DATA GOVERNANCE AND
MANAGEMENT TOOLKIT

January, 2021
Version 1

# Data Archiving

## Table of Contents

## INTRODUCTION

Data archiving refers to the long-term storage of data that are no longer needed for immediate access. Archives are stored separately from the primary system in a system that is designed for long term retention. Data archiving focuses on both maintaining the safety of the data and integrity of the data. This includes producing trustworthy metadata for archives (descriptors for what has been archived). Data archiving applies to both physical and digital data assets. Physical assets refer to physical items in hand such as equipment or hard copies of documents. Digital assets refer to data that is stored on a device. This document primarily focuses on digital assets.

## KEY POINTS IN THIS DOCUMENT

- Data archiving refers to the long-term storage of data that are no longer needed for immediate access.
- Archives are intended to help manage space requirements and tend to be more cost effective than storing data on the primary system.
- Your data archiving policy should address:
    - What data will be archived?
    - How will the data be documented?
    - How and where will the archives be stored?
    - How will the data be protected to ensure the privacy and security of the data will be maintained?
    - How will requests to changes in archived data be handled?
    - How will archived data be destroyed when it is no longer needed?

## WHY ARCHIVE DATA?

Archives are intended to help manage space requirements. They tend to be more cost effective than storing data on the primary system. As data sets grow in size, the management of large data sets becomes complex. Access requirements for certain data can also decrease in frequency. In some cases, data retention may be required for legal or organizational requirements.

## DATA ARCHIVING POLICY

Data archives tend to be read-only as they are long term records that do not change. This creates unique challenges as the data selected for archiving must have a low chance of being used or changed in the future.

For data archives to remain organizationally useful, processes and documentation are required. Processes and documentation ensure that data can be appropriately indexed, searched and retrieved for use as needed.

The checklist in table 1 below establishes key items to consider in your data archiving policy.

TABLE 1: DATA ARCHIVING POLICY CHECKLIST

| | |
|---|---|
| What data is going to be archived?<br>(See Preparing Data for Archiving) | ✓ Select data for archiving<br>✓ Select appropriate data formats |
| How will the data be described so that it can be used and found?<br>(See Documentation of Metadata in the downloadable document “Data Quality”) | ✓ Develop and maintain metadata describing the data<br>✓ Establish a location for metadata |
| How and where will the archives be labelled and stored?<br>(See Archive Options and Archive Documentation, Labelling and Storage) | ✓ Select the type of archiving media and storage solution<br>✓ Select a storage strategy location<br>✓ Select an appropriate labelling media and strategy |
| How will the data be protected to ensure the privacy and security of the data will be maintained?<br>(See Preparing Data for Archiving and Accessing Archived Data) | ✓ Select appropriate encryption strategy for archive<br>✓ Establish access rules for archives<br>✓ Establish a logging and audit policy for data access requests |
| How will requests to changes in archived data be handled?<br>(See Supporting Data Change Requests) | ✓ Develop a data change/deletion policy<br>✓ Establish a logging and audit policy for data change requests |

| | |
|---|---|
| How will archived data be destroyed when it is no longer needed?<br>(See Physical Data Destruction) | ✓ Develop an archive destruction policy<br>✓ Establish a logging and audit policy for archive destruction |

## Preparing data for archiving

It is critical to store all data in a standard format, rather than a format that is software-dependent. File formats that are platform or software-dependent may become obsolete over time. Examples of standard file formats include HTML, PDF, and CSV Open standards will ensure that data will be accessible in the future.

Once data has been selected for archiving, the privacy and security of the data must be considered. The data should be encrypted to ensure that it cannot be accessed in an unauthorized fashion. An appropriate encryption algorithm needs to be selected. Data may be encrypted on a record or document level, or the entire media may be encrypted. Encryption keys must be kept in a secure local. Loss of an encryption key will mean that access to the archived data is not possible.

When data is selected for archiving, it must be made clear in the primary system that the items have been archived.

## Archive options

There are different options for how and where to store your archived data.

Your government will need to consider the required archival life span of the data when choosing an archival media. The archive media you choose must be stable enough that data will not degrade within the needed lifespan. Some media formats will physically degrade with time, which means that after a given period of time, data stored on the device will be corrupt or unusable. Organizations have lost important information due to this type of failure by selecting low quality, non-archival grade media. Archival-grade media are designed to be resistant to this failure for many years.

Table 2 describes different archive media options, and factors to consider when choosing media type. The size, desired retention period, likelihood of needing to access your data, and costs are all considerations in the choice of your archive solution. Key concepts to compare across options include:

- **Write-once/read-many:** This means that once the data is written to the disc, it cannot be changed in place.
- **Density:** Volume of data supported.
- **Retention Period:** Length of time data can be stored before risk of data loss or corruption. Regardless of the choice of physical media selected for archiving, the environmental conditions of the storage will also impact the data retention lifetime.

TABLE 2: ARCHIVAL MEDIA OPTIONS

| Media | Data Retention Period | Density | Access | Costs | Other Factors/considerations |
|---|---|---|---|---|---|
| **Mechanical Hard Disk**: uses a magnetic field to store data on a spinning disk.<br>e.g., desktop computers and laptops that use hard disk drives (HDD) | 3 - 5 years | Medium | Medium | Low | • The magnetic field will degrade over time leading to a potential loss in data integrity.<br>• Utilize moving mechanical parts that can lead to failure and the loss of availability of the data. |
| **Solid State Storage:** uses non-moving storage cells to store data.<br>e.g., memory card, USB flash drive, solid state drive (SSD) laptop or computer | < 10 years | Medium | Fast | High, although rapidly decreasing | • Do not suffer the same issues with mechanical failure as with other devices<br>• Data retention period varies depending on the manufacturer of the device. |
| **Magnetic Tape**: cartridges hold a long strip of magnetic media that is used to record data in a similar fashion to a mechanical hard drive.<br>e.g., cartridges and cassettes | < 10 years | High | Slow, difficult,<br>sequential access limitations. | Tape is a low cost media with very high capacity per cartridge.<br>Upfront cost for infrastructure may be substantial. | • Requires specific hardware, but large scale robotic tape archives are available with archiving control software<br>• Used for system backup, data archive and data exchange<br>• Over time, tape can start to mechanically fail leading to failure and loss in the availability of data.<br>• Stability is affected by temperature, humidity and time.<br>• If left unused for a long period of time (years), mechanical failure in the tape may occur. |

| Media | Data Retention Period | Density | Access | Costs | Other Factors/considerations |
|---|---|---|---|---|---|
| **Optical Disc**: stores read-only data on a physical disk such as a CD or DVD. A laser is used to read and write data to the disc.<br>e.g., CD, DVD, Blu Ray | 2 - 5 years | Low | Slow | Low | • Requires an optical drive to read and write the discs.<br>• Write-once/read-many strategy.<br>• More stable compared to a mechanical hard disc.<br>• Relies on an external mechanical drive that will deteriorate over time, leading to the potential unrecoverable loss of data. |
| **Archival Optical Disc**: is a type of optical disk similar in fashion to a recordable CD or DVD but the construction of the disk is different so that it will prevent the disk from degrading over time.<br>e.g., advanced versions of CDs, DVDs, Blu Ray, such as Verbatim's M-Disc, Sony's Optical Disk Archiving. | 50-100 years | 500 MB to<br>> 5TB<br>media dependent | Slow | Low | • Some solutions offer robotic archiving with archiving control software<br>• Write-once/read-many media.<br>• Requires an optical drive to read and write the archival optical discs |

| Media | Data Retention Period | Density | Access | Costs | Other Factors/considerations |
|---|---|---|---|---|---|
| **Cloud-Based:** archives data with a third-party provider.<br>e.g., Amazon Web Service (AWS), Google Cloud Platform (GCP), Microsoft Azure, Oracle Archive Storage | Vendor specific | High | Varies | Depending on the volume of data and the access requirements (less than once a year, yearly, monthly), this solution can be cost effective. | • Reduces the dependency on site-based hardware<br>• Service Level Agreement (SLA) will detail the service levels that are provided.<br>• Becoming more commonplace due to the increase in internet performance and the volume of storage available in 3rd party data centers.<br>• Risk is moved to the vendor for maintaining integrity and availability of data<br>• Offer high-availability and high-durability (integrity and lifetime) through a software application programming interface.<br>• Data will be required to transit across a network that is potentially insecure or traverses through different jurisdictions.<br>• Data may be stored in a physical location that may violate organizational, local or jurisdictional privacy laws. |

## Archive documentation, labelling, and storage

Independent of the choice of media used for archiving, a strategy also needs to be developed for storing, securing, and searching the data as well as accessing it. Metadata describing what the data is and where it is archived needs to be available so that users will know what is available. Important practices for documentation of archives include:

- Actively maintain a structured system detailing information about what data is stored in each archive along with the age of the archive and how/where it is stored.
- For cloud-based archives, maintain active records of what has been archived, when it was archived, and access requests.
- Archives media should be labelled with descriptive data that can be cross-reference to metadata. Labelling should be done with acid-free archival label material to ensure that labels can be read in the future.
- For each archive created, record information regarding the data formats used and how the data was secured on the device.
- Physical archive media should be stored in a secure location with physical access controls and logging so that access to the media can be monitored and controlled.
- Physical archive media should be stored in an area with the proper environmental conditions to ensure the proper lifespan of the media. Storage locations need to have proper safety protection for fire and water damage to preserve the media.
- As archives are generally considered to be read only, organizations may choose to maintain multiple archives in different locations for asset protection. If data requests are made to change archived data, multiple copies can increase the complexity on how this is managed. See Supporting Data Change Requests.
- Organizations will need to develop a policy regarding how long archives will be maintained. A cost analysis will need to be considered when developing these policies as the long-term archiving of data on site may be cost prohibitive. Third party archival storage agencies can provide options for off-site storage but increase the complexity of record retrieval. Cloud- based archiving can help offset long term physical storage requirements.

## Accessing archived data

When a user requests archived data, the data will need to be brought back online for the user. As the data should be encrypted, actions are required to decrypt and make the requested data available for the user. Depending on the media and the storage solution, this can take a significant amount of time.

Access options include:

- An online read-only copy directly from the archive for a specific time period with read access rights for a specific user.
- Requested data records are brought into the live data system as read-only data from the archive by data administrators.
- An offline read-only copy of the data (risk of data loss/multiple copies)

- Hardcopy record of the data.

Archive access should be controlled and monitored. The process of managing data access in an archive depends on the strategy implemented:

- For small local archives, the individual responsible for maintaining and managing archives would process the request, locate the physical media, bring the media online, bring the desired data into the active storage area for the user or provide the information to them directly, and log the access. The speed of the process is limited by the operator and choice of media.
- For larger local archives using an automated tape or optical cartridge archive solution, the individual responsible for processing the archive request will instruct the device to bring the target archive online. The automated solution will locate the archive and bring the required media online and locate the target data. The speed of the process is limited by the choice of media.
- For cloud-based archives, the individual responsible for processing the archive request will initiate the data access request with the cloud-based provider. Software integration is required between the organization’s active data store and the cloud archive solution. The speed of the process is limited by the vendor's SLA.

## Supporting data change requests

Generally, archive data is considered to be read-only, which means that records do not change. There are cases where an organization receives requests from persons requesting the removal or deletion of data that has been archived. If it is desirable to support data modification requests, your government will require:

- A formal request process to allow for the tracking of the data modification request.
- A documented process for data modification/removal from archives.
- A confirmation process.

## Physical data destruction

Over time data storage devices will fail or are no longer needed. When these devices are removed from the organization's infrastructure, the organization must ensure that security and privacy concerns are addressed. For digital assets, this involves the physical deletion of data and the destruction of the device in a secure fashion.

When data is destroyed, it is critical to maintain a record of the destruction so that if in the future the data is requested, there is an audit path indicating the proper removal of data from the system.

### Best practices for destroying digital media

When destroying digital media, follow these practices:

- Utilize a secure erase utility to wipe data from disks. A normal erase can leave data accessible on both mechanical and solid disks. This is critical regardless of whether the media is going to be reused internally or sent for destruction.
- Send all media for confidential destruction. **Do not send hard drives, tapes, or USB storage devices to electronics recycling as they can end up in uncertain locations and have no security or privacy guarantees**

  - Third party organizations offer confidential media destructions services with a certification of destruction and a secure destruction path
- Ensure all materials slated for destruction are held in properly labeled containers in a secure location while awaiting destruction.
- Record and audit all media destruction.

**Best practices for destroying physical records**

Destroy physical records using secure shredding. Disposing of confidential paper records via paper recycling violates privacy and security policies. **Data breaches have occurred from individuals targeting paper recycling and unsecured shredding locations**. For physical record destruction, follow these practices:

- Utilize a secure shredding mechanism, such as a cross-cut shredder, for in-house shredding.
  - Shreds should be stored in a secure container before being destroyed.
- Utilize a secure shredding service (preferred).
  - Services will provide a secure and locked material receptacle for on-site use and perform curbside destruction.
  - Services will offer certificates of destruction.
- Ensure all materials slated for destruction are held in properly labeled containers in a secure location while awaiting destruction.
- Record and audit all media destruction.