

International Journal of Population Data Science

Journal Website: www.ijpds.org



Understanding the challenges researchers face in accessing sensitive data across multiple trusted research environments in one federated study

Jano van Hemert¹, Kostas Kavoussanakis¹, Kara Moraw¹, Chris Wood¹, Ioanna Lampaki¹, Roger Halliday², Ryan Bremner², Jennifer Muir², Emily Symmons³, and Sophie Elfar⁴

¹University of Edinburgh, Edinburgh, United Kingdom

²Research Data Scotland, Edinburgh, United Kingdom

³Office for National Statistics, Titchfield, United Kingdom

⁴Office for National Statistics, Newport, United Kingdom

Accredited UK researchers can access anonymised data about UK citizens through dozens of Trusted Research Environments (TRE). Each TRE specialises on national or regional data and specific disciplines such as social, health and administrative services. We investigated the challenges researchers face when they want to access data linked across these specialised TREs. We designed and ran a study on sensitive data from two national TREs: Public Health Scotland's National Safe Haven and Office for National Statistics' Integrated Data Service. We recorded process steps and identified issues in and discrepancies between TRE processes from start to finish. We recorded issues and discrepancies in the following aspects of the service and information governance models: lack of in-depth metadata, divergence in requirements for approval of the study,

approval serialisation, divergence and duplication in accreditation processes for the researchers, divergence in charging models and incompatible agreements with data owners. As a result, our study suffered many delays. The most notable cause was time needed to agree a plan for data linkage and integration. This plan had to satisfy simultaneously the purpose of the study, the data access decision makers of both TREs and the technical feasibility of data movement and integration. To enable research on federated data, researchers will need publicly available metadata that provide non-disclosive statistical and quality properties of sensitive data. To reduce time and effort to gain federated access, TREs will need to align and unify their processes, most importantly to facilitate cross-TRE linkage processes compatible with data owner requirements.

