

Compte rendu des discussions / avancées sur la déclaration des variables.

C Pommier, P Neveu, I Alic, F Tardieu

Février 2022

1. Règles "grammaticales"

1.1 Le quadruplet s'écrit avec un _ entre chaque terme du quadruplet, sans blanc à l'intérieur de chaque terme. Si il y a plusieurs mots, utiliser le "camel cases" :

DailyMinTemperature, cf aussi exemples au § 2.

1.2. Texte ASCII uniquement (Alphanumérique minuscule ou majuscules)

- Pas de chiffre en premier
- Pas d'espace, 'camelcase' à la place des espaces
- Pas de signes mathématiquement ou informatiquement interprétable : +, =, /, :, \$, %, ,, <, >, !, -, *, ", #, &, ^, | ; ! [, { , (,) ,] , } , ~
- Pas de ponctuation, en particulier de ' !
- Pas de caractères accentués
- Les noms de variables commencent avec une lettre (pas de chiffre, underscore etc).

1.3. Pour les unités, ne pas utiliser les exposants (Kg m⁻²) ni le "/" (Kg/m2).

Utiliser le PER, KgPERm2

μmol m⁻² s⁻¹ s'écrit donc "micromolPERm2PERs". (Note, ce n'est pas idéal...)

2. Règles pour l'entité

2.1 Cas général: l'entité est l'objet sur lequel la caractéristique est mesurée, on met la plus "petite" entité possible. Une règle serait rentrée dans la machine organe<plant<canopy

Exemples :

- "humidité du grain" : Grain_Moisture_sample_Percent
- "concentration en carbone du grain" : Grain_CarbonContent_calculation_Percent
- "date de levée" : Plant_EmergenceDate_Record_JulianDay
- Leaf area index : Canopy_LAI_ImageAnalysis_Unitless
- NDVI: Canopy_NDVI_Measurement_Unitless
- Note d'aristation: Awn_Type_Expert_Unitless

Notes:

1 Dans le dernier cas, l'entité pourrait être "grain" et la caractéristique "AwnType", mais comme toute ariste appartient à un grain, il vaut mieux mettre "Awn". Cela facilite la cartographie avec des ontologies, "awn" et "type" se trouvent, "awn type" non.

2. Le poids moyen d'un grain est bien grain_weight, à la différence du rendement qui serait canopy_GrainYield (cf point 2.2)

3. A ce stade, on ne renseigne pas l'échelle à laquelle la mesure a été prise CF point 3 : Le quadruplet correspondant à "humidité du grain" est le même (i) si les données se réfèrent à un grain (par exemple pour un tableau de données avec une ligne par grain, collecté dans une plateforme post récolte type Mauguio), ou au grain moyen d'une parcelle expérimentale (par exemple si le tableau de données comporte une ligne par parcelle).

2.2. Cas particulier des "nombre de.. " ou "poids cumulé de...".

Il faut renseigner à la fois la nature des objets comptés et l'objet sur lequel ils sont comptés.

Proposition : l'entité est l'objet sur lequel d'autres objets sont comptés. La nature des objets comptés passe dans la caractéristique.

"Nombre de grains par mètre carré": canopy_GrainNumber_calculation_PERm2

"nombre de grains par épi": Ear_GrainNumber_Manual_Unitless

"Nombre de grains par plante": Plant_GrainNumber_Manual_Unitless

2.3. Règle supplémentaire: Utiliser des noms botaniques, pas des noms attachés à une espèce, exemple "grain" au lieu de "kernel" pour le maïs. Utiliser des entités biologiques, par exemple "grain" et pas "1000grains" pour le poids de mille grain : grain_weight.

3. Prise en compte de l'échelle de mesure (individus dans le tableau considéré)

Une information importante ne se trouve pas dans le quadruplet: l'échelle à laquelle les résultats sont présentés, c'est-à-dire ce que représente chaque individu placé en ligne du tableau de données. Le même quadruplet peut s'appliquer à "humidité du grain" si il s'agit de la moyenne pour les grains d'une micro parcelle (une ligne du tableau de données = 1 microplot) ou si ce sont des valeurs individuelles pour une série de grains (une ligne du tableau de données = un grain). De même une variable peut s'appliquer à un objet (une ligne = 1 plante ou microparcelle), ou à une moyenne génotypique (une ligne = un génotype). Enfin, dans les séries temporelles, les mesures se font à un certain pas de temps, qui définit chaque ligne dans le tableau correspondant.

La règle générale est que la même échelle doit s'appliquer à un tableau de données

Règle pour la capture des données: Les données produites ont vocation à être groupées en tableaux. Il est donc nécessaire de mentionner une échelle afin que les tableaux aient un sens. Cette échelle peut être renseignée au moment de l'acquisition, elle est en générale unique pour une plateforme donnée (par exemple "plante" pour les plateformes conditions contrôlées, "microplot" pour le champ). Cependant, cette règle peut se complexifier à l'avenir, avec distinction de plantes différentes dans une micro parcelle ou de feuilles différentes dans une plante. Dans ce cas, des tableaux "fils" sont constitués, par exemple des tableaux "par feuille" pour toutes les feuilles qui appartiennent à une certaine plante.

Règle pour le partage des données, par exemple sous data.inrae: dans ce cas, les tableaux qui sont inclus dans le dataset ont une métadonnée du type "each line, one day" ou "each line, one genotype".

4. Prise en compte des couverts multi espèce ou multi variété.

Pour une parcelle comprenant une espèce 1 et une espèce 2, les variables peuvent concerner l'espèce 1, l'espèce 2 ou le couvert entier. Par exemple, la surface foliaire du couvert entier n'est pas toujours la somme des espèces 1 et 2 si il y a des mauvaises herbes. Certaines variables ne peuvent être définies que pour le couvert (hauteur, rayonnement intercepté).

Au moins dans un premier temps, on utilise le champ "espèces", en renseignant pour chaque variable, par exemple "Wheat", "Bean" ou "Whole". Ceci demande qu'une espèce virtuelle, "Whole" soit ajoutée à la liste d'espèces ("Whole" pourrait être l'option par défaut si rien n'est précisé).