

第22回「統計データの二次的利用促進に関する研究会」

# 政府統計における個票データの利用の あり方と分析結果のチェックについて

2016年5月31日  
中央大学経済学部  
伊藤 伸介

# 1. はじめに

・わが国では、政府統計(公的統計)の調査票情報(個票データ)に関するリモートアクセス型のオンサイト利用が現在議論されているが、個票データの利用後に利用者が「安全な分析結果」を得る上で、集計表や回帰分析の結果をどのようにチェックするかについての具体的なガイドラインが求められている。

・本報告では、イギリスの事例を中心に、政府統計における個票データの利用のあり方と分析結果のチェックに関する考え方について述べることにしたい。

## 2. 政府統計における個票データの利用のあり方—Five Safes Model—

政府統計マイクロデータの提供において、秘匿性を確保するための枠組みとして、侵入者モデル(Intruder Model)と人間モデル(Human Model)がある。

①侵入者モデル: データに含まれる秘匿情報を故意に漏えいしようとする侵入者(intruder)を想定したモデル

⇒最悪なシナリオを考え、露見リスクを回避するために、様々な匿名化技法が適用される。

→侵入者モデルを適用することによって作成された匿名化マイクロデータは、利用者にとっては想定以上に有用性が低くなる可能性がある。

②人間モデル: (人間(human)は)秘匿情報を漏えいする意図はなくても、誤って漏えいさせる可能性があることを想定したモデル

⇒個票データの提供システムおよび個票データを利用する研究者の訓練の必要性

⇒**Five Safes Model**の適用

\*最近、オーストラリア統計局においても、データ提供の考え方として、侵入者モデルから人間モデルへの転換が行われている。

# Five Safes modelについて(Ritchie(2008),Desai *et al.*(2016))

**Five Safes Model:**2003年に考案されたモデルで、ヨーロッパだけでなく、オーストラリアでも現在適用されているモデル

**\* 個票データ(confidential data)のアクセスは、以下の5つの基準に基づいて行われる。**

## ①安全なプロジェクト(safe projects)

→妥当な統計目的のために個票データのアクセスが行われること。

## ②安全な利用者(safe people)

→研究者は個票データを適切に利用し、利用手続きに従うことについて、信頼されていること。

## ③安全なデータ(safe data)

→データそれ自体が、個人情報が見えないデータであること。

## ④安全な施設(safe settings)

→個票データのアクセスに関する技術的な管理措置によって、容認されていないデータの移動が回避可能なこと。

## ⑤安全な分析結果(safe outputs)

→統計分析の結果に個人情報が見えるような結果が含まれないこと。



**安全な利用(safe use)**

## 安全なプロジェクト(safe projects)

- 個票データの利用については、特定の**研究目的**に対して、特定の**期間**に特定の**研究グループ**で行う必要がある。

## 安全な利用者(safe people)

個票データのアクセスにおいて、利用者に対する「信頼」を重要視している。

→イギリスの場合、国家統計局(The Office for National Statistics=ONS)が実施している統計調査の個票データを利用するためには、「承認された研究者(approved researcher)」の資格を取得する必要がある。

⇒「承認された研究者」については、イギリスの「統計登録サービス法(The Statistics and Registration Service Act 2007)」の第39条「個人情報(personal information)の秘密保護」の条文に明記(伊藤(2012))

・個票データにアクセスするためには、個票データの利用に関する法的根拠、利用手続き、分析結果の秘匿処理について特別のトレーニングコースを受けることが求められる。

⇒ 個人情報情報を漏えいした場合の罰則規定に関する説明

# 安全な分析結果(safe outputs)

・分析結果に関する最終成果物(final output)については、イギリス国家統計局(ONS)のオンサイト施設であるVirtual Microdata Laboratory(VML)やエセックス大学のリモートアクセス施設であるSecure Labの担当者がチェックを行った上で公表することが可能。

①一次秘匿(primary disclosure)

閾値ルール(threshold rule)

占有ルール(dominance rule)(ex. p%ルール等)

②二次秘匿(secondary disclosure)

\* 個票データの提供者側と利用者側の協力のもとで秘匿性が確保されるだけでなく、それに基づきながら、分析結果が公共の利益(public benefit)に資することが可能な形で個票データのアクセスが行われる。

→分析結果のチェックに関する自動化は行わない。

## Rule-Based Approach と Principles-Based Approach (Ritchie and Welpton(2015))

分析結果に関する最終成果物のチェックについては、Rule-Based Approach と Principles-Based Approachの2つの方法がある。

①Rule-Based Approach: 分析結果が公表されるかどうかを決定するための一連のルールにしたがって判断

Ex. セルに含まれる度数は3以上

利点: 単純な方法であるが、分析結果のチェックにおいてあいまいさがない。

欠点: 1) 秘匿の誤り(Confidentiality Error)

2) 有用性の誤り(Efficiency Error)



## ② Principles-Based Approach

- Rule of Thumbモデルの適用

- 原則として分析結果の公表の可否は、チェックを行う担当者に委ねられていること

- 研究者は、良い分析結果を出す責任があること

- チェックを行う担当者は、分析結果の意義を考慮すること

- チェックを行う担当者は、リソースの制約を考慮すること



研究者と分析結果のチェックを行う担当者との交渉によって、分析結果が公表可能かどうかが決まる。

利点：あらゆるリスクを取り除くことが可能で、かつ有用性を高めることができる。  
研究者に秘匿の認識を高めることができる。

欠点：不確実性があり、一貫性がない。リソースの必要性

→ チェックを行う担当者の訓練の必要性

# Rule of Thumb モデルについて

Rule of Thumb モデル: Principles-based approachにおける実践的原則

→秘匿の誤りを減らし、効率性の誤りを減らすモデル

Rule of Thumbモデルの全般的な基準(overall rule of thumb)

1. 全ての結果表において、セルの度数が10以上であること
2. 全てのモデルにおいて、自由度が少なくとも10あること
3. 全ての結果表において、特定のセルの度数が、それを含む行(row)ないしは列(column)の合計の90%を超えないこと
4. 全ての結果表において、特定のセルの度数が表全体の総計の50%を超えないこと

# 「安全な」統計(safe statistics)

「安全な」統計量: 分析結果の安全性が保証されない事例が例外的であること

⇒ 理論的な観点よりも、むしろ研究活動の実践的な経験に基づいて、(個人情報が漏えいされるような)特定の状況が「ありえない」かどうかの判断がなされる。



安全かどうかの判断は、露見リスクに関する理論に基づくというよりも、主観的な判断(subjectivity)

# 「安全な」統計について

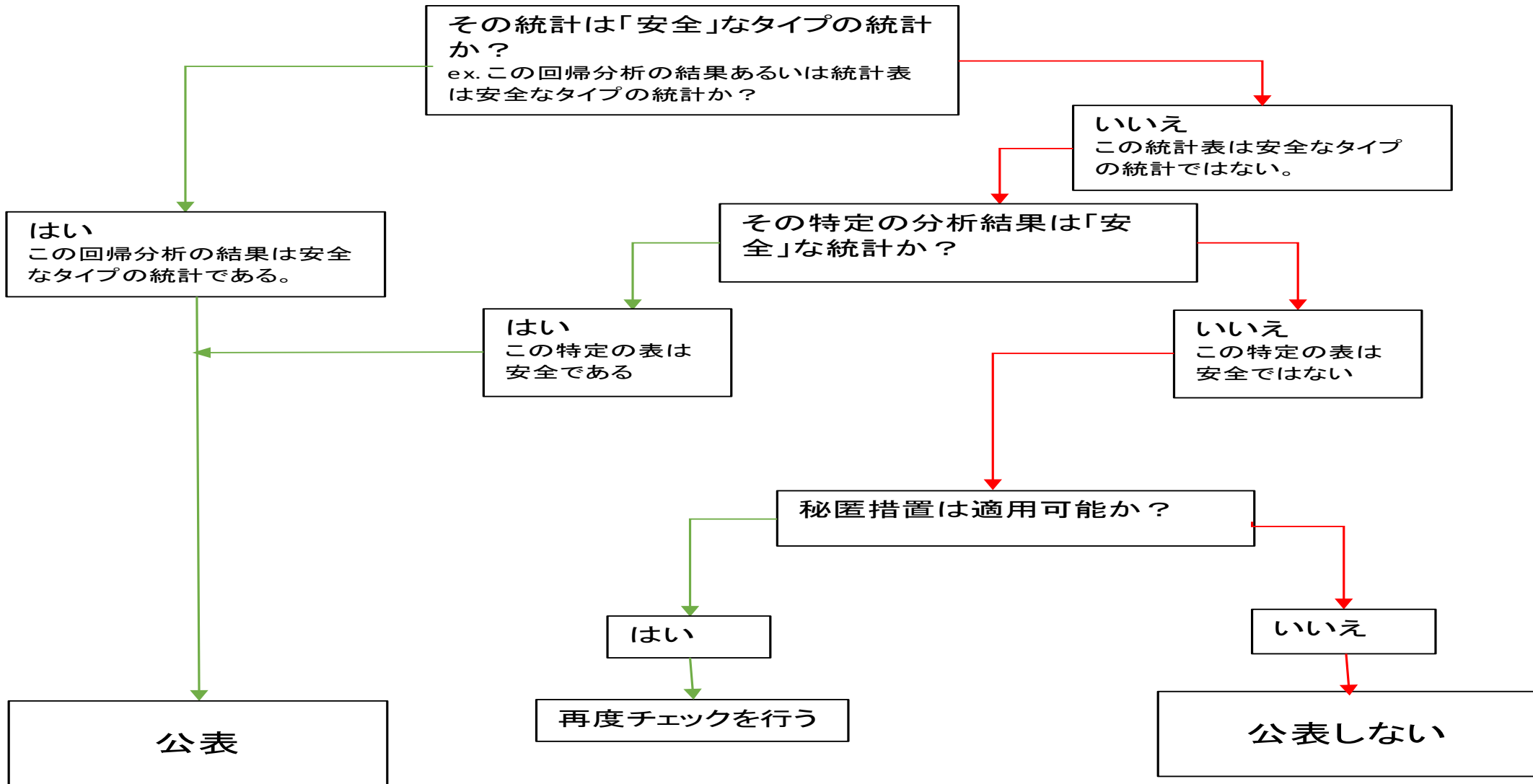
分析結果のタイプによって「安全な」統計か「安全でない」統計を類別することが可能

統計のタイプ	分析結果のタイプ	安全か安全でないか
	度数表(Frequency tables)	安全でない
	数量表(Magnitude tables)	安全でない
	最大値, 最小値, パーセンタイル(メディアンを含む)	安全でない
	最頻値	安全
	平均値, 指数, 比率, 指標	安全でない
	集中度(Concentration ratios)	安全
	分布に関する高次のモーメント(分散, 共分散, 尖度, 歪度を含む)	安全
	グラフ, 実際のデータに関する絵入りの表現 (pictorial representation)	安全でない
	相関係数と回帰分析	線形回帰モデルにおける回帰係数
非線形回帰モデルにおける回帰係数		安全
推定値の残差		安全でない
推定値に関する要約統計量および検定統計量( $R^2$ , カイ2乗値等)		安全
相関係数		安全

\* 重み付きの度数表については、安全な統計として位置付けられている。

出所 Brant et al.(2010)

# 図1 「安全な」統計の決定に関するチャート (decision tree)



## 分析結果をチェックする担当者の育成の必要性について

分析結果のチェックを行う担当者は、秘匿処理に関する専門的な知識と計量分析に関する知識の両方を備えている必要がある。

→計量分析の経験をもつ職員がチェックを行う必要性(あるいは計量分析の経験が豊富なPDを雇用するか)。

→ESSnetのガイドライン(2010)をもとに、Eurostatにおいて、チェックを行う担当者のための講習会が行われている。

### 3. イギリスのリモートアクセス—Secure Lab—

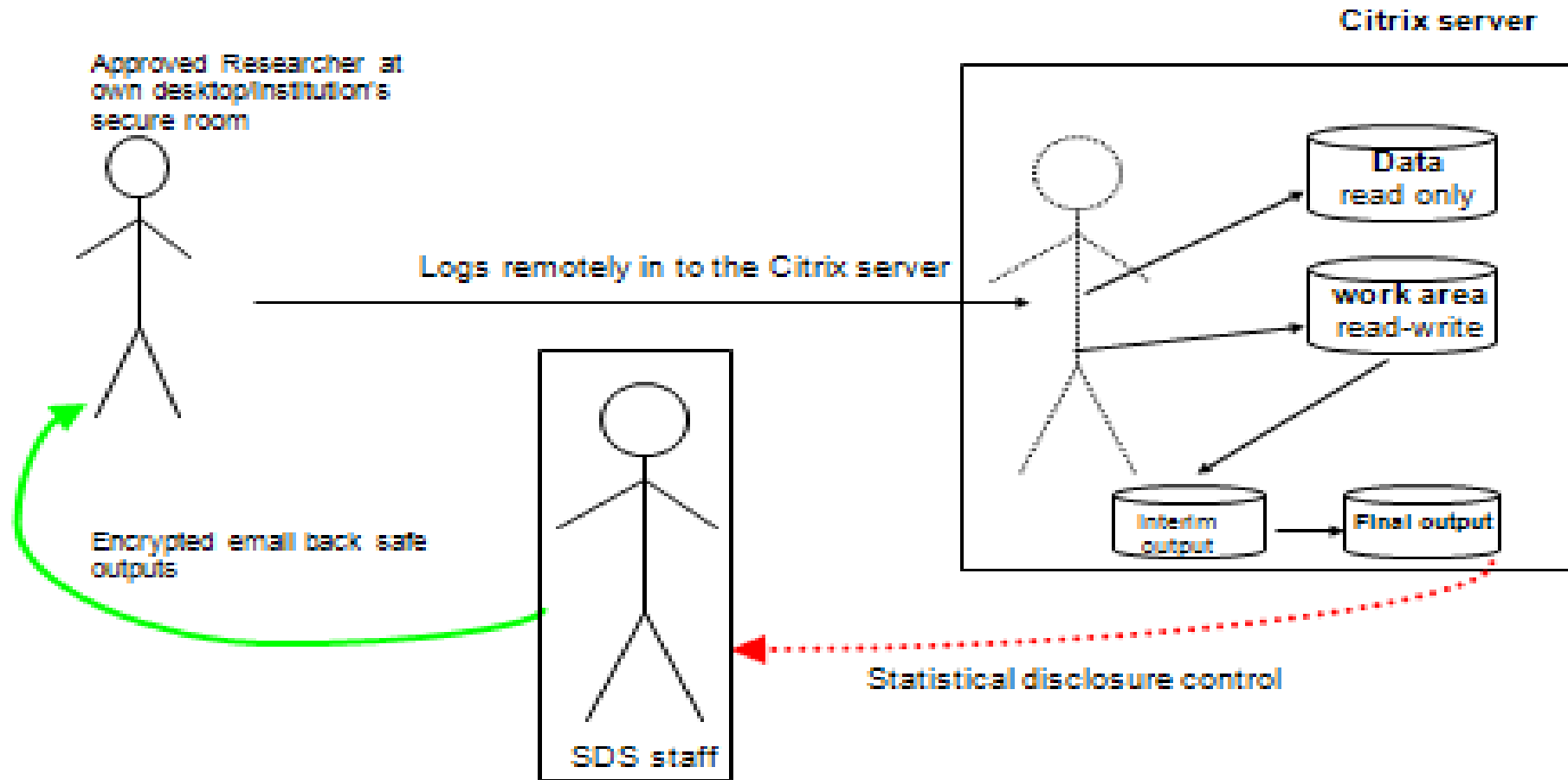
エセックス大学のUKデータアーカイブにおいて、ESRC(=Economic and Social Research Council)の資金提供を受けたthe Secure Data Service(=SDS)というリモートアクセス施設が創設⇒2011年にその運用が開始

・SDSは、UKDAにおいてデータの寄託、利用者の登録・管理、さらにはデータの提供サービスを担ってきたESDS(=Economic and Social Data Service)等と統合

→2012年にUK Data Serviceが創設

⇒現在は、UK Data Service のSecure Labがリモートアクセスのサービスを行っている(伊藤(2016))。

図2 Secure Labにおけるリモートアクセスの概略図



- \* 研究室にはカメラは設置されていない。
- \* \* 研究室で使用するPCは、所属機関のIPアドレスで管理されている。

出所 Afkhamai(2013)



図3-1 Secure Labにおけるログイン画面

SECURE  
DATA  
SERVICE  
enabling the  
research community

Service delivered by: UK • DATA ARCHIVE

Logon Messages Preferences

### Welcome to the Secure Data Service

Please enter your details to log in:

User name:

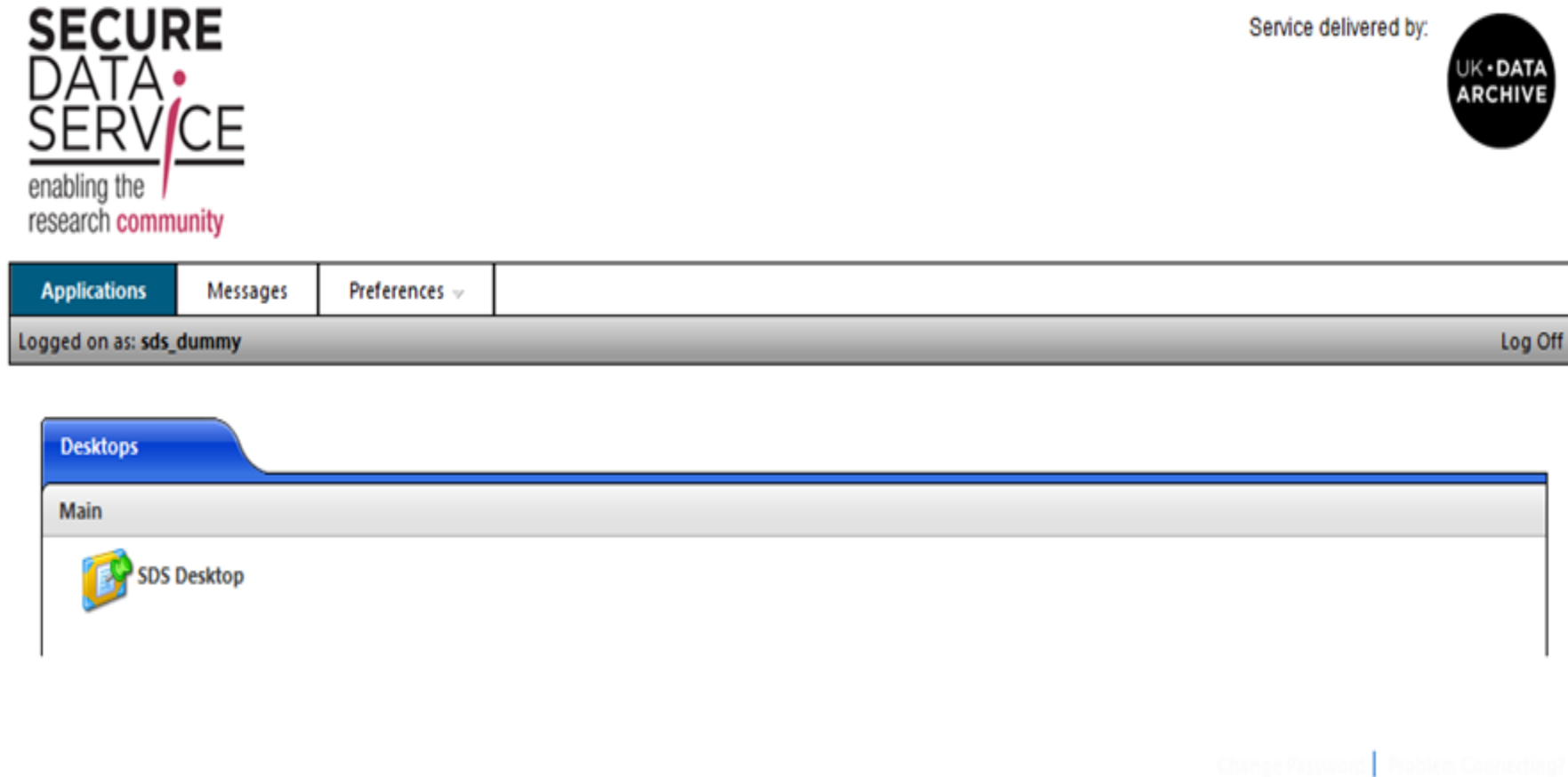
Password:

Log On

**This service is for authorised members only.**  
Anyone attempting unauthorised access will be considered for appropriate legal action.

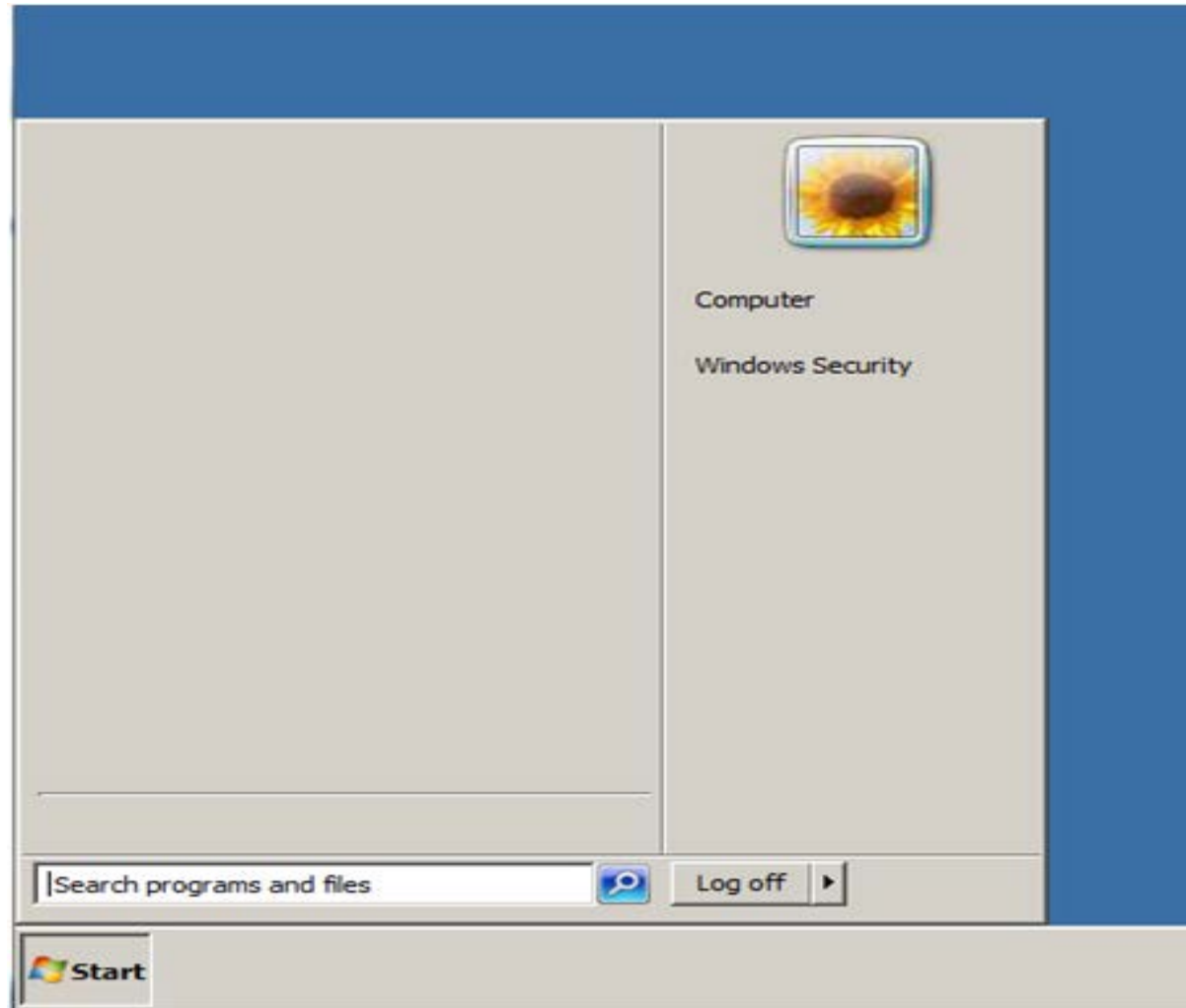
出所 UK Data Service(2014)

図3-2 Secure Labにおけるログイン後の画面1



出所 UK Data Service(2014)

図3-3 Secure Labにおけるログイン後の画面2



出所 UK Data Service(2014)

図3-4 Secure Labにおけるログイン後の画面3

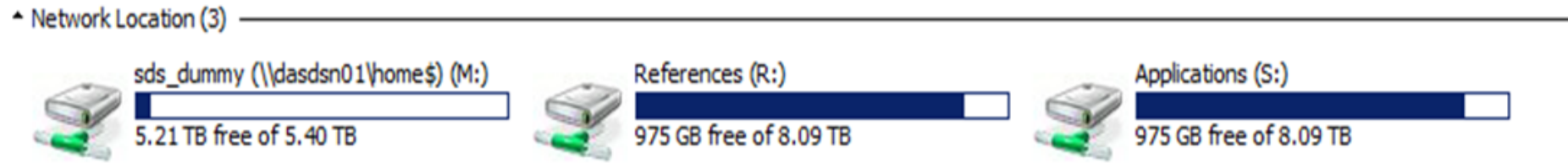


図3-5 M:ドライブの内容

Organize ▾ New folder

Name ^	Date modified	Type
Original_data	08/11/2011 11:18	File folder
SDC	03/11/2011 18:29	File folder
Syntax	25/11/2011 15:47	File folder
Uploads	03/10/2011 17:44	File folder
Working	25/11/2011 09:16	File folder

出所 UK Data Service(2014)

図3-6 Secure Labにおけるログイン後の画面4  
S:ドライブの内容

Name ^	Date modified	Type	Size
 ArcGis	23/11/2011 13:56	File folder	
 Toolkit	07/01/2014 17:30	File folder	
 Adobe Reader X	17/05/2011 12:09	Shortcut	2 KB
 explorer	11/05/2011 13:37	Shortcut	2 KB
 Microsoft Office Access 2007	13/01/2011 10:58	Shortcut	3 KB
 Microsoft Office Excel 2007	18/11/2010 12:43	Shortcut	3 KB
 Microsoft Office Powerpoint 2007	19/05/2011 11:51	Shortcut	3 KB
 Microsoft Office Word 2007	18/11/2010 12:43	Shortcut	3 KB
 MLwiN	17/05/2011 11:46	Shortcut	2 KB
 Notepad++	17/05/2011 11:47	Shortcut	2 KB
 OpenGeoDa	17/05/2011 11:42	Shortcut	1 KB
 R 2.12.2	17/05/2011 11:51	Shortcut	2 KB
 R 3.0.2	26/11/2013 10:23	Shortcut	2 KB
 SPSS Stastics	23/05/2011 12:45	Shortcut	2 KB
 StataSE 12 (64-bit)	15/10/2012 11:59	Shortcut	1 KB
 TeXnicCenter	25/11/2011 16:04	Shortcut	1 KB
 TeXworks	25/11/2011 16:09	Shortcut	2 KB

出所 UK Data Service(2014))

## Secure Labの特徴

・ONS以外の省庁が実施している統計調査のマイクロデータや行政記録のデータをリモートアクセスで利用することが可能。

- Longitudinal Study of Young People in England(Department for Education)
- English Housing Survey(Department for Communities and Local Government)
- Family Resource Survey(Department for Work and Pensions)
- Crime Survey of England and Wales(ONS/Home Office)

→ESRC Accredited Researcherの資格取得の必要性

・ONS以外の他省庁のデータの場合、個票データによってそのアクセスにおける法的根拠は異なる。

Ex. 1998年データ保護法(Data Protection Act 1998)等

- ・イギリスでは、個票データのリモートアクセスにおいて、特別のトレーニングコースを受けることが求められる。

- ・個票データのリモートアクセスにおける禁止事項

- ex.ディスプレイの写真を撮らない、紙媒体等にメモを取らない等

- ・個人情報を漏えいした場合の罰則規定に関する説明

- ・個票データのアクセスに関する利用者契約(User Agreement)に署名をすることが求められる。

→契約に違反した場合には、研究費の使用停止といった措置がとられる。

Secure Labにおける分析結果の持ち出しについて

チェック済みの最終成果物については、持ち出し可能であり、Eメールで研究者に送付することも可能である。

← 中間的な成果物(intermediate output)の持ち出しはできない。

→ リモートアクセスは24時間利用可能

⇒ リモートアクセスのサーバ上でMicrosoft Officeが利用可能であることから、リモートアクセス上で分析結果に基づいてペーパーを書くことができる。



## 4. わが国における分析結果の持ち出し審査のあり方について

- ・わが国における調査票情報(個票データ)の提供において、リモートアクセス型オンサイト利用における事後チェック型の調査票情報の利用方式(小林(2012))を検討し、個票データを用いたさらなる探索的な研究が可能になることは、個票データの利用者側だけでなく、提供者側にとっても有益であると考えられる。⇒ Five Safes Modelはわが国においても参考になるモデルだと思われる。
- ・わが国の場合、現時点では、24時間リモートアクセス施設が使用できるような状況は想定されていないと思われる。したがって、中間的な成果物については、こういった形であればアクセスすることが可能かを今後検討する必要がある。
- ・ヨーロッパの状況を踏まえると、一定のサンプル数があれば、回帰分析の結果については安全な統計だと考えて良いと思われる。他方、記述統計量については、「安全な」統計の決定に関するチャート等を踏まえながら、個別具体的なチェックが必要であろう。

## 主要参考文献

Afkhamai, R. (2013) “Statistical Disclosure Control Practice in the Secure Access of the UK Data Service”, Paper presented at Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Ottawa, Canada, pp.1-7.

Brandt, M., Franconi, L., Guerke, C., Hundepool, A., Lucarelli, M., Mol, J., Ritchie, F., Seri, G., Welpton, R. (2010) *Guidelines for the Checking of Output Based on Microdata Research*, Final Report of ESSnet Sub-Group on Output SDC, Eurostat

Desai, T., Ritchie, F., Welpton, R.(2016) “Five Safes: Designing Data Access for Research”, Economics Working Paper Series 1601, University of the West of England.

伊藤伸介(2012)「政府統計マイクロデータの提供における匿名化措置—イギリス統計法における法制度的措置と攪乱的手法の適用可能性を中心に—」明海大学『経済学論集』Vol.24, No.3, 1～14頁

伊藤伸介(2016)「政府統計におけるリモートアクセスと秘密保護について—イギリスを例に—」, 『経済学論纂(中央大学)』第56巻第5・6合併号, 1～19頁

小林良行(2012)「公的統計マイクロデータ提供の現状と展望: 一橋大学での取り組みをもとに」『日本統計学会誌』第41巻, 第2号, 401～420頁

Ritchie, F.(2008) “Secure Access to Confidential Microdata: Four Years of the Virtual Microdata Laboratory”, *Economic & Labour Market Review*, Vol.2, No.5, pp.29-34.

Ritchie, F. and Welpton, R. (2015) “Operationalizing principle-based output statistical disclosure control”, mimeo

UK Data Service(2014) User Guide: For Your Secure Lab Account.