UK Data Archive

Secure Data Service

Secure Data Service; Secure Approaches to Data Management

Melanie Wright, Reza Afkhami
Secure Data service
UKDA
University of Essex

University of Essex JISC ESRC Economic & Social Research Council

UK Data Archive

Secure Data Service

Outline

- SDS Mission
- Restricted Access vs. Restricted Data
- Data security Model
- Different access modalities
- Challenges
- SDS System Architecture
- Legal/ethical framework
- Access conditions and registration
- Disclosure control

University of Essex JISC ESRC Economic & Social Research Council

UK Data Archive

Secure Data Service

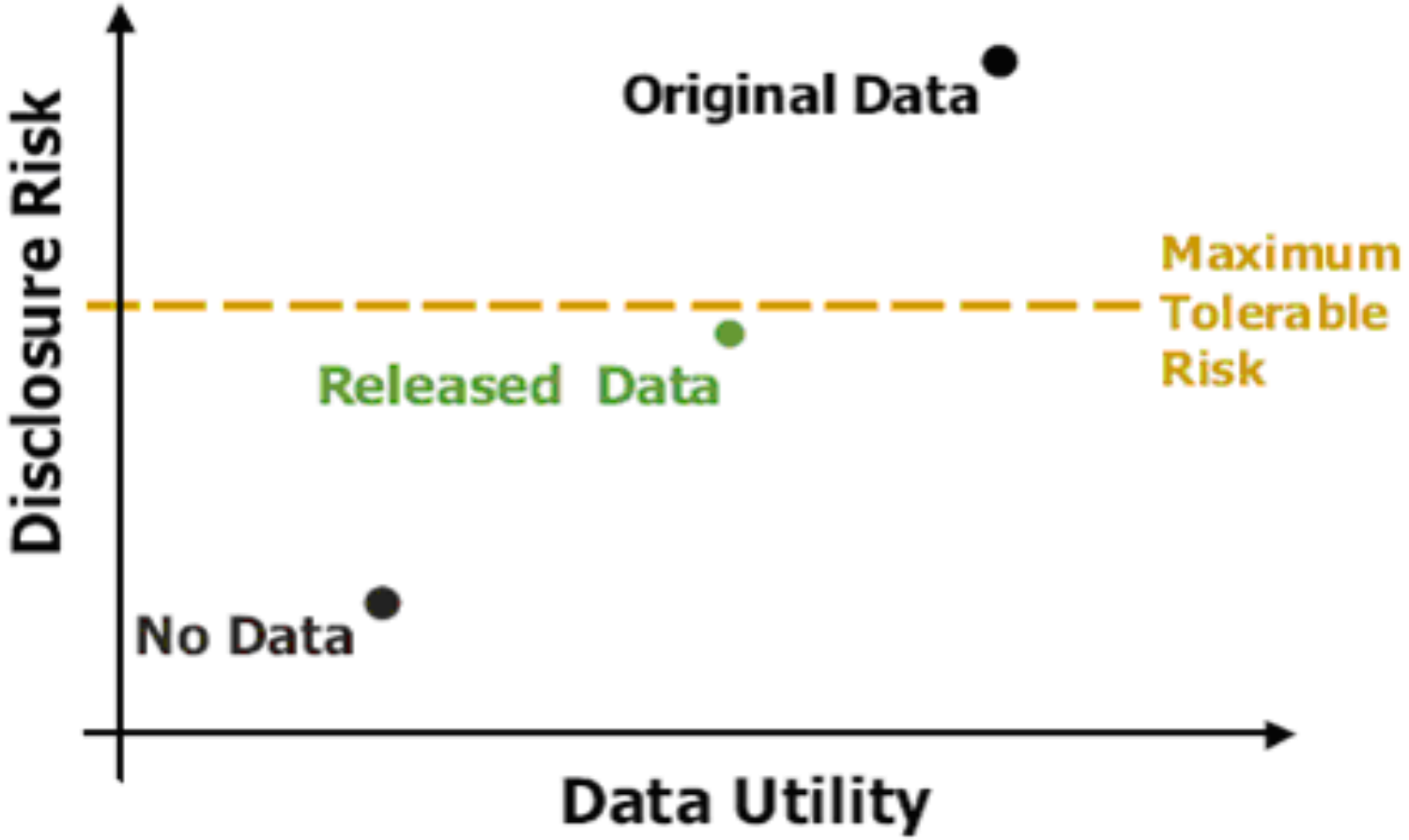
SDS Mission

- promote researcher access to sensitive micro data (maximizing data access and utility)
- protect confidentiality (minimizing disclosure risk)

UK Data Archive

Data Utility and Disclosure Risk Trade-off

Keller-McNulty & Duncan



UK Data Archive

Access Criteria

Secure Data Service

Based on

- Purpose
- Users
- Output
- Location
- Licensing

UK Data Archive

Secure Data Service

Data security Model

- valid statistical purpose → Safe project
- trusted researchers → Safe people
- anonymisation of data → Safe data
- technical controls → Safe setting
- disclosure control of results → safe output

⇒ **safe use**

UK Data Archive

Secure Data Service

- **Different Access Modalities**
  - Anonymised public use files
  - Special Licences
  - Data deposited in enclaves/research data centres
  - Remote execution
  - Synthetic data
  - secure remote access to virtual data centres over a public network

UK Data Archive

What makes SDS data access safe?

Secure Data Service

- Data design
- Law and contract
- Governance
- Environment

UK Data Archive

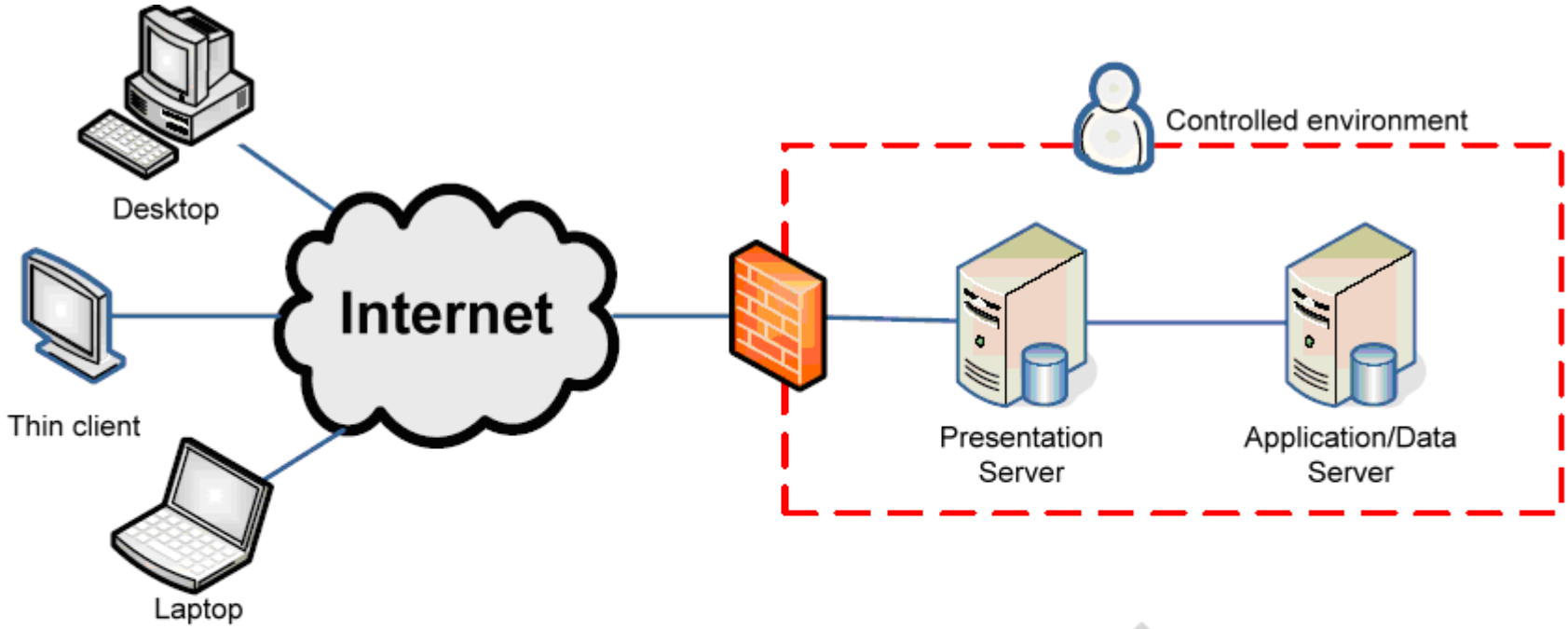
Challenges of Secure Remote Access

Secure Data Service

- Technical
- Organisational
- Educational
- Statistical
- Legal

UK Data Archive

SDS Architecture



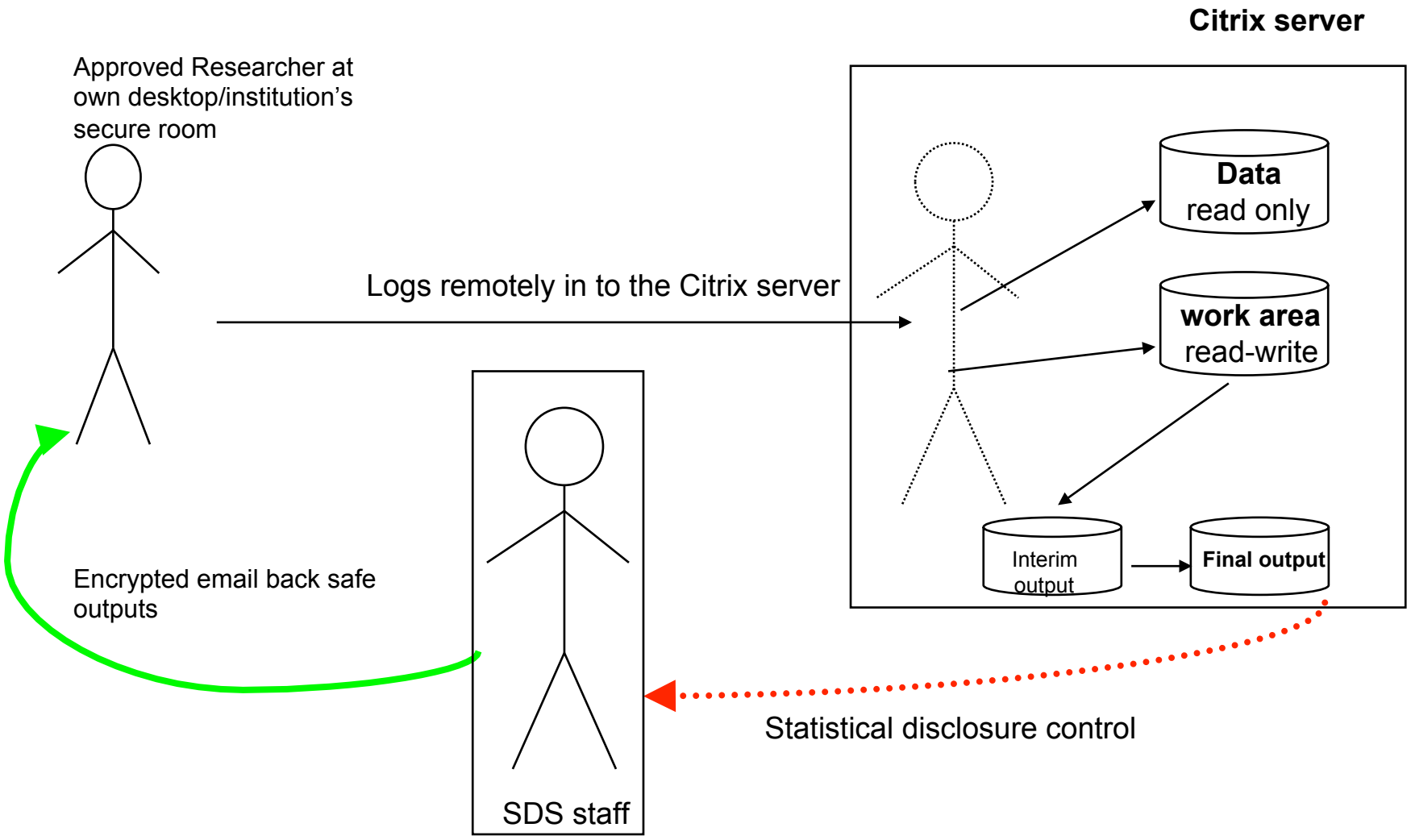
UK Data Archive

Citrix System Specifications

Secure Data Service

- Clients cannot remove data
- Absolutely no outbound traffic
- Clients cannot import data
- Data transfers are logged
- All traffic is encrypted
- Auditing
- Security patches are applied quarterly

User's access cycle



UK Data Archive

How It Works: The Back Office

Secure Data Service

- Data held securely on separate, firewalled SDS servers (farmed for expansion) in secure machine room
- System, premises and procedures compliant with ISO 27001, formal accreditation in Spring 2010. UK Data Archive is already an official Place of Deposit for The National Archives
- User access can be from desktop, remote secure room, or remote secure machine, depending upon the choices of data owners
- Connection via CITRIX, secure remote access technology used by banking and military
- SmartAuditor allows highly sophisticated user monitoring and audit trails
- Remote secure room standards set and audited by SDS and data owners
- No data allowed out; all outputs SDC vetted before release

How It Works:
The User Journey I

- User identifies SDS data they wish to access, via the UK Data Archive catalogue or specialist data support pages
- User registers with UK Data Archive, authenticate via Shibboleth and sign standard End User License
- User fill out forms to become Approved Researcher (for data covered under Statistics legislation) or ESRC Accredited Researcher (for other secure data) wherein they describe their credentials, their institutional setting, and the research they wish to conduct with the data
- Data owners grant or deny permission for access for purpose described
- User completes training session which covers both how to use the system, but also describes principles of statistical disclosure control, and covers penalties for breaches and responsibilities in law
- User signs agreement to terms and conditions of use of service and gets user id and password for remote access

UK Data Archive

How It Works:

The User Journey II

Secure Data Service

- Users access the system remotely, either from their desktop on an approved network (ie JANET) or, for some data, from a remote secure room
- CITRIX presents them with a “home away from home” familiar desktop with their data, the statistical and office tools they are familiar with (SPSS, Stata, Word, Excel, etc)
- Projects allotted common collaborative spaces for drafting papers, sharing interim outputs (all project members must be approved for same data sources)
- Users allowed to bring in data from standard Data Archive collection
- Ability to use SDS as secure space for Administrative Data linkage
- Users encouraged to leave everything on the server until final outputs for publication required, which are then vetted by SDS staff (and data owners, if they wish)

SDS research access & legal framework

- **Legal Framework**
  - Hannigan requirements
  - Statistics Act
  - Public good
  - Private harm

- **Laws under which data was collected**
  - Statistics of Trade Act 1947
  - Census Act 1921
  - health laws/administrative requirements
- **Laws concerning management of data**
  - Data Protection Acts
  - Statistics and Registration Service Act 2007 (SRSA)
- **Duty of confidentiality**
- **Data suppliers' rules**
  - including medical ethics, survey pledges

UK Data Archive

Access conditions

Secure Data Service

- All SDS users must be Approved/Accredited researchers
  - person-specific ("fit and proper")
  - project-specific
  - time-specific
- All users must be trained in
  - Legal responsibilities
  - Access procedures
  - disclosure control
  - No training – no access

SDS Security Philosophy

Big Carrots and Big Sticks

Carrots:

- Providing remote access is a positive security measure because it minimises the likelihood of data removal for convenience sake
- Providing familiar tools in a familiar environment reduces the likelihood of breaches
- Allowing both secure and EUL data furthers convenience
- Training includes impressing upon users the unprecedented access SDS provides, contrasting with other countries far more limited access regimes.

SDS Security Philosophy

Big Carrots and Big Sticks

Sticks:

- Penalties policy with real teeth
- Penalties dependent upon severity of offence, but range from suspending access to the system, to denying access to all data from the Data Archive, to denying access to any ESRC-funded research resource, to denying future ESRC research funding, to fines and custodial sentences (if in breach of statistics legislation)
- Penalties can be imposed both on individuals and on their entire institution

Confidentiality Breaches

- Major disciplinary offences
  - attempting to remove data
  - attempting to identify individuals, households, or firms
  - using data which they are not allowed to
  - using data for anything other than the proposed project
  - Strict liability: all major offences are treated as deliberate actions

Disclosure Control

- Disclosure control methods for
  - Microdata
  - Tabular data
- **Assessment of outputs for Statistical disclosure**
  - Safe: No risk / very low risk of disclosure
  - Unsure outputs: Low or medium risk of disclosure
  - Unsafe: High risk of disclosure – output will be blocked in its current form and won't be released.

UK Data Archive

Disclosure Control

Secure Data Service

- Three types of disclosure
  - Identity disclosure
  - Attribute disclosure
  - Inferential disclosure
- Disclosure Control Techniques
  - Reducing information
  - Perturbing information

Automated disclosure control

- **“ARGUS is the name of a guardian monster in the Greek mythology with hundred eyes to watch”**
- **τ ARGUS**
  - For tabular data
  - Tables can be redesigned (rows and columns combined)
  - Sensitive cells can be suppressed (primary suppression)
  - Additional safe cells can be suppressed (secondary suppression)
- **μ ARGUS**
  - For microdata

Examples of Disclosive Data

- More detailed variables from existing ESRC-funded data resources
- More previously unavailable detailed variables from government social surveys
- Detailed Census microdata files (CAMS)
- Business data which has commercial sensitivity
- Administrative data; the SDS may be able to provide a secure environment for data linkage activities
- longitudinal data, medical data, etc.

SDS Data

Initially:

- Fully geographic grid-referenced version of British Household Panel Study
- PLASC linked education data from the Millenium Cohort Study
- Highly detailed versions of a variety of ONS social surveys, currently held in VML
- Business microdata currently held in ONS VML

Future:

- More data from ESRC-funded longitudinal studies, including verbatim text responses to qualitative questions, linked medical data, linked administrative records, data from the new Understanding Society
- Census CAMS / other sensitive Census products
- Other administrative data for linkage (eg patient records, benefits data etc)

Contact information

- Director; Melanie Wright
- Melanie@essex.ac.uk

- Support service; Reza Afkhami
- Rafkhami@essex.ac.uk

- SDS helpdesk
- securedata@ukda.ac.uk
- http://securedata.ukda.ac.uk/