

ESA's 2025 Big Data Foundations for Earth Observation Training Course

Introduction to Big Data for EO

Anca Anghel
ESA, EOP-SGD
22/09/2025



European
Commission

researchLatvia[★]
Value Through Knowledge

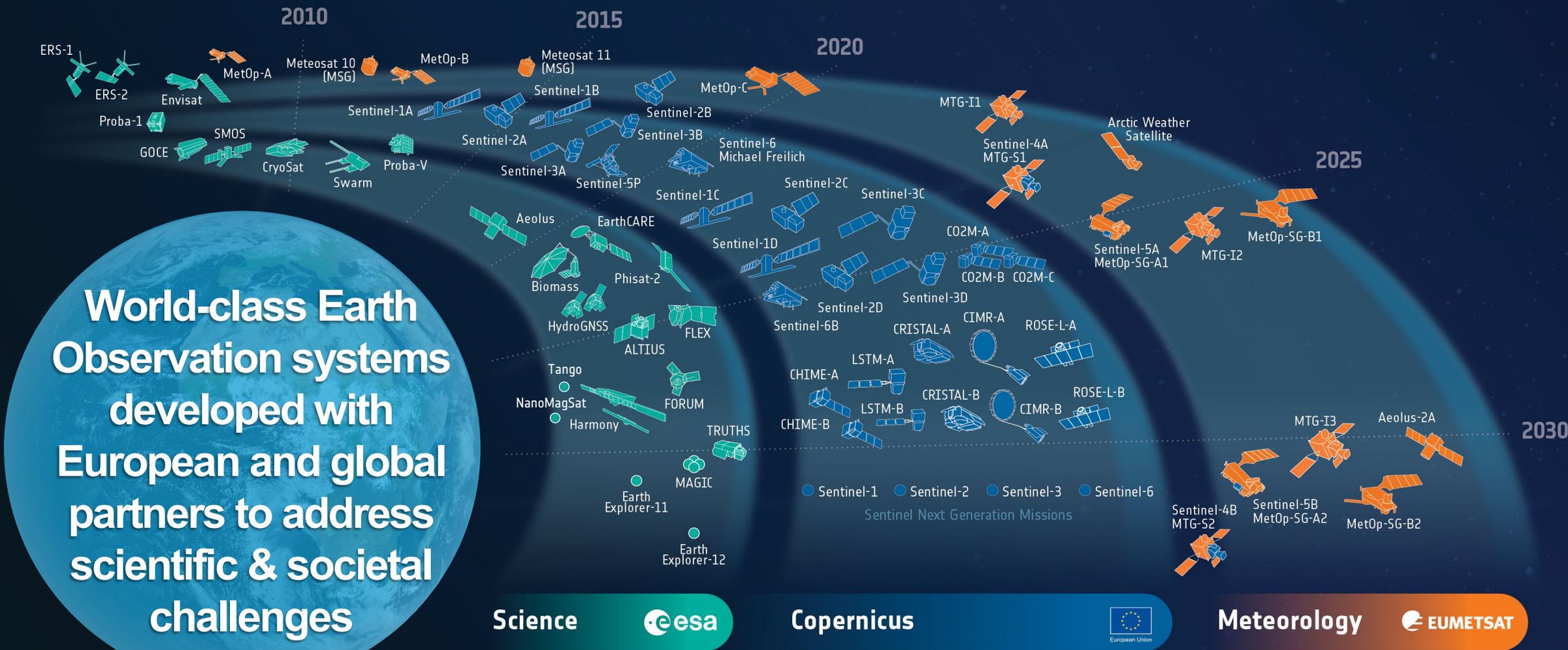


Ministry of
Education and Science
Republic of Latvia



ESA's Earth Observation Missions

Satellites: Heritage 08 / Operational 16 / Developing 40 / Preparing 22 / Total 86



HOW BIG IS A PETABYTE?

11,000 4k movies



It would take you over 2.5 years of nonstop binge watching to get through a petabyte's worth of 4k movies



20+ PB of data
in the Library of Congress



If you took a petabyte's worth of 1GB flash drives and lined them up end to end, they would stretch over

92 football fields



4,000 digital photos
every day for the rest of your life

complex datasets that traditional systems cannot handle.



Google

IBM

ORACLE

How large?

~ PB

Extremely large and complex datasets that traditional systems cannot handle.



How large?

~ PB

How complex?

Diverse collections of structured and unstructured data



Copernicus Programme data: **45.51 PB end of 2023**

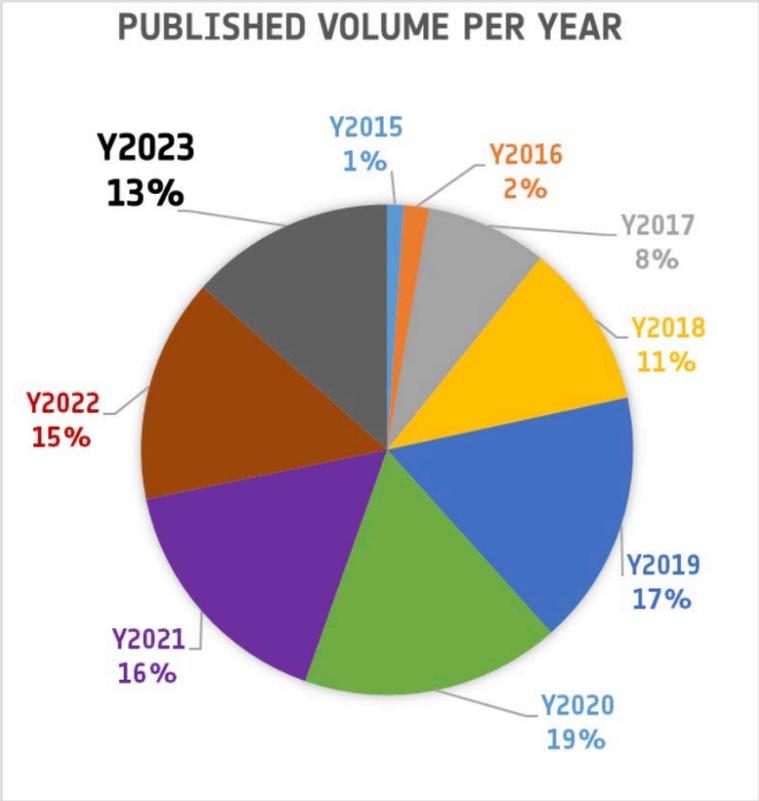
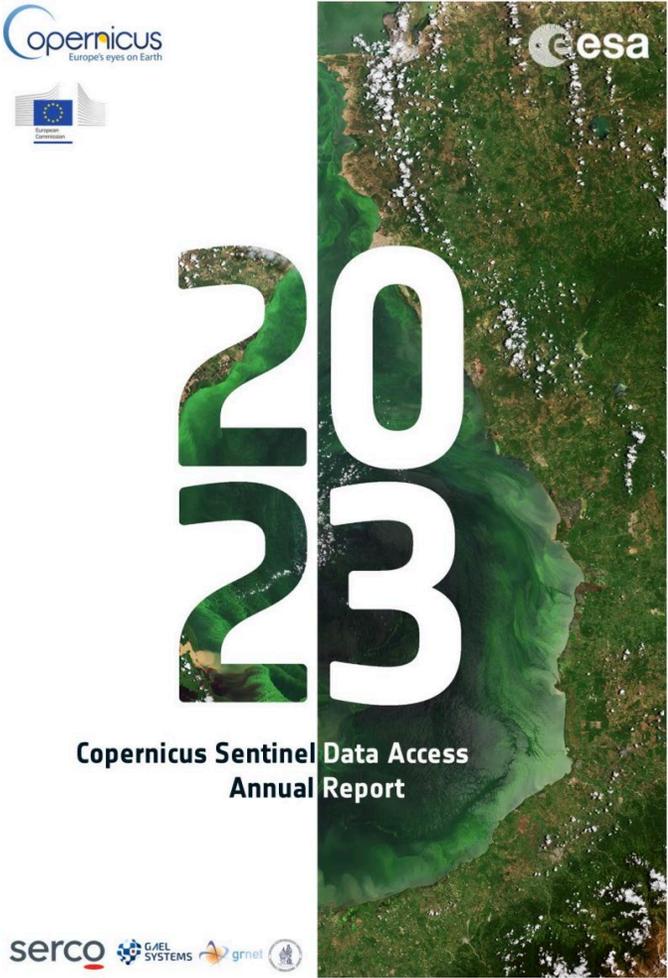
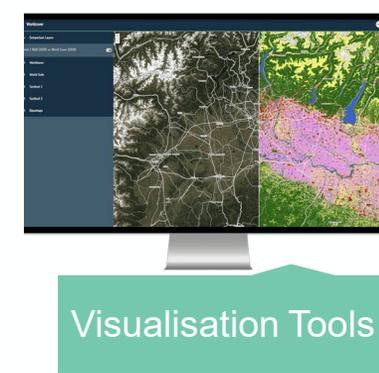
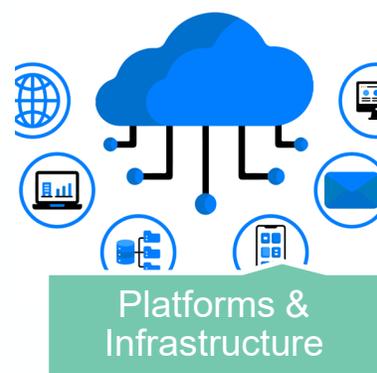
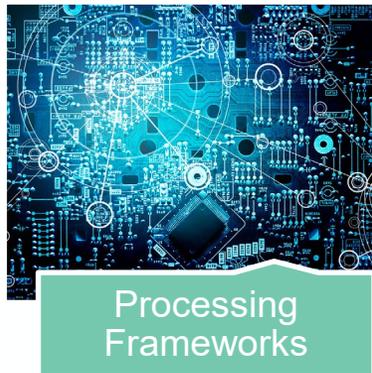


Figure 18: Percentage per reporting year of the total published volume of data since the start of operations (Y2015-Y2023)

Extremely large and complex datasets that traditional systems cannot handle.

What tools and technologies are suitable for Big Data?





DATA ▾

ANALYSIS ▾

SERVICES >

ECOSYSTEM ▾

Explore Data [↗](#)

SUPPORT ▾

LOGIN ▾

Welcome to Copernicus Data Space Ecosystem

Welcome to the Copernicus Data Space Ecosystem, an open ecosystem that provides free instant access to a wide range of data and services from the Copernicus Sentinel missions and more on our planet's land, oceans and atmosphere.

The Copernicus Data Space Ecosystem not only ensures the continuity of the open and free access to Copernicus data but also extends the portfolio for data processing and data access possibilities. Delve into the data immediately via the Copernicus Browser or register to create an account and have an even better comprehensive exploration experience.



Get started

TRUE-COLOUR

FALSE-COLOUR

NDVI

NDWI

SENTINEL-2 QUARTERLY MOSAIC ▾

Systems that allow fast and effective data access and analysis

- **Cloud object storage (s3):** S3 is an object storage service with which you can retrieve data over HTTPS using REST API.
 - **Copernicus Data Space Ecosystem (CDSE)** offers S3 API as one of the main access methods for EO data. It is suitable for Third Party applications that require high-performance parallel access and scalability. Moreover, any user who wants to connect from an external infrastructure to the Copernicus Data Space Ecosystem collection can do so through the S3 protocol.
 - <https://documentation.dataspace.copernicus.eu/APIs/S3.html>



Systems that allow fast and effective data access and analysis

- **File formats optimized for Earth Observation:**

- **Cloud-optimized Geotiff (COG)** : fast access to raster data
- **ZARR/NetCDF/HDF5** : multi-dimensional arrays for time series and climate models



COG
CLOUD OPTIMIZED
GEOTIFF



Zarr

- **Catalogues & APIs**

- **STAC** (spatio-temporal assets catalogue): FAIR metadata for datasets
- **OGC API** (Features, Coverages, Records, Processes) : standardized data discovery and access
- **openEO API**: unified interface to run the same analysis code across different EO backends (Euro Data Cube, Wekeo, GEE, etc.)



Sentinel Data in Cloud optimized format



An initiative to futureproof data from Copernicus Sentinel satellites missions by adopting a common, cloud-native and interoperable solution

-  Enable scalable processing on cloud platforms.
-  Ensure direct compatibility with modern data science programming frameworks.
-  Promote user adoption by providing accessible, user-friendly tools and notebooks.
-  Harmonise data formats across Sentinel missions.

<https://zarr.eopf.copernicus.eu/>

Sentinel Zarr Samples Service STAC API

Search

tion

of the EOPF Sentinel Zarr Samples Service

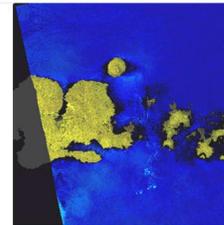
Additional Resources

- OpenAPI service description
- OpenAPI service documentation

Sort by: Tiles | List | Ascending | Descending

Sort by title, description or keywords

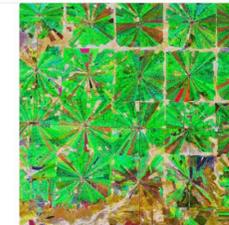
Select keywords



Sentinel-1 Level-1 GRD

Sentinel-1 Level-1 Ground Range Detected (GRD) products focused SAR data that has been detected, multi-looked to ground range using the Earth ellipsoid model...

2022-09-06 13:54:39 UTC - 2025-09-21 08:30:37 UTC



Sentinel-2 Level-1C

The Sentinel-2 Level-1C product is composed of 110x110 km² tiles (ortho-images in UTM/WGS84 projection). Earth is subdivided on a predefined set of tiles, defined in UTM/WGS84 projection and...

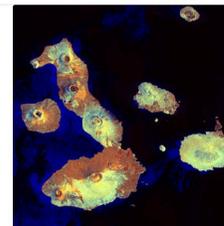
2015-10-22 10:10:52 UTC - 2025-09-21 08:35:31 UTC



Sentinel-3 OLCI Level-1 ERR

The Sentinel-3 OLCI L1 ERR product provides TOA radiances at reduced resolution for each pixel in the instrument grid, each view and each OLCI channel, plus annotation data associated to OLCI...

2025-04-28 07:36:10 UTC - 2025-09-21 07:37:33 UTC



Sentinel-1 Level-1 SLC



Sentinel-2 Level-2A



Sentinel-3 OLCI Level-2 LEP



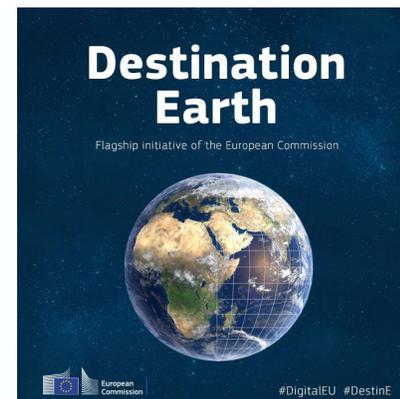
EO Datasets are too large for a single machine, so distributed & parallel computing is needed

- Apache Spark / Dask : scalable data analytics on clusters
- Hadoop (older , less used now for EO)
- Flink / Kafka : real-time streaming, useful for near-real-time EO data, e.g. disasters
- GPU acceleration: CUDA, Tensorflow, pyTorch for AI on satellite imagery



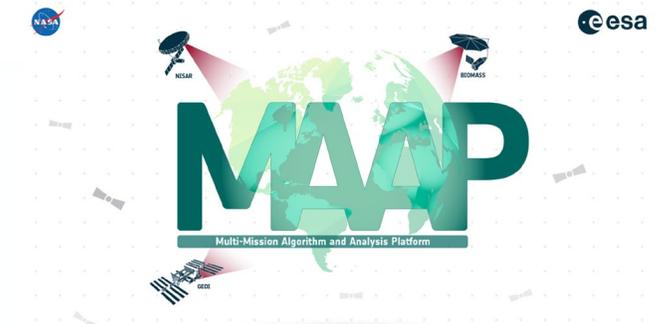
There is a growing ecosystem of cloud-based platforms for EO, some for general purpose, other specialised for applications. Federated approaches are emerging

- **EU:** Copernicus Data Space Ecosystem, Wekeo, DestinE,



There is a growing ecosystem of cloud-based platforms for EO, some for general purpose, other specialised for applications. Federated approaches are emerging

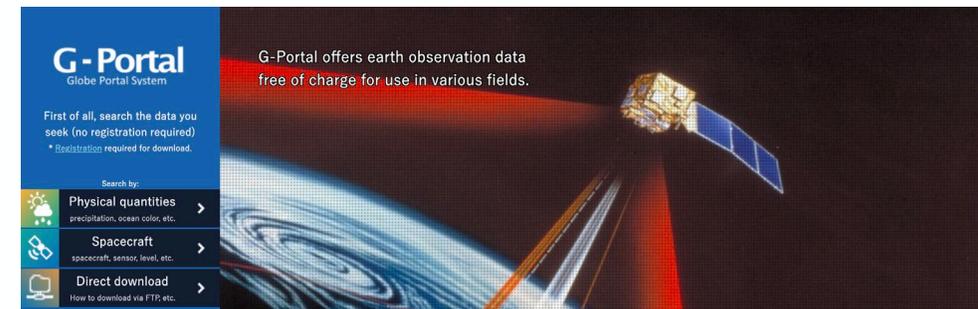
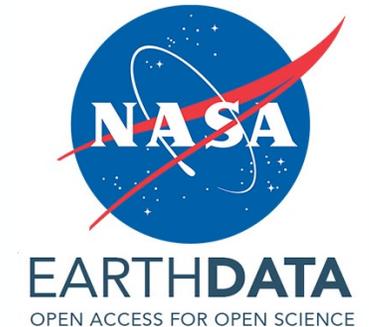
- **EU:** Copernicus Data Space Ecosystem, Wekeo, DestinE,
- **ESA:** Euro Data Cube, Earth System Data Lab, openEO Platform, Geohazards TEP, MAAP, etc.



There is a growing ecosystem of cloud-based platforms for EO, some for general purpose, other specialised for applications.

Federated approaches are emerging

- **EU:** Copernicus Data Space Ecosystem, Wekeo, DestinE,
- **ESA:** Euro Data Cube, Earth System Data Lab, openEO Platform, Geohazards TEP, MAAP, etc.
- **Globally:** NASA EarthData, MAAP, JAXA G-Portal
- **Community frameworks:** Pangeo (xarray + Dask + Jupyter for EO)



EO Datasets are massive, so tools must support scalable geospatial analyses, and automation of workflows

- **Geospatial Libraries:**

- **Xarray** : handles labeled multi-dimensional arrays (time, latitude, longitude, bands)
- **Dask**: parallel & distributed computation, works with xarray to scale from laptop to cloud/cluster
- **Rasterio**: reads/writes raster data (GeoTIFF, COG)
- **GeoPandas**: extends Pandas to handle vector geospatial data (shapes, regions)
- **GDAL**: core geospatial processing library (almost every EO tool depends on it!)



- **APIs:** openEO API, OGC API Processes

- **Interactive Analysis Environments:** e.g. Jupyter Notebooks/ Jupyter Hub – standard environment for EO Big Data exploration.



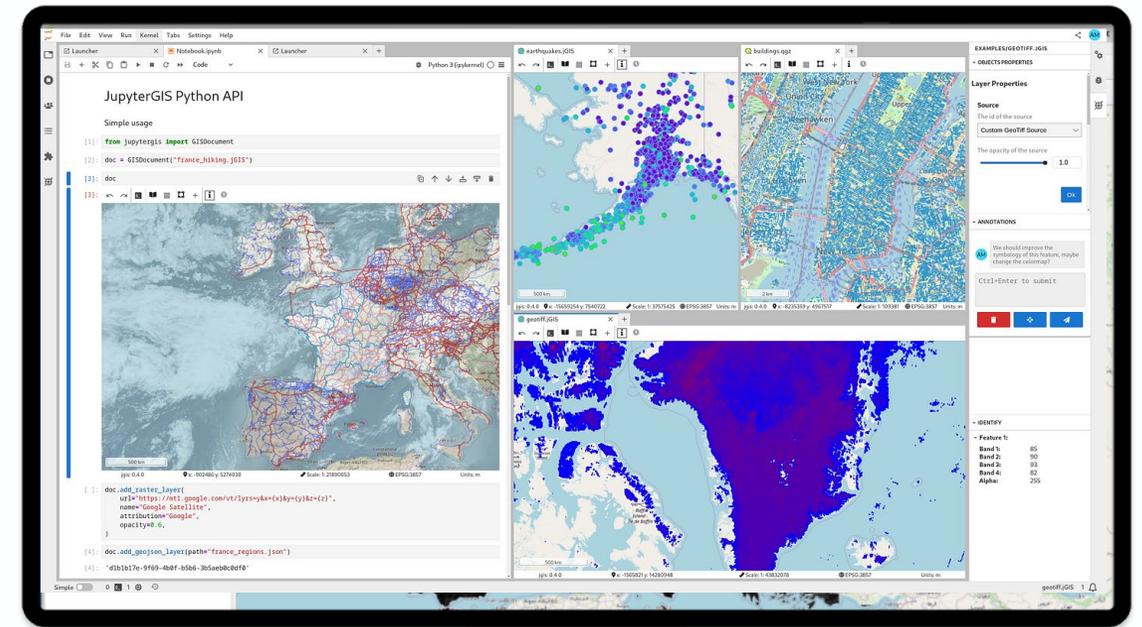
Tools to see, interact with and understand EO Big Data after it has been processed, used by scientists, decision-makers, and general public

- **Notebook based interactive visualization**

- **QGIS for Jupyter:** run QGIS algorithms and visualization inside a notebook

Evolution of the QGIS open-source project (robust tools for spatial data analysis and visualization): enhanced collaboration and accessibility, and a shift towards web-based technologies

[JupyterGIS](https://blog.jupyter.org/real-time-collaboration-and-collaborative-editing-for-gis-workflows-with-jupyter-and-qgis-d25dbe2832a6), a web-based, collaborative, and extensible interface for GIS, leveraging the JupyterLab application framework and integrating seamlessly with the Jupyter notebook interface.



<https://blog.jupyter.org/real-time-collaboration-and-collaborative-editing-for-gis-workflows-with-jupyter-and-qgis-d25dbe2832a6>

Tools to see, interact with and understand EO Big Data after it has been processed, used by scientists, decision-makers, and general public

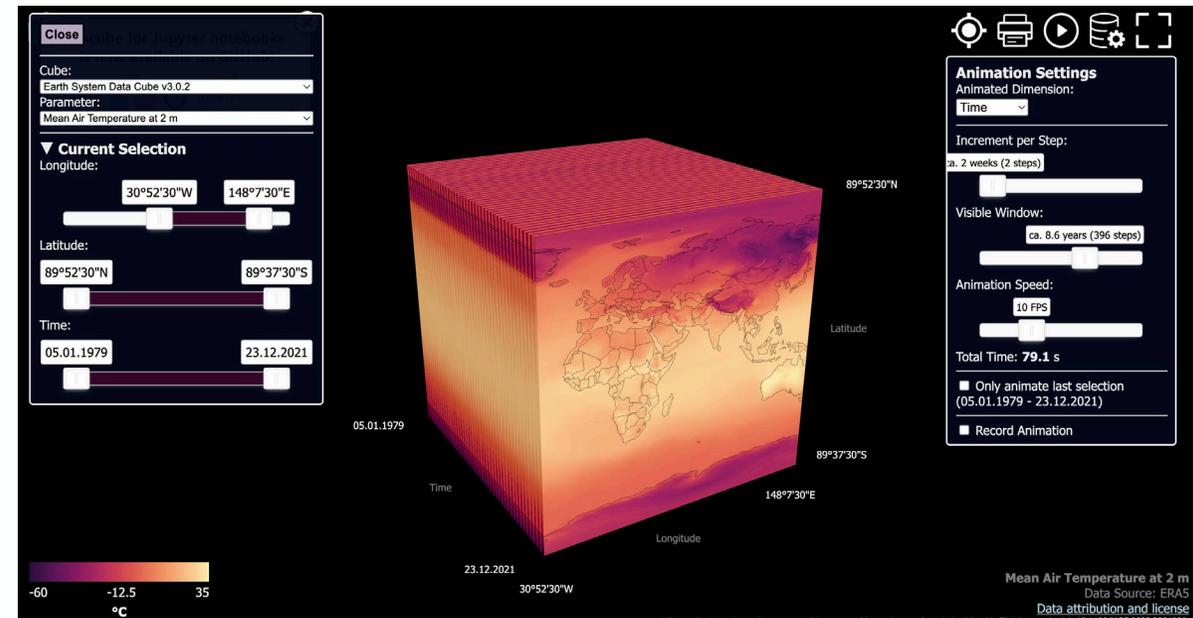
- **Web-based EO Exploration tools (explore global-scale data without download)**
 - **Copernicus Browser (Sentinel Hub):** explore Sentinel data on the fly, compute indices and time series
 - **Xcube viewer:** explore and visualize EO data cubes (zarr, xcube) → integrates also with Jupyter
 - **Lexcube:** interactive 3D EO data cube viewer → integrates also with Jupyter
 - **EODASH.org :** customizable EO data visualization through a series of configurable widgets
- **Web formats:** Leaflet, CesiumJS: to build interactive EO apps
- **Dashboards and storytelling:** bridging science, policy and citizens

Tools to see, interact with and understand EO Big Data after it has been processed, used by scientists, decision-makers, and general public

- **Web-based EO Exploration tools (explore global-scale data without download)**

- **Lexcube:** interactive 3D EO data cube viewer → integrates also with Jupyter

<https://www.lexcube.org/>



EARTH OBSERVING DASHBOARD Home News & Updates Resources Dashboard About

Select indicator (+)

Layers

- Overlay Layers
- Data Layers
 - NO2_daily

Datetime

20.01.2025 02:01

2019 2020 2021 2022 2023 2024

Legend

Tropospheric NO2 [$\mu\text{mol}/\text{m}^2$]

0 50 100 150

Opacity

100%

Base Layers

Riga LATVIA

2025-01-20

Air Quality (tropospheric NO2 concentrations)

atmosphere

Satellite: Sentinel-5P Agency: ESA
Sensor: TROPOMI Extent: 2018-04-30 - 2025-06-23

- Description
- Providers (1)
- Assets (5)
- Links

22/09/2025

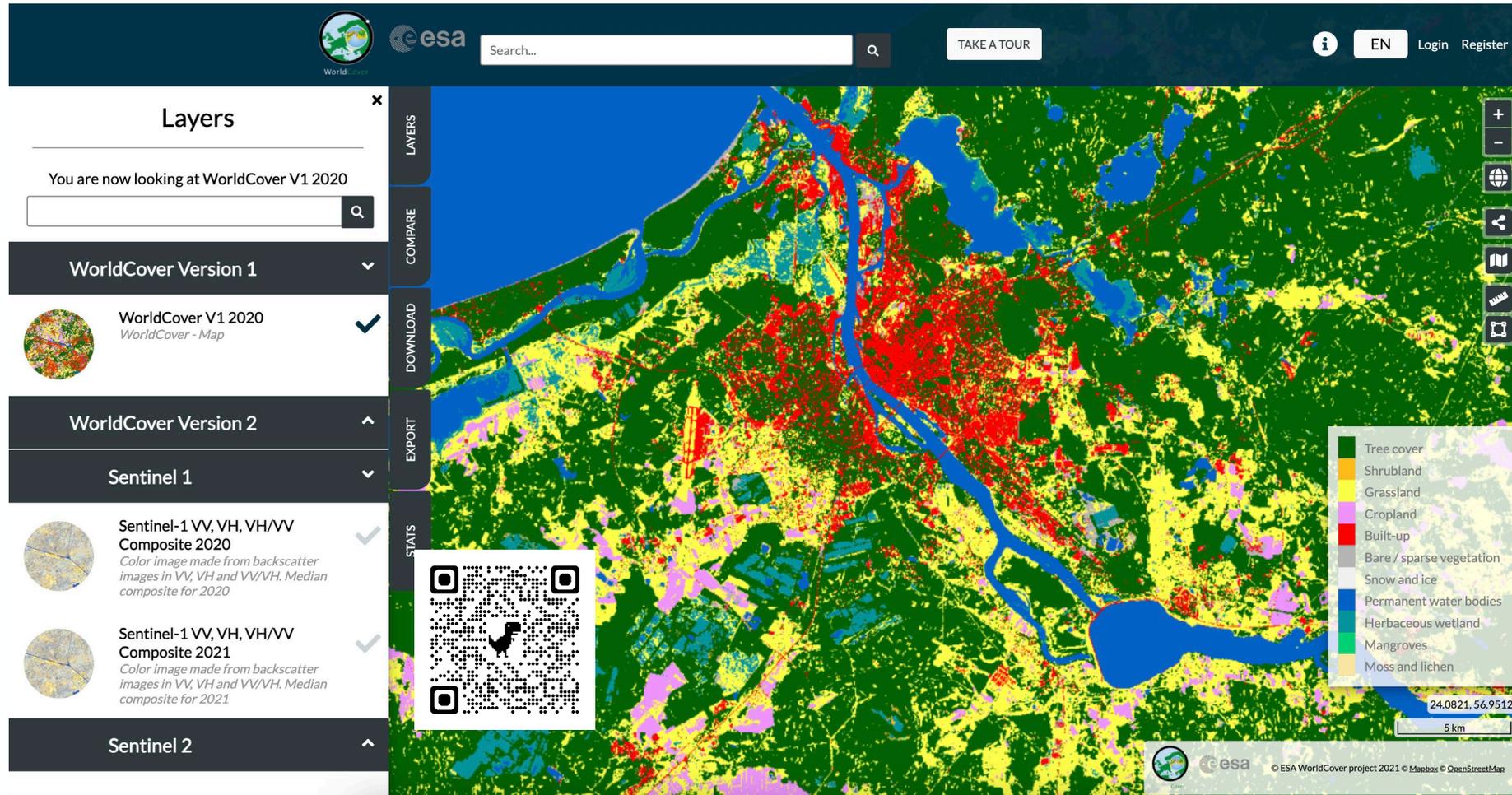
Fast evolving, as the applications mature and the underlying technology matures. Machine Learning, Deep Learning and Foundation Models are widely used to extract knowledge from EO at scale

- Traditional Machine Learning:
 - Random Forests / Gradient Boosting, SVM (Support Vector Machines), Clustering (e.g. k-Means) : mostly classification tasks, e.g. crop type mapping

Example:

ESA's **WorldCover** land cover product (first released at 10 m resolution in 2020 and updated in 2021/2023) is based on **Sentinel-1 and Sentinel-2** data.

For the classification step, the team used an **ensemble of machine learning classifiers** based on **Random Forests (RF)**.



Fast evolving, as the applications mature and the underlying technology matures. Machine Learning, Deep Learning and Foundation Models are widely used to extract knowledge from EO at scale

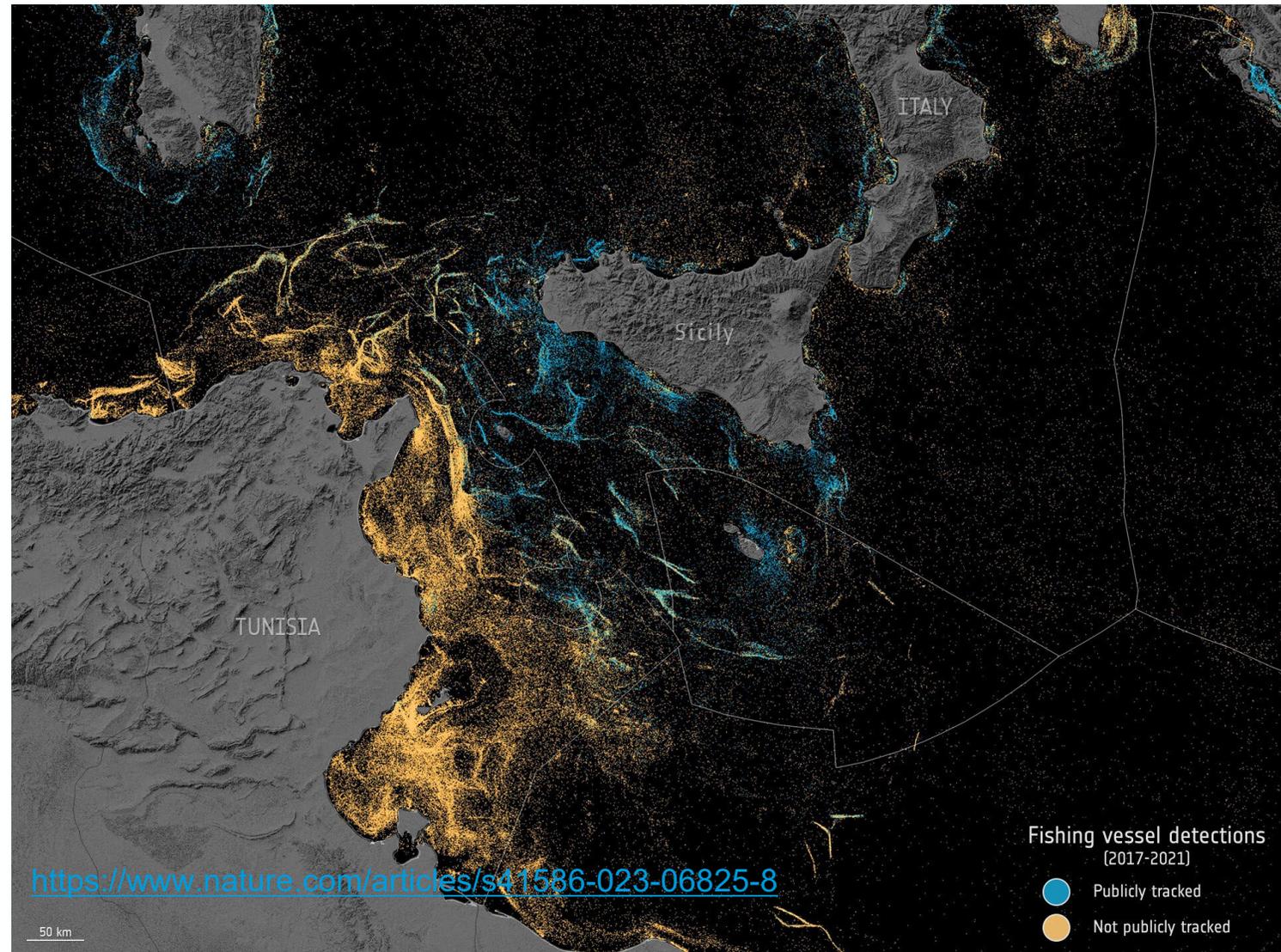
- **Deep Learning for EO**
 - **CNNs (Convolutional Neural Networks):** used for object detection (e.g. ships, buildings), segmentation (e.g. floods, deforestation)
 - **Transformers:** image classification, multi-sensor fusion, spatio-temporal modelling
 - **Generative Models:** super-resolution, gap-filling, synthetic EO data generation

Example:

Sentinel-1 and AI reveal 75% of global fishing vessels not tracked

Use of CNNs for different tasks:

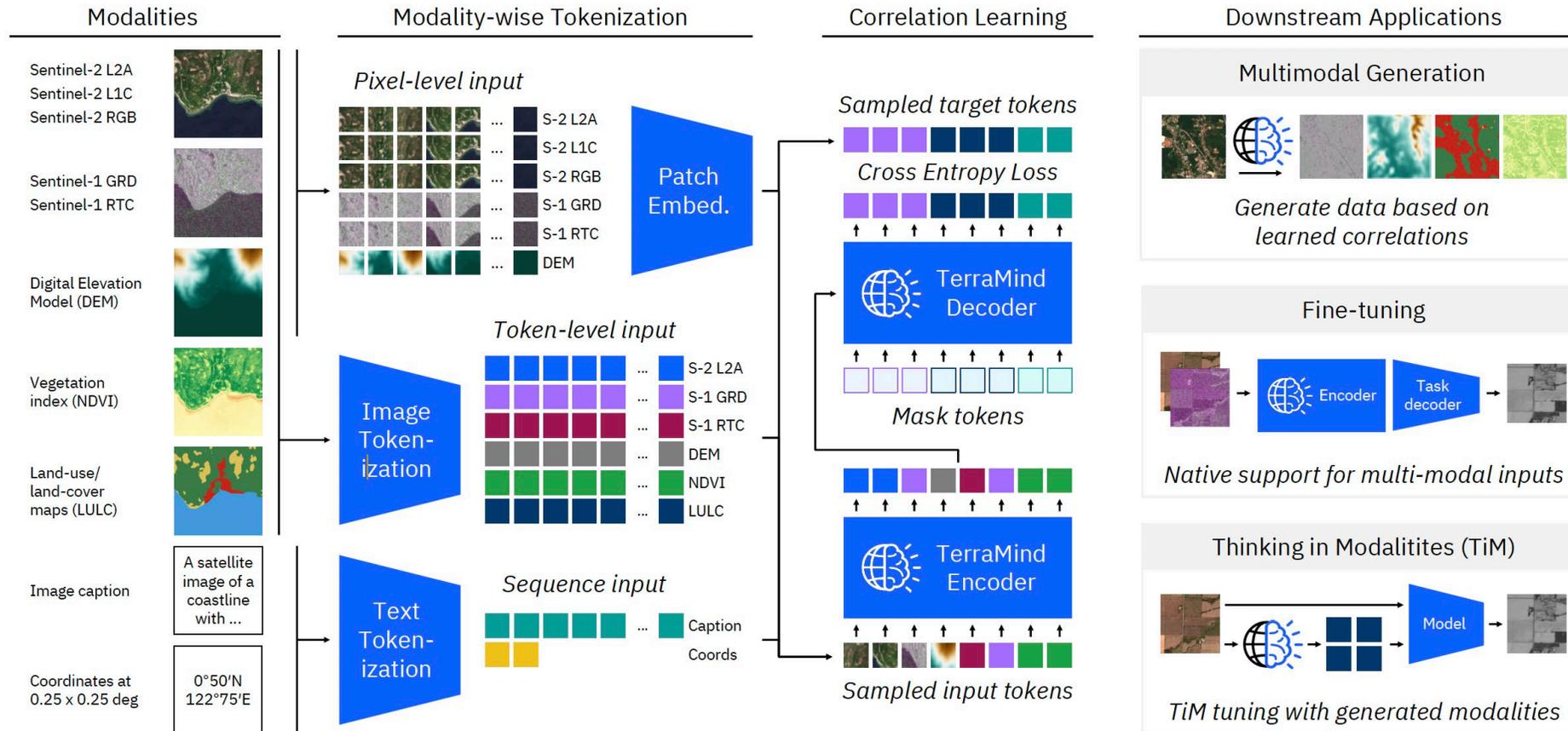
- **Vessel detection + length estimation**
- **Infrastructure classification**
(classifying detected offshore fixed infrastructure into categories (e.g. oil infrastructure, wind infrastructure, other))
- **Fishing vs non-fishing vessel classification** (uses contextual/environmental information to decide whether a vessel is fishing or not.



Fast evolving, as the applications mature and the underlying technology matures. Machine Learning, Deep Learning and Foundation Models are widely used to extract knowledge from EO at scale

Foundation Models for EO: pre-train once, fine-tune for many EO tasks

- **NASA Prithvi** : vision transformer trained on multi-sensor EO data
- **Copernicus FM**: development for Copernicus Sentinels data
- **ESA's TerraMind** best performing generative AI model for Earth observation
- **Climate-FM and Geo-FMs**: bridge EO and Earth System Models



In practice, to predict the risk of water scarcity, researchers need to consider many different factors like land use, climate, vegetation, agricultural activities, and location. Before TerraMind, all of this data was locked away in separate places. Bringing this information together enables users to predict the potential risk of water scarcity informed by a larger, more accurate picture of conditions on Earth.

Overview of TerraMind, first any-to-any generative large scale multimodal model

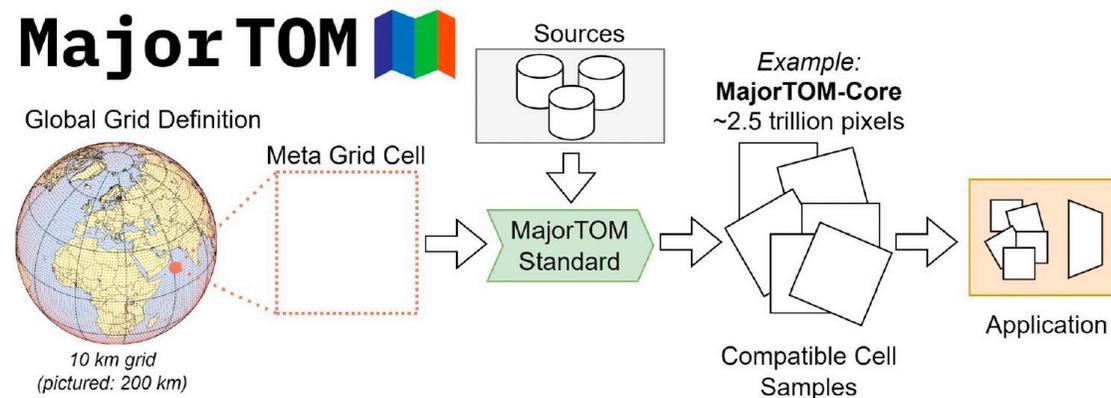
IBM and ESA have released [TerraMind](#), a new earth-observation model that the group has open-sourced on [Hugging Face](#). It was pre-trained on [TerraMesh](#), the largest geospatial data set available, built by researchers as part of the TerraMind project.

Embeddings – AI ready data

- Vector embeddings of Sentinel-1/2 fragments, aligned to a common grid (~10 km cells)
- Generated with self-supervised / foundation models
- ~62 TB raw data processed
- Stored as GeoParquet, aligned spatially and temporally

Embeddings sit between Foundation Models (producing them) and Analysis & Exploration (consuming them).

They are a bridge that make PB EO archives usable for ML at global scale.



2025 ESA–NASA International Workshop on AI Foundation Models for EO



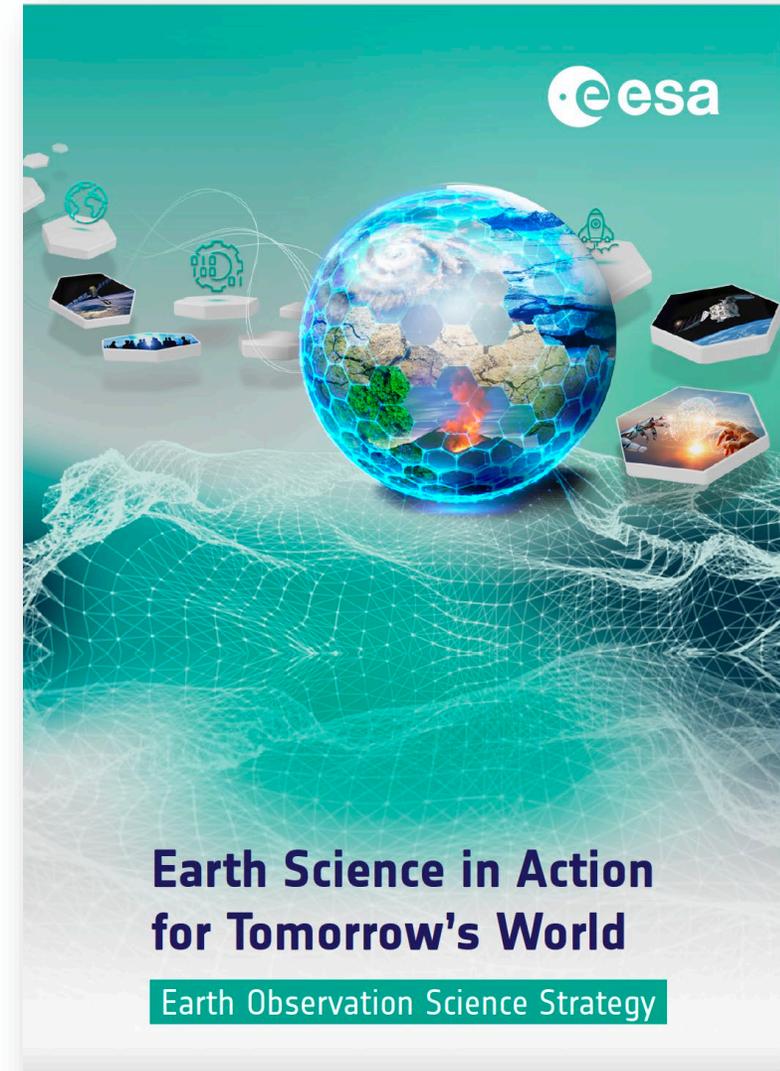
Nicolas Longepe, Hamed Alemohammad, Anca Angheloa, et al. Earth Action in Transition: Highlights from the 2025 ESA-NASA International Workshop on AI Foundation Models for EO. Authorea. July 25, 2025. DOI: 10.22541/au.175346055.53428479/v1

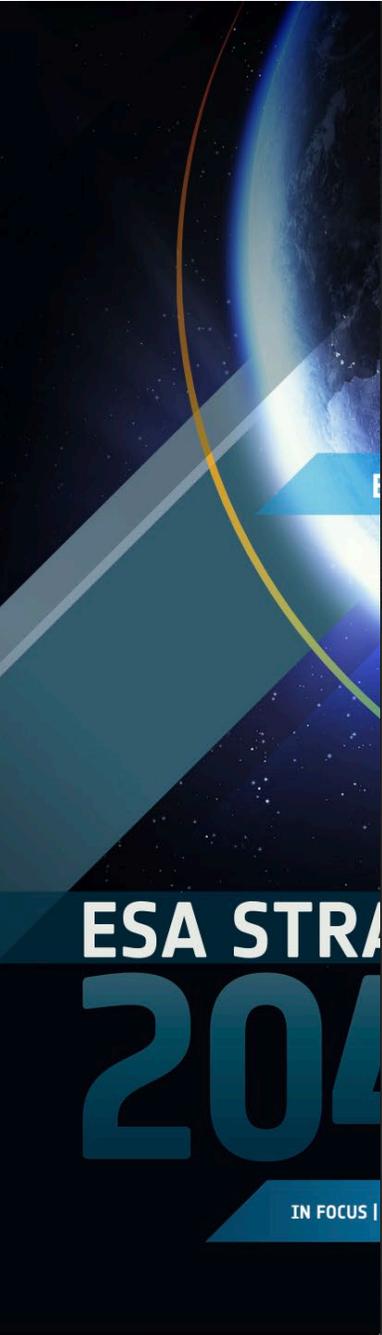
Why do we need Big Data in EO?

“to increase the collective understanding of our evolving planet and the development of actionable information in tackling serious global environmental issues and the resulting challenges.”

ESA Earth Observation Science Strategy, October 2024, <https://doi.org/10.5281/zenodo.13819557>

- Solve complex problems
- Fast & actionable
- Interdisciplinary → Interoperable





We face profound, existential risks driven by sustainability challenges and climate change. The World Economic Forum^b has named three key critical climate issues facing humanity: extreme weather events, such as recent floods in Spain that killed 230 people and caused billions of euros of damage; critical change to Earth's systems; and biodiversity loss and ecosystem collapse. However, the problems go beyond the biosphere as sustainability challenges are arising in space. The number of active satellites in orbit could rise from about 11 000 in 2025 to 60 000 by 2030 meaning constant vigilance is required. Congested orbits, space weather events, and asteroids can endanger satellites and astronauts, and disrupt communications, power grids, and satnav-reliant services on Earth. Pursuing this goal will ensure ESA continues to lead in the use of space, responsibly and sustainably, to address the protection and management of Earth.

^b Source: Global Risks Report 2024

OBJECTIVE 1.1

Develop the technologies, missions, applications and services that will enable progress in acting to address climate change, degradation of the environment, and pressure on natural resources.

By 2040, ESA will support Europe through space-based systems and digital twins of our planet to foster scientific understanding of the processes of Earth's natural systems.

STRATEGIC ACTIONS

- Support Europe as a driver of a thriving Earth observation ecosystem powered by cutting-edge tech, next-generation satellites, and comprehensive pan-European system architecture.
- Identify climate and environmental tipping points by the early 2030s via new observation missions and enhanced predictive modelling capabilities, including AI-powered digital twins.
- Support the development of climate change mitigation and adaptation strategies through data-driven approaches and accurate environmental forecasting, thereby contributing to climate change policies and the UN Sustainable Development Goals.
- Reinforce collaborative environmental efforts and strengthen partnerships with international organisations, national space agencies, and commercial entities.

OBJECTIVE 1.2

Spearhead a greener, circular economy in space with standards for sustainability and a zero-debris environment.

ESA will pursue a zero debris future by 2030 via stringent target collision probabilities for spacecraft below 1 in 1000 and maintain a 99% or greater likelihood of successful disposal, and a sustainable use of space with the establishment of a circular economy by 2040.

STRATEGIC ACTIONS

- Contribute to a net zero debris space environment through technologies and global frameworks, to safeguard the future of Earth orbits.
- Promote a cleaner and greener space sector by compliance with environmental legislation, assessing the life cycle impact of missions, adopting sustainable practices in all areas, and preserving dark and quiet skies in the interests of astronomy.
- Champion a circular space economy via sustainable satellite design and developing in-orbit servicing, assembly, manufacturing and recycling capabilities.

OBJECTIVE 1.1

Develop the technologies, missions, applications and services that will enable progress in acting to address climate change, degradation of the environment, and pressure on natural resources.

By 2040, ESA will support Europe through space-based systems and digital twins of our planet to foster scientific understanding of the processes of Earth's natural systems.

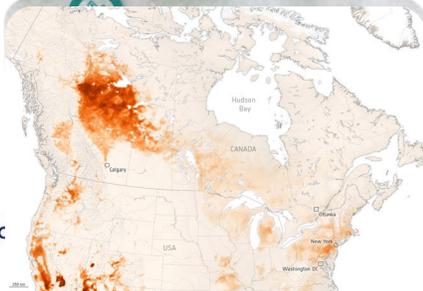
STRATEGIC ACTIONS

- Support Europe as a driver of a thriving Earth observation ecosystem powered by cutting-edge tech, next-generation satellites, and comprehensive pan-European system architecture.
- Identify climate and environmental tipping points by the early 2030s via new observation missions and enhanced predictive modelling capabilities, including AI-powered digital twins.
- Support the development of climate change mitigation and adaptation strategies through data-driven approaches and accurate environmental forecasting, thereby contributing to climate change policies and the UN Sustainable Development Goals.
- Reinforce collaborative environmental efforts and strengthen partnerships with international organisations, national space agencies, and commercial entities.



SCOPE OF THE SPACE MARKET

01 Market segment



02 Applications

Market segment

- Satellite communications
- Earth observation
- Satellite navigation
- Exploration
- Science
- Security
- Space logistics
- Technology
- Space tourism

Applications



Satellite communications*

Considers revenues associated with the manufacturing, launch and ground systems of communication satellites. Includes revenues from satellite communication operators (FSS and MSS) as well as subscriber revenues from satellite radio. Includes the sales of user terminals to receive the signal.



Earth observation*

Considers revenues associated with the manufacturing, launch and ground systems of EO satellites. Includes revenues from the sale of EO imagery as well as value-added services extracted from the raw data.



Satellite navigation*

Considers revenues associated with the manufacturing, launch and ground systems of navigation satellites. Includes revenues from GNSS device sales, as well as the added-value service revenues including all GNSS-enabled services, location-based applications (e.g., navigation), GNSS-attributable revenues of smartphone apps including sales revenue, advertisements, and in-app purchases.



Other applications

Includes other commercial activities such as SSA, space logistics applications (e.g., last-mile delivery, life extension services, active debris removal...), technology demonstration missions, Ground Segment as a Service, etc.

Source: [Euroconsult](#)

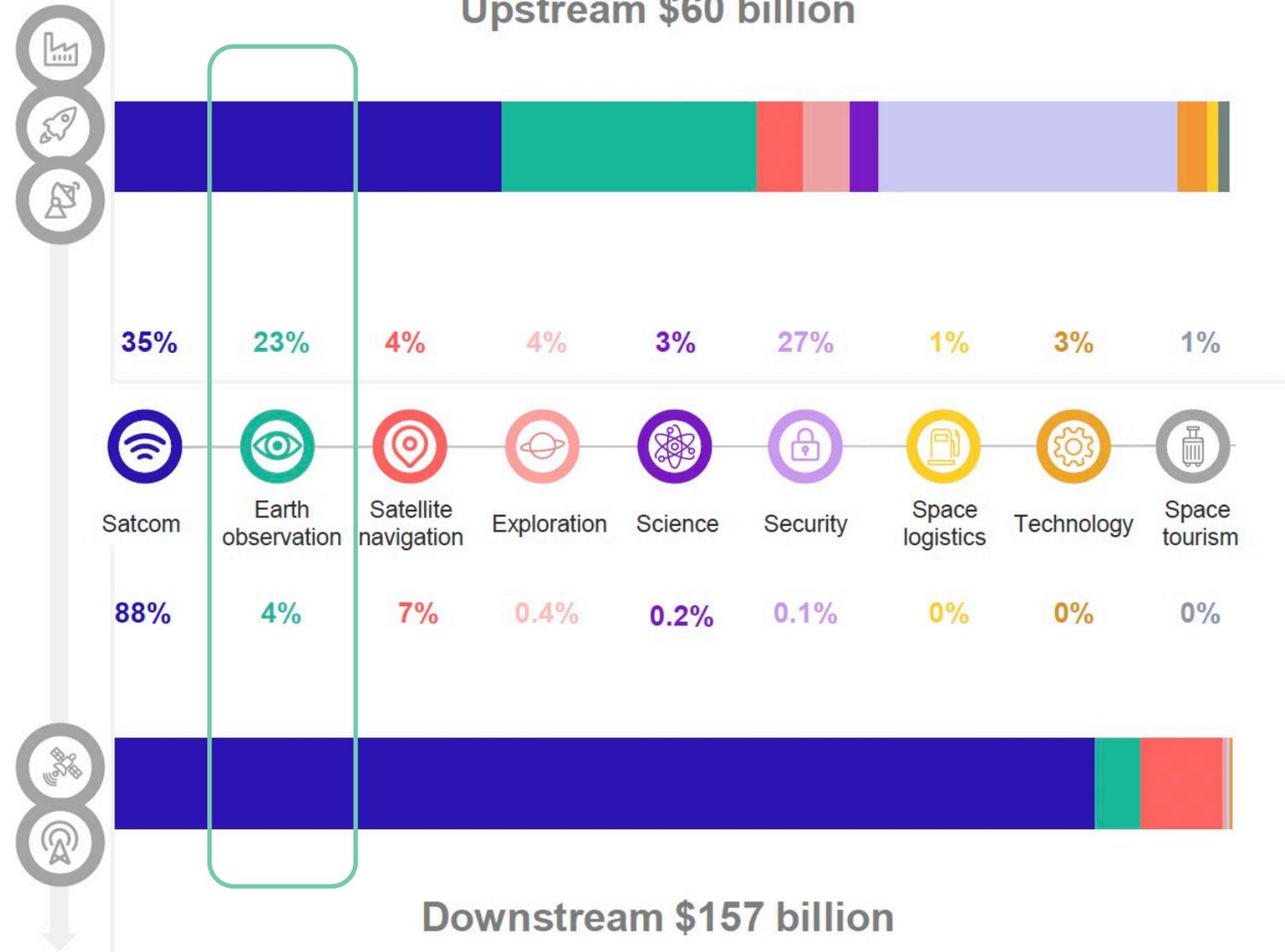
Global space market by application – 2024



In USD

Upstream \$60 billion

Downstream \$157 billion

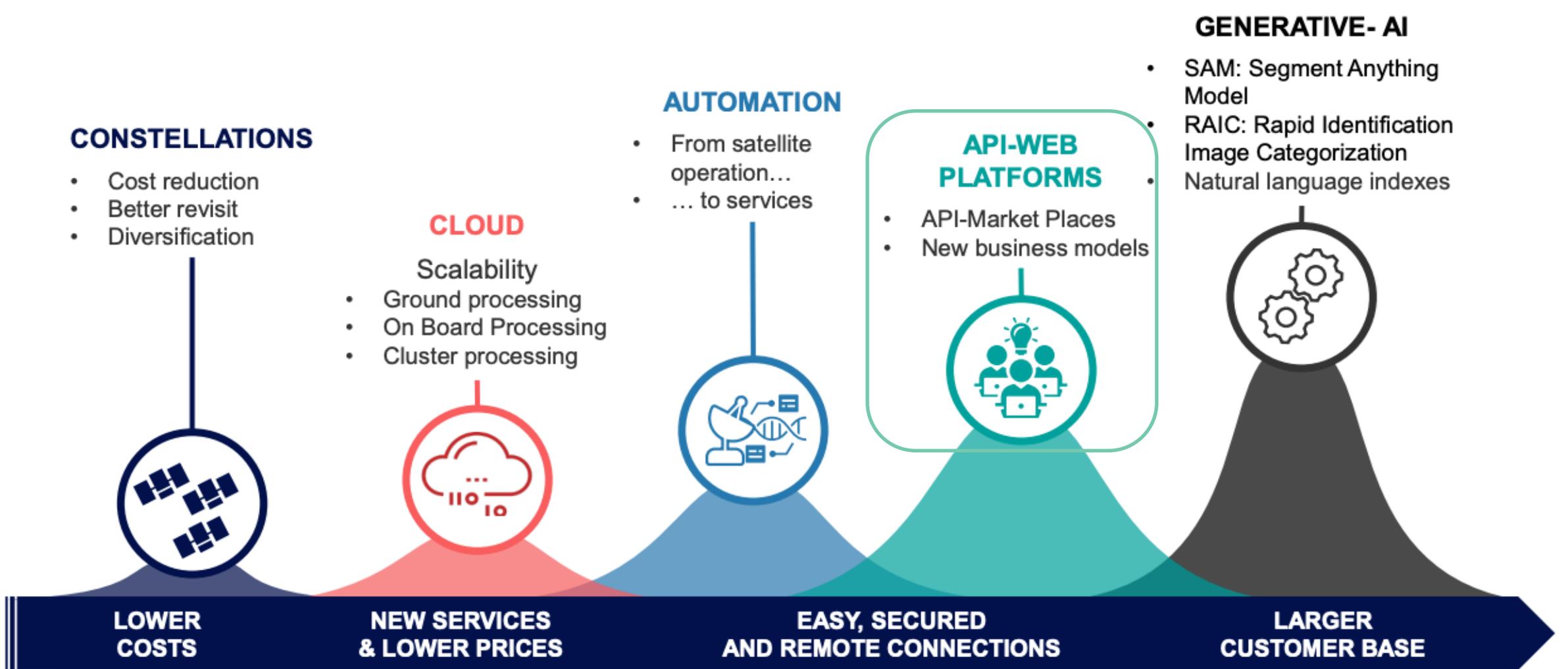


Downstream EO:

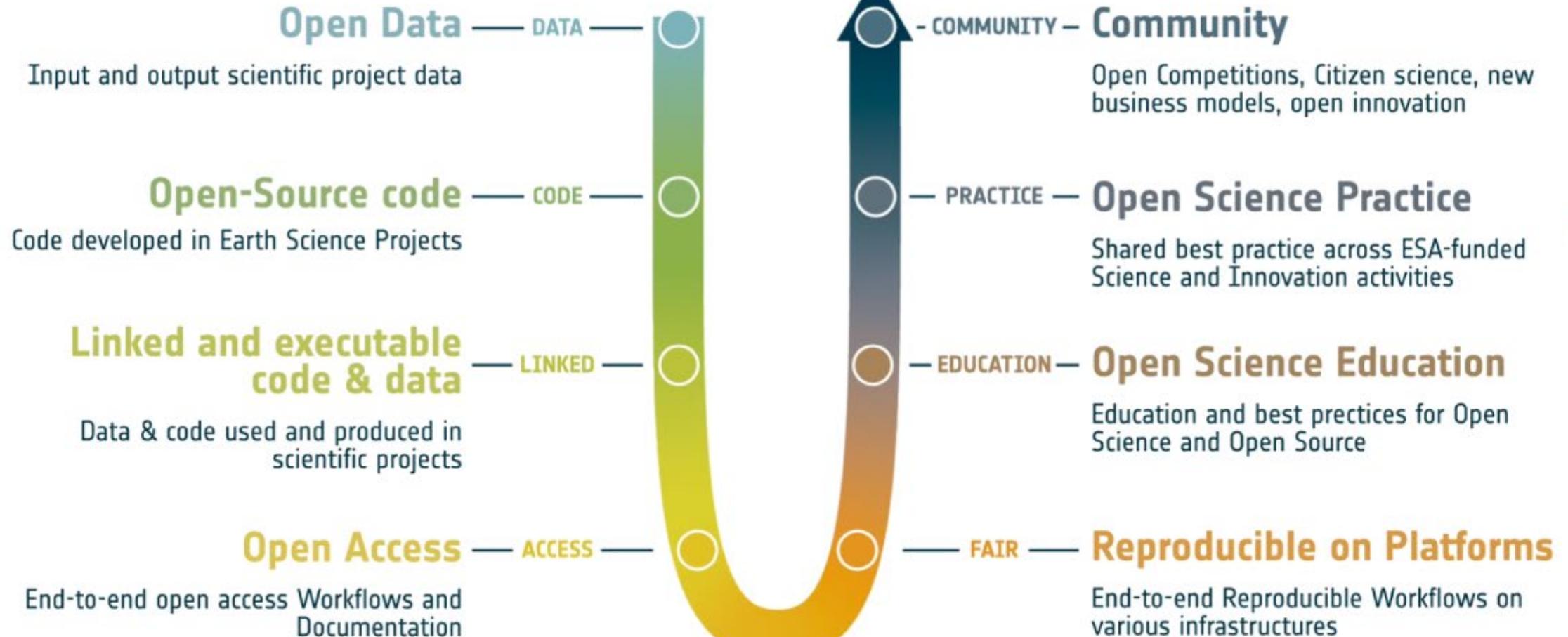
- Commercial data sales (38%)
- Value-added services (62%)

Source: [Euroconsult](#)





EARTH OBSERVATION OPEN SCIENCE AND INNOVATION





EarthCODE **empowers scientists** to adopt FAIR and Open Science, when conducting research **on integrated cloud environments or on a different infrastructure of their choice (e.g.. national infrastructures)**.

It grows the impact of FAIR scientific data and methods by providing a **hub for dissemination and collaboration**.

It **supports Earth Action, accelerating scientific discovery** by maximizing reuse and reproducibility of FAIR and Open Earth science data and methods.

EarthCODE aims to bridge the gap



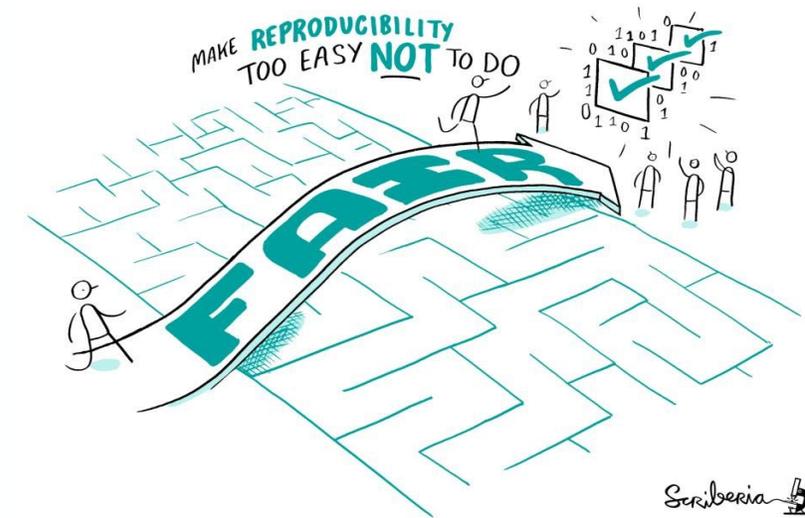
earthcode.esa.int



Making ESA Projects' outputs widely useable is possible if FAIR & Open is **practiced by the community**

EarthCODE offers:

- The tools
- The means
- The training
- The support



...for science teams to **do FAIR & Open Science**



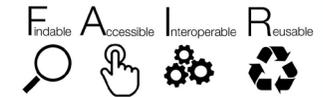
What is EarthCODE?



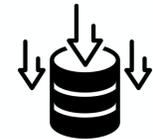
A strategic ESA initiative that develops technology for scientists to:

Adopt FAIR and Open Science:

- Ensure their developed **scientific workflows are FAIR**
- Ensure their produced **datasets are FAIR**



Publish **their research safely for the long-term** in an ESA repository (ESA PRR)



Get **ESA-sponsored access to cloud computing** platforms



Discover and reuse each other's research (ethically)



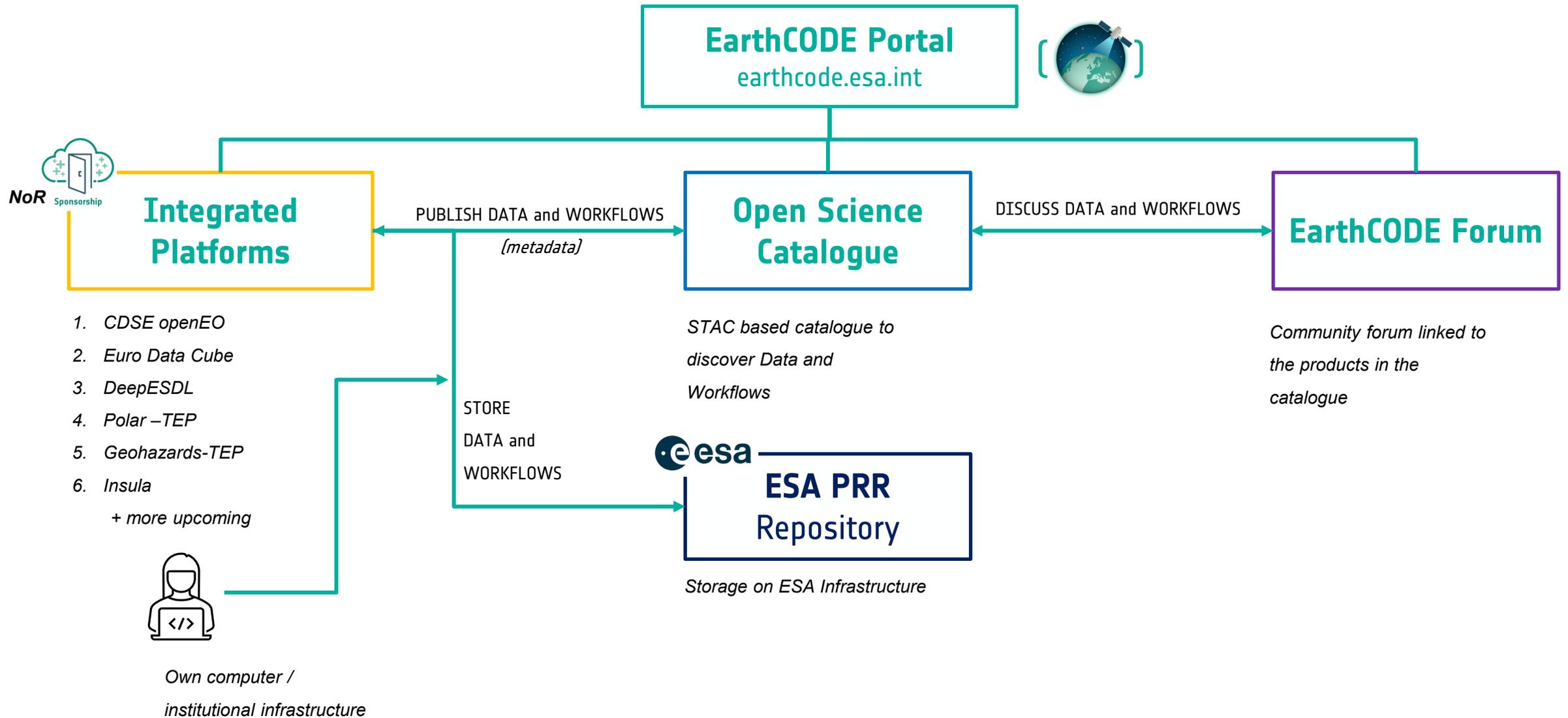
Collaborate on research



Learn Open Science practices



How does it work?





Engagement with National Infrastructures

YEAR 1

2024

ONGOING

Foundational Development

OBJECTIVE:

Establish core architecture and user-centric design

OUTCOME:

outcome="Develop the EarthCODE architecture catering to the needs of the ESA Earth System Science activities; identify and procure services and technologies; pursue synergies with open-source communities and and coordination with other initiatives (EOEPCA+, APEX)

YEAR 2

2025

COMING UP

Initial Platform Deployment

OBJECTIVE:

Deploy an MVP (minimum viable product) to selected users

OUTCOME:

Launch the Portal with core features such as data discovery, workflow management, and integration with cloud computing services; develop a library of tutorials, documentation, and guides to ensure ease of use; engage early users to provide feedback; demonstrate the platform's value

YEAR 2+

2025-2026

COMING UP

Community Expansion and Open Access

OBJECTIVE:

Foster collaboration and broaden community access

OUTCOME:

Open the platform to the wider scientific community with features for open access, publishing, and citation management; create a library of tutorials, documentation, and guides including video content; engage with the community through webinars, forums, and social media

YEAR 3+

2026-2028

COMING UP

Enhance scalability and functionality

OBJECTIVE:

Expand features and refine scalability

OUTCOME:

Incorporate additional platform services and applications to meet diverse project needs; integrate more advanced tools for automation, complex analyses, and machine learning workflows; enhance reproducibility capabilities; demonstrate reproducible workflows across multiple platforms

YEAR 4+

Beyond 2028

COMING UP

Long-Term Sustainability and Growth

OBJECTIVE:

Ensure sustainability and adaptability for future needs

OUTCOME:

Establish a governance model for long-term maintenance and evolution; continuously add new platform services and refine existing tools based on user feedback and evolving scientific challenges; integrate advanced technologies such as AI and advanced visualisation, enhance scalability and capabilities

Welcome to EarthCODE's User Documentation

Learn how to publish, access, integrate, and reuse EarthCODE's data and workflows to advance open science.

Getting started

Community and Collaboration



Computational Research

Learn about advanced computational tools for Earth system science in collaborative research environments.



Datasets

Learn about accessing diverse, high-quality Earth observation datasets for scientific analysis and discovery.



Workflows

Learn how to find, publish, and reuse FAIR workflows to enhance reproducibility and collaboration.



Community

Learn more about interacting with EarthCODE's community, dedicated to FAIR Open Science and sustainable innovation.



<https://esa-earthcode.github.io/documentation/>

Creating a STAC Item Catalog

Context

Purpose

The purpose of this tutorial is to learn how to share research outcomes with the wider research community in Open Science Catalog. This can be done by creating STAC Catalogs that describe a specific dataset you are willing to share.

This tutorial provides steps necessary to Create STAC Item Catalog in a semi-automated way, using the PySTAC library. By following these steps you will be able to create a **self-contained** STAC Catalog with individual items in a JSON format. This catalog should be hosted in your own (or institutional) public GitHub repository to ensure it is accessible. (See more on the requirements about this Catalog in the [documentation](#)).

In this example we will upload it to an open-access repository on GitHub. In the next tutorial we will create the actual Open Science Catalog entry, where we will create a full metadata description of our dataset with a link to this Item Catalog.

STAC Items

A STAC Item is the lowest level component of a STAC catalog. All STAC Items must have an associated data **Asset**, in addition to the Asset (which you can think of as a data file), the Item also contains metadata about the data itself, such as:

- SpatioTemporal extent including start and end time and geographical extent (coordinates)
- Variables
- File type
- File size

⚡ Important

Think about the persistence of your data! If your data files are not currently stored in an open-access and persistent storage, you can [contact the ESA team](#) who will assist you to upload your data to the ESA Project Results Repository (PRR). The same applies for the repository we will upload our STAC Item Catalog to!

<https://esa-earthcode.github.io/examples/>

Integrated platforms (selection)

PLATFORM

Euro Data Cube

One-stop-shop for browsing, analysis and processing of EO data, from source up to the final product. A combination of several services: harness the power of the data cube, access and analyse all the most important Earth Observation data in one application

ACCESS →

<https://eurodatacube.com/>

PLATFORM

Polar TEP

The Polar Thematic Exploitation Platform (Polar TEP) provides a complete working environment where users can access algorithms and data remotely, obtain computing resources and tools that they might not otherwise have, and avoid the need to download and manage large volumes of data.

ACCESS →

<https://polartep.hub.eox.at/>

PLATFORM

Deep ESDL

Virtual laboratory providing data, tools, and computational resources to efficiently implement comprehensive processing workflows for Earth System data

ACCESS →

<https://www.earthsystemdatalab.net/>

PLATFORM

CDSE OpenEO

Standardised interfaces for easy access and processing of Earth observation data. With its versatile tools, you can effortlessly create new workflows or integrate them into existing ones. Discover how to unleash the full potential of Earth observation data with minimal code and maximum efficiency

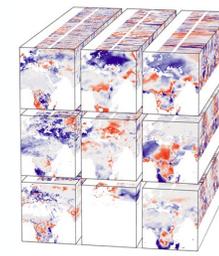
ACCESS →

<https://dataspace.copernicus.eu/analyse/openeo>

EURO DATA CUBE



polar
tep



EARTH
SYSTEM
DATA
LAB



Taking a closer look at a platform example

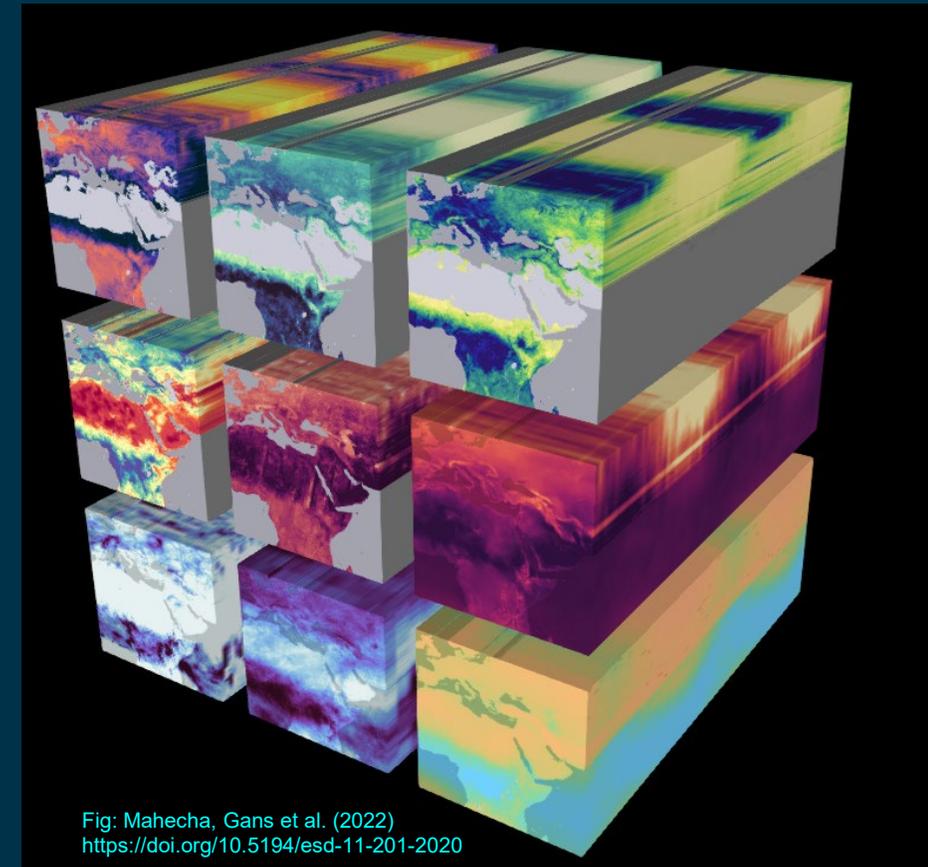


DeepESDL (Earth System Data Lab)

an AI-ready, Open Platform for research
and collaboration in Earth Sciences



EarthSystemDataLab.net

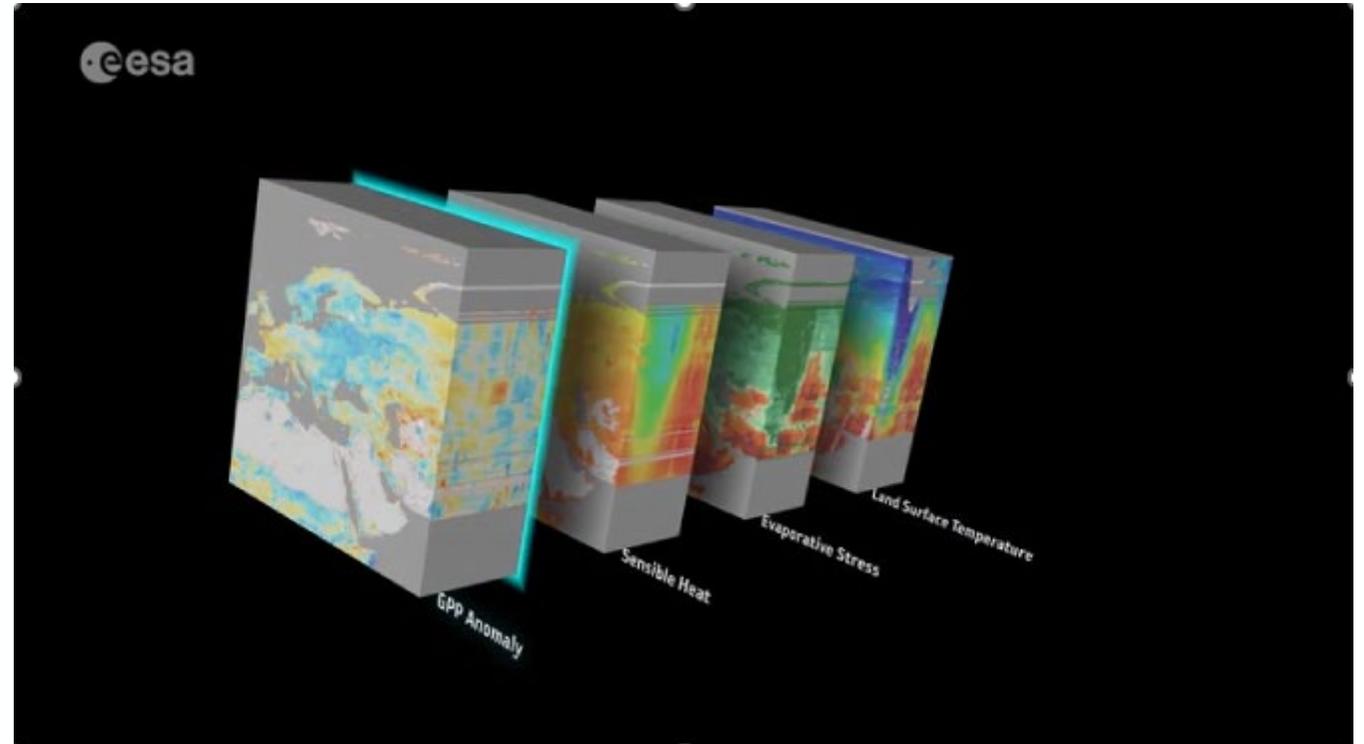


An open-source virtual Laboratory for Earth System Sciences with ...

- analysis-ready data cubes composed of multiple sources
- scientific computing facilities for exploratory analytics
- example workflows
- Machine learning tools

...to support Open Earth System Science:

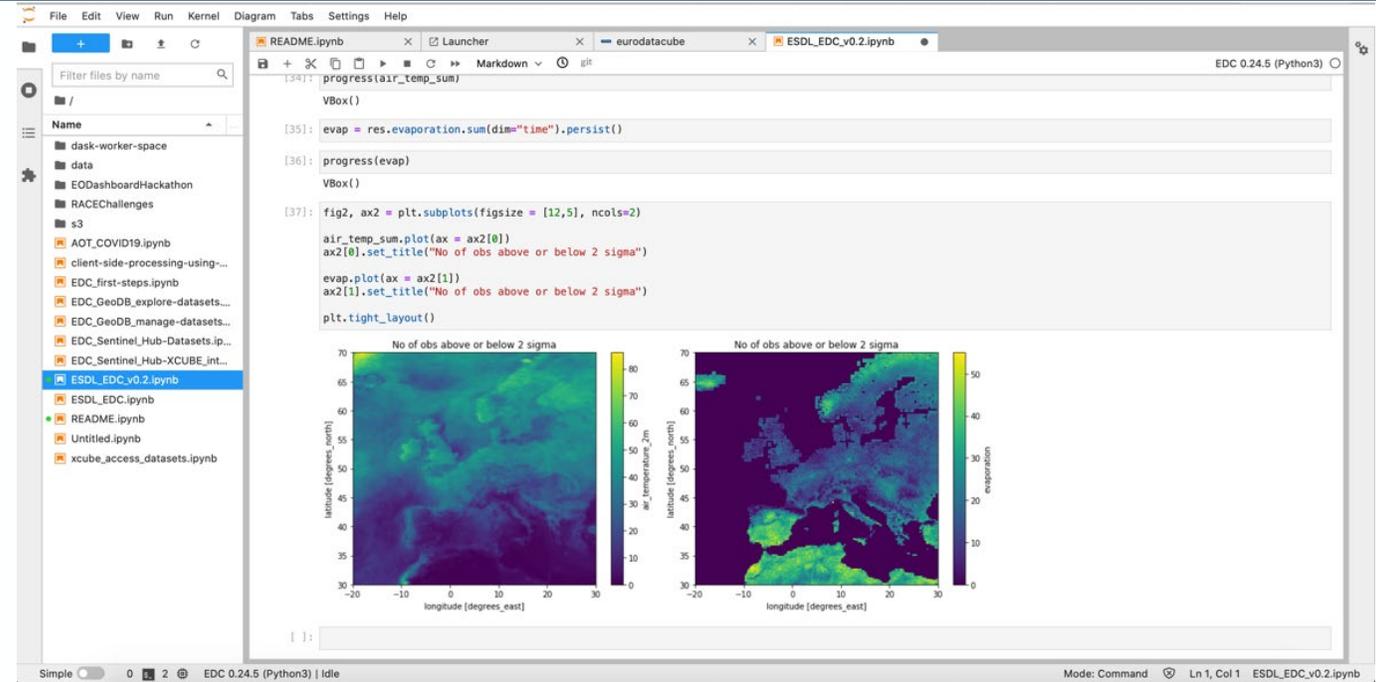
- Reproducible workflows
- Collaborative tools



- Python and Julia
- Jupyter Hub Profile for each scientific team
- Shared team resources, e.g. team s3 cloud storage
- Access to Machine Learning tools i.e. MLflow
- Each team member has own workspace
- Easy sharing of data or jupyter notebooks between team members
- Scalable compute resources

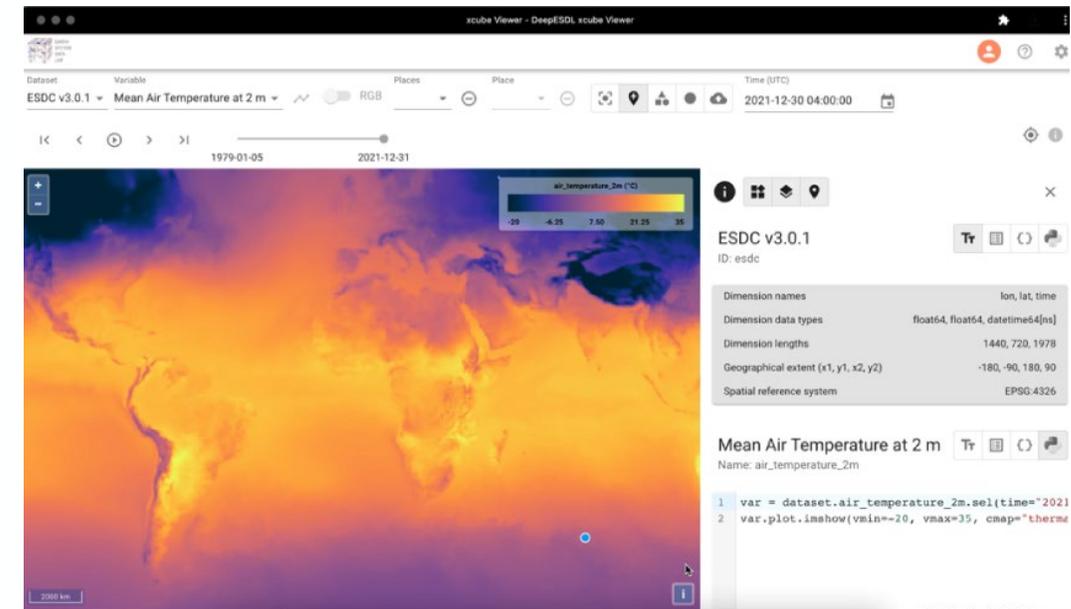
The screenshot shows the Earth System Data Lab website. At the top left is the Earth System Data Lab logo. To its right are two buttons: 'JupyterLab' and 'Viewer'. Below these is a navigation menu with links: Home, Data, Services, EarthCODE, Science Cases, Become a User, References, About, and User Guide. The main content area features a large heading 'Services' and three service cards. The first card is 'DeepESDL Hub' with an icon of a data cube and the text: 'Analyse, publish and collaborate in the DeepESDL Hub for reproducible research.' The second card is 'EarthCODE' with an icon of a globe and satellite and the text: 'Share reproducible workflows and data in EarthCODE.' The third card is 'Viewer' with an icon of a heatmap and the text: 'Publish and visualize data directly in a personalized Viewer.'

- Python and Julia
- Jupyter Hub Profile for each scientific team
- Shared team resources, e.g. team s3 cloud storage
- Access to Machine Learning tools i.e. MLflow
- Each team member has own workspace
- Easy sharing of data or jupyter notebooks between team members
- Scalable compute resources



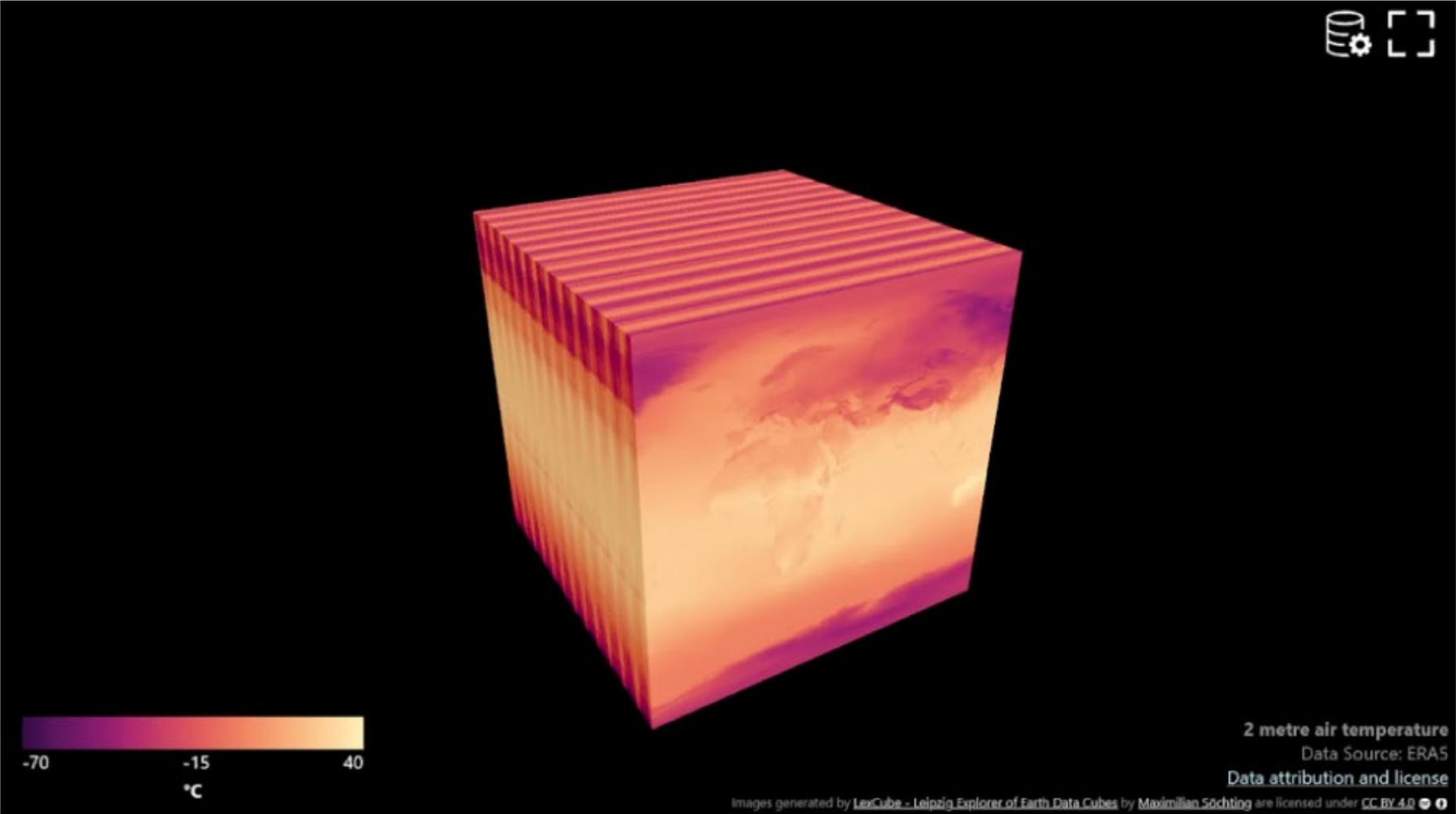
FAIR Data → visualisation-ready data

	Copernicus Climate Data Store (CDS)	<code>new_data_store("cds")</code>	Example Notebook	Github
	Copernicus Marine Service (CMEMS)	<code>new_data_store("cmems")</code>	Example Notebook	Github
	Copernicus Land Monitoring Service (CLMS)	<code>new_data_store("clms")</code>	Example Notebook (external link) Example Notebook (external link)	Github
	DeepESDL Public Data Cubes	<code>new_data_store("s3", root="deep-esdl-public")</code>	Example Notebook	Github
	EOPF Sample Service	<code>new_data_store("eopf-zarr")</code>	Example Notebook (external link)	Github
	ESA Climate Data Centre (CCI)	<code>new_data_store("cciudp")</code> <code>new_data_store("ccizarr")</code>	Example Notebook	Github
	ESA SMOS	<code>new_data_store("smos")</code>	Example Notebook (external link)	Github
	Global Ecosystem Dynamics Investigation (GEDI)	<code>new_data_store("gedi")</code>	Example Notebook (external link)	Github
	Sentinel Hub	<code>new_data_store("sentinelhub")</code>	Example Notebook	Github
	SpatioTemporal Asset Catalog (STAC)	<code>new_data_store("stac")</code> <code>new_data_store("stac-cdse")</code> <code>new_data_store("stac-cdse-ardc")</code>	General Example Notebook, CDSE Example Notebook	Github
	Zenodo	<code>new_data_store("zenodo")</code>	Example Notebook	Github

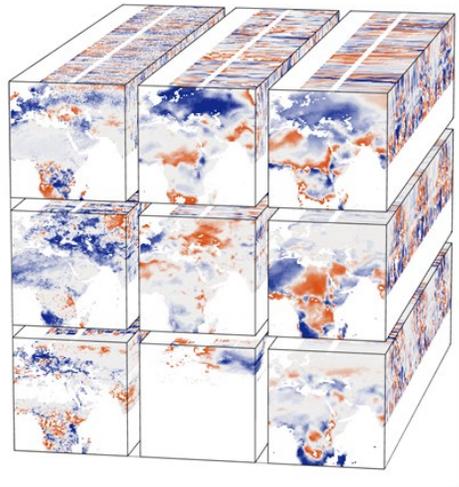


FAIR Data → visualisation-ready data

- Lexcube - 3D Viewer of datacubes



DeepESDL Publisher



EARTH
SYSTEM
DATA
LAB

<https://www.earthsystemdatalab.net/>



github.com/deepesdl/deep-code

README Code of conduct MIT license Security

deep-code

Unittest deep-code **passing** codecov **86%** code style **black** license **MIT**

`deep-code` is a lightweight python tool that comprises a command line interface(CLI) and Python API providing utilities that aid integration of DeepESDL datasets, experiments with EarthCODE.

The first release will focus on implementing the publish feature of DeepESDL experiments/workflow as OGC API record and Datasets as an OSC stac collection.

Setup

Install

`deep-code` will be available in PyPI for now and will be available in conda-forge in the near future. Till the stable release, developers/contributors can follow the below steps to install deep-code.

Installing from the repository for Developers/Contributors

To install deep-code directly from the git repository, clone the repository, and execute the steps below:

```
conda env create
conda activate deep-code
pip install -e .
```

This installs all the dependencies of `deep-code` into a fresh conda environment, and installs deep-code from the repository into the same environment

The screenshot shows the 'OPEN SCIENCE CATALOG [DEVELOPMENT VERSION]' website. The header includes the URL 'https://Opensciencedata.esa.int' and the ESA logo. A dark sidebar on the left contains navigation links: Home, Catalog, Metrics, Search, and API Access. The main content area features a large heading 'Welcome to the Open Science Catalog' and a paragraph explaining the catalog's purpose. Below this, there is a section titled 'What products can I find here?' followed by a paragraph about metadata and data availability. A row of buttons for 'THEMES', 'PROJECTS', 'VARIABLES', and 'PRODUCTS' is displayed. Further down, there are links for 'API Documentation', 'register here', and 'opensciencedata@esa.int'. At the bottom of the main content, there are 'METRICS' and 'SEARCH' buttons, and a row of three satellite imagery thumbnails. The footer contains copyright information '© 2023 by ESA', 'Terms and Conditions | Privacy Notice', and version information 'open-science-catalog v2.0.4 by EOX'.

OPEN SCIENCE CATALOG [DEVELOPMENT VERSION]

SeasFire Cube: A Global Dataset for Seasonal Fire Modeling in the Earth System

[Overview](#)

Description

The SeasFire Cube is a scientific datacube for seasonal fire forecasting around the globe. Apart from seasonal fire forecasting, which is the aim of the SeasFire project, the datacube can be used for several other tasks. For example, it can be used to model teleconnections and memory effects in the earth system. Additionally, it can be used to model emissions from wildfires and the evolution of wildfire regimes. It has been created in the context of the SeasFire project, which deals with Earth System Deep Learning for Seasonal Fire Forecasting and is funded by the European Space Agency (ESA) in the context of ESA Future EO-1 Science for Society Call. It contains 21 years of data (2001-2021) in an 8-days time resolution and 0.25 degrees grid resolution. It has a diverse range of seasonal fire drivers. It expands from atmospheric and climatological ones to vegetation variables, socioeconomic and the target variables related to wildfires such as burned areas, fire radiative power, and wildfire-related CO2 emissions

Item 1

SeasFireCube_v.0.3.zarr

1/1/2001, 12:00:00 AM UTC - 12/28/2021, 12:00:00 AM UTC

[LOG IN](#)

Powered by STAC Browser 3.0.1

[SUGGEST CHANGES](#)

© 2023 by ESA [Terms and Conditions](#) | [Privacy Notice](#) open-science-catalog v2.0.4 by

Examples | SeasFIRE Cube



Available on DeepESDL Viewer: <https://viewer.earthsystemdatalab.net/>

Leaf Area Index (m²/m²)
SeasFire Cube: A Global Dataset for Seasonal Fire Modeling in the Earth System

0 1.75 3.5 5.25 7

SeasFire Cube: A Global Dataset for Seasonal Fire Modeling in the Earth System
ID: seasfire

crs EPSG:4326

The SeasFire Cube is a scientific database for seasonal fire forecasting around the globe. It has been created for the SeasFire project, that addresses 'Earth System Deep Learning for Seasonal Fire Forecasting' and is funded by the European Space Agency (ESA) in the context of ESA Future EO-1 Science for Society Call. It contains almost 20 years of data (2001-2021) in an 8-days time resolution and 0.25 degrees grid resolution. It has a diverse range of seasonal fire drivers. It expands from atmospheric and climatological ones to vegetation variables, socioeconomic and the target variables related to wildfires such as burned areas, fire radiative power, and wildfire-related CO₂ emissions.

Leaf Area Index
Name: lai

The MCD15A2H Version 6 Moderate Resolution Imaging Spectroradiometer (MODIS) Level 4, Combined Fraction of Photosynthetically Active Radiation (FPAR), and Leaf Area Index (LAI) product is an 8-day composite dataset with 500 meter pixel size. The algorithm chooses the best pixel available from all the acquisitions of both MODIS sensors located on NASA's Terra and Aqua satellites from within the 8-day period. LAI is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and as one-half the total needle surface area per unit ground area in coniferous canopies. FPAR is defined as the fraction of incident photosynthetically active radiation (400-700 nm) absorbed by the green elements of a vegetation canopy.

Units	m ² /m ²
Data type	float32
Dimension names	time, latitude, longitude
Dimension lengths	966, 720, 1440
Time chunk size	966

	Array	Chunk
Bytes	3.73 GiB	3.96 MiB
Shape	(966, 720, 1440)	(1, 720, 1440)

FAIR and visualization-ready data



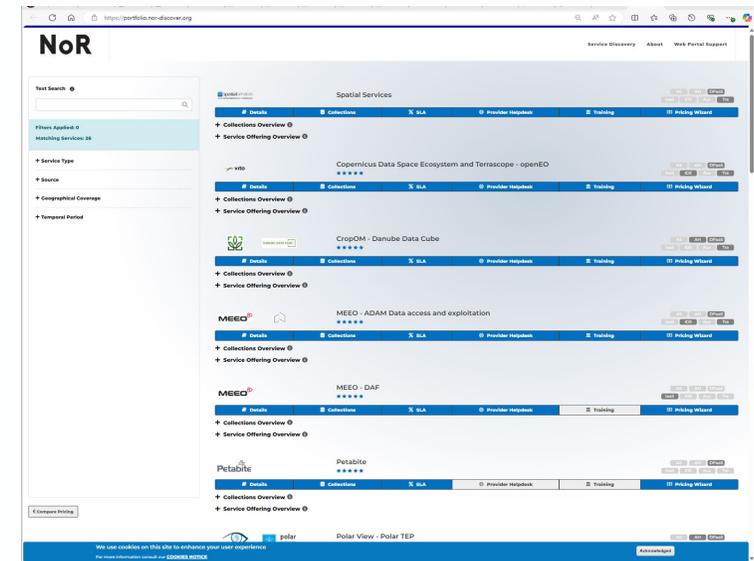
- ❑ Projects onboarded to EarthCODE:
 - ❑ Have their NoR sponsorship requests managed by EarthCODE
 - ❑ Have a guaranteed* allocation of NoR budget



NoR = Network of Resources

A portfolio of platform services for data access and computing resources for which ESA offers sponsored use (only for non-commercial use)

- NoR Website: <https://nor-discover.org/>
- NoR Portfolio: <https://portfolio.nor-discover.org>

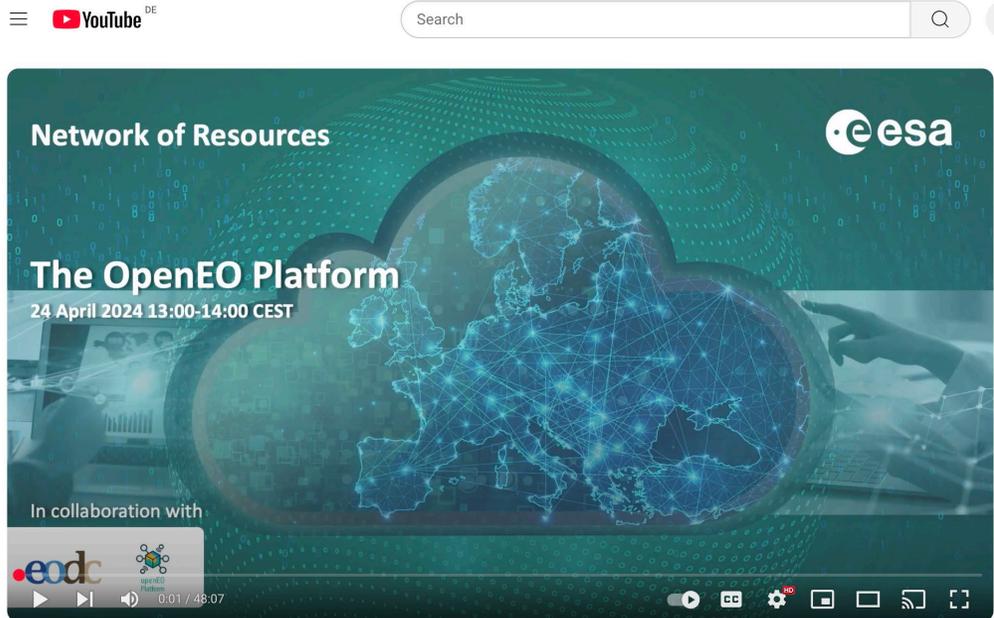


Approval of your proposal is usually less than 1 week!

* Priority is for integrated platforms and within available budget

Getting started with Platforms

- ESA sponsors training on a large number of platforms available on the Network of Resources portfolio
 - Open access webinars: <https://nor-discover.org/en/outreach/#webinars>
 - On-demand training, on site and online for beginner to advanced



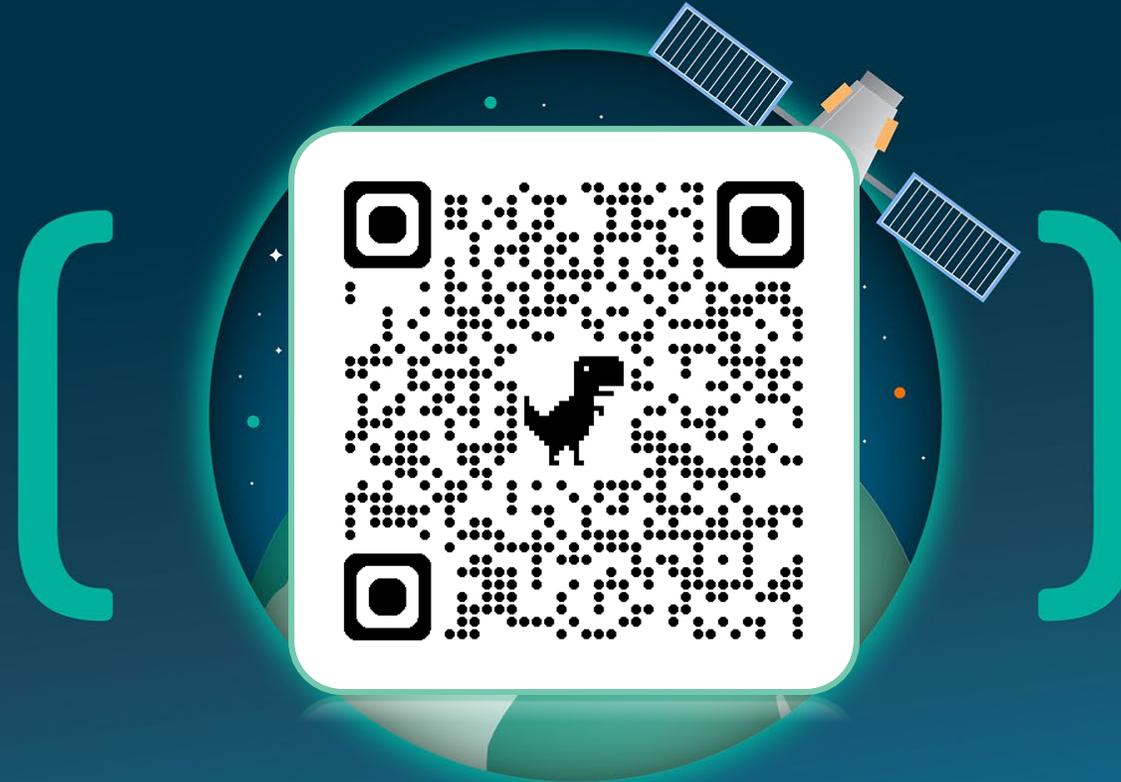
The OpenEO Platform

A screenshot of the NoR (Network of Resources) web portal. The page title is "Copernicus Data Space Ecosystem and Terrascope - openEO" with a 5-star rating. The search bar contains the word "training". The left sidebar shows "Filters Applied: 0" and "Matching Services: 20". The main content area lists "Training Services" with two offerings: "Online dedicated training to user" (2 hours) and "Online webinar" (1 hour). Both offerings list "Service Types Covered" as AH, DPaaS, and IDE. A red box highlights the "Training Services" section.

- Explore EO-based indicators from ESA, NASA and JAXA
- Use the datasets in Platform Environments to generate insights
- Create interactive narratives with Dashboard elements
- Open for community contributions



Sign up for ESA's Open Science Newsletter



Monthly updates on ESA tenders, new open datasets, upcoming events, and other opportunities

