

Reply to Reviewer 1 for following GMDD paper :

Geosci. Model Dev. Discuss., 6, 1–35, 2013
www.geosci-model-dev-discuss.net/6/1/2013/
doi:10.5194/gmdd-6-1-2013

Title: Development and exploitation of a controlled vocabulary in support of climate modelling

Authors: M.-P. Moine, S. Valcke, B. N. Lawrence, C. Pascoe, R. W. Ford, A. Alias, V. Balaji, P. Bentley, G. Devine, and E. Guilyardi

Review:

[07 October 2013 : Author Comments \(AC\) to reviewer comments:](#)

RC C1291: 'Referee comments on Moine et al, 2013', Alison Pamment, 23 Aug 2013

Authors express their sincere acknowledgements to Alison Pamment for this meticulous and well-explained review.

RC:> Specific Comments

RC:> The phrase “aware data-mining” (line 6, page 2969) is technical jargon and needs to be explained.

AC:> Agreed. The sentence is modified:

old sentence : “... is essential for aware data mining and unambiguous data interpretation by end-users”

new sentence: “... is crucial to guide end-users through data mining, data interpretation or data comparison tasks”

RC:> The last few sentences of section 1 (lines 19 – 25, page 2969) describe the structure of the paper itself. Section 6, “The information pipeline” appears to be omitted from this structure.

AC:> Agreed. Section 6 has been renumbered into 5.2 and the last sentence of introduction is augmented:

old sentence: “Finally, we explain how this CV was used to construct the “CMPI5 questionnaire”

new sentence: “Finally, we explain how this CV was used to construct the “CMPI5 questionnaire” and how it was ingested by other metadata systems like ESGF one”

RC:> I suggest that the title of section 3 be modified to read “Existing Metadata Standards for Weather Forecast and Climate”. The inclusion of the word “Standards” more accurately reflects the content of this section.

AC:> Not agreed. Actually it was deliberate to avoid the use of “Standards” word was when talking about existing metadata so long as they never turn out to be “real recognized standards”.

RC: > Paragraph 2 of section 3 (lines 15 – 16, page 2971) could be clarified by explaining that CF is a metadata convention while NetCDF is a binary file format before explaining that together they form the CF-NetCDF data format.

AC:> Agreed. The sentence is modified:

old sentence: One important standard widely adopted by the climate modelling community is the Climate Forecast (CF) convention (<http://cf-pcmdi.llnl.gov/>). Integrated within the selfdocumented NetCDF format it forms the CF-NetCDF data format.

new sentence: The Climate Forecast (CF) metadata convention (<http://cf-pcmdi.llnl.gov/>) and the NetCDF binary file format are standards widely adopted by the climate modelling community. Together they form

the CF-NetCDF data format.

RC:> In paragraph 2 of section 3 (line 17, page 2971), it would be more accurate to say that CF standard names are for “geophysical variables” rather than the narrower term of “climate variables”. This reflects the fact the standard names can equally well be applied to NWP data or indeed observations as well as climate data.

AC:> Agreed. “climate variables” changed into “geophysical variables”

RC:> Section 3 refers to “high level” and “low level” metadata (line 22, page 2971 and lines 3 – 4, page 2972, respectively). From the text I infer that “high level” metadata about the model and simulation would apply to whole data sets, whereas “low level” metadata apply to individual data files and describe their content. I feel the paper would benefit from a clearer explanation of this choice of terminology.

AC:> Agreed. A footnote is added form low-level/high level metadata:

“low-level metadata” term refers to metadata that applies to individual data sets and describes their content (i.e. what the data is), while “high-level metadata” refers to metadata that applies to whole datasets and addresses how the data were produced.

RC:> In addition, I do not think that the reference to “discovery” metadata in line 22 of page 2971 adds much to the description of the overall structure and organisation of the metadata, and the term is not explained, so I suggest its omission from the text.

AC:> Agreed. “discovery” term is removed.

RC:> The last line of Section 3 (line 16, page 2972) refers to an extract of the CMIP3 questionnaire in Appendix A. However, the CMIP3 questionnaire is in fact presented in Appendix B, therefore the reference needs to be changed, or the material in Appendices A and B needs to be interchanged.

RC:> Subsection 4.1 (line 19, page 2974) refers to “Appendix 7”. The list of contributing scientists is currently presented in Appendix A. Clearly the reference needs to be corrected and, as stated above, perhaps the content of the two appendices needs to be interchanged.

AC:> I agree, but I cannot manage to renumber/re-reference correctly the appendices despite corrected in the .tex document. I may have a bug with my latex version. My colleges told me that this kind of problem can be corrected by the editor.

RC:> Subsection 4.1.4 (line 2, page 2978) mentions three project acronyms: ENES2, EUFP7 and IS-ENES. All of these need to be explained.

AC:> Agreed. References/details are added:

IS-ENES2: <https://verc.enes.org/ISENES2/>

EU-FP7: EU’s Seventh Framework Programme for Research

IS-ENES: InfraStructure for the European Network for Earth System Modelling

RC:> I suggest that subsection 4.1.5 be renumbered to 4.2. The introduction to section 4 states that the “Model Controlled Vocabulary” and the “Simulations and Experiments Controlled Vocabulary” will be presented in turn. The former is presented in subsection 4.1 and it seems logical to number the latter as 4.2.

AC:> Agreed. Section renumbered into 4.2

RC:> In the current subsection 4.1.5 (line 14, page 2978) the acronym “AMIP” should either be fully explained or omitted as in this context it is not essential to understanding the purpose of performing different types of climate simulation.

AC:> Agreed. “AMIP” removed.

RC:> Subsection 4.1.5 (line 22, page 2978) introduces the “Conformance” concept but does not elaborate on how the CV helps to describe whether a simulation conforms to experimental requirements. A fuller

explanation is not given until section 5 of the paper. I suggest adding some explanation to section 4.1.5 regarding which metadata attributes are used to determine conformance, or at least adding a reference to section 5 to reassure the reader that this point will be addressed later in the paper!

AC:> Agreed. Details are given for “conformance”:

In current state, CV for conformance is quite restricted, asking meanly how experiment requirements are meet (if so). Possible choices are “via standard configuration”, “via model modifications”, “via inputs”, “via combination”, “not applicable” or “not conformant”. Its main function is to enforce a conformance check by metadata providers.

RC:> Paragraphs 2 and 3 of subsection 4.1.5 (line 23, page2978 – line 23, page 2979) describe the structure of the simulations and experiments CV, mostly by means of an example. I understand from the text that this CV was laid out in the CMIP5 documentation and therefore was not captured as a mind map. However, a Figure in the form of a tree diagram would significantly aid the reader in following this part of the text. Perhaps two tree diagrams are needed, one for the experiment attributes and another for the simulation attributes. In particular, I would like the diagram to illustrate the relationship between an ensemble simulation and the individual ensemble members which I found difficult to follow in the text. Does the whole ensemble have a “rip” value and then each member have another “rip” value or is it assigned only to the members?

AC:> Agreed. 2 diagrams are added (see figures A and B below – to be renumbered and referred to in the article).

A sentence to refer to figure B is added at the end of section 4.2: “Figure B illustrates how these attributes are filled in for an ensemble simulation labelled “decadal1959” that is an instance of the “1.1 decadal experiment”.”

RC:> Section 5, beginning on line 1 of page 2980, has only one subsection. I suggest, therefore, that the subsection 5.1 heading is redundant.

AC:> 5.1 section numbering is kept as long as section 6 is renumbered into section 5.2

RC:> Section 6, lines 12-13 of page 2983 refers to “a Schematron based validation”. I am unfamiliar with this concept and would like there to be at least one example of how or in what sense this approach checks for “deeper level coherency between the parameters”.

AC:> Agreed. More detail for schematron coherency check is given.

“The schematron validation ensures that parameters relevant only for a given condition are only filled when this condition is met. For example, in the description of the vertical grid, a SurfaceReference is asked only if the VerticalCoordinateType is mass-based. The pages of the Questionnaire being non-dynamic, the schematron function is to check coherency between responses given by the person filling the questionnaire”.

RC:> Section 7, line 2, page 2984 refers to “CMIP phase 5”. I suggest that this be replaced with “CMIP5” for consistency with the rest of the paper.

AC:> Agreed. “CMIP phase 5” replaced with “CMIP5”

RC:> Technical Corrections

AC:> Agreed. All technical corrections (vocabulary, spelling, etc.) have been taken into account.

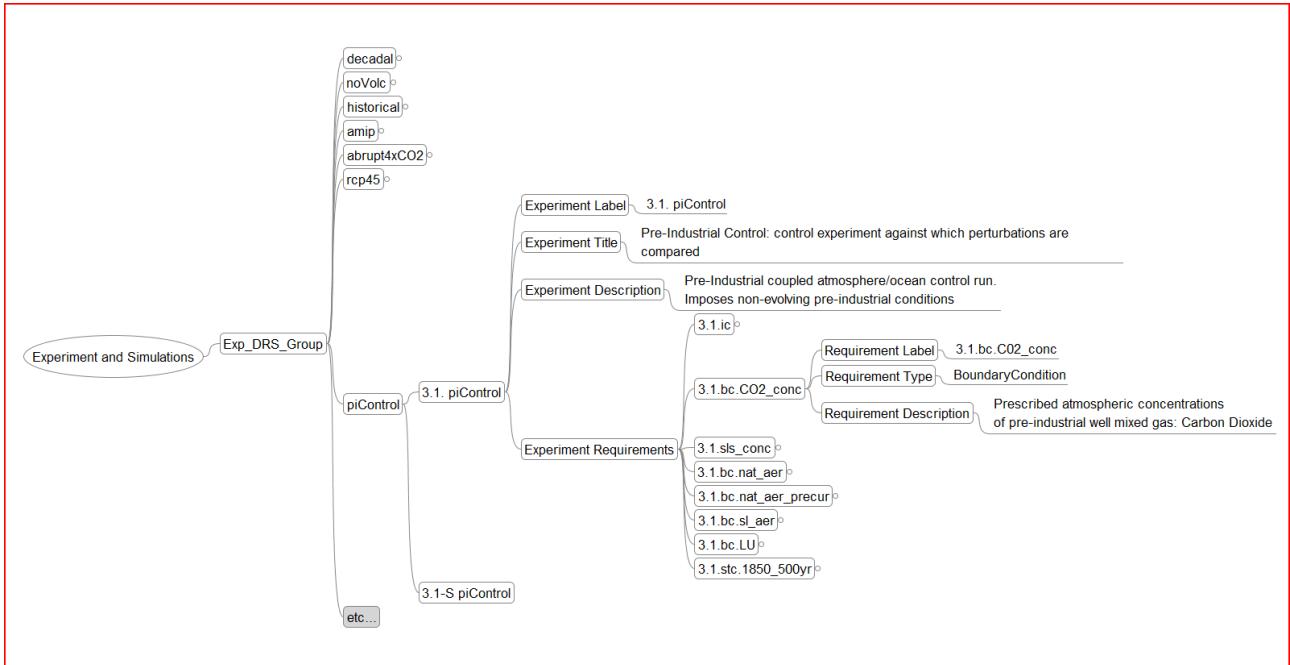


Figure A: Tree-diagram showing information necessary to identify and document an experiment. Example shown is the CMIP5 pre-industrial experiment. The experiment is identified by a label, a title, an associated description and the list of requirements to be fulfilled by the simulations that instantiate this experiment. Each requirement is in its turn identified by a label, a type and a description. Value (text) for these attributes is fixed once for all by the CMIP5 experiment protocol. Notice that this tree-diagram is just illustrative (it is not a CV-mindmap).

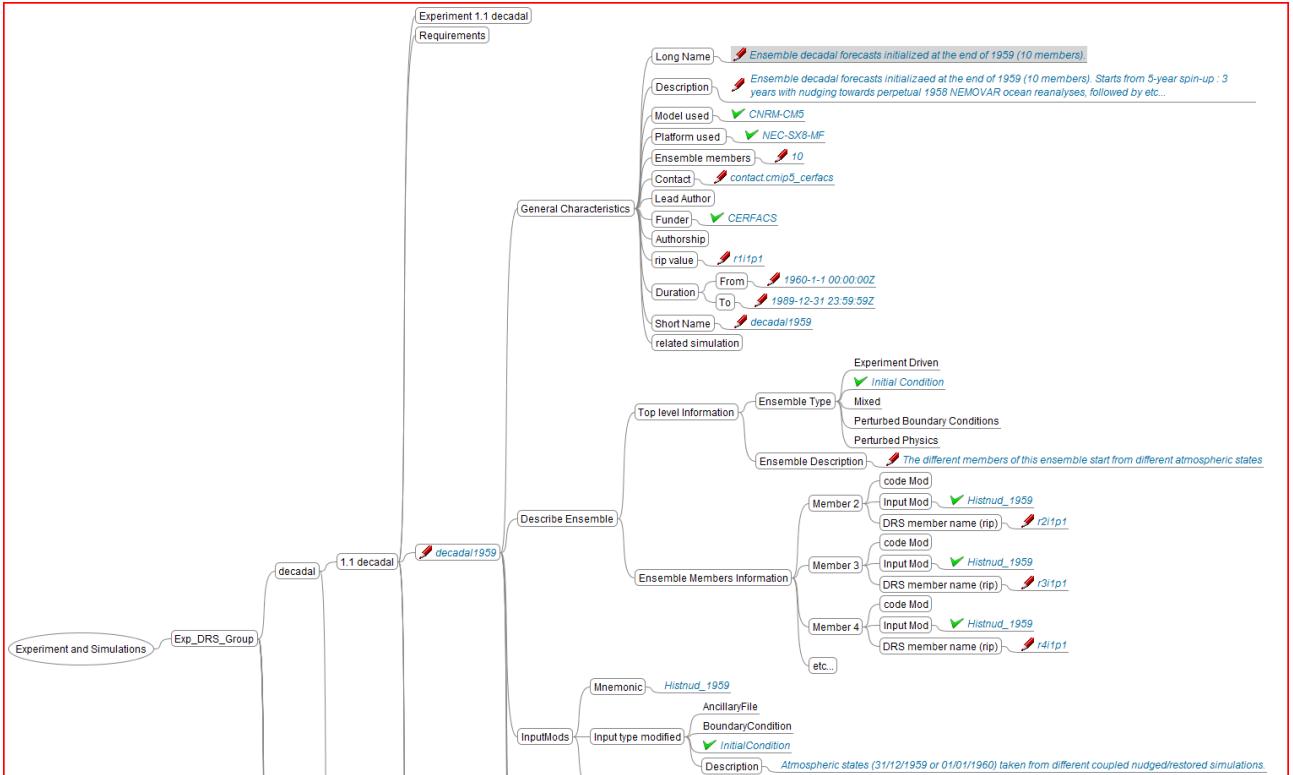


Figure B: Tree-diagram showing attributes used to describe an ensemble simulation. Here is reproduced what CERFACS group filled-up (blue text with a red pencil icon) or what he selected (blue text with a green tick-mark icon) through the CMIP5-questionnaire interface (see section 5). What can be deduced from the information given is that simulation labelled “decadal1959” realizes a “1.1 decadal” experiment using the “CNRM-CM5” model and was run on “NEC-SX8-MF” plateform. The simulation duration is 30 years (from beginning of 1960 till the end of 1989). The ensemble is made of 10 members, each being identified by a unique “rip value”; rip of the first member is used as identifier of the ensemble. Members can be distinguished by their initial condition. “Histnud_1959” is a mnemonic that refers to an “Input modification” the users has previously registered. It provides details about the difference between members of the ensemble (here different initial atmospheric states). Notice that this tree-diagram is just illustrative (it is not a CV-mindmap).