

Descriptif des formats de fichiers CSV de données à insérer dans la base de données PHENOPSIS DB



Sommaire

<u>1.Description et règles générales.....</u>	3
<u>2.Gestion des erreurs.....</u>	4
<u>3.Insertion d'un fichier de consignes environnementales.....</u>	5
<u>4.Insertion d'un fichier de géotypes.....</u>	7
<u>5.Insertion d'un fichier de pots d'une expérimentation.....</u>	9
<u>6.Insertion d'un fichier d'organes.....</u>	13
<u>7.Insertion d'un fichier de variables.....</u>	15
<u>8.Insertion d'un fichier de données de mesures de plantes.....</u>	18
<u>9.Insertion d'un fichier de données de stades phénologiques.....</u>	21
<u>10.Insertion d'un fichier de données météorologiques.....</u>	22
<u>10.1 Insertion de données météorologiques provenant de l'automate PHENOPSIS.....</u>	23
<u>10.2 Insertion de données météorologiques provenant des cabinets de culture.....</u>	23
<u>11.Insertion d'un fichier de données d'irrigation.....</u>	24
<u>12.Annexe.....</u>	28

1. Description et règles générales

Ce document décrit les formats requis des fichiers à insérer dans la base de données PHENOPSIS DB via l'interface Web.

Respectez de manière précise et exhaustive ces formats afin que vos fichiers soient validés !

Si vos fichiers contiennent des erreurs et ne sont pas validés au moment de l'insertion, les erreurs en question seront répertoriées dans un fichier texte que vous pourrez télécharger afin de corriger votre fichier.

L'URL de l'interface pour l'insertion des données est :

<http://lps-junior.supagro.inra.fr/phenopsis/>

- Les fichiers doivent être au format **CSV**. Sous Open Office, il faut sélectionner :
 - Jeu de caractères Europe Occidentale (Windows-1252/WinLatin 1)
 - séparateur de champ **point-virgule**
 - séparateur de texte **vide**.
- **Le séparateur étant le point-virgule, aucune cellule ne doit contenir de point virgule (attention aux champs texte !) !!**
- La **décimale** est indiquée par un « . » et non par une « , ».
Vous pouvez modifier ce paramètre dans Excel, « Outils », « Options » et paramètres linguistiques, ou bien dans « Panneau de configuration », « Options régionales et linguistiques », « Personnaliser ».
- Respectez scrupuleusement le **nombre** de colonnes (même si vous en avez des vides), leur **nom** imposé ainsi que l'**ordre** des colonnes.
- Respectez également le **format** des données : formats des dates, formats texte ou numériques...
- Le fichier ne doit contenir aucun graphique.
- Les valeurs manquantes doivent être indiquées par une cellule **vide**.
- Le fichier doit être **propre** : pas de donnée qui « traîne » dans une colonne non définie, à la fin du fichier ou autre...
- Vérifiez que les données n'existent pas déjà dans la base en consultant l'interface PHENOPSIS DB !
- Attention lorsque vous enregistrez vos fichiers Excel en fichiers CSV : Le format CSV enregistre les données telles qu'elles sont affichées dans le fichier Excel ! Ainsi, si vos données numériques ne sont affichées qu'avec un seul chiffre après la virgule, un seul chiffre après la virgule sera enregistré dans le fichier CSV !!!

Pensez-donc à vérifier la précision que vous souhaitez conserver dans vos données, et à afficher dans vos fichiers Excel un nombre suffisant de chiffres

après la virgule !!!

- Enfin, tous les fichiers type sont disponibles sur l'interface PHENOPSIS DB, sur les pages d'insertion de données. Ces fichiers sont fournis au format Excel, il vous faut ensuite les remplir, puis les enregistrer au format CSV en prenant bien garde au séparateur (« ; »), à la décimale (« . »), et au séparateur de texte (aucun).

2. Gestion des erreurs

Si lors de l'insertion d'un fichier CSV de données, un message d'erreur comme celui-ci apparaît :

Echec de la requête !! Les données n'ont pas été insérées !! [...] Si le problème persiste veuillez contacter l'administrateur !!

Votre fichier CSV a peut-être été corrompu lors de son enregistrement. Suivez les conseils suivants pour déceler les problèmes :

- Ouvrez le fichier CSV avec un éditeur de texte classique.
- Vérifiez que les champs ne sont pas entourés de guillemets. S'ils le sont, enregistrez à nouveau votre fichier Excel en CSV en précisant que le séparateur de texte est vide.
- Vérifiez que les champs sont bien séparés par des points-virgules. S'ils ne le sont pas, enregistrez à nouveau votre fichier Excel en CSV en précisant que le séparateur de champ est le point-virgule.
- Vérifiez que la dernière colonne est bien prise en compte. En effet, si la dernière colonne du fichier est vide, Excel peut ne pas la reconnaître. Si c'est le cas, sélectionnez la colonne dans Excel, et appliquez lui le format correct (Format → Cellules → Texte par exemple). Enregistrez à nouveau votre fichier Excel en CSV.

Si aucune de ces actions n'aboutit, contactez l'administrateur !

Les paragraphes suivants décrivent les formats de chaque fichier CSV. En vert sont écrits les noms précis des colonnes et leurs descriptions suivent.

3. Insertion d'un fichier de consignes environnementales

Ce fichier doit contenir les caractéristiques des consignes environnementales d'une expérimentation.

- Nombre de colonnes : 6

- Colonne 1 : **idManip**
 - Contient l'identifiant de la manip concernée par les consignes environnementales, de la forme CiMj (+ une éventuelle lettre majuscule)
 - L'identifiant doit déjà exister dans la base de données
 - Ne peut contenir de valeurs manquantes

- Colonne 2 : **idVariable**
 - Contient l'identifiant de la variable environnementale. Les valeurs possibles sont les identifiants des variables environnementales à consulter sur l'interface Web (« temperatureJourConsigne », ...)
 - Ne peut contenir de valeurs manquantes

- Colonne 3 : **dateDebut**
 - Contient la date de début de la consigne environnementale
 - Doit être de format **aaaa-mm-jj hh:mm:ss** (année-mois-jour espace heure:minute:seconde)
 - Doit être postérieure ou égale à la date de début de la manip
 - Ne peut contenir de valeurs manquantes
 - Attention !! Les dates doivent être saisies en **heure solaire** pour être cohérent avec les données météo réelles !!

- Colonne 4 : **dateFin**
 - Contient la date de fin de la consigne environnementale
 - Doit être de format **aaaa-mm-jj hh:mm:ss** (année-mois-jour espace heure:minute:seconde)
 - Doit être antérieure ou égale à la date de fin de la manip
 - Facultatif
 - Attention !! Les dates doivent être saisies en **heure solaire** pour être cohérent avec les données météo réelles !!

- Colonne 5 : **valeurNum**
 - Contient la valeur numérique de la consigne
 - Contient des valeurs manquantes si la consigne n'est pas une variable numérique (ex : « debutHeureJourConsigne »)
 - Doit être de type numérique

- Colonne 6 : **valeurHeure**
- Contient la valeur d'heure de la consigne (ex : « debutHeureJourConsigne »)
- Contient des valeurs manquantes si la consigne n'est pas une variable heure
- Doit être de format **hh:mm:ss** (heure:minute:seconde)

idManip	idVariable	dateDebut	dateFin	valeurNum	valeurHeure
C3M1	photoperiodeConsigne	2007-01-01 00:00:00	2007-03-01 00:00:00	12	
C3M1	heureDebutJourConsigne	2007-01-01 00:00:00	2007-03-01 00:00:00		08:00:00
C3M1	rayonnementConsigne	2007-01-01 00:00:00	2007-01-21 08:00:00	150	
C3M1	rayonnementConsigne	2007-01-21 08:00:00	2007-03-01 00:00:00	100	
C3M1	temperatureJourConsigne	2007-01-01 00:00:00	2007-01-15 08:00:00	25	
C3M1	temperatureJourConsigne	2007-01-15 08:00:00	2007-02-15 00:00:00	20	
C3M1	temperatureJourConsigne	2007-02-15 00:00:00	2007-03-01 00:00:00	15	

4. Insertion d'un fichier de génotypes

Ce fichier doit contenir les caractéristiques de nouveaux génotypes.

Attention !! Toute déclaration de nouveaux génotypes doit être impérativement validée auparavant par un responsable scientifique !

- Nombre de colonnes : 12
- Colonne 1 : **idGenotype**
- Contient les identifiants des génotypes, c'est-à-dire son nom courant (Col0, ..). Cet identifiant est celui utilisé dans la table des pots de la manip.
- Consultez la nomenclature chez *Arabidopsis thaliana*:
 - x Sur le site <http://www.arabidopsis.org/portals/nomenclature/guidelines.jsp#genbank>
 - x Dans la publication de Meinke Koornneef (consultable sur le site PHENOPSIS sur les pages de déclaration de génotypes)
- Evitez tant que possible les caractères spéciaux dans l'identifiant ('', '_', apostrophes, ...). Les espaces, les accents et le point-virgule ne sont pas autorisés. Consultez la liste des génotypes.
- Si un même génotype provient de différents laboratoires, son identifiant sera suivi d'indices entre parenthèses. Ex : Col-0(1), Col-0(2), ...
- Chaque identifiant doit être unique
- L'identifiant de génotype du pot servant au calibrage doit être **'calibrage'**
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 50

■ Colonne 2 : **nomGenotype**

- Contient les noms complets des génotypes.

Dans le cas de population de lignées recombinantes le nom doit être au format RILF $x[A \times B]-n$ avec x le nombre de croisement effectués pour obtenir la population, A et B les parents et n le numéro de la RIL.

- Evitez tant que possible les caractères spéciaux dans le nom
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 50

■ Colonne 3 : **code**

- Contient l'identifiant du génotype tel qu'il est référencé (NASC, ...)
- Nombre maximal de caractères : 20
- Facultatif

■ Colonne 4 : **type**

- Contient le type de génotype : RIL, mutant, accession... (consultez la liste des génotypes pour obtenir la liste des différents types)
- Nombre maximal de caractères : 50
- Facultatif

■ Colonne 5 : **fondGenetique1**

- Contient le nom du parent 1
- Nombre maximal de caractères : 50
- Facultatif

■ Colonne 6 : **fondGenetique2**

- Contient le nom du parent 2
- Nombre maximal de caractères : 50
- Facultatif

■ Colonne 7 : **provenance**

- Contient le laboratoire donneur du génotype
- Nombre maximal de caractères : 50
- Facultatif

■ Colonne 8 : **référence**

- Contient une ou plusieurs références dans lesquelles le génotype est décrit
- Nombre maximal de caractères : 100
- Facultatif

■ Colonne 9 : **fonction**

- Contient la fonction en anglais du génotype
- Nombre maximal de caractères : 100
- Facultatif

■ Colonne 10 : **idGoupe**

- Contient le nom du groupe d'utilisateurs qui auront des permissions sur les données du génotype. Si aucun nom de groupe n'est saisi, les données de ce génotype seront a priori consultables par tous les utilisateurs.

Consultez la liste des groupes sur la page d'insertion.

Si vous avez besoin de créer un nouveau groupe d'utilisateurs, contactez l'administrateur ou si vous en avez la permission rendez vous dans la partie « Gestion des Groupes et Utilisateurs ».

- Nombre maximal de caractères : 20
- Facultatif

■ Colonne 11 : **genotypage**

- Contient le nom du fichier Excel de génotypage du génotype, dans le cas de RIL. Ce fichier contient les valeurs aux différents marqueurs, la carte génétique, ...

Ce fichier doit soit déjà exister dans la base de données, soit être téléchargé en même temps que le fichier CSV sur la page d'insertion des génotypes.

On ne peut télécharger qu'un seul fichier Excel à la fois sur cette page, c'est pourquoi le fichier CSV ne doit contenir qu'**un seul** nom de fichier de génotypage.

Dans le cas de plusieurs nouveaux fichiers de génotypage à insérer dans la base, il faut donc faire autant de fichiers CSV que de fichiers Excel de génotypage.

- Nombre maximal de caractères : 100
- Facultatif

■ Colonne 12 : **commentaire**

- Nombre maximal de caractères : 500
- Facultatif

	A	B	C	D	E	F	G	H	I	J	K	L
1	idGenotype	nomGenotype	code	type	fondGenetique1	fondGenetique2	provenance	reference	fonction	groupe	genotypage	commentaire
2	Cesa2			mutant T-DNA KO	Col-Q(7)		G. Mouille (INRA)	Desprez et al 2007 PN	Cellulose Synthase subunit	LEPSE		
3												

5. Insertion d'un fichier de pots d'une expérimentation

Attention ! Cas d'expérimentation dans les PHENOPSIS :

Les identifiants des pots sont déclarés **automatiquement** dans la base de données dès l'insertion du fichier data_in (à condition que l'expérimentation ait été déclarée au préalable !).

Pous pouvez ensuite **compléter** les informations sur ces pots : génotypes, poids, etc .. via la page de consultation des pots (Consulter l'expérimentation => Consulter la liste des pots).

Les informations suivantes ne concernent donc que l'insertion de pots d'expérimentations dans les cabinets de culture, ou d'anciennes expérimentations dans les PHENOPSIS (avant 2010).

Ce fichier doit contenir les caractéristiques des pots concernant **une manip et une seule**. L'identifiant de la manip en question est saisi sur la page d'insertion.

- Nombre de colonnes : 18

- Colonne 1 : **idPotManip**
 - Contient les identifiants des pots, de la forme CiMj-k avec CiMj l'identifiant de la manip à laquelle appartiennent les pots et k le numéro du pot
 - Chaque identifiant doit être unique
 - Ne peut contenir de valeurs manquantes
 - Nombre maximal de caractères : 20

- Colonne 2 : **idManip**
 - Contient l'identifiant de la manip à laquelle appartiennent les pots, de la forme CiMj (+ une éventuelle lettre majuscule)
 - L'identifiant doit déjà exister dans la base de données•
 - Ne peut contenir de valeurs manquantes

- Colonne 3 : **idGenotype**
 - Contient l'identifiant du génotype des plantes contenues dans le pot. Cet identifiant est celui utilisé dans la table des génotypes (champ idGenotype)
 - Chaque identifiant doit déjà exister dans la base de données
 - Ne peut contenir de valeurs manquantes
 - L'identifiant associé aux pots de calibrage est « **calibrage** »

- Colonne 4 : **positionPot**
 - Contient la position du pot dans le robot de la forme $P_i_A_j$ avec P_i le plateau et A_j l'alvéole
 - Si la manip a lieu dans un automate PHENOPSIS, la colonne ne peut pas contenir de valeurs manquantes, sinon la colonne doit être vide.
 - Nombre maximal de caractères : 10

- Colonne 5 : **bloc**
 - Contient le numéro de bloc auquel appartient le pot
 - La colonne doit être vide si la manip a lieu dans un cabinet de culture
 - Facultatif

- Colonne 6 : **poidsPotNonTroue**
 - Contient le poids du pot non troué vide
 - Ne peut contenir de valeurs manquantes
 - Doit être de type numérique

- Colonne 7 : **poidsPotTroue**
 - Contient le poids du pot troué vide
 - Ne peut contenir de valeurs manquantes

- Colonne 8 : **poidsSolSec**
 - Contient le poids de sol sec contenu dans le pot
 - Ne peut contenir de valeurs manquantes
 - Doit être de type numérique

- Colonne 9 : **dateSemis**
 - Contient la date de semis des plantes dans le pot
 - Doit être de format **aaaa-mm-jj hh:mm** (année-mois-jour)
 - Ne peut contenir de valeurs manquantes

- Colonne 10 : **nbPlante**
 - Contient le nombre de plantes contenues dans le pot (ces plantes seront ensuite identifiées dans les mesures par les lettres A, B, C, ..)
 - Doit être de type numérique
 - Facultatif

- Colonne 11 : **substrat**
 - Contient le substrat utilisé dans le pot («*Balsera*» ou «Phenoarch »)
 - Nombre maximal de caractères : 20
 - Facultatif (par défaut «*Balsera*»).

- Colonnes 12 à 16 : **humiditeNiveau**,
 - Valeurs booléennes : contient 1 si l'humidité cible du pot est fixée à cette valeur à un moment de l'expérimentation, 0 sinon. Ces colonnes contiennent les **consignes** d'humidité qui ne sont pas forcément les valeurs réelles d'humidité des pots.
 - Facultatif (par défaut 0)
 - Les niveaux d'humidité dépendent du substrat utilisé. Pour le substrat «*Balsera*» les correspondances sont :
 - x humiditeNiveau1 : $40 \leq \text{humidité} < 50 \%$
 - x humiditeNiveau2 : $30 \leq \text{humidité} < 40 \%$
 - x humiditeNiveau3 : $20 \leq \text{humidité} < 30 \%$
 - x humiditeNiveau4 : $10 \leq \text{humidité} < 20 \%$
 - x humiditeNiveau5 : $0 \leq \text{humidité} < 10 \%$

Ex : Un pot ayant pour humidité cible 40% dans un premier temps puis 25% devra avoir 1 dans les colonnes humiditeNiveau1 et humiditeNiveau3.
 - Pour le substrat «*Phenoarch*» les correspondances sont :
 - x humiditeNiveau1 : 1.5-1.8
 - x humiditeNiveau2 : 1-1.5
 - x humiditeNiveau3 : 0.8-1
 - x humiditeNiveau4 : 0.5-0.8
 - x humiditeNiveau5 : 0-0.5

- Colonne 17: **typePot**
 - Type de pot utilisé («a» : anciens pots phenopsis ; «b» : nouveaux pots ; «u» : indéfini).
 - Nombre maximal de caractères : 1
 - Facultatif (par défaut «u»).

- Colonne 18 : **grille**
 - Utilisation d'une grille sur le substrat (« OUI » ou « NON »)

- Nombre maximal de caractères : 3
- Facultatif (par défaut «NON»).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
idPotManip	idManip	idGenotype	positionPot	bioc	poidsPotNontrou	poidsPotTrou	poidsSolSec	dateSemis	nbPlante	substrat	humiditeNiveau1	humiditeNiveau2	humiditeNiveau3	humiditeNiveau4	humiditeNiveau5	typePot	grille
C1M20A-1	C1M20A	calibrage	P1_A1		15,4	15,4	###	2007-09-14	0	Thieux	1		1			a	OUI
C1M20A-2	C1M20A	Spk1	P1_A7		15,2	15,8	161,1525356	2007-09-14	1	Thieux	1		1			a	NON

6. Insertion d'un fichier d'organes

Ce fichier doit contenir les caractéristiques de nouveaux organes étudiés, c'est-à-dire de nouvelles parties de la plante.

Attention !! Toute déclaration de nouveaux organes doit être impérativement validée auparavant par un responsable scientifique !

- Nombre de colonnes : 6
- Colonne 1 : **idOrgane**
 - Contient l'identifiant de l'organe.
 - Cet identifiant doit être en **minuscules** et ne doit contenir **ni accent ni espace ni caractère spécial** (tiret, slash, point, virgule, ...). ex : rosette, feuille06, petiole02.. (Consultez la liste des organes)
 - Chaque identifiant doit être unique
 - Ne peut contenir de valeurs manquantes
 - Nombre maximal de caractères : 50
- Colonne 2 : **nomAnglais**
 - Contient le nom anglais de l'organe. La nomenclature est la même que pour idOrgane.
 - Ne peut contenir de valeurs manquantes
 - Nombre maximal de caractères : 50
- Colonne 3 : **definition**
 - Contient la définition de l'organe
 - Ne peut contenir de valeurs manquantes
 - Nombre maximal de caractères : 500
- Colonne 4 : **definitionAnglais**
 - Contient la définition en Anglais de l'organe

- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 500
- Colonne 5 : **rang**
- Contient le rang de l'organe : s'il s'agit de la feuille n°61, le rang sera 61. S'il s'agit d'un organe sans rang (ex rosette), alors le champ est vide
- Facultatif
- Numérique
- Colonne 6 : **accessionTerm**
- Contient la correspondance de l'organe dans les ontologies existantes, principalement la Plant Ontology (PO) : http://www.gramene.org/plant_ontology/
- Facultatif
- Nombre maximal de caractères : 100

idOrgane	nomAnglais	definition	definitionAnglais	rang	accessionTerm
rosette	rosette	rosette	rosette		PO:0008034
plante	plant	plante entière	whole plant		PO:0000003
feuille01	leaf01	feuille 1 de la rosette	1 st rosette leaf	1	PO:0000014
feuille02	leaf02	feuille 2 de la rosette	2 nd rosette leaf	2	PO:0000014
feuille03	leaf03	feuille 3 de la rosette	3 rd rosette leaf	3	PO:0000014
feuille04	leaf04	feuille 4 de la rosette	4 th rosette leaf	4	PO:0000014

7. Insertion d'un fichier de variables

Ce fichier doit contenir les caractéristiques des nouvelles variables mesurées.

Attention !! Toute déclaration de nouvelles variables doit être impérativement validée auparavant par un responsable scientifique !

- Nombre de colonnes : 11
- Colonne 1 : **idVariable**
- Contient l'identifiant de la variable.
L'identifiant ne doit contenir **ni espace ni caractère spécial** (tiret, point, ..) **ni accent**.
L'identifiant doit être en minuscules. S'il contient plusieurs mots, ces mots doivent être juxtaposés, et l'initiale d'un nouveau mot en majuscule.

Consultez la liste des variables !

- Chaque identifiant doit être unique
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 50

■ Colonne 2 : **nomAnglais**

- Contient le nom en Anglais de la variable
Ce nom peut contenir des espaces.
Seule la première lettre est en majuscule, ex : Rosette leaf number
Consultez la liste des variables!
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 50

■ Colonne 3 : **nature**

- Contient le type des variables : organe, robot, météo..
Ce type doit déjà exister dans la base de données.
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 20

■ Colonne 4 : **definition**

- Contient la définition de la variable.
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 500

■ Colonne 5 : **definitionAnglais**

- Contient la définition en Anglais de la variable.
- Ne peut contenir de valeurs manquantes
- Nombre maximal de caractères : 500

■ Colonne 6 : **unite**

- Contient l'unité de la variable si la variable est numérique
- Nombre maximal de caractères : 20
- Facultatif

■ Colonne 7 : **min**

- Contient la valeur minimale de la variable si la variable est numérique

- Facultatif

- Colonne 8 : **max**
- Contient la valeur maximale de la variable si la variable est numérique
- Facultatif

- Colonne 9 : **type**
- Contient le type de la variable, qui doit être *numerique*, *texte* ou *fichier*.
- Obligatoire

- Colonne 10 : **protocole**
- Contient le nom d'un fichier Excel de protocole de mesure de la variable, s'il existe.
Ce fichier doit soit déjà exister dans la base de données, soit être téléchargé en même temps que le fichier CSV sur la page d'insertion des variables.
On ne peut télécharger qu'un seul fichier Excel à la fois sur cette page, c'est pourquoi le fichier CSV ne doit contenir qu'**un seul** nom de fichier de protocole.
Dans le cas de plusieurs nouveaux fichiers de protocoles à insérer dans la base, il faut donc faire autant de fichiers CSV que de fichiers Excel de protocoles.
- Nombre maximal de caractères : 100
- Facultatif

- Colonne 11 : **accessionTerm**
- Contient la correspondance de la variable dans les ontologies existantes, principalement :
 - x dans la Trait Ontology (TO : http://www.gramene.org/plant_ontology/). Ex : epaisseurFeuille <=> TO:0000258
 - x en combinant des termes de la Plant Ontology (PO : http://www.gramene.org/plant_ontology/) et de la phenotypic qualities ontology (PATO : http://obofoundry.org/wiki/index.php/PATO:Main_Page) sur le modèle EQ (Entity Qualifier). Ex : epaisseurFeuille <=> PATO:0000915 (épaisseur) + PO:0000014 (feuille de la rosette)
- Facultatif
- Nombre maximal de caractères : 200

idVariable	nomAnglais	nature	definition
surfaceFeuille	Leaf area	organe	Surface de la feuille ...
matiereSecheRacine	Root dry weight	organe	Masse du système racinaire ...
stadePhenologiquePlante	Plant phenological stage	organe	Stade phénologique de ...
nbFeuilleRosette	Rosette leaf number	organe	Nombre de feuilles ...
matiereSecheLimbeRosette	Rosette lamina dry weight	organe	Masse des limbes ...

definitionAnglais	unite	min	max	type	protocole	accessionTerm
Measurement of the total ...	cm ²	6		numerique		TO:0000540
Root system weight ...	g	0		numerique		TO:0000078
Plant phenological ...				texte	protoStadePheno.xls	
Number of leaves ...		0	60	numerique		
Rosette lamina weight ...	g	0		numerique		
Pot weight measured ...	g	0		numerique		

8. Insertion d'un fichier de données de mesures de plantes

Ce fichier doit contenir les données mesurées sur les plantes d'**une et une seule manip.** L'identifiant de la manip en question est saisi sur la page d'insertion.

Le fichier peut concerner une ou plusieurs variables.

L'insertion de données mesurées sur les plantes peut donc être réalisée en **autant d'étapes que nécessaire** (une variable à la fois, toutes les variables en même temps, etc ...).

L'insertion du fichier et des éventuels scans ou autres fichiers d'images de plantes peut être longue : moins de 10 minutes pour un fichier CSV contenant 10000 données.

Attention : lors de l'insertion de mesures météorologiques, de mesures de plante ou de mesures du robot, il se peut que vos fichiers soient validés et que pourtant **les données ne soient pas entièrement insérées dans la base de données!** Dans ce cas vous obtiendrez le message d'erreur suivant : «Echec de la requête! Le fichier a bien été validé mais les données n'ont pas toutes été insérées dans la base de données!! Vous devez contacter l'administrateur!! ». Dans ce cas, il vous faut **absolument** contacter l'administrateur pour lui signaler le bug et pour observer quelles données ont été insérées ou non!

- Nombre de colonnes minimum : 4

- Colonne 1 : **idPotManip**
 - Contient les identifiants des pots, de la forme *CMj-k* avec *CMj* l'identifiant de la manip à laquelle appartiennent les pots et *k* le numéro du pot. La manip doit correspondre à celle sélectionnée sur la page d'insertion.
 - Ne peut contenir de valeurs manquantes
 - Les idPotManip doivent déjà exister dans la base de données !

- Colonne 2 : **idPlante**
 - Contient l'identifiant de la plante, qui doit être entre A et Z
 - Ne peut contenir de valeurs manquantes

- Colonne 3 : **date**
 - Contient la date des mesures
 - Doit être de format **aaaa-mm-jj hh:mm:ss** (année-mois-jour espace heure:minute:seconde)
 - Ne peut contenir de valeurs manquantes
 - Chaque triplet idPotManip / idPlante / date doit être unique dans le fichier

Les colonnes suivantes contiennent les traits mesurés sur les plantes. Ces colonnes peuvent contenir des valeurs manquantes.

Attention : vérifiez bien dans quelle unité les variables doivent être exprimées (consultez l'interface !) !!

- Colonnes 4 et + :
 - **Pour les dates de floraison ou de stade F2 d'une plante : dateF2 ou dateF1**
 - x Ces colonnes contiennent les dates en question, qui doivent être de format **aaaa-mm-jj** (année-mois-jour)
 - x La date des mesures correspondantes (colonne **date**) est la date à laquelle on a saisi les différentes dates de floraison ou de stade F2.

 - **Autres mesures (longueurFeuille, surfaceFeuille, matiereSechePetiole, scanRosette, ...)** :
 - x Les noms de colonnes doivent être de la forme **variable_organe_stade**, où :
 - **variable** est la variable étudiée, existant dans la base de données. Ex : *surfaceFeuille*.
 - **organe** : l'organe concerné, lui aussi existant dans la base de données. Ex : *feuille06*.
 - **stade** : stade de la mesure. Peut être un stade phénologique (ex : *6.00* (floraison)),

1.6 (feuille 6 initiée)), *Dynamic* (en croissance sans plus de précision), *28DAS* (28 days after sowing), etc ..

Consultez la définition des stades phénologiques de Boyes : <http://lps-junior/phenopsis/InfoBDD.php>

=> Ainsi, la longueur de la feuille n°6 mesurée au bolting sera déclarée comme *longueurFeuille_feuille06_5.10*.

x Cas particuliers des variables de fichiers (scans, images de coupes de feuilles, ...) :

- Les colonnes contiennent le nom des fichiers. Tous les types de format sont acceptés (.jpg, .tif, etc).
- Si ces colonnes ne sont pas vides, et que les fichiers référencés ne sont pas déjà présents dans la base de données, **il faut alors avoir auparavant transféré les fichiers correspondants sur le serveur ! Seuls les administrateurs peuvent effectuer cette tâche !**
- Un même fichier peut concerner plusieurs plantes ou pots et donc être répertorié plusieurs fois dans la colonne.
- Remarques :
 - Si des fichiers sont manquants, votre fichier ne sera pas validé, et vous aurez à disposition un fichier CSV contenant les noms des fichiers manquants sur le serveur.
 - Il vous faudra alors soit transférer ces fichiers sur le serveur, soit les supprimer des colonnes en question (cf Annexe).
 - Au contraire, si vous transférez sur le serveur des fichiers dont les noms n'apparaissent dans aucune des colonnes, les noms de ces fichiers vous seront communiqués lors de l'insertion de votre fichier de données, et les fichiers en question seront automatiquement supprimés du répertoire TMP après 1 mois.

idPotManip	idPlante	date	nbFeuilleRosette_rosette_6.00	surfaceRosette_rosette_6.00
C1M7-1	A	2006-02-20 00:00:00	12	4.96418
C1M7-2	A	2006-02-10 00:00:00	7	1.86385
C1M7-3	A	2006-02-24 00:00:00	10	5.16518
C1M7-4	A	2006-02-24 00:00:00	10	3.10642
C1M7-5	A	2006-02-14 00:00:00	9	4.3429

longueurFeuille_feuille06_6.00	dateF2	dateFI	scanRosette_rosette_6.00
0.97	2006-01-23	2006-02-15	C1M7-1.JPG
1.48	2006-01-25	2006-02-10	C1M7-2.JPG
1.25	2006-01-23	2006-02-15	C1M7-3.JPG
0.91	2006-01-26	2006-02-18	C1M7-4.JPG
1.51	2006-01-24	2006-02-12	C1M7-5.JPG

9. Insertion d'un fichier de données de stades phénologiques

Ce fichier doit contenir les données de stade phénologiques mesurées sur les plantes d'**une et une seule manip**. L'identifiant de la manip en question est saisi sur la page d'insertion.

Attention : lors de l'insertion de mesures météorologiques, de mesures de plante ou de mesures du robot, il se peut que vos fichiers soient validés et que pourtant **les données ne soient pas entièrement insérées dans la base de données!** Dans ce cas vous obtiendrez le message d'erreur suivant : «Echec de la requête ! Le fichier a bien été validé mais les données n'ont pas toutes été insérées dans la base de données!! Vous devez contacter l'administrateur !! ». Dans ce cas, il vous faut **absolument** contacter l'administrateur pour lui signaler le bug et pour observer quelles données ont été insérées ou non !

- Nombre de colonnes minimum : 4

- Colonne 1 : **idPotManip**

- Contient les identifiants des pots, de la forme $CMj-k$ avec CMj l'identifiant de la manip à laquelle appartiennent les pots et k le numéro du pot. La manip doit correspondre à celle sélectionnée sur la page d'insertion.
- Ne peut contenir de valeurs manquantes
- Les idPotManip doivent déjà exister dans la base de données !

- Colonne 2 : **idPlante**

- Contient l'identifiant de la plante, qui doit être entre A et Z
- Ne peut contenir de valeurs manquantes

- Colonne 3 : **date**

- Contient la liste des dates des relevés de stades phénologiques : chaque date doit être unique !
- Doit être de format **aaaa-mm-jj hh:mm:ss** (année-mois-jour espace heure:minute:seconde)

- Colonnes 4 et + : **date_i**

- Contiennent les stades phénologiques mesurés sur les plantes à chaque date de la colonne *date*. Ainsi la colonne *date_1* contient les stades relevés à la 1ère date de la colonne *date*, *date_2* contient les stades relevés à la 2ème date de la colonne *date*, etc.
- Les valeurs de stade phénologique possibles sont :
 - x « imbibition » : imbibition des graines
 - x « semis » : semis des graines dans le sol

- x « radicule » : émergente de racines
 - x « c2 » : cotylédons ouverts
 - x le nombre de feuilles > 1 mm, ex : 2, 3, ..
 - x « Bt » : 1ers boutons floraux visibles
 - x « Fl » : 1ère fleur ouverte
 - x « Flf » : toutes fleurs ouvertes
- Peuvent contenir des valeurs manquantes.

idPotManip	idPlante	date	date_1	date_2	date_3	date_4	date_5
C1M8-1	A	2006-03-17 00:00:00	c2	2	3	6	Bt
C1M8-2	A	2006-03-20 00:00:00	c2	c2			
C1M8-3	A	2006-03-24 00:00:00	c2	c2	2	4	4
C1M8-4	A	2006-03-29 00:00:00		c2	2	6	7

10. Insertion d'un fichier de données météorologiques

Ce fichier est issu des fichiers météo fournis par les chambres de culture.

Il peut prendre 2 formats : celui des fichiers météo fournis par les automates de culture PHENOPSIS, ou celui des fichiers météo fournis par les automates de culture.

Attention! De manière à pouvoir convertir les dates en jour julien en dates classiques à partir de la date contenue dans le nom du fichier, le fichier météo **ne doit pas contenir de données à cheval sur 2 années !!** Séparez les données en x fichiers par année !

Attention ! Les données doivent être à l'**heure solaire !!!** A partir du 01/04/2009, les données ont été récoltées au bon format. Les données récoltées avant le 01/04/2009 doivent donc être mises à l'heure solaire avant d'être insérées !!

Heure solaire = heure GMT = heure - 1 en hiver et heure - 2 en été

L'insertion de ce fichier est relativement brève : une minute pour un fichier de 4000 lignes.

Attention : lors de l'insertion de mesures météorologiques, de mesures de plante ou de mesures du robot, il se peut que vos fichiers soient validés et que pourtant les données ne soient pas entièrement insérées dans la base de données ! Dans ce cas vous obtiendrez le message d'erreur suivant : «Echec de la requête ! Le fichier a bien été validé mais les données n'ont pas toutes été insérées dans la base de données !! Vous devez contacter l'administrateur !! ». Dans ce cas, il vous faut absolument contacter l'administrateur pour lui

signaler le bug et pour observer quelles données ont été insérées ou non !

10.1 Insertion de données météorologiques provenant de l'automate PHENOPSIS

Le fichier à insérer est le fichier DAT fourni par les chambres de culture PHENOPSIS, sans modification.

A priori vous n'avez pas à insérer les données météorologiques !! Ces données sont insérées automatiquement dans la base de données 2 fois par jour !!

Le **nom du fichier** doit être de la forme *Ciaammjj.DAT* avec *Ci* l'identifiant de la chambre de culture PHENOPSIS (C1, C2 ou C3) et *aammjj* la date du fichier. L'année contenue dans la date sera utilisée pour convertir les dates en jour julien en dates classiques.

Un fichier ne concerne donc qu'**une seule chambre de culture** à la fois !

10.2 Insertion de données météorologiques provenant des cabinets de culture

Le fichier à insérer est le fichier DAT fourni par les cabinets de culture, sans modification.

Le **nom du fichier** doit être de la forme *CAaammjj.DAT* avec *aammjj* la date du fichier. L'année contenue dans la date sera utilisée pour convertir les dates en jour julien en dates classiques.

11. Insertion d'un fichier de données d'irrigation

Ce fichier doit contenir les données issues du robot PHENOPSIS d'**une et une seule manip**. L'identifiant de la manip en question est saisi sur la page d'insertion.

Par extension, ce fichier s'applique également aux manips en **cabinet de culture non automatisés**.

A priori vous n'avez pas à insérer les données robot des automates PHENOPSIS !! Ces données (ainsi que les images) sont insérées automatiquement dans la base de données en temps réel !

Concaténation et mise en forme de fichiers CSV d'irrigation :

Utilisez le script R 'ConcatenationIrrigation.R' (en téléchargement sur la page d'insertion de données d'irrigation) pour concaténer plusieurs fichiers CSV d'irrigation et les mettre au format requis pour leur insertion.

Attention ! Les données doivent être à l'**heure solaire** !!! A partir du 01/04/2009, les données ont été récoltées au bon format. Les données récoltées avant le 01/04/2009 doivent donc être mises à l'heure solaire avant d'être insérées !!

Heure solaire = heure GMT = heure - 1 en hiver et heure - 2 en été

L'insertion de ce fichier et des éventuelles photos issues du robot est très longue : plus de 2h pour valider et insérer un fichier de 20.000 lignes avec 13.000 photos ! Vous pouvez éventuellement scinder le fichier en plusieurs fichiers de 5000 lignes que vous insérerez un à un.

Attention : lors de l'insertion de mesures météorologiques, de mesures de plante ou de mesures du robot, il se peut que vos fichiers soient validés et que pourtant **les données ne soient pas entièrement insérées dans la base de données** ! Dans ce cas vous obtiendrez le message d'erreur suivant : «Echec de la requête ! Le fichier a bien été validé mais les données n'ont pas toutes été insérées dans la base de données !! Vous devez contacter l'administrateur !! ». Dans ce cas, il vous faut **absolument** contacter l'administrateur pour lui signaler le bug et pour observer quelles données ont été insérées ou non !

- Nombre de colonnes : 15

- Colonne 1 : **date**
 - Contient la date des mesures
 - Doit être de format **aaaa-mm-jj hh:mm:ss** (année-mois-jour espace heure:minute:seconde)
 - Ne peut contenir de valeurs manquantes

- Colonne 2 : **idPotManip**
 - Contient les identifiants des pots, de la forme $CiMj-k$ avec $CiMj$ l'identifiant de la manip à laquelle appartiennent les pots et k le numéro du pot. La manip $CiMj$ doit correspondre à celle sélectionnée sur la page d'insertion.
 - Ne peut contenir de valeurs manquantes
 - Les idPotManip doivent déjà exister dans la base de données !

- Colonnes 3 et + :
 - Les colonnes suivantes contiennent les différentes variables fournies par les robots ou remplies manuellement par les expérimentateurs. **Ces colonnes peuvent contenir des**

valeurs manquantes.

- Le triplet idPotManip-variable-date ne doit pas déjà exister dans la base de données.
- Colonne 3 : **numPot**
- Contient les numéros des pots, qui doivent correspondre à ceux contenus dans les identifiants des pots correspondants.
- Colonne 4 : **positionPot**
- Contient la position des pots dans le robot, de la forme $P_i_{A_j}$ avec P_i le plateau du pot et A_j l'alvéole du pot.
- Doit être vide si la manip a lieu en cabinet de culture.
- Colonne 5 : **nbPot**
- Contient le nombre de pots : 1 ou 2.
- Colonne 6 : **poidsAvantArrosage**
- Contient les valeurs de poids des pots avant l'irrigation.
- Colonne 7 : **poidsCible**
- Contient les valeurs de poids cible des pots définis par l'utilisateur.
- Colonne 8 : **eauDelivreeA**
- Contient soit les valeurs de quantité de solution nutritive A délivrée par le robot, soit les valeurs de quantité de solution nutritive délivrée manuellement par un expérimentateur.
- Colonne 9 : **eauDelivreeB**
- Contient les valeurs de quantité de solution nutritive B délivrée par le robot, uniquement dans la chambre de culture C3.
- Colonne 10 : **poidsApresArrosage**
- Contient les valeurs de somme du poids avant l'arrosage et de la quantité de solution nutritive délivrée.
- Colonne 11 : **eauDelivreeCumulee**
- Contient la somme cumulée des quantités de solution nutritive délivrées depuis le début de l'expérimentation.

- Colonne 12 : **inhibe**
- Contient 0 si le pot n'est pas inhibé ou 1 s'il l'est. Un pot inhibé n'est pas pris en compte lors d'un cycle d'irrigation de l'automate.
- Doit être vide si la manip a lieu en cabinet de culture.

- Colonne 13 : **poidsEstimePlante**
- Contient la valeur estimée du poids de la plante quand elle est trop importante.

- Colonnes 14 et 15 : **photoV** et **photoIR**
- Contient les noms des images numériques prises dans le visible (photoV) et l'infra-rouge (photoIR, uniquement dans la chambre de culture C3)
- Les noms sont de la forme *CiMj-k-V-aaaammjj-xx.JPG* (photoV) ou *CiMj-k-IR-aaaammjj-xx.JPG* (photoIR) avec *CiMj-k* l'identifiant du pot, *aaaammjj* la date de la prise d'image et *xx* le numéro du cycle. **L'extension peut être JPG ou jpg sans distinction.**

Si les photos sont encore au format Potk-aaaammjj-xx.JPG (Phenopsis 1), il faut :

1/ Mettre les noms des images au format correct dans les colonnes photoV et photoIR en remplaçant *Pot* par *CiMj-* puis en remplaçant *-aaa* par *-V-aaa* ou *-IR-aaa*

2/ Renommer les images concernées en utilisant le script *RenommerImages.R* (téléchargeable sur la page d'insertion de données d'irrigation).

- Si ces colonnes ne sont pas vides, **il faut alors avoir auparavant transféré les images correspondantes sur le serveur (cf Annexes). Seuls les administrateurs peuvent effectuer cette tâche, contactez alors Juliette, Vincent, Myriam ou Christine.**

- **Attention : Toutes** les images listées dans les colonnes *photoV* et *photoIR* doivent avoir été transférées sur le serveur avant de télécharger le fichier CSV des données du robot !

Si des images sont manquantes, votre fichier ne sera pas validé, et vous aurez la possibilité de télécharger les fichiers CSV contenant les noms des images manquantes sur le serveur.

Il vous faudra alors soit transférer ces images sur le serveur, soit les supprimer des colonnes *photoV* et *photoIR*. Pour les supprimer du fichier CSV, utilisez le script *SupprimerImagesNonPresentes.R* téléchargeable lors de la non-validation du fichier (cf Annexes).

- Au contraire, si vous transférez sur le serveur des fichiers dont les noms n'apparaissent dans aucune des colonnes *photoV* et *photoIR*, ces fichiers seront automatiquement supprimés du répertoire TMP après 1 mois.

date	idPotManip	numPot	positionPot	nbPot	poidsAvantArrosage
2007-01-08 08:30:00	C1M16-1	1	P1_A1	2	263.20
2007-01-08 08:30:00	C1M16-13	13	P1_A2	2	260.94
2007-01-08 08:30:00	C1M16-25	25	P1_A3	2	264.54
2007-01-08 08:30:00	C1M16-37	37	P1_A4	2	264.67
2007-01-08 08:30:00	C1M16-49	49	P1_A5	2	276.45

poidsCible	eauDelivreeA	eauDelivreeB	poidsApresArrosage	eauDelivreeCumulee
262.00	1.1	0	264.99	0
264.00	0	0	264.54	0
267.00	2.3	0	266.99	0
279.00	2.5	0	278.99	0

Notes d'insertion des données d'irrigation

idPotManip	numPot	positionPot	nbPot	poidsAvantArrosage	poidsCible	eauDelivreeA	eauDelivreeB	poidsApresArrosage	eauDelivreeCumulee
C1M16-1	1	P1_A1	2	263.20	262.00	1.1	0	264.99	0
C1M16-13	13	P1_A2	2	260.94	264.00	0	0	264.54	0
C1M16-25	25	P1_A3	2	264.54	267.00	2.3	0	266.99	0
C1M16-37	37	P1_A4	2	264.67	279.00	2.5	0	278.99	0

12. Annexe

Pour transférer les images sur le serveur : photos des automates PHENOPSIS, scans de plantes, coupes histologiques, etc :

1/ Pour les images des automates PHENOPSIS, merci de trier auparavant les photos : supprimez les photos noires ou sans pot.

2/ Dans **Filezilla** connectez-vous à Junior : host : sftp://lps-junior, login : granier, mot de passe.

3/ Sélectionnez en bas à droite dans la fenêtre correspondant à votre ordinateur les fichiers que vous souhaitez transférer.

4/ Faites glisser les images en bas à droite dans le répertoire **/home/phenopsis/TMP** en faisant glisser la sélection. Attention, l'opération peut être longue.

5/ Si à la fin du transfert, des photos sont présentes dans l'onglet « Failed transfers », sélectionnez-les, cliquez droit et choisissez « reset and requeue selected files ». Les photos apparaissent à nouveau dans l'onglet « Queued files », cliquez droit et choisissez « Process queue ».

Pour supprimer les images des automates PHENOPSIS listées dans la colonne photoV et non présentes sur le serveur :

1/ Lors de la non-validation du fichier CSV de données d'irrigation, on vous propose en téléchargement :

- le fichier CSV contenant la liste des images dans le visible non présentes sur le serveur,
- le script *SupprimeImagesNonPresentes.R* permettant de supprimer du fichier CSV de données d'irrigation les images dans le visible non présentes sur le serveur.

2/ Téléchargez le fichier CSV et le script R, et suivez les instructions du script pour nettoyer votre fichier CSV de données d'irrigation.