Running an ESGF node for CORDEX

An outline of resource levels required, based on experience within the IS-ENES and IS-ENES2 projects.

Background

This document describes the effort and resources needed to run and ESGF data node, publish data, provide a search interface, register users and provide some support. ESGF also provides the option of providing server side computational resources, but the resources required to support this operationally are not included here.

The ESGF software is available here http://esgf.org/releases . The current release (Nov. 2013) is 1.5. Experience within IS-ENES2 suggests that the installation script for version 1.5 will not work for installation from scratch, but will work if version 1.4 is installed first. Version 1.6 is expected soon, and it is hoped that some problems in the installation script will be addressed.

Hardware

All the functionality described above can be run on a single machine, but splitting the search interface from the data node creates a more robust system.

The resources needed depend on the volume of data to be served, the number of catalogues and, of course, the number of concurrent user sessions. Requirements can also vary between different versions of the system software.

Some indicative numbers based in experience with European data nodes. There are many configuration options, and we have not yet done a systematic evaluation. The hardware costs are generally small compared to the manpower costs.

Data node: 1-2Gb of memory per 1000 catalogues published.

Search interface: experience at DKRZ: with 48Gbyte of memory need to restrict number of concurrent searches to 10.

People

Software installation and upgrades

Has complex set of dependencies; correct configuration of network; estimated resource: 0.25 full-time equivalent (FTE).

The system is open source – the resolution of known problems is limited by the availability of people to work on the software. To contribute to the development effort (optional), a further 0.75 FTE would be needed. Contributing to the development effort will give you a stronger voice in determining the direction of future developments.

Publishing data

When data is ready for publication, it needs to be placed in the correct file structure and “published” to the system. Within the CMIP5 project many problems occurred during this process, and fixing

these problems required significant resource. During publication, the file level catalogue information is compiled. Estimated resource for CORDEX: 0.25 FTE.

Preparing data

The ESGF system is fairly flexible about the quality of data published, but the data will be of little value to users if it does not have a high standard of file metadata. This process can be time consuming, especially for an institute with no previous experience of dealing with standardisation of climate model data. Recommended resource 0.75 FTE.