Exploring meaning in data: a hands-on course in semantics and analysis for FAIR Research

Module 4

- 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

Agenda

09:00 – 11:00 Challenges in data analysis and introduction to DataLabs

11:00 – 11:15 Coffee break

11:15 – 13:30 Creating a project in DataLabs

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.

https://doi.org/10.1371/journal.pone.0300333

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.

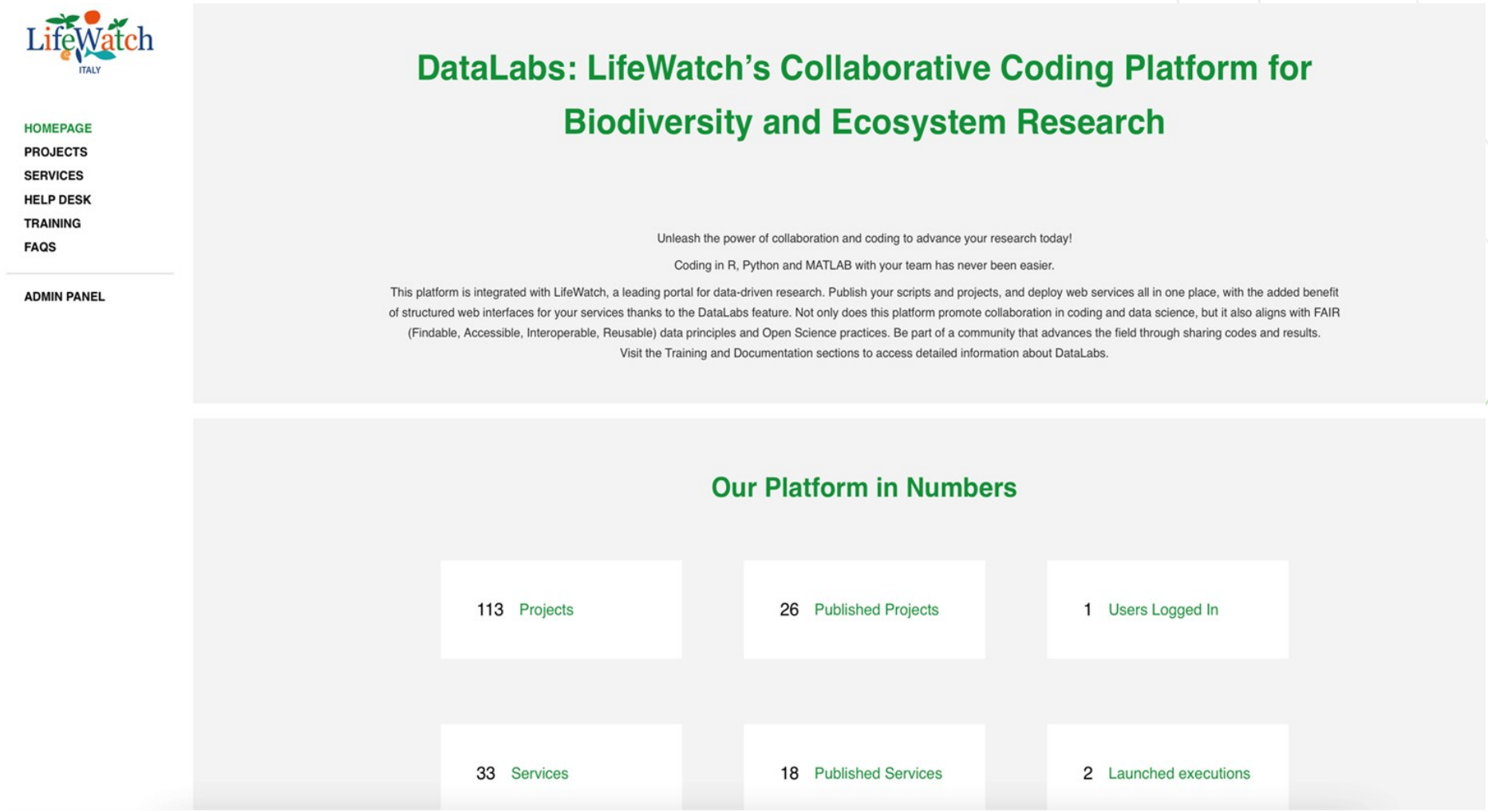
- global and seamless access to the data and services of interest, with no temporal or spatial constraints

- 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

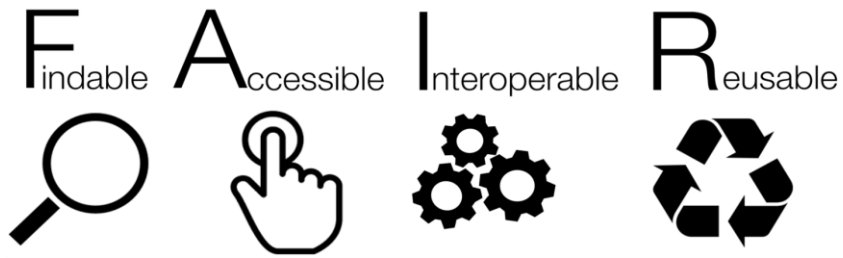
- promote fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance and attribution

- 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.



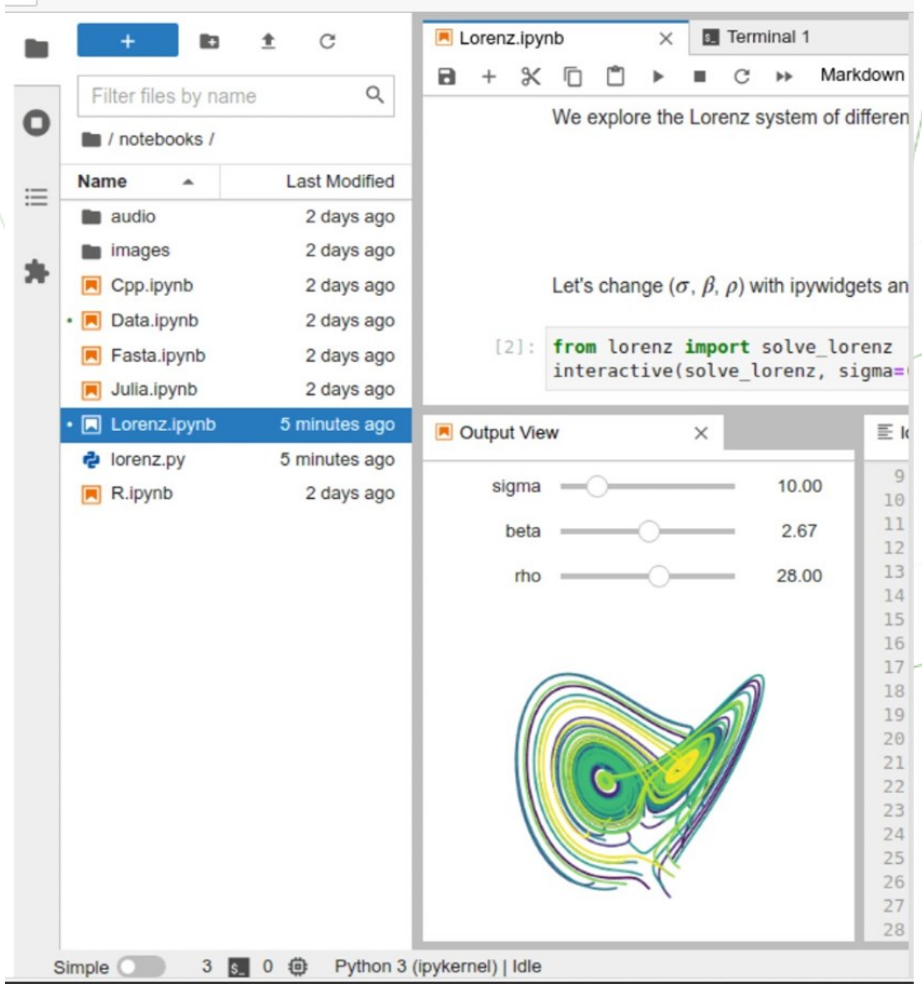
https://doi.org/10.1038/s41597-022-01710-x
https://doi.org/10.1038/s41597-025-04451-9

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

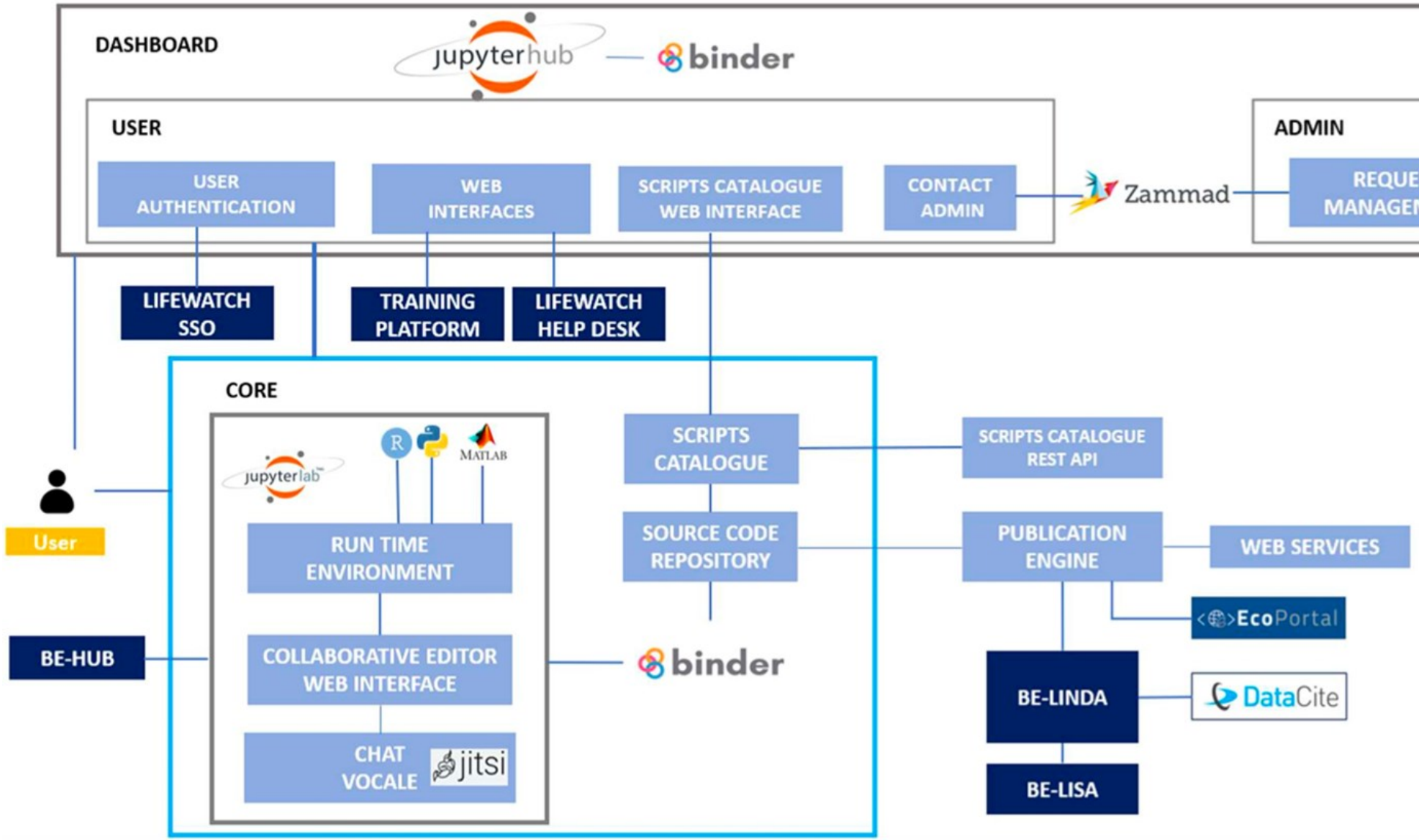
Jupyter is an open-source project, born out of the original IPython Project in 2014, that provides a web interface for interactive data science and scientific computing across multiple programming languages.

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

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