eurostat

# Strategy to Modernise Agricultural Statistics: New Pathways for the Future

## 2023 edition

STATISTICAL
WORKING PAPERS

# Strategy to Modernise Agricultural Statistics: New Pathways for the Future[1]


## Nicolas Lampach[ab2] and Domingo Maríınez-Solano[b]

[a]*Centre for Legal Theory and Empirical Jurisprudence, KU Leuven (Belgium)*

[b]*Agricultural and Fishery Statistics ( Estat E1), Eurostat (Luxembourg)*



1 We thank colleagues from ESTAT B4 for their support and their help to access web monitoring reports and social media coverage. Thanks to our colleagues from ESTAT E1 for the constructive comments in the early drafting phase of the article.



2 *Corresponding author* : Nicolas Lampach, European Commission, Eurostat, ESTAT E1, Agricultural and Fishery Statistics ;
E-mail: Nicolas.lampach@ec.europa.eu; Tel:+352 4301 33152

Manuscript completed in August 2023



Luxembourg: Publications Office of the European Union, 2023



Theme: Agriculture and fisheries

Collection: Statistical working papers

PDF: ISBN 978-92-68-06621-8 ISSN 2315-0807 doi: 10.2785/860462 KS-TC-23-003-EN-N

# Abstract

The growing number of challenges posed by emerging user needs, new data sources and technological progress calls for a rethinking of the way agricultural statistics are produced in the near future. We outline the current endeavours to modernise the core business processes of agricultural statistics within European statistical system. Moreover, we develop a user-integrated dissemination process by proposing a modern approach to evaluate the statistical products as a function of user attention. Based on user metrics recorded between January 2020 and 2023, we explore page views and user engagements from social media of most accessible statistical dissemination products in the realm of agriculture. The exploratory analysis manifest that a short number of tables account for more than half of user hits within all domains. Linking user engagements from social media posts with page views of news items reveals a heterogeneous distribution of issue attention. While there is a strong and positive relationship between statistics explained articles and the associated tables, our results indicate that users prefer crosscutting rather than specialised topics. These findings are discussed within the modernisation framework.

*Keywords*: Automation, Agricultural statistics, Content analysis, Eurostat, Dissemination, EU law, Modernisation; Unsupervised machine learning; Webanalytics

*JEL Classification:* C10, Q1, Q19

# 1. Introduction

The recent transformation of regulatory environment by reforms of common agricultural policy (CAP), the continuous structural changes in EU farming and recent market disruptions of agricultural commodities carry over new requirements for agricultural statistics. Alongside these changes in user needs, new data sources (such as national registers, farm management systems, administrative data, earth observation, modelling and big data) have become readily available to reduce the administrative burden and improve data collection methods (Castano, 2021, Desbois and Adam, 2007, Heerschap and Willenborg, 2006, Selenius et al., 2021, Shekhar et al., 2017, Snijkers et al., 2021, Villani et al., 2019, Whitcraft et al., 2019.

The Agricultural Statistics Strategy 2020 strives to embrace all these elements into one modernization framework with the key objective to produce data that meets the current and future user needs (European Commission, 2019, Jaakkonen et al., 2021). Although, the flexible ways of data collection seem promising to fill data gaps, it requires the design of a more adjustable data processing system with the inclusion of automated validations and improved timeliness.

Moreover, there exist a general agreement to improve the coherence between statistical domains (e.g economic accounts, farm structure survey, animal and crop, agri-environmental, wine and orchards) by harmonising definitions, standards and concepts at an intra-institutional level. In particular, a low level of coherence might lead to user misleading, cause wrong interpretation of the data and cast doubt on the overall quality of statistics. It becomes evident that these wide-ranging requirements challenge producers of statistics, but in return opening new avenues to policymakers, researchers and broader public to expand their understanding on current problems along the EU agricultural food chain (Hill, 1998).

The EU statistical office (Eurostat) is processing annually more than 8000 datasets in the realm of agriculture with the main aim to publish EU-wide agricultural statistics about agri-environmental indicator, agri-monetary data, farm structure, crop and livestock production. While a key role of agricultural statistics is to inform the public about trends, state and impacts of the sector, integrating user preferences of statistical products into the dissemination process is a relatively new concept.

Over the last years the usage of social media and information technologies are gaining widespread attention by international organizations to reach stakeholders, build relationships, but also to identify user preferences and design future communication strategy (Liu and Yuan, 2015, Macnamara and Zerfass, 2012, Pollach, 2011, Shin et al., 2015). User attention metrics can also be a viable feedback mechanism to learn the preferred types of statistical outputs and monitor the performance of statistical production systems.

While many studies explore the content of specific types of online media by public institutions, few attempts have been made to link user preferences with the improvement of public services (Porumbescu, 2016, Shin et al., 2015).

From the broad agricultural survey and statistics literature, several studies discuss the challenges and opportunities to better integrate statistical activities and improve cost-effective data collection (Carletto et al., 2021, Castano, 2021, Heerschap and Willenborg, 2006, Keita and Gennari, 2014, Selenius et al., 2021, Wallgren and Wallgren, 2016). Other investigate new data sources and methods to fill the data-gaps for informed decision-making on agriculture (De Jager, 2015, Karlsson and Widén, 2008, Milics et al., 2022, Snijkers et al., 2021, 2023). However, little effort has been made to understand the necessary adaptations of entire statistical system -ranging from collection to dissemination- to accommodate these future needs and requirements.

We seek to shed light on this gap by presenting the current endeavours of modernisation framework to make the production of EU agricultural statistics fit for purpose. Furthermore, this paper contributes to the vast agricultural statistics literature by proposing a modern approach to evaluate the production of statistical outputs as a function of user attention. Designing a more user-integrated dissemination strategy could be also useful for national statistical office and international organisations releasing information through distinct channels and publishing a wide range of statistical outputs. Linking user preferences with products might also provide an opportunity to streamline and prioritize data dissemination in a more cost-effective way.

The paper is organised as follows. Section 2 gives an overview of the EU legislative framework to devise a more integrated and flexible statistical system. The necessary structural adaptations of the core business processes covering data collection, processing, analysis and dissemination are explained in Section 3. Section 4 presents key findings of the user preferences of most accessible dissemination products within and across domains. Additionally, we introduce a use case of farm structure survey domain to prioritize the dissemination of Eurostat tables based on page views, request from other EU institutions and needs for internal publications. Section 5 discusses limitations and further research needs.

# 2. EU legislative framework

Given the considerable efforts made in the last years to implement the Strategy for Agricultural Statistics 2020 and beyond, several EU legislative acts have been revised to allow the application of any data source that fulfils the necessary statistical quality requirements. This new legal provision sets out the option to retrieve data from available sources, such as administrative records, earth observation, farm management systems or advanced modelling tool (e.g machine learning, Bayesian inference, imputation methods) complementing traditional agricultural surveys.

In general, three pillars form the EU legislative framework for agricultural statistics:

i) Integrated Farm Statistics (IFS)[1] covering structural data on farms

ii) Economic Accounts for Agriculture (EAA) Regulation[2] representing a satellite account of the European System of Accounts by providing complementary information to the agricultural industry

iii) Statistics on Agricultural Input and Output (SAIO) Regulation[3] integrating five statistical domains into one harmonised system

The IFS Regulation entered into force in 2018 to provide the necessary legal basis to collect, process and disseminate farm-level micro data on all farms in a country. The EAA Regulation was amended in 2022 to provide regional breakdown. The SAIO Regulation entered into force in 2022 to integrate the animal production, crop production, organic farming, agricultural prices, and pesticides and nutrients statistical domains into one system. Until now, those domains have been produced in parallel data flows with different legislative frameworks. It is expected that SAIO will have the greatest effect on the harmonisation and coherence of the production of agricultural statistics across the different domains.

[1] Regulation (EU) 2018/1091 of the European Parliament and of the Council of 18 July 2018 on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011.

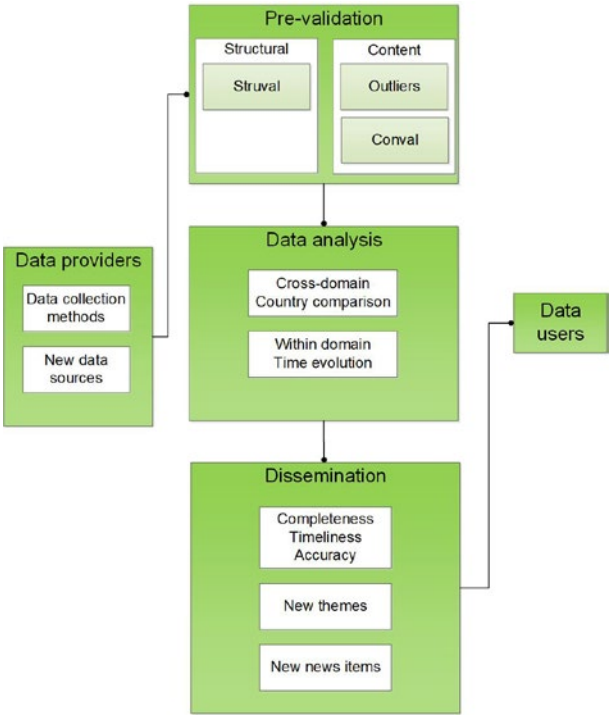
[2] Regulation (EC) No 138/2004 of the European Parliament and of the Council of 5 December 2003 on the economic accounts for agriculture in the Community.

[3] Regulation (EU) 2022/2379 of the European Parliament and of the Council of 23 November 2022 on statistics on agricultural input and output, amending Commission Regulation (EC) No 617/2008 and repeal- ing Regulations (EC) No 1165/2008, (EC) No 543/2009 and (EC) No 1185/2009 of the European Parliament and of the Council and Council Directive 96/16/EC.

# 3. Business processes

The modernised agricultural statistical system covers all phases ranging from data collection at provider level (e.g national statistical office, other governmental agencies) to data users (see Figure 1). At the core of the modernisation stands the integration of an automated data validation and processing system to reduce primary the time, but also to improve the efficiency in detecting errors.

**FIGURE 1:**
**Flowchart of core business processes**




## 3.1. Data collection

The main idea of future data collection is to replace traditional surveys with alternative data sources to decrease the administrative burden of farmers and excessive costs of data providers.

A commonly used source in many countries is administrative data. The massive use of data from administrative registers in IFS has led to the partial discontinuation of surveys and visits to farms in many countries. Many domains under the SAIO framework are progressively substituting or complementing surveys with information from administrative registers to reduce time and effort. EAA use information from administrative data sources either directly from agricultural and non- agricultural registers or from other statistical data collections, such as national accounts or governmental finance statistics. Available information from sources other than official statistics such as the farm sustainability data network are relevant inputs to the EAA.

More innovative data collection methods involves the application of earth observational data and artificial intelligence. Statistical offices and other national data providers are increasingly applying earth observation methods to compile data on land use for the IFS. This leads to higher accuracy of the data, better timeliness and less burden to farmers.

Further important aspect is not only to confine the data collection to traditional methods, but also to leverage new modelling techniques. SAIO regulation lays down the necessary provisions to allow more advanced prediction methods to compute the data from other available sources. This is, in particular, important for areas requiring high level of details. This granularity is difficult to collect with traditional surveys without severely affecting cost and burden. Using models with historical data and current registers as main inputs can provide highly accurate data to users. In particular, the animal production domain is applying this practice with the methodological guidance of Eurostat to ensure quality and harmonisation.

Unlike national registers provide a useful alternative to surveys, retrieving data directly from new digital data sources such as farm management systems or Application Programming Interfaces (APIs) through which data are made available are still at the development stage (Snijkers et al., 2023). However, many data are electronically available and might be accessed with IT technologies. For instance, a Dutch research team explores the use of sensor-based data of precision farming machines (e.g John Deere) in the completion of agricultural surveys.

## 3.2. Data processing

Eurostat has introduced an automated data processing system that pre-validates the transmitted dataset by performing all validation steps of level 0, 1 and part of level 2 according to ESS validation between the same dataset (European Commission, 2022).

This automation has considerably shorten the processing time between transmissions from national data providers to the release of official statistics for the public. Furthermore, it

also results in a more efficient data processing flow, where potential basic errors are systematically detected without human intervention and timelier (immediate) feedback is provided to the data sender.

Outlier detection is another high-level automated validation procedure that has been recently implemented. The data processing tool compares the transmitted data with the historical series. If a certain value falls out of a confidence interval (tailor - made by geographical entity and variable), the dataset needs to contain an explanatory note in order to be processed.

The number of rules usually amounts to hundreds per dataset or even more than 500 in the case of farm structure survey.

**FIGURE 2:**

## Excerpt of validation rules in farm structure survey

| MD_CHR_FIELDID | MD_CHT_NAME | SEVERITY | Precondition (IF...) | Condition (THEN...) | Error message (no commas AND no semicolon AND no quotation marks) (ELSE...) |
|---|---|---|---|---|---|
| VR_IFS2023_LAFO033 | LAFO033 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_NFAM','PER_LEG_NEG','PER_LEG_EG') AND (SEX_MAN = 'F') AND (WH_MAN_AWU_PC = 'PC25T49') | FLF_D_RNFAM_F_PC25T49 > 0 | IF (the manager of the holding is not a family member of the holder OR holding is a legal person (PER_LEG_NEG or PER_LEG_EG)) AND the manager is a female AND the manager work time is in 25-49 percent band THEN the number of female non-family members (directly employed by the holding) with Work time 25-49 has to be > 0 |
| VR_IFS2023_LAFO034 | LAFO034 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_NFAM','PER_LEG_NEG','PER_LEG_EG') AND (SEX_MAN = 'F') AND (WH_MAN_AWU_PC = 'PC50T74') | FLF_D_RNFAM_F_PC50T74 > 0 | IF (the manager of the holding is not a family member of the holder OR holding is a legal person (PER_LEG_NEG or PER_LEG_EG)) AND the manager is a female AND the manager work time is in 50-74 percent band THEN the number of female non-family members (directly employed by the holding) with Work time 50-74 has to be > 0 |
| VR_IFS2023_LAFO035 | LAFO035 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_NFAM','PER_LEG_NEG','PER_LEG_EG') AND (SEX_MAN = 'F') AND (WH_MAN_AWU_PC = 'PC75T99') | FLF_D_RNFAM_F_PC75T99 > 0 | IF (the manager of the holding is not a family member of the holder OR holding is a legal person (PER_LEG_NEG or PER_LEG_EG)) AND the manager is a female AND the manager work time is in 75-99 percent band THEN the number of female non-family members (directly employed by the holding) with Work time 75-99 has to be > 0 |
| VR_IFS2023_LAFO036 | LAFO036 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_NFAM','PER_LEG_NEG','PER_LEG_EG') AND (SEX_MAN = 'F') AND (WH_MAN_AWU_PC = 'PC100') | FLF_D_RNFAM_F_PC100 > 0 | IF (the manager of the holding is not a family member of the holder OR holding is a legal person (PER_LEG_NEG or PER_LEG_EG)) AND the manager is a female AND the manager work time is in 100 percent band THEN the number of female non-family members (directly employed by the holding) with Work time 100 has to be > 0 |
| VR_IFS2023_LAFO037 | LAFO037 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('PER_LEG_NEG','PER_LEG_EG', 'HLD_GRP', 'UNIT_CHL') | FLF_D_RFAM_M_PC1T24 + FLF_D_RFAM_M_PC25T49 + FLF_D_RFAM_M_PC50T74 + FLF_D_RFAM_M_PC75T99 + FLF_D_RFAM_M_PC100 + FLF_D_RFAM_F_PC1T24 + FLF_D_RFAM_F_PC25T49 + FLF_D_RFAM_F_PC50T74 + FLF_D_RFAM_F_PC75T99 + FLF_D_RFAM_F_PC100 = 0 | IF the holding is a legal person (PER_LEG_NEG or PER_LEG_EG) OR the holding is a group holding (HLD_GRP) OR the holding is a common land unit (UNIT_CHL) THEN the farm work of family members of the holder is 0 (zero) |
| VR_IFS2023_LAFO042 | LAFO042 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('UNIT_CHL', 'PER_LEG_NEG', 'PER_LEG_EG') | NOGA_HLD_RH = 1 | IF holding is of type (Holding is a legal person (PER_LEG_NEG or PER_LEG_EG) OR Holding is common land (UNIT_CHL)) must have NOGA_HLD_RH = 1 |
| VR_IFS2023_LAFO043 | LAFO043 | Error | EXTPOL_FACT('LAFO <> 'null' AND MOGA_HLD_RH = 1 | (SOGA_HLD_RH = 0) AND (NOGA_HLD_RH = 0) | For OGA of the holder (MLFO023) which are directly related to the holding, only one option can be true out of MOGA or SOGA or NOGA |
| VR_IFS2023_LAFO044 | LAFO044 | Error | EXTPOL_FACT('LAFO <> 'null' AND SOGA_HLD_RH = 1 | (MOGA_HLD_RH = 0) AND (NOGA_HLD_RH = 0) | For OGA of the holder (MLFO023) which are directly related to the holding, only one option can be true out of MOGA or SOGA or NOGA |
| VR_IFS2023_LAFO045 | LAFO045 | Error | EXTPOL_FACT('LAFO <> 'null' AND NOGA_HLD_RH = 1 | (MOGA_HLD_RH = 0) AND (SOGA_HLD_RH = 0) | For OGA of the holder (MLFO023) which are directly related to the holding, only one option can be true out of MOGA or SOGA or NOGA |
| VR_IFS2023_LAFO047 | LAFO047 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('PER_LEG_NEG', 'PER_LEG_EG', 'HLD_GRP', 'UNIT_CHL') | (MOGA_FAM_RH = 0) AND (SOGA_FAM_RH = 0) | Legal holdings (PER_LEG_NEG OR PER_LEG_EG), holding-groups (HLD_GRP) or common land units (UNIT_CHL) must have MOGA_FAM_RH = 0 AND SOGA_FAM_RH = 0 |
| VR_IFS2023_LAFO052 | LAFO052 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_HLD', 'FARM_HLD_SPOUFAM', 'FARM_SPOU', 'FARM_FAM', 'FARM_NFAM', | SOGA_NFAM_RH >= 0 | Other gainful activities of non-family members (related to the holding) as secondary activity (MLFO027) are collected for sole holder holdings and holding-groups |
| VR_IFS2023_LAFO053 | LAFO053 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('PER_LEG_NEG', 'PER_LEG_EG', 'UNIT_CHL') | MOGA_NFAM_RH = 0 | Other gainful activities of non-family members (related to the holding) as main activity (MLFO026) are NOT collected for legal holdings ((PER_LEG_NEG or PER_LEG_EG) NOR common land units ((UNIT_CHL)) |
| VR_IFS2023_LAFO054 | LAFO054 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('PER_LEG_NEG', 'PER_LEG_EG', 'UNIT_CHL') | SOGA_NFAM_RH = 0 | Other gainful activities of non-family members (related to the holding) as secondary activity (MLFO027) are NOT collected for legal holdings ((PER_LEG_NEG or PER_LEG_EG)) NOR common land units ((UNIT_CHL)) |
| VR_IFS2023_LAFO057 | LAFO057 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_HLD_SPOUFAM', 'FARM_SPOU', 'FARM_FAM', 'FARM_NFAM', 'PER_LEG_NEG', 'PER_LEG_EG', | MOGA_FAM_NRH = 0 | Other gainful activities of family members (not related to the holding) as main activity (MLFO029) are only collected for sole holder holdings where the holder is the manager |
| VR_IFS2023_LAFO058 | LAFO058 | Error | EXTPOL_FACT('LAFO <> 'null' AND LEG_FORM IN ('FARM_HLD_SPOUFAM', 'FARM_SPOU', 'FARM_FAM', 'FARM_NFAM', 'PER_LEG_NEG', 'PER_LEG_EG', | SOGA_FAM_NRH = 0 | Other gainful activities of family members (not related to the holding) as secondary activity (MLFO030) are only collected for sole holder holdings where the holder is the manager |

## 3.3. Data analysis

Agricultural statistics are integrated into more coherent and interoperable IT systems expecting to improve the efficiency of data flows and the quality of statistics (i.e timeliness, internal coherence). This innovation allows performing faster cross-domain checks and leads to higher coherence between the different agricultural domains. The data analysis from different data sources was usually carried out with a time consuming procedure of data extraction and the calculation of defined indicators for data publications.

All these tasks are automated, while human intervention is required only for tasks with highest benefit.

## 3.4. Dissemination

All these improvements have a direct effect on the dissemination products in terms of timeliness, accuracy and completeness. Under the new SAIO regulation, data collection will no longer be on voluntary basis implying that the data will be disclosed for all countries at a specific time moment. In return, EU aggregates -commonly used by EU decision-makers- will also be readily available in all data publications.

Moreover, the automated processing and cross-domain validation will increase the level of accuracy and ensure the production of harmonized data sets across domains. For instance, the considerable discrepancies of organic statistics between Farm Structure Survey (FSS) and Agri-Environmental (AE) domain might be explained by differences in the legal basis, definition on organic farming and the dissimilar actors involved in the data transmission.

While traditionally the dissemination of agricultural statistics has focused on themes related to farm structures, prices of agricultural inputs and outputs, labour or production, the modernisation framework paves the way to cover cross-cutting themes and a wide range of environmental topics (e.g nutrients, pesticide use, nitrogen balance, emissions) related to agriculture.

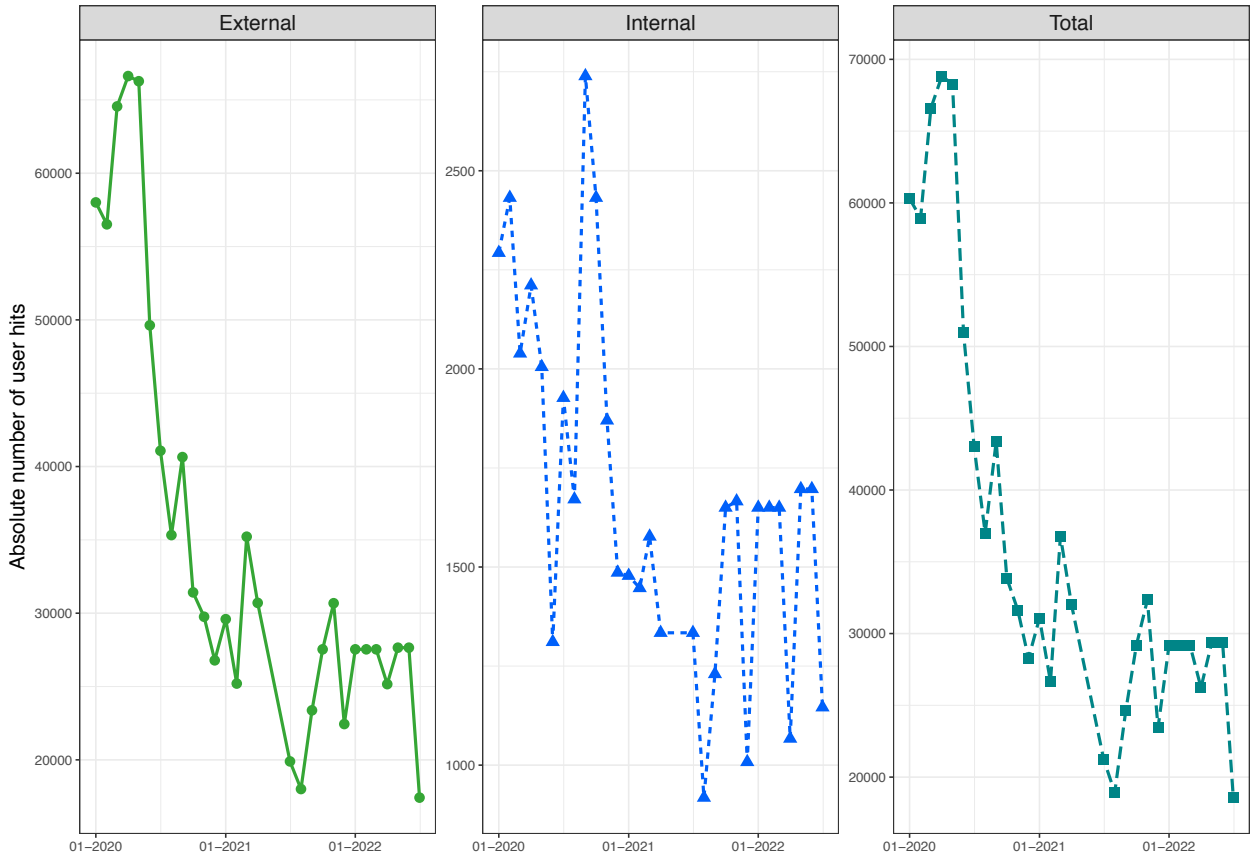
# 4. User attention of dissemination products

## 4.1. Web and social media analytics

To identify the user's need and interest in farm statistics, we retrieved user preferences of all dissemination tables (N=593) of six agricultural domains (AEI, ANI, APRO, EAA, FSS, VIT) from web monitoring reports between January 2020 and July 2022 published on monthly basis at an internal website of Eurostat[4]. Displayed in Figure 3 are the monthly variations of user clicks on data publications across distinct groups. While external users, such as national administrations, researchers, NGOs, public, tend to be mostly affected by COVID SAS-19 pandemic showing a steep drop in April 2020, the European Commission internal users follow a more consistent pattern with peaks in most busy months (January, June, and September). It can also be seen that external users represent the largest share in the demand for European farm statistics.

**FIGURE 3:**

### Monthly records of user statistics for distinct user group



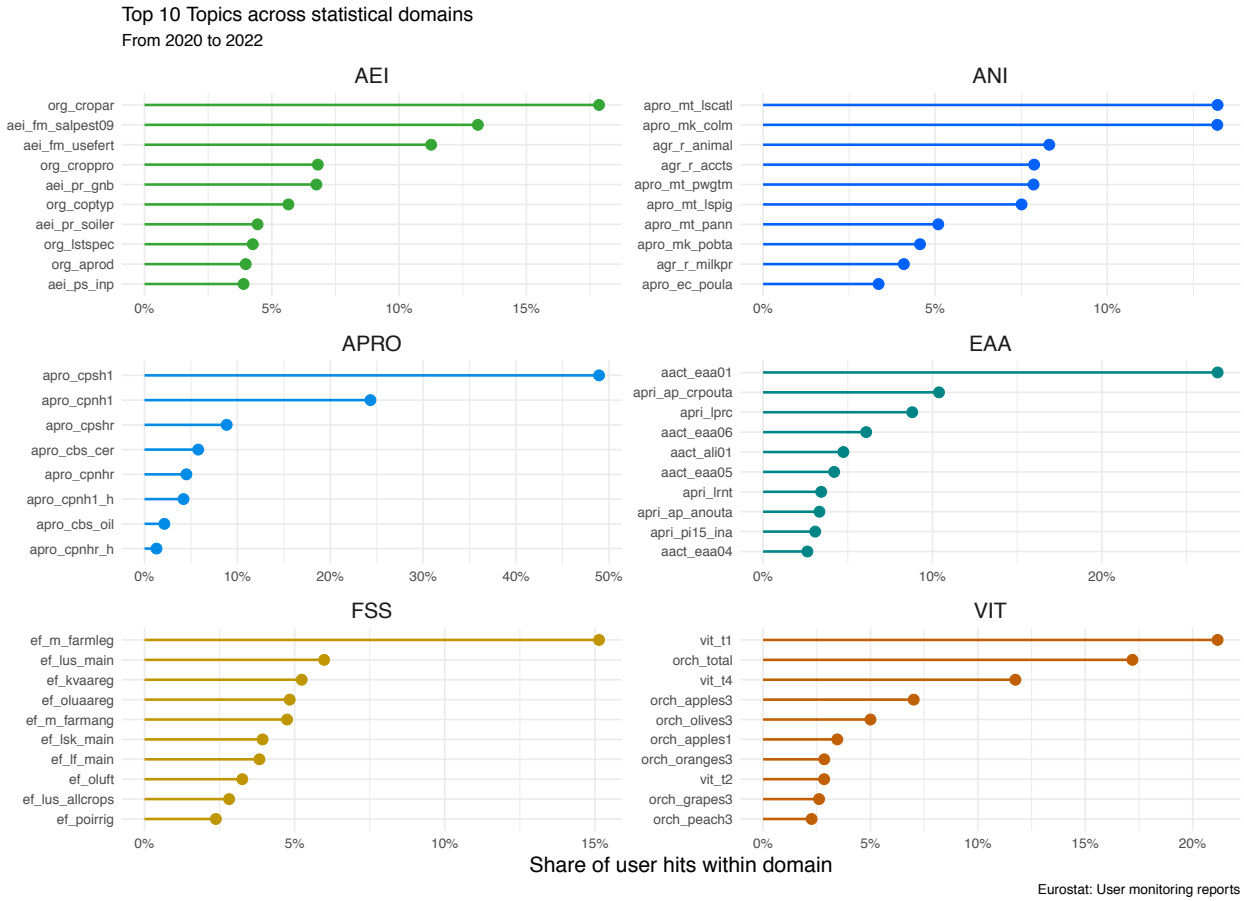

4 An overview of the number of tables produced by each domain is given in Table 6 in Appendix.

Ranking the ten most salient tables across domains reveals that external users prefer mostly generic tables summarizing general information on organic farming, economic accounts, livestock population, crop production, farm indicators, wine-growers and tend to favour less so specialized tables with detailed information on a specific topic (see Figure 4). In most domains, the share of the ten most salient tables accounts for more than half of user clicks.

Alternative text: Bar chart ranking the ten most salient data publications within the six statistical domain in the field of agriculture. More than 50% of total user clicks are attributed to the dataset apro_cpsh1 in the domain of crops statistics (APRO).

**FIGURE 4:**

## Ranking the ten most popular tables within the domain




*Note.* AEI, ANI, APRO, EAA, FSS and VIT denotes agri-environmental indicator, animal, agricultural production of crops, economic-account of agriculture, farm structure survey and viticulture, respectively. Interested user can navigate user hits of data publications and other products with an interactive web app (https://euagristat.shinyapps.io/user_attention_tables)

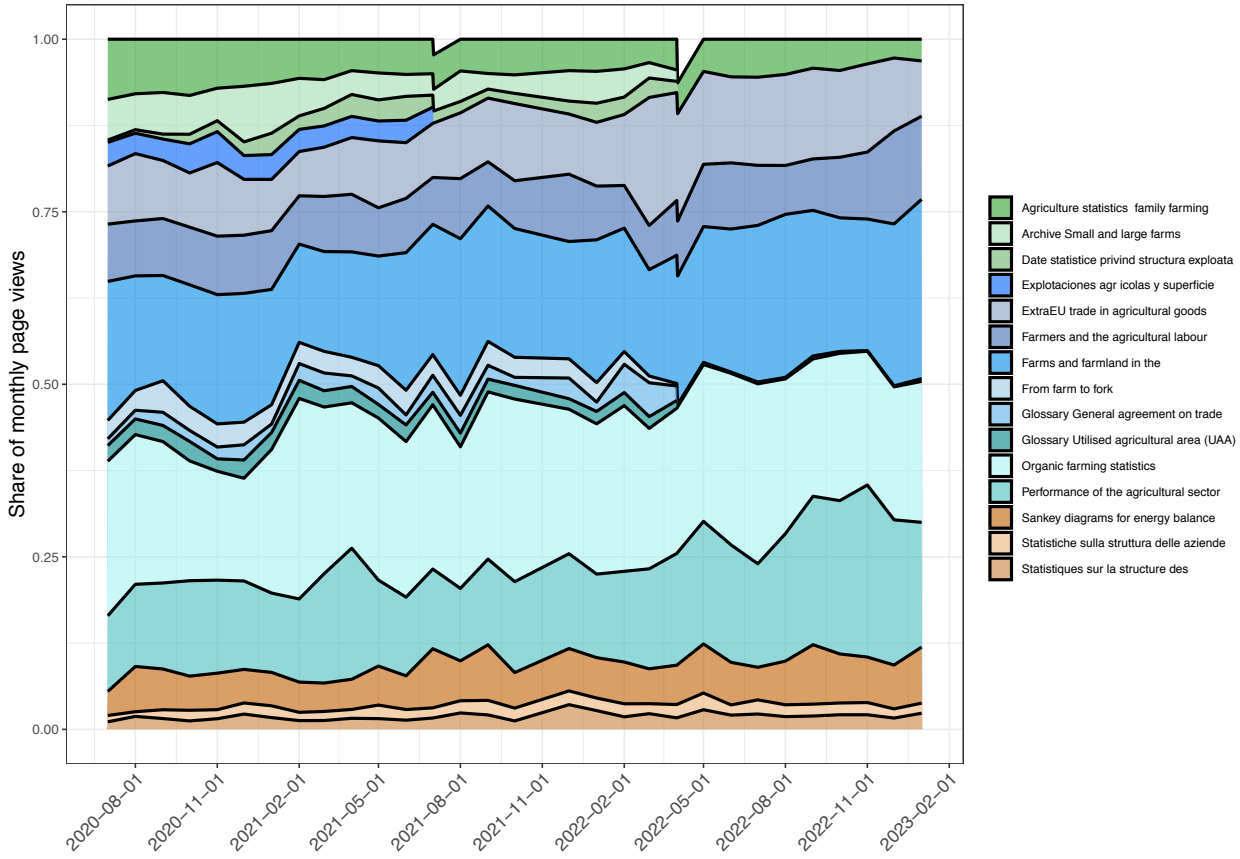
Besides data dissemination, Eurostat publishes a wide range of articles on a monthly basis in different EU languages supplying public with general information on statistical glossary, agricultural themes and methodologies. To get a sense on how much users are engaged with these web services, we measure user attention as the total number of page views on the website of the article.

The data is automatically downloaded from corporate platform provided by European Commission (2023) to

track website audience[5]. From the large bulk of online publications (N=1350), the most preferred articles across time spin around traditional topics ranging from performance of agricultural sector to family farming. While also Statistics Explained (SE) articles on organic farming and Sankey diagrams for energy balance constitute a large share of page views from users, technical details and definition of concepts (e.g glossary Utilised agricultural area) seem to be less viewed in comparison to other.

**FIGURE 5:**

## Monthly page views of the 15 most clicked statistics explained articles




Because certain SE articles contain a number of graphs derived directly from latest data publications, we can put into relationship the daily user attention of SE articles and the associated Eurostat tables mentioned in the publication. We expect that both dissemination products are positively correlated. The main idea of SE articles is to present statistical topics in easily understandable way to users by providing links to the latest data.

Displayed in Figure 6 is the relation of daily page views on SE articles and clicks on the related data publication of four prevalent SE articles[6]. The size of the dots captures the total number of unique data publications used in the article providing some indication on the degree of crosscutting issues. It can be seen that there is a strong positive relationship between both products. Another viable instrument to measure public engagement with the content of dissemination products is social media. Eurostat

[5] Staff from EU institutions retrieving information from EU websites are excluded from the data analysis due to cookie policy. APi-retriever are also suppressed from the data as they could biased the data analysis

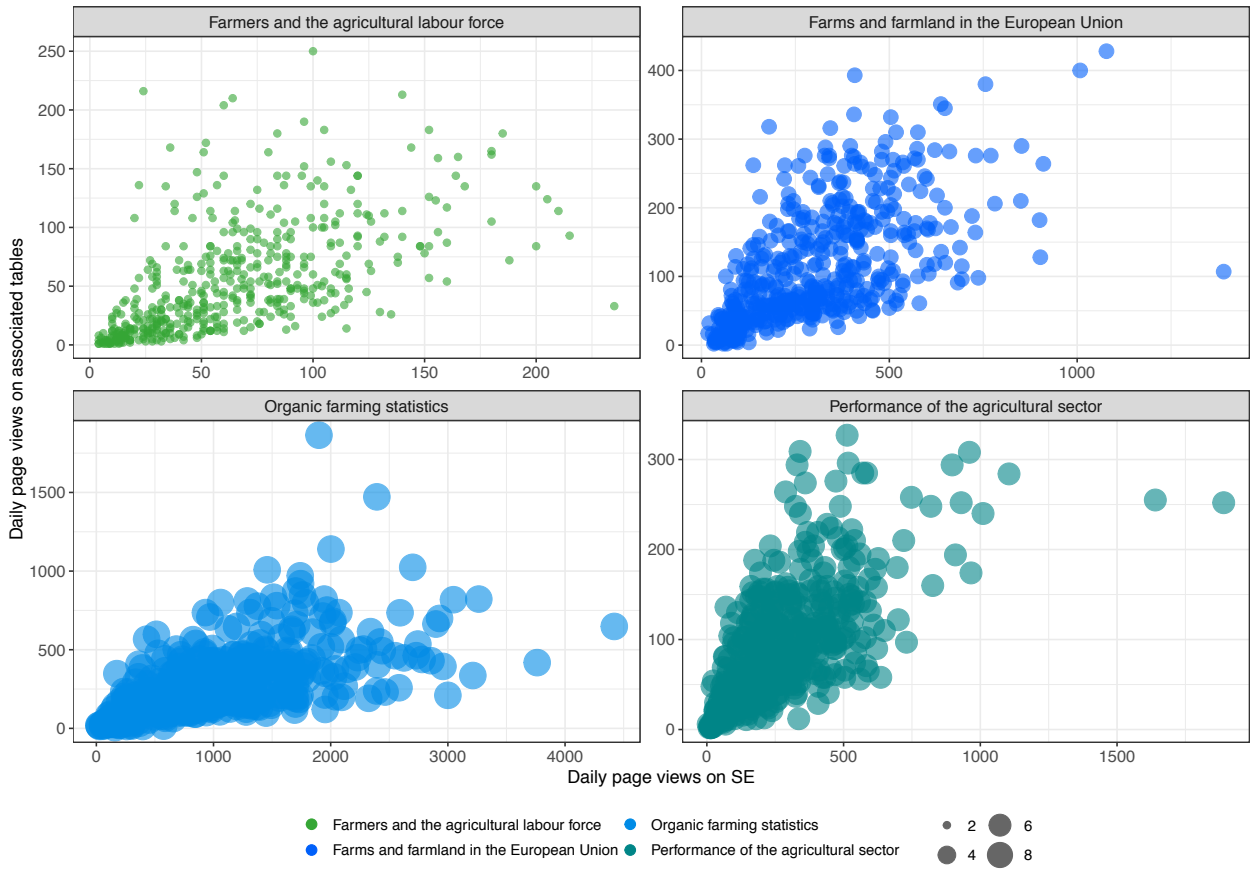
[6] Not all SE articles include direct links to data publication and therefore the analysis is limited.

is active on twitter (since 2011), Facebook (since 2017), Instagram (since 2020) and LinkedIn (since 2022) with high number of followers and interactions with the posts.

It is clear that the audience (i.e public vs. specialist) might vary considerably across channels and therefore the promotion of dissemination products require a slightly distinct approach. Yet, news items comprising short visuals explained in an understandable way are posted in all four channels.

**FIGURE 6:**

## Relationship between SE article and the associated data publication

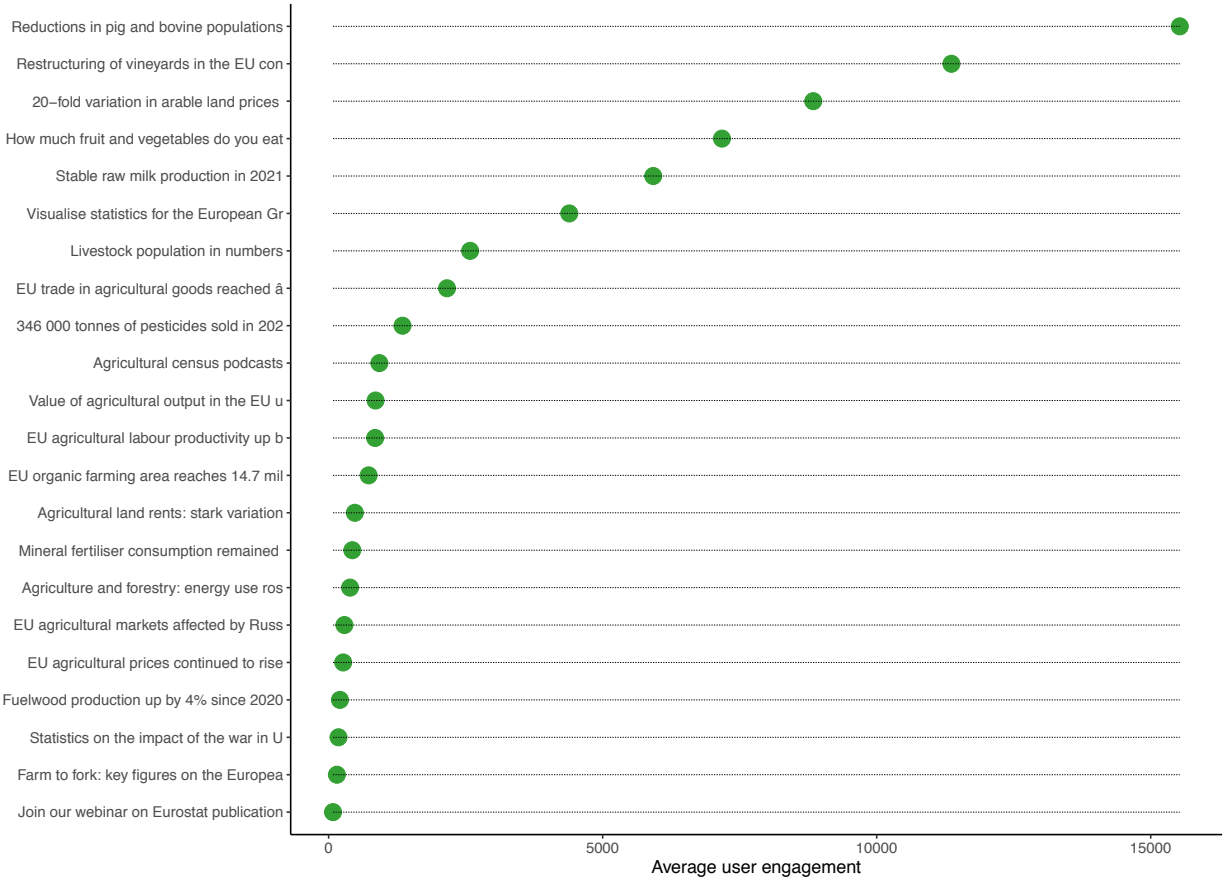



Several social media metrics are available to gauge the impact of the activity on the media promotion ranging from total number of impression to reach engagement rate. A very common and popular metric to pay attention is engagement[7] reflecting how much the user interacts with the content. Ranking the complete list of news items published in 2022 by the average number of engagement manifests that various general topics about livestock, crops and prices received a high level of engagement (see Figure 7). At the bottom, we observe low public engagement with different features, such as publications (farm to fork: key figures on the European food chain) and events (webinar on Eurostat publication), but also news items related to the current Ukraine crisis. However, it is crucial to stress that several determinants can affect the engagement rate. May be the posts are perceived less acutely by users, because the content of the post simply does not match the expectations or profile of the audience or wrong timing of the post on the channel.

[7] Social media engagement is the measurement of all possible interaction with the post such as likes, shares, mentions, saves or comments.

**FIGURE 7:**

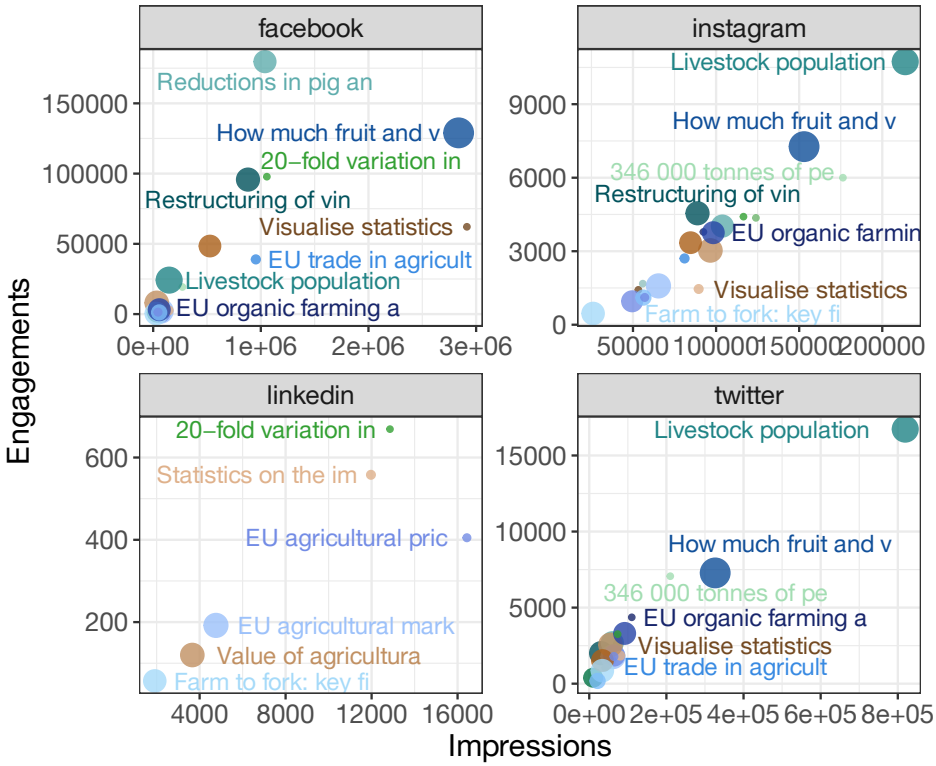
## Ranking news items by user engagements, average 2022




Viewing a single metric might not be informative enough to judge the impact of the media content on users. To build relationship and dialogue through the measurable engagements by users, it requires visibility and awareness. With an increase in the number of impressions – number of times a post is served to the audience– the potential pool of users engaging with the content is rising. Besides this fact, it might be of general interest to bring into relationship both concepts, dialogue and awareness. Figure 8 maps each news item on the engagement and impression dimension by distinct channels. While we observe that the pattern coincides with the presumption that dialogue increases with awareness, the distribution of news items is distributed heterogeneously across channels. Although, Instagram and Twitter seem to share a similar distribution of news items, it is considerably different for content posted on Facebook and LinkedIn. However, this is not surprising as each channel features contrasting audience and disparate approach to engage users. The results highlight to some extend the attentiveness of topics and contents of new items.

**FIGURE 8:**

## Dialogue and awareness of news items published in 2022 by distinct social media channels




*Note.* The size of the dot captures the total number of page views of the news item on a given day.

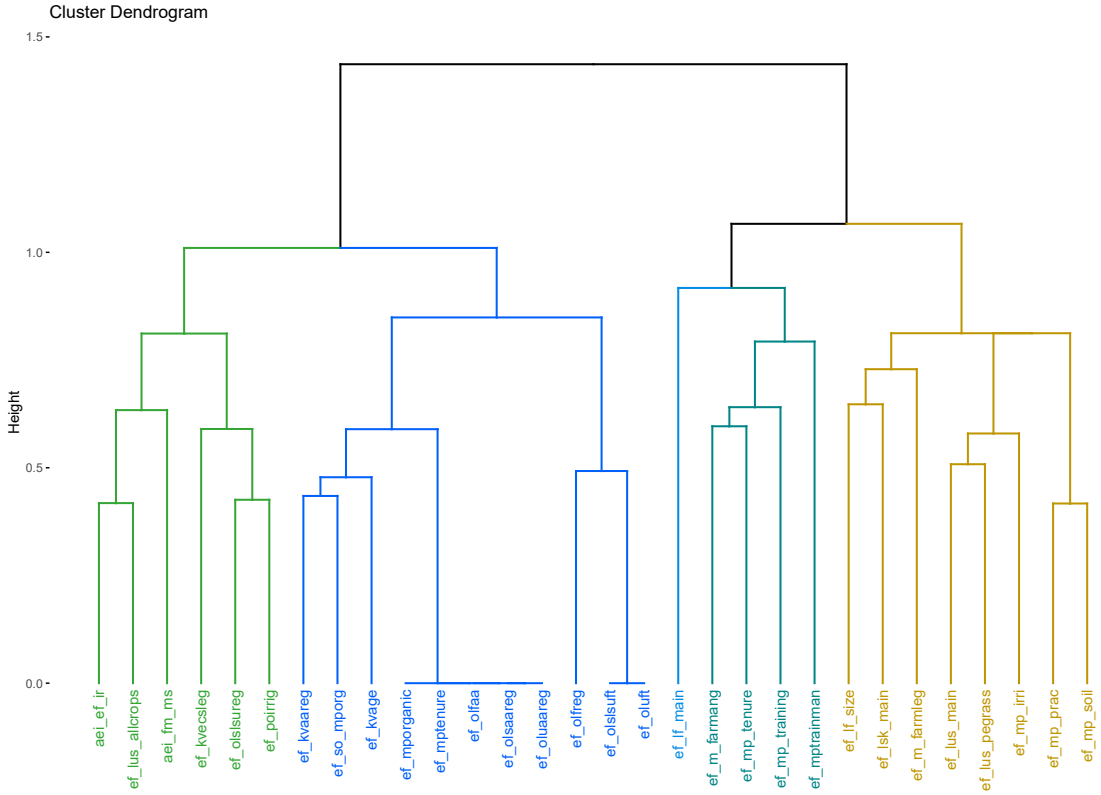
## 4.2. Modernising data dissemination

In light of the challenge to produce a considerable number of dissemination tables at the end of every survey year and fulfil emerging user needs, we identify the most prevalent tables based on page views, request from other EU institutions and needs for internal publications. We use farm structure survey as a use case and apply content analysis methods to explore the strength of corelationship between tables (N=234). This allows not only reducing the sheer number of data publications, but also to restructure and design new tables accommodating emerging user needs.

We design a table-dimension matrix of 30 most salient tables to determine the degree of similarity between tables based on matching coefficient of (Sokal, 1958) calculating distance matrices for binary data for each possible pair of combination. Let be the contingency table of binary data such as (n11 = a), (n10 = b), (n01 = c) and (n00 = d). Distances are of type () with s a similarity coefficient expressed as . The main idea of this technique is to identify similar tables sharing identical features. The similarity measure can be visualized by the means of hierarchical clustering and network analysis.

**FIGURE 9:**

## Hierarchical clustering of most important tables



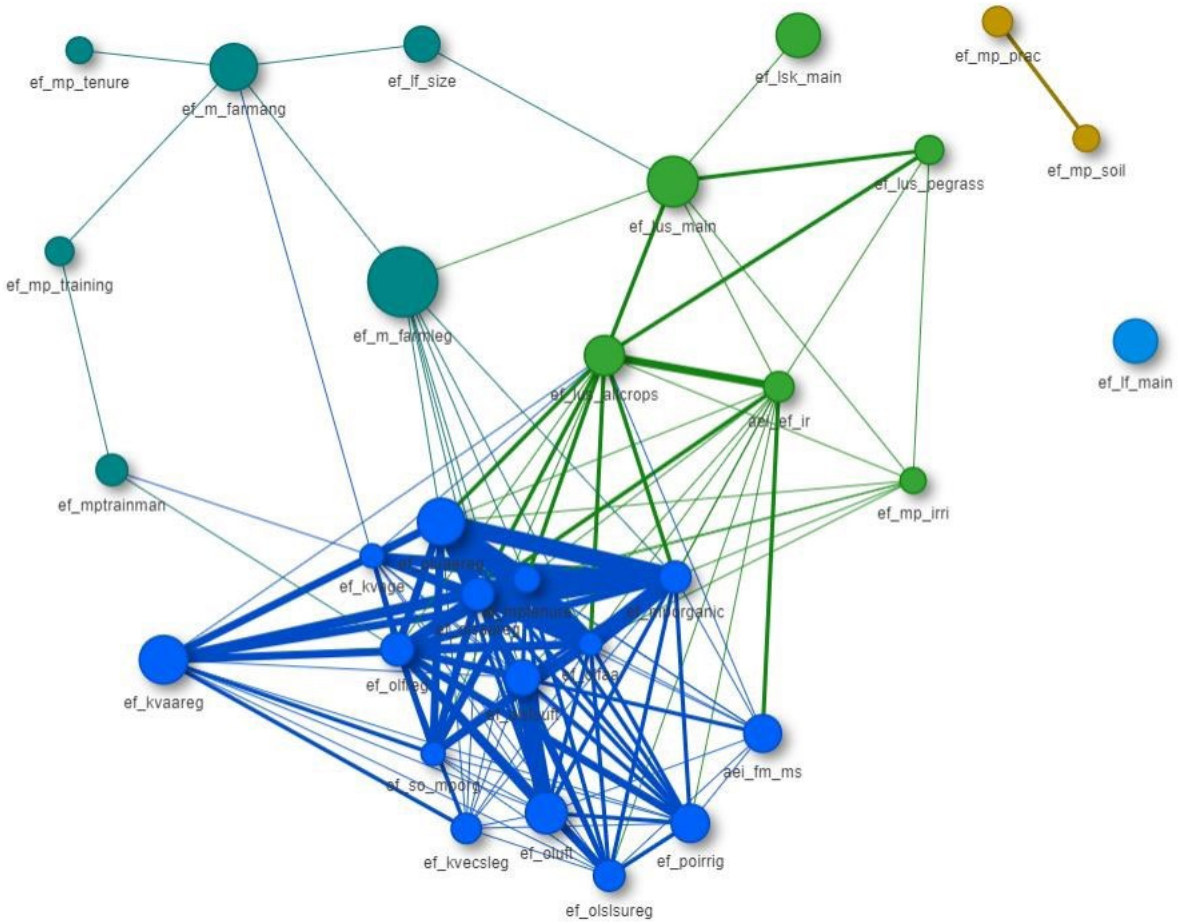

Depicted in Figure 9 is the divisive hierarchical clustering that is an unsupervised learning method applying the inverse of agglomerative clustering technique. It reveals that the tables can be classified into five main groups: "Agri-environmental indicators" (greenish), "Main Structural Indicators" (dark bluish), "Labour force" (light bluish) and "Crops and Management Practices" (goldish). The right part is formed by three subgroups whereas the group of "Crops and Management Practices" seems to be more heterogeneous in terms of topics such as organic farming, agricultural practices, land use and crop production.

To explore further the association between tables, network analysis provides additional information to the hierarchical clustering by visualizing the strength of corelation and the saliency of each table. Each node represents a table. The size of the node is based on the total sum of views on the webpage. The thicker the line is between nodes, the higher is the degree of similarity. Note that the network shows only links between nodes exceeding a similarity measure of 0.70. The main intention behind this condition is to diminish complexity and spot the strongest ties. Moreover, we leverage Louvain clustering algorithm to partition the data into different groups.

To some extent, the inference of the findings are similar to the hierarchical clustering. The two big clusters (forest green and cobalt blue) seem to coincide with the "Farm manager" and "Structural indicator" group displayed in Figure 9. There are very strong ties in the latter highlighting some needs to revise the content and structure of these tables. Besides that, we can observe a clear pattern in the network. With the highest level of user attention, the main tables, such as *ef m farmang* and *ef m farmleg*, connect different groups. Furthermore, they constitute the central node in the group linking more specialised tables. However, this is less obvious in the other parts of the network whereas the structure is more fragmented. The ideal network would be to generate a mixture of interconnected main tables that are subsequently divided into more specialised tables. In this way, users could benefit from a clear hierarchical structure and comparable consistent tables.

**FIGURE 10:**

## Network analysis



# 5. Conclusion

The modernisation of agricultural statistics is expected to engender many benefits for users ranging from more accurate and complete data publications to a wide range of articles and news items. It is evident that automation and a well-functioning IT system are the prerequisites to increase the volume of statistical outputs without impeding the overall quality of statistics. In particular, a stronger cooperation between domains will complement data gaps, alleviate inconsistencies and overall lead to a more efficient use of resources. In consequence, this might also create opportunities to release articles on crosscutting themes, such as sustainable or digital farming. The findings from webanalyics and social media of most accessible dissemination products reveal that user prefer generic rather than specialised themes irrespective on how the information is disclosed. This source of feedback can be viable in the design of dissemination strategy, especially for new areas, such as nutrition and pesticides where the data is scant and fragile.

# 6. References

Carletto, C., A. Dillon, and A. Zezza (2021). Agricultural data collection to minimize mea- surement error and maximize coverage. In *Handbook of Agricultural Economics*, Volume 5, pp. 4407–4480. Elsevier.

Castano, J. (2021). Integrating agricultural statistical operations for optimal data collection. Statistical Journal of the IAOS 37 (1), 333–339.

De Jager, T. (2015). Big data becomes big in agriculture: from the watchtower-current affairs. *FarmBiz 1* (4), 6–7.

Desbois, D. and A. Adam (2007). Stakes and pitfalls of the development of the statisti- cal systems in agriculture: the web of sustainable development. In *Annales-Economies, Soci´et´es, Civilisations*, Volume 3, pp. 553–570.

European Commission (2019). Strategy for european statistics and beyond 2020. Retrieved on 07.10.2020 at: https://ec.europa.eu/eurostat/documents/749240/749310/Strategy+on+agricultural+statistics+Final+version+for+publication.pdf/9c7787ca-0e00-f676-7a64-7f56e74ec813

European Commission (2022). Collaboration in research and methodology for official statistics. Validation in the ESS. Retrieved on 14.10.2022 at: https://cros-legacy.ec.europa.eu/content/validation-levels_en

European Commission (2023). Europa analytics. Retrieved on 02.02.2023 at: https://webanalytics.europa.eu/analytics

Heerschap, N. and L. Willenborg (2006). Towards an integrated statistical system at statistics netherlands. *International statistical review 74* (3), 357–378.

Hill, B. (1998). The implications for agricultural statistics of changes in the system of national accounts. *Journal of Agricultural Economics 49* (3), 359–377.

Jaakkonen, A.-K., J. Kyyrä, P. Mattila, E. Katajamäki, M. Yli-Heikkilä, H. Kastikainen, and M. Sulkava (2021). Modernisation of agricultural statistics.

Karlsson, A.-M. and M.-L. Widén (2008). Using administrative registers for agricultural statistics-methodologies, techniques and experiences. In *Proceedings of FAO Fourth International Conference on Agriculture Statistics, Beijing*.

Keita, N. and P. Gennari (2014). Building a master sampling frame by linking the population and housing census with the agricultural census. *Statistical Journal of the IAOS 30* (1), 21–27.

Liu, S. M. and Q. Yuan (2015). The evolution of information and communication technology in public administration. *Public Administration and Development 35* (2), 140–151.

Macnamara, J. and A. Zerfass (2012). Social media communication in organizations: The challenges of balancing openness, strategy, and management. *International journal of strategic communication 6* (4), 287–308.

Milics, G., I. Matečný, F. Magyar, and P. M. Varga (2022). Data-based agriculture in the v4 countries – sustainability, efficiency and safety. *Scientia et Securitas 2* (4), 491–503.

Pollach, I. (2011). The readership of corporate websites: A cross-cultural study. *The Journal of Business Communication (1973) 48* (1), 27–53.

Porumbescu, G. A. (2016). Linking public sector social media and e-government website use to trust in government. *Government Information Quarterly 33* (2), 291–304.

Selenius, J., C. Wirtz, D. Florescu, and A. C. Lazar (2021). Agricultural census 2020–how to reduce costs and burden? the european statistical system approach. *Statistical Journal of the IAOS 37* (1), 327–332.

Shekhar, S., P. Schnable, D. LeBauer, K. Baylis, and K. VanderWaal (2017). Agriculture big data (agbd) challenges and opportunities from farm to table: a midwest big data hub community whitepaper. *White paper for the US National Institute of Food and Agriculture*.

Shin, W., A. Pang, and H. J. Kim (2015). Building relationships through integrated online media: Global organizations' use of brand web sites, facebook, and twitter. *Journal of Business and Technical Communication 29* (2), 184–220.

Snijkers, G., T. Punt, S. De Broe, and J. G. Pérez (2021). Exploring sensor data for agricultural statistics: The fruit is not hanging as low as we thought. *Statistical Journal of the IAOS 37* (4), 1301–1314.

Snijkers, G., J. Pérez, and T. De Jong (2023). Innovating business data collection: System- to-system data communication applied to an agricultural survey. Retrieved on 03.04.2023 at: https://cros-legacy.ec.europa.eu/sites/default/files/book_of_abstracts.pdf.

Sokal, R. R. (1958). A statistical method for evaluating systematic relationships. *Univ. Kansas, Sci. Bull. 38*, 1409–1438.

Villani, R., T. Sabbatini, O. M. Perez, N. Guiomar, and M. Debolini (2019). An open dataset about georeferenced harmonized national agricultural censuses and surveys of seven mediterranean countries. *Data in brief 27*, 104774.

Wallgren, A. and B. Wallgren (2016). Administrative data and agricultural statistics–what strategy and methods should we adopt? In *Proceedings of the Seventh International Conference on Agricultural Statistics (ICAS-VII)*. FAO Rome.

Whitcraft, A. K., I. Becker-Reshef, C. O. Justice, L. Gifford, A. Kavvada, and I. Jarvis (2019). No pixel left behind: Toward integrating earth observations for agriculture into the united nations sustainable development goals framework. *Remote Sensing of Environment 235*, 111470.

# 7. Appendix

**TABLE 1:**

## Number of Eurostat tables by statistical domain

| Name of domain | Domain | Number of tables |
|---|---|---|
| **Agri-environmental indicator** | AEI | 49 |
| **Animal** | ANI | 61 |
| **Agricultural production of crops** | APRO | 27 |
| **Economic account of agriculture** | EAA | 84 |
| **Farm structure survey** | FSS | 234 |
| **Viticulture** | VIT | 55 |

**Getting in touch with the EU**

**In person**
All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

**On the phone or in writing**
Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
— at the following standard number: +32 22999696,
— via the following form: european-union.europa.eu/contact-eu/write-us_en.

**Finding information about the EU**

**Online**
Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

**EU publications**
You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

**EU law and related documents**
For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

**EU open data**
The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

PDF: KS-TC-23-003-EN-N

# Strategy to Modernise Agricultural Statistics: New Pathways for the Future

The growing number of challenges posed by emerging user needs, new data sources and technological progress calls for a rethinking of the way agricultural statistics are produced in the near future. This report outlines the current endeavours to modernise the core business processes of agricultural statistics within European statistical system. Moreover, we develop a user integrated dissemination process by proposing a modern approach to evaluate the statistical products as a function of user attention. Based on user metrics recorded between January 2020 and 2023, we explore page views and user engagements from social media of most accessible statistical dissemination products in the realm of agriculture. The exploratory analysis manifest that a short number of tables account for more than half of user hits within all domains. Linking user engagements from social media posts with page views of news items reveals a heterogeneous distribution of issue attention. While there is a strong and positive relationship between statistics explained articles and the associated tables, our results indicate that users prefer crosscutting rather than specialised topics. These findings are discussed within the modernisation framework.

**For more information**
**https://ec.europa.eu/eurostat/**

Publications Office
of the European Union

PDF: ISBN 978-92-68-06621-8