DLR’s terrabyte platform
An EO high performance data analytics platform

Cooperation with Leibniz Supercomputing Center (LRZ)

- 16 Million Euro costs for hardware (storage, compute)
- **LRZ** provides in-kind contribution for setup, maintenance, and operations of infrastructure and compute services
- **DLR** provides EO and geospatial data, services for data discovery & analysis, and user support
- **Accessible** to all DLR and LRZ users as well as project partners and collaborators

Highlights of terrabyte

- Data is described with **STAC** metadata and provided with STAC API
- **Analysis Ready Data** are generated for data archives
- Ingestion of global Sentinel, Landsat, and MODIS data
- Virtual **Data Cubes** based on STAC, xarray, and Dask
- **Interactive applications** in the browser (e.g., JupyterLab, QGIS)

Compute
44.000 vCPUs
188 GPUs
333 TB RAM

Online storage
50 Petabyte (net)

lrz
Leibniz Supercomputing Centre
of the Bavarian Academy of Sciences and Humanities

DLR’s terrabyte platform

Dr. Jonas Eberle, Max Schwinger (Project leads)

CEOS WGISS-56 – 24.10.2023

Activities of DLR‘s German Remote Sensing Data Center

EO payload ground segment

as part of DLR‘s space infrastructure

Implementation of national missions
Copernicus and int. mission participation
Services and industry support

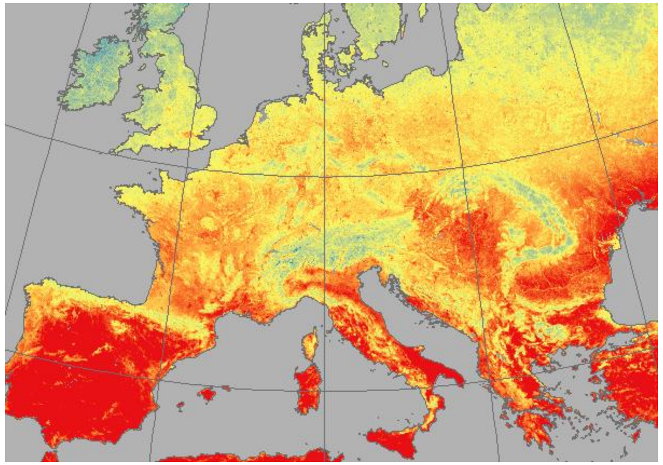
EO informatics

for research infrastructures

Efficient big data processing,
EO exploitation platforms,
Modern ground segments

Research with EO data

on relevant social topics



Knowledge gain, products, transfer and neutral political advice

Activities of DLR‘s German Remote Sensing Data Center

EO payload ground segment
as part of DLR‘s space infrastructure

EO informatics
for research infrastructures

Sentinel-5p PDGS

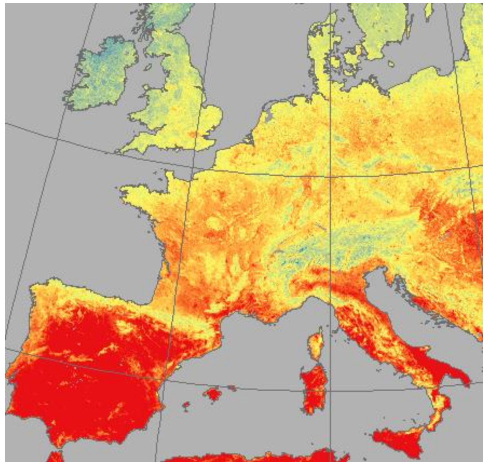
eoc geo service

eoc EOWEB® GeoPortal

DSDA
German
Satellite Data Archive

terrabyte

Research with EO data
on relevant social topics



terrabyte
Overview

An EO high performance data analytics platform
Cooperation with Leibniz Supercomputing Center (LRZ)

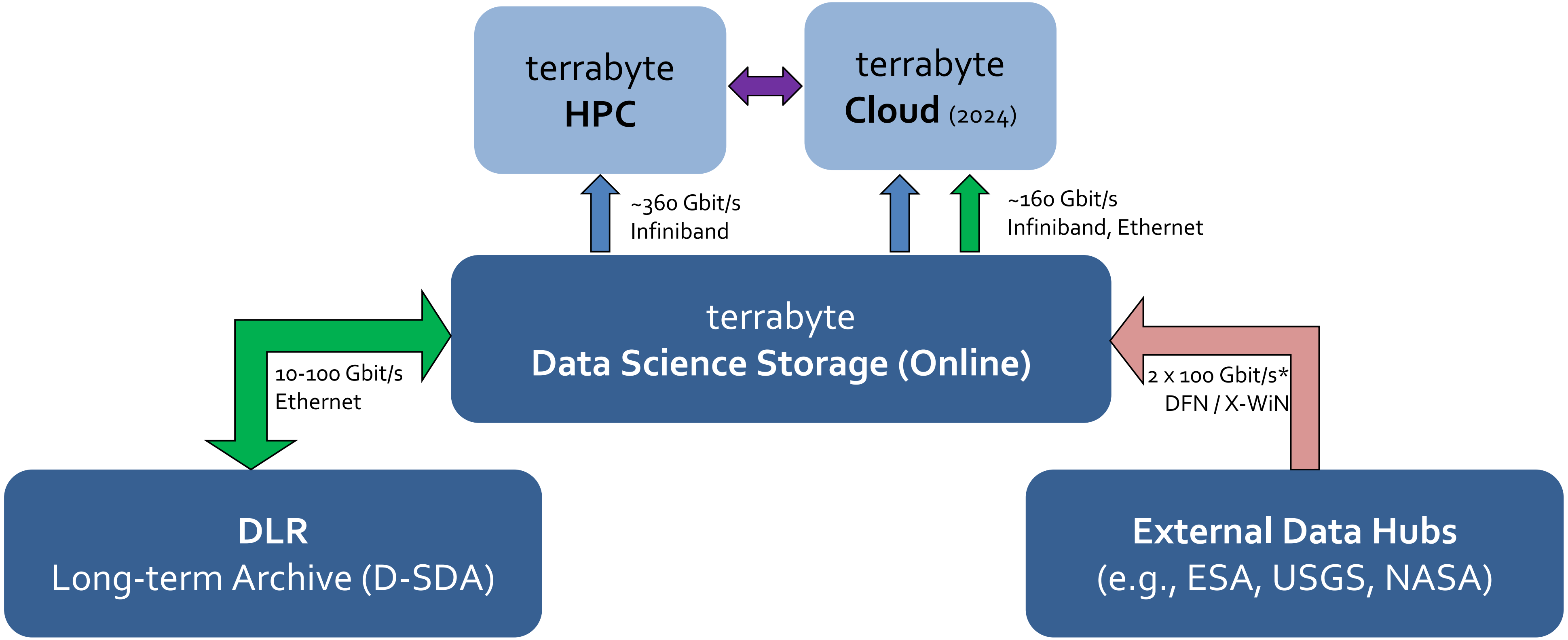
- 16 Million Euro costs for hardware (storage, compute)
- **LRZ** provides in-kind contribution for setup, maintenance, and operations of infrastructure and compute services
- **DLR** provides EO and geospatial data, services for data discovery & analysis, and user support
- **Accessible** to all DLR and LRZ users as well as project partners and collaborators

Compute
44.000 vCPUs
188 GPUs
333 TB RAM

Online storage
50 Petabyte (net)

lrz
Leibniz Supercomputing Centre
of the Bavarian Academy of Sciences and Humanities

terrabyte Infrastructure
A hybrid environment for compute and services



terrabyte HPC
terrabyte Cloud (2024)
~360 Gbit/s Infiniband
~160 Gbit/s Infiniband, Ethernet
terrabyte Data Science Storage (Online)
10-100 Gbit/s Ethernet
2 x 100 Gbit/s* DFN / X-WiN
DLR Long-term Archive (D-SDA)
External Data Hubs (e.g., ESA, USGS, NASA)

* Limited to the capabilities of the external data hubs and shared with other users at LRZ!

terrabyte Services

Software stack for base services

terrabyte DLR

Metadata catalogue

eoAPI

STAC

Interoperable services

Open Geospatial Consortium

openEO

Compute Interface

Open OnDemand

jupyter

QGIS

Workflow engine

CAMUNDA

kubernetes

docker

Charliecloud

slurm workload manager

Base computing resources

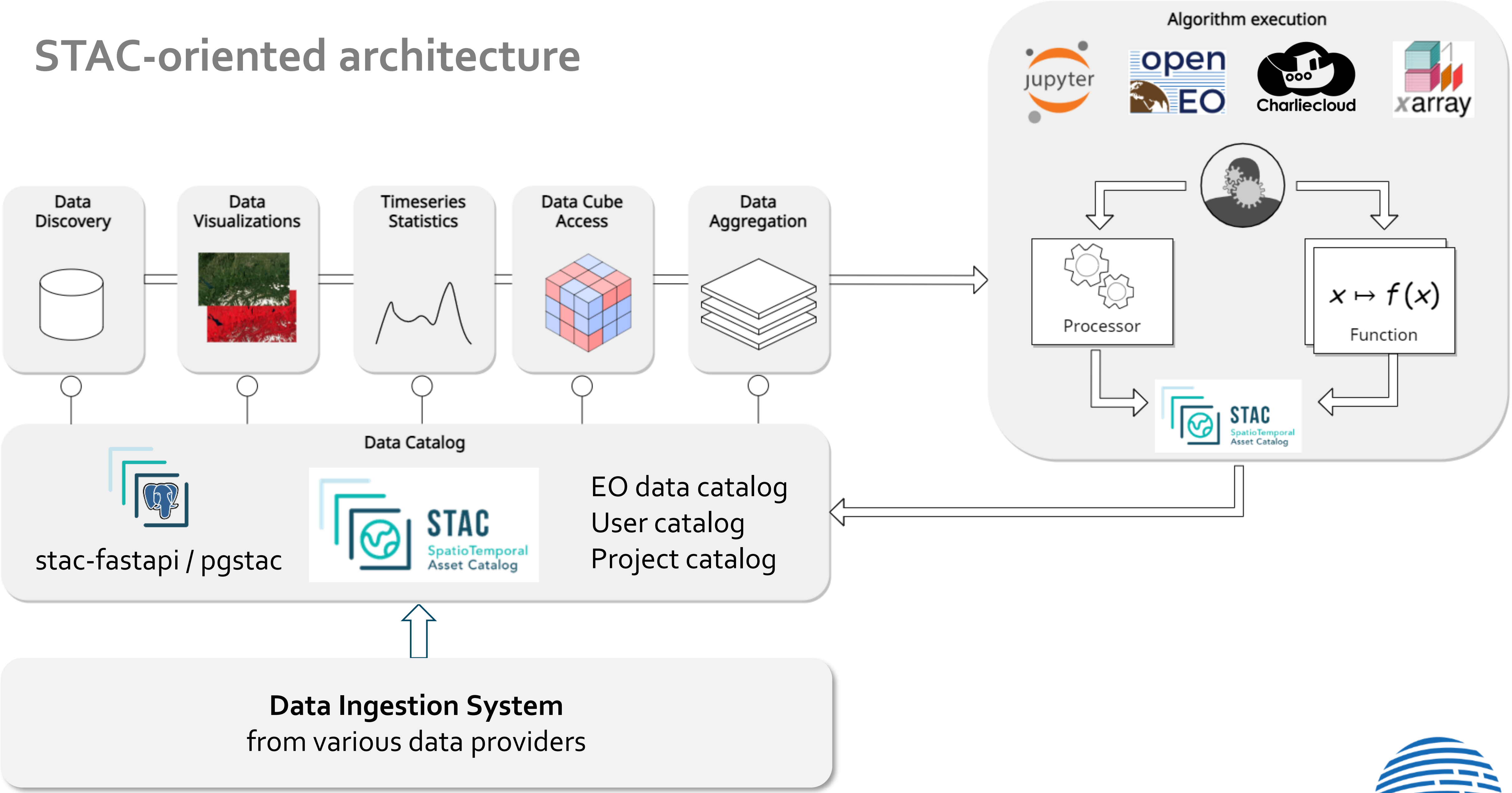
KEYCLOAK

Discourse

User authentication & user support

terrabyte
Data Discovery & Analysis

STAC-oriented architecture



Algorithm execution

jupyter
open EO
Charliecloud
xarray

Processor
$x \mapsto f(x)$
Function
STAC SpatioTemporal Asset Catalog

Data Discovery

Data Visualizations

Timeseries Statistics

Data Cube Access

Data Aggregation

Data Catalog

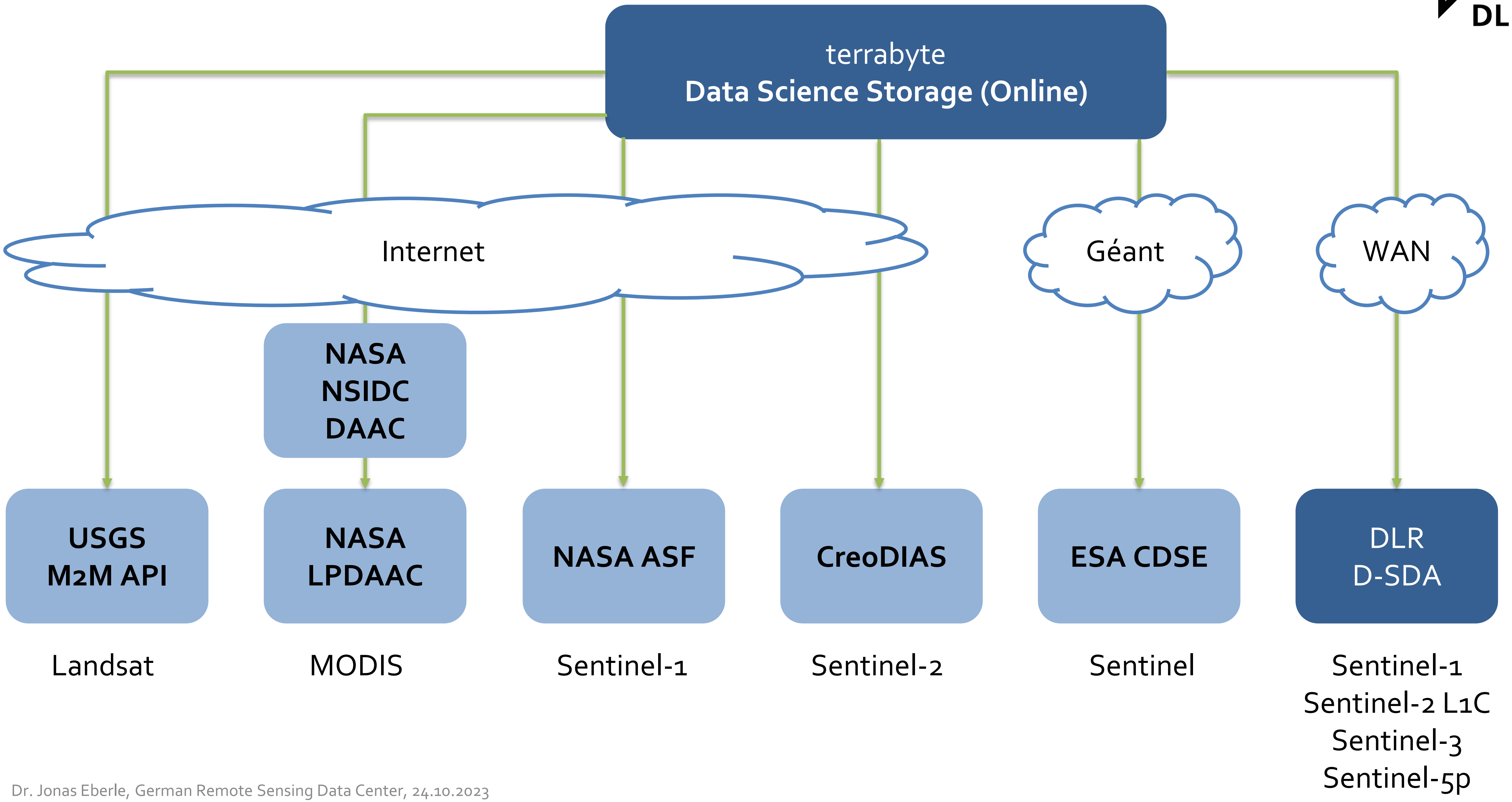
stac-fastapi / pgstac

STAC SpatioTemporal Asset Catalog

EO data catalog
User catalog
Project catalog

Data Ingestion System
from various data providers

Data ingestion


terrabyte
Data Science Storage (Online)

Internet

Géant

WAN

NASA
NSIDC
DAAC

USGS
M2M API

NASA
LPDAAC

NASA ASF

CreoDIAS

ESA CDSE

DLR
D-SDA

Landsat

MODIS

Sentinel-1

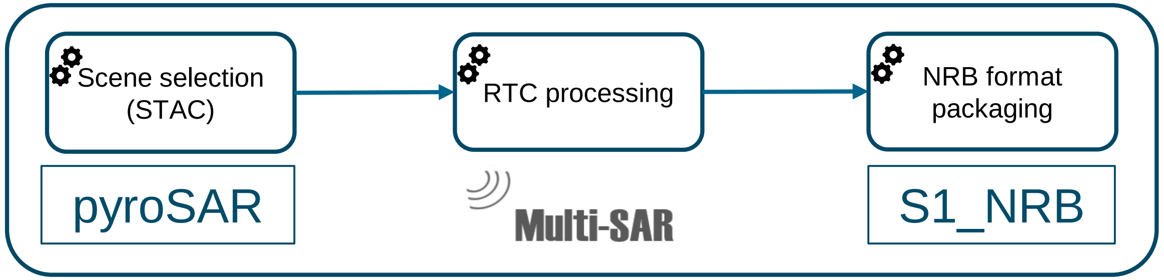
Sentinel-2

Sentinel

Sentinel-1
Sentinel-2 L1C
Sentinel-3
Sentinel-5p

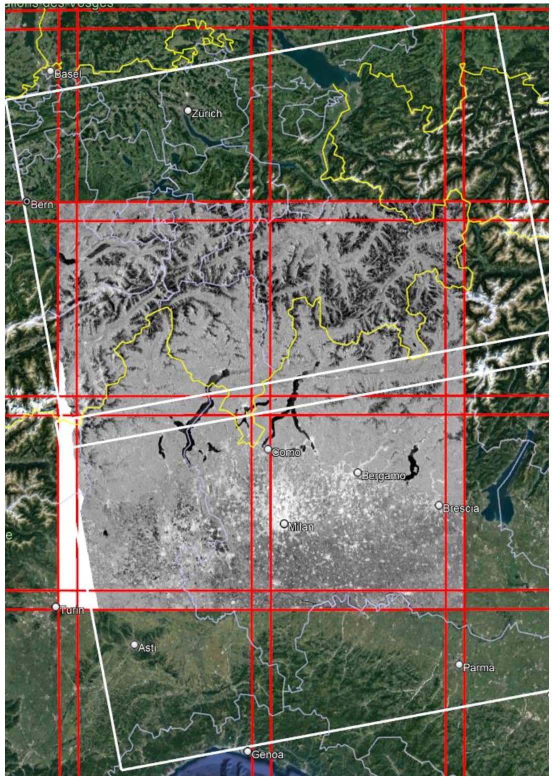
Generating and providing Analysis Ready Data

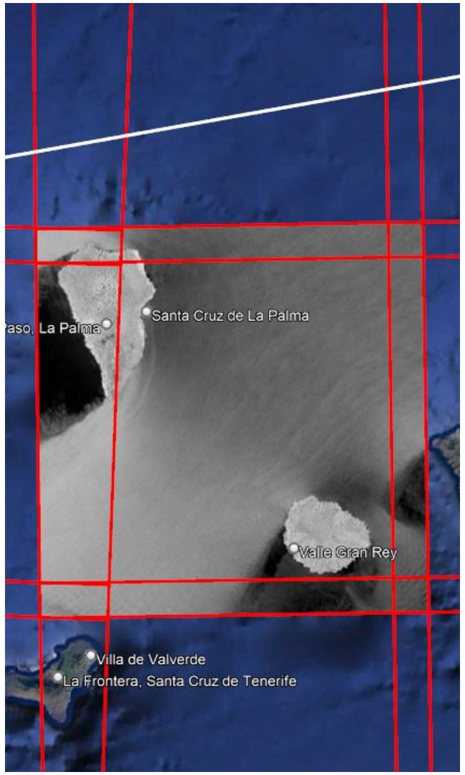
- **Sentinel-1 Normalized Radar Backscatter**



  - Complete archive based on Sentinel-1 SLC
  - Based on CEOS ARD NRB specification
  - Reprocessing from Level-0 to Level-1 useful
  - Conduct processing as “background missions”

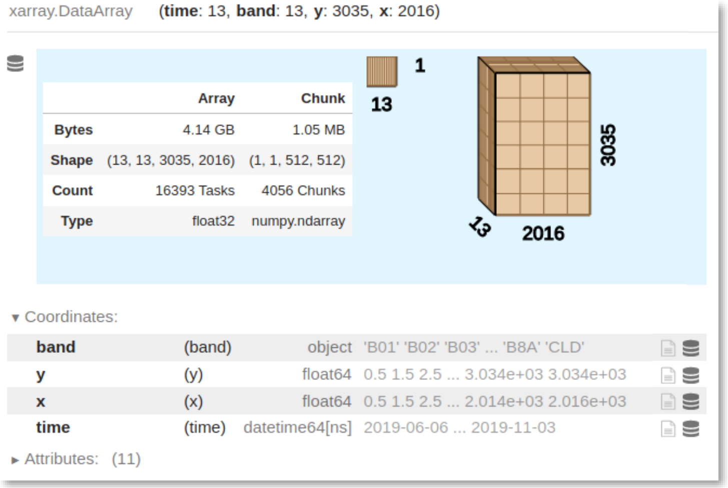
- **Sentinel-2 L2A MAJA (DLR-IMF)**
  - Currently: Europe 2018 – 2022





terrabyte Data Cube
based on STAC, xarray, Pandas, and Dask

- No pre-defined data cubes → „virtual data cubes"
- Build on top of Open Data Cube STAC (odc-stac)
- Usage of terrabyte HPC out of the box
- Initial use case: World Settlement Footprint



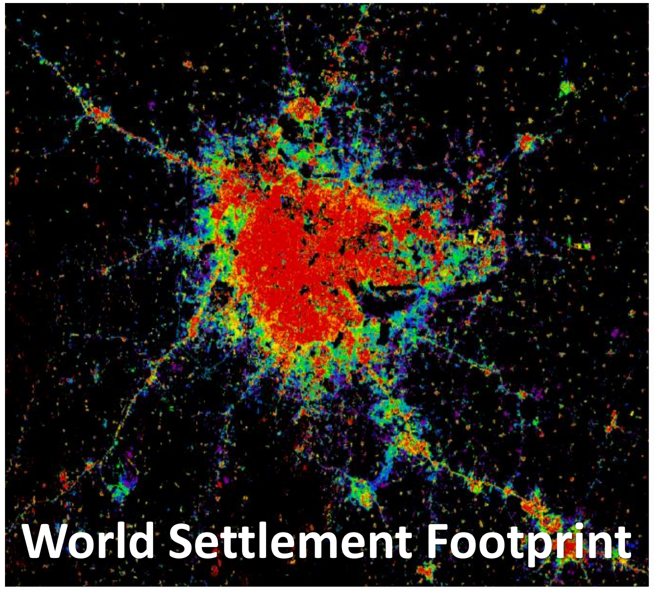
```
xarray.DataArray    (time: 13, band: 13, y: 3035, x: 2016)
```

	Array	Chunk
Bytes	4.14 GB	1.05 MB
Shape	(13, 13, 3035, 2016)	(1, 1, 512, 512)
Count	16393 Tasks	4056 Chunks
Type	float32	numpy.ndarray

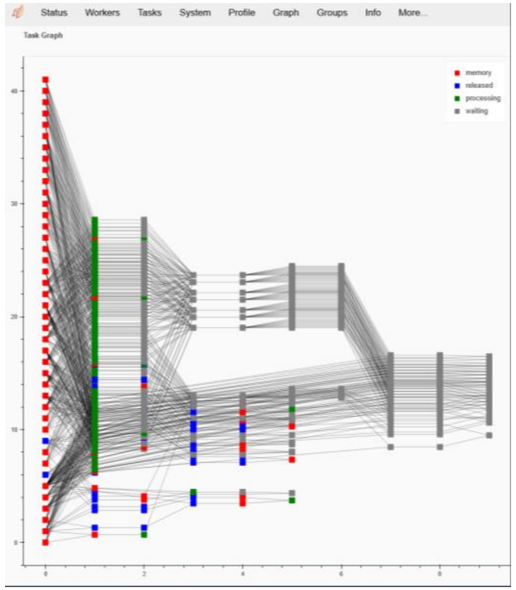
```
▾ Coordinates:
   band     (band)           object  'B01' 'B02' 'B03' ... 'B8A' 'CLD'
   y        (y)             float64  0.5 1.5 2.5 ... 3.034e+03 3.034e+03
   x        (x)             float64  0.5 1.5 2.5 ... 2.014e+03 2.016e+03
   time     (time)   datetime64[ns]  2019-06-06 ... 2019-11-03
▸ Attributes:  (11)
```

OpenEO

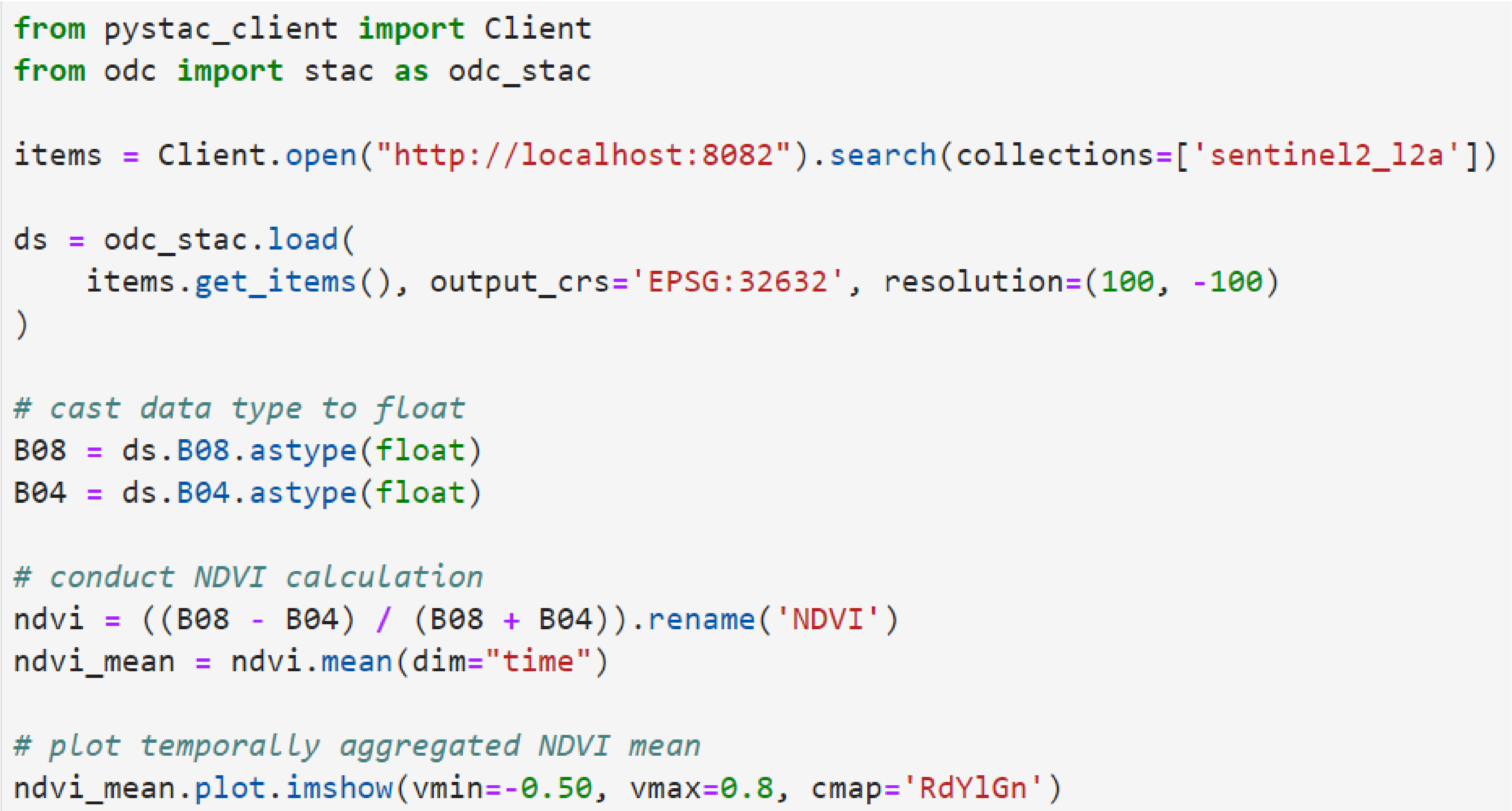
- Backend in development (in exchange with EODC, Eurac Research, IBM Research)
- Based on terrabyte Data Cube



World Settlement Footprint



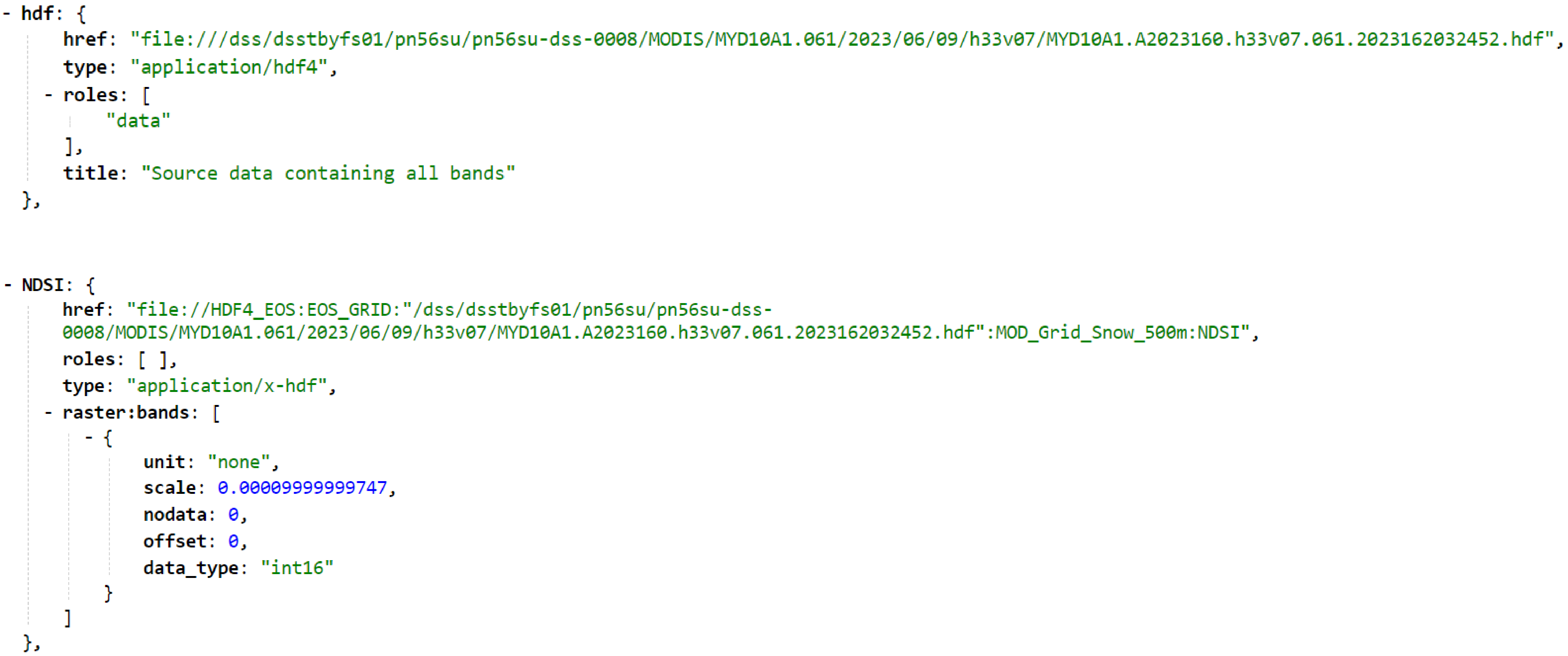
terrabyte Data Cube
based on STAC, xarray, Pandas, and Dask



```python
from pystac_client import Client
from odc import stac as odc_stac

items = Client.open("http://localhost:8082").search(collections=['sentinel2_l2a'])

ds = odc_stac.load(
    items.get_items(), output_crs='EPSG:32632', resolution=(100, -100)
)

# cast data type to float
B08 = ds.B08.astype(float)
B04 = ds.B04.astype(float)

# conduct NDVI calculation
ndvi = ((B08 - B04) / (B08 + B04)).rename('NDVI')
ndvi_mean = ndvi.mean(dim="time")

# plot temporally aggregated NDVI mean
ndvi_mean.plot.imshow(vmin=-0.50, vmax=0.8, cmap='RdYlGn')
```

STAC Item

Assets of a MODIS SnowCover tile (HDF format)

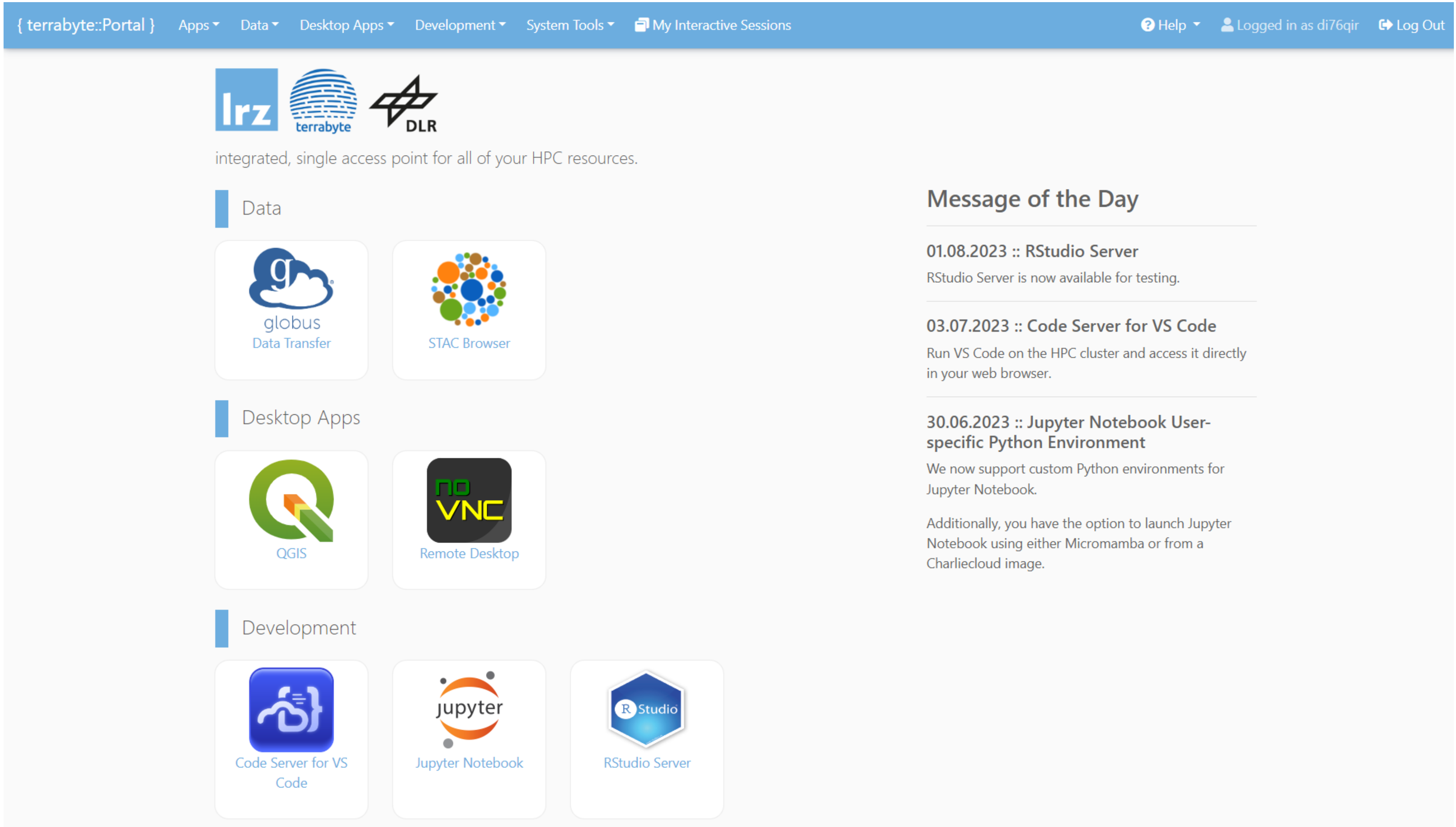


```
- hdf: {
      href: "file:///dss/dsstbyfs01/pn56su/pn56su-dss-0008/MODIS/MYD10A1.061/2023/06/09/h33v07/MYD10A1.A2023160.h33v07.061.2023162032452.hdf",
      type: "application/hdf4",
    - roles: [
          "data"
      ],
      title: "Source data containing all bands"
  },


- NDSI: {
      href: "file://HDF4_EOS:EOS_GRID:"/dss/dsstbyfs01/pn56su/pn56su-dss-
      0008/MODIS/MYD10A1.061/2023/06/09/h33v07/MYD10A1.A2023160.h33v07.061.2023162032452.hdf":MOD_Grid_Snow_500m:NDSI",
      roles: [ ],
      type: "application/x-hdf",
    - raster:bands: [
        - {
              unit: "none",
              scale: 0.00009999999747,
              nodata: 0,
              offset: 0,
              data_type: "int16"
          }
      ]
  },
```

terrabyte
Interactive applications

Compute Portal in the browser

JupyterLab, QGIS, Visual Studio Code, R-Studio Server



integrated, single access point for all of your HPC resources.

Data

Desktop Apps

Development

Message of the Day

01.08.2023 :: RStudio Server

RStudio Server is now available for testing.

03.07.2023 :: Code Server for VS Code

Run VS Code on the HPC cluster and access it directly in your web browser.

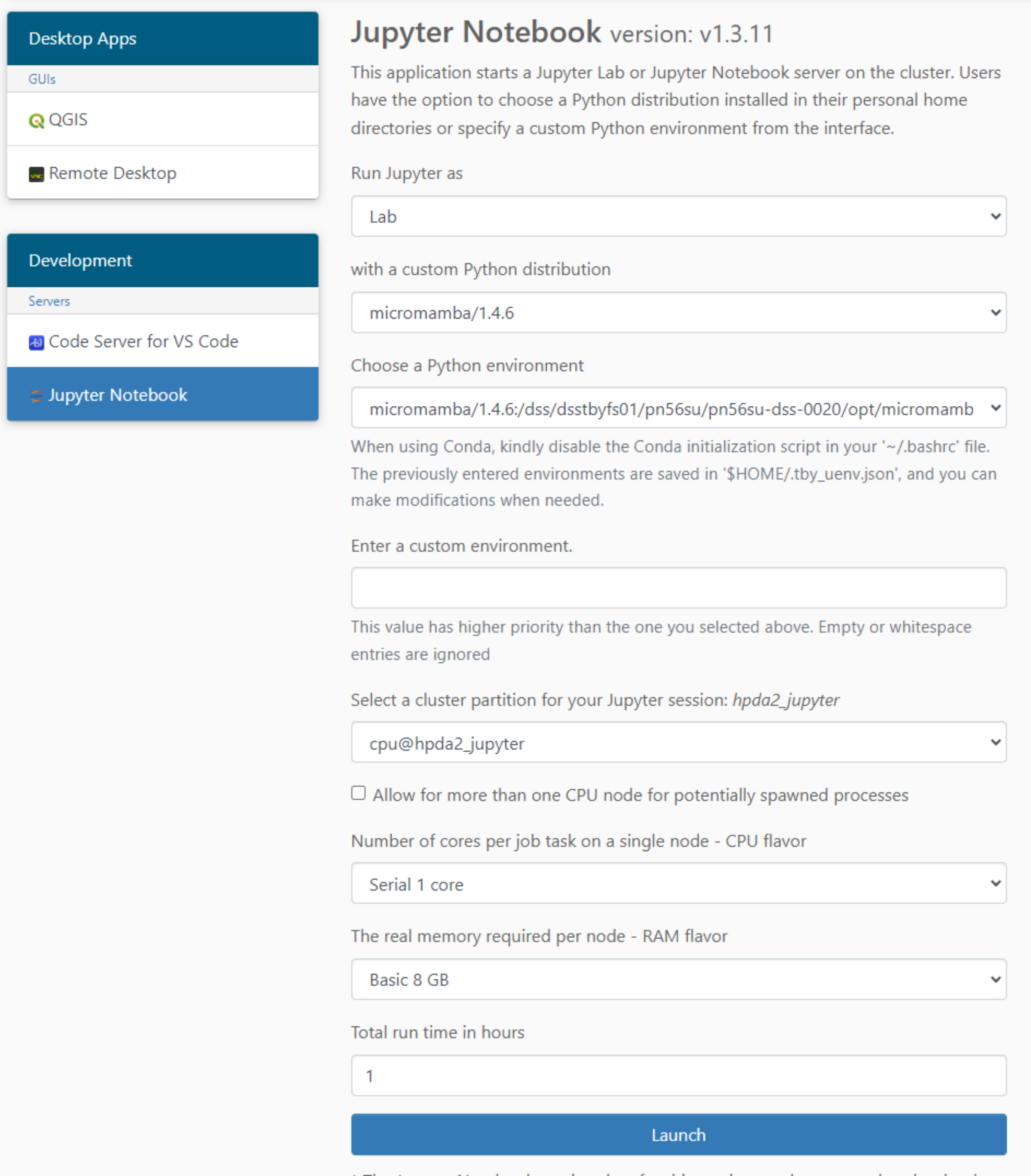
30.06.2023 :: Jupyter Notebook User-specific Python Environment

We now support custom Python environments for Jupyter Notebook.

Additionally, you have the option to launch Jupyter Notebook using either Micromamba or from a Charliecloud image.

Compute Portal in the browser

JupyterLab



Jupyter Notebook version: v1.3.11

This application starts a Jupyter Lab or Jupyter Notebook server on the cluster. Users have the option to choose a Python distribution installed in their personal home directories or specify a custom Python environment from the interface.

Run Jupyter as

Lab

with a custom Python distribution

micromamba/1.4.6

Choose a Python environment

micromamba/1.4.6:/dss/dsstbyfs01/pn56su/pn56su-dss-0020/opt/micromamb

When using Conda, kindly disable the Conda initialization script in your '~/.bashrc' file. The previously entered environments are saved in '$HOME/.tby_uenv.json', and you can make modifications when needed.

Enter a custom environment.

This value has higher priority than the one you selected above. Empty or whitespace entries are ignored

Select a cluster partition for your Jupyter session: *hpda2_jupyter*

cpu@hpda2_jupyter

☐ Allow for more than one CPU node for potentially spawned processes

Number of cores per job task on a single node - CPU flavor

Serial 1 core

The real memory required per node - RAM flavor

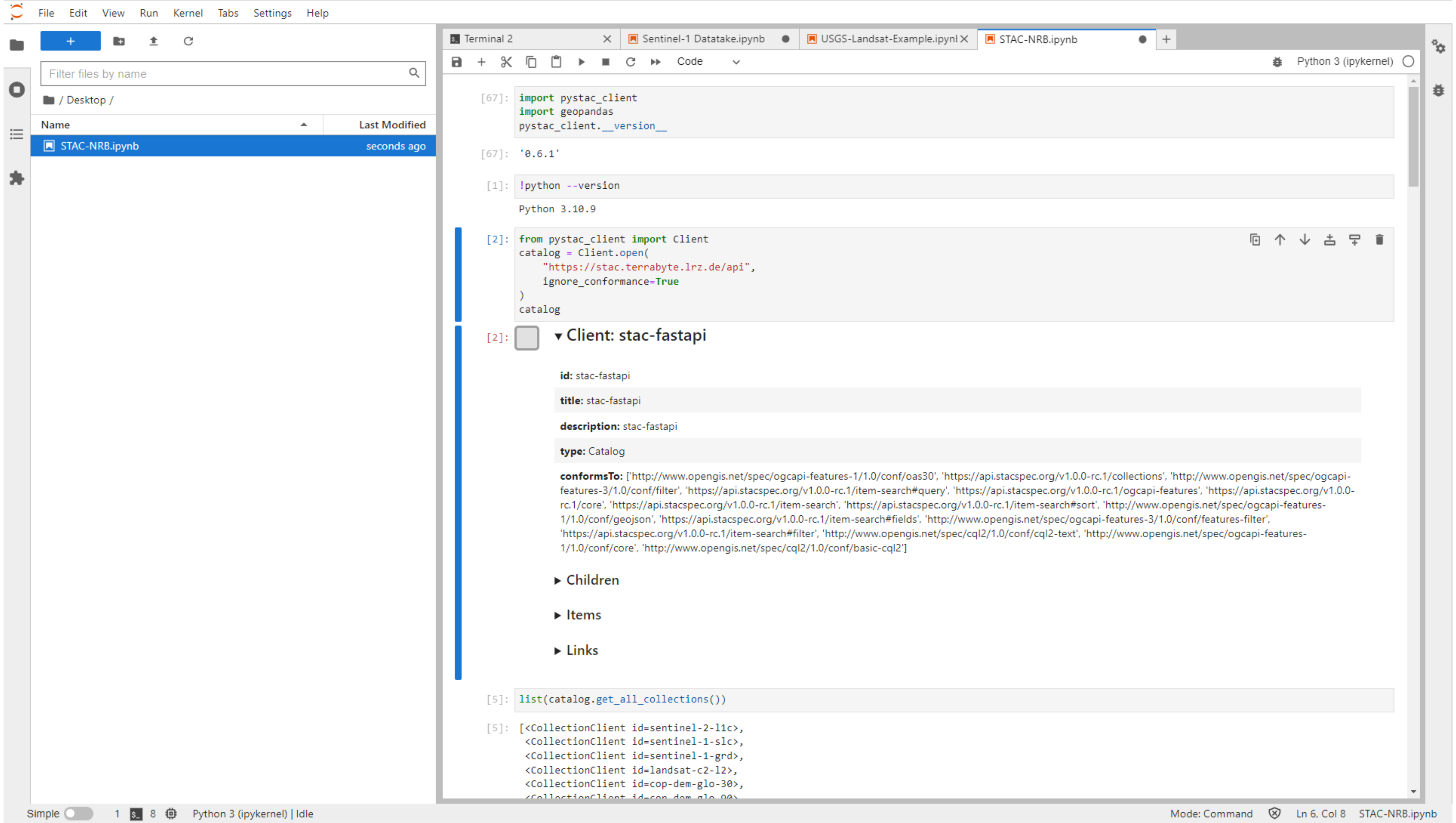
Basic 8 GB

Total run time in hours

1

Compute Portal in the browser

JupyterLab



```
[67]: import pystac_client
      import geopandas
      pystac_client.__version__

[67]: '0.6.1'

[1]:  !python --version

      Python 3.10.9

[2]:  from pystac_client import Client
      catalog = Client.open(
          "https://stac.terrabyte.lrz.de/api",
          ignore_conformance=True
      )
      catalog
```

[2]: ▾ Client: stac-fastapi

id: stac-fastapi

title: stac-fastapi

description: stac-fastapi

type: Catalog

conformsTo: ['http://www.opengis.net/spec/ogcapi-features-1/1.0/conf/oas30', 'https://api.stacspec.org/v1.0.0-rc.1/collections', 'http://www.opengis.net/spec/ogcapi-features-3/1.0/conf/filter', 'https://api.stacspec.org/v1.0.0-rc.1/item-search#query', 'https://api.stacspec.org/v1.0.0-rc.1/ogcapi-features', 'https://api.stacspec.org/v1.0.0-rc.1/core', 'https://api.stacspec.org/v1.0.0-rc.1/item-search', 'https://api.stacspec.org/v1.0.0-rc.1/item-search#sort', 'http://www.opengis.net/spec/ogcapi-features-1/1.0/conf/geojson', 'https://api.stacspec.org/v1.0.0-rc.1/item-search#fields', 'http://www.opengis.net/spec/ogcapi-features-3/1.0/conf/features-filter', 'https://api.stacspec.org/v1.0.0-rc.1/item-search#filter', 'http://www.opengis.net/spec/cql2/1.0/conf/cql2-text', 'http://www.opengis.net/spec/ogcapi-features-1/1.0/conf/core', 'http://www.opengis.net/spec/cql2/1.0/conf/basic-cql2']

▸ Children

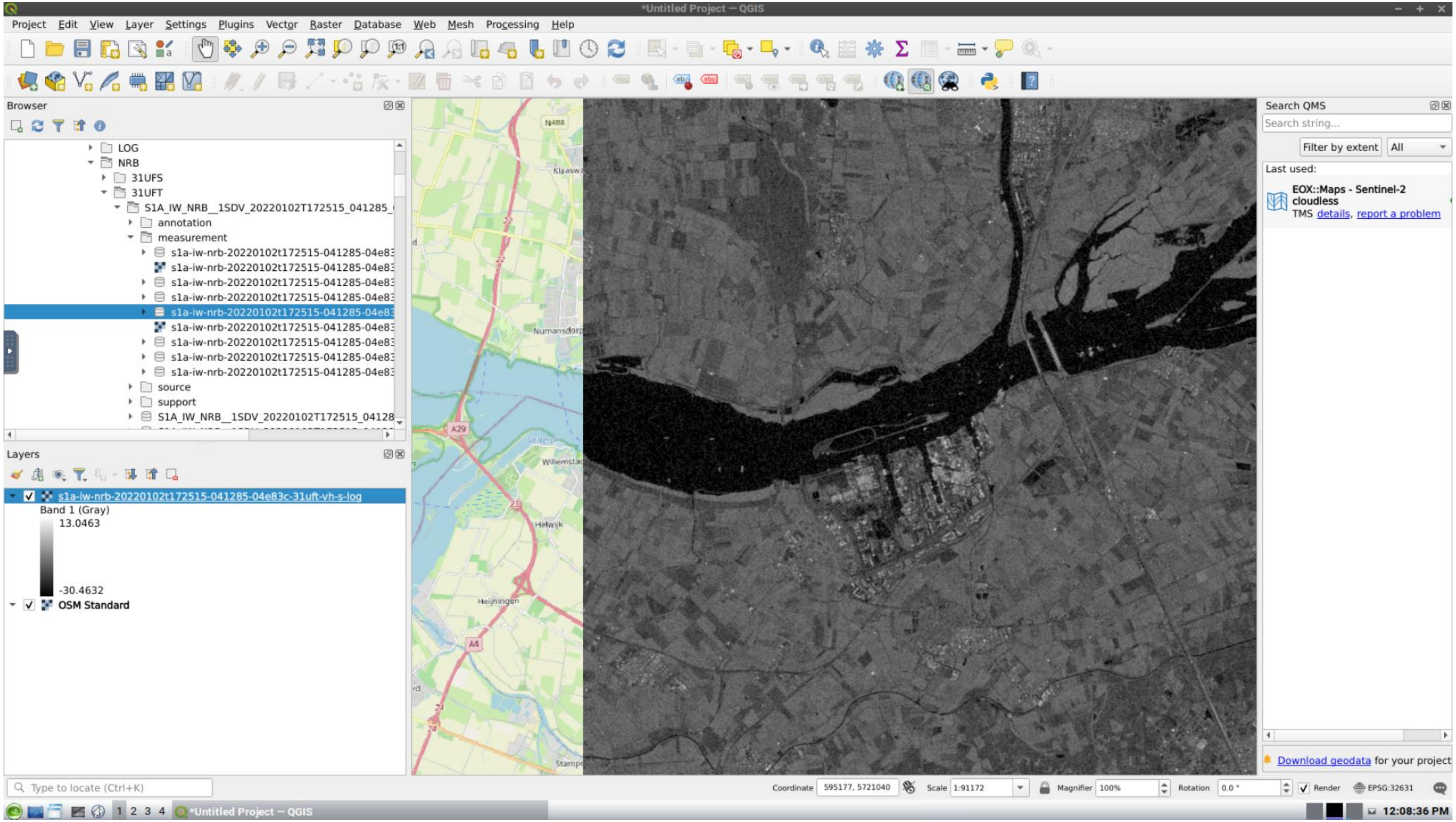
▸ Items

▸ Links

```
[5]:  list(catalog.get_all_collections())

[5]:  [<CollectionClient id=sentinel-2-l1c>,
       <CollectionClient id=sentinel-1-slc>,
       <CollectionClient id=sentinel-1-grd>,
       <CollectionClient id=landsat-c2-l2>,
       <CollectionClient id=cop-dem-glo-30>,
       <CollectionClient id=cop-dem-glo-90>,
```

Compute Portal in the browser

QGIS



```
LOG
NRB
  31UFS
  31UFT
    S1A_IW_NRB__1SDV_20220102T172515_041285_
      annotation
      measurement
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
        s1a-iw-nrb-20220102t172515-041285-04e83
      source
      support
    S1A_IW_NRB__1SDV_20220102T172515_04128
```

☑ s1a-iw-nrb-20220102t172515-041285-04e83c-31uft-vh-s-log

Band 1 (Gray)

13.0463

-30.4632

☑ OSM Standard

Search string...

Last used:

EOX::Maps - Sentinel-2 cloudless

TMS details, report a problem

Download geodata for your project

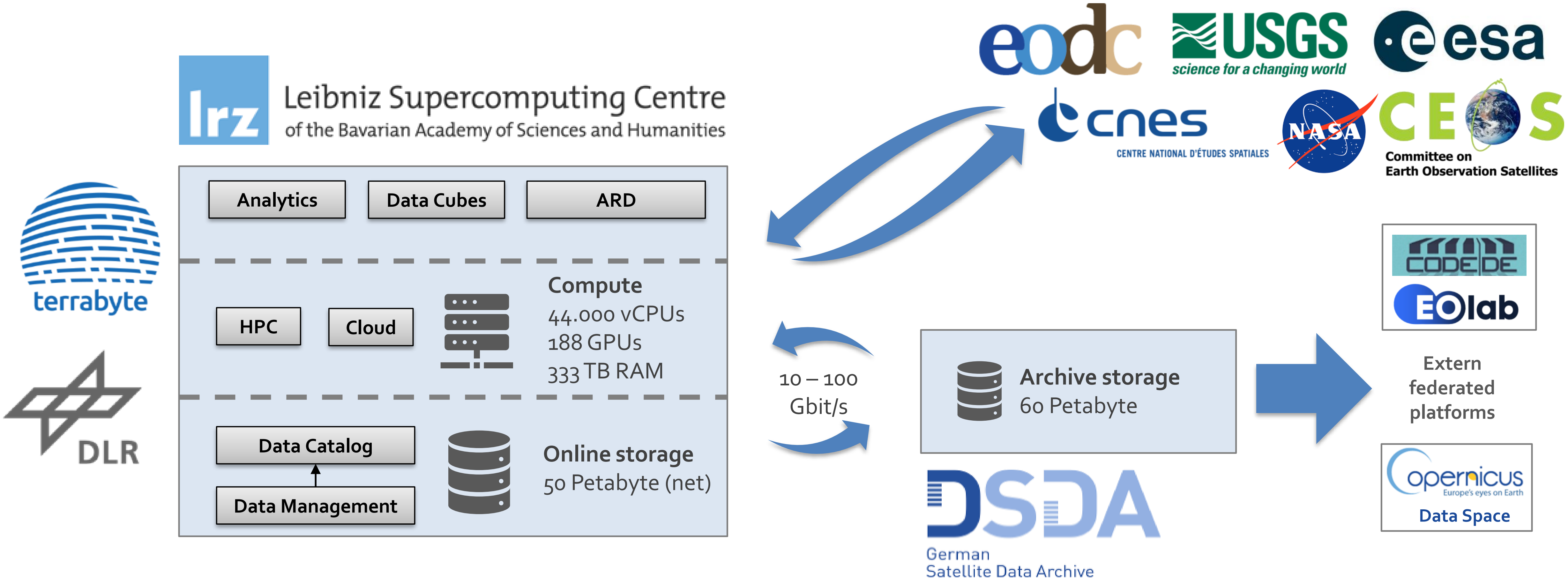
Type to locate (Ctrl+K)

Coordinate 595177, 5721040 | Scale 1:91172 | Magnifier 100% | Rotation 0.0 ° | ☑ Render | EPSG:32631

1 2 3 4 *Untitled Project — QGIS 12:08:36 PM

terrabyte
Platform cooperation

Federation of services and platforms



Leibniz Supercomputing Centre
of the Bavarian Academy of Sciences and Humanities

terrabyte

DLR

Analytics

Data Cubes

ARD

HPC

Cloud

Compute
44.000 vCPUs
188 GPUs
333 TB RAM

Data Catalog

Data Management

Online storage
50 Petabyte (net)

10 – 100
Gbit/s

Archive storage
60 Petabyte

DSDA
German
Satellite Data Archive

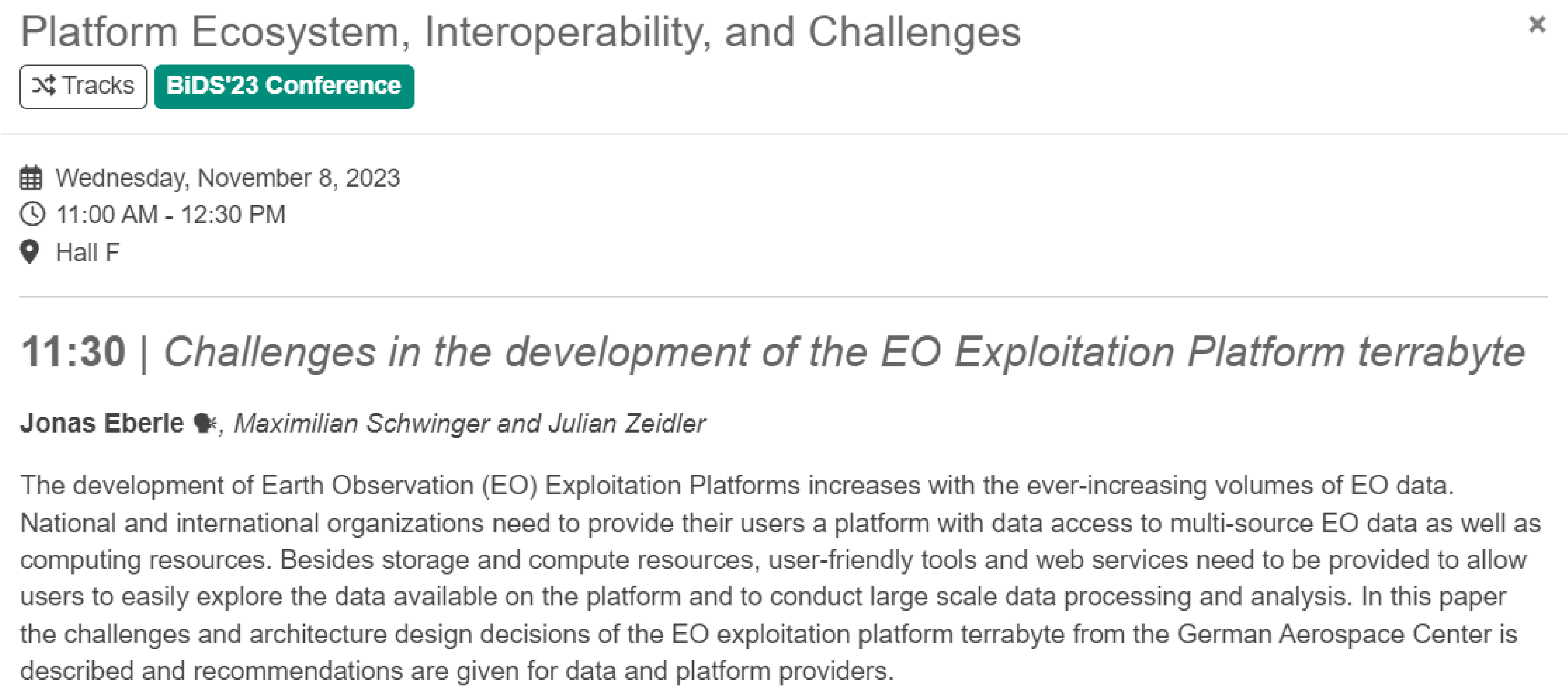
eodc
USGS
science for a changing world
esa
cnes
CENTRE NATIONAL D'ÉTUDES SPATIALES
NASA
CEOS
Committee on
Earth Observation Satellites
CODE-DE
EOlab

Extern
federated
platforms

copernicus
Europe's eyes on Earth
Data Space

Thank you for your attention!

BiDS | BIG DATA FROM SPACE 2023
6–9 November 2023 | Vienna, Austria



Platform Ecosystem, Interoperability, and Challenges

Wednesday, November 8, 2023
11:00 AM - 12:30 PM
Hall F

11:30 | *Challenges in the development of the EO Exploitation Platform terrabyte*

Jonas Eberle 🗣, *Maximilian Schwinger and Julian Zeidler*

The development of Earth Observation (EO) Exploitation Platforms increases with the ever-increasing volumes of EO data. National and international organizations need to provide their users a platform with data access to multi-source EO data as well as computing resources. Besides storage and compute resources, user-friendly tools and web services need to be provided to allow users to easily explore the data available on the platform and to conduct large scale data processing and analysis. In this paper the challenges and architecture design decisions of the EO exploitation platform terrabyte from the German Aerospace Center is described and recommendations are given for data and platform providers.