

# Developing a Statistical Disclosure Standard for Europe

Tanvi Desai

LSE Research Laboratory Data Manager

## Why?

- Not my project

BUT...

- My area of research
- Proof read and commented on paper
- Useful information for IASSIST as archives were not involved in project
- Upcoming DwB project framed before ESSNet reported

# Cross National Collaboration: Where are we now?

- No data available
- Data only available to nationals
- Data only available in collaboration with nationals
- Data only available within national boundaries
- Data available but only as anonymised files

# Barriers to cooperation

## Varying international standards

- Data security
- Statistical Disclosure Control [SDC]
- Attitude to dissemination
- Available resources
- (Legislation – less important see Ritchie)

➔ barriers to data sharing and trust between countries

# SDC Standards

- Common standards for confidential outputs improves trust and facilitates data sharing
- Agreed standards, must be flexible
  - Unpredictable range of outputs
  - National variations
- ESSNet project developed
  - Full model
  - 'Rule of thumb' model

# Full SDC model

- ‘Principles-based model’
- Classification of outputs into safe/unsafe
  - based upon type of analysis, not data
- NSI staff trained in flexible models
  - Nothing ruled in or out, explicitly
- Researchers are also trained

However

- Requires good understanding of data and statistics
- Requires devolution of responsibility
- Cost of training researchers (and staff)

## Safe v. Unsafe

- All output checked, whether safe or unsafe
- SAFE eg regression coefficients
  - will normally be released, unless NSI takes an active decision not to release
- UNSAFE eg tables
  - will not be released unless researcher demonstrates to NSI why output is safe
- No unconditional yes/no

## ‘Rule of thumb’

- Is a set of ‘hard and fast’ rules that can be applied automatically
- Is useful for
  - Naïve researchers
  - Inexperienced NSIs
  - Automated SDC

However

- It is necessary to set high fixed thresholds
  - Loss of data utility
  - Less cooperative relationship with researchers
  - Not risk free



# Examples

	Rule of thumb	Principles-based
Regression coefficients	Release	Release
Tables	Release if >10 unweighted units in each cell	Release if researcher demonstrates safety
Minima & maxima	Don't release	Release if non-disclosive

# Summary

- Principles-based model

- change in thinking
- some initial effort
- researcher training
- but generates efficient, safer clearances

Safe and efficient, requires effort

- Rules of thumb

- easily implemented
- protects confidentiality in most circumstances
- can stop if not confident of PBM
- but is mechanical

Safe (ish) and easy, but inefficient

# Future

European agreement on a Statistical Disclosure Control Standard will provide a framework for data sharing.

# Now

We only need to work out how and who pays!!!

# Thankyou

Tanvi Desai [t.desai@lse.ac.uk](mailto:t.desai@lse.ac.uk)

Thanks to Felix [felix.ritchie@ons.gov.uk](mailto:felix.ritchie@ons.gov.uk)

Guidelines for the checking of output based on microdata research

[http://neon.vb.cbs.nl/casc/ESSnet/guidelines\\_on\\_outputchecking.pdf](http://neon.vb.cbs.nl/casc/ESSnet/guidelines_on_outputchecking.pdf)