

SYNTHETIC DATA

Public sector organisations, like government departments, are responsible for managing people's health, education and other types of important data. To help researchers, these organisations sometimes create 'synthetic' versions of real data.



But what exactly is synthetic data, and why is it useful?

What is it?

Synthetic data is created to reflect real data without using real people's information



It is made by analysing patterns in real data and creating new data that follows the same trends – but without including any actual individuals.

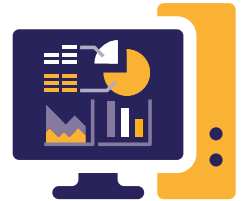
It looks like real data but keeps personal information safe



For example, a synthetic version of hospital visits might look similar to the real data but would not contain any real patients. This makes it safer to use.

Why is it useful?

Researchers can start working with data while waiting for access to real datasets



When researchers request access to real data, their applications must be carefully reviewed by public sector organisations. This can take time, but sharing synthetic data allows researchers to begin their work sooner.

It helps researchers test their methods and improve their studies



Researchers can use synthetic data for training, testing code, and developing research plans. This makes them better prepared when they receive the real data, leading to more effective research that benefits the public.

This infographic focuses on "low fidelity" synthetic data, developed as part of the Discussing Data project with 39 members of the UK public. We thank all those involved for their time and contributions.

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

