



Project Number 017841

NitroEurope IP

The nitrogen cycle and its influence on the European greenhouse gas balance.

Sixth Framework Programme

Priority 6.3

Global Change and Ecosystems

D7.1

NitroEurope Data Management Plan

Due date of deliverable: **01/04/2006**

Actual submission date: **00/00/0000**

Start Date of Project: **01/02/2006**

Duration: **60 months**

Organisation name of lead contractor for this deliverable :
NERC

Revision: **final**

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	<input checked="" type="checkbox"/>
PP	Restricted to other programme participants (including the Commission Services)	<input type="checkbox"/>
RE	Restricted to a group specified by the consortium (including the Commission Services)	<input type="checkbox"/>
CO	Confidential, only for members of the consortium (including the Commission Services)	<input type="checkbox"/>

NitroEurope Data Management Plan

Contents

1. Introduction
2. Types of data generated by NitroEurope
3. Data management infrastructure, data centres, web portal and operations
4. Database contents
5. Formats
6. Metadata
7. Data file names
8. Data submission
9. Data validation and quality checking
10. How to access data
11. Support to the NitroEurope researchers
12. List of acronyms

Annex. NEU Conditions of Use

1. Introduction

This document sets the way to implement principles developed in the NitroEurope Data Policy, to which it refers. It is intended to be a working document and some of its sections will be updated in the course of the project.

2. Types of data generated by NitroEurope

Data gathered, compiled, collected or produced by the science components of the project are listed in the Measurement and Modelling Protocols. Each science component will be involved in some of the data activities summarised in Table 1, and will deliver data to one of the three DCs as listed in the last column of Table 1.

Data collected under Activities (1) and (2) (see Table 1) can present different levels of processing. Raw data, i.e. source measurements in the form that they have when they are first produced (more detailed *ad hoc* descriptions of raw data are given in the Measurement Protocol), will in general not be stored at the DCs but it is each PI's responsibility to ensure that they are stored safely with the relevant processing software or, alternatively, with documentation on retrieval algorithms, at least for the retention period as defined in the NEU Data Policy.

Processed data, i.e. observation data that have been subject to some treatment or formatting, or data derived from these, will be stored at the DCs and made available by them to the NEU community.

TABLE 1 – TYPES OF DATA GENERATED BY NITROEUROPE

Science component		Gathering of historic and external data (1)	NEU measurements (2)	Data synthesis/ processing (3)	Model output (4)	DC
C1	L1 sites - Flux Network-Inferential	✓	✓	-	-	CEH
	L2 sites - Flux Network-Regional, low cost	✓	✓	-	-	
	L3 sites - Flux Network-Supersites	✓	✓	-	-	
C2	Ecosystem Manipulation, Core sites	✓	✓	✓	-	
	Ecosystem Manipulation, Associated sites					
C3	Modelling	✓	-	-	✓	
C4	Landscape Analysis	✓	✓	✓	✓	INRA
C5	5.1 European Integration -harmonised GIS-based development	✓	-	✓	-	JRC
	5.2 European Integration -scenario derivation	✓	-	✓	-	
	5.3 European Integration -multi component model	-	-	✓	-	
	5.4 Analysis of scenarios with detailed ecosystem models	-	-	-	✓	
	5.5 analysis of scenarios with multi component INTEGRATOR model	-	-	-	✓	
C6	6.1 Verification NEU bottom-up models	✓	-	✓	✓	
	6.2 Inverse modelling N ₂ O, CH ₄	✓	-	✓	✓	
	6.3 Uncertainties in N and GHG flux	-	-	✓	✓	
	6.4 Verification of inventories	✓	-	✓	-	

Table 1. Data activities and data centres of the NEU science components.

3. Data management infrastructure, data centres, web portal and operations

The NEU science activity is divided into six components that will deliver different types of data to the NEU distributed database located at three dedicated data centres (DCs), namely CEH Edinburgh (in future managed by the CEH Environmental Informatics Programme (EIP) at Lancaster), UK, the French National Institute for Agricultural Research (INRA) at Thiverval-Grignon, France, and the Joint Research Centre (JRC) at Ispra, Italy (AFOLU DATA Portal, http://afoludata.jrc.it/data_fs.cfm). Table 1 shows which main DC is allocated to each science component. The DC's responsibilities include storing, checking for completeness, maintenance and distribution of NEU data.. Data centres liaise with the different science components through the Component Data Managers (CDMs), who will organise the data handling and supervise data submission within their component, but data can be submitted to the DCs directly by the individual investigators.

Each science component has a Component Data Manager (CDM), whose responsibilities include liaison with the scientists and the Data Management Committee (see below), overseeing data collection and data quality checks, providing support to investigators in issues related to data formatting and submission.

The NEU Data Management Committee (DMC) co-ordinates and supervises all data management activities, liaises with the science Component Data Managers, ensures that the Data Policy is applied and makes decisions regarding its implementation. The DMC issues a NEU Data Management Plan where details of this implementation are given. The DMC members confer at regular intervals. The DMC is composed of

- the six component data managers,
- a representative of each DC,
- the web portal responsible,
- the NEU Scientific Project Manager,
- the NEU Co-ordinator or his representative.

The DMC issues a formal written annual report to the SSC. The DMC may form Task Forces to aid its work (cf. Activity 7.4 of the Description of Work).

A central NEU data web portal has been developed at CEH (<http://www.nitroeuropa.eu/>). The portal includes a **Data** thumbnail leading to a page that links to the data centres, where submission deadlines and data related news will be posted. There are links to related sites and to the online support offered by the data centres. The website is the primary source of information on NEU data management issues.

4. Database contents

The NEU databases will host all processed observational data produced by the project, together with any documentation pertaining to the data.

The NEU database will also host data resulting from plot scale modelling (C3). Model development is documented by individual partners using version control tools like Subversion. This tool allows also to freeze model versions. Frozen model source code, documentation as well as selected significant model simulation results as well as corresponding model input data and drivers will be stored first by the individual partners and later in the JRC database (AFOLU data portal for Terrestrial Ecosystem Modeling).

In recognition that raw data represent a potentially valuable source of future science developments, including possible revision of the processed data, the Principal Investigators (PIs) agree to ensure that they are stored safely with the relevant processing software or, alternatively, with documentation on retrieval algorithms, at least for the retention period as defined below. Although not necessarily stored at one of the DC, they should be documented on the web portal. The DMC will examine the fate of raw data on a case per case basis and will ensure that in all cases they are kept for the long term in a way allowing future access and advise PIs on questions of raw data storage. The definition of 'raw data' will be detailed in the Measurement and Modelling Protocols and the DMP.

5. Formats

Data hosted by CEH and data hosted by INRA will be stored in the form of relational databases. Data are supplied in templates and will be uploaded to the CEH database by C1-C3 site managers, and to the INRA database by C4 site managers, prior to checking and validating by DMs.

Data hosted by JRC will be file-based and no specific restrictions are made on the file format. However, for geographic data in Component 5 data files should be compatible with the ESRI ArcGIS software and preferably be stored as shapefiles or raster (grid) based on the projection system Lambert Azimuthal Equal Area (ETRS/LAEA, centre of projection: 52N, 10E) as

proposed under INSPIRE. Images, text files and model output can be stored in the original format.

6. Metadata

Metadata (i.e. data about the data) are a crucial element of a data archive. They allow the data to be searched (“discovery” metadata), read by humans or software, understood, interpreted and used. Metadata includes supporting documentation (collection methods, algorithms, model parameterisations, references, plots, pictures, etc.) which will be stored alongside the data. Source codes can be stored if they help the understanding of the stored model output; in this case, it is preferable to store also a set of standard model input.

Metadata for C1 and C2 are defined in the data templates and will be stored on the CEH data centre database with the data.

Metadata for C3 will be stored in the JRC database (AFOLU) data portal for Terrestrial Ecosystem Modeling.

Metadata for C4 will be defined in the data templates and will be stored on the INRA data centre database with the data.

Meta-Data for C5 and C6 will be stored in the JRC database (AFOLU), which is designed specifically for NEU spatial metadata. Upload of a dataset will be restricted to those data with complete metadata information (based on ISO 19155). To ensure completeness of the metadata, a specifically designed metadata editor will have to be used, which is offered in two versions, one linked to ArcCatalog facilitating automatic retrieval of geographical information and a stand-alone (more information at: http://afoludata.jrc.it/NEU/editor_help.html). The metadata format is XML.

7. Data file names

A common file name convention provides the database some homogeneity and ease of use. These are listed in Table 2.

TABLE 2 – DATA FILE NAMING CONVENTIONS

	Sub-division	Raw data file naming convention for EXCEL data templates
C1	Level 1 sites	Each lab manages between 3-12 sites, measurements on a monthly basis, data to be submitted on a 6-monthly basis (March 31 st for July-Dec data; Sept 30 th for Jan-June data) File names: countrycode_sitename_YYYYMMDD.xls where sitename is full name of the site e.g. NI_carlow_20080206.xls, and YYYYMMDD is the date of data submission.
	Level 2 sites	eco_sitename_YYYYMMDD_extra.xls eco = ecosystem type (F = Forest, A = Arable, S = Shrubland/wetland, G = Grassland sitename = your sitename e.g. UK_EBu for Easter Bush. See list in Frequently Asked Questions if unsure of your sitename YYYYMMDD = the earliest date when data reported in this file were recorded.
	Level 3 sites	extra = optional field that can be used to add some file specification (e.g. a version or part number) or distinguish two files with the same name beginning e.g. G_UK_EBu_20060801_01.xls would be the first file in a set from the Grassland site at Easter Bush, where the data was collected began on 1st August 2006. Other files in the set may contain other variables which were not ready by the submission deadline. See cell A35

		below for a list of site ecosystem codes and sitename codes.
C2		eco_sitename_YYYYMMDD_extra.xls e.g. A_Petrodolinskoe_20061201_ver3.zip would be the first file in a set from the Arable site at Petrodolinskoe, where the data was collected began on 1st December 2006. Other files in the set may contain other variables which were not ready by the submission deadline. ver3 indicates that this was the third version of the file submitted.
C3		Conventions will be decided until next annual meeting. However, it is more complex compared to C1 and C2 since C3 will store a full set of files (e.g. input, code, results for the sites simulated by application of different models).
C4		Conventions for file name were described in previous versions of the Data Management Plan. Since the DMP was changing and no C4 data have been submitted yet, no new conventions have been defined for C4. This point will be discussed at the next C4 meeting on landscape data (June 2008) prior the first submission of C4 data (September 2008). As far as possible, conventions will be harmonised with other NEU components.
C5		No naming conventions specified for incoming files, to avoid rejection of files due to inadvertent mis-naming. On receipt, files to be renamed, probably “dataset_name” & “version” & “upload_date” & “extension” & “time_frame”
C6		Naming conventions will be decided in the coming half-year

8. Data submission (largely for measurement data)

For components 1 and 2, data and the associated meta-data are generated and collated by site managers, and will be entered onto the EXCEL templates specific for each component. The EXCEL data templates will be uploaded to the CEH database according to the schedule in Table 3. The Data Managers will support the site managers in this activity, and will check the data further before making it available as “validated data” on the NEU website.

For component 4, data and the associated meta-data are generated and collated by site managers, and will be entered onto the EXCEL templates for field verification measurements and onto the ACCESS templates for farm data. The EXCEL and ACCESS data templates will be uploaded to the INRA database according to the schedule in Table 3. The Data Manager will support the site managers in this activity, and will check the data further before making it available as “validated data” on the NEU website.

TABLE 3 – SCHEDULE FOR DATA SUBMISSION AND VALIDATION DEADLINES

Month	Month #	COMPONENT - MANAGER - DEADLINE ACTIVITY
Apr 2007	15	COMPONENT 3: Observation data synthesis → NEU web AND → core models COMPONENT 5: DB set up by WP 5.1 → JRC AND → C3 European upscaling of detailed ecosystem models
May 2007	16	COMPONENT 1.5: Site Manager: L1 Aug-Dec06 submission → CEH COMPONENT 1.3: Site Manager: L3 Aug-Dec06 submission → CEH
Jun 2007	17	COMPONENT 1.5: Data Manager: L1 Aug-Dec06 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Aug-Dec06 data verification → NEU web

Sep 2007	20	COMPONENT 1.5: Site Manager: L1 Jan-Jun07 submission → CEH COMPONENT 1.3: Site Manager: L3 Jan-Jun07 submission → CEH COMPONENT 2 : Site Manager: Historical and Manipulation data Jul06-Jun07 submission → CEH
Oct 2007	21	COMPONENT 1.5: Data Manager: L1 Jan-Jun07 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jan-Jun07 data verification → NEU web COMPONENT 2 : Data Manager: Jan-Jun07 Manipulation data verification
Mar 2008	26	COMPONENT 1.5: Site Manager: L1 Jul-Dec07 submission → CEH COMPONENT 1.3: Site Manager: L3 Jul-Dec07 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jul-Dec07 submission → CEH
Apr 2008	27	COMPONENT 1.5: Data Manager: L1 Jul-Dec07 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jul-Dec07 data verification → NEU web COMPONENT 2 : Data Manager: Jul-Dec07 Manipulation data verification → NEU web
Sep 2008	32	COMPONENT 1.5: Site Manager: L1 Jan-Jun08 submission → CEH COMPONENT 1.3: Site Manager: L3 Jan-Jun08 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jan-Jun08 submission → CEH COMPONENT 4 : Site Manager: Jan-Jun08 data submission → INRA COMPONENT 5 : Data Manager: Results from first model runs → JRC
Oct 2008	33	COMPONENT 1.5: Data Manager: L1 Jan-Jun08 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jan-Jun08 data verification → NEU web COMPONENT 2 : Data Manager: Jan-Jun08 Manipulation data verification → NEU web COMPONENT 4 : Data Manager: Jan-Jun08 data verification → NEU web
Mar 2009	38	COMPONENT 1.5: Site Manager: L1 Jul-Dec08 submission → CEH COMPONENT 1.3: Site Manager: L3 Jul-Dec08 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jul-Dec08 submission → CEH COMPONENT 3 : Data Manager: European upscaling - Detailed ecosystem model results → JRC COMPONENT 4 : Site Manager: Jul-Dec08 data submission → INRA
Apr 2009	39	COMPONENT 1.5: Data Manager: L1 Jul-Dec08 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jul-Dec08 data verification → NEU web COMPONENT 2 : Data Manager: Jul-Dec08 Manipulation data verification → NEU web COMPONENT 4 : Data Manager: Jul-Dec08 data verification → NEU web
Jun 2009	41	COMPONENT 3 : Data Manager: Core manipulation site simulations → JRC
Sep 2009	44	COMPONENT 1.5: Site Manager: L1 Jan-Jun09 submission → CEH COMPONENT 1.3: Site Manager: L3 Jan-Jun09 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jan-Jun09 submission → CEH COMPONENT 4 : Site Manager: Jan-Jun09 data submission → INRA
Oct 2009	45	COMPONENT 1.5: Data Manager: L1 Jan-Jun09 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jan-Jun09 data verification → NEU web COMPONENT 2 : Data Manager: Jan-Jun09 Manipulation data verification → NEU web COMPONENT 4 : Data Manager: Jan-Jun09 data verification → NEU web
Mar 2010	50	COMPONENT 1.5: Site Manager: L1 Jul-Dec09 submission → CEH COMPONENT 1.3: Site Manager: L3 Jul-Dec09 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jul-Dec09 submission → CEH COMPONENT 4 : Site Manager: Jul-Dec09 data submission → INRA
Apr 2010	51	COMPONENT 1.5: Data Manager: L1 Jul-Dec09 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jul-Dec09 data verification → NEU web COMPONENT 2 : Data Manager: Jul-Dec09 Manipulation data verification → NEU web COMPONENT 3 : Data Manager: European upscaling - High-temporal resolution Model results → JRC COMPONENT 4 : Data Manager: Jul-Dec09 data verification → NEU web
Sep 2010	56	COMPONENT 1.5: Site Manager: L1 Jan-Jun10 submission → CEH COMPONENT 1.3: Site Manager: L3 Jan-Jun10 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jan-Jun10 submission → CEH
Oct 2010	57	COMPONENT 1.5: Data Manager: L1 Jan-Jun10 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jan-Jun10 data verification → NEU web COMPONENT 2 : Data Manager: Jan-Jun10 Manipulation data verification → NEU web
Jan 2011	60	COMPONENT 3 : Core sites Simulation results → JRC
Apr 2011	63	COMPONENT 1.5: Site Manager: L1 Jul10-Jan11 submission → CEH COMPONENT 1.3: Site Manager: L3 Jul10-Jan11 submission → CEH COMPONENT 2 : Site Manager: Manipulation data Jul10-Jan11 submission → CEH COMPONENT 3 : Data Manager: GHG Scenario results → CEH
May-11		COMPONENT 1.5: Data Manager: L1 Jul10-Jan11 data verification → NEU web COMPONENT 1.3: Data Manager: L3 Jul10-Jan11 data verification → NEU web COMPONENT 2 : Data Manager: Jul10-Jan11 Manipulation data verification → NEU web

Footnote to Table

In order to warrant continuity to the research activities within NEU, it is essential that data produced by one NEU team be delivered as soon as possible to the other NEU groups via the dedicated DC. Datasets collected over periods of time of six months will be submitted within the following three months, according to the schedule defined in the Table and published on the NEU website. It will be the component data managers' responsibility to remind data suppliers of submission deadlines, to ensure that these are observed and to chase possibly late or missing data.

Delivery dates for C5 and C6 are “internal” and not prescribed as for other components. C5 dates decided on a case-by-case basis depending on component deliverables.

9. Data validation and quality checking

This account is adapted from the “Modality Solutions Specification” for the CEH database, primarily for C1 and C2 data, and it will be adapted for C4 data:

Excel upload

Each data entry form will support the upload of data from an excel document, as long as it is in the same format as a provided template.

Excel documents will be uploaded to the web database application, saved and processed.

The excel import procedure will:

1. Validate the workbook against the expected format.
2. Work through every form field, reading the data from the uploaded excel sheet
3. Each field will be checked for field format (numeric, date/time, etc).
4. Each field will then be checked against any validation rules (required / range).
5. Each field will then be checked for uniqueness – the same value uploaded for the same date for the same site
6. No Value - Excel documents with values in a field, or within a range of fields, will be able to specify “no value” – the system will allow the customization of the “no value” entry. Proposed options are: “NaN” or “No Value”. When a no value field is uploaded, it is not validated, and is not available for reporting.
7. Finally, the field value will be saved

An automatically generated report will be presented to the user, showing and validation errors (distinguishing between errors that require action and warnings), or successes on import. Finally, a web-based version of the form will be shown to review the data.

All excel documents uploaded will be saved by the application, so the original data will be stored.

In addition, the following checks will be implemented in the future:

1. The EXCEL data file name and extension complies with the file naming conventions
2. Column header names will be checked (to detect if extra columns have not been added between them.)
3. Check units
4. Cross check that methodology has been recorded

It is the PIs’ responsibility to perform required calibration and validation prior to data submission to the NEU database, and ensure that the data are of the best possible quality and include error estimates and/or flags as defined in the Measurement and Modelling Protocols. It will be one of the roles of the Component Data Managers to support and supervise quality checks.

Completeness of the data centres’ databases, in terms of completed templates and predefined aspects of template contents, will be checked by the designated DC at the time of submission. Completeness checking will be automated where appropriate. Suitable software will be developed at the DC for this purpose, with detailed guidelines from the NEU Component Data Managers and additional support from the PIs. A submission report will be generated automatically, whereby defective datasets, accompanied by error diagnostics, will be provided to the data originators for improvement, and missing data will be signalled. The ultimate responsibility for data quality lies with the originators of the data.

No data checks on the data content will be performed on the spatial data stored in the DC at JRC. However, metadata are required to be compiled using the NEU-AFOLU DATA metadata editor, which checks whether all obligatory fields have been filled. Only datasets which passed this check will be uploaded.

10. Access to data

Access to NEU datasets in general will be restricted to NEU participants and particular collaborators during a retention period of 5 years after the submission due date or 2 years after the project end date, whichever occurs first.

The NEU web portal (<http://www.nitroeuropa.eu>) is the central entry point to access all NEU datasets. A special “DATA” tag will lead to a centralised page providing information about the DMC and data management related issues and news, on the web file manager built into the portal to share documents between NEU partners, templates and other information based on this DMP.

In addition, this will be the entry point to the databases hosted by the data centres (see Section 3).

For the general public, information about which data is collected and held in the data centres will be made available through the option to browse metadata describing the database content. In addition to that, contact information for data originators and IPR holders will be provided, in case NEU external parties are interested in accessing specific datasets. Access to NEU datasets will only be granted to external parties on a by-case basis and by the IPR holder of the specific dataset.

After the retention period, the NEU data will be released to the public domain, but the NEU Conditions of Use (see Annex below) will continue to apply.

11. Support to the NitroEurope researchers

The DCs and the NEU website administrator, advised and assisted by the DMC, will endeavour to provide support to the NEU scientists in all data related issues. These may include the following.

- Negotiation, retrieval, provision of 3rd party data (e.g. NRT data to support field campaigns).
- Set up of data uploading/downloading system (ftp, web).
- Online documentation (scientific and technical). This includes the Cookbook, and guides for using the databases. <http://www.nitroeuropa.eu/>
- Data catalogue and search engine, including links to the database.
- Data extraction, comparison, visualisation tools.
- Format conversion (e.g. into Excel spreadsheets).
- Web based protected workspace that would provide a forum for discussions, collaboration, exchange of preliminary data, etc.

12. List of acronyms

CEH	Centre for Ecology and Hydrology
CDM	Component data manager
DC	Data centre(s)
DMC	Data management committee
EIP	Environmental Informatics Programme
JRC	Joint Research Centre
NEU	NitroEurope

Conditions of Use of NEU IP Data

These data were generated or collected within the framework of the NitroEurope IP. The user agrees.

- to restrict the use of the data to the context of the research topic specified at the time of the application, when this application was made to access data still restricted from the public domain;
- not to disclose the data to other parties;
- not to use the data for commercial purposes;
- that the Intellectual Property Right remains with the Data Originator;
- to contact the Data Originator prior to any use of the data;
- to offer the Data Originator(s) co-authorship of any publication or communication based on NitroEurope IP data; in the event that the offer is declined or when the Data Originators cannot be contacted, Data Originators must be duly acknowledged.