D5.4 - Agro-climatic & 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 & Economic Modelling Pilot Evaluation Report
RESPONSIBLE AUTHOR	Rob Lokers, 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	Mr. Christophe Doin

WORKPACKAGE N. | TITLE	WP5 | Agro-climatic & Economic Modelling Community
WORKPACKAGE LEADER	Wageningen Environmental Research
DELIVERABLE N. | TITLE	D5.4 | Agro-climatic & 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/D4.2.pdf
DATE OF DELIVERY (CONTRACTUAL)	30 June 2018 (M18), 31 December 2019 (M36 Updated version)
DATE OF DELIVERY (SUBMITTED)	12 July 2018 (M19), 27 January 2020 (M36 Updated version)
VERSION | STATUS	V3.0 | Final
NATURE	ORDP (Open Research Data Pilot)
DISSEMINATION LEVEL	Public (PU)
AUTHORS (PARTNER)	Rob Lokers, Rob Knapen (Wageningen Environmental Research)
CONTRIBUTORS	-
REVIEWER	Mathias Filter (BfR)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
1.1	Initial version 2nd evaluation	04-07-2019	Rob Lokers
1.2	Evaluation 2 reporting	27-08-2019	Rob Lokers, Rob Knapen
2.1	Initial Version 3rd evaluation	14-11-2019	Rob Lokers
2.2	Evaluation 2-3 reporting	21-11-2019	Rob Lokers
2.3	validation definition	19-12-2019	Rob Lokers
2.4	validation – reporting	20-12-2019	Rob Lokers
2.5	Evaluation 4 – reporting of final evaluation results Draft for review	24-12-2019	Rob Lokers, Rob Knapen
2.9	Reviewed version	26-12-2019	Rob Lokers, Rob Knapen
3.0	Final version	27-12-2019	Rob Lokers, Rob Knapen

PARTICIPANTS		CONTACT
Agroknow IKE (Agroknow, Greece)		Nikos Manouselis Email: nikosm@agroknow.com
Stichting Wageningen Research (DLO, The Netherlands)		Rob Lokers Email: rob.lokers@wur.nl
Institut National de la Recherché Agronomique (INRA, France)		Pascal Neveu Email: pascal.neveu@inra.fr
Bundesinstitut für Risikobewertung (BfR, Germany)		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)		George Kakaletris Email: gkakas@di.uoa.gr
Stichting EGI (EGI.eu, The Netherlands)		Tiziana Ferrari Email: tiziana.ferrari@egi.eu
Pensoft Publishers Ltd (PENSOFT, Bulgaria)		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
GACS	Global Agricultural Concept Scheme
GPU	Graphics Processing Unit
VRE	Virtual Research Environment
WOFOST	World Food Studies (crop growth simulation model)
WPS	Web Processing Service

EXECUTIVE SUMMARY

In the past three years, the AGINFRA project has introduced virtual research environments as a collaborative tool for open data science in three communities in agri-food research. For the agro-climatic modelling community that was targeted in WP5 of the project, it was for most members the first time they learned about virtual research and the opportunities of using VREs for collaborative data science. In the past years, many community members have heard about the backgrounds, seen demonstrations of the VRE and its agro-climatic modelling pilot applications and had the opportunity to provide their opinions and thoughts through evaluation and validation surveys. A group of evaluators also had the chance to get a more in-depth view, did guided hands-on work in the VRE and responded to the evaluation survey in the three evaluation events that were organised.

This document describes the evaluation procedures and evaluation outcomes performed in the AGINFRA+ project to assess the agro-climatic modelling Virtual Research Environment (VRE) and the pilot applications developed for the agro-climatic modelling community.

Analysing the evaluation feedback, we can conclude that in general the impressions and opinions about the AGINFRA+ VREs and the developed pilot application are overall positive. We observe that community members state that VREs can be a useful tool for digital open science, and we see a broad willingness and enthusiasm to start using such environments for research. There is also a lot of interest for the developed pilot applications, both from the perspective of novel technologies used in the modelling pilots and from the more scientific perspective of used methodologies. The developed pilot applications show very well the potential and opportunities of collaborative, open science for the agro-climatic modelling community and how to create valuable applications and scientific workflows. Because we have demonstrated the VRE and the pilot applications to a broad audience that also includes data scientist that (partly) work on the broader agro-environmental domain, we can also conclude that concepts and applications are well received and useful beyond the agro-climatic modelling community. Many seem to acknowledge that virtual research offers unique options to collaborate, co-develop and share, responding to the demand for working together over different domains with heterogeneous data that is required to tackle the grand interdisciplinary challenges in agri-food and environment.

Nevertheless, we also need to state that there are clear signs of shortcomings and request to improve features, both from evaluators and from the developers and modellers that were involved in developing the pilot applications. These are reported and have been expressed as recommendations for future improvements where possible.

For the agro-climatic modelling community we can conclude that through AGINFRA+ a number of valuable steps have been taken to introduce virtual research environments to groups of scientists that were generally not aware of the opportunities of virtual research. As a project, supported by previous outcomes from infrastructural and innovation actions, we were able to show the general concept of virtual research, and how VREs can support collaborative teams in performing data science. Moreover, we have demonstrated through the developed community driven pilot applications how a VRE and its components can be used to develop scientific workflows and catalyse open science. Steps like this could only be made because of the collaborative and iterative process implemented in AGINFRA+, where infrastructure developers collaborated with developers and data scientists from scientific communities in agri-food.

TABLE OF CONTENTS

LIST OF TABLES

1 INTRODUCTION

This document describes the evaluation procedures and evaluation outcomes performed in the AGINFRA+ project to assess the agro-climatic modelling Virtual Research Environment (VRE) and the pilot applications developed for the agro-climatic modelling community.

The three agro-climatic community specific pilot evaluations that were described in the assessment plan for the agro-climatic modelling pilots were aligned with the stage of evolution and maturity of the pilot applications under development. This allowed us to better engage the community users and attract attention for the pilots. To cope with the differences in the evaluation approach between the different research communities (besides agro-climatic modelling, also the food security and the food safety community were evaluating community specific pilots), an additional validation round was added to the project activities. This was developed and performed in the format of more holistic and harmonized validation exercise.

The first section of this document describes the objectives of the evaluation and the applied approach for the different evaluation stages and the validation.

The two subsequent sections describe in more detail how the three originally planned pilot evaluations were developed and performed. Here, the used methodology is described, distinguishing between the first and the second and third evaluation respectively, which used different approaches, aligned with the maturity stage of the available VRE functions and pilot applications. Also, a definition of the evaluation (sub)indicators is provided and the evaluation session scenarios are specified. The outcomes of the evaluation surveys are reported per sub indicator and a short summary and conclusions related to these outcomes is provided. The associated evaluation exercises are provided as part of the Annexes of the report.

The last two sections of the report provide the details of the additional validation that was performed as part of the third evaluation period, using a harmonized methodology, which was developed to be able to better compare evaluation outcomes over de three use cases. The outcomes of the survey, which was coupled to an online webinar are reported and a short summary and conclusions regarding the agro-climatic modelling community are provided.

2 PILOT EVALUATION AND VALIDATION – OBJECTIVES AND APPROACH

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, evaluation and validation, in parallel with the proposed three-phase piloting scheme in the project proposal, was designed and performed.

Evaluation within AGINFRA+ has been a challenging task, considering the variety of possible evaluation items (scientific experimentation workflows, technology tools, content/data, computing resources). It was even more difficult, since the different communities involved in the individual use cases had different research and data science requirements and needs. A so-called mixed-method approach was implemented to deal with this. While the assessed indicators were in fact similar, the three communities implemented different evaluation approaches, aligned with the needs and opportunities of the community and its users. While this proved to be an effective way to engage the individual communities and to collect their feedback, this also led to a quite heterogeneous outcome with a variety of tools and methods used to collect and report feedback back to our project work.

To also make sure that a comparison over communities can be made, we have extended the assessments planned for the 3rd evaluation period with an additional validation. This validation (described in the AGINFRA+ deliverable Harmonized Use Case Validation methods, guidelines and materials) uses a more holistic and harmonized approach. The validation approach that we chose aims to look at a concrete and well-defined outcome of the project (the software demonstration of a complete data-powered workflow), as well as to use the same evaluation instrument to collect feedback from the different stakeholders.

2.1 EVALUATION

The performed evaluations to assess the agro-climatic pilot applications developed in the VRE were aligned with the three piloting phases defined in AGINFRA+, to subsequently assess the provided VRE capabilities and the developed pilot applications that were available after the 1st, 2nd and 3rd period of pilot development. Obviously, the level of "readiness" of the VRE and pilot applications to cope with community requirements over these periods increased. The setup and ambition of the three evaluation assessments were aligned with this evolution, and they were adapted to the maturity of developed components and research workflows.

The main objective of the VRE development process in the first pilot phase was 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. Consequently, the agro-climatic pilot applications were not yet advanced enough to be assessed by a broad audience and to be used for hands-on exercises. This first assessment was therefore performed with a small group of experts, that were knowledgeable enough to "see through" the limitations and lack of integration and workflow of the phase 1 VRE components and pilot implementations. The focus was on 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. Main objective was to collect feedback and requirements for the phase 2 and 3 improvements.

In the second piloting phase, the available VRE capabilities and the developed agro-climatic pilot

applications were advanced enough to be tested independently by modelers and data scientists. The amount of VRE and pilot functionality, the stability of developed applications and the integration of components were sufficient to allow evaluators of different expertise levels to work with the VRE and applications in hands-on exercises. Main objectives were to get a good impression of the usability of the applications and their support for open science, to rigorously test the components and to collect feedback for improvements to be implemented in the third round of pilot implementation.

The third piloting phase aimed at delivering fully functional applications for the use cases selected for the agro-climatic modelling community, with issues solved and improvements implemented that were identified during the second pilot evaluation. The evaluation was performed with a diverse group of evaluators from both the agro-climatic and food-security communities, to also get some cross-community attention to different VRE applications. Besides further testing the data science and open science capabilities, the main objective was to obtain feedback and suggestions for improvement, to be implemented in the last months of the project, but particularly also for future, post-project follow-up.

The evaluations of the VRE pilots with the agro-climatic modelling community were, aligned with the common evaluation strategy and the individual evaluation plans of the three piloting communities, aimed at evaluation of a relevant set of indicators. The methodology and the indicators used for the agro-climatic community evaluations have been documented in the Assessment Plan of WP5 (deliverable 5.3). These evaluation indicators are also shown in Table 1. It concerns a list of both qualitative and quantitative evaluation indicators, the possible method for their assessment (e.g. quantitative benchmarking, expert review, end user survey) as well as the piloting phase in which these indicators are particularly relevant and need to be evaluated.

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

2.2 VALIDATION

The additional validation of the agro-climatic modelling pilots that was included in the 3rd evaluation period has been performed according to the approach described in the deliverable Harmonized Use Case Validation methods, guidelines and materials (D1.2). For details on the background of the validation methodology, the indicators that are assessed and the specific survey that was developed to collect the required data, we refer to this report.

In terms of this deliverable D1.2 a "light-weight, remote validation trial" was implemented. Main reason for choosing a remote validation trial is that in general, but specifically due to the late decision to extend the evaluations with the validation, it was expected that this would have the highest impact. The agro-climatic user community is a world-wide community, and opportunities to have physical meetings are scarce. A remote evaluation is less time-consuming and by providing the webinar recording and evaluation resources on-line, it allows also the persons that cannot attend live to participate.

3 PILOT EVALUATION – METHODOLOGY AND MATERIALS

3.1 METHODOLOGY FOR FIRST EVALUATION

3.1.1 Introduction

The first evaluation was planned at a relatively early stage in the project. Because of the at that time still limited functionality available in the VRE and the first phase pilot applications, the methodology described above could not yet be applied in the first pilot evaluation. The VRE for agro-climatic modelling at the time of this first pilot evaluation did not yet (as foreseen in the assessment plan) support the deployment of full modelling workflows. It supported the hosting of specific modelling components, but full and transparent integration of these components was not yet supported. Therefore, the evaluation focused on the assessment of the capabilities of the VRE to support the development, deployment and execution of these components.

3.1.2 Evaluation programme

This evaluation had the format of a guided walkthrough and review with senior experts. The rationale was that in this stage, this was the most effective way to collect clear suggestions for improvement of module and integration issues that would be critical for developing successful pilot applications in the following evaluation cycles. In the review, 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.

<u>Evaluation programme</u>

1. *General principles and structure of a VRE*
   - What is a Virtual Research Environment?
   - Relation with Open Science, FAIR principles
   - Functions of a VRE
     - 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.
2. *VRE walkthrough & demonstration*
   - Collaboration functions: Shared workspace, mail, timeline/newsfeeds, activity tracking, catalogue
   - Modelling environment: DataMiner, Jupyter notebooks, RStudio
   - Visualization: viewing and creating graphs, visualizing spatial data
   - Running the WOFOST model
   - Requirements and adaptations for running a model in the VRE
3. *Hands-on guided exercise*
   - *Deploying and running the WOFOST model: preparing data, running the model, tracking progress,* analyzing outputs, visualization
   - Hello world exercise

    - Create a small model (Jupyter notebooks)
    - Organize input data
    - Use SAI to deploy the model in DataMiner
    - Execute the model
    - Find and analyze output
4. *User evaluation survey*
    - Evaluation survey
    - Short interview of the evaluator to collect further experiences and comments.

3.2 METHODOLOGY FOR SECOND AND THIRD EVALUATION

3.2.1 Introduction

The second and third evaluation follow the evaluation design that was defined in the Assessment Plan for WP5 (deliverable 5.3). To cope with evolving insights, the definition of a few sub indicators was adapted to better fit the different aspects that needed to be evaluated per indicator.

The concept of the evaluation sessions was to spend a full day. The first half of the day was dedicated to providing the evaluators with sufficient background to understand the objectives of AGINFRA+, the concepts of virtual research and the technology behind it, and the objective behind the developed pilot applications. It also contained demonstrations of the pilot applications. The second half of the evaluation programme was a hands-on workshop. Evaluators were asked to perform two practical exercises that touched the main aspects of the agro-climatic modelling VRE and working with the pilot applications. Short exercise descriptions (see Annex 1) were prepared as a manual, guiding the participants through the work. At the end of the day, participants were asked to fill in the evaluation survey.

3.2.2 Evaluation programme

The evaluation programme introduces participants in the background of the AGINFRA+ project and its objectives and it explains the main functions available in the AGINFRA+ agro-climatic modelling VRE. Wageningen UR developers of the AGINFRA+ project demonstrate the pilot applications that were developed to demonstrate the use of VREs for the agro-climatic community. The hands-on part of the workshop contains two exercises that allow the participants to run through some basic workflow scenarios. It is concluded with an evaluation survey.

<u>Evaluation programme</u>

1. *Introduction*
    - The European Open Science Cloud, objectives and state of play
    - AGINFRA+ project, user communities, infrastructure
    - Virtual research, VREs and the D4Science environment
    - Introduction to main functions of a VRE: collaborate, develop, publish & share, upscaling
    - Introduction to the evaluation methodology
2. *AGINFRA VRE architecture, open science functions and tools*
    - Data analytics
    - Semantics
    - Data visualisation

- Publishing

3. *Demonstration 1 – Crop phenology*
   - Introduction
   - Co-developing agronomic algorithms using VRE Jupyter notebooks
   - Deploying and sharing algorithms via DataMiner and VRE catalogue
   - Demonstration of the crop phenology estimation algorithms
     - Running the algorithm for a single agricultural parcel
     - Parallel/distributed running of the algorithm for a region or country
4. *Demonstration 2 – Crop growth Modeling*
   - Introduction
   - WOFOST model and implementing WOFOST-CLOUD for distributed modelling
   - Co-developing agronomic algorithms using VRE Jupyter notebooks
   - Running the algorithm for a single agricultural parcel
   - Agro-climatic modelling dashboard, integration of data provision, modelling and visualisation
   - Parallel/distributed running of the algorithm for a region or country
5. *Hands-on workshop*
   - Exercise 1 – Crop phenology
     - Exploring the used VRE tools
     - Exploring and running the developed Jupyter notebooks
     - Running the crop phenology estimation algorithms
   - Exercise 2 – Crop growth modelling
     - Exploring the used VRE tools
     - Demonstration of the crop growth modelling algorithms
       - Modelling, visualisation and analysis using the agro-climatic visualisation dashboard
       - Running the WOFOST DataMiner algorithms for a region – distributed computing
     - Publishing model outputs in the VRE catalogue
6. *Evaluation*
   - Participants fill in the evaluation survey

3.2.3 Evaluation Indicators and survey questions

With the defined scope of the first pilot evaluation, the selected pilot evaluation indicators mainly consider 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.

<u>Indicator - Ease of use</u>

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 collaboration & communication	Qualitative scoring Expert review	Scoring the individual model development features of the VRE + open questions	Modeler, Software Developer
Ease of use of the VRE for algorithm & model development	Qualitative scoring Expert review	Scoring the modeling configuration and execution features of the VRE + open questions	Senior modeler, Developer
Ease of use of the VRE for algorithm & model execution	Qualitative scoring Expert review	Scoring the modeling analysis and visualization features of the VRE + open questions	Modeler
Ease of use of the VRE for data analysis & vizualisation	Qualitative scoring Expert review	Scoring the collaboration features of the VRE+ open questions	Modeler, Software Developer

Indicator - 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 make an existing model run on the VRE	Qualitative scoring Expert review	Scoring on an ordinal scale the estimated learning curve against the "conventional" environment and open questions	Modeler, 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	Modeler

Indicator - 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 Algorithm & Model Development	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Senior modeler, Software developer
Usefulness of the VRE for Algorithm & Model Execution	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Modeler
Usefulness of the VRE for visualization	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Modeler, Software developer
Usefulness of the VRE for data analysis	Qualitative scoring Expert review	Scoring on an ordinal scale and open questions	Modeler

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.

Table 5 – Evaluation metrics for the indicator "Openness"

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

Indicator - 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	Modeler

Accessibility	Qualitative scoring	Scoring on an ordinal scale	Modeler
Interoperability	Qualitative scoring	Scoring on an ordinal scale	Modeler, Software developer
Reusability	Qualitative scoring	Scoring on an ordinal scale	Modeler, Software developer

Indicator - Performance, Scalability

Agro-climatic modelling is a data and resource intensive activity, so performance and the option to scale up modelling jobs on demand using (cloud) storage and computing clusters are relevant indicators for any research environment for modelling and associated analytics.

In practice, this indicator can only be well judged by evaluators with extensive experience with the VRE and the development of scalable modelling components. Assessment of this indicator will therefore be performed through interviews with the involved model and software developers.

Indicator - Reliability

Reliability is an important indicator, because VREs support researchers in their main work processes and they depend on such systems to be stable and available. This is particularly valid for modelling, where jobs might have to run for a longer time, and unexpected downtime is not acceptable.

In practice, this indicator can only be well judged by evaluators with extensive and long-term experience with developing and running models and performing data science through the VRE. Assessment of this indicator will therefore be performed through interviews with the involved model and software developers.

Indicator - Transparency

For open science and particularly to ensure the reusability of research results, it is important that the provenance of data, models and model runs is clear, and that model runs are repeatable.

In practice, this indicator can only be well judged by evaluators with extensive and long-term experience with developing and running models and performing data science through the VRE. Assessment of this indicator will therefore be performed through interviews with the involved model and software developers.

3.2.4 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.

Indicator - Ease of use

The indicator "Ease of Use" determines if the D4Science environment and the agro-climatic modelling VRE functions are clearly understandable and easy to use. It includes the following aspects:

- *Collaboration & communication (messaging & timeline, mail, issue management etc.)*
- *Development and execution of algorithms and models*
- *Data analysis and visualization*

Please score the ease of use of the AGINFRA+ VRE on the following types of activities:

Collaboration and communication

1- insufficient	2	3- sufficient	4	5 - very good

Algorithm & Model Development

insufficient	moderate	sufficient	good	very good

Algorithm & Model execution

insufficient	moderate	sufficient	good	very good

Data Analysis & Vizualisation

insufficient	moderate	sufficient	good	very good

Additional explanations, comments and suggestions for improvement regarding the ease of use:

_____

Indicator - Usefulness

This indicator evaluates how useful the D4Science environment and the agro-climatic modelling VRE and its functions are for modelers and data scientists to support their daily research and especially the work in collaborative activities, with remote researchers and research teams (e.g. in European or global cooperation).

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

Algorithm & Model Development

1 - not useful	2	3	4	5 - very useful

Algorithm & Model Execution

1 - not useful	2	3	4	5 - very useful

Data analysis

1 - not useful	2	3	4	5 - very useful

Visualization

1 - not useful	2	3	4	5 - very useful

Additional explanations, comments and suggestions for improvement regarding usefulness:

_____

Indicator - Openness

Openness is an important asset of any collaborative research environment, also in the perspective of the European Open Science Cloud. It concerns among others the ease of adding data, algorithms and models to the VRE and publishing and sharing data, analytics and workflows among fellow researchers and with the outside world.

Please score the openness of the AGINFRA+ VRE regarding the following:

Finding & accessing algorithms, models and data registered in the VRE

1 – very hard	2	3 - moderate	4	5 - very easy

Registering and sharing of data, algorithms and models

1 – very hard	2	3 - moderate	4	5 - very easy

Adding and sharing new functionality and components (e.g. a new model, visualization)

1 – very hard	2	3 - moderate	4	5 - very easy

Additional explanations, comments and suggestions for improvement regarding openness:

_____

Indicator FAIR-ness

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

Please score how the AGINFRA+ VRE complies with the following FAIR principles?

Findability

1 – insufficient	2	3 - sufficient	4	5 - excellent

Accessibility

1 – insufficient	2	3 - sufficient	4	5 - excellent

Interoperability

1 – insufficient	2	3 - sufficient	4	5 - excellent

Reusability

1 – insufficient	2	3 - sufficient	4	5 - excellent

Additional explanations, comments and suggestions for improvement regarding FAIRness:

_____

Indicator - Learning Curve

Virtual Research Environments should not add to the further complexity of data science and modeling. This indicator assesses the required effort to learn new concepts and understand the working of VRE tools and their integration before the VRE can be effectively used for research.

This section is specifically added for participants that have more extensively used the AGINFRA+ VRE to develop analytics, models and workflows. Questions in this section are therefore optional. However, you are welcome to provide your opinion in case you have only attended a workshop and have specific thoughts on this matter.

How do you score the learning curve of the AGINFRA+ VRE for the following activities?

Estimated effort to make an existing model run on the VRE

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

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

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

Additional explanations, comments and suggestions for improvement regarding Learning Curve:

_____

3.2.5 Used materials

The material presented in Table 7 was used to support the evaluation programme that was presented above. With the shown reference the specific item can be accessed or downloaded.

Table 7 – Reference material used in the AGINFRA+ agro-climatic modelling pilot evaluations

Material	Reference	Type
Introduction presentation	https://data.d4science.net/eq99	PPT

Agro-climatic modelling VRE (trial version)	https://aginfra.d4science.org/explore	VRE
Crop Phenology pilot demonstration - introduction	https://data.d4science.net/an1D	PPT
Crop Phenology exercises	https://data.d4science.net/3vmD	PDF
Crop Modelling pilot demonstration - introduction	https://data.d4science.net/K12k	PPT
Crop Modelling exercises	https://data.d4science.net/Jvzp	PDF
Evaluation workshop folder on VRE (auxiliary data and information)	http://data.d4science.org/workspace-explorer-app?folderId=Ynh0V0R0STZmL0diZnFwK0hvZnJaamg3Z2tEa1NtM21hd2xDRllrQnh4Wm5SalhPVlpmVFM5OEh1NU8vZjNrWQ	VRE folder

4 EVALUATION RESULTS

4.1 INTRODUCTION

This section provides and overview of the results of the three performed evaluations. It distinguishes between the first evaluation on one side and the second and third evaluation on the other side. This is mainly because of the different approach. At the first evaluation the VRE and pilot applications were assessed by just two senior experts, through a guided walkthrough and review, combined with interviews. The second and third evaluation had a different setup, with demonstrations and a hands-on session. Moreover, they were more structured, providing the same (Google Forms) questionnaire to all participants.

4.2 EVALUATION OUTCOMES – EVALUATION 1

4.2.1 Evaluation details

Location:	Wageningen, the Netherlands
Date:	10th of July 2018
Method:	guided walkthrough and review, interview
Participants:	2
Nr. of responses:	2

4.2.2 Assessing 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).

4.2.3 Running an 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.

4.2.4 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.

4.2.5 Interviews

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 to support 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.

4.2.6 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.

Remark	Function(s)
Why use a dedicated visualisation components 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),	Integration

that they need to know about WPS and 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 more easy 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

4.3 EVALUATION OUTCOMES – EVALUATION 2 AND 3

4.3.1 Evaluation details

Evaluation 2:

Location: Wageningen, the Netherlands
Date: 2 July 2019
Method: pilot application demonstrations, hands-on workshop, on-line survey
Participants: 18
Nr. of responses: 14

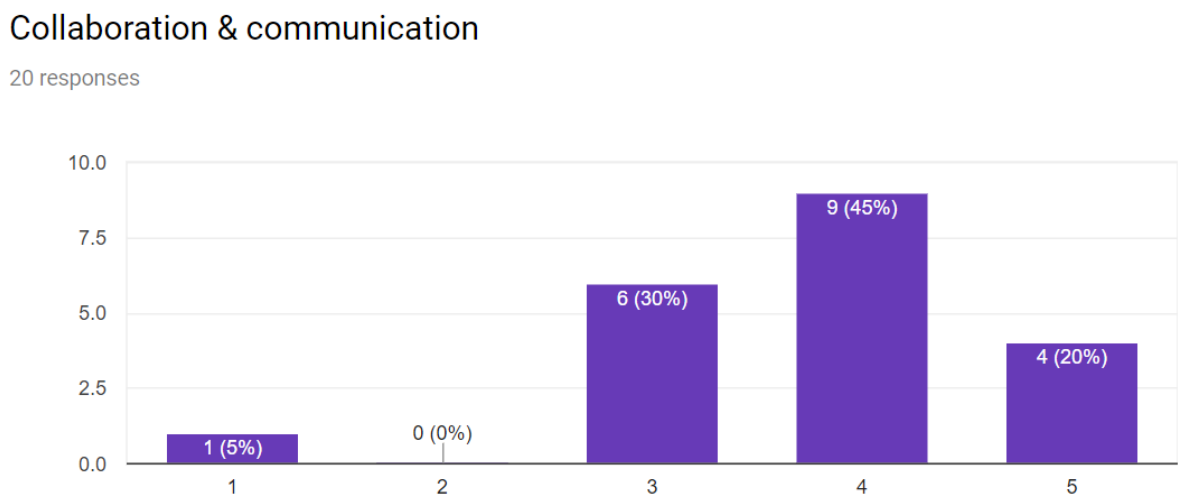
Evaluation 3:

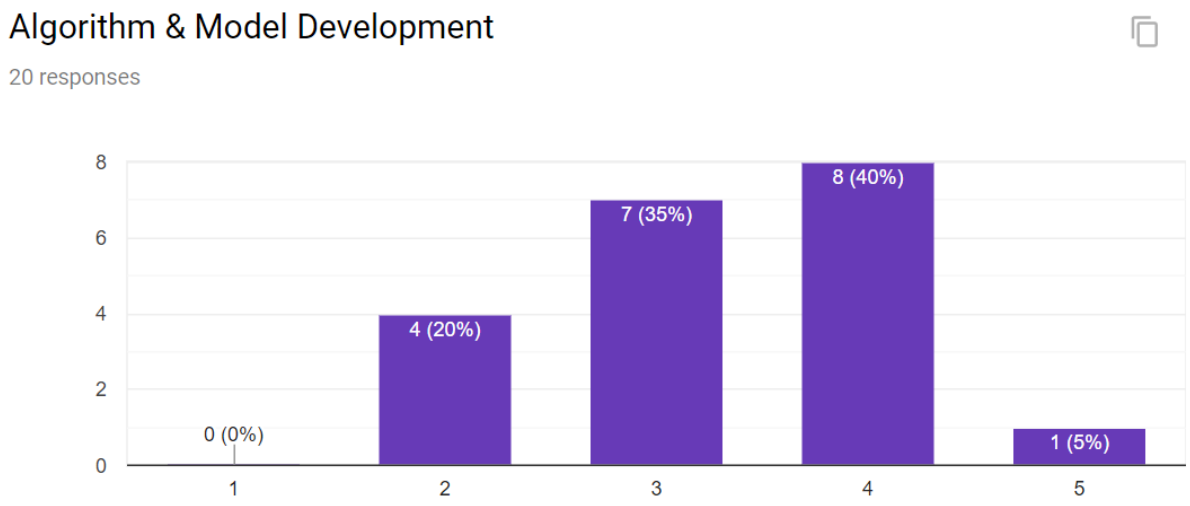
Location:	Paris, France
Date:	18 September 2019
Method:	pilot application demonstrations, hands-on workshop, on-line survey
Participants:	12 (from two user communities: agro-climatic modelling and food security)
Nr. of responses:	6

4.3.2 Indicators – Evaluation Scores

Indicator – Ease of Use

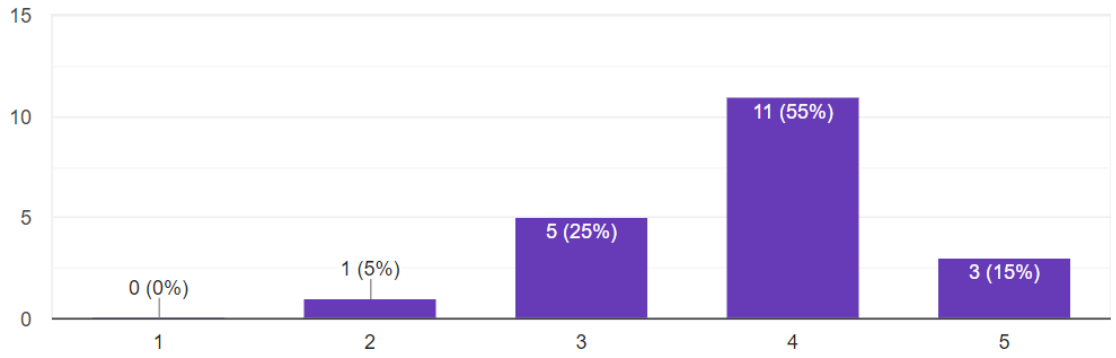
Indicator	Evaluation 1	Evaluation 2	Total
Collaboration & Communication	3.9	3.5	3.8
Algorithm & Model Development	3.4	3.0	3.3
Algorithm & Model Execution	3.8	3.8	3.8
Data Analysis & Visualisation	3.4	3.7	3.5
Total average – Ease of Use	3.6	3.5	3.6





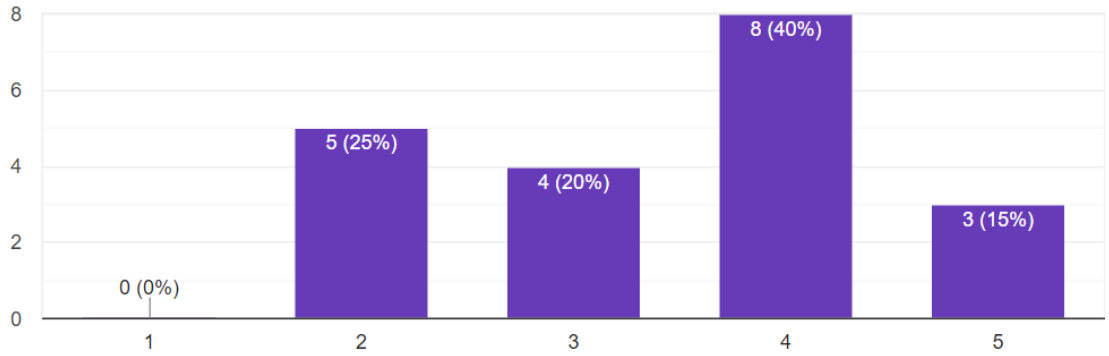
Algorithm & Model Execution

20 responses



Data Analysis & Vizualisation

20 responses



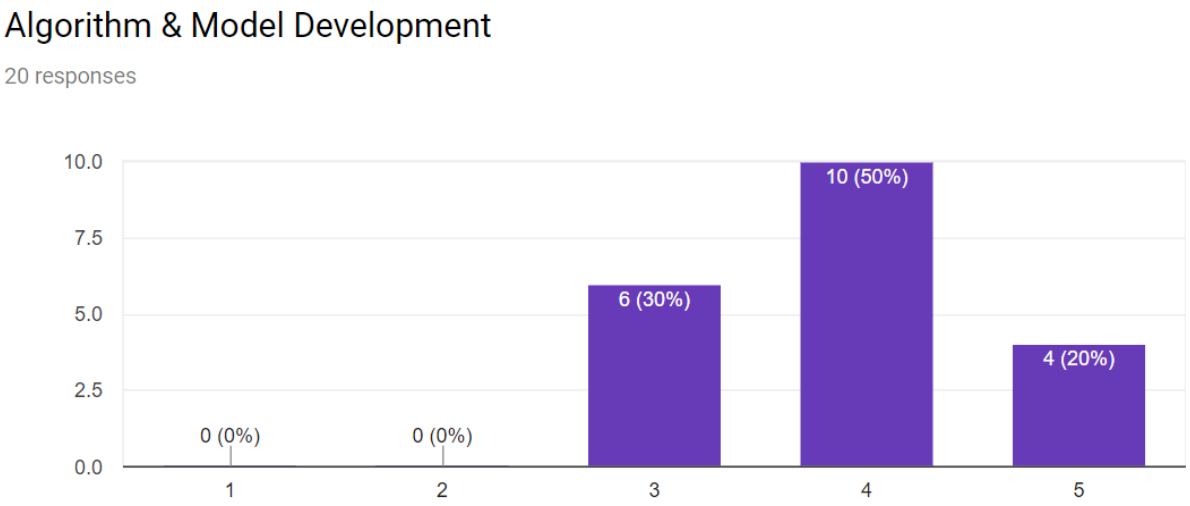
Additional evaluation remarks on the Ease of Use:

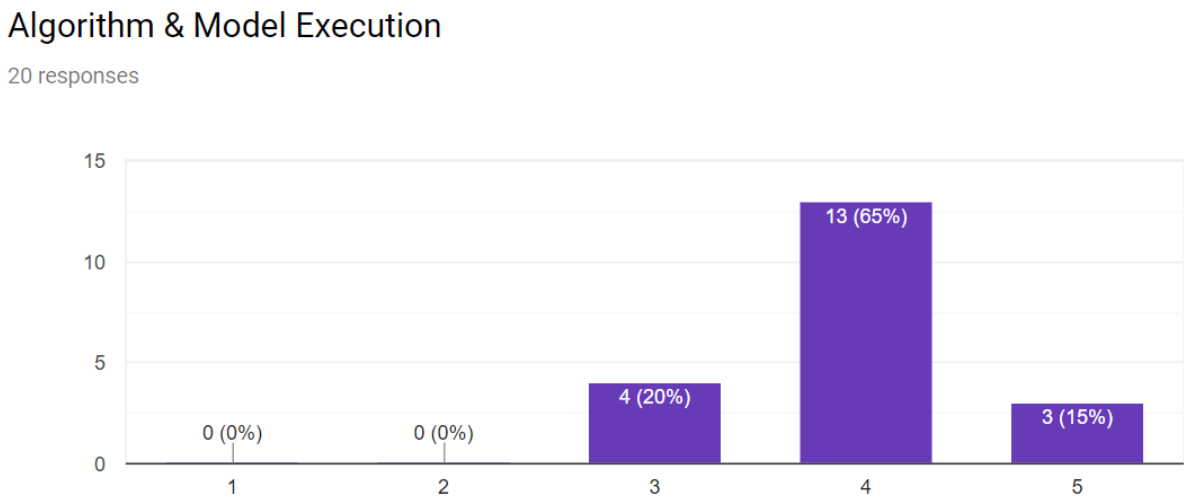
To further improve the co-development options, it would be good if researchers could truly co-work together at real-time in the same notebook (compare Google Docs) The standard visualisation options are difficult to use and too simple. More configuration and integration options would be required to make it useful. Many researchers will still choose to use their own visualisation tools. They offer a lot more functionality and are easier to integrate in e.g. notebooks
More documentation on how to start using VRE, YouTube tutorials
Seems to need a lot of technical skill/knowledge on environment
There is some learning curve, but after that it is fairly good. Integration should be improved, and ease of use of sharing data and types of data understood by the VRE. E.g. JSON, GeoJSON, HTML.
I am not very confident that all the social-media-like features will be used in practice. Teams/communities typically would already have their communication tools (Slack, Jira, etc).
Once output is written to the workspace, it should not be difficult to make use of output which is already there.
I have not use much of the features on this VRE. So I will respond trying to avoid biais but still providing feedback. 3 means "'not used"

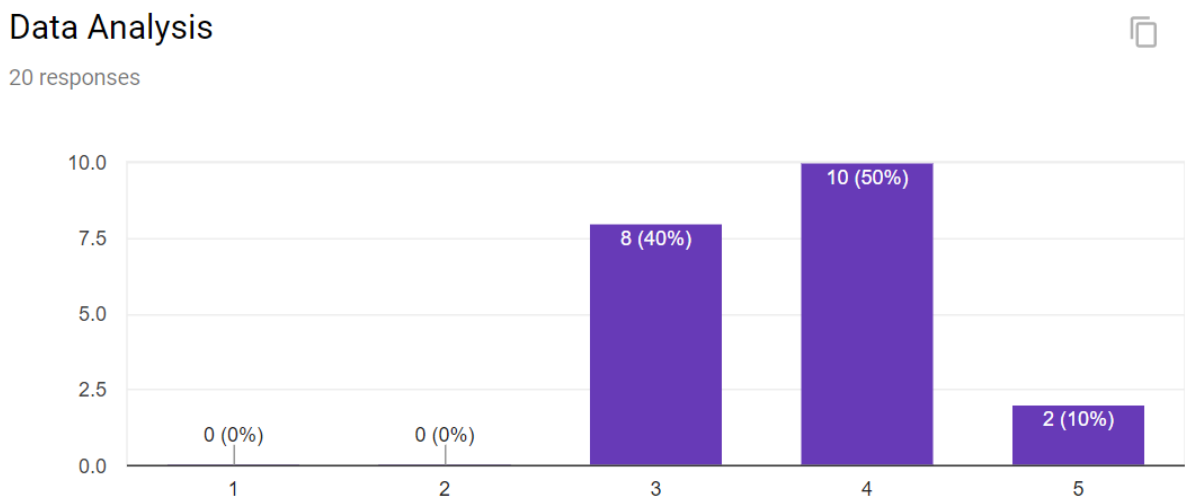
NDVI presentation: default the NDVI standard deviation is presented. Better to combine average and standard deviation in one graph or take the average as default. Reduce WOFOST output to the main ones, with self-explaining names and/or add tool tip

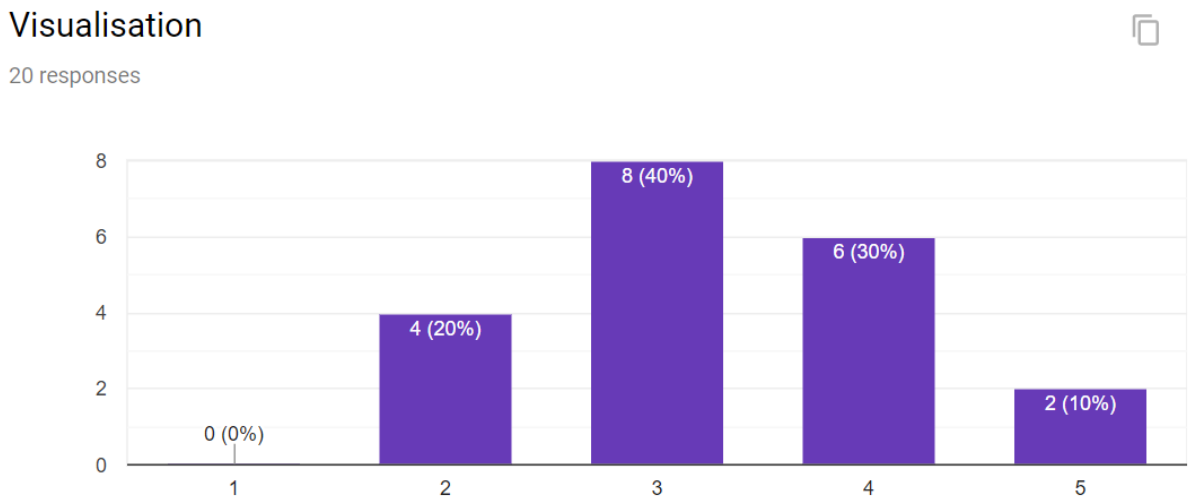
Indicator – Usefulness

Indicator	Evaluation 1	Evaluation 2	Total
Algorithm & Model Development	3.8	4.2	3.9
Algorithm & Model Execution	3.9	4.0	4.0
Data Analysis	3.7	3.7	3.7
Visualisation	3.2	3.5	3.3
Total average – Usefulness	3.7	3.8	3.7









Additional evaluation remarks on Usefulness:

for collaborative modelling it looks perfect; for execution the examples worked on can all easily run locally; advantage of doing that cloud-based unclear
Needs some more data analysis tools (e.g. usable standard DataMiner operators), and more powerful visualisations (that work out-of-the-box).
The primary use I see for VREs is to help easy access to infrastructures. Data analysis and visualization, most probably will also happen with dedicated external tools.
Some specific remarks on the use case of applying WOFOST on field scale in the VRE: - Avoid WOFOST runs without results - Improve parameterization and check input data of WOFOST runs - Add water-limited results of WOFOST

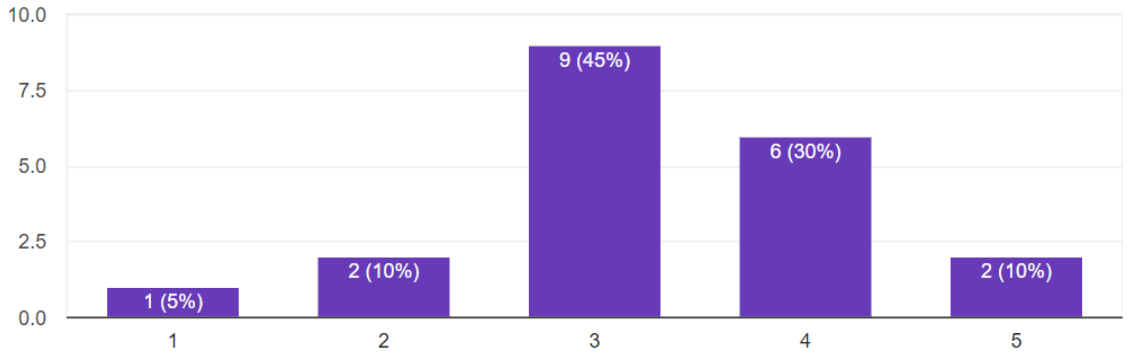
<u>Indicator – Openness</u>

Indicator	Evaluation 1	Evaluation 2	Total
Finding & accessing algorithms, models and data registered in the VRE	3.1	3.7	3.3

Registering and sharing of data, algorithms and models	3.5	3.5	3.5
Adding and sharing new functionality and components (e.g. a new model, visualization)	3.3	3.2	3.3
Total average – Openness	3.3	3.4	3.3

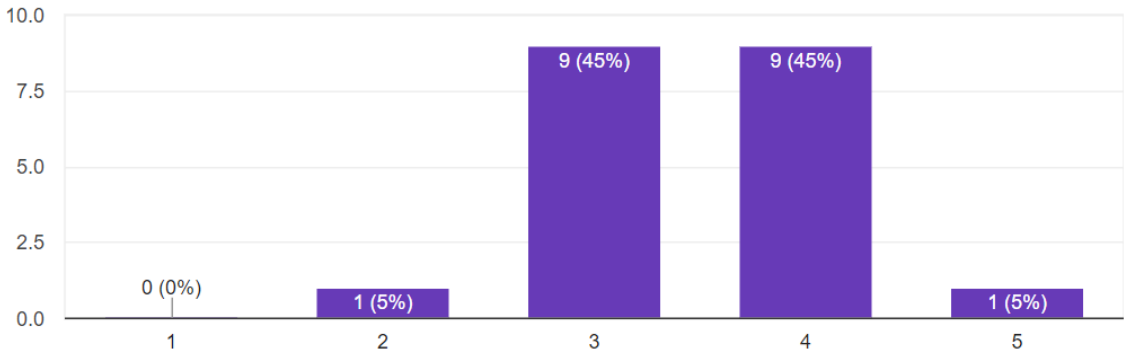
Finding & accessing algorithms, models and data registered in the VRE

20 responses



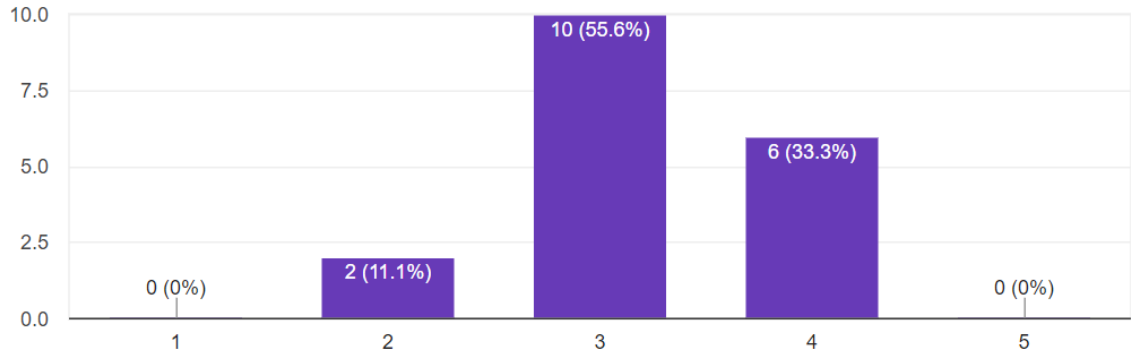
Registering and sharing of data, algorithms and models

20 responses



Adding and sharing new functionality and components (e.g. a new model, visualization)

18 responses



Additional evaluation remarks on Openness:

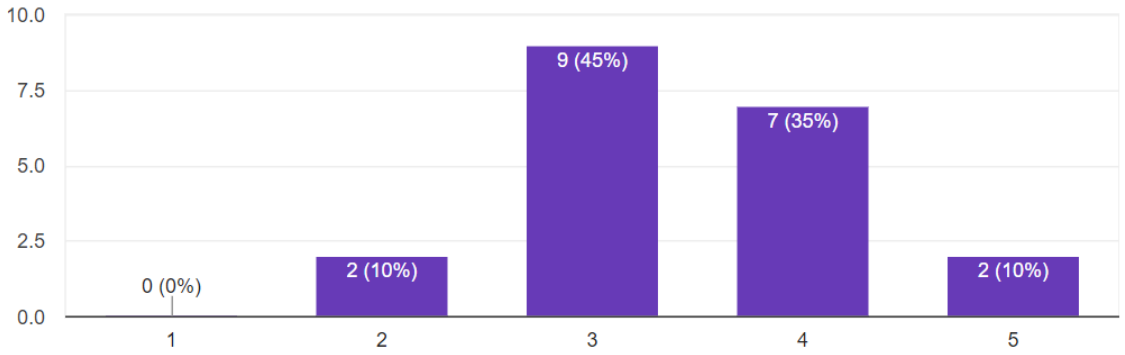
Unexperienced user, answers based on first impression at workshop
did not try to add/register anything
Make better use of existing functionality, and more of semantics
Unfortunately, co-workers are often facing restrictions which have to be removed - sometimes by means of a ticket - before they can access data and / or algorithms.
Updating the help-wiki might help to improve the ease of adding new scripts.
Missing documentation and e.g. online tutorials.
The VRE definitively facilitates this kind of things.
Can't really say it is easy to add a new model algorithm. Moreover if we want it to be distributed. Maybe add some documentation

Indicator – FAIRness

Indicator	**Evaluation 1**	**Evaluation 2**	**Total**
Findability	3.5	3.3	3.5
Accessibility	3.7	3.7	3.7
Interoperability	3.1	3.5	3.3
Reusability	3.4	4.2	3.7
Total average – FAIRness	3.4	3.7	3.5

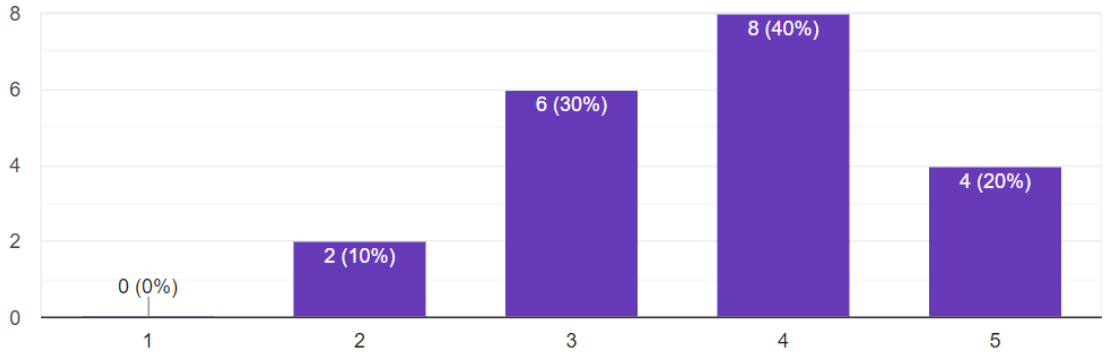
Findability

20 responses



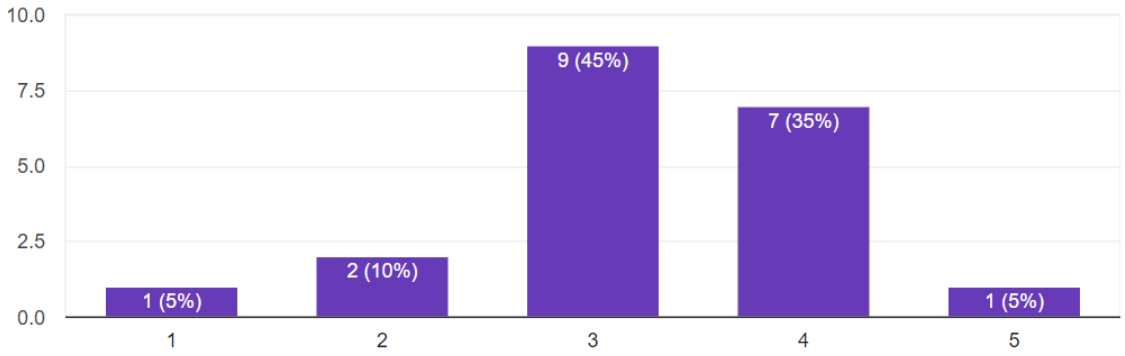
Accessibility

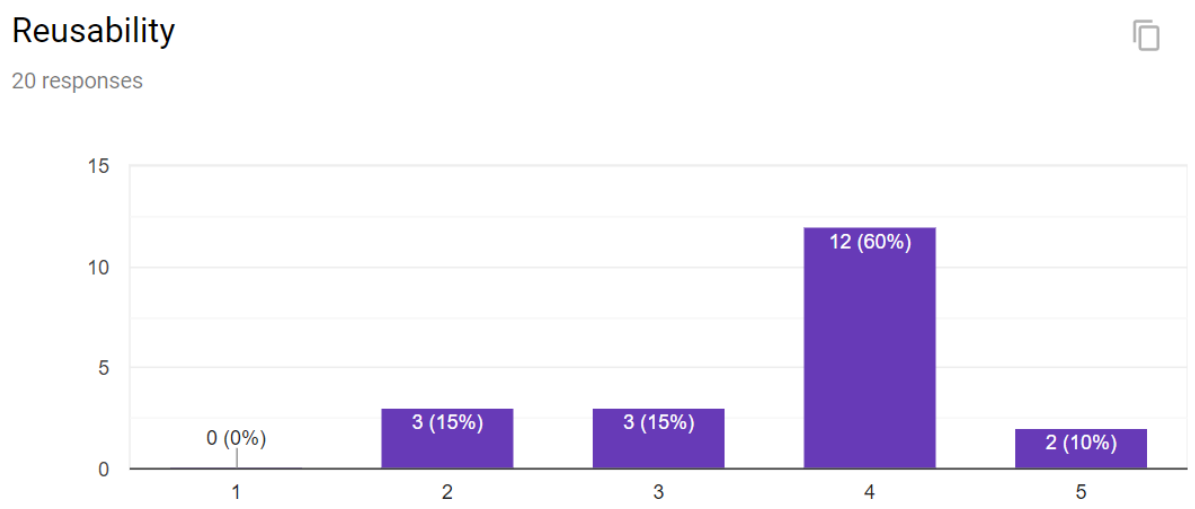
20 responses



Interoperability

20 responses





Additional evaluation remarks on FAIRness:

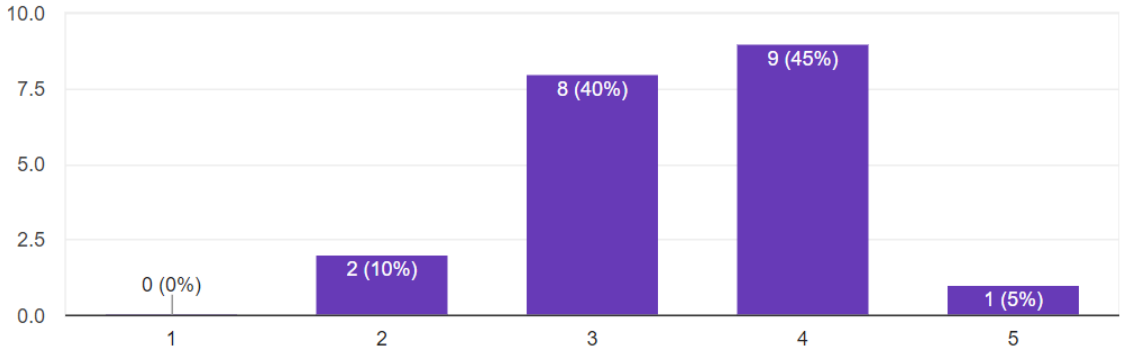
Sharing between catalogues and to the outside world should be possible. More standardisation of data types and formats, and more use of semantics. Use a fair-by-design approach.
Accessibility may be improved by making it easier for users to change permissions on files. Permissions are now granted at the level of individual users. Things would be easier if a user could grant permissions to all users in a group. And as far as I am concerned, it is a major problem that a user's Jupyter notebooks cannot run unless they are hosted at EGI while at the same time such notebooks can apparently not be shared with other users. Sharing of data and code would also be much easier if a user could zip a complete folder in the file browser so that he / she can then make the zip archive available for other users.
I think it needs more semantics to be FAIR

Indicator – Learning Curve

Indicator	**Evaluation 1**	**Evaluation 2**	**Total**
Estimated effort to make an existing model run on the VRE	3.4	3.5	3.5
Estimated effort to configure and run a model and analyse its outputs	3.4	3.2	3.4
Total average – Learning Curve	3.4	3.3	3.4

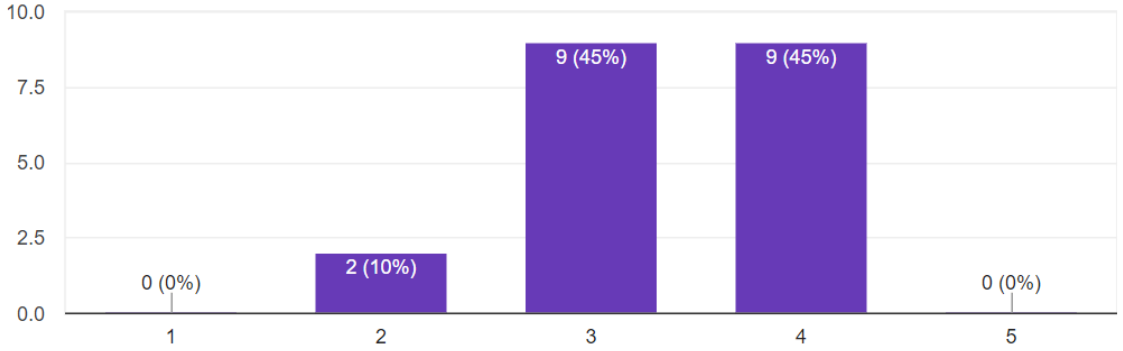
Estimated effort to make an existing model run on the VRE

20 responses



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

20 responses



Additional evaluation remarks on Learning Curve:

Hard to judge yet
The GUI is a bit tricky to learn, but once mastered the process is fairly straightforward. Adapting an existing model to make (better) use of parallel and/or distributed processing takes more effort. But is similar for any kind of distributed computing environments.
I think it needs more semantics to be FAIR
There are indeed a lot of thingies to learn and remember before one can be successful with making an existing model run and with making it available, findable and usable for fellow researchers, such as: fixed names for output files, the need to log off and log on again when an algorithm is published, how to make use of StorageHub facilities from code in a flexible way, the almost constant need to change permissions on files and folders as well as the fact that it is not possible to change algorithm category names once they are entered.
The learning curve differs for type and complexity of the model.
cannot say (don't have much experience in making models)

4.4 SUMMARY AND OBSERVATIONS

4.4.1 General impressions and conclusions

It can be concluded that the VRE functionality and the maturity of the pilot applications after the first phase of pilot development was insufficient for reviewers (even expert reviewers) to really see the added value and appreciate the advantages of using VREs for agricultural modelling. This was, according to the reviewers partly due to the immaturity of the provided VRE components and the lack of integration between different components. More fundamental "showstoppers" were the version and configuration differences between the development environment and the run-time environment and the incompatibility of its file systems, which prohibited the reviewers from deploying and running their model. It resulted in rather low scores for almost all assessed (sub)indicators.

The responses that were collected during the second and third pilot evaluation were in general quite positive. Responding to the feedback from the first evaluation, the major critical issues with VRE compatibility had been solved. Furthermore, the two pilot applications that were developed, were mature and offered a full experience of the agro-climatic modelling workflow. We see the majority of scores now being in the range of 3-4, with little negative (1-2) or extremely positive (5) scores for all indicators.

Clear differences between the scoring of the evaluation indicators after the second and third evaluation cannot be identified. In general, there are slight differences, both in the positive and negative direction. This is probably at least partly due to the relatively short time span of 3 months between the second and third evaluation, so not many additions and improvements were available. Moreover, in general, but particularly for the 3th evaluation, the number of individual evaluations might be too low to draw reliable conclusions.

4.4.2 Observations

In this section we provide a short, qualitative summary of the result of the evaluation survey over the three evaluations. Participants have received an extensive explanation of VREs and the available functions. Community specific agro-climatic modelling pilot applications were demonstrated to show how the VRE can be used to co-develop modelling workflows. Also, the participants had the opportunity to work with the VRE themselves and perform two exercises that made them familiar with practical work in the VRE.

A qualitative analysis of the evaluation survey outcomes leads to the following observations:

<u>Ease of use and usefulness</u>

- Compared to the first evaluation, this indicator is much better scored in the second and third evaluation.
- On average, over the ease of use and usefulness categories, the scores are a bit lower for visualisation compared with the other sub-indicators. We consider this remarkable, because a lot of effort was put in developing a visualisation dashboard for the crop growth modelling pilot demonstrator. There is no clear feedback on possible causes for these relatively low scores, but it might relate to the fact that the options for defining and templating of customizable graphs are limited and below expectation of the average modeler and data

scientist. It is apparently hard to beat freely available visualisation toolkits for R and Python, and dashboarding web applications.

- There seem to be some doubts regarding the ease of use with respect to model and algorithm development for use with DataMiner.
- While evaluators appreciate the options to collaborate, especially model/software developers seem to have higher expectations about being able to collaborate on data analytics and coding. Being able to really access and edit content at the same time (compare Google docs) seems to become the standard. Again, freely available tools like Google Colab and Kaggle notebooks are raising the bar for what users expect. We feel, however, that we need to make the remark here that in this case, considering how notebooks function, easy sharing might make more sense than multiple people trying to edit a single notebook at the same time.
- We observe that some evaluators doubt the usefulness of particularly the collaboration, analytics and visualisation features, because there are already many (good) environments available and in use by researchers. This might be a barrier for researchers to decide on using a VRE and/or to fully exploit its capabilities.
- Several evaluators noted that it was not always obvious how components were to be used, and hinted for easily accessible and easy to use user documentation.

<u>Openness</u>

- We can see a substantial positive difference between the second and third evaluation with regard to the aspect of finding and accessing models, algorithms and data. Moreover, there is quite some dispersion of scores over this sub indicator.

<u>FAIRness</u>

- We can see an average lower score on the sub indicator of interoperability, including some more than average low scoring. There are indications that evaluators link this to a lack of use of semantics in the pilot applications. This is partly due the fact that there has not been much emphasis on semantics in the use cases (semantic resources are still rather scarce in agro-climatic modelling), and partly due to the fact that the semantic functionality that is included (tagging resources with GACS terms) is rather invisible and not very user-friendly.

<u>Learning curve</u>

- Obviously, it is hard to say something on the learning curve of a VRE after half-a-day of demonstrations and a few hours of hands-on work. While evaluators seem to recognize the complexity of a VRE and its many tools, it does not seem to be a big issue given the scoring on this indicator. This might be related to the fact that modelling in general is a quite complex and technical domain, where one is used to having to learn tools and methods with some complexity.

<u>Performance and Scaling</u>

- In both agro-climatic modelling pilot applications, we made use of the option to distribute model runs over the offered computing cluster. While the learning curve for developing such distributed processes is steep, it appeared that the infrastructure could very well handle the submitted processes and associated system loads. Inherent to the use of the WPS (Web Processing Service) standard by DataMiner careful tuning of the batch sizes and job scheduling is however required to avoid overloading the cluster.
- It needs to be stated that the pilot applications have not exploited the VREs gCube features for thin client MapReduce-type of computing. This requires writing Java software using specific gCube interfaces, and we found that this is a technology that is hard to be generally applied outside the VRE (i.e. it would reduce software portability). Also, while it should be possible,

fitting the crop simulation model into a MapReduce workflow would require more Big Data handling capabilities than available in DataMiner. We therefore decided to use the platform agnostic technology (WPS), also part of gCube, to develop this aspect of the pilot.

Reliability

- The overall reliability of the VREs individual components (as far as these were used in the development of pilot applications) was considered good. While several issues were identified during the first evaluation, these were solved in the second pilot development phase.
- A weak point of the VRE configuration we have used seems to be the complexity of the infrastructure and specifically the dependency on distributed infrastructure components that are under the responsibility of different organisations (e.g. CNR, EGI, UoA). In practice, we have experienced critical reliability issues at various points in the process, among others (and most notably) around the pilot evaluations. These were in most cases associated with cross-organisational issues like incompatibility of environments or unscheduled downtime of infrastructure components.

Transparency

- The VRE's catalogue allows for the publication of algorithms, models and data with sufficient metadata to document these for reuse by other researchers.
- The provenance of modelling jobs is registered and stored in such a way that they are traceable and repeatable, which is a valuable asset for making the research process transparent and repeatable and improves reusability.

5 4TH PILOT EVALUATION (VALIDATION) – METHODOLOGY AND MATERIALS

5.1 METHODOLOGY

5.1.1 Introduction

As explained before, the consortium decided to extend the 3rd validation activities with an additional evaluation round. In the first three evaluations a mixed-method approach was used to comply with different needs, and outcomes were not fully harmonized. While this proved to be an effective way to engage the individual communities and to collect their feedback, this also led to a heterogeneous outcome over the pilots with a variety of tools and methods used to collect and report feedback back to our project work. To also make sure that a comparison over communities can be made, the 3rd period assessments were extended with an additional evaluation activity. This validation used a more holistic and harmonized approach, looking at a concrete and well-defined outcome of the project, and using the same evaluation instrument over all pilots to collect feedback from the different stakeholders.

The validation of the agro-climatic modelling pilots developed with the VRE has been performed according to the approach described in the deliverable Harmonized Use Case Validation methods, guidelines and materials (D1.2). For details on the background of the validation methodology, the indicators that are assessed and the specific survey that was developed to collect the required data, we refer to this report. The pilot demonstrators used in the fourth evaluation were mostly the same as the versions used in the third evaluation, due to the short time between evaluations and a shift in focus from development to dissemination activities (also for the technical work packages), no new or updated services were made available to be included in the fourth evaluation).

In terms of this deliverable two "light-weight, remote validation trials" were performed. The main activity was a remote validation trial in a webinar format. Main reason for choosing a remote validation trial is that in general, but specifically due to the late decision to extend the evaluations with this validation, it was expected that this would have the highest impact. The agro-climatic user community is a world-wide community, and opportunities to have physical meetings are scarce. A remote evaluation is less time-consuming and by providing the webinar recording and evaluation resources on-line, it allows also the persons that cannot attend live to participate. The second validation activity was a face-to-face validation workshop with a small group of scientists from the agri-food domain.

5.1.2 Evaluation programme

<u>Evaluation programme - validation webinar</u>

1. *Welcome and contents of the webinar*
   - Including the announcement of the validation survey
2. *Short introduction on AGINFRA+, virtual research and VREs (5 min)*
   - Contents of the webinar
   - AGINFRA+ - project, communities and infrastructure
   - Short walkthrough of major VRE functions for agro-climatic modelling
     - Collaborate: access to VREs, timeline, shared workspace
     - Develop & Automate: Jupyter notebooks, DataMiner, agro-climatic modelling dashboard, issue tracking

    - Publish & Share: VRE Catalogue
    - Scale: VRE as a service, computing cluster, shared storage, D4Science environment
  - Resources: AgroDataCube, WOFOST model
3. *Demonstrator – Crop phenology estimation with remote sensing data*
  - Co-developing data science algorithms with Jupyter notebooks
    - Jupyter notebooks for crop phenology estimation
    - Dutch (AgroDataCube) and European (SentinelHub) implementations
  - Deploying and running the algorithms
    - Deploying developed algorithms in DataMiner
    - Running crop phenology estimation for a single agricultural parcel
    - Running crop phenology estimation for a region (distributed computing)
4. *Demonstrator – Crop growth modelling*
  - Running the WOFOST model for a single field
  - The agro-climatic modelling dashboard
    - Visualisation of geospatial data and timeseries
    - Running of the DataMiner algorithm from using dashboard applications
    - Analysing model outputs
  - Distributed crop growth modelling
    - Running the WOFOST model for a region or country
    - Viewing and analysing model output
  - Sharing results – Publishing models and data through the VRE catalogue
5. *Question & Answers session*
  - *Live Q&A session at the end of the webinar*
  - *Option to submit questions after the webinar*
6. *Evaluation Survey*
  - (to be performed by participants after the webinar)

<u>Evaluation programme – validation workshop</u>

1. *Introduction (30 min)*
  - Digital science tools in food & agriculture: the AGINFRA+ digital infrastructure
2. *Demo (30 min)*
  - Digital science tools to support simulation of crop growth: the WOFOST model case
    - AGINFRA+ WOFOST Software Demonstration Scenario
    - WUR Software Demonstrator (video)
3. *Questions & Answers (15 min)*
  - Discussion about the WUR Software Demonstrator
4. *Wrap up, Validation Survey (15 min)*

5.1.3 Validation indicators and survey questions

As part of the validation methodology developed for this phase, a set of 32 evaluation questions was provided to the participants. Table 8 shows the questions that were part of the performed validation survey. The detailed definition of the developed validation methodology, the survey questionnaire and the indicators that they support can be found in the deliverable Harmonized Use Case Validation methods, guidelines and materials (D1.2).

Table 8 – List of validation questions for validation

Survey Question	Answer Type
I would find such a virtual research environment useful in my job.	1-2-3-4-5 (strongly disagree – strongly agree)
Using such a virtual research environment would enable me to accomplish tasks more quickly.	1-2-3-4-5 (strongly disagree – strongly agree)
Using such a virtual research environment would increase my productivity.	1-2-3-4-5 (strongly disagree – strongly agree)
If I used a virtual research environment, I would increase my chances of getting a better position or salary.	1-2-3-4-5 (strongly disagree – strongly agree)
My interaction with such a virtual research environment would be clear and understandable.	1-2-3-4-5 (strongly disagree – strongly agree)
It would be easy for me to become skilful at using such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
I would find such a virtual research environment easy to use.	1-2-3-4-5 (strongly disagree – strongly agree)
Learning to operate such a virtual research environment would be easy for me.	1-2-3-4-5 (strongly disagree – strongly agree)
Using such a virtual research environment is a good idea.	1-2-3-4-5 (strongly disagree – strongly agree)
A virtual research environment makes work more interesting.	1-2-3-4-5 (strongly disagree – strongly agree)
Working with such a virtual research environment is fun.	1-2-3-4-5 (strongly disagree – strongly agree)
I would like working with such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
People who influence my behaviour would think that I should use such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
People who are important to me would think that I should use such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
The senior management of my organisation would be supportive of using such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
In general, my organization would support the use of such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
I have the resources necessary to adopt and use a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
I have the knowledge necessary to adopt and use such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
The virtual research environment does not seem compatible with other systems I use.	1-2-3-4-5 (strongly disagree – strongly agree)
In my organisation, a specific person (or group) would be available to assist me with difficulties in using such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
I could complete a job or task using this virtual research environment if there was no one around to tell me what to do as I go.	1-2-3-4-5 (strongly disagree – strongly agree)
I could complete a job or task using this virtual research environment if I could call someone for help if I got stuck.	1-2-3-4-5 (strongly disagree – strongly agree)
I could complete a job or task using this system if I had a lot of time to complete the job for which the software was provided.	1-2-3-4-5 (strongly disagree – strongly agree)
I could complete a job or task using this virtual research environment if I had in my organisation a facility for assistance.	1-2-3-4-5 (strongly disagree – strongly agree)
I feel apprehensive about using such a virtual research environment.	1-2-3-4-5 (strongly disagree – strongly agree)
It scares me to think that I could lose a lot of information using such a virtual research environment by hitting the wrong key.	1-2-3-4-5 (strongly disagree – strongly agree)
I hesitate to use such a virtual research environment, fearing to make mistakes I cannot correct.	1-2-3-4-5 (strongly disagree – strongly agree)
Such a virtual research environment looks somewhat intimidating to me.	1-2-3-4-5

	(strongly disagree – strongly agree)
I intend to use such a virtual research environment in the next 12 months.	1-2-3-4-5 (strongly disagree – strongly agree)
I predict I would use such a virtual research environment in the next 12 months.	1-2-3-4-5 (strongly disagree – strongly agree)
I plan to use such a virtual research environment in the next 12 months.	1-2-3-4-5 (strongly disagree – strongly agree)
My gender is:	Male – Female – Prefer not to say
My age is:	Free text

5.1.4 Used materials

The material presented below in Table 9 was used to support the validation programme that was presented in this section. With the shown reference the specific item can be accessed or downloaded from the Internet.

Table 9 – Reference material used in the AGINFRA+ agro-climatic modelling pilot evaluations

Material	Reference	Type
Webinar recording	https://channel.royalcast.com/webcast/wur/20191219_1/	video
Validation survey	https://forms.gle/vf9h9Er8cgcwaS1A7	GF
Introduction presentation	https://www.slideshare.net/RobLokers/introduction-webinar-cloud-based-agricultural-modelling-207721375	PPT
Agro-climatic modelling VRE (trial version)	https://aginfra.d4science.org/explore	VRE
Crop Phenology Estimation demonstration video (part 1)	https://www.youtube.com/watch?v=UzONcaKK4So&feature=youtu.be	video
Crop Phenology Estimation demonstration video (part 2)	https://www.youtube.com/watch?v=UzONcaKK4So&feature=youtu.be	video
Crop growth modelling demonstration video	https://www.youtube.com/watch?v=2uuGDn0oWU8&feature=youtu.be	video
AgroDataCube	http://agrodatacube.wur.nl/	PDF
WOFOST crop growth model	https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Facilities-Products/Software-and-models/WOFOST.htm	VRE folder

6 VALIDATION RESULTS

6.1 INTRODUCTION

This section provides and overview of the results of the additional evaluation round and the harmonized validation exercise. It provides an overview of the characteristics and statistics of the evaluation session and provides the results of the performed survey. The detailed definition of the developed validation methodology, the survey questionnaire and the indicators that they support can be found in the deliverable Harmonized Use Case Validation methods, guidelines and materials (D1.2). Survey results have been qualitatively analysed and are reported as a series of observations.

6.2 EVALUATION OUTCOMES

6.2.1 Evaluation details

1) Validation webinar

Location: on-line, via webinar facilities Wageningen UR
Date: 19 December 2019
Method: light-weight validation, webinar and validation survey
Registrations: 371
Live participants: 133
Total viewers: 261 (of which 100+ watched the full webinar)
Survey responses: 42

A mail was sent to all registered persons (separately to the live participants and the remaining registrants) with links to the recorded webinar and related material, and a reminder to fill in the validation survey. A reminder for the survey was sent again 2 weeks after the webinar.

2) Validation workshop

Location: Agricultural University of Athens, Athens, Greece
Date: 19 December 2019
Method: light-weight validation, short workshop and validation survey
Participants: 13
Survey responses: 13

6.2.2 Indicators – Evaluation Scores

Table 10 below, shows the summarized responses for the validation webinar per validation survey question, including the number of votes and the mean score over all respondents.

Table 10 – Survey results of the validation webinar

Question	1	2	3	4	5	No answer	Total nr of votes	Mean score
I would find such a virtual research environment useful in my job.	1	0	2	12	27	0	42	4.52

Question	1	2	3	4	5	No answer	Total nr of votes	Mean score
Using such a virtual research environment would enable me to accomplish tasks more quickly.	1	0	2	21	18	0	42	4.31
Using such a virtual research environment would increase my productivity.	1	0	4	21	16	0	42	4.21
If I used a virtual research environment, I would increase my chances of getting a better position or salary.	0	4	18	12	8	0	42	3.57
My interaction with such a virtual research environment would be clear and understandable.	1	2	11	20	8	0	42	3.76
It would be easy for me to become skillful at using such a virtual research environment.	1	1	6	20	14	0	42	4.07
I would find such a virtual research environment easy to use.	0	2	17	15	8	0	42	3.69
Learning to operate such a virtual research environment would be easy for me.	1	2	12	19	7	1	41	3.71
Using such a virtual research environment is a good idea.	1	0	0	14	26	1	41	4.56
A virtual research environment makes work more interesting.	1	1	5	16	18	1	41	4.20
Working with such a virtual research environment is fun.	0	1	10	15	14	2	40	4.05
I would like working with such a virtual research environment.	0	1	3	8	5	25	17	4.00
People who influence my behavior would think that I should use such a virtual research environment.	1	7	12	15	7	0	42	3.48
People who are important to me would think that I should use such a virtual research environment.	1	8	11	14	8	0	42	3.48
The senior management of my organisation would be supportive of using such a virtual research environment.	0	2	17	15	7	1	41	3.75
In general, my organization would support the use of such a virtual research environment.	1	2	12	19	8	0	42	3.74
I have the resources necessary to adopt and use a virtual research environment.	4	11	9	12	4	2	40	3.03
I have the knowledge necessary to adopt and use such a virtual research environment.	3	7	11	12	8	1	40	3.37
The virtual research environment does not seem compatible with other systems I use.	4	9	18	6	4	1	40	2.93
In my organisation, a specific person (or group) would be available to assist me with difficulties in using such a virtual research environment.	8	9	14	5	5	1	40	2.76
I could complete a job or task using this virtual research environment if there was no one around to tell me what to do as I go.	3	7	18	8	5	1	40	3.12
I could complete a job or task using this virtual research environment if I could call someone for help if I got stuck.	2	2	13	16	8	1	40	3.63
I could complete a job or task using this system if I had a lot of time to complete the job for which the software was provided.	3	3	13	13	8	2	40	3.50
I could complete a job or task using this virtual research environment if I had in my organisation a facility for assistance.	2	1	12	16	9	2	40	3.73
I feel apprehensive about using such a virtual research environment.	11	5	9	11	3	3	39	2.74
It scares me to think that I could lose a lot of information using such a virtual research environment by hitting the wrong key.	13	15	9	2	2	1	41	2.15
I hesitate to use such a virtual research environment, fearing to make mistakes I cannot correct.	12	17	8	2	2	1	41	2.15

Question	1	2	3	4	5	No answer	Total nr of votes	Mean score
Such a virtual research environment looks somewhat intimidating to me.	16	14	7	2	2	1	41	2.02
I intend to use such a virtual research environment in the next 12 months.	1	2	14	10	14	1	41	3.83
I predict I would use such a virtual research environment in the next 12 months.	1	2	13	14	11	1	41	3.78
I plan to use such a virtual research environment in the next 12 months.	1	2	14	11	13	1	41	3.80

Table 11 shows the summarized responses for the validation workshop per validation survey question, including the number of votes and the mean score over all respondents.

Table 11 – Survey results of the validation workshop

Question	1	2	3	4	5	No answer	Total nr of votes	Mean score
I would find such a virtual research environment useful in my job.	0	0	4	5	4	0	13	4.00
Using such a virtual research environment would enable me to accomplish tasks more quickly.		1	4	6	2	0	13	3.69
Using such a virtual research environment would increase my productivity.		1	5	6	1	0	13	3.54
If I used a virtual research environment, I would increase my chances of getting a better position or salary.	1	4	5	2	1	0	13	2.85
My interaction with such a virtual research environment would be clear and understandable.		1	5	7		0	13	3.46
It would be easy for me to become skillful at using such a virtual research environment.			4	8	1	0	13	3.77
I would find such a virtual research environment easy to use.			2	8	3	0	13	4.08
Learning to operate such a virtual research environment would be easy for me.			2	9	2	0	13	4.00
Using such a virtual research environment is a good idea.				10	3	0	13	4.23
A virtual research environment makes work more interesting.		2	2	7	2	0	13	3.69
Working with such a virtual research environment is fun.		1	8	1	3	0	13	3.46
I would like working with such a virtual research environment.			3	9	1	0	13	3.85
People who influence my behavior would think that I should use such a virtual research environment.	1	1	5	5	1	0	13	3.31
People who are important to me would think that I should use such a virtual research environment.	1	1	4	6	1	0	13	3.38
The senior management of my organisation would be supportive of using such a virtual research environment.		1	3	6	3	0	13	3.85
In general, my organization would support the use of such a virtual research environment.			6	6	1	0	13	3.62
I have the resources necessary to adopt and use a virtual research environment.		1	3	7	2	0	13	3.77
I have the knowledge necessary to adopt and use such a virtual research environment.			4	8	1	0	13	3.77
The virtual research environment does not seem compatible with other systems I use.		5	5	3		0	13	2.85
In my organisation, a specific person (or group) would be available to assist me with difficulties in using such a virtual research environment.		2	4	7		0	13	3.38

Question	1	2	3	4	5	No answer	Total nr of votes	Mean score
I could complete a job or task using this virtual research environment if there was no one around to tell me what to do as I go.		3	3	5	2	0	13	3.46
I could complete a job or task using this virtual research environment if I could call someone for help if I got stuck.			4	9		0	13	3.69
I could complete a job or task using this system if I had a lot of time to complete the job for which the software was provided.			4	7	2	0	13	3.85
I could complete a job or task using this virtual research environment if I had in my organisation a facility for assistance.			4	8	1	0	13	3.77
I feel apprehensive about using such a virtual research environment.	5	4	3	1		0	13	2.00
It scares me to think that I could lose a lot of information using such a virtual research environment by hitting the wrong key.	2	7	2	2		0	13	2.31
I hesitate to use such a virtual research environment, fearing to make mistakes I cannot correct.	1	9	3			0	13	2.15
Such a virtual research environment looks somewhat intimidating to me.	2	8	3			0	13	2.08
I intend to use such a virtual research environment in the next 12 months.		6	5	2		0	13	2.69
I predict I would use such a virtual research environment in the next 12 months.	1	2	7	3		0	13	2.92
I plan to use such a virtual research environment in the next 12 months.	1	4	5	3		0	13	2.77

6.3 SUMMARY AND OBSERVATIONS

In this section we provide a short, qualitative summary of the result of the validation survey. Participants have received a short introduction of VRE features and its opportunities to work together on data science. Moreover, cases were demonstrated that show how the VRE was used to develop community specific pilot applications. This has provided participants with a good impression of what working with a VRE means, how data science tasks and workflows can be performed and what kind of applications can be developed. It should however be noted that it probably remains hard for most participants to really see the technical, scientific and organisational consequences of using VREs. Therefore, we have chosen not to draw decisive conclusions here, but rather to present a number of observations.

A qualitative analysis of the validation survey outcomes leads to the following observations:

- After learning about the features and possible ways to collaborate with a VRE and seeing the community specific pilot applications that were developed with support of the VRE, respondents seem to be positive regarding the usefulness of a VRE for their work. They generally support the idea that it would help them getting more productive.
- We also observe that respondents seem to be willing to use a VRE for their work from a more personal perspective. They indicate that working with a VRE would be interesting and fun, which can be seen as an indication of their personal motivation and is a good basis for any new challenge to start from.
- Answers regarding the “external factors” influencing the use of the VRE are less pronounced. Answers regarding how relevant external persons (people influencing their behaviour,

important persons, senior management) or their organisations would motivate or support the respondent to use a VRE show quite some spread of scores, although tending to the positive side.

- We might conclude that there is no strong convincement that the use of a VRE would influence the position or salary of the respondents. However, this might also be a side effect of respondents (mostly researchers) having less interest in promotion and being content with their salary and/or position.
- Respondents seem to differ in their opinions on how easy it would be to adopt VRE in their work. In general, the answers regarding availability of knowledge and of resources required to adopt and use a VRE show a large spread. Many seem not to be sure that they would be able to easily adopt a VRE and quite some respondents indicate that they lack sufficient knowledge and resources. A same pattern can be observed regarding the need for support. They seem to have doubts on being able to adopt a system like this without support and many indicate that it would become less difficult if sufficient support and sufficient time was available.
- In line with the previous observation, there are quite some respondents who indicate feeling apprehensive about using a VRE. This could well be associated with the impression that they expect a steeper learning curve and they lack knowledge, resources and support to get productive with a VRE. This can be underpinned by the observation that they seem not to be intimidated by a VRE and are not afraid to make critical mistakes using the VRE. We must note here here that the specific question might have not been well understood by respondents, and that this explains the remarkable difference between the answers regarding "feeling apprehensive" and "feeling intimidated".
- Respondents do not seem to be sure about the compatibility of a VRE with the other systems they work with. Few strongly agree or disagree, while many choose to score the "middle option".
- It is remarkable that respondents answer quite positively on their expectation with regard to use a VRE in the next 12 months. It reconfirms the positive attitude towards the VRE, its usefulness and opportunities, and might be fuelled by the rising popularity of many other web based working environments such as Google Docs, Google Colab, Kaggle Notebooks, Google Earth Engine, etc. It is a bit harder to explain from the reservations that quite some respondents seem to have with regard to the learning curve and the knowledge and resources that would be required to maser working with the VRE.

7 DISCUSSION AND RECOMMENDATIONS

7.1 DISCUSSION AND ACHIEVEMENTS

In the past three years, the AGINFRA project has introduced virtual research environments as a collaborative tool for open data science in three communities in agri-food research. For the agro-climatic modelling community that was targeted in Work Package 5 of the project, it was for most members the first time they learned about virtual research and the opportunities of using VREs for collaborative data science. In the past years, many community members have heard about the backgrounds, seen demonstrations of the VRE and its agro-climatic modelling pilot applications and had the opportunity to provide their opinions and thoughts through evaluation and validation surveys. A group of evaluators also had the chance to get a more in-depth view, did guided hands-on work in the VRE and responded to the evaluation survey in the three evaluation events that were organised.

For the evaluation, the community specific pilot evaluations that were described in the assessment plan for the agro-climatic modelling pilots were aligned with the specific stage of evolution and maturity of the pilot applications under development. This allowed to better engage the community users and attract attention for the pilots. Therefore, reviews of separate components were performed with experts in the first evaluation, while in the later evaluations full research workflows could be demonstrated. These later evaluations allowed to organize hands-on workshops where data scientists practically worked with the VRE pilots and obtained a better perspective on the practicalities of working with such an environment. To cope with the differences in the evaluation approach between the different research communities (besides agro-climatic modelling, also the food security and the food safety community were evaluating community specific pilots), an additional validation was developed and performed in the 3rd evaluation period in the form of more holistic and harmonized validation exercise.

The outcomes of the pilot evaluations clearly show higher appreciation of the VRE pilots in the later stages of development. This can in our view be at least partly explained by the maturity level of the evaluated applications, from separate modules in the early stages of the project to full research workflows in the later stage. Early evaluators rated the various evaluation indicators relatively low, and they observed substantial shortcomings when it came to the potential to develop real-world applications of agro-climatic modelling and data science. In the later evaluation rounds, evaluations turned out to be more positive. We believe that this was due to the fact that at that stage, evaluators could see the VRE supporting the full modelling workflow, and supporting open science with collaborative, analytical and sharing facilities and scaling.

From the evaluation feedback that we received we can conclude that in general the impressions and opinions about the AGINFRA+ VREs and the developed pilot application are overall positive. We observe that community members in general state that VREs can be a useful tool for digital open science, and we see a broad willingness and enthusiasm to start using such environments for research. There is also a lot of interest for the developed pilot applications, both from the perspective of novel technologies used in the modelling pilots, e.g. for distributed modelling, and from the more scientific perspective of used methodologies.

The developed pilot applications show very well the potential and opportunities of collaborative, open science for the agro-climatic modelling community. They also allowed community members to see how different VRE modules can be connected to create valuable applications and scientific workflows. Because we have demonstrated the VRE and the pilot applications to a broad audience that also includes data scientist that (partly) work on the broader agro-environmental domain, we can also

conclude that concepts and applications are well received and useful beyond the agro-climatic modelling community. Many seem to acknowledge that virtual research offers unique options to collaborate, co-develop and share, responding to the demand for working together over different domains with heterogeneous data that is required to tackle the grand interdisciplinary challenges in agri-food and environment. Participants in evaluations were in general impressed with the broad range of tools and underlying infrastructure becoming available from one access point, and the opportunity to collaborate and share over a broad range of activities of data science workflows.

For the agro-climatic modelling community we can conclude that through AGINFRA+ valuable steps have been taken to introduce virtual research environments to groups of scientists that were generally not aware of the opportunities of virtual research. As a project, supported by previous outcomes from infrastructural and innovation actions, we were able to show the general concept of virtual research, and how VREs can support collaborative teams in performing data science. Moreover, we have demonstrated through the developed community driven pilot applications how a VRE and its components can be used to develop scientific workflows and catalyse open science. Steps like this could only be made because of the collaborative and iterative process implemented in AGINFRA+, where infrastructure developers collaborated with developers and data scientists from scientific communities in agri-food.

With regard to the developed pilot applications, it is worth mentioning that several features were specifically appreciated:

- The offered opportunities for encapsulating and publishing algorithms through DataMiner and the D4Science catalogue add to the development of digital open science. Towards the future this could add FAIR algorithms to the already known concept of FAIR data.
- The generic mechanisms for distributed modelling, developed to showcase the exploitation of compute clusters in agro-climatic modelling and the application of that mechanism in the pilot applications to demonstrate upscaling of modelling. They show how high-resolution field level modelling can be upscaled to larger regions or even to global scale.
- The opportunity to implement the concept of "explorative modelling", where data scientists can use virtual, shared literate programming environments to co-develop, test and improve their work as a team. Moreover, the option to subsequently isolate the final version of algorithms and models and deploy and publish it with DataMiner adds to the support of the full data science workflow.
- The opportunity to use and combine different approaches to developing applications. For example, the option to use the VRE graph module for simple visualisations, but also more sophisticated options, from using visualisation libraries as part of shared programming environments (Jupyter, RStudio) to developing a customized app (e.g. the modelling and visualisation dashboard developed as part of one of the agro-climatic modelling pilots).

7.2 RECOMMENDATIONS

Many researchers, both agro-climatic modellers and more general data scientists in the area of agro-environmental research, business analysts and information intermediaries have received demonstrations and worked hands-on with the AGINFRA+ VRE and the developed pilot applications. A clear distinction between different groups cannot be made. In many cases evaluators have mixed profiles, and the breadth and depth of the hands-on work they did differs. We have therefore not made the distinction between different groups of evaluators. It seemed useful, however, to distinguish between evaluation results and impressions that were collected from the evaluation and

validation sessions and the experiences of the VRE developers who have worked on a daily basis with the VRE as data scientists and engineers developing the pilot applications.

7.2.1 The modeler and data science perspective

Modelers and data scientists have been able to interact with the AGINFRA+ work at two levels. In the evaluation exercises they had the opportunity to hear about and work with the VRE and pilot applications for a whole day. In the validation events this was limited to one to two hours and the focus was on demonstration, rather than on getting practical experience through hands-on exercises. This was important to keep in mind, both in interpreting the survey results and in drawing conclusions. It is also good to keep this in mind when reading the recommendations given in this section.

We believe that the following issues, either directly resulting from feedback, or indirectly interpreting from discussions with evaluators are relevant for follow-up. Specific recommendations are provided in case relevant.

- There are several indications that the user experience can be improved by better integration of individual VRE components and services. From the perspective of evaluators, a clear message was given regarding the interface between some of the data science components and the shared file system (e.g. from Jupyter notebooks). Evaluators found the access to the shared workspace cumbersome. Recommendation: better integration of tools and shared workspace, so it acts as an easily accessible file system in all cases.
- Many evaluators called for what we could describe as a "truly collaborative" environment where more users can for instance work together in the same code and see changes in real-time. Expectations seem to go strongly into the direction of "Google docs or Colab like" behaviour.
  Recommendation: to consider if collaborative features of data science tools like Jupyter and RStudio can be improved.
- Some evaluators question if specific features, like the social or issue management functions would be used, as there are already systems in place in their working environment that are used for that purpose (e.g. Jira, Slack). This seems to match with the answers on the validation survey question on compatibility.
- Evaluators in general expect to have a steep learning curve when starting to use VREs. There is also a demand for more and better accessible support information and tutorials.
  Recommendation: to provide more and clearer support information and to make it more easily discoverable.
- Quite some evaluators question the usefulness of the standard VRE visualisation modules in practice, with lack of configuration options and difficulty of defining, templating and reuse as main issues. They expect rather to use options in Jupyter or RStudio, or visualisation tools that are currently outside the VRE, e.g. Plotly.
  Recommendation: to improve the functionality and configurability of graphs, to allow defining graph templates and parameterisation, or to add a "richer" visualisation module.
- There are indications that evaluators expect more integrated semantics, not only isolated tools, for a VRE to be FAIR. Particularly the function to link data and algorithms to vocabularies or ontologies was missed.
  Recommendation: to allow linkages between the catalogue and ontologies, so they can be tagged and linked, and semantic search can be implemented.

7.2.2 The developer perspective

In contrary to most of the evaluators, the application developers and data scientists that were involved in the AGINFRA+ project, have had the chance actually to work with the D4Science VREs for a longer period. Consequently, they had a much more in-depth experience with the functional and technical features of VREs. While they had the chance to steer some of the VRE developments in the past years, and many features were added or adapted to fit the needs of the user communities, there are relevant issues that have not been tackled in the project. Like the participants in the evaluation, they had the chance to fill in the validation and evaluation surveys, but due to the numbers their opinions and results will not stand out. We believe that it is therefore useful to also separately document their main experiences and recommendations separately.

We believe that the following issues, either directly resulting from feedback, or indirectly interpreting from discussions with the involved developers and data scientists are relevant for follow-up. Specific recommendations are provided in case relevant.

- Involved developers and data scientists note that there are still integration issues to be solved. Several tools that are offered through the VRE are still "stand alone" tools. Examples are the lack of linkage between semantic tools and "content", like datasets and algorithms and the fact that there is no option to create graph templates and call these from other tools or applications.
  Recommendation: continue to improve tool integration (from the user perspective).
- A specific point of attention regarding integration concerns the shared workspace. It is still cumbersome to access the shared workspace as a real file system from many tools. Other issues that are mentioned are: being able to change file/directory permissions on group level, and being able to move/copy directories.
  Recommendation: continue to improve integration of the shared workspace with other VRE components and improvement of file management functions
- It is currently not possible to properly structure the output of algorithms and models. It is "single silo based", which complicates the development of algorithms and the further processing of data.
- The support for storing, processing and visualising geospatial data is currently not sufficient and does not meet the needs of the user community. Tools are currently not integrated, important functions for visualisation of spatial data are not working properly and (large) files cannot be handled with acceptable performance.
  Recommendation: continue to improve the functionality and integration of geospatial tooling.
- From the technical perspective and specifically compatibility and portability considerations, developers felt that several often-used technologies are currently not available for VRE developers, e.g. running a Spark cluster, deployment of Docker containers, access to GPU resources e.g. for machine learning applications. As an example, we decided to not exploit the VREs gCube features for thin client MapReduce-type of computing when developing the pilot applications. This requires writing Java software using specific gCube interfaces, and we found that this is a technology that is hard to be generally applied outside the VRE (i.e. it would reduce software portability).
  Recommendation: to consider if such commonly used technologies would be useful for a broader range of user communities and applications and if these can be added to the infrastructure.
- The ease of use of deploying and reusing algorithms could be increased by improvements like the option to use named parameters, offering a list of available (versions of) interpreters that

are supported, instead of a free-text field. For non-experienced users, a mechanism that is simpler than the current WPS interface would make the reuse much easier.

ANNEX 1 – EVALUATION EXERCISES

This annex contains the manuals that were supporting the AGINFRA+ evaluation exercises for the agro-climatic modelling community. It shows which aspects of the VRE and pilot applications have been used in the hands-on work sessions. Also, future users that are interested can request access to the D4Science agro-climatic VRE, and perform these exercises to get a feel of the benefits of the VRE for their purposes.