

# **COMET NERC EO Centre of Excellence**

## **Draft Data Management Plan**

### **NEODC**

**August 2005**

#### **Introduction**

NERC's Data Policy requires the curation of data generated by the research they fund. This means ensuring the long-term archiving and widespread use of the data, and ensuring best practice to achieve this. NERC are implementing this policy through a set of designated data centres, which in the case of Earth Observation, is the NEODC.

A survey of NERC EO Centres of Excellence was carried out (Jan – March 2005) in order to establish: (i) what data is used within the NERC EO Centres and whether there are common requirements best organised centrally, and (ii) to develop each Centre's plan and policy for data management.

Questionnaires were sent to all COMET researchers to determine their needs in terms of data support (provision of third-party data sets or other services). The enquiry also addressed issues related to the data generated by the projects (nature, volume, flow, etc.). The main purpose is to consider data with long term importance and/or use to the wider scientific community.

This draft Data Management Plan is the result of discussions between and response to data questionnaires from:

- The COMET Director
- COMET PI's and researchers
- The NEODC

#### **COMET structure**

The Centre for the Observation and Modelling of Earthquakes and Tectonics (COMET) is a collaborative project funded by NERC, involving scientists from the University of Oxford, University of Cambridge, and University College London. The COMET scientists use satellite observations to help them quantify seismic hazard and interpret how past environments control the distribution of natural resources. For more information see <http://comet.nerc.ac.uk/>.

#### **Scope**

The purpose of the COMET data management plan is to set up a coherent approach to data issues for the Centre. Its objective is to ensure that

- Appropriate data support is provided to the scientists within the Centre.
- COMET datasets are archived and distributed in a suitable manner

- Distribution conditions and data usage do not infringe on the individuals' rights to publish their own work.
- Potentially scientifically valuable data are kept for the long-term.
- A high quality documented COMET data archive is created.
- Data and documents can be distributed more widely to the scientific community.

At present there is no funding to provide full data support and archival for all Centre of Excellence datasets. The NEODC can currently only provide these where there is not a resource issue. However, the aim is to identify what the Centres of Excellence's needs are, in order, in a next step, to ascertain what funding is required to meet them.

The following sections cover the main data management issues: provision of a data management plan and a data protocol, setting up an archive, monitoring of data access, data distribution, publication of results based on COMET data and support offered to data providers.

## 1. Data management plan and data protocol

The present draft data management plan should lead, after discussion with COMET PIs, to a final Data Management Plan. It is suggested that a data protocol be adopted for the Centre (a proposed draft is attached to this document).

## 2. Third-party data

### 2.1 *Third-party data external to COMET*

Third-party data required for the development of the projects and held at the NEODC or BADC (e.g. Met Office data, Landsat images), will be made available to COMET scientists, subject to current access conditions. If required, NEODC will endeavour to retrieve data sets from other sources at no cost or will negotiate their acquisition at the best possible cost.

## 3. Sharing COMET data and model results

Data and model results generated by individual COMET groups or researchers could be made available to other groups through the NEODC. Publication issues are dealt with in Section 6.

## 4. COMET data archive

### 4.1 *Archive location*

COMET archives will be located at NEODC (and/or Oxford).

Where it is considered that data are of wider interest to the community and a long-term archive is appropriate the data should be located at the NEODC. The data provider is also responsible for providing documentation, metadata and possibly software to decode, interpret and visualise the data. The data provider may also be expected to field some user queries: science questions should be directly addressed to

**Comment:** Depending on resources, NEODC can archive some or all of the COMET data sets, and/or the data can be held elsewhere, e.g. Oxford, if it is set up to deal with backups, access control and documentation adequately. The details of which data are archived where can be finalised at a later stage.

the responsible scientist, and questions about the data availability, format, etc. to the NEODC helpdesk.

#### ***4.2 Archiving policy***

In recognition that validated raw data (i.e. QA/QC'ed data prior to additional processing) potentially represent an invaluable source of information for the future, the Centre's scientists will archive them in a way that guarantees longevity and accessibility. Although not necessarily located at NEODC, validated raw databases and their access should be fully documented at the NEODC. Processed (final) data will be archived at the NEODC. In addition, investigators are encouraged to submit model results which will have been the basis of theoretical studies or that illustrate the model capabilities.

#### ***4.3 Format***

All data produced by COMET should be stored in standard (commonly used by the community) file formats. When deciding on an output format COMET scientists should consider accessibility and future use. If non-standard data formats cannot be avoided, comprehensive format descriptions and read software should be provided.

#### ***4.4 Documentation***

Metadata (i.e. information on the data) are a crucial part of any data archive since they ensure the accessibility and readability of the data. It is therefore essential that metadata be submitted at the same time as the data sets to which they pertain. Metadata documenting the existence of all COMET data not archived at the NEODC should also be supplied to the NEODC.

To guarantee the data archive quality, full documentation on all validated raw and processed data, as well as on models and model results, must be provided to the NEODC. Standard metadata will be archived within data files. For an example of the sort of metadata that should be provided see the BADC metadata help pages at <http://badc.nerc.ac.uk/help/metadata>. An equivalent NEODC page with specific notes for EO data will be produced in the near future, but any questions in the meantime may be directed to [neodc@rl.ac.uk](mailto:neodc@rl.ac.uk).

In addition to the standard metadata, investigators are encouraged to archive all relevant information, including model and experiment descriptions, references, papers, reports, etc.

#### ***4.5 Supporting collaboration with Collaborative Workspaces***

If requested, the NEODC can set up a collaborative workspace dedicated to COMET. This would be a secure web space available to registered users only, where scientists can share results, documents and preliminary data files.

#### ***4.6 Data submission***

Preliminary data should be made available to other COMET scientists, where appropriate, as soon as possible.

Via NEODC or internal transfers – state which/how

Processed data and model results should be supplied to the NEODC/chosen data archive location as soon as they are ready, and no later than the project end date.

If using NEODC – describe upload method here, e.g. web based file uploader or ftp.

**Comment:** Details in this section can be filled in once archive details are known

## 5. Data distribution

Different access restrictions are appropriate for different COMET datasets, although the duration of the “data validation period” during which access is restricted may be a common feature. A password-protected access system can be set up at the NEODC to reflect the defined permissions. Distribution of COMET data held at the NEODC will take place via the Internet and FTP. During any restricted period, entitled COMET scientists who have applied for access to the data will be allocated an account at the NEODC allowing them to directly download the data from the archive. This facility can be extended to external collaborators who will have been personally authorised to access the data by COMET PIs.

At the end of the retention period, the data will be released to the public domain. The Intellectual Property Rights (IPR) to those data need not be transferred. After release, NEODC will make the data available to other bona fide researchers. Anonymous users will be requested not to use the data for commercial purposes; they will be asked to contact the relevant data providers before using the data and to acknowledge COMET and the data suppliers in any publication using COMET data. If required, a system can be put in place by which users will be asked to indicate agreement to these (possibly amended) terms prior to being given access to the data.

A COMET Web page will be set up at NEODC with links to datasets at NEODC and elsewhere, publications, data access rules etc.

## 6. Publication

Results coming out of COMET projects will be published in the usual way. During the data validation period, each investigator will have the right to refuse the use of his/her results in a publication or a presentation prior to the investigator’s own publication of that work. If measurements or model results from other groups within COMET are used in a COMET participant’s publication during or after the project, joint authorship must be offered. This will not necessarily have to be accepted, particularly in cases where due credit and acknowledgement can be given in other, possibly more appropriate, ways. References of publications should be communicated to the NEODC where a list of published works will be held.

## 7. Liaison between NEODC and COMET scientists

The COMET web page at NEODC will be the primary source of information regarding the COMET archive.

The NEODC will keep in touch with the PIs and their collaborators, e.g. to exchange information on the submission procedure, relevant WWW links, the Data Management Plan and on the population of the COMET archive using this website.

### **8. Support to COMET scientists**

Any other services NEODC could provide?

## Appendix 1 – Summary of CLASSIC data

### 1. Third party datasets

Dataset	Access	Who	When
ESA SAR data	Already have access	Oxford	n/a
Landsat TM/ETM	Already have access	Oxford	n/a
SRTM	Already have access	Oxford	n/a
Weather satellite data	Already have access	Oxford	n/a

### 2. COMET datasets for long term archival

Dataset	Size	Data Producer	When available to archive
Raw ERS/Envisat SAR data		ESA	Already archived at ESA and Oxford
Interferograms for earthquakes and deforming areas	~500 Gb/year	Oxford	When validated and suitable for others to use
DEMs from single satellite sensor models (ASTER, SPOT etc)	25-50 Mb	UCL	When validated and suitable for others to use
Mosaiced DEMs of many models.	300 Mb		
GPS data			Can't be archived outside COMET for due to access restrictions for the foreseeable future
Cambridge data (derived products using fieldwork and satellite data)	?	Cambridge	?

**Comment:** I understand the GPS data are restricted, but are any processed dataset based on these data likely to be usefully archived at and/or distributed by the data centre?

### 3. COMET produced data shared between groups/researchers

Dataset	Size	Share with whom?	How
Interferograms (InSAR) - Oxford	~500 Gb /year	Inside and outside COMET.	COMET (ftp)?
DEMs – UCL	300 Mb	Inside COMET and in future outside	COMET (ftp) ?

		COMET	
GPS data – Oxford		COMET	COMET (ftp) ?
Cambridge data	?	?	?

## Appendix 2 - COMET Draft Data Protocol

The aims of the Data Protocol are

- to encourage rapid dissemination of scientific results from COMET;
- to protect the rights of the individual scientists funded by COMET;
- to have all the involved researchers treated equitably;
- to ensure the quality of the data in the COMET data archive.

These aims conflict at times, and it is hoped that the provisions of the protocol resolve these conflicts fairly. It is recognised that this cannot always be achieved to everyone's complete satisfaction; there are bound to be cases where individual interests clash with those of the COMET Centre. Therefore, to try to meet these aims, all PIs involved in COMET, in accordance with and on behalf of their co-investigators, must agree to abide by the following conditions:

1. COMET data and model results produced during the lifetime of the Centre will be made available to all COMET scientists, and COMET scientists only, during a dataset specific *restricted access period* ending no more than one year after the concerned project end date, after which data and model results will be released to the public domain. At a principal investigator's request, access may be extended to personally authorised collaborators.
2. The designated COMET data centre is the NEODC.
3. The longevity of validated raw data must be ensured in a secure archive, if possible at NEODC. Details pertaining to the validated raw data (i.e. metadata), whether or not archived at NEODC, must be sent to the NEODC, as well as information on how to access the data.
4. When relevant, preliminary data must be made available to COMET collaborators as soon as possible. Any corrections or amendments to the preliminary data should be announced as soon as possible.
5. Validated processed data (i.e. data sets in their final form) must be archived at the NEODC. Archival must take place no later than the end of the concerned project.
6. Results of model studies feeding other COMET projects or using data acquired during COMET can be made available via the NEODC.
7. Data submitted to the NEODC must be in the data format agreed between COMET principal investigators and the NEODC. All agreed metadata describing data, models and model results, regardless of their archival location, must be supplied to NEODC. Format and metadata are documented at NEODC.
8. It is each principal investigator's responsibility to ensure that the data used in publications are the best available at that time.
9. If measurements or model results from other COMET research groups are used in a publication by a COMET participant, joint authorship must be offered. This does not necessarily have to be accepted, particularly in cases where due credit and acknowledgement can be given in other, possibly more appropriate, ways.
10. Whilst the data are restricted from the public domain (see Clause 1), each principal investigator has the right to refuse to allow his/her work, whether measurement or calculation, to be used in a publication or presentation prior to the PI's own publication of that work.
11. Whilst the data are restricted from the public domain, no data should be transferred to a third party without the originator's consent.
12. In the event of dispute the final decision rests with the COMET Steering Committee.