

Interactive comment on “Coordinating an operational data distribution network for CMIP6 data” by Ruth Petrie et al.

Anonymous Referee #2

Received and published: 21 November 2020

The paper presents the CMIP6 data storage tools and strategy. The new infrastructure is design to account for the large increase in data volume produced by MPIS.

The paper describes the data challenge in preparation of the deployment of the ESGD data nodes. This challenges is divided in 5 steps that increasingly test the node capabilities (from install to deployment in the production environment).

This infrastructure is remarkable for multiple reasons. The first is the coordinated effort to create an efficient distributed storage infrastructure. The second is the proposed architecture and provided functionalities to account for the MIP constraint and operational capabilities. For instance the creation of the “Errata Service” or the “Persistent Identifier” are providing crucial operational capabilities. The technologies chosen for the project appears to be very appropriate.

C1

The organisation of a “data challenge” in order to prepare to the operational deployment of CMIP6 looks like a very adequate and structure approach.

The paper covers a lot of ground because the complexity (technical and governance-wise) and richness of the distributed infrastructure. The readers, frequently would like to know more but I believe the authors have found a good tradeoff.

Because the amount of data is at an unusual scale, one of the points that could be better explained in the paper is related with performance issues (especially between centers/nodes) and backup strategies (with redundancy considerations).

It would have been nice to compare the proposed architecture with other international effort such as SeaDataNet (<https://www.seadatanet.org/>) to better understand the choices made for CMIP6. This initiative <https://www.openstoragenetwork.org/> could also be considered.

Overall the paper is extremely interesting and well written.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2020-153>, 2020.

C2