

**UWE
Bristol**

University
of the
West of
England

DRAGoN

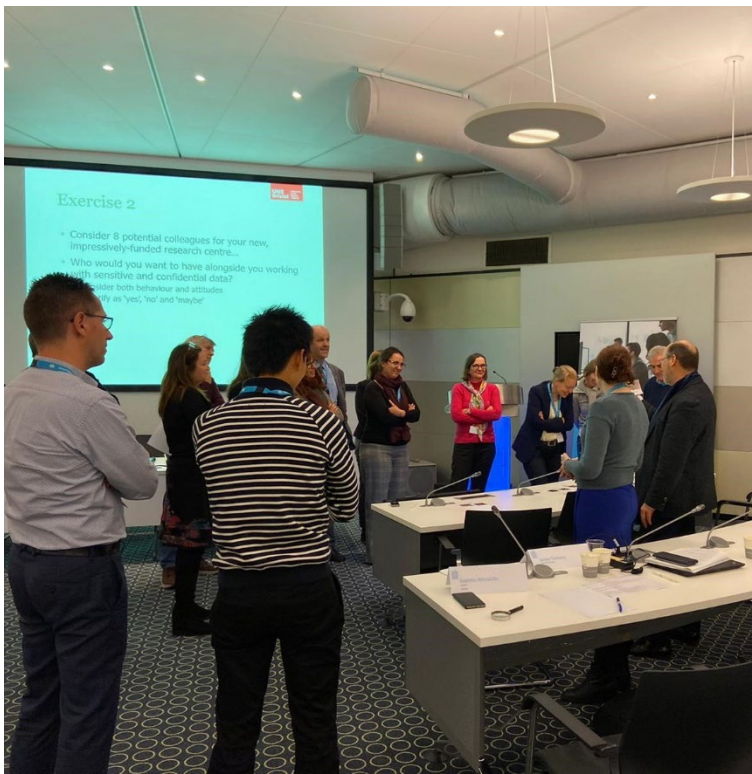
Data Research, Access,
and Governance Network

*What do we know
about good practices
in microdata access
what will be the
challenges for the
coming decade?*

A week long
practitioners
workshop to consider
where we are and
where we are going

**The present
and future of
confidential
microdata
access**

5th- 9th July 2021



WHO SHOULD ATTEND?

This will be a highly interactive workshop. We want contributions from all participants in the microdata access spectrum: those who are managing microdata access, those who are thinking about this or who take responsibility, researchers, analysts and other users of confidential data.

From **Practitioners** we want contributions on good practices and gaps

From **Users** we want views as to how systems work and how they might be improved

From **Planners and Commissioners** we want your input 'How does this help me?'

The workshop will predominately focus on government data access, but we welcome **contributions** from all sectors.

We will prepare in advance a **briefing document** covering all topics to provide all attendees with an overview of the current state of play.

We don't expect everyone to attend every session; simply **register** for the ones you are interested in.

If you are interested in **facilitating** or **chairing** a session- please get in touch with the organisers.



WORKSHOP CONTENT

2x 90 minute sessions per day, timed for Eastern and Western hemispheres. Each session will have:

- General topic introduction
- Breakout sessions
 - Introductory presentation: state of play, gaps
 - Facilitated discussion
- Report back to main session

A summary will be drafted for each session by the workshop organisers.

Breakout sessions will consider

- Best/ Good practice
- Areas for development
- Sustainability and Affordability
- Implementation in LMICs
- International data sharing

SESSIONS

TECHNOLOGY

3 sessions discussing; Research data centres, Remote job servers/table servers, Other tech solutions

STATISTICAL DISCLOSURE CONTROL

3 sessions discussing; Input SDC, output SDC, synthetic data

ORGANISATION MANAGEMENT

4 sessions discussing; User/staff training, Access arrangements, FAIR data/data stewardship, Organisational/ structural issues

SOCIETAL CONTEXT

3 sessions discussing Regulatory regimes, Public engagement, Ethics/benefits and costs



PRACTICALITIES

This will be an online workshop hosted on Microsoft Teams. There is no fee for attendance, but registration is required and available at <https://uwedragon.weebly.com/conferences.html> The timings for the week are as follows (a calendar invite will be sent and should adjust to your timezone)

	Monday 5th July 2021	Tuesday 6th July 2021	Wednesday 7th July 2021	Thursday 8th July 2021	Friday 9th July 2021
Eastern 08:00-09:30 London 09:00-10:30 Brussels 17:00-18:30 Canberra	Technology 1. Research data centres 2. Remote job servers/table servers 3. Other tech solutions	Statistical Disclosure Control 1. Input SDC 2. Output SDC 3. Synthetic data	Organisation 1. User/staff training 2. Access arrangements 3. FAIR data/data stewardship 4. Organisational/structural issues	Societal context 1. Regulatory regimes 2. Public engagement 3. Ethics/benefits and costs	Summary and looking forward
Western 16:00-17:30 London 17:00-18:30 Brussels 12:00- 13:30 Ottawa	Organisation 1. User/staff training 2. Access arrangements 3. FAIR data/data stewardship 4. Organisational/structural issues	Societal context 1. Regulatory regimes 2. Public engagement 3. Ethics/benefits and costs	Technology 1. Research data centres 2. Remote job servers/table servers 3. Other tech solutions	Statistical Disclosure Control 1. Input SDC 2. Output SDC 3. Synthetic data	Summary and looking forward

Topic	Details	Chairs (East/West)
Technology		
RDCs	<p>Research data centres, particularly remote, have been the great success story of the 21st century, allowing detailed access to highly sensitive data in secure but usable environments. Can we identify good practices? What lessons have been learned about what works, and what doesn't? How has homeworking changed the role of RDCs? Where do we see RDCs developing? Is this an appropriate technology for small government agencies or lower income countries?</p>	tbc
Table and remote job servers	<p>Remote job servers have been relatively rare as solutions to data access. Table servers are more common, and recent developments such as cell-key anonymisation increase security and flexibility. What limits development, and what makes a good RJS or table server? What user groups do they work best for? Is this more or less suitable for data holders with limited funding? Are there off-the-shelf solutions?</p>	Heba Omar, Data Innovation Coordinator and Research Associate LISSY
Other technological developments	<p>A range of technical solutions have been proposed for sharing access to data whilst protecting confidentiality: distributed processing, homomorphic encryption, etc. Which problems are they solving, compared to more traditional solutions such as anonymisation or access control? Which technologies are closest to widespread application? Or are these likely to remain a niche solution?</p>	tbc
SDC		
Input SDC	<p>Anonymisation and detail reduction has been the bedrock of statistical data management for fifty years. There is a huge canon of work. Are there big areas still to be developed? What are the risks anonymisation and de-identification to increasing computer power, administrative and social media data, or machine learning models?</p>	Josep Domingo-Ferrer, Universitat Rovira i Virgili
Output SDC	<p>There is a well established set of theories and techniques for producing tables of official statistics, but checking for residual risk in analytical outputs is much less well understood. Is 'output SDC' more of a statistical or operational concern? Has implementation has moved ahead of theory? What are the conceptual gaps? How are output checkers trained?</p>	Eric Schulte Nordholt, Statistics Netherlands

Topic	Details	Chairs (East/West)
Synthetic data	Synthetic data offers the possibility of increasing accessibility of data while reducing risk, at the cost of reduced utility. General-purpose open-source software tools can support both expert and non-expert data creation. Can we synthesise any data including longitudinal or highly skewed data? What are the gains for TRE users? How do we retain value in direct use? Where are the main research gaps? Is ML the future of synthetic data creation?	Gillian Raab, University of Edinburgh
Organisation		
Training	Training of users and data stewards in good data management practices vary widely. What training is necessary, and for whom? What are the costs and benefits of training staff and users? What is good practice in training? Is training developed in high-income countries suitable for LMICs?	John Newman, Australian Bureau of Statistics
Access arrangements	Many data providers release data in multiple ways – under license, in controlled facilities, unrestricted on the internet. Licensing dominates usage of confidential data, particularly for academic use. Is this expected to continue, and why? What determines an effective licensing model? Do data distributors have a role to play in educating users about good data management or use?	Steve McEachern, Australian Data Archive Matthew Woolard, UK Data Archive
FAIR data, data stewardship, sustainability and reproducibility	Sustainable stewardship is becoming increasingly important. How do data holders ensure that their data is findable and accessible? How can reproducibility in research be encouraged? What makes a usable, sustainable data resource? What role does metadata play in supporting data use? What metadata standards are missing?	Stefan Bender, Deutsches Bundesbank & INEXDA
Societal context		
Regulatory regimes	Legislation and regulation has traditionally been based on simple ideas of anonymous/identifiable data. More recent models view data protection as a portfolio model. Legislation and regulation are both moving towards the principles-plus accreditation model, where statutory instruments specify outcomes, and accreditation is used to meet those outcomes. What are the pros and cons of these new approaches? How does this shift responsibility for data management? Does this simplify or complicate sharing between regulatory regimes? What lessons can be drawn for countries developing data legislation?	Pete Stokes, Office for National Statistics Maggie Levenstein, ICPSR

Public engagement	Much of the academic literature concerning public attitudes towards data sharing comes from research regarding people's healthcare data. It's clear from the literature that people have an intuitive wariness of institutions that want to use and share their data. People are less wary of the healthcare industry and police, and people are more wary of Tech companies, Insurance companies and the media. How do we encourage public engagement? What examples are there of (good or bad) public engagement? How does public engagement differ between Low Middle Income Countries and High Income Countries (HIC)?	Tbc
Ethics, benefits and costs	Public good, data sharing between HIC and LMIC, autonomy, opt-out provisions, third party use of data, role of ethics committees,	Lynne Woolfrey, DataFirst University of Cape Town
Summary	In the concluding session, we will review the individual session reports. In breakout groups we will invite views on the findings, and on next steps. The second summary session (Western hemisphere) will also reflect on the conclusions of the first summary session (Eastern hemisphere).	Taeke Anton Gjaltema, UN Economic Commission for Europe

Organising committee

Elizabeth Green, Head of organising committee

Felix Ritchie, UWE

Azeem Haroon, UWE

Juan Carlos Mondragon Quintana, UWE

For enquires email Elizabeth7.green@uwe.ac.uk

Advisory committee

Stefan Bender, Deutsche Bundesbank and INEXDA

Aleksandra Bujnowska, Eurostat

Taeke Gjalterna, UN Economic Commission for Europe

Adam Harris, Australian Bureau of Statistics

Tina Hotton, Statistics Canada

Wim Kloek, Eurostat

Steve McEachern, Australian Data Archive

Eric Schulte Nordholt, Statistics Netherlands

Pete Stokes, UK Office for National Statistics

Steven Thomas, Statistics Canada

Lynn Woolfrey, DataFirst South Africa

Scan/ click below for
DRAGoN website



Scan/ click below for
conference registration
page

