What does a user look for in a survey dataset?

(What does this user look for in a survey dataset)

(A few thing that this user would like you to think about)

Vernon Gayle, FRSE FAcSS
(University of Edinburgh)

Challenges and Opportunities for Social Survey Data Collection in Scotland

Survey Practice Forum

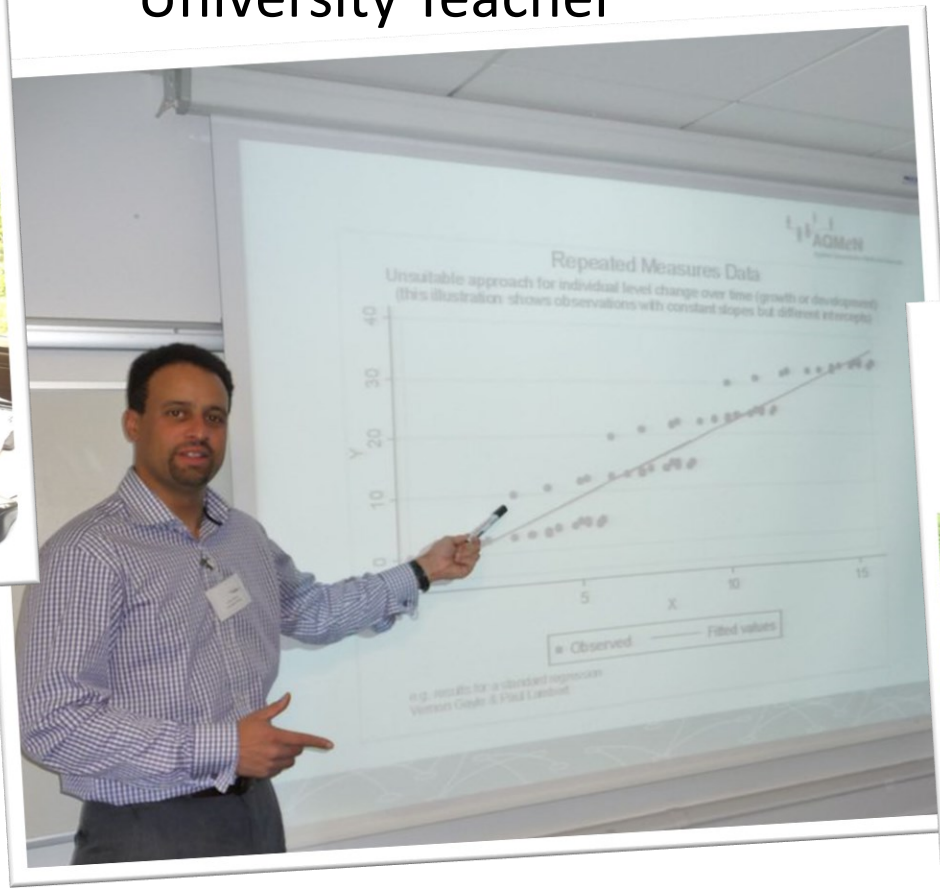
Royal Society of Edinburgh

6th December 2024

THE UNIVERSITY of EDINBURGH

Survey Users / Researcher

University Teacher



Postgraduate Supervisor

Things have improved for survey data analysts…

1994

new datasets,
the internet,
faster computers,
better storage,
improved software,
better documentation,
some linked admin data

2024

What does a user look for in a survey dataset?

A Hiaku

(three lines – 5, 7 and 5 syllables)

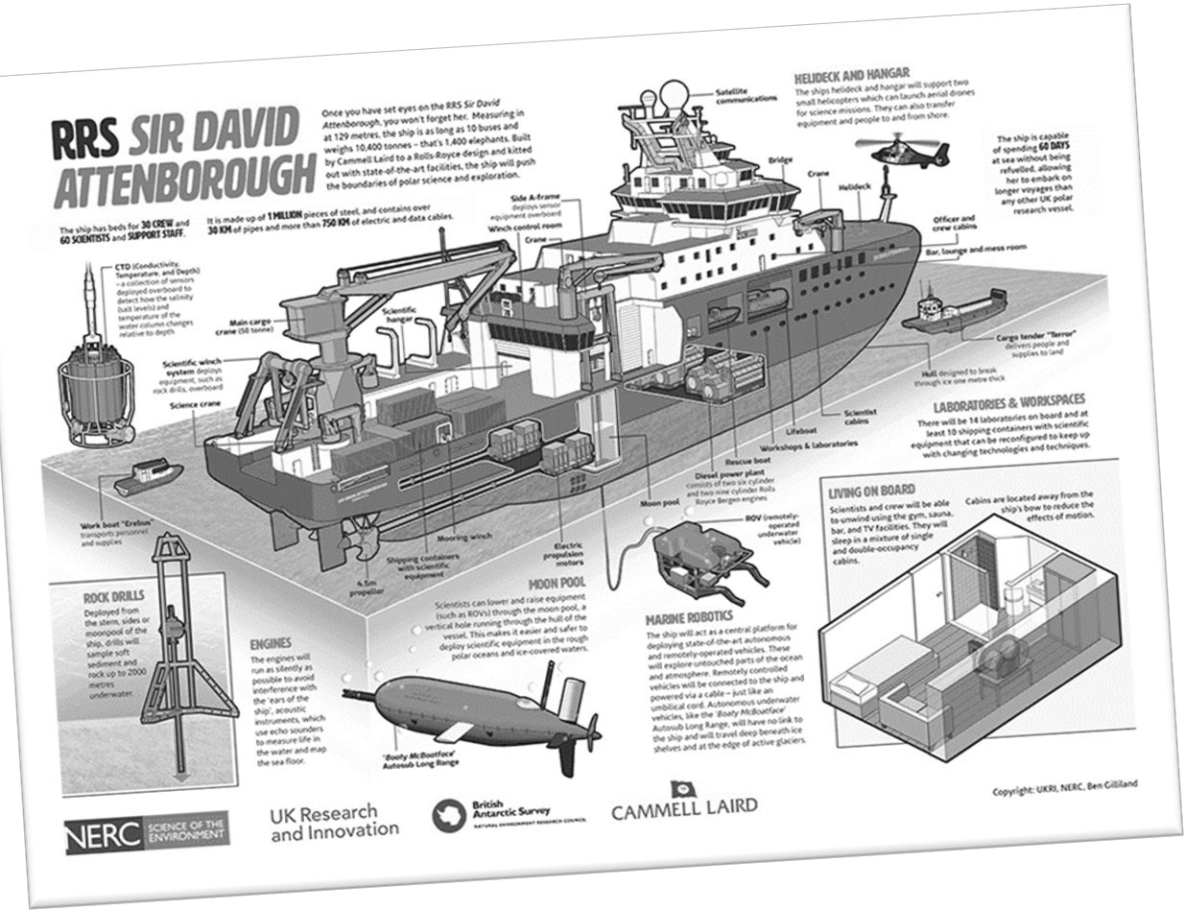
A big sample size
Many good variables
Easy access too

A Hiaku

(three lines – 5, 7 and 5 syllables)

A big sample size
Many good variables
Easy access too

Big n
Large *k*
Multiple *t*
Suitable *j* (geography)



The Omnibus Surveys

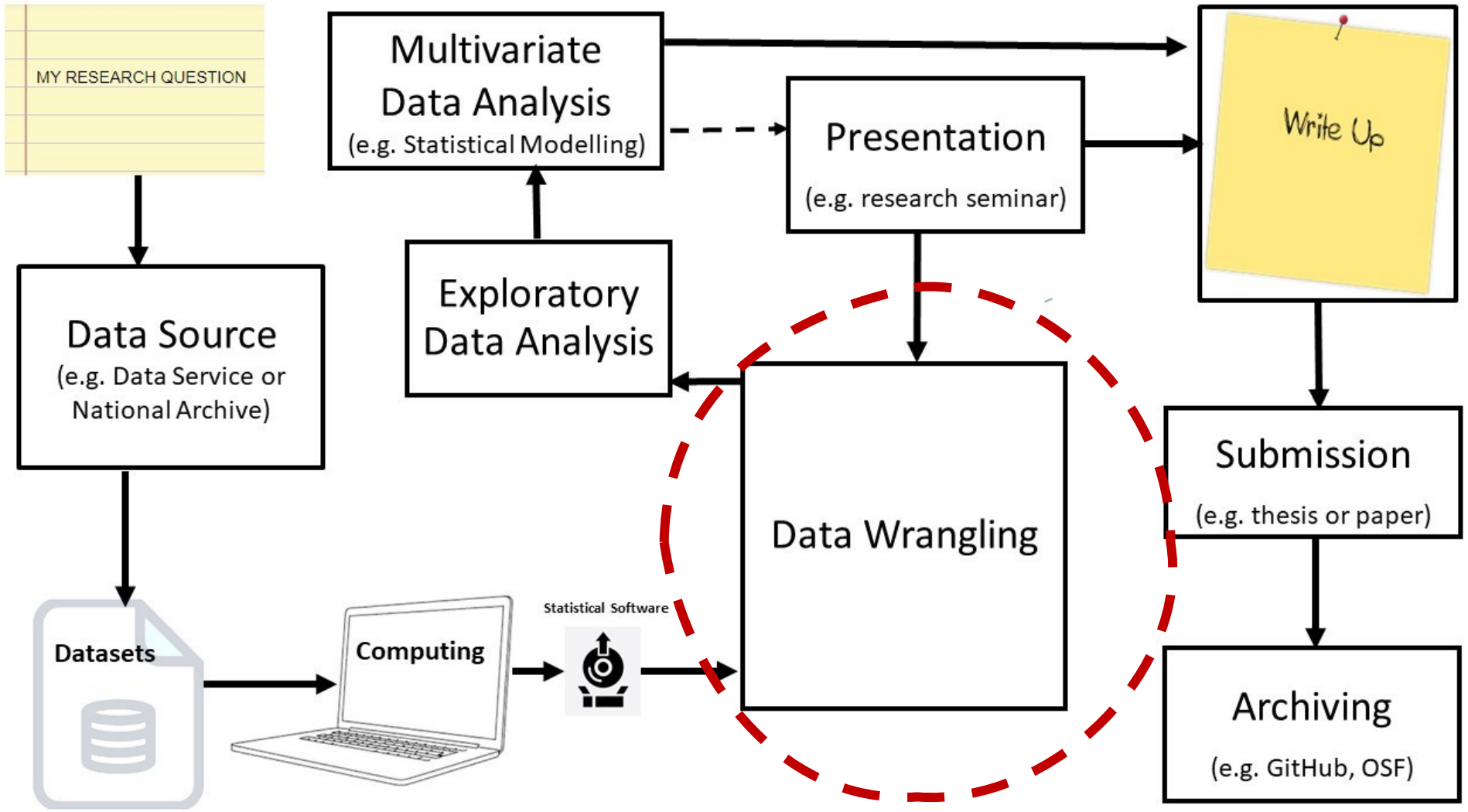
Household Panels

Birth Cohorts

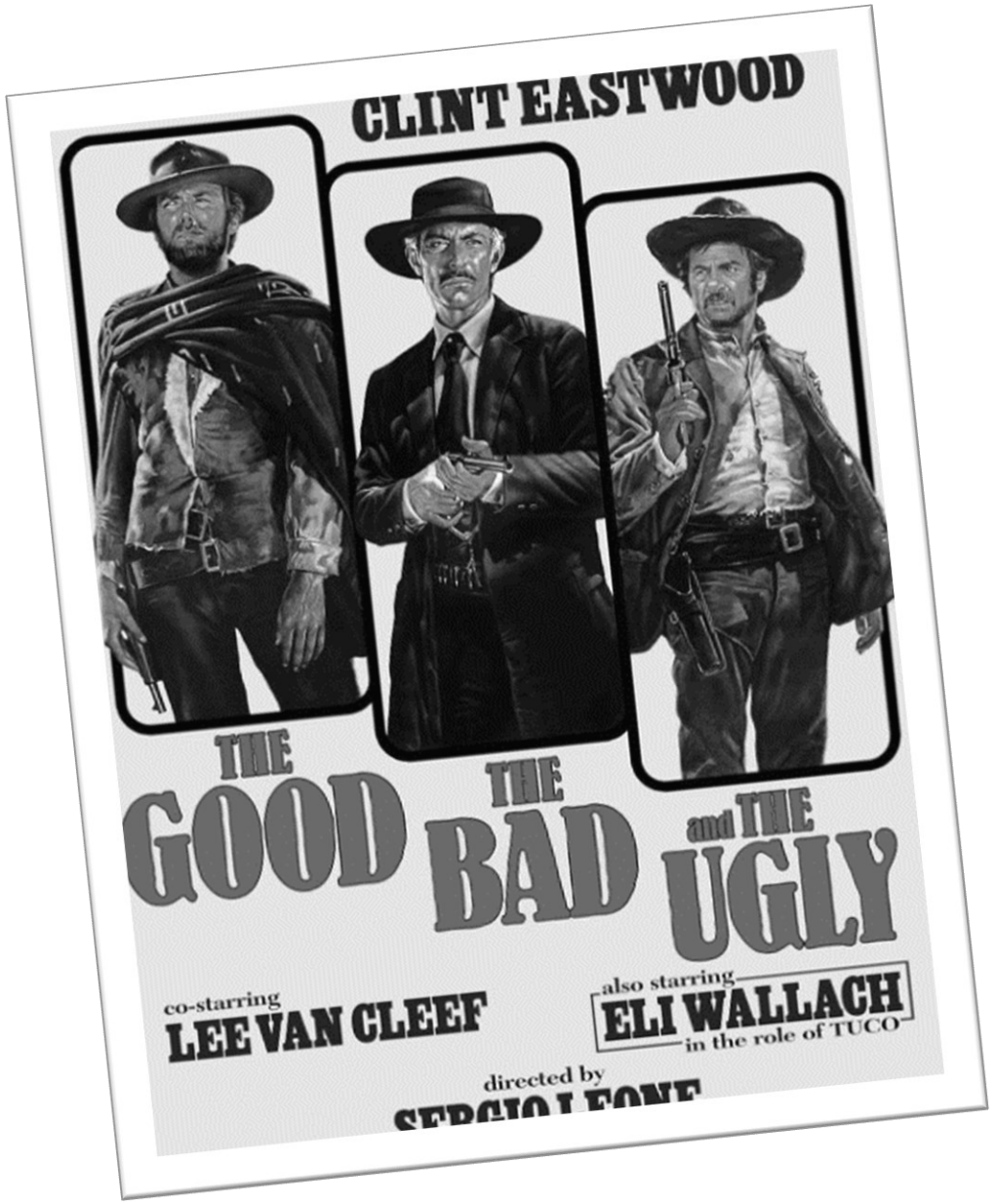
Annual Population Surveys

Challenge to encourage users to be more tolerant –
these surveys are not designed to address a specific hypothesis or for researching single topics or research field

Producing (nearly) Research Ready Datasets



Complex Designs and Selection Strategies



The Good

Assists fieldwork challenges

Provide territorial data (e.g. Scotland)

Facilitate the geographical analyses (e.g. areas)

Facilitate analyses of under-represented groups (e.g. ethnic minority)

The Bad

Analysts ignoring survey designs and selection strategies

The Ugly

Not adequately representing design in analysis

Not being able to provide some measures (e.g. model fitting statistics)

Missing data methods often trickier (e.g. svy m.i. models)

Complex Designs and Selection Strategies

Challenges

Are complex designs and selection strategies necessary?

Can some designs be simplified to help data analysts?

Opportunities

Provide more training

Show routine examples

Show tricky examples (e.g. svy m.i. models)

Documentation

Some Other FAIR Principles

Findable – can the user find in things at the helicopter level

Accessible – can the use get to the information in the weeds

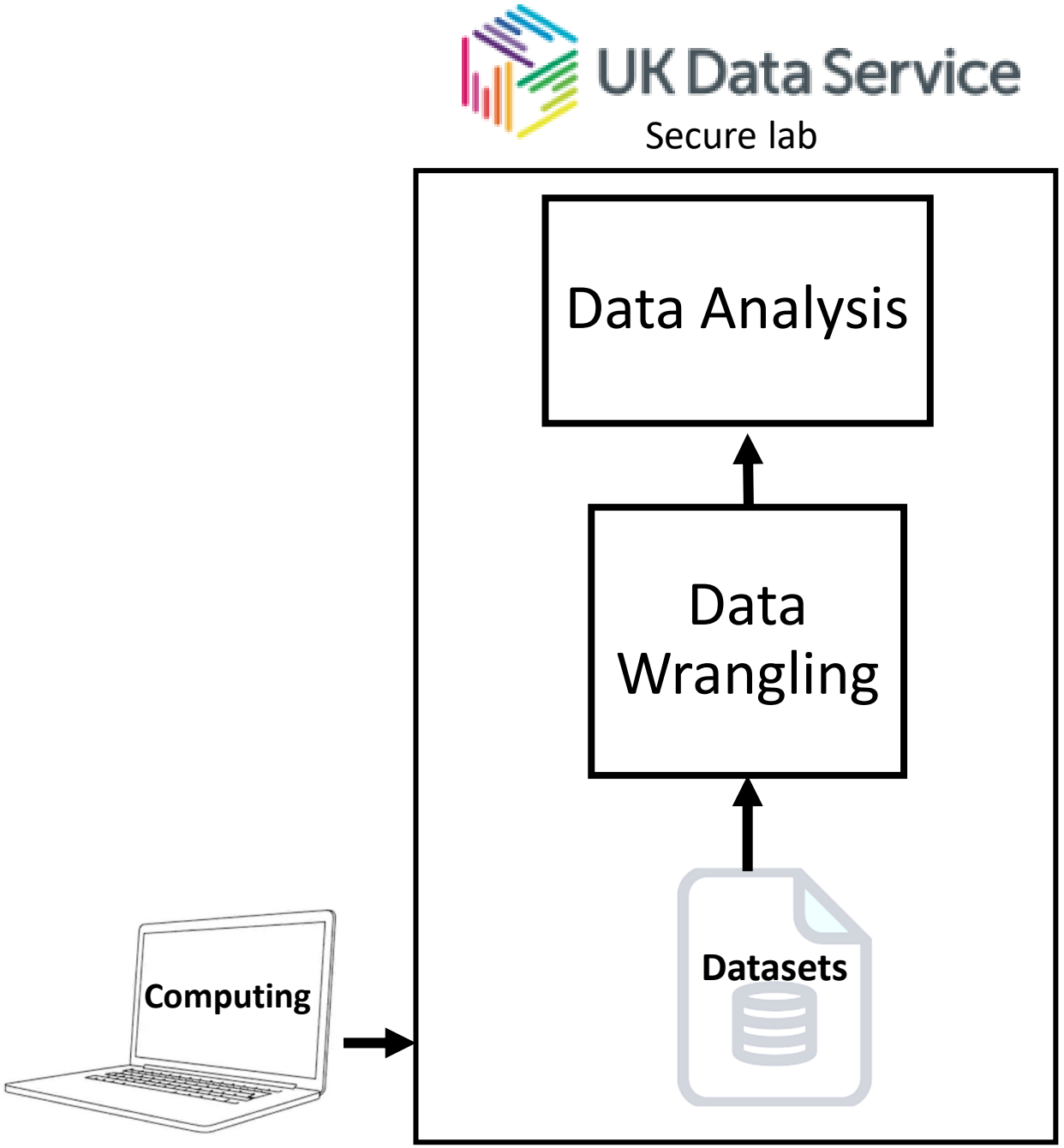
Informative – does it tell the user what they need to know

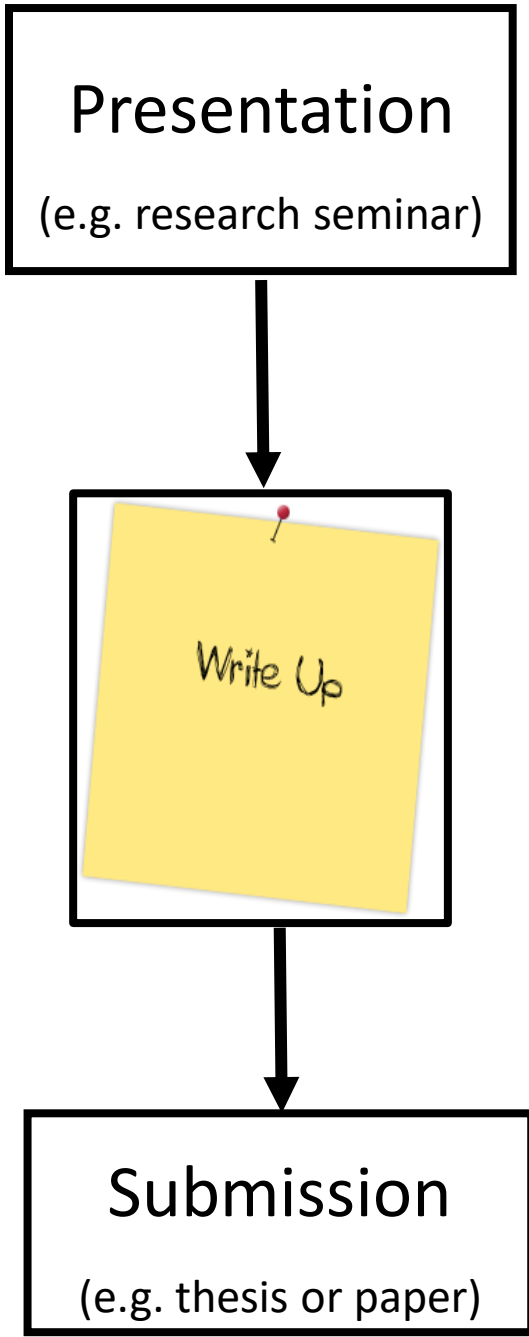
Robust – is the info stable (e.g. 3 waves later on)

Data Access

- End User Licenses have been very effective

- Increasing used of Special License (e.g. Occupational SOC codes) is this really necessary?


UK Data Service
Secure lab
Data Analysis
Data Wrangling
Datasets
Computing


Presentation
(e.g. research seminar)
Write Up
Submission
(e.g. thesis or paper)

Data Access

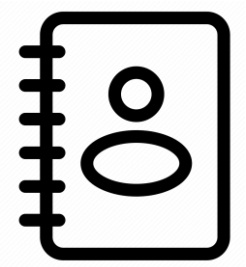
- Secure lab work is more time consuming than desktop work

- Is too much linked admin data unnecessarily held within the secure lab?

Linking to Administrative Data Sources

A characteristic of micro-level administrative social science data sets is that they usually have a large number of observations (n), for example individuals, but a smaller number of social science related explanatory variables (k) than would be the case for social surveys

Playford, C.J., Gayle, V., Connelly, R. and Gray, A.J., 2016. Administrative social science data: The challenge of reproducible research. *Big Data & Society*, *3*(2), p.2053951716684143.



An investigation of the consistency of GCSE qualifications data in administrative educational records and a national social survey

Sarah Stopforth, University of York

Roxanne Connelly, University of Edinburgh

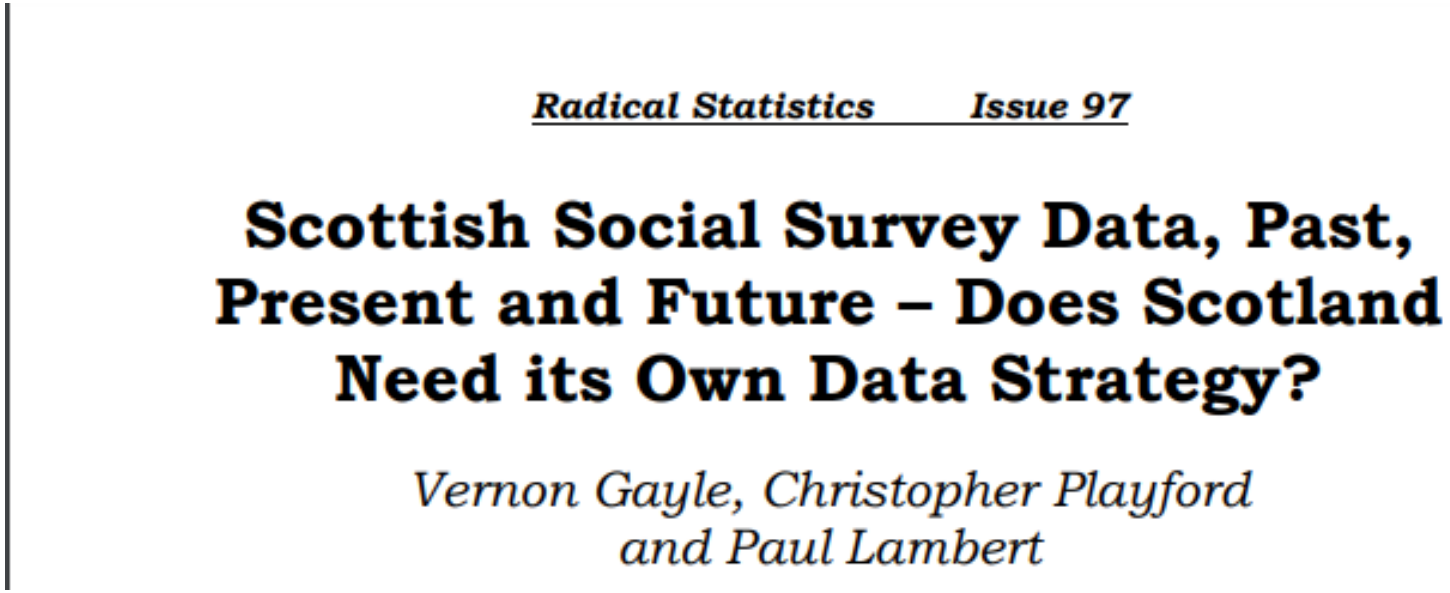
Vernon Gayle, University of Edinburgh

- We advise researchers to access the linked GCSE data from the National Pupil Database (NPD)
- Beware the raw NPD data requires a large amount of data wrangling to prepare – measures are unsuitable for immediate use in analyses (e.g. in the pupil-level dataset)

Is the quality of the admin data = to the quality of the survey data?

Is the quality of documentation of the admin data = to the quality of the survey documentation?

Do We Need A Scottish Strategy?



Radical Statistics *Issue 97*

Scottish Social Survey Data, Past, Present and Future – Does Scotland Need its Own Data Strategy?

Vernon Gayle, Christopher Playford and Paul Lambert

Gayle, V., Playford, C. and Lambert, P., 2008. Scottish Social Survey Data, Past, Present and Future–Does Scotland Need its Own Data Strategy?. *Radical Statistics*, *97*, pp.82-97.