Software metadata extraction, consolidation and evaluation

> ⚠️ under development. Currently restructuring the whole project to align it with a clean architecture and to make it more modular and scalable.

We have developed a pipeline to **gather metadata about research software specific** to Computational Biology, **harmonize** and **integrate** it and to then be able to monitor certain features and **evaluate** their compliance with ** FAIRsoft indicators**. *FAIRsoft* are a set of research software *FAIRness* indicators, specifically devised to be assesed automatically.

This repository contains the code for:

- Harmonization of raw metadata.
- Integration of pieces of metadata belonging to the same software: `integration` use case.
- Calculation of *FAIRsoft* indicators compliance and FAIRsoft scores.

The code for the **previos steps** can be found in the respotories specified as follows:

- Data extraction: each importer, which is responsible for extracting metadata from a specific source, has a repository of its own:

    - [Bioconda importer](#)
    - [Bioconductor importer](#)
    - [Galaxy Toolshed importer](#)
    - [OpenEBench tools importer](#)
    - [OpenEBench metrics importer](#)
    - [Sourceforge importer](#)
    - [Repositories importer](#)

    > ❗ Given the complexity of the installation of the requirements of some of the importers, it is highly recommended to use the dockerized version of the different importers (in `registry.bsc.es/inb/elixir/software-observatory/`).

Installation

Install the dependencies

```
pip install -r requirements.txt
```

Usage

Data transformation

This is one [Use case](#) and can be executed from the [CLI](#) in the following way:

```
python3 src/adapters/cli/transformation.py -l INFO
```

Data storage

During the whole process, metadata is stored in a Mongo Database (INB Mongo oeb-research-software). The database connection is configured through environment variables.