International Journal of Population Data Science

Journal Website: www.ijpds.org

The Ontario Data Safe Haven: Bringing High Performance Computing to Population-wide Data Assets

Victor, JC[1], Paprica, PA[2], Brudno, M[3], Virtanen, C[4], Wodchis, W[3], Goldenberg, A[6], and Schull, M[1]

[1] Institute for Clinical Evaluative Sciences

[2] Vector Institute

[3] University of Toronto

[4] HPC4Health

[5] University of Toronto

[6] The Hospital for Sick Children

Introduction

Canadian provincial health systems have a data advantage – longitudinal population-wide data for publicly funded health services, in many cases going back 20 years or more. With the addition of high performance computing (HPC), these data can serve as the foundation for leading-edge research using machine learning and artificial intelligence.

Objectives and Approach

The Institute for Clinical Evaluative Sciences (ICES) and HPC4Health are creating the Ontario Data Safe Haven (ODSH) – a secure HPC cloud located within the HPC4Health physical environment at the Hospital for Sick Children in Toronto. The ODSH will allow research teams to post, access and analyze individual datasets over which they have authority, and enable linkage to Ontario administrative and other data. To start, the ODSH is focused on creating a private cloud meeting ICES' legislated privacy and security requirements to support HPC-intensive analyses of ICES data. The first ODSH projects are partnerships between ICES scientists and machine learning.

Results

As of March 2018, the technological build of the ODSH was tested and completed and the privacy and security policy framework and documentation were completed. We will present the structure of the ODSH, including the architectural choices made when designing the environment, and planned functionality in the future. We will describe the experience to-date for the very first analysis done using the ODSH: the automatic mining of clinical terminology in primary care electronic medical records using deep neural networks. We will also present the plans for a high-cost user Risk Dashboard program of research, co-designed by ICES scientists and health faculty from the Vector Institute for artificial intelligence, that will make use of the ODSH beginning May 2018.

Conclusion/Implications

Through a partnership of ICES, HPC4Health and the Vector Institute, a secure private cloud ODSH has been created as is starting to be used in leading edge machine learning research studies that make use of Ontario's population-wide data assets.