EUMETSAT

Funded by the European Union

Destination Earth

IMPLEMENTED BY EUMETSAT esa ECMWF

Destination Earth Data Lake – Efficient Use of Big Earth Data & Information

Miruna Stoicescu
Destination Earth Data Lake Services Engineer

EODC Forum, May 9th 2023

Agenda

Introduction to DestinE

Destination Earth Data Lake Concepts & Services

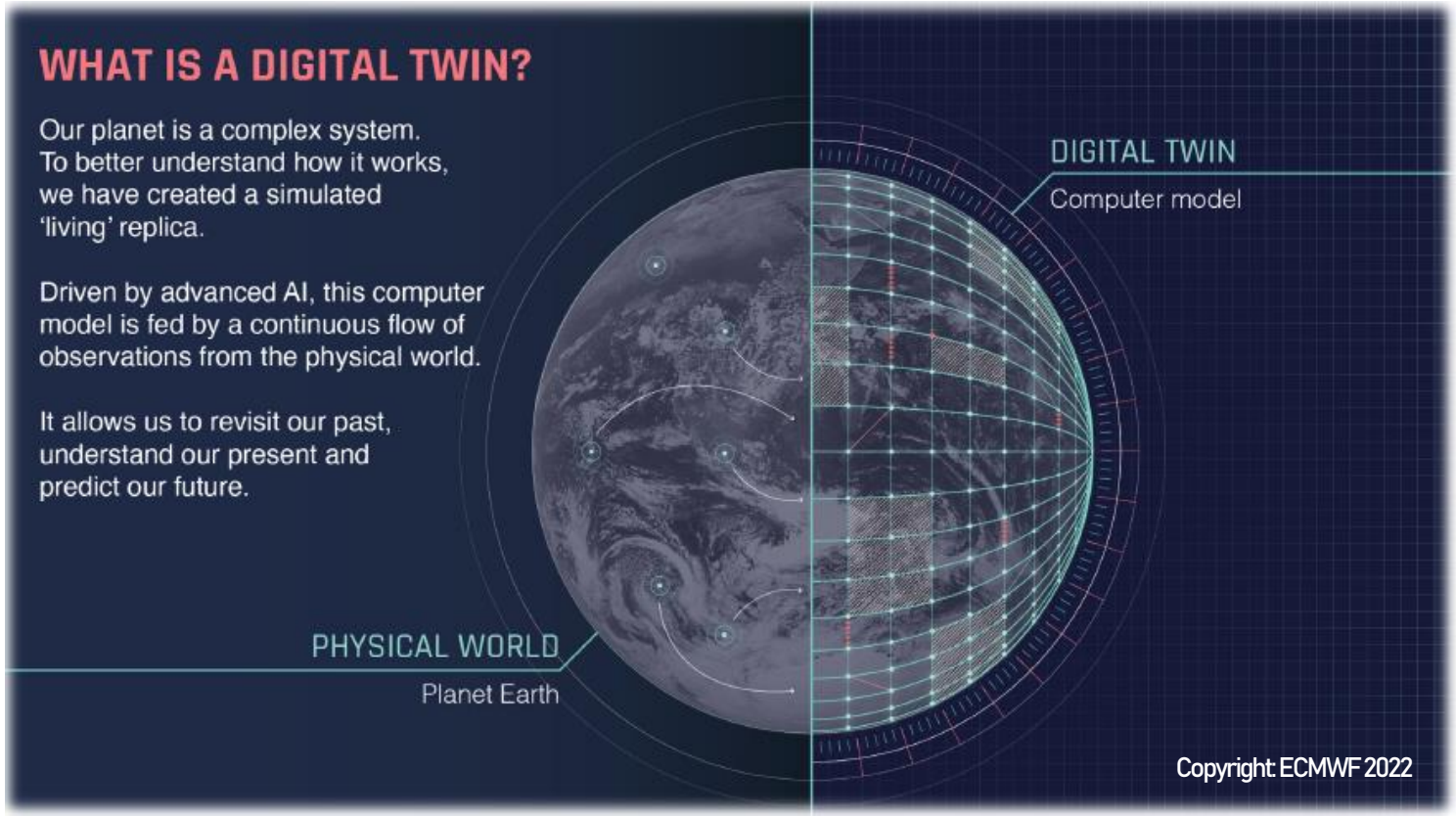
Introduction to DestinE

Destination Earth

- "*Destination Earth (DE) aims at developing a very high precision digital model of the Earth (Digital Twin of the Earth) to enable end-users to assess not only the impact of environmental and other societal challenges but also the efficiency of the proposed solutions, incl. EU legislative measures*."

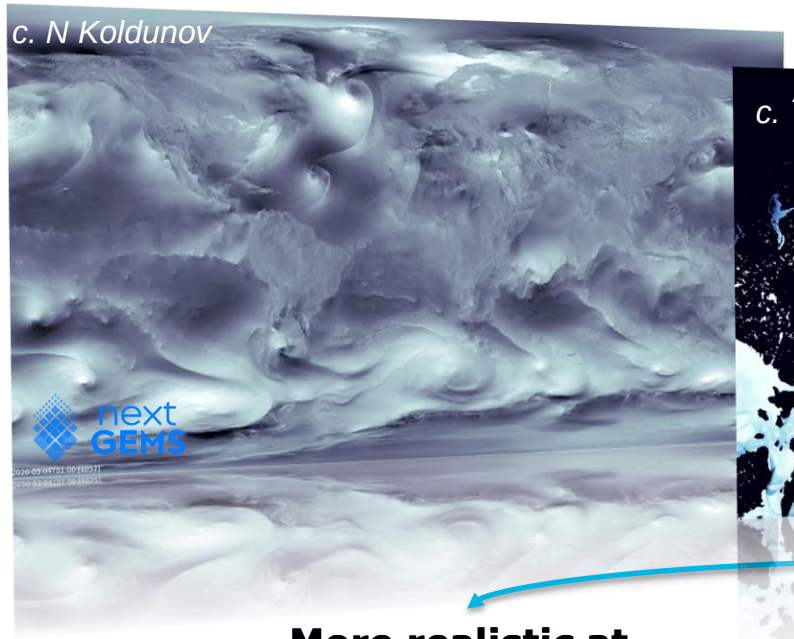
Part of the EU's

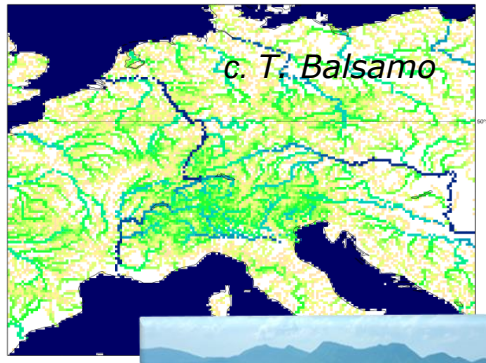
- Green Deal
- Europe's Digital future
- European strategy for Data



DestinE's Digital Twins: Quality + Impacts + Interaction

1. **Better simulations** based on **more realistic models**
2. Better ways of **combining all observed and simulated information** from entire Earth system = physical + food/water/energy/health **supporting action scenarios**
3. Information on scales **where the impact of climate change is measured and observed**
4. **Interactive** and configurable **access to all data, models and workflows**





More realistic at *global* scale

More realistic at *local* scale

Include impacts where they matter

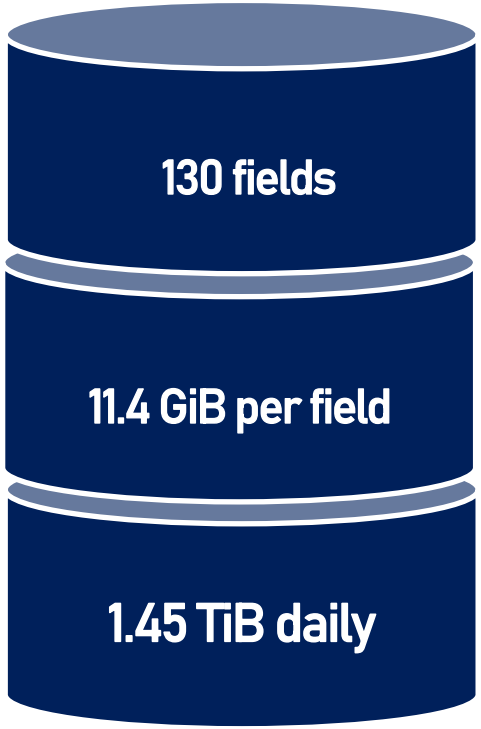
Trial different adaptation and mitigation scenarios

See talk of DestinE science team at *https://events.ecmwf.int/event/241/contributions/3461/attachments/2093/3717/WGNE-WS_Sandu.pdf*

DestinE Digital Twins Data Volume / Data Portfolio

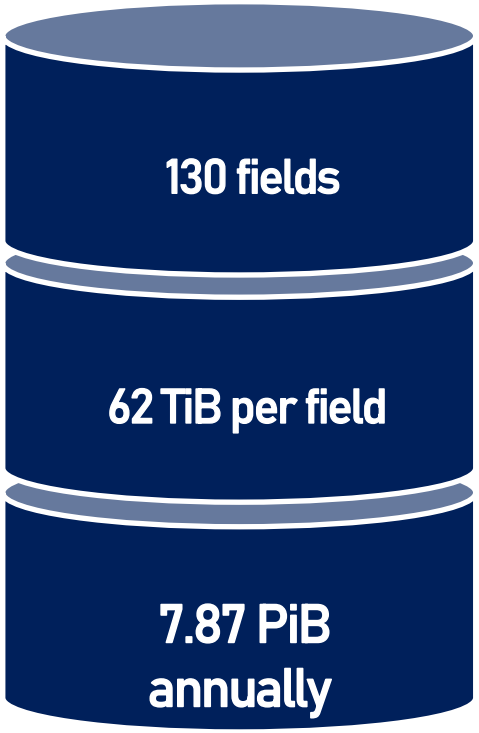
DT on Weather-induced Extremes

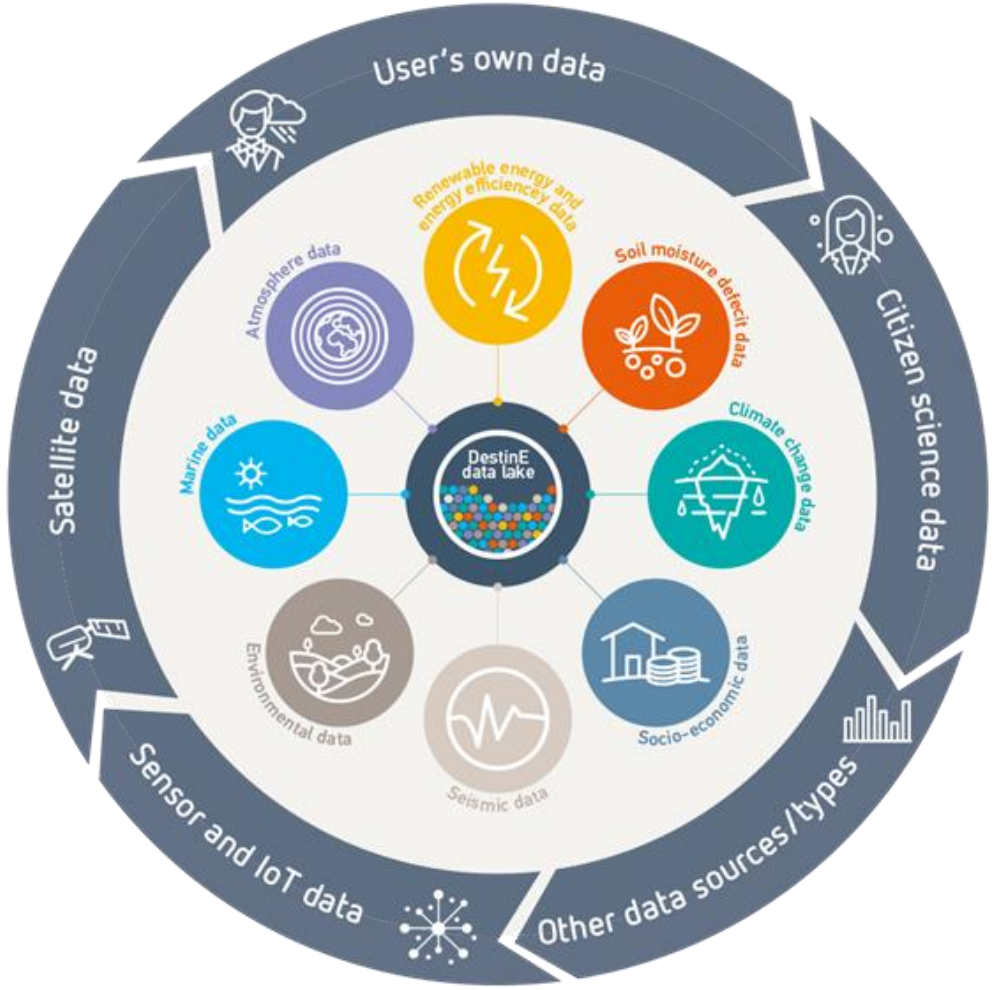
Temporal resolution: 15 minutes to 1 hour
Time horizon: 4-7 days forecast
Horizontal resolution: 4.4/2.8/1.4 km
Number of instances: 1



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)

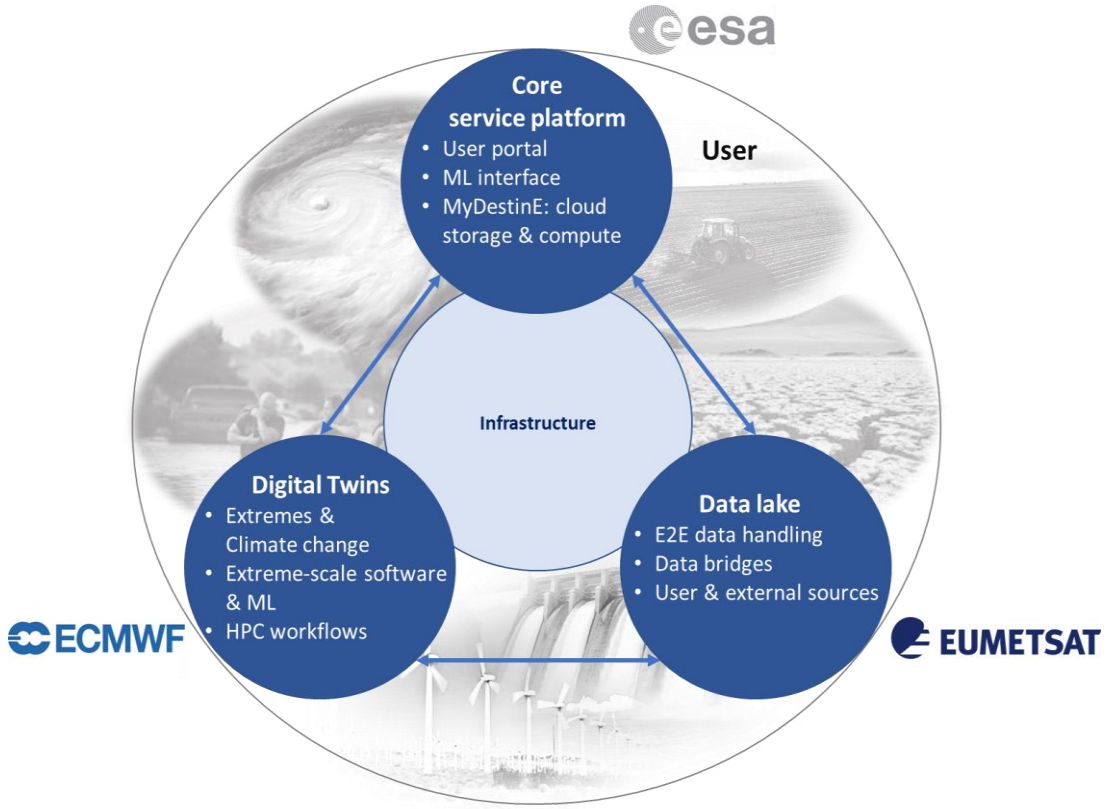




DestinE: A joint undertaking of ESA, ECMWF and EUMETSAT

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)



NOTE:

Three self-standing components

Components do not use common infrastructure

Destination Earth Data Lake Concepts & Services

Destination Earth Data Lake (DEDL)

Self-standing component

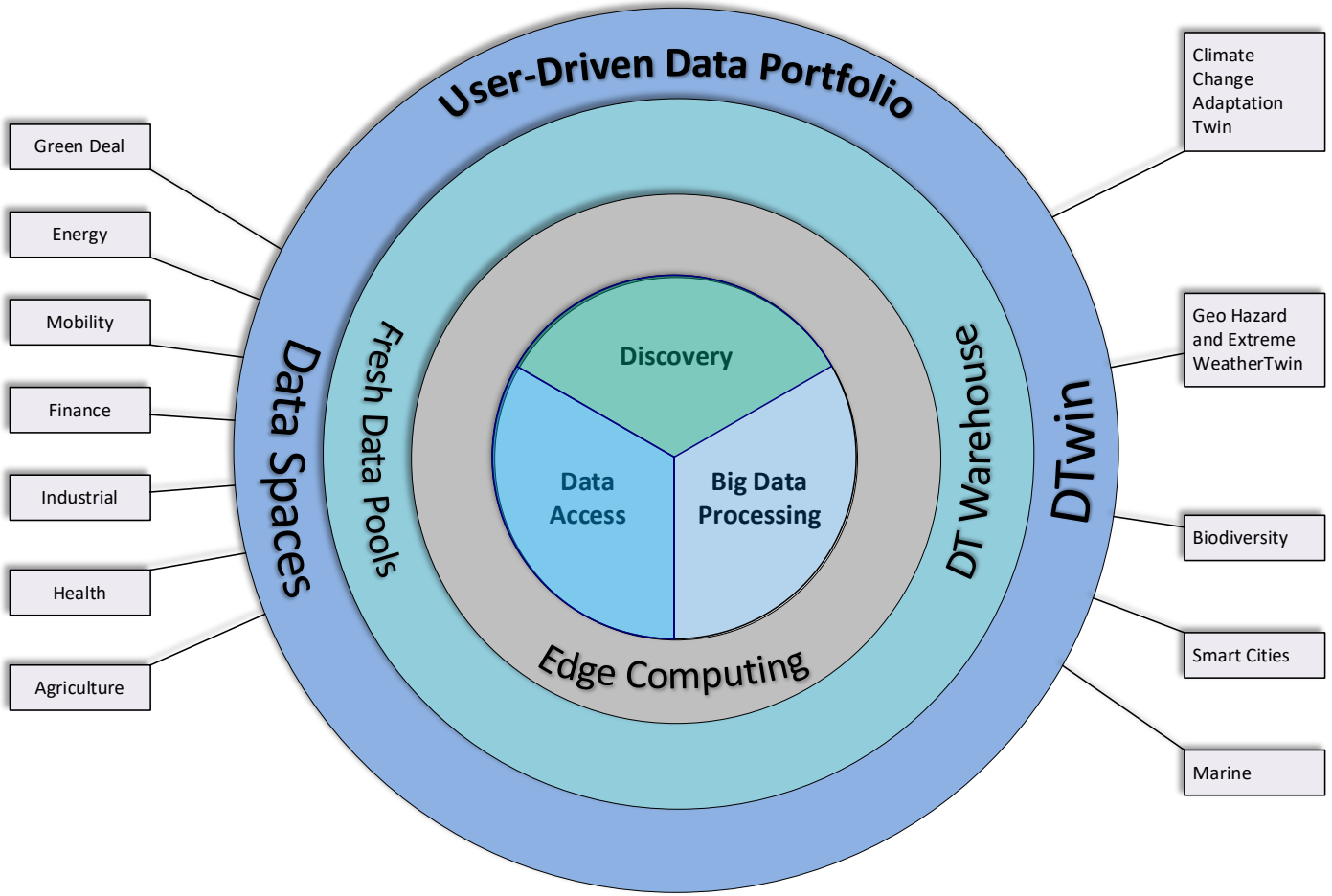
- Built from geographically distributed physical elements (central & edges)
- Distributed services – seamless access

Discovery & Data Access

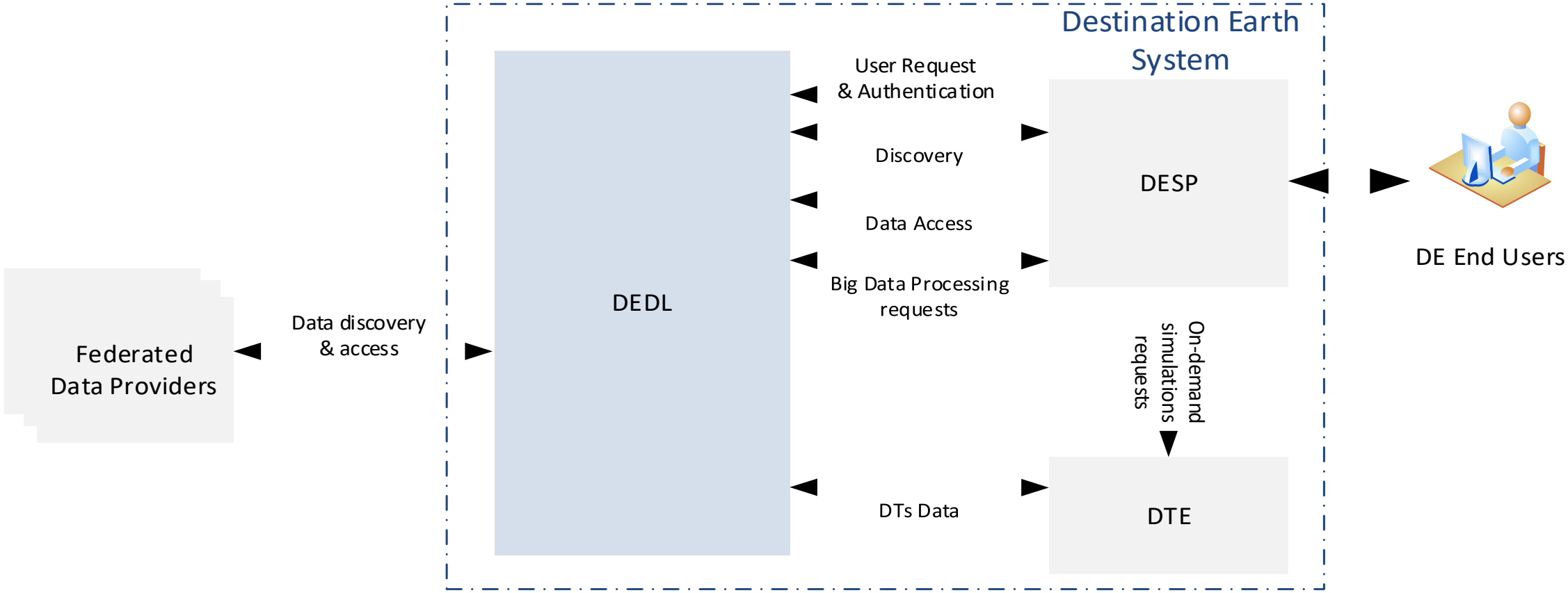
- Harmonisation of data access (HDA) to simplify data discovery & access
- Initially two Digital Twins (ECMWF):
  - Geo Hazard and Extreme Weather
  - Climate Change Adaptation
- External federated data spaces
- User-generated data

Big Data Processing

- Processing near data including distributed computing & workflows

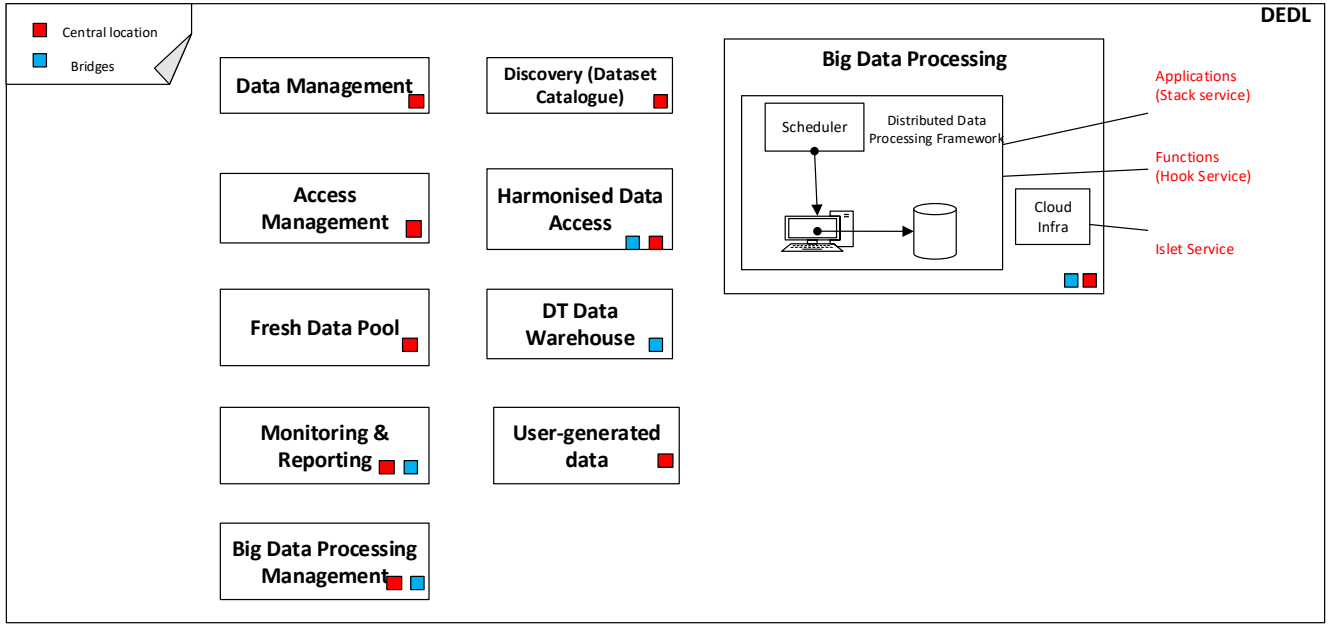


DEDL System Context

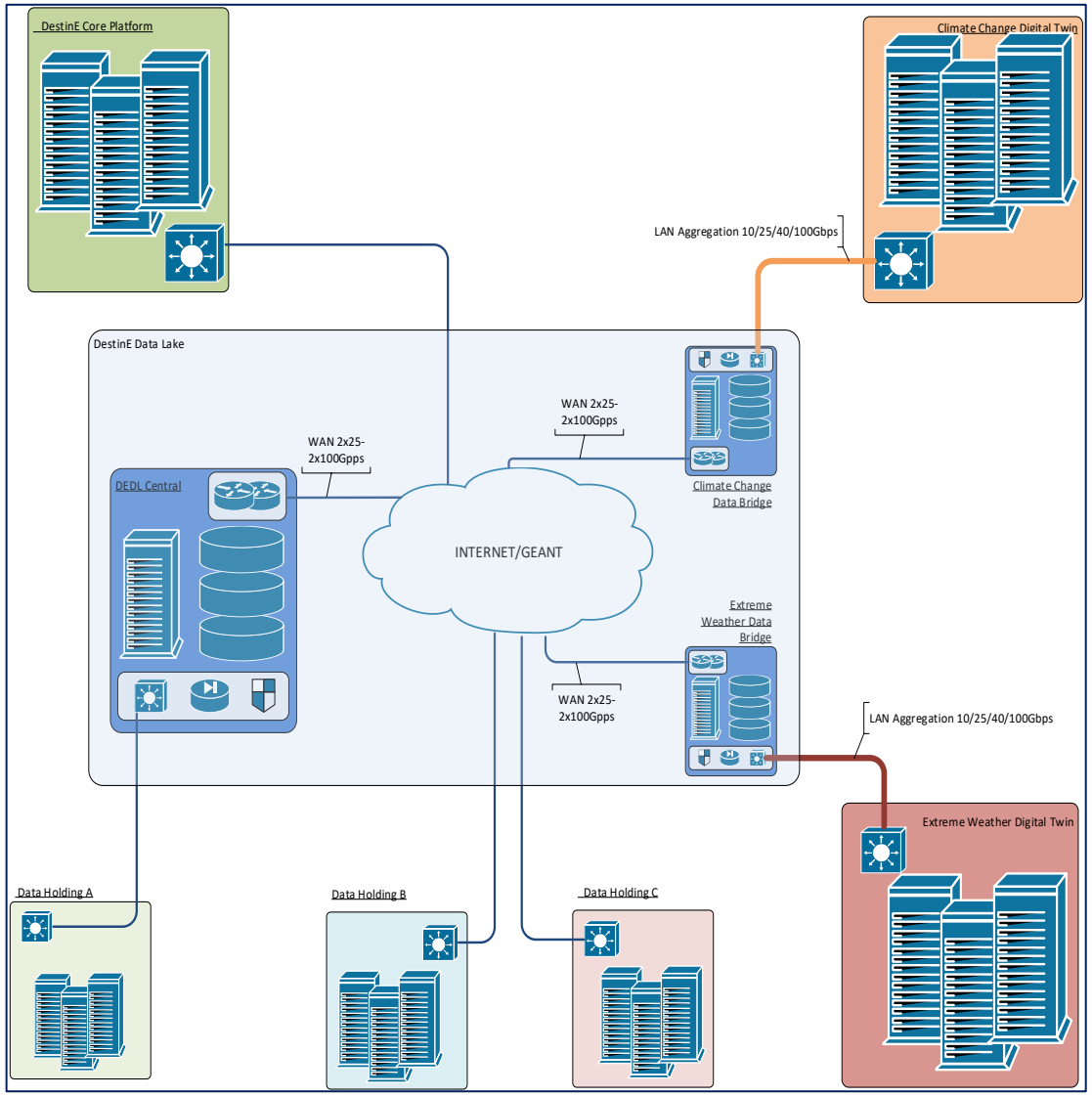


DEDL Service Deployment

Service components



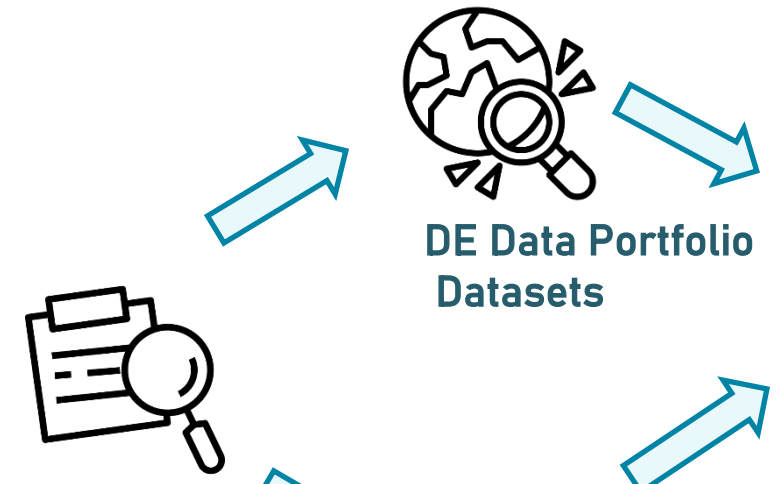
on geographically distributed infrastructure



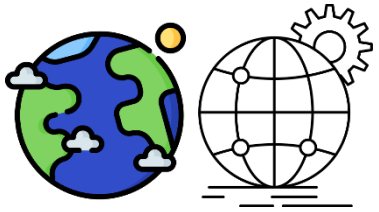
Funded by the European Union Destination Earth

Destination Earth Data Lake Discovery & Data Access Services

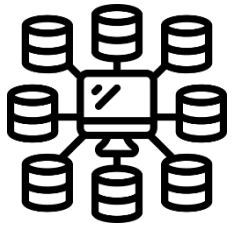
Discovery Services



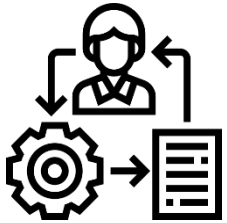
Data Access Services



Digital
Twin Outputs



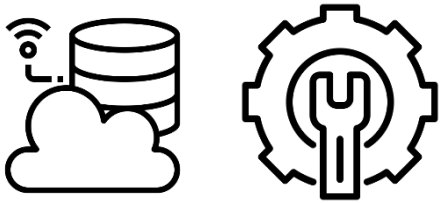
Federated
Datasets



User-
Generated Data

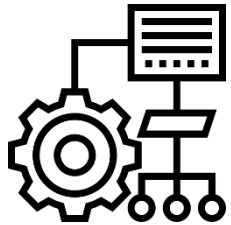
Fresh
Data Pool

Big Data Processing Services



Islet Service
(Infrastructure& tools)

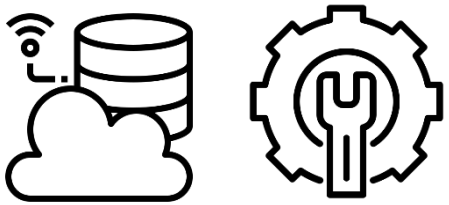
Stack Service
(Hosted Applications)



Hook Service
(Functions)

DEDL Big Data Processing Services

Islet Service



Infrastructure & tools

- 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**

Stack Service

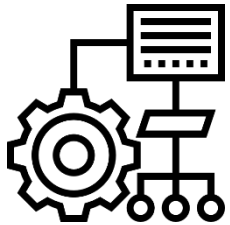
Hosted Applications

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

For Users who

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

Hook Service



Ready-to-use Functions

Predefined processing workflows/ functions

User-defined workflows

System or User-defined data cubes

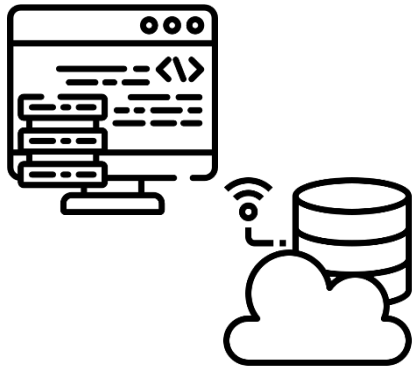
For Users who

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

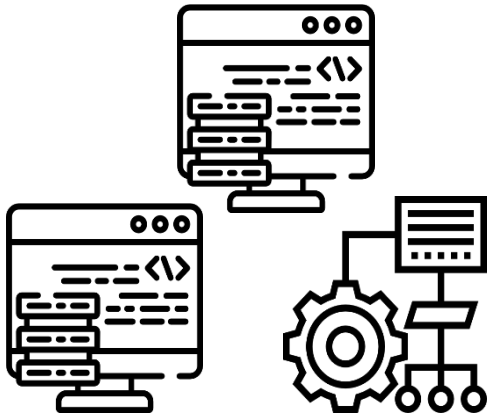
Icons made by Phatplus , Superndre, Freepik, Eukalyp from www.flaticon.com

DEDL Big Data Processing Services

Users can pick and mix big data processing service offerings:

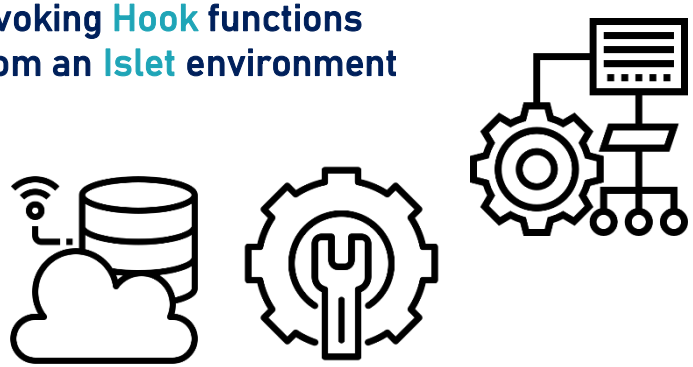


Stack (JupyterHub) + Islet-Storage (Uploading own data & Storing results)



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

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

Destination Earth Data Lake Timeline

- Service Increment 1 (Minimum Viable Service) – Q3 2023
  - Climate DT Bridge (LUMI)
  - Central Site
  - Big Data Processing Services on both sites
  - Harmonised Data Access first version
  - Discovery and Access service – Data Portfolio subset
- Service Increment 2 – Q4 2023
  - Extreme DT Bridge (Leonardo)
  - Additional service features
  - SLA and capacity increase
- Service Increment 3 – Q1 2024
  - MareNostrum Bridge
  - Additional service features
  - SLA and capacity increase

Thank you!

Questions are welcome.