



Experiences of using preventively the data-driven risk analyses experience for non-IACS EAFRD measures and piloting the use of ARACHNE for daily processes and look to the future opportunities of using AI and machine learning-based data analysis.

A visit report from the seminar on the 18th and 19th of January 2022 in Estonia (Tartu), organized by the Estonian Paying Agency (ARIB)

Table of content:

| | |
|--|-----------|
| ABSTRACT | 3 |
| PART I MAPPING THE AREAS, METHODS, SOFTWARE AND RESOURCES USED FOR THE DATA-BASED RISK ANALYSIS | 6 |
| Subsection 1: Prerequisites of general data-driven approach | 6 |
| Background and starting point | 6 |
| Our solution | 7 |
| Lessons learnt | 9 |
| Subsection 2: Data-based risk assessment process, developments and integration to daily processes | 10 |
| Risk assessment as a process | 10 |
| Risk-based administrative control: Control Unit view | 12 |
| IT developments: integrating risk-based approach | 16 |
| Lessons learnt | 19 |
| Data challenges | 19 |
| Organisational challenges | 20 |
| Challenges with new IT developments | 20 |
| Future challenges | 20 |

| | |
|--|-----------|
| Subsection 3: The future of data use, based on AI and machine learning opportunities for sustainable agriculture and for result-based CAP, and for protection of the EU financial interests | 22 |
| Background and starting point | 22 |
| Our solutions | 24 |
| Data-driven future of farming, carbon farming and carbon credits | 24 |
| Machine learning as a new approach to data analysis. Four innovative approaches to data analysis for EU support schemes for CAP | 26 |
| Machine learning-based data analysis for designing interventions - as smart prevention of the EU financial interests | 27 |
| Machine -learning based data analysis for increasing accuracy of OTC control sample - as smart detection of the EU financial interests | 31 |
| Agricultural big data project in ARIB | 35 |
| Sharing best practices by the initiative of AC DG AGRI according to HzR art 13. | 36 |
| Lessons learned | 37 |
| PART II EXPERIENCES OF USING ARACHNE RISK SCORING TOOL | 39 |
| The problem | 39 |
| Our starting point | 40 |
| Our solutions | 43 |
| Lessons learnt | 45 |
| CONCLUSIONS | 49 |

ABSTRACT

On the first day of Smart Pro CAP's visit to Estonia, the ARIB introduced its experience at the beginning of the development of a risk assessment system for CAP rural development support since 2012, in which a risk assessment system has been developed as part of the overall management system. When introducing the risk assessment system for rural support, it is important to point out that.

- Its main focus was on the prevention of possible irregularities and fraud, i.e. the identification of higher risk applications before deciding to grant support.
- Proving the risks of creating artificial conditions has proved to be a challenge, given the short procedural timeframe.
- After five years of testing risk-based management controls outside the IT system, i.e. based on Excel, we are integrating this approach into the IT system, which also required rethinking the business process and the entire risk assessment and feedback system.
- We also shared Estonian experience in developing automatic controls already at the stage of preparing an electronic application and automatic evaluation, which requires the interconnection of many national data but is an effective measure to prevent violations.

Through risk-based administrative control of applications, i.e. the so-called red and green corridor approach, grant applications in the amount of 236 million euros have been processed since 2017 for riskier support schemes. The amount of EUR 22.4 million in the Red Corridor applications, which has been rejected, has made it possible to award risk-free or low-risk grant applications in the same round. The amount of grant applications withdrawn as a result of risk-based administrative controls is also EUR 12.8 million. The number of both rejected and withdrawn grant applications has been higher in the red corridor than in the green corridor.

As for the ARACHNE risk scoring tool, we see it being more valuable at the moment in the functionality of individual queries in the relationships between companies and

its ability to visualization. This is especially important when checking the background and connections of foreign persons in the context of administrative control.

We have mostly been able to automate the compilation of the data needed for ARACHNE.

We have also started testing the suitability of the risks calculated by ARACHNE and their coincidence with the risks calculated by the ARIB. There are currently no very good results in this area. At present, the ARIB's own risks have proved to be more practical both for supporting administrative control and for compiling mandatory control samples. But we have an idea how ARACHNE risk scoring functionality may provide in the future the opportunity to automated risk scoring integration as part of our risk assessment system if risk corresponding to our needs and reality are developed and risk scoring information interfaces are developed as well.

We have a vision, of how the integration of ARACHNE risk scoring functionality into our risk assessment system - may provide in the future the opportunity to free up our risk analysts workflow from doing routine and repetitive tasks for calculating risks. A prerequisite of this vision is an improvement of the ARACHNE risks to correspond to our needs and reality and interfaces of sharing risk scoring information with our IT system are developed as well.

But there is also the alternative approach that ARACHNE will focus more narrowly on specific risks as a tool of providing enhanced transparency of EU-funded projects and will not challenge to cover all possible risks- which might be obviously impossible.

The second day of the Start Pro CAP visit to Estonia was dedicated to gaining an overview of the latest developments and application possibilities of modern data analytics in agriculture in general, as well as for the development of EU CAP support systems and the protection of EU financial interests. In this context, it is important to distinguish classical human rule-based data analysis and methods from data use and analysis based on machine learning and artificial intelligence models. On the second day, we were working on mapping the possibilities of using machine learning and artificial intelligence in the context of CAP rural support.

During the day we mapped at least 4 tested or debated ideas for using machine learning models to design CAP rural development support, developing modern proactive procedures, or for improving the accuracy of office sampling, which would help reduce the overall administrative burden and provide at the same time smarter protection of the EU financial interests.

We are also very grateful to DG AGRI's anti-fraud correspondent (Gianluca Frinzi), for his active and fruitful participation, valuable feedback, and presentations of the expectations of more effective protection of the EU's financial interests for the upcoming Result Based CAP support period.

PART I Mapping the areas, methods, software and resources used for the data-based risk analysis

- *Improves the awareness of the PAs' staff on how the Estonian paying agency developed in the 2014-2020 programming period created a risk management system based on data analysis, focused on the risks of rural development support to identify possible irregularities or suspected fraud at the earliest possible stage. This was the starting point of elaboration of the overall comprehensive risk management system, which is also integrated as part of the entire management system.*
- *Points out main learning points of which components to develop and the challenges to be addressed in developing and integrating classical rules-based risk analysis into the IT system for processing grant applications. Also about the fact that it is one thing to identify risks, but it is much more difficult to prove them in a limited time and what are the alternative options.*
- *Shares an idea of possible/desirable future state on how to use modern AI and machine learning-based data analysis models, at different stages of the CAP rural life cycle (from designing interventions to smart control sample) and to protect the EU's financial interests during both the detection and prevention phases.*

Subsection 1: Prerequisites of general data-driven approach

Background and starting point

At the beginning of programming period of 2014-2020 ARIB was at the point when we decided to start using more proactive approach to fraud detection and bring risk based sample to new levels through promotion of data-based risk analysis.

Also the need for data-based risk management in grant management processes came from the developments in fraud detection processes - in order to detect fraud we needed smarter processes since there is never enough resources to deal with

every application in depth fraudulent cases require. So finding risks and scoring applications led us to the red-corridor approach and helped finding fraud more easily.

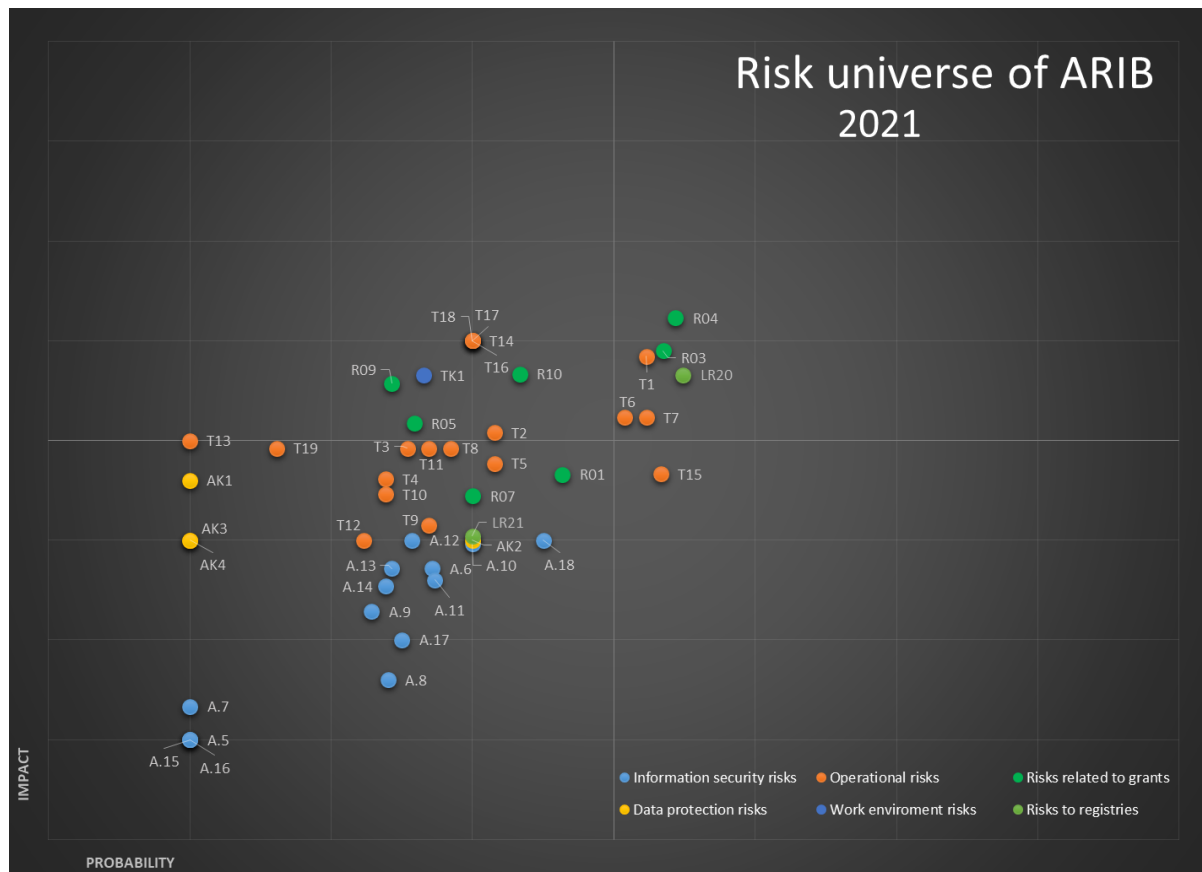
At this point in time the whole risk management process in ARIB was functional but fragmented, risks were assessed formally and they were not the real basis for management decisions. Making new developments in grant related risk management gave us the chance to gradually improve also the whole system.

Our solution

By 2020 ARIB has risk management system that consists of six elements:

- Work environment risks - based on national law
- Operational risks - based on accreditation criteria
- Information security risks - based on ISO standard
- Data protection risks - based on GDPR requirements
- Risks related to grants
- Risks to registries

The risks related to grants and registers are based solely on data analyses, the rest of elements use data when it is available (e.g. information security has input data on the number of data security incidents). All these elements are included in annual risk analyses showing to the organisation the developments and findings of all risk management areas. Annual risk analyses also binds these elements into one unified risk universe of ARIB which shows in a visualized way all areas and risks of ARIB, also changes between years can be observed in order to evaluate the changes occurring and it also helps to prioritize mitigation activities.



Annual risk analysis gave us chance to integrate all the elements of risk analysis. Concise and visualised manner of this document grew interest of all parties in the organisation to understand the nature and results of risk management. Therefore in order to raise awareness all elements of risk management they were all gradually included into one. It was easier to do so since we had had a principle of using the same scale in risk assessment for all elements prior to integration into annual risk analyses.

The second step in making integrated risk management more practical was to integrate mitigation activities into the annual working plans of the units. This process was supported by a project with the aim to map and describe the entire management system in a single management system description. Through this document, all elements of the management system like strategy, budgeting, work plans, risk management etc. were tied together and described as management year clock and it is the basis for all managers working plan. This added the actual requirement to use the results of risk management in daily processes.

Lessons learnt

The main issue of the data-based approach is the issue of **data availability**. On the one hand, the IT systems designed for processing of grant applications and saving audit trails may not be as good for data analyses - the data may be saved in a format that is not analysable or all kinds of real-life special situations are stored differently so the data cannot be used uniformly and directly for analyses. And on the other hand, in order to do quality data-based risk analyses you need more data than just data from grant implementing IT systems have.

Also, there is an issue of constant **change of actual risks** - risks cannot be constants once made up and only weighed regularly. Especially in fraud prevention where fraudsters always think of new ways to unduly received funds from grants. It means that risk management systems must be very flexible and always look for new trends and information. This also makes it difficult to automate risk scoring as by the time something is automated it might already be expired.

Processes of integrating all risk management elements and **integration of the risk management** as a whole into a comprehensive management year clock have made the whole risk management system a practical working tool but the data-based and visualized presentation have made it more easily acceptable for managers to use.

In future we see the annual risk analysis to grow even wider in range as not only data analysts are building its contents but the results and finding will be analysed by wider range of parties who have used data-based input in their work.

Subsection 2: Data-based risk assessment process, developments and integration to daily processes

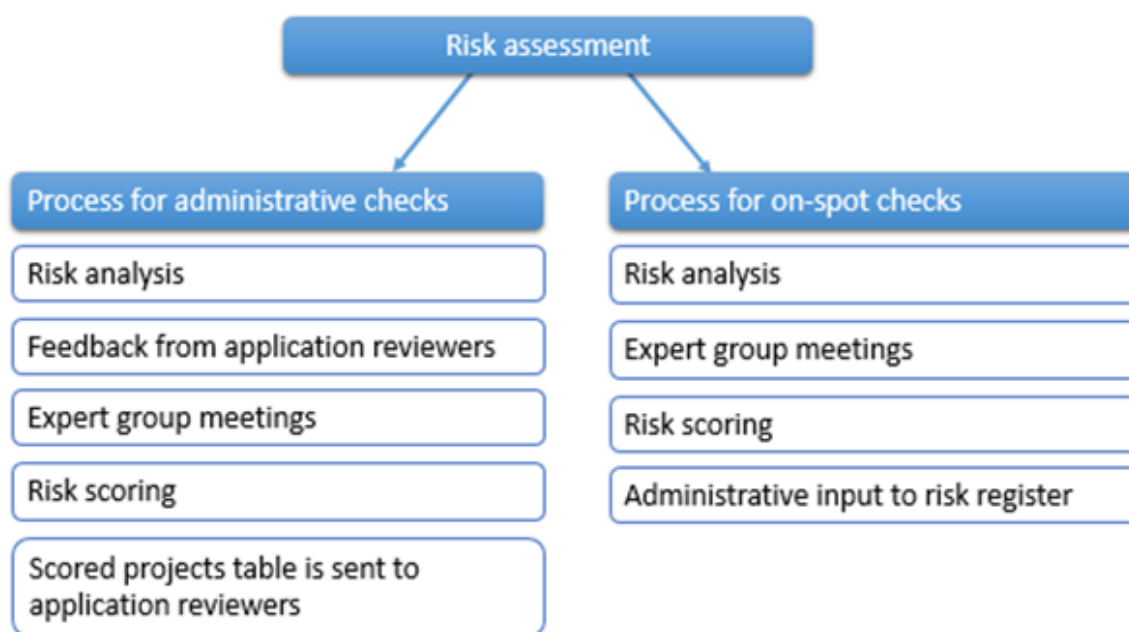
Background and starting point

Risk assessment as a process

In this chapter ARIB risk assessment AS-IS process is being defined from the risk management point of view. It is important to find out the current situation and developments in order to describe how ARIB has developed and is developing the protection of EU's financial interests.

At the moment there are two different processes for risk assessment in ARIB:

1. Process for assessing risks for administrative checks („red corridor“ approach)
2. Process for assessing risks for on-spot checks.



The chart above shows the main steps of both processes. The general steps of the two processes are quite similar: analysing on historical data how previous risk factors have worked, are they statistically relevant, gathering knowledge and experience from expert groups and lastly scoring the applications.

The risk-based administrative checks

As we do not have a lot of resources, it is important to prioritize for which measures the administrative check risk assessment is needed. We use the heatmap method for pointing out riskier measures and to also show where the budget deficit risk is the highest. The purpose of that is to point out riskier projects to application reviewers in order to observe more closely riskier projects before grant decisions. From the results of analysing the sum of refused projects (w/o budgetary and ranking reasons) and withdrawn projects percentage, the administrative check risk assessment has been effective. The main difference from the on-spot process is that administrative checks risk assessment is fully outside of our IT system. This is our main pain point at the moment, which is why we have started to develop an integration process into our IT system.

Risk-based sample of the on-spot checks

With on-spot checks for investment measures we deal with one measure at a time depending on its time schedule. We compare project risk scores with on-spot checks' results and look at each risk factor individually – does it sort out riskier projects. For IACS measures we're comparing the project's risk scores with on-spot control results. On-spot checks risk assessment process is in our IT system, except risk scoring part.

The methods we use for analysing risk factors' effectiveness are descriptive statistics, frequency tables for each risk factor in order to see if the risk factor is identifying riskier projects, Chi-square test. For IACS measures we use additionally correlation, regression and dispersion, as we have more data available.

Data sources

We use data mostly from our own warehouse and registers, data from the Estonian Tax and Customs Board and data from Business Register.

Software used for risk analysis

Here is the list of most used software and functionalities that are used for risk management purposes.

- SAP Business Objects – data inquiries
- Microsoft Excel – data cleaning, analysis, tables, pivoting, risk scoring
- R studio/R – data cleaning, analysis, risk scoring
- Microsoft Power Query – appending, merging data tables
- Microsoft Power BI – visualisation, used in 2021 for annual general risk report.

Examples of ad-hoc horizontal risk analysis to test new approaches of risk scoring

1. Geographical cumulation of RD grant
2. Disappearing client risk analysis

From the horizontal analyses the main results were that quite often it is important to analyse some hypothesis horizontally, because horizontal analyses help to generate new horizontal risk factors to take into use for different measures. For example, from the geographical cumulation analysis the main result was that we took cadastral parcel risk factor into use in both processes and for different measures.

From disappearing client risk analysis we found out that we should use more external data in order to prevent fraudulent behaviour and one way of doing that is through data sharing with other public sector organisations. At the moment we are working on it and also developing a new business process for that.

While developing our risk analysis capabilities we also saw that some new daily business processes should be developed. If the ordinary approach of paying agency processes is designed to react to the client's applications, meaning that interaction is initiated by clients. In the case of proactive actions, the interaction should be initiated by the agency and quite often it becomes obvious that we have to deal with the client as a whole not only with one specific issue. In case of risk of disappearing, all running projects or commitments with IACS or non-IACS measures are under the threat and possible financial losses should be proactively prevented or mitigated in trustful cooperation with the client.

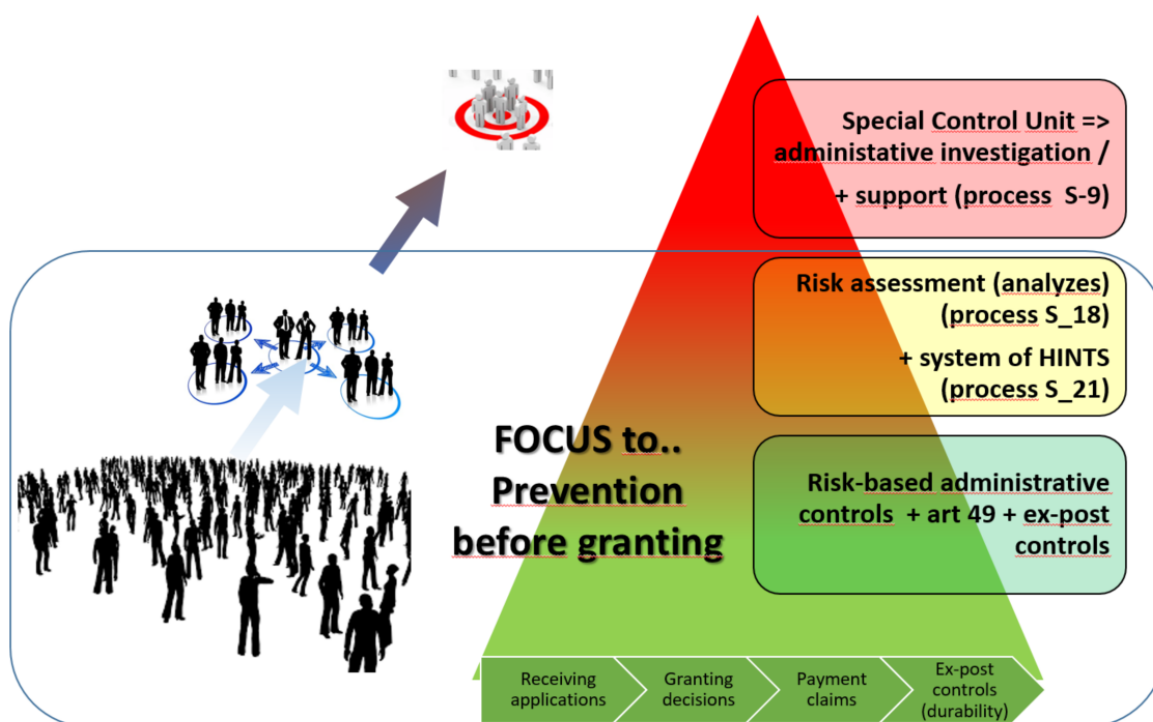
The most important is prevention, so we should make our processes more supportive towards discovering and detecting the signs of disappearance as soon as possible, which is also one of the processes for protecting the EU's financial interests.

Risk-based administrative control: Control Unit view

Developing anti-fraud system in ARIB

The anti-fraud system in ARIB starts with risk-based administrative controls which is supplementing traditional mandatory check-list based control system of all conditions.

Anti Fraud System elements in ARIB (Estonia)



A risk-based approach to administrative controls takes into account risk assessment and hints system and then if needed for more complex issues, cases can be taken to assistance advice to the Special Control Unit.

The Special control unit

- is carrying out special administrative investigations,

- is the contact point in the communication of law enforcement institutions,
- has an internal advising-assisting function to support units in complex or suspected fraud cases.

Additionally, the Special control unit is responsible for carrying on the ex-post checks and handling Irregularities Management System reports to OLAF.

Over time the Control Unit has been developing new practices and taking more actions in the stage of applying in order to prevent possible fraud before granting or payments are made. A positive side effect of this preventive approach is directing grants to non-risky applications in the same round. As ARIB has noticed mostly riskier applications apply the maximum possible amount of EU grant and less risky applications only according to their needs. This means that rejecting one risky application gives the opportunity to finance more than one non-risky project.

Example of the risk-based new approach of modifying practice of proceeding applications. In 2021 the new risk-based approach was granting with the additional ancillary condition to those applications where the self-financing capacity was lower and who could not sufficiently prove the self-financing capacity even during the administrative control. An additional condition was the obligation imposed on the applicant by the ARIB to submit additional documents certifying the ability to self-finance by the specified time.

Consulting and assisting application reviewers

The main goal of advising is to increase the skills and knowledge of application reviewers in controlling risks and use the experiences and knowledge of the Control Unit investigators in handling complex cases and collecting evidence.

The statistics show also that the number of time-consuming special controls cases has decreased over the last 2 years number of consultation cases has increased.

We have found that learning through doing daily practices - is the most effective way of training and risk-based administrative control still needs to develop further in order to prevent possible infringements and fraud even more effectively.

Based on the ARIB's experience, risk analyses help to successfully identify support applications that indicate the risks of creating artificially created support conditions.

However, proving artificially created conditions is a very difficult task due to the short deadline for processing the grant application.

Possible new risk-based preventive approach

As the ARIB has the ability to identify the risks related to the application before processing the applications, we have considered the possibility to inform the applicants about the risks related to the application before they start processing, with the possibility to withdraw the application. If the person has been informed in advance of the risks associated with the application and the application has been withdrawn before a formal decision is taken on the application, this would allow the person to withdraw the application without any possible additional sanctions or consequences. If a person continues to apply despite prior notification of the risks, there would be no possibility of excluding administrative penalties at later stages of the proceedings if the risks are proven. Restricting a person's access to EU funding for a certain period of time can also be an administrative penalty.

The advantage of this approach is that the applicant has the opportunity to withdraw the illegal application honestly and without detrimental consequences. In addition to the deterrent effect, it avoids unnecessary administrative burdens for both the applicant and the paying agency. The negative aspect of pre-notification of risks can be the emergence of false-positive risks, leading to inaccurate pre-notifications to persons who are not actually exposed to the risks. The negative effects of false positives can be mitigated by well-thought-out wording of risk announcements and by informing the public about the new approach. Such an approach would also be in line with the new results-based grant system, where the EC has also favoured a more "preventive, guiding and advisory" approach, rather than the current "control and punishment" approach.

Shortcomings of the reporting system (IMS)

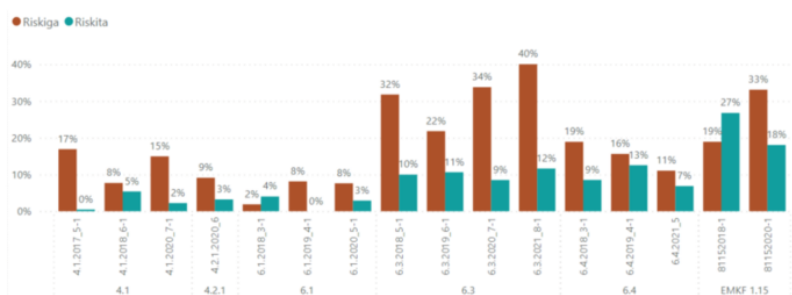
The cases of infringements, which are reported to the IMS system, can decrease over time, but there is no actual statistics about the results of prevention at the European level.

In the Estonian case, we have analysed the results of prevention achieved during consulting application reviewers during administrative checks before granting (prevention in early stages). For example for measure 4.1 in 2020 the prevented damage reached 6,8 million euros - while rejecting 51 risky projects during the consultation process of risk-based administrative checks which gave the opportunity to finance 82 financing non-risky projects. As this result has been achieved before granting decision is made, the IMS report doesn't include this kind of countable achievement against fraud.

The red corridor or risk-based administrative control approach has been used since 2016, but based on a common methodology, it has been used for riskier support measures since 2017 in the 16th round of applications, in which a total of 236.98 million euros has been requested.

Results of rejected or withdrawn risky applications during the administrative check have been typically higher than non-risky applications.

Rejected risky (red) vs non-risky (green) applications by application calls of riskier measure schemes

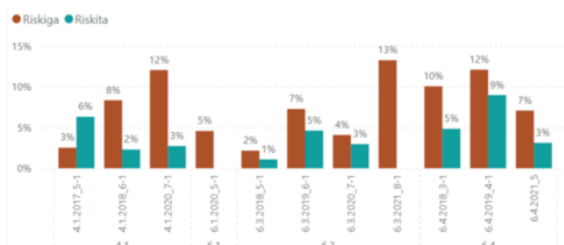


Rejected risky (red) applications

22 440 791 €

- REJECTED (red) = POSITIVE
- REJECTED (green) = FALSE NEGATIVE

Withdrawn risky (red) vs non-risky (green) applications



Withdrawn risky (red) applications

163

12 850 230 €

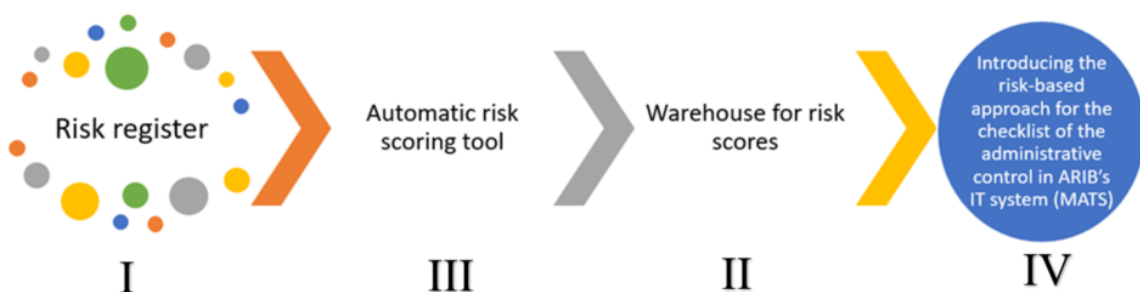
In summary, EU IMS reporting only collects information on irregularities detected after the grant has been awarded, but does not allow the collection of information on potential violations that have been prevented. It is therefore not possible to assess whether or to what extent the reduction in infringement statistics may be linked to successful prevention.

Our solution

IT developments: integrating risk-based approach

As we have practiced a lot of risk assessment and using results for supporting administrative checks outside of our systems, it is becoming more and more crucial to take the process into our IT system in order to have a more concise system for managing risks in order to decrease manual administration workload.

The risk assessment process at the moment is not flexible enough and does not support functionalities that we will be needing in the long term (such as automatic risk scoring for example).



The process starts from the basics: upgrading the risk register for more flexible administration. This is the first step we are taking. The next one will be a warehouse for risk score tables, then the risk scoring process should be automated and in the end, the scored applications risk information should go to checklists for application reviewers. This is the long-term plan.

We started by rearranging the risk register, whose purpose is to manage valid risk factors.

Here the important step was to group risk factors into two categories: 1. measure risks – related to the project and 2. horizontal risks – related to the client. We also had to keep in mind to keep the system as flexible as possible, because we have experienced that risk factors are constantly changing and we have to keep this possibility in mind. Another important aspect was to ensure that the new system should be able to interface with different other systems or external data sources if needed.

The second step in the project was to have a warehouse of risk scores, where we can see all the up-to-date and historical risk scores and make inquiries if needed. In the system, the warehouse shows the most recent scores, but you can make data inquiries about some historical scores if needed in the data warehouse system.

The purpose is to make the scoring more automatic and independent from application rounds and less dependent on risks calculated at different times and with different frequencies.

Furthermore, the system should be more flexible and the system has to be ready for automatic risk scoring in the future.

The third step is actually the one that was eventually left out of the project but it is the most important one in the long-run – automatic risk scoring tool in order to decrease manual risk scoring using up-to-date data. The reason why it was left out was that not enough preconditions were filled and also it would have been quite resource-intensive for this project. The preconditions were that there is an interface between external data sources, it must have some flexibility – not be needing further developments in every step. At the moment we have started negotiations with the external data sources and after that, we will start piloting risk scoring automation in order to know how it would be reasonable to continue.

The last step is making risk factors available to application reviewers in the checklists. The main purpose is for risk management to be part of a core process in order to prevent fraudulent behaviour as soon as possible. If the risks would be effective enough, it can reduce application reviewers' checklist, therefore reducing manual workload.

Checklist for risks - added functionality into IT system

The goal for bringing risks into checklists is to make risk analysis results transparent for application reviewers so that risks would be mitigated. Another purpose is to have a transparent audit trail to ensure focusing on the risks throughout the whole lifecycle of the project.

At the moment, there are already cross and logic checks, which are done against external registers while the client is filling in the application. If the checks fail, the system does not let one submit the application.

There are three types of administrative checks: fully automatic, semi-automatic and manual. For example, in the administrative checks process, there are automatic checks against Land Register, Business Register, Population Register, Criminal Records Database, etc. In ARIB if the application reviewer suspects something suspicious, the hints system is used. Hints are connected to either the application or client and are shown on every checklist.

Integrating risks into daily procedures is relevant because it is crucial that application reviewers should take into consideration all the risks behind the project in order to make the final decision. Another important objective is to get feedback for each risk factor in order to analyse what risk factors were effective and also to get the application reviewer's direct feedback. For application reviewers, it also gives the opportunity to make proposals for changing the risk factors and have a complete overview of previously rated risks and feedback.

Lessons learnt

There are different challenges involved in risk assessment for administrative checks. This chapter focuses on the general challenges that have an effect on the process and mitigating risks. There are four sub-categories for general challenges: data challenges, organisational challenges, IT development challenges and future challenges that we have to take into consideration.

Data challenges

With data we are facing similar challenges that Flemish PA does. Data availability, more specifically collecting and analysing historical data, is an essential part of the whole support lifecycle. Challenges arise when **there is a new programme period and/or measure regulations change, then also data needs change**. Therefore, it is important that the procedural process supports collecting the right data with the right structure from the beginning in order to create and maintain consistency of data structure.

There are some risks that are supported by the data analysis, but are difficult to prove and find evidence on. For example, creating artificial conditions in order to fit into the measure scheme. The difficult part of that is how to prevent it from happening and how to prove artificially created conditions if there are limited resources such as limited time and skills.

Organisational challenges

The challenge that we tackle with in risk assessment is organisational understanding and point of view - **risk analysis is not part of the core process in the whole application life-cycle**. We have developed the process outside of our IT system, but we have not been integrating it into our IT system. Therefore, the next steps would be integrating risk analysis results more into our IT system with the purpose of making risk assessment more as part of a core process.

In addition, when we are making new discoveries horizontally then there are obstacles to overcome - for example proactively taking action when some events take place. **If there are new business processes involved then there are often a lot of unanswered questions: who is responsible and who should decide or deal with the claims.** The main goal in the end is to prevent fraudulent behaviour and deal with problems as soon as possible to avoid dealing with more harmful consequences. That is why this is a challenge for us on the organisational level.

Challenges with new IT developments

Another challenge comes along, which is keeping in mind the long-term plan and the big picture. **A question that arose was how to automate the process, but also to keep it as flexible as possible for changing the risk factors.** This made the elaboration of the requirements description more complex.

Additionally, another challenge we face is **legal considerations** - in order to conduct new developments, there might be legal issues to be resolved. These cases can be challenging and take a long time. Additionally it can be difficult to find common priorities, especially when there is cooperation required with external parties.

Future challenges

There are definitely many challenges ahead in the future that we have to take into consideration. At the moment we have to think about how risk management system would look like in the upcoming years - will it be a rule-based version as it is now or using machine learning models or something in between. Then the questions arise such as **how will the transition take place and how to implement machine learning models into risk management.**

One idea to test is if the risks calculated by ARACHNE and only those that suit us - can be used as ready-to-go input into our risk management processes. If this were possible, the need for ongoing calculation of these risks would be reduced and only the updating of our basic data in ARACHNE should be ensured. This would save risk analysts time for more complex and sophisticated data models and analyses.

Subsection 3: The future of data use, based on AI and machine learning opportunities for sustainable agriculture and for result-based CAP, and for protection of the EU financial interests

Background and starting point

In today's world, there are many factors that are currently forcing all walks of life to look for ways to be more environmentally friendly, more efficient and more accurate. The same applies to agriculture and rural life as a whole, but also to the protection of the EU's financial interests.

Both the European Green Deal and related initiatives, policies and strategies like Farm to Fork or Biodiversity - that have pushed for it are forcing us to look for new modern opportunities for more natural and sustainable practices, while maintaining economic competitiveness and food security for the population. In addition to ensuring the transparency of the use of public money, one of the most important is its effective use and the smart protection of financial interests.

To achieve all this, digital solutions are seen as one of the great enablers. Smart and intelligent digital solutions are like the third or fourth major breakthrough in the world's life after the invention of the wheel, the industrial revolution or the invention of electricity. **However, digital solutions are based on data. Turning data into useful wisdom is a challenge for farmers who want to develop sustainable agriculture without losing their competitiveness, as well as for paying agencies that provide EU agricultural and rural support to protect the EU's financial interests.** For example, the EU 2019 anti-fraud strategy also highlights the importance of data-based risk analysis for the effective protection of the EU's financial interests.

It is important to understand that data-based risk management is also evolving. While the traditional approach to data analysis is based on risk criteria and rules based on human experience and knowledge, the new digital and data-driven society wants to replace the role of human-created rules in data analysis with more

powerful machine learning models that can find much more connections in data than human rules. Even EU supported ARACHNE risk scoring tool is widely based on a lot of rules-based risk analysis rather than modern machine learning analysis.

One of the general observations of this chapter may be - that many **exchanges of smart solutions and practices (including data analysis) may not be initiated on the basis of the OLAF / HERCULE program at all, but may be initiated by DG AGRI** or different other directorates in different formats involving also member states.

It is also certain that data scientists and analysts need to be more involved in policy-making as well as in the development of support systems and IT systems. It is likely that changes will soon be seen in the competencies of the paying agencies staff. The share of the process- and supervision-oriented workforce is likely to decline and be replaced by data analysts and data researchers.

The second-day presentations of Smart Pro CAP in Estonia looked at opportunities in which the digitalisation of agriculture, the development of carbon agriculture or a results-based EU support system or the growing awareness of data analysis and machine learning, could support both societal goals and the protection of EU financial interests.

As a result of the 2nd day of visit to Estonia and the Smart Pro CAP project generally, it can be concluded that data-driven rural life and agriculture is new normality as also data-driven protection of the EU financial interests and these cannot be seen apart from each other. Both spheres are needed to be wisely smartly integrated in order to get the maximum benefit from the data-driven world which smartly enables the prevention of harmful practices against EU financial interest.

Second-day presentations of the Smart Pro CAP Estonian visit were focused on the future and were found to be very inspirational and this subchapter

- Improves the awareness of PAs` staff on how the modern data-based farm management IT solutions with AI-supported data analysis can help farmers to build long-term partnerships with nature and raise competitiveness and promotes

the development of "carbon farming" as well increases the credibility of the market of the carbon credits. This approach promotes focusing on achieving results instead of getting only benefited from EU funds.

- Points out opportunities how to use data-driven machine learning methods by PA-s in different phases of the EU grant lifecycles. These provide a completely new era of introducing data-driven approach also for better protection of the EU financial interest for control sampling (detection) and as well for proactive activities (prevention).

- Shares 7 ideas or key points coming up from 2nd-day presentations and discussion how data-driven approach may be developed or supported on EU or Member states level and in the Paying agencies.

Our solutions

Data-driven future of farming, carbon farming and carbon credits

Supporting on MS and on EU level data-driven carbon farming and by creating trustful common and widely accepted carbon credit system which also urges farmers act more transparently and implement sustainable nature-friendly farming practices.

eAgronom which is a rapidly growing Estonian start-up providing innovative e-services and IT solutions for farmers shared its ideas about what they see to be important to keep in mind while supporting developing the digitalization of farming and supporting carbon farming in the framework of the CAP.

Summary thoughts about farm digitalisation by eAgronom

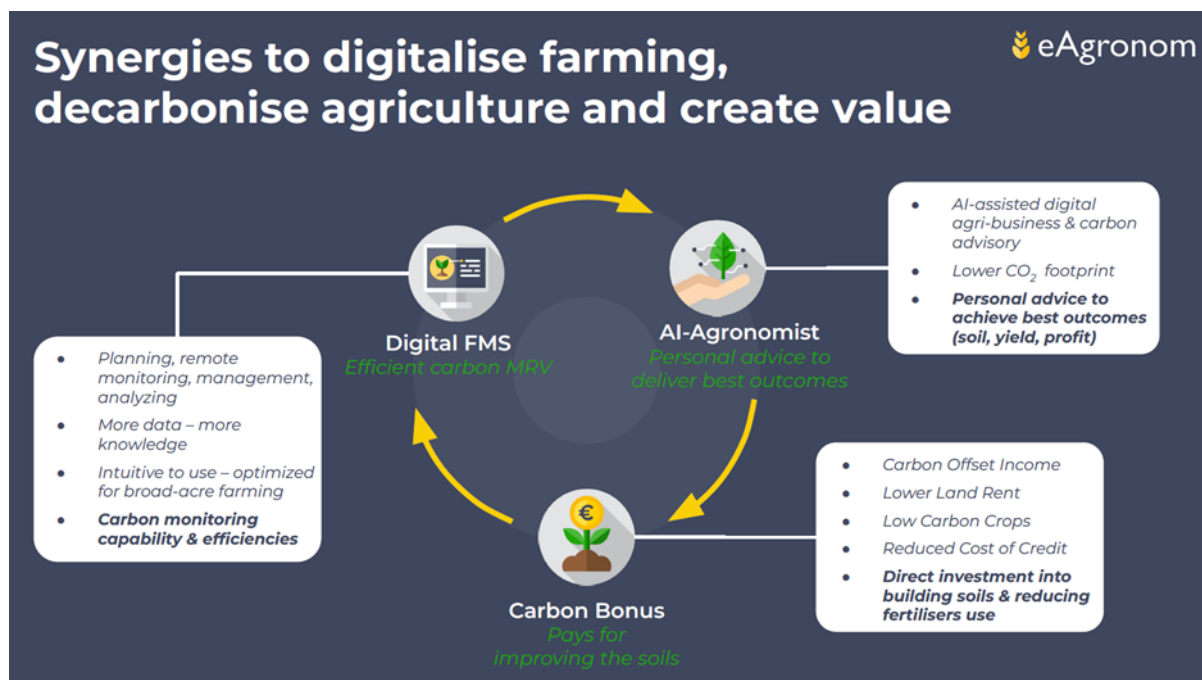
- Farmers need holistic digital tools. Building effective software solutions is expensive. Maintaining costs 4x more than building.
- Farmers only enter minimum data into government tools.
- Allow commercial farm management system (FMS) solutions to integrate and simplify all tracking of practices for government reporting.
- Standardise reporting needs across EU.

Summary thoughts - CAP and carbon farming

- Build the infrastructure for market-based carbon programs.
- Standards for project validation, methodologies, baseline calculation, and MRV (monitoring, reporting and verification).
- Support thorough soil sampling.
- Sync requirements with carbon farming MRV standards.
- Fund costs to support the transition from action-based to results-based carbon sequestration.
- Support capital investments into conservation agricultural equipment.
- Guarantee a floor price for certified credits.
- Direct payments to farms already practicing conservation agriculture.
- Advisory support to educate about regenerative farming.
- R&D in conservation agriculture.

Digitalization of farm management gives the opportunity to use data to enhance transparency farming practices and it has a positive environmental impact and its transparency mitigates risk for harmful practices against EU financial interest.

Financing regenerative farming practices through direct payments at the expense of EU taxpayers could discourage farmers from developing environmentally friendly practices and entering a market for carbon credit that requires a high level of transparency but could provide additional income financed by private companies that are not carbon neutral.



Machine learning as a new approach to data analysis. Four innovative approaches to data analysis for EU support schemes for CAP

In order to make grants of EU CAP more results-based and more secure against misuse, we will examine the possibilities of whether and how the new approaches of data analysis (machine learning) can be applied through different stages of the EU support lifecycle - from the design of the support schemes till the forming the control sample for inspections.

According to Wikipedia, machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

It is not rocket science it is enhanced use of mathematics with the help of machines.

Conventional data analysis examines data based on man-made rules - human expertise and experience. More modern data research is replacing man-made rules with machine learning. While the classical results of data research are limited to human experience and knowledge, there are no such limitations to human experience in the use of machine learning models, and unrestricted analysis of data using mathematical algorithms allows us to see relationships between data that cannot be achieved based on the human experience. The logical explanation of the relationships found as a result of machine learning can therefore often be difficult, but the results found on the basis of machine learning can be much more accurate than the results of classical rules-based data analysis.

Machine learning-based data analysis for designing interventions - as smart prevention of the EU financial interests

As first example, it can be established from the human rules that in order to achieve the target of job creation, it is reasonable to give preference to persons who have historically had more experience in job creation. It is quite easy to prove or disprove this assumption on the basis of historical information and statistics and correlations on job creation. In order to use machine learning models, a sufficient amount of historical data can be input to find the job creators, which does not have to contain any information about the jobs created. The aim is to allow machine learning models to find groups that predict, for example, a low or high job creation capacity based on unnoticed connections.

In ARIB the machine-learning data analysis was tested to support the design of a new CAP period for RD interventions where the target was job creation in rural areas at the determined salary level. For this, the similar grant scheme data of the 2014-2020 CAP period was analysed.

This pilot project was carried out in cooperation with the STATS Unities team, according to the SEMMA model - the sequential methods to build machine learning models incorporated in 'SAS Enterprise Miner', a product by SAS Institute Inc. SEMMA is an acronym that stands for **Sample, Explore, Modify, Model, and Assess**. SEMMA steps are explained according to Wikipedia as follows ¹

- **Sample.** The process starts with data sampling, e.g. selecting the data set for modelling. The data set should be large enough to contain sufficient information to retrieve, yet small enough to be used efficiently. This phase also deals with data partitioning.
- **Explore.** This phase covers the understanding of the data by discovering anticipated and unanticipated relationships between the variables, and also abnormalities, with the help of data visualization.
- **Modify.** The Modify phase contains methods to select, create and transform variables in preparation for data modelling.
- **Model.** In the Model phase, the focus is on applying various modelling (data mining) techniques on the prepared variables in order to create models that possibly provide the desired outcome.
- **Assess.** The last phase is Assess. The evaluation of the modelling results shows the reliability and usefulness of the created models

Following sets- data for analysis were used

- **External data for years 2015-2020**
 - **Data from Business register**
 - Query of simple company data => Output data:
 - Name of the company
 - Legal form of the company
 - Invalid business names of the company
 - Company registry code
 - Status of the entrepreneur in the commercial register
 - Registry area

¹ <https://en.wikipedia.org/wiki/SEMMA>

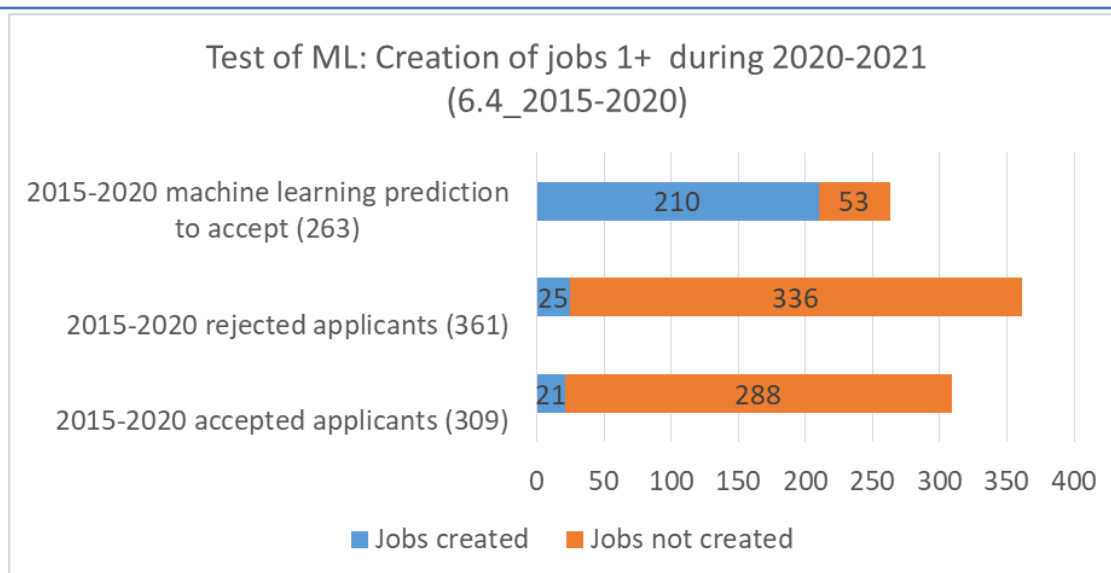
- Company capital
 - Business address
 - Date of first registration (i.e. date of first entry in the court register)
- Representation rights of all persons related to the undertaking
=> Output data:
 - Business name
 - Commercial register code
 - Status
 - Identity number
 - Country of identification code
 - Country of identification code in the text
 - Date of birth
 - First name
- Service of issuing breakdowns of sales revenue according to the annual report according to EMTAK (NACE code) => Output data:
 - Legal form
 - Reporting time (time accuracy)
 - Reporting year
 - EMTAK code (The Estonian Classification of Economic Activities (EMTAK) is the national version of the international harmonised NACE classification.)
 - Main activity True / False
 - Amount (in EUR)
 - Amount%
 - Amount coefficient
 - "Operating company but has not received sales revenue" (True / False)
- **Open Tax data**
 - Taxes paid, turnover and number of employees - on quarterly basis.

- **Internal data from ARIB** - all information available in the ARIB IT system regarding submitted applications of the sub-measure 6.4 - Investments in the development of non-agricultural activities 2015-2020.

TARGET determined as => Entrepreneurs who created at least +1 additional jobs to existing jobs between 2020-2021 with an average salary of 800 EUR per month.

Results demonstrated that the machine learning model predicted accuracy for the same target was 80% instead of the classic selection model accuracy of 53%.

| Accuracy comparison of models | Correct selection | Total | Accuracy |
|--|-------------------|-------|----------|
| Classic model selection accuracy (correctly accepted and rejected) | 357 | 670 | 53% |
| Machine learning predicted selection accuracy | 210 | 263 | 80% |



The use of machine learning presupposes the existence of a sufficient amount of historical data, and for a much larger group of persons than the selected

beneficiaries. Often, the lack of historical data is an obstacle to the use of machine learning models.

In human-controlled processes, such as the delivery of EU CAP grants, the challenge is whether or to what extent the results of machine learning models are trusted in decision-making.

However, the advantage of the machine learning model is the lack of human subjectivity, which makes it quite impossible to change the results in the desired direction by changing the criteria or artificially creating them.

For the new CAP period 2023-2027, a significant EU initiative on area payments is to move to a full area monitoring system (AMS) and the artificial intelligence based analysis and control system. **EU initiative in promoting AI enhanced AMS will encourage the use of machine learning models in other areas and phases of EU CAP support.**

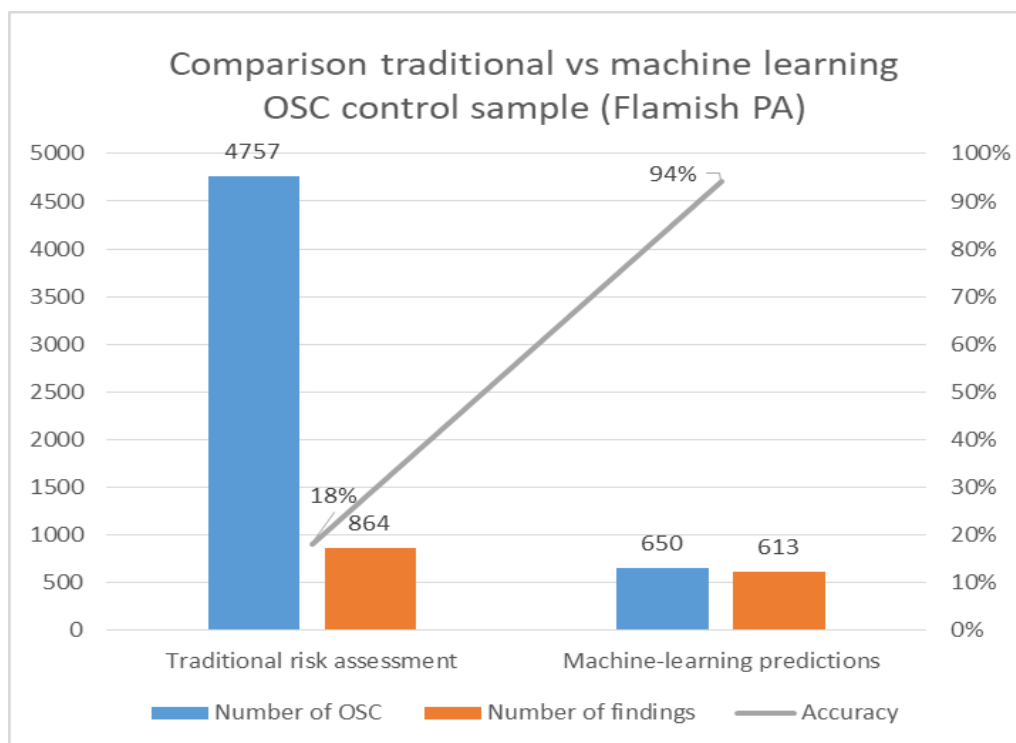
Machine -learning based data analysis for increasing accuracy of OTC control sample - as smart detection of the EU financial interests

Another example is the Flemish Paying Agency's attempt to improve the accuracy of risk analysis using machine learning models. On-the-spot sampling of the population is usually based on risk criteria established on the basis of human expertise, which is the usual approach to data analysis. **In the modern approach, instead of man-made risk rules, potential control objects are allowed to be found by machine learning models based on historical data.**

The Flemish PA test showed that, based on historical data, the accuracy of the classical control sample was 18% and the accuracy of the control samples predicted by the machine learning model was 94%. **This would allow with fewer controls and higher accuracy to achieve the result which is similar to the classical approach.** In numbers, this means that

- on the basis of classical risks control sample on basis of classical risk analysis **4757 objects** were controlled and 3893 inspections were completed without findings, and 864 inspections with the findings.
- With the machine learning model, the number of applications to be inspected **would have been 650 instead of 4757**, of which 94% or 613 would have resulted in a finding and 37 without an invention.
- Decrease of the number of OSC-s by 4107 pcs (4757-650) but increasing of accuracy from 18% up to 94% would still have secured a deterrent impact of the inspections, because 613 findings instead of 864 would have been found with the help of the new machine learning approach at the same time with the robust declining of the administrative burden for clients and for PA. This example demonstrates that the use of machine learning models can help to make control samples more accurate and thus reduce the control burden at the expense of those who do not need control.

| Traditional vs machine learning risk assessment (Flamish PA mini POC) | Number of OSC | Number of findings | Accuracy |
|---|---------------|--------------------|----------|
| Traditional risk assessment | 4757 | 864 | 18% |
| Machine-learning predictions | 650 | 613 | 94% |



In the previous two examples, we have looked at examples of the use of machine learning models in the prevention development phase as well as in the detection of controls.

Machine-learning model for application selection process - as smart prevention of the EU financial interests

The third idea that could be considered in the future is using a machine learning model as part of the application selection process, similarly as it is widely known in the private credit institutions, where AI-based credit ratings are widely used and also the final decisions for smaller loans are trusted to be made by AI.

In the application selection process, the use of risk analysis based on machine learning models can be seen as a useful automated process for pre-or post-evaluation of applications, where machine learning models would be trained to find grant applications with a higher probability of achieving the desired results. Such

an approach would also be directly in line with the new EU 2023-2027 performance-based CAP principle.

As AI-powered systems can appraise customer credit histories more accurately to avoid losses for banks, or create tailored loans and provide scoring free of human biases - the same ideology can be introduced for developing the AI-based automated evaluation and selection processes of grant applications.

Use machine learning models for new proactive targeted support campaigns - as smart prevention of the EU financial interests

Testing of machine learning models has shown that the effective use of these results can also mean completely new approaches to grant-making processes. In a situation where machine learning models have a high ability to predict the target group of people who are more likely to achieve results, support campaigns targeted at a higher potential target group could be considered. This would have a positive effect on the more prudent use of CAP funds. It would also have a deterrent effect, as it restricts access to "fishing" the grants by those who may want nothing more than financial support at the expense of EU funding.

With the help of machine learning models, it is possible to target grants more precisely and more consciously to the target groups for whom they are intended, instead of passively waiting for suitable target groups to find the grants intended for them. **Through a proactive approach, such as the active provision of grants, there is a growing likelihood that grants will reach a more targeted, credible and effective target group, leaving fewer opportunities for those whose intentions may not be honest.**

Agricultural big data project in ARIB

According to our new strategy, ARIB has taken the approach of developing data-driven knowledge services for all clients or other actors who have any relations with living or making business in rural areas or in agriculture.

Basis for this approach are very good IT framework, integration all available registers under ministry of Rural affairs with other state registers and very good competence of analysing and visualizing of data.

According to our new strategy, ARIB has taken the approach of developing data-driven knowledge services for all clients or other actors who have any relations with living or making business in rural areas or in agriculture.

The basis for this approach is a very good IT framework, integration of all available registers under the ministry of Rural affairs and with other state registers, and very good competence in analysing and visualizing data. Of course, information from various sources presupposes a very good system of data interfaces, which ARIB is also developing within the framework of the agricultural big data project.

We believe that the agricultural big data project will be the basis for developing modern data-driven knowledge services, raising awareness of our client's investment needs and their financing options. One public service we wish to develop on basis of knowledge services is 3 in 1 financing service.

Once the client, with the support of AKIS and the advisory system, has identified their investment needs and the specific goals it wants to achieve with this, one could get an answer on the financing of a complete project from external sources by preparing and submitting a single investment financing application. The 3 in 1 financing service could cover co-financing through both CAP grants and a financial instrument and, if necessary, additional loans or guarantees from other credit institutions. The institution dealing with the financial instrument, as well as various credit institutions, have shown interest in the development of such a service.

In conclusion, we found that

- Data-based knowledge services help develop smart support schemes and interventions.
- Services that are based on big data project is supporting digitalization of rural life and agriculture and achieving the goals of the EU Green Deal on one hand and also help to meet the changing requirements more easily and therefore prevent the need to commit irregularities or fraud.
- Smart support schemes and interventions are one step ahead of preventing fraud.

Sharing best practices by the initiative of AC DG AGRI according to HzR art 13.

The Smart Pro CAP project has repeatedly confirmed that the protection of the EU's smart financial interests may not be directly linked to activities to detect, investigate or sanction irregularities or fraud. Smart protection of financial interests often starts even earlier than risk mapping. Smart protection of financial interests starts with a smart design of support schemes and processes, based on objective data, simplicity and transparency.

For example, all paying agencies involved in the project agreed that the use of the standard 3 offers rule in rural investment measures is always problematic and a much smarter way is to introduce a much wider usage of simplified costing methods that supports also much more the concept of performance-based EU supports.

Therefore, the paying agencies of the participating countries, as well as the representative of DG AGRI, considered that the sharing of best practices in the fight against irregularities and fraud should not always take place at the initiative of OLAF or under the HERCULE program. It is probably smart for exchanges of experience and smart practices in different formats to take place on DG AGRI's initiative, for example in the Learning Network or similar format. Due to the number of participants

and the in-depth treatment of the topic, the participants also considered the cooperation format of this project to be very effective and valuable.

The last-mentioned initiative is supported also by EU regulation 2116/2021 where article 13 states that the Commission shall promote the exchange of best practices between the Member States, in particular as regards the work of the governance bodies under this Chapter.

Lessons learned

To summarize the visit of the Smart Pro CAP project to Estonia, at least 4 initial ideas can be presented on how data analysis based on machine learning in different stages of rural development support can contribute to better achievement of support objectives and better prevent and detect activities detrimental to EU financial interests.

- In data analysis, including the use of machine learning models in the design of interventions to define the target group
- The possibility of new proactive targeted support campaigns to ensure better effectiveness of support and to reduce "fishing" on support. Preventive approach
- Implementation of machine learning models in the pre- or post-evaluation phase of the grant application selection process to ensure better performance of funded projects and the independence of the procurement process from human subjectivity.
- The use of a machine learning model for more accurate control sampling, which would reduce the volume of controls and thus the administrative burden for all parties. At the same time, well-targeted controls would ensure the deterrent effect of controls and the protection of the EU's financial interests.

Additional lessons learned from the 2nd day presentations of Staff Pro CAP visit to Estonia.

- Smart support of development transparent and data driven carbon farming and develop EU standard for carbon credit system and standardise reporting systems across EU.
- Develop data availability (for example, big data projects in Estonia and Belgium). Precondition to develop a uniform legal framework or guidelines for data protection that would allow and support the wider use of machine learning models.
- EC initiative to enhance better cooperation between MS in different formats to find and share best practices also in using or piloting modern machine learning methods to rise effectiveness of processes of EU RD support schemes. General acceptance of using machine learning methods by EC.

PART II Experiences of using ARACHNE risk scoring tool

- *Maps the challenges of using ARACHNE for its` risk scoring functionality of rural development support and for individual inquiries for detecting external or internal conflict of interests.*
- *Shares ARACHNE tested solutions (tips) for the compilation and interpretability of the data to be uploaded in order to ensure the maximum usefulness and usability of the data to be uploaded and introduces the most useful functions of ARACHNE according to practice.*
- *Gives ideas of using ARACHNE as the tool in complementing the paying agency's risk management system to a suitable extent and options to save PA`s analytical resources in the future, unless ARACHNE's risk scoring functionality is modernized and adapted to the CAP needs.*

The problem

The European Commission has developed the ARCHNE programme for fraud prevention and risk assessment to protect financial interests. The programme is already available to Member States under the current programming period. The European Commission plans to make it compulsory for agricultural funds in the next programming period. So far, we have set up our risk management system according to our own knowledge and the means at our disposal. The system developed by the Commission could provide an opportunity to further develop and only to complement the national risk management system rather than a substitute to better protect financial interests.

The problems that ARIB found needed to be addressed while starting piloting the use of ARACHNE for the CAP were whether and to what extent the data required by ARACHNE matched the data collected during the CAP procedures and whether they were also automatically inherited from the ARIB's procedural systems.

Secondly, it was also necessary to find out how in practice ARACHNE supports the processing of CAP rural support and the protection of the EU's financial interests.

Our starting point

Preliminary work for the use of ARACHNE started already in 2019 with the establishment of contacts, the first instructions for the use of ARACHNE and the mapping of data.

ARIB has seen the use of ARACHNE mainly in 3 functionalities:

- worldwide individual inquiries concerning the business connections and commercial background information of the persons involved in the support applications (manual searches on persons entities /groups),
- an automatic risk scoring tool of applications, applicants and involved businesses,
- case management.

We set ourselves the following goals to pilot ARACHNE. The first steps were:

- to get acquainted with the structure of data necessary for ARACHNE and their availability from ARIB systems and to compile and upload the first test data to ARACHNE,
- to get feedback on the quality of the first test data and repetition of the data request regarding all possible ARIB data and the question of how to automate the collection of these data as much as possible,
- to extend the number of ARACHNE users to all officials involved in the application process who wish to receive and receive feedback on which functionalities are most useful to users,
- to test whether and to what extent the risks of ARACHNE overlap or harmonize with the risks calculated in the ARIB's own risk management system.

According to DG AGRI ARACHNE expert group meeting on 20th of Dec 2020 European Commission found that ARACHNE can be used for the CAP.

Generally to identify:

- for the most risky projects/contracts through risk scoring;

- potential links between stakeholders;
- potential conflicts of interest.

More specifically for:

- **Project selection** - where Risk calculations help identifying the most risky projects/contracts
- **Administrative checks on eligibility conditions** -
 - through single inquiry functionality helping identify company status (e.g. number of employees, operating revenue , NACE classification, viability based on ORBIS database information) and
 - reports on related companies, subsidiaries, shareholders etc. (e.g. SME condition).
 - Also possible double-funding risk can be checked if Arachne includes data of all possible EU fundings of beneficiary from different EU funds.
- **Administrative checks on the reasonableness of costs**- as Arachne helps identifying concentration of projects at contractors' level.
- **Checking on potential conflicts of interests** - as Arachne helps identifying links between beneficiaries and contractors or other stakeholders
- **Fraud prevention** - while World Compliance data helps identify red flags of persons political involvements or history of criminal involvements etc.
- **Ex post checks** - as ORBIS data helps detect risks as to the 5 year durability clause for investments (e.g. company status and viability after project

Estonia started testing ARACHNE in March 2020 with the first data uploads for two measures. The ARACHNE testing and deployment team will initially consist of 4 people. An anti-fraud advisor and 3 risk analysts. With the loading of the first data, we also received feedback on data quality to address further. During the year, a number of data loads were carried out and efforts were made to find automated solutions for data aggregation. In April, users from different units (from all implementing and investigating units and also analysis department) were added to test ARACHNE. In December 2020, ARACHNE introduced in procedures of measure schemes 6.4, 4.1, 4.2.1. Focus was on background and connections information especially regarding foreign connections. In 2021 we continued testing ARACHNE

risks vs ARIB risks. We asked for a Web-based ARACHNE training for all users in ARIB and it was organised by Wim Smets (EC) in early 2021.

By the end of 2021, we will have made a total of 5 data uploads and will have uploaded all CAP and EMFF projects. In total 7116 applicants for 11303 projects.

The most difficult part of the data loading was the structural matching of the data with the structure provided by Arachne.

Out of the 17 data sheets provided by ARACHNE, we have only been able to complete 7. There was also a need to find a way to merge data in predefined fields.

For example in project level:

- Project ID – application file ID
- Project name – application file ID + application round ID
- Beneficiary ID – beneficiary's registration number
- Turnover, Number of employees – filled if the data is available (depends on measure)
- Start date – approval date (signed)
- End date – durability period end date

Several fields we do not have the data to include (for example Income and Thematic Objective).

In contract level we found the following solution:

- Contract ID – object ID
- Contractor ID – won contractors' register number
- Contract name=Contract description – object description
- Amount – object's cost
- Procurement Valid tenders – nr of contractors per object

There were even more fields at this level for which we have no data.

For example:

- Contract Type
- Contract modified
- Signature date

- Initial end date
- Final end date
- Addenda Amount
- Number of addenda
- Procurement Type
- Exclude tenders
- Total tenders
- Publication date

The biggest problem with the data was with the entities where we don't have data about all contractors. In the case of related persons, information is available on the representatives but not on the board members of the beneficiary.

A big issue is the interpretation of the data - what field do we have that could fit the given criteria. The data will also vary depending on the different measures and what data are collected under specific measures. And data quality is always a problem too.

Our solutions

Data interpretations

The challenge is that ARACHNE's data needs follow the natural data structure and logic of the public procurement process for EU SF funds, but for rural development support most projects are not related to the public procurement process but the conventional 3 tendering system. Through a system of price catalogues for machines that are reasonably guaranteed.

In Estonian conditions, we also foresaw the problem that while ARACHNE collects information on already funded projects that have passed the selection procedure, the ARIB's own risk management system for prevention assesses the risks of applications before and during the selection procedure.

If you want to use the ARACHNE risk scoring functionality to support the selection procedure as well, you can compare ARACHNE's risks with the ARIB's own

calculated risks. For this purpose, we also had to add to the ARACHNE database the data of the applications that were submitted to the ARIB but which did not reach the financing decision, i.e. were rejected. In order to differentiate between rejected applications, we also marked these applications in the ARACHNE datasets as completed projects in subsequent data uploads.

Implementing ARACHNE functionalities

- **Manual single searches on persons/entities /groups and related business data and connections.**

We found that ARACHNE's most valuable functionality at the moment is to have the ability to monitor by single case-based inquiries global-wide business connections and relevant business data in order to support administrative checks. Especially valuable and quick in the cases that have foreign contractors, stakeholders and persons are involved.

Therefore ARACHNE was officially introduced as an additional source of information introduced to support administrative checks in the working procedures of measure schemes 6.4, 4.1, 4.2.1.

- **Searching conflict of interests by Internal Audit**

As another option, we have tested ARACHNE in the context of internal conflict of interest prevention in the internal audit department. We checked the connections of the employees with the applications during several audits. ARACHNE works for such a check but this is one by one checking. We also tested ARACHNE for reporting ancillary activities to our employees.

Problem we found was that through the ARACHNE, you can only see that you are on the board, but for example, other contractual employment relationships are not reflected there.

- **Testing ARACHNE vs ARIB risk scoring**

We tested the ARACHNE risk assessment on the basis of 2 measures. The aim of action 6.1 was to see if the risks of ARACHNE and our risks are different. 9 projects were examined. The main differences, or two major differences, are as follows. First of all: for 6 projects, no risks were identified at all or risks were low according to

ARACHNE. In our data, the projects had a high risk score. As reasons for the difference, we identified that ARACHNE does not have enough scored risk factors using in-house data (beneficiaries' board members network, cadastral parcels cumulation etc.). And the second reason was that there is no historical data about new enterprises in ARACHNE. The second one was just the opposite. Low risk score from our scoring and high risk in ARACHNE with other 3 projects.

Results of measure 4.2.1 examination. 38 of 102 risks were scored in ARACHNE and we identified 8 risks that were similar to our risk list. As an example of the reasons why only 8 risks were added up, it can be pointed out that the risks of the paying agency are often more detailed or vice versa. There were situations where one ARACHNE risk was matched by our 3 risks or vice versa.

After some testing, we created a preliminary SWOT using ARACHNE where we were able to record all the strengths, weaknesses, opportunities and threats.

| | |
|--|--|
| STRENGTHS <ul style="list-style-type: none"> ✓ Added value (new datasets, more data) ✓ Automatic risk scoring ✓ Internationally approved risk scoring tool ✓ Data is visualised | WEAKNESSES <ul style="list-style-type: none"> ☒ Have to update and upload source data manually (no linkage) ☒ Not every risk factor has defined data origin ☒ Not very user friendly ☒ Not enough experience in house ☒ Not enough flexibility with data sources (needs some extra data) ☒ Built up for public procurement ☒ Not for simplified cost method ☒ Not all internal data is included |
| OPPORTUNITIES <ul style="list-style-type: none"> ✓ Detailed client view ✓ International relationships | THREATS <ul style="list-style-type: none"> ☒ Not trustworthy, because not widely used ☒ Threats with interfacing |

The SWOT analysis should certainly be updated as new information becomes available following testing and ARACHNE developments.

Lessons learnt

DG AGRI already mentioned at the ARACHNE expert group meeting in December of 2020, there are several challenges to the full implementation of ARACHNE for the CAP:

- Data availability and data collection process from PA databases (e.g. data on contractors)
- Data availability in company registers (depending on national rules on financial data publication)
- Small project size in the agricultural sector
- Interpretation of scores

Overall, their conclusion can be accepted. However, based on practical experience, they can be specified. For instance from the practical point of view we can point out that:

- ARACHNE input data and risk calculation logic were not originally developed for CAP rural support and do not support the application evaluation and selection process.
- There is a complete lack of input and output information on CAP direct payments
- ORBIS data update interval after 3 months is too slow. For example, for certain support schemes, it is also possible to apply to a company set up less than 3 months ago
- ARACHNE cannot identify applications related to the same persons in the same or different grant rounds or grant schemes, although the identification of the link between the beneficiary and the supplier works
- There are many rules-based risks and they are too broad, leading to a very large number of false-positive matches.
- Many risks remain uncalculated as far as rural support is concerned because of lacking data in CAP field or risks are too much focused on Public Procurement specific risks.
- Missing flexibility to adjust risks more suitable for the paying agency needs or to create or test new risks.

ARACHNE's strength can certainly be considered to be part of individual queries, although here too, some data quality issues may prevent you from finding the right companies due to the different spellings in different languages. There may also be situations where ARACHNE is unable to identify all of a person's business

relationships because the same person may appear in ARACHNE's data in a number of different forms with partially incomplete data.

Overall, ARACHNE has a strong strength in the global capabilities of all related EU grants and interpersonal links and primary business data.

The latest developments in ARACHNE known to the ARIB, which partially solve the above concerns, can also be pointed out as positive.

As we know in January 2022 - expected developments, modules, functions of ARACHNE are following.

- Data updating of ORBIS data => 1 month instead of 3 months
- Tuning ARACHNE more suitable for CAP needs including EAGF (direct payment schemes) => DG AGRI ARACHNE development team is in place
- New module to support selection process (ex-ante module) ready to be launched => 1Q 2022
- Final beneficiary detection function (up to the individuals) => 4Q 2022 => Supports HzR (2021/2116) art 59 (4) => Member States shall ensure that beneficiaries of the EAGF and EAFRD provide them with the information necessary for their identification, including, where applicable, the identification of the group in which they participate, as defined in Article 2, point (11), of Directive 2013/34/EU of the European Parliament and of the Council.

As explained by DG AGRI's representative at the Smart Pro CAP seminar in Tartu, the claim of the beneficial owner and the functionality of ARACHNE are aimed primarily at increasing the transparency of EU grants in the eyes of society and do not currently have a restrictive or favourable legal effect.

The long-term goal of the risk management of ARIB grants is to rely on the risk scores calculated by ARACHNE and to integrate and interface them with the IT systems of the ARIB procedure.

This presupposes in the future,

- The suitability of ARACHNE risks with the risks in the ARIB risk register.

- The ARACHNE risk calculation system should become more flexible as risks are constantly evolving and make more use of machine learning and artificial intelligence.
- Possibility of ARACHNE calculated risk score data exchange machine interfaces with ARIB IT systems.
- ARIB's ability to automatically keep updated support data in ARACHNE.

As a result, we see the potential to reduce the time spent by data analysts on repetitive risk assessments and the detection of 'risky' grant applications once reviewing process of applications has started. Instead, we could use the competence and skills of data analysts to perform more specific analyses in the earlier stages of the support lifecycle for prevention.

We have also learnt that it is necessary to further test the use and further development of ARACHNE with different measures and we will do this next with measure 6.4. Through testing, we can provide feedback for further developments that are more suited to our needs. Using ARACHNE for searching networks case-by-case and single search is very good and useful. As of today, ARACHNE is not suitable for risk assessment. Based on our experience, we can see that ARACHNE could learn more from existing information and use machine learning techniques to display new information. ARACHNE could also be more flexible to meet national needs. If Member States were able to set their own risks according to the specificities of their country and measures, ARACHNE would offer greater possibilities for its use.

Conclusions

A general idea of the visit to Estonia was to share experiences of using preventively the data-driven risk analyses experience for non-IACS EAFRD measures and piloting the use of ARACHNE for daily processes. Secondly also a look into the future of data-driven agriculture and the DG AGRI expectations for the next CAP period regarding the protection of the EU financial interests.

Day 1 was dedicated to covering Estonian paying agency experiences of using data-driven risk analysis system and especially developing risk-based administrative checks (red-green corridor approach since 2016). Statistics show that the rejection and withdrawal rate of the EU RD support applications in the red corridor is higher than in the lower-risk (green corridor). This all brings positive and measurable fraud and irregular preventive effects, although this cannot be reported or evaluated on the EU level due to the absence of such reporting values.

Latest IT developments help to integrate online risk-based administrative checks and risk feedback as part of the daily proceeding processes all over the EU support lifecycle.

Also, Estonian PA experience of using the ARACHNE risk scoring tool was covered, and what included small tests of comparing ARACHNE risk factors with agencies own risk factors.

Day 2 was an inspirational day watching to the future of data-driven agriculture, looking for options for implementing a data-driven approach enhanced with machine learning methods and to processes of designing CAP support schemes and in different phases of proceedings. It came out that new data-driven approaches may encourage beneficiaries and farmers to share more data and be transparent and honest if provided data becomes useful for them. For instance, helping them be more competitive in participating in a beneficial carbon market or if paying agencies can create data-based proactive actions-processes to avoid possible troubles in the future. Supporting instead of punishing. On the second day also DG AGRI anti-fraud correspondent shared DG AGRI views and expectations of designing the new anti-fraud framework in the context of the next CAP period 2023-2027. On 1st and

2nd day also representatives from Latvian and Lithuanian paying agencies were involved.

To generalize the main findings of Estonian staff exchange visit.

Maximum benefit of modern data-driven (incl machine learning methods) can be achieved when the whole organization has data-driven culture, framework, and competence. A data-driven approach needs much more data than only IMS reports. It must be introduced to the whole process from designing interventions to detecting irregularities.

The smart way is to use a data-driven approach to lead and keep beneficiaries on the green corridor (fast track) - also to create data-driven proactive support actions rather than detection of fraud actions.

The ideas of machine-learning-analysis-based targeted campaigning of EU funding or other options of proactive approaches and processes will be the new normality. As shown in the next diagram, more efforts must be made to maximize the use of machine-learning analysis opportunities to protect EU financial interest in a preventional manner.

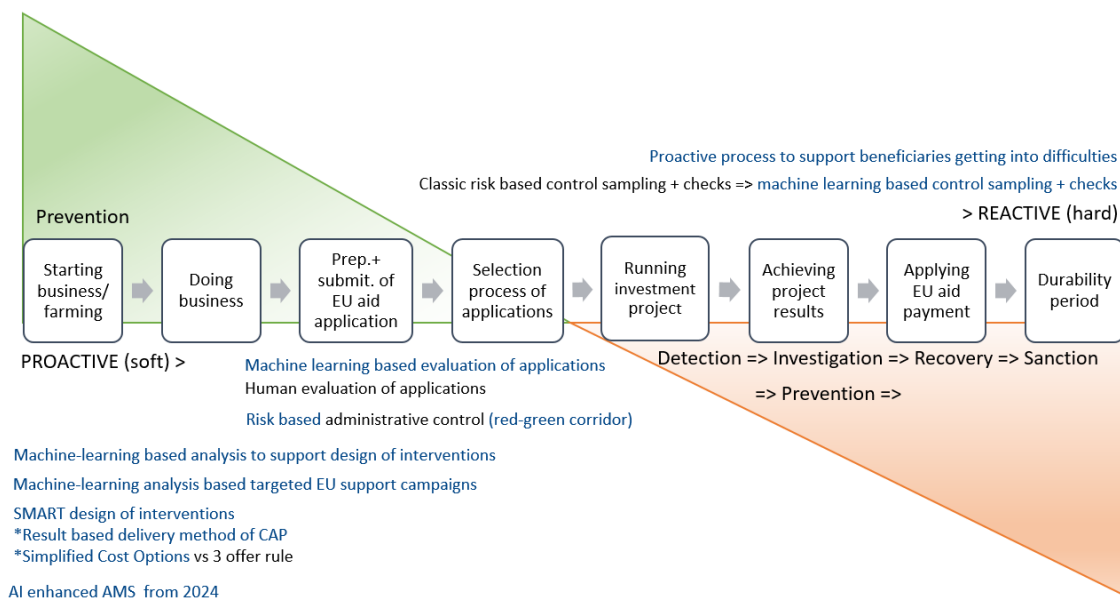


Mapping classic and modern anti-fraud actions for CAP measure schemes 2022

Classic and **modern** anti-fraud actions?

*New Smart options

*Classic options



Prevention starts much sooner than receiving grant applications, it starts from smart designing, targeting, and framing of the processes of the EU grants system. A new result-based approach as the New Delivery Model for CAP 2023-2027 states, is clearly supporting wider use of the modern data-driven approach.

Such staff exchange visits format on certain focused topics in a series of meetings is a very effective tool for possible solutions on chosen thematic topics. We also experienced that involving a representative from DG AGRI from EC was really useful to create a trustful and collaborative atmosphere between EC and MS.