

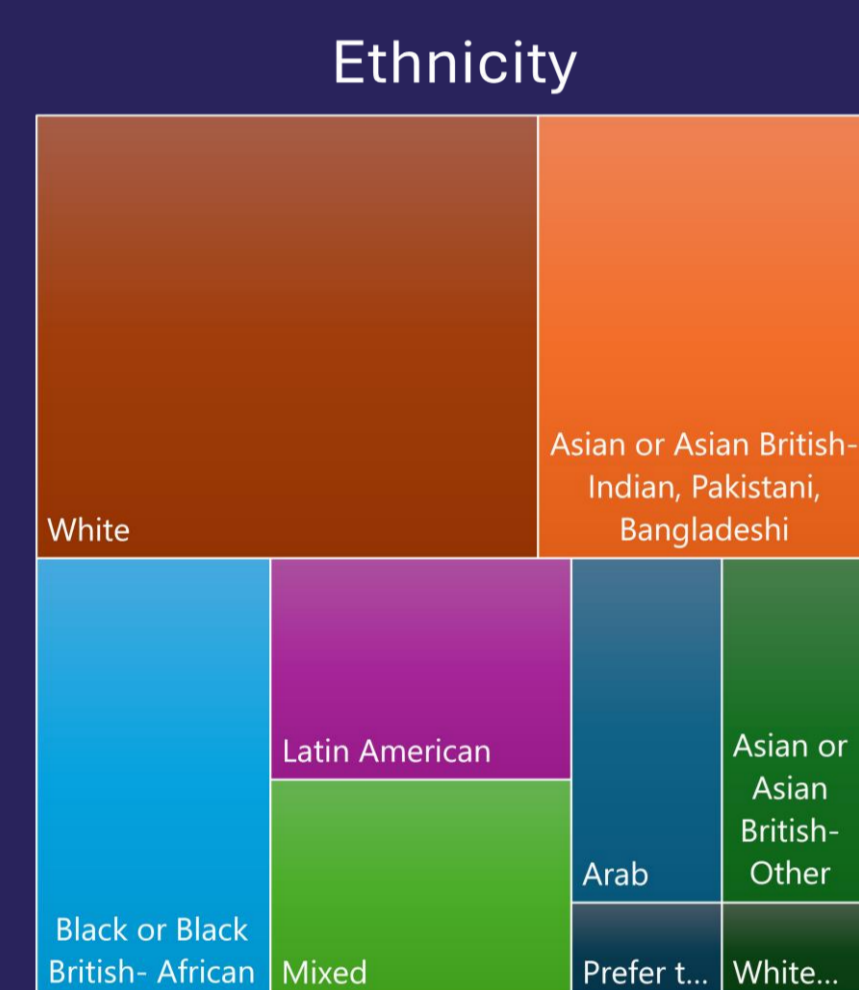
Exploring public views on the use of synthetic datasets for research

Fiona Lugg-Widger¹, Claire Nollett¹, Joshua Stock¹, Jim Fitzgibbon², Farheen Yameen², Lucy Brookes-Howell¹, Mike Robling^{1,3}, Rob Trubey¹
¹Centre for Trials Research, Cardiff University, Cardiff, UK. ²Public Contributors. ³DECIPHER, Cardiff University, Cardiff, UK.

Synthetic data is increasingly being generated and made available as an efficient way for researchers to learn about the properties of datasets and inform data request applications.

Little is understood about how the UK public feel about the creation and use of such datasets for research purposes.

We worked with a community engagement agency (Egality) to recruit a diverse UK public panel, to join a series of workshops to discuss the use of synthetic data in research.



138 individuals expressed an interest

44 were invited to contribute to the workshops

Attendees represented a diversity of views and experiences from all four nations of the UK. We were especially keen to ensure a diversity in ethnicity, age, working status and include individuals from all four nations of the UK.

What we've heard so far:

The use of synthetic data to support researchers was generally accepted

"what's the point?"

Requests for real life examples of when and how synthetic data has been used

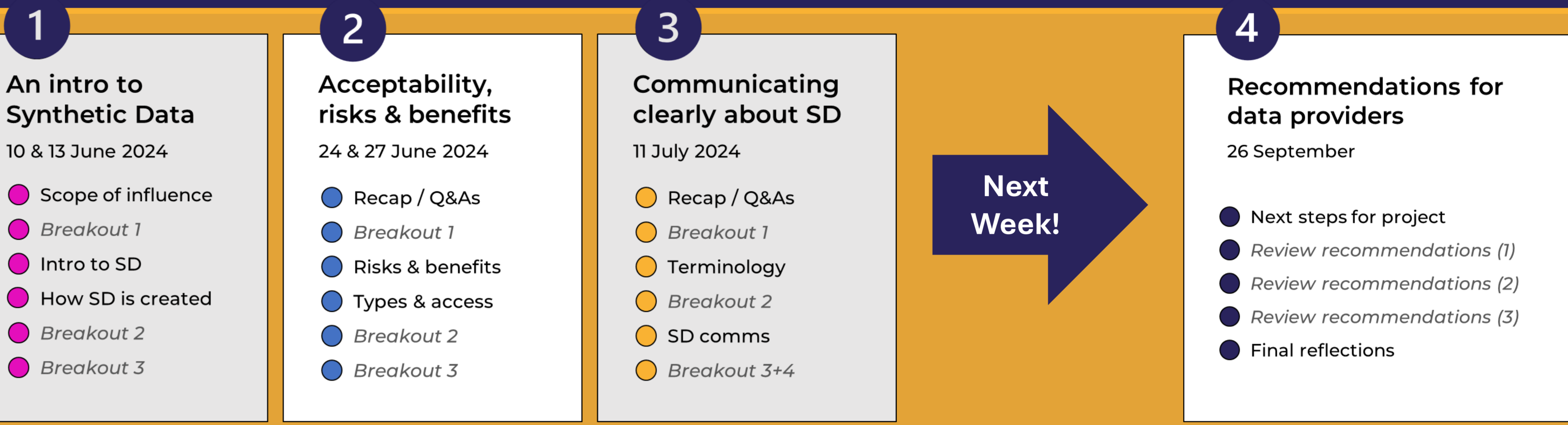
Misuse – what that looked like and how it could be monitored was a recurring query

Data checks and privacy related concerns were raised

Inaccuracies and bias of the synthetic data were discussed at length

These will form a set of recommendations for researchers, data custodians and policy makers.

As with any research that seeks to draw on the public's data, it is vital that researchers and policy-makers engage with and involve the public in any decision making around synthetic data.



Across four workshops, participants listened to presentations from topic experts, and deliberated on perceived benefits and risks of synthetic data, ethical issues and reflections on how to communicate clearly about synthetic data.

- Mirrored Data
- Virtual Data
- Dummy Data
- Artificial Data
- Generated Data
- Test Data
- Mock Data
- Simulated Data
- Synthetic Data
- Fake Data
- Example Data
- Imitation Data

We heard that the term synthetic data is not well understood and does not mean a lot to the public.

An alternative word could not be agreed on but we summarise here the words suggested throughout the workshops.

We recommend choosing words with minimal negative connotations and which sound credible (e.g. not 'fake data').



See our study page for outputs and a PDF of this poster!



PRINTED BY **SCIENCEPOSTERS**.co.uk