EOF On-boarding and Federation Scenarios

4-5 March 2025
Lisbon

Agenda

- CDSE – Data On-boarding scenarios
- OpenEO – Updates
- DestinE – On-boarding scenarios
- On-boarding and Federation - Conclusions

Copernicus Data Space Ecosystem

Proposed EO Data Onboarding scenarios

Key CDSE features from data provider perspective

- Expanded User reach & visibility
  - The biggest user base among all European earth observation platforms – almost **300k registered** and more than **50k active users** every month
- Seamless discoverability & accessibility
  - **Highly available data catalogue** covering Copernicus native and contributing missions
  - **Scalable data access layer** handling daily millions of catalogue requests and hundreds of millions of access requests
  - De facto **reference implementation of API** standards
- Integrated synchronous and asynchronous processing
- Ready to use visualization services with **well-established UIs**
- **Algorithm prototyping** and large-scale **execution** environment

Onboarding scenarios

Which services will support my data?

What are the requirements for my data and metadata?

Where should my data physically reside?

How are access rules enforced - public, private, or custom?

What kind of staging environment do I need to integrate with CDSE?

How does my data appear in CDSE's data catalogue and services?

Do I need to integrate with CDSE Identity and Access Management (IAM)?

What are the key considerations?

Key concepts

Base APIs

- STAC / OData – querying the metadata (search and discovery)
- S3 – accessing the data directly (object download and upload)
- ZIPPER – accessing the packaged data (file download)

Streamlined access and processing APIs

- Sentinel Hub – OGC services, OTF/batch processing, visualization, statistical analysis
- openEO – workflow execution, data cubes,

Supporting APIs

- Traceability – data lifecycle registry

EO Data Provider

- a third party willing to onboard their data into CDSE data offering

Data Policy

- Public data can be accessed by any user
- Private data requires specific access control policies

CDSE IAM

- provides user authentication / authorization and data access control

Data location

- external to CDSE
- CDSE EOData
- CREODIAS Object Storage
- CF Object Storage

Scenarios

“Quick-wins”

I.A: CDSE references EO Data Provider data access services

I.B: CDSE Indexes new data from EO Data Provider (Public Data)

I.C: CDSE Indexes new data (Private Access controlled by Provider)

“Full integration”

II: Standard CDSE Collection Onboarding

“Flexible”

III.A: Authorized EO Data Provider delivers Public Data via CF Object Storage

III.B: Authorized EO Data Provider delivers Data through EOData

Scenario I.A

CDSE references EO Data Provider data access services

PORTAL

STAC

S3

ZIPPER

Sentinel Hub

openEO

Type of Data

Public

Identity and Access

Provider

Data Location

External storage (object or other)

EO Data Provider responsibility

- Maintain the data in their own external storage environment
- Make the data publicly accessible
- Maintain the data search / discovery / download services
- Keep the metadata consistent and current and ensure the data URLs remain valid
- No advanced identity federation is needed because the data is fully public

Usability and performance

- no integration with the CDSE services
- no guarantee about data throughput or performance
- no CDSE quotas or SLAs apply
- data curation remains with the provider

Why?

- Simplest way to showcase new data in CDSE without any integration effort
- Easy to set up and maintain if the data is already public and the data provider has existing search / discovery / download services available

Scenario I.B

CDSE Indexes new data from EO Data Provider (Public Data)

PORTAL

STAC

S3

ZIPPER

Sentinel Hub

openEO

Type of Data

Public

Identity and Access

CDSE

Data Location

External / CREODIAS storage

EO Data Provider responsibility

- Deliver metadata definition / schema for each collection
- Generate metadata compliant with metadata definition, STAC standard and CDSE requirements
- Maintain the data in their own external storage environment
- Keep the STAC metadata consistent and up to date and ensure the data access URLs remain valid
- No advanced identity federation is needed because the data is fully public

Usability and performance

- limited integration with the CDSE services
- no guarantee about data throughput or performance
- no CDSE quotas or SLAs apply
- data curation remains with the provider

Why?

- A “quick win” for showcasing new data in CDSE without investing in integration or new access control measures.
- Easy to set up and maintain if the data is already public.
- Minimal overhead for the data provider - just generate STAC metadata and let CDSE index external links

Scenario I.C

CDSE Indexes new data from EO Data Provider (Private Data)

PORTAL

STAC

S3

ZIPPER

Sentinel Hub

openEO

Type of Data

Public

Identity and Access

CDSE / Provider

Data Location

External / CREODIAS storage

EO Data Provider responsibility

- Deliver metadata definition / schema for each collection and generate metadata compliant with STAC standard, delivered schema and CDSE requirements
- Maintain the data in their own external storage environment, and implement and maintain the private access logic
- Keep the STAC metadata consistent and up to date and ensure the data access URLs remain valid
- Provide access instructions for CDSE users, and manage or federate identities (potentially bridging to CDSE IAM if desired)

Usability and performance

- limited integration with the CDSE services
- no guarantee about data throughput or performance
- no CDSE quotas or SLAs apply
- data curation remains with the provider
- complicated and confusing access to data for CDSE

Why?

- Allows data providers to preserve control and benefit from CDSE's indexing and discoverability
- Potential stepping stone if full ingestion is not feasible
- Limited overhead for the data provider – access provided using provider's own services

Scenario II

Standard CDSE Collection Onboarding

PORTAL

STAC

S3

ZIPPER

Sentinel Hub

openEO

Type of Data

Public / Private

Identity and Access

CDSE

Data Location

CDSE EOData

EO Data Provider responsibility

- Deliver metadata definition / schema for each collection, generate metadata compliant with STAC standard, delivered schema and CDSE requirements and keep it consistent, up to date and with clear version information
- Prepare data in a recommended format and make it available for retrieval by CDSE (optional: Set up and maintain data access API)
- Provide clear version information in the metadata, inform CDSE about corrupted / obsolete products, and decide on the retention policy
- Ensure continuity of versioned products

Usability and performance

- Requires integration with the CDSE EOData involving definition of storage rules, access and retention policies, IAM roles, ingestion pipelines
- Collection contents and data volume is moderated by ESA/Copernicus

Why?

- Data is officially integrated into the CDSE ecosystem (Copernicus Browser, documentation, support)
- EO data providers can rely on CDSE's uptime guarantees, high performance and data governance
- Data lifecycle management is needed

Scenario III.A (not available before Q3 2025)

Authorized Provider (Public Data in CF Object Storage)

PORTAL

STAC

S3

ZIPPER

Sentinel Hub

openEO

Type of Data

Public

Identity and Access

CDSE

Data Location

CF object storage

EO Data Provider responsibility

- Deliver metadata definition / schema for each collection
- Generate metadata compliant with metadata definition, STAC standard and CDSE requirements
- Keep the STAC metadata consistent and up to date and ensure the data access URLs remain valid
- Manage the data in the public bucket (e.g., ensure availability)

Usability and performance

- Throughput and processing performance may be lower than for data stored in EOData
- CDSE SLAs do not apply
- The provider must handle any desired versioning or updates

Why?

- Allows a provider to “self-manage” the data while still benefiting from CDSE base services
- Enables access using Sentinel Hub and openEO
- Public availability simplifies user access and encourages broad use

Scenario III.B (in planning)

Authorized Provider (data in CF Storage via CDSE EOData)

PORTAL

STAC

S3

ZIPPER

Sentinel Hub

openEO

Type of Data

Public / Private

Identity and Access

CDSE

Data Location

CF object storage

EO Data Provider responsibility

- Manage the bucket's privacy settings, data integrity, updates, etc.
- Manage the data in the public bucket (e.g., ensure availability)
- Generate and maintain correct STAC metadata
- Keep the STAC metadata consistent and up to date and ensure the data access location remains valid
- Define data access rules if the data is private

Usability and performance

- Provider must ensure correct access control
- Throughput and processing performance may be lower than for data stored in EOData. CDSE SLAs do not apply
- The provider must handle any desired versioning or updates
- In case of the CDSE disaster recovery event, the data will not be accessible

Why?

- Provider benefits from the CDSE access control and traffic limitation mechanism, a balance between fully external and fully ingested data
- Potentially better performance than hosting data on an entirely external provider's platform due to proximity of the CDSE backend components

Conclusion

The choice of scenario depends on multiple factors:

- **Urgency**: When will the data be available?
- **Location of Data**: Do you want (or need) to store data externally, or fully ingest it into CDSE?
- **Access Control**: Is your data public, private, or do you have custom rules for who can see what?
- **Tools Integration**: Do you want to leverage Sentinel Hub, openEO, or a simple STAC-based indexing?
- **Lifecycle and Versioning**: How often do you update or replace your datasets?
- **Performance Requirements**: Do you need guaranteed throughput and low-latency processing, or is external hosting sufficient?
- **Complexity and Effort**: Are you ready to adapt your data to strict requirements?

Copernicus Data Space Ecosystem

Thank you!

acesarz@cloudferro.com

dataspace.copernicus.eu

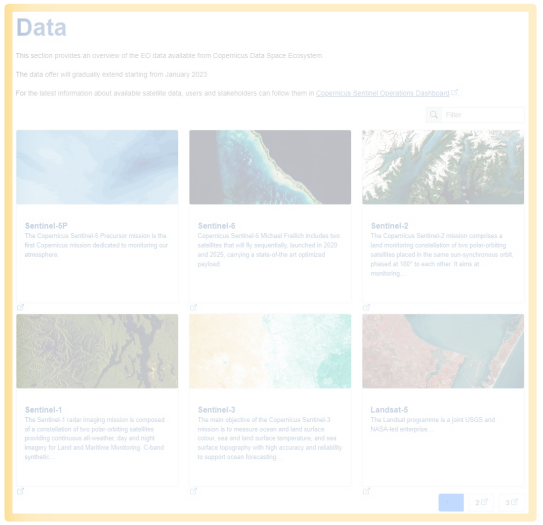
Copernicus Data Space Ecosystem

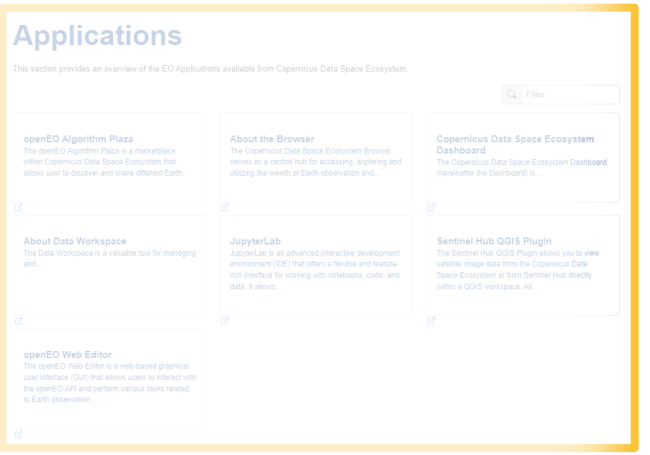
OpenEO updates 2024 - 2025

dataspace.copernicus.eu

PROGRAMME OF THE EUROPEAN UNION

Data Access and Processing API: openEO





dataspace.copernicus.eu

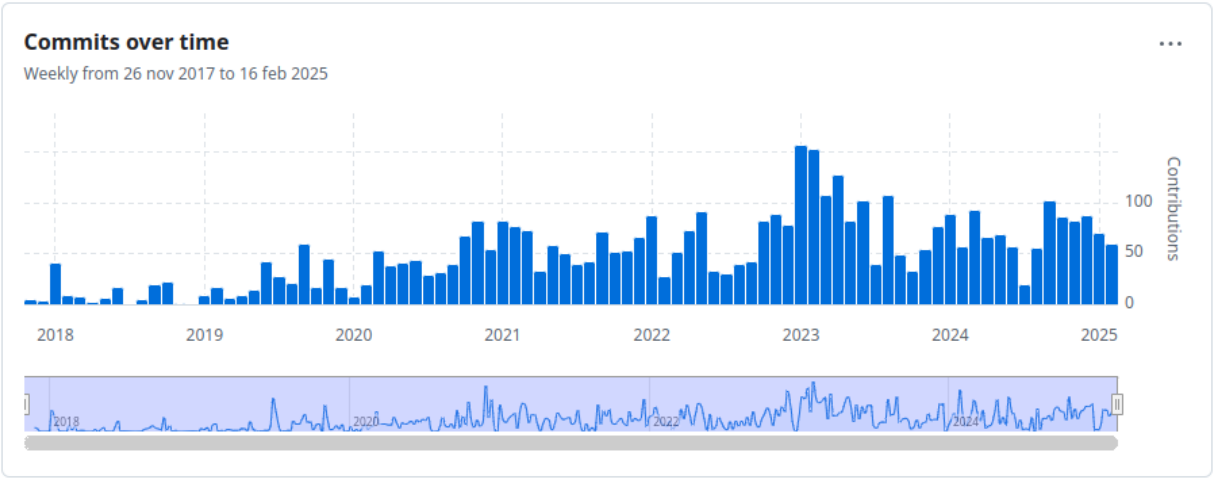
PROGRAMME OF THE EUROPEAN UNION

With openEO

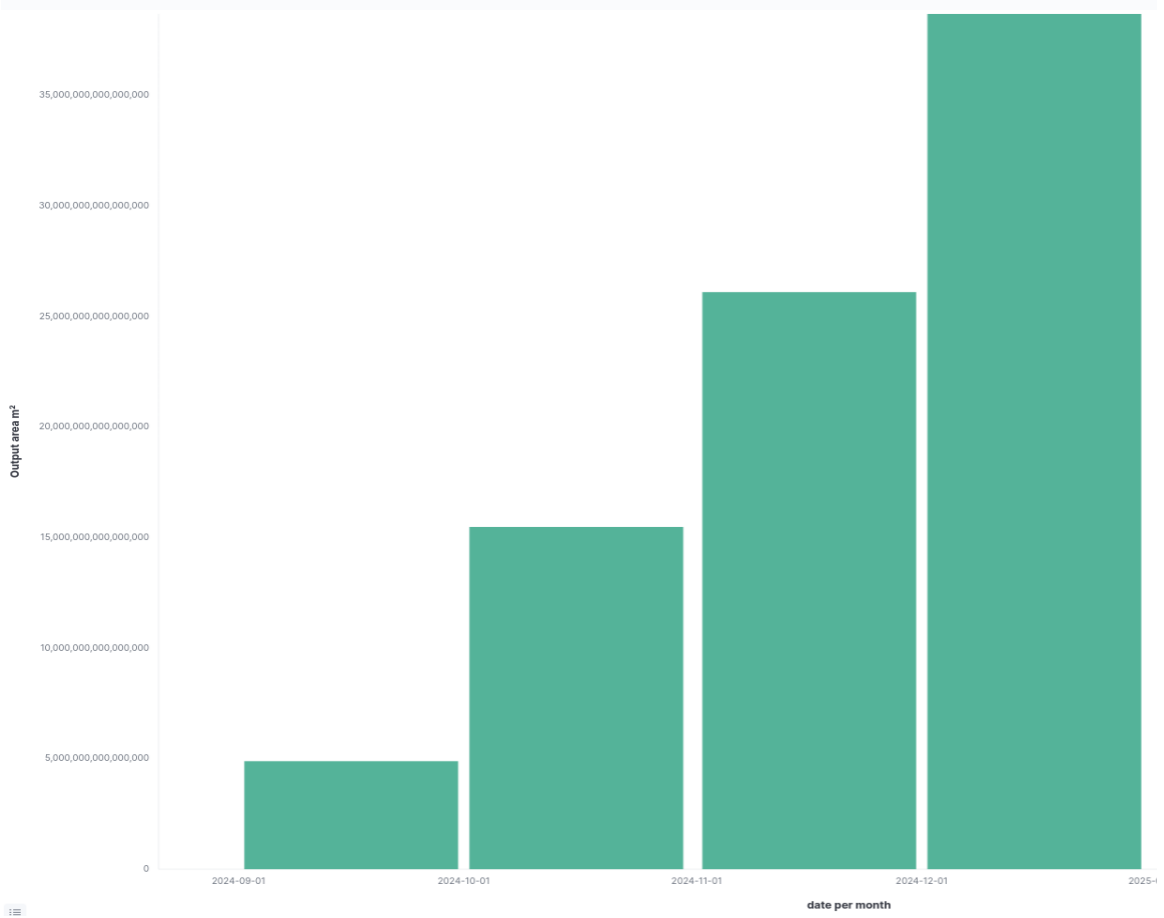
- Simple data access & processing for multiple Earth observation datasets
- Scalable and efficient processing capabilities
- A standardized system that works across different platforms
- Independence from underlying technologies and software libraries
- Supporting principles of FAIR (Findable, Accessible, Interoperable, and Reusable) data and Open Science, e.g. transparent workflows

openEO: An open-source ecosystem

- Active community
- 60+ contributors to the Github Repositories
- Governed by a Project Steering Committee (https://openeo.org/psc.html)



Large scale processing ramping up in 2024



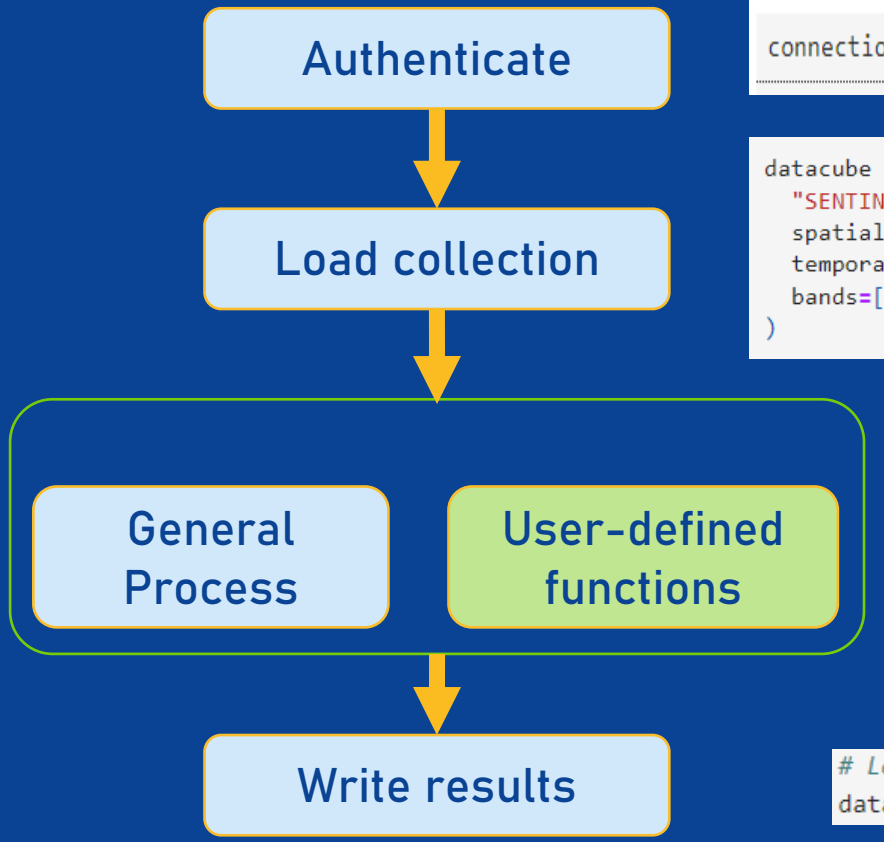
An equivalent of 4 Million Sentinel-2 tile equivalents was delivered in December 2024!

JupyterLab | Copernicus Data Space Ecosystem

openEO Workflow

Built-in OpenEO processes

User Defined Functions (UDF)




```
connection = openeo.connect("openeo.dataspace.copernicus.eu").authenticate_oidc()
```

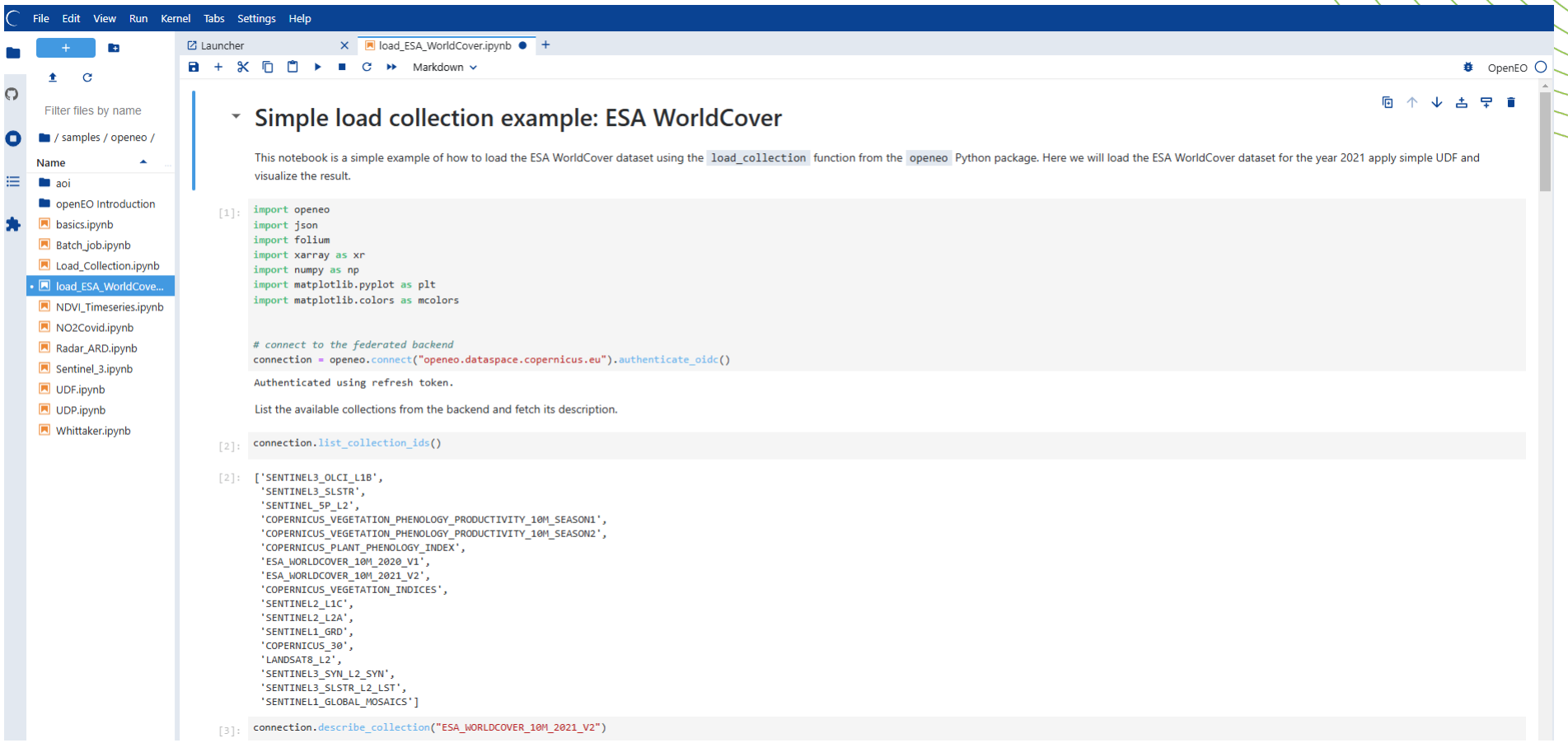
```
datacube = connection.load_collection(
  "SENTINEL1_GRD",
  spatial_extent={"west": 16.06, "south": 48.06, "east": 16.65, "north": 48.35},
  temporal_extent=["2017-03-01", "2017-04-01"],
  bands=["VV", "VH"]
)
```

https://processes.openeo.org/
https://open-eo.github.io/openeo-python-client/udf.html

```
datacube = datacube.min_time()
```

```
# Let's download the data
datacube.execute_batch(title="Simple execution", outputfile="Sentinel1.nc")
```

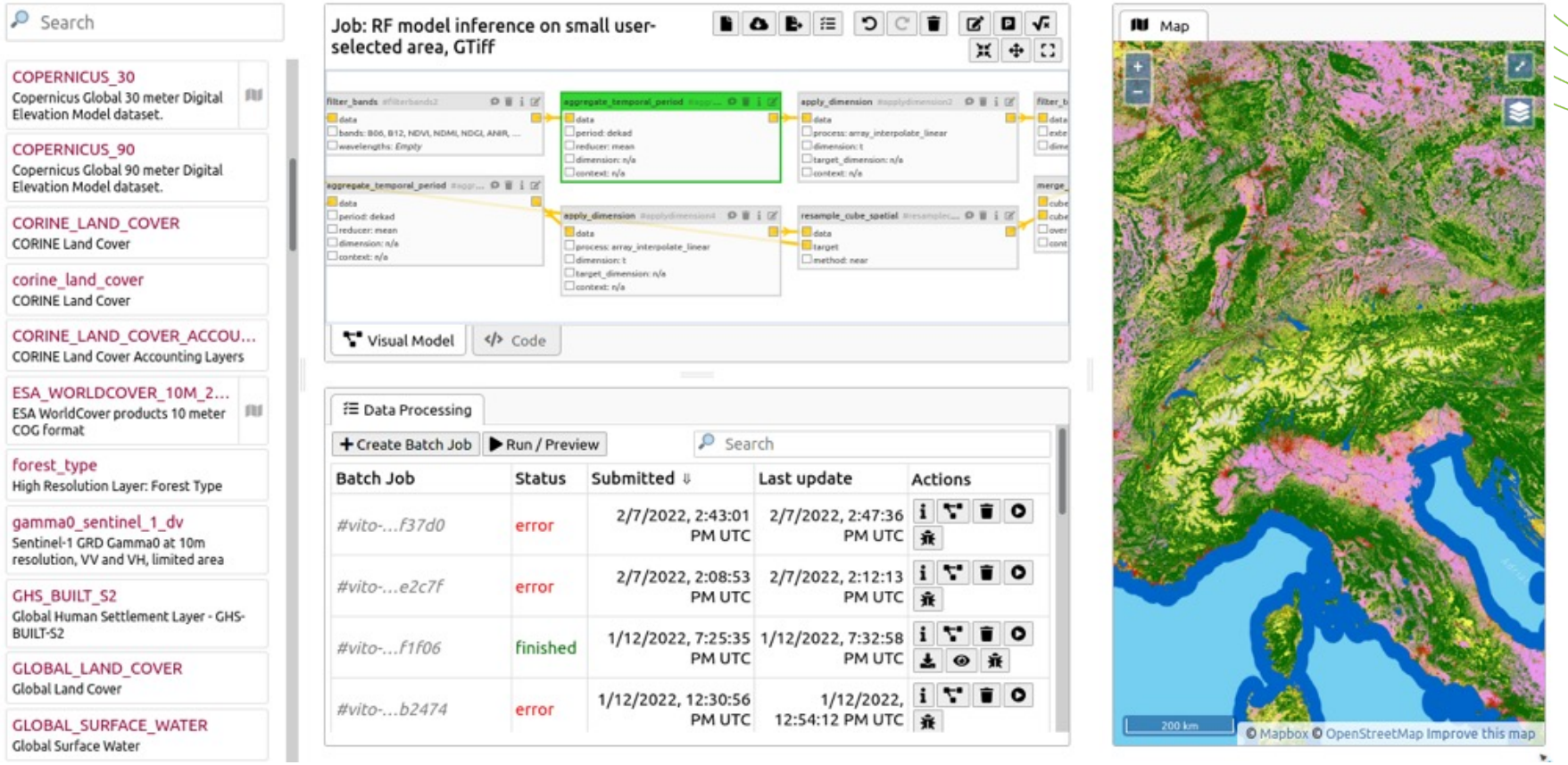
openEO Capabilities



- For interactive prototyping, programming and visualization
- Most convenient way for Python programmers to interact with the openEO Platform.

openEO Web Editor

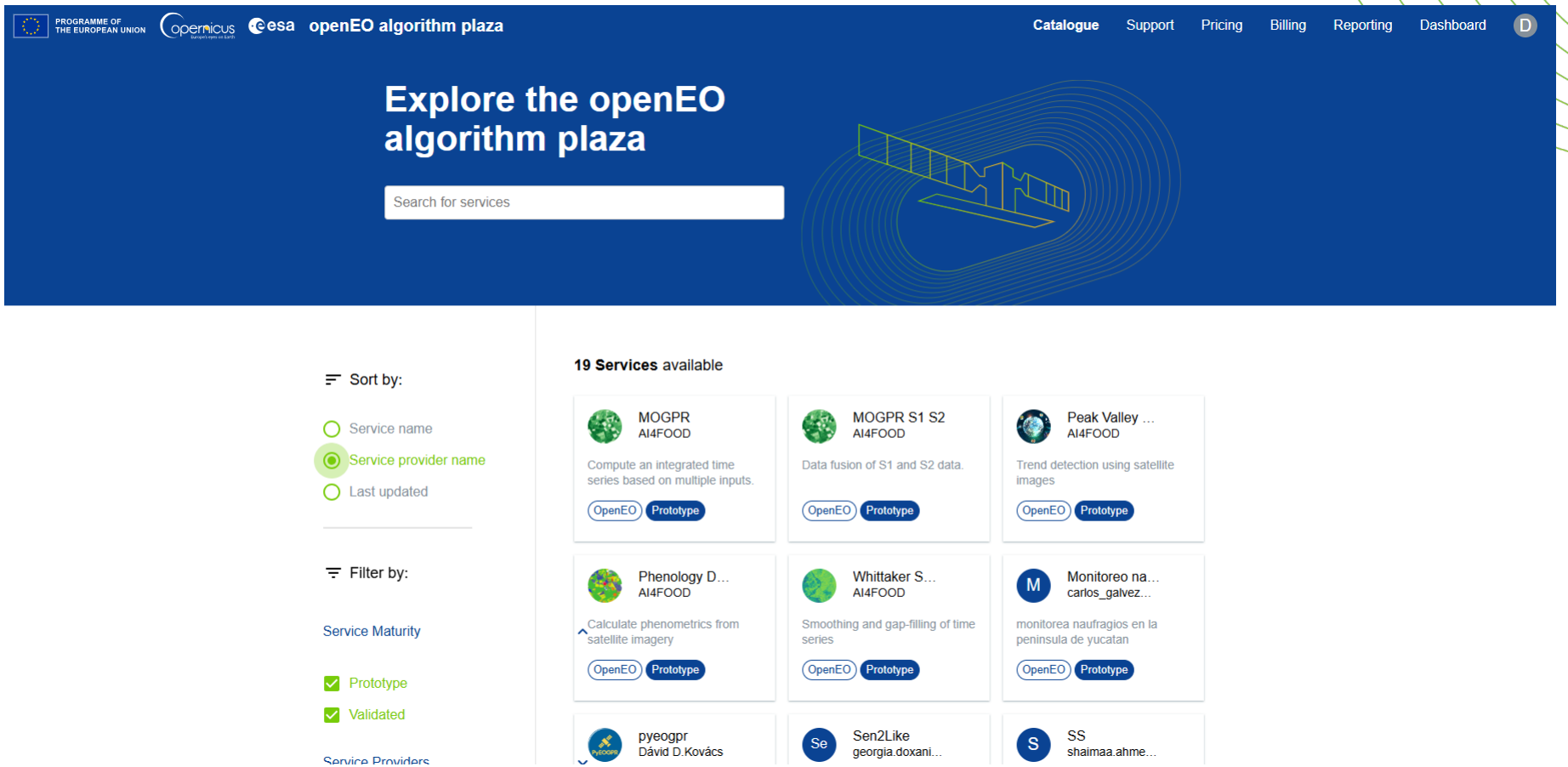
openEO Capabilities



- An interactive and visual user interface for a block-based workflow editor.
- Get an overview of available data sets and processes or monitor the status of their processing workflows.

openEO Capabilities

openEO Algorithm Plaza



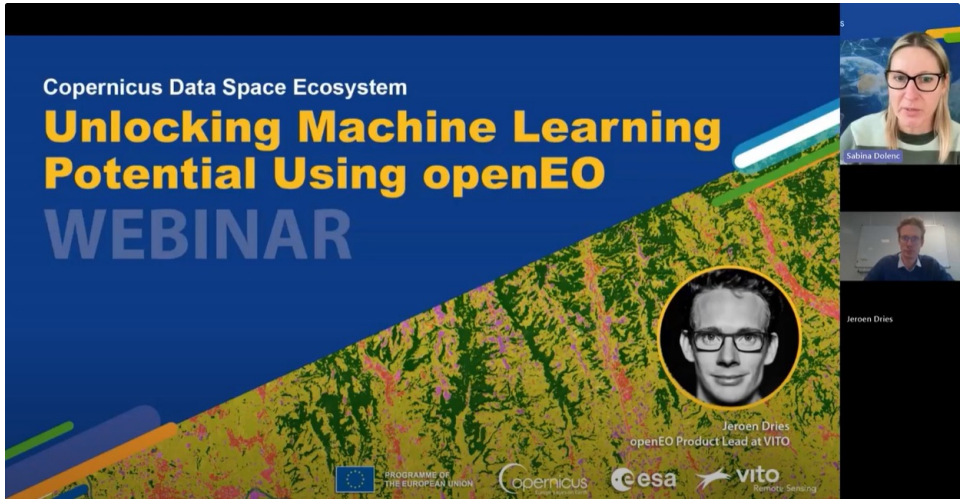
- Share user-defined processes
- For users looking for a simple workflow without necessarily knowing the details of creating workflows in openEO.

Road map 2025

What’s new since 2024?

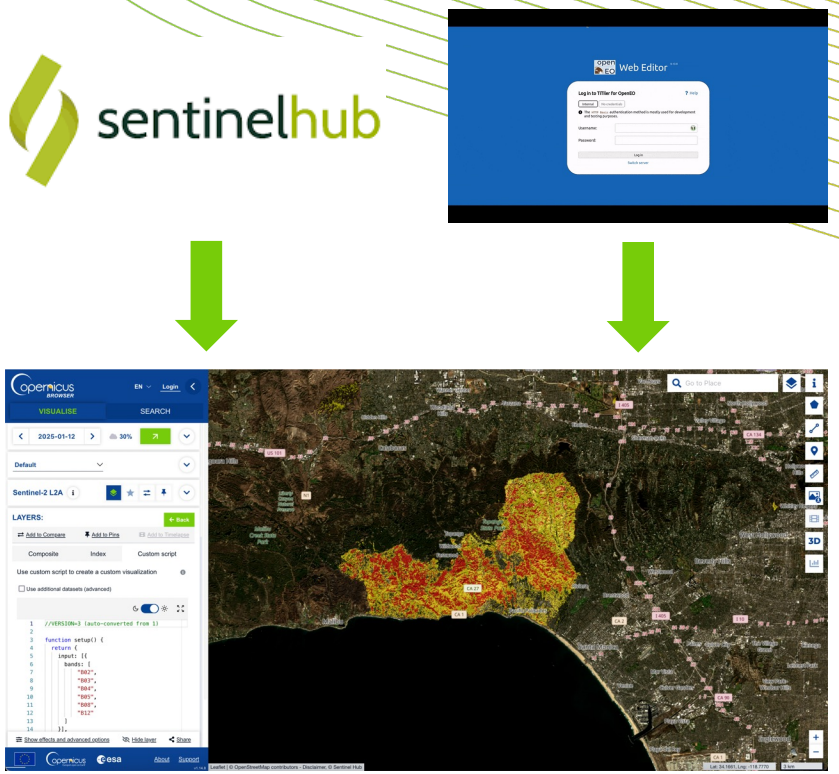
- Handling of multiple jobs
- Load external STAC for cross back-end processing
- openEO federation governance
- New machine learning examples (Random forest)
- R openEO notebooks
- Multiple performance improvements
- Increase of free tier quota

- Supported by (more to come):
  - Webinars
  - Hackathons
  - Use cases
  - Notebooks



openEO integration into Copernicus Browser

- Rationale
  - Enhance reusability of EO algorithms
  - Path from interactive prototyping to large-scale processing
  - Interface consistency
- Road map & timing
  - Operational Sentinel Hub openEO back-end (Q2 2025)
  - Open-source sync openEO back-end prototype (Q2 2025)
- Benefits
  - Improved synchronous process graphs execution
  - (hopefully) more shared code-base (call to action)



More info…

Various

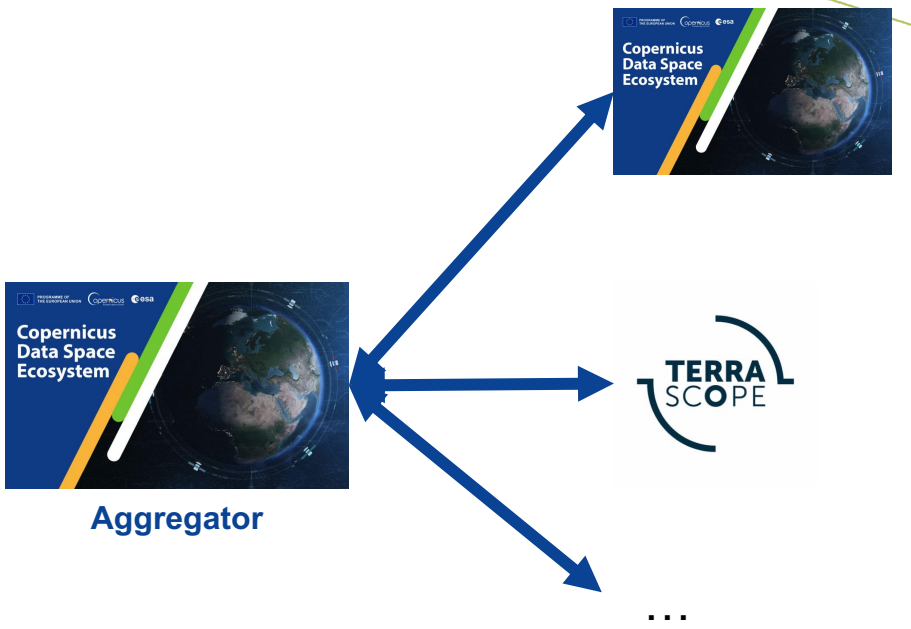
- Support for Docker-based processes via CWL/OGC application package (Q1)
- Export to STAC API (Q1)
- Export of data to object storage (Q1)
- Investment in reuseable functions (2025 – 2026)
- RTC gamma0 (Q4)
- Extension towards using EO embeddings (Q3-4)

Status federation

Update 2025

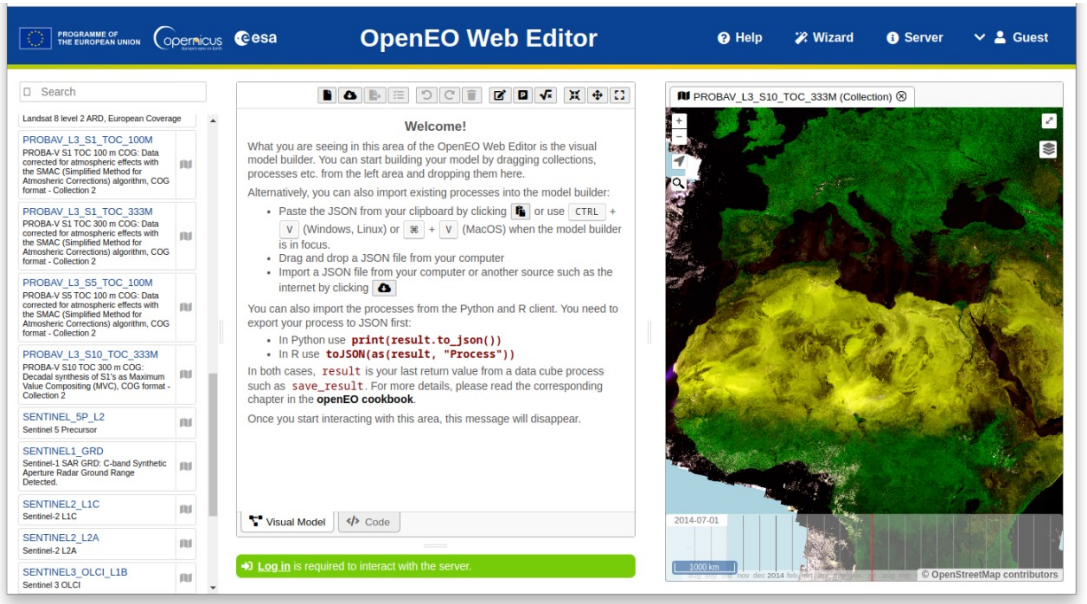
openEO federation concept

- An approach to realize the 'Ecosystem'
- Based on strong standardization with openEO and STAC
- Federates data access and processing capacity



Current Status

- Endpoint online
  https://openeofed.dataspace.copernicus.eu/



Copernicus Data Space Ecosystem openEO Aggregator

Federated openEO Aggregator instance, provided by Copernicus Data Space Ecosystem.

Supported Functionalities (6/8)

- ✔ Basic functionality
- ✔ Synchronous processing
- ✔ Batch processing
- ☐ Estimate processing costs
- ✔ Additional web services
- ☐ File storage
- ✔ User-defined processes
- ✔ User-defined functions (UDF)

Billing

Plans

No plans available.

Currency: credits

Federation

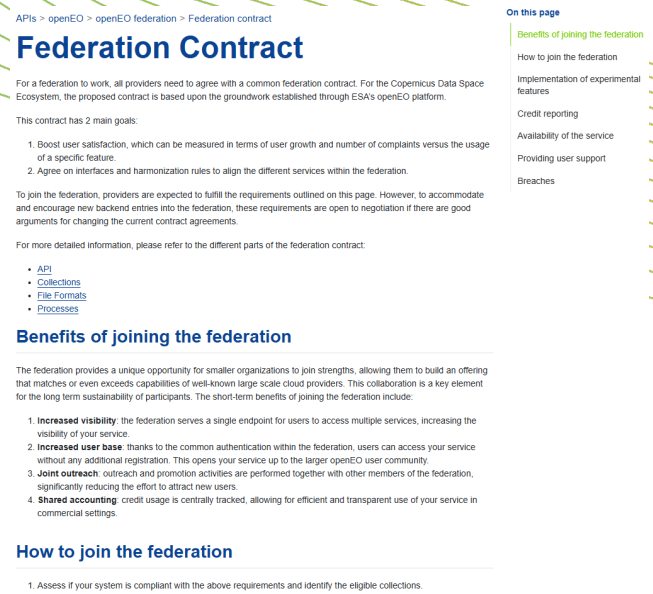
This service is a federation of multiple services, which are all listed below:

- **cdse** ONLINE
  URL: https://openeo.dataspace.copernicus.eu/openeo/1.1/
- **vito** ONLINE
  URL: https://openeo.vito.be/openeo/1.1/

TERRA SCOPE

Call for backends as part of the Ecosystem

- Governance & backend guidelines available
- Backend providers may perform self-review, and apply to join
- Getting started guides available for service providers
- Looking for openEO backends with collections that are
  - Complementary to CDSE offering
  - No experimental collections
  - Preference for full archives or at least large scale (e.g. continental)



APIs > openEO > openEO federation > Federation contract

Federation Contract

For a federation to work, all providers need to agree with a common federation contract. For the Copernicus Data Space Ecosystem, the proposed contract is based upon the groundwork established through ESA's openEO platform.

This contract has 2 main goals:

1. Boost user satisfaction, which can be measured in terms of user growth and number of complaints versus the usage of a specific feature.
2. Agree on interfaces and harmonization rules to align the different services within the federation.

To join the federation, providers are expected to fulfill the requirements outlined on this page. However, to accommodate and encourage new backend entries into the federation, these requirements are open to negotiation if there are good arguments for changing the current contract agreements.

For more detailed information, please refer to the different parts of the federation contract:

- API
- Collections
- File Formats
- Processes

Benefits of joining the federation

The federation provides a unique opportunity for smaller organizations to join strengths, allowing them to build an offering that matches or even exceeds capabilities of well-known large scale cloud providers. This collaboration is a key element for the long term sustainability of participants. The short-term benefits of joining the federation include:

1. **Increased visibility**: the federation serves a single endpoint for users to access multiple services, increasing the visibility of your service.
2. **Increased user base**: thanks to the common authentication within the federation, users can access your service without any additional registration. This opens your service up to the larger openEO user community.
3. **Joint outreach**: outreach and promotion activities are performed together with other members of the federation, significantly reducing the effort to attract new users.
4. **Shared accounting**: credit usage is centrally tracked, allowing for efficient and transparent use of your service in commercial settings.

How to join the federation

1. Assess if your system is compliant with the above requirements and identify the eligible collections.

On this page

Benefits of joining the federation
How to join the federation
Implementation of experimental features
Credit reporting
Availability of the service
Providing user support
Breaches

Known implementations: WASDI/EO4EU/IBM CH/DLR/EODC/CNES/EUMETSAT

Why join openEO federation?

- Official part of the Copernicus Data Space Ecosystem as member
- Increased visibility of your datasets and processing capacity to a larger userbase
- Attract more projects towards your backend
- Enjoy the shared outreach and promotion within CDSE & other federation members
- Take advantage of the shared accounting and expose your service in a commercial setting
- Joined EU effort against non-EU big tech companies

Flagship projects

Building on top of openEO

Continental & global scale projects powered by openEO

ESA WorldCereal

ESA World Agrocommodities

JRC Copernicus Land Cover Forest Monitoring (LCFM)

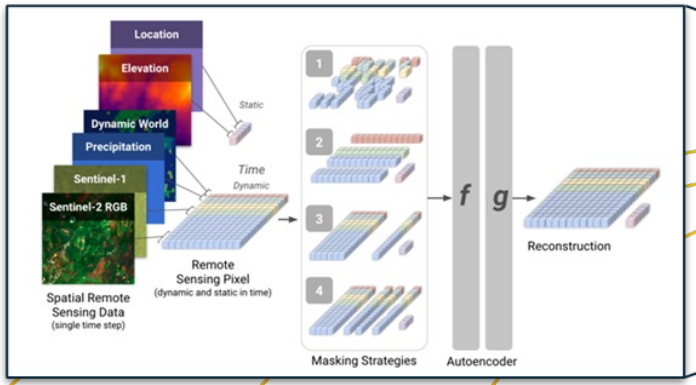
EC
Grasslandwatch

ESA World Ecosystem Extent Dynamics

Default cloud environment

Lightweight, Pre-trained Transformer for Remote Sensing Timeseries

Extracts general-purpose features useful for diverse EO applications



Trained on diverse, global, unlabeled EO data

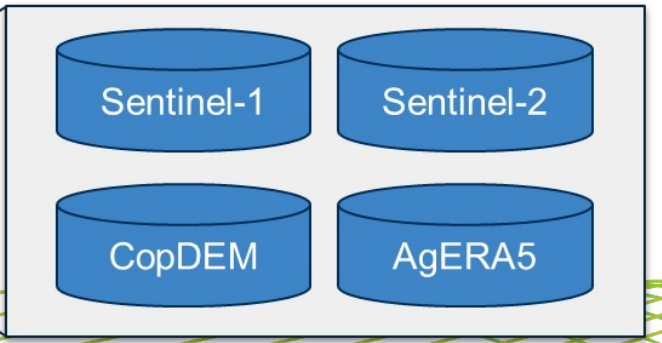
Scalable & customizable crop mapping system

Data loading

Preprocessing

Feature extraction

Classification



Crop type maps

PROGRAMME OF THE EUROPEAN UNION

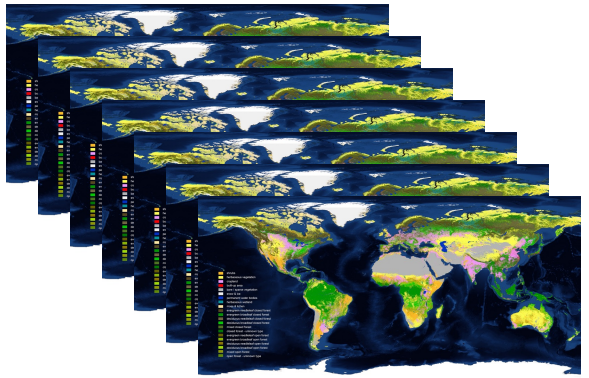
Copernicus

esa

LCFM Objectives: A dynamic global land cover service

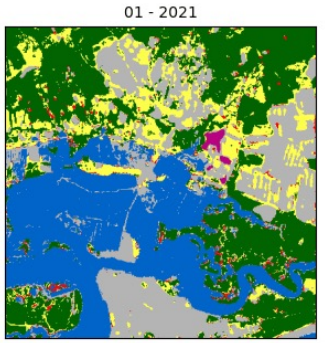
CONTINUE

- Continuation of 100m global LC
- Yearly updates (2020-2026)



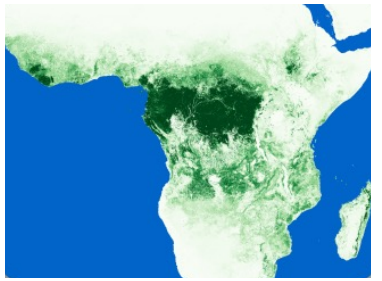
ENHANCE

- Spatial resolution: towards 10 m
- Temporal resolution: towards monthly products and NRT
- Improved accuracy
- Consistent change mapping

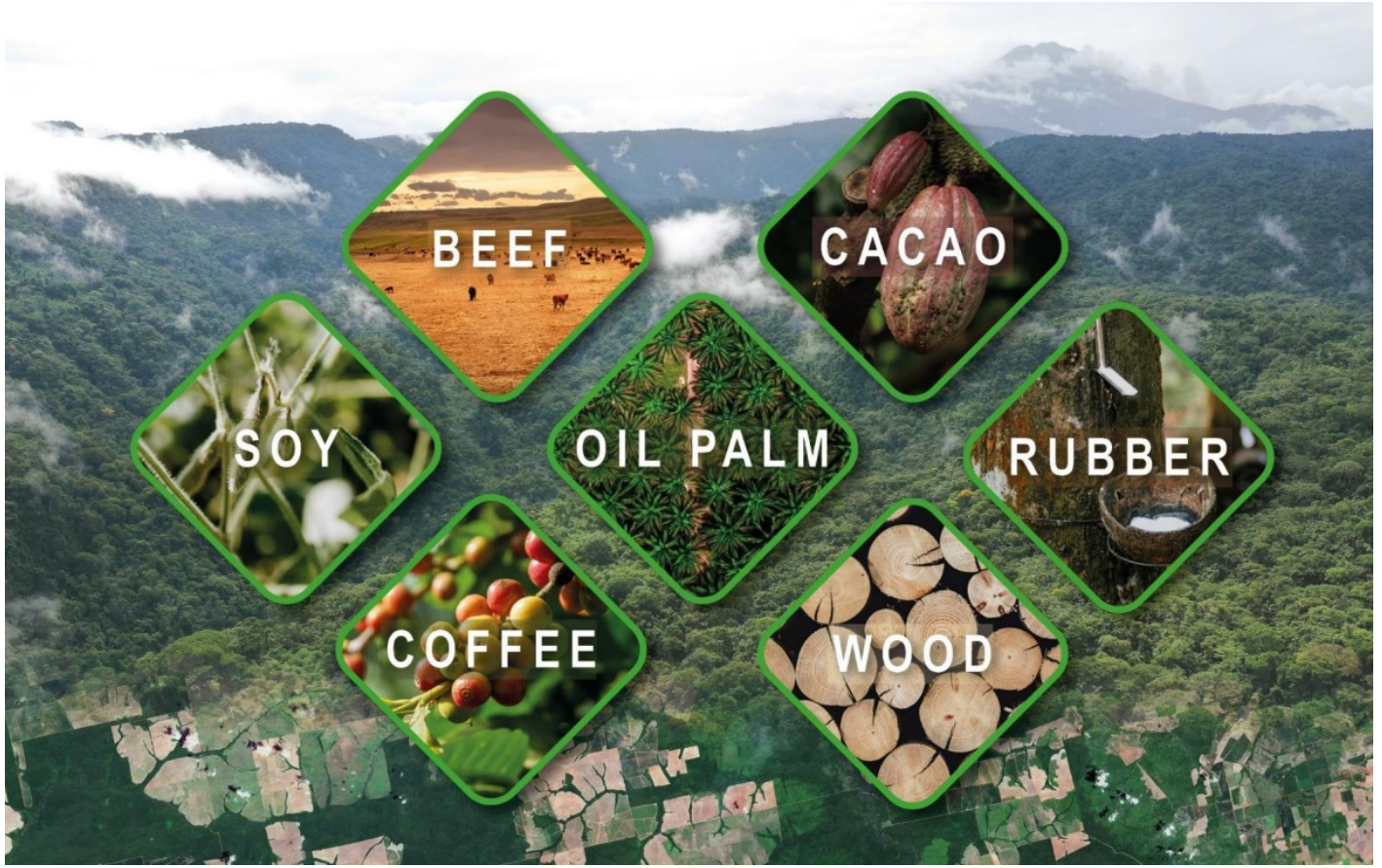


EXTEND

- Specific Tropical Forest Products (TCD, TCPC)



World Agro commodities



- Agricultural commodities detection and monitoring system.
- Supporting the European Union's global deforestation reduction efforts.
- Piece of the puzzle to validate no deforestation occurred for products which reach the EU market

PROGRAMME OF THE EUROPEAN UNION

Conclusion

Conclusion

- openEO on CDSE is a mature solution providing solutions from prototyping to global upscaling, while standardizing and promoting FAIR open science
- Vibrant ecosystem with continuous open-source contributions, advancing new (AI) features, improve efficiency and cost
- Increased free tier availability with sponsored or higher tier options available
- The federation concept of openEO is enabling building a strong EU ecosystem and increased visibility to provide alternatives next to GEE and PC due to the combined, homogeneous data offering
- Large scale projects are being enrolled on top of openEO

DestinE Platform

DESTINE PLATFORM ONBOARDING

Najla Said

Serco Italia S.p.A.

Destination Earth Funded by the European Union Implemented by ECMWF esa EUMETSAT

DESTINE PLATFORM OVERVIEW

Destination Earth

Platform Management services

Identity Access management
- Access authorization
- User admin, federation

Data management
- Data privacy
- GDPR Compliance

Orchestration platform
- Security capabilities
- Resource management

Monitoring dashboard
- Manage resources
- Track performance

Onboarding
- Streamlined process
- Dedicated support

User support
- Service Desk
- Support, FAQ, Forum

Data Management services

Visualization
- 2D/3D/4D data visualization
- Export feature

Data workflow
- Access to DestinE data
- Data Cache service

Development environment
- Ready-to-use environments
- AI notebooks

Processing environment
- Advanced processing via UI/API
- Build code & configuration via UI

Traceability

Tenancy

Advanced services

Sandbox

Collaborative services

Dashboard

e.g. Flood monitoring

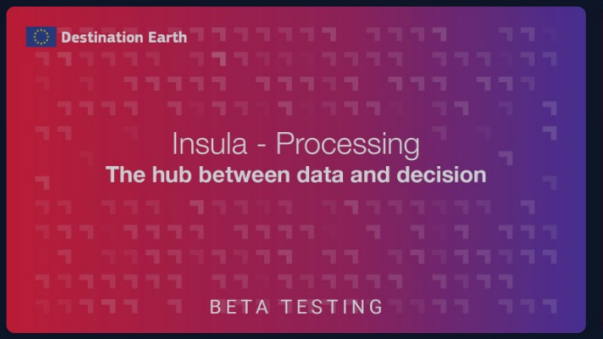
e.g. Climate risks assessment

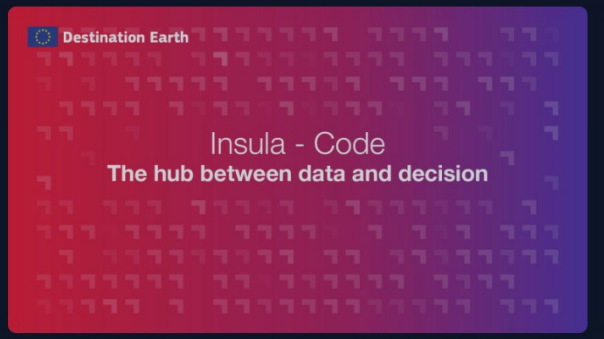
External services

Digital Twins

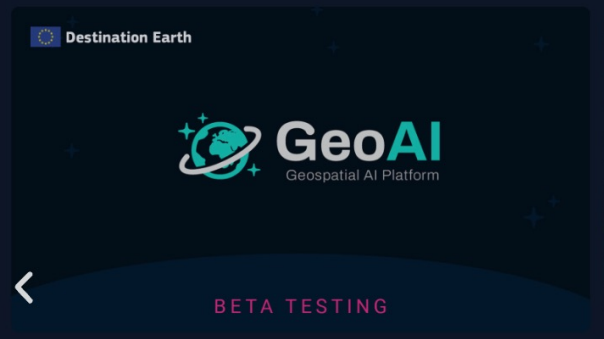
EUROPEAN CLOUD INFRASTRUCTURE

Access and exploit data



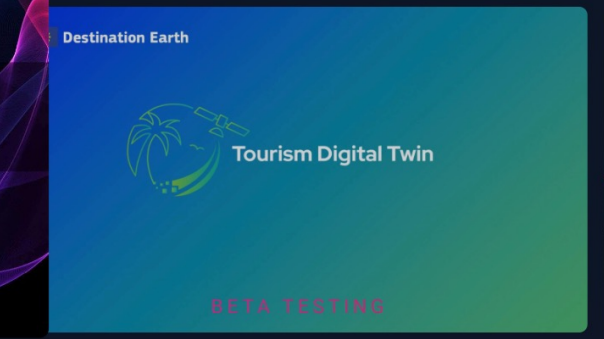


Discover DestinE applications



miniDEA
DATA VISUALISATION

Alia Space Systems | Interactivity | Climate | DestinE | Data Visualization



ONBOARDING
PROCESS

Destination Earth

01 REQUEST

DIRECTLY ON THE PLATFORM, FILLING A DEDICATED FORM

- Provide all information and documentation about the service
- Acceptance T&Cs and Code of Conduct
- Compliance with security requirements

02 EVALUATION

GOVERNANCE BOARD SELECTS SERVICES BASED ON DEFINED CRITERION

- Alignment with DestinE objectives
- Innovation, market need
- Long-term sustainability

03 INTEGRATION

SELECTION AND IMPLEMENTATION OF INTEGRATION SCENARIO

- Integration with IAM
- Deployment solution
- Optional integrations

PUBLICATION ON SERVICE REGISTRY

SUPPORT BY DESTINE PLATFORM ONBOARDING TEAM

platform.destine.eu/onboarding

DestinE Platform

FOCUS ON
ONBOARDING REQUEST

Destination Earth

01

Service Description & Resources

- Provide Service description
- Provide Service documentation
- Provide required information

02

Communication setup

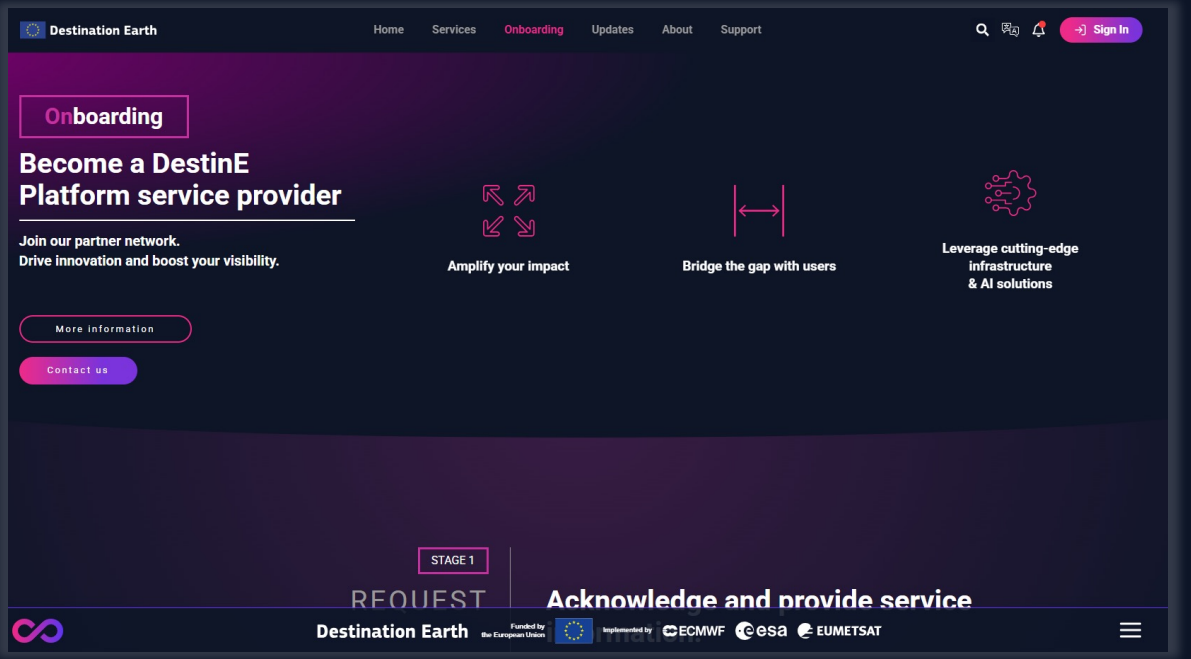
- Define operational Point of Contact for operation & integration activities
- DestinE Platform Service Desk

03

Compliance acknowledgement

- Acceptance of Platform T&Cs
- Code of Conduct
- Security requirements (privacy policy)

https://platform.destine.eu/onboarding

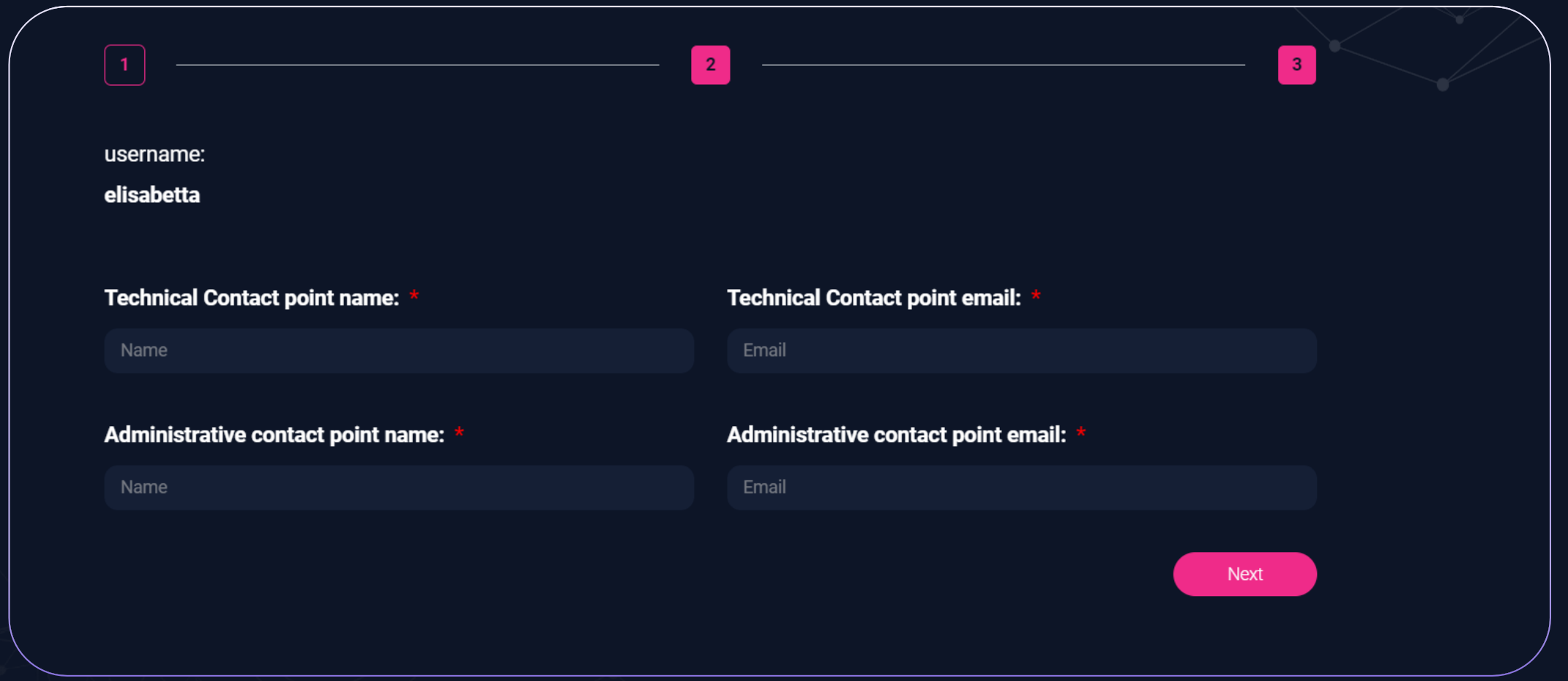


DestinE Platform

SERVICE ONBOARDING REQUEST

PLATFORM.DESTINE.EU/ONBOARDING

Destination Earth



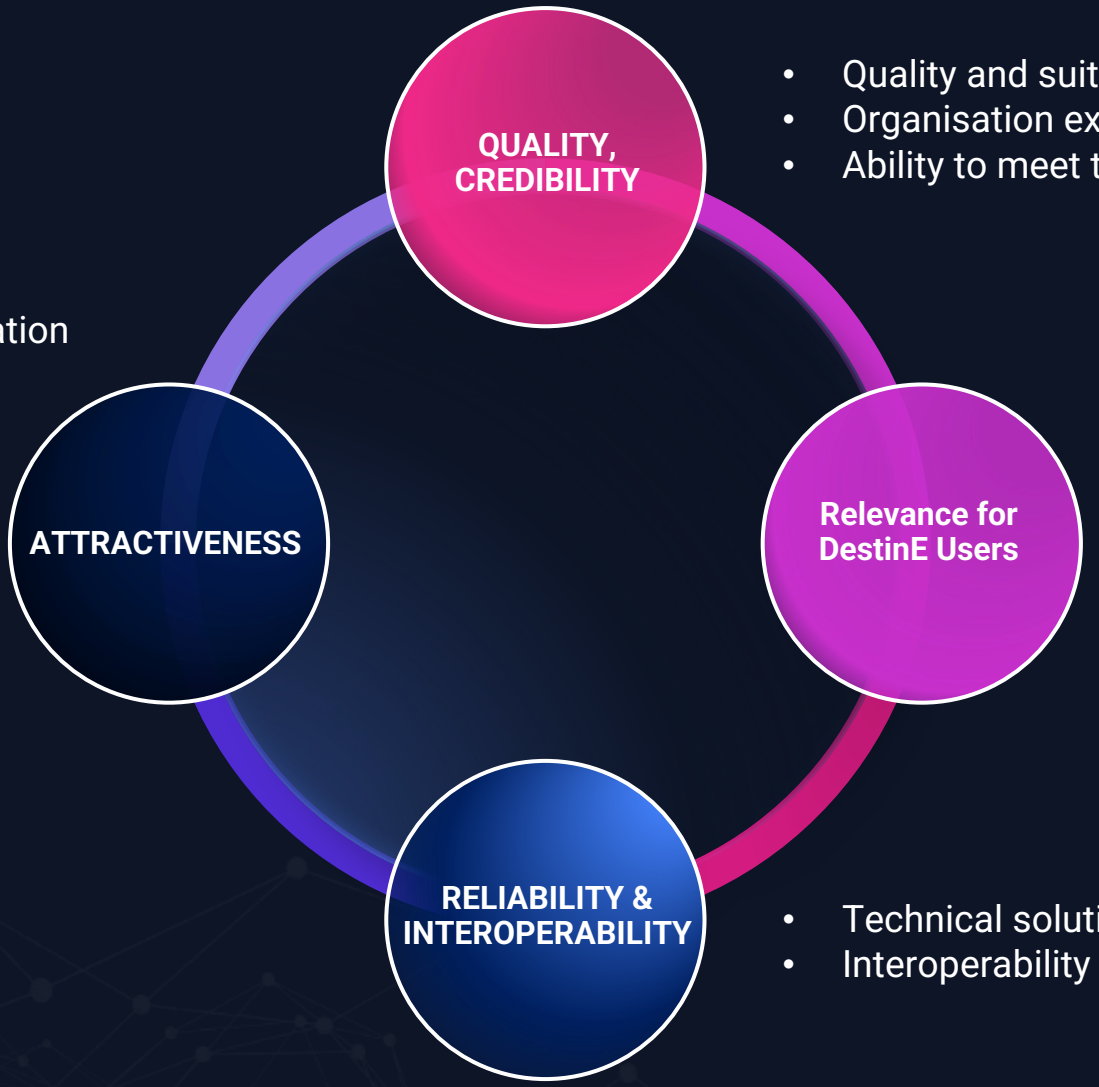
The applicant must be registered on the platform and logged in

DestinE Platform

SERVICE EVALUATION

MAIN CRITERION

Destination Earth



- Quality and suitability of proposed solution
- Organisation experience & capacity in providing operational service
- Ability to meet the DestinE Platform SLA

- In line with EU Green Deal and EU Digital Strategy objectives
- In line with T&Cs and CoC

- Technical solution maturity, scalability
- Interoperability to ensure high performance level

- Technological innovation
- Market demand

SERVICE INTEGRATION
OVERVIEW

Destination Earth

LIVE COMMUNICATION | ASYNCHRONOUS COMMUNICATION

TECHNICAL ENGAGEMENT → PLANNING → MANDATORY/ OPTIONAL INTEGRATIONS → SERVICE VERIFICATION → SECURITY ASSESSMENT → PUBLICATION ON SERVICE REGISTRY

Starter Kit

Jira **setup**

JIRA Kanban Board:

- Dedicated JIRA Project (named DOSI)
- Integration Tasks created by Serco
- Tasks assigned to technical contact point
- Users can update tasks statuses / insert comments / attach files
- Notifications sent for any action
- JIRA Epic for each new Service
- Unique Board (filter possible via Epic)

Regularly review the tasks board for insights/feedback provisioning

Email communication channel for questions/requests

TAILORED INTEGRATION

MAIN CHOICES

Destination Earth

Identity Access management

- Integration with DestinE Platform IAM
- Federation

Service Registry

- Integration with DestinE Platform service registry

Deployment

- Integration on Runtime Platform
- Usage OVH tentant DESP
- Your own infrastructure / cloud provider

Data Management Services

- Data Cache
- Data access
- Data Processing
- Workflow orchestration
- Visualization

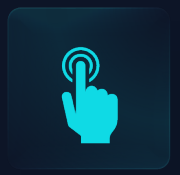
Additional integrations

- Monitoring System
- Tenancy

DestinE Platform

FOCUS ON
RUNTIME PLATFORM BENEFITS

Embedded Security
- Secure Architecture
- DevSecOps pipeline to undergo automated Security Checks



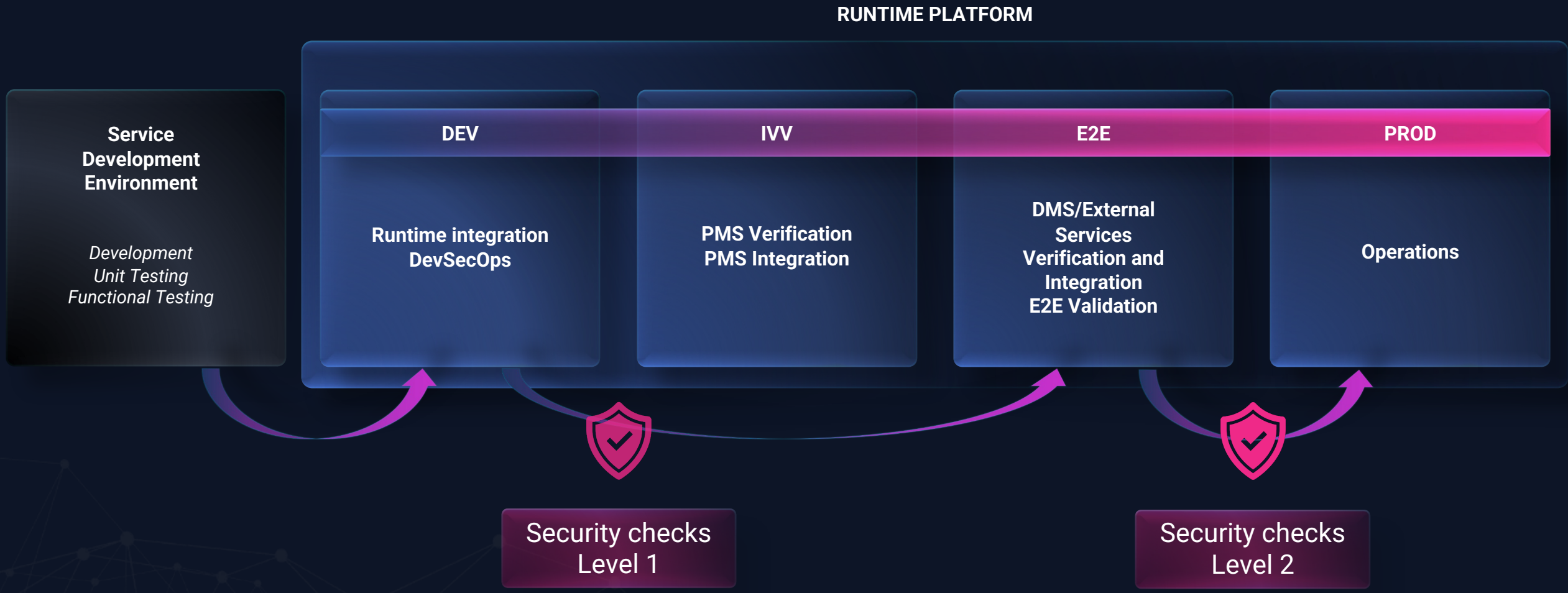
High-Availability
- Microservices orchestration to build flexible and modular Applications
- Kubernetes management of multiple replicas to boost availability

Resource optimization
- Node scalability allows to consume only what is needed
- Non-operational platforms leverage shared resources

Infrastructure operations
- Centralized infrastructural management and monitoring
- Backup and Recovery management

FOCUS ON
DEVSECOPS PIPELINE

Destination Earth



Four environments to run security checks, integration and E2E tests and finally move to Operations

DestinE Platform