Data harmonisation and integration

Module 1

- Mariantonietta La Marra

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

Research life cycle - data analysis

Studying ecological processes, making forecasts and giving support to decision making increasingly relies on the capability to process and analyse large collections of data held electronically.

Researchers make use of various forms of analyses, programming, modelling, specialty tools, custom-built simulation, scientific visualizations and analytical models .

A research lifecycle that includes (big) data analysis and various modelling approaches, if not managed correctly, can lead to discrepancies in the results.

Dealing with this challenge requires both data and analytical codes to be **available** and **reusable** by other researchers providing the capacity to easily and quickly access, re-use and eventually build upon existing information and scientific evidence and to generate novel knowledge.

VREs and Collaborative Coding Platforms

Research and development in Ecoinformatics focus on implementing new approaches, such as Virtual Research Environments (VREs) and collaborative coding platforms, to deploy these tools into an **easy-to-use** way, **accessible** via the web, and to **document** the set of processes used to reach scientific conclusions.

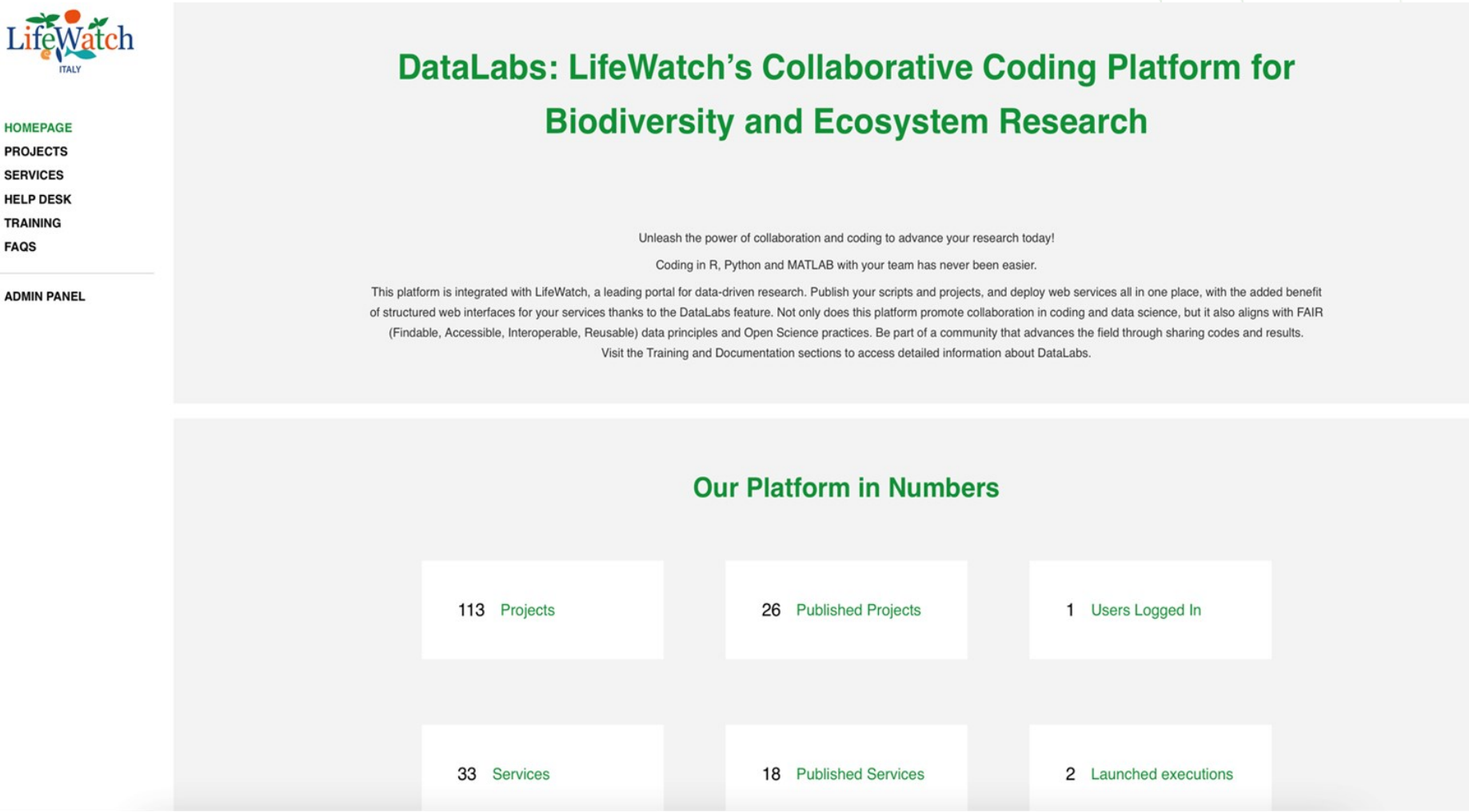
- global and seamless access to the data and services of interest, with no temporal or spatial constraints

- connected to the related infrastructures with both roles of consumer, benefitting from the services offered, including datasets, computing power and hosting machines, and provider, contributing with the results produced to the infrastructures offering

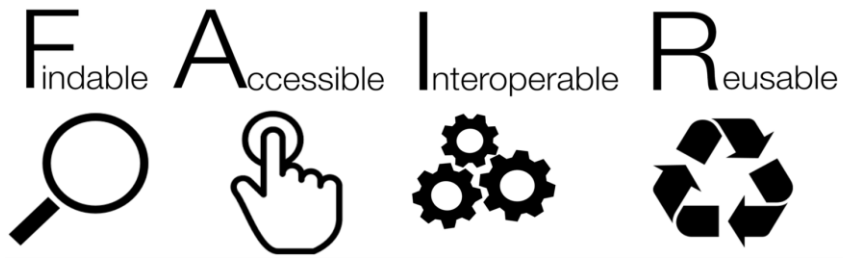
- promote fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance and attribution

- It is also important that these approaches enable the analytical components to be run on powerful distributed computing systems, such as Cloud-computing environments, to put at scientists' disposal a high **computational power** that can be utilized to perform collaborative analysis, visualisations or other computing intensive processes for data interpretation.

DataLabs: LifeWatch Italy Collaborative Coding Platform



DataLabs serves as a secure and standardized collaborative environment for designing and developing analysis and modeling software. The platform integrates several sections and tools to support the scientific community, in line with FAIR (Findable, Accessible, Interoperable, Reusable) data principles and Open Science practices.

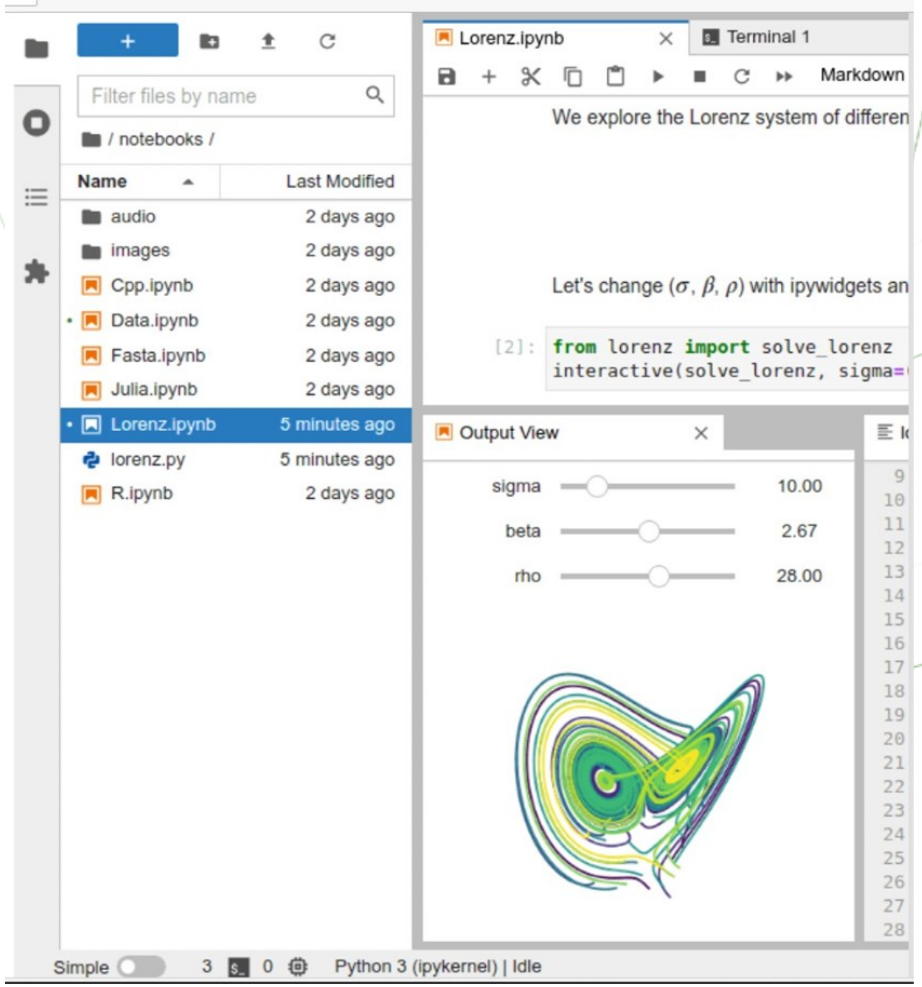


Key features include:

- User access and authentication system.
- Code documentation and metadata management
- Multiple programming languages commonly used in ecological modeling (R, Python and Matlab) within a Jupyter environment
- Web interface for collaborative code editing and execution

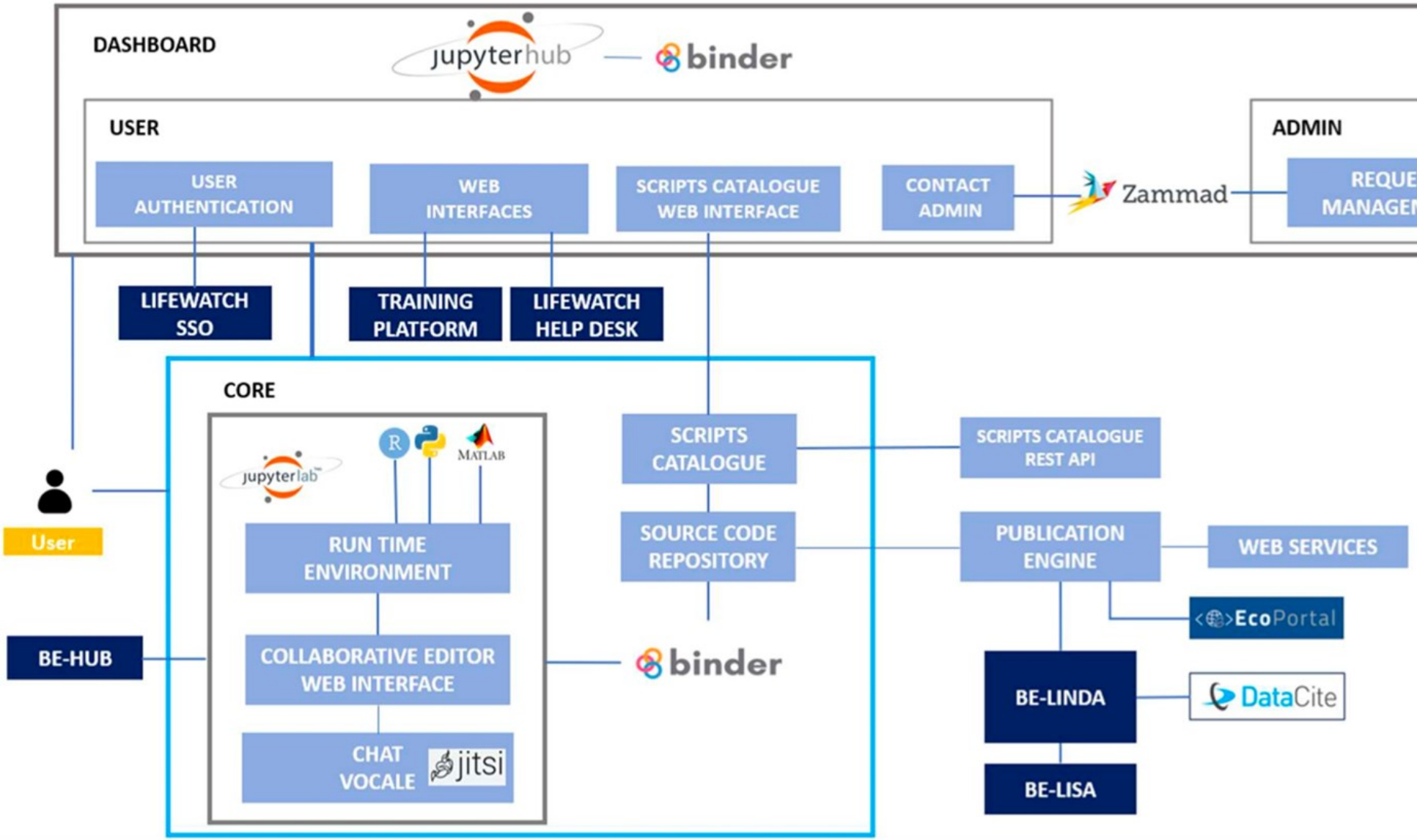
Jupyter is an open-source project, born out of the original IPython Project in 2014, that provides a web interface for interactive data science and scientific computing across multiple programming languages.

The notebook with cells allows to organize programs in different steps that interleave with visual outputs and markdown information.



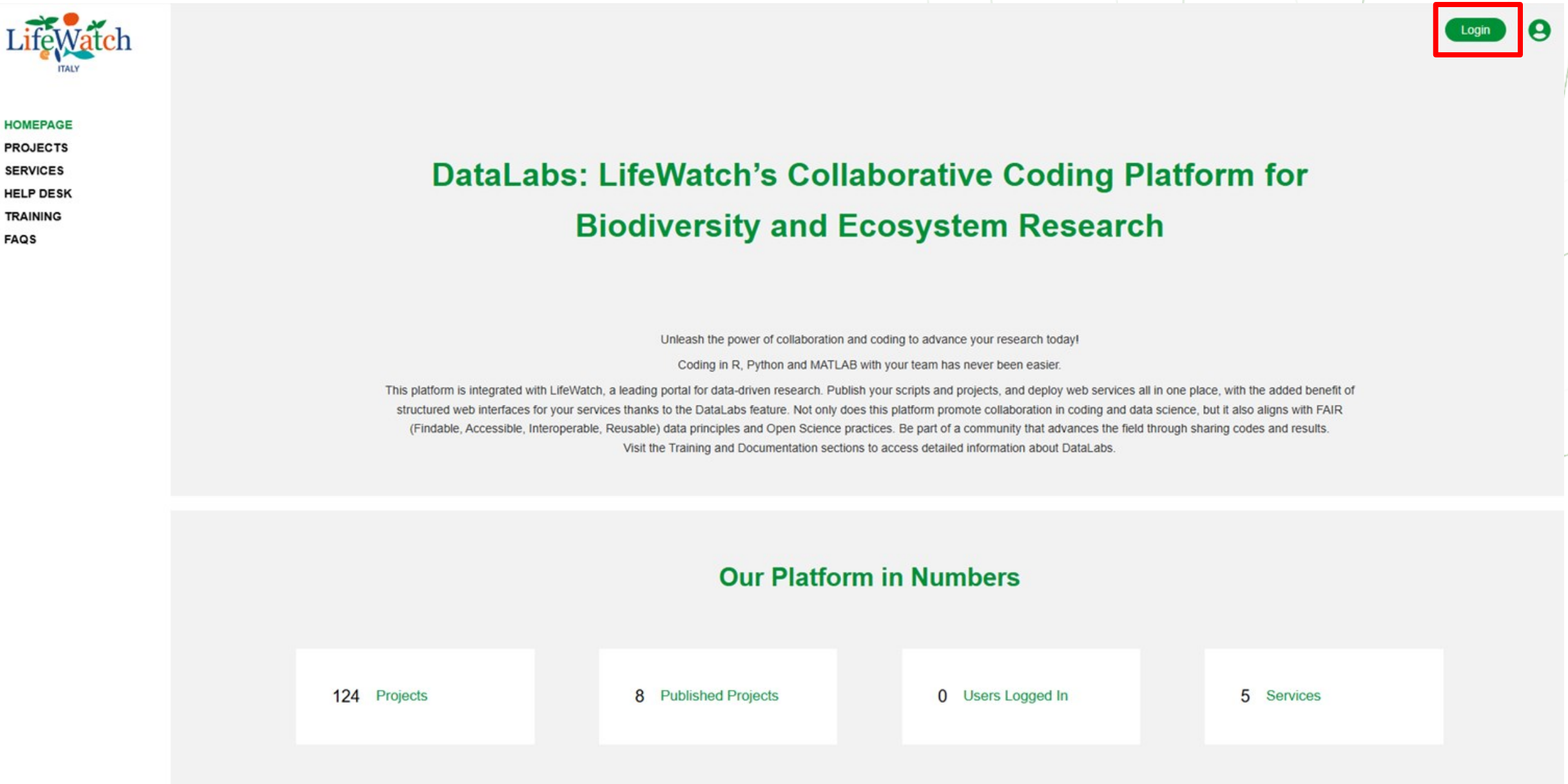
- Integrated version control with Git, enabling code sharing and collaboration.
- Support for a wide array of libraries and tools, with mechanisms for users to request library installations or updates.
- Access to help desk and training platform
- Project deployment as a web service.

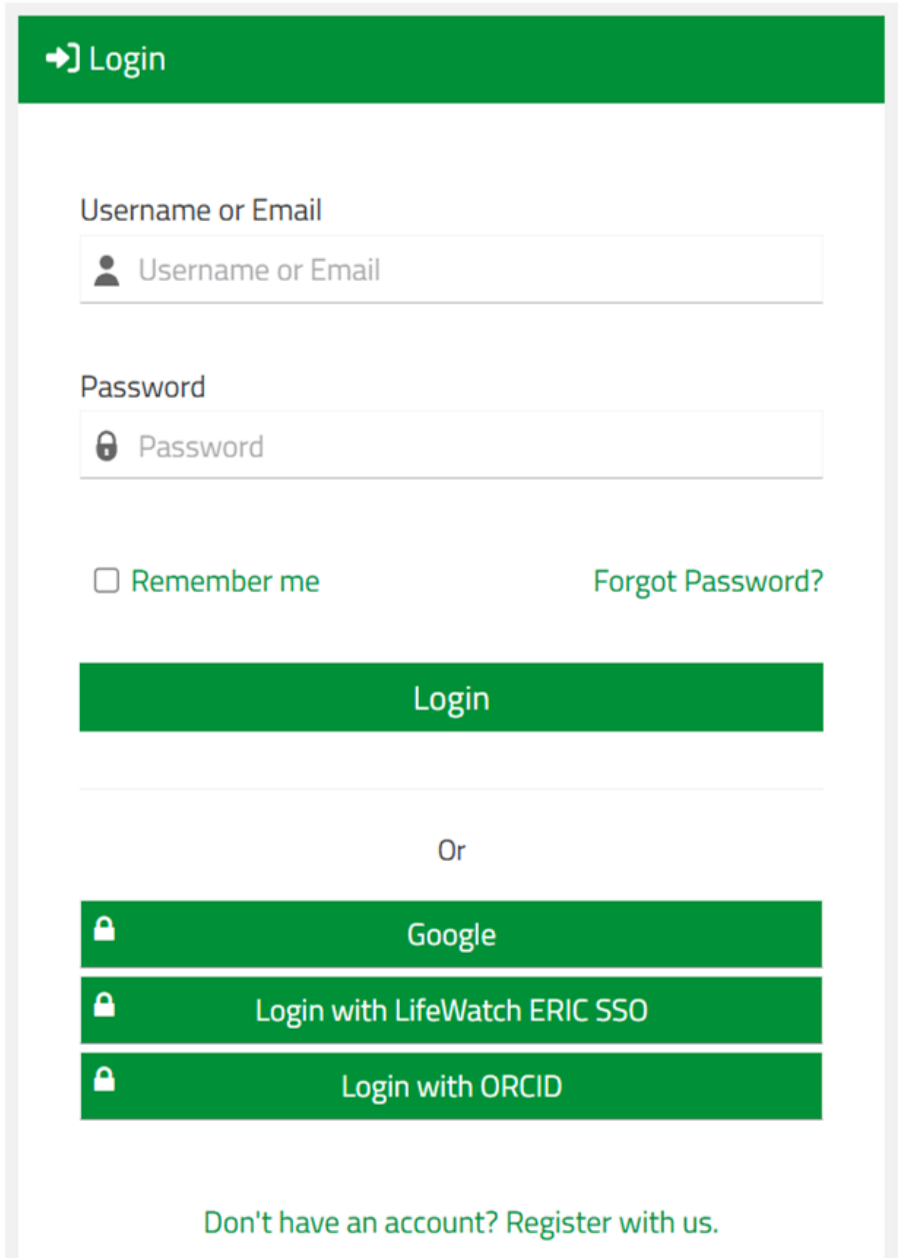
- Integration with existing LifeWatch components, such as Data Portal and Metadata Catalogue, allowing you to publish your research products (e.g., scripts, datasets, etc.) and ensuring consistency and interoperability.

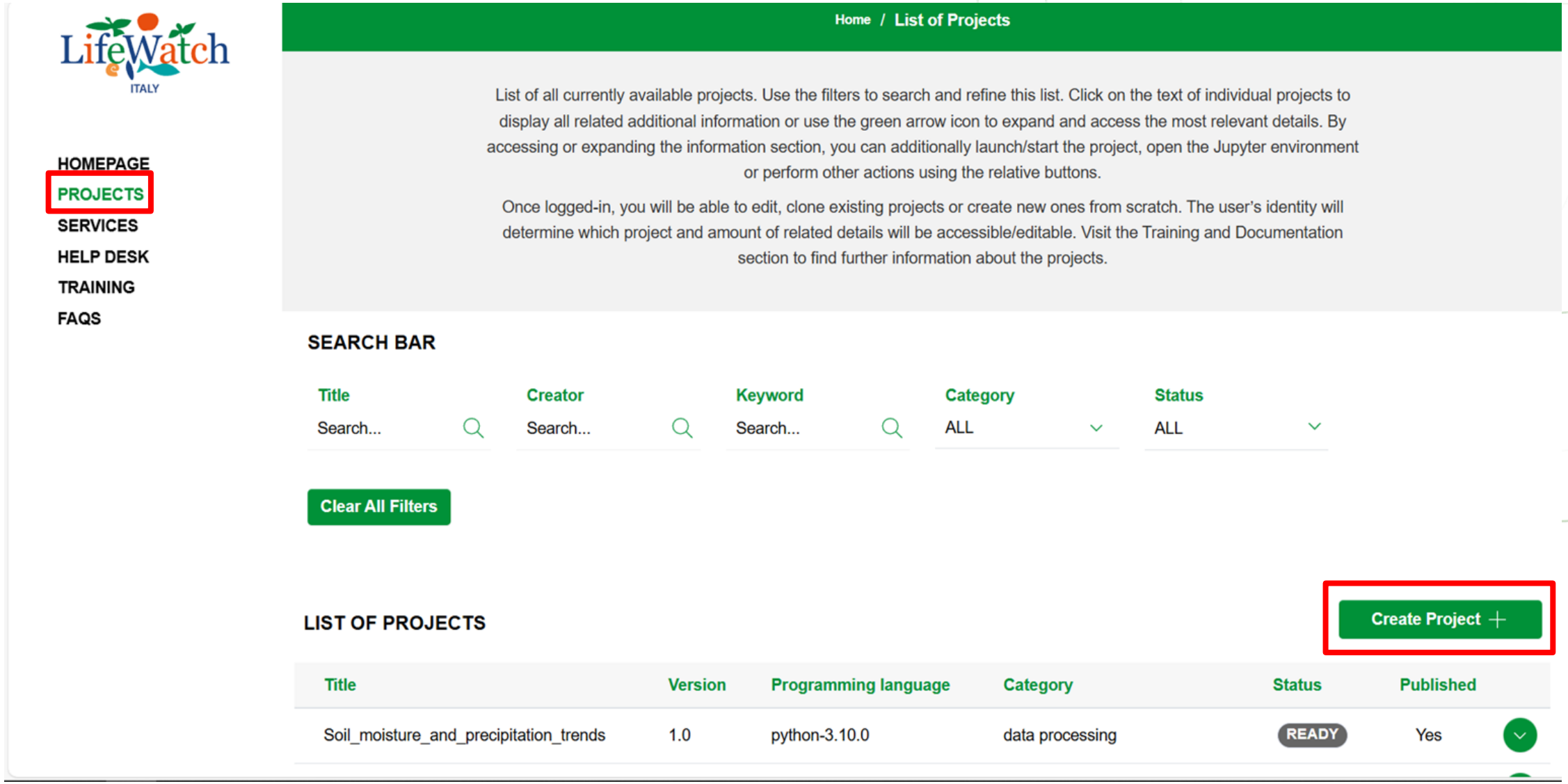

DASHBOARD
jupyterhub
binder
USER
USER AUTHENTICATION
WEB INTERFACES
SCRIPTS CATALOGUE WEB INTERFACE
CONTACT ADMIN
Zammad
ADMIN
REQUEST MANAGEMENT
LIFEWATCH SSO
TRAINING PLATFORM
LIFEWATCH HELP DESK
CORE
jupyterlab
MATLAB
RUN TIME ENVIRONMENT
COLLABORATIVE EDITOR WEB INTERFACE
CHAT VOCALE
jitsi
SCRIPTS CATALOGUE
SOURCE CODE REPOSITORY
binder
SCRIPTS CATALOGUE REST API
PUBLICATION ENGINE
WEB SERVICES
EcoPortal
BE-LINDA
DataCite
BE-LISA
User
BE-HUB

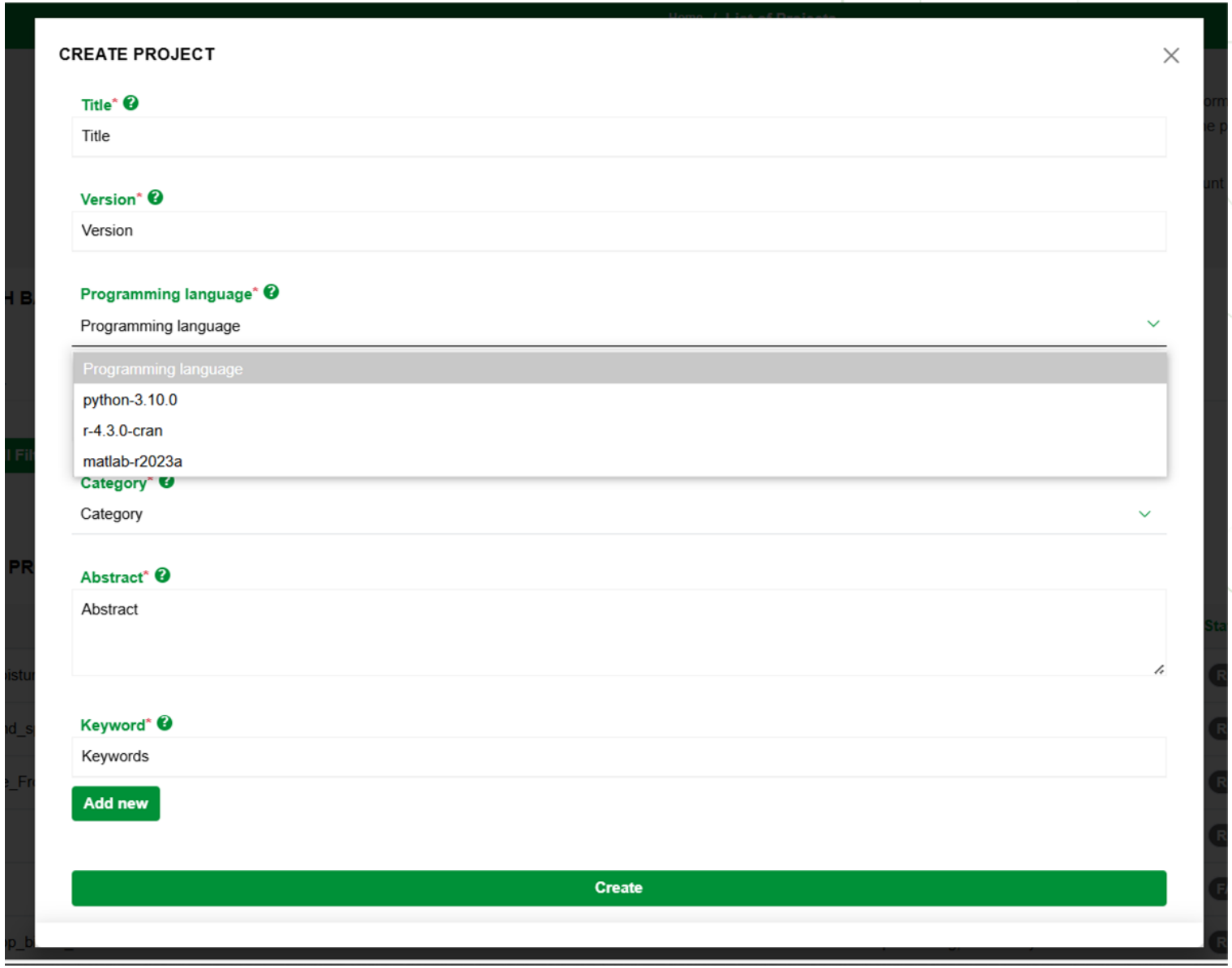
Join DataLabs

https://datalabs.lifewatchitaly.eu/


LifeWatch
ITALY
Login
HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS
DataLabs: LifeWatch's Collaborative Coding Platform for Biodiversity and Ecosystem Research
Unleash the power of collaboration and coding to advance your research today!
Coding in R, Python and MATLAB with your team has never been easier.
This platform is integrated with LifeWatch, a leading portal for data-driven research. Publish your scripts and projects, and deploy web services all in one place, with the added benefit of structured web interfaces for your services thanks to the DataLabs feature. Not only does this platform promote collaboration in coding and data science, but it also aligns with FAIR (Findable, Accessible, Interoperable, Reusable) data principles and Open Science practices. Be part of a community that advances the field through sharing codes and results.
Visit the Training and Documentation sections to access detailed information about DataLabs.
Our Platform in Numbers
124 Projects
8 Published Projects
0 Users Logged In
5 Services


Login
Username or Email
Username or Email
Password
Password
Remember me
Forgot Password?
Login
Or
Google
Login with LifeWatch ERIC SSO
Login with ORCID
Don't have an account? Register with us.


LifeWatch
ITALY
HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS
Home / List of Projects
List of all currently available projects. Use the filters to search and refine this list. Click on the text of individual projects to display all related additional information or use the green arrow icon to expand and access the most relevant details. By accessing or expanding the information section, you can additionally launch/start the project, open the Jupyter environment or perform other actions using the relative buttons.
Once logged-in, you will be able to edit, clone existing projects or create new ones from scratch. The user's identity will determine which project and amount of related details will be accessible/editable. Visit the Training and Documentation section to find further information about the projects.
SEARCH BAR
Title
Search...
Creator
Search...
Keyword
Search...
Category
ALL
Status
ALL
Clear All Filters
LIST OF PROJECTS
Create Project
Title
Version
Programming language
Category
Status
Published
Soil_moisture_and_precipitation_trends
1.0
python-3.10.0
data processing
READY
Yes


CREATE PROJECT
Title*
Title
Version*
Version
Programming language*
Programming language
Programming language
python-3.10.0
r-4.3.0-cran
matlab-r2023a
Category*
Category
Abstract*
Abstract
Keyword*
Keywords
Add new
Create

LifeWatch ITALY

ITINERIS

HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS

SEARCH BAR

Title	Creator	Keyword	Category	Status
Search...	Search...	Search...	ALL	ALL

Clear All Filters

LIST OF PROJECTS

Create Project +

Title	Version	Programming language	Category	Status	Published
Soil_moisture_and_precipitation_trends	1.0	python-3.10.0	data processing	READY	Yes
LCC_and_species_distribution_analysis		python-3.10.0	data processing	READY	Yes
distance_From_river	v1.1	r-4.3.0-cran	data processing	READY	Yes
Pr2_2	1.1	python-3.10.0	data processing	READY	No
Pr2	1.0	python-3.10.0	data processing	FAILED	No
gam_npp_biodiv_2	v1.2	r-4.3.0-cran	data processing, data analy…	READY	Yes

LifeWatch ITALY

HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS

Creator: Mariantonietta La Marra (mariantonietta.lamarra@unisalento.it)

Input

Name	Source	Size	Date	Author	Actions
CCS_2024-10-09061335am_2021.nc	LOCAL	52.78 MB	13/02/2025 13:09	Mariantonietta La Marra (mariantonietta.lamarra@unisalento.it)	
sm_combined_2021.zip	LOCAL	672.26 MB	13/02/2025 13:10	Mariantonietta La Marra (mariantonietta.lamarra@unisalento.it)	

Attach new file input

Select your file from your device or Data Portal

Local | Data Portal

Output

Warnings and Alerts

ITINERIS

LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

Local | Data Portal

Output

Warnings and Alerts

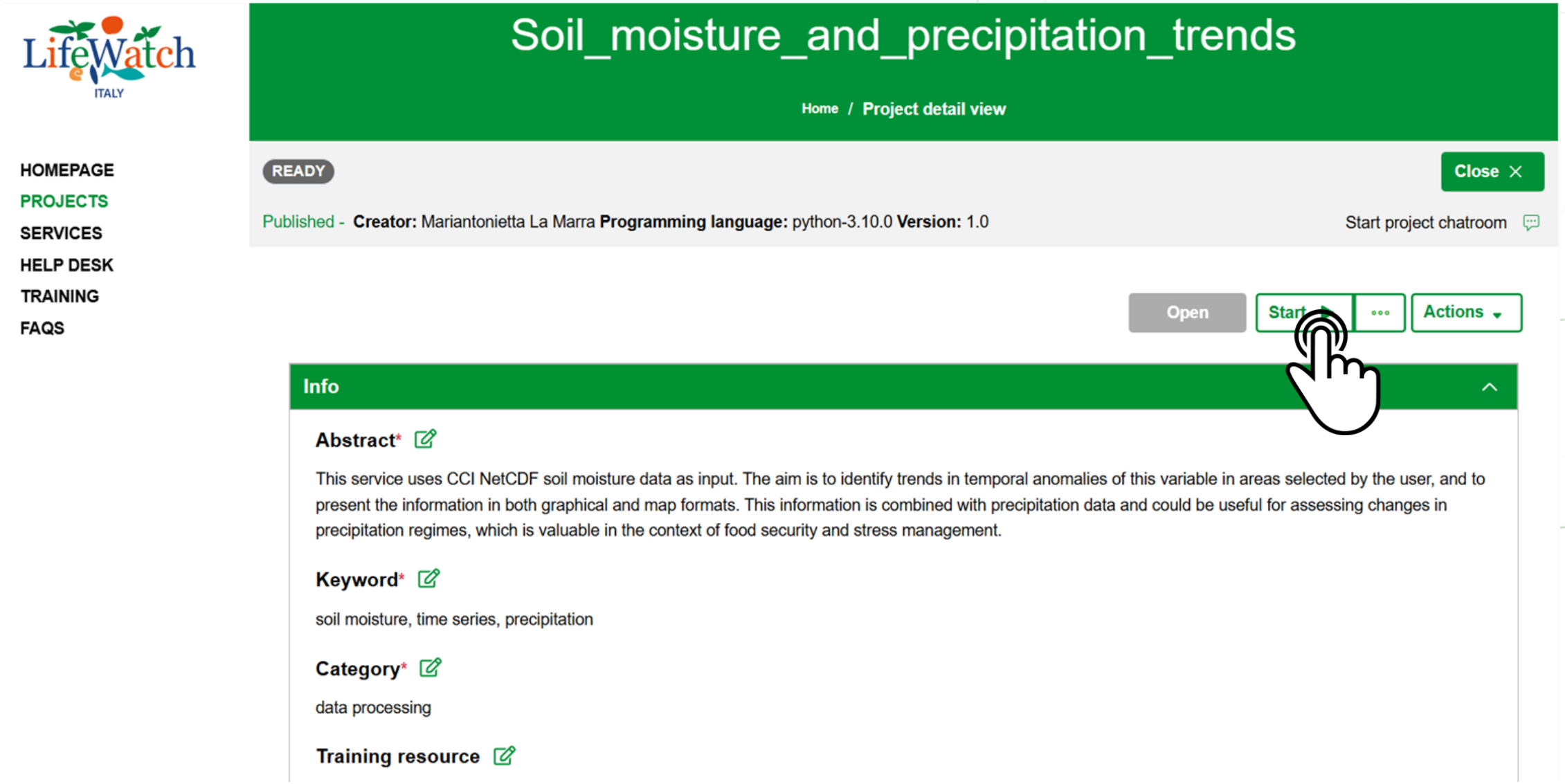
Team

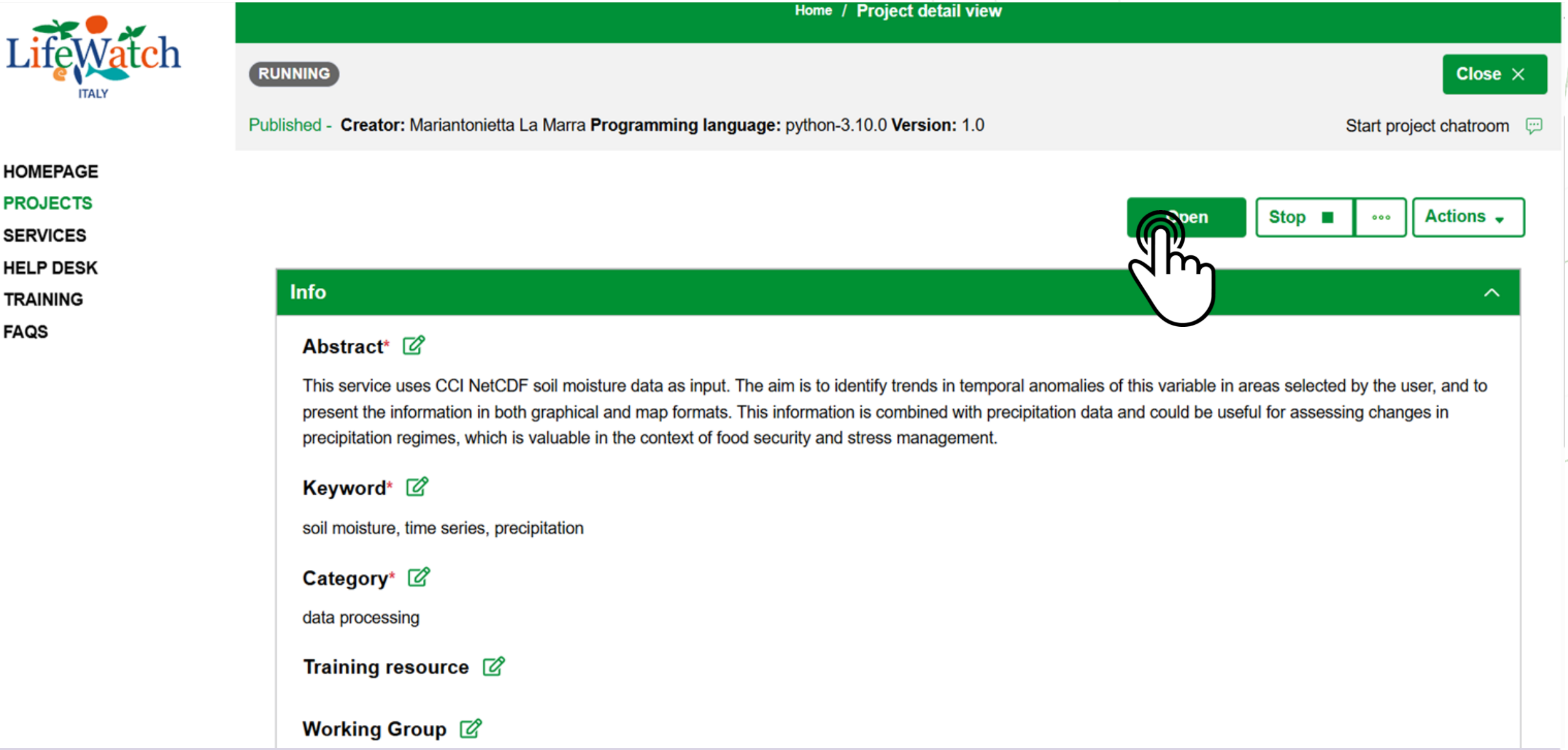
Creator: Mariantonietta La Marra (mariantonietta.lamarra@unisalento.it)

Name	Email	Date	Actions
Francesco De Leo	francesco.deleo@cnr.it	14/02/2025 13:31	

Statistics

Provenance


LifeWatch ITALY
HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS
Soil_moisture_and_precipitation_trends
Home / Project detail view
READY
Close
Published - Creator: Mariantonietta La Marra Programming language: python-3.10.0 Version: 1.0
Start project chatroom
Open
Start
Actions
Info
Abstract*
This service uses CCI NetCDF soil moisture data as input. The aim is to identify trends in temporal anomalies of this variable in areas selected by the user, and to present the information in both graphical and map formats. This information is combined with precipitation data and could be useful for assessing changes in precipitation regimes, which is valuable in the context of food security and stress management.
Keyword*
soil moisture, time series, precipitation
Category*
data processing
Training resource


Home / Project detail view
RUNNING
Close
Published - Creator: Mariantonietta La Marra Programming language: python-3.10.0 Version: 1.0
Start project chatroom
HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS
Stop
Actions
Info
Abstract*
This service uses CCI NetCDF soil moisture data as input. The aim is to identify trends in temporal anomalies of this variable in areas selected by the user, and to present the information in both graphical and map formats. This information is combined with precipitation data and could be useful for assessing changes in precipitation regimes, which is valuable in the context of food security and stress management.
Keyword*
soil moisture, time series, precipitation
Category*
data processing
Training resource
Working Group

LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

Stop ■

File Edit View Run Kernel Git Tabs Settings Help CPU: 0% Mem: 164 / 5859 MB

index.ipynb | install.conda | install.apt

Code | git | Python3.10.0

Filter files by name

/

Name	Last Modified
buildlog	4 months ago
input	4 months ago
output	4 months ago
tmp	a year ago
example.ip...	4 months ago
index.ipynb	4 months ago
install.apt	4 months ago
install.conda	4 months ago
README.md	4 months ago
userinput.y...	4 months ago

[9]:
```
# open multiple files
ds = xr.open_mfdataset('output/sm_combined_2021/sm_combined_2021/*.nc', combine='nested', concat_dim=
```
Last executed at 2024-11-14 12:39:39 in 16.43s

[10]:
```
# open the yml file containing the user inputs
with open('userinput.yml', 'r') as f:
    coord = yaml.load(f, Loader=yaml.SafeLoader)
```

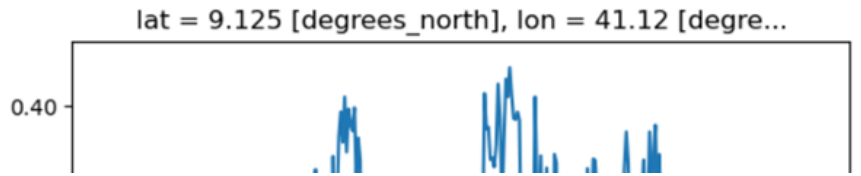
[ ]:
```
# contents of the yml file
coord
```

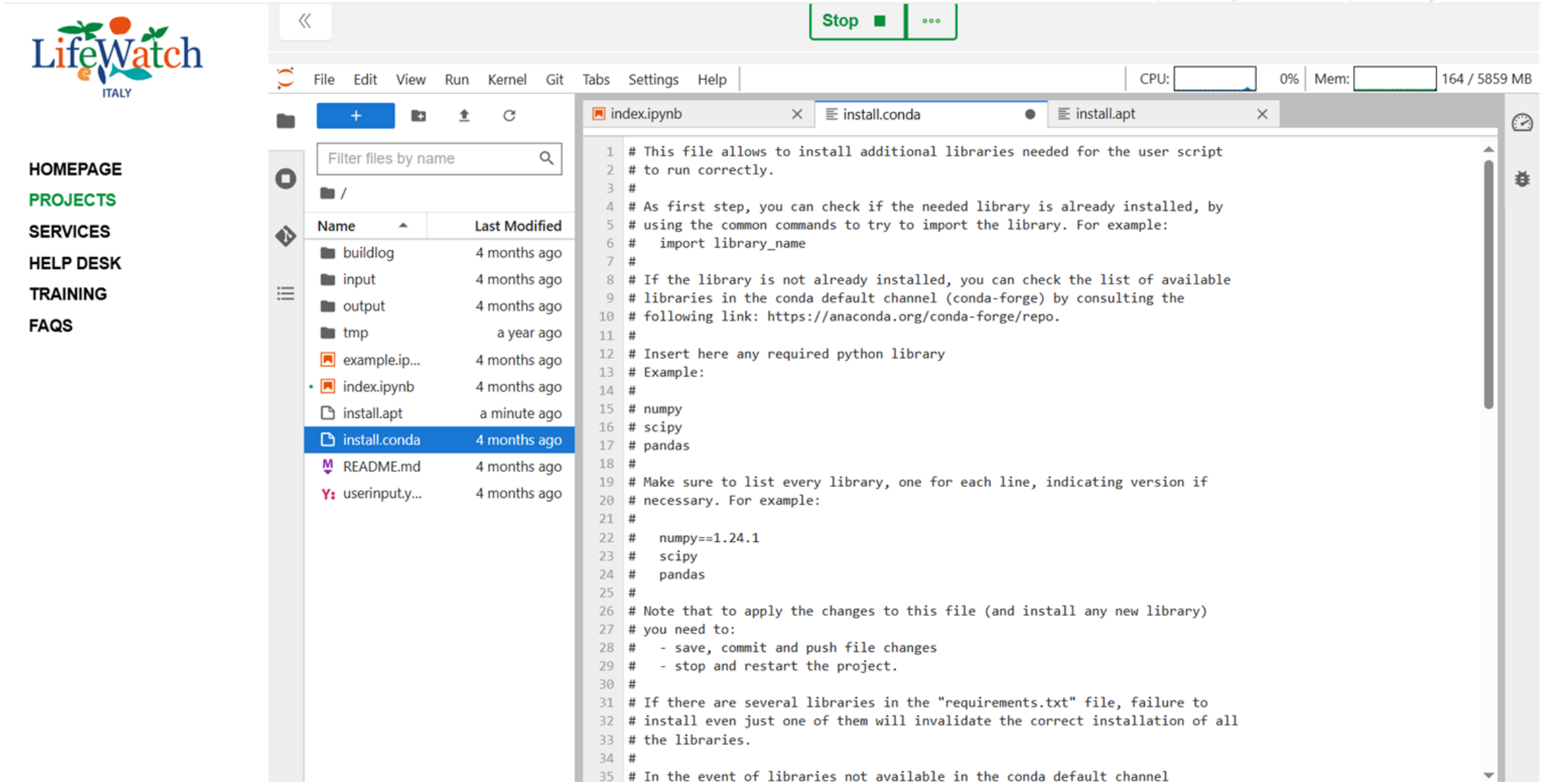
[11]:
```
{'coordinates': [9.0, 41.0], 'boundingbox': [26.0, 3.0, 43.0, 16.0]}
```

[12]:
```
# extract coordinates chosen by the user
latp = coord['coordinates'][0]
lonp = coord['coordinates'][1]
```

[13]:
```
# extract values from the .nc file for the chosen coordinates
dsloc = ds.sel(lon=lonp, lat=latp, method='nearest')
```

[ ]:
```
# plot the soil moisture values for the chosen coordinates
dsloc['sm'].plot();
```




LifeWatch ITALY
HOMEPAGE
PROJECTS
SERVICES
HELP DESK
TRAINING
FAQS
Stop
File Edit View Run Kernel Git Tabs Settings Help
CPU: 0% Mem: 164 / 5859 MB
Filter files by name
/
Name Last Modified
buildlog 4 months ago
input 4 months ago
output 4 months ago
tmp a year ago
example.ip... 4 months ago
index.ipynb 4 months ago
install.apt a minute ago
install.conda 4 months ago
README.md 4 months ago
userinput.y... 4 months ago
index.ipynb
install.conda
install.apt
This file allows to install additional libraries needed for the user script
to run correctly.
#
As first step, you can check if the needed library is already installed, by
using the common commands to try to import the library. For example:
import library_name
#
If the library is not already installed, you can check the list of available
libraries in the conda default channel (conda-forge) by consulting the
following link: https://anaconda.org/conda-forge/repo.
#
Insert here any required python library
Example:
#
numpy
scipy
pandas
#
Make sure to list every library, one for each line, indicating version if
necessary. For example:
#
numpy==1.24.1
scipy
pandas
#
Note that to apply the changes to this file (and install any new library)
you need to:
- save, commit and push file changes
- stop and restart the project.
#
If there are several libraries in the "requirements.txt" file, failure to
install even just one of them will invalidate the correct installation of all
the libraries.
#
In the event of libraries not available in the conda default channel

LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

Stop ■

File Edit View Run Kernel Git Tabs Settings Help

CPU: 0% Mem: 169 / 5859 MB

Changes History

Staged (0)

Changed (1)

index.ipynb M

Untracked (0)

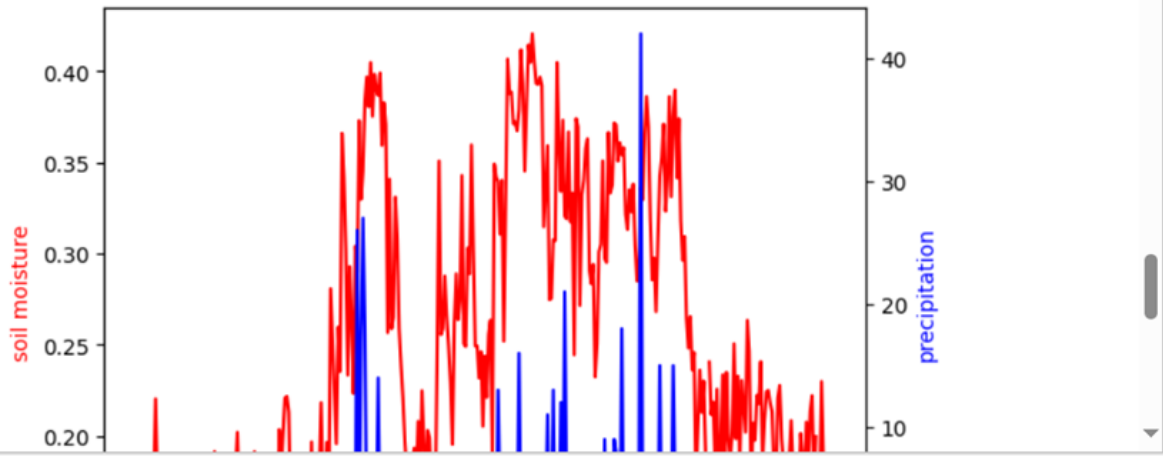
Summary (Ctrl+Enter to commit)

COMMIT

index.ipynb | install.conda | install.apt

Code git Python3.10.0

```
[ ]: # plot soil moisture and precipitation data together

fig, ax1 = plt.subplots()

ax1.set_xlabel('time')
ax1.set_ylabel('soil moisture',color='red')
ax1.plot(dsloc['time'],dsloc['sm'],color='red')

ax2 = ax1.twinx()  # instantiate a second Axes that shares the same x-axis


ax2.set_ylabel('precipitation',color='blue')  # we already handled the x-label with ax1
ax2.plot(preciploc['datetime'],preciploc['precip'],color='blue')

plt.show()
```



LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

Stop ■

File Edit View Run Kernel Git Tabs Settings Help CPU: 0% Mem: 170 / 5859 MB

Changes History

Staged (1)

index.ipynb M

Changed (0)

Untracked (0)

c1

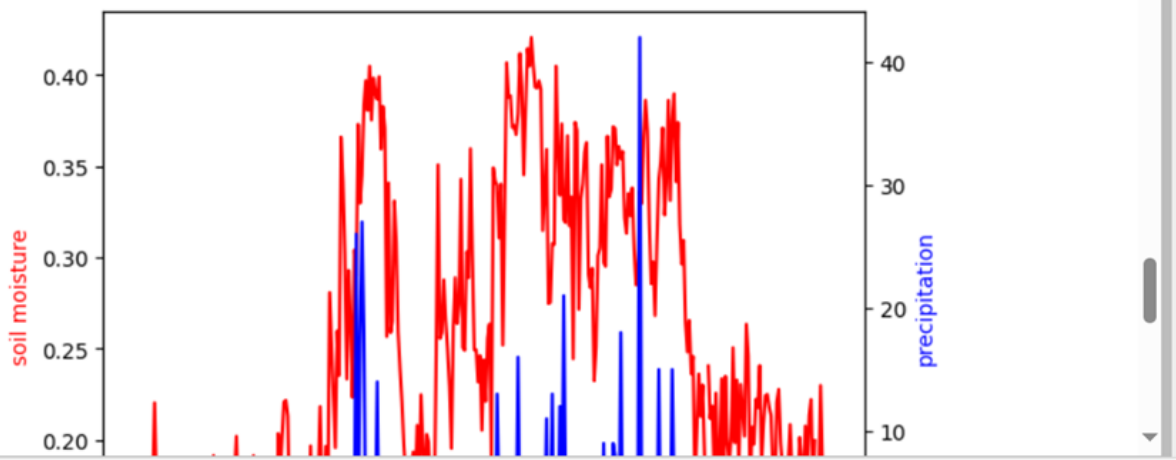
Description (optional)

COMMIT

index.ipynb install.conda install.apt

Code git Python3.10.0

```
[ ]: # plot soil moisture and precipitation data together

fig, ax1 = plt.subplots()

ax1.set_xlabel('time')
ax1.set_ylabel('soil moisture',color='red')
ax1.plot(dsloc['time'],dsloc['sm'],color='red')

ax2 = ax1.twinx()  # instantiate a second Axes that shares the same x-axis


ax2.set_ylabel('precipitation',color='blue')  # we already handled the x-label with ax1
ax2.plot(preciploc['datetime'],preciploc['precip'],color='blue')

plt.show()
```



LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

Stop ■

File Edit View Run Kernel Git Tabs Settings Help

CPU: 0% Mem: 170 / 5859 MB

Changes | History

- Staged (0)
- Changed (0)
- Untracked (0)

Summary (Ctrl+Enter to commit)

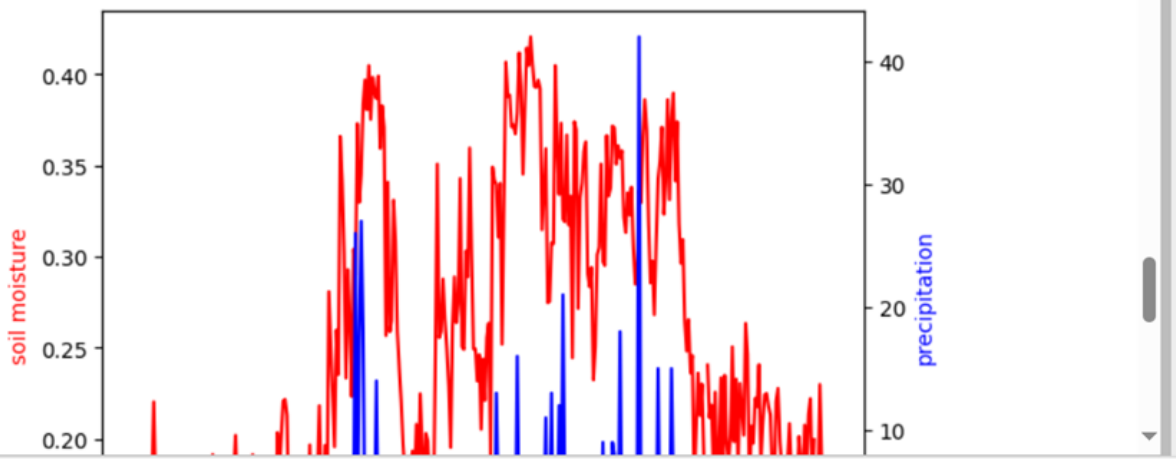
Description (optional)

COMMIT

index.ipynb | install.conda | install.apt

Code | git | Python3.10.0

```
[ ]: # plot soil moisture and precipitation data together

fig, ax1 = plt.subplots()

ax1.set_xlabel('time')
ax1.set_ylabel('soil moisture',color='red')
ax1.plot(dsloc['time'],dsloc['sm'],color='red')

ax2 = ax1.twinx()  # instantiate a second Axes that shares the same x-axis


ax2.set_ylabel('precipitation',color='blue')  # we already handled the x-label with ax1
ax2.plot(preciploc['datetime'],preciploc['precip'],color='blue')

plt.show()
```



ITINERIS

LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

This service uses CCI NetCDF soil moisture data as input. The aim is to identify trends in temporal anomalies of this variable in areas selected by the user, and to present the information in both graphical and map formats. This information is combined with precipitation data and could be useful for assessing changes in precipitation regimes, which is valuable in the context of food security and stress management.

Creator: Mariantonietta La Marra (mariantonietta.lamarra@unisalento.it)

Open | Start ▶ | ••• | Actions

LCC_and_species_distribution_analysis	1.0	python-3.10.0	data processing	READY	Yes

This service uses satellite data on land use provided by ESA's CCI programme, in particular a new high resolution (30m) land cover change (LCC) product available for sub-Saharan Africa. This product is combined with the spatial distribution of IUCN Red List species. The purpose of this service is to allow users to select a bounding box of interest and retrieve information on land use change, the density of Red List species in the area, their conservation status, and a "risk" map highlighting areas particularly vulnerable to change and the distribution of threatened species.

Creator: Mariantonietta La Marra (mariantonietta.lamarra@unisalento.it)

Open | Start ▶ | ••• | Actions

- Clone the Project
- Create a Service
- Publish a Dataset
- Publish the Script
- Download .zip
- Archive the Project
- Delete the Project

distance_From_river	v1.1	r-4.3.0-cran	data processing	READ

calculate distance from river mouth

Creator: Francesco de leo (francesco.deleo@cnr.it)

Pr2_2	1.1	python-3.10.0	data processing	READ

temporal trends in soil moisture and precipitation data

Group activity

How do extreme weather events (e.g. heat waves or late frosts) affect the distribution of pollinating insects?

Group activity

How do local variations in temperature and humidity affect interactions between host plants and phytophagous insects?

THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani