

QUEST Data Management presentation at QUEST ASM, 21/4/06 by Wolfgang Knorr

Notes from K Marsh, BADC.

Wolfgang Knorr, the QUEST Deputy Leader and responsible for data management gave a presentation describing the QUEST data policy which outlined the key elements. The full data policy in its latest draft version is available from the main QUEST web site.

Summary of Presentation:

Effective data management is very important for the success of QUEST. The QUEST data policy (based on the NERC data policy) states that 3 months after “acquisition”, the data should be made available to the chosen NERC designated data centre (DDC), along with the required metadata. For QUEST, the intention is that the DDC providing this support will be the British Atmospheric Data Centre (BADC). The BADC will facilitate the sharing of data through the various projects, as well as obtaining other required datasets.

The data producers can request that their data are embargoed for 1 year from acquisition, to allow them time to exploit the data. Other QUEST researchers will need to seek the permission of the data producer before access can be given during this period. All data will be publicly accessible 1 year after acquisition. The intention is that the IPR of the data will remain with the data producer.

For the model software that will be developed, it will be necessary to store different versions of the code, boundary conditions for the runs, and metadata describing the model configuration for each run. The amount of actual model output that will be required to be archived has not yet been finalised. Indeed, it may not be until some time after the runs have been completed before judgements can be made as to what output data are worth archiving.

The other important source of data will be produced from data synthesis work, and it is envisaged that these will go directly to the DDC.

A “to do” list was presented for QUEST data management. This stated that 1 data management representative per project should be appointed, to ensure that the issues were addressed and information could be fed back to the QUEST data manager. A data working group may also be established to provide a forum for feedback from the individual projects.

The BADC have been funded by NERC to perform a scoping study to determine the precise data management requirements for QUEST, and will be contacting the relevant participants in the near future. This will form the basis of the QUEST Data Management Plan (DMP). The BADC can also provide advice on all aspects of data management.

Outcome of Discussions:

It is likely that the QUEST Earth System Atlas (QESA) produced by the project will be mirrored at the BADC while under development, and may also reside there in its final form. As the QESA is a form of publication and is subject to the ICSU policy of open access, these data will be freely available as soon as they are published.

Model output from GENIE currently resides on its own project data area. These data (and any future output) can remain there for further analysis, however any runs that require long term storage should be archived at the BADC.

Access to the BADC archive would be via ftp, http, or in the case of model output, by means of the BADC data extraction service. This allows users to subset the data on-line and download the required selection in a variety of formats.

There were also the suggestions from the audience that 1) a data processing server at the DDC would be useful, where users could upload and run their own programs (such as IDL, Ferret) on the data, and 2) the use of OpenDAP to access the data remotely.