

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2022-73-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2022-73

Karl E. Taylor (Referee)

Referee comment on "Twenty-five years of the IPCC Data Distribution Centre at the DKRZ and the Reference Data Archive for CMIP data" by Martina Stockhause and Michael Lautenschlager, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-73-RC1>, 2022

General comments:

The infrastructure that enables scientists and others to access, analyze, and cite climate model simulation output may be as important in advancing the science as the development of new models and the design of new diagnostic experiments. This manuscript addresses one aspect of the coordinated international effort to serve CMIP data, and in particular the use of that data in support of the IPCC assessment process. That the subject is perhaps somewhat at the periphery of GMD's focus does not detract from its importance in documenting the advances being made in sharing, curating, archiving, and enabling citation of climate data. It highlights the DDC's and the IPCC's increasing emphasis on FAIR principles (Findable, Accessible, Interoperable, Reusable), principles championed by CMIP long before the acronym appeared in the published literature. It provides an important review of DKRZ's ongoing leadership in establishing high standards for the the preservation and curation of climate data that will be of use to researchers for many years and will ensure availability of the data underpinning the IPCC assessments.

The paper overall should be of interest now and in the future as a record of the evolution in the DDC's hosting of climate model data serving the IPCC.

Specific comments:

- This article tells the history of the DDC from the perspective of one of its original partners, DKRZ. The DKRZ group is the one with primary responsibility for the DDC's collection of climate model data, which is produced primarily by the CMIP and CORDEX projects. Thus, the article's focus is on this. For readers unfamiliar with the DDC,

however, there are some gaps in the summary that need filling.

If the focus is meant to be on the CMIP and CORDEX simulation data hosted by the DDC, then a much more complete discussion is needed describing how the DDC partners with the responsible projects (CMIP and CORDEX) to meet the special needs of the IPCC.

Among the questions that should be addressed are:

- What distinguishes the DDC from the distributed CMIP and CORDEX data archives? Why is the DDC necessary, given that much of the model output would already be hosted and served by the projects generating the data? Does the DDC provide added-value products? (There is mention of the IPCC "atlas", but the description does not indicate what's contained in it. From the website it appears that except for a small amount of observed precipitation, temperature, and surface wind data, the data is otherwise based entirely on the CMIP and CORDEX data. It would be good to highlight how the atlas adds value to the original CMIP and CORDEX data.)
- Have there been special DDC requirements for metadata or data services that, were it not for the leadership provided by the DDC, might have been overlooked by CMIP and CORDEX? What data requirements and services were already part of CMIP and CORDEX planning, and what role did the DDC play in making sure any special needs of the DDC were met? Were there any DDC requirements that CMIP and CORDEX were unable to implement? Did DKRZ need to supplement the metadata or services already delivered by the CMIP/CORDEX projects before data could be served by the DDC?

If the intent of the article is to provide a balanced history of the DDC and how the role of DKRZ has evolved, then much more context is needed. The reader should not be left with questions like:

- What have the other DDC partners contributed? Do they hold replicas of the data hosted by DKRZ? Do they host data not available at DKRZ?
- If there is data hosted by the DDC other than CMIP/CORDEX data, is it structured in a common way and are uniform metadata requirements imposed? If so, does the DDC rewrite the data to meet the requirements or insist that the contributors meet the requirements? Who came up with any requirements for data other than the CMIP/CORDEX data?
- For in-text references like "<https://ipcc-data.org/>, last access: 10 March 2022", could this be made into a hidden link? If not, please delete the "last access:" part. The long reference to a date of marginal interest interrupts the sentence and makes it much more difficult to digest. If you need to indicate a date, do it only once and note that "this and all subsequent URL links were last accessed 10 March 2022".
- Similarly, in documenting meetings, the reading would be improved if the actual date were omitted (i.e., only include month and year). I don't think anyone cares what days the meetings were held.

- Near line 55, why is there no mention of CMIP6? The section is labeled "The Reference Data Archive at the DDC at DKRZ". Isn't CMIP6 data included in the reference data archive? Also wasn't there CORDEX data archived too?
- Line 139: It wasn't clear what was meant by "bidirectional references between ..."
- Lines 150-155: The explanation here of what the procedure is. Please try to be as specific as possible.
- Line 156: Weren't ESGF, the WIP, and PCMDI essential in already having established metadata standards, catalogs, search engines, and more, along with long term preservation, which had gone a long way toward making CMIP data "FAIR"? I think that has to be recognized.
- Lines 180-193: There seems to be little point of including this list here. Add some content describing why the individual collaborations are needed, or eliminate the list (or move it to the "Acknowledgments").
- Line 216: It would be good to recognize the work over 30 years at PCMDI putting in place the "standardization of CMIP" data.
- Lines 245-255: The list of "gaps" needs to provide a bit information about what exactly needs improvement. It would be good to say for each item how improvement in the area will help users or data managers, or whomever.

Technical corrections:

I will separately send the authors a marked up version of their manuscript that includes editorial suggestions for improving clarity and readability, which they might consider.