The Integration of Administrative Data for the Identification of the Ownership of Agricultural Land

**Keywords:** agricultural land ownership, administrative data sources, geospatial modelling, data integration, geospatial data sources

# Introduction

Ownership structure of the agricultural land is one of the indicators for functional and justified orientation of agricultural policy-making as agricultural holdings (AH) often rent agricultural land due to insufficient own land. Despite adequate administrative registers in Slovenia, this is not explicitly captured in a single data source. Therefore, proportion of rented utilised agricultural area (UAA) was formerly estimated using statistical questionnaires. In the scope of the 2020 Agriculture Census (AC), administrative data source modelling process was developed that will substitute survey questions. The presented work is a case of integrating geospatial and tabular administrative data sources and a step towards a more complete exploitation of readily available data.

# Methods

## Used administrative data sources

**Land Parcel Identification System (LPIS)**

It is a polygon based ESRI shapefile managed by the Ministry of Agriculture, Forestry and Food (part of Integrated Administration and Control System – IACS database). Polygons represent (subsidized) agricultural plots (AP). The attribute part contains information on land use and an identifier of subsidy receiver for each unit (LPIS, 2019).

**Land Cadastre (LC)**

It is a polygon based ESRI shapefile managed by the Surveying and Mapping Authority of the Republic of Slovenia (Ministry of the Environment and Spatial Planning). It is the official register with different layers available, depending on the purpose of use. We used a layer that continuously covers the area of the whole country where polygons represent informative borders of land parcels (LP). In Attributive table, LPs identifiers are included but not owner IDs (Ferlan & Vugrin, 2013).

**Register of Agricultural Holdings (RAH)**

It is an administrative source managed by the Ministry of Agriculture, Forestry and Food (MKGP, 2011). Information on AP subsidy receiver and AH household members (15 or more years old) was used.

**Real Estate Register (RER).**

It is a multipurpose administrative data collection managed by the Surveying and Mapping Authority of the Republic of Slovenia that reflects the actual state of real estate in nature. A real estate can either be a parcel, a parcel with a building, a building or a part of a building. It contains identifiers on real estate and their owners (Ferlan & Vugrin, 2013).

## Process

The data integration process ran in two phases: geospatial modelling and tabular analysis (Figure 1). In the former, geospatial data sources were intersected. In the latter, LP owner and AH user IDs were compared. Ownership of an LP was assigned to an AH if any of the AH household members (<15 years old) was the LP owner. It was followed by data aggregation and data validation.

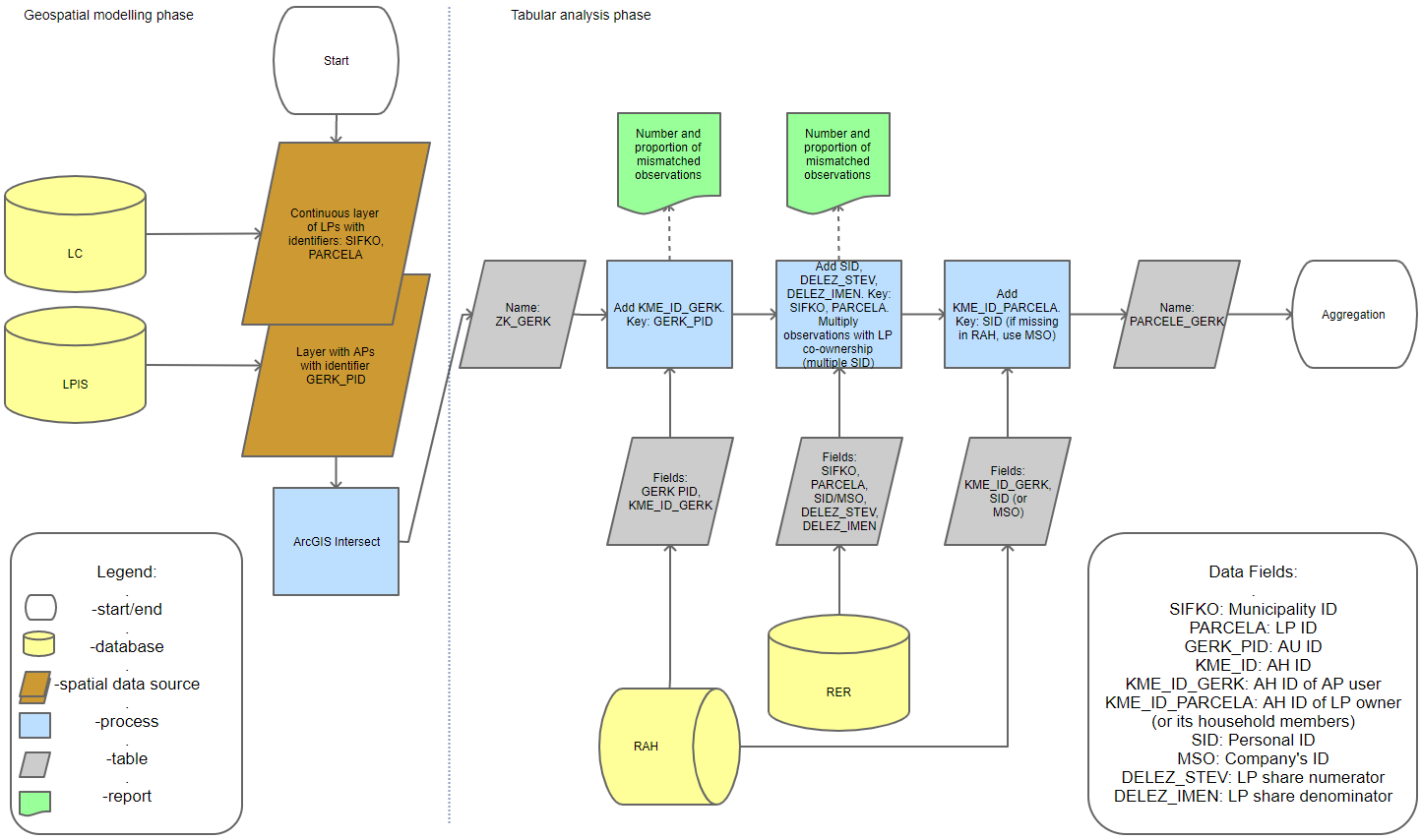


Figure 1. Process workflow

**Aggregation**

In the aggregation step, three sequential aggregations of rented UAA were made: 1) aggregation by LP, 2) aggregation by AP and 3) aggregation by AH, using seven different aggregation methods (Table 1). Different aggregation possibilities arose due to joint LP ownership and multiple LPs (with different owners) on one AP.

Table 1. Aggregation methods*[[1]](#footnote-1)*

|  |  |  |  |
| --- | --- | --- | --- |
| **Ownership share** | **LP Minimal share** | **LP Major share** | **LP Real share** |
| **AP Minimal share** | Aggregation 1 | Aggregation 4 | *omitted (illogical)* |
| **AP Major share** | Aggregation 2 | Aggregation 5 | *omitted (illogical)* |
| **AP Real share** | Aggregation 3 | Aggregation 6 | Aggregation 7 |

**Validation**

Our aim was to match the model data as closely as possible to self-reported data. The results of different aggregation methods were validated through comparing to historical data (AC 2010 – full survey, Farm Structure Survey - FSS - 2013 and FSS 2016 – sample surveys) as ground truth (due to lack of more recent data). Despite expected discrepancies, we assumed aggregation that provided results closest to the FSS data to be the most correct. It is not plausible that the renting structure during the gap period changed in such a way to make this assumption false.

Results were compared on aggregate level, differences at the AH level were accounted for using two *Indicators*. For each AH, only observations with data on both model and FSS were taken into account.

Two variations of the *Indicator* were calculated:

1. Sum of absolute differences of total UAA (*IndicatortotalUAA*) does not depend on the model but estimates differences between LPIS and FSS data:
2. Sum of absolute differences of rented UAA (*IndicatorrentedUAA*) to estimate the comparability of reported to model rented UAA.

# Results

Table 2. Difference between proportion of rented UAA (as calculated with the model) and proportion of rented UAA (as collected with AC/FSS questionnaires).

| **Model\Ground truth** | **AC 2010** | **FSS 2013** | **FSS 2016** |
| --- | --- | --- | --- |
| **Aggregation 1** | 0.063 | 0.043 | 0.040 |
| **Aggregation 2** | 0.141 | 0.121 | 0.118 |
| **Aggregation 3** | 0.166 | 0.146 | 0.143 |
| **Aggregation 4** | 0.084 | 0.064 | 0.061 |
| **Aggregation 5** | 0.162 | 0.142 | 0.139 |
| **Aggregation 6** | 0.186 | 0.166 | 0.163 |
| **Aggregation 7** | 0.198 | 0.178 | 0.175 |

Compared to all three FSS data, discrepancies between proportions of total rented UAA were the smallest in the case of aggregation 1 by a relatively big margin (Table 2). According to aggregation 1, relative differences in rented UAA are 6%, 4% and 4% compared to AC 2010, FSS 2013 and FSS 2016, respectively.

Table 3. Indicator total UAA.

| **Model\Ground truth** | **AC 2010** | **FSS 2013** | **FSS 2016** |
| --- | --- | --- | --- |
| **LPIS data** | 0.255 | 0.207 | 0.146 |

As expected, the more recent the ground truth data the smaller the average differences in total UAA on individual AHs (Table 3).

Table 4. Indicator rented UAA

| **Model\Ground truth** | **AC 2010** | **FSS 2013** | **FSS 2016** |
| --- | --- | --- | --- |
| **Aggregation 1** | 0.199 | 0.227 | 0.207 |
| **Aggregation 2** | 0.215 | 0.229 | 0.206 |
| **Aggregation 3** | 0.228 | 0.232 | 0.212 |
| **Aggregation 4** | 0.204 | 0.224 | 0.204 |
| **Aggregation 5** | 0.224 | 0.234 | 0.214 |
| **Aggregation 6** | 0.237 | 0.237 | 0.219 |
| **Aggregation 7** | 0.246 | 0.242 | 0.224 |

Aggregation 1 proved to be closest to the 2010 AC data. For 2013 and 2016 FSS data aggregation 4 is ahead of aggregation 1 by a small margin (Table 4).

Aggregation 1 was chosen for future use since it was closest to historical data (except for small margin difference on AH level comparison in case of sample surveys). It implies that if at least one part of the plot is at least partially owned (or in co-ownership) by subsidy receiver (or one of their household members), the AP is considered owned and not rented.

# Conclusions

The used administrative data sources represent slightly different time cross sections. For development we chose versions that are less than a month to each other (June-July 2019). The choice of the versions in the production phase has to be done according to the survey methodology (reference date).

Model data are, compared to the ground truth data, of higher precision, allowing also land use differentiation. Accuracy was assessed through presented comparisons. Compared to the FSS 2016 data (most recent), there is in total 4% more rented UAA. Comparisons reflect that more recent FSS data are closer to the model data. This indicates that the discrepancies are a consequence of changes in ownership and renting structure over time rather than a consequence of a model error. Apart from real changes in structure, there are several possible reasons for discrepancies. (1) AC and FSS had a threshold that was not accounted for in the model (full capture). (2) For the testing phase, only household members that were 15 or more years old were accounted. These data are readily available in RAH, while household members of all ages need to be identified from additional data sources. This is planned for the production phase. (3) In the case of deceased LP owners, the legal inheritance procedures are often lengthy. Therefore, the legal state (registers) and perceived ownership by the managers (self-reported data) may differ. (4) Additional reason for discrepancies is survey error in AC and FSS.

Using the procedure described above we managed to make a full capture of a state for a reference date. It was developed for the needs of the AC 2020 but the results can be reused in other areas of agricultural statistics.

Most time and effort went into development and testing, while re-running the process is less work intensive. This enables us to make a full capture quickly and efficiently at any given time to obtain recent data without creating respondent burden.

Utilisation of administrative data ahead of surveys is what statistics aims for; geospatial data sources were, however, not commonly used before in agricultural statistics. Although extracting information from a single geospatial data source can be done by tabular analysis, geospatial modelling requires specific software and knowledge. In order to fully utilise available data, it needs to become an integral part of the statistical data collection process.

# References

1. LPIS technical description. https://www.sinergise.com/sites/all/themes/sinergise/images/LPIS-Technical\_description-WEB.pdf (visited Oct 2020)
2. M. Ferlan, M. Vugrin, Javne nepremičninske evidence: Gradivo za strokovni izpit – geodetska stroka, Inženirska zbornica Slovenije, Ljubljana (2013).
3. MKGP, Aplikacija za vodenje Registra kmetijskih gospodarstev: Priloga 1 k Priročniku o vodenju registra kmetijskih gospodarstev (2011). <http://rkg.gov.si/GERK/documents/RKG_prirocnik_julij_2010.pdf> (visited Oct 2020)

1. *Minimal share:* if at least one part of LP/AP was owned, it was counted as own. *Major share:* if at least one half of LP/AP was owned, it was counted as own. *Real share:* real share of LP/AP ownership was used for aggregation. [↑](#footnote-ref-1)