Data access and analysis with distributed federated data servers in climateprediction.net

Neil Massey, Tolu Aina, Myles Allen, Carl Christensen, David Frame, Daniel Goodman, Jamie Kettleborough, Andrew Martin, Stephen Pascoe, David Stainforth
Computing Laboratory, Oxford University, neil.massey@comlab.ox.ac.uk

climateprediction.net is a large public resource distributed scientific computing project. Members of the public download and run a full-scale climate model, donate their computing time to a large perturbed physics ensemble experiment to forecast the climate in the 21st century and submit their results back to the project. The amount of data generated is large, in the order of terrabytes. Access and analysis of the data is further complicated by the reliance on donated, distributed, federated data servers. This talk will present the problems encountered when the data required for even a simple analysis is spread across several servers and how webservice technology can be used; how legacy climate analysis tools can be wrappered with webservices, specifically using CDAT within a service oriented architecture; how different user interfaces with varying levels of complexity and flexibility can be presented to the application scientists, how using existing web technologies such as HTTP, SOAP, XML, HTML and CGI can engender the reuse of code across interfaces; and how application scientists can be notified of their analysis' progress and results in an asynchronous architecture.