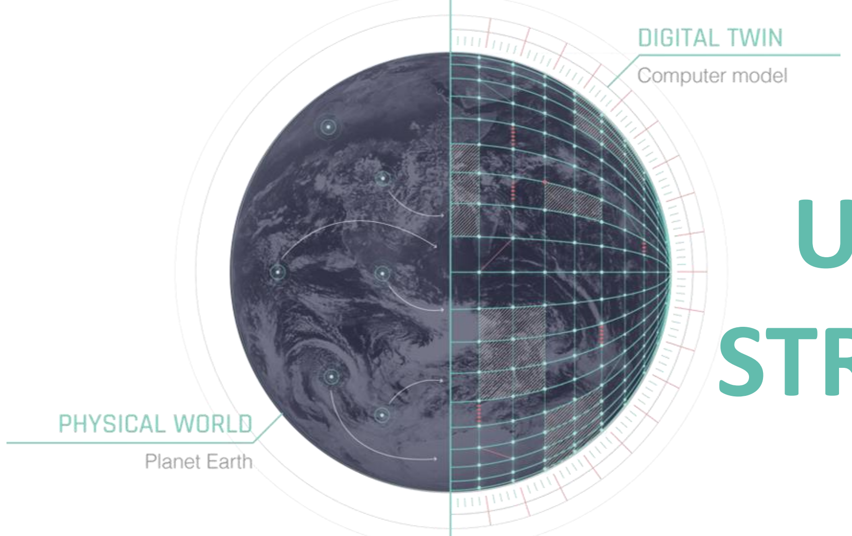
DESTINATION EARTH



USER-DRIVEN CLIMATE MODEL DATA STREAMING FOR CLIMATE ADAPTATION

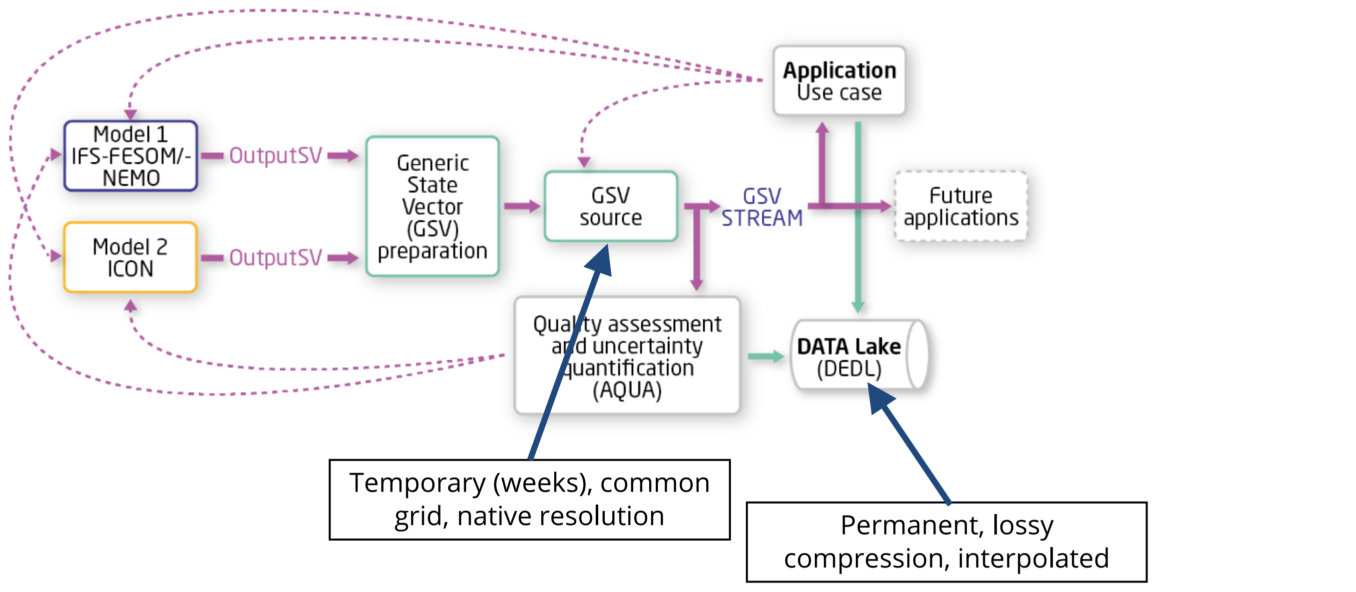
Francisco Doblas-Reyes[1,2], Christian Steger[3], Sami Niemelä[4], Jenni Kontkanen[5], Barbara Früh[3], Bjorn Stevens[6], Heikki Tuomenvirta[4], Roberto Chavez[1], Aleksander Lacima[1], Miguel Castrillo[1], and Katherine Grayson[1]

[1]BSC, [2]ICREA, [3]DWD, [4]FMI, [5]CSC, [6]MPI-M

The Destination Earth Climate Adaptation Digital Twin (Climate DT) will design and implement a climate information system running on pre-exascale high-performance computing platforms to support climate adaptation efforts. An overview of the overall Climate DT is given by Kontkanen et al. (this session). This poster focuses on the climate data flow and the streaming concept to address the requirements of a scalable number of users in the climate adaptation space.

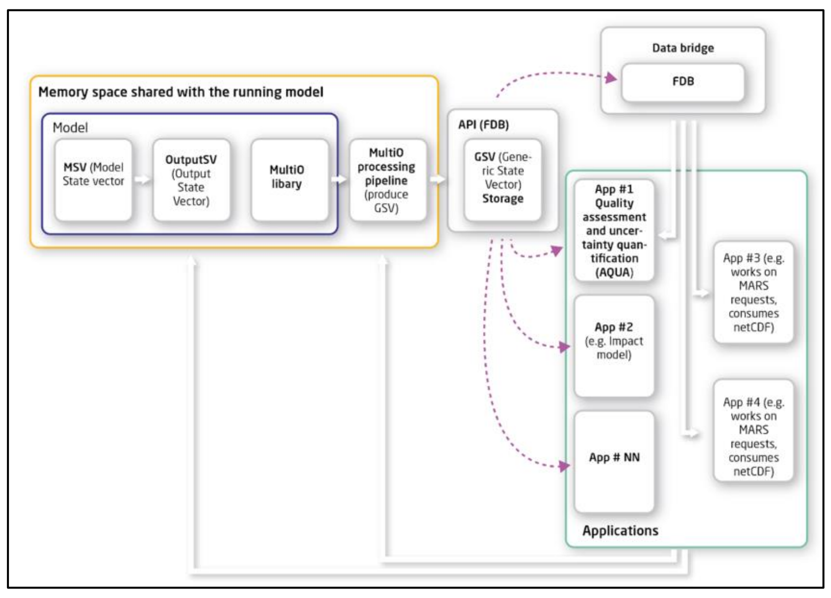
- The Climate DT will provide global climate data for both the historical period and the near-term future with unprecedented spatial and temporal resolution using a new generation of climate models.
- The full model state vector (MSV) will be available at runtime to satisfy the requirements of all the interested applications (also called data consumers), ranging from unprecedented climate analyses, the training of machine learning models for full interactivity through emulators and a range of climate impact models. Climate simulations will be running continuously, increasing the ensemble size and the experiment variety.
- The Climate DT approach will lead to an interactive system where data consumers will be able to harness the MSV in real time by configuring their access to the MSV of the running experiments, with the possibility of modifying their requests at will, and where new data consumers can be continuously added to the workflow or formulate their requirements to disseminate new variables and frequencies.
- The streaming concept is at the heart of this radically new approach. Data consumers will be able to make requests of the variables and data frequencies needed and the Climate DT workflow will stream this data to them in a similar way as more familiar streaming services work. The MSV will be fully accessible for a limited period of time (depending on the size of the HPC storage), to allow for restarting applications, determined by the capacity of the central database. The data of the simulations performed will be periodically removed to make space for the MSV of new members and experiments, not without having stored a number of key variables at reduced resolution for the long term.

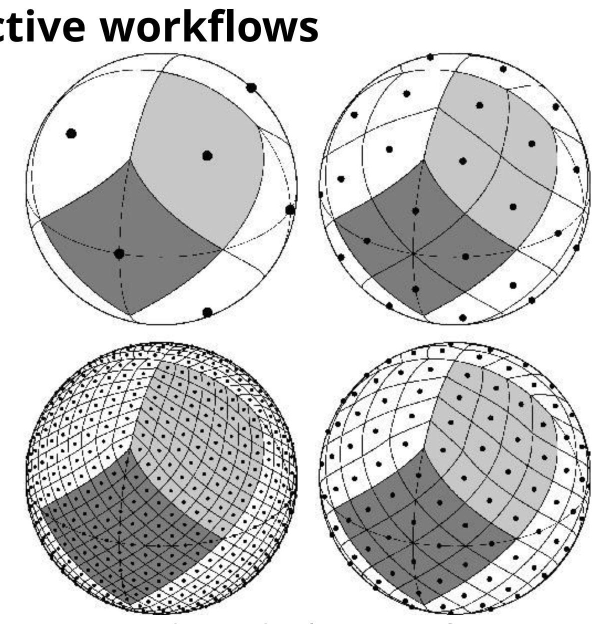
1. The Climate DT data flow



The Climate DT uses climate data from simulations performed by a set of new, high-resolution global models and allows for full interaction with the simulations as they run. The interaction is facilitated by the streaming concept, whereby data consumers are offered a configurable access to the MSV to generate unprecedented climate information for adaptation. The production of historical simulations and projections is continuous and ensemble members are staggered in the HPC to accommodate new user requirements. One (5 km) member takes around 45 days of wallclock time (using the equivalent of around 100,000 cores).

2. Unified model output and user-interactive workflows





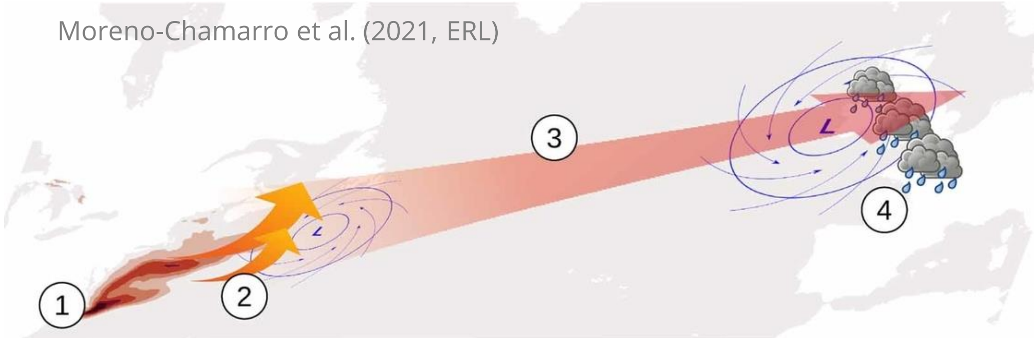
Source: https://healpix.sourceforge.io/

Climate models write the MSV in a centralised database called Generalised State Vector (GSV) in GRIB format and with the data interpolated in a common grid (the HealPIX grid closest to the native model resolution). All tasks are controlled by a configurable workflow common to all models and extended to the data consumers that ensures traceability and recoverability. The MSV is accessible through the GSV API. Data consumers access the data they require through this API via streaming before the data disappears from the database to make space for new simulations. The streaming is configurable as a client, on the data consumer side, for full scalability of the access by a large number of consumers.

3. On-the-fly, flexible diagnostics

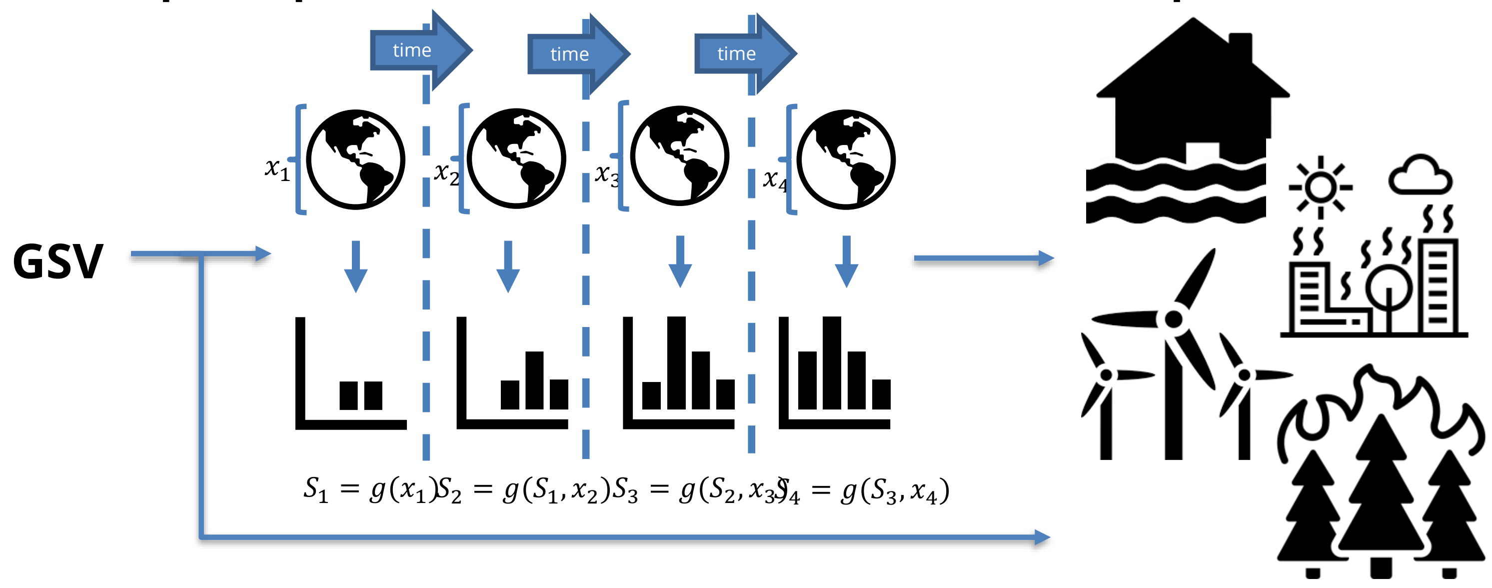
Diagnostic	Realm and frequency
Sea surface height variability	Ocean, daily
Eady growth rate	Atmosphere, 3-hourly
Tropical rainfall	Atmosphere, hourly
Tropical cyclone detection and 3D analysis along track	Atmosphere and ocean, hourly
Extratropical cyclones air-sea interaction	Atmosphere and ocean, hourly
Mesoscale ocean eddy detection and 3D analysis	Atmosphere and ocean, hourly

Moreno-Chamarro et al. (2021, ERL)

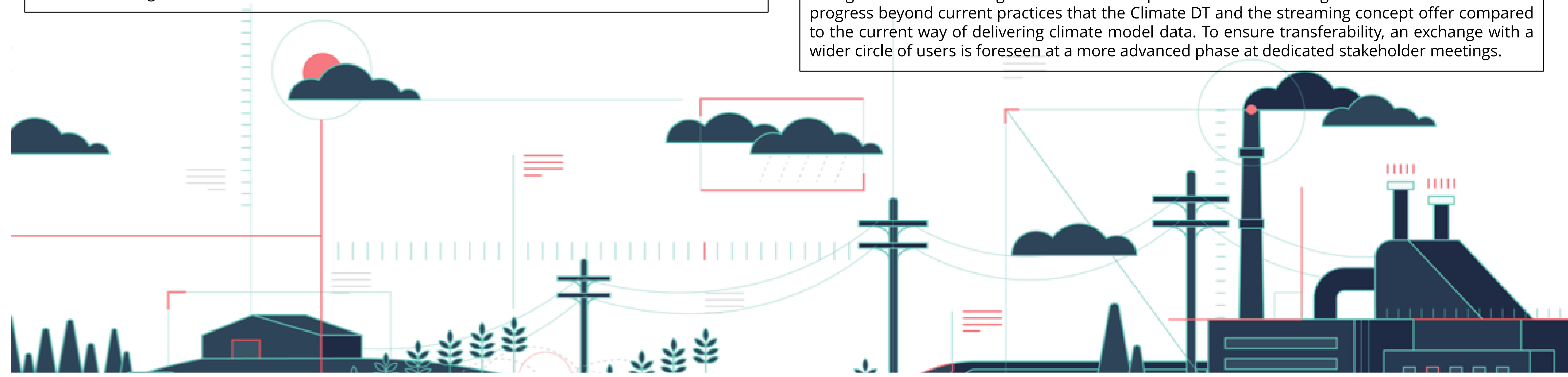


The streaming concept allows to perform climate model diagnostics that have been impossible so far, of which the list above is a selection. These diagnostics will be computed on-the-fly as the data is streamed. They allow for unprecedented analyses (similar to the one illustrated above from the PRIMAVERA EU project) from unprecedented simulations by considering the full 3D climate system from a range of models, at native resolution and with the highest data frequency. This implies reconsidering the way diagnostic software has been written so far to perform these frontier analyses as the models run their simulations. Traditional diagnostics (mean climatologies) will be also possible using the long-term archive where key variables will be stored at reduced resolution in the data bridge.

4. One-pass operators and case studies for climate adaptation



As the MSV is not indefinitely exposed, and the data at native resolution and high frequency is hard to handle by most data consumers, the streaming benefits from a set of one-pass operators. The one-pass operators help reducing both data complexity and volume according to typical user requirements like the calculation of arbitrary time averages, variances or histograms, or even bias-adjusted data. One-pass operators can buffer the MSV data of interest beyond its residence time in the central database to facilitate the synchronisation with the execution of the data consumer. The applications, initially represented by the case studies included in the Climate DT such as urban climate, forest management, energy production and hydrology, can choose to receive streams of MSV data or data from a suite of one-pass operators. The applications have the possibility to iteratively contribute to the design of the experimental set up and request as many variables and indicators as they need each time. These applications will offer technical recipes for other users to configure their streaming and link their impact models to the digital twin and illustrate the progress beyond current practices that the Climate DT and the streaming concept offer compared to the current way of delivering climate model data. To ensure transferability, an exchange with a wider circle of users is foreseen at a more advanced phase at dedicated stakeholder meetings.



Funded by the European Union

Destination Earth implemented by ECMWF esa EUMETSAT