

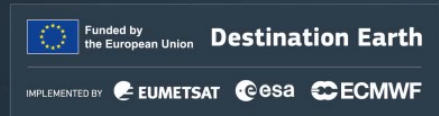
Facilitating Artificial Intelligence Activities in the DestinE Data Lake

*3rd DestinE User eXchange, 15-16 October 2024
Miruna Stoicescu (EUMETSAT)*





DestinE Data Lake – a possible journey





DestinE Data Lake – Working Topic(s)

Enhance Near-Data Processing Services

Improve digital twin outputs integration with other DEDL data (ease of use, data fusion capabilities)



Evolve AI/ML Preparedness

- Evolve DEDL big data processing framework to support AI activities (small model development, training and inference)
- Provide tools and services to generate AI/ML ready data
- Explore satellite-based training data (for AI/ML, data curation)

Contribute to User Engagement

Participate in user engagement, training and capacity building activities (MOOC, User Exchange)

Support Co-Design activities



Support integration and interoperability

Prepare for integration with SIMPL (first PoC expected early 2025)



Focus for 2025

Inject new features from 2024 user exchange into incremental developments

Increase service catalogue and capacities

Continue supporting 3E & EC use cases and emerging end user services



Based on DestinE Data Portfolio, establish the following:

- **AI ready data:**

- facilitate preparation of data to be usable for AI (fuse, tailor & harmonise)

- Geo-referencing and sub-setting, Resampling, Rearranging and reshaping, Normalisation, Dimensionality reduction, Data labelling

Edge Service Users

- **AI model generation:**

- support the development of small to medium scale machine learning models

- Splitting the data, Subset generation, Training data, Test/ Validation data, Build the model, define model architecture/ hyper parameters, Fitting the model, Evaluate the model, Published

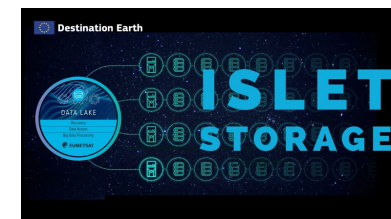
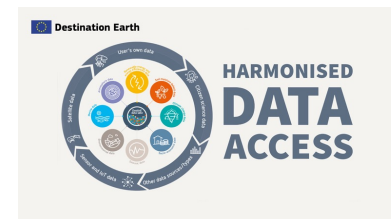
- **AI model usage:**

- facilitate execution of Inference on existing models

- **AI ready infrastructure**

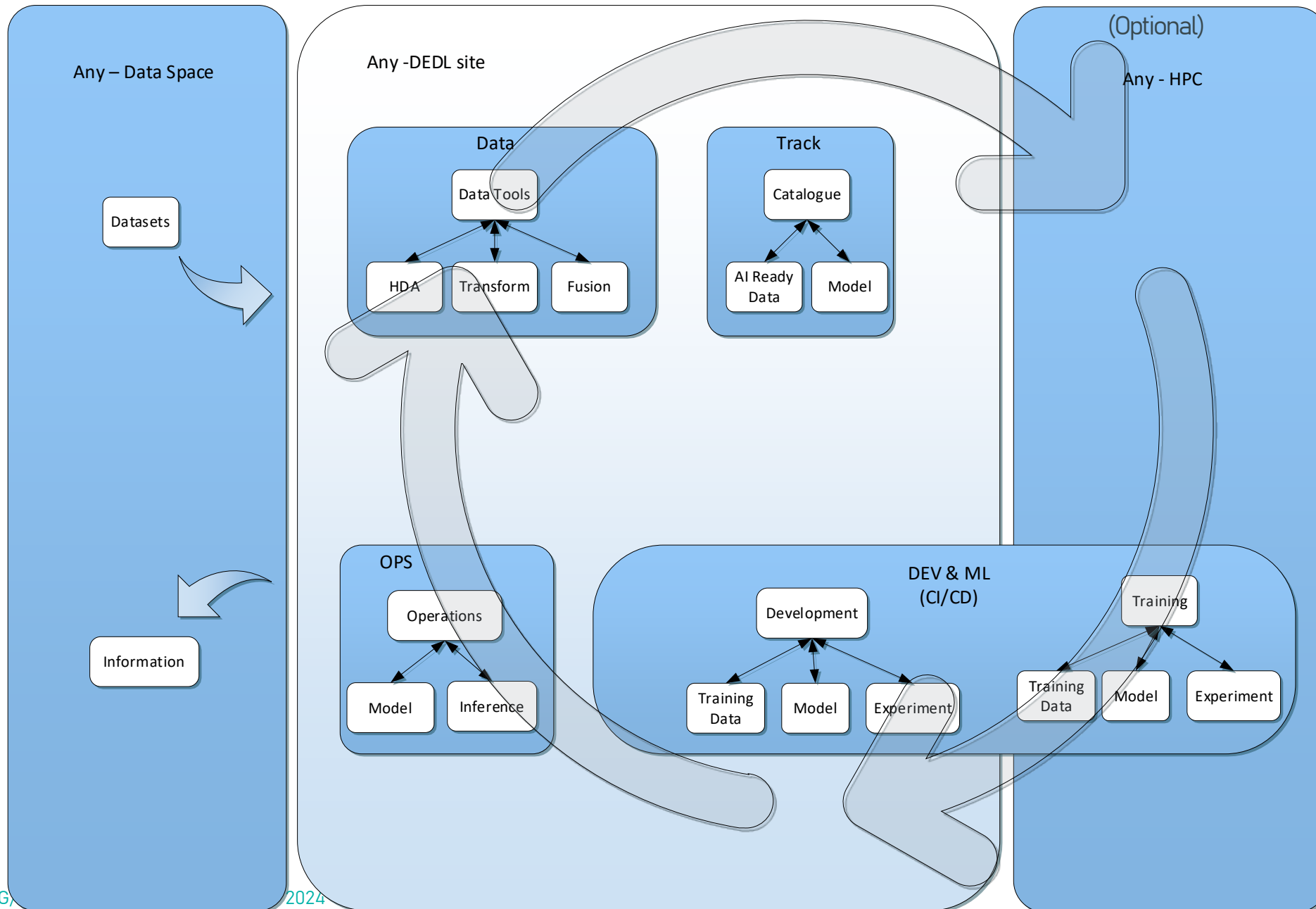
- Openstack IaaS making use of GPU/ vGPU

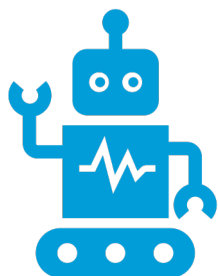
Provide services/ tools or functions in reference to DestinE Data Portfolio





DestinE Data Lake – MLOps





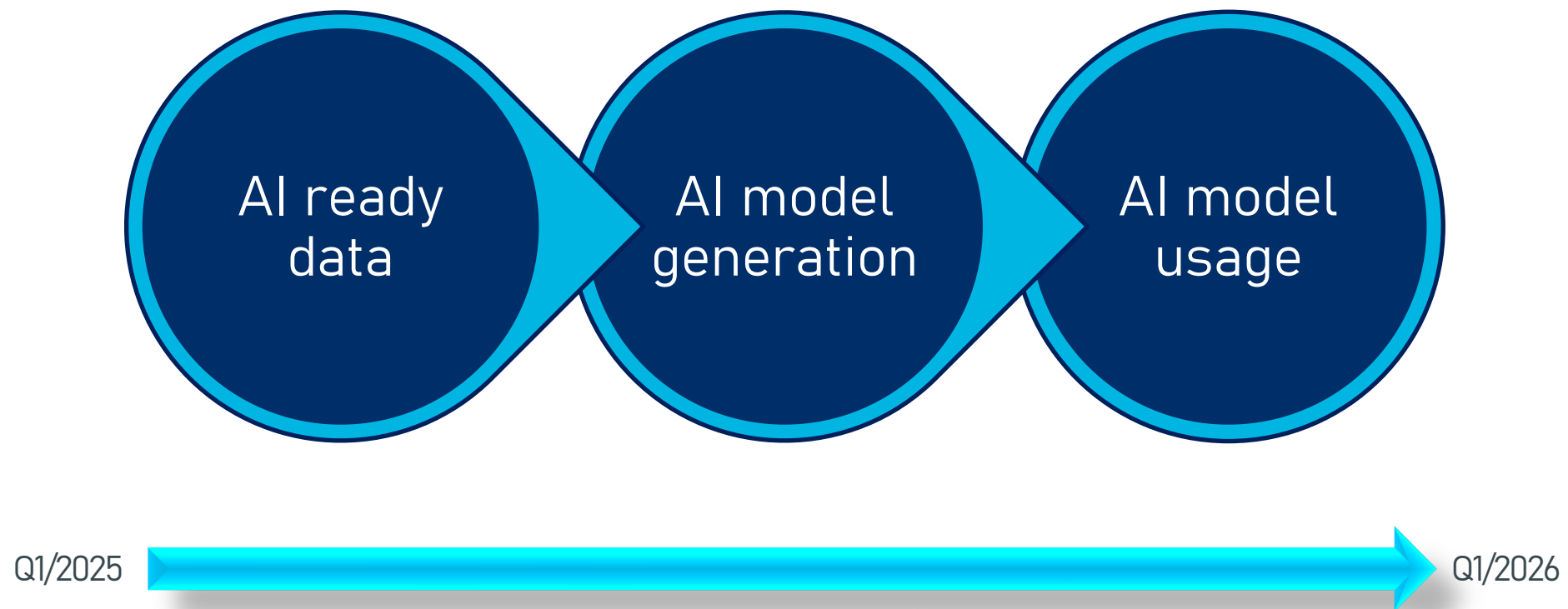
AI/ML application demonstrators using DestinE – ITT Q4 2024

Objective: Develop, validate, and implement AI/ML applications using Destination Earth Data Lake (DEDL) data & services. Focus on data originating from EUMETSAT, meet downstream end-user needs and integrate into operational DestinE services. Open-Source assets.



Tool for generation of AI Ready data - ITT Q4 2024/Q1 2025

Objective: procurement of a toolset/ library enabling the transformation of data provided by the Destination Earth Data Lake in view of efficient exploitation in AI/ML frameworks and activities (generation of “AI-ready data”).





Thank you!
Questions are welcome.