International Journal of Population Data Science

Journal Website: www.ijpds.org

Secure data analysis environments: can we agree on criteria for "Appropriate secure access" to linked health data?

Jorm, LR[1], McGrail, K[2], Victor, JC[3], Jones, K[4], Ford, D[4], and Churches, T[5]

[1]Centre for Big Data Research in Health, University of New South Wales

[2]University of British Columbia

[3]Institute for Clinical Evaluative Sciences

[4]Swansea University

[5]Ingham Institute for Applied Medical Research and South Western Sydney Clinical School, UNSW Sydney

Overall objectives or goal

Many health data linkage ecosystems across the world have designed and implemented secure data analysis environments as one of their controls to protect patient privacy and confidentiality. These have been shaped by local legislation and data governance policies, available IT infrastructure and resources, and the skills and imagination of their architects. However, at present their various features and functionalities have not been reviewed, synthesised or contrasted. Burton et al [1] have proposed 12 criteria for Data Safe Havens in health and healthcare, which they conceptualise broadly as encompassing data governance and ethics, quality and curation of data repositories, and data security. Under this definition, secure analysis environments, which may or may not be integrated with data repositories, are a component of a Data Safe Haven, addressing the criterion "Appropriate secure access to individually identifying data". To guide those building and operating these environments, and data custodians and stewards who need to assess their fitness-for-purpose, it would be of great value to discuss and agree an aggregate term (e.g. "Secure Data Lab") that describes them, and to develop a more detailed set of criteria for what entails "Appropriate secure access" to linked health data.

The goal of this session is to describe and document the approaches that have been taken by flagship secure data analysis environments internationally, including their approaches to authentication, assigning permissions, managing the ingress and egress of files and auditing transactions, and their responses to emerging opportunities, including cloud computing and national and international data sharing. We will explore how the interplay of physical, technical and procedural controls have been combined to create existing models, and the extent to which these can balance each other and be applied with flexibility depending on perceived risk and regimes.

Session structure

Prior to the session, we will develop a draft set of criteria for "Appropriate secure access" to linked health data. The session will comprise presentations describing existing secure analysis environments against the draft criteria, followed by a facilitated discussion. The secure data analysis environments that will be presented include:

- UNSW Sydney E-Research Institutional Cloud Architecture (ERICA)
- PopData BC Secure Research Environment (SRE)
- Institute for Clinical Evaluative Sciences (ICES) Data and Analytic Virtual Environment (IDAVE)
- Secure Anonymised Information Linkage (SAIL) Gateway

Intended output or outcome

We will write up the outcomes of the session as a scientific paper that proposes an aggregate term for secure data analysis environments for linked health data and a set of criteria for what entails "Appropriate secure access" to linked health data.

Presenters and Facilitators

Professor Louisa Jorm, Centre for Big Data Research in Health, UNSW Sydney, Australia

Dr Tim Churches, South Western Sydney Clinical School, UNSW Sydney, Australia

Professor Kim McGrail, Population Data BC, The University of British Columbia, Vancouver, Canada

J. Charles Victor, Institute for Clinical Evaluative Sciences, Toronto, Canada

http://dx.doi.org/10.23889/ijpds.v3i4.836

Dr Kerina Jones, Swansea University Medical School, Wales, United Kingdom

Professor David Ford, Swansea University Medical School, Wales, United Kingdom

1. Burton PR, Murtagh MJ, Boyd A, et al. Data Safe Havens in health research and healthcare. Bioinformatics 2015; 31(20): 3241–3248