

Enabling Climate Scientists to Access Observational Data

David Woollard¹, Chris Mattmann, Amy Braverman, Rob Raskin, Dan Crichton
NASA's Jet Propulsion Laboratory
California Institute of Technology

One research challenge facing both software engineers and climate researchers today is the utilization of large-scale observational datasets to rectify climate models. At NASA's Jet Propulsion Laboratory, we have spend over 40 years developing Earth-orbiting spacecraft capable of making scientific measurements salient to climate modeling, including temperature, atmospheric particulate and water vapor concentration, cloud coverage, and weather phenomena to name a few.

The challenges faced by the climate modeling community are to:

- Locate and gain access to relevant datasets
- Transfer large-scale datasets (often in the *giga-* to *peta-byte* range)
- Translate disparate data and metadata formats
- Understand sampling biases and other instrument nuances

A recent research and technology development initiative at JPL, known as the Climate Data eXchange, or CDX [1,2], has been tasked with developing software technology to help climate scientists begin to overcome these technological and scientific hurdles. CDX is a data grid that enables access to disparate data, both organizationally and geographically distributed.

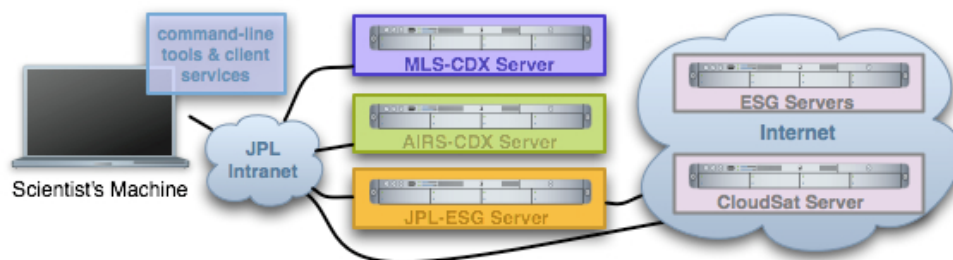


Figure 1: Current CDX hardware configuration.

Using OODT [3, 4], a data grid technology developed at JPL, we have developed Product and Profile Servers that map mission-specific metadata to a common ontology and allow internet-based access to data housed in mission archives via web services. Further, leveraging OODT, we have successfully integrated several technologies including OPeNDAP, and custom mission software toolkits, to provide data subsetting across the climate mission data warehouses shown in Fig. 1.

Additionally, CDX provides conversion services that translate data and metadata to the NetCDF format, allowing for easy manipulation by the climate modeling community. Further distribution of these data sets is facilitated by CDX's partnership with the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the Earth System Grid [5]. Currently, CDX is serving AIRS, MLS, and MISR observational data sets and in the next quarter, we expect to bring CloudSat data online as well.

In the next year, we plan to provide ontology-driven, cross-mission product search capabilities as well as a number of command-line tools and web services for integrating into existing modeling and model verification efforts both at JPL and also at Lawrence Livermore National Laboratory.

References:

- [1] CDX Website. <http://cdx.jpl.nasa.gov>
- [2] D. Crichton, et. al. Facilitating Climate Modeling Research and Analysis via the Climate Data eXchange. *Proc of the Workshop on Global Organization for Earth System Science Portals (GO-ESSP)*, Seattle, WA, 2008.
- [3] OODT Website. <http://oodt.jpl.nasa.gov>
- [4] C. Mattmann, D. Crichton, N. Medvidovic and S. Hughes. A Software Architecture-Based Framework for Highly Distributed and Data Intensive Scientific Applications. *Proc. ICSE* pp. 721-730, Shanghai, China, May 20th-28th, 2006.
- [5] Earth system Grid Website. <http://www.earthsystemgrid.org/>

¹ Corresponding author. E-mail address: woollard@jpl.nasa.gov