D5.4 - Agro-climatic and Economic Modelling Pilot Evaluation Report

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

DELIVERABLE NUMBER	D5.4
DELIVERABLE TITLE	Agro-climatic and Economic Modelling Pilot Evaluation Report
RESPONSIBLE AUTHOR	Rob Lokers (Wageningen Environmental Research), Rob Knapen (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	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	Wageningen Environmental Research
DELIVERABLE N. | TITLE	D5.4 |Agro-climatic and Economic Modelling Pilot Evaluation Report
RESPONSIBLE AUTHOR	Rob Lokers, Rob Knapen (Wageningen Environmental Research)
REPLY TO	rob.lokers@wur.nl
DOCUMENT URL	http://www.plus.aginfra.eu/sites/plus_deliverables/D5.4.pdf
DATE OF DELIVERY (CONTRACTUAL)	30 June 2018 (M18)
DATE OF DELIVERY (SUBMITTED)	12 July 2018 (M19)
VERSION | STATUS	V1.0 | Final
NATURE	ORDP (Open Research Data Pilot)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Rob Lokers, Rob Knapen (Wageningen Environmental Research)
REVIEWER	Matthias Filter (BfR)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Initial Version	14/5/2018	Rob Lokers (Wageningen Environmental Research)
0.2	Assessment description	17/05/2018	Rob Lokers (Wageningen Environmental Research)
0.3	Metrics and measurement process definition	18/06/2018	Rob Lokers (Wageningen Environmental Research)
0.4	internal review	20/06/2018	Rob Lokers, Rob Knapen (Wageningen Environmental Research)
0.5	Draft for review	21/06/2018	Rob Lokers (Wageningen Environmental Research)
0.6	Review	29/06/2018	Matthias Filter (BfR)
0.9	Evaluation results added	11/07/2018	Rob Lokers, Rob Knapen (Wageningen Environmental Research)
1.0	Final version	12/07/2018	Rob Lokers, 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
VRE	Virtual Research Environment
WOFOST	World Food Studies (crop growth simulation model)

EXECUTIVE SUMMARY

This document describes the evaluation indicators and evaluation procedure of the first AGINFRA PLUS evaluation of the VRE pilot developed for the agro-climatic modelling community and reports on the outcomes of the evaluation.

An evaluation procedure was developed that is specifically aimed at the group of users that are most likely to use the components that are currently available, but also looks forward to the next evaluation rounds, where a broader group of less technically experienced users will be addressed. The evaluation procedure has also taken into account the current state of development of both the capacities of the D4Science environment and the specific VRE set up to support the agro-climatic use cases and the associated modelling components developed on these VRE functions.

The first pilot evaluation show that the VRE concept is seen by the evaluators as a very useful for future collaborative research in the area of agro-climatic modelling, especially where it concerns international collaborations that require sharing and reuse of resources. It also reveals that the current implementation of the agro-climatic VRE on the D4Science platform has substantial shortcomings when it comes to real-world applications of agro-climatic modelling. These reside in the limited functionality that is currently available for essential analysis and visualization components, and in the lack of connectivity between individual VRE components. The evaluation results provide useful comments and recommendations for implementation towards better support of the modelling process.

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

This document describes the evaluation indicators and evaluation procedure of the first AGINFRA PLUS evaluation of the VRE pilot developed for the agro-climatic modelling community and reports on the outcomes of the evaluation. The first evaluation phase focuses on the evaluation of development, deployment and application of VRE modelling components.

The report starts with a short explanation of the pilot evaluation's main objectives, considering the specific scope that was defined in the community centered assessment plan (deliverable D5.3) for the first pilot phase. This section also recaptures the indicators that were defined as particularly relevant with regards to these objectives. The assessment plan was defined in an early phase of the project, with little sight on the functional and technical features of the VRE that could be exploited and the directions and amount of effort required for component development. Therefore, a section of this report is dedicated to further specification of the still rather global indicator definitions towards more directed evaluation metrics. This further specification takes into account the current state of progress of both the development of the VRE and its features and the agro-climatic modelling components required to setup the workflows for the use cases in a later phase of the project.

Using the defined metrics as the basis, an evaluation procedure has been developed that is specifically aimed at the envisaged groups of VRE users from the community active in this phase of piloting. It also takes into account the current state of development of both the capacities of the D4Science environment and the specific VRE set up to support the agro-climatic use cases and the associated components developed on these VRE functions.

Finally, the document provides a summary of the evaluation results collected through the performed pilot evaluation. The full outcomes of the evaluation per individual evaluator are provided as an Annex to this report.

2 PILOT EVALUATION – OBJECTIVES AND PROCESS

To assess the effectiveness of the AGINFRA PLUS paradigm of using Virtual Research Environments for research in the agro-climatic modelling community, an iterative approach of assessment and evaluation, in parallel with the proposed three-phase piloting scheme, will be performed. This chapter describes the objectives, the defined indicators and the evaluation process for the evaluation of the first piloting phase the agro-climatic modelling use cases developed in that phase the AGINFRA PLUS project.

The main objective of the VRE development process in the first pilot phase has been to get familiar with the VRE environment and its functions, to develop individual components required in the defined use cases, and to develop the associated VRE capacities for running these components on the D4Science VRE. In the second and third piloting phase, these components will be further extended and integrated into virtual research workflows that can support researchers in their modelling work.

The "Community Centered Assessment Plan" (AGINFRA PLUS deliverable D5.3) describes that the first pilot trials focus on the assessment of some of the major components being developed for the agro-climatic modeling use cases. They can in that sense 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. Consequently, the following evaluation objectives were defined for the first pilot evaluation:

- Available components in the first stage of the pilot will be used and evaluated by an independent agro-climatic modeler employed at Wageningen Environmental Research to acquire first impressions regarding data accessibility and user experience to (1) evaluate the main pilot components developed and deployed on the VRE and (2) to collect issues and RFC's that feed into the further refinement of the components and the full prototype 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 researcher that will run these trials will have capacities in modelling in the broad sense as well as a substantial background in ICT technology related aspects of model development and modeling. He will have experience in areas like the required data and data fusion for modelling, the configuration of models and the assessment of model output. Besides, he will have previous experience with running models in cloud environments.
- Moreover, some indicators that are associated with the development of modelling components rather than with configuring and executing model runs will be evaluated by software developers.

In this phase no specific workflow and integration aspects, e.g. of 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. Thus, the first evaluation phase focuses on the evaluation of modelling components and how they are developed for and deployed on the VRE. As such, it emphasizes the main components for the use case on regional crop modelling and specific elements of the use case on crop phenology. As part of the components of these two use cases it also integrates specific aspects of the third use case, AgroDataCube, and specifically the data access and preprocessing functions.

The assessment plan also provides a list of qualitative and quantitative evaluation indicators, the method of assessment (e.g. quantitative benchmarking, expert review, end user survey) and the piloting phase in which these indicators are particularly relevant. The table below summarizes the indicators that were identified in the assessment plan for the agro-climatic modeling pilots. The indicators printed in black are associated with the first pilot evaluation and will be further detailed in the next chapters of this document.

Some indicators that are particularly relevant for more advanced VRE applications and that will only be evaluated in the next piloting phases are grayed out.

Table 1 – List of evaluation indicators for the Agro-climatic VRE use cases

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?	End user survey	1, 2, 3
Learning Curve	How much time is needed to learn new concepts etc. before the VRE can be used?	Expert review	1, 2
Usefulness	How does using the VRE compare to current hardware and software in use? E.g. considering costs, functionality?	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?	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 algorithms FAIR (Findable, Accessible, Interoperable, Reusable). What are advantages and disadvantages compared to e.g. current research environments and data management practices?	Expert review	1, 2

3 EVALUATION INDICATORS

With the defined scope of the first pilot evaluation, the selected pilot evaluation indicators mainly take into account the perspective of the model (software) developer and the more experienced and ICT knowledgeable type of modeler. Their roles would respectively be (1) the development of new or redesign of existing models able to run on a VRE and to exploit the advantages of a VRE with respect to e.g. cloud computing, high performance computing and collaborative development and testing and (2) the deployment, configuration and running of such components on the VRE and the development of new and application of existing data analytics and visualization features.

In the remainder of this chapter, the indicators defined in the assessment plan are further specified to a form that makes them feasible for integration into the upcoming evaluation.

3.1 INDICATOR - EASE OF USE

This indicator was defined in the assessment plan by indicative questions like:

- 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?

The indicator will be assessed through the following metrics:

Table 2 – Evaluation metrics for the indicator "Ease of Use"

Metric	Type of assessment	Means of assessment	Evaluator
Ease of use of the VRE for model development	Qualitative scoring Expert review	Scoring the individual model development features of the VRE + open questions	Software developer
Ease of use of the VRE for modelling	Qualitative scoring Expert review	Scoring the modeling configuration and execution features of the VRE + open questions	Senior modeler
Ease of use of the VRE for analysis and visualization	Qualitative scoring Expert review	Scoring the modeling analysis and visualization features of the VRE + open questions	Senior modeler
Ease of use of the VRE for collaborative model development	Qualitative scoring Expert review	Scoring the collaboration features of the VRE+ open questions	Software developer

3.2 LEARNING CURVE

This indicator was introduced to assess how much time is needed to learn new concepts before the VRE can be effectively used.

Table 3 – Evaluation metrics for the indicator "Learning curve"

Metric	Type of assessment	Way of measuring	Evaluator
Estimated effort to get develop a model	Qualitative scoring Expert review	Scoring on an ordinal scale the estimated learning curve against the "conventional" environment and open questions	Software developer
Estimated effort to configure and run a model and analyze its outputs	Qualitative scoring Expert review	Scoring on an ordinal scale the estimated learning curve against the "conventional" environment and open questions	Senior modeler

3.3 USEFULNESS

This indicator compares the VRE and its usability, e.g. the comparison to current hardware and software in use. Characteristics that could be compared are for instance functionality and costs.

Table 4 – Evaluation metrics for the indicator "Usefulness"

Metric	Type of assessment	Way of measuring	Evaluator
Usefulness of the VRE for - Model runs (single or multiple)	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Senior modeler
Usefulness of the VRE for - Data analysis	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Senior modeler
Usefulness of the VRE for - Visualization	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Senior modeler
Usefulness of the VRE for - Collaborative modeling exercises	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Software developer

3.4 OPENNESS

Openness should be considered an important asset of any environment in the light of the European Open Science Cloud. To assess openness, the assessment plan states that ease of adding new data and functionality to the VRE or to share workflows, components, and data with the outside world could be important issues.

Table 5 – Evaluation metrics for the indicator "Openness"

Metric	Type of assessment	Way of measuring	Evaluator
Ease of accessing models and data registered in the VRE	Qualitative scoring	Scoring on an ordinal scale	Senior modeler
Ease of registering and sharing of datasets	Qualitative scoring	Scoring on an ordinal scale	Software developer, Senior modeler
Ease of adding and sharing new functionality and components (e.g. model, visualization,	Qualitative scoring	Scoring on an ordinal scale	Software developer, Senior modeler

3.5 FAIR-NESS

The FAIR principles (Findability, Accessibility, Interoperability, Reusability) are key principles for open science on any VRE. This indicator should therefore assess how the VRE helps in making research data and algorithms FAIR and what the advantages and disadvantages are compared to current research environments and data management practices.

Table 6 – Evaluation metrics for the indicator "FAIR-ness"

Metric	Type of assessment	Way of measuring	Evaluator
Findability	Qualitative scoring	Scoring on an ordinal scale	Senior modeler
Accessibility	Qualitative scoring	Scoring on an ordinal scale	Senior modeler
Interoperability	Qualitative scoring	Scoring on an ordinal scale	Software developer, Senior modeler

Reusability	Qualitative scoring	Scoring on an ordinal scale	Software developer, Senior modeler

4 EVALUATION SCRIPT

4.1 INTRODUCTION

This chapter describes the detailed evaluation procedure that has been used for the first pilot evaluation. The evaluation starts with a general introduction to the principles and structure of the VRE and the available modelling, analysis, visualization and collaboration tools in the agro-climatic modelling VRE. As this evaluation focuses on functions for model developers and technically knowledgeable senior modelers, this part is rather extensive. The second part on VRE model development is held concise, as in this particular case, the evaluator has worked as a developer on the VRE and the section is mainly aimed at introducing the senior modeler to the various aspects of model development.

The next section describes the exercises that have been set up for the senior modelers to work with the VRE and to get some first experiences on what it means to configure and run models and to analyze and visualize results as a virtual researcher.

The last section of this chapter describes the actual evaluation survey that was provided to the evaluators after the VRE introduction and after having performed the evaluation exercises on the agro-climatic modeling VRE.

4.2 EVALUATION PROGRAMME

4.2.1 Introduction to the VRE

Setup of the evaluation

This first pilot evaluation will evaluate some of the more complex aspects of developing and deploying VRE components. Therefore, the evaluation will be setup as a workshop. Evaluators will get a short introduction into the VRE and its functions and will then independently work through the evaluation sections. An experiences VRE developer will however be available to guide and support the evaluators if needed.

General principles

The general principles of a virtual research environment are explained to the evaluators by a VRE expert:

- What is a Virtual Research Environment?
- Relation with Open Science, FAIR principles

Structure of the VRE

A VRE is not only meant as a research environment to develop and run your software in the cloud, but also to facilitate collaboration between researchers and research groups. To support that, the VRE has a range of functions available.

- Collaboration functions – to facilitate communication, cooperation and co-development of (distributed) teams.
- Development functions – to develop operational components, e.g. developing executable models, constructing workflows
- Operational functions – to “perform research”, in the case of the agro-climatic modelling VRE to be able to run models and analyze their outputs.

The D4Science environment allows the setup of dedicated VRE's and provides user registration and management functions, so users can be granted access and rights to specific VRE's and VRE functions. For the AGINFRA+ use case for agro-climatic modelling, a dedicated VRE was created.

Collaboration functions

The VRE provides a range of collaboration functions, the most relevant ones being:

- Shared workspaces - These allow the creation of a folder structure on the VRE and the sharing of folders and files.
- Mail - the VRE offers a mailing environment. Users can exchange mail messages (as in common mail applications) among members of a VRE.
- Newsfeeds: News feeds are another way for VRE communities to communicate, acting more as a message forum-like function, with options to comment and address VRE users.
- Activity tracking: is the way to register and track (development) issues and to collaborate/ communicate on progress regarding development of VRE components, applications and workflows.
- Data catalogue: allows the publication of dataset metadata to make datasets discoverable.

4.2.2 Model development and model execution components

Modelling environments

- Data Miner:
    - Explain the features of the Data Miner tool
        - Shared data space
        - Experiment execution
        - Computation status
    - Show how the WOFOST model has been integrated and explain the steps that were followed to accomplish this.
- Jupyter notebooks
- R Studio

Data analysis and data visualization

- Graphs
    - Viewing existing graphs
    - Creating a graph
- Geo-explorer
- Geo-analytics

4.2.3 Developing for the VRE

- What prerequisites before a model can run on the VRE
- Adaptations to the (WOFOST) model
- Preparing and accessing data sources, specifically the AgroDataCube
- Configuring the model

4.2.4 Hands on exercise – deploying and running the WOFOST model

Collaboration

- Set up communication with fellow modellers
    - Through the mail facility

- Through the news feeds
- Create a folder in the VRE, upload file and open file.

Running the WOFOST model

- Determine the dataset(s) to be used as input data
- Start a run
- Check progress of the model run
- Find and analyse the output datasets
- Visualizing output

VRE - Hello world exercise

- Create a small Python "model", using the Jupyter notebooks environment
  - Read from a file in the shared workspace, or read some data from AgroDataCube
  - Implement some simple processing of the data
  - Write results to the shared workspace

Adding your own model to the VRE

- Use SAI to import your 'Hello World' model into DataMiner
- Use DataMiner to execute the model
- Find and view the output

4.2.5 User evaluation survey

- Filling in the user survey
- Short interview of the evaluator to collect further experiences and comments.

4.3 EVALUATION SURVEY

This section describes the setup of the short survey that will be used in the evaluation to collect structured feedback from the evaluators. It addresses the previously mentioned indicators and allows to score them and provide additional open feedback per indicator.

4.3.1 Indicator - Ease of use

The indicator "ease of use" will assess how clear and easy to use the D4Science environment and the agro-climatic modelling VRE are to different types of users. It tries to answer the following types of questions:

- *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 modelling components and workflows?*

How do you score the ease of use of the VRE from the following perspectives?

Collaboration and communication

1- insufficient	2	3- sufficient	4	5 - very good	Not applicable

Model development

insufficient	moderate	sufficient	good	very good	Not applicable

Further explanation and comments:

Model execution

insufficient	moderate	sufficient	good	very good	Not applicable

Further explanation and comments:

Data analysis and visualization

insufficient	moderate	sufficient	good	very good	Not applicable

Further explanation and comments:

Further explanation and comments regarding ease of use in general:

4.3.2 Learning Curve

This indicator was introduced to assess how much time evaluators think will be required to learn new concepts and understand the working of VRE tools and their integration before the VRE can be effectively used.

How do you score the learning curve of the VRE from the following perspectives?

Estimated effort to make an existing model run on the VRE

1 - very low		3 - moderate		5 - very high	Not applicable

Can you provide a rough estimate in hours, day or months you think it would take to adapt an existing model to run on the VRE?

Estimated effort to configure and run a model and analyze its outputs

1 - very low		3 - moderate		5 - very high	Not applicable

Can you provide a rough estimate in hours, day or months you think it would take to perform a single model run and do a quick analysis of its outputs?

Further explanation and comments regarding learning curve:

4.3.3 Usefulness

This indicator compares the VRE and its usability, e.g. compare to current hardware and software in use. Characteristics that could be compared are for instance functionality and costs.

How useful do you think that the current agro-climatic VRE and its tools are from the following perspectives?

Usefulness for model execution

1 - not useful	2	3	4	5 - very useful	Not applicable

Usefulness for data analysis

1 - not useful	2	3	4	5 - very useful	Not applicable

Usefulness for visualization

1 - not useful	2	3	4	5 - very useful	Not applicable

Collaborative modelling

1 - not useful	2	3	4	5 - very useful	Not applicable

Further explanation and comments regarding usefulness of the VRE for modelling:

4.3.4 Openness

Openness should be considered an important asset of any environment in the light of the European Open Science Cloud. To assess openness, the assessment plan states that ease of adding new data and functionality to the VRE or to share workflows, components, and data with the outside world could be important issues.

How open do you think the agro-climatic VRE and its tools are from the following perspectives?

Ease of accessing models and data registered in the VRE

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable

Ease of registering and sharing of datasets

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable

Ease of adding and sharing new functionality and components (e.g. model, visualization)

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable

Further explanation and comments regarding openness:

4.3.5 FAIR-ness

The FAIR principles (Findability, Accessibility, Interoperability, Reusability, https://www.go-fair.org/fair-principles/) are key principles for open science on any VRE. Here you are asked to assess how the VRE

helps in making research data and algorithms FAIR and what the advantages and disadvantages are compared to current research environments and data management practices.

How do you think the current version of the VRE complies with the following FAIR principles?

Findability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable

Accessibility

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable

Interoperability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable

Reusability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable

4.3.6 Interview

The evaluation will be completed by having a short interview with the evaluator(s). They will have the option to comment on their general experiences of working with the VRE. Although the interview will not have a fixed format, the following questions could be guidance:

- What is your general impression of the VRE and its usability after doing the evaluation tasks?
- Do you think the current guidance and documentation available with the VRE is sufficient to independently perform some of your research tasks?
- What do you think are the advantages and disadvantages of using a VRE in comparison with doing the work on your own computer or the closed environment of your organisation's infrastructure?
- If applicable: How do you compare the current usability of the VRE compared to cloud-based environments you are currently working with?

5 EVALUATION SUMMARY

5.1 INTRODUCTION

The VRE for agro-climatic modelling at the time of this first pilot evaluation does not yet (as foreseen in the assessment plan) support the deployment of full modelling workflows. It supports the hosting of specific modelling components, but full and transparent integration of these components is not yet supported. Therefore, the evaluation concentrates on the evaluation of the capabilities of the VRE to support the development, deployment and execution of these components. Nevertheless, evaluators have also assessed the status of the current options to connect components, as this will in our view lead to clear suggestions for improvement of integration issues that are critical for successful evaluation in the following evaluation cycles planned.

In the first pilot evaluation, the agro-climatic modelling VRE was evaluated by two evaluators. The evaluators had the following profiles:

Evaluator 1: A senior modeler, with a broad experience in the development of agro-climatic models and the application of models in large research projects in a European and global context.

Evaluator 2: A senior software engineer, with broad experience in programming agro-climatic models and one year of experience of adapting existing modelling software and deploying and running these on the agro-climatic modelling VRE.

The sections below provide a short summary of the outcomes of the pilot evaluation for the different parts of the evaluation procedure. It also provides some general conclusions and recommendations, as well as a short list of comments, remarks and suggestions that were provided by the evaluators while performing the evaluation.

5.2 SUMMARY OF EVALUATION OUTCOMES

5.2.1 Introduction to the VRE and its components

The evaluation started with a comprehensive introduction to the D4Science VRE and the Agro-climatic modelling VRE and the available collaboration, model development and model execution capabilities. This included among others, a quick run through all VRE components and a short demonstration of the configuration and execution of an existing model and the various visualisation features.

In general, the features for collaboration were appreciated, and are acknowledged as a major asset of the D4Science VRE. Evaluators could see the benefits of a shared working environment that allows to share and reuse resources (also broader scientific resources from European research). This would definitely make specific collaborative tasks more efficient.

At the same time, when going through the existing components available in the agro-climatic modelling VRE, they stated that there are still substantial flaws in the current implementation. These concern mostly (1) the usability and user-friendliness of some of the current components, lacking required minimal functionality or having a cumbersome interface or implementation and (2) the lack of integration between the offered components and (3) specifically the lack of functionality available for data analysis and visualisation (where researchers seem to prefer to use existing solutions, instead of the VRE functions).

5.2.2 Running the existing (WOFOST) agro-climatic model

In the exercise of running a pre-configured simulation model in the VRE, the evaluators did not encounter specific problems. The configuration of the model, including the use of model input files from the shared workspace are transparent. The model executes without problems and output is generated at the expected location and is accessible.

Evaluator 1 (senior modeller) concludes that the current organisation of the input and output files in one folder, and the lack of options to structure this will be a barrier for serious modelling exercises. More in general he expects more guidance and examples on how to configure such “real-world” modelling exercises.

Evaluator 2 (model developer) remarks that the preparation of the model, deployment and configuration on the VRE is rather time-consuming, and adds substantially to the time required to setup a modelling exercise for any research project and that the SAI interface available to configure the model is not straightforward. Moreover, he notes that available documentation is fragmented, not easy to find and sometimes outdated or insufficient to complete tasks.

5.2.3 Development and execution of a simple model

The last part of the evaluation exercises was the development, deployment and execution of a small (Python) model on the agro-climatic modelling VRE. The pilot simulation model that was implemented was a simple Penmann-Monteith algorithm that calculates the net evapotranspiration from a set of basic weather parameters from a weather data file available at the shared workspace of the VRE and writes the results to an output file.

Despite the relative complexity of the VRE and its features, we observed that Evaluator 1, although he had no previous experience in working with the VRE, was able to independently develop a simple model, using input data from and writing output data to the shared workspace. After learning the specifics of accessing the VRE’s shared workspace, it appeared relatively straightforward to implement a small model using the Jupyter/Python environment of the VRE. Unfortunately, when deploying this model on Data Miner, the evaluator ran into compatibility issues, because the Data Miner used an outdated (v 2.7.6) version of Python under the hood, that doesn’t allow to run the model developed under de Jupyter notebook component. This reveals again that currently not all components are aligned, connected and interoperable with respect to the versions of underlying OS, software, libraries etc.

5.2.4 Interviews

The interviews with the evaluators revealed some additional opinions on working with the VRE and suggestions for improvement of its functions.

Both interviewees acknowledge the added value of virtual research environments as a useful concept to advance collaborative research, co-development and the sharing and reuse of scientific resources.

With regard to the comparison of the agro-climatic VRE with some of the currently used cloud environments that are suitable for modelling, a few remarks are relevant. These environments in general are considered offering better integration of functional components, something that is currently still lacking in the agro-climatic modelling environment under D4Science. Evaluators reported specific concerns regarding the expected sustainability of the currently available solutions and consequently if efforts currently invested will pay-off in reusability towards the future. A specific point related to integration that was mentioned is the transparency of access to data services, which is generally well-

developed in commercial cloud solution offers by e.g. Amazon or Google, but still immature in the D4Science environment.

Both evaluators concluded that for some functions better options are available when considering the current state of technology. Both have mentioned the use of Docker containers as a more common and better suitable mechanism to deploy components, that might complement (or even replace) the current GCube infrastructure. This would among others prevent compatibility issues that were encountered during the evaluation when deploying the developed Python model on the VRE. They strongly suggest supporting this technology stack in order to better facilitate the deployment, sharing and reuse of models, and algorithms. The need for easier and more transparent data access, which is available in most cloud solutions on the market (Google Big Query was mentioned as an example) was specifically suggested. Evaluators feel that there is still a gap between data access (mostly through metadata searches, providing links to individual datasets, but not offering data services) and the processing of data in models or analytics.

From the perspective of developing scientific resources for integration in the VRE, the current fragmentation and lack of alignment of documentation was mentioned as a barrier for component development and efficient use of the VRE. Also, not all available documentation is currently complete and up to date. A central access point to all relevant documentation, most useful when separately served for component developers and "end users" would help in adopting the VRE by broader communities and becoming more productive.

5.2.5 Comments, remarks and suggestions

The following list provides a summary of comments, remarks and suggestions that were collected from the evaluators during the evaluation. Although most of them are already integrated as part of the previous sections, they are provided for completeness and for future reference.

Table 7 – Comments, remarks and suggestions retrieved from the evaluation

Remark	Function(s)
Why to use a dedicated visualisation component, if ready-to-use solutions that are commonly used in research environments like plotly are available? There's a high chance that such visualisation components remain unused and that might for some be a reason to not use the VRE at all. The suggested "work-around" that evaluators would use in the VRE is calling a library like Plotly via Jupyter/Python scripts.	VRE components
It seems not very efficient that researchers cannot adapt the virtual research environment themselves on-the-fly. The example mentioned was a modeller requiring additional (Python) libraries on the fly in their daily research practice.	VRE components
The input dataset and output dataset folders are currently too flat. There's no option to structure file sets and this will almost immediately lead to an abundance of files and lack of overview of "what belongs to what", e.g. distinguishing the file sets used for different model runs.	Integration
The evaluators question the feasibility of the current Data Miner approach to run massive simulation exercises that use heterogeneous input data sets (e.g. separate datasets/services for weather, soil, crop characteristics etc.) and that exploit parallel computations and cloud storage in an effective way.	Cloud / parallelisation options
Under the interpreter options (Data Miner), the version is "free-text", which is a very possible source of problems. Besides, e.g. Java interpreters are backwards compatible. Why not provide a list with interpreters that are supported by the VRE.	VRE components
Analytics / Data Miner: No possibility to use named parameters, this might result in errors because the order of parameters is essential.	VRE components
It's not logical that an output file is mandatory. In many cases a model or algorithm would write its outputs to a database directly. In that case, you currently need to write a dummy output file.	VRE components
It's currently cumbersome, and not expected to be workable for less experienced modellers towards the future (and the upcoming 2nd and 3rd pilot evaluation), that they need to know about WPS and	Integration

how to call processes through WPS, before they can do (explorative) research. A user-friendly interface or library function, where e.g. only the process and the file(s) location has to be provided is suggested.	
The input/output is "single silo and cannot be organized. The option to structure I/O is needed as soon as serious model exercises (e.g. more than a single run) are performed.	Integration
The concept of having a VRE where distributed, international research groups can perform collaborative research and share and reuse resources like data and algorithms is highly appreciated.	Collaboration
Jupyter: In common notebook deployments there's a "parallel option". A question of the evaluator was why this is not available, as this is typically an option you would expect to be available and would allow to organize workload.	VRE components
A way to easily access and query data (like Google Big Query), which would make working with (distributed) data easier and not transparent is not available.	Integration
The fact that the modelling environment (Jupyter) is implemented on another infrastructure than the workspace, visualisation etc., makes it very hard to construct a full workflow and leads to compatibility issues when deploying	Integration
Currently, the offered spatial options are just loose components with no integration. This is the case for both the Geo Explorer and the Geo Analytics widget. For example, there's no way to directly visualize the spatial output of an output file (e.g. in GeoJSON) format. Even if the geoexplorer or analyzer would support that, the files would probably first have to be registered in the GeoServer behind the widget. Besides, it seems to have some obsolete features (e.g. available functions).	Integration, VRE components
Versions: When deploying the evaluation model, we run into the fact that the Data Miner runs with a relatively old version 2.7.6, which did not allow the developed model to be executed. Moreover, we suggest to add support for Java version higher than v8.	Integration
Versions: Version input (Python 3.5) into the SAI is not handled correctly (accepting 3.5 but using and reporting errors from 2.7.6)	VRE components
In the Wiki documentation it's unclear which version of Python is running on which environment as there are references to 2.7.6 (on computational machines) but also to version 3.5 (still under development)	Documentation

ANNEX 1 – EVALUATION DATA

Evaluator 1 - senior modeller, with a broad experience in the development of agro-climatic models and the application of models in large research projects in a European and global context.

Indicator - Ease of use

The indicator "ease of use" will assess how clear and easy to use the D4Science environment and the agro-climatic modelling VRE are to different types of users. It tries to answer the following types of questions:

- *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 modelling components and workflows?*

How do you score the ease of use of the VRE from the following perspectives?

Collaboration and communication

1- insufficient	2	3- sufficient	4	5 - very good	Not applicable
	X				

Further explanation and comments:

The communication aspect is OK, the file sharing aspect is OK but there is no clear interface to provide shared access to large data sets which are often required for model runs with different teams on the same data set. Something like OpenDAP, google bigQuery or something like S3 buckets need to be integrated in order to allow efficient access to large datasets within the VRE.

Model development

insufficient	moderate	sufficient	good	very good	Not applicable
X					

Further explanation and comments:

The environment is insufficient for model development because it does not provide any capabilities that modern IDE's have for inspecting variables and understanding the model execution path.

Model execution

insufficient	moderate	sufficient	good	very good	Not applicable
	X				

Further explanation and comment:

When executing a simple python model, we ran into problems related to differences in python environments between the environment where the model was built and the environment installed in the VRE (e.g. the SAI). This could be solved relatively easily by focusing on Docker containers where the model developer provides a container that includes all software required for running the model. Further, the

abilities to scale and parallelize workflows in the VRE are not very clear and rather difficult to be implemented.

Data analysis and visualization

insufficient	moderate	sufficient	good	very good	Not applicable
	X				

Further explanation and comments:

Currently, the opportunities for explorative data analysis are limited. Everything that is based on R (integrated R Studio) seems to work well, but the use of the Jupyter notebook is cumbersome as there is no clear access to the VRE environment and all calls have to be done through WPS. Further, the current possibilities of the built-in visualization tools are limited and to me it is unclear why those have been chosen/developed given that open-source tools like plotly/matplotlib are already available.

Further explanation and comments regarding ease of use in general:

None

Indicator - Learning Curve

This indicator was introduced to assess how much time evaluators think will be required to learn new concepts and understand the working of VRE tools and their integration before the VRE can be effectively used.

How do you score the learning curve of the VRE from the following perspectives?

Estimated effort to make an existing model run on the VRE

1 - very low		3 - moderate		5 - very high	Not applicable
			X		

Can you provide a rough estimate in hours, day or months you think it would take to adapt an existing model to run on the VRE?

It would probably take several days to adapt an existing model to run on the VRE because of the difficulties related to connecting input/output. Adapting the model to simulate using large data sets would be much more difficult.

Estimated effort to configure and run a model and analyze its outputs

1 - very low		3 - moderate		5 - very high	Not applicable
			X		

Can you provide a rough estimate in hours, day or months you think it would take to perform a single model run and do a quick analysis of its outputs?

No

Further explanation and comments regarding learning curve:

None

Indicator - Usefulness

This indicator compares the VRE and its usability, e.g. compare to current hardware and software in use. Characteristics that could be compared are for instance functionality and costs.

How useful do you think that the current agro-climatic VRE and its tools are from the following perspectives?

Usefulness for model execution

1 - not useful	2	3	4	5 - very useful	Not applicable
		X			

Usefulness for data analysis

1 - not useful	2	3	4	5 - very useful	Not applicable
	X				

Usefulness for visualization

1 - not useful	2	3	4	5 - very useful	Not applicable
X					

Collaborative modelling

1 - not useful	2	3	4	5 - very useful	Not applicable
	X				

Further explanation and comments regarding usefulness of the VRE for modelling:

See comments made under “ease of use”

Indicator - Openness

Openness should be considered an important asset of any environment in the light of the European Open Science Cloud. To assess openness, the assessment plan states that ease of adding new data and functionality to the VRE or to share workflows, components, and data with the outside world could be important issues.

How open do you think the agro-climatic VRE and its tools are from the following perspectives?
Ease of accessing models and data registered in the VRE

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable
		X			

Ease of registering and sharing of datasets

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable
			X		

Ease of adding and sharing new functionality and components (e.g. model, visualization)

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable
	X				

Further explanation and comments regarding openness:

Accessing/sharing data is easy, but the challenge is to efficiently access and share large datasets that are often required for model simulation runs. I do not see how the VRE tackles this as there is no interface or API.

Indicator - FAIR-ness

The FAIR principles (Findability, Accessibility, Interoperability, Reusability, https://www.go-fair.org/fair-principles/) are key principles for open science on any VRE. Here you are asked to assess how the VRE helps in making research data and algorithms FAIR and what the advantages and disadvantages are compared to current research environments and data management practices.

How do you think the current version of the VRE complies with the following FAIR principles?

Findability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
	X				

Accessibility

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
	X				

Interoperability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
X					

Reusability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
	X				

Further explanation and comments regarding openness:

None

Evaluator 2 - senior software engineer, with broad experience in programming agro-climatic models and one year of experience of adapting existing modelling software and deploying and running these on the agro-climatic modelling VRE.

Indicator - Ease of use

The indicator "ease of use" will assess how clear and easy to use the D4Science environment and the agro-climatic modelling VRE are to different types of users. It tries to answer the following types of questions:

- *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 modelling components and workflows?*

How do you score the ease of use of the VRE from the following perspectives?

Collaboration and communication

1- insufficient	2	3- sufficient	4	5 - very good	Not applicable
		X			

Model development

insufficient	moderate	sufficient	good	very good	Not applicable
	X				

Further explanation and comments:

Developing and deploying a model (via the Data Miner) facilities is not very easy. Development has to be done locally, so the development environment might differ from final deployment environment (e.g. different version of Java or Python, different versions of software libraries, etc.). SAI user interface is confusing. Steps to import a developed model into Data Miner, defining inputs and outputs, packaging, running, and debugging are time consuming.

Model execution

insufficient	moderate	sufficient	good	very good	Not applicable
	X				

Further explanation and comments:

There's quite some overhead before data and algorithm are on computational node and can be executed, this might not be very efficient for fast running algorithms. It is currently difficult to use VRE for parallel computations.

Data analysis and visualization

insufficient	moderate	sufficient	good	very good	Not applicable
X					

Further explanation and comments:

Not many options yet for data visualization. What is available is a separate component in the VRE and not very integrated with the rest. Can only define a few simple types of charts, based on a single dataset. The dataset is copied into the component which is not very efficient and might not work for large datasets. Spatial data is not easy to display yet.

Further explanation and comments regarding ease of use in general:

Components are there (though as rough version) but integration between the components is lacking. File and algorithm access via web calls, and have to copy data between components in the VRE to use it with the different available functionalities.

Indicator - Learning Curve

This indicator was introduced to assess how much time evaluators think will be required to learn new concepts and understand the working of VRE tools and their integration before the VRE can be effectively used.

How do you score the learning curve of the VRE from the following perspectives?

Estimated effort to make an existing model run on the VRE

1 - very low		3 - moderate		5 - very high	Not applicable
			X		

Can you provide a rough estimate in hours, day or months you think it would take to adapt an existing model to run on the VRE?

If the model runs on the pre-configured computation nodes (e.g. with the available version of Java, Python, software libraries, etc.) and has only a few simple inputs and outputs it might be quick (hours). If not, it can get very time consuming (weeks, maybe months). Wrapping the model for Data Miner and the debugging process to get it running can be slow.

Estimated effort to configure and run a model and analyze its outputs

1 - very low		3 - moderate		5 - very high	Not applicable
	X				

Can you provide a rough estimate in hours, day or months you think it would take to perform a single model run and do a quick analysis of its outputs?

If the model is already implemented and can be executed on the VRE this can be quick (hours-day).

Further explanation and comments regarding learning curve:

Available documentation can be hard to find and is not always up-to-date. There can also be a difference in the documentation about what is available on the production release of D4Science, and the development release.

Indicator - Usefulness

This indicator compares the VRE and its usability, e.g. compare to current hardware and software in use. Characteristics that could be compared are for instance functionality and costs.

How useful do you think that the current agro-climatic VRE and its tools are from the following perspectives?

Usefulness for model execution

1 - not useful	2	3	4	5 - very useful	Not applicable
			X		

Usefulness for data analysis

1 - not useful	2	3	4	5 - very useful	Not applicable
X					

Usefulness for visualization

1 - not useful	2	3	4	5 - very useful	Not applicable
X					

Collaborative modelling

1 - not useful	2	3	4	5 - very useful	Not applicable
		X			

Indicator - Openness

Openness should be considered an important asset of any environment in the light of the European Open Science Cloud. To assess openness, the assessment plan states that ease of adding new data and functionality to the VRE or to share workflows, components, and data with the outside world could be important issues.

How open do you think the agro-climatic VRE and its tools are from the following perspectives?

Ease of accessing models and data registered in the VRE

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable
			X		

Ease of registering and sharing of datasets

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable
		X			

Ease of adding and sharing new functionality and components (e.g. model, visualization)

1 – very hard	2	3 - moderate	4	5 - very easy	Not applicable
	X				

Further explanation and comments regarding openness:

None

Indicator - FAIR-ness

The FAIR principles (Findability, Accessibility, Interoperability, Reusability, https://www.go-fair.org/fair-principles/) are key principles for open science on any VRE. Here you are asked to assess how the VRE helps in making research data and algorithms FAIR and what the advantages and disadvantages are compared to current research environments and data management practices.

How do you think the current version of the VRE complies with the following FAIR principles?

Findability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
X					

Accessibility

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
	X				

Interoperability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
X					

Reusability

1 – insufficient	2	3 - sufficient	4	5 - excellent	Not applicable
	X				