ACCESS
Arctic Climate Change
Economy and Society

Project no. 265863

ACCESS

Arctic Climate Change, Economy and Society

Instrument: Collaborative Project
Thematic Priority: Ocean.2010-1 “Quantification of climate change impacts on economic sectors in the Arctic”

D6.31 – Creation and updating of a data management system – climate

Due date of deliverable: **31/08/2011**
Actual submission date: **02/12/2011**

Start date of project: **March 1st, 2011** Duration: **48 months**

Organisation name of lead contractor for this deliverable: **METNO**

Contents

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

1 Background

The Norwegian Meteorological Institute (METNO) developed and operated the data management system for the EU FP6 project Damocles. This is still available at http://damocles.met.no/data_management/. The Damocles data management system supports data compliance check, data upload, data discovery, data download and metadata exchange. The Damocles data management system was integrated with IPY metadata catalogues. It standardised the file format on NetCDF/CF which is a fully self describing file format which can be handled directly in many visualisation and analysis systems as it follows a standardised structure and documentation convention. This convention is outlined at http://cf-pcmdi.llnl.gov/. In addition to the documentation standard a number of global attributes were added to support data discovery. These attributes fulfils the minimum requirements of metadata standards like ISO19115 and GCMD DIF.

METNO is supporting ACCESS data management for climate-related data with the data management system originally developed for Damocles (and other projects). Further details on the system is provided below.

The data management system set up by the METNO constitutes the ACCESS data management system together with the system set up by the National Oceanography Centre, Southampton, (NOC), representing the Natural Environmental Research Council (NERC, ACCESS Partner 3). The system set up by NERC handles non-climate data that will be used in the ACCESS project while the system set up by METNO handles climate related data. The METNO system is described below.

The linkage between the NERC and METNO subsystems of ACCESS data management is still under development, some elements of linkage is briefly discussed below.

This deliverable is delayed due to a new computer building being phased in at METNO. During this process, creation of new systems were not allowed, focus was on safe migration of existing machines and software. The new building is built according to green computing norms.

2 System implementations

2.1 Data management

2.1.1 Overview

The ACCESS data management system has the following main functional components:

1. Data upload
2. Metadata handling
3. Data access
4. Backup

These components are briefly presented in the remaining chapters. To facilitate data upload, metadata handling and data access the data and metadata management system METAMOD (http://wiki.met.no/metamod/) is used in combination with a THREDDS Data Server (http://www.unidata.ucar.edu/projects/THREDDS/). Backup is facilitated using internal METNO specific systems.

2.1.2 Data upload

The ACCESS Data Management System hosted at METNO have standardised on NetCDF[1] files following on the Climate and Forecast (CF) convention. A full description of the file format and required global attributes including a description of the controlled vocabularies used is found at http://access.met.no/.

In addition to the requirements implied by CF convention, the project has added its own metadata requirements. One reason for this is to satisfy international metadata standards (e.g. ISO 19115, GCMD DIF etc).

Data can be uploaded interactively or automatically. Files uploaded manually are automatically checked to see whether they confirm to the standard or not, and then metadata are automatically extracted into the metadatabase before the files are placed in a data repository served through a THREDDS Data Server.

Data can also after agreement be uploaded automatically using FTP. Currently, the only data automatically inserted into the data management system are METNO datasets or external datasets ingested through METNO implemented systems like e.g. Ice Thethered Platforms or WMO GTS datasets with free access (WMO Resolution 40 datasets).

The data management system do also facilitate a metadata editor. This is by no means complete, but may be used to link to data hosted externally to the system maintained by METNO. It follows a basic structure using Global Change Master Directory (GCMD) science keywords to describe the data and allows specification of a URL linking directly to the data described. A sample view is provided in Figure 1.

[1] CDL files are also supported.

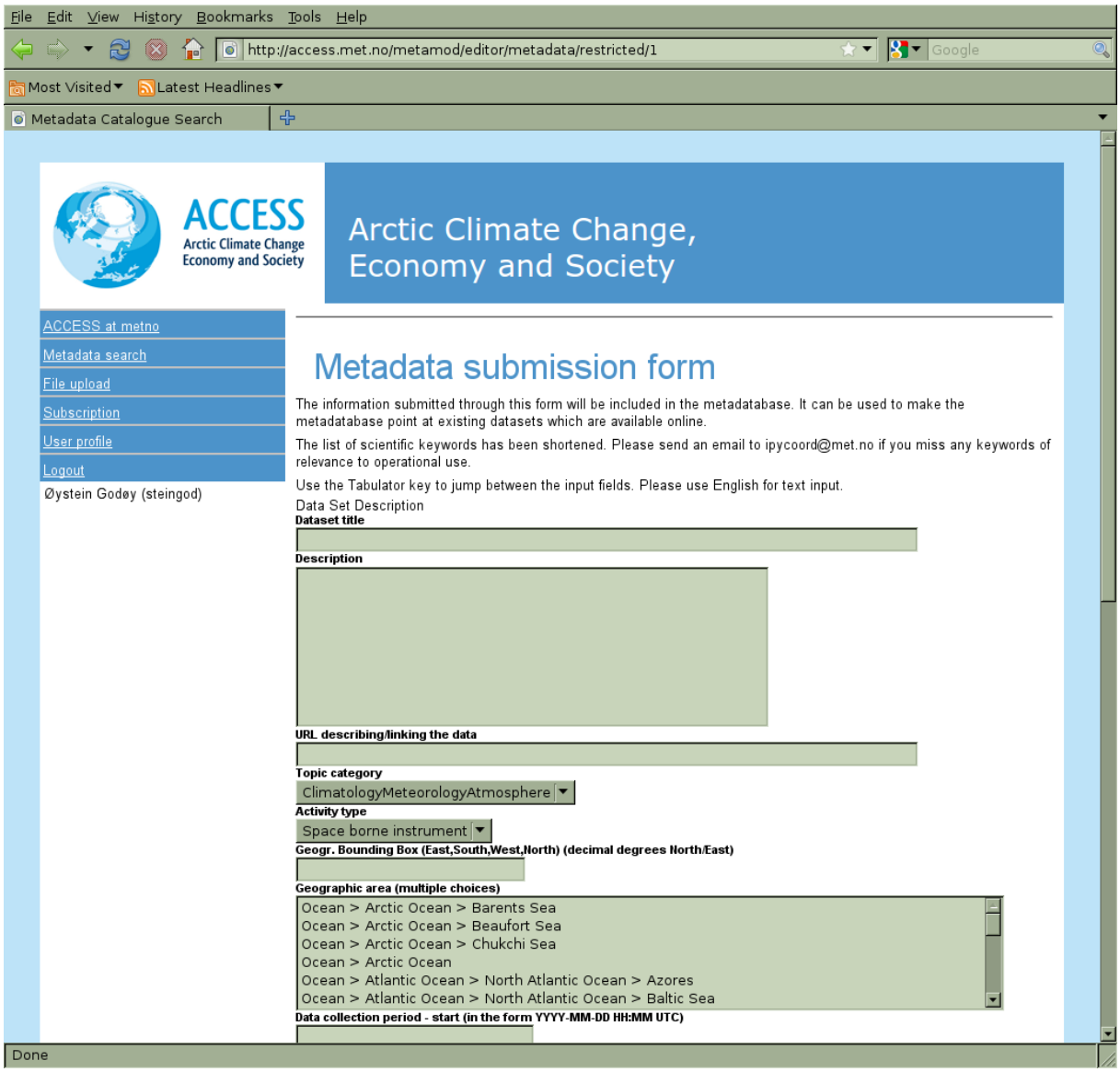

Figure 1: The metadata submission form supported by the system. this may be used to point at data stored within other data management systems.

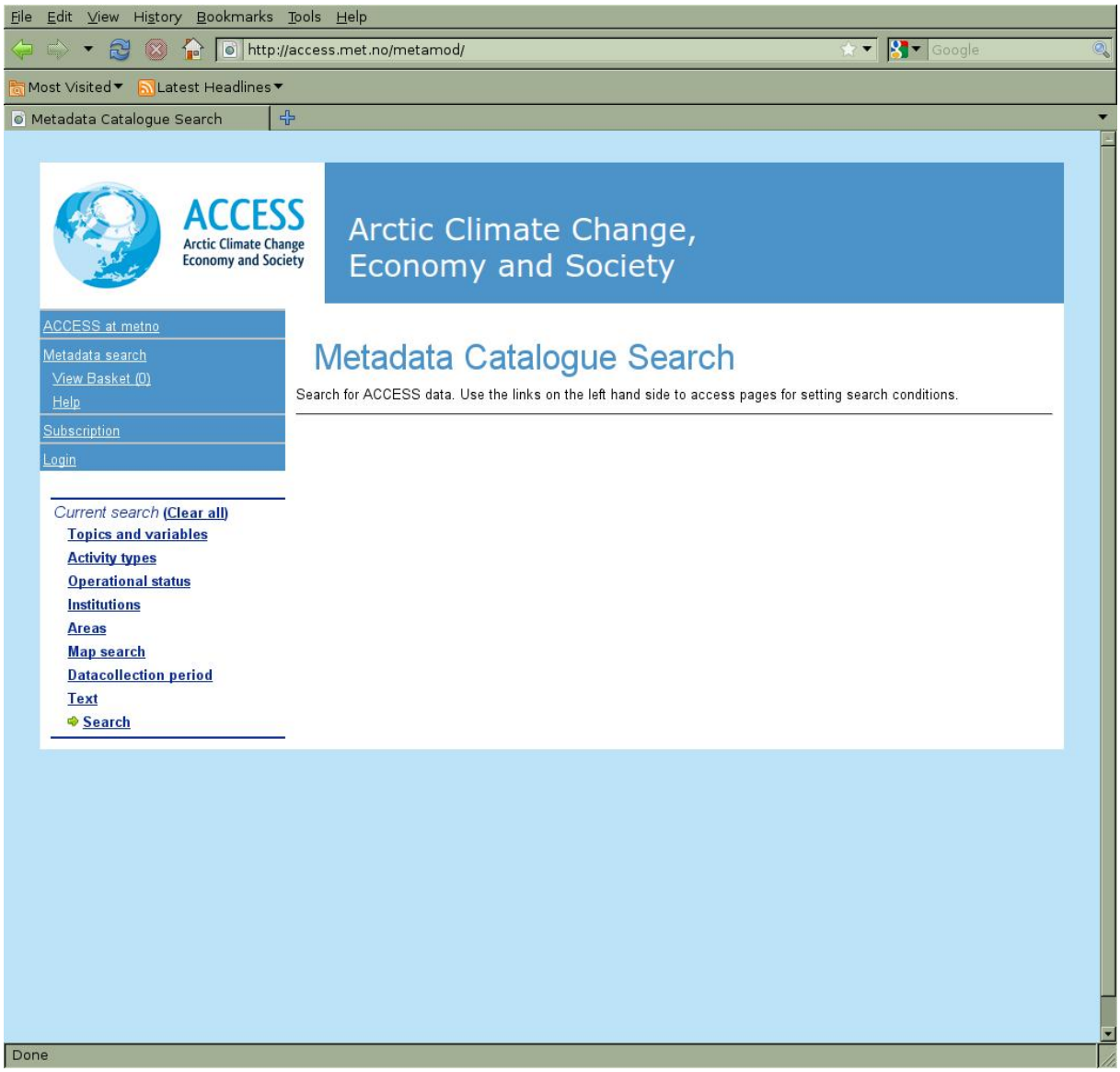

Figure 2: Screen dump of the interactive data discovery interface for ACCESS supported through METAMOD.

2.1.3 Metadata handling

As mentioned above, metadata are automatically extracted from the data sets uploaded. These metadata are made available for interactive search/browsing using a web interface. Furthermore, a machine interface is available for exchange with partners using the Open Arcive Initiative Protocol for Metadata Harvesting (OAI-PMH) serving Global Change Master Directory (GCMD) Directory Interchange Format (DIF) or ISO19115[2].

The interactive data discovery interface supporting ACCESS using METAMOD is illustrated in Figure 2.

[2] ISO19115 is strictly a container, mapping to various ISO19115 profiles is yet not fully supported.

2.1.4 Data access

Data uploaded to the ACCESS data management system are processed to extract metadata and then placed in a data repository which is served using a THREDDS Data Server implementation. This server is configured to deliver products through HTTP and OpeNDAP. For some datasets OGC WMS and OGC WCS may be activated, but this is not done default.

Some data sets are freely available, some require a user name and password for access. This depends on the distribution statement attached to each dataset. Within the ACCESS community all data are freely available and a common username/password will be provided to all ACCESS partners. Any misuse of data is handled by the management structure of ACCESS and the ethical norms within the scientific community. Datasets tagged as restricted to ACCESS only is not available to scientists not belonging to ACCESS. However, metadata describing who measured what, where and when is freely available. Currently exposer of ACCESS metadata towards other catalogues have yet not been agreed.

2.1.5 Backup of data repository

The data repository used to serve the data is implemented in a regular backup scheme, with incremental backups every night and a full backup once a week. Currently all available data are online, which they will continue to be for the coming years. At some point they may however be transferred into a tape archive, but as technology advances, more and more data are kept online. The data archive will be included in a WMO Information System Data Colleection and Production Center which METNO is establishing for the Arctic. This system is yet not operational, but the portal is available at http://arcticdata.met.no/.

2.2 Internal communication

A wiki solution to support internal communication within the project has been setup. This is a DokuWiki solution including a number of plugins supporting mime-types, tags, tasklists, sortable tables and much more. A comprehensive list of supported plugins is available at https://wiki.met.no/plugins. The wiki itself is available at https://wiki.met.no/access/.

This wiki is only accessible (and viewable) to registered users belonging to the ACCESS project. A dedicated area for the Steering Committee have been identified and users are divided in two groups to support authentication/authorisation. Members of the Steering Committee have access to all parts of the wiki, all other project members have access to all areas, except for the steering committee area.

The frontpage of the wiki is illustrated in Figure 3. The wiki will be used for documentation, discussion, description of best practises etc within the project. It is suitable as a tool integrating the climate-related and non-climate related communities within ACCESS as it opens for more detailed documentation/information than the normal metadata standards used.

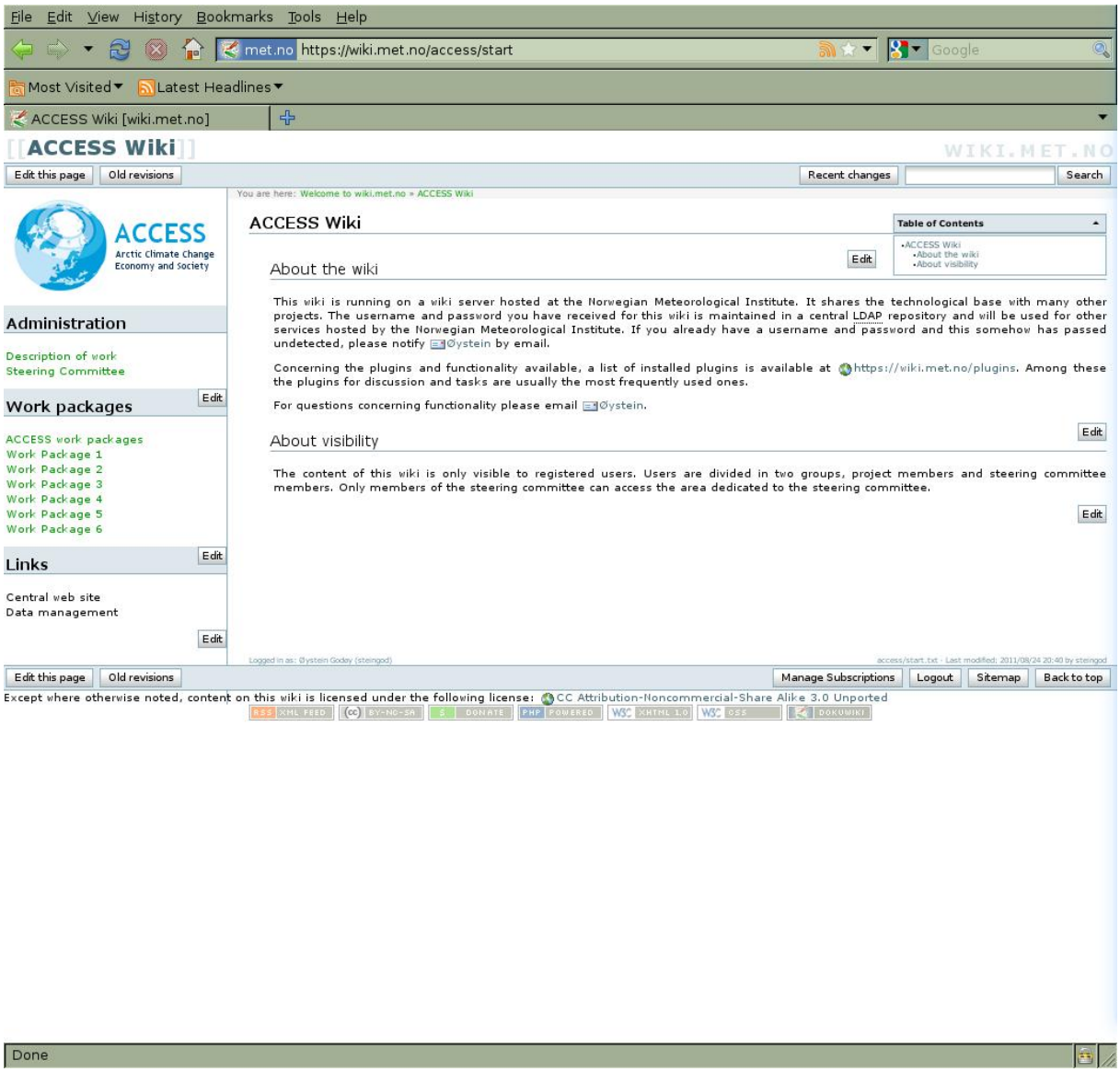

Figure 3: The wiki solution set up to support internal communication within the ACCESS community.

3 Status and next steps

The data management system set up to support ACCESS is known to many ACCESS partners from the FP6 project Damocles. It has been developed further and new functionality developed external to ACCESS will be implemented for ACCESS at a best effort basis as no development is funded by ACCESS. In the current setup, all datasets collected and generated through Damocles are also made visible in the ACCESS data management system.

For METNO to maintain the integrity of the project information, a close collaboration between the partners, the project office and the data management team at METNO is required. METNO will need regular updates of the list of project members in order to approve or reject users requesting access to the information stored within the system or wanting to upload information. For this to work, METNO need the following information on each user:

- Name

- Institution affiliation
- Work package affiliation
- Phone number
- Email address

A regular stream of this information from the project office to the data management team is required. users not listed will be refused access.

Within ACCESS, data management is split in a repository for climate related data hosted by METNO and a non-climate related repository hosted by NOC/NERC. Currently these two systems are not connected, but they may be in the future. How is yet not determined but some possible scenarios are:

1. Users submitting non-climate related data utilise the metadata editor hosted at the climate-related node to document the data and the URL of the actual data is entered in the metadata. this would make all ACCESS data searchable within the same portal, but modifications of the metadata editor may be required and with no funding for this adaptation, a cost benefit analysis is required first.
2. Users submitting non-climate related data document data within the NOC/NERC node and metadata records following an agreed standard (e.g. GCMD DIF) is exchanged (using e.g. OAI-PMH or HTTP) between the two systems.