EUMETSAT

Funded by the European Union

Destination Earth

IMPLEMENTED BY EUMETSAT esa ECMWF

The importance of data lakes in the context of digital Twins

Jordi Duatis, Michael Schick, Danaële Puechmaille
EUMETSAT

Advancing in the digital transformation of our community, from the European Weather Cloud to AI/ML and Big Data

EuoGeo Workshop 2023 – Bolzano, Italy 2-4 October 2023

Destination Earth

A Highly Accurate Digital Model of the Earth

European Commission

To monitor, simulate and predict natural phenomena and the impact of human activity on Earth

To assist in designing accurate adaptation strategies and climate change related mitigation measures

To accelerate the EU's green and digital transition

To leverage existing and new data sources and EU's advanced digital and computing infrastructure

To create and test "what if" scenarios and to integrate impact sector applications for more sustainable development

To support near real-time decision-making at various levels (e.g. EU, national, regional, local)

To go beyond the current complex systems designed mainly for expert use

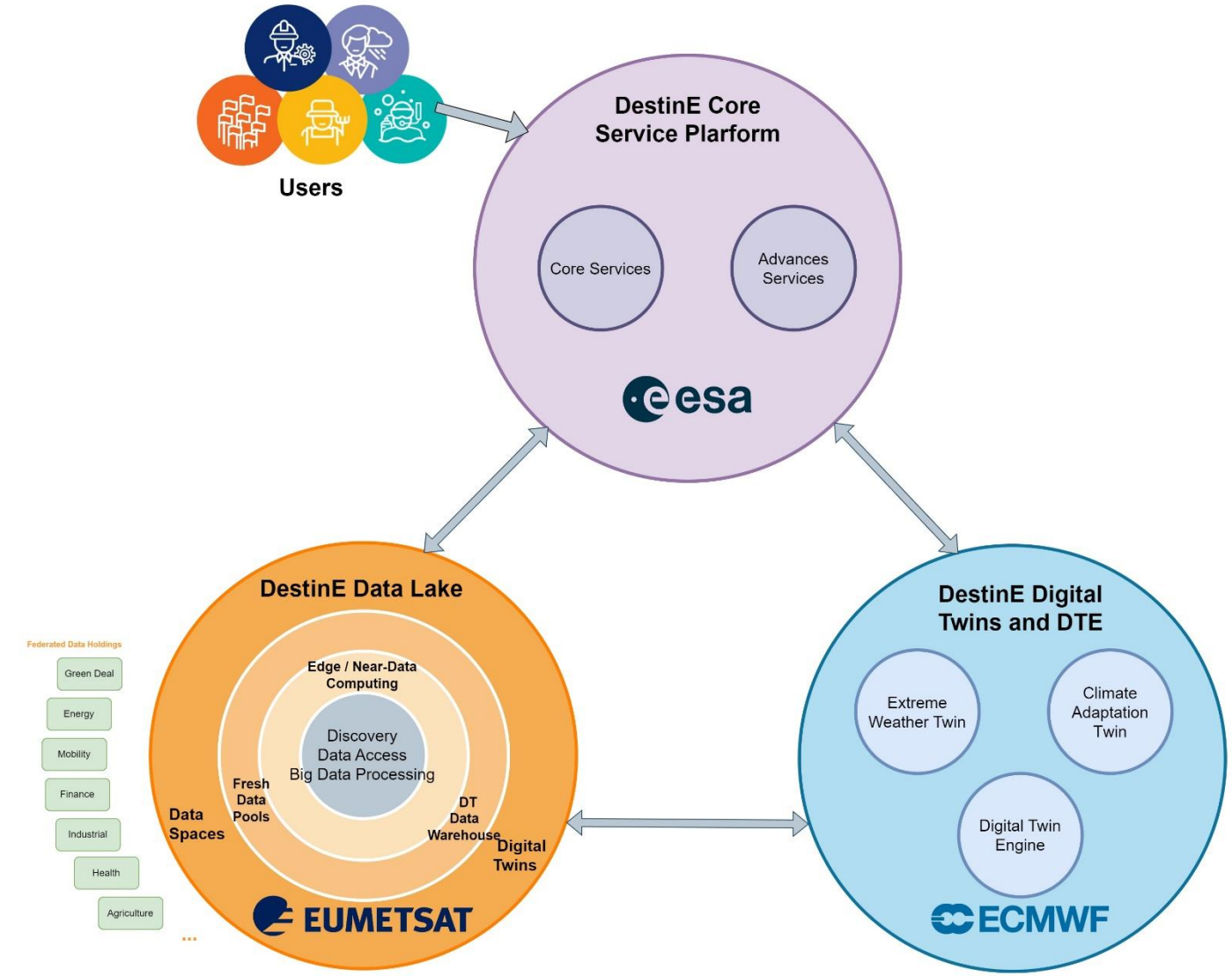
To scale up existing models and fuse simulation with observation

Implemented by ECMWF esa EUMETSAT

DestinE: A joint undertaking of ESA, ECMWF and EUMETSAT

Three entrusted entities implementing DestinE

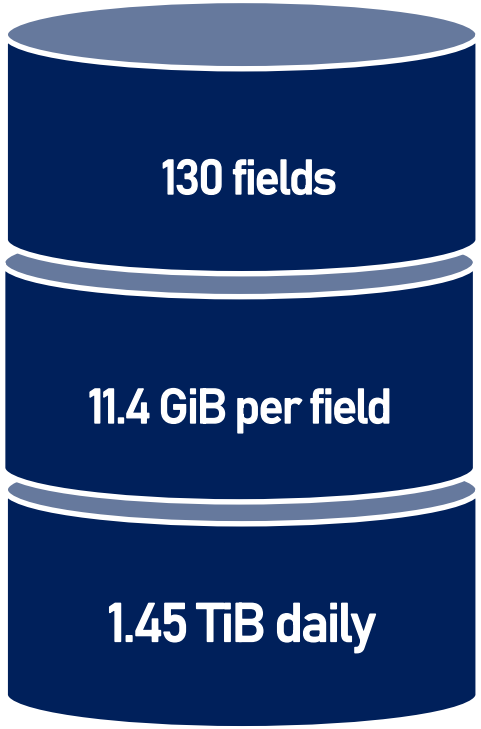
- Core Service Platform interfacing DestinE users (ESA)
- Two Digital Twins. Extreme Weather and Climate Change Adaptation (ECMWF)
- Destination Earth Data Lake (EUMETSAT)



DestinE Digital Twins Data Volume / Data Portfolio

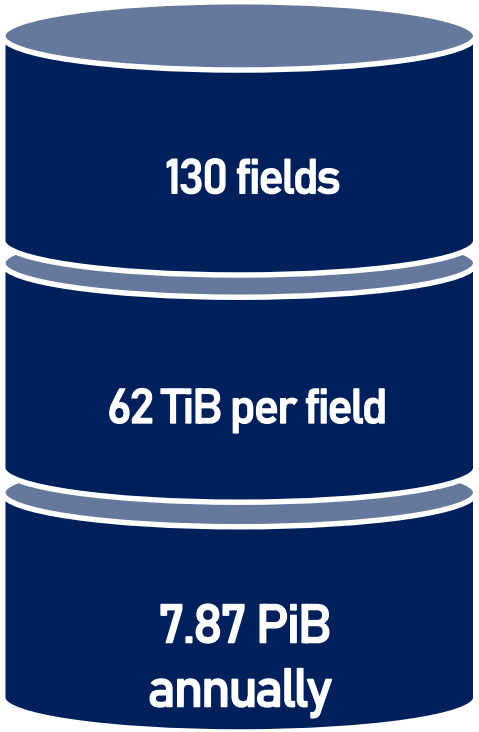
DT on Weather-induced Extremes

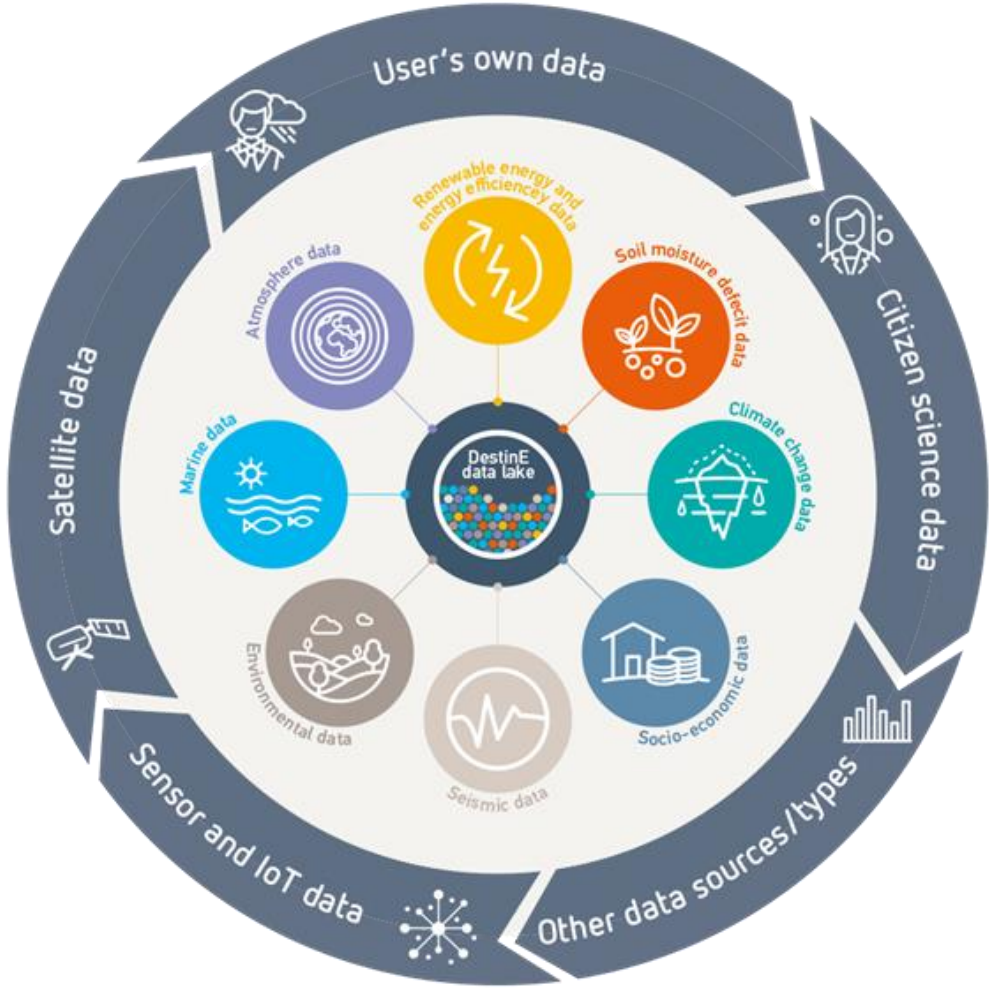
Temporal resolution: 15 minutes to 1 hour
Time horizon: 4-7 days forecast
Horizontal resolution: 4.4/2.8/1.4 km
Number of instances: 1



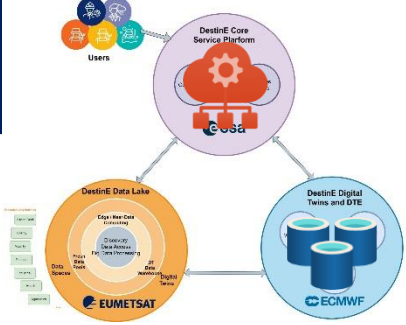
DT on Climate Adaptation

Temporal resolution: 1 hour to monthly
Time horizon: Multi-decadal
Horizontal resolution: 9/4.4/2.8 km
Number of instances: 2-3 models x 70 years (control, historical, future years)

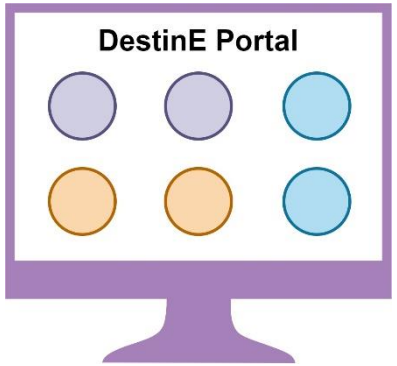




DestinE Services offer



Experts



Solicited experts have available through DestinE:

DATA

- DT Output (beyond what exists today):
  - Routine execution (global)
    - Higher temporal resolution
  - On-demand (regional/ local)
    - Higher spatial resolution
  - new fields
- Easy access to federated "data spaces":
  - Earth Observation data (satellite)
  - Socio-economic statistic data
  - IoT, Sensors
  - And many more
- Frequently used data available in Fresh Data Pool
- Possibility to bring own data
- Data Cubes

COLLABORATIVE RESEARCH ENVIRONMENT

- Processing near data
  - Edge computing to work efficiently with DT Data
- Portfolio of Services/ Toolboxes
- AI/ML libraries
- On-demand resources (CPU, vGPU, Storage)

ADVANCED SERVICES

- Visualisation tools
- Climate change applications and tools for end user
- AI/ML Frameworks
- …

DestinE Data Lake

Destination Earth Data Lake (DEDL)

Self-standing component

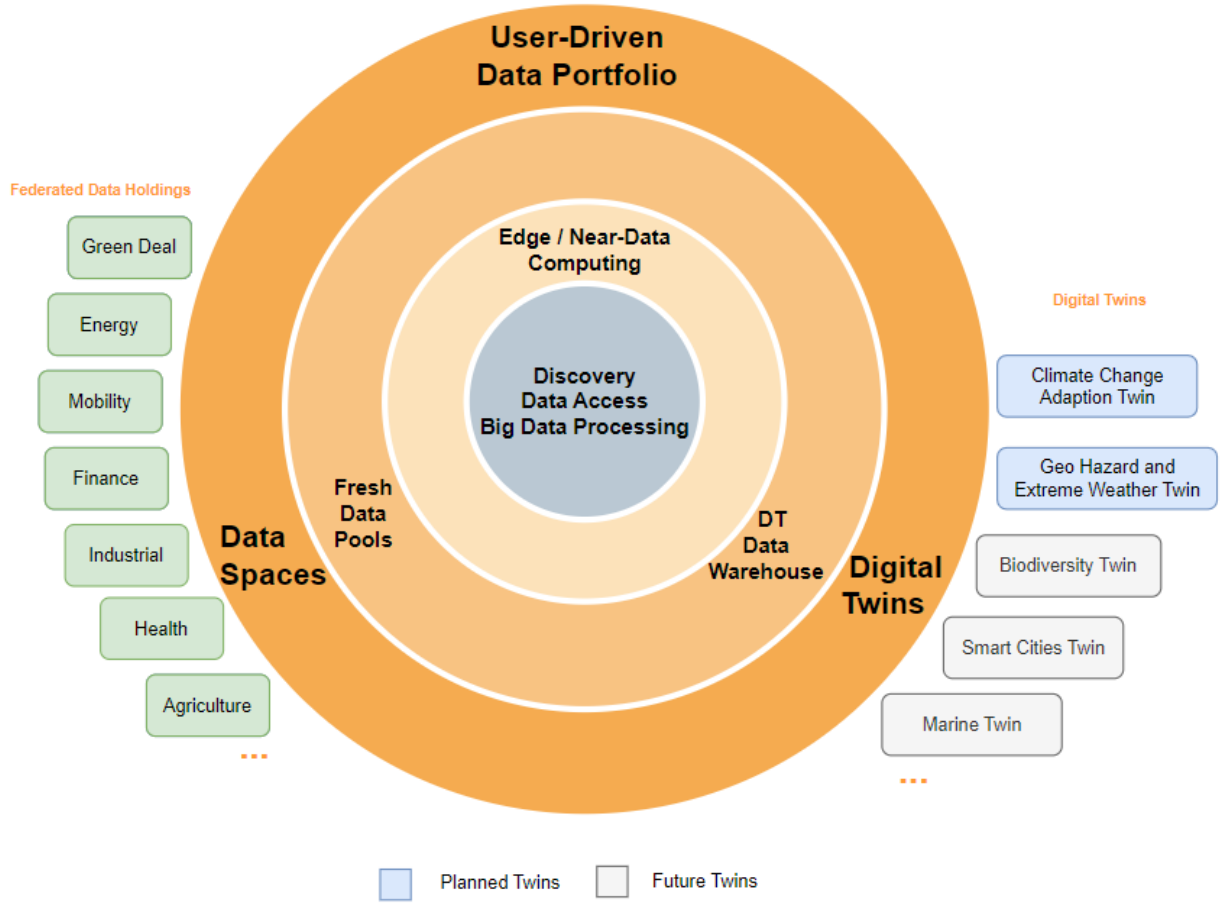
- Built from geographically distributed physical elements (central & edges)
- Distributed services – seamless access

Discovery & Data Access

- Harmonisation of data access (HDA) to simplify data discovery & access
- External federated data spaces
- Digital Twin data (ECMWF):
  - Extreme Weather and Climate Change Adaptation
- DestinE User generated data

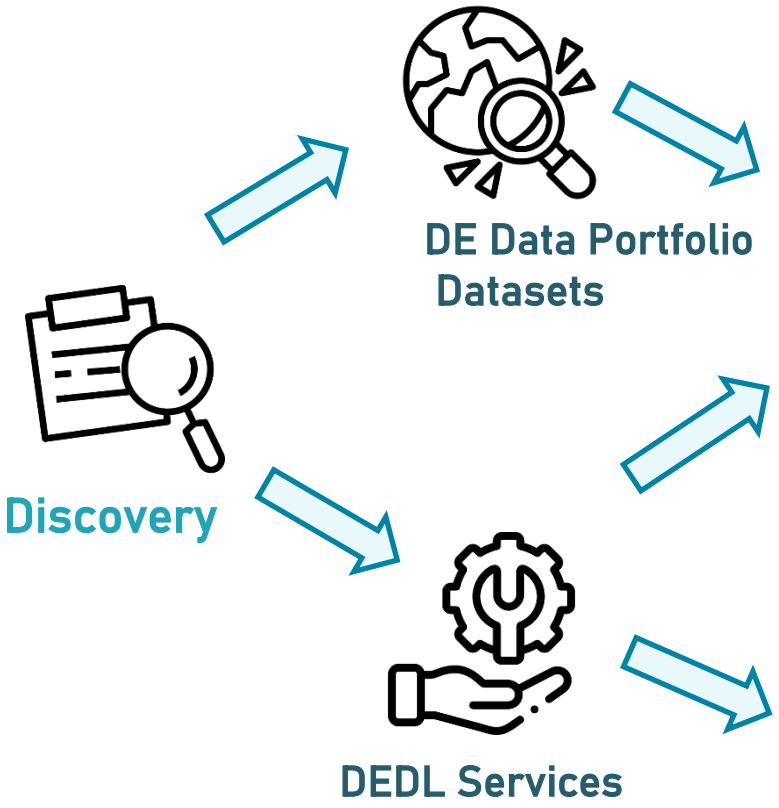
Big Data Processing

- Processing near data including distributed computing & workflows

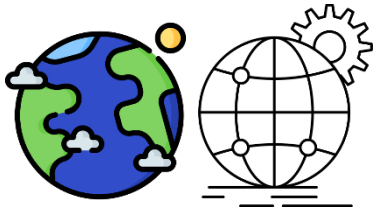


Discovery & Data Access Services

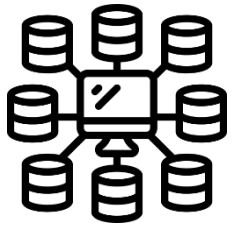
Discovery Services



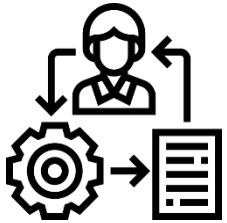
Data Access Services



Digital
Twin Outputs



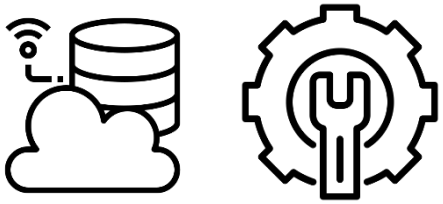
Federated
Datasets



User-
Generated Data

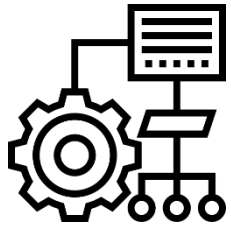
Fresh
Data Pool

Big Data Processing Services



Infrastructure& tools
(Islet Service)

Hosted Applications
(Stack Service)

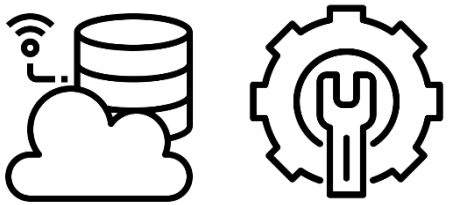


Functions
(Hook Service)

Icons made by DailyPM Studio, Phatplus, surang, juicy_fish, Superndre, Freepik, Eukalyp, flatart_icons, Muhammad_Usman from www.flaticon.com

DEDL – Big Data Processing Services

Infrastructure & Tools



Islet Service

- VMs, GPUs, Object Storage, k8s clusters
- blueprints (VMs, libraries & tools for data science and AI/ML)

For Users who

- **set up and manage their own development environment**
- **deploy already existing processing chains**

Hosted Applications

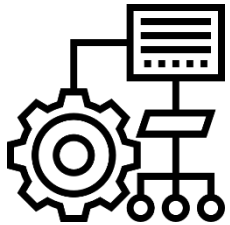
Stack Service

DEDL-provided off-the-shelf working environments and applications
(JupyterHub ecosystem, DASK Gateway)

For Users who

- **want ready-to-use applications and environments**

Functions



Hook Service

Predefined processing workflows/ functions
User-defined workflows
System or User-defined data cubes

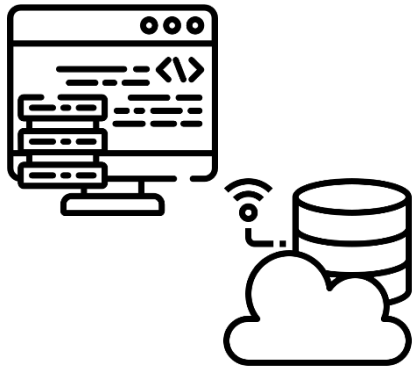
For Users who

- **want ready-to-use building blocks for their applications**
- **want advanced processing services**

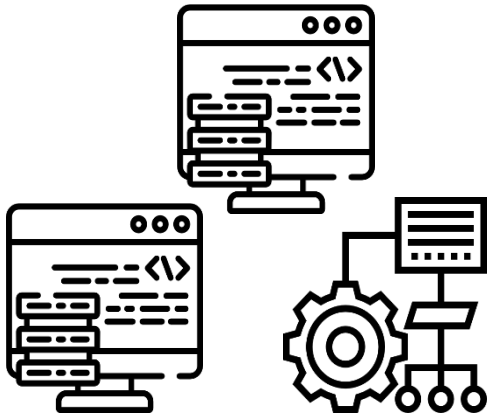
Icons made by Phatplus , Superndre, Freepik, Eukalyp from www.flaticon.com

DEDL – Big Data Processing Services

Users can pick and mix big data processing service offerings:

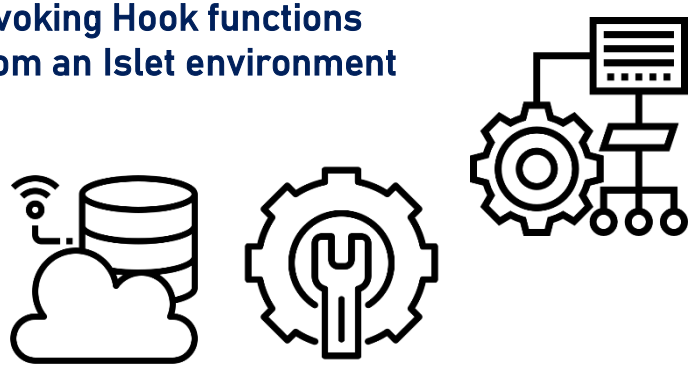


Stack (JupyterHub) + Islet-Storage (Uploading own data & Storing results)



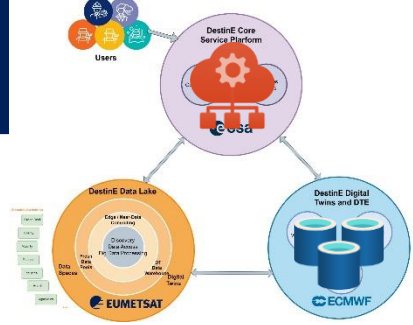
Using Stack application (e.g. DASK Gateway) + Hook functions in a Stack environment (JupyterHub)

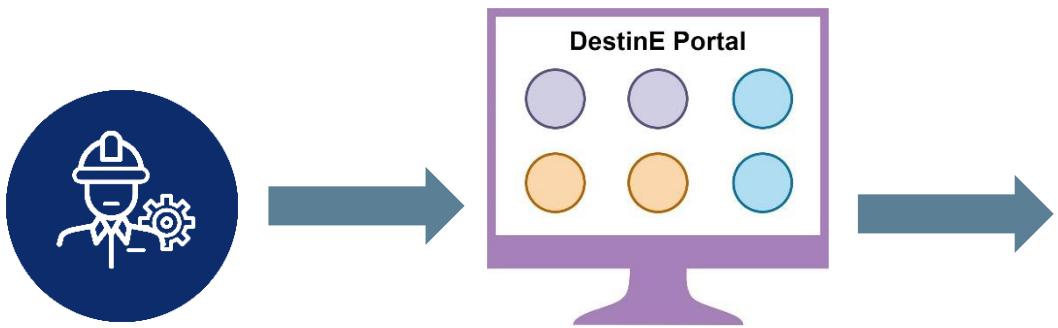
Invoking Hook functions from an Islet environment



Invoking Stack applications (e.g. DASK Gateway) in an Islet environment

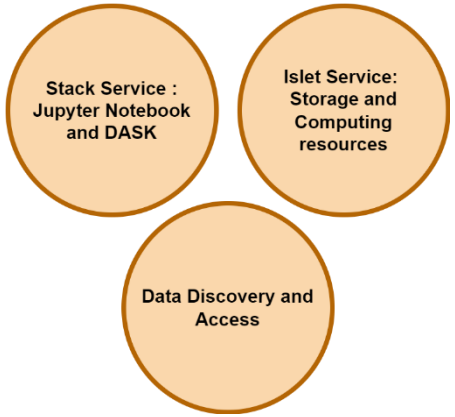
User select services (e.g. DEDL)





Experts

DestinE
Services offer



Discovery and Selection of suitable data and services

Data

- Predicted variables (rain intensity…)
- Sociographic Information (human settlement)
- Imagery of the area under investigation
- Algorithms (Flood predictions)

Benefits:
- Ready to use Apps and Functions
- Processing near data.
  Avoids transfer of data → reduce resource usage and speed up obtaining results
- Harmonized way to access data
- Collaboration development

DEDL Big Data Processing

Instantiation of DASK Cluster via gateway for **distributed workflows capabilities**:
- Processing DT Data – where they are stored
- Work with own and/or federated data – where they are stored

Thank you!

Questions are welcome.

Jordi.Duatis@eumetsat.int
Michael.Schick@eumetsat.int
Danaele.Puechmaille@eumetsat.int

Funded by the European Union Destination Earth