



Digital Preservation Maturity Model

Organisations are realising that it is business critical to retained digital information for several years. For whatever reason it was retained – avoiding risk or creating value - ensuring this information can be accessed when needed is key.

However, the terms “Digital Preservation” or “Digital Archiving” are used to cover many different types of solution to this problem. This paper provides a way of categorising these different approaches to enable organisations to understand the differences and to select the best solution for them.

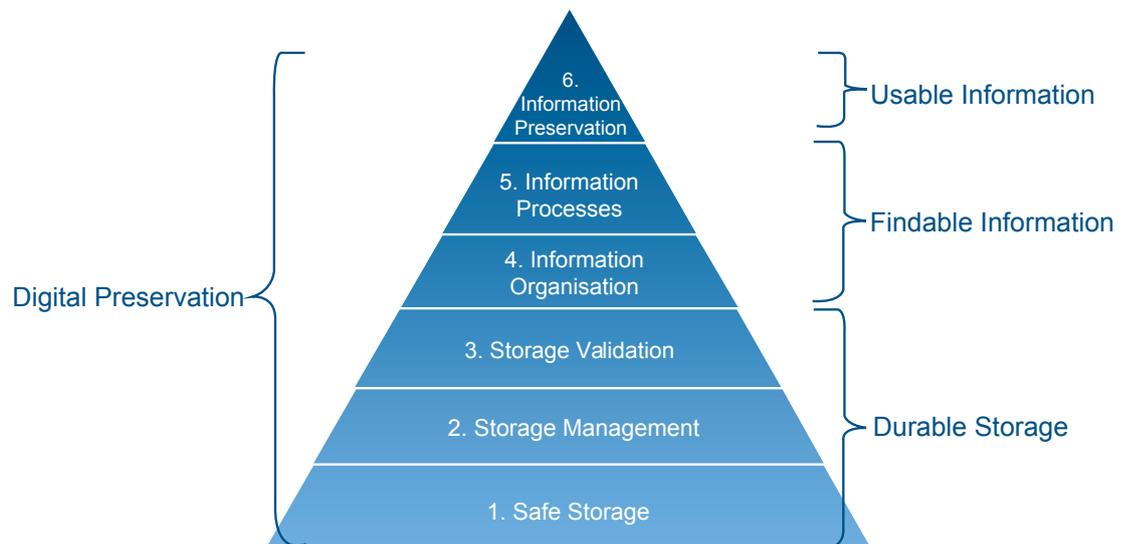
Why “Maturity Model”

The term Maturity Model is used to imply layers of sophistication in processes, the first of which must be complete before graduating to the next. This is true for Digital Preservation – there is no point having a clever information management system if you do not have secure storage.

Key Components

The Digital Archiving Maturity Model has three main sections:

- Durable Storage – Layers 1-3 provide increasing levels of sophistication in the safety and security of the storage of the raw bits used to hold information. By the time you have a Level 3 compliant system you can be confident your information will not be lost and that it has not been tampered with.
- Information Management – Layers 4-5 ensure that the preserved raw bits are organised. They have a hierarchy, descriptive metadata, and security, and they have a set of powerful tools to allow upload, management, search, browse and download.
- Information Preservation – Layer 6 is critical for information that must be retained for more than the lifetime of the application that created it. It ensures the file formats in which the information is held remain relevant to the applications available at the time the information is required, thus enabling it to be use immediately.



Level 1 : Safe Storage

The simplest layer incorporates simple bit-level storage on magnetic or optical storage with some level of reassurance that the bits are protected against simple storage failure. This includes storage on protected spinning disks using RAID techniques, optical media with long-term storage and managed locations, or long-life tape storage.

Key characteristic of Safe Storage

- Demonstrable resilience to common threats such as minor physical damage.

Level 2 : Storage Management

At level 2 we add active storage management which moves the bits to the most appropriate location. The decision on which bits are located where may be done on the basis of storage durability, cost reduction or performance. The criteria are flexible to balance all of these drivers. Information must be held in at least two locations, for example disc and backup tape or multiple disc replicas.

Key characteristics of Storage Management

- Rules to move objects between storage locations depending on criteria of cost, or durability, performance
- Object is held in at least two locations and optionally more
- Can change rules at any time to move information to new locations

Level 3 : Storage Validation

The final storage sophistication adds multiple object storage plus fixity checking to validate storage durability. Object fixity is checked on storage, access and at regular intervals to confirm objects have not been tampered with. If bit failure is identified, self-healing from an alternative copy will occur.

- Information copies are held in multiple safe storage locations, ideally geographically remote.
- Checksums created for all information items under storage
- Use checksums to check information integrity at regular intervals and on every access
- Self-heal from alternate copy if information corruption is detected.



The Best Storage System

A Level 3 Storage System appropriate to Digital Preservation will combine intelligent storage virtualisation with the digital object assurance. This will make sure it is possible to store the information efficiently and retrieve it as required, sure in the knowledge that it has not been tampered with during this process. In short you are entirely confident that what you put in is what comes back, and that it will come back when you want it.

Durable storage systems of this type are increasingly available as an off the shelf

offering from prominent suppliers. These can be disc or tape based and hosted in your data centre, a combine appliance, a remote storage service, or in the cloud, or a combination of all of these. The key point is to ensure they are fully compliant with the requirements of Durable Storage.

Level 4 : Information Organisation

The first level of information management incorporates the ability to turn raw bit level files into information. This is a dramatic increase in the usefulness of the information.

The key characteristics are:

- Information organised into hierarchies, with the concept of information objects comprising many files, and collections comprising many information objects.
- Information tagged with metadata giving it a description, history, and context.
- Tools to upload new information to the system
- Security model to limit access to permitted information
- Can edit the metadata, modify the logical hierarchy and delete content with suitable permission.
- Can traverse the hierarchy to locate information of interest
- Can search both the metadata and context for information of interest
- Can download information of interest, whether files or groups of files, and the metadata associated with them.

Level 5 : Information Processes

The next level of sophistication adds efficient and flexible business processes to automate the activities associated with information management. These include interfaces to the information sources and dissemination to information consumers using flexible workflows and programmer interfaces. They also include high-throughput capabilities and integration with a third party identity management system. Non-archiving processes such as object versioning, should be excluded.

The key characteristics are:

- Tools such as workflows to allow business process automation
- Programmer interfaces to allow system extension
- Integration with content source and disposal locations
- Demonstration of suitably high throughput

Exclusion of features damaging to Digital Preservation such as object editing and versioning

The best Information Management system

Information Management can be performed by a wide range of systems, depending on the information under management, for example systems for Content Management, Product Lifecycle Management, Laboratory Information Management, and Transaction Management. The key things to be aware for when considering a Preservation choice are the need to integrate with Durable Storage and exclude features usual in information management but counterproductive in Digital Preservation.

Level 6 : Information Preservation

To be useful digital information needs to be usable by the person who wants it at the time it is needed. If retaining information for a few years this is not a challenge. When needing to use information in 10 years or more in the future, the application landscape may have changed so much as to make its easy use impossible. A mature system must thus incorporate Information Preservation tools to counter this threat.

The key characteristics of Information Preservation are:

- Identification of the file formats and conceptual objects (information object comprising many files) that make up the ingested information
- Extract characteristics of these objects for future validation
- Ability to identify file formats at risk and in need of attention
- Perform loss-less migration to new “Digital Master” files, and to validate this migration
- Perform lossy migration to presentation formats at suitable quality for dissemination
- Keep the original information plus all relevant representations created later
- Continually add new tools to keep Information Preservation active.

The best Information Preservation system

Fully fledged Information Preservation systems are often based on the pioneering work of National Archives, Libraries and Academia. In fact incorporation of the preservation functions or principles from these initiatives is a key feature of these systems. Another is the acceptance that change will happen – migrating to a “durable” format now such as PDF is naive as it is likely to be a lossy migration and PDF is far from assured as a future format.

The best solutions are described as “Active Preservation” – instead of waiting for disaster to strike they actively make sure the information is usable as required, constantly reviewing it to find the best way of making sure it is always available.

What do I Need?

A key question to ask is what level of sophistication do I need? If it is raw storage you are looking for levels 2-3 might be good enough. If you just need to keep information for a couple of years but need to be able to find it quickly level 4 might be sufficient. If you have high volumes of information and few staff you will need a level 5 System. If you are concerned about being able to find information and to be able to use it more than 5-7 years after it was created you need a level 6 System. Make sure you use the right tools to assess how vulnerable you are to digital loss and whether your digital value is at risk.

Contact Preservica to find out about our world leading technology, fully compliant to level 6 and available on-premise and in the Cloud.