Workshop inter-CATI autour des données omics
Paris 17 Oct. 2019

Session 01 - Bases et entrepôts de données, ontologies

Les ressources en Métabolomique

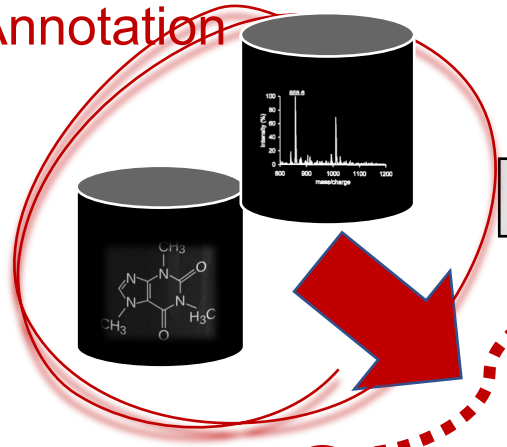
Franck Giacomoni

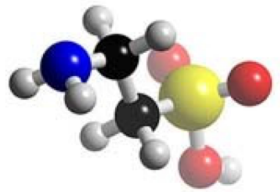
CATI EMPREINTE

Analyses des données en Métabolomique

name	namecustom	mz	mzmin	mzmax
M100T183	M100.07529T183	100.075290217332	100.073962458109	100.075856347639
M101T64	M101.017412T64	101.017411813327	101.016363119337	101.025598611889
M101T1090	M101.094497T1090	101.094497373251	101.093825596321	101.094931330369
M102T67	M102.121467T67	102.121466917005	102.118620434328	102.128045805211
M103T60	M103.032369T60	103.032369394794	103.031358949596	103.040632681061
M103T61	M103.120075T61	103.120074855461	103.118630961785	103.129325340649
M104T1162	M103.950053T1162	103.950052934034	103.948946676433	103.95133690961
M104T60	M104.037434T60	104.037433793914	104.036015958492	104.046229685927
M104T47	M104.100665T47	104.100665065741	104.099665732044	104.108679356773
M104T88	M104.095422T88	104.095422407145	104.093036006622	104.100116358936
M104T42	M104.195599T42	104.195598628802	104.187807623186	104.197644370905
M105T61	M105.045229T61	105.045228536653	105.036480671616	105.046334818564
M105T242	M105.069935T242	105.069934637469	105.06894296829	105.071143736243
M105T48	M105.103611T48	105.103611067744	105.102427529644	105.111698113321

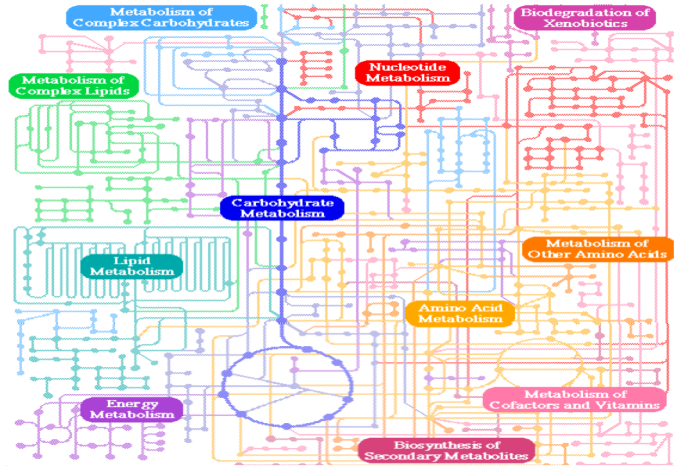
Annotation





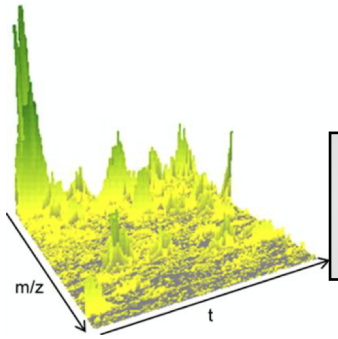
Identification

Interprétation

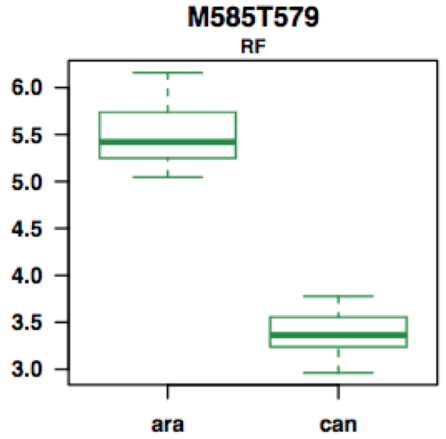


Extraction Data

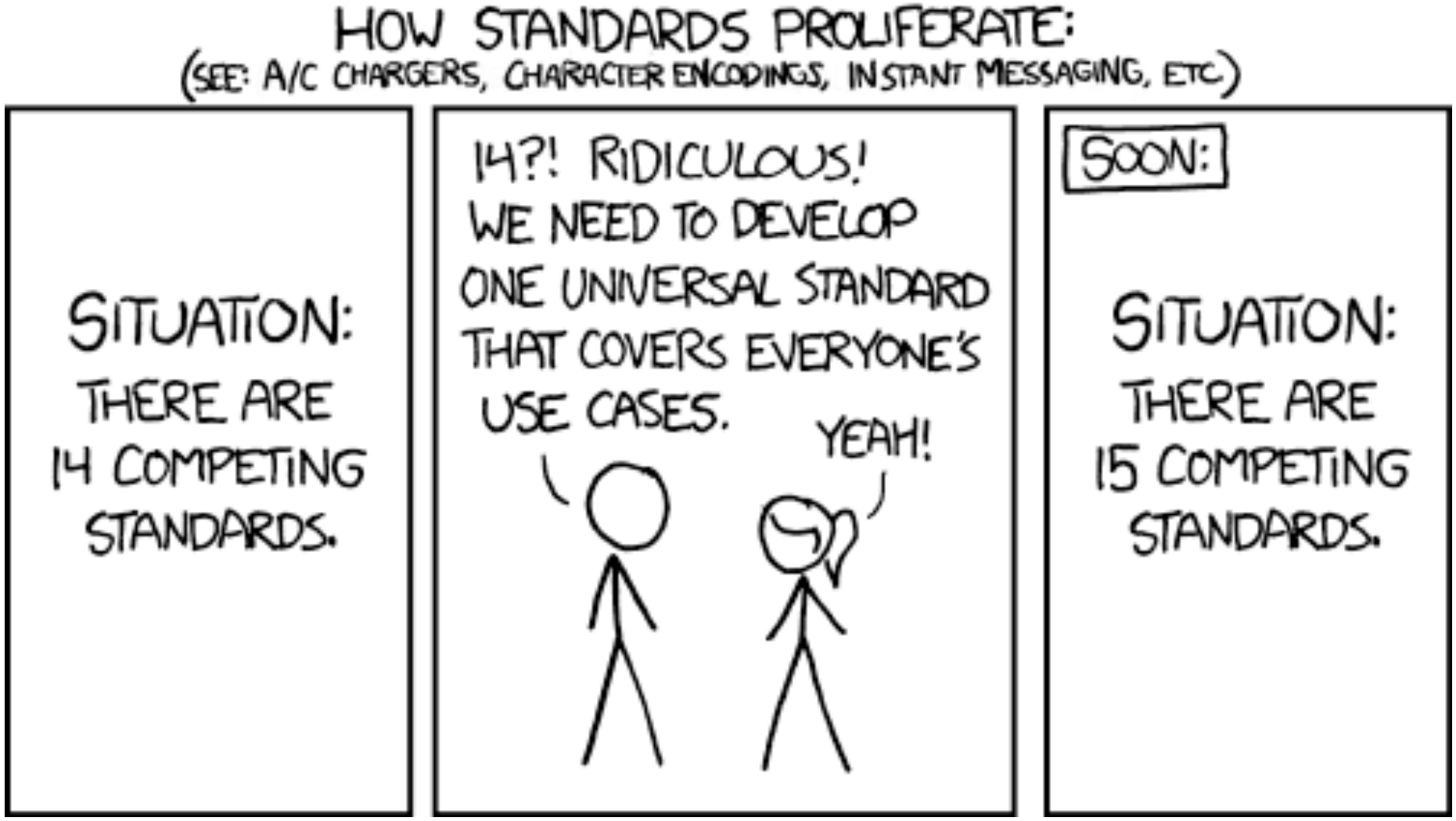
Analyses statistiques



Méthode analytique



RAW Data Formats – Cf wikipedia



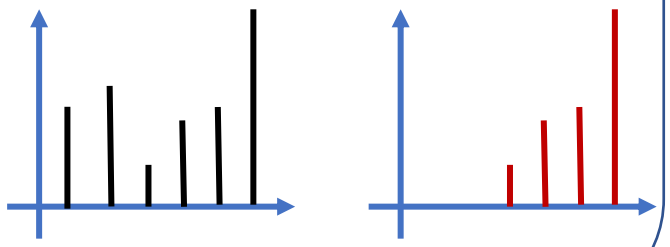
L’identification, étape complexe…

Diversité instrumentale

1 molécule

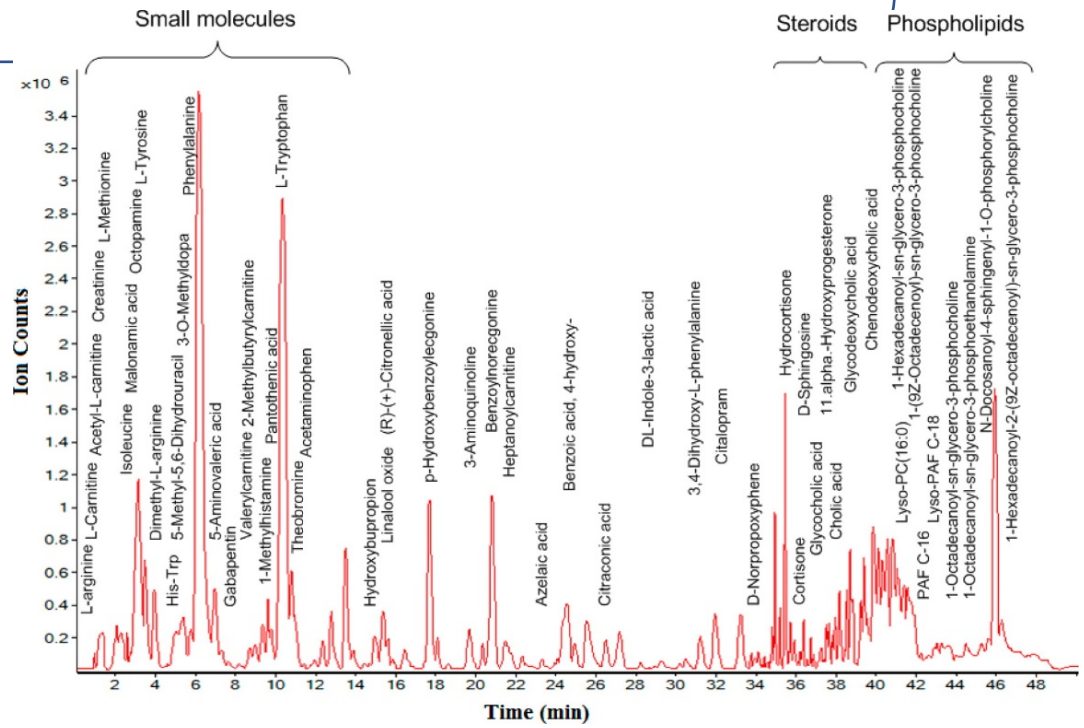
Technique 01

Technique 02



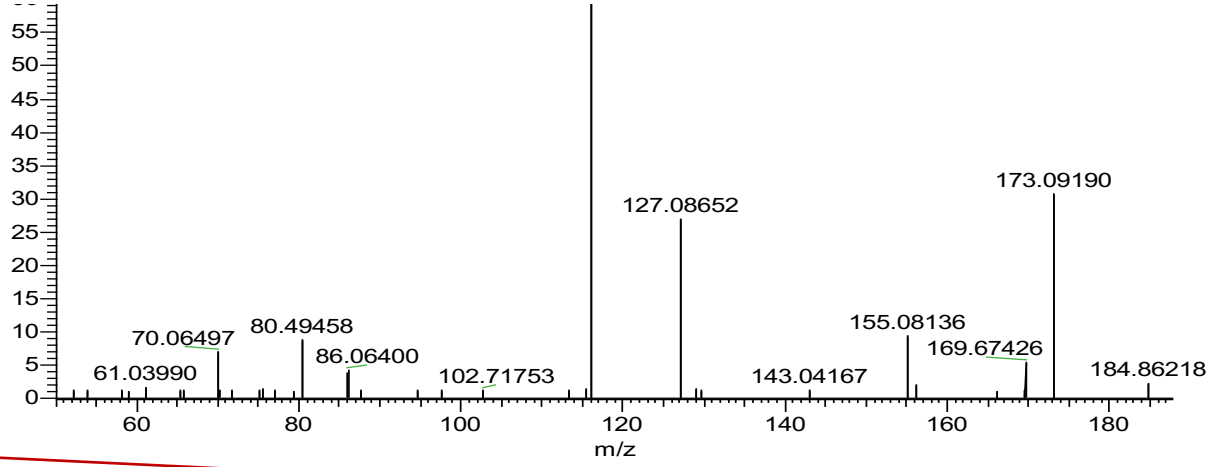
Métabolites en différentes concentrations

Diversités des familles chimiques de Métabolites



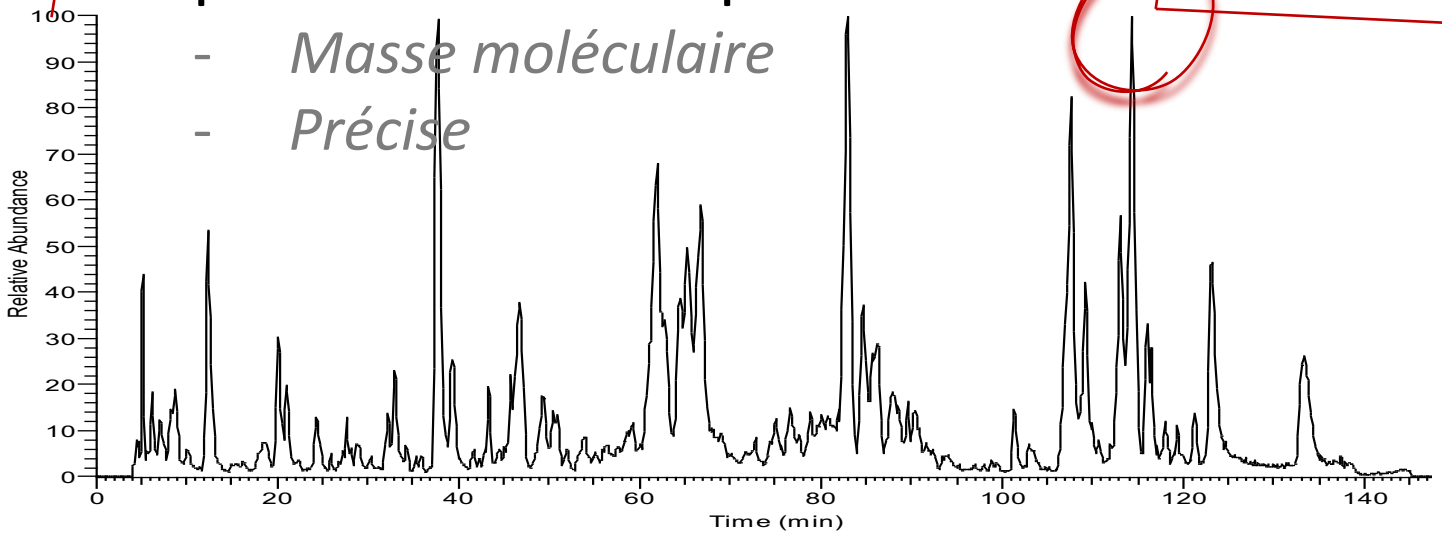
Des empreintes métaboliques à annoter

Exemple de la LCMS



Empreinte métabolique

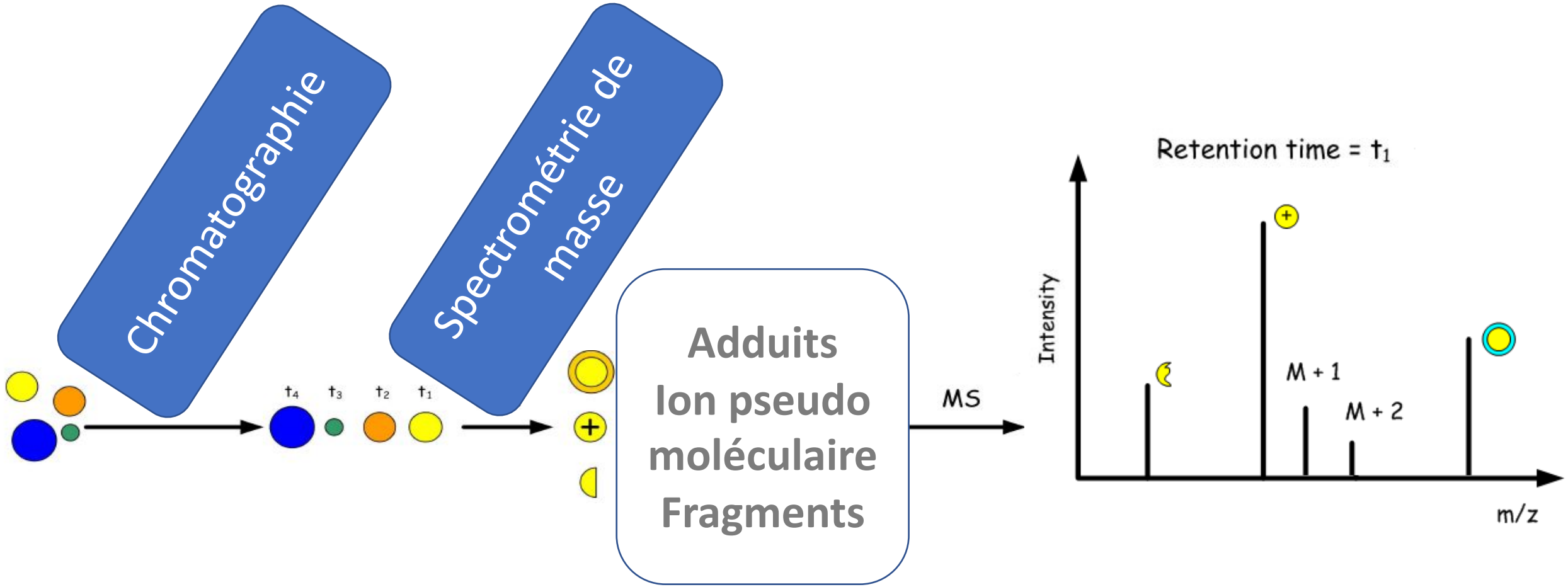
- *Masse moléculaire*
- *Précise*



Spectre

- *M/Z*
- *Intensité*
- *À une temps de rétention donné*

Une molécule, c'est plusieurs ions…



Banques en Métabolomique

Banques de composés chimiques

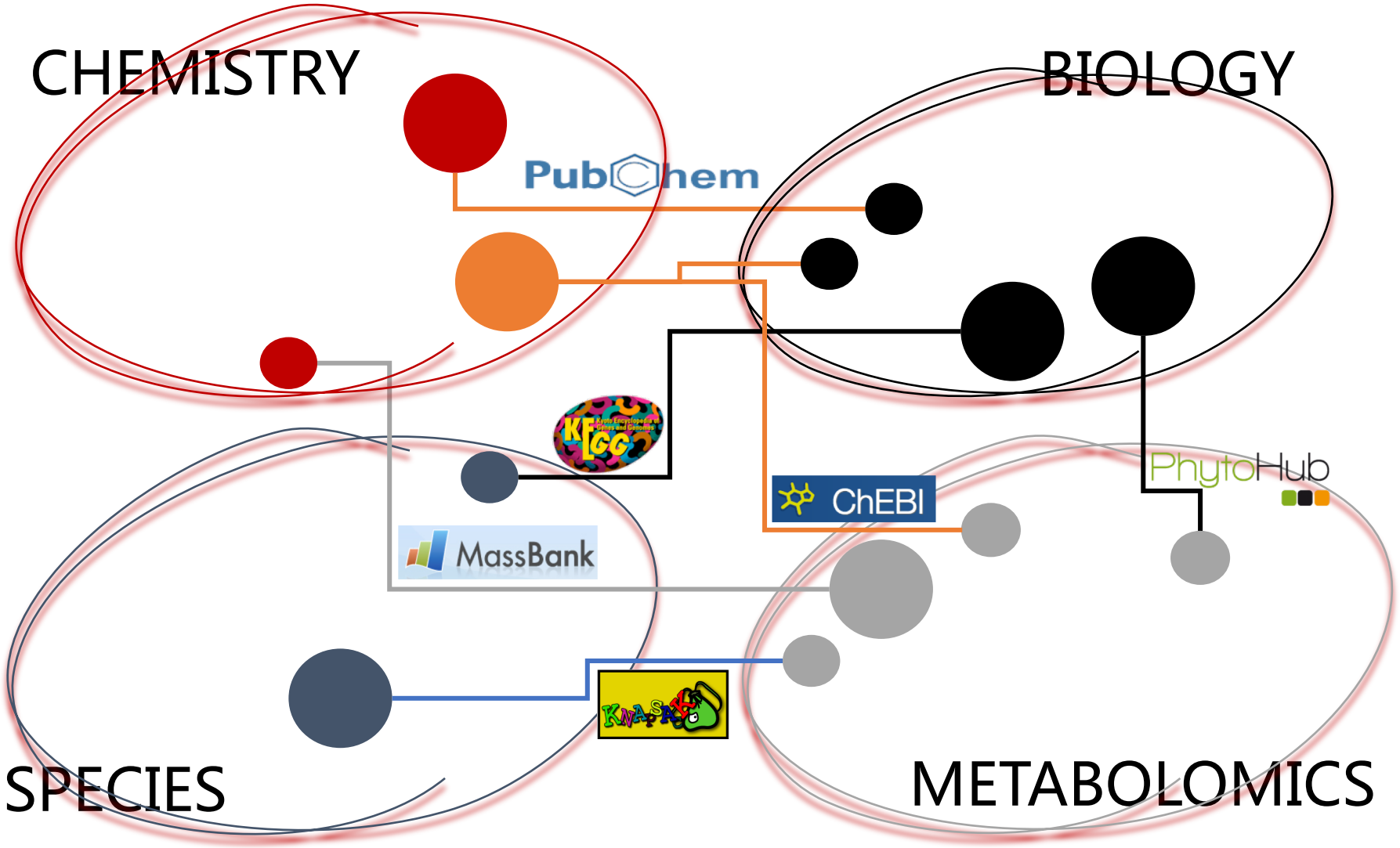
Librairies de spectres

Bases de données de réseaux métaboliques

Dépôts de références

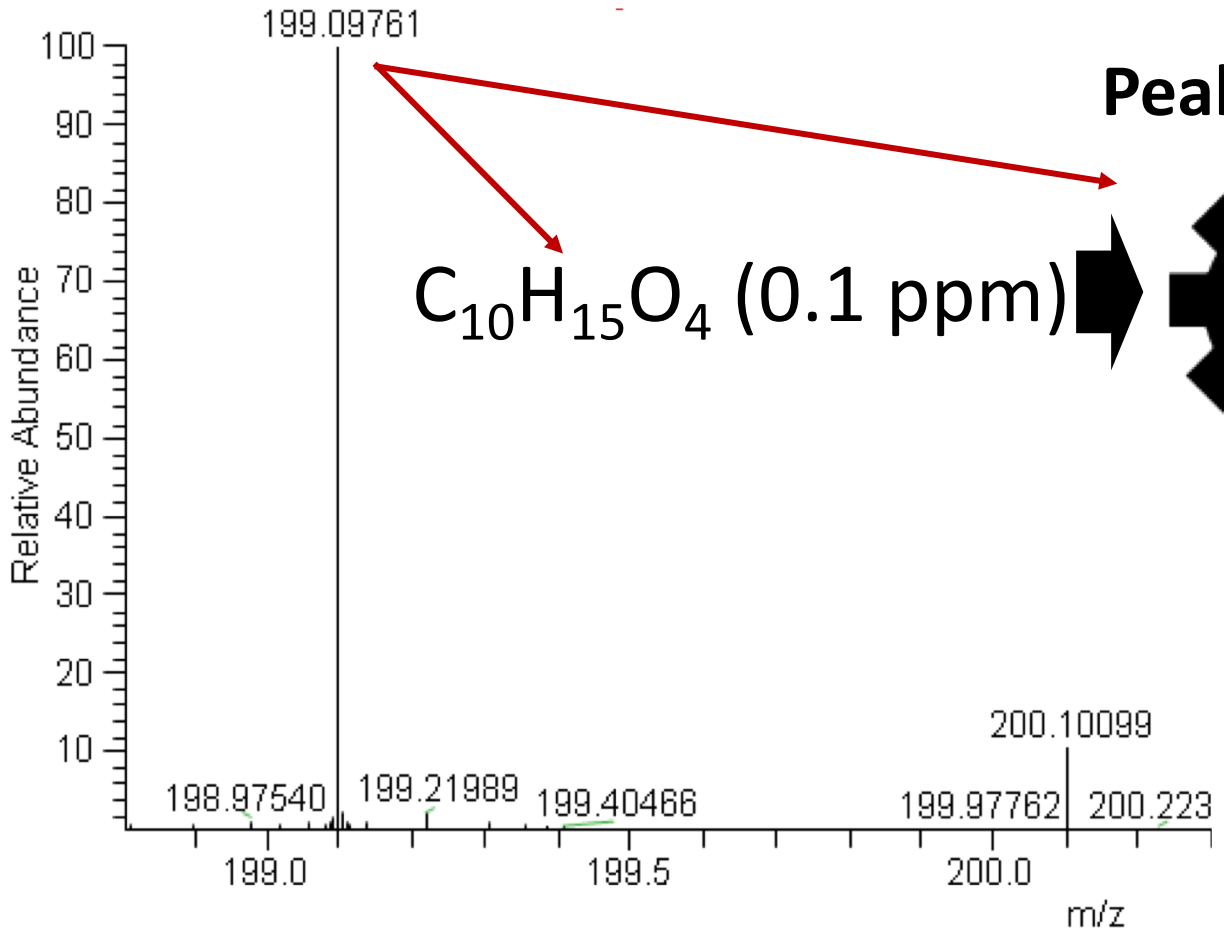
PSM V43 D169 Alfalfa fields in the desert, 1893

Banques en Métabolomique



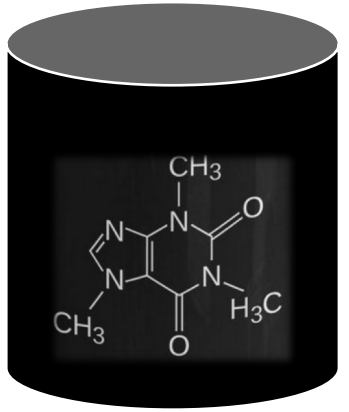
Stratégies d'annotations

Précision en masse



$C_{10}H_{15}O_4$ (0.1 ppm)

Peak matching



Banques de composés chimiques

Banques de composés chimiques

PubChem

A public repository of information on small molecules and their biological activities

157,000,000 entries

VS

ChEBI

Chemical databank

Name

Chemical Entities of Biological Interest

Freely available dictionary of molecular entities focused on 'small' chemical compounds.

Address

EMBL – EBI

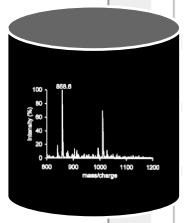
www.ebi.ac.uk/chebi

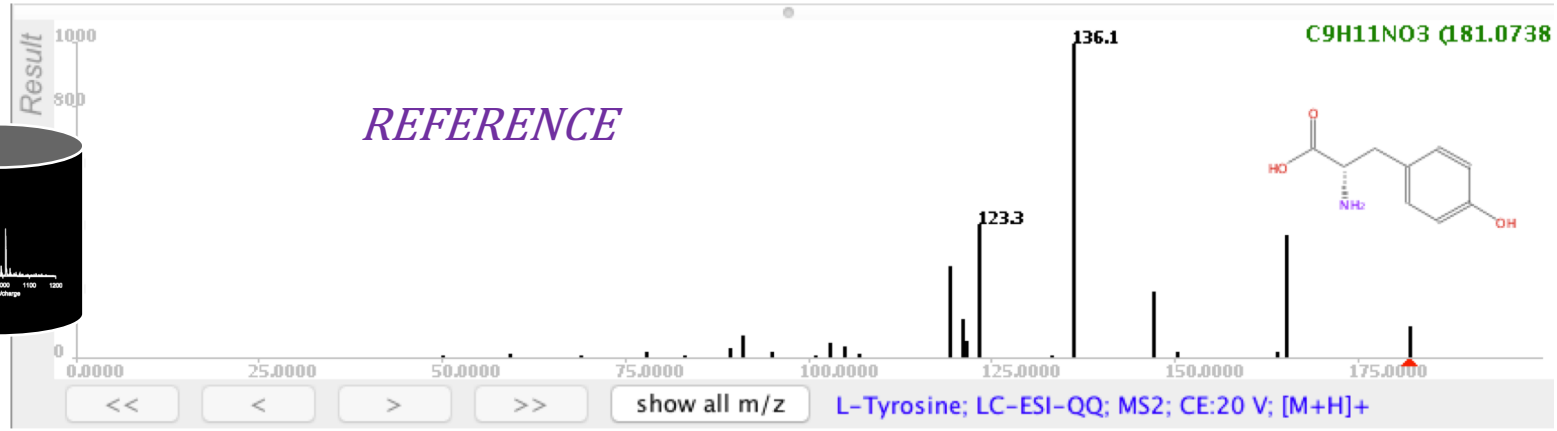
Entries

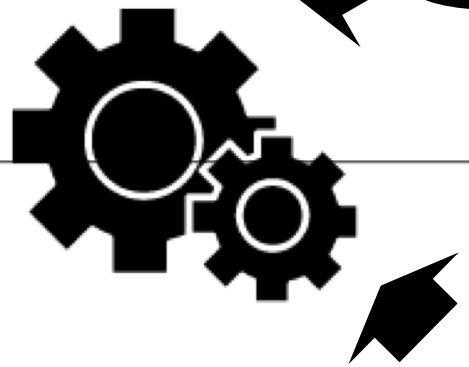
41,000 entities fully annotated

Des banques… des stratégies d'annotations

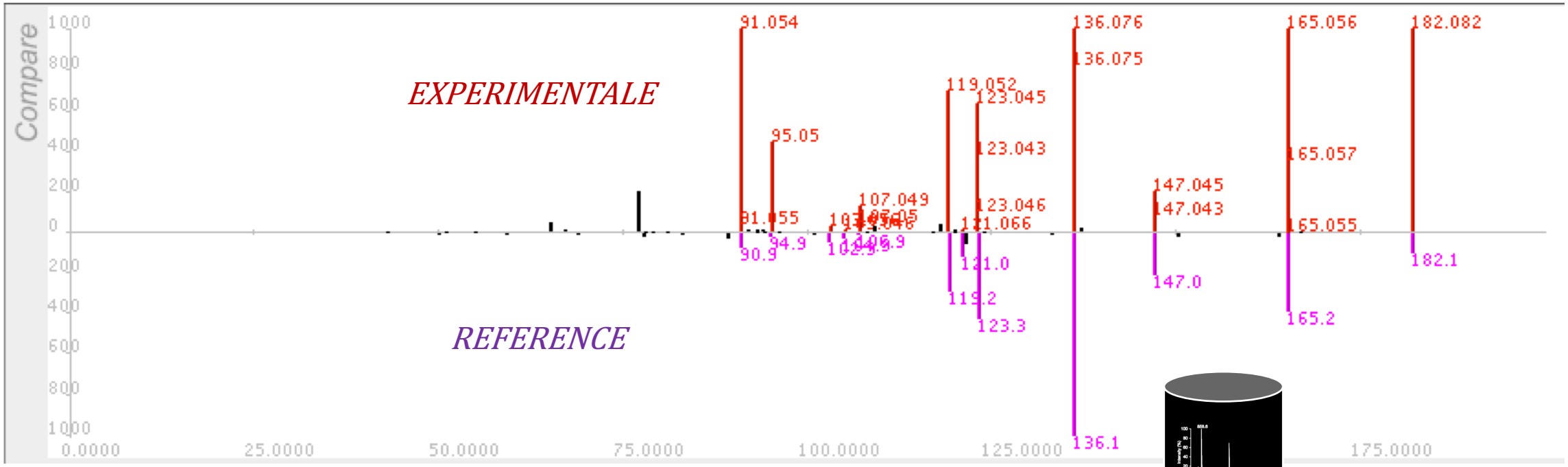
Spectral matching







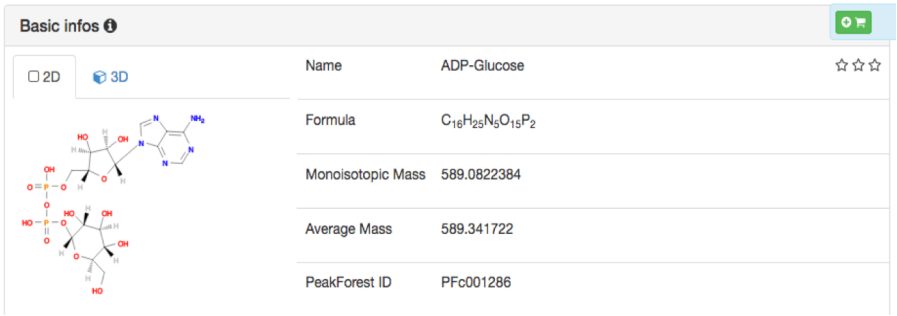
Des banques… des stratégies d'annotations



Spectral matching

Score of 0.83630…

PeakForest – collection de métabolomes



Fonctionnalités

Dépôts de métabolomes
SOP d'acquisition des données
Annotation de matrices biologiques
Curation de données, visualisation

Indicateurs

- **6 000** profiles métaboliques (MS & NMR) and **2 500** métabolites
- *Contributeurs* : > 20 (4 nœuds FR MetaboHUB)

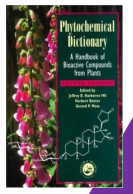
Améliorations en cours

- Ajout de métabolomes complets d'organisme modèles
- **Interopérabilité** avec W4M, MetExplore & dépôts européens.
- Devenir un dépôt de référence

PI : Franck Giacomoni

https://peakforest.org

Bases de connaissances



chemicalize.org
PubChem
ChEBI
HMDB
ChemAxon

Chemical Structures

In-house tool
Expert knowledge of food phytochemicals biotransformations

Predicted metabolites

Physico-chemical properties

Chemaxon Marvin
OpenBabel

PhytoHub

Phenol-Explorer
+ Literature survey

Known metabolites

Dietary sources

Books: Phytochemical dictionary…
Databases:
Phenol-Explorer FooDB
Dictionary of Food Compounds
USDA Dr. Duke's
+ Literature survey

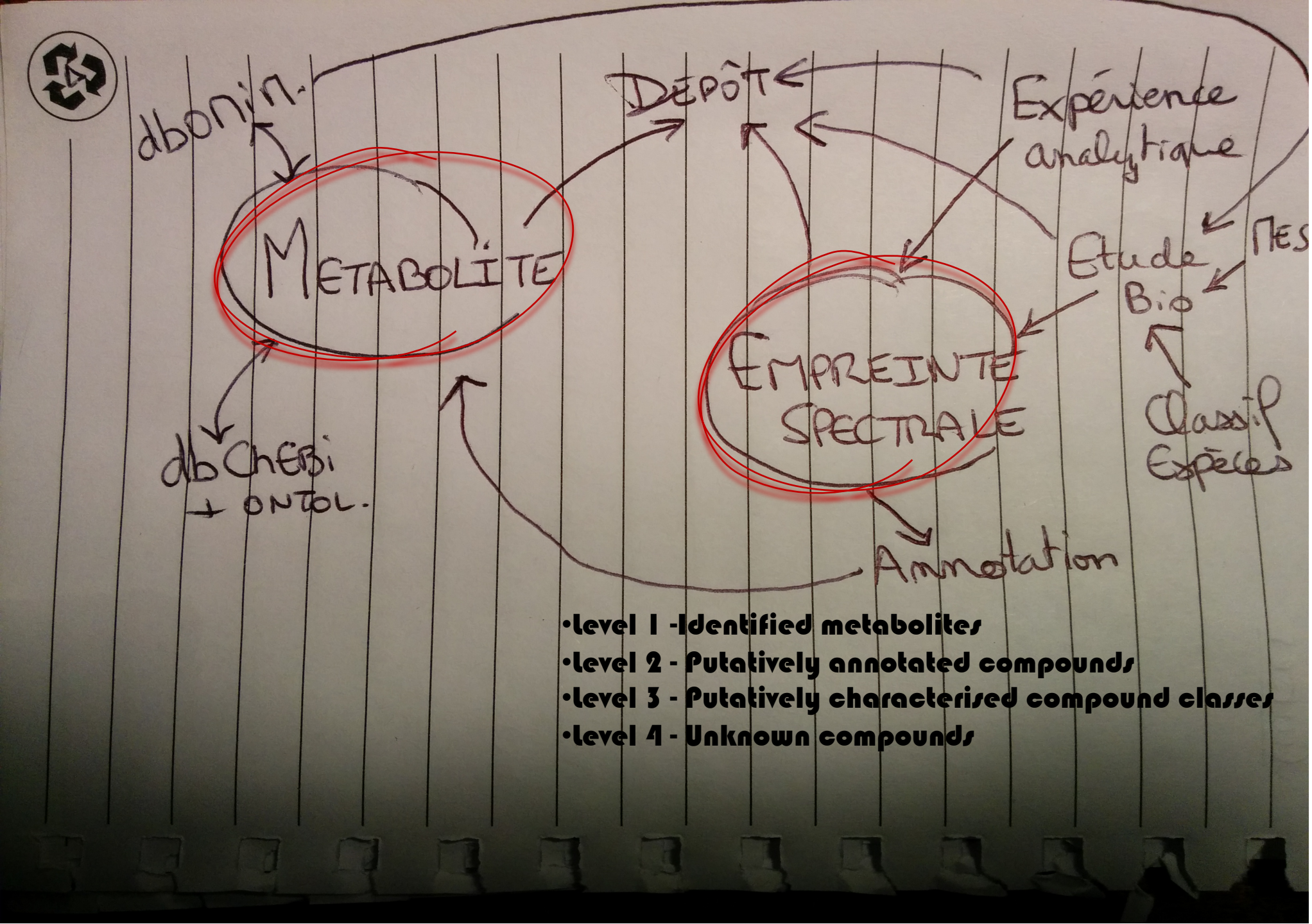
Spectral data

MassBank
ReSpect

- ✓ « Food Metabolome »
- ✓ Une base de données en ligne et en « open access »
- ✓ Dédiée aux phytomicronutriments et leurs métabolites

PI : Claudine Manach

www.phytohub.eu



- level 1 -Identified metabolites
- level 2 - Putatively annotated compounds
- level 3 - Putatively characterised compound classes
- level 4 - Unknown compounds

Enriched Data Format – Results and annot.

mzTab - *Reporting MS-based Proteomics and Metabolomics Results*

build passing

General

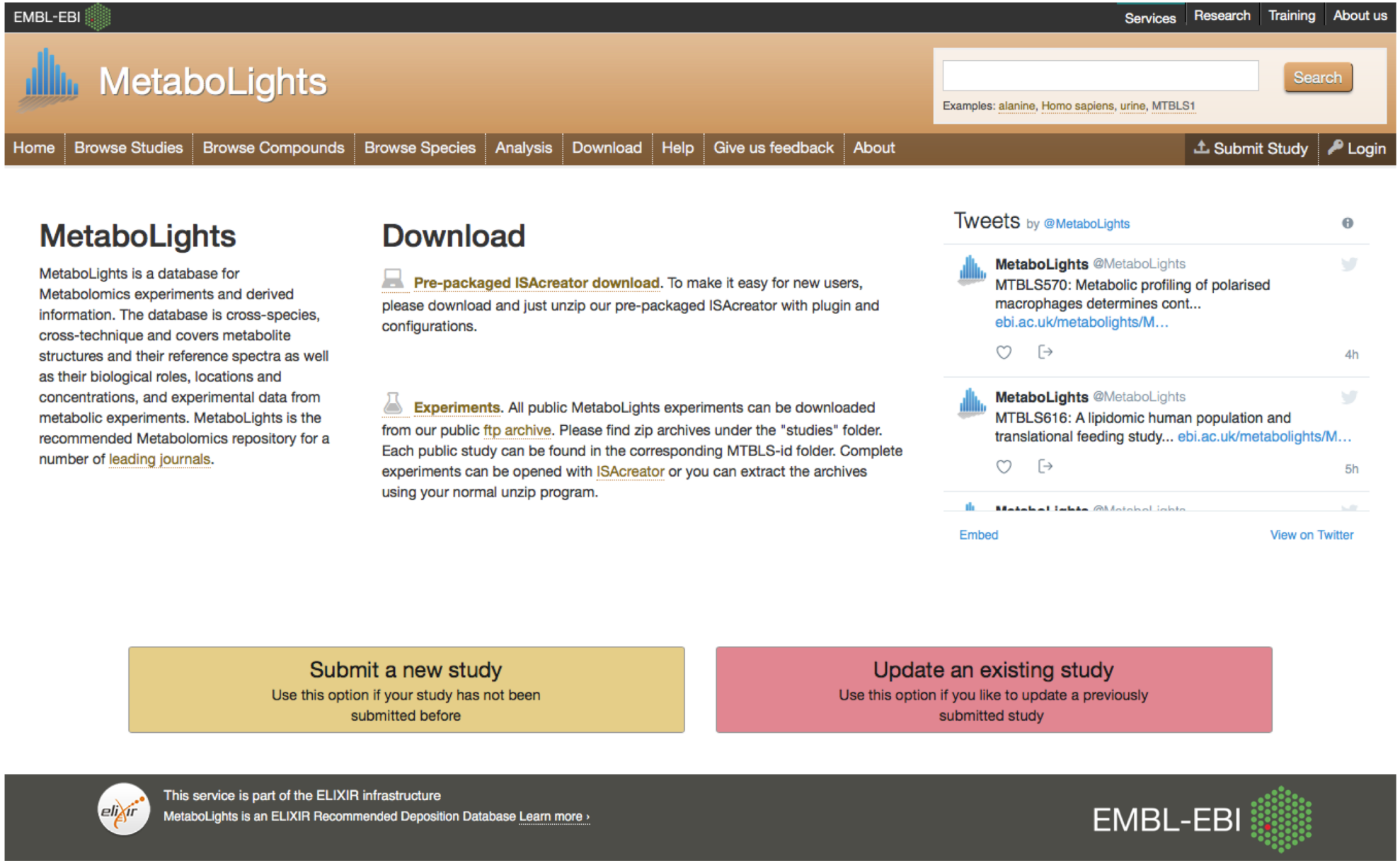
mzTab has been designed to act as a lightweight, tab-delimited file format for mass spec-derived omics data. It was originally designed for proteomics with limited support for metabolomics (version 1.0). The metabolomics aspects are undergoing further development towards full support in a planned version 2.0 release.

One of the main target audiences for this format is researchers outside of proteomics/metabolomics, such as systems biologists. It should be easy to parse and only contain the minimal information required to evaluate the results of an experiment. One of the goals of this file format is that it, for example, should be possible for a biologist to open such a file in Excel and still be able to "see" the data. This format should also become a way to disseminate proteomics and metabolomics results through protocols such as DAS (http://www.biodas.org).

The aim of the format is to present the results of an experiment in a computationally accessible overview. The aim is not to provide the detailed evidence for these results, or allow recreating the process which led to the results. Both of these functions are established through links to more detailed representations in other formats, in particular mzIdentML and mzQuantML for proteomics ID and quantitation.

Enriched Data Format – Study plan – ISA-TAB

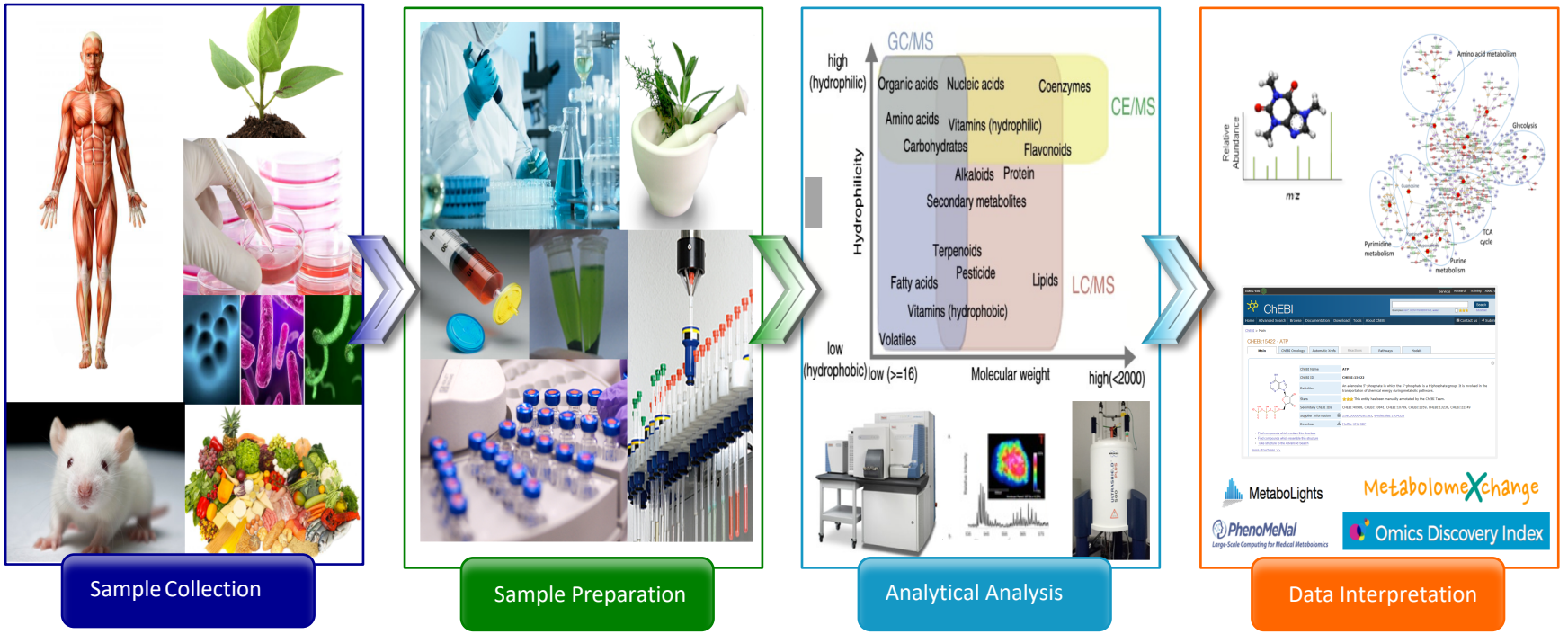
- Trois niveaux de descriptions : Investigation (1.1) / Study (1.n) / Assay (1.n)
- Configuration ISA spécifique à la biologie (**CIMR)**
  - ***e.g. in-vitro / plant / clinical study***
- Gestion des différents appareils d'acquisition
  - Profil ciblé VS métabolomique ouverte
  - LC-MS / GC-MS / DIMS / 1D - 2D NMR…
- Outils, packages R
- Format d'entrée des dépôts de référence (MetaboLigths)



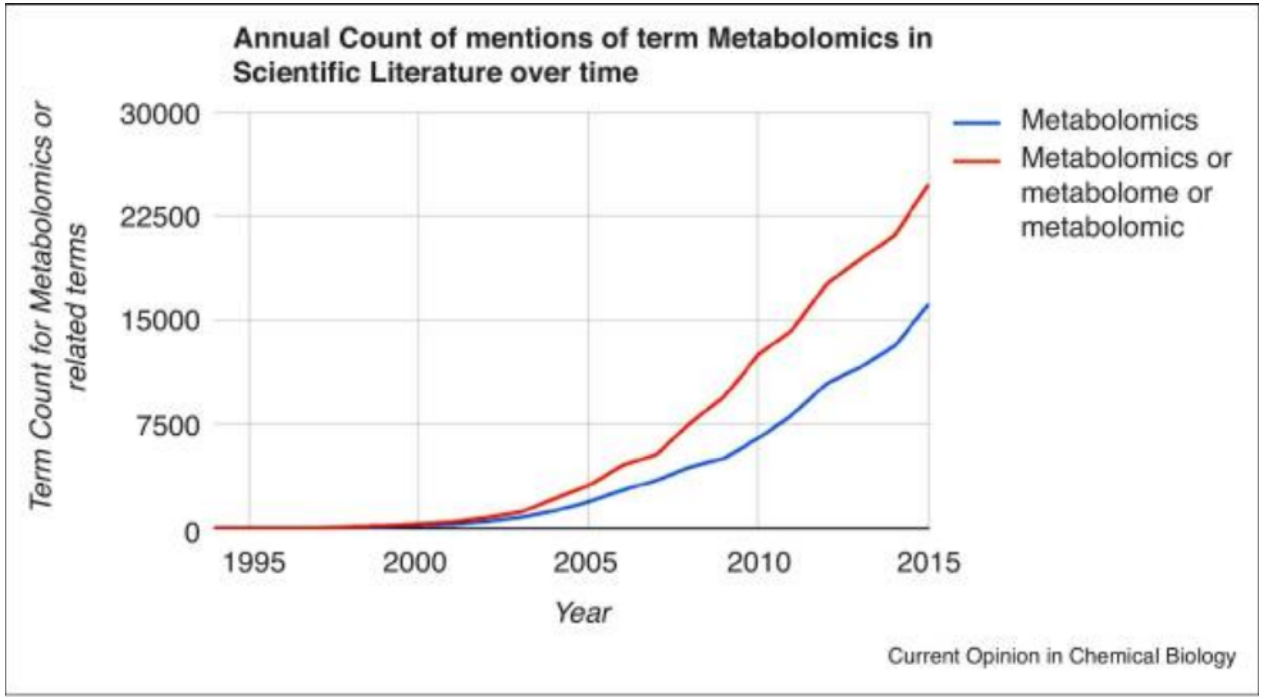
http://www.ebi.ac.uk/metabolights

Email: metabolights-help@ebi.ac.uk

MetaboLights - Open Source Database



- « Cross » espèces, « cross » techniques
- Collection d'information la plus complète possible
- « Open access », accessible via différents portails



Haug *et al* (10.1016/j.cbpa.2016.12.024)

17 900 publications (from 2012-2018)

497 public studies in the European MetaboLights

Including around 80 studies from the French community

917 studies in US NIH Metabolomics Workbench

less than 8.5% of submitted and publicly available studies

Personal work - INRA 2019

Ontologies / Informations en Métabolomique

- Décrire son étude — El biólogo
- Décrire son phénotype / modèle / échantillon — El biólogo
- Décrire ses conditions analytiques — Химик (*khimik*)
- Décrire son empreinte (son spectre) — Химик & 统计学家 (*Tǒngjì xué jiā*)
- Décrire ses métabolites chimiquement — Химик & הביואינפורמטיקאי
- Lier Métabolites et Biologie — הביואינפורמטיקאי & El biólogo & Химик & 统计学家

ChEBI ontology (EBI) - doc

- Molecular structure, in which molecular entities or parts thereof are classified according to composition and structure, e.g. hydrocarbons, carboxylic acids, tertiary amines;
- Role, divided into three sub-categories: 'chemical role' which classifies entities on the basis of their role within a chemical context, e.g. as ligand, inhibitor, surfactant; biological role which classifies entities on the basis of their role within a biological context, e.g. antibiotic, antiviral agent, coenzyme, hormone; and 'application' which classifies on the basis of their intended use by humans, e.g. pesticide, antirheumatic drug, fuel;
- Subatomic Particle, which classifies particles which are smaller than atoms, e.g. electron, photon, nucleon.

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies

Search for a class

Enter a class, e.g. Melanoma

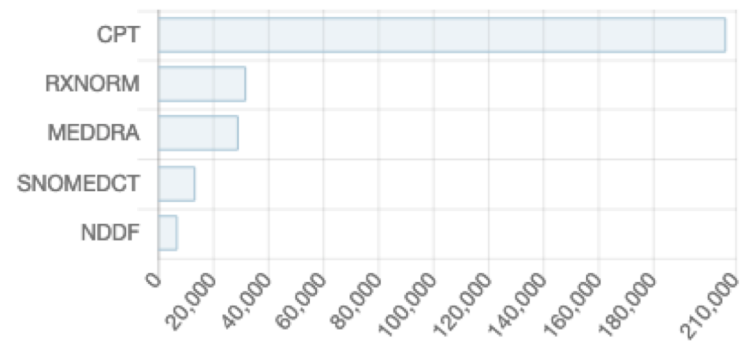
Advanced Search

Find an ontology

Start typing ontology name, then choose from

Browse Ontologies

Ontology Visits (September 2019)



More

BioPortal Statistics

Ontologies	816
Classes	8,808,907
Resources Indexed	48
Indexed Records	39,537,360
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932

GoDaddy

Welcome to metabolomics-msi.org

This Web page is parked for FREE, courtesy of GoDaddy.com.

Search for domains similar to metabolomics-msi.org

Get Started

Is this your domain? Let's turn it into a website!

Get Started

Would you like to buy this domain?

Learn More

$4.99* .COM

THE domain at THE price.

GET YOURS

or use this code at checkout GPPTCOM

*Restrictions apply. See website for details.

http://cosmos-fp7.eu/msi.html

The Metabolomics Standards Initiative (MSI) and Core Information for Metabolomics Reporting (CIMR)

Presentations:

Attachment	Size
In vivo /mammalian biology context	168.05 KB
Plant biology context	36.26 KB
In vitro biology/microbiology context	144.3 KB
Environmental analysis context	100.58 KB
Chemical Analysis Working Group	38.99 KB
Data analysis standards in metabolomics	120.12 KB
Ontology Working Group (OWG) road map	81.05 KB
Exchange format WG.pdf	

The Metabolomics Standards Initiative (MSI) and Core Information for Metabolomics Reporting (CIMR)

SCIENTIFIC DATA

OPEN

Comment: A decade after the metabolomics standards initiative it's time for a revision

Rachel A. Spicer[1], Reza Salek[1] & Christoph Steinbeck[1,2]

Received: 19 June 2017
Accepted: 22 August 2017
Published: 26 September 2017

A recent analysis of publicly available metabolomics data shows that the MSI guidelines are not well abided to in publicly shared metabolomics studies. We propose that the MSI guidelines should be revisited and revised, as has been done in other communities, to fit the current community needs.