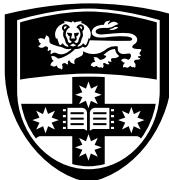


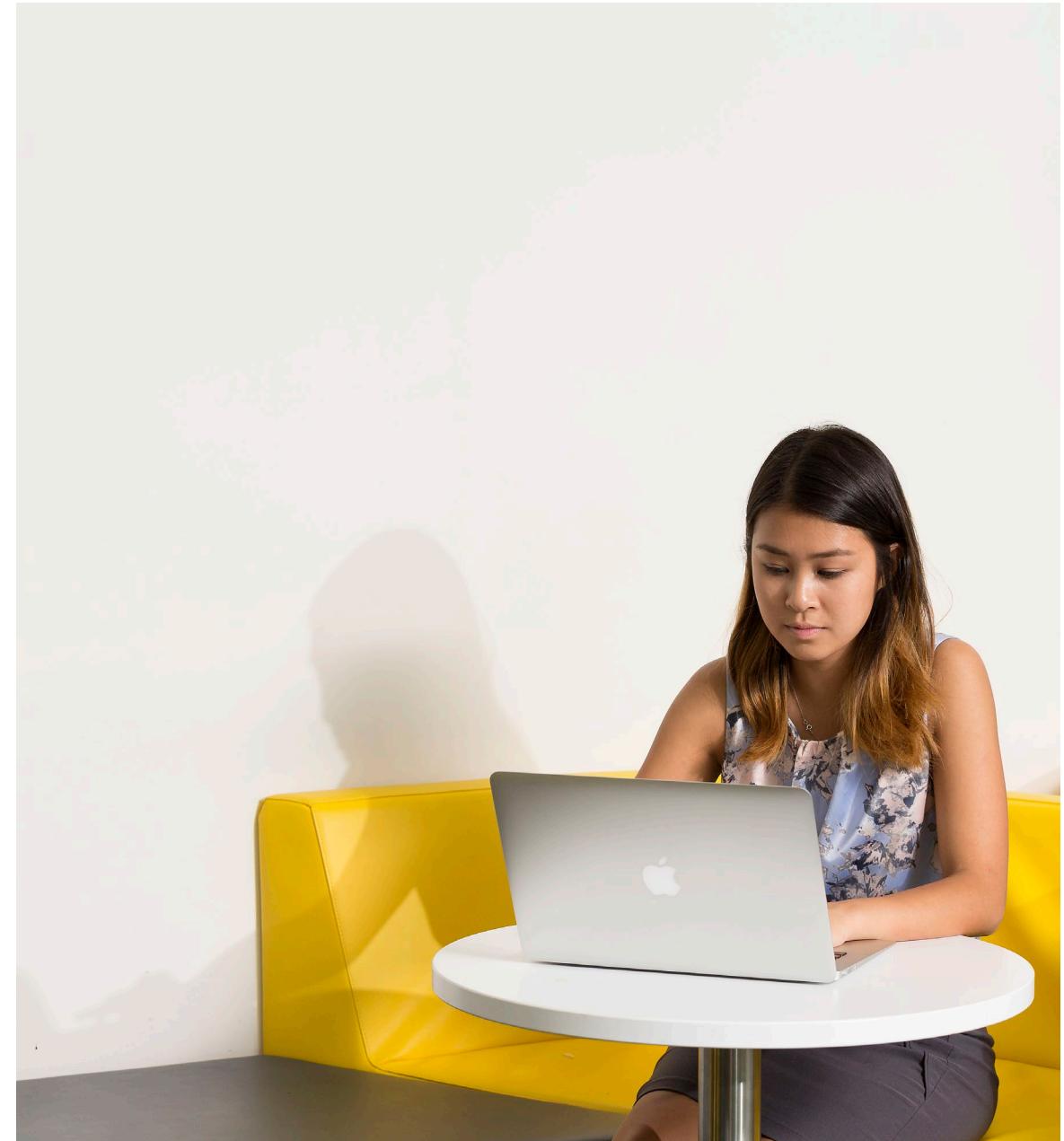
# Research Data Management

Enhancing your research with  
University-supported digital tools

Research Data Consulting team  
Sydney Informatics Hub



THE UNIVERSITY OF  
**SYDNEY**



# What is research data?

## Research data

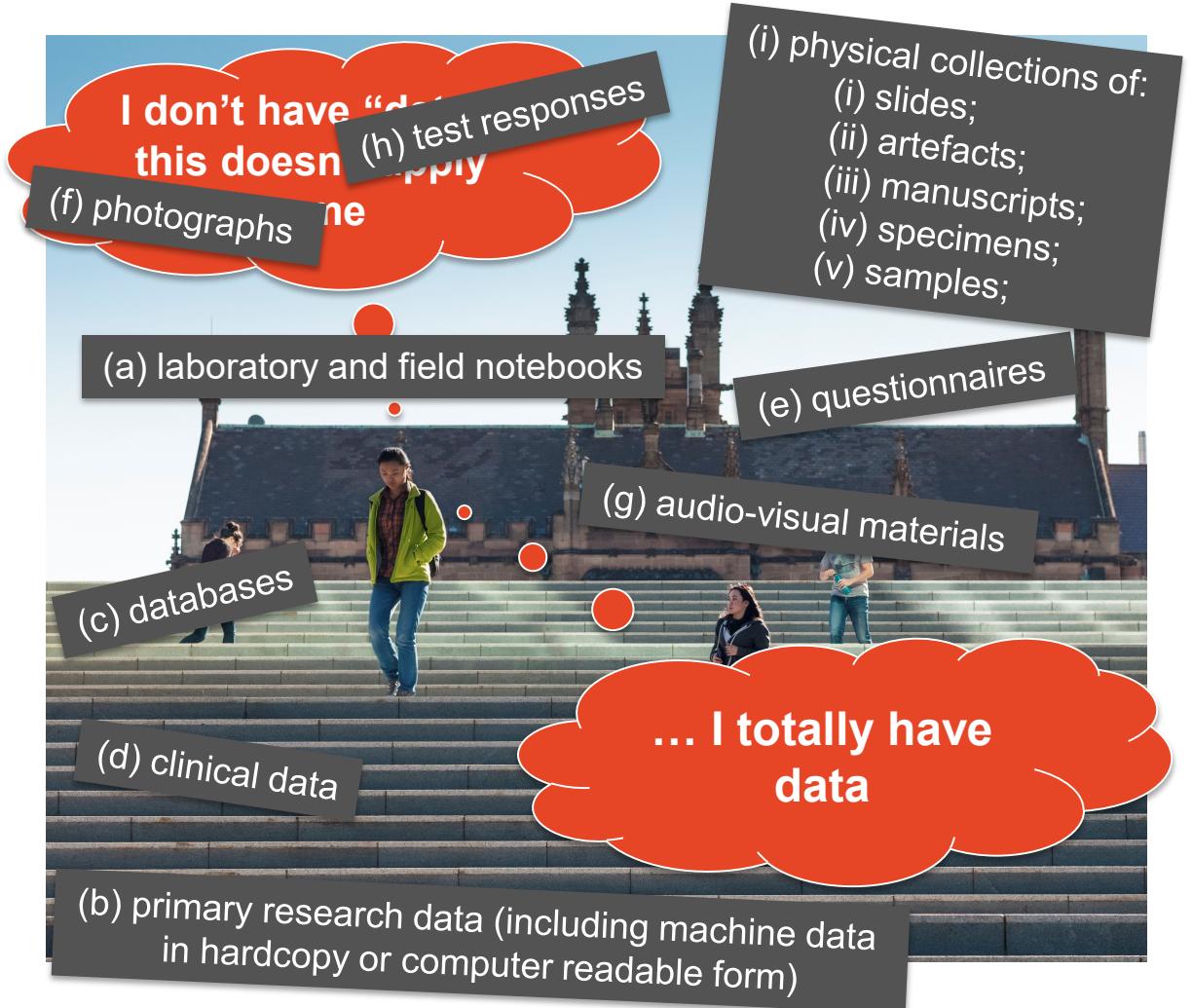
any information needed to validate original research findings.

Focus on *digital* research data

## Research data management (RDM)

the ways in which research data is collected, documented, organised and stored.

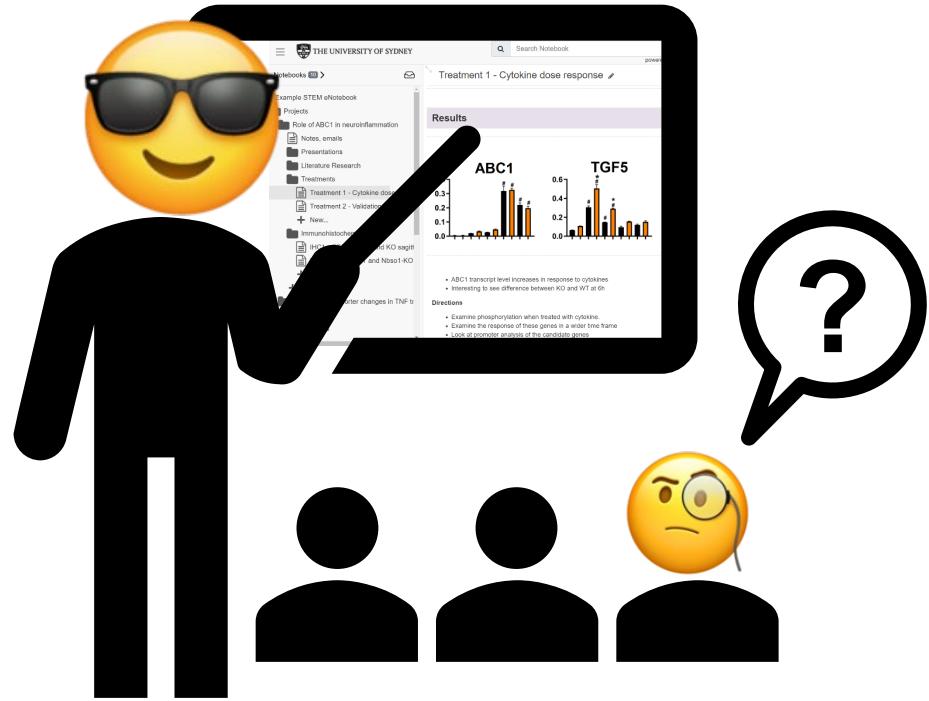
*Research Data Management Policy 2014*  
“Research data may include...”



# **Best practice RDM involves achieving some specific goals**

The University provides researchers with  
**tools and services** to help you achieve these

**“Ok great ... but what’s in it for me?”**



You are presenting at a conference.

Your favourite researcher has a question, but you don't have the data to answer it in your presentation.

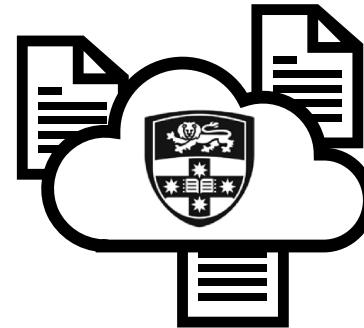
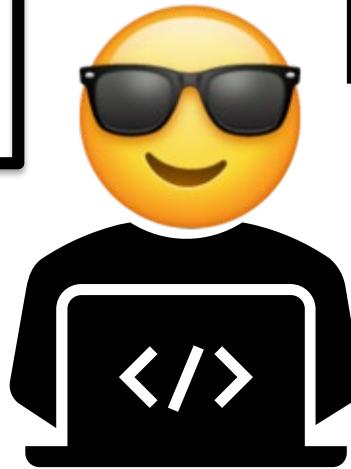
Your data is on the eNotebook, so you can quickly find it and show them. It is the beginning of a beautiful collaboration.

Research data should be  
**Accessible**  
for those who need it

Accessing remotely but securely.

Finding things easily.

Access to different parts for different people.



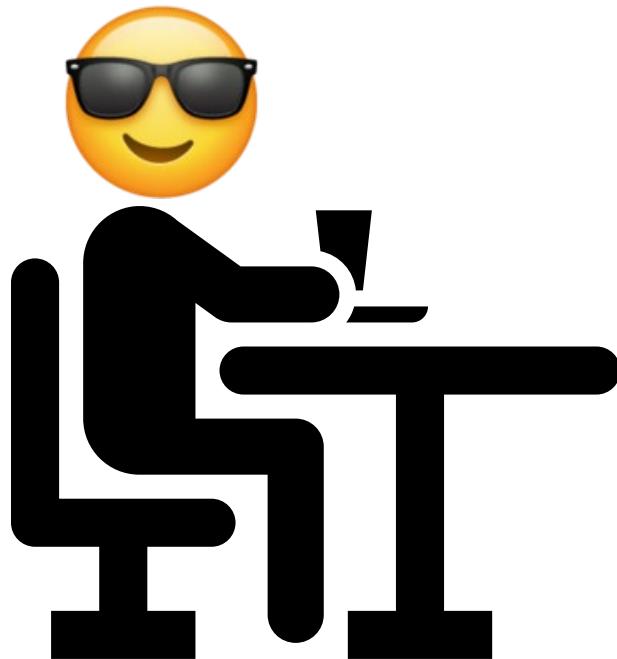
You open your laptop to do some world-class research.

Your hard drive has failed.

Your data is on a University-supported platform, so you never lost any of your research.

Research data should be  
**Secure**  
and protected from loss

Protection against loss and corruption.  
Automatic backup.  
Easy recovery.



After a day of working with ‘Highly Protected’ data on

- patients’ health
- commercially-sensitive research

you head out for dinner with friends

Your laptop goes missing

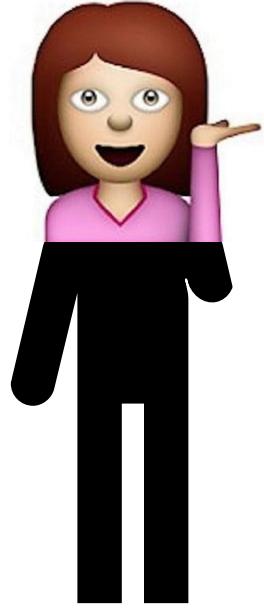
Your laptop and files are encrypted. Also, the data was never stored on your computer.

Research data should be  
**Secure**  
and protected from unauthorised access

An understanding of data sensitivity and the University’s data classifications.

Data kept on secure platforms.

Encryption of devices and individual files.



Date and Time	Entry version #	Revised by	Revised by IP	Revision Action
Feb 03, 2020 @04:17 PM AEDT	14	Taylor Syme	129.78.56.129	file uploaded or modified by Partner application
Feb 03, 2020 @04:15 PM AEDT	13	Taylor Syme	129.78.56.129	file uploaded or modified by Partner application
Feb 03, 2020 @03:20 PM AEDT	12	Taylor Syme	129.78.56.129	file uploaded
Jan 31, 2020 @04:26 PM AEDT	11	Taylor Syme	129.78.56.133	file uploaded or modified by Partner application
Jan 23, 2020 @04:42 PM AEDT	10	Taylor Syme	129.78.56.140	file uploaded
Jan 16, 2020 @03:37 PM AEDT	9	Taylor Syme	129.78.56.130	file uploaded or modified by Partner application
Jan 16, 2020 @03:29 PM AEDT	8	Taylor Syme	129.78.56.130	file uploaded
Jan 16, 2020 @03:28 PM AEDT	6	Taylor Syme	129.78.56.130	moved to entry
Jan 16, 2020 @03:28 PM AEDT	7	Taylor Syme	129.78.56.130	moved from entry
Jan 16, 2020 @11:29 AM AEDT	5	Taylor Syme	129.78.56.130	file uploaded or modified by Partner application
Jan 16, 2020 @10:23 AM AEDT	4	Taylor Syme	129.78.56.130	file uploaded or modified by Partner application
Jan 16, 2020 @10:19 AM AEDT	3	Taylor Syme	129.78.56.130	file uploaded or modified by Partner application
Jan 15, 2020 @05:25 PM AEDT	2	Taylor Syme	129.78.56.130	file uploaded or modified by Partner application
Jan 15, 2020 @05:03 PM AEDT	1	Taylor Syme	129.78.56.130	file uploaded or modified by Partner application



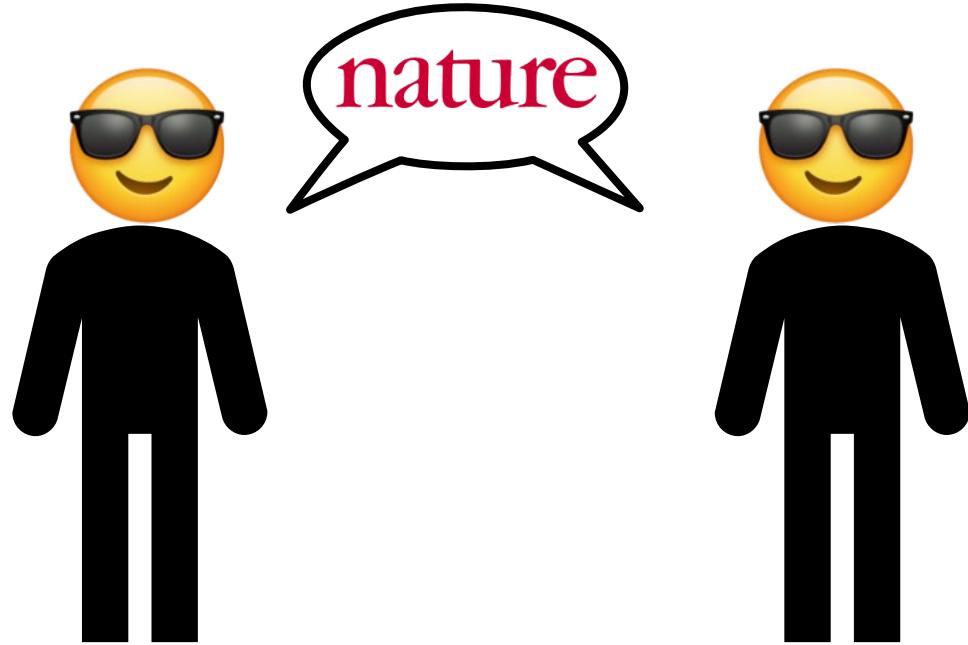
You've submitted a paper.

A reviewer accuses you of fabricating the data and there is talk of academic dishonesty proceedings.

You have an audit trail from the eNotebook proving that *you* did the research *when* you said you did it and *how* you said you did it. The accusation is quickly dealt with.

Research data should be  
**Attributed**  
to keep track of who did what and when

Clear audit trail of the changes made, when they were made, who made them.  
Maintenance of data provenance.  
Version history allows you to “go back” to any previous revision.



Your former research group wants to include your data in an upcoming publication.

It has been years and you can barely remember a thing about it.

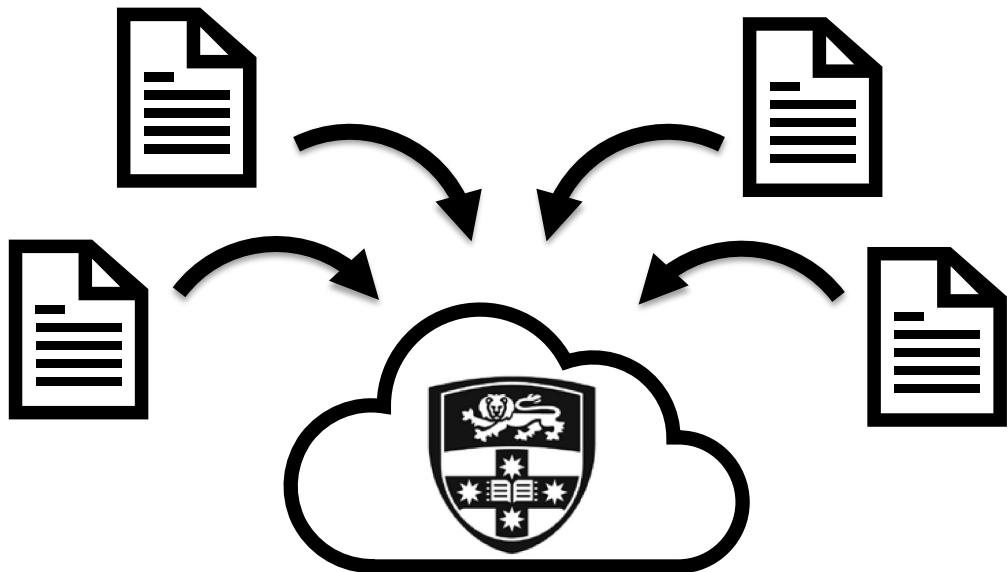
You kept the data together with info on why the data was created, what the methodology was, how the data was collected, so you can all understand the old research.

## Research data should be **Contextualised**

to form a narrative so that research can be understood

Providing a rich background for the data: Why? How? What?  
Data is self-explanatory and able to be understood by others.

Research data should be  
**Consolidated**  
to keep it all together



Data located in as few places as possible makes it easy to keep it...

- accessible.
- contextualised.
- consistent.

Storing data on University-supported platforms means it is...

- secure.
- backed up.

# Why manage research data?

## To ensure it is

- Accessible, secure, attributed, contextualised, consolidated

## To increase research quantity and quality

- save time with efficient & modern processes
- enhanced collaborations inside and outside of own research group

## To prevent costly mistakes

## To prevent loss

- loss of knowledge and data when researchers leave

## To fulfil obligation to research participants

- keeping personal & health information safe

## To maintain integrity

Since that comment, his published scientific work has been under intense international scrutiny.

Now, six research papers he co-authored have been withdrawn or retracted from publication due to unresolved concerns over missing or manipulated data.

Behind the scenes, a series of high-stakes confidential inquiries and secret reports has threatened to derail his stellar career.

Late last year after more questions from journal editors about other images in the paper, the authors withdrew the research, admitting they couldn't locate the raw data needed to prove their work.

In an unusual statement, the journal noted that despite conducting an independent investigation clearing Professor of research misconduct, " was unable to locate any electronic record of the original images and was unable to locate any records of the images in original lab books."

"Based on the unresolvable concerns with these figures, the editors, therefore, hereby retract the article."

ABC News, 4 Dec 2019



Analysis.xlsx



Analysis  
re-done.xlsx



Analysis  
re-done -  
final.xlsx



Analysis  
re-done -  
final (2).xlsx

# Why manage research data?

**Because you have to (!)**



**For example:**

“Detailed descriptive metadata ... should be associated with qualitative and quantitative data”

*Research Data Management Policy 2014 (USYD)*

“Data and datasets created as part of research activities...Retain minimum of 5 years after project completed, then destroy”

*Education: Higher & further education and research records (GA47)*

“The ARC [requires] researchers to outline how they plan to manage research data arising from ARC-funded research”

*Research Data Management – Australian Research Council*

“Nature backs the Enabling FAIR Data initiative and requires authors to deposit data in community repositories”

*Editorial policies – Nature Research*

# The Research Data Management Declaration

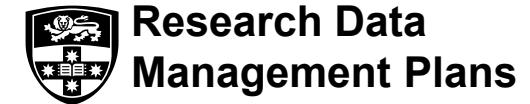
- A concise summary of your responsibilities when managing your research data as set out in University policy
- Can be found at <https://dashr.sydney.edu.au/agreements>

- Intellectual Property**
  - All information should have author attribution and time/date stamping
  - There should be a full revision history of all items
  - If you are working collaboratively, an agreement about who owns the intellectual property of the data should be made in writing
- Curation and Contextualisation**
  - Your records/data must be structured in a form that enables easy interpretation by others, and contain descriptive information that provides context to the data
- Storage and Security**
  - Data must be managed on University-supported platforms
  - Health data containing identifiers must be kept on servers located in Australia
  - Data/research records must be securely stored and be accessible upon request for the duration of the minimum retention period that applies to your data
  - You must understand and adhere to the requirements regarding **protected/highly protected data**
  - Sensitive data must be encrypted at rest and in transit, to standards specified by **ICT Security**
- Collaboration and Off-Boarding**
  - Chief Investigators or their nominees must be in full control of the permissions/access given to collaborators
  - When leaving the University, a copy of your data should be left at the University and a current staff member should be nominated as data custodian
- Data sharing and Publication**
  - Data should be **made openly available** where possible, usually when research is published
  - If data are confidential or sensitive, then either a **de-identified version of the dataset** or a metadata-only record should be published

# University-supported digital tools/platforms

**Q:** Now that you know your responsibilities, *how* are you going to fulfil them?

**A:** The University provides tools, platforms, and services to assist you!



**... and more!**

[sydney.edu.au/research-data-platforms](http://sydney.edu.au/research-data-platforms)

# Some of Sydney's Platforms

# eNotebook

A web-based electronic notebook for recording your research process, data, documents and collaborative interactions in one place

The screenshot shows a web-based eNotebook interface for 'Treatment 2 - Validation of treatment response'. The left sidebar lists notebooks, projects, treatments, and other lab resources. The main content area has sections for 'Rationale' (dose responsiveness of target genes), 'Method' (RNA extraction procedure), and 'Results' (bar charts for ABC1, TGF5, and GAF). A 'Directions' section at the bottom contains notes about transcript levels and differences between KO and WT at 6h.

**Rationale**

- Determine the dose responsiveness of target genes to treatment with cytokine
- Determine optimal dose to use to induce all genes without using excessive amounts

**Method**

RNA was extracted following standard lab procedure. Briefly, for RNA extraction from cell culture, media was removed, TRI Reagent was added directly to the cells and the lysate transferred to microcentrifuge tubes and processed according to the manufacturer's instructions. For RNA extraction from tissue, organs were homogenised in TRI Reagent and processed according to the manufacturer's instructions. Extracted RNA was dissolved in RNase-free TE buffer. Purity and concentration of RNA was assessed using a Nanodrop-2000 spectrophotometer.

**Results**

ABC1, TGF5, GAF

Directions

- ABC1 transcript level increases in response to cytokines
- Interesting to see difference between KO and WT at 6h

Runs in browser, securely, from Sydney

The University of Syd

Example STEM eNotebook - Lab/ au-mynotebook.labarchives.com

THE UNIVERSITY OF SYDNEY

Notebooks 7 >

- Example STEM eNotebook
- Projects
- Role of ABC1 in neuroinflammation
  - Notes, emails
  - Presentations
  - Literature Research
  - Treatments
    - Treatment 1 - Cytokine dose
    - Treatment 2 - Validation of tr
    - + New...
  - Immunohistochemistry
    - + New...
  - Glutamate transporter changes in
    - + New...
- Lab Resources
  - + New...
- Deleted Items

Flexible structure

Rationale

Add Entry Rich Text Heading Attachment More

Treatment 2 - Validation of treatment response

Searchable

Method

Links

RNA was extracted following standard lab procedure. Briefly, for RNA extraction from cell culture, media was removed, TRI Reagent was added directly to the cells and the lysate transferred to microcentrifuge tubes and processed according to the manufacturer's instructions. For RNA extraction from tissue, organs were homogenised in TRI Reagent and processed according to the manufacturer's instructions. Extracted RNA was dissolved in RNase-free TE buffer. Purity and concentration of RNA was assessed using a Nanodrop-2000 spectrophotometer.

Results

GEL

180408\_RPA53\_2d-\_Phosphor\_\_1.gel(73.4 MB) Phosphorscreen scan after 2 day exposure. Scanned using Typhoon FLA 9500. GEL image file produced can be opened with ImageJ.

ABC1

Condition	Value
WT	~0.02
KO	~0.02
WT + Cytokine	~0.32
KO + Cytokine	~0.22
WT + 6h	~0.22
KO + 6h	~0.22

TGF5

Condition	Value
WT	~0.05
KO	~0.05
WT + Cytokine	~0.55
KO + Cytokine	~0.35
WT + 6h	~0.10
KO + 6h	~0.10

GAF

Condition	Value
WT	~0.05
KO	~0.05
WT + Cytokine	~2.8
KO + Cytokine	~2.8
WT + 6h	~1.5
KO + 6h	~1.5

Directions

Searchable

Shareable

Research data kept together with context

Author attributed, time stamped, versioned

Page Tools

Taylor Syme • Aug 28, 2019 @10:25 AM AEST

Taylor Syme • Aug 28, 2019 @10:26 AM AEST

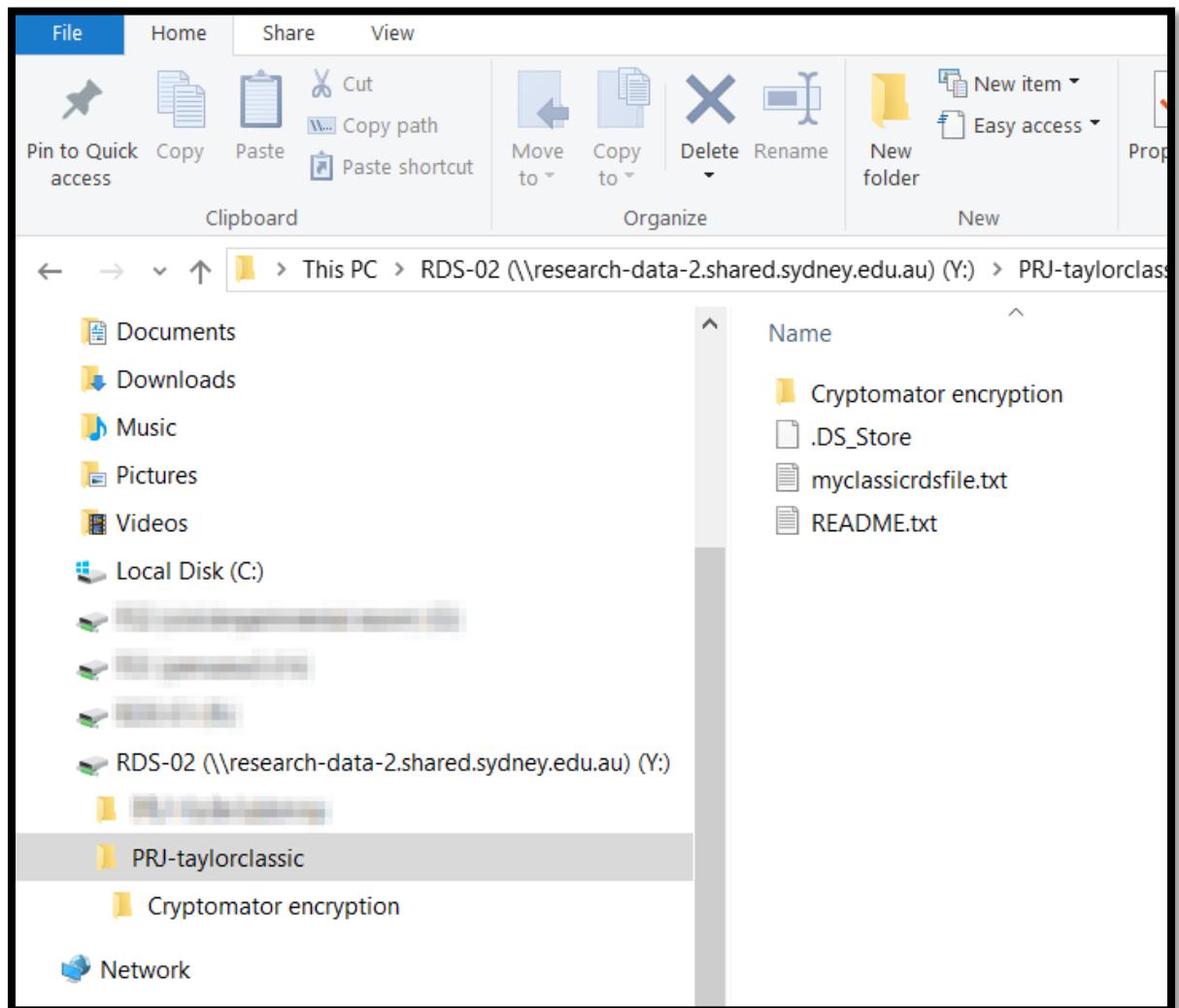
Taylor Syme • Aug 28, 2019 @10:26 AM AEST

Taylor Syme • Aug 28, 2019 @10:29 AM AEST

# Research Data Store (RDS)

A centralised network drive specifically for research

- Ideal for large files and/or series of files
- Access
  - Managed via DashR  
(dashr.sydney.edu.au)
  - Off-campus requires use of VPN
- Security
  - Backed up daily
  - Disaster recovery
- Considerations
  - No automatic version control
  - Contextualization requires README files



# REDCap

A secure, web-based database for collecting clean data

The screenshot displays two browser windows side-by-side. The left window shows the 'Data Collection' interface for a 'Test Project'. It includes a sidebar with project management links like 'My Projects', 'Control Center', and 'Record Status Dashboard'. The main area shows a form for 'Editing existing Participant ID 1' with fields for 'Event Name' (Final Data), 'Participant ID' (1), and 'Has patient completed study?' (Yes). The right window shows a 'Patient Morale Questionnaire' survey with questions about pill intake, dependence level, and satisfaction levels for various aspects of life.

**Runs in browser, securely, from Sydney**

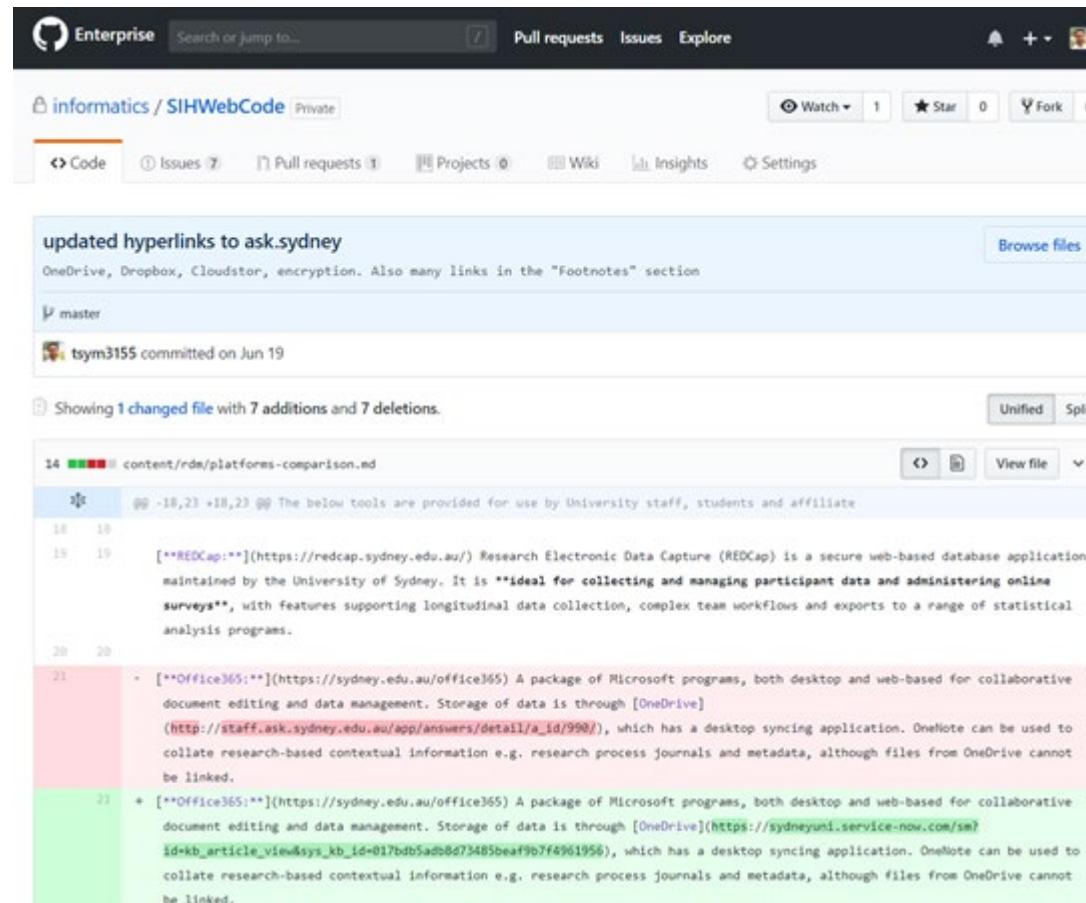
Data entry forms  
(req. researcher to log-in)

Defined field types and validation  
minimise data entry errors

Surveys (via public URL  
or email invite)

# GitHub

A web-based Git repository management platform using an enterprise edition of GitHub. Ideal for code management, version control, review, and collaboration.



**Services Portal:** [What is the University code repository?](#)

# Quick guide to storing and managing your research data

## University supported and licensed platforms

Platform/Tool	eNotebook	REDCap	RDS (via DashR)	OneDrive/ Sharepoint (Enterprise)	Dropbox (Enterprise)	CloudStor	Imaging Data Service (IDS)	Unsuitable as primary storage for research data	Prohibited for protected research data
								local storage (laptop, desktop, USB drives)	other cloud tools (e.g. GoogleDrive, SurveyMonkey, personal Dropbox)
function	electronic notebook	survey and data capture, including Clinical Trials	networked data storage, large files, HPC access	cloud storage	cloud storage	large file transfer, cloud storage	imaging repository and analytics	removable media, local storage	cloud storage
suitable for data classification	● ● ●	● ● ●	● ● ●	+ ● ●	● ● ●	● ●	● ●	●	●
stored in Australia	✓	✓	✓	✓	✓	✓	✓	various	✗
external collaborator access	✓	✓	✓	✓	✓	✓	✓	✗	✗
context and commentary supported	✓	✗	✗	✓	✓	✗	✗	✗	✗
syncing with local copy	not applicable	not applicable	not applicable	✓	✓	✓	not applicable	not applicable	not applicable
unlimited storage	✓	✓	✓	3 TB	✓	1TB	✓	✗	✗
backup and disaster recovery	✓	✓	✓	✓	✓	✓	✓	✗	✗
audit trail/ version control	✓	✓	✓	✓	✓	✓	✓	✗	✗
versioning retained	✓	manual	limited	7 years	10 years	1 month	✗	✗	✗

Version 1.0

Published May 2020



Endorsed by CIO



THE UNIVERSITY OF  
**SYDNEY**

### Data classification key:

- highly protected
- protected
- public
- + highly protected data needs additional file encryption

Highly Protected data may require additional encryption depending on the platform. Protected data may benefit from encryption.

For more information about research data classifications, go to [sydney.edu.au/research-data-classifications](https://sydney.edu.au/research-data-classifications)

For research data management enquiries, email Sydney Informatics Hub [digital.research@sydney.edu.au](mailto:digital.research@sydney.edu.au)

# “Why shouldn’t I use [platform X]?”

- Maybe you can, but you likely have to consider...
  - Adherence to policy and expectations
  - Maintenance of research and data integrity
  - Retention periods
  - Access – both control of and your own
  - Cybersecurity
  - Data ownership
  - Data sovereignty
  - Attribution and versioning
  - Privacy and ethics
- University-supported platforms have already addressed these issues 

# Classifying your research data

Types of research data	Highly Protected*	Protected	Public
data containing health or private information with personal identifiers	data classified as <b>Highly Protected</b>		
data subject to regulatory or contractual controls	data classified as <b>Highly Protected</b>		
data containing culturally or environmentally sensitive information	data classified as <b>Highly Protected</b>		
data containing personal or confidential information		data classified as <b>Protected or higher</b>	
unpublished research data		data classified as <b>Protected or higher</b>	
open access data, published data			data classified as <b>Public</b>

## Services Portal:

[What is the security classification of my research data?](#)

[How do I handle research data classified as Highly Protected?](#)

# eNotebook tips

# Your eNotebook

- **The obvious**
  - Rationale/background
  - Methodology
  - Results (raw and processed)
  - Discussions, notes, thoughts
- **The not-so-obvious**
  - Literature research
  - Important emails
  - Meeting notes
  - Presentations
  - Software installation files
  - Thesis drafts
  - Anything and everything!
- Check with your supervisor/group if there is a preferred structure
  - especially if you are unsure how you will work
- Its up to you!
- What are your work-flows?
  - By aim?
  - By topic?
  - By experiment start-to-finish?
  - By technique?
- As long as it is clear to follow for another person
- Move away from structuring primarily by chronology

# eNotebook: Tips for using it and for managing research data in general

- Context is key
  - Build your **narrative** (how, why, what, etc). It is invaluable for one to understand your research *process*.
  - Include everything someone would need to **replicate** your research.
  - Your main collaborator is **you** six months from now. Your past-self does **not** answer emails.
  - Use **informative** titles and headings.
  - Use the Comments feature to keep discussion about the data with the data instead of your inbox
  - Make good use of **hyperlinks**.
  - Ideally, have a ‘home page’ giving an **overview** of the entire project.
  - Do not make someone **hunt** for things in attachments.
- Keep **raw data** raw
  - Place primary data on the eNotebook immediately and leave it **unmodified**.
  - Process your data in **separate entries/files** and make it very clear how this was done.
  - This ensures the **provenance** of the data.
- Use **paper** if you want – just take a photo/scan of it and pop it in
- Got data elsewhere (RDS, GitHub, etc)? Just provide the path e.g.  
<\\shared.sydney.edu.au\\research-data\\PRJ-projectname\\file.abc>
- Don’t use with identifying data

# eNotebook Further Resources

- Article: [What is the eNotebook?](#)
- Article: [What are some tips for using the eNotebook?](#)
- Video: **Convenient Collaboration** - The eNotebook has a range of features which can enhance collaborations both inside and outside of your own research group.
  - Click [here](#) for firstname.lastname@sydney.edu.au account holders
  - or [here](#) for unikey@uni.sydney.edu.au account holders.
- Video: **Superb Supervision** - The eNotebook can improve supervisor oversight of research students and help satisfy retention period requirements after a student has completed.
  - Click [here](#) for firstname.lastname@sydney.edu.au account holders
  - or [here](#) for unikey@uni.sydney.edu.au account holders.

# **Research Data Management Plans**

# Research Data Management Plans (RDMPs)

A Research Data Management Plan (RDMP) is a document in which a researcher outlines how their data is to be managed both during a project and after it is completed.

## Why?

- To assist researchers **engage with the RDM process**
- To act as a record of data management and as a guide for members of a research group
- Required by University policy and as a part of your HDR Milestones

## How?

- Through the Researcher Dashboard (DashR)
- To be updated as a project progresses

*“Research data will be stored on the University-licensed eNotebook (LabArchives) platform during the project. The eNotebook is an electronic notebook that stores data on secure servers within NSW. The eNotebook will be named ‘Sample Study’ will be owned by the Lead Investigator, Jane Smith (UniKey: jsmi1234)”*

## Further information

[What is a Research Data Management Plan \(RDMP\)?](#)

[What is the Researcher Dashboard \(DashR\)?](#)

## Video: “Pondering Research Data Management Plans”

Click [here](#) for `firstname.lastname@sydney.edu.au` account holders or [here](#) for `unikey@uni.sydney.edu.au` account holders.

# Further RDM Resources

## Services Portal: How do I manage my research data?

Our main help page with links out to other key resources



## Ask a question or book a consult

Email us directly

[digital.research@sydney.edu.au](mailto:digital.research@sydney.edu.au)



## Services Portal: How can I learn more about research data management?

### Digital Research Cafe

Short (5 minute) videos

Using University platforms for collaboration, supervision, managing research data, and so on



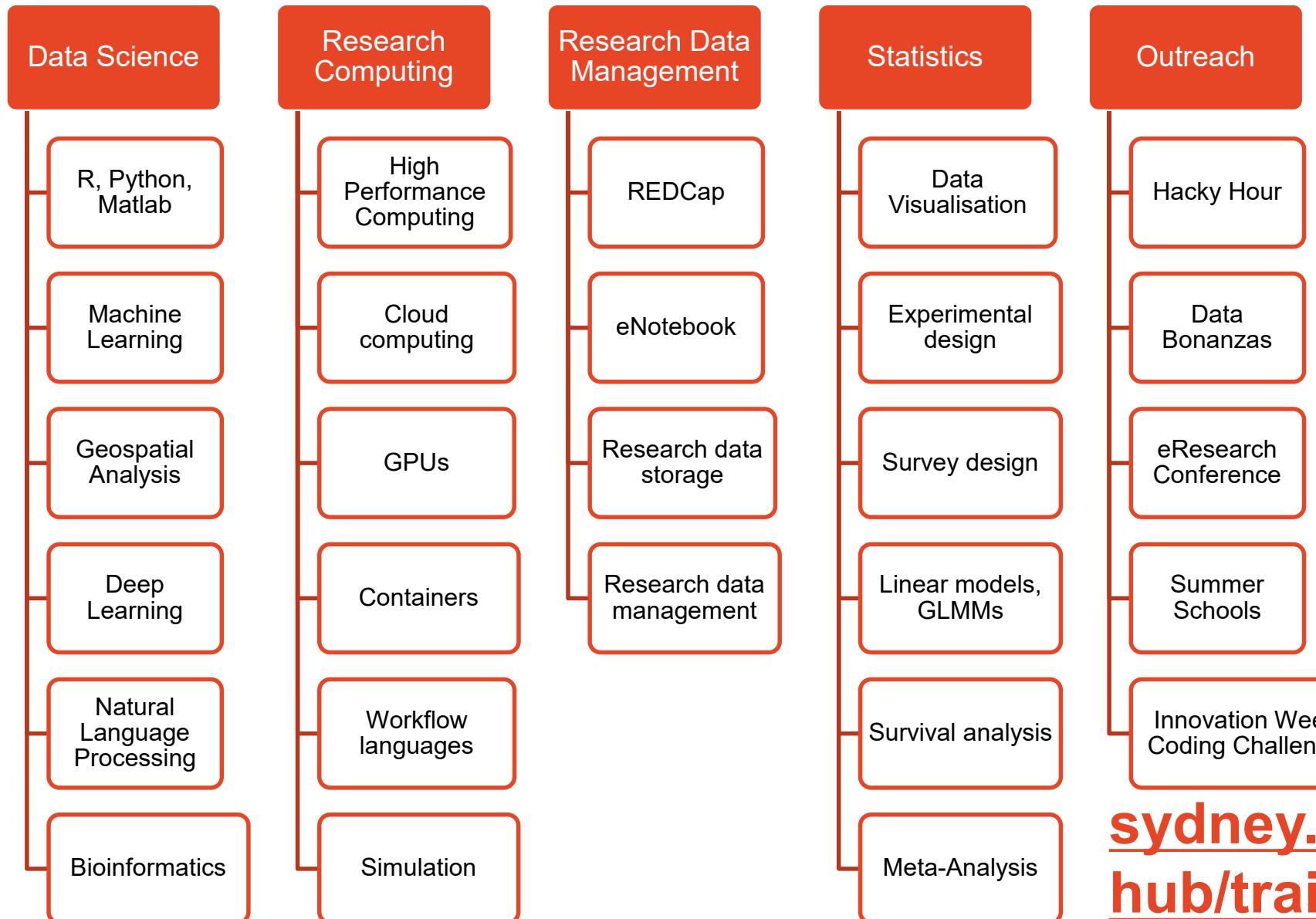
# Sydney Informatics Hub

“ SIH is a Core Research Facility enabling excellence in computational and data-driven research through advanced digital infrastructure, expert data consultancy and analytics training ”

## Activities



# SIH Training and Outreach



USyd Core Facilities @Sydney\_CRF · Jul 30

A rare **#HackyHour** at **#SydneyInformaticsHub**'s home offices. Thanks to Louis Mercorelli (@LouisMercorelli), Nathaniel Butterworth, Kristian Maras, Olya Ryjenko and Tracy Chew.



USyd Core Facilities @Sydney\_CRF · Mar 17

#SydInformatics' Dr. Darya Vanichkina running a **#MachineLearning** in **#R** workshop this rainy Monday morning ☔️ Sky water ain't stoppin no coding! 🌧️💻



[sydney.edu.au/informatics-  
hub/training](http://sydney.edu.au/informatics-hub/training)

**Take home:  
using good RDM practices and tools  
will enhance your research experience**

**Good luck with your research!**

**Sydney Informatics Hub website**  
[sydney.edu.au/sydney-informatics-hub](http://sydney.edu.au/sydney-informatics-hub)