



HAL
open science

Développement d'un système d'information de phénotypage d'*Arabidopsis thaliana*

Juliette Fabre, Anne Tireau, Pascal Neveu, Myriam Dauzat, Christine Granier

► **To cite this version:**

Juliette Fabre, Anne Tireau, Pascal Neveu, Myriam Dauzat, Christine Granier. Développement d'un système d'information de phénotypage d'*Arabidopsis thaliana*. Cahier des Techniques de l'INRA, 2008, 65, pp.31-46. hal-02662730

HAL Id: hal-02662730

<https://hal.inrae.fr/hal-02662730v1>

Submitted on 3 Sep 2024

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Développement d'un Système d'Information de phénotypage d'*Arabidopsis thaliana*

Juliette Fabre¹, Anne Tireau², Pascal Neveu², Myriam Dauzat¹, Christine Granier¹

Résumé : L'évolution de la démarche scientifique dans la plupart des champs d'investigation se traduit par une exigence de haut débit dont on ne peut tirer le meilleur profit sans un effort considérable d'archivage et de mise à disposition des données. Au laboratoire d'écophysiologie des plantes sous stress environnementaux (LEPSE) du centre Inra de Montpellier, le développement récent d'une plate-forme de phénotypage automatisée dédiée à la plante modèle *Arabidopsis thaliana*, la plate-forme PHENOPSIS, a permis d'augmenter de façon considérable les analyses phénotypiques effectuées sur cette espèce. Dans ce contexte, il est devenu nécessaire de développer une base de données associée à la plate-forme. Le travail présenté ici décrit les choix technologiques et le développement de la base de données et de l'interface Web qui établit le lien entre cette base de données et ses utilisateurs. Avec cet ensemble le travail d'insertion des données et de leur stockage ainsi que leur consultation et/ou téléchargement se fait de manière contrôlée.

Mots-clés : Système d'Information, Système de Gestion de Base de Données, Interface Web, Phénotypage Haut-Débit, *Arabidopsis thaliana*, MySQL, PHP, HTML, Architecture Client/Serveur

Introduction

L'agriculture du futur doit être plus économe en eau et en fertilisant tout en maintenant un objectif de production raisonnable. Les travaux du LEPSE visent à analyser et à modéliser les réponses des plantes aux stress abiotiques, en particulier au déficit hydrique du sol. La cible principale sur laquelle le LEPSE porte ses efforts est la croissance foliaire qui est affectée de manière importante et précoce au cours d'un stress. Sa diminution réduit les consommations en eau mais réduit également la production de biomasse. Dans ce contexte, l'objectif des expérimentations du LEPSE est de bâtir, d'évaluer et d'utiliser des modèles qui rendent compte des réponses de la croissance foliaire aux stress et de leur variabilité génétique.

Depuis plusieurs années, l'équipe SPIC (Stress environnementaux et Processus Intégrés du contrôle de la Croissance) du LEPSE a fait le choix d'investir lourdement sur la plante modèle *Arabidopsis thaliana*. Ce choix donne accès à des gammes génétiques larges (mutants, transformants, lignées recombinantes) pour tester des hypothèses de modélisation. Pour répondre à ces objectifs, le LEPSE s'est équipé d'une plate-forme de phénotypage, la plate-forme PHENOPSIS, qui est maintenant composée de 3 chambres de culture automatisées (un prototype construit en 2002 décrit dans Dauzat *et al.*, 2004 a été multiplié par 3 en 2007), permettant chacune de cultiver dans 504 pots des plantes d'*Arabidopsis thaliana* dans des conditions environnementales contrôlées : humidité, température de l'air et rayonnement. Dans chacune de ces chambres, un automate gère les pesées et l'irrigation de chaque pot, et

¹ Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux (LEPSE), UMR 759, Inra/Agro-M, 2 place Viala 34060 Montpellier, France. ✉ juliette.fabre@supagro.inra.fr, christine.granier@supagro.inra.fr

² Analyse des Systèmes et Biométrie (ASB), UMR 729, Inra/Agro-M, 2 place Viala 34060 Montpellier, France.

l'acquisition d'une image de la surface des pots. Au cours des expérimentations successives, les données micro-météorologiques sont enregistrées, ainsi que les données caractérisant l'état hydrique du sol et différentes variables sont collectées sur diverses parties de chaque plante, à différents stades de leur développement. Ces dernières variables sont des variables décrivant les stades phénologiques de la plante, des variables de croissance, des variables physiologiques, biochimiques et/ou moléculaires.

Avec l'utilisation de la plate-forme nous avons multiplié par 100 le nombre de génotypes d'*Arabidopsis thaliana* analysés au LEPSE entre 2002 et 2007.

L'objectif du travail présenté ici était d'une part, de développer une base de données regroupant des données de phénotypage de grandes collections de plantes, des données de micro-climat, des données génétiques et biochimiques relatives à ces collections pour procéder à des analyses à haut-débit et identifier les processus responsables du contrôle de la croissance en conditions de stress environnementaux et d'autre part, de développer une application autour de cette base de données permettant aux utilisateurs d'interagir avec elle et notamment d'assurer sa maintenance grâce à des outils automatisés et simples.

1. Contexte de la plate-forme PHENOPSIS et cahier des charges de la base de données

1.1 La plate-forme PHENOPSIS et ses expérimentations

La plate-forme PHENOPSIS est composée de 3 automates (PHENOPSIS 1, 2 et 3) qui permettent de peser, d'irriguer précisément et de prendre des photos des plantes d'*Arabidopsis thaliana* individuellement dans des conditions environnementales rigoureusement contrôlées. Ces 3 automates sont placés dans 3 chambres de culture distinctes.

1.1.a Les conditions micro-météorologiques

La régulation du climat dans les chambres de culture est réalisée par une acquisition de données CR10 (Campbell) contrôlant des actionneurs commandés par les consignes programmées. Les conditions micro-météorologiques (température de l'air, humidité de l'air et rayonnement PAR) sont mesurées en continu par différents capteurs, avec un pas de temps de 10s pendant toutes les expérimentations, et elles sont moyennées toutes les 10 minutes.

1.1.b Les pots et les plantes d'une expérimentation

L'automate est composé d'un cadre en acier supportant 14 plateaux composés chacun de 36 alvéoles pouvant contenir chacune 1 pot (**figure 1.a**). Chaque expérimentation contient donc au maximum 504 pots. Même si il peut y avoir plusieurs plantes par pot, dans chaque pot, un seul génotype identifié est semé. Le nombre de plantes de ce génotype par pot est noté.

1.1.c L'irrigation

Au démarrage d'une expérimentation, les poids des pots vides sont déterminés ainsi que le poids de substrat sec dans chaque pot. Ainsi par la suite une simple pesée nous fournit le taux d'humidité du substrat à chaque instant dans l'expérimentation, et permet d'ajuster le poids du pot avec une solution nutritive pour atteindre l'humidité cible prédéfinie par l'expérimentateur. Ces pesées et les ajustements de poids sont effectués automatiquement par l'automate à des heures programmées (**figure 1.b**). Dans PHENOPSIS 1 et 2, il y a une seule source de solution nutritive pour irriguer les plantes. PHENOPSIS 3 est équipé de deux sources de solutions nutritives et donc chaque pot peut être irrigué soit par la solution A, soit par la

solution B selon le choix programmé de l'expérimentateur.

1.1.d La prise d'images

Une caméra est embarquée sur le bras de chacun des automates pour obtenir des photos des pots individuellement. Sur le bras de l'automate de PHENOPSIS 3, une caméra infrarouge peut également prendre des images thermiques dans les mêmes conditions (**figures 1.c et 1.d**).

1.1.e Les mesures non destructives sur les plantes

Certaines mesures sont effectuées de façon régulière sur les plantes et sont non-destructives. Ceci comprend les mesures de surface des rosette ou surface de feuilles au cours du temps par analyses d'images sur les photographies des plante et les notations des stades phénologiques (décrits dans Boyes *et al.*, 2001) par l'expérimentateur soit à partir des photographies des plantes soit directement sur la plante, et les mesures de photosynthèse et/ou de conductance stomatique.

1.1.f Les mesures destructives sur les plantes

Certaines mesures nécessitent une récolte de la plante et sont donc destructives. Elles sont en général réalisées à un stade donné. Ce type de mesures comprend les poids secs et les poids frais de la rosette, de certaines feuilles, des pétioles ou du système racinaire, les mesures de surfaces et de nombre de cellules épidermiques sur les feuilles ou de densité stomatique, les mesures d'épaisseur de feuilles et de composition cellulaire des tissus internes sur des photos de coupes de feuilles, les surfaces foliaires éclatées mesurées sur des scans des feuilles individuelles, la longueur totale des racines mesurée sur des scans du système racinaire et les mesures biochimiques ou moléculaires (dosage d'une protéine, d'un transcrit, etc.).



Figure 1 : **a-** Automate dans PHENOPSIS 1, **b-** Pesée et irrigation d'un pot, **c-** Photo numérique prise dans le visible par l'automate, **d-** Photo prise dans l'infrarouge par l'automate dans PHENOPSIS 3

1.2 Les données produites par la plate-forme

Chaque jour, la plate-forme peut produire jusqu'à plus de 1500 photos (format JPG et d'une taille d'environ 25 ko pour PHENOPSIS 1 et 450 ko pour PHENOPSIS 2 et 3), et plus de 500 photos dans l'infrarouge (format JPG et d'une taille d'environ 10 ko).

Les noms de ces images ainsi que les données de sortie des différents cycles d'irrigation (poids des pots avant et après l'irrigation, quantité d'eau déversée...) sont récupérées dans des fichiers au format texte (CSV).

La plate-forme produit également chaque jour plus de 1500 données météorologiques (4 variables enregistrées toutes les 10 minutes tous les jours) organisées dans des fichiers texte au format DAT.

1.3 Cahier des charges pour la mise en place de la base de données

Cette importante quantité d'informations générée par les chambres de culture, les automates et l'expérimentateur nécessitait un outil pour (i) stocker les données et les métadonnées permettant leur traçabilité, (ii) les consulter, (iii) les extraire et les exporter pour les analyser. L'utilisation de la base de données par des utilisateurs distants et non-francophones qui partagent des données issues de la plate-forme (partenariats dans des contrats de recherche) devait également être possible.

2. Choix et mise en place du système d'information

La première étape a consisté à choisir un système de gestion de base de données (SGBD). Après examen des différentes possibilités, notre choix s'est rapidement porté sur le SGBD relationnel MySQL. MySQL est en effet un SGBD libre³, open-source⁴, portable⁵ et multi-utilisateurs⁶, et il est par ailleurs le SGBD le plus utilisé sur le campus de l'Inra de Montpellier. Ce SGBD utilise le langage SQL (Structured Query Language) qui est un langage de type requête standard et normalisé destiné à interroger et à manipuler une base de données relationnelle, et à en assurer la cohérence et la fiabilité grâce à des contraintes d'intégrité référentielle⁷. MySQL permet ainsi d'envoyer des requêtes SQL de création de tables, d'insertion, de modification, de sélection ou de suppression de données. La base de données PHENOPSIS a été créée avec MySQL 5.0.

Dans un deuxième temps, il s'agissait d'une part d'établir les modèles conceptuels et logiques des données, c'est-à-dire la structure des tables de la base de données, les contraintes d'intégrité et le type des données que ces tables devaient contenir et d'autre part de créer la base de données et d'importer les premières données par des requêtes SQL.

Enfin, la dernière étape a consisté à construire une interface entre cette base de données et les utilisateurs. La base de données devant être consultable à distance, une architecture Client/Serveur s'est imposée. Nous avons choisi pour l'interface Web d'utiliser PHP plutôt que Java, dans un souci de simplicité et afin que la maintenance de l'application puisse être évolutive et réalisée par des non-informaticiens. L'interface Web comprendrait ainsi des programmes alliant les langages HTML et PHP. Le langage HTML (HyperText Markup Language) est un format de données conçu pour la mise en forme d'informations par un navigateur Web. Le langage PHP (PHP Hypertext Preprocessor) est un langage de scripts généraliste et open-source, principalement utilisé pour traiter et produire des pages Web dynamiques. Il peut ainsi être intégré facilement au HTML. On utilise pour l'interface PHP 5.0.22.

³ Se dit d'un logiciel dont la licence dite libre donne à chacun le droit d'utiliser, d'étudier, de modifier, de dupliquer, de donner et de vendre ledit logiciel.

⁴ Se dit d'un logiciel dont la licence respecte des critères précisément établis par l'Open Source Initiative, c'est-à-dire la possibilité de libre redistribution, d'accès au code source et de travaux dérivés.

⁵ Se dit d'un logiciel pouvant être utilisé dans différents systèmes d'exploitation.

⁶ Qui autorise l'utilisation simultanée de ce logiciel par plusieurs utilisateurs.

⁷ Une intégrité référentielle garantit le respect des règles induites par les relations entre les tables d'une base de données.

2.1 La base de données PHENOPSIS

2.1.a La structure de la base de données

Les caractéristiques des chambres de culture sont regroupées dans une table *Chambre*, et celles des expérimentations dans une table *Manip*. Le champ *fichier* de cette table correspond au nom d'un fichier Excel contenant le protocole de l'expérimentation, le plan d'expérimentation et les données sur les poids des pots et le poids de sol sec qu'ils contiennent chacun au départ de l'expérimentation. Le champ *idGroupe* contient l'identifiant du groupe d'utilisateurs ayant les droits de consultation des données issues de l'expérimentation.

Chaque expérimentation comprend 504 pots maximum listés dans une table *PotManip*.

Chaque pot contient un certain nombre de plantes d'un même génotype. Les différents génotypes et leurs caractéristiques sont répertoriés dans une table *Genotype*. Le champ *genotypage* de cette table correspond au nom d'un fichier Excel contenant les caractéristiques de génotypage d'un génotype (dans le cas de populations de lignées recombinantes). Le champ *idGroupe* contient l'identifiant du groupe d'utilisateurs ayant les droits de consultation des données du génotype.

Toutes les variables stockées sont définies dans une table *Variable*. Elles sont de 3 types :

- des variables micro-météorologiques mesurées plusieurs fois par heure dans chacune des 3 chambres par des capteurs : table *MesureMeteo* ;
- des variables collectées automatiquement par les automates sur les pots : poids des pots avant arrosage et quantités d'eau apportée, et photos des pots prises dans le visible et/ou l'infrarouge : table *MesureRobot* ;
- des variables mesurées sur différentes parties de la plante à divers moments de son développement (stades phénologiques, nombre de feuilles, surface de la rosette, coupes de feuilles, etc.) : table *MesureOrgane*. Les parties étudiées de la plante sont définies par ailleurs dans une table *Organe*.

Le fait que toutes les variables soient regroupées dans une même table, plutôt que d'apparaître en tant que champs dans les tables *MesureMeteo*, *MesureRobot* et *MesureOrgane*, permet d'ajouter simplement de nouvelles variables, et d'en étudier certaines de manière ponctuelle sans stocker dans la base de données des colonnes contenant énormément de données manquantes. De plus, le champ *type* de la table *Variable* et les champs *valeurNum*, *valeurTexte* et *valeurFichier* des tables *MesureMeteo*, *MesureRobot* et *MesureOrgane* donnent la possibilité de déclarer et d'importer dans la base de données tous les types de variables, qu'elles soient numériques, texte ou sous forme de fichiers (photos, etc.). Cette structure offre donc une grande flexibilité et évolutivité dans l'étude des variables.

Tous les événements et les remarques (incidents météorologiques ou commentaires sur la croissance d'une plante par exemple) sont archivés dans une table *Commentaire*. Le champ *fichier* de cette table correspond au nom d'un optionnel fichier Excel, Word, texte ou image concernant le commentaire.

Enfin, les tables *Groupe*, *Utilisateur* et *GroupeUtil* permettent de gérer les utilisateurs ainsi que les groupes auxquels ils appartiennent et qui leur donnent ou non des droits de consultation sur les données des expérimentations ou des génotypes.

2.1.b Modèle physique des données

L'organisation des différentes tables les unes par rapport aux autres est représentée dans le modèle physique des données (**figure 2**). Ce modèle est simple, compréhensible et utilisable par un grand nombre d'utilisateurs, informaticiens ou non.

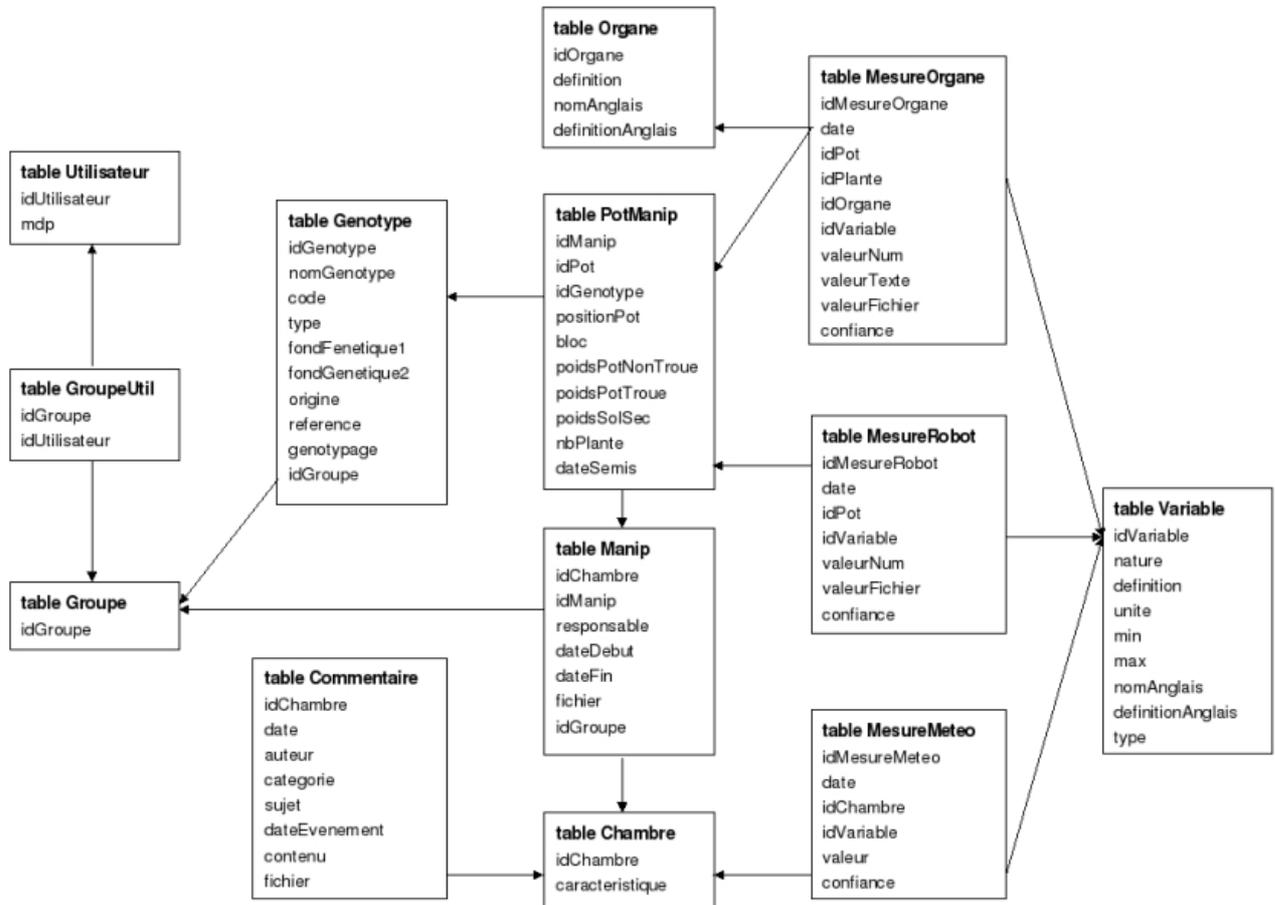


Figure 2 : *Modèle physique des données*

2.1.c Remarques sur le stockage des fichiers

Les fichiers liés aux données, tels que les fichiers Excel concernant les expérimentations, les fichiers Excel de génotypage, les fichiers de commentaire, les photos prises par les automates, les scans de plantes ou les coupes de feuilles produits par les expérimentateurs, sont stockés sur le serveur. Les noms des fichiers et un optionnel chemin sont stockés dans la base de données dans les champs appropriés. Chaque jour, des scripts shell⁸ vérifient l'intégrité des fichiers entre le serveur et la base de données.

Les fichiers auraient pu être stockés directement dans la base de données à l'aide du format BLOB (Binary Large Object). Cette structure aurait entre autres énormément ralenti les opérations sur la base de données, c'est pourquoi nous avons préféré stocker uniquement les noms des fichiers.

⁸ Le shell est un logiciel fournissant une interface pour un utilisateur.

2.2 L'application PHENOPSIS DB

L'interface Web PHENOPSIS DB est pour les utilisateurs un outil convivial, ergonomique, fonctionnel et simple pour interagir avec la base de données, pour consulter, pour extraire des données et pour importer des données dans la base de manière contrôlée en les validant ou non.

Cette interface facilite l'insertion et la consultation de données grâce à 2 méthodes : (i) en se connectant directement à la base de données par le biais du langage PHP et en envoyant des requêtes SQL de sélection et d'insertion de données, (ii) indirectement en appelant par PHP des scripts R de mise en forme de données, et de validation et d'insertion de données.

R est un langage et un environnement libre et open-source pour l'analyse statistique de données et leur représentation graphique. La librairie RMySQL permet de lier le logiciel R à MySQL et d'envoyer des requêtes de sélection et d'insertion de données. On utilise la version 2.4.1 de R, qui possède cette librairie. Il sera possible de migrer vers la librairie RODBC, qui offre plus de possibilités et est présente dans les versions récentes de R. Les scripts R hébergés sur le serveur sont appelés par les programmes PHP pour effectuer les traitements de données. R permet en effet d'une part de manipuler des données beaucoup plus aisément qu'avec PHP, et d'autre part de réaliser des analyses statistiques.

Une version anglaise de l'interface était nécessaire pour nos partenaires non-francophones. Plutôt que de dupliquer les scripts HTML/PHP afin d'obtenir une interface française et une interface anglaise, il a été choisi de créer 2 fichiers PHP contenant chacun l'ensemble des textes de l'interface dans l'une ou l'autre des 2 langues. Selon la langue choisie par l'utilisateur à sa connexion sur l'interface (cf §.3.1), les scripts PHP vont chercher dynamiquement les textes de chaque page Web dans l'un ou l'autre des 2 fichiers de texte. Pour créer une version de l'interface dans une autre langue, il suffit alors d'écrire la traduction des textes de l'interface dans cette langue dans un troisième fichier PHP.

2.3 Le Système d'Information

Le fonctionnement du Système d'Information est décrit sur la **figure 3**. Sur la plate-forme PHENOPSIS, les paramètres des automates et des conditions météorologiques des chambres sont contrôlés par l'ordinateur central. Les automates envoient leurs mesures dans des répertoires spécifiques sur l'ordinateur, de la même manière que les capteurs par le biais de la Campbell, le tout sous la forme de fichiers texte. Les mesures effectuées sur les plantes sont quant à elles saisies manuellement dans des fichiers Excel.

La base de données PHENOPSIS - c'est-à-dire les tables de données et les fichiers (photos et fichiers divers) - est hébergée sur le serveur intranet du LEPSE. Ce serveur héberge également l'interface Web PHENOPSIS DB (formulaires HTML, scripts PHP et R) qui sert de lien entre la base de données et les utilisateurs. Les fichiers hébergés sur le serveur sont en quantité très importante. En effet, la plate-forme PHENOPSIS peut produire entre autres quelques milliers de photos par jour (1512 pots pouvant être photographiés dans le visible et/ou l'infrarouge une ou plusieurs fois par jour !). La quantité de données produite chaque année par la plate-forme peut ainsi aller d'environ 10 Go à 150 Go.

L'administrateur de la base de données et les utilisateurs ayant des droits suffisants gèrent les utilisateurs et groupes d'utilisateurs et les modifications et les suppressions de la base de données ; l'administrateur seul s'occupe de la gestion de l'interface.

Les utilisateurs ayant les droits suffisants transforment les fichiers de données Excel ou texte en fichier CSV selon des règles précises. Ils insèrent ensuite dans la base ces fichiers CSV de données ainsi que les fichiers divers attachés aux données (fichiers Excel d'expérimentation,

de génotypage, fichiers de commentaires et scans de plantes) par le biais de l'interface Web. Les photos issues des chambres de culture ainsi que les autres photos de plantes prises par les expérimentateurs (coupes de feuille, etc.) constituent des dossiers très lourds (au moins 200 Mo), et ne sont donc pas téléchargées par le biais de l'interface mais transférées sur le serveur par les administrateurs via un protocole SFTP.

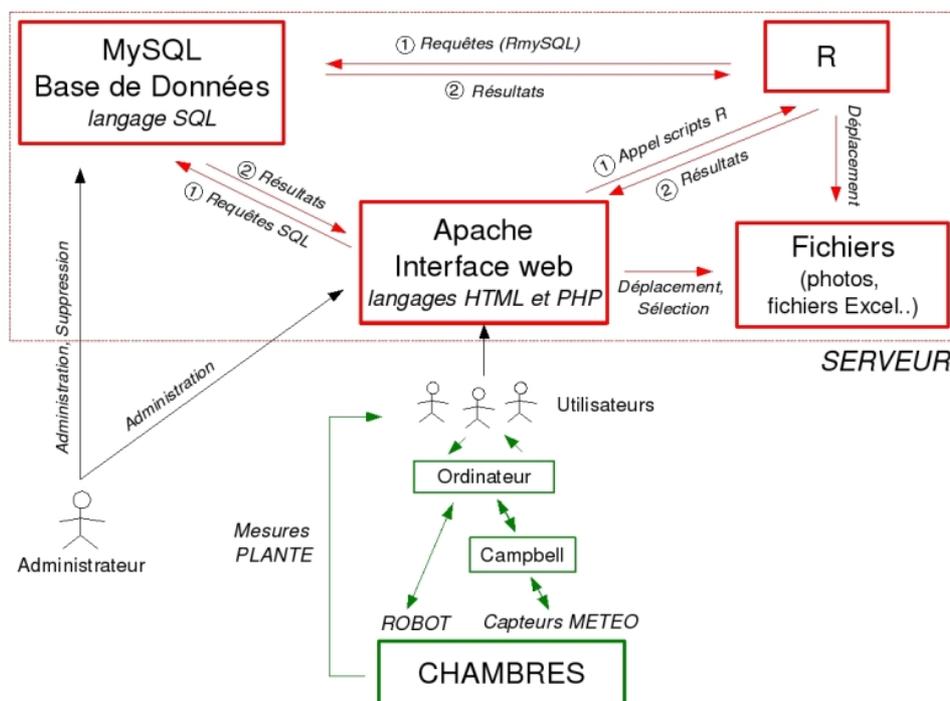


Figure 3 : Schéma du Système d'Information

3. Présentation de l'application PHENOPSIS DB

3.1 Gestion des utilisateurs et des sessions

Sur la page d'accueil de l'interface, l'utilisateur choisit la langue (Français ou Anglais). L'accès à l'interface s'effectue ensuite avec un nom d'utilisateur et un mot de passe fournis par l'administrateur (figure 4). Chaque utilisateur appartient à différents groupes, qui lui ouvrent différentes possibilités sur l'interface. L'insertion des données est ainsi réservée aux administrateurs et à certains groupes, tandis que la consultation des données est permise à tous les utilisateurs. Des restrictions seront établies sur certaines données pour les utilisateurs qui n'en sont pas propriétaires.

A chaque connexion, une session est créée, qui enregistre la langue choisie et les caractéristiques de l'utilisateur. Une fois que l'utilisateur est connecté, on lui propose 3 thèmes : (i) des informations sur le Système d'Information PHENOPSIS, (ii) la consultation et le téléchargement de données, (iii) et l'insertion de données et l'administration.

3.2 Description du Système d'Information PHENOPSIS

La partie *Le Système d'Information* PHENOPSIS décrit l'ensemble du système, des plateformes de phénotypage à l'interface en passant par la base de données.



Figure 4 : Page d'accueil : Connexion de l'utilisateur

3.3 Consultation des données

3.3.a Consultation de tableaux

L'utilisateur a la possibilité de soumettre des critères de sélection des données (**figure 5.a**). Ses critères sont soumis à validation, et s'ils sont corrects la requête SQL correspondante est affichée, avec le tableau de résultats, qui est téléchargeable sous forme de fichier CSV (**figure 5.b**).

Pour les données des chambres, des expérimentations, des pots, des génotypes, des variables, des organes et des commentaires, les données sont récupérées directement dans la base de données par les programmes PHP, puis affichées sous forme de tableaux, et enfin exportées dans des fichiers CSV de 60000 lignes maximum (plusieurs fichiers sont créés si nécessaire). Quand ils existent, les noms des fichiers Excel d'expérimentation ou de génotypage sont affichés en tant que lien. On peut alors cliquer sur ces liens et télécharger ces fichiers.

Pour les données météorologiques, issues des automates ou mesurées sur les plantes (tables *MesureMeteo*, *MesureRobot* et *MesureOrgane*), les programmes PHP extraient dans un premier temps les données demandées grâce à une requête SQL, et les stockent dans un fichier CSV. Ils appellent ensuite des scripts R prenant en argument ces fichiers CSV. Ces scripts transforment les tableaux de données de manière à ce que les variables étudiées apparaissent en colonnes, puis les exportent sous forme de fichier CSV. Les programmes PHP récupèrent alors ce fichier, en affichent les premières lignes (en général ces fichiers contiennent plusieurs milliers de lignes), et créent à partir du fichier autant de fichiers CSV de 60000 lignes maximum que nécessaire (**figure 6**).

Le nombre de données consultables et téléchargeables est limité à 1 million. Si le résultat de la requête contient plus d'un million de données, l'utilisateur est prié d'affiner ses critères de recherche.

Consultation et Téléchargement de Variables

Pour consulter et télécharger la liste des caractéristiques des variables étudiées

MENU ACCUEIL MENU CONSULTATION MENU SAISIE SE DÉCONNECTER

Les caractéristiques disponibles sont : l'identifiant de la variable, son nom, sa nature (robot, météo ou organe), sa définition, son unité, et ses valeurs minimum et maximum. [Plus de détails](#)

1 - Sélectionnez les variables

Les variables **météorologiques** sont mesurées dans les chambres de culture par des capteurs.

Les variables issues des **robots** concernent les données d'humidité du sol dans les pots et les photos prises par les robots dans les chambres de culture.

Les variables **organe** sont les variables mesurées par l'expérimentateur à différentes échelles de la plante : plante entière, rosette, feuille, pétiole, racine...

Variable

Toutes
Variables météorologiques
Variables issues des robots
Variables mesurées sur les plantes

Pour choisir plusieurs éléments dans une liste multiple, maintenez la touche CTRL enfoncée.

2 - Validation

Pour soumettre vos critères de sélection, appuyez sur

Réalisation du site web : Juliette Fabre

lepse

5.a

Consultation et Téléchargement de Variables

Pour consulter et télécharger la liste des caractéristiques des variables étudiées

MENU ACCUEIL MENU CONSULTATION MENU SAISIE SE DÉCONNECTER

Résultat de votre requête

Votre requête SQL est :

```
SELECT idVariable, nature, definition, unite, min, max FROM Variable WHERE nature="météo" ORDER BY nature, idVariable
```

Le nombre de variables correspondant à vos critères de sélection est : 4

[Exporter le tableau](#)

Variable	Nature	Définition	Unité	Min	Max
humidite	meteo	mesurée par la waisala au milieu du robot au niveau des plateaux	%	0	100
rayonnement	meteo	mesuré par un capteur Licor au milieu du robot au niveau des plateaux	µmol m ⁻² s ⁻¹	-10	500
temperature	meteo	mesurée par la waisala au milieu du robot au niveau des plateaux	°C	0	50
vpd	meteo		kPa	0	8

Réalisation du site web : Juliette Fabre

lepse

5.b

Figure 5 : Pages de consultation des variables :
a - saisie des critères de sélection des données, b- résultats

Téléchargement de Données Météorologiques

Pour télécharger des données météorologiques des chambres de culture

[MENU ACCUEIL](#) [MENU CONSULTATION](#) [MENU SAISIE](#) [SE DÉCONNECTER](#)

Résultat de votre requête

Votre requête SQL est :

```
SELECT DISTINCT MesureMeteo.idChambre, date, idVariable, valeur FROM MesureMeteo WHERE date=""2008-01-01 00:00:00"" ORDER BY MesureMeteo.idChambre, date
```

Le nombre de mesures météorologiques correspondant à vos critères de sélection est : 14460

Ceci correspond à 3615 lignes de mesures par date, et donc à 1 fichier(s) CSV de 60000 lignes à télécharger.

[Exporter le tableau 1](#)

Seules les 30 premières lignes sont affichées.

Chambre	Date	humidite	rayonnement	temperature	vpd
C1	2008-01-02 12:00:00	76.3	51.49	20.69	0.58
C1	2008-01-02 12:20:00	76.3	51.12	20.73	0.58
C1	2008-01-02 12:40:00	76.3	50.71	20.75	0.58
C1	2008-01-02 13:00:00	76.4	50.13	20.68	0.58
C1	2008-01-02 13:20:00	76.3	51.41	20.74	0.58
C1	2008-01-02 13:40:00	76.4	51.22	20.66	0.57
C1	2008-01-02 14:00:00	76.4	51.65	20.7	0.58
C1	2008-01-02 14:20:00	76.5	104.5	20.82	0.58
C1	2008-01-02 14:40:00	76.4	107.4	20.85	0.58

Réalisation du site web : [Juliette Fabre](#)

leperse

Figure 6 : Page de consultation de données météorologiques : résultats

3.3.b Download de fichiers

En ce qui concerne les photos issues des automates (table *MesureRobot*) et les scans de plantes ou les photos prises par les expérimentateurs (table *MesureOrgane*), 2 pages sont réservées à leur consultation et à leur téléchargement (download⁹) sur le poste de l'utilisateur. L'utilisateur peut, comme pour les autres données, soumettre des critères de sélection (**figure 7**).

Les programmes PHP extraient ensuite de la base de données, les noms des fichiers correspondants, archivent et compressent ces fichiers en autant de fichiers ZIP de 2000 fichiers maximum que nécessaire. L'utilisateur peut ensuite télécharger ces fichiers ZIP.

Le nombre de fichiers consultables et téléchargeables est limité à 40000. Si le résultat de la requête contient plus de 40000 fichiers, l'utilisateur est prié d'affiner ses critères de recherche.

⁹ Le download consiste en un téléchargement **depuis** un ordinateur distant (téléchargement descendant)



Figure 7 : Page de téléchargement de lots de photos : saisie des critères de sélection

3.3.c Sécurité des données

La sécurité des données est gérée au niveau de l'application. L'accès à la consultation des données est permise à tous les utilisateurs, mais des restrictions sont appliquées lors des requêtes SQL : l'utilisateur ne peut consulter et télécharger des données que s'il fait partie du groupe d'utilisateurs spécifié dans les expérimentations et génotypes concernés par les données.

3.4 Insertion des données

L'insertion des données est réservée à certains groupes d'utilisateurs et aux administrateurs. L'accès aux autres utilisateurs est refusé.

3.4.a Insertion d'un enregistrement à l'aide d'un formulaire

L'utilisateur peut insérer à l'aide de formulaires HTML des données enregistrement par enregistrement, c'est-à-dire n'insérer qu'une ligne à la fois dans une table de la base de données. Ceci est possible pour l'insertion d'une expérimentation, d'un génotype, d'un nouvel organe mesuré, d'une nouvelle variable étudiée ou d'un commentaire. Il peut également, pour une nouvelle expérimentation, un génotype ou un commentaire, télécharger (upload¹⁰) sur le serveur un fichier concernant ces données (fichier de protocole d'expérimentation, fichier de génotype, etc.) (**figure 8**).

Le script PHP de la page vérifie la validité des données et de l'éventuel fichier lié téléchargé. Si les données sont correctes, elles sont insérées dans la base de données et l'éventuel fichier est déplacé dans le répertoire approprié sur le serveur ; si les données ne sont pas validées, les erreurs sont affichées sur la page du formulaire.

¹⁰ L'upload consiste en un téléchargement vers un ordinateur distant (téléchargement montant)



1 - Remplissez tous les champs :

Tous les champs sont obligatoires!

A - Sélectionnez la **chambre** dans laquelle a eu lieu l'expérimentation [Plus de détails](#)

B - Entrez l'identifiant de l'**expérimentation** [Liste des expérimentations existantes](#)
 L'identifiant doit être de la forme *CiMj* avec *Ci* l'identifiant de la chambre et *j* le numéro de l'expérimentation.
 L'expérimentation n°4 de la chambre n°2 s'appellera ainsi C2M4.

C - Entrez le nom du **responsable** de l'expérimentation ou sélectionnez le dans la liste :

D - Entrez les **dates** de début et de fin d'expérimentation dans le robot, c'est-à-dire les dates d'entrée des premiers pots et des sortie des derniers pots.
 Date de début Date de fin
 Les dates doivent être de la forme JJ-MM-AAAA (ex 01-01-2008).



2 - Téléchargement d'un fichier Excel

Téléchargez si vous le souhaitez un **fichier Excel** contenant des caractéristiques de l'expérimentation (protocole, données d'humidité du sol dans les pots, ..).

Ce fichier sera renommé automatiquement avec le nom de l'expérimentation saisi!

Réalisation des pages : [Juliette Fabre](#)



Figure 8 : Page d'insertion d'une expérimentation



Insertion d'une liste de génotypes

ACCUEIL MENU CONSULTATION MENU SAISIE SE DÉCONNECTER



Téléchargement du fichier CSV

Téléchargez le fichier CSV contenant la liste des génotypes à insérer dans la base, et éventuellement un fichier Excel de génotypage (dans le cas de RILs).

Merci AU PREALABLE de lire le document décrivant le format approprié des fichiers et des données qu'ils contiennent.

[Lire le document](#)

Attention : Les génotypes figurant dans votre liste dont l'identifiant existe déjà dans la base de données ne seront pas insérés dans la base!

Fichier CSV de la liste des génotypes à insérer :

Fichier Excel de génotypage dans le cas de RILs :

Le nom du fichier Excel de génotypage doit **correspondre** au nom de fichier spécifié dans la colonne 'génotypage' du fichier CSV!

Soumettre le(s) fichier(s) :

Réalisation des pages : [Juliette Fabre](#)



Figure 9 : Page de téléchargement d'un fichier CSV de nouveaux génotypes

3.4.b Insertion d'une liste d'enregistrements à l'aide de fichiers CSV

L'utilisateur peut également insérer une liste d'enregistrements à l'aide de fichiers CSV (**figure 9**). Ceci concerne les génotypes, les organes et les variables (pour ces données, l'utilisateur a ainsi le choix d'utiliser un formulaire s'il a peu de données à insérer, ou un fichier CSV pour une plus grande quantité d'enregistrements), ainsi que les pots d'une expérimentation, les données météorologiques, les données issues des automates et les données mesurées sur les plantes. Pour les génotypes il est possible de télécharger parallèlement un fichier Excel de génotypage, et pour les données mesurées sur les plantes un fichier ZIP contenant des scans (cf 3.4.c).

Les formats des différents fichiers CSV sont décrits de manière très précise dans un document PDF téléchargeable depuis l'interface. L'utilisateur peut également télécharger des fichiers Excel type à remplir.

Une fois les fichiers téléchargés, le script PHP de la page appelle un script R prenant en paramètre le fichier CSV téléchargé et l'optionnel fichier lié. Ce script R est chargé de vérifier le format des données du fichier CSV. Si les données des fichiers ne sont pas valides, le script crée un rapport d'erreurs sous forme de fichier texte, téléchargeable par l'utilisateur.

Si les données des fichiers sont validées, alors il existe 2 situations. (i) Dans le cas de données de génotypes, d'organes, de variables ou de pots, les fichiers CSV correspondent exactement au format des tables correspondantes de la base de données. Le script PHP effectue donc simplement une requête SQL d'insertion de fichier dans la table appropriée. (ii) Dans le cas des données météorologiques, des données issues des automates ou des données mesurées sur les plantes, les fichiers CSV n'ont pas le même format que les tables correspondantes : dans les fichiers CSV les variables mesurées apparaissent en effet en tant que colonnes, alors qu'elles sont en ligne dans les tables de la base de données. Le script PHP appelle donc un second script R chargé d'insérer les données de manière adéquate dans les tables de la base de données.

3.4.c Upload et transfert de fichiers

On a vu précédemment qu'il était possible de télécharger (upload) un fichier Excel ou autre via les formulaires HTML d'insertion d'une nouvelle expérimentation, d'un génotype ou d'un commentaire. On peut également joindre un fichier Excel de génotypage à un fichier CSV de nouveaux génotypes à insérer, ou un fichier ZIP de scans à un fichier CSV de données de mesures de plantes à insérer. Enfin, on peut déposer sur le serveur dans un répertoire temporaire des lots de photos prises par les automates ou par les expérimentateurs.

Ces fichiers ou lots de fichiers sont validés ou non en même temps que les données saisies dans le formulaire ou présentes dans les fichiers CSV.

Si un seul fichier a été téléchargé (formulaire de saisie d'un seul enregistrement ou insertion d'un fichier CSV de génotypes), il est vérifié dans le script PHP au niveau de son type (extension du fichier), de son nom, et de sa taille. Dans le cas d'insertion d'un fichier CSV de génotypes, le script R de validation des données vérifie également l'adéquation du nom du fichier téléchargé avec le nom spécifié dans le fichier CSV. Si les données sont validées, les fichiers sont déplacés sur le serveur dans le répertoire adéquat via PHP.

Dans le cas d'insertion de données mesurées sur les plantes et de téléchargement d'un fichier ZIP de scans de plantes, ce fichier est décompressé. Un script R vérifie les données du fichier CSV et leur concordance avec les fichiers de l'archive ZIP. Si les données sont validées, un autre script R insère les données dans la base et déplace les scans dans le répertoire approprié sur le serveur. Les scans existants dans l'archive ZIP mais ne figurant pas dans le fichier CSV de données sont signalés à l'utilisateur.

Enfin, dans le cas de lots de photos déposées sur le serveur avant le téléchargement du fichier CSV de données à insérer (données issues des automates ou données de mesures de plantes), un script R vérifie les données du fichier CSV et leur adéquation avec les fichiers présents sur le serveur. Si les données sont validées, un autre script R insère les données dans la base et déplace sur le serveur les fichiers dans les répertoires adéquats. Les fichiers restants dans le répertoire temporaire seront ensuite supprimés par les administrateurs.

3.4.d Validation de données

Les données saisies dans les formulaires ou contenues dans des fichiers CSV sont validées au niveau de leur type (numérique, texte, données manquantes autorisées ou non, etc.), de leur format (format des dates, format des identifiants, nombre de caractères maximal pour une chaîne de caractères, etc.) et on vérifie l'adéquation des données entre elles. Dans les fichiers CSV, on vérifie également le nom des colonnes et leur nombre. Pour chaque champ, des règles strictes doivent être rigoureusement respectées avant que l'enregistrement ou le fichier CSV ne soit validé.

On vérifie également pour chaque donnée et chaque fichier s'il n'est pas déjà présent dans la base de données.

3.5 Modification et suppression des données

La modification et la suppression des données sont pour l'instant réservées à l'administrateur et aux utilisateurs ayant des droits suffisants. L'accès aux autres utilisateurs est refusé. Ce module n'est pas encore fonctionnel.

3.6 Gestion des groupes et utilisateurs

La gestion des groupes et des utilisateurs est réservée à l'administrateur et aux utilisateurs ayant des droits suffisants. L'accès aux autres utilisateurs est refusé. Ils peuvent ainsi créer, modifier ou supprimer des utilisateurs ou des groupes d'utilisateurs.

Conclusion et perspectives

La mise en place de la base de données et de son interface Web a été réalisée en 8 mois environ avant d'être opérationnelle. Une partie du temps non négligeable consiste maintenant à insérer dans la base de données l'ensemble des données accumulées depuis la construction de la première plate-forme PHENOPSIS en 2002. Avec une première étude exploratoire des données mesurées sur les plantes de 6 expérimentations nous avons déjà repéré des anomalies dans les mesures d'une expérimentation (**figure 10**).

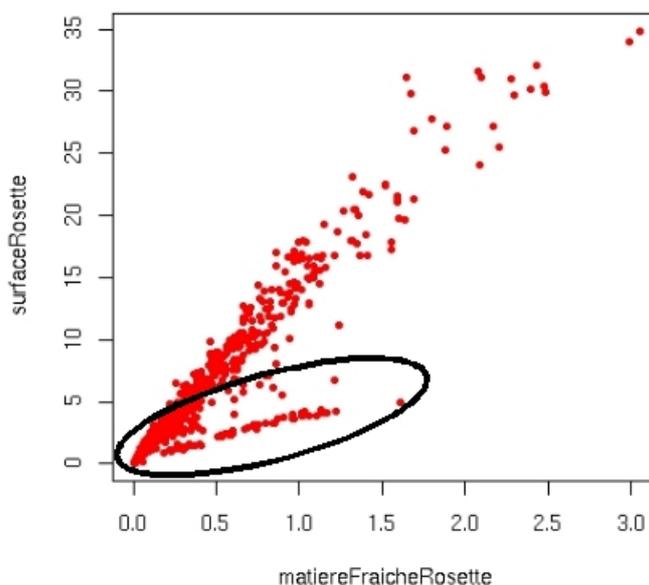


Figure 10 : Surface de la rosette en fonction du poids de la matière fraîche de la rosette

En important dans la base la totalité des données produites par les chambres de culture nous analyserons notamment des données de croissance à grande échelle.

Dans le futur, afin de limiter le temps nécessaire à cette insertion des données, nous envisageons de mettre en place un système d'acquisition en temps réel des données météorologiques, des données issues des automates et des photos, depuis les ordinateurs reliés aux chambres de culture et aux automates jusqu'au serveur hébergeant la base de données, l'interface Web et les fichiers.

Ce serveur appartenant au LEPSE, l'interface Web n'est accessible qu'aux ordinateurs connectés au réseau du LEPSE. Afin d'ouvrir l'accès à l'interface aux différents laboratoires de recherche partenaires, la base de données, l'interface Web et les fichiers doivent être déplacés prochainement sur le serveur du centre Inra de Montpellier.

De plus, pour que les utilisateurs extérieurs puissent effectuer des analyses statistiques simples sur les données et produire rapidement des graphiques pour visualiser les données de façon synthétiques, il est prévu d'ajouter à l'application Web un module d'analyses de données par le biais d'appel à des scripts R.

Enfin, il est également prévu d'ajouter un module d'analyses d'images qui permettrait d'extraire automatiquement des photographies et des scans des données de croissance sur les plantes.

Remerciements : Nous remercions Virginie Rossard pour le partage de son savoir-faire sur la gestion des bases de données et sur MySQL, ainsi que le projet intégré européen AGRON-OMICS (contrat LSHG-CT-2006-037704) qui a financé ce travail.

Bibliographie

Boyes DC, Zayed MA, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Görlach J (2001) Growth stage-based phenotypic analysis of *Arabidopsis* a model for high throughput functional genomics in plants. *Plant Cell* 13: 1499-1510.

Dauzat M., Simon E., Granier C., Hamard P., Muller B., Christophe A., Combes D., Simonneau T. (2004) Un automate de phénotypage pour cultiver des plantes à des états hydriques du sol contrôlés: Un outil d'aide à la caractérisation de la réponse de la croissance et de la transpiration à la sécheresse. *Le Cahier des Techniques de l'Inra*, 53, 21-33.

Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux JJ, Rolland, G, Bouchier-Combaud S, Lebaudy A, Muller B, Simonneau T, Tardieu F (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist*, 169 (3): 623-635.

Darmaillac Y, Rigaux P (2005) Maîtriser MySQL 5, *O'Reilly*

Sklar D (2004) Introduction à PHP 5, *O'Reilly*

Niederst J (2002) HTML précis et concis, *O'Reilly*