

Data management based on Metadata





Data Management Plan

How does the management of data is it funded, especially in the long term?

Resources

What does the project consist of?
Who are the partners?
What policy on data management?
Who is responsible for the management of data?

Responsibilities in the project

What data will be produced/used during the course of the project (type, format, volume and increase...) ?
How will they be produced?

Data collection

Who will be the owner of the data produced?
How will they be used?

Intellectual Property

How, where, by whom, will be stored, backed up and secured the data?

Data backup

Who will be able to access the data? The data will they be shared? published? With whom? How? How long does it take? Under which license?

Data Access and Data sharing

How will the data be identified, described? What metadata standards will be used? How will the metadata be generated?

Data Documentation

What is the plan for long-term archiving and preservation?

Data Archiving

Planning must be followed by implementation and therefore concrete actions



Data management



Being able to easily describe your data (descriptive metadata)

- with its professional vocabulary
- without tedious entries
- without having to re-enter the same information each time
- by associating external resources (links)



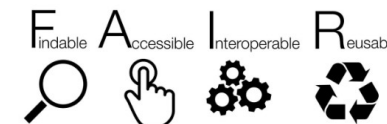
Being able to easily manage your data

- by limiting data loss (after the departure of temporary staff)
- by sharing only metadata
- be able to easily find data (from metadata)
- be able to provide access to data if necessary
- be able to distribute them without having to re-enter everything



Ensuring metadata follows FAIR principles

- Respect a standard (metadata schema)
- Use controlled vocabulary consistent with your domain (thesaurus, ontologies)
- Be at least “Findable, Accessible & Interoperable”





Data management

The Ariadne's thread ...

Organize

Your workspace

(storage, backup, naming, ...)



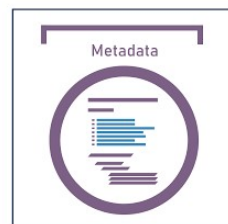
Share & Search data/metadata

(metadata)



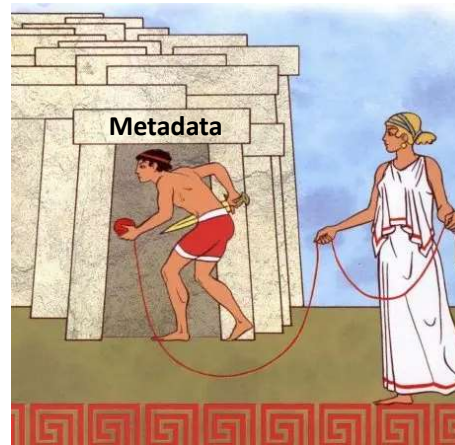
Publish your data

(with metadata)



Describe your data

(metadata)

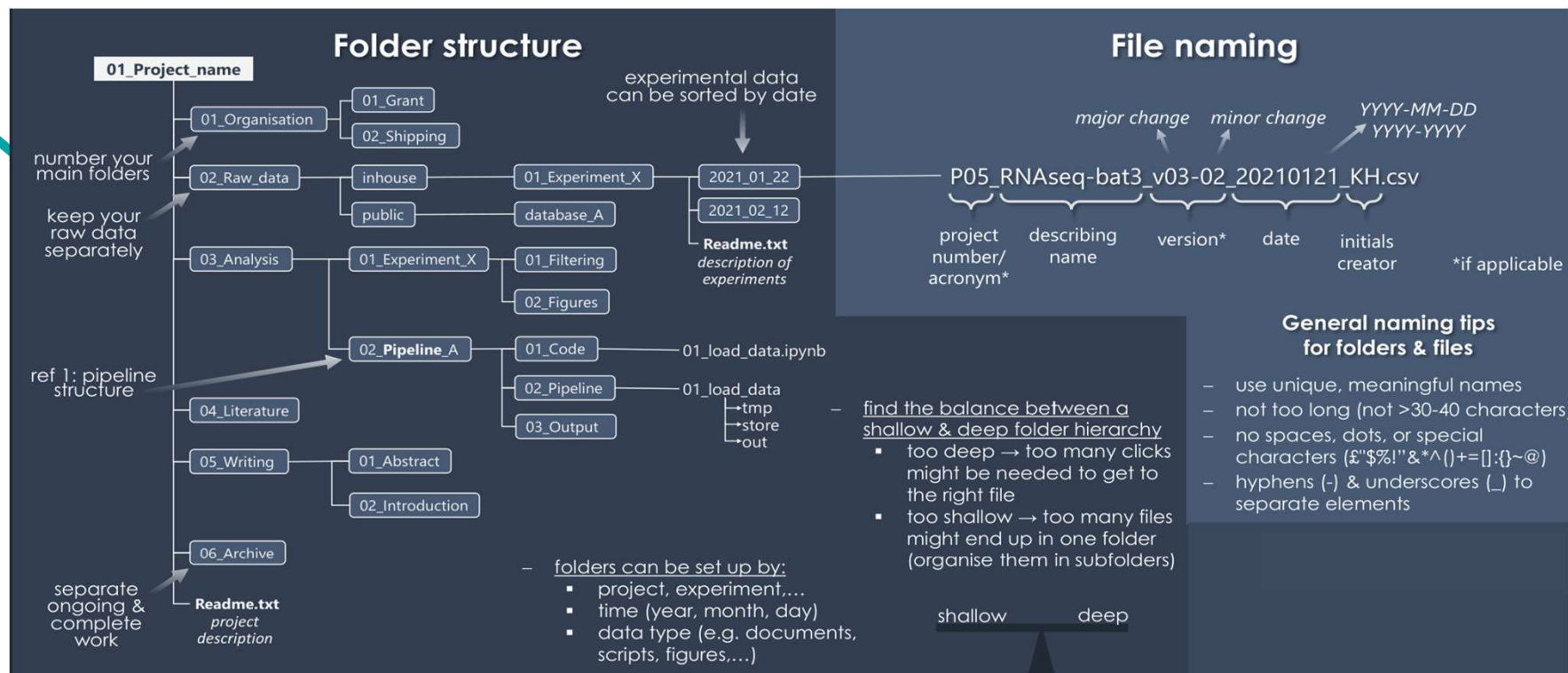




Data management

The Ariadne's thread ...

Organize
Your workspace
(storage, backup,
naming, ...)





Data management

The Ariadne's thread ...

Organize

Your workspace

(storage, backup, naming, ...)



Share & Search (meta)data

(metadata)



Publish your data

(with metadata)



Describe your data
(metadata)

readme.txt files

can be used to describe
projects, folders, and files

What to put in it ?

How to use them effectively when we want to find information ?

With what vocabulary ?

Metadata : descriptive data about the data, which
provide the essential contextual information for
interpreting and reusing the data.



Data management

The Ariadne's thread ...

Organize

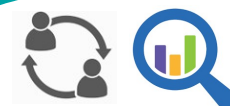
Your workspace

(storage, backup, naming, ...)



Share & Search (meta)data

(metadata)

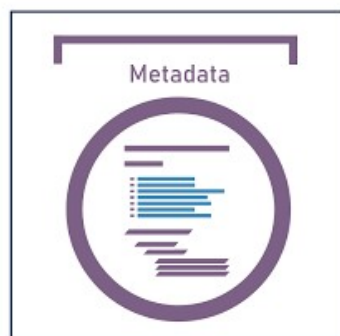


Publish your data

(with metadata)



Describe your data
(metadata)



Metadata Schema

(what)

+



Controlled Vocabulary

(with what)

How ?



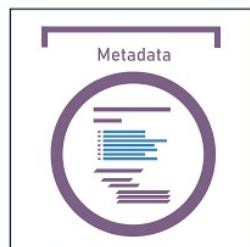
Metadata creation (1/4)

1 *Metadata schema (standard)* ⇔ Template / Metadata model

(container)



Describe your data
(metadata)

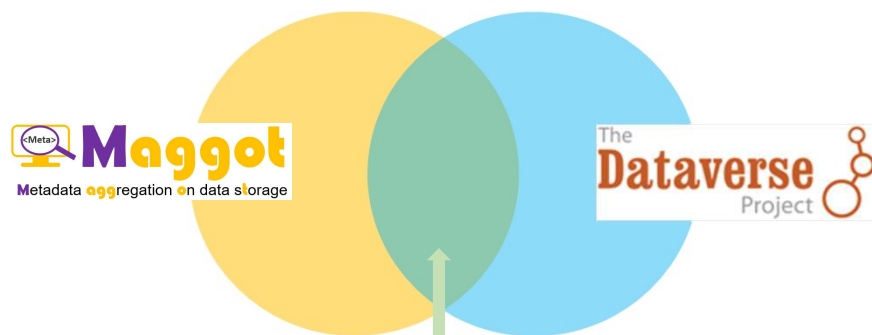


What metadata ?

What information ?



Choose a *schema* common to « Dataverse »



Metadata schema (standard)



Data Document Initiative

DEFINITION
MANAGEMENT
STATUS
DESCRIPTORS
OTHER

Common Metadata

title, description, alternativeURL, contacts, authors, collectors, curators, members, depositor, leader, subject, keywords, topics, kindOfData, dataOrigin, lifeCycleStep, publication, otherReferences, grantNumbers, project, ...

Life Sciences Metadata

Geospatial Metadata

Journal Metadata

Semantic resource

...

Specific Profile Metadata

...



Citation Metadata	
Common Metadata	
Dataset Persistent ID	doi:10.82233/FK2/1TMOX3
Title *	Test API sur un jeu de Métadonnées complètes
Link to data	http://pmb-bordeaux.fr/maggot/metadata/frim1
Other ID	Maggot: frim1
Contact *	Use email button above to contact.
Author *	Bourafai-Aziez Asma (EVEAR-Extraction) Bourafai-Aziez Asma (EVEAR-Extraction) - ORCID: 0000-0001-5974-7641 Deborde Catherine (INRAE) - ORCID: 0000-0001-8189-9238 Jacob Daniel (INRAE) - ORCID: 0000-0002-6687-7169
Producer *	Rousselot Guillaume (EVEAR-Extraction) INRAE
Description *	Development, validation, and use of 1H NMR spectroscopy for evaluating the quality of acerola-based food supplements and quantifying ascorbic acid English (2022-08-05)
Subject *	Medicine, Health and Life Sciences
Keyword *	Acerola ascorbic acid Metabolomic profiling vitamin C NMR
Topic Classification *	metabolic process (Plant Trait Ontology) http://purl.obolibrary.org/obo/GO_0008152 fruit (Plant Trait Ontology) http://purl.obolibrary.org/obo/PO_0009001
Kind of Data *	Dataset; Other
Other Kind of Data *	1H NMR spectra
Data Origin *	experimental data
Life cycle step *	Original release
Related Publication	Development, validation, and use of 1H NMR spectroscopy for evaluating the quality of acerola-based food supplements and quantifying ascorbic acid, Asma Bourafai-Aziez , Daniel Jacob, Gwladys Charpentier, Emmanuel Cassin, Guillaume Rousselot, Annick Moing and Catherine Deborde. Molecules. 2022; 27(17):5614 doi: 10.3390/molecules27175614 https://doi.org/10.3390/molecules27175614
Other References *	Experimental data tables: https://pmb-bordeaux.fr/dataexplorer/?ds=frim1 ; Simulation/Modélisation: https://hal.science/hal-02611223 ; Article: https://doi.org/10.1104/pp.113.231241
Grant Information *	ANR: ANR-11-INBS-0010
Project Information *	(INRAE-EVEAR-Extraction Research Project "Quantification de l'acide ascorbique dans les extraits d'acérola par 1D 1H-RMN" (2019-2020).)
Depositor	Jacob, Daniel
Deposit Date	2023-02-13

What metadata ?

- * Mandatory fields
- * Recommended fields
that can be linked to an ontology
(or to a C.V)
- * Desirable fields

Schema common to « Dataverse »

Common Metadata

title, description, alternativeURL, contacts, authors, collectors, curators, members, depositor, leader, subject, keywords, topics, kindOfData, dataOrigin, lifeCycleStep, publication, otherReferences, grantNumbers, project, ...

Mandatory fields

Recommended fields

Desirable fields

DEFINITION *
STATUS
MANAGEMENT *
DESCRIPTORS *
OTHER
RESOURCES

Short name * ⓘ

Full title * ⓘ

Subject * ⓘ
☐ Agricultural Sciences
 ☐ Arts and Humanities
 ☐ Astronomy and Astrophysics
 ☐ Business and Management
 ☐ Chemistry
 ☐ Computer and Information Science
 ☐ Earth and Environmental Sciences
 ☐ Engineering
 ☐ Law
 ☐ Mathematical Sciences
 ☐ Medicine Health and Life Sciences
 ☐ Other
 ☐ Physics
 ☐ Social Sciences

Description of the dataset * ⓘ

* mandatory fields

DEFINITION *
STATUS
MANAGEMENT *
DESCRIPTORS *
OTHER
RESOURCES

Kind of Data * ⓘ
☐ Audiovisual
 ☐ Collection
 ☐ Dataset
 ☐ Event
 ☐ Image
 ☐ Interactive Resource
 ☐ Model
 ☐ Other
 ☐ Physical Object
 ☐ Service
 ☐ Software
 ☐ Sound
 ☐ Text
 ☐ Workflow

Keywords ⓘ

Search a value: enter the first letters

Topic Classification ⓘ

Search a value: enter the first letters

Data origin ⓘ
☐ Other
 ☐ aggregate data
 ☐ analysis data
 ☐ audiovisual corpus
 ☐ computer code
 ☐ experimental data
 ☐ observational data
 ☐ simulation data
 ☐ survey data
 ☐ text corpus

Specific Profile Metadata

Experimental Factor ⓘ

Search a value: enter the first letters

Measurement type ⓘ

Search a value: enter the first letters

Technology type ⓘ

Search a value: enter the first letters



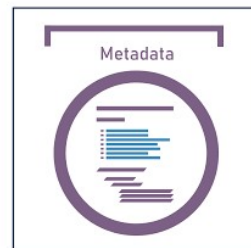
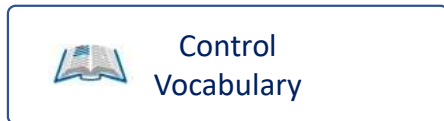
Metadata creation (2/4)

1 Metadata schema (standard)

(container)



2 Vocabulary (content)



From what vocabulary ?

The use of **dictionaries within Maggot** has no other purpose to **facilitate the entry of metadata**, entry which can be long and repetitive in generalist data warehouses (such as repository based on Dataverse).

LAST NAME (*)	FIRST NAME (*)	INSTITUTE (*)	ORCID	EMAIL	
			5828	bordeaux.fr	Edit Del
Dai	Zhanwu	UMR 1287 EGFV, INRAE			Edit Del
Deborde	Catherine	UMR 1332 BFP INRAE	0000-0001-5687-9059	catherine.deborde@inrae.fr	Edit Del
<input type="text" value="Dussarrat"/>	<input type="text" value="Thomas"/>	<input type="text" value="UMR 1332 BFP INRAE"/>	<input type="text" value="0000-0001-6245-365"/>	<input type="text" value="thomas.dussarrat@inrae.fr"/>	Save Cancel
Eveillard	Sandrine	Biologie du Fruit et Pathologie Facility, France <i>BFP</i>	002-8078-	sandrine.eveillard@inrae.fr	Edit Del
Fouillen	Laetitia				Edit Del
Gautier	Roselyne	National Research Institute for Agriculture, Food and Environment			Edit Del
Giauffret	Catherine	Government, France	002-1469-		Edit Del

Dictionaries allow you to record **multiple information** necessary to **define an entity**, such as the names of people, or even the funders.

Its information, once entered and saved in a file called a dictionary, can be **subsequently associated with the corresponding entity**.

Example : people

Contacts

Canlet Cécile, Deborde Catherine

Add a value: enter the first three letters

Authors

Canlet Cécile

Add a value: de

Giraudeau Patrick

Project Lead Deborde Catherine

Cahoreau Edern

Add a value: enter the first three letters

Data Collector

Add a value: enter the first three letters

Data Curator

Add a value: enter the first three letters

Data Member

Add a value: enter the first three letters



Contact ? Use email button above to contact.

Canlet, Cécile (INRAE)
Deborde, Catherine (INRAE)

Author ? Canlet, Cécile (INRAE) - ORCID: 0000-0002-6389-0712
Deborde, Catherine (INRAE) - ORCID: 0000-0001-5687-9059

Contributor ? Project Leader : Giraudeau, Patrick (Univ. Nantes) - ORCID: 0000-0001-9346-9147
Data Collector : Deborde, Catherine (INRAE) - ORCID: 0000-0001-5687-9059
Data Collector : Canlet, Cécile (INRAE) - ORCID: 0000-0002-6389-0712
Data Collector : Gautier, Roselyne (INRAE)
Data Collector : Jousse, Cyril (Univ. Clermont Auvergne) - ORCID: 0000-0002-5899-8243
Data Collector : Lacaze, Méliia (INRAE)
Data Collector : Martineau, Estelle (Univ. Nantes) - ORCID: 0000-0001-5093-2138
Data Collector : Peyriga, Lindsay (INRAE) - ORCID: 0000-0002-6138-7961
Data Collector : Richard, Tristan (Univ. Bordeaux) - ORCID: 0000-0002-5308-8697
Data Collector : Silvestre, Virginie (Univ. Nantes)
Data Collector : Traïkia, Mounir (Univ. Clermont Auvergne) - ORCID: 0000-0002-4595-0400
Data Curator : Deborde, Catherine (INRAE) - ORCID: 0000-0001-5687-9059
Data Curator : Canlet Cécile (INRAE) - ORCID: 0000-0002-6389-0712
Data Curator : Moing, Annick (INRAE) - ORCID: 0000-0003-1144-3600
Data Curator : Jacob, Daniel (INRAE) - ORCID: 0000-0002-6687-7169
Project Member : Cahoreau, Edern (INRAE) - ORCID: 0000-0001-8637-0448
Project Member : Da Costa, Grégory (Univ. Bordeaux) - ORCID: 0000-0002-0336-5828
Project Member : le Mao, Inès (Univ. Bordeaux)



Thus, entering (by autocompletion) just the name of a person will allow the ORCID number, email address and institutional assignment to be associated when **distributing metadata in Dataverse** for example.

List of well-chosen and limited CVs (according to a reference e.g. Data Document Initiative)

Kind of Data * ?

☐ Audiovisual☐ Collection☐ Dataset☐ Event☐ Image☐ Interactive Resource☐ Model☐ Other☐ Physical Object☐ Service☐ Software☐ Sound☐ Text☐ Workflow

Keywords ?

Search a value: experimental

AgroPortal

BioPortal

Topic Classification

Search a value:

experimental

Model of Disease (NCBITAXON)

experimentally modified cell in vitro (OBI)

experimental_feature (OBI)

experimental infection of cell culture (OBI)

experimental disease induction (OBI)

Experimental measurement (EDAM)

List of ontologies to choose according to your domain

Use of dictionaries to target the CV by mixing thesaurus and ontologies

Thesaurus

SKOSMOS

(VOINRAE, LOTERRE, ONTOSTACK, ...)



VOINRAE

VOCABULAIRES OUVERTS

Construire un vocabulaire

https://vocabulaires-ouverts.inrae.fr/construire/

NAME (*)	ONTOLOGY	URL	Add new	
NMR spectroscopy assay	OBI	http://purl.obolibrary.org/obo/OBI_0000623	Edit	Del
agricultural science	EDAM	http://edamontology.org/topic_3810	Edit	Del
amino acid	IOBC	http://purl.jp/bio/4/id/200906089657456524	Edit	Del
analyte assay	MS	http://purl.obolibrary.org/obo/OBI_0000443	Edit	Del
biochemical analysis	IOBC	http://purl.jp/bio/4/id/200906072808564316	Edit	Del
biochemical characterization	IOBC	http://purl.jp/bio/4/id/201306093820876862	Edit	Del
biochemical composition	IOBC	http://purl.jp/bio/4/id/201106016579695836	Edit	Del
biochemistrv	EDAM	htto://edamontology.org/topic_3292	Edit	Del



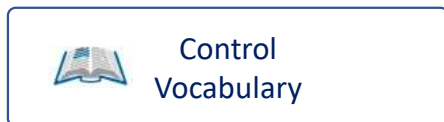
Metadata creation (3/4)

1 Metadata schema (standard)

(container)

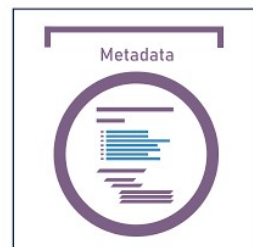


2 Vocabulary (content)



Thesaurus SKOSMOS

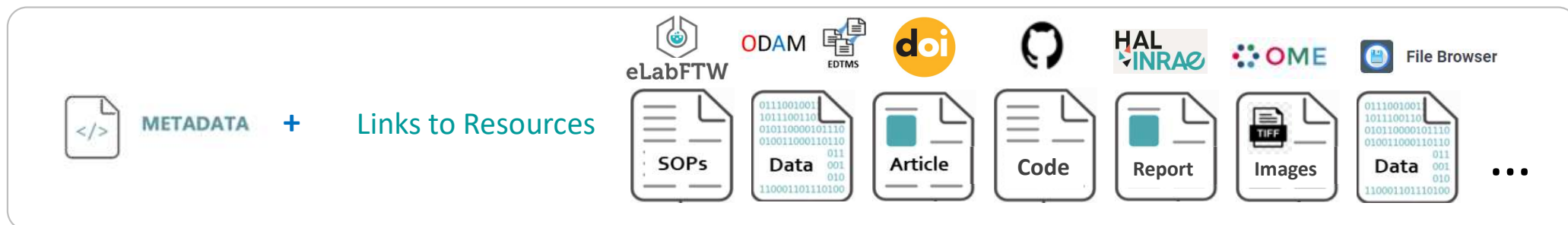
(VOINRAE, LOTERRE, ONTOSTACK, ...)



3

External
resources
(related data)

Additional information ?
Other metadata / data ?



RESOURCES

Type	Media	Description	Location
JournalArticle		Journal of Experimental Botany, Oxford University Press, 2020	http://doi.org/10.1093/jxb/eraa302
Collection		ODAM Experimental data tables	https://pmb-bordeaux.fr/dataexplorer/?dc=Frimouss
Report	application/pdf	Fruit Growth Modelling	https://pmb-bordeaux.fr/getdata/pdf/Frimouss/FruitGrowthModelling.pdf
Software		Growth modeling applied to several fruit species	https://github.com/djacob65/growthmodel



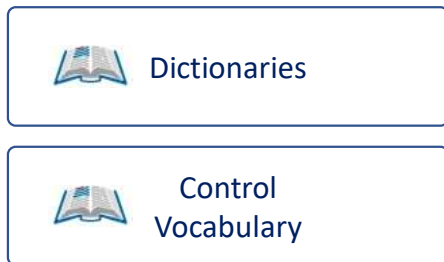
Metadata creation (4/4)

1 Metadata schema (standard)

(container)

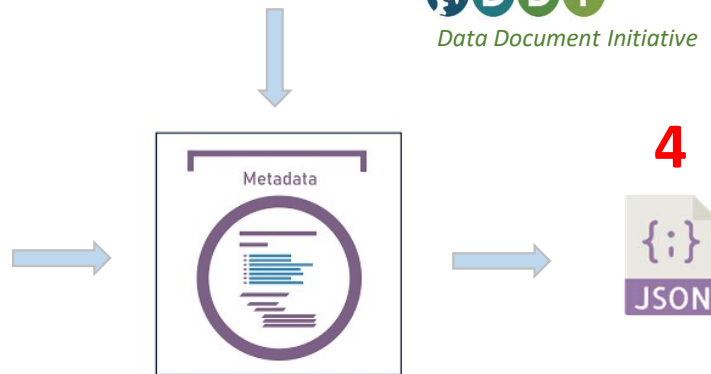


2 Vocabulary (content)



Thesaurus SKOSMOS

(VOINRAE, LOTERRE, ONTOSTACK, ...)



4



As an output, we produce a file in JSON format that can be read by both humans and machines.

3 External resources

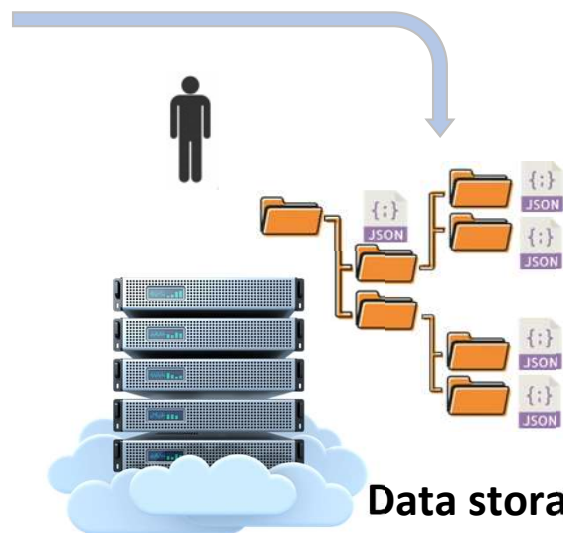
(related data)



What to do with this file?



The deposit must be done by hand because the Maggot tool must only have access to the storage space in read mode. So this guarantees that the user has writing rights to this space without having to manage user accounts on the Maggot side.



Local (meta)data repository
Storage space becomes the data repository

DESCRPTORS

Kind of Data

☐ Audiovisual
 ☐ Collection
 ☒ Dataset
 ☐ Event
 ☐ Image
 ☐ Interactive Resource
 ☐ Model
 ☐ Other
 ☐ Physical Object
 ☐ Service
 ☐ Software
 ☐ Sound
 ☐ Text
 ☐ Workflow

Keywords

Search a value:

enter the first letters

Topic Classification

Search a value:

enter the first letters

Data origin

Search

Empty the form

(kindOfData => Dataset)

Short name	Full title	Status of the dataset	Access rights to data	Metadata
AmaizingEnzymes	Leaf enzyme activities and total proteins of maize hybrids cultivated in the field	Processed	Private	
AmaizingNMR	NMR metabolomic and starch data of young leaf of maize hybrids cultivated in the field with normal sowing in 2013	Processed	Private	
Atacama	Atacama	Processed	Public	
Frimouss	FRuit Integrative MOdelling for a Unified Selection System	Processed	Public	
Frimouss-PeppEgg	1H-NMR metabolomic profiling data of eggplant or pepper fruit during its development	Processed	Public	
Metabofla1	Infection response and susceptibility reduction in the pathosystem Grapevine / Flavescence dorée	Processed	Public	
NMRmetoboRing	NMR metabolite quantification			

Atacama

DESCRIPTION

Predictive metabolomics performed on 24 extremophile plant species in 19 different sites along an elevation transect in the Atacama Desert. (2021-06-01)

DEFINITION

Full title	Subject
Atacama	Earth and Environmental Sciences

STATUS

Status of the dataset	Access rights to data	Language	Life cycle step
Processed	Public	English	<ul style="list-style-type: none"> Original release Deposit

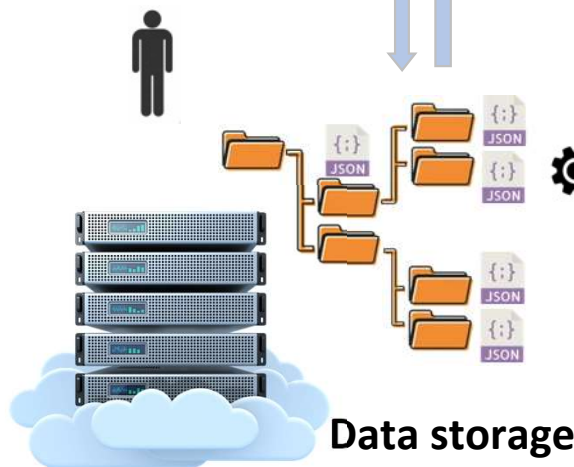
MANAGEMENT

Contacts	Authors	Data curators	Project members
Dussarrat Thomas	<ul style="list-style-type: none"> Dussarrat Thomas Gibon Yves Gutierrez Rodrigo Petriaq Pierre 	Jacob Daniel	Cassan Cédric

Project leader	WP leader	Depositor	Producer	Grant Information
Gutierrez Rodrigo	Gibon Yves	Jacob Daniel	<ul style="list-style-type: none"> Bordeaux Metabolome Plant Systems Biology Lab 	<ul style="list-style-type: none"> MetaboHub Phénome

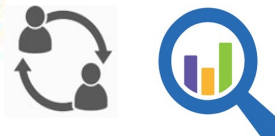
Metadata export

Publish your metadata
... along with data



scan

Share & Search
(meta)data



Institutional data repositories



...

Allow machines to collect metadata

Publish your metadata
... along with data



Push



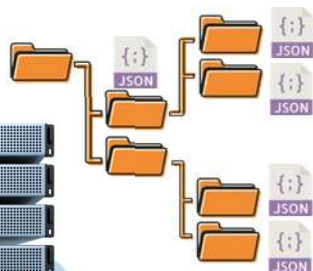
Metadata



Data storage

Local (meta)data repository

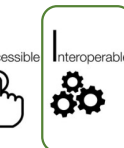
Storage space becomes the data repository



scan



Share & Search
(meta)data

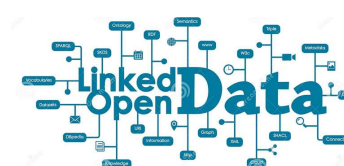
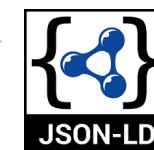


Interoperability



*Open Archives Initiative
Protocol for Metadata
Harvesting*

Dublin Core



« **Climb the LOD mountain** »
gently, and step by step.