Received: 18 March 2022 | Revised: 18 April 2022 | Accepted: 21 April 2022

DOI: 10.1111/1556-4029.15053

JOURNAL OF FORENSIC SCIENCES AAFS

ORIGINAL PAPER

Digital & Multimedia Sciences

Crowdsourcing forensics: Creating a curated catalog of digital forensic artifacts

Eoghan Casey PhD | Lam Nguyen MIS | Jeffrey Mates MSSI | Scott Lalliss

DoD Cyber Crime Center (DC3), Linthicum Heights, Maryland, USA

Correspondence
Eoghan Casey PhD, DoD Cyber Crime Center (DC3), 911 Elkridge Landing Road, Linthicum Heights, Maryland 21090, USA.
eoghan.casey.1@us.af.mil

Funding information
American Academy of Forensic Sciences

Abstract

The increasing volume, variety, velocity, distribution, structural intricacy, and complexity of use of digital evidence can make it difficult for practitioners to find and understand the most forensically useful information (Casey E. Digital evidence and computer crime: Forensic science, computers, and the Internet. Academic Press; 2011. p. 31; Pollitt M. The hermeneutics of the hard drive: Using narratology, natural language processing, and knowledge management to improve the effectiveness of the digital forensic process [PhD dissertation]. University of Central Florida; 2011). Digital forensic practitioners currently search for information and solutions in an ad hoc manner, leading to results that are unstructured, unverified, and sometimes incomplete. As a result, certain digital evidence is being missed or misinterpreted. To mitigate risks of knowledge gaps, there is a pressing need for a systematic mechanism that practitioners can use to codify and combine their collective knowledge. This work presents the design and development of a solution that catalogs crowdsourced knowledge of digital forensic artifacts in a well-structured, easily searchable form to support efficient and automated extraction of pertinent information, improving availability and reliability of interpretation of artifacts (general acceptance). Technical implementation and artifact curation are discussed with illustrative examples and recommendations for future work.

KEYWORDS
crowdsourcing forensics, digital forensic artifact, digital transformation, forensic technology innovation, general acceptance, tool testing automation

Highlights

- Queryable crowdsourced Artifact Catalog created by digital forensic practitioners for practitioners.
- Increasing general acceptance and reliable interpretation of digital forensic artifacts and methods.
- Precise definitions of *atomic artifact* and *dependent artifact*, clarifying understanding of artifacts.
- A conceptualization and model of elements necessary and sufficient to represent an atomic artifact.
- An API (application programming interface) to support automated queries by digital forensic tools.

Presented at the 74th Annual Scientific Conference of the American Academy of Forensic Sciences, February 21-26, 2022, in Seattle, WA.

1 | INTRODUCTION

Digital forensic science is a complex domain, complicated by computer technology that evolves rapidly and an influx of new applications on various computing devices. Additionally, criminals use advanced tactics to conceal their activities, making it even more difficult to detect digital evidence [1]. A natural consequence of these trends is the ever-increasing volume and variety of digital artifacts, many of which require decoding or decrypting. The growing compleixty and scale of digital evidence can make it difficult for practitioners to find and understand the most forensically useful information [2,3]. In general, the term *artifact* refers to an observable object resulting from human or automated activities on a computer [4]. The present work defines more specifically an *atomic artifact* as "a singular unit of interpretable data that can be extracted from a given data source" [5]. A *dependent artifact* is then defined as "one or more atomic artifacts needed to expose the atomic artifact of interest, such as a key and initialization vector required to expose encrypted data."

Maintaining the necessary expertise to handle these challenges is beyond the budget, time, and expertise of most organizations in this field, which has led to current digital forensic knowledge being shared among practitioners around the world. The lack of a consolidated reference catalog poses an increasing risk that practitioners might miss or misinterpret digital artifacts. There is an urgent need for the systematic sharing of knowledge across the digital forensic community to ensure reliable, consistent results and to reduce the risk of errors and omissions in digital investigations.

Furthermore, practitioners and decision makers can benefit from a means to assess the interpretation of digital evidence and its reliability [6]. Assessing reliability in a forensic context depends on a deeper shared understanding of digital artifacts and associated meanings, often coming in the form of peer review and past experience. This shared understanding can be supported by a system that enables practitioners to confirm or correct others' interpretations of digital artifacts. A way to show that the interpretation of a given digital artifact has crossed the line between experimental and demonstrable is extremely valuable to decision makers in criminal justice [7]. Rather than relying on the forensic analysis of an individual practitioner, decision makers seek community consensus as demonstrated by admissibility requirements such as the Frye test, Daubert standard, and U.S. Federal Rules of Evidence. Establishing general acceptance of a digital forensic method is challenging due to rapid technological changes and the need to develop novel solutions to handle new types of data sources and devices. The crowdsourcing approach presented here supports this need and mitigates the associated risks, improving reliability of the meaning and interpretation of artifacts.

Digital forensic practitioners value personal trust and are team-oriented so, to be successful, the community must have ownership in the knowledge management solution, which requires a community-based approach [8]. This work presents a multi-organizational, crowd sourced catalog of atomic artifacts created by practitioners for practitioners to support digital investigations. The resulting Artifact Catalog is in a user-friendly, online knowledge management repository that is intended for sharing with the digital forensic community at large.

The Artifact Catalog mitigates risks of knowledge gaps associated with practitioners not knowing which forensic capabilities are most likely to advance their inquiries, missing relevant digital evidence, and misinterpreting forensic findings. Crowdsourcing digital artifacts augments and reinforces existing software capabilities, tool testing, and practitioner training, which are expensive and often outpaced with the rapidly growing number of applications on a variety of computing devices. Crowdsourcing digital forensic knowledge also allows the relevant scientific community to establish general acceptance for specific atomic artifacts. The Artifact Catalog presented in this work uses crowdsourcing to confirm the validity of generally accepted digital artifacts and to alert the community when an artifact has been misinterpreted.

This work precisely defines digital artifacts and their composition in the context of forensic science, providing the groundwork for designing and implementing the Artifact Catalog. Formalizing this conceptualization and model clarifies the elements necessary and sufficient to represent an artifact, and informs the critical review of prior related work. The design and implementation of the prototype Artifact Catalog are described in detail with specific examples of artifact composition provided. This work concludes by discussing uses of the Artifact Catalog and future work.

1.1 | Atomic artifacts

For forensic purposes, it is generally not sufficient to treat an entire file as an artifact; it is usually necessary to be more specific about the atomic artifact(s) contained within the file. Consider a file like the SQLite database `external.db` on Android devices that contains multiple atomic artifacts (Figure 1), including filenames, associated timestamps (datetaken, date_modified, date_added), and geocoordinates when available. It is necessary to treat each singular unit of interpretable data in `external.db` as a separate atomic artifact to distinguish between the distinct interpretations.

An atomic artifact is not just any piece of data extracted from a data source; it is one that answers questions in forensic investigations. As such, the Artifact Catalog was designed to represent only units of data that might be useful for investigations. Atomic artifacts provided in Table 1 are examples, not a comprehensive list of potential artifacts.

It is also necessary to represent the context of an artifact, including what container encapsulates it (registry hive, database, configuration file, memory dump, unallocated space) and the higher level category to which the artifact belongs (e.g., geolocation, user account). Through this structured representation of artifacts and associated context, practitioners have the ability to search the Artifact Catalog for different purposes, including the following:

_data	_size	datetaken	date_...	date_...	latitude	longitude
/storage/emulated/0/DCIM/Camera/20181203_184436.jpg	3851595	1543859076950	1543859077	1543859076	41.9000015258789	12.5022220611572
/storage/emulated/0/DCIM/Camera/20181203_184444_001.jpg	3157664	1543859084028	1543859084	1543859083	41.9000990692762	12.502339049938
/storage/emulated/0/DCIM/Camera/20181203_184604.jpg	3943010	1543859164295	1543859164	1543859164	41.9000015258789	12.5022220611572
/storage/emulated/0/DCIM/Camera/20181203_194104.jpg	4494483	1543862464136	1543862464	1543862464	41.7949981689453	12.255277633667
/storage/emulated/0/DCIM/Camera/20181203_194114.jpg	5032327	1543862474920	1543862474	1543862474	41.8088874816895	12.2349996566772
/storage/emulated/0/DCIM/Camera/20181203_194135.jpg	4258670	1543862495648	1543862495	1543862495	41.7949981689453	12.255277633667

FIGURE 1 Entries in "files" table of external.Db SQLite database containing metadata about files within the path of external storage media or emulated storage (\emulated\0) on android device.

TABLE 1 Atomic artifact examples

Atomic artifact	Description	Category
Geocoordinates of file on Android external storage	Longitude and latitude embedded within a file that was present on Android emulated/external storage media	Geolocation
Chrome location	Geolocation coordinates (longitude and latitude) where browser was last used	Geolocation
Cash App cashtag	Unique identifier that a person can post or share for others to send payments, linked to user account within Cash App	User account
Discord App email	Unique email configured in the Discord app	User account
Echo Dot SSID	SSID of network(s) to which the device was connected	WiFi access point
uTorrent IP address	Last routable IP address associated with the uTorrent client	Network connection

- Searching for a specific item to understand its meaning
- Searching for a specific application to determine all associated artifacts
- Searching for all items on a device related to a certain category of information (such as user accounts)

2 | BACKGROUND

Lacking a structured repository of digital artifacts, practitioners often search the Internet for needed information. Sometimes that works, but this approach does not provide a comprehensive set of data within the field and has the following setbacks:

- Unstructured – search results are unorganized and not uniform
- Unverified – untrusted/unqualified source of information
- Unsupported – weak documentation to explain methods and meaning
- Unstable – unavailability of information in the future
- Incomplete – no information available for new/uncommon artifacts

Prior work demonstrated the value and difficulties of designing a repository of digital artifacts [4,10]. On one hand, the complexity of an ontology-based approach makes implementation more difficult. On the other hand, the over-simplification of treating digital artifacts as files with tagging lacks precision and leaves it to individual users to infer what piece of data within the file is relevant and what the data represents. To strike a balance between these two extremes, practitioners have conceived and populated the Artifact Catalog based on their observations and insights from actual forensic examinations.

Figure 2 illustrates the importance of distinguishing the representation of an artifact, its context, a general recipe describing how to extract it, and a specific implementation of the recipe using a given programming language or tool.

3 | APPROACH AND STRUCTURE

Each entry in the Artifact Catalog requires specific information and context as defined in Table 2.

Figure 3 depicts the structured composition of an entry in the Artifact Catalog, not including dependent artifacts that are described later in this work.

3.1 | Artifact recipes

Within this structure, an artifact requires details that specify what the *singular unit of interpretable data* looks like so that a person or tool

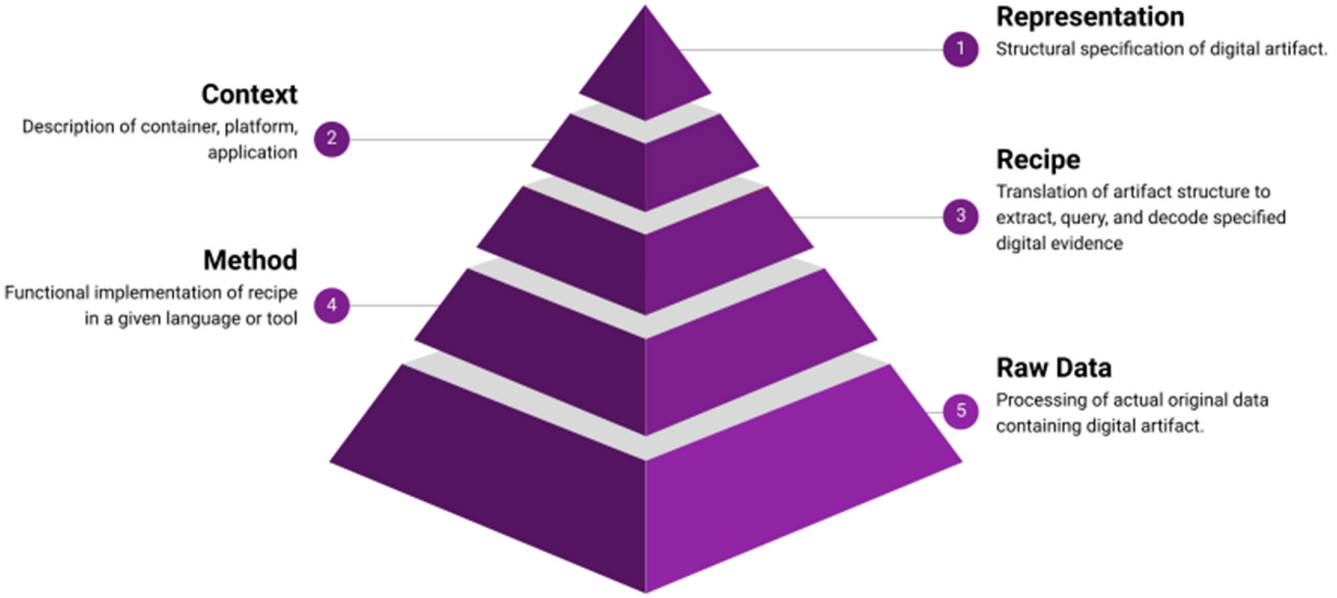

FIGURE 2 Distinguishing an artifact's representation, context, general recipe, and specific implementation.

TABLE 2 Definition of items that make up an entry in the artifact catalog

Item	Definition
Identifier	Deterministic globally unique identifier for reference
Atomic Artifact	Singular unit of interpretable data that is useful for answering a specific question
Dependent Artifact	Other atomic artifact(s) needed to expose the atomic artifact (for example, decryption key and iv)
Category	Higher-level class to which the artifact belongs, when applicable
Container	Full path or data structure of the container in which the actual piece of data is found
Platform	Operating system or platform, including its version
Application	Most closely associated application, including its version
Method	Method used to extract artifact (include versions when applicable)
Reference	Pointer to a shared resource with more information about the artifact, explanation of its meaning/interpretation, and details about the extraction method or analysis

can find it. To specify these details in a general manner independent of any specific method or tool, the Artifact Catalog uses "recipes." A recipe can be a specific XPath Expression, Plist entry, JSON name, Regular Expression, file offset, protobuf item, or SQL Select statement. These recipes specify the XML label, Plist key, JSON name, configuration field, byte range, protobuf item, or position in a SQL statement. Table 3 provides examples of each recipe type.

Table 4 provides an example of an artifact entry with an XML recipe.

3.2 | Method

Different methods or tools can be used to extract a given artifact by following a recipe or similar process. Practitioners can list all tools that support a given digital artifact, which is useful for validating forensic findings and tools. When practitioners find digital artifacts that are not supported by existing tools, they can highlight a gap that developers can fill.

3.3 | Context

Associated Platform, Application, and Container information captures the context of an artifact. The Platform specifies the operating system and version on which the artifact exists, such as Android version 10. If the same artifact exists on multiple versions of an operating system, that is indicated in the version field (for example, 9–11 or All). Similarly, if the same artifact exists on more than one operating system, then each one can be added to the artifact entry. The Application specifies the software application and version that is most closely associated with the artifact. A Container is a data structure that contains artifacts such as Windows Registry hives, SQLite databases, Plists, and configuration files (XML, JSON). Multiple artifacts can be contained within the same Container. For example, Figures 4, 5, and 6 show the following artifacts are all contained within the same SQLite database.

3.4 | Artifact categories

The Category assigned to an artifact ascribes it to a particular class of observable object, such as a User Account, Bank Account, or Geolocation. For example, the contents of a row within a SQL Select

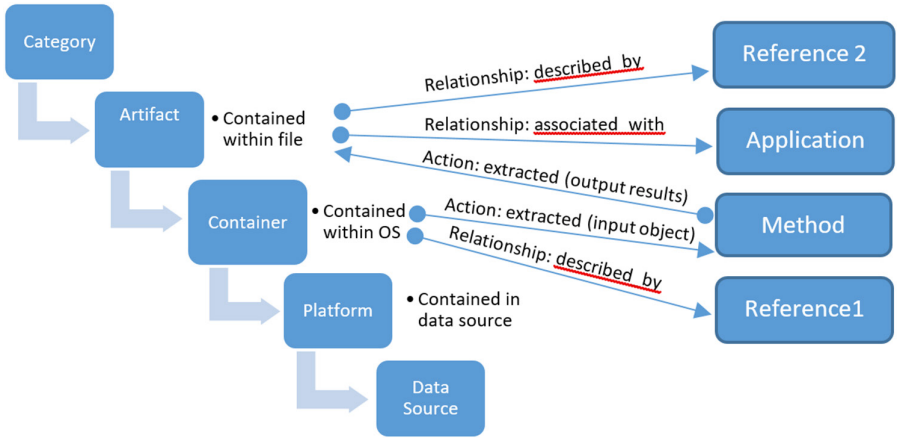

FIGURE 3 An artifact's structured composition within nested containers is illustrated. If there is more than one *singular unit of interpretable data* in a container, each unit is treated as a distinct artifact.

TABLE 3 Recipe type examples

Recipe type	Recipe example	Recipe description
SQLite	SELECT icc_id FROM siminfo	Value stored in "icc_id" field of "siminfo" table
XML	//LOG_CACHE_KEY_USER_ID	Value assigned to XPath "LOG_CACHE_KEY_USER_ID" within Discord preferences XML file
Plist	key: iccid	Value stored in "iccid" key of Plist
Regular expression	^(\+\d{1,2}\s)?\(?\d{3}\)?[\s.-]\d{3}[\s.-]\d{4}$	Regular expression to find phone numbers in a data stream
Protobuf	field: 2	Value stored in second field of protobuf
File offset	offset: 42574	Offset in bytes from start of data stream

TABLE 4 Atomic artifact components for email address registered in discord app on android

Item	Definition
Identifier	Deterministic globally unique identifier for reference
Atomic Artifact	Discord registered user email address
Dependent Artifact	Null
Category	EmailAddress
Container	`/data/data/com.discord/shared_prefs/com.discord_preferences.xml`
Platform	Android 7
Application	Discord
Recipe	`XPATH //LOG_CACHE_KEY_USER_EMAIL`
Method	DC3 script
Reference	DC3 Technical Series Publication 2020–001 describing the most recent email registered in Discord

statement or a JSON object contain username and password information associated with a User Account.

The Category of an artifact can reference the associated `ObservableObject` in CASE/UCO, including `ApplicationAccount`, `EmailAccount`, `WifiAddress`, `Location`, `Identity`. Each of these objects has associated properties and can capture artifact details.* The Platform can also be mapped to `OperatingSystem` in CASE/UCO.

3.5 | Nested artifacts

Entries in the Artifact Catalog support nesting of containers, which is required when an artifact exists in a container that is nested within another container. For instance, Figure 7 shows the SnapChat username is contained in a BLOB, offset 76 bytes, within the `userinfo_coreuserdata` table of the `primary.docobjects` database [11].

3.6 | Dependent artifacts

Sometimes, it is necessary to combine multiple artifacts to obtain digital evidence, particularly when encryption is involved. The Artifact Catalog uses the concept of a dependent artifact to represent an atomic artifact that is required to expose the atomic artifact of interest. For example, decrypting the database `gallery.encrypteddb` in the SnapChat app requires the decryption key within the iOS keychain. This key is represented as an atomic artifact that is required to extract additional encryption keys stored in the `snap_key_iv` table of the `gallery.encrypteddb` database (Table 5).

Another dependent artifact is an encryption key stored within the `snap_key_iv` table which itself is represented as an atomic artifact (Table 6) and is needed to decrypt SnapChat media (snaps) in the Documents\gallery directory (Table 7).

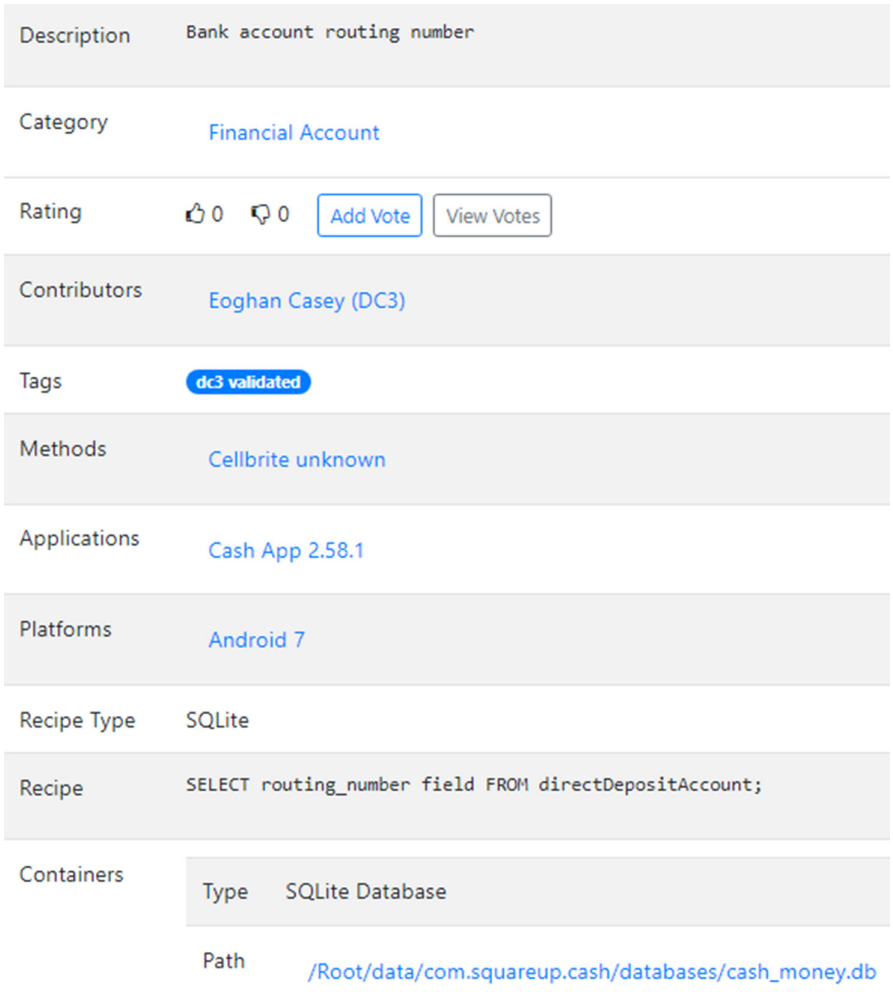

FIGURE 4 Atomic artifact entry for routing number in cash app on android, which is a numerical value of routing_number field in directDepositAccount table.

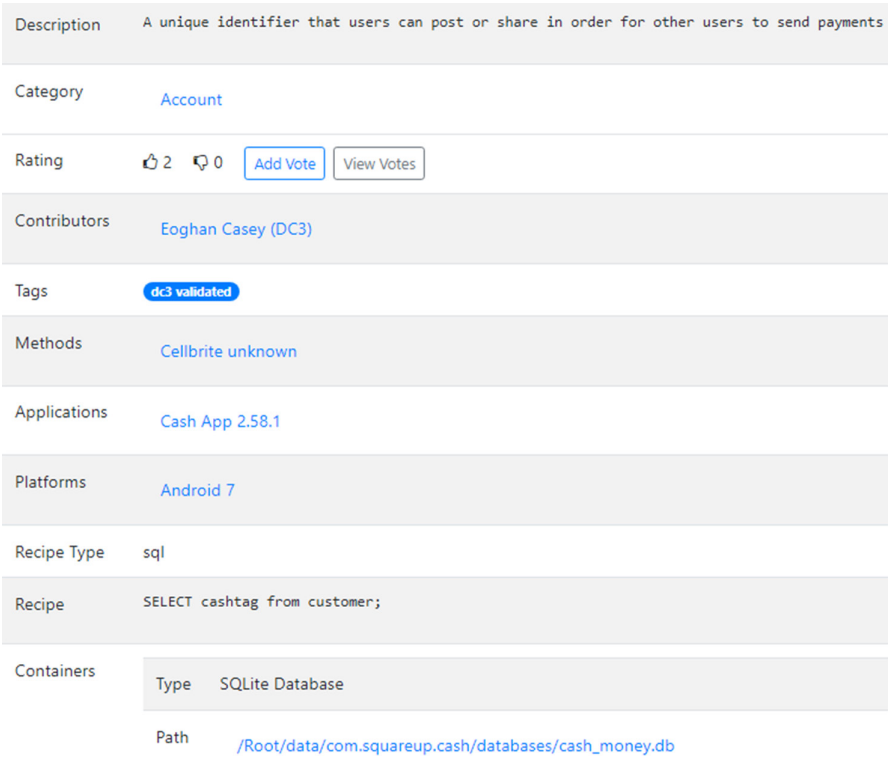

FIGURE 5 Atomic artifact entry for cash app Cashtag on android, which is a numerical value of cashtag field in customer table.

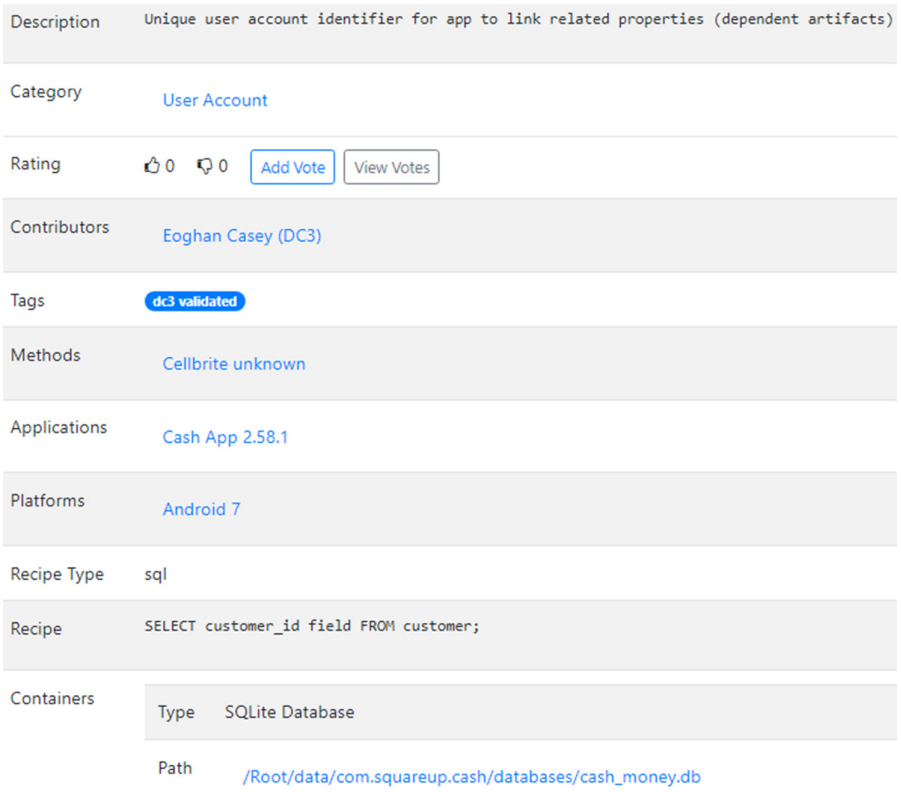

FIGURE 6 Atomic artifact entry for cash app customer ID on android, which is a numerical value of customer_id field in customer table.

In this example, the dependent artifact captures the association among three atomic artifacts: the decryption key of the SQLCipher database, the artifacts within that database, and encrypted snaps on a device. Dependent artifacts are not required for all atomic artifacts. Notable, the CashApp examples in Figures 4, 5, and 6 are not dependent artifacts because each is accessible without requiring the others. There is also no need to define a dependent artifact for multiple artifacts in a particular app or file because these items can already be associated with one another using the value set in the Application field or Container field.

4 | IMPLEMENTATION

This section describes the implementation of the Artifact Catalog.

4.1 | Technology

The Artifact Catalog was implemented as an AWS-hosted MariaDB with user accounts managed by AWS Cognito. The web server was hosted within an ECS cluster with access managed using an application load balancer. General users can query the Artifact Catalog without a user account to support automated processes and tool testing; however, a user account is required to submit artifacts, vote on artifacts, and curate submitted artifacts.

4.2 | Creating artifacts

When creating an entry in the Artifact Catalog, simply adding a file is not sufficient generally. Since an atomic artifact is defined as *a singular unit of interpretable data*, it is usually necessary to be more specific about the artifacts contained within the file.

Consider a database file that contains multiple artifacts such as `external.db` on Android, as shown in Figure 1. Within the `external.db` SQLite database, Android stores metadata about files within the path of external storage media or emulated storage (`\emulated\0`). The files table within the `external.db` database stores filename, timestamp it was added, modified, and other metadata. Because `external.b` contains multiple artifacts, it is not sufficient to treat the file `external.db` as an artifact. Each singular unit of interpretable data in `external.db` must be added to the Artifact Catalog. Figure 8 illustrates how to add the modified timestamp of a file listed in `external.db`.

Figure 9 demonstrates how to specify the characteristics of a new atomic artifact.

When creating a new entry in the Artifact Catalog, users can click the Platform and Application fields to select an option from a list of previously entered terms. The Method field can be populated from a list of previously entered terms as well, but it also allows users to add a new method or tool if necessary. For consistency, users should look for a suitable term on the list before adding a new term.

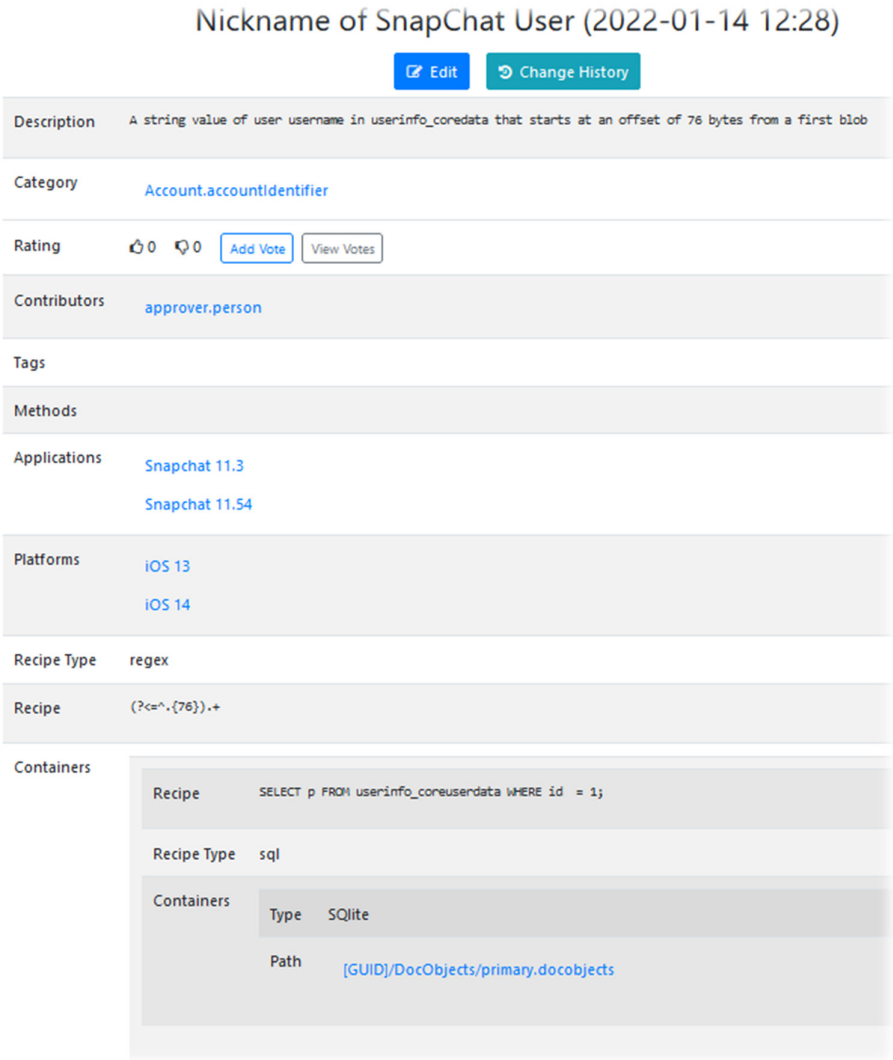

FIGURE 7 Atomic artifact entry for SnapChat user name on iOS, which is a string value within a BLOB stored within the userinfo_coreuserdata table of the primary.Docobjects database.

TABLE 5 Atomic artifact components for encryption key for SnapChat database on iOS

Item	Definition
Identifier	Snapchat-Key-GUID
Atomic Artifact	Key for decrypting SnapChat gallery SQLCipher database
Dependent Artifact	Null
Category	EncryptionKey
Container	`graykey\[UDID]_keychain.plist`
Platform	iOS
Application	SnapChat
Recipe	key: com.snapchat.keyservice.persistedkey
Method	Magnet Axiom
Reference	Decrypt app data using the iOS Keychain and GrayKey (https://support.magnetforensics.com/s/article/Decrypt-app-data-using-the-iOS-Keychain-and-GrayKey)

4.3 | Curating artifacts

Crowdsourcing digital artifacts involves some upfront curation of inputs to ensure that each entry meets required criteria, allowing users to assess and update information themselves collectively. The most common issue encountered in this process is an entire file being proposed as an artifact, which actually contains multiple atomic artifacts that each requires separate entries in the catalog. Therefore, before new or updated artifacts are formally published in the Artifact Catalog, an administrative review is required to maintain consistency. This review process involves correcting spelling errors, removing redundant items from lists, ensuring that the correct Category is specified, adding suitable tags, and checking referenced materials. Figure 10 shows the review area, where newly entered and updated information is held for curation.

When information about an atomic artifact is corrected or updated, the Artifact Catalog retains a record of the changes so practitioners can learn from the updated information and rectify

TABLE 6 Atomic artifact components for encryption key for SnapChat media (snaps) on iOS

Item	Definition
Identifier	Snaps-Key-GUID
Atomic Artifact	Encryption key for SnapChat media (snaps) within SQLCipher encrypted database
Dependent Artifact	Snapchat-Key-GUID
Category	EncryptionKey
Container	`Documents/gallery_encrypted_db/<digit>/<uid>/gallery.encrypteddb`
Platform	iOS
Application	SnapChat
Recipe	SELECT key, iv FROM snap_key_iv WHERE id = snap_id
Method	Magnet Axiom
Reference	Decrypting and extracting juicy data, Snap! (https://xperylab.medium.com/decrypting-and-extracting-juicy-data-snap-17301aa57a87)

TABLE 7 Atomic artifact components for encrypted SnapChat media (snaps) on iOS

Item	Definition
Identifier	Snaps-GUID
Atomic Artifact	SnapChat media (snaps) encrypted using keys and ivs
Dependent Artifact	Snapchat-Key-GUID, Snapchat-IV-GUID,
Category	Picture
Container	`Documents/gallery`
Platform	iOS
Application	SnapChat
Recipe	Directory
Method	Magnet Axiom
Reference	Decrypting and extracting juicy data, Snap! (https://xperylab.medium.com/decrypting-and-extracting-juicy-data-snap-17301aa57a87)

earlier misunderstandings in their own work. The Artifact Catalog provides a revision history for each entry, including which user made the revision.

4.4 | Artifact voting

Since it is not feasible for curators to validate hundreds or thousands of artifacts before publication, the Artifact Catalog uses a simple voting mechanism to mark unreliable information and reinforce reliable artifacts. When practitioners agree with the interpretation of an artifact, they click Add Vote and vote with a thumbs up. When practitioners disagree with the interpretation of an artifact, they click Add Vote and vote with a thumbs down. If an artifact has a high number of thumbs-ups votes, it has greater level of general acceptance.

5 | USAGE

Practitioners can use the Artifact Catalog to support forensic examinations; tools can use it to support testing and validation. Decision makers can use the Artifact Catalog to assess the general acceptance of a specific digital artifact and its interpretation.

5.1 | In-field decisions

At the scene of a crime, investigators can use the Artifact Catalog to query devices found to understand what types of information those devices might contain. If this is too difficult for non-experts, investigators at the crime scene can submit device photographs, which the Artifact Catalog can then attempt to recognize and query. The query results can support decisions on whether to preserve a device for forensic examination.

5.2 | Forensic focus

Practitioners can use the Artifact Catalog to search for pertinent artifacts for a given digital device and investigative focus. Practitioners can perform more refined searches by combining multiple components of an artifact and its context. When practitioners are conducting a forensic examination on a smartphone, they can query all user account artifacts using a combined query such as UserAccount AND iOS. Similarly, practitioners can use a combined query to retrieve all artifacts containing geolocation information on an Android device (Geolocation AND Android).

When a specific application is of interest, practitioners can use a combined query to retrieve related artifacts on a given kind of device (such as Cash App AND Android). Forensic examiners can use the query results to look for the artifact using their preferred forensic method/tool.

5.3 | Investigative questions

Combined queries of the Artifact Catalog can retrieve all artifacts related to investigative questions (5WH), such as all entries that contain information about the user of the system or the location of the device:

- Who = User Accounts OR Contacts OR ...
- Where = Geolocation OR Addresses OR ...

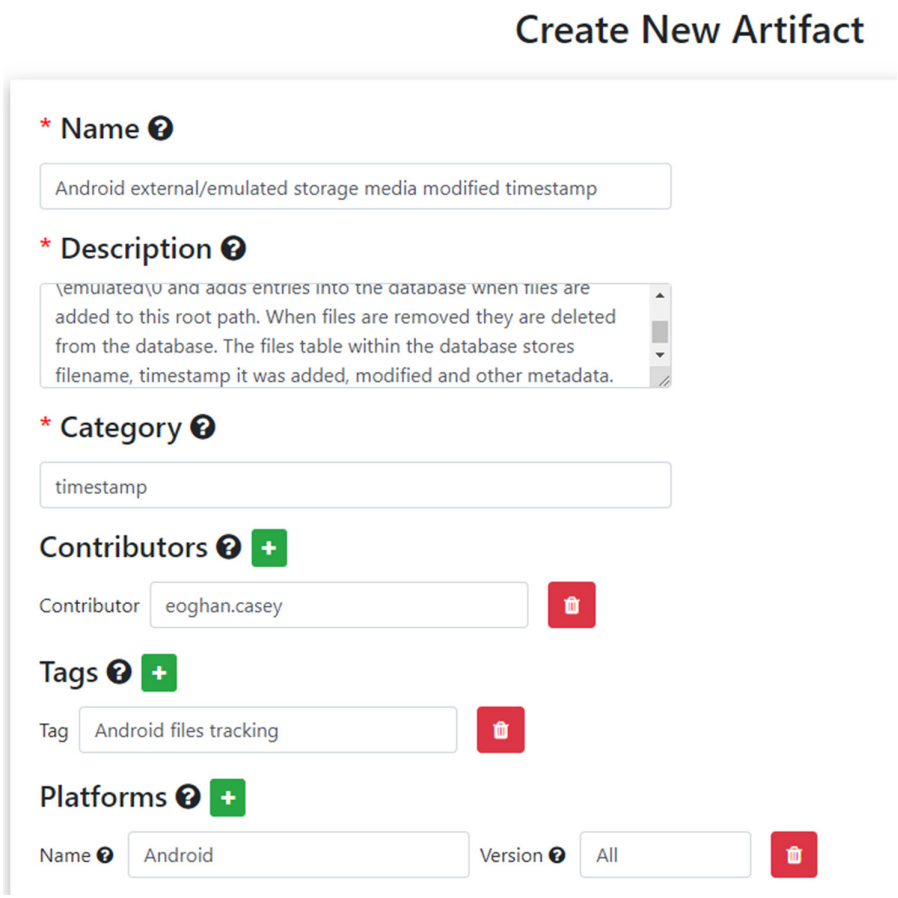

FIGURE 8 Creating a new entry in the artifact catalog

5.4 | Link discovery

Practitioners can query the Artifact Catalog for connective artifacts indicating that a given device is connected with other devices. This link discovery can drive additional inquiries to obtain other devices that might contain related digital evidence. Link discovery can also be useful in cases involving a number of devices, giving practitioners insight into connections between devices. For example, when a device connects to a Wireless Access Point, it stores the associated SSID, and the Wireless Access Point might have captured the MAC address of the connected device. This type of digital evidence can indicate the presence of a given device in a particular area during a time of interest.

5.5 | Tool testing automation

Automated queries of the Artifact Catalog can augment tool testing (artifacts detected versus missed). Digital forensic tools can be tested to determine which indexed entries in the Artifact Catalog are fully/partially supported by each tool. For each artifact, the Artifact Catalog cites (links to) supporting documentation and/or research results (such as CCoE resources, external publications), which provide additional information for developers and practitioners to learn more about specific digital traces. These resources are useful for developing tools as well as explaining digital evidence in court.

5.6 | Forensic value and general acceptance

Aggregated voting results can help determine which artifacts are most widely used and confirmed across the community. In addition, the value and acceptance of artifacts can be assessed by tracking the frequency of specific artifacts occurring and being used to address associated investigative questions (5WH). This aggregated information can assess the reliability and stability of artifacts, which helps to establish general acceptance.

6 | CONCLUSIONS AND FUTURE WORK

The Artifact Catalog provides the community with a robust framework for organizing and sharing knowledge about digital atomic artifacts. Each entry in the Artifact Catalog cites supporting documentation or research results, which provides additional information for practitioners to learn more about the specific item of data. Automated processes within forensic tools can query the Artifact

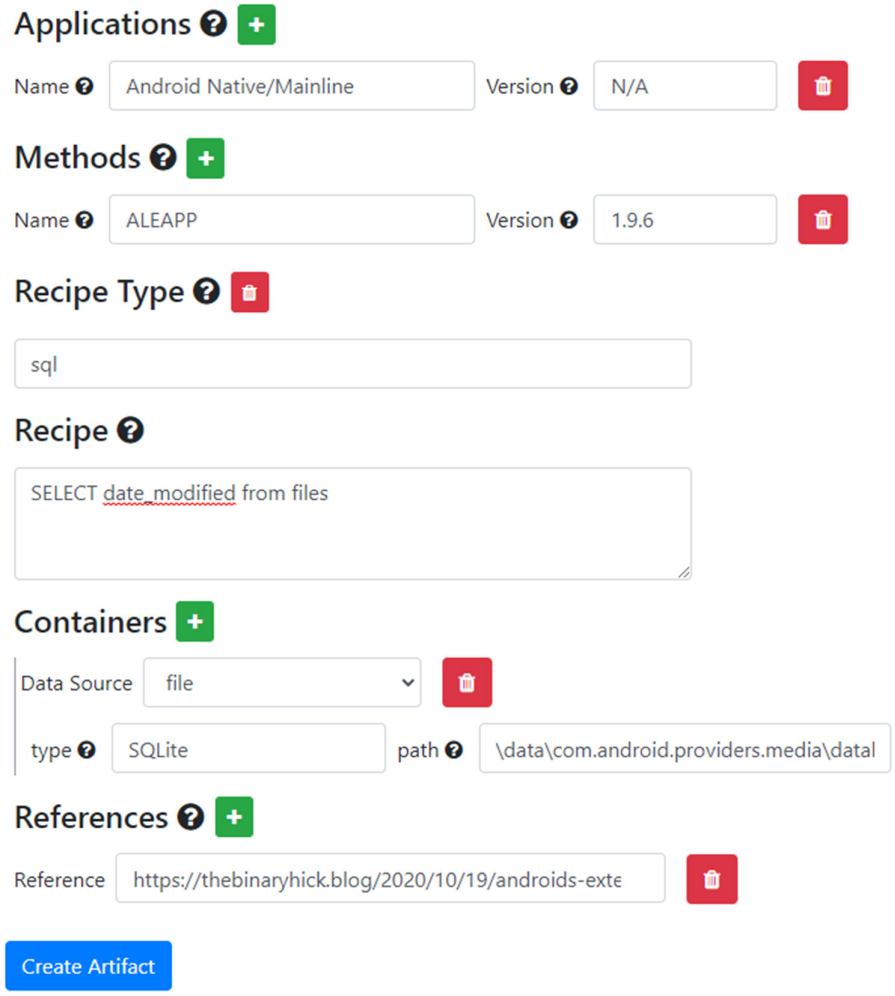

FIGURE 9 Adding methods, containers, recipe, and references to a new artifact entry

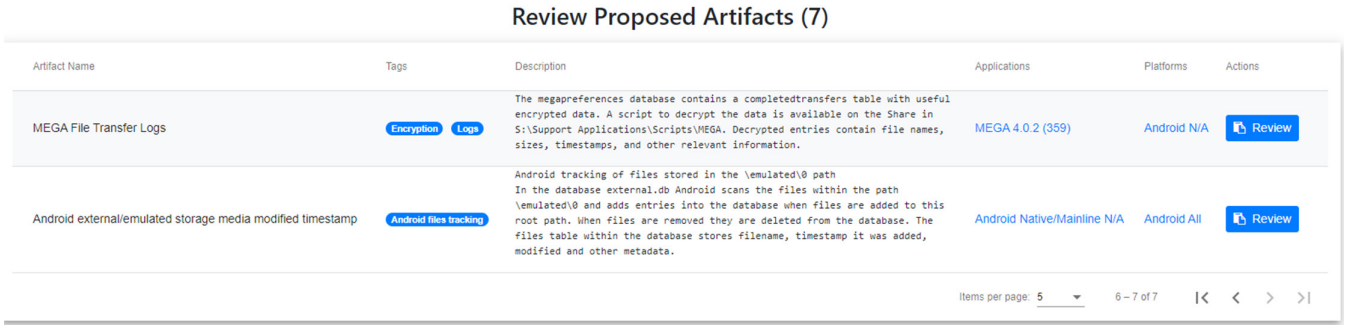

FIGURE 10 Reviewing new/updated artifacts

Catalog and report on or bookmark pertinent artifacts for a given digital device, forensic focus, and investigative question.

This initiative greatly reduces the amount of time and effort that individuals and organizations expend studying digital artifacts independently. It also allows the relevant scientific community to establish general acceptance for specific digital artifacts. In addition, community curation of digital artifact knowledge promotes thorough, consistent, and repeatable digital forensic results.

To support interoperability, automation, and semantic analysis, future enhancements to the Artifact Catalog might include mapping artifact data to existing ontologies, such as CASE/UCO (technical details) and DESO (axiomatic/semantic structure under 5WH investigative questions).

A longer range vision for the Artifact Catalog might involve aggregating statistics about atomic artifacts to determine which ones are most useful in answering source-level and activity-level questions.

ACKNOWLEDGMENTS

Many thanks to Keith Chason and Rebekah Qu, the Artifact Catalog prototype developers. Our deep appreciation to Alyssa Lisiewski, Jon Mayes, Lisa Sharpe, Steve Uder, Kevin Westerman, and others in DC3 for their ongoing contributions to this initiative.

ENDNOTE

* https://github.com/ucoProject/UCO/blob/master/uco-observable/observable.ttl

ORCID

Eoghan Casey https://orcid.org/0000-0001-7219-317X

REFERENCES

1. Casey E. Reinforcing the scientific method in digital investigations using a case-based reasoning (CBR) system [PhD dissertation]. Belfield, Dublin: University College Dublin; 2013.
2. Casey E. Digital evidence and computer crime: forensic science, computers, and the internet. 3rd ed. Waltham, MA: Academic Press; 2011. p. 31–3.
3. Pollitt M. The hermeneutics of the hard drive: using narratology, natural language processing, and knowledge management to improve the effectiveness of the digital forensic process [PhD dissertation]. Orlando, FL: University of Central Florida; 2011.
4. Brady O. Exploiting digital evidence artefacts finding and joining digital dot [PhD dissertation]. Kings College London: London, U.K.; 2018. https://kclpure.kcl.ac.uk/portal/files/95696488/2018_Brady_Owen_0966104_ethesis.pdf
5. Nguyen L, Casey E. Crowdsourcing forensics. Proceedings of the 74th annual scientific conference of the American Academy of forensic sciences; 2022 Feb 21–25; Seattle, WA. Colorado Springs, CO: American Academy of Forensic Sciences; 2022.
6. Biedermann A, Kotsoglou K. Digital evidence exceptionalism? A review and discussion of conceptual hurdles in digital evidence transformation. Forensic Sci Int Synergy. 2020;2:262–74. https://doi.org/10.1016/j.fsisyn.2020.08.004
7. Casey E, Zehnder A. Inter-regional digital forensic knowledge management: needs, challenges, and solutions. J Forensic Sci. 2021;66(2):619–29. https://doi.org/10.1111/1556-4029.14613
8. Biros D, Weiser M, Witfield J. Proceedings of the 5th Australian digital forensics conference; 2007 Dec 3–4. In: Valli C, Woodward A, editors. Managing digital forensic knowledge an applied approach. Perth, Western Australia. Perth, Western Australia: Edith Cowan University; 2007.
9. Grajeda C, Sanchez L, Baggili I, Clark D, Breitinger F. Experience constructing the artifact genome project (AGP): managing the domain's knowledge one artifact at a time. Digit Investig. 2018;26(Suppl):S47–58. https://doi.org/10.1016/j.diin.2018.04.021
10. Casey E, Brady O. Digital artifact reference library. Proceedings of the 70th annual scientific conference of the American Academy of forensic sciences; 2018 Feb 21–25; Baltimore, MD. Colorado Springs, CO: American Academy of Forensic Sciences; 2018.
11. Duffy J. SnapChat – A false sense of security. https://www.magnetforensics.com/resources/mvs-recording-snapchat-a-false-sense-of-security/. Accessed 1 April 2021.

How to cite this article: Casey E, Nguyen L, Mates J, Lalliss S. Crowdsourcing forensics: Creating a curated catalog of digital forensic artifacts. J Forensic Sci. 2022;00:1–12. https://doi.org/10.1111/1556-4029.15053