FAIR metabolomics e-resources in web components

O. Filangi [1], N. Paulhe [2], F. Giacomoni [2]

Séminaire INRAE “Semantic Linked Data” – 2021

- *[1] Plateau de Profilage Métabolique et Métabolomique (P2M2, IGEPP, Rennes)*
- *[2] PlateForme Exploration du Métabolisme (PFEM, Clermont-Ferrand)*

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

P2M2
Plateau de Profilage Métabolique & de Métabolomique

Plate-Forme
EXPLORATION du METABOLISME
des gènes aux métabolites

METABOHUB

Plan

I. Contexte
II. Projet FORUM
III. Metabolomic Semantic Datalake [management des ressources RDF]
IV. Discovery / Web Component [utilisation des ressources RDF]

Context

MetaboHUB 2.0 (2021 – 2025)
Infrastructure nationale de métabolomique et fluxomique
⇨ a pour objectif de fournir des outils technologiques de pointe et des services en métabolomique et fluxomique aux équipes de recherche académiques et à des partenaires industriels dans les domaines de la santé, de la nutrition, de l'agriculture, de l'environnement et des biotechnologies.

WP5 "Creating FAIR e-resources for knowledge mining"

- T.1 MetaboHUB Data management and stewardship
- T.2 Ensure MTH2.0 e-resources sustainability
- **T.3 Interoperability between e-resources: creating study specific e-resources**
- **T.4 Linked data for advanced knowledge extraction**

FORUM / CATI EMPREINTE (F. Giacomoni, C. Frainay)

2019 d'un financement de la part du copil Schéma Directeur du Numérique (SDN)

=> **Preuve de concept de l'apport des technologies du web sémantique dans l'ouverture, la diffusion et l'exploitation d'empreintes métaboliques et de réseaux métaboliques autour de questions scientifiques touchant à la Nutrition et à la Santé**

Portail http://semantic-metabolomics.org/

FORUM - Constat

- **Problématique production** (Utilisation dans les SI/IHM)
  - RDFstore versus base relationnelle
    - Interopérabilité / robustesse
    - API / maintenabilité

- **Problématique recherche/intégration/inférence**
  - Gestion bases de connaissances massives PubCHEM ~ 8 Milliards de triplets
  - Accèssibilité/Updates des bases de connaissances externes (NCBI/EBI/…)
    ⇒ Update des bases de connaissances internes
  - Intégration des bases non RDF (Metabolights, Global Natural Products Social Molecular Networking, Golm Metabolome Database, etc...)

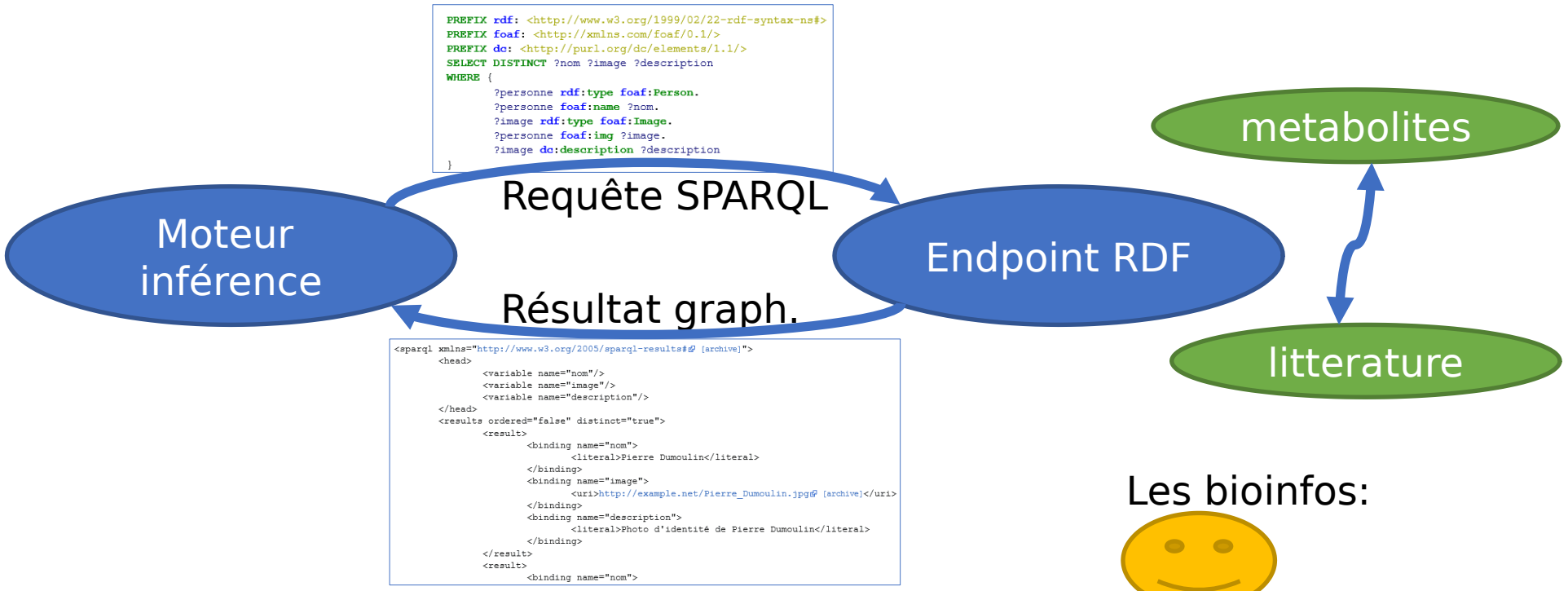
2 exemples

⇒ Mise à disposition des résultats du projet (relations entre des composés et des concepts biomédicaux) via un site web

⇒ Retour d'exp. mise à disposition d'un RDF store (8 Millards triplets)

Forum – Problématique 1 / utilisation

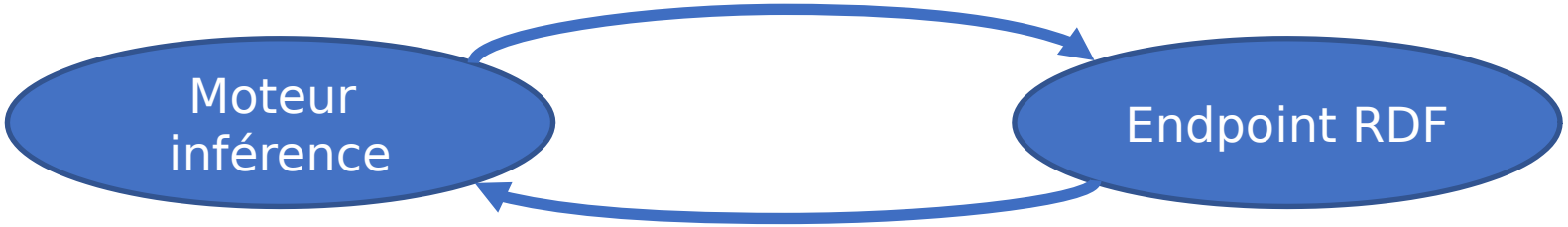
Web sémantique et bioinformatique

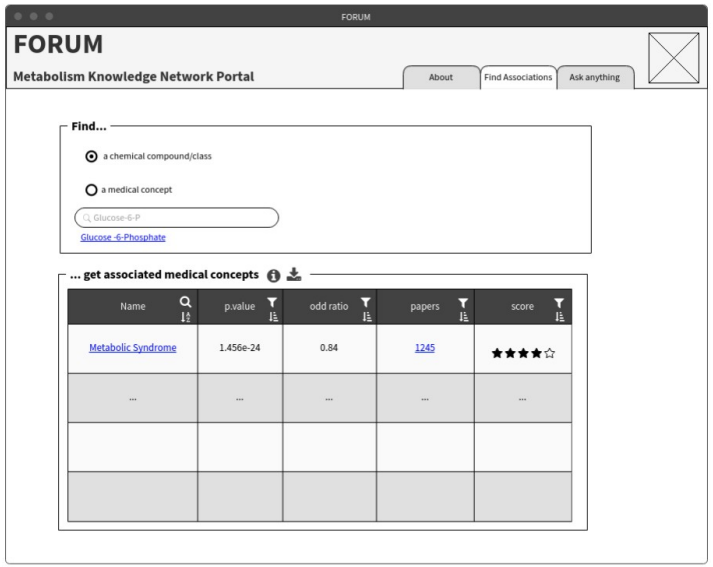


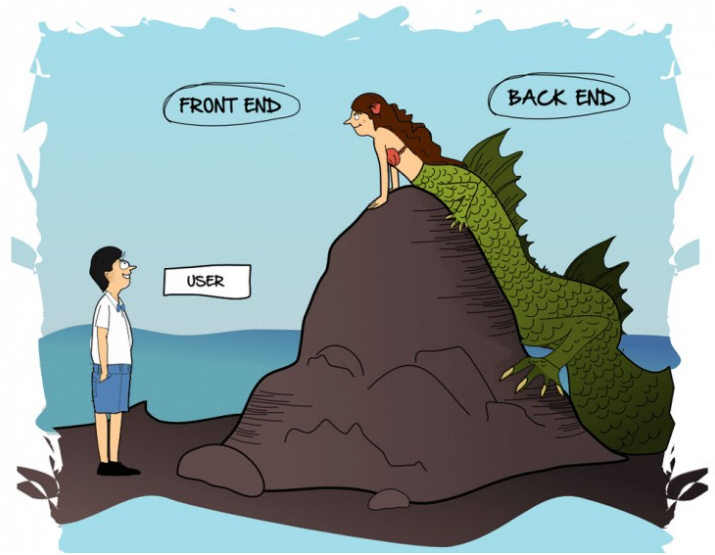
Source: Wikipedia

Forum – Problématique 1 / utilisation

Web sémantique et bioinformatique



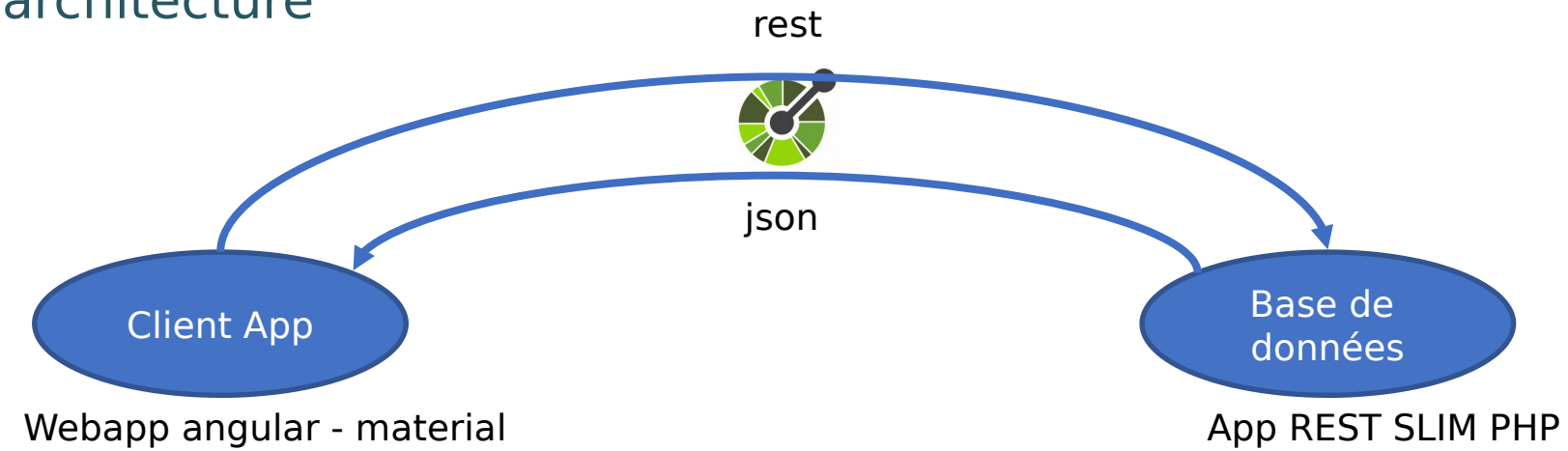


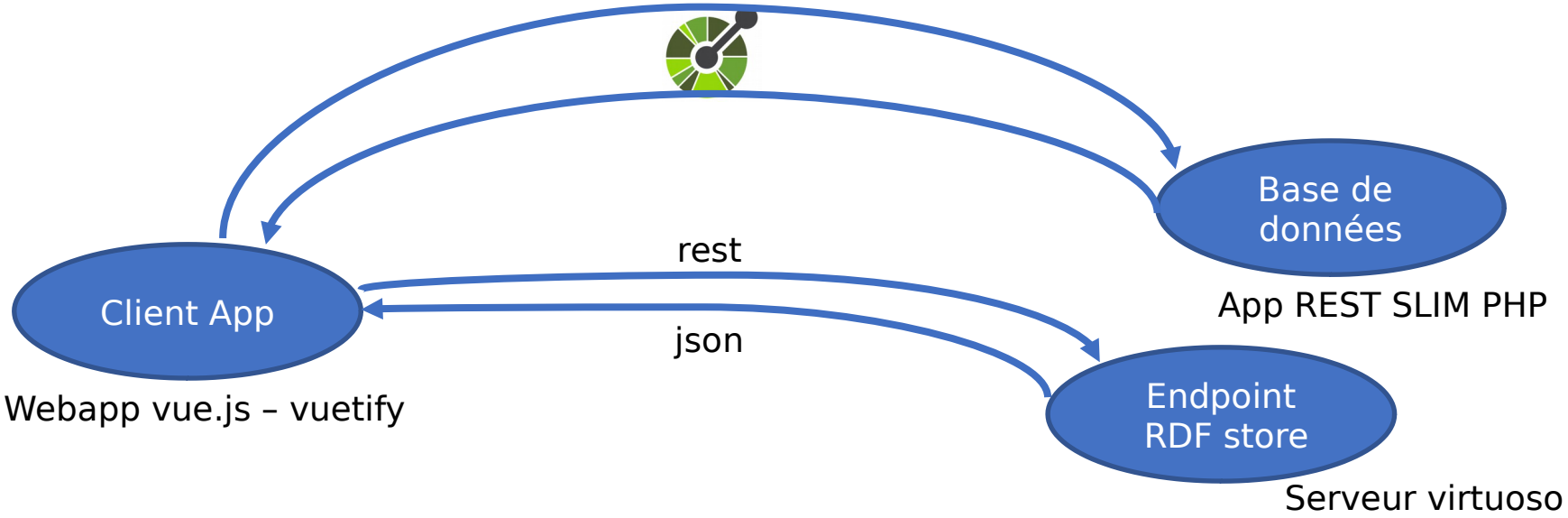


Les biologistes:

Forum – Problématique 1 / utilisation

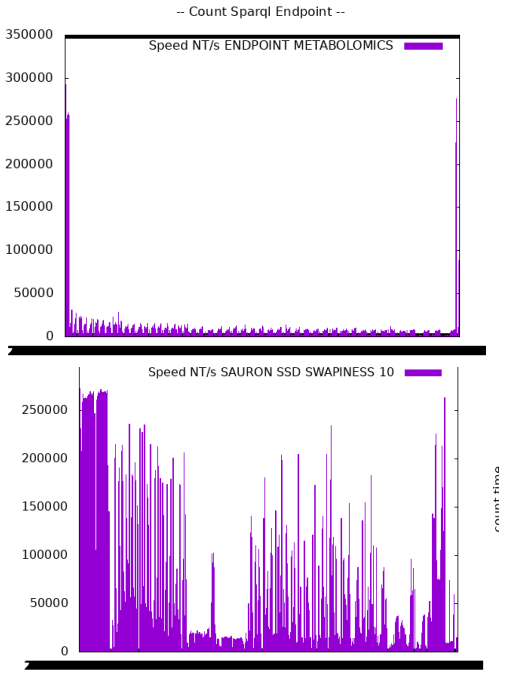
« Meta » architecture





Forum – Problématique 2 / mise à disposition

- **Bases : PubChem, ChEBI, Enrichment**
- **8 Milliards de triplets (8,256,324,376)**
- **Virtuoso**



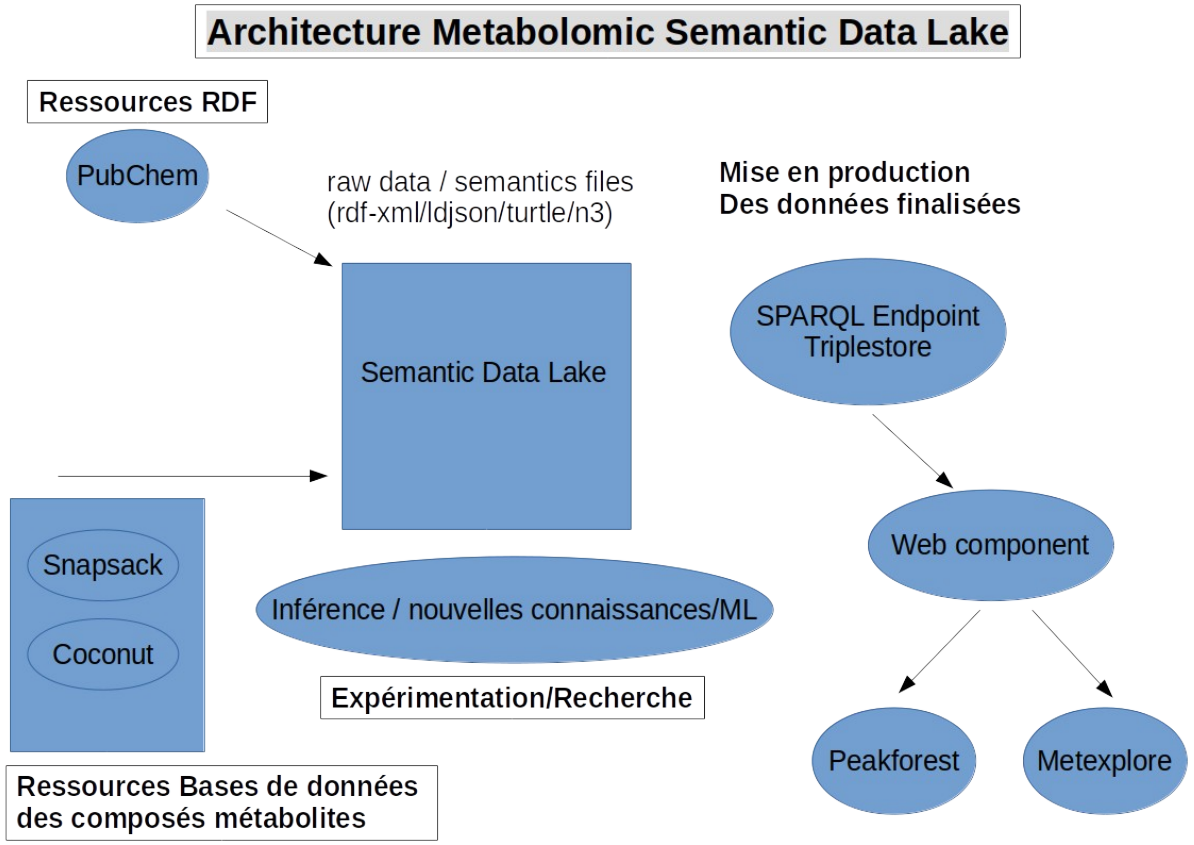
	ENDPOINT-METABOLOMICS	SAURON (DEV)
CORE	16	48
MEM	128Go	192Go
Env Test	Disque Local(SSD)/RBD-SATA	RBD-SATA/RBD-SSD/ NetAPP

	EM - LOCAL(SSD)	EM - RBD-SATA	SAURON-RBD-SATA	**SAURON-RBD-SSD**	SAURON-RBD-SSD **(swapiness10)**	SAURON-NETAPP
vitesse moy(T/sec)	40996.4	7189.96	12428.9	71366.2	64883.3	20243.3
vitesse std	63130.8	25868.6	37922.2	82599	75743.8	47900
durée	~2,8j	~22j	~12,4j	~1,74j	~1,81j	~5,57j

Metabolomic Semantic Datalake

Évaluation d'une infrastructure BIG Data pour la gestion et l'inférence de ressources RDF

- **Financement**
  - **AAP DiPSO 2021 « Soutien à Projets Innovants et Structurants sur le numérique » (20 k€)**
  - MetaboHUB (1k€)
  - ALIMH (1k€)
  - Plateforme Métabolomique P2M2 (1k€)
- **Matériel**
  - 1 Namenode (VM, 4 CPU, 250 Go)
  - 4 Datanodes
    - Serveur PowerEdge C6525 (2CPU,64 RAM, 2x1To SSD)
- **Formations**
  - Big Data / Administration d'un cluster SPARK 4 ½j - Septembre 2021
  - **Big Data / Spark env. / Scala – 6 ½ j Janvier-Février 2022**



- **PROSODIe** : C. Duperier (PFEM), D. Benaben (PMB)
- **BARIC** : M. Boudet (IGEPP)
- **EMPREINTE** : F. Giacomoni (PFEM), O. Filangi (IGEPP)

Metabolomic Semantic Datalake / Technologies

Gestion / Création de ressources

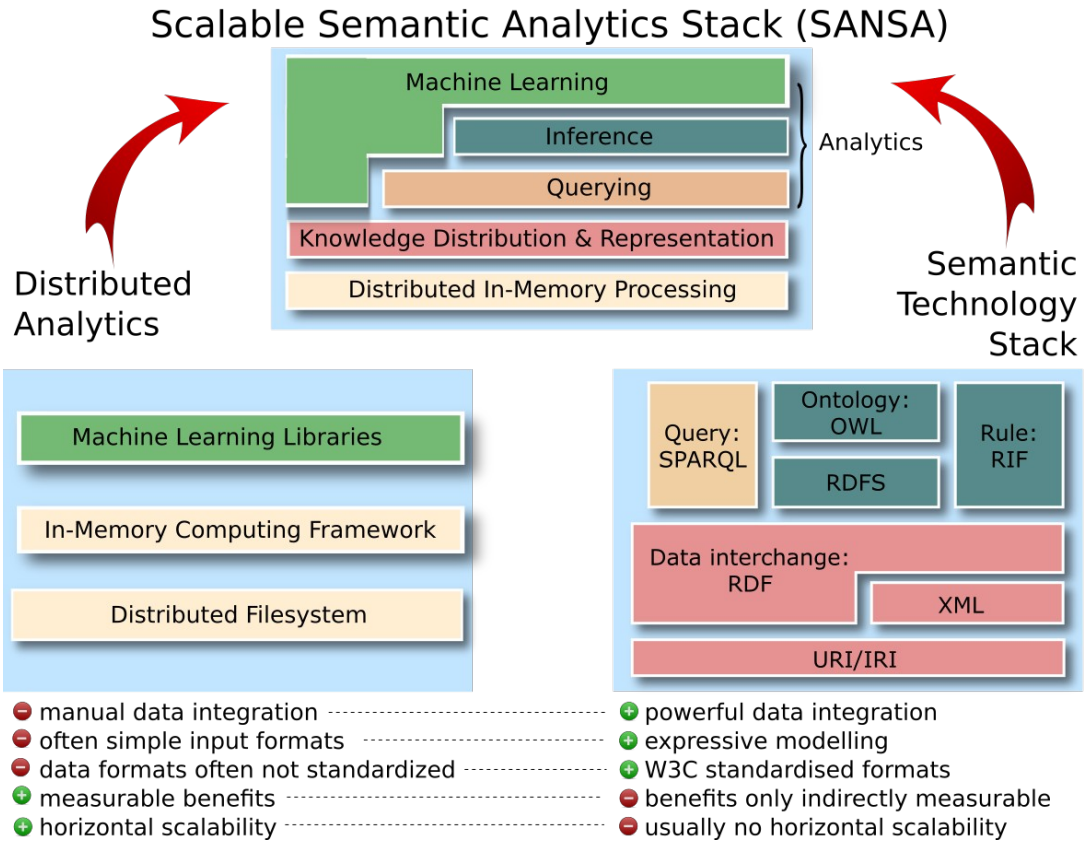
Apache Spark · Scala · SANSA · hadoop · Jena

Management des ressources sémantiques

Ask0mics

Utilisation des ressources RDF

Discovery Vue.js



SANSA-Stack : http://sansa-stack.net/

Smart Data Analytics (SDA) research group

Discovery

https://p2m2.github.io/discovery

DSL (Domaine Specific langage) pour faciliter l'utilisation des ressources RDF et l'implémentation de fonctionnalités dans les SI MetaboHUB. La construction se base sur les concepts de **navigation à facette** (notion de focus et de contexte)

```
SWDiscovery(config)
        .something("chebiId")
          .set(URI("http://purl.obolibrary.org/obo/CHEBI_4167"))
            .isObjectOf(URI("https://www.ebi.ac.uk/metabolights/property#Xref"),"study")
              .datatype(URI("http://www.w3.org/2000/01/rdf-schema#label"),"labelStudy")
          .focus( "chebiId" )
            .datatype(URI("http://www.w3.org/2000/01/rdf-schema#label"),"labelChebiId")
        .select("study","labelStudy","labelChebiId")
```

Scala rdf4j

Fortement typé / langage fonctionnel
Robustesse et maintenabilité

Ecosystème RDF / JVM

JVM (Java Virtual Machine)

Ecosystème NodeJS/Browser

Scala-JS

COMUNICA

Ecosystème RDF / Web

node

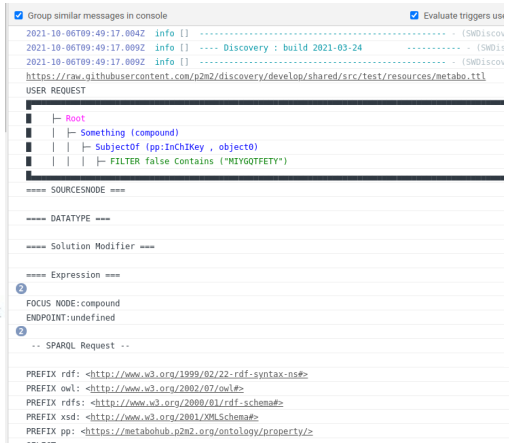
COMUNICA : Flexible SPARQL and GraphQL over decentralized RDF on the Web. Ghent University, https://comunica.dev/

Discovery / exemple HTML / 1

```html
<script type="text/javascript" src="https://cdn.jsdelivr.net/gh/p2m2/discovery@0.2.0/dist/discovery-web.min.js"> </script>
<script>
      let config = SWDiscoveryConfiguration.setConfigString(`
          {
          "sources" : [{
                  "id"        : "file_metabo",
                  "url"       : "https://raw.githubusercontent.com/p2m2/discovery/develop/shared/src/test/resources/metabo.ttl",
                  "mimetype"  : "text/turtle"
           }],
          "settings" : {
           "logLevel" : "info"
           }
           }
          `)
        SWDiscovery(config)
                    .prefix("pp","https://metabohub.p2m2.org/ontology/property/")
                    .something("compound")
                      .isSubjectOf("pp:InChIKey")
                         .filter.contains("MIYGQTFETY")
                    .focus("compound")
                      .console()
                      .helper() // display information on the web page and propose new building block
                      .select("compound")
                      .commit().raw().then((response) => {

                        for (let i=0;i<response.results.bindings.length;i++) {
                          let iValue =response.results.bindings[i]["compound"].value;
                            console.log(iValue );
                        }

                      }).catch( (error) => {
                        console.error(" -- catch exception --")
                        console.error(error)
                      } );
</script>
```



Discovery

- build : build 2021-03-24

New step on focus

- **target node** : Something@compound - [List(SubjectOf@object0 - [List(false Contains ("MIYGQTFETY"))])]
- **regex** :
- **Number of values** : **1**

Values 10 .set(value) .setList(value1,value1,..)

- {"compound":{"value":"https://foodb.ca/compounds/FDB000013","type":"uri"}}

Classes .isA("uri")

- https://metabohub.p2m2.org/ontology/class/Compound

Forward property .isSubjectOf("uri","myreferencevar")

[X]

Datatype property .datatype("uri","myreferencevar")

- https://metabohub.p2m2.org/ontology/property/db_source
- https://metabohub.p2m2.org/ontology/property/cas_number
- https://metabohub.p2m2.org/ontology/property/InChI
- https://metabohub.p2m2.org/ontology/property/iupac_name
- https://metabohub.p2m2.org/ontology/property/mass_minush
- https://metabohub.p2m2.org/ontology/property/mass_plush
- https://metabohub.p2m2.org/ontology/property/exact_mass
- https://metabohub.p2m2.org/ontology/property/InChIKey
- https://metabohub.p2m2.org/ontology/property/can_smiles

Backward property .isObjectOf("uri","myreferencevar")

-

```
-- SPARQL Request --                                                    discovery-web.min.js:1570
                                                                        discovery-web.min.js:1570
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>               discovery-web.min.js:1570
PREFIX owl: <http://www.w3.org/2002/07/owl#>                            discovery-web.min.js:1570
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>                    discovery-web.min.js:1570
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>                         discovery-web.min.js:1570
PREFIX pp: <https://metabohub.p2m2.org/ontology/property/>              discovery-web.min.js:1570
SELECT                                                                  discovery-web.min.js:1570
                                                                        discovery-web.min.js:1570
WHERE {                                                                 discovery-web.min.js:1570
    ?compound pp:InChIKey ?object0 .                                    discovery-web.min.js:1570
    FILTER ( contains(str(?object0),"MIYGQTFETY") )                     discovery-web.min.js:1570
}                                                                       discovery-web.min.js:1570
https://foodb.ca/compounds/FDB000013                                    example_with_urlfile.html:28
```

https://p2m2.github.io/discovery/user_docs_file_example.html

Discovery / exemple Scala / ChEBI Discovery

ScalaJs functional implementation based of the ontology based matching describe in the "Improving lipid mapping in Genome Scale Metabolic Networks using ontologies." Poupin et al - 2020 using the Chemical Entities of Biological Interest (ChEBI)
https://doi.org/10.1007/s11306-020-01663-5

https://github.com/eMetaboHUB/ChEBI-discovery

```
@JSExportTopLevel(name="ChebiDiscovery")
case class ChebiDiscovery(
                          config_discovery : String = """{
              "sources" : [{
                "id"  : "Forum-Desease-Chem",
                "url" : "https://forum.semantic-metabolomics.fr/sparql"
              }],
              "settings" : {
                "logLevel" : "off"
              }
            }""".stripMargin

                         ,groupBySize : Int = 1000) {


val instDiscovery =
  SWDiscovery(StatementConfiguration.setConfigString(config_discovery))
    .prefix("c","http://purl.obolibrary.org/obo/")
    .prefix("cp", "http://purl.obolibrary.org/obo/chebi#")
```

```
def get_has_role(chebi: URI)
: Future[Seq[URI]] = {

  instDiscovery
    .something()
     .set(chebi)
    .isSubjectOf(URI("rdfs:subClassOf"), "ac")
      .isSubjectOf(URI("owl:onProperty"))
        .set(URI("http://purl.obolibrary.org/obo/RO_0000087"))
    .focus("ac")
      .isSubjectOf(URI("owl:someValuesFrom"),"idChebi_role")
    .select(List("idChebi_role"))
    .commit()
    .raw
    .map(json =>
      json("results")("bindings").arr.map(row => SparqlBuilder.createUri(row("idChebi_role"))).toSeq
    )
}
```

```
sbt:chebi-discovery> testOnly com.github.inrae.metabohub.Metabolites2NetworkTest
------------------------------ Running Tests -----------------------------
2021-10-07T13:23:25.232Z  info []  ------------------------------------------------ - (SWDiscovery.scala:26)
2021-10-07T13:23:25.233Z  info []  ---- Discovery : build 2021-05-25         ---------- - (SWDiscovery.scala:27)
2021-10-07T13:23:25.233Z  info []  ------------------------------------------------ - (SWDiscovery.scala:28)
Map(<http://purl.obolibrary.org/obo/CHEBI_15756> -> Map(is_conjugate_acid_of -> <http://purl.obolibrary.org/obo/CHEBI_7896>), <http://purl.obolibrary.org/obo/CHEBI_7896> -> Map(is_conjugate_b
ase_of -> <http://purl.obolibrary.org/obo/CHEBI_15756>))
Map(<http://purl.obolibrary.org/obo/CHEBI_90488> -> Map(is_conjugate_acid_of -> <http://purl.obolibrary.org/obo/CHEBI_57880>))
Map(<http://purl.obolibrary.org/obo/CHEBI_30828> -> Map(is_conjugate_base_of -> <http://purl.obolibrary.org/obo/CHEBI_36023>))
Vector((<http://purl.obolibrary.org/obo/CHEBI_7896>,is_conjugate_acid_of,0.1), (<http://purl.obolibrary.org/obo/CHEBI_7896>,is_conjugate_base_of,-0.1))
Vector((<http://purl.obolibrary.org/obo/CHEBI_57880>,is_conjugate_acid_of,1.1))
Vector((<http://purl.obolibrary.org/obo/CHEBI_30828>,is_conjugate_base_of,-1.1))
+ com.github.inrae.metabohub.Metabolites2NetworkTest. chebi -  Metabolites2NetworkTest examples  1097ms
[info] Tests: 1, Passed: 1, Failed: 0
[success] Total time: 3 s, completed 7 oct. 2021 à 15:23:26
sbt:chebi-discovery>
```

Discovery – POC avec PeakForest

Implémentation de Discovery dans PeakForest – actuel & futur

Requête: « recherche toute les publication liées à un ChEBI id »

```
function listMetabolightsStudies(chebiid) {
    // init
    const config = SWDiscoveryConfiguration.setConfigString(`
    {
      "sources" : [{
        "id"  : "metabolights",
        "url" : "https://metabolights.semantic-metabolomics.fr/sparql"
      }]
    }`);
    const query = new SWDiscovery(config);
    // let chebiid = "CHEBI:4167";
    chebiid = chebiid + "";
    chebiid = chebiid.toLowerCase().indexOf("chebi") < 0 ? "CHEBI:" + chebiid : chebiid;
    let chebi = chebiid.replace( ':' , '_');
    let r = SWDiscovery(config).something()
                .set("http://purl.obolibrary.org/obo/" + chebi)
                .isObjectOf("https://www.ebi.ac.uk/metabolights/property#ref","study")
                .filter.contains("MTBLS")
                .focus("study")
                .datatype("http://www.w3.org/2000/01/rdf-schema#label","label")
                .datatype("https://www.ebi.ac.uk/metabolights/property#technology_type","technology_type")
                .datatype("https://www.ebi.ac.uk/metabolights/property#comment","comment")
                .datatype("https://www.ebi.ac.uk/metabolights/property#instrument_platform","instrument_platform")
                .datatype("https://www.ebi.ac.uk/metabolights/property#Organism_Part","Organism_Part")
                .datatype("https://www.ebi.ac.uk/metabolights/property#study_design","study_design")
                .datatype("https://www.ebi.ac.uk/metabolights/property#study_factor","study_factor")
                .datatype("https://www.ebi.ac.uk/metabolights/property#organism","organism")
//              .datatype("https://www.ebi.ac.uk/metabolights/property#omics_type","omics_type")
//              .datatype("https://www.ebi.ac.uk/metabolights/property#publication","publication")
//              .datatype("https://www.ebi.ac.uk/metabolights/property#author","author")
                .select("study", "label", "technology_type", "comment", "instrument_platform",
                        "Organism_Part", "study_design", "study_factor", "organism"); // "omics_type", "publication",

    // run
    r.commit().raw().then((response) => {
        let strData = "";
        for (let i = 0; i < response.results.bindings.length; i++) {
            let study               = response.results.bindings[i]["study"].value;
            let label               = response.results.datatypes["label"][study][0].value;
            let technology_type     = response.results.datatypes["technology_type"][study][0].value
            let comment             = response.results.datatypes["comment"][study][0].value;
            let instrument_platform = response.results.datatypes["instrument_platform"][study][0].value;
            let organism_part       = response.results.datatypes["Organism_Part"][study][0].value;
            let organism            = "";
            if (Object.prototype.hasOwnProperty.call(response.results.datatypes["organism"], study)) {
                organism            = response.results.datatypes["organism"][study][0].value;
```

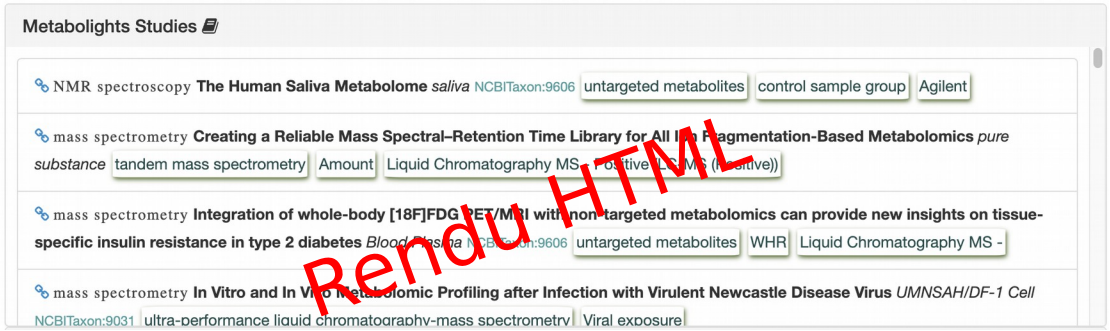
Appel du endpoint

Résultats à traiter

```
            let study_factor        = "";
            if (Object.prototype.hasOwnProperty.call(response.results.datatypes["study_factor"], study)) {
                study_factor        = response.results.datatypes["study_factor"][study][0].value;
            }
            // console.log(label, "=>", study, "; ", comment);
            comment = $.parseHTML(comment); //parseHTML return HTMLCollection
            comment = $(comment).text(); //use $() to get .text() method
            // new line
            strData += '<li class="list-group-item">' +
                '<a href="' + study + '" target="_blank"><i class="fa fa-link"></i></a> ' +
                '<span class="mtbls-technology-type">' + technology_type + '</span> ' +
                '<span class="mtbls-label" title="' + comment + '"> ' + label + '</span> ';
            // organism_part
            if (organism_part !== "") {
                strData += '<span class="mtbls-organism_part"> ' + organism_part + '</span>';
            }
            // organism
            if (organism !== "") {
                let shortOrga = organism.replace('http://purl.obolibrary.org/obo/NCBITaxon_', '');
                strData += '<a class="mtbls-organism" href="' + organism + '" target="_blank">NCBITaxon:' + shortOrga + '</a> ';
            }
            // study_design
            if (study_design !== "") {
                strData += '<span class="mtbls-keyword">' + study_design + '</span> ';
            }
            // study_design
            if (study_factor !== "") {
                strData += '<span class="mtbls-keyword">' + study_factor + '</span> ';
            }
            strData += '<span class="mtbls-keyword">' + instrument_platform + '</span> ' +
                '</li>';
        }
        // display or hide
        if (strData !== "") {
            document.getElementById('cardSheet_metabolights').innerHTML=('<ul class="list-group">'+strData+'</ul>');
        } else {
            $('#panel-metabolights-studies').hide();
        }
    });
```

Construction rendu HTML

Rendu HTML



Source: https://developer.mozilla.org/fr/docs/Web/Web_Components

MetaboHUB & webcomponents

Groupe de travail – partage de composents applicatifs

Qu'est ce qu'un webcomponents?

- Ensemble de technologies permettant de créer des composants d'interface graphique **personnalisés** et **réutilisables.**
- Ces éléments sont **encapsulés**, **versatiles** et **sans risquer une collision avec d'autres morceaux de code**

Problématiques posées:

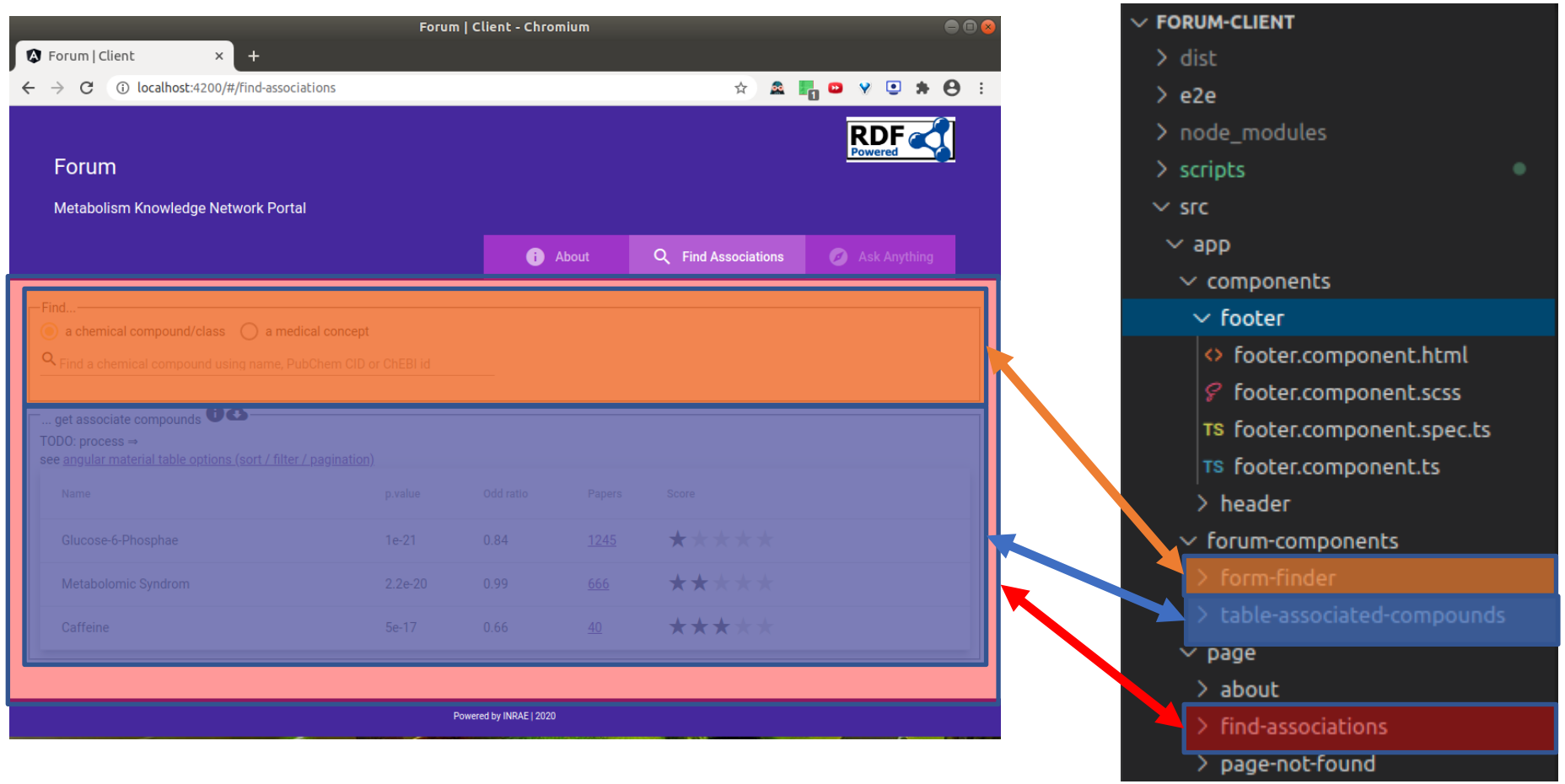
- Uniformisation des choix technologiques (vue3, typescript, **OpenAPI**)
- Uniformisation des pratiques de dev. / framework graphiques (vuetify3)
- Autre (licences, repo de partage, …)

	Pros	Cons
jQuery & …	Bien connu des dev	Conflits de lib; « lourd »;
webcomponents	Réutilisable; TS; archi micro-api;	Transpliage JS; incompatibilité vue/angular/reac

Source: https://developer.mozilla.org/fr/docs/Web/Web_Components

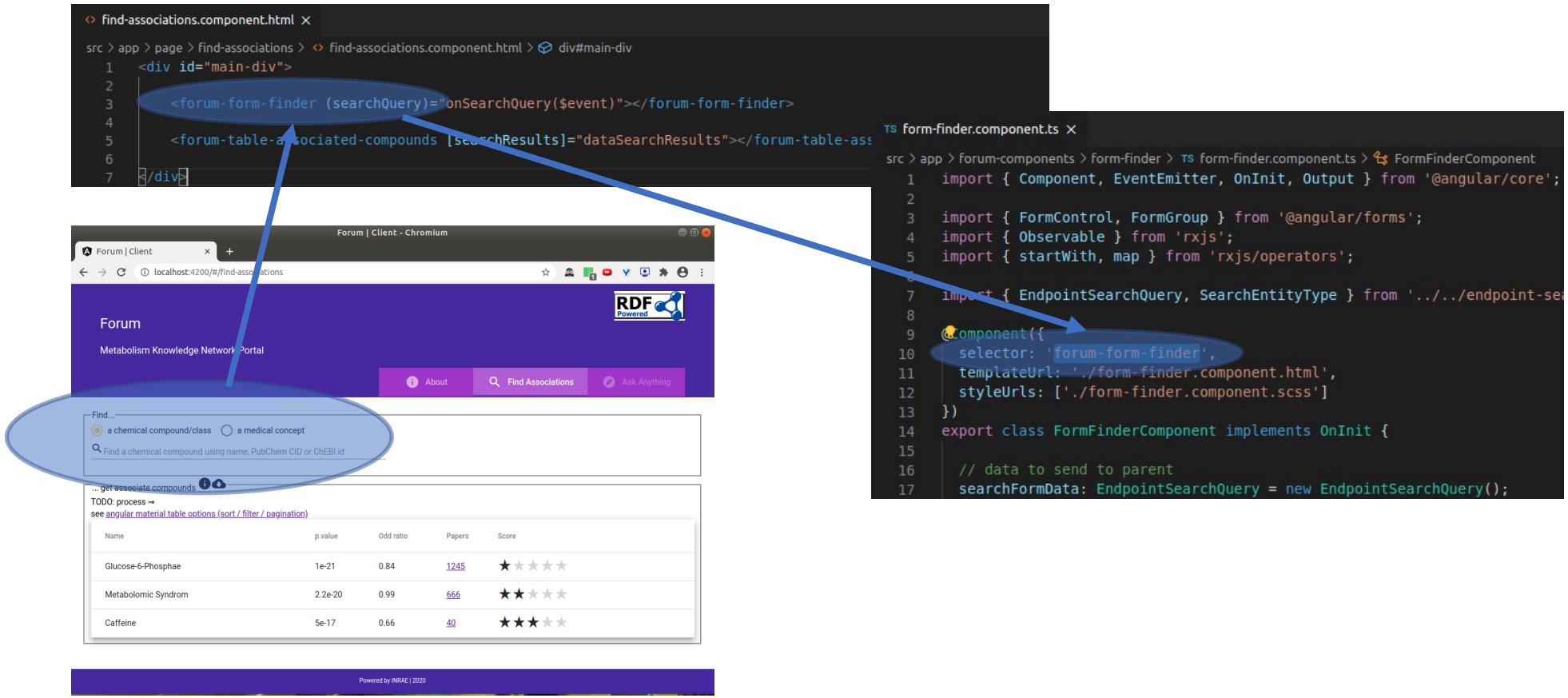
Projet Forum - Architecture

« App » architecture



Forum Client - Code

Web Components



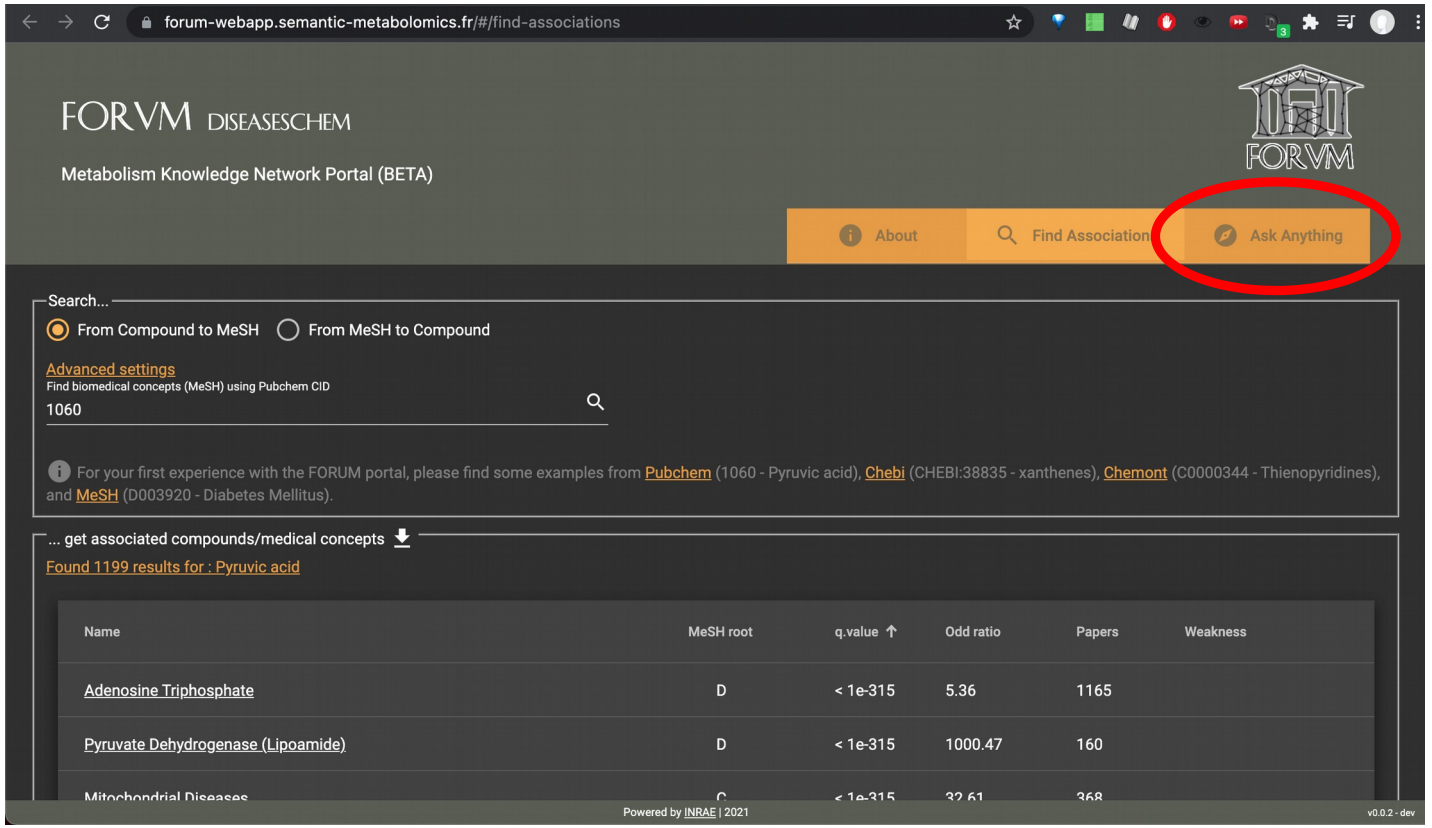
Projet Forum - Perspectives

Done

- ✓ Fuzzy search

TODO

- Afficher version base
- « Ask anything »
  - YASGUI ?
    - https://yasgui.triply.cc/
  - AskOmics ?
    - https://askomics.org/
  - Sparklis ?
    - http://www.irisa.fr/LIS/ferre/sparklis



https://forum-webapp.semantic-metabolomics.fr
https://doi.org/10.1093/bioinformatics/btab627

Questions

- **FORUM**
  - http://semantic-metabolomics.org/
  - *« FORUM: Building a Knowledge Graph from public databases and scientific literature to extract associations between chemicals and diseases » M.Delmas et al*
    https://doi.org/10.1093/bioinformatics/btab627
  - *Contact : clement.frainay@inrae.fr*

- **Metabolomics Semantic Datalake**
  - **Janvier/Février 2022 : formation Big Data (Spark/Scala) 6 demi-journée en visio**
  - *Contact : olivier.filangi@inrae.fr, franck.giacomoni@inrae.fr*

- **Discovery**
  - https://p2m2.github.io/discovery : (doc/tutoriels/templates/exemples d'application)
  - *Contact : olivier.filangi@inrae.fr*

- **Web Component**
  - *Contact : nils.paulhe@inrae.fr*

- **SPARQL Endpoints**
  - https://peakforest.semantic-metabolomics.fr/sparql/
  - https://metabolights.semantic-metabolomics.fr/sparql/
  - https://forum.semantic-metabolomics.fr/sparql/