Piloting a Safe Health Researcher course

Louise Corti and Deb Wiltshire
UK Data Service

ADR Conference
9-11 December 2019
Cardiff

Copyright © [2019] UK Data Service. Created by UK Data Archive, University of Essex
BY-NC-SA

My organisation: the UK Data Archive

- Department of the University of Essex
- Running ESRC's 'Data Bank' since 1967
- Leads the national UK Data Service for ESRC
- Over 50 years of curating and providing access to data for research and teaching
- A Trusted Digital Repository (TDR), accredited to ISO27001 Info Security Management standard, Digital Economy Act Processor application in process
- Work closely with research funders and key data producers /institutions: research centres; UK NSIs, govt departments, British Library etc.

UK Data Service

Health data

UK government surveys

- Health Survey for England, Adult Psychiatric Morbidity Study etc.

ESRC birth/cohort & longitudinal studies incl.

- National Child Development Study (NCDS): 1958+
- 1970 British Cohort Study (BCS): 1970+
- Millennium Cohort Study (MCS): 2000+
- English Longitudinal Study of Ageing (ELSA): 2002
- Understanding Society (UKHLS): 2009+
- Growing Up in Scotland (GUS)

MRC, Wellcome Trust, CRUK

- MRC cohorts: Whitehall II –available soon
- legacy clinical trials

Longitudinal & cohorts studies: linked data

Linked data

Administrative records, biomarkers/genetics, a… meters, pollution, R…

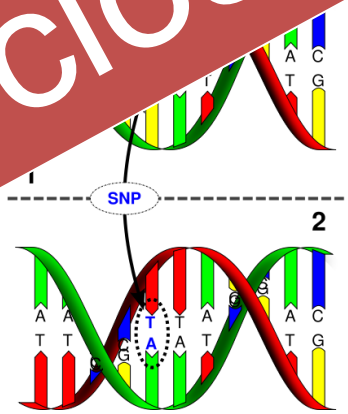
Increased disclosure risk

http://www.actigraphy.com/solutions/actiwatch/actiwatch2.html

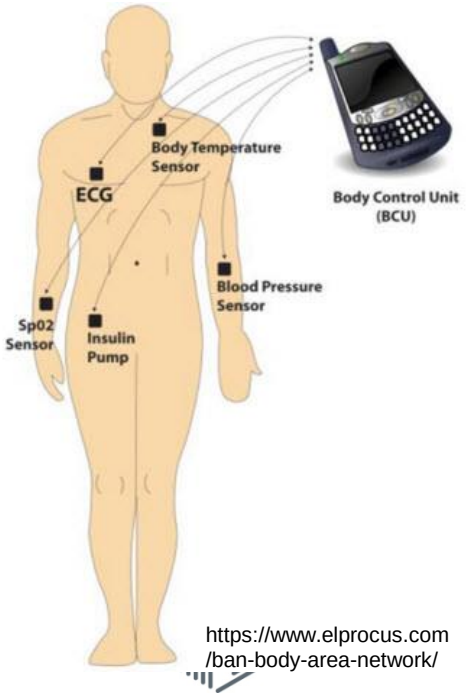
	after blast in ED or
	death (≥4 h after blast)
	Injury types
	Injury locations
	Sought emergency care at ED from scene
	Transported to ED by EMS from scene
	Transferred from ED to other hospital
	Transferred from other hospital to ED
	Discharged home from ED
	Admitted to hospital
	Received operative care
	Types of operations received
	Received critical care (admitted to ICU)
Time course of emergency needs	Time interval to ED arrival (min)
	Time and date of death

ED, emergency department; EMS, emergency medical services; ICU, intensive care unit.

https://www.researchgate.net/figure/Data-Abstracted-From-Hospital-Records_tbl1_7663074



https://knowgenetics.org/snps/



https://www.elprocus.com/ban-body-area-network/

Spectrum of Access: UK Data Service

Open

- No disclosure risk
- Open licence; few restrictions on reuse

Safeguarded

- Zero to low disclosure risk
- User agreement
- Authentication, authorisation & auditing

Controlled

- Disclosure risk/personal data/legal gateway
- Authentication, authorisation & auditing
- Added safeguards (Five Safes)

Access points at UK Data Service

Whole studies can be made available as multiple datasets under different access conditions, with DOIs

Controlled Access: detailed geographies of locations and variables deemed too sensitive for standard release

Fives Safes Framework

Fives Safes enables safe access to data that meet the needs of data protection yet fulfils the demands for open science and transparency

- Safe data - treat data to protect confidentiality
- Safe people - educate researchers to use data safely
- Safe projects - research projects for 'public good'
- Safe settings – Secure Lab environment for personal data
- Safe outputs – Secure Lab projects outputs screened

5 Safes Animation

UK Data Service

Backdrop to Safe Health Researcher: fusion

UK Data Service

About us | Get data | Use data | Manage data | Deposit data | News and events

Home > Manage data

- Data lifecycle
- Plan to share
- Legal and ethical
- Rights
- Document your data
- Format your data
- Store your data
- Collaborative research
- Training
- Tools and templates
- Handbook

Prepare and manage data

"Good data habits from the moment you start planning your research"

SHARE

Data created from research are valuable resources that can be used and reused for future scientific and educational purposes.

Good data management practices are essential in research, to make sure that research data are of high quality, are well organised, documented, preserved and accessible and their validity controlled at all times. This results in efficient and excelling research. Well managed data are easily shared and can thus be used for new research or to duplicate and validate existing research.

Data management needs to be planned early on in research, so that practices can be implemented throughout the research cycle.

LOGIN / REGISTER

DATA CATALOGUE

GO

GUIDE TO GOOD PRACTICE

Get the handbook on Managing and Sharing Research Data: a Guide to Good Practice from Sage publications

QUICK ACCESS TO

FAQ about managing data

Regular depositors

How to deposit

- Data offers
- Regular depositors
- Self-deposit to ReShare

Preparing data

Owners and producers

ORCID

Depositor stories

"What you need to know to deliver a dataset"

For regular depositors of large-scale social surveys or government data series there are five key steps to depositing a collection:

- Plan - what to do before fieldwork begins
- Prepare - what to do during and after fieldwork
- Negotiate - discuss deposit with us
- Deposit - transfer data to us
- Ingest - what we do with your data

Consult our succinct 2014 guide on Depositing shareable survey data, preferably before the survey has been commissioned and fieldwork has started.

Plan	Prepare	Negotiate	Deposit
Ingest			

LOGIN / REGISTER

DATA CATALOGUE

GO

DEPOSITING SHAREABLE SURVEY DATA

QUICK ACCESS TO

FAQ about data deposit

Preparing your data

Advice and training
Guides
Video tutorials
Data in use
Citing data
Secure Lab
Teaching with data
Student resources
Data Skills Modules
Secondary analysis

Secure Lab

"Enabling research through responsible use of detailed and sensitive data"

The UK Data Service is one of the few organisations who have the experience and expertise to build a world-leading secure access system.

What is the Secure Lab? | Our collection | Five Safes framework | How it works | How do I apply?

SHARE

The UK Data Service Secure Lab 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. Our specialised staff apply statistical control techniques to ensure the

LOGIN / REGISTER

DATA CATALOGUE

GO

HOW TO LOG IN TO THE SECURE LAB

View a video turorial to show researchers who have been granted approval to use secure access level data how to log into their Secure Lab account.

QUICK ACCESS TO

About our data
Business microdata

Training areas covered

Research Data Management & Sharing Training	Safe Researcher Training (SRT)
Topics – all areas of social science and health	Topics – limited to economics
Legal, ethical, consent and confidentiality	
Data management skills, data cleaning (QAMyData) data documentation	
Safe Data: input disclosure review, (sdcMicro), anonymization, access pathways	**Safe People** - Good research conduct, **Safe Settings** behavior, Creating **Safe Outputs**
Safe storage and transfer: encryption	
Reproducibility – good code	

Exploring combined training aspects

Focuses on the data lifecycle

Training combines expertise in:

research data management /sharing

- ✓ Safe Data
- ✓ Safe Projects

and

AR Safe Researcher Training (SRT)

- ✓ Safe People
- ✓ Safe Settings
- ✓ Safe Outputs

Safe data Input sdc for data publishing

provided at less detail. For example, if ethnicity is required at 16-category level, then age could be banded, or geography could be “England & Wales” only.

Demographic variables

Variable	Level of detail
Place of residence	Nothing lower than GOR
Sex	Male/female
Ethnicity	Grouped into 6 categories
Date of death	Year only
Age at death	In years

Health-related variables

Variable	Level of detail
Date of Cancer registration	Year only
Diagnosis	ICD-10 code to 4 digits (or ICD-9 code for historical data)
Age at Cancer registration	In years
Histology	Coded only, i.e. not free text
Outcome	Died = yes/no

https://www.ons.gov.uk/file?uri=/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/healthstatistics/proceduresforthereleaseofcancermicrodatatcm77252515.pdf

\+

Sdc Micro combinations: k anonymity

Goal = risk vs. utility

UK Data Service

Safe people

Accredited Researcher status

Social /economic sciences: Accredited Researcher concept

- ✓ Data owner requirement e.g. ONS, ESRC
- ✓ Experience/expertise
- ✓ Good attitude/applies common sense
- ✓ Have understanding of disclosure risk and producing Safe Outputs
- ✓ Attend a one-day training course and pass a test

Medical research: focus on 'bona fide researcher' concept*

- *the professional expertise and experience to conduct bona fide research*
- *a formal relationship with a bona fide research organisation that requires compliance with appropriate research governance and management systems*

UK Data Service

* https://mrc.ukri.org/documents/pdf/data-sharing-from-population-and-patient-studies/

Creating releasable research outputs

Primary Disclosure: Threshold Rule

- No counts/cells with less than <n> units
- Different data providers apply different thresholds e.g. commonly 10
- Aggregation protects against subtraction attacks or differencing (multiple outputs)

Secondary Disclosure

e.g. small number in rare diseases

Handbook best practice on Safe Outputs:
https://ukdataservice.ac.uk/media/622521/thf_datareport_aw_web.pdf

Safe Data Access Professionals

Home

Welcome to the Working Group for Safe Data Access Professionals.

Established in 2011 by Tanvi Desai (formally of the London School of Economics and Administrative Data Service), the purpose of the Group is to facilitate sharing of expertise, best practice and knowledge between organisations engaged in providing secure access to confidential sources of data from the health and social science research sector.

Find out more about what we do.

Safe Health Researcher Training

- ✓ Concept aimed at data managers whose role it is to manage disclosive data and its access
- ✓ Interest from health data managers, data governance folk, and safe haven administrators
- ✓ See benefits from 5 Safes Framework lifecycle training
- ✓ Value better understanding of disclosure review, appropriate data access pathways and risk in research outputs

Case study 1: sharing clinical trials data

- MRC, Wellcome Trust, CRUK bought into into metadata portal (Clinical Study Data Request)

- Desire to archive legacy trials – little take up

- ✓ Explored running day training sessions with clinical trials data managers
  - London School of Hygiene and Tropical Medicine
  - Cochrane Collaboration, early Childbirth group, Liverpool

Case study 2: Sharing legacy MRC cohorts

- Expensive per centre
- Few research-ready datasets
- Little potential for reproducibility, no DOIs

- Over-reliance on Safe Projects
- Limited, but increasing use of Safe Settings
- Little or no use of Safe Data or Safe People

- ✓ Benefit from 5 Safes Framework approach
- ✓ Training ALSPAC Jan 2020

Case study 3: THF and CRUK

- The Health Foundation, Cancer Research UK
- Running SRT style Safe Outputs training with health research examples, plus some data awareness for non academics

Data Awareness Training

Cancer Research UK (Cancer Intelligence team) and The Health Foundation have worked together on a set of training slides which they would like to share with others. Based on the Five Safes framework, these cover the following aspects of accessing data in a Safe Setting:

- **The Five Safes**: how this framework (originally developed by Professor Felix Ritchie at the ONS) provides a useful information governance framework
- **About data**: why are some data more 'sensitive' than others? how does the law apply? why are different ways of accessing the data? (covering **Safe Data** and **Safe Projects**)
- **Safe People**: what do we expect from people accessing sensitive data?
- **Safe Outputs**: what is Statistical Disclosure Control (SDE)? Why is it important for analytical/research outputs released from a Safe Setting? How is SDC undertaken (this links to our SDC Handbook).

Training platform: aspirations

- ✓ Collaborative model for health research
- ✓ Experts are early adopters of 5 Safes, safe haven administrators and RDM, data curation/publishing
- ✓ Use of the Safe Data Access Professionals group
  - ✓ Run growing ‘community’ group, website and events
  - ✓ Build on successful handbook on Outputs
  - ✓ Deliver training and syllabus advice
  - ✓ Offer consultancy/trainers

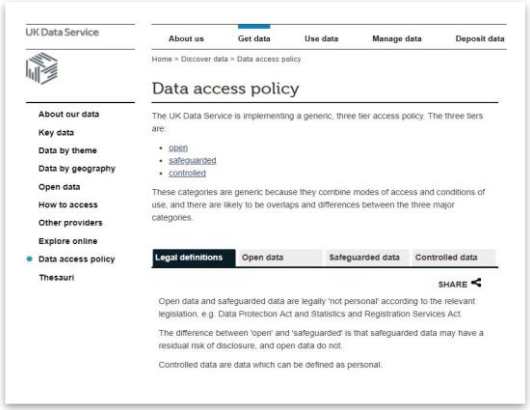
Thanks

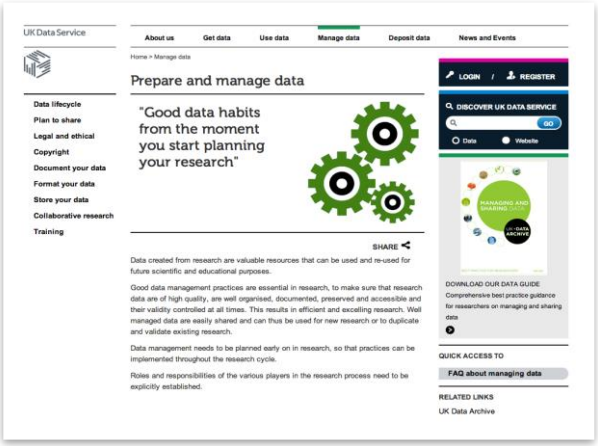
Richard Welpton, THF for his continuing inspiration on 5 Safes and running the SDAP group

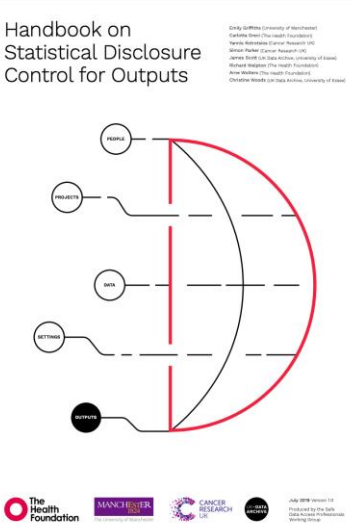
Simon Parker and Iannis Kotrotsios, CRUK for their expertise and enthusiasm in health data safe haven work

Our Secure Lab team at Essex: Chris Woods, James Scott, Valerja Kolbas and Beate Lichtwardt for training Safe People and processing Safe Outputs

ukdataservice.ac.uk







UK Data Service

Secure Lab services

Enabling secure research access to data while protecting confidentiality

The Five Safes Framework

- Safe data
- Safe projects
- Safe people
- Safe settings
- Safe outputs

As promoted by the Office for National Statistics

ukdataservice.ac.uk
help@ukdataservice.ac.uk

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

Keep connected

Copyright, University of Essex, 2019

corti@essex.ac.uk
UK Data Service
University of Essex, Colchester, UK
UKdataservice.ac.uk

Subscribe to UK Data Service list:
www.jiscmail.ac.uk/cgi-bin/webadmin?A0=UKDATASERVICE

Follow UK Data Service on Twitter: @UKDataService, @UKDSRDM
Youtube: www.youtube.com/user/UKDATASERVICE

Clinical trials data managers workshop

Programme: Assessing Data Quality and Disclosure Risk in Clinical Trials Data Sharing

9.30	Registration and Refreshments
9.45	Presentation: Principles and practicalities of appraising data for sharing: quality, risk and pathways to publishing
10.15	Debate: To share or not to share - your world view
10.45	Exercise: Checking a CT file (the way you usually do) and feedback
11.15	Presentation and demo of QAMyData tool: What elements to assess in data and how to run the software
11.45	QAMyData Data hands-on exercises
12.30	**Lunch**
13.00	Evaluating 'risk' in data: principles, rules and tools
13.45	Group exercise: Assessing risk in datasets to be shared
14.00	Presentation and demo of sdcMicro: What aspects of disclosure risk to review in data and how to run the software (content)
14.30	sdcMicro hands-on exercises
15.15	**Coffee break**
15.30	Safe Researchers: Safe People and Safe Outputs (context)
16.00	Group exercise: Assessing Safe Outputs
16.30	Feedback and queries about tools and use
16.45	Close

http://blog.ukdataservice.ac.uk/get-to-it-clinical-data/

UK Data Service

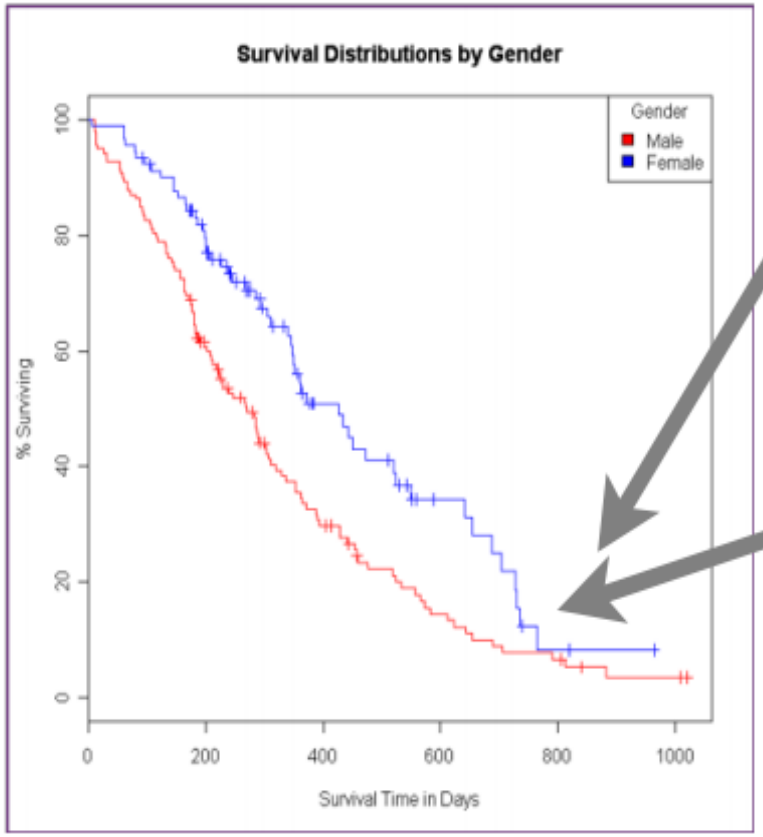
Safe outputs

Survival Analysis: Kaplan Mayer curve

- Graphical representation of survival
- Shows number of observations ‘surviving’ at each time point
- Each step – single individual or group of people
- ‘N’ at each step must meet threshold

- Could band the step changes
- Or remove the axes scales
- Careful when rare disease

FIGURE 8: *Kaplan Mayer curve*



Clear labels

Are there sufficient numbers of observations between step changes?

Could a small number of individuals be identified via a step change?