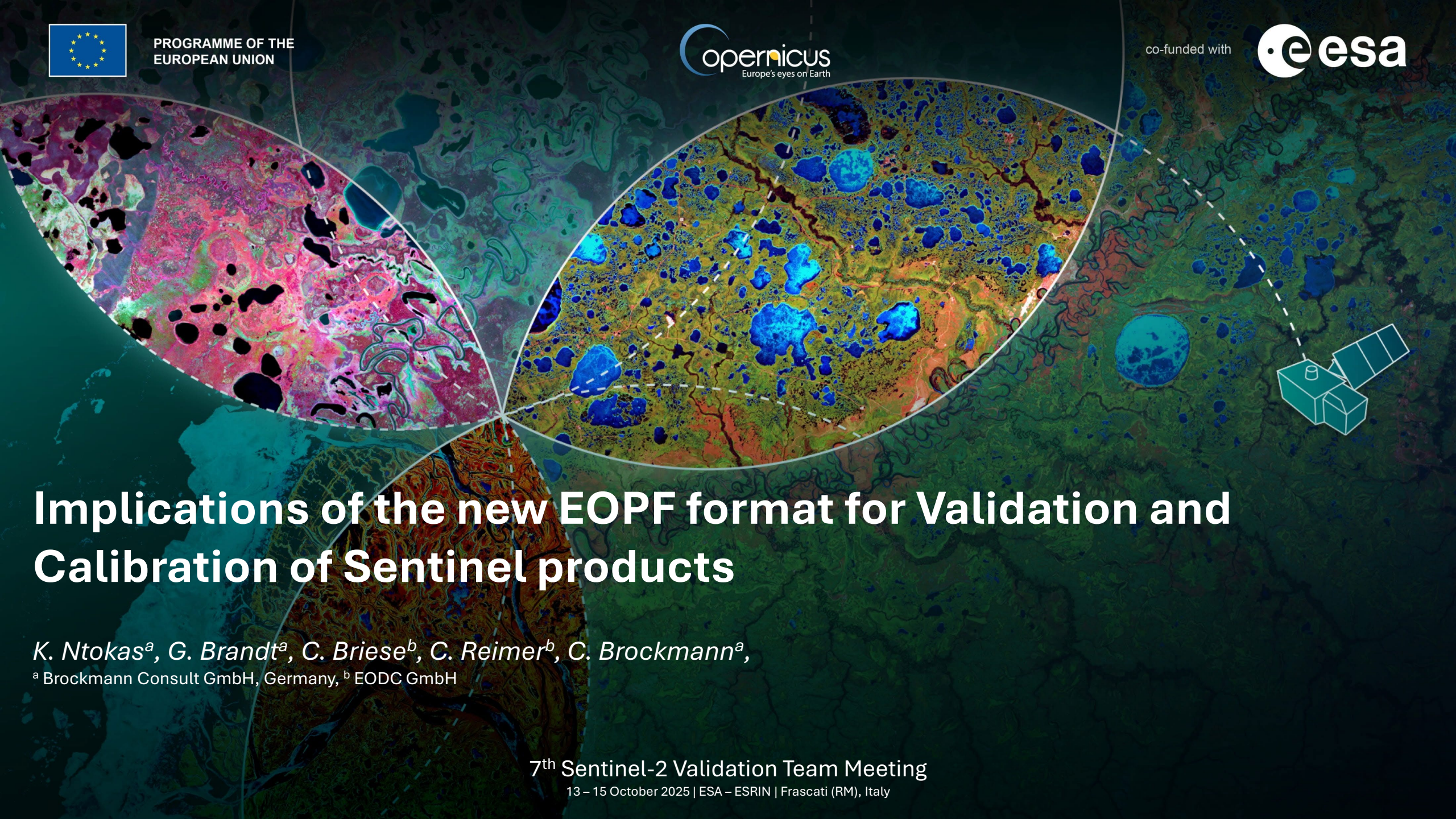


Implications of the new EOPF format for Validation and Calibration of Sentinel products

K. Ntokas[a], G. Brandt[a], C. Briese[b], C. Reimer[b], C. Brockmann[a],
[a] Brockmann Consult GmbH, Germany, [b] EODC GmbH

7th Sentinel-2 Validation Team Meeting
13 – 15 October 2025 | ESA – ESRIN | Frascati (RM), Italy

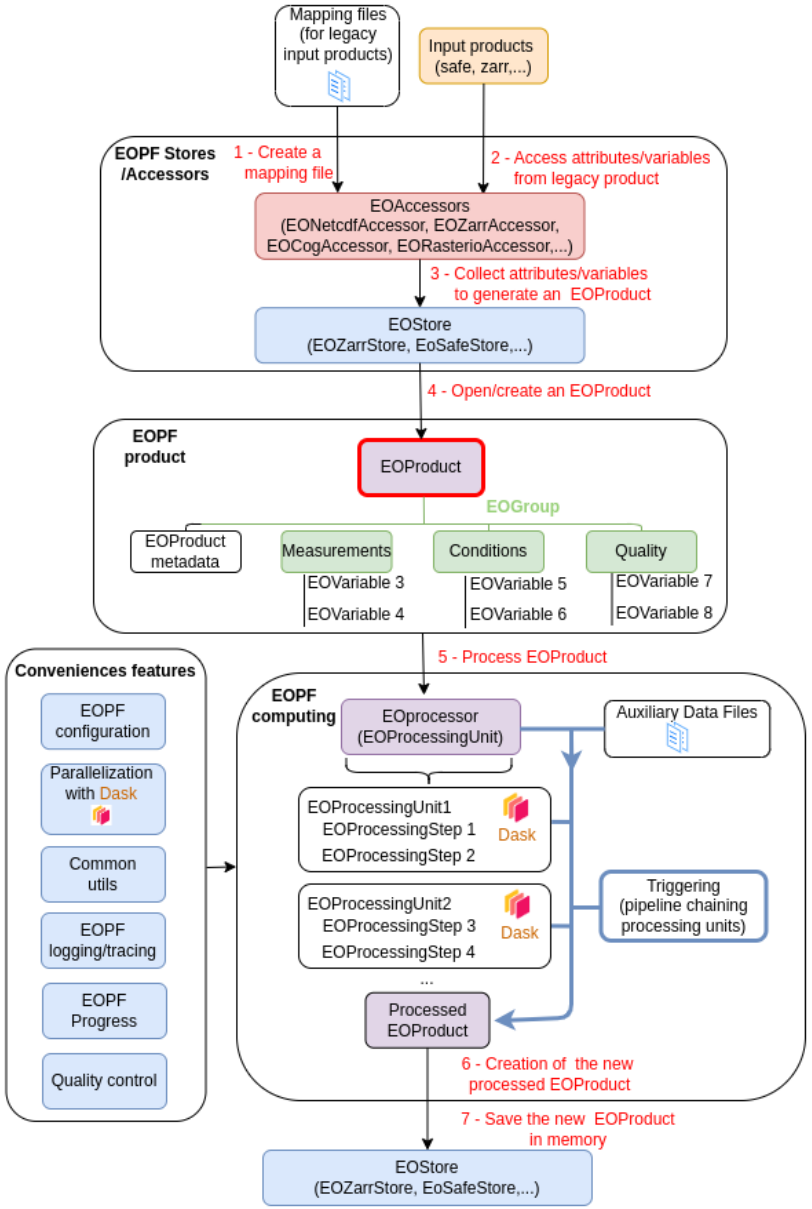
Introduction

ESA developed new Earth Observation Processor Framework (EOPF):

- Unified, up-to-date processor architecture for all Sentinel user level products
- Supporting cloud-based operations and storage, facilitating on-demand processing, also by users
- New standard format for all Sentinel User Level products based on Zarr

Through different activities ESA delivers

- The EOPF software to generate Sentinel products (https://cpm.pages.eopf.copernicus.eu/eopf-cpm)
- Numerous samples of user-level products for Sentinels 1,2, and 3 (L0-L2)
- Tools, plugins, notebooks, webinars supporting users to access and exploit the new data format (https://zarr.eopf.copernicus.eu)



EOPF Zarr Format – Setting the scene

Current Situation:

- Each mission delivers data in different formats with unique metadata structures:
  - Sentinel-1 → GeoTIFF
  - Sentinel-2 → JPEG2000
  - Sentinel-3 → NetCDF
- Some of these formats are *not cloud-optimized*, making it difficult to:
  - Access data efficiently
  - Retrieve only relevant subsets

ESA's Response: CSC Data Processors Re-engineering Initiative

- Introduction of the **EO Processing Framework (EOPF)**
  - Modernisation of L0–L2 data processors for Sentinel-1, -2, -3
- **Core Python Modules (CPM)** enable:
  - Unified data model across missions
  - Automatic conversion of SAFE files into **hierarchical Zarr data trees** with a consistent root structure

Result: A Paradigm Shift

→ From full-file downloads to **on-demand, selective data access and real-time pre-processing**

Why Zarr

Shift to Cloud-Native & Interoperable Architectures

- To improve **data accessibility**, **scalability** and **integration** with modern processing frameworks

How EOPF CPM Enables This

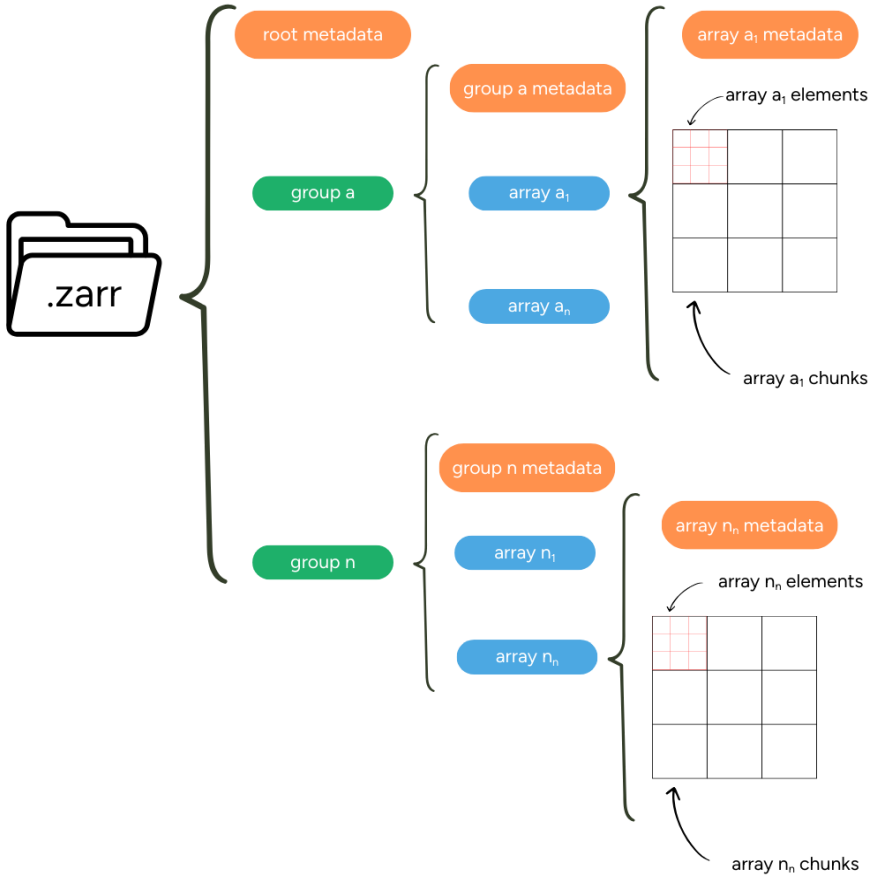
- **Dask** → Parallel & distributed computation
- **Zarr** → Preferred storage format for large multi-dimensional arrays

Why Zarr + Cloud Object Storage?

- ✅ Efficient storage & retrieval (chunking and compression)
- ✅ Parallel read/write operations
- ✅ Lazy loading & partial access — *retrieve only what you need*
- ✅ Embedded metadata stored in a structured, searchable format (catalogue)

Data structure

General structure of a Zarr data tree

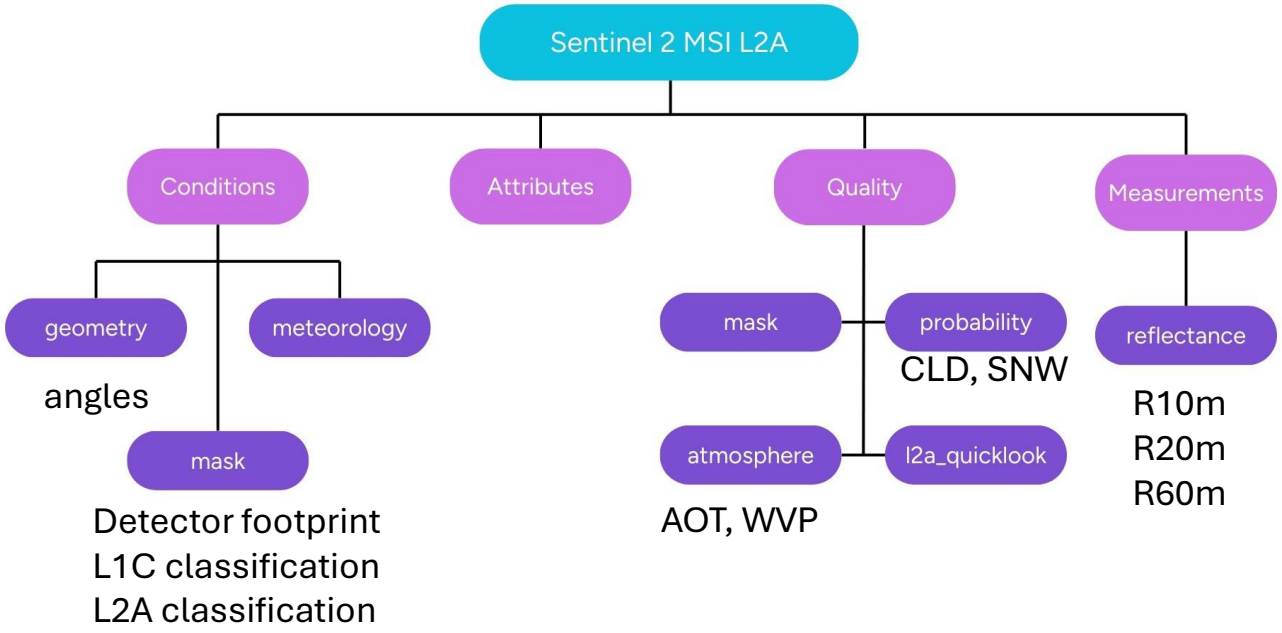


Unified Data Structure Across All Missions

Top-Level Groups:

- **Conditions**
- **Measurements**
- **Quality**

Subgroups contain the actual data arrays.



PROGRAMME OF THE EUROPEAN UNION
Copernicus Europe's eyes on Earth
co-funded with esa

Current Availability of EOPF Product Structures

- Product structures for **Sentinel-1, Sentinel-2, and Sentinel-3** are already published in **EOPF format** as reference examples for processor re-engineering.
  (https://eopf.copernicus.eu/eopf-products-and-adfs/eopf-product-structure-examples/)
- **Level-1 and Level-2 sample products** are accessible via the **EOPF Sample Service**
  (https://zarr.eopf.copernicus.eu/)
- **Additional missions and product levels** will be supported progressively in future releases.

	Sentinel-1	Sentinel-2	Sentinel-3	
Level-0		S2 L0 MSI S2 L0P MSI	S3 L0_ ALT S3 L0 DOP S3 L0 NAV S3 L0 GNS S3 L0 HKM2 S3 L0 HKM S3 L0 MWR	S3 L0 NAT S3 L0 OLCI CR0 S3 L0 OLCI CR1 S3 L0 OLCI EFR S3 L0 SLSTR S3 L0 SR CAL
Level-1	S1 L1 IW GRD S1 L1 IW SLC S1 L1 WV SLC	S2 L1A MSI S2 L1B MSI S2 L1C MSI	S3 L1 OLCI EFR S3 L1 OLCI ERR S3 L1 SLSTR RBT	
Level-2		S2 L2A MSI	S3 L2 OLCI LFR S3 L2 SLSTR LST S3 L2 SYN AOD S3 L2 SYN SDR	S3 L2 SYN V10 S3 L2 SYN VG1 S3 L2 SYN VGK S3 L2 SYN VGP

Relation to Cal/Val – Opportunities

Time saving

- Every new sensor requires time for writing the code to open the images, dealing with the different coordinate systems, looking into the metadata for the scaling factors, rather than doing the validation analysis (e.g. match-ups) itself.
- Having S2 MSI, S3 OLCI, and for radiometry also S3 SLSTR, in zarr is that is ready to use, in a uniform structure, which is easy to prototype

Cloud-native data analysis

- No download, no data transfer costs
- Direct access to S2 products immediately after processing by the ground segment
- Fast, chunk-based processing
- Recommendation to ESA:
  - maintain and curate reference data collections at CDSE (Aeronet, Hypernets, GBOF, OCDB,, ...)
  - migrate calibration site databases to CDSE (RadCalNet, SarCalNet)
  - adapt / develop specific cal- and val tools (database from DIMITRY, match-up extraction, ...)

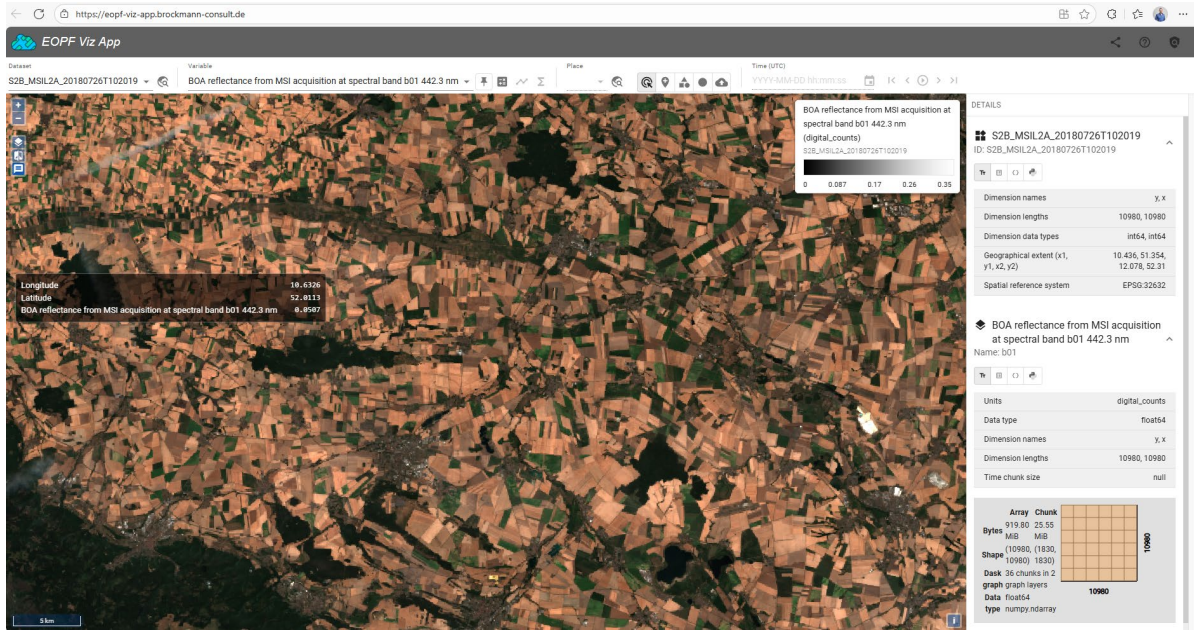
Ease the analysis of large time series stacks

- Very simple to construct data cubes with a time-axis
- Re-chunking in the time domain possible (support tools in EOPF sample service)

Relation to Cal/Val – Opportunities

- Multi-mission data analysis
  - Seamless integration of Sentinel-1, -2 and -3 data into unified data cubes
  - Harmonized product structure enables cross-sensor algorithm interoperability and streamlined metadata access
  - Compatible with interactive visualization tools (e.g. xcube Viewer, or new development by ESA EOPF viewer, …)
- Native Compatibility with the Pangeo Python Stack
  - Built on open-source standards such as:
    - **xarray** – labelled multi-dimensional data handling
    - **dask** – scalable parallel and distributed processing
  - Reduced memory limitations thanks to **lazy loading and chunk-based execution**
- Match-up analysis
  - Fast processing of 100's GB of input products
  - Match-up identification (protocols for selection criteria, filtering and producing match-ups)
  - Computation of metrics (APU, confusion matrix, …)
  - Generation of plots and animations

(https://eopf-viz-app.brockmann-consult.de)

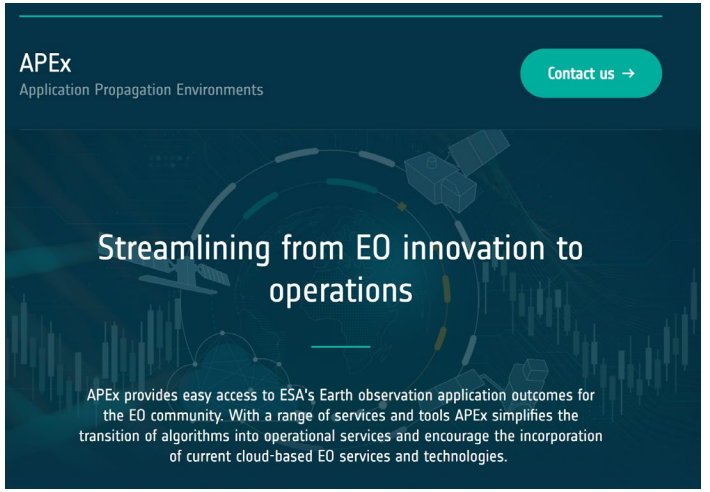


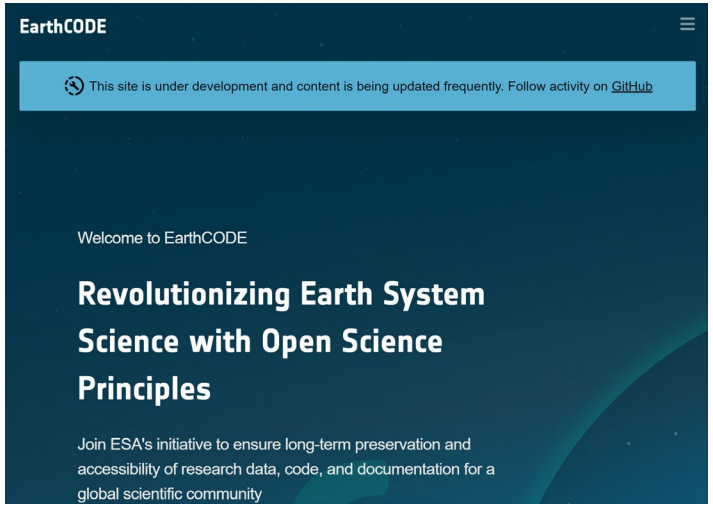
Relation to Cal/Val – Required adaptations

- Knowledge
  - Handling of cloud data sets (S3 buckets, chunking, …)
  - Setting-up own cloud environment (largely supported by virtual labs but still …)
  - Costs
- Adaptation of existing tools
  - Readers
  - Meta data handling in the EOPF format
  - Data model
  - Writers

Relation to Cal/Val – Support

- Sharing of tools
  - **APEx** - access to ESA's Earth observation application outcomes for the EO community; support to users to transfer prototype code into scalable and reusable EO services
  - **EarthCODE** - ESA's initiative to ensure long-term preservation and accessibility of research data, code, and documentation for a global scientific community (https://earthcode.esa.int)
    - Open Science Catalogue - collection of publicly available geoscience products, datasets and resources
- ESA currently maintains an **EOPF Sample Service** including documentation, software tools and Demo Notebooks (next slides)





EOPF Sample Service

PROGRAMME OF THE EUROPEAN UNION

co-funded with

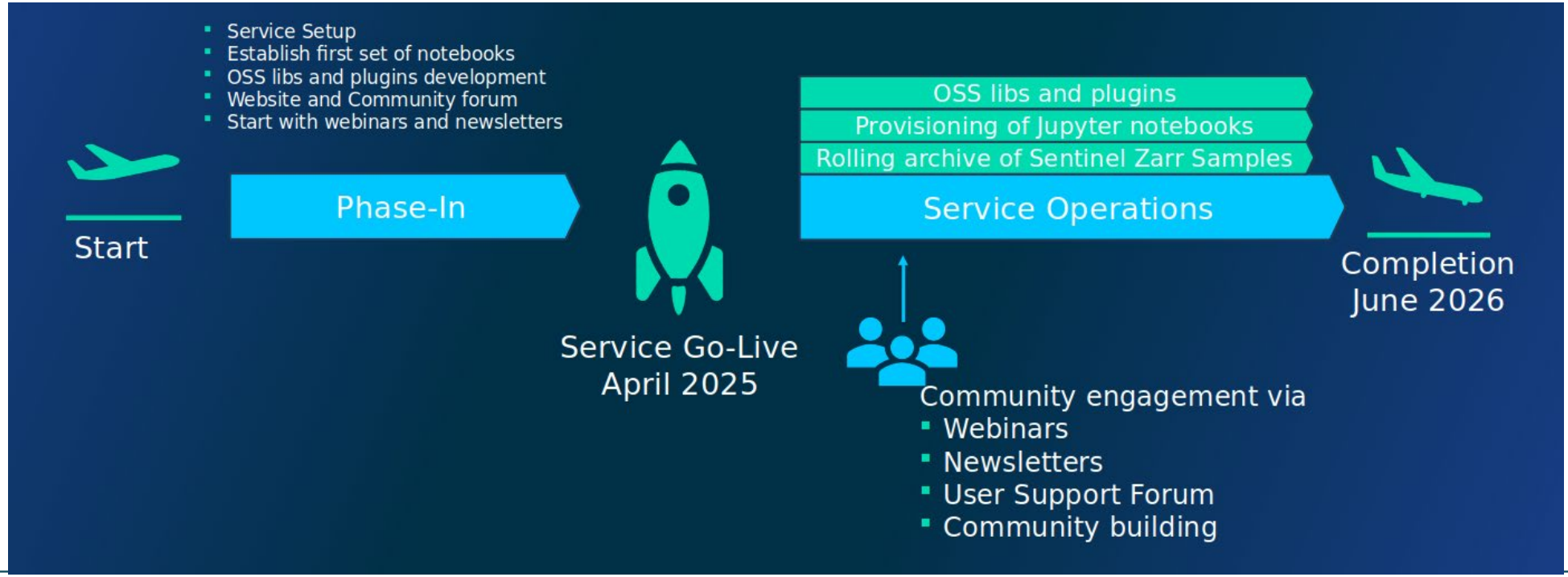
Goal: Enable early and frictionless adoption of the new data format by offering access to live sample data, along with tools and interfaces that demonstrate efficient access to data subsets and time series.

→ This is addressed through three workstreams:

- Building a rolling archive of EOPF Zarr sample data, searchable via a STAC API
- Developing open-source libraries and plug-ins
- Engaging and supporting the user community

EOPF SENTINEL ZARR SAMPLES

(https://zarr.eopf.copernicus.eu)

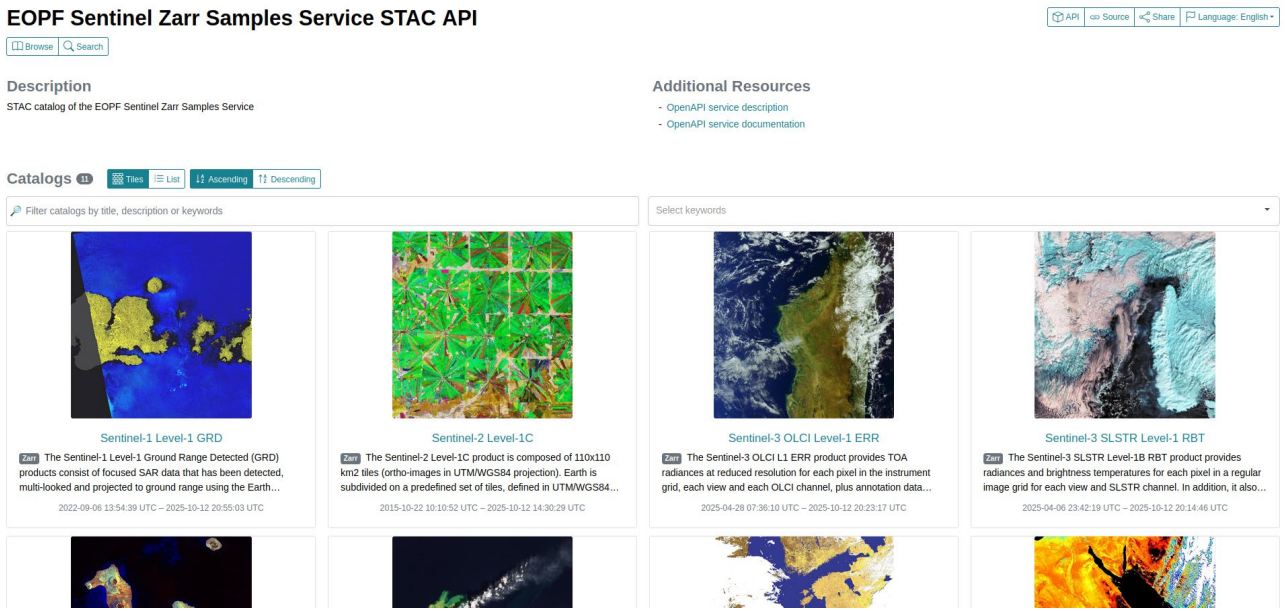


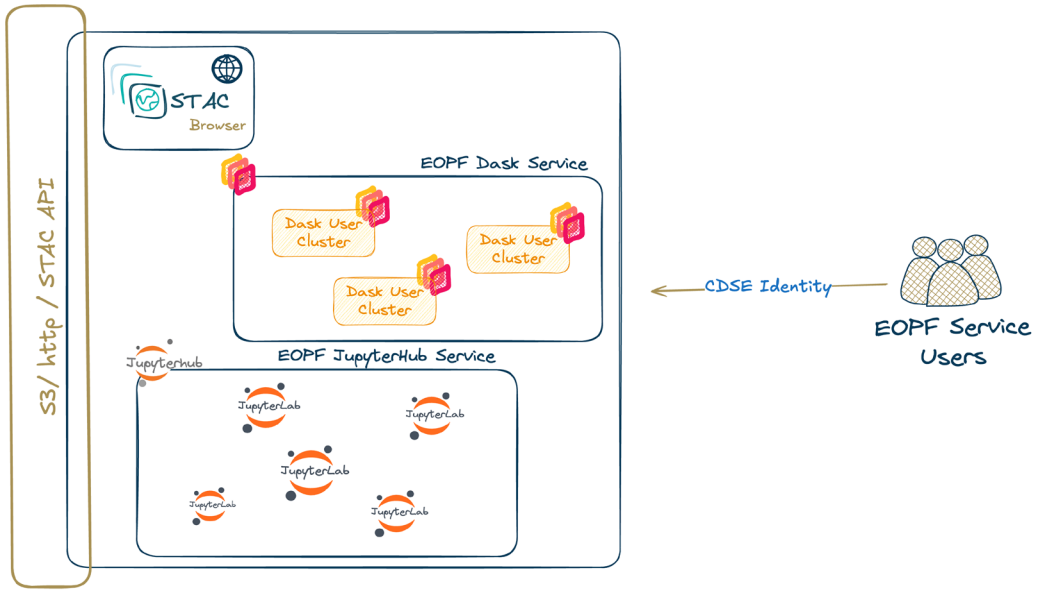
EOPF Sample Service – Data Access

PROGRAMME OF THE EUROPEAN UNION

co-funded with

- **Rolling archive** of recent Sentinel-1, -2 and -3 Level-1 and Level-2 products (≈750 TB), openly accessible via cloud
- **JupyterHub** environment for easy execution, accessible with CDSE credentials
- **Dask cluster** integration available for users





STAC API

- Data is discoverable via a STAC API (stac.browser.user.eopf.eodc.eu)
- Designed to closely mirror the CDSE STAC API and provides equivalent metadata content

EOPF Sample Service – Plugins

xarray EOPF backend

- Read individual observations

xcube EOPF data store

- Build 3D spatial-temporal data cubes from multiple observations

GDAL EOPF reader

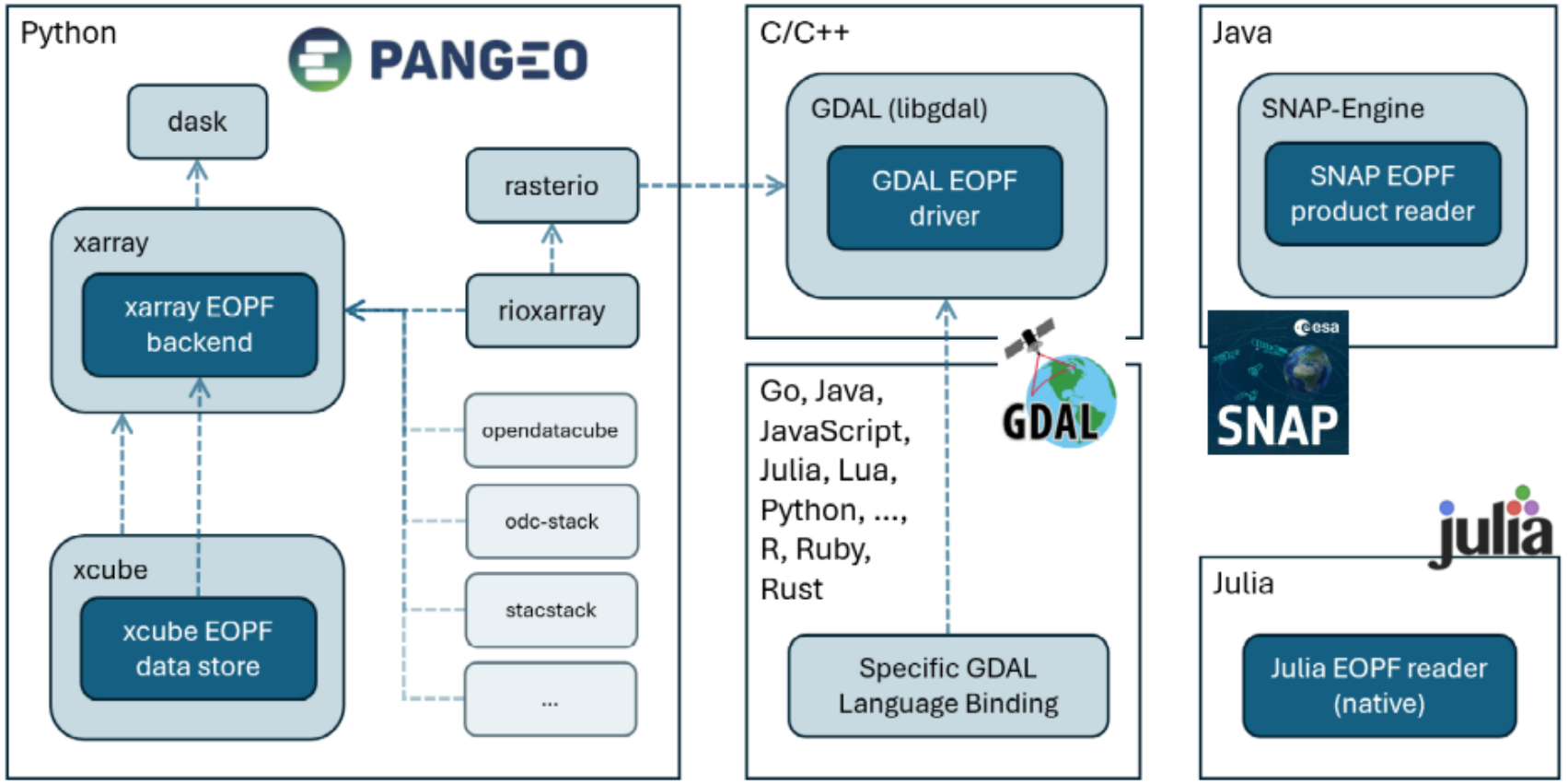
- Integrates with QGIS for geospatial analysis

Julia EOPF reader

- Provides support for the Julia programming community

SNAP EOPF Prototype Reader

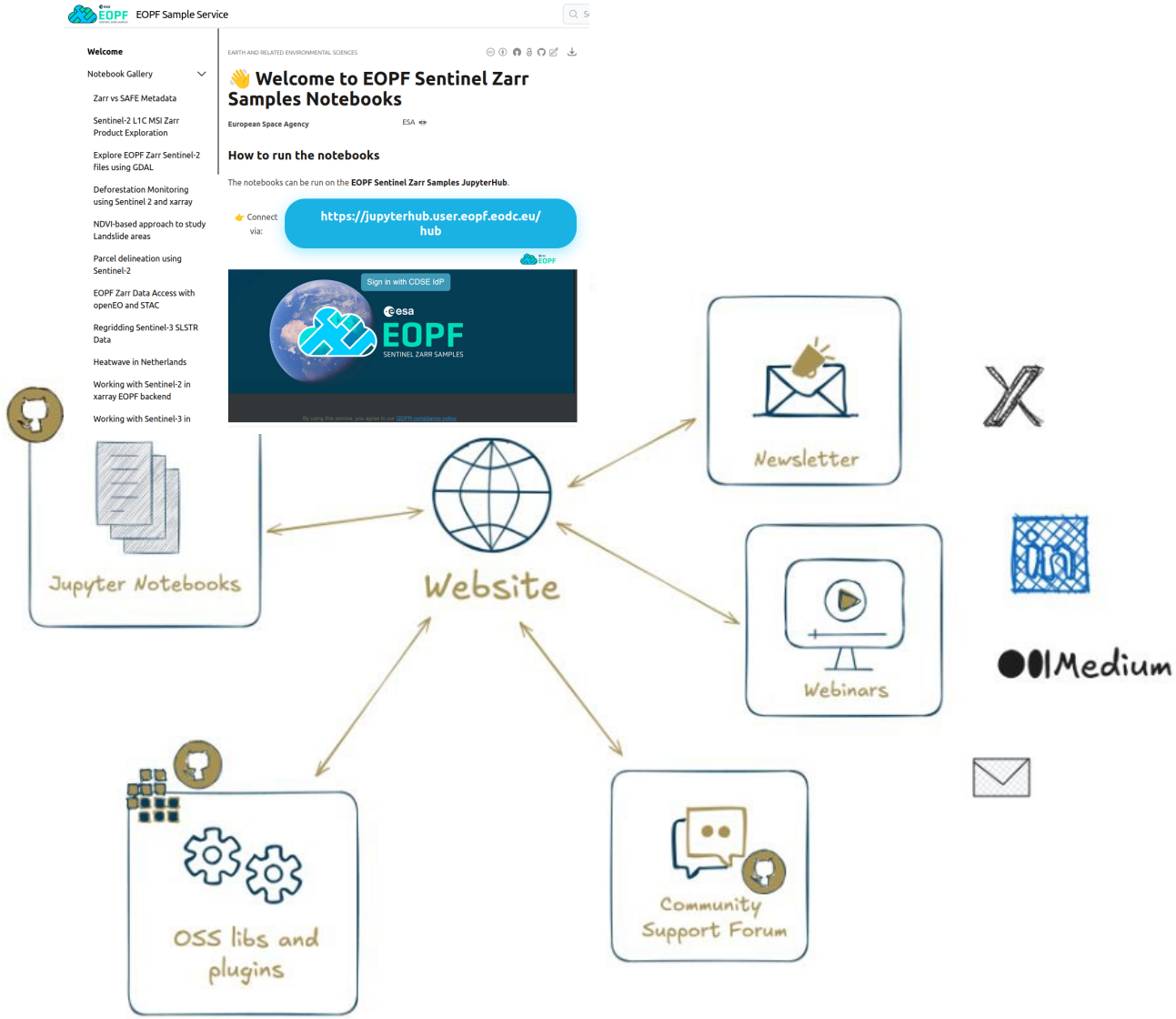
- Proof of concept implementation
- Currently supports only Sentinel-2 L2A products



EOPF Sample Service – Community Engagement

PROGRAMME OF THE EUROPEAN UNION
Copernicus Europe's eyes on Earth
co-funded with esa

- **Easy onboarding** with CDSE user ID
- **JupyterHub** with all required plugins and dependencies pre-installed, plus easy integration of example notebooks (https://jupyterhub.user.eopf.eodc.eu/hub)
- **Example Notebooks** showcase CDSE examples as well as new use cases and recently developed plugins (https://eopf-sample-service.github.io/eopf-sample-notebooks)
- **Newsletter** with updates and news (https://zarr.eopf.copernicus.eu/knowledge-hub/#news)
- **Discourse channels** for discussions and community support (https://discourse.eopf.copernicus.eu/)
- **Monthly webinars** for training and updates and showcasing new developments (https://zarr.eopf.copernicus.eu/knowledge-hub/#webinars)



Take away

- EOPF is the new Earth Observation Processor Framework
  - Unified, up-to-date processor architecture for all Sentinel user level products
  - Supporting cloud-based operations and storage, facilitating on-demand processing, also by users
  - New standard format for all Sentinel User Level products based on Zarr

- EOPF format facilitates working with Sentinel data and especially the analysis of large amounts of data as well as multi-mission analyses

- Adaptation of user tools and experience of working in the cloud are required

- ESA is providing support with EOPF Sample Service, APEx and EarthCODE

- ESA specific support for CalVal would be welcomed: curated reference data collections at CDSE, adapt / develop specific cal- and val tools, …