Integration of external metadata into the Earth System Grid Federation (ESGF)

Katharina Berger (1), Guillaume Levavasseur (2), Martina Stockhause (1), and Michael Lautenschlager (1)

(1) German Climate Computing Center (DKRZ), Hamburg, Germany (stockhause@dkrz.de), (2) Institute Pierre Simon Laplace (IPSL), Paris, France

International projects with high volume data usually disseminate their data in a federated data infrastructure, e.g. the Earth System Grid Federation (ESGF). The ESGF aims to make the geographically distributed data seamlessly discoverable and accessible. Additional data-related information is currently collected and stored in separate repositories by each data provider. This scattered and useful information is not or only partly available for ESGF users. Examples for such additional information systems are ES-DOC/metafor for model and simulation information, IPSL's versioning information, CHARMe for user annotations, DKRZ's quality information and data citation information.

The ESGF Quality Control working team (esgf-qcwt) aims to integrate these valuable pieces of additional information into the ESGF in order to make them available to users and data archive managers by
(i) integrating external information into ESGF portal,
(ii) integrating links to external information objects into the ESGF metadata index, e.g. by the use of PIDs (Persistent IDentifiers), and
(iii) automating the collection of external information during the ESGF data publication process.

For the sixth phase of CMIP (*Coupled Model Intercomparison Project*), the ESGF metadata index is to be enriched by additional information on data citation, file version, etc. This information will support users directly and can be automatically exploited by higher level services (human and machine readability).

DKRZ: http://www.dkrz.de
IPSL: http://www.ipsl.fr/
ESGF: http://esgf.llnl.gov/
IS-ENES: http://verc.enes.org/
CMIP: http://cmip-pcmdi.llnl.gov/