Best practice guidance on digital storage and preservation

October 2024

Te Rua Mahara o te Kāwanatanga
Archives New Zealand
Part of the Department of Internal Affairs

New Zealand Government

Document details

Document Identifier: 20/G17

Version	Date	Description	Revision due
0.1	Dec 2019	Development Draft	
0.2	Jan 2020	Added preservation storage criteria based on international best practice	
0.3	Feb 2020	Internal review	
1.0	Feb 2020	Final publication version	Feb 2021
2	July 2021	Reviewed	July 2024
2.1	May 2022	Updated broken links	
3	Aug – Oct 2024	Reviewed	Oct 2027

Contact for enquiries

Government Recordkeeping Directorate
Archives New Zealand
Phone: +64 4 499 5595
Email: rkadvice@dia.govt.nz

Licence

Contents

1 Introduction

How and where digital information and records (digital records) are stored will affect their viability over time. Public offices and local authorities (public sector organisations) need to manage their digital records to meet the requirements of the *Public Records Act 2005* and the principles of the mandatory *Information and Records Management Standard* (16/S1). This helps to ensure digital records remain authentic, reliable, discoverable, accessible, usable, protected and preserved for as long they are required. This also enables public sector organisations to meet their business needs and legal requirements, particularly for digital records identified as high-risk or of archival value.

If your organisation is a public sector organisation, this document provides best practice guidance about how and where your digital records should be stored, as well as criteria to consider if you are intending to use or provide digital preservation storage systems or solutions.

Given the rapidly changing and evolving nature of digital storage and preservation, this guidance is deliberately generic and high level. You should first consult with your IT specialists to gain an understanding of your organisation's technical environment before contacting us for advice.

Please note that this guidance does not cover any contractual, jurisdictional or funding issues that may also be part of your organisation's decisions on digital storage and preservation solutions.

2 What is digital storage

Previously, your organisation may have been using discrete media such as individual CDs, tapes, etc. to store digital records which then need to be migrated periodically to address format and media degradation and obsolescence.

Nowadays it is becoming more common for organisations to use resilient IT storage systems for the growing volume of digital records that need to be preserved, and more importantly, that need to be easily and quickly retrievable in a culture of online access. In this way, management of the digital content can be decoupled from the mechanism of its storage, i.e. the media or technologies and the supporting IT infrastructure. This has the added benefit of allowing you to handle different preservation activities independently.

A resilient IT storage system consists of storage media contained within a server that provides built in resilience to various failure modes by using inbuilt redundancy and recovery. For example, it might be data tapes in a tape library, remote cloud storage, or automated replication of digital records across multiple sites and systems.[1]

How and where your organisation stores its digital records is key to ensuring they remain accessible and trustworthy for the entire time they need to be retained.

[1] See Storage - Digital Preservation Handbook (dpconline.org) (accessed 27/09/2024).

2.1 How and where can digital records be stored

There are several ways you can store digital records, including:

Online: This can be locally on your organisation's server infrastructure, or by hosted storage through the internet, for example, in cloud storage. Networked online storage is where data is stored on multiple virtual servers that are generally hosted by a third party, which may be offshore. Digital records held in online storage devices are immediately accessible to users and are more likely to be identified and included in changes such as system wide migration processes, and in regular integrity checks and back-ups.

Offline: This allows your digital records to be relatively mobile, for example, on removable storage media such as magnetic tapes, CDs, DVDs, memory cards, flash drives (USB sticks). You need to be aware of the risks with using removable media, for example, data security, malware infections, loss and hardware failures. Also, removable media is often overlooked when systems are upgraded, and digital records migrated to new formats.

Near-line: This is where your digital records are stored separate to and are not directly accessible by your organisation's systems, but can be quickly retrieved and brought online for access, for example, from a local tape library or a cloud storage service.

You should consult with your IT specialists about your organisation's specific digital storage requirements.

2.2 Storage approach considerations

You should consider the following to guide your selection of digital storage systems:

- Security – How will specific storage systems allow you to manage and enforce your digital records' privacy and/or security requirements.
- Access and availability – Select storage media with fast retrieval times if your digital records need to be accessed often and/or quickly. Also consider what specific combinations of hardware and software are needed.
- Longevity – Select storage media with an appropriate proven lifespan if your digital records need to be retained long term. Longer lifespans will also reduce the need for you to migrate, refresh the storage media or undertake other preservation activities to reduce the risk of data loss.
- Viability – What error detection and integrity checks does the storage system have in place to monitor and ensure against inadvertent change, deterioration or loss of your digital records over time, and/or when storage media is refreshed, or data is migrated.
- Obsolescence - Most digital storage media will only last 5 to 7 years before it will be necessary to refresh or update it.[2] Select storage systems that are robust with a regular, clearly defined migration path and widespread industry support to allow for storage media and their technical infrastructure becoming obsolete or unsupported.

[2] "Select storage for digital records", last updated 5 May 2021, Digital records, storage media and systems | For government | Queensland Government (Accessed 31/07/2024), Queensland State Archives.

When selecting or designing storage systems for preservation storage, you should also consider the following principles from the Digital Preservation Coalition's *Digital Preservation Handbook*[3]:

- Redundancy and diversity - For example, make lots of copies stored in different locations; use a combination of online storage systems and offline media; use different types of storage technology to spread risk and balance data safety with easy access.
- Fixity, monitoring, and repair - For example, use fixity measures such as checksums to record and regularly monitor the integrity of each digital record and each copy; store fixity information alongside the digital records as well as in separate systems; if corruption or data loss is detected, use one of the copies to create a replacement.
- Technology and vendor watch, risk assessment and proactive migrations - For example, understand that storage technologies, products and services all have a short lifetime; keep an eye on new and changed technology and the viability of storage vendors or storage solutions; be proactive, migrate storage before your digital records become at risk.
- Consolidation, simplicity, documentation, provenance and audit trails - For example, minimise the proliferation of legacy media types and consolidate your digital records onto a minimum number of storage systems; document how your digital records have been acquired and transferred into the storage system(s) as well as how these are set up and operated; use this documentation to provide audit information on data authenticity.

3 Digital preservation storage categories

Preservation storage supports digital preservation which is defined by the Digital Preservation Coalition as "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary…beyond the limits of media failure or technological and organisational change".[4]

The following nine categories or characteristics of preservation storage are based on international best practice. They are intended not only to help your organisation with developing requirements for digital preservation storage systems or solutions, but also to help with evaluating digital storage options and services and informing your IT infrastructure design and planning. You should adapt the criteria to suit your organisation's individual requirements, practices, legislation and environment.

3.1 Content security

The digital preservation storage solution (the solution) provides and/or supports features and methods that prevent harm (intentional or unintentional) to your digital records. For example, the solution:

- provides remediation actions for content found to have malware (for example, quarantine, notification, etc.)

[3] Digital Preservation Handbook, revised 2nd Edition, https://www.dpconline.org/handbook, Digital Preservation Coalition © 2024.

[4] Glossary, ibid.

- supports permanent deletion by authorised users in a way that prevents recovery, in accordance with your organisation's policies and rules
- provides the additional level of security required for personal, sensitive or confidential data according to your legal or organisational needs.

3.2 Flexibility

The solution is adaptable, interoperable and customisable to your organisation's preservation requirements or preferences. For example, the solution:

- is able to adjust storage infrastructure in response to changing requirements (for example, legal requirements, audit results, etc.)
- includes storage components that can be easily integrated with other systems and applications, i.e. plug and play (for example, uses standard file access protocols and file system semantics, etc.)

3.3 Infrastructure security

The solution has controls and safeguards that protect the storage system's infrastructure from interference or intrusion. For example, the solution:

- provides role-based, access controls to ensure that your digital records cannot be easily altered or inappropriately accessed
- includes software that regularly conducts checks to identify malware.

3.4 Preservation actions

The solution supports your organisation's preservation requirements by allowing you to take an active role in monitoring, managing, and preforming interventions on your digital records. For example, the solution:

- performs verifiable and/or auditable integrity checks to detect changes or loss in or across copies (for example, checksum recalculation, fixity checking, missing files) at regular interval, during transfers, etc.
- allows or supports the use of tools to perform preservation actions both at the individual object level and in bulk.

3.5 Resilience

The ability of the solution to resist, remediate or recover quickly from threats, errors or other difficulties. For example, the solution:

- provides sufficient backup and disaster recovery functionality to ensure continuity of repository functions
- has failure tolerance measures in place to enable continuous operation without interruption for a long period of time (for example by eliminating single points of failure with effective monitoring)
- replaces or repairs missing or corrupt files in acceptable timeframes or provides the ability and tools for your organisation to perform these actions independently.

3.6 Scalability and performance

The ability of the solution to meet the diverse and changing storage, architectural and computational needs of your organisation. For example, the solution:

- supports the entire export of your content and metadata for any reason, within an acceptable timeframe (for example, as part of an exit strategy)
- is able to support long file, path or directory names and diverse character encodings.

3.7 Support

The assistance provided to your organisation related to its use of the solution. For example, the solution:

- supports periodic performance reviews, assessments, validations and audits required by your organisation (for example, reports, technical documentation, transaction history, performance data, and continuity practices)
- supports contingency plans and strategies to stop using the solution (for example, the ability to transfer content to another solution without loss)
- provides appropriate training to your staff across all relevant operational and maintenance tasks.

3.8 Sustainability

The financial, environmental and/or other impacts on your organization or its broader context (society, the natural environment, etc.). For example, the solution:

- costs relatively less overall than other comparable solutions, by being designed with cost efficiencies (for example, has resource pooling and sharing, multi-tenancy (i.e., multiple users share the same applications))
- takes advantage of energy conservation principles and techniques (for example, prefers green computing options that require less cooling, consume less power, or use less rack space).

3.9 Transparency

This characteristic refers to the assurance, evidence and visibility that your organisation has into the activities and status of the solution and its content. For example, the solution:

- provides reports about content (for example, number of objects/files/formats, average file size, types of objects, etc) as well as custom configurable and on-demand reporting of content or activity
- captures and documents all actions relating to the content (for example, information about integrity check failures, deletions, modifications, additions, preservation actions) and who or what performed the actions
- provides full, complete, current and available documentation of key processes, services, systems, procedures, known limitations and functions, and changes that have been made to them.