Collaborative Cancer Cloud demonstrates feasibility of data-sharing model, transitioning proof of concept to projects aimed at technology deployment

Groundwork is laid for further development; tools released open source and will be further developed by OHSU through national consortium

OHSU, Intel and its partners have successfully validated the Collaborative Cancer Cloud platform architecture, finalizing their work on the project.

The Collaborative Cancer Cloud is an advanced, federated analytics platform that enables clinicians and researchers to securely study large data sets across multiple institutions for potentially lifesaving discoveries and treatments. The project launched in 2015 and expanded in 2016 with the establishment of additional partners Dana Farber Cancer Institute and Ontario Institute for Cancer Research.

By completing the proof of concept, the organizations demonstrated the viability of securely sharing genomic information on a massive scale using a distributed architecture that protects privacy and local, institutional control of data. The model system made it possible for disparate research and medical institutions to connect over a secure network to conduct analyses of pooled data without having to download it.

Work done on the CCC lays the groundwork for other organizations to develop it further as part of an integrated infrastructure of tools and systems required to support scalable, widespread use in precision medicine.

“This has been an extremely fruitful project with Intel,” said Mary Stenzel-Poore, Ph.D., who as senior associate dean for research in the OHSU School of Medicine was instrumental in cementing the partnership with Intel. “Now that we have successfully proven the concept of a secure, federated platform to aggregate and analyze large data sets, the stage is set for the precision medicine ecosystem to take forward the concepts demonstrated by the Collaborative Cancer Cloud. By publicly releasing the software developed for the project we have opened up the platform to the entire community.”

Tools and expertise developed in the collaboration are being put to use at OHSU. A high-profile example is the $15 million Kids First Data Resource Center, a National Institutes of Health-funded effort to build the world’s largest database on pediatric cancer and birth defects. Adam Margolin, Ph.D., professor of biomedical engineering, and director of computational biology in the OHSU School of Medicine and the OHSU Knight Cancer Institute, is co-principal investigator. Dr. Margolin and collaborators from the Ontario Institute for Cancer Research – one of the CCC partners – decided to form a consortium to compete for the pediatric project, based on their work together on CCC.

“Our work with Intel has helped us become a major player in a national consortium for pediatric cancer,” said Dr. Margolin. “It’s also been a crucial building block for the OHSU School of Medicine’s Computational Biology Program. Through our work together, we’ve gained valuable experience and developed impactful software tools enabling large-scale data integration and data management. This experience is already being applied to other research initiatives underway on campus.”

Much of the software developed for the project has been released on GitHub, the popular web repository for hosting open-source software projects.

In the months ahead, researchers who worked on the project expect to report their scientific findings in peer-reviewed journals. One study, for example, used the data-sharing platform to search for mutation

hotspots among thousands of breast cancer genomes available from public data sources as well as from OHSU, Dana-Farber Cancer Institute and the Ontario Institute for Cancer Research, constituting the largest breast cancer genomics study performed to date.

Additionally, a number of shared tools have been developed as a result of the CCC collaboration.

- Genomics kernel library (GKL): a collection of kernels to speedup compression and decompression of genomics data, open sourced directly by Intel and used in The Broad Institute's GATK and Illumina's Isaac.
- GenomicsDB: a scalable database capable of supporting hundreds of thousands of genomes and open sourced directly by Intel.
- Graphical user interface (GUI): the interactive graphical mode for users to submit and monitor workflows will be open sourced by Intel and is used in the Broad-Intel Genomics Stack.
- Task execution system (TES): an API permitting deployment of computational tasks across multiple cloud environments, adopted by the Global Alliance for Genomics and Health (GA4GH), The Broad Institute, Google, and Seven Bridges Genomics, among others.

The CCC also established the technical capability required for Dana-Farber, Ontario Institute and OHSU to continue sharing aggregate data among themselves in order to conduct analyses as part of any research projects that are jointly defined and undertaken together in the future. And Intel and OHSU continue to collaborate on other projects as part of a shared goal to accelerate scientific progress in understanding complex diseases.

Beyond the CCC collaboration, the strong relationship with Intel has resulted in additional benefits to OHSU's research community. The ExaCloud, for instance, at OHSU's West Campus Data Center is a resource developed in close collaboration with Intel to support very large scale computation. The primary ExaCloud cluster includes over 6,600 of Intel's Xeon microprocessors. And Intel experts are getting ready to work with OHSU faculty to advance a new MRI technology developed at OHSU.

"The CCC is a strong, successful beginning," said Dr. Stenzel-Poore, who is now chief of research operations in the OHSU Knight Cancer Institute. "OHSU will continue to pursue partnerships and opportunities in the data sharing and data management space as part of our ongoing efforts in precision medicine. By joining forces with health care and industry partners, we will move the vision forward of 'All in One Day' cancer care."

Related Reading

- Collaborative Cancer Cloud welcomes new institutions
- Collaborative Cancer Cloud launches
- A downpayment on the future of precision medicine
