TITLE

MANAGEMING AND ANALYZING ANALYTICAL CHEMISTRY DATA SETS

TEAM *Sana Tfaili*[1], *Ali Tfayli*[1] and *Pierre Chaminade*[1].

[1]Lip(Sys)²- Chimie Analytique Pharmaceutique[§], Univ. Paris-Sud, Université Paris-Saclay, F-92290 Châtenay-Malabry, France.

[§](FKA EA4041 Groupe de Chimie Analytique de Paris-Sud)

ABSTRACT

Analytical chemistry is a scientific domain which has been revolutionized by recent technological and instrumental advances: both the resolution of acquisitions has been increased and the time needed to obtain them has been drastically reduced, contributing to the **generation of massive and complex data at an unprecedented scale**.

Therefore and as **data providers, Lip(Sys)²** domain scientists **participated in CDS 1.0** through the project **MoDALMI** "Management of Data for the Analysis of Lipids, Metabolites and Isotopes". The project focused on making it possible to manage produced data sets individually.

Thank to rapid development of **technology and open-source community**, related to Linked Data like the Linked Data Platform (LDP) of Apache Marmotta, R, Matlab and thank to the **support of BorderCloud**, it is **now possible to archive data** from different analytical tools in the laboratory and **perform automatic conversions to a common format**.

An intern server at the informatic's direction of Université Paris Sud permits the data storage. When new data is detected, plugins are deployed on the server automatically. **MoDALMI** was also **materialized by** a platform called **DAAP** for "Data Acquisition in Analytical Platform", (daap.eu). DAAP is a repository of all devices, and permits to link data to the device and the corresponding projects. To create this repository, **researchers** have been **familiarized to defining and using ontology, SPARQL queries and wikimedia**.

Besides our **need to optimize some points in MoDALMI** results and to **enrich our database**, we now focus on a **new challenge** for implementing **user friendly software VRAC for 'Virtual Research environment for Analytical Chemistry'**.

The development of this software will permit the collect of experimental, protocol and artefact data. It will supply existing software in the laboratory to get the data. **VRAC must allow** the **distinction** between **private** and **public data**. It will feed open access data of the laboratory intern database, the University intern database and finally public database open to Linked Open Data. **Another challenge** is to **conjointly consider** and **deeply analyze data** sets **of various origins** to fully **exploit** their **complementary aspects**.

For this, **we need support from data scientists**. We aim to investigate new kinds of computational analyses of analytical chemistry data to automate some actual data treatments and maybe define new ones. In CDS 1.0 and thank to the help of data scientists, a RAMP was organized around cancer drugs for qualitative identification and quantitative prediction. This was a sample of data that we produce in the laboratory. Many other examples are also available. We wish to continue collaborating with data scientists, challenges are the heterogeneity of the data sets that we want to manage via VRAC and their analyzing. The **advantage** is the **possibility to transpose this project** in **UPSaclay** but also to **other structures** ie: companies **in pharmaceutical domain**.