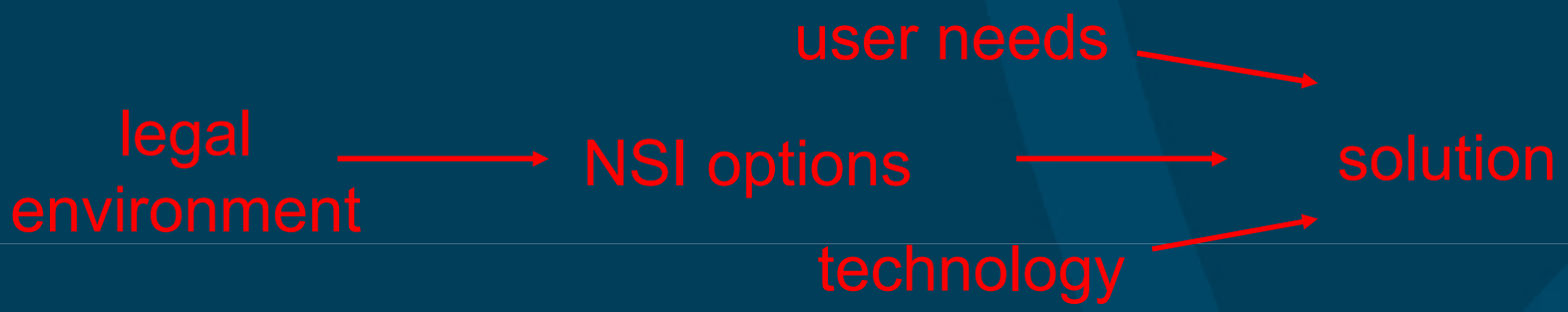
Office for National Statistics

Access to sensitive data: Thoughts from the UK experience

Felix Ritchie

The framework principle

- A model of data access



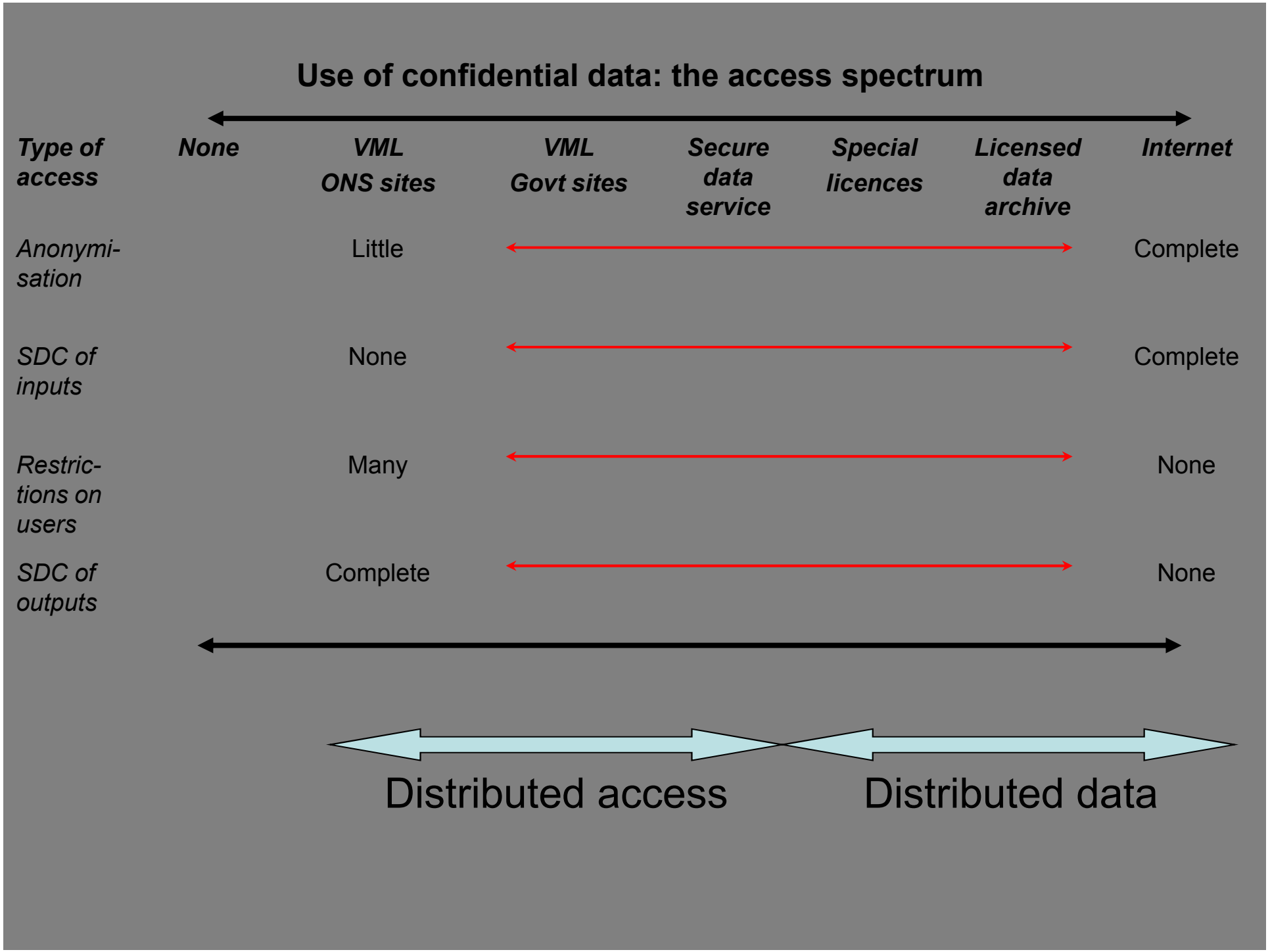
Another model…

Law as an enabler

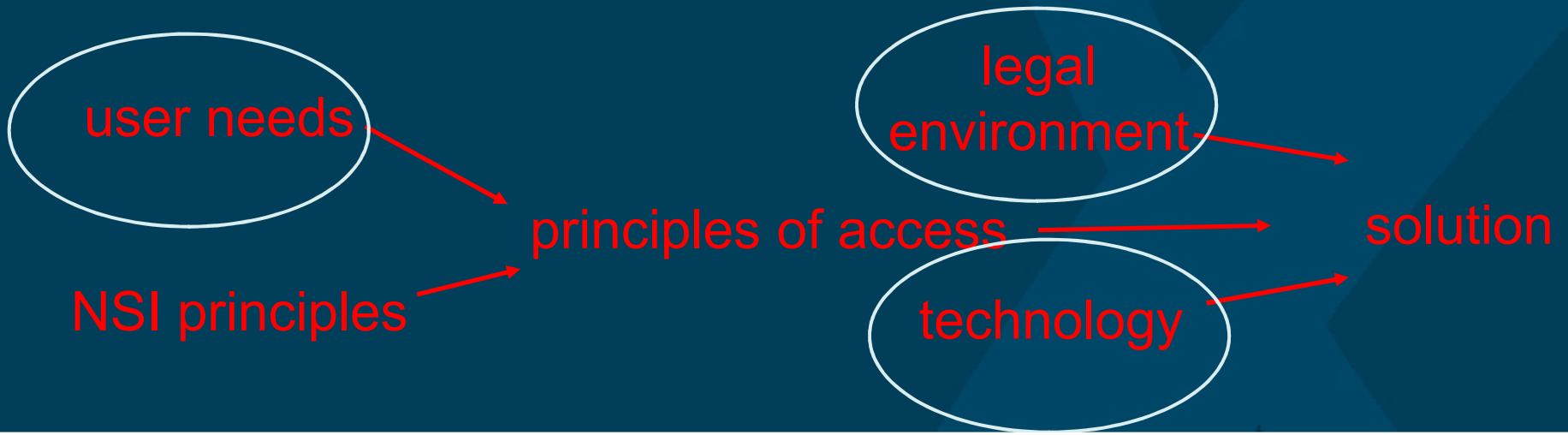
- Up to 2002: various dubious practices
  - £1 contracts
  - Researchers using own equipment
  - Poor records of microdata use
- 2002-2008
  - New recording system for applications
  - Review and rationalisation of legal gateways
  - But still many hurdles to cross
- 2008 –
  - Experience led to significant provision in law for research use

Technology as an enabler

- ‘Spectrum’ of access points balancing
  - value of data
  - ease of use
  - disclosure risk
- for a given level of confidentiality, maximise data use and convenience
- no ‘one-size-fits-all’ solution
  - no absolute prohibitions
  - trade-off is made explicit
  - users determine appropriate level of access


Use of confidential data: the access spectrum
Type of access
None
VML ONS sites
VML Govt sites
Secure data service
Special licences
Licensed data archive
Internet
Anonymi-sation
Little
Complete
SDC of inputs
None
Complete
Restric-tions on users
Many
None
SDC of outputs
Complete
None
Distributed access
Distributed data

The framework principle

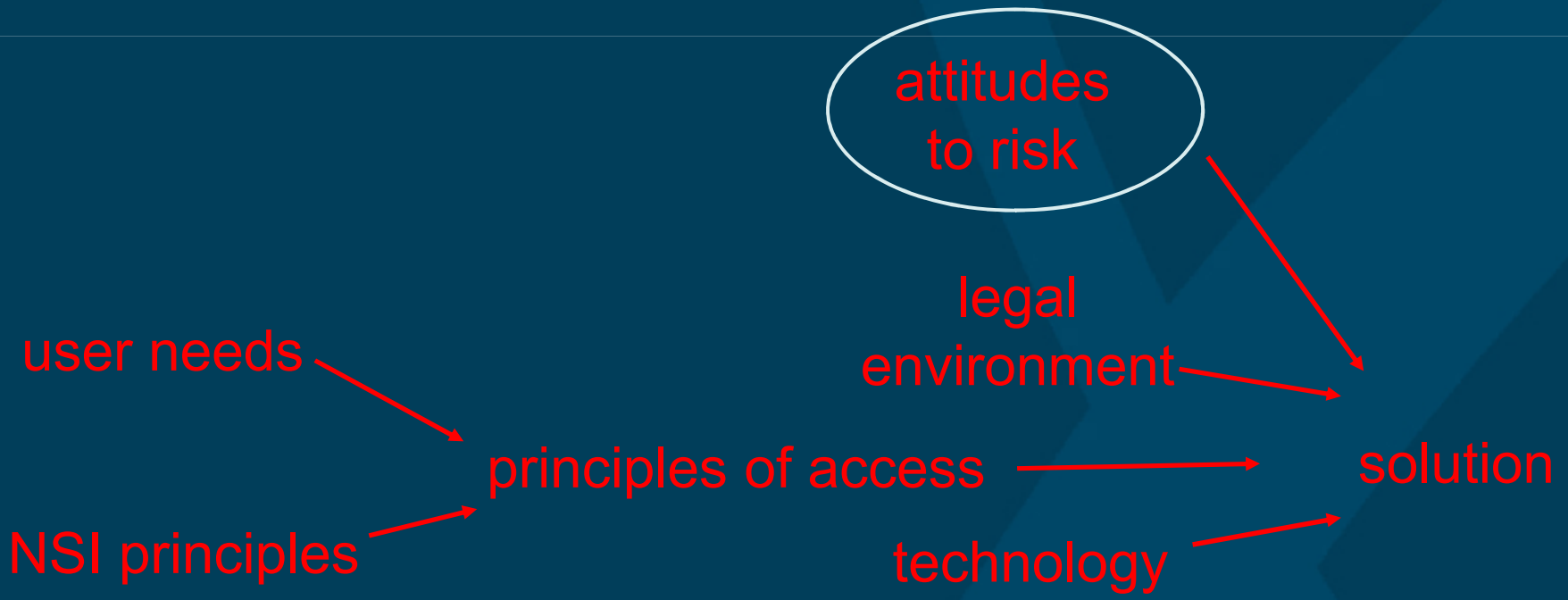


Perceptions of risk and value

- Data access spectrum => all access is doable
  - No more conceptual problems
  - Only question: does risk outweigh value?
    - But if data is already collected, is re-use pure benefit?
    - Does risk assessment include society's cost not using data?

- Where do risks/benefits arise?
  - Risk of non-use borne by public
  - Benefit of use accrues to public
  - Risk of use borne by NSI?

The framework principle



Risk management in practice

- valid statistical purpose
- trusted researchers
- anonymisation of data
- technical controls around data
- disclosure control of results

Active Researcher Management

- Researchers will engage with NSI if given a chance
- Actively engage with researchers
  - In explaining NSI goals
  - In explaining disclosure control
  - in understanding researcher needs, working practices
  - In securing cooperation minimise sensitive output
- Responsibility for data security shared between NSI and researcher (NSI always get final say)
- Certify researchers as part of the security model

Output disclosure control

- Disclosure control at the point of release
- Trained NSI staff and researchers
- Agreement on principles and purpose
- Emphasize co-operation in training

What have we learnt?

- Design based on first principles…
  - made design slow but robust
  - helped identify failings in current approaches
  - showed where new models were needed
- Technology is the easiest problem to solve
- Changes in attitude don't come easily
- Hindsight is wonderful…

Next stages

- Is the micro/macro distinction still useful?
  - If the ‘how’ is solved, shall we revisit the ‘why’?
- Is short-term vs long-term analysis a more useful distinction?

Questions?

- Felix Ritchie
- Microdata Analysis and User Support
- felix.ritchie@ons.gsi.gov.uk

- Virtual Microdata Laboratory (VML)
- Microdata Analysis and User Support
- maus@ons.gsi.gov.uk