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 5.1.2 Methodology for deriving farm-scale input data for models at the European scale

Due date of deliverable: **31/07/2006**

Actual submission date: **10/07/2007**

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

Duration: **60 months**

Organisation name of lead contractor for this deliverable :**ALTERRA**

Revision: **Final version**

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

Executive Summary

The impact assessment of past and future changes in agricultural management on the emissions of N and GHG, requires information on rates and timing of manure and fertilizer application. Consequently, there is a need for consistent and updated information on livestock production, excretion rates, manure and fertilizer application in time and space, dates for sowing, ploughing, harvesting etc. This report describes the work that was carried out to assess available data sources and the methodologies to derive data in the temporal and spatial resolution required for the various models used in NEU. More specifically, the report contains a (i) review of data needs, divided in farm production and management data and model parameters, (ii) assessment of available data sources, divided in statistical agricultural data, agricultural models datasets and data for model parameters and (iii) methodologies to derive data, divided in re-sampling methodologies, statistical methods, expert knowledge, harmonisation and historical reconstruction of agricultural data.

Objectives
The main aim of this report is to present an overview of available farm production and farm management data. A secondary aim is to review the parameters that have been identified as model requirements. The report does not include either forest production or management or land use, which will be considered within the environmental data.

Activities
This work is part of Activity 5.1 “Harmonized GIS-based data development and assessment”. Within Activity 5.1, Task 5.1.2 focuses on “Analysis of farming data”. To evaluate of changes in agricultural management in the past and in the future on N and GHG emissions, we require, amongst others, information on rates and timing of manure and fertilizer application. This implies information on livestock production, excretion rates, manure and fertilizer application in time and space, dates for sowing, ploughing, harvesting etc. This report describes the work that was carried out to asses available data sources and the methodologies to derive data in the required temporal and spatial resolution for the various models used in NEU. The report is based on discussions during: (i) two internal meetings by Alterra in the Netherlands, (ii) a Meeting of component 5 in Garmisch-Partenkirchen (NEU kick-off meeting), March 2006 (ii) a meeting of components 5 and 6 in Wageningen, September 2006.Garmisch: General Assembly NEU (March 2006) and a focused meeting on this deliverable during a meeting of in Wageningen, November 2006.

Results
The report contains a
1. Review of data needs, divided in (i) farm production and management data and (ii) model parameters.
2. Assessment of available data sources, divided in (i) statistical agricultural data (ii) agricultural models datasets and (iii) data for model parameters.
3. Methodologies to derive data, divided in (i) resampling methodologies, (ii) statistical methods, (iii) expert knowledge, (iv) harmonisation and (v) historical reconstruction of agricultural data.

Finally, it contains a systematic overview for the data collection that has been done already and is still needed in NEU context

Milestones achieved
The attached report

Deviations and reasons
This deliverable has been delayed by six months. The reason for this is that more aspects were included in this report than initially foreseen, because they were considered crucial for an adequate application of models. This includes a description of: (i) reconstruction of agricultural livestock data in the past: (from 1900 – 2000 focusing on the period 1970-2000) and (ii) downscaling (disaggregation) of agricultural of livestock data to a 1km x1km grid-cell using an updated CLUE model approach.

Publications
The attached report:

Meetings
1 Internal meetings in the Netherlands (2)
2 Garmisch: General Assembly NEU (March 2006)
3 Wageningen (Sept 2006); part of C5
4 Wageningen (Nov 2006)

Methodology for deriving farm-scale input data for models at the European scale

Marta Pérez-Soba (Alterra)
Adrian Leip (JRC)
Wim de Vries (Alterra)
Gerard Velthof (Alterra)
Berien Elbersen (Alterra)
Klaas van de Hoek (RIVM)
Nicholas Hutchings (DIAS)
Magda Kahl (IMK-IFU)
Claire Campbell (CEH)

Table of contents

1. Review of data needs

This review is based on the input from D 5.1.1 'Report on consistent and comprehensive database requirements', which provides a complete overview of the data needs for all the NEU models and on the report about the model INTEGRATOR (in preparation). The review focuses firstly on the farm production and management data and does not include either forest production and management or land use, which will be considered within the environmental data. Secondly, it reviews the parameters that have been identified as model requirements.

1.1 Farm production and management data

1.1.1 Livestock production

Only the models INTEGRATOR and EPIC need information on livestock production systems. In general terms, the data requirements are: feeding situation of livestock, housing situation of livestock, and manure management systems (Table 1).

Table 1 Data requirements for livestock production systems

LIVESTOCK PRODUCTION Model	Animal numbers	Feeding situation	Housing situation
DNDC-regionalscale-Europe	0	0	0
DNDC-regionalscale-Germany	**x**	**0**	**0**
Forest-DNDC	**0**	**0**	**0**
INTEGRATOR	**+**	**+**	**+**
RothC	**0**	**0**	**0**
SUNDIAL-MAGEC	**0**	**0**	**0**
EPIC	**+**	**+**	**+**
CERES-EGC-Region	**0**	**0**	**0**
PaSim-plotscale	**0**	**0**	**0**
FASSET	**x**	**x**	**x**
DAYCENT-plotscale	0	0	0
DNDC-plotscale	0	0	0
SUNDIAL-MAGEC-plotscale	0	0	0
BASFOR-plotscale	0	0	0
COUP-plotscale	0	0	0
CERES-EGC	0	0	0
EMEP	0	0	0
PaSIM	0	0	0
EFM/BasFor	0	0	0

The following list indicates in detail the parameters required:

I. Animal manure excretion in housing systems and in grassland by grazing animals

- animal numbers per animal category
- N and C excretion factors per animal category
- number of housing days per animal category
- number of grazing days per animal category (complementary to number of housing days)

II. Housing and storage emissions

- animal housing system
- manure storage system
- type of manure in storage (solid manure or slurry)
- N emission fractions (fraction NH_3, NO_x or N_2O compared to N excretion) per animal category

- volume of manure excretion or manure excretion factor (kg/animal/yr) and the reciprocal of the bulk density of the manure (m^3/kg)
- CH_4 emission factors per animal category (kg CH_4/animal/yr)

III. N and C output
- animal production system
- animal nutrition per animal production system

1.1.2 Crop and grassland production and management

The most extensive needs of the models regard the *management of crop and grass production systems*. The data requirements cover nitrogen input through mineral fertilizer and manure, water management (irrigation, flooding), tillage, and data related to crop phenology (sowing date, harvest date, yield). Furthermore, management of crop residues is an important element for the simulation of the carbon balance. Table 2 to Table 4 show the data requirements submitted by the different models participating in NEU Component 3 and/or Component 5. Those models which explicitly have reported data needs for a regional-scale simulation run are shown in bold. From these tables it becomes clear that the requirement of GHG fluxes from soils process-based models that run at the regional scale do not need less information than those that run at the plot scale. EPIC and SUNDIAL-MAGEC require most information, especially for the parameterization of carbon input or return to the soil through management of crop residues and water management. Less input is required for the models RothC and INTEGRATOR, which both do not consider water management.

Table 2 Data requirements for crop and grassland models regarding mineral fertilizer and manure application.

CROP AND GRASSLAND NITROGEN MANAGEMENT		MINERAL FERTILIZER				MANURE					
Model	# of crops considered	TYPE	RATE	DATE	INCORP. DEPTH	TYPE	RATE	DATE	INCORP. DEPTH	TECHNIQUE	Manure C/N ratio
DNDC-regionalscale-Europe	**majorcrops**	**o**	**x**	**x**	**0**	**x**	**x**	**x**	**0**	**0**	**x**
DNDC-regionalscale-Germany	**x #15**	**0**	**x**	**x**	**0**	**x**	**x**	**x**	**0**	**0**	**x**
Forest-DNDC	**0**	**0**	**-**	**-**	**-**	**0**	**-**	**-**	**-**	**0**	**-**
INTEGRATOR	**0**	**+**	**+**	**0**	**0**	**+**	**+**	**0**	**0**	**+**	**0**
RothC	**0**	**0**	**0**	**0**	**0**	**0**	**+**	**0**	**0**	**0**	**0**
SUNDIAL-MAGEC	**0**	**+**	**+**	**+**	**0**	**+**	**+**	**+**	**+**	**0**	**0**
EPIC	**+**	**+**	**+**	**+**	**0**	**+**	**+**	**+**	**+**	**+**	**0**
CERES-EGC-Region	**0**	**0**	**X**	**X**	**X**	**0**	**X**	**X**	**X**	**0**	**0**
PaSim-plotscale	**0**	**0**	**-**	**-**	**-**	**0**	**-**	**-**	**-**	**0**	**-**
FASSET	**circa 20**	**x**	**x**	**x**	**(x)**	**x**	**x**	**x**	**(x)**	**x**	**(x)**
DAYCENT-plotscale	0	0	x	x	x	0	x	x	x	0	x
DNDC-plotscale	0	0	x	x	x	0	x	x	x	0	x
SUNDIAL-MAGEC-plotscale	0	0	x	x	x	0	x	x	x	0	0
BASFOR-plotscale	0	0	-	-	-	0	-	-	-	0	-
COUP-plotscale	0	0	x	x	x	0	x	x	x	0	x
CERES-EGC	0	0	X	X	X	0	X	X	X	0	0
EMEP	0	0	0	0	0	0	0	0	0	0	0
PaSIM	0	0	0	0	0	0	0	0	0	0	0
EFM/BasFor	0	0	0	0	0	0	0	0	0	0	0

Table 3 Data requirements for cropland models regarding water management, tillage and herbicide applications

CROPLAND WATER MANAGEMENT	IRRIGATION					TILLAGE		
	DATE	AMOUNT	N CONTENT	DRAINAGE	FLOODING	TYPE	DATE	DEPTH
DNDC-regionalscale-Europe	**0**	**x**	**0**	**0**	**x**	**x**	**x**	**0**
DNDC-regionalscale-Germany	**0**	**0**	**0**	**0**	**0**	**x**	**x**	**x**
Forest-DNDC	**-**	**0**	**0**	**0**	**0**	**0**	**-**	**0**
INTEGRATOR	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
RothC	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
SUNDIAL-MAGEC	**+**	**+**	**+**	**+**	**0**	**+**	**+**	**+**
EPIC	**+**	**+**	**+**	**+**	**0**	**+**	**+**	**+**
CERES-EGC-Region	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
PaSim-plotscale	**-**	**0**	**0**	**0**	**0**	**0**	**-**	**0**
FASSET	**x**	**x**	**(x)**	**0**	**0**	**x**	**x**	**(x)**
DAYCENT-plotscale	x	0	0	0	0	0	x	0
DNDC-plotscale	x	0	0	0	0	0	x	0
SUNDIAL-MAGEC-plotscale	0	x	0	0	0	0	x	0
BASFOR-plotscale	-	0	0	0	0	0	-	0
COUP-plotscale	x	0	0	0	0	0	x	0
CERES-EGC	0	X	0	0	0	0	X	0
EMEP	0	0	0	0	0	0	0	0
PaSIM	0	0	0	0	0	0	0	0
EFM/BasFor	0	0	0	0	0	0	0	0

Table 4 Data requirements for cropland models regarding sowing and harvest dates, actual and potential yield, and crop residue management

CROPLAND MANAGEMENT	SOWING DATE	HARVEST DATE	ACTUAL YIELD	POT. YIELD	CROP RESIDUES	C INCORP.
DNDC-regionalscale-Europe	**x**	**x**	**0**	**x**	**x**	**0**
DNDC-regionalscale-Germany	**x**	**x**	**0**	**0**	**x**	**0**
Forest-DNDC	**-**	**-**	**0**	**-**	**0**	**0**
INTEGRATOR	**0**	**0**	**+**	**0**	**+**	**0**
RothC	**0**	**0**	**0**	**0**	**0**	**0**
SUNDIAL-MAGEC	**+**	**+**	**0**	**x**	**+**	**+**
EPIC	**+**	**+**	**+**	**(x)**	**+**	**(x)**
CERES-EGC-Region	**0**	**0**	**0**	**0**	**0**	**0**
PaSim-plotscale	**-**	**-**	**0**	**-**	**0**	**0**
FASSET	**x**	**x**	**0**	**0**	**0**	**0**
DAYCENT-plotscale	x	x	0	x	0	0
DNDC-plotscale	x	x	0	-	x	0
SUNDIAL-MAGEC-plotscale	x	x	x	x	x	0
BASFOR-plotscale	-	-	0	-	0	0
COUP-plotscale	x	x	0	-	x	0
CERES-EGC	X	X	X	-	x	0
EMEP	0	0	0	0	0	0
PaSIM	0	0	0	0	0	0
EFM/BasFor	0	0	0	0	0	0

The following list indicates in detail the parameters required:

I. Crop production systems
- Number of crops
- Type of crop production systems

II. Animal manure and fertilizer input
- type of manure
- manure application date (to calculate ammonia emissions from soil)
- manure application rate
- manure application incorporation depth
- manure application technique (to calculate ammonia emissions from soil*)*
- manure C/N ratio
- type of fertilizer
- fertilizer application date
- fertilizer application rate
- fertilizer application incorporation depth

III. Water management, tillage and herbicide applications
- irrigation date
- irrigation amount
- irrigation, N content in water
- drainage
- flooding
- tillage type
- tillage depth
- tillage date
- herbicide application

IV. Sowing and harvest dates, actual and potential yield, and crop residue management
- sowing date
- harvest date
- dry matter yield (crop removal)
- max potential yield
- crop residues

1.2 Model parameters

Table 5 gives a list of parameters that have been identified as model requirements for all models. Model parameters that are specifically needed in INTEGRATOR to make the various N excretion and N application calculations used by the detailed ecosystem models include (these parameters belongs to both livestock and management data):
- Excretion factors for N and C per animal category
- N housing emission fractions (fraction NH_3, NO_x or N_2O compared to N excretion) per animal category and manure storage systems. (Belongs to livestock management data)
- CH4 emission factors per animal category (Belongs to livestock management data)
- N application emission fractions (fraction NH3, NO_x or N_2O compared to N input) for manure, grazing and fertilizer (depending on fertilizer type).

The list of data in Table 5 is much too extensive to realistically make a thorough comparison for all models. Furthermore, some of these parameters are built-in into the modelling concept of the various models and thus a comparison without the respective framework would not make sense. However, we aim to minimize differences in some key parameters for the nitrogen budget: nitrogen input via manure and nitrogen export by the crop. The relevant data required by all models include C and N content in manure and crops or crop residues and the biomass of crops (including grass) (see Table 6). For these data, a database will be put in place that facilitates the comparison of data used by the various models and highlight differences. However, particularly crop nitrogen content and partitioning over the plant compartments can be an inherent result of some models. These results can thus best be used as input for those models that can not calculate these data or have a less detailed methodology.

Table 5 Requirements for 'parameters' by models

Manure	
	DM content of manure; C content in DM of manure; N content in DM of manure; Ammonium content in manure; Organic N content in manure
Grassland	
	grassland functional types (fraction of legumes) (grass); shoot biomass (grass); root biomass (grass); leaf biomass (grass); peak biomass (grass); max LAI (grass); mean/ max rooting depth/ root distribution (grass); leaf, shoot, root C/N ratio (grass); max canopy height (grass);
Cropland	
	Nitrogen in plant (crops); N content in above-ground (crops); N content in below-ground (crops); N content in residues (crops); N content in harvest (crops); N content in debris (crops); N content in seed (crops); Biomass, shoot (crops); Biomass, root (crops); Biomass, grain (crops); Biomass, leaf (crops); Biomass, peak (crops); Dry matter in residues (crops); C content in residues (crops); C content in debris (crops); C/N ratio leaf, shoot, root, grain (crops); DPM/RPM (crops); Decomposition rate, biomass (crops); Decomposition rate, humus (crops); Rooting depth, actual (crops); mean/ max rooting depth/ root distribution (crops); Rooting depth, max. (crops); Rooting depth, increase rate (crops); LAI, max (crops); Light absorption (crops); Photosynthesis (crops); Yield, max (crops); Height, max (crops); Water requirement (crops); Harvest index (crops); Crop development, vegetative (crops); Crop development, productive (crops); Yield / crop ratio (crops); Leaf (crops); Stem (crops); CO2 (crops); Crop senescence, period (crops); Leaf development (crops); Crop senescence, max fraction (crops);

Table 6 Data requirements for cropland models for which a database will be made to ensure use of comparable data in the assessment of N input by manure and N output by crops

Data required	Model parameters
Manure C and N contents	DM content, C content; N content; Ammonium content; Organic N content
Crop yields/ biomass	Biomass of shoots, roots, grains, leaves, crop residues.
Crop C and N contents	C and N content in above-ground crops or shoots/ C and N content in harvested crops; C and N content in below-ground crops or roots, C and N content in crop residues/ crop debris; C and N content in seed/grain (or everywhere only N contents and then C/N ratio in shoots, roots, grain, crop residues).

2. Assessment of available data sources

2.1 Statistical agricultural data

2.1.1 Livestock and manure management data

At EU level the main available statistical data sources with information on (i) livestock (animal numbers per animal category), (ii) animal housing and (iii) manure/fertilizer application are FAOSTAT, REGIO, the Farm Structural Survey (FSS) and the Farm Accountancy Data Network (FADN). In addition, databases from existing models might also provide useful information for NEU, complementary to the formerly mentioned statistical datasets, i.e. CAPRI and RAINS/GAINS. A detailed description of the mentioned datasets, which are summarised in Table 8, is given below.

FAOSTAT is a database by the Food and Agriculture Organization of the United Nations (FAO). It contains a wealth of agricultural data, such as animal numbers for different animal categories, the amount of animal products (e.g. milk, meat and wool), fertilizer consumption, hectares of arable crops and the amount of crop products. The data are given for all countries and the time series ranges from 1960 till present. The fertilizer data concern also mineral nitrogen fertilizer.

Further, FAO publishes jointly with the International Fertilizer Industry Association (IFA), the International Fertilizer Development Centre (IFDC), the Phosphate and Potash Institute (PPI) and the International Potash Institute (IPI) a report on the country-wise fertilizer use by crop for 88 countries including all member states of the EU (FAO/IFA/IFDC/IPI/PPI, 2002). The sources of data in the publication are the following:

- FAO questionnaires sent to member countries;
- IFA questionnaires sent mainly to industry company members, research and development institutes and national and regional fertilizer associations and institutes.
- IFDC questionnaires sent to agronomists, economists and other participants in various fertilizer training courses, professional meetings and other seminars conducted by IFDC;
- IPI and PPI special contacts with specialized experts.

The data are a consensus of the involved organizations and therefore reflect the general magnitude maintaining consistency with the national FAO consumption data.

REGIO (regional databank of Eurostat) consists of the regional socio-economic statistics produced annually by Eurostat, which cover the main economic and social agricultural facts related to the EU. REGIO is a domain of Theme 1: General Statistics of the New Cronos database. The member States are spatially divided into statistical territorial units NUTS (*Nomenclature des Unités Territoriales Statistiques* in French; Nomenclature of Territorial Units for Statistics). The NUTS situation in 1995, before the large changes around 1999, subdivides the territory of the EU into 15 NUTS-0 regions, 77 NUTS-1 regions, 206 NUTS-2 regions and 1031 NUTS-3 regions, as shown in Table 7. The cattle, pig, sheep and goat populations are taken from the *EU livestock surveys* carried out by the Member States in December each year. However, for the Netherlands, United Kingdom and Denmark, the results of the December survey have been regionalised based on another survey carried out during that year.

Table 7 Regional land divisions of 15 EU countries. REGIO (Regional databank of Eurostat), FSS (Farm Structure Survey/EUROFARM of Eurostat) and FADN (Farm Accountancy Data Network of the European Commission, DG Agriculture).

Territorial codes used by different agricultural statistics	Source	Code used in source	Number of territorial units in EU15 (2003)
NUTS-0	REGIO	Member State	25
NUTS-1	REGIO	NUTS-1	89
NUTS-2	REGIO	NUTS-2	254
NUTS-3	REGIO	NUTS-3	1214
FSS-0	FSS/EUROFARM	Member State	25
FSS-1	FSS/EUROFARM	Region	89
FSS-2	FSS/EUROFARM	District	254
FADN-0	FADN	Member State	25
FADN-1	FADN	Region	140

Farm Structure Survey (FSS) is a domain of the Eurostat database provided by Eurostat. This source of information provides figures on structural characteristics of agriculture in the EU, classified by farm type, farm size and region. Data are available since 1975. The statistical unit is the agricultural holding, which is a single unit, both technically and economically, having a single management, producing agricultural products, and having either (i) a utilised agricultural area of 1 ha or more; or (ii) a utilised agricultural area less than 1 ha, provided it produces on a certain market scale or that its production units exceed certain natural thresholds. They are collected in agricultural census surveys every ten years (1990 and 2000), or intermediate sample surveys every two years (1993, 1995, 1997, 2003 and 2005). Data for basic surveys are available in a three level geographical breakdown of the whole country (FSS-0), the regions (FSS-1), and the sub-regions (FSS-2) as shown in Table 7 However, data for intermediate surveys are only available for FSS-0 and FSS-1 levels. The calendar for the basic and intermediate surveys is broad, the time intervals being, respectively 27 and 15 months. An interesting point is that organic farming is now also covered by the FSS and also the Member state returns on production areas and livestock numbers with respect to Regulation 2092/91.

Farm Accountancy Data Network (FADN): the European Commission (DG Agriculture) created the FADN database in order to provide information on the level of farm incomes and analyses of the effects of policy options. It includes information on costs and revenues, income generated from agriculture, including subsidies. Figures are available on an annual basis for the European Union as a whole, distinguishing between about 100 regions by farming type (NUTS 0 or 1). FADN is based on the annual accounting results for a sample of commercial farms in the EU Member States. Farms are classified as ‘commercial’ when they exceed a minimum economic size, measured in European Size Units (ESU). The total sample covered about 60,000 holdings in EU15 in 2000, but the sample is growing every year and the database is now being extended to the rest of the EU with new data expected fro a selection of new MS in 2006 already. Users of this database can work with individual farm data. The farms in the sample are rather heterogeneous. FADN stratifies farms according to region, economic size and farming type to reflect this heterogeneity adequately. Data at an individual farm level are confidential, so that only results aggregated for a group of farms are available. Nine types of farming are used in FADN publications based on the more detailed farming types specified in the Community typology of agricultural holdings. In 1995, there were 104 FADN regions, which varied in size from whole Member States in small countries (e.g. Denmark, Ireland and the Netherlands) to regions in large countries (e.g. four regions in Greece and 22 in France). The time period for data collection relates to a period of 12 months. Member States have accounting years starting on different dates and in some Member States the beginning of the accounting year is not the same for all farms.

The SEAMLESS project has recently reviewed some of the disadvantages connected to the FADN database (Elbersen at al. 2006). Firstly, the most important limitation is that the sample farms that occur in FADN might not represent all farming systems in the EU very well (or not at all). In total the FADN represents 52% of the farms and 86% of the Utilised agricultural area in EU-15, when compared to the data in the FSS. Secondly, a comparison with the FSS data also reveals that mixed livestock farms and beef cattle farms are not very well represented in FADN, though considerable differences occur between the different Member States. Thirdly, the major unit of data collection is the Utilisable Agricultural Area (UAA), not the area actually occupied by the agricultural business. Seasonal lets (common in some countries, such as Ireland) or wintering/summering arrangements, as well as the use of common land and the grazing of fallows, are excluded from consideration. For the spatial allocation of farms this aspect should be taken into account as it means that the influence on land use cannot be limited to the land owned or rented by the farm.

RAINS/GAINS: The RAINS/GAINS model has available an agricultural dataset at country level to predict NH_3, NO_x, N_2O or CH_4 emissions. The dataset includes: (i) number of housing days per animal category; (ii) volume of manure excretion per animal category; (iii) animal housing system; (iv) manure storage system; (v) manure categories (solid or slurry); (vi) type of fertilizers used and (vii) manure application techniques. In addition, it contains factor data such as: excretion factors for N per animal category; N (NH_3, NO_x or N_2O) emission fractions per animal category and manure category, CH_4 emission factors per animal category and N (NH_3, NO_x or N_2O) emission fractions for manure, grazing and fertilizer categories. These data are partly based on general IPCC factors but partly on country information, thus being more specific.

Table 8 NEU relevant statistical agricultural data sets at European level

Title data set	Description	Coverage (time, space & scale)	Sources & data holders
FAOSTAT	The FAO Statistical Database is an on-line multilingual database currently containing over 1 million time-series records from over 210 countries and territories covering statistics on agriculture, nutrition, fisheries, forestry, food aid, land use and population Relevant data for NEU include: (i) Animal numbers in 7 animal categories and (ii) Crop yields for over 100 crop types for the period 1961-2004.	*Time period and interval* Since 1960 up to present: yearly interval *Space and scale* Global (EU25+5) at country level	FAO
REGIO (EU Livestock Surveys)	The Regional Agricultural Statistics contains information on - land use (area); - crop production (area, production, yields); - livestock (nr animals and nr of Livestock Units (LSU)[1]; - production of cows' milk on farms; - agricultural accounts at regional level; - structure of agricultural holdings by region and main indicators.	Time period and interval Since 1977 up to 2004: yearly interval Space and scale EU 15- EU25 at NUTS2 level	Eurostat
Farm Structural Survey (FSS)	FSS is periodically conducted in order to collect data on farm structure in the EU. The NEU relevant indicators are: Land use (different temporary and permanent crops, permanent and temporary grass, rough grazing, woodland) Types of livestock Stocking density/animal numbers – Farm types (per sector, size class in hectares, Livestock Units)	*Time period and interval* Census data since 1990: 10 year interval Sample surveys since 1993: 2 year interval *Space and scale* EU 15- EU25; Census data at district level Sample surveys at a less detailed regional level	Part of the NewCronos database Dataholder: Eurostat
Farm Accountancy Data Network (FADN)	FADN contains farm level data on the structure of the farm (economic size, labor input, agricultural area and livestock population), total output, intermediate consumption, a balance sheet account and a profit and loss account. Users of FADN can work with individual farm data. Data are only representative at large regional level (NUTS 0 or 1). Main NEU relevant indicators are related to: – Land use – General characteristics of the structure of the holding (size, type of crops, animal categories) – Inputs/Outputs	*Time period and interval* Data since 1990 for seven years (1990, 1993, 1995, 1997, 200, 2003 and 2005) *Space and scale* EU15-EU25; Scale differs strongly per member state: varying from the whole member state to regional level	Dataholder: DG-Agriculture
RAINS/ GAINS	RAINS includes agricultural data to predict NH_3, NO_x, N_2O, CH_4 emissions from housing systems and manure application, including: - housing days per animal category - volume of manure excretion per animal category - animal housing system - manure storage system - manure categories (solid or slurry) - manure application techniques - type of fertilizers used	*Time period and interval* Data for the year 2000 *Space and scale* Pan European data on country level	Dataholder: IIASA
FAOSTAT	Fertilizer use by crop types expressed in plant nutrients for nitrogen (N), phosphate (P205) and potash (K20). Data on the application rates, total cultivated area, percentage of area fertilized and total consumption by individual crop. Coverage: 88 countries.	Time period and interval Update every 3 years since 1992. 5th edition from 2002 referring to data between 1995-2000.	FAO/IFA/IFDC/IPI/PPI

[1] LSU is a concept permitting a comparison between livestock of different feed requirements. For example, a diary cow is LSU 1.0, while a breeding boar, a horse, a sheep and a chicken are, respectively, 0.5, 0.8, 0.1 and 0.01.

2.1.2 Crop management data

Relevant datasets on crop N uptake are FAOSTAT, OECD, Regio and CAPRI-DYNASPAT for actual yield data, CGMS/MOCA for sowing and harvest dates and LUCAS/ CAPRI-DYNASPAT for crop shares in arable land for the distinguished HSMUs. A summary description of these datasets is given in Table 9, whereas a more elaborate description is given below.

Yield databases: There are various databases that can be used to derive historic data on actual crop yields for different crop types. The FAO has annual data on country level for EU 25+ 5 (all countries available) since 1960 for more than 100 crop types, The OECD has annual data on country level (for OECD countries) for the period 1985-1998 for 30 crop types, REGIO has annual data on NUTS2 level for the period 1974-2003 for 24 crop types. However, there are many data gaps so it is difficult to use these data without a thorough check. CAPRI has done this already and completed the database to obtain a consistent series for 30 crops. For the year 2000 and the current base year (2002) these data have been downscaled HSMUs, using additional information on crop distribution (from LUCAS survey) to regionally derive a crop suitability model. Estimates are consistent with the CAPRI REGIO database. Data are available for EU 25+ 5 except for the REGIO and CAPRI data (only EU 25).

CGMS study: performed under the MARS STAT Action, which is part of the Agriculture and fisheries Unit of the Institute for the Protection and Security of the Citizens of the Directorate General Joint Research Centre of the European Commission. Under the CGMS, a crop phenology database has been established giving sowing and harvesting or emergence/maturity dates for 11 crops. For the new Member States (EU-10), more detailed information to be used in crop growth modelling has been collected under the MARS CROP YIELD FORECASTING SYSTEM (MOCA study). This is available by CDROM. It covers 11 countries: Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia and Turkey (Table 10). The information cover: sowing, fertilizer application date, ploughing date, irrigation (yes/no), flowering data and harvest date. All the gathered data are structured in an ACCESS database which is available in JRC (contact mars-stat@jrc.it).

LUCAS, which stands for Land Use/Cover Area Frame Statistical Survey, provides statistical data on land cover/land use and environmental features representative at EU level. It is based on the visual observation of a sample of geo-referenced points, with the objective to provide representative and harmonised early estimates of land cover and land use at EU level as well as environmental information. LUCAS has been initiated by DG Agriculture and EUROSTAT. The LUCAS programme is very much biased towards agricultural crops and no time-series are yet available. LUCAS Interviews of LUCAS Farmers offer additional data on farm characteristics and agricultural practices of 5000 holdings. In 2001, LUCAS was carried out in 13 EU Member States. The survey was implemented again in 2003 in all 15 EU Member States plus Hungary. In response to the statistical quality problems appeared in the implementation of the surveys in 2001 and 2003, a new methodology an sampling have been designed and the number of sample points increased to cover the 10 new Member States. A new survey has been carried out in 2006 using the new methodology on 11 countries (BE, CZ, DE, ES, FR, HU, IT, LU, NL, PL and SK) representing around 163,000 points. The sampling design proposed for the surveys from 2005 onwards is a two-phase sampling of unclustered points with stratification after the first phase. The base sample is linked to a 1 km grid, and the LUCAS master sample is a subset of the base sample corresponding to a 2 km grid. Each point of the master sample is photo-interpreted with the most recent orto-photos or with satellite imagery, and classified into seven strata (arable land, permanent crops, permanent grassland, wooded areas, shrubland, bare land, low or rare vegetation, water and artificial land). From the stratified master sample, a field sample (sub-sample of points) is extracted to be classified by filed visit according to the full land nomenclature in the period of March –June 2006. The data collected using the 2005 methodology are: aerial imagery and mapping for each point, survey data at each point (land cover, land use, 6 landscape photos) and transect of 250m E from survey point, recording linear features/land cover types.

Table 9 NEU relevant statistical agricultural data sets at European level focusing on crop uptake

Title data set	Description	Coverage (time, space & scale)	Sources & data holders
FAOSTAT OECD Regio CAPRI	Yield data for crops at various spatial scales and for different time periods	*Time period and interval* Annual data for different crop types since 1960 (FAO), 1985-1998 (OECD), 1995-2003 (Regio) and for 2003 CAPRI *Space and scale* Data at country level (FAO, OECD), NUTS2 level (Regio) and for HSMUs derived from NUTS2 (CAPRI) for Pan Europe (FAO and OECD) and for EU 25 (regio and CAPRI)	FAO OECD Eurostat
CGMS/MOCA	CGMS contains sowing/emergence dates and harvest/maturity dates for 8 crops. MOCA information covers: fertilizer application date, ploughing date, irrigation (yes/no), flowering data, sowing date and harvest date.	*Time period and interval* 1975-2007 *Space and scale* Pan European at a 50 km x 50 km grid It covers 11 countries, also gridded at 50x50 km[2].	JRC JRC (contact mars-stat@jrc.it).
Land Use / Cover Area Frame Statistical Survey (LUCAS)	Statistical data on land cover/ land use and environmental features that are representative at EU level (currently in a pilot phase) LUCAS 2001-2003 also contains information on farm management (presence of drainage and irrigation systems, presence of a winter cover crop, passages of tools, sowing method, quantities and types of fertilisers and treatments, etc.) from farm interviews (about 4000 in EU15). Due to some problems these data are not available for research. The LUCAS classification contains 25 arable, 11 permanent crops and 2 permanent grassland classes.	*Time period and interval* 2001 data available, next survey scheduled for 2003. *Space and scale* EU 15 and 3 Accession countries (HU, ES, SI) at NUTS2 level	Eurostat: difficult to get permission to use
CAPRI-DYNASPAT	Geographic data on crop use shares for 30 crops per HSMU in consistence with REGIO NUTS2 data. Input data are also estimated at the HSMU level, such as mineral and manure N application, crop N export, N losses through denitrification + leaching (surplus).	*Time period and interval* 2002 data available; other years (1990, 1995, 1997) in preparation *Space and scale* EU 15 available; EU10+ 2 (Bulgaria and Romania) in preparation.	JRC

LUCAS may serve as a complementary data source for monitoring stock and changes of land cover and agricultural practices. Landsis has been entrusted to assist Eurostat in the training of the managers of the surveys as well as in the preparation of the technical reference documentation (http://www.landsis.lu/projects).

CAPRI: The model CAPRI has a data based on NUTS 2 level derived from Eurostat sources and FADN. CAPRI includes: N balances at NUTS 2 level per crop (>30). Manure nitrogen is calculated using animal number and animal performance (energy requirements) in combination with N content in feed and in animal products, Mineral N application is derived from FAO/IFA N fertilizer consumption data at national level. N deposition data and crop specific N fixation data externally derived are country-specific; N uptake through harvested products is based on crop specific N contents and region (NUTS2) specific yields. Recently, a procedure has been developed to downscale all parameters calculated in CAPRI to the HSMU level.

---

[2] More information under http://agrifish.jrc.it/marsstat/Crop_Yield_Forecasting/MOCA/00000000.HTM

Table 10 Selected target crops per country

crop	Bulgaria	Czechia	Estonia	Hungary	Latvia	Lithuania	Slovakia	Slovenia	Poland	Romania	Turkey
durum wheat											
soft wheat											
spring barley											
rye											
oats											
grain maize											
oil seed rape											
sunflower											
soybean											
filed peas											
field beans											
potato											
sugar beet											
total	7	10	6	11	7	8	9	5	9	11	10

2.2 Expert knowledge on farm management

The expert knowledge methods that can be followed in a project as NEU are literature review, organizing of workshops or, most importantly, the design and distribution of a questionnaire. In many cases, the use of more than one method might be required, for example, to assemble information available in literature, to complement this information with a questionnaire covering the geographic extent of interest, and to organize a workshop in order to assess the results and make it useful for modelling purposes. Several EU projects have developed questionnaires mainly focused in deriving information which is hardly available in the existing statistical databases, e.g. ELPEN and GREENGRASS.

In **ELPEN** a questionnaire was designed to link the FADN economically defined farm types to real farms and to add further information from different sectors (i.e. technical, environmental and socio-economic). Data were collected from selected farms to furnish a REFERENCE FARM DATABASE that enabled the linking of the computer modelled potential impacts of agricultural policy changes to “the real world”. The number of collected reference farms per region is shown in Figure 1. This map indicates that the dataset only covers specific farm types and is not consistent across EU-25.
The questionnaire had 93 questions about 10 aspects of the farm and 7 policy questions:

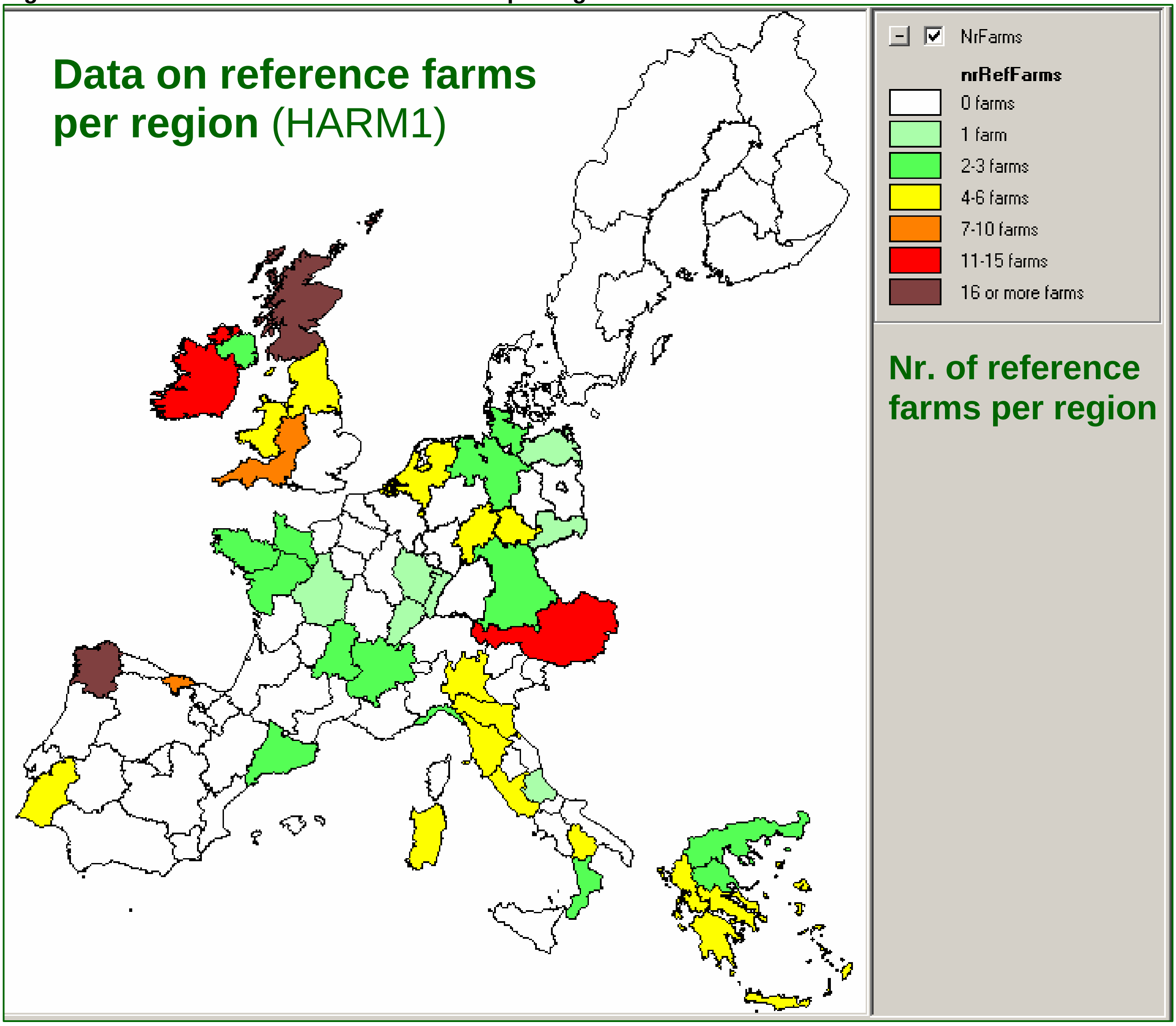
A) General questions
B) Land related questions
C) Grass and Fodder
D) Mechanisation
E) Dairy cows
F) Beef cattle
G) Sheep
H) Goats
J) Other grazing livestock
K) Livestock housing / welfare
L) Questions about Agricultural Policy Changes

Examples of the questions are:
B) LAND RELATED QUESTIONS
18) altitude < 300m, 300-600m, > 600m
19) gradient of land (estimate
20) soil quality (1 excellent - 5 poor)
21) layout of farmland (scattered - consolidated)
22) farm size
23) policy status
24) use of land
25) ownership of farm: land owned/rented/leased
26) types of crops (excluding fodder and grass)
27) fodder production (roughage other than grass)
28) conservation of grass
29) arable land with vegetation cover

30) fertiliser use
31) pesticide use

K) LIVESTOCK HOUSING / WELFARE
86) type of livestock
87) duration of housing
88) housing type
89) housing
90) grazing management
91) transport
92) visits by
93) main health problem

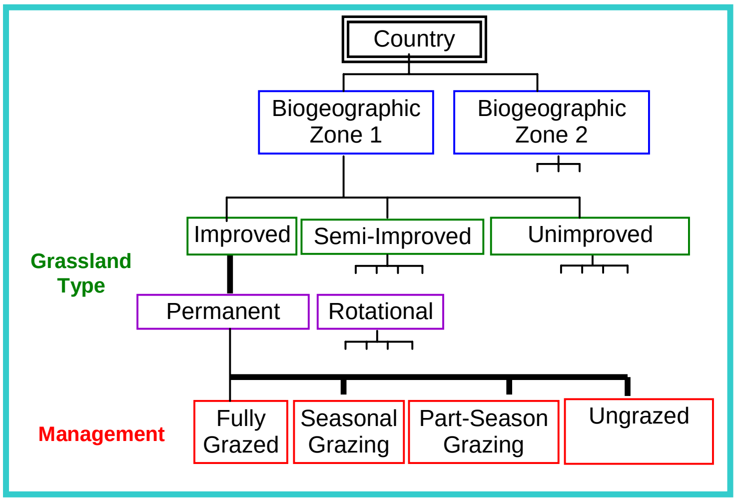
Figure 1. Number of ELPEN Reference farms per region



Each farm selected as a reference farm covered the same set of questions throughout Europe. Different livestock types were considered in separate sections of the questionnaire, so that each farmer only had to fill in the sections appropriate for their farm.

In **GREENGRASS** the main objective was to obtain management data for all combinations of country/biogeographic zone/grassland type (CBG) within the European grid, including EU25 + 5, in order to create a timeline of management practice for each biogeographic zone. The approach should ensure that the grassland management activities (cutting, fertiliser application, manure application and grazing) will take place in a logical order. The database structure of the questionnaire is hierarchical (Figure 2).

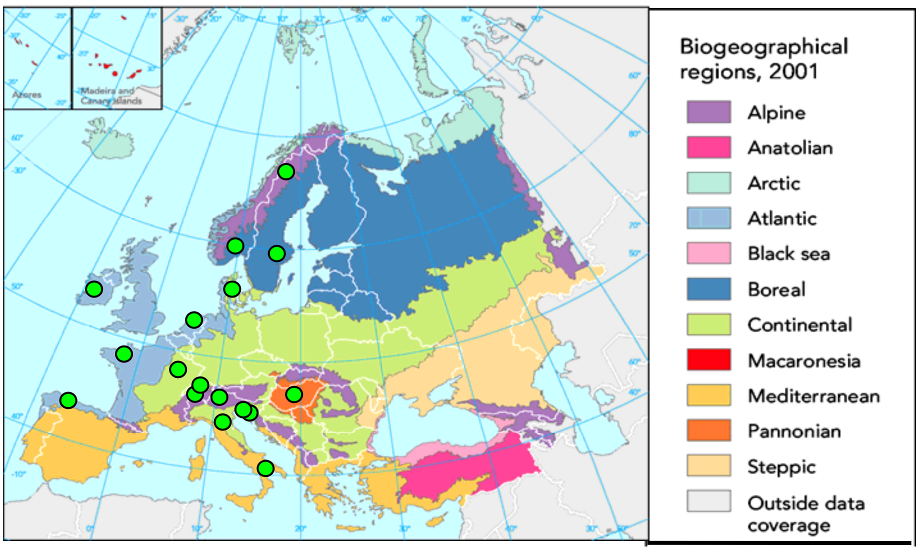
Figure 2. Database structure of NEU questionnaire



The grassland types considered are: permanent intensive grassland (> 5 yrs old), rotational intensive grassland (< 5 yrs old), semi-improved grassland and unimproved and rough grassland. Experts were identified through local contacts and approached by e-mail. The questionnaire has one worksheet for each grassland type, which includes (i) general details (sward area, species composition, irrigation and reseeding practice) and (ii) management details by grazing practice e.g. type of stock, grazing system (%continuous grazing, % rotational grazing).

Existing questionnaire data, collected in GREENGRASS, is currently available in an Access database. 19 questionnaires were received in GREENGRASS; geographic and biogeographic coverage is shown in Figure 3. Data gaps have been identified; consequently a data collection strategy has been devised for the next phase of the project with the intention of filling gaps where possible. The green dots denote each combination of country and biogeographical zone where a completed survey was received.The white lines denote boundaries between countries; biogeographical regions are denoted by the colour scale. Source: EuropeanEnvironmentAgency. http://dataservice.eea.eu.int/atlas/viewdata/

Figure 3. Biogeographic regions, political boundaries & survey responses considered in the NEU questionnaire.



Methodology of synthesis & conversion to model-ready input files

The questionnaire is structured as a hierarchy of increasing detail, beginning with broad biogeographic zones and becoming increasingly specific through country, grassland type and management regime. When the database is fully populated with data for every combination of these four levels then the modeller will be able to select the resolution, down to the sub-grid cell level.

However, at this point, the database has gaps, which prevents the modeller from running the model to encapsulate grid- or sub-grid-cell heterogeneity. In GREENGRASS, DNDC was run with just one management regime for each biogeographic zone. This necessitated averaging received grassland management data. Where the data is numeric, e.g. quantity of mineral fertiliser applied, a simple mean may be calculated. However, a simple mean is not appropriate in synthesising the dates of cutting, fertilising and onset of grazing. In addition, fertiliser quantity is not necessarily simple, as it may need to take into account a range of practice in the combination of mineral and organic fertilisers. Therefore, a subjective average was created in each case, in order to provide a logical chronological sequence. This was done by examining the timeline of management events in each questionnaire relevant to the area of interest, then devising a best-fit timeline. During GREENGRASS, the subjective averaging was carried out by Mark Sutton.

Data from completed questionnaires is entered into an existing Access database, which was developed during the GREENGRASS project. The database has been designed to make it fast and simple to produce files which are ready for use as model input files. The database can be queried using pre-existing report structures; currently these output tables which may be exported and used immediately as DNDC input files. Other formats may also be added at a later date, if format requirements are provided

Timeline of future plans

Existing questionnaire data, collected in GREENGRASS, is currently available in an Access database. However, the database needs to be simplified; the structure streamlined, the procedure for adding new data and querying improved, and a process developed to automate averaging of data such as timing of management events, which may not be simply averaged in the standard way.

CEH will hold a workshop concurrent with the next European Grassland Federation (EGF) meeting, in Ghent, during September 2007. CEH will contact the executive committee of the EGF to ask if they could forward the questionnaire to relevant experts in the region of Europe that they represent on the committee. It is also planned for the questionnaire to be made available for download from the British Grassland Society website. A preliminary database of N flows in agricultural systems will be compiled from the existing grassland management database. This information may be used to derive empirical rules to produce feasible estimates of grassland management timelines where data is not available.

In NEU the GREENGRASS questionnaires on grassland management will be used. Collection of newly-completed questionnaires is currently at an early stage. The main objective of the completed questionnaire is to supply input to the model requirements for the plot scale process model PaSim and also compatible with inputs required by DNDC and Daycent for grassland. There is a need for an additional questionnaire to extend the assessment to cropland management as input to cropland models, i.e. CERES-EGC, MAGEC, Sundial and DNDC-Crop.

In addition, the survey data will be linked to land cover using the Land Cover map of the Stockolm Environment institute (SEI) that has explicit land cover classes for the three main grassland types. Experts filling in the questionnaire will be asked to define the size of the area for which they can provide average data, e.g. province or national level.

The information gathered by the DNDC model regarding timing of fertilizing input and tillage for each crop (see table 11), which is derived from planting and harvest dates obtained from the LUCAS project could be used as expert knowledge (Kahl & Butterbach-Bahl, pers. commun.).

Table 11 Information used by the DNDC model regarding timing of fertilizing input and tillage for various crops

Crop type	1st	2nd	1st	2nd
	Fertilizer date		Tillage date	
spring wheat rapeseed rye sunflower beans/pulses	planting day	-	planting day	5 days after harvest
corn	planting day	-	planting day	15 days after harvest
barley oats	1 month after planting	-	planting day	5 days after harvest
soya	planting day	-	5 days after harvest	15 days before planting
winter wheat	planting day	4 months after planting	5 days after harvest	15 days before planting
potato sugar beets rice cotton tobacco	planting day	1 month after planting	planting day	5 days after harvest
vegetables	planting day	3 months after planting	5 days after harvest	-
nursery flowers berries	1 month after planting	4 months after planting	15 days before planting	1 month after harvest
grapes	1 month after planting	4 months after planting	2 months after planting	1 month after harvest
fruittrees citrus	2 months and 20 days after planting	5 months and 20 days after planting	25 days after harvest	-
legume hay	1 month after planting	-	planting day	15 days after harvest
grassland	-	-	5 days after harvest	-
fallow	-	-	-	-

2.3 Model parameters

The list of model parameters required by all the models used in NEU (described in section 1. 2) are much too extensive to realistically make a thorough comparison of all these data for all models; also, some of these parameters are built-in into the modelling concept of the various models, so a comparison without the respective framework would not make sense. However, we aim to minimize differences in some key parameters for the nitrogen budget: nitrogen input via manure and nitrogen export by the crop. Therefore, (i) a preliminary literature review will be done by ALTERRA, JRC and IMK in order to collect as much as possible model parameters needed; (ii) a comparison will be made between those tackling the same data requirements; and (iii) results will be upload in the general NEU database. NEU modellers using model parameters not included in the review, will be invited to add their own data.

For the INTEGRATOR prototype, most data can be derived from the RAINS/GAINS database. Exceptions are the N contents in crop removal and crop residues, which are also needed by other models. An overview of N contents in harvested products and crop residues used INTEGRATOR, based on a literature review of Velthof & Kuikman (2000) from Dutch studies of the nineties, is presented in Table 12.

Improvement of nitrogen excretion and CH_4 emission factors per animal category. Over time milk production per dairy cow is increasing as well as meat production per animal and crop production per hectare. Increased production per animal might influence the nitrogen excretion rate per animal and also the composition of the ration will have an impact on the excretion rates. Principally the feed intake is calculated on the basis of energy demand of the animal for maintenance, milk production, growth, pregnancy and activity like grazing. Increased milk production or growth rate will influence the energy demand and hence the feed intake. In the second phase of NEU this will be elaborated in more detail in order to derive excretion rates which reflects differences in milk and meat production levels as well as differences in local feed composition. So the effect of different scenario's for future years might influence the mineral composition of grassland products and the availability of feedstuffs and the subsequent consequences for the nitrogen excretion rates might be assessed. Methane emission factors are driven by the energy intake of the animal, so in the second phase of NEU they also will reflect the country, year and scenario specific situation.

Table 12 Average nitrogen contents and nitrogen indexes in arable crops used in the uptake calculation for INTEGRATOR using the various CAPRI crops

CAPRI LUCAS	CAPRI ID	Description	N content (g.kg-1)	N index (Ncr/Nre)	fup (-)
Arable land					
SWHE	1	Common wheat	20	3.02	0.9
DWHE	2	Durum Wheat	20	3.02	0.9
BARL	3	Barley	17	2.37	0.9
RYEM	4	Rye	14	1.79	0.9
OATS	5	Oats	17	2.08	0.9
MAIZ	6	Maize	13.9	1.53	0.9
PARI	7	Rice	20	3.02	0.9
OCER	8	Other cereals	15	2.02	0.9
POTA	9	Potatoes	3.5	3.13	0.8
SUGB	10	Sugar beet	1.8	0.67	0.8
ROOF	11	Other root crops	1.9	1.78	0.8
SUNF	12	Sunflower	32	1.79	0.8
RAPE	13	Rape and turnip rape	35	1.79	0.8
SOYA	14	Soya	58	2.05	0.8
TEXT	15	Fibre and oleaginous crops; Cotton	4	1.11	0.8
TOBA	16	Tobacco	30	2.05	0.8
OIND	17	Other non permanent industrial crops	4	1.11	0.8
PULS	18	Dry pulses	42	4.97	0.8
OFAR	22	Fodder other on arable land; Temporary grasslands	5.8	2.43	0.8
FALL	23	Fallow land		0.0001	0.8
OCRO	29	Other crops; Permanent industrial crops	5	2.05	0.8
Arable land and Horticulture					
TOMA	19	Tomatoes	1	3	0.8
OVEG	20	Other fresh vegetables	2.5	1.16	0.8
FLOW	21	Floriculture	5	2.05	0.8
FRUI	24	Other Fruit	0.5	3	
CITR	25	Citrus fruits: Oranges	0.5	2.05	0.8
OLIV	26	Olive groves	0.5	2.05	0.8
TWIN	27	Vineyards /table wine	0.5	2.05	0.8
NURS	28	Nurseries	5	2.05	0.8
APPL		Apples	0.5	2.1	0.8
OWIN		Other wine	0.5	2.1	0.8
OFRU		Other fruit	0.5	3	0.8
Gras		Gras	30	2	1.0
MAIF		Fodder maiz	15	4.9	0.8
TAGR		Table grapes	1.9	2.1	0.8
TABO		Table olives	5	2.1	0.8
OOIL		Other oil	34	1.3	0.8

3. Methodologies to derive data

3.1 Data needed versus data available: data gaps

When comparing the data needs and the data available reviewed in sections 1 and 2, data needs can be classified according to their availability as follows:

- data that are available from statistics, e.g. animal number per animal type
- data that are not directly available from statistics but that can be derived from statistics using statistical methods, e.g. fertiliser consumption at national level, nitrogen export with harvest, excretion factors
- data that are not available and cannot be derived from statistics. These data refer mainly to specific farm management, go beyond any statistics (now), and in most cases are specifically relevant for detailed ecosystem models with a high time resolution (e.g. daily), such as fertilizer timing and manure application techniques

In addition, the following needs are identified:

- data available from statistics might not fulfil the model requirements regarding spatial and time scales. European modelling of nutrient and greenhouse gas fluxes is particularly challenging with respect to spatial needs: on one side the data have to cover the whole area of Europe, which in the case of the NEU project, has been defined as EU-25 + 5: the three accession countries at the time of the project preparation (Bulgaria, Romania, Croatia), as well as Switzerland and Norway. On the other side, the data have to be detailed enough to reflect all the spectrum of conditions that the European landscape offers (different soil conditions, different climates, different traditions), which are found at a very detailed scale (from 1 km^2 to several 100 km^2, depending on the homogeneity of the landscape). Time periods vary between yearly, seasonally, monthly and daily
- need for harmonisation of the different databases used and assessment of the quality of the data
- reconstruction of agricultural data from 1960 at NUTS2 level

Therefore, the generation of input data for models is dependent on (i) the availability of 'basic' information and (ii) on the specific methodology to make this information useful for the models.

3.2 General methodological approaches for data generation

We can distinguish different methodologies to derive the input data depending on the availability of the data (Table 13)

Table 13 Types of data availability and the corresponding methodologies to generate the input data required

Data availability	Data generation methodology
A = Available from statistics or remote sensing	**Resampling** U =upscaling; D = downscaling; S = side-scaling
B =Not (directly) available from statistics but derivable	**Expert knowledge** L = literature review Q = questionnaire Ex = direct expert consultation St = Statistical methods M = Models specific for data requirement

I. For data available from statistics, data can be directly used or used after resampling:
   - Upscaling: aggregating smaller to larger units (e.g. NUTS3 to NUTS2, grid-data to NUTS 2)
   - Downscaling: disaggregating smaller to larger units (e.g. NUTS2 to NUTS3, administrative units to 1 km^2 grid)
   - Side-scaling: changing the spatial support from the one system of units to another comparable average size (e.g. landscape units to NUTS regions)

II. For data not directly available from statistics but derivable, data can be used after processing through statistical methods and with the support of GIS, e.g. estimation of fertilizer input. In this group no sufficient information exists to derive directly a meaningful coverage for Europe, but some additional information can be obtained (or must be gathered) to come up with realistic numbers. To this group belong many of the management practices. An example is the estimation of fertilizer application rates by using information on standard fertilizer application rates per crop and country, nitrogen content in plant compartments that are removed from the system, a link of expected yield with soil conditions, etc.

III. For data not available and not derivable from statistics, data need to be estimated by targeted modelling, e.g. simulation of farmer's behaviour for fertilizer application timing. This group is evidently the most difficult one as no reliable information is available to support the generation of European-wide data. To this group belong all "soft" management data for which no appropriate survey has yet been made for the whole of Europe. In this case, only crude guesses can be made or, if the techniques are available, models can be used.

All these methodologies can use information obtained by expert-knowledge based on:
- Literature review
- Direct expert consultation (e. g. through a workshop)
- Questionnaires (e. g. GREENGRASS questionnaire)

3.3. Upscaling and downscaling methodologies

There are known issues about data manipulation, which should always be kept in mind when performing up-down-side scaling procedures. The first is that there is no standard, i.e. the decision on which method to be applied is left to the expert performing the operation. The second is the ecological fallacy, i.e. assumptions made about individuals based on aggregate data might not be true because variability in the underlying individual data is lost by using averages. The third is that statistical inference changes with scale because by definition aggregate data has less variance than dis-aggregate data (Paracchini, 2006).

A common characteristic to aggregation/disaggregation issues is the Modifiable Areal Unit Problem (MAUP). It was defined by Heywood (1998) as “a problem resulting from the imposition of artificial units of spatial reporting on continuous geographical phenomenon resulting in the generation of artificial spatial patterns”. An extensive literature exists on methods to control –or try to control- the drawbacks of the MAUP (Bivand, 1998; Morphet, 1997).

3.3.1 Upscaling

This includes aggregation rules for upscaling of site (farm) specific information from farm level (e.g. farm management practices) to NUTS/NUTS 3 region or NitroEurope calculation units (NCUs). This upscaling may lead to ((MAUP) a) a “scale effect”, i.e. different results and inferences are obtained when the same set of data is grouped into increasingly larger areal units and b) a “aggregation effect”, i.e. variability in results and inference results from alternative formations of the areal units. SENSOR decided to use as reference units a mix of NUTS2 and NUTS3 regions (called NUTS-X regions) having comparable surfaces. being the approach used in FSS by Eurosat This choice partially solves the MAUP problem, because the distribution of region surfaces shows a smaller standard deviation than either at NUTS2 or NUTS3 level, and this makes most of indicators more easily comparable. Nevertheless the difference of the “political meaning” of these two levels among countries must not be forgotten, especially when the considered NUTS level has a high impact on the regional planning process.

Within the FP6 IP SEAMLESS, the upscaling of modelling results is based on the The Agri-Environmental Zonation (AEnZ) (Hazeau et al., 2006). AEnZ is a hierarchical and flexible subdivision of the European landscape into 238 relatively homogeneous units from an agronomic perspective. The AEnZ is based on 13 environmental zones, 6 organic carbon content classes and an agrimask that consist of three classes. The 238 classes of the AEnZ are described in terms of land cover, climate and biophysical parameters (altitude, growing season, temperature range, summer drought, slope, AWHC texture and rooting depth). AEnZ) can be used at different levels of detail. A relatively coarse division of Europe can be obtained by looking only at the environmental zones. A more detailed biophysical subdivision of Europe can be generated by incorporating soil (organic carbon content) information. The Agrimask is adding additional information to the AEnZ through indicating possible restrictions for arable agriculture. The upscaling of modelling results with use of the AEnZ typology is possible when a (spatial) relation between model outputs and the AEnZ is established. The allocation of a dominant soil type to an AEnZ class makes it possible to upscale model outputs which have a relation with the soil attributes describing the soil type. Other examples of upscaling are the use of allocation of farm types and the allocation of land use.

In NEU, it is foreseen that the aggregation of the questionnaire results from the local to a higher spatial level (regional or national) will be done with the same upscaling methodology used in GREENGRASS, i.e. based on the dimensions of the averaging area determined by the expert. The regional variation in management practice occurs within biogeographic zones within countries. Data aggregation only takes place when several surveys per combination CBG are submitted.

3.3.2 Downscaling

Downscaling includes rules for disaggregation of information from NUTS2/NUTS 3 to NitroEurope calculation units (NCUs), which are homogeneous spatial mapping units. It is recognised that not all data can/need be downscaled, but will be used at a regional or even national level. Therefore the downscaling procedure will not be feasible or required for all parameters. Following, we shortly describe different downscaling procedures for animal numbers used in different projects.

In the FP5 ELPEN, a methodology was developed to allocate land-based dairy systems based on FSS and Corine Land Cover (CLC) on 1km^2. National experts were asked to formulate rules to allocate the correct statistical information to the correct grid information. The rules were then translated into so-called “allocation-algorithms” that indicated the chance that a specific livestock activity or a specific type of livestock farm would appear in a grid cell (or other spatial entity). With these algorithms maps could be produced of the most probable location of dairy systems within the EU. The allocation procedure was as follows:

- (i) The FSS data were used to determine the number of farms (per farm class) and livestock numbers within the EU per region. FSS data were used because they provide consistent European information.
- (ii) The statistical data asked from the national experts was then used to refine the available FSS data, as these were normally available at a higher regional level than national statistics.
- (iii) A further refined allocation (detached from administrative boundaries) was obtained by using rules, given by experts, indicating (per country) the suitability of the land cover on the CORINE map for each livestock and/or farm type; e.g. by comparing national statistical data with the CLC data one can estimate that e.g. 80% of the dairy cows or dairy farms occur in the CLC class “pastures”, while only 20% occur in the class “annual crops associated with permanent crops”.
- (iv) When possible other expert rules were used for more refinements on the basis of other data like climate, altitude, slope, distance to big cities, etc.
- The probability rules in steps (iii) and (iv) were obtained per country by statistical analysis of available detailed information and/or by expert judgement.

Within the FP6 IP SEAMLESS, the downscaling of modelling results will be also based on the The Agri-Environmental Zonation (AEnZ) (see upscaling section above). The use of the AEnZ for downscaling/upscaling has not yet been implemented.

In the Dutch project Eururalis, a methodology was developed for downscaling of animal numbers per livestock type (Elbersen et al., 2006). The input data was the FSS database, which contains harmonized EU wide regional data on livestock types, their numbers, and their distribution over farm types. These data were taken for the years 2000 to 2003, depending on the available variables and their spatial resolution per region. The rule was to take all information at the most spatially and thematically detailed level possible. Since the livestock types needed to be disaggregated using additional information on the diversity in their spatial distribution according to livestock types, it was important that the livestock units were specified per livestock type but also per farm type (EU farmtype). For the farm types land use information was also collected in order to establish a link between the land use types and the different livestock types. The livestock data that were disaggregated were therefore available at the regional level of NUTS1-2 (or FSS 2), which is mostly national level for the smaller Member States (MS) and province or autonomy level for the larger countries. The first spatial data layers used for identifying the locations where livestock is most likely to occur is the CLUE land cover information, based on an adaptation of the Corine Land Cover (CLC, 2000) information. The CLUE land cover information is available at a 1*1 km^2 resolution and was used to identify the locations within the FSS regions where the livestock occurs. A further downscaling was made using the statistically reported livestock numbers and average regional stocking densities per livestock and land use type. This statistical information was further combined using expert based judgment with spatial data sources on land use, slope, altitude and soil characteristics influencing the livestock carrying capacity of land within the FSS regions. A major distinction is made between different animal types: dairy cows, beef cattle, sheep and goats are assumed to be highly dependent on local land resources for grazing or feed production. Pigs and poultry are assumed to be held in more land independent systems which implied that carrying capacity estimates were assumed to be less influential in allocating them spatially. For more detailed information on the creation of the initial livestock distribution, we refer to Elbersen et al. (2006).

In NEU, it is suggested to use a disaggregation of livestock numbers based on the Eururalis approach, in which the FSS database will be disaggregated to a 1km x 1km grid for further aggregation to NCUs. The changes in the spatial distribution of livestock in relation to land use will be assessed using the Eururalis livestock density map previously described.

3.4 Harmonisation and classification of data

Harmonisation of available data is an essential part of the regional research. However, this becomes an extremely demanding task when it is applied across 30 countries, trying to reach full coverage at the lowest possible geographical level. Therefore, it is expected that the NEU project partners will put collective efforts in this activity and jointly form the NEU database. The agricultural database will be part of the NEU database. It will firstly include basic regional statistical data and secondly specific, model-orientated data and computed indicators. In order to limit work and to avoid the use of different data for the same factual information, these will be centrally prepared and made available to the models on request.

Basic data for the 25 EU member States as well as for Bulgaria, Romania and Croatia will be provided by the Statisitcal Office of the European Communities (Eurostat) and FAOSTAT. These will be complemented by data from the national statistical offices of Switzerland and Norway.

Preliminary proposal of a methodology for the integration of heterogeneous databases

A multidimensional approach of databases

The information which is useful for the NEU models can be described as a set of coherent databases in which the three following important dimensions are more or less developed:

- Thematic dimension (V), i.e. Livestock production, manure and fertilizer application and crop management
- Spatial dimension (S), i.e. NUTS 0 - NCUs
- Temporal dimension (T), i.e. historical data (1970 – 1999), baseline (2000)

In most cases, it seems impossible to develop one of those dimensions without reducing the others. This can be illustrated by the following examples:

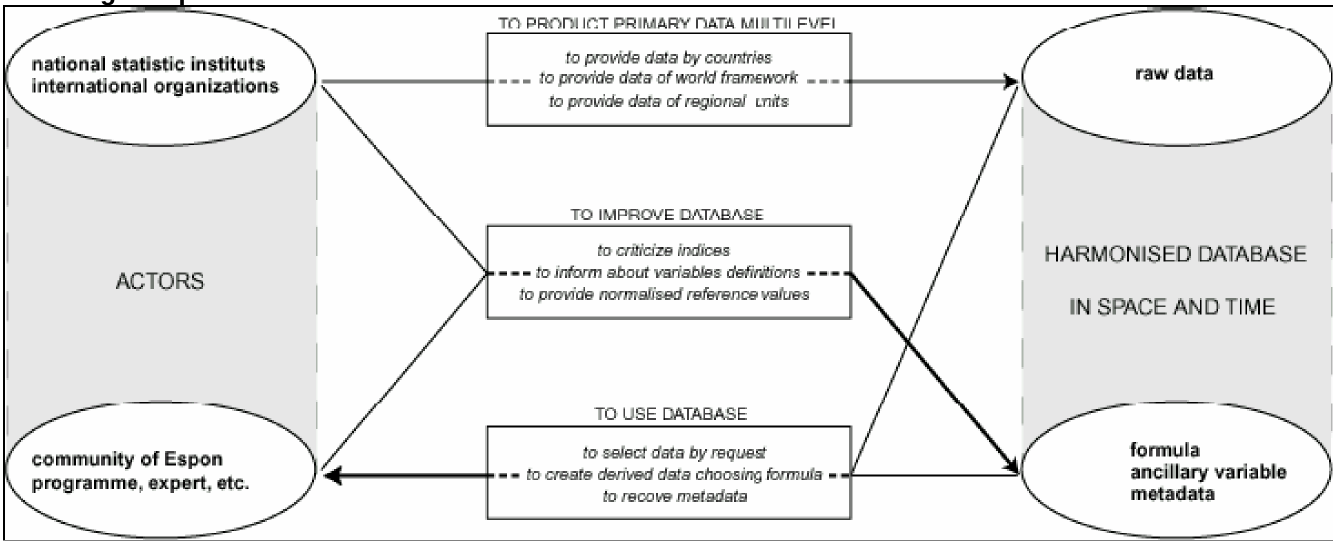
The **FSS agricultural database** is characterised by a relative diversity of variables but those variables are available at a medium spatial level (NUTS2 or NUTS 3) and can generally not easily be transposed at a lower level of aggregation (NUTS 5). But the most important weakness of this database is the temporal dimension, which is limited to the very short 1990-2003 period in most cases.

The **CLUEs/ Corine Land Cover** database is characterised by an opposite situation, with a very high level of spatial resolution (100-200 m in vectorial format, 1 to 2 km in grid format), but with a very low amount of information (land cover), not always very well harmonised, although strongly improved in the new version of CLC 2000.

The **FAOSTAT** database of the FAO (V=Medium; S=Low; T=High) is characterised by a high level of time resolution (yearly time series for most indexes) with a relatively important diversity of information. But the main weakness is the low level of spatial resolution which provides very few solutions for the description of regional differences in Europe. As a consequence, this multidimensional approach is a crucial challenge for the NEU project because the implementation of the models and development of good political scenarios assume precisely a high level of quality in each of the three dimensions. The aim of the present work package will be to effectively examine the possible solutions that may be applied to existing databases in order to improve their quality and their political significance.

Together with EUROSCOPE, the ESPON project has developed an experimental database for European Spatial Planning and political scenarios to improve data harmonization that could be used in the NEU project as well (Figure 4).

Figure 4. General objectives of the ESPON-EUROSCOPE experimental database for European Spatial Planning and political scenarios



The main principles of the ESPON-EUROSCOPE methodology, adapted to the NEU project could be:

Principle 1: The starting point is distributions over EU 30 at national level in 2000: whatever the variable considered (livestock numbers, fertilization application rate), it will first be collected at national scale in 2000. The choice of states (nations) as initial spatial divisions is also compulsory as far as all statistics are normally produced by states and cannot exist at regional level if they are not available at national level.

Principle 2: Times series are first established at state level for the target period (1970-1999): whatever criteria we want to analyse, the first step is to examine if it is possible to establish coherent time series at state level. Indeed, the establishment of time series at state level will quickly reveal the possibility or impossibility to obtain some criteria (e.g. fertilizer application for East European countries before 1989). Furthermore, it will provide normalised values of reference which can be used for the correction of regional statistics which are not necessarily easy to compare through time. This collection of series at state level is of course a good opportunity to distinguish between real information (census) and indirect estimations or interpolations.

Principle 3: The regional distributions are established in the official units of each state at each time period. It is important to collect the data produced by states 'as they are', without trying in a first step to perform some spatial or temporal harmonisation. What is important is to identify precisely the 'STU' (Spatial-Temporal Units), which are the territorial divisions available in a given state during a given period. And this should be done very carefully as there are many examples of STUs which are subject to small changes, not visible on maps but susceptible to produce important mistakes in statistical analysis[3]. For each state and, in certain cases, for each statistical board of each state, the prior work is to establish the precise dictionary of the different STUs which are used for statistical collection during the period of interest. For practical reasons, the collection of the STUs may be limited to a minimum level (like NUTS3) but the database should stay open to the possible use of a lower level of territorial division (like NUTS5).

[3] For example, when a NUTS2 or NUTS3 unit is increased/reduced by some local NUTS 5 units, this might not imply any change in the name of the units or in their mapping, but it can give the false impression of very high increase/decrease in population.

Principle 4: The development of harmonised time series can be achieved in different ways which are not pre-defined. Indeed, the great mistake of most projects of long-term database is to propose a priori one solution for the production of harmonised time-series, whatever this solution may be (aggregation/disaggregation procedures, smoothing methods, AI, expert system, etc.). In our point of view, as there are many possible solutions for the development of harmonised time-series of indicators, it is important to allow them. It is crucial to avoid the confusion between the original information (data as they are collected by states) and the derived information (methods which are applied to these data for the achievement of harmonised series). The database should store the different methodologies which have been elaborated to solve, among others, the Modifiable Area Unit Problem[1], instead of storing the results of a given methodology.

Principle 5: The problem of quality is not to reject approximations but to control them by confidence levels. It would not be satisfying to propose dogmatic rules of statistical harmonisation without a certain level of approximation. When territorial units are absolutely different from one period of time to another (e.g. Portugal, Poland, etc) it is impossible, or at least very expensive, to establish perfectly coherent time series. However, it is possible to reach relatively good approximations by using ancillary variables like surface, population, land use, common to the units of both periods. In the framework of the NEU project, it will be necessary to be pragmatic and to use these approximations because we do not have sufficient resources for real harmonisation. But in such a situation it is crucial to very strictly control the approximation which is introduced in the calculated indicators and to have confidence levels of the results.

3.5 Data quality

It is essential to assess the quality of the data used as input in the NE models.
We will examine the possible solutions that may be applied to existing databases in order to improve their quality. This process is suggested to be guided by five principles:
1. The starting point is distributions over Europe in 2000.
2. Times series are firstly established at State level for the target period.
3. The regional distributions are established in the official units of each State at each time period.
4. The development of harmonised time series can be realised in different ways which are not pre-defined.
5. The problem of quality is not to refuse approximations but to control it by confidence levels.

The agricultural farm databases have different problems regarding accuracies as described below. Therefore, a thorough quality control will take place before including them in the NEU databases.

Note on data quality FSS 2000
Most data published in the statistical tables of the Eurofarm system are produced from sample surveys, while only some are derived from full census. The size of the samples depend upon the country and the survey year. It varies between 3% and about 40% of the total population of agricultural holdings. Consequently, most published data contain sampling errors. Users of Eurofarm data should therefore take into account that especially for cells based on very fine classifications, the estimated sampling error can be in the order of 20% or higher. In addition, a further error is introduced during the rounding treatment to guarantee confidentiality during the production of statistical tables from survey data. The extent of this error for a given table cell depends mainly on the value of the cell or the number of holdings contributing to this cell, the smaller the value or the number of holdings, the higher will be the rounding error.

Note on quality data FADN
The reliability of the data depends not only upon the sample size but also upon the frame from which the sample is selected. Many of the variables about which data are collected are highly skewed. The Court of Auditors discovered that different Member States use different methods for drawing samples, and has questioned whether the samples are drawn randomly. Participation in the survey is voluntary. In the United Kingdom, it has proved difficult to get farmers to take part; there is a 70 percent refusal rate and only 10 percent of the sample is replenished each year. Renewal rates differ among Member

States. The commitment to provide data is demanding, and participants may need to gain experience of what is required before they are able to supply data of the requisite quality. Farmers tend to remain in the survey for long periods of time because of the difficulties in finding replacements. There may be biases in terms of those who are willing to participate - for example, the length of time a farm is in the sample could create an age bias. There may also be an element of self-selection in the sample, in which the individuals in the sample select themselves rather than being selected randomly.

Farm Accountancy Data Network sample, 2002/2003

Country	Sample (percent)
Austria	2.3
Belgium	3.1
Cyprus	…
Czech Republic	10.2
Denmark	4.4
Estonia	6.3
Finland	1.8
France	2.0
Germany	3.0
Greece	0.8
Hungary	2.1
Ireland	1.0
Italy	2.6
Latvia	2.0
Lithuania	3.9
Luxembourg	25.9
Malta - Netherlands	1.7
Poland	…
Portugal	1.1
Slovakia	14.9
Slovenia	…
Spain	1.2
Sweden	2.5
United Kingdom	2.3
EU	1.9

Note: … = not available.

Different types of farm accounts exist and could be used as the basis for FADN. The British survey, for example, uses business rather than tax accounts. This has the advantage of avoiding distortions introduced by tax avoidance strategies. The use of tax accounts can overcome problems with farmers refusing to cooperate because such accounts are available to governments. Different Member States use different dates of accounting years for FADN, so the results are not strictly comparable.

A growing number of farms are run by companies. The Court of Auditors has questioned whether FADN is representative of these or has a suitable methodology for reporting on them. It has questioned whether the profits of a limited company can be compared with the income of an individual farmer, which is likely to include returns on the labour, capital and land (s)he has supplied, whereas a company will buy the use of these. This is particularly pertinent because in 2001 the International Accounting Standards Committee adopted an accounting standard, IAS 41, on farm accounting. Accounting standards tend to have a greater impact on companies than on non-corporate bodies, such as family farms, which tend not to have to register their accounts. The introduction of IAS 41 has been controversial with regard to such features as the use of fair value rather than historical cost in valuing biological assets and the inclusion of holding gains on stocks as income. There have been questions raised as to the extent to which FADN is compatible with this standard (Elad, 2004).

Note on quality of LUCAS
Precision is expected to be around or better than 2% for main categories like wheat, cereals, arable land, permanent grassland, permanent crops, forests, urban areas and inland waters. In response to the statistical quality problems appeared in the implementation of the surveys in 2001 and 2003, a new methodology an sampling have been designed and the number of sample points increased to cover the 10 new Member States. A new survey has been carried out in 2006 using the new methodology on 11 countries (BE, CZ, DE, ES, FR, HU, IT, LU, NL, PL and SK) representing around 163,000 points.

3.6 Historical reconstruction of agricultural data

One of the aims of NEU is to reconstruct the past up to 1970 (and if possible even up to 1900), in terms of past management and land use to gain insight in the impact of historic changes on the emissions of N and green house gases. For the period up to 1960, use can be made of FAO data on animal numbers, fertilizer application rates and crop yields on country level. More detailed data are available for: (i) livestock (animal numbers in more detailed animal categories) in terms of geographic resolution (e.g NUTS2 level) since 1977 (LSS) or 1990 (FSS) and (ii) for manure/fertilizer application and crop yield. Specifically the CAPRI model has derived spatially detailed data (HSMIs comparable to the NCU) on manure and fertilizer application and uptake assessments with dataset at country and NUTS2 level for the year 2003.

Possible approaches to reconstruct data since 1960 are various. For INTEGRATOR, that uses most of these data, it is foreseen that available past agricultural data for life stock at country level in FAO will be allocated to NCUs by (i) using data on annual country statistics from 1960 onwards (FAO) and; (ii) allocation these data at 1 km x 1km level (livestock data; see above) and aggregating them to NCU level.

3.7 Extending the aerial coverage of the data

The data available from statistics do not always fulfil the model requirements regarding spatial and time scales. The European modelling of nutrient and greenhouse gas fluxes by INTEGRATOR (and other ecosystem models) covers the EU-25, the three accession countries at the time of the project preparation (Bulgaria, Romania, Croatia), as well as Switzerland and Norway. Most important is information on life stock and crop uptake, however. By using the FAO databases at country level, these countries can be included on the basis of these data in view of manure and fertilizer input and crop uptake. Information on timelines has to be derived by rules and models that can be validated on databases for a limited number of countries.

4. Systematic overview for data collection

The collection of ‘basic’ data in NEU will consist in gathering existing data sources. This will require in many cases expert knowledge and resources to evaluate which data are accessible, which data contain the information that is required, and, in the case that the data are not exactly what is required (as will happen in most cases), which methodologies van be used to made the data useful. The assessment of data gaps will provide input into the design of future surveys (e.g. FSS survey on production methods planned for 2010).

In order to facilitate the data collection, Annex 1 provides a systematic overview of the former sections concerning data needs in NEU models, data availability, methodologies available and their interaction, the priorities and the responsible partner. The table should be made for (i) present and (ii) historical data (1960 -2000).

The specific information to be collected is as follows:

1. Data needs: parameter needed to be surveyed,

2. Data availability
    - Data source
    - Spatial scale available (which countries in the EU 25+5), which countries are missing
    - Time scale available (time period and interval)
    - Availability type (A, NDA, NA)
3. Priority for data gaps
    1 = need to be collected and processed under any circumstances
    2 = preferably be collected and processed
    3 = not really necessary
4. Responsibility = responsible partner assigned for priority 1
    - Data collection
    - Data processing

References

Bivand R (1998). A review of spatial statistical techniques fro location studies, available at http://www.nhh.no/geo/gib/gib1998/gib98-3/lund.pdf

Elbersen B, Kempen M, van Diepen K, Hazeu G, Andersen E and Verhoog D (2006). Protocols for spatial allocation of farm types. PD 471. SEAMLESS project report, pp 104.

FAO/IFA/IFDC (2003) Food and Agriculture Organization (FAO)/International Fertilizer Industry Association (IFA)/International Fertilizer Development Center (IFDC), Fertilizer Use by Crop, 5th ed., Rome.

Heywood (1998) Introduction to Geographical Information Systems. New York: Addison Wesley Longman.

Morphet C.S. (1997) A statistical method for identification of spatial clusters. Environmental and Planning A29, 1039 -1055.

Velthof, G.L. and P.J. Kuikman (2000) reduction of nitrous oxide emissions from crop residues. A systems analysis. Alterra-rapport 114-3, 81 pp. In Dutch.