CADRE

CADRE and the Five Safes

Assoc. Prof. Steven McEachern,

Director, Australian Data Archive

Indigenous Data Access Workshop, July 2023

Partners

CADRE Partners

Universities

ADA AUSTRALIAN DATA ARCHIVE

UNSW SYDNEY

SWINBURNE Social Data Analytics Lab

THE UNIVERSITY OF MELBOURNE

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Government agencies

Australian Government
Australian Institute of Health and Welfare

Australian Government
Australian Institute of Family Studies

eResearch Providers

ARDC
Australian Research Data Commons

AUSTRALIAN ACCESS FEDERATION
AAF.EDU.AU

aarnet
Australia's Academic and Research Network

ResearchGraph Foundation

AURIN
aurin.org.au

Acknowledgements

Members of the CADRE Conceptual Working Group and WP1 Working Groups:

- Ingrid Mason, Dr Heather Leasor & Marina McGale, The Australian National University
- Prof Julie McLeod & Nicole Davis, University of Melbourne
- Dr Kate O'Connor, La Trobe University
- Assoc. Prof. Amir Aryani, Swinburne University
- Peter Vats, ResearchGraph
- Jatender Mohal, Murdoch Children's Research Institute (formerly Australian Institute of Family Studies)
- Kirsty Leslie, Australian Institute of Health and Welfare

The CADRE project received $1.9M investment (https://doi.org/10.47486/PL106) from the Australian Research Data Commons (ARDC), with matching co-investment funds from project partners. The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS).

Australian Research Data Commons

National Research Infrastructure for Australia

An Australian Government Initiative

Why CADRE

- Australian context for Five Safes (DAT Act 2022)
- Information requirements based approach to support operationalisation of Five Safes for quantitative and qualitative social science research data
- Multiple sources of information that enable decision-support via a dashboard

Inspiration for CADRE

- DAT Act and Office of National Data Commissioner work program as a new foundation for access to government data
- Integrated infrastructure for new framework does not currently exist (gap)
- There are substantial challenges to:
  - Scaling access procedures
  - Coordinating secure services
  - Connecting the the full stack
  - Coordinating underpinning NRI
- Value can be gained from improved access realised and trust built (fill the gap)

CADRE PROJECT

Coordinated Access for Data, Research and Environments – A Five Safes Implementation Framework for Sensitive Data. (ARDC, May 2021-June 2023)

Core aims are to operationalise the Five Safes framework and establish a shared and distributed sensitive data access management platform for the social sciences and related disciplines.

Five Safes Framework:

1. Safe People
2. Safe Data
3. Safe Settings
4. Safe Projects
5. Safe Outputs

Best Practice Guide to Applying Data Sharing Principles, 15 March 2019, Dept Prime Minister and Cabinet, Australian Government.

www.cadre5safes.org.au/about/five-safes/

CADRE is a step change

Current State	Future State
Data users (people) complete one-off, ad hoc applications using **standalone, manual** processes	Users can access, link and reuse information from their projects, data and outputs in a coordinated and integrated way
Data custodians (people) must make stand-along decisions **in isolation** from other data, projects or people	Projects, people and settings are visible and auditable by custodians in evaluation of applications
To conduct ***research projects,*** users and custodians must manage multiple applications for multiple datasets with multiple custodians	Projects can integrate people, projects and data into a coordinated application
Secure access facilities (settings) set up individual, one-off identities independent of institutional or national identity services	Identities are established and managed through single sign-on services leveraging national identity providers (AAF, Vanguard)
Outputs of sensitive data projects are disconnected from how they were approved or produced	Outputs can be tied to the projects, people and settings that generated them

CADRE FRAMEWORK: PURPOSE

The CADRE framework includes the conceptual underpinnings and the guardrails for sensitive data access management and the range of information associated with the Five Safes principles that can be operationalised in a decision-support system.

DOI 10.5281/zenodo.5748611

https://doi.org/10.5281/zenodo.5748610

What are the Five Safes?

Safe dimension	**Ritchie et al. assessment**	**ONDC Principle**
Safe projects	Is this use of the data appropriate?	Data is shared for an appropriate purpose that delivers a public benefit
Safe people	Can the researchers be trusted to use it in an appropriate manner?	The user has the appropriate authority to access the data
Safe data	Is there a disclosure risk in the data itself?	Appropriate and proportionate protections are applied to the data
Safe settings	Does the access facility limit unauthorised us?	The environment in which the data is shared minimises the risk of unauthorised use or disclosure
Safe outputs	Are the statistical results non-disclosive?	The output from the data sharing arrangement is appropriately safeguarded before any further sharing or release

The Safes are “Joint and severable”

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

CADRE Five Safes framework: Key areas for review

1. Overall Approach (Framework Structure)
2. Five Safes Application to Qualitative Data
3. Extensions to the Five Safes
4. Joint and Several Application of the Five Safes
5. Information and Data Models

Qualitative Data implications (McLeod, O'Connor and Davis, 2022)

Critique of big data and impact, governance, and measurement

Differentiated from governance of 'small data'

Contextually rich and detailed data typically generated by qualitative inquiries

Its governance signals reconfiguration of knowledge practices

Brings political economy and ethics into sharp focus

Signals transformations in how researchers imagine, conduct, experience research work

Extensions

Organisations

- Affiliations
- Roles of people in organisations
- Organisations provide:
  - Legal status
  - Resources and infrastructure
  - Legal and ethical control systems for affiliates

- Some data sharing agreements require an organisational affiliation

Groups

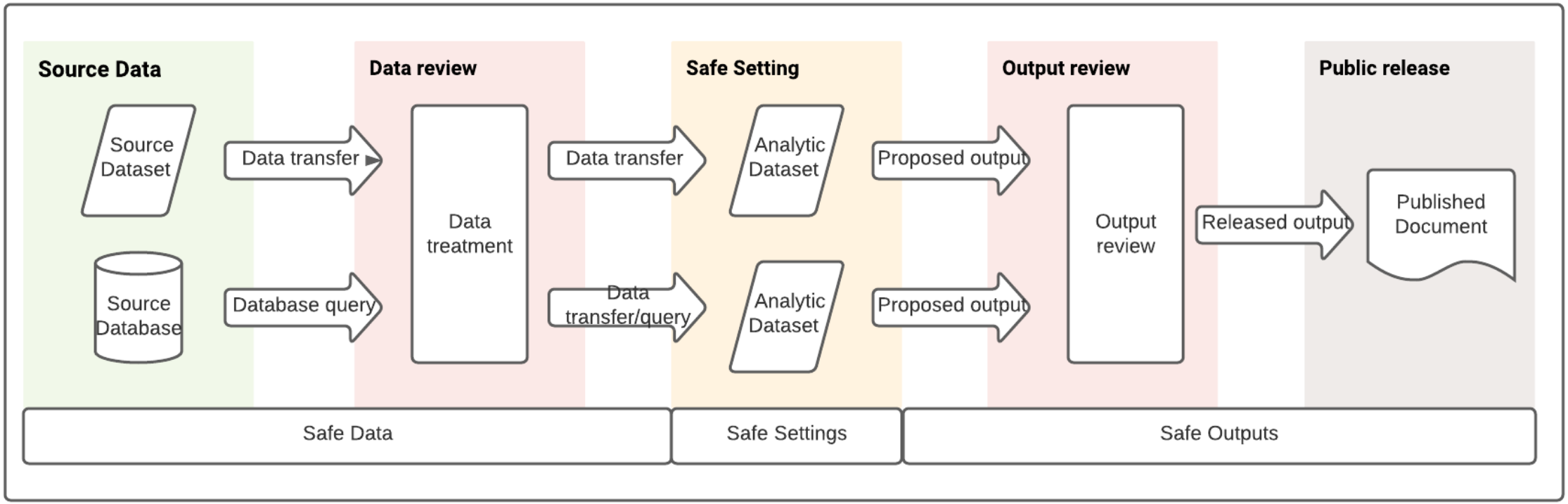
- **People** in teams/organisations
- **Projects** in programs
- **Data** files in datasets
- Datasets in collections
- Use of multiple **settings**
- Collected **outputs** from analysis

- Likely to be CILogon

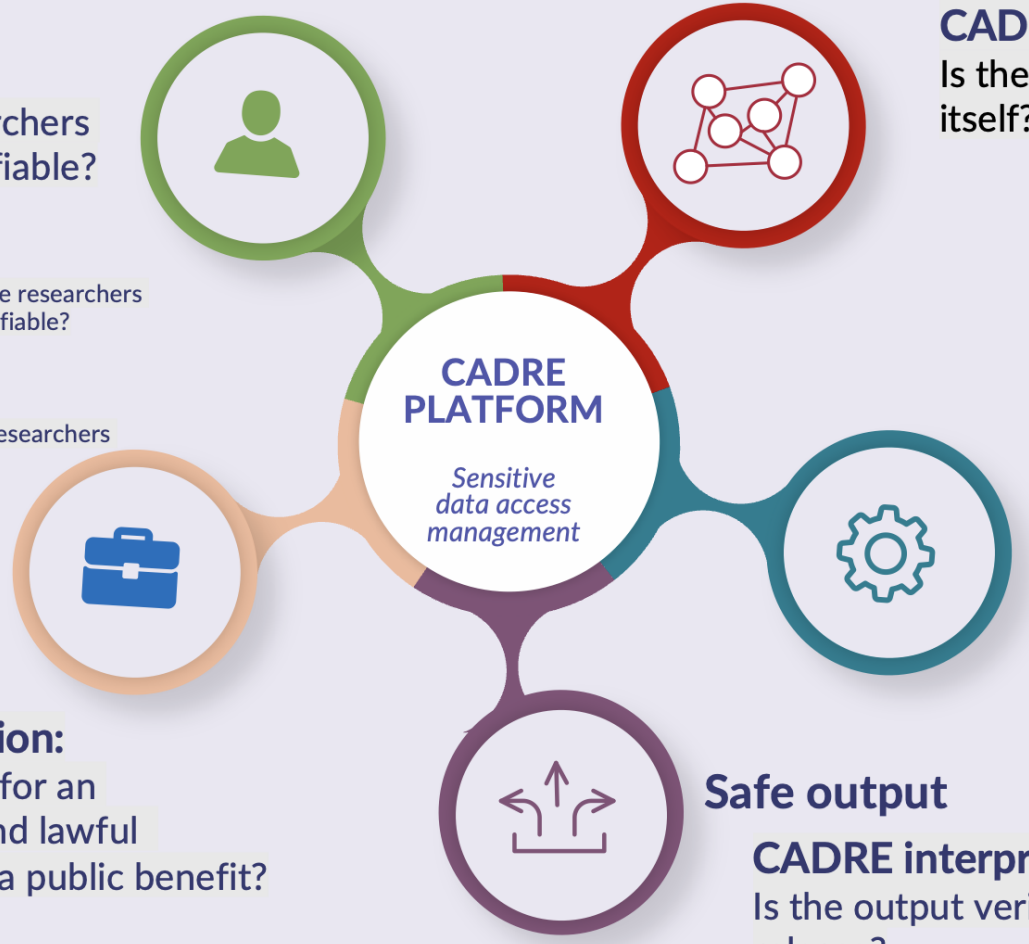
Joint and Severable: Interactions

Intersection	Example	People	Projects	Data	Settings	Outputs
People and Projects	Data custodians assess the characteristics of the person (such as the type of organisation they work in) in assessing the project. E.g. Can a researcher working in a for-profit company conduct research that is primarily for public benefit, or are the benefits largely private?	■	■			
Projects and data	Data custodians will assess the project requirements and research questions in determining the content of the dataset to make available for the project		■	■		
Data and Settings	A data provider will assess the setting and how the data will be accessed in determining what level of treatment to apply to the data prior to transferring it to the setting			■	■	
Settings and outputs	A data custodian will assess the capacity of the setting to allow review of the outputs in assessing it's suitability for use				■	■
People and settings	Users of settings are required to undertake training in the Five Safes and the specific setting before access to the setting is provided.	■			■	
People and outputs	Users of settings are trained in suitable outputs for release as part of training programs – a "virtuous circle" model (Ritchie et al., 2017_	■				■

Safe data, settings and outputs



Conceptual Framework



Safe people

CADRE interpretation:
Are the identities of researchers and their track record verifiable?

Safe organisations

CADRE interpretation:
Is the identity of the organisation, the researchers affiliation, and their track record verifiable?

Safe groups

CADRE interpretation:
Is the identity of the group and the researchers membership verifiable?

Safe data

CADRE interpretation:
Is there a disclosure risk in the data itself?

Safe settings

CADRE interpretation:
Will the data be used in an appropriately safe and secure environment?

Safe projects

CADRE interpretation:
Will the data be used for an appropriate, ethical and lawful purpose that delivers a public benefit?

Safe output

CADRE interpretation:
Is the output verified as suitable for release?

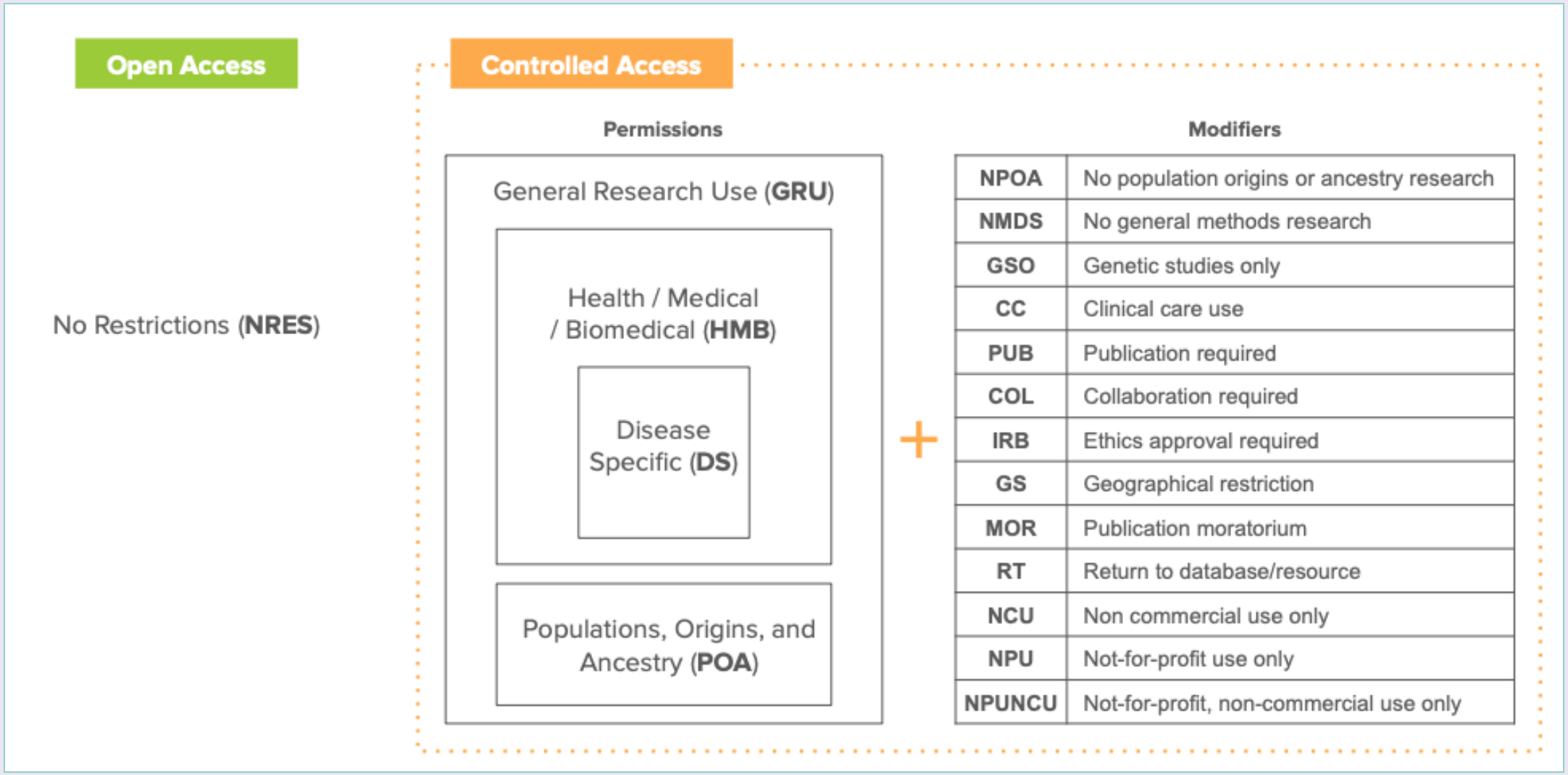
CADRE Conceptual Framework: https://doi.org/10.5281/zenodo.5748610

Operationalisation of the Five Safes

- Data Use Ontology
- Data Tags Suite

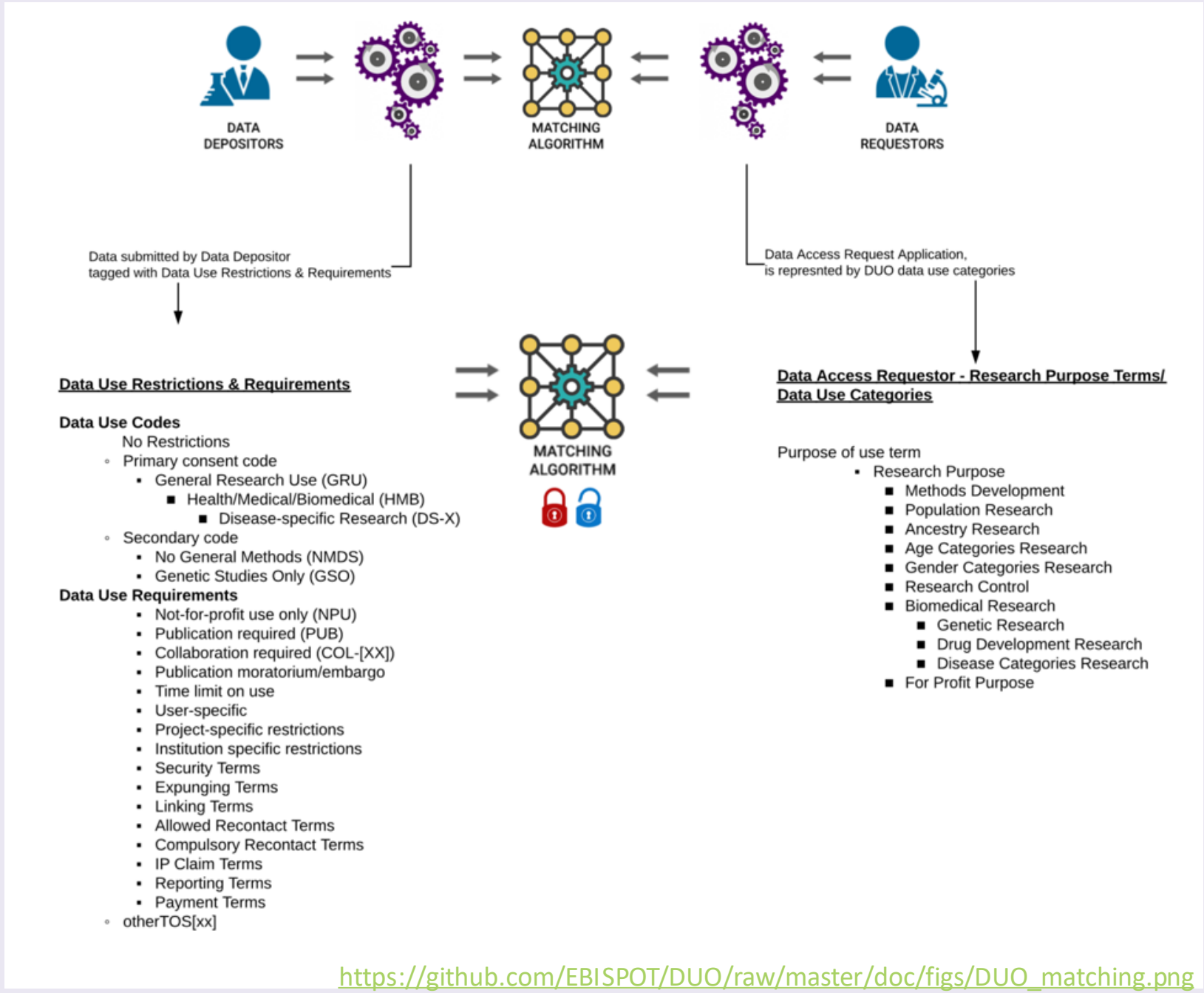
- AAF attributes
- CILogon

Data Use Ontology (GA4GH, ELIXIR)



https://github.com/EBISPOT/DUO

Matching Restrictions and Requests (DUO)



https://github.com/EBISPOT/DUO/raw/master/doc/figs/DUO_matching.png

DUO and the Five Safes … and the ADA catalogue

Australian Survey of Social Attitudes (2020)	Longitudinal Survey of Australian Youth
DOI: http://dx.doi.org/10.26193/C86EZG	DOI: http://dx.doi.org/10.4225/87/PJO7GB
Data Use Limitations • GRU - General Research Use - DUO_0000042	Data Use Limitations • GRU - General Research Use - DUO_0000042
Modifiers • PS - Project specific restriction – DUO_0000027 Custom metadata block?	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 RT – Return to database/resource

TK Labels mapping – maybe DUO too?

Local Contexts

MARC Mapping

 # Notices

```
'traditional_knowledge' => {'tag_number' => '506',  "indicator" => '1',
''}, {'tag_number' => '540', "indicator" => '', 'subfield' => 'a'},
'biocultural' => {'tag_number' => '506',  "indicator" => '1', 'subfield' =
{'tag_number' => '540', 'subfield' => 'a'},
'attribution_incomplete' => {'tag_number' => '561', 'subfield' => ''},
'open_to_collaborate' => {'tag_number' => '500', 'subfield' => ''},
```

 # BC Labels

```
'provenance' => {'tag_number' => '506',  "indicator" => '1', 'subfield'
{'tag_number' => '540', 'subfield' => ''},
{'tag_number' => '561', 'subfield' => ''},
'commercialization' => {'tag_number' => '540', 'subfield' => ''},
'non_commercial' => {'tag_number' => '540', 'subfield' => ''},
'collaboration' => {'tag_number' => '540', 'subfield' => ''},
'consent_verified' => {'tag_number' => '506', "indicator" => '1', 'subfi
{'tag_number' => '540', 'subfield' => ''},
'consent_non_verified' => {'tag_number' => '506', "indicator" => '1', '
''},
{'tag_number' => '540', 'subfield' => ''},
'multiple_community' => {'tag_number' => '506',  "indicator" => '1', 's
''}, {'tag_number' => '540', 'subfield' => ''},{'tag_number' => '561', 'su
''},
'research' => {'tag_number' => '540', 'subfield' => ''},
'clan' => {'tag_number' => '506',  "indicator" => '1', 'subfield' => ''}, {
=> '561', 'subfield' => ''},
'outreach' => {'tag_number' => '540', 'subfield' => ''}
```

TK Labels

Local Contexts

EAD/EAD3 Mapping

 # Notices

```
'traditional_knowledge' => 'accessrestrict', 'userestrict',
'biocultural' => 'accessrestrict', 'userestrict',
'attribution_incomplete' => 'custodhist',
'open_to_collaborate' => 'odd',
```

 # BC Labels

```
'provenance' => accessrestrict', userestrict', 'custodhist',
'commercialization' => 'userestrict',
'non_commercial' => 'userestrict',
'collaboration' => 'userestrict',
'consent_verified' => 'accessrestrict',' userestrict
'consent_non_verified' => 'accessrestrict', 'userestrict
'multiple_community' => accessrestrict', 'userestrict',  'custodhist',
'research' => 'userestrict',
'clan' => accessrestrict', 'userestrict', 'custodhist',
'outreach' => 'userestrict'
```

 # TK Labels

```
'attribution' => 'custodhist',
'clan' => 'accessrestrict', 'userestrict'
family' => 'accessrestrict', 'userestrict',
'outreach' => 'userestrict',
'tk_multiple_community' => 'accessrestrict', 'userestrict',
'non_verified' => 'accessrestrict', 'userestrict'
'verified' => 'accessrestrict', 'userestrict'
'non_commercial' => 'userestrict',
'commercial' => 'userestrict',
'culturally_sensitive' => 'accessrestrict', 'userestrict', 'custodhist'
'community_voice' => 'custodhist',
'community_use_only' => 'accessrestrict', userestrict',
'seasonal' => 'accessrestrict', 'userestrict'
'women_general' => 'accessrestrict', 'userestrict'
'men_general' => 'accessrestrict', 'userestrict'
'men_restricted' => 'accessrestrict',  'userestrict'
'women_restricted' => 'accessrestrict', 'userestrict'
'secret_sacred' => 'accessrestrict', 'userestrict'
'open_to_collaboration' => 'userestrict',
'creative' => 'custodhist', 'odd'
```

Data Tags Suite (Alter et al., 2020)

- Authorisation
- Authentication
- Access

- Aligns with DUO and other standards

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

Applying the 3 As to ADA

Dimension	**ADA Open**	**ADA General**	**ADA Special (e.g. Ten to Men)**
Authorisation	None	Registration*	DUA signed by an individual
Authentication	None	Simple login	Simple login
Access	Download	Download	Download

Combining 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

Sample questions for custodians – defining data access conditions (Source: DUO specification)

(DUL question in Catalog Wizard)	Answer	DUL (Data Use Limitation)	SocialS
Data is available for future general research use [GRU] (required)	Yes	GRU	YES
	No		
Future use is limited for health/medical/biomedical research [HMB] (required)	Yes	HMB	ADAPT
	No		
Future use is limited to research involving the following disease area(s) [DS]	autocomplete	DS={node}	NO
Future commercial use is prohibited [NCU] (required)	Yes	NCU	YES
	No		
Future use by for-profit entities is prohibited [NPU] (required)	Yes	NPU	YES
	No		
Future use for methods research (analytic/software/technology development) outside the bounds of the other specified restrictions is prohibited [NMDS] (required)	Yes	NMDS	YES
	No		
Future use of aggregate-level data for general research purposes is prohibited [NAGR] (required)	Yes	NAGR	MAYBE
	No		
	Unspecified		
Future use as a control set for diseases other than those specified is prohibited [NCTRL] (required)	Yes	NCTRL	MAYBE
	No		
Future use is limited to research involving a particular gender [RS-G] (required)	Male	RS-M	MAYBE
	Female	RS-FM	
	N/A		
Future use is limited to pediatric research [RS-PD] (required)	Yes	RS-PD	MAYBE
	No		

https://cloudstor.aarnet.edu.au/plus/f/5938010800

Sample questions for users - meeting conditions

	USER QUESTIONS/INFORMATION REQUIREMENT					
SHORTHAND	PERSON/(	PROJECT	OUTPUT	COMBINE	Condition	CUSTODIAN_ASSESSMENT
General		Please describe your intended use of this data?				Is this acceptable use?
				Does your project include any of the following uses?		
Health		->		Health/medical/biomedical research		
Disease		->		Specific di	If yes - Is the research focussed on a specific	
Non-commercial		Is this project being undertaken for commer		Commercial benefit		
Non-profit	Is your organisation a for-profit entity (e.g. private business)					
Methods		->		Methods research		
No aggregates			What forms of output will you be producing?			
ControlGroup		Will you be using content from this data set as a control group for comparison purposes?				
Gender		->		Specific ge	If yes - please indicate which gender(s)	
Pediatric		->		Pediatric research		

Aligning standards and the Five Safes

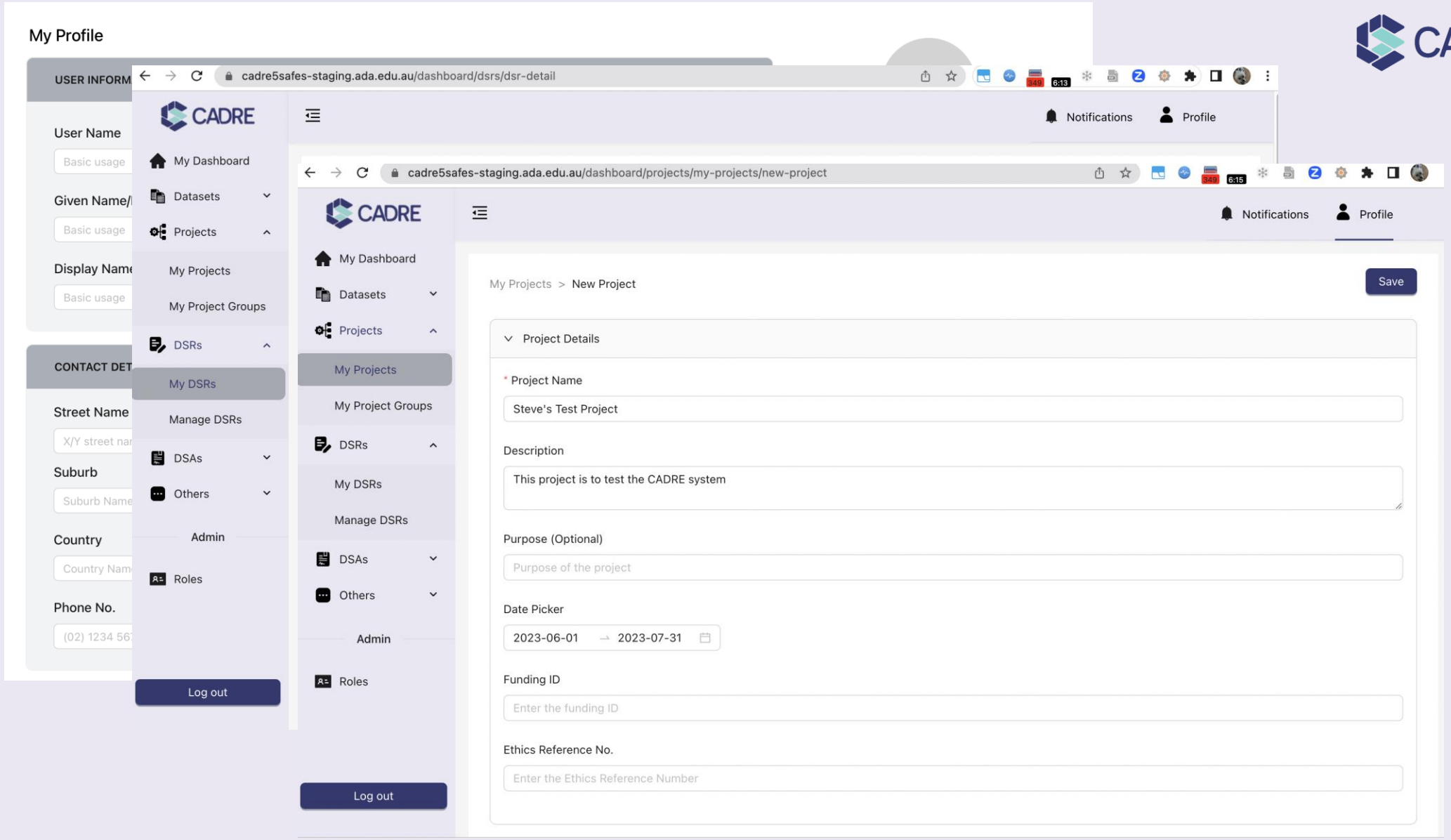
Five Safes dimension	Proposed identifier/PID	Custodian requirements specification	Information source for provision
People	ORCID??	Data Use Ontology (DUO)	AAF attributes, Scholix/ResearchGraph??
Projects	RAID	Data Use Ontology (DUO), Data Tags Suite (DATS)	(CADRE specification)
Data	DOI	???	DataCite, 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

Bringing together the model and the tech

- CILogon: https://www.cilogon.org/
  - AAF partnering in Australia, in pilot with HASS-I Commons, CADRE and Human Genome Project (Aust. BioCommons)
- REMS: https://github.com/CSCfi/rems
  - Production version at Garvan Institute (Aus) and ELIXIR (EU), pilots at Human Genome Project and CADRE
- CADRE Dashboard: ADA/CADRE new development (with Swinburne, Research Graph, AAF and friends)

- Next steps: bringing it all together (ADA Technical Design)
- More on this soon – watch this space!!


My Profile
USER INFORM
User Name
Basic usage
Given Name/
Basic usage
Display Name
Basic usage
CONTACT DET
Street Name
X/Y street nar
Suburb
Suburb Name
Country
Country Nam
Phone No.
(02) 1234 56
cadre5safes-staging.ada.edu.au/dashboard/dsrs/dsr-detail
CADRE
My Dashboard
Datasets
Projects
My Projects
My Project Groups
DSRs
My DSRs
Manage DSRs
DSAs
Others
Admin
Roles
Log out
Notifications
Profile
cadre5safes-staging.ada.edu.au/dashboard/projects/my-projects/new-project
CADRE
My Dashboard
Datasets
Projects
My Projects
My Project Groups
DSRs
My DSRs
Manage DSRs
DSAs
Others
Admin
Roles
Log out
Notifications
Profile
My Projects > New Project
Save
Project Details
* Project Name
Steve's Test Project
Description
This project is to test the CADRE system
Purpose (Optional)
Purpose of the project
Date Picker
2023-06-01 → 2023-07-31
Funding ID
Enter the funding ID
Ethics Reference No.
Enter the Ethics Reference Number

Training Development + Engagement

- Review of the existing training landscape - Gap analysis
- Learning management systems
- Development of content for Five Safes foundation training

Future directions

- Governance
- Coordination
- Technology
- Data

- Establish a national network of safe settings
- Link into a global network of safe settings

Questions and feedback

The framework is out and available through our website and Zenodo, and we would welcome comments.

Contact:

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

https://www.cadre5safes.org.au

Twitter: @cadreaus

DOI 10.5281/zenodo.5748611

https://doi.org/10.5281/zenodo.5748610

The CADRE project received investment (https://doi.org/10.47486/PL106) from the Australian Research Data Commons (ARDC). The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS).