Forensic Science International: Synergy 6 (2023) 100323

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: www.sciencedirect.com/journal/forensic-science-international-synergy

Enhancing research and collaboration in forensic science: A primer on data sharing

1. General information on collaboration

Public (government) forensic laboratories are largely dependent on commercial vendors and academic institutions to drive the development of new and innovative technologies and methods. While larger government and private laboratories often have departments, sections, or personnel devoted to research, development, testing & evaluation (RDT&E), or validations, small to mid-size laboratories often do not have the same capacity to perform large-scale validations or execute transformative research projects. Therefore, successful RDT&E activities depend on engagement between public forensic laboratories and corporate or academic institutions. This may come in a variety of forms, for example, funded and non-funded collaborations. These collaborations commonly involve the need to share data and, particularly in the DNA and latent print disciplines, may be subject to additional requirements related to data privacy, human subjects research protections, and data security. These requirements may differ based on the different parties involved and their disparate corporate, institutional, or government policies. Therefore, successful collaboration depends on a clear understanding of the administrative and technical roles and responsibilities of the parties and the expectations for communicating information and sharing data which. This is necessary to facilitate the communication and review of broad-scope information such as legal matters (non-disclosure agreements (NDAs) or Institutional Review Board (IRB) applications) and more detail-oriented data sharing (what data is needed, file naming schemes, data organization, and determining how data will be transferred). This paper will outline the process and components of effective data sharing within the forensic science community, which will lead to more successful research partnerships and more impactful research.

2. Establishing a data sharing agreement

Data generated by forensic laboratories may have several layers of confidentiality, including but not limited to, data associated with non-adjudicated casework or identifiable private information of biospecimens. Therefore, ensuring the ethical and confidential use of this data by collaborators is essential. Establishing formal data sharing agreements in advance of the transfer of data ensures that all parties including the researchers, scientists, administrators, and legal teams from each institution agree on the terms, use, transfer, and storage of the data. This prevents potential misunderstandings or miscommunications that could compromise the partnership.

Data sharing agreements come in several forms; more generally these agreements can be initiated within the terms of a confidential disclosure agreement (CDA) or non-disclosure agreement (NDA), and they provide a standard legal framework to ensure information or sensitive data remains protected. These agreements include the general terms, disclosure period, disclosing party or parties, disclosure coordinators, confidential information to be shared, and the purpose of the disclosure. All information, except for the general terms, should be entered and reviewed by the disclosure coordinators at the initiating and collaborating institutions (e.g., principal investigator and laboratory director or section supervisor). A generalized process for completing a CDA is shown in Fig. 1. The general terms (legal.

framework) and the information entered by the Principle Investigators are then reviewed by the institution's designated approval authorities or signatory officials (e.g., legal or sponsored program office). The agreements can be initiated by either collaborator using an institutionally approved CDA or NDA template that has undergone prior legal review. Once that is completed by the initiating party, the agreements should be examined by both parties' reviewing officials (e.g., legal department or sponsored programs office). During the legal review, additional considerations and details may be added to further define and clarify information within the data sharing agreement. Once legal reviews are complete, the data sharing agreement will be sent to the designated signatory authority of each party for final approval. At the conclusion of this process, data sharing can commence. If the agreement involves multiple instances of data transfers over a long period of time, it should be reviewed periodically to ensure each party is satisfied with the progress and compliance. The logistics of the data sharing process, under the umbrella of the CDA or NDA, can be addressed through drafting memos or, more formally, with a data sharing agreement (DSA). In either instance, data remains protected under the CDA/NDA agreement.

Parties should also consider the logistical details of data sharing including the format of the data, transfer mechanism, and organizational structure of the data (metadata—sample names, description). This information can be organized in a single bulleted list in a memo or email. For example, the sample naming scheme may be critical since it often contains highly descriptive information about the sample preparation and parameters that were used in the analysis. It should never be assumed that the data provider will send all associated metadata with the samples unless it is requested. By clearly outlining the logistical details of data sharing, both parties can efficiently and effectively transfer and review data.

https://doi.org/10.1016/j.fsisyn.2023.100323
Received 20 December 2022; Received in revised form 10 February 2023; Accepted 21 February 2023
Available online 24 February 2023

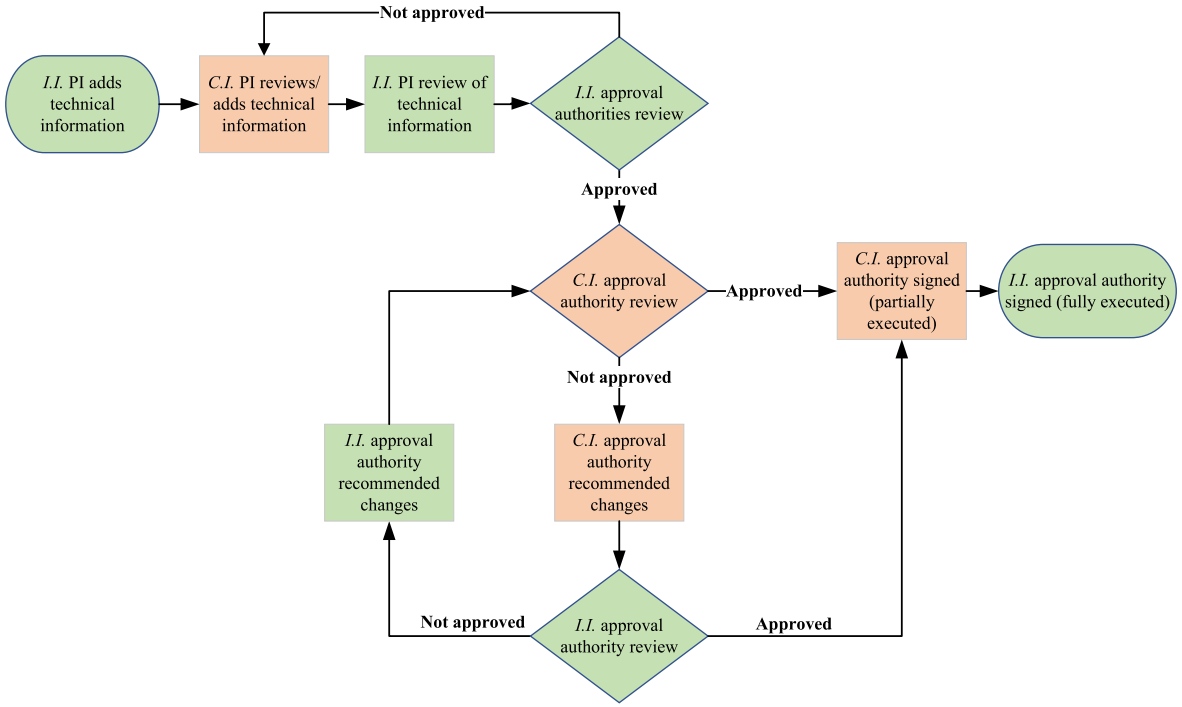

Fig. 1. A generalized process map of the execution of a CDA, where the Initiating Institution has a pre-approved CDA template. I.I. – initiating institution; *C.I.* – collaborating institution.

3. Additional considerations within a data sharing agreement

3.1. Publication

Publication of results should be discussed during the CDA review to prevent the potential release of confidential or private information. Forensic laboratories may be apprehensive about the release of data or results but must not lose sight of the ultimate benefit of sharing data: research enables new and improved forensic capabilities. Publishing research is an important step in the development, acceptance, and transition of new tools and technologies into forensic operations. Publication is also critical to academic researchers who rely on publications as a metric of success and contributions to the field. To ensure results can be published, the party sharing the data should guide the discussion to ensure that sensitive data is safeguarded to their standards. For example, DNA-based data may contain genotypic data, which — although it may not be classified as personally identifiable data — may require privacy protections. If necessary, the researcher can omit genotypic information from publications or presentations. At a minimum, the researcher must not include any information that unites the data with a particular individual.

3.2. Human subjects research considerations

Projects involving human subjects research may be subject to specific requirements outlined by the United States Department of Health and Human Services Office of Human Research Protections in the Code of Federal Regulations, Title 45 Public Welfare Department of Health and Human Services Part 46 Protection of Human Subjects [1]. This code is referred to as the Common Rule and is applicable to a group of 20 federal agencies including the Departments of Justice, Homeland Security and Defense, and the National Science Foundation [2]. When engaging in human subject research, federally funded institutions are required to have a Federal Wide Assurance (FWA) with the government in which the institution pledges to adhere to the Belmont Report and Federal Regulations 45 CFR 46-PartA (Common rule) [1].

Institutional Review Board (IRB) approval for a project involving human subjects research may be required for sharing of data that contains identifiable private information or biological specimens (e.g., genetic data or fingerprints) Generally, IRB approval is required for a project involving an interaction or intervention with a human subject where identifiable private information or biological specimens are collected or analyzed (e.g., collecting buccal swabs for DNA validation projects). The IRB application requires that the principal investigator identify how and why the samples will be analyzed, including the final disposition of the samples or data (e.g., samples saved for later use or destroyed). The level of privacy or security considerations is largely dependent on how directly the specimen or data can be linked to a specific individual. In this context data sharing can be handled in several ways: (1) the interinstitutional partnership and use of the samples or data by both institutions can be proposed in the IRB application, (2) an amendment can be requested to allow another institution to use the data and (3) data can be provided to a partner performing secondary research, where no IRB is required by the partnering institution when the data or samples cannot be linked to a specific individual. Additional information can be found at the United States Health and Human Services Guidance website [3] and through the institution's legal counsel or Office of Research Protections. Note, the General Data Protection Regulation (GDPR) may apply if the project involves members of the European Economic Area [4].

4. Data sharing platforms and data security

Because there are many different data sharing platforms, researchers should consider the data type, quantity (size), and security requirements before choosing one. Funded projects may also include data sharing requirements. For example, grants awarded through the Department of Justice require data to be uploaded to a public repository—the National Archive of Criminal Justice Data (NACJD). It is important to note that only data that can be publicly shared can be uploaded to repositories such as the NACJD; data protected under CDA or NDAs or IRBs should not be uploaded unless it is expressly permitted under the terms of the

agreements.

4.1. Data security

The required level of data security is also an important consideration when selecting a file sharing platform. Institutions or agencies may have specific legal or privacy-based security requirements while specific government partners may have additional requirements to ensure compliance with applicable federal, state, department, and organizational rules and regulations. It is best practice to discuss each party's requirements prior to engaging in a research partnership or exchanging data. If the government is providing data to another party, the data and the intended use of the data may require a review or multiple reviews by authorized individuals for ethics, security, intellectual property, and human subjects' protection. Academic researchers who have limited experience working with government partners should become familiar with the following standard data security measures:

Operations Security (OPSEC) – "Systematic and proven process by which potential adversaries can be denied information about capabilities and intentions by identifying, controlling, and protecting generally unclassified evidence of the planning and execution of sensitive activities. The process involves five steps: identification of critical information, analysis of threats, analysis of vulnerabilities, assessment of risks, and application of appropriate countermeasures" [5].

Information Security (INFOSEC) – "The protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, and availability" [6].

During the review process, reviewers may request specific actions to be taken (e.g., changes to the language in the agreement). For example, reviewers may require encryption of the data, de-identification of participant information, removal of personally identifiable information (PII), or even require a specific government file-share application. Additionally, most institutions have signatory officials who are responsible for the execution of the agreement while some government partners may have an individual that has designated release authority for government information and data. Government partners may also be required to provide information to their Public Affairs Officer for approval.

4.2. Data sharing platforms

Generally, the simplest method is typically the best method; for example, if file sharing is the primary task, choose a method that focuses on file sharing (e.g., Microsoft OneDrive, Google Drive (Google One), Dropbox, and Box). These platforms will typically have the most file storage space available and are easily accessible for collaborators in different institutions. If team-based collaborative work is integral to the project, one can choose collaborative communication platforms such as Microsoft Teams, Slack, or Discord. However, these platforms do not typically have the storage space required to transfer large data sets.

4.3. Recommended data sharing methods and platforms

4.3.1. Microsoft OneDrive

Microsoft OneDrive is a cloud-based storage service providing real-time version control and editing of files that permits multiple users to access and edit files from any compatible device. OneDrive is included in most institutional Microsoft Office suite licenses; however, a Microsoft account can also be created free through a collaborator's institutional license. Sharing files with other users is simple—a file or folder can be moved or copied into the OneDrive directory (Windows – File Explorer or Apple – Mac Finder) or to the OneDrive website. Folders or files can then be shared with select individuals by either sending an email directly to collaborators through OneDrive or by generating a link that can be emailed through the default server to the intended collaborators. The recipient of the data does not require a Microsoft account to access the data.

Microsoft OneDrive secures data by several means: (1) the systems are access-controlled, limiting access to Microsoft engineers unless there is an incident that requires such access. In this case, permission to access these files must be granted by the customer or the customers' designated representatives; (2) Microsoft has real-time monitoring of the security of the system to prevent attacks on the system; (3) Microsoft servers are access controlled, limited to essential personnel, and verified through multifactor authentication—this includes physical means of security such as security officers, video and motion surveillance; (4) User networks are not accessible through the Microsoft corporate network and include firewalls that further control access; (5) All files are encrypted with unique AES256 keys, which are further encrypted in the Azure Key Vault; (6) OneDrive features virus scanning on all downloads; (7) Accounts are monitored for suspicious sign-in attempts; (8) Data can be recovered in the event of malicious attacks, inadvertent deletions, or file corruption; (9) Files can be password protected and shareable links can be assigned an expiration date; (10) OneDrive has an option for two-factor authentication [7]. In addition, OneDrive has not had a major security breach to date, and institutional information technology groups will likely have additional security including firewalls and updated virus and malware detection. Academic and government agencies with business accounts may have client-side encryption where encryption keys are maintained by the institution and, therefore, if Microsoft experienced a security breach, the encryption keys would be inaccessible, and the data would remain secure.

4.3.2. Dropbox

Dropbox, established in 2007, is one of the most used cloud-based storage and file sharing platforms. Basic Dropbox accounts are free with up to 2GB of storage, but additional storage space is available for a fee. Dropbox can be accessed through an account portal on the web or through a desktop application (located in Windows File Explorer or Apple Finder). File upload is performed by either selecting upload and choosing the files of interest, by "dragging and dropping" or by using the copy and paste features. Folders can also be created within Dropbox to organize files and share groups of files with collaborators. The recipient of the data does not need a Dropbox account to access the shared data.

Dropbox security measures include the use of AES256 encryption and Secure Sockets Layer (SSL)/Transport Layer Security (TLS) to secure data when transferring data from the application to the servers—this can prevent data from being intercepted while in transit to the servers. Like OneDrive, Dropbox regularly scans for viruses and attacks and offers the option of two-factor authentication. While Dropbox does not provide encryption on the client-side, users are able to encrypt the data through third-party vendors. However, doing so would only prevent access to the files by Dropbox employees; encrypted data would still be vulnerable to security breaches of the third-party vendors [8]. While Dropbox was the target of several hacks in 2011 and 2012, its security protocols have since been improved by the aforementioned methods [9].

4.3.3. Google Drive/Google One

Google Drive/Google One is a cloud-based storage and collaboration platform; a Google account is required for access. Google Drive features real-time version control and editing of files as well as 15GB of free cloud file storage. Files or folders can be "dragged and dropped" onto the My Drive page or by "right clicking" and choosing file or folder upload. To share a file, the user can "right click" the file or folder and select "Share". The share menu then allows the user to enter email addresses for those individuals they wish to have access to the files or folders. The link to the file location can also be shared or restricted based on the collaborators' preferences. Access can be selected in the share menu, providing access as a "Viewer," (the recipient can only view the file), "Commenter," (the recipient can leave comments but cannot make changes, and sharing is restricted), or "Editor," (the receiver can make changes, leave

comments, accept/reject suggestions, and share files). Ownership changes and access removal can also be managed in this menu. The recipient does not need a Google account to access shared data.

Google Drive/Google One security features are similar to those of Dropbox, including HTTPS (Hypertext Transfer Protocol Secure), two-factor authentication, and 256-bit AES encryption. Google Workspace (which includes Google Drive) can also have client-side encryption enabled; however, this is an exclusive feature of the Enterprise and Education Standard and Education Plus packages [10].

4.3.4. Hard and USB drives

External hard drives represent a simple and secure means to transfer data; however, these must be purchased and mailed to the collaborator. These drives can be encrypted (recommended – AES 256-bit encryption [11]), electronically password protected, and they have physical password keys on the exterior of the device. The primary risk associated with these drives is the potential for data corruption and/or the potential for damage or loss in the mail. Therefore, a backup of the data should be made to ensure that it can be recovered in the event data is corrupted or lost. It is strongly recommended that new external hard drives are used, thereby mitigating any risk that a previously used hard drive contains viruses, malware, or ransomware [12].

See Table 1 for a summary of the data sharing methods and platforms.

4.3.5. Publicly available data

A data set obtained through collaborative agreements may not meet the sample size needs of the project therefore it may be possible to supplement the data set with publicly available data sets. Examples of such data sets or data repositories are the PROVEDIT database (electronic DNA profile data) [13], the American Society of Crime Laboratory Directors Forensic Research Committee Validation and Evaluation Repository [14], the National Archive of Criminal Justice Data (NACJD) [15], the NIST Science Data Portal for Forensic Science [16], the Center for Statistics and Applications in Forensic Evidence Forensic Science Data Portal [17] and the NIST Special Database 300 (fingerprint data) [18]. However, there is a lack of widely available public forensic data sets that can be leveraged for research and evaluation purposes, and we strongly encourage funding agencies, researchers, and industry to consider adding strategic initiatives to help build upon the limited amount of publicly available data.

5. Final thoughts

The use of formal collaborative or data sharing agreements is highly recommended when engaging in inter-institution projects. This ensures that proper roles, expectations, and security protocols are in place and understood. Informal, "hand-shake" agreements can lead to potential issues if all parties are not in complete agreement or do not fully understand the scope and use of the shared data. Therefore, we recommend the use of a CDA to ensure data remains secure and all parties understand their basic roles and responsibilities. Beyond a CDA, a formal data sharing agreement will ensure a comprehensive understanding of the roles and responsibilities while also including logistics-related information which will ensure efficient and effective data sharing. Information about the type and size of data, preferred platforms, and security requirements should be used to determine the data sharing method. It is the authors' hope that this article will help forensic organizations successfully establish data sharing agreements and select effective data sharing platforms that will ultimately help advance collaborative forensic science research.

Funding

No funding

Table 1
An overview of data sharing methods. *Additional security features may be implemented by the institution.

Service	Storage/Cost	Encryption	Client-side encryption	Account required (to share/to access)	Real-time version control
Microsoft One Drive	5 GB – Free (OneDrive Basic 5 GB) 50 GB – $9.99/year (OneDrive Standalone 50GB) 1 TB – $69.99/year (Microsoft 365 Personal)	256-bit AES	Yes	Yes/No	Yes
Dropbox	2 GB – Free 2 TB – $9.99/month (Individual – 1 user) 3 TB – - $16.58/month (Professional – 1 user) 5 TB – $15.00/month/per user (Standard 3+ users)	256-bit AES	No	Yes/No	No
Google Drive/ Google One	15 GB – Free 100 GB – $19.99/year (Basic) 200 GB – $29.99/year (Standard) 2 TB – $99.99/year (Premium)	256-bit AES	Yes (for Enterprise & Educational licenses)	Yes/No	Yes
Hard Drives/ USB Drives	Various (64 GB USB Drive - $9.99+)	Encrypted & Non-encrypted	No	No/No	No

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Code of federal Regulations Title 45 public Welfare parts 1 to 199. https://www.gpo.gov/fdsys/pkg/CFR-2017-title45-vol1/pdf/CFR-2017-title45-vol1.pdf, 2017. (Accessed 6 September 2018).
[2] US Department of Health and Human Services, Common rule departments and agencies. https://gov.ecfr.io/cgi-bin/ECFR, 2018. (Accessed 8 August 2022).
[3] Guidance | HHS.gov (n.d.), https://www.hhs.gov/ohrp/regulations-and-policy/guidance/index.html. (Accessed 6 February 2023).
[4] General data protection regulation (GDPR) – official legal text (n.d.), https://gdpr-info.eu/. (Accessed 6 February 2023).
[5] NIST computer security resource center - operations security (n.d.), https://csrc.nist.gov/glossary/term/operations_security. (Accessed 22 November 2022).
[6] NIST computer security resource center - information security (n.d.), https://csrc.nist.gov/glossary/term/information_security. (Accessed 22 November 2022).
[7] How OneDrive safeguards your data in the cloud (n.d.), https://support.microsoft.com/en-us/office/how-onedrive-safeguards-your-data-in-the-cloud-23c6ea94-3608-48d7-8bf0-80e142edd1e1. (Accessed 3 August 2022).
[8] Dropbox help center | Dropbox help (n.d.), https://help.dropbox.com/. (Accessed 3 August 2022).

[9] Cloud storage: how secure are Dropbox, OneDrive, Google drive, and iCloud? | IT PRO (n.d.), https://www.itpro.com/cloud-security/34663/cloud-storage-how-secure-are-dropbox-onedrive-google-drive-and-icloud. (Accessed 3 August 2022).
[10] About client-side encryption - Google workspace admin help (n.d.), https://support.google.com/a/answer/10741897?hl=en. (Accessed 9 November 2022).
[11] National Institute of Standards and Technology, Standards publication 197 advanced encryption standard (AES). https://doi.org/10.1201/9781439833032.ch89, 2001.
[12] P. Walters, The risks of using portable devices, 1–5, http://www.uscert.gov/reading_room/cyber_threats_to_mobile_phones.pdfandaboutprotectinglaptopsathttp://www.us-cert.gov/cas/tips/ST04-017.html, 2012.
[13] L.E. Alfonse, A.D. Garrett, D.S. Lun, K.R. Duffy, C.M. Grgicak, A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt, Forensic Sci. Int. Genet. 32 (2018) 62–70, https://doi.org/10.1016/j.fsigen.2017.10.006.
[14] ASCLD - validation & evaluation repository (n.d.), https://www.ascld.org/validation-evaluation-repository/. (Accessed 6 February 2023).
[15] National archive of criminal Justice data (NACJD) (n.d.), https://www.icpsr.umich.edu/web/pages/NACJD/index.html. (Accessed 22 November 2022).
[16] NIST data repository page (n.d.), https://data.nist.gov/sdp/#/search?q=topic.tag%3DForensics. (Accessed 23 November 2022).
[17] Open source forensic data: CSAFE forensic science dataset portal (n.d.), https://forensicstats.org/data/. (Accessed 23 November 2022).
[18] G. Fiumara, P. Flanagan, J. Grantham, B. Bandini, K. Ko, J. Libert, NIST special database 300: uncompressed plain and rolled images from fingerprint cards. https://doi.org/10.6028/NIST.TN.1993, 2018.

Michael A. Marciano*
Forensic & National Security Sciences Institute, Syracuse University, 100 College Place 120 Life Science Building, Syracuse, NY, 13244, USA

Henry P. Maynard III
American Society of Crime Laboratory Directors 5 Glen Road, Suite 123, Garner, NC, 27529, USA
E-mail address: HenryPMaynard@gmail.com.

* Corresponding author.
E-mail address: mamarcia@syr.edu (M.A. Marciano).