

# « FAIR metabolomics e-resources in web components »

O. Filangi<sup>13</sup>, N. Paulhe<sup>23</sup>, F. Giacomoni<sup>23</sup>

<sup>1</sup> Plateau de Profilage Métabolique et de Métabolomique (P2M2, Rennes) - Centre de Rennes (INRAE)

<sup>2</sup> Plateforme "Exploration du Métabolisme" (MetaboHUB-PFEM, Clermont-Ferrand) - Centre de Theix (INRAE)

<sup>3</sup> CATI eMPREInTE (Molecular PhEnotyping and blochemical daTa Engineering)

La métabolomique est l'étude des petites molécules (métabolites) constituant le métabolome, qui est l'ensemble des métabolites présents dans un échantillon biologique à un moment et dans des conditions données. Au cours des dernières années, plusieurs ressources ont été développées et exploitées par les membres de la communauté métabolomique pour analyser les métabolites et leur contexte biologique. L'une des principales lacunes des flux d'analyse actuels est qu'il devient difficile de comparer et d'analyser des données d'entrée et de résultats d'expérience vis-à-vis de la connaissance déjà acquise et intégrée dans plusieurs bases de connaissances curée et annotées. Pour déduire les caractéristiques et les produits des nouvelles analyses, il est nécessaire d'exploiter pleinement les mécanismes d'interopérabilité fournis par les technologies actuelles. Il s'agit là d'un défi majeur lorsqu'on envisage d'intégrer les données métabolomiques à d'autres études ou à d'autres approches -omiques. Ce défi peut être relevé par l'utilisation de ressources définies par le Web sémantique qui est une extension du Web standardisée par le World Wide Web Consortium (W3C). Les ressources en métabolomique couvrent diverses disciplines scientifiques telles que la chimie structurale et analytique, les statistiques ou la médecine. Les principaux fournisseurs comme le « National Center for Biotechnology Information » (NCBI) et le « European Bioinformatics Institute » (EBI) produisent, fédèrent et fournissent ces données issues de la recherche quotidiennement dans un format conforme aux modèles du Web Sémantique. Ainsi, cette information est librement disponible selon plusieurs stratégies :

- Points d'accès SPARQL qui permet la construction de requêtes complexes émises via le protocole HTTP
- API dédiés pour la construction de requêtes spécifiques à travers le Web
- Fichiers structurés dans un format spécifique défini par le W3C (OWL, RDF/XML, Turtle, JSON-LD)

Au sein de l'infrastructure nationale de métabolomique et fluxomique - MetaboHUB, les membres du work package « Creating FAIR e-resources for knowledge mining » ont pour missions de structurer l'information et les systèmes d'information du consortium dans le respect des principes FAIR (Faciles à trouver, Accessibles, Interopérables et Réutilisables). Dans ce cadre, nous développons une architecture logicielle en web composant intégrant l'utilisation des ressources définies par le web sémantique. Cette architecture permet de faciliter d'une part la maintenabilité et la réutilisation de composants graphiques, mais aussi de développer des outils d'aide à la décision s'appuyant sur les possibilités d'inférence entre les graphes RDF du consortium et celles fournies par l'EBI et le NCBI. Afin de faciliter le développement des « web composants » métiers, nous proposons une couche logicielle de management des ressources RDF, appelée « Discovery » (API discovery<sup>1</sup>) qui permet de faciliter la construction de requêtes « métiers » et leurs gestions. Ce développement s'appuie sur les outils mis à disposition par le *RDF JavaScript Libraries Community Group*<sup>2</sup> et sur le langage *Scala* associé au compilateur *Scala.js*. Cet environnement permet de générer une API robuste exploitant des graphes RDF sur plusieurs environnements cibles (navigateur web - javascript ou serveur distant - nodejs/scala). Nous proposons de présenter cette architecture avec deux cas d'usage utilisant les ressources *NCBI organismal classification*<sup>3</sup> et l'ontologie « *Chemical Entities of Biological Interest* » (*ChEBI*)<sup>4</sup> pour enrichir les fonctionnalités du système d'information *PeakForest*<sup>5</sup>.

<sup>1</sup> <https://p2m2.github.io/discovery/>

<sup>2</sup> <https://www.w3.org/community/rdfjs/>

<sup>3</sup> <http://www.obofoundry.org/ontology/ncbitaxon.html>

<sup>4</sup> <https://www.ebi.ac.uk/chebi/init.do>

<sup>5</sup> <https://peakforest.org/>