



# Progress Report (for period: 09/2022 – 08/2023)

# Action title: Administrative data sources for agricultural statistics

#### Modernisation of agricultural statistics

Grant Action Number	101036501
Grant Action Acronym	2020-CZ-AGRI
Starting date	01/09/2021
End date	31/08/2024
Duration in months	36
Call (part) identifier	ESTAT-2020-PA8-E-AGRI
Topic	B5476-2020-AGRI

**Contact person:** Radek Matějka (radek.matejka@czso.cz)

#### 30 August 2023

This document has been produced with the financial assistance of the European Commission. The views expressed herein are those of the author and can therefore in no way be taken to reflect the official opinion of the European Commission.

### **Table of contents**

Table of contents	2
Introduction & Short summary of interim results	
Tasks covered by the second interim Progress Report	
Description of work done	
- Animal production statistics	
- Crop production statistics	
- Follow-up in the Farm Register	
Overall evaluation & Further work envisaged	

#### **Introduction & Short summary of interim results**

The Czech Statistical Office (CZSO) is implementing a Grant Action (further called also a project) aimed at modernizing the system of agricultural statistics by establishing and implementing a sustainable system for the regular use of data from administrative data sources (ADS).

The primary aim of the project is to reduce the response burden of agricultural units and rationalize available resources while maintaining the high quality of agricultural statistics. Another equally important goal is to establish new or deepen existing cooperation between CZSO and other state institutions that own and administer the administrative data sources, which are to replace existing data collection systems - statistical sample surveys.

Project activities are divided into 3 Objectives:

Objective A (Tasks A.1 – A.7): Mapping and analyses of ADS

Objective B (Tasks B.1 - B.6): Identification and specification of requirements for a sustainable system for the use of administrative data

**Objective** C (**Tasks** C.1 - C.4): Establishing and implementation of a sustainable system for export and integration of administrative data

The second interim Progress Report describes the present status of the Grant Action and activities realized in the period from September 2022 to August 2023.

- Short summary of interim results:
- In the area of crop production statistics, the use administrative data from IACS (LPIS, IS SAIF) reduced the administrative burden on respondents; the initial sample was reduced from 10 thousand to 350 reporting units.
- In the area of animal statistics, the *Livestock Survey*, which served as a national source for updating the Farm Register, was ceased. The administrative burden decreased by 10 thousand reporting units. The Farm Register will be updated from administrative sources (Integrated Animal Register, IAR) in the future.

Tasks A.1 to A.6 - finished

Tasks B.1 to B.6 – started in 2022 and still ongoing

Tasks C.1 to C.4 - started

### Tasks covered by the second interim Progress Report

Task No.	Task name	Description	Animal statistics	Crop statistics
A.5	Updating of ADS	<ul> <li>Description and frequency of the ADS updating process.</li> <li>The timeliness of recording changes to ADS and their impact on data quality.</li> <li>For crop production statistics: Assessment of data processing and publishing systems (with or without regular data revisions).</li> <li>For animal statistics: Setting of a sufficient degree of data updating for provisional and/or final data, the alignment of reference date with time limits as set out in the legislation in force.</li> <li>Determining the optimal date for data extraction from ADS to match data quality requirements and deadlines laid down by the legislation in force.</li> </ul>	<b>V</b>	V
В	Requirement formulation	A follow-up procedure of formulating requirements based on analyses described in Objective A:  • for external working group (state institutions owning the administrative data sources)  • for internal working group (involved departments of the CZSO)	ongoing	✓ ongoing
B.1	Requirements for external working group	<ul> <li>Proposal of requirements for the supplementation and modification of data collection systems for IS SAIF (administration of subsidies) and IAR based on Objective A analyses.</li> <li>Defining the requirements for ADS data usage: structure and description of the data record, data format. Determination of a suitable method for data usage.</li> <li>Contract on the data usage and on the provision of access to micro-data.</li> </ul>	☑ ongoing	<b>☑</b> ongoing
B.2	Output table preparation	<ul> <li>Defining the structure of output tables according to the requirements of national data users' needs and of the legislation in force.</li> <li>Mapping of the CZSO and ADS code list items.</li> <li>Setting/programming of aggregate items in output tables.</li> <li>Basic checks setting/programming.</li> <li>Prevention of duplicated records.</li> <li>Basic data checks between LPIS (main area) and IS SAIF (individual crop area).</li> <li>Data cross-validation against data recorded in the previous period.</li> <li>Setting rules for the confidential data protection.</li> </ul>	☑ ongoing	ongoing

B.3	Preparation of adjustments to the Statistical Metainformation System	Preparation of adjustments to be implemented in the CZSO Statistical Metainformation System (SMS), related to the use of ADS (code lists, definitions, data processing, output tables, preparation of the system for the data use from ADS, modification of the statistical survey system).	<b>☑</b> ongoing	☑ ongoing
B.4	Preparation of the solution for missing items	<ul> <li>Identification and proposing other data sources for missing variables / items.</li> <li>Determining the suitable data source for missing variables / items:</li> <li>Supplementary sample survey - elaboration of methodology.</li> <li>Expert estimation of missing variables / items - elaboration of methodology.</li> <li>Modelling of missing variables / items - elaboration of methodology.</li> </ul>	<b>☑</b> ongoing	☑ ongoing
B.5	Methodology for updating the Farm Register	Elaboration of the methodology for updating the Farm Register (kept by the CZSO) with data exported from ADS.	<b>☑</b> ongoing	<b>☑</b> ongoing
B.6	Update of the methodology for linked statistical surveys	Modification and extension of the methodology for linked statistical surveys in the crop production statistics ( <i>Harvest Estimates, Final harvest survey</i> ) and animal production statistics (surveys on cattle production, pig production and poultry production).	<b>☑</b> ongoing	☑ ongoing
B.7	Study visits	Exchange of experience, studying the methodology and good practice applied in the implementation of pilot data exercise in more experienced Member States.	×	X
С	Implementation	Simultaneous implementation of sample statistical surveys in full and use of data from administrative data sources.	<b>☑</b> started	<b>✓</b> started
C.1	Pilot data processing	<ul> <li>Export of data from administrative data sources and their processing according to the developed methodology.</li> <li>Preparation of output tables.</li> </ul>	<b>☑</b> started	<b>☑</b> started

C.2	Comparative data analysis from the two data sources	<ul> <li>Data analysis.</li> <li>Evaluation of differences between the two data sources at the level of statistical units and aggregated data in the structure of output tables.</li> <li>Evaluation of possible non-compliances (time series interruption) due to source changes.</li> </ul>	<b>☑</b> started	<b>☑</b> started
C.3	Quality assessment of data from administrative sources	Overall quality assessment of data from administrative data sources (availability, coverage, completeness, punctuality and coherence, a comparison with outputs from crop and animal statistics) of data coming from the administrative data sources.	<b>☑</b> started	☑ started
C.4	Final decision	Final decision on the use of administrative data sources.  4 options for follow-up scenarios:  • full replacement of statistical surveys for crop and animal statistics in the statistical system  • partial replacement of statistical surveys for crop and animal statistics in the statistical system  • continuation of the system of statistical surveys  • update of the Farm Register - support for the selection of basic and sample sets for statistical surveys in crop and animal statistics	<b>☑</b> started	☑ started

#### **Description of work done**

#### - Animal production statistics

The discussion on the content, form and dates of the data transmissions with the Ministry of Agriculture continued. Based on the data analyses, a break in the series can be expected due to the use of administrative data sources; for cattle numbers, the break is assumed insignificant, but for sheep and goats, the data series may be interrupted. For this reason, the verification of an alternative method combining a sample statistical survey with the use of data recorded in the IAR was launched.

#### **A5 Updating of ADS**

A was noted in the first interim Progress Report, the time schedule for updating records in the administrative data source (ADS) affects the completeness of the statistical data, because the quality of extracted data can differ depending on the time lag between the reference date and the extraction date. Optimization of the timeline "reference date - extraction date - publication date" with the reference date between 1 November to 31 December 2022 was repeatedly tested.

Data on the state of IAR were extracted at on average one-month long intervals for the reference days 1/11/2022, 30/11/2022 and 31/12/2022. Based on the analysis of changes in data between consecutive extractions, the progress in IAR updating for cattle, sheep and goats was revealed. The three reference days were considered as repetitions of the extraction schedules.

#### **Results:**

Based on the comparison of the withdrawn datasets, suitable reference dates and IAR extraction dates were determined to meet SAIO's data publication requirements:

#### For cattle

- **preliminary** data to be published no later than 15 February N+1
  - o reference day: 31 December of year N
  - o optimal date for IAR extraction: 31 January N+1
  - O According to the analysis, on average **80.1% of changes** (77.7% to 82.7%), which could affect the cattle numbers at a specific reference date, were recorded in the IAR **within 1 month** following the reference date (where 100% = number of changes recorded within 6 months following the reference date)
- **final** data to be published no later than 15 May N+1
  - o reference day: 31 December of year N
  - o optimal date for IAR extraction: 31 March N+1
  - O According to the analysis on average **94.1% of changes** (94.1% to 94.2%), which could affect the cattle numbers at a specific reference date, were recorded in the IAR **within 3 months** following the reference date (where 100% = number of changes recorded within 6 months following the reference date).

o To improve the results, the processing time could be shortened and the date of the IAR extraction can be postponed to 30 April N+1 (4 months after the reference date).

#### For sheep

- **final** data to be published no later than 15 May N+1
  - o reference day: 31 December of year N
  - o optimal date for IAR extraction: 31 March N+1
  - o According to the analysis, on average **77.8% of changes** (71.5% to 83.7%), which could affect the sheep numbers at a specific reference date, were recorded in the IAR **within 3.5 months** following the reference date (where 100% = number of changes recorded within 7 months following the reference date).
  - O To improve the results, the processing time can be shortened and the date of the IAR extraction can be postoponed to 30 April N+1 or 15 May N+1 (4 or 4.5 months after the reference date). Alternatively, the reference date can be moved to an earlier date and thereby extend the interval between the reference date and the IAR extraction date.
  - o Results for sheep were less consistent than for cattle.

#### For goats

- **final** data to be published no later than 15 May N+1
  - o reference day: 31 December of year N
  - o optimal date for IAR extraction: 31 March N+1
  - o According to the analysis, on average **70.7% of changes** (62.1% to 77.6%), which could affect the sheep numbers at a specific reference date, were recorded in the IAR **within 3.5 months** following the reference date (where 100% = number of changes recorded within 7 months following the reference date).
  - O To improve the results, the processing time can be shortened and the date of the IAR extraction can be postoponed to 30 April N+1 or 15 May N+1 (4 or 4.5 months after the reference date). Alternatively, the reference date can be moved to an earlier date and thereby extend the interval between the reference date and the IAR extraction date.
  - o Results for sheep were less consistent than for cattle nd sheep.

#### **B1.** Requirements for External working group

1. As part of task A1, the transferability of ADS data into SAIO variables was addressed (direct trasfer, by derivation, by modeling) and factors that limit this transferability were identified. As part of task B1, these factors were further analysed (for details see the first interim Progress Report). Possible ways of their elimination were found in cooperation with the owner and administrator of IAR.

A meeting of the CZSO and the owner (Ministry of Agriculture, MoA) and the administrator (Czech Moravian Breeders' Corporation) of the animal register was held in November 2022, where the limiting factors identified in previous analyses were discussed. The MoA provided the CZSO with a list of agricultural farm excluding the inactive enterprises. The list was compiled as of 15 December 2022 when applying an internal MoA algorithm, where the

criterion for inactivity was no report received from the given enterprise during the last 12 months.

#### Results of the analysis of the list:

- satisfactory results for cattle
- in a situation where the sources (statistical survey x animal register) are comparable, i.e. without major breaks in the series for the main categories, the exclusion of inactive enterprises did not result in significant changes in data
- less satisfactory to unsatisfactory results for sheep and goats
- with large differences between sources (as described in the first interim Progress Report) there was no mitigation of the differences in data when eliminating inactive enterprises

The whole process of excluding inactive enterprises and monitoring its effect on animal numbers extracted from the animal register will, after making slight adjustments in the algorithm, be repeated on the reference date of 30 June 2023.

These findings were followed up by activities falling under task B6 (Update of the methodology for linked statistical surveys)

- 2. Based on the meeting, the MoA made internal interventions in the algorithm for determining the category of *dairy cows* and *other cows*. These adjustments were reflected in the reduction of differences between the data resources in numbers of both cow categories as proven in a comparative analysis made as of 31 December 2022.
- 3. The CZSO was also involved in the amendment of the Decree No. 136/2004 Coll., which sets out the list of obigatory data to be recorded in the animal register. The amendments to the decree will affect the quality of data on parturitions of females (distinction of cows x heifers; ewes x lambs). The changes are effective from 1 June 2023 (Decree No. 130/2023 Coll.).

## **B2.** Output table preparation & **B3.** Preparation of adjustments to the Statistical Metainformation System

Preparatory works for the modification of the Statistical Metainformation System to be able to receive data from the animal register and to compile new aggregations for the animal codebook corresponding to the categories in the animal register were ongoing during the period evaluated by this Progress Report.

#### **B.4 Preparation of the solution for missing items**

Models were designed and programmed to produce the missing categories of sheep and goats. In the period evaluated by this Progress Report, the numbers of these animals were modeled at the level of the Czech Republic (NUTS0) and NUTS2.

#### **Results:**

The use of auxiliary or supporting data from the IFS (Integrated Farm Survey) is considered the most suitable model. IAR allows sheep and goats to be divided into categories according to sex, and with applying selected attributes the females can be further divided into milk and non-milk sheep/goats. The proportion of breeding females (ewes and ewe-lambs put to the ram) should be taken from the IFS results and would be valid for the entire period until the next IFS.

#### **B6** Update of the methodology for linked statistical surveys

#### Sheep and goats

Due to the unsatisfactory results of the first round of processing data selected according to the activity/inactivity of agricultural enterprises, an additional, combined method for determining the numbers of sheep and goats was planned.

#### Combined method description:

- Sample survey, stratification by enterprise size group, specialization of the agricultural production (production sub-area), administrative region to which the enterprise belongs
- Microdata from the IAR (only available for some of the categories required by the legislation)
- Sample weighting coefficients will be used to estimate the total populations of sheep and goats
- Categories, which are unavailable from the IAR, will be modeled (see A1, B1)

Testing of the combined method on 2022 data started in the monitored period and is still ongoing. Templates for modeling were prepared.

#### Problems that were faced:

• To link the sample file from the Farm Register and the data file from IAR, it was necessary to find a common identifier. Unlinked units were tested for non-systematic errors blocking the connection.

Results of the first, testing round of the combined method (with 2022 data) are expected to be available by the end of 2023.

In 2024, a second round of testing is planned to verify the combined method, using the 2023 data (data as of 31 December 2023) and supporting data from the IFS 2023.

#### - Crop production statistics

## B2. Output table preparation & B3. Preparation of adjustments to the Statistical Metainformation System

In 2022, the following stages took place:

- The Central subsidy crop code nomenclature was modified.
- Items necessary for the crop production statistics were added.
- An option to tick the attribute *purpose of growing the crop* (for grain/green/seed/other product) was added to the declaration of crops (in the subsidy application).
- A converter (mapping) between items listed in the Central subsidy code nomenclature (323 items) and the CZSO's crop nomenclature (82 items) was defined.
- For each crop or group of crops, a default (prevailing) value of the attribute *purpose* of growing the crop was defined.
- In LPIS, a special output called *Report on sown areas* was programmed.

• A pilot verification of these new functionalities set in LPIS was carried out.

In the period evaluated by the second interim Progress Report, the following activities were carried out:

- The Ministry of Agriculture updated the Central subsidy crop code nomenclature according to the new CAP subsidy rules.
- CZSO has carried out a detailed analysis of items of the Central subsidy crop code nomenclature and the assigned the attribute *purpose of cultivation* (for grain, for green, for seed, other products).
- Items and definitions of CZSO's crop nomenclature (SMS) were accordingly modified.
- The mapping (converter) of the Central subsidy crop code nomenclature and crop production statistics items was checked and updated on the basis of a detailed analysis of the results of the pilot project implemented in 2022.
- The structure of output tables was defined according to the requirements of the Regulation (EC) No 543/2009 concerning crop statistics, future requirements of the SAIO regulation and according to the needs of national data users.
- Aggregation items in output tables were defined.
- Basic checks of data were programmed including the prevention of duplicate records.
- A cross-validation against data recorded in the previous period.
- Adjustments to be implemented in the CZSO's SMS related to the use of administrative data sources (code nomenclatures, definitions, data processing, output tables), were prepared.
- Preparation of the SMS for the use of data from administrative sources.
- Modification of the statistical survey system.

#### **B4** Preparation of the solution for missing items

List of items missing in IACS, which are needed for the crop production statistics, was defined in 2022. It was decided that the most appropriate source for collecting missing items is a supplementary standard statistical survey. The sample included units that are not included in IACS (i.e. they did not apply for subsidies) or grow special crops, which are not currently subject to subsidies in the Czech Republic (e.g. Christmas trees, flowers or seeds and seedlings). A total of 350 units were included in the supplementary survey.

#### **B6** Update of the methodology for linked statistical surveys

A methodology for crop production statistics was developed and is now published on the CZSO website: https://www.czso.cz/csu/czso/crop-production-statistics-methodology

During the year, the sample survey *Areas under Crops Survey* (survey of sown areas and main areas of crops) is followed by the *Harvest Estimates* and the survey on final harvest data (*Final Harvest Survey*).

Data from the administrative source (IACS) have much wider use; they will also be used in the statistical survey of IFS (sown areas and main areas - agricultural land use).

#### $\mathbf{C}$

On the basis of very good results of the pilot project carried out in 2022 (for details see the first Prgress Report, task A.6), it was decided that the second pilot project, where data would be obtained simultaneously from a sample statistical survey and from the ADS, would not be repeated in 2023. Thus, for the first time, data collection exercise in 2023 was only done by extracting data from the IACS (LPIS, IS SAIF) and combined with a supplementary statistical survey.

#### C.1 Pilot data processing & C.2 Comparative data analysis from the two data sources

Instead of a second pilot project - simultaneous data collection from two sources - only administrative data sets were extracted, but at different dates. After detailed analyses of data sets for various extraction dates, a schedule was set for the next period, starting from 2024 onwards.

Output tables for linked statistical surveys (*Harvest Estimates* and *Final harvest survey*) were prepared in SMS.

All activities planned in Task C.2 were carried out in the period evaluated by this Progress Report:

- Data analysis data sets from different extraction dates.
- Evaluation of differences between various data extractions at the level of statistical units and at the level of aggregated data in the structure of output tables.
- Evaluation of possible non-compliances.

#### C.3 Quality assessment of data from administrative sources

In the annual crop production statistics the overall assessment of the quality of data from administrative data sources was realized to evaluate its availability, coverage, completeness, punctuality and coherence.

#### C.4 Final decision

When using ADS for agricultural statistics, there are 4 options envisaged for follow-up scenarios:

- full replacement of statistical surveys
- partial replacement of statistical surveys
- continuation of the system of statistical surveys
- only for updating the Farm Register support for the selection of basic and sample sets for statistical surveys

The final decision on the use of administrative data sources has so far been determined for sown areas and main areas:

- Replacement of statistical surveys with data from administrative data sources IACS (LPIS, IS SAIF) for data on sown area and main area was decided based on the results of pilot data processing in 2022.
- Statistical surveys will continue to be used only as supplementary statistical surveys for units not included in administrative data sources and for units growing special crops (as mentioned above).

#### - Follow-up in the Farm Register

Ongoing work carried out during the period evaluated by this Progress Report:

- Program procedures for updating the Farm Register with data from the Ministry of Agriculture were integrated into the software for operating the Farm Register.
- A database with unit identifiers was created, which serves as a converter between the
  individual identification codes used in the registers of the Ministry of Agriculture (i.e.
  LPIS ID of the user, SZR ID code for IAR an IS SAIF) and the identification code
  DKOD used in the Farm Register.
- With the database for converting different unit codes, it is possible to clearly determine
  which units in the Farm Register will be updated with data from the MoA, which units
  are new and which units should be marked in the Farm Register as no longer valid.
  Units with an ID number can be processed automatically. Units without an ID number
  must be processed with human intervention and each case must be assessed
  individually.
- The work on expanding the number of updated attributes depending on the requirements of plant and animal production statistics and IFS is ongoing.
- When updating the Farm Register with data from the animal register, it is necessary to
  evaluate individually whether also data on the age of the animals should or should not
  be incorporated.

#### Overall evaluation & Further work envisaged

- In the period evaluated in this second interim Progress Report, the pace of work followed the planned timetable.
- Cooperation with other institutions (External working group) was excellent and further negotiations will continue.
- A framework agreement on the transfer of administrative data for statistical purposes between the CZSO, the MoA and the SAIF is still under preparation.
- Within 60 days following the end of the Grant Action a Final Report will be submitted to Eurostat.

#### Follow-up work planned in the next phase of the Grant Action:

• Modernisation of agriculture statistics – an international workshop to be held on 2-3 October 2023 in Budapest.

#### In animal statistics

- The negotiation between the CZSO and the MoA will continue to find a sound solution for factors limiting the use of IAR data for statistical purposes (inactive farms, overaged animals, missing mandatory and non-mandatory data) and to increase the quality of data and the completeness of the animal register.
- Simultaneous use of statistical survey and processing of data from the IAR, supplemented by modeling of unavailable categories (Tasks C1, C2).
- For cattle: Simultaneous implementation of sample statistical survey *Questionnaire on cattle production* and use of data from IAR data extracted as of 31 December 2023.
- For sheep and goats: The sample design combined with IAR data extraction as of 31 December 2023 and supported with IFS results for modelling categories unavailable in IAR.
- A comparison of data from both sources during pilot processing and quality checks.
- A plan for sample statistical survey of livestock production (*Questionnaire on cattle production*) for the year 2025 will be prepared, where the section on cattle numbers will already be replaced. Following the change in the dates when the preliminary and final numbers of cattle will be known, it will be necessary to adjust the schedule of supplementary statistical survey of animal production and productivity (Task B6).
- A second round to verify the combined method (sample survey based on the Farm Register and data from the IAR) (Task B6).
- Continuous adjustments in SMS, according to the need and applicability of verification processes (Task B3).

#### In crop statistics

- If needed, editing and updating the Central subsidy crop code nomenclature of MoA and the converter (mapping) for crop statistics items,
- implementation of any changes in the subsidy administration system of the MoA into the CZSO SMS.
- Editing and programming the SMS for two statistical surveys (*Harvest Estimates* and *Final harvest survey*) linked to *Areas under Crops Survey*: output tables, schedules, data processing.
- Seting up the methodology for linked statistical surveys (*Harvest Estimates* and *Final harvest survey*) for sampling in the Farm Register updated with IACS data.
- Data presentation at CZSO website (public database).
- Use of IACS geo-data for geo-referenced presentation of sown areas and main areas.
- Use of IACS data for IFS 2023.
- Data source for permanent grasslands within crop statistics wil be sought to provide data divided into permanent pastures and meadows, permanent rough grazings and permanent grassland no longer used for production purposes and eligible for the payment of subsidies according to the SAIO regulation requirements.

- Verification of the available data sources:
  - Data from IFS.
  - Data from IACS (IS SAIF) subsidies linked to treatment of permanent grasslands.
  - SAIF project data geospatial/satellite data, analysis of existing SAIF projects, definition of intensive/extensive permanent grasslands

#### • Use of data from the Orchards Register

- Define the structure of the data to be extracted from the IACS (LPIS):
  - unambiguous identification of the subject,
  - identification of a part of the *Farmer's block* (DPB),
  - cultivated fruit species (according to the Central subsidy crop code nomenclature of the MoA and the CZSO SMS),
  - the area of fruit species,
  - the number of fruit trees/bushes,
  - fruit production.
- Implementation of the obtained data into the CZSO's Farm Register.
- Using the orchard data in the crop statistics and for updating the Farm Register.
- Regular extraction of micro-data will be ensured in the framework agreement between the MoA and the CZSO.

#### • Organic farming according to SAIO Regulation

- Using data from the IACS IS SAIF and LPIS (*organic farming* attribute).
- Harmonisation of data sources for crop statistics and IFS.
- Ensure the transfer and processing of data in the structure required by SAIO regulation and according to the needs of the MoA.
- A framework agreement ensuring transfer of data and its further use.

#### Work on the Farm Register

- Ongoing work on expanding the number of updated attributes in the Farm Register, depending on the requirements of crop and animal statistics and IFS.
- In animal statistics: Continual evaluation of whether to incorporate data on the age of animals into the structure of the Farm Register or not when updating the Farm Register from the IAR.