

Healthcare Data Commons Cloud Use Case

The Problem

The sheer volume and variety of digital data, especially from mobile devices, has exploded over the last few years, creating data lakes. Data privacy and security, and privacy regulations challenge the adoption and success of healthcare Big Data. Furthermore, real advances depend on better ways to exploit the disconnected lakes of data (e.g., health records, clinical trial data) as well as better ways to capture, analyze and utilize the streams of new kinds of data (genomics, sensor readings, population, and disease tracking) that are flooding healthcare.

The PHEMI HEALTH DataLab Solution

PHEMI's HEALTH DataLab consolidates data silos into a data commons on Microsoft Azure Cloud: A data collaboration platform with de-identification, privacy management & fine-grained access controls.

Opportunities For Data Commons

- Major Canadian Health Authority consolidates data from multiple sources for advanced analytics
- Enable advanced research and analysis of data collected by government agency
- Improve population health & healthcare delivery through analytics

Healthcare Data Commons Use Case

- Flexible ingest to integrate data from a variety of sources and formats
- Data is annotated with common metadata based on the ingest source and context
- Policy-based data de-identification, privacy protection, and data classification based on access controls
- Multiple views to data; perspective is based on user attributes
- Automated versioning and lifecycle management of data, related models and code
- Integration with data profiling and cleanup toolchains
- Data is derived into a variety of application-specific models for analysis
- Automated datasets manufacturing and export to downstream systems