

# User-centred, evidence-based, risk-managed access to data

Hans-Peter Hafner<sup>1</sup>, Rainer Lenz<sup>1,2</sup>,  
Felix Ritchie<sup>3</sup>, Richard Welpton<sup>4</sup>

*1 Technical University of Saarland*

*2 University of Dortmund*

*3 University of the West of England, Bristol*

*4 Valuation Office Agency*

# Audience participation time

---

- Consider these two statements
  - We should release research data unless it can't be done safely (**open by default**)
  - We should not release research data unless it can be done safely (**closed by default**)
- Which better describes your attitude?
- Which better describes your organisation's attitude?

# Audience participation time

---

- Does the evidence suggest that researchers are
  - intruders?
  - idiots?
  - lazy?
  - liable to make mistakes?

# Audience participation time

---

- “Maintaining confidentiality is our highest priority”
  - Yes or no?
- Is it important to incorporate user preferences when designing data release methods?
  - essential/important

# Audience participation time

---

- Can data access be made safe?
  - yes/no
- Should we plan for worst cases or likely ones?
  - worst/most likely
- Can we make decisions objectively?
  - objective/subjective

# Audience participation time

---

- Finally... all other things being equal, is damaging the data a good thing?
  - so why do we like doing it so much?

# Producing safe data for research use: what good is SDC?

---

- well-established, stable theories
- multiplicity of methods to address problems
- solid knowledge of characteristics of different techniques
- automatic tools to implement and evaluate

⇒it's very good

⇒it's very badly used

# SDC in practice

---

- Assumption of no-release unless proven
- Almost no analysis of user needs
  - or user behaviour
- No reference to non-statistical tools
- Quantified risk measures with no basis in fact
- No recognition of ‘uncertainty’
- No reference to evidence



# SDC in practice, in summary

---

- we mistake constraints and objectives
- we plan and assess in a vacuum
- we treat the user as the enemy
- we ignore hard decisions about real-world uncertainty in favour of easy pseudo-scientific ones

# Data access 2.0

## (a) User-centred

---

- Default-open , not default-closed
- user value is the objective
  - maintaining confidentiality is a constraint
  - acting lawfully is a constraint
  - using technology is a constraint, etc etc
- non-statistical solutions must be considered
  - data damage is the residual

# Data access 2.0

## (b) Evidence-based

---

- We know what fails in user environments
  - no intruders (keep worst cases as checks)
  - researchers are well-intentioned
  - researchers choose the path of least resistance
  - everyone makes mistakes
  
- we can adjust the world to our liking
  - good training demonstrably effective
  - researchers willing to collaborate

# Data access 2.0

## (c) Risk managed

---

- Problems can arise from many sources
  - apply non-statistical approaches to non-statistical problems
  - ‘risks’ are not directly comparable
- The world is uncertain, not risky
  - our views are necessarily subjective
  - no monopoly on truth
  - balance of subjective probabilities

# The chances of change?

---

- institutions matter
- traditions matter
- power structures matter
- incentives matter
  - Especially in government

...but it is happening

# Reasons for change?

---

- cheaper
- safer
- more efficient
- popular with users
  
- grounded in the real world
  - able to stand up to scrutiny
  - prepared for the future

# Questions?

---

- Attitude: **default-open**/default-closed
- Researchers: intruders/idiots/lazy/error-prone/**human**
- Confidentiality: is/**is not** the objective
- User preferences: **essential**/important
- Safe data access: possible/**not possible**
- Scenario planning: worst-case/**likely threats**
- Decision-making: objective/**subjective**
- Damaging data: a good thing/**bad thing**