Maggot : Designed to foster good data management, with " data sharing " in mind.

Daniel Jacob, François Ehrenmann, Philippe Chaumeil

▶ To cite this version:

HAL Id: hal-04256711

https://hal.inrae.fr/hal-04256711

Submitted on 25 Oct 2023

Maggot
Metadata aggregation on data storage

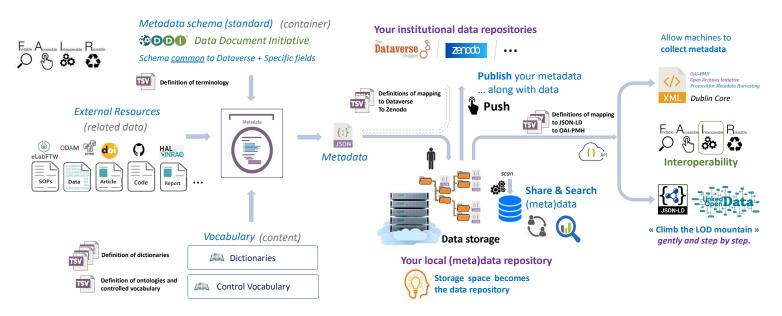Designed to foster good data management, with « data sharing » in mind.

Daniel Jacob[1], François Ehrenmann[2], Philippe Chaumeil[2]
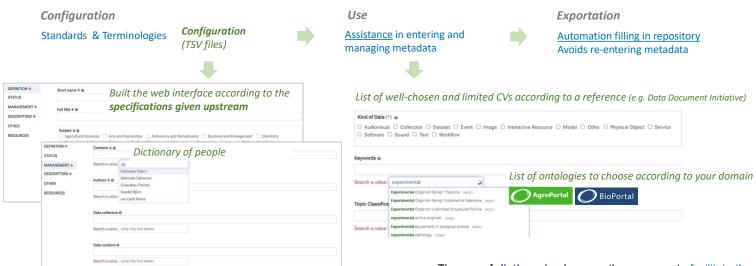(1) INRAE UMR Biologie du Fruit & Pathologie
(2) INRAE UMR BioGECO

An ecosystem for sharing metadata

The **Maggot** tool allows users to describe the Metadata of their datasets produced within a collective (research unit, platform, multi-partner project, etc.), thus making it possible
- to answer certain questions of the **Data Management Plan** (DMP) concerning the **organization, documentation, storage and sharing of data** in the data storage space,
- to meet certain data and metadata requirements, referred to by the **Open Research Europe** (ORE) in accordance with the FAIR principles.

Metadata schema (standard) *(container)*
DDI *Data Document Initiative*
Schema common to Dataverse + Specific fields

TSV — Definition of terminology

External Resources *(related data)*
eLabFTW · ODAM EDTMS · doi · GitHub · HAL INRAE
SOPs · Data · Article · Code · Report · …

TSV — Definition of dictionaries
TSV — Definition of ontologies and controlled vocabulary

Vocabulary *(content)*
Dictionaries
Control Vocabulary

Metadata · JSON · Metadata

Your institutional data repositories
Dataverse Project · zenodo · …

TSV — Definitions of mapping to Dataverse To Zenodo

Publish your metadata … along with data
Push

TSV — Definitions of mapping to JSON-LD to OAI-PMH

Data storage
scan
Share & Search (meta)data

Your local (meta)data repository
Storage space becomes the data repository

Allow machines to collect metadata
OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting
XML Dublin Core
Interoperability
JSON-LD · Linked Open Data
« Climb the LOD mountain » gently and step by step.

Metadata thus produced can be **associated with corresponding data directly in the storage space** which is then scanned and then referenced in a database. We can thus **carry out a search on the metadata** in order to find one or more sets of data. **Only metadata is accessible by default**.

Configuration
Standards & Terminologies
Configuration *(TSV files)*

Use
Assistance in entering and managing metadata

Exportation
Automation filling in repository
Avoids re-entering metadata

Built the web interface according to the *specifications given upstream*

Dictionary of people

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
Experimental Organism Benign Thecoma (NCIT)
Experimental Organism Benign Endometrial Adenoma (NCIT)
Experimental Organism Luteinized Unruptured Follicle (NCIT)
experimental animal engineer (ERMC)
Experimental equipments in biological science (ERMC)
experimental pathology (ERMC)

Topic Classification

List of ontologies to choose according to your domain
AgroPortal · BioPortal

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

NAME (*)	ONTOLOGY	URL	
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/200906086965745824	Save Cancel
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/201306093820876662	Edit Del
biochemical composition	IOBC	http://purl.jp/bio/4/id/201106016579699836	Edit Del

Thesaurus SKOSMOS
(VOINRAE, LOTERRE, ONTOSTACK, …)

V.O. INRAE VOCABULAIRES OUVERTS
Construire un vocabulaire

The use of **dictionaries** 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).

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.