Standardisation, modèles de données et ontologies en métabolomique

SAPI MetaSaurus - Séminaire de rentrée – Tours

26-27-28 Sept. 2022

Nils Paulhe, **Franck Giacomoni**, Olivier Filangi (EMPREINTE)
Sophie Aubin, David Benaben (PROSODIe)
Magalie Weber (Dpt Transform, DipSO)

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

INRAE

Introduction - What is Metabolomics?

Metabolomics is the large-scale study of small molecules, commonly known as metabolites, within cells, biofluids, tissues or organisms. Collectively, these small molecules and their interactions within a biological system are known as the metabolome.

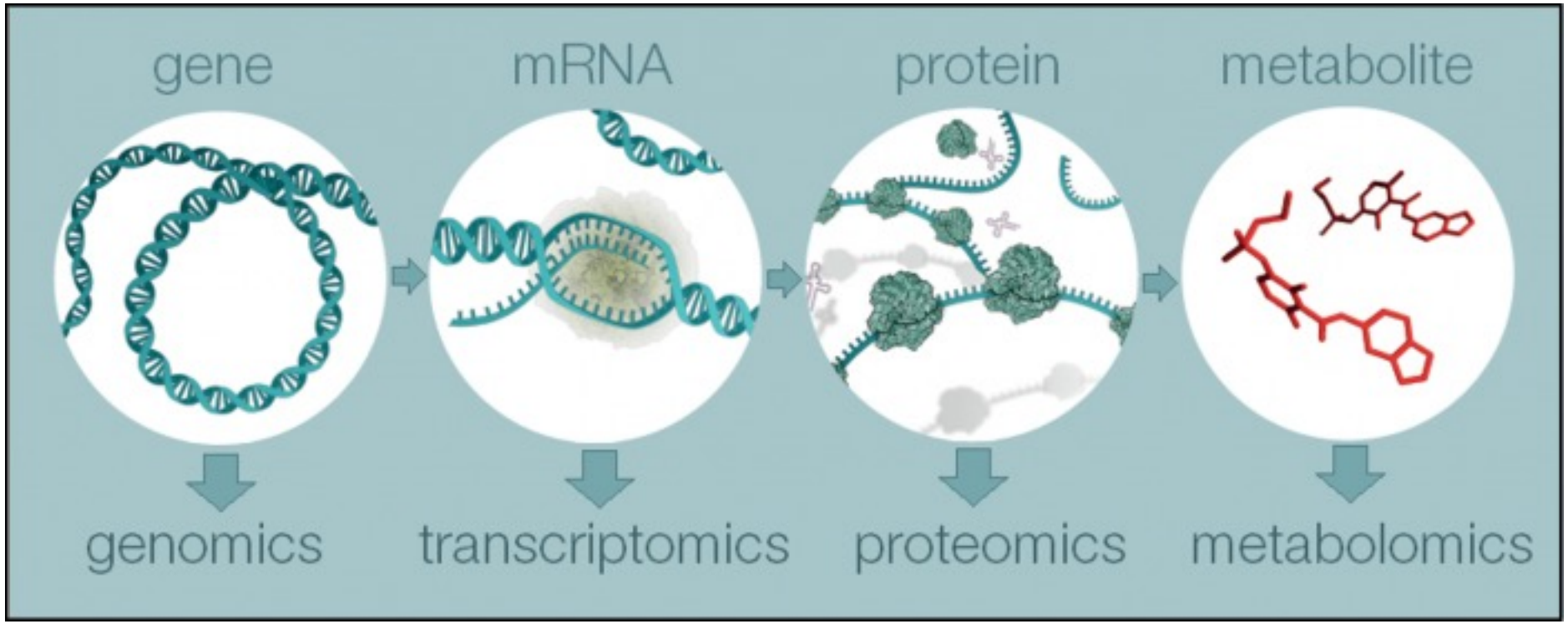
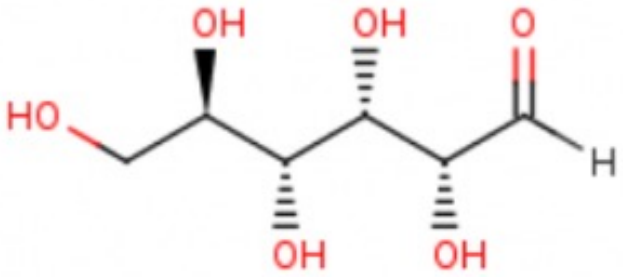


Figure from www.ebi.ac.uk/training/

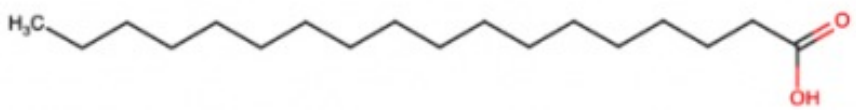
Small molecules

A small molecule (or metabolite) is a low molecular weight organic compound, typically involved in a biological process as a substrate or product. Metabolomics usually studies small molecules within a mass range of 50 – 1500 daltons (Da).



glucose
(a sugar)

Figures from www.ebi.ac.uk/training/



stearic acid
(a fatty acid)

Great deal variation in metabolites between biological species

200,000 across the plant kingdom, … 50,000 for *homo sapiens*

Very large underestimation

Figure from www.ebi.ac.uk/training/

Metabolomics study

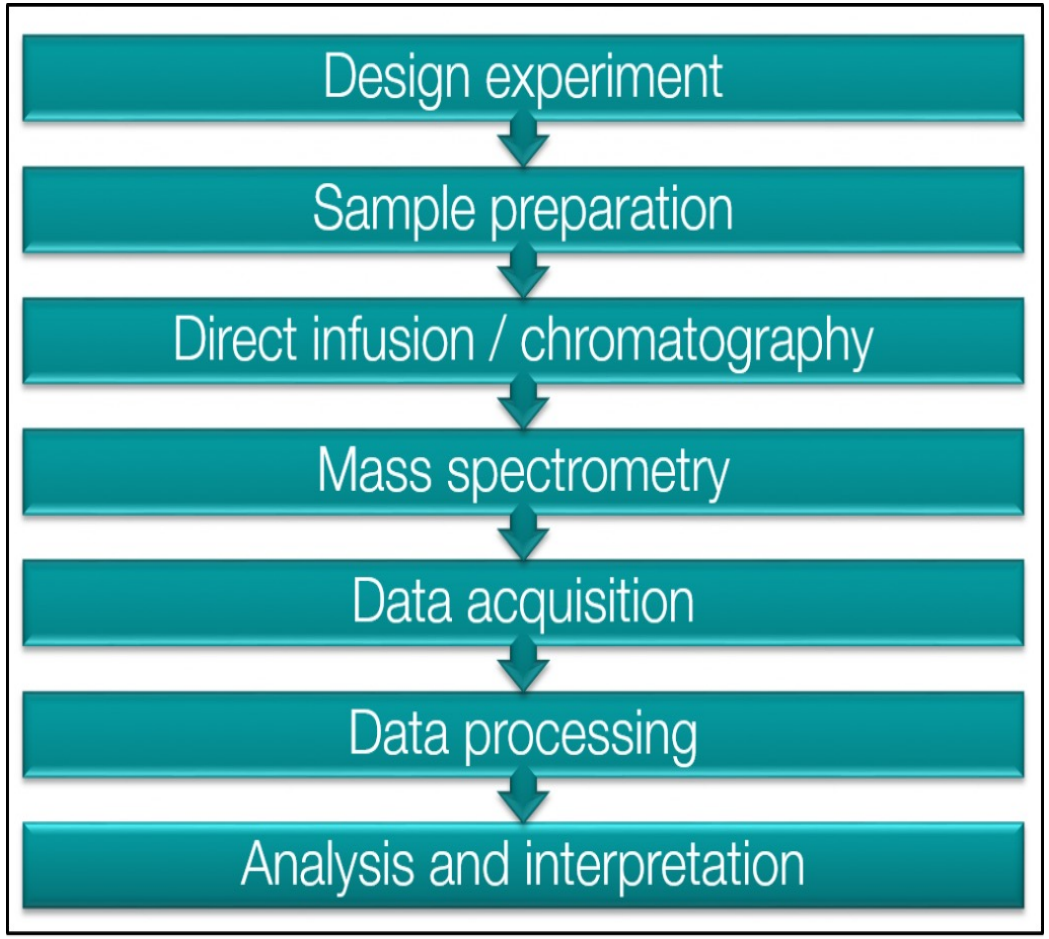


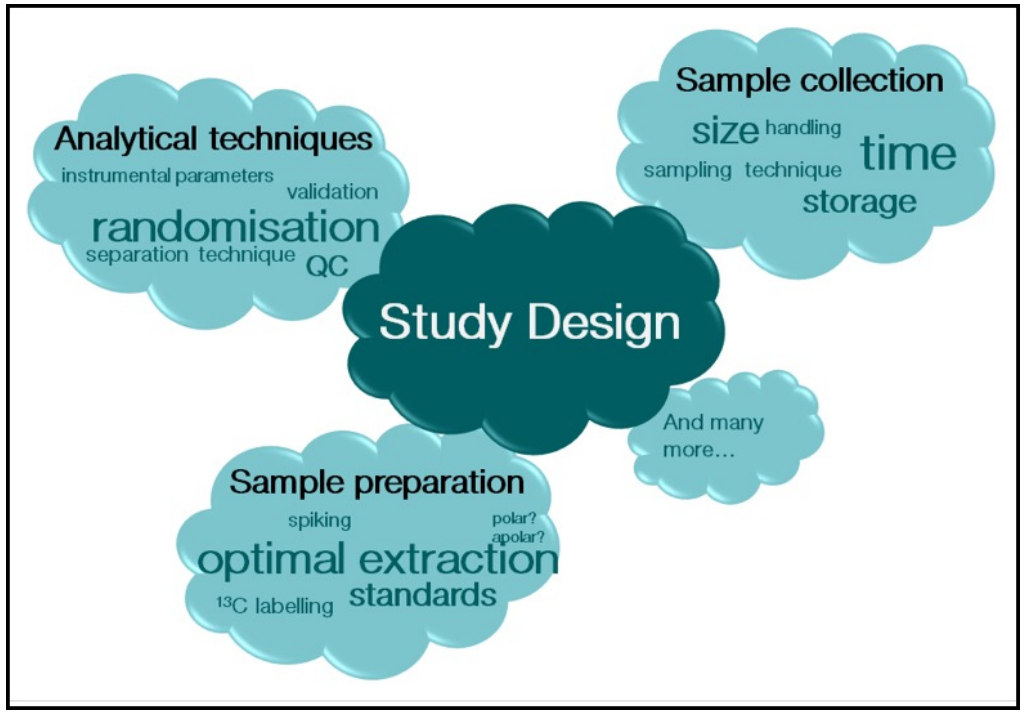
Figure from www.ebi.ac.uk/training/

Targeted approach:

This method is used when you want to measure sets of metabolites and have a specific biochemical question that you want to answer.

Untargeted (global) approach:

This method measures as many metabolites as possible from a range of biological samples without any (intended) bias.



Nuclear magnetic resonance (NMR) / Mass spectrometry (MS)

Bruker AvanceCore solution

NMR is an analytical technique that is used to measure organic and some inorganic compounds inside biological samples (as solid tissue or extracted metabolite). When a sample is exposed to a magnetic field and radio frequency (rf) pulse, the nuclei absorb and re-emit this electromagnetic radiation. The energy that is emitted has a specific resonating frequency, which depends on several factors including the magnetic properties of the atoms' isotopes and the strength of the magnetic field (usually referred to as chemical shifts). In the case of metabolomics, proton atoms from small molecules are usually investigated (^{1}H-NMR).

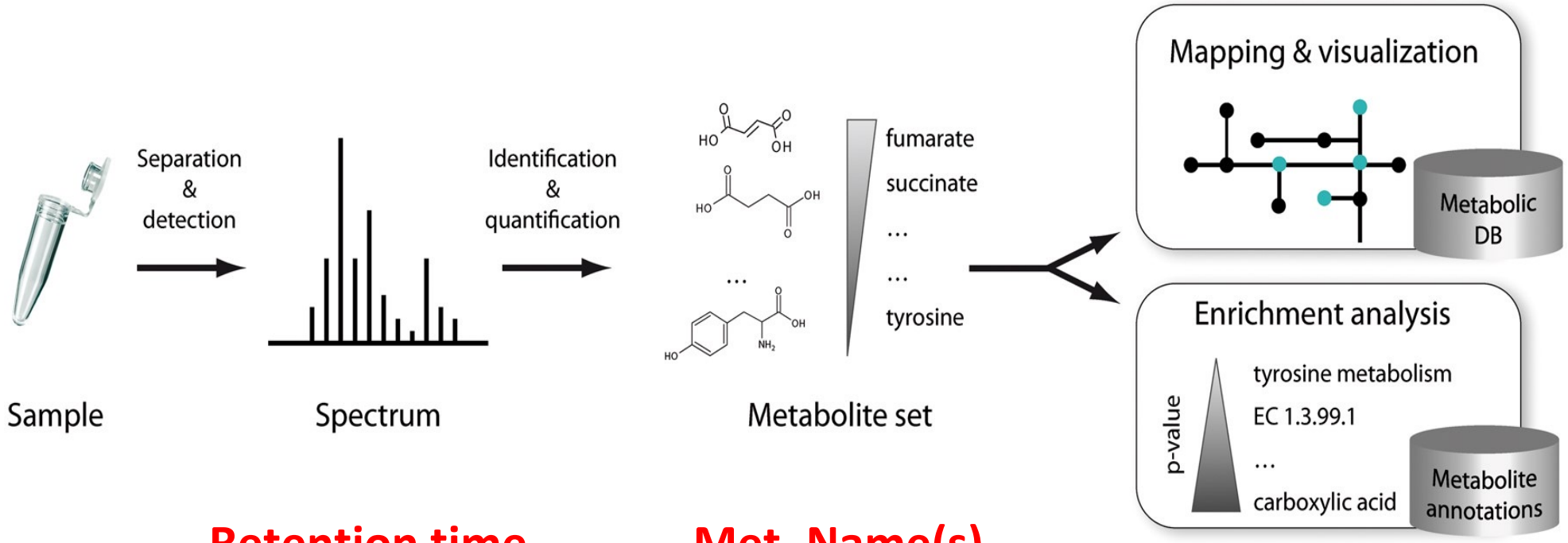
MS is an analytical technique used to measure small molecules. The small molecules may be either directly injected into the mass spectrometer (direct infusion) or through a coupled chromatographic system. The analytes are ionised at an ion source before they can be detected in a coupled mass detector. The resulting data typically consists of mass-to-charge (*m/z*), time, and intensity triplets that describe – for every detected ion mass – the strength of the ion beam and the time it is detected by the spectrometer.

Text from www.ebi.ac.uk/training/

TIMS TOF Pro LC-MS/MS System

Metabolomics data (MS example)

Figure: Metabolomics analysis workflow. A typical metabolomics experiment involves sample preparation, data ...



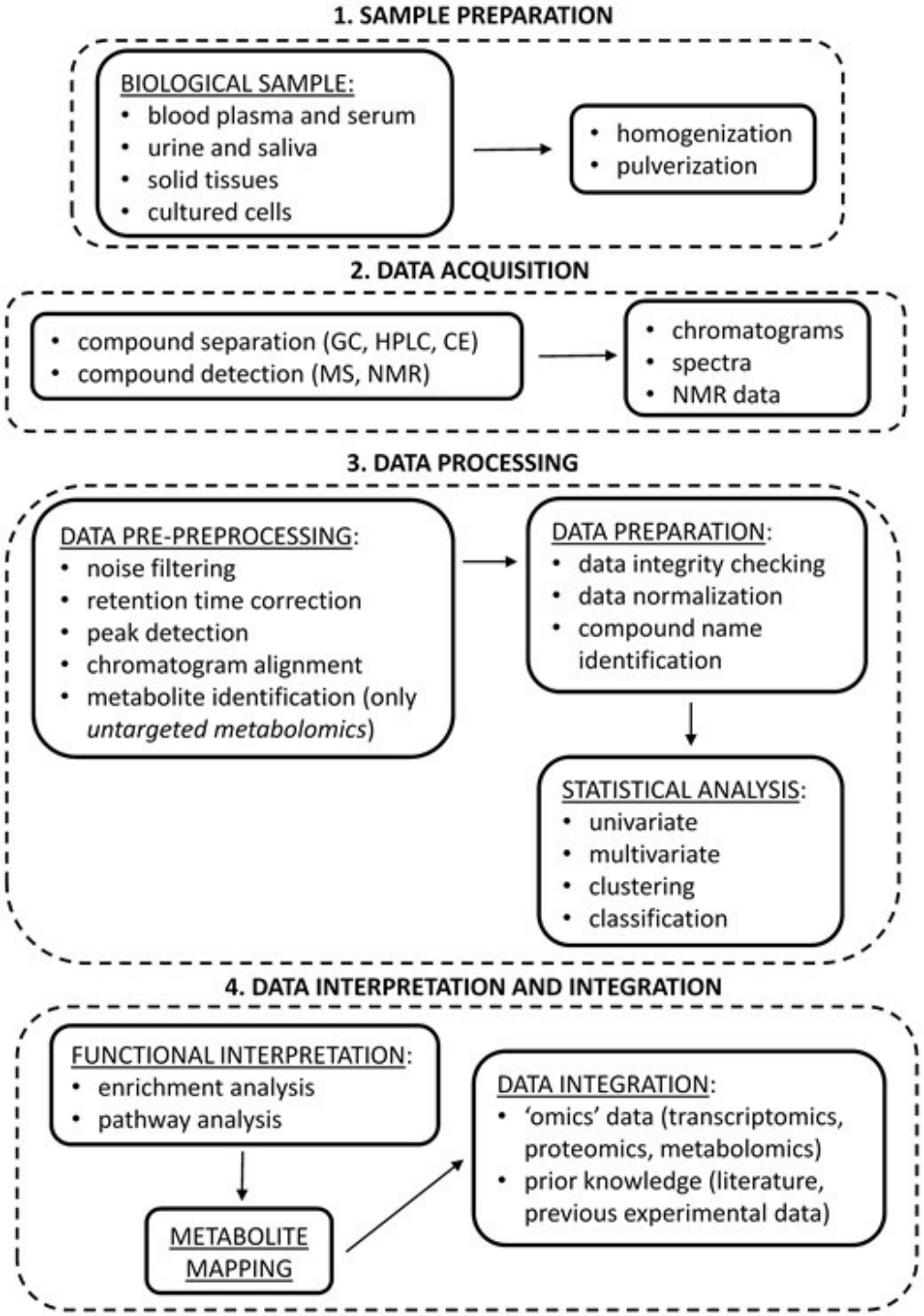
Retention time
Mass/charge
Relative intensity

Met. Name(s)
Met. Structure
Met. Properties

Brief Bioinform, Volume 14, Issue 6, November 2013, Pages 737–744, https://doi.org/10.1093/bib/bbs055

Metabolomics data

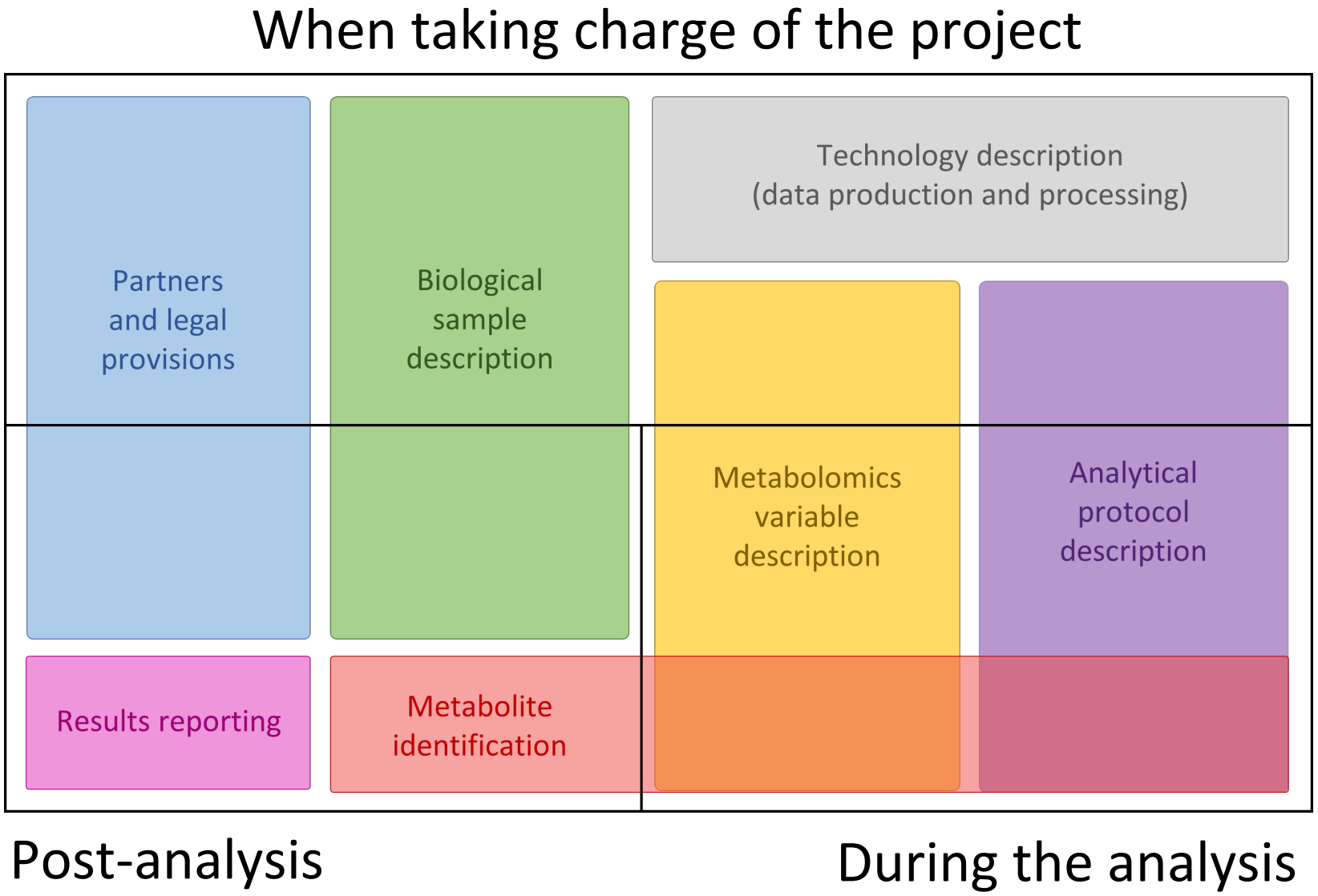
Flowchart of a typical metabolomic study. After sample preparation, specific metabolic signals are acquired using heterogeneous analytical platforms (**DATA ACQUISITION**). Raw signals are then pre-processed to produce data in a suitable format for univariate and multivariate statistical analyses. For *untargeted* studies, metabolites have first to be identified from spectral information (**DATA PROCESSING**). Significantly expressed metabolites are then linked to the biological context, through enrichment and pathway analysis, and mapped into networks. Finally, metabolomic data are integrated with other 'omics' data and with prior knowledge to gain a comprehensive view of the molecular processes involved (**DATA INTERPRETATION** AND INTEGRATION).



Alice Cambiaghi, Manuela Ferrario, Marco Masseroli, Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration, *Briefings in Bioinformatics*, Volume 18, Issue 3, May 2017, Pages 498–510, https://doi.org/10.1093/bib/bbw031

Metabolomics metadata

Based on poster "A project-scale map of metadata to improve future data management"



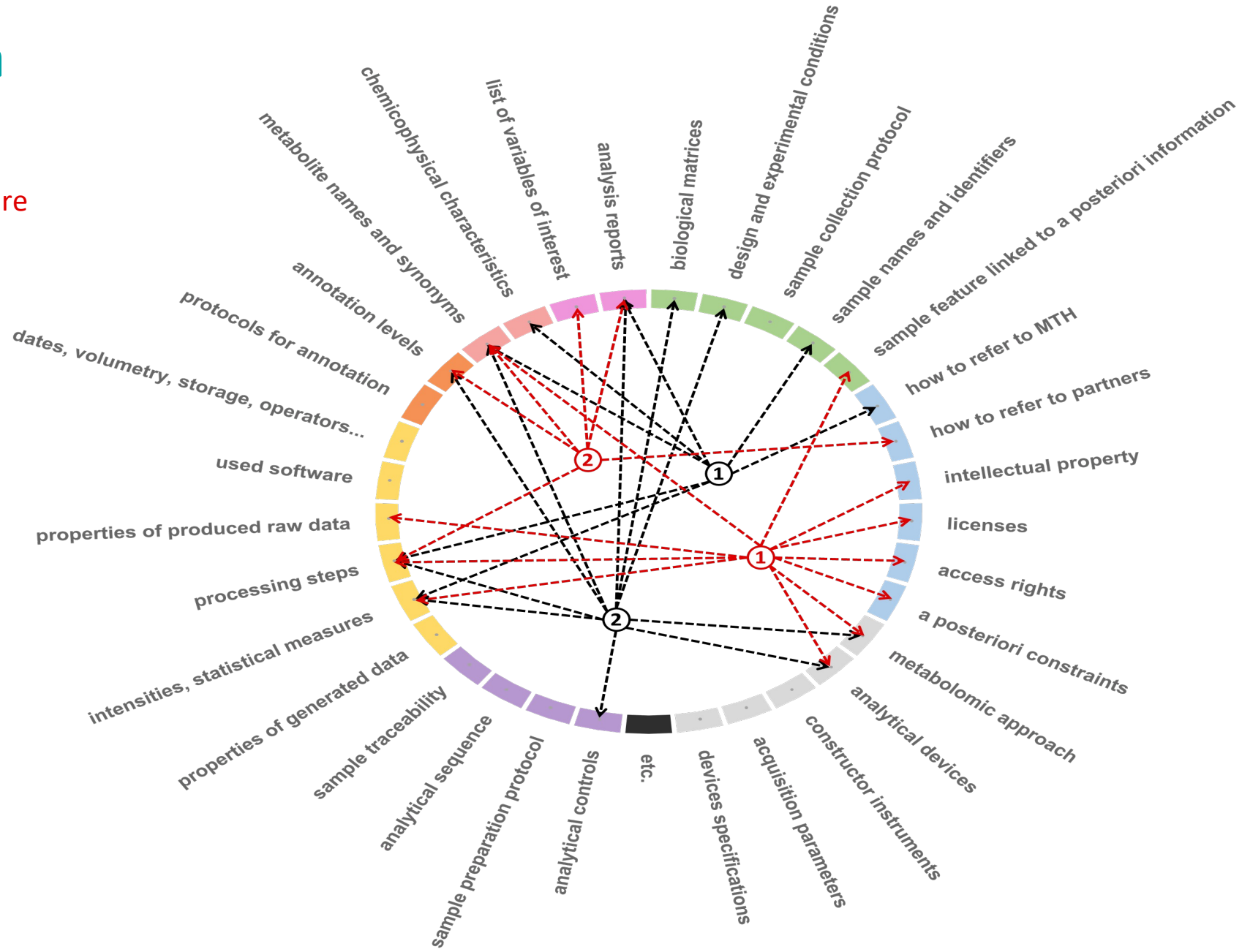
https://hal.inrae.fr/hal-03776494/document

Metabolomics metadata

Examples of potential working topics:

1. Naming standards inside the infrastructure
2. Harmonised use of ontologies

A project-scale map of metadata to improve future data management

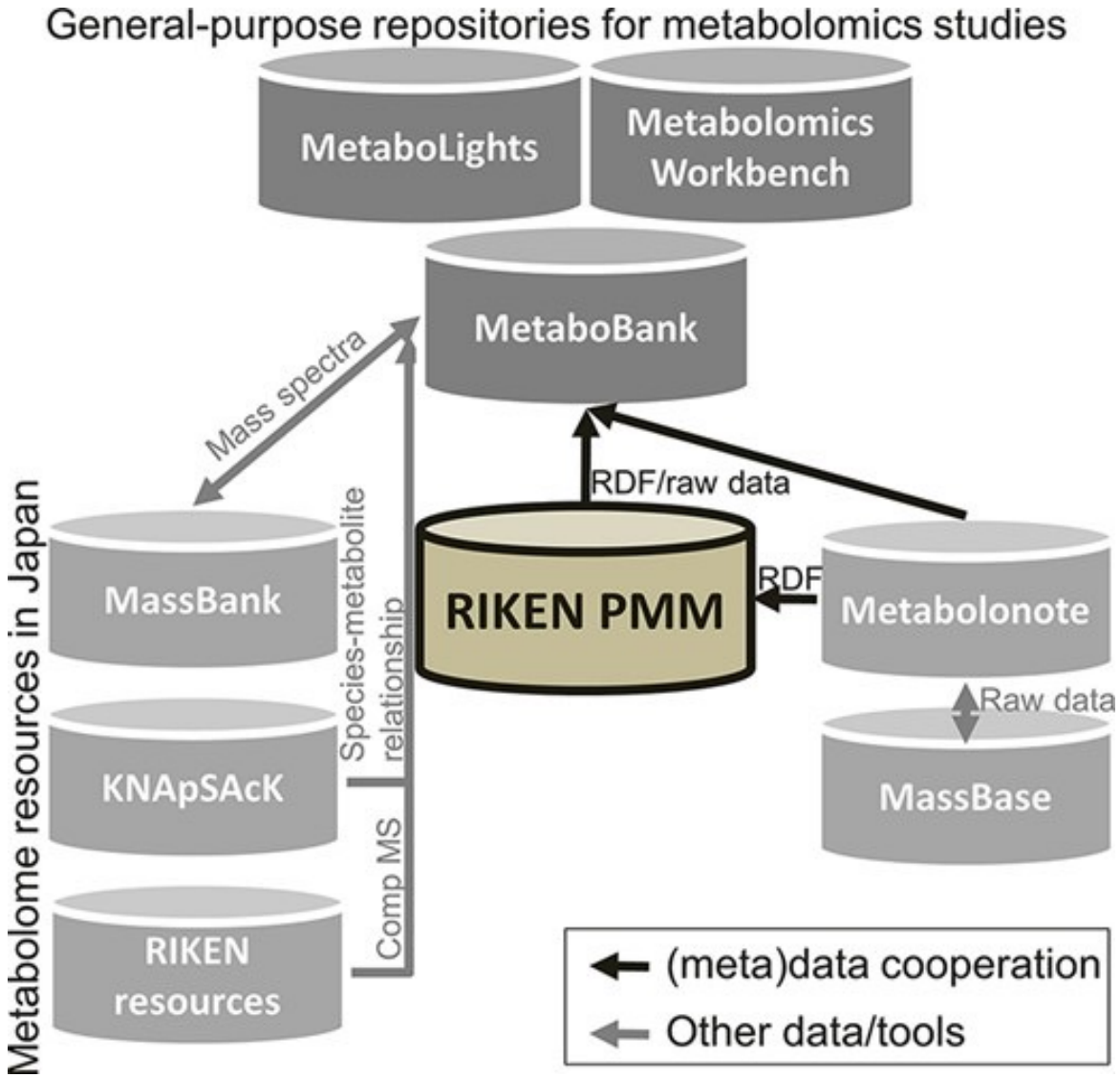


Poster : Pétéra *et al*, 2022 Analytics (https://hal.inrae.fr/hal-03776494/document)

Metabolomics Data & Information repositories

EBI - MetaboLights

MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments.



NHI – Metabolomics Workbench

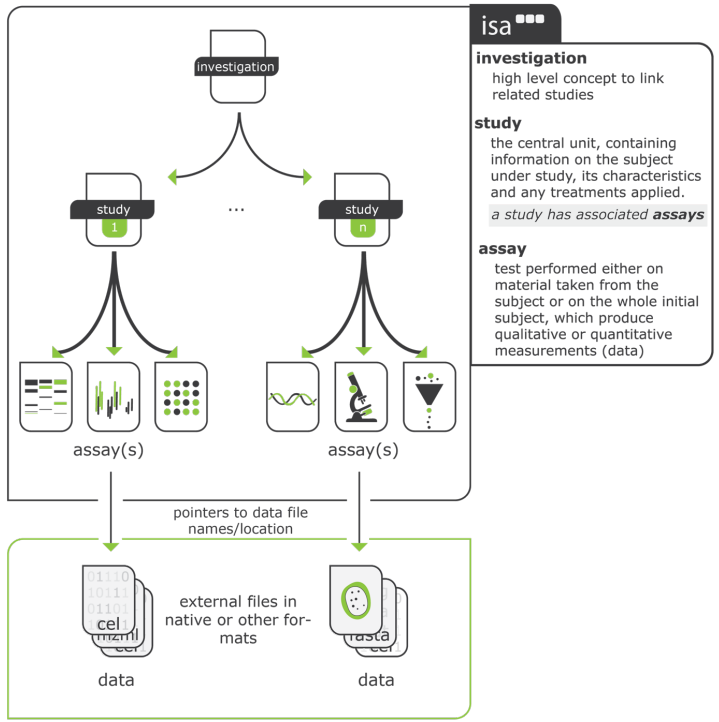
The Metabolomics Workbench serves as a national and international repository for metabolomics data and metadata and provides analysis tools and access to metabolite standards, protocols, tutorials, training, and more.

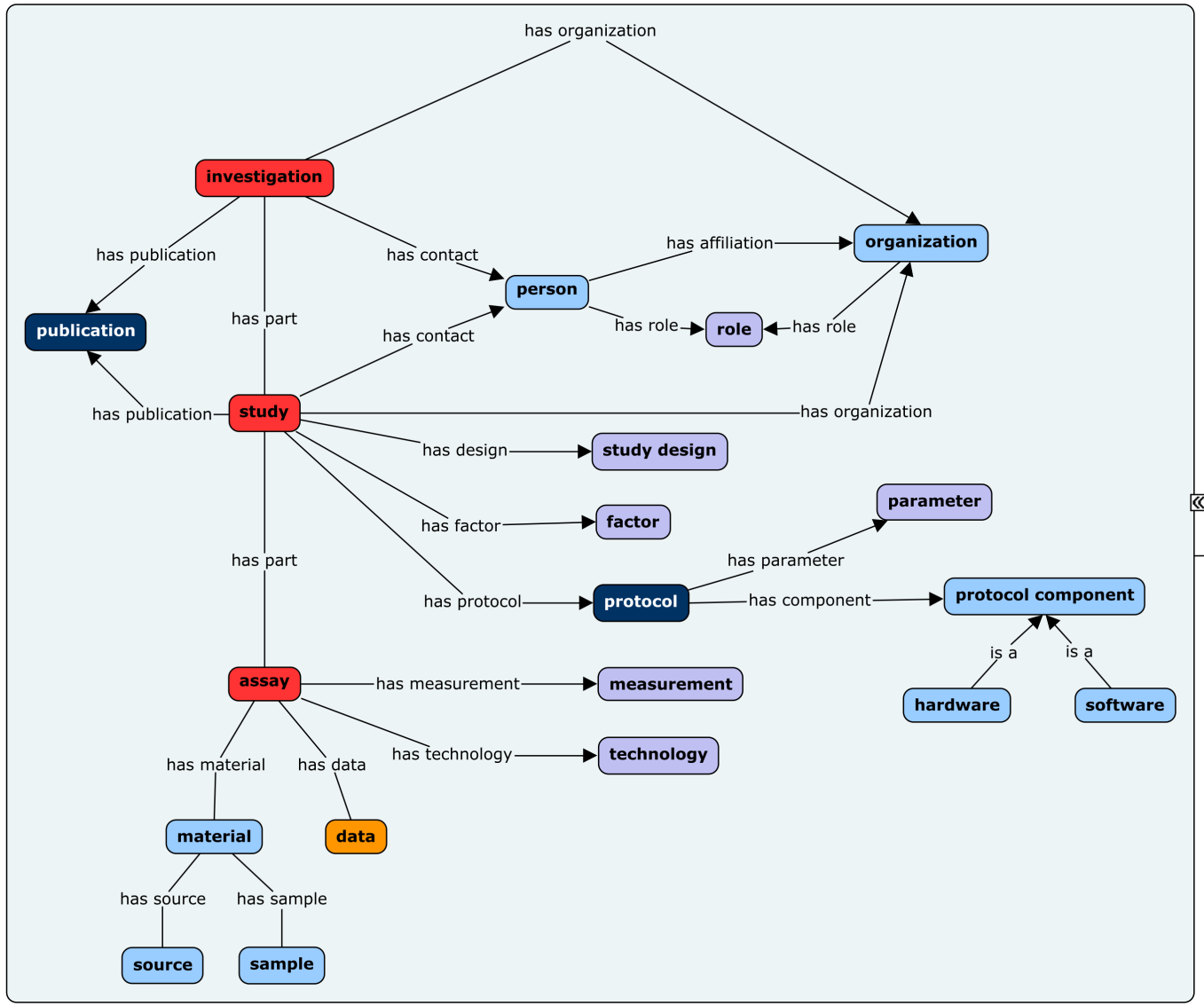
From Fukushima et al 2022 - https://doi.org/10.1093/pcp/pcab173

MTBLS model

EBI - MetaboLights

Powered by **ISA**
Abstract Model





https://isa-specs.readthedocs.io/en/latest/

MTWB model

NHI – Metabomics Workbench

Powered by **mwTab Format**

Version 1.5 (Mar,2022)

MS:INSTRUMENT_NAME
MS:INSTRUMENT_TYPE
MS:MS_TYPE
MS:ION_MODE
MS:MS_COMMENTS
MS:CAPILLARY_TEMPERATURE
MS:CAPILLARY_VOLTAGE
MS:COLLISION_ENERGY
MS:COLLISION_GAS
MS:DRY_GAS_FLOW
MS:DRY_GAS_TEMP
MS:FRAGMENT_VOLTAGE
MS:FRAGMENTATION_METHOD
MS:GAS_PRESSURE
MS:HELIUM_FLOW
MS:ION_SOURCE_TEMPERATURE
MS:ION_SPRAY_VOLTAGE
MS:IONIZATION
MS:IONIZATION_ENERGY
MS:IONIZATION_POTENTIAL

MS fields

MS:MASS_ACCURACY
MS:PRECURSOR_TYPE
MS:REAGENT_GAS
MS:SOURCE_TEMPERATURE
MS:SPRAY_VOLTAGE
MS:ACTIVATION_PARAMETER
MS:ACTIVATION_TIME
MS:ATOM_GUN_CURRENT
MS:AUTOMATIC_GAIN_CONTROL
MS:BOMBARDMENT
MS:CDL_SIDE_OCTOPOLES_BIAS_VOLTAGE
MS:CDL_TEMPERATURE
MS:DATAFORMAT
MS:DESOLVATION_GAS_FLOW
MS:DESOLVATION_TEMPERATURE
MS:INTERFACE_VOLTAGE
MS:IT_SIDE_OCTOPOLES_BIAS_VOLTAGE
MS:LASER
MS:MATRIX

MS:NEBULIZER
MS:OCTPOLE_VOLTAGE
MS:PROBE_TIP
MS:RESOLUTION_SETTING
MS:SAMPLE_DRIPPING
MS:SCAN_RANGE_MOVERZ
MS:SCANNING
MS:SCANNING_CYCLE
MS:SCANNING_RANGE
MS:SKIMMER_VOLTAGE
MS:TUBE_LENS_VOLTAGE
MS:MS_RESULTS_FILE

(required fields in red)

https://www.metabolomicsworkbench.org/data/mwTab_specification.pdf

https://mwtab.readthedocs.io/en/latest/

MassBank project

MassBank Record Format 2.6.0 –

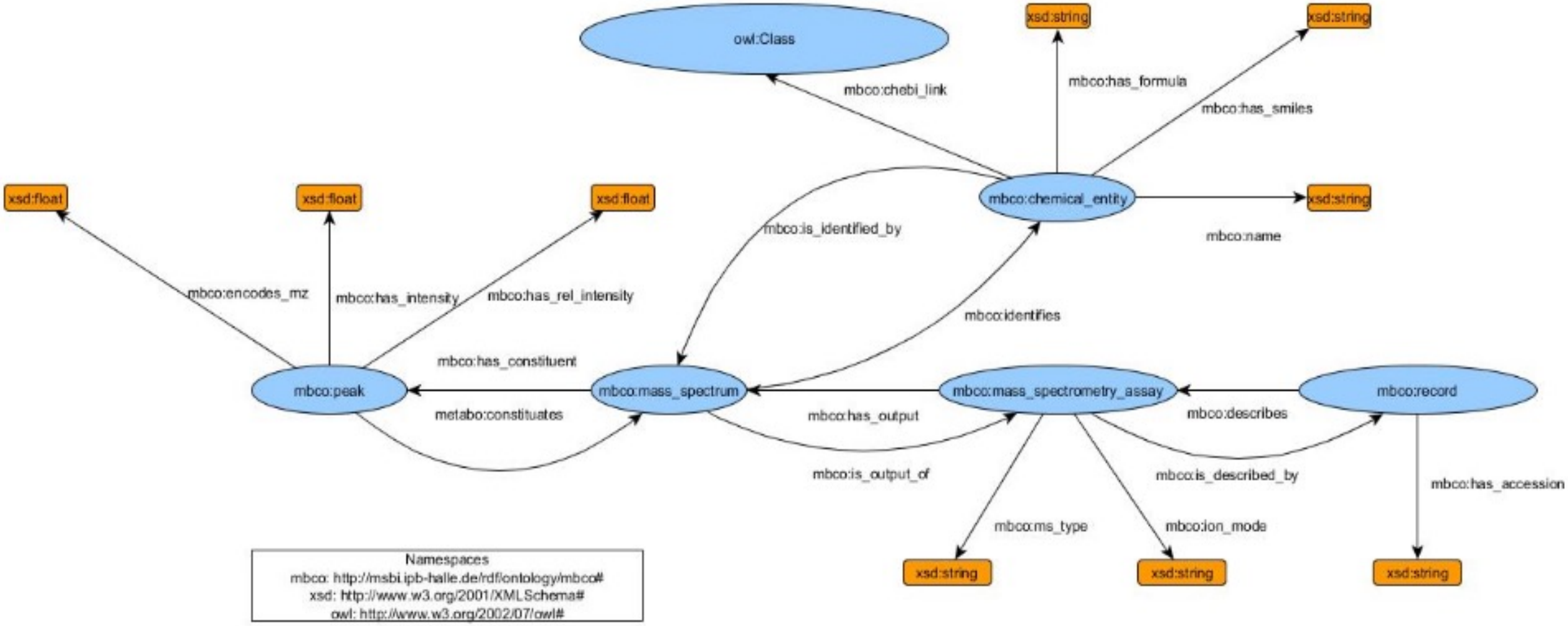
MassBank Consortium (May 13, 2022)

https://github.com/MassBank/MassBank-web/blob/main/Documentation/MassBankRecordFormat.md

Tag	Mandatory/ Optional	Unique/ Iterative	Single line/ Multiple line	Description	Subsection in manual
Record Specific Information					
ACCESSION	M	U	S	Record identifier	2.1.1
RECORD_TITLE	M	U	S	Short title of the record	2.1.2
DATE	M	U	S	Date of creation or last modification of record	2.1.3
AUTHORS	M	U	S	Name and affiliation of authors	2.1.4
LICENSE	M	U	S	Creative Commons License or its compatible terms	2.1.5
COPYRIGHT	O	U	S	Copyright	2.1.6
PUBLICATION	O	U	S	Bibliographic information of reference	2.1.7
PROJECT	O	U	S	Information on a related project)	2.1.8
COMMENT	O	I	S	Comments	2.1.9
Information of Chemical Compound Analyzed					
CH$NAME	M	I	S	Chemical name	2.2.1
CH$COMPOUND_CLASS	M	U	S	Chemical category	2.2.2
CH$FORMULA	M	U	S	Chemical formula	2.2.3
CH$EXACT_MASS	M	U	S	Exact mass	2.2.4
CH$SMILES	M	U	S	SMILES code	2.2.5
CH$IUPAC	M	U	S	InChI code	2.2.6

...

The « old » current state of RDF MassBank (From source)



Metabolomics standards advances

MIAMET (Minimum Information About a METabolomics experiment) - 2004

Bino, RJ et al. Potential of metabolomics as a functional genomics tool, Trends in Plant Science, Volume 9, Issue 9, 2004, Pages 418-425, ISSN 1360-1385, 10.1016/j.tplants.2004.07.004.

The metabolomics standards initiative (MSI) – 2007

Fiehn, O., Robertson, D., Griffin, J. et al. The metabolomics standards initiative (MSI). Metabolomics 3, 175–178 (**2007**). 10.1007/s11306-007-0070-6

MSI Board Members:. The Metabolomics Standards Initiative. Nat Biotechnol 25, 846–848 (**2007**). 10.1038/nbt0807-846b

Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI) – 2007 [http://msi-workgroups.sourceforge.net/]

Sumner LW,et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. **2007** 10.1007/s11306-007-0082-2. PMID: 24039616;

EU FP7 COordination of Standards in MetabOlomicS (COSMOS)

Salek, R.M. et al. COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. Metabolomics (**2015**). 10.1007/s11306-015-0810-y

Schober D., et al. NmrML: A Community Supported Open Data Standard for the Description, Storage, and Exchange of NMR Data Ana.Chem. (**2018)** 10.1021/acs.analchem.7b02795

EU H2020 Phenome and Metabolome aNalysis (PhenoMeNal)

Opinions

Salek RM, Haug K, Steinbeck C. Dissemination of metabolomics results: role of MetaboLights and COSMOS. Gigascience. **2013** May 17;2(1):8. doi: 10.1186/2047-217X-2-8. PMID: 23683662;

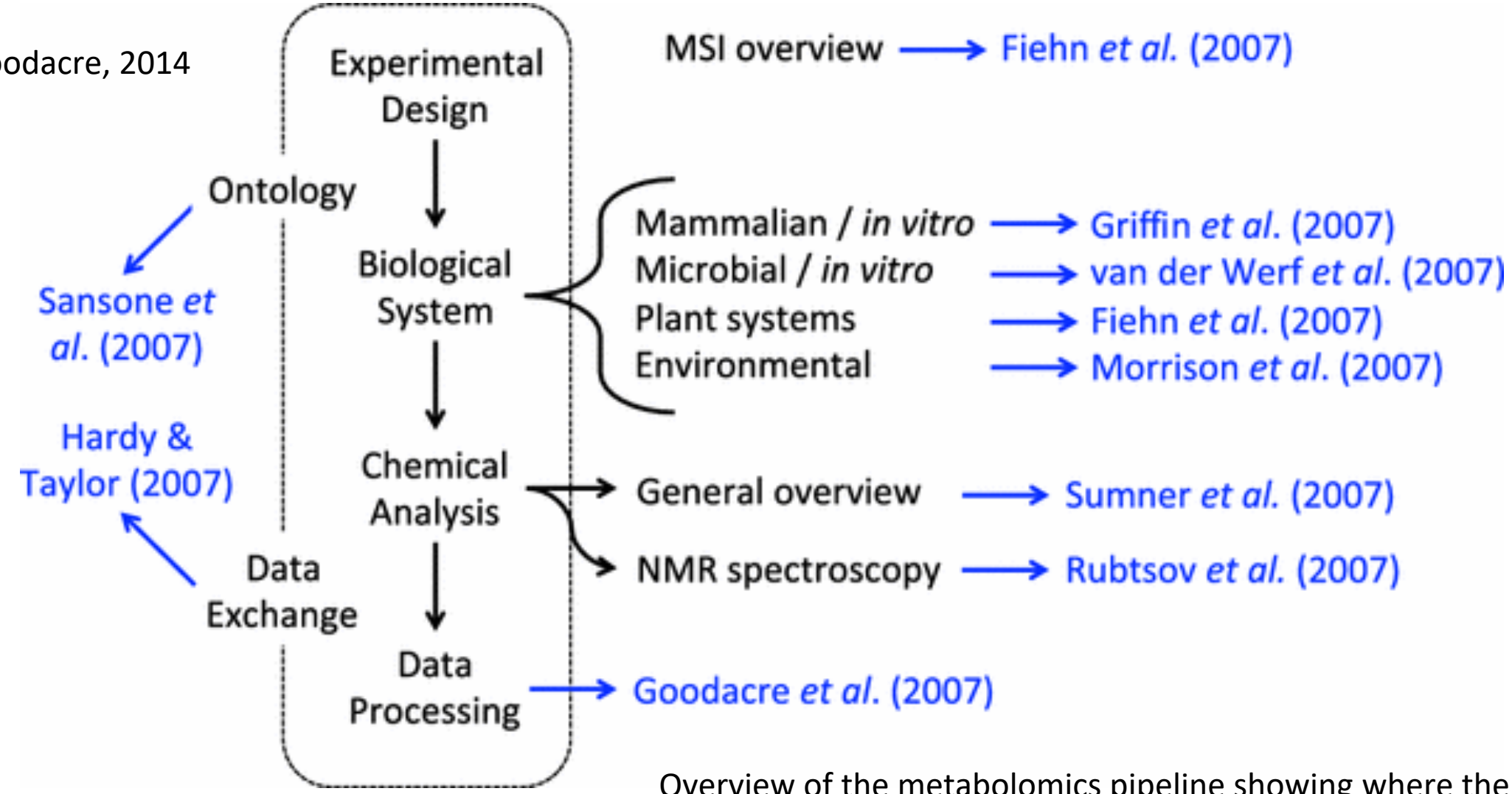
Goodacre, R. Water, water, every where, but rarely any drop to drink. Metabolomics 10, 5–7 (**2014**). 10.1007/s11306-013-0618-6

Spicer RA, Salek R, Steinbeck C. A decade after the metabolomics standards initiative it's time for a revision. Sci Data. **2017** Sep 26;4:170138. doi: 10.1038/sdata.2017.138. PMID: 29989594;

Viant, M.R. et al. Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. Nat Commun 10, 3041 (**2019**). 10.1038/s41467-019-10900-y

Aiming of MSI reporting standards

From Goodacre, 2014

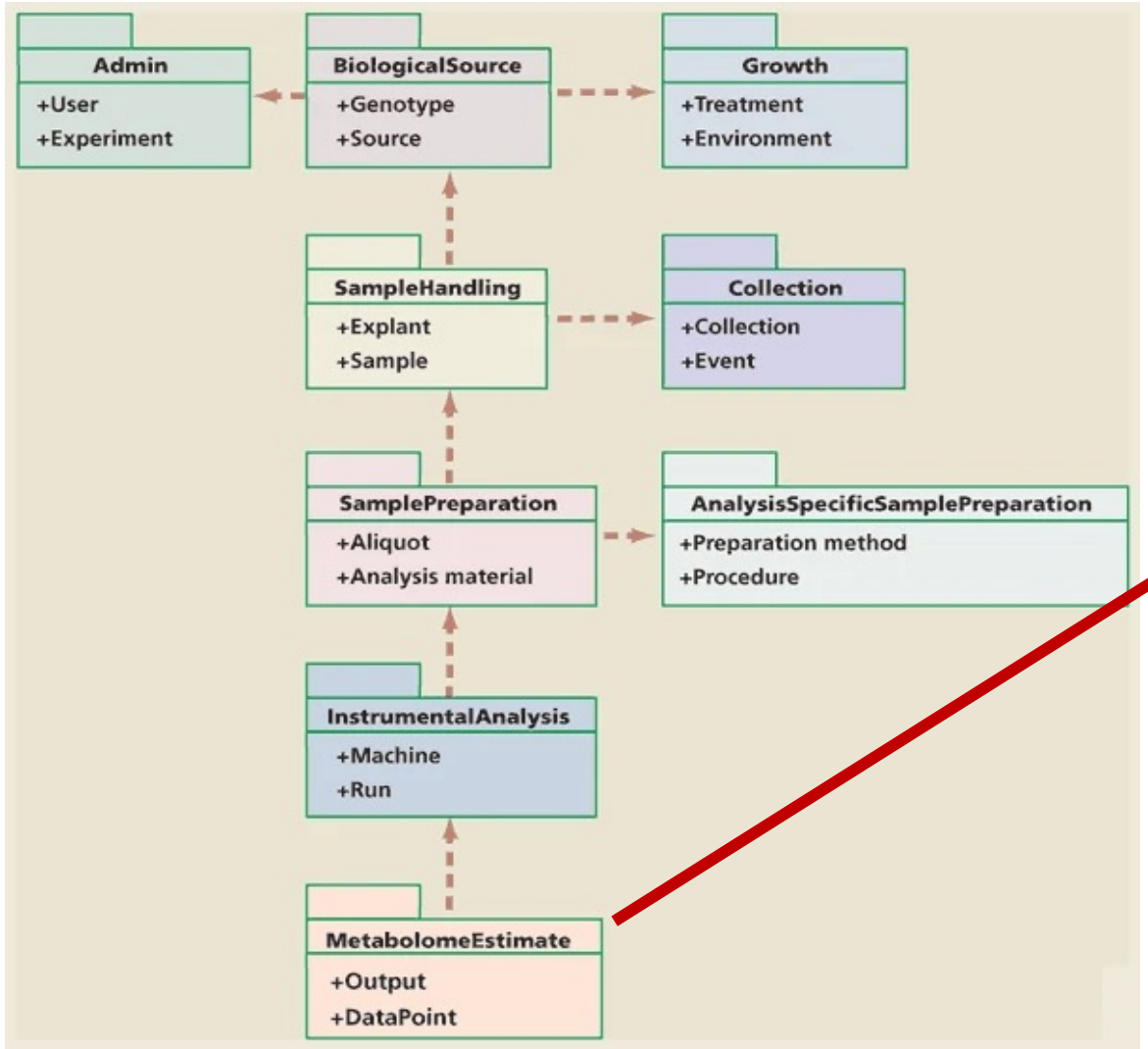


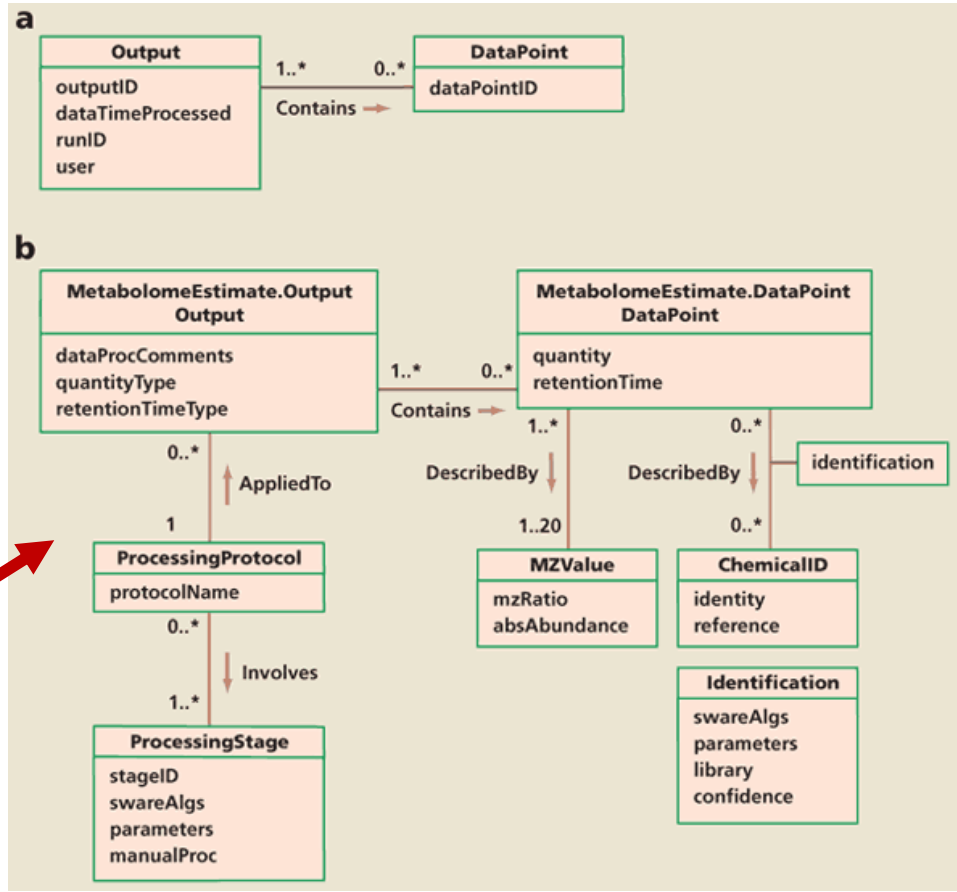
Overview of the metabolomics pipeline showing where the MSI reporting standards are aimed (*highlighted in blue*) (Color figure online)

Designed data metabolomics models

- **Summary recommendations for standardization and reporting of metabolic analyses**
  - The Standard Metabolic Reporting Structures working group. Nat Biotechnol 23, 833–838 (2005). https://doi.org/10.1038/nbt0705-833
- **ArMet: A proposed framework for the description of plant metabolomics experiments and their results**
  - Jenkins H, Hardy N. Nat Biotechnol. **2004** doi: 10.1038/nbt1041. PMID: 15583675.

Standard Metabolic Reporting Structures (SMRS)
Architecture for metabolomics (ArMet)





The 'MetabolomeEstimate' component.

(**a**) The core definition. (**b**) An example subcomponent to support metabolomics data, as produced by GC-MS

Ontologies | controlled vocabulary

OWG - Sansone, SA., Schober, D., Atherton, H.J. *et al.* Metabolomics standards initiative: ontology working group work in progress. *Metabolomics* **3**, 249–256 (2007). 10.1007/s11306-007-0069-z

Experimental metadata

http://www.ncbi.nlm.nih.gov/Taxonomy/

http://www.plantontology.org/

http://environmentontology.org/

…

Analytical metadata

…

Spectral data

…

Compounds

…

1,017 entries

277 entries

https://www.ebi.ac.uk/ols/ontologies

March 24, 2022

Philip Strömert*, Johannes Hunold, André Castro, Steffen Neumann and Oliver Koepler

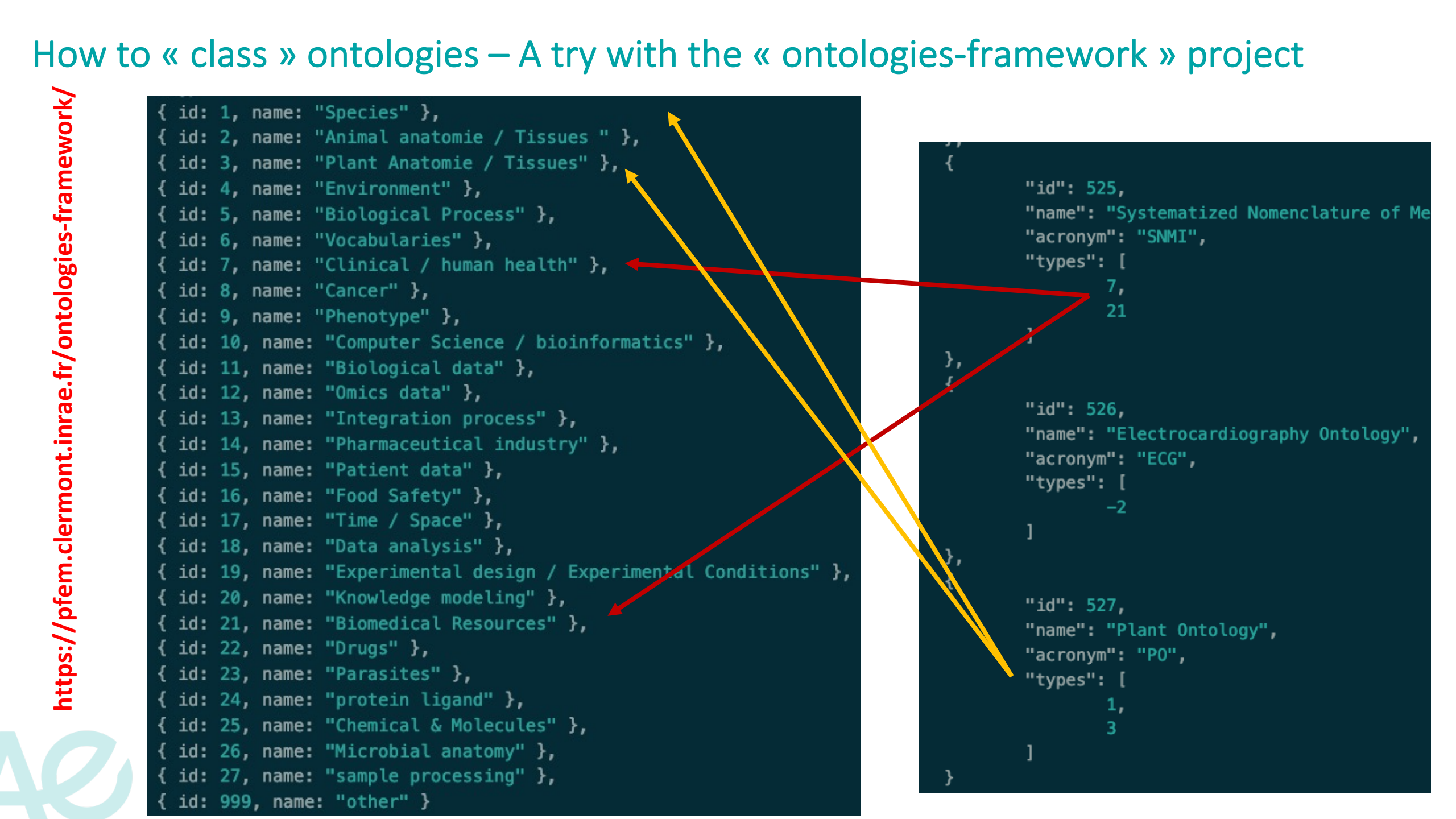
Ontologies4Chem: the landscape of ontologies in chemistry

https://doi.org/10.1515/pac-2021-2007

How to « class » ontologies – A try with the « ontologies-framework » project

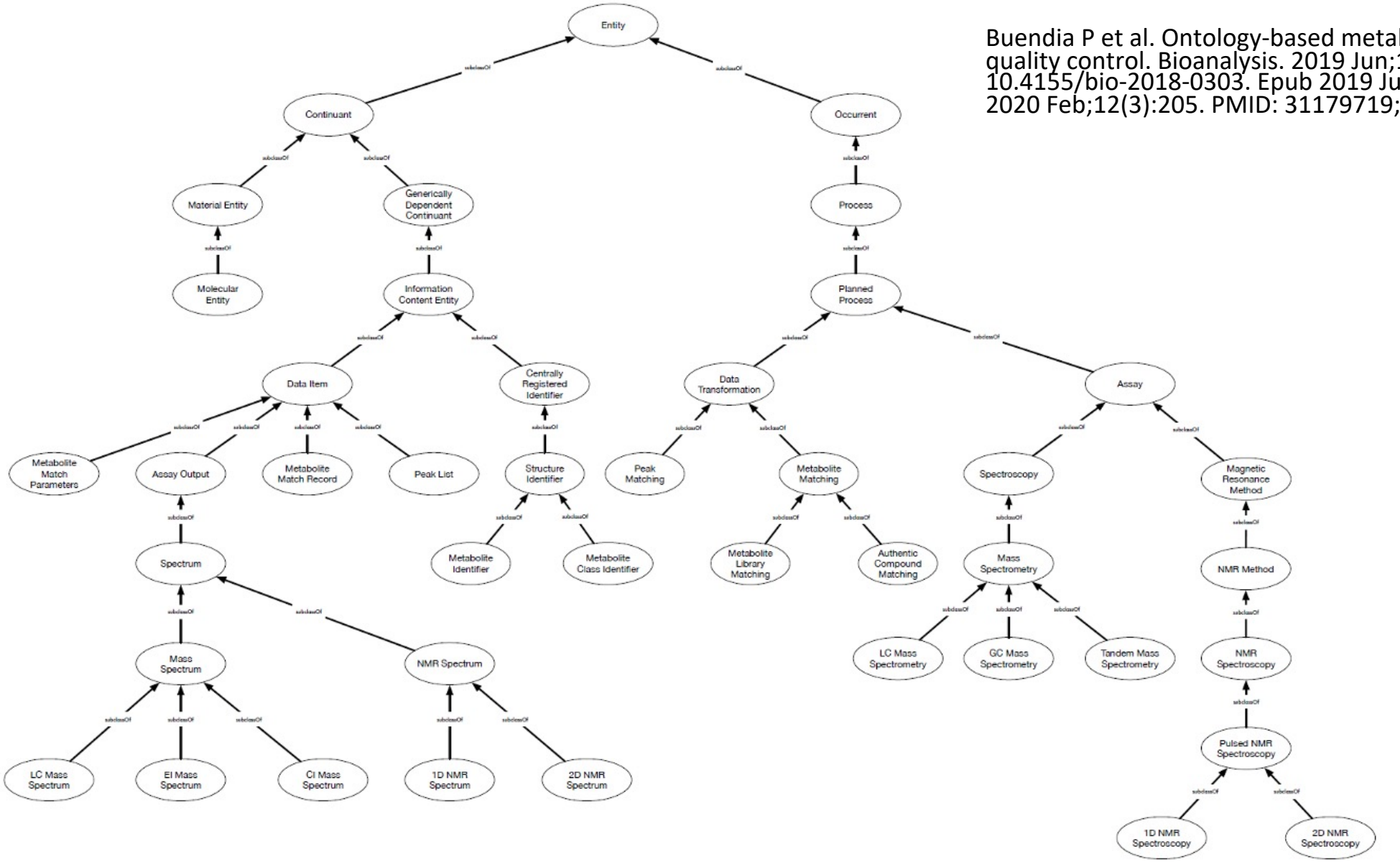
https://pfem.clermont.inrae.fr/ontologies-framework/

```
{ id: 1, name: "Species" },
{ id: 2, name: "Animal anatomie / Tissues " },
{ id: 3, name: "Plant Anatomie / Tissues" },
{ id: 4, name: "Environment" },
{ id: 5, name: "Biological Process" },
{ id: 6, name: "Vocabularies" },
{ id: 7, name: "Clinical / human health" },
{ id: 8, name: "Cancer" },
{ id: 9, name: "Phenotype" },
{ id: 10, name: "Computer Science / bioinformatics" },
{ id: 11, name: "Biological data" },
{ id: 12, name: "Omics data" },
{ id: 13, name: "Integration process" },
{ id: 14, name: "Pharmaceutical industry" },
{ id: 15, name: "Patient data" },
{ id: 16, name: "Food Safety" },
{ id: 17, name: "Time / Space" },
{ id: 18, name: "Data analysis" },
{ id: 19, name: "Experimental design / Experimental Conditions" },
{ id: 20, name: "Knowledge modeling" },
{ id: 21, name: "Biomedical Resources" },
{ id: 22, name: "Drugs" },
{ id: 23, name: "Parasites" },
{ id: 24, name: "protein ligand" },
{ id: 25, name: "Chemical & Molecules" },
{ id: 26, name: "Microbial anatomy" },
{ id: 27, name: "sample processing" },
{ id: 999, name: "other" }
```



```
{
        "id": 525,
        "name": "Systematized Nomenclature of Me
        "acronym": "SNMI",
        "types": [
                7,
                21
        ]
},
{
        "id": 526,
        "name": "Electrocardiography Ontology",
        "acronym": "ECG",
        "types": [
                -2
        ]
},
{
        "id": 527,
        "name": "Plant Ontology",
        "acronym": "PO",
        "types": [
                1,
                3
        ]
}
```

MODIS, a prototype implementation of a minimum reporting standard for collecting information about a metabolomics experiment (2019)

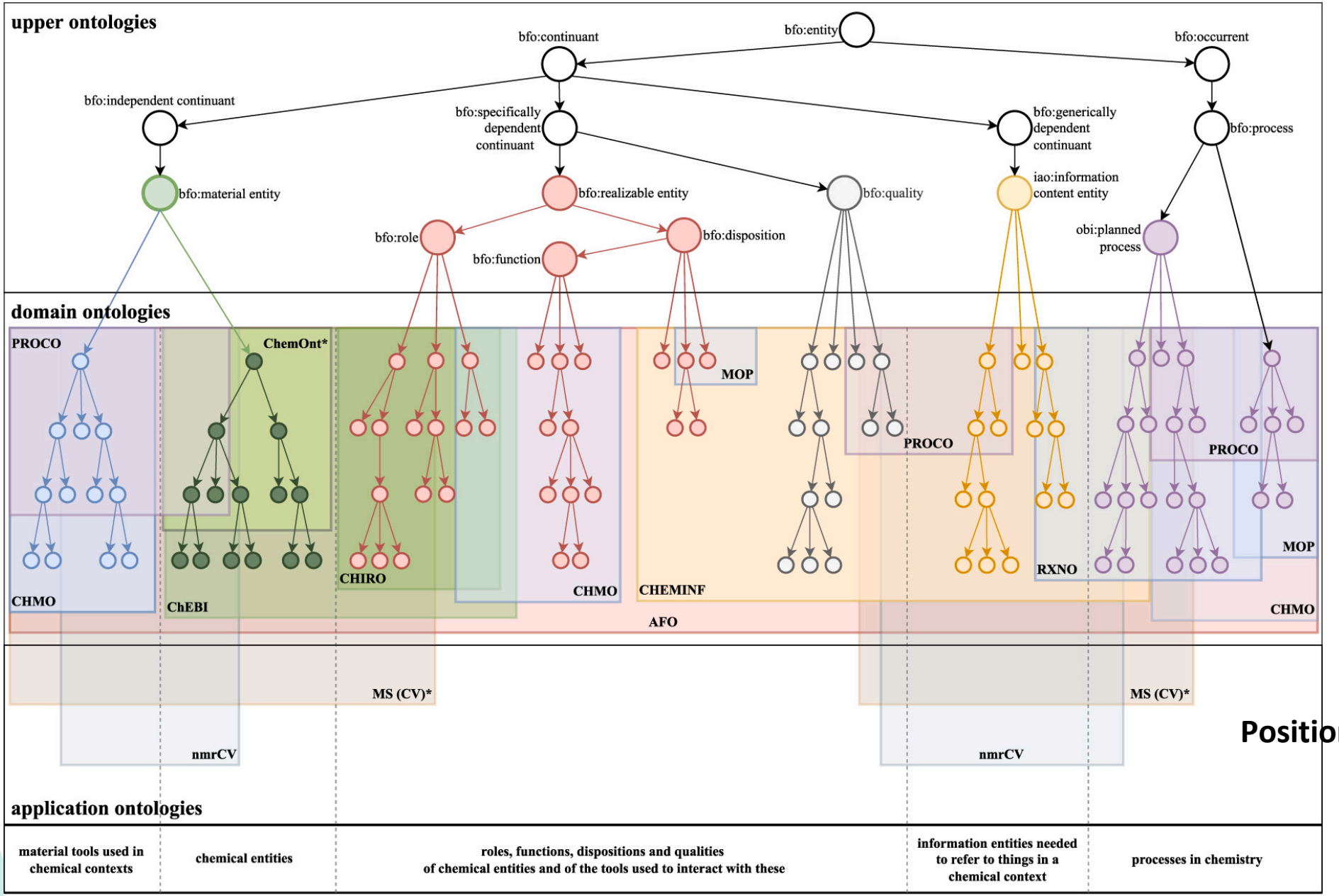


Buendia P et al. Ontology-based metabolomics data integration with quality control. Bioanalysis. 2019 Jun;11(12):1139-1155. doi: 10.4155/bio-2018-0303. Epub 2019 Jun 10. Erratum in: Bioanalysis. 2020 Feb;12(3):205. PMID: 31179719; PMCID: PMC6661928.

Landscape of ontologies in chemistry (Strömert et al, Pure and Applied Chemistry, 2022, 10.1515/pac-2021-2007)

Ontology	Domain	License	Modularity	Used in
ChEBI	Chemistry	CC-BY 4.0	BFO & OBO based	YMDB, HMDB, PubChem, MassBank, KNApSAcK, UM-BBD, GMD, SMID-DB
CHIRO	Chemistry	CC0 1.0	BFO & OBO based	Unknown
ChemOnt	Chemistry	Custom OA license	Subsumable under BFO's *Material entity*	YMDB, HMDB, T3DB, ECMDB, DrugBank, PubChem, ChEBI, LIPID MAPS, MoNA
CHEMINF	Chemistry	CC-BY 3.0	BFO & OBO based	PubChem, Open PHACTS
CHMO	Chemistry	CC-BY 4.0	BFO & OBO based	Chemotion, Allotrope™
MOP	Chemistry	CC-BY 4.0	BFO & OBO based	RXNO
RXNO	Chemistry	CC-BY 4.0	BFO & OBO based	NameRXN, Wikipedia, Chemotion
OntoKin	Chemistry	Unknown	OntoCAPE upper level & modules	J-Park Simulator
AFO	Chemistry	CC-BY 4.0	BFO classes & relations, many AFO- some custom OBO-modules	Allotrope™
PROCO	Chemistry	CC-BY 4.0	AFO & OBO based	Allotrope™
MS	Chemistry	CC-BY 4.0	BFO & OBO mapping possible	mzML
nmrCV	Chemistry	Public Domain Mark 1.0	BFO & OBO mapping possible	MetaboLights, HMDB
BFO	Upper level (classes only)	CC-BY 4.0	OBO backbone	¿300 ontologies & ¿50 organizations, PubChem
RO	Upper level (relations)	CC0 1.0	BFO & OBO based	Monarch Initiative, OBO Foundry, Gene Ontology, PubChem
IAO	Information artifacts	CC-BY 4.0	BFO & OBO based	OBO Foundry, Allotrope™, PubChem, ISA tools
OBI	Biomedicine	CC-BY 4.0	BFO & OBO based	OBO Foundry, Allotrope™, PubChem
UO	Scientific units	CC-BY 4.0	BFO & OBO based	OBO Foundry, UOM, PubChem
QUDT	Scientific units	CC-BY 4.0	BFO & OBO based mapping possible	Open PHACTS
PATO	Phenotypic & physical qualities	CC-BY 3.0	BFO & OBO based	OBO Foundry, Allotrope™
SIO	Upper level	CC-BY 4.0	BFO alignment	PubChem, Bio2RDF, SADI Semantic Web Services, DisGeNET's gene-disease associations, EBI's Gene Expression Atlas, Graph4Code
EDAM	Life-sciences & data management	CC-BY 4.0	BFO & OBO mapping possible	EMBOSS, Bio-jETI
OntoCAPE	Upper level & engineering	GNU GPLv2	Provides upper level concepts	J-Park Simulator

Landscape of ontologies in chemistry (Strömert et al, Pure and Applied Chemistry, 2022, 10.1515/pac-2021-2007)

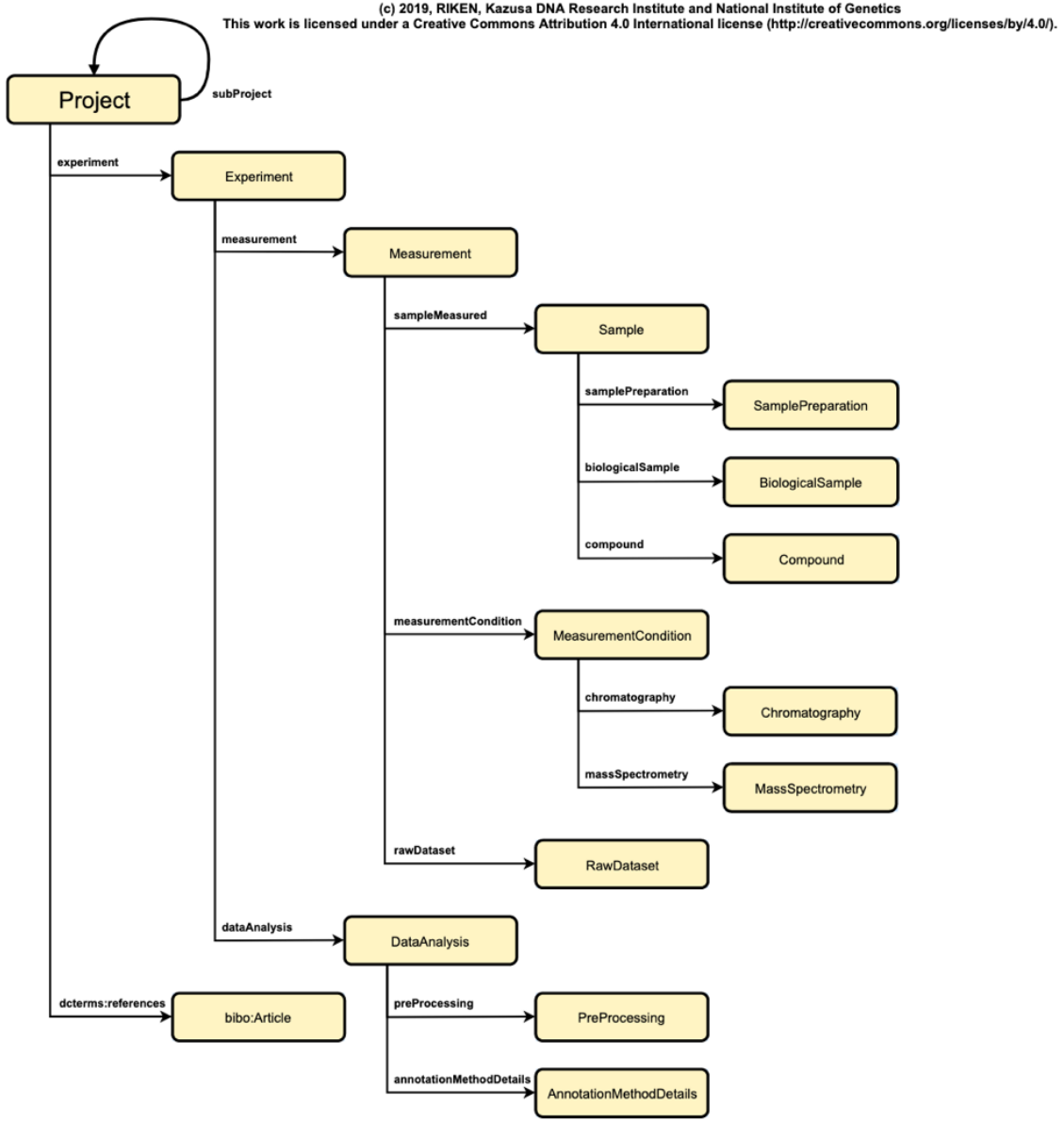


10 ontologies that cover general scientific domains and 12 chemistry domain-specific ontologies (see Table 1). Of these 12, 10 are domain ontologies and two are application ontologies.

Position in the OBO framework.

RIKEN Plant Metabolomics Metadata (RPMM)

(Fukushima A, et al. Development of RIKEN Plant Metabolome MetaDatabase, *Plant and Cell Physiology*, Volume 63, Issue 3, March 2022, Pages 433–440, https://doi.org/10.1093/pcp/pcab173)



https://github.com/afukushima/rpmm-metadata

master ▾ rpmm-metadata / ontology / core / RPMM_CORE.ttl

RIKEN-DKO typo修正

1 contributor

1721 lines (1252 sloc) | 64.7 KB

```
1  @prefix : <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/> .
2  @prefix owl: <http://www.w3.org/2002/07/owl#> .
3  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
4  @prefix xml: <http://www.w3.org/XML/1998/namespace/> .
5  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
6  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
7  @base <http://metadb.riken.jp/ontology/plantMetabolomics/> .
8  @prefix expType: <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/ExperimentType/> .
9  @prefix techType: <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/TechnologyType/> .
10 @prefix expFac: <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/ExperimentFactor/> .
11 @prefix meaType: <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/MeasurementType/> .
12 @prefix techPlat: <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/TechnologyPlatform/> .
13 @prefix massAys:  <http://metadb.riken.jp/ontology/plantMetabolomics/0.1/MassAnalyser/> .
14 @prefix obo: <http://purl.obolibrary.org/obo/> .
15 @prefix bibo: <http://purl.org/ontology/bibo/> .
16 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
17 @prefix dcterms: <http://purl.org/dc/terms/> .
18 @prefix qudt: <http://qudt.org/schema/qudt#> .
19 @prefix org: <http://www.w3.org/ns/org#> .
20 @prefix cc: <https://creativecommons.org/ns#> .
```

based on the semantic web technology

* Resources

- https://www.ebi.ac.uk/training/online/courses/metabolomics-introduction
- http://www.rsc.org/learn-chemistry/collections/spectroscopy/introduction#MassSpectrometry
- https://edu.rsc.org/resources/collections/analytical-chemistry-introductions#NMRSpectroscopy

- https://github.com/sneumann/SemanticMetabolomics/wiki/04-RDF-MassBank-Resource-Module

- Kenneth Haug, et al, MetaboLights: a resource evolving in response to the needs of its scientific community, *Nucleic Acids Research*, Volume 48, Issue D1, 08 January 2020, Pages D440–D444, https://doi.org/10.1093/nar/gkz1019
- Sud, M. et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **44**, D463–D470 (2016).