Part 2:- Quality assurance mechanisms for digital forensic investigations: knowledge sharing and the Capsule of Digital Evidence (CODE)

Graeme Horsman

Teesside University Middlesbrough Tees Valley TS1 3BX

Email: g.horsman@tees.ac.uk

Abstract
Despite potential numerous benefits, the field-wide sharing of knowledge in digital forensics is arguably still yet to be attained. Achieving this has attracted much practitioner and academic debate, yet solutions to two fundamental hurdles have yet to arguably be addressed; ‘how do we share knowledge’, and ‘what do we share’. Currently there a few viable protocols in place which tackle either of these issues forming a barrier to field-wide sharing. The focus of this work is to address the latter issue and guide practitioners on what content must be shared for any data to be of value to fellow professionals. This paper proposes the Capsule of Digital Evidence (CODE), a framework designed to set out the required elements for the sharing of reliable digital forensic knowledge. The CODE structure and its requisite contents are examined along with its applicability for supporting field-wide knowledge sharing in digital forensics.

Keywords: Digital Forensics; Quality Assurance; Investigation; Knowledge sharing; Digital Evidence

1 Introduction
The need for knowledge sharing in digital forensics (DF) has been mooted by academics, researchers and practitioners for over 15 years (see comments from Bruschi et al., 2004; Ruibin et al., 2005; Schatz and Clark, 2006; Biros et al., 2007; Kahvedžić and Kechadi, 2009; Huang et al., 2010; Tanner and Dampier, 2010; Ćosić and Čosić, 2012; Horsman et al., 2014; Casey et al., 2015; Weiser et al., 2016). As a discipline, those in DF recognise the need to share knowledge and the benefits that it can offer, which include uses such as training aids for practitioners (Karie and Venter, 2014; 2015) and for investigation quality assurance measures. With the subject of forensic analysis being that of technology, its rapid pace of change at both a software and hardware level means that no one individual DF practitioner is ever likely to possess sufficient understanding to tackle everything that can be faced in their role.

The sharing of DF knowledge offers one way to support field development, where the only option to understand all aspects of our digital society from a forensic perspective, is arguably via a collaborative effort. Given the diversity of activities and data types in DF there are a range of knowledge-types which can potentially be shared, where the ‘knowledge sharing’ becomes a vague term, without further description being applied. The question then remains - ‘what type of knowledge should be shared?’ with previous suggestions including past DF investigation findings (Horsman, 2014) and datasets for testing (Garfinkel et al., 2009). This article proposes CODE for the sharing of ‘new knowledge’, as discussed below.
World-wide, hundreds of digital investigations take place daily where it remains likely that in many of these cases a practitioner will encounter digital traces on a system which ‘they have not encountered before, and no documented or reliably documented interpretation of the digital trace exists. As noted in ‘Part 1:- Quality assurance mechanisms for digital forensic investigations: introducing the Verification of Digital Evidence (VODE) framework’, this is a scenario where through robust testing, the practitioner must generate their own interpretation of this data and then apply it to the data found in their case. The product of this work is coined ‘new knowledge’ which the field may not have formally encountered before and this is the type of knowledge CODE is designed to capture and share.

A practitioner who shares this ‘new knowledge’ is providing the following potential benefits.

1. **Practitioner benefit:** The sharing of newly discovered accurate knowledge by a practitioner (practitioner of knowledge origin - ‘P1’) provides a benefit to those who this content is shared with. This may occur on multiple levels. First, where another practitioner (practitioner in receipt of knowledge - ‘P2’) has encountered the same digital trace in a case, the shared ‘tested and validated interpretation’ of it supports P2 in their current case. If we assume that P1’s interpretation is free from error, the potential exists for P2 to apply this to their current case. Whilst methodological validation should still occur, P1’s interpretation provides support for P2 in their current case, preventing them retracing the same steps (Casey et al., 2013).

2. **Consistent application:** Again, assuming that P1’s shared interpretation of a digital trace is accurate, this should be regarded as a benchmark standard for interpreting future occurrences of this trace. By sharing this information the application of this meaning by other practitioners is encouraged when they encounter the same digital trace (and the interpretation is applicable given all the facts of the case). This helps to uphold quality standards in DF examinations, preventing divergent interpretations from occurring (where there is no need), arguably reducing the chances of reliance on any existing or created misinterpretation. In addition, where a certain digital trace can be consistently interpreted, this helps those who are outside of DF but still engage with digital evidence are part of legal processes (criminal justice system employees, law enforcement officers) to understand the meaning of this content.

3. **Quality assurance:** Sharing newly discovered knowledge allows peer review to occur. In an ideal world, any shared knowledge is robust, yet human error can still occur. Shared data and the associated test methodologies undertaken by P1 can be scrutinised by others in the DF field with suitable expertise (Grajeda et al., 2017). This can lead to three outputs:

   a. **Confirmation:** Other practitioners can confirm the accuracy of any provided interpretation.
b. *Dispute:* Other practitioners may question the accuracy of the provided interpretation and offer evidence which refutes either in whole or part of the interpretation.

c. *Additions:* Other practitioners may add additional content to the provided interpretation where the interpreted digital trace has since developed or been updated (in the case of software artefacts).

4. *Supporting a common goal:* Particularly in criminal capacities, those operating in DF are supporting a common goal in providing evidence to enable criminal justice systems to apply their applicable laws appropriately and reliably. Knowledge sharing helps to maintain and promote investigatory standards.

In addition, though not seen as field-wide benefits, practitioners should also consider that the creation and dissemination of 'new knowledge' is an activity of continued professional development, supporting the development of both themselves and the field.

Whilst there are multiple benefits to sharing knowledge, this is arguably yet to happen on a wide-scale consistent basis. Presently, there are barriers to knowledge sharing in existence.

1.1 Barriers to sharing
It has been suggested that those in DF are susceptible to a ‘silo mentality’ (Rogers and Seigfried, 2004; Horsman, 2018), where a reluctance to share information exists. The following are arguably obstacles which must be overcome before field-wide knowledge sharing is likely to occur.

1. *Engaging those in the field:* In most cases, the sharing of knowledge will be an activity which a practitioner will engage with in addition of their professional job role; in essence it is extra-curricular. This makes engagement with knowledge sharing a personal burden, with little or no monetary incentive, where in most instances, personal kudos may be the only motivation (Van Baar et al., 2014). As a result, there may be limited enthusiasm to engage. There are likely to be few, if any personal benefits which encourage engagement. Whilst some may consider it an ethical and moral obligation to share content, this cannot be an expected blanket-mentality. Further, individual circumstances may prevent those who even want to engage in sharing from doing so (a lack of time, resources or ability). As a result, there are few incentives to consistently engage those in the DF field, and whilst occasional or short bursts of engagement may occur, for maximum value from knowledge sharing to be obtained, it must be sustainable. Arguably where possible, knowledge-sharing should be incentivised to support engagement.

2. *The ‘silo mentality’:* As noted above, the siloed mentality is a barrier to sharing and this can occur on both an organisational and practitioner level. At an organisational level, knowledge can create monetary reward where the monopoly of certain forensic techniques can be seen. In such cases, it is difficult to argue that such approaches
should be swapped for the open, transparent sharing of data which would ultimately impact financial reward. At an individual practitioner level, harbouring expert knowledge may lead to the acquisition of an ‘expert status’ and an increased reputation in the field of DF or further job opportunities. Arguably this could be achieved through those who share knowledge also, but this position may not always be adopted.

3. A reluctance to share: Shared knowledge allows others to learn and develop through the use of this content. However, for the author of any knowledge, the fear and anxiety of public scrutiny may deter those from making a contributing (Garfinkel et al., 2009). As a result, it is important for a specific code of conduct to be put in place with any knowledge sharing protocol, prevent individuals being compromised in terms of their reputation of job or trolled. Disputed information must be dealt with formally and professionally.

4. How do we do it?: Assuming that wide-scale engagement could be obtained, there are few mechanisms to deliver this and this is a concern. There are two questions which must be addressed:-

a. How do we share it?: A suitable protocol for sharing knowledge is also required. A suitable vessel and format for shared knowledge must be developed which can be adopted by the field. Currently this is absent.

b. What do we share?: Determining what to share is a crucial aspect. Whilst sharing ‘new knowledge’ forms the crux, there are also elements to consider which allow other practitioners to trust and validate the work, including any testing methods undertaken or developed, how the knowledge was interpreted, who is responsible for it and is it reliable. The challenge of ‘what do we share?’ forms the focus of this work.

Securing buy-in from the DF field to contribute maintains its own challenges beyond the confines of this paper where developing a suitable protocol for sharing knowledge remains the focus. This work presents the concept of a Capsule of Digital Evidence (CODE), for sharing knowledge in DF where the CODE structure is present and discussed. Section 2 provides a discussion of CODE with Section 3 demonstrating how it may be implemented in practice. Finally concluding thoughts are offered.

2 The Capsule of Digital Evidence
The CODE structure is a concept based on the requirements needed for effective knowledge sharing in DF, where practitioners share knowledge-cases in individual ‘Capsules’ (analogous to a digital vessel). Every time a practitioner has discovered and validated ‘new knowledge’, CODE defines the requirements for them to capture and share this information in Capsule form with all the elements for this information to be reliably used by other practitioners. The CODE schema provides formalised guidance on the elements needed for shared knowledge to be of value to all practitioners.
The CODE schema is presented in Figure 1 and followed with a full description of its components.

CODE defines the requirement for three key categories of data descriptors; ‘Submission Metadata’, ‘Core Continuity Elements’ and ‘Core Digital Data Descriptors’ which must be present in each Capsule. All comprise of a series of sub-criteria (discussed in Sections 2.1-2.3) which help to demonstrate the reliability of any ‘new knowledge’ contained. At a high-level, any practitioner seeking to rely on a Capsule’s contents must be able to determine the following three points:

- **Continuity requirements for accountability:** All Capsules submitted must contain information for the purposes of maintaining a chain of continuity. This includes the details of the original author of the Capsule data and any other engagement with the Capsule from subsequent practitioners who may have used the information, added to it or refuted findings. This is addressed through CODEs ‘Submission Metadata’ and ‘Core Continuity Elements’ data descriptors.

- **Transparency of the work carried out:** The Capsule must contain a record of all processes involved in the production of the knowledge it contains. This includes details of testing, the test data used, any validation processes carried out and records of peer review carried out on the Capsule. This is addressed through CODEs ‘Core Digital Data Descriptors’ and ‘Core Continuity Elements’ data descriptors.

- **Main contribution of the Capsule:** The purpose of the Capsule is to share a practitioner’s interpretation of ‘new knowledge’ which they have discovered and decided to share. This must be clearly defined and contextualised. This is addressed through CODEs ‘Core Digital Data Descriptors’ data descriptor.

Each of the three categories of data descriptors are discussed in depth below.

2.1 Submission metadata

Each Capsule must be accompanied with submission metadata which describes the origin of the Capsule. This provides accountability for the Capsule’s content and allows for a dialog between the Capsule author and any subsequent users of the information, if required. The following three aspects must be submitted:

1. **Details of CODE submission:** A CODE must not be anonymously submitted to permit accountability for its contents. This is important for two reasons, first, to establish the origin of the capsule’s content and second, to ensure a channel of communication exists between the submitter and any subsequent user of the capsule’s data. This ensures that the original submitter can be contacted if issues with capsule contents are found. Submission metadata must include author details including contact information and
location, an overview of the contents of the Capsule and the date and time of the creation of the Capsule.

2. **Proof of engagement with both the Digital Evidence Reporting and Decision Support (DERDS) framework & the Verification of Digital Evidence (VODE) framework:** CODE is designed to house reliable knowledge generated through engagement with robust forensic methodologies designed to support accurate knowledge creation. Any capsule submitted must contain knowledge which has been developed through engagement with the VODE (discussed in Part 1) and DERDS frameworks which are designed for quality assurance.

3. **CODE Submission Reference:** Each Capsule is provided a unique reference number for the purposes of identification and archiving.

2.2 **Core digital data descriptors**

There are six core digital data descriptors which must be present in order to create a complete and reliable set of information needed for contents in the Capsule to be used by others.

1. **Disclosure of initial hypotheses and assumed impact on the case where the artefact/data originated from:** Ultimately, each Capsule will provide a practitioner’s interpretation of some form of digital trace, encountered as part of their case work. To understand why a practitioner has created a Capsule of this data, the Capsule must include details regarding the case circumstances (not case/suspect specific details, which may breach legal policy) in which it was found. This includes a description of what the specific artefact/data was suspected of being when initially found, and how its usage/function/presence was perceived and why. In addition, its impact on the case in which it originated should be highlighted (for example, it may have proved specific illicit user actions).

   a. A practitioner should also provide a sanitized (case specific/identifying content removed) version of the original data/artefact which has been interpreted. In addition, the practitioner should provide details regarding how the sanitisation process has been completed, so as to not compromise the value of the data in further testing/validation works.

2. **Record of artefact/data surrounding case circumstances and associated metadata:** The practitioner must disclose all circumstances which surrounded the artefact/data including case type, suspected offence, file paths, naming conventions, internal structure indicating file type and associated application and the software or service causing the presence of it on a system.

3. **Iterative testing complete with extracted artefact/data relating to each sub-test:** In order for others to rely on the interpretation provided in the Capsule, fully documented and transparent testing practices must be provided in order to show how any digital trace in
question was interpreted. This also includes the submission of the test data used in any of the tests undertaken and a description of the test actions carried out by the practitioner. As testing will be iterative, each artefact copy must also be disclosed with any associated ‘test actions’ documented to allow for the explanation of any modifications which might be present in the data. The practitioner should also demonstrate the repeatability of any results.

4. **Complete disclosure of test methodology followed and all the steps involved and undertaken:** The full design and implementation of the testing methodology undertaken should be disclosed, allowing for the transparent scrutiny and any methodological weaknesses to be identified in any future peer review of the work. The practitioner should make reference to the Framework for Reliable Experimental Design (FRED) for methodology design support (Horsman, 2018).

5. **Disclosure of any competing hypotheses found to have a potential impact on findings or limitations of the work:** Any competing finding which may limit the confidence of any interpretation should be noted and disclosed by the practitioner. In turn, any limitations of testing must be stated to allow future work to be undertaken.

6. **Final interpretation of findings:** One of the benefits of sharing this type of knowledge is that other DF practitioner benefit from others who have carried out robust work to understand how a particular digital artefact/data works and have shared this interpretation. Therefore disclosure of the final interpretation of findings and any developed iterative methodology used to parse and display findings is core to the Capsule.

### 2.3 Core continuity elements

The sharing of knowledge in DF must be a dynamic and sustainable process, allowing constant iterations and development. As suggested by Horsman (2019) digital artefact research has a specific lifespan where many of the digital artefacts themselves are subject to frequent structural change following software updates. As a result, data inside of each Capsule must be capable of being updated where additional information becomes available, with further validation records added by the practitioner who has undertaken this supplementary work. Such actions form the ‘core continuity elements’ of each Capsule.

1. **Chain of Iteration:** A chain of iteration is a list of all those who have carried out additional testing and development in regards to the data submitted in the original Capsule and furthered understanding of the artefact/data. Capsule content must be dynamic in order to retain its applicability when the data/artefact is updated, for example by a software vendor. However, this process of updating must be monitored, with each iteration of an interpretation of data clearly attributable to its author. Each iteration must also have all six core ‘digital data descriptors’ for any new testing which has taken place.
2. **Chain of Distribution**: A chain of distribution is a ledger of all those who have accessed and utilised the knowledge in the Capsule and how. This includes practitioner details and geographical location. Further, if any iteration of the data within the capsule occurs, those who engage with the capsule must note which iteration they used. This allows ‘knowledge-tracing’ to occur, which has two benefits. First, it helps to quantify how impactful each particular Capsule has been by determining the level of engagement with the DF field. And second, if a Capsule if later found to be compromised and its content is identified are incorrect, all those who have relied upon this knowledge can be traced and any further spread of this incorrect information can be prevented. To note, it should be considered that all levels of knowledge sharing should be treated with caution but seen as a beneficial.

3. **Chain of Validation**: A chain of validation is a list of all those who have validated the findings presented in the Capsule, as per the original submitter findings. This is a formalised record of peer-review, helping to ensure the accuracy of the information provided in the CODE model.

In an ideal world, Capsule data should be reliable if correctly created. However, this is not always the reality, and therefore the core continuity elements are mechanisms designed to validate shared knowledge and to prevent the sharing of bad practices and inaccurate content.

### 3 How to create and share Capsules, and when to submit

Whilst Section 2 has defined what a Capsule must contain, the next issues concern how to create a Capsule, when to submit a Capsule, and how to share it.

#### 3.1 Creating a Capsule

Whilst the contents of a Capsule is described in Section 2, gathering and housing this data in a ‘Capsule format’ provides a challenge which must be addressed for CODE to be operational. The following format difficulties must be considered.

1. **Collecting content**: A practitioner must extract data from their case and testing processed in order to populate a Capsule. This work may be undertaken on a range of tools, each with a different functionality and data export format. Further, some tools may not provide a suitable export procedure which is easily captured for the purposes of creating a Capsule. The issue this presents lies with the burden falling onto the practitioner to manually populate their Capsule, a likely burdensome task, which may deter practitioners from participating with CODE. As a result, one of the requirements for CODE to become fully operational it the need for a ‘plugin’ styled process which can quickly and efficiently automate the capture of requisite information for a Capsule, which may originate from any one (or multiple) of the many tools currently in use by practitioners world-wide.

2. **A consistent, queryable format**: A Capsule must be consistent in structure. This allows the automation of Capsule content validation to prevent incomplete Capsules being
submitted to the project, and to allow the mass querying of Capsule content for analytical, and search and retrieval purposes.

Capsule creation remains the next stage of the CODE project and whilst the contribution of this work is to outline what CODE must contain, future work involves the design and implementation of it. Engagement and consultation with relevant DF communities is required to establish the requirements needed for CODE to function within the industry laboratory environment.

3.1 When to submit a Capsule
Determining when the submission of a Capsule is appropriate depends on a number of factors outlined in Figure 2. The first stage occurs as the practitioner is undertaking case work as part of their main employed role as a DF practitioner (submissions would also be encouraged by researchers and academics where appropriate). When undertaking an investigation, if a practitioner encounters a digital trace on a system which ‘they have not encountered before, and no documented or reliably documented interpretation of the digital trace exists’, and they have had to interpret this data themselves, then a Capsule should be created to capture this interpretation. Providing the interpretation has been generated following robust testing, then a practitioner can share this interpretation in a Capsule. Second, a practitioner, as part of their case work, may encounter a known digital trace and verify an existing interpretation via robust testing. This work is also of value, where the practitioner has two options with how to capture this information. If a Capsule already exists for the known artefact, they can update the ‘core continuity elements’ of that Capsule to reflect that they have engaged with this knowledge and validated/refuted it. If a Capsule does not yet exist, then the practitioners work can be placed within a Capsule.

3.2 How to share a Capsule
The sharing of knowledge, regardless of form should be seen as beneficial in DF, however two main options exist with CODE (see Figure 3).

Internal sharing: Internal sharing takes place when an organisation chooses to store practitioner Capsules and make them available for their practitioners to utilise. Capsule information is not shared beyond the remit of the organisation. Whilst this option does not offer field-wide knowledge sharing benefits, it is currently the only logistically viable option. Management and storage of Capsules can be controlled by the organisation internally.

Field-wide sharing: Field-wide sharing is arguably the goal of any knowledge-sharing schema, with the potential for maximum benefit to the DF field. However, until like internal sharing where an organisation can take responsibility for governing Capsule submissions, there are currently few equivalent structures in place who could take this role in a field-wide capacity (Weiser et al., 2016). This approach requires a governing body to manage Capsules, storing and keeping track of submissions and those who interact with them. One solution is to host such a project online (subject to data protection and associated legal concerns, similar to what is seen with the
‘Artifact Genome Project (Grajeda et al., 2018)) where its function should aim to be autonomous, lessening the burden on the management of its data (Buang and Daud, 2012).

4 Conclusion
The Capsule of Digital Evidence schema has been offered as a means of facilitating the sharing of ‘new knowledge’ in the DF field. The requirements for effective sharing have been outlined, and the three categories of data descriptors; ‘Submission Metadata’, ‘Core Continuity Elements’ and ‘Core Digital Data Descriptors’ have been outlined in detail. The CODE project and Capsule structure has been discussed, highlighting the requirements needed for shared knowledge to be reliably utilised by others within the DF. Capsules are designed to be a quality assurance mechanism for digital forensic investigations, supporting those who undertake case work by providing access to reliable information.

There are no conflicts. I am the section editor of the digital forensic theme.
Reference


Weiser, Mark; Biros, David P.; and Mosier, Greg, "Development of a National Repository of Digital Forensic Intelligence" (2016). Annual ADFSL Conference on Digital Forensics, Security and Law. 2. https://commons.erau.edu/adfsl/2006/session-i/2
Figure 1: The CODE framework (a full quality image has been submitted as a separate file due to size.)

Figure 2: Decisions to submit to CODE.
Figure 3: Options for knowledge sharing with CODE.