

# The Climate Data eXchange: Bringing NASA's Observational Data to the IPCC Community

Amy Braverman, Dan Crichton, Chris Mattman, Rob Raskin and Mike Gunson  
Jet Propulsion Laboratory, California Institute of Technology

The Climate Data eXchange (CDX) is a new initiative at JPL designed to streamline access to and analysis of NASA's observational data sets, with a particular eye towards the role these observations can play in the next IPCC report. Multi-model ensembles are typically used to assess the reliability of climate projections by equating the variance across ensemble members to uncertainty in future climate conditions. Increasingly, the community has recognized that models are not independent of one another, so simple variance calculations or histograms, in which each model contributes equally, may not be appropriate. Different models have different strengths, and their output should be weighted differently depending the questions being asked. Comparisons of model predictions to observations over the period of the observational record provide a way to determine these weights, but significant challenges exist in using NASA data for this purpose. Those challenges arise from the heterogeneity and distributed nature of the data, as well as from the differences among access and manipulation technologies (sophisticated web services such as OPeNDAP all the way to simple HTML ordering carts), and the variety of formats. Most importantly, using these data sources effectively requires understanding of complex methodologies and assumptions used to create them.

CDX addresses these issues with an architectural paradigm that recognizes that data access and data analysis intimately are intertwined. Significant efficiency gains can be achieved by deploying certain "building-block" services in a carefully choreographed strategy specifically designed for observational data. Our guiding architectural principles are 1) standardize interfaces to heterogeneous back-end systems such as science computing facilities using Object Oriented Data Technology, 2) encode complex information about the data (e.g. sampling patterns) directly into data access services, and 3) optimize the architectural topology of the CDX network by quantifying data movement patterns through observational studies and controlled experiments. In this talk we describe the CDX project, its motivation, enabling technologies, and relationship to the Earth System Grid.