Geophysical Research Abstracts,
Vol. 10, EGU2008-A-06342, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-06342
EGU General Assembly 2008

A service oriented framework for automatic discovery metadata management in earth science workflows

S. Kindermann, K. Ronneberger (1)

(1) German Climate Computing Centre, DKRZ (kindermann@dkrz.de)

Standardized discovery metadata is essential to search and discover heterogeneous data products in earth sciences. A rapidly growing amount of data at earth science data centers around the world is described using ISO 19115/ISO 19139 discovery metadata profiles. This data is input to specific scientific data analysis workflows to produce derived data products, which are oftenly not annotated by formal metadata descriptions. Data reuse is hindered by the fact that no automatic metadata generation facility is integrated into scientific workflows. Publication and archiving of workflow output data oftenly requires by hand annotation with discovery information.

We present a reusable automatic discovery metadata generation and metadata management service, which can be easily integrated into scientific workflows. The metadata management service is based on web and grid standards and hides most of the complexity of handling ISO 19139 discovery information by providing a simple update web service interface. Updates are quality checked with respect to community profiles of the ISO 19139 standard using an adaptable set of schematron constraints. The checked metadata files are managed in an XML database and published based on an OAI-PMH interface, which is used by portals for metadata harvesting. Secure integration into e-science (e.g. GRID-) infrastructures is possible by adopting the OGSA-DAI wrapping technology.

Workflows operating on ISO 19139 described input data can thus be easily adapted to automatically provide ISO discovery metadata for their output data. This includes data provenance information which is oftenly needed to allow data reuse. The described service is currently being integrated into the Collaborative Climate Community Data

and Processing Grid (C3Grid) data discovery, data access and processing service architecture. We present the design and implementation of the metadata service and discuss first experiences with application scenarios in other e-science environments.