AN E-INFRASTRUCTURE BASED ON CLOUD AND EDGE COMPUTING FOR DATA ANALYSIS IN METABOLOMICS

Researches from CIBER, URV and IISPV have developed a new method of data analysis that allows the identification of metabolites from complex samples in an accurate and fast way.

The Need

In the era of Big Data, the tsunami of massive 'omics' data is revolutionizing the way we do science. The future success of the biomedical and life sciences will depend on the ability of investigators to properly process, analyze, and interpret large-scale, multidimensional data sets that are generated by high throughput technologies. Among these, metabolomic analyses are essential for biomedical, environmental, and biotechnological research. Despite the broad use of Cloud computing in genomics, there are few cloud-based solutions available for metabolomics data analysis, and none of these, however, include computational tools covering the entire spectrum of mass spectrometry-based metabolomic technologies.

The Solution

The aim is to build a novel Cloud computing solution and ecosystem, called DATOMA, based on multi-cloud serverless programming, for metabolomics data analysis. We aim to run our novel cloud computing infrastructure as a fee-for-service platform.

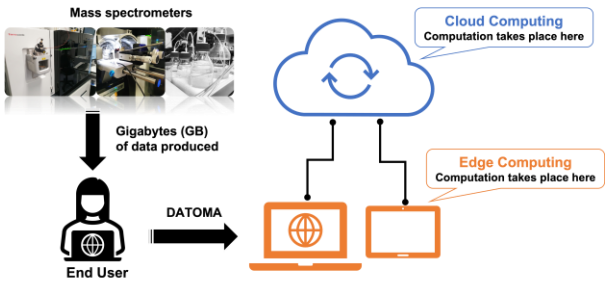
Initially, DATOMA will implement the computational tools developed by Dr. Oscar Yanes' group, which cover the entire spectrum of mass spectrometry (MS)-based metabolomic technologies, and are published, protected via patent application, tested and validated by the international community.

Innovative Aspects

- DATOMA will provide much needed resources enabling faster and cheaper processing of much larger LC-MS/MS, GC-MS, and MS Imaging datasets than would be possible in most individual public and private organizations.
- The Cloud computing solution will enable users anywhere in the world to easily set-up and run metabolomics data processing workflows over distributed/federated resources in the Cloud/Edge continuum, facilitating easier collaboration between remote sites and introducing a standardised, reproducible, and accepted data analysis process.
- DATOMA will cover the needs of users and devices that lack the computational power and storage capacity to analyze metabolomics data, and requirements precluding data movements to the Cloud.

Intellectual Property

- Spanish patent granted in 2020, PCT applied in 2021.



DATOMA's e-infrastructure will run through two complementary architectures: (i) **Central Cloud Computing (CCC)**; and (ii) **Edge Cloud Computing (ECC)**. DATOMA is based on: (a) HERMES for LC-MS/MS (Giné R. et al. *Nature Methods*, 2021) (patent ES2767375B2); (b) eRah for GC-MS (Domingo-Almenara, X. *Anal Chem*, 2016); and (c) rMSIproc for Imaging MS/single cell metabolomics (Ràfols, P. *Bioinformatics*, 2020). Our long-term objective is to expand DATOMA to new functionalities in a collaborative effort with the scientific community.

Contact details

Consorcio Centro de Investigación Biomédica en Red *(CIBER)*
otc@ciberisciii.es
https://www.ciberisciii.es/en