

Interactive comment on “Development and exploitation of a controlled vocabulary in support of climate modelling” by M.-P. Moine et al.

A. Pamment (Referee)

alison.pamment@stfc.ac.uk

Received and published: 23 August 2013

General Comments

This paper explains clearly the motivation for developing new controlled vocabularies (CV) to describe climate models and the climate simulations they produce. The processes, both human and technical, by which the vocabularies were developed are clearly set out. The paper then shows how the CVs are used to help standardize climate model metadata and aid the collection of those metadata. The text is well supported by appropriate Figures and references which assist the reader in assimilating the technical content of the paper. The references and internet links included in the paper allow the reader to access a considerable amount of material that sets the current work into context and provides further detail on the technical approach. The work

C1291

reported in this paper fills an important gap in the ability to document climate simulations in a standardized manner. The scientific significance of the work is: (1) it allows climate scientists and others using the outputs of climate models to more readily find data relevant to their work; (2) it facilitates easier identification than was previously possible of whether the data from two or more climate simulations are directly comparable. I recommend that this paper be published following minor revisions and technical corrections as detailed in the following.

Specific Comments

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

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.

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.

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.

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.

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

C1292

apply to individual data files and describe their content. I feel the paper would benefit from a clearer explanation of this choice of terminology. 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.

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.

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.

Subsection 4.1.4 (line 2, page 2978) mentions three project acronyms: ENES2, EU-FP7 and IS-ENES. All of these need to be explained.

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.

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.

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

C1293

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!

Paragraphs 2 and 3 of subsection 4.1.5 (line 23, page 2978 – 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?

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

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”.

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.

Technical Corrections

Page 2968, line 15: “increase” should read “increases”

Page 2968, line 16: “with increasing number” should read “with an increasing number”

Page 2968, line 24: “ data produced growing number of . . .” should read “data produced and the growing number of . . .”

C1294

Page 2969, line 1: “is in CMIP5 ensured by” should read “in CMIP5 is ensured by”

Page 2969, line 21: “systems in climate area” should read “systems in the climate area”

Page 2969, line 24: “CMIP15” should say “CMIP5”

Page 2970, line 17: “parameterizations turned on” would be better phrased as “parameterizations selected”

Page 2970, line 20: “aerosols types” should read “aerosol types”

Page 2970, line 26: “or solar irradiance by applying appropriate forcing . . .” should read “or solar irradiance and by applying appropriate forcing . . .”

Page 2970, line 23: “millennium” should read “millennial”

Page 2971, line 7: “utterly important” would be better phrased as “vitaly important”

Page 2971, line 20: “among other, short names . . .” should read “among other things, short names . . .”

Page 2972, line 1: “previous projects, among which” should read “previous projects, such as”

Page 2972, line 24: “constitutive” should be replaced by “constituent”

Page 2974, line 10: suggest replacing “intercomparison concern” with “the requirements of the model intercomparison”

Page 2974, line 10: “reach a level of details” should read “to reach a level of detail”

Page 2974, line 14: “consultation of number of climate scientists” should read “consultation with a number of climate scientists”

Page 2975, line 5: “giving at end” would be better phrased as “giving ultimately”

Page 2975, line 16: “in land surface realm” should read “in the land surface realm” and

C1295

“in atmospheric chemistry realm” should read “in the atmospheric chemistry realm”

Page 2976, line 17: “so to feed” should read “so as to feed”

Page 2977, line 3: “according the CMIP5 guidance” should read “according to the CMIP5 guidance”

Page 2978, line 6: “Unlike for” would be better phrased as “In contrast to”

Page 2978, line 17: “consisted firstly in encoding” should read “consisted firstly of encoding”

Page 2978, line 21: “experiment requirements it pretends to fit” would be better phrased as “experiment requirements it is intended to fit”

Page 2979, line 1: “In such CV, Experiments are . . .” should read “In this CV, experiments are . . .”

Page 2979, line 11: “One or several” would be better phrased as “One or more”

Page 2980, line 8: “2 simulations descriptions” should read “two simulation descriptions”

Page 2980, line 9: “simple-linear text-bases” should read “simple, linear text based”

Page 2980, line 19: “at end” would be better phrased as “ultimately”

Page 2980, line 18: I think “he” should say “he or she”

Page 2980, line 19: I think “included in” should be “including”

Page 2981, line 3: “hierarchy presented order of the parameters” should read “the hierarchy presented and the order of the parameters”

Page 2981, line 7: “taken as example” should read “taken as the example”

Page 2982, line 5: “capture” should read “captured”

C1296

Page 2982, line 8: “is targeted for” should read “is targeted at”
Page 2982, line 20: “on Fig. 7” should read “in Fig. 7”
Page 2983, line 2: “step passed” should read “step is passed”
Page 2983, line 9: “if a text is provided” should read “if text is provided”
Page 2983, line 21: “At end” would be better phrased as “Finally” and “conforming to CMIP5 DRS” should read “conforming to the CMIP 5 DRS”
Page 2983, line 22: “were broadcasted” should read “were broadcast”
Page 2984, line 2: For readability I suggest starting a new sentence after “several petabytes”
Page 2984, line 7: “that guarantees” should read “to guarantee”
Page 2984, line 19: “of large number” should read “of a large number”
Page 2984, line 20: “The CV collection raised” would be better phrased as “The CV collection produced”
Page 2984, line 21: “thousand of terms” should read “thousands of terms”
Page 2984, lines 21 – 22: “which hierarchical arrangement is not less important than the terms themselves” would be better phrased as “for which the hierarchical arrangement is equally as important as the terms themselves”
Page 2984, line 22: “Even though perfectible, METAFOR CV. . .” should be phrased as “Even though it can be improved further, the METAFOR CV. . .”
Page 2984, line 25: “user” should say “users”
Page 2984, line 28: “a complete round-tripping” could be better phrased as “the ability to convert between all CV formats”
Page 2985, line 6: “feedbacks” should say “feedback”

C1297

Page 2985, line 7: “these feedbacks” should read “this feedback”
Page 2985, line 8: “In minded focus” should read “The intended focus”
Page 2985, line 9: “ass” should say “as”
Page 2985, line 22: “as SPECS project” should read “such as the SPECS project”
Page 2985, line 24 (heading of Appendix A: “scientist” should say “scientists”
Page 2985, line 25 – 26: “all of the climate scientists” should read “all the climate scientists”
Page 2985, line 26: “to METAFOR” should read “to the METAFOR”
Page 2987, line 8: the entry for G. Mann is missing the country affiliation
Page 2987, line 14: R. Slawitch, Univ. Maryland is listed as being in the UK. Shouldn't this be USA?
Page 2998, legend of Fig. 5a, line 1: “inherits” should say “inherit”
Page 2998, legend of Fig. 5a, line 2: “designs” should read “determines”

Interactive comment on Geosci. Model Dev. Discuss., 6, 2967, 2013.

C1298