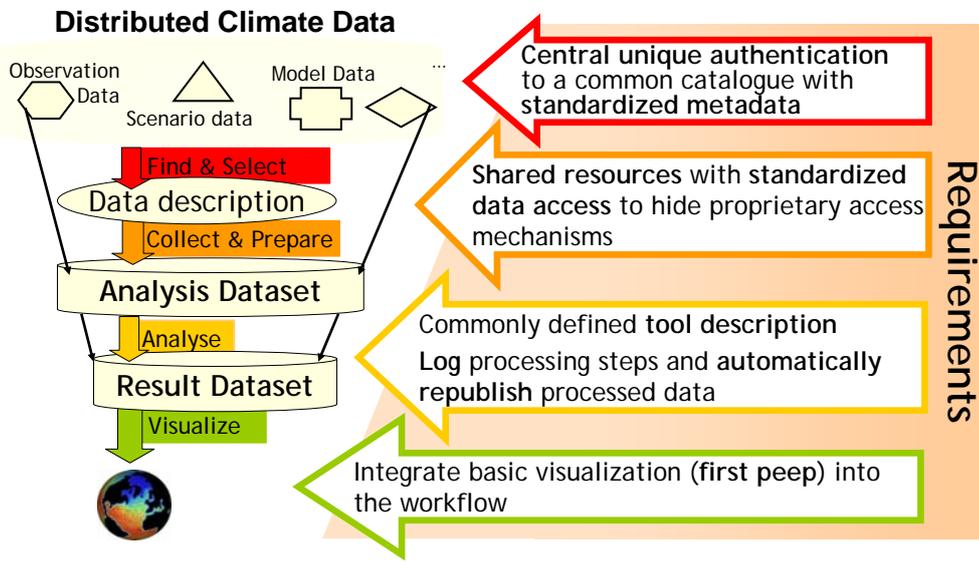




Intelligent, distributed climate data management

K. Ronneberger (DKRZ), S. Kindermann (DKRZ), T. Brücher (ZAIK), M. Stockhause (MPI-M), H. Ramthun (M&D), and B. Bräuer (AWI)

Motivation

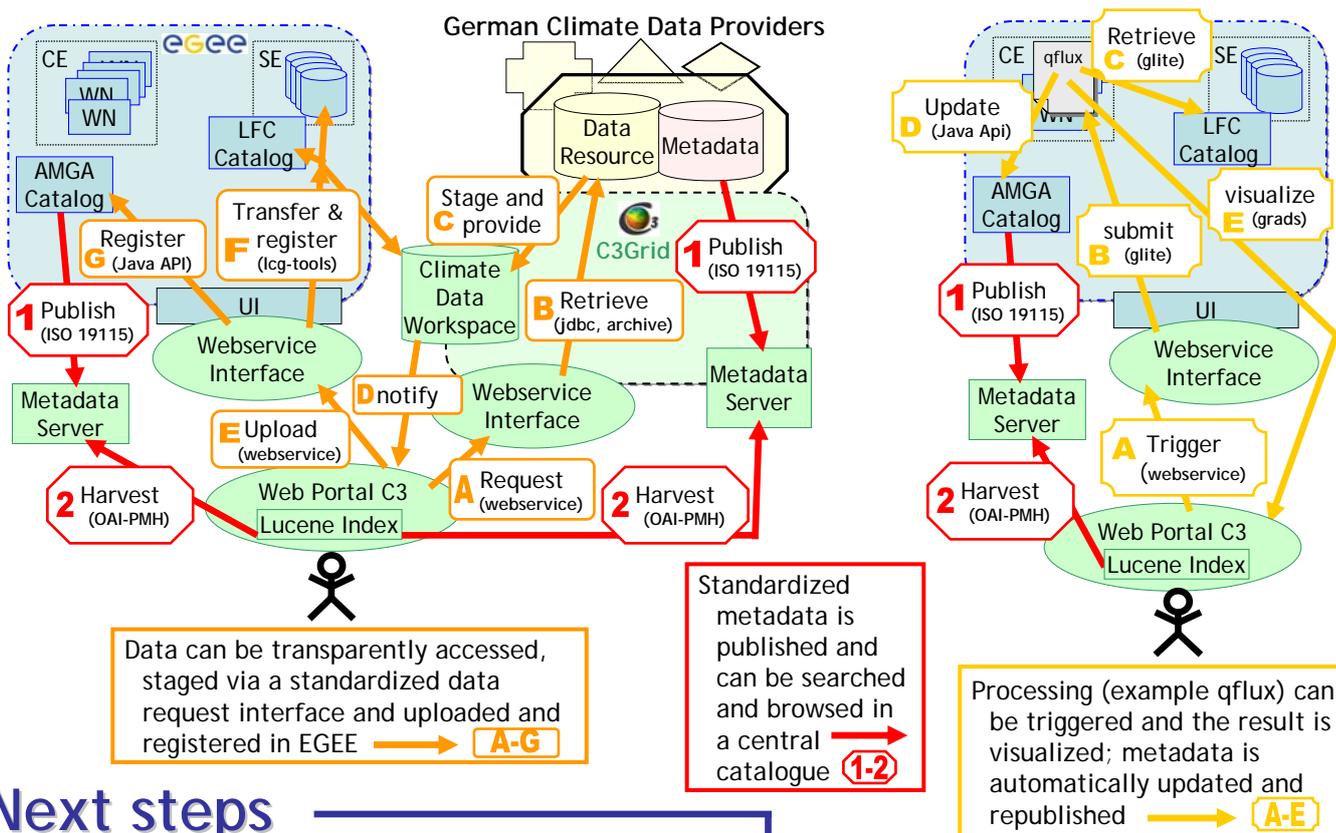


Grid solution

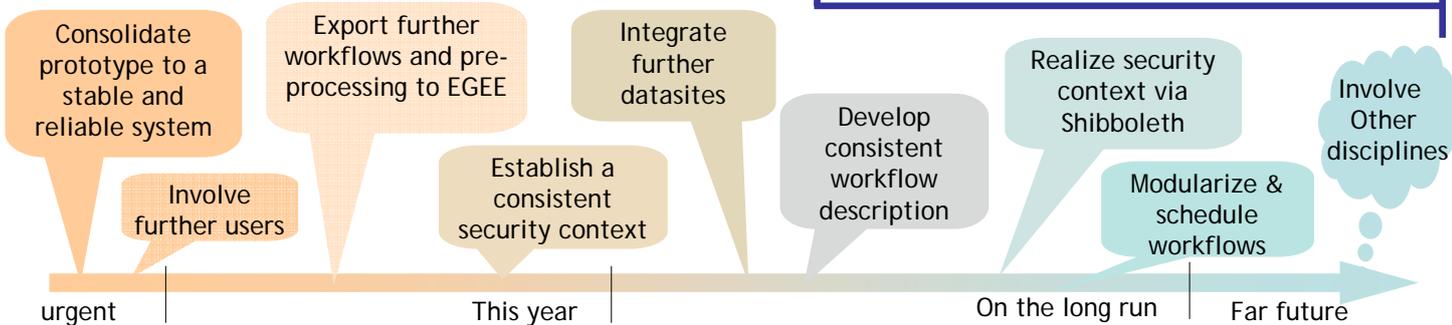
Background

In collaboration with the German C3Grid Project* a system has been developed to ease and accelerate climate data workflows. The system is built modular and based on international standards to be expandable by further data sites, partners and disciplines

*[http://www.c3grid.de]



Next steps



Introduction

Climate research is generally very data intense. Data from observation, analysis and output of climate simulations are traditionally stored in large archives and central databases. They vary highly in quality and in accessibility. Thus, searching, finding, and retrieving the data is often highly inefficient.

The presented infrastructure aims to enable a coherent and intelligent data discovery and data access for climate and earthsystem scientists. A central portal offers a unique entrance point to the established central metadata catalogue. The ISO standard 19115 is adopted to describe the characteristics of geographic data, required for their discovery (discovery metadata) and for their usage (use metadata). Webservice technology is used to hide the details of heterogeneous data access mechanisms and preprocessing implementations.

The components were developed in the German C3Grid project (<http://www.c3-grid.de>, part of the German E-Science initiative D-Grid). In collaboration with C3Grid the system was adapted and extended to enable transparent exchange of data among EGEE and traditional German climate data providers and to trigger an example workflow to process data in EGEE. The Amga catalog is used as a central instance to receive and update the necessary runtime information of data produced or altered in the processing jobs. It is extended by an interface to enable automatic harvesting and republishing of the resulting metadata to the central Webportal in ISO format.

The commitment to international standards and the modular character of the approach facilitates the expandability of the infrastructure as well as the interoperability with international partners and other communities.

Added value

Currently, to find, retrieve, and process climate data mostly complex individual solutions are used. Processed data is commonly stored locally and undocumented. Thus, identical analyses are redone by various scientists. The developed system eases and accelerates the daily work of earthsystem scientists, leaving more time and capacity for the actual science:

Enabling searching and browsing of the various data in a central catalog according to content, quality and processing history eases the discovery of data. An intelligent, transparent data access simplifies the data retrieval. Selectable basic processing options and an automatic republishing of the processed data supports the daily workflows of climate scientists and facilitate further processing or usage of the results. The functionality of some of the components, as well as the shared use of resources necessary to effectively use the infrastructure, can only be realized by means of grid technology. By integrating EGEE into this infrastructure, a proven and mature grid infrastructure enters the system, offering functionalities and services not or not yet implemented in C3Grid. Thus, it does not only advance the interoperability of the two grid projects, but also stimulates synergy effects by combining the strengths of the two systems. For instance the direct republishing of a processed file is possible only in EGEE so far, because EGEE offers the necessary storage and service facilities to upload, register and retrieve files.

EGEE Experience Summary

The presented system is a prototype and still under construction. Thus, so far no detailed experience with the operability on EGEE could be gathered. Yet, it originates from, and builds on the experiences collected in the earthsystem community over the past years. So far, using EGEE as a common platform for data processing in climate or earthsystem science was hampered mainly by insufficient data management: accessing data in EGEE from traditional storage systems, and managing, browsing and updating related metadata was possible only with tedious, individual solutions. The developed system offers a smooth solution for this problem. At the same time, by building on the C3Grid - a community driven project - it directly connects to the climate community and provides a comfortable interface to try out EGEE without getting into the complexity of EGEE usage.

Contact:

ronneberger@dkrz.de
kindermann@dkrz.de