



MONASH
University

Research data tools and resources update

Research Data Strategy

Office of the PVC-RI

18 September 2023



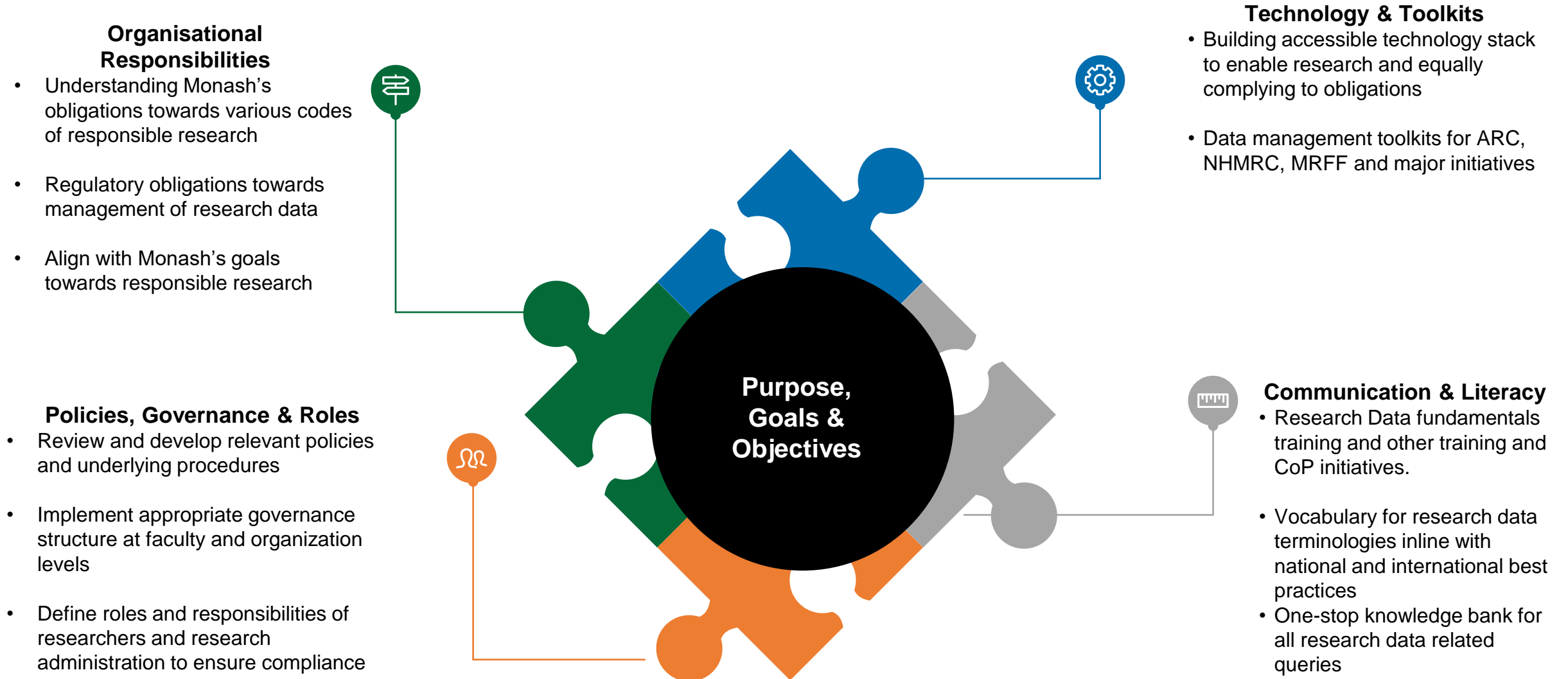


I wish to acknowledge the people of the Kulin Nations, on whose land we are gathered today. I pay my respects to their Elders, past and present.



Addressing the complexity of research data

Range of programs to tackle the complexity



Research data governance post-presentation survey

- Lack of clarity on research data governance **roles and responsibilities**
- Challenges related to **data loss, breach and financial penalties**
- Growing need for **data storage and sharing environments**
- Lack of visibility into **digital research capabilities and support services**
- Lack of faculty **visibility over researcher projects and data**
- Clear guidelines and assistance in processes of **data retention and disposal** are required
- Staff and PhD student **exit procedures** don't clearly address research data handover
- Research Data Management and Data Governance **training is absent or ad hoc**
- Constraints in **faculty access to Ethics applications**
- **Impact on research** has not been sufficiently considered in **technology decisions**

—| **Plan for Research Data**
Research Data website

—| **Plan for Research Data**
Data Classification Tool

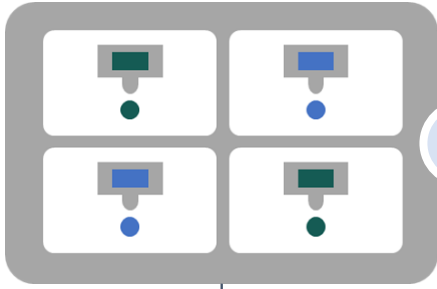
—| **Research Data Solutions**
Data Classification Tool

—| **Plan for Research Data**
Research Data website

—| **Plan for Research Data**

—| **Research Data Solutions**
Catalogue of Research Data
Services

Foundational Work in Research Data Governance



1 Research Data Catalogue

Where is the Research Data and is it being handled appropriately?

Plan for Research Data

Where is **your** Research Data?

CoRDS

Whose Research Data is on this system/ service?

2

3

4

5

6

Research Data Classification Tool

What **Classification** is **your** Research Data?

Research Data Solutions

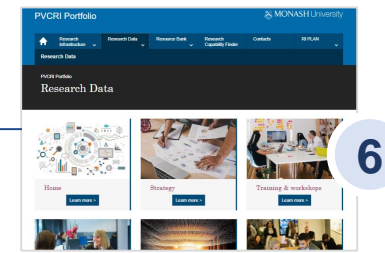
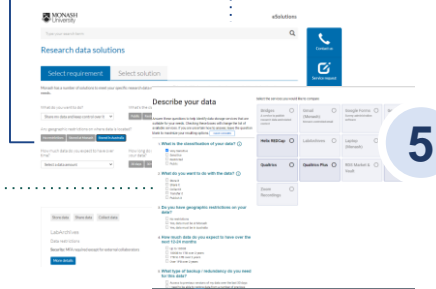
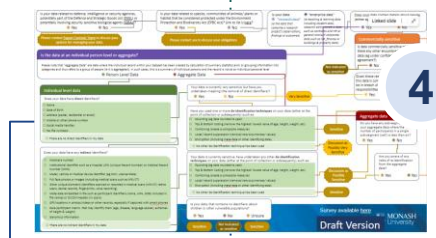
What are the **solutions available** for my Research Data, given my needs?

Comms Strategy

Development of PVC-RI intranet site for research data governance

Catalogue of Research Data Systems/ Services (CoRDS)

#	System Name	Form	Provider	Description
CRD0001	RDS: Market	Digital	MeRC	Market is a storage service that leverages eResearch file storage. Metadata is collected via the Data Vault.
CRD0002	RDS: Vault	Digital	MeRC	Vault is a storage service that leverages eResearch file storage for long term storage of collections. Researchers are on-boarded.
CRD0003	RDS: MURDA	Digital	MeRC	The Monash University Research Data Archive (MURDA) is a collector long-term repository of research data leveraging eResearch infrastructure. Metadata is collected via the Data Dashboard when researchers are on-boarded.
CRD0004	RDS: Object	Digital	MeRC	Object is a storage service that leverages eResearch (Research Data) file storage for unstructured data. Metadata is collected via the Data on-boarded.
CRD0005	RDS: Compute	Digital	MeRC	Compute is a storage service that leverages eResearch (Research Data) active file storage that is used with NECTAR VMs. Metadata is collected when researchers are on-boarded.
CRD0006	M3 HPC	Digital	MeRC	High performance computing data cluster. Also known as MASSIVE (cluster). Files are stored as follows: Home & Utility directories are infrastructure. The backup of their working is on 'RDS: Vault'. The files available for use on M3 HPC are their Scratch storage area.
CRD0007	MonARCH HPC	Digital	MeRC	Monash Advanced Research Computing Hybrid for High Performance is kept via a dashboard. Files are stored as follows: Home & Utility of infrastructure. The backup of their working is on 'RDS: Vault'. The files available for use on MonARCH into their Scratch storage area.
CRD0008	Store.Monash (MyTardis)	Digital	MeRC	Metadata is about data itself and not about collection.
CRD0009	MBI-XNAT	Digital	MBI	Monash Bio Imaging (MBI) instance of XNAT - an open-source imaging dedicated to imaging-based research. Imaging data is stored on RDS Nectar Cloud (Monash).
CRD0010	MeRC-XNAT	Digital	MeRC	Monash eResearch (MeRC) instance of XNAT - an open-source imaging dedicated to imaging-based research. Data is stored on RDS: Market.
CRD0011	Nectar	Digital	MeRC	Australian Research Cloud computing offered by the Australian Research Monash hosted node.



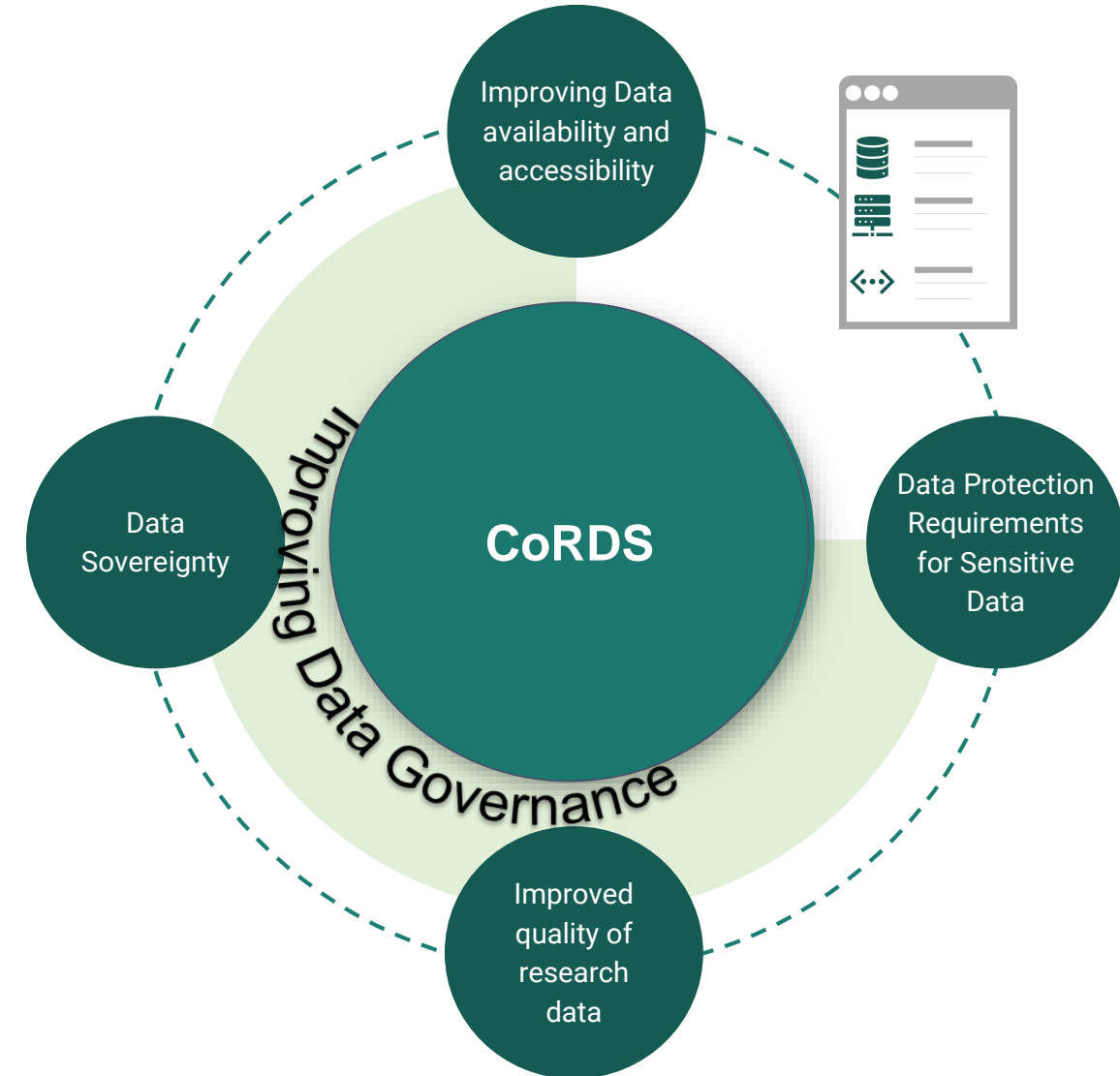
CoRDS



The concept of a Catalogue of Research Data Systems/ Services (CoRDS) is to identify the Monash controlled services (or systems) where research data resides

Benefits include:

- ▶ Awareness of how systems are being used to ensure security and data protection/ privacy requirements are suitable
- ▶ Improve visibility of emerging needs and requirements of researchers
- ▶ Opportunity to collect metadata from systems when access/ allocation of storage is given
- ▶ Single source of information to create resources for research data governance (training, guidance, analysis of usage, etc)



New Research Data website (in progress)

The screenshot displays the PVCRI Portfolio website's 'Research Data' section. At the top, a blue navigation bar features the 'MONASH University' logo and a menu with items: Home, Research Infrastructure, Research Data, Resource Bank, Research Capability Finder, Contacts, and RI PLAN. Below this, a dark grey header contains the text 'PVCRI Portfolio' and 'Research Data'. The main content area is a grid of six cards, each with a representative image, a title, and a 'Learn more >' button. The cards are: Home (with a data visualization icon), Strategy (with a meeting photo), Training & workshops (with a workshop photo), Services & support (with a service desk photo), DATA scheme (with a data visualization photo), and About us (with a meeting photo).

PVCRI Portfolio

MONASH University

Research Infrastructure Research Data Resource Bank Research Capability Finder Contacts RI PLAN

Research Data

PVCRI Portfolio

Research Data

Home [Learn more >](#)

Strategy [Learn more >](#)

Training & workshops [Learn more >](#)

Services & support [Learn more >](#)

DATA scheme [Learn more >](#)

About us [Learn more >](#)

CoRDS - Catalogue of Research Data Systems/ Services



1

Which Data Systems/ Services hold research data?

Identify the Monash controlled services (or systems) where research data resides.



Catalogue of Research Data Systems/ Services (CoRDS)

LAST UPDATED 25/07/2023

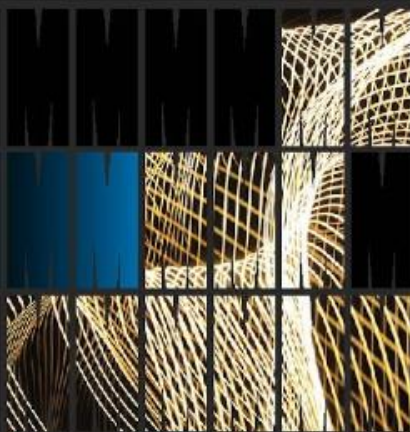
Source: [Research Data Catalogue](#)

#	System Name	Form	Provider	Description
CRD0001	RDS: Market	Digital	MeRC	Market is a storage service that leverages eResearch (Research Data file storage). Metadata is collected via the Data Dashboard when rese
CRD0002	RDS: Vault	Digital	MeRC	Vault is a storage service that leverages eResearch (Research Data S file storage for long term storage of collections. Metadata is collecte researchers are on-boarded
CRD0003	RDS: MURDA	Digital	MeRC	The Monash University Research Data Archive (MURDA) is a cold-stor long-term retention of research data leveraging eResearch (Research infrastructure. Metadata is collected via the Data Dashboard when re additional governance processes ensure long-term oversight of the di
CRD0004	RDS: Object	Digital	MeRC	Object is a storage service that leverages eResearch (Research Data t file storage for unstructured data. Metadata is collected via the Data on-boarded
CRD0005	RDS: Compute	Digital	MeRC	Compute is a storage service that leverages eResearch (Research Dat active file storage that is used with NECTAR VMs. Metadata is collect researchers are on-boarded
CRD0006	M3 HPC	Digital	MeRC	High performance computing data cluster. Also known as MASSIVE; dashboard. Files are stored as follows: Home & Utility directories are infrastructure. The backup of their working is on 'RDS: Vault'. The fil available for use on M3 HPC into their Scratch storage area.
CRD0007	MonARCH HPC	Digital	MeRC	Monash Advanced Research Computing Hybrid for High Performance is kept via a dashboard. Files are stored as follows: Home & Utility dir infrastructure. The backup of their working is on 'RDS: Vault'. The fil available for use on MonARCH into their Scratch storage area.
CRD0008	Store.Monash (MyTardis)	Digital	MeRC	Metadata is about data itself and not about collection
CRD0009	MBI-XNAT	Digital	MBI	Monash Bio Imaging (MBI) instance of XNAT - an open-source imagin dedicated to imaging-based research. Imaging data is stored on RDS Nectar Cloud (Monash)
CRD0010	MeRC-XNAT	Digital	MeRC	Monash eResearch (MeRC) instance of XNAT - an open-source imagin dedicated to imaging-based research. Data is stored on RDS: Market
CRD0011	Nectar	Digital	MeRC	Australia's Research Cloud computing offered by the Australian Reses Monash hosted node.

Catalogue of Research Data Services (CoRDS)

DEMO

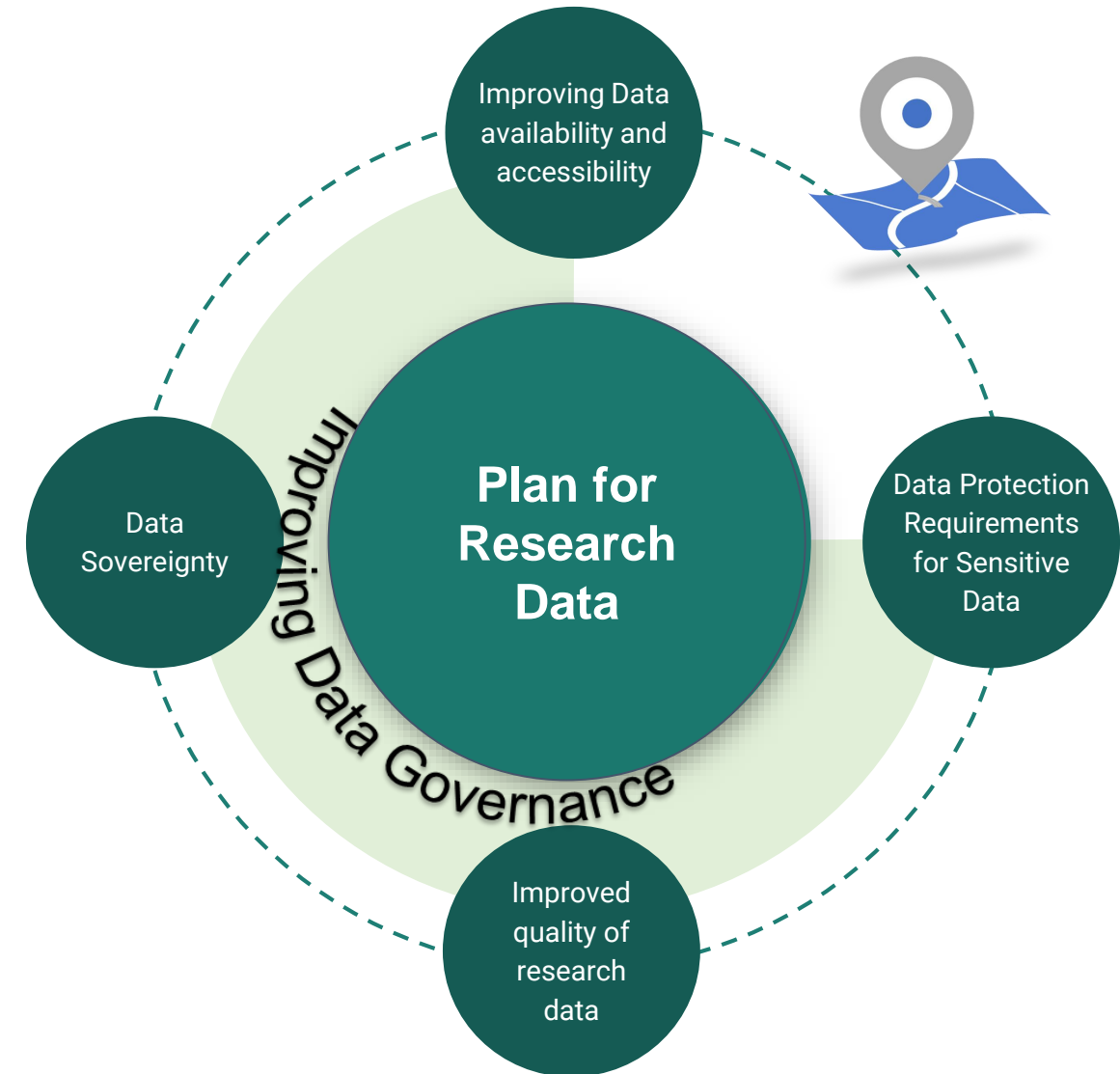
Plan for Research Data



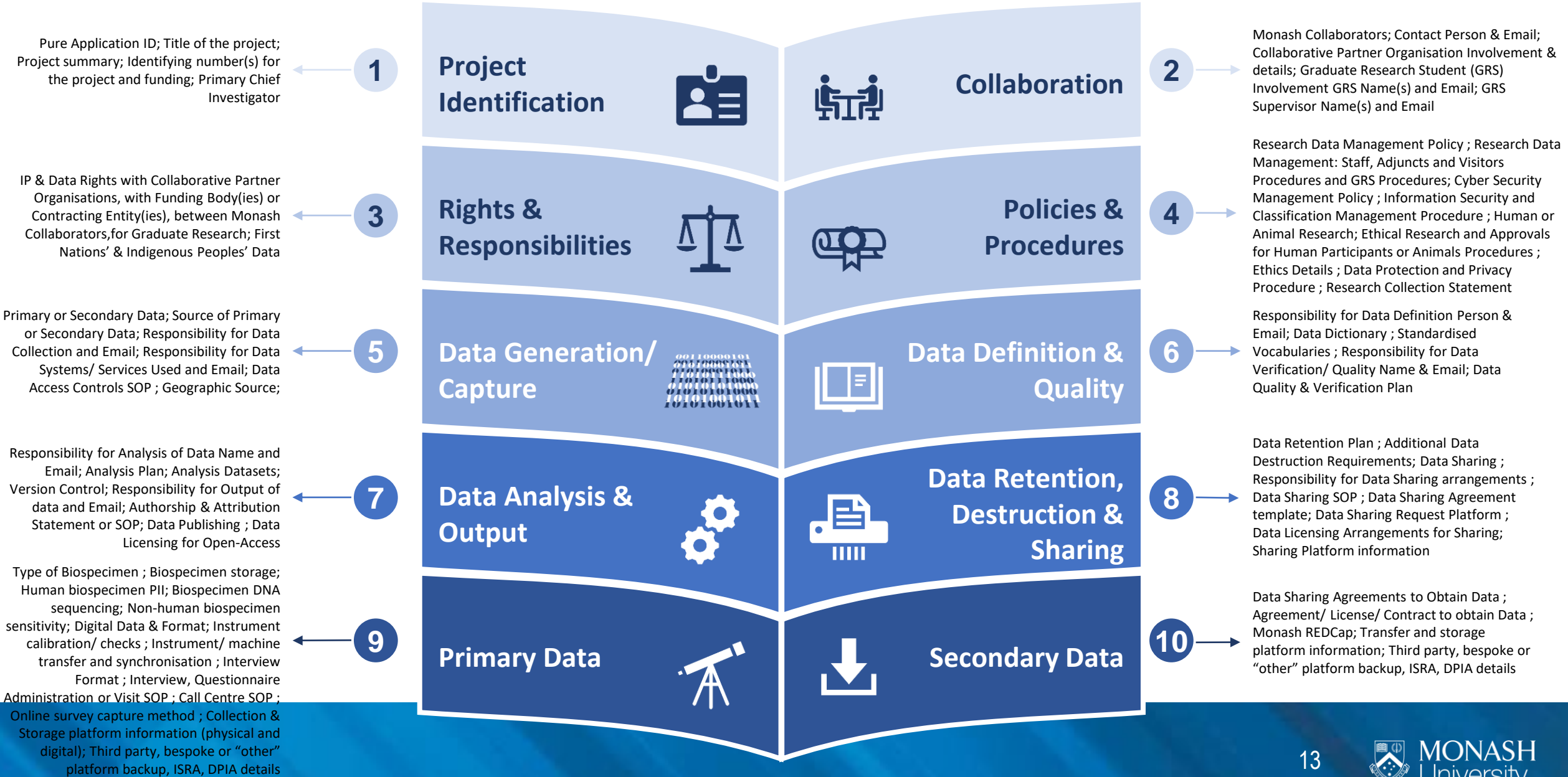
Plan for Research Data

Benefits include:

- ▶ A standard format for recording how research data is managed at Monash, allowing overview by Faculties and Schools
- ▶ Providing junior researchers and PhD candidates an overview of the research project
- ▶ Providing a mechanism for researchers to comply with funder and publisher requirements
- ▶ Standardised reporting for Monash on:
 - Rates of compliance
 - Completion differences across faculties
 - Timelines of completion
 - etc



Plan Content



Four key principles of Plans for Research Data

Structured

Data in DMP needs to be collected in as structured a format as possible

- Allows for useful reporting and analysis
- Ease of compliance

Updatable

Researchers need to be able to review and update their DMP easily

- Allows for annual review
- Provides key resource within project for data governance

Usable

DMPs should be highly usable - guiding researchers to make good decisions

- “Self” guiding to information needed to make good data management choices
- Specific tools to inform decisions

Reusable

Wherever possible, data is reused from other systems and pre-populated

- Minimise re-typing/ re-entering of data into DMP by uploading existing information wherever possible

Report to MRC, DVC-R, Faculties, Schools ...

Structured

ILLUSTRATIVE

1. Purpose of Data

Search by Identifiers

Title of Project

PURE Id

Chief Investigator

Faculty_School...Dept

CrossRef/ DataCite

ORCID

Project Identification

Other Search

Start Date

End Date

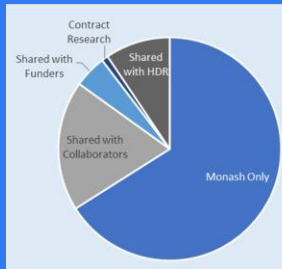
Ethics Type

Data Type

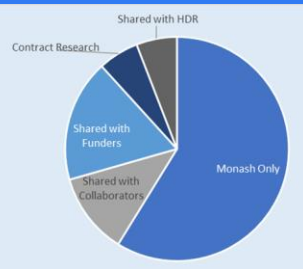


IP & "Ownership"

By Project



By \$ Funding



Decision Making Authority

Proportion of Projects



Policy & Procedure

% DPP Procedure Applicable

51%

% HREC Procedure Applicable

71%

% HDR Procedure Applicable

13%

Has Data System Access SOP

36

Has Data Sharing SOP

21

Has Authorship/ Attribution SOP

11



2. Data Element Management

Missing Data Dictionary

12

Processes

3. Data Capture

5. Data Repository Management

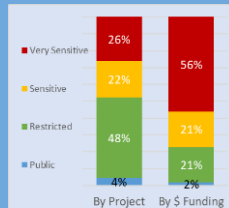
BackUp Routine Unknown

8

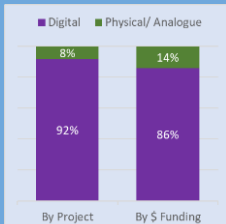
Data Retention Unknown

14

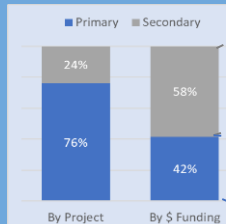
Data Classification



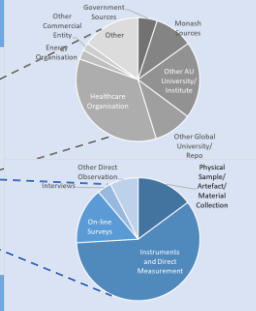
Data Form



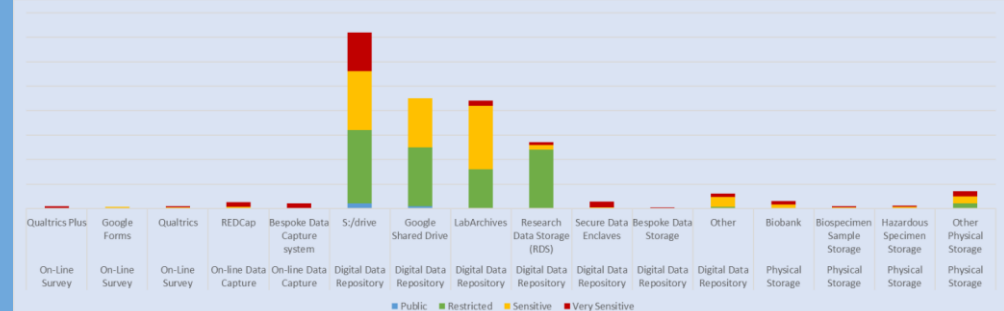
Data Type



Source of Data



Systems/ Services Used



Plan for Research Data DEMO

Research Data Classification Tool

Is your data related to defence, intelligence or security agencies, or a primary part of the Defence and Foreign Security Act 2003, or primarily handling security sensitive information under 2003?
 Yes No

Does your data involve records concerning defence or security agencies?
 Yes No

Is your data related to police, community or personal safety or health care by a Commonwealth or State or Territory law enforcement or health care provider?
 Yes No


Is your data "integrity data" in respect of a service or activity provided by a Commonwealth or State or Territory law enforcement or health care provider?
 Yes No

Does your data involve records concerning defence or security agencies?
 Yes No

Is data commercially sensitive or is there any other circumstance in which data is confidential or otherwise agreed to?
 Yes No

Given these circumstances, would the data be considered to be in breach of the legal & regulatory responsibilities?
 Yes No

Are you aware of any other circumstances in which the data is confidential or otherwise agreed to?
 Yes No

Survey available from  Draft Version

Is the data at an individual person level or aggregate?
 Individual Person Level Data Aggregate Data

Individual Person Data

Does your data have been identified?
 Yes No

Does your data have an identifier?
 Name Date of birth Address (postal, residential or work) Mobile or other phone number Social media profiles ID or the number There are no identifiers in my data

Does your data have an identifier?
 Identifier number Reference number such as a Health Care Unique Number or Health Record Identifier (HRI) Other number or unique access identifier (eg. PIN, control pass) Full or partial or unique identifying medical or health identifier (eg. ICD, other clinical identifier, diagnosis code, laboratory test result, medical device product name, date, serial number, registration, case number) Other unique identifier (eg. full or partial name, date of birth, email address, the number of a person's record in a system) Any other identifier (eg. unique number, email address, IP address, address of a health facility) Geographic information There are no identifiers in my data

Are you using one or more de-identification techniques on your data either at the point of collection or subsequently?
 Masking (eg. redacted or greyed out) Tag & Bucket coding (where the higher number value of tags, height, weight, etc.) Changing dates to approximate values Using pseudonyms (where the pseudonyms are used) Other de-identification techniques have been used

Are your data in a format that is not identifiable by other de-identification techniques on your data either at the point of collection or subsequently?
 Tag & Bucket coding (where the higher number value of tags, height, weight, etc.) Changing dates to approximate values Using pseudonyms (where the pseudonyms are used) Other de-identification techniques have been used

Is your data being classified, either internally or otherwise, as confidential or otherwise agreed to?
 Yes No Unsure

Aggregate Data

Does your data have an identifier?
 Identifier number Reference number such as a Health Care Unique Number or Health Record Identifier (HRI) Other number or unique access identifier (eg. PIN, control pass) Full or partial or unique identifying medical or health identifier (eg. ICD, other clinical identifier, diagnosis code, laboratory test result, medical device product name, date, serial number, registration, case number) Other unique identifier (eg. full or partial name, date of birth, email address, the number of a person's record in a system) Any other identifier (eg. unique number, email address, IP address, address of a health facility) Geographic information There are no identifiers in my data

Are you using one or more de-identification techniques on your data either at the point of collection or subsequently?
 Masking (eg. redacted or greyed out) Tag & Bucket coding (where the higher number value of tags, height, weight, etc.) Changing dates to approximate values Using pseudonyms (where the pseudonyms are used) Other de-identification techniques have been used

Are your data in a format that is not identifiable by other de-identification techniques on your data either at the point of collection or subsequently?
 Tag & Bucket coding (where the higher number value of tags, height, weight, etc.) Changing dates to approximate values Using pseudonyms (where the pseudonyms are used) Other de-identification techniques have been used

Is your data being classified, either internally or otherwise, as confidential or otherwise agreed to?
 Yes No Unsure



How sensitive is your research data?

Information Security and Classification Management Procedure (6 April 2023)

Security Objective	Impact	Information Classification and Labelling
Confidentiality	Catastrophic (A)	Information that may result in a Catastrophic or Major impact from disclosure is classified as Very Sensitive. (C = 4)
	Major (Ma)	
	Moderate (Mo)	Information that may result in a Moderate impact from disclosure is classified as Sensitive. (C = 3)
	Minor (Mi)	Information that may result in a Minor impact from disclosure is classified as Restricted. (C = 2)
	Insignificant (I)	Information that may result in an Insignificant impact from disclosure is classified as Public. (C = 1)

Is your data related to defence, intelligence or security agencies, potentially part of the Defence and Strategic Goods List (DSGL) or potentially involving security sensitive biological agents (SSBAs)?

Yes No

Please contact [Export Controls Team](#) to discuss your options for managing your data.

Is your data related to species, communities of animals/ plants or habitat that be considered protected under the Environment Protection and Biodiversity Act (EPBC Act)? Link to list is [here](#)?

Yes No

Please contact Ethics & Compliance Team to discuss your obligations

Is your data:

"research data" (ie the data that comprise a research project's observations, findings or outcomes)

"enterprise data" (ie teaching & learning data including student data, research administrative data such as contracts and HR or general Monash corporate data such as HR, finance or buildings & property data)

Does your data contain details about people (either aggregated or at a person level)?

Yes No

Commercially sensitive

Is data commercially sensitive or are there any other encumbrances on the data (eg under confidentiality or other agreement?)

No Yes

Not indicated as sensitive

Given these restrictions/ encumbrances, if this data is compromised, would Monash be in breach of its legal & reg responsibilities?

Yes No

Is the data at an individual person level or aggregate?

Please note that "Aggregate Data" are data where the individual record within your dataset has been created by calculation of summary statistics and/ or grouping information into categories and thus refers to a group of people (ie it is aggregated). In such cases, this is a summary of individual persons and the record is not at an individual personal level

Person Level Data

Aggregate Data

Individual level data

Does your data have **direct** identifiers?

- Name
- Date of Birth
- Address (postal, residential or email)
- Mobile or other phone number
- Social media handles
- Tax file numbers

There are no direct identifiers in my data

Does your data have any **indirect** identifiers?

- Medicare number
- Institutional Identifier such as a hospital URN (Unique Record Number) or Medical Record Number (MRN)
- Meter, Vehicle or medical device identifier (eg NMI, License Plate)
- Full face photos or images (including medical scans such as MRI/CT)
- Other unique biometric identifiers scanned or recorded in medical scans (MRI/CT, retina scans, dental records, fingerprints, voice recording)
- Meta-data embedded in file such as participant identifiers (name, URN, DOB) included in file names or DICOM headers (in scans)
- GPS locations in photos/videos or other records, especially if captured with smart phones
- Rare participant metric that may identify them (age, disease, language spoken, extremes of height & weight)
- Genomics information

There are no indirect identifiers in my data

Your data is currently very sensitive, but have you undertaken masking (the removal of direct identifiers)?

Yes No

Very Sensitive

Have you used one or more **de-identification techniques** on your data (either at the point of collection or subsequently) such as:

- Rounding (eg date rounded to year)
- Top & Bottom Coding (remove the highest/ lowest value of age, height, weight, etc)
- Combining (create a composite measure)
- Local/ record suppression (removal rare occurrences/ values)
- Encryption (including meta-data or other identifying data)

No other de-identification technique has been used

Sensitive

Discussion as Possibly Very Sensitive

Your data is currently sensitive, have undertaken any other **de-identification techniques** on your data (either at the point of collection or subsequently) such as:

- Rounding (eg date rounded to year)
- Top & Bottom Coding (remove the highest/ lowest value of age, height, weight, etc)
- Combining (create a composite measure)
- Local/ record suppression (removal rare occurrences/ values)
- Encryption (including meta-data or other identifying data)

No other de-identification technique has been used

Discussion as Possibly Sensitive

Sensitive

Aggregate data

Do you have any sub-segments in your aggregate data where the number of participants in a single sub-segment (cell) is less than six?

Yes No

Are you aware of any risks of re-identification from the aggregate data?

Yes No

Is your data, that contains no identifiers, about children or other vulnerable populations?

Yes No Unsure

Sensitive

Not indicated as sensitive

Sensitive

Survey available [here](#)

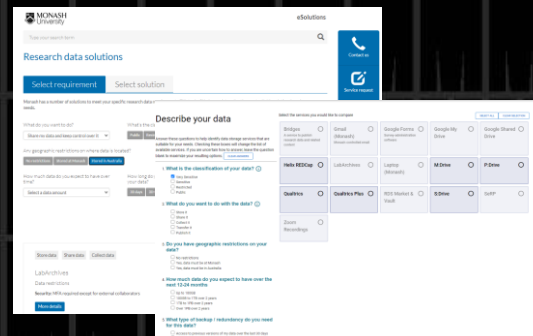
Draft Version



How Sensitive is Your Research Data?

DEMO

Research Data Solutions Tool



I know my data classification, now what?

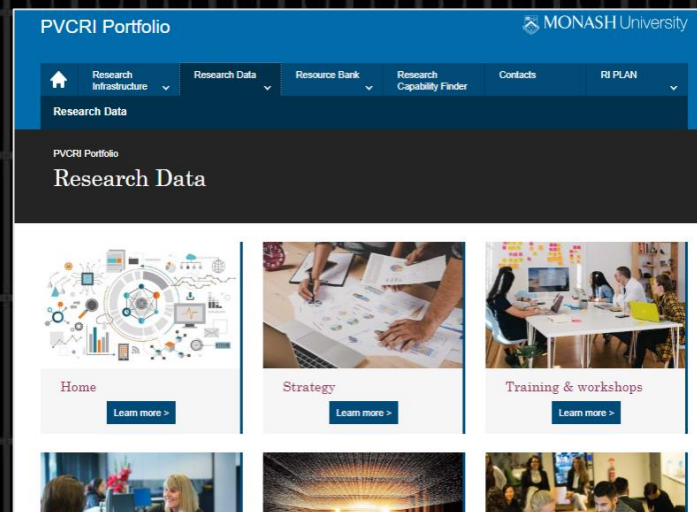
Research data considerations:

- What do you want to do with your research data?
 - Store, share, collect, transfer, publish
- What is your data classification
 - Public, restricted, sensitive, very sensitive
- Are there any geographic restrictions?
 - None, at Monash, in Australia
- Do you need multi-factor authentication?
 - Yes, no
- How much data will be generated?
 - Up to 100GB, 100GB to 1TB, 1TB to 1PB, 1PB+
- How long do you need to be able to access and restore your data?
 - Up to 30 days, 30 days+
- Will you be accessing Australian Commonwealth data?
 - Yes, no

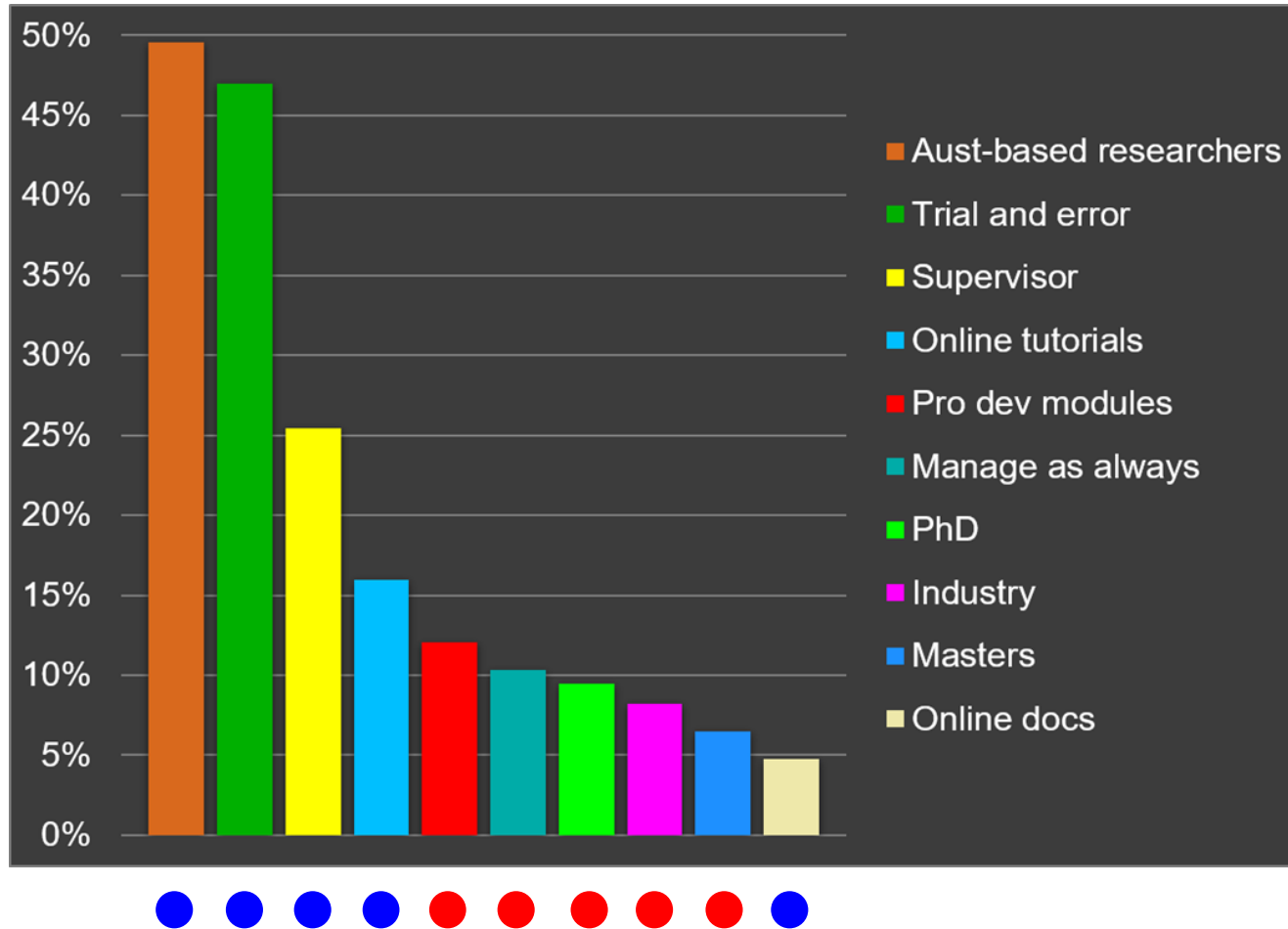


Research Data Solutions DEMO

Research Data Website



New Research Data website (in progress)



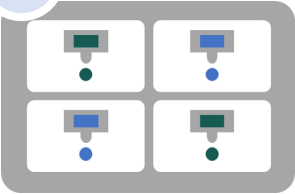
Research data information is highly dispersed:

- eSolutions
- Monash eResearch Centre
- Monash Research Office
- Monash University Library
- Multiple intranets

Research Data Website DEMO

Next steps

1 Research Data Catalogue



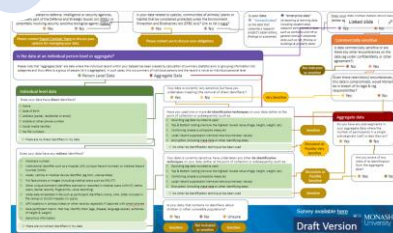
2 CoRDS



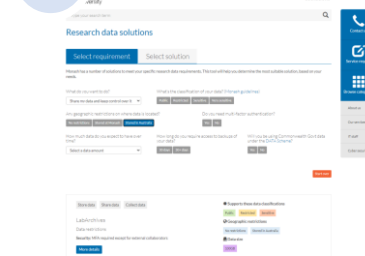
3 Plan for Research Data



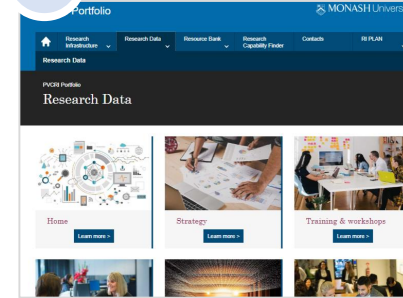
4 Research Data Classification Tool



5 Research Data Solutions Tool



6 Comms Strategy



If you become aware of systems/ services that researchers are using, **let us know**

Phased testing after piloting - **we will contact you directly**

Available for feedback now, please contact us if you'd like to be involved in testing

Initial eSol page available for **feedback shortly**, please contact us if you'd like to be involved in testing.

Available for **feedback shortly**, will contact you when it is live

Overall feedback on RDG activities...

Spread the word....

Contact us

Komathy Padmanabhan

Director, Research Data Strategy

Komathy.Padmanabhan@monash.edu

ResearchDataGovernance@monash.edu

Dianne Brown

Research Data Governance Consultant

Dianne.Brown@monash.edu

Patrick Splawa-Neyman

Research Data Governance Specialist

Patrick.Splawa-Neyman@monash.edu