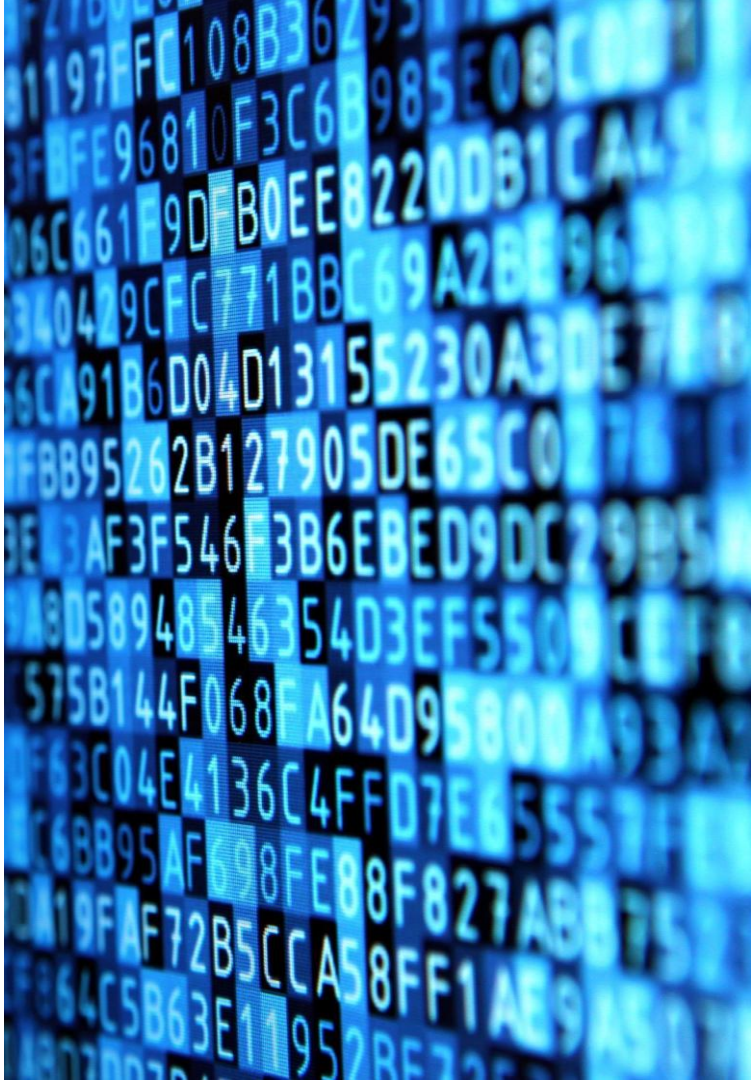
Your future self will thank you

Encouraging good data management practices at project closure

Dr Vanessa Crosby, Executive Manager, Research Services



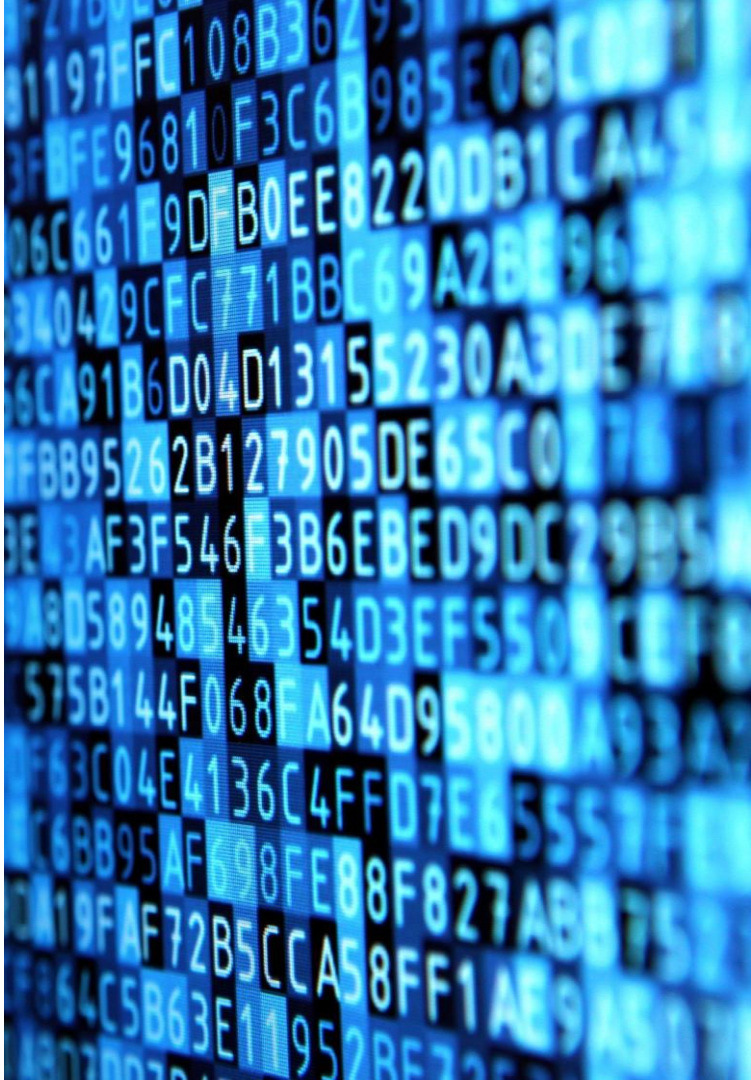
Or, how to get people to do boring things.

Your future self will thank you

Encouraging good data management practices at project closure

By an ADHDer who once spilled computer coolant liquid on the external hard drive containing the sole copy of her PhD data.

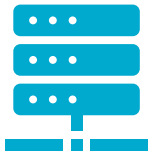
Dr Vanessa Hoskyn, Executive Manager, Research Services



CSIRO Research Data at-a-Glance

Active storage (Cloud object storage)

- Includes auto-tiering and lifecycle management to optimise costs.
- Storing ~ 8.47 PB and growing ~ 10 to 20% annually.



Active storage (on-premises)

- NetApp FAS systems provide high-speed access for 350+ active projects.
- Storing 19.27 PB of active research data.

Tape-based archives

- Long-term storage for post-project data
- 336 million files archived in first six months of operation
- Storing ~ 40 PB currently

Data Access Portal

- Open Access Data Repository
- 11,521 published data collections
- 5,683.87 TiB


THE SUPER-DOOPER REINDEER POPPER
diamond
STRIKE ON BOX MATCHES
LARGE KITCHEN MATCHES
250
COUNT
2292


2292
Diamond
250
COUNT
STRIKE ON BOX MATCHES
LARGE KITCHEN MATCHES


OD 236404

Project Goals

- Research data management seen as a **core** part of research, not an admin or compliance activity
- Research data routinely managed as an organisational asset
- Automation and capture of essential metadata at point of provisioning through the Resource Management System

Assess data holdings

Publish shareable data

Archive underpinning data according to retention requirements

Change location to a new storage space for data that will be used in future projects

Delete working data, duplicates and scratch

"How long do I need to retain the raw data? Funder and publisher guidelines all say something different."

"I only had a small role in this project. I can't recall the names of anyone apart from the leader of the group and it seems he's no longer at CSIRO."

I'm listed as admin, but I have no idea what the data is – it definitely doesn't belong to me ☹

"We need somewhere to store the software just in case the client misplaces their copy or there's a question

"Is there a classification tool to help figure out where to store sensitive data?"

"This storage space has at least a decade's worth of data from different commercial and research projects."

"We may have to do some metadata prep to publish the data. That may take a little time."

What detective work have you had to do?

What were the funder requirements for Open Data?

How was this data collected? There's no-one here from the project team any more

How long do we need to keep this for?

Who owns the IP?

Can I reuse this data in another project?

What constraints around this Human Subject data were in the Ethics Approval?

How do I open and read this weird file format?

What did the contract say about sharing the data?

Where's that dataset? It would be really useful for this new project!

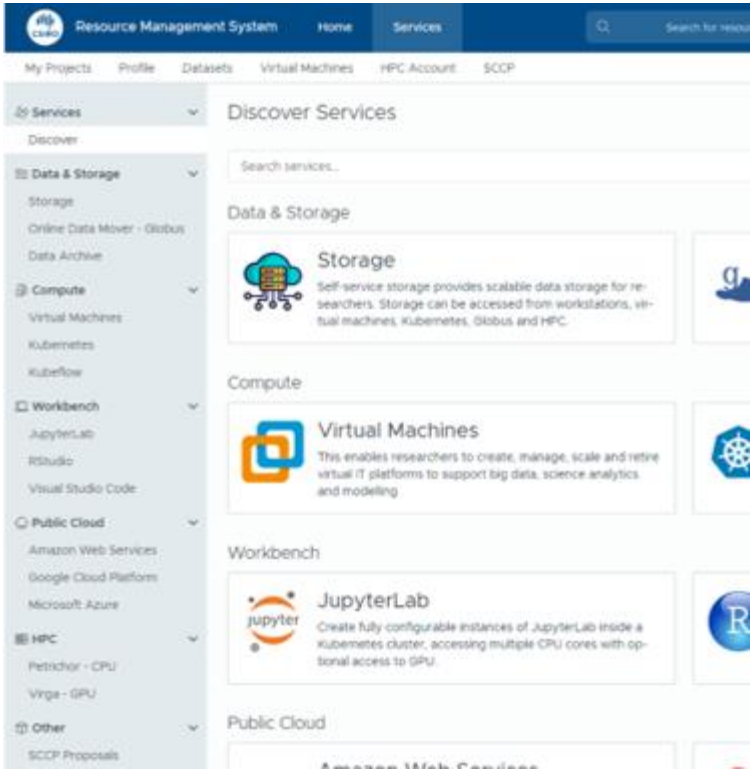
Project Closure data Management: Preliminary Activities

ID	Preliminary Activities	YES	NO
1	Have project data storage spaces been cleaned up for files that are no longer required?	☐	☐
2	Have project data and outputs been audited to determine what it is, where it is located and size? *Refer to Research Project Closure Guide: Data Management for places where data can be located	☐	☐
3	Have data and outputs been assessed for retention?	☐	☐

Artist's impression of retention and disposal schedules

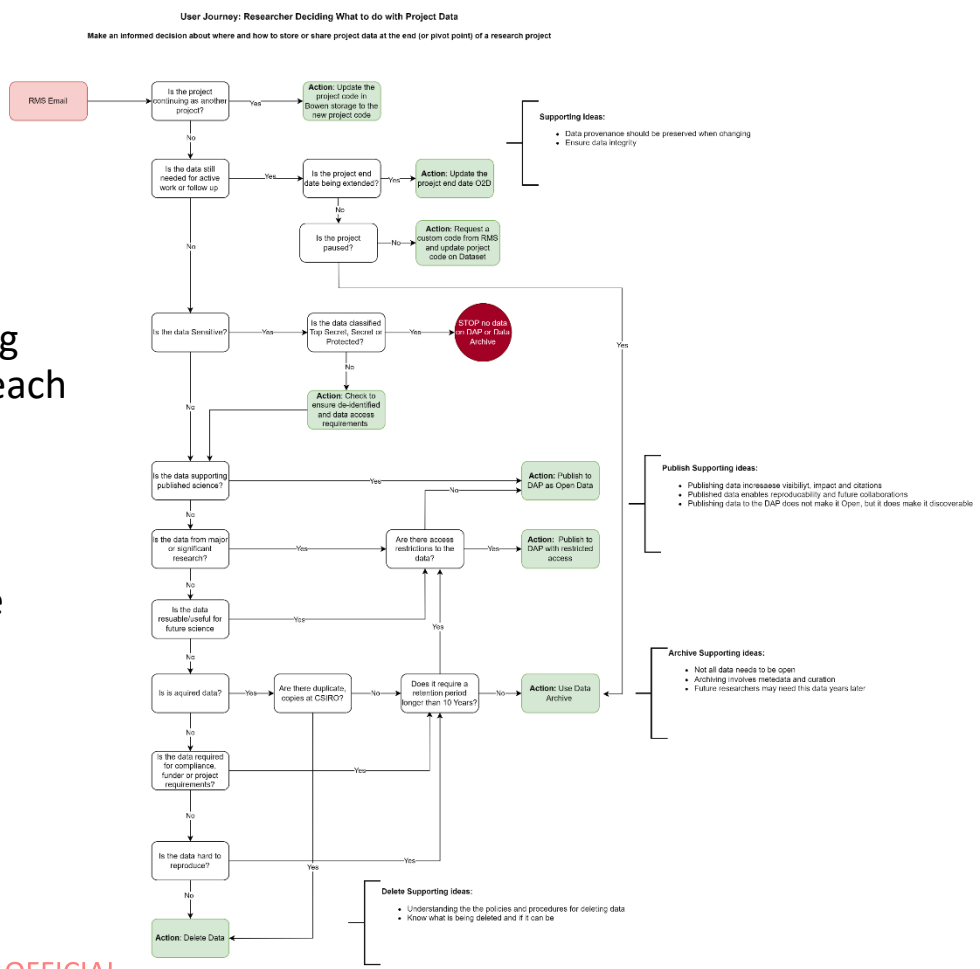
RDM by Design

- Identify systems researchers are motivated to engage with and build in compliance and metadata capture.
- Identify natural checkpoints in the research process where users interact with systems e.g. requests for increases in data storage quotas.
- Incremental improvement over perfection!



Journey Mapping

- Map the complete user journey from email trigger, through systems and guidance.
- Identify "what's in it for me" decision-making points and technical information needed at each stage.
- Create navigation paths with clear "if/then" scenarios.
- Determine compelling reasons that will drive user decisions.



- Have clear “What’s in it for me?” messages targeted to different groups
- Capture key metadata at natural points in the research data management cycle, ideally at the point of provisioning not in a separate system
- Start where you are - small, low-tech changes can have a big impact e.g. prototyping automated email messages with users
- Apparently, memes are only used by unfashionable Gen X-ers and Elder Millennials, which really cramped my change management style.

Project Team:

Julia Anticev (Project Manager)
Paul Crawley, Research Data Specialist
Katie Hannan, Senior Research Data Specialist
Adam Kennedy, Process Lead, Research Technology Solutions
Ryan Meintjes, Team Leader, Research Data Services
Sharmila Rahman, Research Data Advisor
Christina Street, Research Data Specialist
With support from the Data Storage and Protection Team, Enterprise Information Management Team, Digital Office and RMS Project Team

Questions