Session: C21 Poster: T258A

Optimized and scalable climate data services

John del Corral[†]; Benno Blumenthal; Michael Bell; Remi Cousin; Haibo Liu
[†] Columbia University, USA
Leading author: jdcorral@iri.columbia.edu

The IRI Climate Data Library (http://iridl.ldeo.columbia.edu) is designed to optimize the display, analysis, and retrieval of climate datasets. These datasets range from simple station observations, to multi-ensemble climate model results, to high-resolution satellite measurements, to GIS representations of geographic entities. These datasets are represented in a consistent multi-dimensional framework. Station observations can easily be compared with climate model results, and satellite measurements. Gridded data can be spatially averaged over discrete geographic entities. The Climate Data Library is accessible with a browser connected to the internet. The data selection, processing, and analysis are performed by the Climate Data Library servers. The resulting images or data files are sent back to the client's desktop. This model optimizes the use of internet bandwidth. The software required is open source and the hardware is readily available off the shelf. The Climate Data Library only reads the data from the archive that the user requests in an analysis, view, or data download, rather than the entire dataset. After the user has viewed the data and is focused on a subset, that subset (or any other analysis) can be delivered to the user's desktop in multiple formats. Some are self-describing such as netCDF or OpenDAP. Others are convenient for users of particular languages or applications. Standard data formats like ascii, binary, csv, and tsv can also be delivered. The views of the data can be exported in multiple image formats (PDF, postscript, GIF, and JPG) and as services (WMS and KML). Online data analysis of climate data is one of the Climate Data Library's strengths. Analytical functions frequently used by climate and weather researchers are built into the software. We have used this to create tailored analyses in the form of climate maprooms, where data and displays are chosen for a particular audience. The IRI Climate Data Library consists of caching servers and compute servers that are load balanced with failover. Each compute server has a copy of the SQL database that stores Open Geospatial Consortium (OGC) GIS objects and smaller datasets. The IRI Climate Data Library has features that support federated data services. The Data Library implements Thematic Realtime Environmental Distributed Data Services (THREDDS) cataloging. It supports the OpenDAP protocol for data transfer. The Climate Data Library uses THREDDS and OpenDAP as both client and server to organize, access, and display datasets from multiple OpenDAP servers. THREDDS catalog entries can come from IRI Climate Data Library mirror sites. Mirror sites of the Climate Data Library are deployed in regions of the world where local internet access may be intermittent and/or where local institutions desire to add their own climate datasets. The IRI has developed a portable version of the Climate Data Library for training purposes and rapid mirror site deployment. Summary Our system addresses the issues of format compatibility, optimized access to data, interoperability, and scalability of climate data services. Our data is seamlessly available to browsers, programs, and applications, particularly applications that use OpenDAP, WMS, and KML protocols.