INRAE

Atelier Big Data INRAE
10-12 janv. 2023 Sète (France)

Semantic web uses case in metabolomics

*S. Aubin, D. Benaben, M. Boudet, C.Dupérier, **O. Filangi**, C.Frainay, F. Giacomoni, M. Delmas, N. Paulhe, F. Vinson, M. Weber*

P2M2 Plateau de Profilage Métabolique & de Métabolomique

Toxalim Research Centre in Food Toxicology

Bordeaux Metabolome

Plate-Forme Exploration du Metabolisme des gènes aux métabolites

DipSO

METABOHUB

Context

- 2019 FORUM (C. Frainay, F. Giacomoni) – Financement 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

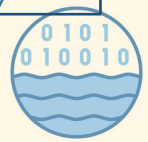
- 2021-2025 MetaboHUB 2 .0 - Infrastructure nationale de métabolomique et fluxomique
  - WP5 "Creating FAIR e-resources for knowledge mining"
    - Task 4 Linked data for advanced knowledge extraction

Projets

2019

Construire un graphe de connaissances à partir de bases de données publiques et de littérature scientifique pour extraire les associations entre les produits chimiques et les maladies

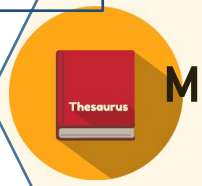
2021



Metabolomics Semantic Datalake

Mettre à disposition une infrastructure distribuée, orientée « Big Data de type « Semantic Datalake Engine » pour l'intégration massive de contenus sémantiques (ontologies et graphes de données) et hétérogènes en métabolomique

2022



Metasaurus

Bâtir un modèle de connaissance adapté aux études métabolomiques en s'appuyant sur les ontologies, thésaurus et vocabulaires contrôlés publiés

FORUM : Building a knowledge graph from public databases and scientific literature to extract associations between chemicals and diseases

Florence Vinson

Clément Frainay

Maxime Delmas

Thèse M. Delmas : « Construire, exploiter et étendre un graphe de connaissances pour l'étude des liens entre métabolisme et santé »

FORVM

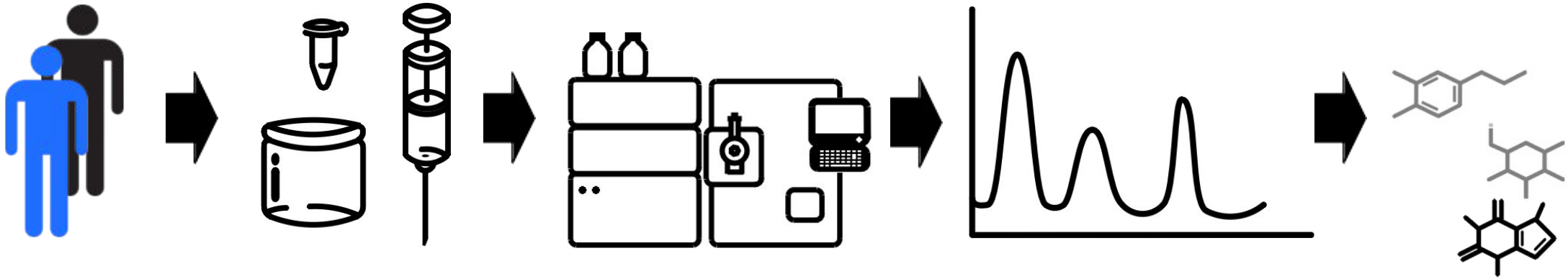
Olivier Filangi

Franck Giacomoni

Nils Paulhe

Christophe Dupérier

FORUM : Building a knowledge graph from public databases and scientific literature to extract associations between chemicals and diseases

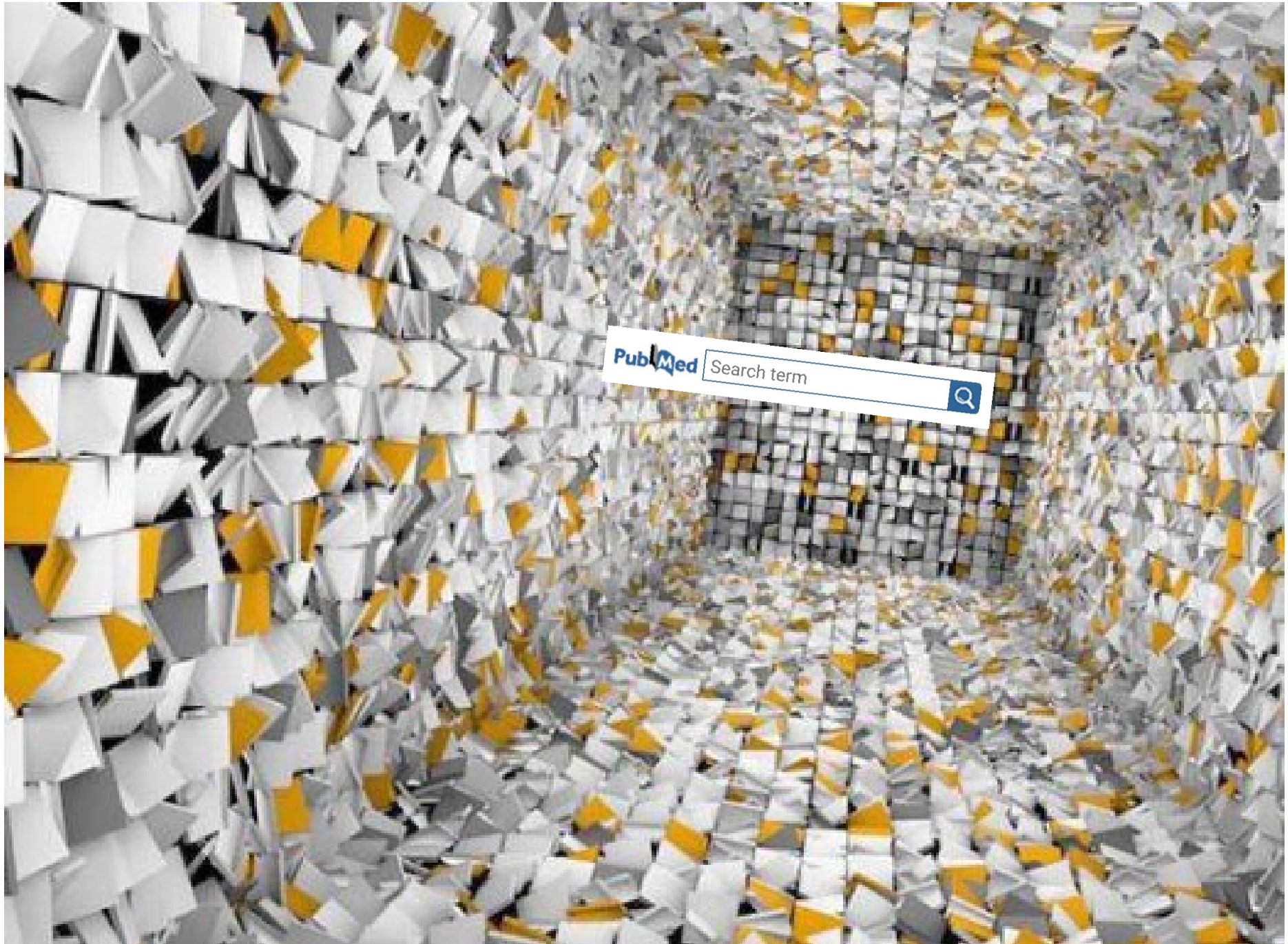


Context :

Metabolomics

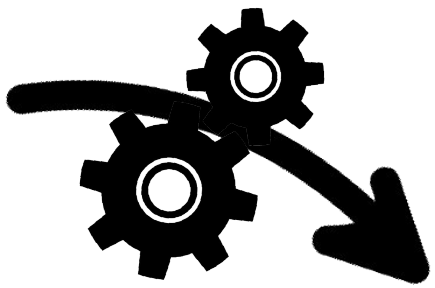
mechanistic

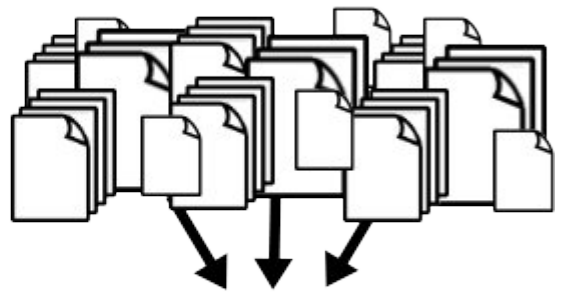
interpretation


PubMed
Search term

Information Overload

Knowledge representation





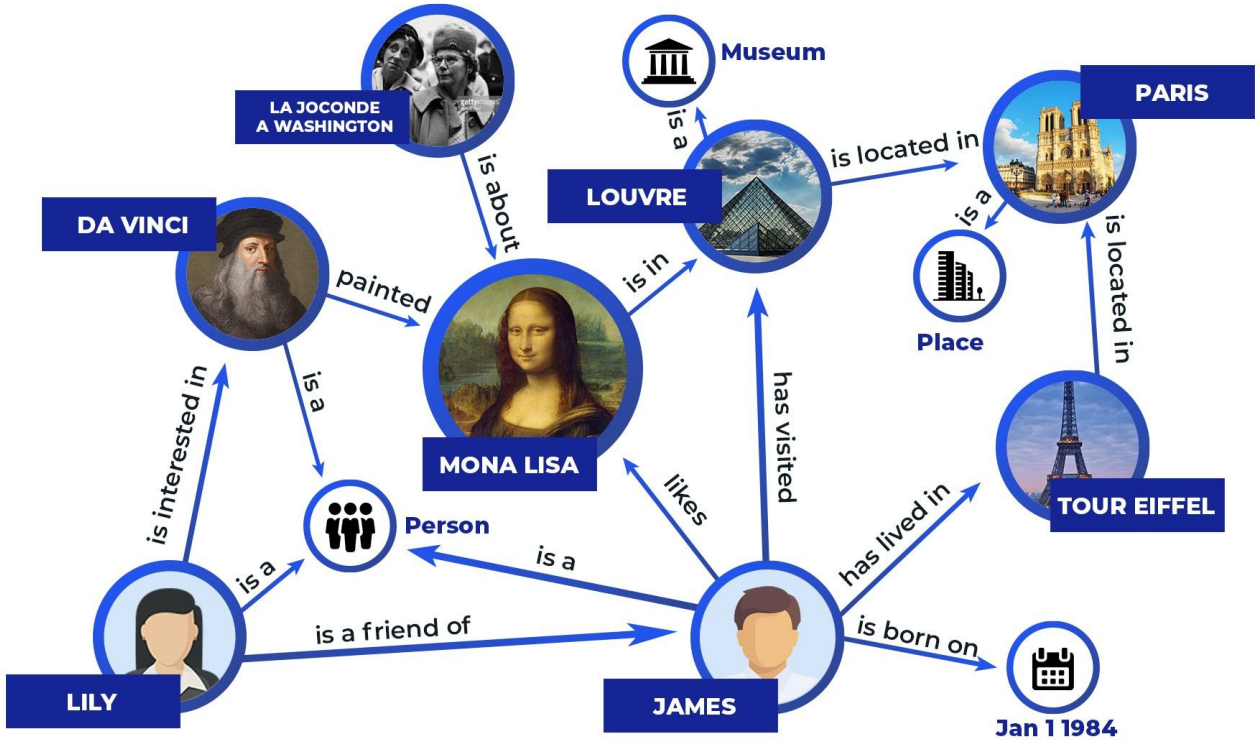
Automated reasoning

FORUM : Building a knowledge graph from public databases and scientific literature to extract associations between chemicals and diseases

What's a Knowledge Graph?

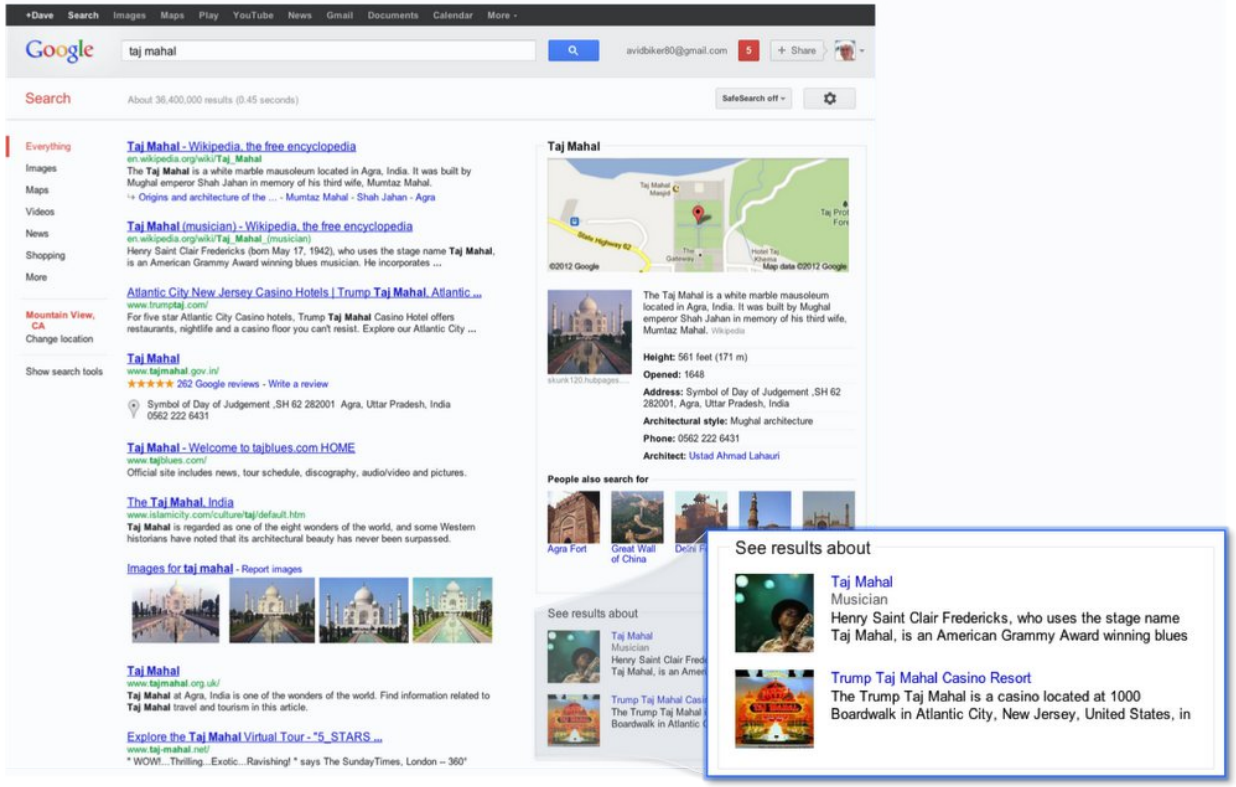
formal & structured representation of knowledge

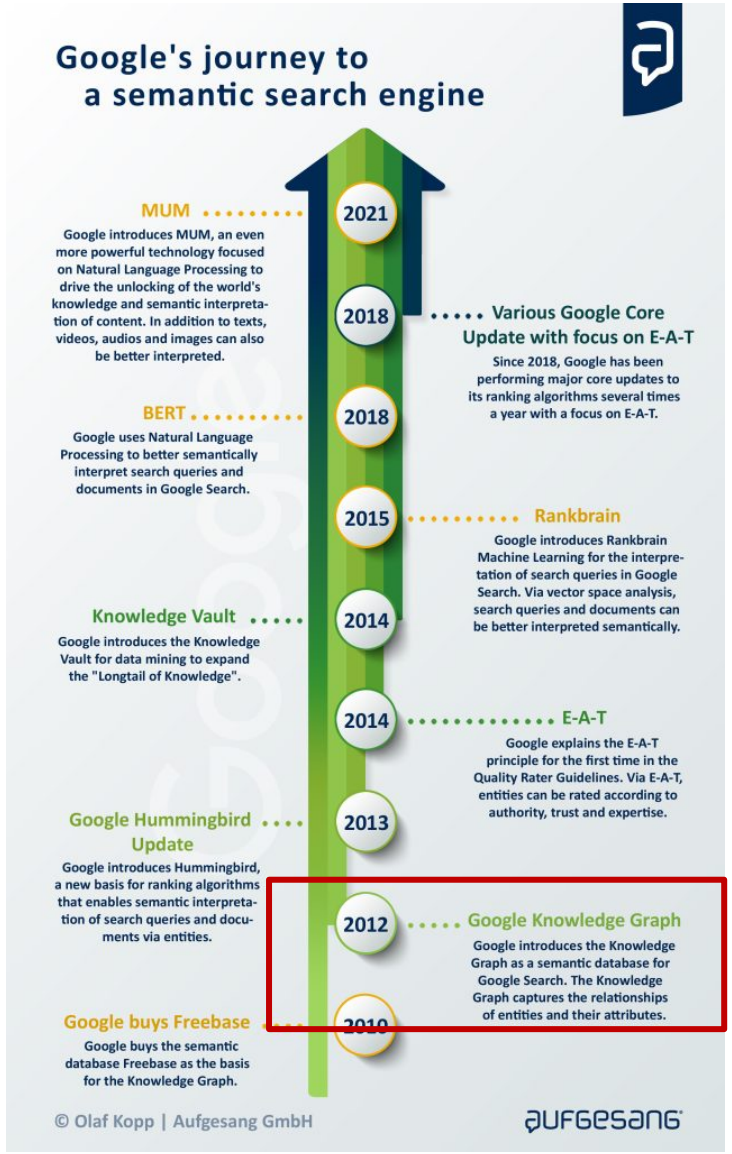
entities
+
relations
+
semantic descriptions
+
rules (axiomatic knowledge)



2012 - Knowledge Graph into Google Search: « things, not strings »

Do you mean Taj Mahal the monument, or Taj Mahal the musician?



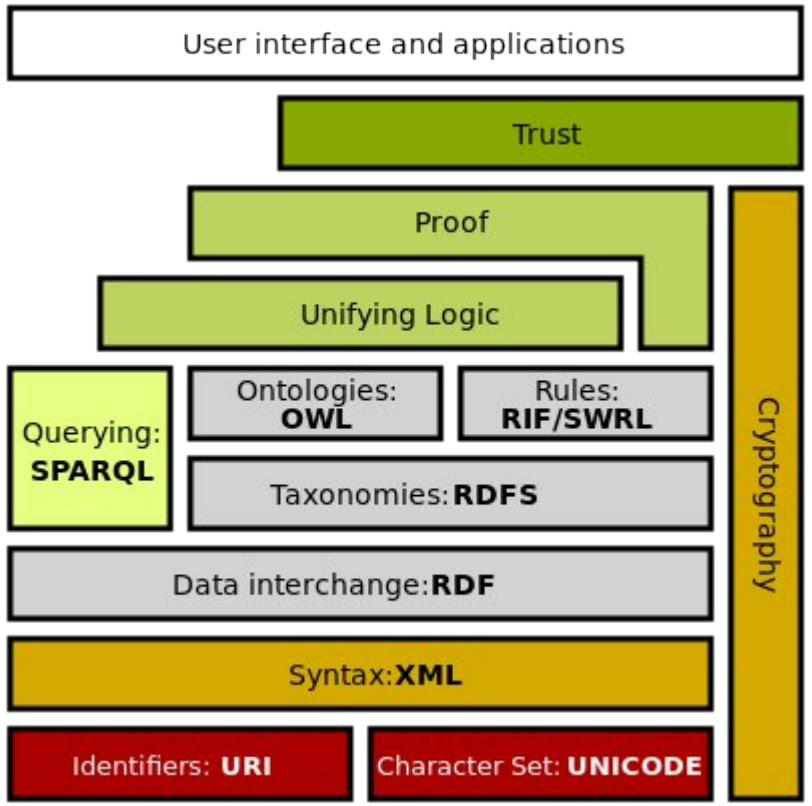


FORUM : Building a knowledge graph from public databases and scientific literature to extract associations between chemicals and diseases

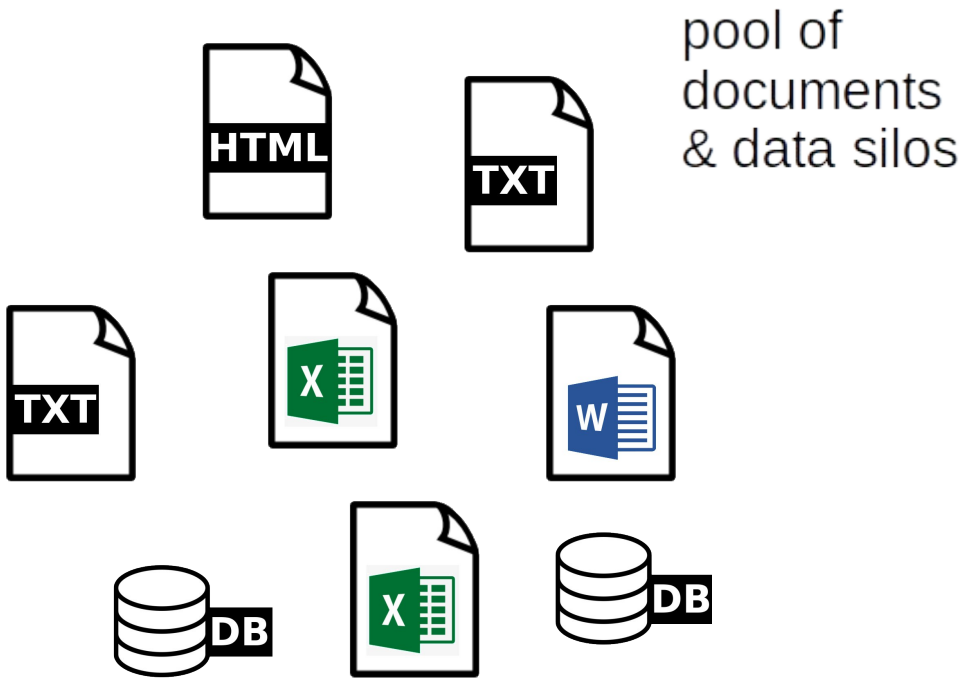
Semantic Web

https:/ / www.w3.org/ standards/ semanticweb/

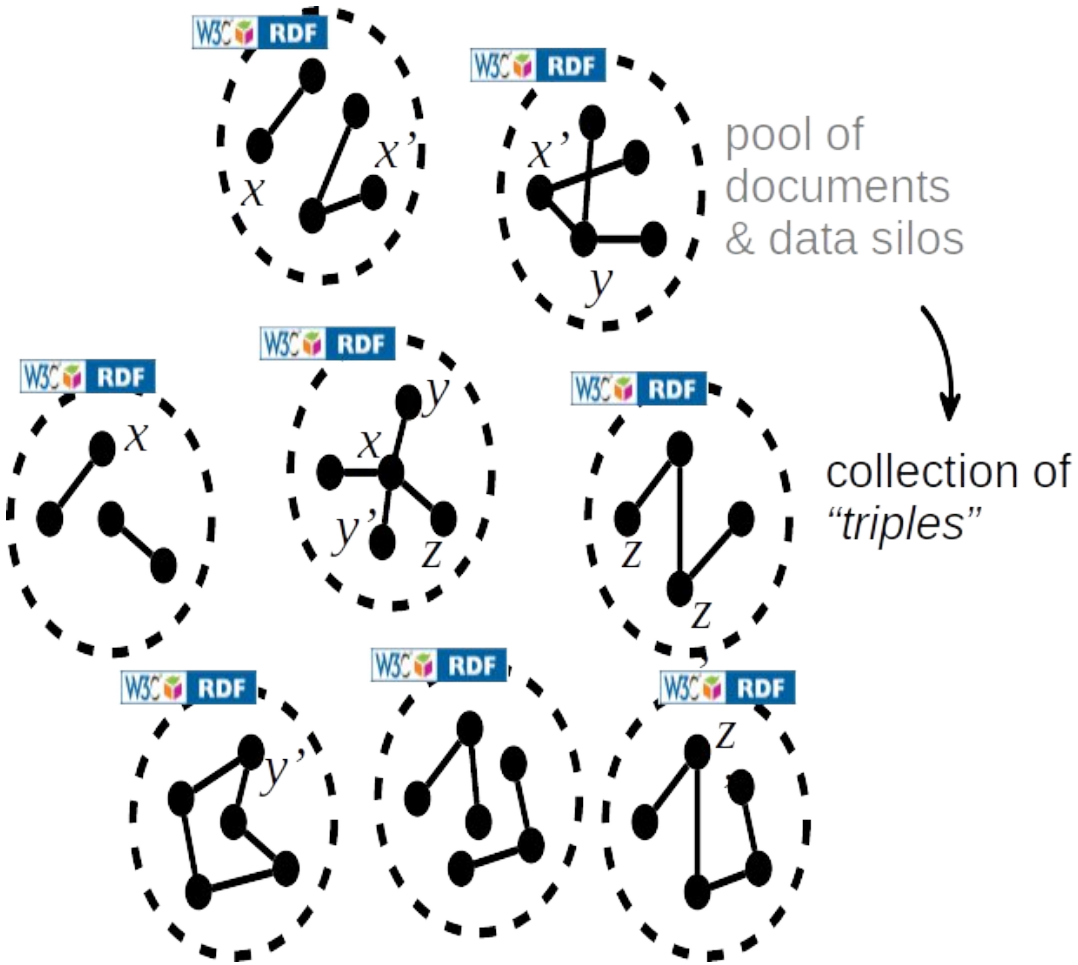
Semantic Web Stack



The semantic web : A technology stack

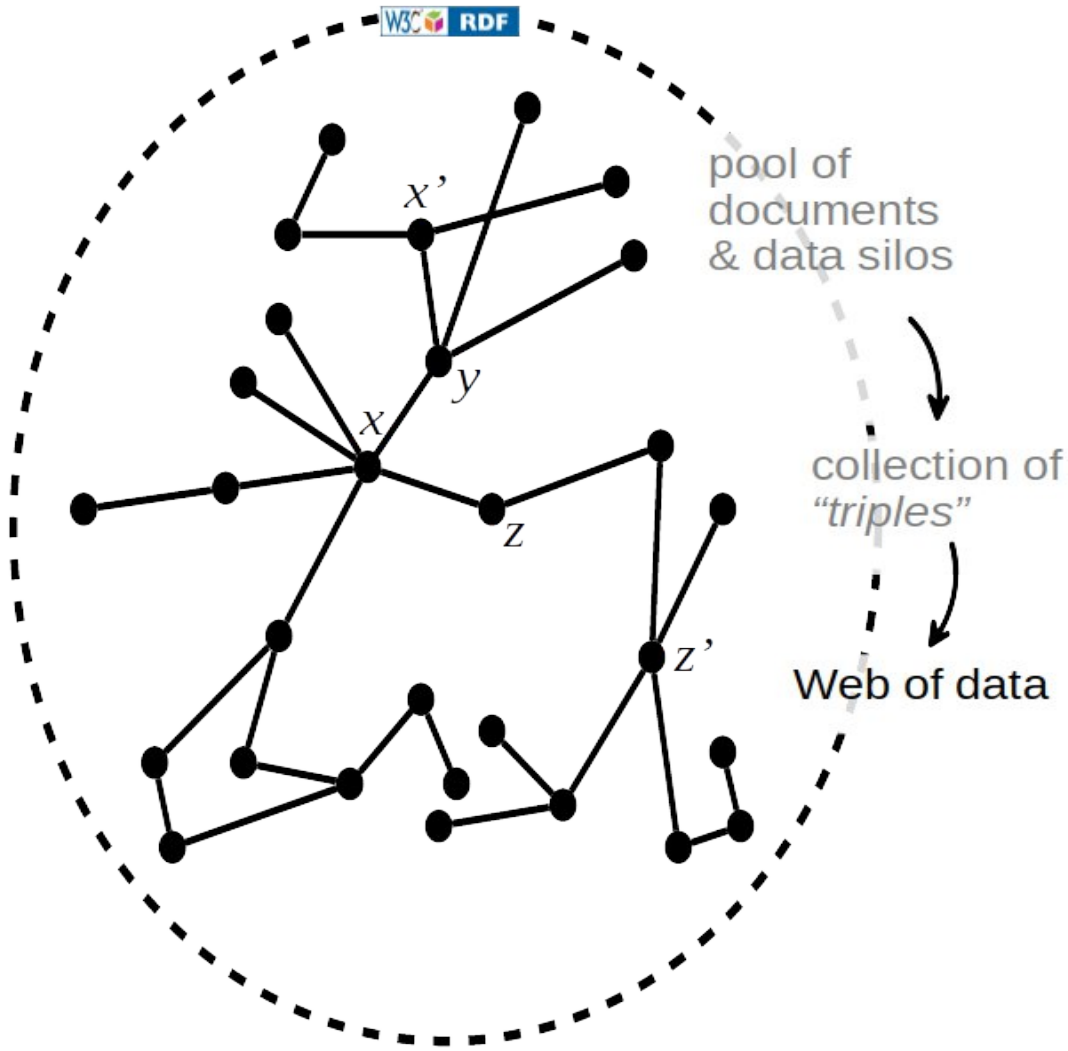


The semantic web : A technology stack



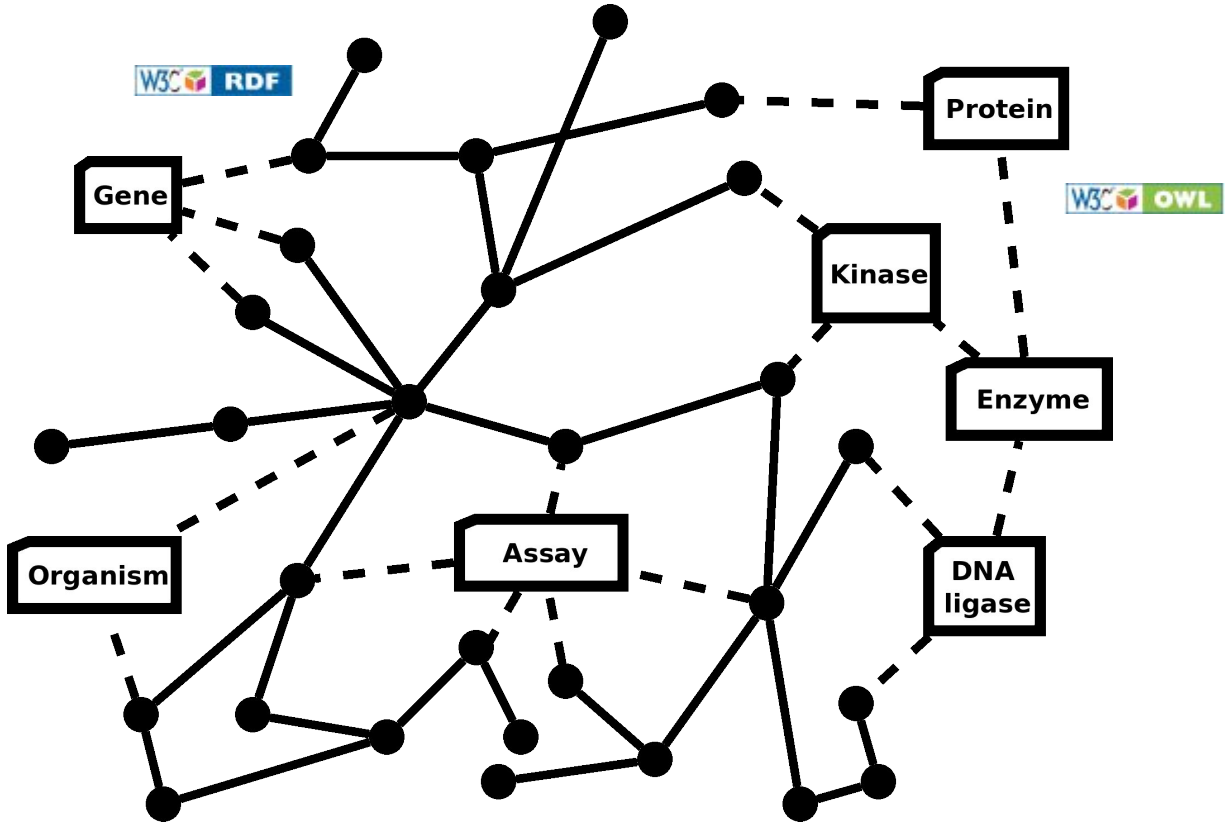
- **A framework for unified description of documents' content (RDF)**

The semantic web : A technology stack



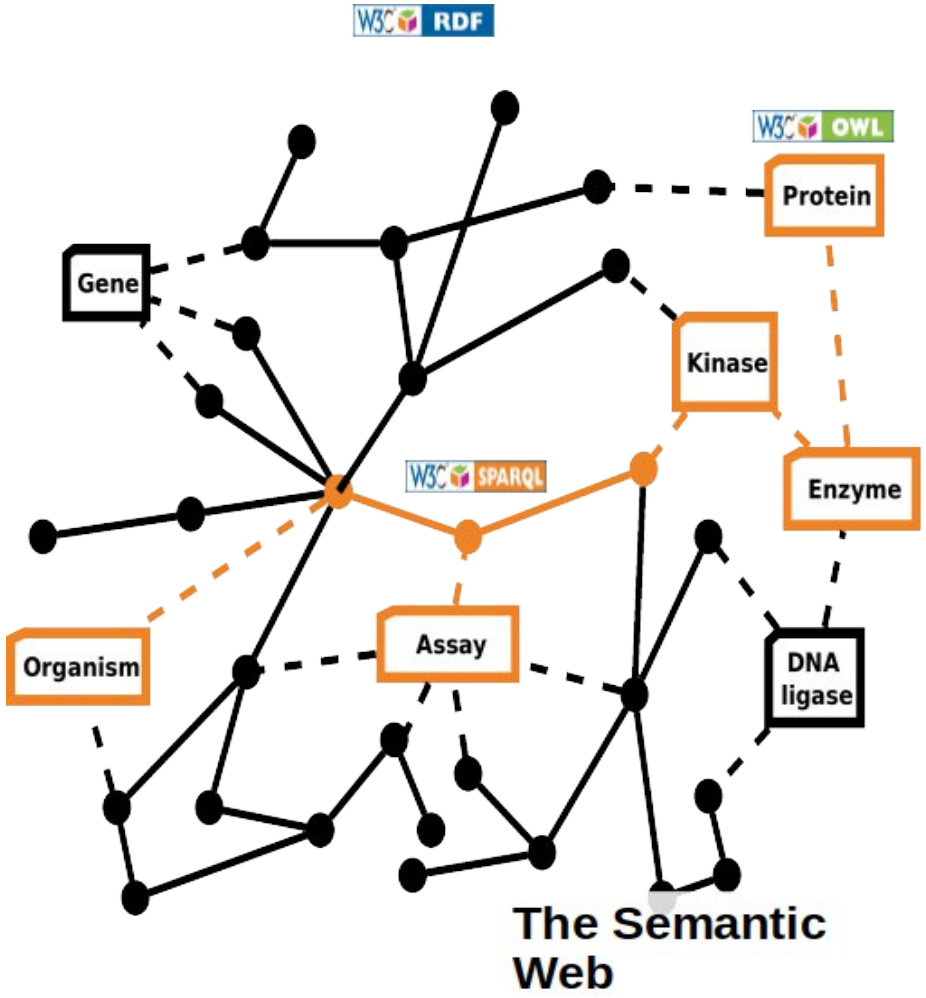
- A framework for unified description of documents' content (RDF)
- **A system for uniquely identify things (URI)**

The semantic web : A technology stack



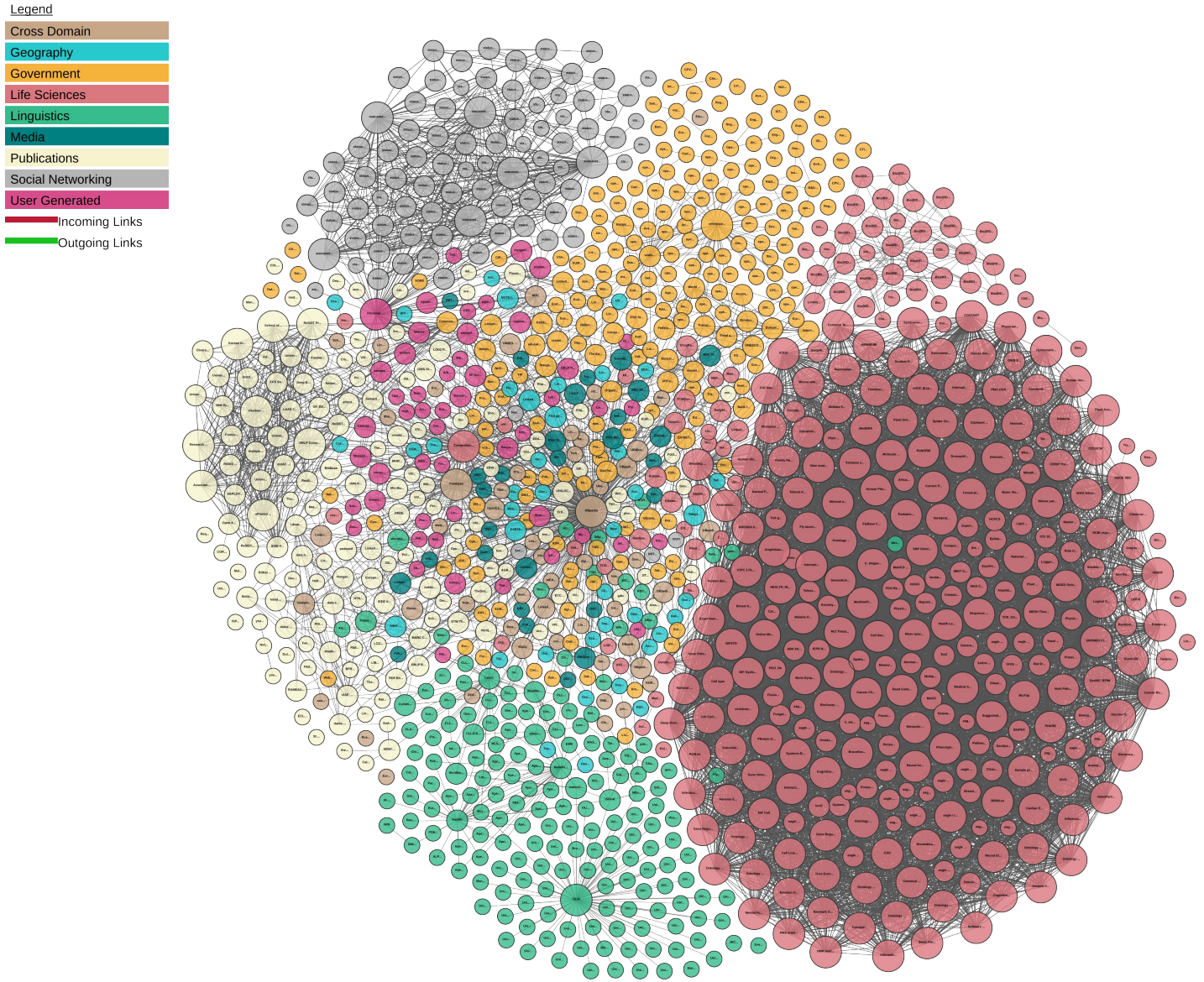
- A framework for unified description of documents' content (RDF)
- A system for uniquely identify things (URI)
- **A semantic description of data using ontologies (OWL)**

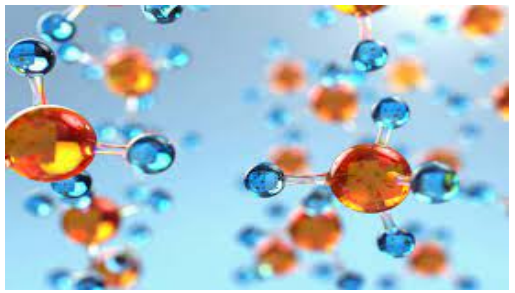
The semantic web : A technology stack



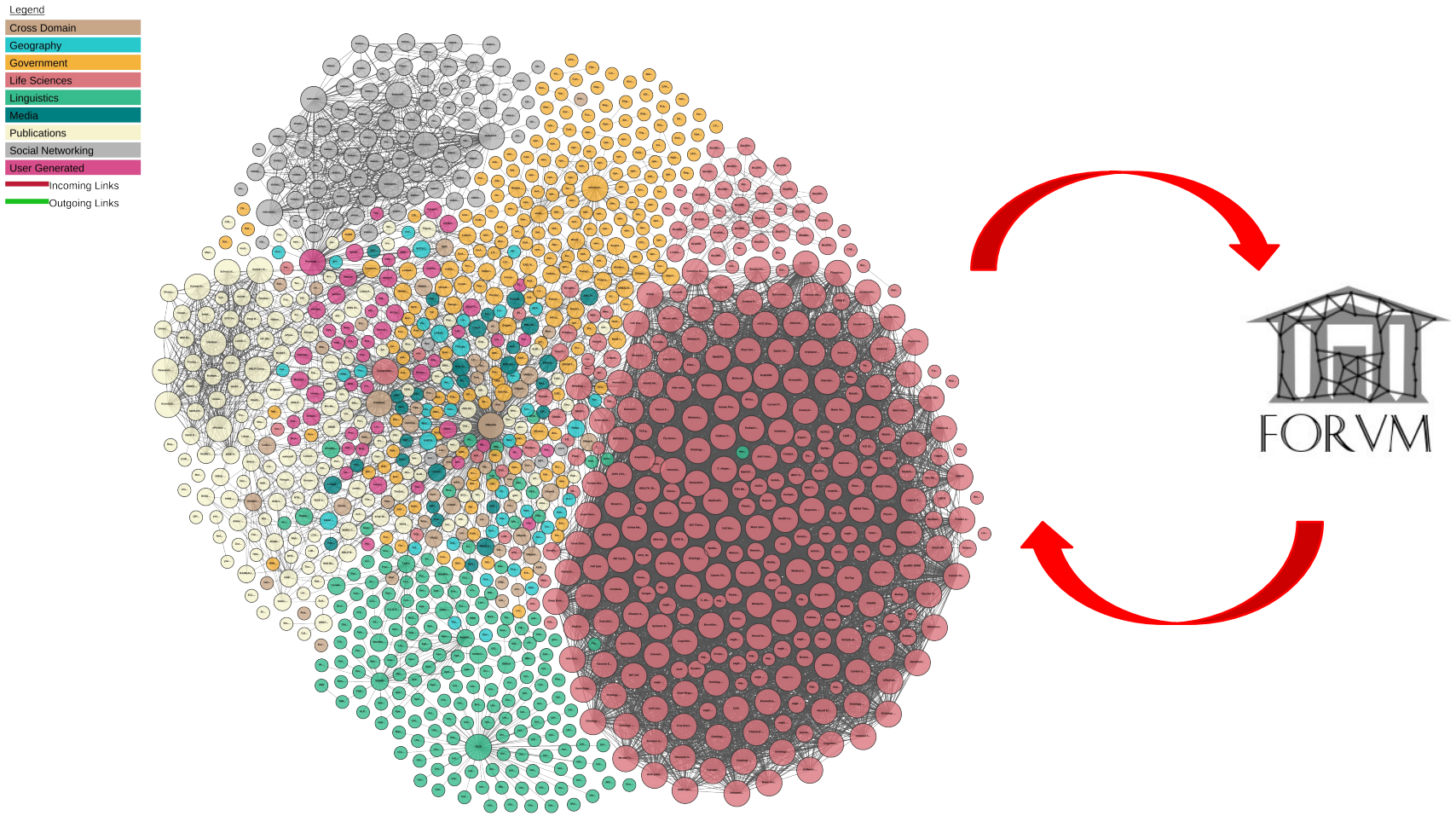
- A framework for unified description of documents' content (RDF)
- A system for uniquely identify things (URI)
- A semantic description of data using ontologies (OWL)
- **...feeding powerful query engine (SPARQL)**

The semantic web : A web of Linked Data

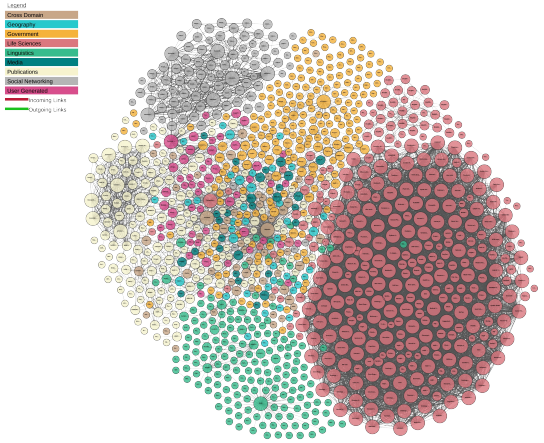




FORUM : Building a knowledge graph from public databases and scientific literature to extract associations between chemicals and diseases



Scientific literature as data



NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed | Advanced

Format: Abstract | Send to

Liver Int. 2015 Dec;35(12):2487-94. doi: 10.1111/liv.12956. Epub 2015 Sep 28.

Current diagnosis and management of post-transjugular intrahepatic portosystemic shunt refractory hepatic encephalopathy.

Pereira K[1], Carrion AF[2], Martin P[2], Vaheesan K[3], Salsamendi J[1], Doshi M[1], Yrizarry JM[1].

Author information

Abstract

Transjugular intrahepatic portosystemic shunt has evolved into an important option for management of complications of portal hypertension. The use of polytetrafluoroethylene covered stents enhances shunt patency. Hepatic encephalopathy (HE) remains a significant problem after TIPS placement. The approach to management of patients with refractory hepatic encephalopathy typically requires collaboration between different specialties. Patient selection for TIPS requires careful evaluation of risk factors for HE. TIPS procedure-related technical factors like stent size, attention to portosystemic pressure gradient reduction and use of adjunctive variceal embolization maybe important. Conservative medical therapy in combination with endovascular therapies often results in resolution or substantial reduction of symptoms. Liver transplantation is, however, the ultimate treatment.

KEYWORDS: hepatic encephalopathy; portal hypertension; spontaneous portosystemic shunt; transjugular intrahepatic portosystemic shunt

PMID: 26332169 DOI: 10.1111/liv.12956

[PubMed - indexed for MEDLINE]

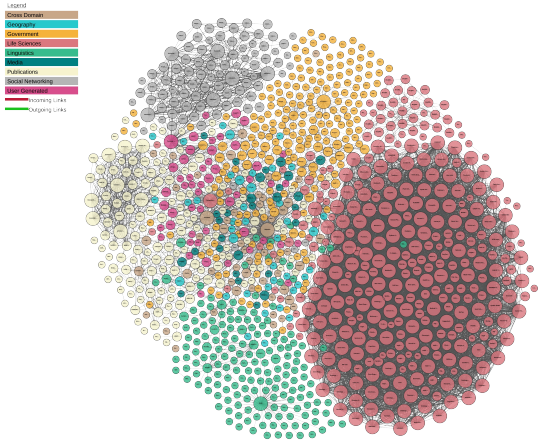
Publication Types, MeSH Terms, Substances

Publication Types
Review

MeSH Terms
Disease Management
Hepatic Encephalopathy*/diagnosis
Hepatic Encephalopathy*/etiology
Hepatic Encephalopathy*/prevention & control
Humans
Hypertension, Portal/surgery*
Polytetrafluoroethylene/pharmacology
Portasystemic Shunt, Transjugular Intrahepatic/adverse effects*
Portasystemic Shunt, Transjugular Intrahepatic/instrumentation
Portasystemic Shunt, Transjugular Intrahepatic/methods
Risk Adjustment

Substances
Polytetrafluoroethylene

Scientific literature as data



Metadata is everywhere !

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed Advanced

Format: Abstract Send to

Liver Int. 2015 Dec;35(12):2487-94. doi: 10.1111/liv.12956. Epub 2015 Sep 28.

Current diagnosis and management of post-transjugular intrahepatic portosystemic shunt refractory hepatic encephalopathy.

Pereira K[1], Carrion AF[2], Martin P[2], Vaheesan K[3], Salsamendi J[1], Doshi M[1], Yrizarry JM[1].

Author information

Abstract

Transjugular intrahepatic portosystemic shunt has evolved into an important option for management of complications of portal hypertension. The use of polytetrafluoroethylene covered stents enhances shunt patency. Hepatic encephalopathy (HE) remains a significant problem after TIPS placement. The approach to management of patients with refractory hepatic encephalopathy typically requires collaboration between different specialties. Patient selection for TIPS requires careful evaluation of risk factors for HE. TIPS procedure-related technical factors like stent size, attention to portosystemic pressure gradient reduction and use of adjunctive variceal embolization maybe important. Conservative medical therapy in combination with endovascular therapies often results in resolution or substantial reduction of symptoms. Liver transplantation is, however, the ultimate treatment.

KEYWORDS: hepatic encephalopathy; portal hypertension; spontaneous portosystemic shunt; transjugular intrahepatic portosystemic shunt

PMID: 26332169 DOI: 10.1111/liv.12956

[PubMed - indexed for MEDLINE]

Publication Types, MeSH Terms, Substances

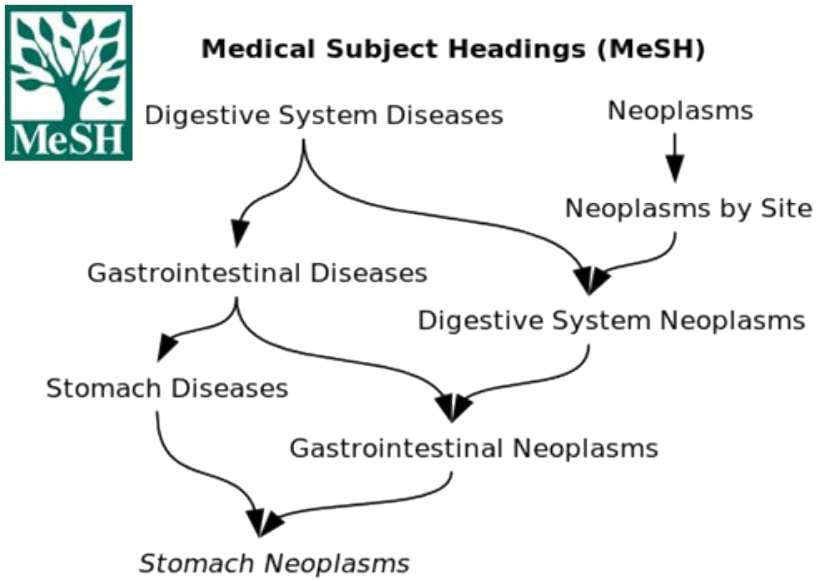
Publication Types
Review

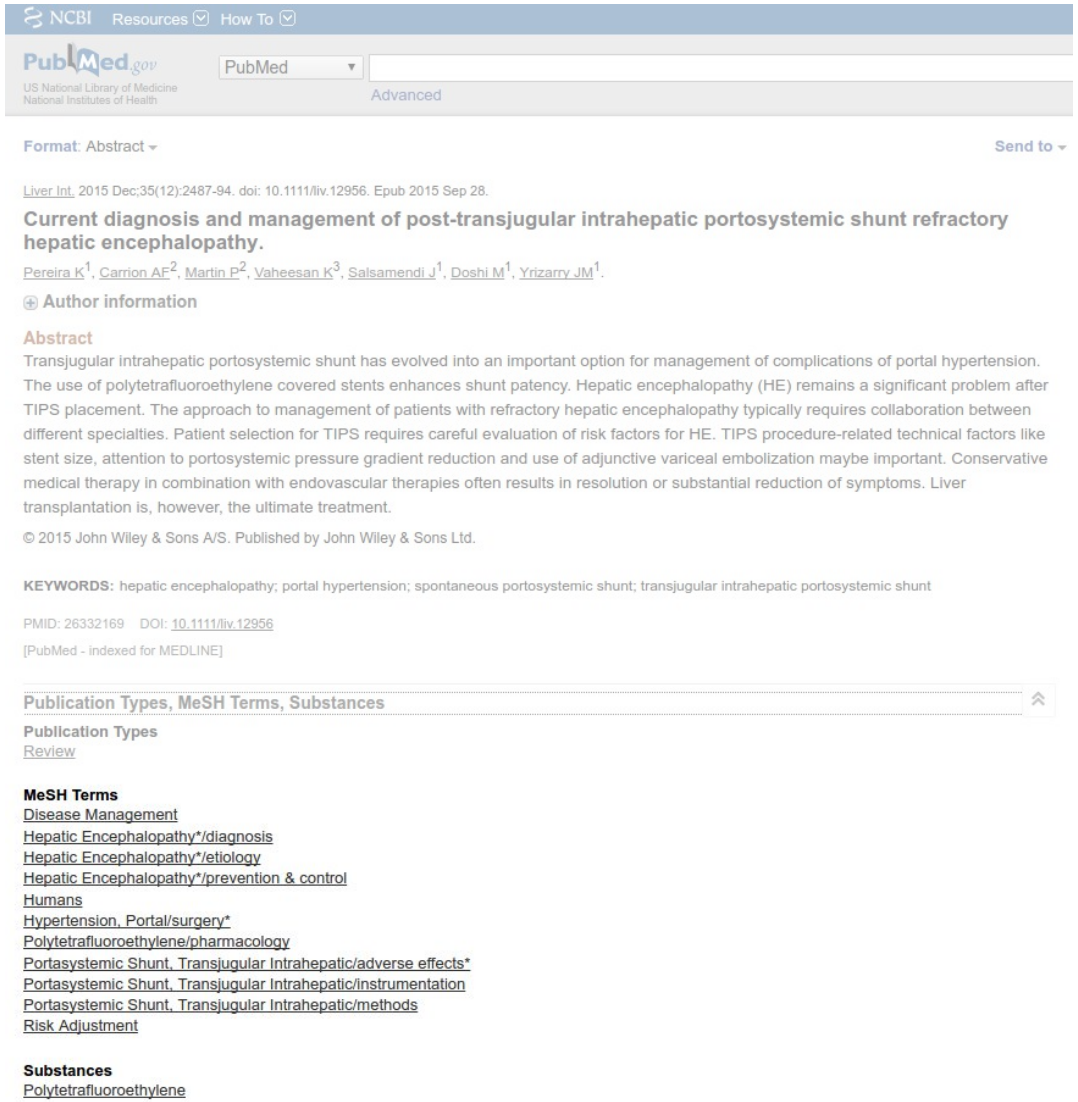
MeSH Terms
Disease Management
Hepatic Encephalopathy*/diagnosis
Hepatic Encephalopathy*/etiology
Hepatic Encephalopathy*/prevention & control
Humans
Hypertension, Portal/surgery*
Polytetrafluoroethylene/pharmacology
Portasystemic Shunt, Transjugular Intrahepatic/adverse effects*
Portasystemic Shunt, Transjugular Intrahepatic/instrumentation
Portasystemic Shunt, Transjugular Intrahepatic/methods
Risk Adjustment

Substances
Polytetrafluoroethylene

Scientific literature as data

Semantic Description





NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed Advanced

Format: Abstract Send to

Liver Int. 2015 Dec;35(12):2487-94. doi: 10.1111/liv.12956. Epub 2015 Sep 28.

Current diagnosis and management of post-transjugular intrahepatic portosystemic shunt refractory hepatic encephalopathy.

Pereira K[1], Carrion AF[2], Martin P[2], Vaheesan K[3], Salsamendi J[1], Doshi M[1], Yrizarry JM[1].

Author information

Abstract

Transjugular intrahepatic portosystemic shunt has evolved into an important option for management of complications of portal hypertension. The use of polytetrafluoroethylene covered stents enhances shunt patency. Hepatic encephalopathy (HE) remains a significant problem after TIPS placement. The approach to management of patients with refractory hepatic encephalopathy typically requires collaboration between different specialties. Patient selection for TIPS requires careful evaluation of risk factors for HE. TIPS procedure-related technical factors like stent size, attention to portosystemic pressure gradient reduction and use of adjunctive variceal embolization maybe important. Conservative medical therapy in combination with endovascular therapies often results in resolution or substantial reduction of symptoms. Liver transplantation is, however, the ultimate treatment.

KEYWORDS: hepatic encephalopathy; portal hypertension; spontaneous portosystemic shunt; transjugular intrahepatic portosystemic shunt

PMID: 26332169 DOI: 10.1111/liv.12956

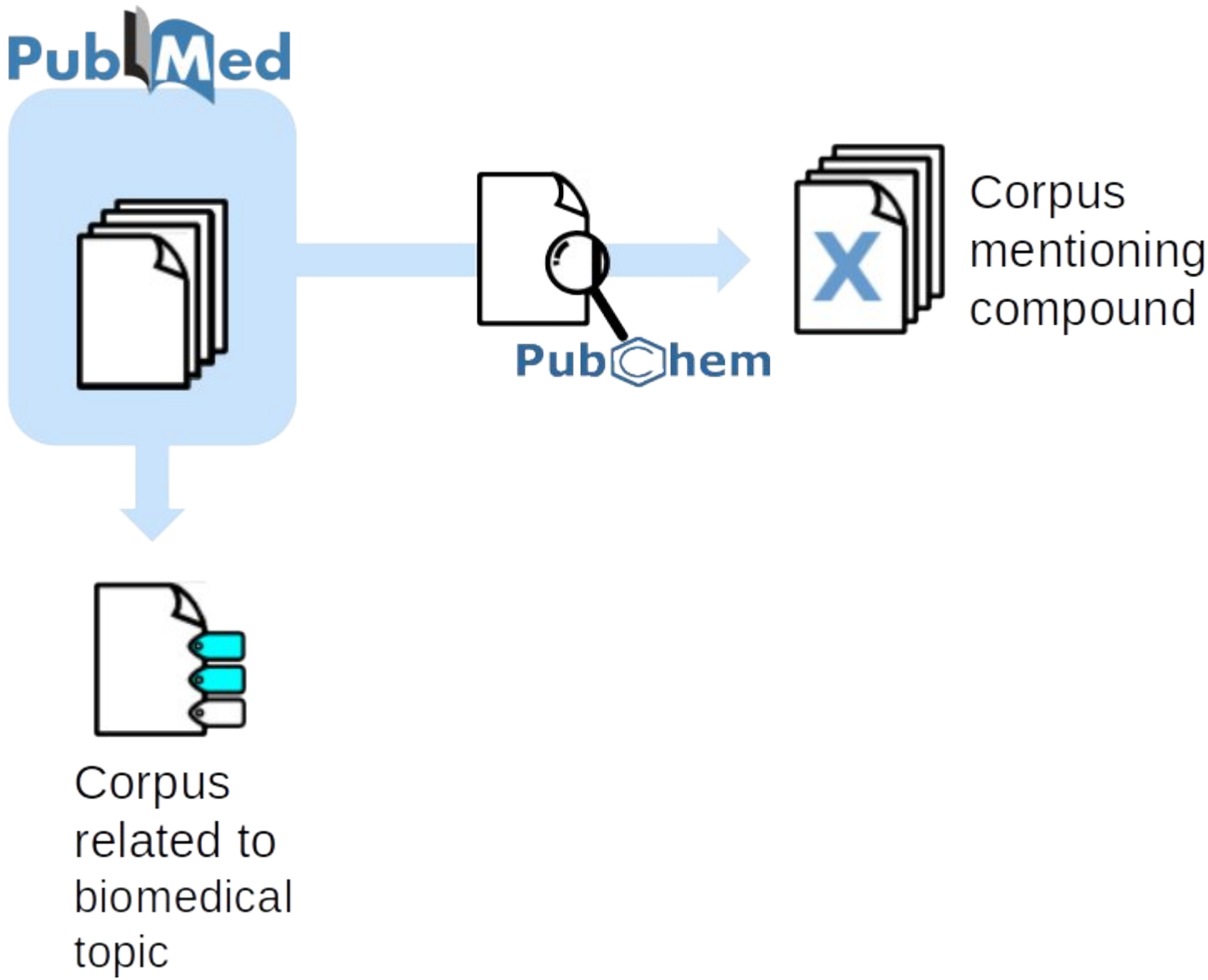
[PubMed - indexed for MEDLINE]

Publication Types, MeSH Terms, Substances

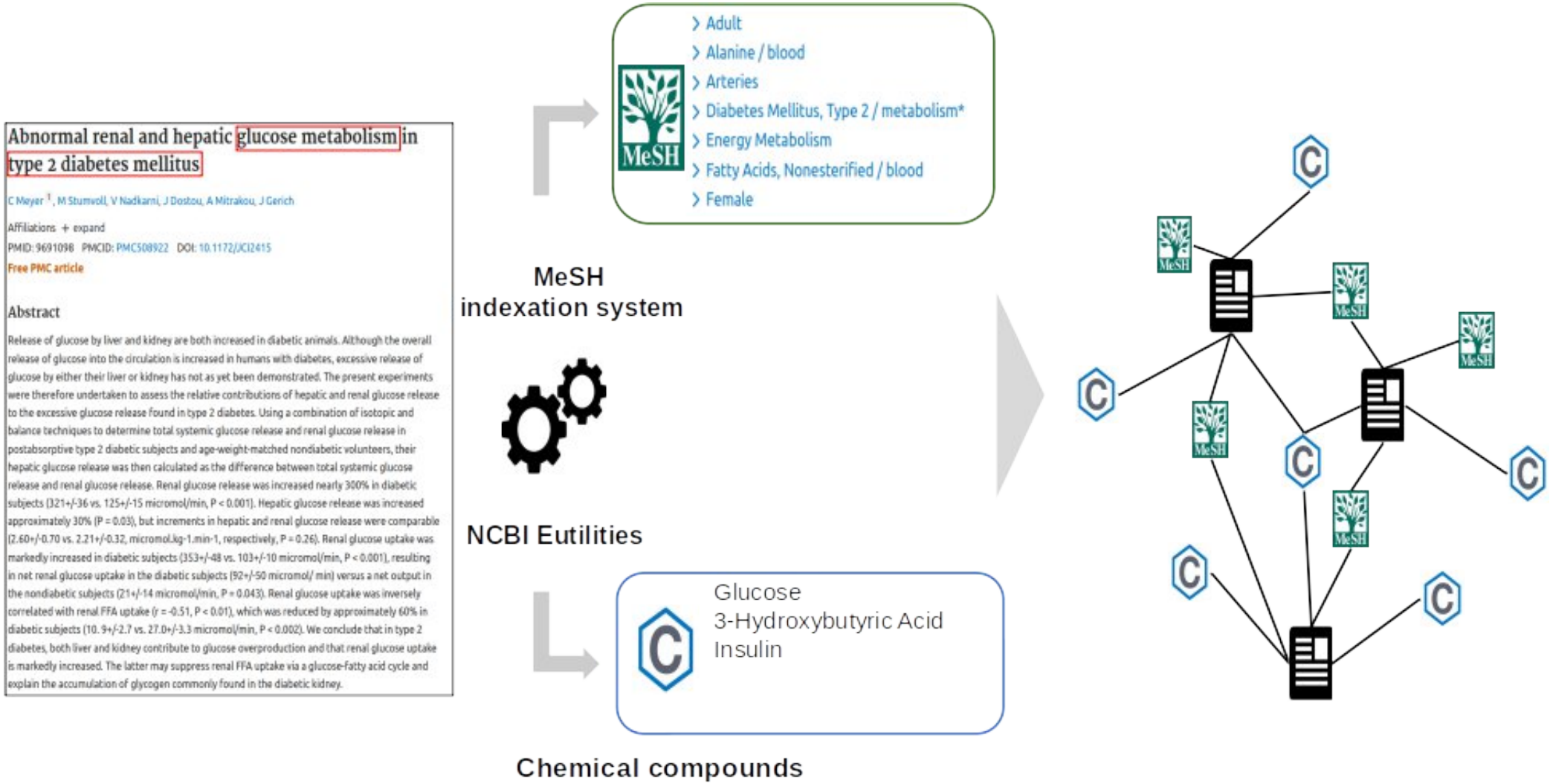
Publication Types
Review

MeSH Terms
Disease Management
Hepatic Encephalopathy*/diagnosis
Hepatic Encephalopathy*/etiology
Hepatic Encephalopathy*/prevention & control
Humans
Hypertension, Portal/surgery*
Polytetrafluoroethylene/pharmacology
Portasystemic Shunt, Transjugular Intrahepatic/adverse effects*
Portasystemic Shunt, Transjugular Intrahepatic/instrumentation
Portasystemic Shunt, Transjugular Intrahepatic/methods
Risk Adjustment

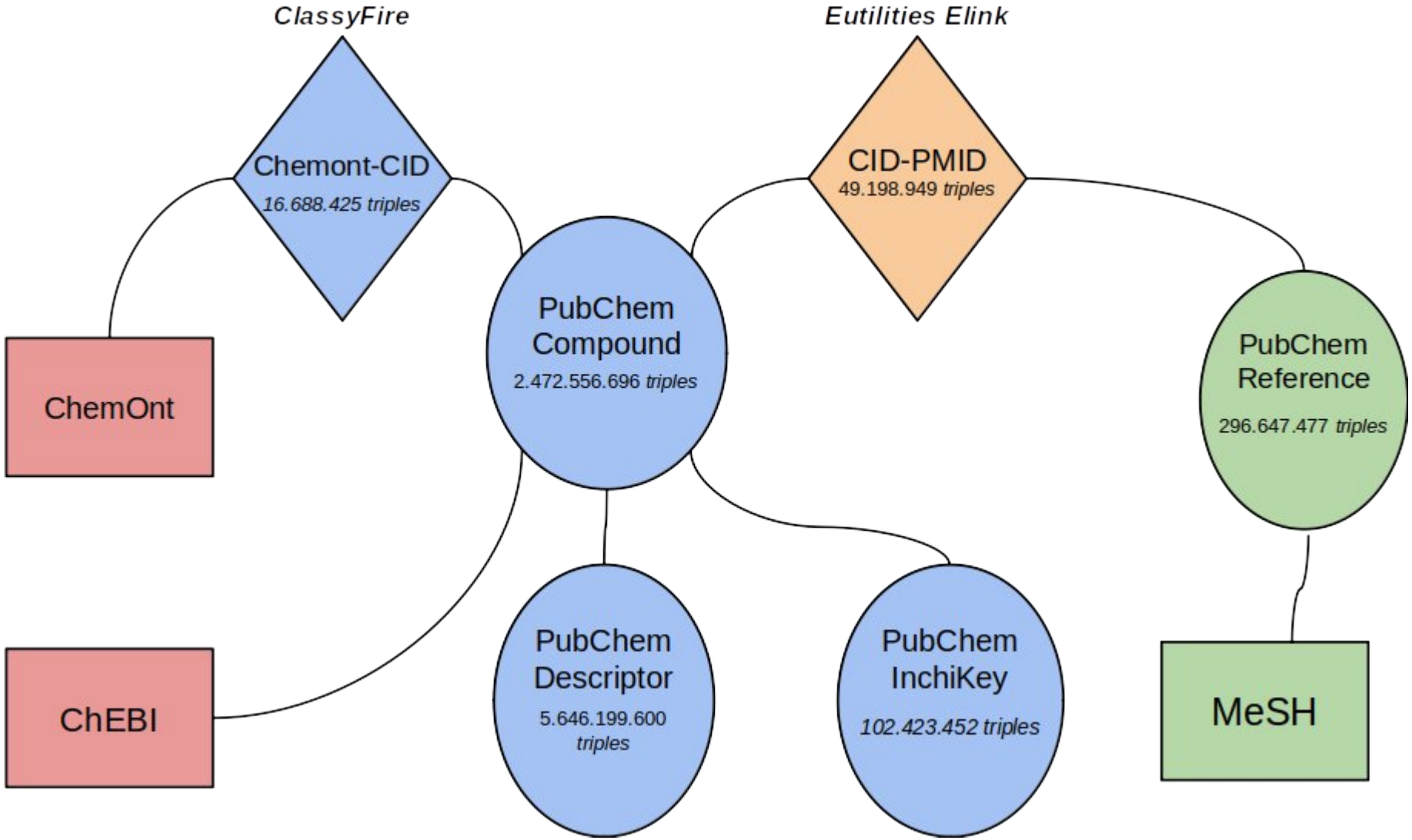
Substances
Polytetrafluoroethylene


PubMed
PubChem
X
Corpus mentioning compound
Corpus related to biomedical topic

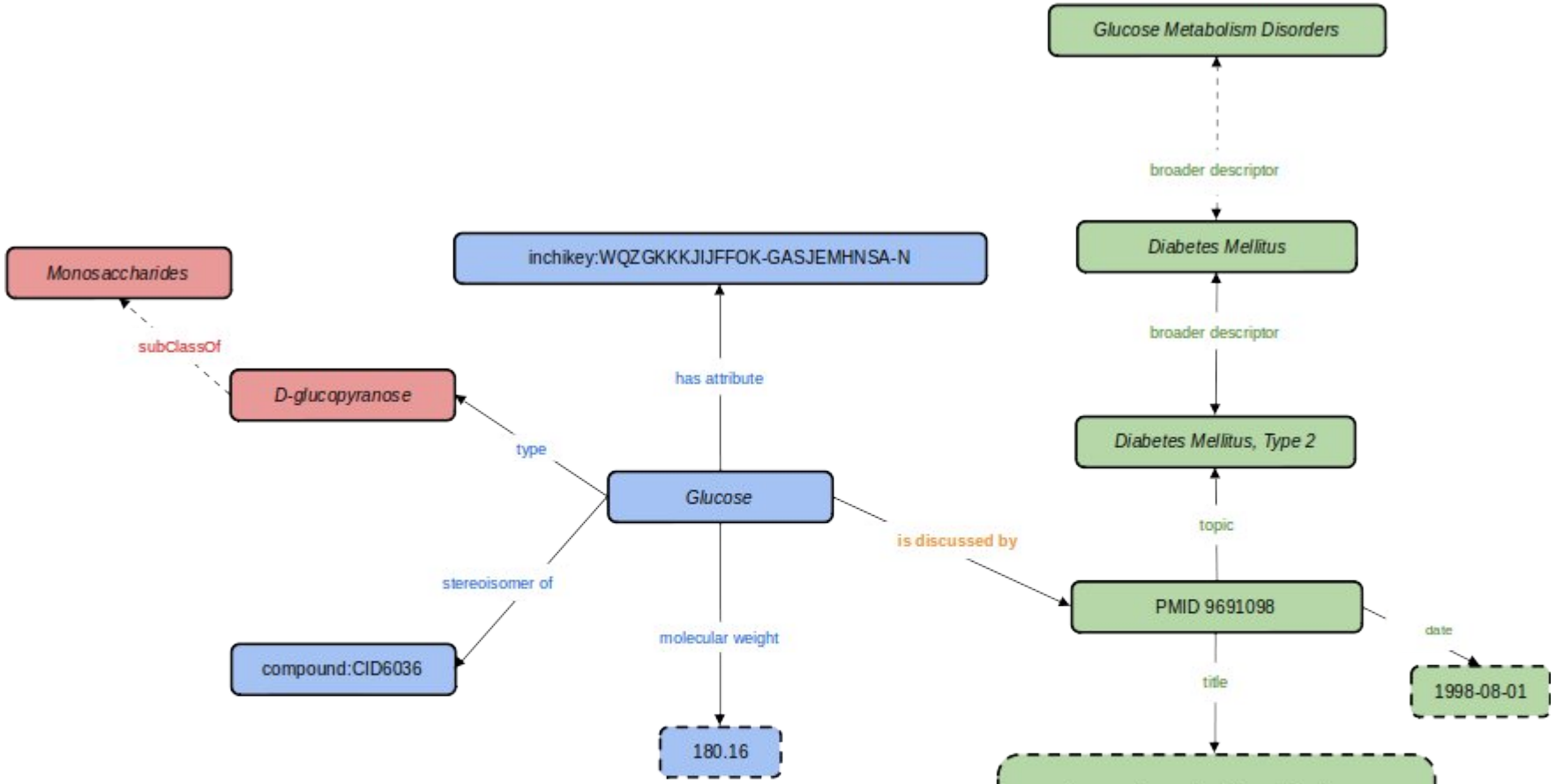
Biomedical Concepts



The FORUM Knowledge Network

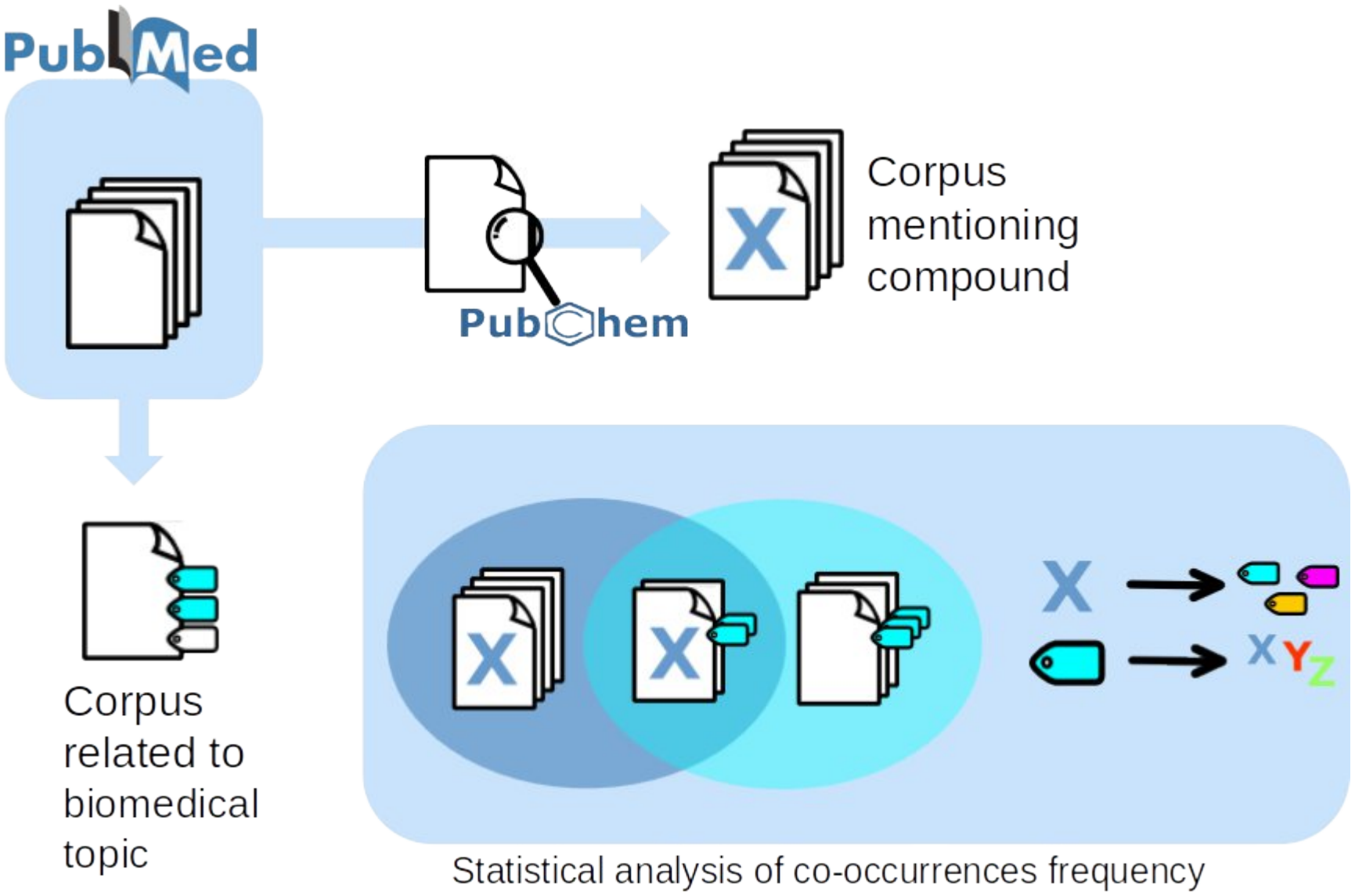


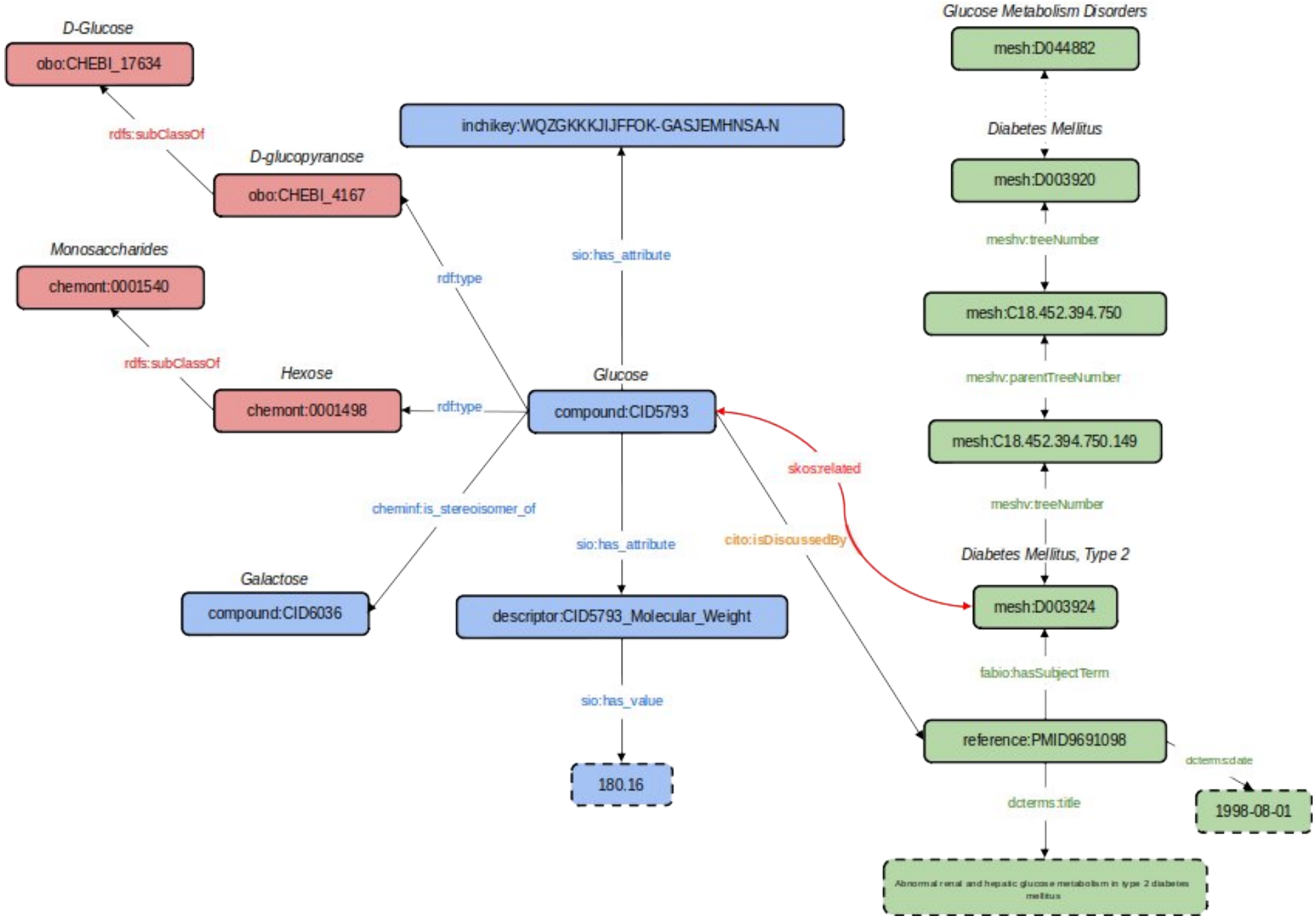
The FORUM Knowledge Network



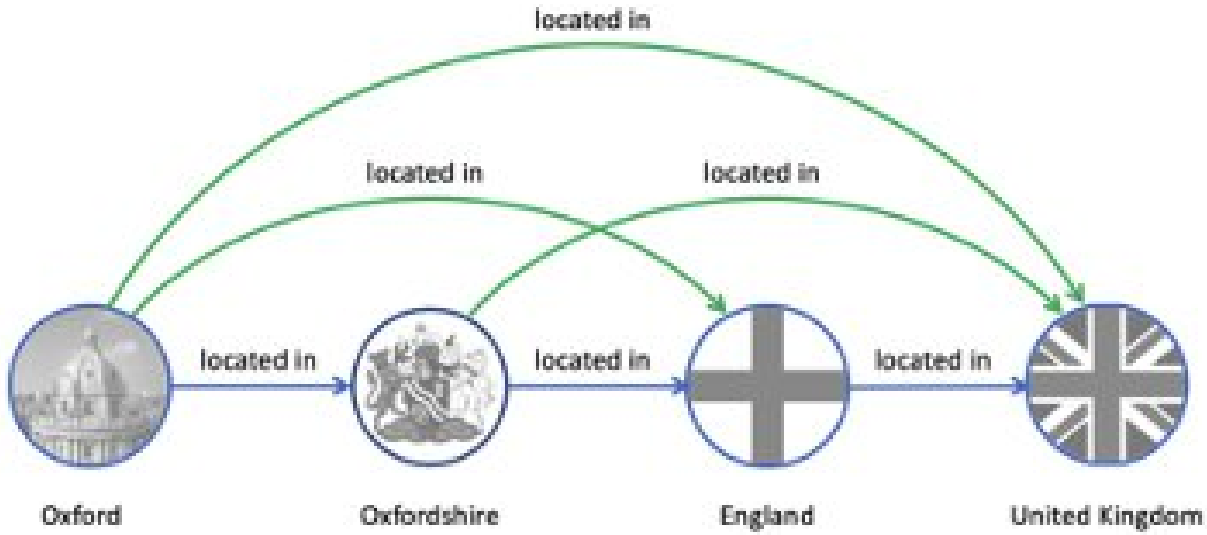
- 13 484 751 scientific articles
- 102 704 476 chemical compounds
- 8 969 553 compound - articles links
- 24 353 MeSH descriptors
- 6 689 ChEBI classes
- 2 712 ChemOnt classes

Extracting relevant associations

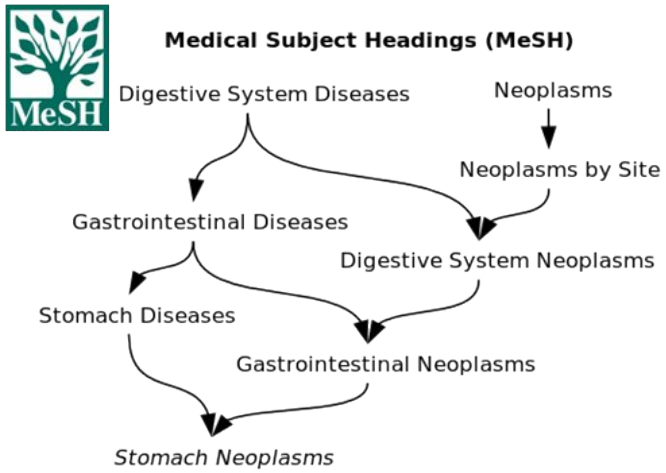


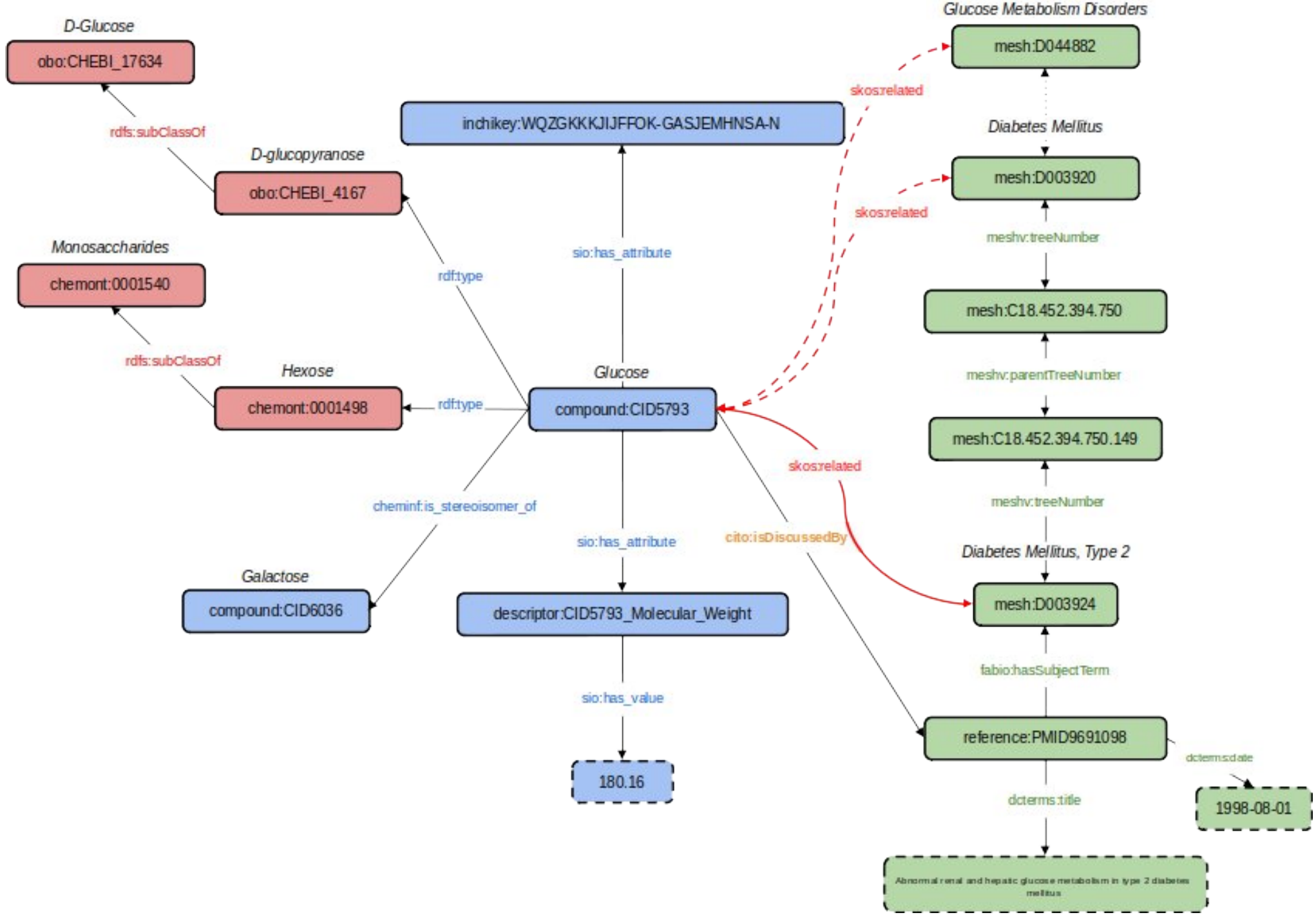

D-Glucose
obo:CHEBI_17634
rdfs:subClassOf
D-glucopyranose
obo:CHEBI_4167
inchikey:WQZGKKKJIJFFOK-GASJEMHNSA-N
Monosaccharides
chemont:0001540
rdf:type
sio:has_attribute
rdfs:subClassOf
Hexose
chemont:0001498
rdf:type
Glucose
compound:CID5793
cheminf:is_stereoisomer_of
sio:has_attribute
skos:related
cito:isDiscussedBy
Galactose
compound:CID6036
descriptor:CID5793_Molecular_Weight
sio:has_value
180.16
Glucose Metabolism Disorders
mesh:D044882
Diabetes Mellitus
mesh:D003920
meshv:treeNumber
mesh:C18.452.394.750
meshv:parentTreeNumber
mesh:C18.452.394.750.149
meshv:treeNumber
Diabetes Mellitus, Type 2
mesh:D003924
fabio:hasSubjectTerm
reference:PMID9691098
dcterms:date
1998-08-01
dcterms:title
Abnormal renal and hepatic glucose metabolism in type 2 diabetes mellitus

Automated reasoning using the semantic level



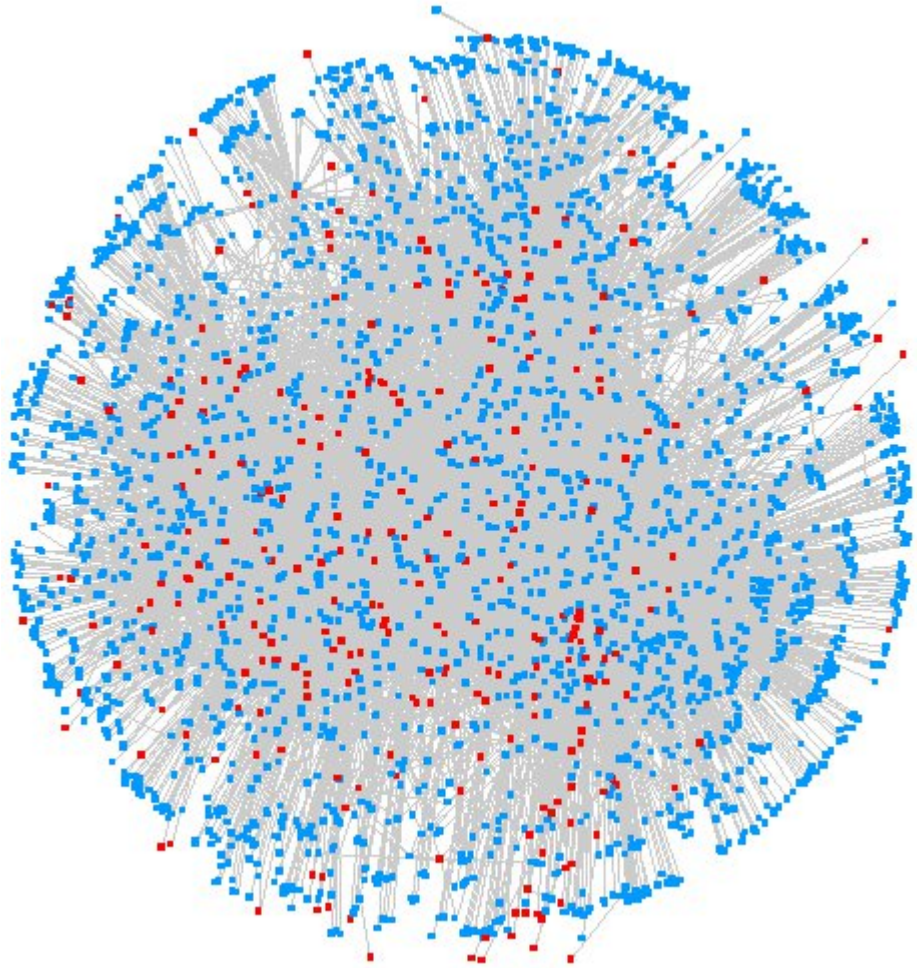
- CHEBI:78616 carbohydrates and carbohydrate derivatives
  - CHEBI:16646 carbohydrate
    - CHEBI:35381 monosaccharide
      - CHEBI:15693 aldose
        - CHEBI:33917 aldohexose
          - CHEBI:17608 D-aldohexose
            - **CHEBI:17634 D-glucose**




D-Glucose
obo:CHEBI_17634
rdfs:subClassOf
D-glucopyranose
obo:CHEBI_4167
Monosaccharides
chemont:0001540
rdfs:subClassOf
Hexose
chemont:0001498
rdf:type
inchikey:WQZGKKKJIJFFOK-GASJEMHNSA-N
sio:has_attribute
rdf:type
Glucose
compound:CID5793
cheminf:is_stereoisomer_of
Galactose
compound:CID6036
sio:has_attribute
descriptor:CID5793_Molecular_Weight
sio:has_value
180.16
skos:related
skos:related
skos:related
cito:isDiscussedBy
Glucose Metabolism Disorders
mesh:D044882
Diabetes Mellitus
mesh:D003920
meshv:treeNumber
mesh:C18.452.394.750
meshv:parentTreeNumber
mesh:C18.452.394.750.149
meshv:treeNumber
Diabetes Mellitus, Type 2
mesh:D003924
fabio:hasSubjectTerm
reference:PMID9691098
dcterms:date
1998-08-01
dcterms:title
Abnormal renal and hepatic glucose metabolism in type 2 diabetes mellitus

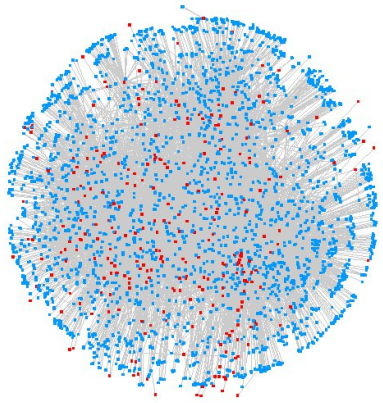
The Knowledge Network - Extended

- 343 435 compounds and 9 401 chemical classes against 24 353 biomedical concepts
- 82 248 460 of relations chemical - biomedical concepts
- **8 915 642 statistically significant ones**
- **1 132 217 related to disease concepts**



RDF Triplestore : Provision of the knowledge graph

https://forum.semantic-metabolomics.fr/sparql?default-graph-uri=&query=....



RDF

OPENLINK VIRTUOSO UNIVERSAL SERVER

Jena

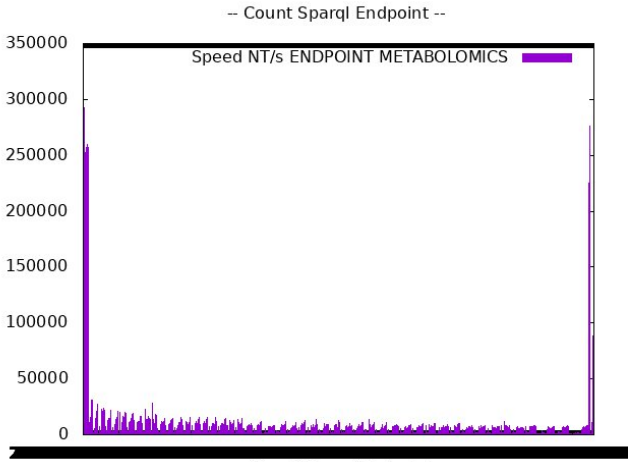
SPARQL

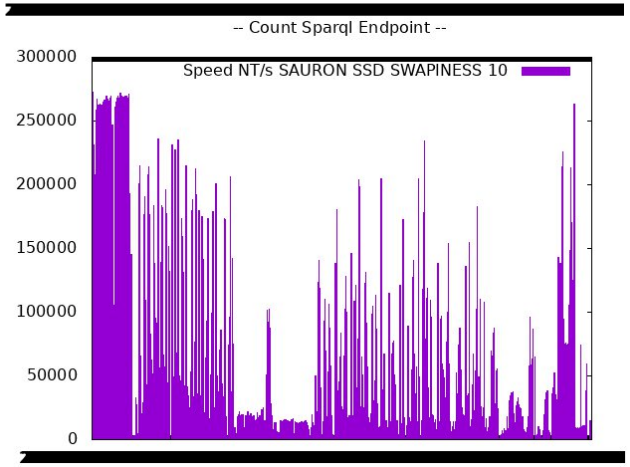
SPARQL | HTML5 table

property	value
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Class
http://www.w3.org/2000/01/rdf-schema#subClassOf	http://purl.obolibrary.org/obo/CHEBI_24995
http://www.w3.org/2000/01/rdf-schema#subClassOf	http://purl.obolibrary.org/obo/CHEBI_52898
http://www.w3.org/2000/01/rdf-schema#label	"N-[(4S,7R,8R)-5-[2-(dimethylamino)-1-oxoethyl]-8-methoxy-4,7,10-trimethyl-11-oxo-2-oxa-5,10-diazabicyclo[10.4.0]hexadeca-1(12
http://www.w3.org/2000/01/rdf-schema#label	"chebi:106584"
http://www.w3.org/2000/01/rdf-schema#comment	"N-[(4S,7R,8R)-5-[2-(dimethylamino)-1-oxoethyl]-8-methoxy-4,7,10-trimethyl-11-oxo-2-oxa-5,10-diazabicyclo[10.4.0]hexadeca-1(12
http://www.w3.org/2002/07/owl#sameAs	https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:106584
http://www.w3.org/2002/07/owl#sameAs	https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:106584
http://www.geneontology.org/formats/oboInOwl#hasDbXref	"LINCS:LSM-17942"^^<http://www.w3.org/2001/XMLSchema#string>
http://www.geneontology.org/formats/oboInOwl#hasOBONamespace	"chebi_ontology"^^<http://www.w3.org/2001/XMLSchema#string>
http://www.geneontology.org/formats/oboInOwl#id	"CHEBI:106584"^^<http://www.w3.org/2001/XMLSchema#string>
http://purl.obolibrary.org/obo/chebi/charge	"0"^^<http://www.w3.org/2001/XMLSchema#string>
http://purl.obolibrary.org/obo/chebi/formula	"C25H38N4O5"^^<http://www.w3.org/2001/XMLSchema#string>
http://purl.obolibrary.org/obo/chebi/inchi	"InChI=1S/C25H38N4O5/c1-16-12-29(23(30)14-27(3)4)17(2)15-34-21-10-9-19(26-24(31)18-7-8-18)11-20(21)25(32)28(5)13-22(16)33-6/h9
http://purl.obolibrary.org/obo/chebi/inchikey	"JUIHLWHRNYURSY-JLHGSKIFSA-N"^^<http://www.w3.org/2001/XMLSchema#string>
http://purl.obolibrary.org/obo/chebi/mass	"474.594"^^<http://www.w3.org/2001/XMLSchema#string>
http://purl.obolibrary.org/obo/chebi/monoisotopicmass	"474.28422"^^<http://www.w3.org/2001/XMLSchema#string>
http://purl.obolibrary.org/obo/chebi/smiles	"C[C@@H]1CN([C@H](COC2=C(C=C(C=C2)NC(=O)C3CC3)C(=O)N(C[C@@H]1OC)C)C(=O)CN(C)C"^^<http://www.w3.org/2001/XMLSchema#string>

https:/ / fr.wikipedia.org/ wiki/ Triplestore

FORUM : A very Big Knowledge Network





	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

OPENLINK® VIRTUOSO UNIVERSAL SERVER

FORUM : A very Big Knowledge Network

+

+

⇒

« Metabolomics Semantic Datalake »

DipSO

SAPI 2021

Franck Giacomoni

Christophe Dupérier

Plate-Forme EXPLORATION du METABOLISME

CLERMONT FOOT 63

BORDEAUX METABOLOME

GIRONDINS DE BORDEAUX

David Benaben

CATI

eMPrEInTE

PROSODie

BARIC

Metabolomics Semantic Datalake

igepp

Matéo Boudet

STADE RENNAIS FOOTBALL CLUB

Olivier Filangi

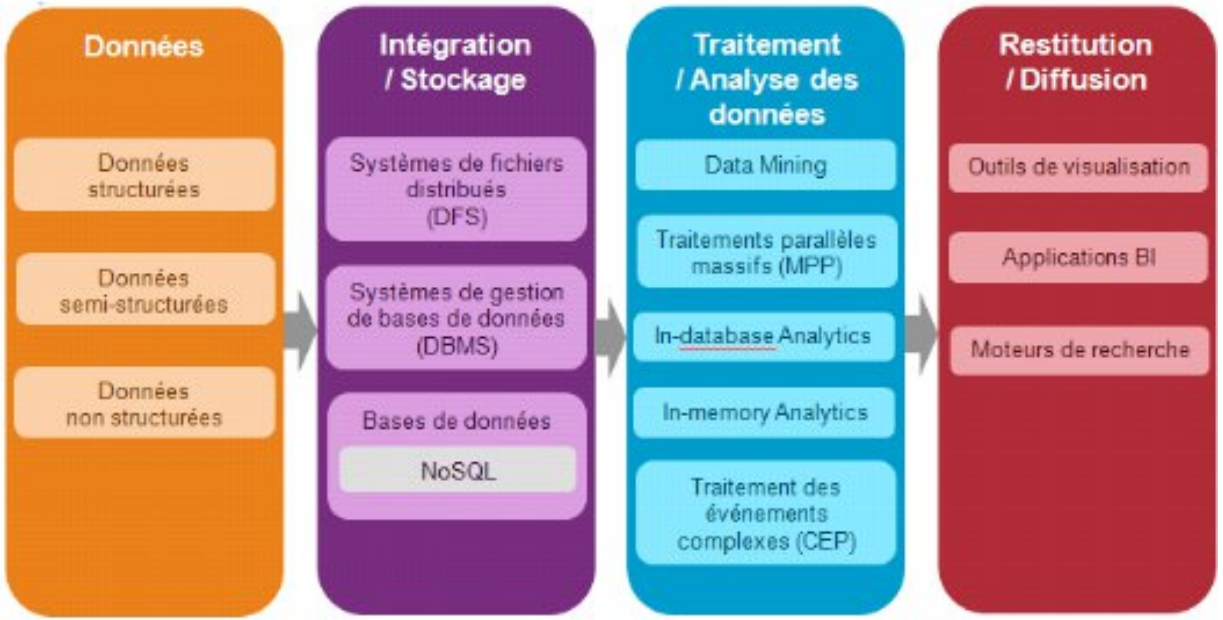
AAP DiPSO 2021 « Soutien à Projets Innovants et Structurants sur le numérique »

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

Objectifs de la plateforme MSD

- Fédération de données à l'échelle d'un consortium
- Renforcer l'interopérabilité de nos données avec les autres omiques / ressources
- Garantir la traçabilité des données et des traitements
- Automatiser l'enrichissement en métadonnées des données
- Doter MetaboHUB d'un dépôt central unifié
- Centraliser autour d'une PF les compétences et les métiers en « science des données »
- Evaluer les technologies du « Big Data » pour les omiques

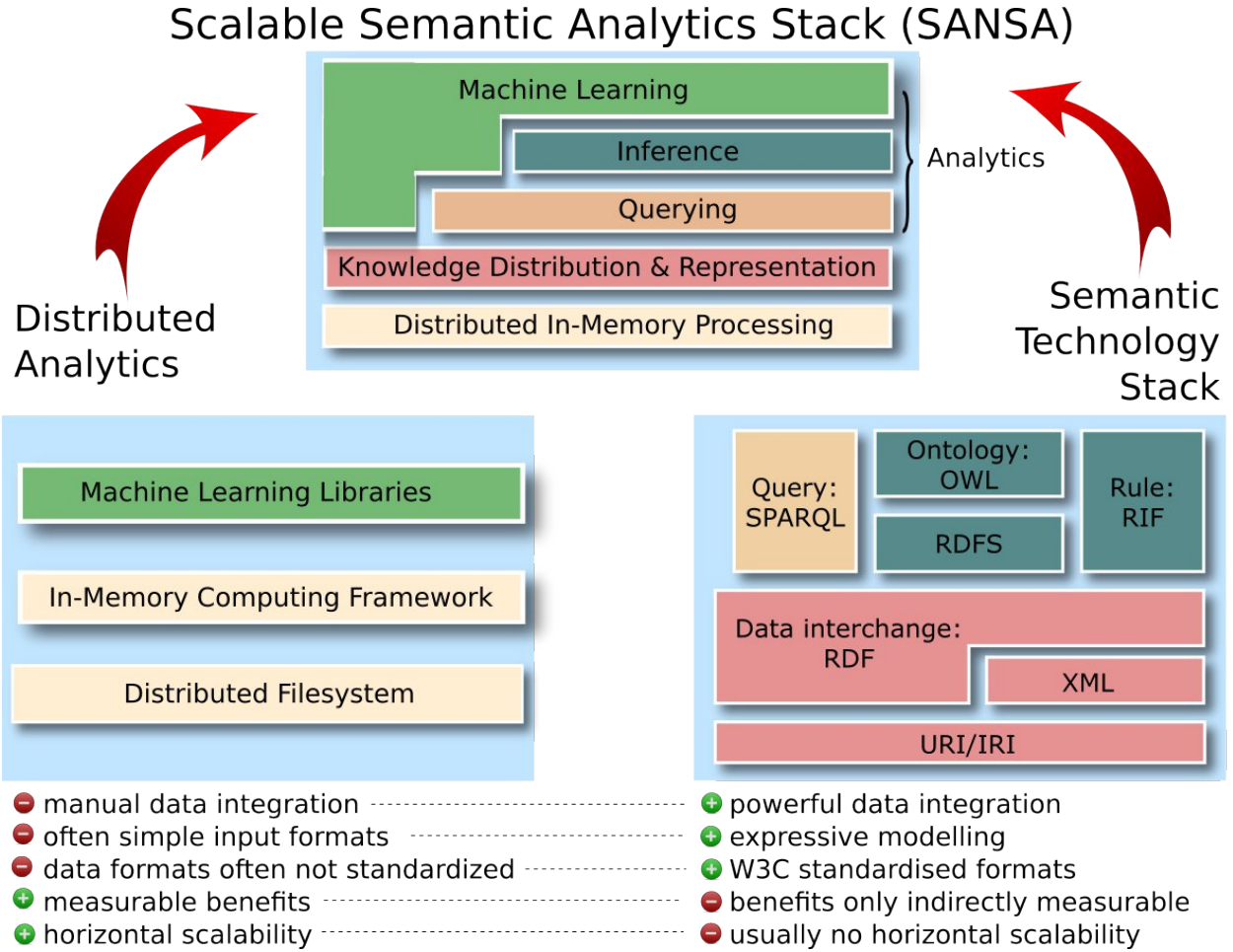
The usual big data…



SANSA : big data engine for scalable processing of large-scale RDF data

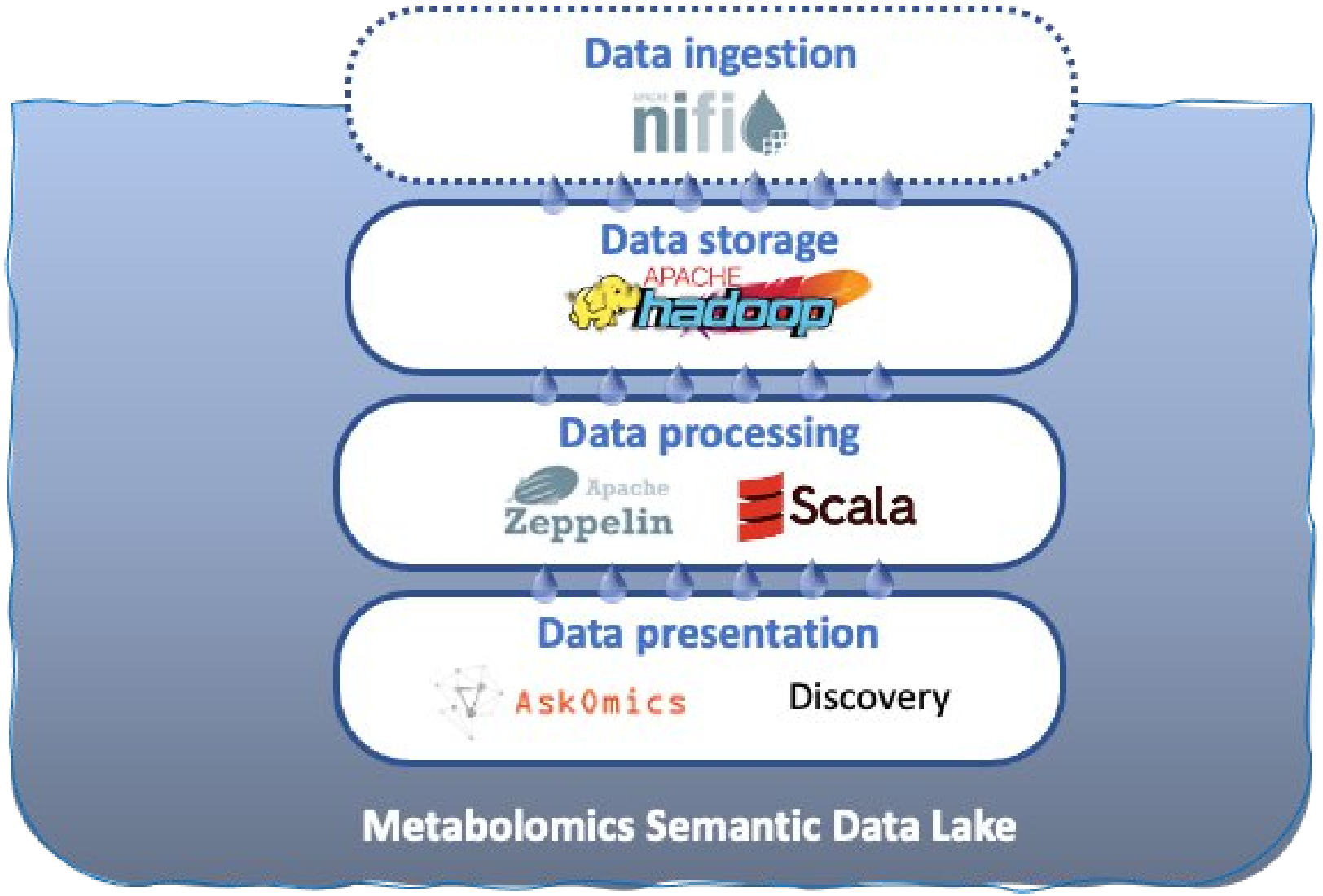
SANSA API

- Read/ Write
- Inference
- Querying
- Machine Learning



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

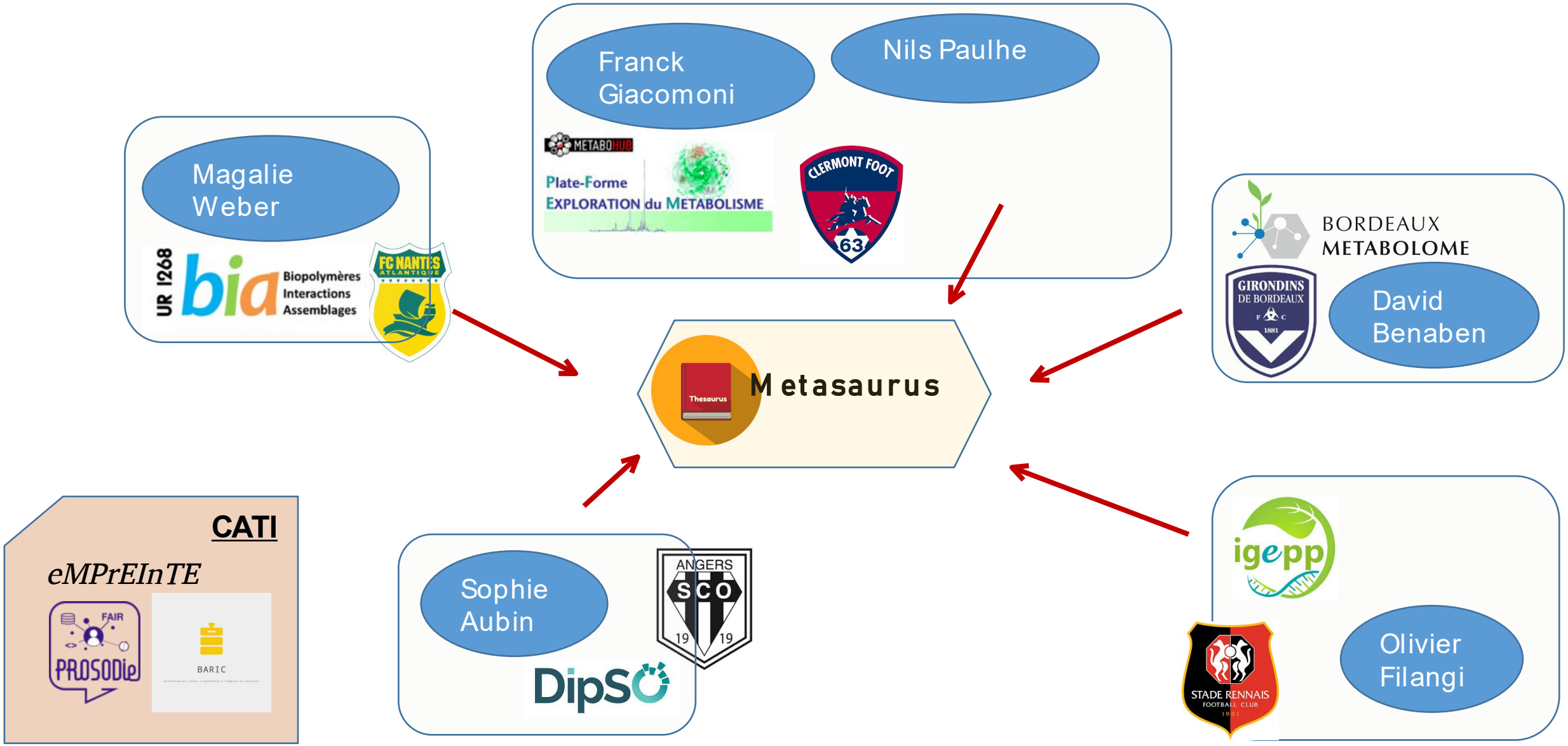
Architecture



Data flows in a metabolomics semantic datalake infrastructure

Bâtir un modèle de connaissance adapté aux études métabolomiques
en s'appuyant sur les ontologies, thésaurus et vocabulaires contrôlés publiés

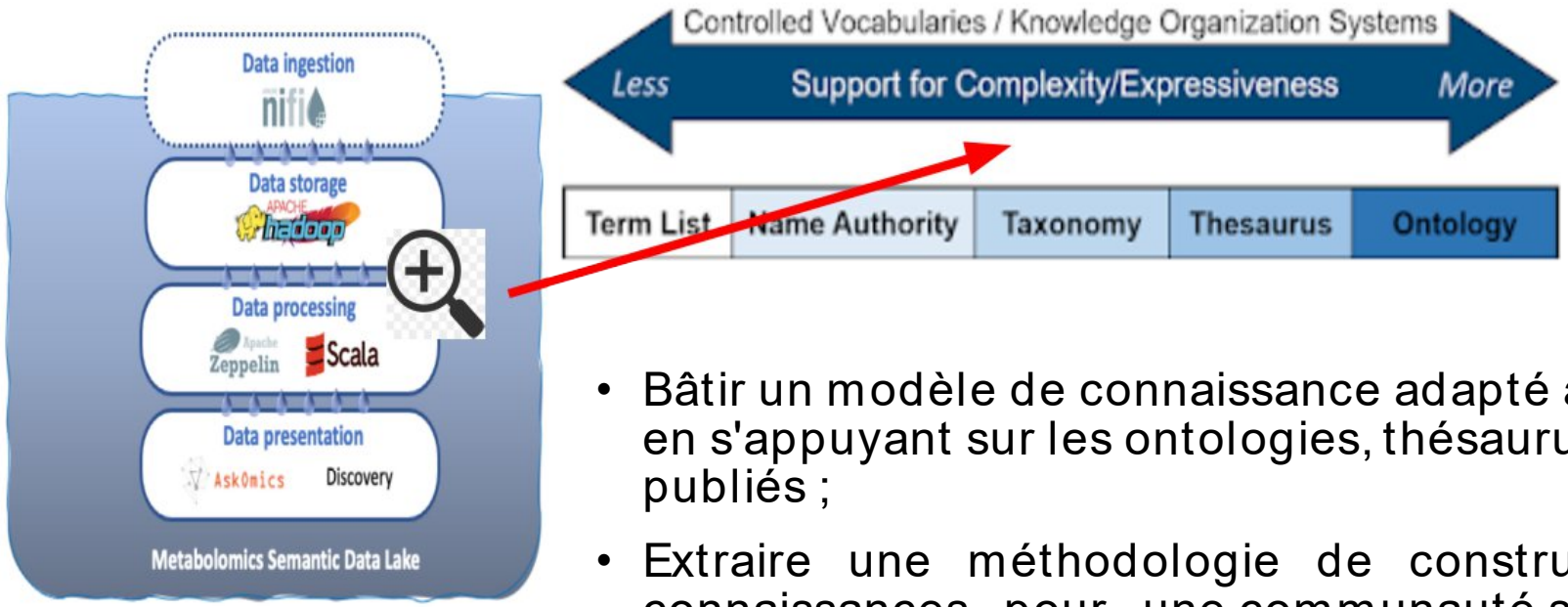
Metasaurus



Modèle pivot avec MetaSaurus

DipSO

SAPI 2022



- Bâtir un modèle de connaissance adapté aux études métabolomiques en s'appuyant sur les ontologies, thésaurus et vocabulaires contrôlés publiés ;
- Extraire une méthodologie de construction d'un modèle des connaissances pour une communauté scientifique à partir de ce cas d'usage ;
- Enrichir le thésaurus INRAE des concepts de ce modèle de connaissance qui ne sont pas encore formalisés.

Thanks ! Questions ?

Florence Vinson

Clément Frainay

Maxime Delmas

Toxalim RESEARCH CENTRE IN FOOD TOXICOLOGY

Thèse M. Delmas : « Construire, exploiter et étendre un graphe de connaissances pour l'étude des liens entre métabolisme et santé »

Franck Giacomoni

Nils Paulhe

Christophe Dupérier

METABOHUB Plate-Forme EXPLORATION du METABOLISME

CLERMONT FOOT 63

FORVM

BORDEAUX METABOLOME

GIRONDINS DE BORDEAUX 1881

David Benaben

Magalie Weber

UR 1268 bia Biopolymères Interactions Assemblages

FC NANTES ATLANTIQUE

Metasaurus

Metabolomics Semantic Datalake

CATI

eMPrEInTE

FAIR PROSODie

BARIC

Sophie Aubin

ANGERS SCO 19 19

DipSO

igepp

Matéo Boudet

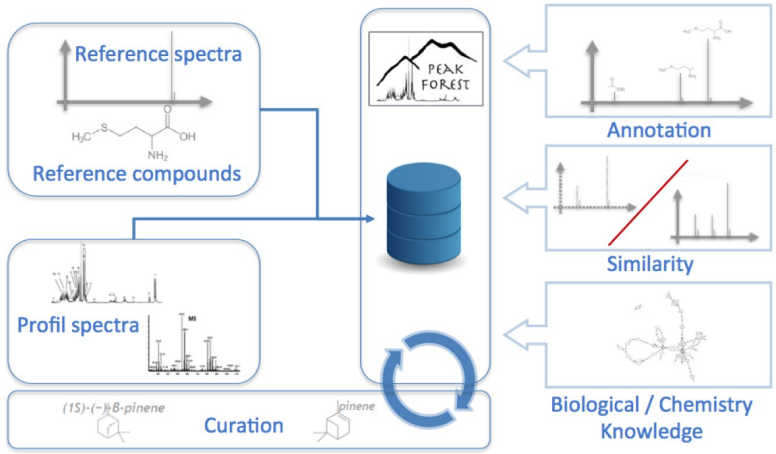
Olivier Filangi

STADE RENNAIS FOOTBALL CLUB 1901

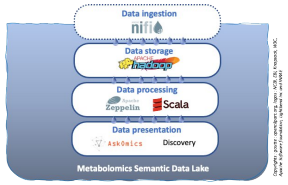
Des besoins -> des techno adaptées

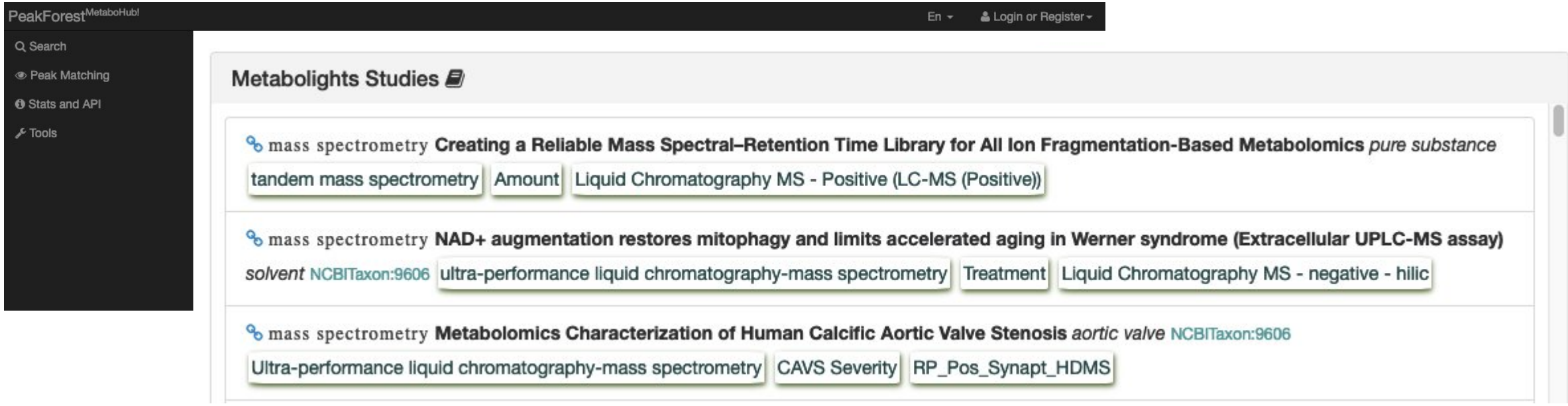
- **Besoins** :
  - Données en métabolomique = problématiques « Big Data »
  - Hétérogénéité des formats (métabolites, spectres, ...)
  - Besoins de contrôler les flux de métadonnées (diversité des sources, ...)
  - Générer des bases de connaissances à partir de sources multiples
- **Les solutions technologiques :**
  - ***Web sémantique / RDF / OWL*** : formalisation et structuration de la donnée
  - ***Spark / Sansa / Scala*** : exploitation de l'infrastructure distribuée
  - ***Notebook / AskOmics / Discovery*** : exploitation des données par les utilisateurs finaux

Use case 01 : PeakForest & MetaboLights (1/ 2)



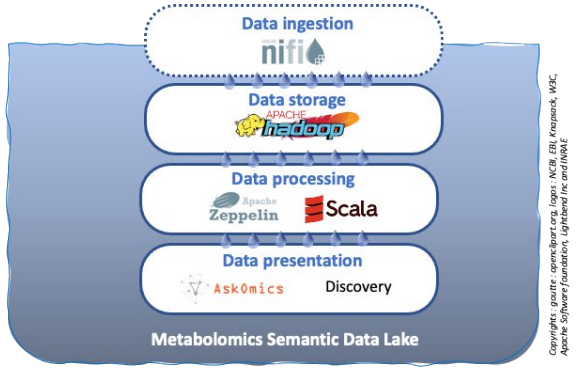
MetaboLights





Use case 01 : PeakForest & MetaboLights (2/ 2)

MetaboLights



Plateforme MSD

MetaboLights

RDF

DISCOVERY

P2M2

demo PEAK FOREST

Metabolights Studies

mass spectrometry **Creating a Reliable Mass Spectral–Retention Time Library for All Ion Fragmentation-Based Metabolomics** *pure substance*
tandem mass spectrometry | Amount | Liquid Chromatography MS - Positive (LC-MS (Positive))

mass spectrometry **NAD+ augmentation restores mitophagy and limits accelerated aging in Werner syndrome (Extracellular UPLC-MS assay)**
solvent NCBITaxon:9606 | ultra-performance liquid chromatography-mass spectrometry | Treatment | Liquid Chromatography MS - negative - hilic

mass spectrometry **Metabolomics Characterization of Human Calcific Aortic Valve Stenosis** *aortic valve* NCBITaxon:9606
Ultra-performance liquid chromatography-mass spectrometry | CAVS Severity | RP_Pos_Synapt_HDMS

Use case 2 : Travaux de Ghina Hajjar – MetaboHUB-Clermont

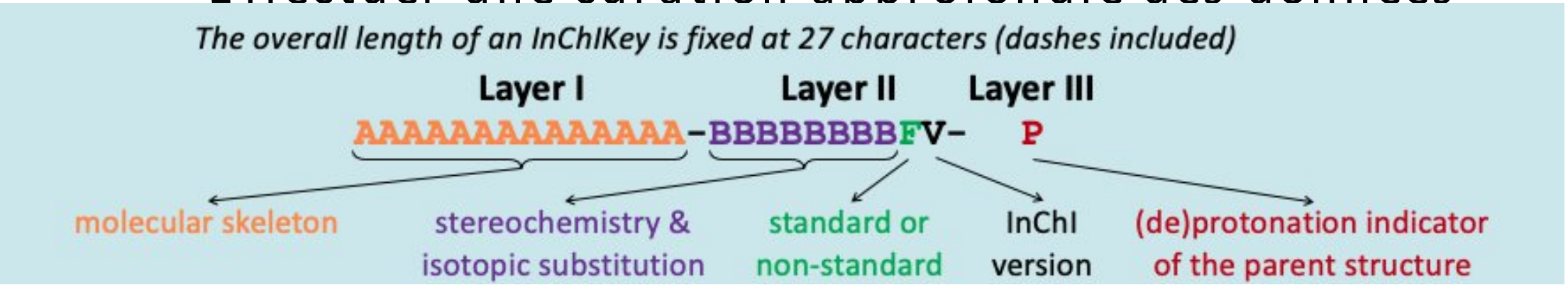
Metabolite reporting in large-scale studies within different metabolomics communities:

DO WE SPEAK THE SAME LANGUAGE?

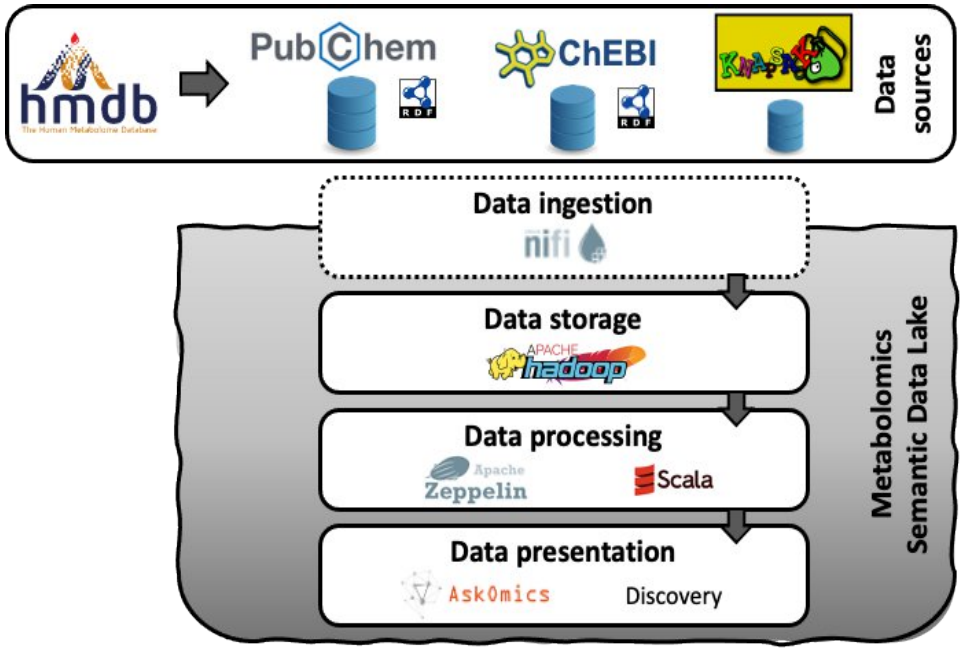
Ghina Hajjar [a], David Benaben [b,c], Nils Paulhe [a,c], Christophe Duperier [a,c], Olivier Filangi [c,d], Franck Giacomoni [a,c], Blandine Comte [a], Estelle Pujos-Guillot [a]

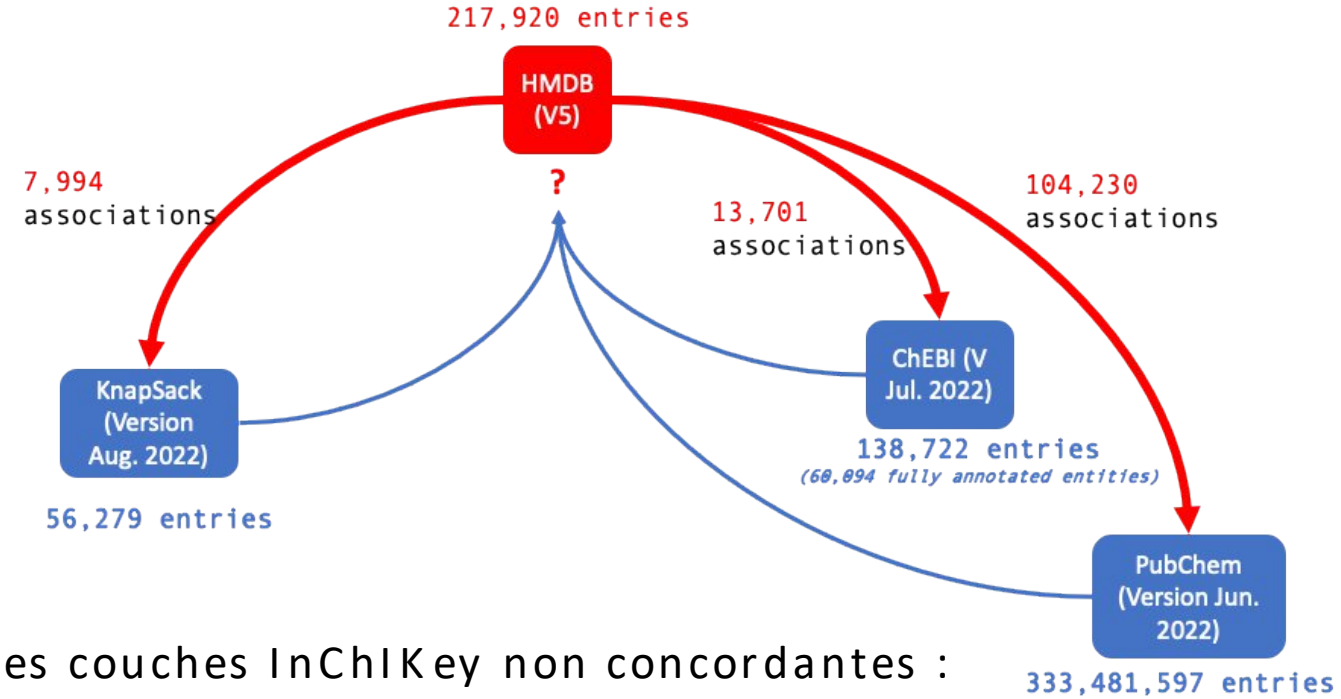
MetSoc Valencia, 2022
Analytics Nantes, 2022

- **Objectif**
  - Guidelines pour un partage plus efficace et reproductible des données métabolomiques
    - Utiliser des identifiants communs tels que InChIKey,
    - Effectuer une curation approfondie des données



Use case 2 : Travaux de Ghina Hajjar





Mismatchings ont été classées en fonction des couches InChIKey non concordantes :

- (De)Protonation of the core parent structure

e.g. N6,N6,N6-Trimethyl-L-lysine : HMDB0001325 (a) linked to ChEBI :17311 (b)

(a) MXNRLFUSFKVQSK-QMMMGPOBSA-**N (parent structure)**

(b) MXNRLFUSFKVQSK-QMMMGPOBSA-**O (protonated structure)**

- Isomerism, stereochemistry

e.g. Methylcysteine : HMDB0002108 (c) linked to PubChem CID 24417 (d)

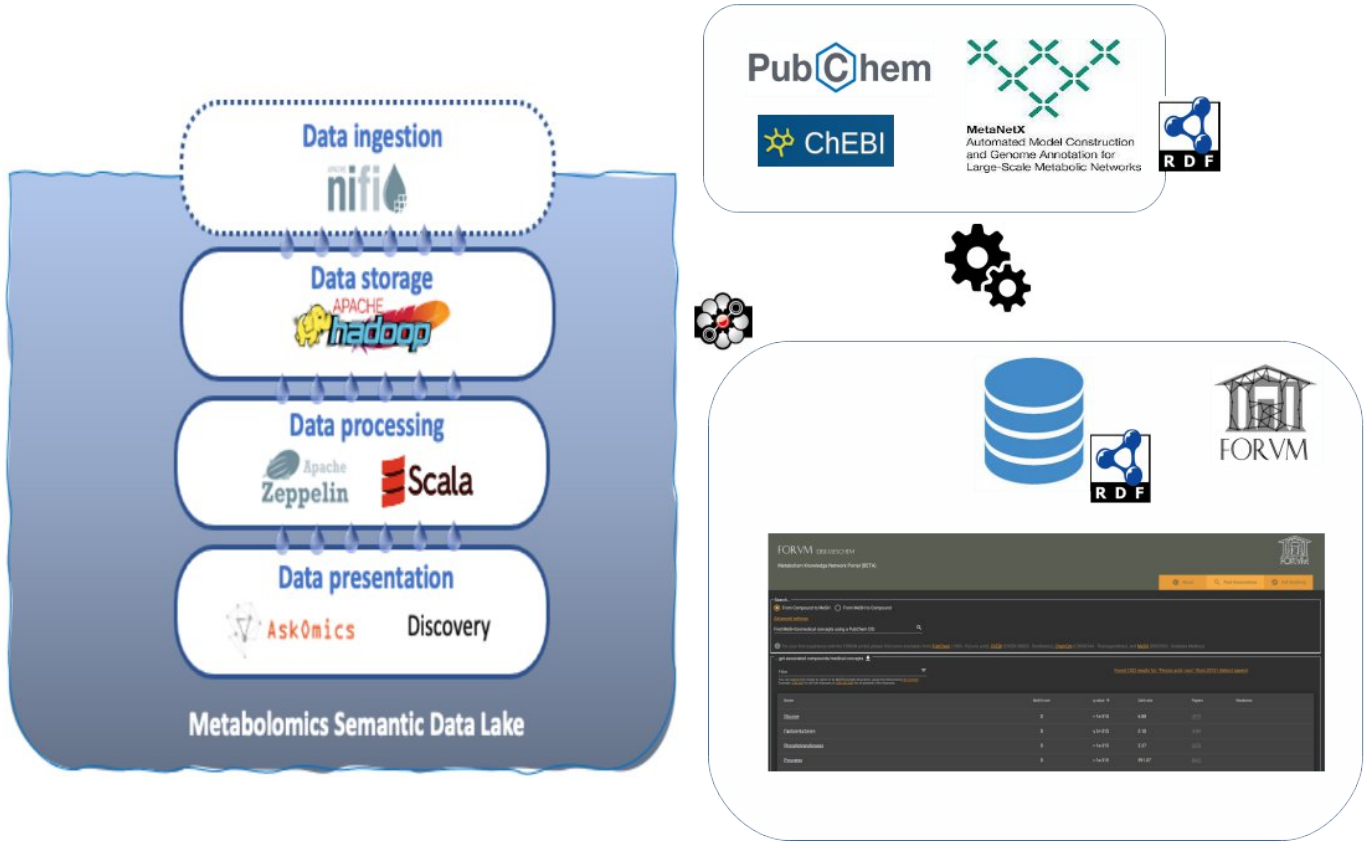
(c) IDIDJDIHTAOVLG-**GSVOUGTGSA**-N → **D-isomer**

(d) IDIDJDIHTAOVLG-**VKHMYHEASA**-N → **L-isomer**

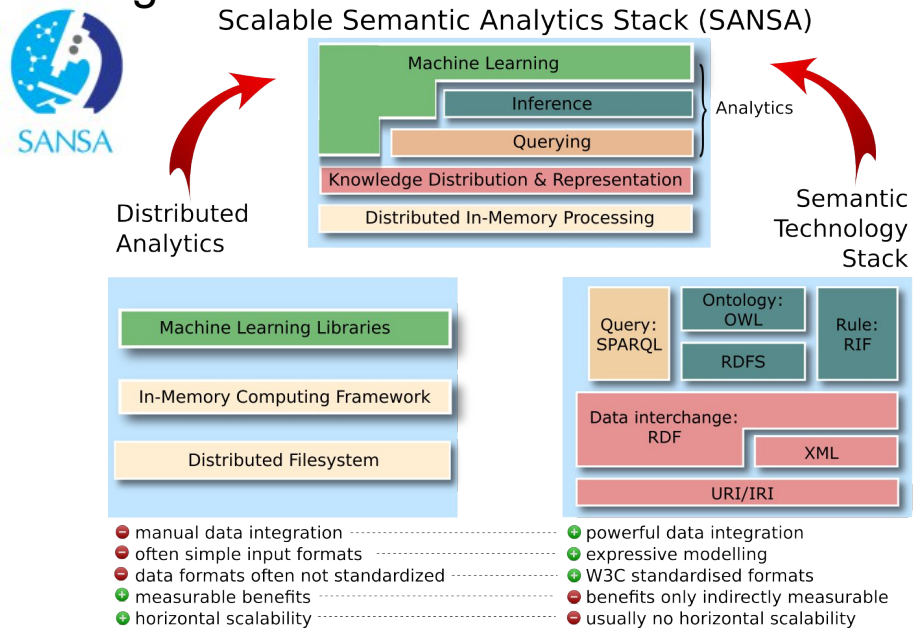
- Mismatch between structurally different compounds

e.g. 4,4-Dimethyl-5a-cholesta-8-en-3b-ol (HMDB0006840) linked to 5,6,7,8-tetrahydrofolylglutamic acid (CHEBI :27650)

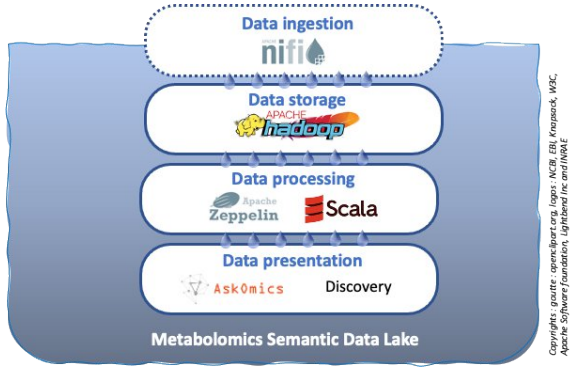
Use case 3 : FORUM (1/ 2)



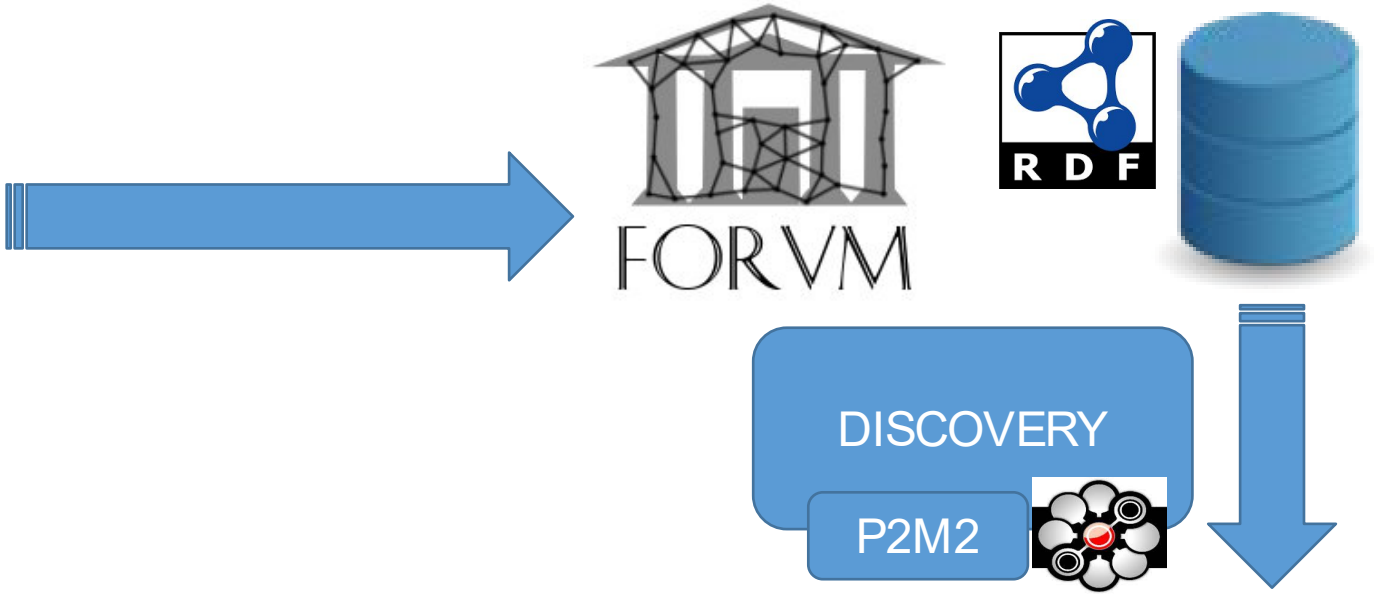
- 8 billion statements federating data from several life-science resources such as PubMed, ChEBI and PubChem
- Extract associations between compound and biomedical concepts, using literature metadata enrichment

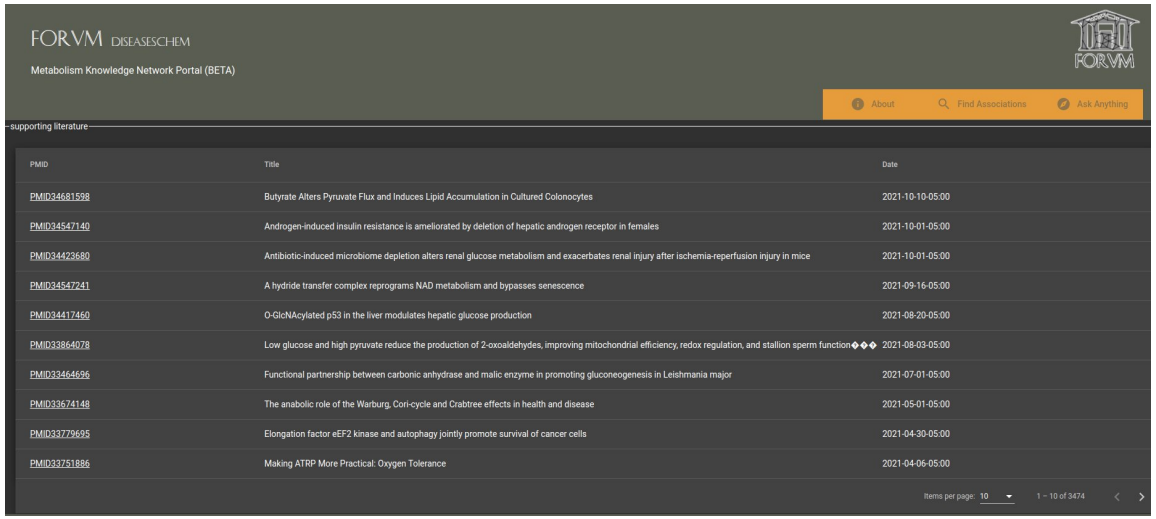


Use case 3 : FORUM (2/ 2)



Plateforme MSD





https:/ / forum-webapp.semantic-metabolomics.fr

Take home message

- Synergie autour de MSD des compétences et des métiers en « science des données »
- Infrastructure qui centralise les données et les traitements à l'échelle du consortium national MetaboHUB
- Innovation IT « Discovery » pour faciliter la mise à disposition des données structurées dans des Web Composants

- Atelier Big Data - 10-12 Janvier 2023 à Sète / / Datalake / / Base No SQL
  - https:/ / atelier-bigdata.sciencesconf.org/