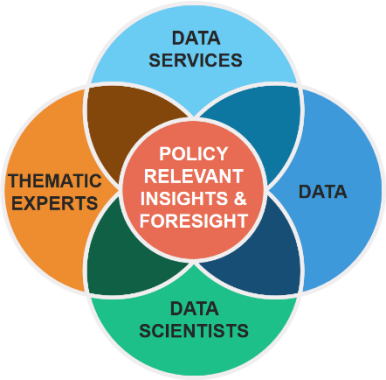
Big Data Analytics Platform – Quick Guide



HIGHLIGHTS

→ **Scope**: The JRC Big Data Analytics Platform (BDAP) links data, data services, data scientists, and thematic experts for generating policy relevant insights and foresight.

→ **Target audience**: any scientist with data and data analytics needs.

→ Create a BDAP account by *registering here*.

→ *More information on the registration service here*.

SERVICES:

JEO-desk

JEO-desk is a data science desktop terminal service accessible from a web browser. JEO-desk runs on BDAP servers with fast connection to the BDAP multi-petabyte scale data pool.

JEO-lab

JEO-lab is a JupyterLab notebook environment, intended to interactively analyse and visualise data via a dedicated API and to create notebooks that leverage on modern GPUs for machine learning algorithms.

JEO-cloud

JEO-cloud provides users with an easy way to transfer data (datasets, documents, scripts, etc.) between their computer and the BDAP infrastructure.

JEO-batch

JEO-batch addresses heavy and long computation needs. It relies on HTCondor job scheduler and leverages on 1,432 CPU cores. Access is through JEO-desk.

Storage

EOS provides multi-petabyte storage as a single volume. Users can link available data or store/share their own.

Catalog

The BDAP catalog provides a web interface to explore and browse all data shared on BDAP.

Contacts

The BDAP ***GitLab instance*** serves as the main communication channel between BDAP users and the BDAP team.
It is also used for storing and sharing code.

BDAP Homepage *Mailbox* *Connected* *SharePoint site*

Can't find what you need?

DOCUMENTATION

OPEN A SUPPORT TICKET

JEO-desk

What it is

- **JEO-desk** is a **data science desktop terminal service** that provides a graphical Linux desktop terminal in your web browser.
- **JEO-desk** offers a wide variety of *data science software*.
- **JEO-desk** runs on BDAP servers with fast connection to the BDAP multi-petabyte scale data pool.

How to access the service

1. *Create* a user account (*reset your password* if you forgot it)
2. Download and install a *user certificate*
3. Enroll an *OTP token* on your smartphone
4. Go to *https://jeodpp.jrc.ec.europa.eu/apps/terminal/*

Use JEO-desk for

- Prototype in your favourite data science environment
- Visualise and analyse results with desktop software
- Perform some light processing
- Access data available in BDAP and share your data

Do not use JEO-desk for

- Perform heavy and long computation (see JEO-batch for that)
- Perform scheduled tasks (e.g. daily download from Internet)

Notes

- Use Chrome or Chromium for enhanced user-experience
- Secure a **reliable** internet connection (*check here* to check that you have latency values < 150ms and jitter < 30ms)
- A lightweight text-based version is also available

LINKS:

→ ***Access JEO-desk***

→ ***See JEO-desk Documentation***

JEO-lab

What it is

- **JEO-lab** offers a **JupyterLab notebook environments.**
- **JEO-lab** environments meet the requirements of (i) deep learning leveraging on GPUs and (ii) interactive analysis and exploratory visualization of geospatial data based on a dedicated API.

How to access the service

1. *Create* a user account (*reset your password* if you forgot it)
2. Enroll an *OTP token* on your smartphone
3. Go to *https://jeodpp.jrc.ec.europa.eu/apps/jhub/*
4. Choose the best suitable environment from the dropdown menu:
    - *GPU Machine Learning with conda*
    - *Interactive Processing - JupyterLab - Python 3*

Use JEO-lab for

- Perform **machine learning** on your data while benefitting from GPU computational power
- Interactively analyse and visualise geospatial data via a dedicated **interactive library and API (see *manual for interactive library*)**

Do not use JEO-lab for

- Perform non-GPU based machine learning (use JEO-desk for that)
- Perform heavy and long computations (see JEO-batch for that)
- Perform scheduled tasks (e.g. daily download from Internet)

Notes

- Within the conda-based environment, users can *create their own conda environment* (also for R)
- Ad-hoc configurations of the environment upon request

LINKS:

→ ***Access JEO-lab***

→ ***JEO-lab Documentation***

JEO-batch

What it is

- **JEO-batch** addresses **heavy and long computation** requirements.
- **JEO-batch** leverages on 1,432 CPU cores.
- **JEO-batch** is accessible from **JEO-desk**.

How to access the service

1. *Create* a user account (*reset your password* if you forgot it).
2. Compile the *use-case form* describing the intended JEO-batch use and return it as indicated.
3. Login to JEO-desk (see JEO-desk card).
4. *Create a submit file* following HTCondor syntax.
5. *Submit your jobs.*

Use JEO-batch for

- Tasks that can be decomposed into many smaller problem which are essentially independent (High Throughput Computing).
- Tasks that are not easy to parallelise and decompose but require long and heavy computations.

Do not use JEO-batch for

- GPU-based applications (GPU is not available from JEO-batch, see JEO-lab for this).
- MPI based computations. Use JRC HPC environments for this (experimental service based on Kubernetes under test on BDAP).

Notes

- HTCondor requires a Docker image to be run. Feel free to contact us for the set-up a dedicate Docker image for your use-case.

LINKS:

→ ***See a list of useful commands***

→ ***See JEO-batch Documentation***

JEO-cloud

What it is

- **JEO-cloud** provides an easy way to transfer data (datasets, documents, scripts, etc.) between user's devices and the BDAP infrastructure.

How to access the service

1. *Create* user account (*reset your password* if you forgot it).
2. Access at least once either JEO-desk or JEO-lab.
3. Connect to *https://jeodpp.jrc.ec.europa.eu/apps/cloud/* or *map JEO-cloud to a network drive*.

Use JEO-cloud for

- Upload/download of files and folders to/from BDAP.
- Share files/folders with other BDAP users or with external users.

Do not use JEO-cloud for

- Share data that are not related to your activities on the BDAP (use JRC-Box instead).
- Transfer files with sizes in the of range of gigabytes. If this is the case, contact us to set-up an FTPS service for your project.

Notes

- The data stored on JEO-cloud are **not backed up**

LINKS:

→ ***Access JEO-cloud***

→ ***JEO-cloud Documentation***

BDAP Data

Data Catalog

- BDAP catalog provides a web interface to **explore and browse** all the collections stored in BDAP
- Try it at *https://jeodpp.jrc.ec.europa.eu/eu/data/stac-browser/*

Collections Explorer

- Dashboard for exploratory visualization and rendering of the main geospatial datasets available at BDAP
- Try it at *https://jeodpp.jrc.ec.europa.eu/eu/dashboard/voila/render/CollectionsExplorer.ipynb*

Storage

- Based on CERN EOS open storage file system
- Accessible from JEO-desk, JEO-lab, and JEO-batch
- Actual data are stored under */eos/jeodpp/data*:
    - /eos/jeodpp/data/base for base datasets of common interest
    - /eos/jeodpp/data/SRS for Satellite Remote Sensing data
    - /eos/jeodpp/data/products for selected data layers produced by JRC projects
    - /eos/jeodpp/data/products/JRC-OpenData for *JRC Open Data* stored on BDAP

Upload your data

- Upload your data to /eos/jeodpp/home/users/USERNAME
- Create a use-case to have a folder to share data with your colleagues
- Place data under the *OpenData repository*

LINKS:

→ ***Access BDAP CatalogAccess JEO-cloud***

→ ***BDAP Storage Documentation***

Icons made
by *ultimatearm*, *freepik* and *itim2101* from *www.flaticon.com*

CONTACT INFORMATION
Pierre.Soille@ec.europa.eu
JRC-JEODPP@ec.europa.eu

The European Commission's science and knowledge service
Joint Research Centre

EU Science Hub: *ec.europa.eu/jrc*
EU Science Hub
EU Science, Research and Innovation
@EU_ScienceHub
EU Science
EU Science Hub - Joint Research Centre