

International Journal of Population Data Science

Journal Website: www.ijpds.org



Live research dialogue on the benefits, costs and utility of synthetic data for researchers

Emily Oliver¹, Fiona Lugg-Widger², Maureen Haaker³, Cristina Magder³, and Emma Gordon¹

¹ADR UK

²Cardiff University

³University of Essex

Objectives

- To explore use cases and characteristics of synthetic data that make it useful for research.
- To discuss governance frameworks essential for the routine creation, dissemination, and use of synthetic data.
- Drawing on the experience of international participants, explore measures to mitigate risks to data privacy and public perception relating to synthetic data creation and use whilst maximising its utility.

Approach

Presenters provided context and insights on definitions and interpretations of synthetic data, including debates about fidelity and disclosure risk; known benefits of synthetic data and use cases; and its costs and challenges from the perspective of data owners, data providers and the public. This informed discussions in breakout groups aligned to the objectives. Participants self-identified as data service providers/processors, academic researchers, data owners, and/or 'other'.

Results

- Characteristics identified as important across a variety of use cases included accessibility, structure and format to match the real data, and documentation.
- The most popular measure to mitigate risks was clear and detailed documentation (metadata, codebooks, user guides, limitations, creation methods etc).
- Training was identified as an important benefit and use case of synthetic data across the participant groups.
- The most popular challenge identified by data service providers/processors and 'other' was lack of governance and standards; whilst for researchers, this was the verification of the synthetic data and the potential for it to be incorrectly interpreted.

Conclusions

There is a demand across stakeholder groups for synthetic data for a range of uses. Synthetic datasets should be accompanied by clear and detailed documentation and provided within an agreed governance framework.

