MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION

*Liberté*
*Égalité*
*Fraternité*

# Modernize the sampling frame to allow frequent updates from administrative data

Thierry GUILLAUME (*)

(*) *Ministry of Agriculture and Food, Service de la statistique et de la Prospective, Bureau des méthodes et de l'informatique statistiques*

Worskhop Budapest October 2023

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
*Liberté*
*Égalité*
*Fraternité*

# Presentation outline

2010 - 2020 findings

The need for a new sampling frame

Greater use of sources

Encouraging initial results

Decision rules that are closer to reality

Reengineering the sampling frame application

Conclusions and short-term outlook

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
*Liberté*
*Égalité*
*Fraternité*

# 2010 – 2020 findings

A sampling frame (called BALSA) initialized with the 2010 agricultural census and updated only with surveys

Problem with surveys: 15% of the field

From 2016, start of use of administrative agricultural databases to update the sampling frame

From 2019, greater use of administrative databases to prepare the 2020 agricultural census

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Why a reengineered application ?

The aim of this project is :

- Having a more frequently and accurately updated sampling universe
- Improving the quality (better identification of farm, update with creations and cessations of units, identification of restructuring, merger or break-up and better quality of the units characteristics, better management of contact information, ……..)
- Reduction of the statistical burden
- a lower burden for managing the sampling frame

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Updating the sampling frame

Databases are numerous and complement each other:

- French **business register** (Sirene) : use of the available API, once a month

- **Bovine identification national database and other animal species databases** : integrate the information on bovine cattle sizes (yearly or monthly), goat and sheep cattle sizes, coupled with information on cattle moves (yearly), pig cattle sizes, coupled with cattle moves (yearly)

- **Common Agricultural Policy subsidies dataset**: integration of information on culture surfaces, and cattle information (bovine, goat and sheep subsidized farming's) (yearly)

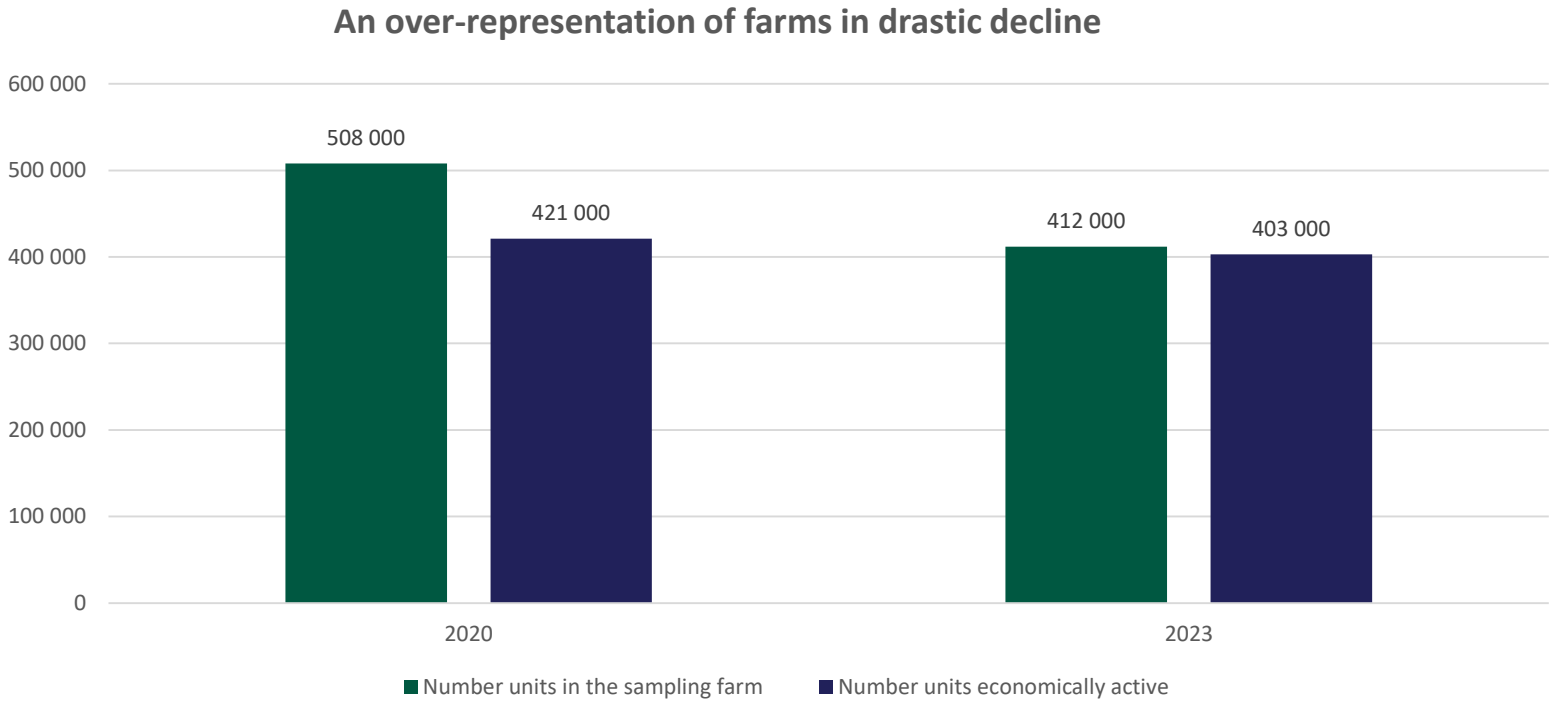
MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Updating the sampling frame

- **customs administration vineyard dataset** (CVI) with use of an additional dataset on wine crops: once a year.
- **Mandatory Social security system for paid and independent Agricultural workers** (Mutuelle Sociale Agricole) : integration of annual information on contributors and paid workers
- **Fiscal information** (benefits and micro-benefits): integration of the dataset (yearly)
- **Other datasets** for activities with poor coverage in administrative registers noted above, such as market gardening, horticulture, organic farming, horse farming and apiaries.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Encouraging initial results




*The number of economically active units (= number of farms) is estimated for 2023*

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Decision rules that are closer to reality

The definition of decision rules for all datasets was tested but due to the specificities of each administrative dataset and the multiple cases of farms, each dataset have its own update decision rules.

To improve the quality, decision rules will allow for the multiplicity of datasets.

Rules about cessation have been defined and allow automatic updates

For bigger units (due to their local economic importance), it is completed by operator's expertise.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Decision rules that are closer to reality

Some results:

During the previous decade (2010-2020), there was a slight increase in the number of units in the database following updates.

Since 2020, there has been a steady decline in the number of farms in the database (-3% between 2020 and 2023), which partly reflects reality (although the real decline is probably more pronounced).

# Better quality samples

Sampling designs require a great deal of stratification.

We used stratification variables (whose data are close to reality) :

- culture surfaces
- number of livestock animals
- number of employees
- composite variables such:
    - standard gross production (SGP), which is a potential of production of farms and allows to classify farms according to their economic size (small, big farms…)
    - technical-economic orientation (TEO) (calculated from SGP), which allows to classify farms according to their specialisation

# Better quality samples

Most of the variables selected for inclusion in the sampling frame are chosen because they meet a number of needs:

- to be a necessary variable for stratifying a survey
- to be a variable used in the calculation of the standard gross production
- to be updatable by administrative databases

The sampling frame is designed so that variables can be added to meet a specific need.

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Reengineering the sampling frame application

Three main changes:

- the use of the ARC (from the French: Acquisition - Réception - Contrôles) software from INSEE.
- the use of a matching engine
- an approach by UX Design

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Reengineering the sampling frame application

- The ARC allows receiving (administrative) data supplied by the providers, to control the compliance of the received files, and to transform administrative data to elementary statistical data.

The software enables to define and apply controls and mappings, to test them in a sandbox environment (linked to the software), and to put them into production without frequently calling on a developer.

ARC was discussed in the ESSnet I3S (Implementing Shared Statistical Services)

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Reengineering the sampling frame application

- the use of a matching engine tool like « Rapsodie » (Reconciliation of Social Data, Taxes and Surveys), which is an application developed by the Tax and Social Income Division (RFS) of INSEE.

**The aim is to draw inspiration from the Rapsodie tool (R scripts) to design our own tool, capable of performing exact and inexact matching and to adapt to the specific needs of businesses/farms.**

Previously, when matching was not exact, we did not make decisions without the expertise of operators. The workload was so great that only a small part of it was carried out.
**The use of scores methods with an automatic matching decision for higher scores should reduce the expertise of operators.**

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Reengineering the sampling frame application

- The approach by UX Design

The aim is to design the operator validation tool with UX (User eXperience)

The operator validation tool will be directly integrated into the sampling frame application, which was not the case in the past

Direct integration should enable real-time updates and provide operators with real-time information for their expertise.

# Conclusions and short-term Outlook

More efficient and regular updating of the sampling frame

Reduction in the integration of new units with no economic reality (entry in a file with no real activity thereafter)

An increase in the number of units leaving the register (to reduce the number of "non-active" units).

Geolocation (Land Parcel Identification System) of Common Agricultural Policy on agricultural plots to identify restructuring

MINISTÈRE
DE L'AGRICULTURE
ET DE L'ALIMENTATION
Liberté
Égalité
Fraternité

# Conclusions and short-term Outlook

Thoughts on a possible calibration of the entire sampling frame with surveys (for example:  pig, goat and sheep cattle sizes)

In the short term, produce a rigorous approach to the demography of farms and agricultural businesses and use the sampling frame to answer specific questions

In the short term, data from the sampling frame could be used to calculate annual agricultural statistics.

# Thank you for your attention

For further information, please contact

damien.babet@agriculture.gouv.fr

magalie.dinaucourt@agriculture.gouv.fr