

EGU2020-10341

<https://doi.org/10.5194/egusphere-egu2020-10341>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



NetCDF in the Cloud: modernizing storage options for the netCDF Data Model with Zarr

Ward Fisher¹ and Dennis Heimbigner²

¹NCAR/UCAR, Unidata, Boulder, United States of America (wfisher@ucar.edu)

²NCAR/UCAR, Unidata, Boulder, United States of America (dmh@ucar.edu)

NetCDF has historically offered two different storage formats for the netCDF data model: files based on the original netCDF binary format, and files based on the HDF5 format. While this has proven effective in the past for traditional disk storage, it is less efficient for modern cloud-focused technologies such as those provided by Amazon S3, Microsoft Azure, IBM Cloud Object Storage, and other cloud service providers. As with the decision to base the netCDF Extended Data Model and File Format on the HDF5 technology, we do not want to reinvent the wheel when it comes to cloud storage. There are a number of existing technologies that the netCDF team can use to implement native object storage capabilities. Zarr enjoys broad popularity within the Unidata community, particularly among our Python users. By integrating support for the latest Zarr specification (while not locking ourselves in to a specific version), we will be able to provide the broadest support for data written by other software packages which use the latest Zarr specification.