Geosci. Model Dev. Discuss., 5, C317–C319, 2012
www.geosci-model-dev-discuss.net/5/C317/2012/

Geoscientific
Model Development
Discussions

Interactive comment on "Quality assessment concept of the World Data Center for Climate and its application to CMIP5 data" *by* M. Stockhause et al.

Anonymous Referee #2

Received and published: 23 June 2012

This paper provides an overview of the federated quality control process developed for the CMIP5 Model Intercomparison Project. The process aims to overcome weaknesses of earlier attempts to provide data access to the climate science community, by introducing a version control system, and the ability to provide DOIs to uniquely identify published datasets. The paper includes some experiences in the initial roll out of the system.

The paper provides a fascinating glimpse into the work that goes into creation of data intermediaries who operate the Earth System Grid, and provide various quality control checks on the served data. The paper deserved to be published, although it would

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

be improved with a few minor clarifications, and a careful read through to improve the sentence structure, which is awkward in places.

I would suggest the authors attend to the following:

In the introduction, when you say that the data volume for CMIP5 is expected to be 100 times that of CMIP3, could you provide the actual numbers for this (even if they're still estimates)?

In section 2, when you distinguish SQA from TQA, you state that TQA can only be applied by the data intermediaries. Why? It's not clear to me why this couldn't alternatively be done by the data creator, by applying standardized data consistency checks. A few words about the rationale would be helpful. Would this have avoided some of the problems described later in the paper (or perhaps made them worse?)

Fig 3 is a little hard to parse - perhaps because it's not clear who generates the CIM metadata - it's shown as being outside the local data centre. Should there be a larger box representing the originating science lab that includes both the green and grey boxes?

The description of the problems experienced with the system is useful, but seems to lack a little context. I think the problems (especially with respect to delays in replication and DOI publication) should be understood as a trade off between meeting the timeline of the CMIP5 & IPCC process, versus completing the entire intended QC process. When it became clear that the process of providing the data would be slower than anticipated, one of these goals had to be weakened. Hence, I think the discussion would be much more useful if you include a brief summary of the expected CMIP5 timeline, and some of the reasons for the delays. Is it the case that the entire process turned out to be too ambitious, given the size of the set of CMIP5 experiments, and the large number of data fields required from the centres? I think a little more explanation of these causes are needed, especially to support the claim in the conclusion that that QC procedure is not slowing down the data publication process.

GMDD

5, C317–C319, 2012

Finally, in the process described, it's possible that a large proportion of the data never makes it through level 3 quality control (and therefore is never given DOIs). Yet such data might still be widely used and cited within the science community. How much does this matter, and would this result be viewed as a failure of the process as designed?

Interactive comment on Geosci. Model Dev. Discuss., 5, 781, 2012.

GMDD

5, C317–C319, 2012