Journal of Digital Forensics, Security and Law, Vol. 6(1)

Column

Putting the Science in Digital Forensics[1]

By Fred Cohen

In a recent study,[2] digital forensics was found to lack a consensus around even the most basis notions and terminology of the field. To quote: "These two preliminary studies individually suggest that (1) scientific consensus in the area of digital forensic evidence examination is lacking in the broad sense, but that different groups within that overall community may have limited consensus around areas in which they have special expertise, and (2) that the current peer-reviewed publication process is not acting to bring about the sorts of elements typically found in the advancement of a science toward such a consensus. ... perhaps the most significant challenge may be in the development of a common language to describe the field..."

If we are to progress as a scientific discipline applicable to legal proceedings, digital forensics has to recognize and reasonably apply history and precedent, use common language for effective communication, and limit our findings to what the scientific understanding of the day justifies. As a starting point, a short history of diplomatics and archival science may be helpful, and is provided here. Following that, a set of usages and definitions of terms is proposed for JDFSL and the broader digital forensics community.

Some history

Legal systems over several millennia have had to deal with issues related to the admission and use of informational evidence in legal matters. This ranges from documents associating ownership of property through the emergence of fingerprints as evidence and their near demise. As an overarching science, the areas of archival science and diplomatics are among the oldest and most deeply embedded in the legal systems of the World, and are thus a good starting point.

Archival science started as a scientific body of knowledge at least in ancient Rome, were the records of government were written on wax tablets and transported through underground passageways to the central archives for permanent archival preservation. Such records were tracked and made

[1]This editorial piece is extracted and modified from F. Cohen, "Digital Forensic Evidence Examination", ASP Press, 2011

[2]F. Cohen, J. Lowrie, C. Preston, "The State of the Science of Digital Evidence Examination", IFIP Seventh annual IFIP WG 11.9 International Conference on Digital Forensics, 2011/01/30, also published as a chapter in in "Advances in Digital Forensics VII".(pending).

available to the public only in certified copies produced by the archivists who were government employees trusted to diligently perform their duties (quaestores).[3]

The Justinian code codified the definition of archives as "the place where public records are deposited" … so that "they remain uncorrupted and might be found rapidly by those who request them", and so they "preserve perpetual memory [of] the acts [to which they relate]". These principles and approaches have been taught since 1158 in all of the legal educational systems associated with "common law" and formed the foundation for admissibility of records and reliance upon them. By the 1500s these ideas became a far more widespread subject of research and implementation, and various facets of understanding relating to the trustworthiness of records were studied and put in to practice over the centuries.

In 1681, the archival science was codified into a legal framework[4] which focused on individual documents, their characteristics, genesis, and treatment.[5] Archival science and diplomatics were developed together, and in the 1800s laws were increasingly being formulated taking into account their concepts and methodologies. By the late 1800s, rules of evidence and their foundation were explained in detail and by the early 1900s, they were clearly codified into laws globally. Today, diplomatics is being updated and applied to information age records around the World for public and private archival organizations. It remains the basis for much of the legal system, and as such, forms a scientific basis for understanding digital evidence.[6]

"According to modern diplomatics, a record is a document created (i.e., made or received and set aside for action or reference) in the course of activity as an instrument and by-product of it."[7] The field of diplomatics focuses on the **assessment of the trustworthiness of records**, which is done retrospectively for existing records (and in digital forensics), and prospectively for designing record systems and types. Classic diplomatics associates **trustworthiness** with **authenticity of the records** (they were written at the time claimed and signed by a person competent to produce them). Modern diplomatics defines and assesses "**trustworthiness**" in terms of **reliability**, **authenticity**, **accuracy**, as a basis to **authenticate** a record.

There is a great deal more to know about these issues, and the reader is encouraged to read the references.

---

[3]L. Duranti, "Archival Science", Article in Encyclopedia of Library and Information Science.

[4]Dom Jean Mabillion, "De Re Diplomatica", 1681, Saint-Maur, France.

[5]L. Duranti, "Diplomatics: New Uses for an Old Science", Archivaria 28. 7-27, 1989.

[6]L. Duranti, "Diplomatics", Encyclopedia of Library and Information Sciences, Third Edition DOI: 10.1081/E-ELIS3-120043454, 2010, Taylor & Francis.

[7]Ibid.

Careful use of defined terms

No matter how many tests are performed, except for special cases, DFE results cannot prove a broad claim true.[8] The best that can be done is to show that tests failed to refute hypotheses and to show the extent to which tests were thorough. Reasonably, the most authoritative claim in [opposition] support of a hypothesis regarding DFE is therefore something like:

> "The results of [the tests I did] were [in]consistent with [the hypotheses]."

To the extent that some set of these statements then combine together with logical reasoning, an overarching statement may be made with regard to the claims, perhaps of the form:

> Based on [the basis], I found [traces and events] to be [in]consistent with [claim(s)].

Or in some cases, when this is true:

> In my examinations of [traces and events], everything I found was consistent with [claims] and nothing I found was inconsistent with [claims].

On the other hand, a single refutation disproves a hypothesis, and the least that can be reasonably said if such a refutation is identified is something like:

> "The [procedures I performed] demonstrate that [traces and events] are [inconsistent with / refute] [the hypothesis]."

Thus the methodology of the science of DFE when working on ay particular matter consists of:

- Devising testable hypotheses ($h \in E$)
- Testing those hypotheses against the evidence (T and E) using forensic procedures (P) and logic to determine type C and D consistency by attempting to refute the hypotheses.
- Making properly limited statements about the results of those tests, typically using wording such as that identified above.

There are some other wordings that may apply in other circumstances, and some of the more commonly misused ones are identified below, along with definitions suited to use by the DFE examiner.

By the careful use of these terms and their consistent application, the field of DFE examination may move forward more quickly, and peer reviews undertaken in the field may be able to create a body of work that is meaningful across time and endeavors. But if, as a field, DFE examination is inconsistent,

[8] K. Popper, The Logic of Scientific Discovery (1959), Hutchins and Company, London. ISBN10: 0415278449.

or if the peer review process fails to force compliance with such terminology, then the science is unlikely to proceed as a normal science or at a rapid pace.

Proposed usage

I propose that the following usages be required for all future JDFSL submissions, to be augmented over time only as justified by a demonstrated community consensus. These terms are intended to be mandatory for submissions, enforced in refereeing and editorial processes, and applied uniformly to all who seek to publish in JDFSL.

Traces, events, and records

Trace := (digital forensics) A set of bit sequences produced from the execution of a finite state machine.(FSM)

> **Structured trace** := A trace that follows a particular defined pattern.
>
> **Unstructured trace** := A trace that is not structured. [Typically image data such as from sound, vision, or other external sensors.]
>
> **Derived trace** := A trace generated by the examiner from another trace.
>
> **Constructed trace** := A trace constructed from a reconstruction process.
>
> **C-trace** := Constructed trace.
>
> **Original trace** := A trace produced from evidence in the matter.
>
> **O-trace** := Original trace.
>
> **Complete trace** := A trace containing all inputs, states, and outputs of a finite state machine (FSM).
>
> **Partial trace** := A trace that is not a complete trace.
>
> **Incomplete trace** := A partial trace from which a complete trace cannot be uniquely reconstructed.

Event := (forensics) A claimed, asserted, or stipulated state of affairs or act.

> **Anchor event** := An event asserted by the examiner based on personal experience or other authority and that can be linked to the issues in the case. [e.g., A time stamp from an external mechanism that the examiner has personal knowledge of.]

Record := A document created (i.e., made or received and set aside for action or reference) in the course of activity as an instrument and by-product of it. [All digital records are traces, but not all traces are records]

> **Internal record** := A record meant for transmission over time.
>
> **External record** := A record record meant for transmission across space.
>
> **Legal record** := A record whose existence in writing is required by the juridical and/or administrative system within which it is created.
>
> **Public record** := A record issued by a public person. [see below]
>
> **Nonlegal record** := A record whose written form is discretionary.

Supporting record := A record that helps to carry out activities in which it participates (e.g., a map, note, plan, presentation, etc.) [Does not provide evidence that any such act was actually carried out]

Narrative record := Free-form communications of information (e.g., memos, messages, etc.) [Is not adequate to show that any such act was actually carried out.]

Instructive record := A record that indicates the form in which something is to be presented or done (e.g., manuals, regulations, instructions for filling out forms, etc.)

Enabling record := Records that either (1) enable performance of a mechanism (e.g., firmware or an operating system), (2) execute business instructions (e.g., a workflow application), (3) conduct experiments (e.g., a control program for a robotic mechanism), or (4) data used in or produced by analysis or observation.

Original record := The first manifestation of a complete and effective record, either received or stored, depending on whether the record is external or internal. [This is essentially never available for DFE examination because of it's physical nature.]

Draft := A document prepared for the purpose of correction, and meant to be provisional and temporary.

Copy := A reproduction of another document. [The other document could be an "original", "draft", or another "copy"]

Copy in the form of original := A copy that is identical to the original in all respects, but produced at a later time. [This is a physical copy of the media, which is outside of the realm of digital forensics.]

Imitative copy := A reproduction of both the form and content of a record. [This is what is typically available and called an "exact", "bit image", or "forensically sound", copy in digital forensics.]

Exact copy := (forensics) imitative copy.

Bit image copy := (forensics) imitative copy.

Forensically sound copy := (forensics) imitative copy.

Simple copy := A transcription of the record content. [The text]

Inserts := A copy of a record or part of it contained within another original record.

Medium := (diplomatics) The physical carrier of a record.

Form := (diplomatics) The rules governing the representation of an act in writing.

Archive := (diplomatics) Sedimentations of the natural documentary residue of activities.

Archives := The whole of the documents made or received in the course of activity and kept for action or reference. [In archives, there is one archive for each physical or juridical person, or creator. Therefore, each archives (or archival fonds, the terms being synonyms) is a whole of the records made by one creator and their interrelationships.]

Archival bond := (diplomatics) The relationship of a record to the other records within the archives in which it exists.

Provenance := from the Latin "prōvenīre", which means "to come forth", (pro-, convene, -ant). Identification of the origins and path by which something came to be.

Procedures and processes

Procedure := (diplomatics) A formal sequence of steps by which a transaction is carried out.

Procedure := (forensics) A formal sequence of steps by which an examiner examines traces.

Transaction := an act aimed to create, modify, maintain, or extinguish relationships between two or more physical or corporate persons. [Some acts, especially transactions, occur in writing or other documentary forms, thereby resulting in records.]

Process := (diplomatics) is a series of motions by which a person carries out acts, including those acts involved in a procedure. [These are the physical acts undertaken]

Process := (computers) a sequence of programmed instructions and related data executing within an operating environment. [There is typically a process identification number within the operating system structures, and there may be “threads” of execution by which multiple execution streams are simultaneously available to execute]

Persons

Person := The subject of a right or duty. [They are recognized by the legal system as capable of acts.]

> **Physical person** := A human being.
>
> **Juridical person** := A corporation or similar legal entity.
>
> **Succession** := A position or title. [e.g., The President]
>
> **Public person** := A person with responsibility for the administration of matters regarding the people as a whole [i.e., A person authorized to issue a public record.]
>
> **Private person** := Any person not a public person.
>
> **Author** := The person with the competence (i.e., authority and capacity) to issue the record.
>
> **Writer** := The person competent for the articulation and disclosure of the record.
>
> **Addressee** := The person for whom the record is intended.
>
> **Creator** := The person in whose archives a record exists.
>
> **Originator** := The person responsible for the electronic account or space in which the record was generated or from which it is sent.

Mens rae := A guilty state of mind.

Examination and computers

Analysis := Methods used to determine consistency or inconsistency of traces and events. [Typically, trace typing, generating derived traces, making various comparisons, and other similar processes.]

Interpretation := A cognitive process used by the examiner to understand the nature of traces and events in context and associate them with issues at hand. [It may be thought of as associating meaning with traces and events.]

Attribution := An interpretation of causality. [Typically identifying plausible (cause effect) sequences consistent with available traces and events. Particularizing or individualizing traces to candidate causes.]

Reconstruction := An experiment testing hypothesized causal chains. [Used to demonstrate consistency or inconsistency with hypothesized sequences.]

Presentation := A method by which traces (i.e., latent evidence) are make into something that can be sensed and observed by humans.

Characteristic := Trace type, syntax, and structure.

Feature := Trace content [e.g., Sequences of words, types of spelling errors, etc.]

Symbol set := A mapping between bit sequences and symbols they represent in an alphabet.

Octet := An 8-bit sequence.

Byte := An 8-bit sequence at a defined boundary.

Trace type := The thing that a trace is intended to represent when generated.

Typing := (forensics) A process by which the type of a trace is hypothesized for examination. [Traces may be retyped after further examination based on consistency analysis.]

Particularization := A process by which a typed trace is associated with a specific use or source.

Individualization := A process by which a trace is associated with an single specific person, process, or mechanism.

Identifier := A trace placed in records intended to associate the trace with a particular person, process, or other thing.

Indicator := Traces and/or events often associated with or produced by other known traces, events, or mechanisms.

Equivalent content := (inexact matches) The same content in different format.

Normalization := Conversion into a common commensurable format.

Nominal metrics := Lists of things with no basis for formal comparison.

Ordinal metrics := Implies a partial ordering.

Interval metrics := Implies the ability to count things not against any standard.

Ratio metrics := Additive, comparable, and normalized to a common zero value.

Wording in reports

Suggests := imply as a possibility ("The [traces / events] suggests ...") - calls to mind - propose a hypothesis or possible explanation.

Indicates := a summary of a statement or statements or other content codified ("His statement indicates that ...") OR a defined set of "indicators" are present and have, through some predefined methodology been identified as such ("The presence of [...] (smoke) indicates [...] (fire)")

Demonstrate := exemplify - show - establish the validity of - provide evidence for ("The reconstruction demonstrates that ...")

Correlates := a statistical relation between two or more variables such that systematic changes in the value of one variable are accompanied by systematic changes in the other as shown by statistical studies ("Based on [statistical analysis method(s)], the use of the "KKJ" account is correlated (p=95%) with ...")

Match := an exact duplicate ("These two documents have matching publication dates, page counts, ...")

Similar := A correspondence or resemblance as defined by specified and measured quantities or qualities ("The 18 files were similar in that they all had syntax consistent with HTML, sizes under 1000 bytes, ...")

Relate := A defined and specified link ("The file system is related to FAT32 in that FAT32 was derived from ...")

Associate := Make a logical or causal connection with basis provided. ("I associate these bit sequences with program crashes because …")

A final comment

I can propose, but I cannot dictate. At the end of the day, the enforcement of this or any approach to using defined terms carefully and consistently can only be carried out by the authors and editors of the journal and across the field. The editorial board of JDFSL has the final say, and I encourage all who participate in the field and in this journal to enter the debate and let your voices be heard. This is the only way we will achieve the consensus needed to move forward. At least that's how I see it.