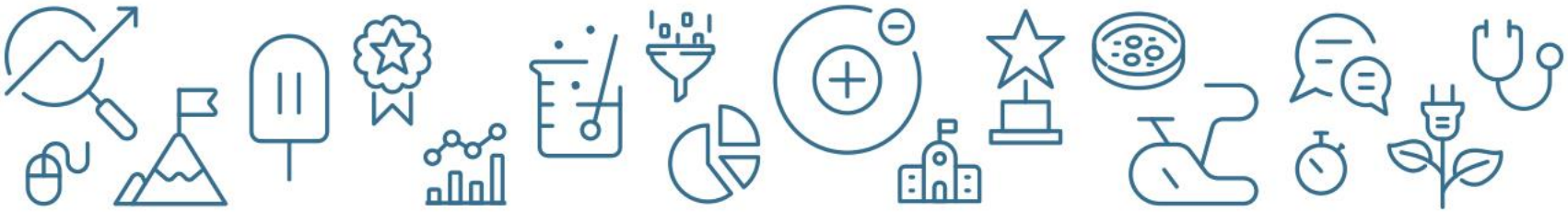
Automating the provisioning and user permissions of institutional research data allocations through the research data management platform ReDBox

Presented by

Jason Bell
j.bell@cqu.edu.au

Andrew Brazzatti
andrew.brazzatti@qcif.edu.au

QCIF



CQUniversity AUSTRALIA

RESEARCH WITH IMPACT

PERSONAL OVERVIEW

Jason Bell

- Senior Research Technologies Officer
  Digital Services Directorate
  CQ University (Australia)

- eResearch Analyst
  Queensland Cyber Infrastructure Foundation (QCIF) http://www.qcif.edu.au/

Helps researchers achieve better research outcomes through the use of technology.

Andrew Brazzatti

- Software Solutions Technical Lead
  Queensland Cyber Infrastructure Foundation (QCIF) http://www.qcif.edu.au/

One of the main developers of the ReDBox platform since 2012

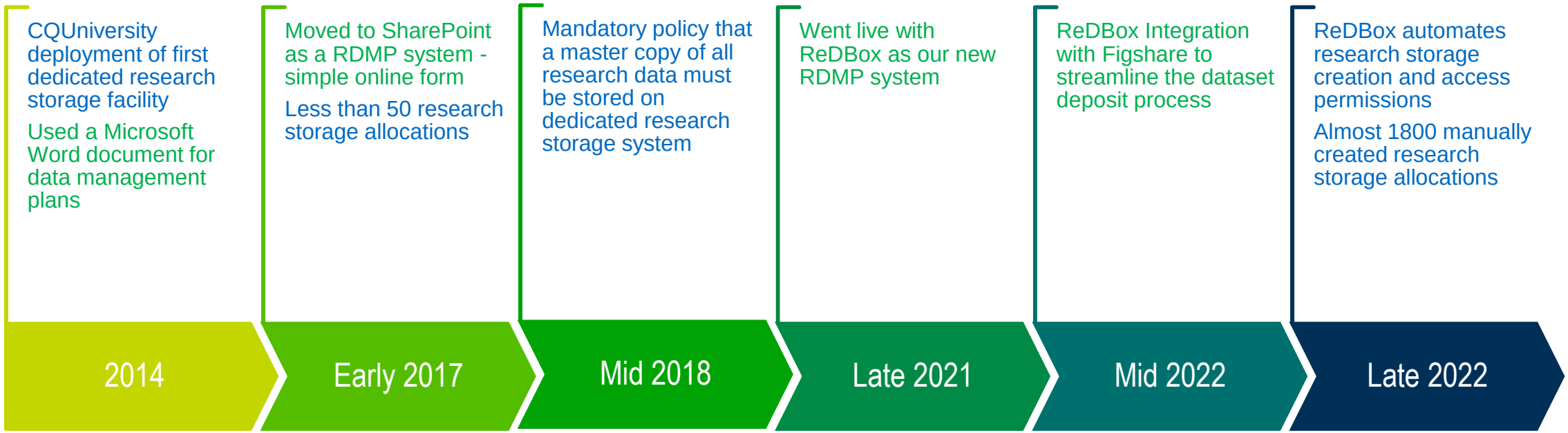
ACKNOWLEDGEMENT OF COUNTRY

I respectfully acknowledge the Traditional Custodians of the land on which we live, work and learn. I pay my respects to the First Nations people and their Elders, past, present and future.

THE MOST DANGEROUS PHRASE IN THE HUMAN LANGUAGE IS… "WE'VE ALWAYS DONE IT THIS WAY"

GRACE HOPPER, AMERICAN COMPUTER SCIENTIST AND US NAVY REAR ADMIRAL WHO IS CREDITED WITH DEVELOPING THE FIRST COMPILER

CQUNIVERSITY RESEARCH DATA STORAGE HISTORY



- Blue text indicates research storage system items
- Green text indicates RDMP items

RESEARCH DATA STORAGE ACCESS METHOD

Researchers access their research data by using the Samba/SMB/CIFS network sharing protocol.

Users mapped each project share as an individual network drive, but some users were running out of drive letters, so we created a generic/umbrella network share to simplify things.

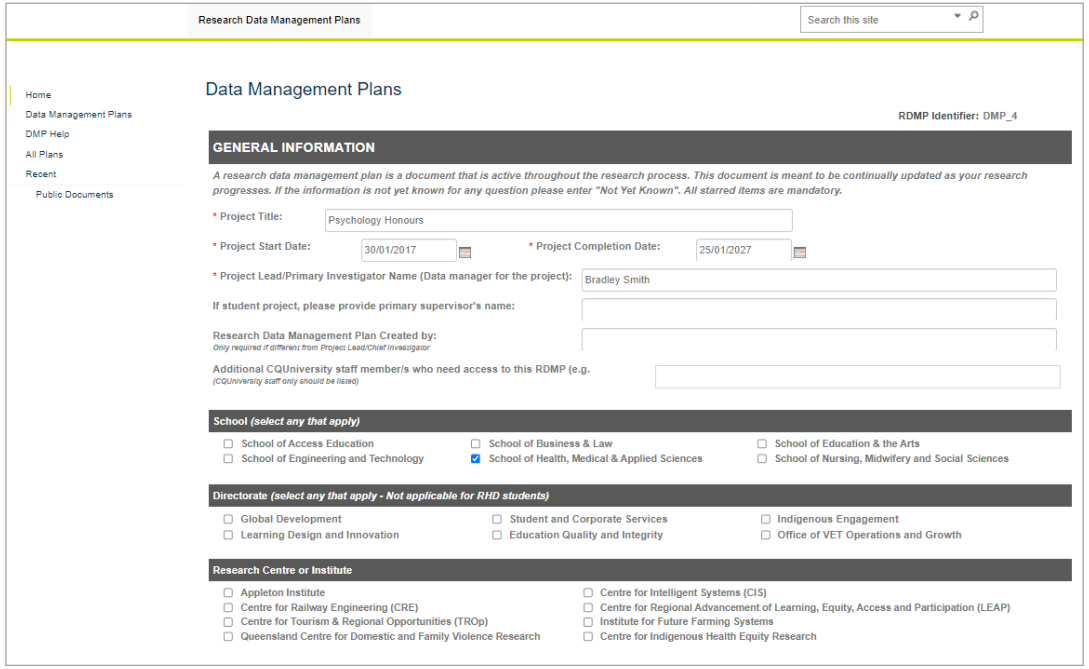
Users maps a generic network share – where each project a user has access to can be seen as separate folder/directory.

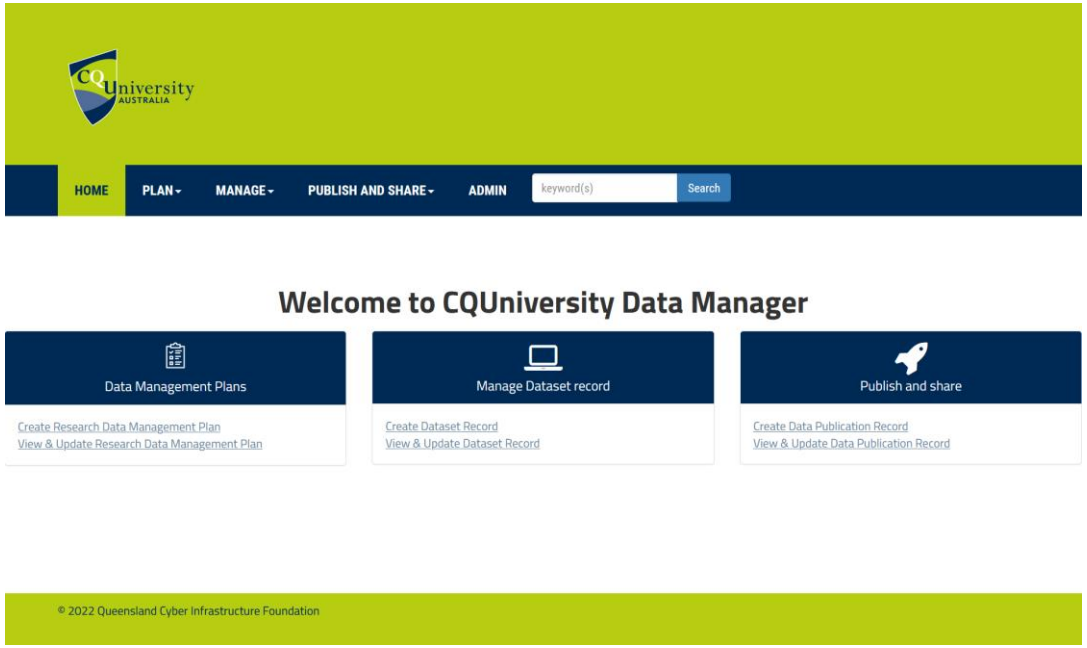
We may consider other access methods in the future.

```
[myshares]
path = <location of research data>
comment = Allows users to see each research
          share as separate folders
valid users = <General Group>
read only = false
inherit acls = true
browseable = yes
hide unreadable = yes
dmapi support = true
hide unreadable = yes
create mask = 0660
create mode = 0660
directory mask = 2770
directory mode = 2770
```

WHY DID WE CHOOSE REDBOX?

- The previous RDPM system (SharePoint) was a “temporary solution / band-aid” that provided an online form, however it was hard to update and users often found it confusing
- ReDBox offered a much better online form, with significantly improved layout, integration capabilities and customised workflows
- This enabled us to:
  - Integrate with other institutional systems, Research Master, HR systems, Figshare etc
  - Autofill's some of the fields, using data obtained from integrated systems (ie, names would autofill email address, etc)
  - Streamline research data deposits into our Figshare repository
  - Automate the provisioning and user permissions of institutional research data allocations
- We are currently exploring more ReDBox development work for future integrations/projects

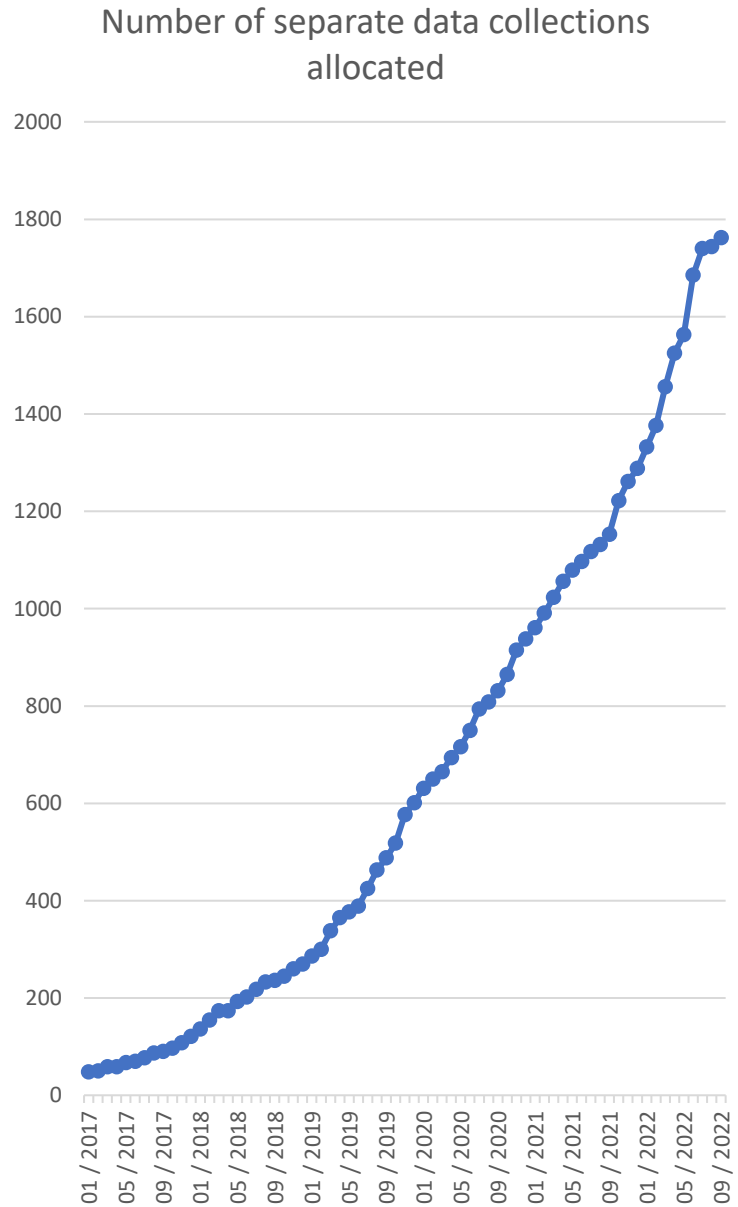




Why Automate and empower User Access permission control?

- **Share creation**
  - Remove manual time consuming work
  - Reduce human error
  - Speed up time for data share creation

- **Permissions**
  - Removes need for IT support to manually give and remove access permissions
  - Users no longer have to wait for access to be granted
  - Users can see who has access at any time
  - Enables users with permission, to add or remove users access to the research share assigned to the project



HOW DID IT USE TO WORK?

- Research users log on to our research data management system (ReDBox – which is MFA'ed), create and save an initial Research Data Management Plan
- Share creation
  - Research IT support would have to monitor newly created Research Data Management Plans and do them in a bunch, so users may have had to wait up to a week before the share was created
  - A system admin would then have to:
    - Request our digital infrastructure team to create a LDAP group based on information provided in the RDMP
    - Create directory on research storage system
    - Create and smb.conf configuration
    - Populate permissions group based on users who have been assigned access on the newly created RDMP
    - Emails instructions to people who have been given access
- Permissions
  - CQUni has a system where if an account expires, permissions would be removed
    - This means if the account is activated again (ie, contract is renewed), the user would need to be given access again.
  - Users would have to contact support to get a user added or removed
  - If the PI wasn't the one making the request, support would contact PI to ensure access is okay to be given

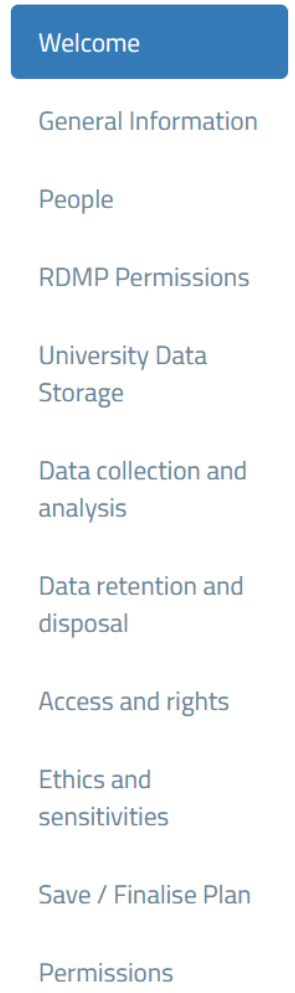
HOW DOES IT WORK NOW? (OR SHORTLY!?)

- Research users log on to our research data management system (ReDBox – which is MFA'ed), create and save an initial Research Data Management Plan
- Share creation
  - As soon as a new research data management plan is created and saved, a research share is automatically created
  - The process then:
    - Creates a write and a read-only LDAP group
    - Creates directory on research storage system
    - Populates permissions based on users who have been assigned access on the newly created RDMP
    - Emails instructions
- Permissions
  - CQUni has a system where if an account expires, permissions would be removed
    - This means if the account is activated again (ie, contract is renewed), the user would need to be given access again
  - If a user had "write" permissions to RDMP, the user could simply open the associated RDMP, go to the "University Data Storage" section and add or remove users as required!

HOW DOES IT WORK – CREATION OF NETWORK SHARE

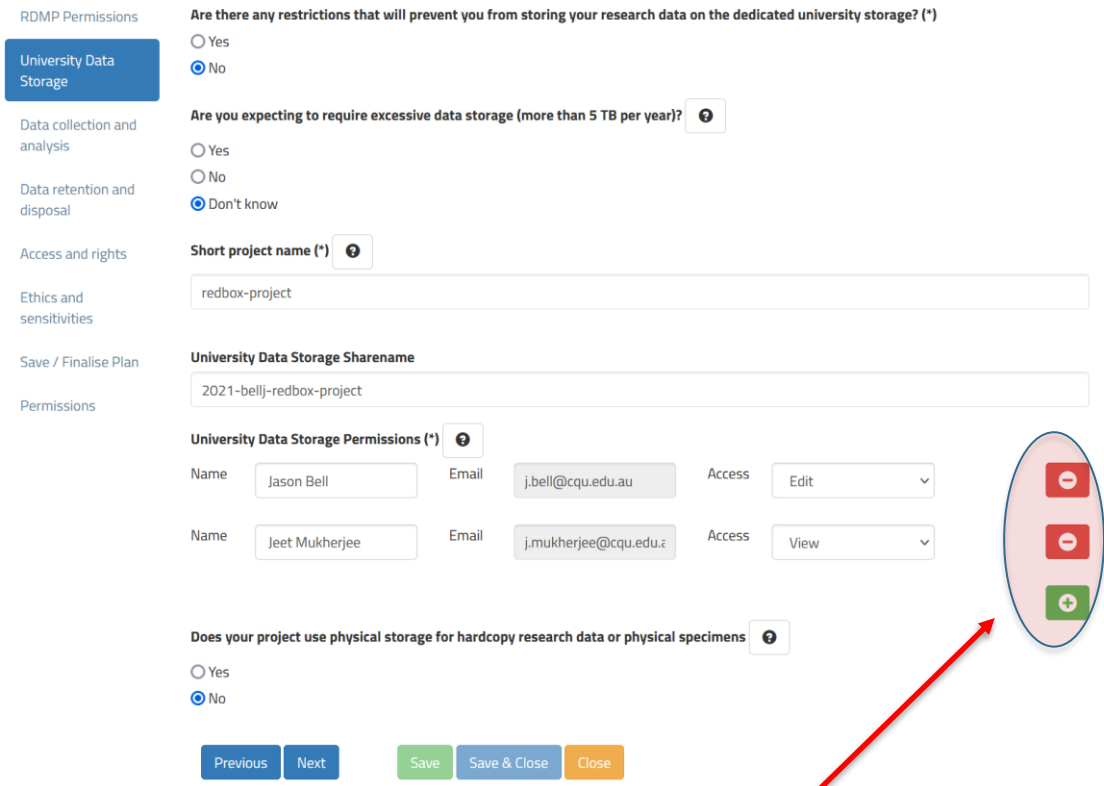
- A new RDMP is created – initially only ten fields needs to be filled out to be able to save the plan, in which a research share is then created.

- A “ReDBox Sync scripts (python)” process detects via the API that a new plan has been created
  - Creates two new LDAP Groups, a read-write and a read-only group
  - Populates the groups with users listed in “University Storage Access” section within ReDBox
  - Creates a new directory that will act as the location for data storage allocation
  - Creates a smb.conf configuration item
  - Sends a notification email with instructions

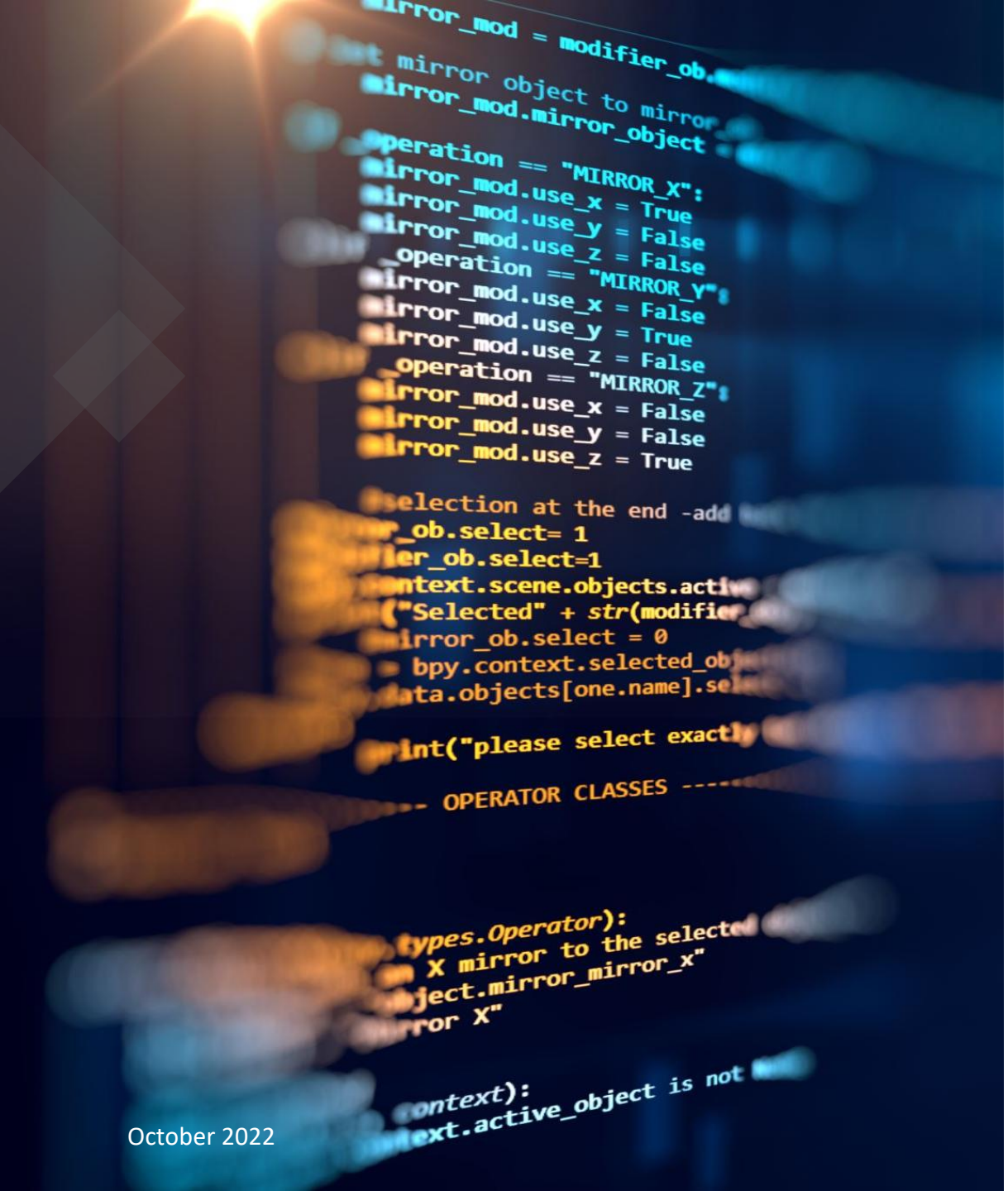


HOW DOES IT WORK – CHANGE IN RESEARCH STORAGE ACCESS PERMISSIONS

- Users with RDMP edit permissions:
  - Edits associated research data management plan for the assigned research project
  - Within the “University Storage Section” add or remove users access permissions
- “ReDBox Sync scripts (python)” process detects via the API that user permissions have changed, then replicates changes to LDAP groups permissions
- Inactive accounts on the CQU side are removed from ReDBox RDMP and Data Share permissions to stop accidental data access if the person returns in a different role
- ReDBox is considered the source of truth for data storage access permissions



Click on these buttons to add and remove user's access



WHAT WE HAVE LEARNT

- Moving from unstructured data to structured data is *interesting*!
- Write code that checks what you want to do, before you do it.
  - This uncovered a lot of things we didn't realise

THE PROJECT TEAM

Automation Project Team

- Jason Bell (CQUniversity)
- Simon Coggins (CQUniversity)
- John Conelius (CQUniversity)
- Andrew Brazzatti (QCIF)
- Alejandro Bulgaris (QCIF)

General CQUniversity ReDBox Team

- Nicole Healy (CQUniversity)
- Jeet Mukherjee (CQUniversity)
- Amanda Flanders (CQUniversity)
- Lachlan Hoole (QCIF)

THANKYOU - QUESTIONS?