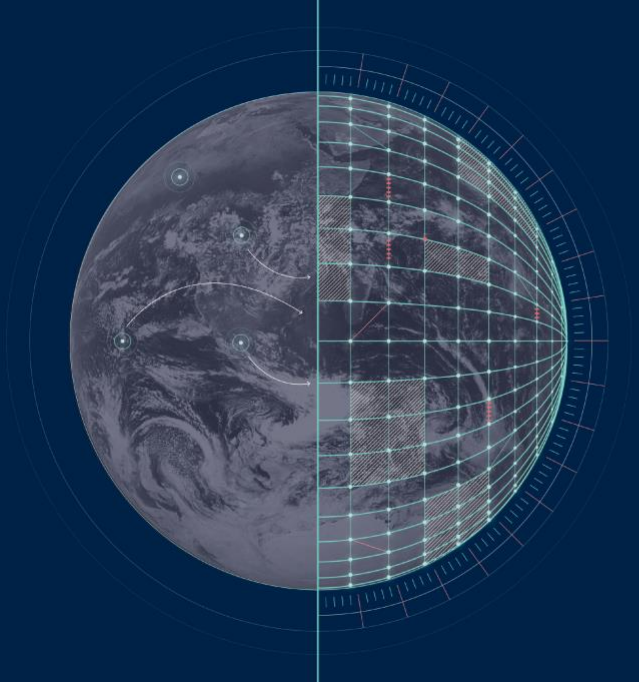

Destination Earth

Deliver DT Test Data

Issued by: ECMWF
Date: 15/12/2022
Version:1.0
REF.: DE-DTE-2-22.D4
Status: Confidential

Funded by the European Union

ECMWF

Distribution List

Name	Organisation
DG CONNECT: cnect-destine@ec.europa.eu Grazyna Piesiwicz Tobias Wiesenthal Charalampos Tsitlakidis	European Commission
Peter Bauer Irina Sandu Nils Wedi Daniel Thiemert Thomas Geenen Joern Hoffmann	ECMWF
Lothar Wolf Michael Schick	EUMETSAT
Eric Monjoux Ines Sanz Morere	ESA
Exprevia Gianluca Palumbo	Contract 350

Change Log

Version	Date	Description
0.1	06/12/2022	Initial draft
1.0	16/12/2022	Internal review
1.1	14/02/2023	Updated data and catalogue links and retrieval keywords

Contents

Glossary

A glossary of relevant terms used in the context of DestinE is available at https://stories.ecmwf.int/destination-earth-glossary/index.html.

1 Executive Summary

DestinE will deploy several highly accurate thematic digital replicas of the Earth system or Digital Twins (DTs) to monitor and simulate natural and human activities as well as their interactions with the aim to support European policy making.

Deliverable DE-DTE-2-22.D4 is a dataset deliverable marking the availability of Digital Twin test data via components of the Digital Twin Engine. This document is a supporting report which explains the contents of the test data and how it may be accessed.

For a higher-level view of the data management strategy and the Digital Twin Engine, refer to DT-DTE-2-22.D3; for more technical information on underlying I/O systems refer to DE-DTE-2-22.D1; for more detail about the prototype catalogue of DT production data refer to DE-EDT-5-22.D3.

References

[DTE-TST-001] System Integration and verification test plan

[DTE-ADD-100] Data Lake - High Level Description & Architecture

[DTE-ADD-200] Digital Twins - High Level Description & Architecture

[DTE-USR-010] DT on Climate Adaptation – Architecture & Models

[DTE-USR-020] DT on Weather-induced and Geophysical Extremes – Architecture & Models

[DTE-ICD-200] Digital Twins – Interface Control Document

[DTE-OPS-200] Digital Twins – Operations Concept

[DE-ECMWF-AWP-2022] ECMWF Annual Workplan 2022

[DE-DTE-2-22.D3] DT Data Management Plan and Preparation for Data Interfaces

[DE-DTE-2-22.D1] Description of the High-Performance Data API and other Technical I/O Developments

[DE-EDT-5-22.D3] First prototype catalogue of DT production data

2 Introduction

Simulations were performed using the Integrated Forecasting System (IFS) at 9 km and 1.4km resolution. The November to February (NDJF) dataset is publicly available and provided here as a prototype DE dataset to simulate storing Extremes DT data in the DEDL. The goal is to store as much of this dataset as possible within the FDB/MARS high-performance data storage system, using the GRIB2 data format, and served via the Polytope Service.

Note that this initial dataset is made available using existing ECMWF infrastructure. This document supplements the dataset by describing the contents of the dataset and the means of access. It will also briefly explain how this technology will map to the EuroHPC machines and associated Destination Earth Data Lake (DEDL) Data Bridges.

This document supplements the dataset by describing the contents of the dataset and the means of access, as well as giving some context into how the relevant components will be utilized in Destination Earth.

3 Test Dataset Description

The following provides a summary of the **raw** dataset, including the packing, levtype, and the number of messages which can currently be processed into GRIB2. The data volume is significantly decreased by using compression, and this will be a critical requirement for storing DT output on the DEDL. We observe a **71%** saving per step in the surface data and an **84%** saving per step in the upper air data.

The GRIB2 conversion and processing of this data is critical to ensure the metadata description adheres to the WMO conventions and is therefore understandable in a systematic and centre agnostic manner. The ECMWF software ecCodes is used to do this, and the GRIB2 encoding updates described previously will allow this data to be understood by any user.

	Gaussian Grid Fields	**Spherical Harmonic fields**	**Upper Air Fields**	**Totals**
Packing	grid_simple grid_ccsds	spectral_complex	grid_simple grid_ccsds	**-**
No. of vars with GRIB2 encoding	24 out of 96	15 out of 16	8 out of 8	**47 out of 120**
Levtype	surface	model & pressure levels	model & pressure levels	**-**
Total no of steps	961 (4 months)	241 (1 month)	241 (1 month)	**1443**
Typical uncompressed data size per step	45 GiB	115.9 GiB	258.5 GiB	**419.4 GiB**
Typical compressed data size per step	13 GiB	-	41 GiB	**54 GiB**

A subset of the data has been archived on the MARS Scratch server at ECMWF. The catalogue for this dataset is provided on a public Confluence page here[1]. This will be

[1] https://confluence.ecmwf.int/display/DDCZ/SUMMIT+1.4+km+NDJF+seasonal+run+dataset%3A+Catalogue

replaced by an interactive data catalogue in the future. This links to deliverable "DE-EDT-5-22.D3: First prototype catalogue of DT production data". Access to this data in Destination Earth is provided by the Polytope web service, as described in the following section.

4 Data Access via Polytope Service

Polytope[2] is a horizontally scalable service which provides a RESTful data access API over HTTPS, allowing access to high-performance data storage over the internet. Its purpose within Destination Earth is to expose the data stored in the FDB/MARS (on the DEDL Data Bridge) to the DESP and to other parts of the DEDL. Polytope implements access control; and is also capable of federating access to data between different DEDL data bridges.

An instance of Polytope is currently hosted on ECMWF infrastructure, providing access to ECMWF FDB/MARS where the DT test data is stored. It is possible to directly interact with the RESTful API[3] in order to retrieve this data. However, it is more convenient to download the polytope client and use the Polytope Python API or Command Line Interface (CLI).

4.1 Using Polytope Client

The polytope client can be downloaded from PyPi:

```
pip install polytope-client
```

Retrieving data using the Python API is straightforward. Below is an example script which will retrieve one field of the test dataset. The email and api-key must be replaced with the email and api-key associated with an ECMWF account[4]:

```
#!/usr/bin/env python3
from polytope.api import Client
c = Client(user_email='<email>', user_key='<api-key>')
request = {
    'class' : 'rd',
    'date' : '2018-11-01',
    'expver' : 'hs3g',
    'levelist' : '500',
    'levtype' : 'pl',
    'param' : '129',
    'step' : '24',
    'stream' : 'oper',
    'time' : '00:00:00',
    'type' : 'fc'
}
c.retrieve('ecmwf-mars', request, "output.grib")
```

[2] https://polytope-client.readthedocs.io/en/latest/
[3] https://polytope.ecmwf.int/openapi/
[4] A (free) ECMWF account can be created at www.ecmwf.int. An API key can be obtained from https://api.ecmwf.int/v1/key/ once logged in.

4.2 Selecting Data

The catalogue, referenced in section 3.3, shows the currently archived set of DT test data (e.g. above *gets a 1-day forecast of the 500hPa geopotential height field, a standard meteorological forecast parameter*). The page gives information of the different types of data available, such as surface, model level and pressure level fields, as well as the distinction between data provided on octahedral and spherical harmonic grids.

Output fields

Surface fields

The parameters at surface level are provided on the **O8000** grid and are as follows:

paramId	shortName	Name	units
31	ci	Sea ice area fraction	(0 - 1)
33	rsn	Snow density	kg m**-3
34	sst	Sea surface temperature	K
47	dsrp	Direct solar radiation	J m**-2
78	tclw	Total column cloud liquid water	kg m**-2
79	tciw	Total column cloud ice water	kg m**-2
129	z	Geopotential	m**2 s**-2
136	tcw	Total column water	kg m**-2
137	tcwv	Total column vertically-integrated water vapour	kg m**-2
145	bld	Boundary layer dissipation	J m**-2

Several examples of retrieve requests are given on an additional public Confluence page here[5], which can be substituted in the example script in section 4.1.

Surface fields

Retrieve "Sea ice area fraction" (param 31; grid **O8000**) and "Sea surface temperature" (param 34; grid **O8000**) at step 0 (0 hours after run start: 2018-11-01, 00:00:00):

```
request = {
  'class' : 'rd',
  'date' : '2018-11-01'
  'expver' : 'hs3g'
  'levtype' : 'sfc',
  'param' : '31/34',
  'step' : '0',
  'stream' : 'oper',
  'time' : '00:00:00',
  'type' : 'fc'
}
```

Hint: double-click to select code

5 https://confluence.ecmwf.int/display/DDCZ/SUMMIT+1.4km+NDJF+seasonal+run+dataset%3A+Example+Polytope+requests

5 Data Handling on EuroHPC Infrastructure

The test data above is hosted on and served from ECMWF infrastructure. The FDB/MARS (for data storage) and Polytope Service (for data access), which enable this, are both components of the Digital Twin Engine. The Digital Twin Engine components will be deployed on EuroHPC machines and their connected DEDL Data Bridges, to build the data pipelines for Digital Twins running on this infrastructure.

Error! Reference source not found. shows the DTE components that relate to the data pipeline and data management in the context of a Digital Twin running on a EuroHPC platform. The FDB/MARS will be deployed both in the HPC partition (to store the large output of the Digital Twins) as well as on the DEDL Data Bridge (for long-term storage of reduced DT data). Polytope Service will be deployed on the DEDL Data Bridge, connected to the collocated FDB/MARS, just as it is currently deployed at ECMWF for this test data. Access to the Polytope Service will also be augmented by the Polytope Feature Extraction library, which is being developed in Destination Earth.

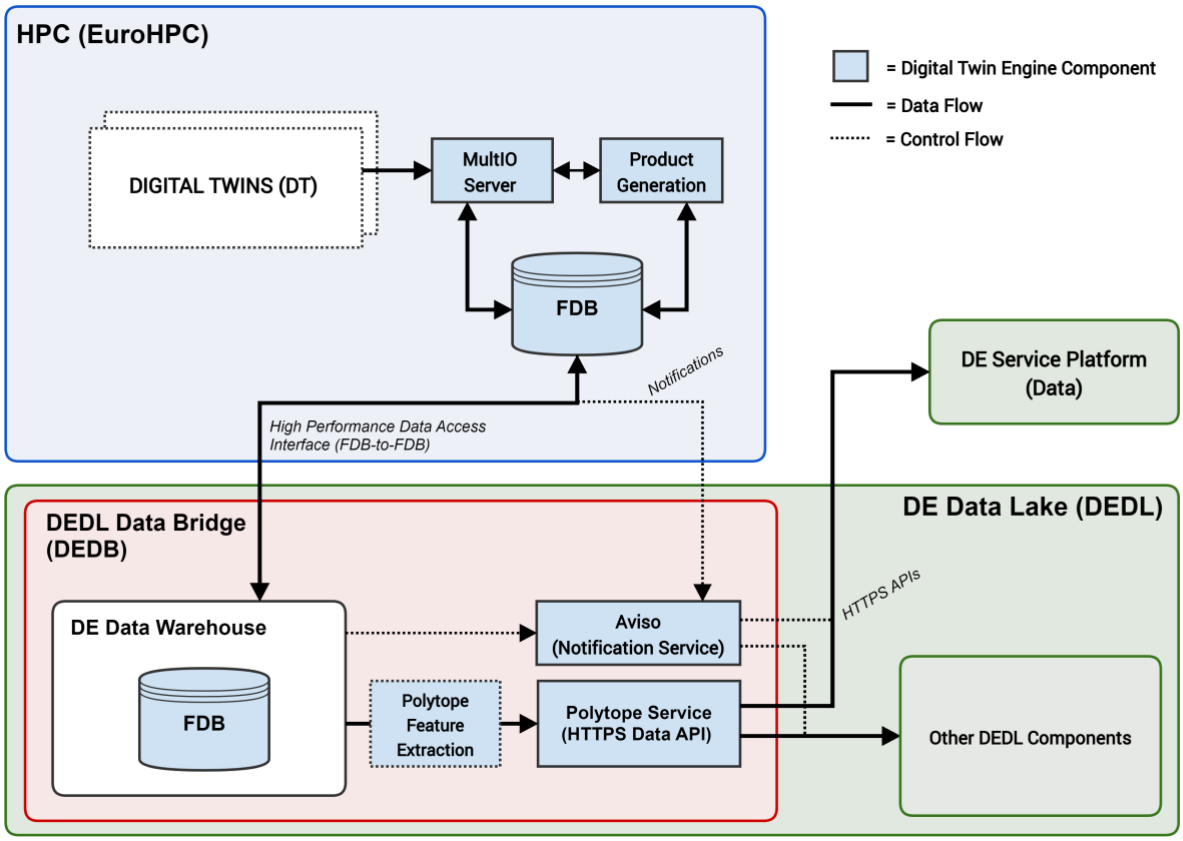

Figure 1: DTE Components involved in the data pipeline.

6 Conclusion

Digital Twin test data obtained from a 1.4km IFS simulation has been made available for download from ECMWF, using the Digital Twin Engine components MARS/FDB (for data storage) and Polytope Service (for data access). This test data gives a flavour of the DT Extremes data which may generated in Destination Earth, and therefore shows that these systems are capable of storing and delivering high-resolution meteorological data.

This document has described the DT test data that has been delivered for Destination Earth, and the process for accessing this data.

UK: (Headquarters) ECMWF, Shinfield Park, Shinfield Road, Reading, RG2 9AX, UK

Italy: ECMWF, Tecnopolo di Bologna, Via Stalingrado 84/3, 40128 Bologna, Italia

Germany: ECMWF, Robert-Schuman-Platz 3, 53175 Bonn, Deutschland

Destination Earth

This document has been produced in the context of the Destination Earth Initiative and relates to tasks entrusted by the European Union to the European Centre for Medium-Range Weather Forecasts implementing part of this Initiative.

This document is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. The European Centre for Medium-Range Weather Forecasts is not liable in respect of this document and gives no warranty for the information provided.

UK: (Headquarters) ECMWF, Shinfield Park, Shinfield Road, Reading, RG2 9AX, UK

Italy: ECMWF, Tecnopolo di Bologna, Via Stalingrado 84/3, 40128 Bologna, Italia

Germany: ECMWF, Robert-Schuman-Platz 3, 53175 Bonn, Deutschland