CADRE

Unlocking the Five Safes

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

Partners

ADA Australian Data Archive
Australian Access Federation AAF.EDU.AU
aarnet Australia's Academic and Research Network
UNSW Sydney
The University of Melbourne
Australian Government Australian Institute of Family Studies
Australian Government Australian Institute of Health and Welfare
The University Of Queensland Australia
ResearchGraph Foundation
AURIN aurin.org.au
Australian Research Data Commons
Swinburne Social Data Analytics Lab

Acknowledgement of Country

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

Housekeeping

- Phone calls
- Toilets
- Questions
- Break

Schedule

Aim: For you to understand the Five Safes and the role they play

1. Background
2. Safe Project
3. Safe People
4. Two additional safes: Organisations + Groups
5. Break 15:00 – 15:30

6. Safe Data
7. Safe Settings
8. Safe Outputs
9. Q + A

Image: Flaticon.com

Polling

CADRE

Respond at **PollEv.com/cadre**

Text **CADRE** to **+61 427 541 357** once to join, then **A, B, C, or D**

Which best describes you?

HDR student	**A**
Academic	**B**
Professional	**C**
Other	**D**

Polling

When poll is active, respond at **PollEv.com/cadre**

Text **CADRE** to **+61 427 541 357** once to join

My level of understanding of the Five Safes is

- Boss level. I can help facilitate if you need.
- I've heard of it somewhere. Refresh my memory.
- Am I in the right workshop? This isn't a locksmiths convention, is it?
- non-existent.

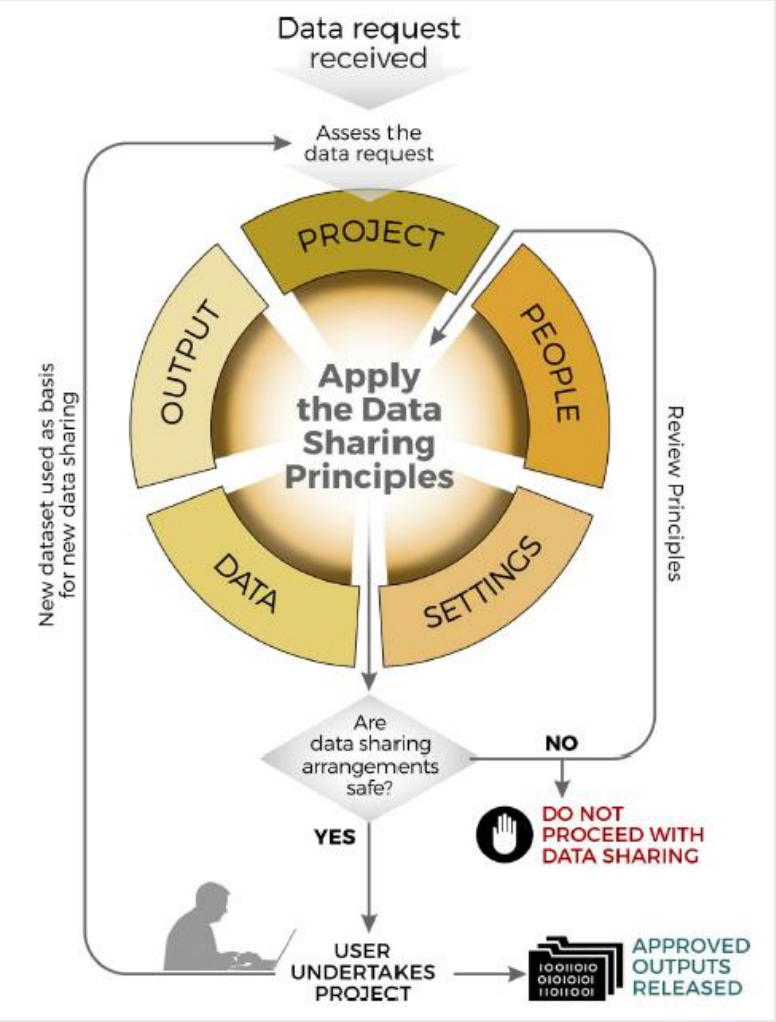
Frameworks for data sharing

Academic:

- FAIR data

Government:

- DATA Scheme (from the DAT Act)
- Equivalent state schemes (e.g. South Australia)
- Agency-specific frameworks (e.g. ABS)



Source: ONDC (2019)

Protection vs. utility

- Custodians have a responsibility to minimise disclosure risk
  - Domain of “statistical disclosure control” established to support this
- Utility of data considers how usable and fit-for-purpose the data is for the analysis task at hand
- Trade-off between utility and protection – higher protections generally result in lower utility
- Application of the Five Safes – allows for management of these trade-offs across multiple dimensions of risk

The Five Safes

Safe Project

Safe Person

Safe Data

Safe Setting

Safe Output

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 use?	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

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.

Five Safes Video

Access the video online here.

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.

CADRE

In your current organisation, what mechanisms, frameworks, policies or training is in place to manage data safely?

Join by Web

1. Go to **PollEv.com**
2. Enter **CADRE**
3. Respond to activity

Join by Text

1. Text **CADRE** to **+61 427 541 357**
2. Text in your message

Total Results: 0

Powered by Poll Everywhere

Why does this matter?

- Researchers/Academics are experts in their fields but not necessarily in all things data.
- Five Safes provides guidance on how sensitive data accessed safely.
- Builds confidence & trust among the whole community – data users, data owners, data custodians and the public.
- Upskilling and supporting those who use sensitive data.

- What can we influence or change in our own organisations?

Safe Project

Is this safe use of the data for public benefit?

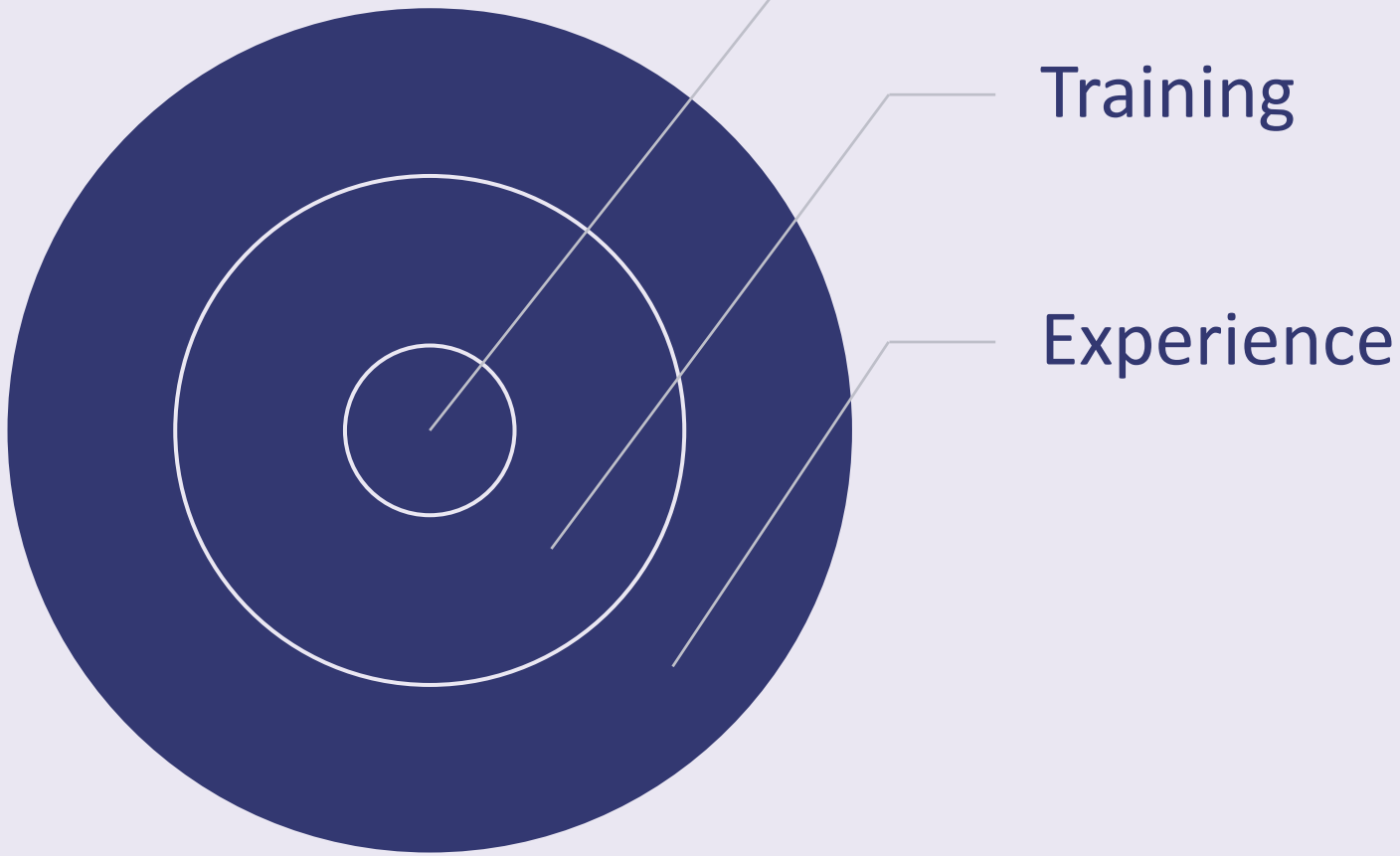
CADRE Considerations

- **Intended use**: Fitness for purpose; public benefit; academic contribution and commercial benefit
- **Context**
- **Ethics**
- **Risks**: Confidentiality risks; reputational risks and commercial risks
- **End user of the project**
  - Who is the sponsor of the research?
  - Who is the research being conducted for?
  - Who is funding the research?

(McEachern, 2021)

Safe People

Appropriate authority to access the data



Safe people

Desai, Ritchie and Welpton (2016):

- Do the users have the necessary technical skills?
- Do the users need training in handling confidential data?
- Are users likely to follow procedures?

CWG additions:

- Past track record
- Institutional affiliation
  - Institutional rules
  - Institutional support
  - Institutional legitimacy
  - Considered further under “Organisations”
- Examples: GA4GH and ICPSR “Researcher Passports”

Safe People

What role does credentials, training and experience play in determining someone is ‘safe’?

Two Additional Safes

Organisations

Groups

Safe Organisations + Groups

CADRE Considerations

Organisations:

- **Legal implications**
- **Resources and infrastructure**
- **Legal and ethical controls**
- **Data sharing**

Groups:

Five Safes Dimension	Grouping
Safe people	People working in Research teams
Safe Projects	Projects completed as part of larger Work Programs
Safe Data	A linked dataset resulting from linkage or integration of multiple upstream Datasets
Safe Settings	Use of data from multiple sources, located in multiple settings
Safe Outputs	Multiple publications outputs resulting from a completed analysis

(McEachern, 2021)

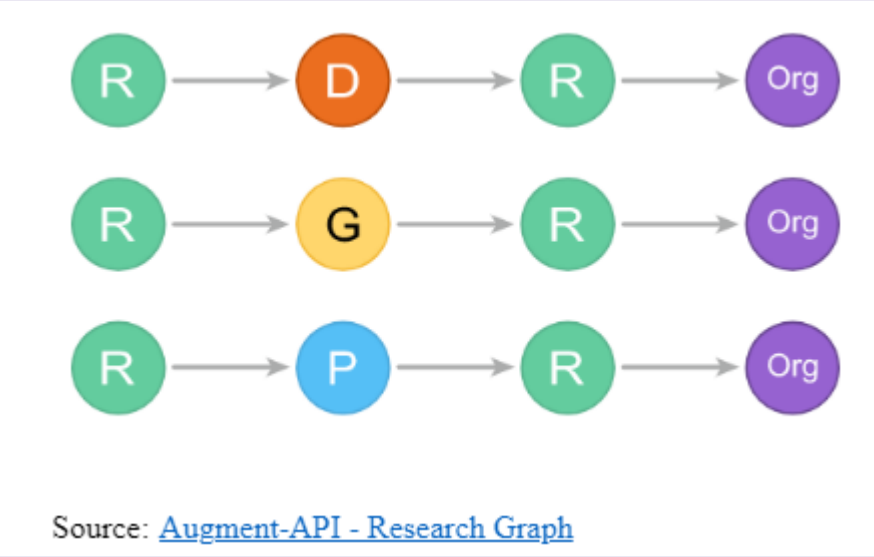
Research Graph Augment API

Research Graph (researchgraph.org) provides information about the following research-related objects:

- Researcher
- Publication
- Dataset
- Grant
- Organisation

Research Graph Augment API

Research Graph Augment API is a RESTful API that transforms metadata of research objects into a collaboration network. The API transforms a single Open Researcher and Contributor ID (ORCiD) or Digital Object Identifier (DOI) to a connected graph consist of the following connections:

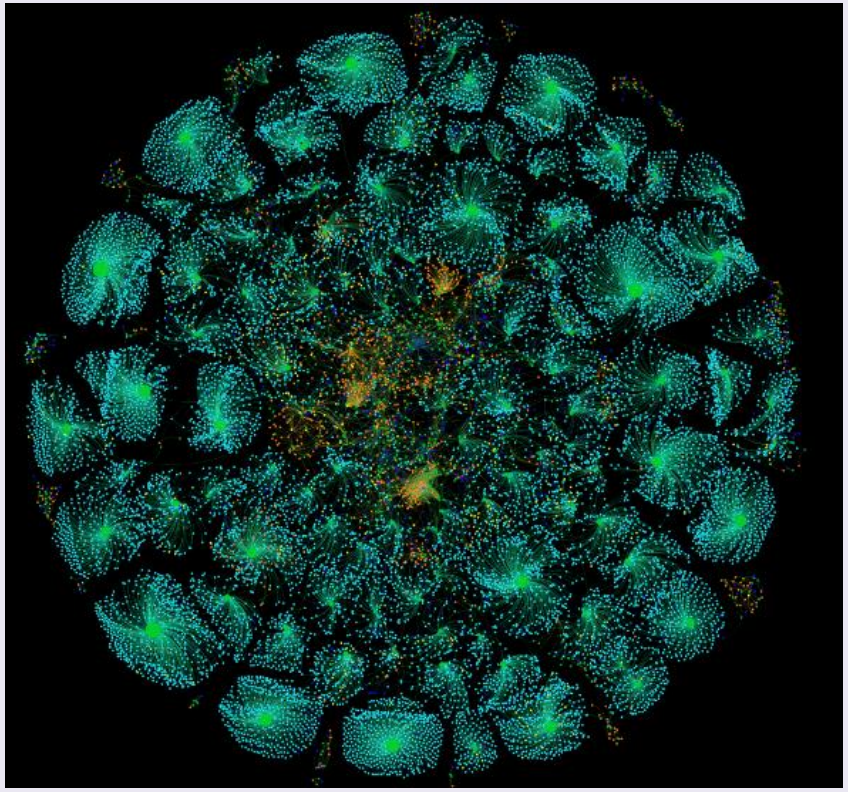


Source: Augment-API - Research Graph

https://researchgraph.org/augment-api/

Research Graph Augment API

Source: Research Graph ADA – augmentation with cross reference



BREAK until 15:30

Safe Data

Appropriate and proportionate protections are applied to the data

CADRE

Data Lifecycle

Safe Data

When you think of safe data, what issues come to mind for you?

Safe Data

Usability of the data

CADRE Considerations

Usability:

- Have different versions of the data available
- Availability of structural and contextual metadata (data dictionaries, project descriptions etc.)
- Availability of ‘test’ or synthetic versions of the data

Integrity:

- **Contextual information**
- **Data treatment**
- **Curatorial process**

Access:

- User characteristics
- Limit on purposes
- Time-limited access
- Notification – review of outputs
- Licenses and other relevant documentation

(McEachern, 2021)

Safe Data

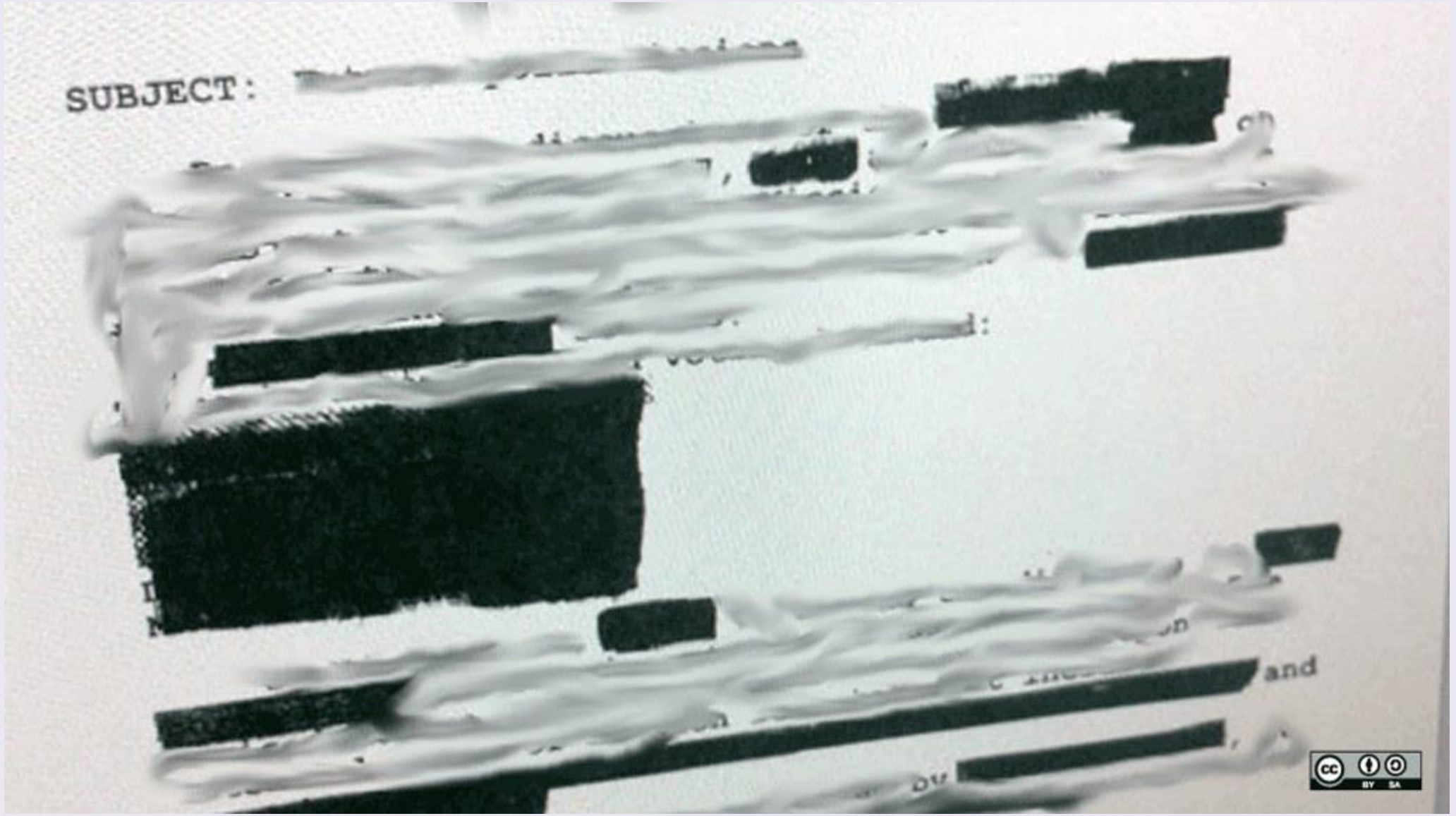
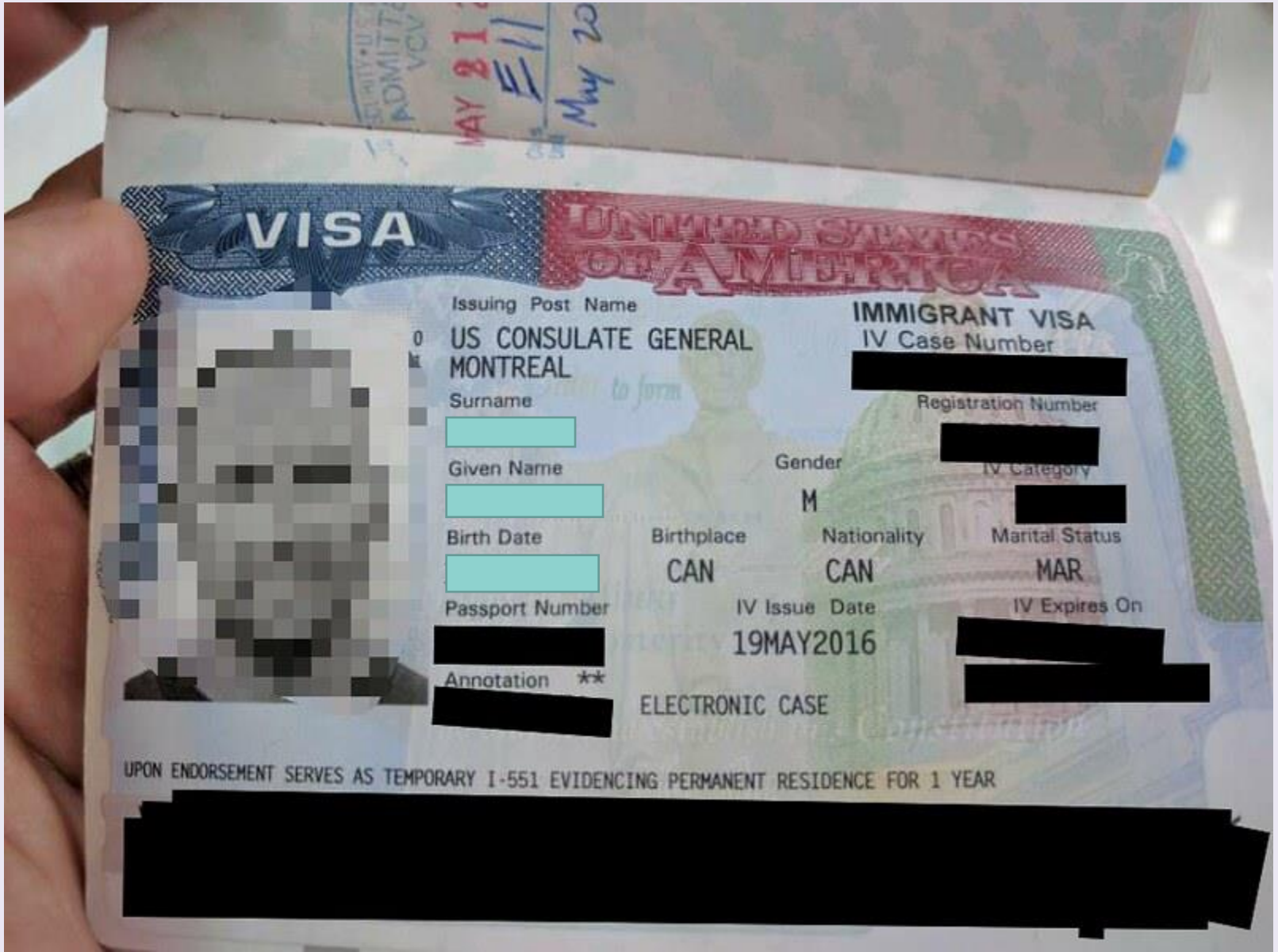


Image: opensource.com

Safe Data

CADRE



Safe Setting

Physical, technical and informational context for data access

- “whether all parties have taken reasonable steps to ensure data will be used in an appropriately safe and secure environment, i.e. one that minimises unauthorised use, access or loss of data”
- Physical environment
- IT environment
- (Training in use of settings)

DATS Descriptors for 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

Safe Setting

Audience Questions:

Does your organisation provide a ‘safe setting’?
What is it & why is it safe?
Can you name some other safe settings within Australia?

Safe Output

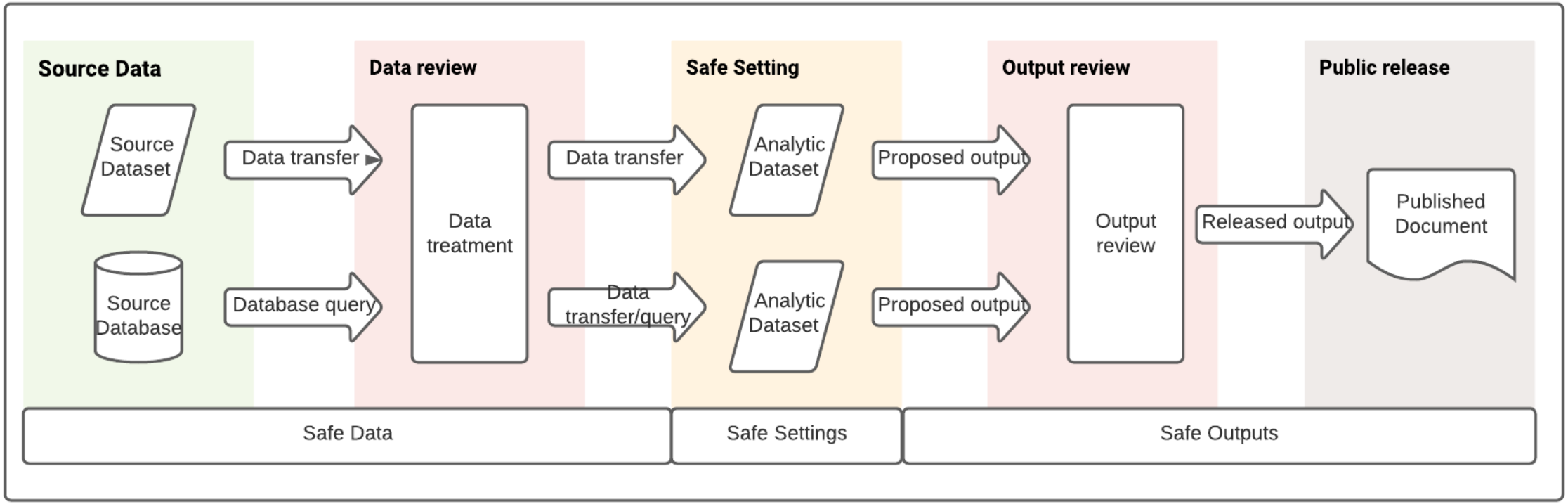
Characteristics of a principles-based output statistical disclosure control

CADRE Considerations

- Researchers and output checkers both trained in SDC
- Rules-of-thumb rather than hard rules
- Freedom to approve any output in principle
- No duty to release any output
- Responsibility for producing good output resting with the researcher
- Output checkers considering the value of the output
- Output checkers considering resource constraints

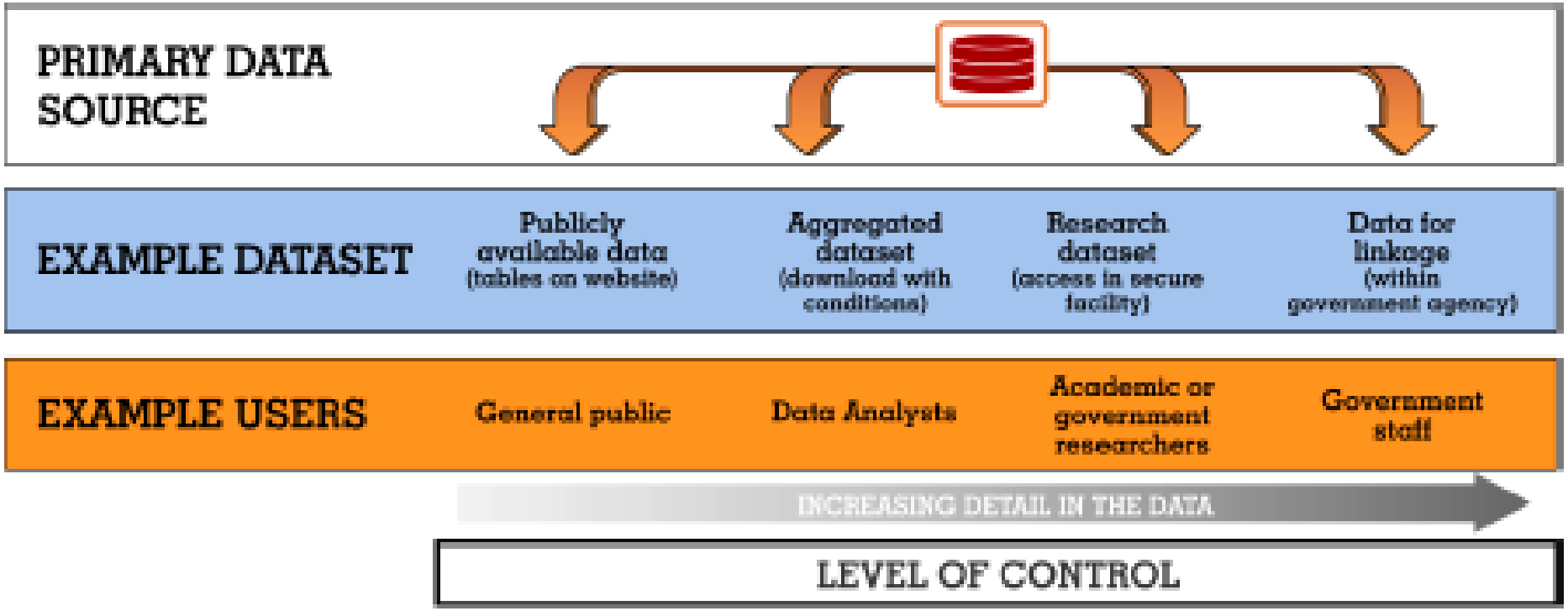
(McEachern, 2021)

Safe data, settings and outputs



Applying the Five Safes model

Diagram 2: the level of project control will depend on the level of detail being shared



PROJECT	None	Medium	High	Very High
PEOPLE	None	Low	High	Very High
SETTING	None	Low	High	Very High
DATA	Very High	High	Low	Very Low
OUTPUT	None	Low	High	Low

Example: Applying the Five Safes across data types

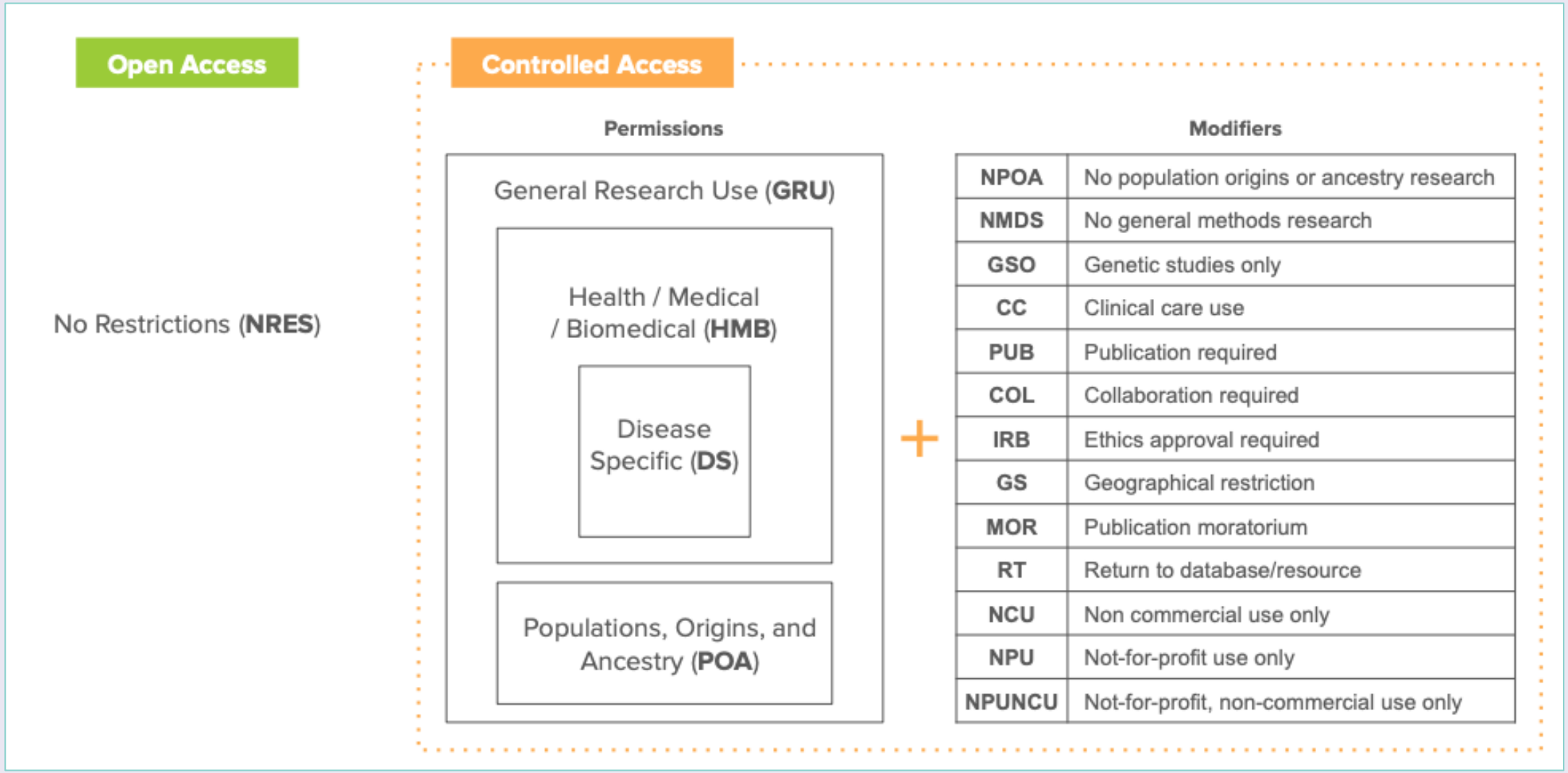
Safe	Census aggregates	ANU Poll unit record files	MADIP
Data	Aggregate count files at ASGS levels	Confidentialised unit record files	Deidentified administrative records
Projects	None required	Specified in application	Specified in application
People	None	User information provided and validated	User information provided and validated, data sharing agreement signed
Settings	ABS website – open access, download	ADA Dataverse – restricted access, download	ABS Lab
Outputs	None	None	Output checking before release

Questions?

Thank You

Follow us on Twitter: @cadreaus

Data Use Ontology



DUO and the Five Safes

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

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

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

Custom metadata block?