

Info about the data base survey Mix-Enable

Frédéric Vanwindekens, Marie Moerman

October 30, 2019

1 Introduction

This file briefly explain the 3 data base generated from the survey (WP2).

2 Type of data

Regarding the GDPR as applied in Europe, we define consider different type of data that were collected :

1. Personnal data : name, adress, ...
2. Sensisitive data : some open ended question and free text (history of the farm, ...)
3. Other data : factor data or numeric data (interger, ...)

The variable considered in the two first categories are reported below, in R-script format. This information was used for producing two version of the data base.

```
personal_variables <- c("Code_farm",  
  "Name_farm",  
  "Date_of_meeting",  
  "Name_farmer",  
  "Adress", "Zip_cod",  
  "City",  
  "pseudo1")  
  
sensitive_variables <- c("Main_events_at_the_farm",  
  "Overall_strenght",  
  "Diversification_Strenght",
```

```
"Overall_Weakness",  
"Diversification_weakness",  
"Overall_opportunities",  
"Diversification_opportunities",  
"Overall_threats",  
"Diversification_threats"  
)
```

3 Data base

Two main data base were build based on the data collected during the survey :

- A **safe** data base, without personal data and without sensitive data
- A **sensitive** data base, without personal data, but **including** sensitive data

After a final check from the board of the project, the safe data base could become **public** (open access data) and the sensitive data base could be kept **inside** the Mix-Enable project, between partners. (It is a proposal to be discussed between leaders).

As expected by the GDPR, we proceed with the pseudonymisation of the data and no personal data are in the two data base. A third data base with the keys must be kept secret (WP leader, project coordinator) as it allows to make the link between personal data and pseudos used in the data base. This file has to be permanently deleted at the end of the projet.

The three tables can be found in 2 formats :

- **'csv'**, in order to be read by R or other scripts
- **'xlsx'**, in order to be read by human in excel.

The names of this three files are explicit :

- `mixenable_data_base_safe_191030210719`,
- `mixenable_data_base_sensitive_191030210726`,
- `mixenable_secret_keys_pseudo191030210702`.

The last part of each of these names are a **date** and **time** when they have been produced (ex. 191030210702 = 30 nov. 2019, 21h07m02s)

4 Extra scripts

We produced a serie of other script in orther to read and basically treat the main data base. These scripts will be shared among the consortium also.