Geophysical Research Abstracts,
Vol. 10, EGU2008-A-04965, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-04965
EGU General Assembly 2008

Metadata organization and lifecycle in earth system science (C3Grid)

M. Stockhause (1), S. Kindermann (2) and S. Plantikow (3)

(1) Max Planck Institute for Meteorology (MPI-M), (2) German Climate Center (DKRZ), (3) Zuse Institut Berlin (ZIB)

Earth system research and the provision of a growing community of non-scientists with derived climate parameters depend on the assimilation of huge data amounts from different origins (numerical models, satellites, observations, research campaigns). For an effective and sustainable usage the continuous annotation of these data and their products with metadata throughout their whole lifecycles is essential.

Apart from their traditional bibliographical information purpose (content and citation), modern metadata have to provide technical information (data identification and access) and provenance information (data history and quality) as well. Furthermre, since the produced data amount in Earth System sciences increase rapidly, data and metadata organization become more and more important. Possible kinds of interdependencies are:

- hierarchical: information about parent-child or superset-subset relations,
- historical: provenance or lineage information, and
- topical: aggregation information.

For grids focussing on data performance like the C3-Grid (Collaborative Climate Community Data and Processing Grid) a uniform metadata description model and a thourough documentation of (meta)data relations are of special importance for data reusability (discoverage and evaluation). The C3-Grid community has established a metadata profile derived from the ISO guidelines 19115/19139 using the CF (Climate

and Forecast) standard names convention and developed a set of tools for metadata creation and handling.

This talk provides a concise description of the metadata handling throughout the whole lifecycle of the data based on a concrete usage szenario. The lifecycle starts with metadata creation and publication, followed by alternation and republication (during the application of a scientific workflow) and (possibly) ends up with long-term archiving. The application of different metadata tools is shown as well as the logic behind the structuring of metadata reflecting interdependencies like hierarchy or data provenance information.