Destination Earth

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

Autorzy: Patryk Grzybowski[1]; Aubin Lambare[2]; Christoph Reimer[3]; Michael Schick[4]; Danaele Puechmaille[4]

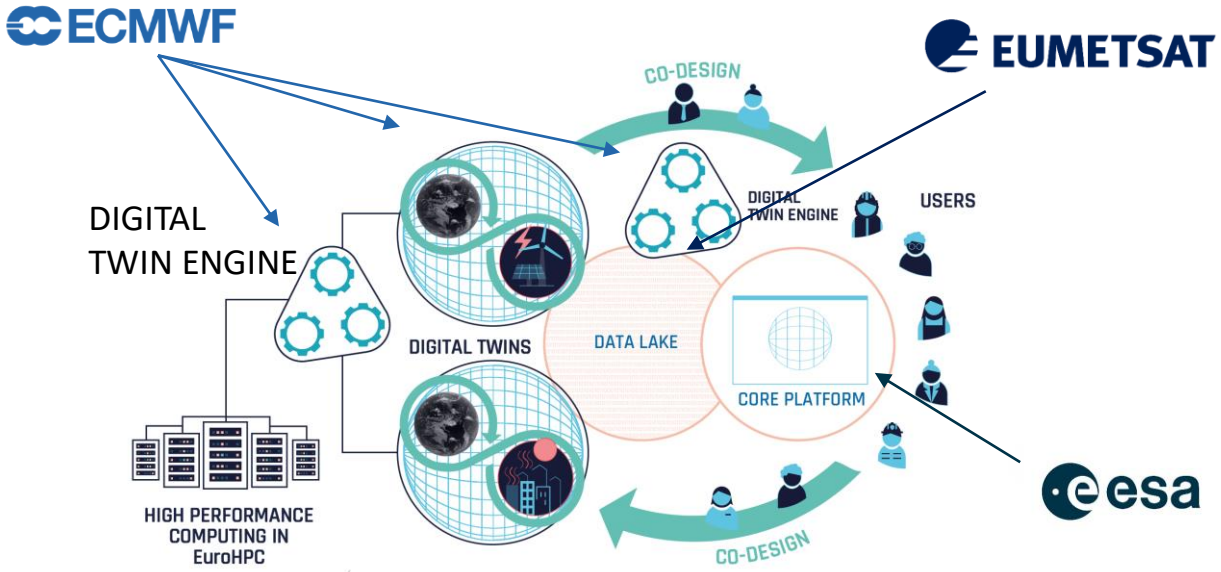
Afiliacje: [1]CloudFerro S.A., Warsaw, Poland; [2]CS Group - Sopra Steria, Le Plessis Robinson, France; [3]EODC, Vienna, Austria; [4]EUMETSAT, Darmstadt, Germany

Destination Earth Funded by the European Union Implemented by ECMWF esa EUMETSAT

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.

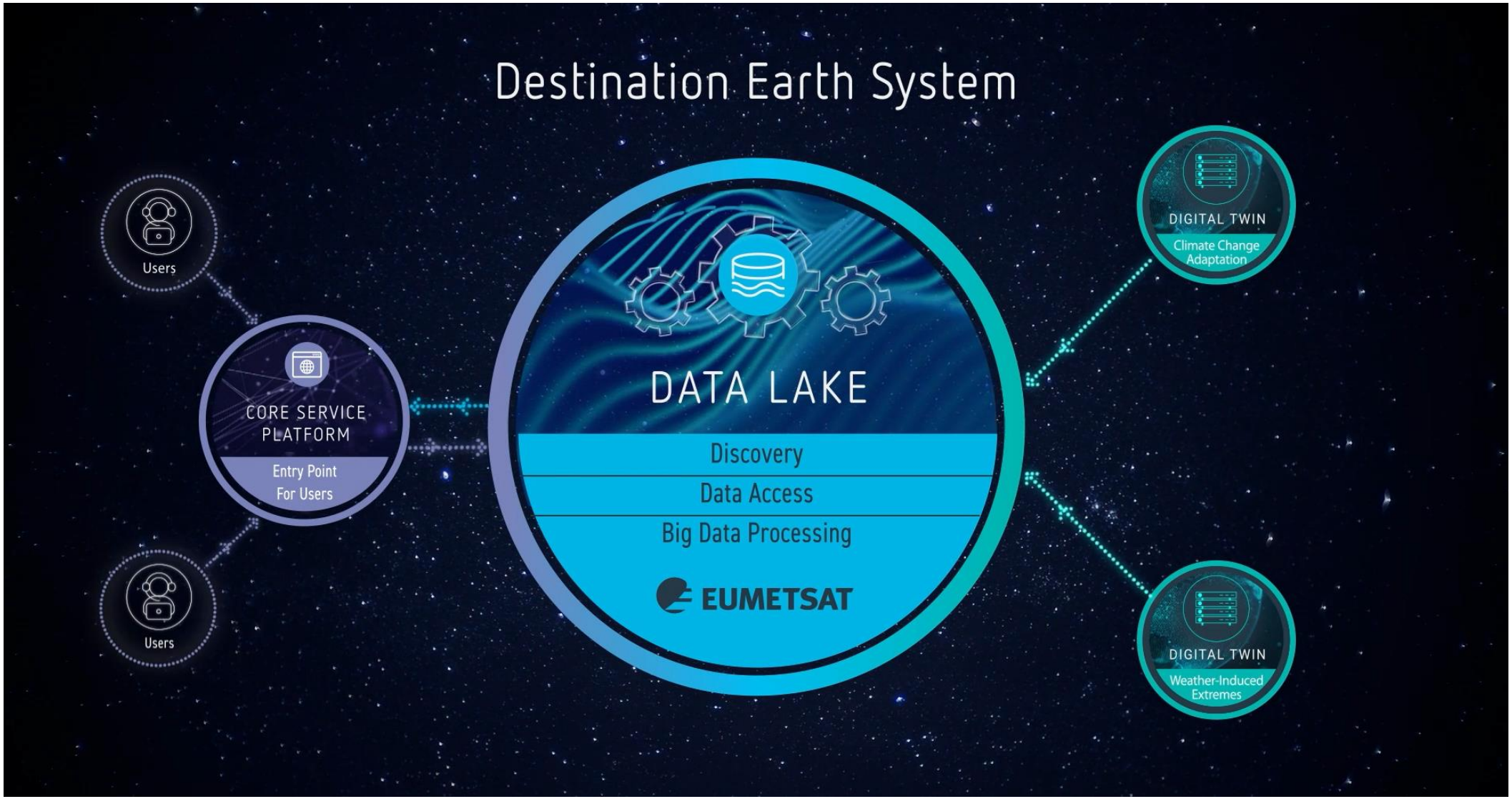
Trzy komponenty DestinE:

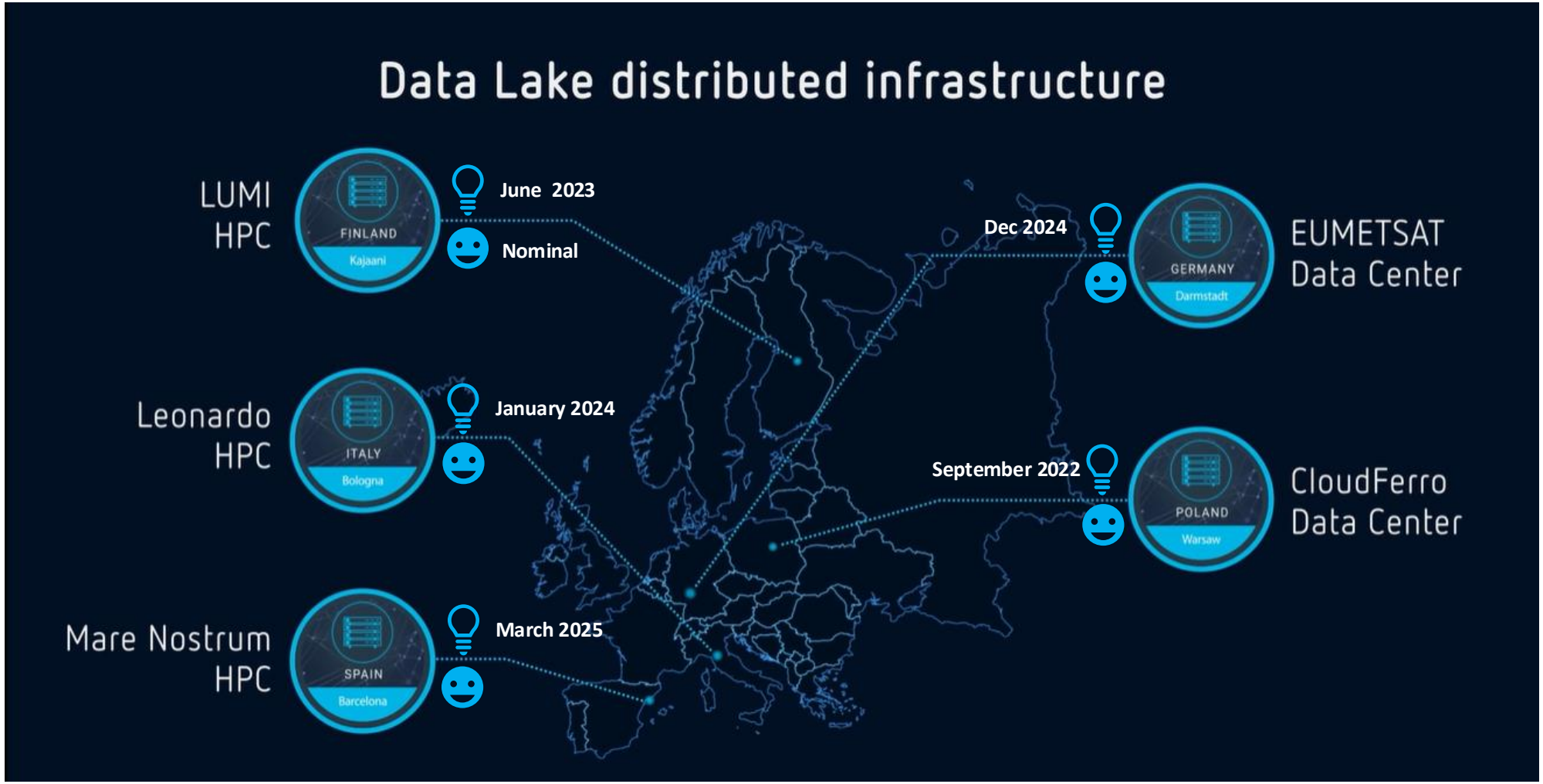
- Core Service Platform (DESP)
- Data Lake (DEDL)
- Digital Twins (DTs)



(ECMWF, 2024)

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


Destination Earth System
Users
Users
CORE SERVICE PLATFORM
Entry Point For Users
DATA LAKE
Discovery
Data Access
Big Data Processing
EUMETSAT
DIGITAL TWIN
Climate Change Adaptation
DIGITAL TWIN
Weather-Induced Extremes


Data Lake distributed infrastructure
LUMI HPC
FINLAND
Kajaani
June 2023
Nominal
Leonardo HPC
ITALY
Bologna
January 2024
Mare Nostrum HPC
SPAIN
Barcelona
March 2025
Dec 2024
GERMANY
Darmstadt
EUMETSAT Data Center
September 2022
POLAND
Warsaw
CloudFerro Data Center

HDA

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

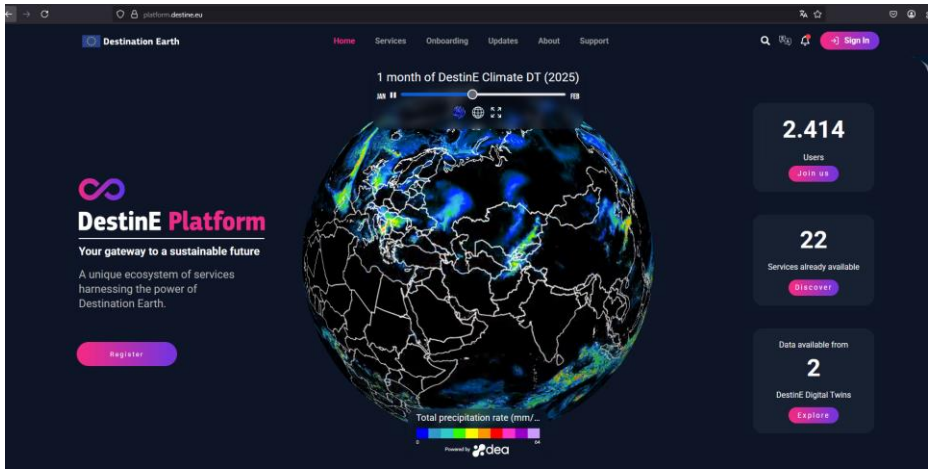
Edge Services – Na wniosek

STACK: usługa PaaS, oferująca obliczenia obok danych, a także:
- JupyterHub,
- Dask/ Dask Gateway

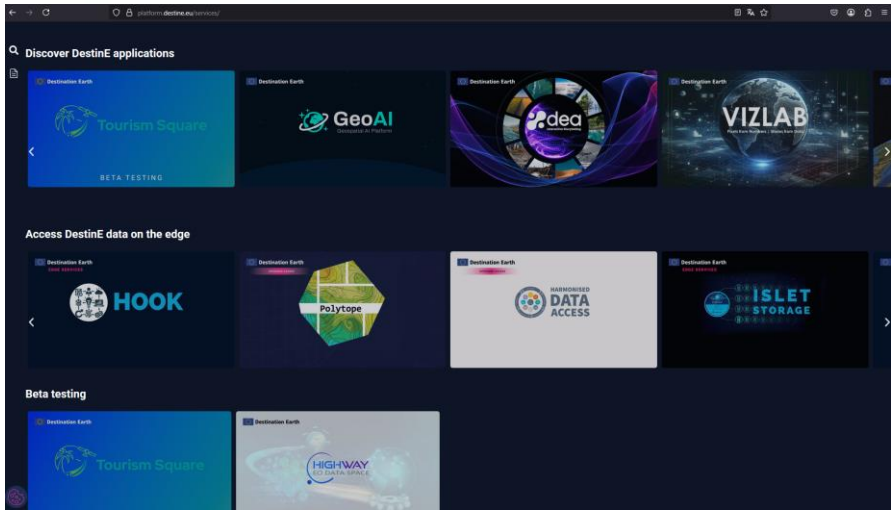
- ISLET Compute: (IaaS/PaaS) umożliwia przetwarzanie danych w ich bezpośrednim sąsiedztwie, pozwalając użytkownikom zarządzać i wdrażać wirtualne środowiska pracy
- ISLET Storage: usługa s3 object storage, do przechowywania danych użytkowników oraz wyników przetwarzania.

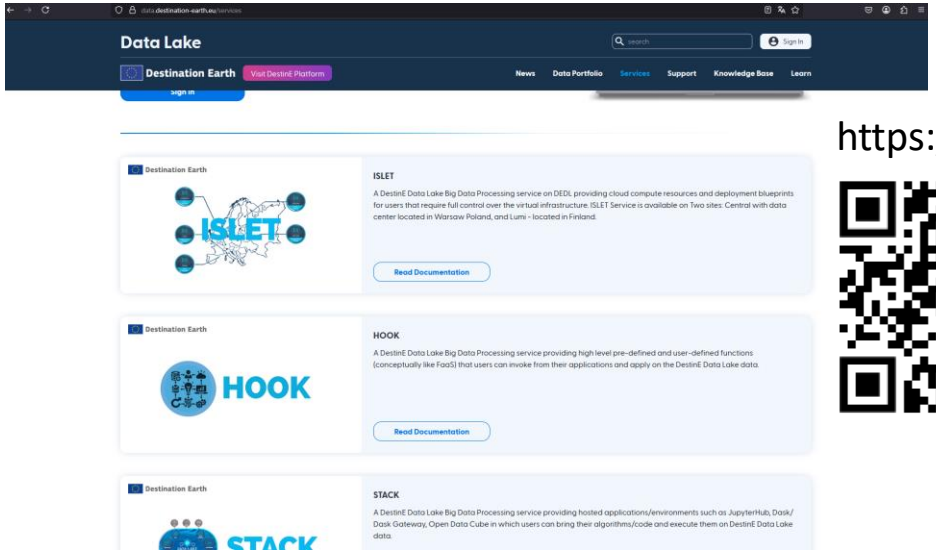
- HOOK - pozwala na uruchamianie złożonych, predefiniowanych lub niestandardowych workflow.

Jak zostać użytkownikiem DEDL?

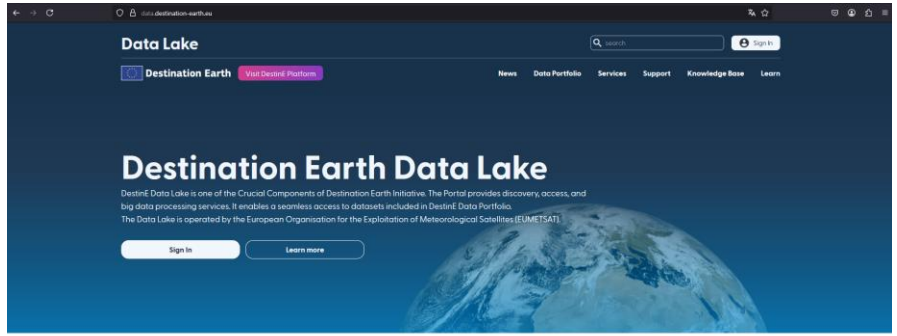


https://platform.destine.eu/





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



HDA – Harmonized Data Access

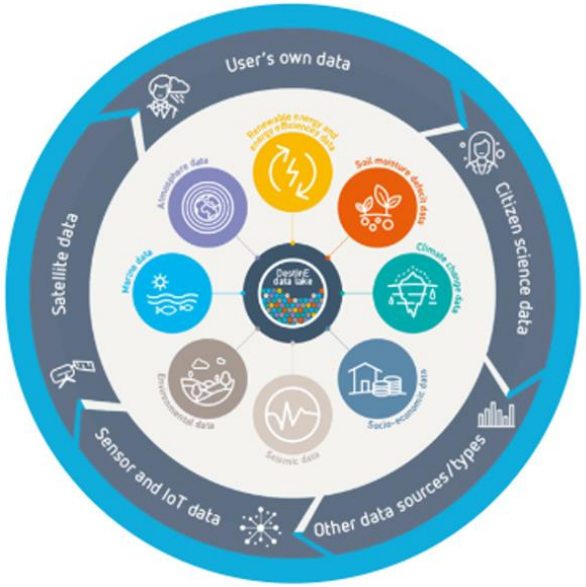
HARMONISED DATA ACCESS

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

Około 200 kolekcji

Jeden zestaw haseł

Różni dostawcy danych

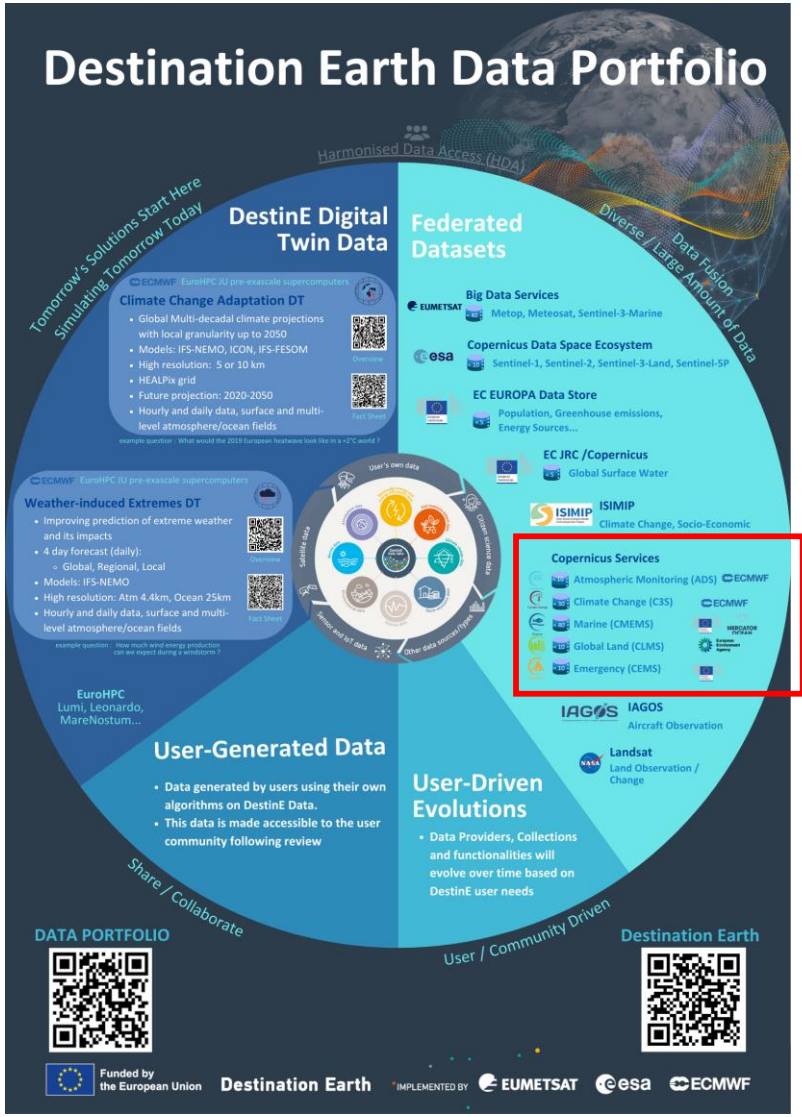


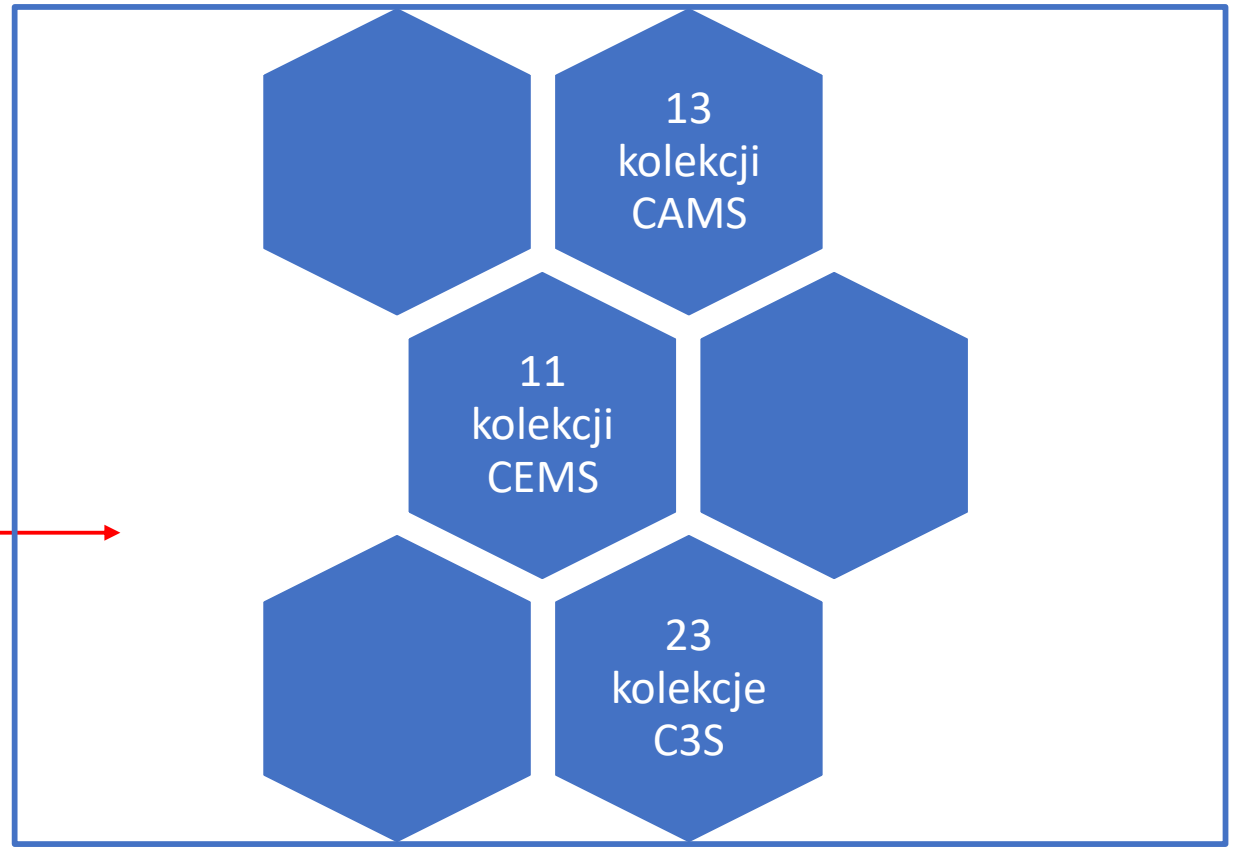
HARMONISED DATA ACCESS

Dane Digital Twins

Przegladanie i dostęp do danych

STAC


Destination Earth Data Portfolio
Harmonised Data Access (HDA)
Tomorrow's Solutions Start Here
Simulating Tomorrow Today
Diverse / Large Amount of Data
Data Fusion
DestinE Digital Twin Data
Federated Datasets
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
Big Data Services
Metop, Meteosat, Sentinel-3-Marine
Copernicus Data Space Ecosystem
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
Climate Change, Socio-Economic
Copernicus Services
Atmospheric Monitoring (ADS)
Climate Change (C3S)
Marine (CMEMS)
Global Land (CLMS)
Emergency (CEMS)
IAGOS
Aircraft Observation
Landsat
Land Observation / Change
EuroHPC
Lumi, Leonardo, MareNostum...
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
User / Community Driven
DATA PORTFOLIO
Destination Earth
Funded by the European Union
Destination Earth
IMPLEMENTED BY EUMETSAT esa ECMWF


13 kolekcji CAMS
11 kolekcji CEMS
23 kolekcje C3S

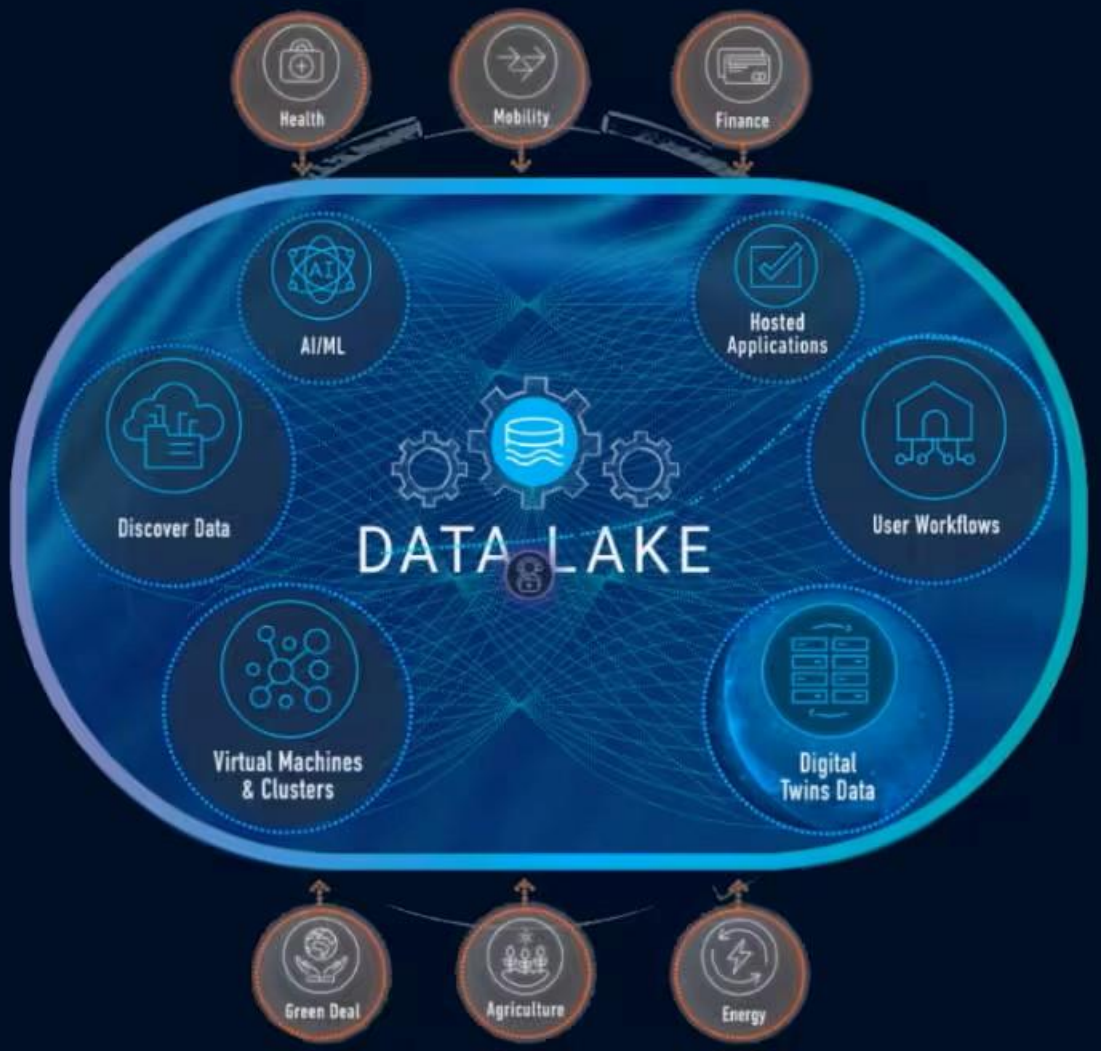
HDA – Harmonized Data Access

HARMONISED DATA ACCESS

- JupyterLab w ramach DEDL STACK Service
- Maszyna wirtualna w ramach DEDL Islet Service
- Środowisko lokalne

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

Destination Earth



STACK

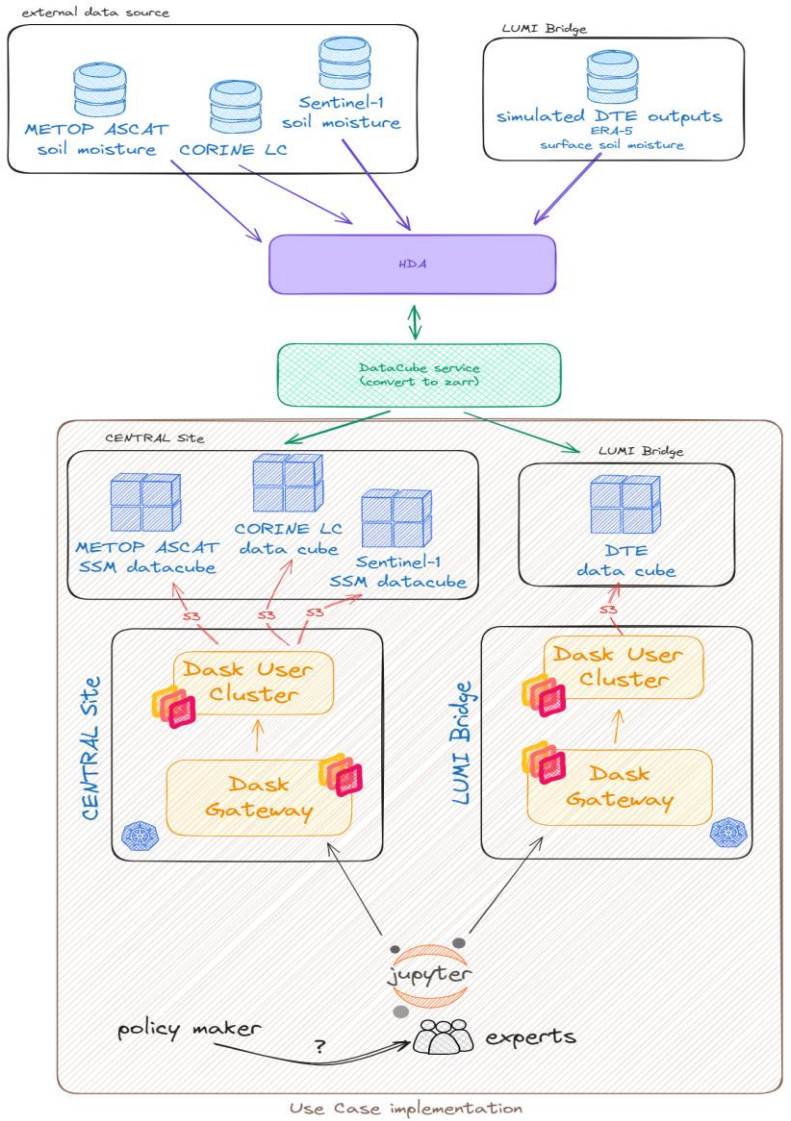
Data Bridges

STACK

JupyterLab

DASK clusters

HDA: restAPI
PySTAC
EODAG

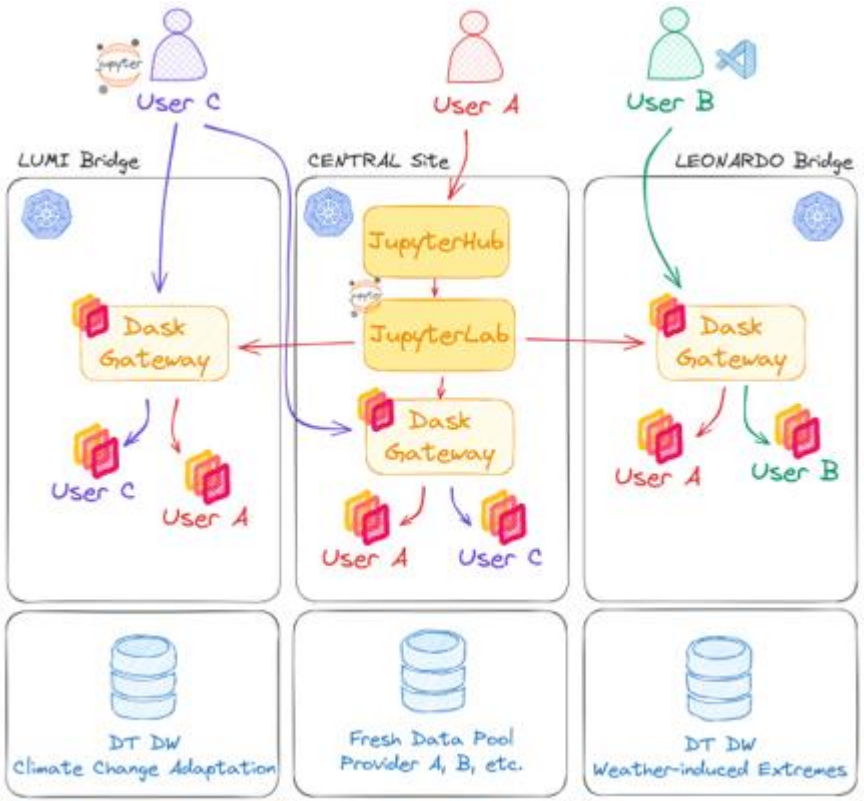


Funded by the European Union

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



dask

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**

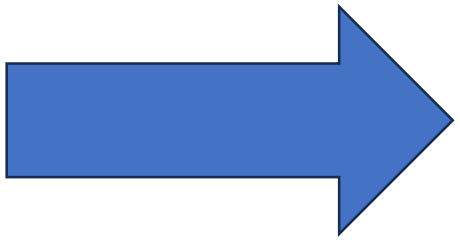
Destination Earth

HOOK

Predefiniowane procesory

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

Procesory tworzone przez użytkowników



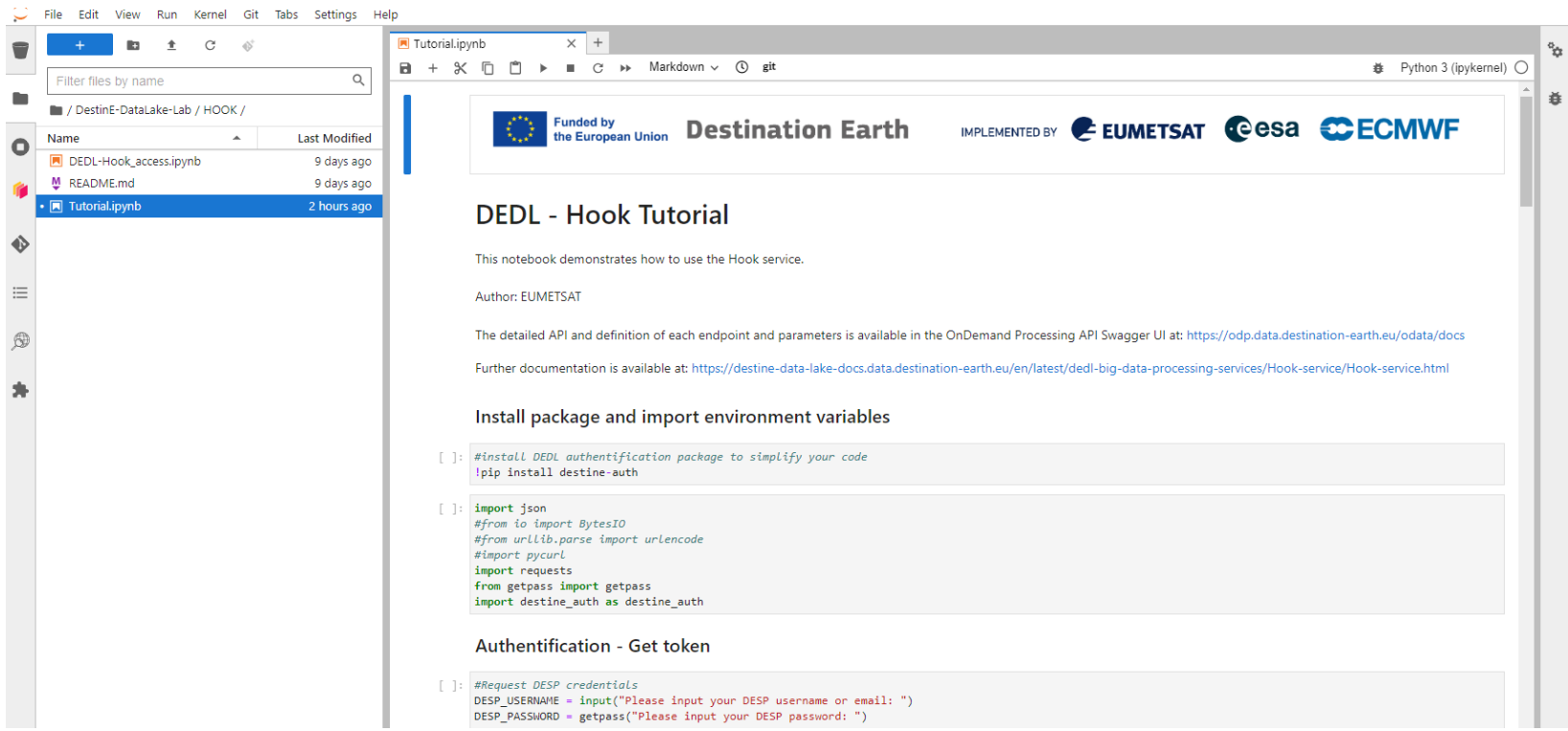
Prywatny s3 storage
OR
Tymczasowy s3 storage

Komunikacja przez API:

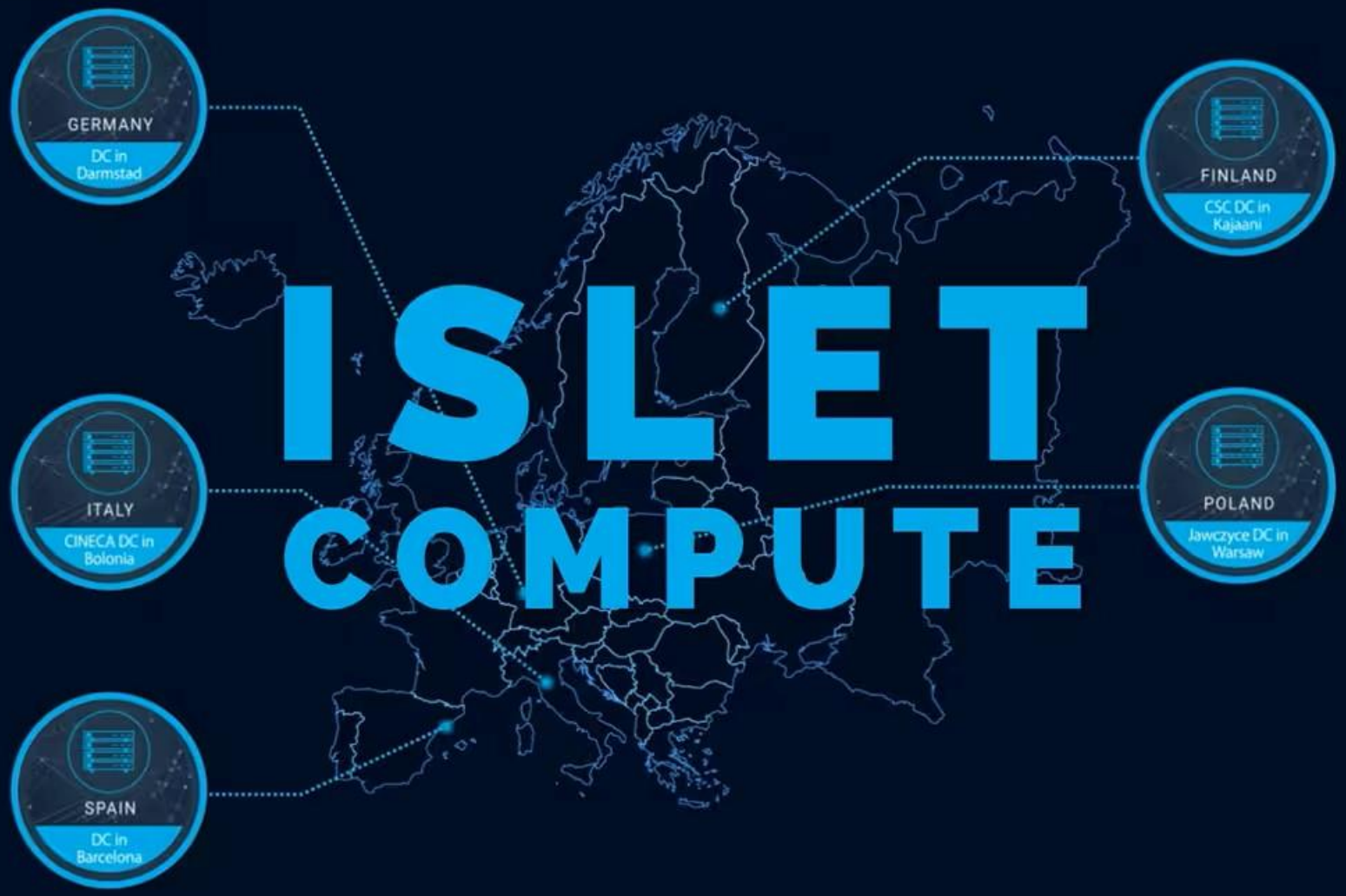
- REST API
- openEO API

Komunikacja przez interfejs graficzny:

- JupyterLab (STACK)
- openEO (UI)



Destination Earth



Wirtualne maszyny i wirtualne środowiska

Gotowe obrazy systemów operacyjnych

CPU i GPU

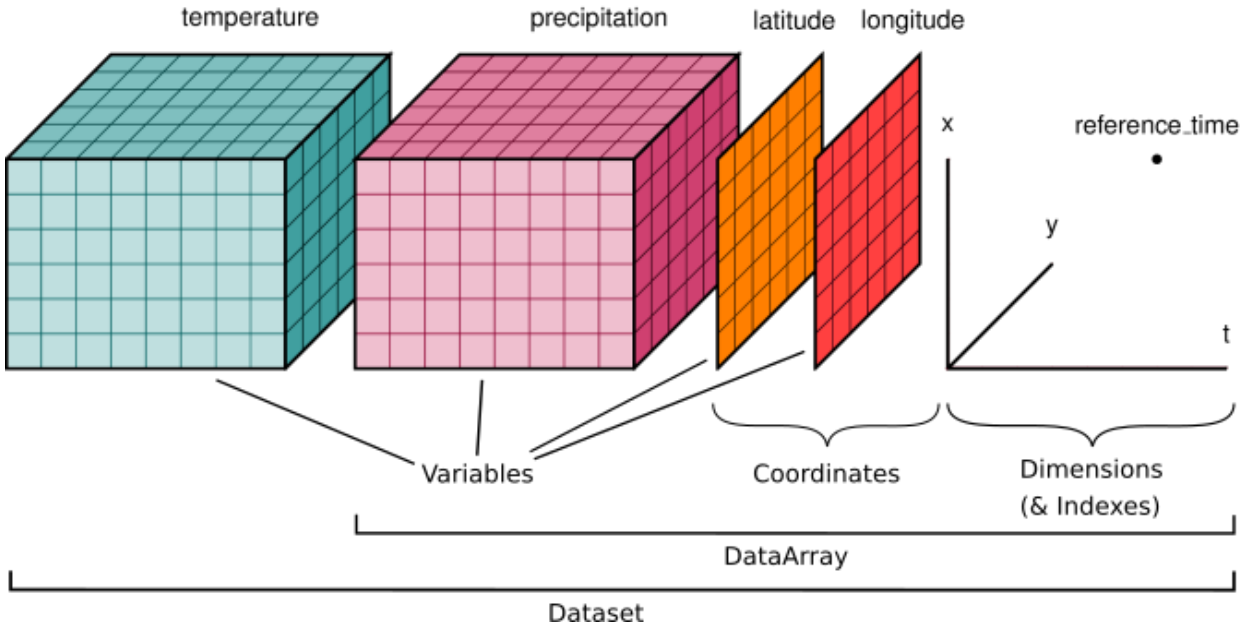
Klastry K8s

Dostęp s3 storage

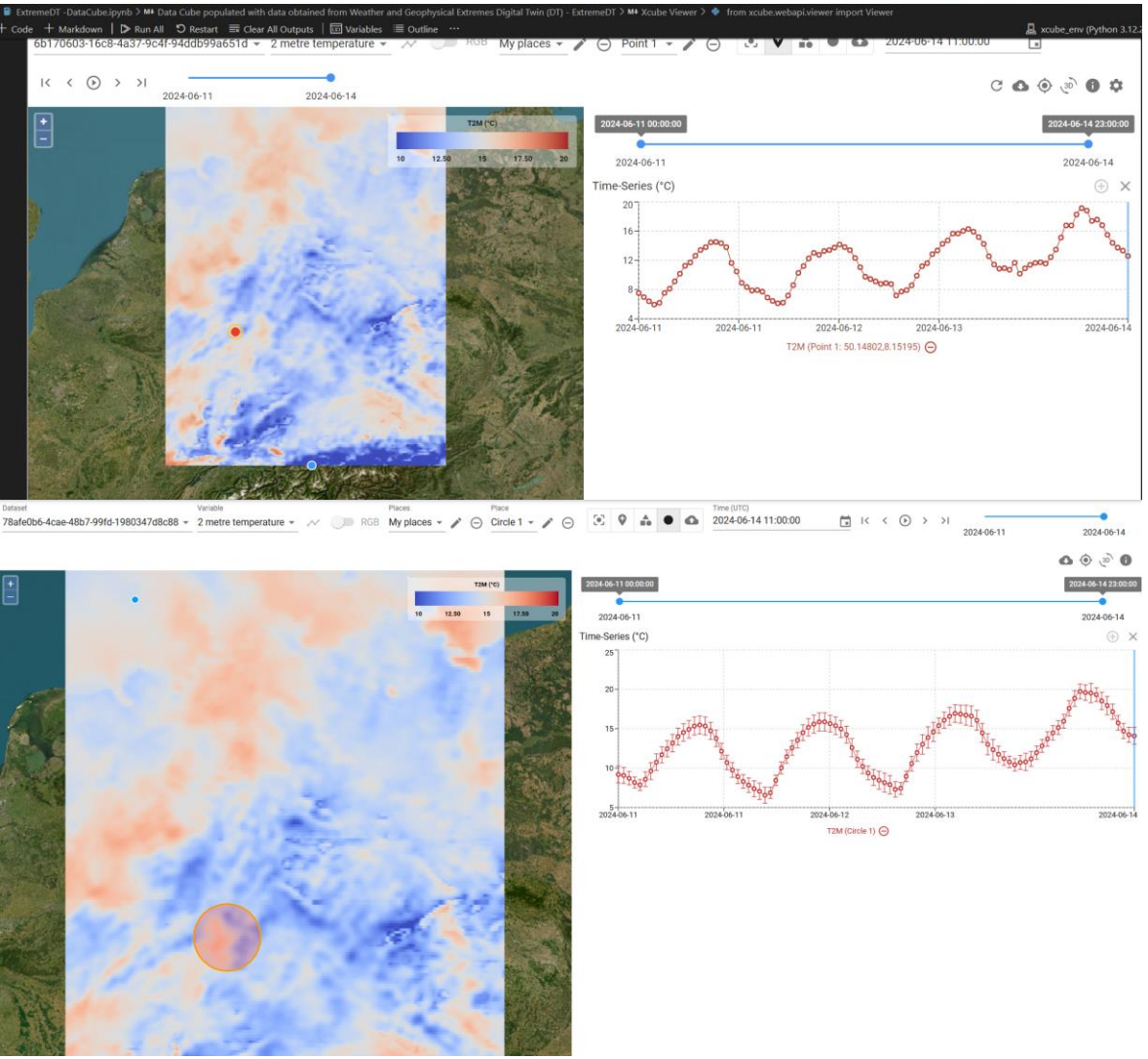
Zarządzanie s3 storage

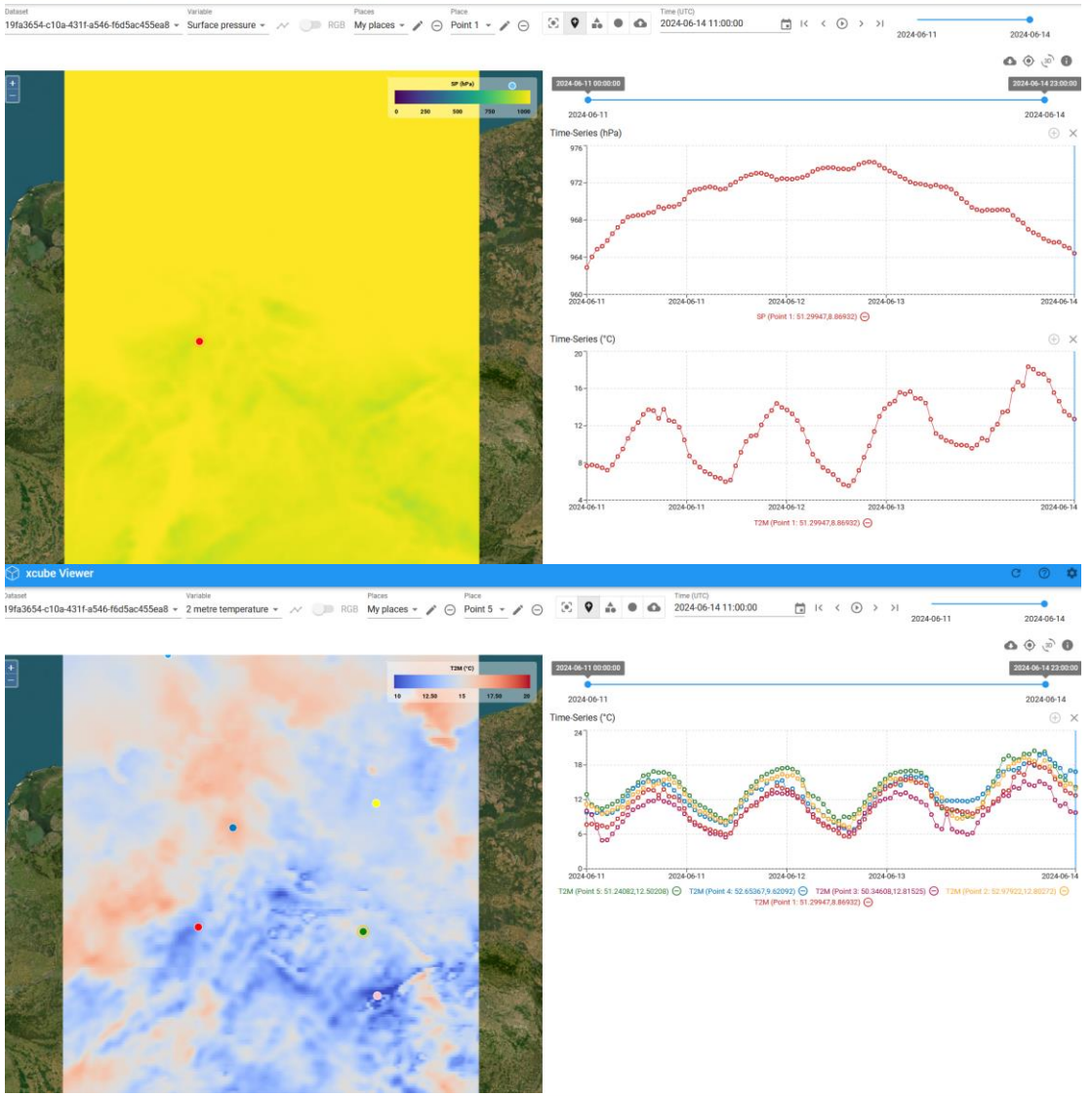
Dlaczego nam to wszystko potrzebne?

xarray



Dane w natywnym formacie → .zarr/.netcdf → Python





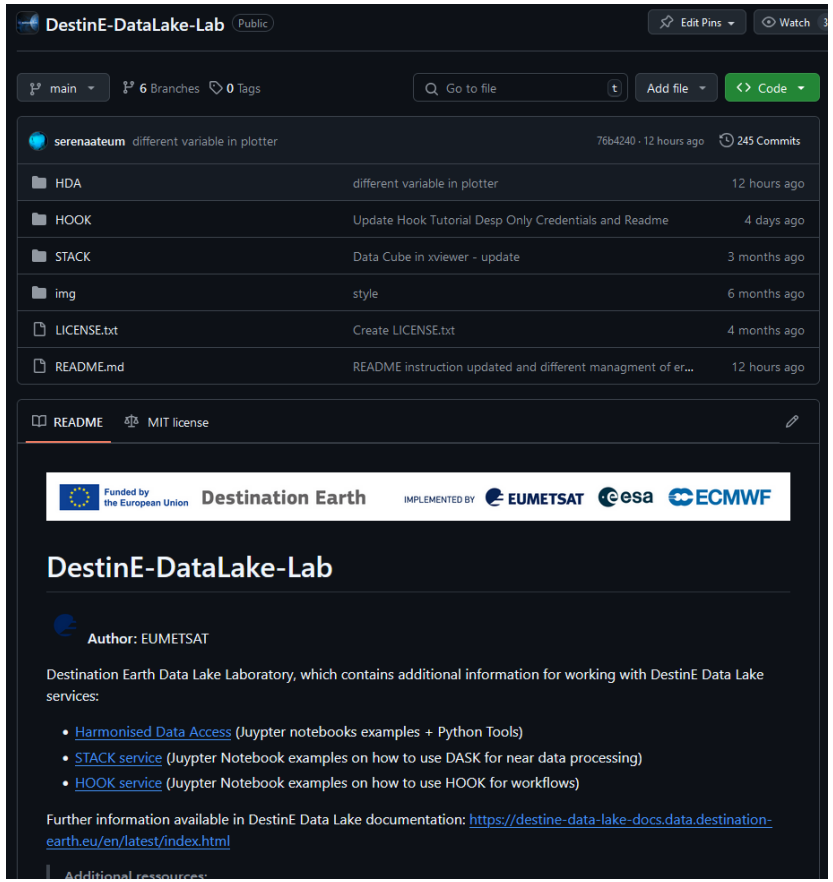
Funded by the European Union

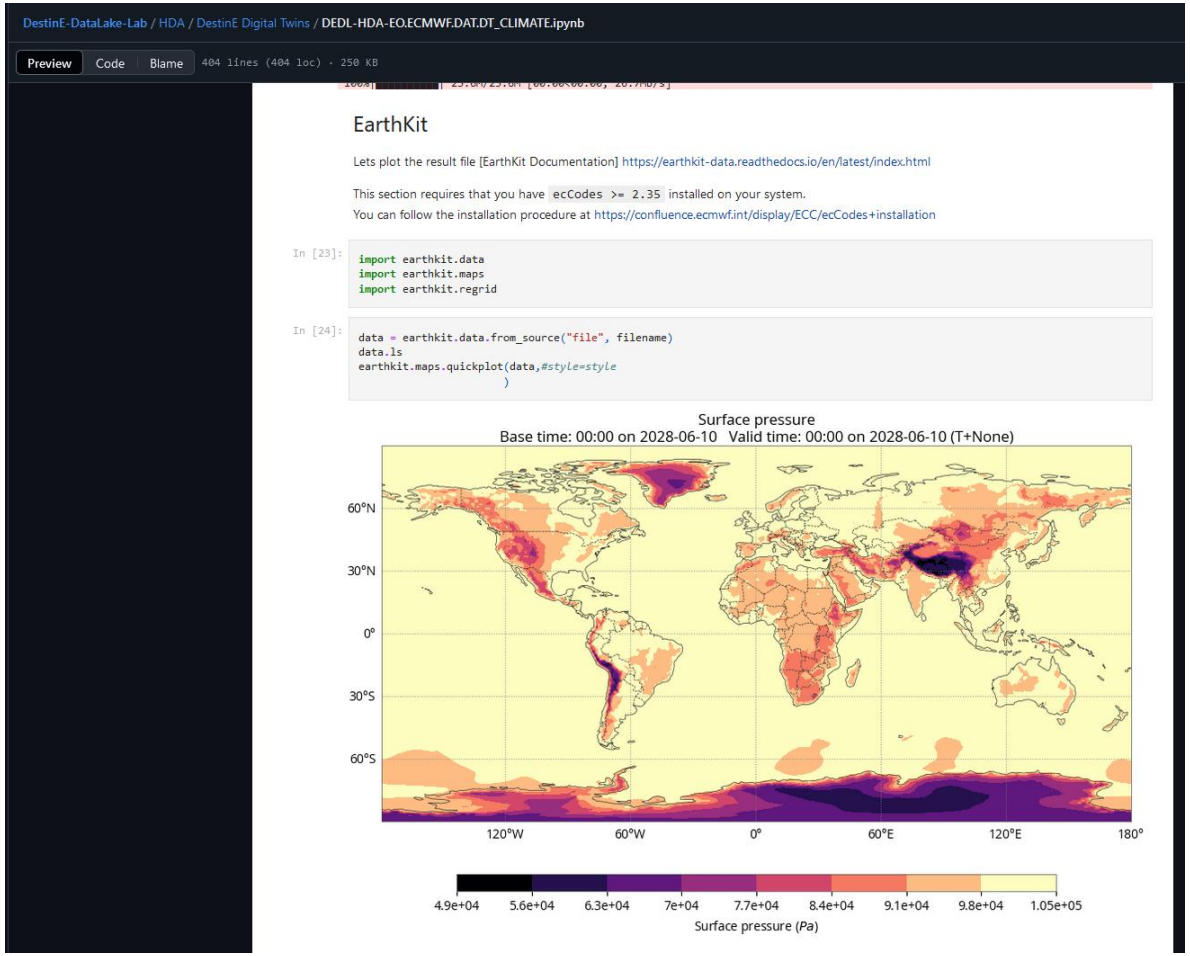
IMPLEMENTED BY EUMETSAT esa ECMWF

Wniosek o dostęp do ‚Edge Services'

Edge services

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





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

Destination Earth Funded by the European Union Implemented by ECMWF esa EUMETSAT