UK Data Archive

Secure Data Service

ESRC Secure Data Service

A new vision for secure data access

Melanie Wright, Director
Secure Data Service
UK Data Archive
University of Essex

University of Essex

ESRC Economic & Social Research Council

UK Data Archive

Secure Data Service

SDS Mission

To promote excellence in research by enabling safe and secure remote access by bona fide researchers to data heretofore deemed too sensitive, detailed, confidential or potentially disclosive to be made available under standard licensing and dissemination arrangements.

University of Essex

ESRC Economic & Social Research Council

Data Access Before SDS

For more than 40 years, UK Data Archive has operated a number of different data access regimes

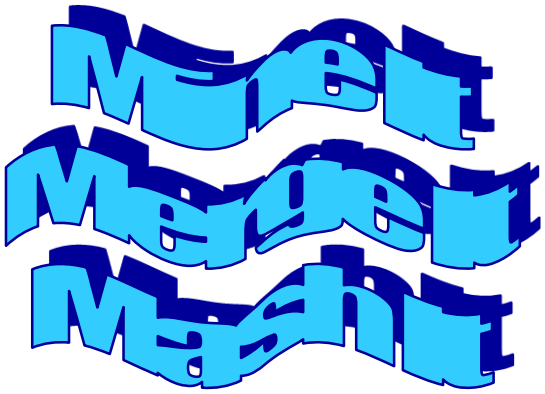
- Open access / anonymous download (very very few datasets) – Public Use data
- End User License (EUL) (vast majority of it 6000+ dataset collection) – Scientific Use data
- Special Conditions (small but significant proportion) – Scientific Use data, sensitive
- Special License / Approved Researcher (a few dozen datasets) – Restricted data

UK Data Archive

The Data Feeding Frenzy

Secure Data Service

Cloud Computing



Let's Google Map It!

Publicly funded should mean publicly available

UK Data Archive

Data Paranoia

Secure Data Service

Millions of benefits records lost

Laptop with official secrets left on train

Data CDs sent in post LOST

Identity Theft on the Rise

Census "resisters" organise

Opportunities and Threats

- New UK Statistics and Registration Act 2007 allows for unprecedented access to official data about individuals to “approved researchers”
- Also provides unprecedented penalties for data confidentiality breaches – hefty fines and custodial sentences

UK Data Archive

ESRC response

Secure Data Service

- A two-year pilot of a Secure Data Service
- Offer remote secure access to sensitive and potentially disclosive data
- Focus initially on the ESRC-funded resources (e.g. longitudinal studies)
- Call for proposals resulted in pilot grant to UK Data Archive beginning October 2008

Data Security Model
Find the Weakest Link

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

⇒ **safe use**

-- After Ritchie, 2006

Security

During the pilot SDS has visited and spoken with a variety of secure data enclaves worldwide about when and why breaches occur

Fundamentally two types:

- Accidental disclosure through ignorance of statistical disclosure control principles and methods for outputs

- Users wanted to take data home with them for convenience sake (to avoid repeat onsite visits or to work with home tools/data)

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, contrasted 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

SDS Security Philosophy

Fundamentally it is about trust

- The most important security measure is to get the researchers to buy into security as their own project
- Training is absolutely central : both how to do it right, why to do it right, and what the penalties are for doing it wrong
- Backed up by appropriate legal licensing framework and agreements
- Backed up by technology to first prevent and second identify misuse and provide reliable audit trails
- Backed up by commensurate penalties

How It Works: The Back Office

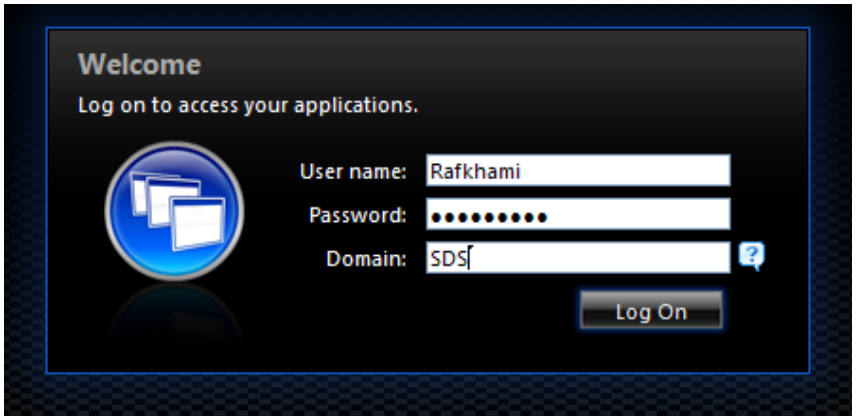
- Data held securely on separate, firewalled SDS servers (farmed for expansion) in secure machine room
- System, premises and procedures compliant with ISO 27001, formal accreditation in Spring 2010. UK Data Archive is already an official Place of Deposit for The National Archives
- All relevant information about UK Data Archive's information security policy will be published to inculcate trust.
- 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

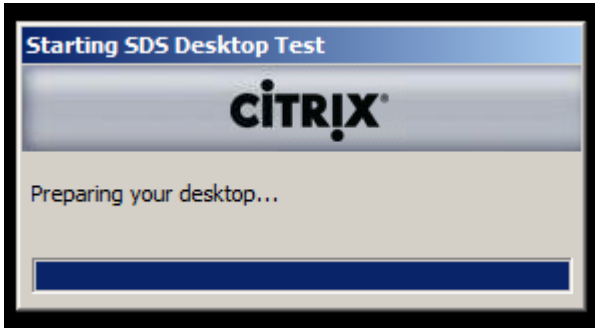
User Journey Part I:
Gaining Access

- 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 (or the authority they have designated) grant or deny permission for access for purpose described
- User completes training session which covers both how to use the system, but also 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 userid and password for remote access

User Journey Part II:
Using the System

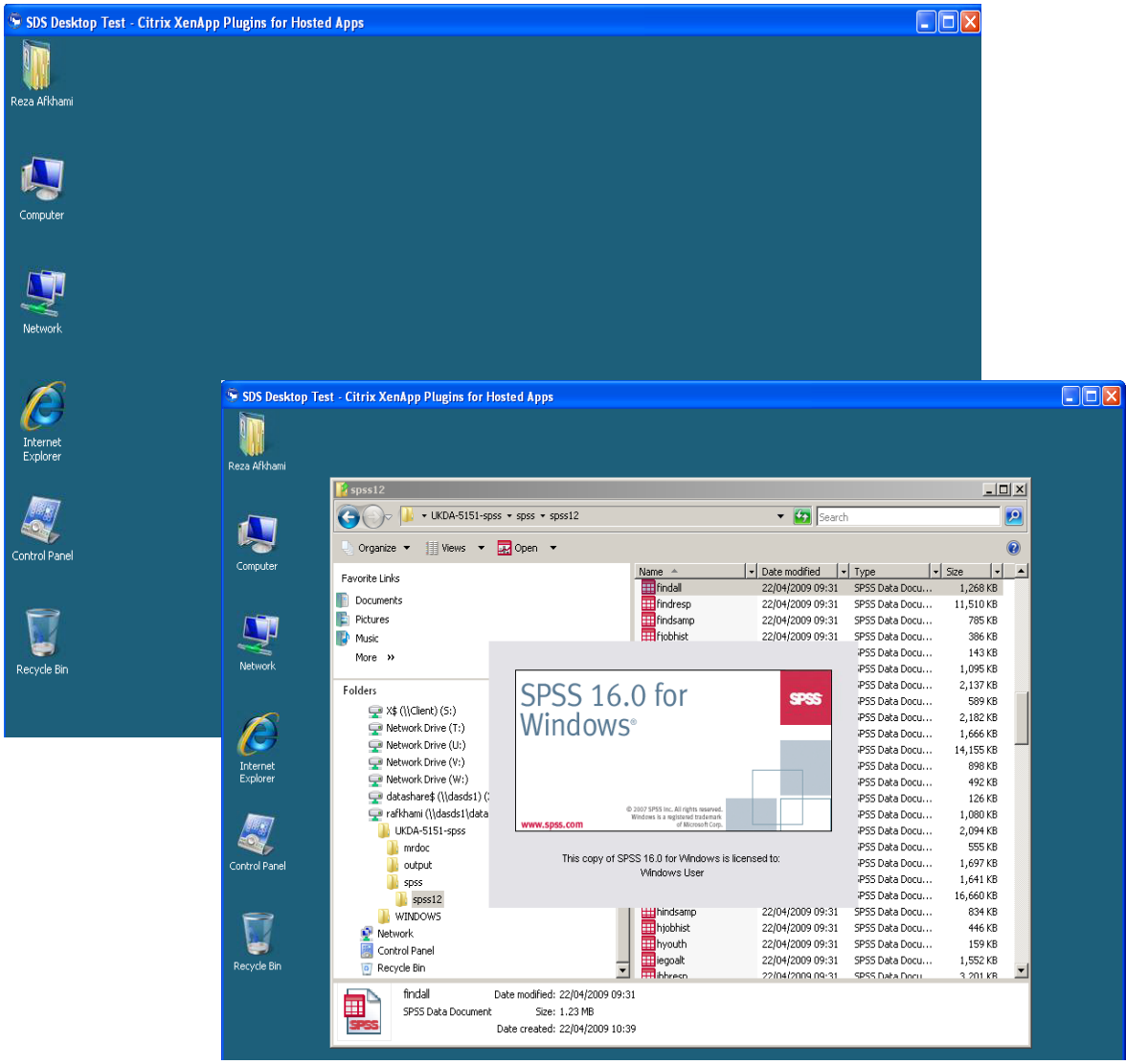
Users access the system remotely, either from their desktop on an approved network (ie JANET) or, for some data, from a remote secure machine and/or secure room





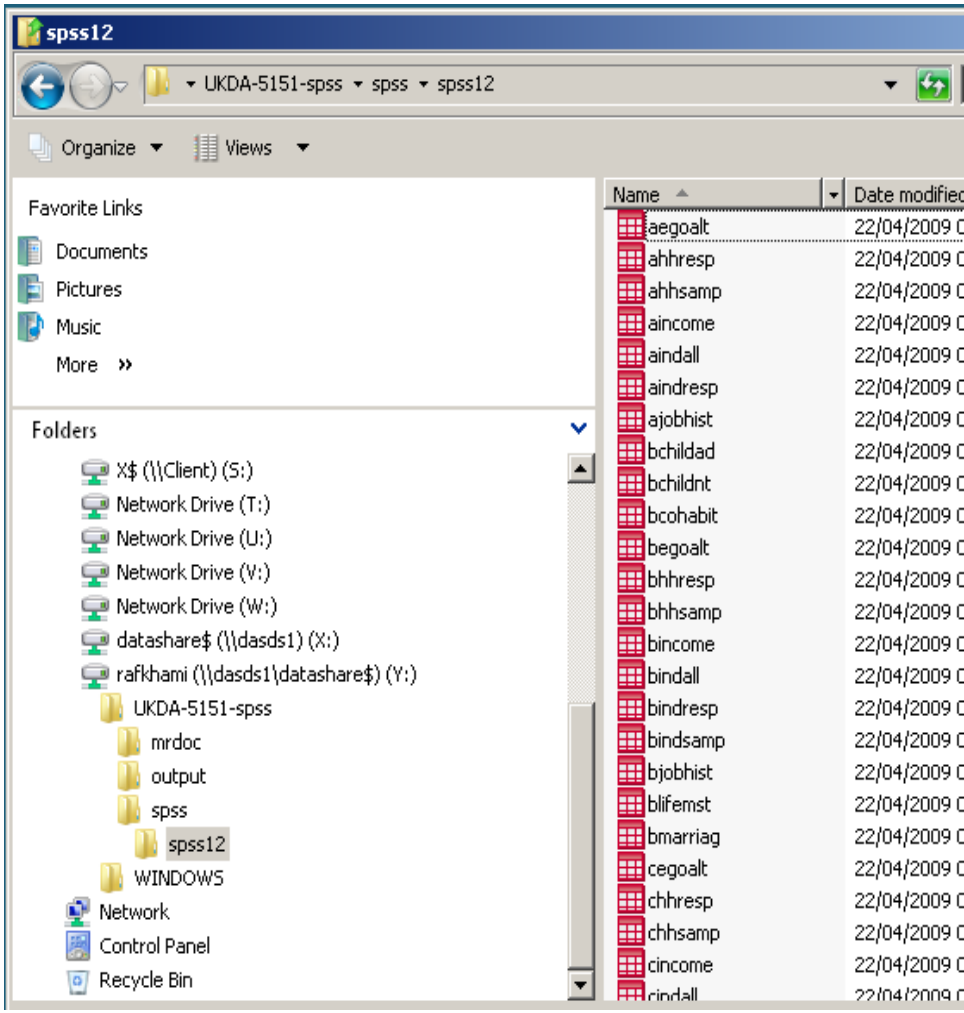
User Journey Part II:
Using the System

Citrix™ presents user 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)



User Journey Part II:
Using the System

- 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



User Journey Part II:
Using the System

- 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)
- System allows remote locking for inactivity, remote shut down for suspicious keystrokes

After the Pilot

- Close working with ONS has led to official approval of the service, agreement in principle to lodge gov't data in the service
- New research council funding allowed for a proposal for service expansion
- Expansion funding approved in principle last month: 3 years, appx. £1.6million
- Data acquisition expanded to include data currently in the ONS's Virtual Microdata Laboratory (including business microdata)

SDS Data

Pilot

- Fully geographic grid-referenced version of British Household Panel Study
- Working to pilot remote room access with a non-disclosive teaching dataset from the Scottish Longitudinal Study
- In negotiation to enable access to PLASC-linked education data from the Millennium Cohort Study, currently only available onsite at CLS

SDS Data

Full service (From Autumn 2010)

- Highly detailed versions of ONS social surveys, currently held in VML
- Business microdata currently held in ONS VML
- More data from ESRC-funded longitudinal studies, including verbatim text responses to qualitative questions
- SDS can provide a safe setting for analysing linked data for users whose home institutions can't provide the requisite IT security infrastructure

SDS Data

Future:

- Data from the new *Understanding Society*
- Census CAMS/other sensitive Census data
- Other survey data linked with administrative data (eg patient records, benefits data etc)
- More extensive qualitative data resources
- Network of European Secure Data Centres – SDS as gateway to gaining access to comparable potentially disclosive or sensitive data from other European countries
- Verbatim qualitative data from longitudinal studies, etc.
- Possible data linkage services as an Honest Broker / Trusted Third Party

Data Partnership

- Over-riding philosophy is that SDS will provide access to datasets either never before available, or available only onsite
- Not intended for delivery of any data currently available under less restrictive access arrangements
- Does not replace Special Conditions or Special License, but allows access to even more sensitive/detailed data
- Data owners (or their designated decision-makers – e.g. MRP) control access decisions on a case-by-case basis
- Data owners choose access modality (standard, secure machine, or secure room)
- If they wish, data owners can be involved in initial user training and output vetting

Timeline

- Pilot Launch December 2009
- Full launch in October 2010, with ONS social surveys, more ESRC-funded data
- Business microdata currently in VML to follow by January 2011
- Service at full speed by Spring 2011
- Service extension contract ends 30 September 2012; likely jointly refunded with the ESDS from October 2012 for 5-10 years dependent on outcome of Review.

UK Data Archive

Secure Data Service

Contact information

SDS helpdesk

securedata@ukda.ac.uk

Public website:

http://securedata.ukda.ac.uk

University of Essex

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL