

International Journal of Population Data Science

Journal Website: www.ijpds.org



Public Perceptions of and Best Practices for Synthetic Data Use in Research

Rob Trubey¹, Claire Nollett¹, Joshua Stock¹, Jim Fitzgibbon², Farheen Yameen², Lucy Brookes-Howell¹, Mike Robling^{1,3}, and Fiona Lugg-Widger¹

¹Centre for Trials Research, Cardiff University, Cardiff, United Kingdom

²Public Member, Cardiff, United Kingdom

³Decipher, Cardiff University, Cardiff, United Kingdom

This project aimed to establish recommendations for data owners on the production, release, and public communication of synthetic data. By engaging members of the UK public in facilitated discussions, we explored perceptions of synthetic data use in research and identified strategies to enhance transparency, trust, and responsible data management.

We worked with community engagement specialists to recruit a diverse group of 39 public members from across the UK. Members took part in four deliberative workshops, held online, to co-develop recommendations for data owners. They heard from subject matter experts and engaged in structured discussions on synthetic data, its applications, and ethical considerations. Key concerns and priorities were identified through analysis of workshop notes, with public feedback incorporated iteratively throughout the process to ensure recommendations aligned with public expectations. The workshops facilitated an in-depth exploration of transparency, accessibility, and governance surrounding synthetic data, highlighting the importance of clear public communication.

Ten key recommendations were developed, categorised into five thematic areas: (1) Introducing synthetic data and its distinction from real data; (2) Clarifying its purpose and role in research; (3) Guidelines for dataset creation and validation; (4) Managing access, usage, and potential misuse; and (5) Strategies for effective public communication and trust-building. Public members emphasised the need for accessible explanations, governance measures, and ongoing engagement to ensure ethical use. Iterative refinement of recommendations in the final workshop highlighted public concerns about data misuse and the importance of organisation's accountability. These findings provide a framework for responsible synthetic data creation and improved public engagement strategies.

This project provides valuable insight into public attitudes towards synthetic data in research, emphasising transparency and trust. The recommendations offer guidance for UK data owners in responsibly managing synthetic datasets. Ongoing research is needed to monitor evolving public perspectives as synthetic data use expands across research domains.

