

Response to reviewer #2

The authors would like to thank the reviewer for their comments, responses are provided here:

1. “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).”
 - The performance of individual nodes is primarily determined by the adherence to infrastructure configuration recommendations as detailed in section 4. If a particular site deviates from the recommendations, which may happen for technical, resource or political issues performance may be impacted. The degree to which it is impacted is reflect in how far the deployed infrastructure deviates from the recommendations. A statement to this effect has been included in the final paragraph of section 4.
 - It is the aspiration that all data published will have a primary copy and at least one secondary copy made by one of the larger sites. For the more popular CMIP6 data provided for the climate research community such as the future climate projections data it is usually the case that a number of copies are made at large sites for their own purposes. These copies however act as backup copies and provide redundancy of data in the federation. As noted in response to Reviewer 1, it is the responsibility of each modelling centre and their allocated publication node to ensure that a back-up copy of the data exists. This has been discussed on p5, para. 2.
2. 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.
 - The other international efforts to distribute Earth System Science data while useful in their own right did not fit the needs of the Climate modelling community. Firstly SeaDataNet is a European research infrastructure which also provides some services exploited by a global community. Data access is via a conventional dataset selection -> add to basket -> download route, it does not scale well to the volumes of data in the use case of CMIP6 where it is not uncommon for users to require thousands of datasets. Metadata standards are expressed as schema documents, so that compliance can be evaluated with generic tools, this is something that could be of benefit if it were able to be integrated into the ESGF infrastructure.
Secondly the OpenStorageNetwork is a US NSF research project developing new approaches to standardizing the design of data storage and analysis clusters. This project is relatively new and development and deployment of the operational CMIP6 ESGF infrastructure was already committed to by the World Climate Research Programme (WRCP) Working Group on Climate Modelling (WGCM) before this project came into being. In addition ESGF offers an additional layer of data discovery that does not seem to be offered by the OSN which is primarily a data delivery solution. That said there is always scope for communities, programmes and projects to collaborate and learn from each other and I would expect the ESGF community to ensure that where any new best practice is identified that this be incorporated. An additional paragraph addressing this issue has been included in the conclusions.