Safe access arrangements in a modern world

Felix Ritchie

Bristol Business School, University of the West of England, Bristol
Outline

• Modern approaches to safe access
• challenges for government
• challenges for academia

• example: principles-based output SDC
Safe access for grown-ups
The traditional govt. view

• Access to data is an option
• Access is fundamentally risky
• Users = intruders
• Statistical tools manage risk
• User agreements manage responsibility
• If in doubt, do not release data/outputs
The traditional academic view

• Access to full data is essential
  – data restrictions limit research potential

• Data owners are excessively risk-averse
  – and don’t really know or care what users do
  – their main aim is to protect themselves

• Academics know what they’re doing
The modern perspective

• Access to data is assumed
• Confidentiality introduces constraints
• This is an uncertain world
  – Let’s aim to make ‘most reasonable’ choices
• We are all human
The modern perspective, cont.

• Non-statistical risks are key concerns
• Non-statistical solutions address them
  – statistical solutions are the residual
• Relationships crucial
  – data subjects, data collectors, ethics committees, governance committees, user support teams, users user users users...
Modern data access assumes…

• life is uncertain and decisions subjective but we must make decisions so...
• default-open: all data is available – but data/access may need to be constrained
• confidentiality concerns: evidence-based
• shared responsibility between all parties
Implications for government

• Seriously scary
  – explicitly acknowledging unmanaged risks
    • not under their control, can’t be signed away
  – data release not under their control
  – no hiding behind statistical tools

• Needs training

• Needs a lot more active engagement
  – Needs a good understanding of users
Implications for academia

• Unappealing
  – no unconstrained right to data
  – having to take responsibility for data safety

• Need training

• Needs a lot more active engagement
  – Needs a good understanding of data owners
Why does this happen?

because we’re not idiots

Implications for government

• Seriously scary
  – explicitly acknowledging unmanaged risks
    • not under their control, can’t be signed away
    – data release not under their control
    – no hiding behind statistical tools

• Needs training

• Needs a lot more active engagement
  – Needs a good understanding of users

Implications for academia

• Unappealing
  – no unconstrained right to data
  – having to take responsibility for data safety

• Need training

• Needs a lot more active engagement
  – Needs a good understanding of data owners
Challenges for government
Challenges for government

- Accepting/selling key ideas:
  - default assumption is data available
  - risk is non-zero
  - academics are not risky users
- devising intelligent training programmes
- demonstrating value from research use
  - generating value from research use
Challenges for academia
Challenges for academia

• Accepting restrictions
• Actively supporting government
• Intelligently critiquing solutions
  – decisions are subjective → everyone’s ideas have validity
Example
Principles-based output statistical disclosure control (PBOSDC)
Output SDC

• statistical analyses of confidentiality data may pose confidentiality risks

• historically only tabular data considered

• recently: ‘generalised output SDC’
PBOSDC

- Define outputs as
  - fundamentally ‘safe’ (eg regression)
  - fundamentally ‘unsafe’ (eg frequency table)
- For the latter, set strict release criteria
- Add flexibility
  - criteria are ‘rules of thumb’, NOT hard rules
PBOSDC and flexibility

• data owners can impose stricter rules
• users can request weaker rules
• both outcomes are
  – occasional
  – relate to specific cases
PBOSDC and understanding

• Both parties need to understand
  – what the rules of thumb are
  – what sort of things lead to exceptions
  – what ‘occasional’ means

⇒ common training
PBOSDC: constraining academic freedom?

• shared goal: quick release of outputs
• protecting researcher from mistakes
• consistent with good statistical practice
• designed for research outputs
• checking: art not science
  ➔ help is valued
  ➔ users are the experts in statistics, not checkers
Why is PBOSDC an example?

• If there’s no communication between data owners and researchers, it doesn’t work
• if there is, build on this to create positive feedback
Summary

• new world view better for govt and ac
  – more efficient, more cost-effective, safer
• need for academia and govt to co-operate
  – realistically
  – positively
Thank you

• Questions?

felix.ritchie@uwe.ac.uk