Addressing emerging training needs in Earth Observation with Jupyter notebooks

Julia Wagemann[1], Simone Mantovani [2], Federico Fierli [3],
[1] Julia Wagemann Consulting, [2] MEEO s.r.l., [3] EUMETSAT

Living Planet Symposium | 26 May 2022

Wagemann, J., Fierli, F., Mantovani, S., Siemen, S., Seeger, B. and J. Bendix (2022): Using Jupyter Notebooks for Earth Observation data education. *Remote Sensing. (under review)*

Earth Observation – Emerging training needs

CHALLENGES

A growing gap between the amount of EO data produced every day and the ability of users to find, access and process the data

Growing data volumes, data discovery and limited processing capacity are major challenges users face, which hinder data uptake and use[1]

TRAINING NEEDS

To teach EO data users about existing and upcoming products and to teach them how to access and work with these data products

EO TRAINING PROVIDER

EO data providers vs. Higher education system

[1] Wagemann et al. (2021): Users of open Big Earth data – An analysis of the current state. Computers and Geosciences.

Earth Observation training @EUMETSAT

Aim to increase use and overall uptake of EUMETSAT and Copernicus data

Target group and audience

Highly diverse group of EO practitioners

- Can be 'expert' users who would like to learn more about upcoming missions and data
- Also users who just started using EO data products for a specific application
- Coming from different geographical regions
- Access to a widely differing range of resources e.g. own computers, high performance clusters, cloud.

Training types and needs

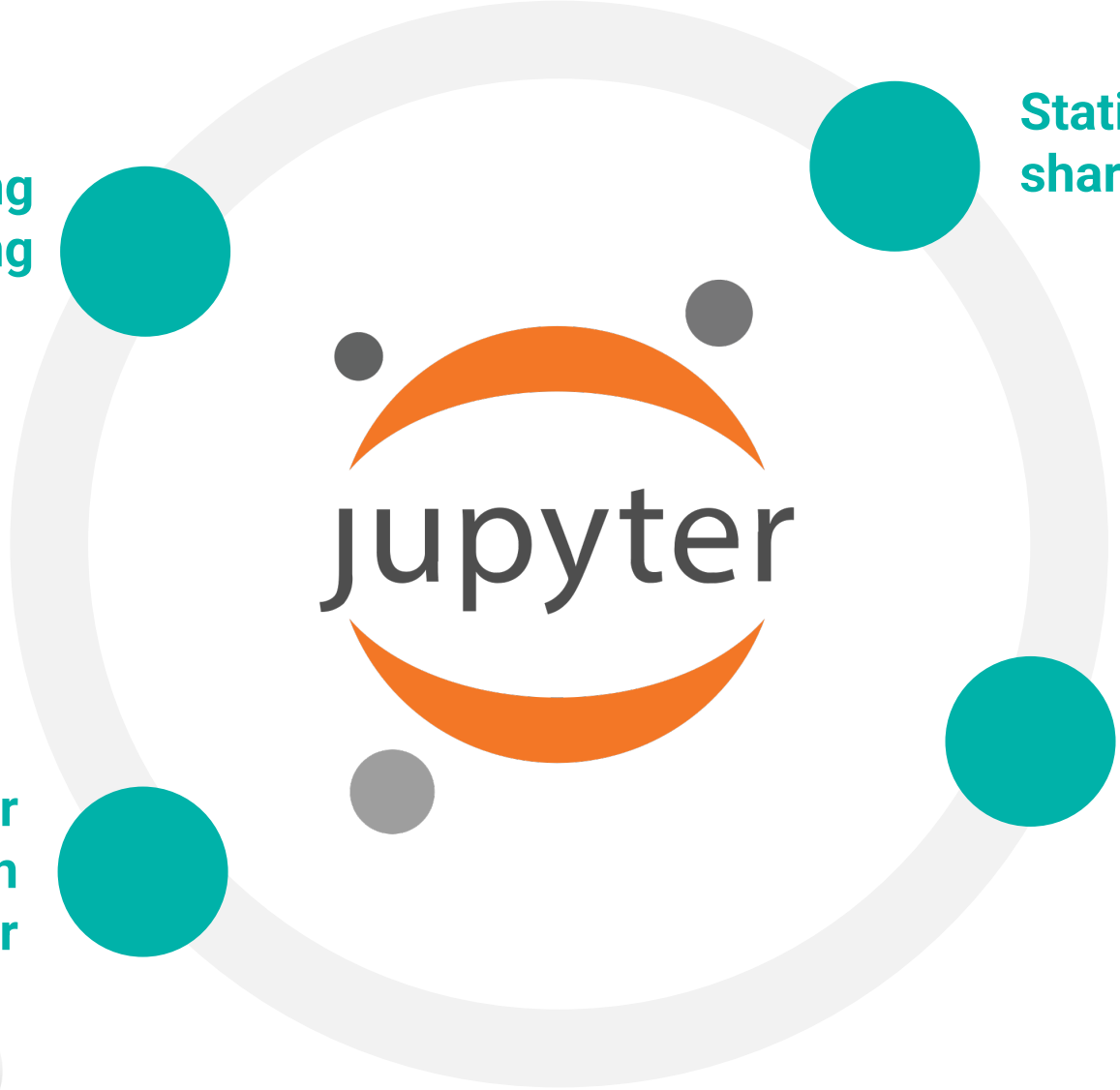
Different training types offered

- Short courses: light-weight introduction to a specific topic, usually 1 to 1.5 hours long
- Thematic expert workshops / training schools: multiple days, aimed at experts and graduate students

Training has to

- cater for different levels of data, thematic and programming literacy
- Offer a high degree of flexibility in terms of how the courses and modules can be taught

Beyond notebooks – the Project Jupyter ecosystem



Version control, hosting and sharing

GitHub GitLab

Static rendering and sharing

jupyter nbviewer

Group and multi-user access to notebooks on a remote server

jupyterhub lab

No-setup, customizable Jupyter notebook environment

colab binder kaggle

On which side are you? Like or dislike Jupyter notebooks?

Perkel (2018, 2021):

Within 10 years, became the de-facto standard for data exploration, analysis and training.

In 2021, there are more than 10 million Jupyter notebooks available on Github

Interactive web application that combines code, computational output, explanatory text and multimedia resources

Grus, J. (2018):

Cells can be executed out of order and fosters poor coding practices

Pimentel et al. (2019):
1 out of 4 notebooks could be re-executed and only 4% produced the same results

[1] Perkel (2018): Why Jupyter is data scientists' computation notebook of choice. Nature.

[2] Perkel (2021): Ten computer codes that transformed science. Nature.

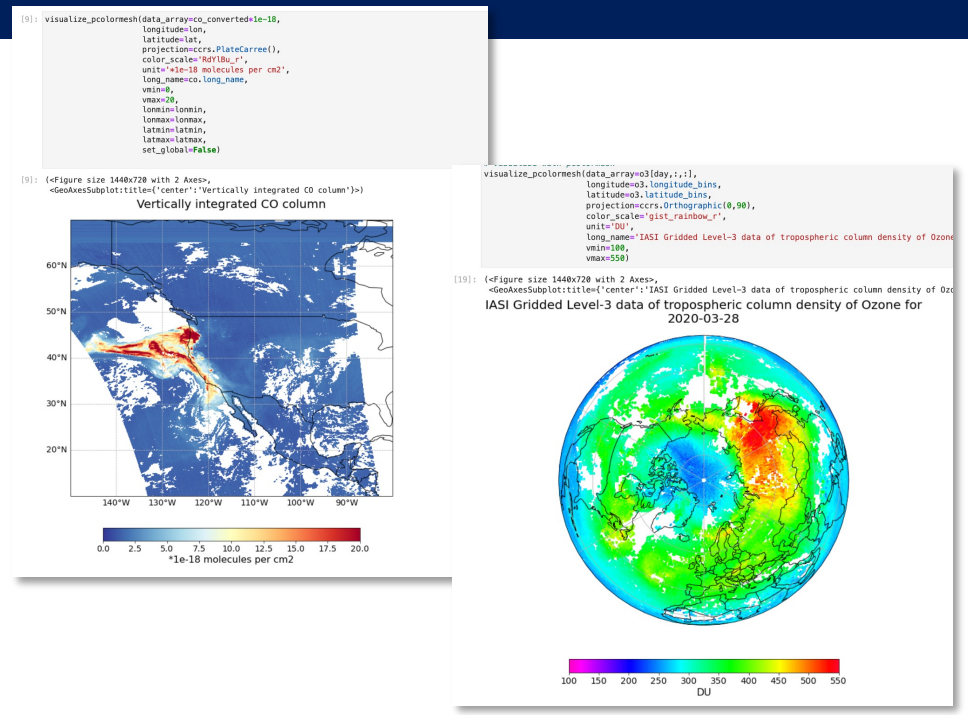
[3] Pimentel et al. (2019): A Large-Scale Study About Quality and Reproducibility of Jupyter Notebooks. IEEE.

[4] Grus, J. (2018): I don't like Notebooks. JupyterCon 2018

Example: Learning Tool for Python on Atmospheric Composition (LTPy)

Jupyter-based training course on open
satellite-, model- and ground-based data on atmospheric composition

- Provide a general overview of different satellite-, model- and ground-based data on atmospheric composition
- Provide code examples and step-by-step guides on how to load, process and visualize these data
- Provide examples of thematic application areas, e.g. fire monitoring, air quality, dust, or stratospheric ozone



Main course

Thematic module on dust

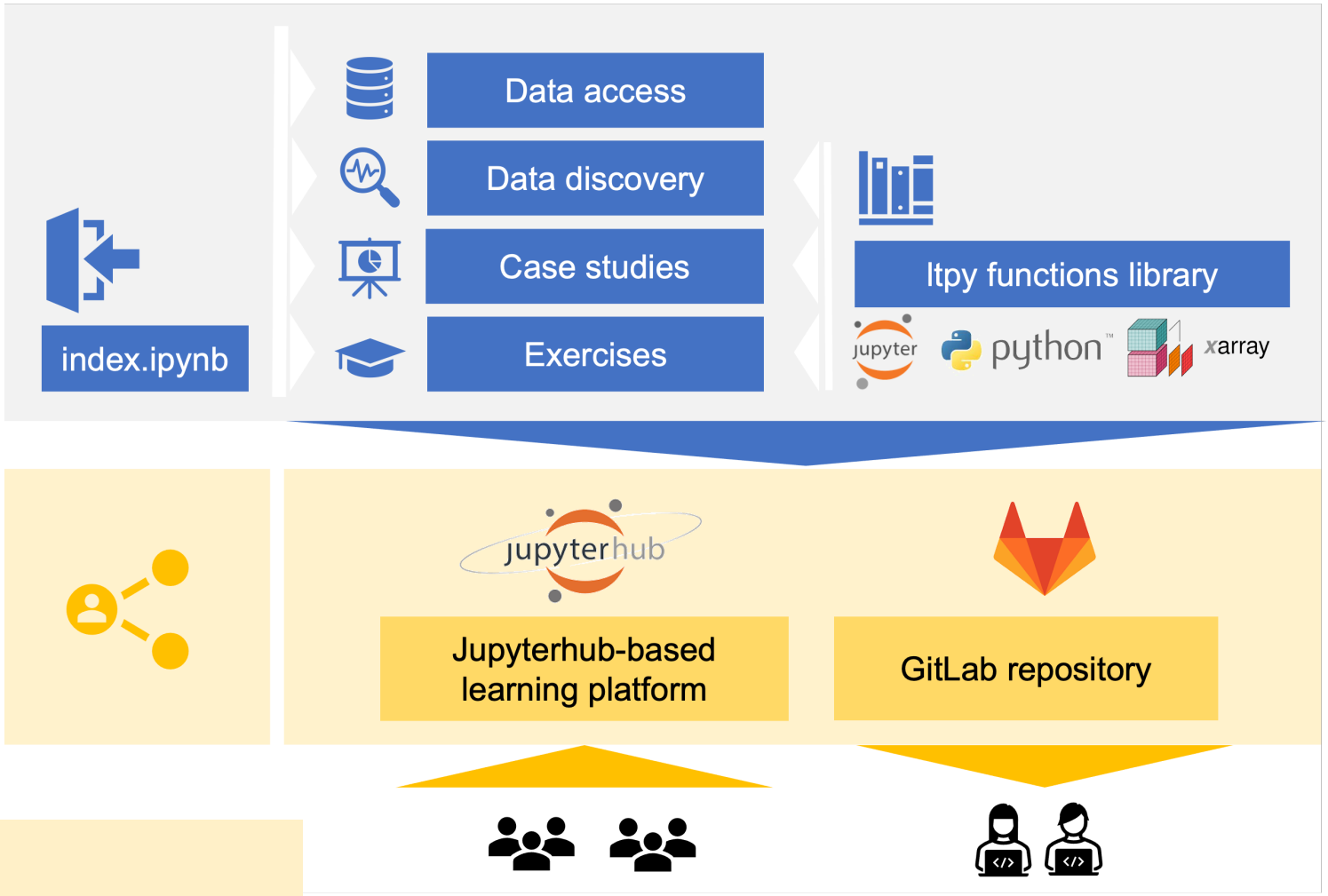
Example: LTPy main course

LTPy main course is aligned with a typical data analysis workflow

Data from 6 different satellites and 5 different model-based products

Over 50 notebooks related to (i) data access, (ii) data discovery, (iii) case studies and (iv) exercises

A collection of 14 reusable functions for effective visualization and data handling



https://ltpy.adamplatform.eu

https://gitlab.eumetsat.int/eumetlab/atmosphere/atmosphere

Example: LTPy thematic module on dust monitoring and forecasting

Self-contained collection of notebooks related to dust monitoring and forecasting

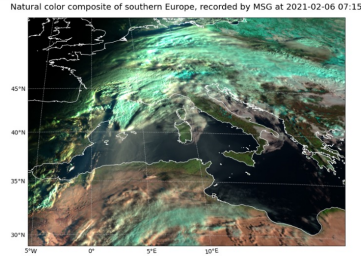
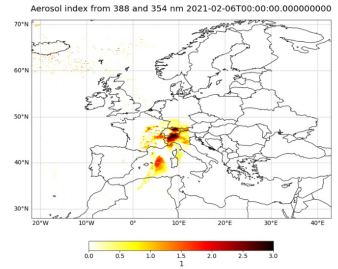
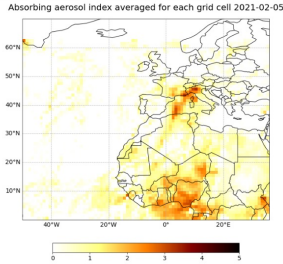
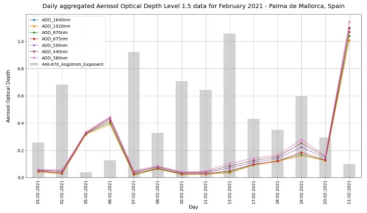
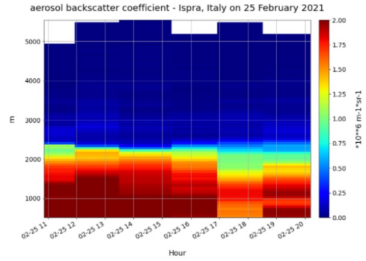
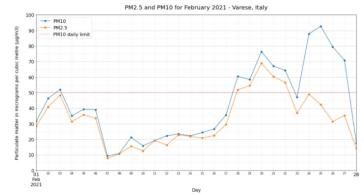
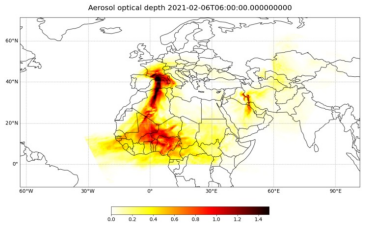
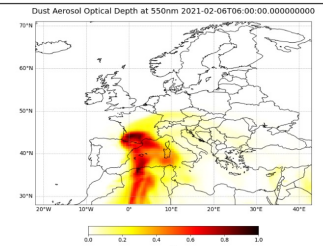
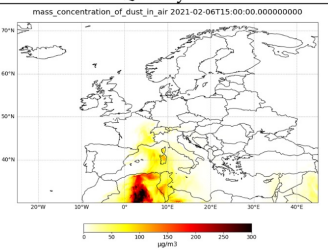
Data from 6 satellites, 3 model-based products and 3 ground-based observations

18 notebooks related divided in two sections: (i) data discovery and (ii) practical assignments

Accessible in form of a Jupyterbook and a dedicated Jupyterhub instance

https://dust.trainhub.eumetsat.int

https://dust.ltpy.adamplatform.eu

SATELLITE-BASED DATA (selection)	 Meteosat Second Generation Level 1B RGB composite	 Sentinel-5P TROPOMI Aerosol Index Level 2 product	 Metop-ABC GOME-2 Absorbing Aerosol Index Level 2 product
GROUND-BASED OBSERVATIONS	 Aerosol RObotic NETwork (AERONET)	 EARLINET Lidar backscatter profiles	 European Environment Agency (EEA) Air Quality Data
MODEL-BASED DATA	 NMMB/MONARCH regional model – Aerosol Optical Depth	 Copernicus Atmosphere Monitoring Service (CAMS) global forecast – Dust Aerosol Optical Depth	 Copernicus Atmosphere Monitoring Service (CAMS) European air quality forecasts – Dust concentration

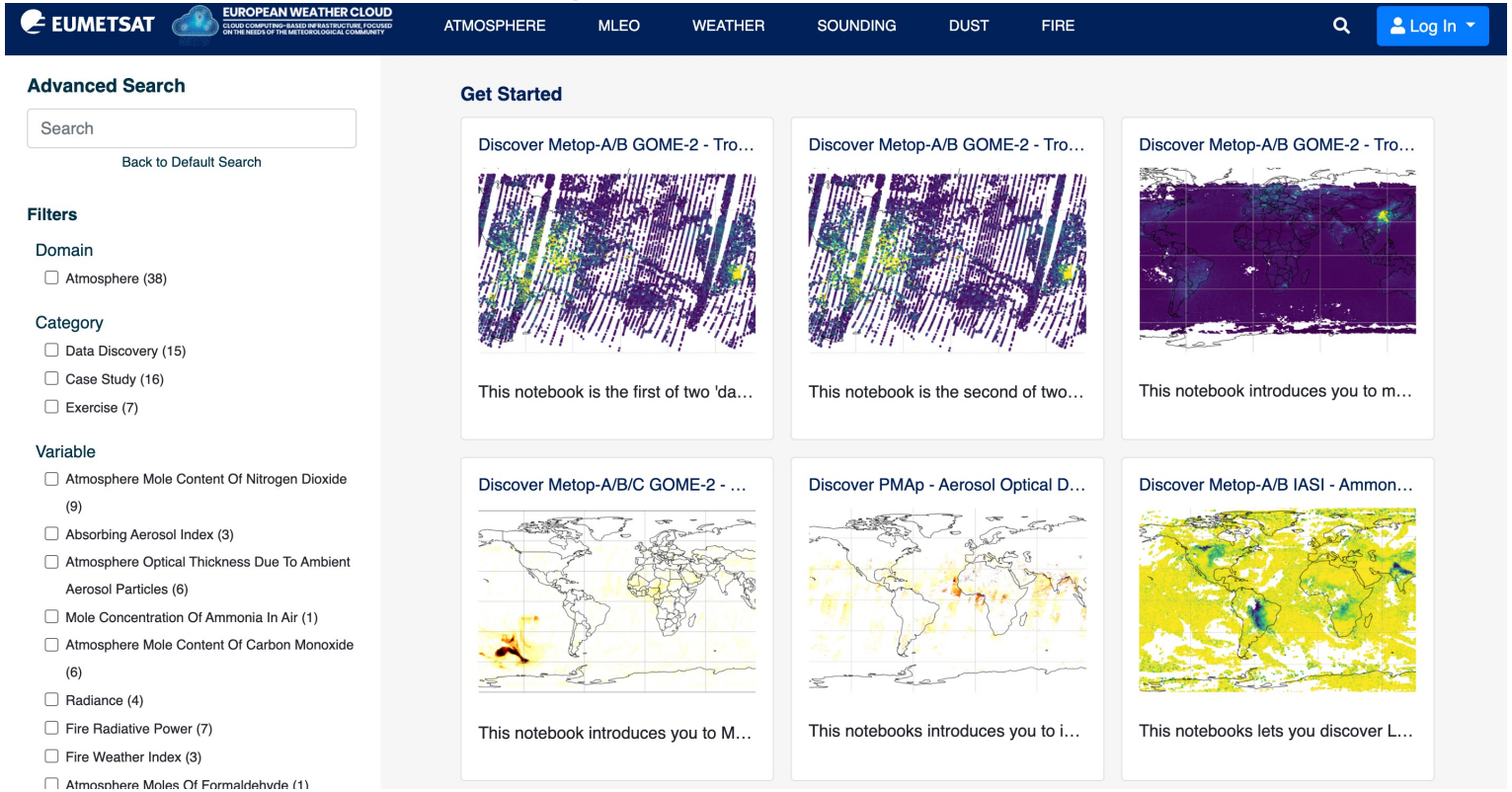
Jupyter notebook examples for EO training @EUMETSAT

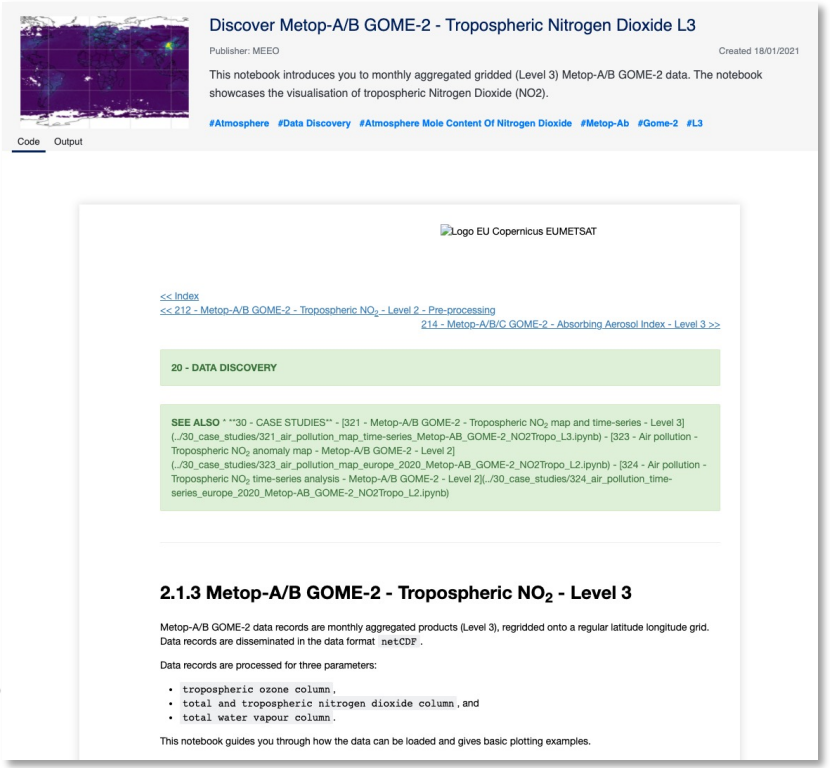
Trainhub – EUMETSAT's Jupyter notebook portal

https://trainhub.eumetsat.int/

Different domains

Login to a Jupyterhub-based processing platform





Search functionality based on notebook metadata

Overview of training notebooks available

Static rendering of the notebook

Best practices: A need for educational and reusable notebooks

- to make notebooks educational and reproducible
- To cater for different levels of data, thematic and programming literacy

Systematic use of recommended Python libraries for handling EO data

Modularisation of content, e.g. to outsource functions

Use of instructional design patterns, such as navigation pane or a table of contents

Provide guided learning pathways, e.g. follow a naming nomenclature to order notebooks in an expected sequence or include an index notebook.

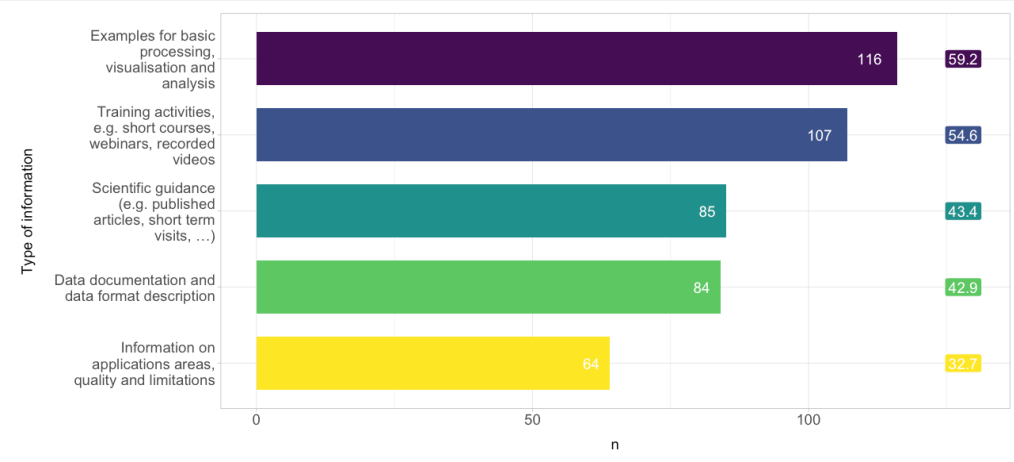
LTPy: Feedback from training participants

Since 2019, 1085 learners trained in 16 training events

Type of training	Number of events	Number of participants
Training schools	6	553
Thematic expert workshops	2	130
Short courses	8	402
Total	**16**	**1085**

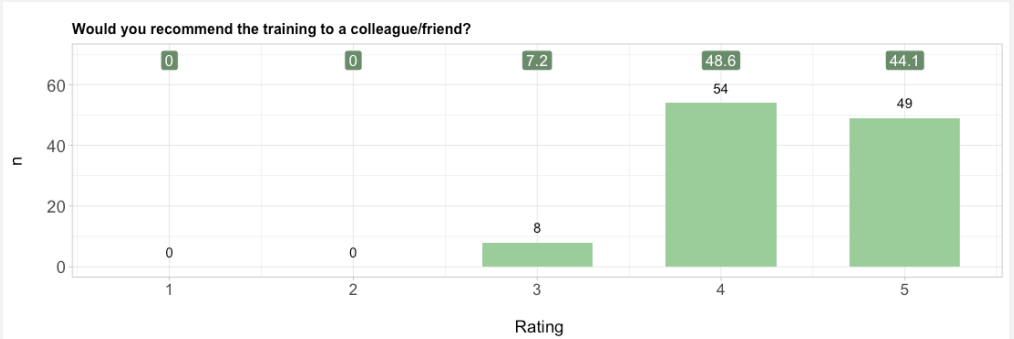
Needs

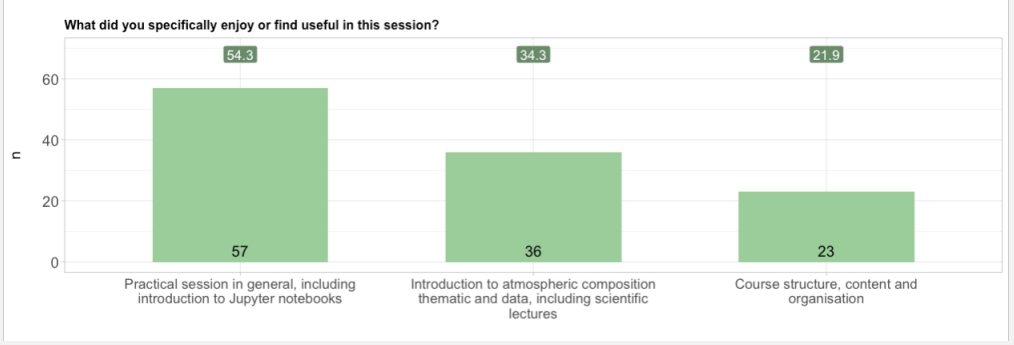
Training in line with user needs. More than half consider *examples for basic processing, visualisation and analysis* and *training activities* as helpful



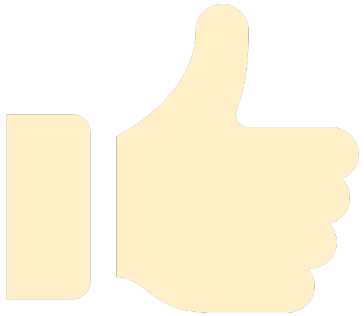
Training feedback

Feedback overall very positive. Practical training part with Jupyter notebooks is particularly useful.





Lessons learned – Things that work well and don't work so well

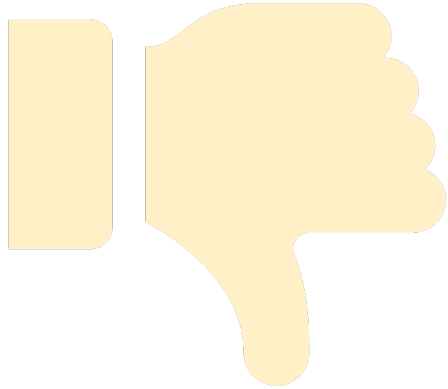


Advantages

- Notebooks remove elements of fear – people are more familiar with a browser than a command line
- Notebooks allow to break up the code into step-by-step workflows and also to describe the workflow in richer ways
- The wider ecosystem, including sharing and hosting notebooks, make notebooks well suited as a teaching tool
- Useful as training material in an instructor-led setting and for self-paced learning alike

- Accessing data and using Python in the browser is a paradigm-change
- Compromise between having a pre-installed Python environment setup vs. learning how to setup the environment
- Compromise between downloading data vs. making data already available on the training platform (fragmentation of data landscape a barrier)
- Maintenance of notebooks – especially with many contributors.

Challenges



Conclusion & Outlook

- In less than a decade, Jupyter notebooks became the de-facto standard to conduct data science, including analysing Earth observation data.
- Jupyter notebooks and the Jupyter ecosystem, such as JupyterLab, will play a significant role in how users of EO data will access and process data in the future
- Using computational notebooks for training/education brings in a set of additional requirements that require the integration of didactical concepts, instructional design patterns and best practices for coding

Next step

Defining, sharing and implementing quality standards and best practices on how to make notebooks effective, reproducible and educational

Thank you!

Wagemann, J., Fierli, F., Mantovani, S., Siemen, S., Seeger, B. and J. Bendix (2022): Using Jupyter Notebooks for Earth Observation data education. *Remote Sensing. (under review)*