

MAKING SYNTHETIC DATA WORK FOR EVERYONE

Practical recommendations for data owners to responsibly share low-fidelity synthetic data in alignment with public attitudes and expectations.

Recommendations

Introducing synthetic data



RECOMMENDATION 1

The term synthetic data is not well understood by the public. Provide a brief definition to explain:

- that synthetic data is not real data
- but that it is based on real data
- that it is created in a way that minimises personal privacy risks

Explaining the purpose of your synthetic data



RECOMMENDATION 2

Make it clear what your synthetic dataset can and cannot be used for e.g. it can be used for training, or writing computer code, but not to answer research questions. Emphasise that real policy decisions are always made with real data.

RECOMMENDATION 3

Explain the benefits and impact of your synthetic dataset for:

- your own organisation
- researchers e.g. training and understanding the data
- the public e.g. if public money is being spent to create it, how do the public benefit?

RECOMMENDATION 4

Explain the personal privacy benefits that using synthetic data offers.

Accountability for creating your synthetic data



RECOMMENDATION 5

Provide a simple explanation for how your synthetic data is created. In particular, you should explain the role of humans vs automation (such as Artificial Intelligence) in the process.

RECOMMENDATION 6

Human oversight in checking personal privacy risks is important. Explain the quality and privacy checks you undertake before your synthetic dataset is released.

Access, use and misuse of synthetic data



RECOMMENDATION 7

There is not widespread support for a fully open access approach to synthetic data:

- use a simple registration process which records the requestor's name, email address and intended use (as a minimum).
- implement a simple user agreement covering the key terms and conditions such as allowed usage and how long synthetic data can be held.

Communicating synthetic data clearly



RECOMMENDATION 8

Real life examples are particularly helpful to the public. Use accessible case studies from researchers to:

- demonstrate what synthetic data is
- report the outcomes from researchers using your synthetic data
- emphasise the positive impact for the public

RECOMMENDATION 9

Use creative communication methods including infographics and engaging videos to convey information about synthetic data to the public.

RECOMMENDATION 10

Work with the public to ensure all of this information is accessible to people with a diverse range of needs.

These recommendations were co-produced with public input through a UK-wide consultation involving workshops with 39 public members. They reflect public perspectives on low-fidelity synthetic data governance, privacy, and utility.

More information and access to the full report is available through the project website.

[Synthetic Data Infographic](#) © 2025 by [Discussing Data Project at Cardiff University](#) is licensed under [CC BY-SA 4.0](#)

