Climate Data Infrastructure

Where Are My Data?

Graziano Giuliani
ggiulian@ictp.it

Fifth Workshop on Water Resources
in Developing Countries
Trieste 2019

WORLD CLIMATE RESEARCH PROGRAMME

ORGANIZATION

JOINT SCIENTIFIC COMMITTEE (JSC)

WCRP MODELLING ADVISORY COUNCIL (WMAC)

WCRP DATA ADVISORY COUNCIL (WDAC)

JOINT PLANNING STAFF (JPS)

WORKING GROUPS ON: COUPLED MODELLING (WGCM), SUBSEASONAL TO INTERDECADAL PREDICTION (WGSIP), NUMERICAL EXPERIMENTATION (WGNE), REGIONAL CLIMATE (WGRC)

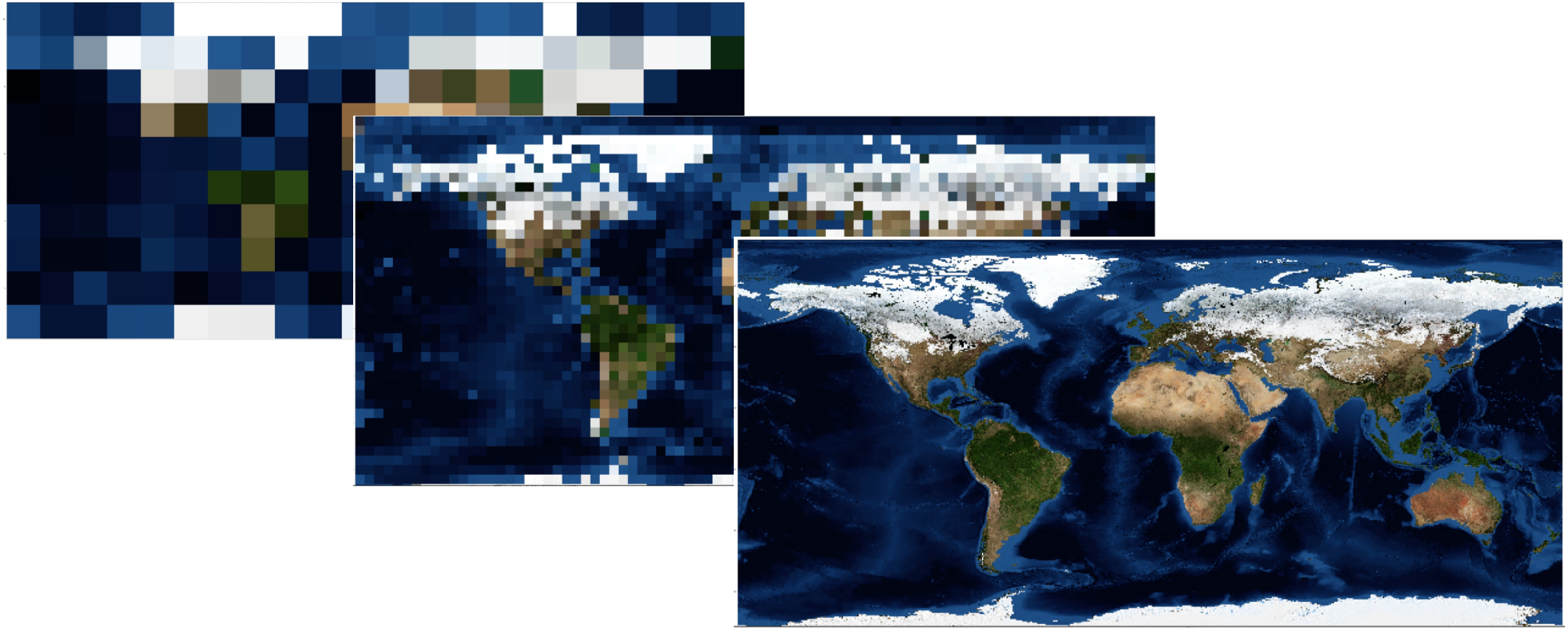
CliC	CLIVAR	GEWEX	SPARC	CORDEX
CRYOSPHERE-CLIMATE	OCEAN-ATMOSPHERE	LAND-ATMOSPHERE	TROPOSPHERE-STRATOSPHERE	REGIONAL CLIMATE DOWNSCALING

GRAND CHALLENGES ON: CLOUDS, CIRCULATION AND CLIMATE SENSITIVITY; MELTING ICE AND GLOBAL CONSEQUENCES; WEATHER AND CLIMATE EXTREMES; REGIONAL SEA-LEVEL CHANGE AND COASTAL IMPACTS; WATER FOR THE FOOD BASKETS OF THE WORLD; AND CARBON FEEDBACKS IN THE CLIMATE SYSTEM

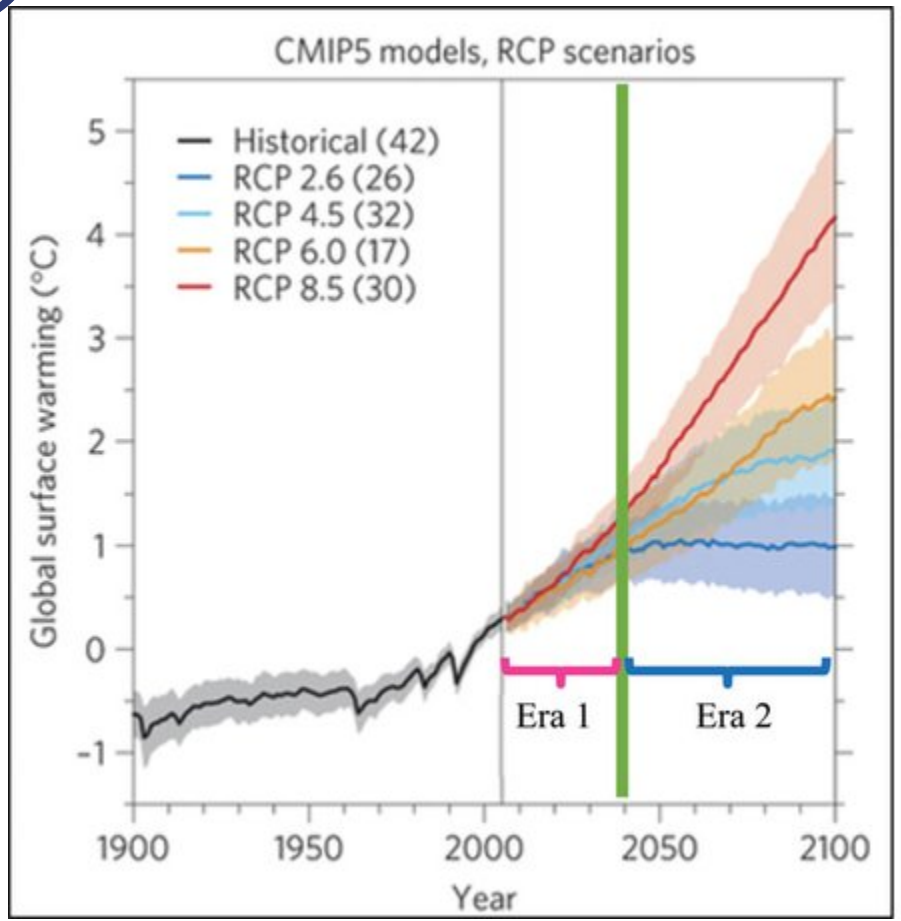
Climate Models

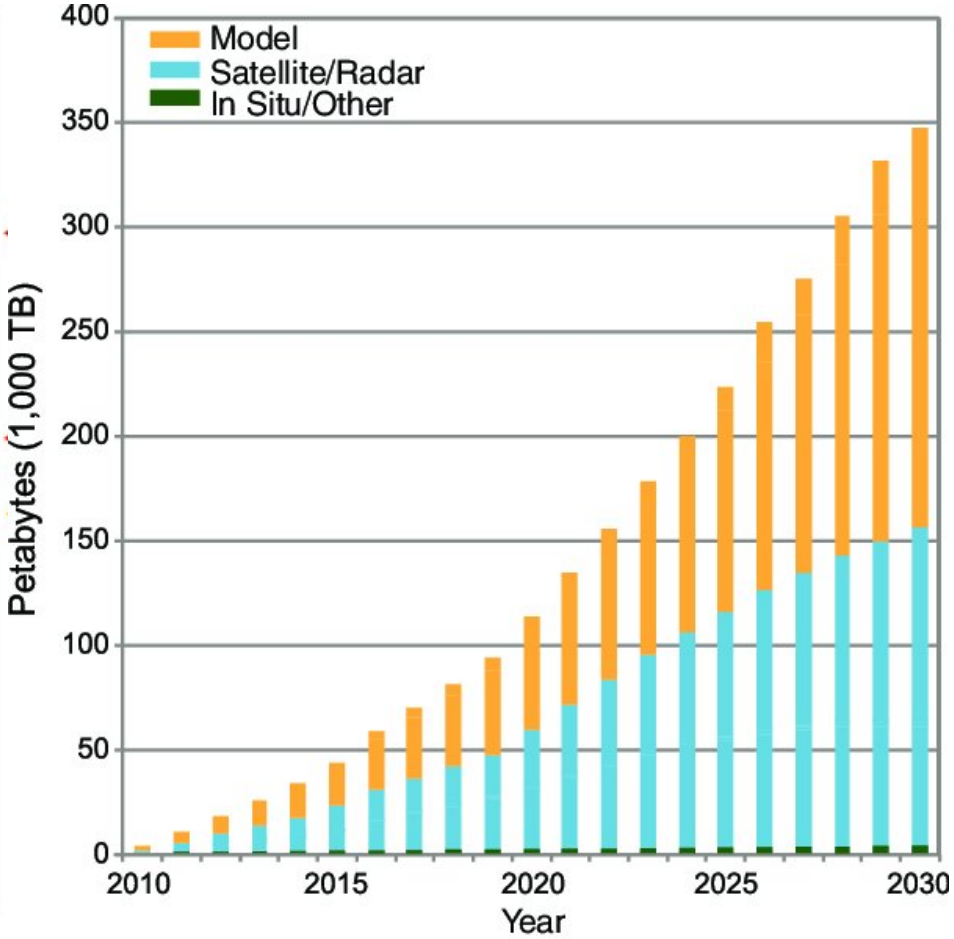
- Numerical models are fundamental tools in climate research:
  - They range from very detailed models of a particular process, through complex global and regional climate or earth system models, to more idealized climate system models.
  - They are used in a wide variety of applications including studies of processes, data assimilation and analysis, attribution, historical and paleo-

Climate Model Resolution



Data Volumes





ESGF

Geosci. Model Dev., 11, 3659–3680, 2018
https://doi.org/10.5194/gmd-11-3659-2018

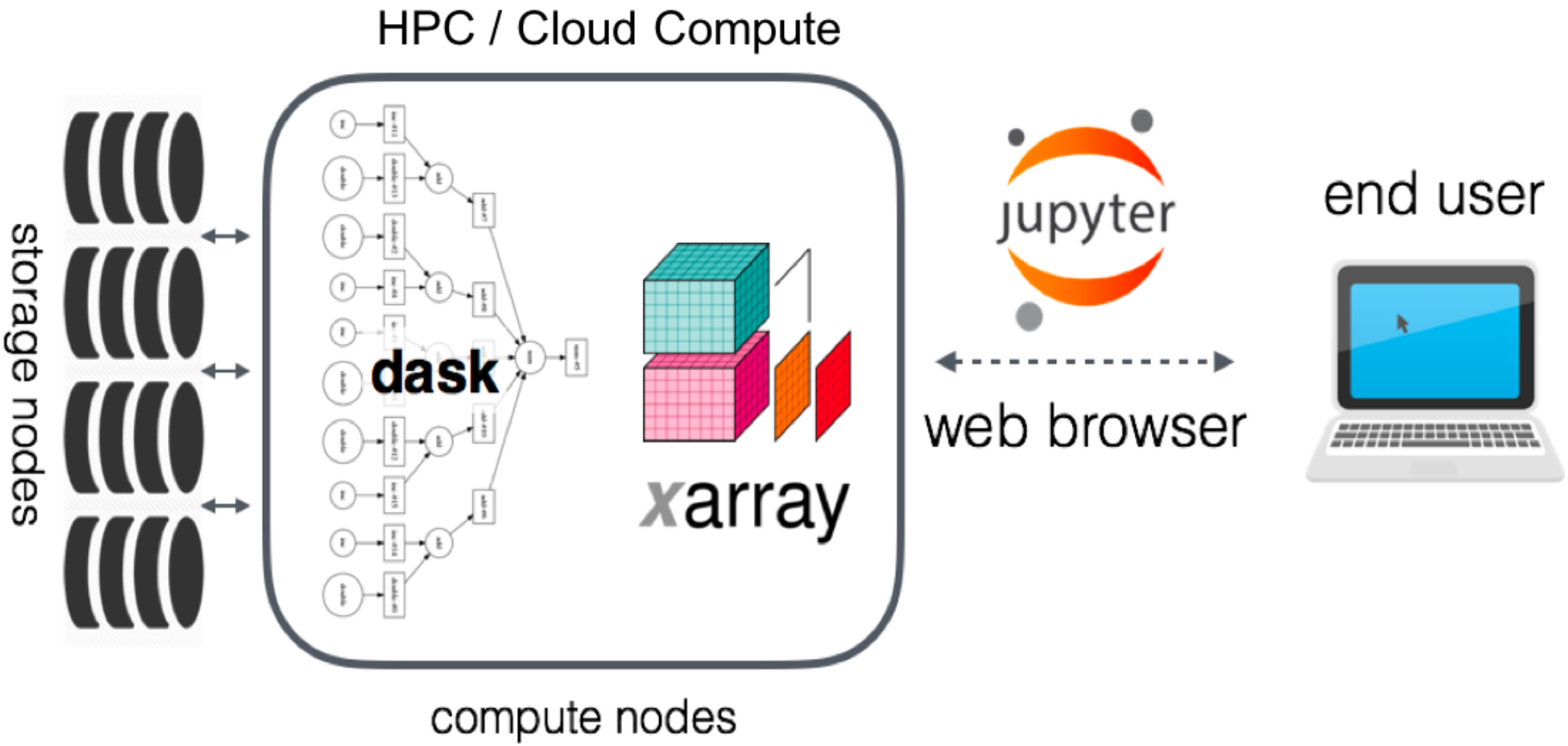
Geoscientific Model Development
Open Access
EGU

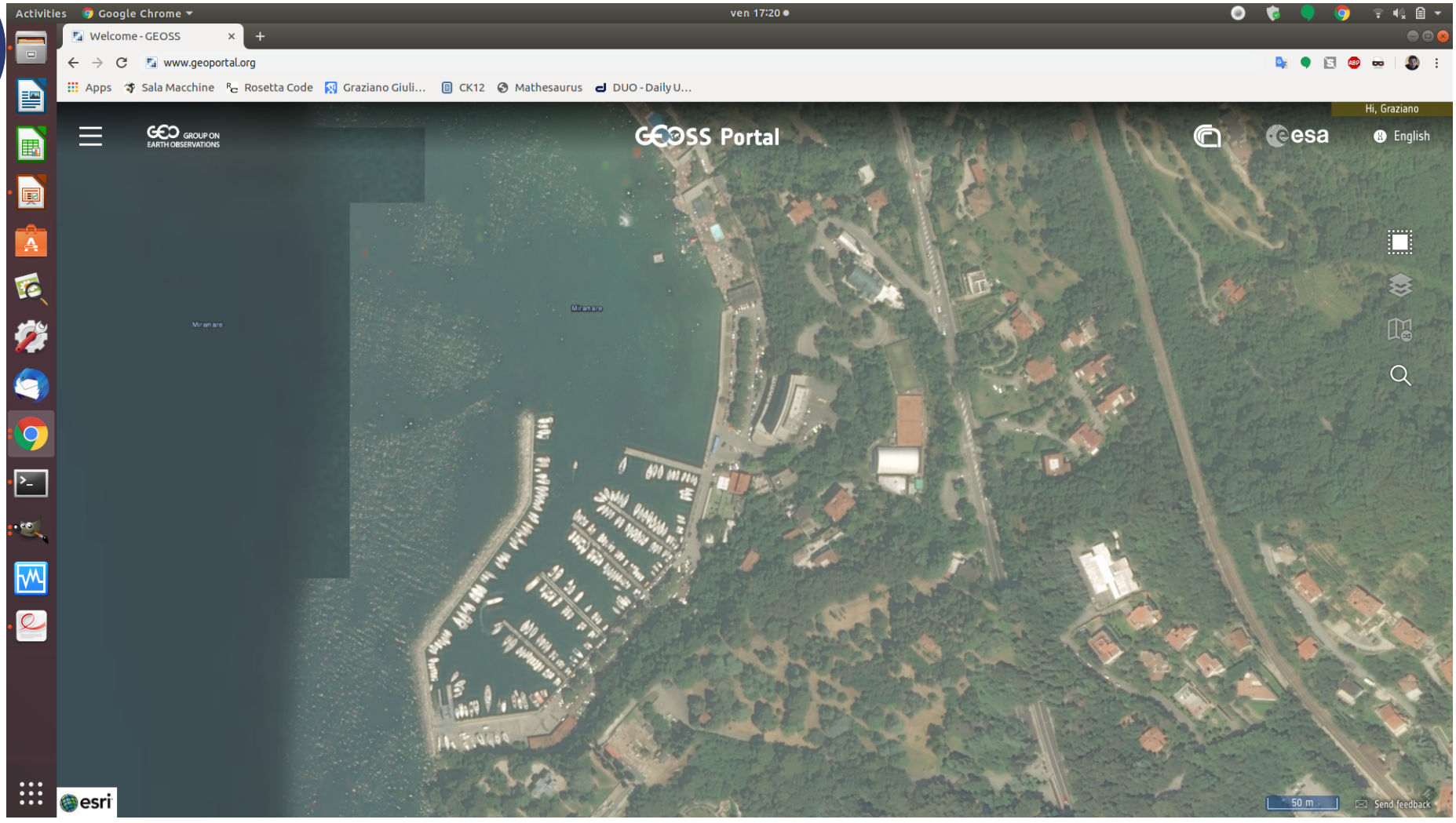
Requirements for a global data infrastructure in support of CMIP6

Venkatramani Balaji[1,2], Karl E. Taylor[3], Martin Juckes[4], Bryan N. Lawrence[5,4], Paul J. Durack[3], Michael Lautenschlager[6], Chris Blanton[7,2], Luca Cinquini[8], Sébastien Denvil[9], Mark Elkington[10], Francesca Guglielmo[9], Eric Guilyardi[9,4], David Hassell[4], Slava Kharin[11], Stefan Kindermann[6], Sergey Nikonov[1,2], Aparna Radhakrishnan[7,2], Martina Stockhause[6], Tobias Weigel[6], and Dean Williams[3]

[1]Princeton University, Cooperative Institute of Climate Science, Princeton, NJ 08540, USA
[2]NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540, USA
[3]PCMDI, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
[4]Science and Technology Facilities Council, Abingdon, UK
[5]National Centre for Atmospheric Science, University of Reading, Reading, UK
[6]Deutsches KlimaRechenZentrum GmbH, Hamburg, Germany

Pangeo

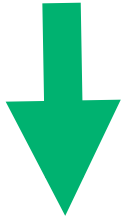

HPC / Cloud Compute
storage nodes
dask
xarray
compute nodes
jupyter
web browser
end user


Activities
Google Chrome
ven 17:20
Welcome - GEOSS
www.geoportal.org
Apps
Sala Macchine
Rosetta Code
Graziano Giuli...
CK12
Mathesaurus
DUO - Daily U...
GEO GROUP ON EARTH OBSERVATIONS
GEOSS Portal
esa
English
Hi, Graziano
Miramare
esri
50 m
Send feedback

Not Quite...

Data Analysis Workflow

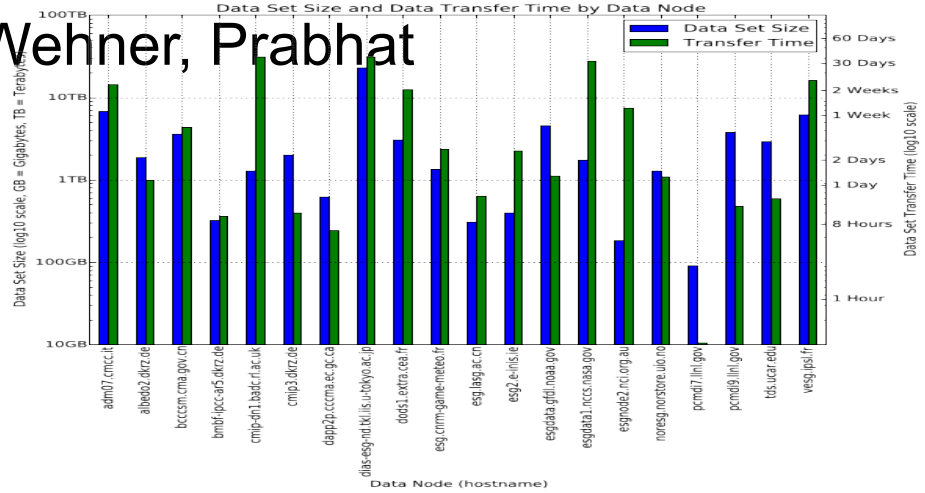
- Data Collection (STAGING)
- Data Pre-Processing (ADAPTATION)
- Scientific work (PROCESSING)
- Result check (VERIFICATION)



- Publication and peer review process

Timings

- An Assessment of Data Transfer Performance for Large--Scale Climate Data Analysis and Recommendations for the Data Infrastructure for CMIP6 - Dart, Wehner, Prabhat



- STAGING – 3 months
- ADAPTATION – 3 weeks
- PROCESSING – 2 days
- VERIFICATION – 10 minutes

CORDEX-CORE

CORDEX Coordinated Output for Regional Evaluations (CORE)

A simulation framework in support of IPCC AR6
CORDEX Scientific Advisory Team

This document presents the simulation framework for the CORDEX-CORE program. As discussed at the ICRC-CORDEX 2016 conference and in subsequent meetings, the main goal of the CORDEX-CORE framework is to provide a core set of comprehensive and homogeneous projections across all, or most, CORDEX domains. The immediate intent is to provide output that can support IPCC AR6 assessments. For this reason, the core framework is intended to be ambitious but sufficiently cost-effective to attract participation and produce timely output for the AR6. This program also stems from the results of a questionnaire send out to the CORDEX community through the CORDEX POCs in the fall of 2016, with feedback received in late 2016.

CORE has a specific focus on the IPCC AR6. It is not intended by itself to be the next phase of CORDEX. However, CORE simulations can provide a foundation in each region for further downscaling activities.

http://www.cordex.org/experiment-guidelines/cordex-core/

CORDEX-CORE

Coordinated Regional Climate Downscaling Experiment

- The primary target is grid spacing in the range of 12.5 -25 km
- The minimum simulation period is 1970-2100
- The GCMs to be downscaled should:
  - include both the RCP8.5 and RCP2.6 scenarios,
  - cover the range of GCM climate model sensitivity (essentially low, medium, high sensitivity),
  - have reasonably good quality of historical climate simulations in the regions for which they supply boundary conditions data
  - have good quality of historical climate simulations for important large-scale features such as ENSO, NAO
  - The format of output data should follow the standard CORDEX protocol.

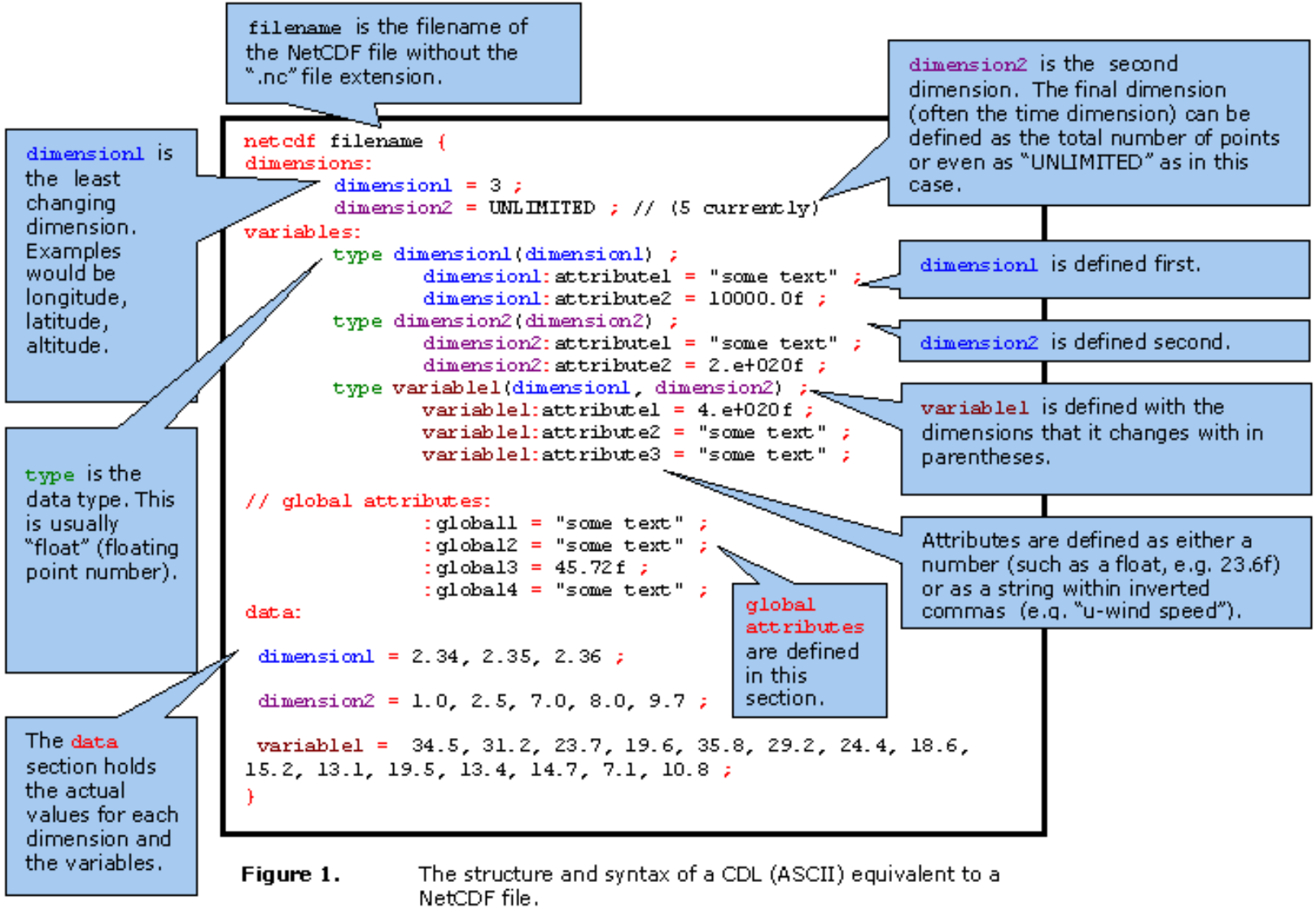
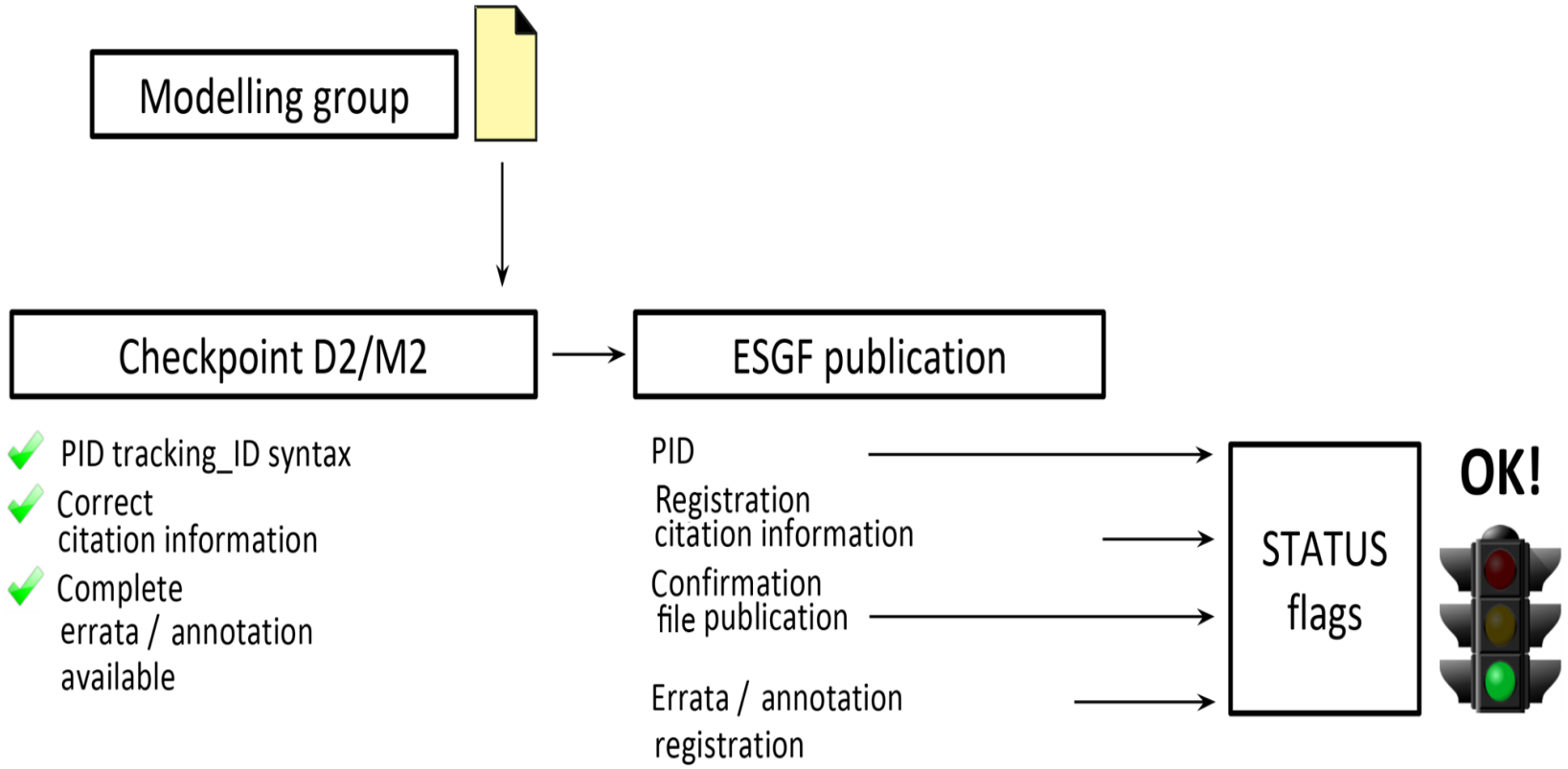


Figure 1. The structure and syntax of a CDL (ASCII) equivalent to a NetCDF file.

CORDEX DRS

- `tas_AFR-22_MPI-M-MPI-ESM-MR_rcp26_r1i1p1_ICTP-RegCM4-7_v0_day_20320801-20320901.nc`


Modelling group
Checkpoint D2/M2
ESGF publication
PID tracking_ID syntax
Correct citation information
Complete errata / annotation available
PID
Registration citation information
Confirmation file publication
Errata / annotation registration
STATUS flags
OK!

Output variables

MANDATORY variables

output variable name	units	frq [1/day]	ag
tas	K	8	i
tasmax	K	1	i
tasmin	K	1	i
pr	kg m-2 s-1	24	a
ps	Pa	8	i
hurs	%	8	i
sfcWind	m s-1	8	i
rsds	W m-2	8	a

frq: frequency [samples per day]
ag: aggregation

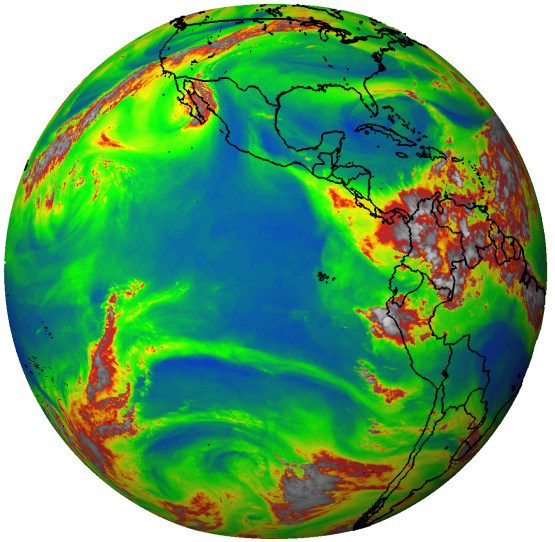
a: averaged over output interval (in model)
i: instantaneous
c: cumulative over sampling period
number: minimum samples per day if not averaged o

Highly Recommended Variables

output variable name	units	frq [1/day]	ag
sfcWindmax	m s-1	1	i
rlds	W m-2	8	a
hfls	W m-2	8	a
hfss	W m-2	8	a
rsus	W m-2	8	a
rlus	W m-2	8	a
evspsbl	kg m-2 s-1	8	a
mrro	kg m-2 s-1	1	a
mrso	kg m-2	4	i
snw	kg m-2	4	i
prc	kg m-2 s-1	1	a
ua100m	m s-1	8	i
va100m	m s-1	8	i

RegCM 4.7-CORDEX

- ICTP Regional Climate Model
  - Physical downscaling tool
    - Multiple GCM input
    - Hydrostatic 25 km atmosphere
    - CLM 4.5 surface Land Model
    - Multiple physical parametrizations



Boundary Conditions

- 3 GCM
  - NorESM2
  - HadGEM2
  - MPI-ESM-MR
- 2 Scenario
  - RCP2.6
  - RCP8.5

ICTP in CORDEX CORE

- CINECA Marconi Tier0
- Runs:
  - AFR-22 (scenario)
  - AUS-22 (OK)
  - CAM-22 (OK)
  - EUR-11 (OK)
  - SAM-22 (OK)

External to ICTP

- LLNL
  - WAS-22 (OK)
- HK
  - SEA-22 (scenario)
- China
  - EAS-22 (RegCM 4.6)
- NCAR

Data @ ICTP

CORDEX phase 1 (50-25 km, previous dataset)

- `/home/netapp-clima/users2/ggiulian/CORDEX/output`

CORDEX-CORE (25-11 km, orginal RegCM data format, monthly means, some daily)

- `/home/clima-archive4/CORDEX2/monthly_original`

Data @ ESGF

- As soon as QA/QC is passed
- Cineca ESGF data node
- ESG will index the data which will be available for
- download through the standard ESGF interface:
  - https://esg-dn1.nsc.liu.se/search/cordex/
  - https://XXXXXXXXXXXX/search/cordex

Data Availability Policy

- Embargo on derived articles until first article is published with defined author list (ask Erika)

Access to Marconi data

- Each of the CORDEX domain runs has one reference person
  - Ask Fabio or Adriano

Future of data

- “In the future, datasets and software with provenance information will be first-class entities of scientific publication, alongside the traditional peer-reviewed article […] Data analytics at large scale is increasingly moving toward machine learning and other directly data-driven methods of analysis, which will also be dependent on data with provenance tracking”

Questions ?