Copernicus and H2020 Programme by the ENES Data Task Force.

S.Denvil\textsuperscript{1}, M.Lautenshlager\textsuperscript{2}, S.Fiore\textsuperscript{4}, F.Guglielmo\textsuperscript{1}, S.Joussaume\textsuperscript{1}, M.Juckes\textsuperscript{2}, S.Kinderman\textsuperscript{2}, M.Kolax\textsuperscript{7}, C.Pagé\textsuperscript{3}, W.Som de Cerff\textsuperscript{6}.

\textsuperscript{1}Institut Pierre Simon Laplace, IPSL, France
\textsuperscript{2}Centre for Environmental Data Analysis, CEDA, United Kingdom
\textsuperscript{3}Centre Européen de recherche et de formation avancée en calcul scientifique, CERFACS, France
\textsuperscript{4}Euro-Mediterranean Center on Climate Change Foundation, CMCC, Italy
\textsuperscript{5}Deutsches Klimarechenzentrum, DKRZ, Germany
\textsuperscript{6}The Royal Netherlands Meteorological Institute, KNMI, Netherland
\textsuperscript{7}Swedish Meteorological and Hydrological Institute, SMHI, Sweden
H2020 is the biggest EU research and innovation programme ever. Almost 80 billion € of funding is available over seven years (2014 to 2020) – in addition to the private and national public investment that this money will attract.

Within H2020 three types of activity are supported:

- The first activities are targeted to the development of new world-class research infrastructures. Support will be provided for the implementation and operation of the research infrastructures listed on the ESFRI Roadmap.

- The second set of activities aims at optimising the use of the national facilities by integrating them into networks and opening their doors to all European researchers. This is a continuity of the so-called Integrating Activities under FP7.

- The third action will support further deployment and development of ICT based e-infrastructures which are essential to enable access to distant resources, remote collaboration, and massive data processing in all scientific fields.
CMIP international coordinated experiments

Evaluate
Pattern correlation with observations

Understand
RCP 2.6
Change in average surface temperature (1986–2005 to 2081–2100)

Predict
RCP 8.5

IPCC AR5 (2013)

CMIP5
3400 simul. yrs up to > 12000 yrs
50 expts up to > 160 expts
2000 Tbytes (CMIP3: 36)

1 Canada
6 USA
7 Europe

1 Russia
4 Japan
5 China
1 Korea
2 Australia

27 modelling groups
58 models
IS-ENES (2009-2017): Infrastructure for ENES

Climate models
Environment software tools
ESM ca 1000 man years

High-performance computers & storage facilities

Data & Metadata
Distributed Database ESGF
Open Access

Climate research & Impact research
Climate services

Support WCRP international experiments
Used in IPCC Assessments Reports
**IS-ENES 2 activities**

Foster the integration of the European ESM community

*Foster interactions, synergies and common strategies*

**ENES Infrastructure Strategy :**

- Infrastructure for model evaluation → Need for routine evaluation
- Mid-term update 2017 → 5 + 2 new recommendations
- Long-term sustained European RI → Still to be done

**Strengthen governance:**

- ENES Scientific Officer → Important support
- ENES Organisation force → Key role of HPC and Data task
- Governance on common software → Clarified levels
- *International governance ESGF, WIP* → *Support from IS-ENES*

**Community building :**

- 2 Training schools on ESM → Worked well – but lot of work
- ENES portal → Important common basis
IS-ENES 2 activities

Facilitate the dissemination of ESM simulation results
Ease use of model results for climate research & for climate impact research

Service around model results:
• CMIP5 & CORDEX on ESGF → Essential role
• Service to providers (data nodes & users) → Essential role (eg. Help)

Metadata
• Upgrades & Interoperability, CMIP6 → Leading role, simplified

Develop more efficient tools for ESGF:
• Core services → Europe : half WGs
• Security issues & CMIP6 → Watch out
• Quality control, monitoring, synchro → leading role

Services for climate impacts - Climate4impact:
• Tools, downscaling, indices → Now a basis for other projects

Societal innovation:
• To corporate (KIC) → Mater classes but limited
• To climate services centres → link with Copernicus
Support CMIP6

### European models in CMIP6

<table>
<thead>
<tr>
<th>Country</th>
<th>name of model (CMIP5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consortium</td>
<td>EC-EARTHE</td>
</tr>
<tr>
<td>France</td>
<td>IPSLCM5</td>
</tr>
<tr>
<td>France</td>
<td>CNRM-Cerfacs</td>
</tr>
<tr>
<td>Germany</td>
<td>MPI-ESM</td>
</tr>
<tr>
<td>Italy</td>
<td>C-ESM</td>
</tr>
<tr>
<td>UK</td>
<td>HadGEM2</td>
</tr>
<tr>
<td>Norway</td>
<td>NorESM</td>
</tr>
</tbody>
</table>

#### Germany: AWI-CM EMAC

<table>
<thead>
<tr>
<th>Short name of MIP</th>
<th>AWI-CM</th>
<th>CMCC</th>
<th>CNRM</th>
<th>EC-Earth</th>
<th>EMAC</th>
<th>IPSL</th>
<th>MPI-ESM</th>
<th>Nor ESM</th>
<th>UK ESM</th>
<th>HadGEM3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AerChemMIP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C4MIP</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CFMIP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>DAMIP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DCPP</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FAFMIP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GeoMIP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>GMMIP</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>HighResMIP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ISMIP6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>LS3MIP</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>LUMIP</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OMIP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PMIP</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>RFMIP</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ScenarioMIP</td>
<td>0'</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>VolMIP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CORDEX</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DynVar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SIMIP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>VIACS AB</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: Veronika Eyring
Multiple type of storage & data interaction

- Earth System Model Simulation
- Final checkpoint
- Multiple Tools, Visualization
- (periodic) output physical variables
- post-processing (& analysis)

Distributed/Federated Archives (Servers/Public Clouds)

- Reformatting, Sub-setting, Downloading, Processing.
- ESGF: Earth System Grid Federation
- Cloud Computing
- Download

Multiple Roles, at least:
- Model Developer
- Model Tinkerer
- Expert Data Analyst
- Service Provider
- Data User
Management of output

**Data Request**

- **CMIP6 protocol**
  - MIPs, experiments, forcings,…
  - IPSL-CM LMDZOR

- **configurations**
  - IP SL - CM LMDZOR

- **ping files XML**
  - parameters, forcing files, restart

- **“file def” XML**

- **IPSLCM6.1-LR**
  - Monitoring of the simulation

- **XIOS2**
  - Simulation responsibles/“runners”
  - Model developers
  - Platform group

- **Output Files**
  - CMIP6 Publication

- **Settings**: experiment, year, MIPs, institution,…

- **« Home vars »**

- **Version 01.00.18**
Dedicated Analysis Facilities

Example: JASMIN – (IPSL and DKRZ operates as well).

- 5 PB of CURATED POSIX archive data (and growing, with 10 PB expected in next 2 years).
- 6 PB of USER POSIX data on disk (and growing)
- Lots of EPHEMERAL POSIX data (need space for analysis!)

Moving PB per day in and out of LOTUS (the batch cluster):

The importance of data gravity; when you have data, more data comes to you!
Storage issues and action arising

Issues:

• **Cost:** Disk prices not falling as fast as they used to.

• **Behaviour:** Larger groups sharing data for longer, which means *data is “hot” for longer.*

• **Performance:** Traditional (POSIX) disk not performant at scale.

• **Software:** *Little software for our domain which can exploit “OBJECT store” disk (hard to use the public cloud).*

• **Tape:** Tape remains important, particularly for *large amounts of “cold” data.*

Community Action: **ESIWACE “Exploitability”** work package:

1. Better understanding of costs and performance of existing and near-term storage technologies.

2. New **“Earth System Middleware”** prototype
   Provides an interface between the commonly used HDF library and storage which addresses both the performance of POSIX and the usability of object stores.

3 New **“Semantic Storage Library”** prototype:
   Python library that uses a “weather/climate” abstraction (CF--NetCDF data model) to allow one “file” to be stored across tiers of, e.g. POSIX disk, OBJECT store, and TAPE.
Cloud computing & Big Data

Three domains of interest to weather and climate community:

• New Fabric and infrastructure (private/public cloud)
  – Exploiting virtualisation to provide flexible and elastic services. Not suitable for large scale simulations, but big role to play in analysis facilities.
  – Large scale use will depend on addressing usability of object stores.

• New compute paradigms emerging (in our community)
  – New ways of arranging data and scheduling compute across hardware (e.g. HADOOP, SPARK) – not used
  – Some small scale experiments reported in the literature. DASK experiments underway at the UK Met Office (http://www.informa6 cslab.co.uk/)

• New ways of exploiting algorithms emerging (in our community).
  – e.g. using machine learning to identifying patterns in data, something we’ve done for decades, but with new and (possibly) better tools.
  – Experiments comparing traditional methods to new methods are underway (e.g. at LLNL in the US) to evaluate potential.
  – Possible use for Quality Control of data (e.g. unusual field) or Parameterisations (e.g. optimal parameters)
Copernicus Climate Change Service (C3S)

Beta version http://climate.copernicus.eu

- Climate Data Store
  - ECVs past, present and future
  - Observed, reanalysed and simulated
  - Derived climate indicators

- Sectoral Information System
  - Monitors quality of C3S products and services
  - Ensures C3S delivers state-of-the-art climate information to end-users

- Evaluation and Quality Control

- Outreach and Dissemination
  - Web content
  - Public outreach
What simulations will be available from the CDS?

★ **Global projections** [from CMIP-5 Core and Tier-1 simulations]
★ Pre-industrial control with prescribed, non-evolving concentrations of atmospheric gases and aerosols;
★ Historical ensemble, 1850 to at least 2005, imposed changing concentrations and forcings, minimum of 3-member ensemble [Tier-1];
★ AMIP ensemble, 1979 to at least 2008, prescribed SST and sea-ice concentration, other forcings as in Historical ensemble above, minimum of 3-member ensemble [Tier-1];
★ Projections following RCP 4.5 and 8.5 concentration scenarios, years 2006-2100, **preferably from models with multi-member ensembles**
★ Optionally: Projection following RCP 2.6 and 6.0 emission scenario, years 2006-2100, **preferably from models with a 3-member ensemble**

★ **Regional projections:**
★ Existing simulations from the Euro-CORDEX and Med-CORDEX projects
★ New CORDEX simulations for a pan-European domain based on an agreed “3-D matrix” of regional climate models, boundary conditions from global models, concentration scenarios (RCPs)
Climate projections contract

Global projection service (C3S 34a)

★ Solution to access and manipulate global climate projection data from the CMIP archive, consistent with the requirements of climate services

★ metrics for fidelity of models in simulating historical climate, to be translated into quality for specific applications
★ interactive tools for generic products (e.g. maps of intra-ensemble variability for different models and scenarios), and tailored products for several economic sectors

★ Studies on how well climate projections address sectoral needs, to guide requirements for the operational phase of C3S. Areas of interest: the benefit of ensemble size versus resolution for global models, and the benefit of initialised decadal predictions, in relation to the specific needs of different economic sectors
Regional projection service (C3S 34b)

**Lot 1:** CORDEX for the Copernicus Data Store (CORDEX4CDS; lead contractor: CNRS (France), start 1 May 2017, end Apr 2021):

- facilitate access and manipulation (via the CDS) of output of regional climate projections over Europe and boundary conditions from GCM simulations needed for future regional projections.

**Lot 2:** Producing regional climate projections leading to European services (PRINCIPLES; lead contractor: SMHI (Sweden), start 1 May 2017, end Apr 2021)

- define and complete a matrix of global/regional model combinations and scenarios, which allows robust assessment of the uncertainties arising from these factors in a multi-model set of regional projections.

Evaluation and quality control for climate projection services

**C3S 51 Lot 4:** Data evaluation for climate models (DECM: lead contractor: FMI, start 1 Aug. 2016, end Oct. 2018).

- Conduct survey of user requirements, evaluate user feedback about services provided by 34a and 34b contracts, provide a gap analysis.
Who are they talking to?
High Level Architecture
What do users gain from C3S?

- Improved reliability in the access to climate projection data through the Climate Data Store
- Products computed from models which show good fidelity in the simulation of climate during the recent decades (as quantified by appropriate metrics)
- Improved estimates of uncertainties allowed by focusing on models that provide ensemble simulations of individual scenarios
- User defined indices and products tailored to specific application sectors
- Quality and usability of products tested by an Evaluation and Quality Control consortium