Enhancing data connectivity of modelling results obtained in the TFHTAP multi-model assessment study

Martin Schultz, Michael Decker, Sebastian Lührs, Sabine Schröder
Forschungszentrum Jülich, Germany
Background

- Jülich hosts the central HTAP data server (htap.icg.kfa-juelich.de) – presently 58 user accounts
- HTAP multi-model study results are accessible only via direct login, i.e. you must have an account and know how to access netcdf files
- OGC (Open Geospatial Consortium) has defined protocols for “unattended“ data delivery via web-URLs
- WUSTL has created a software package that implements (parts of) the OGC concept
- EPA and DE-ENV have an interest to make HTAP results available in an open and unrestricted manner
- Scientific users could benefit from online model intercomparison and validation tools
Objectives of „HTAP server project“

June – Sep 2009

- Install WUSTL prototype software to establish connectivity
- Screen and reformat existing data sets so that they can be made available
- Compile data catalogues
- Develop a demonstration web application
- Analyse performance of data service with respect to hardware and software efficiency
**Status**

- OWS software installed and running
- First basic scripts installed to generate data catalogues
- First basic web interface created
- Several problems with file formats, domain description, etc.
Current issues

- data attributes not always compatible with OWS (missing_axes, unknown_axes, ambiguous_axes, missing_time_units, invalid_time_format, invalid_chars)
  → reprocessing of netcdf files needed
- irregular time coordinate (e.g. monthly means) cannot be extracted into catalogue
- model level extraction and interpolation
- geographic domain for Europe (350 E – 60 E)
- exclude list for variables that are of no interest to user
- list of variables to be automatically included in data file
Next steps

- define web interface layout and functionality
- customize template xml files
- clarify handling of coordinate data (usage of bounds information, ...)
- map data files -> providers
- discuss cache handling for both ows server and frontend: file lifetime, hard quotas and what to do when the cache is full
- clean data files
- extend owsadmin to allow for irregular time series and level extraction
- assess efficiency and performance