HAL
open science

Test et adaptation d'une base de données

Matthias Blanchandin

To cite this version:

Matthias Blanchandin. Test et adaptation d'une base de données. 2015, 35 p. hal-01269080

HAL Id: hal-01269080

https://hal.science/hal-01269080v1

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INRA ORLEANS

Test et adaptation d'une base de données

Rapport de Stage

Blanchandin Matthias

Année Universitaire 2014-2015

Maîtres de stage : Frédérique Santi & Odile Rogier

Enseignant tuteur : Mirian Halfeld Ferrari

IUT Orléans département Informatique

Remerciements

Je tiens à présenter mes remerciements

à Madame Frédérique Santi ainsi qu'à Madame Odile Rogier, mes maîtres de stage, pour m'avoir accepté en tant que stagiaire durant ces dix semaines ;

à Gilles Pilate, directeur de l'unité AGPF de m'avoir accueilli au sein de son unité ;

à Monsieur Yannick De Oliveira pour son aide et ses conseils concernant SHiNeMaS ;

à Monsieur Chabin sans qui ce stage n'aurait pas été possible ;

à Madame Halfeld Ferrari, mon professeur encadrant à l'IUT ;

à l'ensemble des stagiaires et doctorants présents à l'INRA, pour leur accueil et leur gentillesse ;

à l'ensemble des professeurs de l'IUT d'Orléans pour leurs enseignements qui m'ont permis d'acquérir les compétences nécessaires pour réaliser ce stage et continuer mes études.

Sommaire

Remerciements 2
Introduction 4
I Présentation du contexte 5
1. Contexte général 5
1.1. INRA d'Orléans 5
1.2. UAGPF 5
1.3. Projet SPEAL 6
1.4. Projet SHiNeMaS 6
2. Contexte technique 7
3. Sujet du stage 8
4. Description 8
4.1. Données biologiques 8
4.2. SHiNeMaS 9
II Travail réalisé 12
1. Découverte 12
1.1. Découverte du contexte 12
1.2. Découverte sur le terrain 12
2. Découverte de l'outil 16
2.1. Découverte de SHiNeMaS 16
2.2. Prise en main 12
3. Analyse 17
3.1. Des fichiers de données 17
3.2. Des fichiers d'entrée de SHiNeMaS 18
3.3. Lien entre les fichiers d'entrée et les fichiers de données 21
3.4. Fichier de mesure 21
3.5. Gestion de la sélection 22
4. Création 22
4.1. Création des fichiers 23
4.2. Simulation pour les années suivantes 24
4.3. Enregistrement 24
5. Insertion 25
6. Rédaction du tutoriel 26
III Bilan 27
1. Technique 27
2. Humain 27
Conclusion et perspectives 28
Glossaire 29
Bibliographie 30
Annexes 31

Introduction

Dans le cadre de l'obtention du DUT informatique, chaque étudiant en informatique de l'IUT d'Orléans doit réaliser un stage d'une durée minimale de 10 semaines afin de découvrir le monde du travail, valider ses acquis, et de découvrir des nouvelles technologies.

J'ai réalisé mon stage au sein de l'Institut Nationale de la Recherche Agronomique (INRA) d'Orléans situé à Ardon, dans l'Unité de Recherche Amélioration, Génétique et Physiologie Forestières (UAGPF), dans le cadre du projet de Sélection Participative d'Espèces Annuelles ou Ligneuses adaptées aux pratique agroécologiques (SPEAL).

Durant mon stage, j'ai eu pour mission de tester et d'adapter les fichiers utilisés sous Excel par Madame Frédérique Santi, pour la base de données « Seed History and Network Management System» (SHiNeMaS) développée à l'INRA du Moulon (91), par l'équipe de Monsieur Yannick De Oliveira. Cette base de données a été initialement développée pour stocker un volume important de données concernant le programme de sélection participative[1] sur le blé tendre.

Pendant les dix semaines de mon stage, j'ai été encadré par Madame Frédérique Santi, ainsi que par Madame Odile Rogier.

Dans un premier temps, j'expliquerais le contexte du stage, aussi bien général que technique. Dans un second temps j'expliquerais le travail réalisé, et pour finir je ferais un bilan humain et technique ainsi qu'une conclusion.

NB : L'intégralité des photographies dans ce rapport a été prise par moi-même, avec l'accord des personnes pouvant se trouver dessus.

Dans les paragraphes suivants, l'emploi du « je » correspond à un travail et une réflexion venant de moi-même ; l'emploi du « nous » correspond au travail venant de Madame Santi et moi-même.

I Présentation du contexte

1) Contexte Général

1.1) INRA d'Orléans

L'INRA (Institut National de la Recherche Agronomique) est un organisme français de recherche agronomique fondé en 1946. Il est le premier institut de recherche agronomique en Europe et deuxième dans le monde en nombre de publications en sciences agricoles et en sciences de la plante et de l'animal. L'INRA mène des recherches finalisées pour une alimentation saine et de qualité, pour une agriculture durable, et pour un environnement préservé et valorisé. L'INRA a pour missions de produire et diffuser des connaissances scientifiques, de concevoir des innovations et des savoir-faire pour la société, d'éclairer les décisions des acteurs publics et privés de par son expertise, de développer la culture scientifique et technique et participer au débat science-société, et de former à la recherche scientifique.

1.2) UAGPF

L'Unité de Recherche Amélioration, Génétique et Physiologie Forestières (UAGPF) fait partie du centre INRA Val-de-Loire et rassemble des compétences en génétique, génomique et physiologie appliquées à l'étude des arbres forestiers.

L'UAGPF est impliquée dans l'évaluation et la gestion de la diversité génétique ainsi que dans l'étude des interactions entre les variétés améliorées et les populations sauvages correspondantes, aidant ainsi à la définition de stratégies de gestion et de conservation des ressources génétiques forestières.

Les recherches menées visent à valoriser les ressources génétiques forestières en vue d'une production durable de bois d'œuvre et de biomasse[2], tout en prenant en compte l'impact écologique des populations domestiquées sur l'écosystème et un contexte climatique changeant. L'UAGPF conduit des programmes d'amélioration génétique sur 6 espèces forestières : le douglas, le mélèze, le pin sylvestre, le frêne, le merisier et le peuplier ; et s'investit dans des stratégies innovantes en sélection et en diffusion du progrès génétique. Afin de remplir ces objectifs, une approche de biologie intégrative est développée principalement sur le modèle peuplier pour mieux comprendre le déterminisme génétique des caractères complexes d'intérêt participant au développement de l'arbre, à son adaptation aux contraintes environnementales, concourant ainsi à la construction du phénotype.

1.3) Projet SPEAL

Le projet de Sélection Participative d'Espèces Annuelles ou Ligneuses adaptées aux pratiques agro écologiques[3] (SPEAL) est un projet dans lequel des généticiens et agronomes issus des centres de recherche INRA du Val de Loire (Orléans et Tours) et du Moulon, et leurs partenaires, pépiniéristes et forestiers volontaires de la Région Centre travaillent ensemble dans la mise en place d'une démarche expérimentale pérenne de sélection participative[1], co-construite avec les professionnels, de sélection de nombreuses variétés de blé, maïs, tournesol et d'arbres feuillus adaptées à des systèmes agricoles durables et favorables à la biodiversité.

A terme, les variétés obtenues pourront être multipliées dans chaque exploitation agricole ou être homologuées par le ministère de l'agriculture.

Pour résumer, le projet SPEAL a pour finalités :

- de favoriser le développement de systèmes agro-écologiques[3], dont l'agroforesterie ;
- de créer des variétés[4] pouvant être multipliées dans chaque exploitation ou être homologuées, grâce à une méthodologie de sélection participative[1], associant les acteurs à la recherche publique et les professionnels du secteur agricoles ;
- d'établir une expérimentation-système co-construite dans une parcelle drainée représentative du terroir agricole qui pourra servir de vitrine et de lieu d'échanges de savoirs.

1.4) Projet SHiNeMaS

Le projet SHiNeMaS (Seeds History and Network Management System) comprend une base de données qui a été développée par et pour les chercheurs de l'INRA du Moulon, et son application Web, développée avec Django[5]. Ce projet a été mis en place à l'INRA du Moulon pour permettre une gestion durable de la biodiversité. Des agriculteurs regroupés au sein du Réseau Semences Paysannes (RSP) participent à cette gestion en cultivant et en échangeant des variétés-populations de blé, depuis 2003. L'équipe de recherche du Moulon essaye de comprendre l'impact des pratiques et des échanges de semences sur la structure de la biodiversité dans le cas du blé tendre.

Dans ce contexte, une base de données a été développée à partir de 2010 pour stocker l'ensemble des données récoltées dans les différentes fermes impliquées dans le projet (diffusion des semences, mélange, création variétale, reproductions, données sur les pratiques culturales, phénotypiques et génétiques)

La base de données a pour objectifs de :

- Modéliser les réseaux représentant l'histoire de la vie des variétés/lots de semences
- Valoriser et exploiter ces informations à plusieurs niveaux (recherche, associations, paysans)

De plus, elle doit permettre de répondre à un certain nombre de questions telles que :

- ✓ « Qui possède quel lot de graines ?»
- ✓ « Qui a le plus donné sur une période déterminée ? »
- ✓ « Qui a le plus reçu ? »
- ✓ « Combien de croisements, sélections intra-variétales y a-t-il eu ? ».

2) Contexte technique

Durant mon stage, j'ai été amené à travailler principalement avec Excel 2013, qui est un logiciel tableur de la suite bureautique Microsoft Office, développée et distribuée par l'éditeur Microsoft. C'est l'outil le plus utilisé dans le domaine de la biologie.

J'ai aussi travaillé sur une base de données –SHiNeMaS- qui a été développée à l'aide du Framework python Django.

J'ai travaillé avec ses différents outils sur un ordinateur situé dans les bureaux de l'INRA possédant comme système d'exploitation Windows 8, développé par Microsoft.

3) Sujet du stage

A l'heure actuelle, toutes les données relatives à la sélection participative se trouvent dans différents fichiers Excel (construits par Madame Santi). Il devient donc urgent de les regrouper au sein d'une base de données.

Il m'a été demandé durant ce stage d'adapter les fichiers utilisés par Madame Santi à la base de données SHiNeMaS, déjà existante et toujours en développement, puis de les tester sur cette même base de données. Ces tests ont pour but de voir si cette base correspond aux besoins exprimés, c'est-à-dire de permettre une meilleure gestion des données, assurant leur traçabilité. De plus, le stockage en base de données permet de faciliter la gestion de plusieurs manières. Tout d'abord, ce stockage requiert en amont une formalisation des fichiers, de telle sorte qu'ils aient une base fiable et soient facilement compréhensibles par les différents acteurs. De plus, la base de données permet un stockage centralisé et sécurisé, n'obligeant plus Madame Santi à avoir toute une arborescence de dossiers/fichiers à gérer, et permet donc d'optimiser l'accès à l'information, notamment par un système de recherche.

Enfin, pour faciliter l'utilisation et la gestion future de la base de données, il m'a été demandé de réaliser un tutoriel concernant l'insertion d'un fichier de données « arbres » dans la base.

Les utilisateurs seront principalement Madame Santi, ainsi que les différents acteurs de la sélection participative tels que les pépiniéristes qui utiliseront certaines parties des fichiers traités, tels que les plans.

4) Description

4.1) Données biologiques

Les données correspondent à toutes les informations liées au suivi du projet de sélection participative, c'est-à-dire toutes les données par espèce (Aulne, etc.). Cela concerne la description de l'espèce et sa traçabilité : on sait donc pour une espèce donnée, le nombre plants plantés, leurs origines (le fournisseur, la pépinière, etc.), les dates de plantation, leurs places dans les parcelles, les commentaires par plant si besoin, les mesures réalisées sur ces plants ou encore leurs natures. Lors de la sélection participative, les plants sont repartis selon un ordre spécifique : quatre témoins suivis de quatre « choisis », les « témoins » servant de comparaison. Parmi les quatre « choisis », celui qui sera le meilleur, c'est-à-dire celui qui possédera les meilleures caractéristiques (le plus grand, le plus droit, etc.), sera de nouveau « choisi », est nommé « composant », justement pour le différencier des autres « choisis » dans les fichiers.

De même, le but est de suivre leur évolution dans le cadre spatio-temporel, de telle sorte, que l'on souhaite savoir où se trouve tel plant (définissable via un x et un y sur un plan), ainsi que son évolution au cours des années, pour pouvoir comparer les différentes mesures réalisées, et vérifier quels sont les « meilleurs ».

Les données biologiques à adapter et à insérer dans la base de données sont stockées dans des fichiers Excel. Chaque fichier contient entre quatre et cinq onglets. Le premier onglet est celui qui nous intéresse : il stocke toutes les données concernant une espèce et est constitué de 18 en-têtes. Nous avons donc un fichier par espèce (*Figure1*, p9).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Année	Pépinière	Espèce	Origine	Age	Choisis	Fournisseur-n°	N° cert	Destination	Catégories plage témoins	vendables	Taux réel	Taux théo	Commentaires-pr
2	2014	Bauchery	Aulne glutineux	AGL130	1	8	Vilmorin-K71661	F11-04R008	AGL-SULLY	30.50.80	80 000	10 000	2 000	THEILLAY - 10 choisis (et pas
3	2014	Lemonnier	Aulne glutineux	AGL130	1	4	Vilmorin-K71661	F11-04R008	AGL-SULLY	20.40.60.80	4 000	1 000	2 000	
4	2014	Naudet-P	Aulne glutineux	AGL130	1	12	Vilmorin-K71661	F11-04R008	AGL-SULLY	15.30.50.80	18 000	1 500	2 000	Préchac
5	2014	Claireau	Aulne glutineux	AGL130	2	19	Vilmorin-K37119	FR52-12R004	AGL-SULLY	50.80.100	15 000	789	2 000	17 choisis en 2013, dont 12 r +7 nouveaux, = 19 (et pas 8).
6	2014	Wadel	Aulne glutineux	AGL130	2	7	Vilmorin-K37119	FR52-12R004	AGL-SULLY	80.120.>	48 800	6 971	4 000	7 et pas 12 plants retrouvés
7	2014	Wadel	Aulne glutineux	AGL901	2	15	Vilmorin-K40528	FR43-12R025	AGL-CHEU	80.120.>	60 000	4 000	2 300	15 et pas 28 plants retrouvés
8														
9														

Figure 1 - Exemple de fichier par espèce

Le dernier onglet est aussi d'intérêt car il contient les mesures réalisées dans chaque verger. Ce sont ces données que nous souhaitons adapter et insérer dans la base de données.

Le reste des onglets est composé des plans théoriques des vergers dans lesquels cette espèce est plantée ; et d'un plan réel par verger par onglet.

4.2) SHiNeMaS

La base de données SHiNeMaS est une base de données toujours en cours de développement s'insérant dans le projet SHiNeMaS expliqué précédemment. SHiNeMaS est basée sur le système de gestion de base de données relationnelle PostgreSQL. Elle comporte une interface web développée à l'aide du Framework Django[5]. Cette interface web est séparée en deux parties.

La partie administrateur (Cf annexe 3, p33), permet de réaliser plusieurs actions sur les différentes entités disponibles : sur les acteurs, les personnes participant au projet ; sur les analyses, réalisées grâce à un package R ; sur les authentifications ; etc.

La deuxième partie est composée de plusieurs onglets correspondant aux différentes fonctionnalités accessibles, telles que la possibilité de réaliser une analyse grâce au package R, de faire une recherche avancée suivant différents paramètres.

Elle dispose aussi d'une interface d'insertion des données (Cf annexe 2, p32), permettant d'insérer des fichiers de données dans la base. Cette insertion nécessite deux fichiers : un fichier dans lequel se trouvent les données, ainsi qu'un fichier dit *méthode*.

Les fichiers de données actuellement utilisés à l'INRA du Moulon sont définis selon plusieurs types : croisement, reproduction, sélection, mélange, etc. Chacun de ces types correspond à une action effectuée sur un lot de graine (Cf. Figure 2). L'interface d'insertion de données est commune et il est possible de choisir le type de fichier à insérer : cross, diffusion, mixture, etc.

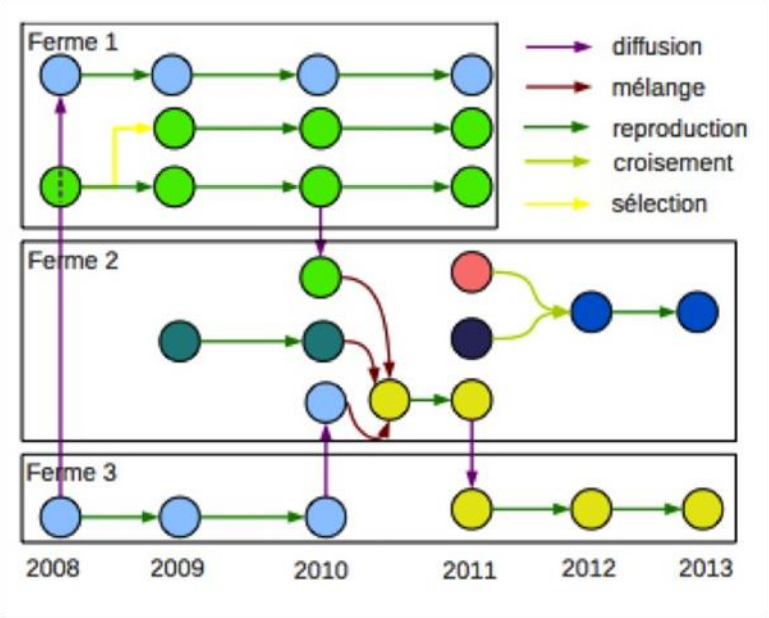


Figure 2 - Modélisation des différentes relations entre les lots de graines (P. Rivière, 2014). Les lots de graines sont représentés par des ronds de couleur. Chaque couleur correspond à une variété

Le type de fichier qui correspondait le plus à notre besoin est de type *reproduction* (*Figure 3*, p11). Je décrirai son format dans la suite du manuscrit.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	project	sown_year	harvested_y	id_seed_lot_	intra_selecti	etiquette	split	quantity_sown	quantity_harvested	block	X	Y	semis	semis$date	desherbage	fertilisat
2	PraiFou	2006	2006	Donata_AVM_2006				20000	15000	1	A	1	Ã la volÃ©e	10/02/2006		N
3	PraiFou	2006	2006	Donata_AVM_2006				20000	15000	1	A	2	Ã la volÃ©e	12/02/2006		N
4	PraiFou	2006	2006	Dact-precoce_AVM_2006				20000	1700	1	A	3	semoir en lig	10/02/2006	Stade 3 feuil	N+P
5	PraiFou	2006	2007	Dact-tardif_AVM_2006				20000	18000	1	A	4	Ã la volÃ©e	10/02/2006		
6	PraiFou	2006	2008	Glomerata-variegata_AVM_2006				20000		1	B	2	Ã la volÃ©e	13/02/2006	levÃ©e	
7																
8																
9																
10																
11																

Figure 3 - Exemple de fichier de reproduction

Les fichiers méthodes permettent quant à eux de réaliser la jonction entre le fichier de données et la base de données. A chaque fichier de données correspond un fichier méthode. Comme nous le verrons plus tard, dans le fichier de données, il y a des en-têtes obligatoires, et d'autres non-obligatoires. Les en-têtes non-obligatoires sont décrits dans le fichier méthodes. L'utilité des fichiers méthodes va être détaillée par la suite.

Voici un schéma (*Figure 4*, p11) pour récapituler les différents fichiers nécessaires pour SHiNeMaS :

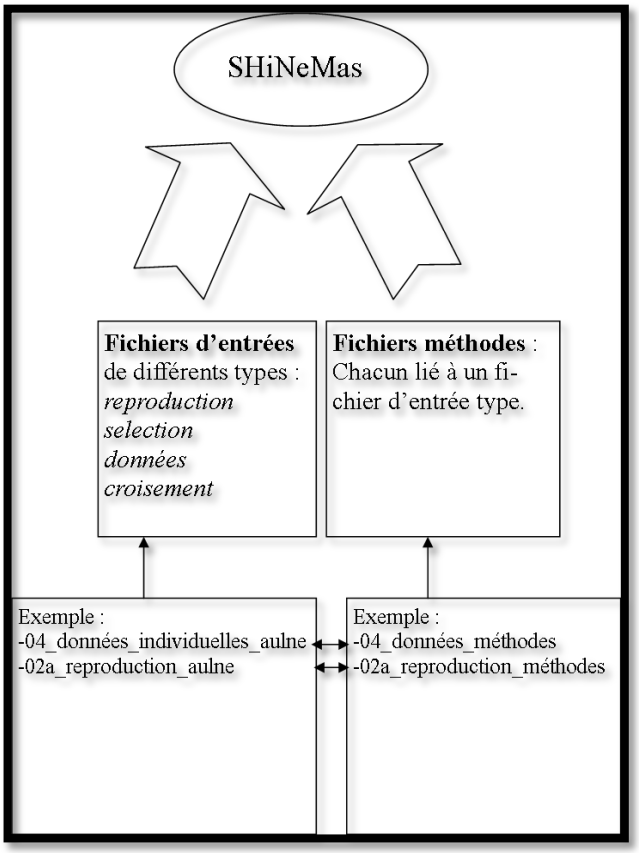


Figure 4 - Schéma des fichiers d'entrée de la base de données (M Blanchandin)

II Travail réalisé

1) Découverte

1.1) Découverte du contexte

Dans un premier temps, j'ai dû m'adapter à mon environnement de travail. C'est-à-dire que j'étais un néophyte dans le domaine de la biologie, j'ignorais l'existence même de la sélection participative, son but et tout ce qui en découle. Durant la première semaine, j'ai donc consacré mes heures à me renseigner sur ces sujets, grâce notamment au site réalisé par Madame Santi (www.selection-participative.fr), mais aussi aux différents documents qui m'ont été fournis, tel que différentes présentations Power Point pour présenter le sujet de la sélection participative[1]; ou encore la « Fiche d'identification du projet » qui résume, tant pour les débutants que les experts, les moyens et finalités du projet SPEAL.

Dans un second temps, j'ai découvert les données que j'allais avoir à traiter, qui étaient stockées au format Excel, lors de la préparation de feuilles récapitulatives pour les pépiniéristes avec ma tutrice. Malgré le fait que je sois novice dans ce domaine, cela m'a permis d'avoir un premier aperçu sur ces données.

Madame Santi m'a alors proposé de l'accompagner lors de sa rencontrer avec un pépiniériste, à Chéu à la pépinière Naudet.

1.2) Découverte sur le terrain

Le 5 Mai, nous nous sommes donc rendus à Chéu, à 2h30 d'Orléans (*Figure 5*, p12).

Lors de cette excursion, j'ai enfin pu mettre des images sur les concepts qui étaient représentés sur les fichiers Excel. Pendant la matinée, nous avons aidé le pépiniériste et ses employés à préparer les plants (élagage des branches ainsi que les racines trop longues) pour ensuite qu'ils les plantent, et installent les protections pendant que nous mesurions ceux qu'ils venaient de planter.

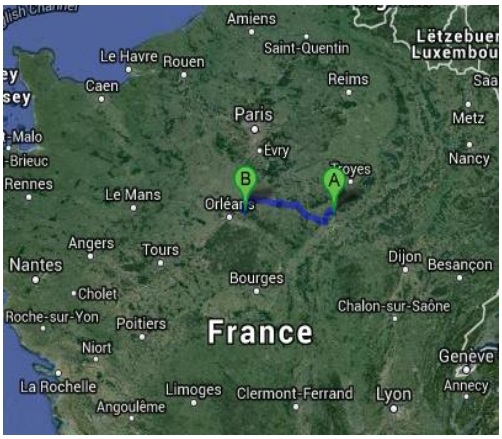


Figure 5 – Plan Orléans-Chéu

L’après-midi, nous avons continué cette même tâche, mais dans une autre parcelle (Photographies 1 et 2).

Photographie 1- Verger d’aulnes à Chéu (Crédit : M. Blanchandin)

Les plants sont disposés de cette façon : 4 choisis, 4 témoins, 4 choisis, 4 témoins, … ainsi que des charmes pour combler les rangées.

Photographie 2 – Verger d’aulne à Chéu (Crédit : M. Blanchandin)

Après avoir réalisé ces tâches, nous sommes allés durant l'après-midi dans la pépinière pour réaliser une présélection parmi les plants, dans quelques planches de semis.

Nous avons donc parcouru plusieurs rangées de différentes espèces telles que l'alisier, le cormier, le tilleul ou encore le tulipier ; pour choisir les individus qui paraissaient être les « meilleurs ». C'est-à-dire que parmi tous les plants, il fallait sélectionner celui qui paraissait le plus grand, le plus large, le plus droit, tout en le comparant avec les plants les plus proches. Une fois l'individu sélectionné, nous l'étiquetions (Photographie 3).

Photographie 3 – Exemple d'individu choisi (Crédit : M. Blanchandin)

Plus tard, le 11 Juin, nous sommes allés à Sully-la-chapelle, à la pépinière Claireau, pour réaliser des mesures sur des plants d'Aulnes et de Cormiers (Photographie 4).

Photographie 4 – Verger à la pépinière Claireau (Crédit : M. Blanchandin)

Nous nous sommes rapidement aperçus que les plants n'avaient pas été répartis comme défini sur le plan confié avec les jeunes plants venant des différentes pépinières.

Cela nous a aussi amenés à rediscuter de la clarté des plans à fournir.

2) Découverte de l'outil

2.1) Découverte de SHiNeMaS

Le 6 Mai, nous sommes allés avec Frédérique Santi ainsi qu'Odile Rogier à l'INRA du Moulon, situé à Gif-sur-Yvette au Sud-ouest de Paris, pour rencontrer Yannick De Oliveira, qui a participé au développement de la base de données SHiNeMaS.

Il nous a présenté la base de données à l'aide d'un tutoriel (*Prairie Fourragères*), pour nous montrer comment elle fonctionnait et quelle pourrait être son utilité pour nous. Par la suite, Monsieur De Oliveira m'a envoyé l'ensemble des tutoriels (*blé*, *maïs*, *praires fourragères*), les présentations PowerPoint réalisées pour présenter la base lors de formations, ou encore les rapports de 3 étudiants/doctorant en bio-informatique qui ont travaillé sur la base (Cf Bibliographie, p30), pour bien m'aider à appréhender et comprendre l'organisation et la construction de la base de données SHiNeMaS.

2.2) Prise en main

Il a donc fallu que je découvre et prenne en main l'outil que j'allais utiliser : SHiNeMaS. Monsieur De Oliveira a mis à notre disposition une version de test de SHiNeMaS, vierge, hébergée sur un serveur situé à l'INRA du Moulon, accessible depuis Orléans, grâce à un compte administrateur. Malgré une coupure d'une journée à cause d'un incident technique au Moulon, l'accès à distance a parfaitement fonctionné.

Pour bien apprendre à me servir de la base de données, j'ai réalisé les tutoriels *Maïs* puis *Pairies Fourragères* envoyés par Monsieur De Oliveira, qui se rapprochait plus de ce que l'on souhaitait faire.

Le déroulement type du tutoriel est (en abrégé) :

- Avant insertion des données, rentrer les données nécessaires dans l'interface administrateur, à savoir le nom des espèces, le nom du projet en cours, le nom de la personne en charge, le nom des personnes recevant les lots de graines diffus, le type de la variété, le type de fécondation.
- Ensuite, rentrer les méthodes, via l'interface administrateur, dans la section *env_pra_phe_method* (Cf annexe 3, p33).

- Insérer le fichier de données (croisement entre variétés, reproduction, sélection, fichiers de données individuelles, mélange de lots de graines, diffusions, etc…) ainsi que le fichier méthode correspondant (*Figure 4*, p11).

A la suite de ces tutoriels, j'ai donc découvert qu'avant de réaliser l'insertion des fichiers, il fallait réaliser plusieurs ajouts manuels via l'interface administrateur.

3) Analyse

3.1) Des fichiers de données

Il a fallu dans un premier temps, que je comprenne et assimile tous les fichiers de données qui allaient être à rentrer en base de données. La sortie à Chéu m'avait déjà permis de comprendre les fichiers qui étaient envoyés aux pépiniéristes, c'est-à-dire à quoi correspondait telle donnée dans telle ligne/colonne, et de même pour le plan.

Ensuite, notre but étant de faire le lien entre les fichiers d'entrée de SHiNeMaS et les fichiers de données que nous voulions insérer, il fallait que l'on prépare nos fichiers. Nous voulions refaire « au propre » nos fichiers de telle sorte qu'ils aient une base commune et fiable. Pour cela, nous avons analysé les en-têtes déjà présents pour voir lesquels seraient à garder, lesquels seraient à détailler, et lesquels seraient à retirer, par rapport à ces besoins. Nous avons donc produit une nouvelle liste d'en-têtes plus détaillés. Puis j'ai réalisé un tableau pour récapituler quelles en-têtes iraient dans quels fichiers (*Figure 6*, p17).

Figure 6 - Aperçu du fichier récapitulatif

Anciens en-têtes	Nouveaux en-têtes	En-têtes formalisés	Exemple			
				02a_reproduction_dactyle_2006	Compilation sél pép	fich sé
	Nom verger	nom_verger		project		
Année	Sélection automne année	annee_selection		sown_year	x	
Pépinière	Pépinière	pepiniere			x	
Espèce	Espèce	espece	Aulne glutineux		x	
	Nom latin	nom_latin				
Origine	Origine	origine	AGL130		x	
	Planches	planches				
	Age pré-sélection	age_preselection				
	Pré-sélectionneur1	preselectionneur1				
	Pré-sélectionneur2	preselectionneur2				
	Nb de pré-sélectionnés	nb_preselectionnes				
Age	Age sélection	age_selection	1		x	
	Sélectionneur1	selectionneur1				
	Sélectionneur2	selectionneur2				
Choisis	Nb de choisis	nb_choisis	12		x	
	Nb de témoins	nb_temoins				
Fournisseur-n°	Fournisseur-n°	fournisseur_num	Vilmorin-K71661		x	
N° cert	N° cert	num_cert	F11-04R008		x	
Destination	Destination initiale	destination_initiale	AGL-SULLY		x	
	Destination finale	destination_finale	SULLY			
Catégories plage témoins	Limites catégories vendables	categories_vendables	15.30.50.80		x	
	Catégorie témoin	categorie_temoin	30.50			
vendables	Nb plants vendables	nb_plants_vendables	18 000		x	
Taux réel	Taux sélection réel	taux_selection_reel	1 500		x	
Taux théo	Taux sélection visé	taux_selection_vise	2 000		x	
Commentaires-précisions	Commentaires-précisions	commentaires	Préchac		x	
Code peuplement	Code peuplement	code_peuplement	source identifiée		x	
nom du peuplement ou	Nom du peuplement ou commune	nom_peuplement	Lot ONF la Joux = 04.0092.A - FD de		x	

3.2) Des fichiers d'entrée de SHiNeMaS

Comme énoncé précédemment, les fichiers d'entrée de SHiNeMaS sont de deux types : les fichiers de données ainsi que les fichiers méthodes (*Figure 1*, p10).

Dans un premier temps, il a fallu comprendre comment étaient structurés les fichiers d'entrée étant donné que le but était d'adapter nos fichiers de données en fonction de ces fichiers d'entrée. Après avoir refait les tutoriels, je me suis donc intéressé aux fichiers d'entrée.

Un premier problème s'est présenté : la lecture des fichiers. Etant enregistrés au format « .csv », les fichiers étaient quasiment illisibles dans Excel (*Figure 7*, p17) :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	projectsown_yearharvested_yearid_seed_lot_sownintra_selection_nameetiquettesplitquantity_sownquantity_harvestedblockXYdensite_semisdensite_semis$datesemissemis$date													
2	DEMO20112012C1_JFB_2011_00011701A12002012-11-01volÃ©e2012-11-01													
3	DEMO20112012C2_JFB_2011_00011711A22002012-11-01volÃ©e2012-11-01													
4	DEMO20112012C3_JFB_2011_00011721A32002012-11-01volÃ©e2012-11-01													
5	DEMO20112012C4_JFB_2011_00011701A42002012-11-01volÃ©e2012-11-01													
6	DEMO20112012C5_JFB_2011_00011691B12002012-11-01volÃ©e2012-11-01													
7	DEMO20112012C6_JFB_2011_00011681B22002012-11-01volÃ©e2012-11-01													
8	DEMO20112012C7_JFB_20111681B32002012-11-01volÃ©e2012-11-01													
9														
10														
11														

Figure 7 - Exemple fichier « 02-a-reproduction_2011-2012 » format .csv

Le format Comma-Separated Values (CSV) est un format informatique ouvert représentant des données tabulaires sous forme de valeurs séparées pas des virgules, des points–virgules ou encore des tabulations. Ici, la tabulation est utilisée comme séparateur de données.

Toutes les données sont donc stockées dans une seule colonne, et sont séparées par des tabulations, rendant l'ensemble illisible. Odile m'a dépanné en m'expliquant comment convertir ces données (c'est-à-dire mettre dans chaque colonne un type de données) à l'aide d'Excel. Il suffit simplement de sélectionner la colonne souhaitée, aller dans l'onglet *Données*, puis cliquer sur *Convertir* et choisir comme séparateur *Tabulation*.

On obtient alors un agencement des données beaucoup plus clair, et facilement interprétable (*Figure 8*, p18):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	project	sown_year	harvested_year	id_seed_lot_sown	intra_selection_name	etiquette	split	quantity_sown	quantity_harvested	block	X	Y	densite_semis	densite_semis$date	semis	semis$date
2	DEMO	2011	2012	C1_JFB_2011_0001			1	70		1	A	1	200	01/11/2012	volÃ©	01/11/2012
3	DEMO	2011	2012	C2_JFB_2011_0001			1	71		1	A	2	200	01/11/2012	volÃ©	01/11/2012
4	DEMO	2011	2012	C3_JFB_2011_0001			1	72		1	A	3	200	01/11/2012	volÃ©	01/11/2012
5	DEMO	2011	2012	C4_JFB_2011_0001			1	70		1	A	4	200	01/11/2012	volÃ©	01/11/2012
6	DEMO	2011	2012	C5_JFB_2011_0001			1	69		1	B	1	200	01/11/2012	volÃ©	01/11/2012
7	DEMO	2011	2012	C6_JFB_2011_0001			1	68		1	B	2	200	01/11/2012	volÃ©	01/11/2012
8	DEMO	2011	2012	C7_JFB_2011			1	68		1	B	3	200	01/11/2012	volÃ©	01/11/2012
9																

Figure 8 - Exemple fichier « 02-a-reproduction_2011-2012 » avec données converties

Ensuite, second problème : à quoi servent les fichiers *méthodes*, qui sont chacun liés à un fichier de données ? (par exemple : le fichier *02-a-reproduction_2011-2012* a un fichier méthode correspondant : *02-a-reproduction_method_2011-2012*, etc…). Monsieur De Oliveira avait expliqué rapidement leur utilité lors de sa démonstration, mais sans expliquer exactement à quoi ils correspondaient.

Après comparaison du fichier méthode (*02-a-reproduction_method_2011-2012)* avec le fichier correspondant (*02-a-reproduction_method_2011-2012)*, j'ai compris plusieurs choses :

Après avoir regardé chacun des fichiers par type (de *reproduction* par exemple), il se trouve que parmi les en-têtes présents, il y en a qui se retrouvent dans chacun des fichiers, et d'autres non. J'en ai déduit qu'il y avait des en-têtes obligatoires. C'est-à-dire que sur les 16 en-têtes présents dans un fichier de type *reproduction*, il y en avait 12 qui étaient récurrents dans les autres fichiers du même type. Pour les fichiers de *reproduction*, les en-têtes obligatoires dans l'ordre

- project
- sown_year (=année de semis)
- harvested_year (=année de récolte)
- id_seed_lot_sown
- intra_selection_name
- etiquette
- split
- quantity_sown
- quantity_harvested
- block
- x
- y

Parmi les en-têtes sont obligatoires, seulement les colonnes *project* et *id_seed_lot_sown* sont obligés d'être remplies par des données pour être acceptées par la base de données. Pour les autres en-têtes, seule leur présence est obligatoire.

Il est ensuite possible de rajouter des en-têtes supplémentaires, définis en fonction des besoins :

- densite_semis
- densite_semis$date (où le symbole $ est utilisé lorsque deux variables portant le même nom ne font pas appel à la même méthode de mesure)
- semis
- semis$date
- …

Dans un second temps, après avoir fait cette distinction, il fut plus facile de comprendre le fichier *méthode*. En fait, il regroupe les en-têtes non obligatoires rajoutés dans le fichier des données.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	project	sown_year	harvested_year	id_seed_lot_sown	intra_selection_name	etiquette	split	quantity_sown	quantity_harvested	block	X	Y	densite_semis	densite_semis$date	semis	semis$date
2	DEMO	2011	2012	C1_JFB_2011_0001			1	70		1	A	1	200	01/11/2012	volÃ©	01/11/2012
3	DEMO	2011	2012	C2_JFB_2011_0001			1	71		1	A	2	200	01/11/2012	volÃ©	01/11/2012
4	DEMO	2011	2012	C3_JFB_2011_0001			1	72		1	A	3	200	01/11/2012	volÃ©	01/11/2012
5	DEMO	2011	2012	C4_JFB_2011_0001			1	70		1	A	4	200	01/11/2012	volÃ©	01/11/2012
6	DEMO	2011	2012	C5_JFB_2011_0001			1	69		1	B	1	200	01/11/2012	volÃ©	01/11/2012
7	DEMO	2011	2012	C6_JFB_2011_0001			1	68		1	B	2	200	01/11/2012	volÃ©	01/11/2012
8	DEMO	2011	2012	C7_JFB_2011			1	68		1	B	3	200	01/11/2012	volÃ©	01/11/2012
9																

	A	B	C	D	E
1	variable	type	method_name	correlation_group	
2	densite_semis	T1	densite_semis		
3	semis	T1	type_semis		
4					

Figure 9 - En haut, fichier de données. En bas, fichier méthodes.

Le fichier de méthodes permet à la base de données de savoir à quoi correspondent les en-têtes rajoutés en plus de ceux obligatoires dans le fichier de données (*Figure 9*, p20). (Il ne faut pas oublier que les en-têtes rajoutées doivent être entrées manuellement dans la section *env_pra_phe_method* de l'interface administrateur pour être reconnues). Il est alors possible de déclarer dans le fichier de données un en-tête plus explicite pour les néophytes – par exemple *ensemencement* au lieu de *semis* -, et ensuite mettre dans la colonne variable du fichier méthode *ensemencement* et dans la colonne method_name, *semis*

Pour l'exemple, la cellule bleue correspond au nom de la variable dans le fichier de données, et la cellule verte correspond au nom de la variable rentrée dans la base de données (*Figure 10*, p20).

	A	B	C	D	E
1	variable	type	method_name	correlation_group	
2	ensemencement	T1	semis		
3					
4					

Figure 10 – Illustration pour les fichiers méthodes

Les variables contenant l'attribut *$date* n'ont pas être rajoutés dans le fichier méthode.

Pour résumer, le fichier méthode fait le lien entre le fichier de données, et la base de données.

3.3) Lien entre les fichiers d'entrée et les fichiers de données

Ensuite, il a fallu faire la liaison entre les fichiers que l'on souhaitait entrer dans la base de données, et les fichiers d'entrée. D'un côté, nous avons tous les en-têtes que nous avons retravaillés (Cf 3.1) et de l'autre, nous avons les en-têtes obligatoires pour la base de données (Cf 3.2). Pour éviter d'avoir trop de colonnes vides, j'ai commencé à chercher les relations entre les en-têtes des deux fichiers. Après analyse et discussion avec ma tutrice, nous avons décidé que six colonnes pouvaient correspondre :

- nom_verger = project
- annee_selection = sown_year
- nom_lot = id_seed_lot_sown
- ligne = x
- colonne = y
- placeau = block

On obtenait alors un fichier avec les 12 en-têtes obligatoires, et 29 en-têtes supplémentaires.

Après avoir une discussion avec Frédérique, nous avons décidé d'utiliser le maximum d'en-têtes parmi ceux qui étaient obligatoires. Nous avons donc choisi d'utiliser *intra_selection_name* à la place de *nature*, et *etiquette* au lieu de *nom_arbre*.

Il ne reste plus que trois en-têtes obligatoires vides : *quantity_sown* et *quantity_harvested*, que nous n'avons pas remplis, et *split* qui va être utilisé ultérieurement.

3.4) Fichier de mesure

Par la suite, nous voulions aussi stocker dans la base de données les mesures, comme celle réalisées à Chéu ou à Sully-la-chapelle par exemple. Pour chacune des mesures, il est nécessaire de remplir trois colonnes dont les en-têtes seront : la mesure en elle-même, la date de la mesure et les commentaires/remarques correspondants. Par exemple, une mesure de hauteur initiale correspond à la hauteur (h0) suivie de la date de la mesure et des remarques, de même pour les circonférences à hauteur de poitrine à 5 ans (c5), suivie des mêmes attributs. Nous avons donc choisi ces deux mesures et créé les en-têtes suivants :

- h0
- h0$date

- h0$comAGL-CHEU (c'est-à-dire h0$commentairesNOMDUVERGER)
- c5
- c5$date
- c5$comAGL-CHEU

Dans les fichiers types des fichiers d'entrée (*reproduction*, *selection*, …), il y avait un type qui nous paraissait correspondre potentiellement à notre besoin : le fichier dit de *données individuelles*.

J'ai donc commencé à analyser les en-têtes de ce fichier pour réaliser la même démarche que précédemment. Suite à une discussion avec Monsieur De Oliveira, il m'a conseillé de plutôt rester sur le fichier de type *reproduction* sur lequel nous avions déjà préparé tous les en-têtes pour les données. Nous avons finalement décidé de suivre son conseil et de rajouter les mesures dans le fichier *reproduction* que nous avons déjà fait, car la plupart des en-têtes étaient similaires, et que l'on en rajoutait seulement six.

Nous obtenions alors un fichier avec quarante-cinq en-têtes.

3.5) Gestion de la sélection

Parmi les besoins exprimés, Frédérique m'a aussi demandé de gérer la sélection des plants en fonction de leur nature (« choisi », « témoin », « composant »). Pour cela, j'ai demandé conseil à Yannick qu'en à leur utilisation de la colonne *split*. Il m'a alors expliqué que la colonne *split* correspondait à une donnée binaire : 0 indique qu'il n'y a pas eu de split (de scission, de partage), donc que la totalité du lot de graine a été utilisée, et donc 1 indique qu'une partie seulement du lot de graine a été utilisée. Pour correspondre à nos besoins, nous avons donc adapté cette colonne en décidant que 1 correspondrait aux « composants » (Cf *I-4.1*), et 0 aux « choisis » et « témoins » de la colonne *intra_selection_name*.

4) Création

Maintenant que j'ai bien compris la structure des fichiers d'insertion, il est nécessaire de tester la base de données SHiNeMaS pour voir si elle correspond bien aux besoins exprimés par les biologistes.

4.1) Création des fichiers

Il a fallu commencer par remplir des fichiers pour les soumettre à la base de données pour réaliser des tests. Nous avons décidé de réaliser le test à partir d'informations réelles et simulées du verger de Chéu.

J'ai donc commencé à insérer les données des années que nous avions déjà, c'est-à-dire les données correspondant aux en-têtes suivants :

- le projet (*project*)
- l'année (*sown_year*)
- l'id (*id_seed_lot_sown*)
- le bloc (*block*)
- les coordonnées x,y ainsi que la place dans le placeau
- la pépinière
- l'origine
- l'âge de la sélection et le sélectionneur
- les données plus détaillées (nb_choisis, nb_témoins, fournisseur_num, ….)
- les mesures réalisées à Chéu sur les plants de 2013/2014

Une première insertion de ces données a fait apparaître une erreur au niveau de la colonne *id_seed_lot_sown*. Nous avions décidé d'entrer les données dans cette colonne suivant ce formalisme : VERGER_ANNEE, or, la base de données attend un formalisme spécial : VERGER_NOMREPONSABLE_ANNEE. Il a donc fallu convertir toute la colonne *id_seed_lot_sown* dans ce format.

La colonne *etiquette*, anciennement *nom arbre*, qui permet de décrire un individu succinctement, résulte de la concaténation (en utilisant la fonction CONCATENER d'Excel) de l'origine, de la pépinière, de l'année, ainsi que de la ligne, du bloc, et de ses coordonnées dans le placeau.

Ensuite il a fallu que je gère la colonne *split*, c'est-à-dire qu'il fallait que parmi chaque lot de quatre « choisis », celui qui avait la plus grande circonférence (c5), c'est-à-dire le « composant » prenne la valeur de 1. Tous les autres prennent la valeur 0.

Pour cela, j'ai découvert la fonction *SI* dans Excel. Elle permet de réaliser une simple condition *IF* comme dans la plupart des langages de programmation.

Je l'ai donc utilisée dans la fonction suivante, dont voici une manière simplifiée :

Si la circonférence sélectionnée parmi les quatre *choisis* est la plus grande :

    Alors la valeur de split devient 1

Sinon

    La valeur de split reste 0

La fonction Excel correspondante est :

SI(A1=MAX(B1:B4) ; «1» ; « 0 »)

4.2) Simulation pour les années suivantes

Pour pouvoir réaliser un test réaliste, nous avons décidé de simuler les années suivantes. Dans un premier temps, nous avons décidé que *l'id* prendrait la forme suivante : AGLXXX_FS_2015, car nous ne connaissions pas encore le nom du verger, et FS correspond à Frédérique Santi.

Dans un second temps, il a fallu générer les mesures de hauteur à la plantation (h0) ainsi que les circonférences à hauteur de poitrine à 5 ans (c5) pour l'année 2015. Nous n'avions aucune mesure concernant la circonférence, Frédérique m'a donc suggéré des mesures de circonférence pour pouvoir simuler les années suivantes.

Pour réaliser cela, j'ai utilisé une fonction mathématique conseillée par Frédérique, la loi normale inverse. J'ai donc dans un premier temps repris toutes les mesures à part dans un fichier annexe pour réaliser les calculs. Grâce aux fonctions Excel, j'ai calculé la *moyenne* ainsi que *l'écart-type* de la hauteur dans les données réelles et la même chose pour la circonférence. La loi normale inverse a permis de simuler les données de hauteur à la plantation pour les plantations futures 2015 et de circonférence à 5 ans :

LOI.NORMALE.INVERSE(ALEA() ;*moyenne* ; *ecart-type*)

Par exemple, la moyenne de toutes les hauteurs est de « 142.18 » , et l'écart-type est de « 10,5 ». Une application de la loi normale inverse nous permet d'obtenir des valeurs simulées probables :

LOI.NORMALE.INVERSE(ALEA() ;142.18 ;10.5) = 152.5 , 137.9 , 147.5 , 151.9 , etc.

4.3) Enregistrement

Une fois le fichier fini, il fallait l'enregistrer au format demandé pour que la base de données l'accepte, c'est-à-dire au format « .csv », avec pour séparateur la tabulation. Or, Excel permet seulement d'exporter les fichiers au format « .csv » avec comme séparateur le point-virgule. Par contre, via Open Office (suite bureautique libre de droits), utilisé à l'INRA du Moulon, il est possible d'éditer les filtres lors de l'enregistrement et de choisir la tabulation comme séparateur.

Dans un premier temps, nous avions enregistré avec Excel au format texte « .txt » avec séparateur tabulation. La base de données acceptait ce format au début, puis

lorsque j'ai essayé avec le fichier complet, j'ai eu l'erreur décrite dans la partie suivante (5).

Apparemment, il est impossible d'enregistrer avec tabulation pour séparateur avec cette version d'Excel. Il faudrait donc installer une version ultérieure d'Excel, ou la suite d'Open Office pour pouvoir enregistrer au format voulu.

Ayant Open Office sur mon ordinateur personnel, j'ai donc pu procéder aux enregistrements dans le format correct pour ensuite faire les insertions.

5) Insertion

Lors de l'insertion, différentes erreurs se sont produites. Dans un premier temps, comme dit précédemment, il y avait une erreur de formalisme pour la colonne *id_seed_lot_sown* qui a été corrigée.

Ensuite, lors des premières insertions de test, avec seulement quelques données, j'ai eu quelques erreurs dues à une inattention de ma part, telles que des oublis d'en-têtes dans le fichier méthode, qui m'ont été signalés par la base de données (*Figure 11*, p25).

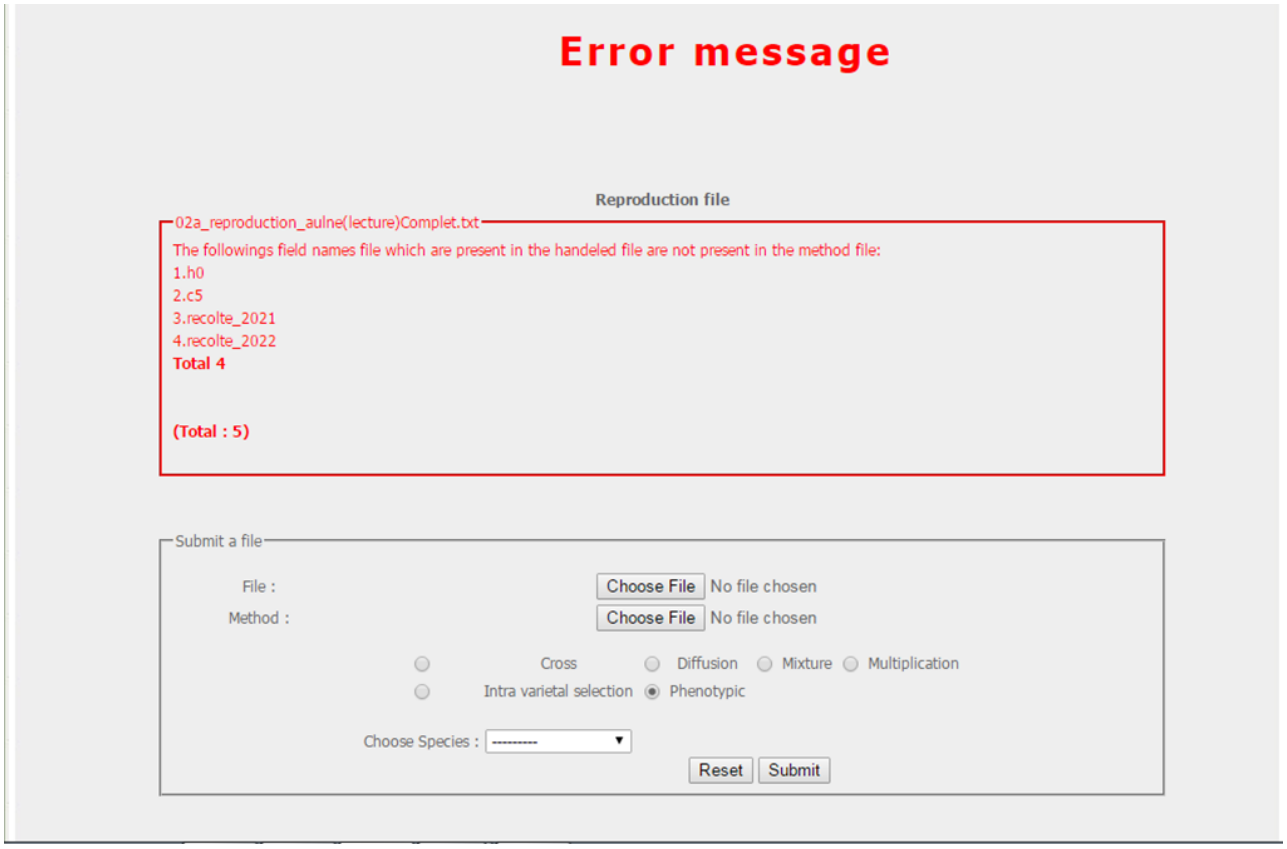


Figure 11 - Exemple de message d'erreur dû au fichier méthode

Dans un second temps, lors de l'insertion du premier jeu de données, une erreur Django s'est produite : « invalid litteral for int() with base 10 : '' », alors que j'enregistrais toujours les fichiers au format « .txt » avec séparateur tabulation.

Après avoir cherché de mon côté pour voir si l'erreur ne provenait pas de mes fichiers (sans prendre en compte l'enregistrement), j'ai contacté Monsieur De Oliveira pour réfléchir à une solution. Suite à des recherches de son côté (sur l'application Django), il s'avère que l'erreur provient d'un bug de la base de données. Néanmoins, une autre erreur de formalisme s'était glissée dans mon fichier de données, au niveau de la date, qui doit être au format « YYYY-MM-DD » (exemple : 2014-04-23), au lieu du format que l'on utilisait avant : « DD-MM-YYYY ».

Pour contourner cette erreur, j'ai donc eu à enregistrer les fichiers au format « .csv » via Open Office.

Au tout début, j'ai commencé à faire des insertions sur des données quelconques. Puis, j'ai réalisé le fichier de données complet (Cf. annexe 1, p31), avec les années connues ainsi que celles simulées pour faire un test « grandeur nature », en rajoutant les mesures, et en réalisant dans ce fichier tout ce qui est décrit dans la partie II.*3)*, c'est-à-dire l'adaptation des en-têtes, etc. Les différentes insertions ont été un succès. La base de données SHiNeMaS répond donc à nos attentes en l'adaptant au besoin.

6) Rédaction du tutoriel

Pour faciliter l'utilisation future de la base de données, il m'a donc été demandé de réaliser un tutoriel sur l'insertion des données dans la base via l'interface web.

J'ai pris pour modèle les différents tutoriels que m'avait fournis Yannick, que j'ai adapté pour le cas « aulne », espèce sur laquelle nous avions fait les fichiers de tests.

Pour cela, j'ai donc commencé à rédiger les différentes étapes à faire pour réaliser une insertion (Cf annexe 5, p35), c'est-à-dire les prérequis à insérer manuellement dans la base comme : le nom de l'espèce, le nom du projet en cours, le nom de la personne en charge, le type de la variété, etc.

La suite du déroulement du tutoriel va être d'expliquer comment bien préparer les fichiers de données et les fichiers méthodes, ainsi que réaliser l'insertion via l'interface web.

Ce tutoriel ainsi que les fichiers exemples associés pourront être réutilisés au besoin par les futurs utilisateurs qui veulent découvrir la base et la tester avec des données arbres.

III - Bilan

1) Technique

Durant ce stage, j'ai pu mettre en application mes connaissances acquises pendant ma formation, telles sur les bases de données par exemple. De plus, ce stage m'a permis de parfaire mes connaissances dans l'utilisation d'outils tels qu'Excel, ou encore le maniement d'interface codée avec Django. En outre, j'ai eu l'occasion de découvrir la rédaction d'un tutoriel, activité qui m'a beaucoup apporté notamment sur la reformulation pour des néophytes du travail informatique à effectuer.

2) Humain

Ce stage m'a appris beaucoup de choses sur le plan humain. Tout d'abord il m'a permis de découvrir le monde du travail, c'est-à-dire d'appréhender les différentes facettes du travail en informatique outre l'aspect technique.

J'ai donc dû m'insérer dans ce qui était pour moi un monde nouveau, autant celui du travail que celui de la recherche, et découvrir une nouvelle thématique qu'est la biologie. Grâce aux rencontres avec les différents acteurs de la sélection participative, et aux différentes sorties que j'ai réalisées, j'ai été à même de mieux comprendre et concevoir le monde dans lequel j'ai évolué. De plus, j'ai pu faire la connaissance, avec les autres membres travaillants à l'INRA, autant les stagiaires que les doctorants, qui ont volontiers partagé leurs expériences, ce qui m'a permis de découvrir d'autant plus l'univers dans lequel j'ai travaillé. En outre, j'ai eu l'occasion d'assister à plusieurs réunions d'équipe, ce qui m'a permis de découvrir l'organisation nécessaire pour la recherche, et a continué d'éveiller ma culture scientifique.

J'ai aussi pu mettre en pratique les différentes connaissances que nous avions travaillées dans le cadre du cours d'Expression et Communication durant ma formation, à savoir l'organisation et la communication, ou encore comment interagir avec une personne néophyte en informatique.

Sur un plan plus personnel, il m'a aussi permis de savoir me remettre en question quant au travail que je venais de réaliser selon les critiques qui m'étaient données.

IV- Conclusion et perspectives

Ces dix semaines de stage m'ont donc permis de découvrir le monde du travail et son fonctionnement, notamment dans un institut de recherche. Durant ce stage, j'ai pu mettre en pratique mes connaissances acquises lors de ma formation, dans le cadre du projet qui m'a été confié. Après les différents tests réussis sur la base de données, j'ai pu en conclure qu'elle répondait à nos besoins. Ayant été en communication avec Yannick tout au long du stage, j'ai pu lui remonter les différents problèmes et erreurs que j'ai rencontrés, lui permettant d'avoir un retour sur une utilisation différente de SHiNeMaS, et déceler différents bugs qu'il n'avait pas encore décelés.

Dans les dernières semaines de mon stage, j'ai encore le tutoriel à finir. De plus, ayant déjà réalisé l'insertion du fichier complet, et à la demande de Madame Santi, je vais devoir séparer ce fichier en sous-fichiers, pour simuler une insertion année par année.

En outre, après la sortie à la pépinière Claireau le 11 Juin, nous avons vu à quel point il était important de réaliser un plan clair et compréhensible par n'importe qui, et nous avons donc l'objectif de retravailler sur ce plan.

Cette première expérience dans le monde du travail m'a été très bénéfique et enrichissante, autant sur le point de vue technique qu'humain, et m'a conforté dans l'envie de poursuivre mes études dans l'informatique.

Définitions

[1] Sélection participative : Sélection variétale réalisée grâce à une importante implication des utilisateurs finaux dans l'évaluation, la réflexion et la décision des schémas et protocoles de sélection.

[2] Biomasse : Masse totale d'organismes vivants dans un biotope déterminé à un moment donné.

[3] Agro-écologie : Façon de concevoir des systèmes de production qui s'appuient sur les fonctionnalités offertes par les écosystèmes.

[4] Variété : Ensemble d'individus d'une espèce (ou hybrides entre deux ou plusieurs espèces) dont l'origine est définie pour leur commercialisation.

[5] Django : Framework (ensemble cohérent de composants logiciels) open-source de développement web en Python.

Bibliographie

-**Rapports de stage** :

- *Développement d'une base de données dédiée à la gestion dynamique locale de ressources variétales végétales* (2014), L. Burlot.

- *Intégration d'un package R et visualisation des résultats dans une interface web* (2014), M. Lefebvre.

- *Gestion de données liées à l'histoire de lots de graines* (2012), M. Darkawi.

- **Webographie** :

- Site présentant la sélection participative, *www.selection-participative-arbres.fr*

- Site de l'INRA d'Orléans, équipe UAGPF, *www6.val-de-loire.inra.fr/uragpf*

- Page présentant le projet SHiNeMaS :

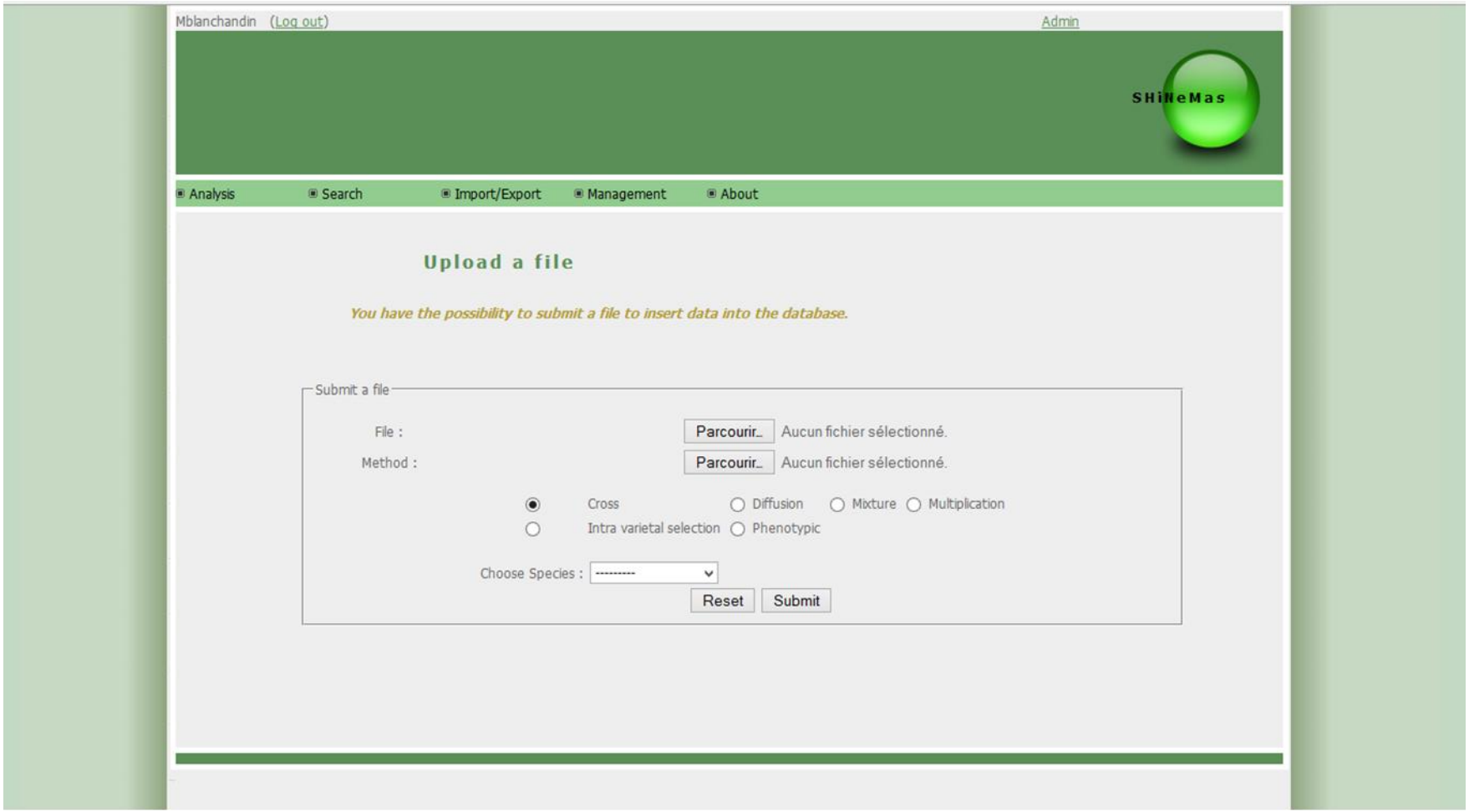
http://moulon.inra.fr/index.php/fr/equipestransversales/atelier-de-bioinformatique/projects/181

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
19	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_2.6.1				6	2	40	1	Wadel	Aulne glutineux		AGL130
20	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_2.6.3				6	2	42	3	Wadel	Aulne glutineux		AGL130
21	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_2.6.4				6	2	43	4	Wadel	Aulne glutineux		AGL130
22	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.1.1				1	3	0	1	Wadel	Aulne glutineux		AGL130
23	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.1.2				1	3	1	2	Wadel	Aulne glutineux		AGL130
24	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.1.4				1	3	3	4	Wadel	Aulne glutineux		AGL130
25	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.2.1				2	3	8	1	Wadel	Aulne glutineux		AGL130
26	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.2.2				2	3	9	2	Wadel	Aulne glutineux		AGL130
27	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.2.4				2	3	11	4	Wadel	Aulne glutineux		AGL130
28	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.3.1				3	3	16	1	Wadel	Aulne glutineux		AGL130
29	AGL-CHEU	2013	2020	AGL130_Wadel_2013	choisi	AGL130_Wadel_2013_3.3.2				3	3	17	2	Wadel	Aulne glutineux		AGL130
30	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_2.1.1				1	2	0	1	Wadel	Aulne glutineux		AGL130
31	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_2.2.2				2	2	9	2	Wadel	Aulne glutineux		AGL130
32	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_2.3.3				3	2	18	3	Wadel	Aulne glutineux		AGL130
33	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_2.4.2				4	2	25	2	Wadel	Aulne glutineux		AGL130
34	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_2.5.1				5	2	32	1	Wadel	Aulne glutineux		AGL130
35	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_2.6.2				6	2	41	2	Wadel	Aulne glutineux		AGL130
36	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_3.1.3				1	3	2	3	Wadel	Aulne glutineux		AGL130
37	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_3.2.3				2	3	10	3	Wadel	Aulne glutineux		AGL130
38	AGL-CHEU	2013		AGL130_Wadel_2013	composant	AGL130_Wadel_2013_3.3.3				3	3	18	3	Wadel	Aulne glutineux		AGL130
39	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.1.5				1	2	4	5	Wadel	Aulne glutineux		AGL130
40	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.1.6				1	2	5	6	Wadel	Aulne glutineux		AGL130
41	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.1.7				1	2	6	7	Wadel	Aulne glutineux		AGL130
42	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.1.8				1	2	7	8	Wadel	Aulne glutineux		AGL130
43	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.2.5				2	2	12	5	Wadel	Aulne glutineux		AGL130
44	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.2.6				2	2	13	6	Wadel	Aulne glutineux		AGL130
45	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.2.7				2	2	14	7	Wadel	Aulne glutineux		AGL130
46	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.2.8				2	2	15	8	Wadel	Aulne glutineux		AGL130
47	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.3.5				3	2	20	5	Wadel	Aulne glutineux		AGL130
48	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.3.6				3	2	21	6	Wadel	Aulne glutineux		AGL130
49	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.3.7				3	2	22	7	Wadel	Aulne glutineux		AGL130
50	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.3.8				3	2	23	8	Wadel	Aulne glutineux		AGL130
51	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.4.5				4	2	28	5	Wadel	Aulne glutineux		AGL130
52	AGL-CHEU	2013	2020	AGL130_Wadel_2013	témoin	AGL130_Wadel_2013_2.4.6				4	2	29	6	Wadel	Aulne glutineux		AGL130

02a_reproduction_aulne -

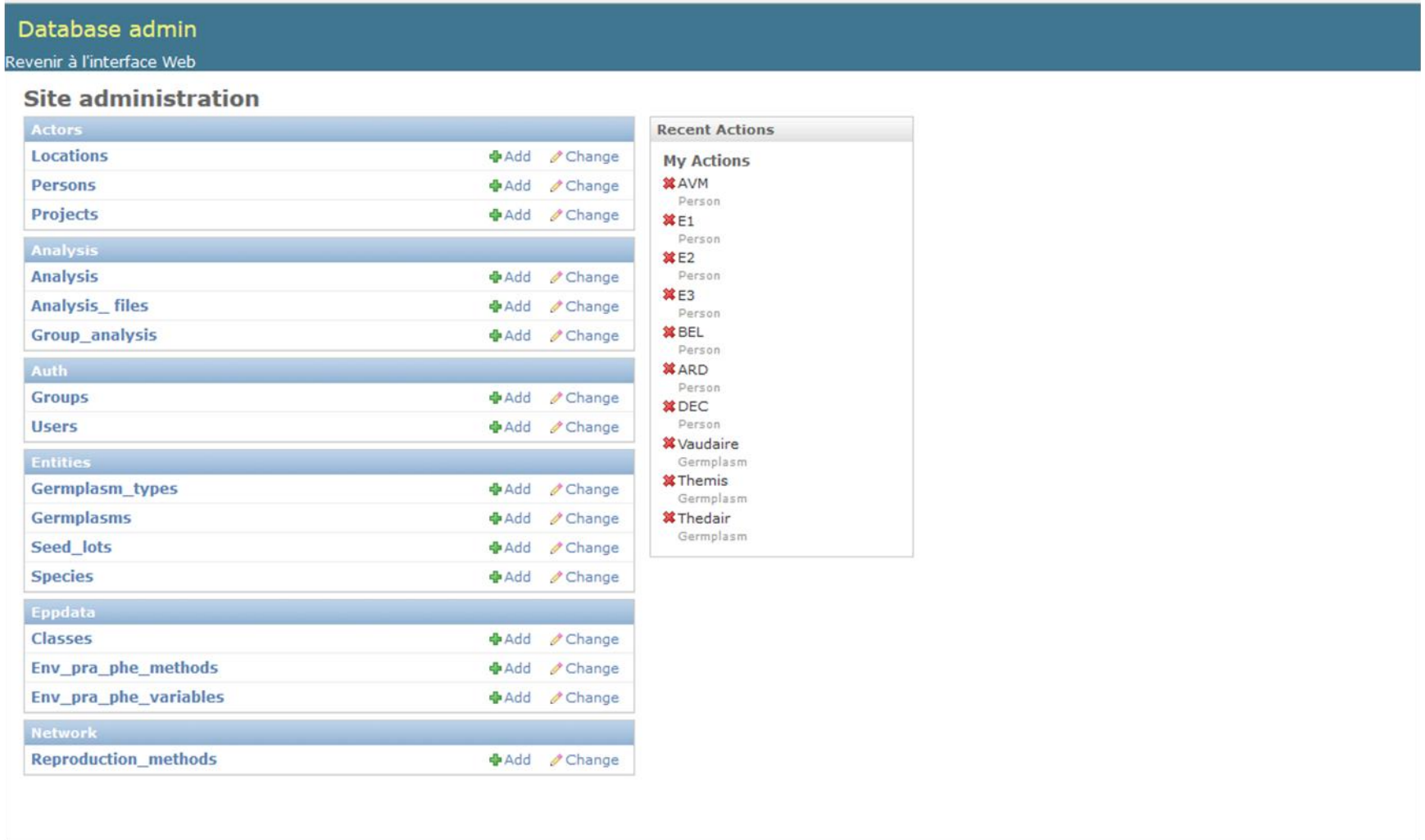
Annexe 1

Exemple Fichier de test



Annexe 2

Interface web de SHiNeMaS



Annexe 3

Interface administrateur de SHiNeMaS

1	**AGL-SULLY (AGL130-ouest)**															
2																
3		Distance (m)														
4	Ligne 8	49		I	I	I	I	I	I	I	I	I	I	I	I	
5	Ligne 7	42		I	I	I	I	I	I	I	I	I	I	I	I	
6	Ligne 6	35		I	I	I	I	I	I	I	I	I	I	I	I	
7	Ligne 5	28	charme	Bauchery2014-AGL130	Claireau2014-AGL130	Naudet-P2014-AGL130	Claireau2014-AGL130	Lemonnier2014-AGL130	Claireau2014-AGL130	Naudet-P2014-AGL130	Claireau2014-AGL130	Bauchery2014-AGL130	Claireau2014-AGL130	Naudet-P2014-AGL130	I	
8	Ligne 4	21		cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	
9	Ligne 3	14		cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	
10	Ligne 2	7		cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	
11	Ligne 1	0		cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	cormiers	
12																
13	Distance (m)			0	8	16	24	32	40	48	56	64	72	80	88	96
14	N°parcelle			Parcelle 1	Parcelle 2	Parcelle 3	Parcelle 4	Parcelle 5	Parcelle 6	Parcelle 7	Parcelle 8	Parcelle 9	Parcelle 10	Parcelle 11	Parcelle 12	
15	Place intra-parcelle			12345678	12345678	12345678	12345678	12345678	12345678	12345678	12345678	12345678	12345678	12345678	12345678	
16																
17	Composition de 10 parcelles Xxxxx2014-AGL130						choisi-choisi-choisi-choisi-témoin-témoin-témoin-témoin									
18	Composition de 1 parcelle Claireau2014-AGL130						choisi-choisi-choisi-charme-témoin-témoin-témoin-charme									
19	(ligne 5 parcelle 10)															
20																
21																

Annexe 4

Plan du verger à Sully-la-chapelle

TUTORIEL POUR L'UTILISATION DE LA BASE DE DONNÉES SHiNeMaS : cas Aulne.

Ce tutoriel va vous permettre de réaliser l'insertion de fichiers de données dans la base de données, de générer automatiquement ces fichiers et de réaliser des requêtes.

Important : Pour chaque étape, reportez-vous au guide utilisateur. L'endroit où vous reporter est indiqué sous la forme (GU-B.1.A.) qui signifie Guide Utilisateur : partie B.1.A.

Dans le cas des aulne, chaque espèce dispose de son propre fichier type.

1. Insertion de fichiers de données

Toutes les données sont recueillies dans le cadre du projet AulneCheu

Avant de démarrer il est nécessaire de rentrer certaines informations dans la base manuellement, en passant par l' interface administrateur, à savoir:

le nom des espèces :
- Aulne glutineux

le nom du projet en cours **(GU-B.1.A.)**:
- AulneCheu

le nom (shortname) de la personne en charge des semis et des croisements **(GU-B.2.b.)**:
- FS

le type de la variété **(GU-B.3.A.)**:
- Cultivar
- Peuplement
- Verger à graines
-

le type de fécondation **(GU-B.4.A.)**: =OK
- fécondation libre
- fécondation manuelle

1.1. Reproduction

Semis : 02-reproduction_espèce_année-de-semis.csv (GU-B.4.c.vii.)

Annexe 5

Extrait du tutoriel