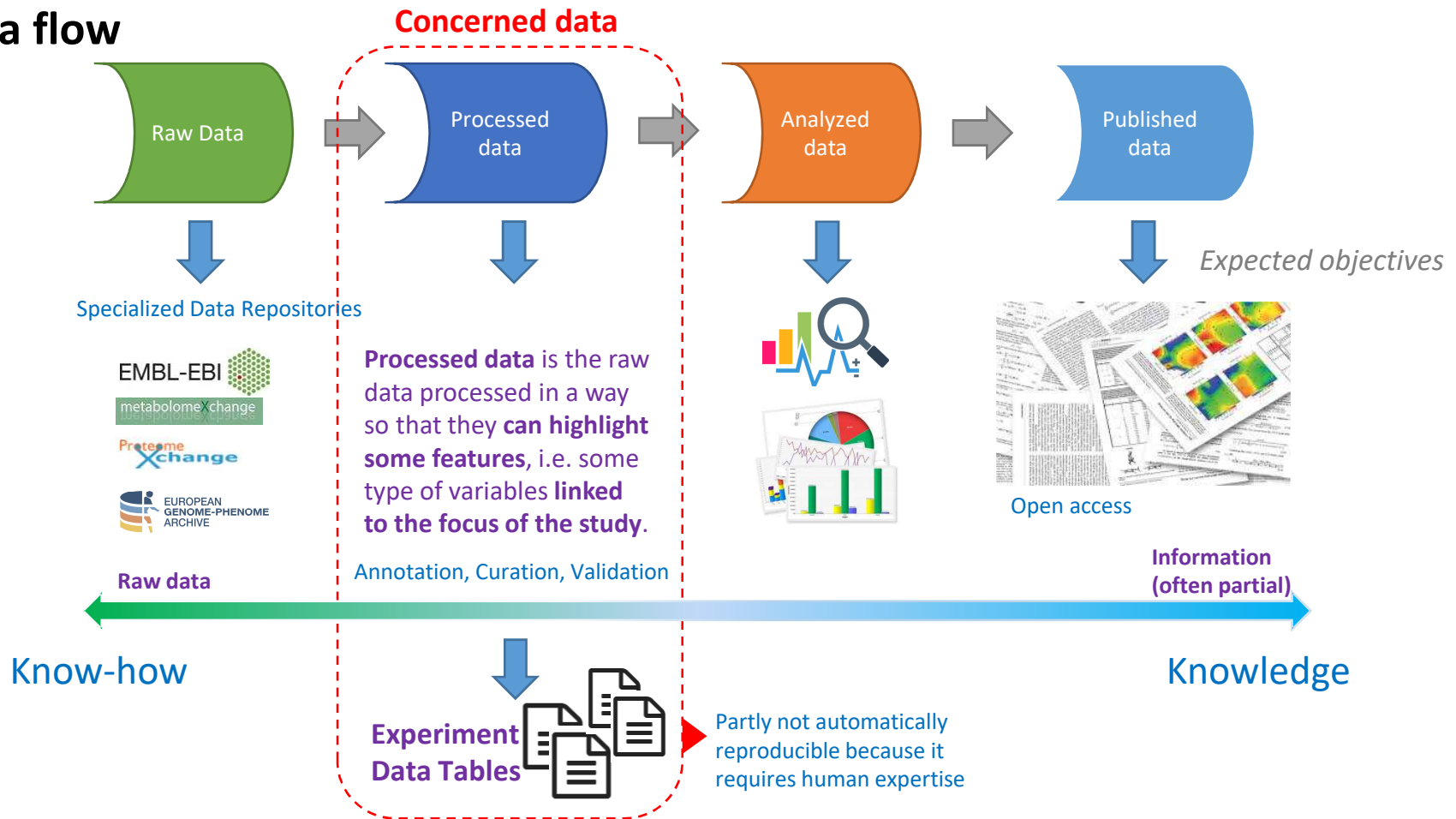
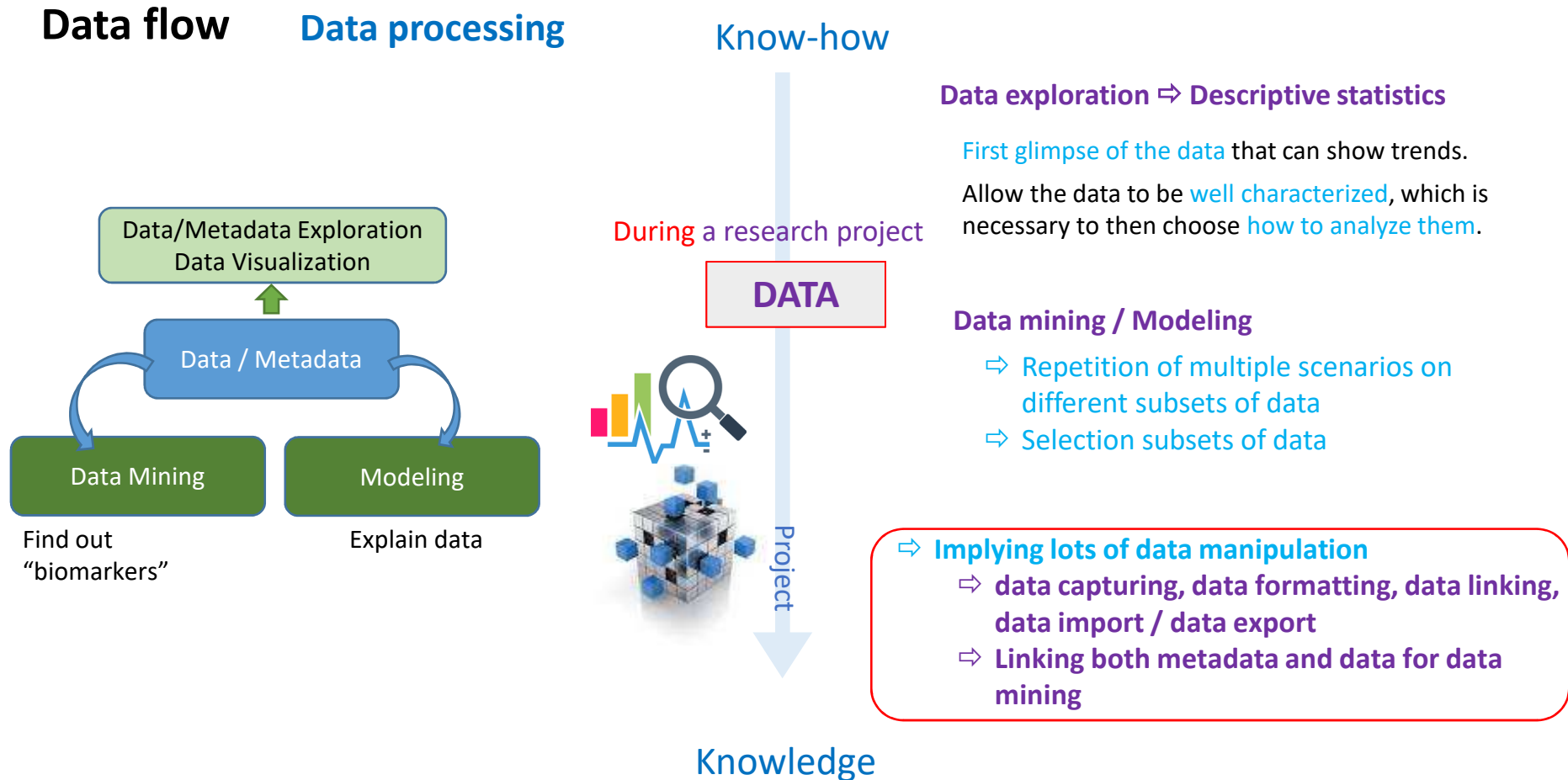


<sup>(\*)</sup> *Centre Automatisé de Traitement de l'Information* -  
Automated Information Processing Center

## Data flow



Studies with a design of experiment (DoE) involving one or more analytical approaches



## Data flow

## Data management



HORIZON 2020

## Publishing policies



A **data management plan** or **DMP** is a formal document that outlines **how data are to be handled** both **during a research project**, and **after the project is completed**.

The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis **before** the project begins

this ensures that data are well-managed in the **present**, and prepared for preservation in the **future**.

During a research project

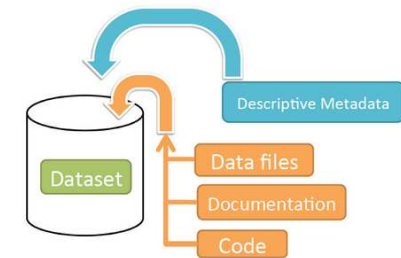
**DATA**



After the project is completed

The **Dataverse** Project

Schematic Diagram of a **Dataset** in Dataverse 4.0

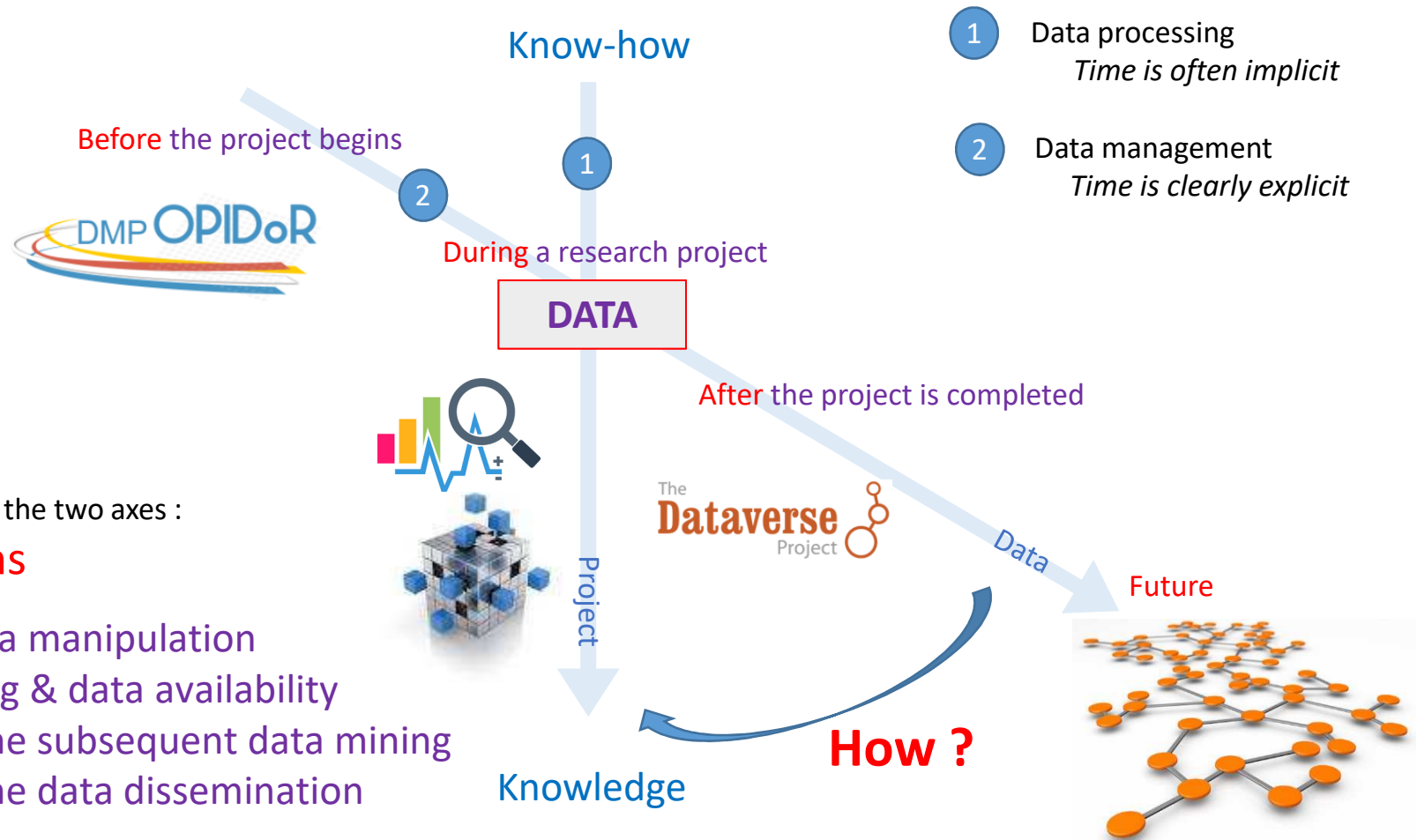


Container for your data, documentation, and code.

**Data** → **Future**

- ⇒ **Description by metadata (Ontology / Controlled Vocabulary)**
- ⇒ **Data capturing, data formatting, data linking**
- ⇒ **Implying data archeology after several months / years**

## Data flow



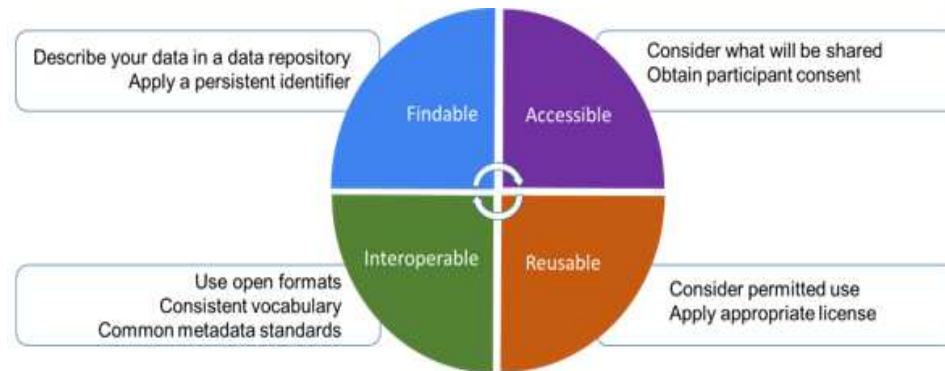
Make consistent the two axes :

### Motivations

- ⇒ Reduce data manipulation
- ⇒ Data sharing & data availability
- ⇒ Facilitate the subsequent data mining
- ⇒ Facilitate the data dissemination

The "data life" must therefore be integrated into the scientific research process

## THE FAIR DATA PRINCIPLES



The "FAIR" principles **define the basis** for data sharing easily to find, accessible, interoperable and reusable.

However, it is to the communities to specify the actions necessary for their implementation



The **implementation** of **FAIR principles** is a **process** that must be thought of in a **progressive and community-oriented way**.

It **must be integrated into existing practices** to ensure that they evolve without interruption and in a way that is **acceptable to the various actors**

## Data capture... as far upstream as possible

Take into account users operating methods and work habits

### Spreadsheet as a central tool

Despite all their drawbacks

e.g. multiple information in a format without internal structure

This does not take away any of their benefits

Universal tool

### ⇒ But: Repetitive and tedious tasks

#### Gathering Data and Preparing Data :

- lot of data manipulation, mainly in the form of table,
- combine data sets according to a common field (identifiers)

#### Modelisation :

- selection of subset of data then many repetitions of complex processing operations
- according to a very varied parametrization (scenarios).

Allow users to gain efficiency where they would like to gain efficiency

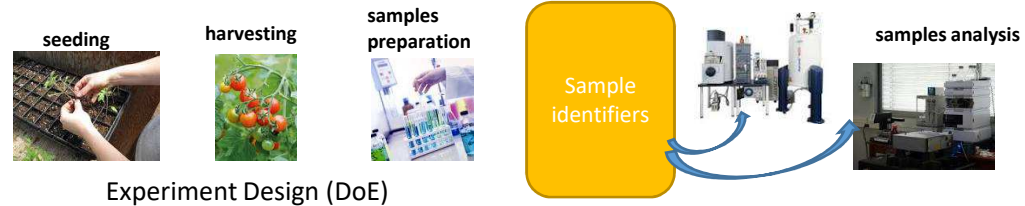
Handling of all these tasks related to **data management**

Promote  
good practices

Provide  
services

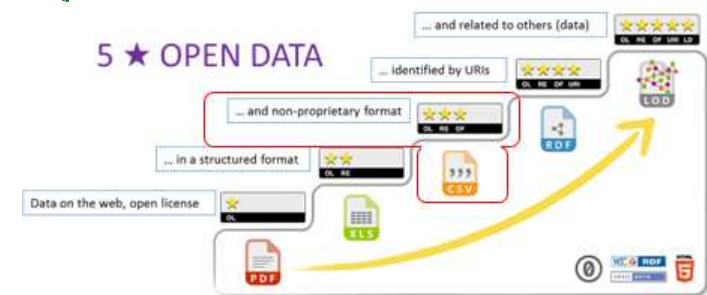


## Promote good practices



samples : Sample features

A	B	C	D	E	F	G	H	I	J
SampleID	Treatment	DevStage	FruitAge	FruitPosition	FruitDiamete	FruitHeight	FruitFW	Rank	Truss
115	Control	FF.01	07DPA	3	11.95	10.42	0.81	A	T7
121	Control	FF.03	22DPA	3	36.13	31.77	21.43	A	T6
164	Control	FR.01	42DPA	2	51.09	46.85	64.05	A	T5
353	Control	FR.04	55DPA	5	48.28	43.35	66.64	A	T5
355	Control	FR.04	55DPA	3	49.84	44.93	66.98	A	T5
413	Control	FR.02	47DPA	1	60.48	54.23	106.13	A	T7
512	Control	FF.03	21DPA	NA	41	35.82	37.22	A	TA
117	Control	FF.01	07DPA	3	13.44	12.39	1.14	A	T7
536	Control	FR.02	47DPA	NA	59.4	49.05	87.28	A	TA
544	Control	FR.03	50DPA	NA	57.31	47.69	92.86	A	TA
158	Control	FF.04	35DPA	5	58.38	49.3	92.86	A	T5
109	Control	FF.03	22DPA	7	43.37	35.77	38.73	A	T5
134	Control	FF.02	15DPA	3	27.89	23.8	9.88	A	T7
31	Control	FF.01	08DPA	4	NA	NA	0.48	A	T6
179	Control	FF.03	28DPA	3	53.68	45.43	65.34	A	T7
383	Control	FF.04	34DPA	5	47.04	41.19	48.96	A	T7
425	Control	FR.04	55DPA	2	62.74	50.27	115.3	A	T7
520	Control	FF.03	30DPA	NA	48.86	41.52	52.94	A	TA
419	Control	FR.03	50DPA	2	55.63	48.02	86.79	A	T7
138	Control	FF.02	15DPA	6	27.96	22.14	9.69	A	T7
143	Control	FF.03	29DPA	4	48.45	42.92	51.35	A	T6
365	Control	FR.02	47DPA	5	55.11	44.9	71.82	A	T6
127	Control	FF.03	27DPA	3	45.71	43.28	47.8	A	T5
188	Control	FR.01	42DPA	3	55.38	47.1	77.39	A	T6



necessary and indispensable step towards « **Linked Open Data** ».

Promote non-proprietary format like CSV or TSV

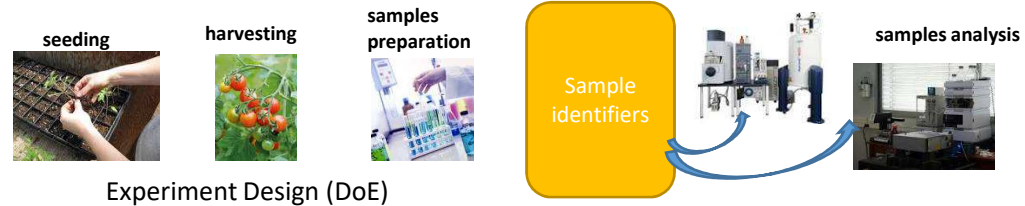


## Data

Well organized data

- Each variable forms a column
- Each observation forms a line
- Each type of "unit observational" forms a table

## Promote good practices



samples : Sample features

A	B	C	D	E	F	G	H	I	J
SampleID	Treatment	DevStage	FruitAge	FruitPosition	FruitDiamete	FruitHeight	FruitFW	Rank	Truss
115	Control	FF.01	07DPA	3	11.95	10.42	0.81	A	T7
121	Control	FF.03	22DPA	3	36.13	31.77	21.43	A	T6
164	Control	FR.01	42DPA	2	51.09	46.85	64.05	A	T5
373	Control	FR.04	55DPA	5	48.28	43.35	66.64	A	T5
373	Control	FR.04	55DPA	3	49.84	44.93	66.98	A	T5
373	Control	FR.02	47DPA	1	60.48	54.23	106.13	A	T7
512	Control	FF.03	21DPA	NA	41	35.82	37.22	A	TA
117	Control	FF.03	21DPA	3	13.44	12.39	1.14	A	T7
536	Control	FF.03	21DPA	NA	59.4	49.05	87.28	A	TA
544	Control	FF.03	21DPA	NA	57.31	47.69	92.86	A	TA
158	Control	FF.04	35DPA				92.86	A	T5
109	Control	FF.03	22DPA				38.73	A	T5
134	Control	FF.02	15DPA	3	27.05	23.0	9.88	A	T7
31	Control	FF.01	08DPA	4	NA	NA	0.48		
179	Control	FF.03	28DPA	3	53.68	45.43	65.34	A	
383	Control	FF.04	34DPA	5	47.04	41.19	48.96	A	
425	Control	FR.04	55DPA	2	62.74	50.27	115.3	A	T7
520	Control	FF.03	30DPA	NA	48.86	41.52	52.94	A	TA
419	Control	FR.03	50DPA	2	55.63	48.02	86.79	A	T7
138	Control	FF.02	15DPA	6	27.96	22.14	9.69	A	T7
143	Control	FF.03	29DPA	4	48.45	42.92	51.35	A	T6
365	Control	FR.02	47DPA	5	55.11	44.9	71.82	A	T6
127	Control	FF.03	27DPA	3	45.71	43.28	47.8	A	T5
188	Control	FR.01	42DPA	3	55.38	47.1	77.39	A	T6

identifier

factors

Quantitative

Qualitative



Whatever the kind of experiment, this assumes a design of experiment (DoE) involving individuals, samples or whatever things, as the main objects of study (e.g. plants, tissues, bacteria, ...)

This also assumes the observation of dependent variables resulting of effects of some controlled experiment factors.

Moreover, the objects of study have usually an identifier for each of them, and the variables can be quantitative or qualitative.

## Data

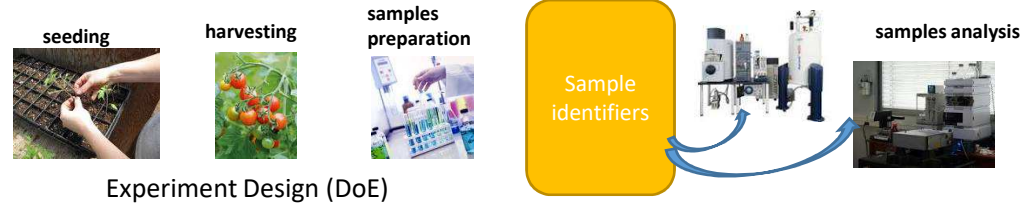
Well organized data

- Each variable forms a column
- Each observation forms a line
- Each type of "unit observational" forms a table

Promote non-proprietary format like CSV or TSV



## Promote good practices



samples : Sample features

A	B	C	D	E	F	G	H	I	J
SampleID	Treatment	DevStage	FruitAge	FruitPosition	FruitDiameter	FruitHeight	FruitFW	Rank	Truss
115	Control	FF.01	07DPA	3	11.95	10.42	0.81	A	T7
121	Control	FF.03	22DPA	3	36.13	31.77	21.43	A	T6
164	Control	FR.01	42DPA	2	51.09	46.85	64.05	A	T5
323	Control	FR.04	55DPA	5	48.28	43.35	66.64	A	T5
512	Control	FF.03	21DPA	NA	3	49.84	44.93	A	T5
117	Control	FF.03	21DPA	NA	3	49.84	44.93	A	T5
536	Control	FF.03	21DPA	NA	3	49.84	44.93	A	T5
544	Control	FF.03	21DPA	NA	3	49.84	44.93	A	T5
158	Control	FF.04	35DPA	NA	3	49.84	44.93	A	T5
109	Control	FF.03	22DPA	NA	3	49.84	44.93	A	T5
134	Control	FF.02	15DPA	NA	3	49.84	44.93	A	T5
31	Control	FF.01	08DPA	4	NA	27.96	22.14	9.69	A
179	Control	FF.03	28DPA	3	NA	48.45	42.92	51.35	A
383	Control	FF.04	34DPA	5	NA	55.11	44.9	71.82	A
425	Control	FR.04	55DPA	2	NA	45.71	43.28	47.8	A
520	Control	FF.03	30DPA	NA	3	55.38	47.1	77.35	A
419	Control	FR.03	50DPA	2	NA	55.38	47.1	77.35	A
138	Control	FF.02	15DPA	6	NA	27.96	22.14	9.69	A
143	Control	FF.03	29DPA	4	NA	48.45	42.92	51.35	A
365	Control	FR.02	47DPA	5	NA	55.11	44.9	71.82	A
127	Control	FF.03	27DPA	3	NA	45.71	43.28	47.8	A
188	Control	FR.01	42DPA	3	NA	55.38	47.1	77.35	A



Description of the different columns within data files

Shortname	Description	Unit	Category
SampleID	Pool of several harvests		Identifier
Treatment	Treatment applied on plants		Factor
DevStage	fruit development stage		Factor
FruitAge	fruit age	Days post-anthesis (dpa)	Factor
FruitDiameter	Fruit diameter	mm	Variable
FruitHeight	Fruit height	mm	Variable
FruitFW	Fruit Fresh Weight(g)	g	Variable
Rank	Row of the individual plant on the table		Feature
Truss	Position on the stem of the truss		Feature

### categories

- identifier
- factor
- qualitative
- quantitative

Metadata

Data

⇒ **Metadata** : not just on the "top" linked to datasets but more deeply linked to the variables.

Minimal effort (PUT)

Promote non-proprietary format like CSV or TSV





# FAIR principles, an new opportunity to improve the data lifecycle

## Promote good practices

### Minimal effort

ODAM Framework

<http://pmb-bordeaux.fr/dataexplorer/>

Experiment  
Data Tables  
+ Metadata

Multiscale deployment  
Local / Intranet / Internet

Storage and  
sharing with all  
partners

F  
A  
I  
N  
T  
E  
R  
O  
P  
E  
R  
A  
B  
L  
E

plants.tsv  
harvests.tsv  
samples.tsv  
compounds.tsv  
enzymes.tsv

Drag & Drop

PUT



VM / docker

GET

Additional data subsets can be  
added step by step, as soon as  
data are produced.

Multi-criteria exploration &  
interactive visualization



Data can be  
downloaded,  
explored and mined

Merging & selection  
of data subsets



Export

Repetition of complex  
treatments according to very  
varied parameters



python

CRAN 0.1.4

API

The  
Dataverse  
Project  
<https://data.inra.fr>

Link

facilitate the  
subsequent data  
publication

F  
I  
N  
D  
A  
B  
L  
E  
A  
C  
C  
E  
S  
S  
I  
B  
L  
E  
I  
N  
T  
E  
R  
O  
P  
E  
R  
A  
B  
L  
E  
R  
E  
-  
U  
S  
A  
B  
L  
E

Provide services  
offered to Data  
Producers/  
Consumers

Maximum efficiency

Develop if needed, lightweight tools

- R/Python scripts (Galaxy), lightweight GUI (R shiny)
- Python (Plotly Dash)



## Data Lifecycle

ODAM Framework

<http://pmb-bordeaux.fr/dataexplorer/>

### Data capture

Experiment  
Data Tables  
+ Metadata

Additional data subsets can be added step by step, as soon as data are produced.

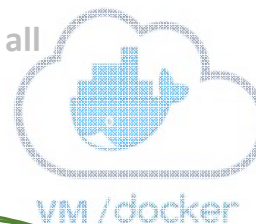
### Data Backup

Storage and sharing with all partners

Multiscale deployment  
Local / Intranet / Internet

Drag & Drop

PUT

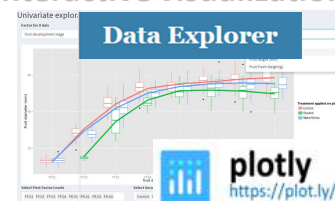


GET

### Data Curation

Remove inconsistent values

Multi-criteria exploration & interactive visualization



Data can be downloaded, explored and mined

### Data Extension

Merging & selection of data subsets



Export

Repetition of complex treatments according to very varied parameters



python

CRAN 0.1.4

API

Link

facilitate the subsequent data publication

The Dataverse Project

FINDABLE  
ACCESSIBLE  
INTEROPERABLE  
RE-USABLE

### Data Publication

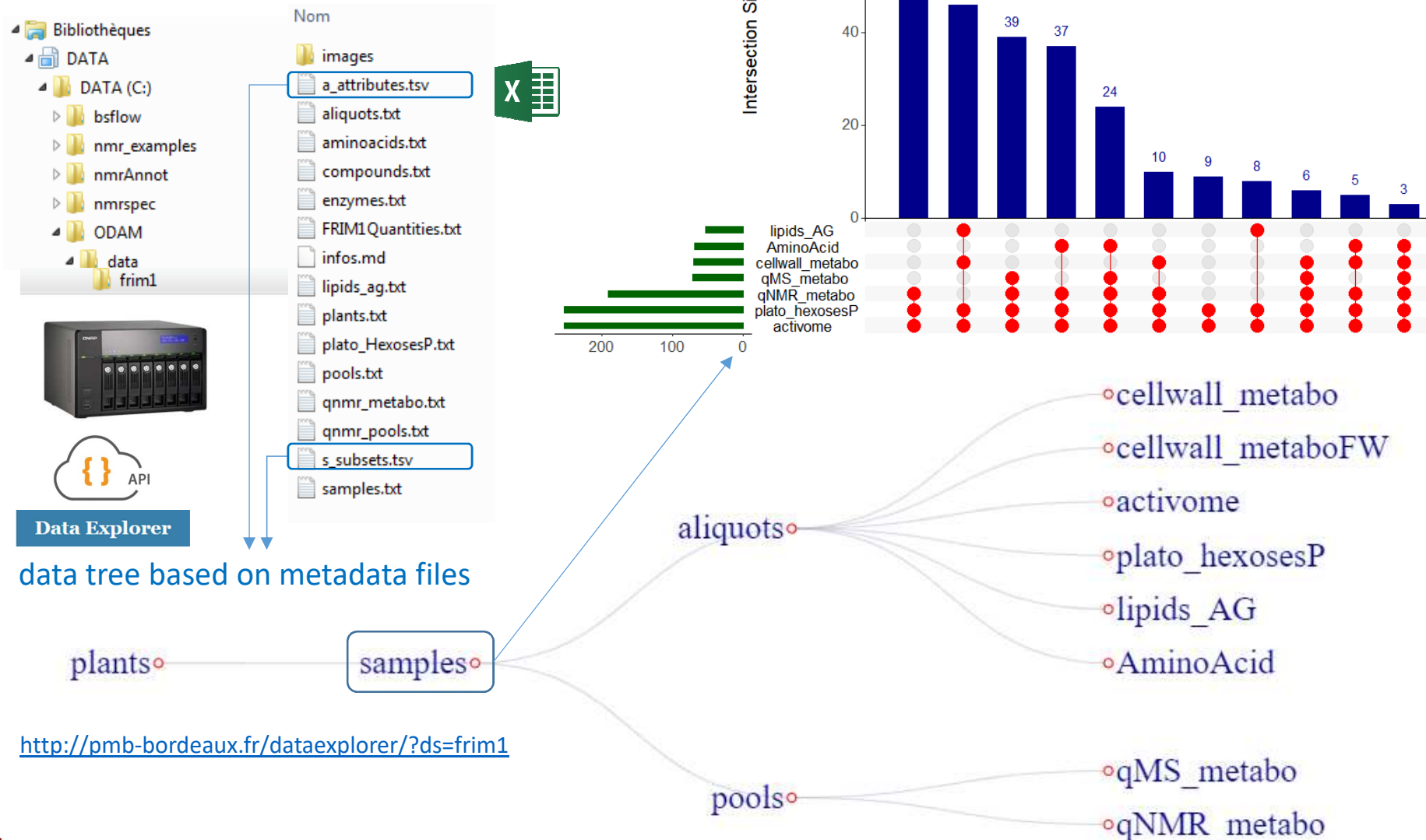
<https://data.inra.fr>

The "data lifecycle" is thus integrated into the scientific research process

Develop if needed, lightweight tools

- R/Python scripts (Galaxy), lightweight GUI (R shiny)
- Python (Plotly Dash)

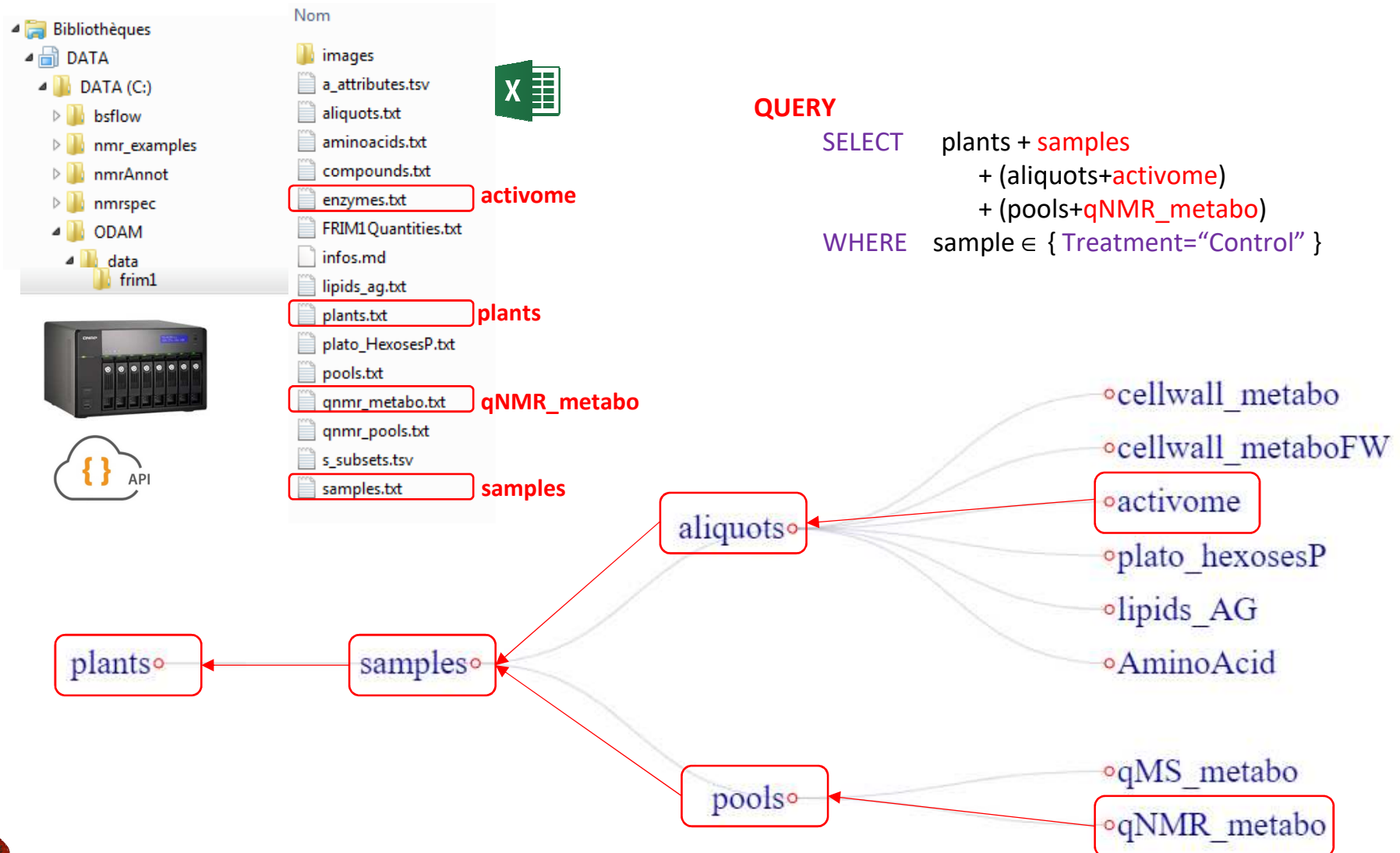
## Merging & selection of data subsets



Example based on FRIM - Fruit Integrative Modelling

## Merging & selection of data subsets

Merge the **activome data** **AND** the **metabolome data** acquire on the same **samples** belonging to the "Control" group




Example based on FRIM - Fruit Integrative Modelling



## Merging & selection of data subsets



Merge the **activome** data **AND** the **metabolome** data acquire on the same **samples** belonging to the "Control" group



API

[http://myhost.org/getdata/xml/frim1/\(activome,qNMR\\_metabo\)/treatment/Control](http://myhost.org/getdata/xml/frim1/(activome,qNMR_metabo)/treatment/Control)

**Data Explorer**

PlantID	Rank	PlantNum	Treatment	SampleID	Truss	DevStage	FruitAge	AliquotID	PGM	F16BP	Cyt	PyRk	CitS	PPI	AcoS	PoolID	glucose	saccharose	fructose	galactose
A1	A	1	Control	115	T7	FF.01	07DPA	397	4924.26	17.89	1327.67	50.79	2225.01	657.4		F1	409.3950483	70.36199274	372.9002229	0
A1	A	1	Control	121	T6	FF.03	22DPA	409	3123.06	8.67	450.85	23.53	1194.12	288.49		F9	725.0013475	33.0778677	871.7036417	0
A1	A	1	Control	164	T5	FR.01	42DPA	493	1597.7	5.16	237.69	34.63	578.74	84.92		F17	723.1068785	27.73666946	970.4150597	0
A1	A	1	Control	353	T5	FR.04	55DPA	1019	779.85	17.04	373.57	NA	395.51	153.66		F23	801.7624659	27.50853514	1103.460076	4.104895086
A2	A	2	Control	355	T5	FR.04	55DPA	1023	635.13	9.08	300.12	NA	338.39	100.22		F23	801.7624659	27.50853514	1103.460076	4.104895086

**plants**

PlantID	Rank	PlantNum	Treatment	SampleID	Truss	DevStage	FruitAge
A5	A	5	Control	134	T7	FF.02	15DPA
A6	A	6	Control	31	T6	FF.01	08DPA
A6	A	6	Control	179	T7	FF.03	28DPA
A6	A	6	Control	383	T7	FF.04	34DPA
A6	A	6	Control	425	T7	FR.04	55DPA
A7	A	7	Control	419	T7	FR.03	50DPA
A8	A	8	Control	138	T7	FF.02	15DPA
A8	A	8	Control	143	T6	FF.03	29DPA
A8	A	8	Control	365	T6	FR.02	47DPA
A9	A	9	Control	127	T5	FF.03	27DPA
A9	A	9	Control	188	T6	FR.01	42DPA
A10	A	10	Control	33	T6	FF.01	08DPA
A14	A	14	Control	99	T5	FF.02	15DPA
A14	A	14	Control	103	T6	FF.02	15DPA
A14	A	14	Control	119	T7	FF.01	07DPA
A14	A	14	Control	119	T7	FF.01	07DPA
A14	A	14	Control	181	T7	FF.03	28DPA
A14	A	14	Control	367	T6	FR.02	47DPA
A15	A	15	Control	97	T5	FF.02	15DPA
A15	A	15	Control	123	T6	FF.03	22DPA
A15	A	15	Control	377	T6	FR.04	55DPA
A15	A	15	Control	423	T7	FR.03	50DPA
A17	A	17	Control	136	T7	FF.02	15DPA

**samples**

SampleID	Truss	DevStage	FruitAge
115	T7	FF.01	07DPA
121	T6	FF.03	22DPA
164	T5	FR.01	42DPA
353	T5	FR.04	55DPA
355	T5	FR.04	55DPA
419	T7	FR.03	50DPA
138	T7	FF.02	15DPA
143	T6	FF.03	29DPA
365	T6	FR.02	47DPA
127	T5	FF.03	27DPA
188	T6	FR.01	42DPA
33	T6	FF.01	08DPA
99	T5	FF.02	15DPA
103	T6	FF.02	15DPA
119	T7	FF.01	07DPA
119	T7	FF.01	07DPA
181	T7	FF.03	28DPA
367	T6	FR.02	47DPA
97	T5	FF.02	15DPA
123	T6	FF.03	22DPA
377	T6	FR.04	55DPA
423	T7	FR.03	50DPA
136	T7	FF.02	15DPA

**activome**

AliquotID	PGM	F16BP	Cyt	PyRk	CitS	PPI	AcoS
397	4924.26	17.89	1327.67	50.79	2225.01	657.4	
409	3123.06	8.67	450.85	23.53	1194.12	288.49	
493	1597.7	5.16	237.69	34.63	578.74	84.92	
1019	779.85	17.04	373.57	NA	395.51	153.66	
1023	635.13	9.08	300.12	NA	338.39	100.22	
113	3997.96	0.7	3053.11	44.2	2330.09	719.39	
365	3836.19	4.16	603.74	31.55	1836.44	330.26	
373	4220.68	31.7	592.57	14.81	1654.27	602.62	
405	3840.64	15.02	1032.17	37.88	1967.72	672.92	
527	1446.93	10.1	404.03	NA	690.03	207.48	
1047	792.75	11.77	247.22	40.65	364.72	165.21	
361	3967.66	12.54	786.54	88.85	1828.44	363.23	
413	2745.15	17.99	382.58	111.43	1027.11	258.78	
1067	515.91	9.28	341.31	101.62	312.49	100.78	
1158	622.88	8.8	123.63	NA	320.27	71.84	
437	3348.39	22.26	444.55	79.2	1206.14	290.49	

**qNMR\_metabo**

PoolID	glucose	saccharose	fructose	galactose
F1	409.3950483	70.36199274	372.9002229	0
F9	725.0013475	33.0778677	871.7036417	0
F17	723.1068785	27.73666946	970.4150597	0
F23	801.7624659	27.50853514	1103.460076	4.104895086
F23	801.7624659	27.50853514	1103.460076	4.104895086
F6	619.3250882	45.83076382	651.0915378	0
F3	356.3376494	48.89175215	353.455502	0
F11	966.8603372	27.47090189	1062.996002	0
F14	873.6366279	22.27264949	1029.943239	0
F25	1128.526505	29.07120456	1339.54916	5.643624441
F22	1050.48663	35.66103031	1243.74997	3.503144469
F6	619.3250882	45.83076382	651.0915378	0
F12	679.2961808	23.94077746	857.8519232	0
F19	756.6885092	33.90751466	975.4853658	1.544911295
F10	623.2729217	27.38271169	790.6343093	0
F18	961.5393478	23.54937901	1125.360688	0
F3	356.3376494	48.89175215	353.455502	0
F4	555.1791836	41.06615447	659.640831	0
F5	628.0378114	38.84344563	701.485209	0
F1	409.3950483	70.36199274	372.9002229	0
F26	363.3532859	64.43921076	343.535269	0
F11	966.8603372	27.47090189	1062.996002	0
F19	756.6885092	33.90751466	975.4853658	1.544911295
F4	555.1791836	41.06615447	659.640831	0
F9	725.0013475	33.0778677	871.7036417	0
F24	889.732794	37.9595702	1144.084393	3.498960704
F22	1050.48663	35.66103031	1243.74997	3.503144469
F6	619.3250882	45.83076382	651.0915378	0

⇒ Avoids lots of data manipulation

(activome,qNMR\_metabo) ⇔ plants + samples  
 + (aliquots+activome)  
 + (pools+qNMR\_metabo)

As far as possible, keep the old way of using the scientist's worksheets ...

Example based on FRIM - Fruit Integrative Modelling

Daniel Jacob, INRA – April 2019



## Retrieving Data within R

CRAN 0.1.4

data and metadata are all available and accessible by scripting languages(R, API)

Merge the **activome** data AND the **metabolome** data acquire on the same samples

```
### Load the R ODAM package
library(Rodam)
```

```
### Get the merged data of two data subsets based on their common identifiers
dh <- new('odamws', wsURL='https://pmb-bordeaux.fr/getdata/', dsname='friml')
setNameList <- c("activome", "qNMR_metabo")
dsMerged <- dh$getSubsetByName(setNameList)
```



```
> dim(dsMerged$data)
[1] 948 83
```

```
> dsMerged$LABELS
```

	Subset	Attribute	Description	Type
1	samples	SampleID	Pool of several harvests	Identifier
2	plants	Treatment	Treatment applied on plants	Factor
3	samples	DevStage	fruit development stage	Factor
4	samples	FruitAge	fruit age (dpa)	Factor
5	activome	PGM	Phosphoglucumutase	Variable
6	activome	F16BP_Cyt	Fructose-1,6-bisphosphatase (cyt)	Variable
7	activome	PyrK	Pyruvate kinase	Variable
8	activome	CitS	Citrate synthase	Variable
9	activome	PPI	Phosphofructokinase (PPi)	Variable
10	activome	AcoS	Aconitase	Variable
...				
43	qNMR_metabo	glucose	glucose (nMol / mg MS)	Variable
44	qNMR_metabo	saccharose	saccharose (nMol / mg MS)	Variable
45	qNMR_metabo	fructose	fructose (nMol / mg MS)	Variable
46	qNMR_metabo	galactose	galactose (nMol / mg MS)	Variable
47	qNMR_metabo	mannose	mannose (nMol / mg MS)	Variable
48	qNMR_metabo	rhamnose	rhamnose (nMol / mg MS)	Variable
49	qNMR_metabo	acetate	acetate (nMol / mg MS)	Variable
50	qNMR_metabo	chlorogenate	chlorogenate (nMol / mg MS)	Variable
...				

data

metadata

... while allowing a way to be more efficient ...



Example based on FRIM - Fruit Integrative Modelling



## Retrieving Data within R

data and metadata are all available and accessible by scripting languages(R, API)

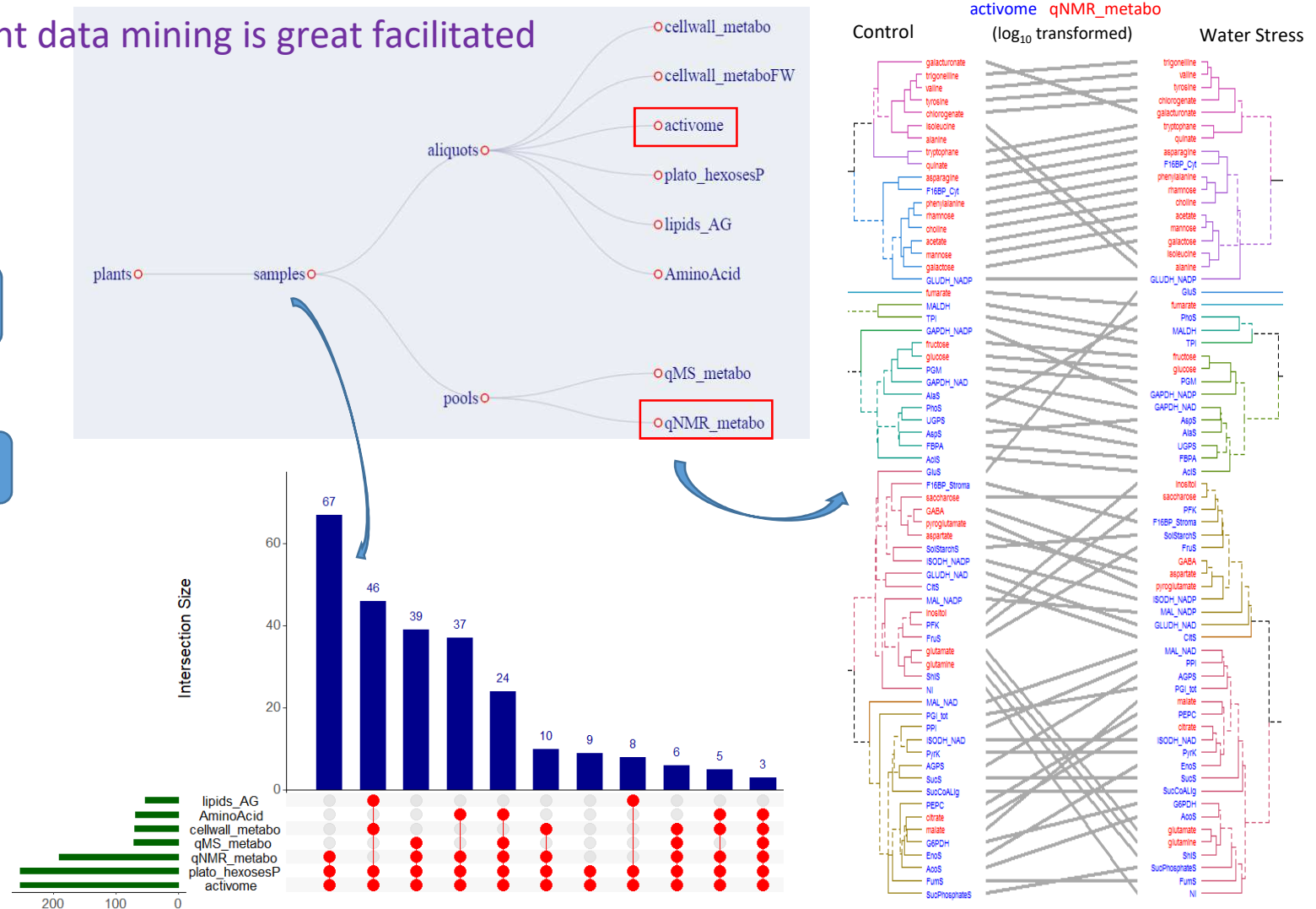
The subsequent data mining is great facilitated

Make both  
metadata and data  
available for  
data mining.

Experimentation  
/ Analysis

Data / Metadata

Data Mining



## Opportunity for Reproducible Research

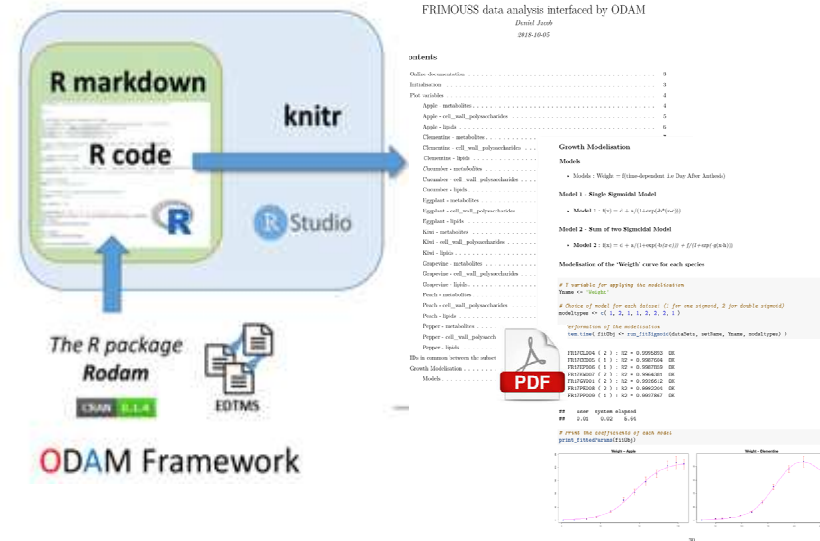
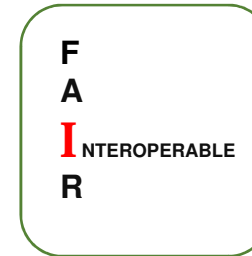
Reproduce the results presented in an article.

Fostered by the fact that data and metadata are all available and accessible by scripting languages(R, API)

Gathering Data and Preparing Data carried out by the ODAM approach

Merging & selection of data subsets

Repetition of complex treatments according to very varied parameters



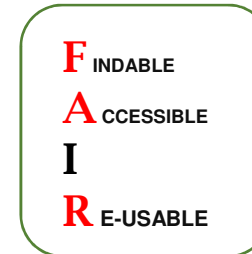
## Data Analysis Reporting

[rescience.github.io](https://rescience.github.io)

## Opportunity for Open Data

### Data dissemination

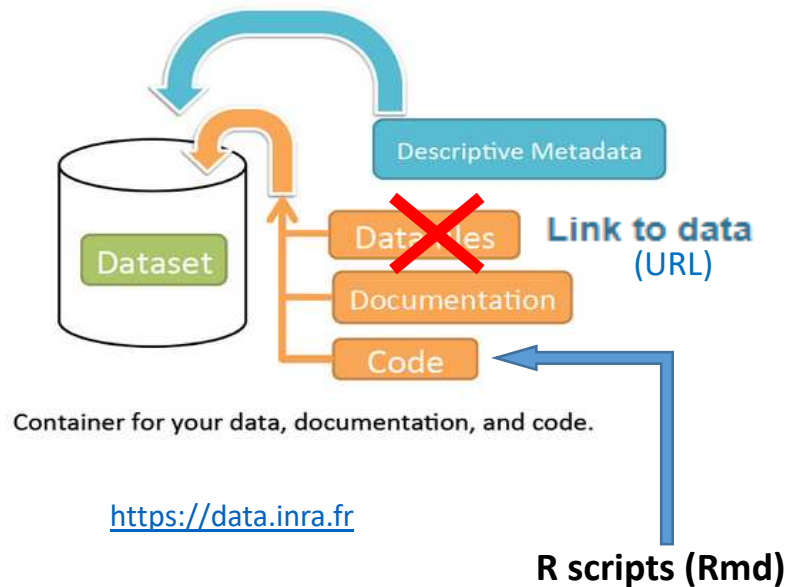
Disseminate their data more quickly  
at any time they wish  
without any additional effort



Schematic Diagram of a **Dataset** in Dataverse 4.0



### Data Publication



## FAIR principles, an new opportunity to improve the data lifecycle



### FRIM - Fruit Integrative Modelling Version 1.1

Bénard, Camille; Bernillon, Stéphane; Blais, Benoit; Maucourt, Mickael; Ballias, Patricia; Deborde, Catherine; Colombié, Sophie; Cabasson, Cécile; Jacob, Daniel; Gibon, Yves; Moing, Annick, 2018, "FRIM - Fruit Integrative Modelling", <https://doi.org/10.15454/95JUTK>, Portail Data Inra, V1

Cite Dataset

Learn about Data Citation Standards.



#### Description

The project aimed to build a virtual tomato fruit that enables the prediction of metabolite levels given genetic and environmental inputs, by an iterative process between laboratories which combine expertise in fruit biology, ecophysiology, theoretical and experimental biochemistry, and biotechnology.

#### Subject

Computer science; Information management; Omics; Plant Health and Pathology

#### Related Publication

Bénard C., Bernillon S., Blais B., Osorio S., Maucourt M., Ballias P., Deborde C., Colombié S., Cabasson C., Jacob D., Vercambre G., Gautier H., Rolin D., Génard M., Fernie A., Gibon Y., Moing A. 2015 Metabolomic profiling in tomato reveals diel compositional changes in fruit affected by source-sink relationships. Journal of Experimental Botany Vol. 66, No. 11 pp. 3391–3404 doi: [10.1093/jxb/erv151](https://doi.org/10.1093/jxb/erv151)

**Link to data (URL)** <http://pmb-bordeaux.fr/dataexplorer/?ds=frim1>

doi:10.15454/95JUTK



ODAM - Data Explorer

frim1

-- Select a Data Subset --

Subset Information

Data Table

Univariate

Bivariate

Multivariate

About

Data Information

Tomato

Culture location

- INVENIO, Ste Livrade, France

ERASysBio++ FRIM Project

Fruit Integrative Modelling

The project aimed to build a virtual tomato fruit that enables the prediction of metabolite levels given genetic and environmental inputs, by an iterative process between laboratories which combine expertise in fruit biology, ecophysiology, theoretical and experimental biochemistry, and biotechnology.

INRA  
SCIENCE & IMPACT

© INRA UMR 1332 BFP - Metabolism Team - Yves Gibon - 2017



**Example based on FRIM - Fruit Integrative Modelling**



## Opportunity for the Web Of Data



Resource Description Framework (RDF)



## Metadata

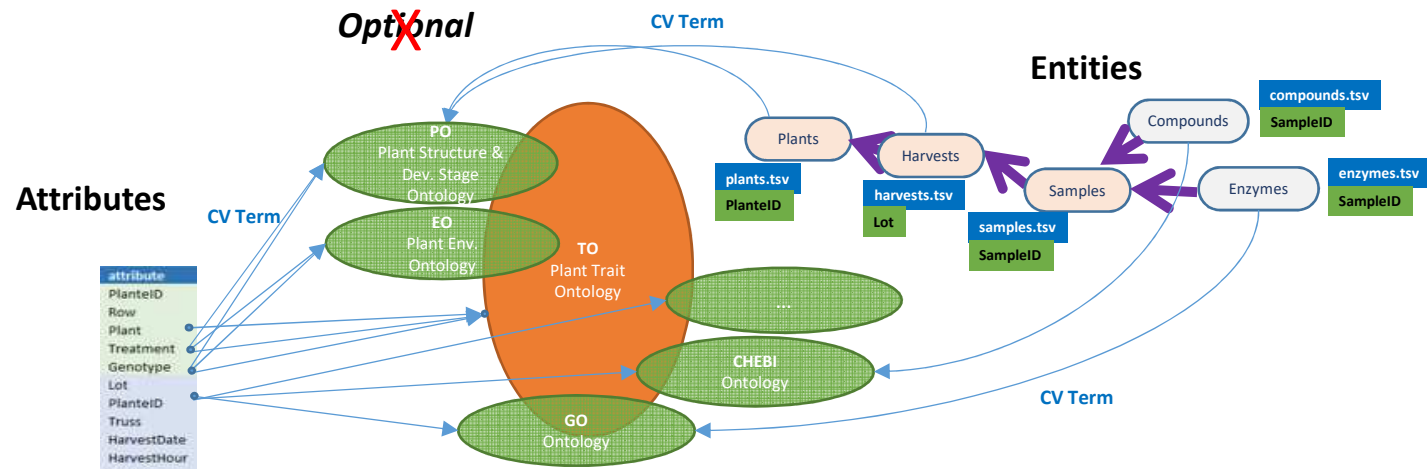
Attributes						CV Term	
	A	B	C	D	E	F	G
1	subset	attribute	entry	category	type	description	CV term id
2	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_0000003
3	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046
4	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0000001
5	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_0000002
6	plants	Genotype	genotype	string	string	Genotype	http://purl.obolibrary.org/obo/PO_0000004
7	harvests	Lot	lot	identifier	numeric	Pool of several harvests	http://purl.obolibrary.org/obo/PO_0000005
8	harvests	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_0000003
9	harvests	Truss	truss	qualitative	string	Position on the stem of the truss	http://purl.obolibrary.org/obo/PO_0000006
10	harvests	HarvestDate	harvestdate	string	string	Harvest date	http://purl.obolibrary.org/obo/PO_0000007
11	harvests	HarvestHour	harvesthour	string	string	Harvest hour	http://purl.obolibrary.org/obo/PO_0000008
12	harvests	FruitAge	fruitage	factor	string	Fruit development stage	http://purl.obolibrary.org/obo/PO_0000009
13	harvests	FruitPosition	fruitposition	qualitative	numeric	Position on the truss of the fruit	http://purl.obolibrary.org/obo/PO_0000010
14	harvests	FruitDiameter	fruitdiameter	quantitative	numeric	Fruit diameter (mm)	http://purl.obolibrary.org/obo/PO_0000011
15	harvests	FruitHeight	fruitheight	quantitative	numeric	Fruit height (mm)	http://purl.obolibrary.org/obo/PO_0000012
16	harvests	FruitWeight	fruitweight	quantitative	numeric	Fruit fresh weight (g)	http://purl.obolibrary.org/obo/PO_0000013
17	samples	SampleID	sampleid	identifier	numeric	Sample identifier	http://purl.obolibrary.org/obo/PO_0000014
18	samples	Lot	lot	identifier	numeric	Pool of several harvests	http://purl.obolibrary.org/obo/PO_0000005
19	samples	NbFruit	nbfruit	quantitative	numeric	Fruit Number per sample	http://purl.obolibrary.org/obo/PO_0000015
20	samples	GellyFW	gellyfw	quantitative	numeric	Gelly fresh weight (g) per sample	http://purl.obolibrary.org/obo/PO_0000016
21	samples	GellyFW	gellyfw	quantitative	numeric	Gelly per Fruit (estimated g)	http://purl.obolibrary.org/obo/PO_0000017
22	samples	BER	ber	string	string	BER	http://purl.obolibrary.org/obo/PO_0000018
23	compounds	SampleID	sampleid	identifier	numeric	Sample identifier	http://purl.obolibrary.org/obo/PO_0000014
24	compounds	DPA	dpa	factor	numeric	Day Per Anthesis	http://purl.obolibrary.org/obo/PO_0000019
25	compounds	MassBefore	massbefore	quantitative	numeric	Mass before (g)	http://purl.obolibrary.org/obo/PO_0000020
26	compounds	MassMIA	massmia	quantitative	numeric	Mass MIA (g)	http://purl.obolibrary.org/obo/PO_0000021
27	compounds	ROT	rot	quantitative	numeric	Rdt (% MIA/DW)	http://purl.obolibrary.org/obo/PO_0000022
28	compounds	Starch1	starch1	quantitative	numeric	Dosage amidon (Spoids/MIA)	http://purl.obolibrary.org/obo/PO_0000023
29	compounds	Starch2	starch2	quantitative	numeric	amidon (g/gDW)	http://purl.obolibrary.org/obo/PO_0000024
30	compounds	RHAMNOSE	rhamnose	quantitative	numeric	RHAMNOSE	http://purl.obolibrary.org/obo/PO_0000025

Reference ontologies

a conceptualization associated with a set of facts (domain-dependent)



<http://agroportal.lirmm.fr/ontologies>



Entity ⇔ unit observational (e.g. samples, compounds, ...)

Attribute ⇔ Variable, Feature, ... (e.g. Plants, Fruits, Glucose, Rank, ...)

## Opportunity for the Web Of Data

### Metadata

		Category		CV Term	
Attributes					
1	subset	attribute	entry	category	type
2	plants	PlantID	plantid	identifier	Plant identifier
3	plants	Row	row	qualitative	Row of the individual plant on the table
4	plants	Plant	plant	string	Code identifier of the individual plant
5	plants	Treatment	treatment	factor	Treatment applied on plants
6	plants	Genotype	genotype	string	Genotype
7	harvests	Lot	lot	identifier	Pool of several harvests
8	harvests	PlantID	plantid	numeric	Plant identifier
9	harvests	Truss	truss	string	Position on the stem of the truss
10	harvests	HarvestDate	harvestdate	string	Harvest date
11	harvests	HarvestHour	harvesthour	string	Harvest hour
12	harvests	FruitAge	fruitage	factor	Fruit development stage
13	harvests	FruitPosition	fruitposition	qualitative	Position on the truss of the fruit
14	harvests	FruitDiameter	fruitdiameter	numeric	Fruit diameter (mm)
15	harvests	FruitHeight	fruitheight	numeric	Fruit height (mm)
16	harvests	FruitWt	fruitwt	numeric	Fruit Fresh Weight(g)
17	samples	SampleID	sampleid	numeric	Sample identifier
18	samples	Lot	lot	numeric	Pool of several harvests
19	samples	NbFruit	nbfruit	numeric	Fruit Number per sample
20	samples	GellyWt	gellywt	numeric	Gelly Free Weight(g) per sample
21	samples	GellyFruit	gellyfruit	numeric	Gelly per Fruit (estimated g)
22	samples	BER	ber	string	BER
23	compound	SampleID	sampleid	numeric	Sample identifier
24	compound	DPA	dpa	factor	Day Per Anthesis
25	compound	MassBefore	massbefore	numeric	mass MIA (g)
26	compound	MassMIA	massmia	numeric	mass MIA (g)
27	compound	ROT	rot	numeric	Rdt (% MIA/DW)
28	compound	Starch1	starch1	numeric	Dosage amidon (%poids/MIA)
29	compound	Starch2	starch2	numeric	amidon (g/gDW)
30	compound	RHAMNOSE	rhamnose	numeric	RHAMNOSE
Entities					
1	link	obtainedfrom	subsets	identifier	file
2	plants	PlantID	plants	plants	Plant features
3	1 harvests	lot	harvests	harvests	Harvest features
4	2 samples	sampleid	samples	samples	Samples features
5	3 compound	sampleid	compounds	compounds	Compound quantifications
6	3 enzymes	sampleid	enzymes	enzymes	Enzyme features

Other Metadata

images,  
geolocation,  
meteorology,  
other data sets,  
...

### Linked Data



Resource Description Framework (RDF)



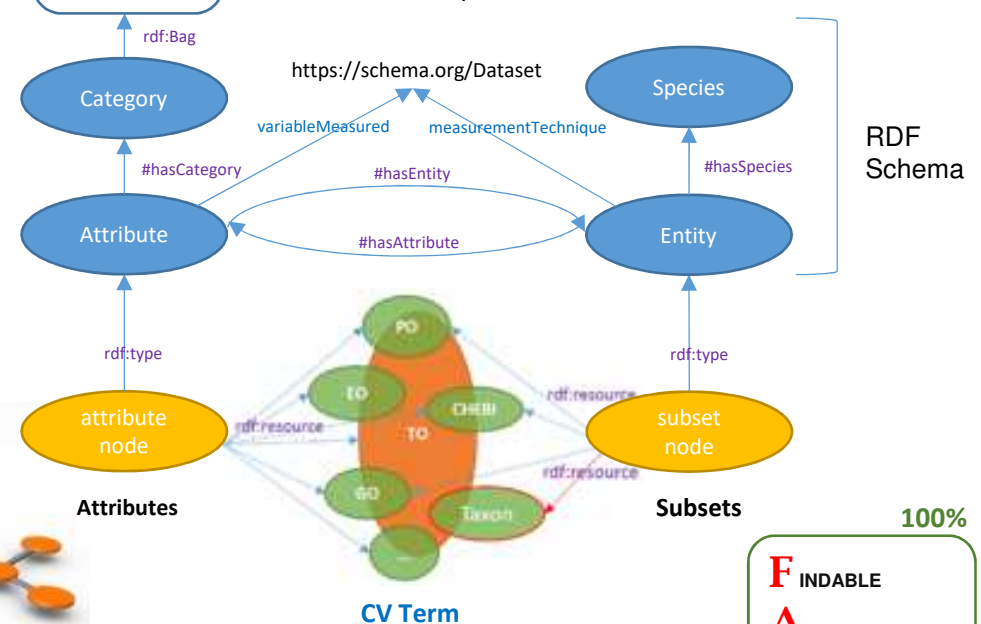
### Application ontologies

a conceptual model or **ontologies** within a **knowledge base**.

### Typical queries:

Search for a particular **Trait**

Entity + Attribute = **Trait**



RDF Schema

100%

**F**INDABLE  
**A**CCESIBLE  
**I**NTEROPERABLE  
**R**E-USABLE



## FAIR principles

- can **improve** the **data lifecycle** by relying on technologies,
- but **only** if they **serve** the concerned **communities by meeting their expectations.**

The role of a data authority is to translate principles into standards in agreement with the concerned communities



ODAM Framework  
Open Data for Access and Mining

<https://fr.slideshare.net>

“Make your data great now”

towards Open Data and Reproducible Research

“Make your data great again”

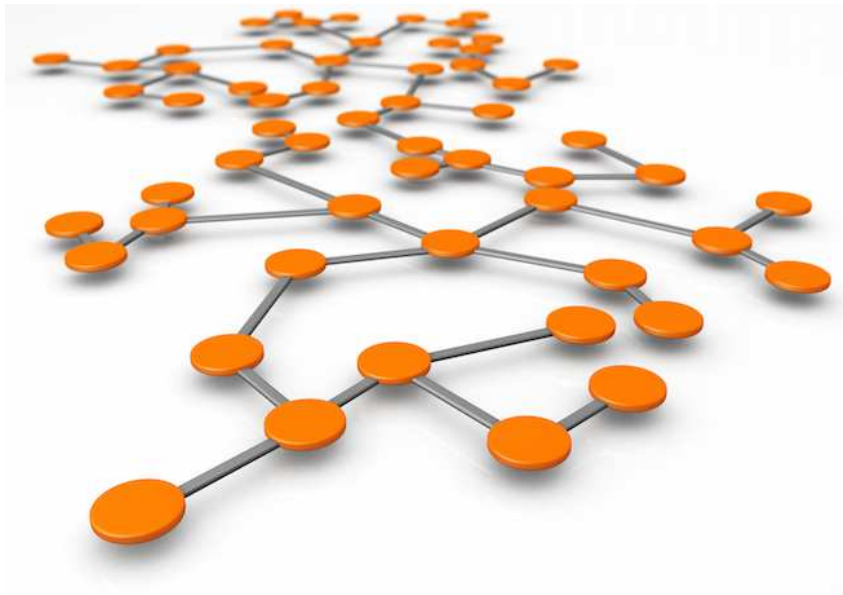
towards Linked Open Data

Data Explorer

<http://pmb-bordeaux.fr/dataexplorer/>



<https://hub.docker.com/r/odam/getdata>



**Thank you for your attention**



# Appendix 1

**Data Explorer**



ODAM - Data Explorer

About




Visual data exploration  
a first key step for deeper analyses


Created by Daniel Jacob - INRA UMR 1332 BFP, PMB - 2016 - Version 1.1


Give an open access to your data and make them ready to be mined - A data explorer as bonus

ODAM (Open Data for Access and Mining) is an Experiment Data Table Management System (EDTMS) that implements a simple way to make research data broadly accessible and fully available for reuse, including by a script language such as R. The main purpose is to make a dataset accessible online with a minimal effort from the data provider, and to allow any scientists or bioinformaticians to be able to explore the dataset and then extract a subpart or the totality of the data according to their needs. For more information/explanation, see [online presentation](#)


- Test online with the [FRIM dataset](#) - DOI [10.5281/zenodo.154041](https://doi.org/10.5281/zenodo.154041)
- To install the ODA software suite on your own hardware (laptop or server), get the docker images and the installation instruction to the [Docker Hub](#) site.
- To prepare your own data subsets, see the [tutorial on metadata files](#)
- Test online the getData API through the web [swagger UI](#) (See [Github ODA/API](#) )
- For open data access through web services within R, see the [R ODA package](#) and [How to use it](#)
- All ontology terms were picked from [AgroPortal](#) and especially by leveraging on their very useful [Annotator module](#)

 **INRA**  
SCIENCE & IMPACT

 **plotly**  
<https://plot.ly/>

 API

lightweight tools

 **R shiny**

## Data exploration ⇒ Descriptive statistics

First glimpse of the data that can show trends.

Allow the data to be well characterized, which is necessary to then choose how to analyze them.



ODAM - Data Explorer

≡ frim1

Activome Features

Example online

<https://pmb-bordeaux.fr/dataexplorer/?ds=frim1>

**Subset Information**

[Data Table](#)

[Univariate](#)

[Bivariate](#)

[Multivariate](#)

[About](#)

**Data Information**



# Tomato

**Culture location**

- INVENIO, Ste Livrade, France

---

## ERASysBio++ FRIM Project

**Fruit Integrative Modelling**

The project aimed to build a virtual tomato fruit that enables the prediction of metabolite levels given genetic and environmental inputs, by an iterative process between laboratories which combine expertise in fruit biology, ecophysiology, theoretical and experimental biochemistry, and biotechnology.

---

 © INRA UMR 1332 BFP - Metabolism Team - Yves Gibon - 2017



ODAM - Data Explorer

≡ frim1



FRIM1

## Using Data

A	B	C	D	E	F	G	H	I	J
SampleID	Treatment	DevStage	FruitAge	FruitPosition	FruitDiamete	FruitHeight	FruitFW	Rank	Truss
115	Control	FF.01	07DPA	3	11.95	10.42	0.81 A		T7
121	Control	FF.03	22DPA	3	36.13	31.77	21.43 A		T6
164	Control	FR.01	42DPA	2	51.09	46.85	64.05 A		T5
353	Control	FR.04	55DPA	5	48.28	43.35	66.64 A		T5
355	Control	FR.04	55DPA	3	49.84	44.93	66.98 A		T5
413	Control	FR.02	47DPA	1	60.48	54.23	106.13 A		T7
512	Control	FF.03	21DPA	NA	41	35.82	37.22 A		TA
117	Control	FF.01	07DPA	3	13.44	12.39	1.14 A		T7
536	Control	FR.02	47DPA	NA	59.4	49.05	87.28 A		TA
544	Control	FR.03	50DPA	NA	57.31	47.69	92.86 A		TA
158	Control	FF.04	35DPA	5	58.38	49.3	92.86 A		T5
109	Control	FF.03	22DPA	7	43.37	35.77	38.73 A		T5
134	Control	FF.02	15DPA	3	27.89	23.8	9.88 A		T7
31	Control	FF.01	08DPA	4 NA	NA		0.48 A		T6
179	Control	FF.03	28DPA	3	53.68	45.43	65.34 A		T7
383	Control	FF.04	34DPA	5	47.04	41.19	48.96 A		T7
425	Control	FR.04	55DPA	2	62.74	50.27	115.3 A		T7
520	Control	FF.03	30DPA	NA	48.86	41.52	52.94 A		TA
419	Control	FR.03	50DPA	2	55.63	48.02	86.79 A		T7
138	Control	FF.02	15DPA	6	27.96	22.14	9.69 A		T7
143	Control	FF.03	29DPA	4	48.45	42.92	51.35 A		T6
365	Control	FR.02	47DPA	5	55.11	44.9	71.82 A		T6
127	Control	FF.03	27DPA	3	45.71	43.28	47.8 A		T5
188	Control	FR.01	42DPA	3	55.38	47.1	77.39 A		T6

Bibliothèques

DATA

DATA (C:)

bsflow

nmr\_examples

nmrAnnot

nmrspec

ODAM

data

frim1

Nom

images

a\_attributes.tsv

aliquots.txt

aminoacids.txt

compounds.txt

enzymes.txt

FRIM1Quantities.txt

infos.md

lipids\_ag.txt

plants.txt

plato\_HexosesP.txt

pools.txt

qnmr\_metabo.txt

qnmr\_pools.txt

s\_subsets.tsv

samples.txt





ODAM - Data Explorer

≡ frim1



FRIM1

A	B	C	D	E	F	G	H	I	J	K
PoolID	glucose	saccharose	fructose	galactose	mannose	rhamnose	acetate	chlorogenate	citrate	fumarate
F1	409.395048	70.3619927	372.900223		0	3.42180438	4.30631345	9.69094266	11.7899795	181.07691
F2	393.12983	88.5173122	372.905171		0	2.71163313	4.52225358	7.25107692	10.5400556	149.685243
F3	356.337649	48.8917522	353.455502		0	3.39882904	4.35861259	10.2973691	10.6132957	165.939757
F4	555.179184	41.0661545	659.640831		0	3.52323262	4.34025402	4.53889647	3.50545496	116.957376
F5	628.037811	38.8434456	701.485209		0	3.46163677	4.41411896	4.2425186	3.71828615	100.613568
F6	619.325088	45.8307638	651.091538		0	3.60635839	5.52636053	3.84391466	5.08350871	97.710292
F7	785.009156	36.3580174	892.165913		0	3.78684739	4.4194075	3.02633059	2.58493159	69.0448901
F8	651.332424	32.7834368	798.614906		0	3.37954329	4.46102826	4.36716646	ND	85.4756868
F9	725.001348	33.0778677	871.703642		0	3.23235017	5.51633293	3.49546404	1.94856918	79.3095839
F10	623.272922	27.3827117	790.634309		0	3.82300439	5.2077975	4.40566989	1.46718766	91.6493046
F11	966.860337	27.4709019	1062.996		0	3.08272287	3.97079791	3.13304376	2.29908263	81.5836755
F12	679.296181	23.9407775	857.851923		0	3.64544893	5.52039645	3.96505573	2.23428282	82.3600813
F13	730.14179	25.7149098	936.609812		0	3.83379924	5.45157429	4.46059786	2.03956778	93.7336956
F14	873.636628	22.2726495	1029.94324		0	3.45122911	5.25520708	3.5987367	1.57125484	85.2389555
F15	751.842434	23.0485025	948.72502		0	3.61697316	4.79080951	4.10768337	1.85324312	94.7446369
F16	867.346037	23.901677	1001.43311		0	4.07176479	5.9466084	3.68065864	2.6234223	102.039955
F17	723.106879	27.7366695	970.41506		0	3.65434337	5.72801805	4.84933228	1.99337061	131.804643
F18	961.539348	23.549379	1125.36069		0	3.33529211	6.12036392	3.48034793	1.9871408	113.240742
F19	756.688509	33.9075147	975.485366	1.5449113	4.15070906	4.02275996	4.29601852	2.42827873	182.777927	0.07999949
F20	965.326168	31.7935423	1155.47423	3.12877425	4.11656458	4.32039138	3.18927544	1.889941	164.5021	0.10338685
F21	979.84239	30.7950267	1201.57086	3.23551003	4.1335882	4.83971718	3.42497797	1.65969829	183.281283	0.07597753
F22	1050.48663	35.6610303	1243.74997	3.50314447	3.73635121	6.01077229	2.98575954	0.89577819	151.452942	0.05611195
F23	801.762466	27.5085351	1103.46008	4.10489509	3.98841525	6.10821394	3.68821995	0.93232461	178.021061	0.08529778
F24	889.732794	37.9595702	1144.08439	3.4989607	3.82695526	5.3292931	3.56671696	1.2809934	184.391814	0.05145507
F25	1128.52651	29.0712046	1339.54916	5.64362444	4.31061033	6.75345563	2.53811062	1.11727033	172.623623	0.05319742
F26	363.353286	64.4392108	343.535269		0	3.02010673	3.85945307	8.58648736	12.0896674	173.150989
F27	372.084474	89.8753681	356.465076		0	2.6601792	5.2144064	7.34018965	11.7821663	151.39492
F28	326.41201	49.0725408	329.515932		0	3.53228489	6.95284913	8.8965536	12.1564146	176.572601
F29	445.215159	43.3136938	541.68046		0	3.97980076	5.25141114	5.90403916	3.95096163	113.303711
F30	515.394042	37.8155094	605.332373		0	4.0024065	5.62557098	4.03544802	4.32860718	101.964138
F31	573.749056	53.1446302	621.01976		0	3.9599372	5.75257178	4.09776668	5.73265078	98.8820964
F32	668.798348	41.8900559	742.338788		0	3.46057104	5.21139103	3.32222007	3.39640084	87.5852769

Bibliothèques

DATA

DATA (C:)

bsflow

nmr\_examples

nmrAnnot

nmrspec

ODAM

data

frim1

Nom

images

a\_attributes.tsv

aliquots.txt

aminoacids.txt

compounds.txt

enzymes.txt

FRIM1Quantities.txt

infos.md

lipids\_ag.txt

plants.txt

plato\_HexosesP.txt

pools.txt

qnmr\_metabo.txt

qnmr\_pools.txt

s\_subsets.tsv

samples.txt



ODAM - Data Explorer

≡ frim1



FRIM1

A	B	C	D	E	F	G	H	I	J	K	L
AliquotID	PGM	F16BP_Cyt	PyrK	CitS	PPI	AcoS	PFK	FruS	F16BP_Stron	GluS	ISODH_NAD
1	NA	8.97	1599.53	64.53	2767.89	1172.28	192.05	876.13	523.57	722.19	508.99
3	6844.78	85.02	1839.39	NA	3373.46	2014.02	263.33	984.08	634.14	622.22	975.77
5	5336.12	14.58	1399.41	65.84	2642.21	980.02	184.31	743.8	445.98	518.01	555.54
7	5376.43	23.81	1333.66	105.67	2663.67	1349.45	147.18	807.87	501.92	915.19	572.33
9	5230	25.21	1271.13	110.25	2575.94	865.05	177.67	792.12	459.77	510.8	421.23
11	6147.04	24.19	754.24	68.66	2272.79	1005.74	144.17	641.46	467.71	598.68	416.04
109	5416.24	35.57	1833.13	168.54	2779.48	869.08	197.73	971.88	494.3	614.79	660.68
111	8054.61	33.14	2128.92	64.59	3128.5	1830.77	270.3	1008.07	778.47	1271.77	965.12
113	3997.96	0.7	3053.11	44.2	2330.09	719.39	250.87	1223.19	360.15	526.31	437.56
115	5459.74	23.73	1944.78	59.71	3338.61	1168.66	246.51	1039.7	526.42	606.25	607.02
117	7509.32	1.71	1593.34	8.34	2597.87	1035.6	260.67	1019.02	663.14	986.63	632.42
119	6018.59	21.54	1624.31	50.1	2755.34	1245.73	243.07	915.78	530.86	832.47	659.72
361	3967.66	12.54	786.54	88.85	1828.44	363.23	131.41	525.24	268.55	337.81	210.84
363	4753.73	17.77	728.5	NA	1869.46	450.13	171.89	593.3	357.74	220.39	247.36
365	3836.19	4.16	603.74	31.55	1836.44	330.26	131.87	499.38	208.85	249.95	190.41
367	4.11	4.87	507.67	7.39	1371.13	573.06	128.65	390.78	283.25	304.36	301.7
369	3798.59	15.14	703.9	82.78	1473.25	286.02	120.11	438.09	260.44	240.08	204.19
371	5610.69	14.69	1092.42	99.09	2384.35	799.55	182.18	812.09	425.56	583.97	366.29
373	4220.68	31.7	592.57	14.81	1654.27	602.62	150.23	480.9	283.92	354.94	256.53
375	3342.92	13.68	557.38	1.32	1612.42	359.41	119.45	407.36	279.22	290.21	202.93
377	3100.01	21.21	467.95	60.43	1095.82	369.89	93.32	206.48	170.71	196.26	146.66
379	NA	NA	584.29	26.11	1787.02	668.52	158.95	438.82	262.62	334.18	291.38
381	3483.95	8.98	771.26	13.32	1514.18	497.3	115.14	337.25	234.69	271.34	216.68
384	4011.16	4.85	590.28	13.75	1245.94	378.87	109.77	344.62	193.74	165.72	196.86
385	1767.14	11.97	351.52	43.95	881.43	174.81	85.89	91.79	127.57	82.19	135.55
387	1907.36	10.54	268.66	36.35	964.63	135.77	70.58	137.41	108.43	151.45	110.26
389	1843.23	10.37	214.02	22.79	837.37	82.88	85.74	141.82	124.58	61.97	120.31
391	2322.24	15.06	349.62	52.12	1014.67	230.62	80.48	185.05	187.63	56.33	151.72
393	3826.41	18.98	466.4	103.64	1485.31	212.96	140.72	304.75	184.27	181.77	169.72
395	2483.08	18.13	556.97	63.22	1154	286.7	103.31	193.56	127.9	138.77	172.3
397	4924.26	17.89	1327.67	50.79	2225.01	657.4	160.33	638.3	398.38	458.04	495.24

Bibliothèques

DATA

DATA (C:)

bsflow

nmr\_examples

nmrAnnot

nmrspec

ODAM

data

frim1

Nom

images

a\_attributes.tsv

aliquots.txt

aminoacids.txt

compounds.txt

enzymes.txt

FRIM1Quantities.txt

infos.md

lipids\_ag.txt

plants.txt

plato\_HexosesP.txt

pools.txt

qnmr\_metabo.txt

qnmr\_pools.txt

s\_subsets.tsv

samples.txt





ODAM - Data Explorer

≡ frim1



FRIM1

A	B	C	D	E	F	
rank	father_rank	subset	identifier	file	description	
1	0	plants	PlantID	plants.txt	Plant features	
2	1	samples	SampleID	samples.txt	Sample features	
3	2	aliquots	AliquotID	aliquots.txt	Aliquots features	
4	3	cellwall_metabo	AliquotID	compounds.txt	Cell wall Compound quantifications	
5	A	B	C	D	E	F
6	subset	attribute	entry	category	type	description
7	plants	PlantID	plant	identifier	string	Plant identifier
8	plants	Rank	row	qualitative	string	Row of the individual plant on the table
9	plants	PlantNum	plantnum		numeric	Code identifier of the individual plant
10	plants	Treatment	treatment	factor	string	Treatment applied on plants
11	samples	SampleID	sample	identifier	numeric	Pool of several harvests
12	samples	PlantID			numeric	Plant identifier
	samples	Truss	truss	qualitative	string	Position on the stem of the truss
	samples	DevStage	stage	factor	string	fruit development stage
	samples	FruitAge	age	factor	string	fruit age (dpa)
	samples	HarvestDate			string	Harvest date
	samples	HarvestHour			string	Harvest hour
	samples	FruitPosition			string	Poistion on the truss of the fruit
	samples	FruitDiameter		quantitative	numeric	Fruit diameter (mm)
	samples	FruitHeight		quantitative	numeric	Fruit height (mm)
	samples	FruitFW		quantitative	numeric	Fruit Fresh Weight(g)
	samples	DW		quantitative	numeric	Percentage of dry matter (% DW), measured a
	aliquots	SampleID	sample		numeric	Sample identifier
	aliquots	AliquotID	aliquot	identifier	numeric	Aliquot Identifier
	cellwall_me	AliquotID	aliquot	identifier	numeric	Aliquot Identifier
	cellwall_me	MassBefore		quantitative	numeric	m av.extraction (g)
	cellwall_me	MassMIA		quantitative	numeric	masse MIA (g)

Bibliothèques

DATA

DATA (C:)

bsflow

nmr\_examples

nmrAnnot

nmrspec

ODAM

data

frim1

Nom

images

a\_attributes.tsv

aliquots.txt

aminoacids.txt

compounds.txt

enzymes.txt

FRIM1Quantities.txt

infos.md

lipids\_ag.txt

plants.txt

plato\_HexosesP.txt

pools.txt

qnmr\_metabo.txt

qnmr\_pools.txt

s\_subsets.tsv

samples.txt

### Metadata files

In order to allow data to be explored and mined, we have to adjoin some minimal but relevant metadata



ODAM - Data Explorer

≡ frim1

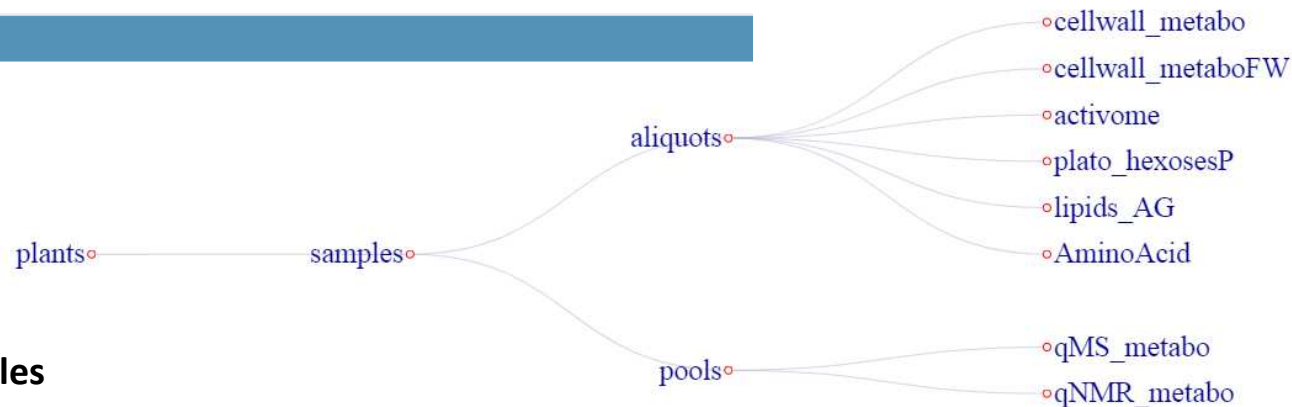
Activome Features

## Data Subset Information

Subset	Description	Identifier	WSEntry	CV_Term
samples	Sample features	SampleID	sample	[OBI_1110046] organ harvesting
cellwall_metabo	Cell wall Compound quantifications	AliquotID	aliquot	[CHEBI_24431] chemical entity
cellwall_metaboFW	Cell Wall Compound quantifications (FW)	AliquotID	aliquot	[CHEBI_24431] chemical entity
activome	Activome Features	AliquotID	aliquot	[CHEBI_24431] chemical entity
qMS_metabo	MS Compounds quantification	PoolID	pool	[CHEBI_24431] chemical entity
qNMR_metabo	NMR Compounds quantification	PoolID	pool	[OBI_1110046] organ harvesting
plato_hexosesP	Hexoses Phosphate	AliquotID	aliquot	[CHEBI_24431] chemical entity
lipids_AG	Lipids AG	AliquotID	aliquot	[CHEBI_24431] chemical entity
AminoAcid	Amino Acids	AliquotID	aliquot	[CHEBI_24431] chemical entity

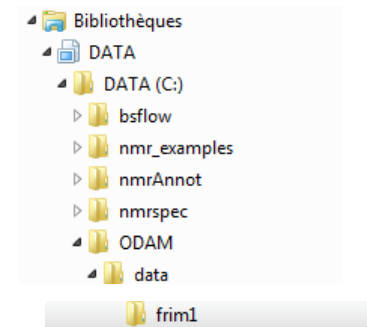
Showing 1 to 9 of 9 entries

## Data Graph



## Metadata files

In order to allow data to be explored and mined, we have to adjoin some minimal but relevant metadata



## Nom

- images
- a\_attributes.tsv
- aliquots.txt
- aminoacids.txt
- compounds.txt
- enzymes.txt
- FRIM1Quantities.txt
- infos.md
- lipids\_ag.txt
- plants.txt
- plato\_HexosesP.txt
- pools.txt
- qnmr\_metabo.txt
- qnmr\_pools.txt
- s\_subsets.tsv
- samples.txt



ODAM - Data Explorer

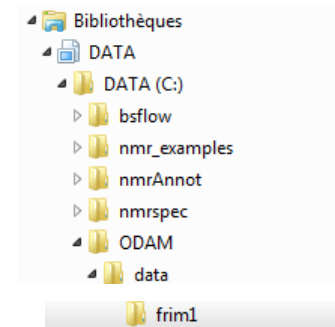
≡ frim1

Activome Features

## Data Subset Information

Subset	Description	Identifier	WSEntry	CV_Term
samples	Sample features	SampleID	sample	[OBI_1110046] organ harvesting
cellwall_metabo	Cell wall Compound quantifications	AliquotID	aliquot	[CHEBI_24431] chemical entity
cellwall_metaboFW	Cell Wall Compound quantifications (FW)	AliquotID	aliquot	[CHEBI_24431] chemical entity
activome	Activome Features	AliquotID	aliquot	[CHEBI_24431] chemical entity
qMS_metabo	MS Compounds quantification	PoolID	pool	[CHEBI_24431] chemical entity
qNMR_metabo	NMR Compounds quantification	PoolID	pool	[OBI_1110046] organ harvesting
plato_hexosesP	Hexoses Phosphate	AliquotID	aliquot	[CHEBI_24431] chemical entity
lipids_AG	Lipids AG	AliquotID	aliquot	[CHEBI_24431] chemical entity
AminoAcid	Amino Acids	AliquotID	aliquot	[CHEBI_24431] chemical entity

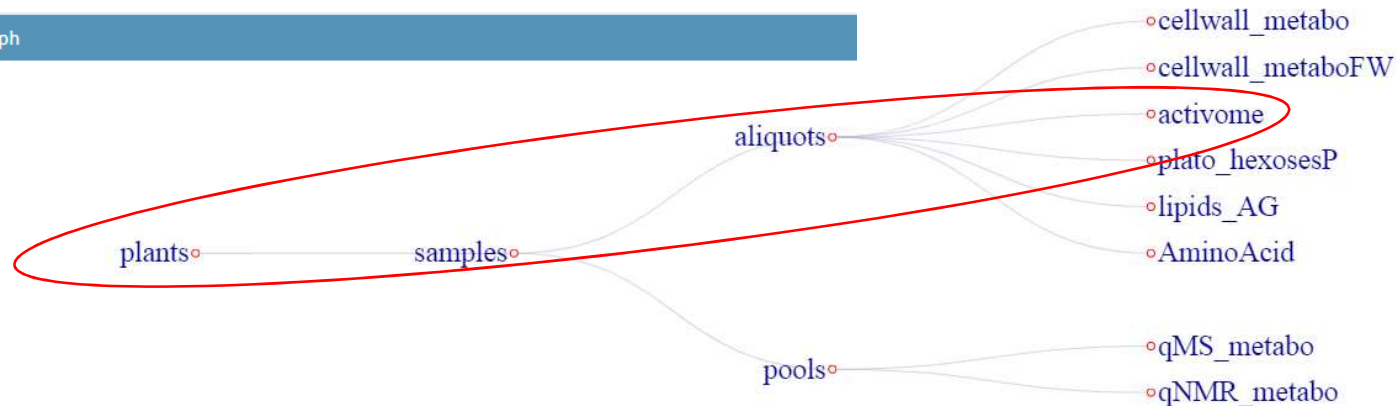
Showing 1 to 9 of 9 entries



## Nom

images  
a\_attributes.tsv  
aliquots.txt  
aminoacids.txt  
compounds.txt  
enzymes.txt  
FRIM1Quantities.txt  
infos.md  
lipids\_ag.txt  
plants.txt  
plato\_HexosesP.txt  
pools.txt  
qnmr\_metabo.txt  
qnmr\_pools.txt  
s\_subsets.tsv  
samples.txt

## Data Graph





ODAM - Data Explorer

≡ frim1

Activome Features

Subset Information

Data Table

Univariate

Bivariate

Multivariate

About

## Univariate exploration

Factor for X Axis

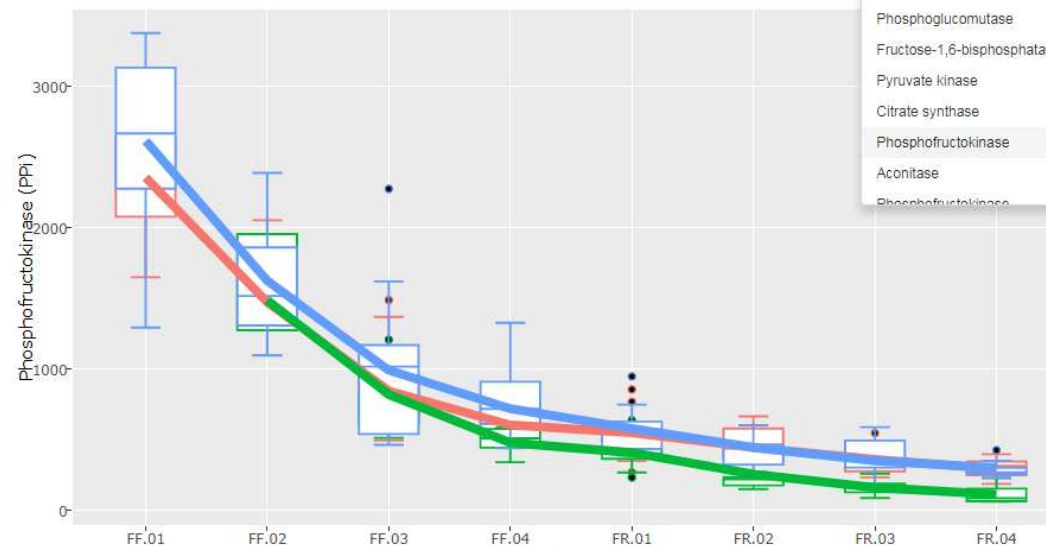
fruit development stage

Factor for Grouping

Treatment applied on plants

Variable to explore

Phosphofructokinase





ODAM - Data Explorer

≡ frim1

Activome Features

Subset Information

Data Table

Univariate

Bivariate

Multivariate

About

## Univariate exploration

Factor for X Axis

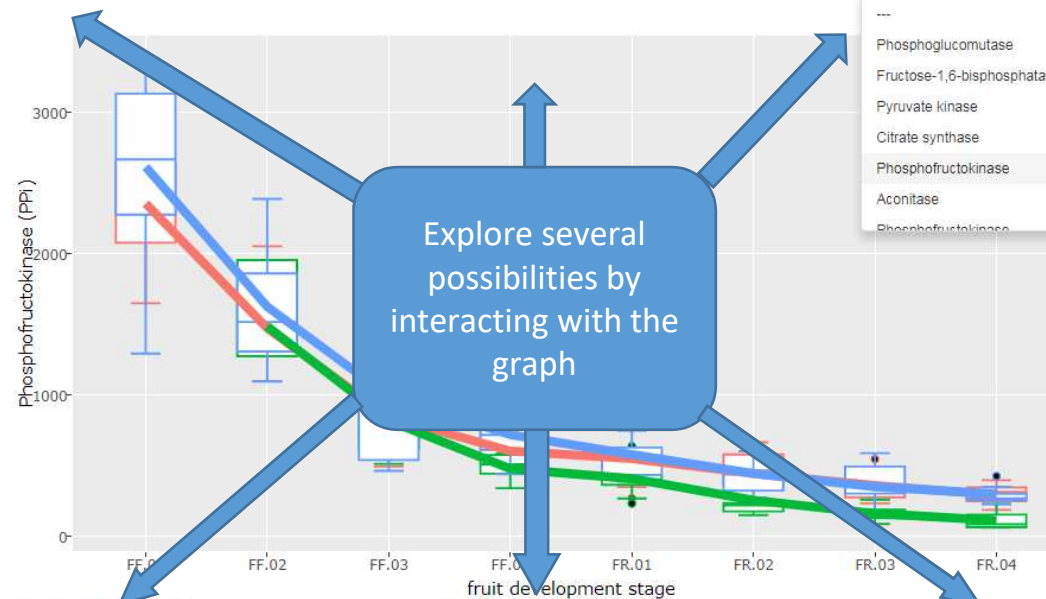
fruit development stage

Factor for Grouping

Treatment applied on plants

Variable to explore

Phosphofructokinase



Phosphoglucumutase  
Fructose-1,6-bisphosphatase  
Pyruvate kinase  
Citrate synthase  
Phosphofructokinase  
Aconitase  
Phosphofructokinase

☒ Curve☐ Log10



ODAM - Data Explorer

≡ frim1

Activome Features

Subset Information

Data Table

Univariate

Bivariate

Multivariate

About

## Bivariate exploration

Factor for Grouping

fruit development stage

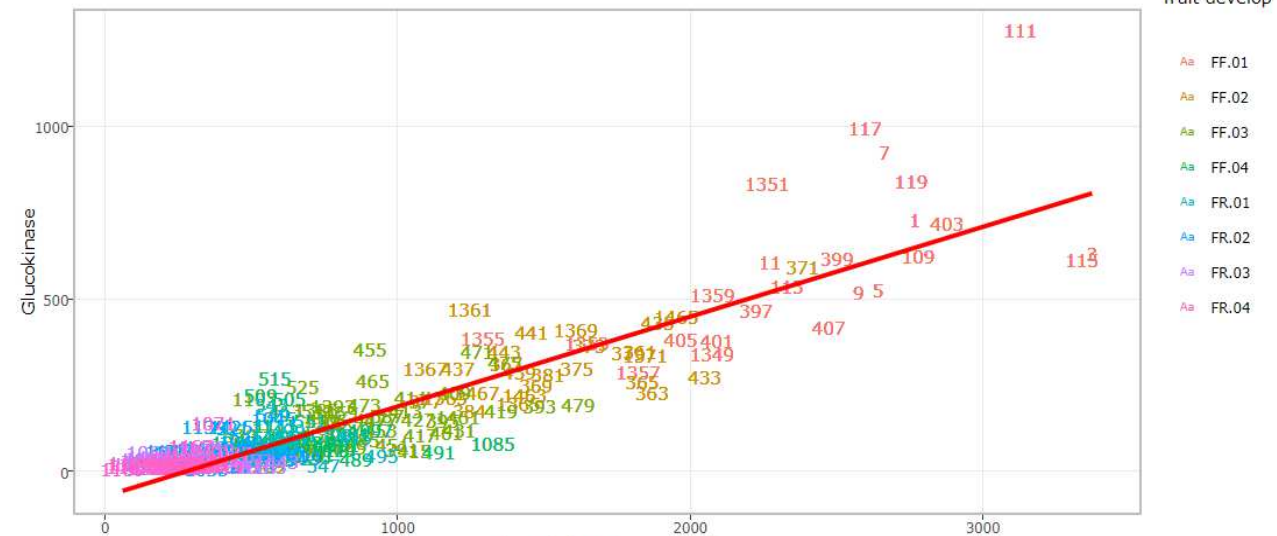
First Variable

Phosphofructokinase

Second Variable

Glucokinase

Adj R2 = 0.82001 Intercept = -72.649 Slope = 0.26042 Pvalue = 0



fruit develop

Aa FF.01  
 Aa FF.02  
 Aa FF.03  
 Aa FF.04  
 Aa FR.01  
 Aa FR.02  
 Aa FR.03  
 Aa FR.04

Select First Factor Levels

FF.01 FF.02 FF.03 FF.04 FR.01 FR.02 FR.03  
 FR.04

Select Features as Labels

---

(Un)Select Items

---

Add On:

☒ None  
☐ Reg. model (LM)  
☐ Ellipse

☒ Labels☐ Log10(X)☐ Log10(Y)☒ Global Reg. model



ODAM - Data Explorer

≡ frim1

Activome Features

Subset Information

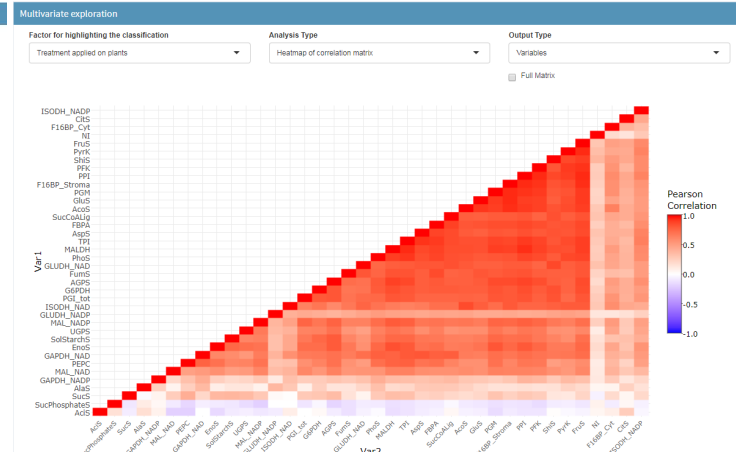
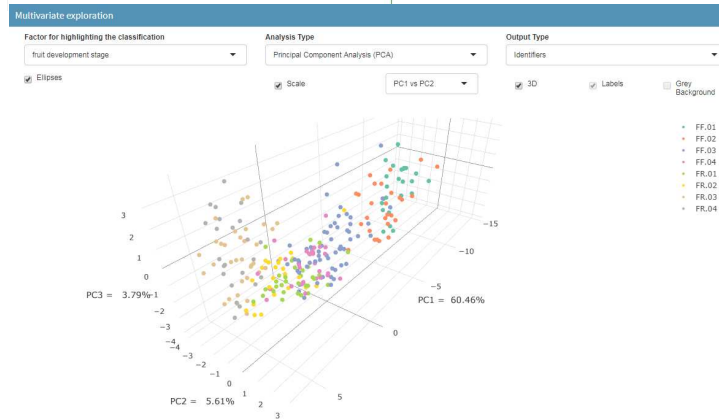
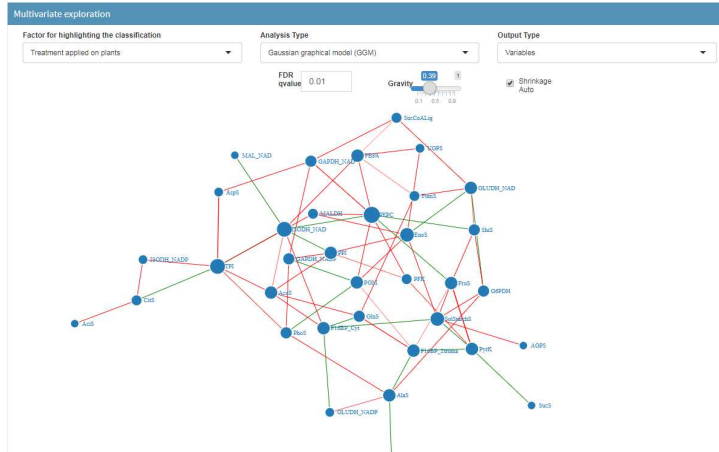
Data Table

Univariate

Bivariate

Multivariate

About







ODAM - Data Explorer

≡ frim1

Activome Features

Subset Information

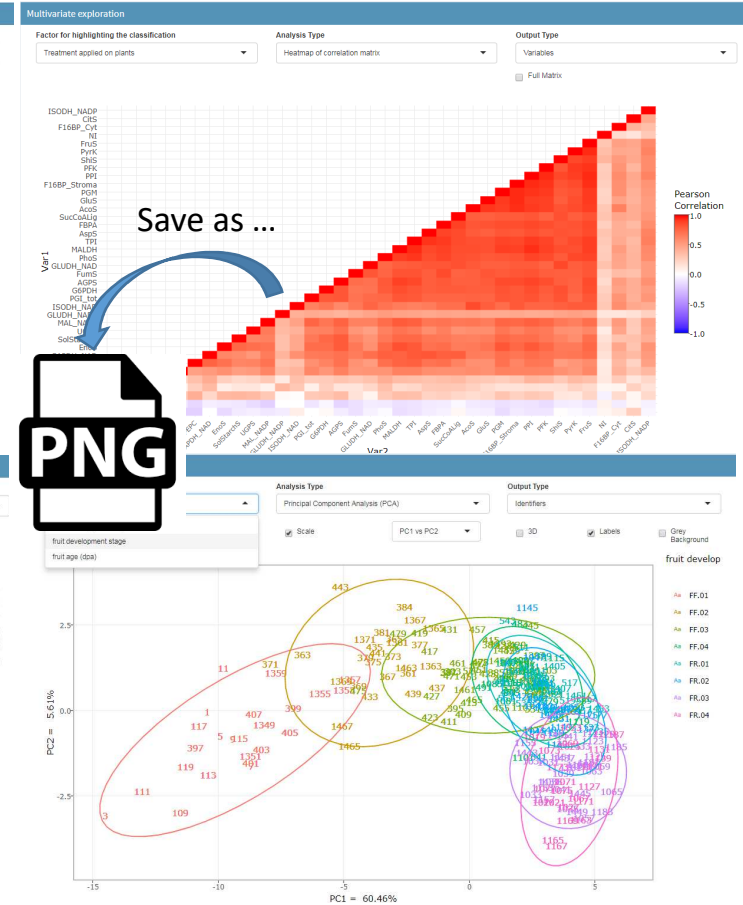
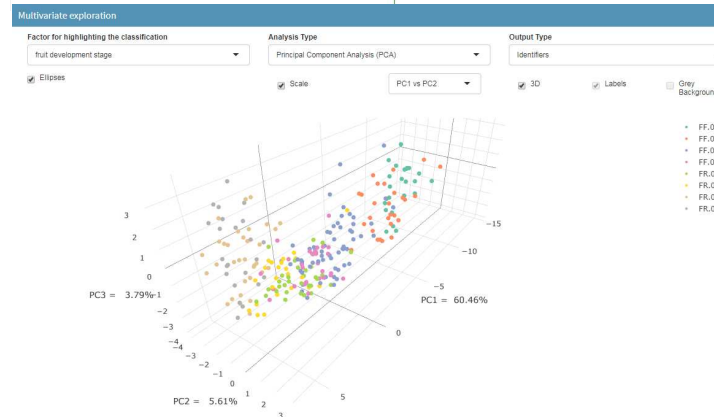
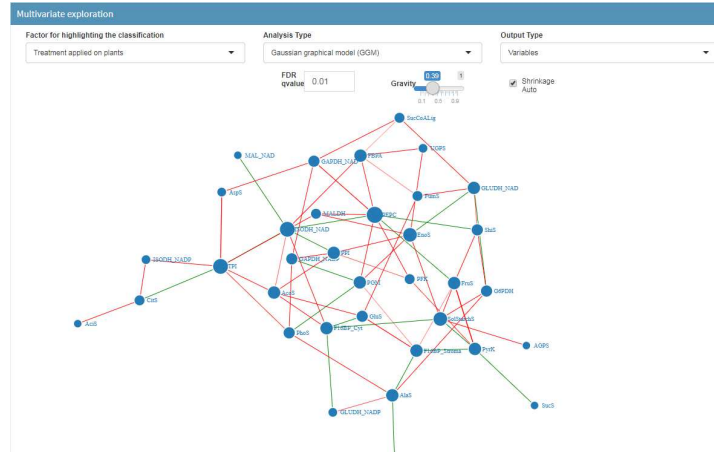
Data Table

Univariate

Bivariate

Multivariate

About







ODAM - Data Explorer

≡ frim1

Activome Features

Subset Information

Columns to show:

- ☐ PlantID
- ☐ Rank
- ☐ PlantNum
- ☒ Treatment
- ☐ Truss
- ☐ SampleID
- ☒ DevStage
- ☒ FruitAge
- ☐ HarvestDate
- ☐ HarvestHour
- ☐ FruitPosition
- ☐ FruitDiameter
- ☐ FruitHeight
- ☐ FruitFW
- ☐ DW
- ☒ AliquotID
- ☒ PGM
- ☒ F16BP\_Cyt
- ☒ PyrK
- ☒ CitS
- ☒ PPI
- ☒ AcoS
- ☒ PFK
- ☐ FruS
- ☒ F16BP\_Stroma
- ☒ GluS
- ☒ ISODH\_NAD
- ☐ EnoS
- ☐ ISODH\_NADP
- ☐ PEPC
- ☐ FBPA
- ☐ SucCoAlig
- ☐ MALDH
- ☐ AlaS
- ☐ FumS
- ☐ AcoS

Copy Excel

Export as ...

ODAM - Data Explorer.xlsx - Excel

Treatment	DevStage	FruitAge	AliquotID	PGM	F16BP_Cyt	PyrK	CitS	PPI	AcoS	PFK	F16BP_Stroma	GluS	ISODH_NAD
Control	FF.01	07DPA	397	4924.26	17.89	1327.67	50.79	2225.01	657.4	160.33	398.38	458.04	495.24
Control	FF.03	22DPA	409	3123.06	8.67	450.85	23.53	1194.12	288.49	125.19	209.7	218.39	188.53
Control	FF.01	07DPA	493	1597.7	5.16	237.69	34.63	578.74	84.92	66.91	72.58	47.22	133.34
Control	FR.04	55DPA										6.47	314.05
Control	FR.02	47DPA										9.9	227.51
Control	FF.03	21DPA										121.14	275.09
Control	FF.01	07DPA										89.87	
Control	FR.02	47DPA										368.98	336.88
Control	FR.03	50DPA										21.28	296.93
Control	FF.04	35DPA										9.47	306.54
Control	FF.03	22DPA										70.67	197.18
Control	FF.02	15DPA										82.19	135.55
Control	FF.01	08DPA										265.13	180.87
Control	FF.03	28DPA										614.79	660.68
Control	FF.04	34DPA										34.81	96.16
Control	FR.04	55DPA										55.6	122.12
Control	FF.03	21DPA										15.32	161.75
Control	FF.01	07DPA										9.84	122.83
Control	FR.03	50DPA										28.09	266.62
Control	FF.02	15DPA										394.7	176.69
Control	FF.03	29DPA										69.56	121.26
Control	FR.02	47DPA										26.46	222.35
Control	FF.03	27DPA										90.63	137.55
Control	FR.01	42DPA										25.77	114.93
Control	FF.01	08DPA										526.31	437.56
Control	FR.02	15DPA											

Showing 1 to 25 of 255 entries

As far as possible, keep the old way of using the scientist's worksheets ...



ODAM - Data Explorer    frim1    Activome Features

Subset Information

Data Table

Univariate

Bivariate

Multivariate

About

Session Information

```
options(width=256)
options(warn=1)
options(stringsAsFactors=FALSE)

library(Rodam)

# Initialize the 'ODAM' object
dh <- new('odamws', 'http://www.bordeaux.inra.fr/pmb/getdata/', 'frim1')

# Get the Data Tree
show(dh)

# Get the data subsets list
dh$subsetNames

# Get 'activome' data subset
ds <- dh$getSubsetByName('activome')

# Show all descriptions of variables
ds$LABELS

# Show all factors defined in the data subset
ds$facnames

# Show all quantitative variables defined in the data subset
ds$varnames

# Show all qualitative variables defined in the data subset
ds$qualnames

# Display a summary for each quantitative variable
summary(ds$data[, ds$varnames ])

# Boxplot of all variables defined in ds$varnames
Rank <- simplify2array(lapply(ds$varnames, function(x) { round(mean(log10(ds$data[, x])), na.rm=T) })))
Rank[!is.finite(Rank)] <- 0
colRank <- Rank - min(Rank) + 1
cols <- c('red', 'orange', 'darkgreen', 'blue', 'purple', 'brown')
boxplot(log10(ds$data[, ds$varnames]), outline=F, horizontal=T, border=cols[colRank], las=2, cex.axis=0.5)
```

The Comprehensive R  
Archive Network  
<https://cran.r-project.org>

*The R  
package  
Rodam*



Copy-Paste

... while allowing a way to be more efficient ...

## Appendix 2



**GET** <http://myhost.org/getdata/<data format>/<dataset name>/< ... >/< ... >>





## REST Services: hierarchical tree of resource naming (URL)

GET <http://myhost.org/getdata/<data format>/<dataset name>/< ... >/< ... >>

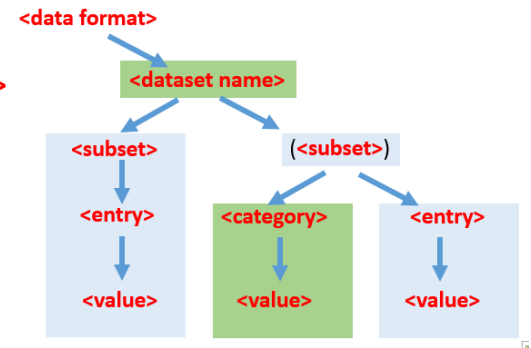
Field	Description	Examples
<data format>	format of the retrieved data; possible values are: 'xml' or 'tsv'	tsv
<dataset name>	Short name (tag) of your dataset	frim1
<subset>	Short name of a data subset	samples
<entry>	Name of an attribute entry (defined by the user in the a_attribute file (column 'entry'))	sampleid
<category>	Name of the attribute category; (assigned by the user in the a_attribute file (column 'category')) possible values are: 'identifier', 'factor', 'qualitative', 'quantitative'	quantitative
(<subset>)	Set of data subsets by merging all the subsets with lower rank than the specified subset and following the pathway defined by the "is_part_of" links.	(samples) ⇔ plants + harvests + samples
<value>	Exact value of the desired entry or category	1, factor



## REST Services: hierarchical tree of resource naming (URL)

GET <http://myhost.org/getdata/<data format>/<dataset name>/< ... >/< ... >>

- **Get the subset list of a dataset**  
<http://myhost.org/getdata/<data format>/<dataset name>>
- **Get all values within a data subset**  
<http://myhost.org/getdata/<data format>/<dataset name>/<subset>>
- **Get values within a data subset for a specific value of an entry**  
<http://myhost.org/getdata/<data format>/<dataset name>/<subset>/<entry>/<value>>
- **Get all values within a set of data subsets**  
[http://myhost.org/getdata/<data format>/<dataset name>/\(<subset>\)](http://myhost.org/getdata/<data format>/<dataset name>/(<subset>))
- **Get values within a set of data subsets for a specific value of an entry**  
[http://myhost.org/getdata/<data format>/<dataset name>/\(<subset>\)/<entry>/<value>](http://myhost.org/getdata/<data format>/<dataset name>/(<subset>)/<entry>/<value>)
- **Get the attribute list within a set of data subsets for a specific category**  
[http://myhost.org/getdata/<data format>/<dataset name>/\(<subset>\)/<category>](http://myhost.org/getdata/<data format>/<dataset name>/(<subset>)/<category>)



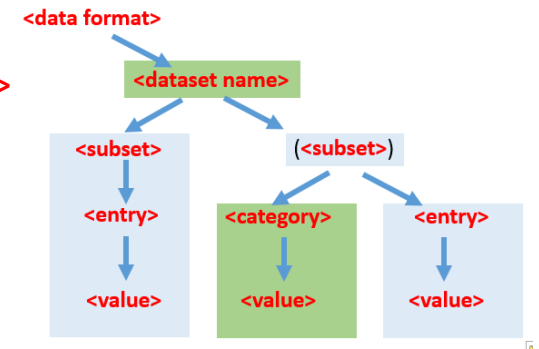



## REST Services: hierarchical tree of resource naming (URL)

GET <http://myhost.org/getdata/<data format>/<dataset name>/< ... >/< ... >>



<http://pmb-bordeaux.fr/odamsw/>




**swagger**

### Open Data for Access and Mining 1.0.0 OAS3

<http://www.bordeaux.inra.fr/pmb/api/ODAM/1.0.0/odam-oas3.yaml>

GET	/getdata/check/{dataset}	test checks
GET	/getdata/tsv/{dataset}	Get the subset list of a dataset
GET	/getdata/tsv/{dataset}/{subset}?limit={limit}	Get all values of a data subset
GET	/getdata/tsv/{dataset}/{(subset)}?limit={limit}	Get all values of a merged data subsets
GET	/getdata/tsv/{dataset}/{(subset)}/entry	Get the entry list of a merged data subsets
GET	/getdata/tsv/{dataset}/{(subset)}/{entry}/{value}?limit={limit}	Get all values of a merged data subsets for a specific value of an (WS)entry
GET	/getdata/tsv/{dataset}/{(subset)}/{category}	Get the variable list within the specified category of a merged data subsets



<http://pmb-bordeaux.fr/odamsw/>

**GET** `/getdata/tsv/{dataset}/{subset}/{entry}/{value}?limit={limit}`

dataset identifier

data subset identifier

entry identifier

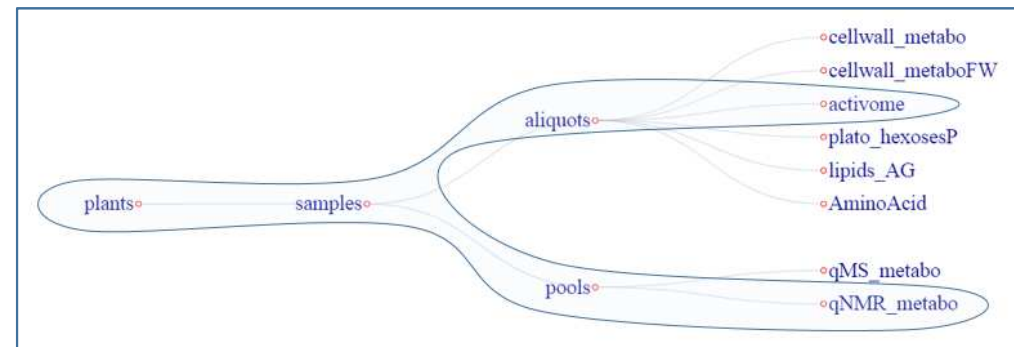
value as filter criteria for entry attribute

limit the number of lines in the output table (0 means no limit)

**Request URL**

`http://pmb-bordeaux.fr/getdata/tsv/frim1/(activome,qNMR_metabo)/sample/365?limit=10`

Get of data subsets by **merging all the subsets** with lower rank than the specified subsets and following the pathway **defined by the “obtainedFrom” links**.



(activome,qNMR\_metabo) ⇔ plants + samples  
+ (aliquots+activome)  
+ (pools+qNMR\_metabo)



Example based on FRIM - Fruit Integrative Modelling







# Retrieving Data within R

CRAN 0.1.4

## The R package Rodam

The Comprehensive R Archive Network  
<https://cran.r-project.org/web/packages/Rodam/index.html>

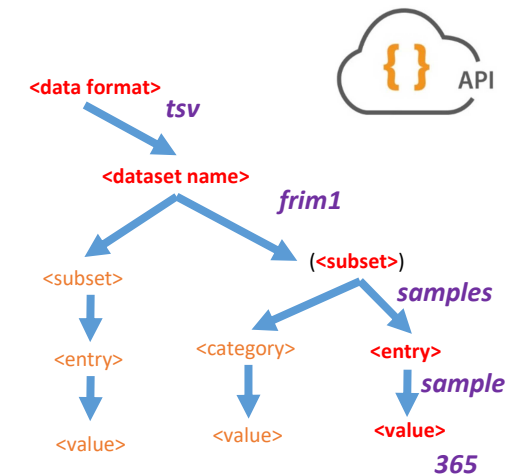
```
library(Rodam)

# Get the Data Tree
dh <- new("odamws", "http://www.bordeaux.inra.fr/pmb/getdata/", "frim1")
dn <- show(dh)
```

levelName	SetID	Identifier	WSEntry	Description	Count
plants	1	PlantID	plant	Plant features	552
--samples	2	SampleID	sample	Sample features	1287
--aliquots	3	AliquotID	aliquot	Aliquots features	2571
--cellwall_metabo	4	AliquotID	aliquot	Cell wall Compound quantifications	343
--cellwall_metaboFW	5	AliquotID	aliquot	Cell Wall Compound quantifications (FW)	343
--activome	6	AliquotID	aliquot	Activome Features	1260
--plato_hexosesP	10	AliquotID	aliquot	Hexoses Phosphate	1260
--lipids_AG	11	AliquotID	aliquot	Lipids AG	272
--AminoAcid	12	AliquotID	aliquot	Amino Acids	342
--pools	7	PoolID	pool	Pools of remaining pools	948
--qMS_metabo	8	PoolID	pool	MS Components quantification	399
--qNMR_metabo	9	PoolID	pool	NMR Components quantification	948

```
# get data from 'samples' subset with a constraint
dh$getDataByName("samples", "sample 365")
```

PlantID	Rank	PlantNum	Treatment	SampleID	Truss	DevStage	FruitAge	HarvestDate	HarvestHour	FruitPosition	FruitDiameter	FruitHeight	FruitFW
1	E35	E	311	Control	365	T6	FR.02	47DPA	40423	0.5	55.46	48.98	83.32
2	A17	A	17	Control	365	T6	FR.02	47DPA	40423	0.5	56.59	47.77	82.02
3	A8	A	8	Control	365	T6	FR.02	47DPA	40423	0.5	55.11	44.90	71.82
4	D3	D	210	Control	365	T6	FR.02	47DPA	40423	0.5	49.28	44.35	58.28
5	H11	H	356	Control	365	T6	FR.02	47DPA	40423	0.5	46.68	38.69	49.25



GET [http://www.bordeaux.inra.fr/pmb/getdata/tsv/frim1/\(samples\)/sample/365](http://www.bordeaux.inra.fr/pmb/getdata/tsv/frim1/(samples)/sample/365)



Example based on FRIM - Fruit Integrative Modelling



## Retrieving Data within R

CRAN 0.1.4

data and metadata are all available and accessible by scripting languages(R, API)

Merge the **activome data** AND the **metabolome data** acquire on the same samples

```
### Load the R ODAM package
library(Rodam)
```

```
### Get the merged data of two data subsets based on their common identifiers
dh <- new('odamws', wsURL='https://pmb-bordeaux.fr/getdata/', dsname='friml')
setNameList <- c("activome", "qNMR_metabo")
dsMerged <- dh$getSubsetByName(setNameList)
```



```
> dim(dsMerged$data)
[1] 948 83
```

```
> dsMerged$LABELS
```

Subset	Attribute	Description	Type
1	samples	SampleID	Identifier
2	plants	Treatment	Factor
3	samples	DevStage	Factor
4	samples	FruitAge	Factor
5	activome	PGM	Variable
6	activome	Fl6BP_Cyt	Variable
7	activome	PyrK	Variable
8	activome	CitS	Variable
9	activome	PPI	Variable
10	activome	AcoS	Variable
...			
43	qNMR_metabo	glucose	Variable
44	qNMR_metabo	saccharose	Variable
45	qNMR_metabo	fructose	Variable
46	qNMR_metabo	galactose	Variable
47	qNMR_metabo	mannose	Variable
48	qNMR_metabo	rhamnose	Variable
49	qNMR_metabo	acetate	Variable
50	qNMR_metabo	chlorogenate	Variable
...			

data

metadata



Example based on FRIM - Fruit Integrative Modelling



## Retrieving Data within R

CRAN 0.1.4

data and metadata are all available and accessible by scripting languages(R, API)

Visualization of data set intersections acquire on the same samples



```

library(Rodan)
library(UpSetR)

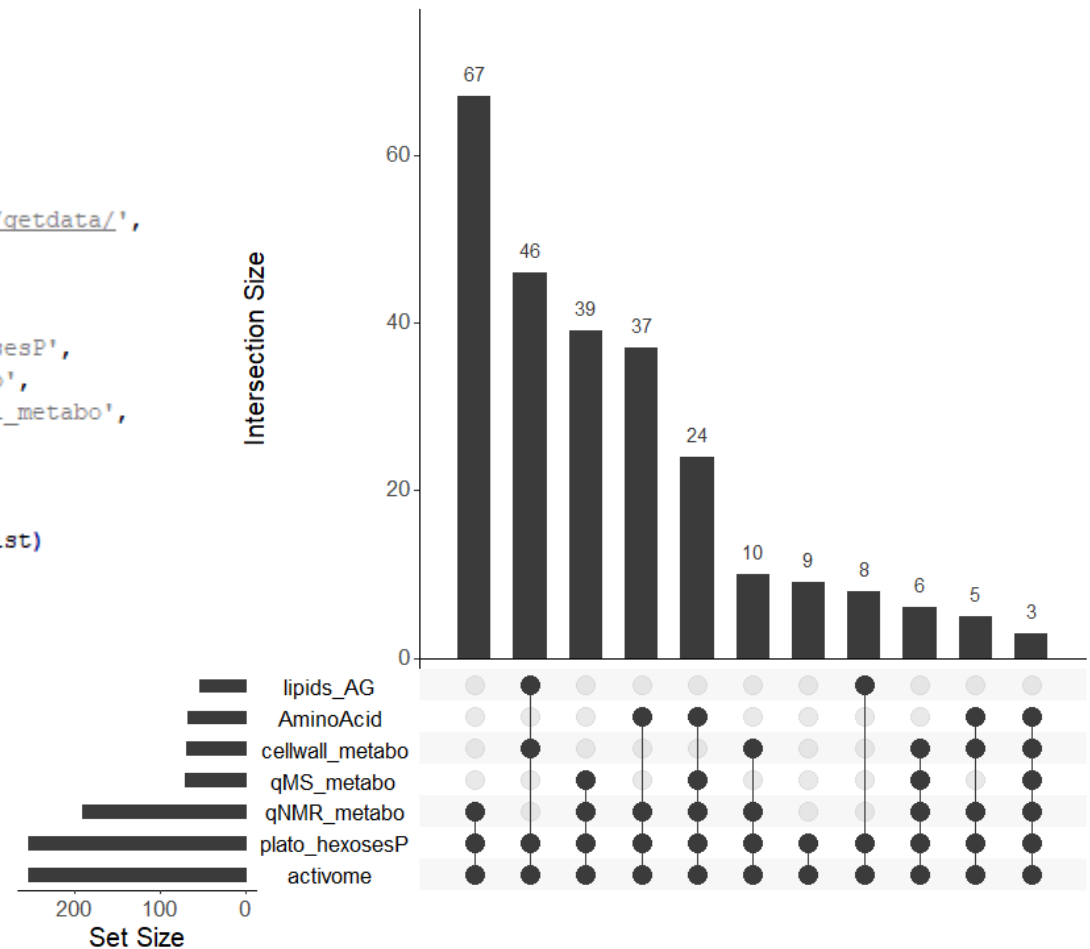
# Initialize the 'ODAM' object
dh <- new('odamws',
          wsURL='https://pmb-bordeaux.fr/getdata/',
          dsname='friml')

# List of data subsets
setNameList <- c('activome', 'plato_hexosesP',
                 'AminoAcid', 'qMS_metabo',
                 'qNMR_metabo', 'cellwall_metabo',
                 'lipids_AG' )

# Get the UpSet Table
upset.table <- dh$getUpSetTable(setNameList)

# Plot the UpSet Graphic
upset(upset.table, sets = setNameList,
      order.by = "freq",
      point.size=5,
      text.scale=1.8)

```



Example based on FRIM - Fruit Integrative Modelling



## ODAM facilitates the subsequent data mining

Make both  
metadata and data  
available for  
data mining.

Experimentation  
/ Analysis

Data / Metadata

Data Mining

MFA  
rCCA  
pLDA  
...

