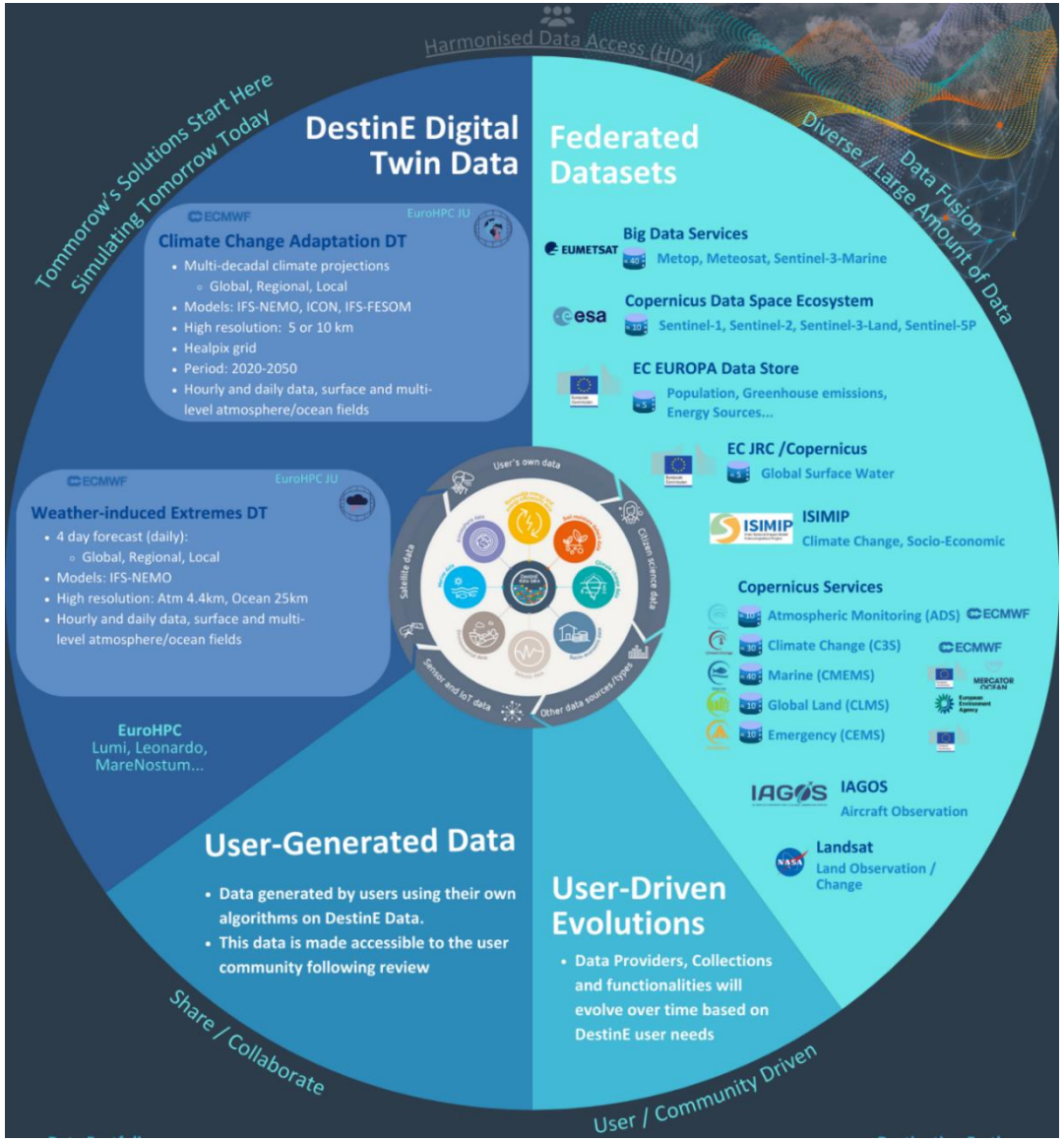
On-Going AI-Related Activities in Destination Earth Data Lake

Miruna Stoicescu on behalf of DestinE Data Lake Team

AI for Preparedness Workshop 17/06/205

Wealth of Data on DestinE Data Portfolio

- Large-scale Heterogenous Data
  - Model data
  - Satellite data
  - Socio-economic data
- ML to derive insights from data
- Challenges:
  - Infrastructure needs
  - Data Access and Harmonisation
  - High-speed data serving
  - Reproducibility



AI-Related Developments on DEDL

- AI-ready data (Under development): DEFAIR

  Python Framework to prepare analysis-ready data

- AI-ready infrastructure:

  Infrastructure as a Service (IaaS) providing GPU/vGPU

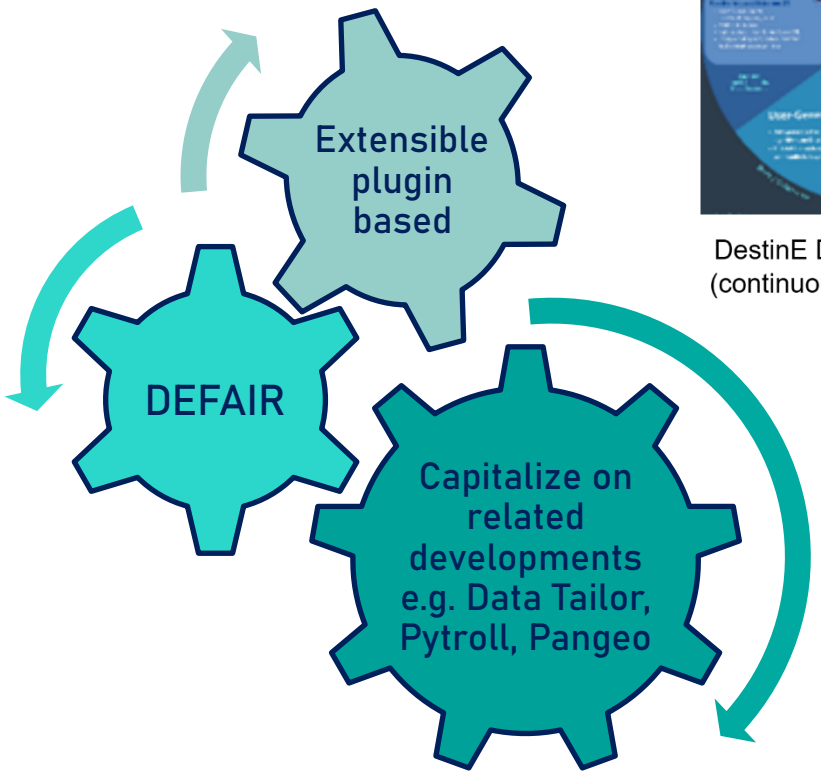
- AI model development:

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

- AI model usage:

  facilitate execution of Inference on existing models

Destination Earth Framework for AI-Ready Data (DEFAIR)



DestinE Data Portfolio (continuously evolving)

DEDL Harmonised Data Access API

Object Storage S3 interface (DEDL Fresh-Data Pool, user bucket...)

Local user storage

....

NetCDF

GRIB

MSG native

EPS native

....

Resample

Reproject

Extract ROI

Reformat

.....

xarray

Zarr

...

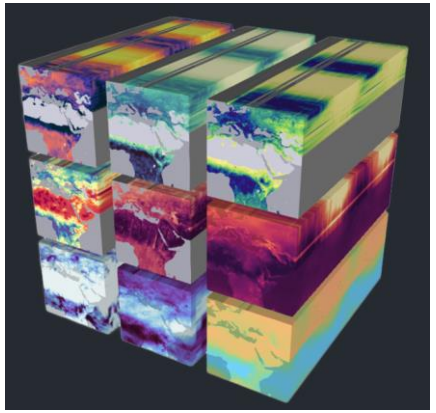
SciPy

Tensorflow

Anemoi

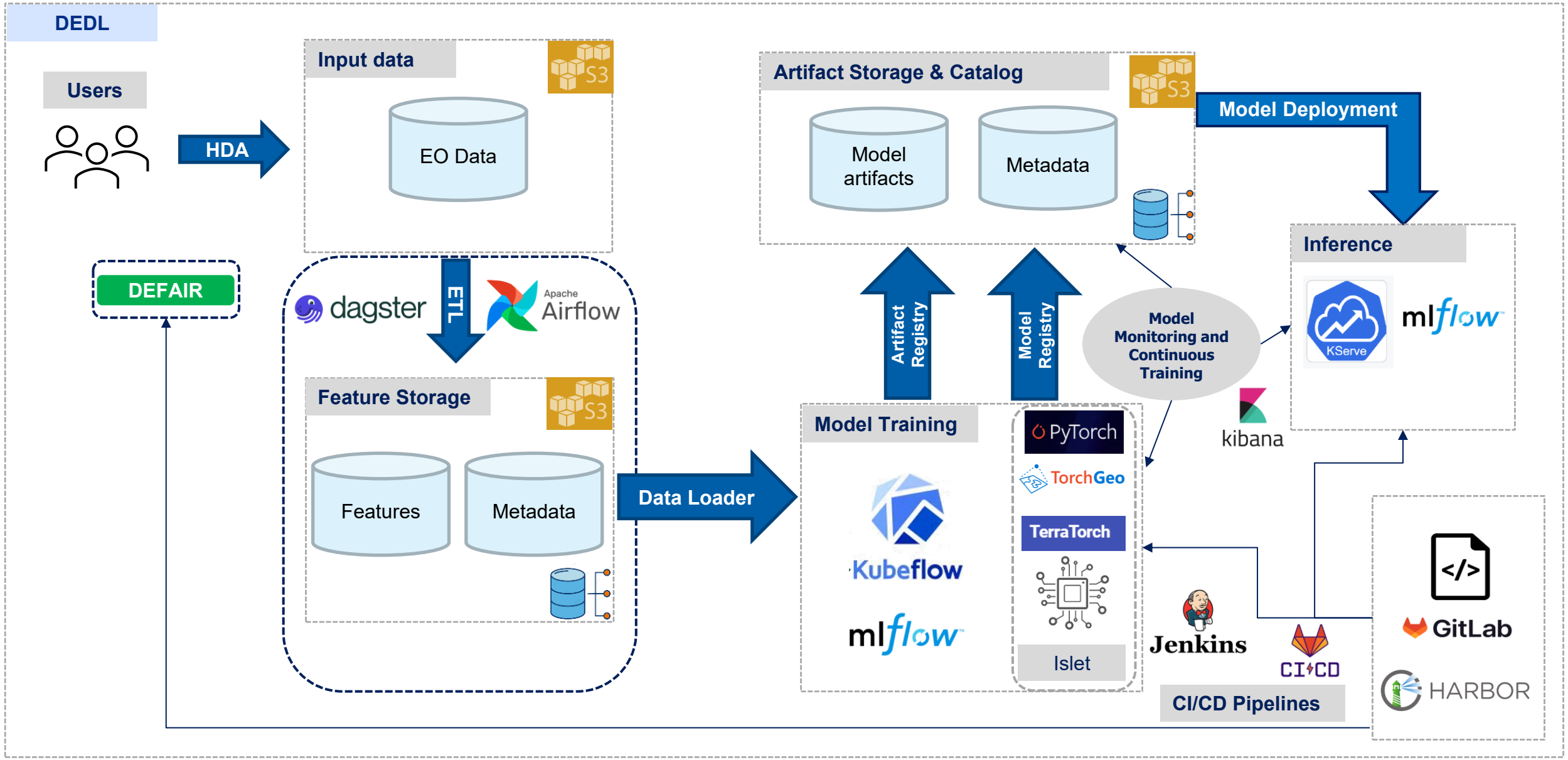
....

- ❑ Dataset versioning
- ❑ Data Lineage
- ❑ Metadata propagation
- ❑ Flexible and Modular



© Environmental Data Science and Remote Sensing group, Leipzig University

Envisaged MLOps on DEDL



AI/ML Application Demonstrators

Goal: Assessing AI and MLOps capabilities of DEDL services via applications exploiting DestinE data and EUMETSAT satellite data

- Selected end users are involved from the beginning of development
- Deliverables, applications and outputs will be open-source

➢ AI4clouds
  - Enhancing short-term (up to 12-hour) cloud forecasts (cloud cover, cloud optical depth & particle phase) by merging satellite observations (MSG SEVIRI/MTG FCI) and Extreme DT data

➢ Enhancement of MTG Lightning Imager (LI) background resolution
  - Combining the strengths of LI (high temporal resolution) and FCI (high spatial resolution)

➢ ML-based super resolution of MTG FCI L1C, Level 2 data (e.g. cloud mask) and RGB products

Proof-of-Concept: Application Development on HPC

- On-going activity on two HPC sites: MeluXina and TU Wien HPC (via EODC)
- Goals:
  - Capacity building to be able to support DEDL users in similar endeavours
  - Identifying and solving challenges such as:
    - efficient data transfer to the HPCs
    - application packaging and execution orchestration
    - efficient resource usage

Thank you!

Questions are welcome.