

Global ESG Technical Workshop

9.00 am until late, Monday 5th of November.

This short note has some background material on meeting aims, and a draft agenda for the meeting, which is to be held in the seminar room of DKRZ (Deutsches Klimarechenzentrum, Bundesstr. 55, D-20146 Hamburg). Further documents will be tabled at the meeting itself (and circulated on go-essp-tech).



Issues

There are a number of issues which need to be resolved for the development of the global federation to deliver CMIP5/AR5, and we certainly can't address them all at this meeting. The issues on the table include, but are not limited to:

1. Issues associated with “procedure”, that is, how we expect data and metadata to flow within the “global ESG consortia”, in particular between centres holding complete copies of the CMIP5 data.
 - It is apparent that the “initial pump-priming” of the ESG core copies will have to occur using physical disk transfer. The details of that pump-priming will be out of scope for this particular day, but a document describing those details will be worked on during the week.
2. Issues associated with “timing”, that is, what the current technical roadmap looks like, and how we can milestone from where we are to where we need to be to deliver against the requirements of the global consortia (and possibly exploit technical resources outside the existing ESG team), and
3. Specific details associated with “replication”, as we have previously identified this as an area of development where our requirements are not well enough scoped, and the software is not yet mature (enough).
 - While we might expect some of the replication to occur via physical disk transfer, not all will do so, so we need to address both “network-replication” and managing the identification of data which needs replication, and of data which has “appeared” from “out-of-band” replication.

The phrase “global consortia” is intended to indicate that these are CMIP5 requirements that go beyond the original ESG technical aims of data nodes and managed gateways at PCMDI and NCAR to be inclusive of a) “core archive” replicates elsewhere, and b), possibly, core data being ingested at more than one point (i.e. core data being “acquired” at BADC and WDCC).

The terms “ingested” and “acquired” will need definition. Currently we expect data to “enter” ESG via publication in an ESG data node and via physical disk replication. Such data will in many cases be a superset of the “core” data expected from each modelling centre. How and where will core “subsets” be identified for the purpose of replication? What would be the constraints on timing of replication? How will we handle versioning, both in response to “routine QA” by PCMDI (or others) and “updated science” (ie. Model fixes). When, where, and how will we tag the AR5 version of the data?

Clearly there are a large number of technical activities underway, and it might be that we have an opportunity at this meeting to resolve some issues, but the primary aim is to establish shared understanding of a) what is going to happen, when, and who does what; along with b) how we will communicate and agree changes to our shared understanding.

Draft Agenda

Chair for the day: Bryan Lawrence

09.00 – 09.15	Welcome, Introduction, General Aims
09.15 – 10.00	Getting to grips with the general requirements of the global consortia
10.00 – 10.30	Replication in the context of data flow. Requirements.
10.30 – 11.00	General discussion, with coffee
11.00 – 12.00	Replication in the context of data flow. Software, Deliverables. Milestones.
12.00 – 12.30	Versioning in the context of replication of replaced data and replicants ¹ . Requirements.
12.30 – 13.30	Lunch
13.30 – 14.30	Establishing milestones and processes for establishing requirements beyond those in the “first” ESG release. Can we exploit a bigger team beyond (the original) ESG team to help deliver? Time and duration of next “team” telecon.
14.30 – 15.00	Registration, Access Control and Authorisation. (30 minutes maximum, those interested in more can pick it up over dinner etc). Concentrate on expected procedures, and what services need to be running (as opposed to their technical details).
15.00 – 15.30	General discussion, with coffee.
15.30 – 17.00	Technical Discussion: ESG data node installation issues. ESG gateway architectures and futures.
17.00 – 17.30	Agree a schedule of telecons on specific issues (and perhaps a schedule of telecons on general issues as well).
17.30 – & on	More Technical Discussion. Beer.

Attendance

Note that while this is an open workshop, space is limited. If you haven't already been asked to come, or advised Dean Williams or Bryan Lawrence that you are attending, please do so asap. If too many take up this offer, we reserve the right to refuse attendance on space grounds. Sorry.

¹ We want to avoid a discussion of DOI management etc at this stage. Not that it's unimportant, but to limit the scope of what we can achieve in a short time.