STATISTICS ESTONIA

Co-funded by the European Union

Project name: Development of Environmental Accounts

Project acronym: 2024-EE-EGD 101197994
WP2 Developing environmental subsidies and transfer's account

Deliverable 2.1.
Final methodological report

Raigo Rückenberg, Kaia Oras, Hans Hõrak, Grete Luukas

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them

30.12.2025

We would like to thank Eurostat for providing the grant to develop the Account and everyone contributing and consulting us on the topic. Special thanks go to Sjoerd Schenau from Statistics Netherlands for consulting us throughout the grant project. We would also like to thank everyone involved from Statistics Estonia for their contributions and all the stakeholders for their inputs and feedback.

1 Introduction

This methodological report describes the development work done under the Work Package 2 Environmental Subsidies and Transfers Account.

During the reporting period, Statistics Estonia completed the development work related to the Environmental Subsidies and Transfers Account (ESST) within the framework of the project Development of Environmental Accounts (2024-EE-EGD). The overall objective of this activity was to advance the methodological, technical, and practical foundations for the regular compilation of the account in line with Eurostat requirements.

The main objectives of the work were to:

Develop the Environmental Subsidies and Transfers Account, addressing methodological challenges and ensuring consistency with national accounts data on subsidies in terms of coverage and ESA transaction codes;

Compile the data and metadata for the Environmental Subsidies and Transfers Account, including the 2023 data and revised 2022 data for the first obligatory Eurostat reporting;

Automate the compilation process, for creating an efficient and sustainable system for future regular production.

All these tasks were successfully completed during the reporting period. The work aimed to streamline the management of environmental subsidies and transfers, ensuring accurate, consistent, and efficient data handling across reporting cycles.

The compilation of the account and the completion of the corresponding Eurostat reporting tables for reference years 2023 and revised 2022 were achieved as planned for the first obligatory reporting. Further cooperation and consultations with colleagues from National Accounts were undertaken. This work also covered the potential reclassification of enterprises by institutional sector, ensuring methodological consistency between the environmental accounts and national accounts frameworks.

To ensure the robustness of the methodology, experiences and approaches from other countries were analysed. Reports and documentation from other statistical organisations were reviewed, and consultations with Statistics Netherlands were held. These exchanges focused on issues related to subsidies classification, methodological harmonisation, and the implications of revisions in national accounts.

The account was compiled according to the Classification of Environmental Purposes (CEP), while maintaining parallel comparability with the earlier CEPA/CREMA framework. This dual compilation approach ensured continuity of time series and facilitated the transition to the new classification system.

Work was also completed on developing an automated data integration process, aimed at establishing a central database that gathers information on subsidies and transfers directly from administrative data sources. Within this system, data were automatically classified according to the rules of the Environmental Subsidies and Transfers Account (ESST), significantly improving the efficiency and reliability of the compilation process.

2 T2.1 Development of the Environmental Subsidies and Transfers Account

Based on earlier methodological work, challenges in compiling the Environmental Subsidies and Transfers Account were analyzed and addressed. Discussions with National Accounts happened throughout the year and consistency with National Accounts data was reviewed, particularly regarding coverage and transaction coding. Tax abatements topic was discussed with stakeholders to identify potentially relevant tax abatements in Estonia. Consultations with Statistics Netherlands supported the refinement of the methodology, especially concerning the classification of subsidies and the impact of revisions in national accounts.

3 T2.2 Compilation of data and metadata

The data for 2023 was compiled and data for 2022 was revised. Corresponding metadata was prepared. Data was submitted to Eurostat via the questionnaire provided by Eurostat and on the required CEP level – CEP level two. The process for the regular production of the account was further developed, including steps towards the

automatic gathering of data from administrative sources. Consultations with Statistics Netherlands were conducted to exchange experience and ensure methodological alignment.

Data for 2022 was mainly revised on the purpose of training machine learning tool for new CEP classification. However, in the process of applying CEP to 2022 data, other aspects were revisited and changed, too. Although the methodology for the compilation of 2022 and 2023 ESST account remained largely the same, certain transfers in 2022 data were corrected. First and foremost, the corrections made to the 2022 data were made to ensure better consistency in the ESST timeline. Furthermore, the data rework revealed error-prone scenarios, providing a basis for improved identification and mitigation in future processes. This already proved to be useful when working the 2023 data.

Whilst working on applying CEP, transfers were spotted that were assigned incorrect CEPA/CReMA. This was fixed simply by applying correct CEP to the transfers. However, some transfers were removed from the compilation of 2022 ESST data after revision. For example, removing transfers which were assigned environmental domain by mistake during the original compilation of the data, but which fall out of scope of ESST or CEP. Conversely, small number of transfers were added to ESST compilations – transfers which were missed during the original compilation.

Additionally, there were some projects recorded in the 2022 dataset that had either ended or not yet started – such cases were related to lump sum projects deriving from CINEA. This means the sum of Rest of the World transfers reduced significantly compared to the original data submission for 2022 (Table 1)

Table 1. Comparison between original and revised data for the year 2022, million EUR

ESA transfer type	Original 2022 data	Revised 2022 data
Subsidies (D.3)	99,66	99,66
Other current transfers (D.7)	143,21	44,14
Capital transfers (D.9)	125,80	108,28

As seen in the table above, significant change occurred in D.7 transfers as a result of 2022 data revision. These were mainly CEP 07 transfers – projects focused on research and development. Integrating CINEA and classifying CINEA data is complicated due to the different structure of the data and lack of relevant information for ESST (such as recipients business registry number). Lump sum funding means that all the financing was divided across the duration of the project – this practice was applied after discussions with Statistics Netherlands. However, this means that discrepancies occur between ESST and NA. Dividing project costs between several recipients and several years caused calculation errors which lead to a significant overestimation of CINEA transfers during previous grant project. As these errors were corrected during this grant project, extra caution was taken when including CINEA projects into 2023 ESST dataset and errors made during the previous grant project were avoided successfully.

Some other issues with 2022 were also corrected when spotted – such as correcting the ESA transaction code or recipients NACE and removing double entries from the dataset. However, such cases were rare and did not have a significant effect on the revised 2022 results.

In general, reworking 2022 data during this grant project proved to be useful. As larger methodological questions were solved during the previous grant project, the focus could shift on quality and processes of the compilation of ESST account. Reworking old data highlighted the bottlenecks of the previous compilation and appropriate approach was taken for the compilation of 2023 data, mitigating similar errors and making the process smoother. Suspicious transfers could be identified early on into the compilation of 2023 dataset and corrections processes were faster. For the better consistency of the ESST timeseries, 2022 final data was linked to 2023 raw data. This allowed multiyear projects to be classified in the same CEP category from year to year. Furthermore, this helps to explain yearly changes in the ESST results, as this allows to observe which subsidies schemes are gradually phased out or launched by rest of the world or general government.

3.1 Compilation of data for 2023 and results

The data compilation for the year 2023 was done manually for the most part. Although the focus was on developing automatic solutions for the compilation of ESST, the IT solutions were not completed in time to fully apply them. New R scrips were written to accommodate data from new data sources or scripts from the last grant projects were used and further developed so that they could be used in the 2023 dataset compilation. The preliminary version of Oracle database created during the previous grant project for the ESST account compilation was used and further developed to accommodate new data. Though not fully developed, the available IT solutions helped to save time during the compilations.

During the grant project new data sources became available and were included in the ESST account. More specifically, subsidies related to forestry became available. However, the data was not available retrospectively, so 2022 data could not be updated. Work was also done to eliminate overlap in administrative data – overlapping data was removed on the grass root level by updating contracts with data holders. This was done to ensure duplicate data does not reach Statistics Estonia in the first place. As less data was requested, it should also lower the burden on data holders in the future. The more detailed data acquired allowed for better allocation of transfers according to CEP, NACE and transfer type. New version of statistical profile was also used, which provides more up to date and accurate data for assigning NACE and institutional sector. This helps to improve consistency with National Accounts.

Additionally, COFOG data was used to determine transfers originating from local governments (S.1313). It was determined during this grant project that local governments transfers are not covered by administrative data, so transfers from S.1313 within COFOG 05 could be included, in addition to D.74 transfers. The addition on S.1313 transfers is marginal, less than 1% of ESST total, but for more accurate coverage, it is now included in ESST account. CINEA data was re-studied and re-worked during 2023 data compilation to assure better quality of reporting. Due to the nature of CINEA data, it remains to be difficult to work with and extra care has to be taken when adding CINEA data to ESST dataset to mitigate issues experienced with 2022 data.

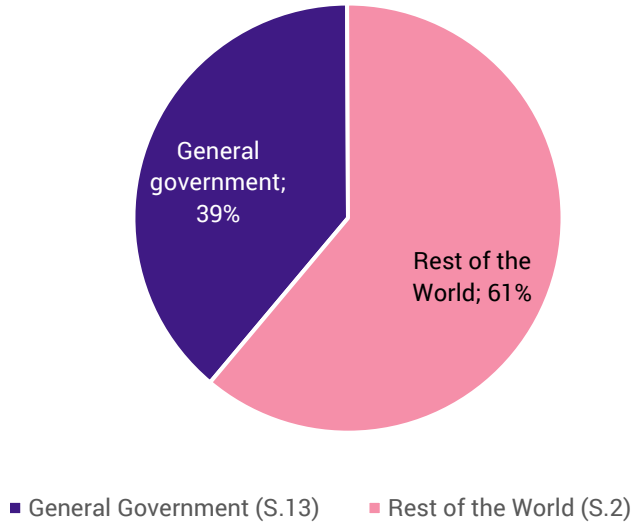
Tax abatements were re-examined as part of this grant project, continuing the approach taken in previous projects aimed at developing the ESST account. As established during the previous grant projects, tax abatements in Estonia have no significant relevance to ESST. Instead, financial mechanism for environmental and resource management are mainly related to direct subsidies in Estonia. Currently, there are no tax abatements to report to Eurostat in the scope of ESST. For the future, SE continues to monitor the situation regarding tax abatements in the scope of ESST.

CEP was used for the first time in Estonia to allocate environmental domain to transfers. Data was submitted to Eurostat on CEP level two, as requested by Eurostat in the questionnaire. As part of the grant objective, data for 2022 and 2023 was compiled on the most detailed level of CEP (third level). As this was the first time CEP was used in Estonia and the allocation was done on the most detailed level, it was quite difficult and time-consuming process. For the analysts required to used CEP, it took some time to get familiar with CEP. Classifying transfers on third, most detailed level, added to the complexity of applying CEP. In general, the detailed administrative data allows for applying CEP on most detailed level. For COFOG data, assigning CEP on third level is problematic – counterpart details were taken into consideration when applying CEP on the most detailed level, e.g., counterpart NACE and institutional sector. It is not a large issue as the COFOG transfers make up a marginal percentage of ESST total in Estonia.

3.1.1 Results

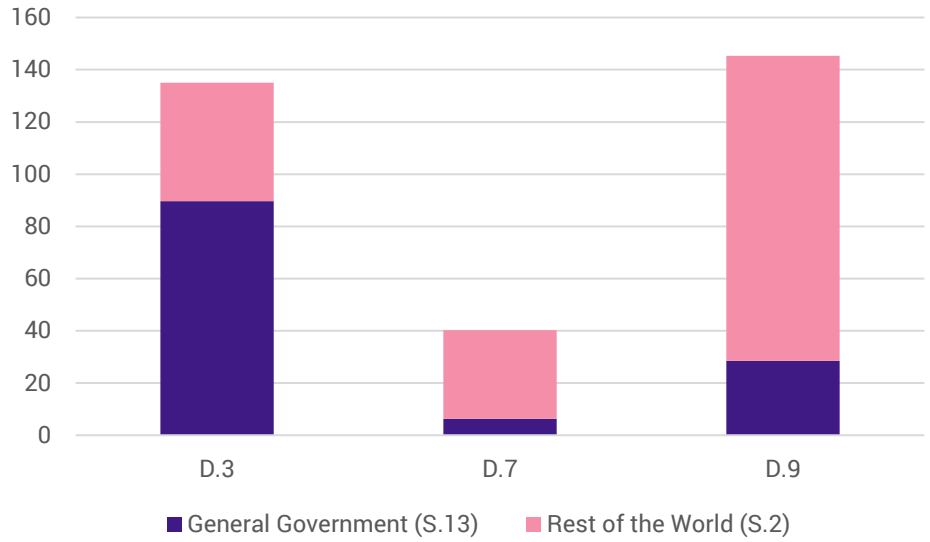
Compilation of the 2023 data followed largely the same methodology as in previous years. Total subsidies for 2023 are 320 million Euros. As in previous years, Rest of the World (S.2) provides over half the subsidies (Table 2). This is consistent with the results from previous grant projects, for the reference years 2020 and 2022. During the development of ESST account in Estonia, more precise data has been acquired, which allows more precise allocation of General Government and Rest of the World subsidies. For most transfers, the final recipient's own contribution is available, too. Final recipients own contribution is not in the scope of ESST, but it is used in the compilation of EGSS and EPEA, when applicable.

Table 2. Transfers by Rest of the World and General Government, 2023, percentage



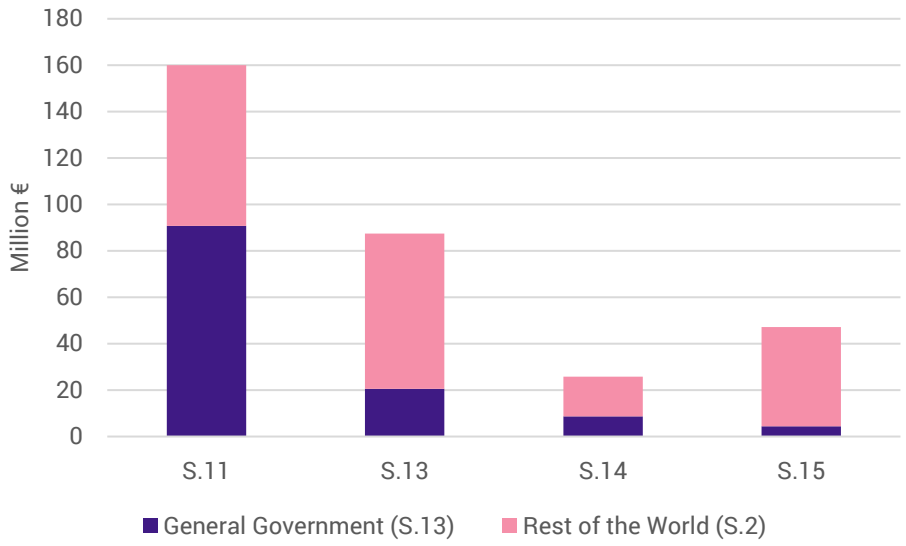
Transfers by ESA transaction type are dominated by subsidies (D.3) and capital investments (D.9) (Table 3). General Government accounts for most subsidies (D.3). These subsidies are largely directed toward production of renewable energy and the combined production of renewable and non-renewable energy in cogeneration and trigeneration plants. Rest of the World provides subsidies (D.3) for agriculture (CAP) but also contributes towards capital investments (D.9), such as energy efficient buildings (renovation/reconstruction). Other current transfers (D.7) are mostly made up of Rest of the World transfers for research and development (CINEA projects).

Table 3. Environmental subsidies and similar transfers by ESA transaction code, 2023, million EUR



Corporations (S.11) received the most subsidies in 2023, both from General Government and Rest of the World (Table 4). This is expected, as corporations receive transfers for a large range of CEP activities. General Government (S.13) and NPISH (S.15) receive largest number of transfers for energy efficiency particularly. These transfers are mostly related to renovation and reconstruction of buildings. While households (S.14) receive notable support for energy efficiency, most of the transfers they receive are related to agriculture – organic farming and soil protections measures.

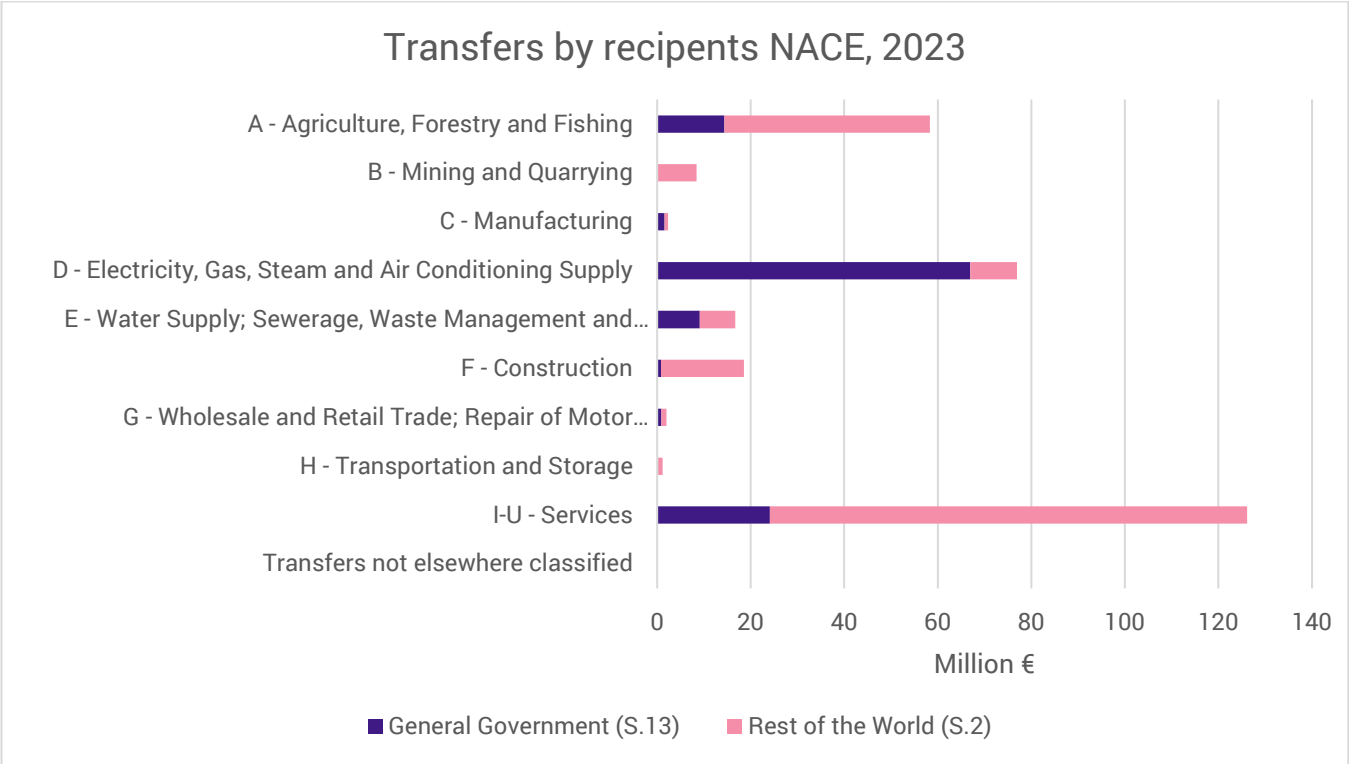
Table 4. Transfers by recipients institutional sector, 2023, million EUR



Agriculture, forestry and fishing (NACE A) and electricity, gas, steam and air conditioning supply (NACE D) receive substantial number of transfers (Table 5). This is logical as agriculture subsidies and subsidies for production of renewable energy are prominent in Estonia. However, the highest number of transfers are to services sector (I – U). This might look odd at first, but it can be explained by capital transfers for energy efficiency – general government offers subsidies to local governments and NPISH for renovations.

Transfers not elsewhere classified is zero for Estonia – that's because of the high detail administrative data available that allows classifying all final recipients by NACE.

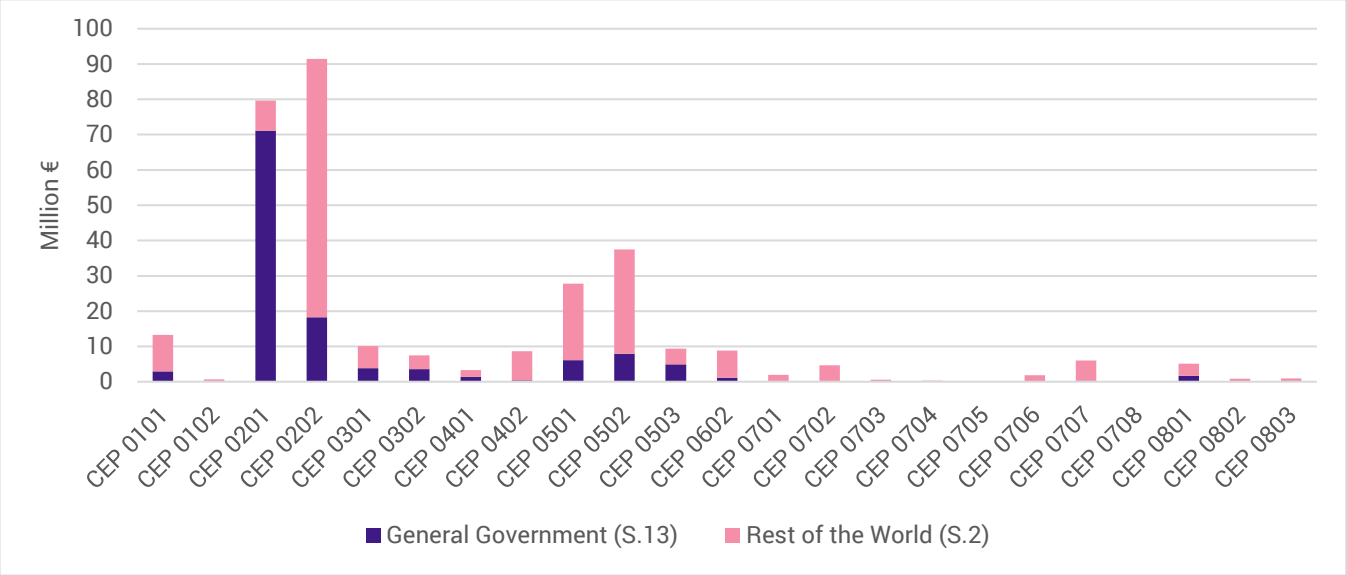
Table 5. Transfers by recipients NACE, 2023, million EUR



Transfers for production of renewable energy (CEP 0201) and energy efficiency (CEP 0202) are most prominent when observing transfers by CEP (Table 6). Subsidies for protection of soil (CEP 0501) and biodiversity (CEP

0502) are also relevant – these originate mostly agriculture subsidies for organic farming or for farming measures related directly to protection of soil, surface or groundwater. Research and development (CEP 07) also receive a fair share of transfers – these originate mostly from Rest of the World and are large multiyear projects. General Government does contribute to cross-cutting environmental purposes (CEP 08) – mostly education and training programs for schools.

Table 6. Rest of the World and General Government transfers by CEP, 2023, million EUR



3.2 Data acquisition and availability in Statistics Estonia databases

As the compilation of ESST account in Estonia relies largely on administrative data, part of this grant project was dedicated to secure proper data flow to and within Statistics Estonia to compile the ESST account. During the previous grant project, data holders were identified and where possible, data was acquired via requests. This meant that the data acquisition relied on voluntary cooperation from data holders.

Although Statistics Estonia has good relations with data holders and data was successfully obtained to compile the ESST account during the previous grant project, it was clear that inquiring data on yearly basis and counting on the voluntary cooperation from data holders was not a sustainable solution for the future. Statistics Estonia would have no guarantees regarding data transmission and data holders would face unnecessary administrative burdens processing and answering Statistics Estonia data requests.

Statistics Estonia has a set of procedures in place to guarantee a sustainable dataflow from data holders, and within SE. For data holders, this meant legally binding contracts and obligations to provide data to SE, but also clear terms for data provision. Within SE, it means that all the data received was described and metadata attached to it. As such, projects were started in Statistics Estonia to acquire data according to the official procedures to guarantee a sustainable solution for data collection for the foreseeable future.

During this grant project, contracts are in progress of signing Environmental Investment Center (EIC) and The register of state aid and de minimis aid (RAR). We made official data request to Elering and Estonian Private Forest Centre but there has been changes of data holders and due to that we have new data provider as RAR and data from Estonian Private Forest Centre we are involving to contract with EIC.

In addition to providing Statistics Estonia a fluent data flow, acquiring data according to the official procedures, it is possible to automize data cleaning and some data processing features for the compilation of ESST account.

3.2.1.1 Gathering of administrative data

If the necessary data was not available at Statistics Estonia or was not with the required frequency, the missing administrative data had to be ordered to Statistics Estonia. Before new data could be ordered, the leading analyst of the environmental statistics team had to make sure that the necessary data was not already available at Statistics Estonia. It was the task of environmental statistics leading analyst to determine from which institutions, from which datasets and which data fields should be ordered.

To process the request for new data, or make changes to the existing datasets, a project had to be created, described and approved first. The project description had to include its purpose, scope, legal basis, planned outcomes and impacts. Furthermore, the description had to address the problems the project aimed to solve and include any limitations or risks that might occur and affect the project. If there is no project or the project does not pass the internal review, SE will not contact data holders for any data.

Once the project was approved, a JIRA task (epic) along with subtasks was created to plan and monitor the progress of the project. All necessary communications and information were exchanged in JIRA, overseen by the leading analyst from the environmental statistics team.

Every data provider will get unique JIRA task (epic). This is important because every dataset needs concreate actions which are necessary to make data collecting correct, smooth and as automated as possible.

Generally, bringing data into the organization took at least 4 months after project approval, and in some cases, even 6-8 months if there were legal issues, particularly related to data protection or change of dataset holder.

3.2.1.2 Description of the requests of administrative data

To order the necessary data for ESST compilation, the leading analyst from environmental statistics team informed the coordinator from administrative data team about which institution or register and from whom (person or department) the data is needed.

The purpose and justification for requesting the data had to be provided. This included the name and number of the statistical work, citation to regulation 691/2011, but also the reasoning why it was not possible to compile ESST without requested data.

The scope of the request had to be described precisely by the leading analyst from the environmental statistics team. This included the frequency of data transmission, the observation period, the date of first data transmission, whether the data was needed retrospectively or for future periods only, fixed term or indefinite request or contract. One big part of input is to define variables – name and metadata description. Metadata descriptions were created in cooperation with data provider.

Along with the terms of the request, the exact data composition needs to be described by the leading analyst from the environmental statistics team. Conditions like description of variables in the dataset and data extraction conditions had to be submitted to the administrative data team designer. More precisely described data allows for easier negotiations with data holders. If the administrative data team coordinator is more related to the gathering of data and communication with data provider. Designer is employee who designs dataset tables, variables (format) and this input is necessary for automated process of data flow.

At first, Environmental Investment Center (EIC), State Shared Services Center (SSSC) and Elering were contacted for data acquisition. However, during the data acquisition negotiations it was discovered that the data in sufficient detail for the compilation of ESST was in fact available from another data source. This meant that the requests had to be resubmitted to the correct data holder and negotiations had to be started with another data holder. This was the case for forestry and renewable energy subsidies. For data related to forestry subsidies, Estonian Private Forest Centre no longer provides the data and data had to be acquired from EIC. Similarly, Elering does not have detailed enough data for renewable energy subsidies, so the focus was turned to the register of state aid and de minimis aid (RAR). This meant that new contracts had to be drafted, new data tables (FOR) had to be created by the designer and metadata had to be created for newly acquired data.

An important aspect of this project was to determine the most convenient data submission channel for providers and, where possible, for Statistics Estonia (SE). During the grant project, multiple submission channels were used:

- Email with attachments
- Web-based secure environment
- Direct export from the data provider’s system

Email and system exports were handled through manual data capture, while the web-based secure environment enabled an automated data capture process.

When a dataset was submitted via the secure environment, it was integrated into the fact table (FOR) within minutes, allowing analysts to begin work immediately. If additional processing or calculations are required, the dataset is automatically moved to a process table (DSA). These process tables are temporary; once processing is complete, the results are stored in the fact table, which serves as the foundation for statistical analysis.

Automatic data submission by data holders was preferred, but in some cases the data holder did not have the resources to enable automatic data transmission to Statistics Estonia and simpler solution had to be used.

3.2.1.3 Confirmation of data composition and agreement

Once the data needs had been documented by the lead analyst of the environmental statistics team and the administrative data team had no further questions, the administrative data team contacted the data holders via email. In the email, SE explained the need for the data/dataset, from which data collection the data was needed, the required frequency, and the data composition. The data provider might have considered SE email as a clarification request and responded according to the law within 30 calendar days.

Once an agreement has been reached to receive sample data and the data has been sent to SE, for example, via email or through secured web environment for data submitting, it was loaded into the sample data source database (Final Observation Register – FOR) and, if necessary, anonymized - transformed so that it cannot be directly identified, i.e., pseudonymized. If the data was received via X-Road, it was parsed (transforming XML data into Oracle flat table format).

From the received data, SE verified whether the data composition was correct and all requested elements, variables and objects, were present; whether the received columns had the agreed-upon titles; whether the values of variables in the data matched the agreed lists (including spelling); whether the values met the requirements, including data types, etc.; and whether the data was available for linking and identification. The presentation formats of date type data were also checked.

Based on the analysis of the sample data, if needed, SE communicated with the data provider to improve data quality and then proceeded with formalizing the data transmission contract.

All the contracts signed between SE and data holders were reviewed by lawyers from both parties to ensure that all the legal basis were covered.

3.3 Description of metadata

For Environmental subsidies and similar transfers accounts Statistics Estonia has compiled ESMS-based quality and metadata report (https://stat.ee/en/find-statistics/methodology-and-quality/esms-metadata/10108).
The base for the report is compiled in Excel by metadata expert based on the information available from existing sources. Afterwards the owner of a ESST statistical process complemented the report with specific information. After reaching the agreement on the content of metadata the report was translated into English. Then the Publication Team published the report on Statistics Estonia's website.

Describing quality and metadata is a core requirement of official statistics and also for ESST implementation.

Agreed principles:

- Metadata document contains:
  - Methodologies
  - Data sources
  - Transformations and adjustments
  - Assumptions (e.g., discount rate, assortment model, mortality model)
  - Accessibility and clarity
  - Quality indicators
- Metadata meets the standard of official statistics, given that forest accounts will be produced regularly under EU regulation.

In Statistics Estonia quality and metadata reports are produced about every statistical process which output data are published in SE statistical database. Quality and metadata reports must be compiled to ensure the transparency, clarity and understandability of the statistics.

In order to ensure comparability of the statistics between member states of the European Union we follow the European Statistical System (ESS). In the ESS there are 3 structures for compiling quality and metadata reports: Single Integrated Metadata Structure (SIMS) and it's underlying reporting structures EURO-SDMX Metadata Structure (ESMS) and ESS Standard for Quality Reports Structure (ESQRS). All of them are output-oriented way of viewing statistical processes but describe other stages of statistical production (e.g design, collection, processing, dissemination) as well. ESQRS is a producer-oriented structure comprising more quality indicators and it's level of detail might be too specific for general users. ESMS is a user-oriented structure which main focus is on the methodology. SIMS is so called all-inclusive report which comprises both ESMS and ESQRS metadata elements (concepts).

Until 2025, SE reports have been based on the Euro-SDMX Metadata Structure (ESMS).

In 2017 the European Commission recommended all member states to implement SIMS in order to streamline and harmonise user and producer reports in the ESS, decrease the reporting burden on statistical organizations by creating the framework that enables once for all purposes reporting, where concepts that are common to user and producer reports are reported upon once for both purposes, create an integrated and consistent reporting framework where the reports can be stored in a single database; create a flexible and up to date system where future extensions are possible by adding new concepts.

As Statistics Estonia is in the middle of the transition period and therefore publishing the SIMS reports on our website is under development, metadata for ESST was created according to both SIMS and ESMS.

4 T2.3 Automation of the compilation process

The development of an IT-based process for automated compilation of the Environmental Subsidies and Transfers Account was completed. The workflow allowed environmental subsidies and transfers to be gathered, processed, and classified within a unified system.

In preparation for machine learning applications, historical datasets were reclassified from CEPA/CREMA to the Classification of Environmental Purposes (CEP), enabling the machine learning model to adapt to the new classification system. This transition supported future automation and improved classification accuracy.

Consultations with Statistics Netherlands provided valuable input for strengthening the automation process and ensuring compliance with international best practices.

4.1 Machine learning

Government subsidies for environmental or other purposes leave a paper trail of documentation where the goals and activities related to the funding are described. These text documents and the related funding provide a basis for statistics on state subsidies by purpose. In the previous grant project 101113157 — 2022-EE-EGD we reported on classifying such documents according to the CEPA/CrEMA classification scheme. We achieved great utility with the fastText algorithm (Bojanowski et al., 2017a; Bojanowski et al., 2017b; Mouselimis, 2024) leveraging pre-trained Estonian word vectors (Grave et al., 2018) and a two-step classification scheme where in the first step

[1] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017a). Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*: Volume 2, Short Papers, 427–431.
Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017b). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5, 135-146.
Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. In *Proceedings of the eleventh international conference on language resources and evaluation* (LREC 2018).
Mouselimis, L. (2024). fastText: Efficient Learning of Word Representations and Sentence Classification using R. R package version 1.0.4, https://CRAN.R-project.org/package=fastText .

any environmental subsidies are separated from all other purposes and in the second step, the environmental purposes are classified according to the classification scheme. As the continuation of this work we explore how well the method applies to the new CEP classification scheme and how much manual annotation or editing is required to achieve acceptable accuracy.

State policies can change quickly and this change can be aligned with the calendar (funding measures opening and closing around the end/beginning of the year). This poses risk for a machine learning-based statistics production method where predictions are made with a model trained on data from a previous year with possibly very different policy goals, funding areas, and (buzz)words used in project proposals (data drift). Here we assess how well a text classification model trained from the previous year's subsidies data can generalize to the next year in CEP classification. We designed a machine learning experiment to assess the quality of CEP classification based on previous year(s) data alone versus including some portion of manually annotated data from the new year. We trained fastText models from 2022 manually labelled data and made predictions on 2023 data. Then we included 10%, 20%, 30% and 40% of unique texts from 2023 to the training data to see model performance on the remaining 90%, 80%, 70% and 60%. We run the same experiment also with a strategy of searching texts of rare categories from earlier years assuming what was rare in CEPA/CrEMA will also be rare in CEP (189 texts from 2020-2021 were found for 22 categories at CEP3 level and 15 at CEP2). Since the sizes and class representations of training and test sets are dynamic across the training runs, a separate hold-out validation set is used in addition to the dynamic test set, but due to extreme class imbalance, this validation set is very small with most classes including only one or two examples, so the results must be interpreted carefully. Two data sources available for 2022 were not available for 2023 for this experiment (PRIA and Elering) so the experiment only covers the document registers which were available for both years (SFOS).

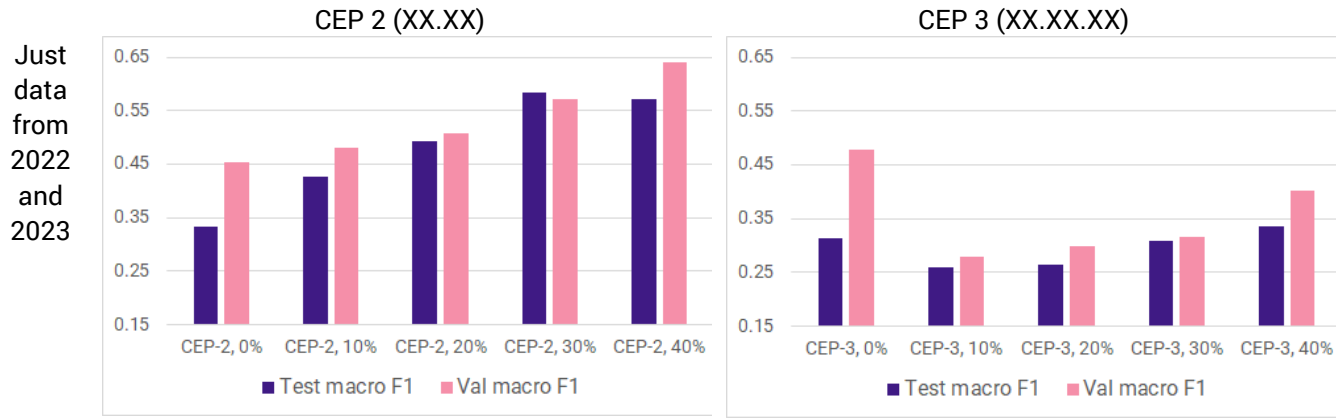
Table 7. Training, test and validation dataset sizes (number of classes).

	Just 2022 and 2023 data			Including rare classes from past		
Run	**Train**	**Test**	**Val**	**Train**	**Test**	**Val**
CEP-2, 0%	1283 (23)	2207 (22)	43 (16)	1472 (24)	2209 (22)	43 (16)
CEP-2, 10%	1503 (23)	1987 (22)	43 (16)	1692 (24)	1989 (22)	43 (16)
CEP-2, 20%	1724 (23)	1766 (21)	43 (16)	1913 (24)	1768 (22)	43 (16)
CEP-2, 30%	1945 (23)	1545 (20)	43 (16)	2134 (24)	1547 (22)	43 (16)
CEP-2, 40%	2166 (23)	1324 (20)	43 (16)	2355 (24)	1326 (22)	43 (16)
CEP-3, 0%	1283 (43)	2133 (47)	43 (27)	1472 (46)	2135 (47)	43 (27)
CEP-3, 10%	1503 (49)	1967 (45)	43 (27)	1692 (51)	1973 (47)	43 (27)
CEP-3, 20%	1724 (50)	1763 (44)	43 (27)	1913 (52)	1766 (46)	43 (27)
CEP-3, 30%	1945 (51)	1544 (41)	43 (27)	2134 (53)	1546 (43)	43 (27)
CEP-3, 40%	2166 (51)	1323 (41)	43 (27)	2355 (53)	1325 (42)	43 (27)

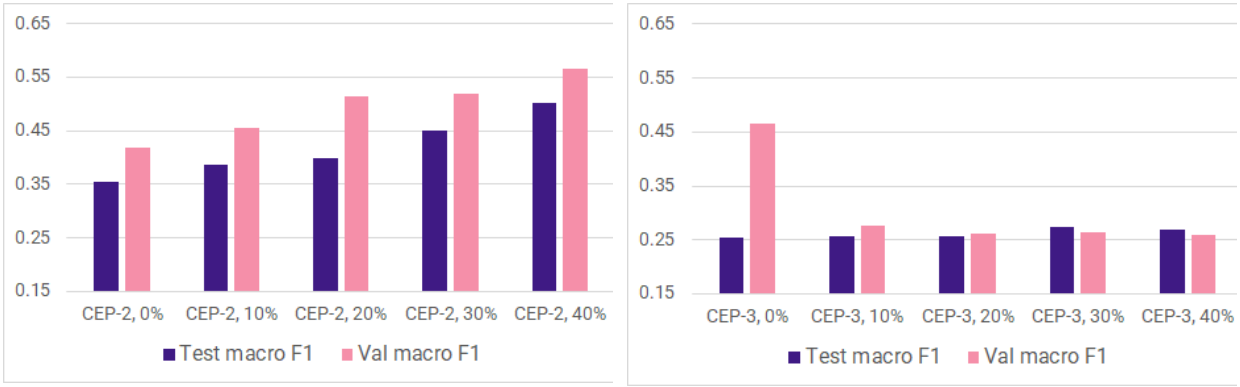
The results are somewhat contradictory: the simpler task of CEP 2 classification (22-24 classes) benefits significantly from including 10% of unique texts from 2023 but the more difficult task of classifying CEP at the most detailed level (CEP 3, 47-53 classes) actually shows slightly better performance when no data from 2023 are included.

Figure 1. Total accuracies (accuracy of binary model × accuracy of the multiclass model).

Figure 2. Macro F1 scores of multiclass models (rows where F1 could not be computed are removed).



Extra 189 rare class samples from past



The confusion matrices (figures 3,4,6,7) show clearly where the confusion appears in CEP 2 classification – without including data from 2023, over a half of CEP_0101 texts (reduction and control of greenhouse gases) are confused with CEP_0401 (waste management). This confusion disappears when including just 10% of the data from 2023. The class distributions on the consecutive years (figure 5) also explain the confusion in CEP 2 classification: the model is initially trained from data where only 1.2% of texts represent greenhouse gas emission reduction while at the prediction year such texts make up a quarter of unique texts. However in CEP3 classification CEP_0101XX categories are not confused with CEP_0401XX categories at all whether any data is included from 2023 or not. Furthermore, the performance on the small hold-out validation set from 2023 actually decreases when including data from 2023.

Figure 3. Confusion matrix for CEP 2 including 0% of unique data from 2023 into the training set.

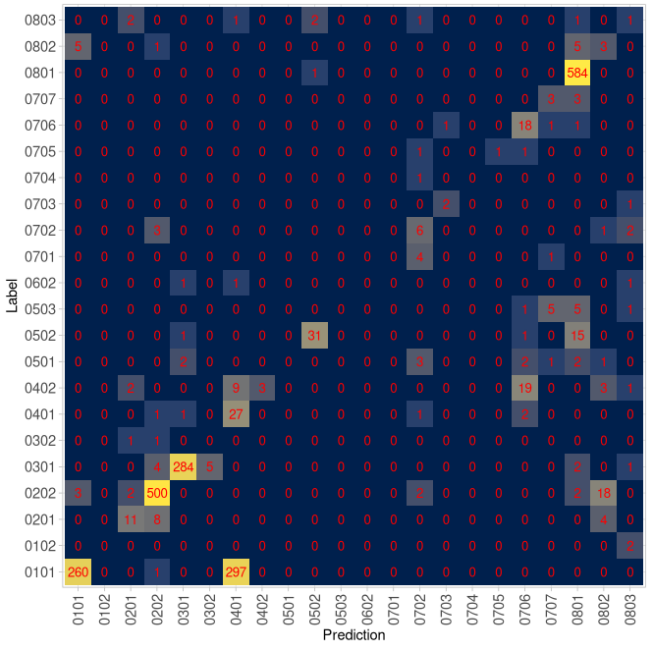


Figure 4. Confusion matrix for CEP 2 including 10% of unique data from 2023 into the training set.

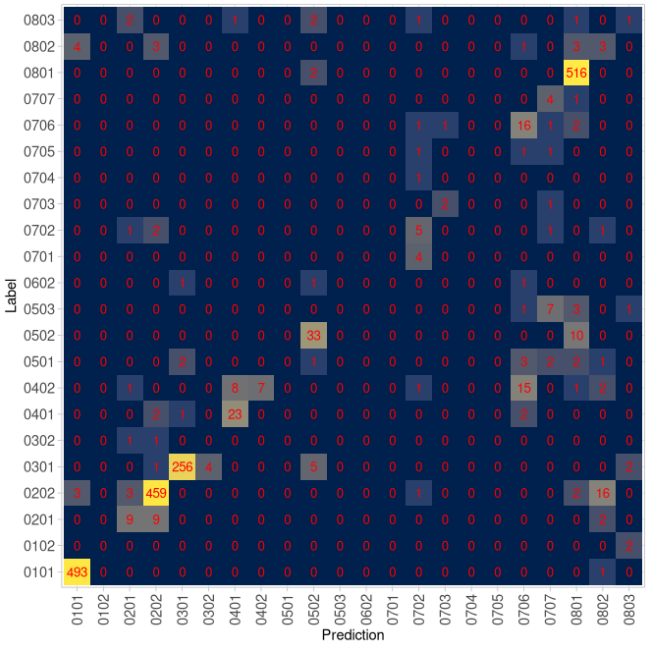
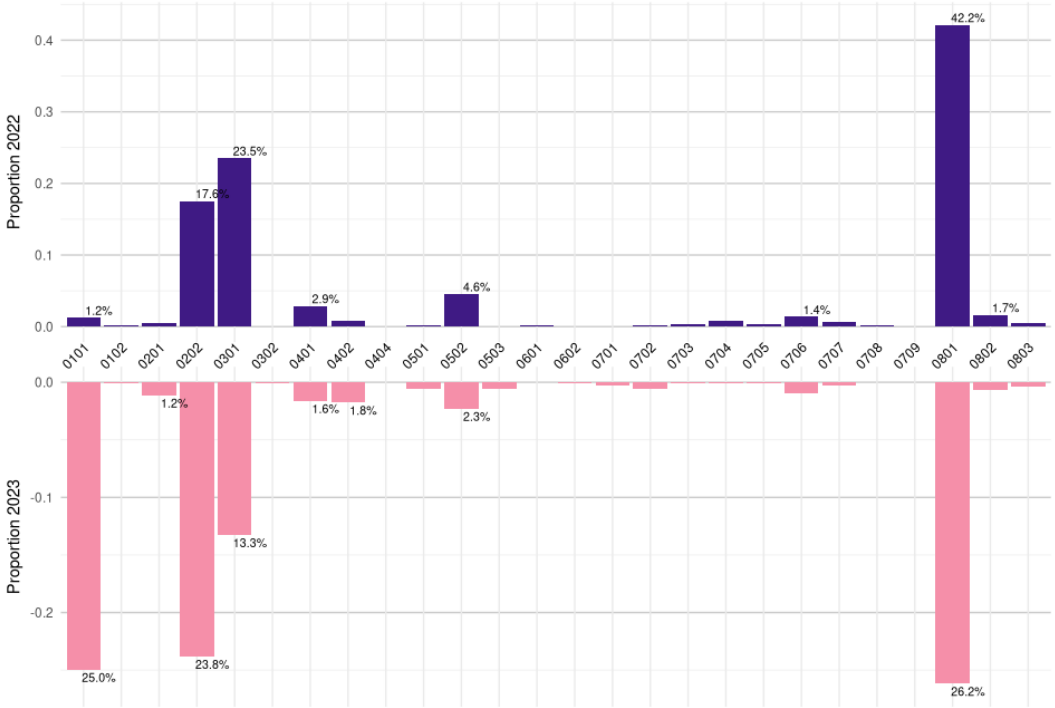


Figure 5. CEP 2 class representations in unique texts 2022 (train) vs 2023 (predict).



One possible explanation to this anomaly could be different people labelling the datasets of 2022 and 2023. If these two people have a similar understanding of CEP 2 categories but diverging understanding at the most detailed level, then the class-text relations could be homogenous across the years at CEP 2 level but not CEP 3. When including data from 2023 for CEP 3 classification, the model can be confused by conflicting labels for similar text patters produced by different annotators thereby reducing overall performance. Such systematic label errors can also influence performance on other classes.

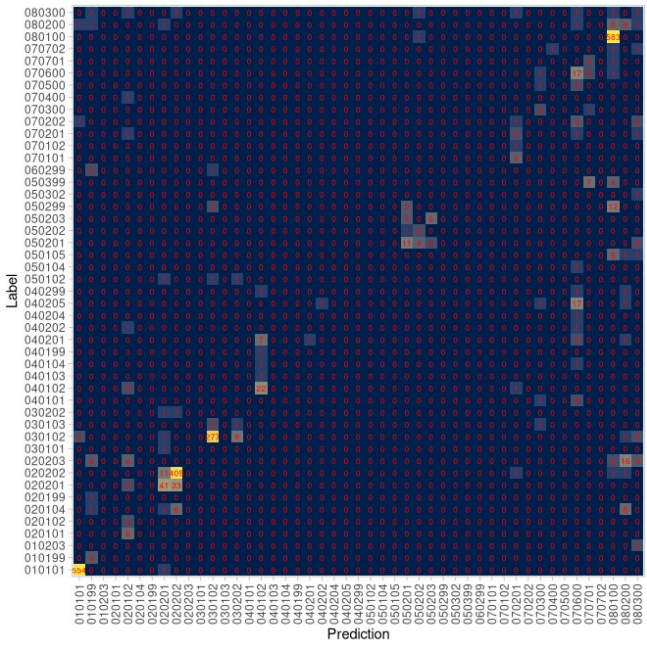


Figure 6. Confusion matrix for CEP 3 including 0% of unique data from 2023 into the training set.

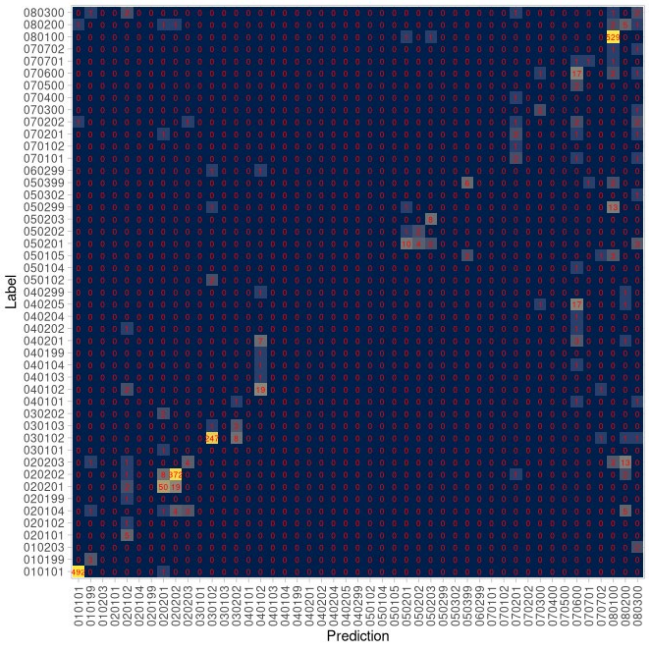


Figure 7. Confusion matrix for CEP 3 including 10% of unique data from 2023 into the training set.

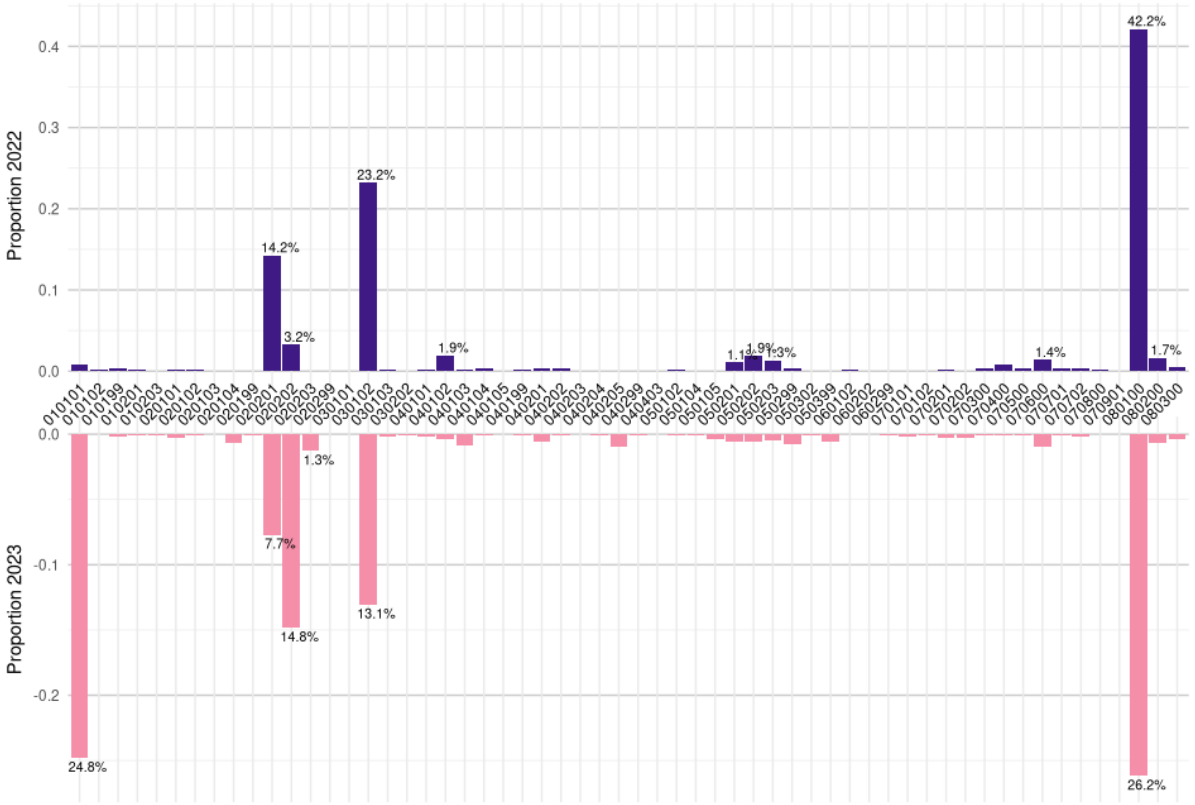


Figure 8. CEP 3 class representations in unique texts 2022 (train) vs 2023 (predict).

In conclusion, production of statistics at CEP 2 level can benefit from the the text classification method for reducing manual work without much risk to data quality. When manually labelling 10% of the new year's data to include in the training set, accuracy of ~90% is achieved. Selective editing informed by class probabilities, confusion matrix and the size of funding should allow accuracy above 95% without manually processing the majority of records. To improve performance on the rare classes (the macro F1 score), the fraction of the new year's data should be sampled wisely such that the 10% would not entail only the most common classes (e.g. removing duplicates, removing large share of projects with similar size of funding or similar length of text or based on computable text pattern indicators – rare text patterns should have a higher likelihood of also belonging to a rare CEP category). But for CEP 3 the class imbalance and annotator/coder effects do not allow acceptable accuracy in the present case. Production of statistics at the CEP 3 level will require more manual processing. This experiment also highlights the importance of attending to the annotator effect. Statistics offices may not have the privilege of having the same person labelling the machine learning data for decades. To prevent problems from multiple and/or changing coders, certain guidelines and rules of thumb regarding difficult categories and borderline cases should allow more homogenous labelling. The concurrent coders should regularly discuss the difficult cases and to understand each others thought processes when assigning a particular label to a text. Similarly the outgoing coder should attempt to train the new coder to perform as similarly to themselves as possible.

4.2 Creating Oracle database for environmental subsidies

The ESST dataset was supplemented with SFOS data, which was transferred to the Data Staging Area (DSA) via the VAIS ETL tool. Developed by Statistics Estonia in collaboration with external partners to meet SE's specific requirements, VAIS is used to perform data loading, transformation, and validation tasks, as well as to log related processes and operations. It also enables the use of reusable templates to streamline the creation of data flows.

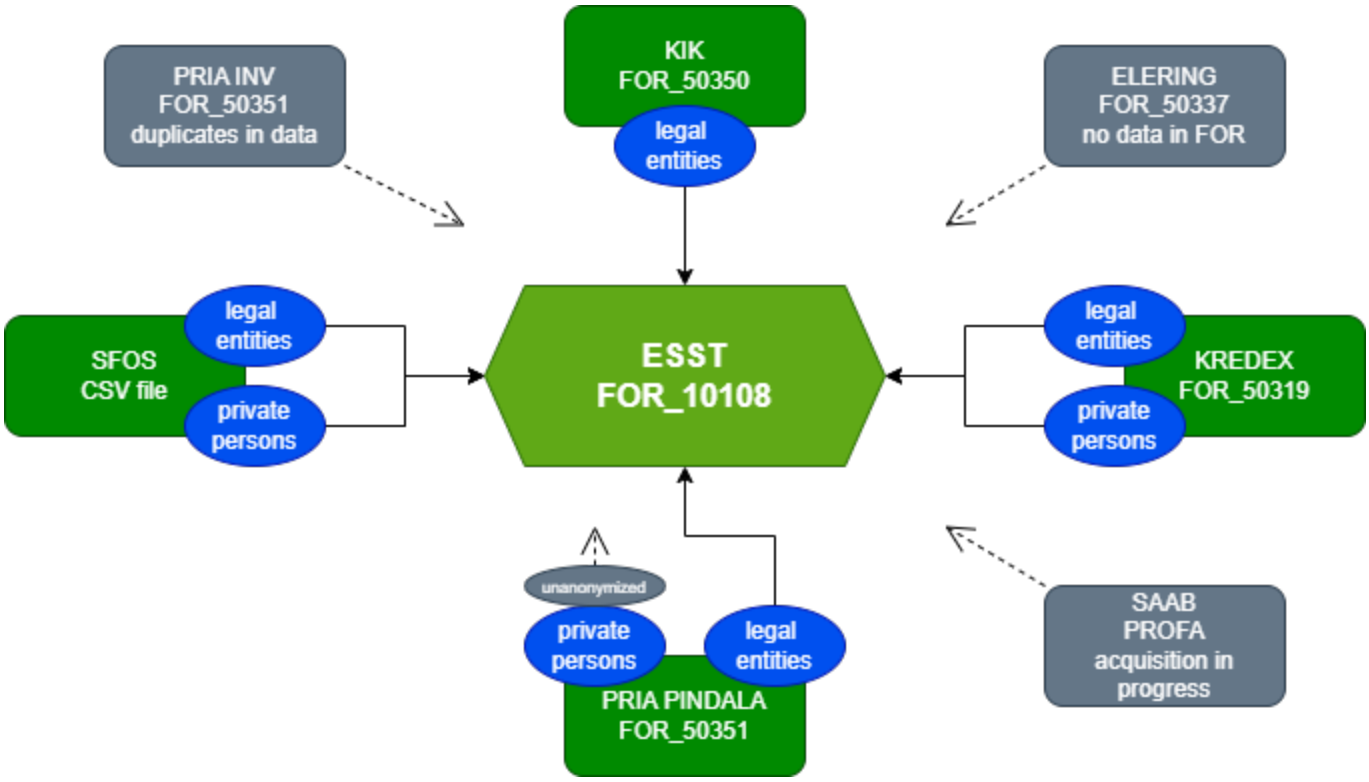
The format and composition of the dataset was checked beforehand based on a correlation table provided by a leading analyst in the environmental statistics field. It was ensured that the variables were valid and in the correct order, the observation period and the transaction period of the data was specified, some technical adjustments were made due to some of the values in the added SFOS data (e.g. maximum field length was modified).

Subsequently, as with data from previous sources, the SFOS data was also enriched with additional information, such as the field of activity, the number of employees, etc. from the economic entities register. Since some of the data required anonymization, a separate step was added to ensure anonymization in the DSA before loading the data into the Final Observation Register (FOR).

During the project, additional variables were described within out metainformation system (iMeta) and added to the subsidies dataset in the DSA, as well as FOR. The automatic reusable workflow (based on VAIS ETL tool packages) developed in the previous grant project was enhanced with steps for loading the SFOS dataset and updated to incorporate the newly added variables. Certain variables that were previously extracted from the statistical register SPI were migrated to the more accurate and current table K_SP. Once the required changes had been implemented, a new version of the ESST dataset was loaded into the FOR.

Work is still ongoing to develop procedures for incorporating various calculated variables into FOR, which are currently added manually by the lead analyst after data extraction. This will ensure the comprehensiveness of the data in the ESST FOR. Further works include integrating additional datasets (Elering, SAAB) into the ESST database, resolving certain issues related to data that, for various reasons, have not yet been incorporated into the ESST FOR. The reasons are briefly outlined in the table below (Table 8), which presents all datasets whose data are already present in the ESST FOR, are planned for future inclusion, or are currently in the process of integration. Green nodes represent datasets from which some or all relevant data are already included in the ESST FOR, while grey nodes indicate datasets that are still being integrated or have outstanding data extraction issues.

Table 8. Environmental subsidies and similar transfers FOR data lineage diagram



4.3 Compiling data using ESST database

Concentrating all ESST relevant data to a uniform database was an important step to ensure the quality and sustainability of the ESST account. As the data was processed automatically and according to the set rules, it eliminated human errors, such as typos, copy-pasting wrong data or inserting data incorrectly. Automatic data processing also speed up processing the data – data was cleaned and prepared for calculations automatically. Missing data, e.g., NACE for the final recipient, was added by linking different datasets within Statistics Estonia. Such automatic actions sped up the process of making final calculations. The data was extracted from FOR in Excel format for final calculations, such as assigning ESA transaction code and CEP but also making calculations to fill the ESST questionnaire.

Currently, the ESA transaction type and CEP was calculated manually. Although machine learning was used to determine CEP, the machine learning results had to be manually inserted to the ESST database. This was done by linking project identification numbers with xlookup function in Excel between the CEP model predictions and ESST dataset. For filling out the questionnaire, pivot tables in Excel were used to make appropriate calculations. The ESST data and the data published in Statistics Estonia database follow similar structure - making it easier to fill out both ESST and SE datasets.

In the future, the development of an automated or semi-automated solution could be considered for filling out the Eurostat questionnaire and SE dataset. That would help to reduce manual workload and help ensure greater consistency and accuracy across reporting cycles.

5 Annex 1 Minutes: Main stakeholders meeting on relevant questions of the project described in work plan

Meeting Minutes

Project Name: Establishment of an Environmental Subsidies and Transfers Accounting Account

Topic: Kick-off Meeting with Stakeholders

Date: January 30, 2025

Time: 9:30–10:30

Recorder: Raigo Rückenberg

Participants (Full Names and Affiliations)

Raigo Rückenberg – Statistics Estonia

Kaia Oras – Statistics Estonia

Grete Luukas – Statistics Estonia

Pauline Kommer – Tallinn University

Aire Rihe – Statistics Estonia

Krisela Uussaar – Affiliation not confirmed

Velda Buldas – Ministry of Finance

Mari Lahtmets – Affiliation not confirmed

Hendrik Hundt – Affiliation not confirmed

Eva Suurkaev – Ministry of Climate

Agenda and Discussion Points

1. Introduction of Objectives and Activities

The project goals and planned activities related to the creation of the environmental subsidies and transfers accounting account were presented.

2. Methodology Overview

An overview of the methodological approach was given, including the use of the CEP classifier.

3. Discussion: Horizon/LIFE and Similar Programs

Raigo Rückenberg, Mari Lahtmets, and Aire Rihe discussed the inclusion of Horizon and LIFE programs and projects, including those related to transport and energy.

All programs except for Cluster 6 are currently accounted for.

HELCOM and Interreg programs are not yet included.

Information on measures can be found on the Ministry of Regional Affairs website.

For multi-year international projects, funding is currently distributed evenly across the project duration.

Cluster 6 will be included in the new account, and the possibility of including HELCOM and Interreg projects will be explored.

4. Discussion: Data Dissemination and Analysis

Raigo Rückenberg and Aire Rihe explained that data is currently submitted to Eurostat. Starting in 2025, data will also be published for Estonian consumers via Statistics Estonia's dissemination database.

Statistics Estonia does not provide recommendations, assessments, or guidelines based on this statistical work.

An analytical/methodological report will be prepared, focusing on methodology and processes. This is a mandatory deliverable for Eurostat as the project is funded by a Eurostat grant.

5. Discussion: Classification of Public Sector Enterprises

Mari Lahtmets and Pauline Kommer discussed the annual review conducted by the Ministry of Finance regarding the classification of enterprises operating in the public sector.

For example, Elron is classified under the general government sector (S.13), while Elering is in the corporate sector (S.11).

The same applies to the non-profit sector (S.15).

6. Discussion: Tax Abatements

Raigo Rückenberg, Aire Rihe, Mari Lahtmets, Velda Buldas, and Eva Suurkaev discussed the term "tax abatements," clarifying whether it refers to specific taxes (e.g., excise duty on special diesel) or the replacement of environmental charges.

Replacement of environmental charges falls under the Ministry of Climate's jurisdiction (contact: Eva Suurkaev).

For specific taxes, the Ministry of Finance can provide support (contact: Velda Buldas).

Tax abatements are not a mandatory part of the environmental subsidies account but will be monitored.

7. Discussion: Wind Farm Compensation Fees

Raigo Rückenberg and Eva Suurkaev discussed wind farm compensation fees, where electricity producers pay municipalities (KOV), which then compensate local residents for disturbances.

The Ministry of Climate does not hold data on this, as the funds are paid directly to municipalities, which distribute them at their discretion.

The maximum compensation to residents is 50% of the fee paid to the municipality.

Currently, no one collects data on how much producers pay municipalities.

This could be a potential area for collaboration with the Ministry of Climate (contact: Eva Suurkaev).

6 Annex 2 Minutes: Seminar 1 on methodological issues.

Meeting Minutes

Project Name: Development of Environmental Accounts (2024-EE-EGD)

Topic: Consultation with Statistics Netherlands – Cooperation under Grant

Date: February 6, 2025

Time: 11:00–12:30 (EET)

Recorder: Raigo Rückenberg

Participants (Full Names and Affiliations)

Kaia Oras – Statistics Estonia

Grete Luukas – Statistics Estonia

Raigo Rückenberg – Statistics Estonia

Sjoerd Schenau – Statistics Netherlands

Marieke van der Veen – Statistics Netherlands

Agenda and Discussion Points

1. CEP Application for ESST

Discussion: Raigo Rückenberg, Sjoerd Schenau, Marieke van der Veen

CEP will be applied at level 2 for the Environmental Subsidies and Similar Transfers (ESST), in line with other environmental accounts.

Statistics Estonia will classify transfers at this level to support machine learning and fulfill grant project requirements.

Historical data will be classified using CEP for machine learning purposes, but previous years' ESST will not be resubmitted.

Any issues arising from CEP implementation will be discussed in detail during a study visit, once both teams have made progress.

2. Agriculture Subsidies

Discussion: Raigo Rückenberg, Sjoerd Schenau

Eurostat raised questions to Statistics Estonia following the 2022 data submission, regarding the methodology for classifying agriculture subsidies and potential challenges with CEP application.

It was previously observed that agriculture subsidies are allocated differently in CEPA/CREMA categories between Statistics Estonia and Statistics Netherlands.

Both institutions will continue refining the methodology to ensure consistent CEP allocation across member states, especially given the large share of funding from the EU's Common Agricultural Policy (CAP).

This topic will be further studied and discussed with Eurostat during the study visit.

3. Transfers from Local Governments

Discussion: Raigo Rückenberg, Sjoerd Schenau

Administrative data does not fully cover transfers from local governments.

There is overlap between administrative and COFOG data, making integration challenging.

Currently, only D.74 transfers are included, as they are not covered by administrative data.

The methodology for identifying and selecting transfers from COFOG data will be reviewed during the study visit.

A National Accounts expert from Statistics Netherlands should be involved in these discussions.

4. Government-Owned Enterprises

Discussion: Raigo Rückenberg, Sjoerd Schenau

In Estonia, government-owned enterprises are classified across sectors: S.13 (general government), S.11 (corporate sector), and S.15 (non-profit sector).

The classification methodology used by National Accounts will be studied and discussed further during the study visit, including comparisons with Dutch practices.

5. Wind Turbine Subsidies

Discussion: Raigo Rückenberg, Sjoerd Schenau

Estonia has introduced a new tax/subsidy for enterprises producing electricity via wind turbines.

The subsidy is paid to local governments and nearby households (S.14) to compensate for disturbances caused by turbines.

This may be comparable to NATURA 2000 subsidies, which compensate for negative side effects such as land use restrictions.

Further discussions with Estonian National Accounts experts will take place, and the topic will be explored in more depth during the study visit.

7 Annex 3 Minutes: Seminar 2 on methodological issues.

Minutes of the Methodological Seminar on the Development of Environmental Subsidies and Transfers Account

Date: 11 December 2025

Participants:

Statistics Estonia:

Kaia Oras, Raigo Rückenberg, Grete Luukas, Taavi Dubinin, Pauline Kommer, Helena Evert, Heleri Gaponenko, Reana Parve, Alice Kase, Grete Allas

Statistics Netherlands; Sjoerd Schenau, Marieke Rensman

Environmental Investment Centre: Tanel Oppi

Ministry of Finance: Mari Lahtmets

Opening Remarks

Kaia Oras opened the seminar by emphasizing the importance of methodological clarity and technical efficiency in compiling the Environmental Subsidies and Similar Transfers (ESST) account. She explained that the seminar aimed to review progress, address unresolved issues, and discuss future developments such as automation and integration with other environmental accounts. Kaia introduced the participants and went through the agenda describing the key inputs.

Methodological Overview and Challenges

Raigo Rückenberg provided a comprehensive update on the project. He explained that this is the third year of ESST development under Eurostat's grant framework. The initial years focused on conceptual and methodological foundations, while the current phase emphasizes technical implementation and automation. He discussed the persistent challenge of aligning ESST with national accounts while avoiding double counting. Strategies include selective use of GOFOG data and manual review at transaction level.

Two still problemtic areas were discussed in detail:

Agricultural subsidies were identified as the largest source of discrepancy, and the current approach assigns subsidies based on their primary purpose rather than splitting across multiple SEEA categories. Raigo emphasized that agricultural subsidies are the single largest source of discrepancy between the Environmental Subsidies and Similar Transfers (ESST) account and the national accounts. This issue has persisted across previous years and was confirmed again during the last data submission to Eurostat.

He explained that Estonia uses administrative data from the Agricultural Register and Information Bureau, which is far more detailed and accurate than the aggregated data typically used in national accounts. This detailed data allows Estonia to represent transfers more precisely in ESST. However, the complexity arises because many agricultural support schemes are multi-purpose. Policymakers often design these schemes to achieve several objectives simultaneously—such as improving biodiversity, protecting soil, supporting climate adaptation, and managing resources.

Raigo noted that this multi-purpose nature creates a methodological challenge: how to allocate these subsidies across SEEA categories. While theoretically, one could split the subsidies among multiple categories based on their different objectives, Raigo argued that this approach would introduce inconsistencies across Member States unless Eurostat provides clear, harmonized guidelines.

For Estonia, Raigo suggested to adopt a pragmatic approach:

Each subsidy is assigned to a single CEP category based on its main purpose.

No attempt is made to divide subsidies across multiple categories, as this would require subjective judgments and could lead to fragmentation and comparability issues.

He acknowledged that this solution is not perfect but is currently the most feasible and consistent approach. Raigo also mentioned that if Eurostat issues more precise instructions in the future, Estonia would consider revising its methodology to align with those guidelines.

Finally, he stressed that agricultural subsidies will remain the biggest methodological challenge for ESST until a standardized EU-wide approach is agreed upon.

Other major topics included wind turbines subisidies

Raigo explained that Estonia has recently introduced a wind turbine fee for local governments and households located within the influence zone of wind turbines. He clarified that both Statistics Estonia and Eurostat regard this fee as an environmental tax, not as a subsidy or transfer.

Because of its nature, the fee is considered a compensation for the negative impact caused by wind turbines rather than a financial incentive to promote environmental protection. Therefore, Raigo decided to exclude wind turbine fees from the ESST account. Instead, these fees will be recorded under the environmental tax account, which is also compiled by Statistics Estonia.

He emphasized that this classification aligns with the principle that ESST should only include subsidies and transfers aimed at supporting environmental objectives, not compensatory payments for adverse effects.

IT Solutions and Automation

Helena Evert explained the transition to automated data flows using the Data Gate system, which replaces email-based submissions. Helena began by explaining her role in the Development Department, where she focuses on organizing and negotiating data flows from external providers to Statistics Estonia. Her presentation centered on the transition from manual data submission to a fully automated process, which she described as a critical step toward improving efficiency and reliability in compiling the ESST account.

She introduced Data Gate, a secure web environment developed by Statistics Estonia to replace the traditional email-based submission system. Helena emphasized that this change is not only about speed but also about ensuring data integrity and security. With Data Gate, once a data provider uploads a dataset, it reaches Statistics Estonia's Oracle database within approximately three minutes. This rapid transfer is accompanied by real-time validation, which checks whether the submitted data meets the agreed structural and format requirements. If errors occur, such as submitting an Excel file instead of the required CSV format, the system immediately provides feedback to the data provider. This instant response helps prevent delays and ensures that corrections are made quickly.

Helena also highlighted additional features of the system, such as automated reminders sent to data providers before submission deadlines and a complete metadata history that records every submission, including timestamps and validation results. This transparency is essential for maintaining trust and accountability. She stressed that security is a major advantage of Data Gate, as it guarantees safe data transfer, which is increasingly important in today's digital environment.

Before automation can function smoothly, Helena explained that agreements with data providers must be carefully negotiated. These agreements cover file formats, extraction conditions, data field formats, value descriptions, and submission protocols. By standardizing these elements, both parties share a common understanding of the dataset structure and metadata, which prevents inconsistencies and errors. Once the structure is agreed upon, any changes must be mutually approved to maintain system integrity.

Helena concluded by explaining the impact of this automation on ESST compilation. Manual tasks such as copy-pasting and using Excel functions like VLOOKUP will be eliminated. Data will arrive in a uniform structure, ready for processing and enrichment, including the automatic addition of NACE codes and institutional sectors. This will significantly reduce compilation time and improve data quality. She emphasized that automation is not only about efficiency but also about creating a system that ensures trustworthy metadata, security, and reliability—key factors for producing high-quality official statistics.

Grete Allas, who is a member of the data processing design and development team at Statistics Estonia and explained her role in improving the technical infrastructure for compiling the Environmental Subsidies and Similar Transfers (ESST) account. Her presentation focused on the advancements made during the current grant period, particularly in automating data processing and creating a uniform database.

Grete Allas elaborated on the Wise ETL tool, which handles extraction, validation, and loading into the ESST database, enabling reusable templates and workflows. A uniform ESST database now exists, significantly reducing manual compilation time.

She began by describing the core tool used in their workflow: Wise ETL, a data extraction, transformation, and loading system developed by Statistics Estonia in collaboration with external partners. This tool is central to managing large and complex datasets. It handles the extraction of data from various sources, validates the data in a Data Staging Area (DSA), and then loads it into the Final Observation Register (FOR), which serves as the main repository for analysis. Grete emphasized that Wise ETL not only performs these tasks but also logs every step, making it easier to trace and resolve issues when they arise.

One of the key advantages of Wise ETL, according to Grete, is its ability to use reusable templates and workflows. These templates streamline repetitive tasks such as loading tables from SQL queries, executing validation checks, and calculating variables. She noted that there are about thirty such templates currently in use, which significantly reduce manual effort and improve consistency. Additionally, the system allows for custom workflows tailored to specific projects. For ESST, they reused the workflow developed in the previous grant period and enhanced it with new steps required for this year's improvements.

Grete then moved on to the ESST-specific tasks completed during the project. She explained that before integrating new datasets, her team carefully reviewed their format and composition. For example, a new dataset added during this period was thoroughly checked using a correlation table provided by Raigo. They ensured that all variables were valid, correctly ordered, and that observation and transaction periods were properly specified. Some technical adjustments were necessary, such as modifying the maximum field length for certain variables.

Once the data was validated in the DSA, it was enriched with additional information from the economic entity register, including fields like the enterprise's activity type and number of employees. For datasets containing sensitive information, Grete's team implemented anonymization steps before loading the data into the FOR. She also highlighted updates to the metadata system (IMETA), which now includes descriptions of new variables added to the subsidies dataset.

Grete explained that the automatic workflow in Wise ETL was enhanced to incorporate these new datasets and variables. Some variables previously extracted from the registers were migrated to a more accurate source table, improving data reliability. After these changes, a new version of the ESST dataset was successfully loaded into the FOR, representing a more uniform and up-to-date database.

However, Grete acknowledged that work is still ongoing. The next steps include developing procedures to incorporate calculated variables directly into the FOR, which are currently added manually. This improvement will make the database more complete and reduce manual intervention. Future plans also involve integrating additional datasets, such as Loring and Saab, and resolving outstanding issues with certain data sources that have not yet been fully incorporated.

Grete also presented a diagram showing the current status of data integration. Green nodes represented datasets already included in the ESST database, while grey nodes indicated those still in progress. She summarized by saying that significant progress has been made toward creating a uniform and automated ESST database, but further refinements are needed to achieve full integration and automation.

Metadata and Quality Reporting

Reana Parve, a methodologist in the Data Governance Team at Statistics Estonia, explained that her responsibility is to coordinate the compilation of quality and metadata reports for all statistical processes whose outputs are published in the official statistical database. She emphasized that these reports are essential for ensuring transparency, clarity, and understandability, which in turn makes the data reusable and trustworthy for users.

Reana explained that the purpose of these reports is to give data users a complete understanding of how the statistics are compiled, which sources are used, and what methodology is applied. Until 2025, Statistics Estonia has based its reports on Eurostat's ESMS (Euro SDMX Metadata Structure), which is widely used across the European Statistical System to ensure comparability between Member States. However, she noted that the organization is currently in a transition phase toward implementing SIMS (Single Integrated Metadata Structure), which combines elements of ESMS and ESQRS (Eurostat Quality Reporting Structure) into a single, harmonized framework.

She described the three main metadata structures used in the European Statistical System: ESMS, which is user-oriented and focuses on methodology, ESQRS, which is producer-oriented and includes detailed quality indicators, SIMS, which integrates both approaches and provides a comprehensive set of metadata elements—approximately 90 compared to ESMS's 65.

Reana highlighted that SIMS was recommended by the European Commission as early as 2017 to streamline reporting and reduce the burden on national statistical authorities. The goal is "once-for-all-purpose reporting," meaning that common concepts are reported only once and can be reused for both user and producer needs. She noted that although SIMS implementation is underway, Statistics Estonia must currently maintain both ESMS and SIMS reports in parallel because ESMS reports are still publicly available on the website, while SIMS reports will only be published starting next year.

Reana gave an overview of the quality and metadata report for the Environmental Subsidies and Similar Transfers (ESST) account. She explained that she first created a base report in Excel, inserting all available information from existing resources. Raigo then complemented the report with additional details, and they worked together to fill in missing metadata elements. Once the content was finalized, the report was translated into English and published on the Statistics Estonia website.

Reana walked the participants through the key sections of the metadata report, starting with contact information for user inquiries, followed by a statistical presentation outlining the published data, clear definitions of concepts such as "waste management" and "wastewater management," and details on observation units specifying which enterprises are included or excluded. She also covered measurement units, legal background, and accessibility provisions, explained the forthcoming links to the online database, highlighted methodological documents and their relevance for both national and international users, addressed comparability information, and concluded with a description of statistical processing, noting that ESST relies entirely on administrative data rather than questionnaires and detailing how data is obtained from registers and compiled.

She concluded by noting that the transition to SIMS will make metadata reporting even more complete and harmonized across Europe. SIMS includes more metadata elements than ESMS, which will improve transparency and usability for both producers and users. Reana also showed examples of how users can access these reports on the Statistics Estonia website.

Reflections from Statistics Netherlands

Sjoerd Schenau from Statistics Netherlands expressed a very positive view of Estonia's progress. He complimented the depth and breadth of the work done, noting that Estonia has achieved a lot in a complex and wide-ranging topic that involves methodological issues, automation, and data quality. He highlighted that the collaboration between Estonia and the Netherlands has been highly valuable, as both sides have learned from each other.

Sjoerd specifically commended Estonia for trying to stay as close as possible to national accounts principles, acknowledging that perfect consistency is not achievable but that Estonia's approach is sound. He also found Estonia's work on automation and metadata reporting impressive and said these developments were very interesting for the Netherlands, as they are planning similar improvements. He appreciated Estonia's efforts with

machine learning and recognized the challenges involved, mentioning that the Netherlands also intends to explore this area in the future.

Overall, Sjoerd characterized Estonia's progress as significant and forward-looking, and he emphasized that the cooperation between the two countries has been fruitful and should continue.

Sjoerd Schenau shared insights from Statistics Netherlands, noting the submission of the ESST questionnaire to Eurostat as a major milestone. He discussed challenges such as classification under CEP, allocation to NACE and households, and improving GOFOG data quality. He emphasized that full consistency with national accounts is unrealistic but alignment on totals is essential. Agricultural subsidies allocation remains complex, and machine learning for CEP classification is promising but requires stable, high-quality source data.

Machine Learning for SEEA Classification

Machine Learning for CEP Classification

Raigo Rückenberg provided an overview of the application of machine learning in assigning CEP (Classification of Environmental Protection Activities) categories to environmental subsidies. He explained that while machine learning had proven successful for previous classifications such as CREMA, applying it to the more detailed CEP structure presents significant challenges. The complexity arises from the large number of categories, which results in fewer examples per class and consequently lower model accuracy.

Raigo emphasized that the quality of source data is critical for the effectiveness of machine learning. Errors in manual classification and inconsistencies in project descriptions—such as changes in scheme names over time—have negatively impacted model performance. For instance, when the name of a support scheme for electric vehicle purchases was altered, the model failed to recognize it as the same category, leading to misclassification.

Despite these challenges, Raigo highlighted the benefits of machine learning, noting that it accelerates the compilation process, reduces subjectivity, and minimizes human error. He stressed that quality checks on machine-generated classifications are significantly faster than manual review of thousands of individual transfers. Furthermore, the model tends to make fewer mistakes than manual classification, provided that the training data is accurate and stable.

Raigo concluded by stating that machine learning is a crucial component of future ESST development. Work is ongoing to improve model accuracy by expanding training datasets and refining algorithms. He affirmed that this approach will play a key role in streamlining production and ensuring consistency in classification.

Future Changes: New NACE and ESA

The discussion covered the mandatory implementation of the new NACE classification for environmental accounts by 2029 and the ESA 2025 update.

Raigo Rückenberg addressed upcoming structural changes that will impact the compilation of environmental accounts, including the Environmental Subsidies and Similar Transfers (ESST) account. He explained that the implementation of the new NACE classification will become mandatory for environmental accounts by 2029. While this transition is technically manageable—since systems already allow for both old and new NACE codes—it will require careful planning to ensure consistency during overlapping periods and to manage source data that may switch to the new classification earlier than required.

Group discussed the forthcoming ESA 2025 update, noting that its overall impact on environmental subsidies is expected to be limited. However, one significant change under consideration relates to the EU Emissions Trading System (ETS). Under the new ESA framework, freely allocated emission permits were classified taxes, which represents a conceptual shift from current practice. Sjoerd emphasized that if this change is adopted, these permits would likely be treated as environmentally harmful subsidies rather than environmentally beneficial ones, and this would have implications for PETS (Potentially Environmentally Damaging Subsidies).

Raigo concluded by stating that Statistics Estonia will continue to monitor developments closely and maintain alignment with national accounts principles during these transitions. Preparatory work will focus on ensuring technical readiness for dual reporting under old and new classifications and on assessing the methodological implications of ESA changes for subsidy accounting. Pauline Kommer promised to share respective minutes of the Eurostat working group of government finance statistics where the ETS1 and ETS2 were discussed.

Integration with Other Environmental Accounts

Raigo Rückenberg explained that the Environmental Subsidies and Similar Transfers (ESST) account is not an isolated dataset but an integral component of the broader system of environmental-economic accounts. He noted that ESST data already provides input for other accounts, including the Environmental Goods and Services Sector (EGSS) and the Environmental Protection Expenditure Account (EPEA). This integration ensures that subsidies are properly reflected in the measurement of environmental protection activities and related economic indicators.

Grete described why currently the integration is partial, with some subsidies already incorporated into EGSS and EPEA, while others remain to be aligned. Kaia emphasized that the next phase of development will focus on achieving full integration within the framework of environmental monetary accounts, which combine multiple datasets to provide a comprehensive view of environmental protection financing. This will involve harmonizing classifications, improving data flows, and ensuring consistency across all related accounts. Kaia concluded by stating that integration is essential for producing coherent and comparable statistics at both national and European levels, and that work in this area will continue as a priority in the coming year.

Publication Plans

Statistics Estonia will publish ESST data and a methodological report by 30 December 2025 on its website. Eurostat will release comparable EU-wide data in early 2026, enhancing cross-country comparability.

Closing

The seminar concluded with an agreement to share methodological reports and issue summaries between Estonia and the Netherlands.

Next steps include continued collaboration and improved automation. Development will focus on achieving full integration within the framework of environmental monetary accounts, which combine multiple datasets to provide a comprehensive view of environmental protection financing

Minutes compiled: Kaia Oras 12.12.2025

8 Annex 4 Study visit “Development of the environmental accounts”

June, 2-3, 2025

Statistics Netherlands

List of participants: Statistics Estonia

Mr Raigo Rückenberg Analyst, responsible for the compiling of the environmental subsidies accounts

Grete Luukas, responsible for the compiling of the environmental monetary accounts

Statistics Netherlands;

Mr Sjoerd Schenau Project leader Environmental Accounts (physical and monetary)

Ms Marieke Rensman Researcher in environmental accounts

Parallel session 2

9:30-12:00 Environmental subsidies:

B3057

Enhancements of the evaluation methodology for the environmental subsidies' accounts: development of environmental subsidies and other transfers (focusing on agricultural (EU CAP) and renewable energy subsidies, local government and EU subsidies inclusion), accounting logic and the methodology, developing further IT solutions for environmental subsidies account.

- NL has a similar approach to Estonia when it comes to identifying transfers to and from local governments. This also means similar problems – some transfers are not visible to analysis. They are classified outside of COFOG 05, but also some transfers listed as COFOG 05 are not in scope of ESST. The solution is a better understanding of COFOG data and examining local governments/municipalities budget reports for relevant transfers. In the case of Estonia, administrative data can fill the data gaps created by the lack of detail in COFOG data. It is assumed that only relatively small portions of transfers are missing from ESST – mainly D.7 transfers to and from local governments/municipalities.

- NL analysed financial budgets of 5 biggest municipalities. This kind of analysis is planned for 2 years and after that would be automatized based on results.

COFOG – finding and classifying env. related transfers, determining transfer code, etc. Please include NA specialist. determining transfer code, etc., How does NL identify and filter ESST related transfers at local governments/municipalities level.

- NL has descriptions of transactions available from COFOG statistics. NL has attributed CEP category to transactions and if these are climate mitigation (including CCM investments) and PEDS. Assigning CCM/PEDS in ESST data could be applied also in Estonia.

Renewable energy and wind turbine fees – included in environment taxes. If and how to include in ESST? Coherence between NA, ESST and env.taxes account – discussion on how to build coherence between different accounts.

- Not sure if these are part of ESST or even part of PEDS – the wind turbine fee is to compensate for the negative effects of production of electricity from wind. Such negative effects include visual, sound and noise disturbance, reduction of property value (land and housing). No data source is available to identify the recipients of wind turbine fees – windmill fees are paid to local governments local governments and local governments distribute it to the final recipients. Data about distributions are not available from local governments.

EU's common agricultural policy (CAP) subsidies – which subsidies and under which CEP to classify? Which agriculture subsidies NL includes in ESST and how are they classified according to CEP?

- There has not yet been significant progress by Statistics Netherlands on the topic of CAP subsidies. Statistics Estonia will continue to apply CEP for CAP subsidies on the same principle as it was done for CEPA/CReMA. Discussions will continue later in the year if there has been significant progress.

Horizon/LIFE projects – what and how to include (specifically cluster 6). Discussing CINEA/CORDIS tables – identifying correct transfers, R-codes with Marieke.

- Marieke will download Estonian data so that Raigo can check if he downloaded correct information. After that Raigo can check R-codes on data.

June 10,2025
