**Working Paper**
ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION**

**STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)**

**Joint UNECE/Eurostat work session on statistical data confidentiality**
(Ottawa, Canada, 28-30 October 2013)

Topic (iii): Modes of access to microdata

# Statistical Disclosure Control Practice in the Secure Access of the UK Data Service

Prepared by Reza Afkhamai, UK Data Archive/UK Data Service, University of Essex, United Kingdom

Abstract for Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality-2013, Ottawa

# Statistical Disclosure Control Practice in the Secure Access of the UK Data Service

Reza Afkhamai

*UK Data Archive/UK Data Service*

Secure access of the UK Data Service has been operating for nearly two years. Output checking of the outputs created by the researchers is one of the prime activities of the service. This short paper has twofold. First to introduce and share some practical issues/ experiences encountered with every day output checking requests. I introduce the current procedure in the SDC. Second, in doing so to suggest how practical/contextual and data specific issues may affect decisions on whether to release an output or not. Examples of the accepted/rejected outputs would be demonstrated. This in turn may raise more questions than answers and again may recall the complicated nature of the statistical disclosure control/limit. The role of researchers' training to minimize the rejection rate and consequently increase researchers' satisfaction would also be discussed.

## Introduction to the UK Data Service-Secure Access

The UK Data Service provides secure access to data that are too detailed, sensitive or confidential to be made available under the standard End User Licence or Special Licence.

Building on the success of other secure data enclaves worldwide, and employing security technologies used by the military and banking sectors, the SDS will allow trained researchers to remotely access data held securely on central SDS servers at the UK Data Archive. The aim of the service is to provide approved academics unprecedented access to valuable data for research from their home institutions with all of the necessary safeguards to ensure that data are held, accessed and handled securely.

Data accessed in this way cannot be downloaded. Once researchers are specially trained, they analyse the data remotely from their institutional desktop or in our Safe Centre. We provide access to statistical and office software to make remote analysis and collaboration secure and convenient.

Our security philosophy is based upon training and trust, leading-edge technology, licensing and legal frameworks, and strict security policies and penalties.

We invite data owners and academic researchers to join our community to enable responsible use of detailed data for high-quality research.

We provide researchers with access to sensitive and confidential business, social and economic microdata.

Our growing collection of approximately 45 datasets which are derived from survey, administrative and transaction sources include:

- Productivity data from the Annual Respondents Database
- Innovation data from the UK Innovation Survey
- Geospatial data from the Labour Force Survey, Understanding Society
- Sensitive data about childhood development

Providing secure access to highly sensitive, confidential, and identifiable data, is based on a five-point security philosophy (which is shared with other services around the world):

- state-of-the-art secure data technology and procedures
- training and convenience for approved researchers
- standards backed by a professional code of practice
- meaningful penalties for breaches
- establishing honest and collaborative relationships between researchers, data providers and the UK Data Service

The UK Data Service fosters a community of trusted researchers and data providers, working together to maximise the full potential of data, whilst protecting the privacy of individuals and organisations, through our secure access routes.

Our Trusted Researchers (our members) come from universities across the UK, and are at the forefront of research in fields such as economics, sociology, healthcare, and education.

Users of the SDS will be required to be either “ONS Approved Researchers” or “ESRC Accredited Researchers.” The first of these is defined by the Statistics and Registration Services Act 2007 as an individual to whom the Board has granted access, for the purposes of statistical research, to personal information held by it.

UK Data Service staff have collective expertise in data management, handling, licensing and security; data analysis and Statistical Disclosure Control; and in helping users to accomplish their research.

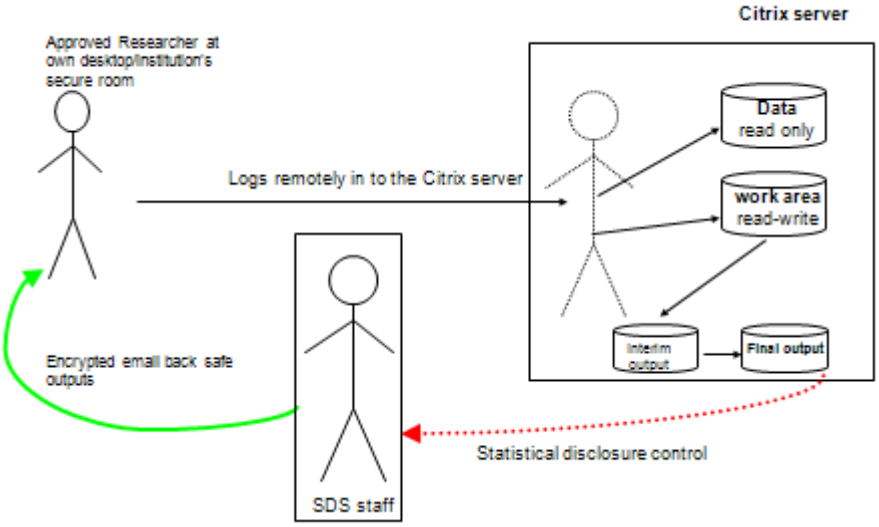
Secure access is available through researcher’s own institution PC or at our Safe Centre at University of Essex, depending on how restrictive (and sensitive) the data are.

Members access the data via a web-based interface that uses secure encrypted Citrix Virtual Private Network technology. The data are never downloaded.

We provide our members with a familiar Windows environment and the statistical tools that they require to achieve their analyses, including Stata, SPSS, R, Gauss, Matlab and ArcGIS.

We return statistical results (rather than data) to researchers, subject to a statistical disclosure control process, to ensure that no individual or organisation can be identified from the results.

**Figure1: Remote Secure Access Operational Cycle**




## Output checking procedures

This section sets out procedures for processing requests by members (users) of the Secure Data Service (SDS) for releasing their statistical outputs. This includes managing the initial request, undertaking the necessary Statistical Disclosure Control (SDC) of the statistical outputs, and returning the outputs to the user as well as keeping records of the output request.

Careful user vetting and the most secure analysis environment in the world cannot on its own ensure that data are not disclosed. The missing piece of the data security puzzle is not what goes into the secure data system, but what comes out of it. For the service to be able to meet the security guarantees placed upon it by the data guardians, it must offer some form of output screening. If an output has been determined to be disclosive, it will be up to the user to determine the best way to render it safe.

SDS adheres to European-wide ESSNet standards on good practice in statistical disclosure control of tabular and other statistical analytical outputs. SDC of outputs are undertaken by the Secure Data Access Manager and Senior Access Officers. These are responsible for undertaking the SDC, and returning the statistical outputs to researchers only if after assessment, the statistical outputs are deemed non-disclosive.

All SDS outputs must be checked by two Senior Access Officers (or one Senior Access Officer and the Secure Data Access Manager or VML Manager) before it can be released to the researcher. This is due to the nature of sometimes lengthy outputs, which can be difficult to check.

Double-checking is a safety mechanism to ensure that a disclosive or potentially disclosive statistic is not overlooked by accident. Sign-off by both members of staff is necessary however this is not required of requests for syntax files.

The SDS distinguishes between two kinds of outputs: Intermediate Outputs and Final Outputs. Intermediate Outputs could include results within a spreadsheet, Stata or SPSS log file etc.

Final outputs are statistical outputs that have been formally written up as, for example, a Report, Working Paper, Presentation etc. The researcher has clearly selected the results they wish to present to the outside world from their intermediate results, and have written up these results.

It is our policy that only final outputs will be released from the secure server. However, exceptions will be made to for researchers who are working with colleagues based abroad.

PhD students often have to request outputs relating to a particular chapter of their Thesis, or may also be required to submit a Data Appendix by their examiners. PhD work may differ from the examples of Final Outputs defined above. For example, they might contain many tabulations of data. These may not be the primary interest of the researcher, but nevertheless are still required by the PhD examiners. Nevertheless, the work should still be considered 'final', in that the researcher no longer needs to change the results, once they have been released from the secure environment.

In order to provide flexibility to researchers who are using the Secure system, and work with project colleagues based abroad (and who cannot therefore access the SDS), it is possible to release intermediate outputs, so that results, not intended for publication, can be discussed.

Where all colleagues are based in the UK, they are able to access the SDS from their institution, and it is therefore not necessary to provide these researchers with intermediate outputs, as they can see all results produced within the SDS using their SDS account.

The following conditions should be met for release of intermediate outputs:

- The researcher based outside of the UK should be an ONS Approved Researcher or ESRC Accredited Researcher for the project that is being undertaken by their colleagues in the UK
- The intermediate results are only intended for discussion purposes with the colleague(s) based overseas
- Releasing these intermediate results should not be considered a means of releasing results which can then be written up and presented outside of the SDS: it is expected that the researchers based in the UK will use the feedback they receive from their overseas project colleague to continue producing a final output in the SDS
- The researcher accessing the data in the SDS and producing the results should apply consideration as to which results they need to have released, to show to their colleague. For example, results they know that they do not wish to discuss, should be omitted
- The results should be neatly presented, in Word or Powerpoint documents. Excel

spreadsheets, Stata log files, and SPSS output files, for example, are not acceptable for release.

- Results should be explained – it should be easy for output assessors undertaking the disclosure control to ascertain the meaning of the results in order to apply statistical disclosure control techniques

Researchers may request syntax files from the SDS in order to write up their methodologies outside the secure confines of the SDS.

However, they should be processed as with other requests for output releases. They should be opened and examined to ensure that the files do not contain data or results, or any identifiers (e.g. person IDs, IDBR reference numbers). An exception is that they may be checked and released by one output assessor.

Researchers will contact the SDS helpdesk with a request for an output to be released. This requested will be saved in a JIRA issue (helpdesk system). The output(s), together with the completed Output Request Form, should have been saved by the researcher in the SDC folder in the relevant project area of their SDS account.

## Secure Access Training

It is well known in data security that people are the weakest links. The training and educating of people prevents them from getting a criminal record. The education couples with stricter legislative protections can offer another potentially efficient means of improving confidentiality—efficient because the probability of disclosure can be decreased without imposing costs on rule-abiding researchers.

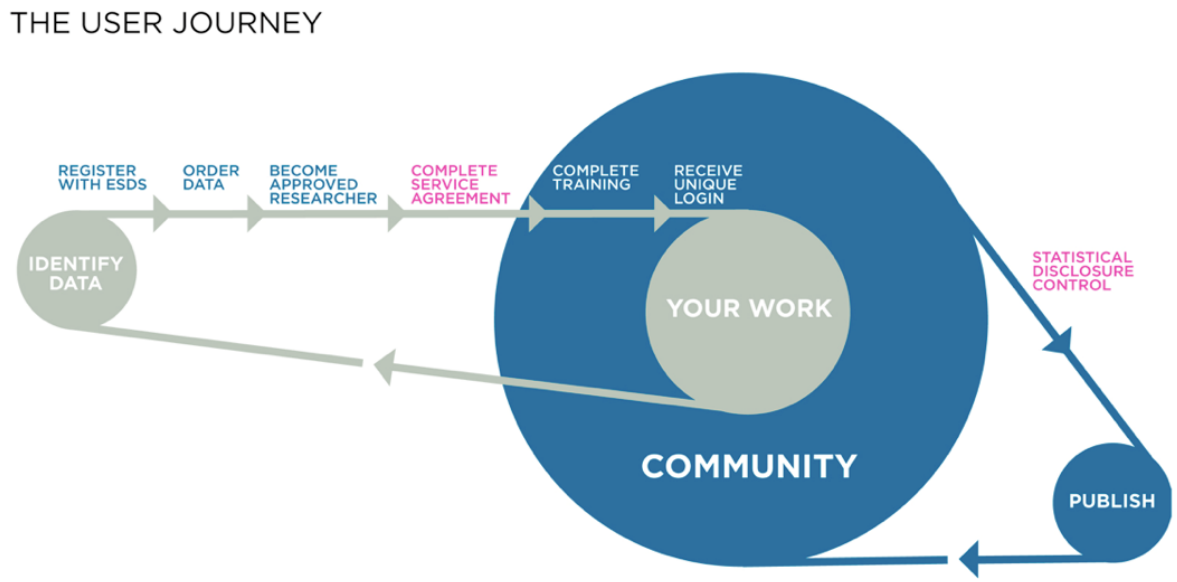
Before becoming an active user of the SDS, users will have to attend a mandatory training session which will focus first on the user's legal and ethical responsibilities within their SDS user license agreement, the mechanics of how to use the SDS, what they can and cannot do in a remote access setting, and the potential of the collaboratory spaces. The second part will focus on principles of the statistical disclosure control, assessment of outputs, and analysis aspects of the particular datasets in the SDS.

Access to the SDS will only be granted after users have attended an SDS training session. There will be vetting of data analysis outputs for disclosure issues by SDS staff, to ensure that nothing escapes the secure data setting which could compromise the data security (safe output). One of the purposes of the training is to give researchers the ability to recognise confidential data in order to distinguish it from statistical results that are safe to remove from the SDS — in effect, the training removes the 'reasonable belief' defence for a disclosure. We believe that penalties will only be an effective deterrent if they are known about, and it should also be clear that we are much more concerned about prevention than punishment.

The training course is designed to provide researchers with an understanding of how to use the Service appropriately. This includes: Understanding their legal obligations of accessing data via the SDS, under the Statistics and Registration Services Act, Data Protection Act, and

other relevant legislation (depending on the source of the data). An introduction to Statistical Disclosure Control, which trains researchers so that they can ensure their statistical outputs, cannot be used to identify observations from the data or disclose any personal information/information relating to a single observation. One module is about how to use the SDS (including software available, key safety tips to remember in terms of using the data appropriately, logging in and logging off).

**Figure2: Different stages in user journey to access to the secure data**



## Summary

The SDS is a secure environment funded by ESRC to provide researcher access to disclosive micro data either from their offices, safe rooms in their institutions or on site at the UKDA. It has two goals: to promote researcher access to sensitive micro data and to protect confidentiality. SDS operation is legally framed by the 2007 Statistics Act, which makes access to confidential data for statistical purposes possible. Researcher access to microdata serves the public good both by leveraging existing public investments in data collection, and by ensuring high quality science through the replication of scientific analysis.

The SDS provides Approved/Accredited researchers with remote access to microdata using the most secure methods to protect confidentiality. This is achieved by implementing technological security (Citrix gateway), applying statistical protections, enforcing legal requirements, and training researchers. The SDS also ensures that valuable data are preserved for the long term by documenting the data using DDI compliant metadata standards. In addition, the SDS aims to engage the research community in using its shared data space to share information which enables collaboration among geographically dispersed researchers.

**References**

Afkhami, R, Wright, M. and Ahmet, Mus (2010) Secure Data Service; an improved access to disclosive data, *IASSIST Quarterly, 15-18*

Brandt, M., (Destatis), Franconi, L., (ISTAT), Guerke, C., (Destatis), Hundepool, A., (CBS), Mol, J., (CBS), Ritchie, F., (ONS), Welpton, R., (ONS), (2010), "ESSNET SDC – Guidelines for the checking of outputs based on microdata research". Available upon request.

Mulcahy, Timothy M, & John Nieszel.: Towards a Secure Data Service at the UK Data Archive. SDS Consultants' Report. (2008)

Ritchie, F .: *Disclosure Control of Analytical Outputs*. Mimeo: Office for National Statistics, UK. (2006)

Ritchie, F. (2008), "Default Procedures for Statistical Disclosure Detection and Control". Available upon request.

Ritchie, F. (2013), "Output-based disclosure control for regressions" forthcoming. Available upon request.

Ritchie, F (2007) Disclosure control for regression outputs, Mimeo : Office for National Statistics, UK.

Ritchie, F. (2007)*Statistical Disclosure Control in a Research Environment*. Mimeo: Office for National Statistics, UK.

SDS Training Course – available upon request, please ask the SDS team.

Welpton, Richard (2013) SDS Output checking Procedures, UK DA internal controlled document.

Wright, Melanie. (2008) Case for Support – Secure Data Service.