

International Journal of Population Data Science

Journal Website: www.ijpds.org



Federating governance, access and infrastructure to support researcher use of synthetic data

Katherine O'Sullivan¹, Jackie Caldwell², Christian Cole³, Kathy Harrison⁴, Charlie Mayor⁵, Antonietta Chaliou¹, Jaroslaw Dymiter¹, Stacey Dawson¹, Diane Brown¹, and Katie Wilde¹

¹University of Aberdeen

²Public Health Scotland

³University of Dundee

⁴University of Edinburgh

⁵NHS Greater Glasgow & Clyde

Synthetic Data has the potential to improve efficiency of data analysis for researchers. However, there is no standard approach to synthetic data governance, access controls or infrastructure requirements, and researchers may face inconsistencies in how they can access or use synthetic data across trusted research environments. We present a federated solution taken by the Scottish Safe Haven Network to address these barriers to facilitate researcher use of synthetic data.

We documented and evaluated existing governance pathways, access controls and infrastructure design for non-synthetic data across the Network, recognising uniformity and establishing equivalence using the 5 Safes framework, ISO27001 standards and the SATRE TRE specification. We also interviewed current and potential researchers using our trusted research environments to identify common use cases for accessing synthetic data. We then mapped researcher requirements against the documented equivalencies, validating with current and prospective users. We identified several use cases: to undertake feasibility studies, to understand dataset structure and format and to write analysis code whilst waiting on the project-specific data to be provided. By mapping the use cases onto existing governance and access processes and infrastructure designs, we were able to agree to a standard application process, access control mechanism, and infrastructure platform across the Network to provide a consistent process for researchers. A federated approach to synthetic data access will improve the speed at which research can be conducted as well as improving the transparency and consistency of data governance and access across organisations, ultimately improving the experience for researchers using TREs.



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.