AI4 copernicus

Annex 2.2:

Technical Documentation

Reinforcing the AI4EU Platform by Advancing Earth
Observation Intelligence, Innovation and Adoption

OPEN
CALL

AI

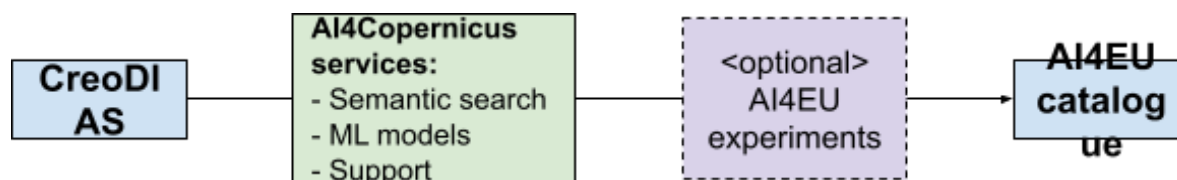# Annex 2: Technical Documentation

Version 2 (September 2021)

Table of Contents

Quick start

Overview

AI4Copernicus provides users with access to DIAS platforms and AI4EU resources in a streamlined way, offering support along the way.

The expected roadmap for interacting with AI4Copernicus resources is:
- **Develop** (CREODIAS or local/private resource)
- *Optional*: **Onboard** onto the AI4EU Experiments platform
    - **Refine** making use of other Experiments resources
    - Deploy the new workflow on CREODIAS or elsewhere
- **Publish** onto the public AI4EU catalogue

(1) We anticipate that, as a developer, you will develop your solution and models on CREODIAS - integration of more DIAS platforms is currently underway.
(Alternative or local resources may also be used for development, however in this case support and the integration of Copernicus data will not be provided by AI4Copernicus.)
CREODIAS will provide you with access to necessary cloud resources as well as to Copernicus datasets and other products.

(2) For exploring and experimenting AI4Copernicus provides a set of additional services and tools, outlined below, e.g. semantic searching.

(3) Once you have designed and built your model you can optionally onboard and publish it on the AI4EU public marketplace. This set of resources provide workflow and sharing functionality.

(4) As a final step, you are encouraged to publish your work on the AI4EU catalogue for other interested parties, researchers and practitioners to be able to discover it.

Getting access

AI4EU

AI4EU can be reached through https://www.ai4europe.eu. The AI4EU catalogue is reachable at https://www.ai4europe.eu/research/ai-catalog. Successful bidders are expected to publish their finished products on this catalogue - more information will be provided upon success.

The AI4EU experiments platform is reachable via https://www.ai4europe.eu/development, where interested parties can register and documentation is provided. Further information and tutorials can be found at the relevant GitHub repository (https://github.com/ai4eu).

CREODIAS

You can register on CREODIAS at https://creodias.eu. As a registered user you are entitled to a free trial. Successful AI4Copernicus bidders will be provided with credits to develop and publish their products.

AI4Copernicus

More information about the participating platforms can be found on the website of AI4Copernicus, under https://ai4copernicus-project.eu/platforms/.

Sample use-case

Let's consider a use case where we calculate the maximum value of a CO column in a given AOI (Area of Interest) and for a certain time period. To achieve this goal the user has access to AI tools to set up a pipeline which runs a workflow in an automated way. The result of this use-case is the value presented on the port of a container deployed in the Kubernetes cluster that is set up in the CREODIAS environment.

Prerequisites:
1. Knowledge base for the example CO processing
   - Finder API, to look for relevant data. Documentation: https://creodias.eu/-/how-to-use-creodias-finder-
   - Help on using the S3 API can be found at the following resources:

---

    i.   HOW TO ACCESS/LIST EO DATA USING BOTO3: https://creodias.eu/-/how-to-list-eo-data-using-boto3-?inheritRedirect=true&redirect=%2Ffaq-s3

    ii.   HOW TO DOWNLOAD EO DATA FILE USING BOTO3: https://creodias.eu/-/how-to-download-a-eo-data-file-using-boto3-?inheritRedirect=true&redirect=%2Ffaq-s3

    iii.   EO DATA ACCESS - S3 OR NFS?: https://creodias.eu/-/eo-data-access-s3-or-nfs-?inheritRedirect=true&redirect=%2Ffaq-s3

    iv.   HOW TO ACCESS EO DATA AND OBJECT STORAGE USING S3CMD (LINUX): https://creodias.eu/-/how-to-access-eo-data-using-s3cmd-linux-v2?inheritRedirect=true&redirect=%2Ffaq-s3

- ○ Interacting with the experimental AI4EU experiments requires knowledge on GRPC communication standards. This is provided at https://developers.google.com/protocol-buffers and https://grpc.io/

- ○ For data acquisition in NetCDF format please refer to: https://gdal.org/index.html

2. CREODIAS Kubernetes cluster
   - ○ kubectl setup for CREODIAS Kubernetes cluster
     Configuration is available after login to CREODIAS under:
     ~/.kube/config
3. Python 3: This use-case has been implemented in Python 3.

Steps:

At the beginning the user needs to prepare a data broker to find relevant data products, and an analyzer to read data from product files. The user then needs to publish the relevant docker images in the image registry.

1. Information on model onboarding is presented at: https://www.youtube.com/watch?v=Ts4KqvvmkMg
2. Pipeline composition and local download: https://www.youtube.com/watch?v=gM-HRMNOi4w&t=6s
3. Configuration of kubectl to point execution environment in CREODIAS
4. Deployment and execution of the solution pipeline

For this particular use-case we prepare:

1. An EO Data broker that searches for Sentinel-5p's carbon monoxide data via the EO Finder API. This broker returns the file metadata to the pipeline designed on AI4EU experiments

2. A CO Max analyzer makes use of the EO data files using S3 API and makes available the calculated value via a REST API

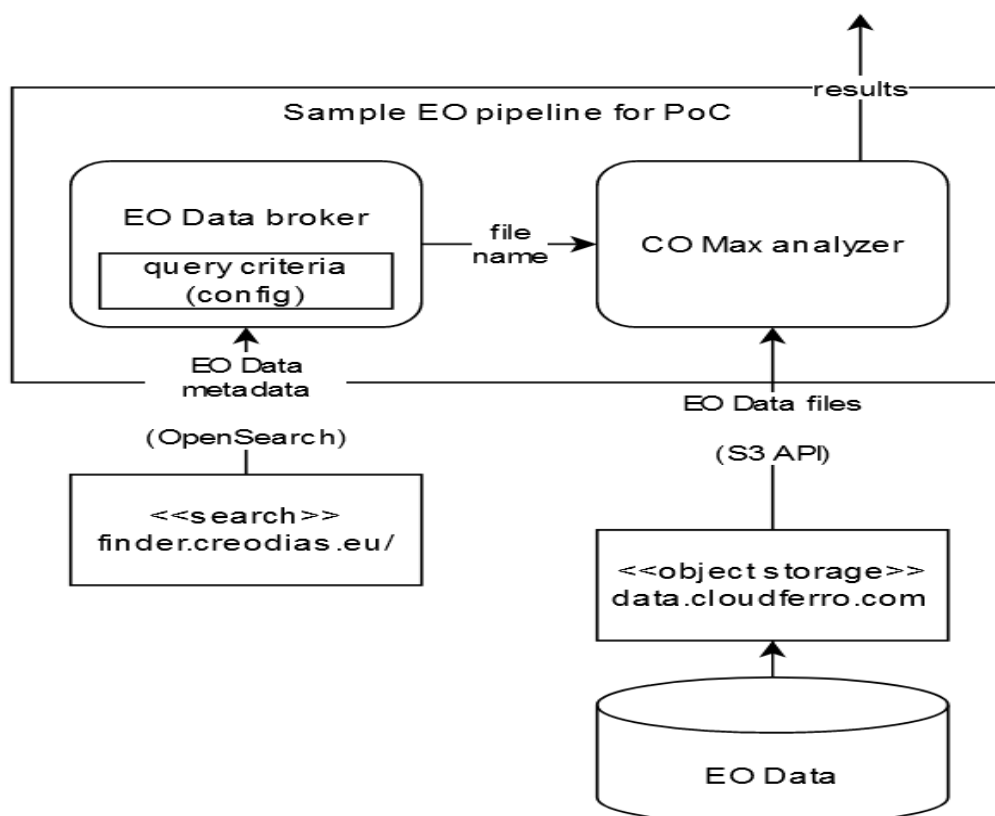The figure below outlines the procedure:

Fig 1. An overview of the pipeline of this example use-case.

AI4Copernicus Services Overview

As part of our technical work in AI4Copernicus we will provide semantic searching over Copernicus data as well as pre-trained ML and related models to be used by our users and bidders. More information on how to access and use them will be provided here soon. Moreover, we will provide linked data tools for transforming, querying, interlinking and federating big linked geospatial data. These linked data tools are the following:

1. GeoTriples (http://geotriples.di.uoa.gr): GeoTriples allows the conversion of various geospatial formats into RDF. The tool can be used to provide widely used geospatial data as RDF resources.
2. Strabon (http://strabon.di.uoa.gr): The spatiotemporal RDF store Strabon can be used for storing and querying big linked geospatial data and enable analytic techniques.
3. JedAI (https://jedai.scify.org): JedAI is a tool for linking geospatial data sources using geospatial relations. The tool can be used to allow interlinking of various datasets.
4. Semagrow (https://semagrow.github.io): The Semagrow federated query processor can be used for creating "virtual datasets" that dynamically integrate heterogeneous linked geospatial datasets.

In addition to these services, several bootstrapping services resources listed in the table below will be provided by AI4Copernicus to support the open calls applications.

Service	Scope	Domain
Sentinel-1 and Sentinel-2 pre-processing pipelines	Pre-processing chains for Sentinel-1 and Sentinel-2 (from Sentinel-1 Level-1 SLC and GRD/Sentinel-2 Level2A), to develop ML algorithms on top of Analysis Ready Data	General, Security
Sentinel-1 and Sentinel-2 processing pipelines	Processing chains for Sentinel-1 and Sentinel-2 (e.g. change detection) to benchmark algorithms	General, Security
OSM-derived vector data	OSM-derived data (e.g. roads, buildings and other features of interest for the security domain) to train and validate ML algorithm	Security
Sentinel-2 time series monthly composite techniques	Harmonization of time-series through monthly composite approach	Agriculture
Supervised classifier based on LSTM deep network	Classification technique based on Long-Short Term Memory (LSTM) deep network optimized for the analysis of time series of	Agriculture

	Sentinel-2 images	
Large and detailed crop training set	Large data set with crop-type labeled multitemporal samples for the training of deep learning architectures	Agriculture
Deep network for pixel-level classification of S2 patches	Backbone network (resNet50) and U-Net model using this backbone. Training binaries for patch pixel segmentation, pretrained backbone (on S2 patches), network and training code to build U-Net for specific applications.	General, Agriculture
Emodnet Human activities Portal	Geospatial data sets of offshore infrastructure (e.g. wind farms, aquaculture etc.) for classification and machine learning processes	Energy
JRC Open Power Plants Database	Geospatial data sets of energy production infrastructure for classification and machine learning processes	Energy
Air quality resource	Downscaling (super-resolution) of CAMS air quality analyses and forecasts using deep learning (e.g. generative adversarial networks)	Health
Air quality resource	High-resolution, targeted air quality and/or atmospheric composition forecasts (e.g. at the city scale) through ML data fusion of CAMS forecasts with local observations, ERA5 / CERRA and Sentinel 5P data.	Health

Contacts for technical information

CREODIAS: https://cloudferro.com/en/why-cloudferro/contact/
AI4EU Experiments: Antonis Koukourikos (NCSR-D): kukurik@iit.demokritos.gr