Control vocabulary software designed for CMIP6 (PrePARE)

Denis Nadeau

Karl Taylor

Sasha Ames

Lawrence Livermore National Laboratory

Control Vocabulary

- Agenda

  — CMIP6 model output publication to ESGF
    - **Required** metadata must be stored along with the data.

  — PrePARE
    - Software developed at Lawrence Livermore Laboratory as an extension to the Coupled Model Output Rewriter (CMOR) in order to validate CMIP6 data using the authorized schemes created by the WIP.

  — ESGF publisher
    - Use Control Vocabulary program to validate file before publication.

  — Conclusion

CMIP6 descriptive information

- ▪ Way to organized information
  - – Organize and facilitate use of complex collection of anticipated CMIP6 output.

- ▪ Specify how the data should be delivered to the scientists.
  - — CMIP6 required metadata (31 Global Attributes)

CMIP6 global attribute	description	examples	Corresponding attribute in CMIP5	Form	When required?	Further information and rationale
activity_id	activity identifier(s)	"CMIP", "PMIP", "LS3MIP LUMIP" see note 3	project_id	CV	always	renamed more generically, since not all activities are projects; also multiple activities may now be listed separated by single spaces.
external_variables	external cell measures	areacella areacello	-	CV	whenever appropriate	list of cell measure variables (separated by single spaces) that are referenced but not included in the file. These variables will be stored independently in the CMIP data archive.

CMIP6 required metadata

- Quality assurance checks
  - https://docs.google.com/document/d/1h0r8RZr_f3-8egBMMh7aqLwy3snpD6_MrDz1q8n5XUk/edit
  - http://cmor.llnl.gov/mydoc_cmip6_user_input/

CMIP6 global attribute	CMOR source of global attribute	Required attribute?	Check that type is	Required check against CV?	Check also in file name?
activity_id	CMIP6_CV.json	yes	string	CV (consistency with experiment_id)	no
external_variables	cell_measure values in CMIP6 data request	conditionally	string	CV (consistency with variable_id and table_id)	no

Registered Activities and Models

PrePARE

- Validate CMIP6 file global attributes against Control Vocabulary file.
  - Validate required attributes (31).
  - Validate registered "institution" and "institution_id".
  - Validate registered "source_type" and "source_id".
  - Validate "experiment, experiment_id" and attributes associated with each experiment.
  - Validate "grid_label and native_resolution".
  - Validate creation time in ISO format (YYYY-MM-DDTHH:MM:SS)
  - Validate "furtherinfourl" from CV internal template.
  - **Validate variable attributes with CMOR JSON table.**

PrePARE

- Each experiment contains specific metadata, which are partly determined by the CMIP6 design.

- PrePARE will tell you when a user input for a specific experiment does not match the one found in this table.

- "source_type" is a combination of:
  - "additional allowed_mode components"
  - "required model components".

```
"piControl":{
    "activity_id":
        "CMIP"
    ],
    "description":"DECK: control",
    "start_year":"",
    "end_year":"",
    "sub_experiment":"none",
    "parent_experiment_id":[
        "piControl-spinup"
    ],
    "parent_activity_id":[
        "CMIP"
    ],
    "experiment":"pre-industrial control",
    "additional_allowed_model_components":[
        "AER",
        "CHEM",
        "BGCM"
    ],
    "required_model_components":[
        "AOGCM"
    ],
    "tier":"1",
    "min_number_yrs_per_sim":"500",
    "sub_experiment_id":"none"
}
```

PrePARE

- ■ "activity_id" (Validation Failed)

!!!!!!!!!!!!!!!!!!!!!!!!
! Warning: Your input attribute "activity_id" with value
! "ISMIP6" needs to be replaced with value "CMIP"
! as defined for experiment_id "piControl".
!
! See Control Vocabulary JSON file.(Tables/CMIP6_CV.json)
!!!!!!!!!!!!!!!!!!!!!!!!

!!!!!!!!!!!!!!!!!!!!!!!!
! Error: The input file is not CMIP6 compliant
! Check your file or use CMOR 3.2 to achieve compliance for
! ESGF publication.
!!!!!!!!!!!!!!!!!!!!!!!!

"piControl":{
        "activity_id":
            "CMIP"
        ],
        "description":"DECK: control",
        "start_year":"",
        "end_year":"",
        "sub_experiment":"none",
        "parent_experiment_id":[
            "piControl-spinup"
        ],
        "parent_activity_id":[
            "CMIP"
        ],
        "experiment":"pre-industrial control",
        "additional_allowed_model_components":[
            "AER",
            "CHEM",
            "BGCM"
        ],
        "required_model_components":[
            "AOGCM"
        ],
        "tier":"1",
        "min_number_yrs_per_sim":"500",
        "sub_experiment_id":"none"
    }

PrePARE

- CMIP6 Validation passed

```
****************************************************************************

* This file is compliant with the CMIP6 specification and can be published in ESGF. *

****************************************************************************
```

PrePARE

- **Pre-Publication Attribute Reviewer for the ESGF**
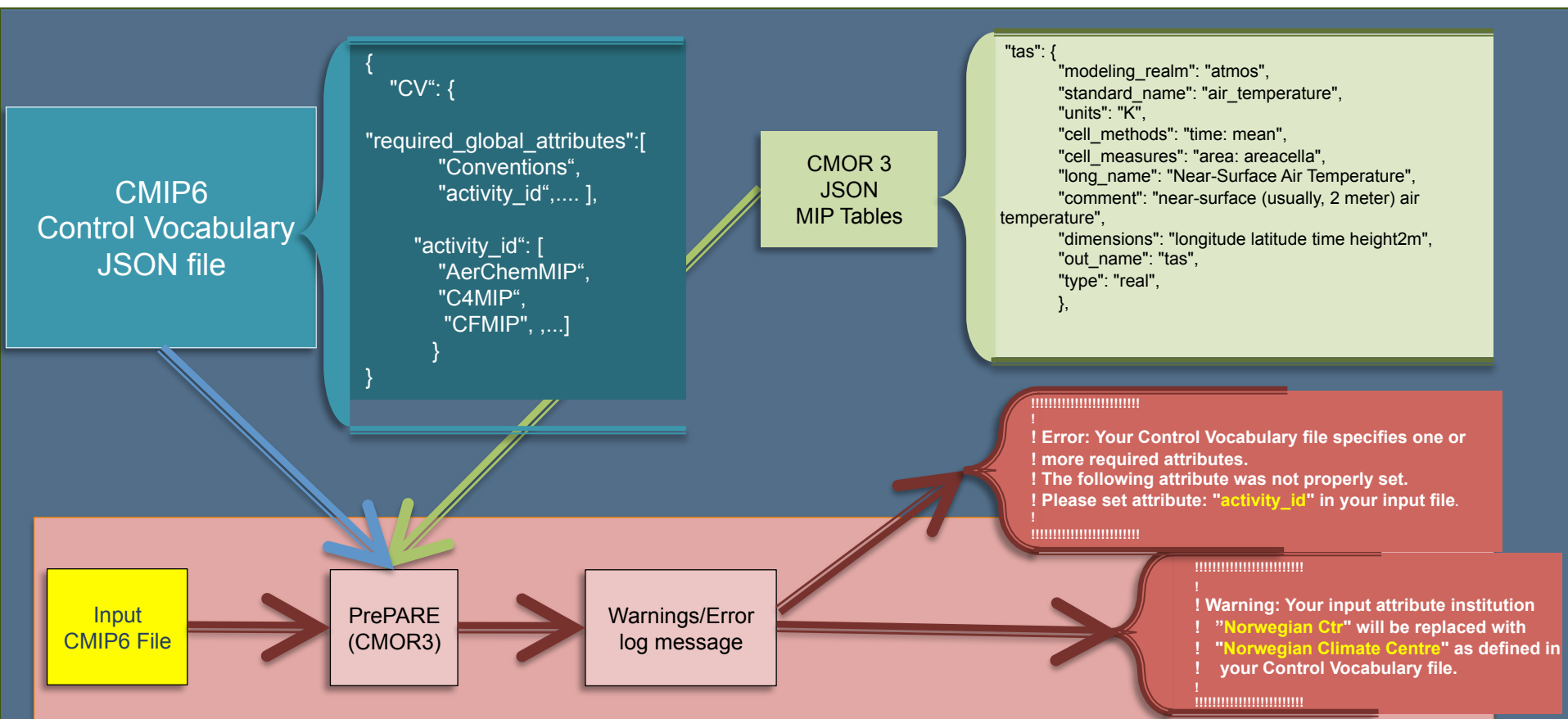
Control Vocabulary JSON file

- PrePARE will read any netCDF file and validate it's Global Attributes according any given Control Vocabulary File (CV).

- Can be modified to validate metadata of other projects.
  — Obs4MIPs (being written)
  — Input4MIPs (done)
  — CREATE (…..)

- Can use list of strings including regular expression.
  — "sub_experiment_id":

  ```
  [
  "^s[[:digit:]]\\{4,4\\}$",
  "none"
  ]
  ```

Conclusion

- CMIP6 Control Vocabulary is embedded in CMOR3.
  - If you use CMOR version 3, your data will be ready for publication to ESGF.

- ESGF publisher is running PrePARE as an attribute reviewer.

- Future work:
  - PrePARE will validate filename consistencies with CMIP6 requirements
  - If given, PrePARE will also validate file path.

  http://cmor.llnl.gov/
  conda install -c conda-forge -c pcmdi -c uvcdat cmor

US Citizen

April 20th 2016

Oakland, CA