SWGDE Best Practices for Archiving Digital and Multimedia Evidence

Disclaimer:
As a condition to the use of this document and the information contained therein, the SWGDE requests notification by e-mail before or contemporaneous to the introduction of this document, or any portion thereof, as a marked exhibit offered for or moved into evidence in any judicial, administrative, legislative or adjudicatory hearing or other proceeding (including discovery proceedings) in the United States or any Foreign country. Such notification shall include: 1) the formal name of the proceeding, including docket number or similar identifier; 2) the name and location of the body conducting the hearing or proceeding; 3) subsequent to the use of this document in a formal proceeding please notify SWGDE as to its use and outcome; 4) the name, mailing address (if available) and contact information of the party offering or moving the document into evidence. Notifications should be sent to secretary@swgde.org.

It is the reader's responsibility to ensure they have the most current version of this document. It is recommended that previous versions be archived.

Redistribution Policy:
SWGDE grants permission for redistribution and use of all publicly posted documents created by SWGDE, provided that the following conditions are met:

1. Redistribution of documents or parts of documents must retain the SWGDE cover page containing the disclaimer.
2. Neither the name of SWGDE nor the names of contributors may be used to endorse or promote products derived from its documents.
3. Any reference or quote from a SWGDE document must include the version number (or create date) of the document and mention if the document is in a draft status.

Requests for Modification:
SWGDE encourages stakeholder participation in the preparation of documents. Suggestions for modifications are welcome and must be forwarded to the Secretary in writing at secretary@swgde.org. The following information is required as a part of the response:

a) Submitter's name
b) Affiliation (agency/organization)
c) Address
d) Telephone number and email address
e) Document title and version number
f) Change from (note document section number)
g) Change to (provide suggested text where appropriate; comments not including suggested text will not be considered)
h) Basis for change

Intellectual Property:

Unauthorized use of the SWGDE logo or documents without written permission from SWGDE is a violation of our intellectual property rights.

Individuals may not misstate or over represent duties and responsibilities of SWGDE work. This includes claiming oneself as a contributing member without actively participating in SWGDE meetings; claiming oneself as an officer of SWGDE without serving as such; claiming sole authorship of a document; use the SWGDE logo on any material or curriculum vitae.

Any mention of specific products within SWGDE documents is for informational purposes only; it does not imply a recommendation or endorsement by SWGDE.
.

Table of Contents

1. Purpose

The laboratory's role in a case does not end with the completion of a forensic examination of digital or multimedia evidence. The evidence and examination results may be needed for review or testimony years after completion of a case. Therefore, those materials need to be retained, protected, and available as needed.

The purpose of this document is to familiarize the reader with issues surrounding data archiving and suggest best practices for establishing and maintaining an archiving system. This document draws from the experience of the public archival community and aims to translate concepts and best practices developed by that community into the digital and multimedia evidence context.'

This document offers reference implementations to aid the reader in implementing an archival management system for digital and multimedia evidence at their organization *(*see *Reference Implementations)* and a discussion of the theory, technical factors, and other considerations involved in developing an effective archival management system.

2. Scope

This document provides basic information on data archiving for organizations performing digital and multimedia forensics. This document is intended as a guide; however, individuals should follow their organizational archiving and evidence retention policies. Presiding law may also govern requirements for retention and disposal. This document is intended to address the archiving of data related to digital forensic analysis that is not retained in another system of records.

3. Limitations

This document is not intended to address preservation or retention of agency records not related to digital and multimedia evidence.

4. Archive Management

4.1 Goals

The goal of archive management for digital and multimedia evidence is to ensure that data is available when needed in the future. The organization must determine to what extent they are required by law, policy, or mission needs to maintain or preserve digital evidence and derivative work product. These requirements may address what items need to be preserved and for what time periods they must be retained. The organization should use applicable retention requirements to develop their retention objectives.

Once these objectives have been set, the organization should determine what levels of risk they are willing to accept when selecting an archival management system, then weigh those risks

against the costs associated with implementation to select the appropriate solution. Generally, archival management systems with greater redundancy, more reliable storage media, and robust data management functionality are less prone to failure, but more expensive to acquire, implement, and maintain. The likelihood and impact of failure should be weighed against the cost of prevention.

4.2 Responsibility

The organization must make a determination regarding what entities within their organization will be responsible, administratively and financially, for building and maintaining the archiving system. Depending upon the structure of the organization, this may fall upon individual examiners in the field, a custodian in a centralized facility, or some combination thereof. These responsibilities must be clearly designated. To avoid confusion and risk potential corruption of the preserved information, there should be only one authoritative repository (archive) of archived digital and multimedia evidence in an organization.

4.3 Summary of Best Practices

1. Define in policy what data the organization requires to be archived and for how long it must be retained
2. Have a system to keep track of what is in the archive, where it is stored, and for validating its integrity
3. Choose storage that appropriately meets the organization's needs
4. Have redundancy, preferably geographically dispersed
5. Have policy, plans, and procedures for:
    a. Identifying personnel responsible for managing the archive
    b. Adding data to the archive
    c. Retrieving data from the archive
    d. Ensuring archived content will be accessible when needed for the full retention period
    e. Removing data from the archive when no longer needed

5. Discussion

5.1 Archive Management System Functions

An "archival management system" is a combination of personnel actions, tools, infrastructure, and policy. Management of a digital evidence archive is an active, ongoing process involving a set of policies, practices, procedures, and tools that collectively ensure archived information is preserved, safeguarded, and remains accessible and usable for its entire lifecycle, from acquisition to final disposition. Some of the system's functions are performed manually, others can be supported by tools. An effective archive management system supports these functions:

5.1.1 Indexing

An index enables an organization to know what is stored in its archive and locate data in that archive. Organizations should maintain an index of what data is in each archive, with sufficient information about the stored information to enable it to be readily located. In this context, "Index" likely won't mean a full content wordlist index of all data contained in the archive, but merely a descriptive catalog of the items archived. Organizational needs will dictate the level of detail and specific information elements indexed. In some cases, it may be sufficient to index stored information merely by a unique identifier (such as a case number) and a brief description.

Organizations should consider whether indexing additional information about the contents of data - e.g., a listing of files and directories in a disk image or phone numbers contained in a mobile device acquisition - would further organizational objectives and improve findability of data. Local policy and compliance requirements may impact what information may be indexed and how it may be queried. Indexing can be managed in simple tools such as spreadsheets, or more complex tools such as databases or evidence management systems.

5.1.2 Retrieval

Organizations must have the ability to retrieve archived data in a timely fashion. Data might not always be directly retrievable from the management system; it could be stored offline on optical media, hard drives, or Linear Tape-Open (LTO) data tape. When data is stored offline, the management system must include the information about the data and where it is stored, so the data may be retrieved as soon as possible after a request. Access control and security protocols must be included in the index record so the access protocols are clear.

5.1.3 Record Keeping, Audit Trail, and Provenance

An effective archive management system relies on the organization collecting and maintaining the information needed to carry out the archive management functions. It typically includes, for each archived item:

- Original source of the archived data;
- Circumstances surrounding the collection or generation of the archived data, including who, what, when, where, why, and how it was collected;
- Authorities for the collection, maintenance, and use of the archived data;
- Physical and logical location of the archived data;
- Index information (see *Indexing section*);
- Digital formats and physical media types used, including file system of the media;
- Software, operating system, hardware used, and legacy hardware dependencies in order to open and render the archived data;
- Fixity information, such as hash values and hash algorithm used (e.g., MD5, SHA-3);

- Record of actions taken with the archived data, including fixity checks, transcoding, migration between physical media, and retrieval;
- Relationships between items (e.g., item B is a transcoded version of item A);
- Access history (who retrieved, when, for what purpose); and
- Disposition of the archived data and storage media, if applicable.

5.1.4 Unique Identifiers (UIDs)

To execute archival management functions, an organization must be able to uniquely identify the information stored in an archive, the media containing the archived information, and the subject matter to which they pertain (e.g., a particular investigation or forensic examination). The organization must establish conventions for the types of unique identifiers needed to implement their archive management system. These unique identifiers may be a combination of existing identifiers (e.g., a case number, the make, model, and serial number of a hard drive containing digital evidence, an evidence ID number from an evidence management system) or specific to the archive management system.

5.1.5 Data Obsolescence Mitigation

For an archive to be effective, organizations must maintain means to access preserved data for the duration of the retention period. As preserved digital evidence ages, the software and hardware needed to read the preserved data may become obsolete and not readily available. In addition, acquired data could be received in an already obsolete format. To ensure continued access to preserved data, organizations should develop policies and procedures for monitoring for the obsolescence of formats of archived data. In order to perform this monitoring, information on the software, operating system, and hardware dependencies of the data must be included in the archive management system.

When obsolescence approaches, or if the data is obsolete when acquired, organizations have several options:

- Archive the software, hardware and legacy dependencies, documentation and procedures necessary to access the archived data.
- Transcode the archived data into a non-obsolete format, while maintaining the original and ensuring, to the extent possible, the transcoding process does not result in a loss of information. Organizations transcoding archived data should strongly consider transcoding into an *open* format - a format for which the specifications are published and freely available and has no restrictions on the use or implementation of the format - to avoid the need to repeatedly transcode in the future. Technical characteristics about the original file must be included in the metadata record. The management system must provide a clear link between the original and the transcoded file, describing the creation process of the transcoded file.

- If neither archiving the hardware and software needed to access the archived data nor transcoding are possible, organizations should continue to maintain a bit-for-bit copy of the data and conduct routine fixity checks, should a solution become available in the future.
- When archiving data stored in proprietary formats, organizations should give strong consideration to also archiving the software needed to read that format or to transcode the archived data using a preferred format listed in the Library of Congress' Recommended Formats Statement [http://www.loc.gov/preservation/resources/rfs/index.html]. The Library of Congress' Sustainability of Digital Formats website [http://www.loc.gov/preservation/digital/formats/index.html] discusses obsolescence factors in general and for specific media formats.

5.1.6 Storage Media Migration

The media on which the data is stored should also be migrated over time. As described in the "Comparison of Physical Storage Options" section, each physical storage option has specific characteristics, including speed and life expectancy, that impact its suitability for particular frequencies of retrievals and retention periods. Every storage option will eventually become obsolete and fail. For data that must be retained more than five years, media migration is unavoidable.

Organizations using multiple physical storage options should develop policies and procedures regarding when to migrate archived data between particular physical storage options and actively monitor for these triggers. These procedures for migrating data between particular tiers of storage, whether manual or automated, are known as hierarchical storage management (HSM). HSM is usually defined as having three tiers: online (data that must be immediately accessible so is stored on servers or computers); nearline (data that the user can wait a few minutes to retrieve, this is usually stored on data tape accessed by a robotic arm or on attached hard drives that must be powered up); and offline (data stored offline on detachable media so must be manually retrieved). Data often is moved through the three tiers as the frequency of its use decreases.

Potential policy triggers for migrating archived data include:

- Case lifecycle events (e.g., digital evidence is moved to tape following initial adjudication of case);
- The last use of that particular digital evidence item (e.g., a digital evidence item is written to tape 18 months after the last time that particular item is used); and
- Frequency of retrieval from a "lower" tier of storage (e.g., all data for a particular case is retrieved from tape when any data from a particular case is retrieved from tape).

Along with policy triggers, storage media must be refreshed due to media obsolescence, infrastructure upgrades, or hardware end-of-life. Organizations should develop policies and procedures to monitor the viability of the physical media used and migrate archived data to new media when the physical media approaches the end of its lifetime or failure appears imminent.

As with all preservation actions, the migration actions should be noted in the archive management system.

Media migration costs can be anticipated. If the organization prefers migrating in-house, the appropriate upgraded hardware and media costs can be budgeted. If media migration cannot be performed in-house due to infrastructure limitations, service providers can be used.

5.1.7 Fixity Checking

Fixity checking is the act of verifying, generally through the use of checksum or hash function, that particular information has not changed. In the digital and multimedia evidence context, computing and verifying the hash value of a file is an example of a fixity check. Fixity checking should occur, at a minimum, on ingest or receipt of information, on retrieval, and when information is transferred between media or systems.

Files should also be checked after events that could potentially adversely impact the storage media; for example, extreme temperature or environment fluctuations, fire or smoke in the immediate area, water damage (sprinklers or flood), or the media has been dropped. Files should also be checked if a particular lot or batch of storage media is known to be prone to failure. This latter example implies the batch number is included in the metadata or catalog record for the storage media. Files stored on live media, such as storage servers, should be checked after any event likely to affect the integrity of the storage platform, such as power failures, disk replacements, or array rebuilds. Hosted storage providers usually do not offer a means to run fixity checks in-situ on data stored on their servers.

An organization may also choose to run fixity checks at a recurring interval if the data has a long-term retention schedule or is on media prone to failure (optical media or external hard drives). Organizations must determine their fixity check interval based on factors including the stability and failure rates of the media involved, volume of data, and desired level of information assurance. Organizations should implement fixity checking by, at a minimum, computing a cryptographic hash value for the stored information using a NIST-approved Hash Algorithm[1].

Organizations should develop policies and procedures to document and store the fixity information (e.g., cryptographic hash value and hash algorithm used) needed to run fixity checks, document the outcome of fixity checks, and handle and recover from fixity check failures.

The National Digital Stewardship Alliance (NDSA), an organization of federal, state, and public entities involved in digital preservation activities, has created a report with additional information and guidance on implementing fixity checking. See NDSA, *Checking your Digital*

[1] Currently, Federal Information Processing Standard (FIPS) 180-4, Secure Hash Standard, and FIPS 202, SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions, specify the NIST-approved hash algorithms. See https://csrc.nist.gov/Projects/Hash-Functions for additional information.

Content [http://www.digitalpreservation.gov/documents/NDSA-Fixity-Guidance-Report-final100214.pdf.](http://www.digitalpreservation.gov/documents/NDSA-Fixity-Guidance-Report-final100214.pdf)

5.1.8 Redundancy and Geographic Dispersal

Having only one copy of data is high risk. If the file becomes corrupt or its storage media fails, the content is lost. Archival standards recommend storing three copies, ideally on different media types (e.g., a mix of HDD and tape), which are dispersed geographically. If three copies are not possible for financial or management reasons, no fewer than two should be maintained. Redundant disk arrays, while providing parity to protect against media failure, do not satisfy this requirement.

Storing multiple copies impacts scheduling fixity checks, especially if the copies are geographically dispersed. The organization might decide to only check the local copy, if it verifies the files were correctly written before sending to other locations.

Copies should be geographically dispersed to regions with different potential disaster factors (e.g., earthquake, hurricane or terrorism). If the organization does not have a presence in different regions, partnerships can be formed with organizations in different regions that can securely store redundant copies. The management system must enable tracking locations of data in the custody of other organizations. Organizations should ensure that appropriate security and access controls are implemented to protect their data in the custody of others.

5.1.9 Security

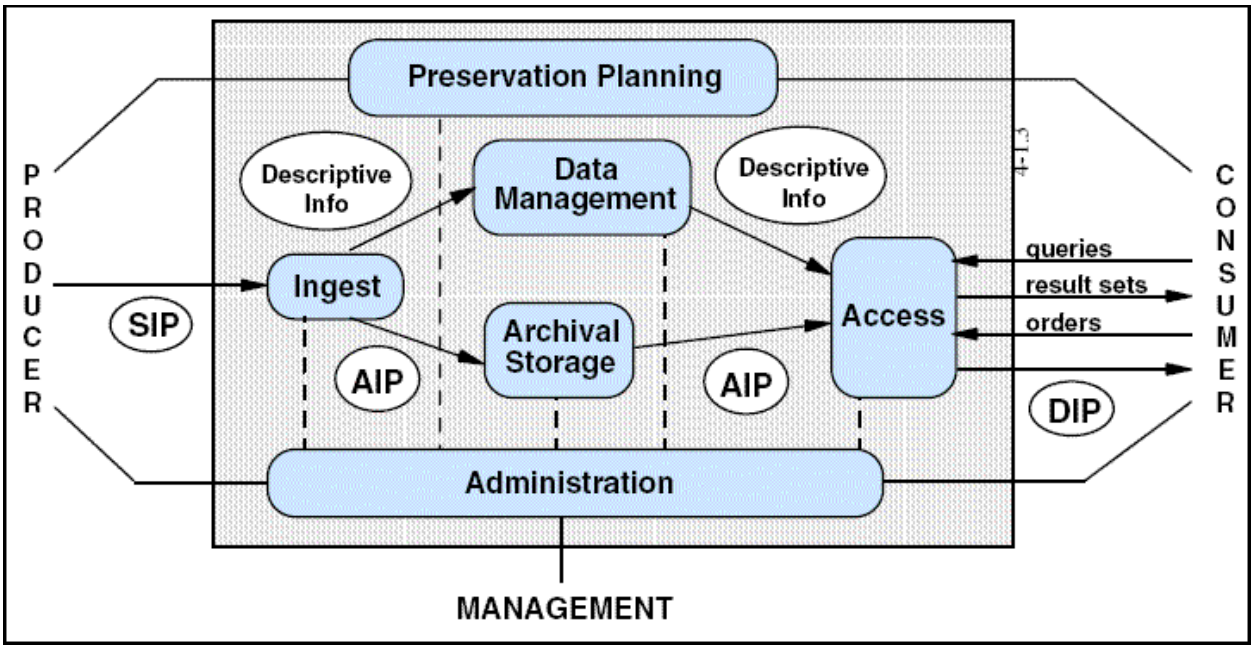
Whether the archive is maintained locally or in hosted storage, organizations should take the necessary steps to ensure the security of the archived data in accordance with law and policy. Encryption is one way of securing archived data. Additional documentation needs to be maintained for the encrypted data and systems, including the type of encryption, encryption keys or passphrases, and procedures for encrypting and decrypting the data. This documentation should be maintained following the same guidelines set forth herein regarding archived data. Organizations archiving contraband or other sensitive data should ensure it is appropriately labeled and secured in accordance with law and policy.

6. The Archive Management System Model

The public archival community follows a standard conceptual model for implementing a system that provides digital preservation functions: the Open Archival Information System (OAIS) model, ISO 14721.[2] This standard was originally developed by NASA and organizations sharing space data, but has been adopted as a reference model by organizations around the world. The OAIS model divides the archival process into specific functional components.

[2] https://public.ccsds.org/pubs/650x0m2.pdf

This is a reference model, making it applicable across all disciplines. It does not specify systems or procedures, but focuses on functions and concepts. It can be adapted to archiving digital evidence workflows, whether the functions are performed by one person documenting processes with spreadsheets or a large team using an enterprise management system.



Credit: Open Archival Information System (OAIS) Reference Model (ISO 14721)

6.1 OAIS taxonomy

While the OAIS model uses a distinct taxonomy, its concepts translate to digital and multimedia archiving contexts. In OAIS' taxonomy:

- The ***Item*** is the information being archived.
- The ***Producer*** is the person or entity generating data to be archived, like the forensic examiner or their organization.
- The ***Consumer*** is the person or entity retrieving or using data retrieved from the archive.
- ***Content Information*** is the actual data or thing to be archived. In this context, it is the digital or multimedia evidence item being archived.
- ***Preservation Description Information (PDI)*** is metadata needed to ingest the archived item into the archive, preserve it for the retention period, locate it when needed, and satisfy compliance requirements. See the *Record Keeping, Audit Trail, and Provenance* section for examples of what this might include in a digital and multimedia evidence context.

- OAIS archives receive, store, manage, and return to consumers *Information Packages*, which are containers with both the Content Information and PDI. OAIS defines three types of Information Packages:
    - The ***Submission Information Package (SIP)*** is what a Producer submits to the archive. It contains the Content Information and whatever PDI the archive needs to ingest the data. For digital and multimedia evidence, this might include the evidence file, a case number, a description of the item, and the acquisition information. Organizations internally define what information is necessary to include in a SIP.
    - The archive stores and manages the ***Archival Information Package (AIP)***, which contains the Content Information and whatever PDI the archive needs to manage the archived items and make it retrievable. This may include PDI submitted in the SIP, information the archive generates incident to accepting the submission, like a listing of files contained in a forensic image file, or information the archive generates incident to managing the item, such as a record of retrievals or fixity checks.
    - The archive returns ***Dissemination Information Packages (DIP)*** to consumers, which contain the Content Information and whatever information is necessary for the Consumer to use the archived item, including technical information and information needed to satisfy legal and compliance requirements.

An OAIS archive has six functional entities to manage the flow of information between producers, the archive, and consumers.[3]

- The ***Ingest function*** receives data to be preserved and prepares it for storage and management in the archive. It receives SIPs, generates AIPs from them, and provides the AIPs to the Storage function. The Ingest function verifies appropriate metadata accompanies the data, provides this metadata to the Data Management function, extracts other technical metadata that may be needed to maintain the archive, and may conduct a fixity check.
- The ***Storage function*** stores, maintains, and retrieves archived data. It receives AIPs from the Ingest function, places them onto appropriate storage media, migrates them between storage media, and retrieves stored data in response to retrieval requests from Consumers. The Storage function also includes disaster recovery planning.
- The ***Data Management*** function maintains an authoritative repository of data about the contents of the archive's collection, including metadata, the index, data needed to monitor for obsolescence, and other information needed to administer the archive. It provides access to this data to the other functions and for reporting. When items are ingested, the Data Management function receives and stores some of the metadata from the

[3] "Reference Model for an Open Archival Information System ... - CCSDS." 2 Jun. 2012, https://public.ccsds.org/pubs/650x0m2.pdf. Accessed 9 Jan. 2018.

accompanying PDI where it is accessible without retrieving the related Content Information.

- The ***Administration function*** manages the day-to-day operation of the archive. This function also serves as an interface between the archive, management, and consumers.
- The ***Preservation Planning*** function conducts the monitoring, analysis, and planning needed to ensure the archive's contents remain usable for the duration of their retention and that the archive system continues to meet the evolving needs of the organization. These functions include evaluating archive contents to identify needs for updates, monitoring changes in technology potentially leading to obsolescence, identifying changes in the archival needs of the organization, planning for migration of archived information between formats and media, policy updates, risk analysis, and adoption of advancing archival technologies.
- The ***Access function*** provides an interface for Consumers to locate, request, and retrieve data from the archive.

6.2 Management Systems

Archive management systems enable organizations to operate their archive. There are several types of data management systems that each have their own purpose. Organizations might find they only use one system, or combine systems to implement a full OAIS-compliant workflow.

Collection Management System (CMS). As the simplest data management system, a CMS is similar to a traditional evidence tracking system in that it does not hold evidence, only records regarding the evidence. A CMS contains a listing of, and metadata describing the data stored in an archive as well as records of where and how the data are stored. As a CMS is separate from and not linked to the data it references or the data storage facility, the chance for orphaned or lost data is greater with this system than the other more complex types of data management systems.

Digital Asset Management System (DAMS). This system is used for managing data to ensure the owner of the data can use it. The DAMS is often utilized from the start of the data creation through usability and ends prior to the destruction of the data. The metadata of the data contained within the DAMS is not as extensive as other systems, but only needed to allow the user to access the data. Files are ingested into a DAMS, which extracts technical metadata from the files and can create checksums upon ingest.

Digital Preservation Systems (DPS). A DPS is an all-in-one archive creation and management system. Digital preservation systems are built to receive SIPs and create AIPs. (ref. OAIS) They extract more extensive technical metadata, and sometimes embedded metadata. They create checksums, can perform fixity checks on the assets over time, and track where the data is stored. Ideally, a DPS manages information on the original and current environment for rendering the object (including operating and file system information), and on the storage environment and location.

7. Comparison of Physical Storage Options

All data is ultimately stored on physical media. Each type of media has strengths and weaknesses and different relative cost. When considering the cost of storage media, one must consider the lifecycle cost of the overall storage environment, not only the media. For example, LTO tape requires at least one drive for writing and playback; storage servers require power and cooling; hosted storage incurs recurring service fees.

Additionally, all physical storage media will eventually be retired from service and replaced with new media, as frequently as every five years. However, physical media replacement costs are commonly lower than the labor cost of staff managing migration.

Storing data on media requires selecting a file system and software for writing the data to the media. For example, hard drives must be formatted with a file system before data is written to them (e.g., NTFS, ZFS, exFAT). LTO tapes are initialized LTFS. The file system determines how the data is retrieved from the media. The system should be open (not proprietary or using a proprietary backup software) so there will not be problems accessing the data in the future. The file system selected can also depend on the operating system (OS) of the computer that will be used to access the files. When restoring archived data, it should be restored to the operating system and file system from which it was created, to avoid changing metadata. For example, if the data was created in a linux environment and then restored to a Windows environment, there is the potential for changing the metadata of the data because of the differing metadata schema of the file systems.

Many of the storage media options described below are detachable media, making them at greater risk for theft, loss, and accidental destruction. Conversely, if the media are kept detached from network storage, the data on them can potentially be more secure than files stored online. Detachable media should be stored in a locked environment with limited and controlled access.

This section offers recommendations for the type of storage media for a particular storage objective, summarized in the table below, and a discussion of the factors motivating these recommendations:

Storage Media	**Recommended Use**
Optical media (CD-R/DVD-R, Blu Ray)	For small-size files with a retention period from years to decades
External hard drives	For archives up to several terabytes, with a retention period less than 3 years

Solid State Drives (SSDs)	Not recommended for archival purposes
Data tape	For long-term storage of infrequently used data
Servers with either Network Attached Storage(NAS) or Direct Attached Storage(DAS)	For large data volumes or frequent access requirements
Hosted storage	Varies

7.1 Optical media (CD-R/DVD-R, Blu Ray)
Recommended use: For small-size files with a retention period from years to decades.

Optical media is the least expensive storage option at the per piece level. Its ease in management, storage, and use makes it appealing for storing data. However, it has long-term archiving issues.

As a physical format, optical media has high obsolescence probability as manufacturers have already begun decreasing support for recorders, players, and blank media. Thus, it should not be used as a long-term archiving medium.

Optical media also has significant storage capacity restrictions requiring large datasets to be written across multiple disks, thus increasing the probability of failure. The medium's portability increases its security risk, so media should be stored in locked areas with limited access.

Environment: CD-R and DVD-R media have varying life expectancy (LE) depending on these factors:

- Metal;
- Dye (CD-R and DVD-R);
- Sleeve and case material;
- Storage environment; and
- Handling by humans.

Blu-ray disc media are a different construction than CD-R and DVD-R media. The laser is blue-violet (higher frequency) rather than the red used in DVD-Rs or infrared used in CD-Rs (hence the name "Blu-Ray"). The dye is inorganic, and is not affected by UV light. BD-R LTH (**B**lu-**R**ay **R**ecordable **L**ow **T**o **H**igh) uses organic dye. BD-R LTH discs should not be used for storing data.

Blu-ray discs have an unknown LE, so the most likely determining factor affecting data access will be obsolescence. As the format has not had as broad consumer adoption as DVD-R, disc drive support may end before the discs themselves deteriorate.

Optical media with the longest LE are those made with gold and, for CD-R and DVD-R discs, phthalocyanine dye. These formulations, if stored in cool and dry conditions, can have a LE of up to 100 years. The availability of readers (drives) to play this media in 100 years is questionable, and as with all media, this risk should be mitigated by an appropriately planned migration. See http://www.loc.gov/preservation/scientists/projects/cd-r_dvd-r_rw_longevity.html

Because optical media can potentially have a short LE (depending on the metal and dye, and the storage environment), any data stored on it with a retention period of over five years should have an annual fixity check, and a planned migration schedule.

Storage: See this guide for recommendations on handling optical media: (*Fred Byers. Care and Handling of CDs and DVDs* at https://www.clir.org/pubs/reports/pub121/pub121.pdf). In general:

- Do not write directly on the media recording area (top or bottom sides);
- Do not write on the media with a pen;
- Only write in the inner plastic hub using a water soluble archival pen;
- Write the unique identifier on the hub;
- Do not affix labels to the top label side;
- Use inert polypropylene jewel cases;
- Remove any paper inserts from inside the case;
- Store upright like books; and
- Store the discs in polypropylene jewel cases in archival boxes, away from light.

7.2 Hard Disk Drives (HDDs)
Recommended use: For files with a retention period less than 3 years.

Their relatively low cost, high portability, and ease of use make hard disk drives (HDDs) such as self-contained drive enclosures with an interface such as USB, or bare drives placed in a drive dock, an appealing option as a short term storage medium. However, these devices are fragile and prone to fail without warning. This medium's unreliability makes it unsuitable for long-term archival storage.

If HDDs are considered, high quality drives are recommended. Hard drive manufacturers specify different classes of drives based on performance, reliability, and intended use. Bare drives marketed for enterprise network applications are generally the most reliable.

If HDDs are the selected storage medium, the organization must have redundancy in case of failure.

A media migration schedule must be followed. HDDs should be swapped out at least every 3 years. Fixity checks should be performed on the migrated files and tracked in the management system along with a record of the migration.

Before writing data to a new HDD, the HDD must be formatted with a file system. Organizations should use an open, or commonly used, well-documented file system supported by the computer systems used to access it.

Environment: HDDs may be stored in a typical office environment. They should not be stored or operated at extremely low temperatures, in extreme heat, or in high-humidity environments. HDDs may fail in ambient temperatures beyond the manufacturer's operating specifications. If left running without adequate ventilation, overheating can cause delamination of the drive's read/write heads. HDDs should not be stored near sources emitting strong vibrations which can cause the drives' head stack assembly to come in contact with the surface of the data platters (crash) and cause irreparable damage. Strong magnetic fields can corrupt data on HDDs. Bare drives should be stored in cases to protect from dust and mishandling.

7.3 Solid State Drives (SSDs)

Recommended use: Not recommended for archival storage

Solid state storage technology can be manifested in a medium as simple as a flash thumb drive, or as a solid state drive (SSD). SSDs store data differently than mechanical drives by storing data in electronic gates. Research shows these gates fail to hold their electrical charges over time if not kept powered periodically.

Due to uncertainties regarding the demonstrated longevity of data storage on unpowered SSDs they are not recommended for long-term storage.

7.4 Data tape

Recommended use: Long-term storage of infrequently used data

Data tape is a stable and relatively low cost storage medium. While best known as part of a systems back-up infrastructure, tape can also be used in standalone tape drives for storing files.

The open file system Linear Tape File System (LTFS) should be used with data tape. The archive management system should track which version LTFS is used when writing files to tape. LTFS is now an ISO standard, ISO/IEC 20919.

Manufacturers estimate their products have an LE of 15-30 years. Data should be migrated to new media well before the manufacturer's expected end of life of the previous media. Additionally, regardless of LE, data should be migrated every two or three generations of Linear Tape-Open (LTO) drive, as drives for older generations may become difficult to obtain.

Migrating data from data tape to a newer generation can be labor-intensive. The data must be copied off the tape to other storage (HDD, server), a fixity check run to be sure the files are healthy, and the files copied to the target generation tape. The files must then be restored from at least one copy (since more than one copy should be made for redundancy) to verify they were correctly written. Policies should address the disposition of the older tapes.

Environment: Tapes should be stored standing upright in their original plastic jewel cases in a climate controlled environment.

7.5 Servers (online)
Recommended use: large data volumes or data with somewhat frequent access requirements

In this context *online* refers to powered-on, running storage facilities, be they locally or remotely hosted. Networked storage must have redundancy, whether on another storage environment or backed up to other storage media. Redundancy requires managing the dispersed copies and knowing their locations. Utilizing RAID or object storage does not constitute geographic redundancy.

Networked storage requires sophisticated access controls to ensure the confidentiality and integrity of the archive.

Networked storage is a relatively expensive storage solution due to the hardware, maintenance, electrical, and climate control costs. Servers and associated hardware require periodic upgrades within manufacturers' recommended lifecycles.

7.6 Hosted storage
Recommended use: varies

Hosted storage (also: Storage As A Service, cloud storage) entails storing data on an external third party's servers. Generally, the service provider provides storage and maintains the infrastructure and the organization is responsible for uploading and managing the data. Hosted storage is a viable option for archiving files of any size or retention period, but requires careful planning, understanding of the technology, and ongoing management to ensure a successful implementation. The falling cost of hosted storage is making it an attractive archiving option for many organizations.

Organizations should evaluate hosted storage providers carefully, considering their product offerings, operational and security practices, experience, reputation, and cost. Existing vehicles, including the ISO/IEC 27001:2013 information security management standard and the U.S. federal government's Federal Risk and Authorization Management Program (FedRAMP), may assist organizations in evaluating some of these factors.

Organizations should consider the logistics of storing data in and retrieving it from a hosted storage provider. Organizations should evaluate whether their existing internet connection is sufficient for uploading and downloading their anticipated volume of digital and multimedia evidence to and from a storage provider. Some providers offer offline data transfer services - e.g., a hard drive mailed to provider, who will then locally upload the data to their servers - which may be appropriate for uploads or retrievals of large volumes of digital or multimedia evidence.

Organizations must ensure - technically and contractually - there are appropriate controls to prevent and detect unauthorized access to or loss of their data in hosted storage. While a detailed discussion of securing hosted storage environments is beyond the scope of this paper, this will typically include:

- Implementing multi-factor authentication for access to the hosted storage;
- Configuring appropriate permissions on hosted storage and routinely auditing these permissions;
- Ensuring the provider's audit logging services are enabled, retaining these logs for an appropriate period, and reviewing them periodically; and
- Considering encrypting data stored in the hosted using keys under the control of the organization but not the provider.

As with other storage options, redundancy is a best practice with hosted storage. Many hosted storage providers have redundant storage offerings, allowing customers to store multiple copies of their data within the same or multiple geographic regions. These redundancy options are typically more expensive than non-redundant storage. Organizations should store multiple copies of their data in different geographic regions.

Many hosted storage providers have multiple tiers of storage offerings, based on the speed and expected frequency of data retrieval. Organizations should determine which tiers best meet their needs and consider using a combination of storage tiers.

Hosted storage involves a recurring cost to the organization. Organizations must ensure these costs are included in their budget planning. Hosted storage costs will increase with the volume of stored data, and organizations may incur additional costs for uploading and retrieving data.

Fixity checking practices vary by provider. Organizations should identify the fixity checking practices of potential providers and assess whether they are sufficient for the organization's risk tolerance. Fixity checking should occur, at a minimum, when data is stored in or retrieved from the hosted storage. Many providers have capabilities in their upload functions to perform fixity checks as part of the upload process; organizations should use these capabilities when they exist. For many providers, it may not be possible for the organization to conduct in-situ fixity checks of data stored in the provider's environment; the organization may be required to retrieve the data from the hosted storage to perform a fixity check and should consider the potential logistic challenges and costs of this operation.

Limits on the size of data objects stored vary by provider. Organizations should determine whether a provider's size limits meet their needs or whether archived data can be segmented into chunks within the size limit.

Organizations should research the process for migrating their data to another system or provider. Sometimes there are hidden costs and difficulties when exiting hosting storage providers. Organizations should assume their data will be migrated in the future and begin planning for this when they first consider using hosted storage.

8. Reference Implementations

Below are examples of how differing organizations may choose to organize a system for archiving digital evidence, forensic work products, and ancillary information. This is not meant to be prescriptive, but informative. The examples below may be implemented as part of, in conjunction with, or entirely separate from an investigative case management system.

8.1 Example 1

Small laboratory. One person managing archiving processes. No dedicated digital preservation system or tools to manage digital evidence; most operations are performed manually. Evidence stored on detachable media.

Requirements: Preservation of closed forensic examination products (reports, notes, data extractions, forensic images, and copies of specialized software or programs used for the analysis) for a few years after case closure.

Responsibilities: Examiner(s) serve most if not all of the archiving functions. They create the Submission Information Package (SIP), Archive Information Package (AIP), and Dissemination Information Package (DIP) (see *Taxonomy*).

Process: At the completion of each forensic exam, the examiner gathers all the related records from the particular case into a single directory tree. The examiner then compiles a spreadsheet listing the records with descriptions of their contents, relevant metadata as detailed in "Record Keeping, Audit Trail, and Provenance," and checksums for all files. This directory is then compressed, hashed, and copied to external storage devices in duplicate and re-hashed. One copy is stored locally in a cabinet, a second copy is stored in a different location, preferably offsite.

The compiled spreadsheet is stored on a backed-up directory and an entry is added to a ledger listing all of the archived examinations, container identification, storage location, and hashes. Backup copies of the ledger are also maintained.

The examiner is responsible for conducting periodic audits of the archives and purging items older than the specified retention period. The examiner will migrate data to new storage (and perform fixity checking) as indicated by the migration schedule for the destination media.

Someone in the examiner's organization (possibly the examiner) is assigned the responsibility for responding to retrieval requests, finding respondent records from the ledger/index, locating the stored files, extracting and delivering copies, running fixity checks when scheduled, migrating storage media, and tracking all actions.

8.2 Example 2

Multi-examiner laboratory. More than one person managing processes. Has a dedicated digital preservation system or tools to manage digital evidence; operations are performed manually or automatically. Evidence stored on NAS with redundancy.

Requirements: Preservation of closed forensic examination products (reports, notes, data extractions, forensic images, and copies of specialized software or programs used in the analysis) for a time consistent with the organization's policy case closure.

Responsibilities: Examiner(s) serve most if not all of the archiving functions. They create the SIP, AIP, and DIP.

Process: At the completion of each forensic exam, the examiner gathers all the related records from the particular case into a single directory tree. The examiner then compiles a spreadsheet listing the records with descriptions of their contents, relevant metadata as detailed in "Record Keeping, Audit Trail, and Provenance," and checksums for all files. This directory is then compressed and moved to an archive storage server. A copy of the archive is automatically uploaded to inexpensive 'deep archive' hosted storage. The archives on the server are backed up to tape in duplicate and stored offsite.

The compiled spreadsheet is stored on a backed-up directory and an entry is added to a ledger listing all of the archived examinations, container identification, and storage location. Backup copies of the ledger are also maintained.

Someone in the examiner's organization (possibly the examiner) is assigned the responsibility for conducting regular periodic audits of the server and tape archives and purging items older than the specified retention period, and migrating archives to new media as indicated by the migration schedule.

Someone in the examiner's organization (possibly an examiner) is assigned the responsibility for responding to retrieval requests, finding respondent records from the ledger/index, locating the stored files, extracting and delivering copies, running fixity checks when scheduled, migrating storage media, and tracking all actions.

8.3 Example 3

Large laboratory with personnel dedicated to perform archive management functions. Enterprise class storage with automated functionalities.

Requirements: Preservation of closed forensic examination products (reports, notes, data extractions, and forensic images) for a variable number of years after case closure, with some cases retained for decades.

Responsibilities: Examiners and dedicated IT personnel work in conjunction to perform the archiving functions. IT Personnel create the SIP, AIP, and DIP.

Process: At the completion of each forensic exam, the examiner gathers all the related records from the particular case into a single directory tree. The examiner then compiles a spreadsheet listing the records with descriptions of their contents, relevant metadata as detailed in "Record Keeping, Audit Trail, and Provenance," and checksums for all files.

SIP creation:
Examiner compiles directory of items to be archived and information regarding it, then identifies it to a person, process, or utility that conducts the ingest procedure and compiles metadata about the contents.

AIP creation:
Someone in the examiner's organization, manually or using an archive management tool, collects the identified data and information and copies it to an archive server (or hosted storage). Automated processes create regular periodic tape backups of the storage archive and the tapes are stored offsite.

DIP creation:
Someone in the examiner's organization is assigned the responsibility for responding to retrieval requests, finding respondent records from the ledger/index, locating the stored files, extracting and delivering copies, running fixity checks when scheduled, migrating storage media, and tracking all actions. Planning for data archival hardware and software upgrades and obsolescence.

Someone in the examiner's organization is assigned the responsibility for conducting periodic audits of the archives and purging items older than the specified retention period.

History

Revision	Issue Date	Section	History
1.0 DRAFT	2019-06-06	All	Initial draft created and voted by SWGDE for release as a Draft for Public Comment.
1.0	2020-09-17	--	Voted for release as Final Publication