

TECHNICAL NOTE

Digital & Multimedia Sciences

A template for creating and sharing ground truth data in digital forensics

Graeme Horsman PhD

Cranfield University, Bedford, UK

Correspondence

Graeme Horsman, Cranfield University,
College Road, Cranfield, Wharley End,
Bedford MK43 0AL, UK.

Email: graeme.horsman@cranfield.ac.uk

Abstract

Ground truth data (GTD) is used by those in the field of digital forensics (DF) for a variety of purposes including to evaluate the functionality of undocumented, new, or emerging technology and services and the digital traces left behind following their usage. Most accepted and reliable trace interpretations must be derived from an examination of relevant GTD, yet despite the importance of it to the DF community, there is little formal guidance available for supporting those who create it, to do so in a way that ensures any data is of good quality, reliable, and therefore usable. In an attempt to address this issue, this work proposes a minimum standard of documentation that must accompany the production of any GTD, particularly when it is intended for use in the process of discovering new knowledge, proposing original interpretations of a digital trace, or determining the functionality of any technology or service. A template structure is discussed and provided in Appendix S1 which sets out a minimum standard for metadata describing any GTD's production process and content. It is suggested that such an approach can support the maintenance of trust in any GTD and improve the shareability of it.

KEYWORDS

datasets, digital evidence, digital forensics, ground truth data, testing

Highlights

- The role and importance of ground truth data (GTD) in the field of digital forensics are discussed.
- GTD for the purpose of evaluating undocumented or previously unseen digital traces is focused upon.
- A template structure to support the creation of GTD is offered and explained.
- The GTD template defines a minimum standard for describing any GTD's production process and content.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Journal of Forensic Sciences* published by Wiley Periodicals LLC on behalf of American Academy of Forensic Sciences.

1 | INTRODUCTION

Ground truth data (GTD) is data that represent the truth for a given state or scenario [1]. This type of data is gathered from direct observations rather than generated from inferences [2] and can help us to understand how something works and the impact and characteristics of any given event. In the field of digital forensics (DF), one of the pivotal roles that GTD plays is helping practitioners and researchers operating within it to evaluate the functionality of undocumented, new, or emerging technology and services and the digital traces left behind following their usage [3]. When previously unseen instances of tech usage are encountered in casework or such occurrences are hypothesized by those who are vigilant of the developing “tech” landscape, then exploratory work must be conducted in order to ascertain the presence and meaning of any digital traces generated as a result of these actions [4]. This process is particularly important if it is considered that establishing an understanding of these acts is likely to impact the casework of practitioners in the field.

When those in DF seek to determine how any tech service or function works, they often do so by testing it under controlled conditions (effectively “using it”), allowing them to observe and monitor how it operates [5]. This is typically achieved using a “black box” approach where test actions are carefully chosen to elicit appropriate responses from any functions or services that a practitioner intends to test [6, 7]. The data that is then generated is acquired using forensic methods, and attempts are made to then identify and interpret the presence of any digital traces created as a result of these actions [8]. In essence, these procedures try to establish whether any activity on a device or interactions with a service leave traces that describe a user's course of digital conduct when using any technology or service. For practitioners to begin to reliably understand any trace, they must examine their presence and the surrounding activity that caused its creation, in conditions that have been designed to produce data considered to describe the ground truth for that particular investigative scenario.

Accessing any technology's source code or underlying algorithms in an attempt to determine its functionality may not be possible due to legal restrictions or code-protective measures [9]. Even in cases where it is, there is a significant challenge involved with the task of reverse engineering and understanding the code, requiring both specialist knowledge and tools. As a result, in most cases, reliably understanding the functionality and impact (digital traces created by this use) of any form of technology or service requires using it under ground-truthing conditions, and capturing data resulting from this usage. Arguably almost all newly acquired knowledge of digital traces in DF and any subsequent interpretations of them must be derived from properly generated and appropriate GTD that depicts the conditions that led to the creation of the trace. Most accepted and reliable trace interpretations must be derived from an examination of relevant GTD (although “good GTD” does not prevent trace misinterpretation, the interpretative process is a separate issue and beyond the scope of this work) and the field of DF can only further its knowledge and understanding of the technological landscape via

the use of appropriate GTD. However, producing it is difficult. It is questionable as to whether any proposal of a trace's meaning or description of a device/service's functionality arising from an evaluation of GTD that does not appropriately provide the ground truth for these tasks can be relied upon. Given this, the importance of GTD for DF cannot be understated.

It should be recognized that generating data is easy, but creating data that describe the ground truth for a given scenario can be difficult [10]. GTD must be distinguished from standard “datasets”; its creation must be robustly planned, any activities methodically conducted and the whole process be rigorously documented [11, 12]. As the OSAC Digital Evidence Subcommittee Task Group on Dataset Development states, “a well-documented dataset facilitates more rigorous testing and reliable results” [11]. Any ambiguity in regard to the production process and content of any GTD should be kept to a minimum (or ideally removed, however, in reality, this may not be feasible), where emphasis is placed upon the need to describe all actions and the surrounding circumstances and scenarios leading to data creation accurately [12]. It is stressed above that GTD must be “properly generated and appropriate,” and whether data can be considered ground truth depends upon the quality of its construction and the level of detail in any documentation that accompanies it. In contrast, standard datasets (non-GTD) may not have accompanying documentation that describes the processes that were used to create it, and its content in a level detail that allows a third party to evaluate their contents.

Despite the importance of GTD to the DF community, there is little formal guidance available for supporting those who need or want to create it within the DF field, to do so in a way that ensures any data is of good quality, reliable, and therefore usable [11]. The OSAC Digital Evidence Subcommittee Task Group on Dataset Development [11] provides one of only a few examples of supporting documentation. In reality, many of the datasets that are publicly available are unlikely to be GTD for any scenario they describe; something that is not troublesome unless any dataset creator is claiming that their data is ground-truthing, but in reality is not. In these instances, any subsequent trace interpretations resulting from these data may be unsafe, and if incorrect conclusions have been drawn, their impact upon all those involved in a case could be significant. For this reason, current levels of trust in shared/distributed “GTD” may be low.

While alluded to above, creating GTD is a challenging task in itself but not the only one that requires attention; properly created GTD can also be beneficial to the whole DF community if it is shared. Distributed GTD provides an opportunity for others to discover knowledge that may have been missed during any original analysis of the data, and perhaps more importantly, it allows others to evaluate the data and any trace interpretations that have been proposed as a result of it—“peer scrutiny” [13, 14]. Yet, there are practical reasons (privacy and legal issues aside) as to why the use of shared GTD is somewhat limited and problematic; those receiving it must be able to trust it in order to be able to use it, and those sharing it must be able to demonstrate that it is trustworthy. These are hurdles that have proven difficult to overcome, and currently, with limited GTD

production standards and methods, there may be little progress in this space.

This work suggests a minimum standard of documentation to accompany the production of GTD when used in the process of discovering new knowledge, proposing original interpretations of a digital trace, or determining the functionality of any technology or service. A template structure (GTD template.xlsx) is provided in Appendix S1 which sets out a minimum standard for metadata describing any GTD's production process and content. It is suggested that such an approach can support the maintenance of trust in any GTD and improve its shareability.

2 | KNOWLEDGE DISCOVERY AND GTD

GTD for DF is described by Göbel et al. [8] as “data whose content is well known and understood by the community, i.e., the exact information or the actual digital traces to be discovered in a data set during the investigation are well documented.” While Woodhouse [1] suggests the origins of the term GTD are unclear, it is often linked with both military and cartography domains, and the field of DF has recently sought to adopt this term to describe data that has been methodically created and its content known and documented. The field of DF concerns itself with the examination of “digital traces” and those within it seek to determine the value and meaning of any that they consider are potentially relevant during the course of an investigation [15]. We have already noted that the production of GTD is pivotal when attempting to understand the meaning of any digital traces, and there are typically two main contexts where GTD for this purpose is produced. Practitioners will produce GTD as part of their working roles when there is a need to interpret any traces encountered as part of their casework that they do not understand—referred to here as “in-practice” GTD. In addition, practitioners and/or researchers in DF may create GTD with the intention of disseminating it to the wider community so that others can engage with it and evaluate its content—“community GTD.” Each is discussed in turn.

2.1 | “In-practice” GTD

While discussions surrounding GTD in DF are typically focused on its use for method/tool evaluation [16–18], its role in the context of interpreting traces that have not yet been described and documented is often overlooked. Practitioners may encounter digital traces that they do not yet understand and no existing and accepted meaning is available, yet they may initially believe that they could be of relevance to their current investigation—“an investigative hunch.” In these cases, in order to fully understand the impact, meaning, and value of any trace, they must interrogate any technology and/or services that they believe to be responsible for them. Using such technology/services under documented conditions and in a controlled environment allows practitioners to observe the creation of any relevant traces resulting from any test actions, and evaluate their

meaning. DF practitioners are constantly required to engage in this type of exploratory research as the technology landscape changes at considerable speed, often impacting the structure of existing traces as well as generating new trace types which must be inspected.

When practitioners need to establish the meaning of any trace that is currently undescribed (or no agreed or reliable understanding exists), this is what this work considers a “new knowledge situation” (NKS). NKSs occur when there is a requirement to establish the meaning of a trace for the first time or where there are questions over the reliability of any existing trace interpretation. For a practitioner to begin to understand any digital trace, they must be able to observe all of the factors that are involved in its creation and modification under controlled conditions.

While some practitioners may not formally recognize this fact, they should be creating GTD every time they encounter an NKS in their casework. Good GTD does not prevent a practitioner from misinterpreting traces, yet reliable inferences in regards to any given trace cannot be drawn without it, and GTD production should be a formalized part of the investigative process in these circumstances. This is particularly important for the purposes of quality control and assurance, as practitioners should be able to demonstrate the validity of any methods used, the reasons why they have reached such conclusions and the details of any tests/experimentations and data used to form any interpretation [19]. Any GTD used should be kept and made available for scrutiny and act as evidence of how any trace's proposed meaning was formed. If a practitioner cannot provide the GTD upon which their testing was performed, then their practices may be called into question. In addition, poorly produced GTD may also undermine any findings offered by a practitioner when the data are scrutinized by any authorized third party. It is within the best interests of all practitioners when interpreting new traces to produce GTD that is to the highest standard as this data should be considered evidence of good and safe interpretive practice.

2.2 | Community GTD

While in-practice GTD is an important aspect of casework conducted by operational practitioners, community GTD is often produced by researchers and “good Samaritans” who seek to make a contribution to the field. The DF community has long since recognized value in the provision of datasets for learning and development purposes [20, 21]; however, it is suggested that there is a need to scrutinize the quality of their production, if their “value” is to extend beyond merely providing an opportunity for others to curiously and informally observe data structures. It is generally accepted that while many of those in the fieldwork for different organizations and in different roles (some as competitors), there is a shared consensus that we can generally learn from one another [15]. As different levels and types of expertise exist, and there are those who have access to technologies that others may not, advances in knowledge are frequently made from a diverse range of sources and occasionally in isolation. In some cases, we see submissions in the form of datasets

derived from research that has been conducted, often in return for kudos and to be considered valued members of the community.

The sharing of knowledge should be encouraged as all practitioners and organizations in this space maintain a common goal of supporting all those involved in the investigation of crime [22]. Community GTD is often intended to be available for all those in the field of DF. It may give access to data types and structures that some practitioners may never have a chance to see, or in some cases, never have the chance to produce their own GTD as they do not have access to any original technology from which to create it. Fundamentally, good community GTD may simply provide an additional source of data that can be used to validate information and support organizational practices. Yet conversely, shared datasets that fall short of ground truthing any scenario that it claims to describe offer little to no value to others, or cause harm if they are wrongly claiming to be GTD.

This work does not intend to deter those who seek to contribute datasets for the benefit of others; however, the value of any dataset that does not describe the ground truth for whatever scenario it intends to provide coverage of is of questionable worth to others. While this may seem a controversial statement to make, if a dataset has not been properly created and described, then those seeking to use it are unable to determine the reliability of any interpretation of traces they gain from it. This does not preclude any casual use of it, but practitioners cannot use it to bolster or validate any information or practices that are going to be relied upon by criminal justice systems.

2.3 | Sharing GTD

Motivations behind sharing GTD may differ depending on the context of any GTD's production. In-practice GTD may typically remain internal to a practitioner's organization, however, there are benefits to sharing it. One of the most powerful techniques for quality assurance in the DF field is peer review, and where GTD has been used to form an opinion of a trace, releasing that GTD to others so that they can evaluate any proposed interpretation can be beneficial.

In contrast, the purpose of community GTD is to provide others with access to data. Those creating community GTD may do so in one of two ways:

1. *"Seed & leave"*: "Seed & leave" refers to the scenario where a GTD creator generates a scenario regarding a piece of technology and the functions they intend to test, invokes them using appropriate test actions, and leaves the task of identifying and interpreting any subsequently generated traces to someone else. In essence, they invoke any software/function in a way that is designed to seed appropriate data of this activity and leave it to others to identify and determine the meaning of any resulting traces. When a seed-and-leave approach is taken, accompanying documentation describing any test actions must be thorough.
2. *"Seed & interpret"*: Unlike a seed-and-leave activity, an individual may seed data, then go on to identify the presence of, and propose an interpretation of the meaning of any generated traces—analogueous to in-practice GTD. In these cases, GTD can be shared in order to allow others to evaluate any proposed interpretation of a trace

2.4 | Problem summary

When producing GTD for the purposes of discovering new knowledge, this work considers two distinct problem areas.

1. *Quality issues*: When we interpret traces, there may not be a formal acknowledgment of the need to ensure that any data we are basing our observations on needs to be ground truthing. In addition, practitioners may not recognize when data are insufficient in quality and therefore not "ground truthing." There is a risk that when testing, any data generated are created using ad hoc approaches and standards, and lacks rigor in the methodological approaches and documentation of it. Simply put, there is a worry that any generated data does not reach the standard of ground truth.
It is important to recognize that any trace's interpretation must originate from data that have had its production controlled to reflect the conditions in which any trace exists. Currently, there are no standards or benchmarks that must be reached during the production of GTD, and to guide the creator of it. GTD is therefore likely created on a case-by-case basis and without guidance could differ in quality between creators.
2. *Trust*: One of the main problems with sharing GTD is trust (where legality and data protection issues are not considered in the remit of this work). Those who are given GTD or want to use GTD that have been produced by another, must have confidence in its quality, that it has been robustly and accurately produced, and that any interpretations derived from it can be relied upon. This involves being able to trust that the data generation process has been done correctly, data seeding has been done correctly, and any accompanying documentation is accurate and in enough detail to allow it to be used. Any recipients of GTD cannot assume that it is of an acceptable standard, GTD creators must be able to provide evidence of the quality of their data

Platforms such as the Computer Forensic Reference DataSet Portal [23] (CFReDS) and the Digital Corpora [24] house datasets that are available for use by those in the DF field. Such initiatives should be welcomed and encouraged, but caution should be exercised by those seeking to rely on some of the datasets these platforms contain when trying to reliably establish the use of technologies and services and the digital traces they subsequently generate and to test forensic software. While some datasets may maintain sufficient documentation that describes both their production processes and the contents of the dataset, allowing them

to potentially be considered GTD, others may not. CFReDS states that their “datasets can assist in a variety of tasks including tool testing, developing familiarity with tool behaviour for given tasks, general practitioner training and other unforeseen uses that the user of the datasets can devise.” Emphasis should be placed on “can assist” where the platform does not claim to be a source of GTD and it is up to contributors to decide whether they intend to produce and submit GTD, and it is their responsibility to make sure their data reaches this standard—both CFReDS and the Digital Corpora do not guarantee the quality of the data. CFReDS also acknowledges that “most datasets have a description of the type and locations of significant artifacts present in the dataset,” suggesting in some cases such documentation may be missing. Even where datasets maintain accompanying documentation, the standard of this documentation should be scrutinized as in some cases, accompanying documentation does not describe the processes involved with seeding data, or describe the seeded data itself in enough detail (e.g., relevant metadata describing the quantity of seeded data types, the data itself, and the location of it (path or physical offset) may be lacking detail).

The OSAC Digital Evidence Subcommittee Task Group on Dataset Development [11] provides one of the only guidance documents for the construction of robust and shareable datasets. While they acknowledge that they do not cover all scenarios, they encourage use of their guidance “to improve consistency and quality among datasets used in digital forensics” and it should be applauded for taking on the task of supporting dataset creators. The author of this work intends this piece to be complementary to that of OSAC and support the goal of increasing the quality of dataset generation in DF by also offering a structure to harmonize the dataset generation process. OSAC provides detailed information discussing the creation of datasets and the intricacies involved in data seeding, and in addition, provides a template document to record such actions which this work considers a base from which to build. OSAC rightly recognizes the importance of recording the metadata of any systems used within their datasets and details of any acquisition methods and verification processes used. There is also provision to describe any data seeding activities.

In an attempt to support this work, we emphasize the need to attribute datasets to their creators and to understand their developmental process, therefore dataset authorship information is important so that future dataset users can contact creators and liaise with them should questions arise. While this work also notes the importance of describing any apparent limitations of a given dataset, we believe it is necessary to define the scope of use of any given dataset along with its specific purpose to ensure it is not misused. In addition, as noted in Section 3.3, it is important to provide documentation that describes the interactions of future users of the data, particularly if they have conducted tests that have yielded important results that should be shared with the community who may seek to interact with the data. It is also important that any actions relating to the seeding process are described in detail, and this work opts to break down the description of data seeding actions to include

details of activity, seeded data, trace location, and trace considered. While this may all be considered under what OSAC terms “Action” and “Content/Details/Location” for a given data seeding event, it is suggested that it is important to distinguish each information type to avoid their omission during the documentation of tests. Finally, we also believe there is value in allowing the structure of any seeded traces within the dataset to be described, if necessary, to support future users when conducting and evaluating their tests.

It is suggested that to improve practices in regard to GTD production and to facilitate the sharing of it, guidance is required to support those producing it. This work offers the following template.

3 | A TEMPLATE FOR GTD

Motivations underpinning the proposed use of a GTD template lie with the opportunity it provides to define a benchmark in regard to the information that any GTD must maintain that describes its contents and production processes and to standardize accompanying documentation. Currently, GTD is produced and disseminated in a variety of ways, and any of the accompanying information that is disseminated with it is produced in a range of formats and is generated to a standard that the creator considers acceptable. This means that some datasets will fall short of requirements needed for GTD, and, some will meet this standard in terms of the data generated, but be unable to evidence that it is in fact GTD via their documentation.

It is suggested that a template provides a formalized threshold for GTD creators to ensure that they can provide the necessary metadata required to describe their data in a way that makes it usable by others. In addition, it can guide wannabe GTD creators by ensuring they work in accordance with the template during the developmental process and informing them of the need to ensure that such metadata are available and required.

The proposed template is split into two main sections; “Dataset metadata” and “GTD schedule of activities”—each described below. The complete template is provided in Appendix S1 as file “GTD template.xlsx.”

3.1 | Dataset metadata

It is suggested that every GTD dataset should be accompanied by a set of information that describes the construction and purpose of it. This is considered the dataset’s “metadata” and helps users to understand more about the creators of it, how it was produced (and any limitations of it), and the purpose and scope of the GTD. The following information fields are suggested as a minimum requirement for accompanying GTD.

1. **Name:** The name of the producer of the dataset should be recorded—it is arguably difficult to trust data that is provided anonymously, particularly if it is to be used in circumstances where any resulting information may need to be relied upon

- by a court of law. A producer may be an individual, group, or organization. All those actively involved in the production of the data should be listed unless a principal producer has had oversight and responsibility for all data-generating actions and can vouch for the accuracy of these processes.
2. *Organization*: The dataset producer's organization should be listed. Alternatively, if the producer is an independent researcher (or acting in this capacity), this should be stated. GTD should be attributable to an owner.
 3. *Contact Email*: It must be possible to contact those who have responsibility for the production of the dataset in order to raise queries and engage in narratives specific to the dataset content. This is particularly important if concerns are identified with the data and this needs to be communicated to prevent the further dissemination of it (or at least to place caveats on it).
 4. *Details of hardware used to create GTD*: Information regarding the hardware used as part of the GTD production should be stated. This includes device makes, models, and relevant identification information.
 5. *Operating system details*: The name of any operating system and version installed on any hardware used should be listed.
 6. *Date and time GTD creation process started*: As per.
 7. *Date and time GTD creation process ended*: Listing both the start and end dates and times of the GTD creation process allows others to understand the extent of any documented activities. It also provides a reference point in regards to what versions of any software and hardware were available at the time of the dataset being produced.
 8. *Name of dataset*: The dataset should have a uniquely identifiable name that is embedded within any image files (to prevent tampering) and attributable to it for its lifespan, allowing it to be referenced in any future work where results have been derived from it.
 9. *Format of dataset*: Any dataset may be captured in a number of ways (e.g., DD, Expert Witness Format (.E01)) and this work seeks not to define a specific capture format requirement. However, whatever the format of the dataset, it should be stated. This allows any user of the dataset to understand any limitations that might be present due to the capture method deployed, and what tools they need to access the dataset.
 10. *Dataset hash type*: Any dataset should be hashed and the type of hash used be recorded.
 11. *Dataset hash value*: The hash value of the dataset should be recorded, and users can utilize this value on receipt of the dataset to ensure its integrity.
 12. *Dataset protection details*: A dataset creator may seek to protect the contents of their dataset via encryption. Information that describes how to gain access to the dataset should be included if it is required, which may include contact being made with the dataset creator for further instruction.
 13. *Method of dataset acquisition*: Any specific hardware used to acquire the dataset (including bespoke methods) must be described and documented.
 14. *Tool used for dataset acquisition*: The name of any tool(s) used for data acquisition should be stated.
 15. *Tool version used for dataset acquisition*: The version number of any tool(s) used for data acquisition should be stated.
 16. *Name of acquisition method*: The acquisition method name should be stated. Some acquisition method names are bespoke to specific vendors and establishing the remit of the acquisition method may require consulting vendor documentation.
 17. *Date and time of acquisition*: The date and time that any dataset was acquired should be stated.
 18. *Purpose of dataset*: All GTD is produced for a specific purpose (it can never be all encompassing), which could include testing a specific function or application, or understanding how a specific artifact behaves under certain conditions. It is important to outline what the purpose of a dataset is, as its construction should have focused on meeting the conditions needed for it. As a result, the GTD may not be fit for any uses that sit outside of its defined purpose, and when deployed for reasons beyond this, any information or results derived may not be reliable. Any recipient of the GTD must understand how it is intended to be safely used.
 19. *Scope of dataset*: While a dataset should have a defined purpose, the scope of this purpose should also be described. For example, a dataset may have a single purpose of testing the function of Chrome's Internet history service. However, the scope of the dataset may be limited to specific versions of Chrome and specific actions attributed to Internet history creation (e.g., only when URLs are typed into the address bar and a visit is initiated).
 20. *Any known limitations/anomalies*: Dataset creators should state any known limitations or anomalies that they are aware of in regard to the dataset that may impact any user of it. As the dataset is disseminated, any subsequent reports of issues should be recorded here and accompany any further distribution of the data.

See [Table 1](#) for the "Dataset Metadata" table with example data.

3.2 | GTD schedule of activities

Most GTD in DF will contain the results originating from a series of test actions conducted in a controlled environment. While the 20 dataset metadata fields described in [Section 4.1](#) define the setup and constraints of the dataset, any GTD should also maintain detailed records of the schedule of activities conducted. A "schedule of activities" is defined as records that describe how data have been seeded within the dataset (those activities conducted by the creator), and information about any subsequently generated traces. The following information fields are suggested as a minimum requirement for describing a schedule of activities.

1. *Trace ID*: When seeding data in a dataset, creators should consider segregation activities into manageable and defined

TABLE 1 Dataset metadata.

Name	Mary smith
Contact Email	m.s@gmail.com
Organization	Forensic Inc.
Details of hardware used to create GTD	Dell Inspiron 242
Operating system details	Windows 11 Home
Date and time of GTD creation process started	15-07-2023 15:00
Date and time of GTD creation process ended	16-07-2023 15:00
Name of dataset	GTD2023.E01
Dataset hash type	SHA256
Dataset hash value	c09fe13f50efa40e595962cad6c609f5f12a9f31b10297c9a5b1b964caabb09e3
Dataset protection details	Email author for password
Format of dataset	Expert Witness Format (.E01)
Method of dataset acquisition	Standard Disk Image using Tableau v 1234
Name of acquisition method	N/A (consider bespoke tool vendor names)
Tool used for dataset acquisition	FTKi
Tool version used for dataset acquisition	23.123
Date and time of acquisition	23.123416/07/2023 15:10:00
Purpose of dataset	GTD for Chrome Internet history
Scope of dataset	To establish how Internet history is generated and stored by the Chrome (v70.1.2.3) browser application when URLs are typed into the address bar and a visit is initiated
Any known limitations/anomalies	None

test cases. This helps to prevent test contamination and also makes it easier to attribute any identified traces to their test actions, arguably reducing the risk of their misinterpretation.

2. *Filesystem Info*: The file system present on any storage media should be recorded.
3. *Service/Software*: The name of any software or service being tested (and therefore the subject of dataset) should be recorded.
4. *Version Info*: The version of any software or service being tested (and therefore the subject of dataset) should be recorded.
5. *Interaction Type*: Details of any activity conducted should be recorded. This may include actions such as "file 'X' executed," "message 'M' sent," or "link 'L' clicked." It should be emphasized that any interaction should be specifically described. For example, the exact number or order of any button presses should be quantified, the exact series of any engagement with menus should be described, and the names of any service "labels" engaged with should be listed. Any third party should be able to take the details describing any interaction type and recreate the interaction type.
6. *Input Data*: Where an action requires the use of specific criteria or data (e.g., the sending of test messages on a chat platform), any test data used should be described in full. Any seed data must be unique and attributable to each test case.
7. *Real Time of Activity*: The real-world time and date of any activities conducted should be recorded. If any activity occurs for a prolonged period of time, consideration should be given as to whether it is necessary to record both the start and end date of any given activity.

At this point, the aforementioned seven information fields describing a schedule of activities are considered a minimum requirement for community GTD where a seed-and-leave approach is adopted. This information describes to any users of the dataset what services/functions have been used and how to provide information for them to search for the presence of any potential traces created by these actions.

For in-practice datasets or community GTD adopting a seed-and-interpret approach, additional information regarding the impact of test actions and any traces created is required. These types of GTD go beyond simply describing test actions and provide additional details that outline any resulting traces.

The following additional information fields are suggested.

8. *Primary Trace Considered*: Any single action conducted on a device may result in the creation of multiple digital traces. Where a GTD creator has gone on to describe the results of any actions they have conducted, they should be clear as to the primary trace they have focused upon, and which traces may be beyond their documented considerations. For example, if a test case involves establishing what happens when a specific website has been visited, they may define that their test has only considered the impact on "Internet history records" traces, rather than also considering cached behaviors (which they may consider secondary traces generated as a result of any test interactions conducted).
9. *Logical Location of Trace*: The file path to any traces created as a result of any test actions should be recorded.

NaN. *Embedded Location of Trace (Optional)*: Any generated trace may be embedded within a specific file structure—for example, a specific table within a database. Where this is the case, relevant details should be stated.

NaN. *Physical Location of Trace (Optional)*: Where it may not be feasible to reference a trace using file path references, the physical location of a trace should be stated (e.g., if a trace for a test case exists in unallocated space).

NaN. *Trace Structural Information*: Where a trace is stored within a format that requires additional interpretative considerations, any relevant instructions should be provided (e.g., if a trace is stored within an encoded format, decode instructions should be provided).

The schedule of activity information fields 9–13 describes the presence of any traces within the dataset that are a result of any actions undertaken.

See Table 2 for the “GTD schedule of activities” table with example data.

3.3 | Additional—Validation tests

The aforementioned “dataset metadata” and “GTD schedule of activities” sections of the proposed GTD template concern information argued as being required in the production of any GTD. It is suggested that a separate third element to any GTD documentation may also be considered; a “validation tests” section. When any GTD is used by someone other than its creator, particularly if they are conducting tests upon it or evaluating its contents, it is argued that the maintenance of a log showing how they have engaged with the data, how this engagement has been conducted, and any subsequent results of any tests conducted would be of benefit to all future users of the GTD.

Particular reference is drawn to any engagement with the data that validate whether any seeded data described by the creator have generated traces and these have been confirmed as being present and accurate. For example, any creator of GTD can describe a specific “interaction type” deployed in their GTD but fail to identify correctly whether any trace is generated. Subsequent users of the data may then explore such an interaction and find a relevant trace. Alternatively, any GTD creator may erroneously describe a trace within their GTD that subsequently cannot be found by others. Users of any GTD should consider documenting any relevant tests they have conducted and the outcomes, and ensure these findings accompany the dataset in any future dissemination of it.

The following “validation test” fields are suggested.

1. *Test ID*: Reference should be made to the “schedule of activities” and which specific seeded activity is being tested.
2. *Result ID*: A results ID should be given so that any outcomes of the test can be referenced.
3. *Test Conductor Name*: The name of those conducting the validation test.

TABLE 2 GTD schedule of activities.

Trace ID	Dataset	Hardware info	Operating system info	Filesystem info	Service/software	Version info	Interaction type	Input data	Real time of activity	Primary trace considered	Logical location of trace	Embedded location of trace (optional)	Physical location of trace (optional)	Trace structural information
ID1	GTD2023. E01	Dell Inspiron 242	Windows 11	NTFS	Chrome	v70.1.2.3	Visit to website	www.google.com	30-12-2023 13:07:00	Internet History Record	C:\Users ...	"Urls" table records		
ID2	GTD2023. E01	iPhone 12	iOS16.6	APFS	Messages	v 12.00	Message Sent	"Hello, Test2023"	06-12-2023 17:34:00	Sent Message	'/private/var/mobile/...	'Messages' table records		

TABLE 3 Validation tests.

Test ID	Dataset	Result ID	Test conductor name	Test conductor contact details	Date and time test conducted	Was trace identified in full?	Was trace manually identified?	Name of tool used	Tool version	Tool configuration details	Other observations
ID1	GTD2023.E01	RID1	Mary Smith	111@111.com	30-12-2023 13:07:00	Yes	No	ChromeCacheView	2.46	No configurable options are available	
ID2	GTD2023.E01	RID2	Terry Jones	222@222.com	06-12-2023 17:34:00	Yes	No	Cellebrite	19.123	Option X & Y selected	
ID3	GTD2023.E01	RID3	Matt Low	333@333.com	23-12-2023 12:09:01	Yes	Yes	N/A	N/A	N/A	

4. *Test Conductor Contact Details*: The contact details of those conducting the validation test.
5. *Date and Time Test Conducted*: As per
6. *Was Trace Identified in full?*: In essence, any user of the GTD is attempting to establish whether any trace generated as a result of data seeding actions is present within the data. A dataset creator may claim the presence of a trace (a trace confirmation exercise), or in the case of "seed & leave" GTD, any resulting traces were never explored (a trace discovery exercise). Any subsequent users of the GTD may conduct a series of tests and provide an independent record that confirms the status of any seeded data and subsequently generated trace. If no trace can be identified in relation to any data-seeding actions following testing, this position should be flagged here. A user may then provide additional information (see field 11 below) which outlines the methods they have deployed as part of their trace discovery processes.
7. *Was Trace Manually Identified?*: A GTD user may manually traverse a file system or data structure and find a relevant trace. This is considered the manual identification of a trace and is one potentially appropriate technique for trace identification. Alternatively, a GTD user may deploy a specific tool to identify a trace (automated identification), where the following three fields are required.
8. *Name of Tool Used*: Where a tool has been used to identify a trace, the name of it should be stated.
9. *Tool Version*: As per.
10. *Tool Configuration Details*: Information describing the name and confirmation settings of any tool function used to discover a trace should be described in detail.
11. *Other observations*: Additional information deemed relevant to the test conducted should be included. As noted above, where methods for trace discovery are deployed and no trace is found, details of these methods should be included. This allows others to evaluate the work conducted. Any interpretation of a trace may be included.

See Table 3 for the "Validation Tests" table with example data.

The role of the "validation tests" feature of the proposed GTD template is to provide a living record of the interactions that have occurred with the dataset and serve as a continuous and dynamic record of its evaluation as more users interact with it. By doing this, potential users of the GTD are able to evaluate the experiences of others with the data and seek to determine whether they can reliably use it.

It is recognized that once any GTD is in circulation, it may be difficult to ensure any of the above-noted validation test information that is generated by third parties is subsequently reflected in all future distributions of the GTD. Therefore, it is suggested that when a future user conducts tests and seeks to provide any validation test information to accompany a dataset, they do so by contacting the dataset provider and providing them with this information so that their GTD can have its GTD template updated. As a result, it becomes important that the date and time regarding when any validation tests were conducted is accurate and maintained, so future

users can establish a chronological timeline of testing and when/if any issues have occurred—akin to version control.

4 | CONCLUSIONS AND DISCUSSIONS

This work has proposed the use of a GTD template for standardizing the documentation accompanying any GTD, while also acting as a guide for defining a minimum benchmark of documentation quality. The fields noted within the GTD template collectively are considered to be the minimum amount of metadata that should be made available regarding the production and makeup of any GTD. It is suggested that this information helps to allow future users of any GTD to evaluate its value and use cases, while also allowing them to risk assess the data and identify any potential reliability concerns that may exist with it. It has been noted that the quality of GTD is dependent upon its accompanying documentation which must allow third parties to understand the processes involved in its creation and to be able to rely upon the data that are generated as a result. GTD documentation must not just be sufficiently detailed but also contain the right type of information, where it is important that those creating GTD are provided guidance on the type of information that must be present for their data to be of maximum value to all those who may use it.

It is recognized that this work suggests that additional meta-data regarding a dataset is required in order for it to be considered reliable and shareable GTD, and this places an additional burden upon the data creator. This extra workload may put off wannabe GTD producers, however, it is suggested that without the inclusion of the GTD metadata suggested in Section 4, any dataset is fundamentally limited in value in any instance. If we are to increase the presence of good quality and reliable GTD in the DF field, it is argued that we need to set a minimum set of requirements and we should seek to focus on producing quality datasets, rather than lots of datasets.

It is necessary to state that the use of a GTD template like that proposed here does not guarantee that any GTD using it is perfect; procedural issues surrounding data seeding and recording may still occur and this is unavoidable. Such a template can never be a silver bullet solution for producing reliable GTD in all instances, however, it is argued that it does provide support to those producing such data, where without it, limited guidance may exist.

CONFLICT OF INTEREST STATEMENT

The author has no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

An Excel file containing all three tables is available as Supplemental Information, enabling any interested party to use it as a template.

REFERENCES

1. Woodhouse IH. On 'ground' truth and why we should abandon the term. *J Appl Remote Sens*. 2021;15(4):041501. <https://doi.org/10.1117/1.JRS.15.041501>

2. Reidy S, Harris R, Gwinnett C, Reel S. Planning and developing a method for collecting ground truth data relating to footwear mark evidence. *Sci Justice*. 2022;62(5):632–43. <https://doi.org/10.1016/j.scijus.2022.09.006>
3. Yannikos Y, Graner L, Steinebach M, Winter C. Data corpora for digital forensics education and research. In: Peterson G, Sheno S, editors. *Advances in digital forensics X: proceedings of the 10th IFIP WG 11.9 international conference, revised selected papers*; 2014 Jan 8–10; Vienna, Austria. Berlin/Heidelberg: Springer; 2014. p. 309–25.
4. Horsman G. Raiders of the lost artefacts: championing the need for digital forensics research. *Forensic Sci Int Rep*. 2019;1:100003. <https://doi.org/10.1016/j.fsir.2019.100003>
5. Horsman G. Forming an investigative opinion in digital forensics. *WIREs Forensic Sci*. 2022;4(6):1460. <https://doi.org/10.1002/wfs2.1460>
6. Hughes N, Karabiyik U. Towards reliable digital forensics investigations through measurement science. *WIREs Forensic Sci*. 2020;2(4):1367. <https://doi.org/10.1002/wfs2.1367>
7. Horsman G, Lyle JR. Dataset construction challenges for digital forensics. *Forensic Sci Int Digit Investig*. 2021;38:301264. <https://doi.org/10.1016/j.fsidi.2021.301264>
8. Göbel T, Baier H, Breiting F. Data for digital forensics: why a discussion on 'how realistic is synthetic data' is dispensable. *Digit Threat*. 2023;4(3):1–18. <https://doi.org/10.1145/3609863>
9. Casey E. *Digital evidence and computer crime: forensic science, computers, and the interne*. 3rd ed. London: Academic Press; 2011.
10. Casey E. Standardization of forming and expressing preliminary evaluative opinions on digital evidence. *Forensic Sci Int Digit Investig*. 2020;32:200888. <https://doi.org/10.1016/j.fsidi.2019.200888>
11. OSAC. Digital evidence subcommittee task group on dataset development. *Guidelines for Dataset Development*. [accessed 2024 Feb 10]. Available from: <https://www.nist.gov/system/files/documents/2022/12/15/OSAC-DE-Guidelines%20for%20Dataset%20Development.pdf>
12. Breiting F, Jotterand A. Sharing datasets for digital forensic: a novel taxonomy and legal concerns. *Forensic Sci Int Digit Investig*. 2023;45:301562. <https://doi.org/10.1016/j.fsidi.2023.301562>
13. Horsman G, Sunde N. Part 1: the need for peer review in digital forensics. *Forensic Sci Int Digit Investig*. 2020;35:301062. <https://doi.org/10.1016/j.fsidi.2020.301062>
14. Sunde N, Horsman G. Part 2: the phase-oriented advice and review structure (PARS) for digital forensic investigations. *Forensic Sci Int Digit Investig*. 2021;36:301074. <https://doi.org/10.1016/j.fsidi.2020.301074>
15. Casey E. The chequered past and risky future of digital forensics. *Aust J Forensic Sci*. 2019;51(6):649–64. <https://doi.org/10.1080/00450618.2018.1554090>
16. Göbel T, Schäfer T, Hachenberger J, Türr J, Baier H. A novel approach for generating synthetic datasets for digital forensics. In: Peterson G, Sheno S, editors. *Advances in digital forensics XVI: proceedings of the 16th IFIP WG 11.9 international conference, revised selected papers*; 2020 Jan 6–8; New Delhi, India. Cham: Springer International Publishing; 2020. p. 73–93.
17. Tully G, Cohen N, Compton D, Davies G, Isbell R, Watson T. Quality standards for digital forensics: learning from experience in England & Wales. *Forensic Sci Int Digit Investig*. 2020;32:200905. <https://doi.org/10.1016/j.fsidi.2020.200905>
18. Du X, Hargreaves C, Sheppard J, Scanlon M. TraceGen: user activity emulation for digital forensic test image generation. *Forensic Sci Int Digit Investig*. 2021;38:301133. <https://doi.org/10.1016/j.fsidi.2021.301133>
19. Forensic Science Regulator. *Forensic Science Regulator guidance: Method validation in digital forensics*. FSR-G-218, Issue 2. [accessed 2024 Feb 10]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_

[data/file/921392/218_Method_Validation_in_Digital_Forensics_Issue_2_New_Base_Final.pdf](#)

20. Garfinkel S, Farrell P, Roussev V, Dinolt G. Bringing science to digital forensics with standardized forensic corpora. *Forensic Sci Int Digit Investig.* 2009;6(Suppl):S2–S11. <https://doi.org/10.1016/j.diin.2009.06.016>
21. Grajeda C, Breiting F, Baggili I. 2017. Availability of datasets for digital forensics—and what is missing. *Forensic Sci Int Digit Investig.* 2017;22(Suppl):S94–S105. <https://doi.org/10.1016/j.diin.2017.06.004>
22. Kari NM, Kebande VR, Swaziland K. Knowledge management as a strategic asset in digital forensic investigations. *Int J Cyber-Sec Digit Forensics.* 2018;7(1):10–20.
23. Computer Forensic Reference Data Set Portal. What is CFReDS? <https://cfreds.nist.gov/>
24. Digital Corpora. Home. [accessed 2024 Feb 11]. Available from: <https://digitalcorpora.org/>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Horsman G. A template for creating and sharing ground truth data in digital forensics. *J Forensic Sci.* 2024;00:1–11. <https://doi.org/10.1111/1556-4029.15524>



DSPACE

<https://dspace.org/>

A template for creating and sharing ground truth data in digital forensics

Horsman, Graeme

2024-07

Wiley

<https://doi.org/10.1111/1556-4029.15524>

Horsman G. (2024) A template for creating and sharing ground truth data in digital forensics. Journal of Forensic Sciences, Volume 69, Issue 4, July 2024, pp. 1456-1466