

Packages importation

1. Data importation

A. Datasets structures

B. Data manipulation

2. Data template

A. Data template: plant_info

B. Data template: endpoint

C. Data template: timeseries

D. ABER data templates

3. Export the data templates in .txt

ABER Data Preparation

Elise

2024-06-08

Set the right working directory.

```
setwd("C:/Users/elise/Documents/Mémoire/Main/Data/Drive/ABER")
```

Packages importation

1. Data importation

The first step in this data preparation process involves importing all the pertinent datasets listed in the Google Sheets “Variables template” document. First we find the files, then import them.

```
## [1] "ABER_Initial Code Draft"      "degree_days.txt"
## [3] "env_data.txt"                 "ISA_EPPN2020_ABER.xlsx"
## [5] "ISATab_a_phe_ZM009.txt"       "root_pixels_lateral_all.txt"
## [7] "root_pixels_lateral_mean.txt" "root_pixels_vertical_all.txt"
## [9] "root_pixels_vertical_mean.txt" "root_ratio_lateral_total.txt"
## [11] "root_ratio_lateral_vertical.txt" "root_ratio_vertical_total.txt"
```

We can extract the coordinates of each plant with the ISA_EPPN.xlsx dataset, using a made-up function “coordinates_isaTAB”.

```
# Get the coordinates
isaTAB <- read_excel("ISA_EPPN2020_ABER.xlsx", sheet = "s_exp")
```

```
## New names:
## • `Unit` -> `Unit...9`
## • `Term Source REF` -> `Term Source REF...10`
## • `Term Accession Number` -> `Term Accession Number...11`
## • `Unit` -> `Unit...13`
## • `Term Source REF` -> `Term Source REF...14`
## • `Term Accession Number` -> `Term Accession Number...15`
## • `Unit` -> `Unit...22`
## • `Term Source REF` -> `Term Source REF...23`
## • `Term Accession Number` -> `Term Accession Number...24`
## • `Unit` -> `Unit...26`
## • `Term Source REF` -> `Term Source REF...27`
## • `Term Accession Number` -> `Term Accession Number...28`
```

```
coordinates <- coordinates_isaTAB(isaTAB)
```

A. Datasets structures

We can take a quick look at all the datasets.

- coordinates
- data
- data_environment

```
head(coordinates)
```

```
##   Sample.Name nrow ncol rep
## 1 ZM009-01711    1    2   1
## 2 ZM009-01712    1   10   2
## 3 ZM009-01713    2    7   3
## 4 ZM009-01714    3   10   4
## 5 ZM009-01715    4    1   5
## 6 ZM009-01716    6    1   6
```

```
head(data)
```

```
##      barcode  X leaf_no shoot_fwt root_fwt shoot_dwt root_dwt shoot_dwt.fwt
## 1 ZM009-01711  9   70.9    52.5    12.3    22.3 17.34838          NA
## 2 ZM009-01712 10   73.6    40.5    28.8    16.2 39.13043          NA
## 3 ZM009-01713  9   75.1    34.0    12.2    10.9 16.24501          NA
## 4 ZM009-01714  8   64.2    46.5    12.1    18.3 18.84735          NA
## 5 ZM009-01715 10   64.0    37.5    11.8    14.5 18.43750          NA
## 6 ZM009-01716 10   59.9    34.4    12.4    16.1 20.70117          NA
```

```
head(data_environment)
```

```
##          logdate c2air c2rh    c2svp    c2vpd
## 1 11/03/2021 00:00  22.1 37.4 2659.515 1664.857
## 2 11/03/2021 00:05  21.7 38.0 2595.412 1609.155
## 3 11/03/2021 00:10  21.4 38.6 2548.224 1564.610
## 4 11/03/2021 00:15  21.0 39.2 2486.474 1511.776
## 5 11/03/2021 00:20  20.6 40.2 2426.037 1450.770
## 6 11/03/2021 00:25  20.2 41.2 2366.887 1391.729
```

B. Data manipulation

This next step standardizes diverse datasets by renaming variables for consistency, converting data into appropriate units, adding necessary columns, and merging the datasets.

```
#####
# COORDINATES
#####
# Unit.ID
coordinates$Unit.ID <- seq_len(nrow(coordinates))
# Reference for Sample.Name et Unit.ID
reference <- coordinates[, c("Sample.Name", "Unit.ID")]
## We can then copy dataset2$Unit.ID <- reference$Unit.ID[match(dataset2$Sample.Name, r
eference$Sample.Name)]
reference$Genotype <- isaTAB$`Source Name`

#####
# DATA
#####
# Time, Date and Timestamp
data$Date <- as.Date("2021-03-11")

# Name of the platform
data$Platform <- "ABER"

# Unit.ID
data$Unit.ID <- reference$Unit.ID[match(data$barcode, reference$Sample.Name)]

# Rename the columns for the template
data <- rename(data,
               Leaf_number = X,
               FW_shoot_g = leaf_no,
               FW_root_g = shoot_fwt,
               DW_shoot_g = root_fwt,
               DW_root_g = shoot_dwt)
```

2. Data template

A. Data template: plant_info

This dataset contains information about the plant: Unit.ID, genotype, replication, row and column location in the greenhouse, and soil treatment.

B. Data template: endpoint

This datasets contains information of the end of the experiment (variables at harvest). It is then linked by the Unit.ID to the plant_info data template.

C. Data template: timeseries

This section in divided in three data templates:

- timeseries
- S_timeseries (variables computed from sideview imaging or image processing)
- T_timeseries (variables computed from topview imaging or image processing)

The time interval between data timestamps varies in each platform. They are then linked by the Unit.ID to the plant_info data template.

D. ABER data templates

- plant_info
- endpoint

##	Unit.ID	Genotype	Soil	Replication	Row	Column	Platform
## 1	1	EPPN1_L	NA	1	1	2	ABER
## 2	2	EPPN1_L	NA	2	1	10	ABER
## 3	3	EPPN1_L	NA	3	2	7	ABER
## 4	4	EPPN1_L	NA	4	3	10	ABER
## 5	5	EPPN1_L	NA	5	4	1	ABER
## 6	6	EPPN1_L	NA	6	6	1	ABER

```
##      Unit.ID Time      Date Timestamp DW_shoot_g FW_shoot_g DW_root_g FW_root_g
## 1      1    NA 2021-03-11      NA      12.3      70.9      22.3      52.5
## 2      2    NA 2021-03-11      NA      28.8      73.6      16.2      40.5
## 3      3    NA 2021-03-11      NA      12.2      75.1      10.9      34.0
## 4      4    NA 2021-03-11      NA      12.1      64.2      18.3      46.5
## 5      5    NA 2021-03-11      NA      11.8      64.0      14.5      37.5
## 6      6    NA 2021-03-11      NA      12.4      59.9      16.1      34.4
##      Leaf_number Plant_height_cm DW_plant_g Root_length_cm Root_number Root_angle
## 1      9          NA          NA          NA          NA          NA
## 2     10          NA          NA          NA          NA          NA
## 3      9          NA          NA          NA          NA          NA
## 4      8          NA          NA          NA          NA          NA
## 5     10          NA          NA          NA          NA          NA
## 6     10          NA          NA          NA          NA          NA
##      Total_wu DW_seed_g FW_seed_g Leaf_area_cmsquared Genotype Soil Replication
## 1      NA      NA      NA          NA      EPPN1_L      NA      1
## 2      NA      NA      NA          NA      EPPN1_L      NA      2
## 3      NA      NA      NA          NA      EPPN1_L      NA      3
## 4      NA      NA      NA          NA      EPPN1_L      NA      4
## 5      NA      NA      NA          NA      EPPN1_L      NA      5
## 6      NA      NA      NA          NA      EPPN1_L      NA      6
##      Row Column Platform
## 1      1      2      ABER
## 2      1     10      ABER
## 3      2      7      ABER
## 4      3     10      ABER
## 5      4      1      ABER
## 6      6      1      ABER
```

3. Export the data templates in .txt

Stock the new data sets in a new folder.

```
setwd("C:/Users/elise/Documents/Mémoire/Main/Data/Templates/ABER")

write.table(plant_info, file = "plant_info.txt", sep = "\t", row.names = FALSE, quote = FALSE)
write.table(endpoint, file = "endpoint.txt", sep = "\t", row.names = FALSE, quote = FALSE)
```