AnimalChange deliverable 15.1 : A detailed description of the database architecture with guidelines for use.

Raphaël Martin

ANIMALCHANGE

SEVENTH FRAMEWORK PROGRAMME

THEME 2: FOOD, AGRICULTURE AND FISHERIES, AND BIOTECHNOLOGIES

SEVENTH FRAMEWORK PROGRAMME

Grant agreement number: FP7- 266018

DELIVERABLE 15.1

Deliverable title: A detailed description of the database architecture with guidelines for use

Abstract:

The aim of this document is to describe the process that allows creating the database and its web interface. Then it describes its architecture and the technical choice that had been made. Finally it explains how to use the database.

Due date of deliverable: M18 Actual submission date: M20

Start date of the project: March 1st, 2011 Duration: 48 months

Organisation name of lead contractor: INRA

Revision: V*1*

Dissemination level: PU

Release	Date	Reason of change	Status	Distribution
V1	31/08/2012	/	To be verified	Animal Change Project Manager
V1	15/10/2012		Validated by coordinator	

Table of Contents

1 INTRODUCTION 3

2 DATA USER AND PROVIDER SPECIFICATION 4

2.1 PRELIMINARY QUESTIONNAIRE 4

2.2 THE DATA EXPERTS COMMITTEE 4

2.3 TASK LEADER QUESTIONNAIRE 5

3 DATABASE AND WEB INTERFACE ARCHITECTURE 6

3.1 INFRASTRUCTURE 6

3.2 DATABASE DEVELOPMENT 7

3.3 WEB INTERFACE 9

4 USER GUIDE 10

4.1 GLOBAL DESCRIPTION 10

4.2 AUTHENTICATION 10

4.3 MODEL DRIVEN INTERFACE 11

4.4 EXPERIMENTAL SITE DRIVEN INTERFACE 13

4.5 DATA TYPE DRIVEN INTERFACE 14

5 CONCLUSION 14

6 ANNEX 1 15

1 Introduction

ANIMALCHANGE requires a systematic, comprehensive and integrative approach, where a multitude of experimental data, of socio-economic data, of GHG measurements, of farm data and of long-term observations along with meteorological observations is brought together with state-of-the-art process modeling approaches. This objective will be achieved by integrating three major types of scientific data, along with meteorological data and scenarios, from: (i) long-term observations spanning several times-scales (e.g. eddy covariance data, crop yields, animal production, long-term remote sensing data on soil moisture and vegetation activity, soil data bases), (ii) a network of established ecosystem and animal manipulation experiments in CP1 and CP2, (iii) a network of highly documented case-study farms in CP3. Moreover, detailed geo-referenced agricultural statistics data will be used in CP1 and in CP4 for the regional modeling of livestock.

In other terms, the database will create the interface between the data collections and the modeling activity, and more generally the link between CP1 to CP4. It will receive data inputs from partner research projects and provide data outputs to the scientific community and to stakeholders.

This work description, extracted partially from the AnimalChange proposal “Description Of Work”, describes the amplitude of the task, illustrating the heterogeneity of the data to handle with. Thus, in a first approach, we carried out an analyses on data to be expected (“Data user and provider specification”), allowing to design a well structured architecture of the database and to implement pre-process treatments, which will allow to control data quality, standardize and harmonize them and provide all the necessary meta-information essential for a correct data use and interpretation.

“Data user and provider specification” analyses quickly revealed that datasets to tackle during the AnimalChange project will be spatially and temporally heterogeneous, ranging from ecosystem manipulation experiments and long term observations at site level (eddy fluxes, yield history, etc.), to regional and continental level (soil maps, farming systems, remote sensing…) in CP1 and CP2, through farm scale in CP3. It will also contain the model outputs. So care was paid during the development phase to stay as flexible as possible, because of the risks of specific needs or modifications in later stages of the project are high.

2 Data user and provider specification

2.1 Preliminary questionnaire

At a first step before the official launch of the project, each key actor of the AnimalChange project has been contacted in order to capture their needs in terms of data storage and pre-processing treatment. This document, named “Functional analysis”, was sent to every WP and/or CP leader in order to get information on:

- Data importation (file format, file size, number of data per file)
- Link with other database (name, type of data, type of connection)
- Data validation (type of treatment)
- Meta-data (documentation, description with correspondant validation)
- Data manipulation (add, delete, modify)
- User roles (permission, visible page)
- Web interface (functionalities)

Because the project was at an early stage of advancement, we received only few answers to our questionnaire. At the 2011 kick-off meeting, a new group (data experts committee) was created, leading from now on the data provider specifications.

2.2 The data experts committee

During kick-off meeting of Opio in March 2011, as a result of the WP15 session, a group of data experts has been created in order to take a proactive approach to data collection. This group is composed by 11 data experts (one per data type) and 4 country experts (one per country outside Europe: South Africa, Tunisia, Brazil and Senegal). During this meeting we decided to setup a functional database by M24.

Table 1: Composition of data experts committee

Members per data type		
Member	**Source of data**	**Data type**
Katja Klumpp	Exp	Soil carbon/GHG flux (CO2, N2O)
Michel Doreau		Animal diet and CH4
Soeren Petersen		Manure/Emissions
François Gastal		Forage
Catherine Cochard		Rain manipulation
Juliette Bloor		CO_2
Jean-Yves Dourmad		Animal Heat stress
Raphaël Martin	GIS	Soil properties / Land use (manure)
Jean-Louis Drouet	Obs	Case study farm
François Gastal		Forage
Member	**Country**	
Eyob Tesfamariam	South Africa	
Hichem Ben Salem	Tunisia	
Paulo Di Faccio Carvalho	Brazil	
Amadou Tamsir Diop	Senegal	

The main objectives of this group are to:

- Provide a detailed data description;
- Check if the implemented database matches known needs;
- Discuss data policy/access rights.

Despite sufficient announcement for the first data experts committee meeting at M6 (planned more than three month before), a large number of experts were absent, indicating that the role of the committee was not well understood at this stage. The minutes of this meeting can be found in the MS60. In order to imply in a more efficient way the scientific community of AnimalChange, the data experts committee decided to contact each task leader separately in order to get a detailed description of the data they use/need for their task. The idea was to focus the development of the database around model and experimental sites.

2.3 Task leader questionnaire

In the questionnaire send, we asked each task leader to fill out three documents, including 3 sections:

- data needs for modeling,
- data provided by experiments,
- model outputs.

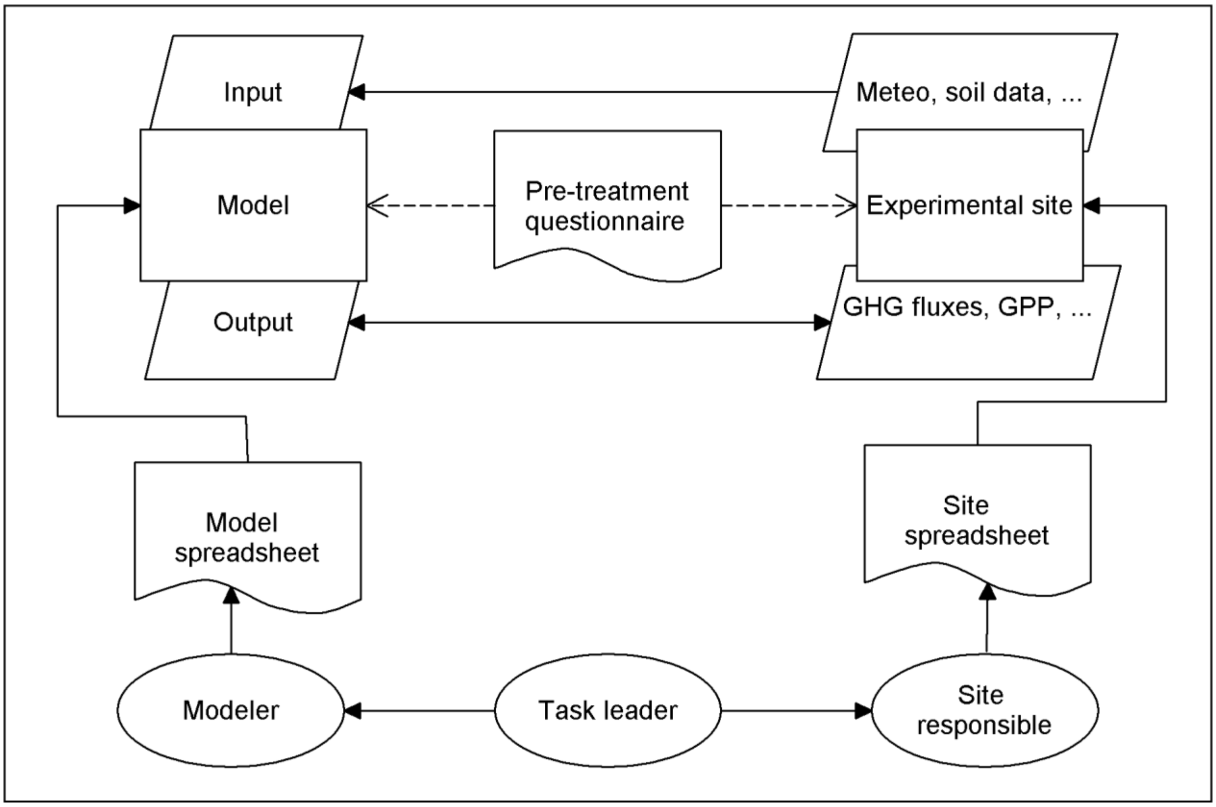

Figure 2: Flowchart of the task leader questionnaire

In order to fill in the corresponding spreadsheet, we asked each task leader to contact site PI's and modelers. Thanks to this description, we were able to characterize in details all models and experimental sites. Moreover, the questionnaire allowed us to specify operations needed to format data.

The evaluation of the questionnaire led to two documents that can be found on the AnimalChange collaborative platform. The collected information was used to develop the database and the website interface.

Table 2: Extract of the AnimalChange_Models spreadsheet

Variable\Model Name	INRA (CERES-EGC)	INRA (FARMSIM)
Optional input are in grey		
Field capacity	X	X
Maximum soil depth		X
NH4+ deposition other than gaseous NH3	X	X
NO3- deposition	X	X
Organic C pool	X	X
Organic N pool	X	X
pH	X	X
Plant wilting point	X	X
Relative root dry matter in different soil layers	X	X
Saturated hydraulic conductivity	X	X
Saturated soil water content	X	X
Texture (clay/silt/sand fractions)	X	X
Water content of lower soil boundary layer in autumn		X
Water content of lower soil boundary layer in spring		X
CLIMATE		
Atmospheric CO_2 concentration		Daily
Atmospheric NH_3 concentration	X	Daily
Average air temperature		
Global radiation	Daily	Daily
Minimun and maximum air temperature	Daily	Daily
Precipitation	Daily	Daily
Temperature		Daily
Water vapor pressure		Daily
Wind speed	Daily	Daily

3 Database and web interface architecture

3.1 Infrastructure

Analyses of spreadsheet from "Data user and provider specification" revealed that the type of the data and their size was very heterogeneous between tasks (from a few kilobytes to a hundred of megabytes per file). No available server was able to host such a database. We thus carried out a study on the best technical solution and took the decision to setup a centralized computing applications infrastructure.

This setup will improve the management of data such as the clustering of model inputs/outputs and experimental data. In collaboration with the INRA Information Technology (IT) group, we developed a general design providing the following infrastructure:

- Services for web applications and web interfaces with high networking features;
- Services for database storage and management with high computing performances;
- Files system storage with large capacity (some terabytes);
- Backup service.

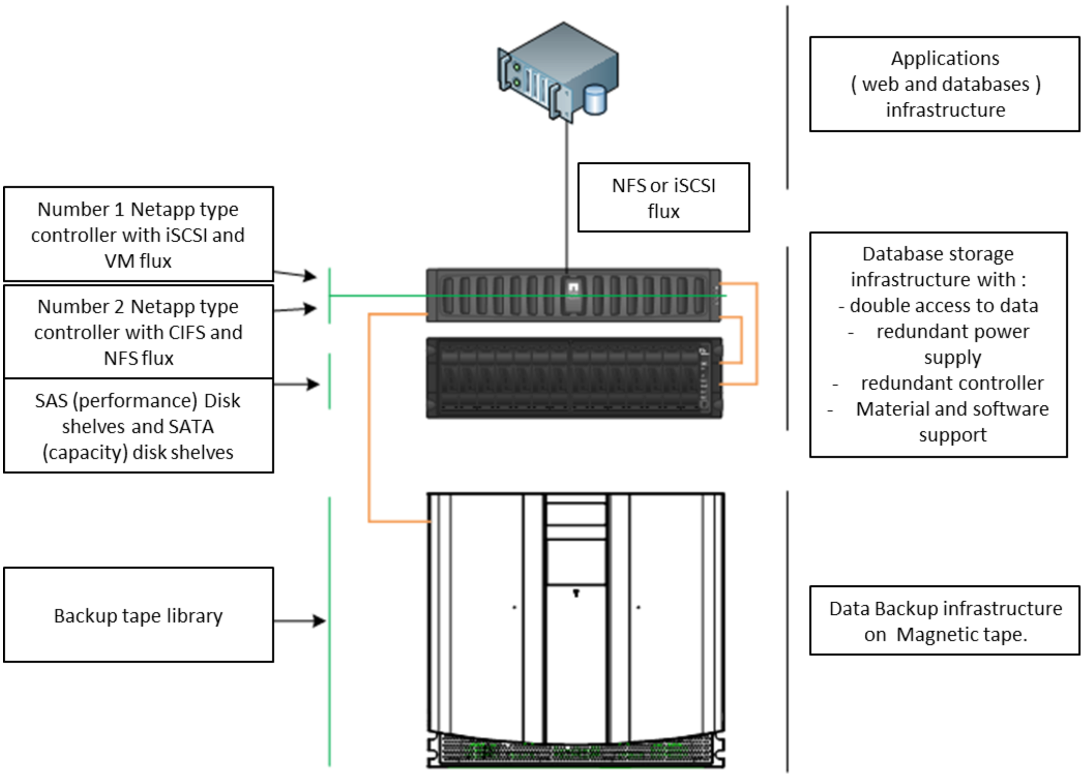

Figure 3: Scheme of the global organization and technical description of the database and web interface infrastructure

The selected technologies of the “database and web interface infrastructure” will give reliability, secured data, performance and a functionally/technically mastering service. Servers and data centralization permits costs reduction in terms of integration, administration and support.

This infrastructure will be based in the large INRA infrastructure at Clermont-Ferrand, dedicated to the scientific data managing. An external company ensures the technical support and INRA teams will administrate the application. A group of IT Experts already estimated the data features.

3.2 Database development

The database is developed under PostgreSQL, a powerful, open source object-relational database system. This database language has more than fifteen years of active development and a proven architecture that has earned it a strong reputation for reliability, data integrity, and correctness. PostgreSQL runs on all major operating systems, including Linux, UNIX and Windows. The language provides a fully ACID compliance. More precisely,

- A for **A**tomicity. Each transaction is "all or nothing": if one part of the transaction fails, the entire transaction fails, and the database state is left unchanged
- C for **C**onsistency. Any transaction will bring the database from one valid state to another
- I for **I**solation. Execution of transactions results in a system state that could have been obtained if transactions are executed serially, i.e. one after the other
- D for **D**urability. Once a transaction has been committed, it will remain so, even in the event of power loss, crashes, or errors.

The database architecture consists of a set of 18 tables. The two main tables are the “site” and “model” ones. Each of them has a description table. For the model table, several versions of the same model can be stored in the database thanks to the table versions. Table documentation refers to the files that explicit the model they are used for. Also, data inputs and outputs are linked to a model of a given version by a “run” operation. Moreover, each of these files is accessible by a given group through rights (table filegroup). A user, describes in the table of the same name, has a role and belongs to groups.

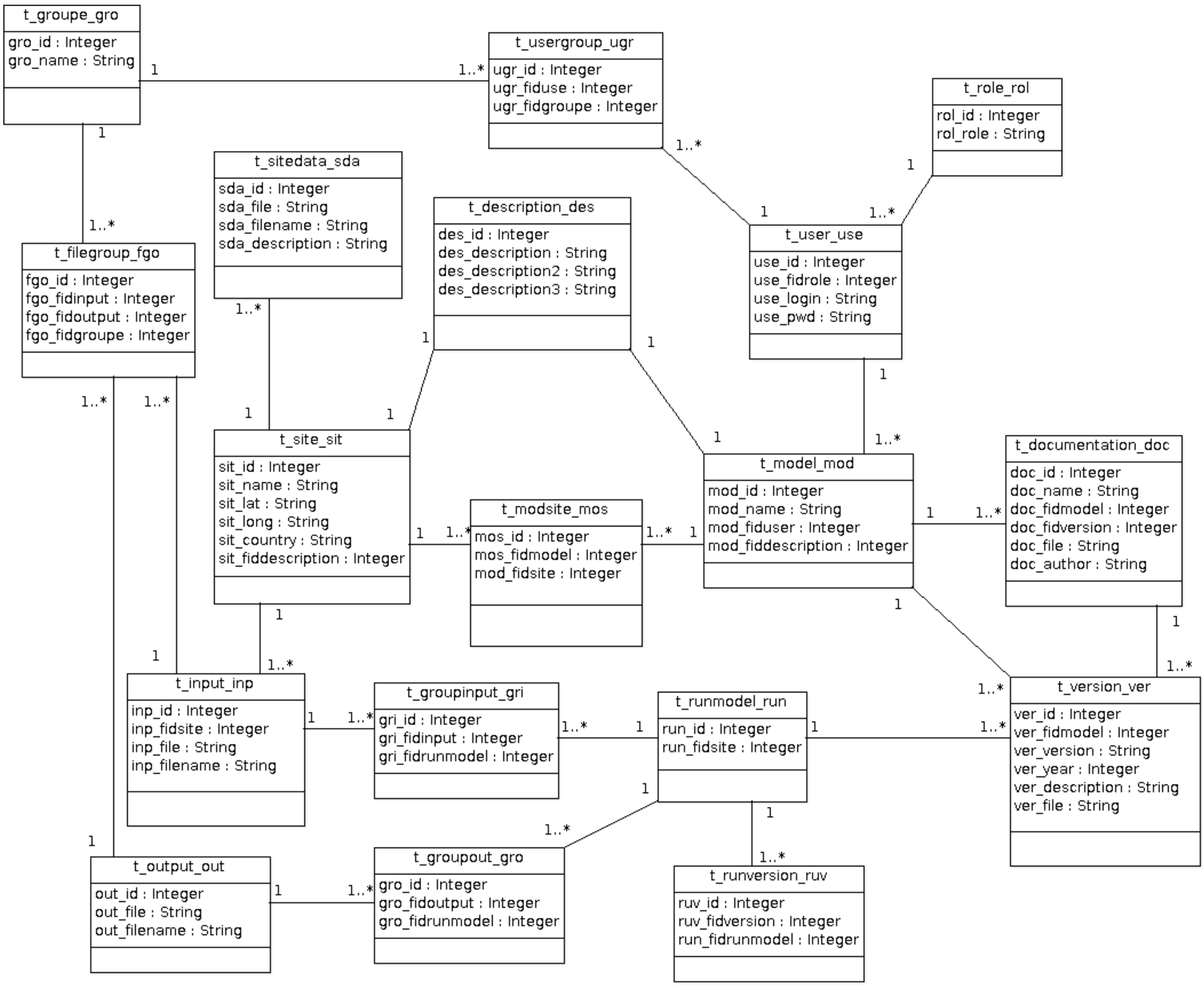

Figure 4: Database architecture

Each table is pre-fixed by the character “t”, its name and three unique letters to prefix its columns. The database is focused on the coupling between models and sites. This physical structuring of the database will be useful to provide an easy access to data through the web interface. Accordingly some tables only contain a description and references to files.

3.3 Web interface

In order to develop the web interface, the PHP programming language (general-purpose server-side scripting language originally designed for Web development to produce dynamic Web pages) was selected. The Model-View-Controller type of computer user interface was chosen as it is a well-known architecture to develop user interface and it allows separating the representation of information from the user's interaction with it. The *model* consists of application data and business rules, and the *controller* mediates input, converting it to commands for the *model* or *view*. A *view* can be any output representation of data, such as a chart or a diagram. The central idea behind MVC is code reusability and separation of concerns as it is shown in the following figure:

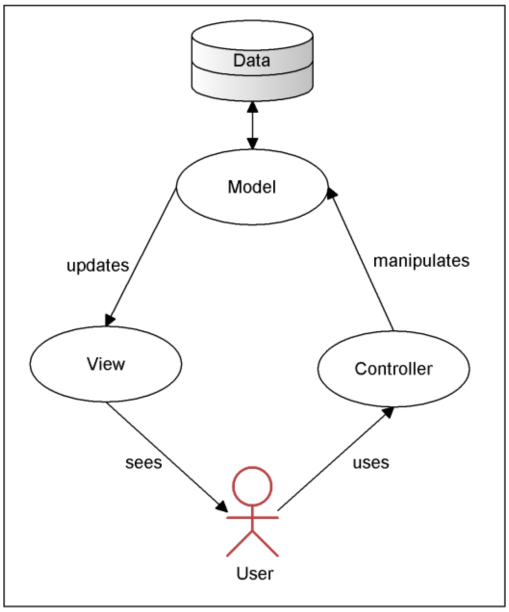

Figure 5: MVC user interface architecture

To implement this architecture, and in order to development and provide a standardized methodology, we used the framework Zend, which provides a lightweight loosely-coupled component library.

4 User guide

4.1 Global description

The main objectives of this chapter is to describe the main functionalities provide by the user interface.

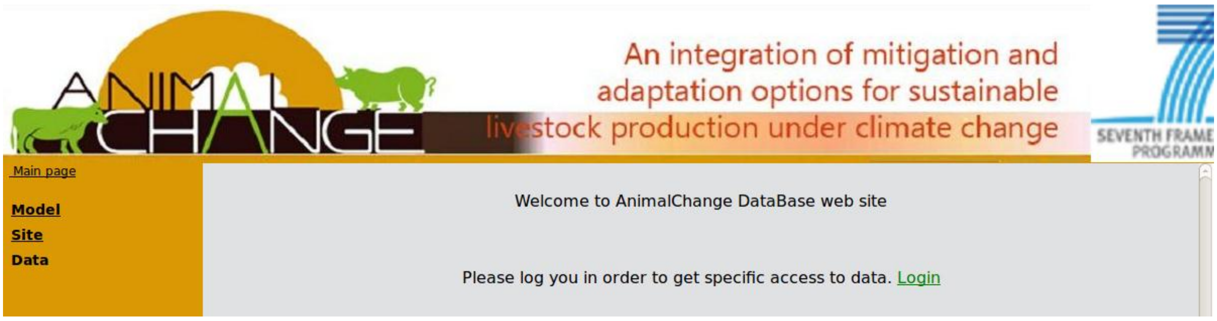

Figure 6: Main page of the AnimalChange user interface

The main page is composed of three parts:

- The upper part containing references to the AnimalChange project and the “Seventh framework programme”.
- A menu line (left part) allowing the user to navigate into the four sections of the website: the main page, the model driven page, the site driven page and the data driven page.
- The central part showing the information chosen by the user.

4.2 Authentication

Authentication is a process closely related to identification. The username identifies the user, while the password authenticates the user. In the following example, a user named “Raphael Martin” is connected by using the login/password functionality.

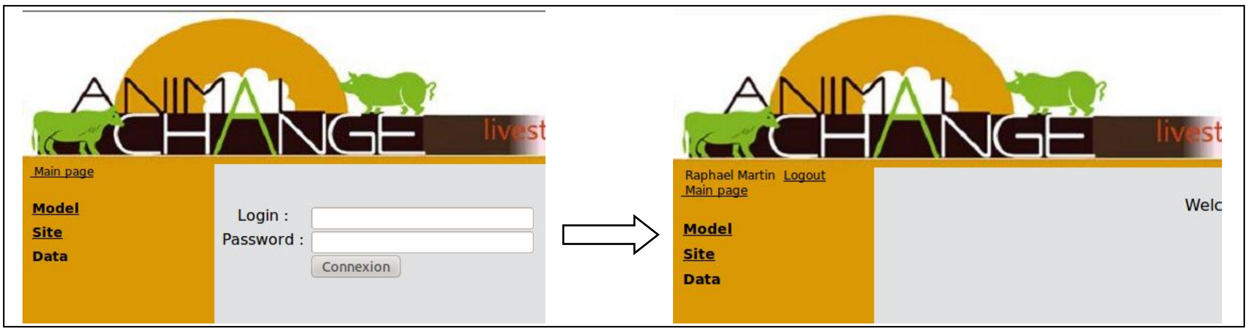

Figure 7: Process of authentication

Each user is provided with rights which allow him/her to navigate within the website (each web page can be shown by a specific list rights). Each user also belongs to “groups”, where files are linked to one or more groups. By this mean, it is possible to know whether a user has the right to download a file, or not.

4.3 Model driven interface

This section of the user interface concerns mainly:

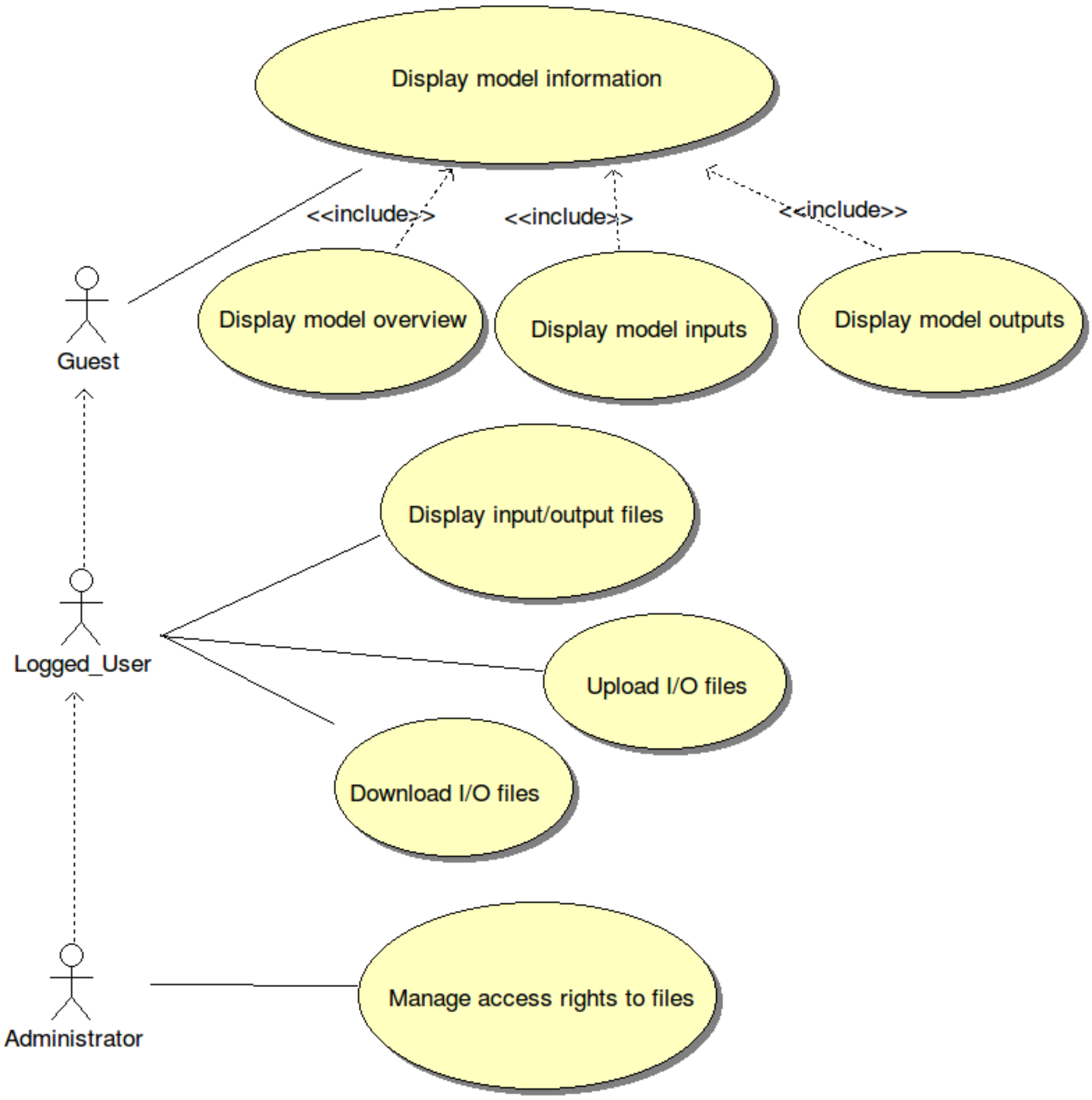

Figure 8: Use case diagram of model driven interface

First of all a list of models implied in AnimalChange (extracted from the AnimalChange_Models.xls) is provided and a model can be selected from this list by the user. Guest users (non-authenticated) have the possibility to see this list.

List of models				
ALTERRA /MITERRA	ASG/DT3	AU/FarmGHG	AU/FASSET	AU/Icafood
AU/Nutrients	IIASA/EPIC	IIASA/GLOBIOM	ILRI /GRANGE	INRA/ADAM-SelAction
INRA/AROPAj	INRA/CERES-EGC	INRA/FARMSIM	INRA/MELODIE	INRA/PASIM
INRA/STICS-GRASSLAND	JRC/CAPRI	LSCE/ORCHIDEE	TEAGASC/BEEFGEM	TEAGASC/Breen
TEAGASC/Dairy	TEAGASC/DNDC	TEAGASC/FAPRI	TEAGASC/GHG	TEAGASC/LCA
TEAGASC/N model	SAC/APSIM	SAC/COUP	SAC/DNDC	SAC/Manure-DNDC
SAC/MoBile DNDC	SAC/FSSIM	SAC/LU-CATCH	SAC/SACLP	SAC/MACC
UP/SWBSCI	UPM/ClimtateCrop			

Figure 9: List of AnimalChange models

Once this list is displayed, the user can pick one model in order to get further information on the model itself as well as on data input and output. If the user is authenticated and belongs to an appropriate workgroup, the user gets access to input and output files. Finally, depending of the status of the logged in user, the administration area also allows uploading data. By doing so the user has to associate tags label to describe the nature of the data, which later allow to data users to do a research on a given variable and year (see section 4.5). The following figure explains in more details the dynamic of the model driven part:

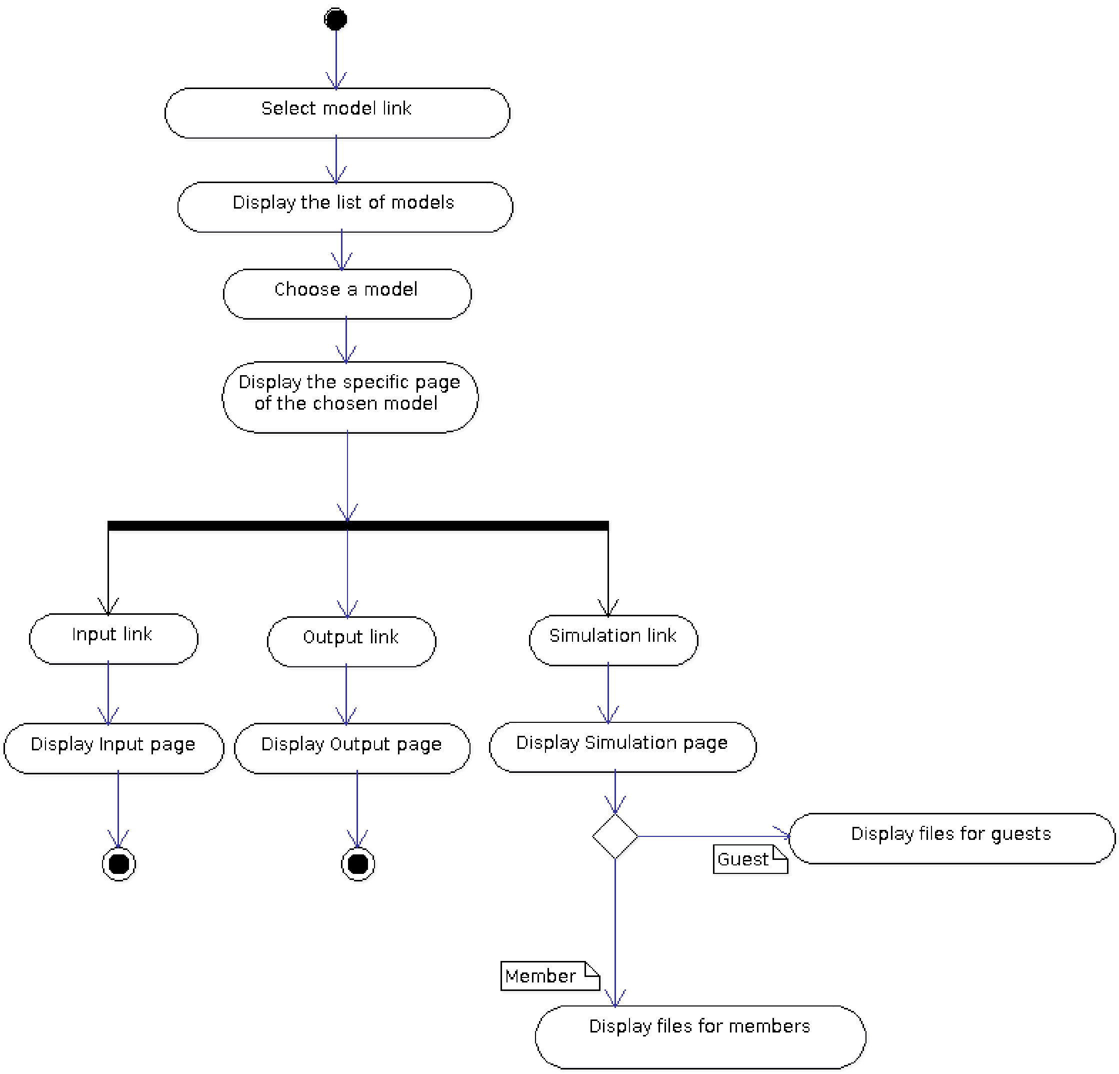

Figure 10: Activity diagram of the model driven model driven interface

To illustrate this diagram, an example (screenshot) data inputs necessary for the MITERRA model:

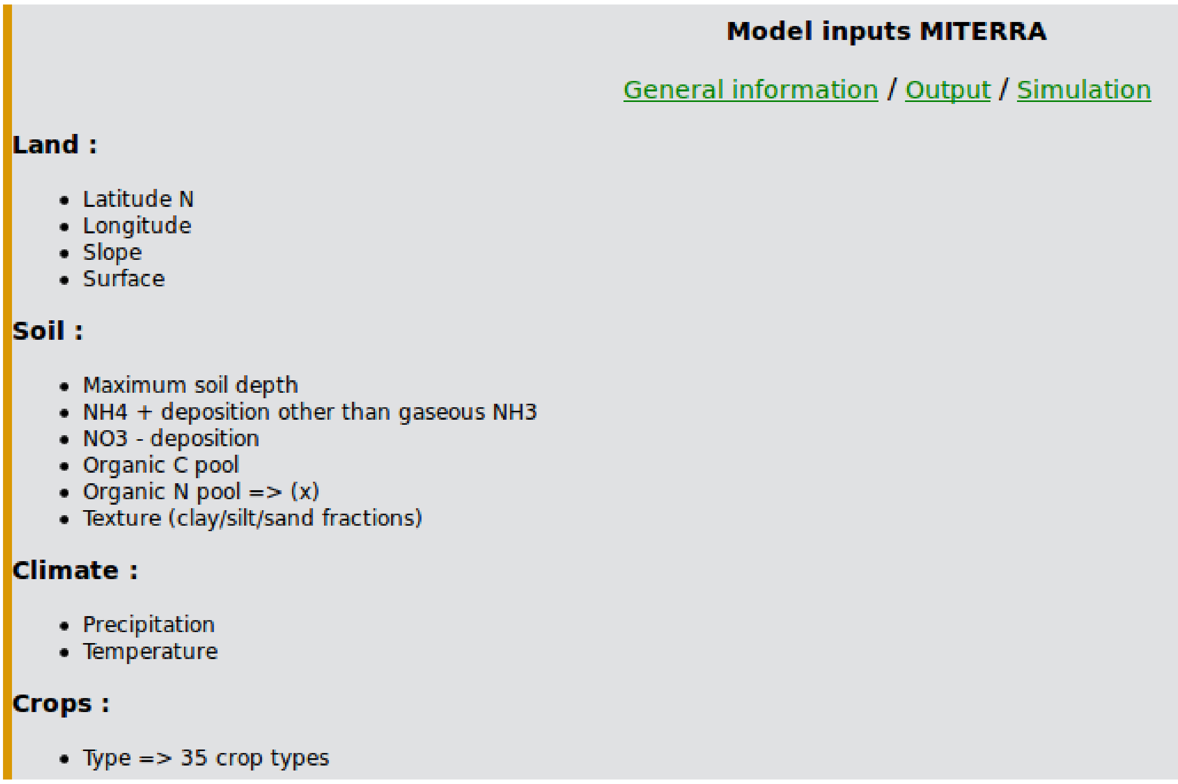

Figure 11: Illustration by a screenshot of the model input page for MITERRA model

4.4 Experimental site driven interface

In the same way as the modeling interface, this part provides the data link to experimental sites. Here the user selects an experimental site from the map. After clicking onto a site, a page appeared with an overview description of the site and a list of corresponding files:

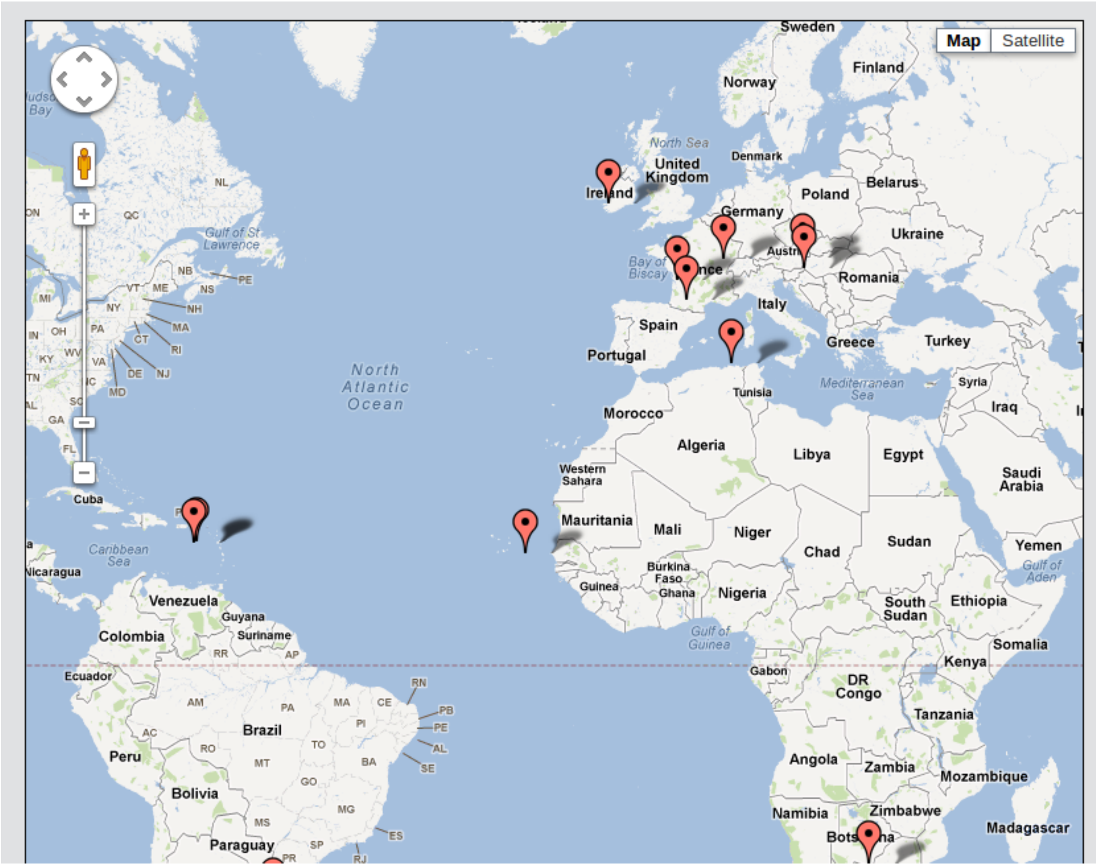

Figure 12: Map to select experimental sites

(Note that some other sites will be added to this map)

4.5 Data type driven interface

In this screen, the user is able to search already uploaded files. To do so, the user has to specify an experimental site, variable and time interval. This part of the development is still in progress.

5 Conclusion

For the moment, the database and user web interface are well developed and installed on the local server. Despite a lack of complete information the database provides a flexible, secure and robust way to store data and share them.

The server will be operational in spring 2013. At that time, the final version of the database will be installed and public access will be provided. Once this operation is done, a training session on the project database (see MS61) will be carried out. This session initially planned for M18 will be facilitate the use of the database. Moreover the present document will be regularly updated in order to contain the latest modifications of the database structure, web interface and the updated user guide for the AnimalChange database.

6 Annex 1

Deliverable Check list (to be completed by the Deliverable Leader)

	Check list		Comments
BEFORE	I have checked the due date and have planned completion in due time		**Please inform MGT team of any foreseen delays**
	The title corresponds to the title in the DOW (workplan)		**If not please inform MGT team with justification**
	The content corresponds to the description in the DoW (workplan)		
	The dissemination level corresponds to that indicated in the DoW (workplan)		
	The contributors (authors) correspond to those indicated in the DoW (workplan)		
	The Table of Contents has been validated with the WP Leader		**Please validate the ToC with the WP leader before drafting the deliverable**
	I am using the AnimalChange deliverable template (title page, styles etc)		**Can be found under Deliverables⇨Template on the collaborative workspace**
AFTER	The deliverable has been reviewed internally in my organization		**Please ask colleagues to review the deliverable for its scientific content**
	The deliverable has been reviewed by all contributors (authors)		**Make sure all contributors have reviewed and approved the final version of the deliverable. You should leave sufficient time for this validation.**
	I have done a spell check and had the English verified		**Ask a colleague with a good level of English to review the language of the text and do a spell-check too.**
	I have sent the final version to the WP Leader for approval		**Please send the final validated draft to the Coordinator (MGT Team) & ExC for validation and then to the EC.**

Deliverable Review Feedback

This deliverable is:

Excellent (the deliverable has fully achieved its objectives and technical goals).

Good (the deliverable has achieved most of its objectives and technical goals with relatively minor corrections to be made).

Unsatisfactory (the deliverable has failed to achieve critical objectives and needs to be significantly revised).

The following modifications should be made:

Page n°	Changes to be made	Response to requested changes