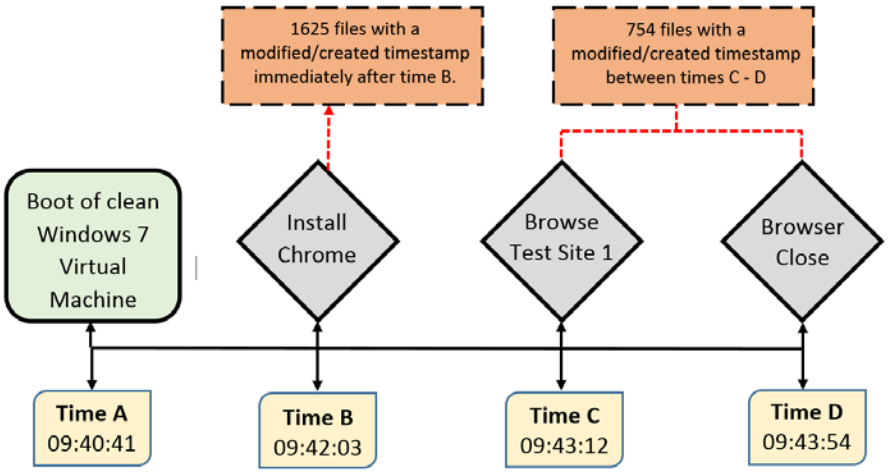

Figure 5: A demonstration of using time stamps to establish files subject to change, corresponding to a process under investigation. Here, Chrome is installed and used to browse a test website, with potentially evidential files being filtered based on timestamp information.

Finally, observing the live process related information associated to an application under investigation can help to identify relevant files. Where the test environment for an application is a Microsoft operating system, Windows Sysinternals's Process Monitor application can provide some support. Process Monitor allows a user to examine 'real-time file system, Registry and process/thread activity' (Sysinternals, 2017), allowing system events to be captured as testing is being carried out. A user can filter events based on a Process Identifier (PID) associated with an application under investigation, allowing system changes associated to to a chosen application to be filtered. Both registry and file system changes associated to a PID can be captured, including where information is being written to disk, suggesting potentially evidential locations. Process Monitor can be installed on a test environment and run as a background process during testing, recording system activity live, which can be collected and examined retrospectively (see Figure 6).

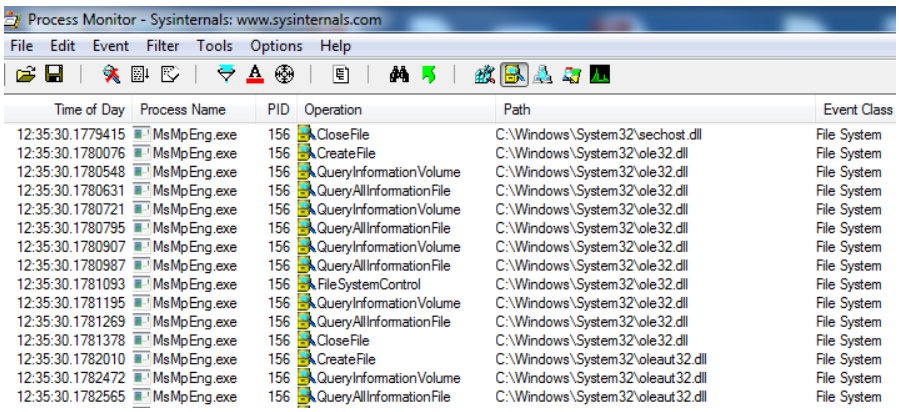

Figure 6: A example of Process Monitor output.

ii. Is the output within the boundaries of your original hypothesis?: It is also important at the evaluation stage to consider whether the result of an implemented test is within the boundaries of what was expected from testing. For example, existing research may have informed the design of the planned test, but the changes which subsequently occurring within a system may be different to those previously documented or what was initially expected. In addition, it may not have been possible to identify system changes after an implemented test. If either situation presents, the following options are available.

1. *Redesign of the test*: The researcher should consider if the research has been sufficiently planned. If weaknesses in methodology are present at this stage, it is necessary to return to the planning stage of FRED and re-design the plan based on the acquired experience of this round of testing.
2. *Re-run the test*: The more times a test is repeated, the greater the chance of establishing consistency in behaviour. Although changes may not be within the confines of initial expectations, it does not mean that the results are incorrect. Establishing consistent behaviour means establishing factual behaviour.
3. *Test data set issues*: Consideration must also be given to the test data used during the implementation phase. If test data is of the wrong type or lack the diversity needed to simulate real-world usage of the application, it may be failing to trigger relevant events on a system.

4. *If changes cannot be detected*: If it is not possible to detect changes within a target artefact of OS, this could be due to the following reasons:
    a. Changed data may be compressed or encrypted. As a result keyword searching may return limited hits (subject to prior decompression / decryption processes being run). However, in such instances, timestamp information should still reveal that changes have occurred. If it is suspected that content may be encrypted, then a greater challenge presents itself. First a practitioner needs to be able to actually identify that content is in fact encrypted and not simply obfuscated or stored in a way which is interpretable by an application under investigation, but not readable without some form of data conversion. If encryption is in use (for example, where the timestamps of a file suggest internal modifications are occurring following test actions), then decryption may be an option during testing. Successfully decrypting a target file may provide access to modified content as a result of testing and help to understand what is occurring following test actions. However, in order to decrypt the content, a practitioner needs to identify the encryption algorithm in use and passphrase or protocol to trigger decryption. Depending on the method and strength of any encryption method, the feasibility of this process may vary. Yet the success of any decryption process may mean the difference between understanding the process under investigation by acquiring access to all data, or missing crucial data linked to an application's usage.
    b. Test data may not be comprehensive / of the wrong type, in order to trigger sufficient events by an application under investigation.
    c. Misunderstanding of the artefact / application under investigation meaning that it functions fundamentally different to what was originally planned for, requiring re-planning of testing methods.
    d. Evidential changes may occur external to any local storage. Consider that information may be stored in cloud / server side of even in physical memory (particularly relevant for privacy-enhancing applications).
    e. User error. Is the researcher doing something fundamentally wrong and as a result, is failing to detect changes which are there to be found.

If changes can be detected and collected, it is crucial to establish whether these events can be consistently reproduced.

3.4 Repeat

Repeatability is the key to establishing robust knowledge as the standard to be achieved through the utilisation of FRED is that of fact, where repeatability is a useful measurement of reliability. Reliability cannot be established from one test alone, and result must be capable of being replicated. However, it must be noted that it is the testing process that must be capable of repetition. Repeating a process should result in consistent outputs; however, even when the same actions are carried out, testing can also reveal an application’s behavioural inconsistencies. Establishing consistency requires running the same test procedures with the same data on multiple occasions. Then taking knowledge this acquired knowledge and running the test with variants of this test data to confirm actions. If a test cannot be repeated, a researcher may need to revisit the planning stage of FRED to revise the test plan.

The goal of any testing is to develop a repeatable structure which offers consistency in application and soundness of decision making. A suitably devised plan must consider repeatability in order to differentiate between results which are reliable and those which have been generated by anomalies, one time events or mere coincidence. Yet the issue of repeatability also raises the question, 'how many times does a test have to be repeated before the results are reliable?'. There is no universal answer, and decision making regarding the number of repetitions of a test is one which is driven by available resources in terms of time equipment, whilst considering the practical feasibility of the research. It is not feasible to perform an infinite number of tests, therefore the threshold of repeatability of any testing must be decided by the researcher and will be influenced by multiple factors including the type of process subject to investigation and the diversity of test data being utilised. In any circumstance, to establish reliability in results, a testing procedures must have been repeated beyond a negligible number of times, noting the requirement of some legal evidential admissibility tests such as the previously discussed Daubert standard which require known error rates.

Repeatability and consistency of results is also a requirement for those who seek to develop an accurate parsing algorithm in order to automate any devised evidence recovery and interpretation procedures. Here consistency (and deviations) in metadata structures must be correctly identified to ensure automated processes are not overlooking potentially relevant content. The development and use of pseudo-code (a simplified description of a program and its structure) to test any identified metadata, file offsets and internal structures will support this process.

If consistency in output can be established through repetition of testing, the results can then be analysed.

3.5 Analyse

The analysis stage involves the interpretation of results generated from testing carried out and collected during the 'evaluate' and 'repeat' stages of FRED. At this point, a researcher should have at least two sets (likely more) of results generated from the same test or associated test processes. The focus of the analysis stage is to be able to explain definitively how an artefact/application functions and what are the results of certain user actions. This task is not straightforward as Chessman (2017) indicates that establishing the true extent of functionality requires a thorough examination of an application's source code, which in most instances is either not possible or not practical. Instead, most DF research is undertaken using functional testing (also referred to as behavioural or black-box testing) (Khan and Khan, 2012), where test actions are carefully selected to examine expected outputs.

It is important to be able to associate actions with their associated effects on a given system (See Figure 7). A researcher should be able to establish that following test action 'A' (for example, when investigating an Internet browser, the act of saving a bookmark), this action results in a series of system changes ('1', '2' and '5', which may be system or application specific artefacts).

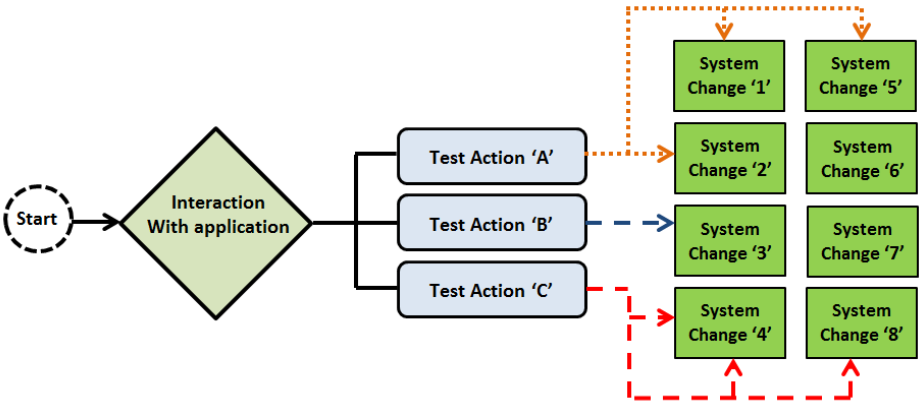

Figure 7: A demonstration of tying test actions to system changes.

Establishing system changes related to test actions is a two way process. Once an action can be associated to a change, the process can be reversed. This is particularly important in investigation scenarios where only post-mortem events are present on a defendant's system. Therefore using Figure 7 again as an example, when an investigator identifies system changes '1', '2' and '5' as evidential, these traits can be correlated to an action 'A'.

The goal of the analysis phase is to determine reliably whether as the result of testing, the findings allow a researcher to confirm the facts associated with testing. If a factual outcome cannot be established, then two options present themselves, a revision of the test plan and its structure or carry out further testing (with a review of test data used) to identify why there is a lack of consistency in results.

3.6 Confirm

The final stage of FRED is the ability to affirm as a matter of fact, the outcome and interpretation of testing which has taken place and to document the process. A researcher should at this stage be able to factually establish that when investigating an application/artefact, say 'X', using FRED, user actions 'Y' result in outcome 'Z'. At which point, testing procedures can be documented and methodologies formalised for peer review, demonstrating rigorous underpinning testing. Consideration must also be given to the fact that the results generated from testing must be defensible, therefore a core requirement on confirmation is to fully document all procedures in line with the stages of FRED.

Transparency in the testing carried out allows for the reliability of any research to be objectively assessed and any potential weaknesses in the validity of results to be highlighted before they are wrongfully incorporated into any investigation. Sallavaci and George (2013) state that new regulatory requirements for experts, particularly in England and Wales are likely to put a DF expert witness in a 'defensive mode', trying to justify every evaluative step, when writing their reports'. Utilising FRED supports a practitioner to demonstrate the use of thorough planning, rigorous testing and valid interpretations, which can be relied upon in a court of law.

4 Conclusion

With the speed of technological developments, DF practitioners will frequently encounter applications and artefacts during an investigation maintaining functionality which is not fully

understood. In these situations, reliance is placed either on existing or the development of valid research into these areas. The reliance on DF research by practitioners is substantial, where often it underpins the findings of an investigation. The consequences of non-valid research findings can be severe for all involved, specifically in a criminal process. This article has offered FRED to support those engaged in DF research at all levels including industry and academia to support the development of best practice in this area. Such methods for supporting validation and testing are increasingly important given current requirements for compliance with ISO/IEC 17025. FRED provides a framework from which to design, develop and implement DF testing and validation, supporting the generation of reliable, repeatable and documentable outputs. FRED also encourages transparency in the research and testing process to allow for effective peer-review and for a thorough assessment of the reliability of any work carried out.

References

ACPO (2012) 'ACPO Good Practice Guide ACPO Good Practice Guide for Digital Evidence' Available at: http://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf (Accessed 1 April 2017)

Atkinson, John S. (2014) 'Proof Is Not Binary: The Pace and Complexity of Computer Systems and the Challenges Digital Evidence Poses to the Legal System' 2 Birkbeck L. Rev. 245-262

Ayers, D., 2009. A second generation computer forensic analysis system. *digital investigation*, *6*, pp.S34-S42.

Barbara, John (2012) 'ISO/IEC 17025:2005 Accreditation of the Digital Forensics Discipline' Available at: https://www.forensicmag.com/article/2012/02/isoiec-170252005-accreditation-digital-forensics-discipline (Accessed 6 July 2017)

Baryamureeba, V. and Tushabe, F., 2004, August. The enhanced digital investigation process model. In *Proceedings of the Fourth Digital Forensic Research Workshop* (pp. 1-9).

Beckett, J. and Slay, J., 2007, January. Digital forensics: Validation and verification in a dynamic work environment. In System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on (pp. 266a-266a). IEEE.

Beebe, N., 2009, January. Digital forensic research: The good, the bad and the unaddressed. In *IFIP International Conference on Digital Forensics* (pp. 17-36). Springer Berlin Heidelberg.

Biros, D.P., Weiser, M., Burkman, J. and Nichols, J., 2008. Information Sharing: Hackers vs Law Enforcement.

Bryant, R. ed., 2016. *Policing digital crime*. Routledge.

Bulbul, H.I., Yavuzcan, H.G. and Ozel, M., 2013. Digital forensics: an analytical crime scene procedure model (ACSPM). *Forensic science international*, *233*(1), pp.244-256.

Carrier, B. and Spafford, E.H., 2004, July. An event-based digital forensic investigation framework. In *Digital forensic research workshop* (pp. 11-13).

Chessman, Christian (2017) 'A Source of Error: Computer Code, Criminal Defendants, and the Constitution' 105 Cal. L. Rev. 179-228

Christensen, A.M., Crowder, C.M., Ousley, S.D. and Houck, M.M., 2014. Error and its meaning in forensic science. *Journal of forensic sciences*, *59*(1), pp.123-126.

Committee on Identifying the Needs of the Forensic Science Community (2009) 'Strengthening Forensic Science in the United States: A Path Forward' ISBN: 0-309-13131-6, 352 pages

Criminal Practice Directions [2015] EWCA Crim 1567

Crown Prosecution Service (2014) 'Expert Evidence' Available at: http://www.cps.gov.uk/legal/assets/uploads/files/expert_evidence_first_edition_2014.pdf (Accessed 1 April 2017)

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993)

Elyas, M., Maynard, S.B., Ahmad, A. and Lonie, A., 2014. Towards a systemic framework for digital forensic readiness. *Journal of Computer Information Systems*, *54*(3), pp.97-105.

Flandrin, F., Buchanan, W., Macfarlane, R., Ramsay, B. and Smales, A., 2014, September. Evaluating digital forensic tools (DFTs). In 7th International Conference: Cybercrime Forensics Education & Training.

Forensic Science Regulator, (2014) Codes of Practice and Conduct Appendix: Digital Forensic Services FSR-C-107 Issue 1

Forensic Science Regulator, (2016) Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System (Issue 3: February 2016)

Garfinkel, S.L., 2010. Digital forensics research: The next 10 years. *digital investigation*, *7*, pp.S64-S73.

Gov.uk (n.d.) 'Application for compensation after a miscarriage of justice' Available at: https://hmctsformfinder.s3.amazonaws.com/forms/guidance/index.htm#jump-2 (Accessed 1 April 2017)

Guo, Y., Slay, J. and Beckett, J., 2009. Validation and verification of computer forensic software tools—Searching Function. digital investigation, 6, pp.S12-S22.

Hitchcock, B., Le-Khac, N.A. and Scanlon, M., 2016. Tiered forensic methodology model for Digital Field Triage by non-digital evidence specialists. *Digital Investigation*, *16*, pp.S75-S85.

Home Office (2016) 'Forensic Science Strategy,A national approach to forensic science delivery in the criminal justice system' Cm 9217

House of Commons Science and Technology Committee, 'Forensic Science Strategy' (Fourth Report of Session 2016–17) HC 501, Published on 17 September 2016. Available at: https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/501/501.pdf

International Organisation for Standardization, n.d.a 'ISO/IEC 17025:2005' Available at: https://www.iso.org/standard/39883.html (Accessed 6 July 2017)

International Organisation for Standardization, n.d.b 'The facts about certification' Available at: https://www.iso.org/certification.html (Accessed 6 July 2017)

International Organisation for Standardization, n.d.c 'Popular Standards' Available at: https://www.iso.org/popular-standards.html (Accessed 6 July 2017)

Ireland, Jane and John Beaumont, (2015) "Admitting scientific expert evidence in the UK: reliability challenges and the need for revised criteria – proposing an Abridged Daubert", Journal of Forensic Practice, Vol. 17 Issue: 1, pp.3-12, doi: 10.1108/JFP-03-2014-0008

Jang, Y.J. and Kwak, J., 2015. Digital forensics investigation methodology applicable for social network services. *Multimedia Tools and Applications*, *74*(14), pp.5029-5040.

Jerman-Blažič, B. and Klobučar, T., 2016. Towards the Development of a Research Agenda for Cybercrime and Cyberterrorism–Identifying the Technical Challenges and Missing Solutions. In *Combatting Cybercrime and Cyberterrorism* (pp. 157-174). Springer International Publishing.

Karie, N.M. and Venter, H.S., 2015. Taxonomy of challenges for digital forensics. *Journal of forensic sciences*, *60*(4), pp.885-893.nist

Karyda, M. and Mitrou, L., 2007, August. Internet forensics: Legal and technical issues. In *Digital Forensics and Incident Analysis, 2007. WDFIA 2007. Second International Workshop on* (pp. 3-12). IEEE.

Khan, M.E. and Khan, F., 2012. A comparative study of white box, black box and grey box testing techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *3*(6).

Köhn, M., Olivier, M.S. and Eloff, J.H., 2006, July. Framework for a Digital Forensic Investigation. In *ISSA* (pp. 1-7).

Levett, L.M. and Kovera, M.B., 2008. The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence. *Law and human behavior*, *32*(4), pp.363-374.

Lillis, D., Becker, B., O'Sullivan, T. and Scanlon, M., 2016. Current Challenges and Future Research Areas for Digital Forensic Investigation. *arXiv preprint arXiv:1604.03850*.

Martini, B. and Choo, K.K.R., 2012. An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, *9*(2), pp.71-80.

Mozilla (2017) 'Firefox Releases' Available at: https://www.mozilla.org/en-US/firefox/releases/ (Accessed 10th April 2017)

Nance, K., Hay, B. and Bishop, M., 2009, January. Digital forensics: defining a research agenda. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* (pp. 1-6). IEEE.

National Institute of Standards and Technology (2015) 'Welcome to the Computer Forensics Tool Testing (CFTT) Project Web Site.' Available at: https://www.cftt.nist.gov/ (Accessed 10th April 2017)

National Institute of Standards and Technology (2016) 'Disk Imaging' Available at: https://www.cftt.nist.gov/disk_imaging.htm (Accessed 4th September 2017)

Naughty, Michael (2007) 'Rethinking Miscarriages of Justice: Beyond the Tip of the Iceberg' Springer, pg. 18

Naughton, M. and Tan, G., 2010. *Claims of Innocence: An Introduction to Wrongful Convictions and How they Might be Challenged*. University of Bristol.

Olivier, M., 2016. Digital forensic science: a manifesto: viewpoint. *South African Computer Journal*, *28*(2), pp.46-49 at p.49

Parsonage, Harry (2012) 'Under My Thumbs – Revisiting Windows thumbnail databases and some new revelations about the forensic implications.' AVailable at: http://computerforensics.parsonage.co.uk/downloads/undermythumbs.pdf (Accessed 10th April 2017)

Pilli, E.S., Joshi, R.C. and Niyogi, R., 2010. Network forensic frameworks: Survey and research challenges. *digital investigation*, *7*(1), pp.14-27.

Quick, D., Tassone, C. and Choo, K.K.R., 2014. Forensic analysis of windows thumbcache files. Available at: https://pdfs.semanticscholar.org/97a4/a135acf7ef534992e18f643f577a6749cb3e.pdf (Accessed 10th April 2017)

Raghavan, S., 2013. Digital forensic research: current state of the art. *CSI Transactions on ICT*, *1*(1), pp.91-114.

Reith, M., Carr, C. and Gunsch, G., 2002. An examination of digital forensic models. *International Journal of Digital Evidence*, *1*(3), pp.1-12.

Rogers, M.K. and Seigfried, K., 2004. The future of computer forensics: a needs analysis survey. *Computers & Security*, *23*(1), pp.12-16.

Sallavaci, O. and George, C., 2013. New admissibility regime for expert evidence: the likely impact on digital forensics. *International Journal of Electronic Security and Digital Forensics*, *5*(1), pp.67-79.

Scanlon, M. and Kechadi, T., 2014. Digital evidence bag selection for P2P network investigation. In *Future Information Technology* (pp. 307-314). Springer Berlin Heidelberg.

Scott, L., 2009. It Never, Ever Ends: The Psychological Impact of Wrongful Conviction. *Crim. L. Brief*, *5*, p.10.

Selamat, S.R., Yusof, R. and Sahib, S., 2008. Mapping process of digital forensic investigation framework. *International Journal of Computer Science and Network Security*, *8*(10), pp.163-169.

Slay, J., Lin, Y.C., Turnbull, B., Beckett, J. and Lin, P., 2009, January. Towards a formalization of digital forensics. In IFIP International Conference on Digital Forensics (pp. 37-47). Springer, Berlin, Heidelberg.

Sommer, Peter (2010). Forensic science standards in fast-changing environments. Science and Justice, 50(1) pp. 12–17 at pp.12

Sommer, Peter, Pat Beardmore, Geoff Fellows (2017) 'UK ISO 17025 Digital Forensics Survey April 2017: Results' Available at: http://digital-evidence.expert/UK%20ISO%2017025%20Digital%20Forensics%20Survey%20April%202017.pdf (Accessed 6th July 2017)

Spyridopoulos, T., Tryfonas, T. and May, J., 2013, October. Incident analysis & digital forensics in SCADA and industrial control systems. In *System Safety Conference incorporating the Cyber Security Conference 2013, 8th IET International* (pp. 1-6). IET.

Sremack, J.C., 2007, January. The Gap between Theory and Practice in Digital Forensics. In *Proceedings of the Conference on Digital Forensics, Security and Law* (p. 85). Association of Digital Forensics, Security and Law.

Scientific Working Group on Digital Evidence (2017a) 'SWGDE Current Documents' Available at: https://www.swgde.org/documents (Accessed 6th July 2017)

Scientific Working Group on Digital Evidence (2017a) 'SWGDE Establishing Confidence in Digital Forensic Results by Error Mitigation Analysis' Available at: https://www.swgde.org/documents/Current%20Documents/SWGDE%20Establishing%20Confidence%20in%20Digital%20Forensic%20Results%20by%20Error%20Mitigation%20Analysis (Accessed 6th July 2017)

Sysinternals (2017) ‘Process Monitor v3.32’ Available at: https://technet.microsoft.com/en-gb/sysinternals/bb896645 (Accessed 10th April 2017)

United Kingdom Accreditation Service (2016) ‘ISO/IEC 17025 Accreditation for Forensic Cell Site Analysis – An overview’ Available at: https://www.ukas.com/download/development_pilot_programmes/Cell%20Site%20Analysis%20Project%20-%20ISO17025%20Accreditation%20of%20Cell%20Site%20Analysis%20-%20An%20overview(2).pdf (Accessed 4th September 2017)

U.S. Department of Justice (2004) ‘Forensic Examination of Digital Evidence: A Guide for Law Enforcement’ Available at: https://www.ncjrs.gov/pdffiles1/nij/199408.pdf (Accessed 10th April 2017)

von Solms, S., Louwrens, C., Reekie, C. and Grobler, T., 2006. A control framework for digital forensics. *Advances in Digital Forensics II*, pp.343-355.