

DRAFT

NOTE: The procedure written down here has been discussed before, and is the procedure proposed for making changes to the CMIP5 data services. It now will be widely circulated among the CMIP5 stakeholders, and may be revised based on their input. The CMIP panel will then be asked to formally approve the finalized procedure. Some aspects of this procedure that seem uncontroversial will be adopted immediately, even without formal approval.

Procedure for Improving Delivery of CMIP5 Model Output and Documentation

15 January 2012

1. Introduction

CMIP was established and is organized under the World Climate Research Organization's Working Group on Coupled Modelling (WGCM). Representatives from several of the major climate modeling centers serve on the WGCM. The WGCM has appointed a CMIP panel to oversee CMIP, and for nearly 20 years the WGCM has charged PCMDI with a role in coordinating CMIP.

This document focuses on a single aspect of CMIP5, namely, the effort to make CMIP model output and model documentation accessible to researchers and other interested parties in a coordinated way. Decisions concerning other aspects of CMIP5 (e.g., the production and rewriting of model output and generation of model and experiment documentation) may be arrived at by different procedures.

The IPCC has always had a keen interest in CMIP and the science it produces. In 1996 the IPCC appointed a Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA), which in subsequent years developed more and more interest in climate model output. The TGICA established a Data Distribution Centre, which has primarily served the needs of scientists contributing to the IPCC's Working Groups II and III. This center is run by a partnership of hosts, including the British Atmospheric Data Centre (BADC) in the UK and the Deutsches Klimarechenzentrum (DKRZ) in Germany. For CMIP3 PCMDI, which traditionally has focused on the needs of the IPCC WG I, began working closely with the BADC and DKRZ to make model output available in a coordinated way across the entire scientific community. For CMIP5 PCMDI continues to partner with the DDC, having established formal memorandums of understanding with BADC and DKRZ.

Thus, at present PCMDI coordinates CMIP5 under the guidance of the WGCM and has partnered with the BADC and DKRZ to address the needs of the IPCC. Day-to-day decisions are made by PCMDI, sometimes based on in-house judgment, but more often seeking input from its partners, the modeling groups, the CMIP panel, and/or other interested parties. This approach has worked very efficiently in the

DRAFT

DRAFT

past, and until recently, reasonable compromises and exceptional cooperation have led to a remarkably smooth running of the project.

Now that the number of users and those with strong interests in CMIP have expanded well beyond the modeling groups and climate scientists, it is proving more difficult to forge compromises and encourage cooperation, which is one reason why this document has become necessary. One thing that cannot change is that ultimately someone must retain primary responsibility for coordinating CMIP, and that responsibility is PCMDI's. This document formalizes, but does not fundamentally alter the way that PCMDI will make decisions, which is to: seek input from all interested parties, consult with its partners at BADC and DKRZ, rely on the CMIP panel (and perhaps a complementary advisory panel that could be established), and attempt to build a consensus. The term "CMIP5 Data Centers" will subsequently be used to refer to PCMDI, BADC, and DKRZ.

The formal procedures proposed below should be followed whenever substantial changes are contemplated or when small changes might have important impacts on any of the stakeholders. In order to proceed expeditiously, other decisions should be made in the traditional, less formally prescribed way.

2. Stakeholder requirements and concerns

All decisions involving the delivery of CMIP5 model output and documentation to users should be sure to take into account the requirements of all stakeholders, including those listed below (nb.: certain requirements are shared by multiple stakeholders, but are listed under a single one):

1. User group 1 (low data volume, typically an individual investigator)
 - a. Easily understood point-and-click interface to files and to generate download scripts.
 - b. Fast search
 - c. Access to model and simulation documentation
 - d. Minimization of number of points and clicks and "sign-in" requests required
 - e. Good "help" tools
 - f. Prompt notification of errors discovered in downloaded files
 - g.
2. User group 2 (high data volume, typically a research organization)
 - a. Option to bypass point-and-click interface and script downloads
 - b. Fast alternatives to http transfer (e.g., gridftp)
 - c. Alternatives to point and click search (e.g., through an API)
 - d. Minimization of impact on any operational scripting capabilities (developed to bypass limitations of current services)
 - e.

DRAFT

3. Modeling groups
 - a. Uninterrupted user-access to their model data and documentation
 - b. Record of usage documenting scientific impact of their model's output
 - c. Terms-of-use controlled access to their data
 - d.
4. Data node managers
 - a. Easy installation and upgrade of node software
 - b. Minimization of number of upgrades required
 - c. Easy publication procedure
 - d. Generally, minimization of resources required to maintain data node
 - e.
5. GO-ESSP and ESGF software developers
 - a. Minimization of duplication of effort
 - b. Justification for funding used to develop software contributing to CMIP
 - c.
6. WCRP's WGCM
 - a. Assurance that all users can easily obtain model output and documentation through a coordinated delivery system
 - b. Minimization of the burden on the modeling groups
 - c.
7. IPCC's TGICA
 - a. Assurance that the IPCC's needs are met
 - b.
8. Metafor and ESMF curator
 - a. Easy access to information collected using the Metafor questionnaire
 - b.
9. CMIP5 Data Centers (PCMDI, BADC, and DKRZ)
 - a. Assurance that the expectations of the WGCM and TGICA are met
 - b. Minimization of resources needed to provide help to users
 - c. Justification and due credit for resources devoted to effort
 - d. Facilitation of replication
 - e.
10. Other gateway managers
 - a. Easy installation and upgrade of gateway software
 - b. Minimization of number of upgrades required
 - c. Generally, minimization of resources required to maintain gateway
 - d.

DRAFT

DRAFT

It is expected that other stakeholders and additional requirements will be added to this list as they become apparent. Note that the above list summarizes general requirements and is not meant to specify how these requirements should be satisfied.

3. Procedure for upgrading CMIP5 data services

In planning a change in CMIP5 data services, the general goals include:

- Improving user ability to easily find and download CMIP5 output and documentation of interest to them
- Ensuring smooth transitions so that users are not inconvenienced
- Minimizing resource requirements and other disruptive impacts on each of the three CMIP5 data centers
- Minimizing efforts required of CMIP5 data *node* managers
- Anticipating future changes in the system and planning for how those can be implemented smoothly
- Encouraging backward compatibility, so that the entire system need not be upgraded simultaneously.

Because the system currently serves hundreds of users whose access to CMIP5 data must not be interrupted, it is imperative that thorough testing be performed before implementing any changes. Thus, the following step-by-step procedure should be followed.

Step 1: Proposal for change

A written summary of proposed changes to the CMIP5 data services should be submitted to the CMIP5 Data Centers. The summary should include what parts of the system will be affected directly and why the changes are being proposed. PCMDI in consultation with its partners will decide whether the CMIP panel should be contacted for input and then will make a decision on whether to move forward. If the proposal merits further consideration, then input will be sought from stakeholders whose interests might be significantly impacted. Assuming that any objections can be adequately addressed or are not judged to be important enough to hold up further consideration of the proposal, approval will be granted to move to the first stage of testing. If two competing proposals for upgrading services are under simultaneous consideration, a list of criteria for assessing the relative merits of the competing systems will be agreed upon by the Data Center partners and will subsequently be used in helping to decide which system (if either one) should become operational.

DRAFT

Step 2: Closed testing stage

A version of the data delivery system that includes the proposed changes should be established at one of the CMIP5 Data Centers, or at another facility agreed upon by the Centers. An appropriate group of “early” (friendly) testers will be invited to evaluate the new system. Next (unless this was already part of the earlier tests), the system should be tested in a “federated configuration.” This may in some cases require test data nodes or test gateways to be established at more than one CMIP5 Data Center. The testing outlined above should in so far as possible exercise the software under realistic operational conditions expected to be encountered in application to CMIP5.

This first stage of testing will be an iterative process in which bugs are uncovered and eliminated. PCMDI, in consultation with the other CMIP5 Data Centers will determine if and when the new system appears ready for testing by a wider group of users.

Step 3: Beta-testing

Beta-testing should not occur until:

- the system has been tested sufficiently by the developers and CMIP5 Data Center partners,
- all (important) bugs discovered have been corrected, and
- any major concerns of the Data Center partners have been dealt with.

During the beta-testing period, the proposed new system will be operated in parallel with the operational system. Users will be invited through the CMIP5 Data Center websites to use the beta version, but will be warned that support (“help”) for the new system will be limited. The stakeholders and users will be asked to report any problems encountered and convey any feedback, concerns, and suggestions to the CMIP5 Data Centers. Bugs in the beta version should be fixed as they are found, and user input should be relied on to improve the system (while remaining within the general scope of the proposed changes).

While beta-testing is going on, the operational system should not be disrupted. During this period, users will be able to access CMIP5 model output via both the old system and the beta test version of the system.

Step 4: Decision to upgrade

After a reasonable period of beta testing, and when all of the problems identified by users have been corrected and all concerns addressed, or when explanations have been provided as to why nothing has been done, PCMDI will consult with the CMIP5 Data Center and make a decision on whether to upgrade. Some of the criteria that

DRAFT

need to be considered in determining when the system should be upgraded are listed in section 4 below. With the understanding that the ESG federation relies on good-faith intentions to work together, every effort will be made to address any objections to a proposed course.

Step 5: Upgrade implementation

Once the decision to upgrade has been made a plan to smoothly implement the transition will be transmitted by PCMDI to all those affected. In particular any actions required by anyone affected by the changes will be made clear well in advance of any deadlines. In particular, if the changes require actions by data node managers, flexibility in scheduling will need to be accommodated.

4. Some criteria affecting decisions to change the CMIP5 data services

The following questions should be considered in deciding when the CMIP5 data delivery system should be replaced by an upgraded system:

1. Is the new system stable and robust enough to provide users with an improvement over the existing system?
2. Will any capabilities found in the existing system be lost, and if so, will users miss them more than they are likely to value improvements in the existing system?
3. Would it be better to delay installation of the new software until further capabilities are ready for deployment?
4. Have all the concerns reported during the testing phase been addressed?
5. Has the testing phase been thorough enough to likely reveal all major problems?
6. Will users have to adapt substantially the way they interact with the system, and if so, will they perceive the improvements worth the investment of their time?
7. Are the developers comfortable that their software is ready for operational service of CMIP5 model output?
8. Will the changes impose additional burdens on data node managers?
9. Will the changes impose additional burdens on the CMIP5 Data Centers?