Destination Earth
Data Lake

15th February 2023– DestinE User eXchange
ESA–ESRIN, Frascati

Michael Schick (EUMETSAT)

Funded by the European Union

Destination Earth

IMPLEMENTED BY EUMETSAT eesa ECMWF

DestinE: A joint undertaking of ESA, ECMWF and EUMETSAT
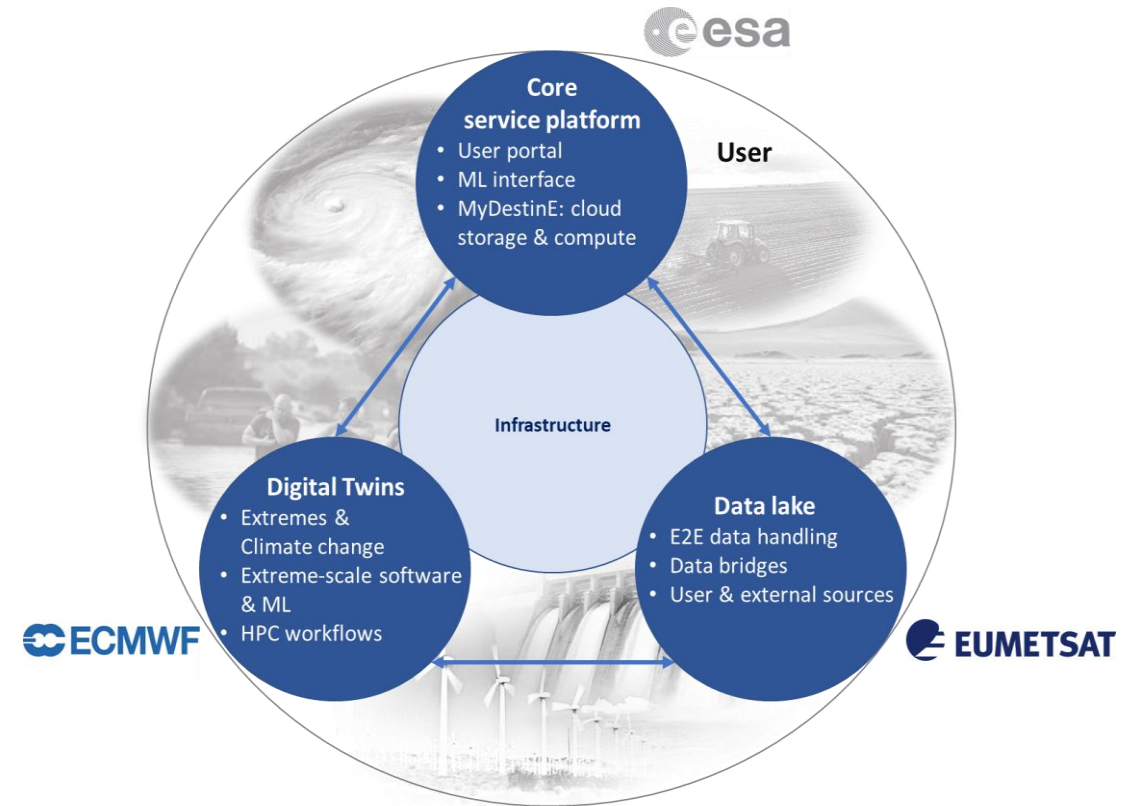
Three entrusted entities implementing DestinE

- Core Service Platform interfacing DestinE users (ESA)

- Two Digital Twins. Extreme Weather and Climate Change Adaptation (ECMWF)

- Destination Earth Data Lake (EUMETSAT)

Self-standing components

Components do not use common infrastructure

DEDL – Edge computing paradigm (Processing near data)

Core service platform
- User portal
- ML interface
- MyDestinE: cloud storage & compute

User

Infrastructure

Digital Twins
- Extremes & Climate change
- Extreme-scale software & ML
- HPC workflows

Data lake
- E2E data handling
- Data bridges
- User & external sources

esa

ECMWF

EUMETSAT

Funded by the European Union **Destination Earth**

DestinE Data Lake

Destination Earth Data Lake (DEDL)

- <u>Self-standing</u> component, built from geographically distributed physical elements, that references and provides seamless access to <u>all DestinE</u> user required <u>data</u>.

- Provision of data & information available from a large number of external <u>data spaces</u> or generated by the DestinE <u>Digital Twins and applications on the Core Service Platform</u>, <u>regardless of data type and location</u>.

- The Data Lake supports <u>near-data processing</u> to maximize throughput and service scalability and implements <u>big data distributed workflows</u>.

- The concepts applied in the DestinE Data Lake Service will provide a harmonisation of data access, <u>beyond</u> anything that exists today.
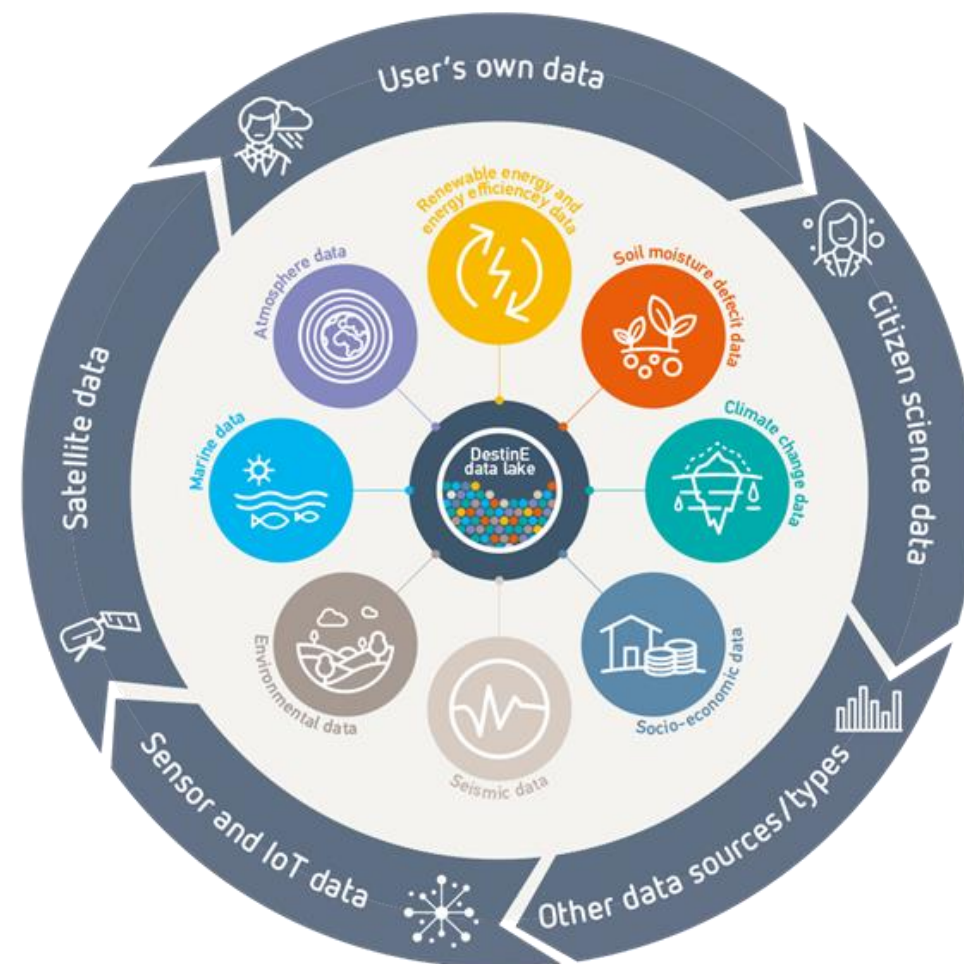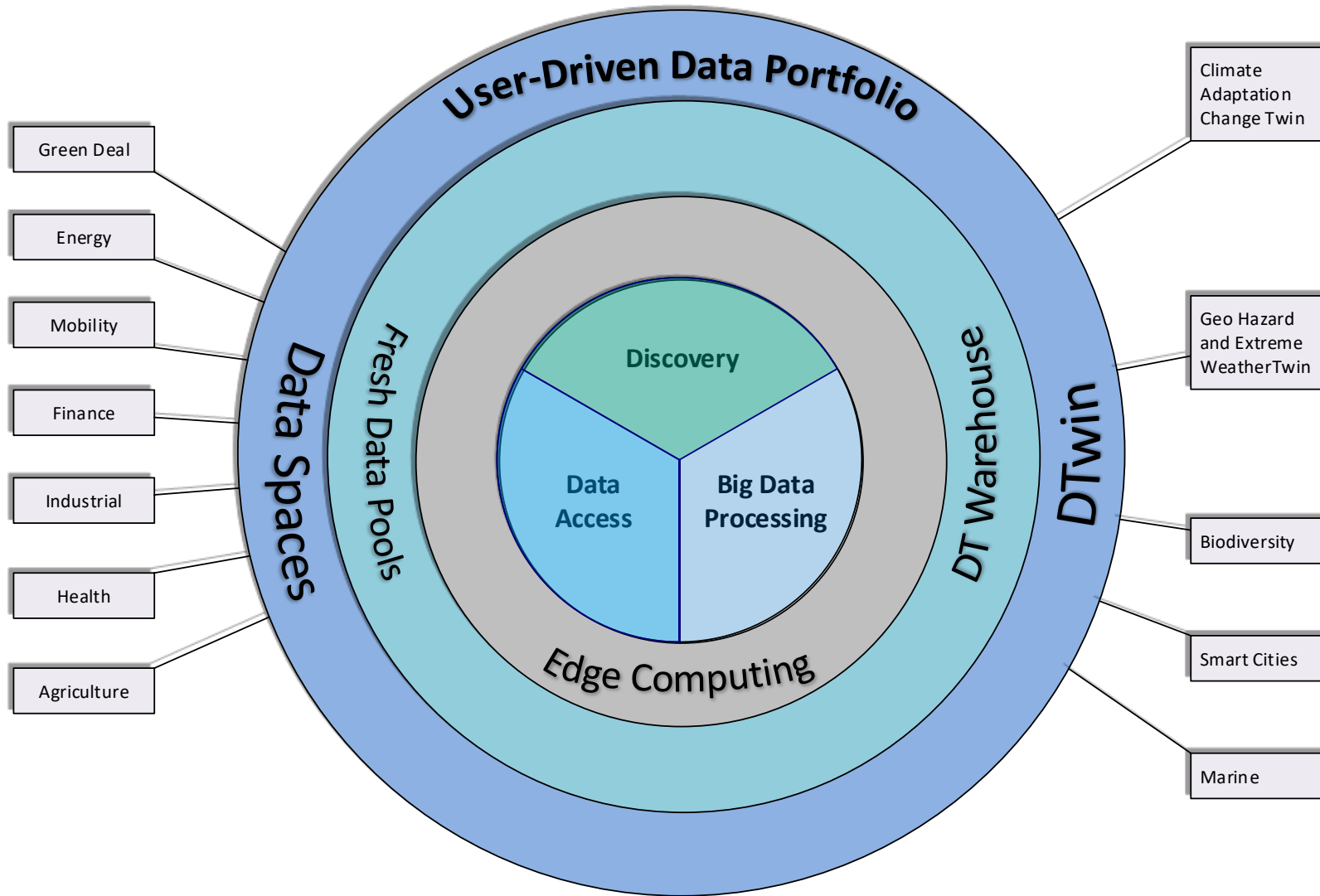
DestinE Data Lake Services

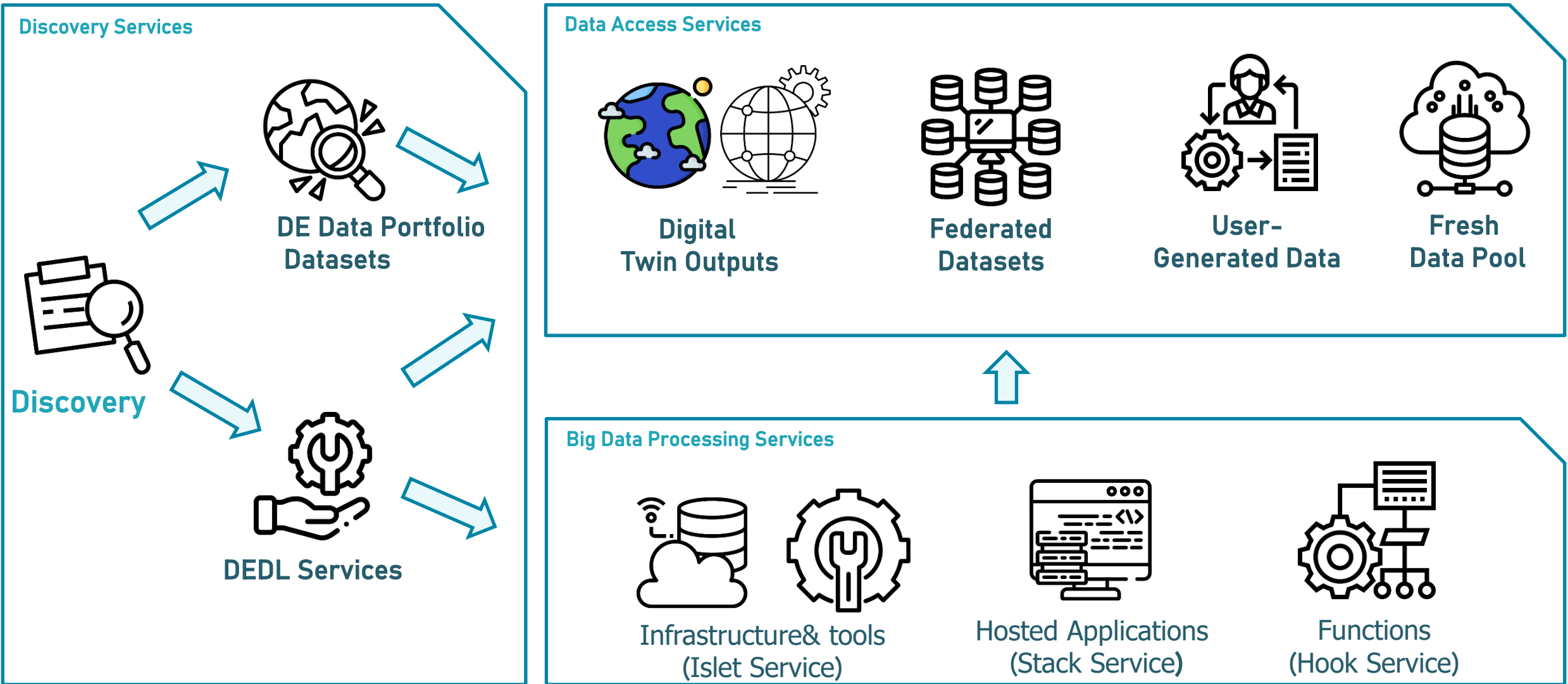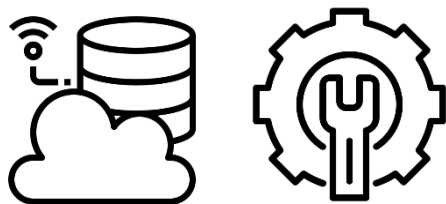| DEDL Services Exposed via DESP				Operator Services	
DEDL Discovery Service	**DEDL Data Access Service**	**DEDL Big Data Processing Service**	**DEDL Service Desk**	**DEDL Management Service**	**DEDL Services for DTs**
Discover Data	Access Federated Datasets	Cloud Infrastructure (Islet)	Service Desk	DEDL Data Management	Cloud Infrastructure (IaaS and PaaS)
Discover DE Services	Access Fresh Data Pool	Application (Stack)	System Status Information	DEDL Access Management	Provision of Inputs data (Phase II)
	Access DT Outputs	Functions (Hook)		DEDL Big Data Processing Management	
	Access User Generated Data			DEDL Monitoring and Reporting	
				DEDL Maintenance	

Key Points: fusion of data, on–Demand, distributed processing near data, extendable reference Architecture, suitable for AI/ML, workflows

Discovery & Data Access Services

Discovery Services

Discovery

DE Data Portfolio Datasets

DEDL 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)

Funded by the European Union   **Destination Earth**

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
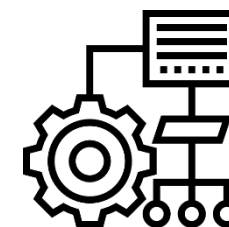
Hosted Applications

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
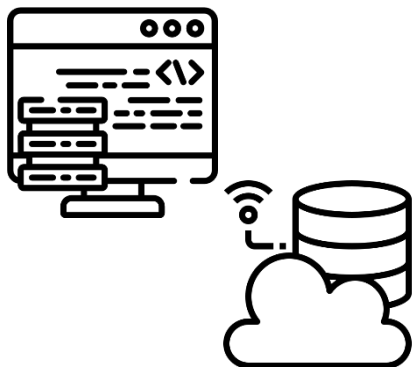
System or User-defined data cubes

For Users who

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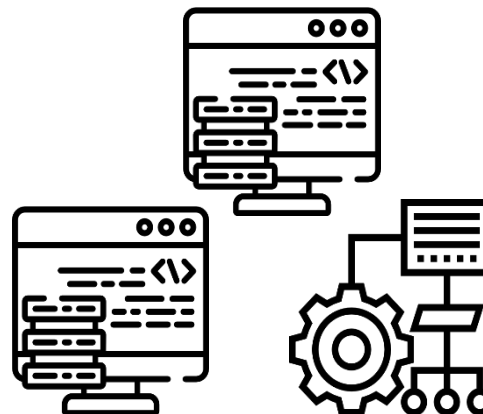- want advanced processing services

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

Funded by the European Union **Destination Earth**

DEDL – Big Data Processing Services

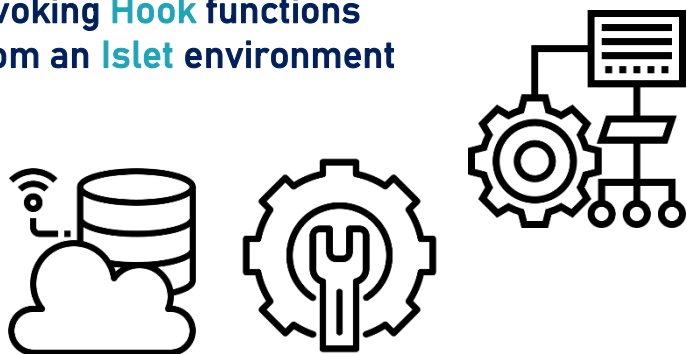## Users can pick and mix big data processing service offerings:

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

Funded by the European Union **Destination Earth**

DestinE Data Lake Timeline

- Service Increment 1 (Minimum Viable Service) – Q3 2023
  - Climate Change Adaptation Twin
  - Priority datasets from Data Portfolio (Use Case based)
  - Initial version:
    - Discovery and Access service
    - Harmonised Data Access (Federation of Data)
    - Big Data processing services (Islet, Stack, Hook)

- Service Increment 2 – Q4 2023
  - Extreme Weather and Geo hazard Twin
  - Data Portfolio (next priority datasets)
  - Enhancements of Increment 1 services
  - SLA and capacity increase

- Service Increment 3 – Q1 2024
  - Enhancements of Increment 2 services
 - Data Portfolio (next priority datasets)
  - SLA and capacity increase

DestinE Data Portfolio

DestinE Data Portfolio

- ## Evolving Data Portfolio
  - User Driven
  - Managed and Controlled by DestinE Data Governance Board

- ## Digital Twins Data
  - Climate Change Adaptation
  - Extreme Weather and Geo hazards

- ## Federated datasets
  - Contributing missions (EUMETSAT, ESA, ECMWF)
  - Copernicus Satellites & Services data
  - Eurostat
  - ISIMIP
  - IAGOS

Funded by the European Union **Destination Earth**
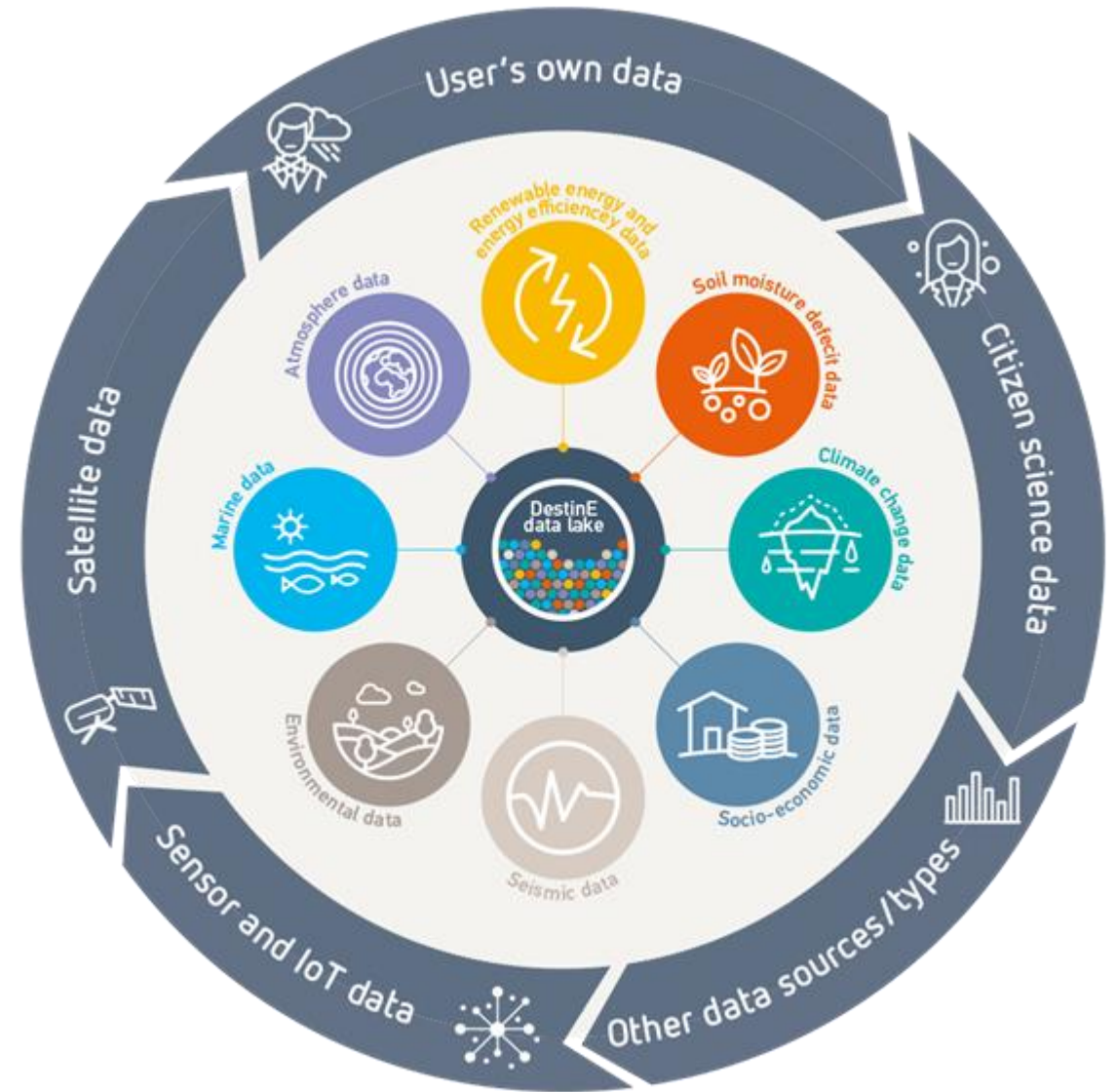
Thank you!

Questions are welcome.

Michael.Schick@eumetsat.int

Funded by
the European Union **Destination Earth**