A data-centric platform for scientific support to EU policies

Chiara Chiarelli, Davide De Marchi, Ilias Ioannou, Pieter Kempeneers, Tomas Kliment, Pierre Soille, **Roberto Ugolotti**

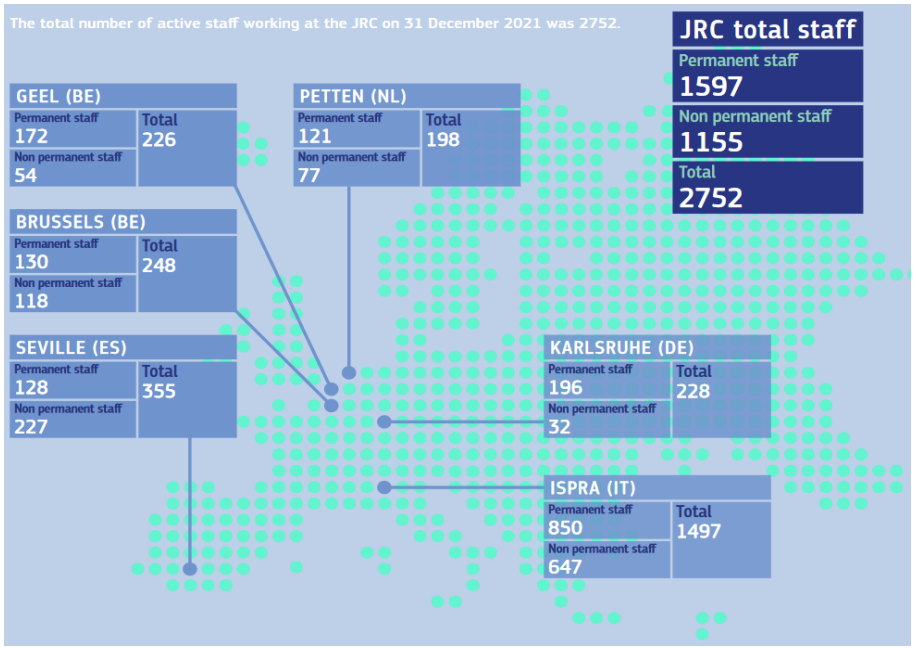
Joint Research Centre

Geneve, 04/05/2023

European Commission

Joint Research Centre

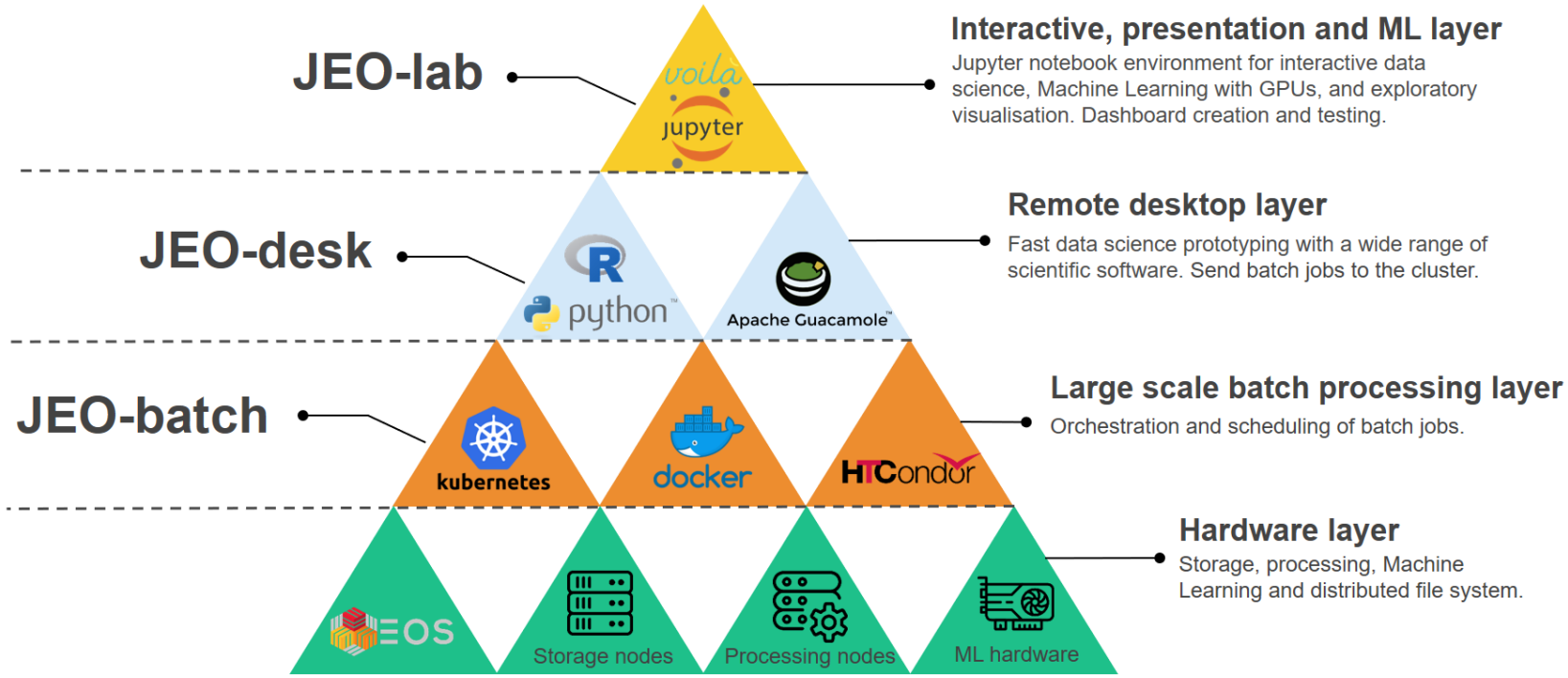
The role of the Joint Research Centre (JRC) of the European Commission is to provide independent, evidence-based science and knowledge, supporting European Union (EU) policies to positively impact society.



European Commission

Big Data Analytics Platform

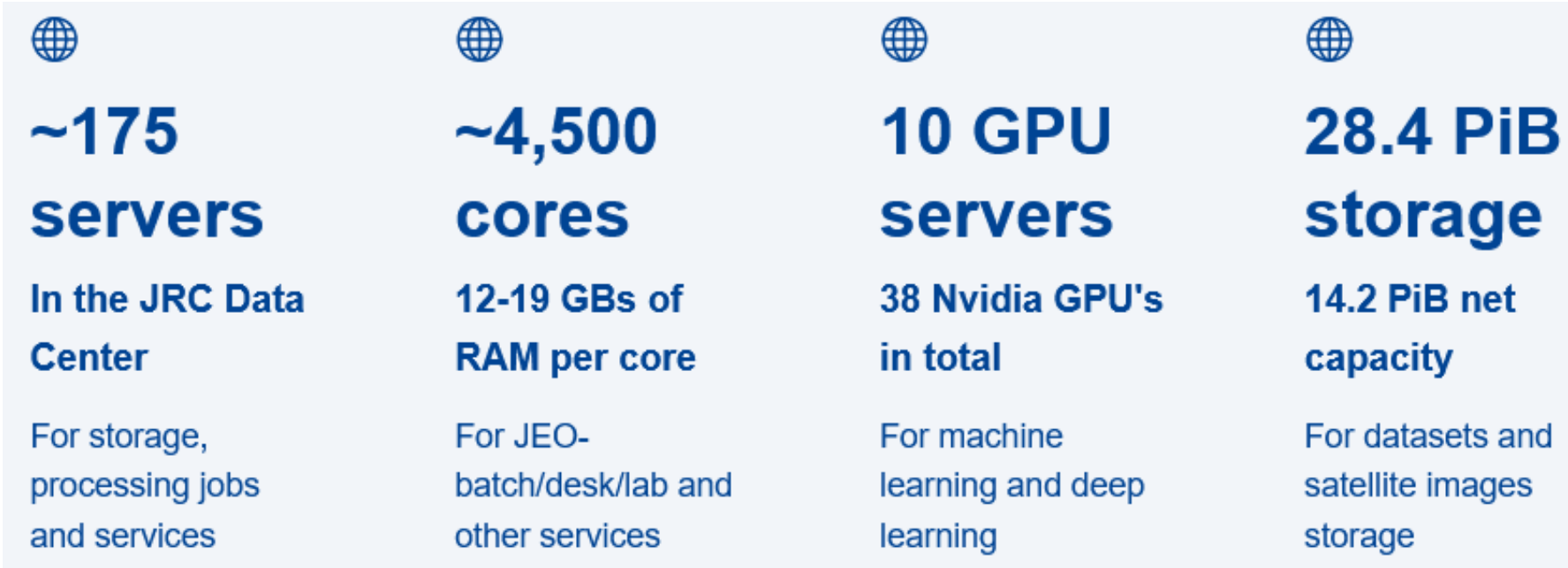
The JRC Big Data Analytics Platform is a cloud platform that provides data, environments, and software to JRC researchers, all in a single place.



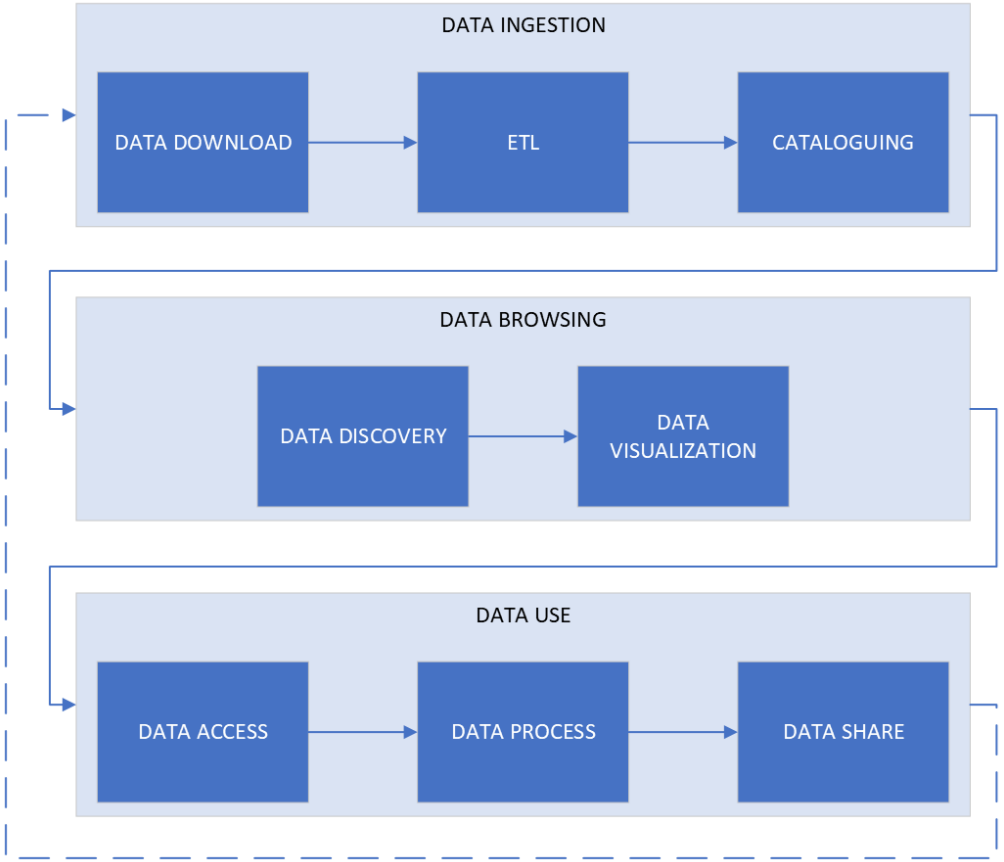
European Commission

Big Data Analytics Platform

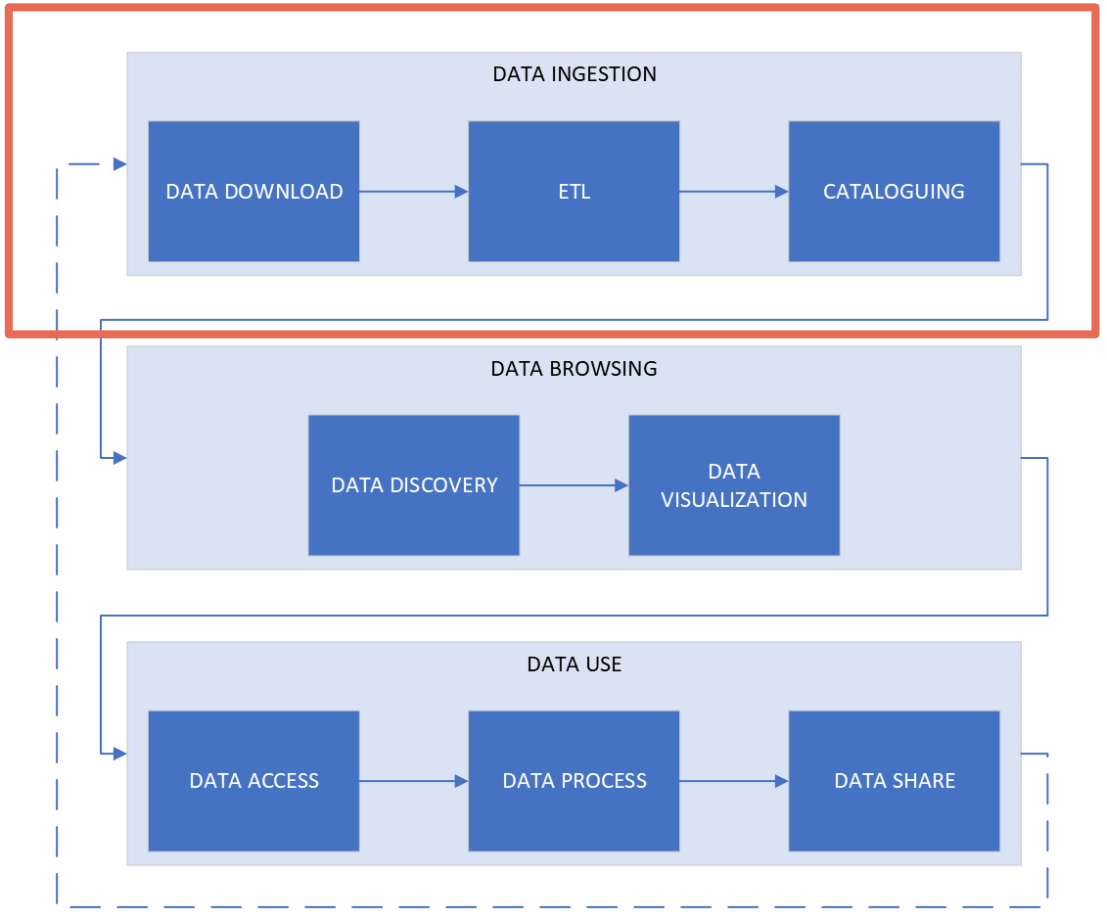
The JRC Big Data Analytics Platform is a cloud platform that provides data, environments, and software to JRC researchers, all in a single place.



A Data Story in BDAP



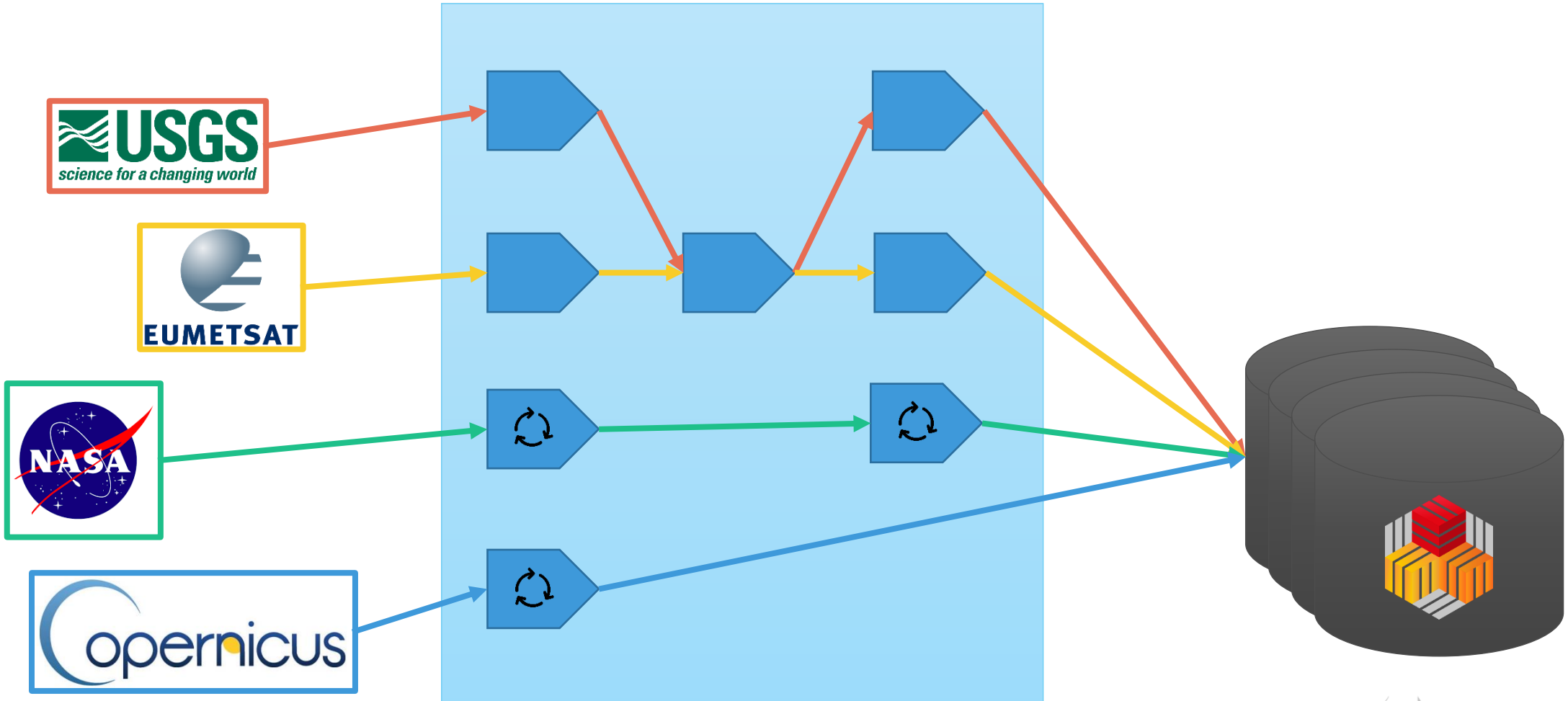
A Data Story in BDAP – Data Ingestion



Some numbers

- > 400 collections
- > 100 providers
- > 20 domains
- > 20 million of metadata records
- > 10 PB of data

Data Download

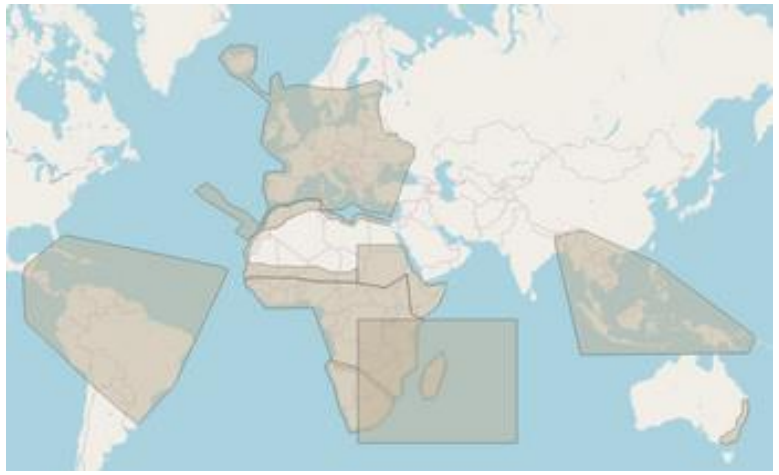


European Commission

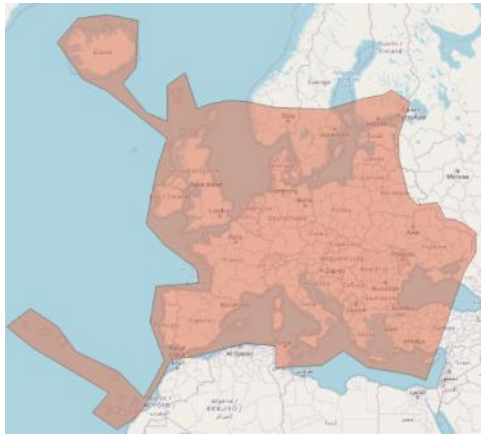
Main Systematic Downloads

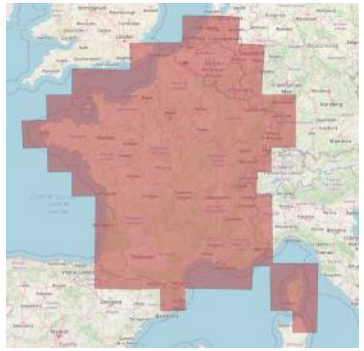
Sentinel-1 GRD

Sentinel-1 SLC



Sentinel-2 L1C and L2A





NASA HLS L30 and S30

Sentinel-3 OL2LRR + OL2WRR and Sentinel-5P: global

European Commission

Quality Checks

Completeness

Accuracy

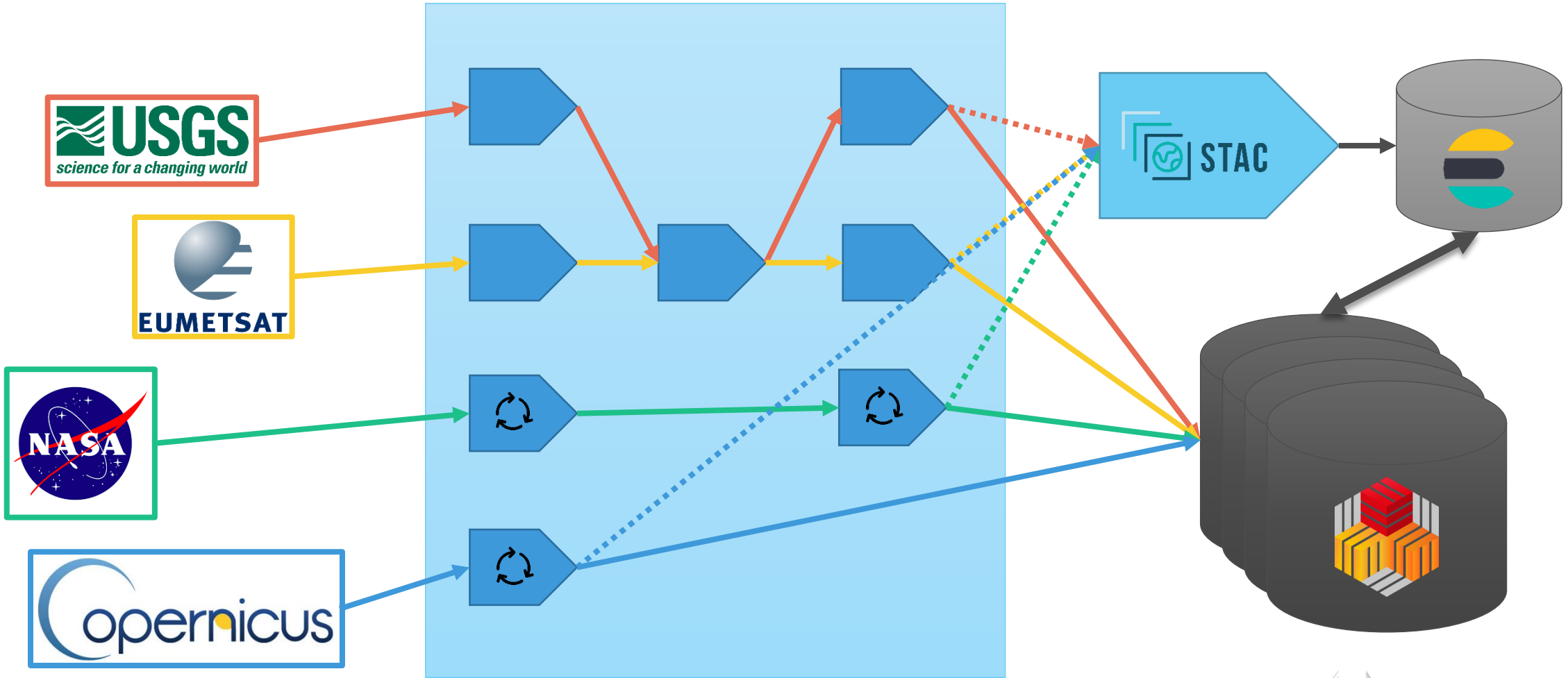
Consistency

Validity

Uniqueness

Integrity

Cataloguing



European Commission

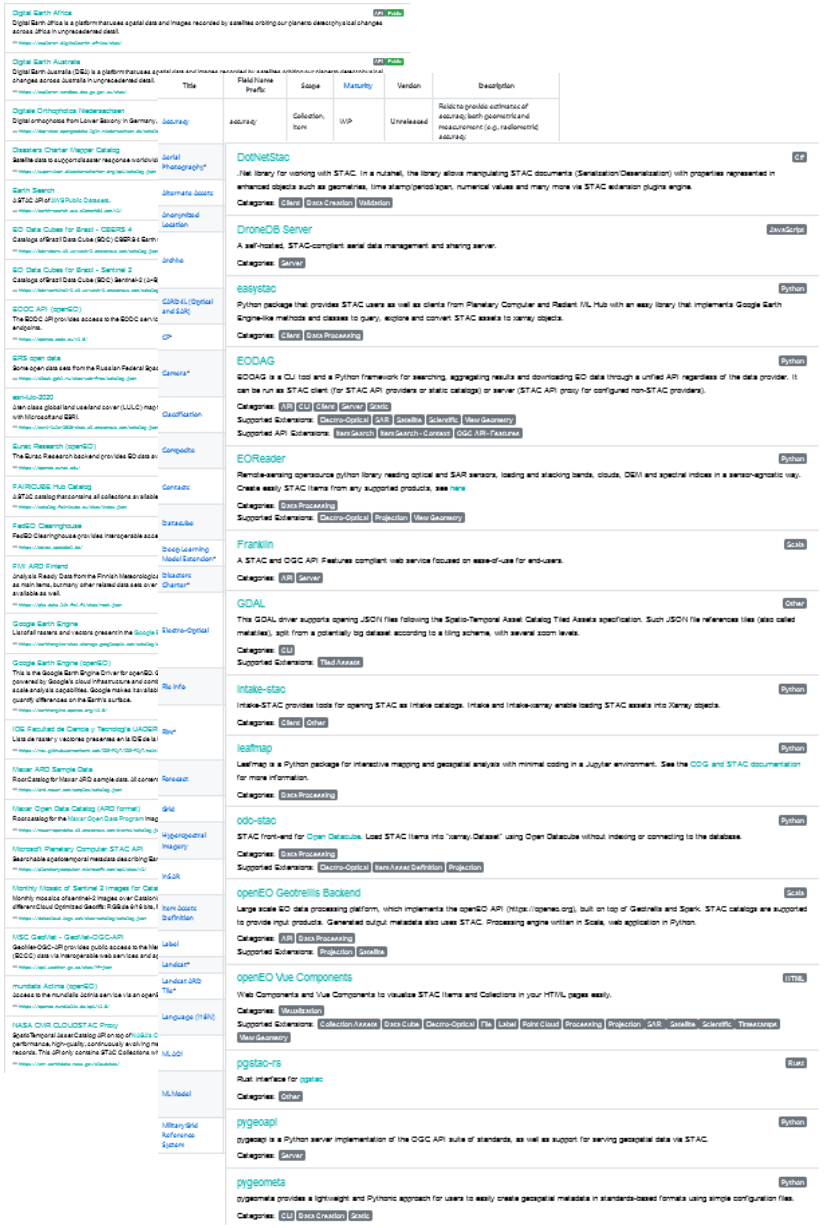
Cataloguing

Each download is associated with a metadata entry

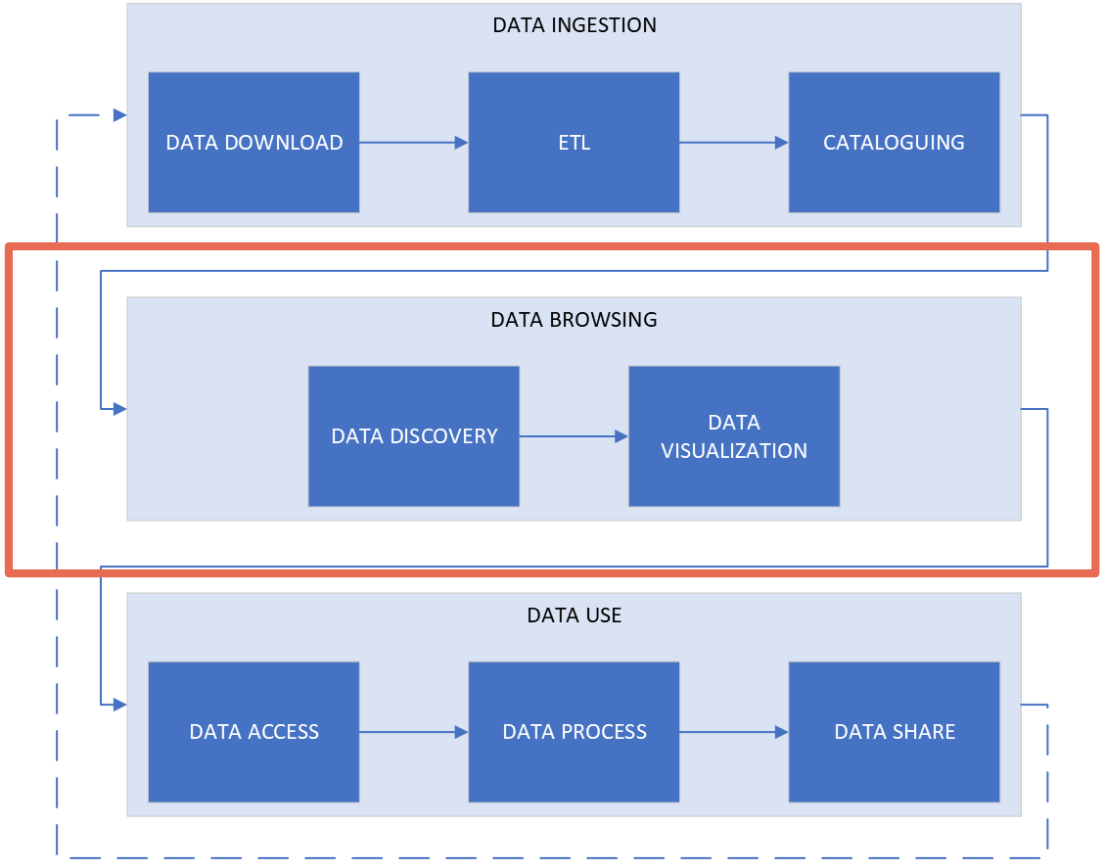
STAC (SpatioTemporal Asset Catalogs) format

- Widely accepted and used by OS community
- Geospatial Information (but not limited to)
- Modular/Easy to extend
- Ecosystem of libraries, tools, and clients

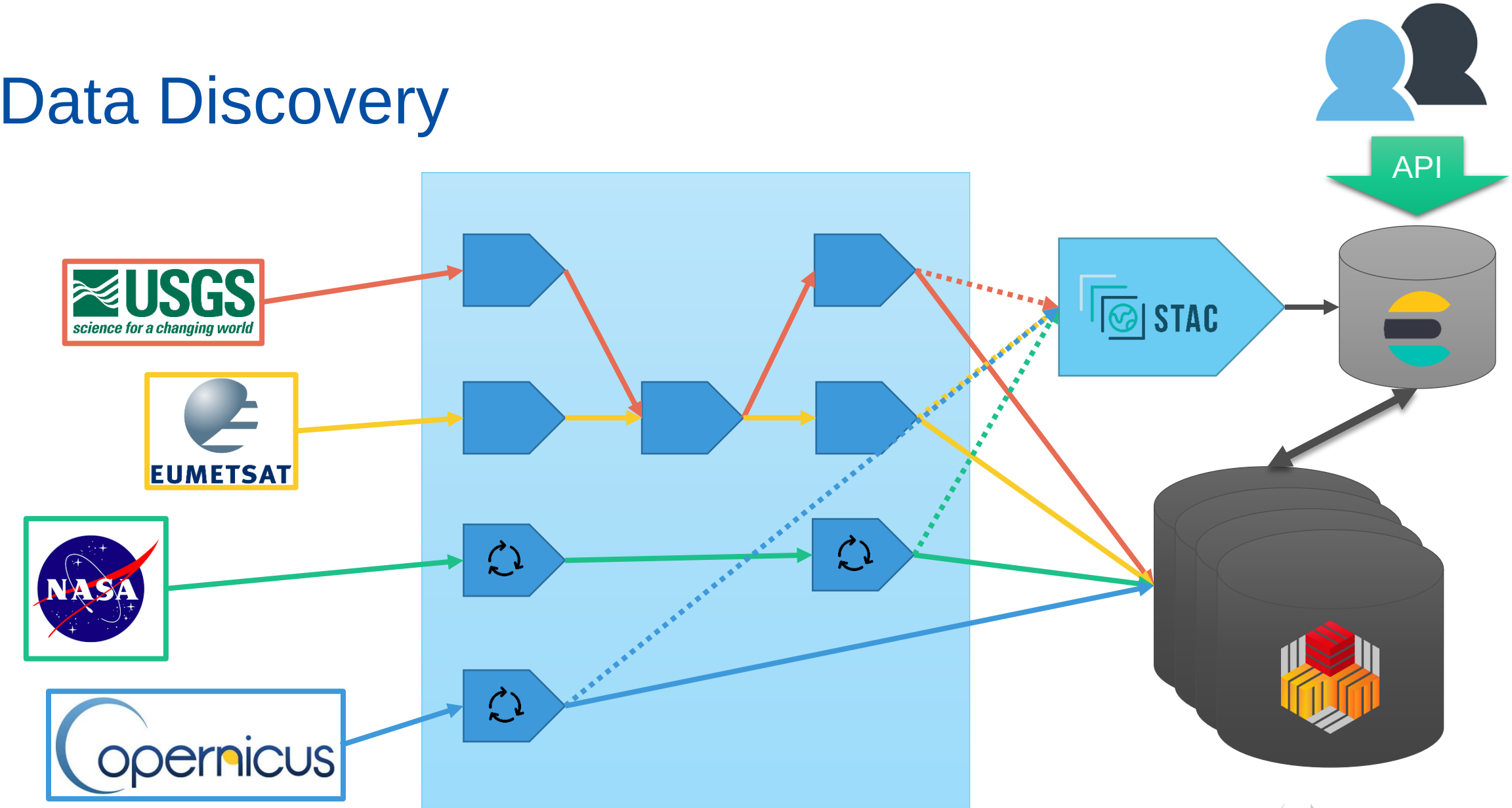
Elasticsearch



A Data Story in BDAP – Data Browsing

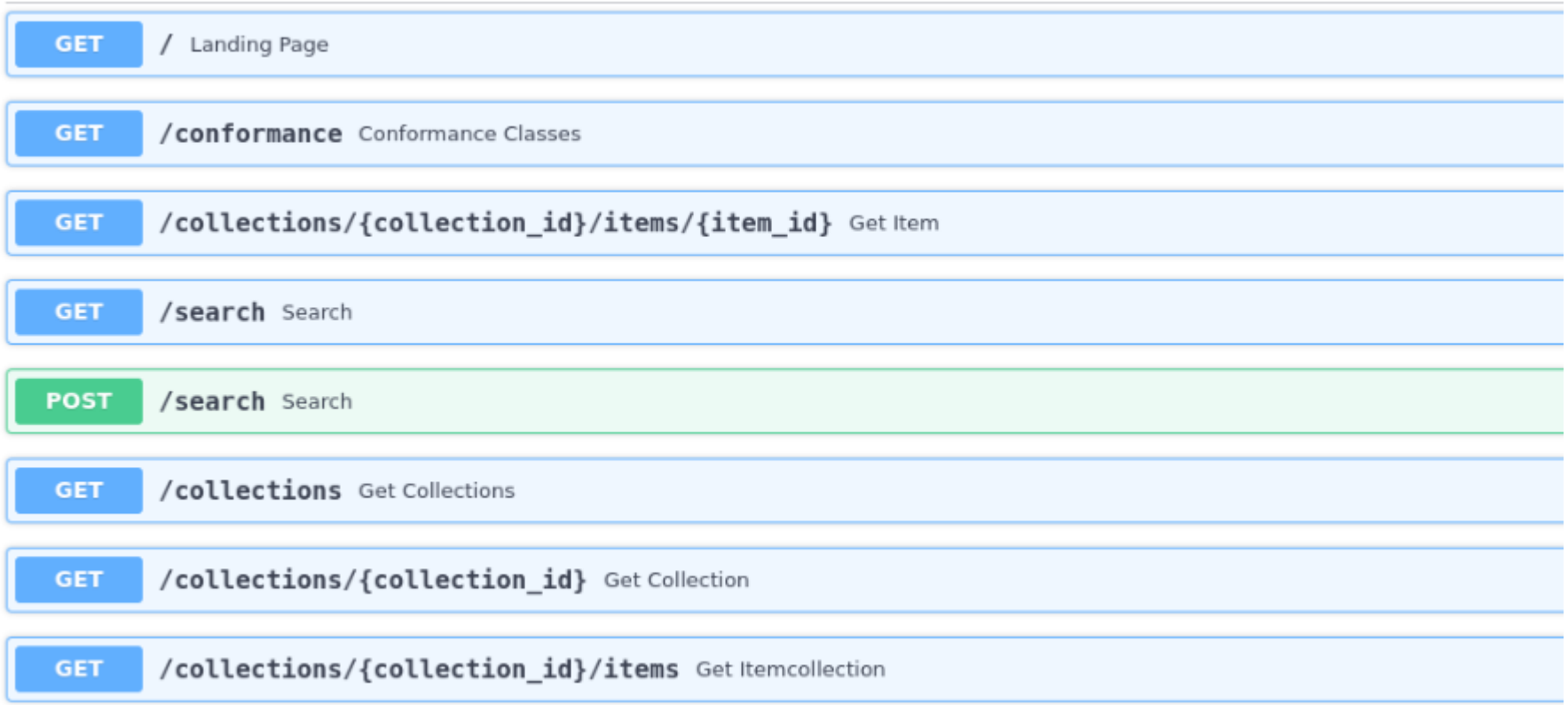


Data Discovery



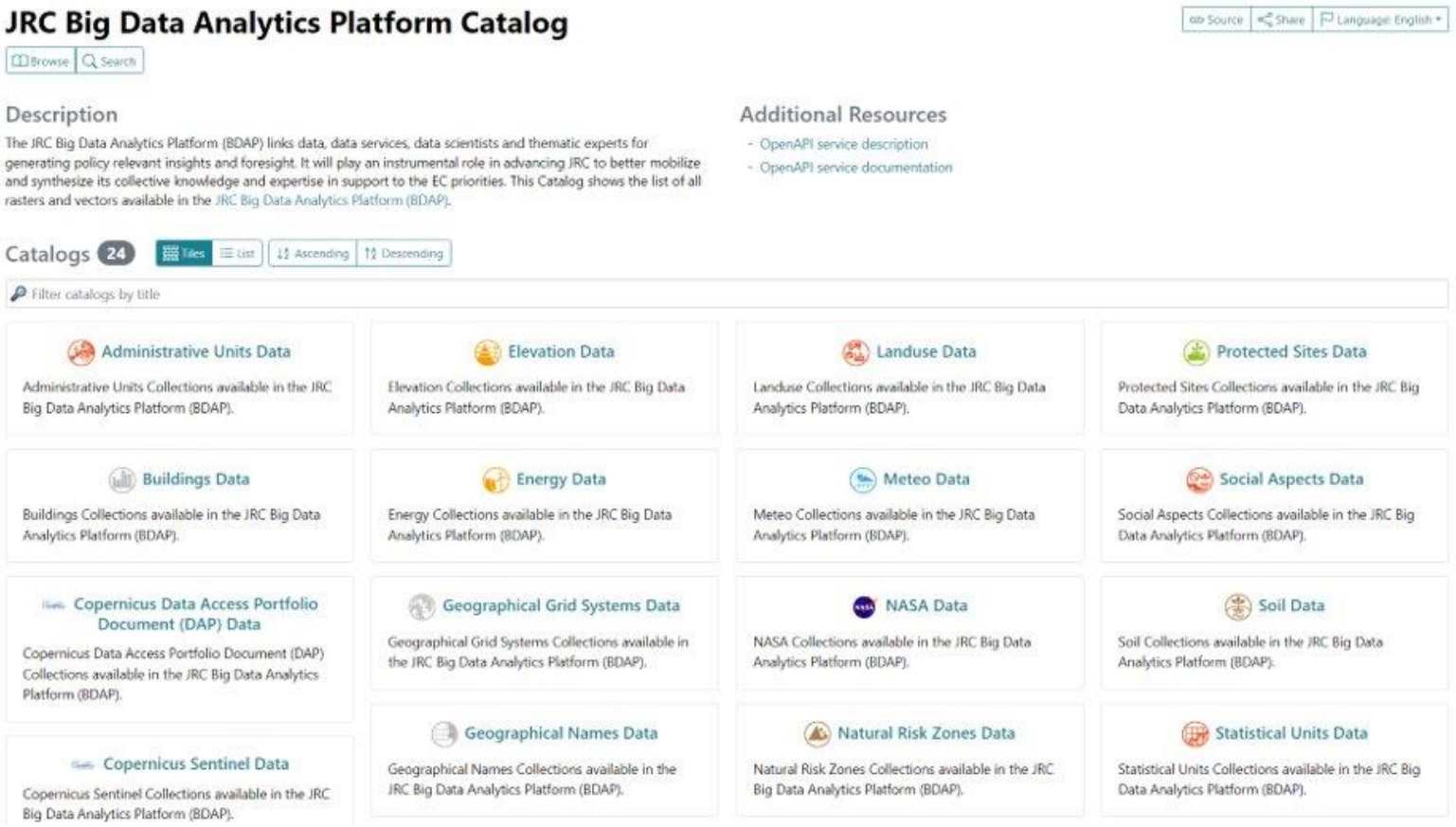
European Commission

REST APIs



Following **STAC API V1.0.0 RC.1** specifications.

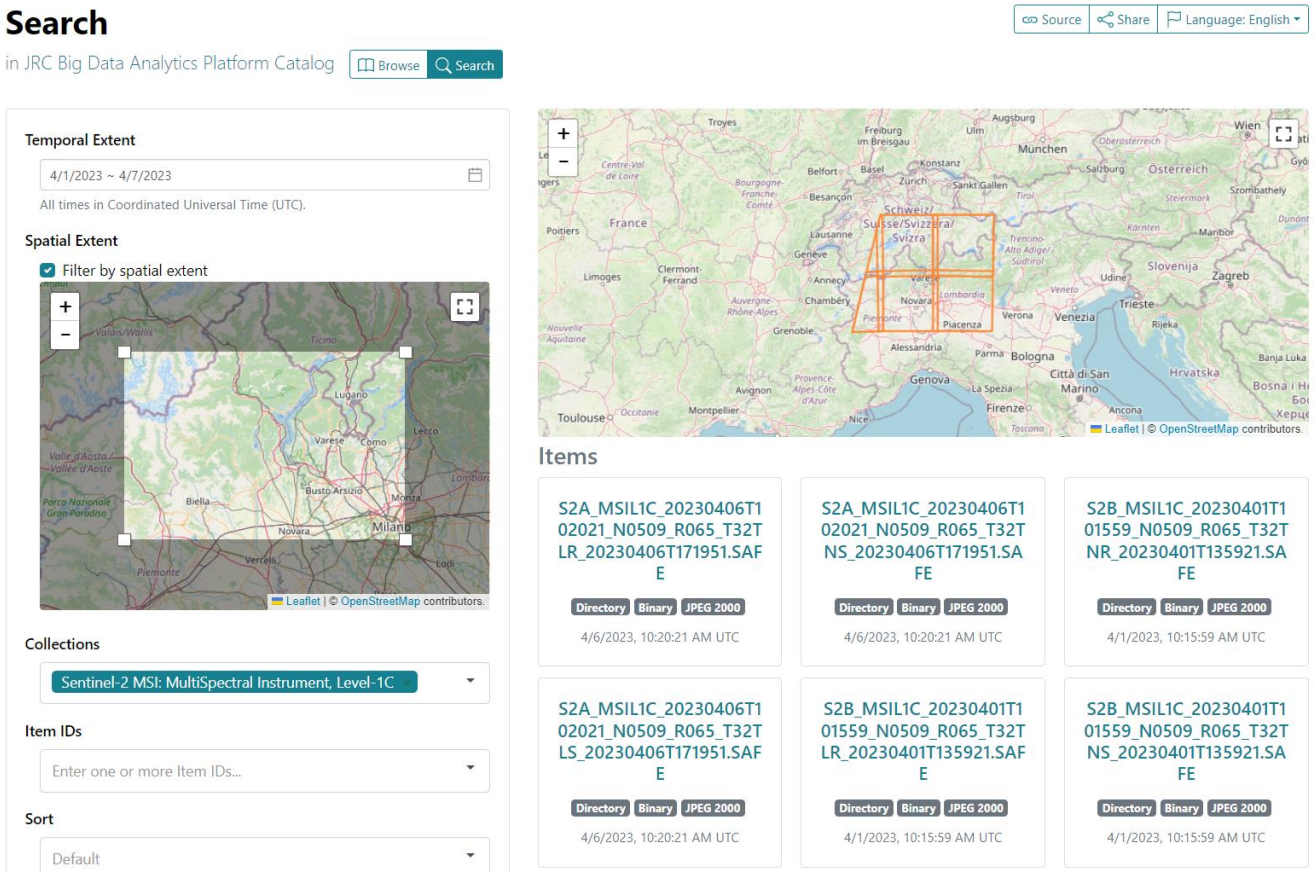
BDAP Data Catalog



JRC Big Data Analytics Platform Catalog

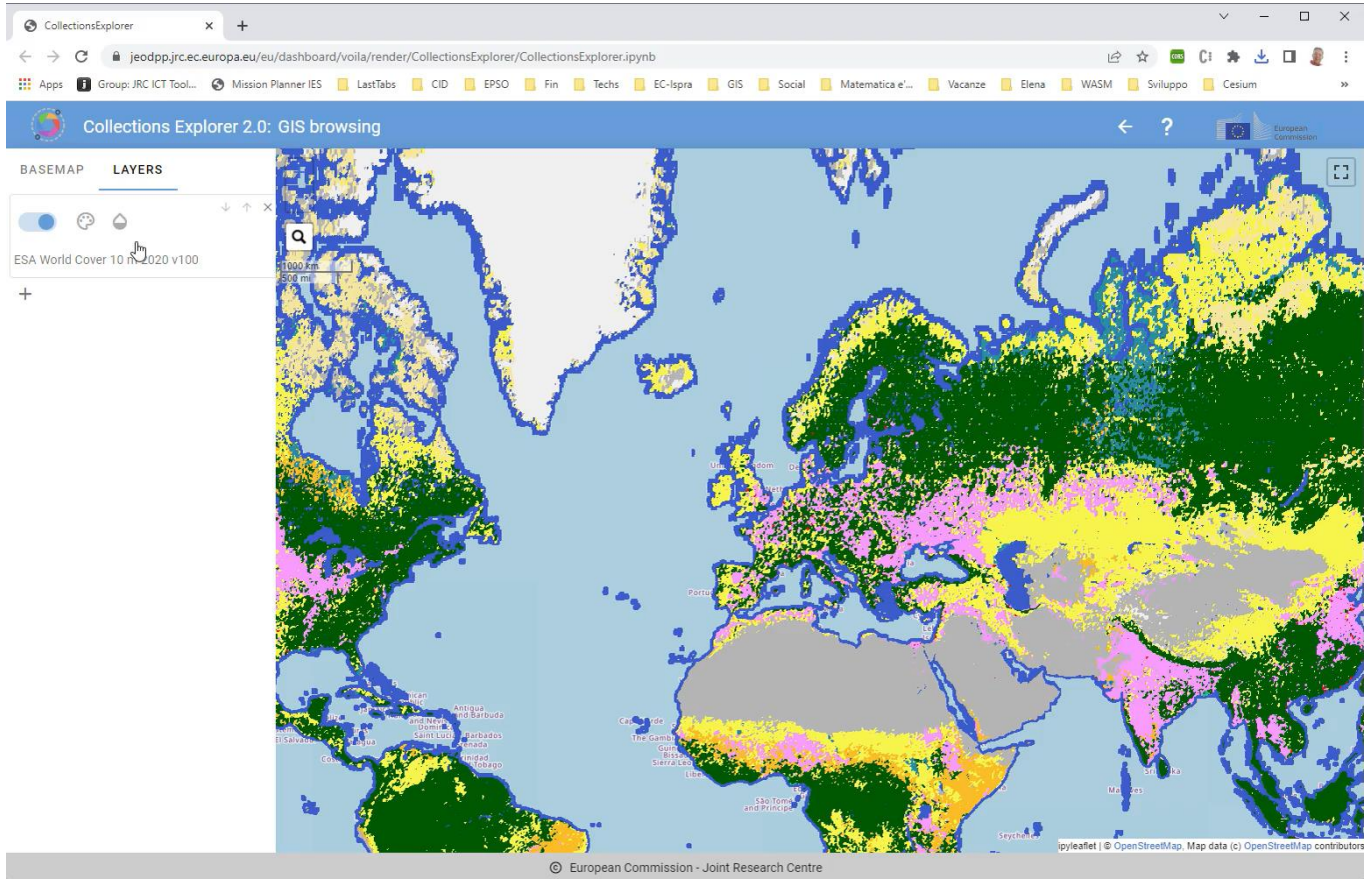
European Commission

BDAP Data Catalog - Search



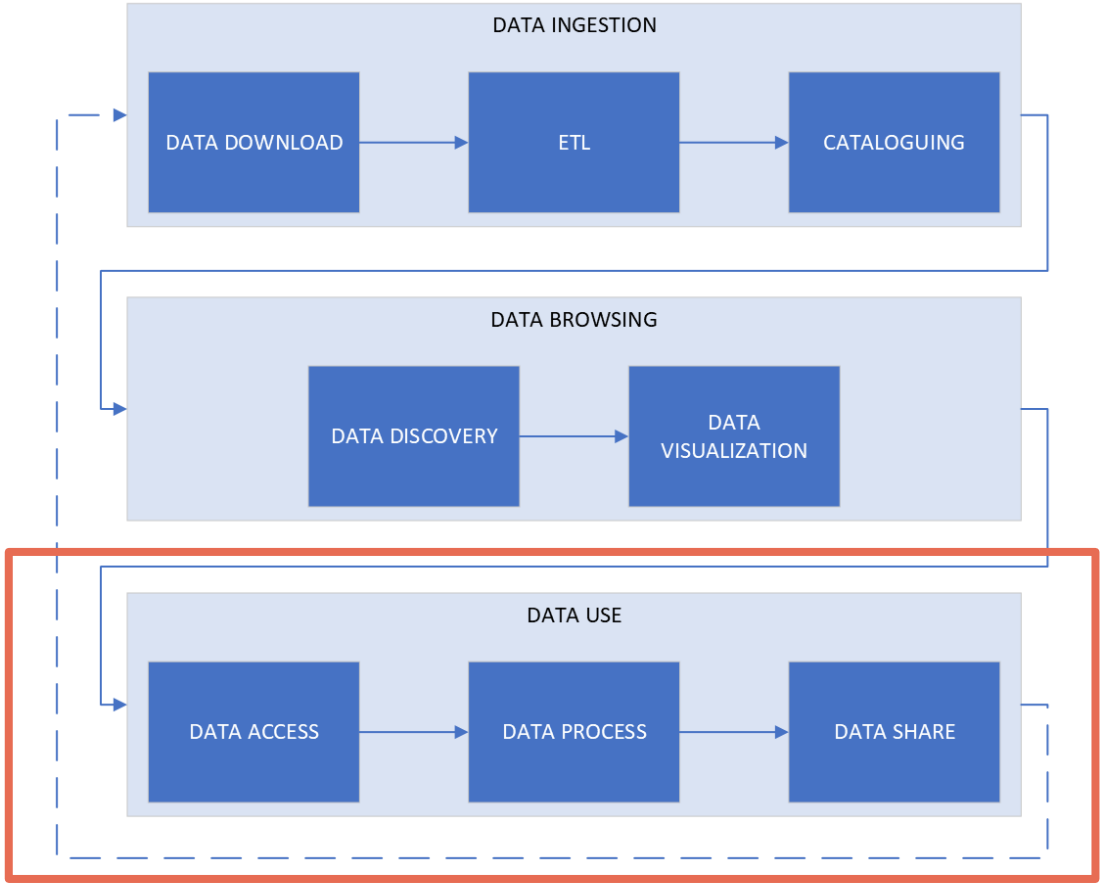
European Commission

BDAP Collections Explorer

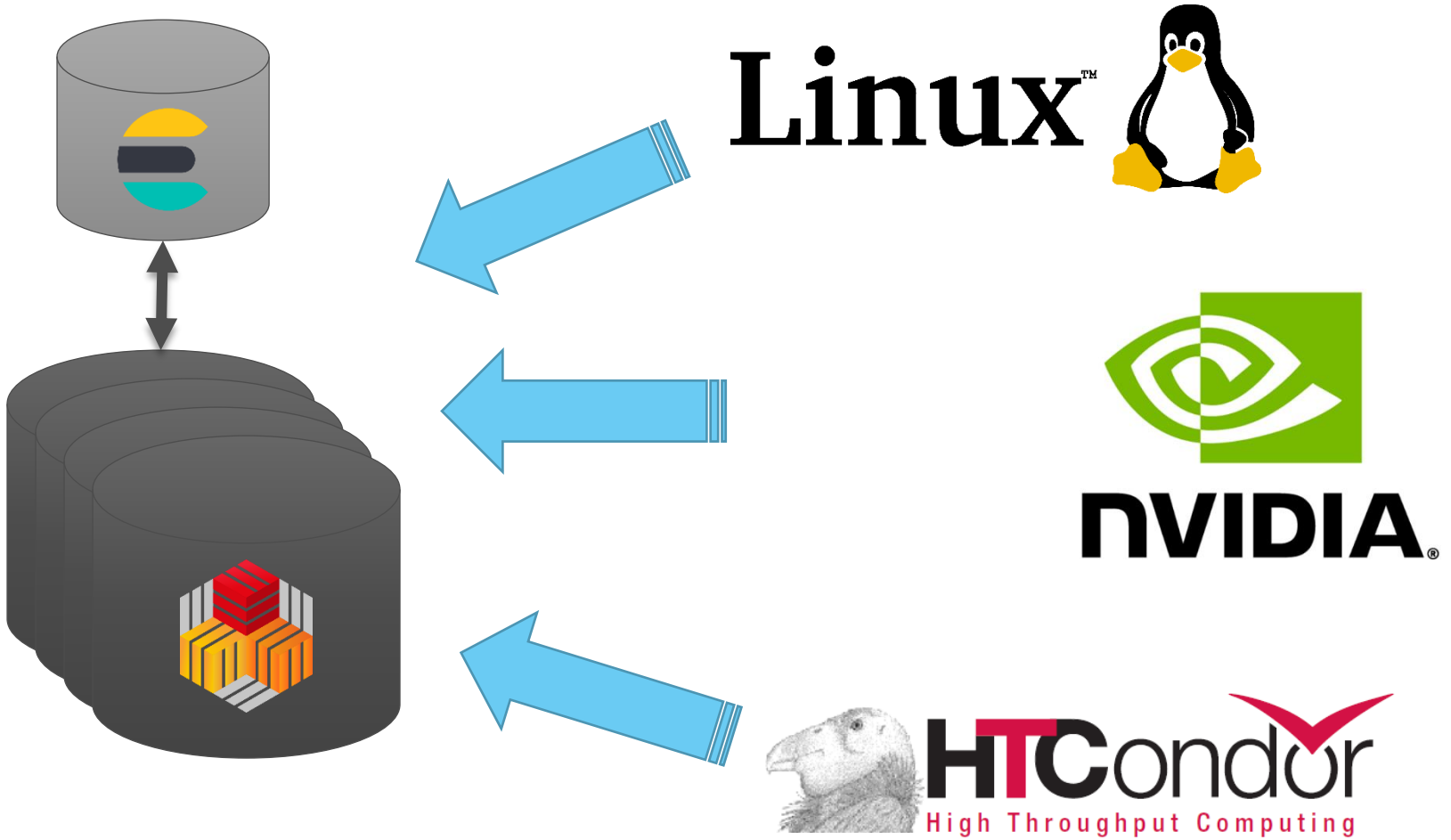


European Commission

A Data Story in BDAP – Data Use



BDAP Development Environment



BDAP software landscape

Operating systems & virtualization

debian
CentOS
Xen Project
KVM
oVirt
docker
HARBOR
kubernetes
RANCHER
HELM

Storage & file sharing

EOS
MINIO
NetApp
Nextcloud
FileZilla
DAV

Web-desktop for data science

GUACAMOLE
xrdp
ubuntu
XFCE
Linux Unified Key Setup

Data science languages & IDEs

python
PyCharm
Spyder
R
RStudio
julia
Java
eclipse
GNU Octave
MATLAB
IDL
jupyter
lab
SWAN
voila
vois library
ATOM
Open for Innovation KNIME

Monitoring & Automation

Nagios
Ganglia
Grafana
FOREMAN
ANSIBLE
GitLab

Search engine & web frameworks/servers

OpenSearch
Gunicorn
NGINX
FastAPI
MapServer

AI & Machine Learning

NVIDIA
CUDA
RAPIDS
mxnet
TensorFlow
PyTorch
Keras
scikit learn
CONDA
CLEARML

HTC/HPC & Workflow engines

HTCondor
DASK
Open MPI
argo

Database engines, clients, and tools

PostgreSQL
PostGIS
Spatial PostgreSQL
mongoDB
DBeaver
redis
SQLite
talend
Elasticsearch
qdrant

Geospatial software applications

QGIS
ORFEO ToolBox
SNAP
GTB
GRASS GIS

Conclusions – How BDAP adds value to data?

- Combine different sources of data
- Quality checks on data
- Based on open-source and widely used standards
- Integrated environment + data

A data-centric platform for scientific support to EU policies

Chiara Chiarelli, Davide De Marchi, Ilias Ioannou, Pieter Kempeneers, Tomas Kliment, Pierre Soille, **Roberto Ugolotti**

Joint Research Centre

Geneve, 04/05/2023

European Commission