D7.4- Food Security Pilot Evaluation Report

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

DELIVERABLE NUMBER	D7.4
DELIVERABLE TITLE	Food Security Pilot Evaluation Report
RESPONSIBLE AUTHOR	Alice Boizet (INRA)

GRANT AGREEMENT N.	731001
PROJECT ACRONYM	AGINFRA PLUS
PROJECT FULL NAME	Accelerating user-driven e-infrastructure innovation in Food & Agriculture
STARTING DATE (DUR.)	01/01/2017 (36 months)
ENDING DATE	31/12/2019
PROJECT WEBSITE	http://www.plus.aginfra.eu
COORDINATOR	Nikos Manouselis
ADDRESS	110 Pentelis Str., Marousi GR15126, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Mr. Christophe Doin

WORKPACKAGE N. | TITLE	WP7 | Food Security Community
WORKPACKAGE LEADER	INRA
DELIVERABLE N. | TITLE	D7.4 | Food Security Pilot Evaluation Report
RESPONSIBLE AUTHOR	Alice Boizet (INRA)
REPLY TO	Alice.boizet@inra.fr
DOCUMENT URL	http://www.plus.aginfra.eu/sites/plus_deliverables/D5.1.pdf
DATE OF DELIVERY (CONTRACTUAL)	29 June 2018 (M18), 31 December 2019 (M36)
DATE OF DELIVERY (SUBMITTED)	02 July 2018 (M19), 27 January 2020 (M37)
VERSION | STATUS	2.0 | final
NATURE	ORDP (Open Research Data Pilot)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Alice Boizet (INRA)

VERSION	MODIFICATION(S)	DATE	
0.1	Initial Version	25/05/2018	Alice Boizet (INRA)
0.2	Objective and procedure description	29/05/2018	Alice Boizet (INRA)
0.3	Review	14/06/2018	Rob Lokers
1.0	Final version	27/06/2018	Alice Boizet (INRA)
1.1	Initialising version 2	22/12/2019	Alice Boizet (INRA)
1.2	Reporting final evaluation	24/12/2019	Alice Boizet (INRA)
1.3	Review	27/01/2020	Rob Lokers (WUR)
2.0	Final Version for submission	31/12/2019	Alice Boizet (INRA)

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

ACRONYMS LIST

FAIR	Findable, Accessible, Interoperable, Reusable, as set of principles acting as an international guideline for high quality data stewardship
SAI	Statistical Algorithms Importer
VRE	Virtual Research Environment
BRAPI	Breeding API, specification of standard API for plant phenotype/genotype databases

EXECUTIVE SUMMARY

This document describes the evaluation procedure performed in the AGINFRA+ project to assess the VRE developed for the Food Security community and reports on the outcomes of the different evaluation phases.

The Food Security VRE features have been implemented by iteration. Three phases of evaluation have been performed in order to assess the VRE at different stages of its development and then being able to take into account the users feedbacks for adding improvements. This also allowed to better engage the community users. The first evaluation consisted in testing each available feature independently while in the two later evaluations full research workflows could be demonstrated and it was possible to organize hands-on workshops where users could follow a real user scenario and get a better view of a VRE potential.

To cope with the differences in the evaluation approach between the different research communities (agro-climatic modelling, food safety and food security community), an additional evaluation was developed and performed in harmonized validation format.

Analysing the evaluation feedback, we can conclude that for the Food Security community, the AGINFRA+ project enabled to introduce virtual research environments to scientists who were generally unaware of the possibilities of virtual research. We were able to demonstrate the general concept of a VRE, and how VREs can support collaborative teams in plant science research. In addition, we demonstrated how a VRE and its components can be used to discover and access data and develop scientific workflows. We observe that community members state that VREs can be a useful tool for open science, and we see an interest to use such environments for research. Many seem to recognize that virtual research environments offer unique opportunities to share heterogeneous data and analysis workflows and thus make them more easily reusable, responding to the demand for collaboration in different fields with heterogeneous data needed to address major interdisciplinary challenges in agri-food and environment.

TABLE OF CONTENTS

INTRODUCTION 9
1 PILOT EVALUATION AND VALIDATION – OBJECTIVES AND PROCESS 10
1.1 INTRODUCTION 10
1.2 METHODOLOGY FOR PILOTS EVALUATIONS 10
1.3 METHODOLOGY FOR VALIDATION 11
2 PILOT EVALUATION – METHODOLOGY AND MATERIALS 12
2.1 INTRODUCTION 12
2.2 METHODOLOGY FOR FIRST EVALUATION 12
2.2.1 Evaluation programme 12
2.2.2 Evaluation Indicators and Survey Questions 13
2.3 METHODOLOGY FOR SECOND EVALUATION 13
2.3.1 Evaluation programme 13
2.3.2 Evaluation indicators and survey questions 13
2.4 METHODOLOGY FOR THIRD EVALUATION 14
2.4.1 Evaluation programme 14
2.4.2 Evaluation indicators and survey questions 14
3 EVALUATION RESULTS 15
3.1 INTRODUCTION 15
3.2 EVALUATION 1 OUTCOMES 15
3.2.1 Evaluation details 15
3.2.2 Global evaluation feedbacks 15
3.2.3 Survey results 16
3.3 EVALUATION 2 OUTCOMES 17
3.3.1 Evaluation details 17
3.3.2 Evaluation results 18
3.4 EVALUATION 3 OUTCOMES 19
3.4.1 Evaluation details 19

3.4.2 Evaluation results 19
3.5 SUMMARY OF THE THREE EVALUATIONS 20
3.5.1 Global Results 20
3.5.2 Global Observations 26
4 VALIDATION (4TH PILOT EVALUATION) METHODOLOGY 27
4.1 INTRODUCTION 27
4.2 EVALUATION PROGRAMME 27
4.3 EVALUATION INDICATORS AND SURVEY QUESTIONS 27
5 VALIDATION RESULTS 29
5.1 INTRODUCTION 29
5.2 EVALUATION OUTCOMES 29
5.2.1 Evaluation details 29
5.2.2 Evaluation results 29
5.3 SUMMARY AND OBSERVATIONS 30
6 DISCUSSION AND RECOMMENDATIONS 32
6.1 DISCUSSION AND ACHIEVEMENTS 32
6.2 RECOMMENDATION 32
ANNEX 1 – EVALUATION 1 USER SCENARIO 34
ANNEX 2 – FILLED EVALUATION 1 SHEETS 36
ANNEX 3 – EVALUATION 2 EXERCISE 39
ANNEX 4 – EVALUATION 3 EXERCISE 50

LIST OF TABLES

Table 1 - List of evaluation indicators 11
Table 2: Evaluation 1 material 13
Table 3: Evaluation 2 material 14
Table 4: Evaluation 3 material 14
Table 5: Evaluation 1 results 16
Table 6: Evaluation 2 results 18
Table 7: Evaluation 3 results 19
Table 8: Validation questions 27
Table 9: Validation survey results 29

INTRODUCTION

As part of the project of AGINFRA+, the virtual research environment called Food Security VRE has been deployed for the plant phenotyping community.

This document describes the evaluation procedures and evaluation outcomes performed in the AGINFRA+ project to assess the Food Security VRE and the pilot applications developed for the community.

This document begins with a description of the main objectives of the assessment and the methodology of the different stages of the evaluation as well as the validation stage. The second part describes in more detail the three pilot evaluations carried out. The third part reports on the results of the three evaluations. The last two sections present the methodology and results of the validation phase, which consisted of a 4th additional evaluation. This was carried out using a validation method harmonised between the three use cases (agro-climatic modelling, food safety and food security).

1 PILOT EVALUATION AND VALIDATION – OBJECTIVES AND PROCESS

1.1 INTRODUCTION

In order to assess the effectiveness of using a VRE for research in the high-throughput phenotyping community, an evaluation approach has been performed in three piloting phases. This part describes the objectives and the process followed for the evaluation of the three piloting phases and for the validation phase.

1.2 METHODOLOGY FOR PILOTS EVALUATIONS

The performed evaluations to assess Food Security 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.

In the first piloting phase, the main objective was to collect the expectations of potential users of the VRE on Food Security This phase was carried out in two stages. First of all, an informal evaluation was made during a workshop organized with 5 complementary persons that were representative of the phenomics community: a statistician, 2 field managers of phenotyping and greenhouse platforms, a remote sensing researcher, an ecophysiologist. The VRE concept and the VRE platform for food security were presented. Then an open discussion followed on their first impressions of the Food Security VRE to gather their needs. In a second step, two people were invited to test the VRE by following a very simple use scenario adapted to the components available at that time. The testers who were selected were people who showed an interest in new technologies and were looking for tools that could make their job easier. After testing the VRE, they responded to a survey based on indicators that are described in Table 1.

In the second piloting phase, the implementation of the functionalities was sufficiently advanced for the VRE to be tested by evaluators with different levels of expertise and to carry out practical exercises. The main objective was to determine whether any important functionality was missing and to identify improvements to be implemented in the third pilot. For this evaluation, a workshop was organized to introduce the VRE and demonstrate a real phenotyping research scenario. Then, each participant was invited to follow this same user scenario while evaluating each item tested. Participants were then able to provide feedback by completing a survey.

The third piloting phase was originally supposed to focus on semantic features as it was described in D7.3 but we chose to focus on the 2nd evaluation scenario in order to deliver fully functional applications for the plant phenotyping 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.

Indicators

The evaluation was performed along a set of assessment indicators and evaluation questions that are targeted to the WP7 target community. Table 1 below presents these indicators.

Table 1 - List of evaluation indicators

Indicator	Description
Ease of Use	Assessing the ease of use of the Food Security VRE consists in answering question: "How easy is it to use the VRE?" In other words: "Is the VRE user-friendly?"compared to existing tools, taking account of the new features interest.
Learning Curve	The learning curve indicator is complementary to the ease of use indicator. It contributes to the evaluation of the time needed to learn how to use the features which are not easy to use right away because the user needs to be trained.
Usefulness	The usefulness indicator scores the functionalities the VRE provides. This indicator is used to check if every VRE feature is really relevant to the use case of high-throughput phenotyping. In order a feature to be accessible outside of the VRE, this indicator should also evaluate the feasibility of having this functionality inside the VRE.
Openness	The openness indicator evaluates the contribution of features to openness which means: "Is it easy to add new data or new functionality?
FAIR-ness	The FAIR principles (Findability, Accessibility, Interoperability and Reusability) are key principles for open science on any VRE. This indicator should therefore evaluate 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.
Uptake potential	This indicator gives us feedback on how likely it is that the person assessing the pilot would also be willing to use it after the end of the project. It is closely correlated to the usefulness indicator.

Contrary to what was planned and described in the Assessment Plan deliverable (D7.3), the "scalability and performance" indicator could not been assessed during the evaluation sessions as it takes more than half a day of VRE usage to assess this indicator.

1.3 METHODOLOGY FOR VALIDATION

The fourth evaluation of the Food Security VRE has been performed according to the approach described in the deliverable Harmonized Use Case Validation methods, guidelines and materials (D1.2). This deliverable describes the validation methodology with the indicators that are assessed and the specific survey that was developed to collect the required data.

Due to the short time available to organize this last evaluation and the difficulties encountered in getting a large number of people to participate in the previous workshops, it was decided to hold a webinar. The high throughput phenotyping community is small, but the resulting data can be used by a larger community. The people directly involved in the production of the data are few in number and insufficiently informed, which made it very difficult to have a large number of people from this community at our various workshops. The webinar was a good opportunity for people who had not been able to attend previous workshops to have a demonstration of the VRE.

2 PILOT EVALUATION – METHODOLOGY AND MATERIALS

2.1 INTRODUCTION

This section provides a description of the used methodology for the three pilots evaluations. The methodology of each evaluation has been adapted to the state of implementation of the VRE functionalities. This explains why the user scenarios initially planned for the 3rd evaluation in particular have been redesigned.

2.2 METHODOLOGY FOR FIRST EVALUATION

2.2.1 Evaluation programme

First Global Evaluation
A workshop has been set up with 5 complementary representative persons of the phenotyping community (statistician, phenotyping platforms managers, remote sensing researcher, ecophysiologist).

First, the AGINFRA+ project and the concept of VRE were introduced. Then an experimented user of a VRE presented the tunaAtlas VRE. This is a VRE that has also been deployed in the D4science platform for the BlueBridge project. This VRE provides services to discover the available datasets from various Regional Fisheries Management Organizations. This presentation helped the other participants to better understand the potential of a VRE. Finally, the Food Security VRE was presented and the participants gave their first opinion and expectations of this VRE.

This a non-exhaustive list of the questions asked to the participants:

- Do you think this VRE could be useful for your work?
- What features are you really interested in? For what purpose?
- Do you think some features are missing?

User Evaluation
This evaluation involved two individuals testing several features of the VRE. The concept of the VRE and the Food Security VRE were first introduced to the two participants: one of them had participated in a previous workshop and the other was briefly briefed just before the evaluation session. A document outlining the steps to follow to test the VRE was provided to both testers.

The more generic components were tested. Some components were not completely integrated to the VRE (for example, the access to the workspace from Jupyter and Rstudio was missing).
The steps of the evaluation were:

- Share and collaborate with others members (message posting, messaging system)
- Share datasets with others members (shared workspace)
- Access to different data sources (ReFindit, dataminer algorithm GetPlantHeight_From_PhenomeAPI)
- Developing analytics scripts and share them as black boxes (Rstudio, Jupyter, SAI)

User #1	User #2
Plant Phenotyping Platform Manager	Statistician in Plant Breeding
Didn’t attend the workshop	Did attend the workshop

Each step was not very detailed on purpose in order to see if the user could easily find how to use the features.

After the hands-on session, the testers responded to a survey to score each tested feature.

2.2.2 Evaluation Indicators and Survey Questions

The survey consisted in scoring the features on a scale from 1 to 5 (insufficient to very good) on five selected indicators (ease of use, learning curve, usefulness, openness, FAIRness). The testers also had to answer to open questions that were helpful to better understand and sometimes temper the indicators scores. The exercise and the survey questions are reported in Annex 1 and Annex 2. The material presented in Table 2 was used to support the evaluation programme that was presented above.

Table 2: Evaluation 1 material

Material	Reference	Type
VRE presentation	https://data.d4science.net/ZyCB	PPT
Tuna atlas VRE presentation	https://data.d4science.net/eVSd	PPT
Food Security VRE	https://aginfra.d4science.org/web/foodsecurity	VRE
Exercises	https://data.d4science.net/iWwC	PDF
Survey questions	https://data.d4science.net/iWwC	PDF

2.3 METHODOLOGY FOR SECOND EVALUATION

2.3.1 Evaluation programme

A workshop was set up within the INRA phenotyping community. The evaluation lasted half a day. First, the AGINFRA+ project and the VRE concept were presented. Then, there was a brief description of the main features available in the Food Security VRE. The INRA team working on the AGINFRA+ project gave a demonstration of the VRE following the user scenario described below. During the hands-on session, participants were invited to follow a similar user scenario to test all relevant functionalities. The hands-on exercise is presented in Annex 3. The workshop ended with an evaluation survey.

2.3.2 Evaluation indicators and survey questions

For this evaluation, the survey covered every feature that users should have tested by following the given user scenario. This evaluation focussed on 2 indicators: the ease of use and the usefulness. Indeed, the first evaluation showed that it was difficult for the testers to assess the other indicators, and the initial feedback we received indicated that developments should focus on the usability of the tools. To assess usefulness, this indicator had been divided into two sub-indicators: the usefulness of the functionality in general and the usefulness of that functionality inside the VRE. The objective was to assess the actual benefit of having the feature fully integrated within the VRE versus having the same feature outside the VRE. In this way, users were able to see that a VRE is not just a platform hosting different tools, but that these tools can actually connect.

Some open questions followed the scoring questions in order to have a more global opinion on the Food Security VRE. For this evaluation, the online survey management tool of the VRE was used.

The material presented in Table 3 was used to support the evaluation programme that was presented above. With the shown reference the specific item can be accessed or downloaded.

Table 3: Evaluation 2 material

Material	Reference	Type
Food Security VRE	https://aginfra.d4science.org/web/foodsecurity	VRE
Exercises	https://data.d4science.net/dxHf	PDF
Survey questions	https://data.d4science.net/fL8W	PDF

2.4 METHODOLOGY FOR THIRD EVALUATION

2.4.1 Evaluation programme

A workshop has been set up collaboratively with WUR partners in order to embrace a larger public to the two respective communities. Twelve people participated in the workshop, which lasted a full day. The first half of the day was devoted to providing the evaluators with sufficient information to understand the objectives of AGINFRA+, the concepts of virtual research and the technology behind it, as well as the purpose of the pilot applications developed. It also included demonstrations of the two respective VREs: Agro-climatic Modelling VRE and Food Security VRE.
The second half of the evaluation programme was a hands-on session. The evaluators were asked to carry out practical exercises that covered the main aspects of VRE and working with the pilot applications. Short exercises were prepared in the form of a manual, guiding participants in their work. The exercise of the Food Security VRE is described in Annex 4. At the end of the day, participants were invited to complete an evaluation questionnaire.

2.4.2 Evaluation indicators and survey questions

For this evaluation, WP5 and WP7 agreed to have separate (one per community) but similar questionnaires in order not to disturb the participants. WP7 therefore decided to align the questions in its survey with the questions in the WP5 survey. In order to limit the number of questions while evaluating the functionalities on more indicators, the functionalities were grouped by category. The indicators evaluated were the following: Ease of use, Usefulness, Openness, FAIRness and Learning curve.

The material presented in Table 3 was used to support the evaluation programme that was presented above. With the shown reference the specific item can be accessed or downloaded.

Table 4: Evaluation 3 material

Material	Reference	Type
Introduction presentation	https://data.d4science.net/aw3T	PPT
VRE presentation	https://data.d4science.net/XhGB	PPT
Food Security VRE	https://aginfra.d4science.org/web/foodsecurity	VRE
Exercises	https://data.d4science.net/pZ9t	PDF
Survey questions	https://forms.gle/D9nnmPnx951nc8fW6	Google form

3 EVALUATION RESULTS

3.1 INTRODUCTION

This chapter provides the results of the three performed evaluations. Given the different evaluation approaches that were followed in the different phases (the tested features and the evaluated indicators could be different), the evaluation outcomes are presented independently for each evaluation. In order to get more valuable results, we combined the 3 evaluations results into one. The last sub-section presents the overall results.

3.2 EVALUATION 1 OUTCOMES

3.2.1 Evaluation details

	Informal evaluation	User evaluation
Location: Date: Method: Participants:	INRA Montpellier, France April 2018 Workshop with open questions 5	INRA Montpellier, France June 2018 Guided walkthrough and review 2

3.2.2 Global evaluation feedbacks

This part reports the workshop participants' feedbacks.

Isabelle Sanchez (Statistician):
She finds the collaborative functionalities very useful.
Having Rstudio deployed inside the VRE is something really important to her. She can have a R environment with all the necessary libraries and access to it from any machine. This is also very convenient to give training in R.
She expressed reserves on the groups organization because the phenotyping community is hierarchical. Some issues could concern only one kind of platforms for example.

Romain Chapuis (Phenotyping platform manager):
According to him, a VRE could be a good tool to share on measurements protocols, problems with sensors, etc. He raised the issue of data ownership.

Philippe Burger (Phenotyping platform manager):
He finds the collaborative functionalities very important.

Frédéric Barret (Remote sensing researcher):
According to him, a VRE is a good way to share information. For example, if they have problems with their phenotyping machines they could share the way they solve it with others.
He could get a lot of data from other platforms in order to be able to do deep learning for example.

Pierre-Etienne Alary (Ecophysiologist, Emphasis project partner):
He thinks that a VRE could be a good tool to collaborate inside the Emphasis project. The collaborative features could be very useful for this community.

3.2.3 Survey results

Table 5: Evaluation 1 results

Feature category	Functionality	Indicator	user #1	user #2	average score
Collaboration & communication	messages publication	Ease of use	5	5	5
		learning curve	5	5	5
		usefulness	5	4	4,5
		openness	4	4	4
		FAIR-ness	3	2	2,5
	messaging system	Ease of use	5	5	5
		learning curve	5	5	5
		usefulness	5	4	4,5
		openness	1	3	2
		FAIR-ness	1	1	1
	shared workspace	Ease of use	5	5	5
		learning curve	5	5	5
		usefulness	5	5	5
		openness	4	4	4
		FAIR-ness	4	2	3
Data Exploration	discovery tool	Ease of use	5	4	4,5
		learning curve	5	5	5
		usefulness	5	3	4
		openness	5	1	3
		FAIR-ness	2	2	2
	data access	Ease of use	3	4	3,5
		learning curve	3	5	4
		usefulness	2	5	3,5
		openness	4	5	4,5
		FAIR-ness	4	4	4
Data Visualization	Visualization tool	Ease of use	-	5	5
		learning curve	-	5	5
		usefulness	-	4	4
		openness	-	4	4
		FAIR-ness	-	3	3
Developing Environments	Developing Environments	Ease of use	2	4	3
		learning curve	2	5	3,5
		usefulness	4	3	3,5
		openness	3	4	3,5
		FAIR-ness	3	2	2,5
Workflow building	SAI	Ease of use	2	2	2
		learning curve	2	3	2,5
		usefulness	4	5	4,5
		openness	4	4	4
		FAIR-ness	3	3	3

3.3 EVALUATION 2 OUTCOMES

3.3.1 Evaluation details

Location:	INRA Montpellier, France
Date:	June 20th 2019
Method:	Pilot application workshop, hands-on session, survey on the VRE
Participants:	5
Survey responses:	5

This assessment was performed during a workshop set up within the INRA phenotyping community. The expected number of participants was 15. Unfortunately, only 5 participants were able to attend the workshop. In spite of this, the workshop was very informative because it allowed for discussion with each participant and it was easier to help them all during the hands-on session.

The evaluation lasted half a day. First, the AGINFRA+ project and the VRE concept were presented. Then, there was a brief description of the main features available in the Food Security VRE. The INRA team working on the AGINFRA+ project gave a demonstration of the VRE following the user scenario described below. During the hands-on session, participants were invited to follow a similar user scenario to test all relevant functionalities. The hands-on exercise is presented in Annex 3. The workshop ended with an evaluation survey.

3.3.2 Evaluation results

Table 6: Evaluation 2 results

Features group	Functionality	Indicator	User #1	User #2	User #3	User #4	User #5	Average score
Collaboration & Communication	Messages posting	ease of use	4	2	5	4	3	3,6
		usefulness	5	2	5	5	2	3,8
		usefulness in the VRE	5	2	1	5	2	3
	Shared workspace	ease of use	2	4	5	3	4	3,6
		usefulness	5	5	5	4	5	4,8
		usefulness in the VRE	5	5	5	4	5	4,8
Data exploration & access	Refindit	ease of use	3	5	4	4	4	4
		usefulness	4	3	5	5	3	4
		usefulness in the VRE	4	2	5	4	2	3,4
	brapi algo to access data	ease of use	3	3	3	2	3	2,8
		usefulness	5	5	5	4	5	4,8
		usefulness in the VRE	5	5	5	5	5	5
Developing environment	Rstudio	ease of use	2	1	5	5	2	3
		usefulness	5	4	5	5	4	4,6
		usefulness in the VRE	5	4	5		5	4,75
	Jupyter	ease of use	3	3	5	3	4	3,6
		usefulness	5	4	5	5	4	4,6
		usefulness in the VRE	5	5	5	5	5	5
		FAIRness	3	3	1	5	3	3
workflow building	SAI	ease of use	1	3	3	3	2	2,4
		usefulness	5	5	3	5	5	4,6
		usefulness in the VRE	5	5	4	5	5	4,8
		FAIRness	3	3	3	5	3	3,4
	Galaxy	ease of use	-	2	2	4	3	2,75
		usefulness	-	5	4	5	5	4,75
		usefulness in the VRE	-	5	4	5	5	4,75
		FAIRness	-	2	3	5	2	3
Catalogue	Catalogue	ease of use	-	4	4	4	3	3,75
		usefulness	-	3	5	4	3	3,75
		usefulness in the VRE	-	3	5	5	4	4,25
		FAIRness	-	3	4	2	3	3
Data Visualization	Data Visualization	ease of use	-	2	3	3	2	2,5
		usefulness	-	4	3	5	3	3,75
		usefulness in the VRE	-	4	3	5	4	4
Semantics tools	Semantics tools	ease of use	-	-	-	5	4	4,5
		usefulness	-	-	-	5	5	5
		usefulness in the VRE	-	-	-	5	5	5

The users were also asked to give their global opinion about the VRE by evaluating these 4 questions:

- The VRE is really easy to use (1 if you disagree, 5 if you agree)
- The VRE helps to collaborate with other researchers (1 if you disagree, 5 if you agree)
- The available tools in the VRE are completely integrated (1 if you disagree, 5 if you agree)
- The VRE helps to make FAIR data or processes (1 if you disagree, 5 if you agree)

They were finally asked if they would be interested to use a VRE. Of the five interviewees, four indicated an interest in using the VRE.

3.4 EVALUATION 3 OUTCOMES

3.4.1 Evaluation details

Location:	Paris, France
Date:	September 18th 2019
Method:	Pilot application workshop, hands-on session, online survey
Participants:	12
Survey responses:	5

3.4.2 Evaluation results

Table 7: Evaluation 3 results

indicator	Feature Group	user #1	user #2	user #3	user #4	user #5	Average score
Ease of Use	Collaboration & communication	4	5	3	4	4	4
	Data exploration & Data extraction	4	3	3	3	3	3,2
	DataMiner algorithm execution	4	3	3	5	4	3,8
	Developing Environments (Rstudio, Jupyter)	4	3	4	4	5	4
	Data Visualization	2	5	3	1	3	2,8
	Galaxy workflow building and execution	4	3	4	4	5	4
Usefulness	Collaboration & communication	4	5	5	5	5	4,8
	Data exploration & Data extraction	4	3	3	3	5	3,6
	DataMiner algorithm execution	4	3	3	4	5	3,8
	Developing Environments (Rstudio, Jupyter)	4	3	4	4	5	4
	Data Visualization	2	5	3	1	5	3,2
	Galaxy workflow building and execution	4	3	5	3	5	4
Openness	galaxy workflows and data registered in the VRE	4	3	4	3	4	3,6
	Registering and sharing of data, analytic scripts or workflows	4	3	4	4	3	3,6
	Adding and sharing new functionality and components (e.g. a new shiny application, new charts)	3	3	3	2	4	3
FAIRness	Findability	3	5	2	4	3	3,4
	Accessibility	3	5	4	3	3	3,6
	Interoperability	4	5	2	2	4	3,4
	Reusability	4	5	4	4	4	4,2
Learning curve	Estimated effort to make an existing dataminer algorithm or galaxy workflow run on the VRE	3	3	2	2	2	2,4
	Estimated effort to configure a new dataminer algorithm or a new galaxy workflow	3	3	2	4	3	3

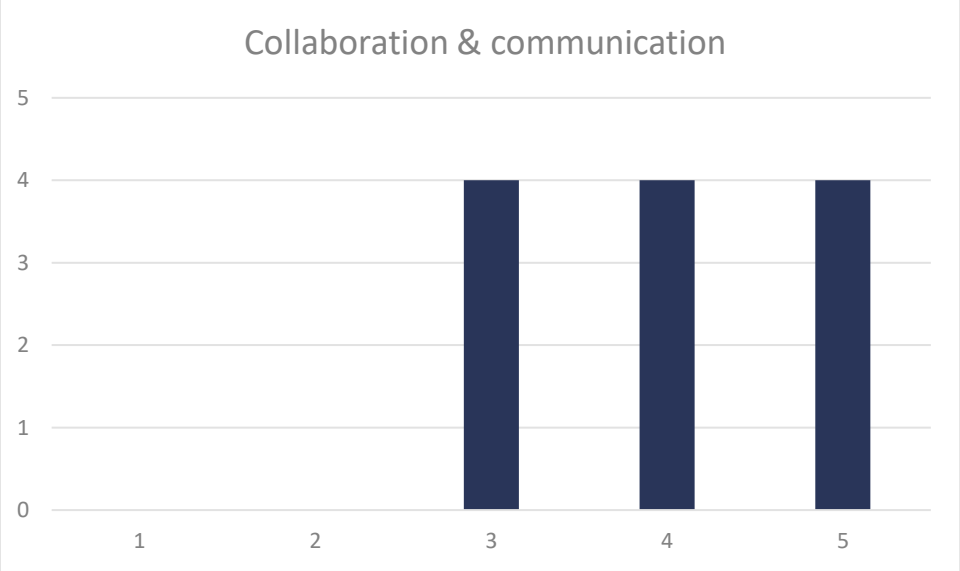
3.5 SUMMARY OF THE THREE EVALUATIONS

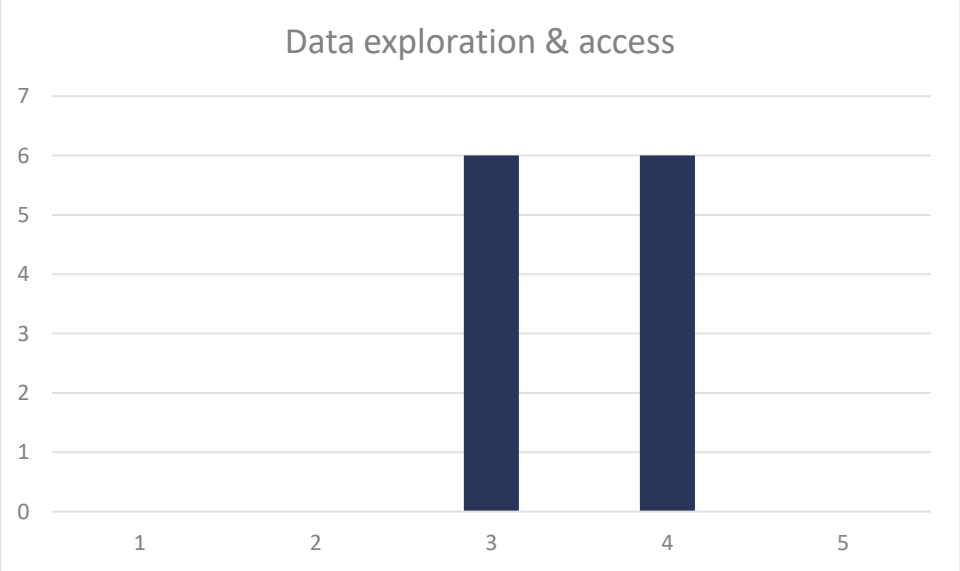
3.5.1 Global Results

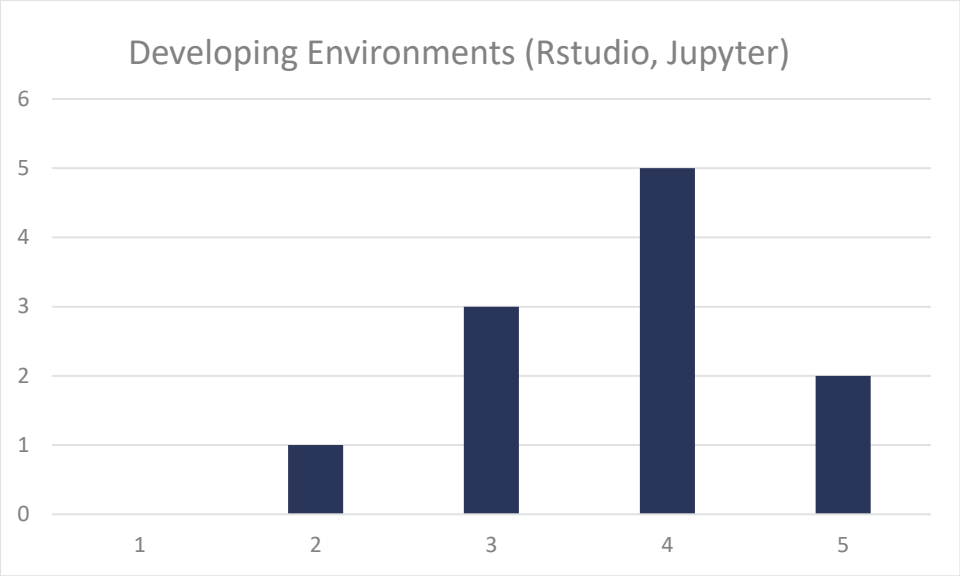
In order to combine the results of the three evaluations, we have grouped the features evaluated in the first 2 sessions under the corresponding categories used in the third session. We calculated the average score of each category for each user, rounded up the score. The detailed scores are presented in the charts below. This method was used for the indicators "Ease of use" and "Usefulness" which were assessed at every evaluation.

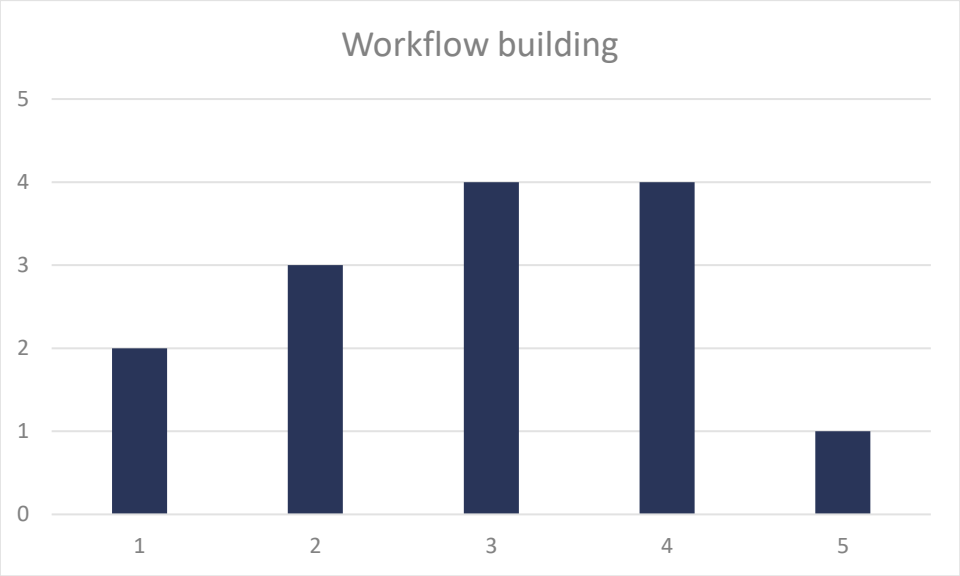
Indicator – Ease of use:

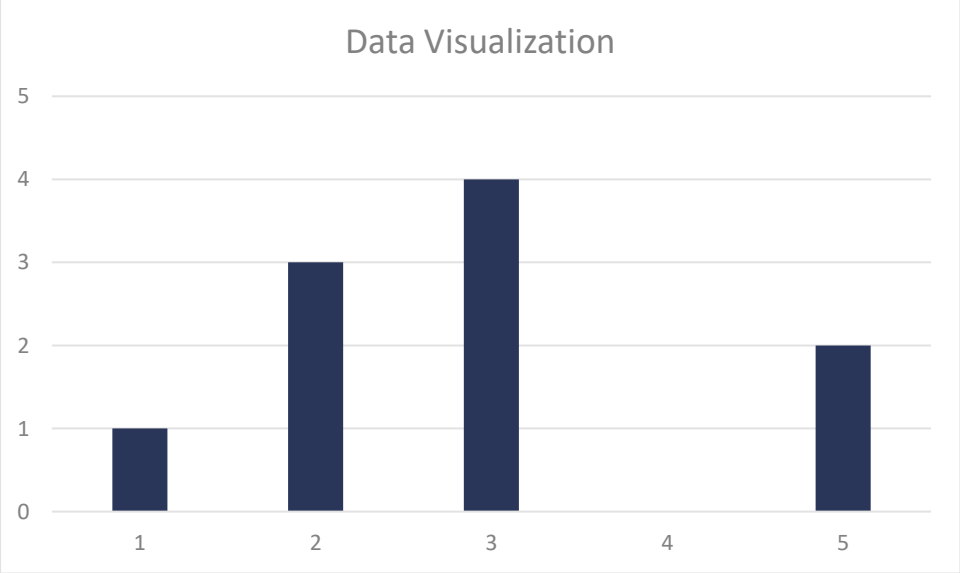
	Evaluation 1	Evaluation 2	Evaluation 3	Combined evaluations
Collaboration & communication	5	3.6	4	4.00
Data exploration & access	2.66	3.4	3.2	3.5
Developing Environments	3	3.3	4	3.45
Workflow building	2	2.6	4	3.42
Data Visualization	5	2.5	2.8	2.9
Total Average Ease of Use	**3.53**	**3.08**	**3.6**	**3.46**






Developing Environments (Rstudio, Jupyter)
6
5
4
3
2
1
0
1
2
3
4
5


Workflow building
5
4
3
2
1
0
1
2
3
4
5

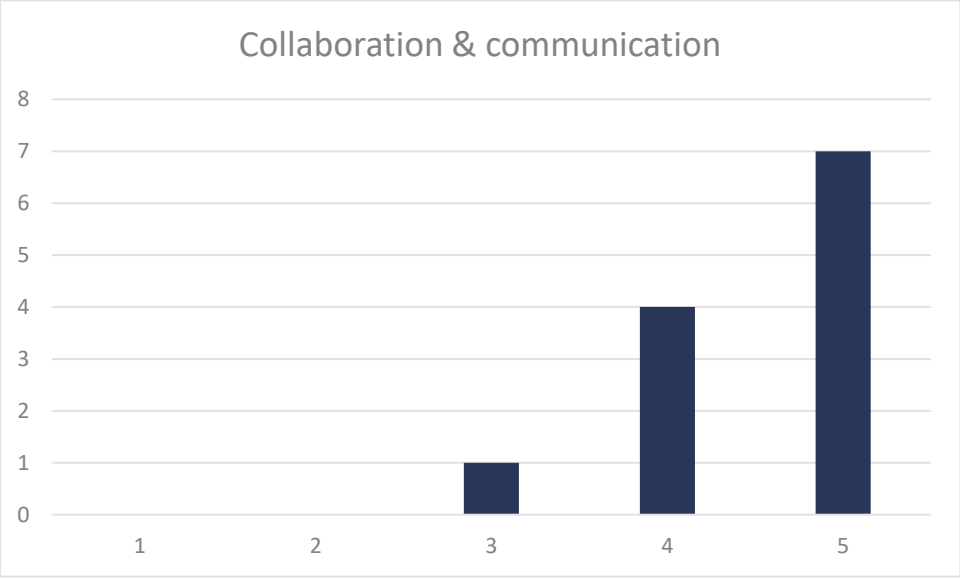

Data Visualization
5
4
3
2
1
0
1
2
3
4
5

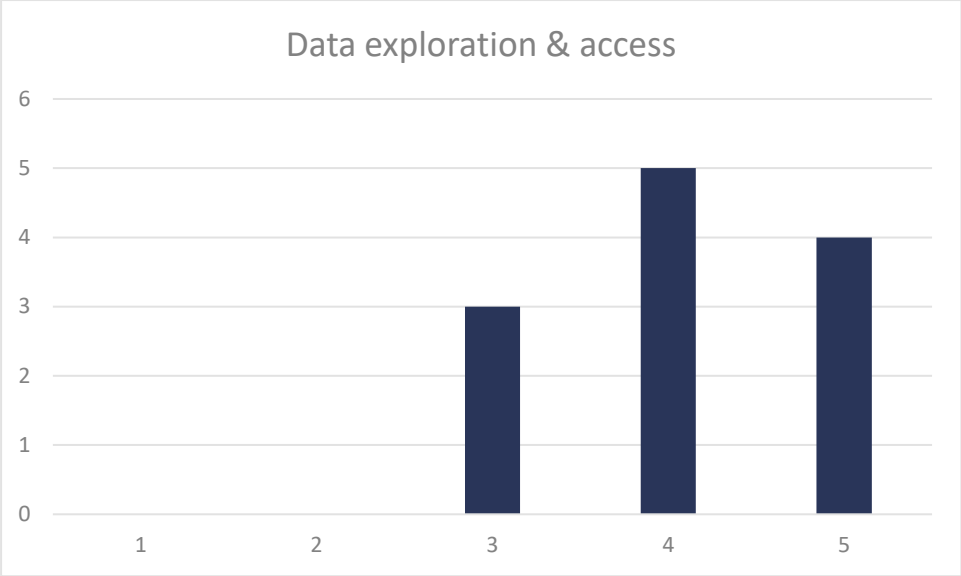
Additional remarks:

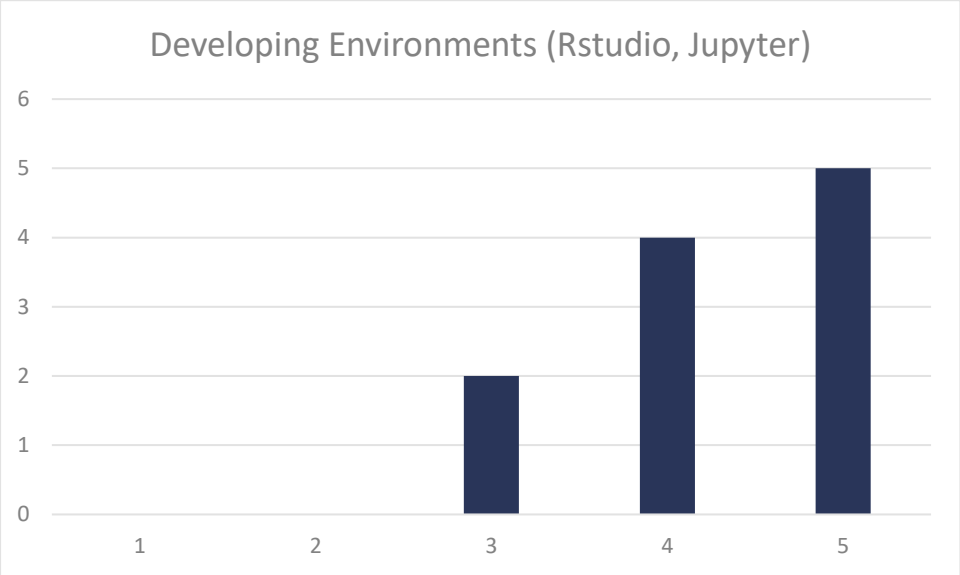
"Ergonomics is far too variable from one tool to another"
"good file management, sometimes I had some problems to rename some files (in Jupyter for example) " about shared workspace
"Enable visualisation of data sets stored in DataMiner without having to download/reupload them"
"I think it is necessary to improve the usability of some features such as the visualization which requires to download the data file from the workspace and to upload it."
"Getting all of this packaged in a unique environment highly improve ease of use. That's to me one of the major aspect (with also unique login)."
"The documentation has to be up to standard, and the first steps made easier, but the functions are interesting" about the statistical algorithm importer
"Some features were difficult to use on a laptop because of the small screen size (SAI)"
"The big blocking point is the reaction time which is really long even for very small data sets."

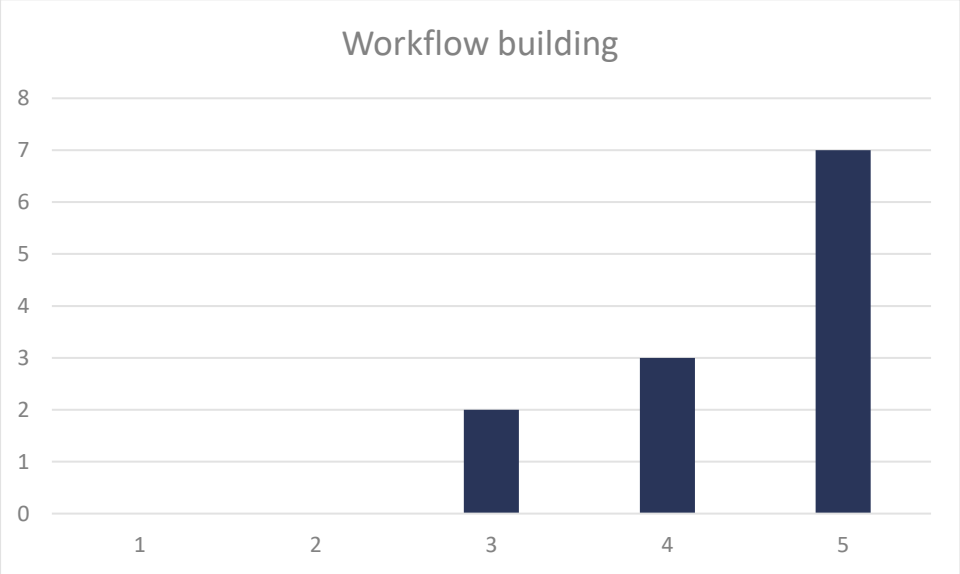
Indicator – Usefulness:

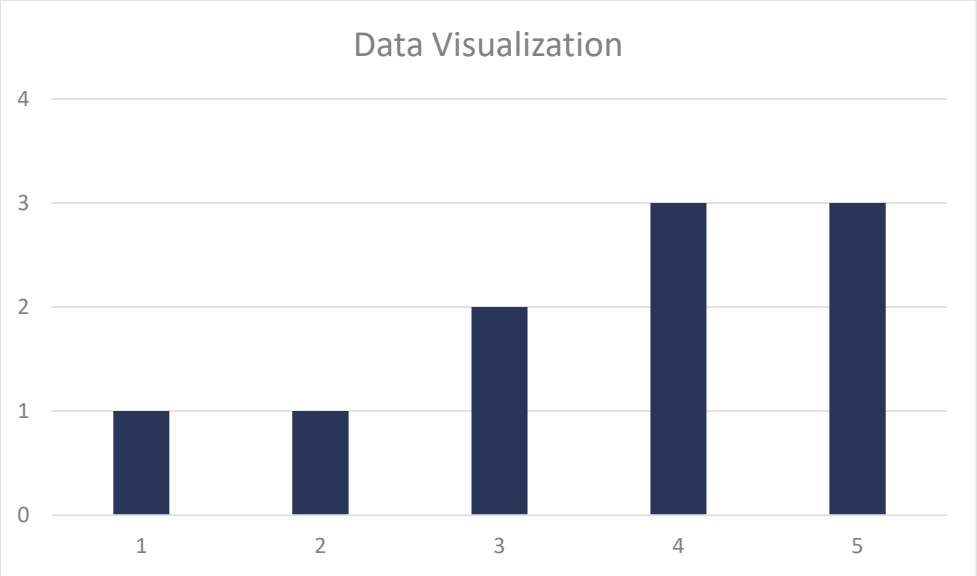
	Evaluation 1	Evaluation 2	Evaluation 3	Combined evaluations
Collaboration & communication	4.66	3.9	4.8	4.50
Data exploration & access	2.33	4.2	3.6	4.08
Developing Environments	3.5	4.4	4	4.27
Workflow building	4.5	4.8	4	4.42
Data Visualization	4	4	3.2	3.6
Total Average Usefulness	**3.80**	**4.26**	**3.92**	**4.07**




Data exploration & access
6
5
4
3
2
1
0
1
2
3
4
5


Developing Environments (Rstudio, Jupyter)
6
5
4
3
2
1
0
1
2
3
4
5


Workflow building
8
7
6
5
4
3
2
1
0
1
2
3
4
5



Additional Remarks:

"The tools selected are judicious and represent a complete panel."
"The visualisation options are not very useful, unless visualization can be better configured and even more important, can be automated e.g. through templates, so not every graph has to be configured individually."
"The visualisation tools are very basic."
"More generally I would say the communication aspects and the ability to visualize vast amounts of data really show how valuable the VRE tool set is as part of the large scale collaborative projects (e.g. EMPHASIS, ELIXIR, etc.)"
"for a statistician or a lambda user, the semantic tools are not very useful"
"Very interesting for those who don't really know the software (it avoids more or less well done installations), the regulars will perhaps take the time and be reluctant." About Rstudio

Indicator – Learning curve:

The learning curve was assessed only during the 3rd evaluation. The average results must be put into perspective because it was calculated from only 5 answers.

	Evaluation 3
Estimated effort to make an existing dataminer algorithm or galaxy workflow run on the VRE	2.4
Estimated effort to configure a new dataminer algorithm or a new galaxy workflow	3
Total Average Learning curve	**2.7**

There was no additional remark regarding this indicator.

Indicator – Openness:

The learning curve was assessed only during the 3rd evaluation. The average results must be put into perspective because it was calculated from only 5 answers.

	Evaluation 3
Finding & accessing dataminer algorithms, galaxy workflows and data registered in the VRE	3.6

Registering and sharing of data, analytic scripts or workflows	3.6
Adding and sharing new functionality and components (e.g. a new shiny application, new charts)	3
Total Average Openness	**3.4**

There was no additional remark regarding this indicator.

Indicator – FAIRness:

The learning curve was assessed only during the 3rd evaluation. The average results must be put into perspective because it was calculated from only 5 answers.

	Evaluation 3
Findability	3.4
Accessibility	3.6
Interoperability	3.4
Reusability	4.2
Total Average	**3.65**

Additional remarks:

"One can find the few items that are published in the VRE, but for accessibility I guess more of it should be public. I see some interoperability in being able to seamlessly access the trial data, which is great. However, I miss out on the link between the available semantic tools and the demonstrated application."
"The good practices concerning FAIR are a second step (following the initial handling) and one should not insist too much on them from the beginning. Good practices should be addressed elsewhere."

Indicator – Uptake potential:

This indicator was assessed by asking participants about their willingness to use the Food Security VRE

	Evaluation 1	Evaluation 2	Evaluation 3
Would you be interested in using a VRE?	2 yes 0 No	4 yes 1 No	5 yes 0 no

Additional remarks:

"Yes, particularly because of the collaboration and social functions. Also for using the individual data science components (e.g. RStudio, Jupyter), developing and publishing/sharing algorithms (DataMiner, Catalogue) and being able to integrate them in workflows (Galaxy). *For visualisation I would most probably use a tool that is more flexible and can be better automated and integrated or even notebook functions. It seems that semantic tools are there, but there is no clue on how they can be used in this VRE. In general it seems that visualisation and semantics components are there as loose components but are not easily integrated."*
"Yes, saves a lot of installation and maintenance troubles leaving more time for research. Also gives fairly easy access to remote compute and storage facilities that are vendor neutral."
"Yes, as a way to improve reproducibility by using the same environnment and to connect data storage and analysis services"
"I have gained a better understanding of how a VRE can positively impact the scientific community that I work with. As such I would have no issues with looking at ways to increase awareness and use of a VRE within the plant phenotyping community."

"Yes, the VRE allows to overcome an initial difficulty which consists in taking in hand the various tools necessary during a project (software, contacts, document sharing, ...)"
"Yes it allows to use some interesting tools (semantic, analysis scripts, extraction, ...) and to save a lot of time in prospecting for appropriate tools. *The data exchange spaces are very interesting and help for the organization as well as for daily interactions."*
"Yes. If really adopted in the context of a project or community, I see a lot of advantages... the problem is getting it the default place for collaborating. As we all go to Google tools much easily and quickly."

3.5.2 Global Observations

The features were generally found to be easy to use. However, some reservations can be noted concerning the implementation of workflow or data visualization. Concerning the implementation of workflow, we can link this result to the "learning curve" indicator which shows that this functionality requires the user to be trained before being able to fully use it. Concerning data visualization, the written comments of the evaluators showed that it is not really a question of too much complexity of the tool but rather a lack of practicality. Indeed, the tool is not fully integrated into the VRE (with direct access to the workspace, for example) and the format used for input data is limited. Moreover, for researchers used to R or Python, they don't really see the advantage of using this tool for basic charts visualization compared to visualisation R libraries for example. More advanced visualizations are expected such as the proposed chart for visualizing data and plant images related to data on the same chart. Unfortunately, configuring this kind of chart in the tool is not very convenient and is not adapted to big datasets.

Overall, the VRE was found to be very useful. In particular, the collaboration and communication features showed the most interest from users. The tools put in place to access data from the phenotyping platforms hosted in the PHIS information system were also greatly appreciated. This also helped to show the community the importance of implementing standard APIs in all phenotypic databases. The developing environments Rstudio and Jupyter are tools that researchers already use so they are easy to use for them. The main advantage they saw of having these tools inside the VRE was to be able to work on the same session from any machine. They can easily stop their work and get back to it somewhere else on another machine. The issue they raised was the fact that they couldn't access to their workspace to open their R script directly inside Rstudio. The semantic tools were of interest only to participants familiar with this technology and they noted a certain lack of integration of some of these tools.

The evaluators admitted having difficulties in assessing the "FAIRness" and "Openness" indicators, but most believe that a VRE can really help them to share scientific data and/or workflows but also to make them FAIRer.

Finally, of the 12 respondents to the various questionnaires, 11 indicated an interest in using a VRE. Moreover, most of the participants who did not respond to any questionnaire showed a high level of interest in using a VRE.

To conclude, impressions and opinions regarding the VRE are generally positive. Members of the plant phenotyping community say that a VRE can be a useful tool for doing open science, and express a broad enthusiasm for using such environments for research.

4 VALIDATION (4TH PILOT EVALUATION) METHODOLOGY

4.1 INTRODUCTION

It was decided with all project partners to carry out an additional evaluation cycle. In the first three evaluations, each "use case" partner chose its own method by targeting the questions to the target community. As a result, the results achieved by the three use case partners could not be harmonized to obtain an overall result. Indeed, although this method allowed for the best possible collection of user feedback, it led to a rather heterogeneous result with a variety of tools and methods used. In order to obtain a more generic result for all the targeted communities, the evaluations were extended with an additional validation This validation uses a more harmonized approach by using the same assessment instrument to collect feedback from different communities.

This validation phase of the Food Security VRE was carried out according to the approach described in the deliverable Harmonized Use Case Validation methods, guidelines and materials (D1.2). For more details on the context of the validation methodology, the indicators that are being evaluated and the specific survey that was developed to collect the required data, we refer to this report.
The demonstration scenario used in this fourth assessment was more or less the same as the one used for the third assessment. Due to the short time between the third evaluation and the end of the project, no new services could be made available for inclusion in the validation).

In order to carry out this additional evaluation, it was decided to do it remotely by means of a webinar for two main reasons. Firstly, the decision to carry out this validation phase was taken late. Second, the webinar allowed people who had not been able to travel to previous events to have access to the information.

4.2 EVALUATION PROGRAMME

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

4.3 EVALUATION INDICATORS AND SURVEY QUESTIONS

As part of this validation phase, a set of 32 evaluation questions was provided to participants in the form of a google form. Table 2 lists these questions. 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: Validation questions

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 skillful 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 behavior 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 VALIDATION RESULTS

5.1 INTRODUCTION

This section provides an overview of the results of the validation step. It provides an overview of the characteristics of the assessment session and presents the results of the survey that was conducted. The results of the survey have been qualitatively analysed.

5.2 EVALUATION OUTCOMES

5.2.1 Evaluation details

	Global evaluation
Date:	December 18th 2019
Method:	Webinar using ZOOM and online survey using Google Form
Participants:	16
Survey responses:	11

5.2.2 Evaluation results

Table 9 shows the summarized responses per validation survey question, including the number of votes and the mean score over all respondents.

Table 9: Validation survey results

Question	1	2	3	4	5	No answer	Number of votes	Mean rating
I would find such a virtual research environment useful in my job.	0	0	1	7	3	0	11	4.18
Using such a virtual research environment would enable me to accomplish tasks more quickly.	0	1	4	5	1	0	11	3.55
Using such a virtual research environment would increase my productivity.	0	1	6	3	1	0	11	3.36
If I used a virtual research environment, I would increase my chances of getting a better position or salary.	2	2	6	1	0	0	11	2.55
My interaction with such a virtual research environment would be clear and understandable.	0	1	2	8	0	0	11	3.64
It would be easy for me to become skillful at using such a virtual research environment.	0	1	2	8	0	0	11	3.64
I would find such a virtual research environment easy to use.	0	1	2	5	3	0	11	3.91
Learning to operate such a virtual research environment would be easy for me.	0	1	4	4	2	0	11	3.64
Using such a virtual research environment is a good idea.	0	0	1	3	7	0	11	4.55
A virtual research environment makes work more interesting.	0	3	3	5	0	0	11	3.18
Working with such a virtual research environment is fun.	1	1	5	2	2	0	11	3.27
I would like working with such a virtual research environment.	1	0	2	4	0	3	7	3.29
People who influence my behavior would think that I should use such a virtual research environment.	0	3	4	2	2	0	11	3.27
People who are important to me would think that I should use such a virtual research environment.	0	3	5	1	2	0	11	3.18

The senior management of my organisation would be supportive of using such a virtual research environment.	0	0	4	4	3	0	11	3.91
In general, my organization would support the use of such a virtual research environment.	0	0	1	5	5	0	11	4.36
I have the resources necessary to adopt and use a virtual research environment.	0	1	2	5	3	0	11	3.91
I have the knowledge necessary to adopt and use such a virtual research environment.	0	0	2	7	2	0	11	4.00
The virtual research environment does not seem compatible with other systems I use.	0	3	3	4	1	0	11	3.27
In my organisation, a specific person (or group) would be available to assist me with difficulties in using such a virtual research environment.	2	2	3	3	0	1	10	2.70
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.	0	3	1	4	3	0	11	3.64
I could complete a job or task using this virtual research environment if I could call someone for help if I got stuck.	0	0	5	4	2	0	11	3.73
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.	0	0	2	8	1	0	11	3.91
I could complete a job or task using this virtual research environment if I had in my organisation a facility for assistance.	0	0	4	6	1	0	11	3.73
I feel apprehensive about using such a virtual research environment.	2	5	1	2	1	0	11	2.55
It scares me to think that I could lose a lot of information using such a virtual research environment by hitting the wrong key.	5	4	0	2	0	0	11	1.91
I hesitate to use such a virtual research environment, fearing to make mistakes I cannot correct.	4	4	0	2	1	0	11	2.27
Such a virtual research environment looks somewhat intimidating to me.	4	5	0	2	0	0	11	2.00
I intend to use such a virtual research environment in the next 12 months.	0	3	6	1	1	0	11	3.00
I predict I would use such a virtual research environment in the next 12 months.	0	2	4	1	1	2	8	3.13
I plan to use such a virtual research environment in the next 12 months.	0	1	5	1	1	2	8	3.25

5.3 SUMMARY AND OBSERVATIONS

This section provides a qualitative summary of the results of the validation survey. Participants in this survey were provided with a brief overview of the VRE's potential for collaborative research. A typical scenario of a researcher's work was presented, beginning with data discovery and access, through data analysis and visualization, to sharing analysis results and processing processes. This gave participants a good first impression of what it means to work with a VRE and the type of applications that can be developed. However, it should be kept in mind that it remains difficult for most participants to really see the technical and organizational consequences of using VRE. This is why we present a qualitative summary of this validation phase.

The qualitative analysis of the results leads to the following observations:

- Respondents appear to be positive about the usefulness of a VRE for their work. They generally support the idea that it would help them be more productive.

- Respondents appear willing to use a VRE for more personal reasons. They indicate that working with a VRE would be interesting and fun.
- Most respondents believe that the VRE is easy to use. In addition, many believe they have the knowledge and resources to use a VRE. However, some respondents show a rather high level of apprehension. It is noteworthy that these same individuals expressed a possible lack of assistance. Their apprehension regarding the use of the VRE would therefore be more related to the absence of someone to assist them.
- The final questions in the survey did not show a strong intention of respondents to use a VRE in the next 12 months. Given the rather positive responses regarding the usefulness of an VRE, it is likely that this result is due more to the possible lack of support in using a VRE but also to the fact that respondents do not seem to be sure of the compatibility of a VRE with the other systems they work with.

6 DISCUSSION AND RECOMMENDATIONS

6.1 DISCUSSION AND ACHIEVEMENTS

The AGINFRA+ project has introduced the concept of Virtual Research Environment to the phenotyping community and the plant science community more generally. Most members of this community were not familiar with virtual research environment and were not aware of how a VRE can help collaborative research. Some members of the community were able to see demonstrations of VREs and their possible applications in the field of high-throughput phenotyping. They were given the opportunity to provide feedback through evaluation surveys.

For the evaluation, the community-specific pilot evaluations were aligned with the stage of implementation of the VRE features. This allowed for a better involvement of community users and for the adaptation of ongoing developments. Therefore, during the 1st evaluation, the available features of the VRE were evaluated independently while during the following evaluation phases, the same features were evaluated following a user scenario corresponding to a typical scenario of this community. Scientists from the community were thus able to participate in practical workshops during which they were able to concretely test the use of the VRE. To address the differences in the approach to assessment between the 3 different research communities of the project, an additional assessment was developed and carried out as a more holistic and harmonised validation exercise.

For the Food Security community, the AGINFRA+ project enabled to introduce virtual research environments to scientists who were generally unaware of the possibilities of virtual research. Even if we had real difficulties to attract a lot of members of the plant phenotyping community to participate to evaluation session, we took opportunities to present our work to the community through European projects dedicated to the plant phenotyping. We were able to demonstrate the general concept of a VRE, and how VREs can support collaborative teams in plant science research. In addition, we demonstrated how a VRE and its components can be used to discover and access data and develop scientific workflows. We observe that community members state that VREs can be a useful tool for open science, and we see an interest to use such environments for research. Many seem to recognize that virtual research environments offer unique opportunities to share heterogeneous data and analysis workflows and thus make them more easily reusable, responding to the demand for collaboration in different fields with heterogeneous data needed to address major interdisciplinary challenges in agri-food and environment.

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

- The opportunity of having a full working environment accessible on-line from any machine, with Rstudio or Jupyter as developing environments.
- The opportunity to explore and have access to different plant phenotyping and breeding databases. This also helped to make them aware of the importance of having interoperable systems to facilitate data exchange.
- The offered opportunities for encapsulating algorithms through DataMiner and designing and publishing scientific workflows with Galaxy.

6.2 RECOMMENDATION

Despite the positive feedback we have received about the VRE, some of the features have been found not very user-friendly or missing. We believe that the following issues, resulting either directly from the feedback or indirectly from the interpretation of the discussions with the developers involved, are relevant for future post-project improvement.

- During the evaluations, it appeared that the use of a VRE is of more interest to the consumers of the data than to the producers. Indeed, the first difficulty in plant phenotyping is accessing the data because they are often very heterogeneous and come from different databases. This was precisely what showed the most interest. That is why it is very important to keep enriching the plant phenotyping information system with BRAPI web services, so that the user would be able to search data with more filters such as the genotype.
- The user experience could be improved by a better integration of individual VRE components and services. For example, accessing and visualizing data should be done in one single step. Downloading a data file from the workspace to upload it to the visualization tool is not convenient. Some users also found the workspace unwieldy.
- Some evaluators expected more integrated semantics to improve the links between ontologies concept and the catalogue items metadata.
- Some evaluators expected features that really enable to collaborate on the development on an algorithm or a paper by working together on the same time on the same file.

ANNEX 1 – EVALUATION 1 USER SCENARIO

I. Objectives

A VRE (Virtual Research Environment) is a collaborative working environment for a research community. The aim of this exercise is to assess the effectiveness of using a VRE for research in high-throughput phenotyping. This evaluation will be made in 3 phases. This is the first one.

The first evaluation of the VRE is based on these 5 indicators:

- Ease of use: Guidance of the environment. How simple is the concept of a VRE to the user; how easy are D4Science or EGI to use? How much effort is needed to define or use workflows? How much effort does it take to develop workflow components? Data preparation: does the VRE have a significant array of coding or noncoding features, such as for data transformation and filtering, to prepare data for modelling? Data exploration and visualization: does the VRE allow for a range of exploratory steps, including interactive visualization?
- Learning curve: How much time is needed to learn new concepts etc. before the VRE can be used?
- Usefulness: How does using the VRE compare to current hardware and software in use? E.g. considering costs, functionality? User interface: does the VRE have a coherent "look and feel," and does it provide an intuitive UI, ideally with support for a visual component for the design of workflows?
- Openness: How easy it is to add new data and functionality to the VRE? How easy is it to share workflows, components, and data?
- 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? How well does the VRE support access and integrate data from various data sources and of different types (textual, sql, RDF, images, location data, etc.)?

II. User scenario

Connect to https://aginfra.d4science.org/group/foodsecurity with your login and password (or sign in if it's not already the case). Follow the instructions below.

1. Communication between members

a. Publish a message for others members in « share updates ». You should see this message in the « news feed » section.

b. Go to the messaging system  (in the top left corner) and send a mail to one member of the VRE (Alice Boizet).

2. Shared Workspace

a. Go to the workspace  (in the top left corner) and import a file inside your personal workspace and upload a file inside the VRE workspace.

b. Download a file from the VRE workspace.

3. Data Access

a. Go to Discovery Tool / ReFindit: test the research tool by entering free text or DOI (the data sources are written below the input field. INRA dataverse should be soon part of the data sources)

b. Dataminer algorithm to get data from phenomeAPI: Go to Analytics / Dataminer. Execute the algorithm Getplantheight Fromphenoarch (in Data extraction category). You should get a csv file with plant heights data (this data come from the ZA17 experiment of the phenoArch platform in Montpellier)

4. Data visualization

Go to Data visualization/Create graphs. Click on Create. Import a csv file with data to visualize. Fill the required field and save.
Go to Data visualization/ View graphs and select your chart to display it.

5. Data Analysis

a. Developing analytics scripts (R or python):
Go to Rstudio or Jupyter and write or copy and paste a script inside the tool and run the script. If you script requires a library which is not in this list:
https://wiki.gcube-system.org/gcube/Pre_Installed_Packages , you can contact a manager of the VRE to install the library.

b. Share a script as a black box with the Statistical Algorithms Importer (SAI):
Import your script inside Dataminer: Go to Analytics/Import a new method. You can click on « help » button to go to the user documentation of this tool.

ANNEX 2 – FILLED EVALUATION 1 SHEETS

Tester n°1. This tester didn't attend the VRE workshop.

Feature type	Features	Ease of use	Learning curve	Usefulness	Openness	FAIR-ness
Collaboration features	Messages publication (1.a)	5	5	5	4	3
	Messaging system (1.b)	5	5	5	1	1
	Shared Workspace (2)	5	5	5	4	4
Data access	Discovery tool (3.a)	5	5	5	5	2
	Data access with Dataminer algorithm (3.b)	3	3	2	4	4
Data analysis	Developing environments (4.a)	2	2	4	3	3
	SAI (4.b)	2	2	4	4	3

N.B. The visualization tool was not available when this user tested the VRE

Open questions:

- *Which features do you find the most useful or even necessary?*

Data Miner looks a good tool (provided that I learn how to use it) allowing to store, analyse and share data and analysis pipelines.

- *Which features miss? What important features should be deployed in the VRE?*

I don't think there is something lacking for the moment.

- *Do you think a VRE could be a good environment for your daily work?*

Not for my daily work, but seems a nice tool for collaborative projects. This may avoid using the number of existing tools (google drive and docs, mails, repositories, etc…) and centralise in a unique and versatile tool.

- *Do you think phenotyping researchers are ready to use a VRE?*

Yes, they are smart people.

- *Other remarks:*

The phenoarch algorithm didn't work because the library was not updated.

When I am in Rstudio, I don’t have access to my VRE folders.

Tester n°2: this tester attended the VRE workshop.

Feature type	Features	Ease of use	Learning curve	Usefulness	Openness	FAIR-ness
Collaboration features	Messages publication (1.a)	5	5	4	4	2
	Messaging system (1.b)	5	5	4	3	1
	Shared Workspace (2)	5	5	5	4	2
Data access	Discovery tool (3.a)	4	5	3	1	2
	Data access with Dataminer algorithm (3.b)	4	5	5	5	4
Data visualization	Create and view a chart (4)	5	5	4	4	3
Data analysis	Developing environments (5.a)	4	5	3	4	2
	SAI (5.b)	2	3	5	4	3

Open questions:

- Which features do you find the most useful or even necessary?
Rstudio and the collaboration features

- Which features miss? What important features should be deployed in the VRE?
Rshiny, more resources in data discovery

- Do you think a VRE could be a good environment for your daily work?
Yes, as an additional working environment, but not a main environment.

- Do you think phenotyping researchers are ready to use a VRE?
Yes, some of them to some extent. Maybe mostly to:

- Discuss with other researchers
- Share scripts and visualize their results directly (better than email)
- Conduct surveys

ANNEX 3 – EVALUATION 2 EXERCISE

Programme de l'atelier

1. Présentation du VRE et du projet AgINFRA+
2. Présentation des fonctionnalités
3. Test et Evaluation des fonctionnalités
4. Retours
5. Focus sur Galaxy

Présentation du VRE Food Security

Qu'est ce qu'un VRE ?

Un VRE (Virtual Research Environment) est un environnement de travail virtuel et collaboratif à destination des chercheurs. "Il est constitué par un ensemble d'outils et de ressources en ligne rendues interopérables dans l'objectif de faciliter les processus de recherche au sein et entre les institutions. La caractéristique essentielle d'un VRE est de faciliter la collaboration entre les scientifiques." source IFB

Le fait qu'il soit virtualisé permet tout d'abord au chercheur de travailler sur son environnement depuis n'importe quel machine : ses données sont stockées dans le VRE et des outils (tels que Rstudio) permettant d'analyser les données sont accessibles directement dans le VRE.

Un espace de travail partagé permet aux chercheurs de partager leurs données ou processus de traitement et ainsi de collaborer plus facilement.

Projet AgINFRA+

AGINFRA + vise à exploiter les e-infrastructures telles que l'EGI et D4Science, pour fournir des VREs adaptés à certaines communautés de chercheurs autour de l' agriculture et de l'alimentation.

À cette fin, le projet a mis en place plusieurs VRE pour 3 communautés :

- la communauté de la modélisation agro-climatique
- la communauté de la sûreté alimentaire
- la communauté de la sécurité alimentaire.

Le but du projet est d'alimenter ces VREs avec différents composants répondant aux besoins de chaque communauté. Ainsi, des fonctionnalités permettant le développement rapide et intuitif de workflows variés d'analyse de données ont été intégrées ainsi que des outils de visualisation de données.

Food Security VRE

Le Food Security VRE est porté par la plateforme D4Science, développée par le CNR.
https://aginfra.d4science.org/group/foodsecurity Pour la communauté "Food Security", le cas d'utilisation choisi est le phénotypage de plante haut-débit qui produit de gros volumes de données variées.

Evaluation des fonctionnalités

Objectifs et déroulement de l'évaluation

Le but de l'exercice est d'évaluer si un VRE et plus particulièrement le VRE "Food Security" est un outil qui peut faciliter les travaux de recherche en phénotypage.

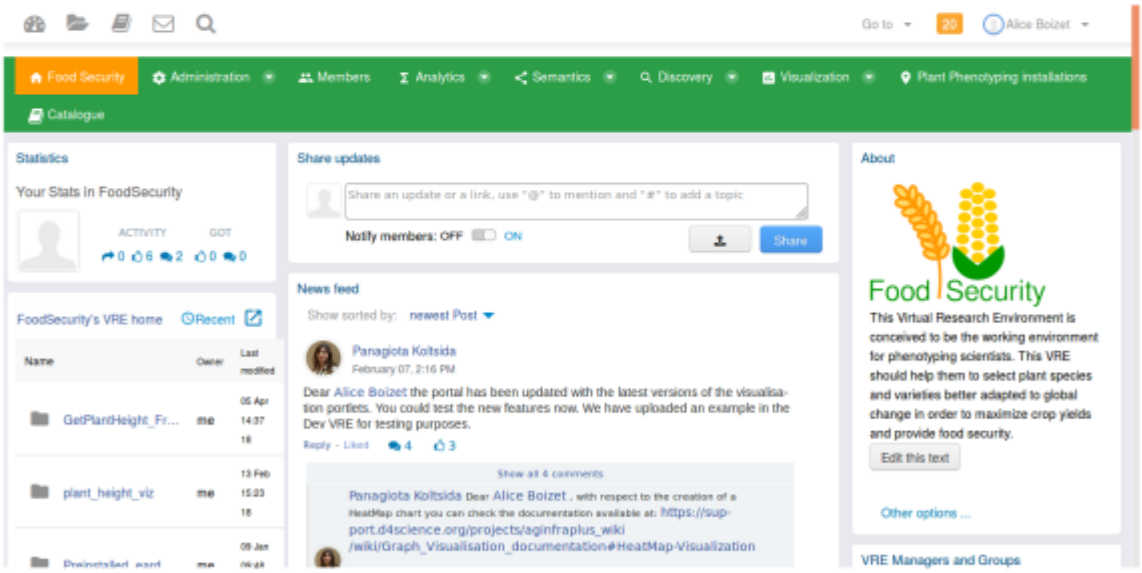
Après la présentation des différentes fonctionnalités du VRE, vous allez être amenés à suivre un scénario utilisateur permettant de tester chacune de ces fonctionnalités. Ensuite vous pourrez répondre à un questionnaire dans lequel il faudra noter chaque fonctionnalité suivant différents indicateurs.

Scénario utilisateur

Inscription/connexion au VRE

Rendez-vous sur https://aginfra.d4science.org/group/foodsecurity. Si vous êtes déjà inscrit, connectez-vous directement et passez directement à l'étape suivante. Sinon, inscrivez-vous au VRE (par la méthode de votre choix)

Accéder à des données



L'espace de travail

- Le workspace Cliquez sur  . Dans le répertoire VRE Folders, vous accédez à l'espace commun de chaque VRE auquel vous êtes inscrit. Les fichiers que vous déposerez dans VRE Folders/Food Security VRE seront donc accessibles aux autres membres du VRE. Les répertoires et fichiers créés à la racine du workspace sont privés.

- La messagerie Cliquez sur  . Vous avez la possibilité d'utiliser cette messagerie pour envoyer des messages privés aux membres du VRE.

- Le catalogue Cliquez sur  . Le catalogue recense les ressources disponibles sur le VRE. Son utilisation sera expliquée dans une autre partie

Refindit

Allez dans Discovery/Refindit. Cherchez une publication Vous pouvez par exemple rechercher la publication sur PHIS dont voici la référence {P. Neveu, A. Tireau, N. Hilgert, V. Nègre, J. Mineau-Cesari, N. Brichet, R. Chapuis, I. Sanchez, C. Pommier, B. Charnomordic, F. Tardieu, L. Cabrera-Bosquet, Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System, New Phytol., 221 (2019) 588-601. doi: 10.1111/nph.15385}

Récupérer des données depuis les Web Services BrAPI

BRAPI (pour Plant Breeding API) est une spécification d'un standard d'API pour les données en phénotypage et génotypage. Toutes les API respectant ce standard pourront donc être appelées de la même manière, ce qui facilite l'accès et l'échange des données.

1. Allez dans l'application Brapi Data Exploration
   Non Disponible pour le moment

2. Récupérer des données de PHIS grâce aux services BrAPI

Allez dans *Analytics/Data Miner*. Cliquez sur *Execute an experiment*. Dans le menu sur la gauche, cliquez sur *Data Extraction* et *Brapi_Get_Studies_Observations*

Remplissez les différents champs avec les paramètres suivants et cliquez sur *Start Computation*

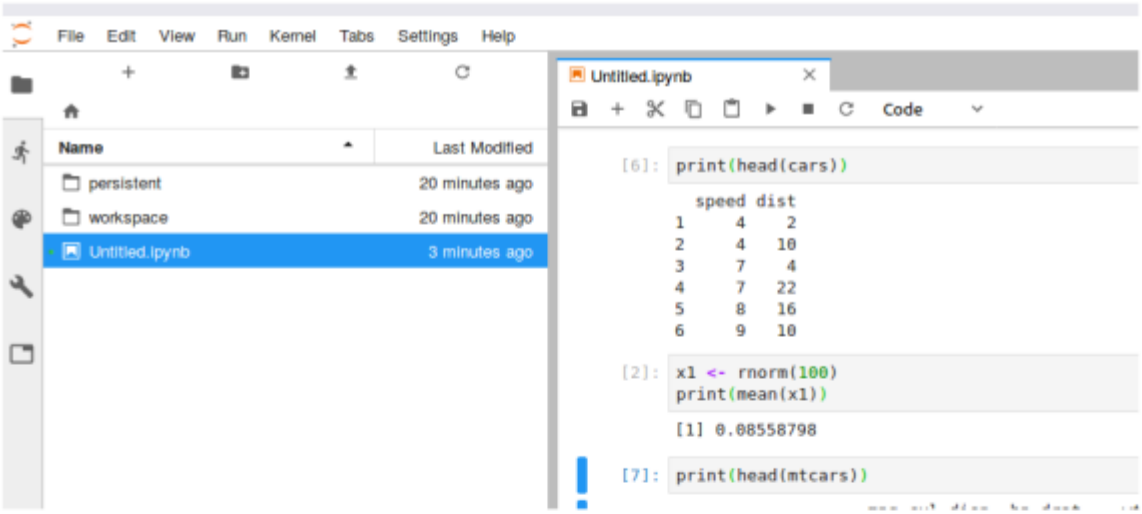
- DBServerURL : http://138.102.159.37:8080/openSilexTestAPI/rest
- studies : http://www.phenome-fppn.fr/test/DMO2019-6
- variables : all
- login : admin@opensilex.org
- password : 21232f297a57a5a743894a0e4a801fc3

Les environnements de développement

- Rstudio Allez dans *Analytics/Rstudio* Dans le cadre en bas à droite, parcourez les répertoires pour retrouver le fichier de données créé par Data Miner. Le répertoire *Workspace* correspond au workspace du VRE. Allez dans *Workspace/DataMiner/output_dataset

Ecrivez un script R court qui lit les données de ce fichier et qui crée un fichier en sortie (*ex : un fichier texte avec la moyenne des mesures*)

- JupyterLab Allez dans *Analytics/Jupyter* Vous accéder alors à l'outil Jupyter qui permet de créer des notebooks dans différents langages.



Créez un notebook R ou python et exécuter quelques lignes de codes simples. (Voir la documentation) Le notebook créé se trouve à côté de 2 répertoires *persistent* et *workspace*. Vous pouvez déplacer votre notebook dans l'un des 2 répertoires.

Le *workspace* correspond au workspace du VRE. Les fichiers stockés dans *persistent* ne seront pas accessible depuis le workspace du VRE. Si vous laissez votre notebook en-dehors de l'un de ces 2 répertoires, celui-ci ne sera pas sauvegardé.

Rendre ses scripts exécutables par n'importe qui (SAI)

Vous allez importer un script R dans le Data Miner afin de pouvoir l'exécuter comme une boîte noire. Ci-dessous, vous avez pour exemple un script qui prend comme paramètres un fichier csv et le nom de la colonne sur laquelle calculer une moyenne et qui renvoie un fichier avec la moyenne

```
inputFile = "test_data.csv"
inputColumnName = "value"
outputFile = "monFichierMoyenne.txt"

myData = read.csv(inputFile, sep=",")
moyenne = mean(myData[[inputColumnName]])
write.table(moyenne, outputFile, append = FALSE, sep = " ", dec = ".",
            row.names = FALSE, col.names = FALSE)
```

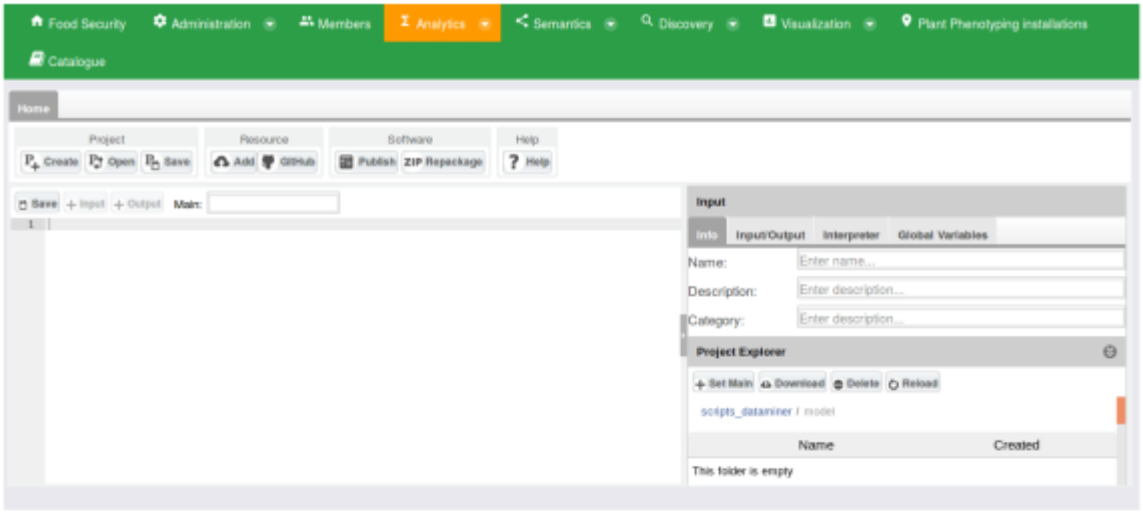
1. Création du répertoire de travail

Allez dans votre workspace et créer un répertoire dans lequel vous déposerez votre script.

2. Création d'un nouveau projet data miner

Allez dans *Analytics/Import a new method* Cliquez sur *Create a project* Sélectionnez le type de projet : R et validez Sélectionnez le répertoire précédemment créé et cliquez sur *Finish*

Vous arriverez alors sur cet écran :

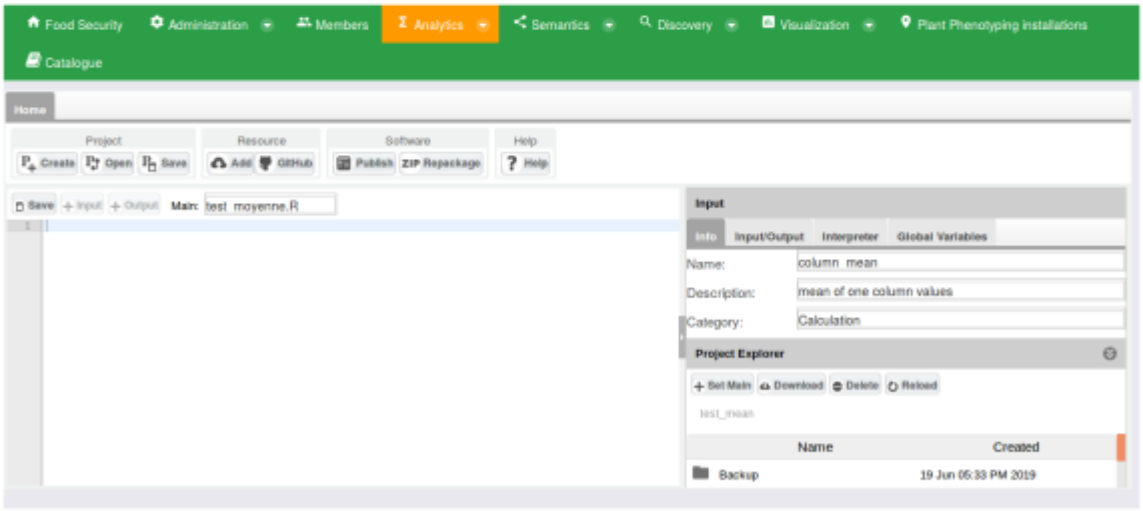


3. Import du script

Dans le panneau project explorer, vous devez voir votre script R. (Lorsque vous créez un nouveau projet, vous pouvez le créer sur un répertoire vide et ajouter ajouter le script directement depuis votre machine avec un cliquer-déposer dans le *project explorer*)

4. Infos de l'algorithme

Remplissez les champs Name, Description et Category du panneau Input/Info (Le champ category correspond à la rubrique dans laquelle se trouvera l'algorithme créé) Attention : Le nom de l'algorithme sert d'identifiant, il doit donc être unique.

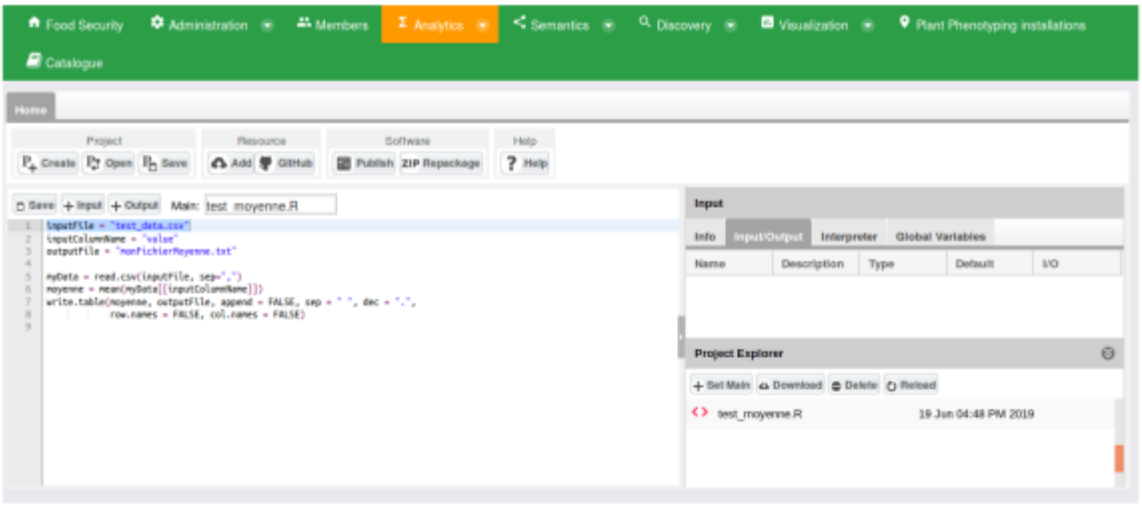


Si vous cocher le champ *Private*, l'algorithme ne sera pas visible par les autres membres du VRE.

5. Définir le script principal

Dans le projet, vous pouvez également importer d'autres fichiers (fichier de fonctions R, fichier de données) Vous devez donc définir quel script doit être lancé lors de l'exécution de l'algorithme. Pour cela, sélectionnez

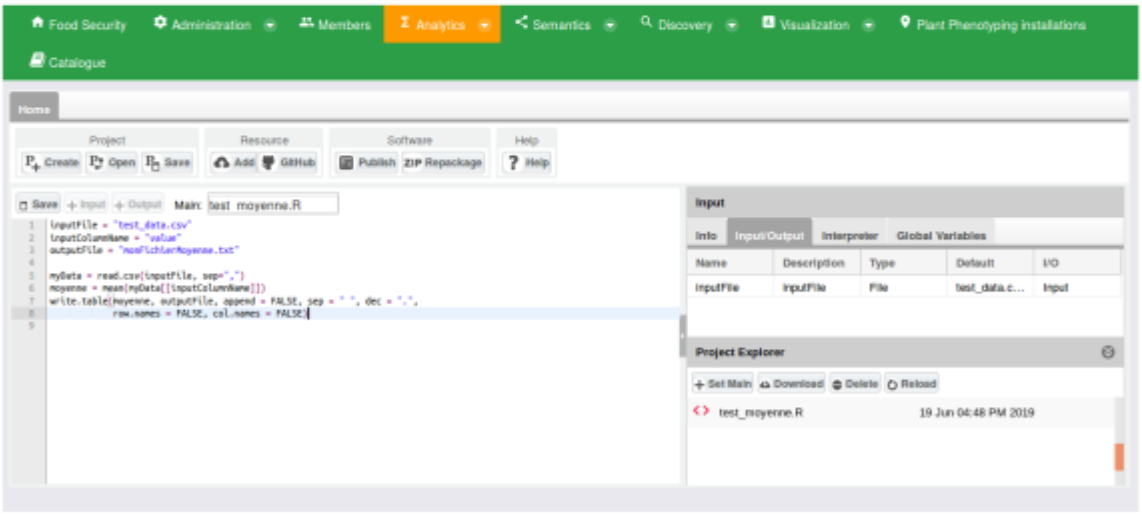
votre script et cliquez sur "Set Main". Le contenu du script s'affiche alors dans le panneau de gauche.



Vous pouvez modifier votre script directement dans cette fenêtre et le sauvegarder avant de publier l'algorithme.

6. Définir les inputs et outputs

Vous devez définir les inputs et outputs du script. Sélectionner la ligne contenant un input puis cliquer sur *+Input*. Un input doit alors apparaître dans le panneau Input (les informations sont entrées de manière automatique).



Faîtes de même pour les autres inputs et outputs.

7. Définir l'interpréteur

Dans le panneau Input/interpreter, précisez "R" dans la version.

8. Publier l'algorithme

Cliquer sur le bouton *Publish* et cliquer *Yes* au message de confirmation A chaque fois que vous modifier une information relative à l'algorithme (description, input/output...), vous devez republier l'algorithme. Dans le cas

où vous ne modifiez que le script, vous pouvez alors juste cliquer sur *ZIP repackage*.

9. Lancer l'algorithme

Déconnectez-vous et reconnectez-vous au VRE. Allez dans *Analytics/Data Miner* > *Execute an Experiment* puis exécutez l'algorithme créé.

Réaliser des workflows de traitement (Galaxy)

Allez dans *Analytics/Galaxy*. Vous arrivez alors sur la page de Galaxy. Dans la barre du haut, cliquez sur *workflow*, cliquez sur le + pour créer un nouveau workflow : renseignez un nom et cliquez sur *Save*.

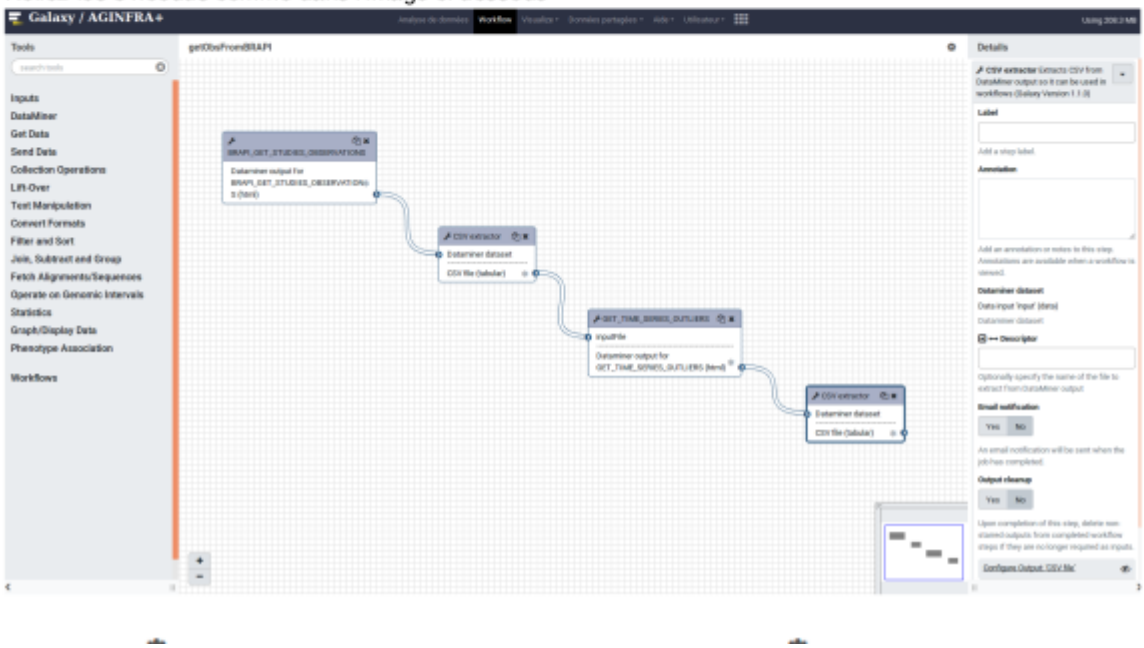
Vous allez créer un workflow qui contient 3 étapes : donc 3 tools La 1ère étape récupère les données d'observations de PHIS grâce à l'algorithme data miner utilisé auparavant. Tous les algorithmes du dataminer sont convertis automatiquement en *"tool"* Galaxy : ceux-ci se trouvent dans la rubrique DataMiner. Cliquez sur le *"tool"* BRAPI_GET_STUDIES_OBSERVATIONS. Dans le panneau *Details* sur la droite, renseignez les différents paramètres :

- DBServerURL : http://138.102.159.37:8080/openSilexTestAPI/rest
- studies : http://www.phenome-fppn.fr/test/DMO2019-6
- variables : all
- login : admin@opensilex.org
- password : 21232f297a57a5a743894a0e4a801fc3

La 2ème étape consiste à convertir la sortie du dataminer en fichier CSV Cliquez sur le *tool CSV extractor* dans la rubrique Data Miner.

La 3ème étape va utiliser l'algorithme DataMiner que vous avez créé. Cliquez sur le tool *test_data*.

Reliez les 3 noeuds comme dans l'image ci-dessous

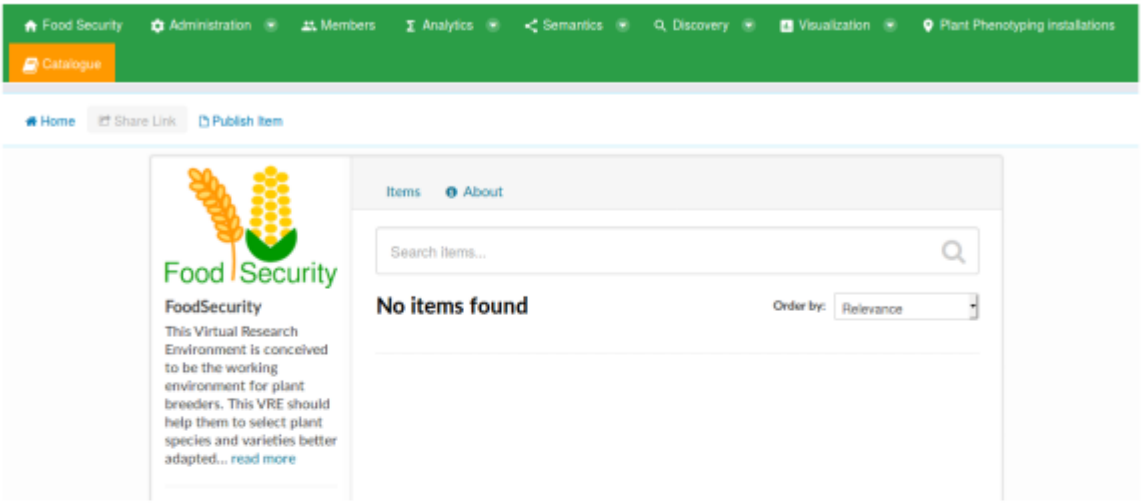


Publier ses données dans le catalogue

Publier des données dans le catalogue

Importez un fichier dans votre workspace.

Cliquez sur *Catalogue*



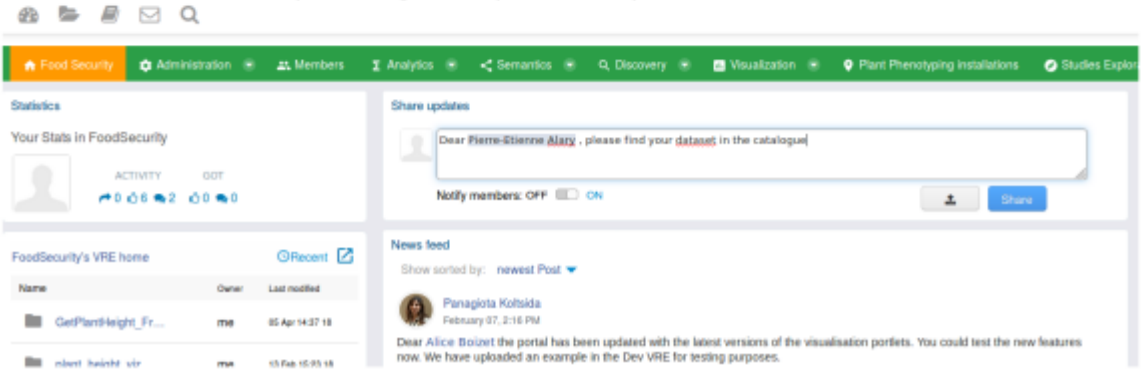
Cliquez sur *Publish Item* Remplissez les informations obligatoires. Dans *Tags*, entrez un terme puis validez avec la touche entrée de votre clavier. Vous pouvez en ajouter plusieurs. En fonction du *type* sélectionné, les informations demandées à la suite diffèrent d'un type à l'autre. Par exemple, pour le type maizeExperiment et wheatExperiment, il vous sera demandé de préciser la ou les variables mesurées. La liste de ces variables provient de la crop ontology (*maize ontology* pour le maïs et *wheat ontology* pour le blé)

Rechercher des données via le semantic search

Non disponible pour le moment

L'envoi de messages

Informez les membres du VRE les membres du VRE de l'ajout d'un nouveau dataset en ajoutant un message dans le fil d'actualités. Vous pouvez taguer une personne en particulier en utilisant @



Visualisation de données

L'outil de visualisation permet de créer des graphiques facilement à partir d'un fichier CSV.

Créer un graphique

Allez dans *Visualization >> Create Graphs* Dans Data, vous pouvez importez un fichier csv avec un cliquer-déposer. Dans les autres onglets, vous préciser le type de graphiques et les colonnes correspondant aux données à visualiser. Pour plus d'informations, veuillez consulter la documentation de l'outil : documentation visualisation

Visualiser des graphiques

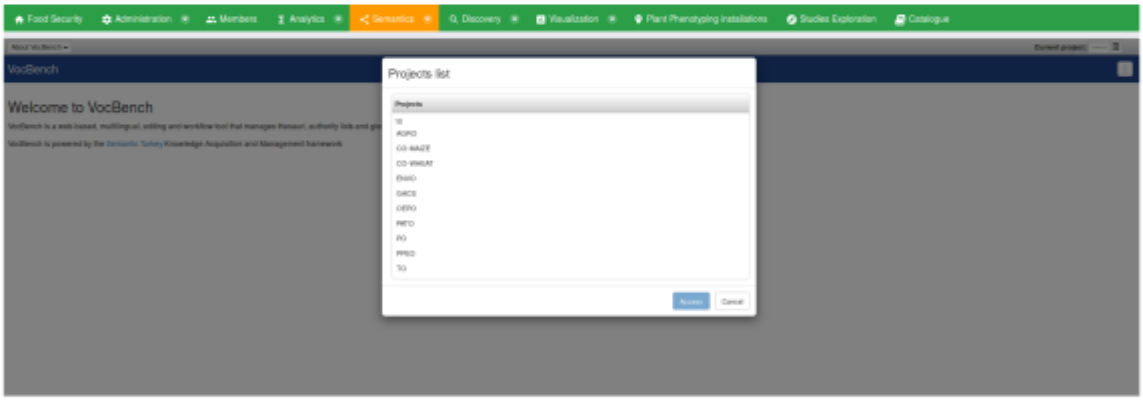
Allez dans *Visualization >> Create Graphs* cliquez sur la liste *Select chart* Sélectionnez le graphique *Plant height with images* Sélectionner un génotype en bas du graphique puis cliquez sur un point du graphique. L'image correspondant à la donnée mesurée appraît alors.

Outils sémantiques

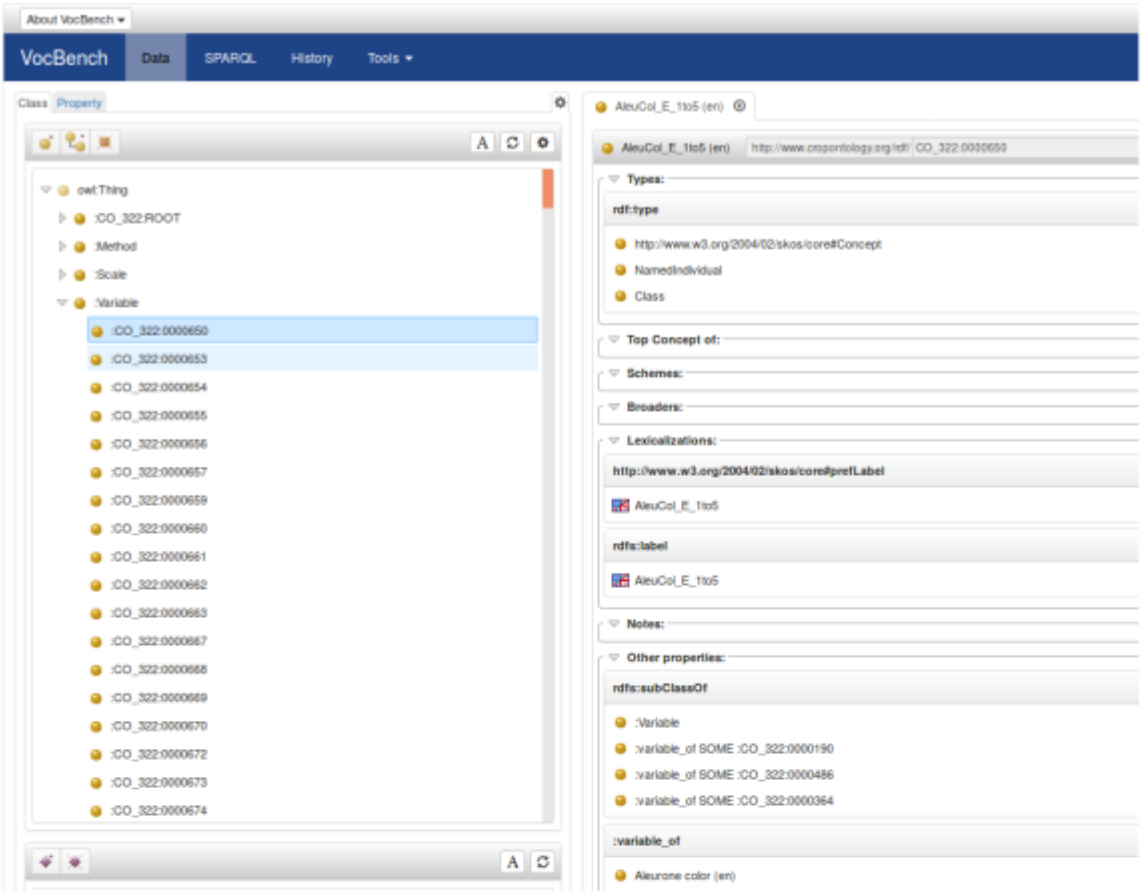
Plusieurs outils de sémantiques sont disponibles

Vocbench

Vocbench est un logiciel permettant de créer des ontologies. Les ontologies GACS, maize ontology, wheat ontology, trait ontology, etc... ont été importées dans Vocbench. Pour accéder à Vocbench, allez dans *Semantics >> Vocbench*. Cliquez en haut à droite sur current project pour sélectionner une ontologie.

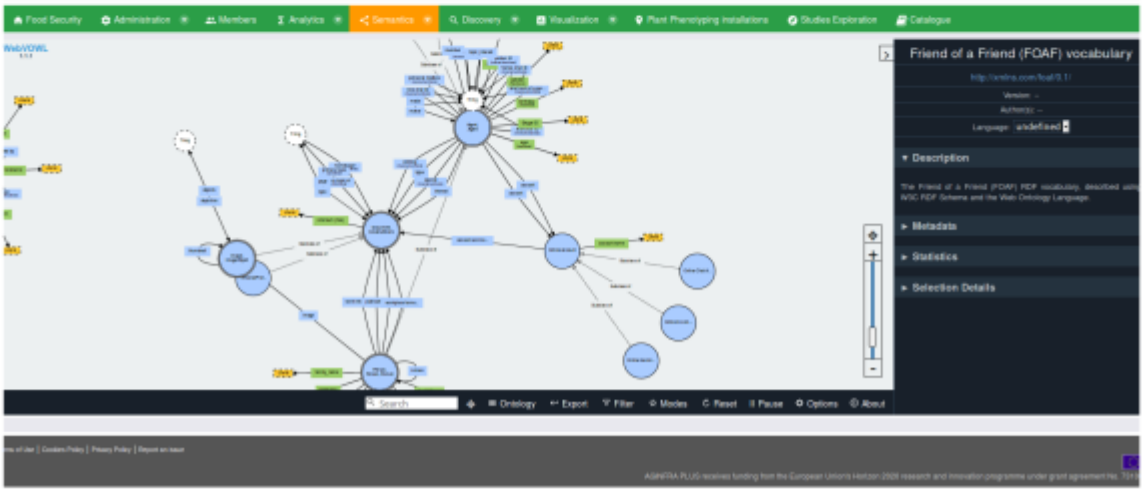


Cliquez sur data pour parcourir l'ontologie



WebVowl

WebVowl est un outil intéressant pour visualiser facilement une ontologie. C'est utile pour des ontologies relativement petites. Allez dans *Semantics >> Ontology visualization*.



Cliquez en bas sur *Ontology* pour sélectionner une ontologie ou en importer une nouvelle

yam++

Allez dans *Semantics >> Ontology matching*. Si vous le souhaitez, vous pouvez comparer 2 ontologies

openrefine

OpenRefine permet notamment de convertir un fichier csv en rdf Allez dans *Semantics >> OpenRefine*

Silk worbench

Silk est un outil permettant de faire du data linking Allez dans *Semantics >> Data Linking*.

Remplir le questionnaire

Vous devez avoir reçu par mail une invitation pour remplir un questionnaire. Cliquez sur *Participate Now* pour ouvrir la page.

Pour chaque fonctionnalité, il vous est demandé de donner une note entre 1 et 5 sur plusieurs critères:

- **Facilité d'utilisation** : 1 si très difficile, 5 si très facile.
- **Utilité de la fonctionnalité** : 1 si inutile, 5 si très utile
- **Intérêt de l'outil dans le VRE** : 1 si l'outil présente peu d'intérêt à être intégré dans un VRE, 5 si le fait que l'outil soit dans le VRE apporte de la plus-value à l'outil.
- **Aide à rendre les données/scripts FAIR** (Findable Accessible Interoperable Reusable) : 5 si la fonctionnalité permet vraiment de les rendre FAIR Il vous est également demandé de donner quelques informations libres sur chaque fonctionnalité.

Une fois le questionnaire rempli, cliquez sur *Send answers*

ANNEX 4 – EVALUATION 3 EXERCISE

AGINFRA+ data science workshop

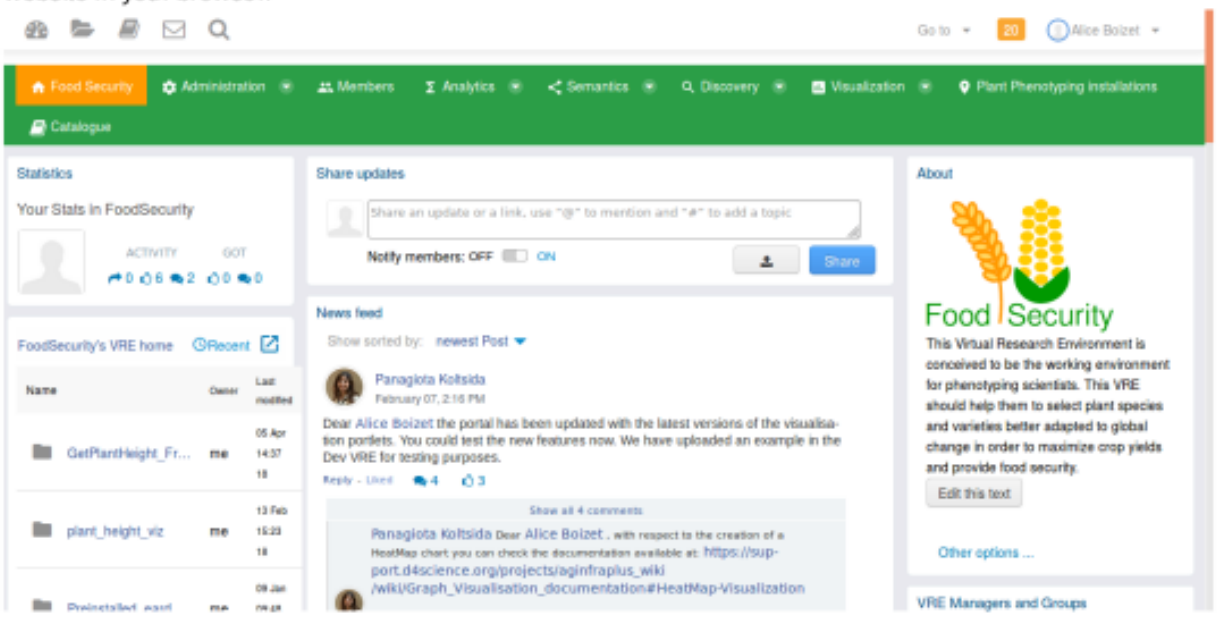
PLANT PHENOTYPING EXPERIMENTS ANALYSIS

1. Accessing the Food Security VRE
2. Exploring phenotyping studies
3. Retrieve data
4. Visualize data
5. Analyze data

1. ACCESSING THE FOOD SECURITY VRE

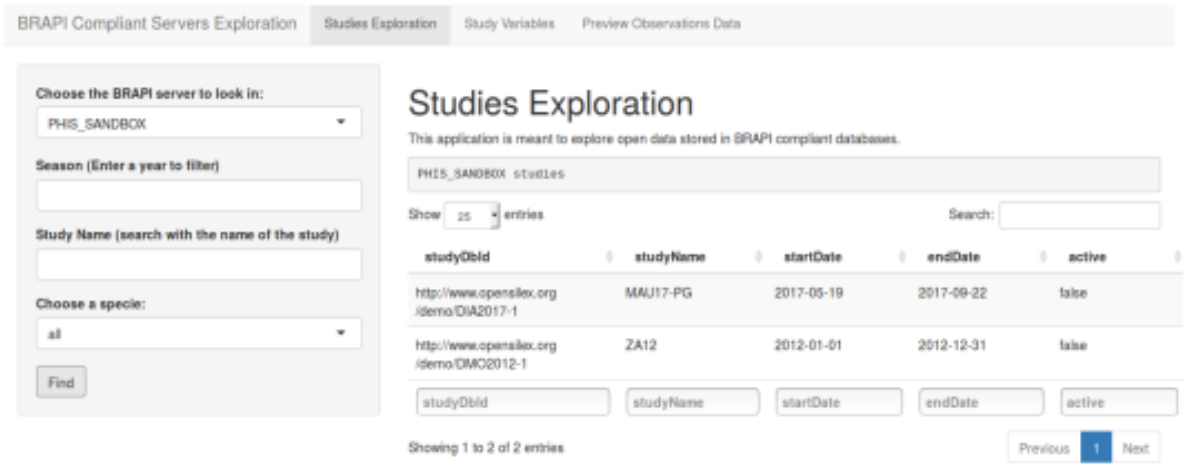
To access the VRE, open a web browser and go to the website : https://aginfra.d4science.org/ On this AgInfra+ Gateway site, please log in with your account details.

You should see the VREs that are accessible to you. Select the Food Security VRE. This will open the website in your browser.



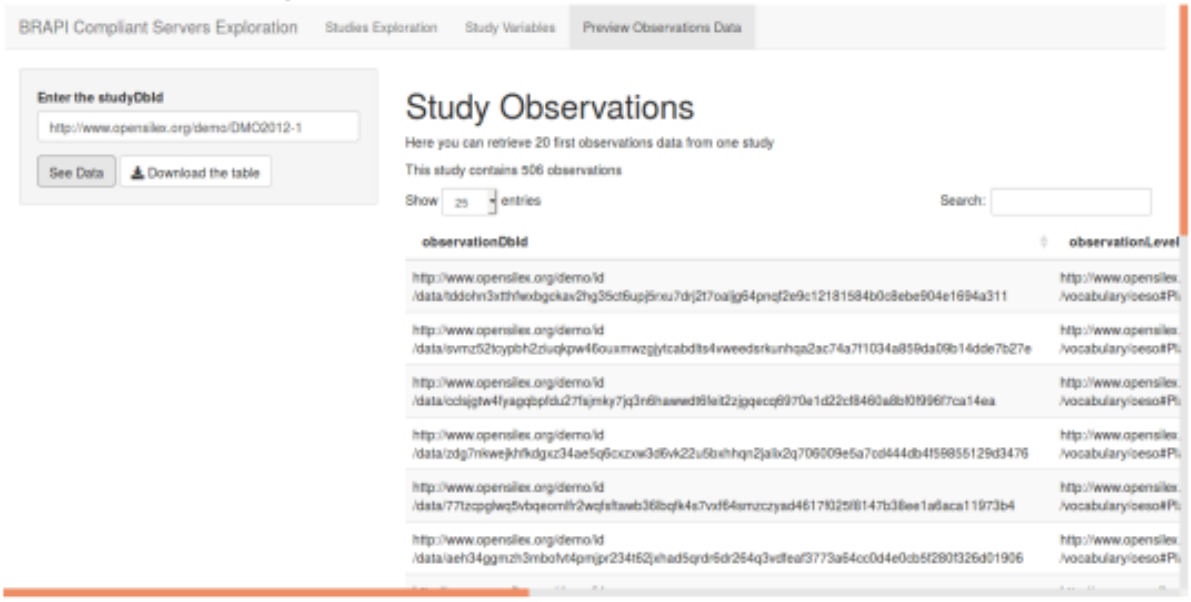
2. EXPLORING PHENOTYPING STUDIES AND RETRIEVING DATA

Open the Studies Exploration. Select the server PHIS_SANDBOX.



You should see at least 2 observations. In this exercise we will work with the study ZA12 data. We are interesting in plant height data. Let's check if there are plant height data on this study.

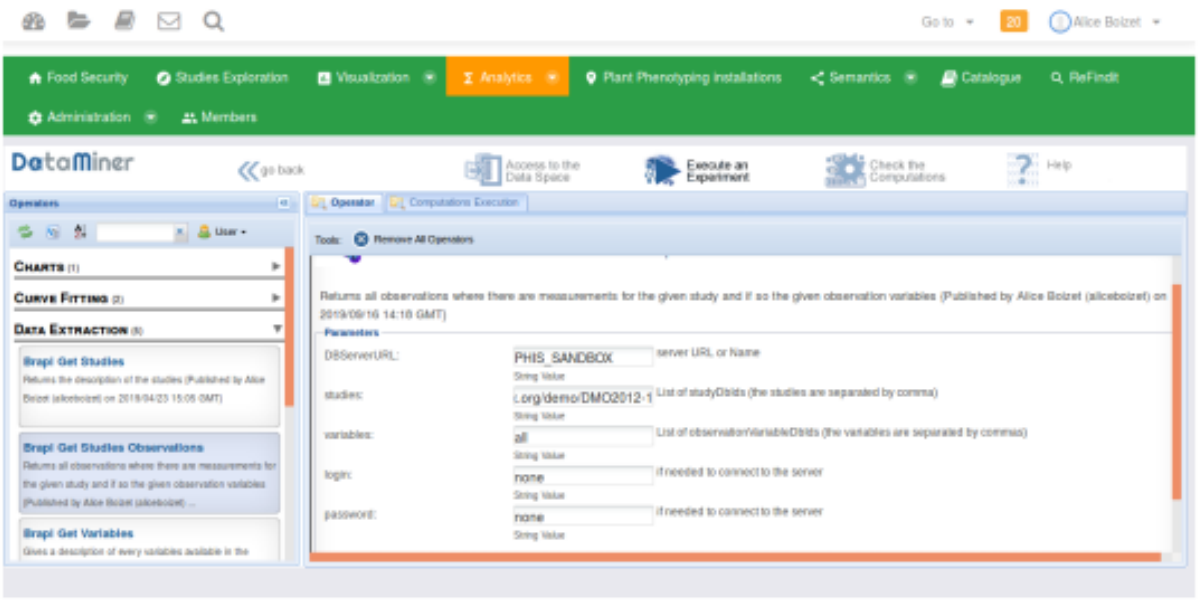
Copy the ExperimentURI and go to the tab "Preview study observations". You can see the first observations measurements of the study.



3. RETRIEVE DATA

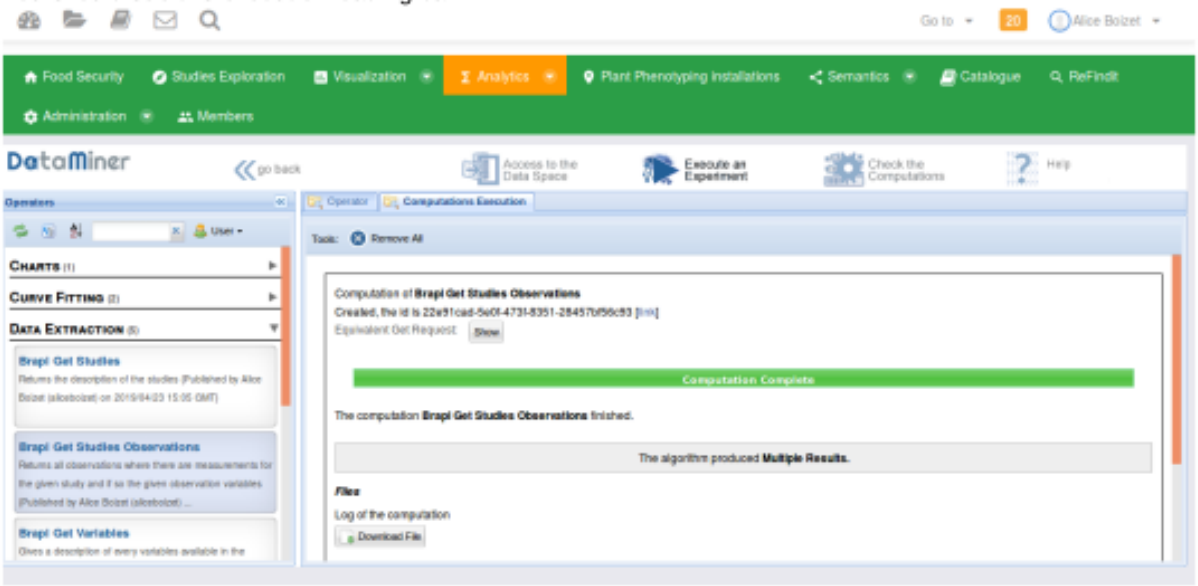
You will use a dataminer algorithm to get all data of the ZA12 study. Go to *Analytics/Dataminer* to execute the algorithm **Brapi Get Studies Observations**. (this algorithm is inside the category "Data extraction")

Fill the different parameters as presented below (the login and password parameters can be used to access to private studies):



Then *Start Computation*

You should see this execution loading bar :



This algorithm creates a csv file and a json file with the study data. They are stored in your workspace under the dataminer folder.

4. VISUALIZE DATA

There are 2 ways to visualize the data. You can do it by using the visualization tools or you can use Rstudio to generate graphs.

Visualize data in the visualization tool

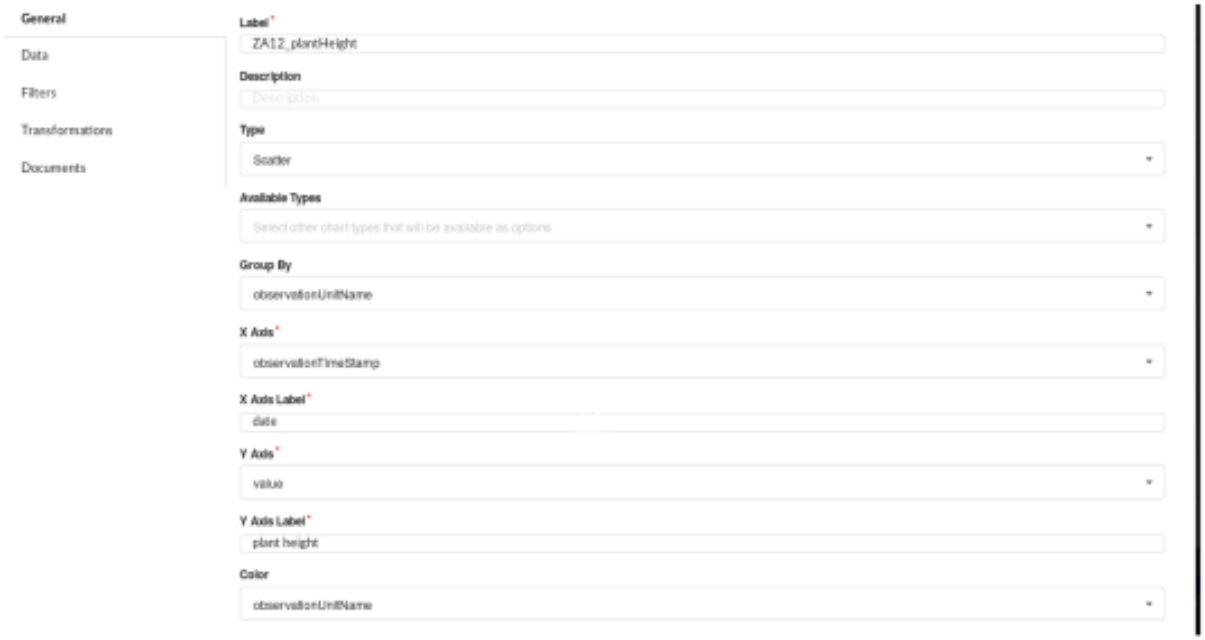
The workspace is not integrated with this tool yet. So you will first need to download the csv file created by the dataminer. If you didn't do it at the end of the dataminer computation, you can go to your workspace:

button  at the upper left corner. The files are stored in DataMiner/Output Data Sets.

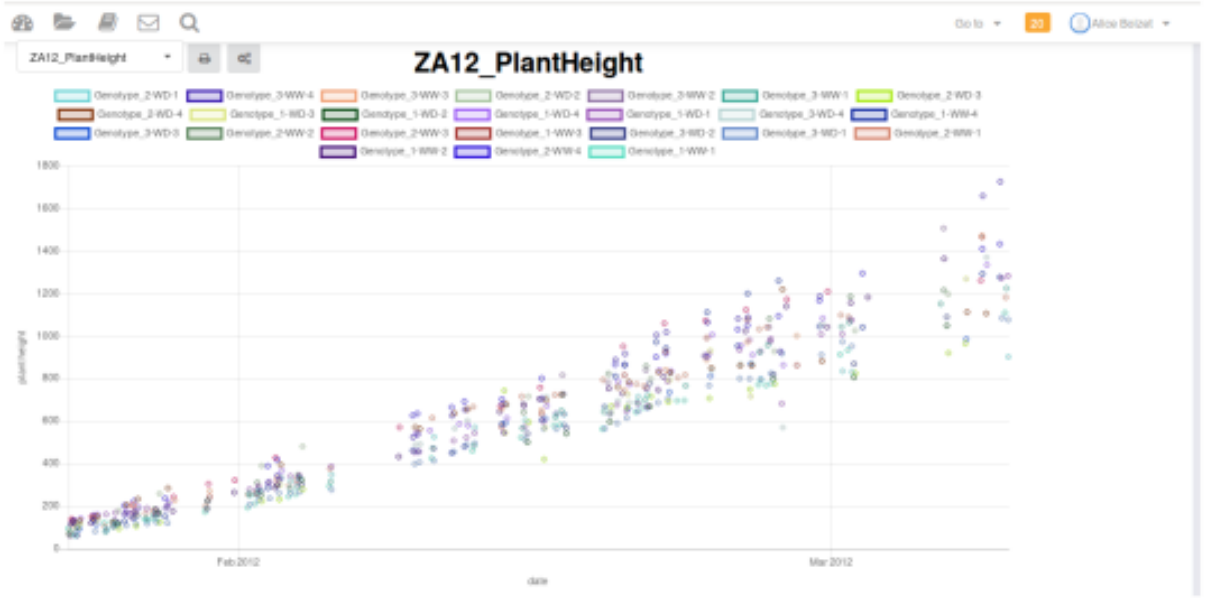
When you have downloaded the csv file, you can go to the *visualization/Create Graphs* to create a new graph.

In Data, upload the csv file you got from the studies exploration tool.

Fill the General tab this way :



You can go to *visualization/Create Graphs* and see your chart.

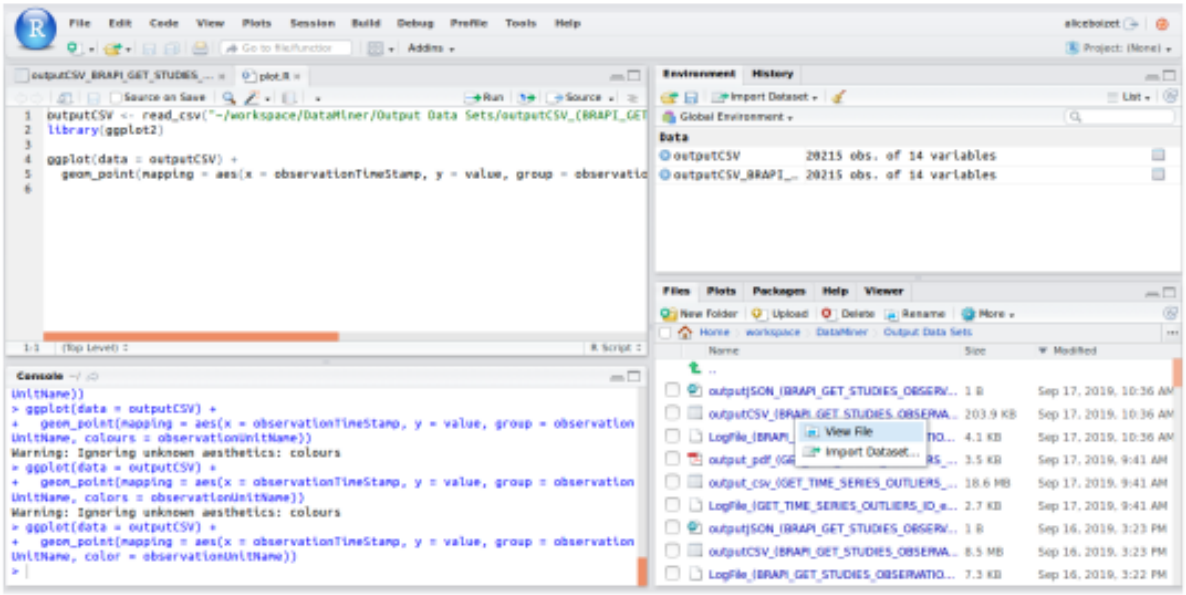


Visualize data in Rstudio

An other good way to visualize is to use R. Indeed there are great libraries such as ggplot2 to build charts. There are also very helpfull galleries with a lot of code examples (https://www.r-graph-gallery.com/)

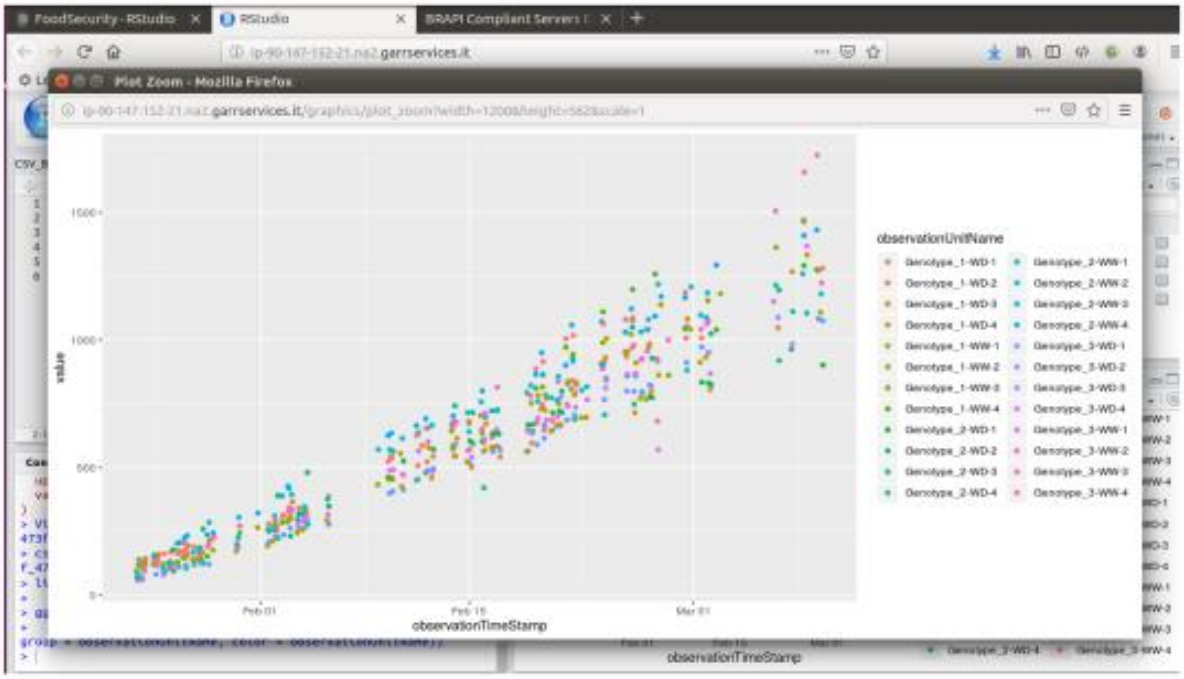
Go to *Analytics/Rstudio* to open **Rstudio**. Find your data csv file (created by the dataminer) in the *Files* tab at the bottom right corner. It is stored in your workspace under *DataMiner/Output Data Sets* folder.

Click on the file to "import Dataset".



You can know use the R console to visualize your data. To create a simple chart, you can use the code bellow which creates the following chart:

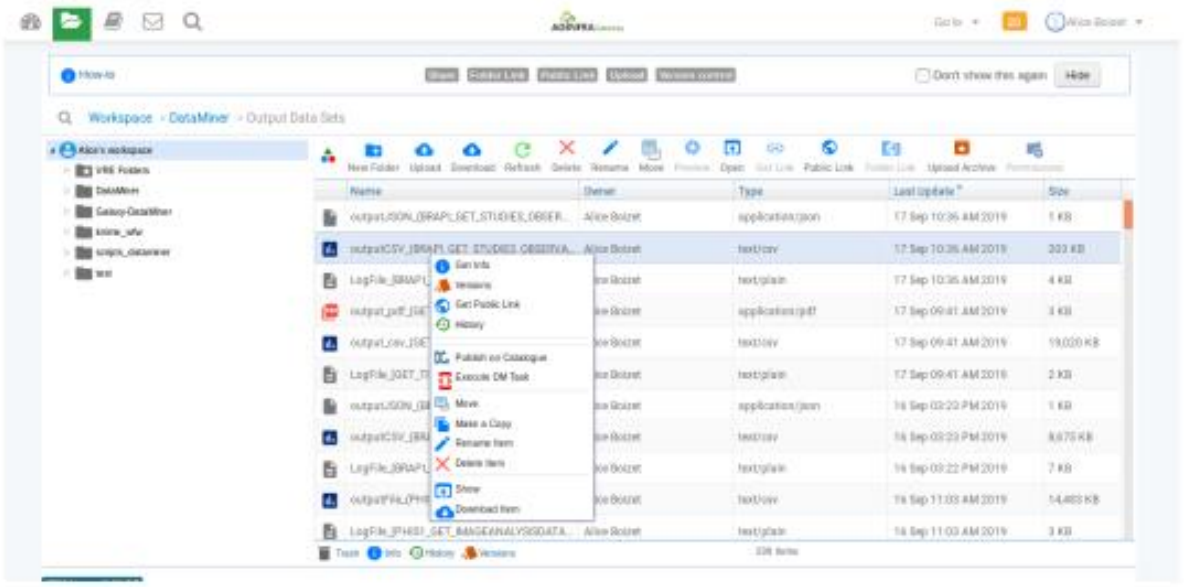
```
library(ggplot2)
ggplot(data = outputCSV) +
  geom_point(mapping = aes(x = observationTimeStamp, y = value, group =
observationUnitName, color = observationUnitName))
```



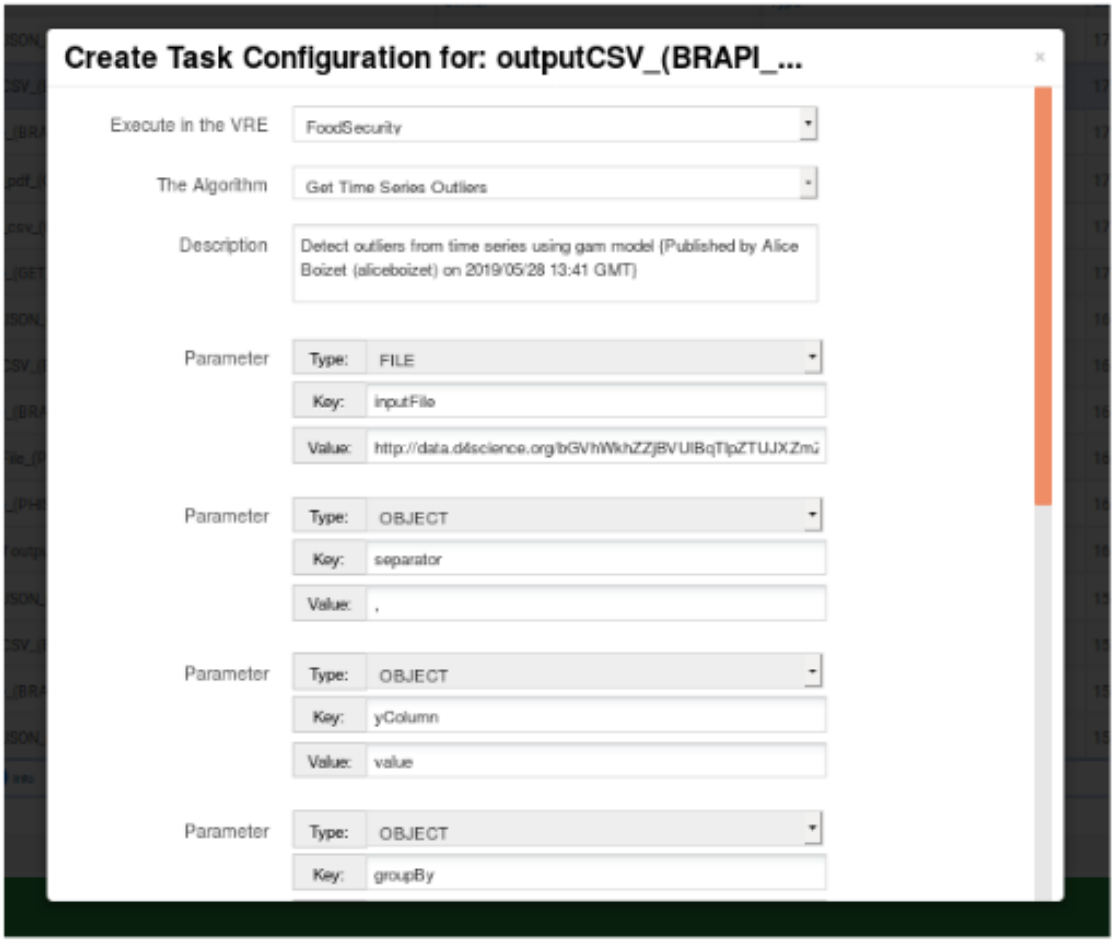
ANALYZE DATA

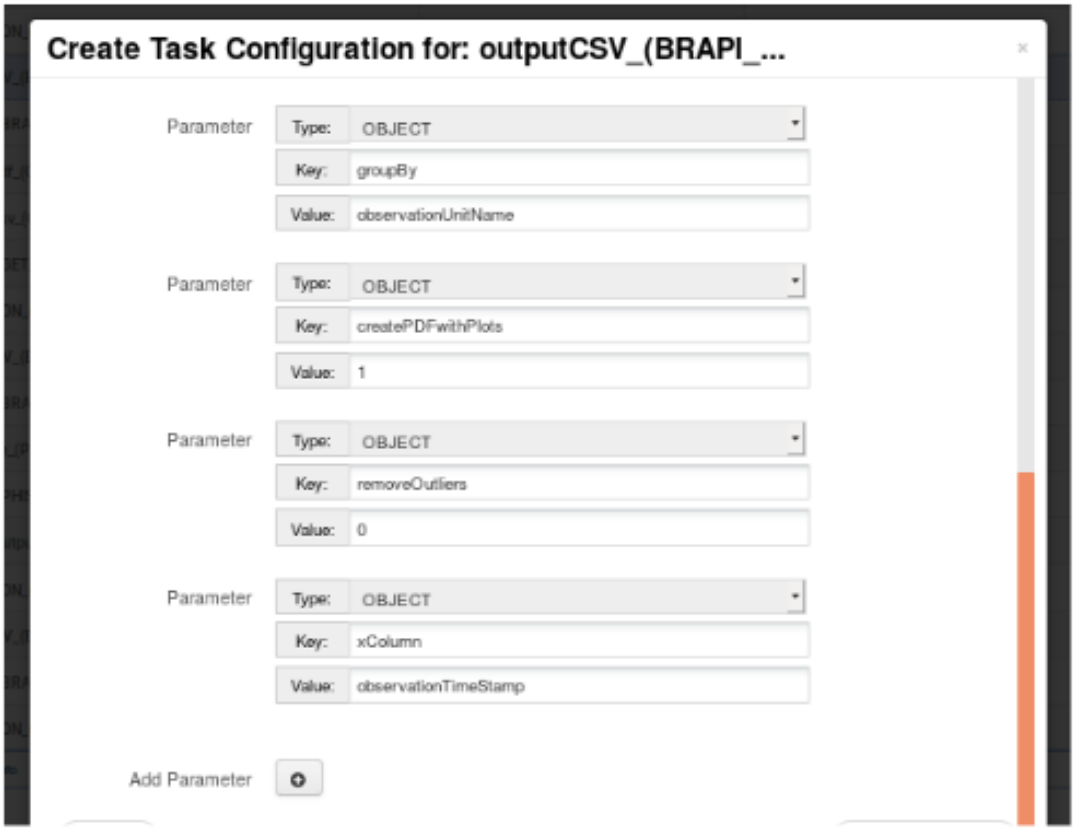
Reuse an algorithm to detect outliers

You would like to detect the outliers on your data. To do that, you could use Rstudio (or even Juppyter) and make a script. But you are very lucky, another VRE member has already done a similar script and he has imported it in the dataminer tool as a black box (the algorithm name is **Get Time Series Outliers**). So all you need is to go to your workspace and right click on your data csv file and select "execute DM task".



Fill the parameters as bellow:


Create Task Configuration for: outputCSV_(BRAPI_...
Execute in the VRE
FoodSecurity
The Algorithm
Get Time Series Outliers
Description
Detect outliers from time series using gam model (Published by Alice Boizet (aliceboizet) on 2019/05/28 13:41 GMT)
Parameter
Type: FILE
Key: inputFile
Value: http://data.d4science.org/bGVhWkhZZjBVUlBqTlpZTUJXZmU
Parameter
Type: OBJECT
Key: separator
Value: ,
Parameter
Type: OBJECT
Key: yColumn
Value: value
Parameter
Type: OBJECT
Key: groupBy


Create Task Configuration for: outputCSV_(BRAPI_...
Parameter
Type: OBJECT
Key: groupBy
Value: observationUnitName
Parameter
Type: OBJECT
Key: createPDFwithPlots
Value: 1
Parameter
Type: OBJECT
Key: removeOutliers
Value: 0
Parameter
Type: OBJECT
Key: xColumn
Value: observationTimeStamp
Add Parameter

Parameters description : If removeOutliers=0 the outliers will be kept and identifiable. If createPDFwithPlots=1 a pdf file with plots of each time serie will be created

Click on *Create configuration*, then run the algorithm.

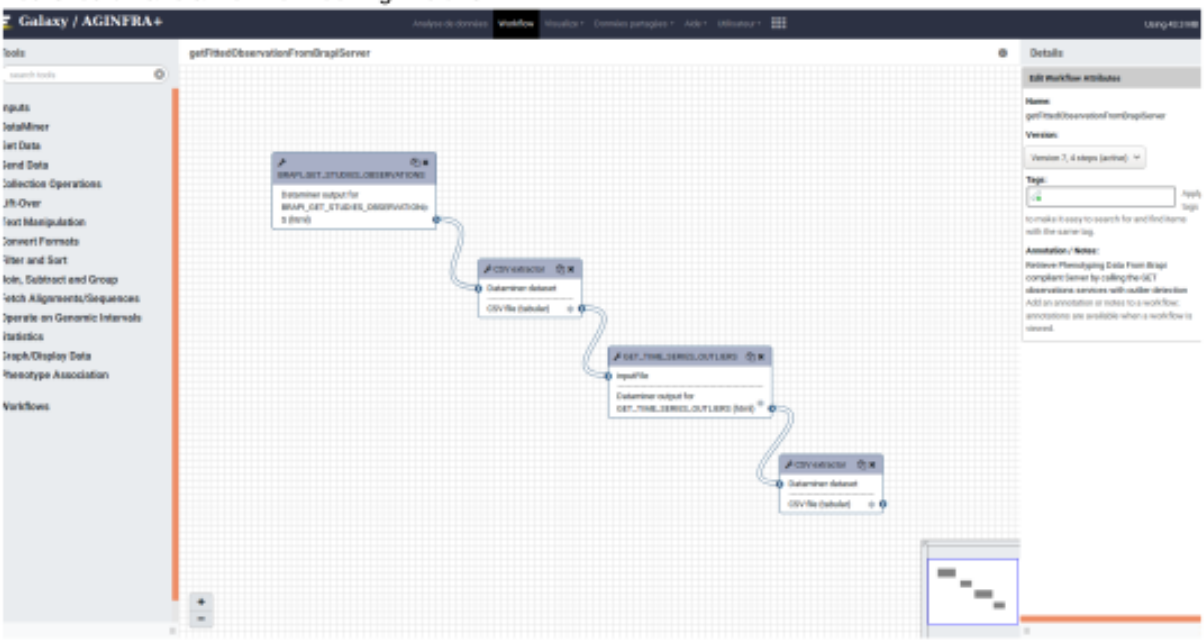
After the execution, you should fine the output files in your workspace (*dataminer/output data sets folder*)

Make a Galaxy workflow

You would like to be able to access data from PHIS studies and detect outliers in a single process as a routine, so that you could easily do the same process on next studies. To do that, you can build a workflow in Galaxy. Galaxy can be found in the analytics tab. The dataminer algorithms are automatically transfered as galaxy tools. Build a workflow which first Retrieve observation data of the study ZA12 and then detects outliers.

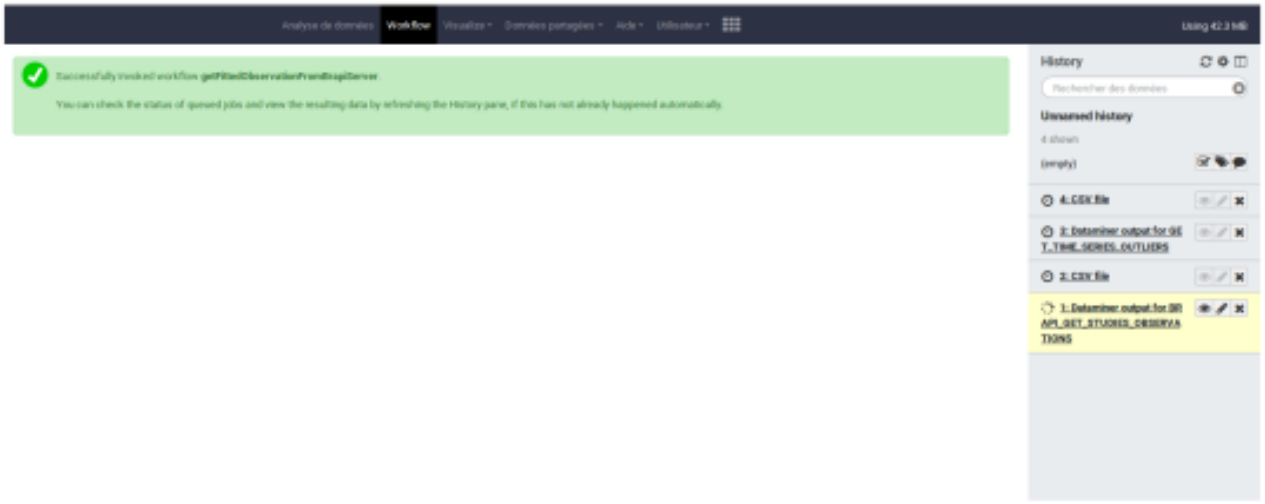
Tip : you need to use node csv_extractor in order to convert the dataminer tools output into csv

You should have a workflow looking like this :



Don't gorget to save you workflow by clicking on 

When your workflow is ready, you can run it. You can see each step execution on the right of the screen.



After the workflow execution, you can view the outputs. You can click on the view button on the 3rd step (Dataminer output for get_time_series) when you can click on the 2 output files to download them.

The csv validator tool enable to convert the dataminer csv output into a galaxy datatable file which means that it can be reused as an input on another galaxy tool and you can visualize the data directly in Galaxy. To do that, you can click on the view button of the last step of the workflow.