Metadata

Exercise : using the Ecological Metadata Language - EML through Galaxy-Ecology

July 2025

Yvan Le Bras

Scientific and technical coordinator

@PNDB

@DataTerra

Olivier Norvez

Animation coordinator

@DataTerra

@PNDB

@THEIA

https://biodiversitydata.github.io/

FRB FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ

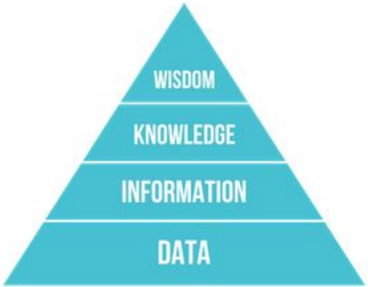
CESAB CENTRE FOR THE SYNTHESIS AND ANALYSIS OF BIODIVERSITY

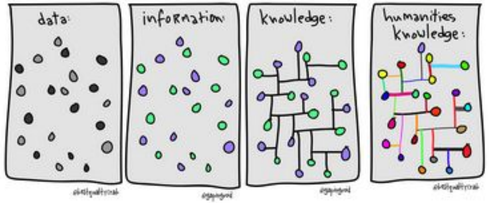
Tables of contents

- Reminders
  - context and issues
  - EML
  - Galaxy ecology
- Connect to Galaxy Ecology Europe
- Creating an EML metadata record from EML metadata template files in text format
- Creation of EML Assembly Line metadata template files
- Creating tab-delimited text metadata template files from an EML record
- Creating an EML record from an ISO19115 metadata record
- Creation of an EML record for genetic data
- Perspectives

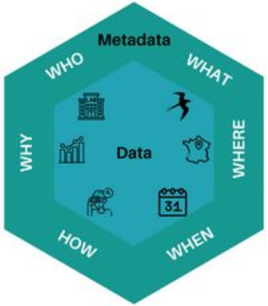
Reminder : context and issues

- **Heterogeneity** (data types, origin, standards) & **diversity** of "objects" to be linked together[1]
- **Loss of information** over time[2]
- Toward a better **open science** and **reproducibility** [3] [4]

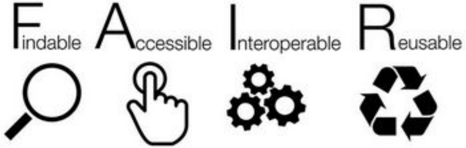




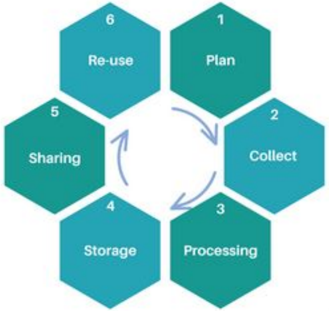
BIODIVERSITY {META}DATA DEFINITION



CONTINIUM BETWEEN DATA AND METADATA



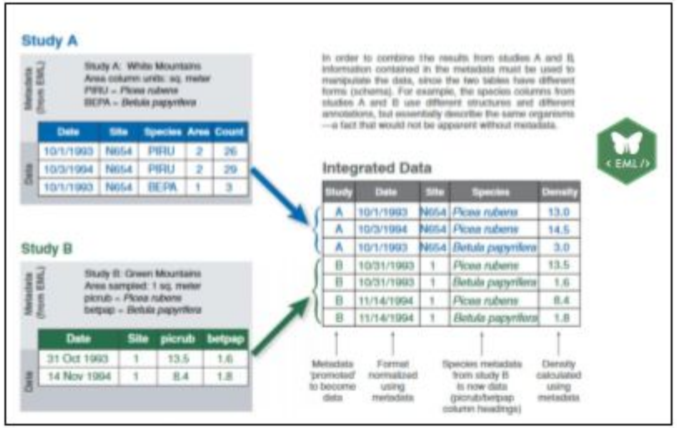
Cycle de vie des données



[1] Page (2016)
[2] Michener *et al.* (1997)
[3] Powers & Hampton (2018)
[4] Genkins *et al.* (2016)

https://biodiversitydata.github.io/

Reminder : Ecological Metadata Language



Study A

Metadata (from EML)

Study A: White Mountains
Area column units: sq. meter
PIRU – *Picea rubens*
BEPA – *Betula papyrifera*

Data

Date	Site	Species	Area	Count
10/1/1993	N654	PIRU	2	26
10/3/1994	N654	PIRU	2	29
10/1/1993	N654	BEPA	1	3

Study B

Metadata (from EML)

Study B: Green Mountains
Area sampled: 1 sq. meter
picrub – *Picea rubens*
betpap – *Betula papyrifera*

Data

Date	Site	picrub	betpap
31 Oct 1993	1	13.5	1.6
14 Nov 1994	1	8.4	1.8

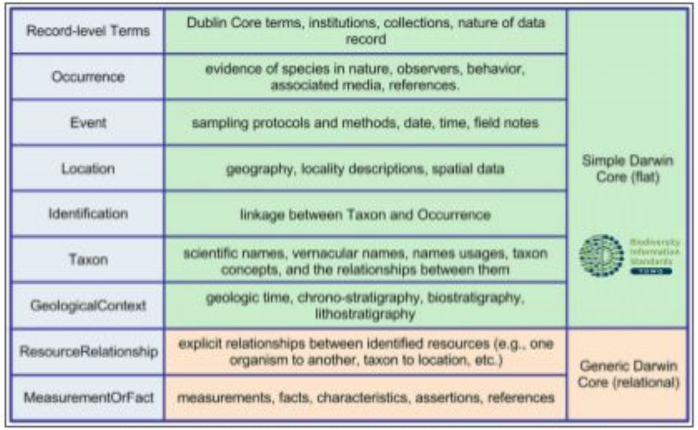
In order to combine the results from studies A and B, information contained in the metadata must be used to manipulate the data, since the two tables have different forms (schema). For example, the species columns from studies A and B use different structures and different annotations, but essentially describe the same organisms —a fact that would not be apparent without metadata.

Integrated Data

Study	Date	Site	Species	Density
A	10/1/1993	N654	*Picea rubens*	13.0
A	10/3/1994	N654	*Picea rubens*	14.5
A	10/1/1993	N654	*Betula papyrifera*	3.0
B	10/31/1993	1	*Picea rubens*	13.5
B	10/31/1993	1	*Betula papyrifera*	1.6
B	11/14/1994	1	*Picea rubens*	8.4
B	11/14/1994	1	*Betula papyrifera*	1.8

Metadata 'promoted' to become data
Format normalized using metadata
Species metadata from study B is now data (picrub/betpap column headings)
Density calculated using metadata
< EML />

M.B. Jones *et al.*, 2006 https://doi.org/10.1146/annurev.ecolsys.37.091305.110031 +
https://eml.ecoinformatics.org/



Record-level Terms	Dublin Core terms, institutions, collections, nature of data record	Simple Darwin Core (flat)
Occurrence	evidence of species in nature, observers, behavior, associated media, references.	
Event	sampling protocols and methods, date, time, field notes	
Location	geography, locality descriptions, spatial data	
Identification	linkage between Taxon and Occurrence	
Taxon	scientific names, vernacular names, names usages, taxon concepts, and the relationships between them	
GeologicalContext	geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy	
ResourceRelationship	explicit relationships between identified resources (e.g., one organism to another, taxon to location, etc.)	Generic Darwin Core (relational)
MeasurementOrFact	measurements, facts, characteristics, assertions, references	

J. Wieczorek *et al.*, 2012 https://doi.org/10.1371/journal.pone.0029715

metadata
data

Raw vs. specific and derived information

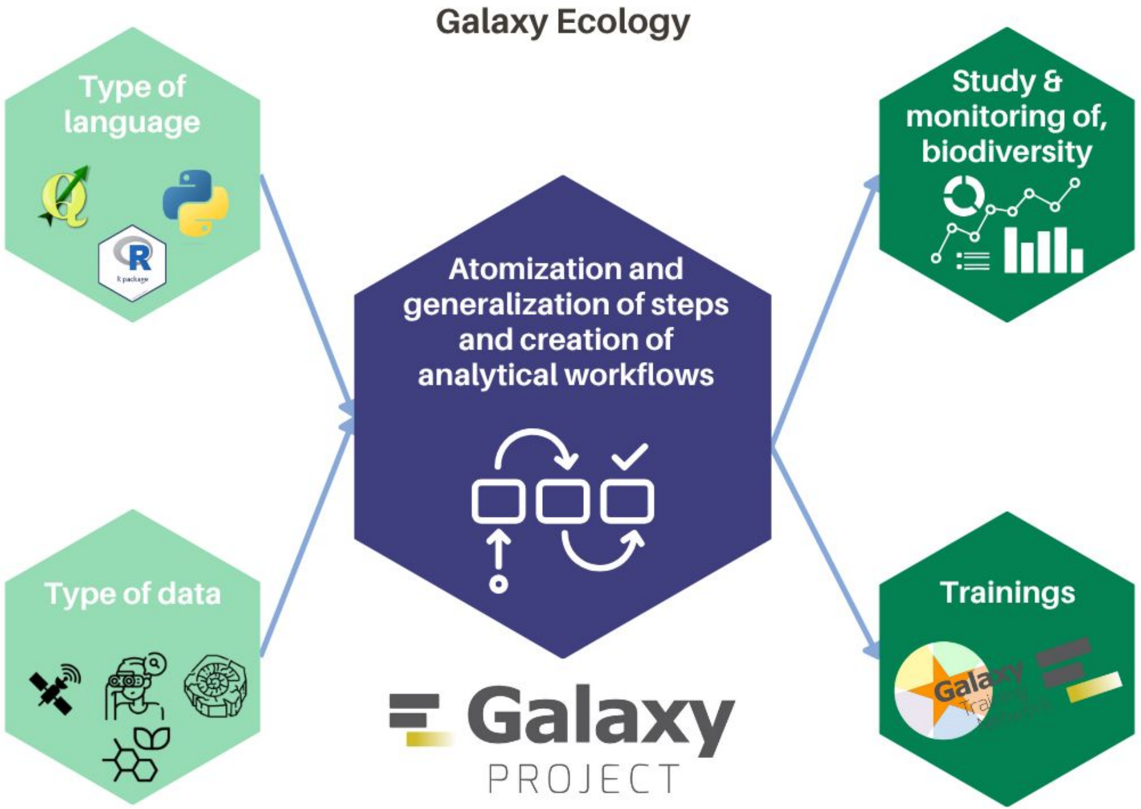
Knowledge of the standard, its formalism, its restrictions

Need to know the time for standardization

Diversity of data types

COMPLEMENTARITY OF THE TWO APPROACHES

Reminder : Galaxy-Ecology

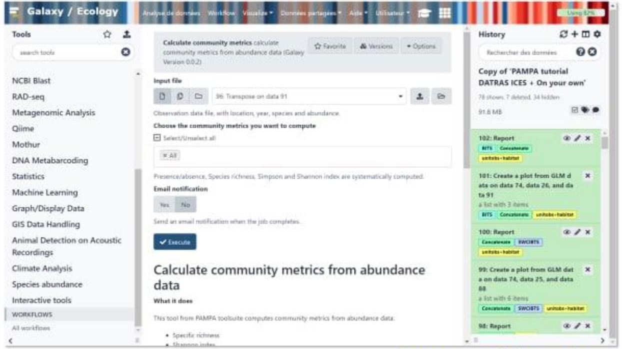

Galaxy Ecology
Type of language
R package
Type of data
Atomization and generalization of steps and creation of analytical workflows
Galaxy PROJECT
Study & monitoring of, biodiversity
Trainings
Galaxy Training Network

See more

- https://www.pndb.fr/pages/galaxy-ecology0/
- Royaux et. al., 2025. Guidance framework to apply best practices in ecological data analysis: lessons learned from building Galaxy-Ecology. Gigascience. https://doi.org/10.1093/gigascience/giae122

Reminder : Galaxy-Ecology

Use scripts on a cluster → **with 0 programming skills!** ← Create, edit, share, reuse workflows



Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.2)

Input file: 96: Transpose on data 91

Observation data file, with location, year, species and abundance.

Choose the community metrics you want to compute

☐ Select/Unselect all

× All

Presence/absence, Species richness, Simpson and Shannon index are systematically computed.

Email notification: Yes No

Send an email notification when the job completes.

Calculate community metrics from abundance data

What it does

This tool from PAMPA toolsuite computes community metrics from abundance data:

- Specific richness
- Shannon index

Copy of 'PAMPA tutorial DATRAS ICES + On your own'

78 shown, 7 deleted, 34 hidden

91.8 MB

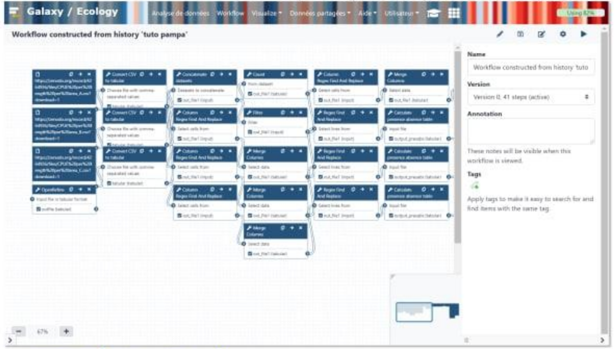
102: Report — BITS, Concatenate, unitobs+habitat

101: Create a plot from GLM data on data 74, data 26, and data 91 — a list with 3 items — BITS, Concatenate, unitobs+habitat

100: Report — Concatenate, SWCIBTS, unitobs+habitat

99: Create a plot from GLM data on data 74, data 25, and data 88 — a list with 6 items — Concatenate, SWCIBTS, unitobs+habitat

98: Report



Workflow constructed from history 'tuto pampa'

Name: Workflow constructed from history 'tuto

Version: Version 0: 41 steps (active)

Annotation

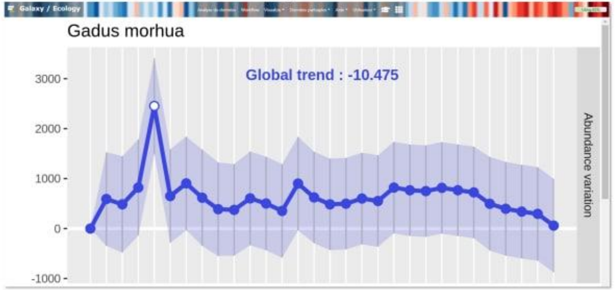
These notes will be visible when this workflow is viewed.

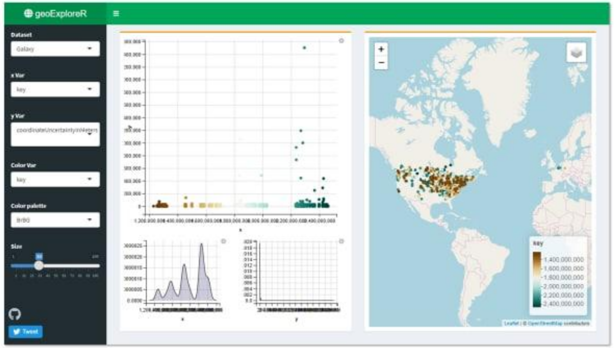
Tags

Apply tags to make it easy to search for and find items with the same tag.

https://ecology.usegalaxy.eu/

Create & share interactive visualizations,
Even deploy and share Jupyter notebooks or R shiny apps





https://biodiversitydata.github.io/

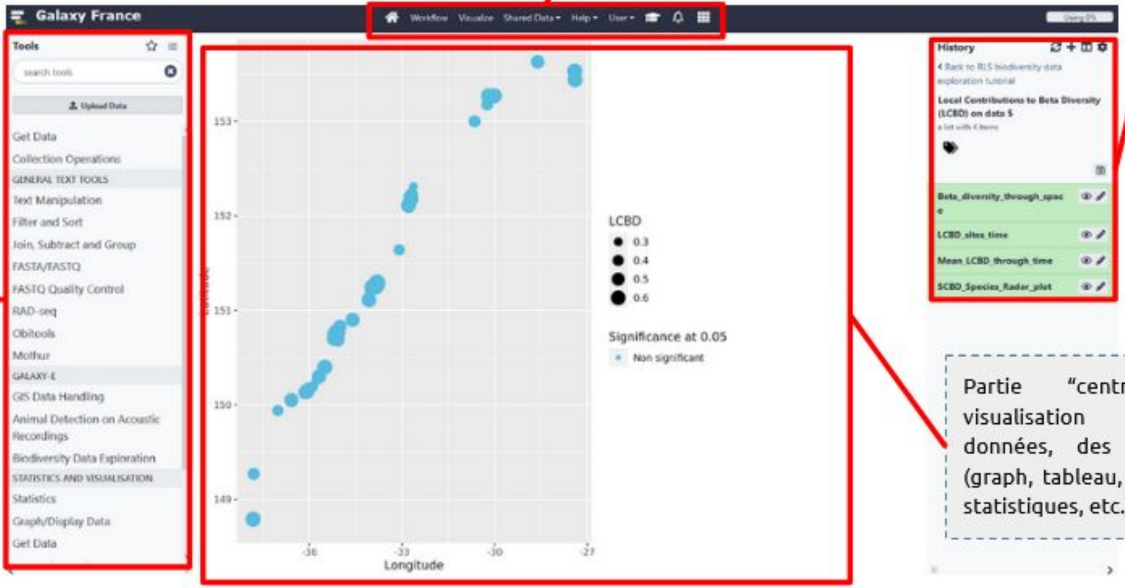
Reminder : Galaxy-Ecology

Galaxy PROJECT

Partie "bandeau" : raccourcis, connection, menu

Partie "history" : Historique des étapes d'analyse

Partie "outils" : liens vers les outils existants, des workflows, avec un moteur de recherche par mots clés



Partie "centrale" : visualisation des données, des analyses (graph, tableau, résultats statistiques, etc.)

https://biodiversitydata.github.io/

FRB FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ

CESAB CENTRE FOR THE SYNTHESIS AND ANALYSIS OF BIODIVERSITY

Connect to Galaxy Ecology Europe

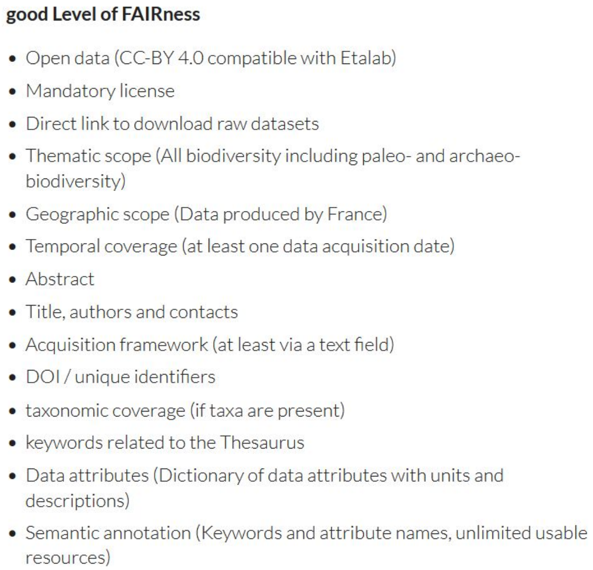
To create a Galaxy Europe account: https://ecology.usegalaxy.eu/login/start?redirect=None or https://usegalaxy.eu/login/start?redirect=None

To access training resources: https://usegalaxy.eu/join-training/eml

Then go to Galaxy Ecology Europe: https://ecology.usegalaxy.eu/

Context and issues

- Rich metadata production (F2 of the FAIR principles, Wilkinson et al., 2016)
- Easy creation/modification of metadata by researchers
  - Metadata inference
  - Data attributes
  - Taxonomic coverage
  - Geographic coverage
  - Personal information
  - Semantic annotations
  - No need to know the specifications of the target standard
  - No coding skills required
  - Web-enabled
  - Possibility of incremental and collaborative development
- Human-machine exchange
- List of minimum criteria chosen by the PNDB



good Level of FAIRness

- Open data (CC-BY 4.0 compatible with Etalab)
- Mandatory license
- Direct link to download raw datasets
- Thematic scope (All biodiversity including paleo- and archaeo-biodiversity)
- Geographic scope (Data produced by France)
- Temporal coverage (at least one data acquisition date)
- Abstract
- Title, authors and contacts
- Acquisition framework (at least via a text field)
- DOI / unique identifiers
- taxonomic coverage (if taxa are present)
- keywords related to the Thesaurus
- Data attributes (Dictionary of data attributes with units and descriptions)
- Semantic annotation (Keywords and attribute names, unlimited usable resources)

https://biodiversitydata.github.io/

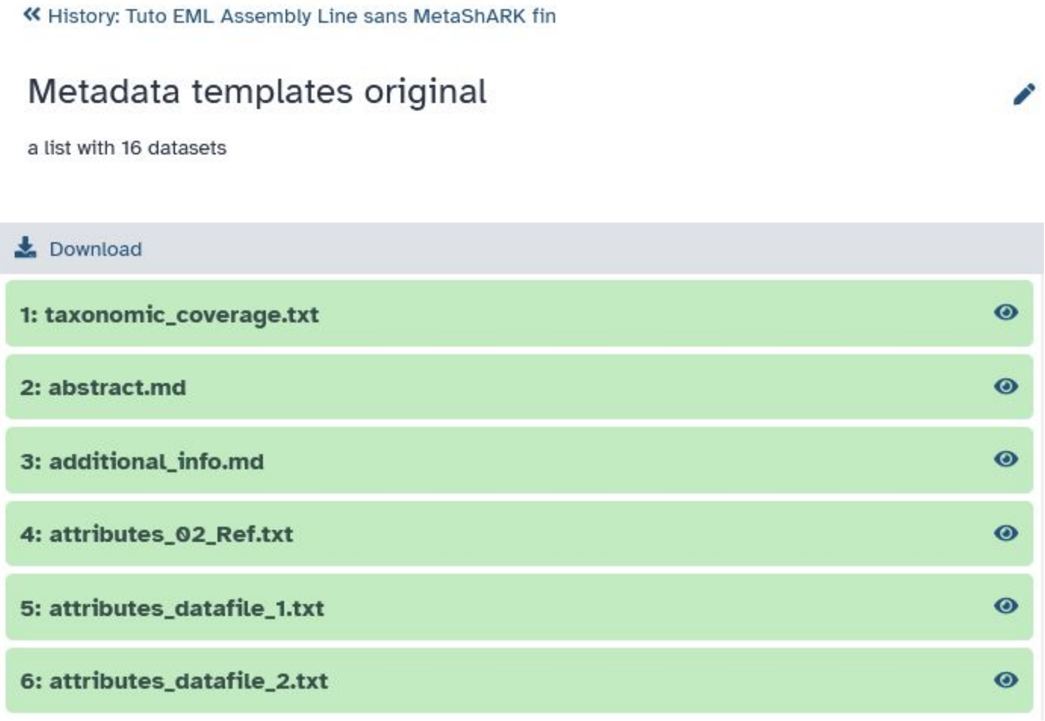
Creating an EML metadata record from EML metadata template files in text format

Click on Galaxy History with the input data: https://ecology.usegalaxy.eu/u/ylebras/h/eml-assembly-line-tape-1--cration-deml

Steps

- **Import the data, then click on the "home" icon at the top of the page and/or refresh the page.**
- **File review / What information is there (by looking at just the file names first, then browsing the contents of the metadata templates)?**

Creating an EML metadata record from EML metadata template files in text format



« History: Tuto EML Assembly Line sans MetaShARK fin

Metadata templates original

a list with 16 datasets

Download

1: taxonomic_coverage.txt

2: abstract.md

3: additional_info.md

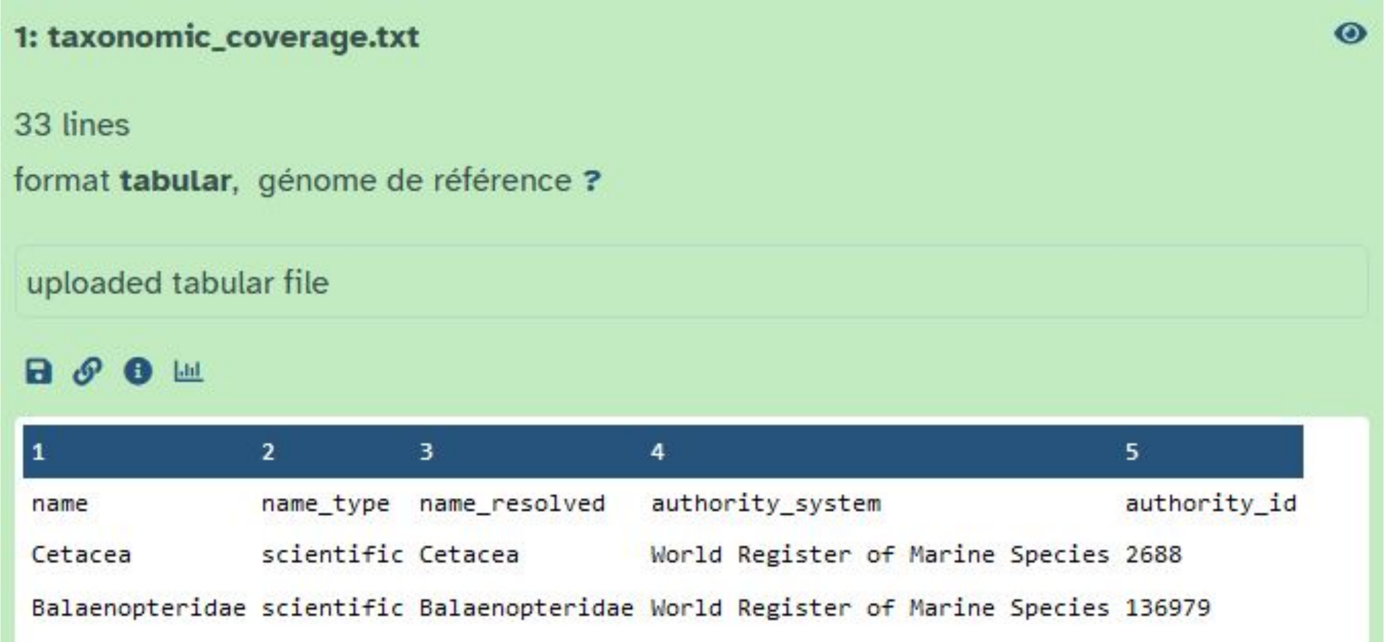
4: attributes_02_Ref.txt

5: attributes_datafile_1.txt

6: attributes_datafile_2.txt

Creating an EML metadata record from EML metadata template files in text format

Taxonomic coverage



1: taxonomic_coverage.txt

33 lines

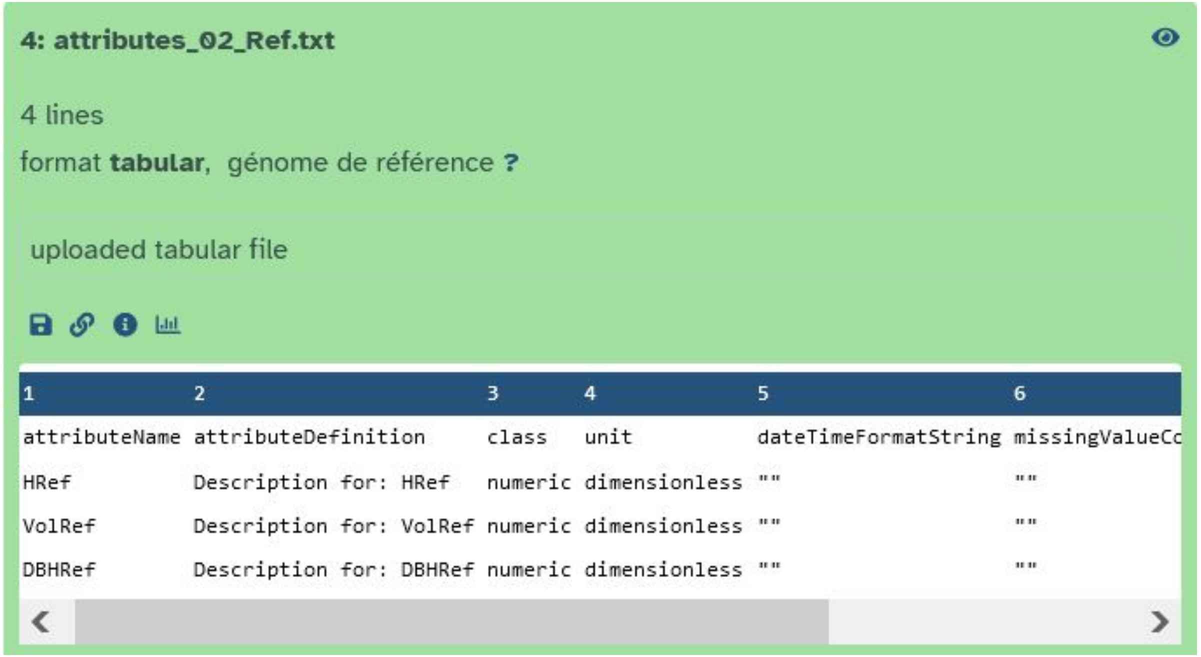
format **tabular**, génome de référence **?**

uploaded tabular file

1	2	3	4	5
name	name_type	name_resolved	authority_system	authority_id
Cetacea	scientific	Cetacea	World Register of Marine Species	2688
Balaenopteridae	scientific	Balaenopteridae	World Register of Marine Species	136979

Creating an EML metadata record from EML metadata template files in text format

Attributes



4: attributes_02_Ref.txt

4 lines

format **tabular**, génome de référence **?**

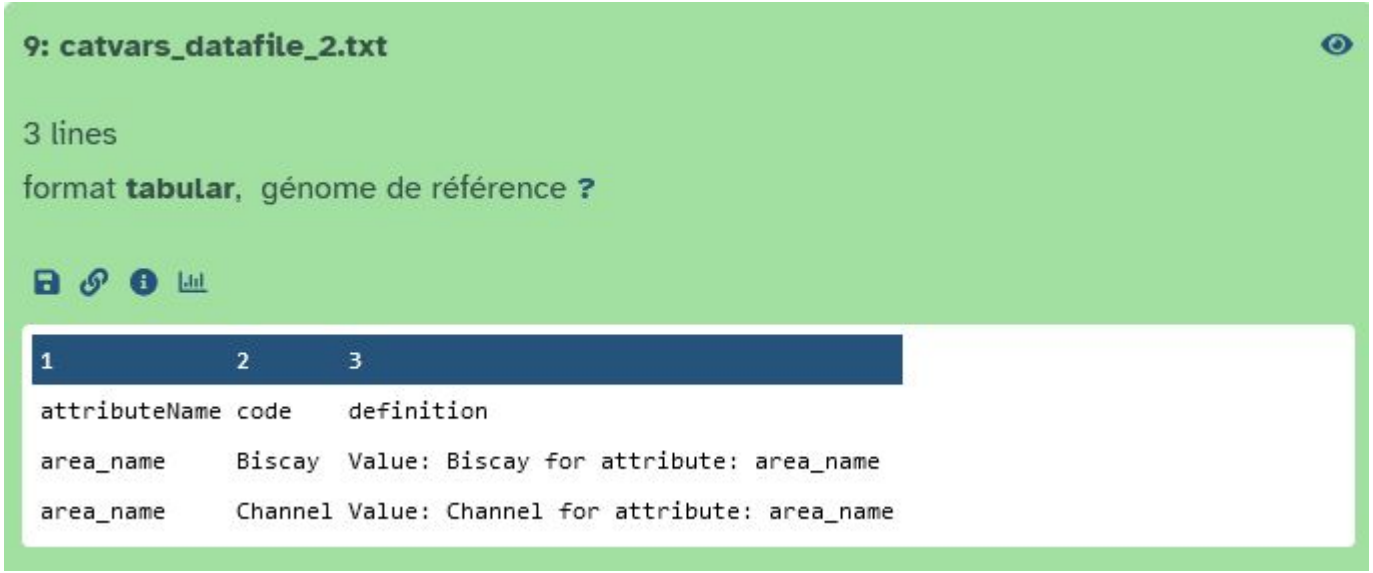
uploaded tabular file

1	2	3	4	5	6
attributeName	attributeDefinition	class	unit	dateTimeFormatString	missingValueCc
HRef	Description for: HRef	numeric	dimensionless	""	""
VolRef	Description for: VolRef	numeric	dimensionless	""	""
DBHRef	Description for: DBHRef	numeric	dimensionless	""	""

The missing value code explanation is missing.

Creating an EML metadata record from EML metadata template files in text format

Categorical variable



9: catvars_datafile_2.txt

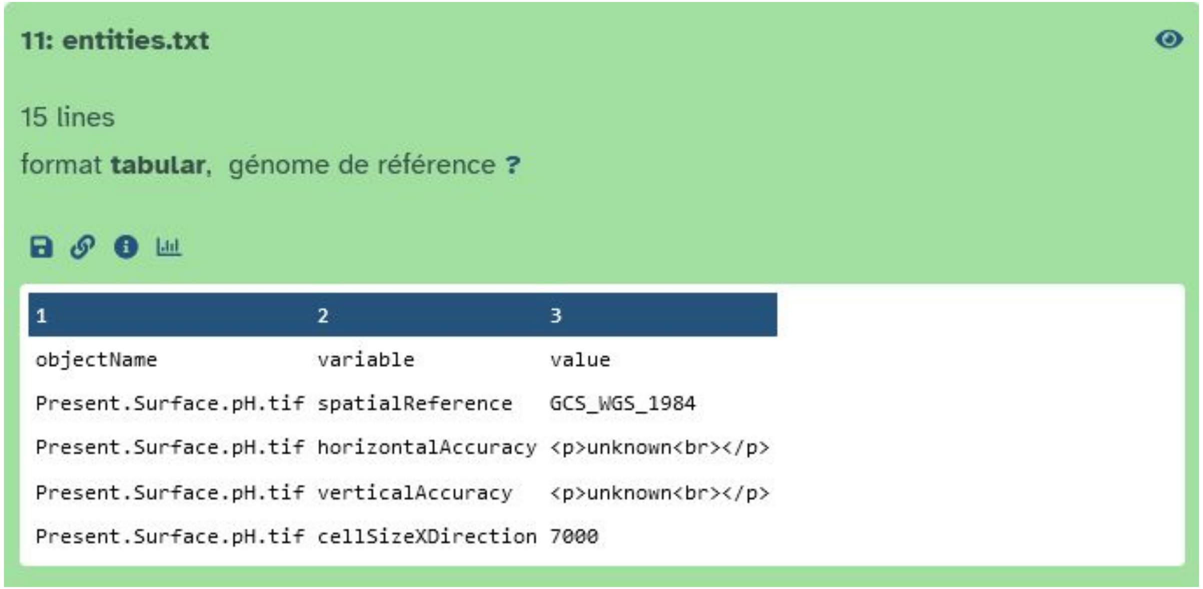
3 lines

format **tabular**, génome de référence ?

1	2	3
attributeName	code	definition
area_name	Biscay	Value: Biscay for attribute: area_name
area_name	Channel	Value: Channel for attribute: area_name

Creating an EML metadata record from EML metadata template files in text format

Entities to describe the spatial information of raster and vector files



11: entities.txt

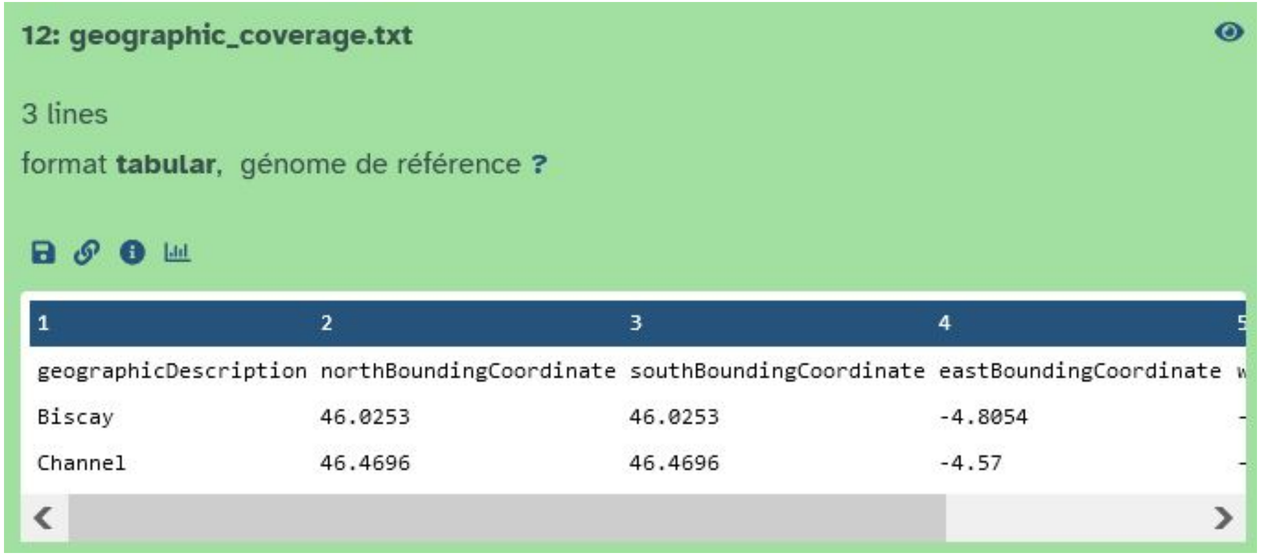
15 lines

format **tabular**, génome de référence **?**

1	2	3
objectName	variable	value
Present.Surface.pH.tif	spatialReference	GCS_WGS_1984
Present.Surface.pH.tif	horizontalAccuracy	<p>unknown </p>
Present.Surface.pH.tif	verticalAccuracy	<p>unknown </p>
Present.Surface.pH.tif	cellSizeXDirection	7000

Creating an EML metadata record from EML metadata template files in text format

Geographic coverage



12: geographic_coverage.txt

3 lines

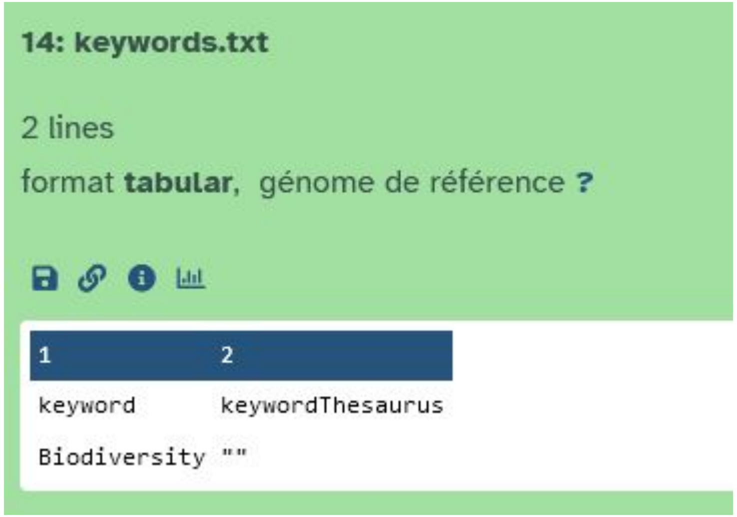
format **tabular**, génome de référence **?**

1	2	3	4	5
geographicDescription	northBoundingCoordinate	southBoundingCoordinate	eastBoundingCoordinate	w
Biscay	46.0253	46.0253	-4.8054	-
Channel	46.4696	46.4696	-4.57	-

westBoundingCoordinate is missing

Creating an EML metadata record from EML metadata template files in text format

Keywords



14: keywords.txt

2 lines

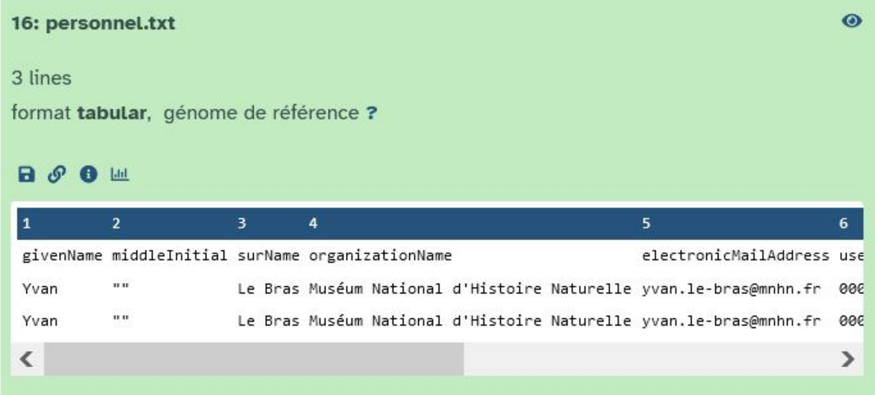
format **tabular**, génome de référence **?**

1	2
keyword	keywordThesaurus
Biodiversity	""

https://biodiversitydata.github.io/

Creating an EML metadata record from EML metadata template files in text format

Contacts of persons (data producer, PI, data manager,...)

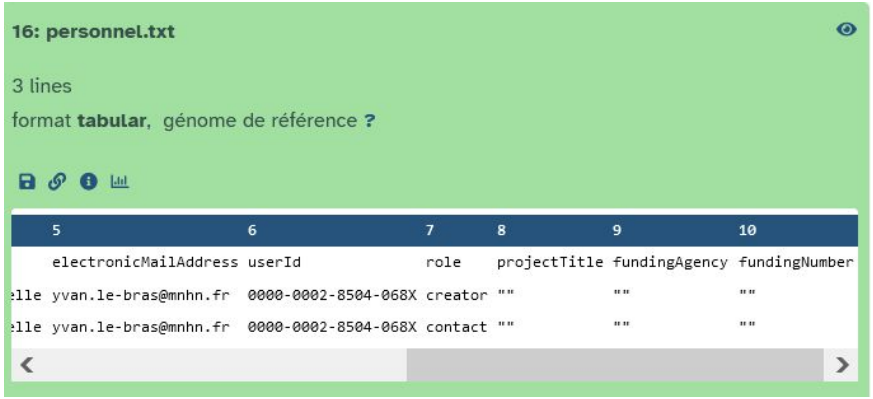


16: personnel.txt

3 lines

format **tabular**, génome de référence **?**

1	2	3	4	5	6
givenName	middleInitial	surName	organizationName	electronicMailAddress	use
Yvan	""	Le Bras	Muséum National d'Histoire Naturelle	yvan.le-bras@mnhn.fr	000
Yvan	""	Le Bras	Muséum National d'Histoire Naturelle	yvan.le-bras@mnhn.fr	000



16: personnel.txt

3 lines

format **tabular**, génome de référence **?**

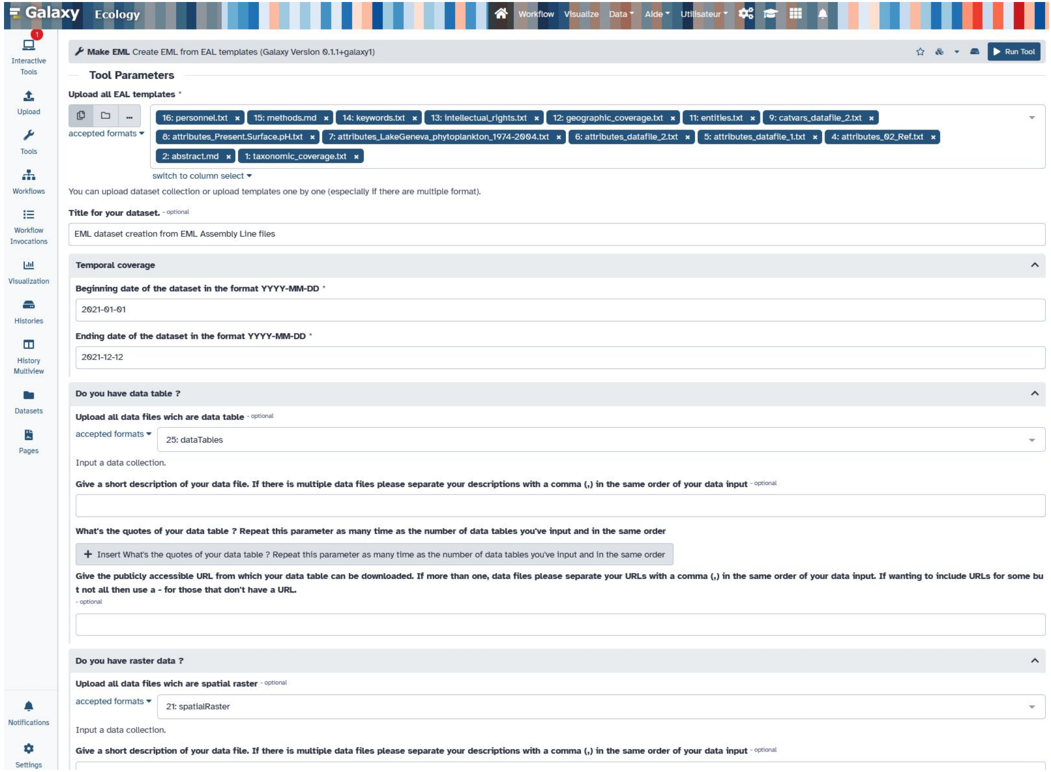
	5	6	7	8	9	10
	electronicMailAddress	userId	role	projectTitle	fundingAgency	fundingNumber
elle	yvan.le-bras@mnhn.fr	0000-0002-8504-068X	creator	""	""	""
elle	yvan.le-bras@mnhn.fr	0000-0002-8504-068X	contact	""	""	""

Creating an EML metadata record from EML metadata template files in text format

Steps

- Import the data, then click on the "home" icon at the top of the page and/or refresh the page.
- File review / What information is there (by looking at just the file names first, then browsing the contents of the metadata templates)?
- **EML creation from EML Assembly Line template files using the Galaxy Make EML tool**
  - Provide the data package title “EML dataset creation from EML Assembly Line files”
  - Temporal coverage “2021-01-01” to “2021-12-12”
  - Select the data collections of type “dataTable”, “spatialRaster”, and “spatialVector”
  - The license is automatically CC-BY 4.0 compatible with Etalab 2.0 open license.
    - make eml /
    - What was written in the metadata?
    - EML validated?

Creating an EML metadata record from EML metadata template files in text format



Make EML Create EML from EAL templates (Galaxy Version 0.1.1+galaxy1)

Tool Parameters

Upload all EAL templates *

accepted formats

16: personnel.txt, 15: methods.md, 14: keywords.txt, 13: intellectual_rights.txt, 12: geographic_coverage.txt, 11: entities.txt, 9: catvars_datafile_2.txt, 8: attributes_Present.Surface.pH.txt, 7: attributes_LakeGeneva_phytoplankton_1974-2004.txt, 6: attributes_datafile_2.txt, 5: attributes_datafile_1.txt, 4: attributes_02_Ref.txt, 2: abstract.md, 1: taxonomic_coverage.txt

switch to column select

You can upload dataset collection or upload templates one by one (especially if there are multiple format).

Title for your dataset. - optional

EML dataset creation from EML Assembly Line files

Temporal coverage

Beginning date of the dataset in the format YYYY-MM-DD *

2021-01-01

Ending date of the dataset in the format YYYY-MM-DD *

2021-12-12

Do you have data table ?

Upload all data files wich are data table - optional

accepted formats

25: dataTables

Input a data collection.

Give a short description of your data file. If there is multiple data files please separate your descriptions with a comma (,) in the same order of your data input - optional

What's the quotes of your data table ? Repeat this parameter as many time as the number of data tables you've input and in the same order

+ Insert What's the quotes of your data table ? Repeat this parameter as many time as the number of data tables you've input and in the same order

Give the publicly accessible URL from which your data table can be downloaded. If more than one, data files please separate your URLs with a comma (,) in the same order of your data input. If wanting to include URLs for some but not all then use a - for those that don't have a URL.
- optional

Do you have raster data ?

Upload all data files wich are spatial raster - optional

accepted formats

21: spatialRaster

Input a data collection.

Give a short description of your data file. If there is multiple data files please separate your descriptions with a comma (,) in the same order of your data input - optional

https://biodiversitydata.github.io/

Creating an EML metadata record from EML metadata template files in text format

Steps

- Import the data, then click on the "home" icon at the top of the page and/or refresh the page.
- File review / What information is there (by looking at just the file names first, then browsing the contents of the metadata templates)?
- EML creation from EML Assembly Line template files using the Galaxy Make EML tool
- **MetaSHRIMPS: FAIRness assessment of metadata and creation of a draft data paper**
    - https://ecology.usegalaxy.eu/root?tool_id=interactive_tool_metashrimps on the generated EML
    - Execute => Draft of data paper + FAIR assessment
        - The draft “Data Paper” allows you to see the different metadata elements presented in the form of a static web page, a bit like in a data/metadata catalogue but without waiting for sending to the catalogue administrators and putting online!
        - We can see that some attributes have definitions added manually, while others have been automatically populated by EML Assembly Line
            - We could modify them afterwards to improve this
        - You can download an editable version of the draft data paper, to have a shareable or modifiable word version that can serve as a basis for writing an internal document or for wide distribution such as a data paper.
        - FAIRness score 30 success / 10 failure / 5 warning
            - On the “warning” side, we note a point on the size of the summary which would facilitate the “findable” aspects, another on the definition of an attribute “Present.Surface.pH” which only contains 3 words and would improve the “reusable” aspects.
            - On the "failure" side, it's worth noting the lack of semantic annotation.

Creating an EML metadata record from EML metadata template files in text format

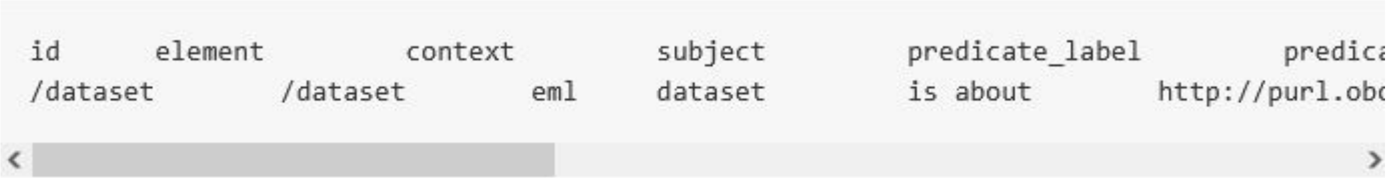
Steps

- Import the data, then click on the "home" icon at the top of the page and/or refresh the page.
- File review / What information is there (by looking at just the file names first, then browsing the contents of the metadata templates)?
- EML creation from EML Assembly Line template files using the Galaxy Make EML tool
- MetaSHRIMPS: FAIRness assessment of metadata and creation of a draft data paper
- **How to improve the files and get a better EML?**
  - Added annotation (EML report) + modified abstract and attributes (MetaShRIMPS report)

Creating an EML metadata record from EML metadata template files in text format

Steps

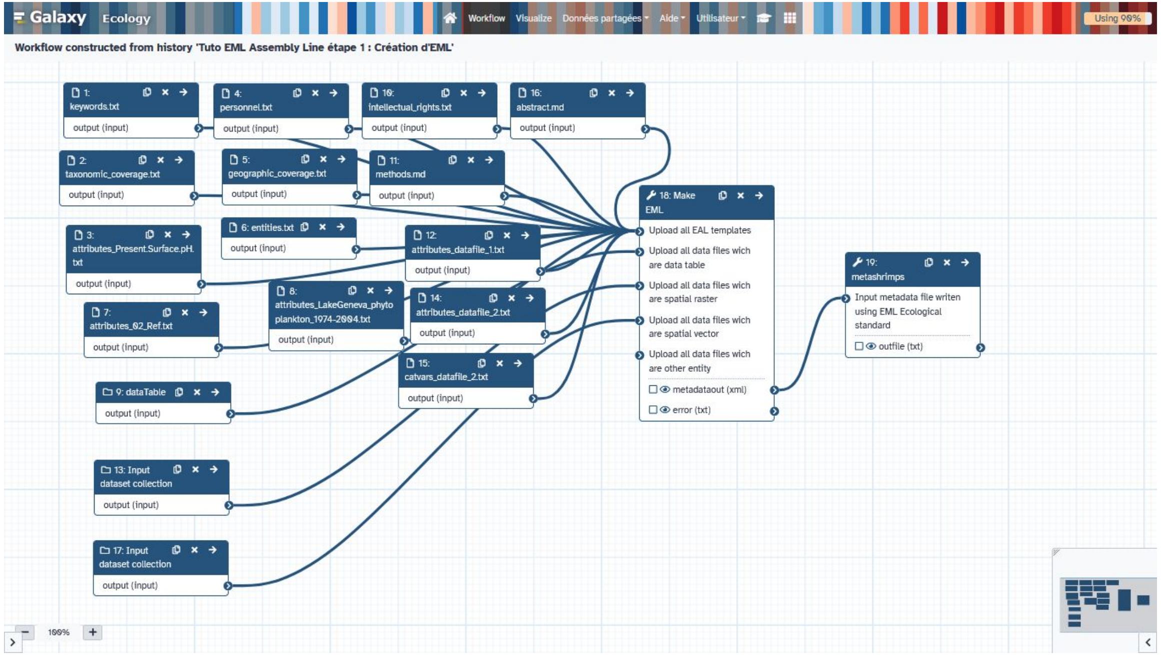
- Import the data, then click on the "home" icon at the top of the page and/or refresh the page.
- File review / What information is there (by looking at just the file names first, then browsing the contents of the metadata templates)?
- EML creation from EML Assembly Line template files using the Galaxy Make EML tool
- MetaSHRIMPS: FAIRness assessment of metadata and creation of a draft data paper
- How to improve the files and get a better EML?
- **Edition de fichiers de template EML Assembly Line pour améliorer le score de FAIRitude**
  - editing the “abstract.txt” file to obtain a summary of more than 100 words, of the “attributes_Present.Surface.pH.txt” file to expand the description of the “Present.Surface.pH” attribute (“surface present pH” instead of “ph” for example).
  - Creation of an “annotations.txt” file to add a keyword to the metadata record, “is about” “biodiversity”
  - => We recreate an EM
  - => We re-evaluate the “FAIRness” of our metadata via MetaShRIMPS (33 successes / 9 failures / 3 warnings)



```
id      element         context         subject         predicate_label         predica
/dataset        /dataset        eml     dataset         is about        http://purl.ob
```

Creating an EML metadata record from EML metadata template files in text format

You can create a workflow from your history by clicking on the “gear” in the top right corner of the history, then “extract workflow”.


Galaxy
Ecology
Workflow
Visualize
Données partagées
Aide
Utilisateur
Using 99%
Workflow constructed from history 'Tuto EML Assembly Line étape 1 : Création d'EML'
1: keywords.txt
output (input)
4: personnel.txt
output (input)
10: intellectual_rights.txt
output (input)
16: abstract.md
output (input)
2: taxonomic_coverage.txt
output (input)
5: geographic_coverage.txt
output (input)
11: methods.md
output (input)
18: Make EML
Upload all EAL templates
Upload all data files wich are data table
Upload all data files wich are spatial raster
Upload all data files wich are spatial vector
Upload all data files wich are other entity
metadataout (xml)
error (txt)
3: attributes_Present.Surface.pH.txt
output (input)
6: entities.txt
output (input)
12: attributes_datafile_1.txt
output (input)
19: metashrimps
Input metadata file writen using EML Ecological standard
outfile (txt)
7: attributes_02_Ref.txt
output (input)
8: attributes_LakeGeneva_phytoplankton_1974-2004.txt
output (input)
14: attributes_datafile_2.txt
output (input)
15: catvars_datafile_2.txt
output (input)
9: dataTable
output (input)
13: Input dataset collection
output (input)
17: Input dataset collection
output (input)
100%

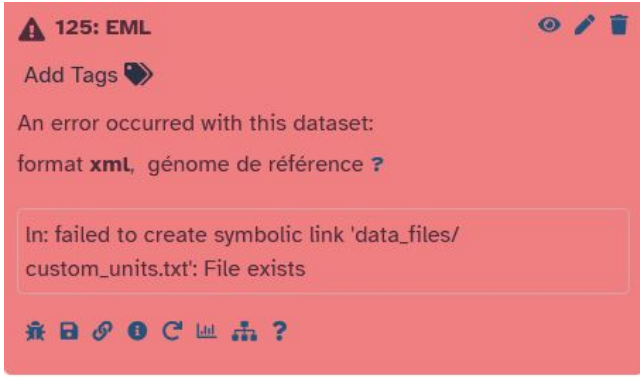
https://ecology.usegalaxy.eu/u/ylebras/w/workflow-constructed-from-history-tuto-eml-assembly-line-tape-1--cration-deml

=> But how do you create these famous template files?

Creation of EML Assembly Line metadata template files

Galaxy history with input data:
https://ecology.usegalaxy.eu/u/ylebras/h/eml-assembly-line-tape-2--cration-des-fichiers-modeles-de-mtadonnes

- Galaxy EML Assembly Line tools for generating:
  - Lists of attributes
    - For data tables
    - For raster GIS files
    - For vector GIS files
      - Surval data from the Quadrige database retrieved on November 30, 2021 for the REPHY program
  - Geographic coverage
    - If this does not appear to work, carefully review the error message (by clicking on the "insect" icon in the bottom left of the data file preview) to determine the source of the problem and suggest a solution.



125: EML

Add Tags

An error occurred with this dataset:

format **xml**, génome de référence **?**

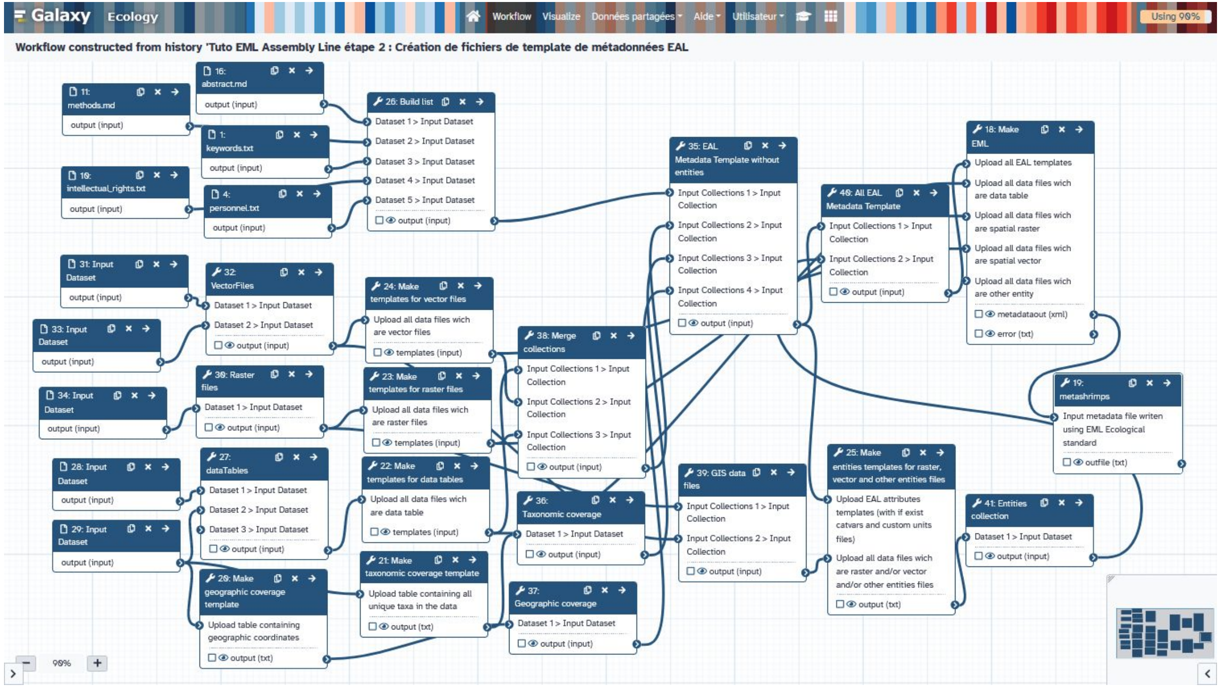
```
In: failed to create symbolic link 'data_files/
custom_units.txt': File exists
```

Creation of EML Assembly Line metadata template files

- Galaxy EML Assembly Line tools for generating:
  - Lists of attributes
  - Geographic coverage
  - **Taxonomic coverage**
  - Regarding the Temporal Coverage => as previously mentioned, this must be provided when creating the EML, along with the title
- We are seeing missing information in the metadata template, which explains why not all data files are populated in the EML.
  - Find and fill in the missing information by navigating to the "Warning message" file.

Creation of EML Assembly Line metadata template files

https://ecology.usegalaxy.eu/u/ylebras/w/copy-of-workflow-constructed-from-history-tuto-eml-assembly-line-tape-1--cration-deml



Workflow constructed from history 'Tuto EML Assembly Line étape 2 : Création de fichiers de template de métadonnées EAL

Galaxy
Ecology
Workflow
Visualize
Données partagées
Aide
Utilisateur
Using 99%
11: methods.md
output (input)
16: abstract.md
output (input)
1: keywords.txt
output (input)
10: intellectual_rights.txt
output (input)
4: personnel.txt
output (input)
26: Build list
Dataset 1 > Input Dataset
Dataset 2 > Input Dataset
Dataset 3 > Input Dataset
Dataset 4 > Input Dataset
Dataset 5 > Input Dataset
output (input)
35: EAL Metadata Template without entities
Input Collections 1 > Input Collection
Input Collections 2 > Input Collection
Input Collections 3 > Input Collection
Input Collections 4 > Input Collection
output (input)
40: All EAL Metadata Template
Input Collections 1 > Input Collection
Input Collections 2 > Input Collection
output (input)
18: Make EML
Upload all EAL templates
Upload all data files wich are data table
Upload all data files wich are spatial raster
Upload all data files wich are spatial vector
Upload all data files wich are other entity
metadataout (xml)
error (txt)
31: Input Dataset
output (input)
32: VectorFiles
Dataset 1 > Input Dataset
Dataset 2 > Input Dataset
output (input)
24: Make templates for vector files
Upload all data files wich are vector files
templates (input)
33: Input Dataset
output (input)
38: Merge collections
Input Collections 1 > Input Collection
Input Collections 2 > Input Collection
Input Collections 3 > Input Collection
output (input)
30: Raster files
Dataset 1 > Input Dataset
output (input)
23: Make templates for raster files
Upload all data files wich are raster files
templates (input)
34: Input Dataset
output (input)
19: metashrimps
Input metadata file writen using EML Ecological standard
outfile (txt)
28: Input Dataset
output (input)
27: dataTables
Dataset 1 > Input Dataset
Dataset 2 > Input Dataset
Dataset 3 > Input Dataset
output (input)
22: Make templates for data tables
Upload all data files wich are data table
templates (input)
39: GIS data files
Input Collections 1 > Input Collection
Input Collections 2 > Input Collection
output (input)
25: Make entities templates for raster, vector and other entities files
Upload EAL attributes templates (with if exist catvars and custom units files)
Upload all data files wich are raster and/or vector and/or other entities files
output (txt)
29: Input Dataset
output (input)
36: Taxonomic coverage
Dataset 1 > Input Dataset
output (input)
41: Entities collection
Dataset 1 > Input Dataset
output (input)
20: Make geographic coverage template
Upload table containing geographic coordinates
output (txt)
21: Make taxonomic coverage template
Upload table containing all unique taxa in the data
output (txt)
37: Geographic coverage
Dataset 1 > Input Dataset
output (input)
99%

https://biodiversitydata.github.io/

Creating tab-delimited text metadata template files from an EML record

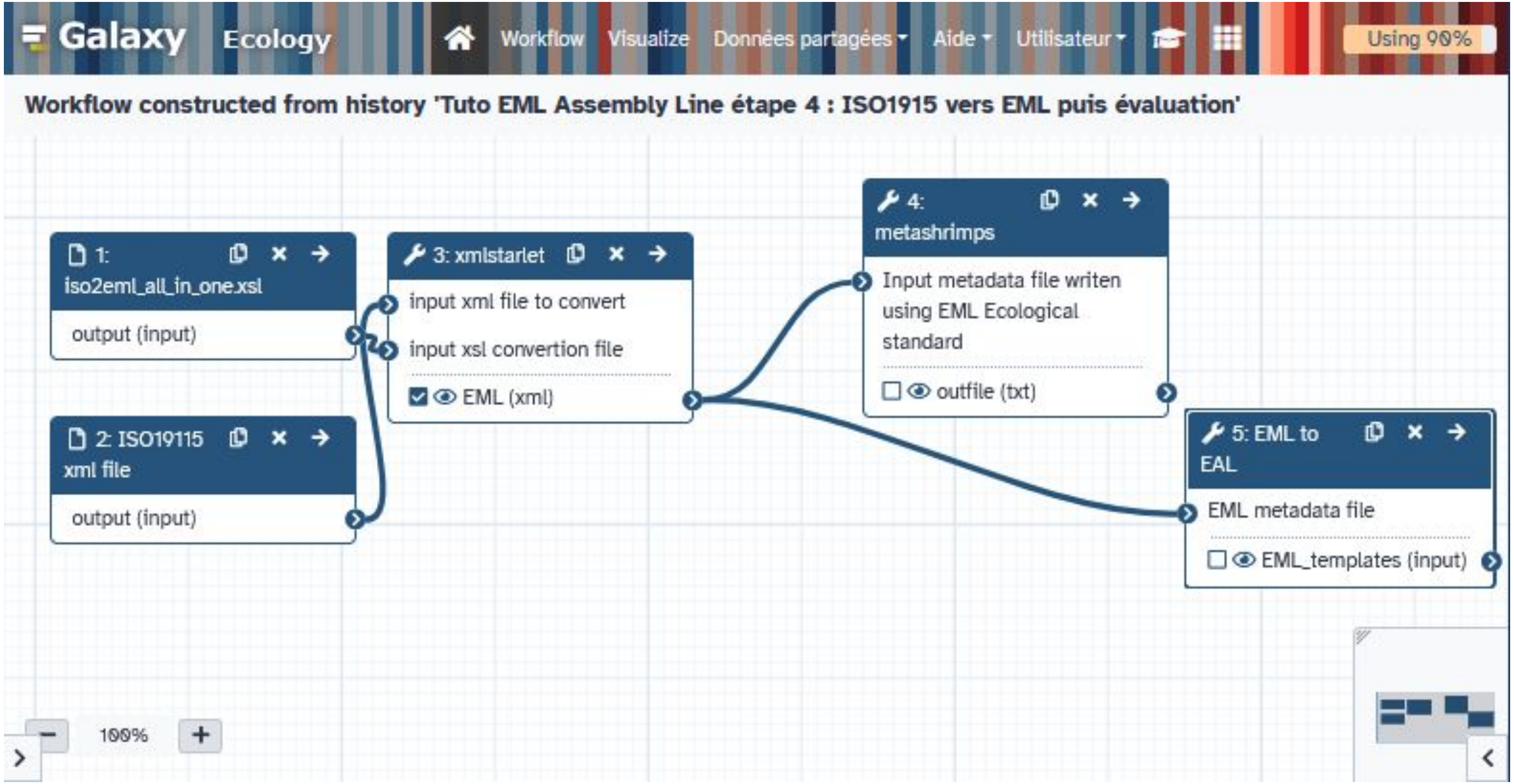
- From the Kakila database's EML metadata record: https://data.pndb.fr/view/doi%3A10.48502%2F8bb5-pk85
  - Copy and paste the direct link URL to the metadata record into the Galaxy upload module / "paste/fetch data" field: https://pndb.fr/metacat/d1/mn/v2/object/doi%3A10.48502%2F8bb5-pk85 Product analysis history
- https://ecology.usegalaxy.eu/u/ylebras/h/tuto-eal-3-eml-to-eal

Creating an EML record from an ISO19115 metadata record

Galaxy history with input data:
https://ecology.usegalaxy.eu/u/ylebras/h/tuto-eml-assembly-line-tape-4--iso1915-vers-eml-puis-valuation-fairitude-et-bauche-de-data-paper

- Using the “xmlstarlet” tool to convert the SURVAL ISO metadata sheet accessible from this webpage https://surval.ifremer.fr/Donnees/Cartographie-Inventaire-du-reseau-REPHY#/metadata/aa8fe568-d2c0-4b53-a8bb-d9fcef2b5293 (direct link to metadata sheet: https://sextant.ifremer.fr/geonetwork/srv/api/records/aa8fe568-d2c0-4b53-a8bb-d9fcef2b5293/formatters/xml) to EML using the conversion file “iso2eml_all_in_one.xsl”
- Creation of a “FAIritude” report and a draft “data paper” via “MetaSHRIMPS”
- Creation of metadata template files in tab-delimited text format via the “EML to EAL” tool

Creating an EML record from an ISO19115 metadata record



Workflow constructed from history 'Tuto EML Assembly Line étape 4 : ISO1915 vers EML puis évaluation'

1: iso2eml_all_in_one.xsl
output (input)

2: ISO19115 xml file
output (input)

3: xmlstarlet
input xml file to convert
input xsl convertion file
☑ EML (xml)

4: metashrimps
Input metadata file writen using EML Ecological standard
☐ outfile (txt)

5: EML to EAL
EML metadata file
☐ EML_templates (input)

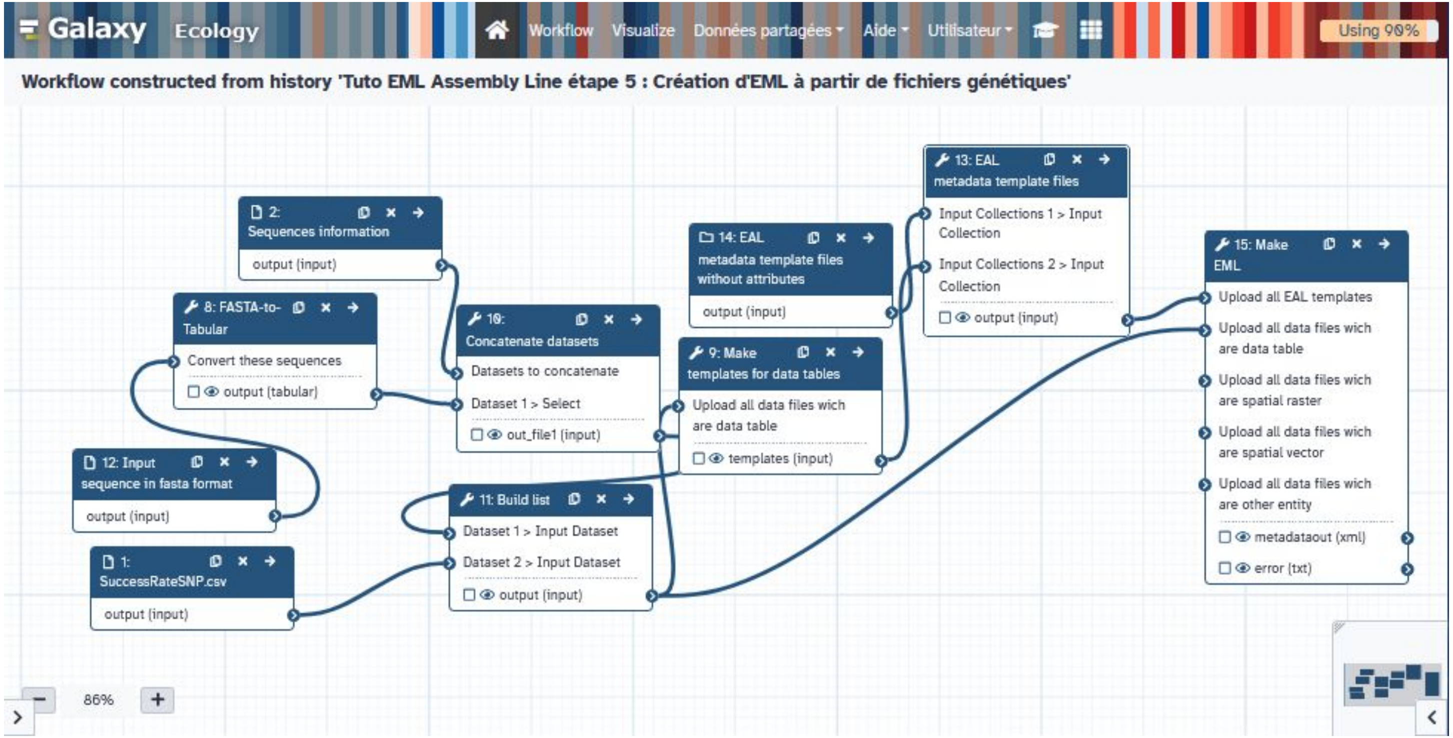
https://ecology.usegalaxy.eu/u/ylebras/w/tuto-eml-assembly-line-tape-4--iso1915-vers-eml-puis-valuation

Creation of an EML record for genetic data

Galaxy history with input data:
https://ecology.usegalaxy.eu/u/ylebras/h/tuto-eml-assembly-line-tape-5--cration-deml--partir-de-fichiers-gntiques

- Using SEANOE data DOI 10.17882/70546 https://www.seanoe.org/data/00593/70546/ Le Cam et al 2019
- From sequence data in FASTA format, identify a natural segmentation of sequence names into information elements, which we will call metadata elements, and create a tabulated file from this sequence file with one column for the sequence and as many columns as there are metadata elements.
- Looking at the FASTA file, we notice that there appear to be 7 metadata elements:
    - >RCL_P_1 ACATAACTTAGAGAAAGGAGCAGAGATCAGGGAAGGGGACAGCAACAAAG 2705094059 RCL_P_1_0_T_U_2705094059 23/07/2018 InfiniumII 1
- From the sequence file, we can therefore create a tabulated file using the FASTA TO TABULAR tool, ultimately obtaining a file with 8 columns.
- Next, you can add a header to the file using the Sequences Information history file and the “Concatenate dataset tail to end” tool.
- Then, you can create EAL metadata template files of type DataTable for this tab-delimited “SequencesSNP_Raja_clavata” file and the “SuccessRateSNP.csv” file, and then modify the content using the metadata provided in the “68857.txt” file, particularly regarding attribute descriptions.
- You can then use the content of the SEANOE metadata record “seanoe_metadata_export_20240312164552.txt” to populate the necessary information in the EAL template files. This step can be done entirely manually, but it is also possible to create a workflow to automate it using this file. If you choose this option, you will need to be or become a Galaxy Jedi, but this will allow you to share it with others and offer an automated workflow to go from a SEANOE txt quote file to EAL template files.

Creation of an EML record for genetic data



Workflow constructed from history 'Tuto EML Assembly Line étape 5 : Création d'EML à partir de fichiers génétiques'

Galaxy
Ecology
Workflow
Visualize
Données partagées
Aide
Utilisateur
Using 90%
2: Sequences information
output (input)
8: FASTA-to-Tabular
Convert these sequences
output (tabular)
12: Input sequence in fasta format
output (input)
1: SuccessRateSNP.csv
output (input)
10: Concatenate datasets
Datasets to concatenate
Dataset 1 > Select
out_file1 (input)
11: Build list
Dataset 1 > Input Dataset
Dataset 2 > Input Dataset
output (input)
14: EAL metadata template files without attributes
output (input)
9: Make templates for data tables
Upload all data files wich are data table
templates (input)
13: EAL metadata template files
Input Collections 1 > Input Collection
Input Collections 2 > Input Collection
output (input)
15: Make EML
Upload all EAL templates
Upload all data files wich are data table
Upload all data files wich are spatial raster
Upload all data files wich are spatial vector
Upload all data files wich are other entity
metadataout (xml)
error (txt)
86%

https://ecology.usegalaxy.eu/u/ylebras/w/workflow-constructed-from-history-tuto-eml-assembly-line-tape-5--cration-deml--partir-de-fichiers-gntiques

Perspectives

- **Transformation and Enrichment**
  - => Offers a simplified and collaborative way to share, convert, and enrich metadata
  - Identified need: Work on mappings and conversions between standards!
- **Automation by humans and/or machines**
  - => Use of the Galaxy API by external services, in particular
- **Populating data warehouses with rich metadata** without modifying the software used by the warehouses
  - => Using EAL template metadata files