



# Destination Earth

## Destination Earth Data Lake: Jeden punkt dostępu, dziesiątki źródeł, setki kolekcji danych



**Autorzy:** Patryk Grzybowski<sup>1</sup>; Aubin Lambare<sup>2</sup>; Christoph Reimer<sup>3</sup>; Michael Schick<sup>4</sup>; Danaele Puechmaille<sup>4</sup>

**Afiliacje:** <sup>1</sup>CloudFerro S.A., Warsaw, Poland; <sup>2</sup>CS Group - Sopra Steria, Le Plessis Robinson, France; <sup>3</sup>EODC, Vienna, Austria; <sup>4</sup>EUMETSAT, Darmstadt, Germany

Destination Earth

Funded by  
the European Union



Implemented by





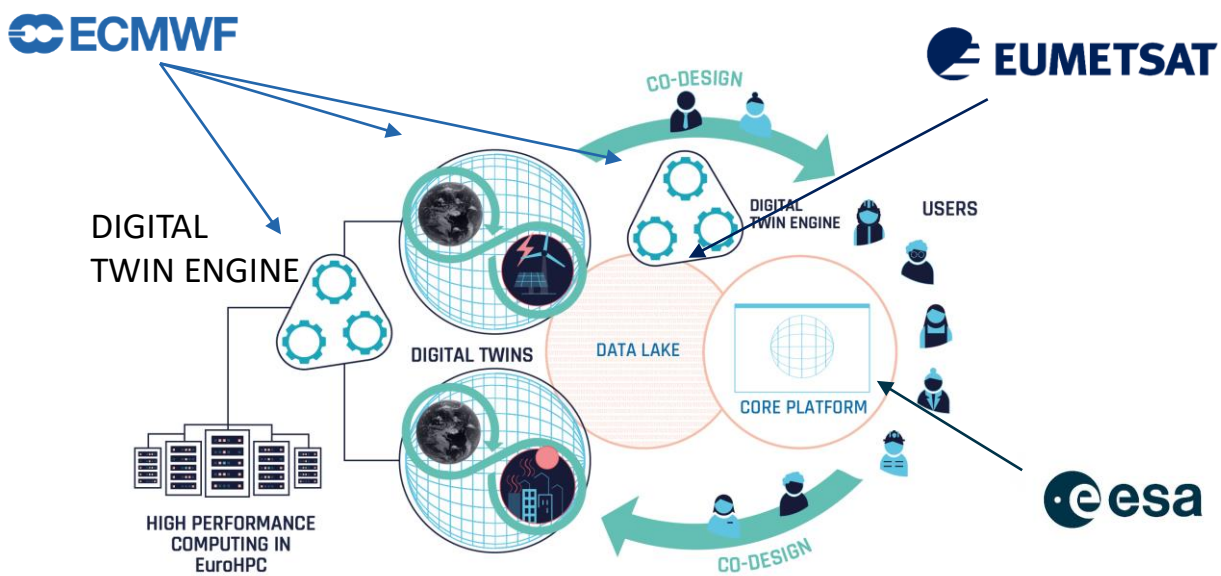
**Destination Earth (DestinE)** - sztandarowa inicjatywa Komisji Europejskiej, której celem jest opracowanie wysoce precyzyjnej cyfrowej repliki Ziemi w celu modelowania, monitorowania i symulowania zjawisk naturalnych, zagrożeń oraz działalności człowieka.

Trzy komponenty DestinE:

Core Service Platform (DESP)

Data Lake (DEDL)

Digital Twins (DTs)

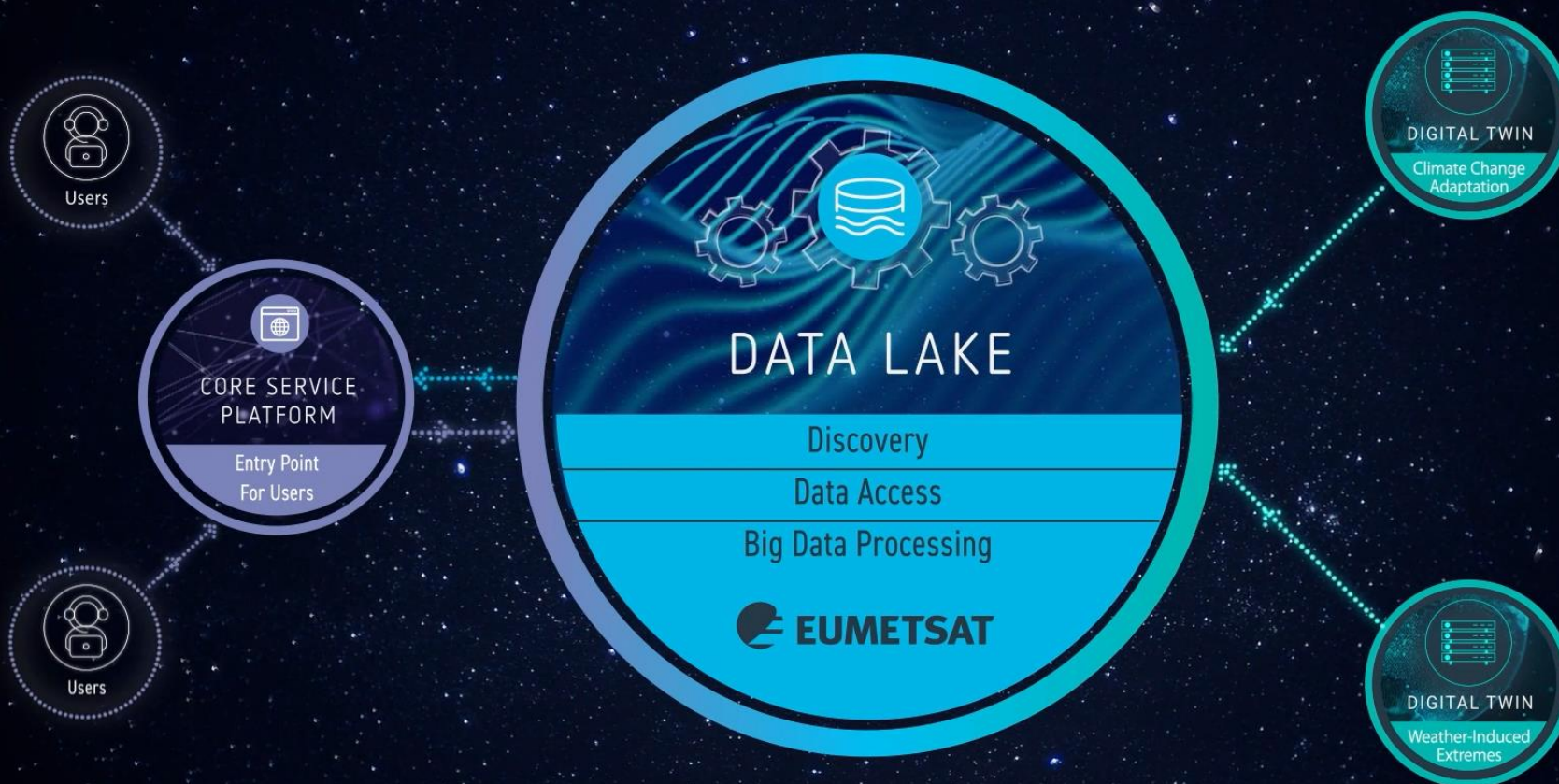


(ECMWF, 2024)



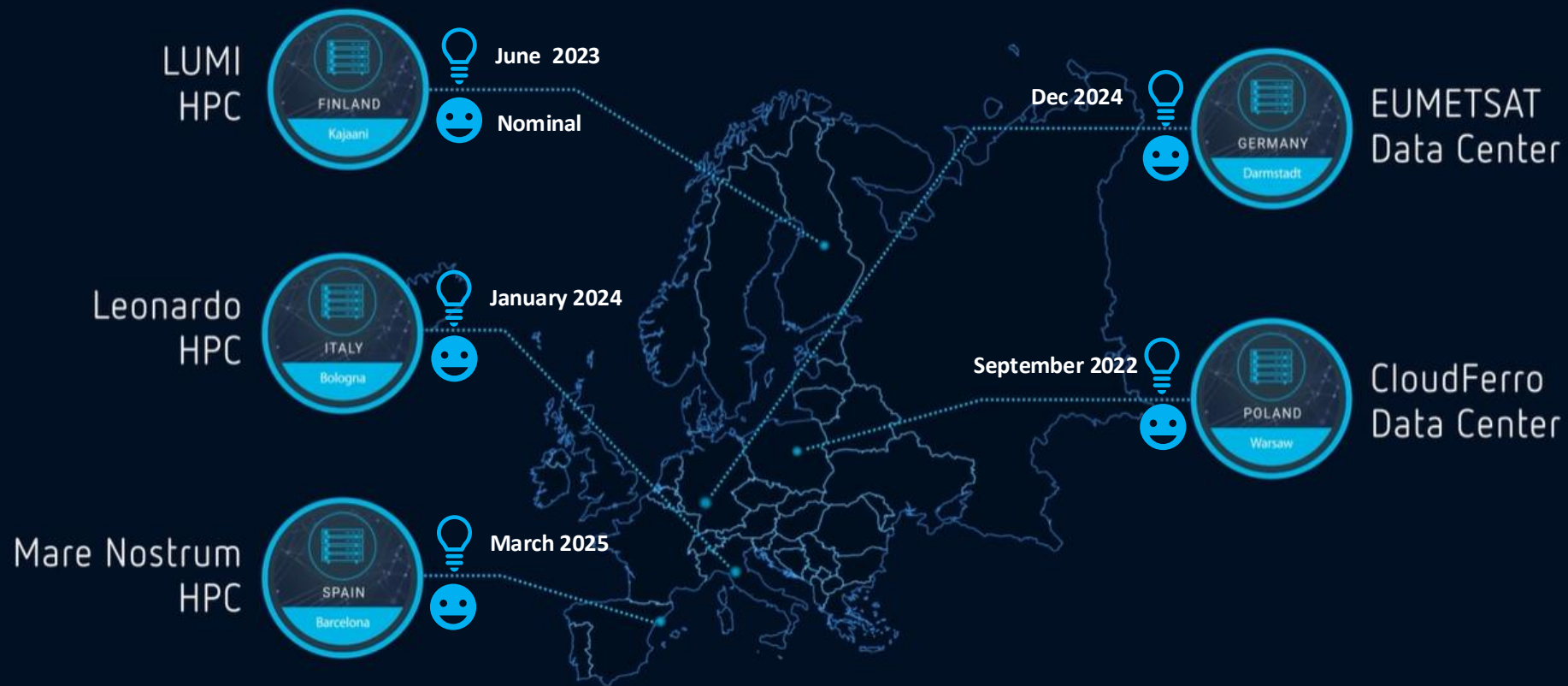


## Destination Earth System



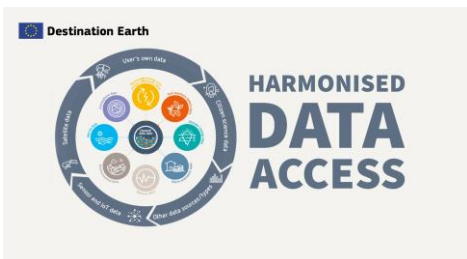


## Data Lake distributed infrastructure





HDA



### Harmonised Data Access:

- DestinE Data – DT Outputs & User Generated Data for DestinE
- Federated Data – Federated Data Providers
- API => Spatio Temporal Asset Catalog (STAC)

Edge Services – Na wniosek



### STACK: usługa PaaS, oferująca obliczenia obok danych, a także:

- JupyterHub,
- Dask/ Dask Gateway

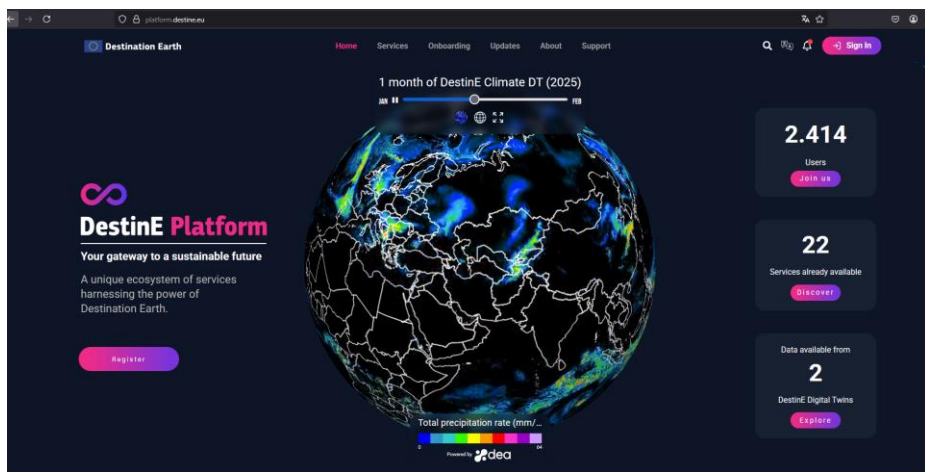


- ISLET Compute: (IaaS/PaaS) umożliwia przetwarzanie danych w ich bezpośrednim sąsiedztwie, pozwalając użytkownikom zarządzać i wdrażać wirtualne środowiska pracy
- ISLET Storage: usługa s3 object storage, do przechowywania danych użytkowników oraz wyników przetwarzania.

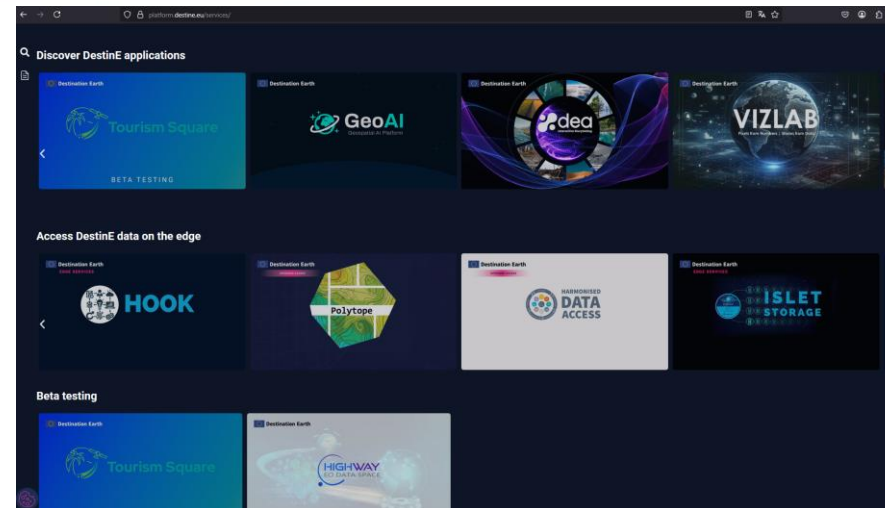


- HOOK - pozwala na uruchamianie złożonych, predefiniowanych lub niestandardowych workflow.

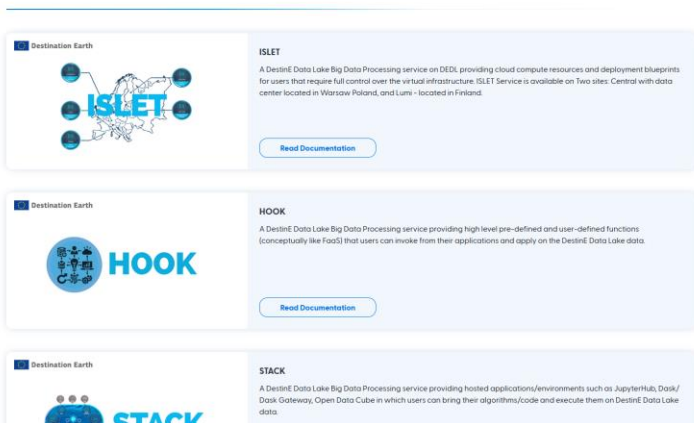
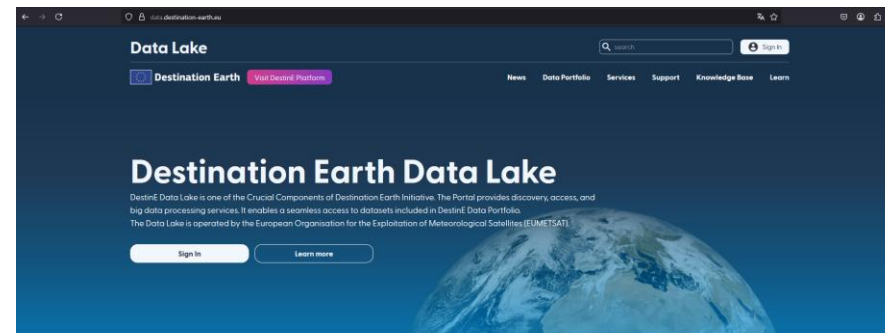




<https://platform.destine.eu/>



<https://data.destination-earth.eu/>



## Destination Earth Data Lake

The Goal of DestinE Data Lake (DEDL) is to help us better understand our Ecosystem, enhance decision-making process, and enable us to build a better, safer and more sustainable future. DEDL grants its users access to huge amounts of EO and Climate Data and makes it available to all users.





# HARMONISED DATA ACCESS



Około 200  
kolekcji

Jeden  
zestaw  
hasel

Różni  
dostawcy  
danych



Dane  
Digital  
Twins

Przeglądanie  
i dostęp do  
danych

STAC

## HARMONISED DATA ACCESS





# Destination Earth

## Destination Earth Data Portfolio

Harmonised Data Access (HDA)

Tomorrow's Solutions Start Here  
Simulating Tomorrow Today

DestinE Digital Twin Data

**Climate Change Adaptation DT**

- Global Multi-decadal climate projections with local granularity up to 2050
- Models: IFS-NEMO, ICON, IFS-FESOM
- High resolution: 5 or 10 km
- HEALPix grid
- Future projection: 2020-2050
- Hourly and daily data, surface and multi-level atmosphere/ocean fields

**Weather-induced Extremes DT**

- Improving prediction of extreme weather and its impacts
- 4 day forecast (daily):
  - Global, Regional, Local
- Models: IFS-NEMO
- High resolution: Atm 4.4km, Ocean 25km
- Hourly and daily data, surface and multi-level atmosphere/ocean fields

Example question: *What would the 2020 European freshwater look like in a +2°C world?*

Federated Datasets

**Big Data Services**

- EUMETSAT**: Metop, Meteosat, Sentinel-3-Marine
- ESA**: Sentinel-1, Sentinel-2, Sentinel-3-Land, Sentinel-5P
- EC EUROPA Data Store**: Population, Greenhouse emissions, Energy Sources...
- EC JRC / Copernicus**: Global Surface Water
- ISIMIP**: ISIMIP Climate Change, Socio-Economic

**Copernicus Services**

- Atmospheric Monitoring (ADS) **ECMWF**
- Climate Change (C3S) **ECMWF**
- Marine (CMEMS) **ECMWF**
- Global Land (CLMS) **ECMWF**
- Emergency (CEMS) **ECMWF**

User-Generated Data

- Data generated by users using their own algorithms on DestinE Data.
- This data is made accessible to the user community following review

User-Driven Evolutions

- Data Providers, Collections and functionalities will evolve over time based on DestinE user needs

Share / Collaborate

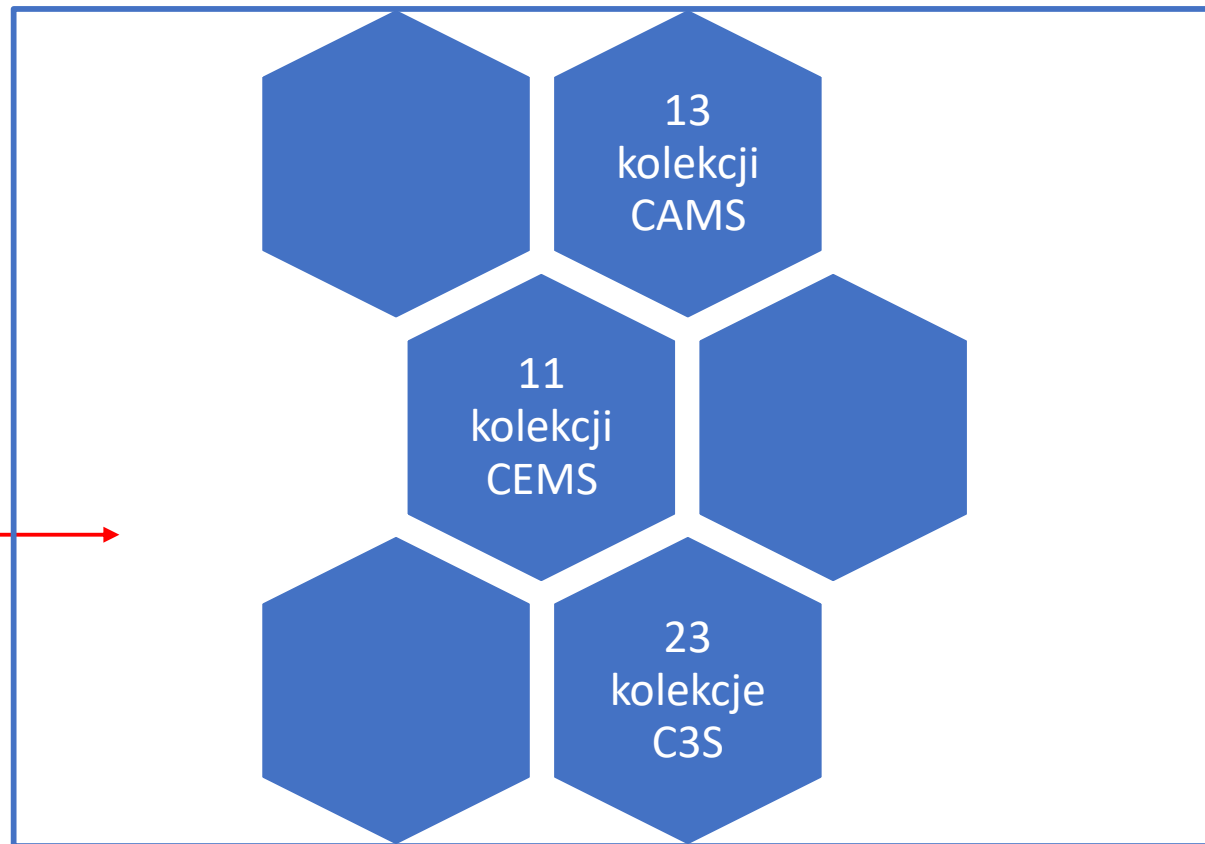
DATA PORTFOLIO

Destination Earth

Funded by the European Union

Destination Earth

IMPLEMENTED BY **EUMETSAT** **ESA** **ECMWF**





JupyterLab w ramach  
DEDL STACK Service

Maszyna wirtualna w  
ramach DEDL Islet Service

Środowisko lokalne



HARMONISED  
**DATA  
ACCESS**



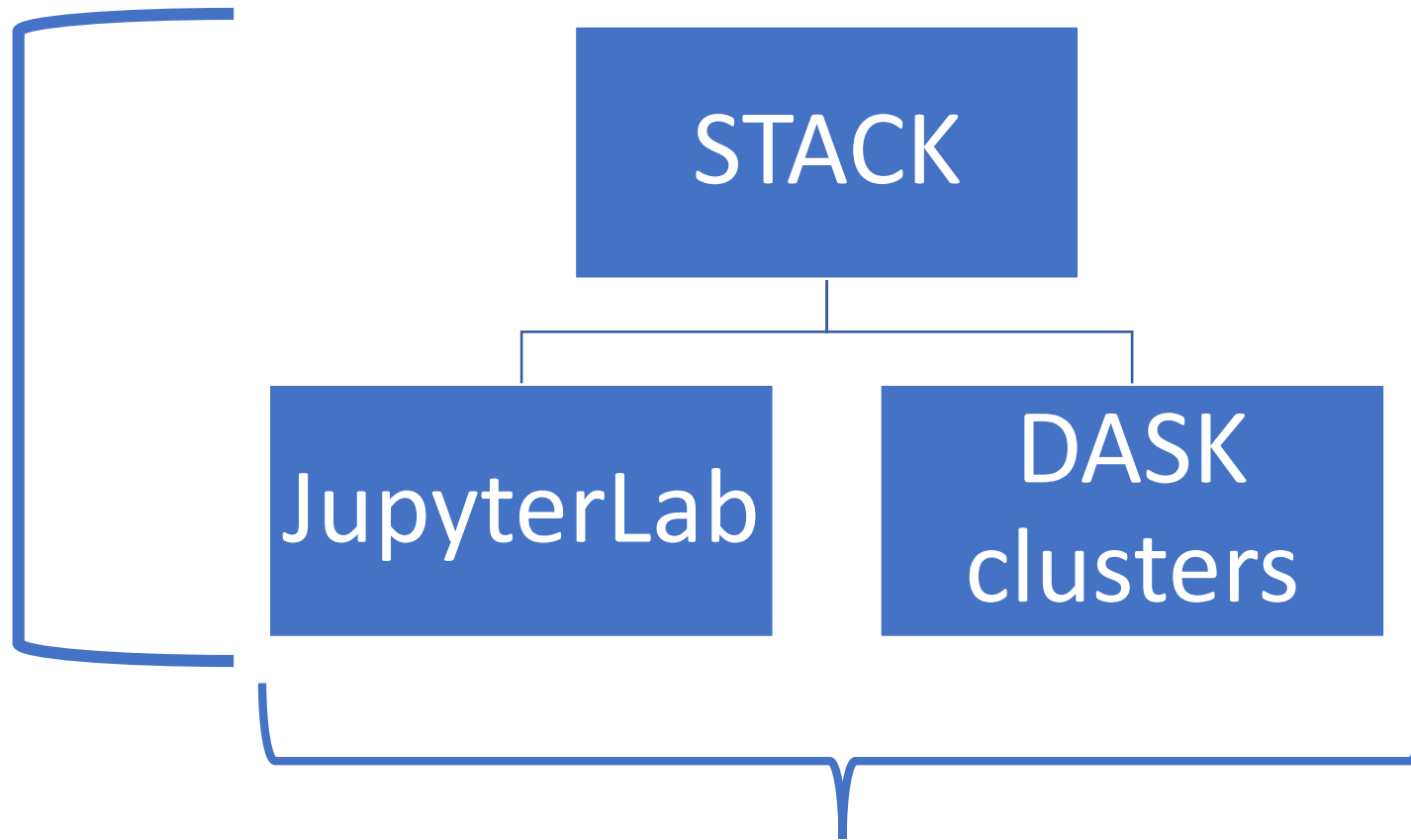
# Destination Earth



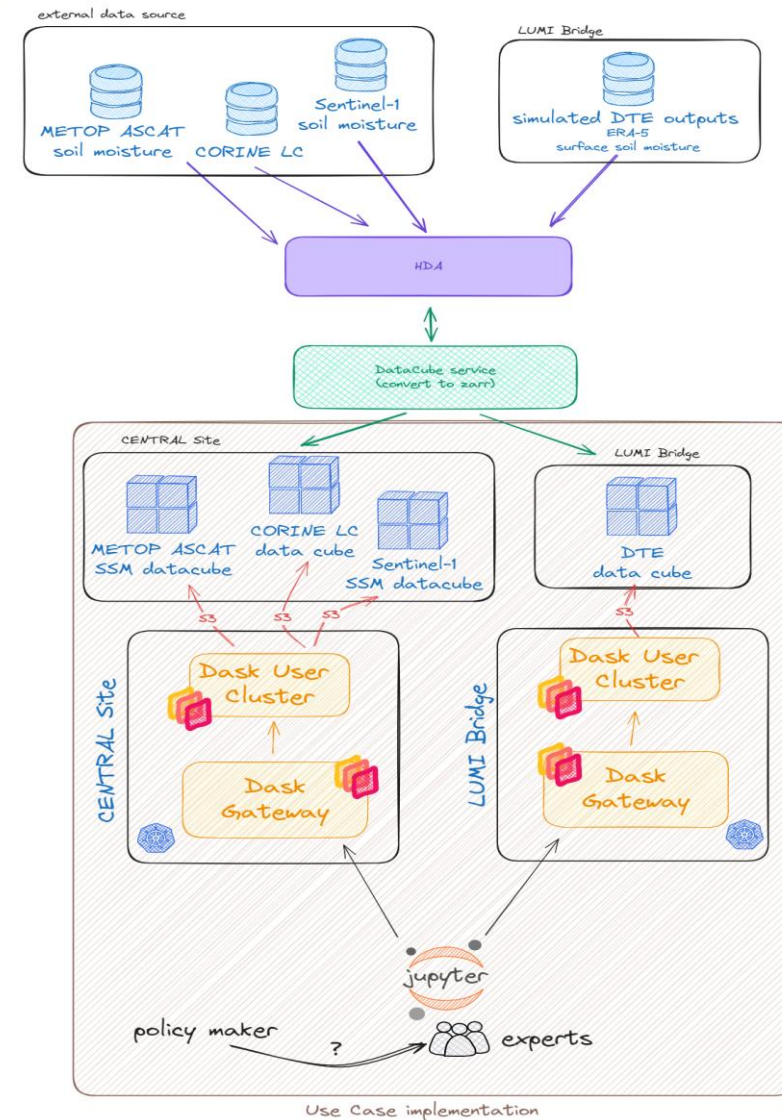
# STACK



## Data Bridges



HDA: restAPI  
PySTAC  
EODAG





## Server Options

- JupyterLab (Full working environment)

DestinE Data Lake JupyterLab

---

- Remote Compute Kernels - Central Site

Remote Kernels to compute at Central Bridge (Jupyter Home not available & use of Object Storage S3 for Data Exchange mandatory)

---

- Remote Compute Kernels - LUMI Bridge

Remote Kernels to compute at LUMI Bridge (Jupyter Home not available & use of Object Storage S3 for Data Exchange mandatory)

---

- Remote Compute Kernels - LEONARDO Bridge

Remote Kernels to compute at LEONARDO Bridge (Jupyter Home not available & use of Object Storage S3 for Data Exchange mandatory)

---

- Remote Compute Kernels - EUMETSAT Bridge

Remote Kernels to compute at EUMETSAT Bridge (Jupyter Home not available & use of Object Storage S3 for Data Exchange mandatory)

---

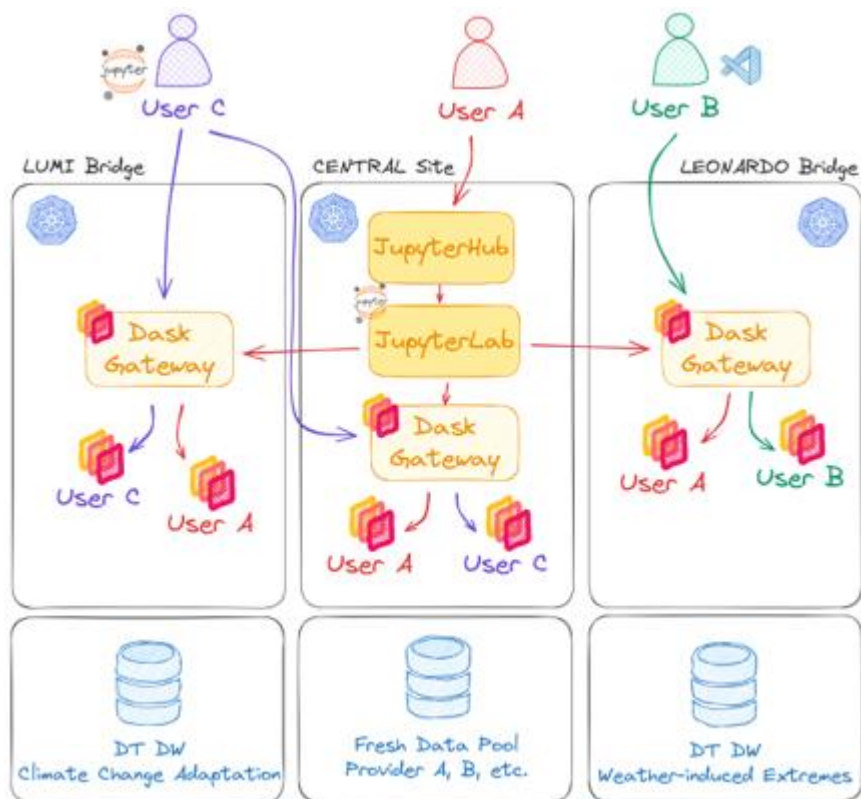
- Remote Compute Kernels - MareNostrum Bridge

Remote Kernels to compute at MareNostrum Bridge (Jupyter Home not available & use of Object Storage S3 for Data Exchange mandatory)

---

Start





Zdalny dostęp do klastrów: zarządzanie i wykonywanie zadań na **zdalnych klastrach** (np. w chmurze lub systemach HPC) bezpośrednio z poziomu JupyterLab

Dashboard'y w czasie rzeczywistym: monitorowanie wydajności, postępu zadań oraz wykorzystania zasobów za pomocą **interaktywnych pulpitów**

Łatwe skalowanie: dynamicznie równoważy obciążenia między węzłami roboczymi, umożliwiając skalowalne, **rozproszone przetwarzanie**





Destination Earth



# HOOK



## Predefiniowane procesory

- Data harvest
- LAI
- CARD BS
- CARD COHINF
- SNAP C2RCC
- SEN2COR
- MAJA

Procesory tworzone przez użytkowników



Prywatny s3 storage  
OR  
Tymczasowy s3 storage





## Komunikacja przez API:

- REST API
- openEO API

## Komunikacja przez interfejs graficzny:

- JupyterLab (STACK)
- openEO (UI)

The screenshot shows a JupyterLab environment with a file browser on the left and a notebook titled 'Tutorial.ipynb' open. The notebook content includes:

- Header: Destination Earth logo and logos for EUMETSAT, ESA, and ECMWF.
- Title: DEDL - Hook Tutorial
- Text: 'This notebook demonstrates how to use the Hook service.' and 'Author: EUMETSAT'
- Text: 'The detailed API and definition of each endpoint and parameters is available in the OnDemand Processing API Swagger UI at: <https://odp.data.destination-earth.eu/odata/docs>'
- Text: 'Further documentation is available at: <https://destine-data-lake-docs.data.destination-earth.eu/en/latest/dedl-big-data-processing-services/Hook-service/Hook-service.html>'
- Section: 'Install package and import environment variables' with code:

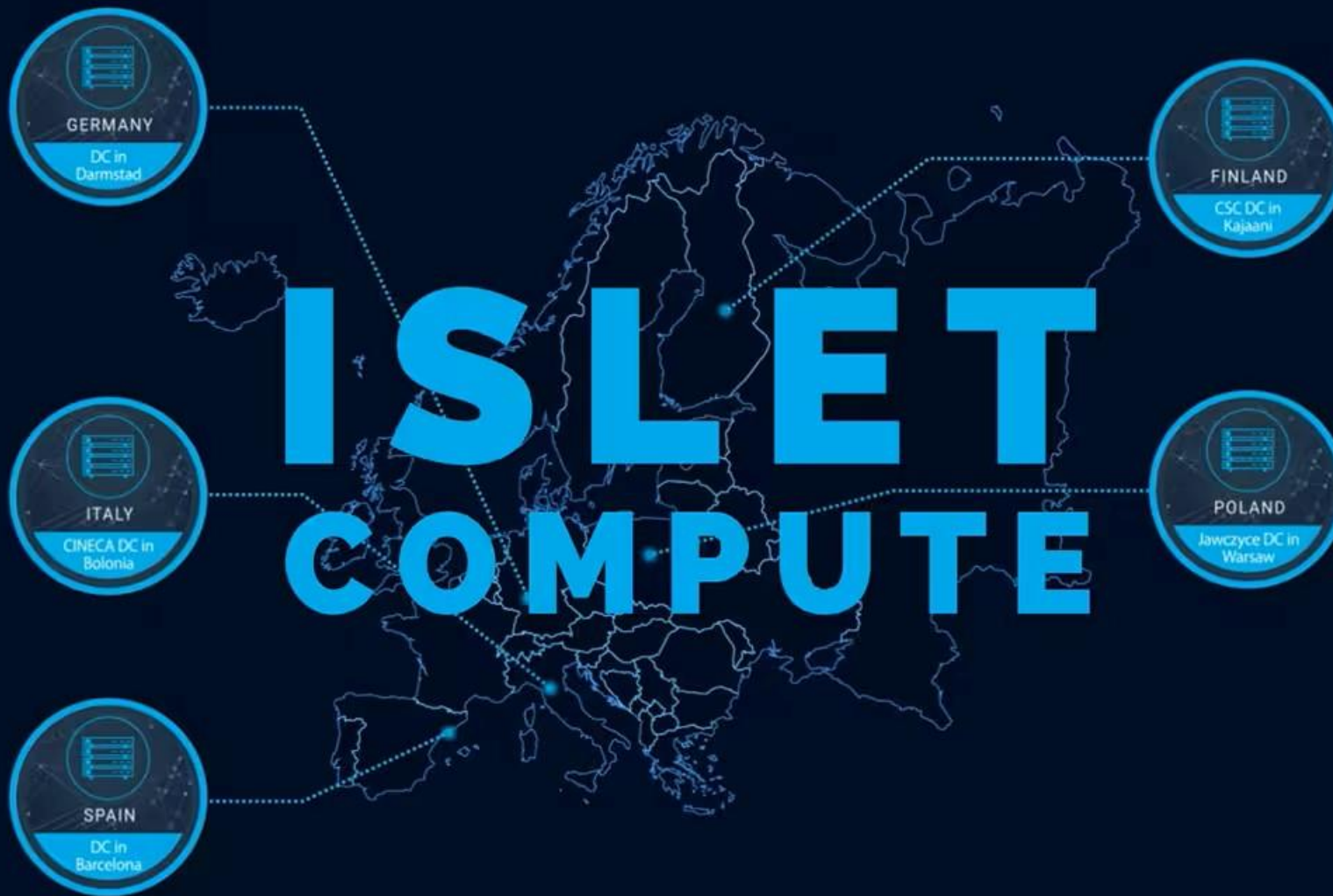
```
[ ]: #install DEDL authentication package to simplify your code
!pip install destine-auth
```
- Section: 'Authentication - Get token' with code:

```
[ ]: #Request DESP credentials
DESP_USERNAME = input("Please input your DESP username or email: ")
DESP_PASSWORD = getpass("Please input your DESP password: ")
```





# Destination Earth





---

Wirtualne maszyny i wirtualne środowiska

---

Gotowe obrazy systemów operacyjnych

---

CPU i GPU

---

Klastry K8s

---



---

Dostęp s3 storage

---

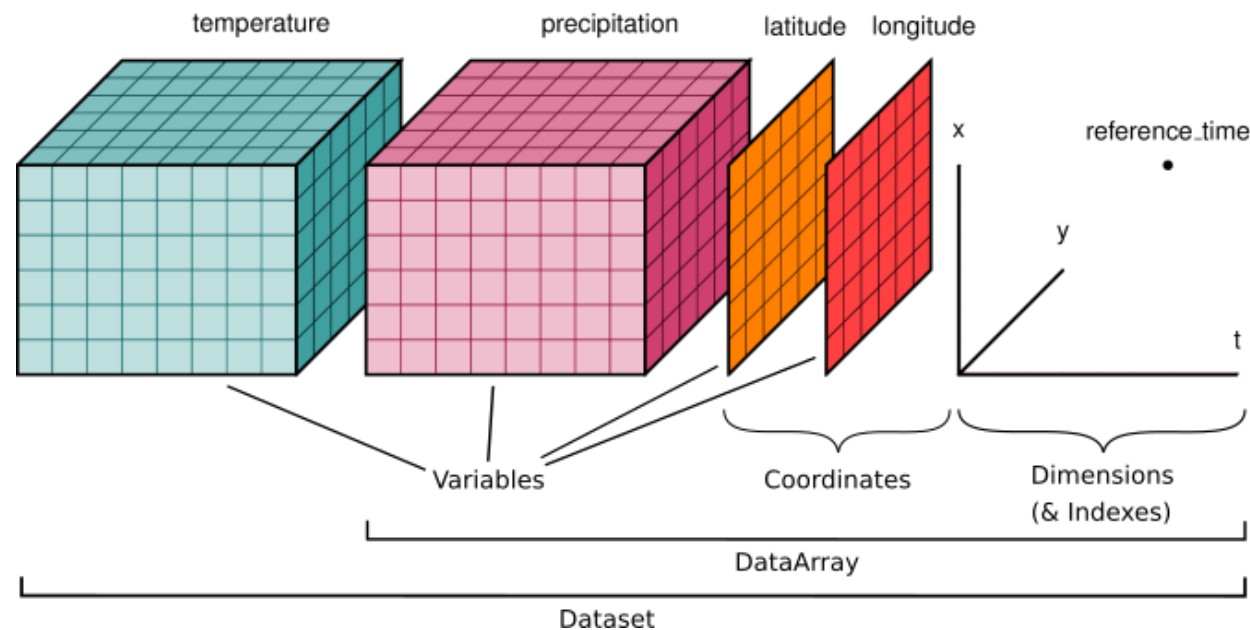
Zarządzanie s3 storage

---





Dlaczego nam to wszystko potrzebne?



Dane w  
natywnym  
formacie



.zarr/.netcdf

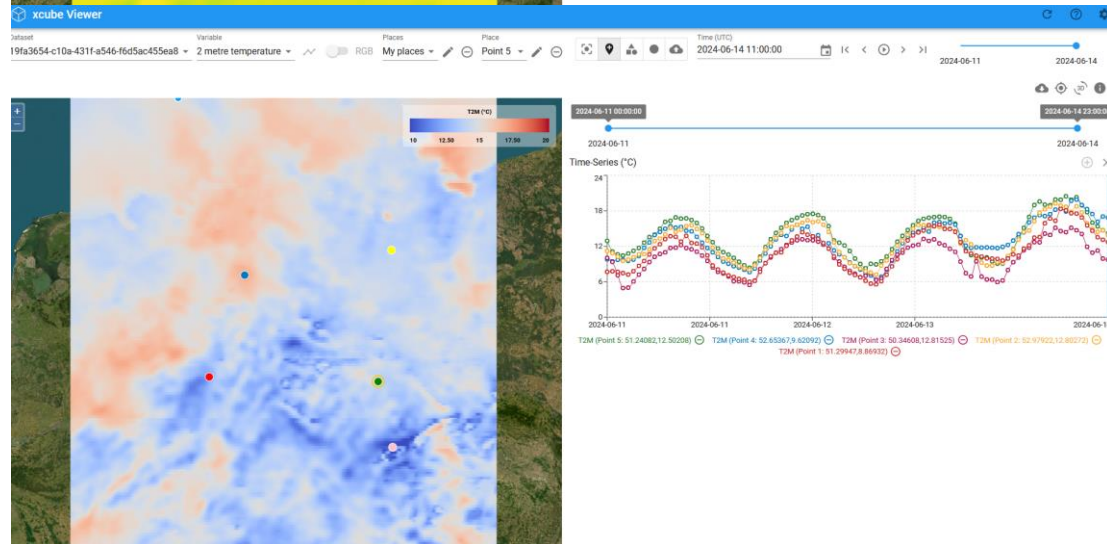
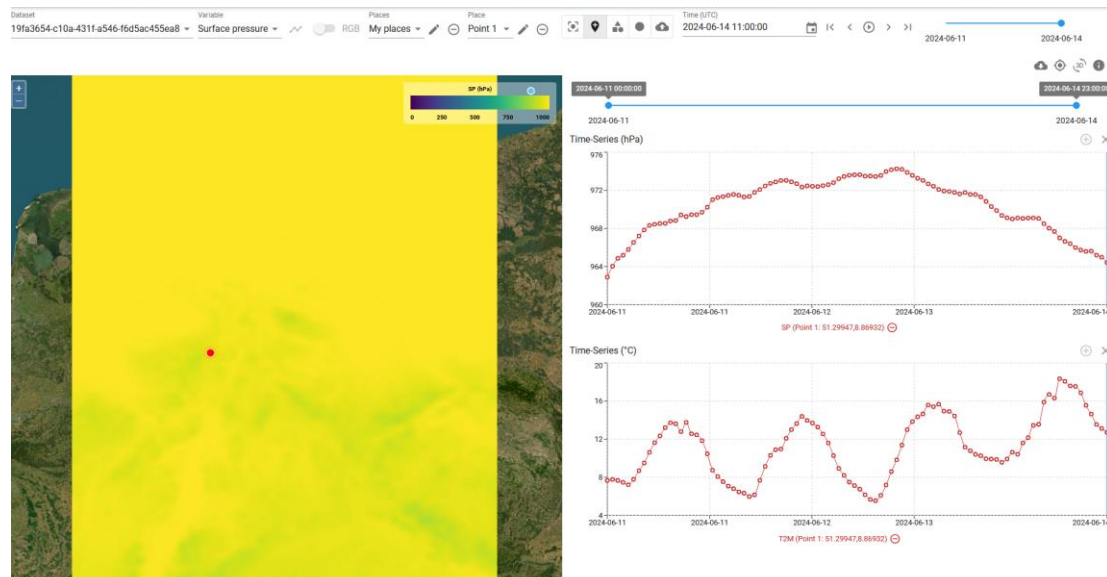
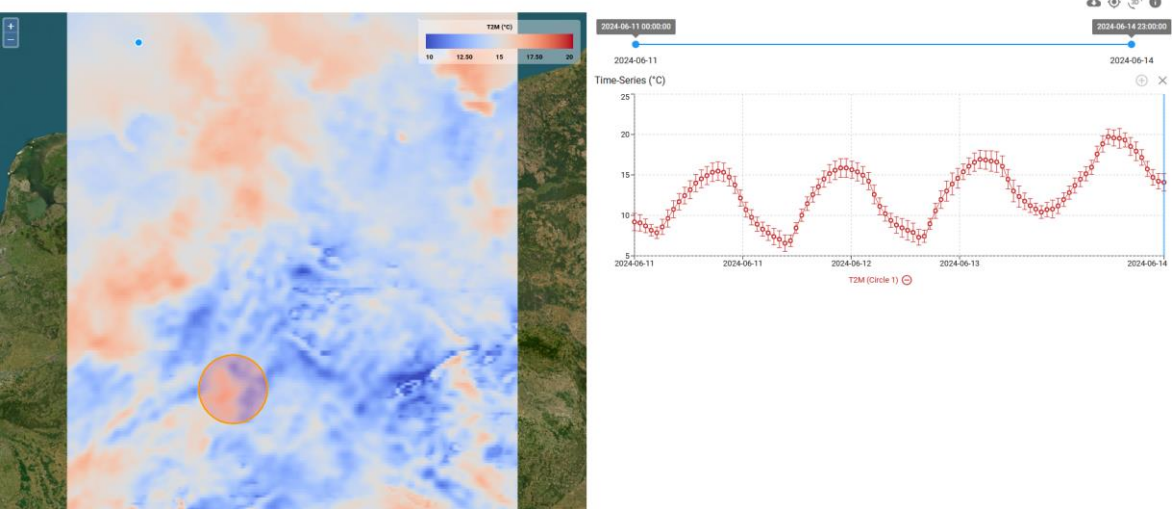
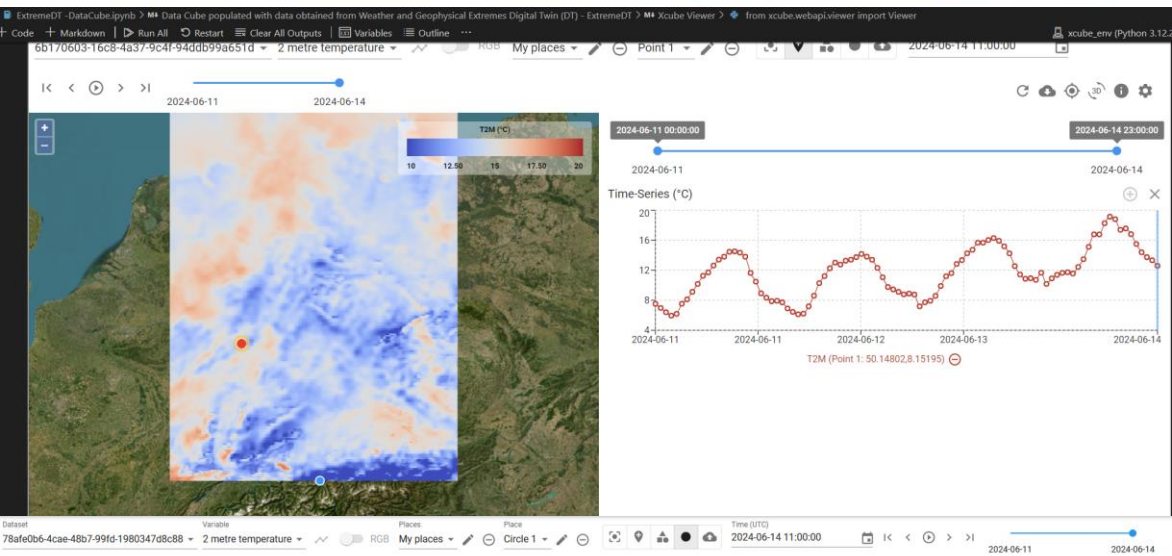


Python



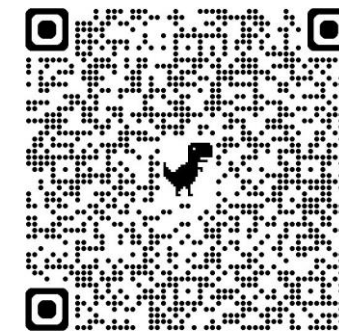


# Destination Earth





# Wniosek o dostęp do ,Edge Services'



## Edge services





## https://github.com/destination-earth/DestinE-DataLake-Lab

DestinE-DataLake-Lab Public

main 6 Branches 0 Tags

serenaateum different variable in plotter 76b4240 · 12 hours ago 245 Commits

- HDA different variable in plotter 12 hours ago
- HOOK Update Hook Tutorial Desp Only Credentials and Readme 4 days ago
- STACK Data Cube in xviewer - update 3 months ago
- img style 6 months ago
- LICENSE.txt Create LICENSE.txt 4 months ago
- README.md README instruction updated and different management of er... 12 hours ago

README MIT license

Funded by the European Union **Destination Earth** IMPLEMENTED BY EUMETSAT esa ECMWF

### DestinE-DataLake-Lab

Author: EUMETSAT

Destination Earth Data Lake Laboratory, which contains additional information for working with DestinE Data Lake services:

- Harmonised Data Access (Jupyter notebooks examples + Python Tools)
- STACK service (Jupyter Notebook examples on how to use DASK for near data processing)
- HOOK service (Jupyter Notebook examples on how to use HOOK for workflows)

Further information available in DestinE Data Lake documentation: <https://destine-data-lake-docs.data.destination-earth.eu/en/latest/index.html>

Additional resources:

DestinE-DataLake-Lab / HDA / DestinE Digital Twins / DEDL-HDA-EO-ECMWF.DAT.DT\_CLIMATE.ipynb

Preview Code Blame 404 lines (404 loc) · 250 KB

### EarthKit

Lets plot the result file [EarthKit Documentation] <https://earthkit-data.readthedocs.io/en/latest/index.html>

This section requires that you have 'ecCodes >= 2.35' installed on your system. You can follow the installation procedure at <https://confluence.ecmwf.int/display/ECC/ecCodes+installation>

```
In [23]: import earthkit.data
import earthkit.maps
import earthkit.regrid

In [24]: data = earthkit.data.from_source("file", filename)
data.ls
earthkit.maps.quickplot(data, #style=style
)
```

Surface pressure  
Base time: 00:00 on 2028-06-10 Valid time: 00:00 on 2028-06-10 (T+None)

4.9e+04 5.6e+04 6.3e+04 7e+04 7.7e+04 8.4e+04 9.1e+04 9.8e+04 1.05e+05  
Surface pressure (Pa)





HDA



- Jeden zestaw haseł
- Jedno API
- Setki kolekcji

Edge Services - Na wniosek



- Wizualizacja
- Rozproszone i równoległe procesy
- Dostęp przez przeglądarkę internetową



- Moc obliczeniowa – **w tym GPU**
- Obliczenia w bezpośrednim sąsiedztwie danych
- Storage s3



- Oszczędność czasu – mniej kodowania





**Destination Earth**

**Dziękuję za uwagę**

[pgrzybowski@cloudferro.com](mailto:pgrzybowski@cloudferro.com)



**Destination Earth**

Funded by  
the European Union



Implemented by

