

International Journal of Population Data Science



Swansea University
Prifysgol Abertawe

Journal Website: www.ijpds.org

Evaluating the Benefits, Costs, and Utility of Synthetic Data for Data Owners and Trusted Research Environments

Melissa Ogwayo¹, Cristina Madger¹, Hina Zahid¹, and Maureen Haaker¹

¹University of Essex, Colchester, United Kingdom

Objectives

This study evaluates low-fidelity synthetic data's benefits, costs, and utility for data owners and Trusted Research Environments (TREs). It examines financial implications, operational efficiencies, and governance challenges, informing best practices for scalable and ethical synthetic data production. The findings will provide actionable insights for the entire research ecosystem.

Methods

A mixed-methods approach evaluated synthetic data adoption among data owners and TREs. A literature review synthesized best practices, ethical considerations, and technical challenges. A survey assessed data owners' perceptions, readiness, and financial concerns. Using semi-structured interviews, policy analysis, and standard operating procedures (SOPs), case studies were conducted with existing synthetic data producers to examine frameworks, governance, and cost structures. Focus groups with TRE representatives explored operational challenges, security risks, and policy gaps. This research was a jointly funded initiative by the Economic and Social Research Council (ESRC) Data & Infrastructure Programme and ADR UK, conducted by UK Data Service.

Results

When submitting this proposal, the study was ongoing and due to conclude in April 2025. Preliminary findings highlight that financial constraints currently hinder synthetic data production. Limited dedicated funding leads organisations to explore collaborations and process optimisation.

Key challenges include data quality assurance, regulatory compliance, and the lack of dedicated training. Improved data access shows efficiency gains, but legal uncertainties slow adoption. Organisations vary in synthetic data sharing and licensing approaches, balancing openness with (perceived) risk management.

Future efforts focus on sustainable funding, standardised governance, automation, and bridging expertise gaps. Addressing public misconceptions and defining set governance

frameworks remain priorities. Organisations continue evaluating the feasibility of synthetic data with a view to expanding its role in research, policy, and innovation.

Conclusion

Synthetic data offers significant potential for secure data sharing and privacy protection in an ever-evolving data landscape. However, governance inconsistencies, financial constraints, and public trust remain key barriers. Standardized policies, improved documentation, and cross-sector collaboration will ensure scalable, ethical, and impactful synthetic data adoption across research and industry.



Temporary page!

\LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because \LaTeX now knows how many pages to expect for this document.