



Contribution ID: 415

Type: **Talk**

Federated Computing for Health Data Science in Africa

There is a great need to develop computing infrastructure to support the increased application of data science and health informatics across Africa which includes robust data sharing and federated computing, whilst fostering research collaboration. The Global Alliance for Genomics and Health (GA4GH; <https://www.ga4gh.org/>) aims to promote responsible data sharing standards through the use of open, community derived standards and APIs such as Data Repository Service (DRS), Workflow Execution Service (WES), Data Connect, Passports and Tool Registry Service (TRS), amongst others. The DRS API provides a generic interface to access data in repositories. The data is discovered through the Data Connect API which supports federated search of different kinds of data. The WES API provides a standardized approach for accessing computing resources with use of reproducible workflows, usually housed in a tools registry service such as Dockstore (<https://dockstore.org/>). The eLwazi Open Data Science Platform (ODSP) has undertaken a pilot implementation of the GA4GH standards with the aim of delivering a federated framework for data discovery and analysis within Africa for the DS-I Africa consortium. The eLwazi GA4GH pilot project was started in June 2023 as an outcome of a training hackathon by the eLwazi ODSP Infrastructure work group in collaboration with the GA4GH. The main goal of the GA4GH pilot project is to enable the findable, accessible, interoperable and reusable (FAIR) principles for data discovery and analysis. Four sites within Africa (Ilifu - South Africa, ACE Lab - Mali, ACE Lab - Uganda and UVRI - Uganda) are currently hosting the different API endpoints for authorized data discovery and analysis. From within the project we can locate DRS datasets using the Data Connect API, use workflows from Dockstore via the TRS API for reproducible analysis, and submit it to the WES API for analysis without the data leaving the actual location, which provides a technical solution for data analysis within legislative data protection constraints. We are now in the process of developing a federated approach for the imputation of African genomics data as a GA4GH implementation forum (GIF) project collaboration based on the lessons from the pilot GA4GH implementation project.

Presenting Author

Takudzwa N Musarurwa

Email

tnmusarurwa@gmail.com

Student or Postdoc?

No. Not a student nor Postdoc.

Institute

University of Cape Town

Registered for the conference?

No

CHPC User

CHPC Research Programme

Workshop Duration

Primary authors: Mr MUSARURWA, Takudzwa Nyasha (University of Cape Town - eLwazi ODSP); Mr MAIGA, Aboubacar (ACE Labs - Mali); MUTEGEKI, Henry (ACE Labs - Uganda); Mr WAMALA, Timothy (Uganda Virus Research Onstitute); Mr SOUMARE, Sidy (NIH ICER Mali); KIMERA, Rodgers; KENNEDY, Dane; Dr BENDOU, Hocine (University of Cape Town); Mr MEINTJES, Ayton (University of Cape Town - eLwazi ODSP); Mr MBIYAVANGA, Mamana (University of Cape Town - eLwazi ODSP); BUNDE, Amos; Mr CUPAK, Miro (DNAS-tack); VIZZIER, Ben; KAYONDO, Jonathan; Dr JJINGO, Daudi (ACE Labs - Uganda); DOUMBIA, Seydou; SIMMONDS,, Rob; Dr PANJI, Sumir (University of Cape Town - eLwazi ODSP); O'CONNOR, Brian (GA4GH); Mr BOTHA, Gerrit (Univeristy of Cape Town - eLwazi ODSP); Prof. MULDER, Nicola (University of Cape Town - eLwazi ODSP)

Presenter: Mr MUSARURWA, Takudzwa Nyasha (University of Cape Town - eLwazi ODSP)

Session Classification: HPC Technology

Track Classification: HPC Technology