CHFP Data Server Guide

V 1.0 11 June 2008

Nomenclature: “Data Server” refers to the organization which is serving CHFP data via an openDAP server.

1. OpenDAP server technology

Each organization should use an appropriate OpenDAP server. It is recommended to use an aggregation server such as THREDDS. So far, aggregations have only been made using local data, but if everyone is able to serve the data to a consistent standard, then it should be possible to form global aggregations of the CHFP data. It is presently our target to reach this level of uniformity, although this may need to be revised.

2. Creating netCDF data suitable for serving

Hopefully the data provided by the data producers is already in a form close to being ready to serve. However, in some cases it may be desired to combine data from different models in aggragate netCDF data files, in which case the Data Server will need to combine the data and set the netCDF headers accordingly. The CHFP netCDF specification is described in a separate document.

3. Testing the OpenDAP server and its contents

We hope that the Data Servers will do a good job of putting the data on a server, and testing that the data is accessible and has the correct metadata. As a starting point, we provide a simple script which attempts to retrieve (one by one, for further processing such as calculating SST indices) all of the monthly surface temperature data from a given http address, using nco operators. Please let Tim Stockdale know of any issues with this script - it is written in ksh, and the syntax of some of the UNIX utilities may not be universal.

4. Data volumes, server capacity and bandwidth

At the moment we have little experience with these issues. Data volumes are expected to be in the range of 0.5-2 Tbytes per model. If all of the model producers submit all of the requested data (particularly the 6h surface data from the ocean), then the total data volumes will be large. Although the “requested output” for CHFP was produced by committee discussions, we suspect that some of the largest parts of the dataset may be very little used. We thus encourage priority to be given to some parts of the data - particularly the monthly means, and daily data for key surface variables such as total precipitation and 2 metre Tmax/Tmin.

The capacity requirements for the server, and the bandwidth of the connection to the internet are expected to have an impact on the performance of the service given to users of the data. But again, we have little experience of this, and in many cases the restrictions might come more from the user end. Once we have services up and running we will be able to assess this.

5. Sharing tools

If you develop any useful tools, eg for authoring the netCDF files and their metadata, and if those tools are based on widely available software, then it might help others if you are able to make them available, eg for download from the CHFP web pages (or a link from them). Unfortunately, the ENSEMBLES data were authored using a system based on “MetPy”, which is a local utility unsuitable for distribution.