5 Safes in Practice

2 November 2022

#QCIF

@qcifltd

@qcif

Safe People

Safe People

"Safe people reviews the **knowledge, skills and incentives** of the users to store and use the data appropriately. It considers the confidence of the data owner that those who will access the data can be trusted to use it appropriately."

Desai, T., Ritchie, F. and Welpton, R. (2016) *Five Safes: designing data access for research*, Economics Working Paper Series no. 1601, University of the West of England, Bristol

Knowledge, skills and incentives

Training

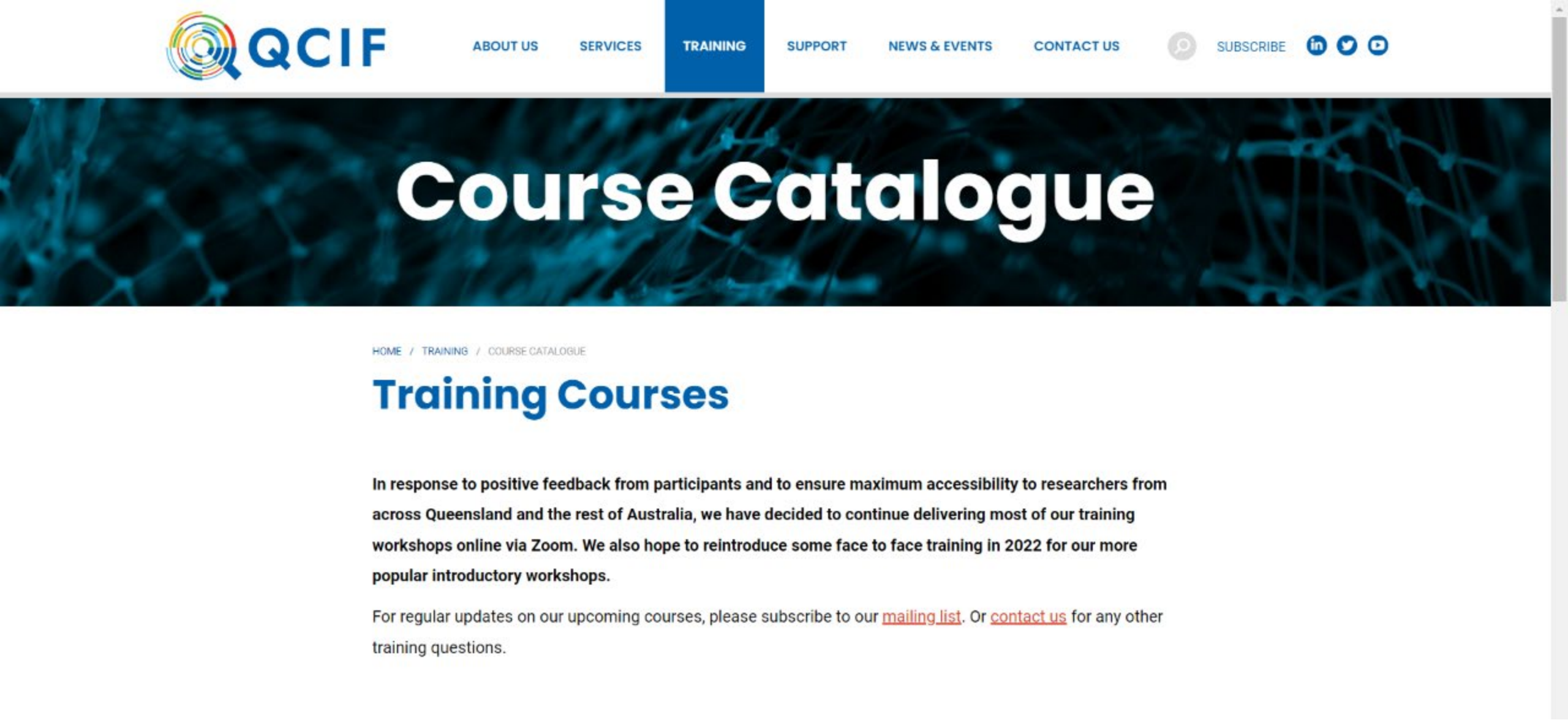
- QCIF runs regular training sessions on Working with Sensitive Data, as well as statistics, statistical tools (r, python) and REDCap
- https://www.qcif.edu.au/training/course-catalogue/

Experience

- Training alone is insufficient to demonstrate a good track record
- Record all **projects** worked on, both successful and unsuccessful (and learnings from these)

Conflicts of Interest

- Always declare!



https://www.qcif.edu.au/training/course-catalogue/

CATEGORIES

- [ ] Bioinformatics Analysis
- [ ] Data Processing and Statistics
- [ ] Data Sharing and Reproducibility
- [ ] Machine Learning
- [ ] Programming / Software Carpentry
- [ ] Surveys and Data Management
- [ ] Visualisation and Research Communication

PROGRAMMING / SOFTWARE CARPENTRY

INTRODUCTION TO UNIX

This workshop will help you get started with using Unix and the command line interface.

Learn More

DATA SHARING AND REPRODUCIBILITY

A GUIDE TO REPRODUCIBLE RESEARCH

A webinar exploring why and how you should make your research findings reproducible

Learn More

DATA SHARING AND REPRODUCIBILITY

MAKING YOUR RESEARCH DATA FAIR

Making your data FAIR helps improve the visibility of your research and encourages collaborations and citation.

Learn More

DATA SHARING AND REPRODUCIBILITY

WORKING WITH SENSITIVE DATA

Sensitive data requires special approaches and management, covered in this practical and discussion-based workshop

PROGRAMMING / SOFTWARE CARPENTRY

INTRODUCTION TO PROGRAMMING: R FOR REPRODUCIBLE SCIENTIFIC ANALYSIS

A novice programmers guide to the R software environment, a

DATA SHARING AND REPRODUCIBILITY

DATA CAPTURE AND MANAGEMENT WITH REDCAP

Our series of modular REDCap training workshops.

Safe Settings

Safe Settings

"Safe settings relates to the **practical controls** on the way the data is accessed."

Desai, T., Ritchie, F. and Welpton, R. (2016) *Five Safes: designing data access for research*, Economics Working Paper Series no. 1601, University of the West of England, Bristol

Practical controls

Technical

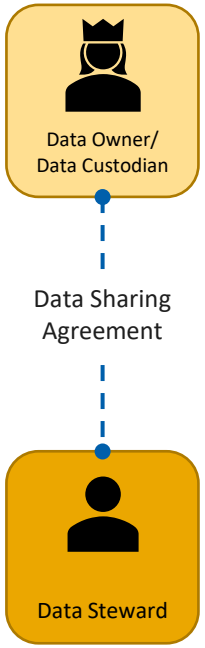
- Data Safe Havens: SeRP, SURE, ERICA, KeyPoint
- QCIF is preparing the launch of KeyPoint; built in-house from the ground up with the specific purpose of delivering a 5-Safes approach to data sharing and analysis
- Secure storage, secure log in, managed data access, audit trail
- Physical security

Administrative

- Governance and procedures
- What happens if/when someone does the “wrong thing”
- How to avoid people doing the wrong thing in the first place?

5 Safes and KeyPoint

Establish the data sharing and use model



Data Sharing Agreement

- It starts with an Agreement to share data (This is external to KeyPoint)
- Established between one or more Data Owners / Data Custodians and a Data Steward
- Prescribes how data can be used and for what purposes:

Safe Projects →
- What types of Research Activities are permitted

Safe People →
- Who is allowed to access the data

Safe Settings →
- Where the data can be stored and accessed

Safe Data →
- What data are allowed to be shared and what data treatments are required (e.g. anonymisation)

Safe Outputs →
- What outputs are allowed and what is the process for approvals

Bringing together the 5 Safes: KeyPoint

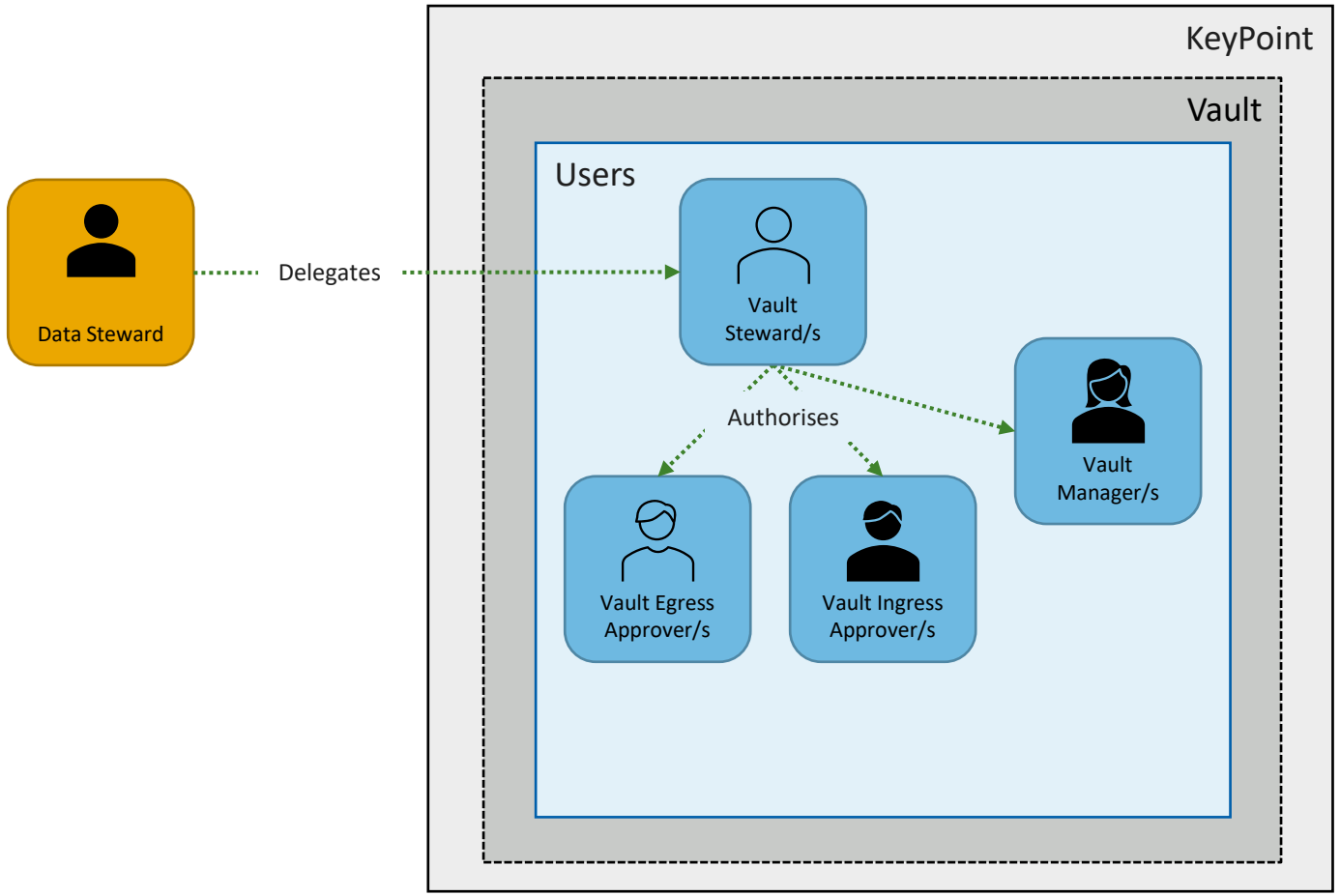
Safe Setting

- From what physical location(s) will the data be accessed?
- Does there need to be any auditing of these locations?
- What IT security needs to be in place?
- Do controls exist to limit misuse, interference, unauthorised access, modification, loss or disclosure?
- Do users understand how to access the data safely in the IT and/or physical environment?
- How will data transfer into and out of the secure environment?

KeyPoint

- Each Data Hosting agreement between the Data Steward and QCIF will result in a ring-fenced, isolated environment that encompasses the data and resources (a Vault)
- Data is encrypted at rest and in transit
- Access is monitored and audited
- Multi factor authentication

Establish data governance in KeyPoint



- The Data Steward can delegate the day-to-day operations to a team, or take on the roles themselves
- Each role has a strictly defined scope of concern

Bringing together the 5 Safes: KeyPoint

Safe Projects

- Is the proposed study in the public interest?
- Is the use of the data approved for the project?
- Is the project based on sound scientific fundamentals?
- Is it ethical?

KeyPoint

- Each project proposal is evaluated on its merits by the Data Steward
- No access to the system or data without approval

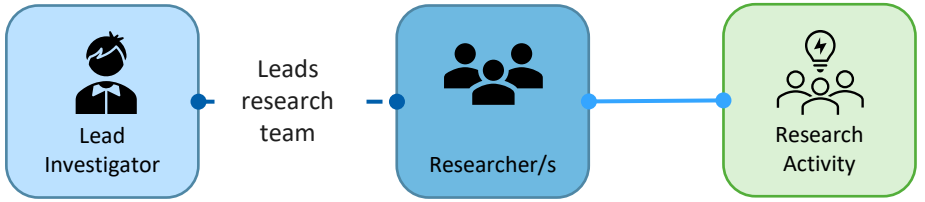
Bringing together the 5 Safes: KeyPoint

Safe People

- Are there any restrictions on who may apply for data access?
- What is the process for people or organisations to apply for data access?
- Does the data user need to meet any specific criteria to access data?
- Does the data user have a history of good data handling practices?
- Does the data user need to be trained in safe use, data storage and technical skills?
- Will a legally binding agreement govern the access and use of the data be required?
- What sanctions (legal and non-legal) need to be available for misuse of data?

KeyPoint

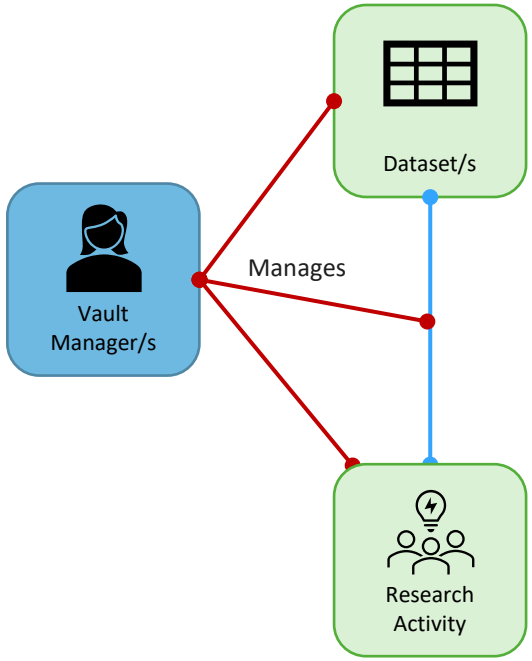
- Access to use the system is **only** by invitation from the Vault Manager (Data Steward's delegate)
- Prospective users are required to demonstrate understanding of 5 Safes and safe data handling
- Prospective users must accept KeyPoint terms of use
- Account and access can be suspended by Vault Manager and/or QCIF



Bringing together the 5 Safes: KeyPoint

Safe Data

- Has the data been treated to appropriately match the disclosure risk and the project, in order to provide the maximum utility at the acceptable risk?
- The degree of data treatment required will become evident once it is clear who will be able to access the data, under what conditions, in what circumstances and how the resulting data will be protected in order to be made public. (ABS)



KeyPoint

- Datasets and access to these by Researchers are managed by the Vault Manager
- Datasets can be treated (subsetted) inside KeyPoint to create new datasets
- Each dataset can be linked / unlinked to multiple research activities (groups of researchers)
- Complete separation of visibility of files between research activities
  - i.e. a researcher working on two research activities cannot access both datasets in the same session

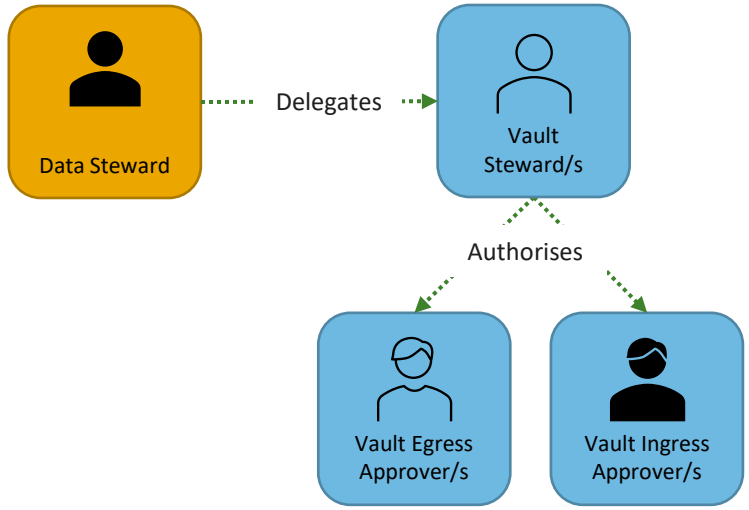
Bringing together the 5 Safes: KeyPoint

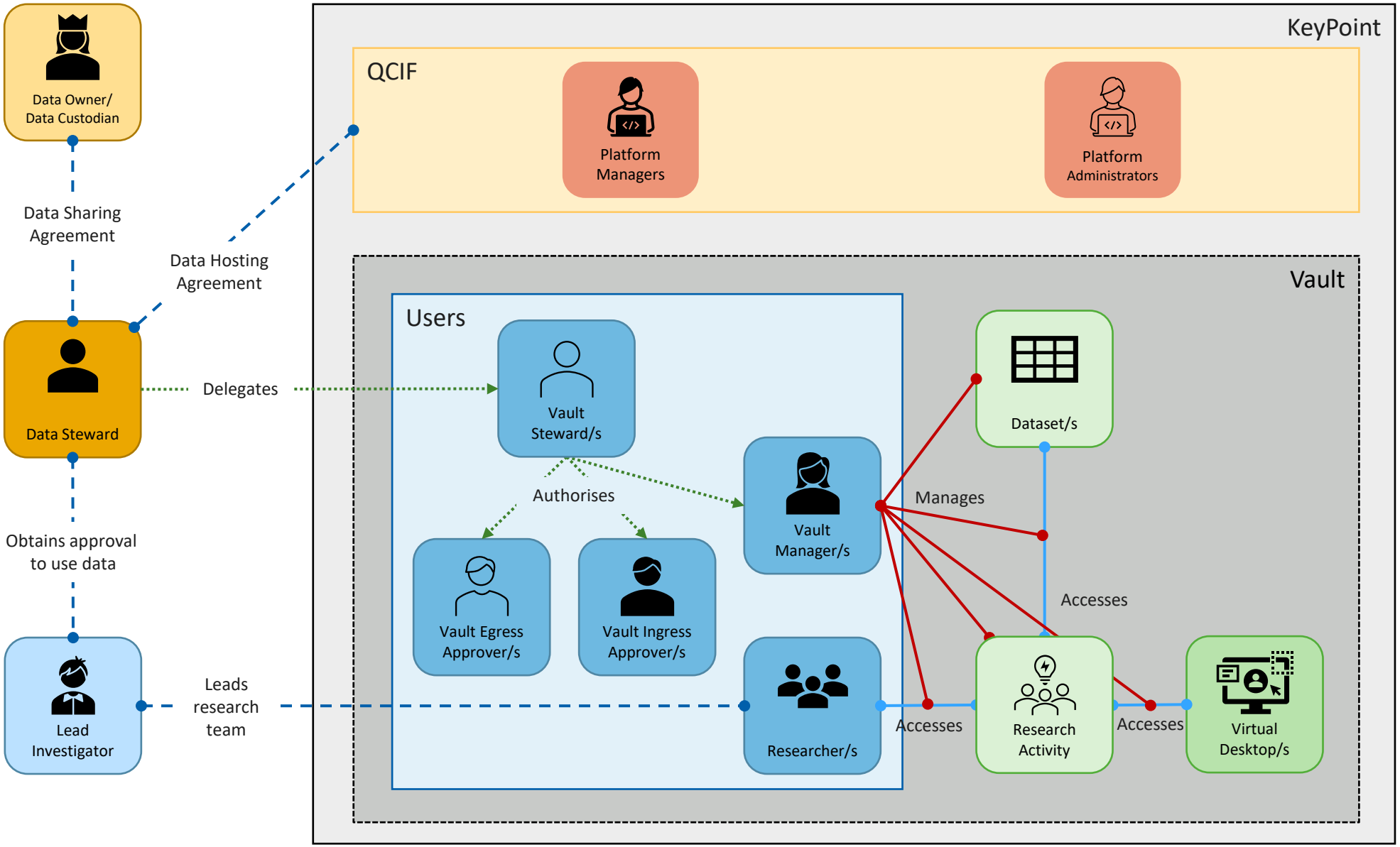
Safe Outputs

- Will detailed statistics be made available, and will these reveal sensitive information in themselves or in previously released information?
- Will qualitative data be screened to ensure that no sensitive information remains in the outputs?

KeyPoint

- All file ingress and file egress requests are evaluated by Ingress and Egress Approvers
- These are nominated by the Data Steward, ensuring adherence to the data governance model set out in the **Data Sharing Agreement**




Data Owner/
Data Custodian
Data Sharing
Agreement
Data Hosting
Agreement
Data Steward
Delegates
Obtains approval
to use data
Lead
Investigator
Leads
research
team
KeyPoint
QCIF
Platform
Managers
Platform
Administrators
Vault
Users
Vault
Steward/s
Authorises
Vault Egress
Approver/s
Vault Ingress
Approver/s
Vault
Manager/s
Researcher/s
Dataset/s
Manages
Accesses
Accesses
Accesses
Research
Activity
Virtual
Desktop/s

Thank you

Mark Hoffmann

m.hoffmann@qcif.edu.au