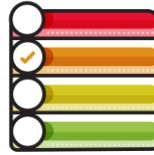


Reproducible research using electronic notebooks



RDM@
UNSW



Classify all Research Data
using **UNSW Data Classification**



Use **UNSW Supported Data Platforms** for Research Data



A '**Living**' **RDM Plan** for each
UNSW Research Project



Complete Core **RDM online Training** Modules



Outline Title

- Topic 1
 - Topic 1A
 - Topic 1A.i
 - Topic 1B
- Topic 2
 - Topic 2A
 - Topic 2B
- Topic 3
- Topic 4
- Topic 5
- Topic 6

General reproducibility

A python script which was used by 150+ papers was found to produce different results depending on operating system. The code was fixed by other researchers who had tried to reproduce the paper

Great that the other researchers could attempt to reproduce the data

But what about the other papers? What now?



<https://arstechnica.com/information-technology/2019/10/chemists-discover-cross-platform-python-scripts-not-so-cross-platform/>

Elements of reproducibility

- Method
- Data
- Process
- Metadata (eg. Definitions, standards)
- Physical things? (we don't care about those at E-Research! :)

Transparency vs Replication



Reproducing Methods vs Results vs Inferences

U.S. National Science Foundation (NSF) subcommittee on replicability in science:

“reproducibility refers to the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results.... Reproducibility is a minimum necessary condition for a finding to be believable and informative.”

(<https://stm.sciencemag.org/content/8/341/341ps12.full>)

E-Notebooks



Lots out there, we use LabArchives but features are similar.

Advantages to E-Notebooks over paper:

- Accessibility: Needs internet, but that's all
- Search and organisation
- Collaboration and sharing
- Auditing
- Multimedia (I feel old using this word...)
- APIs - importing automatically
- Flameproof
- Save the trees!

Accessibility



- Easier to keep records
- easier to be in one place
- Can keep documentation, procedures, even data in the same place
- Generic: usable by any discipline

Search and organisation



- easy to look for existing records
- maintain consistency between experiments and even studies
 - cut and paste
 - referencing
 - copying folder structures
 - Templates

Collaboration and sharing



- keeping data in the same place
- Supervisors can check in
- External people even!
- Can potentially reference this data in publications (DOIs, etc)
- Can be used to facilitate replication

Auditing



The meaner side.

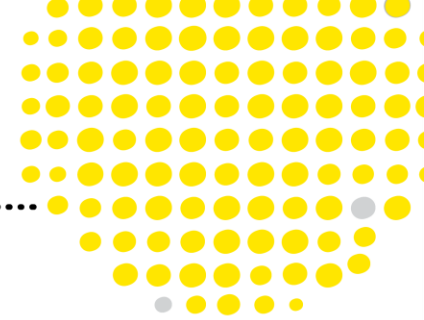
Labarchives doesn't delete data, and doesn't allow editing of logs.

Multimedia

- Can keep documentation, graphs, images, links to other parts of the notebook or external resources.
- Small data can be kept, mostly text, but external data can be linked in.

Flameproof

Not to be underestimated!



Automation - Widgets



UNSW's Ramacciotti centre is using to find what settings and materials to use even before the experiments are started.

Widgets are part of the template for experiments

Settings are then recorded for future reference

Automation- APIs



Exploring this option with a few projects.

API can import arbitrary data into Labarchives

- Instrument settings
- Metadata
- Timestamps
- Maybe data?

HPC

- job scripts,
- input/results files or links

Not a complete solution.



A lot of data is too big for these systems – auditing external files

Still requires discipline from the researchers

Standards needed for metadata to facilitate re-use

Lack of standard templates, different for each discipline/group/method

Internet can be hard to get into some locations eg; labs, developing nations.

Thanks

David Jung



Ramaciotti Centre
for Genomics

