



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

D.7.5

Report on NEU databases and -centres for guidance of site operators and NEU partners

Due date of deliverable: **01/09/2008**

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	X
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 Integrated Project (NEU IP)
Data Management
Annual Report of the Data Management Committee
to the Science Steering Committee

February 2009
Last update 10 Feb 2009

This document summarises the data management activities and data centre developments that took place between February 2008 and February 2009

Compiled by Sue Owen (CEH, Edinburgh)

Contents

I	Data Management Committee activities
II	Data Centre activities
III	Data Web Portal
IV	Data Submission updates
V	Acronyms
	Appendices

I. Data Management Committee Activities

Data Management Committee Members

Sue Owen	CEH	Chair, C1 data manager C1-C3 data centre manager
David Leaver	CEH	C1-C3 data centre consultant
Rory Wilson	Modality Solutions	C1-C3 data centre external contractor
Bill Bealey	CEH	NEU Web portal, C1-C3 data centre consultant
Klaus Larsen	RISO	C2 data manager
Ralf Kiese	FZK	C3 data manager
Frank Neidl	FZK	C3 data manager
Jean-Louis Drouet	INRA	C4 data manager, C4 data centre manager
Olivier Maury	INRA	C4 data centre manager
Gert-Jan Reinds	ALTERRA	C5 data manager
Arnoud Frumau	ECN	C6 data manager
Adrian Leip	JRC	C5-C6 data centre manager
Fulgencio Sanmartin	JRC	C5-C6 data centre consultant

The Data Management Committee met July 3rd, 2008 – July 4th, 2008 at Dragør, Copenhagen, and by videoconference on December 9th 2008. The C1-C3 data centre development team met formally on 12/03/08 and 20/10/08 with several ad-hoc meetings of this team during 2008, to discuss specific issues of the development of the database. In addition, there was a meeting of C1- C3 data managers and the full DMC at the NEU General Assembly at Gothenburg January 26-29th 2009. Action points, decisions and progress from all these meetings are outlined below.

II Data Centre Activity

1. C1-C3 Data Centre

This is currently on the point of release to the NitroEurope community, and the current Technical Specification is provided on the NEU website. Live demonstrations were given at the NEU GA in Gothenburg January 26th-29th 2009. The first release will be for Component 1 level 3 sites (mid-February 2009), soon followed by Component 2 (March 2009). C1 L1, C1 L2, Activity 1.4 and special topics will follow in turn, with full release anticipated by early summer 2009.

Additional features, particularly “cross-form reporting” which will enable reporting from C1L3 and C2 concurrently, are under development and are being tested on the UAT server. Data submitted to Data managers between May 2007 and March 2008 on EXCEL templates are being uploaded by the Data Managers. NEU scientists will continue to complete the same templates for future submissions, and they will use these to upload to the database (submission from September 2008 onwards).

There were specific C2 template problems related to the new database, as 4 forms for the 4 ecosystem sites for C2 can't be cross queried (at present) in the new database structure. Therefore, one common form, and hence, one common template are needed. A developer was contracted to write a macro to transfer all existing templates with data to the new C2 template.

2. C4 Data Centre

Component 4 generates farm data (from farm questionnaires), field measurements data (meteorology, network of atmospheric and hydrological measurements of concentrations and fluxes, vegetation and soil measurements) and spatial data (land use maps, soil maps...). Those measured data will be used to calibrate and validate the C4-NitroScape model. Because the three types of measured data and simulated data for C4 need to be stored at the same location, the new C4 Data Centre has been developed from March 2008 at the INRA/AgroParisTech Environment and Arable Crops Joint Research Unit (EGC). Harmonization between C4 and C1-C3 was an important issue for discussion. The C4 database is at an advanced state of development, and one database server will store farm questionnaire data (questionnaires already distributed), measurement data and spatial data. PHP is used to import EXCEL (measurement) and ACCESS (questionnaire) files onto a relational database, using the same data checks as those used for C1-C3 database. An action was noted to construct a table highlighting where field names and/or units are different between C1, C2 and C4. This will be posted on the NEU website. The following Data Centre Development Activities took place during 2008 and early 2009 (see Appendix A):

2008:

- January - agreement with the C4 activity leaders and the DMC
- March-June - establishing the C4 Data Centre at INRA
- June - agreement on the C4 Data Centre functionalities at the C4 data meeting
- September - final version of Access farm template
- final version of Excel field verification measurements template
- implementation of both final templates into the C4 Data Centre

From end September 2008: 1st submission of datasets (6 landscapes):

- field measurements: 4 out of 6 landscapes (more or less completed)
- farm data: 3 out of 6 landscapes (more or less completed)
- checking data (consistency with templates)

2009:

- Jan.-March - 2nd submission of datasets (6 landscapes): field + farm data

- collecting spatial data (maps)
- integrating data in the C4 Data Centre following checking then define outputs (views and graphs)

3. C5 C6 Data Centre

At this data portal, data (model results) with prescribed meta-data can be archived in any form, along with the model code (executable and source) and its proper description. Modellers may withhold permission to access the model but meta-information is accessible to all. Currently, the data portal stores only spatial data (eg for C3 multi-site model runs) as well as 'upload' of model versions.

A future development for the data portal is a "file handling manager", whereby a dataset can set permissions, add files, upload new versions, etc. This will be a decentralized procedure and under the responsibility of the data set owner.

There was a security issue on the C5 data portal where the site had been hacked through an SQL injection which caused some delay in the database development (though no data security problems arose). To avoid similar problems in the future, the data portal was rebuilt using CodeIgniter framework. It is provisionally available under <http://afoludata.jrc.ec.europa.eu/codeigniter/index.php/login> and will soon replace the "old" data portal under http://afoludata.jrc.ec.europa.eu/data_fs.cfm. Both old and new databases now have multiple security checks. Despite this delay, the AFOLU DATA portal was already used substantially by partners of the NitroEurope-IP Component 5 to download preliminary datasets needed to work on their respective contribution to the work programme. Multifile upload will be implemented soon. Other new functionalities will be announced on the NEU data portal as they become available.

Selected additional features of the data portal:

- further development of the meta-data editor, including the generation of meta-data for models, which must be archived in the C5/C6 database;
- interface for uploading new data sets by registered users of the data portal including selection of NEU partners to whom access to the data is granted;
- automatic generation of a data disclaimer that a user must agree to before downloading data – the data disclaimer is created as a pdf file, and must be stored by the user (a copy of this file is also sent to the AFOLU responsible person).

4. Single Sign-On and Registration for databases

There was discussion on the possible implementation of Single Sign-On (SSO) bespoke system for each of the three NEU databases via the web portal, or use of OpenID. Further discussion via email and a final meeting at the NEU GA at Gothenburg resulted in a decision that SSO is not needed, because (i) the databases are very different and managed independently with different security issues, (ii) relatively few scientists (if any) would need to access more than one database at any given time, (iii) individual scientists can chose to use the same username and password for each of the three databases. However, access to all databases from a single location on the NEU web-pages is necessary for a unified image.

Registration for C4 and C5/C6 databases is already implemented. A registration facility is under development for C1-C3, with automatic transfer of the NEU community user-id and passwords from the NEU database to the C1-C3 database.

5. Training for using databases

It was agreed that there is no need for formal training on the C1-C3 database system as many of the user functions are fairly intuitive. However, demonstrations were given to 35 individuals at the Gothenburg NEU GA where “Guides to using the NEU database” were distributed. These training user guides and a video screencast focusing on the upload and reporting components will be available on the NEU web portal on release of the databases. In addition it was suggested that there should be an ‘awareness’ campaign to all NEU users. This takes the form of a few slides to be presented at any future NEU meetings, and visibility on the NEU web portal.

III. Web Portal

The web portal CMS software was upgraded in July/August to improve the functionality and security of the site. Dissemination activities including the posting of news, events and jobs are carried out regularly. The portal functions as an important repository for reporting documents, templates, component minutes, agendas, bibliographies and presentations. The registration process for the Gothenburg GA was carried out this year again through the portal with no issues to report. The use of the forums is a little disappointing and few members have used this as a medium for communicating within, and across, their programs. News about the database releases is clearly presented on log-in, and also in the “Data” tab. A downloadable file of a description of data available (metadata) in the C1-C3 database is available from the NEU web pages for the general public. Further development this year will include the updating of the Site page to include C2 and C4 sites, and the inclusion of sites to be filtered by ecosystem type. The web portal will also be used by the Gender Action Committee in electing their representatives by the use of an e-poll.

IV. DATA Submission Updates

1) Component 1

L1 sites

All data for 2007 were delivered in April 08. These data will be uploaded to the database before early summer 2009.

L2 sites

These are just starting to operate and a template for this data will be designed based on Level 3 site template

L3 sites

All sites submitted some data for 01 August-31 December 2006, all sites except Speulder submitted some data for 01 January – 30 June 2007, and all sites except Speulder and Easter Bush submitted data for 1 July – 31 December 2007. Speulder and Easter Bush will submit outstanding data shortly. Data for January – December 2008 will be submitted directly to the database shortly. Table 1 gives an overview of Data Submission compliance for C1L3 sites.

Table 2 ranks sites according to a “submission score”. This is based on number of submission dates where some data were submitted, number of time-series sheets with some data submitted, percentage of essential fields across all submissions, and is given by:

$$\text{Score} = \sqrt{(\# \text{ submissions})^2 + (\# \text{ time-series sheets with some data submitted})^2 + (\% \text{ essential fields across all submissions})^2}$$

This score is an *indicator* only of site compliance with data submission targets.

2) Component 2

108 historical templates were received and a contractor employed to help process the templates to fit the new master template. One notable change has been the closure of Pianosa which has not been replaced by site Beano. There is still outstanding data from Logården, Falköping, Männikjärve, Oldebroek and Beano.

Table 3 gives an overview of data submission for C2 up to 2008.

BB to make sure that up and coming C3 workshop/meetings are added to the NitroEurope portal events calendar.

3) Component 3

Model runs will be made for the different sites during 2009. Therefore, data is required from the C1-C3 database, and frozen model code, model output and validation data will be stored at the C5 database. C3 together with C5 is currently developing a first draft of a protocol for the submission of simulation results. A number of workshops were planned wetland sites, arable and forests ecosystems.

4) Component 4

Data from the six NEU-C4 experimental sites are sent to C4 Data Manager (maury@grignon.inra.fr, drouet@grignon.inra.fr) by e-mail or by the document sharing interface according to the schedule defined in the Data Management Plan (Table 4).

Since data are submitted only twice a year, data are uploaded to the database by C4 Data Manager with no “write access” to any other member of the project. The fraction of data delivered for the first submission period (September 2008) in comparison with data required in templates (for field measurements and farm data) varies between landscapes (Table 5). GIS maps (ESRI shapefiles) will be submitted from March 2009 and uploaded on the file sharing space.

The contribution of the Component 4 Data Centre activities to the C4 annual report is included in Appendix A

Components 5 and 6

The “preliminary data sets” with environmental and farm management information are available at the data portal.

The final versions of the data sets (to be delivered by activities 5.1-5.3) are partly concluded, but several parts are delayed and will be made available at the data portal in month 42 at the latest.

V. Acronyms

C1 ... C6 NEU Science Components 1 to 6
CDM Component Data Manager

CEH Centre for Ecology and Hydrology
DC Data Centre
DMC Data Management Committee
DMP Data Management Plan
DP Data Policy
EC European Commission
GA General Assembly
INRA French National Institute for Agricultural Research
IPR Intellectual Property Right
JRC Joint Research Centre
NEU IP NitroEurope Integrated Project
SSC Science Steering Committee

Table 1
Component 1 Submission Compliance 2006 - 2007

Site	C1 L3 mean % of essential data/sheet submitted		
	Aug 2006- Dec 2006	Jan 2007- Jun 2007	Jul 2007- Dec 2007
Bugac	47	48	50
Easter Bush NE	52	50	0
Easter Bush SW	52	50	0
Borgo Cioffi	47	49	57
Castellaro	44	46	53
Gebesee	36	47	42
Auchencorth	36	32	39
Grignon	57	60	57
Höglwald	44	38	0
Hyytiälä	55	50	62
Lompolojänkkä	41	40	56
Oensingen	72	77	82
Soroe	41	58	73
Speulder	17	0	0

Table 2

Component 1 “Submission Score” 2006 – 2007 for each site, based on number of submission dates where some data were submitted, number of time-series sheets with some data submitted, percentage of essential fields across all submissions (January 26th 2009).

	SUBMISSION SCORE
	(Top possible score = 3)
Oensingen	1.60
Grignon	1.52
Bugac	1.49
Castellaro	1.49
Auchencorth	1.42
Gebesee	1.42
Hyytiälä	1.40
Soroe	1.35
Borgo Cioffi	1.35
Lompolojänkkä	1.33
Easter Bush SW	1.25
Easter Bush NE	1.00
Höglwald	0.79
Speulder	0.38

Table 3
Component 2 data submissions up to 2008.

Eco type	Site	Data		Percent of essential data		
		First	Latest	Min	Max	Mean
A	Encin	08/03/2004	30/06/2008	59	72	65
	Foulum	01/01/1996	25/04/2007	49	49	49
	Logården	no data yet	no data yet			
	Maulde	14/09/2006	30/06/2008	34	64	47
	Paulinenaue	31/01/1991	01/01/3908	13	51	41
	Petrodolinskoe	10/10/2006	30/06/2008	25	49	35
	Tulloch	08/12/2002	08/09/2008	3	59	34
	Zimbabwe	26/09/2006	30/04/2008	64	70	66
	Beano	no data yet	no data yet			
S	Brandbjerg	21/09/2004	30/06/2008	48	67	57
	Clocaenog	01/01/1998	18/09/2007	76	76	76
	Fajemyr	01/01/2007	01/08/2008	26	31	29
	Garraf	12/02/2003	18/10/2008	17	38	27
	Mols	25/02/1999	15/05/2007	10	53	22
	Männikjärve	no data yet	no data yet			
	Oldebroek	16/12/1998	21/02/2002	52	52	52
	Storflaket	01/01/2007	01/08/2008	24	28	26
	Whim	01/01/2002	31/08/2008	17	29	23
F	Achenkirch	01/07/2001	30/06/2007	66	67	66
	Alptal	01/01/1994	19/06/2008	66	66	66
	Falköping	no data yet	no data yet			
	Gaardsjoen	01/06/1979	16/09/2008	76	76	76
	Herdada de Mitra	01/01/2003	01/01/2007	33	59	46
	Hoeglwald	01/01/1994	31/12/1997	NA	NA	NA
	Klausenleopoldsdorf	01/01/1994	28/08/2007	12	83	34
	Klosterheden	01/01/1983	12/06/2008	5	33	14
	Skogaryd	14/07/2006	30/10/2007	9	9	9
	Strødam	01/01/2007	24/06/2008	5	12	7
	Tolfa	01/01/2005	30/06/2008	12	19	16
	Vestskoven	27/03/2007	16/06/2008	5	5	5
G	Chrichton	29/03/2006	11/12/2006	NA	NA	NA
	Gödöllö/SZIU	01/10/2002	06/10/2008	28	40	33
	Nafferton	01/10/2006	30/06/2008	23	49	32
	Peaknaze (Plynlimon)	01/10/2002	14/11/2007	60	60	60
	Rzecin	01/01/2004	01/01/2008	34	52	42
	Theix	01/04/2005	30/06/2008	45	57	51

Table 4
 Schedule for Component 4 data submission and validation (DM: Data Manager, DC: Data Centre)

Submission deadline	Project month	Responsibility	Actions on data	Submission / Upload to
September 2008	32	C4 Landscape Manager	January – June 2008 submission	C4 DM
October 2008	33	C4 DM	January – June 2008 validation	C4 DC
March 2009	38	C4 Landscape Manager	July – December 2008 submission	C4 DM
April 2009	39	C4 DM	July - December 2008 validation	C4 DC
September 2009	44	C4 Landscape Manager	January – June 2009 submission	C4 DM
October 2009	45	C4 DM	January – June 2009 validation	C4 DC
March 2010	50	C4 Landscape Manager	July – December 2010 submission	C4 DM
April 2010	51	C4 DM	July - December 2009 validation	C4 DC

Table 5
 Component 4 Data submission status for the first submission period (September 2008)

Landscape	Field measurements data	Test farm data
Italy – Piana del Sele	No data	Yes
Denmark – Bjerringbro	Completed	No data
The Netherlands – NFW	No data	No data
France – Kervidy-Naizin	To be completed	Yes
Poland – Turew	Completed	Yes
UK – Southern Scotland	To be completed	No data

Appendix A

Contribution of C4 Data Centre to the C4 Annual report

Report from the C4 Data Centre – February 2009
Jean-Louis Drouet, Olivier Maury

1. Need for a C4 Data Centre

Component 4 generates farm data (from farm questionnaires), field measurements data (meteorology, network of atmospheric and hydrological measurements of concentrations and fluxes, vegetation and soil measurements) and spatial data (land use maps, soil maps...). Those measured data will be used to calibrate and validate the C4-NitroScape model. Because the three types of measured data and simulated data for C4 need to be stored at the same location, the NEU coordinator proposed to establish a separate C4 Data Centre at INRA, France (November 2007). This was agreed with the Data Management Committee in January 2008. The new C4 Data Centre (Fig. 1) has been developed from March 2008 at the INRA/AgroParisTech Environment and Arable Crops Joint Research Unit (EGC). Its structure and functionalities was agreed with the C4 partners in June 2008. The following reports the main features of the C4 Data Centre (see NEU_C4_DC_functionalities_080605.doc, for more details).

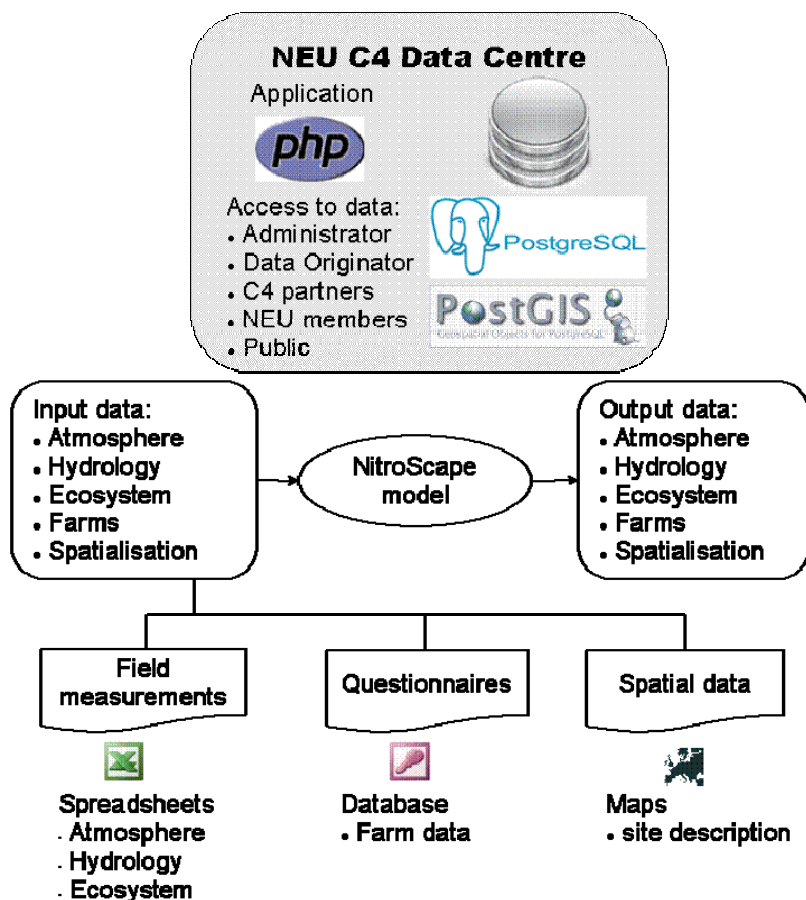


Fig. 1. Types of data generated by NEU-C4.

2. C4 database structure

Data are collected on the INRA/AgroParisTech EGC database server (Fig. 2) into an OpenSource Relational Database Management System (RDMS): PostgreSQL. PostGis GIS functions (<http://postgis.refractions.net/>, <http://postgis.fr/>) will be added to the PostgreSQL database. The database structure matches the one of the template files which were completed in September 2008 for field measurements (Excel file: NEU_C4_DataTemplate_FieldMeasurements_20080901_v2) and farm data (Access file: neu_c4_landscape_db_v7). Export files have the same structure as template files. Each intervention on the database server is consigned in a daily log. A multilingual web portal (in French and in English by default) has been created to consult statistics and access to data under conditions of use (see section 4).

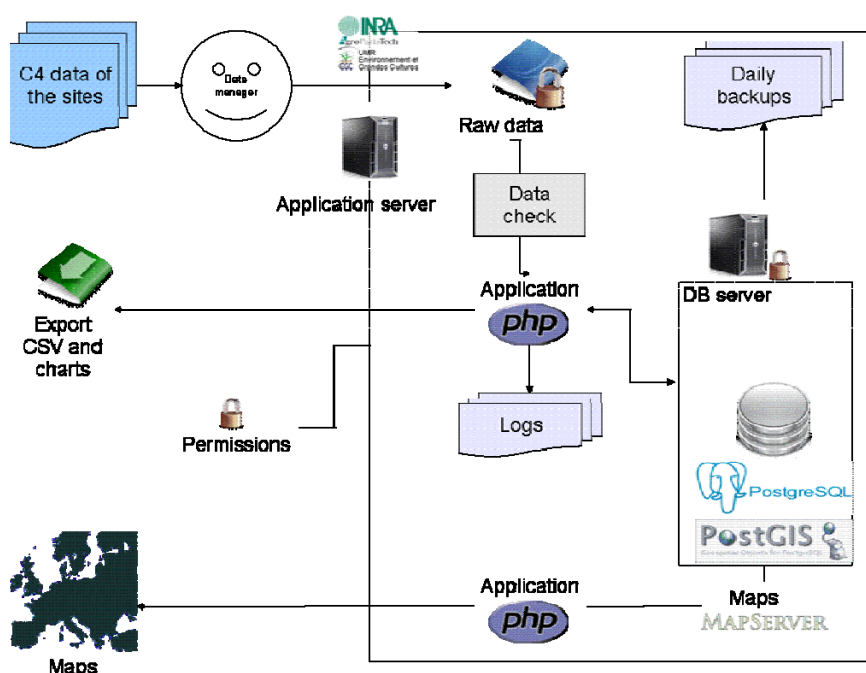


Fig. 2. Data fluxes within the C4 Data Centre.

3. Data submission and data processing (validation, quality checking, uploading)

Data from the six NEU-C4 experimental sites are sent to C4 Data Manager (maury@grignon.inra.fr, drouet@grignon.inra.fr) by e-mail or by the document sharing interface according to the schedule defined in the Data Management Plan (Tab. 1).

Submission deadline	Project month	Responsibility	Actions on data	Submission / Upload to
September 2008	32	C4 Landscape Manager	January – June 2008 submission	C4 DM
October 2008	33	C4 DM	January – June 2008 validation	C4 DC
March 2009	38	C4 Landscape Manager	July – December 2008 submission	C4 DM
April 2009	39	C4 DM	July - December 2008 validation	C4 DC
September 2009	44	C4 Landscape Manager	January – June 2009 submission	C4 DM
October 2009	45	C4 DM	January – June 2009 validation	C4 DC

March 2010	50	C4 Landscape Manager	July – December 2010 submission	C4 DM
April 2010	51	C4 DM	July - December 2009 validation	C4 DC

Tab. 1: Schedule for data submission and validation (DM: Data Manager, DC: Data Centre)

Since data are submitted only twice a year, data are uploaded to the database by C4 Data Manager with no “write access” to any other member of the project. The fraction of data delivered for the first submission period (September 2008) in comparison with data required in templates (for field measurements and farm data) varies between landscapes (Tab. 2). GIS maps (ESRI shapefiles) will be submitted from March 2009 and uploaded on the file sharing space.

Landscape	Field measurements data	Test farm data
Italy – Piana del Sele	No data	Yes
Denmark – Bjerringbro	Completed	No data
The Netherlands – NFW	No data	No data
France – Kervidy-Naizin	To be completed	Yes
Poland – Turew	Completed	Yes
UK – Southern Scotland	To be completed	No data

Tab. 2: Data submission status for the first submission period (September 2008)

Excel and Access datasets are saved and processed.

Following conversion to CSV format and import to the database, the Excel process:

1. Validates the workbook against the expected form,
2. Works through every form field, reading the data from the uploaded Excel sheet,
3. Checks for field format (numeric, date / time...),
4. Checks for field uniqueness: the same value uploaded for the same date for the same site,
5. Checks for fields with “no value”,
6. Saves field values.

The Access datasets are validated against the whole database and integrated.

4. Security, storage and access to data

It is the responsibility of each C4 Site Manager to ensure that raw data are stored safely at least for the retention period. The C4 Data Centre will maintain the datasets and ensure the security of the database by making regular backups, doing the common maintenance tasks and managing the access rights. The C4 data server has autosaving process to local and distant discs. Backup set will be kept on read-only memory (CD-Rom, DVD) in another place than C4 Data Center.

According to NEU Data Policy, 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. (see NEU Data Policy, for more details). Because of confidentiality of data especially farm data, access to C4 datasets are given according to specifications from landscape partners. To access the

database, only the web application can be used. This use is given after requesting an account. The account is validated and access rights are granted according to Data Policy.