Climate data integration and user access

DKRZ - DATA INTENSIVE CLIMATE SCIENCE

DKRZ: A vital data service provider for the networked climate community

Confronted with the climate data deluge, climate scientists around the world need a new generation of climate data services.

Thus DKRZ actively participates in national, European and global e-Science projects which build integrated climate data handling infrastructures for the future. Fundamental components of these infrastructures include:

- **Portals:** providing users access to data discovery, search and access services as well as data processing services.
- **Data Nodes:** Data centers acting as data nodes, expose their data and associated metadata via standardised interfaces.
- **Security:** Providing distributed authentication and authorization services.
- **Data and Job management middleware:** providing services to move, replicate and process data in a distributed environment.
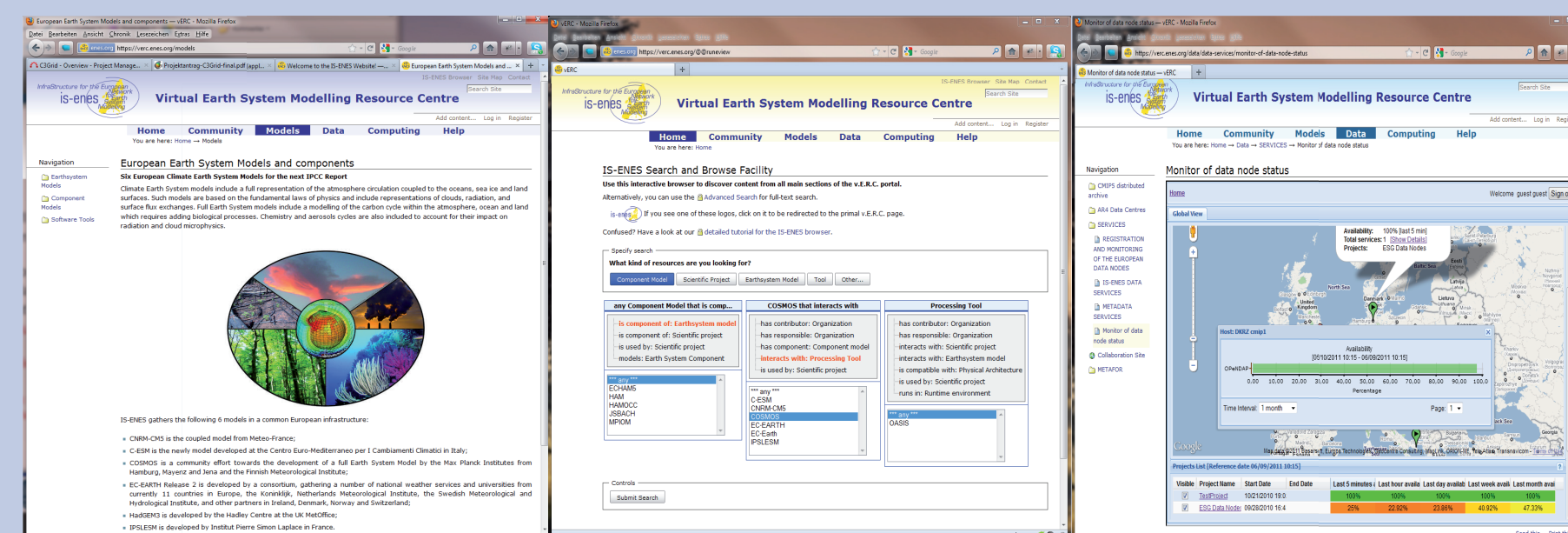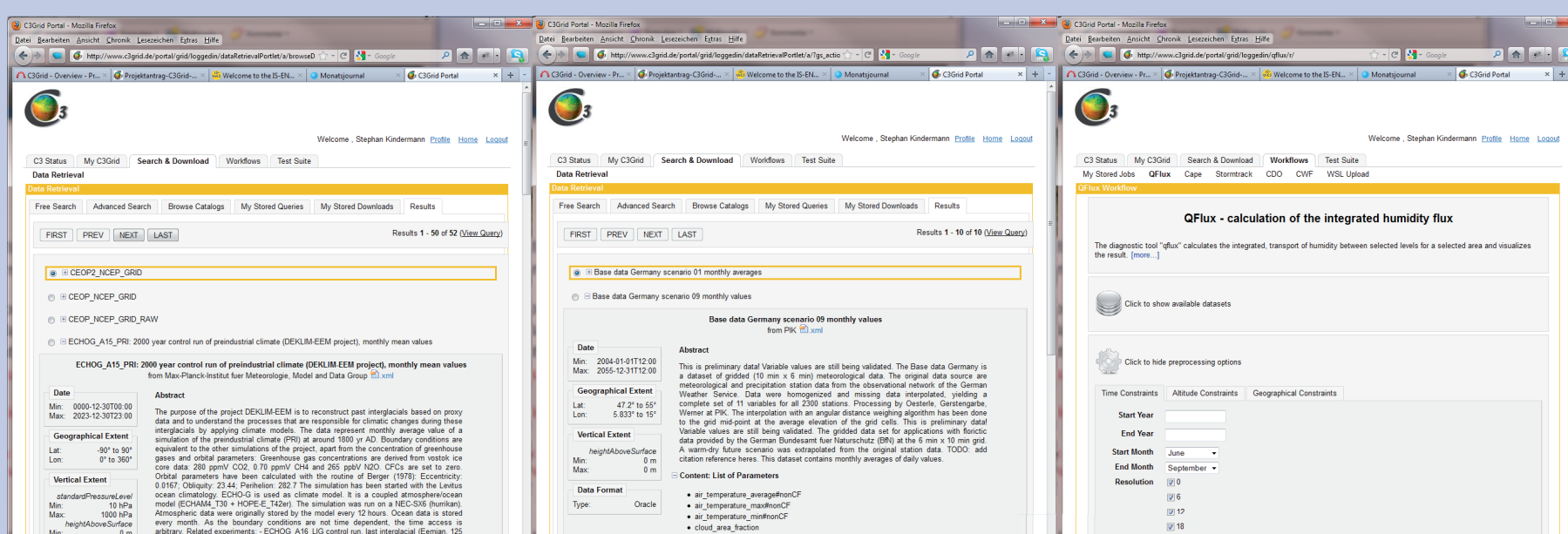
data browse   search   workflow processing   information   cross-linking   resources

C3Grid

In Germany climate community service providers and universities joined in 2005 to build up an integrated climate data access and processing infrastructure: C3Grid. In 2009 the follow up project C3-INAD started, which will provide production level services to the community. This includes developments to make C3–INAD interoperable with the European IS–ENES and the global ESGF infrastructures. DKRZ leads this integration effort and will establish a sustainable support center for C3Grid.
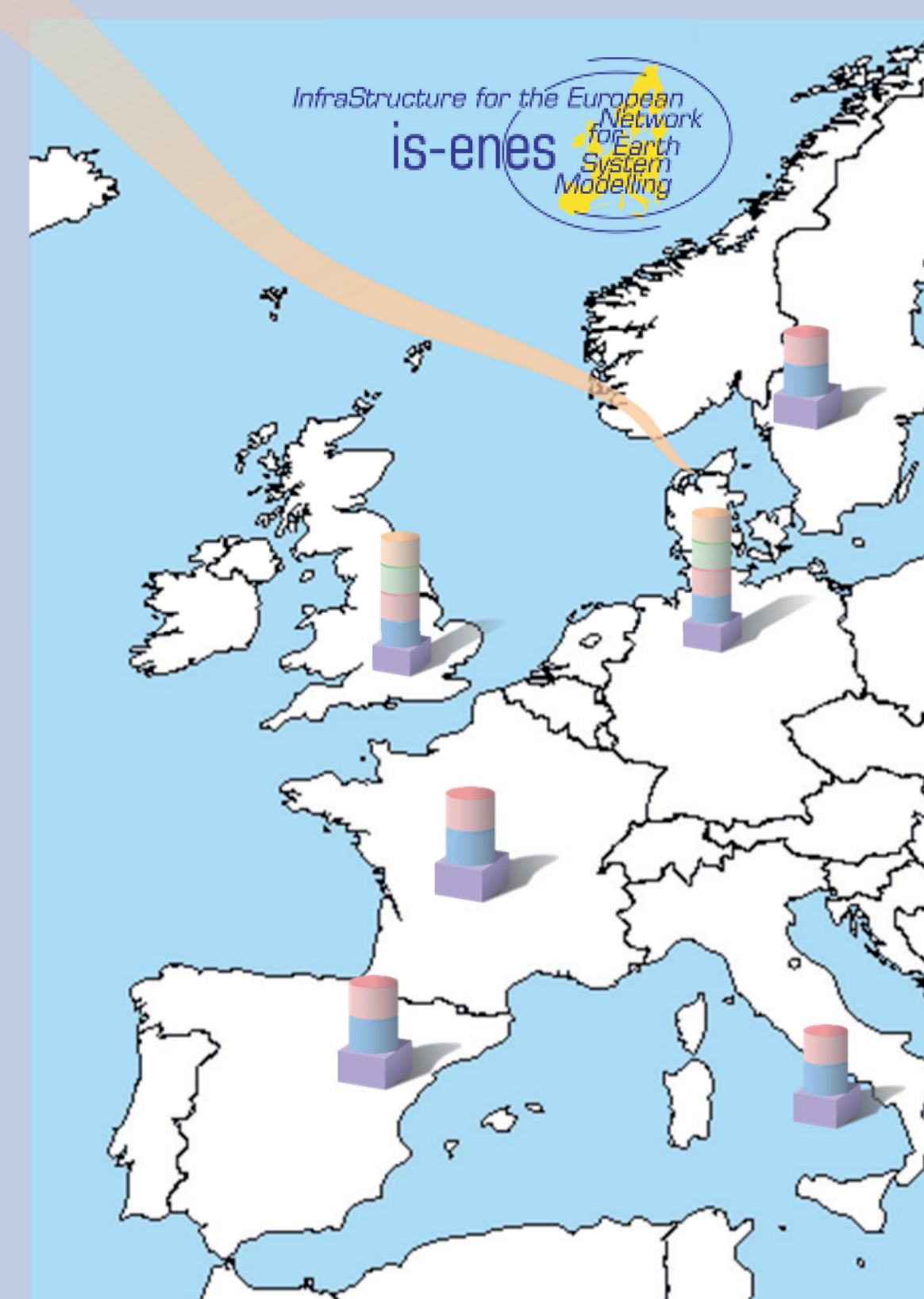
ESGF and CMIP5/IPCC AR4

Key driver of current international infrastructure building efforts is the 5th coupled model intercomparison project (CMIP5), which will provide key input for the next IPCC report. Some numbers illustrate the challenge behind:

- ~20 modeling centers around the world
- Produce ~10`s of PetaBytes of output
- Based on ~60 different experiments
- ~3.5 PetaByte, requested' data is collected centrally
- >1 PetaByte of data is replicated between PCMDI, BADC and WDCC
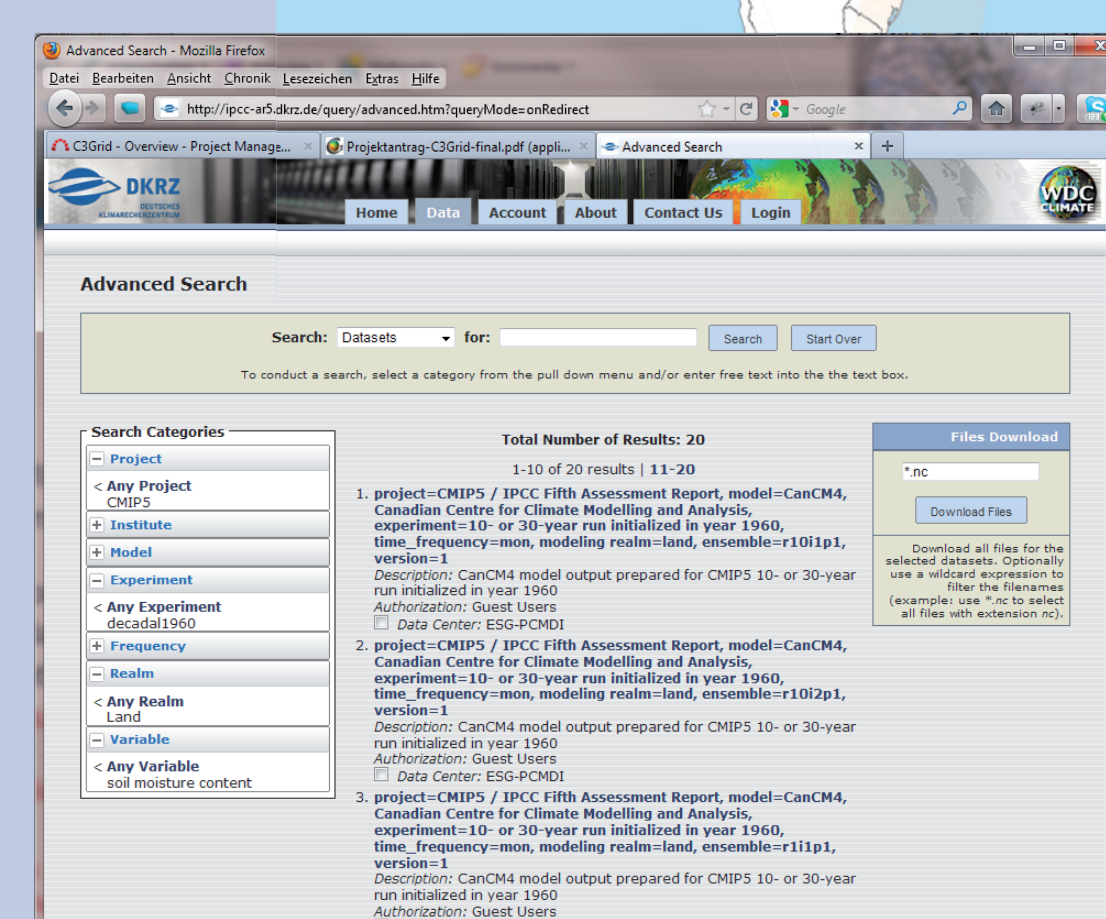
data browse / search / access

The distributed data infrastructure is based on open source software developed within the Earth System Grid Federation (ESGF) initiative.

The **ESGF initiative** is developing a modular, layered software infrastructure to enable distributed data management of Petabytes of climate data. Data providers publishing their data, only have to install the base layer providing data and metadata access. Service providers add data analysis and security features. Portals additionally deploy a flexible search service layer.
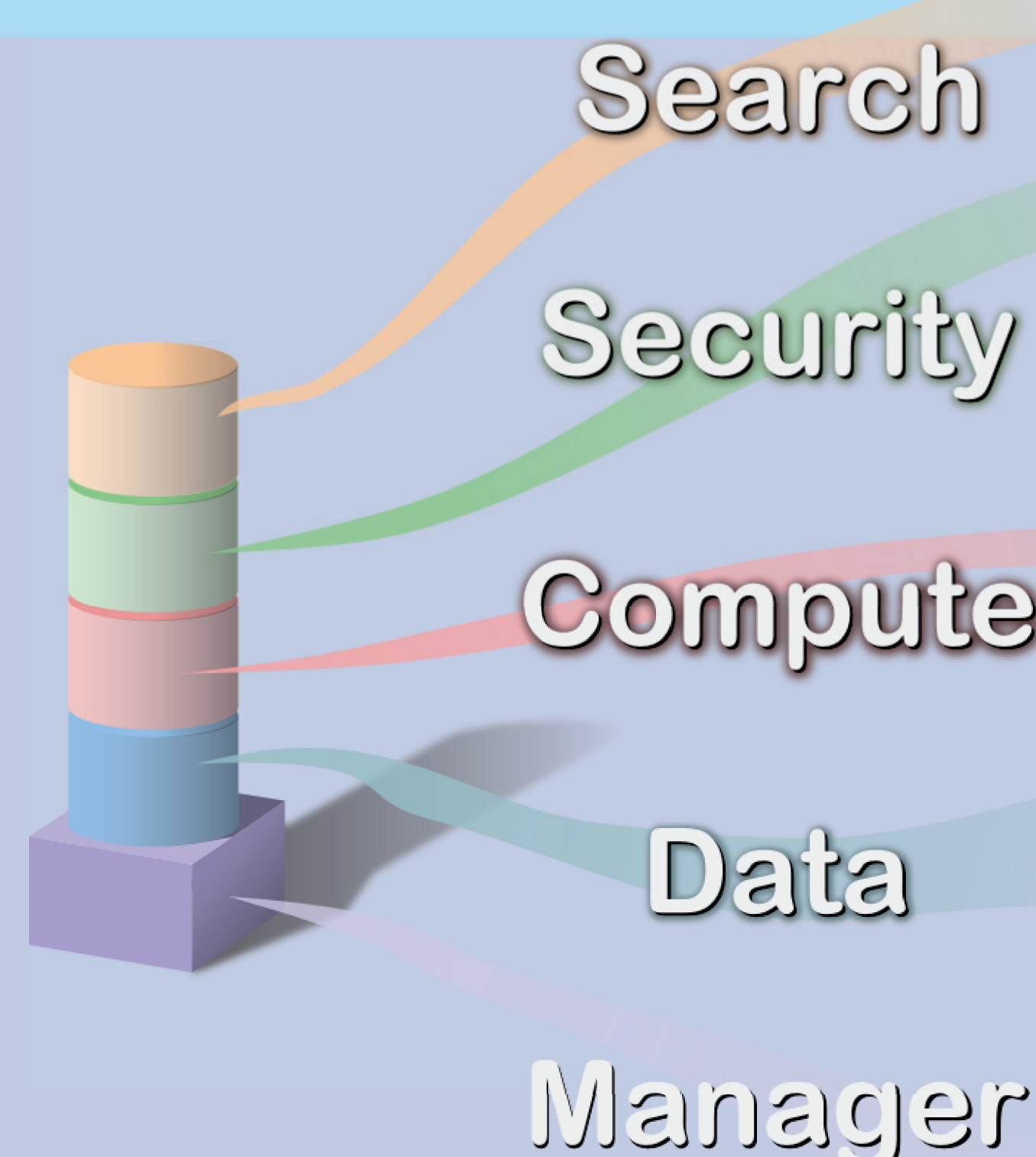
IS-ENES

Within the 7th European framework programme, the IS-ENES (Infrastructure for the European Network for Earth System Modeling) project is funded, gathering 45 institutions comprising the European climate and Earth system modelling community. DKRZ develops and hosts the IS-ENES portal, integrating services and information from the community.

The IS-ENES distributed data network is based on ESGF, extending it to include new processing capabilities as well as new data sources (e.g. regional climate modeling results from CORDEX).

The British athmospheric data center BADC and the World data center for climate at DKRZ act as key data service providers within this network.

IS-ENES is closely collaborating with the **Metafor** FP7 project, which develops the information model and metadata framework to describe climate models, their configuration as well as model data.

References:

ESGF: http://www.esgf.org

IS-ENES portal at DKRZ: http://verc.enes.org

Metafor: http://www.metafor.eu

The CMIP5 Gateway at DKRZ: http://ipcc-ar5.dkrz.de

C3-Grid: http://www.c3grid.de

Search
Security
Compute
Data
Manager

www.dkrz.de

Poster 4

DKRZ

DEUTSCHES KLIMARECHENZENTRUM