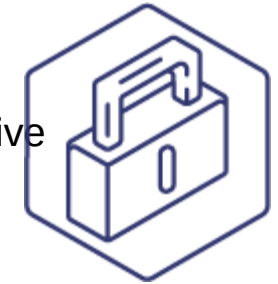
CADRE

Sensitive data management: authorisation decision-support

A/Prof Steven McEachern
Director of Australian Data Archive
Project Lead, CADRE
eResearch 2022, Brisbane

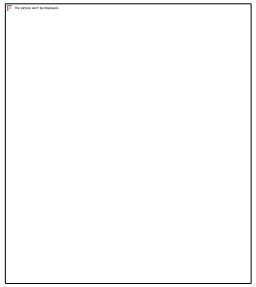


Partners

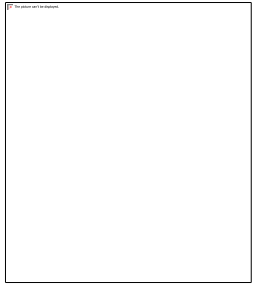
Acknowledgement of Country

I acknowledge the Traditional Owners of the land on which this talk is taking place, the land of the Turrbal and Jagera, and pay my respect to their Elders past, present and emerging. I also acknowledge any Aboriginal people joining us today in person and online.

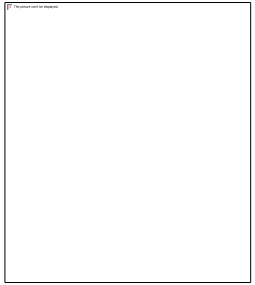
The Five Safes



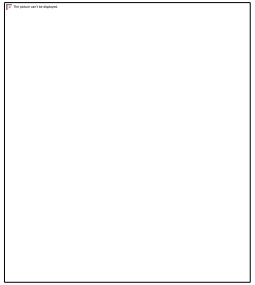
Safe Project



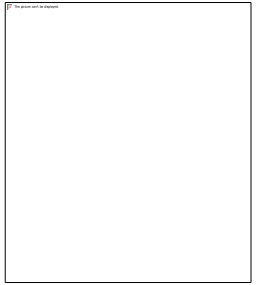
Safe Person



Safe Data



Safe Setting



Safe Output

Two Additional Safes

Organisations

Groups

Joint and severable

- Dimensions are designed so that each can be evaluated independently of the others, as far possible.
- All five dimensions need to be considered jointly to evaluate whether a data access system can provide an ‘acceptable’ solution.

CADRE

A SYSTEM TO	BY MEANS OF	IN ORDER TO
Increase the speed at which social sciences and related disciplines get access to sensitive data. Decrease the risk, time and costs associated with providing access to data (for data holders) and accessing data (for researchers)	The development of a shared and distributed sensitive data management platform using the **Five Safes** framework and common accreditation and information exchange protocols.	Enable data owners and users to address the core concerns around governance, creation, management and sharing of sensitive data for research. Share and move sensitive data safely between higher education, national research and government facilities and services.

The Five Safes PLUS Two PLUS PIDs

RAID

Safe Project

ORCID

Safe Person

DOI

Safe Data

??? WATCH THIS SPACE

Safe Setting

DOI?? HANDLE??

Safe Output

ROR

Organisations

??? (CILogon)

Groups

Operationalisation

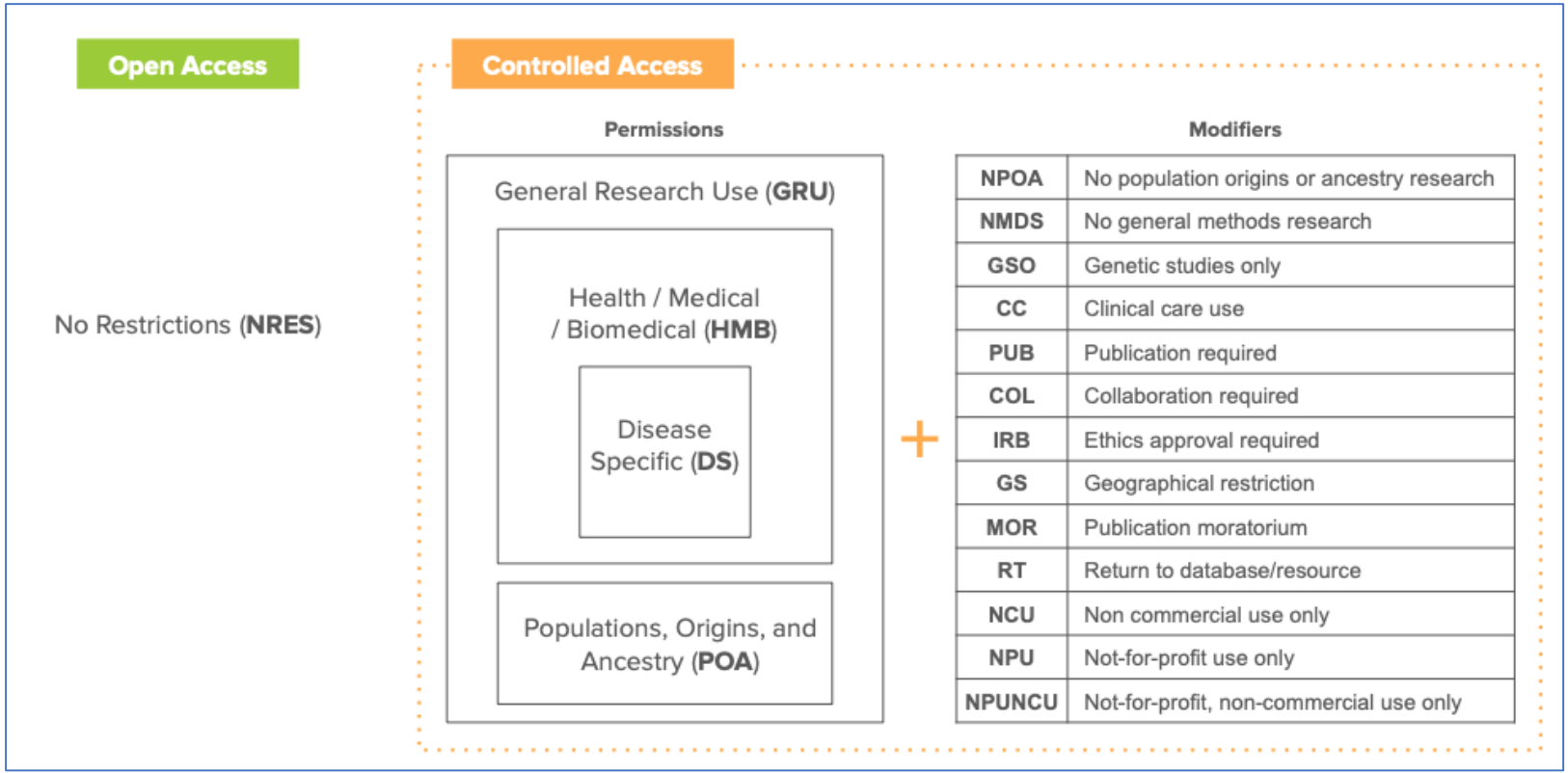
- Solutions architecture

- Data Use Ontology
- Data Tags Suite

- AAF attributes
- CILogon

- Evaluation and POCs -> Beta

Data Use Ontology



Australian Survey of Social Attitudes (2020)

DOI:

http://dx.doi.org/10.26193/C86EZG

Data Use Limitations

- GRU - General Research Use - DUO_0000042

Modifiers

- PS - Project specific restriction – DUO_0000027

Custom metadata block?

Longitudinal Survey of Australian Youth

DOI:

http://dx.doi.org/10.4225/87/PJO7GB

Data Use Limitations

- GRU - General Research Use - DUO_0000042

Modifiers

- PS - Project specific restriction – DUO_0000027
- GS - Geographic restriction – DUO_0000022

Five Safes dimension	DUO permissions	DUO modifiers
(Non-specific)	NRES - No Restrictions	
People		GS – Geographical restriction COL – Collaboration required US – User specific restriction
(Organisations)		Institution specific restriction NPUNCU – Not-for-profit, non-commercial use only*
Projects	GRU – General Research Use Health/Medical/Biomedical (HMB) Disease specific (DS) Populations, Origins, and Ancestry (POA)	NPOA – No population origins or ancestry research NMDS - No general methods research GSO – Genetic studies only CC – Clinical care use IRB – Ethics approval required NCU - Non-commercial use only NPU – Not-for-profit use only NPUNCU – Not-for-profit, non-commercial use only* PS - Project specific restriction TS - Time limit on use
Data	(None)	(None)
Settings	(None)	(None)
Outputs		PUB – Publication required MOR – Publication moratorium

Data Tags Suite (Alter et al., 2020)

- ***Authorisation***
- Authentication
- Access

- Aligns with DUO and other standards
  - ODRL, DPV, …

Data Authorisation

Authorisation type	Description
None	Not covered by a DUA
“Click through” online license	Users must agree to an online agreement without providing additional identification
Registration	Users must register before access is allowed and agree to conditions of use. Registration information may be verified
DUA signed by an individual	An agreement signed by the investigator is required. DUAs may require additional information, such as a research plan and an IRB review (see discussion of licenses below)
DUA signed by an institution	An agreement signed by the investigator’s institution is required. DUAs require additional information, such as a research plan and an IRB review (see discussion of licenses below)

Data Authentication

Authentication type	**Description**
None	No authentication required
Simple login	Single-factor login or the use of an authentication key or registered IP address is required
Multi-factor login	Multiple-factor login using a combination of IP address, password protection, authentication key, or other forms of authentication

Data Access

Access method	Description
Download	The data are available for download. A license may be required
API	Interaction with the data may be automated via defined communication protocols, i.e., APIs
Remote access	Users may access the data in a secure remote environment (“virtual data enclave”). Individual-level data may not be downloaded, only approved results
Remote service	A user may submit program code or the script for a software package to be executed in a secure data center. The remote site returns outputs. It may perform a review before releasing the results
Enclave	Access is provided to approved users within a secure facility without remote access. Results may remain at the enclave or be released after review

DUO and DATS – Ten To Men

DUO Dimensions	Ten To Men Release 3
DOI	http://dx.doi.org/10.26193/JDE1TD
Data Use Limitations	• GRU - General Research Use - DUO_0000042
Modifiers	• PS - Project specific restriction – DUO_0000027 • US – User specific restriction • (Institution-specific restriction??)

DATS Dimension	Ten To Men Release 3
Authorisation	DUA signed by an organisation (or individual?)
Authentication	Simple login
Access	Download

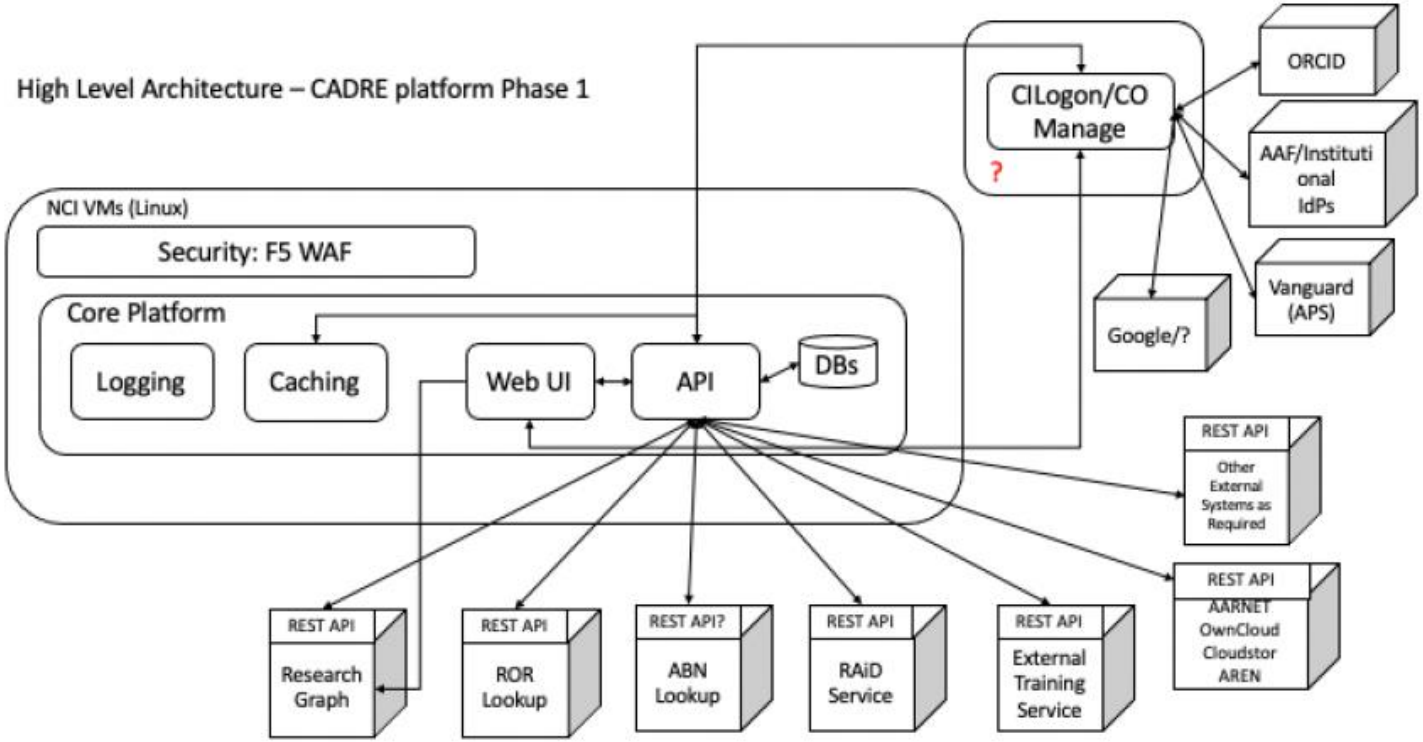
Aligning standards and the Five Safes (Alternative)

Five Safes dimension	Proposed identifier/PID	Custodian requirements specification	Information source for provision
People	ORCID??	Data Use Ontology (DUO)	AAF attributes, Scholix/ResearchGraph??
Projects	RAID	Data Use Ontology (DUO), Data Tags Suite (DATS)	(CADRE specification)
Data	DOI	???	DataCite, DCAT, Scholix/ResearchGraph??, Others??
Settings	(RAID??)	Data Tags Suite (DATS)	(CADRE specification?? Existing standard??)
Outputs	Handle, DOI	Data Use Ontology (DUO)	DataCite, DCAT, Scholix/ResearchGraph??, Others??
Organisation	ROR	Data Use Ontology (DUO)	ROR specification (who is ROR provider?)
Group	???	???	CILogon

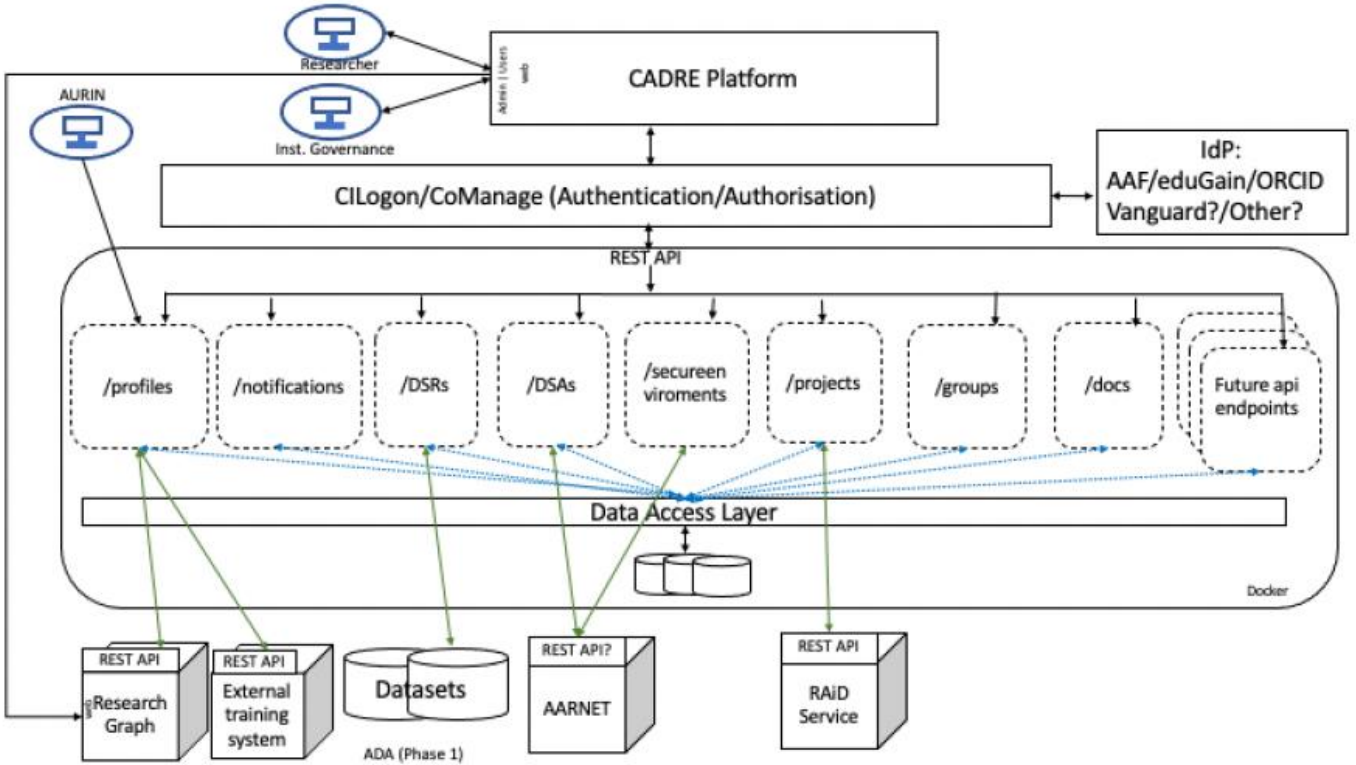
Initial work on information model
https://cloudstor.aarnet.edu.au/plus/s/u6SdXBELo5s08Go

Solutions architecture

3.2.1 High Level Platform Architecture



3.2.1.1 3-Tier web application with API backend

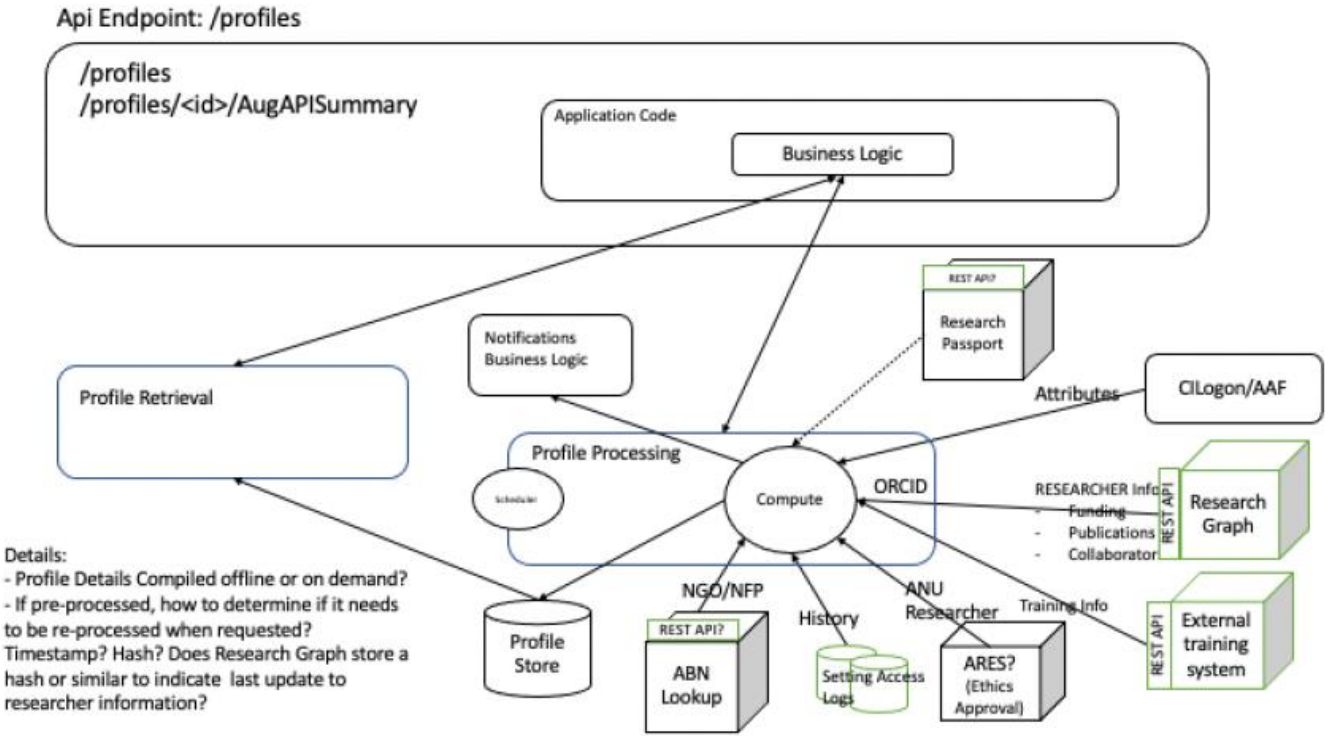


Proposed Technology stack – Phase 1

- Servers: The CADRE platform will be hosted on NCI VMs and servers. NCI can provide a level of security with f5 Web Application Firewall (WAF).
- Proposed tech stack - dependent on developer resources available and their skillset:
- Web-based front end: ReactJS or Bootstrap/Vanilla/JS JQuery
- Back-end API: NodeJS or Django web service
- middle layer: to be determined
- Database: Postgres
- Caching: MongoDB
- Container: Docker

3.2.2 CADRE API Details

3.2.2.1 Proposed Architecture for API endpoint: /profiles



Evaluation of dashboard/access management options

- REMS: CSCFI, Finland
- IMPACT: RENCI, USA
- IXUP: Commercial
- DUOS: Broad Institute, USA

High Level Evaluation – System Options for CADRE Platform – Access Request Management

Tech Solutions	ADA Reqs Fit	CADRE Reqs Fit	Community Support	ADA BAU Impact	Licensing Functions	5 Safes Aligned	DUO Integration	CILogon Integration	Dev Work Required	In/Out	Overall Rating 1-5
RAM app prototype	High	High	None	N/A	Dataverse	Yes	N/A	N/A	High (DV integration)	Out	-
REMS	High	High	Finland AU BioCommons?	Request workflow	In REMs	Implied		Yes	High (in Clojure)	In	
ImPACT	High	High	USA AU QCIF?	Request workflow	Data Policy Store – Notary	Implied		Yes	High (request management less advanced than REMS)	In	
IXUP	?	?	?	?	?	Yes	?	?	NA. Coupled with MS Azure for controlled data processing.	?	
DUOS	?	?	?	?	In DUOS	Implied	Yes	?	?	?	

Overall rating: 1 = Low and 5 = High

Confirmed technology solutions

- CILogon (Monday)
- plus
- REMS (earlier today)
- plus
- Dashboard (locally developed)
- Plus
- ResearchGraph (more tomorrow!!)

Where are we at?

- https://cadre5safes-test.ada.edu.au/
- https://dataverse-demo.ada.edu.au/
- http://learning.cadre5safes.org.au/

Questions

- cadre5safes [at] anu [dot] edu [dot] au

- ada [at] anu [dot] edu [dot] au