ONS-UNECE Machine Learning Group 2022

Building an ML Ecosystem in Statistical Organisations

Infrastructure Theme Group Report

Authors:

Jakob Engdahl (Theme Group Lead)
InKyung Choi
Eric Deeben
Joni Karanka
Anton Karlsson
Matyas Meszaros
Jenny Pocknee
Peter Holroyd
Alison Baily

Contents:

1. Introduction

As we continue to advance in the field of machine learning, it is becoming increasingly clear that the ability to use and scale these technologies requires more than just the underlying ML code. After initial experimentation and proof-of-concept, it is important to establish a structure that allows us to expand the use of machine learning beyond a small group of experts.

In order to do this effectively, it is crucial to avoid creating two separate statistical organizations under the same roof – one focused on traditional methods and the other on machine learning. This would result in duplication of efforts and a fragmented approach to statistical analysis, ultimately hindering our ability to fully leverage the benefits of machine learning. Instead, we must find a way to integrate new methods and technology into our existing processes and skillset, while still allowing these new approaches to influence and improve our existing systems, methods, and infrastructure.

One key aspect of building sustainable capacity for machine learning is the focus on MLOps, or the practice of integrating machine learning models into the production environment. This involves building the infrastructure and platforms necessary to deploy, monitor, and maintain machine learning models at scale. This infrastructure must be flexible and scalable, allowing for the rapid deployment and testing of new models, as well as the ability to integrate with existing systems and processes. Additionally, it is important to have robust ML-platforms that provide tools and services for data preparation, model training and evaluation, and deployment. By investing in these technologies and practices, organizations can ensure that they have the infrastructure and platforms necessary to effectively use and scale up machine learning.

Creating sustainable capacity for machine learning involves considering multiple perspectives and components, including technology, acceptance, infrastructure, data, people, culture, legal and ethical considerations, and the broader ecosystem in which we operate. Building capacity in these areas requires a coordinated and iterative approach, with a focus on continuous learning and improvement.

The following report summarises highlights of discussions held by the IT Infrastructure Theme Group between February and December 2022. This group convened colleagues from different statistical organisations to share experiences of developing open base platforms for ML data processing and analysis and the wider supporting infrastructure.

To help navigate the journey toward establishing the ML-enabling environment in a more systematic way, we introduce a maturity matrix for ML capbility in **Section 2**. In **Section 3**, some of the common challenges that arise during this journey are discussed. The report concludes in **Section 4** with proposals for future work.

2. ML Capability Maturity Matrix

Capability can be defined as “an ability that an organization, person or system possesses”. Capabilities are typically expressed in general and high-level terms and typically require a combination of organization, people, processes, and technology to achieve” (TOGAF 9.1). Machine Learning can be considered as a capability which benefits statistical organisations by making their processes more efficient and their statistics more relevant and timely.

Organisational capabilities do not appear in a fully mature state, but they develop over time as the organisation gains experience and resources. In the following section, we consider four dimensions, namely **institutions/people**, **process/method**, **technology** and **information**. For each dimension we define four levels of maturity starting from an initial stage where ML is emerging in the organisation to a more mature stage where the use of ML is well embedded, efficient and standardised across the organization. A summary table can be found in the Annex.

<u>Institutional setting / people</u>

1. **Initial:** A few individuals become interested in ML and experiment in silos. Most parts of the organisation, including business areas, are not aware of the potential and value of ML. There is no corporate-level strategy for using ML/AI in the organization, thus projects are often opportunistic and ad-hoc without a common approach or long-term plan.
2. **Spreading:** Groups of individuals become aware of others working on similar issues and form one or several informal knowledge-sharing networks. Some ML-related roles are identified and their responsibilities are formalised. Training for staff might be available but is often ad hoc in nature. There is an increasing awareness of the multidisciplinary nature of ML projects. Senior leaders of the organisation show an interest in exploring the potential of ML.
3. **Widespread:** ML work becomes more centralised and formalised (e.g. some organisational restructuring takes place, new divisions are created). Supporting capabilities are identified and start engaging in ML development and operation. There is regular training for staff adapted for their needs and expertise (e.g., senior management, data scientists, ML engineer) as well as aligned with organisational goals. There is a wide awareness of ML across the organization, and a corporate-wide strategy for the use of ML is in place. Ethical and legal issues are taken into consideration for ML projects.

4. **Mature:** Essential capabilities relevant to ML are fully integrated into the organization: recruitment, training, roles, planning, finances, communities. ML is considered as a strategic capability and its developments and financing are planned and managed at a corporate level with responsible senior leaders. All business areas are aware of how to seek support with the implementation of ML projects. The organisation meets clear requirements with regards to data and AI ethics as well as transparency and public responsibility.

Information (data and metadata)

1. **Initial:** Individuals who develop ML models acquire and process necessary data on their own. There are difficulties in obtaining data if it exists in other areas of the business as there is low awareness on the importance of data for ML. Documenting the model and development process is mainly done for the reference of the model developer and is often not a requirement. There is no standard for documentation.
2. **Spreading:** There is an effort to create a common pool of data sets among those in different statistical programmes who work on similar issues (e.g., text descriptions of economic activities). There is increasing awareness of the pre-processing of ML data and the core metadata requirements of the models. There is some guidance on feature engineering and labelled data.
3. **Widespread:** There are standardised means of accessing and storing data for training, testing and inference. There is some standardisation of the quality of training data and labels. A minimum set of core metadata is established, and models and their code are version controlled. There are established processes for preparing ML data for training and insight. There are standardised documentation requirements.
4. **Mature:** It is straightforward to find data and metadata to train, test and deploy ML models. There are central repositories for storing important information (features, models, metadata, code) that can be reused and reproduced. Information on previous versions of models is easily accessible, as are labels used for training and model hyperparameters, making previous runs of the models reproducible.

Process / method

1. **Initial:** There are no clear handover points and individuals are responsible for a large range of ML activities (e.g. a data scientist carries out IT work). Individuals are tasked with multiple projects in an ad-hoc manner. There are no standard development methodologies or common practices, and individuals develop ML models specific for their needs.
2. **Spreading:** There is a common understanding of the general processes involved when developing and deploying ML models, including activities and responsibilities. There is division of work and clear hand-over points that are defined based on a better-defined process. There is an established link between the ML processes and existing production processes. ML projects are scoped, prioritised and rationalised leading to some common models and ML services being identified.
3. **Widespread:** The use of common methods and processes harmonizes the development and deployment of ML models. ML models are developed in

coordination with business areas that have similar needs using established triage and development practices. Models are developed following best practice with the aim to be modular and portable, and to facilitate sharing and re-use. There are some ML-driven services used across several business areas (e.g., linkage, classification or coding tools).

4. **Mature:** Processes are fully or partly automated, reducing manual intervention as far as possible, hence freeing resources for high-value activities. Models are continuously reviewed to monitor quality and improve their efficiency. There is an established governance plan for models, including their retraining, decommissioning and replacement. There are established ML models that are easy to find and test for new users.

Technology (software and hardware)

1. **Initial**: There is no corporate approach to ML software. Individuals either have little access to software or install the programmes and software needed for ML with no corporate support. Practitioners deal with issues in a piecemeal fashion with little experience and support provided by the IT department. There is no suitable hardware or tooling available for practitioners when they need to scale their training or models.
2. **Spreading**: Essential software is white-listed in the organization but the approach is either ad hoc or lacks key tools. Individuals working on non-traditional data types might face difficulties with specific software and packages. Support is limited and there is no common approach for versioning or CI/CD pipelines. The scope of software considered for corporate support in IT is limited. In some instances, specialised hardware or equivalent cloud services (for example, labelling tools or GPUs) are provided to practitioners but on a per-individual or per-project basis.
3. **Widespread**: The organisation has software platforms (either on the business premises or on the cloud) that specifically deal with ML requirements. Environments for different types of ML tasks are available where a core set of libraries are installed by default. IT services are familiar with ML needs and take these into account in their architecture. The software and hardware needed for ML maintenance and operations are supported.
4. **Mature**: The procurement of new hardware (e.g., clusters, GPUs, TPUs) or cloud computing services for ML is part of regular corporate planning. Identification of needs for new software and hardware is done through continuous monitoring and is planned at a corporate level, including regular horizon-scanning exercises. There are agreed approaches and tools to deal with model versioning, testing, monitoring and deployment.

This matrix illiustrates the complexity of building an ML ecosystem. A wide range of operational issues must be considered when seeking to build sustainable capacity for machine learning. Developing capability to fit within an existing system is a multi-stage process that requires a systematic approach tailored to the skills set and maturity level of the particular organisation.

3. Recurring Themes

Building organisational ML capabilities takes long-term investment and requires transformative changes in areas such as culture, organisation, technology and infrastructure. This section describes some of the common challenges that arise during this journey.

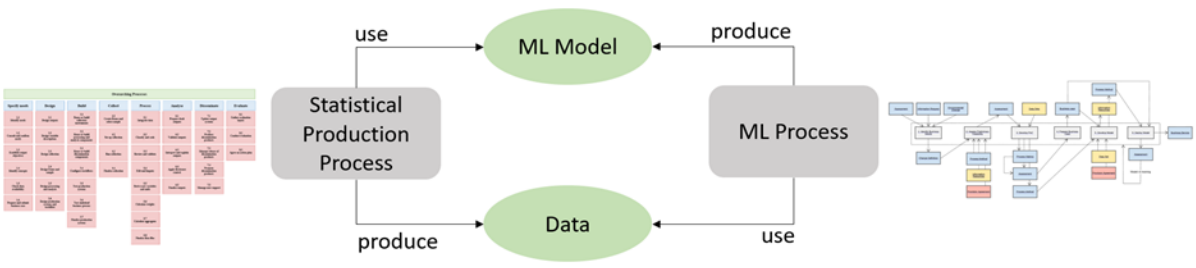
3.1 Connection between the Statistical Production Process and ML Process

ML solutions should address business needs in the organisation. No matter how accurate it is, a model that does not deliver any business value is not suitable for production. Given that statistical production is one of the core businesses of the statistical organisation, it is important to understand the connection between the regular statistical production process (which is often described by GSBPM) and the ML development process.

At a high level (see Figure 1 below), statistical production processes can use ML models and these are, from the point of view of GSBPM, functionally equivalent to traditional statistical methods. For example, if the purpose of the ML solution is auto-coding, it would be used for GSBPM 5.2 (Code and Classify); if the purpose is assisting interactive editing, it would be used for GSBPM 5.3 (Edit and Impute). The ML process, on the other hand, uses data from the statistical production process to build the model. Note that this “data” does not necessarily mean the end-product of the entire production process. ML models are often trained on micro-level data, thus it is more likely to be microdata, for example, from GSBPM 5.2 (micro-level records coded manually) of the previous production cycle, or from a statistical production process in a different programme (for more information about the connection, see this).

As ML models need to be monitored and maintained after a deployment, the touchpoint between ML process and the statistical production processes that use the ML solution does not end at the ML solution hand-over. The ML process will use data from the production process once more to re-train the model, and the production process will use the re-trained model.

Figure 1.

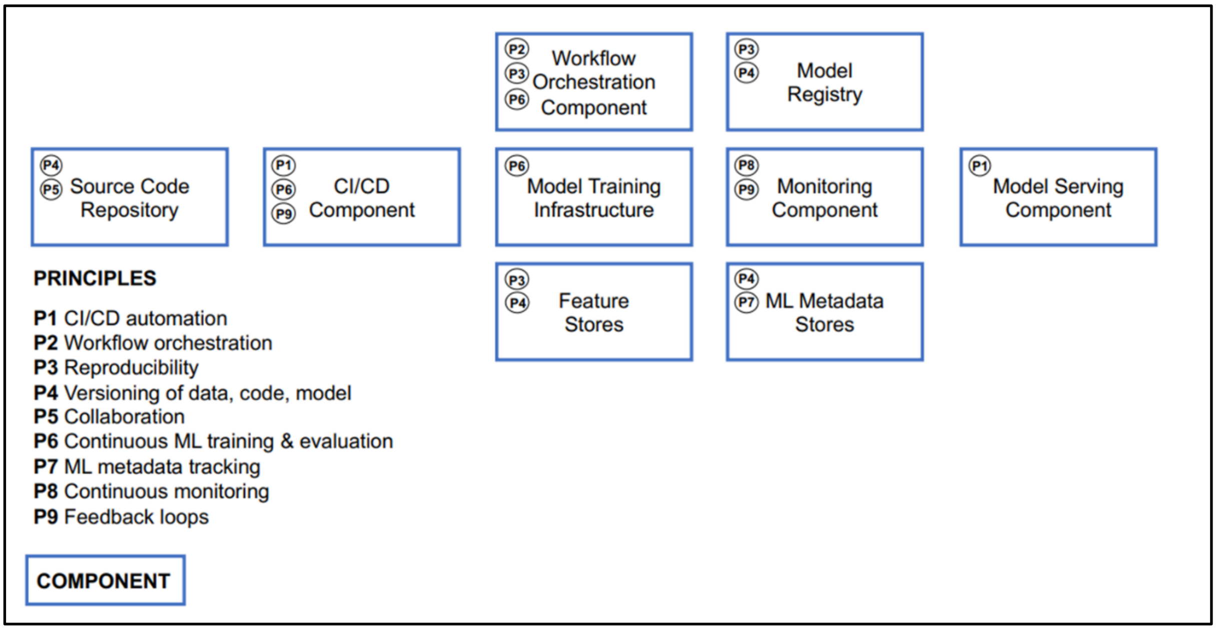


3.2 ML Platforms

ML processes need to be production ready and integrated into the traditional production processes. An important feature for implementation is to have a secure environment where confidential data can be used to test different models. These platforms should not only facilitate the production of models, but should also support data.

A fully-fledged ML platform contains the following components detailed in the figure below.

Figure 2: Technical components of production grade MLOps platform
(source: Machine Learning Operations (MLOps): Overview, Definition, and Architecture)



The components that an organisation will require depends on its level of maturity in using machine learning. In the Initial and Spreading phase of the maturity matrix (see Section 2), it is likely that only a few of these (if any) are organised in a structured way. As maturity increases (Widespread and Mature) more of these components are necessary to run machine learning in a more structured way.

The architecture to host such components can be cloud based or on-prem dependent on the legal possibilities. This creates additional complexity in the alignment to the traditional production systems. Cloud based architecture can be operated at lower cost, but needs an additional layer of security and for governance.

As an emerging technology, cloud capabilities and ML services and components are still under development – they may not yet provide the full functionality needed by NSOs. When deciding on the appropriate technological stack / solution, there can also be trade-offs between platform solutions versus individual components (like CI/CD pipelines, orchestration, model registry, serving of models, endpoint management, GPUs / TPUs for training, auto training, dedicated feature engineering tools, dedicated labelling tools, hyperparameter optimisation) that are easier to implement and support. When making this decision, organisations need to take into consideration the level of control, transparency, explainability, potential lock-in-effects, maintainability, functionality of the components and ensure this fulfils the needs of the business areas. Such ML components as an integral part of regular production are only accepted by the business users when the quality aspects and reliability of those components are proven. A further consideration is whether to build such component / functionality internally (best of suite) and/or integrate with alternative systems (best of breed) that provide the same functionality.

In order to use these systems, a different mindset is needed from the producers perspective. The users have to learn about versioning code, data, models or use open source components available on the internet. Many of the features and skillset for these systems could also be useful in regular statistical production.

Many of the components used for machine learning utilise so called Container Technology which increases quality aspects such as scalability and portability. By utilising container capabilities, it is possible to set up multiple machine learning environments for different purposes in the same organisation. Use cases for having multiple environments could be separation between exploratory innovation and regular automated machine learning production.

Many software components for ML are also Open Source. This increases the opportunities for collaboration on ML platforms within the statistical community. There are already existing ML platforms that are used for experimental research and/or in production for official statistics like the UN Global Platform, Statistics Canada Advance Analytics Workspace, INSEE Onyxia or ABS MDIATE / SADE.

Use Case: ONS Address Index Matching Service

The ONS Address Index Matching Service utilises a BERT (Bidirectional Encoder Representation from Transformers) model used during the 2021 England and Wales Census. Conditional Random Fields are used to parse the input addresses and Elasticsearch to execute queries.

This has been in use for five years and recently further research has been done over the past 12 months utilising SBERT (Semantic BERT) to prepare a trained model in order to improve matched results

Work is currently underway to combine the older BM25 (ranking of results for relevance) and more recent k-NN (k – Nearest Neighbour) by executing queries simultaneously within ElasticSearch for the purpose of improving matched results.

3.3 ML Deployment and Serving

Production-ready patterns for deploying machine learning models can be remarkably different to the way data scientists initially learn to deploy ML models. As addressed in Kreuzberger et al (2022), a complete deployment which can provide inferences to a system or end-user is composed of multiple parts. Doing this as a team collaboration and across multiple devices in a production environment can be quite complex.

Taking a simplified view, we can distinguish several components:

- A feature engineering and data processing pipeline, which prepares the data for the model to utilise (in either training, evaluation or inference).
- A training and evaluation pipeline, which is used to build a trained ML model that meets the requirements of the task.
- A deployment process for the ML model into the target infrastructure.
- A means of serving the ML model; that is, a way for users or processes to interface with the model so that the model can provide inferences on data.

All these key processes are supplemented by repositories for code, models and data, continuous integration and deployment (CI/CD) of models, workflow orchestration, and monitoring of the different processes. These supplemental artefacts support the automation, reproducibility and quality of the models, and therefore are key to reducing effort in the longer term.

Two main options are used for serving the predictions generated by ML models: batch inference and online inference. A good discussion on the relative advantages and disadvantages of both approaches can be found here: Batch Inference vs Online Inference - ML in Production.

Commonly ML practitioners are used to deploying ML models in their own machines and making use of batch inference. In this case, the data to be evaluated and the ML model are co-located and the entire data set can be sent to the model for inferences to be produced. The key advantage of this approach is its ease of deployment, as long as the data can be batched into sizes that the machine can handle, and there is no requirement to provide near-real-time inferences.

Alternatively, ML practitioners can deploy their ML models for online inference. This is often an activity that requires more thought to the architecture of the solution: the model has to be designed to handle one observation at a time and to provide an inference to it. The typical deployment of a model for online inference is for it to have an API, usually as a web service with a REST API, that takes one message and provides a specific output for it. In the case of a typical ML deployment the message could be a string, number, or reference to another type of object (for example, the address of an image or sound file). This typically requires an understanding of the client / server model, in which the ML model is hosted and exposed with an endpoint with a specific url (see UNECE CSPA for comparison).

The **key advantages** of online inference are that: it can provide real time inference; if correctly architected it can also scale horizontally (that is, multiple concurrent requests can be handled); and if the API is clear, it can provide language independence for the end user.

The **main drawbacks** are the greater complexity of deployment and understanding of API design. The reproducibility aspects of statistical production could also become more complex in online inference solutions.

It is worth noting that online inference is becoming a more common pattern for the ML offering of cloud providers. This means that many of the hurdles with regards to building, deploying and exposing models via endpoints have specific solutions provided by each vendor, reducing overall effort for the ML practitioner.

3.4 Model Maintenance

ML models are built based on a training data set. However, as the underlying phenomenon captured in the training data set changes over time (e.g., new types of jobs and products enter the market), the models start to decay. The degradation of the models inevitably affects the quality of the statistical output that is based on them. Therefore they should be continuously monitored and re-trained as needed. The lack of a maintenance plan often leads to a low acceptance of the models from business and IT areas.

Monitoring of the model can be done in two ways:

1. A performance-based approach that monitors the performance metric directly (e.g., comparing prediction error). This approach requires the “true” values of the target variable.
2. A data distribution-based approach that monitors the distribution of input variables (e.g., Pearson’s chi-squared test for categorical variable). While this approach may look conceptually straightforward, the distribution can be often difficult to formulate in practice, in particular for new types of data, and first have to transformed into a vector space.

Once the drift is detected, the decision on whether the model needs to be retrained and on the scope of this re-training (e.g., model parameters, hyperparameters, pre-processing) should be determined based on cost factors (e.g., impact of the model, human resources needed, potential risk of break of time series).

The development of ML solutions requires multidisciplinary collaboration. It is important to note that the level of involvement of different types of expert changes based on the stage of development (e.g., during the proof-of-concept phase, data scientists are more involved; during the transition to production, IT experts are more involved). From the organisational point of view, this change of involvement should be clearly understood, specified and agreed among all actors from the beginning to the end.

Having a minimal set of metadata and proper documentation of this information is particularly critical for model monitoring and retraining as these processes take place after the initial model development (sometimes after a few months or years) and are conducted by those who were not involved in the model's development. During the transition, the set of metadata needed for the selected monitoring and re-training methods also needs to be set up. (For more detail, see the Model Retraining Theme Group report)

3.5 Organisation structure

The organisation of work is changing with the introduction of ML in the production process. The role of IT, methodologist and subject matter experts changes when data scientists step into the production process. There are different approaches for this change, such as:

- Centralised: A central team of IT experts and data scientists help the domain experts to set up the ML process in production and then move on to the next project after hand-over of the finished ML model to the domain experts.
- Distributed: The data scientist is placed in a given domain where he/she cooperates with the domain expert and IT to implement ML in the production processes.

There are a number of related challenges. Development of statistical solutions relies on input from IT areas, as well as business areas (methodology and/or statistical/domain areas) to ensure the solution meets business needs, such as maintainability, control, transparency, and flexibility. When developing ML solutions, the addition of a 'data scientist' role brings key expertise that can help bridge the 'concept and language gap' in discussions with IT and business areas.

When NSOs are considering the organisation structure and job function, they may consider a number of questions:

- Where does this data scientist role fit in the ML solution development structure – within IT areas, business areas, in a multi-disciplinary team, or in an entirely separate team of data-scientists?
- Are these arrangements ongoing or temporary/project-specific?
- To what extent do all teams need to build data science capabilities?

Business areas need to become increasingly familiar with concepts/tools traditionally considered part of IT (such as code version control, app hosting) in order to assist and inform this development work (particularly if using emerging tools / innovation environments) and to undertake ongoing maintenance.

They also need the ability to work with emerging environments and tools and have the language and conceptual background to inform development of ML solutions. Business areas need to build a number of related skills such as:

- machine learning;
- data science more-broadly (including programming languages);
- how to harness algorithms and to adapt / supplement emerging data science tools so they meet the needs of an NSO;

- how to work with big data; and
- how to develop innovations throughout the cycle, and effectively work in multi-disciplinary teams.

IT support teams are needed to support both the development and the ongoing maintenance of an ML solution. During the development of ML solutions, business areas may need assistance from IT areas / data scientists to learn how to use the new tools and environments (and to resolve issues, for example where business areas have a lower level of access / permissions than IT staff).

As a solution matures from being a trial / point solution to a broader organisational solution, onboarding support also needs to transition from servicing a few early adopters, to accommodating a broader and more diverse clientele.

Because every new ML solution and supporting infrastructure requires ongoing IT support, organisation of IT staffing can become increasingly complex, with a growing number of solutions to support, potentially over multiple evolving environments with a diversity of components as the organisation transforms.

To reduce this load, NSOs may also want to give business areas greater control over maintaining their statistical products. In practice, this could mean enabling subject matter areas to directly retrain models, test and implement into production without needing IT teams to do this work. This change in roles for business areas introduces additional challenges.

A system that enables business areas to directly maintain their statistical product needs to consider the following:

- Create a stable, controlled pipeline that also enables flexibility for business areas to explore and adapt systems as needed (e.g. respond to changes in the data).
- ML solutions that are easier to implement and support technically may not be easy to use by business areas, or provide the level of control / transparency / maintainability needed.
- If the maintenance requires coding capabilities, is the ML solution using a programming language that the business area can develop.
- Big data can introduce additional challenges for efficient processing / storage, such as data formats, tools and languages that may be less commonly supported across the organisation.
- Business areas need to build new skills – the design of the ecosystem and the state of evolution (disparate and evolving), as well as the level of support (spread thin), all have the potential to stretch business areas away from core knowledge (methods, data collection, data providers / systems, data users).

The design of organisational structure and job function may be part of the solution to help manage these issues. Questions include: how best to incorporate ‘data scientist’ skills into the structure of the organisation, and how does this need to evolve as the organisation matures.

3.6 Stages of innovation in organisations

Continuous innovation and improvement is an essential part of creating a successful ML ecosystem. To achieve this, it is necessary to identify the key components of that innovation cycle and how to start start introducing them into an organisation's existing eco-system.

Understanding the machine learning life cycle.

The machine learning life cycle is a process of continuous development, analysis, and improvement. The next few years of ML development is likely to lay the groundwork for the growth of new enterprise ML applications.

Today, building a scalable backend for ML involves integrating many tools - a modern stack may include separate vendors for data versioning, labelling, feature stores, model management, monitoring, etc. Provisioning, integrating, and scaling each of these tools requires considerable engineering effort and time, and as a result slows development.

In future, a broader standardisation in the de-facto infrastructure stack is likely to occur as model architectures further standardise and turn-key solutions integrate with adjacent vendors in the MLOps lifecycle.

Successful ML initiatives require taking control of the entire ML Ecosystem. In addition to the factors discussed above, orgainsations also need to consider the importance of staying connected to new developments and having more trusted and explainable business use cases.

Innovation can mean different things to different people, but the focus should ultimately be on what benefit ML brings back to the organisation. Of equal importance for official statisitcs is ensuring the creation of a responsible Machine Learning ecosystem

Innovation in practice

A key challenge for developing statistical products is bringing together secure data with emerging tools into the same environment. One potential approach is to have a number of environments to suit different stages of the innovation cycle; that is, to have separate environments for initial feasibility assessment or for proof-of-concept development and testing.

Initial stages of the innovation cycle may not require full-functionality infrastructure or complete data - it is possible to make good use of samples, public data, fake data, and partial environments. This allows an environment with lower security needs and/or greater access to emerging tools, services and methods.

However, using research environments that are partly-functional and partly-supported can bring challenges for both this stage of the assessment, and when taking forward towards productionisation – for example:

- This approach requires additional effort to build and support these environments, including IT support for the business areas and research teams using the environments.
- When going through the more-advanced stages of the innovation cycle, the solution will need to be assessed and developed on the full dataset (not a sample or fake data) and in a more-complete environment (e.g. may require parallelisation capabilities; different programming languages; different security / roles).

Building in iterations within each stage of the cycle can also be a useful approach. However, this may also require interim solutions that integrate with corporate platforms and processes (adding to the effort and complexity of the build).

4. Moving Forward

Statistical offices seeking to adopt machine learning at scale need to invest significant resources in developing the appropriate enabling environment. This is a complex task involving structural transformation and change. For this reason, it requires a multidisciplinary, strategic-level approach that builds capacity and processes, and which aligns with the organisation's wider ecosystem.

This report has identified a number of different issues across technology, people, data and process that need to be addressed. The statistical community can take these forward with the following next steps:

1. Investigate ways to establish technically-oriented infrastructure within statistical systems, especially
    - connection to open source; and
    - connection to cloud.
2. Exchange knowledge on where to find components and the best way to integrate them into the standard production process.
3. Develop MLOps frameworks and skill-sets tailored to statistical offices' individual requirements and maturity level.
4. Continue work on the capability maturity model, including the question of
    - what are the potential ways to move forward to the next level of maturity.
5. Explore the different organisational structures needed for supporting business areas with the use of machine learning.

The discussion on how to build the enabling environment for ML in statistical organisations has just begun. As there is still a lack of experience and knowledge in the statistical community, it will be important to continue collaboration in this area, through, for example, international expert groups under the umbrella of the UN

As with many innovations, the transition from proof-of-concept to production takes time and investment. Ultimately, each statistical office will need to implement a strategic, coordinated and tailored approach across their organisation if they are to realise machine learning’s significant potential benefits for their statistical production.

5. Acknowledgements

This report was written by the Infrastructure Theme Group of the ONS-UNECE Machine Learning Group 2022 in November and December 2022. The report summarises discussions held by the group during its meetings held between March and December 2022. The group was chaired by Jakob Engdahl, Statistics Sweden, with regular contributions from Matyas Meszaros (Eurostat), InKyung Choi (UNECE), Joni Karanka, Eric Deeben, Alison Baily (ONS), Anton Karlsson (Statistics Iceland), Jenny Pocknee (Australian Bureau of Statistics) and Jason Solomon (Statistics Canada). The views and opinions expressed in this report do not necessarily reflect the official view of the organisations mentioned.

The ONS-UNECE Machine Learning Group 2022 is a platform for international research collaboration, knowledge exchange and capability building for Machine Learning in Official statistics. Bringing together over 400 members from 65 different organisations around the world, the group runs regular presentations, discussions, and research activities to enable colleagues from across the global statistical community to learn from each other’s insights and experiences. Our focus is on understanding the added value of ML for official statistics and how we can develop approaches to successful integration into production.

6. Annex: Table for ML Capability Maturity Matrix

	Dimensions			
	Institutional / people	**Information**	**Process / method**	**Technology**
Initial	A few individuals experimenting in silos. Low awareness of business value of ML. ML projects are opportunistic and ad-hoc	Individuals acquire and process necessary data on their own. There are difficulties in obtaining data. No standard for documentation	There are no clear handover points and individuals are responsible for activities beyond their responsibilities. There is no standard method or common practice	There is no corporate approach to ML software. Individuals install programs and software needed for ML on their own
Spreading	Groups of individuals form networks of knowledge-sharing. Some ML roles and responsibilities are identified and formalised	There is effort in creating a common pool of data sets for similar ML issues. Increasing awareness on the core metadata on models	There is common understanding of general processes involved, including activities, responsibilities. Some common methods and services are identified.	Essential software is whitelisted in the organization but the approach is either piecemeal or lacks key tools. Scope of software considered for corporate support in IT is limited to those that are directly related to ML only (e.g., Python, R).
Widespread	ML work becomes more centralised. There is regular training opportunities for staff. There exists a corporate-wide strategy for ML	There are standardised means of accessing and storing data for training, testing and inference. Minimum set of core metadata is established. Documentation is done in a standard way.	Use of common methods and processes. ML models are developed in coordination with the aim to be modular and portable, facilitate sharing and re-use.	The organisation has software platforms that specifically deal with requirements raised by ML. IT services are familiar with needs for ML and take these into account in their architecture.
Mature	Essential capabilities are fully integrated in the organization providing ML infrastructure. ML development is streamlined and monitoring is automated.	There are central repositories for storing important information (features, models, metadata, code) that can be reused and reproduced. Information on previous runs of the same models is easily accessible	Processes (or parts of them) are automated. Models are continuously reviewed to monitor qualities and improve their efficiency. There is an established governance plan for models	Procurement of new hardware or cloud computing services for ML are part of regular corporate planning. Needs for new software and hardware are continuously monitored and planned at a corporate level