D5.3 - Agro-climatic and Economic Modelling Community-centred Assessment Plan

Co-funded by the Horizon 2020
Framework Programme of the European Union

DELIVERABLE NUMBER	D5.3
DELIVERABLE TITLE	Agro-climatic and Economic Modelling Community-centred Assessment Plan
RESPONSIBLE AUTHORS	Rob Lokers (Wageningen Environmental Research)

GRANT AGREEMENT N.	731001
PROJECT ACRONYM	AGINFRA PLUS
PROJECT FULL NAME	Accelerating user-driven e-infrastructure innovation in Food & Agriculture
STARTING DATE (DUR.)	01/01/2017 (36 months)
ENDING DATE	31/12/2019
PROJECT WEBSITE	http://www.plus.aginfra.eu
COORDINATOR	Nikos Manouselis
ADDRESS	110 Pentelis Str., Marousi GR15126, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Mrs. Georgia Tzenou
WORKPACKAGE N. | TITLE	WP5 | Agro-climatic & Economic Modelling Community
WORKPACKAGE LEADER	Agroknow
DELIVERABLE N. | TITLE	D5.3 | Agro-climatic and Economic Modelling Community-centred Assessment Plan
RESPONSIBLE AUTHOR	Rob Lokers (Wageningen Environmental Research)
REPLY TO	Rob.lokers@wur.nl
DOCUMENT URL	http://www.plus.aginfra.eu/sites/plus_deliverables/D5.3.pdf
DATE OF DELIVERY (CONTRACTUAL)	30 September 2017 (M9), 30 June 2018 (M18, 1st Updated version)
DATE OF DELIVERY (SUBMITTED)	18 October 2017 (M10), 29 June 2018 (M18, 1st Updated version)
VERSION | STATUS	2.0 | Final
NATURE	Report (R)
DISSEMINATION LEVEL	Public (PU)
AUTHORS PARTNERS	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
REVIEWER	Matthias Filter (BfR)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial Version	15-09-2017	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
0.2	Assessment description and planning	06-10-2017	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
0.3	Indicator definition and measurement	11-10-2017	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
0.9	Draft for review	12-10-2017	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
1.0	Review	13-10-2017	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
1.1	Final version	18-10-2017	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)
2.0	Updated version	29-06-2018	Rob Lokers (Wageningen Environmental Research), Rob Knapen (Wageningen Environmental Research)

PARTICIPANTS		CONTACT
Agro-Know IKE (Agroknow, Greece)	Agroknow	Nikos Manouselis Email: nikosm@agroknow.com
Stichting Wageningen Research (DLO, The Netherlands)	ALTERRA WAGENINGEN UR	Rob Lokers Email: rob.lokers@wur.nl
Institut National de la Recherché Agronomique (INRA, France)	INRA SCIENCE & IMPACT	Pascal Neveu Email: pascal.neveu@inra.fr
Bundesinstitut für Risikobewertung (BFR, Germany)	BfR Bundesinstitut für Risikobewertung	Matthias Filter Email: matthias.filter@bfr.bund.de
Consiglio Nazionale Delle Richerche (CNR, Italy)		Leonardo Candela Email: leonardo.candela@isti.cnr.it
University of Athens (UoA, Greece)	National and Kapodistrian University of Athens	George Kakaletris Email:gkakas@di.uoa.gr
Stichting EGI (EGI.eu, The Netherlands)	egi	Tiziana Ferrari Email: tiziana.ferrari@egi.eu
Pensoft Publishers Ltd (PENSOFT, Bulgaria)	PENSOFT	Lyubomir Penev Email: penev@pensoft.net

ACRONYMS LIST

Acronym	Description
FAIR	Findable, Accessible, Interoperable, Reusable, as set of principles acting as an international guideline for high quality data stewardship
RFC	Request for Change
VRE	Virtual Research Environment
WOFOST	World Food Studies (crop growth simulation model)
YAML	Yet Another Markup Language, a human-readable data serialization language.

EXECUTIVE SUMMARY

This Community-centered Assessment Plan describes the detailed plan regarding the procedures to be carried out for assessing the effectiveness of the AGINFRA PLUS paradigm for research in the agro-climatic modelling community. It defines the objectives of the pilot trials and their assessment and the specifics of the different phases defined in the piloting scheme. Starting from that, the assessment objectives for the selected use cases (crop modelling and crop phenology estimation) are outlined and the associated assessment indicators, the method of assessment per indicator and the planning of the assessment of indicators over the iterative piloting scheme are defined.

TABLE OF CONTENTS

1 INTRODUCTION 7

2 ASSESMENT METHODOLOGY 8

2.1 INTERACTION WITH COMMUNITIES 8

2.2 PILOTS TRIALS 9

2.3 PILOTS EVALUATION 10

3 USE CASE 1- CROP MODELLING 13

3.1 TARGET USERS 13

3.2 ASSESSMENT OBJECTIVES 13

3.3 EVALUATION 14

4 USE CASE 2 - CROP PHENOLOGY ESTIMATION 15

4.1 TARGET USERS 15

4.2 ASSESSMENT OBJECTIVES 15

4.3 EVALUATION 15

5 USE CASE 3 - AGRODATACUBE 16

5.1 TARGET USERS 16

5.2 ASSESSMENT OBJECTIVES 16

5.3 EVALUATION 16

ANNEX 1 – PERSONAS 17

LIST OF TABLES

Table 1: Evaluation indicators for agro-climatic modelling 12

1 INTRODUCTION

This document describes the community centered assessment plan for the pilot trials and the pilot application's to be developed, deployed and evaluated as part of the AGINFRA PLUS VRE pilots defined for the agro-climatic modelling community.

It describes the iterative phasing of the piloting scheme for the selected use cases to be developed for the community and the specific objectives for the pilots and the associated pilot assessments in the different phases. As the three piloting phases look upon the use case from different angles (e.g. different development stages of the pilot application, development perspective versus usage perspective), the assessment methodology and planning are aligned to these phases. In this plan, the specifics of every single phase of the pilot are defined and the targeted types of evaluators from the community that are required to assess the pilots in the specific stage are identified. Moreover, the plan describes the different levels of assessment that will be performed to evaluate the heterogeneous set of quantitative and qualitative indicators.

Finally, the plan also clarifies the assessment objectives for the two selected use cases: crop modelling and crop phenology estimation. From these objectives, a set of evaluation indicators per use case is derived that will provide a broad view on the effectiveness of the use of VRE's for these use cases. For every indicator, the level of assessment and the specific phase(s) in the overall piloting scheme of the AGINFRA PLUS project is specified.

2 ASSESMENT METHODOLOGY

This chapter describes the methodology for the assessment of the identified AGINFRA PLUS use cases targeted at the agro-climatic modelling community and its planning throughout the project. It defines how the project intends to expose the use case, how it interacts with the community in the various phases of the project through targeted assessment sessions and events and how the evaluators involved in each phase match the objectives of every phase. Moreover, it defines the setup of the trials and the associated evaluation procedures and metrics used to measure the effectiveness of the AGINFRA PLUS paradigm for research in the targeted research community.

2.1 INTERACTION WITH COMMUNITIES

To assess the effectiveness of the AGINFRA PLUS paradigm for research in the agro-climatic modelling community, an iterative approach of assessment, parallel to the proposed three-phase piloting scheme, will be performed. Following the concept of gradual extension of functionality, intended audience and assessment of this scheme, the pilot will interact with the agro-climatic modelling community and the pilot evaluators accordingly.

First Pilot (intermediate phase, M13-18)

- Available components that have been developed in the first phase of the piloting will be used and evaluated by an independent agro-climatic modeller employed at Wageningen Environmental Research to acquire first impressions regarding data accessibility and user experience. The main objectives are to (1) evaluate the main pilot components developed and deployed on the VRE and the usability of the VRE and (2) to collect issues and RFC's that feed into the further refinement of the components and the full prototype(s) to be developed in the next phase. The researcher selected for the evaluation process will be a person that is not involved in the development of modelling components and their deployment on the VRE.
- The profile of the researcher that will run these trials is that of a senior modeller, highly experiences in modelling and model development and with a broad knowledge of ICT. He/she will have capacities in modelling in the broad sense and experience in areas like the required data and data fusion for modelling, the configuration of models and the assessment of model output. Besides, he/she will have previous experience with running models in alternative cloud and high performance computing environments.

Second Pilot (summative phase, M22-M27)

- The fully-fledged version of the VRE based prototypes developed for the use case within AGINFRA PLUS will be in first instance trialed by a small group of researchers from the AGINFRA PLUS consortium. The principal objective will be to collect a first set of data and user experiences for the pilot evaluation. Moreover, it provides an excellent opportunity to thoroughly test the prototype and signal and resolve issues before deploying it to the broader agro-climatic modelling community in the third piloting trial.
- The researchers that will run these trials will ideally have a range of expertise levels, from junior modeller to expert-modeller, to account for a broad range of potential end users. While most obviously these will include modellers from Wageningen Environmental Research, we will extend the group with researchers from other project partners (e.g. INRA) to incorporate a broader and more independent perspective.

Third Pilot (summative phase, M30-M36)

- In the third pilot trial, the final version of the prototype(s) for the use case will be exposed to and trialed by the broader scientific community of agro-climatic modellers through a dedicated piloting event. The main objective is to complement the set of data and user experiences

collected in the second pilot in order to perform a full evaluation of the developed prototype and its effectiveness for end users.
- The audience will be a broader group of international modellers, where the concept will be to organize a side or back-to-back event with a major international conference or workshop that attracts the foreseen audience.

2.2 PILOTS TRIALS

First Pilot (intermediate phase, M13-18)

The first pilot trials will focus on the assessment of some of the major components being developed for the agro-climatic modelling use cases. They can be regarded as a form of "module testing", where the emphasis will be on the accessibility of the components from the VRE and the required connectivity for further integration of these modules into use case specific workflows to build a full workflow. It will also be a more generic assessment of the ease of use and usability of the VRE as an environment for effective collaborative modelling.

- The evaluation of the accessibility and connectivity (with regard to the foreseen prototype and its application) of data sources required for the use case on the VRE. This includes the assessment of how the FAIR principles are complied with through the VRE.
- The assessment of accessibility, usability of the main application components of the use cases.
- The global assessment of ease of use and expected learning curve of the VRE to develop and deploy modelling components like data access, models, data analytics.

In this phase only a very limited set of specific integration aspects, e.g. data analytics, data fusion, post-processing etc., will be considered. Components will be individually tested, e.g. data sources will be tested by querying them for typical (raw) data required for pre-processing, models and algorithms will be tested using available, ready-for-use datasets.

Where feasible, we will also have evaluators assess the ease of use of the VRE by having them develop a simple (proxy of) a model, with data feeds and data outputs. This might give them additional insights in potential use and opportunities of the D4Science VRE and could provide a broader evaluation perspective.

Second Pilot (summative phase, M22-M27)

As the results of the second piloting phase will be evaluated by a limited group of more experienced scientists, there will be opportunities for a more in-depth evaluation of specific aspects of VRE usage that either require larger efforts and/or a high level of expertise. These pilots will emphasize the assessment of complete scientific agro-modelling workflows on the D4Science VRE. We foresee that this evaluation will include both the assessment of a pre-configured modelling workflow, as well as evaluation from the model and workflow development perspective, where specific components could be configured in real-time by evaluators.

Besides the indicators that were already subject of the first evaluation, the following additional aspects will be assessed in the trials:

- Usability of specific workflow components, e.g. components for model run schematization, data analytics and visualization.
- Performance of components and pre-configured workflows.
- Scalability of workflows, through the assessment of a large-scale modelling experiment that mimics massive modelling exercises that are required e.g. in climate modelling or robust decision-making scenarios.
- Transparency of VRE components and workflows, e.g. repeatability of tasks, version control etc.

This phase seeks to assess the defined agro-climatic workflows in a more integrated manner, looking specifically at linkage between the input data, computing and analytics elements through data fusion, data wrangling, post processing etc. As far as feasible, in-depth experiments will be performed that require amounts of effort and expertise that cannot be expected to be broadly available in the 3rd pilot phase that will reach out to a larger and more diverse group of evaluators.

Third Pilot (summative phase, M30-M36)

The third pilot assessment will be performed by a group of international modellers, which is expected to be broader and more diverse than the evaluators that were involved in the previous two stages of the pilot. As AGINFRA PLUS will have less control on the composition of this final group of evaluators, the performed assessment will be limited to aspects that can be evaluated uniformly by a relatively heterogeneous group of scientists.

The following aspects will be assessed in the trials:

- Usability and performance of pre-configured workflows.
- The assessment of the accessibility, discoverability and connectivity of model outputs (with respect to integration with e.g. post-processing and visualization tools).
- Usability and adaptability of specific workflow components, e.g. components for data analytics and visualization.

This pilot phase will thus focus on indicators that can be more easily assessed in a limited time-frame and with the heterogeneity of expertise levels that can be expected in international community events. It will therefore focus in first instance on assessing the usability and effectiveness of pre-configured workflows and working with its output data, with some attention to perform specific associated analytics and visualization functions. Nevertheless, targeted evaluators will be asked if they are willing to perform a deeper and more time-consuming assessment of a use case, taking into account and complementing the evaluation data for some parts of the assessments that were defined for the 2nd pilot trial phase.

2.3 PILOTS EVALUATION

The evaluation of the pilots will be performed along a set of assessment indicators that are targeted to the specifics of that use case. Due to the difference in the scale, setup and involved evaluators in each of the piloting phases (as described in the previous section), only the indicators fitting the specific phase will be evaluated. Moreover, the way in which indicators and metrics will be measured will differ over the three evaluations. This is inevitable, because of the type and complexity of the components and workflows to be evaluated as well as because of the different profiles of the evaluators that are involved in every single phase.

The assessment of the effectiveness of a VRE for agro-climatic modelling has a broad range of objectives, which again leads to a broad and heterogeneous set of potential assessment indicators. Evaluations will vary from the assessment of expert functions required for developing and deploying components and setting up workflows, to the user experience of the VRE's "end users", researchers that are focused on the (re)use of existing modelling workflows in their daily research activities and that are also working with the input and analyzing the output data of the models. While for some indicators it is expected to be feasible to perform quantitative evaluations, the character of some of the other indicators, and of the group of evaluators involved, will result in qualitative evaluations.

To facilitate the overall evaluation process, we will perform different types of evaluations adapted to the specific indicators and the evaluators:

- *Quantitative benchmarking*: for a limited set of indicators, quantitative measurements can be performed, for example for the indicator performance by measuring performance in a "conventional" modelling environment compared to a VRE environment.
- *Qualitative scoring*: Quite some indicators will not be suitable for quantitative evaluation. In these cases, at least a qualitative evaluation, for example though ordinal scoring will be developed. We propose that in these cases also the option to provide an additional explanation in the form of a short expert review will be added. This will allow getting better insight in the motivation of the individual reviewer.
- *Expert review*: Both in the first and second phase of the piloting scheme, where more experienced evaluators will be involved, a set of more complex, technical and usability indicators will be included in the evaluations. For those indicators that cannot be easily benchmarked or where quantitative assessment alone does not provide sufficient insight, an expert review procedure will be developed. The expert review will allow experts to (1) explain the motivation for their quantitative scoring of indicators, (2) assess indicators, using their past experiences with non-VRE environments and (3) allow evaluators to give their opinion on VRE for modelling through an interview with open questions after an evaluation session.
- *End user survey*: The broader trials in the third phase of the piloting scheme, the associated character and heterogeneity of the group of evaluators, the more end user oriented setup of the trial and the limited time available in events will require a targeted approach. The indicators that are specifically evaluated by this group will be assessed through an end user survey that can be completed in a relatively short time, also by less experienced evaluators. The format will be an on-line survey.

The following table summarizes the indicators from which a selection will be assessed for each of the use cases. It also indicates the method of assessment and the phase(s) in the piloting scheme where the evaluation of these indicators will be performed.

Indicator	Examples	Assessment method	Phase
Ease of Use	How simple is the concept of a VRE to the user; how easy is D4Science to use? How much effort is needed to define or use workflows? How much effort does it take to develop workflow components?	Qualitative scoring Expert review End user survey	1, 2, 3
Learning Curve	How much time is needed to learn new concepts etc. before the VRE can be used?	Qualitative scoring Expert review	1, 2
Usefulness	How does using the VRE compare to current hardware and software in use? E.g. considering costs, functionality?	Qualitative scoring End user survey, Expert review	1, 2, 3
Performance, Scalability	How does performance of the VRE compare to current systems in use? How significant is increased performance? Is the VRE fast enough for day-to-day use? How flexible is the VRE in scaling up compute and storage?	Expert Survey, Benchmarking	2, (3)
Reliability	Is the VRE (and infrastructure) reliable enough for day-to-day use? How much downtime is acceptable?	Expert review	2
Openness	How easy it is to add new data and functionality to the VRE? How easy is it to share workflows, components, and data?	Qualitative scoring Expert review	1, 2
Transparency	How repeatable are workflows? Does the VRE have version control and for workflows, components, data?	Expert review	2, (3)
FAIR-ness	How does the VRE help in making research data and	Qualitative scoring	1, 2

	algorithms FAIR (Findable, Accessible, Interoperable, Reusable). What are advantages and disadvantages compared to e.g. current research environments and data management practices?		

Table 1: Evaluation indicators for agro-climatic modelling

A more detailed description of the case specific user audience and the objectives of the pilot assessment are provided in the next chapters.

3 USE CASE 1- CROP MODELLING

The use case "crop modelling" is defined in deliverable D5.1 (User-driven Requirements &Use Cases - Agro-climatic and Economic Modelling) as a use case where researchers perform large scale regional crop model simulations. Besides, it also includes some of the elements that have been specifically mentioned in the use case specified for information intermediaries and service providers, focusing on the accessibility, the data provisioning and data wrangling of data from the AgroDataCube towards use in agro-climatic models.

3.1 TARGET USERS

Associated persona's (see Annex 1) for this use case are: researcher, information intermediary & service provider

Typical users of this use case will be:

- Researchers in various scientific domains (e.g. agronomy, agro-economy, climate change) that use agro-climatic crop modelling as part of their research.
- Information intermediary & service providers (e.g. extension services, ICT service providers) that either use crop models or the output of crop models as a resource to provide added value services for farm advisory and farm management support.

3.2 ASSESSMENT OBJECTIVES

The principal objective of this use case is to deploy and implement a full crop modelling workflow on a virtual research environment and to be able to run it in an explorative way as well as operationally. Based on the foreseen data availability the focus will be on regional yield forecasting (specifically looking at simulating crop growth and yields at agricultural parcels in the Netherlands). To accomplish that, a set of functional, integrative and technical "sub-objectives" need to be accomplished:

- Deployment of a version of the WOFOST (World Food Studies) crop model on the VRE. There are several implementations of WOFOST, and the most suitable candidate at the moment is the recently developed Java implementation of this model.
- Findability and accessibility of required (raw) input data through the VRE. This includes access to data from the AgroDataCube (a PostGIS spatial/relational database), and sets of (crop) parameter files (currently in YAML format).
- Publication of generated output files (from model, analytics components etc.) through the VRE for reuse by third parties and for data analysis.
- Development of data integration functions to process the mentioned (raw) input data to usable data formats for the WOFOST crop model.
- Data analytics and data visualization options (both for input and output data), particularly focused on handling and analyzing spatial datasets. E.g. display simulated yields per crop per parcel as a geographic map.
- Explorative use (e.g. through the use of a Notebook environment such as Jupyter) of components and assembling them into documented workflows.
- Deployment of modelling workflows on the VRE so that crop model configurations can be operated by less experienced users.
- Performance – the option to access high performance hardware through the VRE so that crop model simulation runs can be executed quicker compared to regular workstations and laptops used by researchers.
- Scalability – the option to distribute crop modelling runs effortlessly over a variable number of computing nodes through the VRE. E.g. running model simulations for a certain crop for all relevant parcels in the Netherlands for multiple years in parallel and combining the results.

3.3 EVALUATION

The development of the evaluation procedures for the use case will, aligned with the iterative approach of the pilot trials, be developed in a phased way. The relevant indicators from the set described in section 3.3 and the methods of assessment in the different phases will then be further specified.

4 USE CASE 2 - CROP PHENOLOGY ESTIMATION

The use case "crop phenology estimation" is defined in deliverable D5.1 (User-driven Requirements &Use Cases - Agro-climatic and Economic Modelling) as the use case targeting business analysts working on crop phenology estimation, combining various heterogeneous data sources to generate policy advice and decision support. Moreover, it includes elements that have been specifically mentioned in the use case for information intermediaries and service providers, focusing on the accessibility, the data provisioning and data wrangling of data from the AgroDataCube for data analytics required to provide farm advisory services.

4.1 TARGET USERS

Associated persona's (see Annex 1) for this use case are: business analyst, information intermediary & service provider

Typical users of this use case will be:

- Analysts that aim at combining various data sources like satellite data, statistics etc. to generate policy advice and decision support.
- Intermediaries that either use crop models or the output of crop models as a resource to provide added value services for farm advisory and farm management support.

4.2 ASSESSMENT OBJECTIVES

The principal objective for the use case is to deploy the AgroDataCube on a VRE, to make its data accessible and reusable, for instance to use the data to determine crop phenology development. This implies that a set of functional, integrative and technical "sub-objectives" are accomplished:

- Making the content of the AgroDataCube(a PostGIS spatial/relational database, containing agro-climatic data for the Netherlands) available (findable and accessible) on the VRE for further processing.
- Using the VRE for the explorative definition of algorithms (e.g. using a Notebook environment such as Jupyter) for assessing crop phenology development.
- Using the VRE to define workflows for assessing crop phenology development based on data from the AgroDataCube,e.g. for different crops and all agricultural parcels in the Netherlands.
- Using the VRE to train machine learning algorithms using crop phenology development and other input data from the AgroDataCube, and subsequently use the trained model to predict yield per crop per parcel, based on forecasted weather.
- Performing data analytics and visualization for e.g. (spatial) input data, the crop phenology, derived crop growth curves, yield forecasting. Preferably in an easy to use and customizable dashboard.
- Performance – the option to access high performance hardware through the VRE so that algorithms can be executed quicker compared to regular workstations and laptops used by researchers.
- Scalability – the option to distribute algorithm calculations effortlessly over a variable number of computing nodes through the VRE. E.g. running machine learning algorithms for yield prediction fora certain crop for all relevant parcels in the Netherlands in parallel and combining the results.

4.3 EVALUATION

The development of the evaluation procedures for the use case will, aligned with the iterative approach of the pilot trials, be developed in a phased way. The relevant indicators from the set described in section 3.3 and the methods of assessment in the different phases will then be further specified.

5 USE CASE 3 - AGRODATACUBE

The use case AgroDataCube concerns the integration of a (Dutch) regional data source for agronomic research and farmer support services into the VRE to facilitate various agronomic research applications, among others the two use cases described previously. Although defined here as a separate use case (realization of the case will be able to facilitate a range of similar use cases), in practice this case will be implemented as part of the development of the functionality for regional crop modelling and crop phenology estimation. This implements a range of functions that are relevant for applications of all types of users (researchers, information intermediaries and service providers, business analysts) in the area of data accessibility, data provisioning and data wrangling for data analytics.

5.1 TARGET USERS

Associated persona's (see Annex 1) for this use case are: researcher, business analyst, information intermediary & service provider.

Typical users of this use case will be:

- Researchers that are collecting, preparing and pre processing data feeds for modelling runs.
- Analysts that aim at combining various data sources like satellite data, statistics etc. to generate policy advice and decision support.
- Intermediaries that either use crop models or the output of crop models as a resource to provide added value services for farm advisory and farm management support.

5.2 ASSESSMENT OBJECTIVES

The principal objective for the use case is enable using a resource like theAgroDataCubefrom a VRE and to make its data findable, accessible, interoperable and reusable (FAIR). This implies that a set of functional, integrative and technical "sub-objectives" are accomplished:

- Making the content of the AgroDataCube (currently a PostGIS spatial/relational database, containing agro-climatic data for the Netherlands) available (findable and accessible) on the VRE for further processing.
- Using the VRE to implement a range of pre processing, data analytics, visualisation and similar jobs that use (among others) the AgroDataCube data services.
- When needed implementation of the AgroDataCube on a scalable infrastructure, which could be the D4Science data infrastructure. This should allow resource demanding applications, like mass scale modelling, to query the data without resource capacity or performance issues.

5.3 EVALUATION

The development of the evaluation procedures for the use case will, aligned with the iterative approach of the pilot trials, be developed in a phased way. The relevant indicators from the set described in section 3.3 and the methods of assessment in the different phases will then be further specified.

ANNEX 1 – PERSONAS

Persona 1 - Researcher
The researcher's work is focussed at scientific and applied scientific definition and application of models. These could either aim at developing and applying innovative models and model applications or at configuring and applying existing models for innovative research on societal challenges like food security or climate change.

Alain Duvall - Senior Agronomic Researcher

"*I would like to find ways for me and my team to work more efficient and to **make it easier to share and reuse data and analytics***"

Name: Alain
Position: senior-researcher at an agronomic research organisation
Age: 46
Education: University, PhD
Location: France
Archetype: Researcher

Biography
Alain has a PhD in Agronomy. He has worked as a researcher in France and the US for an NGO and several research organisations and he's been involved in developing simulation models and performing impact assessments for years. Since a few years he is head of a small research team. Besides, he works on a strategy to better exploit big data and big data analytics in his domain.

Daily Tasks
- Operational management of a small team of researchers and ICT professionals specialized in agro-climatic modelling and impact assessment.
- Coaching his team members in their daily work.
- Setting out directions for future research in the agronomic domain, with a focus on big data and data analytics.
- Scientific publication.
- Data curation.

Motivations
- Doing research that is being used and has impact.
- Progressing his scientific career.

Goals
- To provide good quality scientific outputs
- To extend his team's data analytics capacities
- To produce results that are fit-for-use for policymakers

Frustrations:
- Why is it still so hard to find the data I need for my research?
- All my team members are developing their own data analytics, there's too much redundancy and too little reuse.
- I cannot easily publish my data in a citable manner.

Persona 2 - Information intermediary / service provider

Information intermediaries and service providers usually aim to provide added value, like farm management advice, through the(re)use of existing data and (meta)models to support end users (for instance policy and decision makers or farmers). They aim at providing operational services, using trusted datasets and proven technologies.

Kenneth Brown - ICT developer

*"Despite all efforts it is still hard to **find trusted data and process it efficiently for modelling**"*

Name: Kenneth
Position: programmer and data-analyst at AgroTalk, a small ICT company
Age: 28
Education: University, BSc
Location: Uganda
Archetype: Developer

Biography

Kenneth has finished his BSc in informatics in Uganda and has just joined AgroTalk. AgroTalk is a small ICT company that specializes in providing advice to farmers through among others mobile phone and smartphone apps. With his six colleagues Kenneth is developing a service that provides farm management advice to smallholders through SMS services and through a farm management support system that is used by extensionists in the field. They are providing crop advice using a crop growth model that uses among others weather data. In the future, they plan to also develop a smartphone app for this purpose.

Daily Tasks

- Co-developing a software system that performs automated daily crop model runs and turns the results into advice.
- Manual and semi-automatic analysis of data sources like weather data, crop development data, soil data etc., to provide advice to farmers.
- Liaising with two colleagues, who are agronomists and advise him on his data analysis work and on the implementation of agronomic algorithms for the new IT system.

Motivations

- Programming reliable, robust software for smallholders.
- Using new technologies, with a special interest in big data and data analytics.

Goals

- To deliver a high-quality ICT solution that has added value and can be charged for.

Frustrations:

- The data I need to turn agronomy knowledge into working solutions is hard to find and get, e.g. weather data is available, but often too expensive, and the data I can get is not always of good quality.
- Data is provided in so many formats that are not compatible and there's no software to easily process the data.

Persona 3 - Business Analyst

The business analyst is mainly interested in re(using) existing data sources and modelling outputs for commercial application. Some well-known examples are the use of combination of weather, soil and agronomic data for farm management or the use of yield forecasts and weather data for risk assessment for re-insurance activities.

John Jackson - Policy advisor and analyst

"I know there is scientific work on agricultural data analytics but I do not have access to scientific resources other than papers"

Name: John
Position: Data analyst at the Ghana Ministry of Agriculture
Age: 36
Education: University, BSc
Location: Ghana
Archetype: Data Scientist

Biography

John has been working at the Ministry of Agriculture for 12 years now. He started as a policy officer, but he has always had interest in data analysis and three years ago he has been transferred to the new department for data analytics and data science. He is producing regional crop bulletins, using data collected in the different agricultural regions of Ghana and data provided through the national weather service. The results of his work are published through the open data platform that the government has just set up. Currently he is exploring the opportunities of satellite derived NDVI data that his department has been provided access to recently.

Daily Tasks

- Collecting the data for his work from different sources.
- Analysing, improving and post-processing the crop and weather data as input for his data analysis algorithms and specifically exploring the options of NDVI data streams.
- Publication of his output data on the open data platform and curation of the section on agricultural data.

Motivations

- Working on this new field of data science, using new technologies and analytics.
- Promoting open data and the use of open data by farmers and business.

Goals

- Make his work easier by improving the data flows.
- Operationalize the analysis of NDVI time series
- Becoming a senior data scientist.

Frustrations:

- Data is very scattered and it takes too much effort to collect and integrate it.
- I do not have access to scientific resources for my data analytics.
- Even though we are providing open data, I see that it is hardly used.