C5 Digital Trace Reference Library (DTRL)

Eoghan Casey, PhD, University of Lausanne, Lausanne, Vaud, SWITZERLAND; Owen Brady, PhD, KCL, London WC2R 2LS, UNITED KINGDOM*

Learning Overview: After attending this presentation, attendees will understand how to structure digital traces in a database in a manner that can be readily referenced by practitioners or queried automatically by tools. Attendees will also learn how digital traces are represented using the evolving Cyber-Investigation Analysis Standard Expression (CASE) standard.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by showing how the DTRL can advance digital forensic tool validation and R&D in digital forensics.

The DTRL implements an ontology-based classification system to help users find digital traces that are relevant to their inquiry and to support systematic analysis of digital traces. The structure of the DTRL enables automated queries to augment tool testing and evidence processing.

The importance of reference databases in forensic science was raised in the President's Council of Advisors on Science and Technology (PCAST) report on forensic science. In some respects, the DTRL is comparable to the Ballistics Toolmark Database with standardized representation of characteristics using an open-source data format.

The DTRL is built on the Digital Evidence Semantic Ontology (DESO) and the Cyber-investigation Analysis Standard Expression (CASE). This open-source ontology-based model is designed to standardize and extend emerging proprietary artifact recording systems, including the University of New Haven's (UNH) Artifact Genome Project (AGP), Magnet Forensics' Artifact Exchange, and Google's® Forensic Artifacts.

The DTRL aims to address the following issues: (1) availability—what artifacts are available on a particular digital evidence source?; (2) selection—from the perspective of the investigation, which ones are required?; (3) correlation—allowing the effective comparison of artifacts from disparate sources to make evidential connections; and (4) reliability—what is the basis behind a particular artifact?

Using CASE to represent digital traces in the DTRL allows any tool to parse the standard structures for a variety of purposes. Digital forensic tools can be tested to determine which indexed entries in the DTRL are fully/partially supported by each tool. For each digital trace, the DTRL cites supporting documentation and/or research results, which provide additional information for developers and practitioners to learn more about specific digital traces. This information is useful for tool development purposes and explaining digital evidence in court.

Specifically, the ontology has three classes: (1) location—the actual pieces of data. These could be pertinent to a boot record, file system, operating system or application file, such as a SQLite Database. In much the same way that Advanced Forensic Framework 4 (AFF4) breaks down an image into its component parts and prioritizes, so does DESO. In fact, it's interesting how the two could be paired together; (2) Type Identifier—each of the trace locations is assigned a Type Identifier that allows identification of common artifacts, irrespective of source. This allows the discovery of connections that had not previously been considered. For example, USB serial number, so that those with a common type can be compared. This also allows for a common reporting format to aid comparison; and (3) provenance—the basis for stating that the data at the stated location is the particular Type Identifier

The DTRL overcomes limitations in emerging proprietary systems in the following ways: (1) expressivity—the ontology-based structure of the DTRL clearly differentiates between multiple facets of a digital trace, whereas other systems do not differentiate clearly between the different facets. In UNH's AGP, an entry that is classified as a File object can also contain a username and password. UNH's AGP uses tags on the File object to label such important characteristics rather than distinctly representing these characteristics and their relationship with the File object. Simply put, in general, the AGP does not capture artifacts, but the File in which they are contained. The AGP user is then required to discover what piece of data within the File is relevant and what the data represents. This lacks precision and does not provide the required coordinates for automation; (2) non-ambiguity—the ontology-based structure of the DTRL represents digital traces and their context in a non-ambiguous manner, whereas other systems have ambiguity. For instance, searching UNH's AGP for Media Access Control (MAC) addresses using "MAC" on its keyword and tags function also returned hits relating to "Macintosh." Some users might misinterpret the "MAC" tag as Message Authentication Code; and (3) extensibility—CASE can support many kinds of digital trace, and can be extended to represent new traces as needed.

Digital Traces, Digital Forensic Standards, Tool Testing and Validation

Copyright 2019 by the AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by the AAFS.