METHODOLOGY FOR MAINTENANCE OF FARM REGISTER IN ITALY

Roberto Gismondi
ISTAT - Italy

EU Twinning Project IL/12 CRIS 2015/370-467

Long Term Work Plan for Agricultural Statistics

Jerusalem, 12-15 June 2018

Istat

Contents

PART 1 – Definitions, sources, overall methodology

- The use of administrative sources for statistical purposes
- The farm register project
- The administrative sources used to build up the Integrated Base of Administrative Sources (BIFA)
- Description of input data: sources, contents and definitions
- Data pre-treatment: examples

PART 2 – Building up and data validation

PART 3 – The «permanent» agriculture census

PART 1
Definitions, sources, overall methodology

Why to use administrative sources?

Are typically inexpensive / Help the NSI in presence of budgetary constraints
Can have good coverage / Increase response rates
Often the same statistical requests are already provided to administrative programme / Reduce response burden (and costs)

When to use administrative sources?

✓Survey evaluation: benchmarck
✓Variables editing and imputation
✓Improving sample design and estimation
✓Auxiliary information in data estimation
✓Replacement of data collection

Administrative Sources (AS) contain information that is not primarily collected for statistical purposes:

1) Differences in the definition of the units
2) Differences in the definitions of variables
3) Different classification systems or applied differently
4) AS could not be available in time or could related to a different period
5) Data susceptible to political or/and fiscal regulations changes that cause spurious structural breaks
6) Operators can adapt legal procedures on the base of their owns interests (the administrative declaration do not correspond to the operator's "economic reality")
7) Problems of data matching (unit duplications)
8) Consistency among sources

Logical process in order to use administrative data

Definition of the reference conceptual frame
statistical definitions and classifications

Knowledge of the observed universe
(administrative files) referring to coverage, definition of the units and characters, classification used, time and modalities of updating;

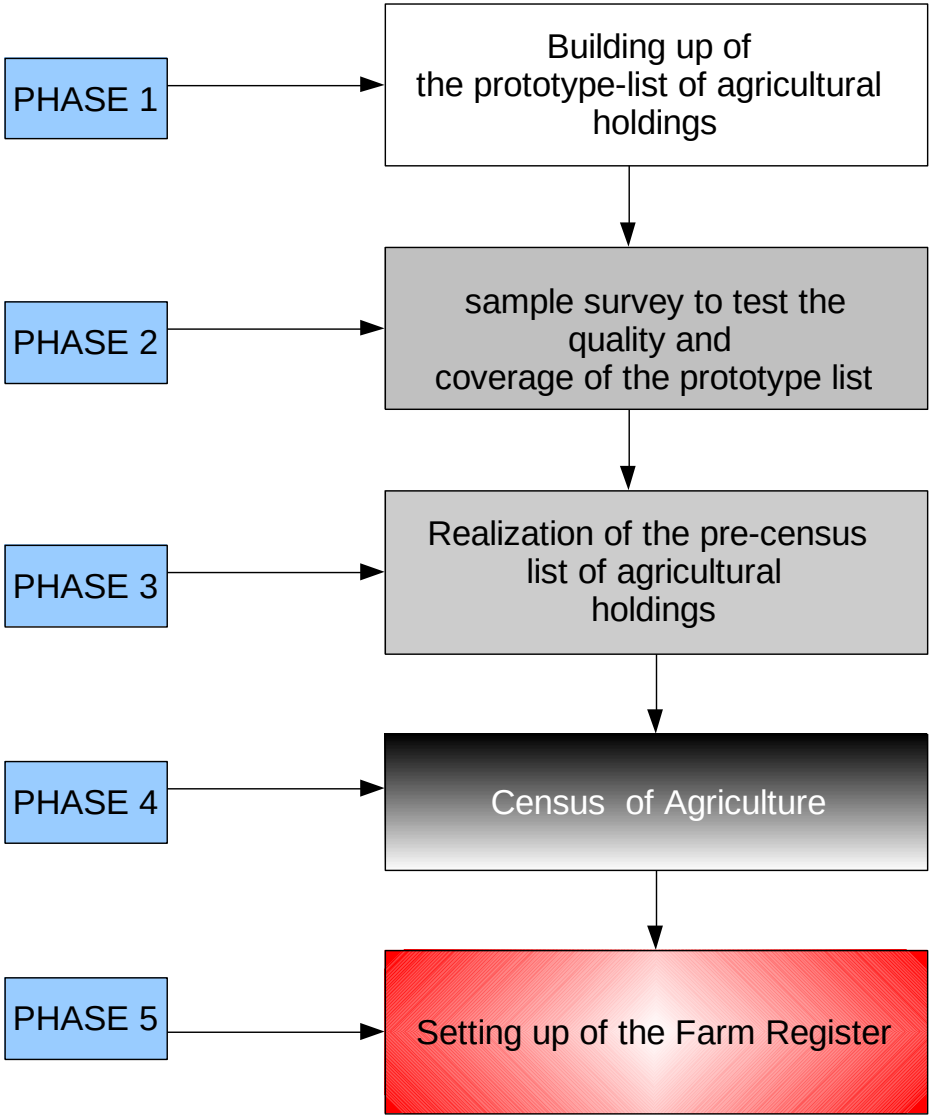
Identification of the MetaData Translator to convert administrative data in statistical ones.

Developing suitable statistical methodologies in order to treat administrative information

Main purposes for setting-up and maintaining a Farm Register

- To enlarge the statistical register system to the sector of agriculture (European Regulations)
- To exploit information available into administrative sources
- To maintain high-quality sampling frames for current statistical surveys conducted in agriculture (random samples, weighting structure of the samples)
- To update the census data between two censuses providing on time information on the current structure of agricultural holdings
- Providing longitudinal analyses on historical trends in the overall structure of agricultural holdings (FRAME approach)

The Project



Istat

Difficulties

Agriculture features

- It is characterized by small and very small productive units, often defined as “small scale food producers”
- Labour force is mainly based on household labour, often seasonal and part-time
- The sector is strictly integrated with other activities such as transformation, trade, tourism, etc..

These items make complex the correct identification of units as well as the estimation of their actual size and their principal activity

Definition of units

At national level as defined by the Census of Agriculture and according to the definition of the FSS regulation n. 1166/2008 (article2), the statistical definition of “Agricultural holding” is*:*

A single unit, both technically and economically, which has a single management and which undertakes agricultural activities listed in Annex I within the economic territory, either as its primary or secondary activity

- Weak connection with the NACE Rev.2 (01.1: non permanent crops; 01.2: permanent crops; 01.3: plants reproduction; 01.4: animals breeding; 01.5: mixed activities; 01.6: activities whioch support agriculture and after harvest activities)
- Not solved problem concerning the threshold to be adopted for deciding whether including or not a unit in the observation field
- No significant changes if compared with the definition to be adopted in the IFS Regulation.

Peculiarity of the available sources of information

Each source uses peculiar definitions and classifications that need to be translated according to a statistical framework before their usage

The main difficulty is to identify easily the statistical unit of reference starting from the units recorded into administrative files

Why Integration of many administrative sources?

- ✓ Lack of a unique benchmark or pivot archive that covers the whole potential universe of agricultural holdings.
- ✓ Information gathered for different purposes such as: the payment of EU contributions (IACS); the identification and registration of animals for public health reason (AA.ZZ.); the income declaration for tax reasons.
- ✓ Recorded variables in each source can present a different degree of complexity in the statistical translation.

- ❑ Beyond administrative sources, current statistical survey can be used as well for building up the register

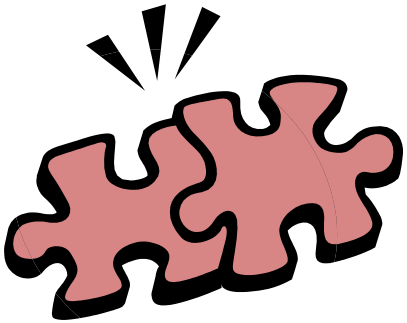
INPUT: Administrative and statistical sources

Administrative sources managed by public agencies or private companies

Statistical sources owned by Istat

LIST FRAME: SFR records all the eligible agricultural holdings and their main structural characteristics by integrating information coming from both sources

AREA FRAME: SFR records lands and livestock linked to holdings



The SFR production and updating process

- **Standardization** of the tax identification codes, territorial variables, units of measurement, etc.
- **Identification** of the holder/farm according to the statistical definition
- Identification and **classification** of the variables related to holder/farm
- Identification of **duplicated** or out of target units, ceased or with duplicated land
- **Quality** checks on outliers e missing values
- **Metadata** analysis

Specific sectoral sources

1) The Integrated Administration and Control System (IACS) –
2) Lands' property Incomes (Tax Agency)
3) The System for the Identification and Registration of Bovine Animals and other species (AA.ZZ.)
4) The Land registry

General sources

5) Chambers of Commerce (CCIAA); a) Business register of Italian Chambers of Commerce (Legal units and primary and secondary economic activities) b) partners-partnership archive
6) Fiscal register (VAT)

Statistical sources

7) PDO (Protected Designation of Origin) and PGI (Protected Geographical Indication) survey,
8) the Farm Structure Survey and other current surveys
9) Business register ASIA
10) Census of Agriculture 2010

1) The Integrated Administration and Control System (IACS)

- The database IACS is managed by AGEA
- AGEA is the Italian Paying Agency set up in accordance with the EC n. 885/2006 that, under the common agricultural policy, acts in the coordination and execution of payments to support farmers
- The Integrated Administration and Control System has been realized in order to record, verify and control data
- The core of this system is made of files containing information on data that each agricultural holding is obliged to present for any aid application
- That file (called "*fascicolo aziendale*") is constituted when for the first time a request of subsidium is presented, then information must be taken updated on the basis of the following yearly requests of financial support

Main definitions using the AGEA files

(a) '*farmer'* means a natural or legal person, or a group of natural or legal persons whose holding is situated within the Community territory and who exercises an agricultural activity

(b) '*holding'* means all the production units managed by a farmer situated within the territory of the Member State

(c) '*agricultural activity*' means the production, rearing or growing of agricultural products including harvesting, milking, breeding animals and keeping animals for farming purposes, or maintaining the land in good agricultural and environmental condition

Advantages

AGEA data are of primary importance:

- to estimate the agricultural area by crops
- to locate farms on the territory

The unique ID code (CUAA) identifies the holding and solves problems of links with different sources having a system of units identification based on fiscal codes

Information content (1)

Identification data:

Unique Code of Agricultural Holdings (CUAA code), e.g. the fiscal code of the holder, holder's name, permanent address or place of residence, VAT number if present, dates of inscription and updating

Territorial data:

Agricultural parcels managed by the farmer, use of each parcel (crops, livestock), hectares invested by type of product (cadastral area and agricultural area utilized for farming), fiscal code of landholders if different from the holder and the type of contract that links farmer to landholder, dates of start and end of conducted parcels

Information content (2)

Building data:

agricultural parcels, type of use of building (stable, warehouse, home), covered surface, owner's and user's fiscal codes

Labour force data:

Full time and part time agricultural labourers, paid employees

Agricultural machinery data:

Type of used machines

Coverage

Each agricultural holding having relations with the PA and that by law should set up the "*fascicolo aziendale*" independently from the type of cultivation

Undercoverage

Coverage is partial, both in terms of recorded crops (there are no aids for some growing crops like orchards) or/and in terms of units, in fact it is possible that not all the holders ask for a financial support

Overcoverage

There are many cases of fascicoli aziendali present in the IACS system but which do not correspond to active units

2) The land's property income (RA)

- Database managed by the Tax Agency
- It covers those tax payers that for a given tax year declare income because of property (landlord's) rights (cadastral income) and/or carry out agricultural activities (agricultural income)
- 4 taxpayers kinds have to fill according to their law status

Type of model	main section	other sections
Natural persons	income from agricultural activities, landlords' rights livestocks	1. participation in partnerships or others 2. dependent relatives
Simple partneships Non profit institutions	income from agricultural activities, landlords' rights holidays farm	
Income earned by employees, retired persons	agricultural income, by declarant and his spose	

Natural persons

Have to declare an income coming from any agricultural activity carried out on lands, or/and income coming from the landlords' rights

Juridical persons

Simple partnerships, non profit institutions

Information content

- ✓ Identification data concerning taxpayer (fiscal code, fiscal address, name)
- ✓ property title
- ✓ percentage of property
- ✓ amount of income (both income deriving from the agriculture activity carried out on the land and income deriving from the landlord's rights)

Coverage

Under-coverage

Some kinds of organizations, like limited liability companies or cooperatives, are not included in this universe as they produce business income (so they have to fill in different models)

Over-coverage

Because of a very big difference in terms of unit definition in comparison to the statistical one, more than one administrative unit can represent only one agricultural holding and therefore over-coverage can occur

For example both the owner and the tenant of the same land have to make tax declaration as well as two or more co-landowners but only one is the holder according to the statistical rule.

Over 3 million tax payers declaring agricultural incomes

3) The System for the Identification and Registration of Bovine Animals and other species (AA.ZZ.)

Data managed by the Ministry of Health

The national database of Bovine Animals realized in accordance with the EC n. 1760/2000 and EC n. 21/2004 for the setting up of a System for the Identification and Registration of Bovine Animals, Sheep and Goats is the only fully operative register up to date

No regulations for others species

Information content

- The farm livestock identification code
- The holder's and owner's fiscal code
- Fiscal address, name, telephone,
- Number of animals (only for bovines, sheep, goats, pigs),
- Dates of beginning and closing activity

- Number of farm livestock units

1) Bovines (monthly data) – two deliveries per year, 180 thousands records
2) Sheep and goats (punctual data) = 130 thousands records
3) Pigs (punctual data) = 144 thousands records

4) The Land registry (LR)

Data organization and content:

- File parcels
- File ownership
- File owners

Specific identification codes to link these file

File parcels

Parcels identify a portion of land. Contents:

- ❑ Parcel identification codes
- ❑ Parcel characteristics like “soil quality” (i.e. land devoted to vineyard or arable land)
- ❑ Hectares
- ❑ Estimated agricultural income

Each land is univocally identified by a unique identification code (inside the municipality)

File ownership

Specifies the link between land and its owner. Contents:

- ❑ Land's rights (property, usufruct)
- ❑ Property share and property settlement

File owners

Natural person:

- Name
- Surname
- Sex
- Birth date
- Fiscal code

Juridical person:

- Name
- Fiscal code
- VAT code

Advantages

✓Full coverage in terms of lands, but risks of over-coverage in terms of units (holdings)

Problems

✓Enormous variety of information difficult to treat
✓The parcel's owner cannot be the holder
✓Updating is not homogenous
✓Soil quality could not be updated
✓Difficult to identify agricultural uses only
✓Difficulty to intercept modification of land in time series

General sources

5) Chambers of Commerce
Selected only those units carrying out an agricultural economic activity, principal or secondary (section A divisions 1 or 2 of Nace rev.2)

6) VAT declarations
Only units having a VAT code to carry out agricultural activity

These sources cover mostly business units

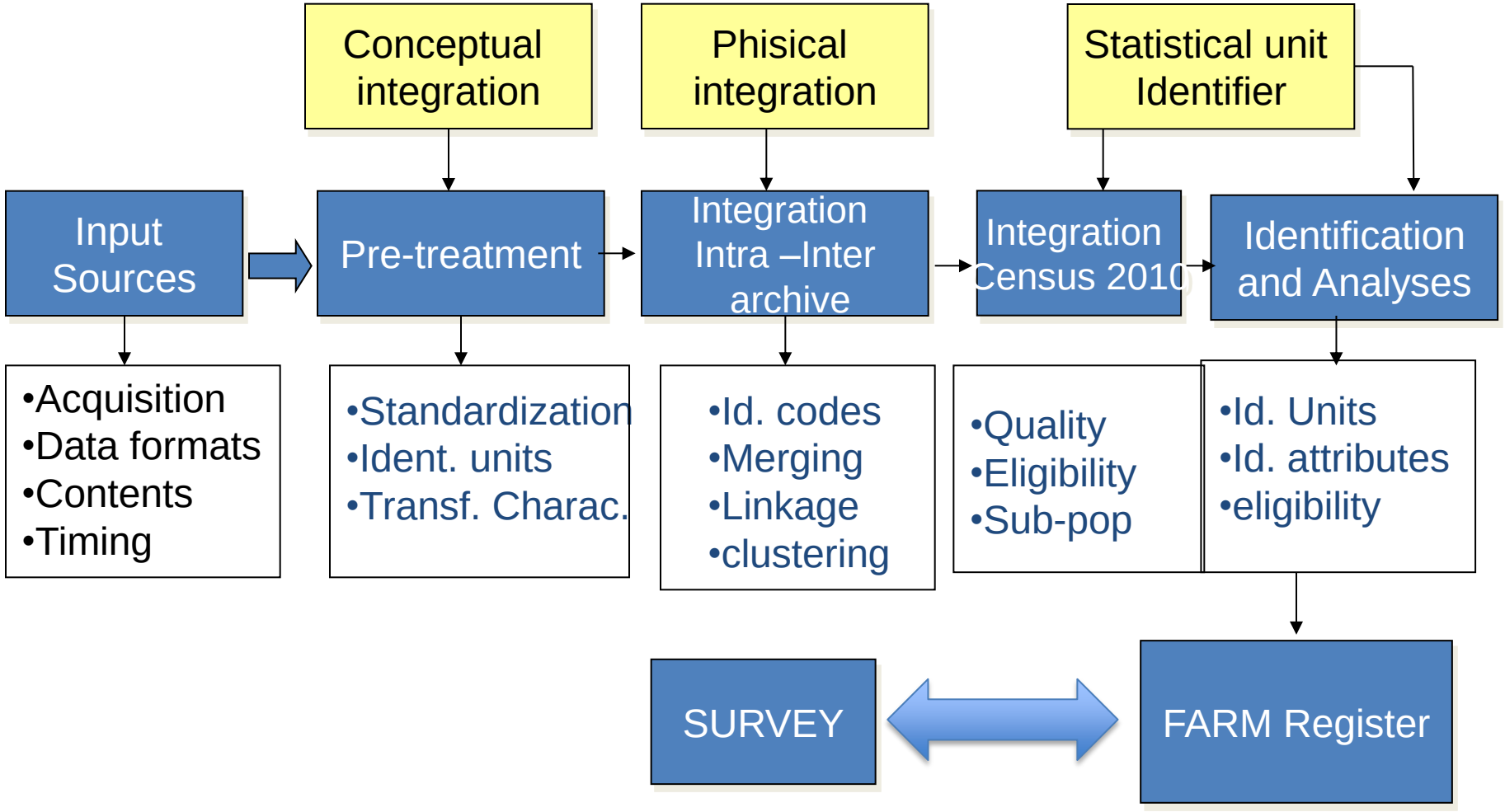
Overlapping between the farm register and the business register can happen

Characteristics of sources – units and coverage

Source	Administrative unit	Statistical unit compliance	Coverage	
			Theory	Real
IACS	Farmer or agricultural holding	High	T	P undercoverage
AAZZ	Livestock holder	High	Ts	Ts
Land's income	Natural/juridical person	Low	T	T overcoverage
Chambers of Commerce	Enterprise /local unit	High	P	P
VAT	Legal unit	High	P	P
Land registry	Parcels' owner	Low	Ts	Ts overcoverage

T = Total
Ts = Total for the reference population
P = Partial

Production Process – Overview



EXAMPLE OF PRE-TREATMENT

- **Check of duplicated lands (intra–archive linkage)**
- **Linkage of lands (inter-archive) between the land registry and IACS - decomposition of cadastral area by owner in land registry**
- **Riconciliation of the land use between Census (or other updated sources) and IACS (classification table)**

PART 2
Building up and data validation

CHARACTERISTICS OF THE ITALIAN STATISTICAL FARM REGISTER

- Administrative/Statistical databases in input: **9 + Census 2010**
- Linkage variable: **ID of the holder**
- Output: **3 databases linked among them**
  - ✓ database of holders/farms with the related main structural data *(Legal form, UAA, Total, area, main crops area, livestock by species, labour force, etc.)*
  - ✓ database with personal data of the holders *(name of the holder, address, postcode, telephone number, email, etc.)*
  - ✓ database with the land parcels of the holders/farms
- **Legal references**

  a) PSN 2018-2020 (IST-02466)

STEP 1

IACS

Aggregate all the parcels and land use by ID of the holder

STEP 2

IACS

Link the dataset with the holder personnel information with the dataset of the land use by ID

Identify the ID without land

Exclusively animal farms?
Not agricultural units?

STEP 3

Animal Register

Aggregate all the animals by ID of the holder

STEP 4

Aggregate the information by ID of the holder for all the other databases

STEP 5

Physical integration of the database by ID of the holder

Databases	**Statistical units**
IACS	1.915.077 *of which 1.696.282 with land*
Animal Register (AA.ZZ)	344.390
Land's property income (RA)/Cadaster (Land Register)	3.030.560
VAT on agricultural income/Chamber of Commerce	969.063
Agritourisms	18.891
Farms with PDO/PGI products	54.509
Agricultural census	2.120.289

OUTPUT: Integrated database includes 4.853.608 rks

Output of the STEP 5

Each record by ID with all the useful information of the databases
Integrated database including:

- ID
- Data to be used for the eligibility rules (land use, livestock, sources, etc.)

Moreover, we can build the matrix containing 1 or 0 depending on the presence of each unit in the various original databases
It is useful for eligibility rules as well

STEP 6

Specific elaborations on the integrated database

- Assignment of the territorial codes of the estimated farm headquarter
- Choice of the land use by main crops in the SFR *(IACS or AC or Cadaster)*
- Assignment of a flag on the quality of the land use *(comparison between ICAS and census)*
- Choice of the livestock by species in the SFR *(Animal Register or AC)*

STEP 7

METHODOLOGY FOR SELECTING THE ELIGIBLE UNITS

38 rules to assign the eligibility or the non-eligibility to the unit according to:

- ✓ the database where the unit comes from *(presence in one or more sources, some sources are considered more reliable then others)*
- ✓ characteristics of the unit *(if the unit is active or ceased in the database or if reaches some physical or economic threshold)*

The units are divided in 4 lists

1. IACS and Animal Register
2. Land's property income and Cadaster
3. Chamber of Commerce and VAT on agricultural income
4. Agricultural census and statistical list of Agritourisms and PDO/PGI farms

Cascade selection of the units

Example: Eligibility rules for the list 1 (deterministic)

Rule	Specification	ELEGGIBILITY
1	unit coming from IACS without duplicated parcels	YES
2	unit coming from IACS with duplicated parcels	NO
3	unit coming ONLY from IACS without parcels	NO
4	unit coming ONLY from IACS without parcels and census	NO
5	unit coming ONLY from IACS without parcels and CCIAA/TVA	NO
6	unit coming ONLY from IACS without parcels, census and CCIAA/TVA	NO
7	unit coming from animal register with livestock	YES
8	unit coming from animal register without livestock but coming also from IACS or CCIAA/TVA or tax/cadaster or DOP/IGP or agritourisms	YES
9	unit coming ONLY from animal register and CA, with livestock for self consumption or already included or ceased in the business register	NO
10	unit coming ONLY from animal register and CA, without livestock	YES
11	unit coming ONLY from animal register without livestock	NO
12	unit coming ONLY from animal register, IACS without parcels and CA, without livestock	NO
13	unit coming ONLY from animal register and IACS without parcels, without livestock	NO

Main kind of information in the Italian Statistical Farm Register

- ID code
- Territorial codes of the farm headquarter
- Territorial codes of the holder residence
- Flag of the databases where the unit comes from
- Legal form
- Land use by main crops
- Flag on the quality of the land use
- Livestock by species
- *Labour force (to be estimated)*
- Flag on farm typology
- Eligibility
- Flag on the selection rule
- Standard output
- Enterprise / Not enterprise

DATA VALIDATION

- Identification of familiar and corporate clusters
  - ❑ Two members of the same family can register the same land in two different databases
  - ❑ The farm land can be split in two by the family members in IACS

- Micro checks with the main comparable surveys (FSS 2013 and 2016,Orchard 2012 and 2016, livestock survey, sowing forecasts, Annual Crop Statistics, FADN, ….)

- Macro and micro comparison with other administrative and statistical sources

- Sample survey for evaluating the quality of the statistical Farm Register

DATA VALIDATION

AREA SAMPLE SURVEY

- Selection of a sample without starting list (all units found within a certain territorial area)
- Selection of an area which can represent a specific profile among the various national areas (ex.: mainly based on small scale producers)
- Interviewing one by one all units found in the area by expert interviewers
- Comparing the area survey results with the Farm Register information in order to assess under-coverage of the register and discrepancy between the data captured for the same unit in the two sources for each variable investigated
- Potential over-coverage estimation as well

DATA VALIDATION

FARM REGISTER vs SURVEY A

- Selection of a sampling survey A which observes the variable y (we label observed sample data as y_A) and uses sampling weights w for grossing up sample data to the universe. The estimated total is T_A
- Linking one by one sample units in survey A with the Farm Register, which reports from variable y as well (say, y_{FR}). The estimated FR total is T_{FR}
- 3 datasets:
  - Matched units (A and FR)
  - Non matched units only in A (A not FR)
  - Not matched units only in FR (FR not A)
- Using sample weights w we can estimate over-coverage and discrepancy decrees

DATA VALIDATION

- The sum of *w* weights of units (FR not A) is W_{FR}
- The sum of *w* weights of units (A not FR) is W_A
- The difference W_{FR}-W_A is the Farm Register over-coverage

For the matched units (A and FR):

- The sum of products $y_{FR}w$ is Y_{FR}
- The sum of products $y_A w$ is Y_A
- The difference Y_{FR}-Y_A is the decree of discrepancy due to different reported y values

For the not matched units:

- The sum of product $y_{FR}w$ is Y_{FR*}
- The sum of products $y_A w$ is Y_{A*}
- The difference Y_{FR*}-Y_{A*} is the decree of discrepancy due to over-coverage and different reported y values

PART 3
The “permanent” agriculture census

THE PERMANENT CENSUS 2021-2029

The **Farm Register** will be the basic pillar for agriculture structural data production within the permanent census framework

Year	Source	Variables
2020	**General Census**	**About 600**
2021	Farm Register	26 in FR + n in FR2
2022	Farm Register	26 in FR + n in FR2
2023	Farm Register + FSS	26 in FR + n in FR2 + FSS
2024	Farm Register	26 in FR + n in FR2
2025	Farm Register	26 in FR + n in FR2
2026	Farm Register + FSS	26 in FR + n in FR2 + FSS
2027	Farm Register	26 in FR + n in FR2
2028	Farm Register	26 in FR + n in FR2
2029	Farm Register	26 in FR + n in FR2

Data released after two years from the reference year

The number of variables released will be lower with respect to the 2020 census

THE PERMANENT CENSUS 2021-2029 – The Extended FR

The project – FR2

- FR2 means «Extended Farm Register
- Additional variables to be added to those already included in the Farm Register.
- Basically new variables concern production input and output:
  - Labour cost (Inps-DMagr) and input of labour typology
  - Economic account (Unico PF, SP – various kinds of Balances for big farms and cooperatives)
  - Import-export (COE)
  - Investments and production tools (IACS, Ministry of Agriculture, Health Ministry)

THE PERMANENT CENSUS 2021-2029 – The Extended FR

The project – FR2

Tentative planning

1. April 2018: feasibility study about sources which may be used and about methodologies for integration with the FR
2. Within December 2018: preliminary version of FR2 including the main variables related to economic account: valued added, incomes, costs of goods and services, (through Fiscal declaratons and Balances), labour costs (Inps: Social security); Import-export (Foreign trade)
3. Estimation of data concerning small-scale food producers
4. From 2019, feasibility study concerning integration od IACS data AGEA on production tools/machineries and investments