Organisational Aspects of Implementing ML Based Data Editing in Statistical Production

Claire Clarke[1], Jenny Pocknee[1], Wesley Yung[2], Jean Le Moullec[2], Stan Hatko[2], Riita Piela[3], Steffen Moritz[4], João Poças[5], Sandra Barragán[6], David Salgado[6], Elena Rosa-Pérez[6], Jens Malmros[7], Daniel Kilchmann[8], Olivier Sirello[9], Bilyana Bogdanova[9], Amilina Kipkeeva[10]

[1]Australian Bureau of Statistics, Australia, [2]Statistics Canada, Canada, [3]Statistics Finland, Finland, [4]German Federal Statistical Office, Germany, [5]Statistics Portugal, Portugal, [6]Statistics Spain (INE), Spain, [7]Statistics Sweden, Sweden, [8]Swiss Federal Statistical Office, Switzerland, [9]Bank for International Settlements (BIS), Switzerland, [10]United Nations Economic Commission for Europe (UNECE), Switzerland

steffen.moritz@destatis.de

I.   Introduction

1.   Data cleaning and editing are essential components of ensuring the quality of official statistics. However, finding and correcting errors in datasets can be a lengthy and time-consuming process. The increasing size of modern datasets makes manual interventions increasingly infeasible, and new types of data may not be well served by traditional methods. Machine learning methods have strong potential to provide solutions to these challenges.

2.   Editing and imputation were identified as some of the most obvious use cases for machine learning in the first High-Level Group for the Modernisation of Official Statistics (HLG-MOS) project on machine learning, conducted in 2019 and 2020. The report from this project concluded that editing and imputation were valid use cases for machine learning in the production of official statistics (1).

3.   However, since then, agencies have been slow to adopt machine learning methods for editing. Therefore, rather than discussing or proposing further technical approaches to editing using machine learning, the **Data Editing Task Team of the UNECE Applying Data Science and Modern Methods Group (ADSaMM)** decided that it would be valuable to examine organisational blockers preventing the adoption of these methods and suggest some guidelines for overcoming them.

4.   Utilizing a structured approach, the task team, that involved multiple countries, developed and refined a template for collecting and analysing use cases from various statistical agencies and derived key organisational themes influencing ML adaption in data editing in the statistical production process.

5.   This short paper gives an overview of the main findings and the used methodology by the task team. The full technical report with detailed use case description and additional materials can be found under (2).

II.   Methodology for Identifying Key Themes

A.   Process taken by the task team

6.   The task team followed a systematic approach to identify the key themes related to the organizational aspects of implementing ML-based data editing in statistical production. The process was structured as follows:

(a) The task team developed a template for gathering use cases. This was initially framed around the steps in the journey from experimentation to development for machine learning methods described in Chapter 5 of Machine Learning for Official Statistics (3). Some adjustments were made to this initial version after gathering a couple of examples and determining which elements the task team found most useful. The team also eventually decided to incorporate a short technical description of the methods used in each use case, as this was deemed of considerable interest. The template is included in Appendix 2 of the full report (2) for reference.

(b) We identified potential use cases and reached out to the agencies involved to fill in the template. Early members of the team provided a small number of use cases that were used as examples to assist subsequent use case development. The most fruitful source of intelligence for identifying potential use cases was the agenda of the UNECE Machine Learning for Official Statistics 2023 Workshop[1]. In most cases, agencies that provided a use case also gave a short presentation to the team about their use case and provided a staff member to join the team. In the end, there were seven use cases overall.

(c) We assessed the use cases to extract key themes and identify areas where there were blockers to the implementation of machine learning for editing. We utilised the use cases to craft brief descriptions of these key issues and provide guidelines for overcoming them. Then, we edited these guidelines along with a select set of use cases (those which agencies agreed could be included) to create a coherent document.

B.   Use case templates

7.   The use case template was designed to gather some insight on why the uptake of ML solutions by NSOs has been slower than in the private sector and how it could be accelerated. Guiding questions were:

(a) Throughout the journey from experiment to production, what has been stopping your organisation from applying data science and modern methods in data editing?
(b) Focusing on the organisational aspects along the productionisation process, how did the project manager overcome those obstacles (e.g., how to explain the processes and methods of data editing to stakeholders within and beyond the organisation and proof that it is a good value for money)? Please identify the problems and stakeholders involved (e.g., the dynamics between senior management, research team, business area, IT team, etc.) and the actions taken to resolve the issues and improve multi-level engagement throughout the different stages listed in the use case template.
(c) To help contextualise the use cases, please discuss the pros and cons of applying these modern methods in data editing (in terms of accuracy, explainability, transparency, cost effectiveness, or other metrics) that help get buy-in from stakeholders.
(d) What are the lessons learned and best practices that would be useful for other NSOs?

8.   Our preparatory efforts culminated in the organization of the use case templates into the following categories, systematically capturing and evaluating key information.

(a) Title of the use case and the name of the organisation

---

[1] See also https://unece.org/statistics/events/ML2023

(b) Project overview
(c) Organisational readiness
(d) Understand business needs (Who needs what)
(e) Assess Preliminary Feasibility
(f) Develop proof of concept
(g) Approach/method used
(h) Prepare a Comprehensive Business Case
(i) Deploy the model
(j) Results
(k) Latest status and next steps
(l) Lessons learned & recommendation
(m) Reference
(n) Contact

9.   The detailed, completed use case templates from all participating countries, along with explanations of the categories, can be found in the full report (2).

C.   Supplied use cases

10.   The following use cases were submitted by various national and international statistical agencies, showcasing diverse applications of machine learning techniques for data editing, anomaly detection, and imputation across different domains

(a) Un-supervised ML for anomaly detection in large and frequent admin data (Australian Bureau of Statistics)
(b) Unit Value (UV) Error Detection and Correction: A Machine Learning Approach (Statistics Canada)
(c) Anomalies detection and imputation on administrative data (Statistics Portugal)
(d) Imputation using missForest (Swiss Federal Statistical Office)
(e) Imputation of Occupation in the Occupational Register (Statistics Sweden)
(f) Time Series Outlier Detection using Metadata and Data Machine Learning in Statistical Production (Bank for International Settlements)
(g) Early Estimates of the Industrial Turnover Index using Statistical Learning Algorithms (Statistics Spain)

D.   Identified key themes

11.   Based on the use cases assembled, the task team identified six influential factors that contribute to the adoption of machine learning editing methods. These were:

(a) The driver of the problem being addressed by a machine learning solution
(b) The lack of labelled data or other suitable training data
(c) The relationship between business area, methodologists/data science staff and IT specialists
(d) The need for input and feedback from subject matter specialists
(e) Domain specific knowledge and the black box nature of machine learning methods
(f) IT issues and Machine Learning Operations and machine learning platform

Each of these will be addressed separately in the following chapter.

III.   Insights into Key Themes for Adopting Machine Learning in Data Editing

A.   Driver of problem

12.   An influential factor in whether or not a machine learning solution will be accepted and progressed into production is the driver of the problem. This is especially the case for editing, where business areas may have confidence in human-led quality assurance, and less confidence in or understanding of the methods underpinning machine learning. Considerable motivation is required to change approaches even where a change may lead to substantial efficiencies. **Business areas may be less open to proposals for change that are not driven by their needs or take an 'if it isn't broken don't fix it' attitude when they are satisfied with the quality of their current methods**.

13.   For these reasons, it may be necessary to look for the right kinds of opportunities to introduce machine learning methods for editing or to reframe a problem so that it presents as the right kind of opportunity. The two most commonly represented in the use cases are:

(a) Acquisition of new data
In this scenario, an organisation receives, or is anticipating the reception of, a new dataset for which traditional approaches to editing will be less than optimal. This is most likely (for example, in the Australian Bureau of Statistics and Statistics Portugal use cases (Appendix 3 of the full report (2), use cases 1 and 3) because the dataset is very large. Editing methods that employ human intervention become infeasible at the scale of some larger administrative datasets that have high frequency and/or high volumes of data. It is also possible that new datasets may contain types of data that cannot be edited via traditional methods, such as text data. The Bank for International Settlements use case (Appendix 3 of the full report (2), use case 6) falls into this scenario - although not a new situation, there were certain kinds of time series that could not be quality assured using traditional methods. These new situations present substantial opportunities to demonstrate the capabilities and benefits of machine learning methods, as these are designed by their nature to handle large amounts of data and non-standard data in ways that traditional methods are not. In other words, in these scenarios, machine learning methods present a very natural solution to the problem at hand.

(b) Improvement to current methods
Situations where there is clear evidence that current methods have some deficiency, whether that be in accuracy, speed, or coverage, also present opportunities to introduce machine learning methods. Statistics Canada, Swiss Federal Statistics Office and Statistics Sweden use cases (Appendix 3 of the full report (2) use cases 2, 4 and 5), fall into this kind of scenario. In these situations, there is less of a case for machine learning to provide a natural solution to the problem at hand. It may therefore be necessary to compare the machine learning method to the old approach and/or to another non-machine learning method in order to convince business areas that the additional complexity of introducing machine learning is worthwhile.

One further scenario, represented by the Statistics Spain use case (Appendix 3 of the full report (2) use case 7), is the opportunity to develop new products or services. In this situation, it may also be less clear that machine learning methods are a natural solution to the problem being addressed (although this depends on the nature of the problem). Again, it may be necessary to compare them to traditional methods to show that they offer superior performance

B.   Lack of training/labelled data

14.   In the realm of data editing, one central pillar that ensures the accuracy, relevance, and integrity of automated edits is the availability of abundant, high-quality training data. However, the reality often presents a quite different picture with issues stemming from insufficient and unlabelled data. These issues often snowball

into formidable challenges affecting the development, efficiency, and effectiveness of (machine learning) models applied in data editing.

(a) Lack of (high quality) labels

Typical problems here are "Absence of labels", "Low quality/biased labels" and "Late availability of labels". Since ML-based solutions for (automated) data editing usually need labels either for model building/training or evaluation, the lack of high-quality labels can have a huge impact. The problem affects ML solutions for detecting errors as well as ML solutions for correcting errors. This is because, bias in the data/labels will often be propagated to the model. In the following, each problem is described in detail with potential mitigation measures:

- Absence of labels
  In supervised learning, the absence of labels hinders the development and evaluation of (machine learning) models. When there are no (or not enough) labels for the target variable, a sufficient model cannot be built. A lack of labels makes it essentially impossible to learn the connection and patterns between predictors and the target variable. Relying more on unsupervised learning often does not necessarily solve the problem. In unsupervised learning, the absence of a ground truth (labels) can turn the identification of automated editing candidates into a complex puzzle. Distinguishing errors from non-errors can be an uphill task, considering not all extreme values amount to errors, and vice versa, not all errors manifest as extreme values. Examples where there are not enough labels are usually missing data problems. Some labels are present (the complete data), which are used for building the imputation model. But the missing data itself is quite often not recoverable (forever unknown), which complicates the evaluation of the imputation results.

  Possible Mitigations: Make an honest approach to derive quality labels, e.g., with the expertise of a human reviewer. Combine unsupervised methods with specialised knowledge, use human-in-the-loop approaches for critical / influential edits, or use simulation studies (overimputation).

- Low quality /biased labels
  In addition to missing labels being a problem, another very common problem is biased labels. That is, the labels themselves are there, but their manifestations can be incorrect or there may be no real consensus about their values. This can be seen when different people labelling the data come to different conclusions. Ultimately, this means models built on these labels will be biased, which affects the subsequent data editing actions.

  Possible Mitigations: Analyse reviewer consensus, find solutions for reviewer consensus, put more effort into label quality, use methods for uncertainty quantification or use human-in-the-loop approaches for some edits.

- Delayed availability of labels
  Delayed availability of labels also affects development and evaluation of models and editing solutions. A key difference to the mere complete/partial absence of labels is that the labels do become available at some point in time. But, usually too late to include them in the data editing process for the currently ongoing statistical production process. In comparison to the absence of labels, late incoming labels at least enable some kind of ex-post evaluation, which would not be possible otherwise (e.g., for missing data).

  Possible Mitigations: See also all suggestions for absence of labels, try to speed up label availability, or work with partial label deliveries.

(b) Lack (quality, amount) of training data

- Not enough data

Not having enough data can affect the performance of the (machine learning) model used for data editing. Very limited data leads to a whole list of problems. For example, overfitting might become a problem, because the model learns (or focuses too much on) the noise and outliers of the limited data instead of generalising. Furthermore, some potentially predictive feature manifestations / variations might not appear in the limited data, preventing the model from using them effectively. Also, model/parameter selection becomes difficult, since there is only a limited number of test/train evaluation combinations. Overall, this causes the underlying models to fall short of the required level of robustness and accuracy.

Possible Mitigations: Obtain more data, leverage known approaches from the ML literature such as data augmentation, bootstrapping or some form of transfer learning.

- Non-representative data
  If the data comes from a non-probability sample, certain groups in the data could be overrepresented. This may lead to bias because the dataset does not represent the entire population. The model trained on the data may not generalise well to the broader population - leading to skewed predictions. Unfavourably, the evaluation metrics may also be misleading since they would be computed on the same biased data and not on the general population. Overall, the models could be skewed, exhibiting partiality towards particular trends, patterns, or classifications as a result.

  Possible Mitigations: Obtain more data, ensure that the training data does not suffer from selection bias, use statistical methods to mitigate sampling bias or control the selection of training data.

- Delayed availability of data
  Delayed availability (of some) of the data basically comes with the same issues as "Not enough data" and "Non-representative data".

  Possible Mitigations: See also all suggestions for "Not enough data" and "Non-representative data", attempt to speed up data availability or work with partial data deliveries.

C.   Relationship among business areas, methodology, data science team(s) and IT specialists

15.   For many years, National Statistical Organisations (NSOs) have been applying statistical methods to produce high quality outputs from data typically obtained from surveys or administrative sources. With the expansion of data science tools, in particular machine learning, NSOs are exploring ways to integrate these tools into their production processes. **The challenge has been that the dynamic between subject matter experts (the "business"), methodologists and IT specialists is already well established in the organisation, and the data science group has had to integrate themselves into this dynamic**. As illustrated in the use cases in Appendix 3 of the full report (2), this has been successfully done in some situations and less so in others. This section will provide some themes pulled out from the use cases and potential best practices

16.   A common thread in the use cases was that the business areas usually came to the data science areas looking for a solution to a particular problem. While this is encouraging, it can lead to a relationship where the data science area is seen almost as a 'consultant' who has been hired for a particular task. Methodology groups have been particularly successful as they are known as an area which can solve many different problems related to statistical methods. If a data science area can gain a broader reputation, it will help with having them consulted on more varied problems. In addition, if a data science area can become familiar with the business area and other problems that they are facing, then they may be able to offer solutions to those problems.

17.   All use cases recognise that close cooperation between the data science group and the business area is essential. Several use cases (Australia, Portugal, and Spain) highlighted that not only does the business area

need to understand what the data science area is putting into place, but the data science area needs to understand requirements of the business area.

18.   In the use cases, the relationship between methodology and the data science areas is not always clear. Most use cases mention the importance of collaboration between methodology and the data science area (if one exists) but did not elaborate on it. At Statistics Canada, the data science area is housed in the same organisational unit as the methodology group to foster collaboration. The methodology group is well integrated into the statistical programmes and steps are underway to leverage this to further integrate the data science group into these programmes. In addition, this arrangement is helpful in sharing knowledge on both sides and, more importantly, identifying potential barriers to fully integrating data science tools into statistical programmes.

19.   This arrangement has also brought up some interesting discussions on the future relationship between methodology and data science. In recent years, new methodology recruits often join with some competencies in data science. If this trend continues, how will the roles and responsibilities evolve going forward? One possible scenario is that "citizen data scientists" will be more common in both methodology and subject matter areas and that a small data science division consisting of more research-oriented data scientists will be established. This scenario would be similar to what probably happened many years ago as statistical sampling techniques or complex statistical analyses were adopted. However, both of those examples took multiple years to occur.

20.   Similar to the relationship between methodology and data science areas, the one between IT and data science has also been a challenge. The major challenge has been the concerns around IT security and the ability to provide the necessary IT infrastructure for the new data science applications such as computing power, data storage. There will obviously be a "feeling out" stage where IT and data science will have to learn about each other and to define roles and responsibilities, but the earlier that this is achieved the better for the organisation. The advent of ML Operations (MLOps) has brought a new dimension to the traditional Development and Operations (DevOps) framework. Often misunderstood as competing approaches, DevOps and MLOps are, in fact, deeply interconnected, each playing a pivotal role in the lifecycle of software and ML development. DevOps and MLOps share foundational principles of automation, iterative processes, and a collaborative ethos. The Continuous Integration/Continus Deployment (CI/CD) pipelines, central to MLOps, are predominantly an extension of DevOps practices, underscoring the interplay between the two. When aligning MLOps with DevOps, it is imperative to delve into the intricacies of MLOps. This understanding is pivotal in recognising how MLOps does not just coexist with DevOps but actively intertwines with it, enhancing and extending its capabilities. MLOps aims for the automation of processes and champions transparency and reproducibility, aligning closely with the core objectives of DevOps.

D.   Input and feedback from subject matter experts

21.   Relations among different profiles in the statistical offices are not always easy to manage and getting to a full understanding that leads to fruitful results can be difficult. However, these multidisciplinary teams are the key to success.

22.   Subject matter experts have accumulated great knowledge in the particular statistical areas for which they are responsible for. Their competencies and skills have been developed through years of training and experience learned while working, even accumulated from former colleagues in the same subject matter. In relation to subject matter experts, the challenge is the lack of time due to the production process. They are focused on the needs of production that sometimes require urgent interventions, so it is difficult for these experts to be engaged in innovation projects. At the same time, they have a great amount of knowledge about the real needs of production, and they know the behaviour of the data better than anyone so they can be helpful with interpretability of intermediate results which gives feedback to improve the methodology. Subject matter experts are vital also in the first steps of the machine learning methods with their description of the manual procedures to be transformed into regressors containing essential information for the model. Another challenge arises from the steep learning curve of new methods, making the project appear too difficult to confront for the subject matter experts who are not familiar with the new methods.

23.   From the point of view of the methodology units, it is important to understand the problems and needs of the business areas but even more important to be able to develop standard solutions that solve not only the problem at hand but similar (of the same nature) problems that could appear in other business areas. Then, a possible fruitful collaboration is not one-to-one but the hub and spoke model to build teams where the methodology unit is in the root and the business areas are in the nodes. Then, the methodologists can understand not only the initial problem but others that are of similar nature.

24.   Potential mitigation measures include:

(a) Incorporate subject matter experts right from the project's inception. Ensure that they are not just participants but actively recognized as integral contributors to the project.
(b) Engage the subject matter experts at the early stage is also important to learn their real needs and incorporate them in the design of solutions.
(c) Explain the methodology to the subject matter experts and give them enough training in order to feel comfortable with the new process that they will have to run.
(d) Transmit to the subject matter experts that these new projects are an opportunity for them to improve and to save time and encourage them to see the time spent for the project as an investment for the future. Starting from recent new methods incorporated to the pipeline process as cases of success, set how the result of this new project will be in the production process and which are the advantages of that.
(e) Work with groups of people within the structure of the hub and spoke model. (See (4) "The Use of Data Science in a National Statistical Office", Erman et al (2022))

E.   Requirements for data science expertise and black box issues

25.   One key aspect related to the implementation of ML pipelines into a production environment relates to the organisation readiness, including the human resources (e.g., availability of expert, trained staff in data science and machine learning). This is critical not only to grasp all the benefits from using ML for statistical production, but also to prevent black-box challenges, that is, the use of obscure ML algorithms

26.   To start with, ML methods often require strong expertise in data science. Several use cases in Appendix 3 of the full report (2) indicate the specific need for knowledge, training, and recruitment in order to keep up with ML advances in a rapidly changing environment. For instance, the Australian Bureau of Statistics mentions the effort to provide staff with knowledge about data science activities and the use of machine learning methods. Statistics Portugal also reports a very similar requirement, with its management encouraging training courses in data science, both to empower employees with new knowledge and to deploy machine learning methods into their daily work activities. Conversely, only few organisations (Statistics Sweden) admit having sufficient knowledge in-house to build and maintain ML-based applications for official statistics.

27.   The transparency of the ML methods chosen is also key to prevent black box issues. This is crucial in order to mitigate operational and reputational risk for the organisation in case ML pipelines generate unexpected results which cannot be explained. To this respect, some organisations promote synergies between data scientists, IT and business areas in order to conduct an in-depth evaluation and test phases of the methods chosen as well as to jointly define validation, consistency and coherence analysis steps (e.g., Statistics Portugal and Statistics Spain). Code sharing is another approach to mitigate the black-box risk followed by several organisations such as the Bank for International Settlements. It aims to eventually foster discussions among experts on the best methods to follow and avoid the use of highly uncertain, complex algorithms. Organisations may also consider the possibility to disclose the full decision-matrix behind the usage of machine-learning techniques, including the rationale behind the selection of specific parameters. Finally, black box machine learning models and their potential model failures can be mitigated by setting up rigorous uncertainty sets (e.g., conformal prediction) for the predictions of the models used in production.

F.   IT issues (ML infrastructure)

28.   As mentioned, IT infrastructure, systems and processes are fundamental to harness data science, machine learning and compute capabilities. While the adoption of emerging data science technologies offers potential opportunities, such as meeting the computation demands of big data, there are also challenges. These challenges are relevant for innovations generally but appear particularly so for machine learning data editing projects. These challenges can also depend on where an NSO is on their IT / data science modernisation and machine learning journeys. This section outlines some of the key issues and potential solutions. For more background on challenges to machine learning projects, please refer to the "Building an ML ecosystem in Statistical Organisations" report from the Office of National Statistics Office (ONS)-UNECE ML Group 2022 (5).

29.   Innovation projects in general require IT systems that <u>support the research and development process</u>. Innovation is more likely to be successful if an organisation has streamlined R&D environments / processes that support the innovation cycle, for example, environments that enable data to be brought together with emerging tools and software in a safe way. Research environments may have less functionality than production systems; so later stages of the innovation cycle might require additional assessment be undertaken, for example, model hyperparameters, and compute performance using full-scale data. It is important to allow for these steps.

30.   Another aspect of the innovation cycle is the importance of <u>streamlined governance processes</u>, such as resourcing different stages of the cycle and go/no go decision making. It is particularly important that "production owners" for the methods, IT, data science and statistical subject matter have been identified and agreed upon early in the process. For example, Statistics Canada uses formal Service Level Agreements for production roles and responsibilities across production staff, IT staff, data scientists and statisticians.

31.   The innovation may also require integration with, and modifications to, <u>the production environment</u>. Productionisation takes effort and resourcing to test, deploy and integrate the model / components into the proof of concept and production systems; including refactoring code to reduce tech debt, automation (e.g., iterative model updating), memory usage, and I/O optimisation. This integration with production may also include components such as pre-processing, Quality Assurance/Machine Learning tools (Statistics Canada), related editing/imputation processes and tools (such as manual editing) and incorporating any necessary system changes to standard outputs such as prediction errors (Statistics Spain). Some components or underlying processes may not yet exist in a production system for an organisation, such as R/Python servers or cloud compute capabilities. It is important to start arranging production IT infrastructure early because of the time and resourcing demands on IT teams.

32.   Many NSOs are undertaking <u>IT / data science modernisation programmes</u>, which provides opportunities for innovation and enables the organisation to meet future needs. However, it also places high demands on IT teams as modernisation programmes can be a long multifaceted journey that stretches IT teams' support over new and existing systems through the transition. Innovations beyond these programmes may compete with these resources, and so need to be seen as complementary. The emerging tools and supporting infrastructure need IT staff to build components and provide ongoing support.

33.   <u>Cloud-based environments</u> provide the potential to manage big data and harness emerging technologies and open-source software. While this can be the catalyst and opportunity for editing and imputation projects, there are some challenges. For example, different cloud providers offer different services / functionality; what an organisation wants (e.g., MLOps_ may not be easily available. Standard production system components and tools may not be easily incorporated, for example, not all programming languages are natively supported. It takes time and resources to adapt cloud environments to meet the needs of an NSO, to build and incorporate these components and aspects (e.g., security). This means that environments under development may not (yet) have all the services and functionality needed for ML data editing.

34.   Acquiring and developing skillsets is essential. Collaborating with cloud providers can be beneficial, although the potential issue of vendor lock-in should be considered. These environments provide access to open-source programming languages with a wide range of packages that may be available, which is useful for ML projects. While in-built machine learning cloud services may be available, organisations need to consider the needs of an NSO for transparency, explainability and control (for more refer to the HLG-MOS Project "Cloud for Official Statistics" (2023)).

35.   <u>Support for programming languages</u>: Each programming language used by an organisation requires a support team, so NSOs may select a set of languages to support. Every programming language has strengths and limitations. For example, SAS is a trusted and well-supported programming language widely used for official statistics. Open-source programming languages such as R and Python offer a wide range of prebuilt packages that are useful for statistics (including those developed by NSOs) and are particularly useful and flexible in the ML space. Being open-source, there is no guarantee over robustness and support for packages. Nevertheless, many R and Python packages do have committed support teams. They also require more effort on the part of the organisation, for example, version management.

36.   Not all functionalities can be met by pre-built software / packages, so some components may be developed or modified in-house. These custom solutions take additional effort to build and maintain. This may especially be the case to adapt emerging approaches such as machine learning, to meet the needs of an NSO such as applying for a statistical product and providing greater control or explainability.

37.   <u>Vendor lock-in</u>: Historical decisions about the IT environment may make it harder to incorporate emerging technologies particularly when adapting or transitioning away from legacy systems. For data editing projects, this could apply, for example, to the introduction of open-source programming languages (and the supporting infrastructure and processes), ML infrastructure (refer to the Appendix 1 of the full report (2)), and cloud environments/tools. For example, the "Cloud for Official Statistics" project noted the importance of having an exit strategy from the start when procuring IT solutions, so that costs are understood (such as egressing data), and that time and resources are already allocated to transitioning at a later stage. For cloud solutions, what is possible in terms of an exit strategy depends on the type of cloud approach that the organisation has adopted. The project team also noted that vendor-agnostic and open-source culture also make it easier to acquire skilled staff. Cloud-related skills are in high demand and take time to build, so it can be useful to work closely with vendors to develop systems. However, one needs to be mindful of the potential for lock-in if vendor-specific systems are embedded and skills are developed in the organisation.

IV.   Conclusion

38.   In previous sections, we have discussed a range of non-technical issues that can present barriers to introducing machine learning methods for editing. Some of these issues may also arise when trying to use ML methods for other parts of the statistical production process. For all these issues, it is clear that careful planning of any machine learning project is necessary for success. This might include factoring in costs to acquire and label training data, planning to bring in different teams with different expertise at the right points in the project, scheduling project milestones to facilitate the involvement of busy subject matter experts, or including time and resources to educate future users. Building good relationships and having clear lines of responsibility between methodology, data science, business and IT teams is vital, as is making sure that appropriate IT environments and resources are available to support the demands of an ML project. Different approaches may be needed to carry out a proof of concept compared to introducing methods into production, but the latter should not be ignored when planning the former in order to facilitate a smoother introduction into production later.

39.   The implementation of ML Operations (MLOps) in statistical offices represents a pivotal transition from traditional ways of implementing methods to more advanced and automated processes. MLOps also provides a structured framework for embedding Responsible AI principles. These principles ensure ethical, transparent, and accountable use of AI and machine learning, covering aspects such as fairness to prevent

biases in models, accountability in development and deployment, transparency in decision-making processes, adherence to ethical standards in data usage and protection of sensitive information. MLOps in statistical offices is not just a technological upgrade but a comprehensive strategy towards more responsible, efficient, and advanced data processing and analysis. Along with responsible AI and other MLOps principles, it covers IT infrastructure, tools, processes, and roles. This transition is essential for statistical offices to remain relevant and effective in a datadriven era, guaranteeing the provision of accurate, reliable, and insightful statistics.

IV. Literature

(1) United Nations Economic Commission for Europe. (2021). *HLG-MOS Machine Learning Project: Project report*. Retrieved from https://statswiki.unece.org/display/ML/Machine+Learning+Project+Report

(2) Data Editing Task Team of the UNECE Applying Data Science and Modern Methods Group (ADSaMM). *Organisational Aspects of Implementing ML-based Data Editing in Statistical Production*. United Nations Economic Commission for Europe, 2024. Retrieved from https://w3.unece.org/stories/2024/09/mlops_wp/

(3) United Nations Economic Commission for Europe. (2021). *Machine learning for official statistics*. Retrieved from https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf

(4) Erman, S., Rancourt, E., Beaucage, Y., & Loranger, A. (2022*). The Use of Data Science in a National Statistical Office*. Harvard Data Science Review, 4(4). https://doi.org/10.1162/99608f92.13e1d60e

(5) ONS-UNECE Machine Learning Group. (2022). *Building an ML ecosystem in statistical organisations: Infrastructure theme group report*. United Nations Economic Commission for Europe. Retrieved from https://statswiki.unece.org/download/attachments/338329823/ML%20Ecosystem%20report%20%282%29.pdf?version=1&modificationDate=1671532803773&api=v2