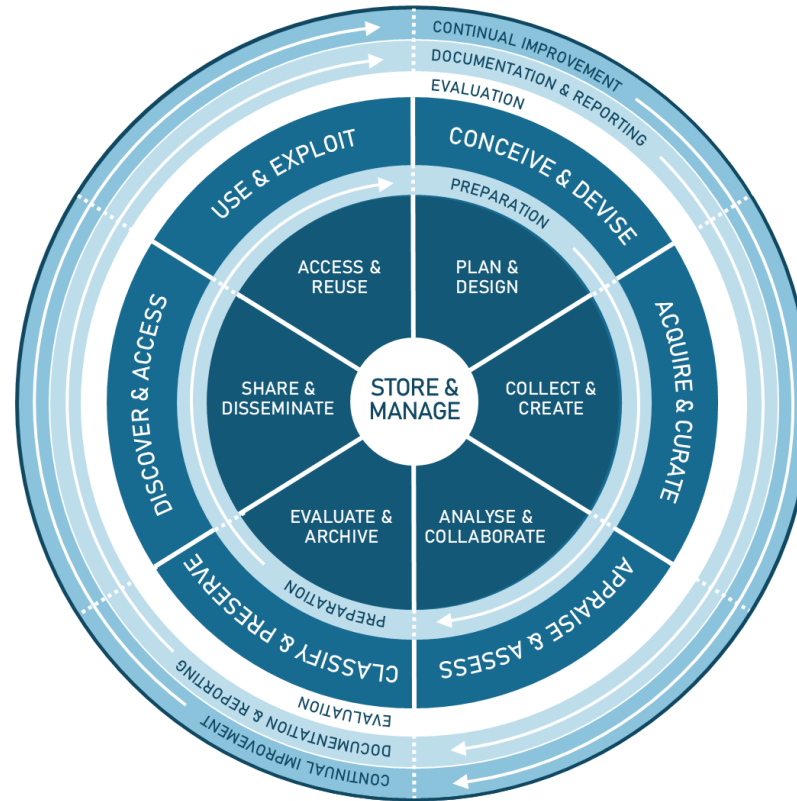# Preservation Starts with Creation

A framework for university libraries to integrate digital curation in scholarly literacies

roger mills

roger.mills@datacomit.com.au

"involves maintaining, preserving and adding value to digital research data throughout its lifecycle" (DCC, n.d).

Other terms include **Digital Stewardship** or **Digital Archiving**

What is Digital Curation ?

Digital preservation is "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary" (DPC, n.d).

Digitisation is not digital preservation!

Digital curation is not just what you plan and implement, but the series of actions you continue to perform over the whole lifecycle of research data.

Digital Curation: Core Activities

- Planning & conceiving data to be created or collected, incl., file formats and types, migration rules, version control, preserved storage options, and discovery and accessibility of data;

- Creating or acquiring digital objects, developing metadata (incl. preservation data) through experimental research process;

- Appraisal and selection: what research data is useful for analysis and storage;

- Ingest to a preserved storage repository with checksums, and relevant retention and disposal information; and

- Discovery, access, use, and exploitation (re-use) of research outcomes.

Importance of Digital Curation in Research Data Management (RDM)?

- Ensure research students are equipped with necessary skills & knowledge to manage their data for maximum impact and future professional academic work;

- Requirement of research funding bodies to ensure the integrity, authenticity, and preservation of project data; and
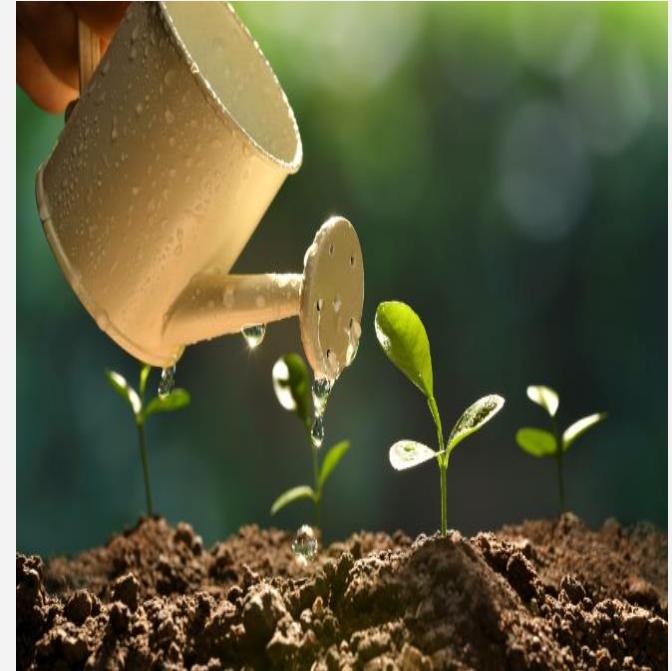
"When data is well managed, researchers can get more out of the data that they collect and can find new value in existing data" (Hicks, 2023, p .iii).
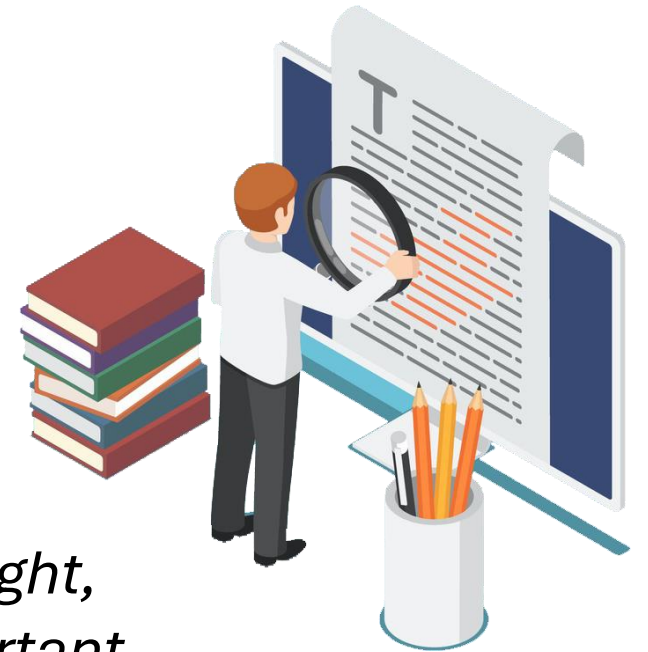
Research

- Review of academic literature; research papers, blogs and forums related to digital curation and RDM;

- Semi-structured interviews with de-identified Research Librarians, research students and ECRs from selected Group of Eight (Go8) Australian universities, and other Australian universities; and

- Examination of university library RDM guides and toolkits, and training across the Go8 Australian universities, other Australian and international universities.

Key Research Findings

- Research Data Management (RDM) programs included high level information about digital curation, but it was not a core feature;

- Research librarians were unsure if their data repository performed file validation, fixity checks and obsolescence monitoring;

- Research librarians felt they lacked the knowledge and skills to support digital curation as part of RDM training and support;

- Research students reported they hadn't received direct tuition in digital curation or preservation, and some expressed a lack of interest in learning about it;

- Research librarians thought resourcing digital curation training could be a problem; and

- Interviewees from both groups expressed confusion about the differences between digital curation, digital preservation, and the role they play in RDM.

In their own studies, Weatherburn and Neish (2018) highlight, digital preservation was not widely understood as an important part of the research lifecycle" (p. 2).

Existing Frameworks

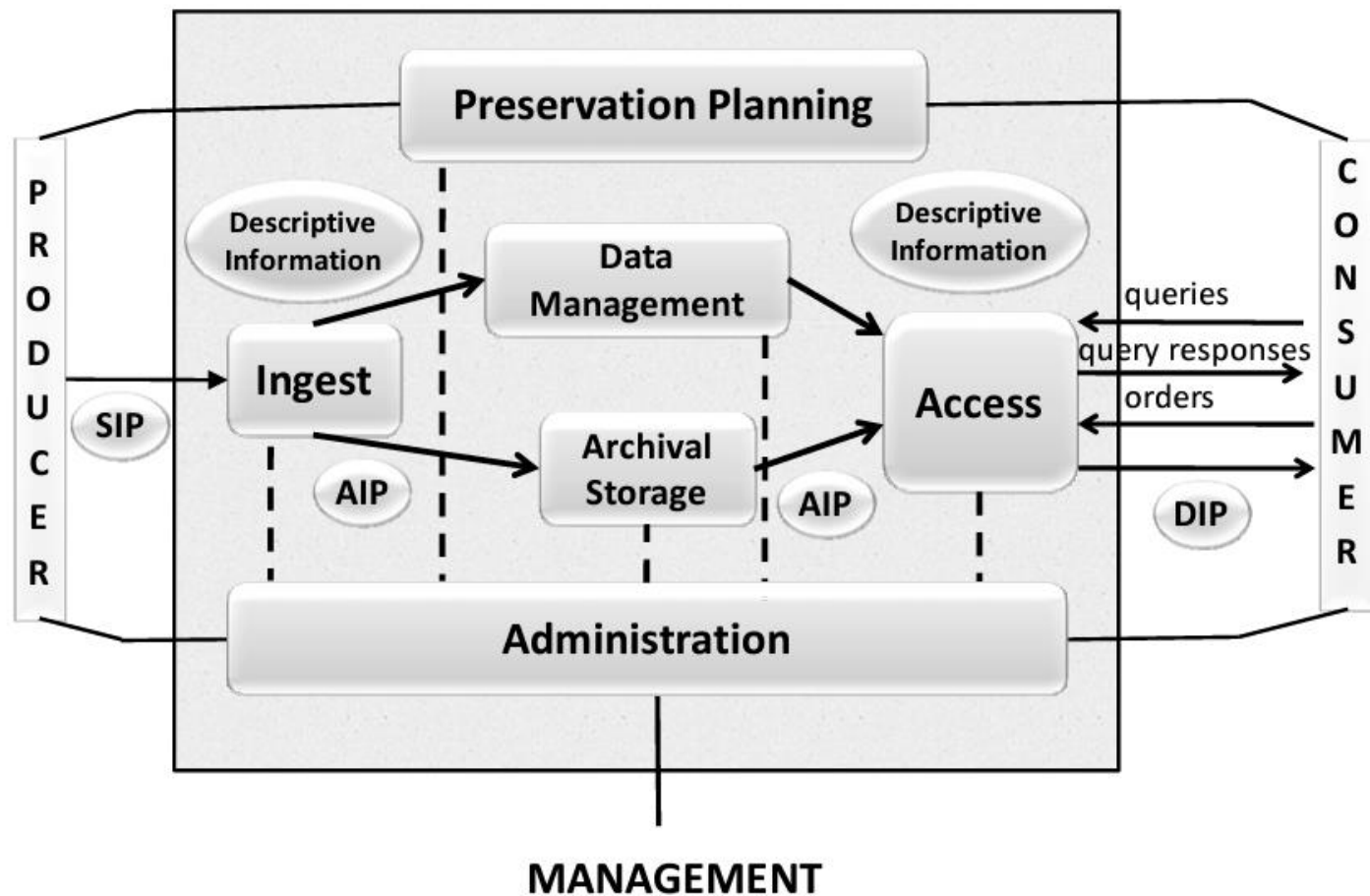OAIS (Open Archival Information System) ISO 14721:2012

Fig. 1 OAIS Functional Model, (CCSDS, 2012)

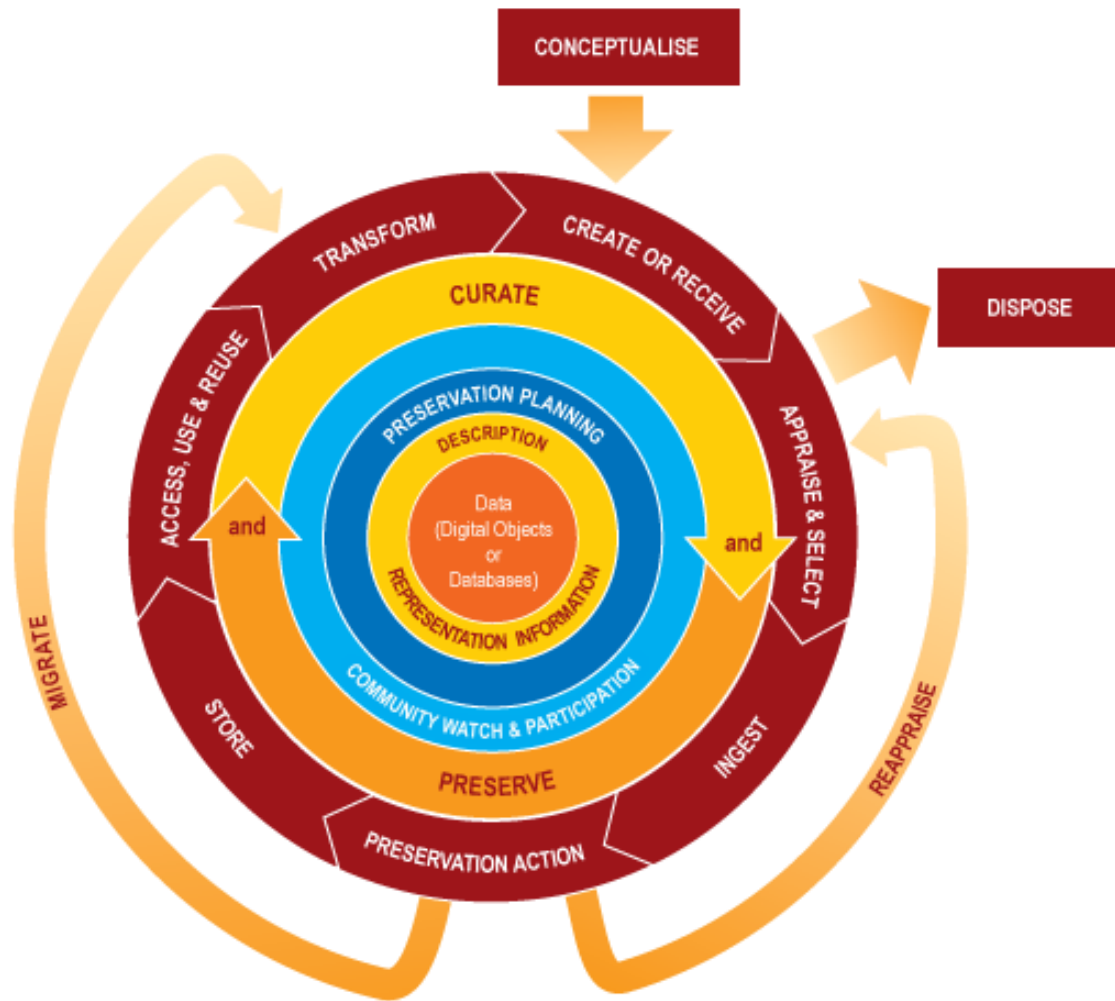# Digital Curation Centre Lifecycle Model

Fig. 2 Digital Curation Centre Lifecycle Model, (Higgins, 2008).

Research Data Lifecycle

Fig. 3 Research Data Lifecycle (LMA Research Data Management Working Group, nd)

Institutional Underpinnings Program: Research Data Management Framework for Institutions

Fig 4. Research Data Management Framework for Institutions mapped onto the Research data life cycle (Australian Research Data Commons, 2023).
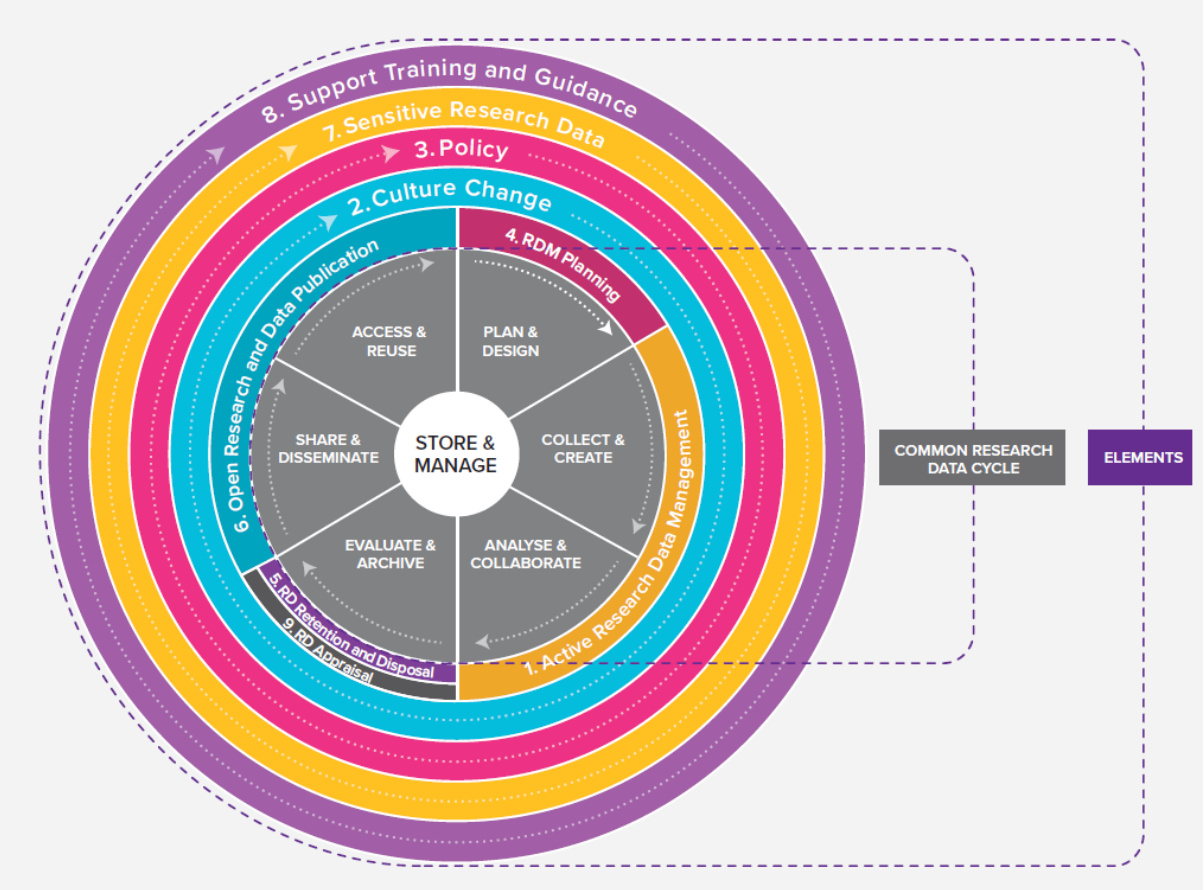
1. Active Research Data Management
2. Culture Change
3. Policy
4. Research Data Management Planning
5. Research Data Retention and Disposal
6. Open Research and Data Publication
7. Sensitive Research Data
8. Support, Training and Guidance
9. Research Data Appraisal

10. Data Sharing and Access
11. Cybersecurity
12. Data Ownership
13. ***Digital Preservation***
14. Funding and Sustainability
15. Governance
16. Identifiers and Metadata
17. Non-Digital Material
18. Standards and Guidelines
19. Indigenous Data Management

The Research Data Management Digital Curation Lifecycle (RDM-DCL)

Fig 5. The RDM-DCL Framework

Provides Research Librarians, research students, and ECR scholars with a 10-stage lifecycle framework that includes:

- A series of actions and tools to integrate digital curation into research data management; and

- Information and resources to learn from, and develop, digital curation as part of a professional practice.
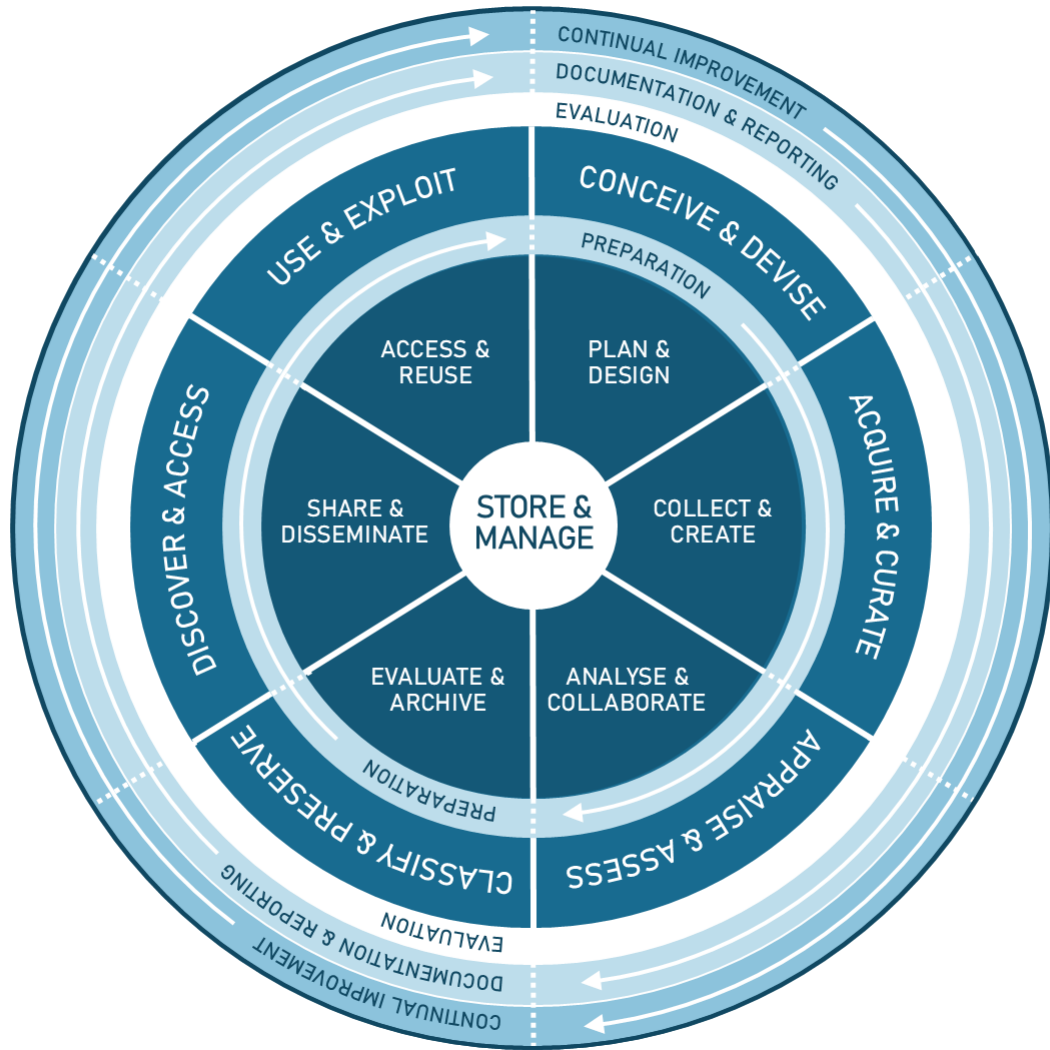
Table. 1 *RDM-DCL Framework*

RDM-DCL Stage	RDM Stage	Description	Actions	Notes	Tools
Preparation	Plan & Design Collect & Create Analyse & Collaborate Evaluate & Archive Share & Disseminate Access & Reuse	Continual preparation Organisational preparedness Awareness and Training	Research librarians undertake digital curation training or self-directed learning Assess organisational preparedness to integrate digital curation in RDM training and support Revise RDM templates to incorporate curation actions at each stage of the RDM lifecycle Develop required policies and procedures to implement RDM-DCL framework	Digital curation staff training - drawn from cross-institutional expertise and external professionals Guide students and ECRs to continually review preparation actions Understand organisational capability	DPC (Digital Preservation Coalition) Digital Preservation Handbook Rapid Assessment Model RISE Research Infrastructure Self Evaluation Framework (Rans & Whyte, 2017) Digital Streams Matrix Workflow decision making tool (Langley, 2022)
Conceive & Devise	Plan & Design	Digital curation training Review curation actions in line with research project requirements	Delivery of digital curation training to HDR students and ECRs as part of RDM planning Integrate digital curation as a core component of RDM templates or guidelines to ensure digital curation considerations are addressed at each RDM stage	Encourage HDRs and ECRs to document preservation strategies, including procedures for data backup, validation, and migration, in their RDM Plans	PRONOM Registry Information about file formats, software products and other technical components required for long-term access to data
		RMP planning incl. data retention, disposal and access.	Encourage HDRs and ECRs to select open, non-proprietary file formats whenever possible, to ensure interoperability and long-term accessibility	Explain risks associated with data loss, format and technology obsolescence, and necessity for preserving research data for future reproducibility and reuse	Read Preservation Planning (Digital Preservation Coalition, nd)

RDM-DCL Stage	RDM Stage	Description	Actions	Notes	Tools
		Institutional or funding body requirements	Introduce HDRs & ECRs to institutional and funding policies related to data management, ethical and legal requirements, rights, and preservation requirements for access, use and exploitation.	Varying ethical and legal requirements will apply based on research sensitivity and funding body policies	Read Understanding the legal implications of data sharing, access, and reuse in the Australian research landscape (Fitzgerald, Pappalardro, Austin, 2008)
		Requirements Analysis Metadata standards Data Management and Digital curation plan development	Support HDR's & ECRs to identify preservation requirements of their research data, considering factors such as data types, formats, and anticipated lifespan Select metadata standards and data documentation approaches Encourage HDRs & ECRs to integrate digital curation actions as part of their RDM plans, outlining strategies for data backup, long-term accessibility, integrity, authenticity, and usability of their research data	Recommend discipline appropriate metadata standard to describe data facilitate data discovery and interoperability Support HDRs & ECRs to document preservation strategies in their RDM Plan	Dublin Core Metadata standard for describing digital or physical resources Darwin Core Metadata standard and framework for compiling biodiversity data from varied sources PREMIS Metadata standard to support the preservation of digital objects and ensure their long-term usability METS Metadata Transmission standard for encoding descriptive, administrative, and structural metadata expressed using the XML schema language

RDM-DCL Stage	RDM Stage	Description	Actions	Notes	Tools
		Risk Assessment	Identify potential threats to the integrity and longevity of research data, such as accidental loss, file degradation, format obsolescence, technological and institutional dependencies, and changes	Plan for periodic data migration to newer formats to prevent format obsolescence and ensure continued accessibility	DiAGRAM Tool to assess and manage risks to data and digital assets DROID Automated batch identification of file formats
Acquire & Curate	Collect & Create	Metadata Creation File integrity	Guide HDR students to create metadata for their research data at time of creation Generate checksums for data and digital assets and record in metadata Instruct in the use of version control systems (e.g., Git) to track changes to research data and ensure the integrity of the dataset over time	Metadata should include descriptive, administrative, technical, preservation and rights metadata, to facilitate discovery and reuse Ensure file naming conventions are followed as part of version control	See previous metadata tools above Checksum Calculator Tool to calculate the checksum of a file or data Git Version Control Records changes to a file or set of files over time so that you can recall specific versions CSV Validator Automated metadata validation tool
		Compliance	Refer to RDM ethical, legal and project funding or Institutional requirements	Ensure compliance with ethical and legal requirements for data creation, acquisition, and storage	
Appraise & Assess	Analyse & Collaborate	Organisation, review and management of research data Authenticity	Apply quality assurance measures to verify the integrity and authenticity of data as it is analysed Convert data into standard formats for long-term preservation	Track and record changes and updates to the data as part of version control and provenance	See Git Version control above Sustainability of Digital Formats (Library of Congress)

RDM-DCL Stage	RDM Stage	Description	Actions	Notes	Tools
Classify & Preserve	Evaluate & Archive	Retention, disposition and storage policy compliance Access controls and permissions to protect sensitive or restricted data Ingest, storage and Infrastructure Monitoring and Maintenance File validation	Encourage HDRs and ECRs to follow naming convention during data creation and classification. Record checksums in metadata before uploading to university storage solution. Perform file validation, and ensure security, permissions and long-term accessibility Implement backup and replication procedures to prevent data loss Review and update preservation plans in RDM to accommodate changes in research data or need for file migration	Most universities require research data to be uploaded to their own repositories Consider data and file packaging formats, e.g., Bagit or Zip	BagIt - hierarchical file packaging format for DP See checksum calculator and CSV Validator above COPTR Multi-platform database tools for file migration and data management
Discover & Access	Share & Disseminate	Data documentation Metadata and finding aids	Guide HDR students and ECRs in the importance of metadata for discovery, access and retrieval of preserved data	Encourage data sharing and collaboration within the research community	COPTR Tools for discovery & access
Use & Exploit	Access & Reuse	Access, use and reuse of data	Provide data access to designated users, which may include access copies of file formats Ensure all legal, copyright, and ethical policies are followed	Maintain appropriate legal and ethical access controls	COPTR Tools for use and reuse

Evaluation and Improvement

Table 2. Evaluation and Improvement

| Lifecycle Stage	Description	Actions
Evaluation	Assessment of Effectiveness	Evaluate the effectiveness of digital curation integration into HDR DMPs through surveys, interviews, or feedback mechanisms from HDR students and faculty members
	Identifying Improvements	Identify areas for improvement and refine practices based on evaluation findings and emerging best practices in digital curation
Documentation and Reporting	Documentation	Document developed processes, strategies, and outcomes to provide transparency and accountability for stakeholders
	Reporting	Report on the implementation and impact of the integration of the RDM-DCL into HDR student and researchers DMPs for internal assessment and external reporting purposes
Continual Improvement	Feedback Incorporation	Incorporate feedback from HDR students, faculty members, and other stakeholders to continually refine and enhance the processes developed for the RDM-DCL framework
	Adaptation to Change	Stay abreast of evolving research practices, technological innovations, and policy developments to ensure that the integration framework remains relevant and effective over time

Benefits

- Upskilling research librarians, research students and ECRs to deal with the challenges of a quickly evolving field;

- Stronger research outcomes as data is managed, described, and more efficiently used and re-used;

- Higher levels of discovery and access to research; and

- Increased research impact and funding opportunities as research outputs become recognised and used by the wider community.

Next steps

- Download the RDM-DCL framework and paper and slides and from the VALA website when it becomes available;

- Familiarise yourself with the tools and activities and think about what training you think you may need to use them;

- Have discussions within your institution about trialing the RDM-DCL framework * as part of the support you provide to research students and ECRs; and

- Reach out to your peers in other university libraries for support, or engage external providers;

* Even a gradual integration of framework components will upskill librarians and benefit research students and ECRs.
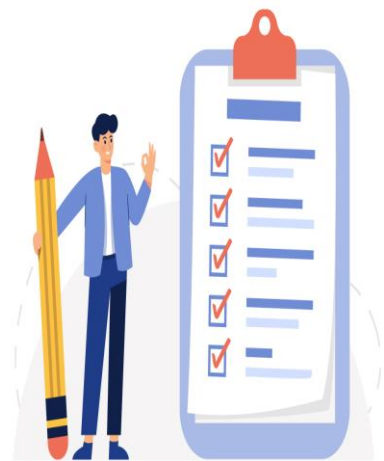
Any Questions ?

Thank you!

References

Australian Research Data Commons. (2023). Research data management framework for institutions (Version 2). https://zenodo.org/records/8433246

Consultative Committee for Space Data Systems. (2012). Reference Model for an Open Archival Information System (OAIS): ISO 14721:2012. http://public.ccsds.org/publications/archive/650x0m2.pdf

DPC Digital Preservation Coalition. (2024). What is Digital Preservation? https://www.dpconline.org/digipres/what-is-digipres

Harvard Medical School. (n.d). Longwood Research Data Management. What is Research Data Management. https://datamanagement.hms.harvard.edu/about/what-research-data-management

Hicks, R. (2023). Australian Research Data Commons: Research Data Management Framework for Institutions. http://doi.org/10.5281/zenodo.6392340

Higgins, S. (2008). The DCC curation lifecycle model. International Journal of Digital Curation 3(1):453. https://doi.org/10.2218/ijdc.v3i1.48

LMA Research Data Management Working Group. (n.d.). Research Lifecycle. Harvard University. https://researchsupport.harvard.edu/research-lifecycle

Oliver, G., & Harvey, D. R. (2016). *Digital curation (2nd ed.)*. Facet Publishing

Weatherburn, J., & Neish, P. (2018, February). Demystifying digital preservation: taking action with a capability maturity model. In Proceedings of the VALA Preserving the Future 2018, Libraries, Technology, and the Future Conference, (pp. 1-14). https://doi.org/10.6084/m9.figshare.6169553.v1