

8 avr. 2019 Toulouse (France)



FAIR principles, a new opportunity to improve the data lifecycle

Preliminary question :

Data Capture ?

Daniel Jacob

INRA UMR 1332 BFP – Metabolism Group
Bordeaux Metabolome Facility





Is a



transversal INRA CATI^(*) whose vocation
is to

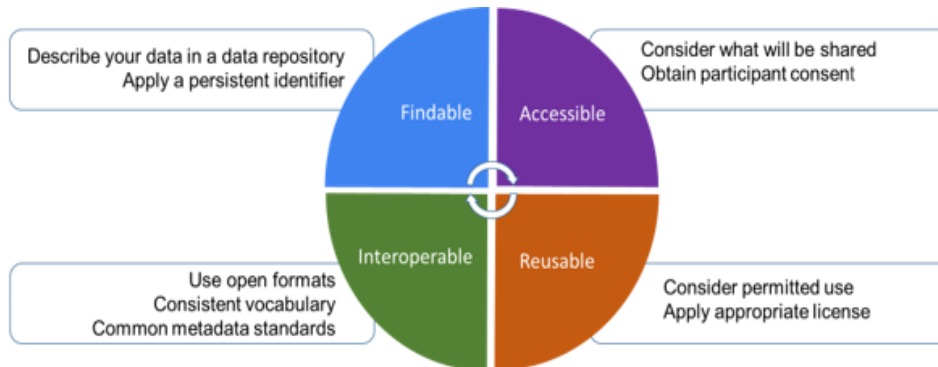
promote, support and valorize
the implementation of
data and code management and
sharing

for the purpose of

reuse and reproducibility of
science.

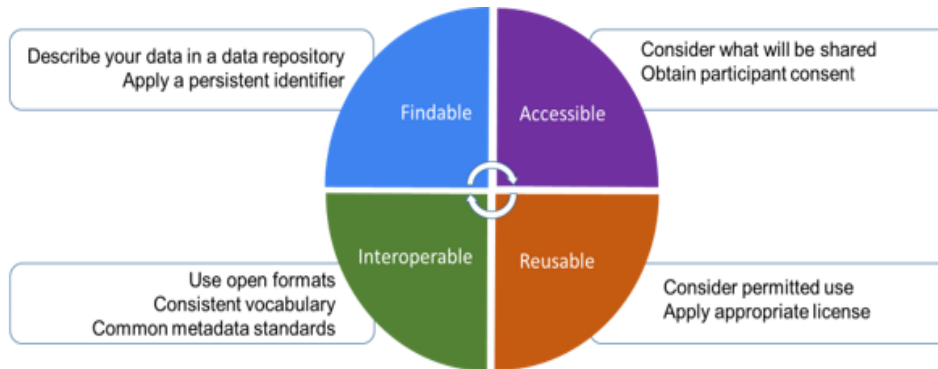
^(*) *Centre Automatisé de Traitement de l'Information* - Automated Information Processing Center

THE FAIR DATA PRINCIPLES



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

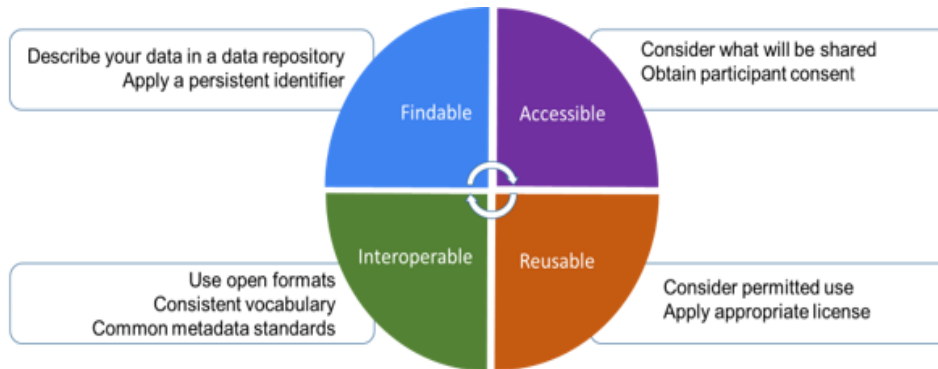
THE FAIR DATA PRINCIPLES



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 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

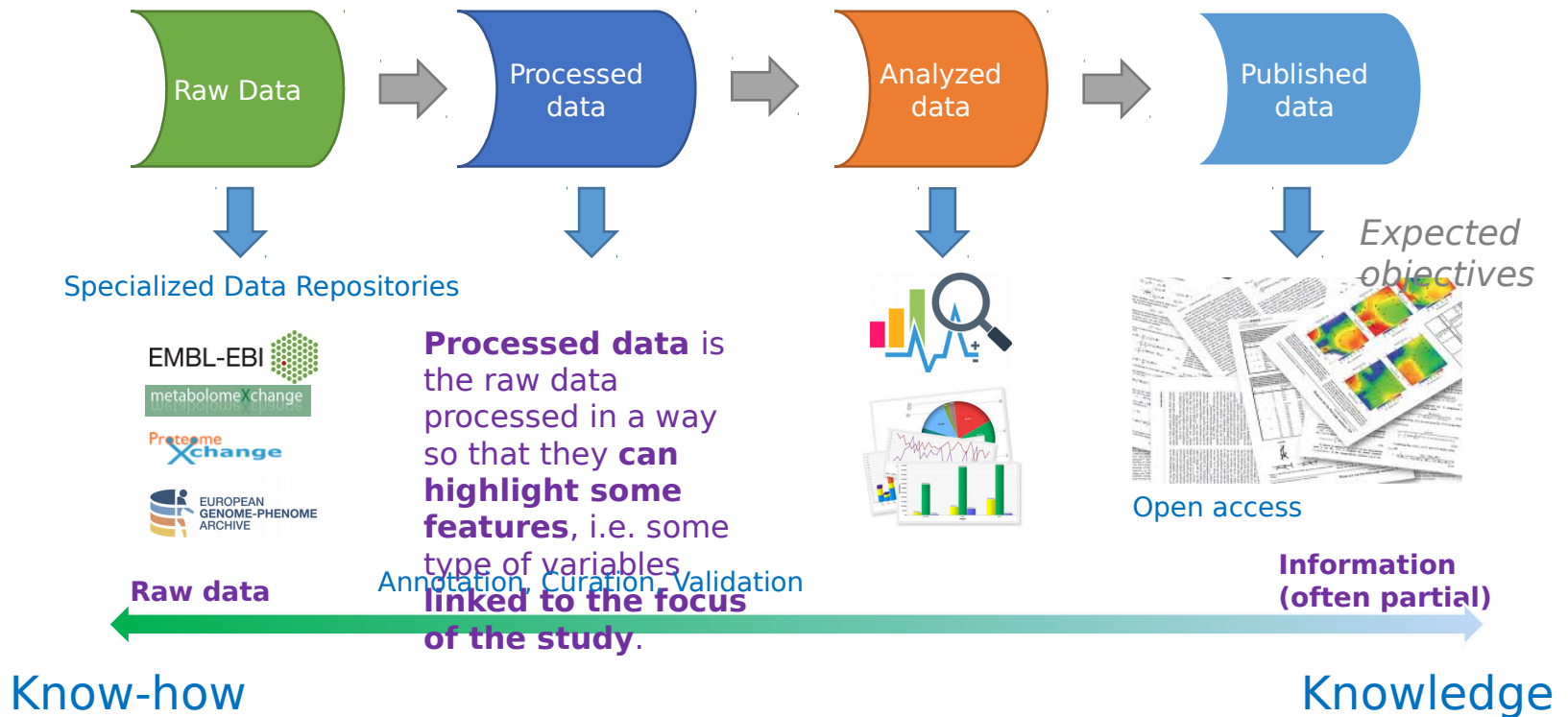


The **implementation** of **FAIR principles** is a process that must be thought of in a **progressive and community-oriented way**.

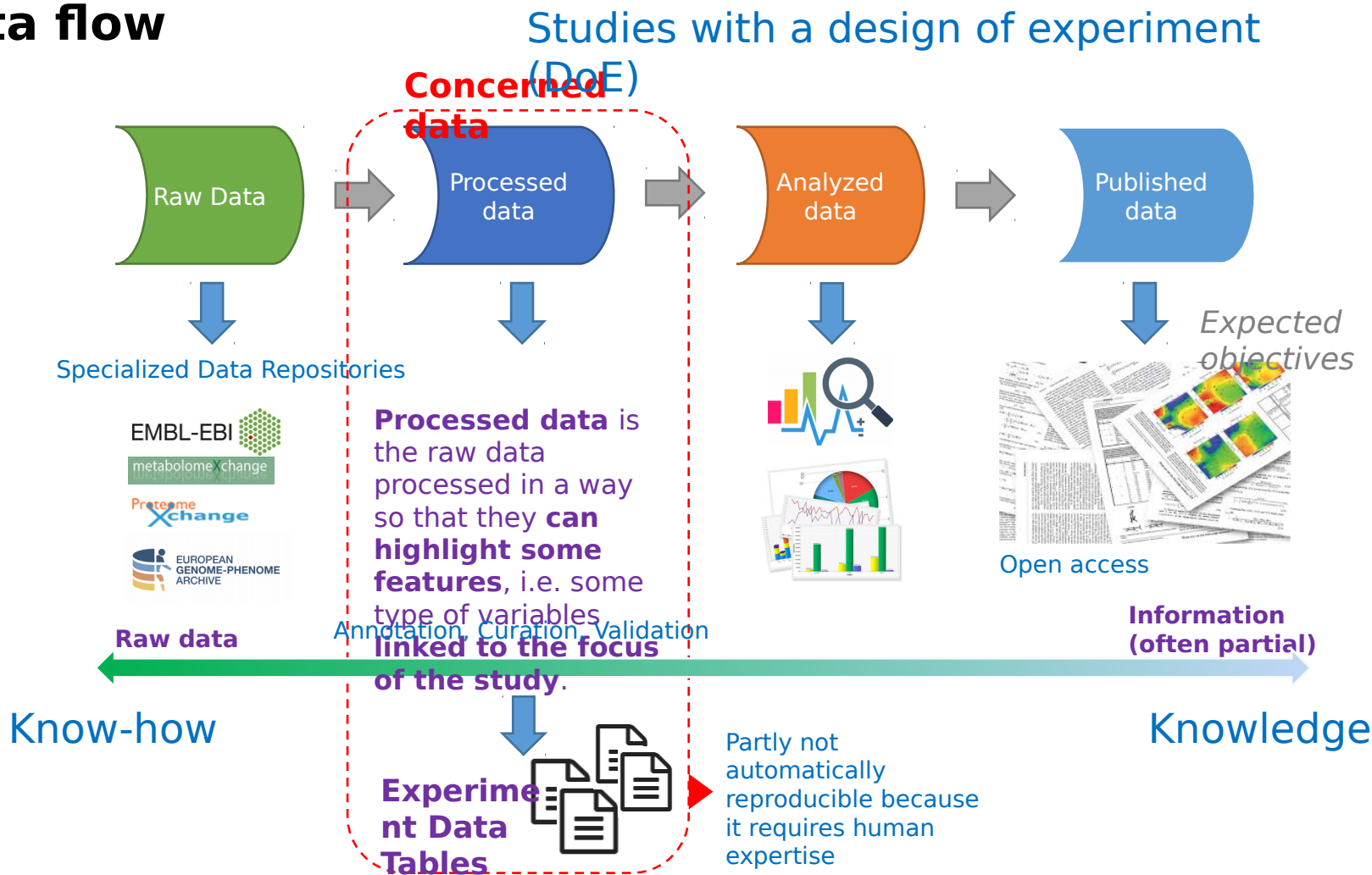
It **must be integrated into existing practices** to ensure that they evolve without interruption and in a way that is **acceptable to the various actors**

Data flow

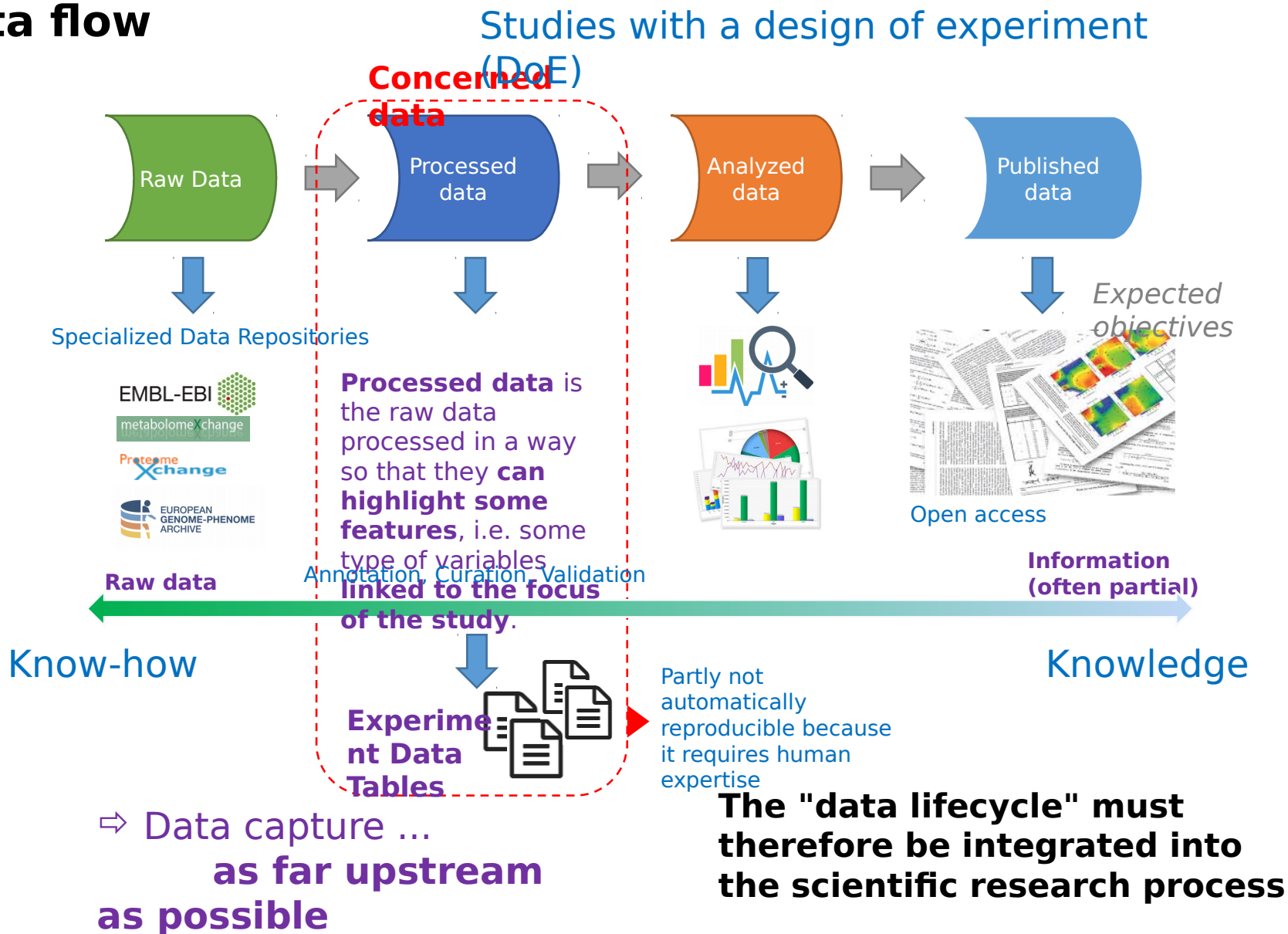
Studies with a design of experiment (DoE)



Data flow



Data flow



Data capture... **as far upstream as possible**

Take into account users [operating methods](#) and [work habits](#)

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

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 tables,
- 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).

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

Provide services



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 :

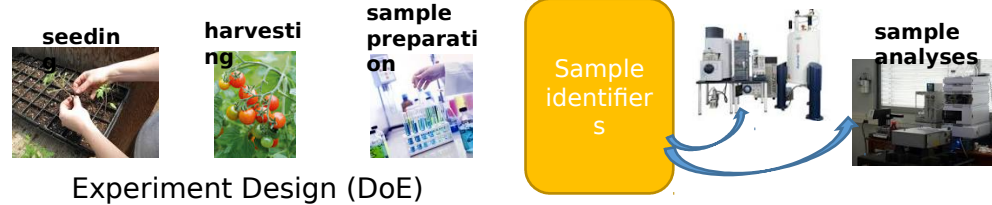
- selection of subset of data then many repetitions of complex processing operations
- according to a very varied parametrization (scenarios).

Allow users to gain efficiency where they would like to gain efficiency
Handling of all these tasks related to data management

Promote good practices

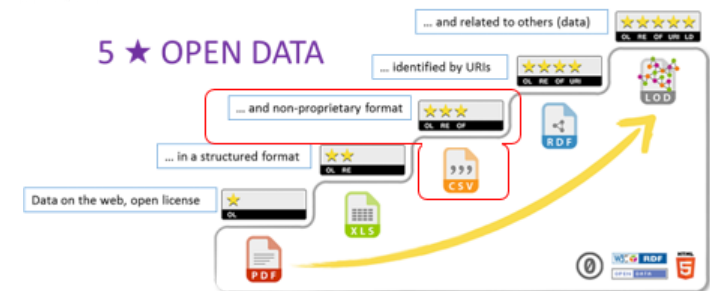
Provide services

Promote good practices



Samples : Sample features

A	B	C	D	E	F	G	H	I	J
SampleID	Treatment	DevStage	FruitAge	FruitPosition	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
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

Promote good practices



seedling



harvesting



sample preparation



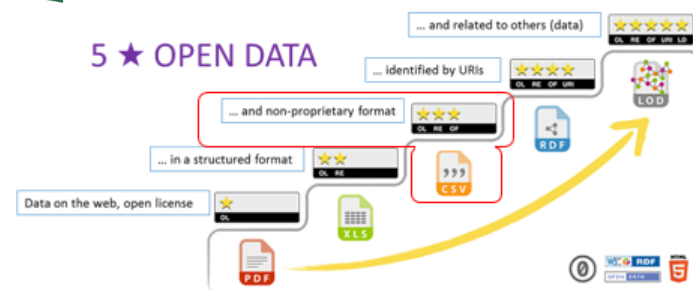
Experiment Design (DoE)

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
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



5 ★ OPEN DATA



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

Promote non-proprietary format like CSV or TSV

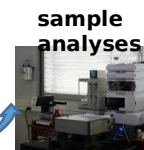
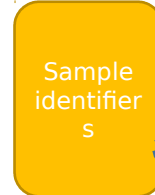


Data

Well organized data

- Each variable forms a column
- Each observation forms a line
- Each "observational entity" forms a table

Promote good practices



Experiment Design (DoE)

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.01	22DPA	3	36.13	31.77	21.43	A	T6
164	Control	FR.01	42DPA	2	51.09	46.85	64.05	A	T5
252	Control	FR.04	55DPA	5	48.28	43.35	66.64	A	T5
353	Control	FF.04	55DPA	3	49.84	44.93	66.98	A	T5
452	Control	FF.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	3	28.6	25.0	2.86	A	T5
109	Control	FF.03	22DPA	3	3.73	3.73	3.73	A	T5
134	Control	FF.02	15DPA	3	27.07	23.0	3.88	A	T7
31	Control	FF.01	08DPA	4 NA	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	40.04	A	T7
425	Control	FR.04	55DPA	2	62.74	50.27	1	A	T7
520	Control	FF.03	30DPA	NA	48.86	41.52	5	A	T7
419	Control	FR.03	50DPA	2	55.63	48.02	86.73	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

Quantitatives

Qualitatives

Data

Well organized data

- Each variable forms a column
- Each observation forms a line
- Each "observational entity" forms a table



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, animal, bacteria, tissues, ...)

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**.

Promote non-proprietary format like CSV or TSV



Promote good practices



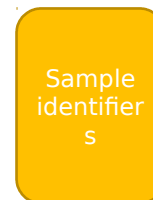
seedling



harvesting



sample preparation



sample analyses

Experiment Design (DoE)

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.01	22DPA	3	36.13	31.77	21.43	A	T6
164	Control	FR.01	42DPA	2	51.09	46.85	64.05	A	T5
252	Control	FR.04	55DPA	5	48.28	43.35	66.64	A	T5
353	Control	FF.04	55DPA	3	49.84	44.93	66.98	A	T5
512	Control	FF.03	21DPA	1					
117	Control			3					
536	Control			NA					
544	Control			NA					
158	Control	FF.04	35DPA						
109	Control	FF.03	22DPA						
134	Control	FF.02	15DPA						
31	Control	FF.01	08DPA						
179	Control	FF.03	28DPA						
383	Control	FF.04	34DPA						
425	Control	FR.04	55DPA						
520	Control	FF.03	30DPA						
419	Control	FR.03	50DPA						
138	Control	FF.02	15DPA						
143	Control	FF.03	29DPA						
365	Control	FR.02	47DPA						
127	Control	FF.03	27DPA						
188	Control	FR.01	42DPA						

Identifier

Factors

Qualitative

Shortname	Description
SampleID	Pool of several harvests
Treatment	Treatment applied on plants
DevStage	fruit development stage
FruitAge	fruit age
FruitDiameter	Fruit diameter
FruitHeight	Fruit height
FruitFW	Fruit Fresh Weight(g)
Rank	Row of the individual plant on the table
Truss	Position on the stem of the truss



Description of the different columns within data files

Identifier
Factor
Factor
Factor
Variable
Variable
Variable
Feature
Feature

categories

- identifier
- factor
- qualitative
- quantitative

Metadata

Data

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

Promote non-proprietary format like CSV or TSV



Promote good practices Minimal but relevant Metadata

Metadata file allowing to associate a key concept to each **data subset file**



	A	B	C	D	E	F	G	H
1	rank	obtainedFrom	subset	identifier	file	description	CV_term_id	CV_term_name
2	1		0:plants	PlantID	plants.csv	Plant features	http://purl.obolibrary.org/obo/PO_0000003	whole plant
3	2		1:harvests	Lot	harvests.csv	Harvest features	http://purl.obolibrary.org/obo/OBI_1110046	organ harvesting
4	3		2:samples	SampleID	samples.csv	Samples features	http://purl.obolibrary.org/obo/PO_0009001	fruit
5	4		3:compounds	SampleID	compounds.csv	Compound quantifications	http://purl.obolibrary.org/obo/CHEBI_24431	chemical entity
6	5		3:enzymes	SampleID	enzymes.csv	Enzyme Features	http://purl.obolibrary.org/obo/OBI_0000427	enzyme

Entity ⇔ Observational entity (e.g. samples, compounds, ...)

Metadata file allowing each **attribute (variable)** to be annotated



	A	B	C	D	E	F	G	H
1	subset	attribute	entry	category	type	description	CV_term_id	CV_term_name
2	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/individual	individual organism identifier
3	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://ncicb.nci.nih.gov/	Row
4	plants	Plant	plant		string	Code identifier of the individual plant	http://ncicb.nci.nih.gov/	Discrete Set Coded String Data Type
5	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://www.ebi.ac.uk/ef/	environmental factor
6	plants	Genotype		qualitative	string	Genotype		
7	harvests	Lot	lot	identifier	numeric	Pool of several harvests	http://www.ebi.ac.uk/ef/	sample pooling
8	harvests	PlantID			numeric	Plant identifier	http://purl.obolibrary.org/obo/individual	individual organism identifier
9	harvests	Truss		qualitative	string	Position on the stem of the truss	http://purl.obolibrary.org/obo/stem	stem node
10	harvests	HarvestDate			string	Harvest date		
11	harvests	HarvestHour			string	Harvest hour		
12	harvests	FruitAge	age	factor	string	fruit development stage	http://purl.obolibrary.org/obo/fruit	fruit development stage
13	harvests	FruitPosition		qualitative	numeric	Position on the truss of the fruit	http://ncicb.nci.nih.gov/	Position Number
14	harvests	FruitDiameter		quantitative	numeric	Fruit diameter (mm)	http://ncicb.nci.nih.gov/	Diameter
15	harvests	FruitHeight		quantitative	numeric	Fruit height (mm)	http://ncicb.nci.nih.gov/	Height
16	harvests	FruitFW		quantitative	numeric	Fruit Fresh Weight(g)	http://ncicb.nci.nih.gov/	Weight
17	samples	SampleID	sampleid	identifier	numeric	Sample identifier	http://purl.obolibrary.org/obo/centrally	centrally registered identifier
18	samples	Lot			numeric	Pool of several harvests	http://www.ebi.ac.uk/ef/	sample pooling
19	samples	NbFruit			numeric	Fruit Number per sample		
20	samples	GellyFW		quantitative	numeric	Gelly Fred Weight(g) per sample		
21	samples	GellyFruit		quantitative	numeric	Gelly per Fruit (estimated g)		
22	samples	BER			string	BER		
23	compounds	SampleID	sampleid	identifier	numeric	Sample identifier	http://purl.obolibrary.org/obo/centrally	centrally registered identifier
24	compounds	DPA		factor	numeric	Day Per Anthesis		
25	compounds	MassBefore		quantitative	numeric	m av.extraction (g)		
26	compounds	MassMIA		quantitative	numeric	masse MIA (g)		
27	compounds	RDT		quantitative	numeric	Rdt (% MIA/DW)		
28	compounds	Starch1		quantitative	numeric	Dosage amidon (%poids/MIA)	http://purl.obolibrary.org/obo/starch	starch
29	compounds	Starch2		quantitative	numeric	amidon (g/gDW)	http://purl.obolibrary.org/obo/starch	starch
30	compounds	RHAMNOSE		quantitative	numeric	RHAMNOSE	http://purl.obolibrary.org/obo/rhamnose	rhamnose

categories

- identifier
- factor
- qualitative
- quantitative

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

Promote good practices Minimal but relevant Metadata

Metadata file allowing to associate a key concept to each **data subset file**



rank	obtainedFrom	subset	identifier	file	description	CV_term_id	CV_term_name
1	1	0:plants	PlantID	plants.csv	Plant features	http://purl.obolibrary.org/obo/PO_0000003	whole plant
2	2	1:harvests	Lot	harvests.csv	Harvest features	http://purl.obolibrary.org/obo/PO_0000004	organ harvesting
3	3	2:samples	SampleID	samples.csv	Samples features	http://purl.obolibrary.org/obo/PO_0000005	fruit
4	4	3:compounds	SampleID	compounds.csv	Compound quantifications	http://purl.obolibrary.org/obo/PO_0000006	chemical entity
5	5	3:enzymes	SampleID	enzymes.csv	Enzyme Features	http://purl.obolibrary.org/obo/PO_0000007	enzyme

Optional

Entity ⇔ Observational entity (e.g. samples, compounds, ...)

Metadata file allowing each **attribute (variable)** to be annotated



subset	attribute	entry	category	type	description	CV_term_id	CV_term_name
plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_0000003	individual organism identifier
plants	Row	row	qualitative	string	Row of the individual plant on the table	http://ncicb.nci.nih.gov/	Row
plants	Plant	plant	qualitative	string	Code identifier of the individual plant	http://ncicb.nci.nih.gov/	Discrete Set Coded String Data Type
plants	Treatment	treatment	factor	string	Treatment applied on plants	http://www.ebi.ac.uk/ef/	environmental factor
plants	Genotype		qualitative	string	Genotype		
harvests	Lot	lot	identifier	numeric	Pool of several harvests	http://www.ebi.ac.uk/ef/	sample pooling
harvests	PlantID		qualitative	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_0000003	individual organism identifier
harvests	Truss		qualitative	string	Position on the stem of the truss	http://purl.obolibrary.org/obo/PO_0000004	stem node
harvests	HarvestDate		qualitative	string	Harvest date		
harvests	HarvestHour		qualitative	string	Harvest hour		
harvests	FruitAge	age	factor	string	fruit development stage	http://purl.obolibrary.org/obo/PO_0000005	fruit development stage
harvests	FruitPosition		qualitative	numeric	Position on the truss of the fruit	http://ncicb.nci.nih.gov/	
harvests	FruitDiameter		quantitative	numeric	Fruit diameter (mm)	http://ncicb.nci.nih.gov/	
harvests	FruitHeight		quantitative	numeric	Fruit height (mm)	http://ncicb.nci.nih.gov/	
harvests	FruitFW		quantitative	numeric	Fruit Fresh Weight(g)	http://ncicb.nci.nih.gov/	Weight
samples	SampleID	sampleid	identifier	numeric	Sample identifier	http://purl.obolibrary.org/obo/PO_0000006	centrally registered identifier
samples	Lot		qualitative	numeric	Pool of several harvests	http://www.ebi.ac.uk/ef/	sample pooling
samples	NbFruit		quantitative	numeric	Fruit Number per sample		
samples	GellyFW		quantitative	numeric	Gelly Fred Weight(g) per sample		
samples	GellyFruit		quantitative	numeric	Gelly per Fruit (estimated g)		
samples	BER		quantitative	string	BER		
compounds	SampleID	sampleid	identifier	numeric	Sample identifier	http://purl.obolibrary.org/obo/PO_0000006	centrally registered identifier
compounds	DPA		factor	numeric	Day Per Anthesis		
compounds	MassBefore		quantitative	numeric	m av.extraction (g)		
compounds	MassMIA		quantitative	numeric	masse MIA (g)		
compounds	RDT		quantitative	numeric	Rdt (% MIA/DW)		
compounds	Starch1		quantitative	numeric	Dosage amidon (%poids/MIA)	http://purl.obolibrary.org/obo/PO_0000007	starch
compounds	Starch2		quantitative	numeric	amidon (g/gDW)	http://purl.obolibrary.org/obo/PO_0000008	starch
compounds	RHAMNOSE		quantitative	numeric	RHAMNOSE	http://purl.obolibrary.org/obo/PO_0000009	rhamnose

Optional

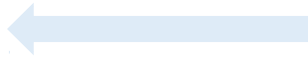
categories

- identifier
- factor
- qualitative
- quantitative

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

FAIR principles, a new opportunity to improve the data lifecycle

Provide
services



Promote
good
practices

Experiment
Data
Tables

+
Metadata

plants.tsv

harvests.tsv

compounds.tsv

enzymes.tsv

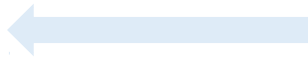
*Additional data subsets
can be added step by
step, as soon as data are
produced.*

ODAM Framework

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

FAIR principles, a new opportunity to improve the data lifecycle

Provide
services



Promote
good
practices

Experiment
Data
Tables

+
Metadata

plants.tsv	harvests.tsv	compounds.tsv	enzymes.tsv
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

Drag & Drop

PUT



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

ODAM Framework

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

FAIR principles, a new opportunity to improve the data lifecycle

Provide
services

Promote
good
practices

ODAM Framework

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

Experiment
Data
Tables

Metadata

The screenshot shows a data table with columns for experiment data, metadata, and compounds. The table is divided into sections: 'plants.tsv', 'harvests.tsv', 'Metadata', 'compounds.tsv', and 'enzymes.tsv'. The 'plants.tsv' section contains columns for plant ID, name, and location. The 'harvests.tsv' section contains columns for harvest date, time, and location. The 'Metadata' section contains columns for metadata ID, name, and location. The 'compounds.tsv' section contains columns for compound ID, name, and location. The 'enzymes.tsv' section contains columns for enzyme ID, name, and location.

Drag & Drop

PUT

Storage and
sharing with
all partners

Multiscale
deployment

/ local / Intranet /
Internet



vm / docker



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

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

GET

Software layer
transparent from
the user's point of
view



Data can be
downloaded,
explored and
mined

FAIR principles, a new opportunity to improve the data lifecycle

Provide services

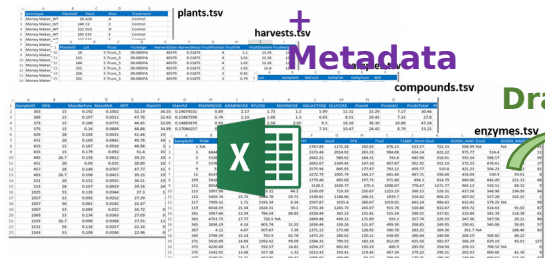
Promote good practices

ODAM Framework

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

Experiment Data Tables

Metadata



The screenshot shows a data table with columns for plants, harvests, compounds, and enzymes. The table is divided into sections: plants.tsv, harvests.tsv, compounds.tsv, and enzymes.tsv. The plants section lists various plant species and their associated harvests. The compounds section lists chemical compounds and their associated enzymes. The enzymes section lists enzymes and their associated compounds.

Drag & Drop

PUT

Storage and sharing with all partners

Multiscale deployment

Local / Intranet / Internet



VM / docker

GET

F
A
I
R
INTEROPERABLE

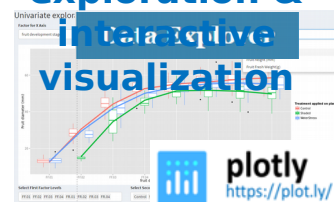
Additional data subsets can be added step by step, as soon as data are produced.

Merging & selection of data subsets



Export

Multi-criteria exploration & interactive visualization



Data can be downloaded, explored and mined

FAIR principles, a new opportunity to improve the data lifecycle

Provide services

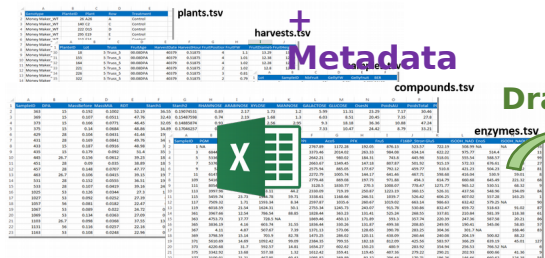
Promote good practices

ODAM Framework

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

Experiment Data Tables

Metadata



The screenshot shows a data table with columns for plants, harvests, compounds, and enzymes. The table is divided into sections: plants.tsv, harvests.tsv, compounds.tsv, and enzymes.tsv. The plants section lists various plant species and their associated harvests. The compounds section lists chemical compounds and their associated enzymes. The enzymes section lists enzymes and their associated compounds.

Drag & Drop

PUT

Storage and sharing with all partners

Multiscale deployment

local / Intranet / Internet

VM / docker

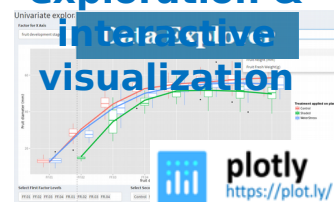


F
A
I
R
INTEROPERABLE

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

GET

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



CRAN 0.1.4



API

Develop if needed, lightweight tools

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

FAIR principles, a new opportunity to improve the data lifecycle

Provide services

Promote good practices

ODAM Framework

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

Experiment Data Tables

Metadata

The screenshot shows a data table with columns: plants, harvests, compounds, and enzymes. The table contains multiple rows of data, including plant names, harvest dates, compound IDs, and enzyme names.

Drag & Drop

PUT

Storage and sharing with all partners

Multiscale deployment

local / Intranet / Internet



VM / docker



F
A
I
R
INTEROPERABLE

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

GET



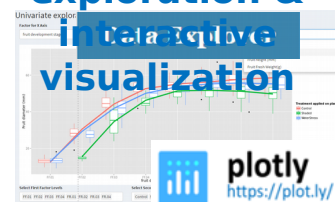
Data can be downloaded, explored and mined

Merging & selection of data subsets



Export

Multi-criteria exploration & interactive visualization



Link

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

Repetition of complex treatments according to very varied parameters



CRAN 0.1.4



API facilitate the subsequent data publication

FINDABLE
ACCESIBLE
I
RE-USABLE

Develop if needed, lightweight tools

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

FAIR principles, a new opportunity to improve the data lifecycle

Promote good practices

Minimal effort

ODAM Framework

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

Experiment Data Tables

Metadata

plants.tsv

id	name	year	location	harvest	compounds	enzymes
1	Arabidopsis thaliana	2010	France	1000	1000	1000
2	Arabidopsis thaliana	2011	France	1000	1000	1000
3	Arabidopsis thaliana	2012	France	1000	1000	1000
4	Arabidopsis thaliana	2013	France	1000	1000	1000
5	Arabidopsis thaliana	2014	France	1000	1000	1000

harvests.tsv

id	year	location	harvest
1	2010	France	1000
2	2011	France	1000
3	2012	France	1000
4	2013	France	1000
5	2014	France	1000

compounds.tsv

id	name	year	location	harvest	compounds	enzymes
1	Arabidopsis thaliana	2010	France	1000	1000	1000
2	Arabidopsis thaliana	2011	France	1000	1000	1000
3	Arabidopsis thaliana	2012	France	1000	1000	1000
4	Arabidopsis thaliana	2013	France	1000	1000	1000
5	Arabidopsis thaliana	2014	France	1000	1000	1000

enzymes.tsv

id	name	year	location	harvest	compounds	enzymes
1	Arabidopsis thaliana	2010	France	1000	1000	1000
2	Arabidopsis thaliana	2011	France	1000	1000	1000
3	Arabidopsis thaliana	2012	France	1000	1000	1000
4	Arabidopsis thaliana	2013	France	1000	1000	1000
5	Arabidopsis thaliana	2014	France	1000	1000	1000

Drag & Drop

PUT

Storage and sharing with all partners

Multiscale deployment

local / Intranet / Internet

vm / docker



F
A
I
NTEROPERABLE
R

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

GET

Multi-criteria exploration & interactive visualization

Data can be downloaded, explored and mined

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

Link

Merging & selection of data subsets

Export

Repetition of complex treatments according to very varied parameters



CRAN 0.1.4



API

facilitate the subsequent data publication

F
I
N
DABLE
A
C
C
E
S
S
I
B
L
E
R
E
-
U
S
A
B
L
E

Provide

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

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

Data Backup

Storage and sharing with all partners
Multiscale deployment
local / Intranet / Internet

Drag & Drop

PUT



vm / docker

GET

F
A
I
R
INTEROPERABLE

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

Data Curation

Remove inconsistent values

Merging & selection of data subsets



Export

Multi-criteria exploration & interactive visualization



R CRAN 0.1.4



API

facilitate the subsequent data publication



Data can be downloaded, explored and mined

The
Dataverse
Project

Link

FINDABLE
ACCESIBLE
I
RE-USABLE

Data Extension

Repetition of complex treatments according to very varied parameters



Develop if needed, lightweight tools

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

Data Publication

<https://data.inra.fr>

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

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

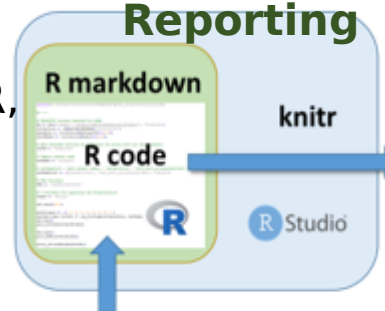
subsets

Repetition of complex

treatments according to

very varied parameters

Data Analysis Reporting



FRIMOUS data analysis interfaced by ODAM

Daniel Jacob
2018-10-05

Contents

Order documentation	3
Initiation	3
Plot variables	4
Apple - metadatas	4
Apple - cell, wall, polysaccharide	5
Apple - lipids	6
Chenopodium - metadatas	6
Chenopodium - cell, wall, polysaccharide	6
Chenopodium - lipids	6
Cucurbit - metadatas	6
Cucurbit - cell, wall, polysaccharide	6
Cucurbit - lipids	6
Eggplant - metadatas	6
Eggplant - cell, wall, polysaccharide	6
Eggplant - lipids	6
Kivi - metadatas	6
Kivi - cell, wall, polysaccharide	6
Kivi - lipids	6
Grapevine - metadatas	6
Grapevine - cell, wall, polysaccharide	6
Grapevine - lipids	6
Peach - metadatas	6
Peach - cell, wall, polysaccharide	6
Peach - lipids	6
Paper - metadatas	6
Paper - cell, wall, polysaccharide	6
Paper - lipids	6
ODAM - metadatas	6
Growth Modulation	6

Growth Modulation

Models:

- Model 1: Weight = (Time-dependent (x Day After Anthesis))

Model 1 - Single Sigmoidal Model

- Model 1: $W(t) = a + b \cdot \exp(-c \cdot (t - t_0)^d)$

Model 2 - Sum of two Sigmoidal Models

- Model 2: $W(t) = a + b \cdot \exp(-c \cdot (t - t_0)^d) + e \cdot \exp(-f \cdot (t - t_1)^g)$

Modulation of the 'Weight' curve for each species

T variables for applying the modulation

Time = "Weight"

Choice of model for each dataset (1 for one sigmoid, 0 for double sigmoid)

modelspecs = c(1, 0, 1, 1, 1, 0, 0, 1, 1)

Application of the modulation

res = lapply(modelspecs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

res = lapply(specs, FUN = function(specs, dataset, species, time, modeltype) {

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

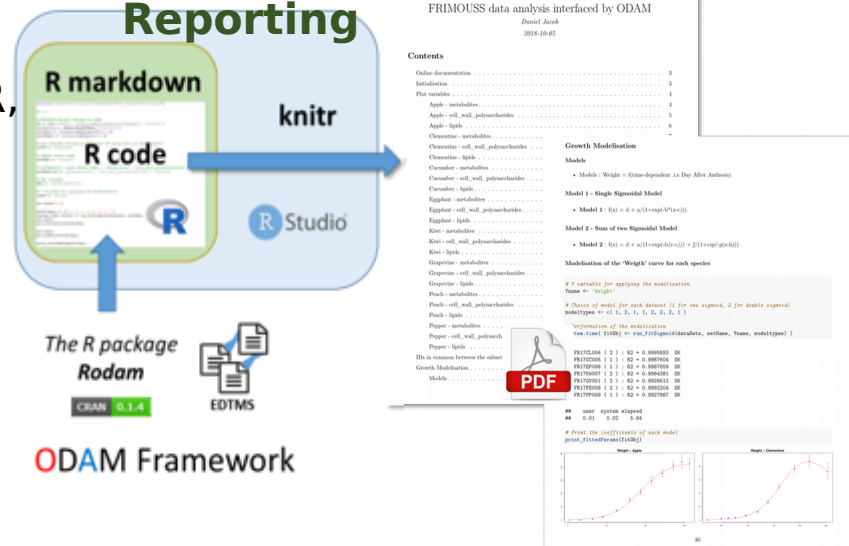
carried out by the ODAM approach

Merging & selection of data

subsets

Repetition of complex treatments according to very varied parameters

Data Analysis Reporting

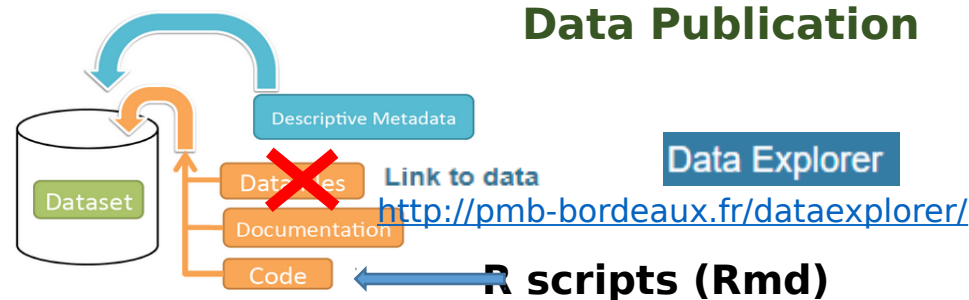


Schematic Diagram of a **Dataset** in Dataverse 4.0

Data Publication



<https://data.inra.fr>



Container for your data, documentation, and code.

FINDABLE
ACCESIBLE
I
RE-
USABLE

Opportunity for the Web Of Data Metadata



Resource Description Framework (RDF)



CV Term

	A	B	C	D	E	F	G	H
	subset	attribute	entry	category	type	description	CV term id	CV term name
Attributes	1	plants	PlantID	plantid	identifier	Plant identifier	http://purl.obolibrary.org/obo/PO_0000003	individual organism identifier
	2	plants	Row	row	qualitative	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_0000004	Row
	3	plants	Plant	plant	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0000005	Discrete Set Coded String Data Type
	4	plants	Treatment	treatment	factor	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_0000006	environmental factor
	5	plants	Genotype	genotype	qualitative	Genotype	http://purl.obolibrary.org/obo/PO_0000007	sample pooling
	6	plants	Lot	lot	identifier	Pool of several harvests	http://purl.obolibrary.org/obo/PO_0000008	individual organism identifier
	7	plants	Truss	truss	qualitative	Position on the stem of the truss	http://purl.obolibrary.org/obo/PO_0000009	stem node
	8	plants	HarvestDate	harvestdate	string	Harvest date	http://purl.obolibrary.org/obo/PO_0000010	Harvest hour
	9	plants	HarvestHour	harvesthour	string	Harvest hour	http://purl.obolibrary.org/obo/PO_0000011	fruit development stage
	10	plants	FruitAge	fruitage	factor	Position on the stem of the fruit	http://purl.obolibrary.org/obo/PO_0000012	Position Number
	11	plants	FruitPosition	fruitposition	qualitative	Fruit development stage	http://purl.obolibrary.org/obo/PO_0000013	Diameter
	12	plants	FruitDiameter	fruitdiameter	quantitative	Fruit diameter (mm)	http://purl.obolibrary.org/obo/PO_0000014	Height
	13	plants	FruitHeight	fruitheight	quantitative	Fruit height (mm)	http://purl.obolibrary.org/obo/PO_0000015	Weight
	14	plants	FruitWeight	fruitweight	quantitative	Fruit fresh weight (g)	http://purl.obolibrary.org/obo/PO_0000016	centrally registered identifier
	15	plants	SampleID	sampleid	identifier	Sample identifier	http://purl.obolibrary.org/obo/PO_0000017	sample pooling
	16	plants	Lot	lot	numeric	Pool of several harvests	http://purl.obolibrary.org/obo/PO_0000018	centrally registered identifier
	17	plants	NbFruit	nbfruit	numeric	Fruit Number per sample	http://purl.obolibrary.org/obo/PO_0000019	centrally registered identifier
	18	plants	GellyFruit	gellyfruit	numeric	Gelly fresh weight (g) per sample	http://purl.obolibrary.org/obo/PO_0000020	centrally registered identifier
	19	plants	GellyFruit	gellyfruit	numeric	Gelly per fruit (estimated g)	http://purl.obolibrary.org/obo/PO_0000021	centrally registered identifier
	20	plants	BER	ber	string	BER	http://purl.obolibrary.org/obo/PO_0000022	centrally registered identifier
	21	plants	SampleID	sampleid	identifier	Sample identifier	http://purl.obolibrary.org/obo/PO_0000023	centrally registered identifier
	22	plants	DPA	dpa	factor	Day Per Anthesis	http://purl.obolibrary.org/obo/PO_0000024	centrally registered identifier
	23	plants	MassBefore	massbefore	numeric	Mass before extraction (g)	http://purl.obolibrary.org/obo/PO_0000025	centrally registered identifier
	24	plants	MassMIA	massmia	numeric	Mass MIA (g)	http://purl.obolibrary.org/obo/PO_0000026	centrally registered identifier
	25	plants	ROT	rot	numeric	Rot (% MIA/DW)	http://purl.obolibrary.org/obo/PO_0000027	centrally registered identifier
	26	plants	Starch1	starch1	numeric	Dosage amideon (Nipoids/MIA)	http://purl.obolibrary.org/obo/PO_0000028	centrally registered identifier
	27	plants	Starch2	starch2	numeric	amidon (g/DW)	http://purl.obolibrary.org/obo/PO_0000029	centrally registered identifier
	28	plants	RHAMNOSE	rhamnose	numeric	RHAMNOSE	http://purl.obolibrary.org/obo/PO_0000030	centrally registered identifier
	29	plants	RHAMNOSE	rhamnose	numeric	RHAMNOSE	http://purl.obolibrary.org/obo/PO_0000031	centrally registered identifier
	30	plants	RHAMNOSE	rhamnose	numeric	RHAMNOSE	http://purl.obolibrary.org/obo/PO_0000032	centrally registered identifier

Optional

Entity ⇔ Observational entity (e.g. samples, compounds, ...)

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

Opportunity for the Web Of Data Metadata



Resource Description Framework (RDF)



CV Term

	A	B	C	D	E	F	G	H
subset	attribute	entry	category	type	description	CV term id	CV term name	
1	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
2	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
3	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
4	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
5	plants	Genotype	genotype	qualitative	string	Genotype		
6	plants	Lot	lot	identifier	numeric	Pool of several harvests		
7	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
8	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
9	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
10	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
11	plants	Genotype	genotype	qualitative	string	Genotype		
12	plants	Lot	lot	identifier	numeric	Pool of several harvests		
13	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
14	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
15	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
16	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
17	plants	Genotype	genotype	qualitative	string	Genotype		
18	plants	Lot	lot	identifier	numeric	Pool of several harvests		
19	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
20	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
21	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
22	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
23	plants	Genotype	genotype	qualitative	string	Genotype		
24	plants	Lot	lot	identifier	numeric	Pool of several harvests		
25	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
26	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
27	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
28	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
29	plants	Genotype	genotype	qualitative	string	Genotype		
30	plants	Lot	lot	identifier	numeric	Pool of several harvests		
31	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
32	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
33	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
34	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
35	plants	Genotype	genotype	qualitative	string	Genotype		
36	plants	Lot	lot	identifier	numeric	Pool of several harvests		
37	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
38	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
39	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
40	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
41	plants	Genotype	genotype	qualitative	string	Genotype		
42	plants	Lot	lot	identifier	numeric	Pool of several harvests		
43	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
44	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
45	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
46	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
47	plants	Genotype	genotype	qualitative	string	Genotype		
48	plants	Lot	lot	identifier	numeric	Pool of several harvests		
49	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
50	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
51	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
52	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
53	plants	Genotype	genotype	qualitative	string	Genotype		
54	plants	Lot	lot	identifier	numeric	Pool of several harvests		
55	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
56	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
57	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
58	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
59	plants	Genotype	genotype	qualitative	string	Genotype		
60	plants	Lot	lot	identifier	numeric	Pool of several harvests		
61	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
62	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
63	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
64	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
65	plants	Genotype	genotype	qualitative	string	Genotype		
66	plants	Lot	lot	identifier	numeric	Pool of several harvests		
67	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
68	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
69	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
70	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
71	plants	Genotype	genotype	qualitative	string	Genotype		
72	plants	Lot	lot	identifier	numeric	Pool of several harvests		
73	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
74	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
75	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
76	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
77	plants	Genotype	genotype	qualitative	string	Genotype		
78	plants	Lot	lot	identifier	numeric	Pool of several harvests		
79	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
80	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
81	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
82	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
83	plants	Genotype	genotype	qualitative	string	Genotype		
84	plants	Lot	lot	identifier	numeric	Pool of several harvests		
85	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
86	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
87	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
88	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
89	plants	Genotype	genotype	qualitative	string	Genotype		
90	plants	Lot	lot	identifier	numeric	Pool of several harvests		
91	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
92	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
93	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
94	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme
95	plants	Genotype	genotype	qualitative	string	Genotype		
96	plants	Lot	lot	identifier	numeric	Pool of several harvests		
97	plants	PlantID	plantid	identifier	numeric	Plant identifier	http://purl.obolibrary.org/obo/PO_000003	whole plant
98	plants	Row	row	qualitative	string	Row of the individual plant on the table	http://purl.obolibrary.org/obo/PO_1110046	organ harvesting
99	plants	Plant	plant	string	string	Code identifier of the individual plant	http://purl.obolibrary.org/obo/PO_0009001	fruit
100	plants	Treatment	treatment	factor	string	Treatment applied on plants	http://purl.obolibrary.org/obo/PO_000427	enzyme

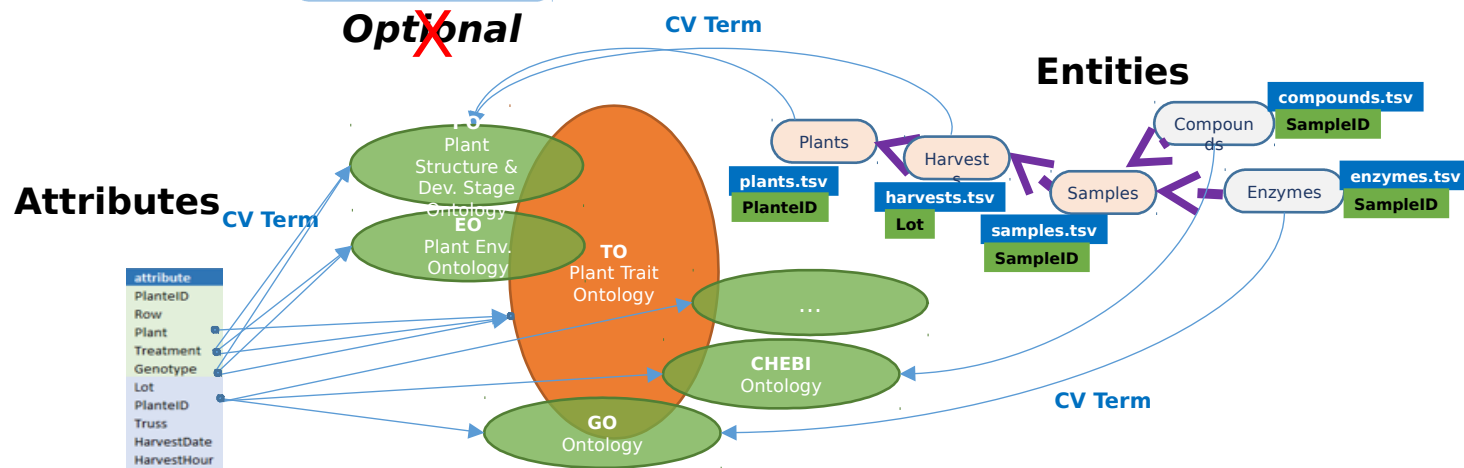
Reference ontologies

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



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

~~Optional~~



Entity \Leftrightarrow Observational entity (e.g. samples, compounds, ...)

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

Opportunity for the Web Of Data



Entities



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

Search for a particular

Trait

Entity + Attribute
= **Trait**



100

FINDABLE
ACCESIBLE
INTEROPERABLE
RE-USABLE

FAIR principles, a new opportunity to improve the data lifecycle

Opportunity for the Web Of Data Metadata

		Category			CV Term		
		Attribute	Entity	Category	CV Term	CV Term ID	CV Term Name
Attributes	1 subset	PlantID	plantid	identifier	numeric	http://purl.obolibrary.org/obo/PO_000003	whole plant
	2 plants	Plant	plant	qualitative	string	http://purl.obolibrary.org/obo/PO_000004	organ harvesting
	3 plants	Plant	plant	factor	string	http://purl.obolibrary.org/obo/PO_000005	fruit
	4 plants	Treatment	treatment	factor	string	http://purl.obolibrary.org/obo/PO_000006	chemical entity
	5 plants	Genotype	genotype	qualitative	string	http://purl.obolibrary.org/obo/PO_000007	chemical entity
	6 plants	Lot	lot	factor	string	http://purl.obolibrary.org/obo/PO_000008	chemical entity
	7 harvests	PlantID	plantid	identifier	numeric	http://purl.obolibrary.org/obo/PO_000009	whole plant
	8 harvests	Truss	truss	qualitative	string	http://purl.obolibrary.org/obo/PO_000010	organ harvesting
	9 harvests	HarvestDate	harvestdate	factor	string	http://purl.obolibrary.org/obo/PO_000011	fruit
	10 harvests	HarvestHour	harvesthour	factor	string	http://purl.obolibrary.org/obo/PO_000012	chemical entity
	11 harvests	FruitAge	fruitage	factor	string	http://purl.obolibrary.org/obo/PO_000013	chemical entity
	12 harvests	FruitPosition	fruitposition	factor	string	http://purl.obolibrary.org/obo/PO_000014	chemical entity
	13 harvests	FruitDiameter	fruitdiameter	factor	string	http://purl.obolibrary.org/obo/PO_000015	chemical entity
	14 harvests	FruitHeight	fruitheight	factor	string	http://purl.obolibrary.org/obo/PO_000016	chemical entity
	15 harvests	FruitWeight	fruitweight	factor	string	http://purl.obolibrary.org/obo/PO_000017	chemical entity
	16 harvests	SampleID	sampleid	identifier	numeric	http://purl.obolibrary.org/obo/PO_000018	whole plant
	17 samples	Lot	lot	factor	string	http://purl.obolibrary.org/obo/PO_000019	organ harvesting
	18 samples	NbFruit	nbfruit	factor	string	http://purl.obolibrary.org/obo/PO_000020	fruit
	19 samples	GellyFruit	gellyfruit	factor	string	http://purl.obolibrary.org/obo/PO_000021	chemical entity
	20 samples	GellyFruit	gellyfruit	factor	string	http://purl.obolibrary.org/obo/PO_000022	chemical entity
	21 samples	BER	ber	factor	string	http://purl.obolibrary.org/obo/PO_000023	chemical entity
	22 samples	SampleID	sampleid	identifier	numeric	http://purl.obolibrary.org/obo/PO_000024	whole plant
	23 compounds	DPA	dpa	factor	string	http://purl.obolibrary.org/obo/PO_000025	organ harvesting
	24 compounds	MassBefore	massbefore	factor	string	http://purl.obolibrary.org/obo/PO_000026	fruit
	25 compounds	MassMIA	massmia	factor	string	http://purl.obolibrary.org/obo/PO_000027	chemical entity
	26 compounds	ROT	rot	factor	string	http://purl.obolibrary.org/obo/PO_000028	chemical entity
	27 compounds	Starch	starch	factor	string	http://purl.obolibrary.org/obo/PO_000029	chemical entity
	28 compounds	Starch	starch	factor	string	http://purl.obolibrary.org/obo/PO_000030	chemical entity
	29 compounds	RHAMNOSE	rhamnose	factor	string	http://purl.obolibrary.org/obo/PO_000031	chemical entity
	30 compounds	RHAMNOSE	rhamnose	factor	string	http://purl.obolibrary.org/obo/PO_000032	chemical entity
Entities	1 link	PlantID	plantid	identifier	numeric	http://purl.obolibrary.org/obo/PO_000003	whole plant
	2 plants	Plant	plant	qualitative	string	http://purl.obolibrary.org/obo/PO_000004	organ harvesting
	3 samples	SampleID	sampleid	identifier	numeric	http://purl.obolibrary.org/obo/PO_000018	whole plant
	4 enzymes	Enzyme	enzyme	factor	string	http://purl.obolibrary.org/obo/PO_000042	chemical entity

Other Metadata

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

Linked Data



Resource Description Framework (RDF)



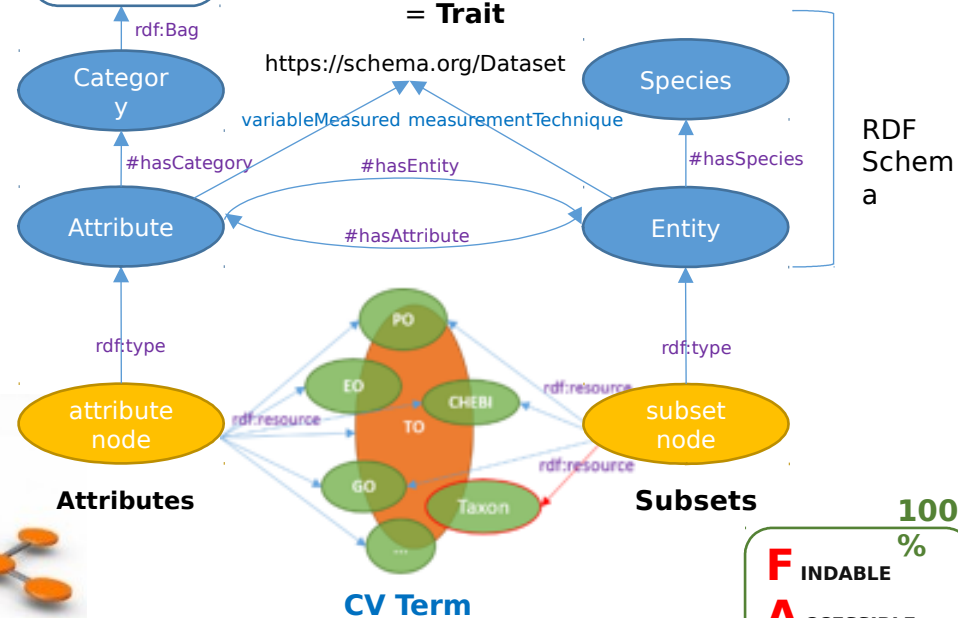
Application ontologies

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

categories
☐ identifier
☐ factor
☐ qualitative
☐ quantitative

Typical queries:

Search for a particular
Trait
 Entity + Attribute
 = **Trait**



RDF
Schem
a

100%
FINDABLE
ACCESIBLE
INTEROPERABLE
RE-USABLE

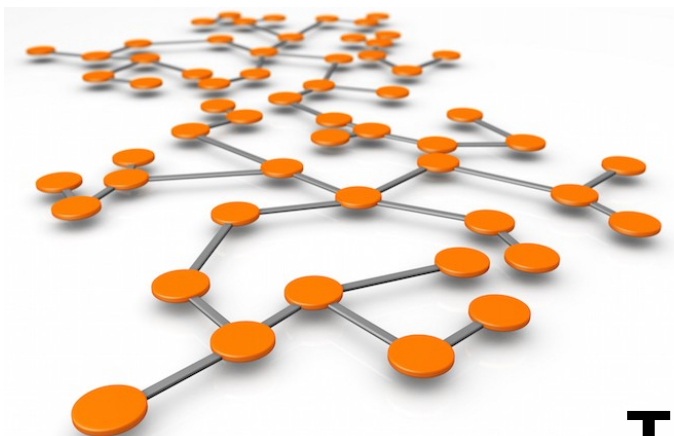
FAIR principles

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

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



Thank you for your attention

