Euroopa Maaelu Arengu Põllumajandusfond:
Euroopa investeeringud maapiirkondadesse

"Long-term programme of knowledge transfer in the area of agricultural big data"

(procurement contract no. 194144)

final report

Persons responsible:

Agricultural Research Centre
Estonian University of Life Sciences
Estonian Crop Research Institute
Tieto Estonia AS
E-Agronom OÜ

August 2019

Contents

Introduction

In 2017, the Ministry of Rural Affairs initiated long-term knowledge transfer programme in the field of agricultural big data (link to the directive).

The aim of the programme is to create economic value added for farming by increased interoperability and efficient use of data, to support the development of precision farming and environmentally friendly agriculture. In order to reach the aim, electronical agricultural Big Data system (tool) will be developed by the end of the programme, which allows relating existing data, linking data to relevant analytical models and practical solutions. The data in the system must be harmonised, compatible, updated, linked or linkable to spatial data, allow two-way traffic (i.e. from farmer to the system and back to the farmer) and be usable in various practical farmer-oriented models/applications.

In the first stage of the programme, ARIB organised public procurement in the first half of 2018 in order to perform **feasibility study** of agricultural Big Data system. The purpose of the study was to ascertain whether it would be possible to develop agricultural Big Data system in legal, infotechnological, economic terms and based on the principle of practicability.

The contents of the programme include the following areas:

1) Nutrients balance, humus balance and fertilisation;
2) Integrated plant protection;
3) Crop rotation;
4) Precision farming.

After winning the procurement, consortium was awarded public contract in autumn 2018, term for completing the feasibility study is 9 September 2019.

Leading partner of the consortium is Agricultural Research Centre and partners are Tieto Estonia AS, E-Agronom OÜ, Estonian University of Life Sciences and Estonian Crop Research Institute. Project is managed by a council, members of which represent various agencies related to the topic (see Chapter 1).

The analyses performed contains infotechnological, scientific, legal and economic view of agriculture-related data and possible solutions for using them in the Big Data system. It defines services that can be realised based on the analysed data (state, software providers). Outcome of Stage I of the programme will serve as a basis for calling Stage II procurement for realising the technical solution of big data system and more significant or "practicable" services. When creating services, it is important to observe which services are intended to be realised under FaST and NIVA projects initiated by the European Commission, to achieve additional synergy and avoid duplication.

Agricultural big data project is not just an IT project as it may seem at first glance. The project aims at making data usable and analyses accessible for farming purposes. Other beneficiaries besides farmers include public authorities and researchers, also third parties such as agricultural software providers. On global scale, big data is already used for the benefit of agriculture to some extent, but in Estonia it is still mostly seen as something to be used in future. Meanwhile, Estonia has great prerequisites for that as we have good public infrastructure for data exchange (X-Road), there are many databases that contain data useful for farmers. The society has good overall electronic capacity.

Another significant result of the analysis performed consisted in suggestions for persons who maintain databases in order to increase data quality. These suggestions are split in terms of improvement and development needs: those necessary for joining big data system and those necessary even without joining it. Individual analysis reports of particular databases are not publicly available, because they contain confidential information, and thus only conclusions and recommendations are made public.

Stage I of the programme lasted for 11 months, during which a feasibility study in the area of using agricultural big data was performed. There was no software created in Stage I, it focused on creating prerequisites instead. The feasibility study mainly aimed at ascertaining the potential solution for electronic agricultural big data system.

Within the framework of the study, 41 thematic databases of the Ministry of Rural Affairs, the Ministry of the Environment, the Ministry of Economic Affairs and Communications, the Ministry of Finance and Estonian University of Life Sciences were analysed.

Feasibility study was compiled by involving parties related to creating, managing and using data. The results of the feasibility study were introduced to farmers through various presentation or information activities. Analysis contains infotechnological, scientific, legal and economic view of data in order to explain whether and what can be used.

In order to involve stakeholders and present them the results, three information days were organised, including one information day for introducing programme activities to agricultural advisers, and a conference on agricultural big data (took place on 2.07.2019 at EULS). Conference presentations are enclosed in Annex 7.

Stakeholders and potential beneficiaries include agricultural producers, public sector institutions, universities and other research institutions, and third parties (primarily agricultural software manufacturers).

1. Programme management and distribution of work between the persons responsible

1.1 Persons responsible for the programme

To implement the activities of Stage I, a consortium was established, consisting of Agricultural Research Centre (hereinafter ARC; leading partner), Estonian University of Life Sciences (hereinafter EULS; research partner), Estonian Crop Research Institute (hereinafter ECRI; research partner), Tieto Estonia AS (hereinafter Tieto; IT partner) and E-Agronom OÜ (hereinafter EA; electronic field data book, agricultural software). Consortium and contracting authority entered into public contract on 10 September 2018. Thus, in view of the volume of work, the time limit was rather critical.

Joint tenderers in the procurement distributed their work so that each one was responsible for planning and proper performance of their activities, whereas, due to interconnected nature of these activities, they were planned and coordinated jointly. Project managers held discussions and weekly meetings. There were regular discussions held by the representatives of the Ministry of Rural Affairs and direct contracting authority ARIB.

The duties of ARC as a leading partner, included representation and management of the consortium and coordinating its activities; monitoring of the performance of works and compliance with contractual objectives; ensuring that necessary information reaches the members of the consortium; reporting to the contracting authority or third parties who needed the information regarding consortium activities; communication with programme council and contracting authority; ensuring timely provision of proper reports and related documentation and providing additional information, supplementations and explanations to the council and contracting authority; organising the work of programme council and drawing the minutes of the meetings.

Duties of the work groups of the programme:

- 1.1. General programme coordination was performed by ARC.
- 1.2. Work group coordinators coordinated the work of relevant groups:
  - 1.2.1. Development of the standard for electronic field data book, coordinated by EA.
  - 1.2.2. Definition of metrics and dimensions; coordinated by ECRI and EULS.
  - 1.2.3. Analysis of databases; coordinated by Tieto.
  - 1.2.4. Development of architectural versions; coordinated by Tieto.
  - 1.2.5. Legal analysis; coordinated by ARC.
  - 1.2.6. Economic analysis; coordinated by Tieto and ARC.
  - 1.2.7. Research group; coordinated by ECRI and EULS.
  - 1.2.8. Information and communication work group; coordinated by ARC.
- 1.3. IT work group:
  - 1.3.1. Project manager – product owner (Tieto):
    - 1.3.1.1. Coordination of the work of IT work group.
    - 1.3.1.2. Management, incl. prioritising the requirements for the solution to be created.
    - 1.3.1.3. Directing works according to requirements and priorities.
    - 1.3.1.4. Economic analysis of the solution to be created.
    - 1.3.1.5. Organising legal analysis of data use (ARC).

- 1.3.2. System analyst (Tieto):
  - 1.3.2.1. Development of metrics and dimensions in cooperation with research group.
  - 1.3.2.2. Analysis of database definitions, searching for data sources based on the needs of particular dimension or metric.
  - 1.3.2.3. Analysis of requirements for big data central system in cooperation with the representatives of stakeholders.
  - 1.3.2.4. Definition of system functionality, definition of services.
  - 1.3.2.5. Description of data structures and standards.
- 1.3.3. Data analyst, big data specialist (Tieto):
  - 1.3.3.1. Visualisation and analytics of database data.
  - 1.3.3.2. Data profiling, including development of quality assessment based on analytics results.
  - 1.3.3.3. Technical work related to data extracts from databases.
- 1.3.4. IT-architect (Tieto).
- 1.3.5. Agricultural software development specialist (EA):
  - 1.3.5.1. Creating visual prototypes of services in eAgronom software.
  - 1.3.5.2. Providing input for system analysts in terms of data required by farmers according to EA experience.
- 1.3.6. Services developer (EA)
  - 1.3.6.1. Creating visual prototypes of services.

1.4. Research group (EULS, ECRI, ARC)

- 1.4.1. Experts, members of work groups to the extent of their competency:
  - 1.4.1.1. Compiled descriptions of metrics and dimensions that served as input for system analysis.
  - 1.4.1.2. Described the relations between metrics.
  - 1.4.1.3. Described and systemised classification data that facilitated interpretation of data in databases.
  - 1.4.1.4. Provided overviews of scientific works and voiced opinions regarding potential future visions, how the data should be used.
  - 1.4.1.5. Assessed data analytics visualisations prepared for databases.
  - 1.4.1.6. Assessed completed system analyses.
  - 1.4.1.7. Compiled reviews of possible data sources.
  - 1.4.1.8. Advised the project team in other questions arising with regard to agricultural sector.

1.5. Information and communication work group, coordinated by ARC – operated in close cooperation with other work groups

- 1.5.1. Updating and dissemination of programme-related information.
- 1.5.2. Preparing the materials for information days and conference.
- 1.5.3. Finding competent lectors.
- 1.5.4. Providing information about information days and conference as required.
- 1.5.5. Collection and analysis of feedback to events, information days, conference.
- 1.5.6. Compilation of programme reports.

Programme core team:

1. <u>General coordination</u>

Urmas Visse (ARC) – programme manager
Maris Kruuse (ARC) – programme coordinator

2. Legal analysis

Jaanus Põldmaa (ARC) – programme lawyer

3. Economic analysis

Martin Paukson (Tieto), Mati Mõtte (ARC)

4. IT work group

Martin Paukson (Tieto) – leader of work group
Andres Lille (Tieto) – chief analyst
Kati Rohtla (Tieto) – analyst
Targo Tennisberg (EA) – project manager (electronic field data book)

5. Research work group

Alar Astover (EULS), Mati Koppel (ECRI) – work group leader
1) nutrient balance, humus balance and fertilisation: Alar Astover, Liia Kukk, Karin Kauer, Henn Raave (EULS), Priit Penu, Livi Rooma, Jaan Kanger, Tambet Kikas, Marje Särekanno (ARC), Valli Loide (ECRI)
2) integrated plant protection: Triin Saue (ECRI)
3) crop rotation: Ilmar Tamm (ECRI), Rainer Roosimäe (ARC), Aret Vooremäe (EULS), Toivo Lauk (ARC)
4) precision farming (including remote sensing): Toomas Tõrra (EULS), Taavi Võsa (ECRI), Kalvi Tamm (ECRI), Tambet Kikas (ARC), Kalev Sepp (EULS), Janar Raet (EULS), Martin Paukson (Tieto)
5) liaison group: 1-2 persons from each thematic group + Martin Paukson
6) information and communication work group
Urmas Visse (ARC), Maris Kruuse (ARC), Krista Kõiv (ARC), Martin Paukson (Tieto), Andres Lille (Tieto), Stenver Jerkku, Targo Tennisberg (EA).

1.2 Programme council

Programme implementation was managed and monitored by programme council consisting of 13 members.

Big data PIP council composition (approved on 17.09.2018, amended on 04.02.2019):
Toomas Kevvai – representative of The Ministry of Rural Affairs (chairman of the council)
Katrin Rannik – representative of The Ministry of Rural Affairs
Roomet Sõrmus – representative of agricultural organisations (Estonian Chamber of Agriculture and Commerce)
Jaak Läänemets – representative of agricultural organisation (Farming society KEVILI)
Janek Rozov – representative of The Ministry of Economic Affairs and Communications
Karin Kroon – representative of The Ministry of the Environment
Jüri Jõema – representative of Estonian Association of Information Technology and Telecommunications
Ahti Bleive – representative of Estonian Agricultural Registers and Information Board (ARIB)
Andres Kukke – representative of Statistics Estonia
Leho Verk – representative of Rural Development Foundation
Andres Oopkaup – sectoral expert (vice chairman of the council)
Märt Riisberg – sectoral expert
Ahto Vili – sectoral expert

2. Methods for analysis

Various analyses were performed in the course of the project, more significant aspects of those analyses are described below.

Data analysis

The purpose of data analysis was to determine whether definitions of databases in RIHA (*administration system for the state information system*) correspond to actual situation, which data the database contains and what is the quality of those data. Data quality analysis was performed on the basis of real data or close-to-real data. Starting point was to get a practical quality assessment.

The purpose of data analysis and visualisation is to present data in comprehensive and graphic manner that allows ascertaining the services that could be developed by using these data in the framework of big data system. Simultaneously, during visualisation, any anomalies found in data were presented in a form that would not usually be visible in character format.

In order to better understand undefined data, we carried out automatic reverse engineering of several databases, resulting in physical data model in EAP/XMI format. Reverse engineering of databases without suitable database engine or links for performing reverse engineering was carried out manually in the course of data analysis by using data analysis software.

Reverse engineering was performed with majority of Postgre databases where definition was missing or outdated.

The outcome of reverse engineering was presented in database analysis reports as data model diagrams.

Geoinformation analysis is slightly different from other types of data analyses, because it uses formats not supported by majority of analysis tools. Both GIS software and classic data analysis was used for analysing geoinformation.

Definition of metrics and dimensions

Operation of big data system is based on numeric information, viewed by the user through various software versions and by different dimensions. View can be numeric or graphic, e.g. on a map. Map allows displaying e.g. fertilising suggestions, soil data or plant mass index. All this is based on numeric information of soil map, which is supplemented with numeric data of soil samples, resulting in graphic representation. That representation will move in numeric form to farm machine, which controls fertiliser dispenser according to given map. For the system, map has been divided into dots, each of them has individual indicators. These indicators are defined in this project as metrics.

A metric is linked to a dimension. For example, each geographical location (square or dot) has a unique soil composition. Content of substances in soil is measured via analyses and expressed as the amount of nutrient (N, P, K, etc.) per certain unit of soil. Such information is treated as indicator for a square within a field or for the entire field. In this case, field or square represents the dimension. Similarly, all other metrics are always used along with dimension. Dimension can be a particular machine and metric its fuel consumption or crop can be seen as metric and enterprise as dimension. Dimensions include basic objects of data in databases, classifications, consolidated lists, as well as derived dimensions, which may be object groups or clusters formed dynamically in the course of data analysis.

Legal analysis

One project goal was to provide legal analysis of whether and how big data system should be regulated. Legal analysis of big data system in the field of agriculture contained conceptual legal opinion regarding the creation of big data system, transferring data to big data system and other related legal issues.

Big data system is technologically and legally complex project as it contains many contemporary infotechnological features and thus requires relevant legal approach. A proposal concerning the draft project has been made with regard to creation of big data system, describing how big data system could be regulated from legislative aspect.

Legal analysis of big data system has been prepared based on legislation that is currently valid or that will enter in force in near future. Additionally, relevant EU law and professional literature was analysed. Legal analysis was also based on the information entered in the administrative system of state information system and the information generated from IT analysis. Besides that, controllers of databases were contacted to clarify some questions.

In the course of entire project, it became clear that big data system is linked to national law, European Union law and, in certain cases, civil law contracts. When performing legal analysis, it was found that big data system relates to more than 100 different legislative acts that should be considered when creating and implementing big data system. Inasmuch legislative drafting and substantial circumstances incidental thereto are constantly changing, future amendments have to be taken into account when implementing big data system.

Economic analysis

In order to determine the economic feasibility of big data system, an economic analysis was performed concerning the following aspects:

1) What are big system services and their interdependencies?
2) How much will it cost to create the services?
3) What are the benefits or savings as a result of using the services?
4) What is the economic model of big data system like?

Services were mapped by using the following methods:

1) Database analysis ascertained data that could be used for creating services.
2) Project consortium members prepared initial list of services in the course of joint brainstorming.
3) List was supplemented by different parties.
4) Consortium ranked services in order of priority.
5) Services were introduced at big data system information days and additional feedback on priorities was gathered from their future users.
6) Economic model was prepared for the first twelve most important services, indicating the costs of creating and maintaining the service and potential revenue by using net revenue calculation method.
7) Comparison of services was performed by using monetary and social metrics and description of significant factors for creating services.
8) Last step consisted in preparing roadmap for developing the services.

It is important to note that the results of economic analysis comprise subjective assessment of different parties given to the services based on the information known and indicators deployed. Economic analysis was performed to ensure comparability of services, but the absolute values calculated as the result of that might not be final (e.g. development costs, net revenue). One must not sum up net revenue of all services, because calculation of net revenue and comparison of services do not use additional calculations for finding absolute total financial synergy.

More detailed overview of analysis methods is provided in Annex 1.

3. Concept of agricultural big data system

There are many databases in Estonia that contain plenty of information useful for agricultural producers. Several of them contain open data (and hence can be freely used by everyone), but majority of them are intended for departmental use only (they can only be viewed and used by an official and the producer submitting the data or a person authorised by such producer).

For example, some data are generated in the course of applying for agricultural subsidies, some during state supervision or consulting activities. There are also many supporting data, such as base maps, aerial photographs, etc. Farmers also create data in the course of their daily activities: either by using some agricultural software such as eAgronom, Terake, Cropio, etc. or “good old” Excel.

Moreover, data could also be retrieved from agricultural machines as many producers have machines with corresponding capability. In any case, there are lots of agriculture-related data generated, but right now they have extremely different meaning, structure and quality, and it is difficult to analyse them in order to get beneficial results.

This is where **big data system** comes to rescue. It is in fact nothing unfathomable or complicated as one may occasionally hear or read. Naturally, it has sophisticated solutions and methods on the background, but in the eyes of a regular user, big data are data that are abundant enough, have complex structure and are distributed in different databases, so that can no longer be analysed or interpreted by using ordinary means.

Big data system allows reaching new conclusions and value-added during data synthesis and analysis, something that was not attainable before without unreasonable costs. Despite being big data system, it may have a very specific output for particular agricultural producer: e.g. suggestions for fertilising specific field based on soil data, nutrient balance, crop grown, estimated yield, weather forecast and satellite data. This information can be sent to fertilising machine with precision farming capacity, which is used to provide soil of particular field with nutrients in the amount that is necessary according to its unique soil status (and not as recommended by fertiliser vendor). We can be sure that not a single agricultural producer wants to use as much fertiliser as possible; they want to use fertiliser to the exact amount necessary based on nutrient balance. Precision farming is beneficial for both agricultural producer due to reduced cost of input and for natural environment.

In the context of big data system, it is crucial to understand the functioning of entire data value chain – it is important to ensure cooperation between different stakeholders, mainly agricultural producers but also public sector institutions, universities, research institutions, third parties (primarily agricultural software manufacturers). Big data value chain consists of activities that start with collecting data and end with using these data in decision-making processes by either agricultural producer, policymaker, researcher. Data are accessible to everyone according to their user rights and level of generalisation of data. In the context of agriculture our current data value chain is very fragmental, because data are difficult to access, and their quality is extremely uneven.

In Estonia, big data systems are created in several areas. However, there are currently no established practices and standards concerning how these systems should be created. There is no common agreement on what the big data are. Therefore, this project aims at cooperation between different administrative areas to find as comprehensive solution as possible, attempting to utilise all knowledge generated in the country so far.

For the purposes of this project, big data are data that comply with the following criteria:

1) Data are added very quickly;
2) Data are very complex;
3) Data have great volume;
4) Data content is not known in detail or data are not structured.

The analysis did not find such data that would comply with all those criteria. It is certainly possible to provide subjective assessment of criteria, but when compared to the rest of the world, the volume of agricultural data generated in Estonia is rather low.

However, in the course of the project we managed to find future sources of big data – mostly data from agricultural machines or fieldwork data, detailed data concerning ground and soil, and indexes calculated based on satellite images, when presented for each square metre of ground. Such data are not yet used on national level and there are also no systems to register and utilise them. First step seems to be systematic registration of field work data in electronic format (electronic field data book) and calculation of indexes based on satellite image (EstHUB), which are common to crop production.

Telemetric data retrieved from agricultural machines represent one of major big data topics that would benefit from technology. Analysis revealed that for agricultural producers, one remarkable problem consists in exchanging data with agricultural machines. Different machine manufacturers have developed their own systems that are not mutually compatible. As a rule, one has to pay for using the full functionality of the system provided by the manufacturer. This means that if a producer owns several agricultural machines from different manufacturers, he has to use more than one software and pay contractual fees. In general, it is impossible to use one and the same control device for machines made by different manufacturers due to their close connection to software used by relevant manufacturer. Figuratively speaking, agricultural producers have to pay for the data they themselves have generated (and which they own).

The aforesaid led to the conclusion that in the course of big data project, it is necessary to create software for exchanging data with agricultural machines that would ensure uniform data exchange with all machines that comply with the requirements of ISO 11783 standard family. This concerns primarily reading the operational data of the machine and recording them in electronic field data book. When using data, one must consider that their quality may vary depending on the machine used. Before recording data in electronic field data book, they need to be tidied up where necessary. Even if the data need to be tidied up, automatic data exchange with the machine would still mean considerable timesaving when adding data to the field data book. Furthermore, one has to consider that every manufacturer has their own rules (contract for using the machine) which limit the use of data. On one hand, such limitations are related to machine safety and reliability, but on the other hand also to manufacturer's copyrights. In any case, involving telemetric data of machines in big data system has key importance. Attention should be paid to the fact that under NIVA project, the project partner from Netherlands deals with telemetric data of machines and their results should be carefully observed. If this project fails to show any progress in this field, Estonia could take initiative on EU level. Reading machine data does not fit in Stage II of the project, but it should be dealt with as soon as possible.

The concept of this big data system is based on data analysis, legal analysis and economic analysis of agriculture-related databases, including analysis of potential big data system services.

Data of big data system are divided into the following categories:

1) Real data concerning objects and subjects. Big data system contains some real data that are not included in interfaced databases and that cannot be included in any other database. Additionally, it is possible to copy data from various databases to big data system when performing major analyses, as to allow performing the analysis in real time and allow repeating it with supplemented algorithms.
2) Descriptions of real data and services or metadata.
3) Information concerning individuals (system users) and their rights and authorisations.

Big data system **will not permanently contain** data of other databases, including spatial information, for which the Land Board has developed strong infrastructure and publicly used components, and there is no point in competing with that. This also applies to the environment for processing ESTHub satellite images, created by Land Board, capacity of which will be used for developing and calculating various indexes in future.

The most likely institution to create and manage agricultural big data system and central system for electronic field data book is Agricultural Research Centre in close cooperation with ARIB and other administrative institutions of the Ministry of Rural Affairs and private sector enterprises. In terms of big data system, electronic field data book is a consolidated database as any other such database, but it is also main "gateway" for agricultural producers to access big data system. It would be reasonable to develop these two systems simultaneously (several services require the existence of electronic field data book in order to be used for organising agricultural production).

The concept of agricultural big data is described in greater detail in Annex 2, which contains in-depth description of big data system. Other extras include roadmap files (Annex 2a) and estimated cost of activities (Annex 2b), also descriptions of metrics for selected services (Annexes 2c, d, e and f) and summary of idea collection regarding the services carried out in the course of the project (Annex 2g).

4. Data analysis

The purpose of data analyses carried out in the framework of feasibility study of big data system was to ascertain the dataset of databases linked to big data system, its data quality and whether it is possible and necessary to include data from database in big data system for creating services.

Analysis was based on the fact that although by law, the dataset of databases is described in RIHA, the reality differs from the description provided in RIHA and in some cases to significant extent. This was confirmed by the analysis. More than half of the descriptions of databases had no description of data composition in RIHA or it differed significantly from actual composition.

During the project, data were viewed by using analytical methods that presume access to real data (so-called *live* data). Additionally, interviews were carried out with specialists acquainted with the content and technology of databases as well as interested parties. Summaries of real data analysis compiled for each database describe whether and how the data of particular database could be used in big data system. Furthermore, the development needs of databases were assessed from substantive and financial aspect.

In analysed databases, large amount of information is oriented towards supporting work processes within or related to various institutions. Such information is not attractive for most parties on rational grounds, as it would not facilitate their business activities. However, evaluations of the data of database and especially data quality assessments may provide significant information for the controller or processor of the database, allowing to optimise the database and plan its development. Therefore, performed analyses help to save state funds.

In the course of data analysis, we analysed 41 databases, majority of which can be merged with agricultural big data system. At the same time, it must be noted that only a few databases are fully ready for merging.

Main problems found in databases:

1) Lack of X-Road or web services that would allow making data inquiries from database.
2) Significant problems with the quality of data in database.
3) Lack of metadata, RIHA description does not correspond to actual situation.
4) Data are present, but database has not been lawfully established.
5) Database lacks special software for data management, which increases the risk of corruption and loss of data.
6) Data of database are not disclosed, or database holders do not wish to disclose data.

Detailed solutions for the said bottlenecks are provided in database analysis documents intended for in-house use and are thus not disclosed.

Summary of data analysis is presented in Annex 4.

5. What are potential services and how are they made accessible?

Main objective of agricultural big data programme is to allow agricultural producers access to data of various databases when making economic decisions, compare the data of own production unit with the data of other production units and receive additional data analysis and recommendations.

It is certainly important to note that the state does not necessarily have to create all services in the agricultural big data system. Instead, the state should take care of the services related to mandatory submission of certain data to the state and for the services based on the analysis results of these data – the rest of services can be realised by other market players. It is also possible to establish automated reporting to the state (e.g. Statistics Estonia, ARIB). Another task of the state consists in taking care of "survival" of the basic components of big data system (e.g. metadata set, authorisation and authentication, certain important services not provided by other system members), standardised data exchange interfaces.

Key importance is ascribed to the creation of national electronic field data book and maintaining it only in electronic environment in the future. Here, the only role of state consists in making available a free environment for electronic field data book, which allows submitting minimum set of data required by legislation. State may create such application by itself or purchase its development and maintenance from private sector via public procurement.

During project idea collection were mapped approximately 60 services that could theoretically be created based on analysed data. Summary of the idea collection is presented in Annex 2g.

6. Economic analysis

Each agricultural big data service has its cost of creation and maintenance and someone has to pay for it. Either in specific or approximate figures, it would be beneficial to agricultural producer, authorities, research institutions, environment. This was considered in agricultural big data project as it comprised as many inputs as possible. The analyses resulted in creating so-called roadmap (Annex 2, Chapter 9), based on which it would be possible to commence creating agricultural big data system (along with investment needs). Once again, it has to be noted that the roadmap suggested based on the project is subject to potential changes with regard to FaST and NIVA projects and services created within framework of those projects (to avoid duplication).

Summary of the economic analysis of services is presented in chapter 8.2 of Annex 2.

7. Legal analysis

Legal analysis is a crucial component of the project. With IT-related innovation projects it sometimes happens that the parties involved have great ideas about which data to collect or share or which services to realise, but legal framework currently in force may become an obstacle to that. It is possible to change or specify regulations in a manner that allows using modern technological solutions, but at the same time it is important to ensure the rights and obligations of different parties arising from prevailing legislation. Therefore, the system to be created must be in accordance with our fundamental rights and relevant European Union law. Legal analysis ascertained the legal framework for creating such system, functions of that framework, owners of big data system and data processors, it is also necessary to establish the criteria for access to data and the options for financing big data system.

During the project, the legal situation of persons submitting data is mapped, whether it is database or in-house database or another technical solution – this is the basis for creating legal solutions for transferring data to big data system. It includes description of general need for changes in legislation or prerequisites for transferring data into a central system.

Furthermore, it also involves mapping and analysis of all other legal issues related to data exchange, such as the rights of agricultural producer, data security, issuing data to third parties, in some cases certain state agencies may have the right to access data.

Summary of legal analysis is presented in Annex 3.

8. Electronic field data book

Creation of electronic field data book and keeping only electronic records of field data book in the future is crucial for the success of agricultural big data system.

In the future, electronic field data book will be main “gateway” for the producer to access the data and analyses present in the agricultural big data system. Electronic field data book represents one of the most important services in big data system and should be realised at the same time. However, creation of agricultural big data system and electronic field data book are also independent of each other in a way. One significant prerequisite consists in establishing the legal framework: it is necessary to create legal bases for the electronic field data book, procedure for transferring data from private sector software and to consider alternative legal bases in order to allow transferring data from state databases to private sector software. Longer analysis of the subjects of electronic field data book from legal point of view is provided in clause 6.6 of the legal analysis.

The role of state upon establishment of the electronic field data book is limited to creating the database of electronic field data book, providing a free user environment of electronic field data book (e.g. via the client portal of the Ministry of Rural Affairs), which allows submission of minimum set of data required in various legislation. Any data beyond such minimum obligation can be realised by agricultural software manufacturers in accordance with market conditions. However, it is possible to develop an option where, in case of submitting additional data (e.g. dataset of agricultural machines), the state can reduce red tape in other proceedings or provide agricultural producer with further recommendations or analyses. By no means can state start competing with agricultural software manufacturers. Another option is to realise the electronic field data book solution in cooperation with private sector, including aforementioned minimum set available for free use.

Expected common practice in the future is that agricultural producer uses a certain third-party manufactured agricultural software in organising its daily activities, generating the data, including those in national electronic field data book, in the course of routine work, without taking up any additional time. Ideally, the electronic field data book “auto-fills” and communicates with state database to the extent established by legislation. Using agricultural software cannot be seen only as cost – it is first and foremost a means for production, just like a tractor or harvester.

Electronic field data book is a digital information system that preserves information about fields and operations carried out in the fields. **Operations** performed in the field represent central keyword in the electronic field data book, they are linked to field identification, area (may vary depending on operation) and materials used, or crops grown. In the course of the project two options were proposed for using the electronic field data book: "big portal" and "small portal". More detailed document on the analysis of electronic field data book and description of standard is presented in Annex 5.

Important **participants** in the use scenarios of electronic field data book are listed below:

1. **Agricultural producers** can add data and view data through either state or third-party applications.
2. **Researchers** use the data in electronic field data book via relevant analysis environment.
3. **Statisticians** use the data in electronic field data book via analysis environment, where they can be imported to suitable systems of e.g. Statistics Estonia or another stakeholder.
4. Previously mentioned statistics can be used for shaping **agricultural policy**.
5. **Surveillance authorities** see the data transferred from electronic field data book to corresponding information system, e.g. allowing to send agricultural producers automated rule-based **recommendations and reminders**.
6. In any case, state provides a “window” for transferring data, nobody is forced to buy agricultural software. If producers need to view or add the data in electronic field data book, they can use an existing web interface (e.g. client portal of the Ministry of Rural Affairs) – if they do not use a particular agricultural software.
7. Electronic field data book is open to **third party applications** for making data inquiries and adding data. This means that producer can use field data book by using a different software, which exchanges data with the database of central electronic field data book in the name of the producer. Viewing and changing the data requires having relevant access rights and needs corresponding legal bases to arrange such data exchange service. Such external applications may keep their data in separate databases as well.
8. The data of electronic field data book can be used for various **analyses and estimations**, combining the data of field data book with the data in other databases where necessary.
9. The data in electronic field data book can be used in the information systems of agriculture-related agencies (e.g. ARIB, Board of Agriculture, Environmental Inspectorate). Here are some instances where it can be used:
    1. **Performance of tasks** (including compliance checks) arising from legislation (primarily Water Act and Plant Protection Act) can be organised more efficiently, because data are available in the same format. In the interest of more successful implementation of electronic field data book it is important to use new opportunities primarily for making suggestions and not sanctions.
    2. More efficient proceeding of various **agricultural subsidies**, e.g. pre-filled forms using the data retrieved from electronic field data book. Once the producer has entered

certain data in field data book or submitted the data to the state for other reason, there should be no need to re-enter them.

10. Besides field-specific analysis, the data in electronic field data book can be processed **statistically**, by exploring topics such as “use of nitrogen fertilisers in Estonia by counties and months”. Here are two main directions: firstly, the models created in the course of **research work** ca be used for improved further analysis of fields; secondly, shaping **official statistics and agricultural policy**.
11. Data are **added** to the national electronic field data book through relevant **services** that may be executed as e.g. X-Road or regular web services (as administrative proceeding does not generally have fixed format). No data are added outside these services. The definition of data used in the services and the list of services are described in Annex 5 Chapter 10. X-Road services of electronic field data book.
12. Data **inquiries** from national electronic field data book are made through relevant **services** that may be executed as e.g. X-Road or regular web services. No data inquiries are made outside these services, except for the purposes of statistical analysis. The definition of data used in the services and the list of services are described in Annex 5 Chapter 10. X-Road services of electronic field data book.
13. **Additional data inquiries** can be made in external databases. In order to facilitate data use, relevant services can execute API equivalent to the one used for national electronic field data book. It is recommended that these services use the same data structures as national field data book, see Annex 5 Chapter 7. Data structures used in services of electronic field data book.
14. Miscellaneous other databases execute their own **data inquiry services** (e.g. X-Road or web services), which can be combined with the data in electronic field data book.
15. Electronic field data book attempts to avoid management of own persons and authorisations, as this is already covered in existing systems. The purpose is to arrange authentication (personal identification) and authorisation (management of authorisations), by using suitable **external system**. See Annex 5 Chapter 5. Implementation of electronic field data book.
16. Data can be **analysed** in corresponding environment by using either web-based visualisations or special software.
17. Analysis is performed on the basis of **analysis packages** exported from various databases (including electronic field data book)**.**
18. **Supplementary data of electronic field data book** may be stored e.g. in third party databases. It must be emphasized that these data are legally speaking not state-related and require establishment of data processing agreement between software owner and agricultural producer, and state has no part in it.
19. **Database of national electronic field data book** is accessible only through relevant services. The details of using the database of field data book will be clarified in corresponding development project, but Annex 5 contains some recommendations: 11. Recommendations for executing the database (logical data model).

As seen from the figure, the “ecosystem” related to electronic field data book consists of three layers: **data**, **services** and **applications**. The analysis performed in the course of the project does not prescribe specific platforms for any of the components, but any solution must follow the principle of separation of data, services and applications. More specifically:

- **Database** of electronic field data book only stores data and does not contain business rules or direct access options
- In this context, **services** indicate methods necessary for adding data and making inquiries, executed e.g. by means of X-Road or regular web services. An example of typical service is “Add field work with parameters X, Y and Z” or “Send inquiry about material consumption in

the predetermined geographical area". The services do not execute other business logic beside simple data validation.

- **Applications** have a "more interesting" business logic, controls, subsidies, estimations, etc.

Electronic field data book would give the following benefits:

1. Benefits for agricultural producers
   1. Less duplication when entering data and pre-filled forms, including:
      1. ARIB subsidy applications can use data in electronic field data book.
      2. There is no need to maintain several different plans (fertilisation, plant protection) and submit it in different places.
      3. Statistics Estonia receives lots of information from the electronic field data book and farmers do not have to complete fewer questionnaires.
   2. Better agronomic decisions – data in electronic field data book allow e.g. ARC to make more comprehensive and precise suggestions for fertiliser use for a particular field and its subunits.
   3. Opportunity to receive better recommendations and reminders to meet the requirements established by support schemes, laws and regulations.
2. National benefits
   1. Easier to check the compliance with legal requirements (e.g. Water Act and Plant Protection Act) – inspectors can use digital information, thus reducing the need for on-site visits.
   2. More efficient statistics and better opportunities for shaping agricultural policy – availability of better and more complete data on fields, field works, and materials used.
   3. Quicker and more accurate information collection due to possibility of using various pre-filled forms.
   4. In case of successful project, the experience can be extended to other EU countries.
3. Benefits for agricultural research
   1. Data on fields, works, and materials used are digitalised and in the same format.
   2. Quicker and easier access to data.
   3. Significantly more data and better data quality.
4. Benefits for agricultural software manufacturers
   1. Standard interface for submitting field and field work data.
   2. In case of mandatory use of electronic field data book, the producers will be more motivated to use special software.
   3. Opportunity to provide farmers with better analyses and estimations.
5. Social benefits
   1. Improved environment protection due to reduced use of fertilisers and other materials due to better planning and estimation options.

9. Links to similar projects.

FaST *(Farm Sustainability Tool*) – digital tool for optimising nutrients use in farming

This project is directly related to agricultural big data project and also to NIVA project.

General objectives of FaST project:

- Achievement of environmental protection targets and economic benefits for agricultural producer
- Sufficient level of accuracy for required effect
- Digitalisation of agricultural sector, digital tools at the disposal of agricultural producer
- Facilitating compliance with requirements for agricultural producer (training and instruction materials)

Various digital tools are planned in the framework of the project, primarily with regard to nutrient balance. More versatile use of satellite information is planned.

Project summary: https://ec.europa.eu/info/news/new-tool-increase-sustainable-use-nutrients-across-eu-2019-feb-19_en

FaST demo: https://github.com/PwC-FaST/

Presentation made at the agricultural big data conference on 2 July 2019: Expectations of the European Commission towards digital agriculture, emphasis on FaST tool – Isidro Campos-Rodriguez (European Commission, DG-AGRI).

NIVA (***N**ew **I**ACS **V**ision in **A**ction*) **-** Integrated Administration and Control System

General purpose of NIVA project is to update and supplement on EU level the IAKS system created for administering EU area and animal subsidies for new program period. It also aims at creating various new technical tools (farm register, photo system with location tags, automated system for detecting field register updates, etc.).

European Commission proposed moving towards a more flexible system that allows simplified and updated implementation of agricultural policy. Main focus should shift from requirements and compliance with requirements to results and performance.

Three main challenges of NIVA project:

- Use innovation to simplify implementing governance/policy;
- Reduce administrative burden of agricultural producers;
- Extend using data of IAKS for "other purposes".

Several pilot projects have been initiated in cooperation of 9 countries in the following areas:

- Monitoring of environment (subsidies);
- Performance of agricultural producers;
- Farm register;
- Data on agricultural machines;
- Automated updating of field register;
- Request-free system;
- Ground monitoring/so-called traffic lights system;
- Pre-filled application;

- Photos with location tag.

NIVA and FaST projects allow providing agricultural big data project with significant value-added (and vice versa), and therefore, subsequent stages should involve close cooperation.

SOSTARE document related to NIVA project: Paracchini et al. 2015 SOSTARE.PDF

Action plan for digitalisation of NPA Dairy Cluster

Stage I of the agricultural big data programme is focused on plant production, but the architecture of the system allows adding any data source that can meet prescribed standards. Action plan for digitalisation of Dairy Cluster fits big data project very well and requires close cooperation in subsequent stages of creating big data system.

Activities of the project for digitalisation of Dairy Cluster are divided into two main topics:

- Creating a solution for automated collection of bovine health data and assessment of its efficiency
- Multilateral mapping of data exchange stories and linking data

Dairy farming is certainly a strategic area in Estonia as it fits perfectly with climatic conditions and many other prerequisites. Increasingly intense production conditions have led to reduced longevity and reproduction, which indicates the need for systemic work with animal health and feeding issues. Particular attention should be paid to the health of young animals. Increasing shortage of labour adds to the tensions in animal husbandry. Wider introduction of digital tools, including development of new tools, is one way to find solutions. In order to use that potential, it is necessary to create a solution for automated collection and sharing of bovine health data for dairy farmers and use these data in the supply chain in combination with other data.

The most significant goal of the action plan is to find new solutions to problems arising from labour shortage. Digital solutions in dairy farming should allow distinguishing Estonian dairy products from those of competitors. This is expected to increase the amount of more expensive dairy products sold by the industries. After all, main goal of the Dairy Cluster consists in increased export capacity.

implementation of collection of health data is a significant step in the development of technologies of precision farming. Animal health data provide a basis for drawing conclusions about the animal, herd, and entire dairy sector (big data). Great future potential can be attributed to the big data system, where transfer and automated processing of health data of an individual animal essentially allows creating electronic bovine health passport. Second stage applications include automated herd- and population-specific information management and interaction with private sector organisations and state agencies (e-ARIB; Veterinary Supervision, Animal Recording, etc.). Resulting savings and new business opportunities arising from linking and exchanging data are measurable in millions of euros, not to mention providing a solution in a situation where some things are neglected merely due to shortage of labour.

Applications for Estonian Sheep and Goat Breeders Association (ELKL)

Estonian Sheep and Goat Breeders Association (ELKL) has made proposals regarding agricultural big data project in two areas: firstly, they want to be contribute to the development of data exchange options with animals register of ARIB by example of ELKL breeding information system "Pässu 2.0" and secondly, they are interested in involvement in the system of classifications of sheep and goat breeds.

ELKL online information system "Pässu 2.0" is intended first and foremost for collecting and managing data on sheep and goats of breeders participating in performance testing and for collection statistical data necessary for breeding.

In future software developments, ELKL will make efforts to facilitate entering of animal data (tagging, delivery, movement, etc.) and ensure that cattle farmer would not have to do extra work when entering the data in “Pässu” system and ARIB animal register.

Thus, ELKL is interested in development of both technical and legal mechanisms in the framework of agricultural big data project in order to allow so-called third-party information systems (such as Pässu 2.0 and eAgronom that was included in this project) to authenticate their users and authorise them to forward data to state registers in a way that ensures technical security at minimum administrative cost. Interface should be two-way, i.e. cattle farmer must be able to both submit the data concerning his animals to ARIB and make inquiries regarding his animals.

With regard to animal register, ELKL wishes to continue involvement in the development of sheep and goat breeds classification system. ELKL has already made proposals directly to ARIB concerning abbreviations used for sheep and goat breeds and these proposals have been taken into account. However, there is another suggestion, according to which the information systems that use breed classifications should be notified about changes or additions in classifications based on a certain procedure; it would be even better to have relevant software options (last change timestamp in the descriptive interface of the classification, etc.).

10. Connections to national IT frameworks and state data governance principles

Development of the concept of agricultural big data system takes into account state data governance principles defined in Estonian data governance management action plan. Pursuant to the action plan, the following subareas of data governance will be managed:

1) Data definition and metadata.
2) Overview of datasets, basic data.
3) Data quality.
4) Disclosing data and open data.
5) Data lifecycle management.
6) Establishment of the function of data stewards to databases.

In the course of this present project, the situation of 41 databases was mapped in terms of data definitions and metadata. Definitions were provided for data and metadata to be linked to the services of big data system. Definition standard in the form of physical data model was created for defining metadata.

Project activities involved compiling an overview of datasets, visualisation of data that may be used for more than just managing the operation of database. Visualisations were turned into an integrated application (tietoanalytics.ee/ARIB), which is available online to any interested parties. This application can also be referred to as agricultural data catalog.

All databases included in the project were analysed in terms of data quality, including the presence of (mandatory) values, logical connections, use of classifications, accuracy and use of personal identification codes, compliance of addresses with ADS, and other aspects described in the data quality standard for big data system. The concept document of big data system described data quality standards that can also be used in other areas. Data quality standard was compiled by using information from document "Guidelines for data quality assurance for database owners".

Project included mapping open datasets of databases and compiling their comprehensive visualisations. It also included determination of technical standards (web services standards) that should be used when disclosing data. Main principle is that in addition to disclosed data files, every database should be equipped with web services that allow automated use of open data. File-based dataset makes automation prone to errors and requires excessive manual work by data consumer. Depending on the nature of data, the services used should be either RESTful (OpenData) webservices, or in case of spatial information WMS, WFS and WCS services.

With regard to the architecture of the planned big data system, we consulted state IT architect. It was agreed that the architecture is defined by using UML *component model*. Detailed UMP deployment views are not defined as they contain information that make the system vulnerable. Deployment views are used only within the institution managing the big data system.

Pursuant to the requirements of the interoperability framework of state information system, all databases must be described in RIHA. Database analyses clearly indicate that RIHA is not functional in its current form, and the work carried out in the course of this project further confirmed it for the State information systems department (RISO). Currently, the state is planning to introduce new RIHA that would be directly linked to the information systems to be created. According to the plan, each application will also contain RIHA information and RIHA will be updated as soon as a new applications installed.

Project did not discuss in depth data stewards' role or data life cycle management principles that according to the Estonian data governance management action plan will be dealt with in the second half of 2019 and in 2020.

ANNEX 1. Used methodology

https://confluence.agri.ee/display/PIP/Lisa+1.+Kasutatud+metoodika

ANNEX 2. Description of big data concept. Services. Economic analysis. Roadmap

https://confluence.agri.ee/pages/viewpage.action?pageId=80708388

ANNEX 3. Legal analysis

https://confluence.agri.ee/pages/viewpage.action?pageId=80708394

ANNEX 4. Summary of data analysis

https://confluence.agri.ee/pages/viewpage.action?pageId=80708396

ANNEX 5. Analysis document of electronic field data book

https://confluence.agri.ee/pages/viewpage.action?pageId=80708398

ANNEX 6. Consolidated outcome of project stages

https://confluence.agri.ee/display/PIP/Lisa+6.+Projecti+etappide+tulemite+koond

ANNEX 7. Background materials

https://confluence.agri.ee/display/PIP/Lisa+7.+Taustamaterjalid