Data Storage

Table of Contents

INTRODUCTION

Data storage refers to where data that is currently in use by your government will reside. Where the data resides depends on the storage solution that has been selected during the implementation and procurement phase of the system. For example, data will reside in the cloud or VPS for cloud based-solutions. If your government is self-hosting, co-locating, or managing leased hardware, direct management of the data will be required.

Most systems will rely on solid state hard drives for fast data access with mechanical hard drives for large quantities of less frequently used data. The data management or operating system will often manage the placement of data in more sophisticated systems to improve data access, but smaller systems may have all data resting on a single hard drive.

Data storage devices need to be accessible to authorized and authenticated users. A database management system (DBMS) (see the document "Types of Software Needed for Data Management") will generally be used to organize data on a disk and allow users to search for data based on different attributes of the data set. To improve the ability for data to be searched, indexing of data can be done. Many database solutions support the construction of indexes to improve the search performance of systems but will require the use of a database administrator. A DBMS will allow users to search for data based on text-based records.

KEY POINTS IN THIS DOCUMENT

- Data storage refers to where the data that is currently in use by your government will reside.
- Most systems will rely on solid state hard drives for fast data access with mechanical hard drives for large quantities of less frequently used data.
- A database management system (DBMS) will generally be used to organize data and allow users to search for data based on different attributes of the data set.
- Avoid data duplication through Single Source of Truth.
- Data storage systems should use some form of redundancy where the system offers a form of physical drive protection in the event of a hard drive failure.
- Your government should have a comprehensive data protection strategy that addresses both data security and privacy considerations and dictates data storage practices.
- Data should be protected physically and digitally.
- Removable storage devices should use full disk encryption.
- There are three types of encryption: full disk encryption, volume/file system encryption, file level encryption.

BEST PRACTICES FOR DATA STORAGE

- When dealing with data in systems, regardless of the location, data duplication will need to be avoided. When multiple copies of data exist, it can be unclear which data is the correct data. Data system design should use Single Source of Truth (SSOT) where all requests and data accesses will direct to one authoritative source (see the document "Data Quality" for further discussion).
- Data storage systems should use some form of redundancy where the system offers a form of physical drive protection in the event of a hard drive failure. The design of the physical storage system will need to be included during the implementation and procurement phase and included in the RFP. The type of storage must also consider the volume of access requests a system will handle and needs to be designed to eliminate potential bottlenecks.
- For semi-structured items such as documents, a document-orientated database or document-store can be utilized that can help with access and retrieval of documents.
- Ensure that for non-text based documents (such as image scans of documents), ensure that correct metadata is captured to allow for searchability of data.

PRIVACY AND SECURITY CONSIDERATIONS

As data stewards, governments collect large amounts of personal information. While single data points may not pose a substantial privacy risk in isolation, it may be possible to link the data with other data, increasing the potential damage of a breach to an individual. Your government should have a comprehensive data protection strategy that addresses both data security and privacy considerations and dictates data storage practices. Data should be considered a valued asset and a potential liability that must be protected both in transit (such as moving across a network or being transported on a USB drive) and at rest.

Principles of data storage privacy and security

Digital data

- Computers and hard drives need to be physically secured to prevent loss of devices.
- Data systems should be properly secured in a locked location.
- If data is stored on removable media, they should be secured in a locked and monitored location.
- In addition to the physical security of data store devices, devices will need to be digitally secured to maintain the privacy and security of data on the device.
- If the organization permits the use of removable media in the data management strategy, removable devices should use full disk encryption (see table 1 below) to maintain the privacy and security of data.
- Security and privacy concerns should be recorded with specific data as different data will have different security concerns. Access to data should be restricted based on access needs.

Physical data

- Physical documentation requires indexing that includes both metadata for the document as well as where documents are located.

- Physical access should be controlled with logging and auditing policies in place.
- The Archive Association of British Columbia provides suggestions and guidelines for maintaining physical records in A Manual for Small Archives

Encryption

Table 1 below outlines different types of encryption that can be used to protect data.

TABLE 1: DATA ON DISK ENCRYPTION STRATEGIES

Type	Description	Factors
Full Disk Encryption (FDE)	Contents of the entire hard drive are encrypted and managed by the operating system and hardware.	A master key can be used to unlock the drive that must be centrally measured. Encryption and decryption of data is transparent to the user. Prevents access to the entire hard drive if lost.
Volume/File System Encryption	Operating system manages the encryption of a portion of the harddrive and is managed by the operating system.	Encryption and decryption of data is transparent to the user. Prevents access to only the encrypted part of the hard drive if lost.
File Level Encryption	Encryption is managed on a file-by-file basis where each file has a separate encryption key	Allows for flexible individual management of access to files. Requires key management strategy to allow access to individual files.

Sulingituk Government: Understanding risks – data loss

Fictional case study

A Sulingituk Government employee has been working on organizing personal data that contains the financial data of individuals. As they have been working at multiple locations, they chose to store the data on a USB device. The data on the USB device has not been encrypted. The user carries the USB device in their pocket every day and accidentally drops the device. The device is found by a stranger.

What are the risks here?

Table 2 describes potential risks, issues, and solutions for a data loss scenario.

TABLE 2: DATA LOSS SCENARIO

	Issue	Potential Solution
Privacy Risk: **Loss of PII and Sensitive Data**	The data contains PII ("personally identifiable information") and financial data about individuals. Loss of this data could mean that someone could gain access to individual banking information which could cause individual harm.	Privacy and security policies must be in place and audited to ensure that users are working within the privacy framework. Persons impacted by the loss of data must be notified.
Security Risk: **Data Security**	As the device is being used to store PII and sensitive financial data which may be lost or compromised, will lead personal harm.	Data must be encrypted to secure the data in the event of data loss. Either the file could be encrypted, or the device could use full disk encryption (FDE) which would prevent unauthorized access to the entire device.
Availability Security Risk: **Data Storage and Availability**	USB devices fail and get lost. The data should reside on a network device where it can be secured but also put under backup protection. In this case, the only copy of the data is on the USB device and losing it represents not only a data privacy breach, but a loss of the time it took to prepare and organize the data. The data is also not available to other users.	Data should be stored on a network device, where backups, auditability and integrity can be assured.
Security Risk: **Data Transport**	Having removable data that can be transported on a USB device creates an opportunity for data loss in transport. USB devices are small and easily lost. Allowing for data to be moved to a USB device creates the opportunity for a disgruntled employee or thief to steal the data, potentially undetected.	Sensitive PII and unique data should not be permitted to be stored on removal devices and should instead be accessed through a secure portal, where auditability and access control are possible.
Data Quality Risk: **Duplication**	As data can be copied to multiple devices, which data is the correct data?	Data should exist in a single location source of truth to prevent confusion and duplication.

Data Security and Quality Risk: **Integrity**	The user retraces their steps and after a few days finds that someone at a local coffee shop found the device. While the user is happy to have recovered the device, the quality of the only copy of the data on the device is now in question. The data on the device may have been altered and there is no way of knowing.	Integrity checks need to be used with data to be able to determine if the data has been changed and can be considered trustworthy.

In this straight-forward scenario, there are numerous **security**, **privacy,** and **data quality** concerns. Losing government data could produce harm for the individuals that are described by the data. It could also represent a loss to the government in terms of the value of the data to the government and the damage to the perceived reliability and reputation of the organization.