



**ABS DATA INTEGRATION
& DIGITAL SERVICES**

**ABS DATA INTEGRATION
& DIGITAL SERVICES**

DataLab

DataLab Safe Researcher Training

Overview

Part 1 - Working together

(60 minutes)

- ABS vision for the DataLab
- Shared responsibility
- Five Safes Risk Framework

Break (10 minutes)

Part 2 - Maintaining data confidentiality

(40 minutes)

- What does that mean?
- Why is it important?
- Your role and the ABS' role

Break (10 minutes)

Part 3 - Statistical disclosure control

(60 minutes)

- How might disclosure occur?
- Making outputs safe
- Output Rules

Training Outcomes



Understand your role and the role of the ABS

Understand how the five safes framework underpins ABS disclosure risk assessment

Know how to apply statistical disclosure control to your output

DataLab 101



Q1. When should data be protected?

- A. Only if the data are sensitive
- B. Unless the data are already in the public domain
- C. When data are deemed to be personal
- D. Always

Q2. Which is the most common reason behind breaches of procedure when sharing data?

- A. Mistakes
- B. Ignorance
- C. Laziness
- D. Malicious intent
- E. Dislike of procedures

Q3. Risks from sharing data should be?

- A. Minimised
- B. Controlled using subjective measures
- C. Absolutely zero
- D. Controlled using objective measures

Q4. Researchers are best supported by?

- A. Giving them data to do with as they please
- B. Only letting them have open data
- C. Ensuring they understand their rights and responsibilities
- D. Ensuring they understand that they will be prosecuted if they do the wrong thing

Video by Dr Felix Ritchie

A video player interface with a black background. At the top center, there is a white audio waveform and a small Australian Bureau of Statistics logo. On the left side, there is a small inset video showing a man in a suit speaking. The main area of the video player shows a presentation slide with a white background and a grey footer. The slide text is "Culture 1: attitudes" in red. The footer contains the UWE Bristol logo and the text "University of the West of England BRISTOL".

Culture 1: attitudes

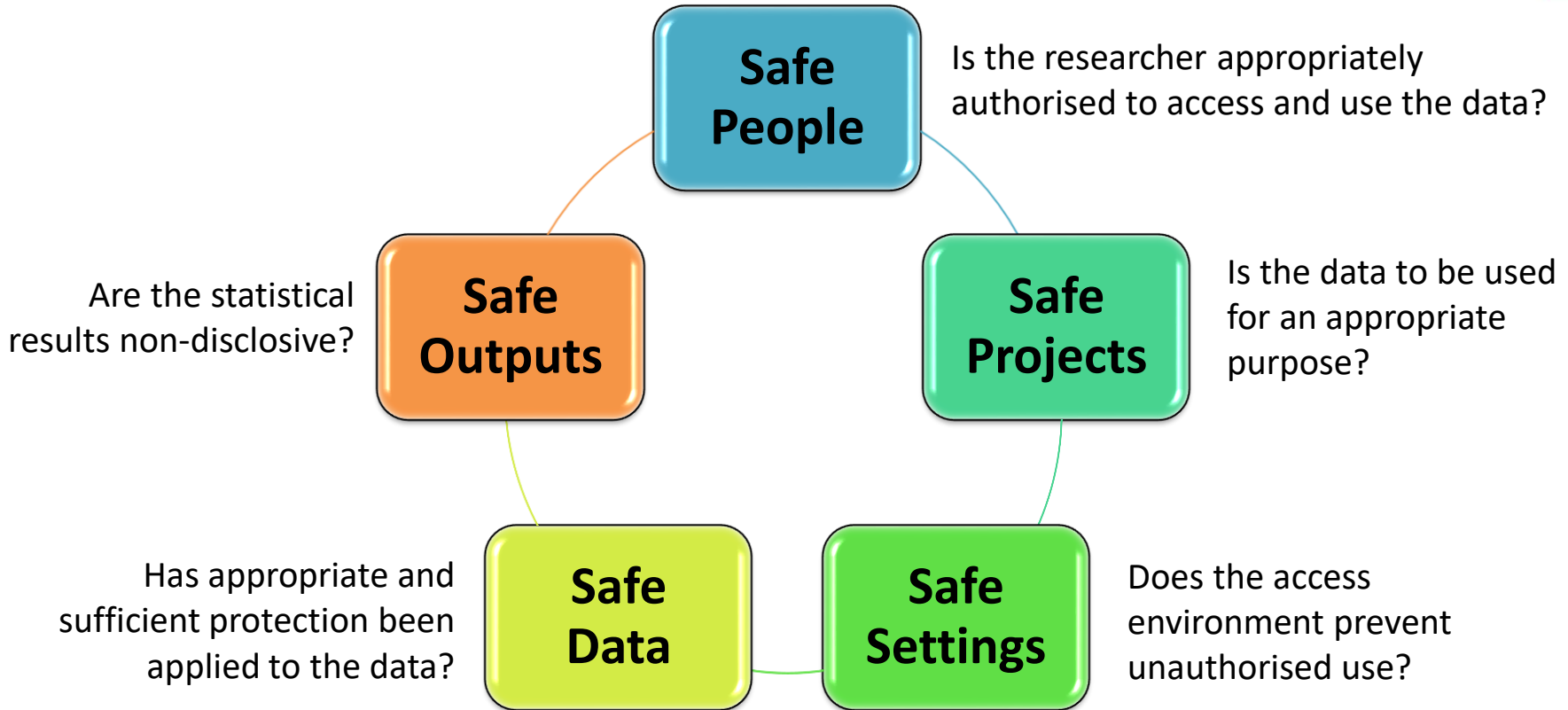
UWE Bristol University of the West of England BRISTOL

Consider what could go wrong



1. A researcher has ethical approval to study outpatient outcomes from hospital data and demographics data. The data are emailed separately and stored on the researcher's laptop.
2. A multi-institutional (university and government) team of researchers access business data in a controlled Federal government facility to investigate sole trader survival.
3. Government staff in 4 agencies linked identified data from their agencies to create an enduring dataset that will be available for policy development.

Five Safes risk framework



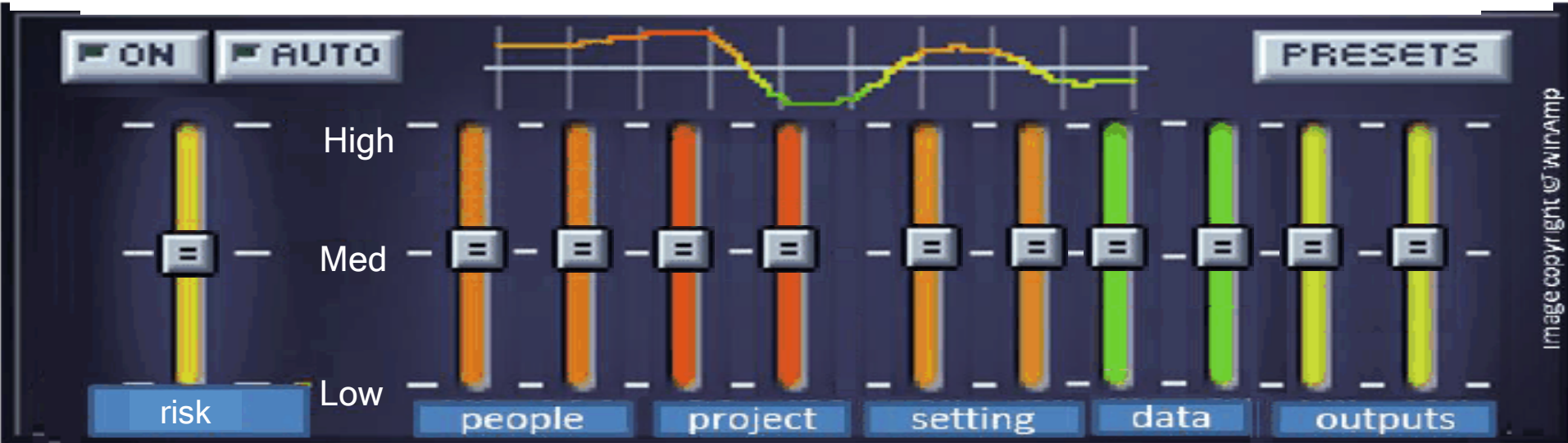
Exercise: Five safes – the controls



Access to detailed microdata in the DataLab

Risk: What is the overall disclosure risk?

Low



Adapted based on image from 'Five Safes: designing data access for research', Desai, T, Ritchie, F and Welpton, R, 2016

Five safes – the data



All data in the DataLab should be treated as potentially disclosive

Disclosure risks can be largely managed by non-statistical processes

Detailed data is available for use by researchers in the DataLab, but the final outputs need to be checked for disclosure before being released for use outside the DataLab

Video: Dr Felix Ritchie

A video player interface with a black background. At the top center, there is a white audio waveform and a small Australian Bureau of Statistics logo. On the left side, there is a small inset video showing a man in a suit speaking. The main area displays a presentation slide with a white background and a grey footer. The slide title is "Culture 2: users as colleagues" in red. Below the title is a bullet point: "• Greater awareness of value of collaboration". The footer contains the University of the West of England logo and the text "University of the West of England".

Culture 2: users as colleagues

- Greater awareness of value of collaboration

University of the West of England
BRISTOL

Recap of part 1



Shared responsibility is critical to safe and effective use of data in the DataLab

The ABS has adopted the five safes framework for managing disclosure risk

Collaboration and feedback are important to ensure the DataLab works for everyone



**ABS DATA INTEGRATION
& DIGITAL SERVICES**

**ABS DATA INTEGRATION
& DIGITAL SERVICES**

DataLab

Break – 10 minutes

Overview

Part 1 - Working together

(60 minutes)

- ABS vision for the DataLab
- Shared responsibility
- Five Safes Risk Framework

Break (10 minutes)

Part 2 - Maintaining data confidentiality

(40 minutes)

- What is it?
- Why it's important
- Your role and the ABS' role

Break (10 minutes)

Part 3 - Statistical disclosure control

(60 minutes)

- How might disclosure occur?
- Making outputs safe
- Output Rules

Discussion



Question: Is confidentiality our highest priority for microdata in the DataLab?

Answer: No

- Our highest priority is providing researchers access to useful microdata
- Confidentiality is now a constraint

But We need a high priority on confidentiality when data leaves the DataLab and is made public

Data confidentiality



All users accessing data in the DataLab are legally required to protect data confidentiality

Legal and non-legal sanctions apply to any DataLab breaches

Not complying with DataLab terms and conditions has consequences

Understanding legal issues and impacts



Looking at the following scenarios

- Is there a breach of confidentiality?
- Is there a breach of procedure?
- Are there any other issues (e.g. ethical)?

Scenario 1



A researcher is accessing data in the DataLab. They take a screen shot of a table they have just created to show their research team in a regular project meeting. They email the screen shot to their supervisor to review the content before the regular meeting.



Are there any legal, ethical or procedural issues here?

Scenario 2



You have just been granted access to the DataLab. You log into the system and notice you have access to data that you did not request. The data looks like it might be suitable for a future project you have in mind, but you have questions about the data. You email your friend who works at the ABS and ask for help.



Are there any legal, ethical or procedural issues here?

Scenario 3

A researcher presents a seminar at a conference using analysis from business unit record data. In addition to their vetted analysis, they describe data about a business with particular characteristics, located in a small geographical area. Someone in the audience believes that the business being described is a manufacturing firm in SA.

Are there any legal, ethical or procedural issues here?



Security Protocols



Do not capture on-screen information in any way during your DataLab session

Do not screen share unless using LabLink with other approved users on the same project

When accessing DataLab you must:

- access the data in a private location
- protect the screen from oversight by other people, and
- use a secure internet connection

Do not share your login details

Statistical experience and referral form



All users including discussants must complete the Declaration of Compliance form

You must have:

- at least three years of either quantitative research experience, or university study with a significant component working with quantitative data, or
- complete the DataLab referral form if you don't think you meet the above criteria

Publish outputs



When do I need to inform the ABS I am publishing?

- Any publication, report and presentation that references BLADE or PLIDA data needs to be provided to the ABS a **minimum of 2 weeks** prior to wider release.
- This is a requirement of our Data Custodians.

Further information available at:

<https://www.abs.gov.au/statistics/microdata-tablebuilder/datalab/using-datalab-responsibly#publishing-and-citing-data>

Your role



Do not disclose any data

Do not attempt to circumvent the system

Follow the rules and processes as outlined in the undertakings and declarations you signed

Complete DataLab refresher training every 2 years or as directed

Our role



Check outputs and provide you with advice on how to make your outputs non-disclosive

Respect your academic independence

Respond to any questions related to the data, processes and systems

DataLab User Guide



- [Logging into the portal and workspace](#) – log into DataLab portal, launch VM, activate VM, launch desktop
- [Using your workplace](#) – accessing your data files, available software, virtual machines, databricks
- [Portal features](#) – VM management options, functions in my projects, recommended browsers
- [Troubleshooting](#) – authentication, errors, code and software

More information



For more information:

- [Conditions of Use](#)
- [Using DataLab responsibly](#)
- [Output and input clearance](#)

If you are ever unsure, always reach out and ask using our [Contact Us](#) web page

Part 2 recap



ABS staff and researchers are legally and ethically required to maintain data confidentiality

You and the ABS play a role in maintaining data confidentiality

Breaches of confidentiality and not complying with an undertaking potentially have legal and/or non-legal consequences

Legislation protects confidentiality and privacy whilst enabling access to detailed information in special circumstances



**ABS DATA INTEGRATION
& DIGITAL SERVICES**

**ABS DATA INTEGRATION
& DIGITAL SERVICES**

DataLab

Break – 10 minutes