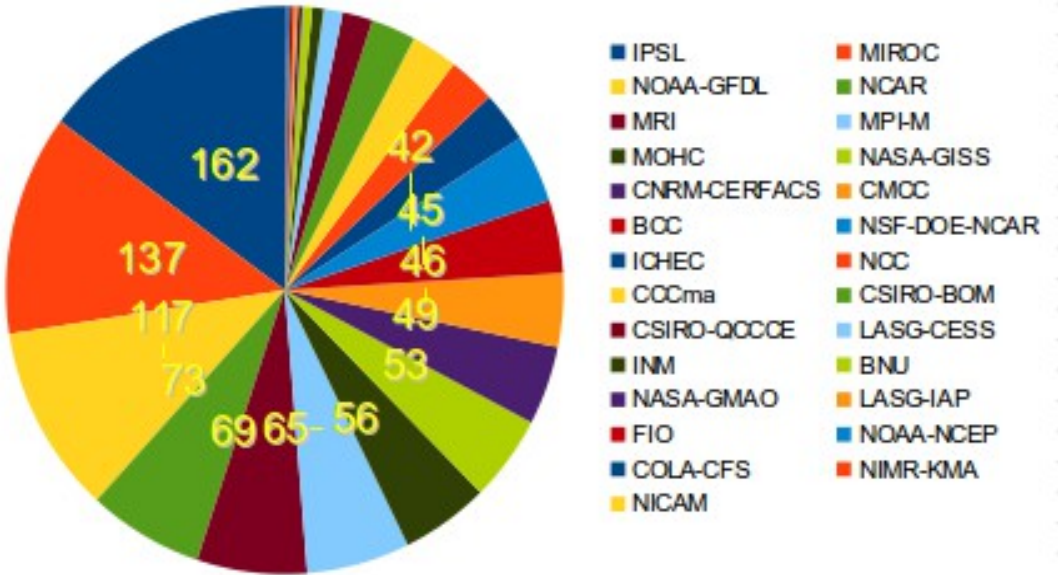
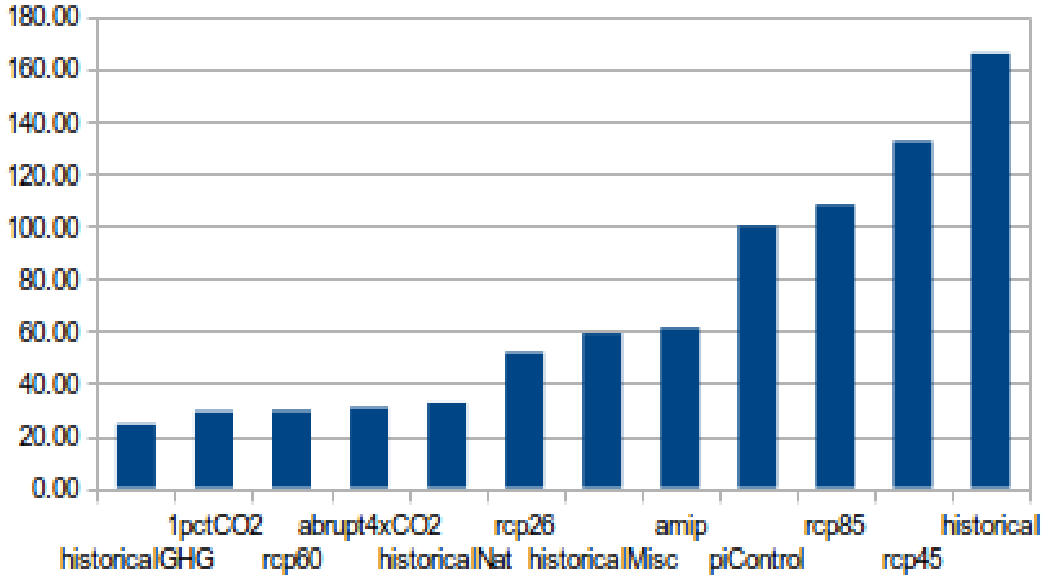
The CMIP6 Data Request

Martin Juckes

October 2015

CMIP5





Volume at CEDA: 1Pb

27 Institutions contributing data;
100 Experiments

Major input to AR5

CMIP3: 30Tb
CMIP5: 1.5Pb
CMIP6: ~$2x10^{16\pm0.8}$ Bytes

Variable Lists
- Definition of physical quantities;
- Output specifications.

Output Requirements
- Experiments and time slices
- Objectives supported
- Priority of variable

Experiment Specifications
- Duration of simulation
- Tier
- Number of ensemble members

Recommendations for output and analysis

CMIP6 Data Request

What is the data request?

The data request is the composite of the **endorsed MIP** requests.

(1) "DECK" is not an endorsed MIP …

(2) *The request from each MIP covers the data that they need from the experiments they define, from the DECK + CMIP6 historical, and from experiments defined by other MIPs where it is needed for the analysis they propose.*

Data Request Target Users

Infrastructure provider (manager)

Infrastructure provider (technician)

Data provider (programme manager)

Data provider (technician)

Data provider (scientist)

Data user (outside modelling centres)

Data user (in modelling centres)

The Software Architecture Components

Consolidated Request
The request as a structured document.

Programming Interface
A python library, facilitating use of the request by software.

Command Line
For flexible access

Web Access (coming soon)
To support a web interface

Greater flexibility

Modelling groups will select

- priority of variables (as for CMIP5);
- which MIPs to support;
- objectives within MIPs;
- Tiers of experiments;
- MIPs can specify groups of variables; variables may be given different priorities for different experiments.

Variable Lists

CMIP5

"standard output" spreadsheet: list of variables organised into MIP tables, 1098 CMOR variables, 536 standard names

CMIP6

~800 standard names

~1000 MIP Variable

~2000 CMOR Variables

~3500 Request Variables

A "MIP Variable" can be re-used at different frequencies, or with different masking options.

A CMOR variable can be re-used at different priorities or in different groups.

Variable Lists

Different MIPs may request the same variable with different priorities

- MIP
- Priority

Request Variable

e.g. tas, Amon, PMIP

- short name
- standard name
- long name
- description
- units

MIP Variable

e.g. tas

CMOR Variable

- CMOR name
- description
- frequency
- MIP table
- realm
- ….

e.g. tas, Amon

A variable may occur in many different guises

Structure

- dimensions
- cell_methods
- cell_measures
- flag_meanings
- flag_values
- ……

Managing duplicate requests

There is a single consolidated list of CMOR variables.

All endorsed MIP requests refer to the same set of variables.

The python API will provide a list of required variables for any combination of MIPs

Use in python code:

```
from dreqPy import scope
sc = scope.dreqQuery()
v1 = sc.volByMip( 'C4MIP', pmax=2 )
v2 = sc.volByMip( {'C4MIP', 'LUMIP'}, pmax=2 )
```

From the LINUX command line:

```
>drq -m C4MIP,LUMIP -p 1 -t 1

    4.20Tb
```

Variable choices

Requests conditional on model configuration.

Some variables are only needed for specific configurations/types of models. E.g.

- time varying ice sheet state …. only needed for models with dynamic ice sheets;
- pressure on model levels …. not needed for models with pressure-based vertical coordinates;

Ranked variables

Where there is clear redundancy between variables, e.g. air temperature on 7 pressure levels at 6 hourly intervals vs. the same variable on the same levels at 3 hourly intervals, these can be given a rank and the API will only select the highest ranked variable.

A modelling group supporting the HighResMIP request for priority 1 variables should provide the 6 hourly data, but another group going up to priority 2 will only supply the 3 hourly data.

Problems …

Preferred grid

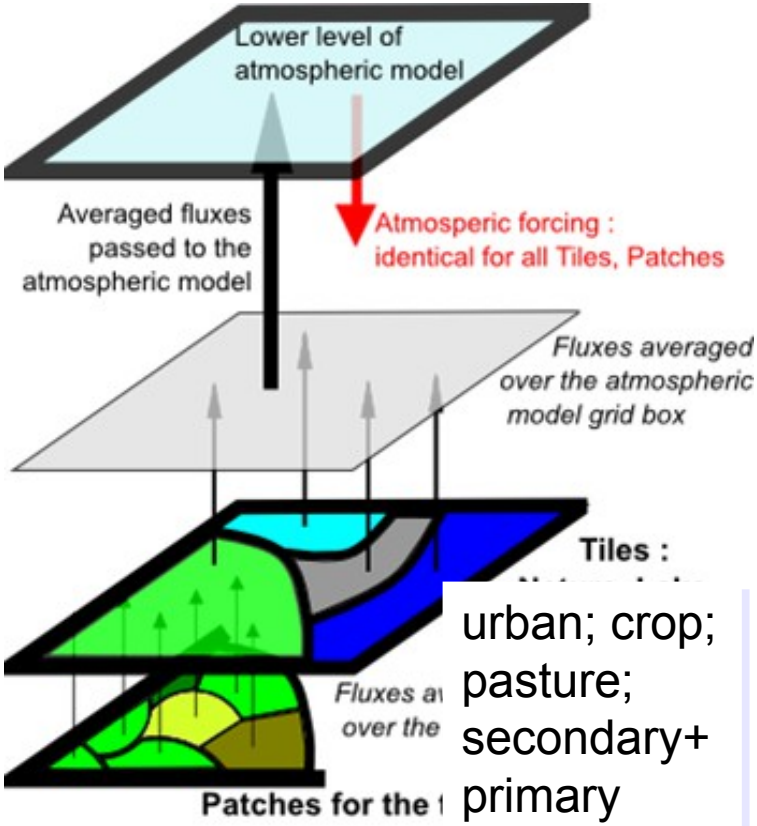
If DCPP asks for data on a 100km grid and FAFMIP asks for the same data on the native grid, should both be considered as required?

Gaps in the request

Only one MIP has requested “areacella” (the area of atmospheric grid cells). Should some fields be specified as required metadata? Are there other gaps?

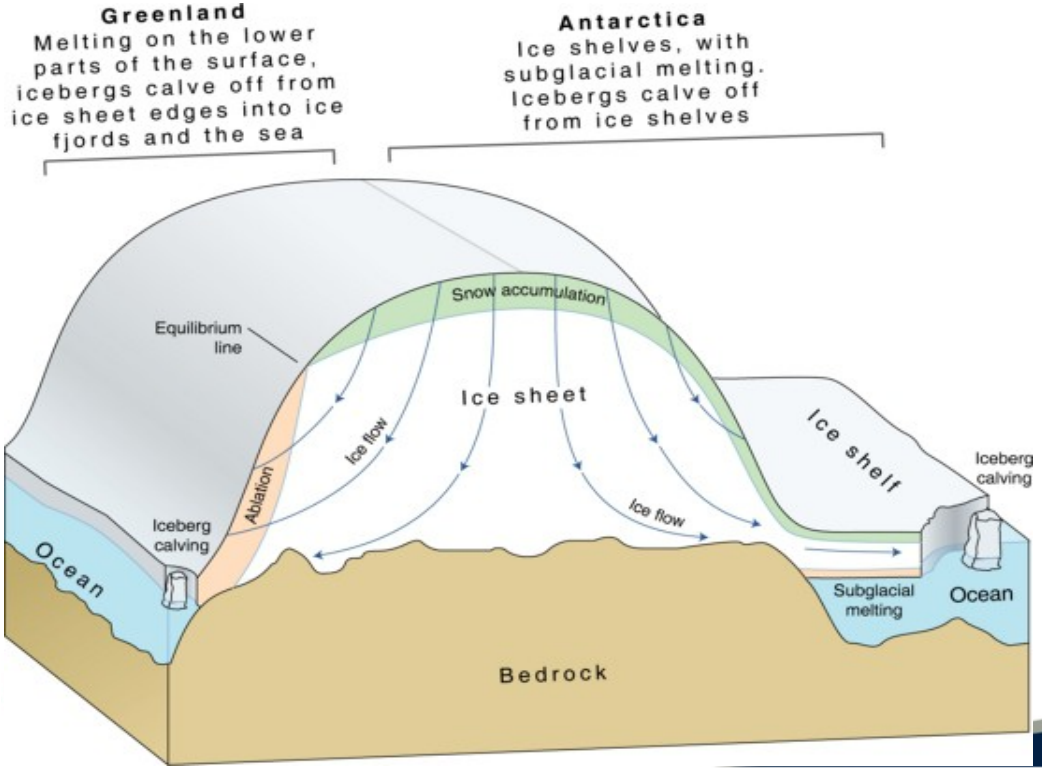
CMIP6 Novelties

LUMIP is requesting output on four Land Use tiles which partially align with C4MIP vegetation fractions.



urban; crop; pasture; secondary+ primary

Adapted from MeteoFrance
http://www.cnrm.meteo.fr/surfex/spip.php?rubrique8



New area types and boundaries:

- surface under ice sheet
- grounding line

Resources

*XML request document and documentation;

*Python library and documentation;

*Repository of document versions;

*Persistent identifiers (e.g. w3id.org/cmip6dr/variable/tas);

*Data request handbook (in preparation);

*Additional views of the request (excel, html, …);

* forum: dreq01.vanillaforums.com

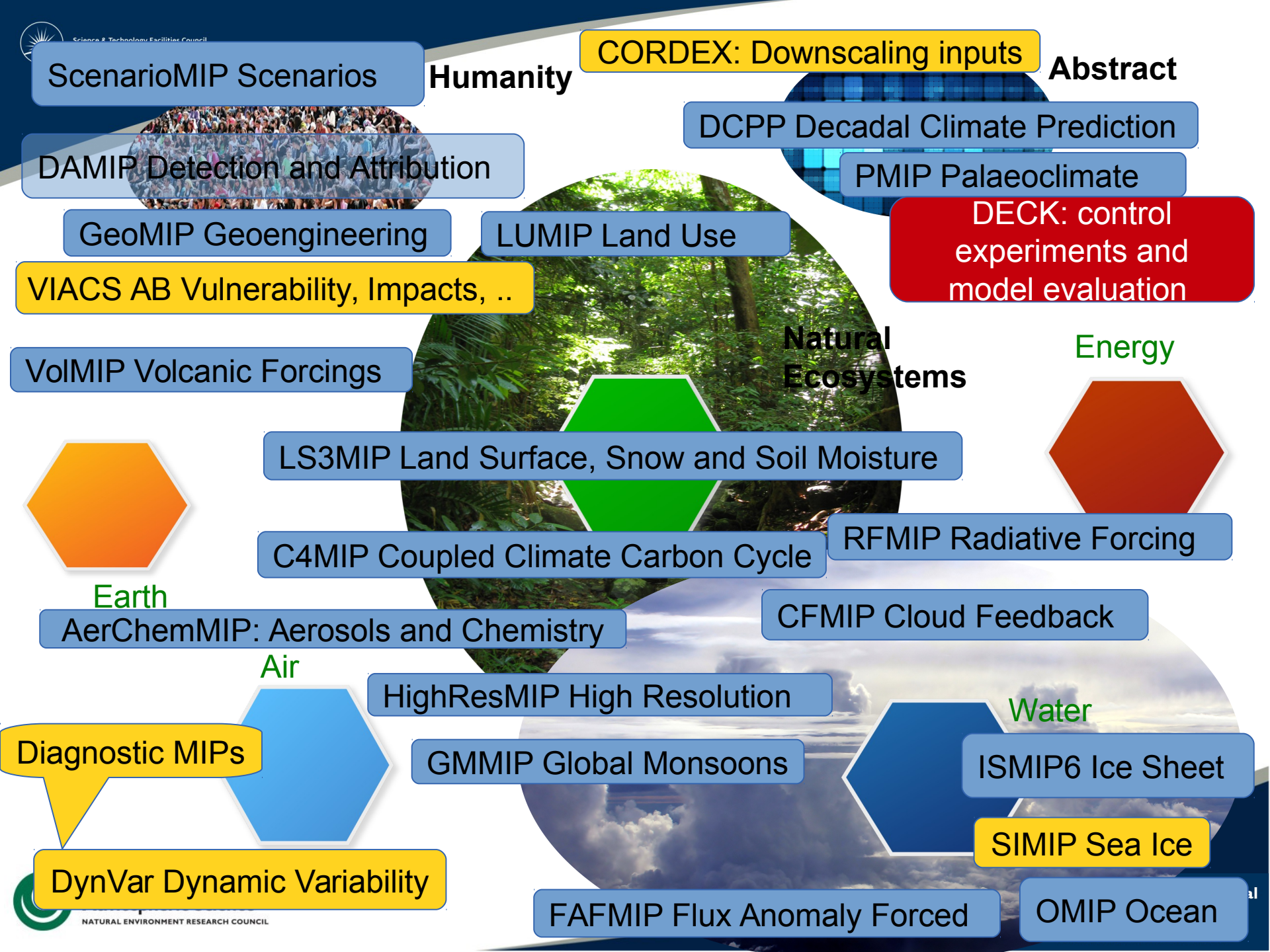
- The end …

Output Request

- Objective
- Time slice
- Experiment group

- MIP
- Priority

e.g. Amon, PMIP, amip


ScenarioMIP Scenarios
Humanity
CORDEX: Downscaling inputs
Abstract
DCPP Decadal Climate Prediction
DAMIP Detection and Attribution
PMIP Palaeoclimate
GeoMIP Geoengineering
LUMIP Land Use
DECK: control experiments and model evaluation
VIACS AB Vulnerability, Impacts, ..
Natural Ecosystems
Energy
VolMIP Volcanic Forcings
LS3MIP Land Surface, Snow and Soil Moisture
RFMIP Radiative Forcing
C4MIP Coupled Climate Carbon Cycle
Earth
CFMIP Cloud Feedback
AerChemMIP: Aerosols and Chemistry
Air
HighResMIP High Resolution
Water
Diagnostic MIPs
GMMIP Global Monsoons
ISMIP6 Ice Sheet
SIMIP Sea Ice
DynVar Dynamic Variability
FAFMIP Flux Anomaly Forced
OMIP Ocean
NATURAL ENVIRONMENT RESEARCH COUNCIL

Organising data

for the Coupled Model Inter-comparison Project (CMIP)

CMIP Panel

WGCM Infrastructure Panel

~20 (proposed) endorsed Model Inter-comparison Projects dealing with different science areas. Science objectives and experimental outlines distributed for open review Dec. 2014

Data requirements for each MIP

Consolidated data request: > 1000 variables, various frequencies, grids, masks and tiles.

~30 modelling centres will participate, each supporting one or more of the endorsed MIPs. All centres should complete the “DECK” experiments: historical, control, steady and abrupt CO2 increase.