

# The Data Divide: from collection to analysis

**Sharon Witherspoon MBE FAcSS**  
**Head of Policy**  
**Academy of Social Sciences**

**Or:**

**Tribes, Sects or  
Neighbouring Communities?**

Previous SRA lecture in 2014, by Professor David Rhind (and me), *The Impact of Big Data on Social Research*

[the-sra.org.uk/wp-content/uploads/rhind-witherspoon.pdf](http://the-sra.org.uk/wp-content/uploads/rhind-witherspoon.pdf)

Campaign for Social Science Sage Annual Lecture 2015,  
*Social science for public good: who benefits, who pays?*

[campaignforsocialscience.org.uk/wp-content/uploads/2015/12/CfSS-Annual-Lecture-2015-Sharon-Witherspoon.pdf](http://campaignforsocialscience.org.uk/wp-content/uploads/2015/12/CfSS-Annual-Lecture-2015-Sharon-Witherspoon.pdf)

## Data ‘analysts’ and data ‘collectors’: who are we?

- **Tribes:** a group of distinct people...largely self-sufficient, and not integrated into the national society
- **Sects:** a group of people with different religious beliefs (typically regarded as heretical) from those of a larger group to which they belong
- **Neighbouring communities:** a group of people having a particular characteristic in common, especially the condition of sharing or having certain attitudes and interests in common?

## Data design/ collection issues affect 'big data' too

- Google flu analytics vs. Food Standards Agency use of social media norovirus measures. (Radio 4, More or Less, [www.bbc.co.uk/programmes/b08343jx](http://www.bbc.co.uk/programmes/b08343jx) )
- **Hypothetical:** benefit fraud and whether community norms make a difference – does knowing other people who falsely claim make people more likely to do so?
  - Or could it be simply a product of high unemployment areas, with higher absolute numbers but similar proportions?
  - And / or artefact of the behaviour of the benefit staff in particular areas?

## Some challenges facing surveys

- Sampling frames possible and legal framework?
- Response rates – down 20-30 percentage points or more in 30 years
  - Differences between different surveys / subject
- Response burden
- Ambition in design (related)
- Lack of good shared evidence about what matters, how much, and cost trade-offs of solutions

## Data design ‘eco-system’ – our social structure

- ‘Analysts’ who use sophisticated methods usually in universities or independent research institutes
- ‘Collectors’ mainly located in some parts of government (ONS), ‘agencies’ (social research departments of market research), or other independent research institutes
- Links between them often restricted by tendering rules
- Academic funders set broad parameters without always brokering relationships
- Government funders often have more substantive expertise but often decide data design behind closed doors

## Implications/ consequences I:

- Respondent burden grows with multi-purpose work and data collection / design sometimes lacks clear sense of priorities
- Sampling issues – span data collection issues and legal gateways – often addressed late in the day, with no early brokerage
- Tendering model often locks out collection / design expertise until late in process
- Too few scale experiments in different ways of sampling, different designs, use of behavioural insights in contact to improve response



## Implications/ consequences II:

- Lack of budget for experiments in how much response rates matter:
  - For different topics
  - For different subgroups
  - For different analytic questions
  - Often find out only post hoc
- Commercial impediments to sharing data
- Lack of a 'total research design' model, lack of iteration

## What can we do? I:

- Funders and others can play a role in brokering early (ideally continuing) and open conversations:
  - about both general issues and those specific to certain surveys, topics, involving analysts and collectors
  - re sampling issues with ICO, legal experts – probably study specific
- Funders and others can ensure budgets for:
  - experiments with MULTIPLE contracts for different designs before academic contracts let by tender
  - experiments in contact (not pilots of content) in early stages of work on projects

## What can we do? II:

- SRA and learned societies and ESRC could deepen brokering roles, help convene seminars.
  - SRA arguably occupies critical position
- Government researchers to resume role (when evaluations return....) in opening up design conversations, brokering sampling
  - GSR role?
- Academics and 'collectors' need more regular joint work outside particular surveys.
- Programme of collaboration on papers on methodology?
  - Budgets and time for collectors for publication of papers with empirical evidence

## Data access:

- ‘Analysts’ and ‘collectors’ to assemble case studies from their own or others’ work to show why this matters
- Engagement with Lords in January but more broadly with policy makers and others in government
- Engagement with work of Wellcome Trust taskforce to link health and social data
- [wellcome.ac.uk/news/independent-patient-data-taskforce-announced](http://wellcome.ac.uk/news/independent-patient-data-taskforce-announced) )  
Speak up for the ‘Five Safes’ model as bridge of data sharing/ privacy debate
- Seek to conduct EMPIRICAL ‘Motivated intruder tests’

## The 'Five Safes' and social consent

- Safe projects – lay involvement in judging public benefit
- Safe people – approved public benefit researchers
- Safe settings – data security
- Safe outputs – for identifiability
- Safe data – appropriate and necessary but no longer sole focus

## **Ultimately... are we....**

- Different tribes?
- Different sects?
- Or neighbouring communities?

## **Contact**

**Sharon Witherspoon FAcSS**  
**Head of Policy**  
**Academy of Social Sciences**  
[policy@acss.org.uk](mailto:policy@acss.org.uk)