Microdata access: where we are? Where we need to go?

Elizabeth Green and Felix Ritchie

DRAGoN

University of the West of England

The workshop

- a virtual 5-day expert workshop
- 130 attendees across 88 different organisations from 26 different countries

Workshop sessions

Statistical Disclosure Control

- Input SDC
- Output SDC
- Synthetic data

Organisation Management

- Access arrangements
- FAIR data/ data stewardship
- User/ staff training

Technology

- Research data centres
- Remote job servers
- Other technologies

Societal context

- Rules and regulation
- Public engagement
- Ethics, benefits and costs

Questions

1. Is there a consensus on good practice?

2. What lessons have we learnt (i.e. things not to do)? In particular, what did we learned from having to respond to Covid-19

3. Are current practices sustainable (what happens if demand increases) and affordable?

4. What are the lessons for implementation in LMICs?

5. What are the lessons for international data sharing?

6. What are the other main challenges for the next 10 years?

Goodbye scientific use files; hello synthetic data?

Scientific use files will eventually become obsolete with synthetic data sets taking their place.

Synthetic data could provide an opportunity for researchers to test, develop and finalize code before sending it to a secure environment to be executed. And as a tool for training and top-level insights, there is enormous potential for synthetic data sets.

Concerns surrounding the loss of detail as synthetic data may wash out findings and trends in smaller populations, resulting in uneasiness that data sets could become ethnocentric focusing on white populations and losing details for marginalized populations.

The use of synthetic data for decision and policymaking should be approached with caution.

Co-creation of community governance models

Successful data governance models are developed in tandem with public and data users. This is driven by the recognition of the need for better public engagement and public understanding of how microdata is being accessed and used, as well as tailoring process to user needs (as users ignoring rules is a key risk).

The issue of data colonization and the need to protect against exploitation is one of concern and the issue of indigenous data governance and sovereignty. HIC/LMIC co-development of training materials for data governance shows that the community model has high transferability, and supports developing capacity and ensuring microdata is retained in the country of origin.

Rise of the machines

With a lack of resources, we are encroaching towards a situation in which technological advances outpace current knowledge and practice of disclosure control.

With the rise in machine learning, reverse engineering, and AI models there will be new concerns for output attacks and training in these models. However, these may also present an opportunity to assess risk better by mimicking real-life attacks.

Sustaining momentum

Covid has acted as a catalyst for action (overriding the typical defensive stance), and advances to data access practices can be attributed to this; but how do we maintain momentum in normal times?

With previous natural disasters, an influx of action and transformation can be seen, partly due to the necessity to meet demand and partly due to extra resource provisions made available. Once normality begins to resume do the old processes and behaviours resume as well?

The future

- Seminar series discussing general topics (Jan/ Feb 22)
  - What is an RDC?
  - What is a remote job server?
  - User training models
- Conference on Disclosure control (Summer 22)
  - Cover wide range of different topics including; Qualitative data, machine learning

The future...

- Network for trainers to share info/good practices/catalogues/ global overviews and link to resources
- Network to consider ethical issues
- Centralised group/ website to share info? Wiki/ linkedin/ launch event to help takeup?
- Software consistency and advice
- Software solution to help manage metadata and dataflow process
- Access to training- how do we do it? Support network/ mentors
- Process for technical workers/coders/developers to discuss/share

Let's stay connected

Elizabeth7.green@uwe.ac.uk

Felix.Ritchie@uwe.ac.uk