MIND STEP

MODELLING INDIVIDUAL DECISIONS TO SUPPORT THE EUROPEAN POLICIES RELATED TO AGRICULTURE

Deliverable D2.3:
Common Processing Plan

Possible Synergies within the AGRICLUSTER projects

AUTHORS	Marc Mueller (MIND STEP, WR), John Helming (MIND STEP, WR), Carlos Leyva (Agricore, IDENER), Pablo Báez González (Agricore, IDENER), Guy Ziv (BESTMAP, Uni Leeds), Jodi Gunning (BESTMAP, Uni Leeds), Alexander Gocht (MIND STEP, Thuenen)

APPROVED BY WP MANAGER:	John Helming (WR)
DATE OF APPROVAL:	28-05-2021
APPROVED BY PROJECT COORDINATOR:	Hans van Meijl (WR)
DATE OF APPROVAL:	31-05-2021

CALL H2020-RUR-2018-2	Rural Renaissance
WORK PROGRAMME Topic RUR-04-2018	Analytical tools and models to support policies related to agriculture and food - RIA Research and Innovation action
PROJECT WEB SITE:	https://mind-step.eu

TABLE OF CONTENTS

ACRONYMS

MIND STEP Modelling INdividual Decisions to Support The European Policies related to agriculture

BESTMAP Behavioural, Ecological and Socio-Economic Tools for Modelling Agricultural Policy

AGRICORE Agent-Based Support Tool For The Development of Agricultural Policies

AGRIMODELS cluster MIND STEP, BESTMAP and AGRICORE forming together the AGRIMODELS cluster in order to in order to coordinate the potential synergies between them. Please find below the description for each one of them.

FADN Farm Accountancy Data Network

EAAE European Association of Agricultural Economics

FAIR Findable, Accessible, Interoperable and Reusable

GIT/GITHUB software for tracking changes in any set of files

SVN Apache Subversion

POSTGRES relational database management system

ARDIT Agricultural Research Data Index Tool

NUTS Nomenclature des Unités territoriales statistiques

LPIS Land Parcel Identification System

IACS Integrated administration and control system

FSS Farm structure survey

GA Grand Agreement

GDPR General Data Protection Regulation

EXECUTIVE SUMMARY

This deliverable 2.3 summarises the two workshops organised within the AGRIMODELS cluster (https://agrimodels-cluster.eu/), that organises the three RUR-04 projects, discussing the possibilities and difficulties of a common data processing plan. On the agenda were among others i) data management plan already developed ii) how to support reproducibility and transparency of the data and how to make the work public accessible, enables easy review, reanalysis, replication and reuse e.g. regarding use of FADN data iii) explore the possibility to setup a tool for continuous integration iv) possible ethical requirements and v) plans for data sharing and publication. It is concluded that the exchange of individual farm data, either from European or national FADN, will not be possible due to data protection regulations. Common data processing is therefore restricted to publicly available data, research findings at sufficiently highly aggregated levels, and scenario narratives, possibly including a common baseline. The exchange of data processing methods and utilities will be possible, provided that the developed utilities can accommodate a sufficiently wide range of datasets. Furthermore, joint dissemination and publication activities were launched and will be continued for the remainders of the project.

1. INTRODUCTION

The MIND STEP project, together with the BESTMAP (http://bestmap.eu/) and AGRICORE projects (https://agricore-project.eu/), are funded under the European Commission's Rural Renaissance (H2020-RUR-2018) call. Together, they form the AGRIMODELS cluster (https://agrimodels-cluster.eu/), which aims at developing innovative approaches to increase modelling capabilities in and for the agricultural sector. The modelling activities in each project of the AGRIMODELS cluster include aspects like the environmental and climatic impacts of farming, the delivery of ecosystem services, linkages between the sector as a whole and individual farming systems, structural change including the transfer of production factors such as land, the integration of agriculture in rural society, and the establishment of links with biophysical models and geo-referenced datasets. The modelling activities capture various geographic scales, ranging from regional to global. A major goal in each project is to build highly modular and customisable suites of tools which will allow flexible use and further improvements as needs arise.

These common aspects of the modelling activities within the three AGRIMODELS cluster projects have the potential to create a wide range of overlapping interests, for instance regarding the identification of innovative modelling approaches, the selection and acquisition of data, or developing joint dissemination strategies. This deliverable 2.3 of the MIND STEP project aims at finding ways to share the burden of identifying and processing all relevant economic and biophysical data as it requires resources, needs repetition and continuation given the continuous availability of new up-dated data. A further goal here is to explore the idea of a public available interface database for farm data to encourage future collaboration and continuous research in that field and easy access to harmonized data for the AGRIMODELS cluster projects.

Consultative meetings of the project coordinators started taking place immediately after the launch of the projects in September 2019. Workshops dedicated to sharing experiences and ideas about data acquisition, dissemination of results, and early publication opportunities were held in November 2020 and March 2021. During this time experiences with data acquisition focussed mainly on difficulties encountered when requesting data from the Farm Accountancy Data Network (FADN). An upcoming event for a first joint publication exercise will be an organized session titled "Use Of Individual Decision Making Models To Support Policies." during the 16th EAAE Congress 2021.

The following sections provide a summary of the two workshops held so far, the main challenges for data processing and the conclusions for further collaboration within the AGRIMODELS cluster.

2. AGRIMODELS CONSORTIUM MEETINGS

2.1. Workshop 02 November 2020

Participants in the first workshop were: John Helming (MIND STEP), Carlos Leyva (Agricore), Pablo Báez González (Agricore), Guy Ziv (BESTMAP), Jodi Gunning (BESTMAP), Marc Mueller (MIND STEP), Alexander Gocht (MIND STEP)

The major topic of this workshop was to operationalize the concept of a data processing plan that can be built together with parallel working consortia, namely to share the burden of identifying and processing all relevant economic and biophysical data. Further, the potential to create a publicly available interface for the exchange of databases for farm data to encourage future collaboration and continuous research was explored. From a MIND STEP perspective, this involves establishing a data management plan which ensures that research data should be findable, accessible, interoperable and

reusable (FAIR) as required by the Programme Guidelines on FAIR Data Management in Horizon 2020. Within MIND STEP, this requirement will be implemented by a number of IT services, namely a continuous integration system and a distributed version control system in combination with an online developer platform, like GITHUB or SVN. This supports reproducibility and transparency of the development and makes the work public accessible, while enabling easy review, reanalysis, replication and reuse of models.

For data exchange, a geo-server in combination with a geo-database such as POSTGRES to provide data- and download services of results obtained from the model toolbox will also be installed by the MIND STEP project. Such publicly accessible data exchange services are constrained by a set of ethical requirements related to the protection of personal data and the prevention of the misuse of research findings. In particular, the risk that individual farms or households can be re-identified by e.g. the combination with additional datasets, has to be prevented. As a consequence, the MIND STEP project has to find a balance between the obligation to publish findings and make them available according to the FAIR principle, and the obligation to protect confidential information and ensure that only entitled users can access the data.

These restrictions are particularly relevant for the use and dissemination of results originating from FADN data.

During this workshop the FADNUtils package/tool was discussed. This package for the programming language R facilitates the efficient handling of FADN data within the R language framework. The tool is open source and will be further developed in the MIND STEP project. It should be noted that the tool is only applicable to FADN, not national FADN. At the time of this workshop, the MIND STEP project had submitted a joint, project-wide request for FADN data from DG AGRI, and was invited to provide a justification for the extent of the requested data during a committee meeting with the responsible staff at DG AGRI, so the data itself was not yet available for the MIND STEP project.

In particular the latter experience to encounter difficulties when requesting FADN data was shared by the BESTMAP project as their request had been rejected and the team was searching for solutions, either by re-submitting the request or by finding alternative databases. A particular challenge was that the coordinating organization of BESTMAP is the University of Leeds, hence based in the UK, which implies the transfer of EU FADN data to a non-EU organisation. The AGRICORE team had not applied for FADN data at that time, partly due to the fact that AGRICORE partners rely dominantly on national-level FADN.

With regard to the development of a data management plan, the AGRICORE project develops an Agricultural Research Data Index Tool (ARDIT, Veneziani et al. 2020) to facilitate the task of identifying relevant and useful data for performing agricultural policy analysis. To do so, the AGRICORE partners devised a characterisation methodology for describing the available data sources and their content to enable proper mapping and searching capabilities over the gathered information. To design such a methodology, one of the key elements is the definition or adoption of an ontology (or a set of them), which allows sharing a common understanding of the structure of information among people or software agents, enabling the reuse of domain knowledge, making domain assumptions explicit, separating domain knowledge from the operational knowledge, analysing the domain knowledge, and securing the interoperability of datasets.

The adoption of a comparable ontology could be of interest to the other projects as well, but depends on specific use cases to become operational.

Web-applications offering IT services for data exchange and management were not established in any of the projects at the time of the workshop. A kick-off meeting for the AGRIMODELS cluster was planned as part of the AGRICORE project already in month 12, but needed to be postponed due to the COVID-19 restrictions. A new date could be found in Summer 2021 around the finalization of the first

reporting phase. The workshop concluded that the pending issues have to be re-visited during a follow-up workshop in early 2021, when there is a chance that new information could be obtained.

2.2. Workshop 29 March 2021

Participants in the second workshop were: John Helming (MIND STEP), Pablo Báez González (AGRICORE), Guy Ziv (BESTMAP), Marc Mueller (MIND STEP), Alexander Gocht (MIND STEP)

2.2.1. Acquisition and use of FADN data

In continuation of the previous workshop, the first topic to be discussed was the state of FADN data requests by the three projects. In the case of BESTMAP, the request had been granted by DG AGRI, but objections regarding the agreement form had been raised by the legal department of Leeds University because of the liability of the project coordinator in case of breaches of data security at the project partners. A request to change the agreement form was rejected by DG AGRI. A possible solution would be that each partner submits an individual request, but that bears the risk that farm identifiers are different for each partner and the exchange and testing of methods will be hampered. In case of changing farm identifiers, it could be possible to request correspondences between the identifiers from DG AGRI to link the farm keys from different requests. Technically, this has to be possible, but there could be legal obstacles. If farm keys can be harmonized among partners of the same project, exchange of data and methods can be facilitated substantially within the restriction given by data protection guidelines. Still, the legal issues related to partner-wise data requests and their linkage by correspondences between farm identifiers are difficult to solve for projects with multiple partners.

In case of the MIND STEP project, there have been delays because additional documentation and justification was requested by DG AGRI. In addition, the agreement forms had to be reviewed by the legal departments of each partner organization prior to signature. Subsequently, the signed originals had to be collected from each partner and sent by ordinary mail to DG AGRI, which took longer than usual due to COVID-19 restrictions, i.e. that most involved persons worked from home. Once the original signatures were received by DG AGRI and remaining questions were clarified with the responsible IT staff, the datasets were made available for the MIND STEP partners.

The AGRICORE project currently relies on national FADN to estimate the probability density functions and can therefore already start working with individual farm-level data. However, according to the grant agreement, AGRICORE will apply for EU FADN before the end of month 24 (i.e. August 2021). A challenge is the selection of FADN variables as it is not obvious which ones are really needed. In the case of MIND STEP, this problem was solved by starting from a minimum list of variable needed to parameterize the IFM-CAP model used at JRC and then expanding the list by additional partner requests.

In general, the use of national FADN data could also be an alternative for BESTMAP in case the EU FADN data cannot be obtained. A drawback is that this approach can prevent the use of exploitation of processing tools like FADNUtils, which is used in MIND STEP, but is based on EU FADN variable definitions.

A further alternative to individual EU FADN data could be the use of aggregated information for farm types at NUTS2 or NUTS3 level and create a sample farm population based on weights for farm-types obtained from other sources like the farm structure survey (FSS). While technically possible, this different time intervals in which the data are collected may damage the reliability of the obtained results.

Concerning interfaces between biophysical databases and farm-level statistics, BESTMAP and MIND STEP have started working on prototypes: In the case of BESTMAP, a link between LPIS/IACS data and farming systems archetypes (economic size and farm specialisation) is being tested. This is comparable to the activities in MIND STEP to create probabilistic spatial matches between socio-economic entities and environmental data. In addition, the MIND STEP project is in an advantageous situation for case studies in the Netherlands because spatial data from AgroDataCube can be linked to census data by farm identifiers. This is not transferable to other EU Member States. Alternative solutions to connect farm-statistics with spatial data have to be explored. In general, the connection of farm statistics and spatial data at NUTS3 level could already be sufficient for the purposes of most analyses.

During the meeting the representativeness of NUTS3 variables in FADN was also discussed. From BESTMAP it was asked if you could recalculate the farm weights based on FSS/agricultural census data. Experience within the MINDSTEP team was, that this is probably not worth the effort. Also because FSS data is only available in time steps of four years. You can still use the NUTS3 data for your research.

2.2.2. Joint exploitation of data obtained in participatory research tasks

The three individual projects involve various participatory research tasks to close information gaps experienced when using solely FADN data or other datasets collected by third parties. In the case of AGRICORE, three different specific participatory research trajectories will start in early 2022 to determine production costs and innovation on small farms not represented in FADN. BESTMAP plans 125 interviews with farmers and MIND STEP plans surveys in the field of adoption of GHG mitigation measures and risk management instruments.

A crucial question is the extent to which such information can be shared between the AGRIMODELS projects. If the same restriction as for FADN data apply, data sharing will only be possible in aggregated form, e.g. shared data should be aggregates (e.g. averages) of at least 15 individual observations. The General Data Protection Regulation (GDPR) will apply for sure and thus exchange at individual farm level is not permitted. However, the exchange of methods (software) to process data without exchange of the data itself is highly encouraged.

2.2.3. Joint dissemination and communication actions

A proposal for an organised session titled "Use Of Individual Decision Making Models To Support Policies." during the 16th EAAE Congress 2021 was re-submitted by John Helming (MIND STEP), which will be a first opportunity for a joint dissemination activity.

Enhancing the collaboration within AGRIMODELS can also involve the application of new models and other tools to the same set of scenarios and a core set of indicators. For this, the MIND STEP deliverable "Report on policies and global drivers affecting IDM units in EU agriculture" could be relevant and will be shared with the partners. An extension to this form of intra-cluster alignment of model-based analyses will be to agree on a common baseline, for instance taken from models used in the SUPREMA projects.

2.2.4. Other topics

The AGRICORE project had planned an official Kick-off meeting of the cluster for the 12th project month (August 2020). This meeting could not be held in person due to COVID19 restrictions and was postponed. A possible new date would be the time around the first reporting period, which was originally moth 18. However, the MIND STEP project has asked for a general postponement of deliverables by 4 month in response to the COVID-19 situation and further 2 month due to the delay of FADN data access. As a consequence, the reporting period was also shifted for 6 month tpo

September 2021. From this perspective, a reasonable new date for the cluster Kick-off metting would be around this time, perhaps involving DG AGRI staff to share the experiences made so far. This will be clarified during the next cluster workshop.

The MIND STEP project, as the other projects, were asked to address certain ethical issues identified during the evaluation of the grant proposal. In the case of MIND STEP, these ethics requirements involved 3 deliverables:

- Involvement of human beings in research - Human H - Requirement No.1 (D 10.1)
- Protection of personal data - POPD - Requirement No.2 (D 10.2)
- Misuse of research findings - M - Requirement No. 3 (D 10.3)

Such ethics deliverables are a fairly new requirement for H2020 projects and there is not much experiences and examples from previous projects to be build upon. The MIND STEP coordinators approached their project officer for guidance and were informed that the Ethics Assessment in the GA should be addressed directly and concisely.

The next cluster workshop will be organized before the summer break in June.

3. FIRST EXPERIENCES AND NEXT STEPS

So far, the two AGRIMODELS cluster workshops have proven to be a useful forum for exchanges of the project coordinators regarding data processing and acquisition, as well as result dissemination. MIND STEP and BESTMAP have experienced several obstacles when applying for FADN data, which will be useful information for AGRICORE once they start this process. It has become clear that exchange of FADN data between projects, or the joint application for FADN data, will be not feasible. However, the exchange of methods and aggregated results will be possible, but requires clear definitions of the levels of aggregation and the required contents. The same applies for survey data collected by the projects themselves as exchange of farm-level information would most certainly violate the GDPR. Consequently, a common database for farm-level statistics will not be possible or at least require substantial amount of security measures and legal counselling, which renders such an effort highly impractical.

A more promising avenue are the development of common scenarios and possibly baselines. This will be further clarified once the respective IT services for data exchange are in place.

The upcoming EAAE congress will be a first opportunity for a joint public appearance during an organized session, which can be the starting point for further publications and dissemination activities.

Due to the positive experiences with the two cluster workshops so far, this format will continue every 3 to 4 months.

4. CONCLUSIONS

The teams of the three projects AGRICORE, BESTMAP, and MIND STEP, which form the AGRIMODELS cluster have met so far twice to exchange experiences with data acquisition and processing to devise a common data processing plan. An important conclusion was that the exchange of individual farm data, either from European or national FADN, will not be possible due to data protection regulations. Common data processing is therefore restricted to publicly available data, research findings at sufficiently highly aggregated levels, and scenario narratives, possibly including a common baseline. The exchange of data

processing methods and utilities will be possible, provided that the developed utilities can accommodate a sufficiently wide range of datasets. Furthermore, joint dissemination and publication activities were launched and will be continued for the remainders of the projects.

5. ACKNOWLEDGEMENTS

6. REFERENCES

Veneziani, Mario, Fedrico Antonioli, Giorgia Eranio, Carlos Leyva, Pablo Baez, Alvaro Fernandez (2020). D1.1 Standardised Methodology and Set of Ontologies for the Characterisation of Data Sources. H2020 project AGRICORE. https://agricore-project.eu/wp-content/uploads/2021/02/D1.1-Standardised-Methodology-and-Set-of-Ontologies-for-the-Characterisation-of-Data-Sources_V1.0.pdf