

# EODC Forum 2022 – Destination Earth and its Data Lake Service

*15<sup>th</sup> of June 2022*





- Introduction
- High Level Technical Concepts
- Service Establishment
  
- Question & Answer

*“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.”*

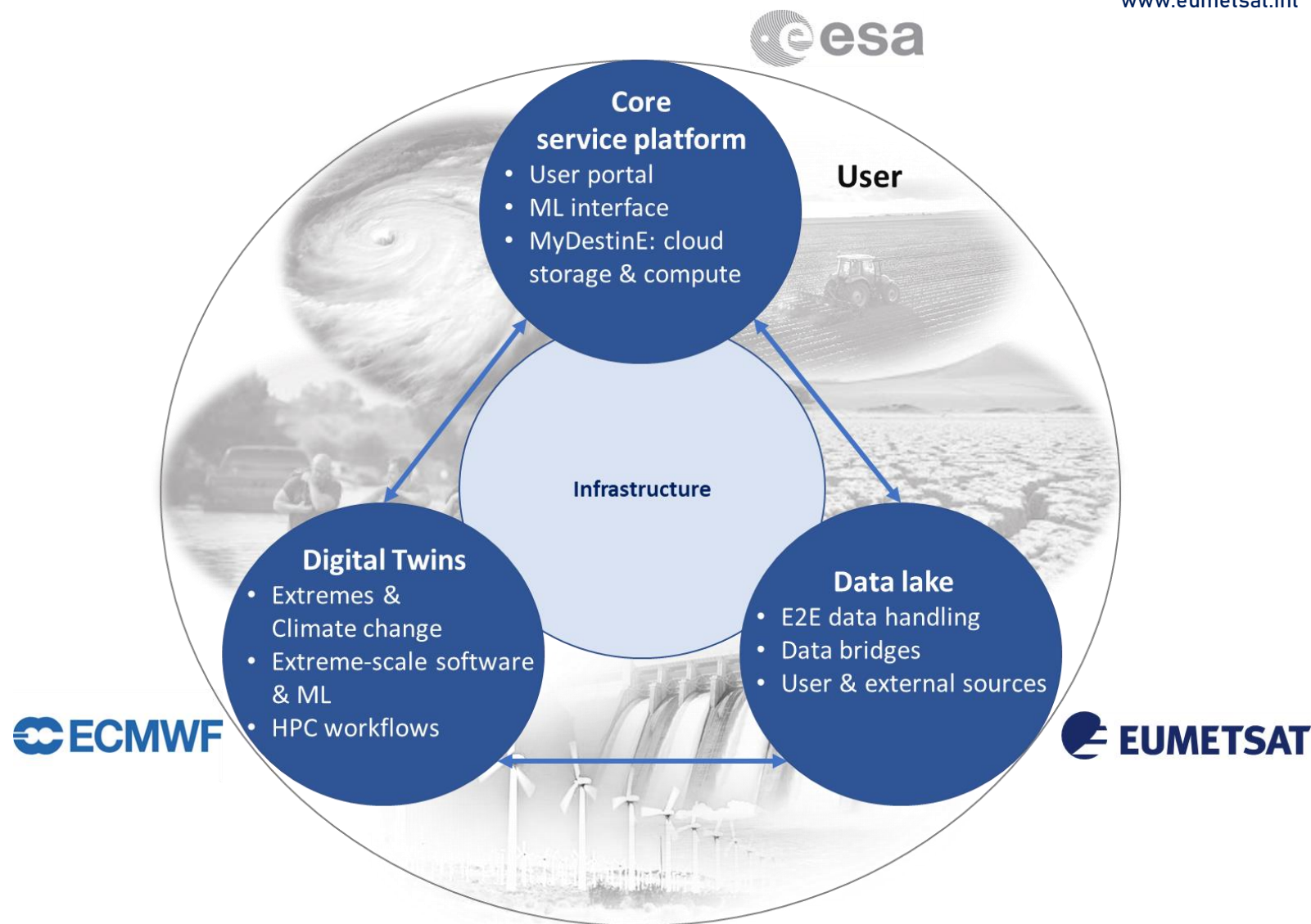
Destination Earth is part of the European Commission implementation of

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



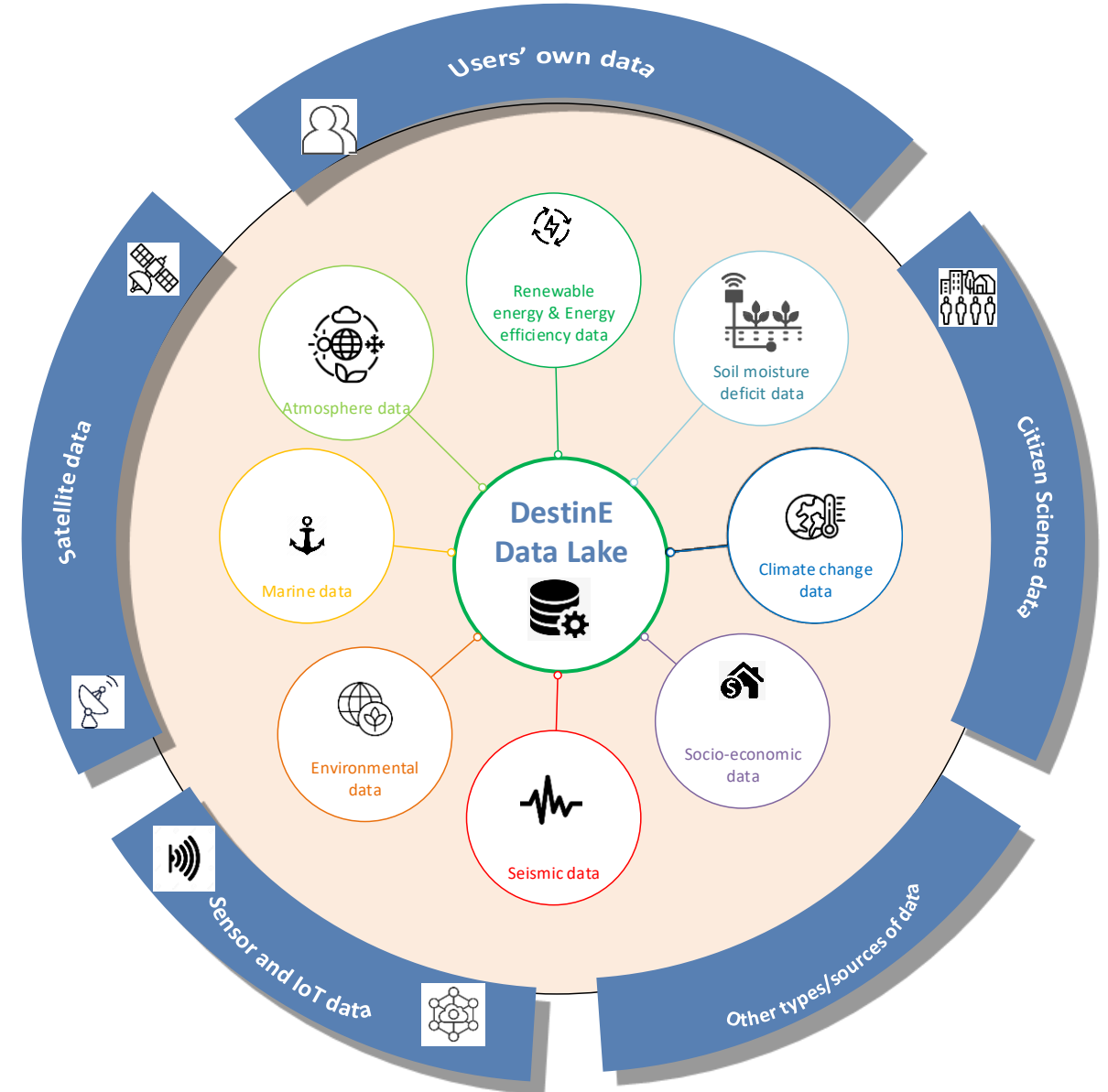
# A joint undertaking of ESA, ECMWF and EUMETSAT

- Three implementing agencies;
- ESA has the responsibility for the Core Service Platform that is interfacing to the DestinE users
- ECMWF is responsible for two Digital Twins: Extreme Weather and Climate Change
- EUMETSAT has end-to-end responsibility for the Destination Earth Data Lake;
- There is no common infrastructure, data exchange is provided via defined Interfaces
- If physical proximity is needed it will be provided by co-hosting of service infrastructure (DEDL bridge and HPC installation for example)



## DestinE Data Lake

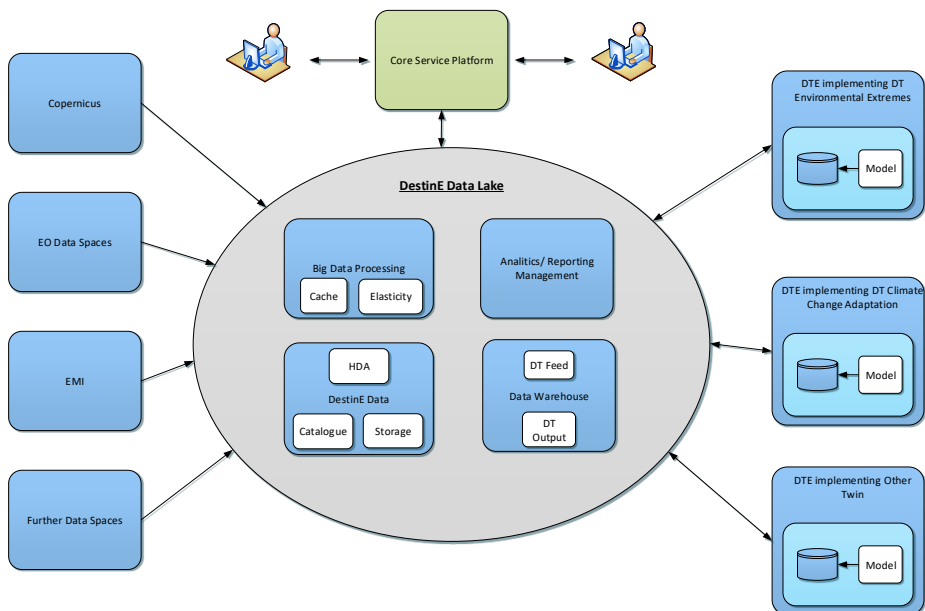
- Self-standing component, built from geographically distributed physical elements that references and provides seamless access to all DestinE user required data.
- Provision of data & information available from a large number of external data spaces or generated by the DestinE Digital Twins and applications on the Core Service Platform, regardless of data type and location.
- The Data Lake supports near-data processing to maximize throughput and service scalability and implements big data distributed workflows.
- The concepts applied in the DestinE Data Lake Service will provide a harmonisation of data federation, beyond anything that exists today.



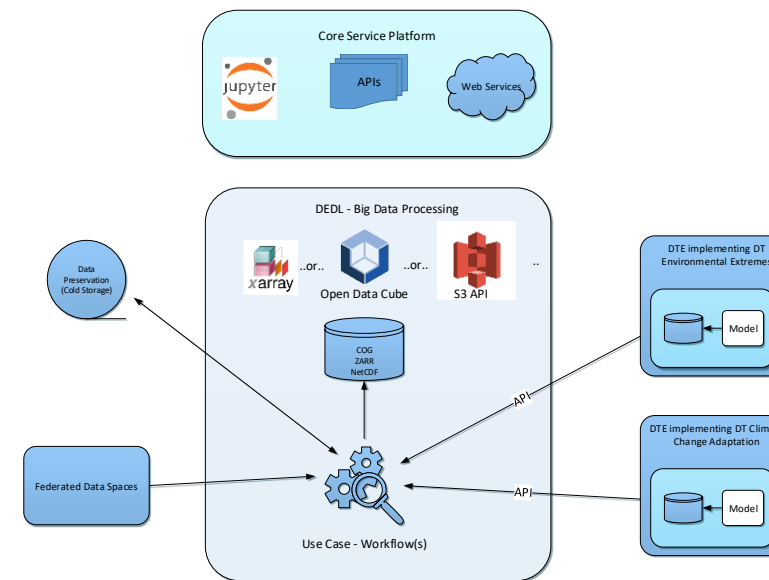


# High Level Technical Concepts

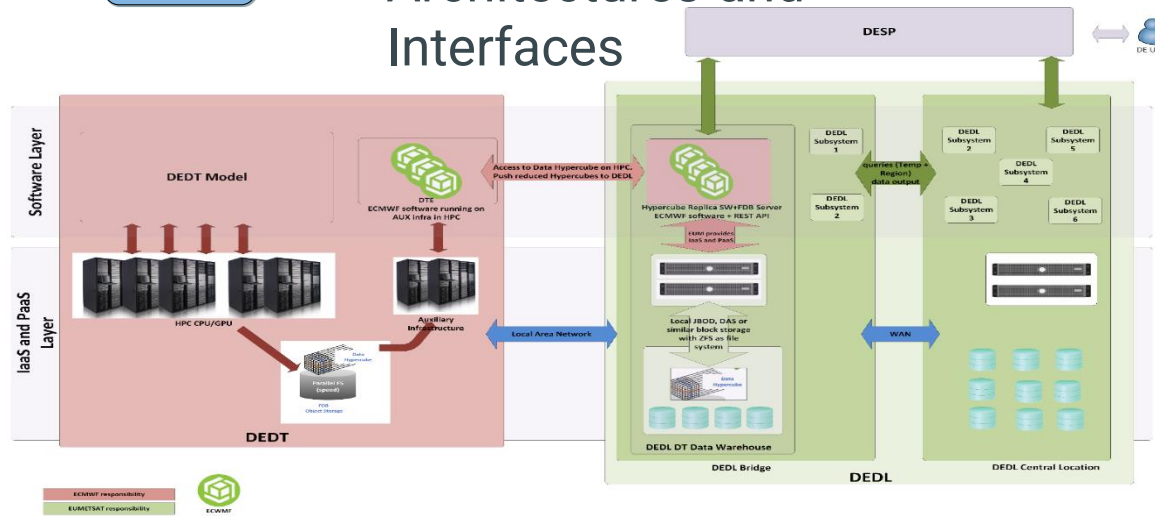
- Scalable Service Framework(s)



- Supporting "What if" Scenarios and Big Data Processing Workflows

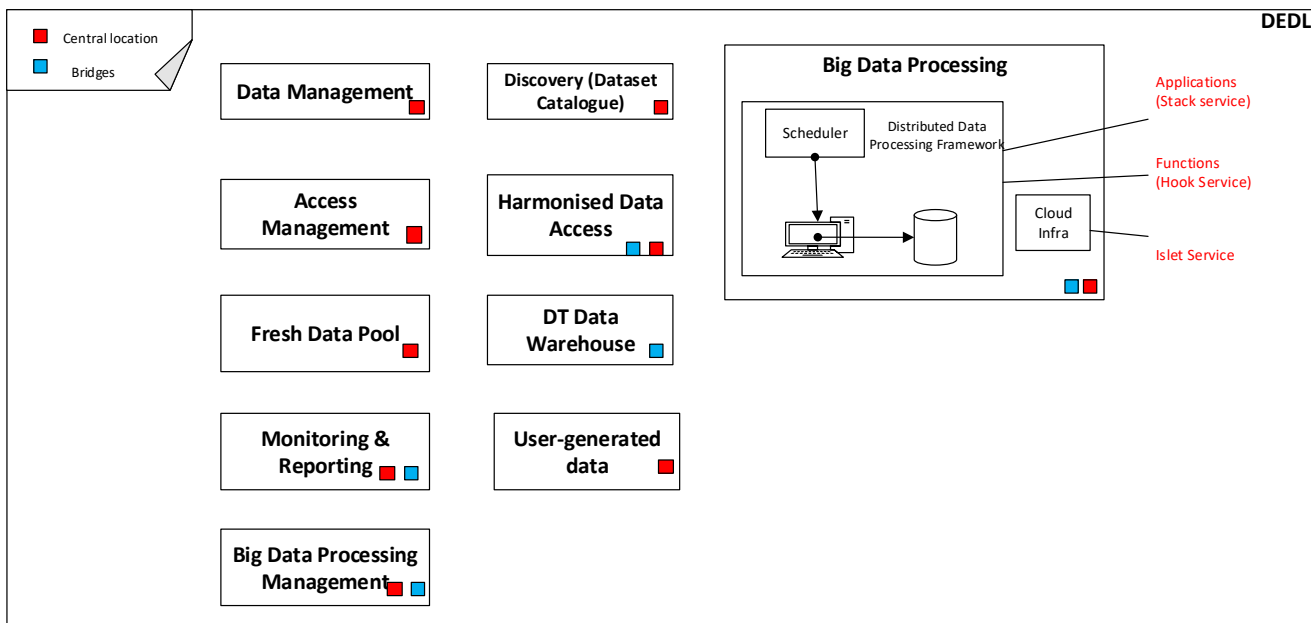


- Establishing Reference Architectures and Interfaces

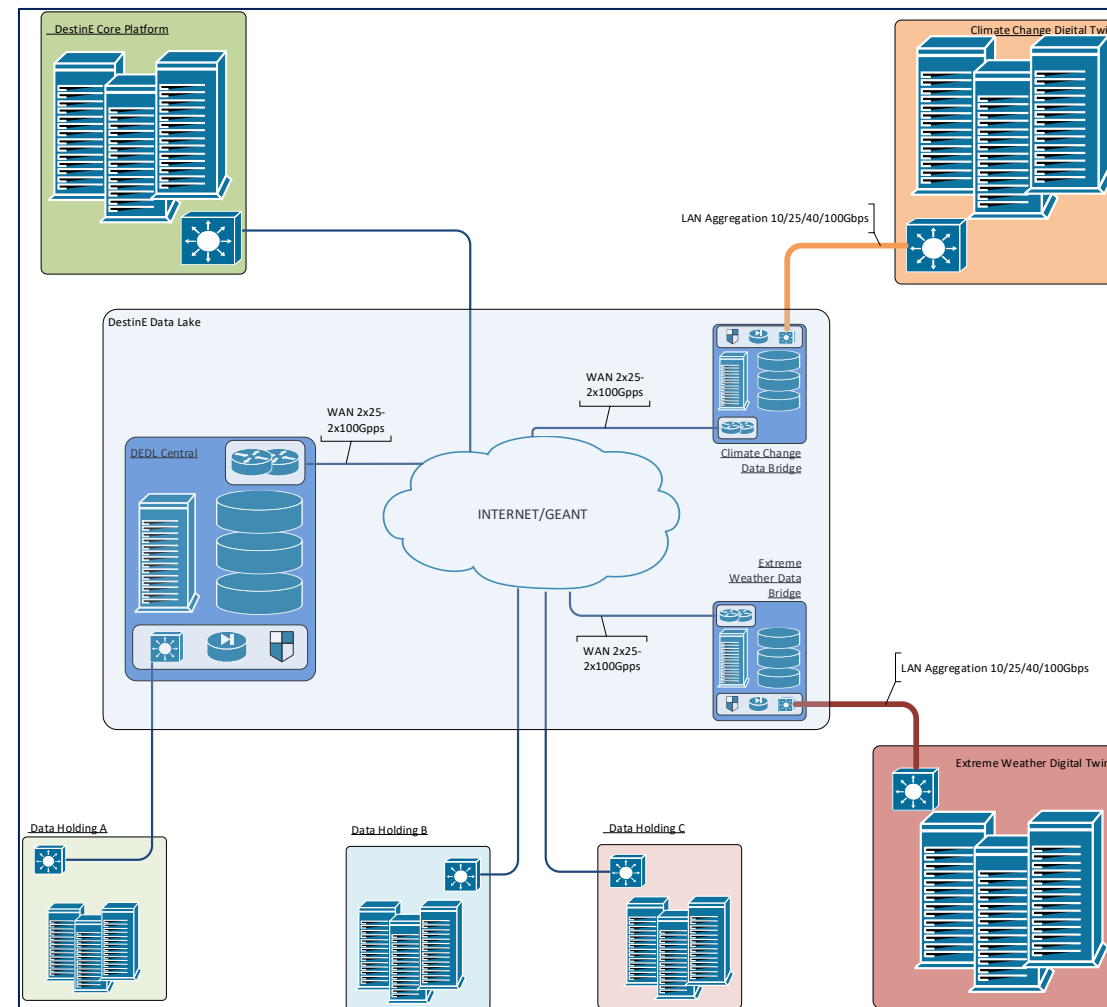




- service components



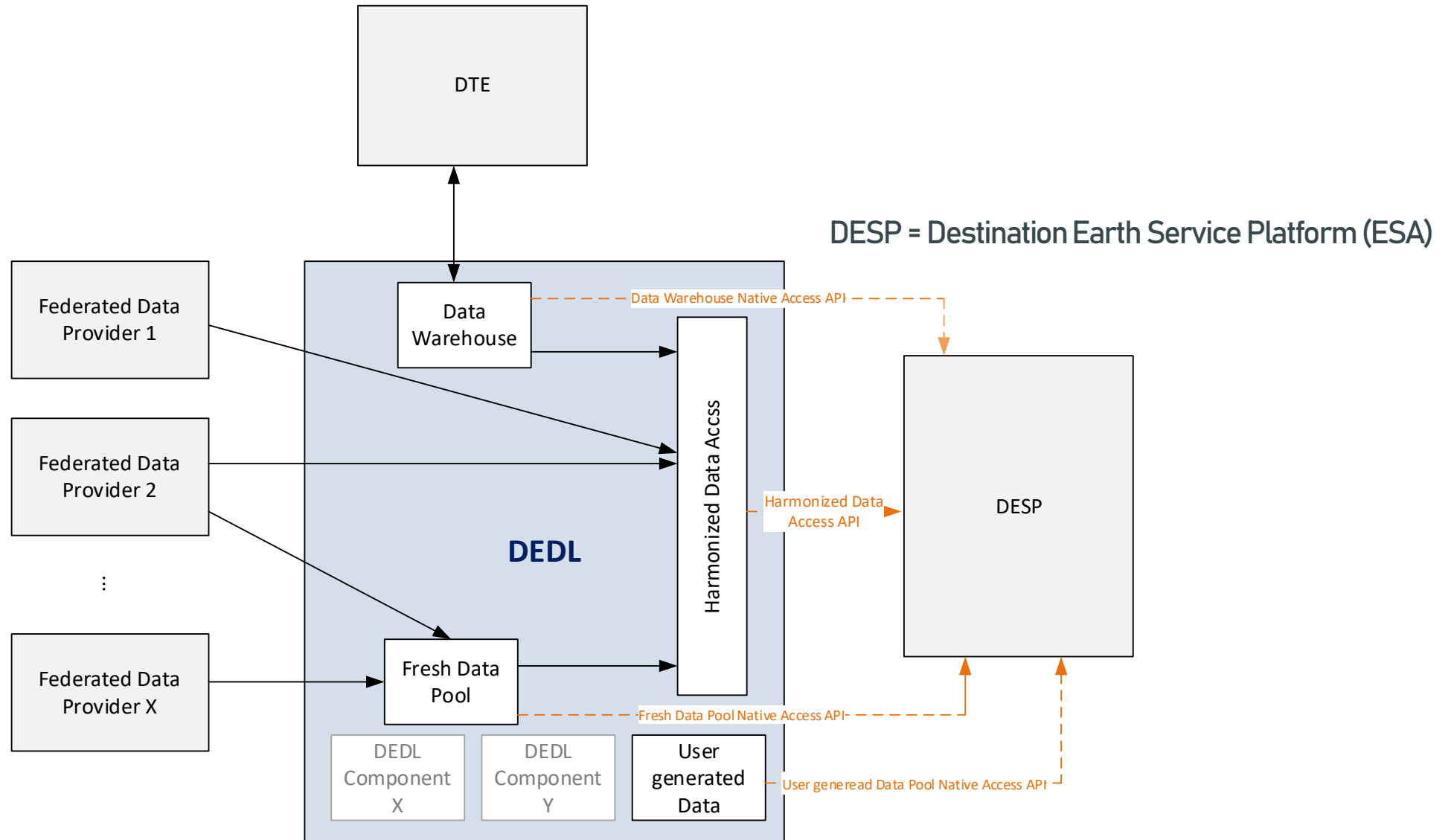
- on geographically distributed infrastructure





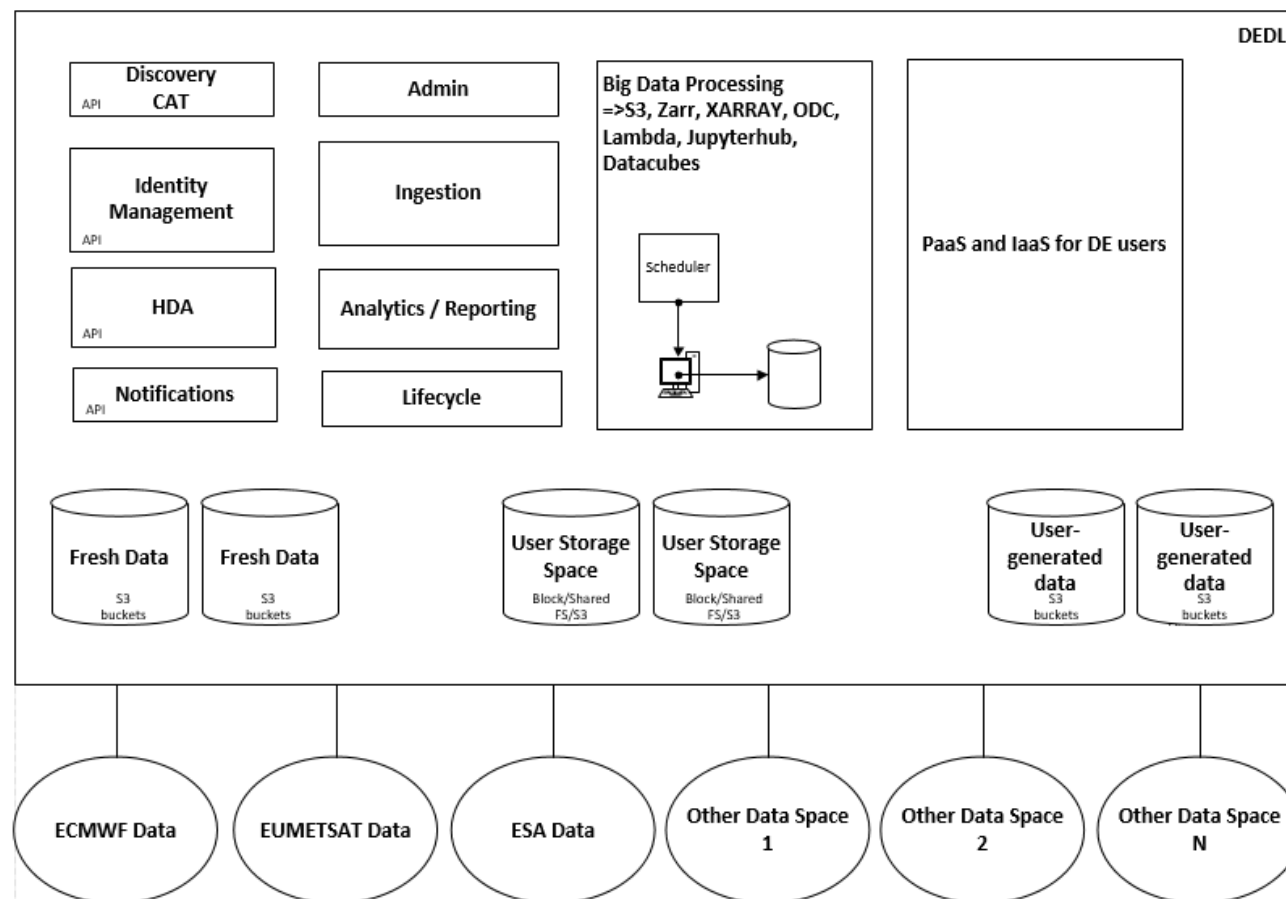


# DEDL Data access flows





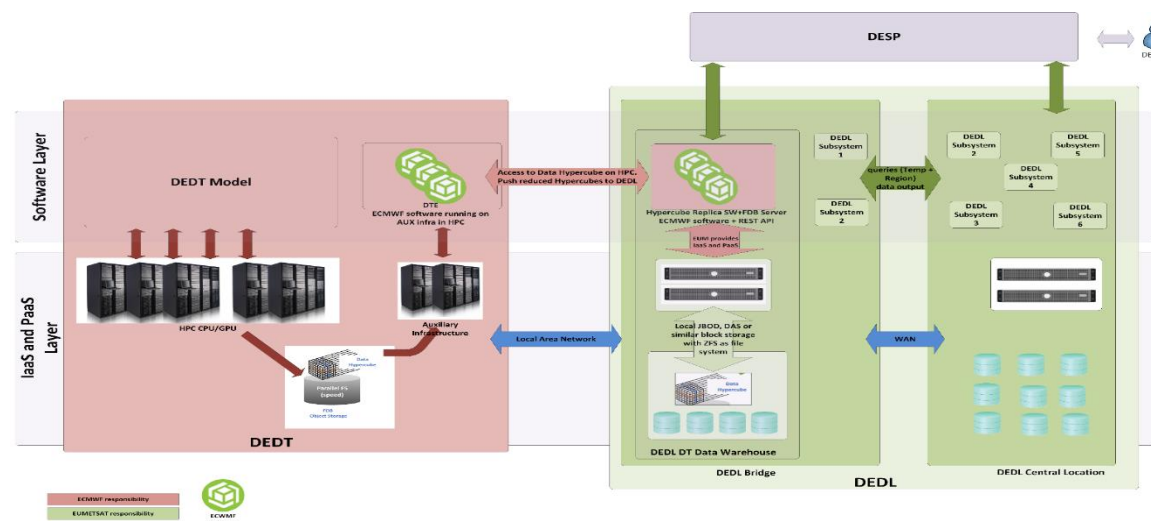
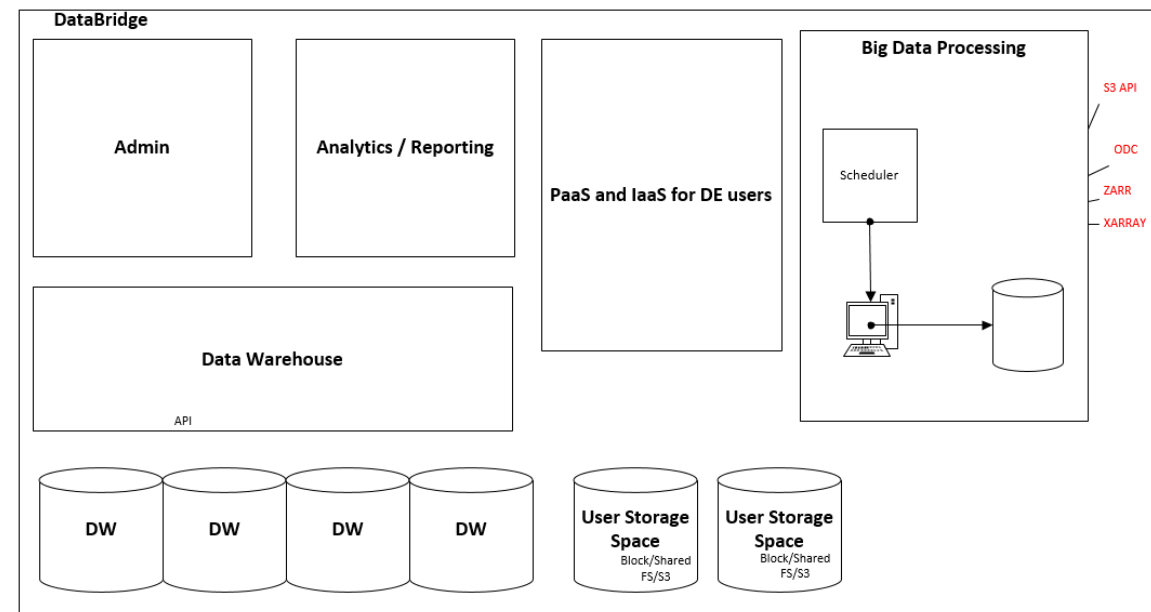
- Location: anywhere in EU, at the Service Provider DC, in public or private cloud infrastructure located within EU
- Integrated with DESP federated identity management
- High performance internet access
- Shall support establishing private WAN links to the data holdings without internet connectivity
- IaaS and PaaS for the users
- Host Big Data Processing Services (examples)
  - FaaS
  - Jupyterhub
  - Dask
  - openEO
  - Support of AI/ML
  - Analysis-ready data (DataCubes)
- Data as a Service
  - Fresh data pools for EO data
  - Data Management and Administration
  - Notification Services
  - Access to federated data sets





# Bridge Service (Edge)

- Edge Cloud Service
- Located on premises of the Digital Twin HPC or other large data holding
- Integrated with DESP federated identity management
- High performance internet access
- High Performance LAN access to internal HPC infrastructure
- High-capacity storage for Data Warehouse
- IaaS and PaaS for the DT interface
- IaaS and PaaS for the users
- Big Data Processing (examples)
  - FaaS
  - Jupyterhub
  - Dask
  - openEO
  - Support of AI/ML
- Data as a Service
  - Data Warehouse (storage of DT results)
  - DT HyperCubes





# Service Establishment



## DEDL Services Exposed via DESP

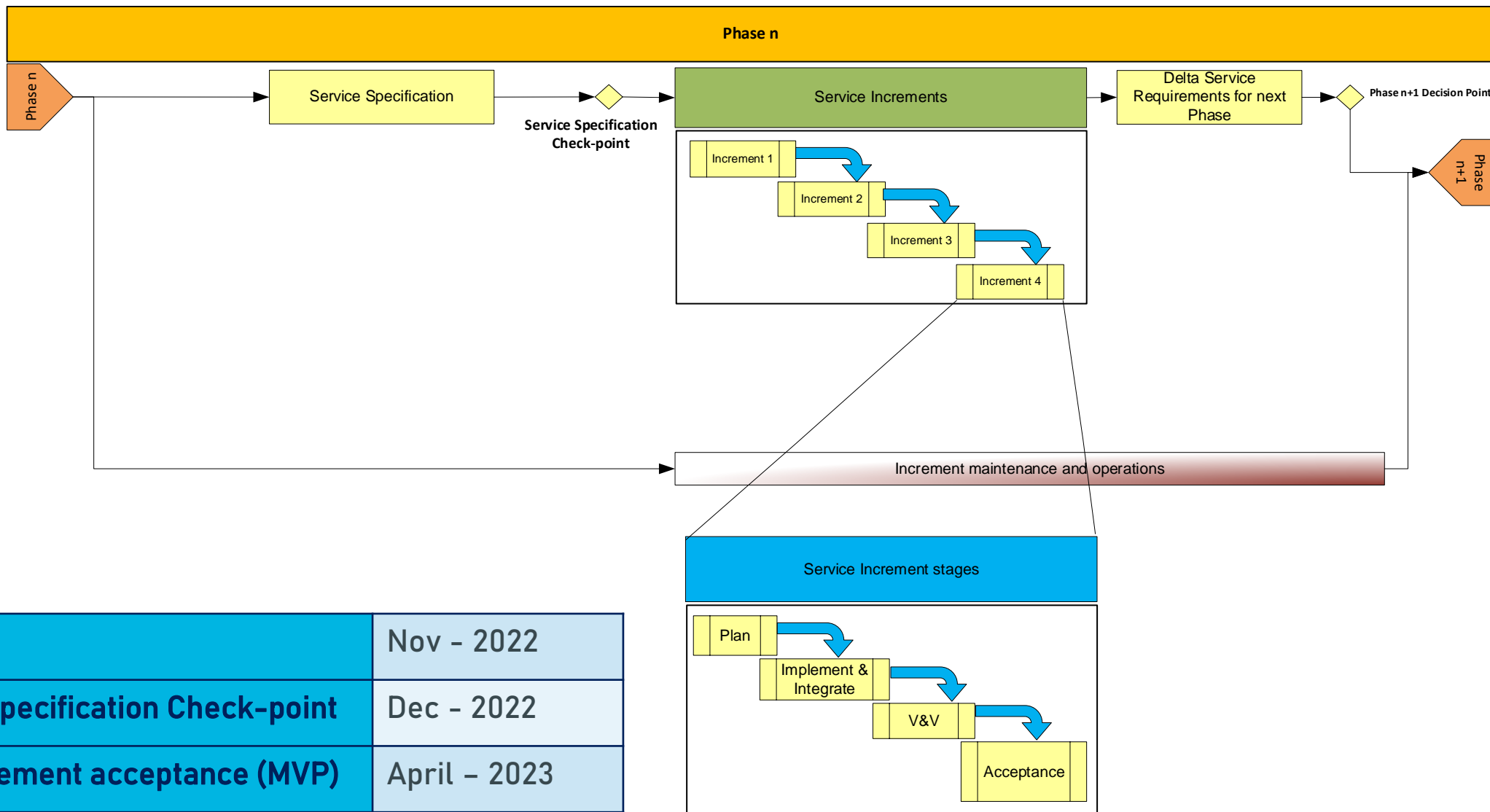
<b>DEDL Discovery Service</b>	<b>DEDL Data Access Service</b>	<b>DEDL Big Data Processing Service</b>	<b>DEDL User Service Desk</b>
Discover Data	Access Federated Datasets	Cloud Infrastructure (Islet)	Help Desk
Discover DE Services	Access Fresh Data Pool	Application (Stack)	
	Access DT Outputs	Functions (Hook)	
	Access User Generated Data		

## Operator Services

<b>DEDL Management Service</b>	<b>DEDL Services for DTs</b>
DEDL Data Management	Cloud Infrastructure (IaaS and PaaS)
DEDL Access Management	Provision of Inputs data (Phase II)
DEDL Big Data Processing Management	
DEDL Monitoring and Reporting	
DEDL Maintenance	



# Service development logic



<b>Kick-off</b>	Nov - 2022
<b>Service Specification Check-point</b>	Dec - 2022
<b>First increment acceptance (MVP)</b>	April - 2023
<b>Last increment acceptance</b>	Dec - 2023



- Allowing the execution of “What if” scenarios by efficient information access and/or processing near data (big data-, AI- and ML workflows) for value adding data generation
- On demand model execution and data generation (several tens of Petabytes)
- Harmonisation of information exchange and service exploitation between diverse European data spaces
  - The diversity of such data spaces is far beyond earth observation data
- Efficient access on Information for decision makers



**Thank you!**  
Questions & Answers