

➤ **Projet MENHIR**

Étude de faisabilité de connexion directe entre deux outils : OpenSILEX et ASTERICS

Elise Maigné (MIAT, Toulouse, CATI B4B)

Isabelle ALIC (MISTEA, Montpellier, CATI CODEX)

Séminaire inter-CATI | PEPI-2G
04, 05 et 06 mai 2026 - Lyon



MENHIR

Etudier la possibilité d'une connexion directe entre OpenSILEX et ASTERICS

- ASTERICS, MIAT, Toulouse et Cellule OpenSILEX, MISTEA, Montpellier
- Données de cas d'usage fournies par le LIPME, Toulouse (C. Donnadiou et N. Blanchet)

Pourquoi ?

- OpenSILEX : intégration d'un outil d'exploration clic-bouton pour les utilisateurs
- ASTERICS : élargissement de la communauté des utilisateurs à la communauté OpenSILEX

4000 euros de budget

- stage de 4 mois (2025 - Paulin Kodjovi)
- déplacements Toulouse-Montpellier

MENHIR



OpenSILEX

OpenSILEX : Système d'Information pour
L'Expérimentation

<https://opensilex.org/>

Instances :

PHIS (phénotypage végétal à haut débit),

Sixtine (unités expérimentales du végétal à INRAE),

SunAgri (agrophotovoltaïque),

Alfis (données de fermentation alcoolique),

Silex-VitiOeno (vin et viticulture),

ENVIBIS (dépollution),

SIDURI (fermentation).



ASTERICS : Application web pour l'exploration
et l'intégration (statistique) de données

<https://asterics.miat.inrae.fr/>

Objectif du projet

Selected experiments 1 Display ▾

Showing 0 to 1 of 1 entries

<input type="checkbox"/>	↕ Name	↕ Start date	↕ End date
<input checked="" type="checkbox"/>	23HP04_HELEX_STRESS	2023-06-27	2023-12-19

« ‹ 1 › »



asterics My workspace Edit Explore Integrate Documentation

Workflow

```

    graph TD
      Index[Index generic Rows: 23365 #col.: 0] --> Editor1[editor_1]
      Editor1 --> Index_edited_1[Index_edited_1 generic Rows: 23365 #col.: 0]
      Editor1 --> Meteo_bee[Meteo_bee generic Rows: 23365 #col.: 21]
      Editor1 --> Impact_phen[Impact_phen generic Rows: 149 #col.: 14]
      Editor8[editor_8] --> Impact_phen_edited[Impact_phen_edited generic Rows: 149 #col.: 14]
  
```

Dataset "impact_phen"

	# rows	# col.	# missing	% missing	# numeric	# cat.	logt	normalized
	162	14	0	0	7	7	no	no

	BE_NB_AI	date_plant_height	PLANT_HEIGHT	date_collar_diam	COLLAR_DIAM	date_leaf_nb_tot	LEAF_NB_TOT	date_leaf_nb_sen	LEAF_NB_SEN	genotype
	Rows	Date	Rows	Date	Rows	Date	Rows	Date	Rows	Rows
23HP0	0.3571	2023-08-21	730	2023-08-21	10.67	2023-08-21	16	2023-08-21	8	ANN1054
23HP0	0.0948	2023-08-14	548	2023-08-14	8.65	2023-08-14	13	2023-08-14	3	ANN1054
23HP0	0.9684	2023-08-29	715	2023-08-29	7.61	2023-08-29	21	2023-08-29	14	ANN1054
23HP0	0.1429	2023-09-01	795	2023-09-01	12.21	2023-09-01	28	2023-09-01	11	ANN1184
23HP0	0.1387	2023-09-07	1045	2023-09-07	12.64	2023-09-07	28	2023-09-07	11	ANN1184
23HP0	0.28	2023-08-29	740	2023-08-29	12.49	2023-08-29	27	2023-08-29	9	ANN1184
23HP0	0.1639	2023-09-07	1080	2023-09-07	12.52	2023-09-07	28	2023-09-07	11	ANN1430
23HP0	0.1735	2023-09-01	920	2023-09-01	11.39	2023-09-01	24	2023-09-01	11	ANN1430
23HP0	1.7104	2023-08-22	780	2023-08-22	12.14	2023-08-22	19	2023-08-22	8	ANN1430
23HP0	0.4314	2023-09-08	730	2023-09-08	13.25	2023-09-08	24	2023-09-08	6	XHQ

Let's explore a dataset!



Explore variables in a dataset

Obtain numerical summaries and plots for a few variables.

Explore a dataset with PCA

Perform Principal Component Analysis on a dataset.

Explore a dataset with a heatmap

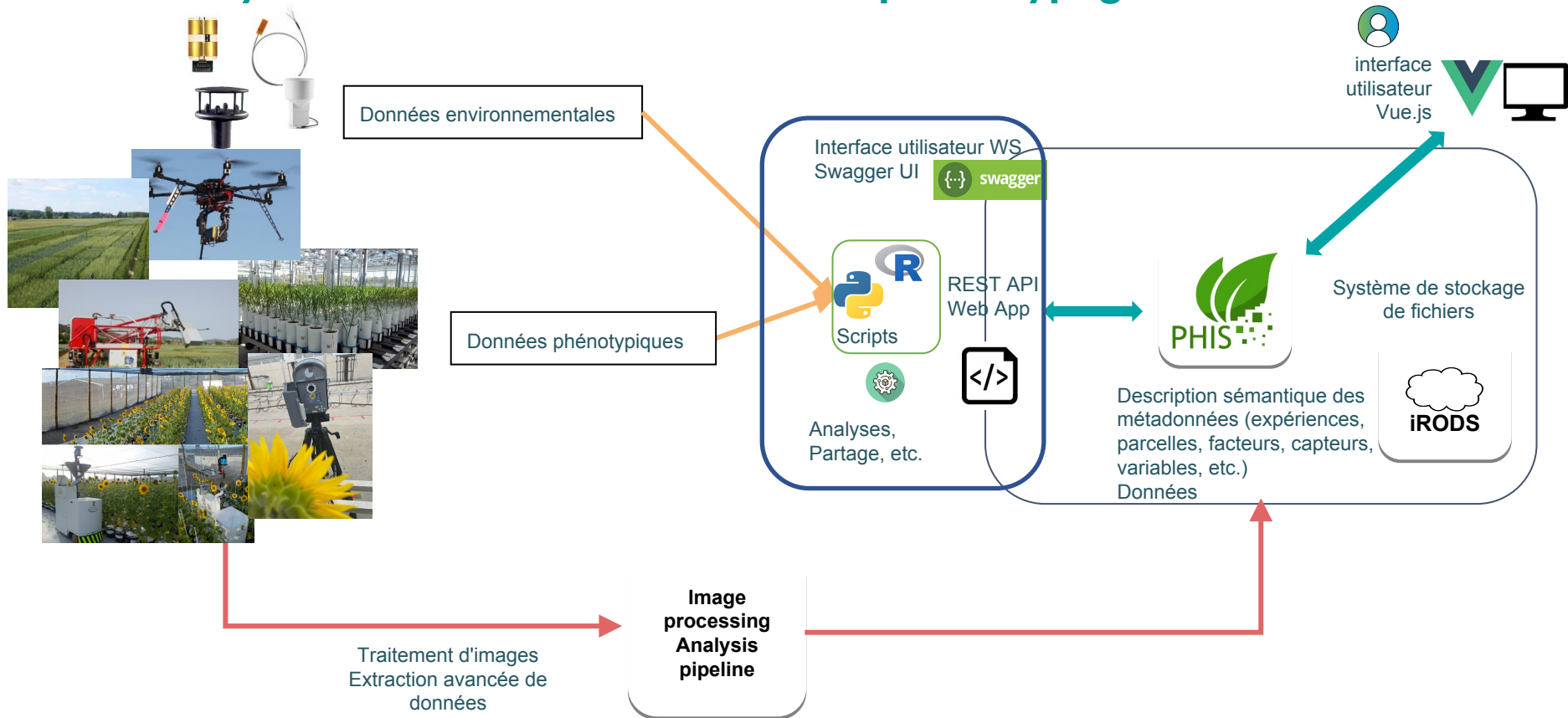
Obtain the heatmap of a dataset.



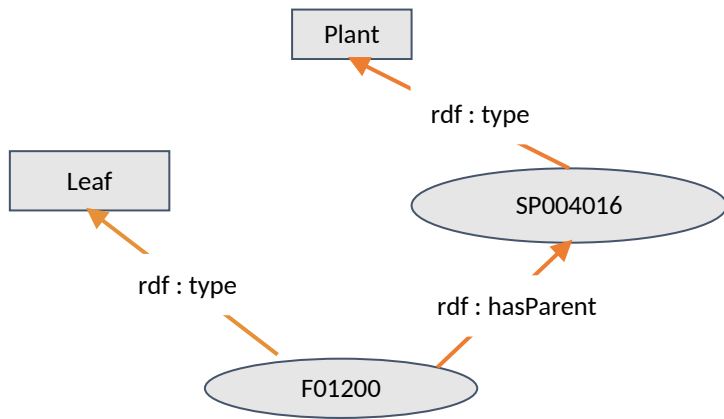
OpenSILEX

OpenSILEX - (Système d'Information pour L'Expérimentation) - <https://opensilex.org/>
Une suite logicielle pour la gestion FAIR des données expérimentales

PHIS : un système d'information dédié au phénotypage



Gestion des données dans PHIS



Triplestore -> métadonnées
(RDF4J, GraphDB)

```
{
  "uri": "opensilex-test:id/data/fbe78545-420a-4be8-b5c3-b2174a1a8023",
  "date": "2018-07-30T20:07:00.000Z",
  "target": "opensilex-test:id/scientific-object/so-sp004016",
  "variable": "opensilex-test:id/variable/plant_maximumquantumyield_computation_unitless",
  "value": 0.82,
  "confidence": null,
  "provenance": {
    "uri": "opensilex-test:id/provenance/fc_plant_sweetpotato",
    "prov_used": null,
    "prov_was_associated_with": null,
    "settings": null,
    "experiments": [
      "opensilex-test:id/experiment/sweetpotatoviruses_2018"
    ]
  }
},
"metadata": null,
"publisher": null,
"raw_data": null,
"issued": null,
"modified": null
}
```

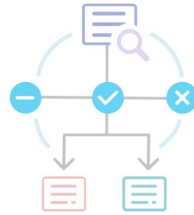


NoSQL -> données, géospatial
(MongoDB)

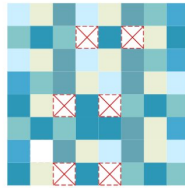


Ce qu'on peut faire avec ASTERICS

Let's edit data!



Dataset edition

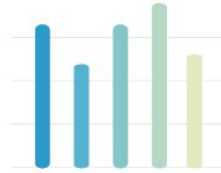


Missing values



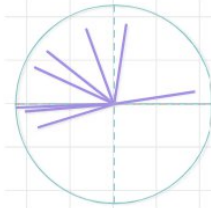
Normalize dataset

Let's explore a dataset!



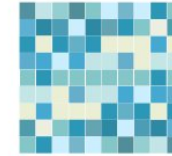
Explore variables in a dataset

Obtain numerical summaries and plots for a



Explore a dataset with PCA

Perform Principal Component Analysis on a



Explore a dataset with a heatmap

Obtain the heatmap of a dataset.



Clustering

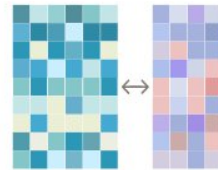
Cluster the individuals of a dataset.



Self-Organizing Map

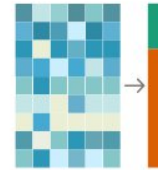
Use SOM as a clustering and visualization method.

Let's integrate data!



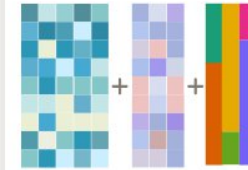
Integrate two datasets with PLS

Performs Partial Least Squares analysis on two datasets.



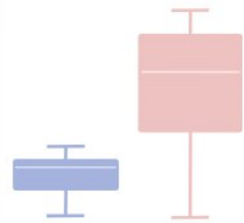
Integrate two datasets with PLS-DA

Performs Partial Least Squares analysis on two datasets.



Integrate datasets with MFA

Perform Multiple Factor Analysis on several datasets.



Differential analysis

Perform differential analysis for all numeric variables of a dataset.

ASTERICS



asterics

FRONTEND



BACKEND

PyRserve



NGINX



ASTERICS USERS SESSIONS

Rserve



The screenshot shows the Asterics web interface. At the top, there's a navigation bar with 'My workspace', 'Edit', 'Explore', 'Integrate', and 'Documentation'. Below that, a workspace summary shows 'toto' created on 'May 19, 2026' with 4 datasets and 2 analyses. A 'Workflow' section displays a flowchart: 'mrna' (microarray, 988 rows, 2000 columns) and 'protein' (generic, 379 rows, 142 columns) feed into 'PCAObj_1' (PCA axes, 379 rows, 5 columns). 'clinical' (metadata, 989 rows, 217 columns) and 'PCAObj_1' feed into 'combinedDF_1' (combined datasets, 379 rows, 217 columns). 'combinedDF_1' and 'PLSObj_1' (PLS axes, 379 rows, 5 columns) are also shown. Below the workflow, there are two tables: 'All datasets' and 'All analyses'.

Name	Type	Rows	Columns	Log.	Norm.
<input type="checkbox"/> mrna	microarray	988	2000	yes	yes
<input type="checkbox"/> protein	generic	379	142	no	no
<input type="checkbox"/> clinical	metadata	989	217	no	no
<input type="checkbox"/> PCA_axes_ncp5_p...	PCAaxes	379	5	nc	nc

Name	Function
<input type="checkbox"/> PCAObj_1	r_famd
<input type="checkbox"/> combinedDF_1	r_combine_datasets
<input type="checkbox"/> PLSObj_1	r_pls

Footer: Asterics v5.0 © 2022 — Terms of service — Legal notice — Optimized for firefox

INRAE

Données cas d'usage : Expérience 23HP04 (LIPME, INRAE, Toulouse)



Comptage de pollinisateurs (abeilles) sous 3 conditions :

- Stress hydrique
- Stress chaud
- Control

Sources de données gérées dans PHIS :

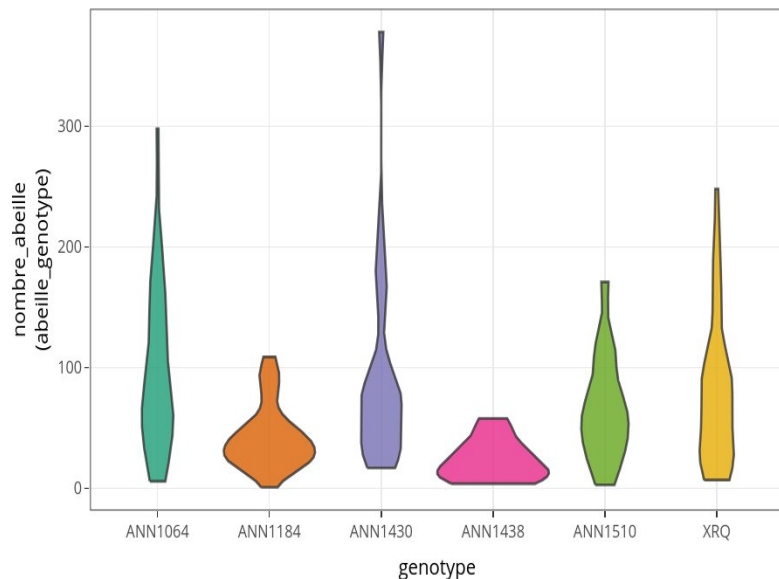
- Phénotypage des plantes
- Comptage de pollinisateurs
- Météo

Etapes du projet

1. **Extraction des données de PHIS** : création de modules python génériques pour extraire les données de PHIS dans un format compatible avec l'analyse dans ASTERICS
2. **Import des données dans ASTERICS**
3. **Étude statistique** : arrive-t-on à répondre aux questions biologiques ?
4. **Recommandations techniques**
 - pour le transfert du module python vers l'API OpenSILEX.
 - pour une compatibilité directe des données de phénotypage dans ASTERICS (nouveaux formats de données, ...)

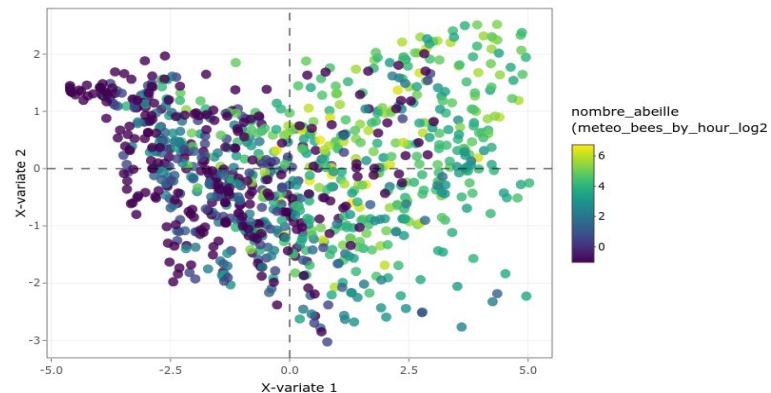
Exemples d'exploration de données

Distribution du nombre d'abeilles en fonction du Genotype

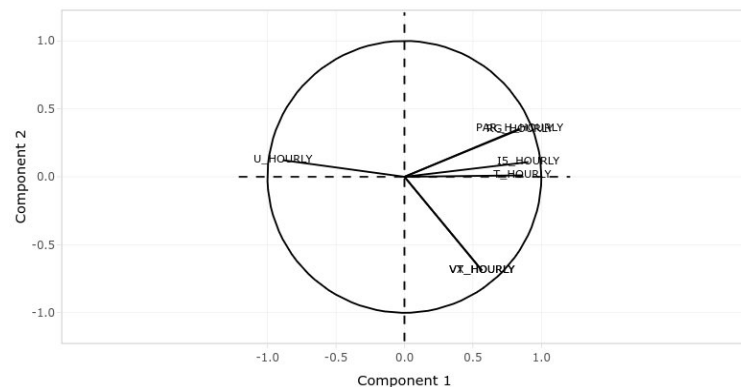


Sélection de variables à intégrer dans des modèles

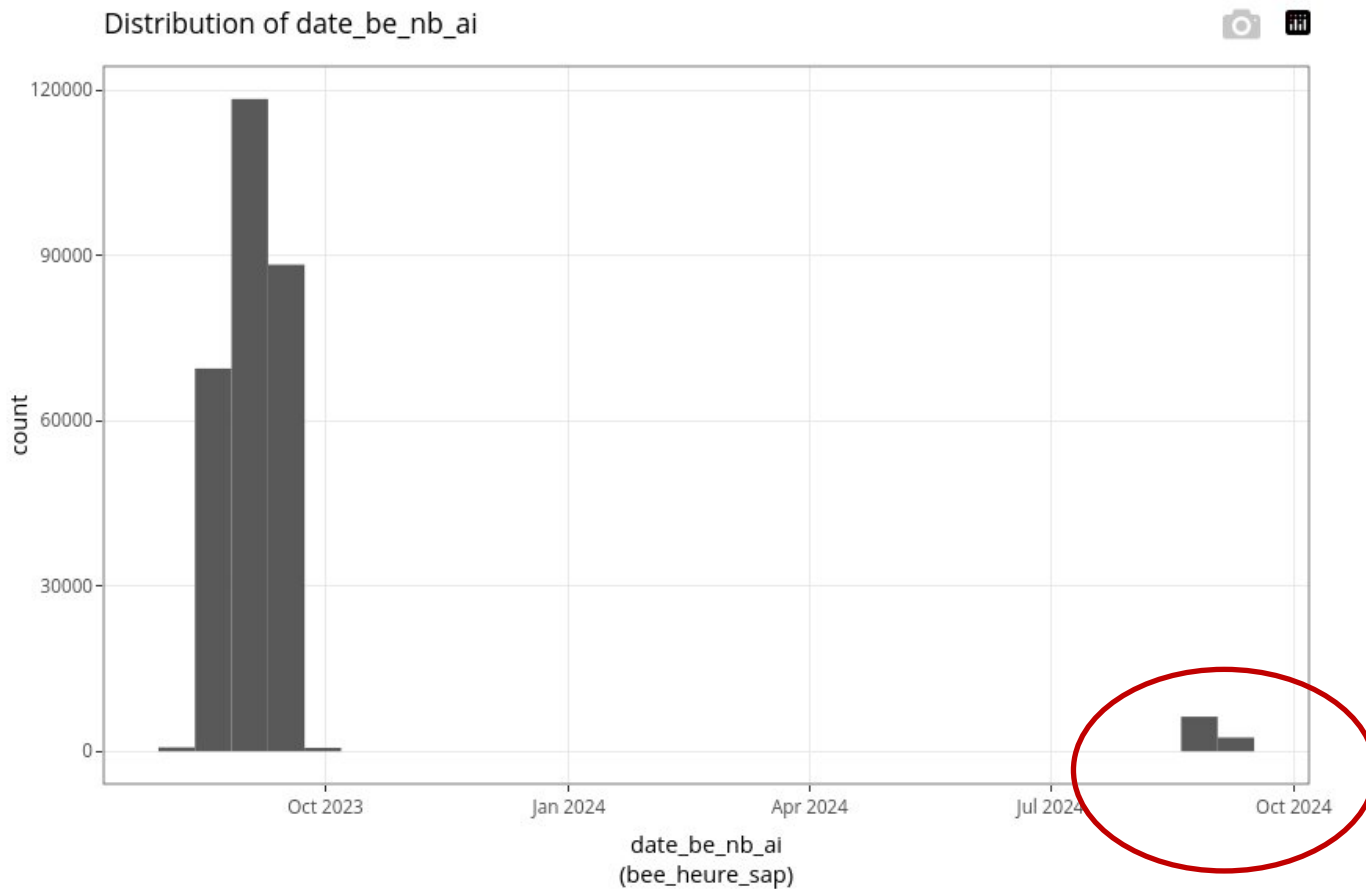
meteo_hour_scaled, nb_bee_hour_scaled - PLS: X-space



meteo_hour_scaled (vs nb_bee_hour_scaled, not represented) - PLS: correlations of
Correlation threshold: 0.8



Détection de valeurs aberrantes



INRAE

Exemple de recommandation technique

Module “Agrégation des données”

```
# date (passage du format str vers date) et on ne garde pas les min et sec
bee["date_photo"] = pd.to_datetime(bee["date_photo"]).dt.tz_localize(None).dt.floor("h")

# agrégation du nombre d'abeille sur l'heure (nombre total d'abeille par heure, toutes plantes confondues)
bee = bee.groupby(['date_photo']).agg(
    nombre_photo = ('nombre_abeille', 'count'),
    nombre_abeille_total = ('nombre_abeille', 'sum')
).reset_index()
```

Colonne sur laquelle se fait l'agrégation

Fonction d'agrégation

Unité

Objectif: pouvoir croiser les données d'abeilles avec les données météo
=> Nécessité d'avoir la même échelle de temps (l'heure)

Exemple de recommandation technique

Module “Agrégation des données”

```
# date (passage du format str vers date) et on ne garde pas les min et sec  
bee["date_photo"] = pd.to_datetime(bee["date_photo"]).dt.tz_localize(None).dt.floor("h")
```

```
# agrégation du nombre d'abeille sur l'heure (nombre total d'abeille par heure, toutes pla  
bee = bee.groupby(['date_photo']).agg(  
    nombre_photo = ('nombre_abeille', 'count'),  
    nombre_abeille_total = ('nombre_abeille', 'sum')  
) .reset_index()
```

Colonne sur laquelle se fait l'agrégation

Fonction d'agrégation

Unité

Quelques vérifications :

- Présence de valeurs manquantes
- Compatibilité entre Fonction d'agrégation et type de colonne (ex. colonne qualitative et fonction SUM)

Dataset edition

1. Select a dataset and an action

Dataset

Dataset ▾

Actions

- Transpose
- Change dataset nature
- Change variable (column) types
- Set individual (row) names
- Subset variables (columns)
- Subset individuals (rows) by name
- Subset individuals (rows) by category
- Rename categories
- Aggregate**
- Create individual (row) names from two columns**

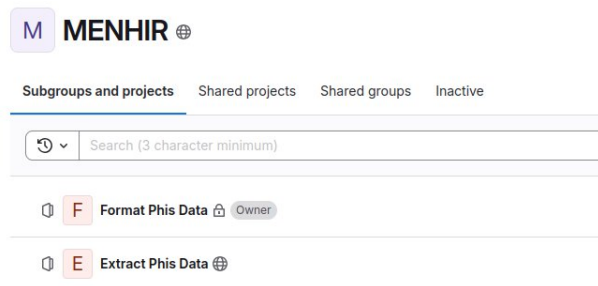


INRAE

Bilan : connexion entre OpenSILEX et ASTERICS faisable

Production de scripts d'extraction / formatage des données

- <https://forge.inrae.fr/groups/menhir>



Mais quelques limites liées au package OpenSILEX (API REST) :

- Temps d'extraction -> **Utiliser le nouvel outil SilexExplorer**
- Certains cas non pris en charge (tester sur d'autres instances avant de généraliser)

Intérêts identifiés sur le cas d'usage :

- Contrôle qualité des données visuel rapide
- Exploration des données aisée
- Intégration de différentes sources de données PHIS possible (cf. SilexExplorer)



Perspectives

Agrégation des données nécessaires :

- A faire par l'utilisateur manuellement dans ASTERICS

OU

- Automatisable si ajout du niveau d'agrégation pertinent côté OpenSILEX (échelle de temps minimum, fonction d'agrégation)

Ajout de modules dans ASTERICS pour :

- Agréger les données
- Créer un identifiant unique par combinaison de deux variables
- Fusionner deux jeux de données

Intégration du module Python dans l'API OpenSILEX

et

Appel de l'API dans ASTERICS (Importer une expérience depuis OpenSILEX)

Merci !

Des questions ?



OpenSILEX

+



asterics

=

MENHIR ♥