

Cloud-based Integration of Causal Modeling and Discovery Tools with a Unified Patient Research Data

Jeremy Espino, MD, MS - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Paul Avillach, MD, PhD - Department of Biomedical Informatics, Harvard Medical School, Cambridge, Massachusetts, USA

Michael Davis, MS - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Jeremy Easton-Marks - Department of Biomedical Informatics, Harvard Medical School, Cambridge, Massachusetts, USA

Michael McDuffie MS - Department of Biomedical Informatics, Harvard Medical School, Cambridge, Massachusetts, USA

David Bernick - Department of Biomedical Informatics, Harvard Medical School, Cambridge, Massachusetts, USA

Gregory Cooper, MD, PhD - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Isaac Kohane, MD, PhD - Department of Biomedical Informatics, Harvard Medical School, Cambridge, Massachusetts, USA

Michael Bech, MD, PhD - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Both the NIH Commons and the BD2K initiatives share a common goal of providing a central computational environment for the sharing, reuse, interoperability, and discoverability of data. The University of Pittsburgh Center for Causal Modeling and Discovery of Biomedical Knowledge from Big Data (CCD) and the Harvard University Medical School (HMS) Patient-Centered Informatics Common: Standardized Unification of Research Elements (PIC-SURE) have implemented a framework that provides a model for biomedical researchers to share data and tools in a cloud environment. An instantiation of this framework allows researchers to apply causal discovery algorithms to biomedical and clinical big data to discover new and significant causal relationships. The PIC-SURE team has adapted their tranSMART tool and extended their application-programming interface (API) to provide secure access to the Autism Simons Simplex Collection dataset running on an Amazon Elastic Compute Cloud (Amazon EC2) instance. Through this service, a secure authentication layer, via Auth0, provides an institutional single sign-on which handles requests for access to the data. This data is then readily available for analysis by an Amazon EC2 implementation of the CCD causal web tool and API. CCD causal web provides a similar authentication method using institutional single sign-on for handling access control. Our proof-of-principle integrated data ecosystem serves as a model for sharing and analyzing biomedical big data in a secure and scalable manner in a cloud environment.

Session to which submitted: Collaborations