

Article

Georeferenced Agricultural Data for Statistical Reuse

Katalin Tóth 

European Commission, Joint Research Centre, Directorate for Sustainable Resources, Food Security Unit—D5, CAPLAND Group, TP266, via E. Fermi 2749, I-21027 ISPRA, VA, Italy; katalin.toth@ec.europa.eu; Tel.: +39-03-3278-6491

Received: 13 April 2018; Accepted: 15 May 2018; Published: 20 May 2018



Abstract: The guidelines to the Public Sector Information (PSI) Directive states: “opening up public sector information (PSI) for reuse brings major socioeconomic benefits”, which has been recognised in various domains. However, the reuse may be limited due to organisational and technical reasons. This study addresses the collaboration between the statistical and the agricultural domain using the example of the Integrated Administration and Control System (IACS) and the Integrated Farm Statistics (IFS). After the comparison of the spatial data requirements in IACS and IFS, a conceptual collaboration model was developed that makes clear how the challenges of interoperability can be resolved by technical arrangements and work organisation.

Keywords: agricultural statistics; IACS; interoperability; collaboration model

1. Introduction

The directive on the re-use of public sector information (Directive 2003/98/EC, known as the “PSI Directive”) [1] and its amendment [2] encourages the member states (MS) of the European Union (EU) to reuse existing documents held by the public sector. It states that, apart from a few exceptions, all content under national access should be reusable beyond the original scope under non-discriminatory conditions. The guidelines to the PSI Directive [3] specify that “opening up public sector information for reuse brings major socioeconomic benefits. Yet, studies [. . .] show that industry and citizens still face difficulties”.

The INSPIRE Directive [4] identifies missing metadata, unsatisfactory discovery services, unclear licensing and a lack of interoperability as main obstacles of sharing spatial information in the context of policies that have an environmental impact. During more than 10 years of its existence, INSPIRE made good progress in overcoming these obstacles within the scope of the 34 data themes defined in its annexes. The technical solutions and the good practices elaborated in the frame of INSPIRE have been picked up by other thematic communities too. The implementation of the Common Agricultural Policy (CAP) as well as European statistics are examples of thematic areas that have already benefited from the achievements of INSPIRE.

The management tool of CAP, the Integrated Administration and Control System (IACS), is operated by accredited national bodies in every MS. It consist of six subsystems:

- System for identification of beneficiaries (farmers’ register);
- Systems for identification and registration of entitlements;
- Identification system for agricultural parcels;
- Aid application and payment claims;
- System for identification and registration of animals; and
- Integrated control system.

The implementation of IACS is financed from budgetary resources. Therefore, data from IACS are potential subjects of PSI, even though some information may be excluded by virtue of protection of personal data or commercial confidentiality [1].

IACS widely uses geospatial data, comprising orthoimagery, the Identification System of Agricultural Parcels (a.k.a. Land Parcel Identification System (LPIS)), the geospatial aid application of farmers, which is a georeferenced delineation of agricultural land declared for EU support and the observations and measurements of authorities that they perform in the course of controlling the farmers. Comparing the data content of IACS and especially the LPIS subsystem with INSPIRE shows that there is a good match in concepts of land cover and land use, area management and statistics. Statistical units were included in Annex III of the INSPIRE Directive [5]. In addition, administrative units that are widely used in statistics have been also harmonised. Consequently, INSPIRE may play an important role in establishing interoperability between agricultural statistics and IACS.

European agricultural statistical data are aggregated from the input of the national statistical offices according to the specifications of the Farm Structure Survey (FSS) established by the Council Regulation (EC) No 571/88 [6]. FSS will be replaced by the Integrated Farm Statistics (IFS). The related proposal for a Regulation of the European Parliament and of the Council [7] is in the final stage of the legislative process. Introducing IFS contributes to the modernisation of the European agricultural statistics as “Information on crop area and land use [. . .] is vital for effective policy planning and designing interventions to fully realize agriculture’s potential strengths” [8].

According to EUROSTAT, in 2016 the degree of covering the agricultural statistical demand with IACS data varied between 0% and 52% in the MS. In Switzerland, the administrative registers of agriculture supply more than 90% of data needed for agricultural statistics [9]. The highly regulated character of IACS is a strong enabler for reusing its data in national and European statistics [10]. Nevertheless, a lack of interoperability between the two systems persists due to the legacies of business processes and the management of the information systems.

Moreover, limited discovery facilities create further obstacles for reuse. Metadata on agricultural statistics, as a rule, are available online on the websites of national statistical offices and EUROSTAT. The situation is quite different with IACS. Intended for internal use in the paying agencies data residing in this system, it receives much less publicity. In the frame of the LPIS quality assessment process, the European Commission requires the delivery of a standardised set of metadata, but these metadata are not provided for public use.

The main data collection method of agricultural statistics are surveys. Depending on the targeted survey population, they may be full (census), or sample-based surveys. Surveys consume a lot of resources, such as designing frames and samples, defining production processes and workflows, building collection instruments, setting up collections, just to mention a few from the components that are listed in the Generic Statistical Business Process Model [11]. Additional resources might be consumed by the limited availability of some respondents and by handling errors in their answers.

The most powerful enabler of reusing geospatial information are spatial data infrastructures (SDI), as they put in place the necessary arrangements for efficient data sharing in legal and technical terms. In addition to the general and transparent rules of data access, they also specify technical solutions for metadata, data and web-services. Use of standards plays a key role. On the one hand, a successful SDI manages to build on existing standards, while on other hand it also contributes to a standardization process setting up and sharing registers and registry services. According to Oosterom and Lemmen, “it is important to reuse existing standards as a foundation and to continue from that point to ensure interoperability in the domain” [12].

With the advances of information and communication technology, many statistical organisations introduced service-oriented architecture to build up their organisational structure and production processes, building up a statistical infrastructure technology [11]. The Global Strategy to Improve Agricultural and Rural Statistics [13] outlines a conceptual framework, which translates policy issues into statistical language by identifying the need for the survey framework to link the

farm as an economic unit, with the household as a social unit, and with the land they occupy in the natural environment. This holistic approach can be implemented by developing an underpinning infrastructure.

The infrastructural approach is also applied in IACS. By the force of law, the subsystems have to be integrated, which means that they share common components and data. Such integration excludes the duplication of data collection and storing.

Even though the main spatial component of IACS, the LPIS, is not included in the annexes of the INSPIRE Directive, it has followed its main principles. Besides ISO TC 211 standards, the IACS domain model [14] uses INSPIRE schemas as its foundation. It was a conscious decision that anticipates its possible insertion in the national and European SDIs.

All in all, the subsystems of IACS can provide valuable information for agricultural statistics, which can be used for:

- filling gaps of surveyed data;
- building up statistical frames;
- pre-fill survey questionnaires for respondent farmers;
- quality control (consistency check) of statistical data;
- calibrate survey estimates;
- replacing surveys with administrative data; and
- georeferencing.

The “administrative data” term in statistics refers to third-party data, which come from public administration. Table A1 in Appendix A summarises the results of analysing the national methodological reports of the FSS [15] for 2013. It shows that with the exception of two of them, the MS of the EU extensively use IACS for producing European statistics. The most popular use cases were building up and updating statistical frames (17 reported such use) and controlling and validating statistical data (20 reported such use).

Administrative data sources are particularly well suited in those cases where the need for data is permanent, as their usage requires an initial investment only that pays off later during its continuous use. Such an approach also allows the creation of more frequent output at the expense of a limited marginal cost [16] or the replacement of samples with full populations.

“Because of the fundamental relationship between agriculture and land, the geospatial aspects of land should be seen as an element of the scope of agricultural statistics” [13]. Even though IACS could be a source of other, mainly socioeconomic variables, (e.g., farmers, their income, or indicators of rural development), this paper will show how the geospatial components of IACS can be reused in the forthcoming IFS, setting the focus on standard-driven interoperability of the information systems.

2. Methods

As the infrastructural approach is applied both in statistics and IACS, it was logical to stay inside this framework when researching the modalities of reusing IACS in agricultural statistics. As the FSS is being phased out, this research targets at reusing IACS data in the IFS.

Developing formal collaboration models between different domains is a powerful method for clarifying similarities and differences between domains [17]. A collaboration model can be prepared when formal models that use the same modelling constructs exist for each of the domains. A model previously developed by Tóth and Kučas [14] was applied for IACS, but for the IFS the first attempt at formal model development is performed in the context of this study.

The initial step in this work was a comprehensive analysis of the related legal text. Methodologies for translating legal requirements into technical construct have been applied in various domains, such as marine policy [18], agriculture [14] and conformance testing [19]. For the IFS model the Generic Statistical Information Model (GSIM) [20] and the Statistical Data and Metadata eXchange model (SDMX) [21] provided further inspiration.

In order to deal with the specificities of spatial information out of the totality of the IFS requirements, this paper selected only those that directly deal with geographic location (e.g., utilised agricultural area), or that can be derived from georeferenced information stored in IACS (e.g., area values, crop types, or agricultural practices). Having translated the requirements of the IFS into technical concepts, a class diagram was then developed in Unified Modelling Language (UML).

The reason for selecting UML is justified by the fact that foundation schemas, ISO TC 211 (geographic information) standards, INSPIRE, GSIM, and SDMX as well as the IACS domain model use this conceptual schema language. Importing the schemas of these standards and models in the modelling tool allowed the direct reuse of their component types and stereotypes. This approach was fundamental in achieving conformance with them. In order to support a wide interoperability in the geographic information realm, the types of variables were specified, in first order, as types of ISO 19103-Geographic Information-Conceptual Schema Language. When the standard did not provide some specific concepts, types were taken from INSPIRE data specifications, or the IACS domain model.

The spatial dimension adds additional challenges to the interoperability aspects of generic information systems. In addition to semantics and encoding, the potential issues of spatial representation geometries, coordinate reference systems and, if used, projection systems have to be sorted out. So far, in statistics the preference has been given to indirect referencing through administrative units or geographic names (addresses). However, natural and even social phenomena do not follow such unit. Therefore, the role of direct referencing with coordinates is coming to the forefront.

The IACS, through its LPIS subsystem or the geospatial application of farmers, can be of considerable help. It is advisable to reuse the geometries defined in the IACS, as they provide very accurate positional data. While the IACS is based on vector spatial schema, the IFS opts for coverages, namely for the equal area grid, as defined by the INSPIRE directive. However, the vector to grid conversion is a routine task that can be performed by the majority of geographic information system (GIS) software.

It should be noted that the use cases listed in the introduction could be implemented only after understanding the similarities and differences between the concepts of the two domains. Therefore, a full mapping exercise was performed, which consisted of the following steps:

- inventory of feature types of the two domains;
- schema mapping;
- clarification of data instance matching; and
- assessment in terms of quality and data transformation.

Before linking the elements of the two domains, it was necessary to reiterate and refine the feature types in the corresponding models, which also helped to specify the necessary operations for the collaboration model. One of the important decisions for modelling was to apply code lists, which is a proven tool for interoperability and for supporting subsidiarity in the implementation of EU law.

3. Results and Discussion

3.1. Geospatial Profile for the Integrated Farm Statistics

The IFS conceptual model (Figure 1) outlined on basis of regulation COM(2016) 786 [7] and its Annexes [22] is the first attempt to tackle its technical implementation. This model is not complete; it focuses on those elements that have a geospatial dimension and may reuse administrative data from the IACS. I used the same Geography Markup Language (GML) profile of UML as in the IACS domain model, in order to guarantee compatibility.

Regulation COM (2016) 786 also defines the categories of statistical variables. They are rigid classifications, which cannot be changed at the local level. This business rule is justified by the requirement of comparability of statistics and uniformity of reports delivered by the MS. It is worth

mentioning that from a technical point of view, the hierarchy level of categories is not homogenous. Some classifications list species of crops (e.g., sorghum, rice), while others point to genera (e.g., peas, beans and sweet lupines in one category) or to mixes of varieties (e.g., grapes for wine of protected denomination of origin).

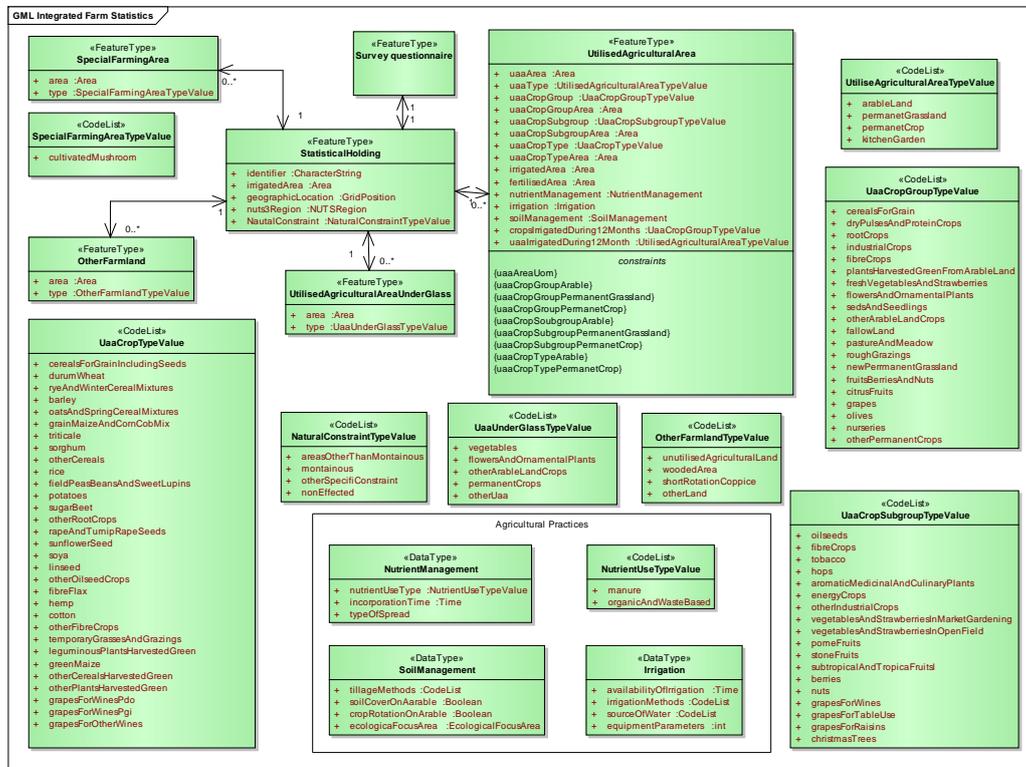


Figure 1. The geospatial profile of the IFS conceptual model.

This issue was addressed by defining hierarchical code lists and constraints that specify the links between the various levels of hierarchy. The code lists governance rules from INSPIRE were applied. The aggregated categories stemming from the regulation should be centrally managed; which means that end users (national statistical offices) cannot add additional values to them. However, allowing extensions of crop classification responds to local needs. For instance, MS may wish to include a detailed classification of fruits by species, or even by varieties, which helps to keep the focus of respondents during the surveys. A farmer may find it easier to deal with straightforward crop names rather than with their collections.

3.2. Inventory and Comparison of IFS and IACS Concepts

After defining which components of the IACS domain model are relevant for IFS, it was necessary to understand the ontology and the semantics of the source and the target domains. First, a comparison of feature types, one after another, was performed. This helped to see which transformations and queries are necessary at conceptual schema level.

The consistency of statistics might be at risk, when administrative concepts differ from statistical concepts. Such conceptual discrepancy can occur both at national and EU level. For example, in agricultural statistics “holding means a single unit, which has a single management and which undertakes economic activities in agriculture [. . .] or maintain agricultural land in good agricultural and environmental condition [. . .] as its primary or secondary activity” [7]. The IACS defines holding as follows: “all the units used for agricultural activities and managed by a farmer situated within the territory of the same Member State; [where] agricultural activity means production, rearing or growing

of agricultural products and maintaining an agricultural area in a state which makes it suitable for grazing or cultivation” [23].

The two definitions are seemingly the same. However, going deep in the details of the aforementioned regulations, it becomes clear that the two holdings are not equivalent. For instance, Article 9(2) of Regulation (EU) 1307/2003 strongly limits the eligibility of secondary agricultural activities, which creates a discrepancy in the scope. Moreover, different thresholds for the minimum areas are applicable. In the IFS, the smallest unit of utilised agricultural area is 5 ha, while in IACS it varies between 0.1 ha (Malta) to 5.0 ha (United Kingdom). Therefore, the selection criterion in the two domains is not the same; consequently, the two holdings are not identical.

By contrast, the “agricultural area” and the “utilised agricultural area” may imply that the two concepts are different because of fallow land. Again, the details tell us that fallow land and agricultural set aside are included in areas of arable land, permanent crops and permanent grasslands.

A similar comparison was performed for each relevant feature type. The description of further instances can be found in Table A2 of Appendix B.

3.3. Schema Mapping

The interactions between the IACS and IFS can be represented through a collaboration model. Naturally, the “collaboration” i.e., which components of the IACS can be reused are defined by the use cases of the IFS that were described in the introduction. A graphical representation, such as given in Figure 2, provides a generic overview of interoperability, where the interactions are represented with the “trace” stereotype of the relationships. This overview also shows the relation to the foundation schemas. The interacting models have the same structures—they have use cases and application schemas (in the IACS in sub-packages) that realise them.

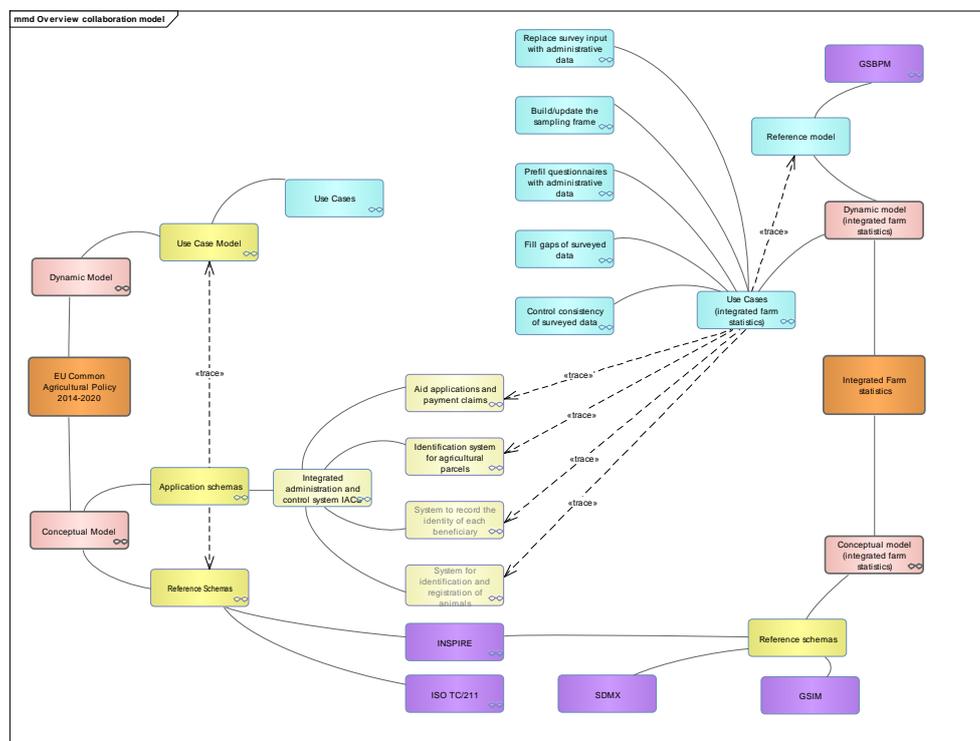


Figure 2. Overview of the collaboration model between the IACS domain model and IFS. Use cases are blue, application schemas yellow, while the foundation schemas of standards are violet. The fonts applied in the components that are not relevant to spatial information are in grey.

The collaboration model was developed finding the relationships between the corresponding feature types of the IFS and IACS. The following paragraphs describe the particularities and considerations of mapping. In all figures the following notation is used:

- Feature types of the IFS are green
- Feature types of the IACS are light brown
- Internal relationships in the IFS are green
- Internal relationships in the IACS are black
- Relationships between the IFS and IACS (that implement the collaboration), are blue.

The overview of the collaboration model is shown in Figure 3.

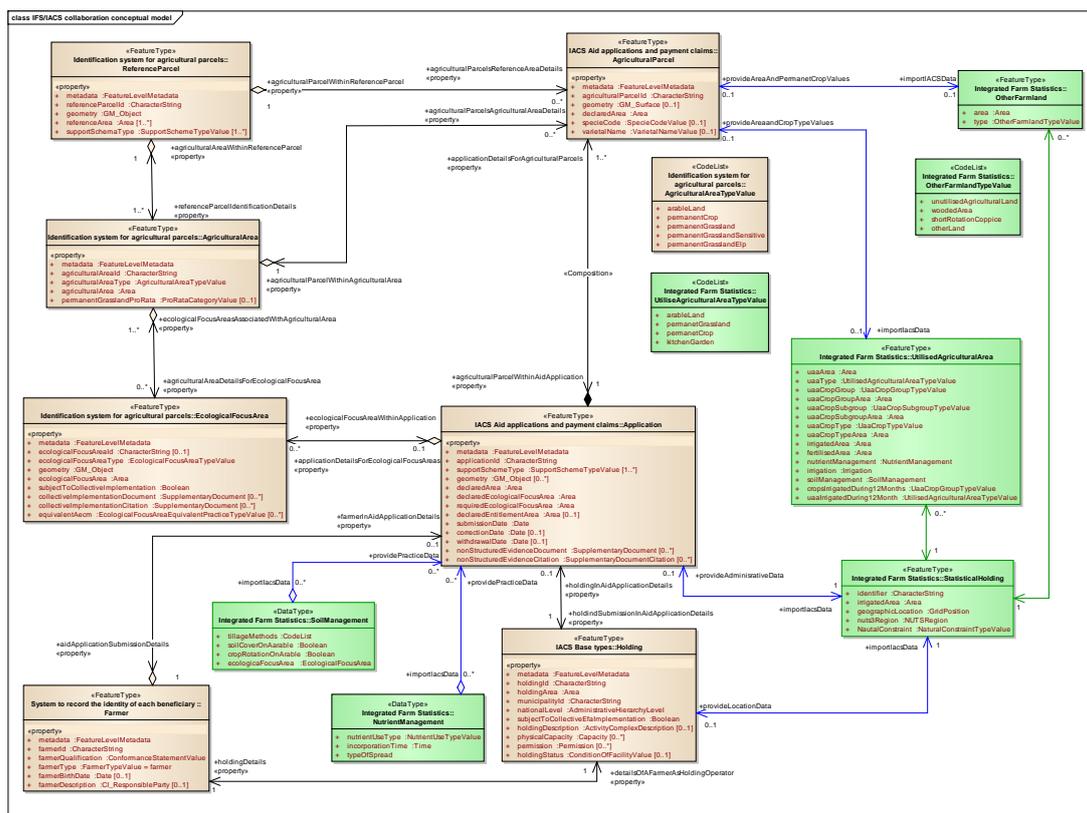


Figure 3. Overview of the information collaboration model.

The central concept of the IFS is the statistical holding (StatisticalHolding), which is the main unit of statistical data collection. This is in line with the recommendation of the Global Strategy [13]. The StatisticalHolding feature type is directly related to the Holding feature type of the IACS. The reason why the multiplicity of association between StatisticalHolding and Holding is not 1 to 1 stems from the ontology of the two domains, which is reflected by the area size threshold, or the agricultural area type, as discussed in the previous section. This means that no administrative data can be found for statistical holdings that are not declared in the IACS, lay in kitchen gardens or pursue agriculture as secondary activity on the land. Therefore, administrative data of the IACS should be complemented by statistical surveys and the two sources have to be merged with data conflation operation.

Even though the relationship between Statistical Holding and Holding brings all properties of Application and Farmer too, for the sake of emphasizing the connection between the Application and the StatisticalHolding, a relationship was created between them too. Such redundancy is allowed at

a conceptual level. The collaboration of the StatisticalHolding feature type with the elements of the IACS is shown in Figure 4.

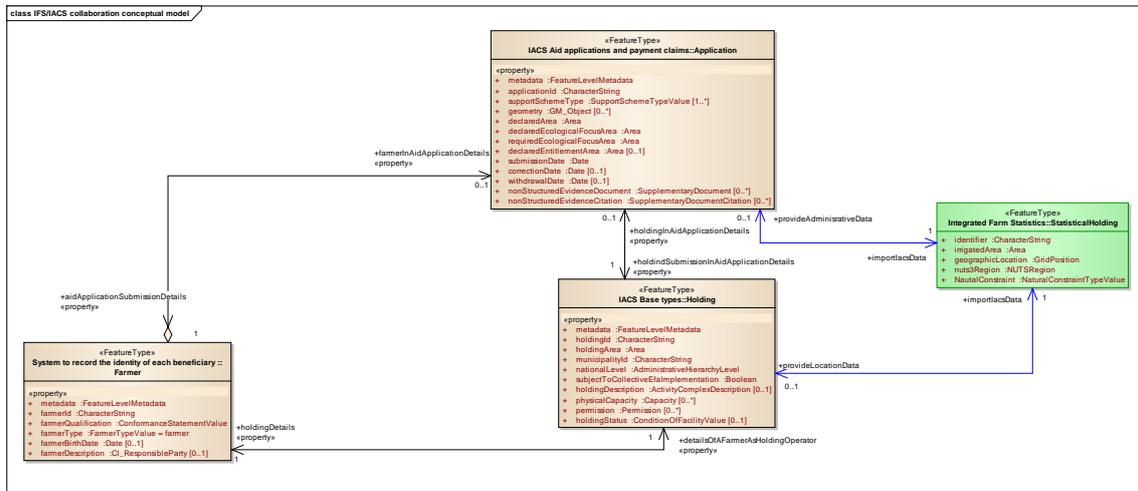


Figure 4. Relationships between the StatisticalHolding feature type with feature types of IACS.

The next collaboration relationship is established between Utilised Agricultural Area (UAA) in the IFS and Agricultural Area feature type in the IACS, which is shown in Figure 5. It is implemented through the Agricultural Parcel feature type, which is the subject of the Application. The classification of UAA seems to be very close to the agricultural area of the IACS. There are three common major categories: arable land, permanent crop, and permanent grassland, which is extended by a fourth category: the kitchen garden in the IFS.

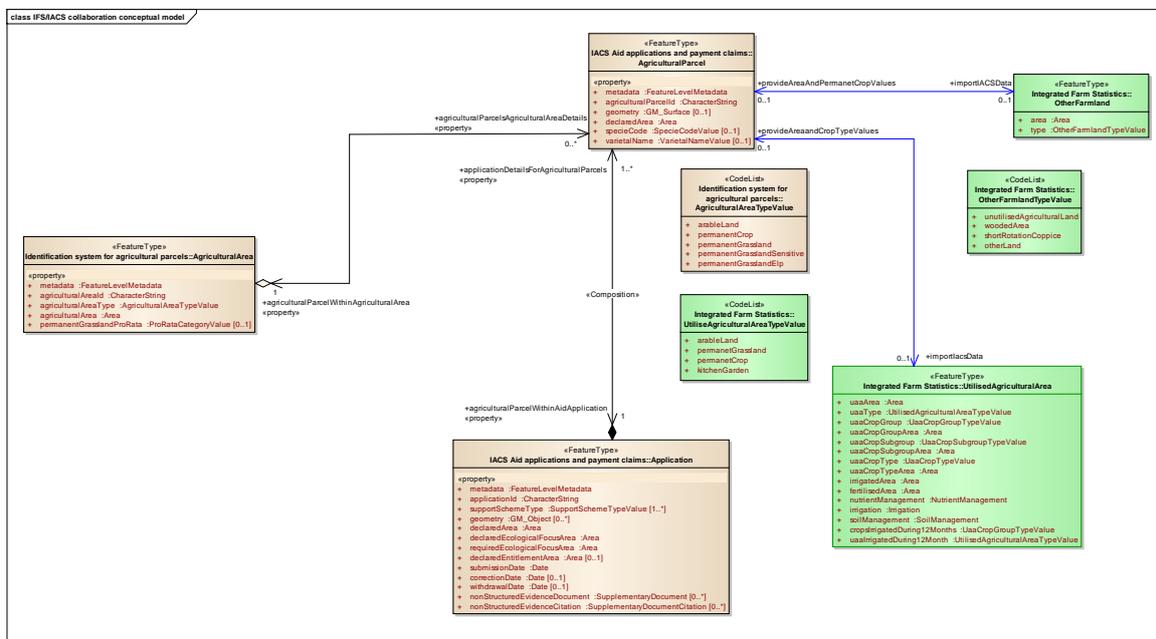


Figure 5. Relationship between UtilisedAgriculturalArea feature type of the IFS and AgriculturalArea of the IACS.

It should be noted, that the category of kitchen garden also exists in the IACS, but only in those countries where the single area payment scheme is applied. However, their typical small size usually

does not classify them for a declarable parcel. In statistics, small-scale farming, the so called subsistence farming data is also relevant. Therefore, administrative data have to be complemented with survey data. The observation here confirms the statement in the global strategy [13], which recommends screening the statistical samples of administrative units or census enumeration areas for small and subsistence farms.

It would be interesting to see, how bigger kitchen gardens appear in declaration in the countries, where the basic payment scheme (BPS) is applied. The BPS, by definition, does not contain this category. If kitchen gardens are declared as “arable” (with vegetables as crop and fruit trees as Ecological Focus Area (EFA) elements), this may lead to an overcount of arable land in statistics.

As attributes in a code lists or attributes of a feature type are equivalent representations in UML, the types of the other farmland feature type may be treated in the same way as the types of Utilised Agricultural Area. Therefore, we can see the relation between wooded area and its partial equivalents in IACS.

Wooded area, in principle, is not eligible for payment, therefore is not subject of the IACS. On the other hand, “afforested area” and “hectares of agroforestry” are EFA categories, which means that their area can be included in that of arable land if the size of the holding is less than 15 ha. Such cases create overcounts of arable and undercounts of wooded area in the IFS. When the size of the holding is bigger than 15 ha, because of the greening obligation the area values of these types are directly recorded in the IACS. We can conclude, therefore, that statistical offices may select those farmers, who do not need to be interviewed. For the rest, a traditional survey should be organised. The results from the two data sources have to be merged in order to embrace the full statistical population.

The “other land” category in the IFS is defined as “land occupied by buildings, farmyards, tracks, ponds and other non-productive areas”. Again, these types are not the subject of the IACS, as they are not eligible for payments. Even though it is not a conceptual requirement, many MS delineate such items in their LPIS implementations as ineligible, in order to help the declaration of farmers and controls by authorities. Therefore, LPIS might be a possible administrative source for the IFS. However, ponds require more attention as they can be included in list of landscape features and as such qualify as EFA. Therefore, problem of double count may occur again for holdings that are smaller than 15 ha.

The problem of potential double counts may be raised for land laying fallow. In principle, it is recorded in the IACS as part of arable land, as the absence of crops is not a discriminator for payments. However, as many MS collect crop data anyway, this information may be available in the declaration. In some countries fallow land is included in EFA, which means that this information can be readily imported for holdings that are bigger than 15 ha.

The case of short rotation coppice is also interesting. In IFS, on the one hand, it can be reported, under the other farmland type. However, there is a “permanent crop” category of UAA too. The regulation does not specify whether the latter should or should not include short-rotation coppice. Perhaps this categorisation will be refined in the course of implementation. Therefore, this ambiguity is not a problem of mapping between the two domains, but more an issue that should be clarified within the IFS.

Both the IFS and IACS deal with agricultural practices, but the scope and classifications in these systems are not aligned. The link to agricultural practices can be established through the Application feature type of the IACS. As shown in Figure 6 the soil cover on arable (soilCoverOnArable) category of the IFS correspond to the winter soil cover (winterSoilCover) and the catch crop (catchCrop) types in the IACS. However, these two categories may not be exhaustive, as the CAP regulations allow the adoption of equivalent agricultural practices at national or regional level, which reflect local conditions.

Crop rotation on arable land is listed in the IFS as an agricultural practice beneficial for the environment and climate. The corresponding category of the IACS domain model is the “beneficial multiannual sequence of crop rotation” (beneficialMultiannualSequenceOfCropRotation).

The fertiliser regime (fertiliserRegime) of the IACS can be mapped in the nutrient management (NutrientManagement) type of the IFS. However, the fertiliser regime does not contain all the properties that are needed in the IFS. Therefore, they have to be collected by traditional surveys.

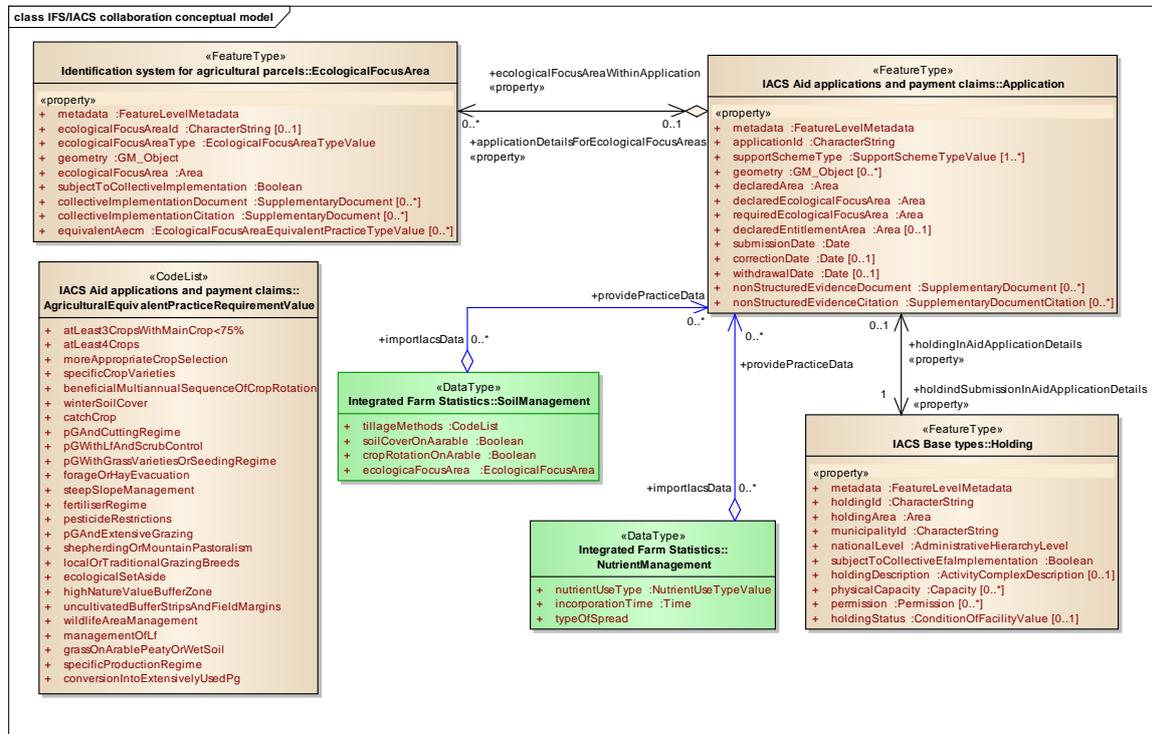


Figure 6. Relationships between agricultural practices in the IFS and IACS.

3.4. Data Matching Considerations

Overcoming the semantic difficulties between the domains is insufficient for achieving full interoperability. An unambiguous matching between the data instances is also needed. Statistical literature refers to the deterioration of data quality, especially in terms of accuracy and comparability with historic data, as a potential consequence of using administrative data. But it also acknowledges that administrative data are less impacted by memory effects (i.e., the respondents of the surveys unintentionally give a wrong answer), or effects of social desirability (i.e., what the respondents or the entities that commissioned the survey would like to see) [9].

In this context, it is worth noting that the IACS is one of the best maintained administrative information systems. In addition to the yearly controls of farmer declarations by the national authorities and the audits of the European Commission and the European Court of Auditors, the geospatial component is governed by strict technical specifications. For the LPIS a rigorous conformance assessment system (LPIS QA) has been put in place [19]. As all area values in the IACS have to be defined with a 100 m² accuracy, the data coming from this administrative source are fit for the purposes of the IFS.

The most efficient way of data matching can be implemented through global identifiers that are unique, persistent, and pick the lifecycle dimension of data instances. In the absence of a country/region-wide identifier (and registry service), there is no guarantee of uniqueness. In an ideal case, unique IDs are assigned according to the rules of national spatial data infrastructure. When no global IDs are maintained, thematic IDs (that are present in both systems) can be considered for data linking. Such thematic IDs might be the personal ID of the farmer, the ID of the agricultural parcels, etc.

As mentioned earlier, IFS requires the use of the Inspire 1 km Equal Area Grid cell code as a spatial reference, which supports data aggregation needed for data publishing. Such aggregation serves the depersonalisation of statistical information. In the IACS/LPIS vector based representations are used. The Holding feature type in IACS has an indirect geographic reference too—the identifier of the municipality. Linked to the holding, through the unique identifier precise vector geometries of reference parcels and agricultural areas are available. Thanks to the geospatial aid declaration of the farmers, vector geometries of agricultural parcels are also provided. As pointed out earlier, all vector geometries can be converted into geographic grid raster representation by using standard GIS software.

Further aspects of data matching are given in Table A3 of Appendix C.

4. Discussion

Producing European statistical data on the basis of different national administrative systems is very challenging in terms of comparability of the result. However, the IACS with its uniform content, data structure and regulated update cycles across the MS provides reliable data for European statistics.

According to the current strategy of EUROSTAT, the geospatial scope for agricultural statistics should focus on the use of land for agriculture and forestry and could be embedded in the broader scope of land-cover and land-use statistics [9].

Statistical data are traditionally reported for administrative units. However, natural phenomena do not follow them. A possible way out could be replacing object referencing (i.e., using administrative units as a basis to link thematic information) with direct geographic referencing, which could be either vector representation or a coverage (e.g., grid). It is expected that more time is needed for the full implementation of the gridded spatial schema in statistics, as such change should be synchronised with the decennial rhythms of the censuses. Once grids are introduced, updating data might be facilitated by advanced technologies, for example using remote sensing to determine land cover and crops. With automatised technology, statistics could be produced over shorter intervals. As Naik et al. pointed out “conventional methods are unable to respond quickly to changes in cropping patterns and therefore do not accurately record the area” [8].

The level of detail of statistical information also needs to be fine-tuned depending on the specific use cases. Aggregating all information at a georeferenced holding does not seem to be enough, for example, in case of agro-environmental statistics. The European Statistical Strategy, therefore, considers splitting holdings into smaller geolocalised units. If this is the case, the spatial units used in the IACS (reference parcels, agricultural parcel declared by the farmers,) would be of good use, which would be one more step towards easy data integration.

The reusability of data also depends on system interoperability, data harmonisation and the usage of standards. Not only it is necessary to standardise the information concepts, but also the business processes (such as data collection and processing methods, documentation with metadata, etc.) in order to be able to judge their fitness for the purpose, or to achieve comparability of data. Integrated statistical systems can resolve many of these problems by avoiding duplications of effort, preventing the release of conflicting statistics, and ensuring the best use of resources. Concepts, definitions, and classifications have to become standardised, allowing more systematic data collection across sources [13]. Such a holistic view can be achieved by the use of enterprise architecture technology which streamlines user requirements, business processes and information systems [14].

In the absence of harmonised definitions, collaboration between the two domains is not straightforward. The comparison of the total utilised agricultural area provided in the FSS with the total agricultural area registered in LPIS of the MS gives an illustration of the issue. The LPIS, by definition, should contain all agricultural land that are potentially subjects of aid declaration. The container of agricultural land in LPIS is the so called reference parcel (RP), which serves for identification and quantification of agricultural land eligible for support payments. Consequently, each RP has a unique identifier and an area value, which is also called as maximum eligible area. LPIS contains the small parcels too, as the area thresholds apply not at parcel, but at holding level. As a

holding may consist of more parcels, small parcels may become relevant for potential payment. Based on these arguments, we could expect that the total agricultural area in statistics cannot exceed the total agricultural area registered in LPIS.

However, as Table A4 in Appendix D shows, such consistency cannot be observed in each case. The root causes of differences lay in the legally defined ontology of the domains. As stated earlier, the statistical utilised agricultural area may contain “ineligible” for LPIS land cover types such as land occupied by buildings, farmyards, tracks, ponds and other non-productive areas as well as areas, where agriculture features as secondary activity (grazing on airports and solar panel farms, etc.). Statistics fully accounts wooded grasslands, while LPIS applies “pro-rata” reductions in order to account the maximum eligible area of RS in a net way.

It should be noted that the FSS results are made based on the extrapolation of sample results and therefore the figure for the total area cannot be expected to be identical the the IACS.

Further discrepancies can be attributed the thresholds that discriminate what instances belong to the universe of discourse. According to the recommendation of the agricultural statistics strategy, thresholds should be described in a common framework [24]. If this principle was implemented, it should be an important step towards interoperability.

A better convergence between the domains could be reached if the maximum eligible area of the RPs could have been replaced by the area calculated from the representation geometries. Meanwhile, the maximum eligible area values are yearly reported to the European Commission in the frame of the LPIS QA in the “Point Zero State” dataset, there is no obligation to report or publish the area values from the spatial representations.

Nevertheless, even the comparison of the currently available data yielded some benefits. The statistical data helped to clean the LPIS data from the errors in the units of measures. When the values differed by the order of magnitude, it could be assumed that instead of hectares square meters were used. On the other hand, when the numbers differed completely, the assumption was that the heterogeneity of units of measures happened at instance level. The reason why some LPIS data is missing from the table in Annex C is this data inconsistency.

5. Conclusions

This study compared the Integrated Farm Statistics and the Integrated Administration and Control System from an information point of view with the purpose of assessing the extent to which it is possible to reuse information in agricultural statistics.

Even though many concepts in the two systems seem to be similar, a number of challenges have to be faced when agricultural statistical surveys, or part of them, are being replaced by administrative data from the IACS.

First, finding out the relevance of administrative data is not a trivial task due to the lack of up to date and standardised metadata. Adopting the standard set of metadata specified in INSPIRE, which is also applied for the LPIS component of the IACS, would be a good solution to better position statistical and other data in the IACS in the national and the European SDIs.

Second, in spite of the overlaps in the two universes of discourse, there are significant differences in ontologies defined by the legal requirements of the domains. These differences originate in legal definitions and different discriminator thresholds that decide whether an apparently similar thing qualifies to be part of one or the other domain. This phenomenon leads to inconsistent populations (in statistics frames) and divergent semantic definitions of feature types. This is the reason why the comparison of data is not trivial. The detailed analysis of data content carried out in this study clarifies which pieces of information can be reused and what operation should be performed for reusing the IACS data in statistics.

The third challenge to be faced is data quality. A good starting point is that IACS implementations in the MS of the EU have to fulfil the same business requirements and are subject of rigorous controls. The strictly defined business cycle of aid applications, controls and payments help to produce current

data on a yearly basis. This frequency is fully satisfactory for IFS, where a two-year periodicity of surveys are foreseen. More attention is needed when the LPIS alone is used as an administrative source for the IFS, as the typical LPIS update cycles varies between 3 and 5 years in the various MS. Synchronisation of IFS survey periods with LPIS updates would improve the timeliness of administrative data for statistical surveys.

Administrative data, such as the IACS, are valuable sources for statistics as they are free from bias coming from respondents and commissioners of surveys. Replacing surveys with other, and especially quality-controlled sources renders statistics more efficient, reduces the burden of the respondents and the processing costs of statistical data [9]. In addition, geospatial technologies such as Global Positioning System (GPS) measurements, remote sensing and GIS could be employed to improve the quality of crop-area statistics [8].

Last, but not least it should be noted that an “infrastructural” approach manifested in applying standards, registers and registry services would greatly increase interoperability. This, and detection of user requirements from a broader perspective in the design phase of systems, would make data reuse more efficient and would lead to economies of public funds (no duplication or parallel work in public organisations). It would also contribute to a decrease in the burden of citizens, in this case of farmers, who are the subject of statistical surveys and declarations in the IACS.

The agricultural statistics strategy foresees a gradual integration of agricultural, forestry, land use and environmental statistics. In order to implement this principle, land-use and land-cover statistics delivered by administrative sources and primary production statistics need to fit together seamlessly. As one of the statistical priorities of the forthcoming 10 years relates to the environment (vulnerability, ecosystem services, economic risks), the relevance of the INSPIRE directive is increasing. The INSPIRE data standard is quite mature, as it is maintained frequently, and it includes ready to use data specifications and guidance, as well as a number of available (web) tools [18].

Administrative data coming from the IACS and other datasets published in frame of INSPIRE may accelerate the creation of data infrastructure due to the existing interoperability solutions. However, a common validation policy and a common language for validation rules (validation levels, terminology and syntax) have to be agreed in the whole European statistical community in order to preserve comparability of data.

Funding: This research received no external funding.

Acknowledgments: This work was carried out within the institutional work programme of the Joint Research Centre of the European Commission. The author acknowledges the work of Vincenzo Angileri and Aleksandra Sima (CAPLAND project) who revised the manuscript. Romuald Franielczyk (CAPLAND project) helped in LPIS data queries. Oscar Gomez Prieto (EUROSTAT) provided valuable background material. The author also expresses her gratitude to the two anonymous reviewers who helped to improve the manuscript.

Disclaimer: This article expresses and shares a methodological approach elaborated with the view of future improvements of sharing data collected under the implementation of CAP. It does not in any manner replace the formal texts neither in CAP nor in statistical legislation.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A.

Table A1. Cases Reported for Reusing IACS Data in the 2013 Edition of the Farm Structure Survey (FSS).

	Build/Update Survey Frames	Prefilling Survey Questionnaires	Control/Validate Micro/Macro Data	Impute Data	Calibrate Survey Estimates	Replace Surveys on Some Characteristics on a Part of the Population	Replace Surveys on Some of the Characteristics on the Whole Population	Georeferencing
Austria	x	x	x	x				
Belgium	x					x		
Bulgaria	x		x					
Croatia			x					
Cyprus								
Czech Republic			x					
Denmark	x		x		x	x		
Estonia		x	x	x		x		x
Finland	x		x	x				x
France		x	x					
Germany	x		x	x		x		
Greece						x		
Hungary	x		x	x		x		x
Italy	x						x	
Ireland	x		x	x			x	x
Latvia		x	x	x	x	x		
Lithuania	x	x	x	x				
Luxembourg				x		x		
Malta			x					x
Netherlands	x		x			x		x
Poland			x			x		
Portugal	x		x					
Romania								
Slovenia	x		x	x		x	x	
Slovakia	x		x					
Spain	x		x					
Sweden	x			x			x	x
United Kingdom	x						x	x

Appendix B.

Table A2. Comparison of Thematic Definitions of Real World Entities Defined by the Regulations of the IFS and IACS.

Concept Defined in the IFS	Selection Criteria	Corresponding Concept in the IACS and Its Comparison with the IFS
Utilised agricultural area	5 ha	The corresponding concept is agricultural area feature type In the IACS there are two thresholds at holding level (ref 1307/2013). In monetary terms: EUR 100 of minimal payable support; In area size: according to Art. 10 (1b) MS may decide according to local conditions. The minimum area unit may vary from 0.1 ha (Malta) to 5.0 ha (United Kingdom). As all area values in the IACS have to be defined with a 100 m ² accuracy, the data coming from this administrative source are fit for the purposes of IFS data
Arable land	2 ha	The corresponding concept is the arable land value from the code list of the agricultural area value types. Thresholds are related to the combination of all agricultural area types. No specific limit to arable land. The size of arable land has to be declared/registered in the IACS in order to assess the fulfilment of greening obligations (Ecological Focus Area (EFA), crop rotation). Therefore, this variable can be reused from the IACS. Another possibility for defining the area of arable land in IACS is to take the area values of arable land from the Land Parcel Identification System (LPIS). The role of the LPIS is to identify and quantify agricultural land. It should also delineate (in GIS tools) all types of agricultural areas. Therefore, the area values of the polygons are available. The only shortfall might be the update cycle—which is dependent on the local conditions of the MS can be up to 5 years. This is usually not synchronised with the periodicity of agricultural surveys and census.
Potatoes	0.5 ha	No specific corresponding concept. MS may opt to define crop type by species. IACS thresholds are related to the combination of all agricultural area types. No specific limit to this crop. For holdings with area less than 10 ha the crop species are irrelevant from point of view of payment. For bigger holdings (because of the requirement of crop diversification) this data is available.
Fresh vegetables and strawberries	0.5 ha	The same condition applies as in case of potatoes
Aromatic, medicinal and culinary plants, flowers and ornamental plants, seeds and seedlings, nurseries	0.2 ha	The same condition applies as in case of potatoes.
Fruit trees, berries, nut trees, citrus fruit trees, other permanent crops excluding nurseries, vineyards and olive trees	0.3 ha	The corresponding concept is the permanent crop value of the agricultural area type value code list. The “permanent crop” type in the IACS also contains vineyards and olive trees. It should be noted, that other administrative data source (vineyard cadastre) is usually available in the MS.
Vineyards	0.1 ha	Is included in the permanent crop type of agricultural area. MS may opt to introduce crop types based on species. Separate from the IACS, vineyard cadastre may be also available in the MS.
Olive trees	0.3 ha	Is included in the permanent crop type of agricultural area. The same conditions may apply as in case of vineyards.
Greenhouses	100 m ²	No specific category is defined; part of arable land.
Cultivated mushrooms	100 m ²	Not eligible for payment, as cultivation take place in buildings that are excluded from the scope of agricultural area.

Appendix C.

Table A3. Comparison of Selected Modelling Constructs Required by the IFS and IACS.

General Characteristics	Units/Categories	Comparison with the IACS
Agricultural holding identifier	holding ID	The type of ID are CharacterString in both domains. However, in absence of a country/region wide identifier (and registry service) there is no guarantee for having the same ID in both systems. In an ideal case, unique IDs are assigned according to the rules of national spatial data infrastructure. In the absence of unique and universal IDs, some thematic IDs (that are present in both systems) can be considered for linking data. Such thematic IDs might be the personal ID of the farmer, the ID of the agricultural parcels, etc.
Geographical location	Inspire 1 km Equal Area Grid cell code	In the IACS/LPIS, vector-based representations are used. The Holding feature type in the IACS has an indirect geographic reference—the identifier of the municipality. Linked to the holding—through the unique identifier precise vector geometries of reference parcels and agricultural areas are available. Also, thanks to the geospatial aid declaration of the farmer, vector geometries of agricultural parcels are provided. All vector geometries can be converted in geographic grid raster representation by using standard GIS software. In case it is needed, through the indirect reference of the municipality point coordinates can be determined that can be converted in grid raster, too.
NUTS¹ 3 region	NUTS 3 code	NUTS 3 regions are not generally used in the IACS. However, through vector geometries, by overlaying with NUTS region the data can be determined.
The holding has areas designated as facing natural constraints under Regulation No 1305/2013	L—areas with natural constraints; M—less-favoured mountainous area; O—Other areas affected by specific constraints; N—normal area	Payment for areas with natural constraints is one of the support schemes of the IACS. Therefore, the declarations contain the necessary alphanumerical data that can be imported into the IFS. In case (for example, for extended studies) geographic representations are needed, they can be obtained from the MS. Delineation of areas with natural constraints are produced by the MS and are approved by the commission services. However, technical specifications concerning spatial representations and the attributes are not standardised. The IFS classification may be adapted in the future in the light of developments in the Common Agricultural policy (CAP).

Appendix D.

Table A4. Comparison of the Total Utilised Agricultural Area Values of the FSS with the Total Area of the Reference Parcels in the Land Parcel Identification System (LPIS) .

MS/Year	Area in FSS ¹ (1000 ha)			Area in LPIS ² (1000 ha)			Difference (1000 ha)			Difference (%)		
	2010	2013	2016	2010	2013	2016	2010	2013	2016	2010	2013	2016
Austria	2878.17	2726.89	2669.75			2667.35			-2.4			-0.09
Belgium	1358.02	1307.9	1354.25	1453.66	1450.37	1482.74	95.64	142.47	128.49	6.58	9.82	8.67
Bulgaria	4475.53	4650.94	4468.5		4005.35	4331.78			-136.72			-3.16
Croatia	1346.34	1571.2	1562.96		1018.48	1120.31			-442.65			-39.51
Cyprus	118.4	109.33	111.93	148.36	136.78	145.61	29.96	27.45	33.68	20.19	20.07	23.13
Czech Republic	3483.5	3491.47	3455.41			3557.26			101.85			2.86
Denmark	2646.86	2619.34	2614.6	2732.51	2723.41	2713.04	85.65	104.07	98.44	3.13	3.82	3.63
Estonia	940.93	957.51	995.1	1122.05	965.86	1006.3	181.12	8.35	11.2	16.14	0.86	1.11
Finland	2290.98	2282.4	2278.61	2457.09	2254.71	2323.44	166.11	-27.69	44.83	6.76	-1.23	1.93
France	27,837.29	27,739.43	27,569.22	1529.48	27,561.28				-178.15			-0.65
Germany	16,704.04	16,699.58	15,166.86	13,252,441		17,196.11			2029.25			11.8
Greece	5177.51	4856.78	4553.83	78,609.7	78,873,247	4634.31			80.48			1.74
Hungary	4686.34	4656.52	4670.56	5650.61	5530.76	5789.56	964.27	874.24	1119	17.06	15.81	19.33
Italy	12,856.05	12,098.89			10,286.2	11,206.33			-1812.69			-17.62
Ireland	4991.35	4959.45	4938.52	4439.53	4133.33	4764.66	-551.82		-173.86	-12.43		-3.65
Latvia	1796.29	1877.72	1930.88	2114.22	1861.46	1793.62	317.93	-16.26	-137.26	15.04	-0.87	-7.65
Lithuania	2742.56	2861.25	2924.6	3559.68	3467.18	3230.8	817.12	605.93	306.2	22.95	17.48	9.48
Luxembourg	131.11	131.04	130.65	133.54	125.31	124.55	2.43	-5.73	-6.1	1.82	-4.57	-4.9
Malta	11.45	10.88	11.12	11.7	9.23	8.23	0.25	-1.65	-2.89	2.14	-17.88	-35.12
Netherlands	1872.35	1847.57	1796.26	2001.31	1938.32	1888.39	128.96	90.75	92.13	6.44	4.68	4.88
Poland	14,447.29	14,409.87	14,405.65	14,414.2	14,597.7	14,116.64	-33.09	187.83	-289.01	-0.23	1.29	-2.05
Portugal	3668.15	3641.59	3641.69	4459.06	2926.08	2991.28	790.91	-715.51	-650.41	17.74	-24.45	-21.74
Romania	13,306.13	13,055.85	12,502.54		12,800.31	13,111.86			-255.54			4.65
Slovenia	482.65	485.76	488.4		471.66	473.63			-14.1	-14.77		-2.99
Slovakia	1895.5	1901.61	1889.82	2083.21	2014.7	1949.07	187.71	113.09	59.25	9.01	5.61	3.04
Spain	23,752.69	23,300.22	23,229.75		24735	22,953.57			1434.78			5.8
Sweden	3066.32	3035.92	3012.64		3063.63	3031.45			27.71			0.9
United Kingdom	16,881.69	17,326.99	16,673.27		14,831.12	16,244.01			-2495.87			-2.64

¹ EUROSTAT http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ef_m_farmleg&lang=en. ² LPIS QA "Point Zero State" data.

References

1. European Parliament and European Council. Directive 2003/98/EC of the European Parliament and the Council of 17 November 2003 on the re-use of public sector information. *Off. J. Eur. Union* **2003**, *L345*, 90–96.
2. European Parliament and European Council. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. *Off. J. Eur. Union* **2013**, *L175/1*, 1–8.
3. European Commission. Commission notice—Guidelines on recommended standard licences, datasets and charging for the reuse of documents. *Off. J. Eur. Union* **2014**, *C240*, 1–12.
4. European Parliament and the Council of the European Union. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Off. J. Eur. Union* **2007**, *L108*, 1–14.
5. INSPIRE Thematic Working Group Statistical Units. *Data Specification on Statistical Units—Technical Guidelines*; European Commission Joint Research Centre: Brussels, Belgium, 2013.
6. European Council Regulation (EEC). No 571 on the organization of surveys on the structure of agricultural holdings for 1985 and for 1987. *Off. J. Eur. Communities* **1984**, *L 056*, 1–14.
7. European Commission. *Proposal for a Regulation of the European Parliament and of the Council on Integrated Farm Statistics and Repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011*; European Commission: Brussels, Belgium, 2016.
8. Naik, G.; Basavaraj, K.P.; Hegde, V.R.; Paidi, V.; Subramanian, A. Using geospatial technology to strengthen data systems in developing countries: The case of agricultural statistics in India. *Appl. Geogr.* **2013**, *43*, 99–112. [[CrossRef](#)]
9. European Commission DG ESTAT. *Use of Administrative Data Sources for Statistical Purposes*; Seminar of the Directors' Group on Agricultural Statistics (DGAS); European Commission: Luxembourg, 2016.
10. European Commission DG ESTAT. *Analysis of Six Country Reports about Pilot Studies on Linking IACS and Other Administrative Registers and Statistics*; European Commission: Luxembourg, 2016.
11. UNECE. *GSBPM v5.0—Generic Statistical Business Process Model—UNECE Statistics Wikis*; UNECE: Geneva, Switzerland, 2013.
12. Van Oosterom, P.; Lemmen, C. The Land Administration Domain Model (LADM): Motivation, standardisation, application and further development. *Land Use Policy* **2015**, *49*, 527–534. [[CrossRef](#)]
13. World Bank. *Global Strategy to Improve Agricultural and Rural Statistics*; Report No. 56719-GLB; The World Bank: Washington, DC, USA, 2011.
14. Tóth, K.; Kučas, A. Spatial information in European agricultural data management. Requirements and interoperability supported by a domain model. *Land Use Policy* **2016**, *57*, 64–79. [[CrossRef](#)]
15. Eurostat National Methodology Reports. Available online: <http://ec.europa.eu/eurostat/web/agriculture/national-methodology-reports> (accessed on 4 May 2018).
16. Vâju, S. ESS.VIP Admin Business Case. Available online: https://ec.europa.eu/eurostat/cros/content/essvip-admin-business-case_en (accessed on 9 August 2017).
17. Inan, H.I.; Sagris, V.; Devos, W.; Milenov, P.; van Oosterom, P.; Zevenbergen, J. Data model for the collaboration between land administration systems and agricultural land parcel identification systems. *J. Environ. Manag.* **2010**, *91*, 2440–2454. [[CrossRef](#)] [[PubMed](#)]
18. Abramic, A.; Gonzalez, D.; Bigagli, E.; Che-Bohnenstengel, A.; Smits, P. INSPIRE: Support for and requirement of the Marine Strategy Framework Directive. *Mar. Policy* **2018**, *92*, 86–100. [[CrossRef](#)]
19. Tóth, K.; Kučas, A. Conformance testing of geographic information. A case study on the Land Parcel Identification System. *Comput. Environ. Urban Syst.* **2018**. [[CrossRef](#)]
20. UNECE. *Generic Statistical Information Model*; UNCE: Geneva, Switzerland, 2013.
21. SDMX Community SDMX—Statistical Data and Metadata eXchange. Available online: https://sdmx.org/?page_id=5008 (accessed on 13 April 2018).
22. European Commission DG ESTAT. *Annexes to the Proposal for a Regulation of the European Parliament and of the Council on Integrated Farm Statistics and Repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011*; European Commission: Luxembourg, 2016.

23. European Parliament and the Council of the European Union. Regulation (EU) No 1307/2013 of the European Parliament and of the Council of 17 December 2013 establishing rules for direct payments to farmers under support schemes within the framework of the common agricultural policy and repealing Council Regulation. *Off. J. Eur. Union* **2013**, *L 347*, 608–670.
24. European Commission DG ESTAT. *Strategy for Agricultural Statistics for 2020 and Beyond*; European Commission: Luxembourg, 2015.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).