A Grid Metadata Service for Earth and Environmental Sciences

Sandro Fiore, Alessandro Negro, and Giovanni Aloisio

Euro-Mediterranean Centre for Climate Change and University of Salento, Lecce, Italy (sandro.fiore@unisalento.it, alessandro.negro@unisalento.it,giovanni.aloisio@unisalento.it)

Critical challenges for climate modeling researchers are strongly connected with the increasingly complex simulation models and the huge quantities of produced datasets. Future trends in climate modeling will only increase computational and storage requirements. For this reason the ability to transparently access to both computational and data resources for large-scale complex climate simulations must be considered as a key requirement for Earth Science and Environmental distributed systems.

From the data management perspective (i) the quantity of data will continuously increases, (ii) data will become more and more distributed and widespread, (iii) data sharing/federation will represent a key challenging issue among different sites distributed worldwide, (iv) the potential community of users (large and heterogeneous) will be interested in discovery experimental results, searching of metadata, browsing collections of files, compare different results, display output, etc.;

A key element to carry out data search and discovery, manage and access huge and distributed amount of data is the metadata handling framework.

What we propose for the management of distributed datasets is the GRelC service (a data grid solution focusing on metadata management). Despite the classical approaches, the proposed data-grid solution is able to address scalability, transparency, security and efficiency and interoperability.

The GRelC service we propose is able to provide access to metadata stored in different and widespread data sources (relational databases running on top of MySQL, Oracle, DB2, etc. leveraging SQL as query language, as well as XML databases – XIndice, eXist, and libxml2 based documents, adopting either XPath or XQuery) providing a strong data virtualization layer in a grid environment.

Such a technological solution for distributed metadata management leverages on well known adopted standards (W3C, OASIS, etc.); (ii) supports role-based management (based on VOMS), which increases flexibility and scalability; (iii) provides full support for Grid Security Infrastructure, which means (authorization, mutual authentication, data integrity, data confidentiality and delegation); (iv) is compatible with existing grid middleware such as gLite and Globus and finally (v) is currently adopted at the Euro-Mediterranean Centre for Climate Change (CMCC - Italy) to manage the entire CMCC data production activity as well as in the international Climate-G testbed.