**DataLab Safe Researcher Training**

**General Housekeeping**

**Overview**

Part 1 - Working together

Break (10 minutes)

Part 2 - Maintaining data confidentiality

Break (10 minutes)

Part 3 - Statistical disclosure control

• ABS vision

• Shared accountability

• Five Safes Framework

• What is it?

• Why it’s important

• Your role and the ABS’ role

• How might disclosure occur?

• Making outputs safe

• Rules of thumb

2

**Training outcomes - At the end of today you will**

▪ **Understand** your and the ABS’ roles in shared

accountability around the access and use of detailed

microdata in the DataLab

▪ **Understand** how the five safes framework underpins ABS

disclosure risk assessment

▪ **Know** how to apply statistical disclosure control

3

**Introduction**

In 30 seconds (or less) introduce yourself and tell us:

• The data you will be working with, i.e., MADIP, BLADE

or other survey data

• Your role on the project, i.e., analyst or manager/

support person

4

**How much do you already know?**

Q1: When should the confidentiality of 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.

|  |  |  |
| --- | --- | --- |
| Q3. Risks from | A. | Minimised. |
| sharing data should | B. | Controlled using subjective |
| be? |  | measures. |
|  | C. | Absolutely zero. |
|  | D. | Controlled using objective |
|  |  | measures. |



|  |  |  |
| --- | --- | --- |
| Q2. Which is the most | A. | Mistakes |
| common reason behind | B. | ignorance |
| breaches of procedure | C. | Laziness |
| when sharing data? | D. | Malicious intent |
|  | E. | Dislike of procedures |

the wrong thing.

|  |  |  |
| --- | --- | --- |
| Q4.Researchers are best | A. | Giving them data to do with as they please. |
| supported by? | 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 |
|  |  |  |

**Evaluation of first exercise**

**Knowledge**

**Generalist**

**Specialist**

Q1 –Q4

**Experience**

Today: mostly obvious, lots of thinking

6

**Video: Dr Felix Ritchie**

Associate Professor, University of the West of England

**Consider what could go wrong**



|  |  |
| --- | --- |
| Group A | A PhD student downloads DSS and DoH datasets from data.gov.au toundertake preliminary study into health outcomes |
| Group B | A researcher has ethical approval to study outpatient outcomes from hospital data and demographics from DHS data. The data are emailed separately and stored on the researchers laptop |
| Group A | A multi-institutional (university and government) team of researchers access business data in a controlled Federal government facility to investigate sole trader survival |
| Group B | 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 framework**

**Safe**

**People**

Is the researcher appropriately

authorised to access and use the data?

Are the statistical results non-disclosive?

**Safe**

**Outputs**

**Safe**

**Projects**

Is the data to be used

for an appropriate

purpose?

Has appropriate and

sufficient protection been

applied to the data?

**Safe**

**Data**

**Safe**

**Settings**

Does the access environment prevent unauthorised use?

**Exercise: Five safes - Control**

Access to detailed microdata in the DataLab

**Risk: What is the overall disclosure risk?**

**Low**

High

Med

**Safe data: Has appropriate and sufficient protection been applied to the data?**

**Safe outputs: Are the statistical results non-disclosive?**

**Safe setting: Does the access environment prevent unauthorised use?**

**Safe project - Is the data to be used for an appropriate purpose?**

**Safe People - Is the researcher authorised to access and use the data**

**appropriately?**

risk

Low

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

**Five safes – Other ABS data types**

Detailed microdata in the DataLab Microdata download

Risk Risk

ABS TableBuilder ABS website

Risk Risk

**Five safes - 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**

Associate Professor, University of the West of England

**Recap of part 1**

✓ Shared accountability is critical to safe and effective use

of microdata

✓ The ABS has adopted the five safes framework for

managing disclosure risk

✓ Collaboration and feedback will improve ABS microdata

and access processes

**Break – 10 mins**

**Overview**

Part 1 - Working together

(50 minutes)

Break (10 minutes)

Part 2 - Maintaining data confidentiality

(40 minutes)

Break (10 minutes)

Part 3 - Statistical disclosure control

(50 minutes)

• ABS vision

• Shared accountability

• Five Safes Framework

• What is it?

• Why it’s important

• Your role and the ABS’ role

• How might disclosure occur?

• Making outputs safe

• Rules of thumb

16

**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 is

made public (covered in Part 3)

**Data confidentiality**

▪ All users accessing microdata are legally required to protect data confidentiality

▪ Legal and non-legal sanctions apply to any breaches and not complying with 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)?

**Exercise - Scenario 1**

A researcher is accessing a Detailed Microdata file in the DataLab

environment. They take some notes on paper from the screen about

records with particular characteristics (e.g. crime records). They do

not show the ABS the paper notes for vetting, and subsequently

misplace the piece of paper.

***Group A: identify the legal, ethical or procedural issues here?***

***Group B: Convince us this scenario is fine.***

**Exercise - Scenario 2**

A government officer is conducting research in the DataLab. Their supervisor is also named on the project, although the supervisor is not actively involved in the research. They discover some really interesting results and are excited to share it with their supervisor. They send an email to their supervisor with a copy of the analysis.

***Group B: identify the legal, ethical***

***or procedural issues here?***

***Group A: Convince us this scenario***

***is fine.***

**Exercise - 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 particular region. Someone in the audience believes

that the business being described is a manufacturing firm in SA.

***Group A: identify the legal, ethical or procedural issues here?***

***Group B: Convince us this scenario is fine.***

**Your role**

▪ Key things to remember:

▪ Do not disclose any microdata

▪ Do not attempt to circumvent the system

▪ Do follow the rules and processes of the DataLab system

▪ For your ongoing reference, there is a longer list of DO’s and

DON’T’s online in the *ABS DataLab User Guide* and the pre-reading

▪ And, if you are ever unsure, always reach out and ask us by emailing: mailto:data.services@abs.gov.au

**ABS role**

▪ Check outputs and provide you with advice on how to make your

outputs non-disclosive

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

▪ Respect your academic independence

▪ For more information, refer to the *ABS DataLab User Guide*

**Security protocols**

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

▪ When accessing DataLab you must:

• access the microdata in a work or private location

• protect the screen from oversight by other people, and

• use a secure internet connection

▪ Do not share login details with anyone

▪ **This is critical for maintaining confidentiality and managing public perception**

**regarding data care**

▪ [For more detail, please read: https://www.abs.gov.au/statistics/microdata-](file:///%5C%5Ccorp%5Cabsdfs%5Cworkgroup%5CDIPAC%5C03.%20Access%5C19.%20DataLab%20Client%20Support%5CDataLab%5CDataLab%20Client%20Training%5CVirtual%20Training%20Slides%5CWeb%20page%20docs%5CFor%20more%20detail%2C%20please%20read%3A%20%20https%3A%5C%5Cwww.abs.gov.au%5Cstatistics%5Cmicrodata-) tablebuilder/responsible-use-abs-microdata/microdata-user-obligations

**Recap of part 2**

✓ ABS staff, inposted officers and researchers are legally and ethically required to maintain 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

✓ You and the ABS play a role in maintaining data confidentiality

**Break – 10 mins**