Digital Preservation System Requirements and DPC's Rapid Assessment Model

A digital preservation system is only one element of a successful digital preservation implementation. As neatly illustrated by Kenney and McGovern's three-legged stool[1], which has one leg devoted to Technology and the other two representing Organization and Resources, it is not possible to create a balanced platform to sustain digital preservation activities if only one of these areas is addressed. In short, procuring a digital preservation system isn't going to solve digital preservation for you, sufficient time must also be devoted to other aspects of digital preservation.

The DPC's Rapid Assessment Model (DPC RAM)[2] is a maturity model that provides a holistic view of the elements that need to be in place to carry out digital preservation successfully. These include organization commitment and resourcing, staff expertise, policies and procedures, legal and ethical frameworks and of course more technical digital preservation concepts such as bitstream preservation and metadata management. Using DPC RAM will enable an organization to better understand where they are and where they would like to be. Carrying out a maturity modelling exercise and in particular, setting target levels, is a useful step to carry out before considering technical requirements for a digital preservation system.

A detailed mapping of requirements against DPC RAM is included in the table that follows and may help users understand the requirements that are most important for them based on RAM target levels they have set. For example, if the RAM target level for section I (Content Preservation) is a 3 and there is no requirement to go higher than this for the foreseeable future, an organization may decide to leave out any system requirements which map to level 4 of this section.

Introduction to mapping

A mapping to DPC RAM from these requirements is included below. There are a few points to note when using this mapping:

- The mapping is based on the illustrative examples that are included in DPC RAM version 3.0. These examples may not be applicable to every organization in every context and there may be alternative ways that an organization could meet a particular level of RAM.
- The system requirements do not cover every section of RAM and RAM does not cover every point included in the system requirements.
- Some mappings are partial – for instance, the requirement covers only one aspect of an example in RAM or covers a greater scope than is included in the RAM example. In some cases, more than one section of RAM is included in a mapping.
- It should not be assumed that having a system that fulfils these requirements will move you to the level of RAM shown in this mapping. There are likely to be additional tasks that will need to be carried out in order to get you there (for example, establishing policies and procedures around a particular task, or training staff).

[1] https://deepblue.lib.umich.edu/bitstream/handle/2027.42/60441/McGovern-Digital_Decade.html?sequence=4

[2] https://www.dpconline.org/digipres/implement-digipres/dpc-ram

System requirement from 'Core Requirements for a Digital Preservation System'	RAM section	RAM level	RAM example text
1. The system must precisely record and manage the integrity and authenticity of the digital content and metadata it holds. It must use checksums to verify that digital content has not accidentally or maliciously been changed over time and metadata to record any actions enacted on the content.			
1.1 The system must record checksums for every file.	H - Bitstream preservation	2	Checksums are generated for all content.
1.2 The system must be able to validate checksums against those supplied with content.	G - Acquisition, transfer and ingest	3	Successful transfer of content is verified by integrity checking.
1.3 The system must support periodic integrity checking, reporting any damaged or missing files.	H - Bitstream preservation	3	Content is managed with a combination of integrity checking and content replication to one or more locations.
1.4 The system must support steps to repair or replace damaged files from replicated copies and report on actions taken.	H - Bitstream preservation	3	Content failing integrity checks is repaired.
1.5 The system must be able to generate an audit log and record event metadata (such as that required by PREMIS) describing all actions enacted on digital content.	I - Content preservation	3	Actions resulting in changes to digital content are recorded, including details of when, what, how, why and who
2. The system must have a comprehensive data model that enables the complex structure of digital objects to be captured on ingest and accurately represented over time as they are managed and preserved.			
2.1 The data model must be able to capture and represent digital objects that are composed of multiple hierarchical components, such as: files, drafts, published versions, or copies subsequently created for preservation or access.	J - Metadata management	3	Structural relationships between the data and metadata elements that form a particular digital object are maintained.
2.2 Digital objects should be assigned unique and persistent identifiers.	J - Metadata management	3	Persistent unique identifiers are assigned and maintained for digital content.

3. The system must provide a clear exit strategy to other systems, without vendor lock in. This must ensure that the inevitable migration to a future digital preservation system is possible. It should also minimise the effort and risk involved in such a migration.			
3.1 The structure of stored digital content could be understandable/interpretable without the preservation system application itself.	J - Metadata management	4	Managed exit strategy is facilitated by standardized content packaging and metadata standards.
3.2 The system must have the ability to batch export digital content and all associated metadata in a manageable format/structure for ingest into another system. 3.3 The preservation system application could be placed in an escrow, providing some reassurance in the event of the failure of the vendor.	K - Discovery and access	3	Established access use case in place for mass extraction of all digital content during invocation of an exit strategy.
4. The system must enable authentic digital content and metadata to be ingested.			
4.1 The system must enable the ingest of digital content and associated metadata at scale.	None		
4.2 Content for ingest should be virus checked, with appropriate facilities for quarantine.	G - Acquisition, transfer and ingest	2	A working area (physical or virtual) is available for pre-ingest and ingest activities (for example to carry out virus checking and file identification).
4.3 The system must be able to retain not just the original bitstream, but other characteristics necessary for preservation and access to the content such as: original folder structure, file information such as date stamps, associated documentation, and associated metadata.	G - Acquisition, transfer and ingest	2	Documentation and metadata may bereceived or captured as part of the acquisition or transfer process.
	G – Acquisition, transfer and ingest	3	Successful transfer of content is verified by integrity checking.
4.4 The system must be able to provide management reports on the success or failure of ingest activities and should also be able to flag any potential issues with ingested content (e.g. deprecated file formats, unrecognized file	A - Organizational viability	3	Metrics and reports can be generated about the digital archive to help inform reporting, planning and management.

formats, missing metadata, password protected or encrypted files etc.).	I - Content preservation	2 & 3	File formats are identified. Content is characterized and assessed for preservation and quality issues such as encrypted, broken or incomplete content and invalid files. Preservation watch activities are carried out and 'at risk' content is identified.
5. The system must have the facility to assess the characteristics of ingested digital content and record them in associated metadata.			
5.1 The system must identify file formats to the level of specific file format version and reference appropriate registries of further information such as PRONOM and/or Wikidata.	I - Content preservation	2	File formats are identified.
5.2 The system must extract technical characteristics (such as size, image dimensions, video codec, audio run time, creating application).	None		
5.3 The system should identify content that cannot be rendered, such as broken, badly constructed, or encrypted content.	I - Content preservation	2	Content is characterized and assessed for preservation and quality issues such as encrypted, broken or incomplete content and invalid files.
5.4 The system should validate file formats against file format specifications or customised profiles.	I - Content preservation	2	Content is characterized and assessed for preservation and quality issues such as encrypted, broken or incomplete content and invalid files.
5.5 The system should capture external dependencies, where content (or software) not present in the digital object is vital for it to be rendered or used (such as non-embedded fonts, non-embedded media such as YouTube videos or software libraries).	I - Content preservation	3	Technical dependencies are detected and documented.

6. The system must support replication and storage management. The system must have the ability to store multiple copies of ingested digital content on different storage systems in different geographical locations.			
6.1 The system must automatically manage the replication of digital content to multiple storage locations (potentially in different geographical locations).	H - Bitstream preservation	3	Content is managed with a combination of integrity checking and content replication to one or more locations. A process of risk assessment is used to evaluate storage risks and appropriate mitigations (such as the number of copies, location, technologies used, frequency of integrity checking).
6.2 The system should perform regular system backups.	None		
6.3 The system should be able to regularly test and report on its backup and restore capabilities.	H - Bitstream preservation	3	Tests are routinely carried out to verify the effectiveness of backups, replication and integrity checking.
6.4 The system should create and retain management reports on replication, storage management, backup and restore activities.	A - Organizational viability	3	Metrics and reports can be generated about the digital archive to help inform reporting, planning and management.
7. The system should support preservation planning, including risk assessment, and the design and testing of plans to preserve digital content.			
7.1 The system should support the identification, management and analysis of preservation risks (e.g. when digital content is in a file format that is no longer supported).	I - Content preservation	3	Preservation watch activities are carried out and 'at risk' content is identified.
7.2 The system should enable the design, development, and management of plans for mitigating identified preservation risks.	I - Content preservation	3	Actions are occasionally carried out to ensure preservation and quality of content such as

			migration, emulation or modification of creation or capture workflows.
7.3 The system should deliver reporting to enable the effective management of digital content. This must include a wide range of functional, operational, and statistical reports and analytics.	A - Organizational viability	3	Metrics and reports can be generated about the digital archive to help inform reporting, planning and management.
8. The system should support preservation actions that fulfil preservation plans designed to mitigate identified preservation risks.			
8.1 The system should enable the migration of files from one file format to another.	I - Content preservation	3	Actions are occasionally carried out to ensure preservation and quality of content such as migration, emulation or modification of creation or capture workflows.
8.2 The system should enable the rendering of digital content via the application of emulation and/or other specialist tools.	I - Content preservation	3	Actions are occasionally carried out to ensure preservation and quality of content such as migration, emulation or modification of creation or capture workflows.
8.3 The system should quality assure the results of any preservation action	I - Content preservation	4	Quality control is in place to assess (and record) the outcome of preservation actions, ensuring that the meaning and/or functionality of the content has been retained as required.
8.4 The system must record preservation actions (and their outcomes) in the associated metadata.	I - Content preservation	3	Actions resulting in changes to digital content are recorded, including details of when, what, how, why and who.
9. The system must provide facilities to manage digital content and metadata over time.			

9.1 The system must provide controls to minimise the risk of accidental or malicious deletion or content change.	H - Bitstream preservation	4	All access to content is logged and reviewed for unauthorized use and/or changes made (for example which content, when and by whom)..
9.2 The system must be able to ingest and manage all required metadata including metadata appropriate for specific content types (e.g. geospatial, audio visual).	J - Metadata management	2	Metadata and documentation acquired with content is retained and preserved.
9.3 The system must facilitate the enhancement of metadata and documentation throughout the lifetime of the digital content.	J - Metadata management	4	Metadata and documentation can be enhanced throughout the lifetime of the content.
9.4 The system must enable the managed disposal of content and permit the retention and maintenance of metadata even when the associated digital content has been removed from the system.	None		
9.5 The system must restrict any actions to manage digital content, based on configurable user roles.	H - Bitstream preservation	3	Authorizations to access the content by staff are enforced and documented.
10. The system must enable the controlled discovery of, and access to, digital content and metadata.			
10.1 The system must ensure that all digital content and any associated metadata is only discoverable and accessible by authorized users.	K - Discovery and access	3	User access rights are displayed and enforced. Clear information is provided to users on permitted uses of the content.
10.2 The system could warn users if they attempt to access any content that may be problematic (e.g. digital content which is held in an unsupported file format, has incomplete metadata, requires additional permissions etc.).	None		
10.3 The system could provide emulation facilities, specialist viewers, migration on-the-fly, or helper applications to enable the use of	K - Discovery and access	4	Different options are available for access, rendering or re-use such

digital content in obsolete or unsupported formats by users.			as migrated, emulated, visualized content.
10.4 The system must provide an access interface for users and/or sufficiently capable API to enable integration with other user access systems or discovery tools.	K - Discovery and access	4	Advanced resource discovery and access tools are provided, such as faceted searching, data visualization or custom access via APIs.
10.5 The system should expose information about digital content on access, including preservation metadata and checksums, to demonstrate a chain of authenticity.	None		
10.6 Metadata and/or digital content should be harvestable.	J - Metadata management	4	Metadata is harvestable and reusable.