

NEODC Annual Report 2006/2007

Victoria Jay, Matt Pritchard, Bryan Lawrence - June 2007

Introduction

The role of the NEODC is, in line with NERC Data Policy, to ensure the responsible stewardship and distribution Earth Observation (EO) data acquired and generated by NERC, to give guidance on the availability and use of EO data and to coordinate the acquisition of new data resources. NEODC work therefore consists primarily of data management activities, data acquisition and user support.

The following metrics give some indication of NEODC usage:

- 152 identifiable users downloaded 22TB of data in 1.2 million files over this period.
- 187 users have registered this year bringing the total to 474. The number of users registered for one or more datasets or services is 178.
- Approximately 245 queries were answered this year.

NEODC Management

The Centre for Environmental Data Archiving (CEDA) hosts both the BADC and the NERC Earth Observation Data Centre (NEODC). This year has shown significant progress in implementing an organisational structure that will enable more efficient sharing of data centre infrastructure. The Open Archive Information Systems reference model (an ISO standard for archives) has been used to ensure best practice in this context.

Developments

Website

The NEODC website has continued to be updated throughout the year, with news items and fact sheets for new datasets.

A new, improved version of the interactive map showing NEODC airborne datasets and related information has been deployed on the NEODC website (URL: http://www.neodc.rl.ac.uk/maps/mapserver_neodc/dbox/neodc/neodc_coverage.html). The map enables several datasets (NEXTMap, ATM, CASI, LiDAR, Photography) to be viewed in spatial context, with a choice of background mapping. Clicking on individual items reveals metadata about the item, and in many cases enables registered users direct download access to the data. Furthermore, the server providing the backend to the map interface exposes the various data layers in the map as OGC Web Feature Service (WFS) layers, enabling users to by import these layers into their own GIS systems.

User Support

Delivery of data to customers has continued to grow, with 21 datasets now available via NEODC. Approximately 245 user enquiries were logged this year, compared to 180 in 05/06. The number of registered users also increased significantly (from 287 to 474), as did the volume of data downloads (e.g. 152 users downloaded 22 Tb of data, compared to 101 users and 5 Tb in the previous year). A breakdown of download statistics by period and dataset is shown in Appendix 1.

Dataset catalogue

A new dataset catalogue has been developed and implemented with BADC. This system uses the Metadata Objects for Links in Environmental Science (MOLES) metadata schema developed by the NERC DataGrid project, and provides a more useful and complete basis for storing dataset information. Staff were trained to use the new system which came into operation in November and

work is underway to populate the new catalogue with information about all NEODC datasets and related entities (e.g. data production tools, satellite missions etc).

Progress on Datasets

- **CASIX** : The CASIX EO Centre of Excellence provided some datasets including gridded CO2 flux climatology and ship data. The dataset is restricted to CASIX partners initially but will be made generally available in due course.
- **NCAVEO field experiment**: NEODC is distributing data from the NCAVEO (Network for Cal/Val of EO data) field experiment in June: aircraft, satellite and ground observations are gradually being added to the archive.
- **Aerial photography**: The project funded by the NERC Data Management Coordination budget to digitise the archive of ARSF aerial photography is approaching its conclusion. Approx. 35,600 images have already been delivered to NEODC and are available on-line alongside other ARSF data. A metadata catalogue is being produced. There are fewer than 6,000 images remaining to be scanned and it is estimated that the entire collection will be digitised by the end of June.
- **ARSF**: The majority of 2006 data were provided to NEODC on disk in May 2007 and delivery of remaining 2006 data is expected in the coming months. Following a major hardware failure in August 06 at NEODC, the ARSF archive was restored from tape backup and was reorganised into the correct file structure.
- **Landsat-7**: The collection of Landsat7 ETM data at NEODC has been re-organised and is now available via the on-line archive.
- **MTCI**: MERIS Terrestrial Chlorophyll Index products produced by Infoterra are hosted at and distributed by NEODC. Data are available to registered users for 2003, 2004, and since May 2006 near-real time data are added as they are produced. In response to BADC and NEODC user demand NEODC commissioned Infoterra to process the data to fill the gaps (end 2002, 2005 and early 2006).
- **ESA (A)SAR**: ESA has authorised NEODC to distribute SAR data held at NEODC to UK students. (A)SAR data previously held at NOCS has been ingested onto the NEODC archive and will be made available to eligible users. Metadata is being collected and will be used to make the data searchable via the web site.
- **ESA Envisat**: MIPAS, SCIAMACHY and MERIS data continue to be added to the NEODC archive (arriving via DDS and ftp). Gaps in the MERIS and SCIAMACHY archive (2004/2005) due to hardware failures at BADC/NEODC are being filled via tape and ftp from ESA. The GOMOS dataset has been removed from the NEODC catalogue as there are no active users, and equivalent data are available for download directly from ESA.
- **(A)ATSR**: The (A)ATSR common-format archive has continued to grow at NEODC, with regular deliveries of AATSR data from ESA, and processing to Envisat-format of ATSR-1 and ATSR-2 data. Official release as Version 1.1 (June 2007) of products processed so far was publicised at the Envisat Symposium (Montreux) in April.

NCEO

Bryan Lawrence attended several meetings relating to the National Centre for Earth Observation, due to start 2008, and a bid was prepared and submitted for the role of NEODC in the NCEO Informatics Theme.

Services and Facilities

A short data management survey was carried out by BADC and NEODC of the relevant NERC Services and Facilities. The findings are included in Appendix 2.

Collaborations and Co-funded activities

The NEODC has a number of co-funded activities and collaborations. Active and funded projects in

2006-2007 include:

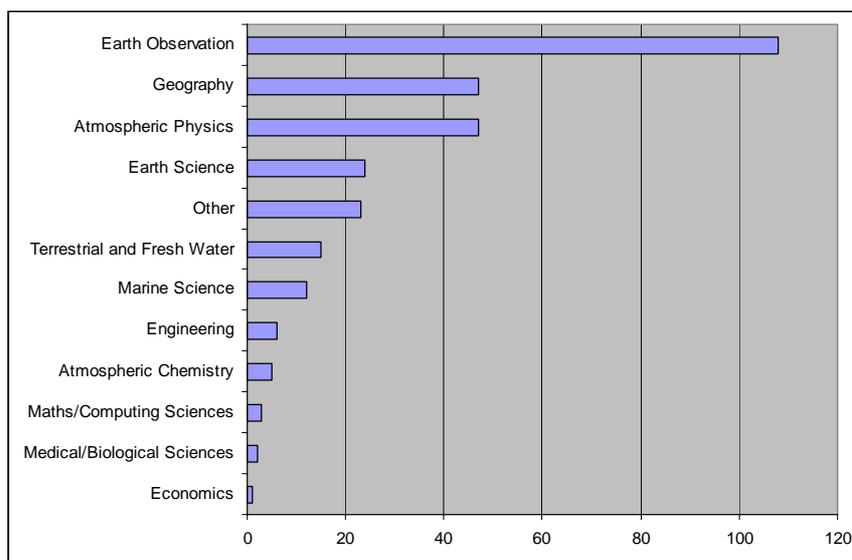
- *(A)RC – (A)ATSR Reanalysis for Climate* is a major Knowledge Transfer project funded by NERC, Defra and the MOD. (A)RC is led by University of Edinburgh and includes the Met Office, University of Leicester, RAL and NOCS. The project started in 2006 and will continue until 2010. NEODC's role is to facilitate processing and archive and disseminate the project's output data.
- *(A)ATSR Archive* - The £480k Defra/SpaceConneXions project to create a common-format archive of (A)ATSR data products started in 2005 and is due to end this year. A proposal for a joint (A)ATSR Operations & Archive follow-on contract has been prepared and submitted to SpaceConnexions for consideration by DEFRA, with the aim of continuing operations and archive support at RAL for the next 2 years. In addition, RAL/NEODC has been asked to cost the additional activity of converting the entire (A)ATSR archive into GHRSSST L2Pcore products.
- *Multimission Vicarious Calibration of Spaceborne Optical Instruments*: ESA project to create a dataset of cal/val products from the (A)ATSR Archive, started 2007
- *ESA SSE Project*: A bid to ESA for the creation of additional web services using the ESA SSE was successful, providing some £44k of funding until the end of 2007. Three NEODC services will be developed, 2 relating to producing derivative products from (A)ATSR, and a further service aimed at enabling SSE users to discover resources made available by the NERC DataGrid.
- *ESA data management strategy*: NEODC has been involved in providing UK feedback to ESA's data management strategy and ESA's Heterogeneous Mission Accessibility (HMA) project
- *AVHRR Rothera Archive*: This project (with BAS and RSDAS), to process and make available the archive of AVHRR data acquired at the Rothera base in Antarctica, was completed in 2006. The NEODC has created metadata (compliant with NERC DataGrid (NDG) and FGDC standards) and assisted in the development and provision of a web interface to make this discoverable and available to the user community.

Other collaborative activities during 06/07 include:

- *JISC Geospatial Data Working Group*: Matt Pritchard (MJP) was nominated by the NERC Data Management Coordinator as NERC's new representative on the JISC Geospatial Data Working Group
- *NCAVEO* : Victoria Jay (VLJ) continued as a member of the NCAVEO steering committee
- *Dundee back-up*: NEODC continues to operate a secure back-up for AVHRR and SeaWiFS data at NEODAAS Dundee.
- *BIGF*: NEODC now holds a backup of British Isles GPS Archive Facility data in the firesafe
- *ARSF*: Several meetings were held with staff at ARSF to discuss data transfer to NEODC and outstanding data management issues. Note that recent staff changes at ARSF and the upcoming retirement (2008) of Andrew Wilson at CEH mean that significant effort (at NEODC and ARSF) will be required this year on handover of information, data and documentation.

Knowledge Exchange

Figure 1: Active NEODC users in 2006/07, by discipline



The NEODC provides data to an entire community of EO data consumers, not just EO scientists. Of the users who registered and accessed data in the year 2006/2007, more than half were not from Earth Observation (see Figure 1). In particular, several were not from NERC disciplines (6 from engineering, 2 from medical and biological sciences, and 1 from economics).

As well as direct communication of data, many of the projects with which NEODC is involved have knowledge exchange as a fundamental outcome:

- (A)RC is a Knowledge Transfer project to generate a climate quality SST dataset for users at the Hadley Centre and wider.
- NCAVEO is a NERC Knowledge Transfer Network to support the calibration and validation of Earth observation data.
- ESA SSE development will provide services for users in the community to access and manipulate NEODC datasets.
- The MTCI dataset, a consolidated regional Terrestrial Chlorophyll Index produced as a Level 3 product by Infoterra Ltd with ESA funding, is publicised and made available for the research community by the NEODC.

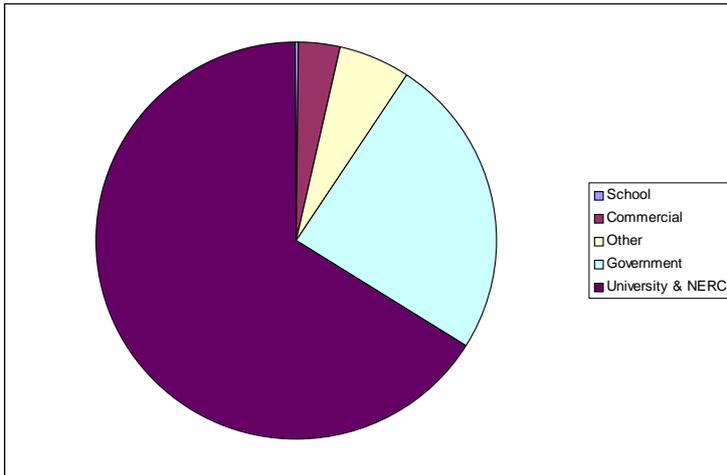
Outreach

The following activities in public engagement, outreach and publicity were carried out this year. In addition, NEODC has a significant web presence, and many users of the NEODC are not from the academic community at all (approximately one third in 2006/2007 – Figure 2).

- VLJ was involved in preparation and presentation of publicity material on Earth Observation for an RCUK event to raise awareness of work on climate change and the environment at Swindon Outlet Mall 16/17 March (part of National Science and Engineering Week)
- NEODC staff supervised two work experience students 8-15 January
- NEODC sponsored the RSPSOC Annual Student Meeting (29-30 March 2007) and provided a publicity poster and leaflets for attendees
- VLJ attended the NAPLIB (National Association of Aerial Photographic Libraries) Annual Meeting at BGS in November and gave a talk about NEODC's data and services

- VLJ presented an overview of Earth Observation to school teachers at a meeting of the Specialist Schools and Academies Trust
- MJP and VLJ attended the NERC EO conference, Edinburgh, in June (display stand with posters and presentation advertising MTCI dataset)
- VLJ and Steve Donegan (SJD) attended the RSPSOC Conference in Cambridge in September 2006 to promote NEODC datasets and activities (display stand with posters) and support the ARSF workshop
- VLJ attended the EO workshop at the Challenger Marine Science Conference in Oban in September (presentation and posters alongside the newly-launched NEODAAS)

Figure 2: Active users in 2006/2007 by institute.



Appendix 1: Access statistics for NEODC

Table 1 : Time breakdown

Period	Users	Datasets	Number (k)	Size (GB)
2006Q2	55	13	75	3971
2006Q3	64	17	239	5634
2006Q4	72	17	452	5234
2007Q1	70	16	471	7003
ALL	152	18	1238	21843

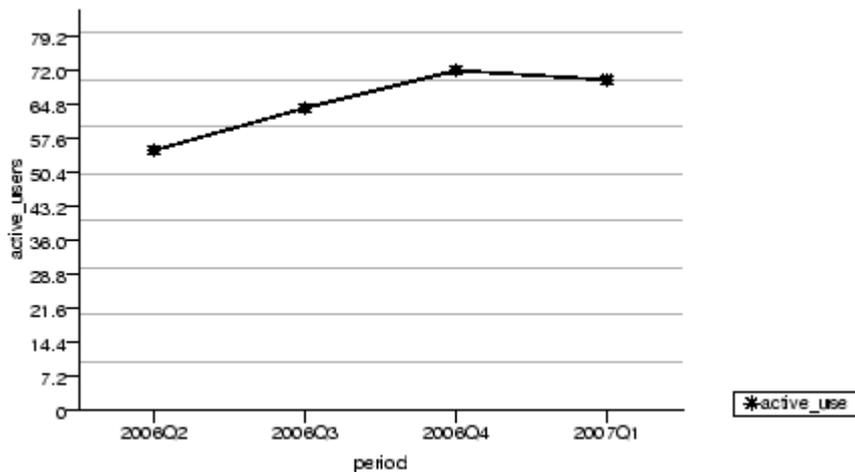


Table 2 : Most popular datasets

Dataset	Users	Number	Size (MB)
NEXMap	31	206489	335020
ARSF	26	1061	76438
AATSR multimission	25	65235	13034685
MERIS	20	929	319291
MIPAS	18	14557	3237486
Scanned Aerial photographs	17	1179	36263
NCAVEO field experiment	14	1580	96927
Landsat7 ETM	11	100	3314
ATSR UBT	11	946398	5153693
SCIAMACHY	11	254	61174
AVHRR_FASIR	7	66	161
MTCI	6	81	5164
NCAVEO LCM2000	5	73	7
SHAC2000	2	5	1671

NERC Services and Facilities – Data Management

14 June 2007

Victoria Jay, Charles Kilburn, Bryan Lawrence (NEODC/BADC)

BADC and NEODC have carried out a survey of data management practices in relevant NERC Services and Facilities. Short policy documents have been drawn up to clarify data management responsibilities between NERC Designated Data Centres and NERC S&F. The following Services and Facilities which have been included in this survey: ARSF, BIGF, CFARR, FSF, MST, MSF, NEODAAS, NSGF and the conclusions are summarised here.

Molecular Spectroscopy Facility

The MSF transfers experimental data to the BADC. Access is initially restricted to the experimenter (12 months) after which any BADC user may apply for access. The MSF must approve applications.

Data centre responsibilities: The BADC ingest data, set up a permissions group for each experiment and add approved users to the groups. Access statistics are tracked.

Funding implication : Minimal work which can be covered under BADC core.

Chilbolton Facility for Atmospheric and Radio Research

Data are uploaded once every few months by instrument scientists and some data streams are uploaded automatically every day. Data are in NetCDF to Climate and Forecasting metadata standards and follow the BADC filename convention. The dataset is split into instrument groups and access permissions are separate for each group. However there is no formal policy governing access to the data and this should be developed.

Data centre responsibilities: The BADC liaise with instrument scientists, provide access statistics and thumbnail image pages for cloud camera images.

Funding implication: Unfunded

The British Isles GPS Archive Facility

BIGF performs delivers archived GPS data to the community from the British Isles network of GPS stations. Data are held in RINEX format (standard for GPS data) and are provided through an online form.

Data centre responsibilities: NEODC keeps a copy of BIGF's data on DVD in the fire safe for off-site backup (but without knowledge about the data held).

Funding implication: None.

The Mesosphere Stratosphere Troposphere (MST) Radar

Data from the MST site arrive at the BADC constantly. Some processing is performed by the facility scientist and the results are redirected back to the BADC deliveries area. There are various data versions in the archive but the aim is for all data to be in NetCDF format. Some supporting data are in NASA-Ames files, others in non-standard formats which are being worked on.

Data centre responsibilities: The BADC ingests data every 15 minutes.

Funding implication: Activities covered by MST radar facility funding

The Field Spectroscopy Facility (FSF)

FSF calibration files are fairly small and FSF maintain them locally as they are frequently updated. Measurement data archival is the responsibility of PIs who use FSF equipment. The development

of a single archive of spectral signatures has been discussed, but lack of suitable methods of quality control and documentation mean that FSF feel this is not worthwhile at present.

Data centre responsibilities: NEODC/BADC only have data management responsibilities where PIs using FSF equipment are associated with projects linked to BADC/NEODC.

Funding implication: None

The NERC Space Geodesy Facility

GPS and GLONASS data are received at Hertmonceux and distributed routinely to various international services and archives. The data are available either directly from NSGF or via the international archives (International GNSS Service – IGS, NASA’s Crustal Dynamics Data Information Service – CDDIS, Eurolas Data Centre – EDC). Data are also archived and made available through BIGF.

NSGF are starting to make accurate weekly measurements of local gravity. At present there is no clear national or international archival facility for this data, but discussions are ongoing within the geodetic community. For the present, NSGF are archiving these data themselves.

Data centre responsibilities: None

Funding implication: None

NEODAAS (NERC Earth Observation Data Acquisition and Analysis Service)

NEODAAS provides EO data acquisition facilities (Dundee) and data processing and analysis services (Plymouth).

Users register on the website to gain access to freely available browse images and some products. They must submit a short proposal for access to high resolution data and other products. Formats depend on user requirements (e.g. 8 or 16 bit raw, GIF, TIF, PNG, NetCDF or GeoTiff)

Data received at Dundee are archived in raw, unprocessed form.

An online database is maintained of the archive catalogue. Only the foreign geostationary satellite data are not backed up at present, though it is likely this will happen as well.

NEODAAS-Plymouth receive raw Level 0 data via FTP from NEODAAS-Dundee, and process these to higher level products according to user requirements. They also receive data via the Internet from NASA subscriptions and orders and the ESA rolling archive. The final Level 2 and 3 products are maintained in an archive accessible by various web interfaces; these and the underlying data versions are archived or backed up onto LTO tapes, one copy held off-site. Certain Level 0-2 products are also archived where difficult or time consuming to obtain again from data providers.

Data centre responsibilities: NEODC stores copies of polar orbiting data from Dundee in firesafe (but without knowledge about data held), includes NEODAAS figures in annual Output Performance Measures for NERC and may have a data management role for data from Rothera project, if it becomes necessary in future (NERC EOEF project completed in 2006)

Funding implication: NEODC effort small and covered by NEODC core funding

ARSF Airborne Research and Survey Facility

ARSF data are processed (and geocorrected) at ARSF and provided to NEODC (by ftp and LaCIE disks) for archival and distribution, but ARSF deliver directly to project PIs, by ftp. The PI embargo period is 1 year after data are provided. Current formats: ATM and CASI Level 1b HDF products (including metadata), other files ‘as they are’ eg raw data, aircraft data, flight logs, aerial photographs, LiDAR in ascii with separate headers

Data centre responsibilities: NEODC

- ingest data provided by ARSF : make secure backups (Atlas tape store), organise data into project directories/instruments and make available on-line to users
- Handle user queries (sometimes with help from ARSF or Andrew Wilson at CEH)
- Handle user registrations and PI embargo period
- “pre-HDF” data needs work. No navigation data available but still useful to process these up to L1a. Work was started in ~2002 but not completed due to insufficient resource at

NEODC. Need A.Wilson's expertise to complete task, while he is still available (until 07/08?)

- Analogue aerial photography digitisation

Funding implication: Most tasks are already carried out within existing ARSF and NEODC funded activities. The current effort at NEODC is ~0.2 FTE. Additional effort is required to improve the archival and ensure proper curation of all ARSF data at NEODC (providing full metadata and consistent formats for all instruments, and coping with increasingly more data from new instruments). We estimate that additional funding for ~1/3 FTE is required.