

Destination Earth Data Lake

Danaële Puechmaille - EUMETSAT
danaele.duechmaille@eumetsat.int

*DestinE Roadshow Webinar: Participation and Upcoming
Collaboration Opportunities Today - 19/07/2023*

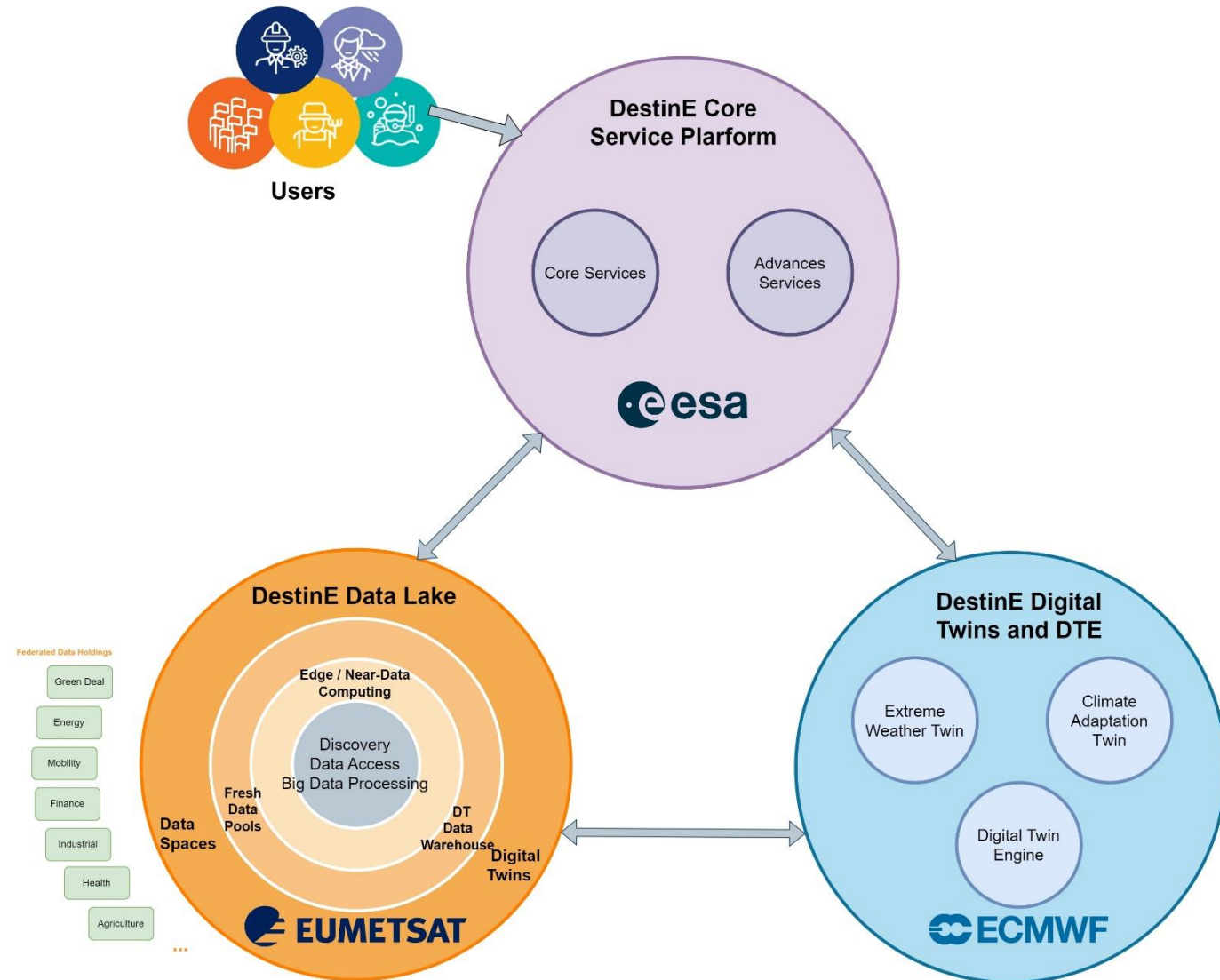




DestinE: A joint undertaking of ESA, ECMWF and EUMETSAT

Three entrusted entities implementing DestinE

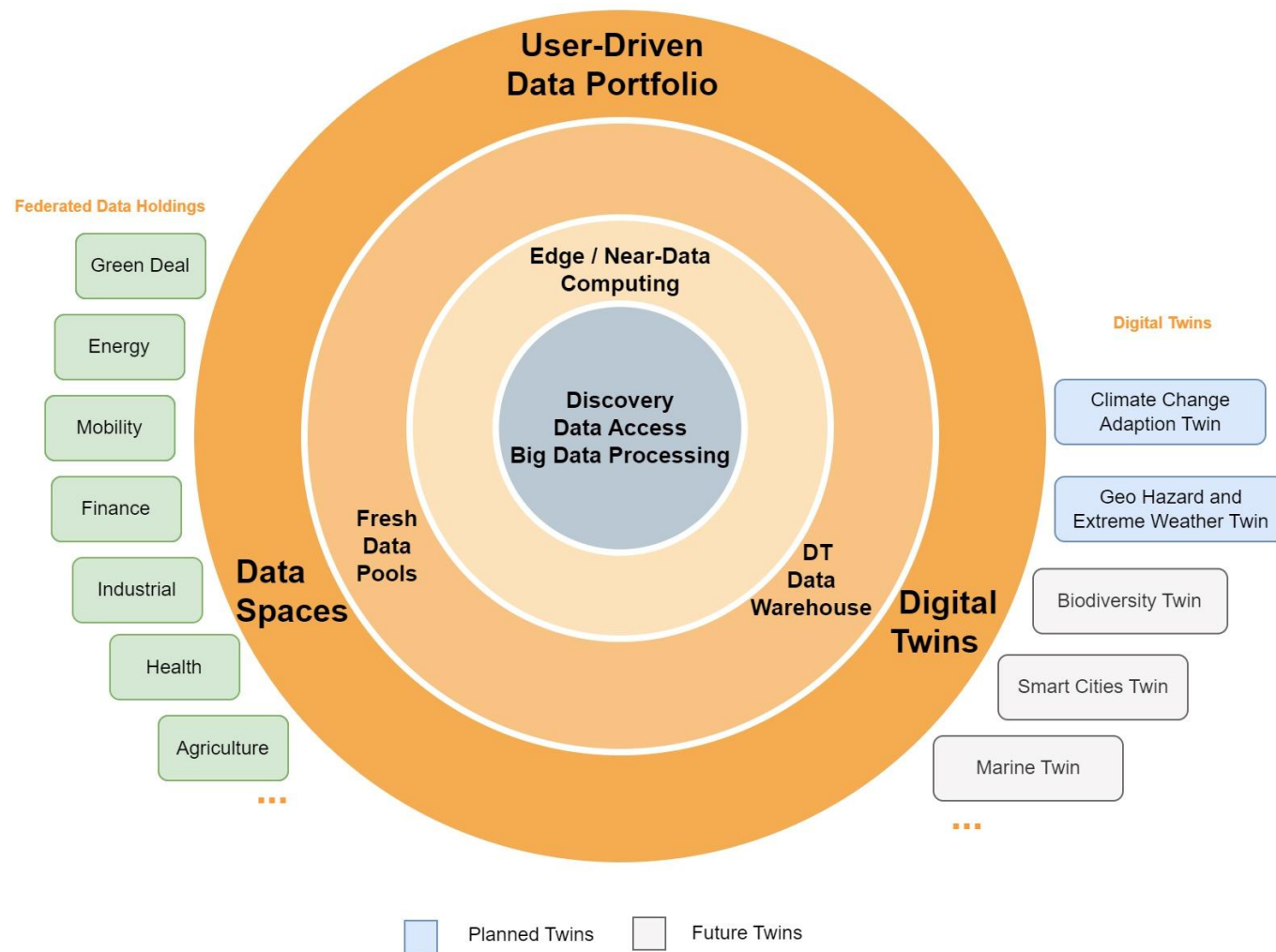
- Core Service Platform interfacing DestinE users (ESA)
- Two Digital Twins. Extreme Weather and Climate Change Adaptation (ECMWF)
- Destination Earth Data Lake (EUMETSAT)





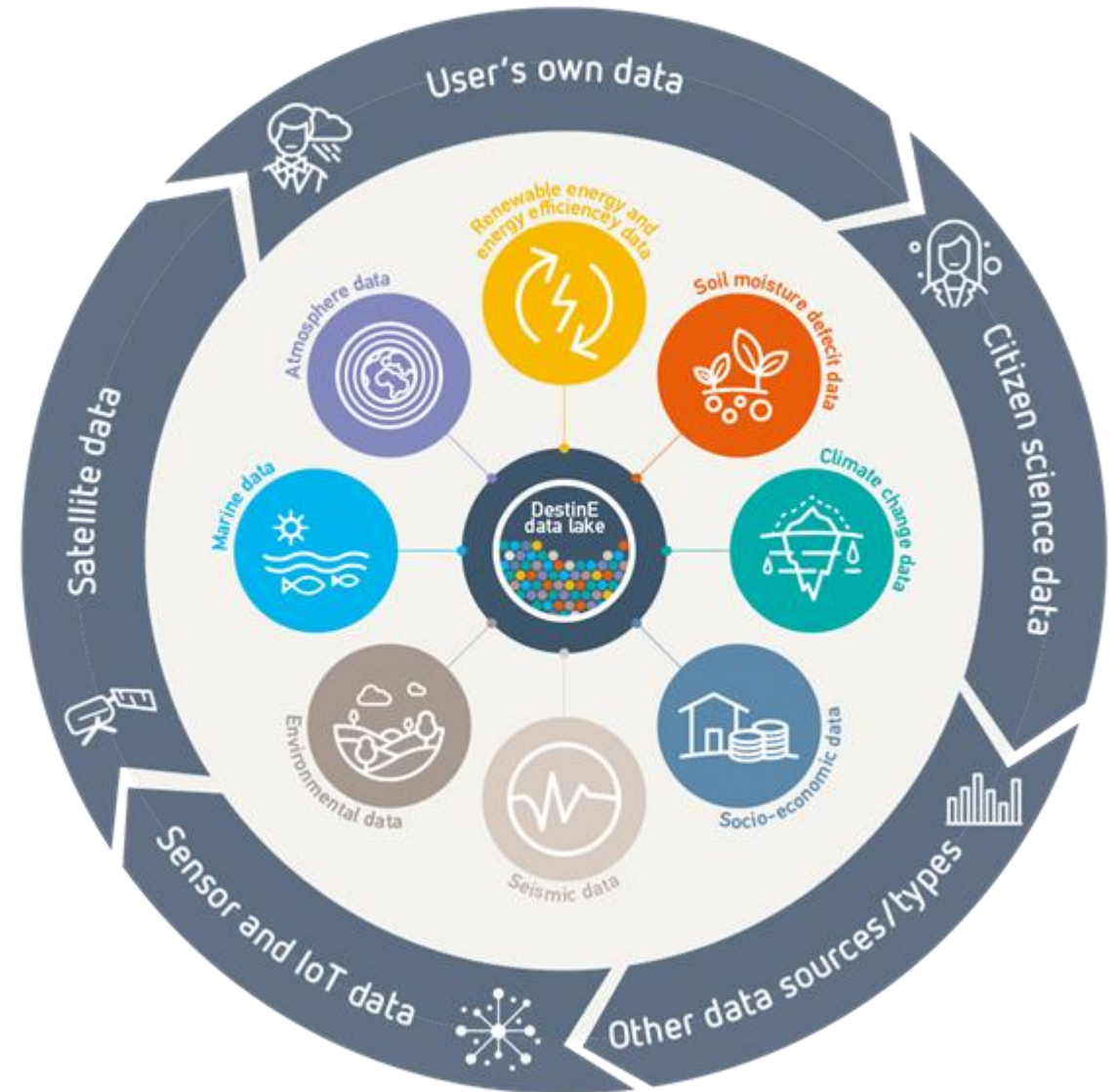
Destination Earth Data Lake (DEDL)

- 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 access, beyond anything that exists today.





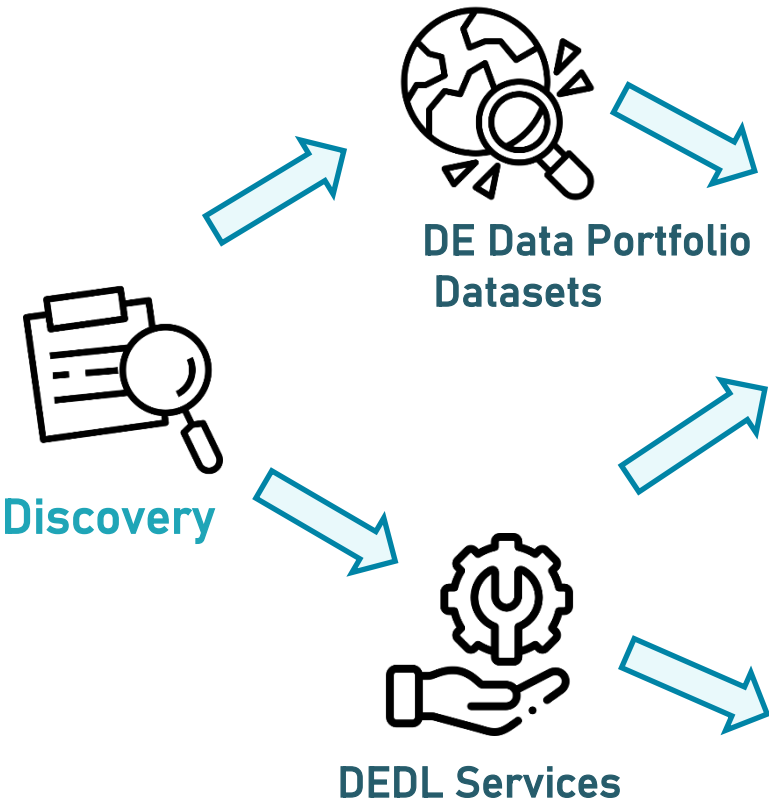
- Evolving Data Portfolio
 - User Driven
 - Managed and Controlled by DestinE Data Governance Board
- Digital Twins Data
 - Climate Change Adaptation
 - Extreme Weather and Geo hazards
- Federated datasets
 - Contributing missions (EUMETSAT, ESA, ECMWF)
 - Copernicus Satellites & Services data
 - Eurostat
 - ISIMIP
 - IAGOS
- Harmonized APIs: STAC compliant



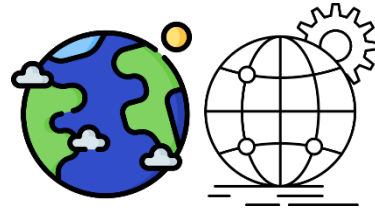


Discovery & Data Access Services

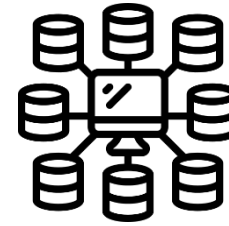
Discovery Services



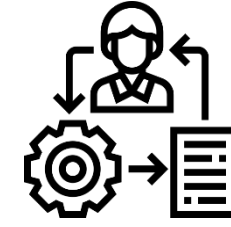
Data Access Services



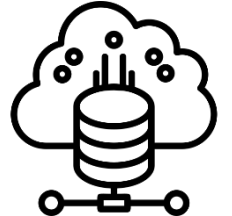
Digital Twin Outputs



Federated Datasets



User-Generated Data



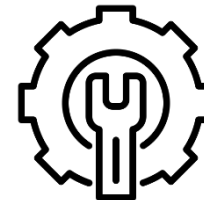
Fresh Data Pool



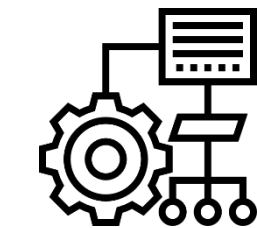
Big Data Processing Services



Infrastructure & tools (Islet Service)



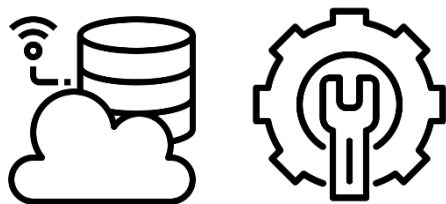
Hosted Applications (Stack Service)



Functions (Hook Service)



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



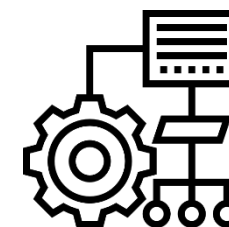
Stack Service

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



DestinE Data Lake Timeline

- **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)
- **Increment 2 – Q4 2023**
 - Extreme Weather and Geo hazard Twin
 - Data Portfolio (next priority datasets)
 - Enhancements of Increment 1 services
 - SLA and capacity increase
- **Increment 3 – Q1 2024**
 - Enhancements of Increment 2 services
 - Data Portfolio (next priority datasets)
 - SLA and capacity increase



Thank you!
Questions are welcome.