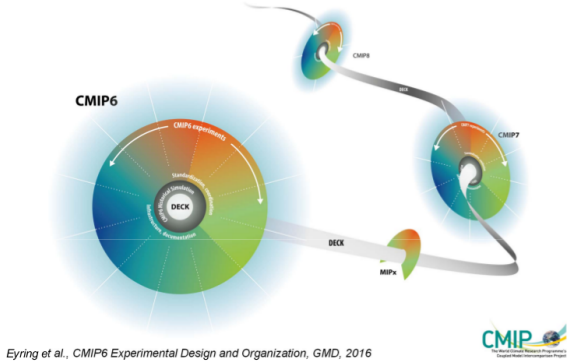


Eyring et al., CMIP6 Experimental Design and Organization, GMD, 2016

Agence Nationale de la Recherche ANR - CONVERGENCE MODÈLES NUMÉRIQUES

METEO FRANCE

Directly driving data and metadata generation by CMIP6 Data Request content thanks to XIOS

S. Sénési, (CNRM, Meteo-France)
Y. Meurdesoif, A. Caubel, S. Denvil (IPSL)
M.-P. Moine (Cerfacs)

Joint final IS-ENES2 workshop on Workflow Solutions in Earth System Modelling and Meta-Data Generation during Experiments - Lisbon, 26-29/09/2016

Motivations

The CMIP6 Data Request (DRQ) :

- **CMIP6 is *pharaonic***: 28 MIPs, 228 experiments, 2280 CMOR variables, 49 tables,…
- DRQ gathers (**heterogeneous**) requirements **from all MIPs**
- **High variability in the DRQ**: from one experiment to the other, from one simulated year to the next one, from a modelling group to an other depending on the MIPs it is engaged in,...

CMIP6 data format & Controlled Vocabulary :

- Standards naming convention for variables (CF), directories and files (DRS)
- Mandatory attributes : a lot of metadata on files and variables (CMIP6 CVs)

Constraints:

Modelling groups have to:

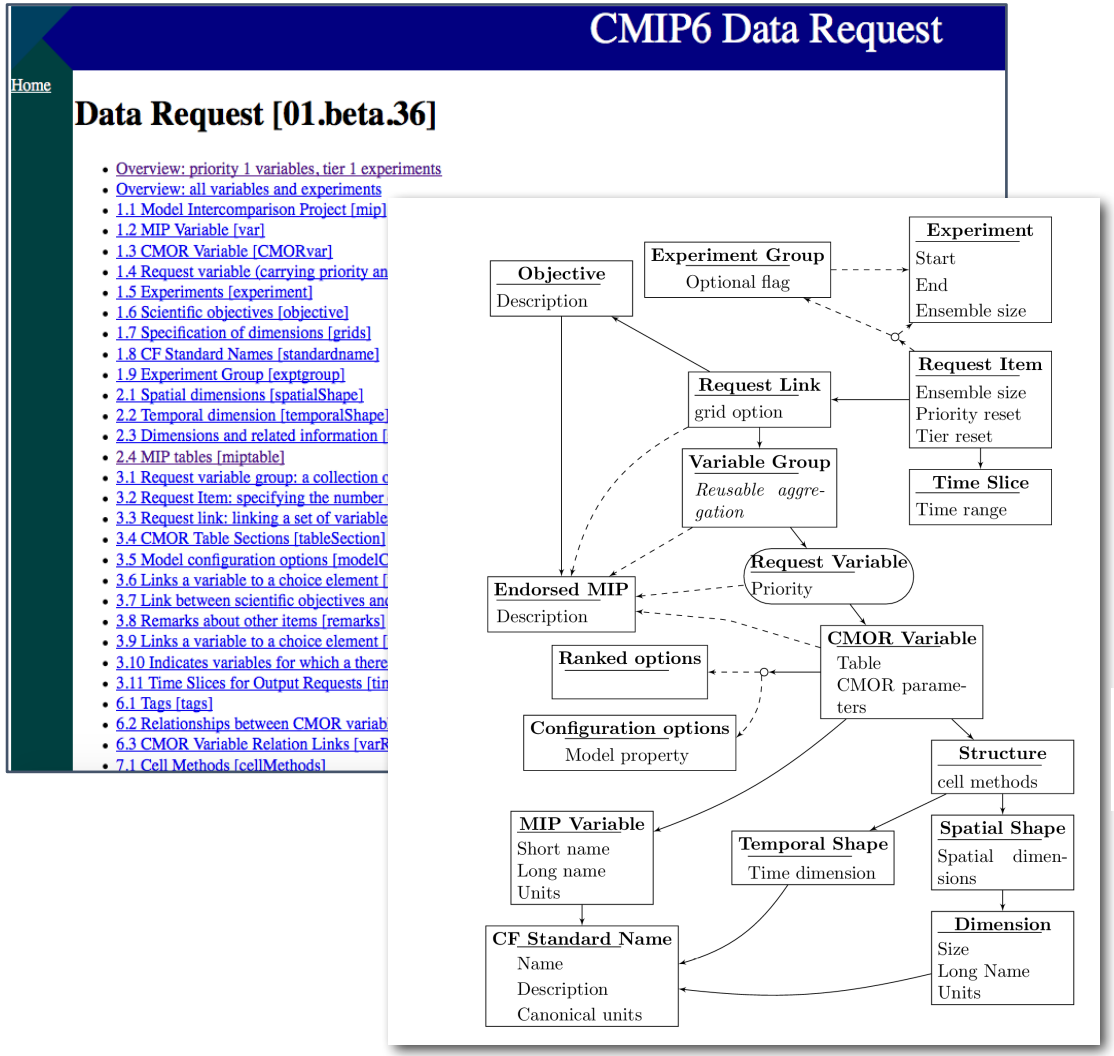
- Configure their model outputs to conform (*at best*) to the DRQ
- Post-process their data to comply with the CMIP6 format (CMORisation) : *change format, names, units, add metadata, compute derived diagnosis…*

Challenges:

- Such a model's output configuration is nearly intractable manually
- Raises some workflow configuration problems *(e.g. how to tell your model to output a variable only over a given time slice ?)*
- Post-processing is experienced as a *critical* step *(additional pipe in the production workflow to monitor, often with limited computing efficiency, generating I/O traffic)*

dr2xml, a tool to automate the configuration of XIOS-enabled models

Data Request python API + XML files
(Martin Juckes)



XIOS-enable model
(e.g. NEMO)

iodef.xml

```
…
<field_definition src="field_def.xml"/>
<file_definition src="file_def.xml"/>
…
```

dr2xml

field_def.xml

```
<field_definition>
 <field id="sst" />
 …
</field definition>
```

Model source code

```
…
CALL xios_send_field("sst", sst)
…
```

```
<file_definition>
 <file id="grid_T" name="grid_T" >
  <field field_ref="sst" name="sosstsst" />
  …
 </file>
 …
</file_definition>
```

file_def.xml

CMIP6-publication-ready data files

dr2xml, in practice

lab_and_model_settings:

settings related to a laboratory and a model

Global attributes

institution_id
model_id
model
source_type
references
contact …

Modeling group choices

mips
max_priority
tierMax
realms_per_context
excluded_vars

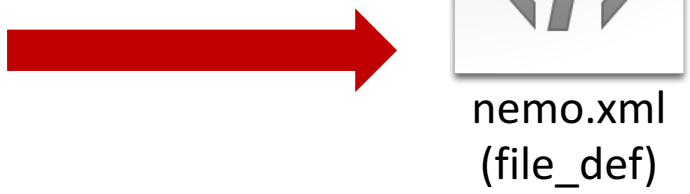
simulation_settings:

settings related to an experiment

Global attributes & choices

experiment_id
activity
realization_index
initialization_index
physics_index
forcing_index
parent_experiment_id …

```
from dreqPy import dreq # ------------------> Import dataRequest package
dq = dreq.loadDreq()
from dr2xml import generate_file_defs # ------> Import dr2xml
my_cvs_path='/home/user/CMIP6_CVs/' # --------> Path to CMIP6 CVs (jason files)
generate_file_defs(dq, lab_and_model_settings, simulation_settings,
                   year=2000, context='nemo, printout=True)
```



nemo.xml
(file_def)

dr2xml, features

1 Exploits the DRQ content and 'scoping' tools to dynamically:

- Identify the **list of relevant CMOR variables** *(given MIP(s), experiment(s), tier(s), output priority(ies) and simulated year prescribed by the user)*
- Collect **CMIP6 metadata** associated to experiment(s) and CMOR variables *(e.g. variable_id, standard_name, long_name, units, frequency, description,…)*
- Get **cell_method*** information *(e.g.: 'time mean over sea ice')*
- Get **spatial_shape*** information *(e.g.: 'global field 7 pressure levels'; 'ocean basin meridional section')*

2 Relies on XIOS2 properties and pluggable processing filters(*):

(*) *parallel and scalable*

- Allows writing **CF-compliant files**
- **File structure** is flexible: multi- or single-variable, splitting period
- **Can glue any attribute**, attached to the files or to the fields
- Can perform :
  - **basic arithmetic** operations *(useful for units conversion)*
  - **time** operations *(e.g., time averaging, min/max)*
  - **spatial** operations *(e.g. zonal mean)*
  - **grid remapping** *(horizontal and vertical)*
- Has **masking** functions

3 Handles exceptions to the DRQ:

Enable to specify a list of **excluded variables** *(that ARE requested by MIPs but the modelling group WILL NOT output because of volumes, N/A, …)*

4 Handles complement to the DRQ:

Enable to specify a list of **additional variables*** *(that ARE NOT requested by MIPs but the modelling group WANT TO output)*

**features upcoming in next version*

file_def.xml example

DRS compliant **file name**

File global attributes (CMIP6 CVs)

field_ref
adopted convention:
CMIP_<var_shortname>

var_shortname

field local attributes (CMIP6 CVs)

dr2xml_DEVEL.py × nemo.xml ×

```
376  <file name="Amon_historical_CNRM-CM6_r1i1p1f1_TBD_%start_date%_%end_date%"
377   freq_output="1mo" append="true" split_freq="10y" timeseries="exclusive" >
378    <variable name="project_id" type="string" > CMIP6/CMIP6 </variable>
379    <variable name="activity"  type="string" > CMIP6 </variable>
380    <variable name="contact"  type="string" > contact.cmip@meteo.fr </variable>
381    <variable name="conventions"  type="string" > CF-1.7 CMIP-6.0 </variable>
382    <variable name="creation_date"  type="string" > 2016-09-19T16:34:42Z </variable>
383    <variable name="data_specs_version"  type="string" > TBD </variable>
384    <variable name="experiment"  type="string" > all-forcing simulation of the recent p
385    <variable name="experiment_id"  type="string" > historical </variable>
386    <variable name="forcing_index"  type="string" > 1 </variable>
387    <variable name="frequency"  type="string" > mon </variable>
388    <variable name="further_info_url"  type="string" > http://furtherinfo.es-doc.org/CM
389    <variable name="grid"  type="string" > TBD </variable>
390    <variable name="grid_label"  type="string" > TBD </variable>
391    <variable name="grid_resolution"  type="string" > TBD </variable>
392    <variable name="initialization_index"  type="string" > 1 </variable>
393    <variable name="institution_id"  type="string" > CNRM </variable>
394    <variable name="institution"  type="string" > Centre National de Recherches Météoro
395    <variable name="license"  type="string" > CMIP6 model data produced by Centre Natio
       Attribution'Share Alike' 4.0 International License (http://creativecommons.org/lice
       https://pcmdi.llnl.gov/home/CMIP6/citation.html.Further information about this data
```

```
<field field_ref="CMIP_bigthetao" name="bigthetao" operation="average" ts_enabled="true" ts_split_freq="10y">
   <variable name="realm" type="string" > ocean </variable>
   <variable name="variable_id" type="string" > bigthetao </variable>
   <variable name="standard_name" type="string" > SeaWaterConservativeTemperature </variable>
   <variable name="description" type="string" > Conservative Temperature is defined as part of the Thermodynamic Equation of Seawater 2010 (TEOS-10)
   2010 by the International Oceanographic Commission (IOC). Conservative Temperature is specific potential enthalpy (which has the standard name
   sea_water_specific_potential_enthalpy) divided by a fixed value of the specific heat capacity of sea water, namely cp_0 = 3991.86795711963 J kg-1
   Temperature is a more accurate measure of the "heat content" of sea water, by a factor of one hundred, than is potential temperature. Because of t
   regarded as being proportional to the heat content of sea water per unit mass. Reference: www.teos-10.org; McDougall, 2003 doi: 10.1175/1520-0485(
   .0.CO;2. </variable>
   <variable name="long_name" type="string" > Sea Water Conservative Temperature </variable>
   </field>
```

Summary

- <DQR-dr2xml-XIOS> pipeline is designed to facilitate the configuration of XIOS-enabled climate models on the road to CMIP6
- Is the best way to conform (as far as we can) to the data request
- Scraps the nightmare of a ‘by hand’ configuration
- By dynamically analysing the simulated period (and adapting output configuration consequently), prevents from stop/restart operations
- Thanks to XIOS pluggable operations, it can help convince groups to output more variables (that requiring simple operations on a model variable)
- Files written by XIOS configured by dr2xml will be CMIP6-compliant
- Avoid the CMORisation step in the data production workflow

Upcoming features (automatic configuration of) :

- spatial regridding (horizontal and vertical)
- various temporal and spatial aggregations (‘cell_method’)
- homemade list of output variables