



Contribution ID: 20

Type: **Oral**

## Climate data storage in e-INIS

*Tuesday, 13 April 2010 16:40 (20 minutes)*

We describe the federated national datastore activity of the e-INIS project, aimed at building a sustainable national e-Infrastructure for the Irish academic research community and how the CMIP5 project is using the datastore to meet their storage requirements. The datastore builds upon existing infrastructure and services, including Grid-Ireland, the National Grid Initiative. Read access to the data is to be offered to international researchers using GeoNetwork and OPeNDAP, requiring that Grid technology be interfaced with community technologies using e-INIS's bridge servers.

### Detailed analysis

Using Grid technologies such as LFC and DPM allows distributed storage and replication of data on inexpensive hardware. The CMIP5 project requires the storage of approx. 100,000 netCDF files on 198 TB of storage, from mid-2009 to 2011 for the IPCC AR5 project. This data is being generated by EC-Earth climate model runs over November 2009 - December 2010. Data will be made available to other scientists and the public via http and OPeNDAP in 2010-2011, using a GeoNetwork catalog server. OPeNDAP enables optimised access to netCDF datasets, making it possible to download parts of files and access metadata for files. GeoNetwork provides a federated catalog service to ISO 19115 standards, fulfilling the INSPIRE directive requirements for public access, and enables data discovery via a web portal and clients such as Google Earth, NASA's Worldwind and its own GeoNetwork utilities. The metadata is then compliant with METAFOR conventions, for intercomparison with other climate models in the CMIP5 intercomparison as part of the IPCC Assessment Report 5. A subset of the data generated in Ireland is forwarded to other centres for comparison, but the majority is available only via e-INIS.

### Conclusions and Future Work

Bridge servers are being used to interface grid technology with servers that comply with user-community specified standards. The CMIP5 work is testing this with a reasonably large quantity of data over a period of time. As the project progresses it is envisaged that the various institutions involved will acquire 10Gb/s paths. The main technical challenges are establishing an integration in the bridge layer and scaling in step with both the size of dataset and speed at which it is being accessed.

### Impact

Integration of gLite and OPeNDAP in the e-INIS datastore is a large-scale demonstration of use of the e-INIS bridge servers to interface the existing back-end storage with community defined protocols. OPeNDAP is a very widely used standard but is better suited to access of subsets of large data-sets than to the bulk transport of data. The gLite middleware is better suited to secure transport of large quantities of data but does not provide the introspection features of OPeNDAP or comply with the relevant international standards for this climate modelling community of users. Other groups are looking at integrating gLite and OPeNDAP but we present a generic architecture that can be applied to other services, illustrated by the specific example of CMIP5 using this architecture to manage scientific data of international interest. The resultant system uses

grid technology for back-end storage and to manage writing of data while providing a read-only front-end with a different security model that is compliant with international standards.

## **Keywords**

Grid, Storage, Data, Climate Model,

## **URL for further information**

[http://www.ichec.ie/research/met\\_eireann#cmip5](http://www.ichec.ie/research/met_eireann#cmip5)

**Primary author:** Dr QUIGLEY, Geoff (Trinity College Dublin)

**Co-authors:** Mr MCKINSTRY, Alastair (Irish Centre For High End Computing); Dr COGHLAN, Brian (Trinity College Dublin); Dr RYAN, John (Trinity College Dublin); Dr ROCHFORD, Keith (Dublin Institute of Advanced Studies)

**Presenter:** Dr QUIGLEY, Geoff (Trinity College Dublin)

**Session Classification:** Earth Science

**Track Classification:** Scientific results obtained using distributed computing technologies