

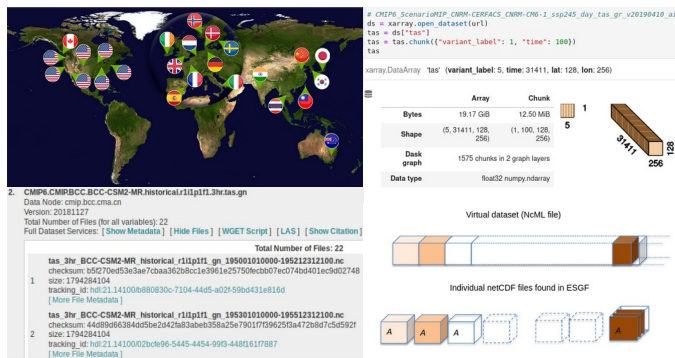
A Science Gateway for climate data analysis based on Virtual Analysis Ready Data

Ezequiel Cimadevilla

Instituto de Física de Cantabria (IFCA, CSIC-UC)
Avenida de los Castros, Santander, Cantabria, Spain
ezequiel.cimadevilla@unican.es

ABSTRACT

The Earth System Grid Federation (ESGF) Virtual Aggregation aims to enhance the accessibility and facilitate climate data analysis of datasets housed within the ESGF, a globally distributed network of research centres collaborating as a federated data archive. The ESGF provides modelling groups with nodes for publishing and archiving their model outputs to make them accessible to the climate community at any time. The federation has been chosen as a platform for dissemination of climate data of international initiatives such as CMIP and CORDEX. Thus, the ESGF has become the reference archive for Assessment Reports (AR) on Climate Change produced by the Intergovernmental Panel on Climate Change (IPCC). However, use cases in scientific research often involve calculations on datasets spanning multiple variables, over the whole time period and multiple model ensembles.



The ESGF Virtual Aggregation extends the federation capabilities, beyond file search and download, by providing out of the box remote climate data analysis capabilities over data analysis ready, virtually aggregated, climate datasets, on top of the existing software stack of the federation. In contrast to other new methodologies based on the creation of Analysis Ready Data (ARD) and physical duplication of the datasets of the ESGF, this work presents a methodology based on logical and virtual aggregations that does not require duplication of the

data. Moreover, this work exploits all the existing data services already available in the federation. Thus, climate analysis ready data is provided in a much more sustainable way.

The ESGF Virtual Aggregation is provided both as a dataset and as a scientific gateway for users interested in climate data analysis. This work describes the methodology, implementation and features of the ESGF Virtual Aggregation. Also, a comparison between additional methods for the generation of climate analysis ready data within the context of the ESGF is provided. The trade-offs of the different approaches are considered and a comparison of climate data analysis use cases will be presented.

Keywords – *esgf, analysis ready data, climate, netcdf, science gateway*

REFERENCES

- [1] R. P. Abernathy et al., “Cloud-Native Repositories for Big Scientific Data,” *Comput. Sci. Eng.*, vol. 23, no. 2, pp. 26–35, Mar. 2021, doi: 10.1109/MCSE.2021.3059437.
- [2] D. N. Williams et al., “A Global Repository for Planet-Sized Experiments and Observations,” *Bulletin of the American Meteorological Society*, vol. 97, no. 5, pp. 803–816, May 2016, doi: 10.1175/BAMS-D-15-00132.1.
- [3] L. Cinquini et al., “The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data,” in 2012 IEEE 8th International Conference on E-Science, Chicago, IL, USA: IEEE, Oct. 2012, pp. 1–10. doi: 10.1109/eScience.2012.6404471.
- [4] C. Stern et al., “Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production,” *Front. Clim.*, vol. 3, p. 782909, Feb. 2022, doi: 10.3389/fclim.2021.782909.
- [5] J. T. Overpeck, G. A. Meehl, S. Bony, and D. R. Easterling, “Climate Data Challenges in the 21st Century,” *Science*, vol. 331, no. 6018, pp. 700–702, Feb. 2011, doi: 10.1126/science.1197869.
- [6] J. L. Dwyer, D. P. Roy, B. Sauer, C. B. Jenkerson, H. K. Zhang, and L. Lyburner, “Analysis Ready Data: Enabling Analysis of the Landsat Archive,” *Remote Sensing*, vol. 10, no. 9, p. 1363, Aug. 2018, doi: 10.3390/rs10091363.