International Journal of Population Data Science

Journal Website: www.ijpds.org

Designing and Implementing a Privacy Preserving Record Linkage Protocol

Gee, T¹, Behan, B², Lefaiivre, S², Azimae, M³, Dharsee, M¹, El Emam, K⁴, Yang, J³, Vaccarino, A¹, Evans, K¹, Victor, JC³, and Theriault, E²

¹Indoc Research

²Ontario Brain Institute

³Institute for Clinical Evaluative Sciences

⁴Electronic Health Information Laboratory

Introduction

The Ontario Brain Institute has developed Brain-CODE, an informatics platform, to support the acquisition, storage, management and analysis of multi-modal data. The standardized research data within Brain-CODE spans several brain disorders, allowing for integrative analyses, while also providing the opportunity to leverage existing clinical administrative data holdings through external linkages.

Objectives and Approach

Within Ontario, the majority of individuals who access the healthcare system have a unique identifier, the Ontario Health Insurance Plan (OHIP) number. The OHIP number can facilitate linkages with administrative data holdings, such as those at the Institute for Clinical Evaluative Sciences (ICES). Given that OBI is not permitted under Ontario's privacy legislation to hold OHIP numbers, identifiers for consented participants are encrypted using a public key mechanism upon entry into Brain-CODE, where the private key is inaccessible. To facilitate linkages involving OHIP numbers between Brain-CODE and ICES, Brain-CODE Link software was co-developed by members of the Indoc Consortium.

Results

Brain-CODE Link allows a deterministic linkage between encrypted identifiers (OHIP numbers), without revealing participant identity. The same homomorphic encryption algorithm applied to identifiers upon entry to Brain-CODE, is applied to relevant identifiers within ICES data holdings. Encrypted identifiers from Brain-CODE are securely transferred to ICES, where a comparison computation calculates differences between the encrypted sets. These differences are sent to a semi-trusted third party, who has no access to the original data, to decrypt the differences using the private key. A zero difference indicates a set of matching identifiers. One of the main challenges during testing and development of Brain-CODE Link

was ensuring the software was capable of scaling to a population level, performing a large number of comparisons, in a computationally efficient manner.

Conclusion/Implications

Ongoing pilot projects within the areas of epilepsy, neurodevelopment disorders, and neurodegeneration will be the first examples of linkages between Brain-CODE and ICES. Brain-CODE Link has successfully performed several billion test comparisons, indicating its suitability to function as a scalable privacy preserving record linkage to support comprehensive analyses.