AIML4OS

Artificial Intelligence and Machine Learning for Official Statistics

Newsletter 6

Welcome to our Newsletter #6 for ESSnet AIML4OS! Here you will find project updates highlighting progress, achievements, events and all news.

PROJECT OVERVIEW

The main objectives of AIML4OS are to explore the use of Artificial Intelligence/Machine Learning (AI/ML) for the production of official statistics and to implement innovative solutions for statistical products and processes. This four-year project started in April 2024, with activities structured in the following work packages

OVERARCHING WORK PACKAGES

WP1
Project management and coordination

WP2
Communication and community engagement

WP3
ESS AI/ML lab: Technical infrastructure and organisational setup

WP4
AI/ML state-of-play and ecosystem monitoring

WP5
Standards, methodological and implementation frameworks

WP6
Knowledge repository and training material

USE CASES

WP7
AI/ML on earth observation data, satellite imagery

WP8
Statistically valid and efficient editing and imputation in official statistics by AI/ML – with a special focus on editing

WP9
Imputation focus - Statistically valid and efficient editing and imputation in official statistics by AI/ML – with a special focus on imputation

WP10
From text to code - Experiences and potential of the use of AI/ML for classifying and coding

WP11
Applying ML for estimating firm-level supply chain networks

WP12
Large language models

WP13
Generation of synthetic data in official statistics: techniques and applications

In this issue: Interview with Frankie Kay, Progress on Supply Chain Network, Land Cover / Crop Mapping updates and AI Standards ready to share quality framework.

READ MORE...

AIML4OS

INSIDE THE AIML4OS PROJECT AND ITS IMPACT ON EUROPEAN STATISTICS

An interview with Frankie Kay

We interviewed Francesca (Frankie) Kay, Management Board sponsor of the AIML4OS project. She has an extensive background in data, technology and transformation, with over 20 years of experience in the private sector before joining the Office for National Statistics. She later held senior leadership roles in both the UK and Ireland, and is currently leading national data and statistical services while continuing to support AIML4OS from its inception through to its completion.

Hi Frankie, can you tell us how the AIML4OS project started?

I first heard about AIML4OS in 2023, when Eurostat announced a grant to set up a One-Stop-Shop for AI/ML in Official Statistics. Seeing an opportunity to accelerate AI/ML adoption at the CSO, I persuaded our Management Board to lead a consortium for the grant. The initiative attracted strong interest, and we now have 16 member states participating.

What made Eurostat decide to roll out this project?

The interest in and use of AI/ML was accelerating back in 2023 as it continues to do so today. It was a topic of strategic interest and therefore Eurostat recognised its importance and looked to find a way to accelerate its adoption across member states. The project was created to provide guidance and support in the implementation of AI/ML solutions for Official Statistics within quality, methodological and implementation frameworks that facilitate new statistical products and processes, leveraging cutting-edge research and developments in the field of AI/ML.

What's the added value of international collaboration in a field like artificial intelligence?

We are all operating in a world with limited, and in many cases, shrinking resources within our National Satistics Offices, so collaborating to make best use of those resources is vital. That is one of the huge strengths of the AIML4OS which is bringing together and sharing knowledge and expertise as well as providing concrete and practical use cases and material. By working together we are able to work on more areas in the field of AI/ML then individual countries might be able to do on their own. The project is providing training material, guides, code and examples of specific use cases which are being made publically available via our GitHub site and Eurostat's CROS portal. We showcase our work at conferences and events to build up the AI/ML community to share that expertise and knowledge as widely as possible.

How can this project contribute to the development of AI in European INSs?

Key to the success of the project is to make sure that practical applications are developed. We are looking to achieve this in a number of ways. For example, we are delivering a number of specific use cases covering different aspects of the statistical process and data.

Can you give us some examples?

We are looking at whether we can apply previous work on Supply Chain and Earth Observation models to other countries and we are also looking at how to use Machine Learning in editing and imputation. Code for these use cases will be shared by uploading it to GitHub for any country to use, and we will provide training materials to support their use. Another practical example is that we are providing a sandbox for experimentation using Insee's Onyxia Platform together with guides on how to create ML platforms within NSO's. We are using both Hackathons and Funathons as ways of helping countries to get practical experience of using Onyxia and LLM's. These practical applications are then supported by providing methodological and quality guidance, as well as implementation frameworks for MLOps.

What are the next steps for the project?

We're half way through the project so the focus over the next 2 years is ensuring we complete all the remaining milestones and deliverables as planned. We will continue to provide updates at events, for example this year we will be running a session on the project at the Quality Conference in Sibenik and a number of other work packages will also be presenting on their progress so far. Alongside completing the project is the need to work with Eurostat to make sure that the AIML4OS is sustainable once the official project has been completed.

What is the most rewarding aspect of coordinating an international project of this kind?

I've really enjoyed sponsoring the AIML4OS. It's a fantastic opportunity to work with many different countries on a topic that is critical to the future of statistics. I've got to work lots of amazing people and learn a lot about AI/ML. Of course it's not without its challenges, but luckily for me the hard work is handled by the coordination team and the work packages teams.

What has been the biggest challenge?

Probably the hardest part was getting the project established in the first year – lots of scoping, structures, and processes. These are essential but not the most exciting statistically! Maintaining momentum is also a challenge, as everyone balances this with their day jobs, though enthusiasm has stayed high.
It's great to see concrete deliverables taking shape as we enter year 3 – hard to believe we're already halfway through. I'm looking forward to seeing everything come together over the next two years, delivering practical resources for countries in and beyond the EU, and of course celebrating the project's completion in Dublin over a pint of Guinness!

PROGRESS ON SUPPLY CHAIN NETWORK RECONSTRUCTION

From model training to pipeline deployment and international adoption

The WP11 team finished a **first round of training several firm-level network** reconstruction models on the Portuguese firm-level supply chain network dataset created in the project in 2025. Training was done by our Polish colleagues on-site in Portugal, due to the confidential nature of the data. We used a software pipeline that was developed last year. **The LightGBM model was selected** for further training and finetuning. XGBoost, Histogram-Based gradient Boosting and Random Forest will be used for helping to prepare the training data and feature engineering as well as the optimal set-up of LightGBM. Meanwhile, work continued on creating the firm-level supply chain network reconstruction pipeline in which the trained model will be implemented. **This pipeline will also include the addition of weights to the reconstructed buyer/supplier relationships**. Poland joined Ireland and the Netherlands as countries where the Statistical Institutes will apply the pipeline. Validating the results will be used for further improving the models and pipelines.

SCALING LAND COVER AND CROP MAPPING PIPELINES

Standardization, validation, and expansion across countries

On 14 and 15 October 2025, the WP7 team **met in person in Warsaw**, Poland. On the morning of the first day, we presented the current status and progress made. Both the land cover and the crop type teams had made impressive progress. Both models were run locally for various countries and regions, as well as on the Copernicus Dataspace Ecosystem (CDSE). This resulted in **a lot of experiences and insights**. In the afternoon, the teams had a deep dive into the pipelines and discussed **how to generalize and standardize** these into common pipelines, including quality metrics, made available on the CDSE. The second day we planned how to implement and test (ground truth) these pipelines. Then we will begin **running these models on larger regions** and even entire countries. Meanwhile, other countries are showing interest in our results and want to apply these models to their territories.

AIML4OS

ADVANCING AI STANDARDS IN OFFICIAL STATISTICS

Insights from the WP5 Workshop

At the WP5 Standard Meeting (17–19 March 2026, Heerlen, Netherlands), partners held key discussions on **developing standards for quality, methodology, and implementation** of AI and ML in official statistics.

The workshop focused on shaping a **unified framework** that can be consistently applied **across statistical models**, machine learning, and generative AI, while aligning with ongoing European efforts to further develop the Quality Assurance Framework (QAF) and the Code of Practice (CoP).

The team has already shared the **quality dimensions with all project partners** to collect feedback. This exercise will be repeated for methodological and implementation principles in May (more information will be shared during the town hall meeting in April).

In addition, it plans an interactive workshop session at the **AIML4OS project meeting in Šibenik to gather input from all partners** and to connect the developed concepts with concrete examples, challenges and best practices from the work packages.

Next steps in WP5:

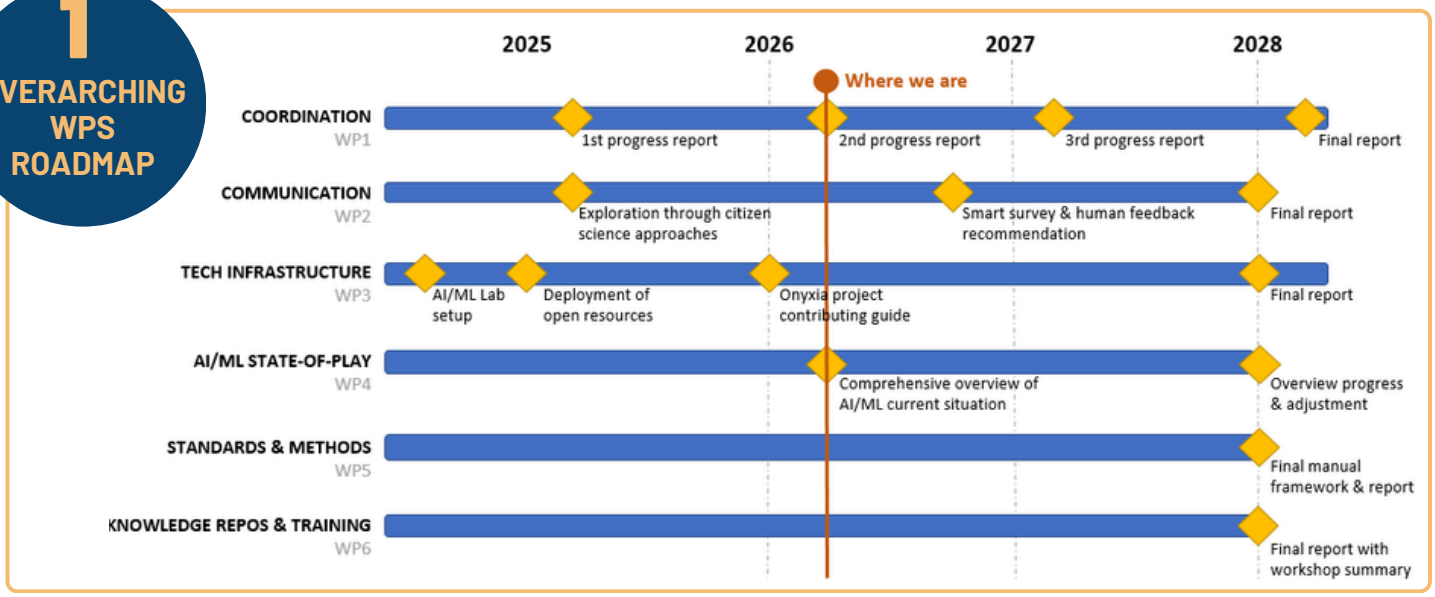
- **Finalisation of quality dimensions**, methodological, and implementation principles by early May 2026 (to be shared with all partners in mid-May after internal WP consultation)
- **Preparation of a session** for the AIML4OS project meeting
- By the end of 2026, most of the framework (covering quality, methodology, and implementation) is expected to be in place, with **further refinement and illustrative examples** planned for the first half of 2027

AIML4OS

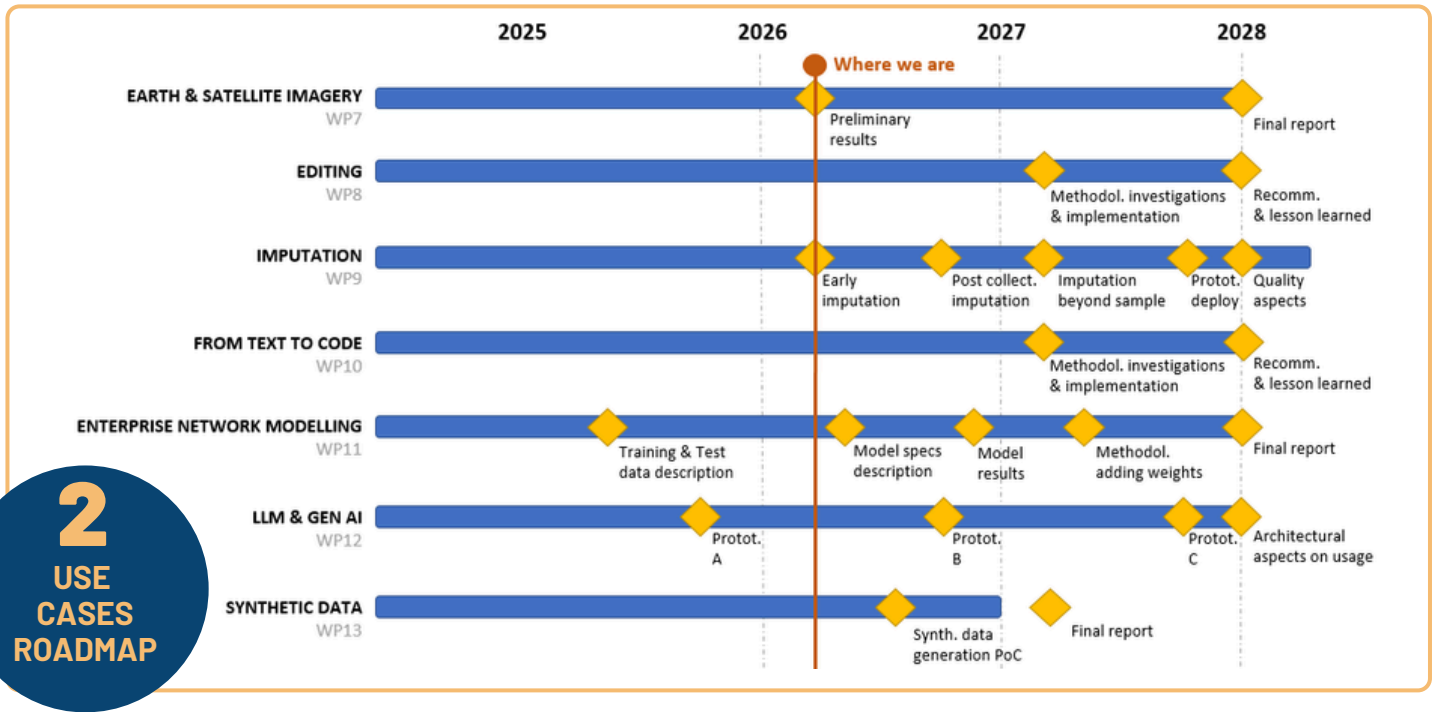
For each WP involved, the project is divided into several phases. Below are descriptions of what will be achieved and when.

PROJECT OVERVIEW

1 OVERARCHING WPS ROADMAP



2 USE CASES ROADMAP



Subscribe Newsletter

To stay informed about the latest developments of the project, please subscribe to the newsletter

AIML4OS WEBSITE

FOLLOW US ON LINKEDIN