SESSION A4 - DATA LAKE AND DATA SPACES

FROM A DEFINED TOWARDS A USER-DRIVEN DATA PORTFOLIO

DATA GOVERNANCE

Michael Schick - EUMETSAT

Destination Earth User eXchange – 14th November 2023

Funded by the European Union
Destination Earth implemented by ECMWF esa EUMETSAT

AGENDA

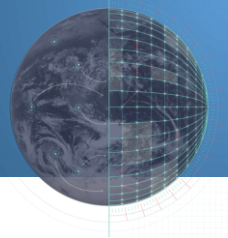
- Data Classification
- Data Policy & Data Stewardship
- Data Access Control
- Data Lifetime and Retention Management
- Data Attribution
- Data Provenance
- Data Quality
- DestinE Data Governance

DATA CLASSIFICATION

DestinE provides access to two categories of data and information DestinE Data and DestinE Metadata Information, which includes:

- **DestinE Digital Twin Outputs**
- **DestinE User-Generated Data**
- **Federated Data**

governed by DestinE Access Governance



DATA POLICY & STEWARDSHIP

Federated datasets must be managed in **accordance** with the **Federated Data Provider** and **EC** policies. Entrusted Entities must comply with pre-defined formats, timeliness, dissemination, licensing conditions, and security limitations. **Ownership** of Federated Data **is not transferred** to the EU when used in DestinE.

1. **DestinE Data and DestinE Metadata Information**

   DestinE Data and Metadata Information policy is based on the definition in the Contribution Agreements concluded with the 3EEs. This policy allows users to reproduce, distribute, communicate, adapt and modify the DestinE Data and Metadata Information. Certain limitations apply such as respecting the formats, timeliness and dissemination characteristics of the data used. Access to the data is ensured through the DestinE Access Governance. Additionally, Member States are responsible for providing warnings for high impact meteorological events.

2. **DestinE Digital Twin Outputs**

   The EU is owner of all DestinE Digital Twin Outputs. ECMWF will oversee and manage the digital twins for Climate Adaptation Change and Extreme Weather Twins in view of managing the data assets, ensuring data quality, while the data policy implementation is a joint 3E responsibility.

3. **DestinE User-Generated Data and Information**

   DestinE users are owners of the data they generate, provided it has been transformed. If they believe their data could be valuable to others, they can submit a Change Request (CR) proposal to the DestinE Joint Configuration Control Board (JCCB). The JCCB will then decide on the ownership of the data and copyright of newly generated datasets.

4. **Federated Data**

   Data Federations are managed by submitting CRs to the JCCB. Owners maintain ownership of Federated Data and 3EEs must ensure DestinE users can access it according to policy and stewardship of the Data Lifecycle.

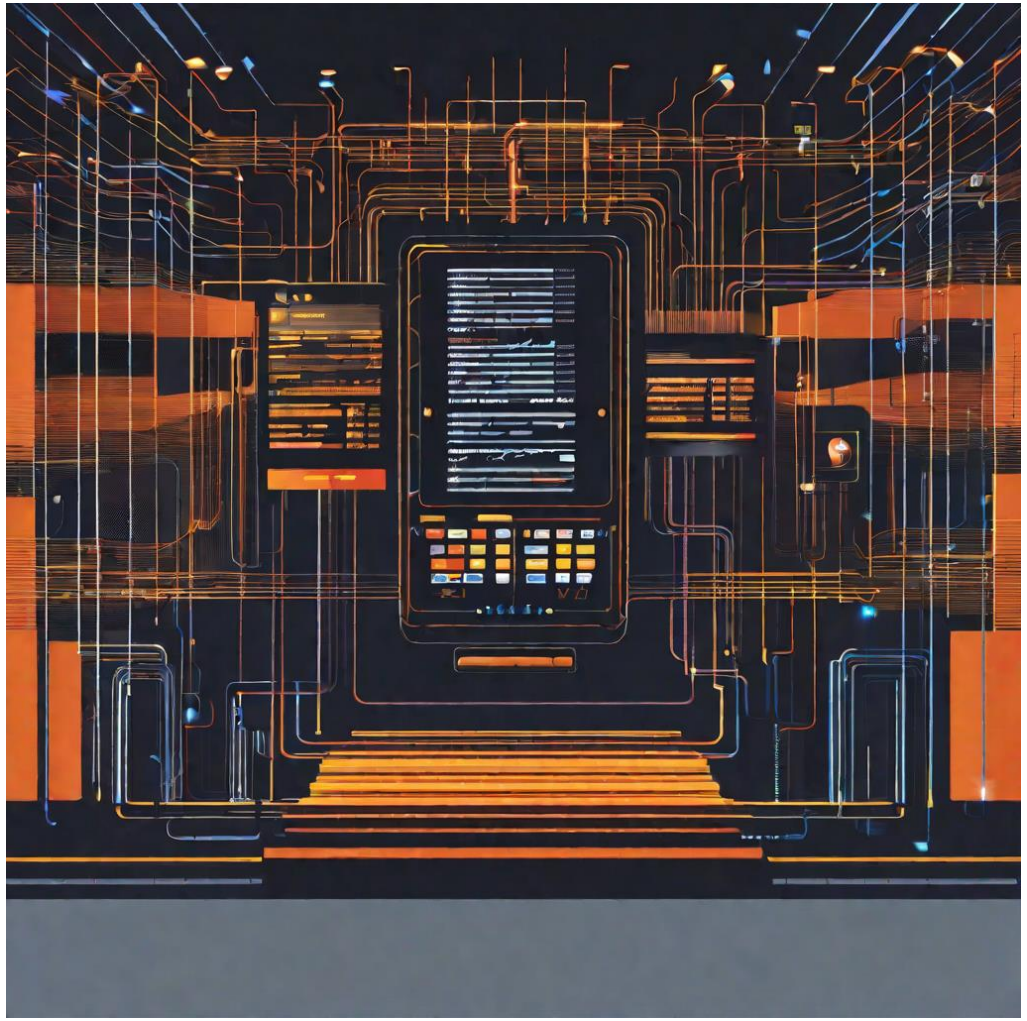
DATA ACESS CONTROL

DestinE Data and DestinE Metadata Information

- DestinE Data and Metadata Information is accessible and downloadable to all DestinE users, in accordance with the FAIR principles and Terms & Conditions. Restrictions to access may be implemented through user access classes on the DestinE platform

Federated Data

- In order to be able to access Federated Data, DestinE users must agree to Terms & Conditions attached to the respective Federated Data



DATA LIFETIME & RETENTION MANAGEMENT

The DestinE Data Portfolio outlines the availability and retention policy for DestinE Digital Twin Outputs, user-generated data and information, and Federated Data.

The DestinE JCCB endorses changes to the content of the Data Portfolio which are implemented by the DEDL.

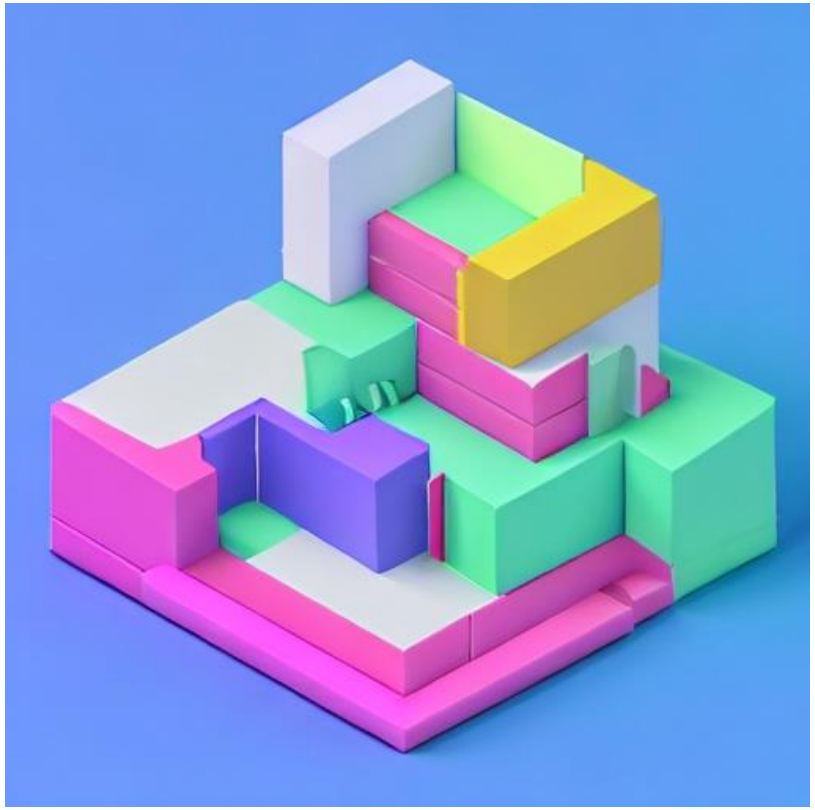
The retention time is defined as the duration in time a dataset is stored, accessible and available in the DEDL.

DATA ATTRIBUTION

Data attribution is the process of recognizing the source of a piece of data, which helps to ensure transparency and accuracy in data sharing.

DestinE users must accept Terms and Conditions attached to the data, and Digital Object Identifiers (DOIs) shall be assigned to DestinE datasets.

Dataset metadata will feature optional a dataset citation text field, allowing the data provider to insert a copyright or ownership statement.



DATA PROVENANCE

DestinE follows FAIR guidelines to ensure data reusability. It uses the Spatio Temporal Asset Catalogue (STAC) family of specifications to standardize geospatial asset metadata. STAC contains a common metadata model and a harmonised dictionary of keywords for metadata information at dataset level.

1. **DestinE Digital Twin Outputs -** Fields such as creation time, originating centre, algorithm, model version and configuration must be specified.

2. **DestinE User-Generated Data and Information -** The DestinE DEDL is responsible for obtaining data provenance information for management of the dataset.

3. **Federated Data -** The Federated Data providers are responsible for providing data provenance information, e.g. in the form of metadata attributes inside the products.

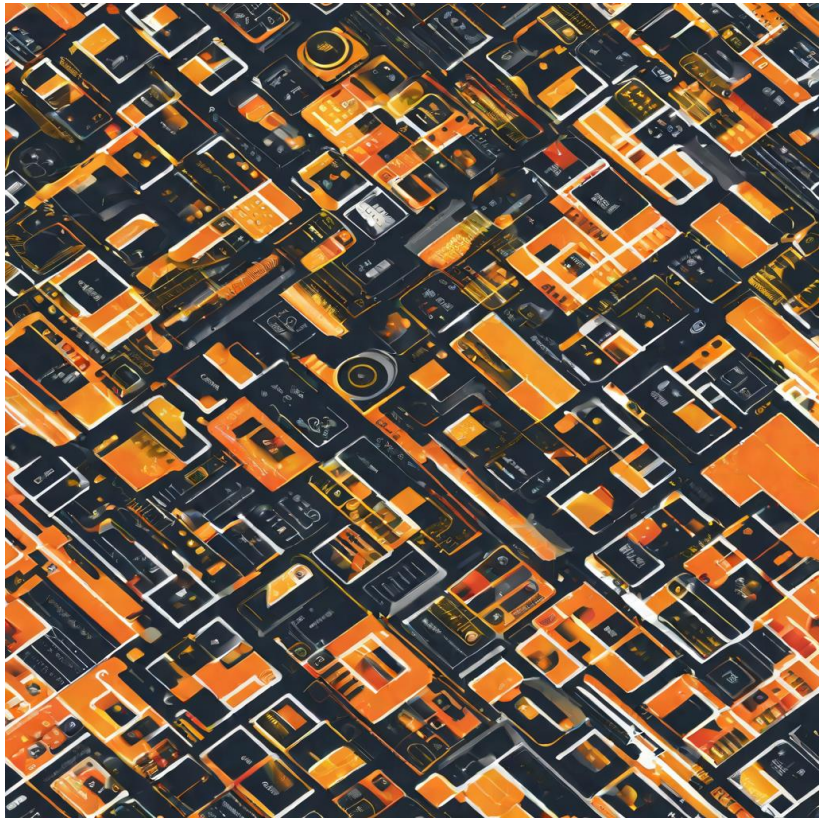
DATA QUALITY

Users of DestinE Data, DestinE Metadata Information and/or Federated Data must accept that these services are provided without any warranties regarding quality, accuracy, availability, reliability, speed and suitability. Users must also renounce any claims for damages against the European Union and the providers. This waiver includes any dispute, be it in court, arbitration or any other form of dispute settlement.

DestinE Digital Twin Outputs - Data quality information will be provided along with the data output to facilitate data usability.

DestinE User-Generated Data and Information - As part of the provided change request for storing a user generated dataset, the user shall provide information about the data quality information in order to facilitate data usability.

Federated Data - The Federated Data providers must provide quality information in various form (e.g. quality levels going between a minimal and a maximal value, number of missing data records, flags specifying whether the data is degraded or not etc.). This information is usually specified as metadata attributes inside the individual data items.



DESTINE DATA GOVERNANCE

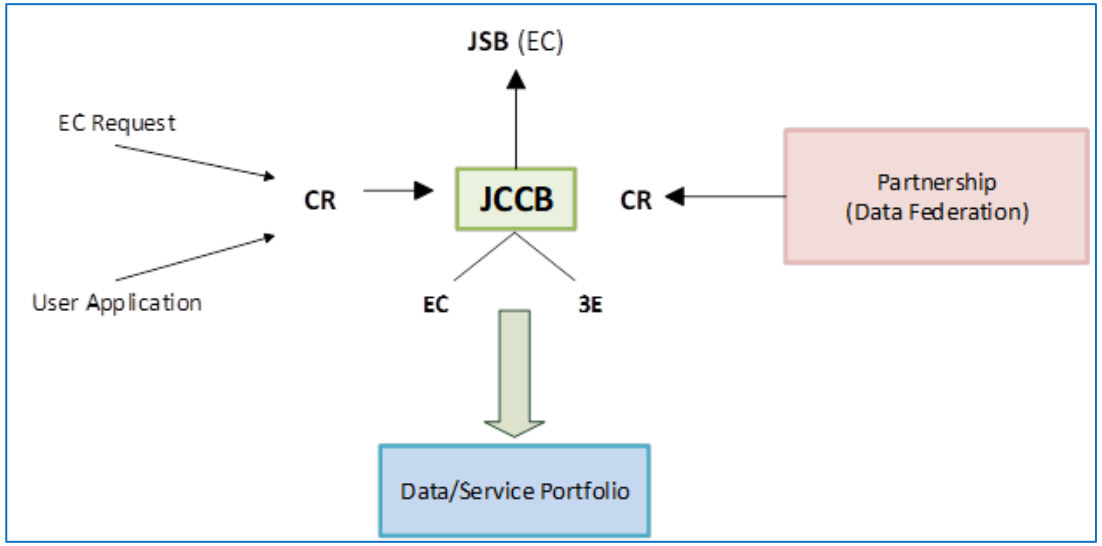
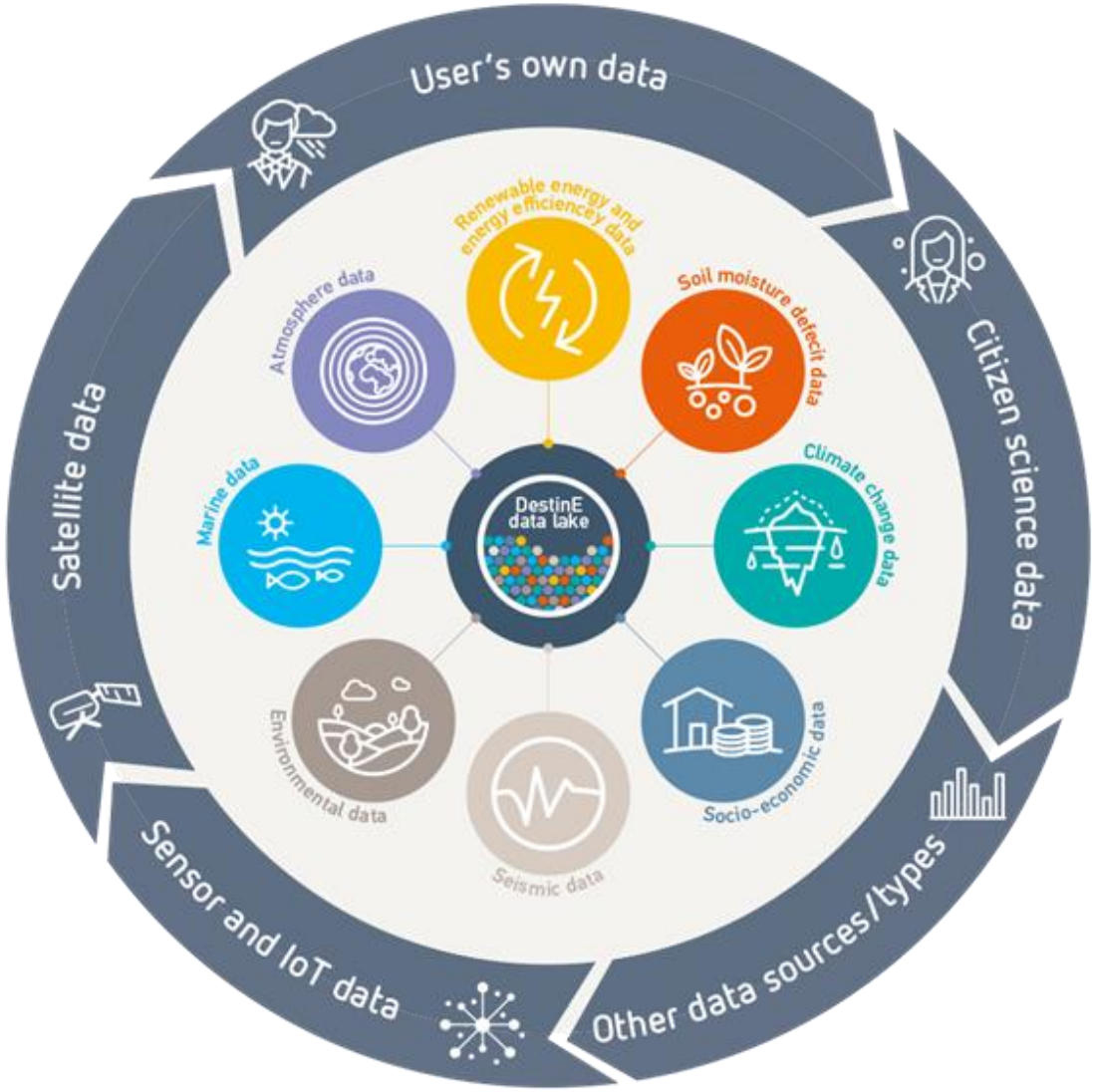
The Joint Change Control Board (JCCB) is responsible for approving change requests for the DestinE Data Portfolio.

It is composed of members of the EC (chair) and the 3EE. Its responsibilities include assessing data access policies, data lifecycle, portfolio updates, user-generated data, compliance to standards and data-related regulations, enhancing usability and uptake, and data quality.

The JCCB assesses change requests made by user requests, the EC, or the 3EE.

A particular case of interest is user-generated data and information, which the JCCB reviews against criteria such as value and usability.

In case of dispute, an issue may be raised to the Joint Steering Board.



THANK YOU VERY MUCH