This is a pre-print version (the Submitted Version) of the paper Assante, M, Boizet, A, Candela, L, et al. Realizing virtual research environments for the agri-food community: The AGINFRA PLUS experience. Concurrency Computat Pract Exper. 2020;e6087. https://doi.org/10.1002/cpe.6087

ARTICLE TYPE

Realising Virtual Research Environments for the Agri-food Community: the AGINFRA PLUS Experience

M. Assante[1] | A. Boizet[2] | L. Candela*[1] | D. Castelli[1] | R. Cirillo[1] | G. Coro[1] | E. Fernández[3] | M. Filter[4] | L. Frosini[1] | T. Georgiev[5] | G. Kakaletris[6] | P. Katsivelis[7] | M.J.R. Knapen[8] | L. Lelii[1] | R.M. Lokers[8] | F. Mangiacrapa[1] | N. Manouselis[7] | P. Pagano[1] | G. Panichi[1] | L. Penev[5] | F. Sinibaldi[1]

[1]Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", National Research Council of Italy, Pisa, Italy
[2]French National Institute for Agricultural Research, Montpellier, France
[3]EGI Foundation, Amsterdam, Neederlands
[4]Federal Institute for Risk Assessment, Berlin, Germany
[5]Pensoft Publishers, Sofia, Bulgaria
[6]University of Athens, Athens, Greece
[7]Agroknow, AThens, Greece
[8]Wageningen University & Research, Wageningen, Neederlands

Correspondence
*L. Candela, Email: leonardo.candela@isti.cnr.it

Summary

The enhancements in IT solutions and the open science movement are injecting changes in the practices dealing with data collection, collation, processing, analytics, and publishing in all the domains, including agri-food. However, in implementing these changes one of the major issues faced by the agri-food researchers is the fragmentation of the "assets" to be exploited when performing research tasks, e.g. data of interest are heterogeneous and scattered across several repositories, the tools modellers rely on are diverse and often make use of limited computing capacity, the publishing practices are various and rarely aim at making available the "whole story" including datasets, processes, and results. This paper presents the AGINFRA PLUS endeavour to overcome these limitations by providing researchers in three designated communities with Virtual Research Environments facilitating the use of the "assets" of interest and promote collaboration.

KEYWORDS:
Virtual research environment, Agroclimatic modeling, Food safety risks assessment, Food security

1 | INTRODUCTION

The developments in information and communication technologies, including big data availability and management, web and cloud technologies, as well as open science related practices are not yet fully embraced by the Agriculture and Food Science research domain [1,2]. The fragmentation of "resources" needed to exploit these new opportunities across several and heterogeneous "places" is certainly one of the major factors hindering this uptake process, e.g. data are heterogeneous and scattered across several repositories, modelling tools and supporting systems are diverse, the amount of available computing capacity varies a lot across teams and laboratories.

The AGINFRA PLUS project has been set up to develop an innovative approach in Agri-food digital science practices aiming at overcoming the limitations stemming from the above settings by leveraging on existing e-Infrastructures and services. In particular, AGINFRA PLUS promotes the exploitation of *Virtual Research Environments* (VREs) [3] to provide designated communities with seamless access to the data, services, and facilities they need to perform their research tasks collaboratively and efficiently. Such VREs are built by relying on an open and distributed platform providing a rich array of services supporting all the phases of an open science research lifecycle from data collection to data analytics and publication.

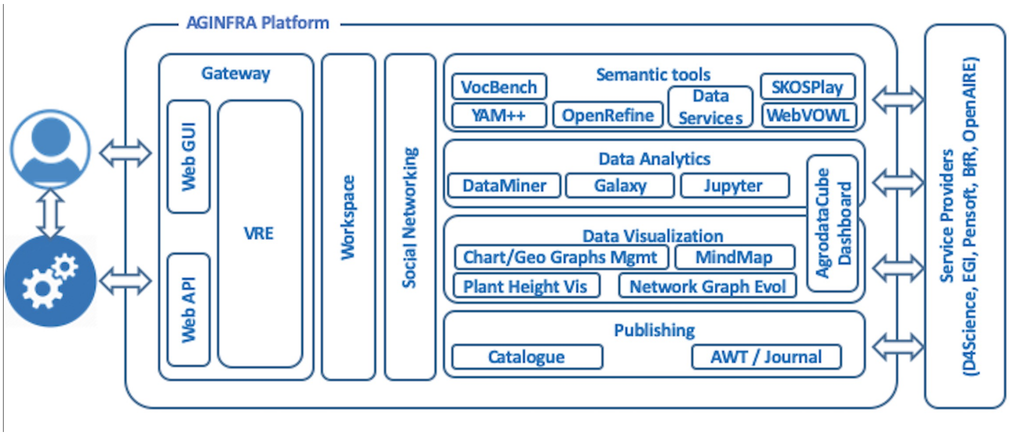


FIGURE 1 AGINFRA PLUS VRE Platform Architecture

AGINFRA PLUS exploited the VRE approach for three prominent agri-food research communities, namely: (*i*) *agro-climatic modelling*, focusing on use cases related to crop modelling and crop phenology estimation, (*ii*) *food safety risk assessment*, focusing on use cases to support scientists in the multidisciplinary field of risk assessment and emerging risk identification, and (*iii*) *food security*, focusing on use cases related to high-throughput phenotyping to support phenomics researchers to select the most suitable plant species and varieties for specific environments.

The remainder of the paper is organised as follows. Sec. 2 presents the major constituents of the AGINFRA PLUS platform. Sec. 3 describes the exploitation scenarios developed by each community and the benefits resulting from the use of the platform. Sec. 4 discusses the achievements. Finally, Sec. 5 concludes the paper by reporting some future works.

2 | THE AGINFRA PLUS VRE PLATFORM

In order to support the AGINFRA PLUS communities, a comprehensive and feature rich platform has been developed and operated supporting the development of several Virtual Research Environments. An overall picture of the platform that build on the "as-a-Service" delivery model is given in Fig. 1.

Such a platform follows the *system of systems* approach[4], where the constituent systems offer "resources" (namely services) for the implementation of the resulting system facilities. In particular, such a platform aggregates "resources" from "domain agnostic" service providers (e.g. D4Science[5], EGI[6], OpenAIRE[7]) as well as from community-specific ones (e.g. AgroDataCube[8], AGROVOC[9], RAKIP model repository[10]) to build a unifying space where the aggregated resources can be exploited via VREs[11]. This system of systems approach is enabled by D4Science. D4Science is at the heart of the overall platform. In fact, this service provider offers the core services to implement the resulting platform, namely: (*a*) the AGINFRA PLUS *Gateway*[12], realising the single access point to the rest of the platform (see Fig. 2); (*b*) the *authentication and authorisation infrastructure*, enabling users to seamlessly access the aggregated services once managed to log in the gateway; (*c*) the *Workspace*, for storing, organising and sharing any version of a research artefact[13], including dataset and model implementation; (*d*) the *Social Networking* area enabling collaborative and open discussions on any topic and disseminating information of interest for the community, e.g. the availability of a research outcome[13]; (*e*) the *Catalogue* recording the assets worth being published thus to make it possible for others to be informed and make use of these assets[13].

These basic facilities are complemented by services for the semantic-oriented management of data, data analytics, data visualization, and publishing.

2.1 | Semantic Tools

The AGINFRA PLUS *Semantics Tools* offer services for managing semantic resources (e.g. ontologies, thesauri, vocabularies) and for benefitting from such resources in tasks related with data management. The supported facilities include: (*a*) an *ontology engineering service* for creating, editing and managing semantic resources and, at the same time, catering for their collaborative

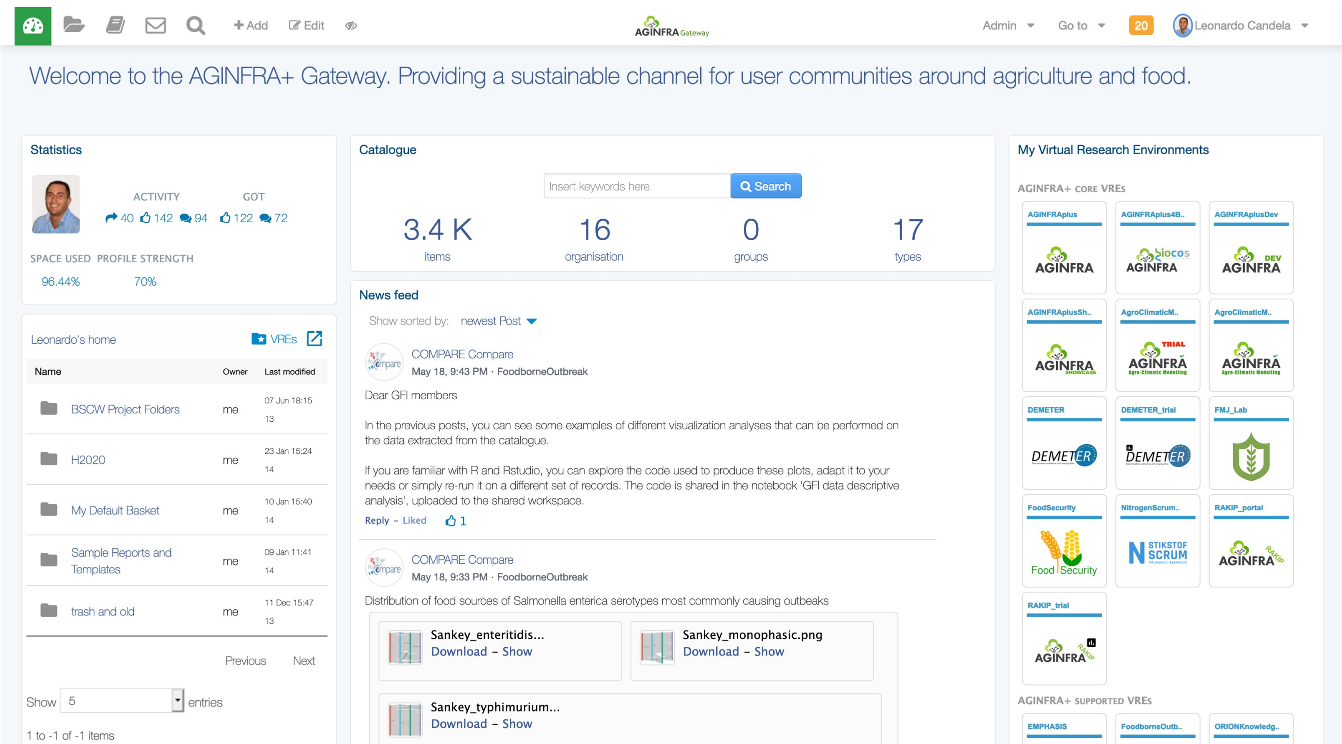


FIGURE 2 AGINFRA PLUS Gateway: the Dashboard

design, editing and management. It is based on VocBench[14], a web-based platform for managing OWL ontologies, SKOS thesauri and RDF datasets; (*b*) a *semantic linking service* supporting the establishment of semantic links between data items belonging to different datasets and different sources. It is based on Silk[15], a web-based platform enabling users to manage diverse datasources, linking tasks and transformation tasks; (*c*) a *data transformation service* promoting the RDF-isation of tabular data, i.e. a user can determine the rules for transforming the data into triples using arbitrary schemas and ontologies. In practice, it supports the building of an RDF skeleton for defining how cell values will be translated in RDF. It is based on the open-source OpenRefine tool[16], a powerful tool for data cleaning and transformation including a plug-in for RDF-isation; (*d*) an *ontology visualisation service* supporting users to upload and / or import ontologies and visualise the graph corresponding to the ontology. Classes and instances are represented as circular nodes and properties are represented as edges between these nodes. A side panel giving information on entity as defined in the ontology completes the offering. It is based on WebVOWL[17], a web-based tool for the interactive visualisation of ontologies; (*e*) an *ontology alignment service* facilitating users in establishing mapping between two diverse ontologies or thesauri. It is based on YAM++[18], a web tool proved to be effective and scalable in ontology matching tasks.

2.2 | Data Analytics

The AGINFRA PLUS *Dnalytics Analytics* offer services for the challenging task of big data analyticsoffer a rich array of services for the challenging task of big data analytics[19]. The supported facilities include: (*a*) a *data analytics platform* to execute analytics tasks either by relying on methods provided by the user or by others[20]. It is endowed with importing and sharing facilities for analytics methods implemented in heterogeneous forms including R, Java, Phyton, and KNIME[21] (largely used by the food safety community). The platform enacts tasks execution by a distributed and hybrid computing infrastructure including EGI resources. Moreover, one of the worth highlighting feature of this platform is its open science-friendliness. All the analytics methods integrated in it are exposed by a standard protocol (the OGC WPS protocol) clients can use to get informed on available methods as well as to start processes, monitor their execution and access results. Every analytics task performed by the platform automatically produces a provenance record catering for the repeatability of the task; (*b*) an RStudio-based *development environment for R* enabling to perform statistical computing tasks in the cloud. The environment provide its users with a powerful

IDE including a console, a source code editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. (*c*) a Jupyter-based *notebook environment* for documenting and recording analytics processes[22]. Every notebook is a rich document that contain live code, equations, visualizations and narrative text aiming at capturing a research activity; (*d*) a Galaxy-based *workflow management workbench* for combining several analytics tasks into workflows[23]. In practice, if offers a means to build multi-step computational analyses by specifying what data to operate on, what steps to take, and what order to do these steps in.

All these platforms and environments are nicely integrated each other as well as are integrated with the rest of services offered by AGINFRA PLUS. For instance, every method integrated in the data analytics platform can be easily executed by a Jupyter notebook or by a Galaxy workflow. All these tools are equipped with solutions facilitating the access to the workspace content thus to make use of it during the processing steps, e.g. to use files as inputs or to store results. It is straightforward to publish every analytics process implemented by these tools into the *Catalogue* to share it with coworkers.

2.3 | Data Visualization and Publishing

The AGINFRA PLUS *Data Visualization and Publishing* facilities provide users with feature-rich, flexible solutions for generating representations (e.g. charts) out of datasets and publishing "research objects", documenting a research activity and its results. Via those solutions not only delivery of knowledge is facilitated but, what is more significant from the perspective of Open Science, the researcher is given the means to guarantee reproducibility and repeatability of the underlying analysis process. Technologically, the AGINFRA PLUS data visualisation and publishing offering is delivered under the SaaS model and operates on the platform that allows the realisation of a pipeline where data is in focus from its discovery and through its processing, to its delivery.

The data visualisation offerings fall under two categories: (*a*) *common data visualisation*, which covers generic types of visualisations offered by various other platforms on common data formats, yet integrated in the landscape of AGINFRA PLUS, as well as special visualisations that can fit several use-cases, and (*b*) *case-specific data visualisation*, which attempts to respond to innovative community driven data visualisation and interaction concepts and combine ad-hoc and generic visualisation facilities with platform' services in order to deliver an integrated user experience.

Under the first category fall the following tools: (*i*) A *charts management workbench* for creating several types of interactive charts ranging from common generic ones (e.g. Spline, Scatter, Bar, Line, Step, Pie, Doughnut, Polar); The workbench provides users with facilities to import datasets of interest, to configure how they has to be utilised to produce the visualisation of interest, to adjust the visualisation styling and to share the produced graphical outcome; (*ii*) A *mind map tool* for creating, manipulating, and visualizing mind maps with features that support interoperable mind-map definitions, visual auto-arrangement of concepts and several styling configuration options; (*iii*) A *geospatial data visualisation framework* (GeoVisualisation) that handles geospatial data ingestion, rendering, and publication through a number of common standards; (*iv*) A *network evolution visualisation tool*, that allows users to create animated series of network structures (nodes and edges) aiming to highlight the changes of the graph over time. The tool offers a variety of formatting options that allow users to and capabilities of video exports.

Under the second category, AGINFRA PLUS offers (*i*) the *Agrodatacube dashboard*, an instrument that binds together data services, that offer geotemporal crop data and metadata, with a geospatial exploration backdrop to support from drilling into crop information and rendering of various crop data charts to the execution of crop models for the generation of new simulation data; (*ii*) A plant height exploration and visualisation tool, which binds together data and images from plant growth monitoring into an easy to navigate and explore front end, that can present graphical representations depicting plant temporal (timeseries charts) and spatial (heatmap) characteristics.

Publishing facilities cover the following perspectives: (*a*) A *catalogue-based information publishing service* to disseminate artefacts according to the FAIR principles[24]. This service[13] makes it possible to customise, per domain, the typologies of items to be published by carefully defining their metadata (attributes, possible values, constraints) and some management triggers (e.g. what values should be transformed in tags, what should lead to groups). Moreover, catalogue items are expected to be endowed with "resources" representing the payload of any item. Therefore, by using catalogue item resources it is possible, for example, to execute a model, to access a dataset, to visualize a graph; (*ii*) A *research community dashboard* realising a domain specific access point to search for research content of interest, be it publications, data, projects. This is based on the OpenAIRE specific service[25] enabling to publish research products and interlink them with the OpenAIRE scholarly communication cloud. (*iii*) A *scholarly publishing platform* integrated with Pensoft infrastructure, namely the ARPHA Publishing Platform[26] to enable the creation of innovative papers including datasets and methods hosted by the AGINFRA PLUS platform (see Sec. 3.4). By relying on this platform, users are allowed to mix the narrative of a traditional paper with links aiming at giving effective access to the digital version of the research products.

3 | EXPLOITATION SCENARIOS

AGINFRA PLUS aimed at the development of technical resources that could bring open science forward in agri-food research. To develop and test such new resources the AGINFRA PLUS project collaborated with a range of scientific user communities from the agri-food domain and implemented and evaluated customized Virtual Research Environments supporting specific pilots for a number of practically relevant use cases in these communities.

3.1 | Supporting Agroclimatic Modeling Scenarios

One of the target communities for AGINFRA PLUS is the agro-climatic modelling community. This group of researchers focusses on developing and calibrating agro-environmental models and algorithms and applying these in research in the agro-environmental and food security domains. They use a variety of agro-environmental data, for instance agronomic data (like crop parameters and crop calendars), weather and climate data, soil data, and remote sensing data to determine the behaviour and development of crops under different conditions. Applications differ considerably depending on the focus of the involved researchers and practitioners. However, some specific characteristics are particularly crucial for the work of this community.

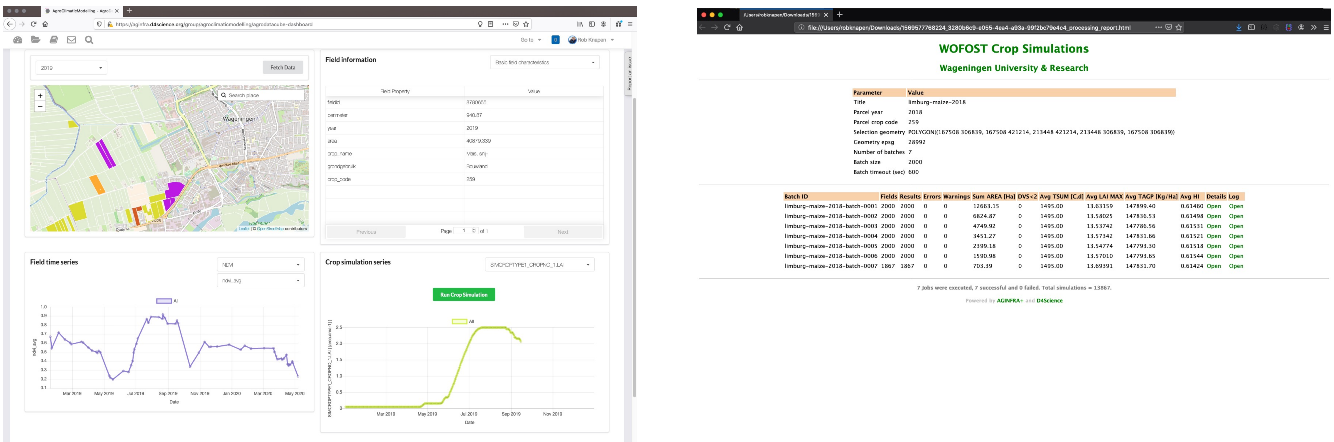
First, data used is generally highly heterogeneous, coming from a variety of sources, in different formats, and implementing different standards. Secondly, it is common that at least part of it concerns spatiotemporal information, requiring specific analytics tools that can handle such data. Moreover, such data typically is large and requires substantial storage and compute resources for processing. Lastly, relevant data for this domain is becoming available at ever increasing speed and finer resolutions, e.g. near real-time remote sensing data from satellites, and from IoT devices, allowing for analytics at higher scales such as at farm crop field levels. Again, this comes at the cost of higher computing demands, surpassing the commonly used computing environments such as laptops, single desktop computers, or isolated modelling servers.

Hence, resources provided by a platform such as AGINFRA PLUS and made available to researchers via Virtual Research Environments could be helpful to cope better with these new challenges, while its services can help promote open science within the community.

One of the AGINFRA PLUS use cases for the agro-climatic modelling community that was explored and for which typical research applications were implemented and deployed into a VRE, with the aim to demonstrate and evaluate the usability for its researchers, is crop growth modelling. Simulations using crop growth models are one of the important components in yield forecasting, used frequently in food security research and related research areas. Currently, the application of crop growth models at large scale with high resolution is often still limited by the available computing resources. The application piloted in AGINFRA PLUS, applying European or global scale crop simulations on the detailed level of agricultural parcels is currently too demanding for many existing research infrastructures in the field. To meet the requirements for such large scale, high-resolution crop growth modelling exercises, the following facilities are indispensable: efficient retrieval of spatiotemporal data streams, spatiotemporal data wrangling and data processing, running models at scale using distributed computing resources and parallel technologies computing, and intuitive spatiotemporal visualisation. In the AGINFRA PLUS crop growth modelling use case, the preprocessing of spatiotemporal data is an integral part of the AgroDataCube infrastructure[8]. This infrastructure provides Dutch agricultural open data as a service to research and business. The AgroDataCube ingests and merges different spatiotemporal data streams that are relevant for agricultural and environmental applications (among others weather data, agronomic data, parcel geometries, Sentinel-2 satellite data, and soil data). It provides a set of well-documented, ready to use REST services that allow retrieval of the merged data on the parcel level in usable packages and a standardised format (GeoJSON). To cope with the requirement to scale up simulations, the widely used WOFOST crop growth simulation model[27] was embedded into a distributed computing framework that facilitates the distribution of computing jobs over a compute cluster[28]. The resulting modules have been integrated into the VRE as DataMiner algorithms, and were published, using the Catalogue service, to make them discoverable and reusable as FAIR algorithms for the whole community. As the requirements for spatiotemporal visualisation in this use case were above what is offered by standard components, a dedicated visualisation dashboard, developed on the basis of various VRE components, was developed.

For visual inspections, the dashboard allows geospatial and temporal visualisation of the various data services provided through the AgroDataCube services that are input to the crop growth simulations. Moreover, it offers its users the opportunity to manually search for, and select a specific agricultural parcel and initiate a WOFOST crop simulation by executing the VRE DataMiner algorithm using input data based on the selected field. Generated simulation results are stored on the VRE's shared workspace, and the simulated parameters can be visualised as graphs and can be compared and analysed side by side with the used input data. After being tested and quality checked, the developed models and algorithms and the generated output data can be shared with the broader user community, by publishing them through the VRE's Catalogue service. Thus, the VRE is complying with the requirements of FAIR data services and open science in general, adding to that the opportunity to also share algorithms and models in a FAIR manner.

The Virtual Research Environment developed for supporting this scenario is available by the Gateway[†]. Figure 3 shows two screenshots of the VRE: Fig. 3a shows the AgroDataCube dashboard for displaying and accessing data from the AgroDataCube, and executing a crop simulation for the selected field; Fig. 3b shows a report of a crop simulation report for several batches of crop simulations.



(a) The Agrodatacube Dashboard

(b) A report of a crop simulation task

FIGURE 3 AgroClimaticModeling VRE Screenshots

3.2 | Supporting Food Safety Risk Assessment Scenarios

In the domain of food safety modelling two application scenarios were identified where scientific data analysis workflows and software based resources for knowledge sharing and integration are of extraordinary importance. Both scenarios nicely complement the activities the community is promoting to harmonise the knowledge produced[29].

In the first application scenario the VRE serves the need of the One Health research community that aimed to create an online One Health EJP Glossary that could be accessed, maintained and curated in a collaborative manner. This application scenario largely builds upon the VRE Catalogue and Social Networking services as well as on a number of dedicated KNIME-based data handling workflows that support the automatic generation of new or maintenance of existing glossary entries. In addition this new VRE-based OHEJP Glossary resource serves as a service endpoint for external web-services that consume its content via a RESTful API.

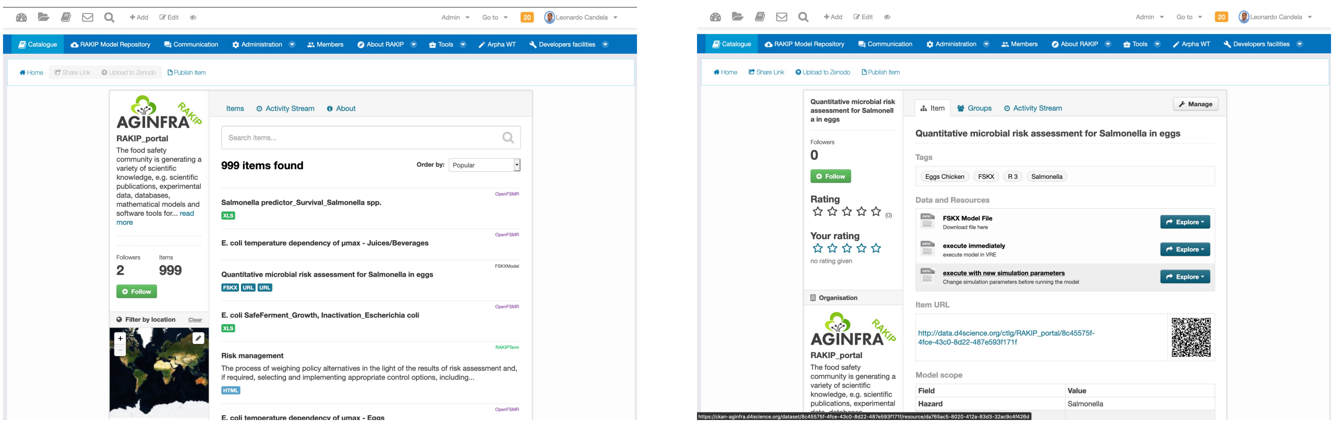
In a second application scenario risk assessors and risk modellers were provided with an integrated environment that facilitates the sharing of data, mathematical models and simulation results relevant for risk assessments in a harmonized way. A distinguishing feature of this environment is a community-driven food safety model repository, that contains mathematical models from the area of predictive microbial modelling and quantitative microbial risk assessment (QMRA). This repository builds

[†] AgroClimaticModeling VRE https://aginfra.d4science.org/web/agroclimaticmodelling

upon the open source FSK-Lab KNIME extension[30] and the Food Safety Knowledge Markup Language, a new community standard to harmonize the annotation and exchange of script based food safety models that allows to include all relevant input and output data, metadata as well as model scripts in a machine-readable format. By providing the option to execute the open source KNIME and FSK-Lab software in AGINFRA PLUS VREs it became possible to even execute FSK-ML compliant models in the new community VRE.

The AGINFRA PLUS platform supported these scenarios by further providing: (*a*) facilities for developing an ontology supporting the FSK-ML standard for model exchange (VocBench, see Sec. 2.1); (*b*) extending the VRE GUI by specific features that allow to trigger dedicated DataMiner workflows from the workspace that in the case of the RAKIP VRE support the automatized publishing of a model metadata into the VRE Catalogue or the execution of a selected model file; (*c*) extending the Catalogue resource such that models can be described according to the FSK-ML community metadata and endowed Catalogue entries with three actionable resources enabling users to respectively download the model as FSK-ML file, perform a model simulation with the default simulation parameters settings or perform a model simulation by customizing the simulation input parameters; (*d*) a mind map visualization component facilitating knowledge representation and exchange among the VRE members; (*e*) a journal-based approach for publishing the models (see Sec. 3.4).

The Virtual Research Environments developed for supporting these scenarios are available by the Gateway[‡]. Figure 4 shows two screenshots showcasing the variety of items published by this community and a catalogue item for a model with links enacting to access the model as well as to execute it.



(a) The RAKIP Catalogue

(b) A RAKIP catalogue item

FIGURE 4 RAKIP VRE Screenshots

3.3 | Supporting Food Security Scenarios

The Food Security Community is focusing on a high-throughput plant phenotyping scenario. This scenario can help to select crop varieties that better adapt to global changes in order to respond to the food security challenges. High-throughput phenotyping produces a large amount of data which need to be integrated and analysed right away. For example, in a greenhouse platform, a lot of images of plants are taken: 13 images per plants per day are taken in the Montpellier platform which works on 1600 plants (more than 20,000 images per day). Field platforms produce and need a lot of images including UAV or satellite. High-throughput phenotyping platforms produce complex data (sensors data, human reading) at different scales (e.g. population, individuals, molecular).

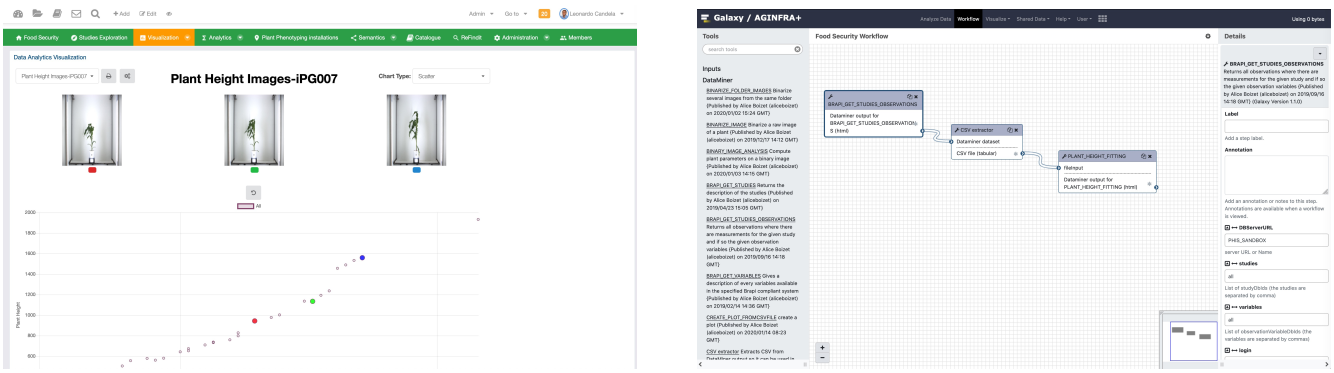
The phenomics community needs tools to easily access to large datasets and to be able to visualize and analyse them. Moreover, sharing data, analytics process and results is essential. The objective of this use case is to develop a VRE for phenomics researchers where these users: have access to relevant ontologies; collaborate on building and share semantic resources; have

[‡]OrionKnowledgeHub https://aginfra.d4science.org/web/orionknowledgehub/ and RAKIP https://aginfra.d4science.org/web/rakip_portal

access to phenomics platforms data from the information system OpenSILEX-PHIS[31]; visualize data; import and run data analytics scripts in different languages (R, Python, etc); import or update and run data analytics workflows (KNIME, Galaxy); share results and work with other users.

The first evaluation results of the Food Security VRE indicated that the VRE is useful for collaborative work. The diversity of tools that are available has also shown interest from the users. However, there were some reservations on the integration of these which made difficult the execution of certain data analysis workflows. Another concern on big data manipulation and data access has also been noted. Considering this, recent work has been made to improve these integration capabilities in order to provide better connected tools. Web Services based on the Breeding API standards[§] had also been implemented into the OpenSILEX-PHIS system in order to easily access phenotyping data in the VRE.

The Virtual Research Environment developed for supporting this scenario is available by the Gateway[¶]. Figure 5 shows two screenshots from this VRE: Fig. 5a displays a plant-heigh graph enabling to easily display and compare plant images while Fig. 5b displays the Galaxy facility integrated into the VRE where the algorithms of DataMiners are handled as tools on pair with the rest of Galaxy tools.



(a) A plant height data visualization graph **(b)** Galaxy embedded int a VRE

FIGURE 5 FoodSecurity VRE Screenshots

3.4 | Realising Executable Papers and Supporting the Food Modeling Journal

It is a fact that traditional publications tend to fail to address the needs of scientific communication, information exchange and scientific quality control these days[32]. Many publishers are therefore implementing advanced forms of publications aiming at making available the narrative of the specific research activity together with the datasets, the processes and the software underlying it[26,33,34,35]. Within AGINFRA PLUS the ARPHA Publishing Platform has been integrated into the VREs (see Sec. 2.3), so it is now straightforward for the users of a VRE to access ARPHA Publishing Platform. Further, a set of new tools and workflows have been developed to address part of these requirements and create novel ways for researchers to automatically import and publish dataset and models in the field of agricultural and food sciences.

Figure 6 displays the process leading to the publication of an *executable paper*, i.e. a paper enacting its readers to seamlessly access the content and execute the processes underlying the research activity described in the paper. The workflow starts with the provisioning of the software code implementing the algorithms and its data via the workspace service (as well as by GitHub). This material is integrated into the data analytics platform (see Sec. 2.2) by using a straightforward process requesting the algorithm owner to instruct the platform on specificities needed to execute the algorithm like input and output parameters, and runtime dependencies. This way the algorithm becomes an executable process that can be invoked by the data analytics platform as well as by clients invoking it programatically via WPS[20]. The overall artefact is now suitable for being published into the catalogue so as to make it possible its discovery and access (every catalogue item representing a DataMiner integrated algorithm contains metadata and links facilitating its use by "one click"). The ARPHA Publishing Platform integrated into VREs has been

§Breeding API https://brapi.org
¶FoodSecurity VRE https://aginfra.d4science.org/web/foodsecurity

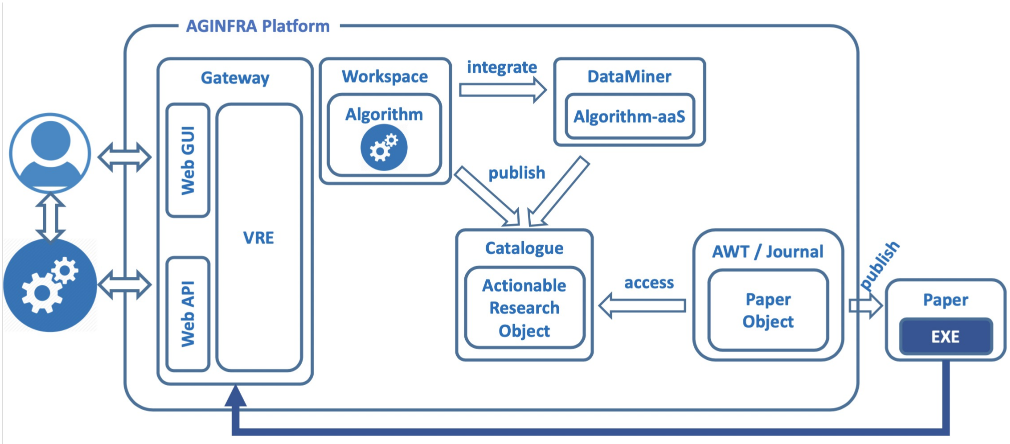


FIGURE 6 AGINFRA PLUS Executable Papers Publishing Process

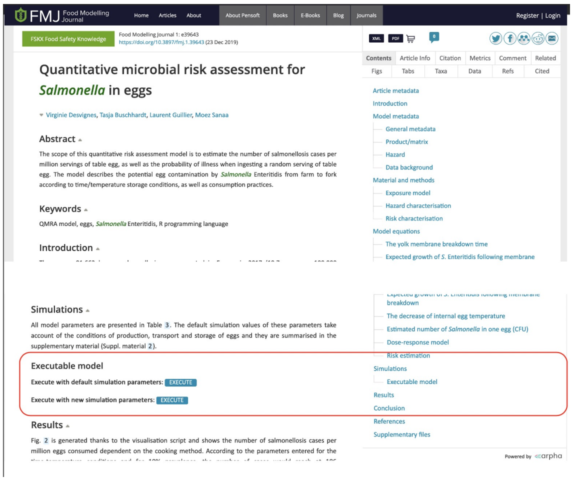
equipped with a facility implementing the transformation of a catalogue item into a paper, i.e. transforming the metadata into the paper content and taking care of the links pointing to the resources. This make it possible to link efficiently and effectively almost any artifact produced during a research activity with the narrative produced by the authors to communicate it.

A specific example of the above approach has been implemented by the Food Modeling Journal (FMJ)[#][36]. This new open access journal has been designed and launched to support the needs of the risk assessment and food modelling community (see Sec. 3.2). It promotes the publishing of scientific research outcomes in novel article formats like the "Executable Model Paper" based on the information exchange standard Food Safety Knowledge Metadata Language (FSK-ML)[29]. FMJ publishes also articles on data analytics pipelines, applied studies, data papers, or software descriptions. In FMJ new journal articles can be started directly from VRE and import data and metadata from any research object provided in a VRE Catalogue. Moreover, the steps needed to integrate every model developed by using the community developed FSK-Lab KNIME extension[30] into the data analytics platform and make the model available and executable have been completely automatised. It is thus sufficient to upload a FSK-Lab-based archive into the workspace and execute a one-click procedure to have this archive transformed into an executable data miner process and a suitable catalogue item. The catalogue item information is automatically provided into predefined FMJ article sections that can then be modified and finalized by the scientists in a collaborative manner using the ARPHA Writing Tool (AWT). AWT allows also to submit articles into other Pensoft's journals, in each case the article passes then a thorough peer review process. It is further possible to include additional files (e.g. simulation results generated by the computational models) into these article. Thanks to the close integration between the ARPHA publishing platform and the VREs, reviewers of an article are now able to control the description of a model by executing the models and comparing own simulation results with the results described by the authors. In a similar way any readers of a so called "Executable Model Paper" can execute the model with default parameters or with custom parameters. This feature allows readers and reviewers to reproduce scientific results, such as is showcased in the exemplar article of Desvignes et al.[37]. An excerpt of this article is in Figure 7 to showcase the level of integration achieved.

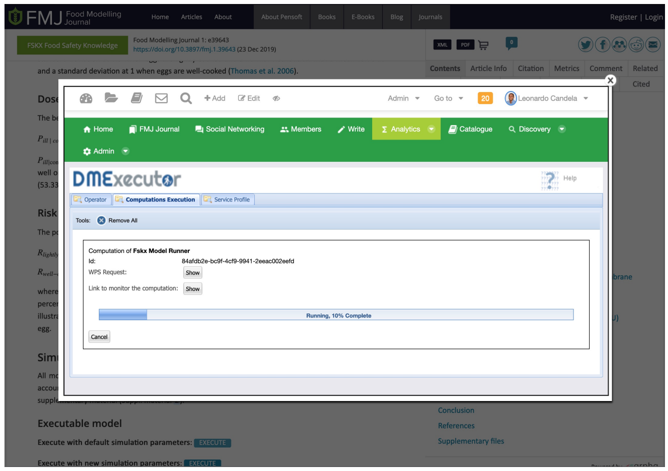
4 | DISCUSSION

The AGINFRA PLUS discussion paper[38] provides a comprehensive view on the current settings characterising the agri-food domain. In particular, this paper recalls the willingness to set up "a pan-European e-Infrastructure for Open Science in Agricultural and Food Sciences". It reports the roadmap agreed by more than 100 stakeholders for the development of this infrastructure based on the "*Share*, *Connect*, *Collaborate*" motto as guide through its evolution: (*i*) "*Sharing: of the resources of relevance to the scientific process (data, models, papers, etc). Open science is only possible if one is able to share one's research first.*"; (*ii*) "*Connecting: available resources need to be connected to allow integration and tackling large scale and more ambitious*

[#]Food Modeling Journal (FMJ) https://fmj.pensoft.net



(a) Executable Model Paper Excerpt

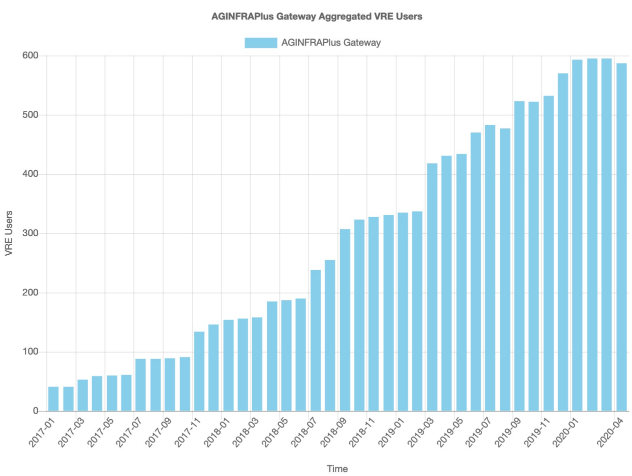


(b) Model Execution by a VRE

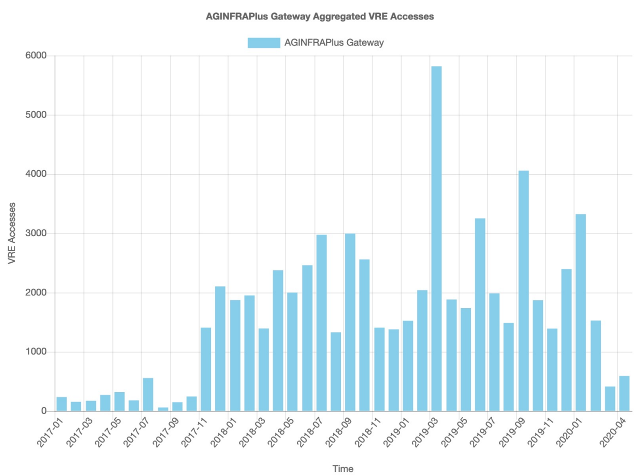
FIGURE 7 FMJ Executable Model Paper

questions in science."; (*iii*) "*Collaborating: the research community itself needs to collaborate beyond ad-hoc arrangements to create, maintain and supply domain specific resources for open science in a network of regional or domain nodes.*".

The platform resulting from the AGINFRA PLUS project has been developed by taking these principles into account. This holistic view of end-to-end scientific workflows powered by an open science platform has been piloted, tested and illustrated with concrete cases by demonstrating the overall suitability and effectiveness. In particular, AGINFRA PLUS strongly promoted a Virtual Research Environment development approach based on *co-creation*. Co-creation is a widely used term to describe participative models where several actors together generate and develop a common "meaning"[39]. It is also known as 'participatory' or 'cooperative' design and one of the important elements characterising the process is the continuous use of prototypes as a mechanism to bring ideas and new features to life and to generate feedback. The overall approach is aiming at innovation by recognising that radical and adaptive creativity can co-exist when generating new products, by recognising that ideas tend to be better developed by groups working together than by selected lead-users. This approach proven to be really effective to deal with the issue affecting the building of systems and infrastructures aiming at providing designated communities with services to use usually described by the "*build it and they will come*" or the negative formulation "*build it, but they don't come*"[40,41,42].



(a) User Base



(b) Working Sessions

FIGURE 8 AGINFRA PLUS VREs User Base and Working Sessions

Overall, the AGINFRA PLUS platform is currently serving hundreds of users (circa 600 in Apr. 2020) by 16 active VREs (Fig. 8). Since January 2017 (when the creation of the platform started) the users served by this platform and its VREs performed: a total of 65,848 working sessions, with an average of circa 1,646 sessions per month; a total of 5,528 social interactions, with an average of circa 138 interactions per month; a total of 13,150 analytics tasks, with an average of circa 328 tasks per month; a total of 3,364 publications of items into the catalogue including models, research objects, methods, services, terms, and datasets.

5 | CONCLUSION

This paper presented the AGINFRA PLUS platform, a science gateway providing the Agri-food community with a rich array of services oriented to promote the implementation of open science practices. Such a platform is currently supporting three designated communities dealing with crops simulation, food safety risk assessment, and high-throughput plant phenotyping scenarios.

The platform brings into these communities and their working practices a number of benefits including (*a*) the *simplicity for coworkers to perform collaborative work*, e.g. the workspace is a working area users can count on to collaborate, the social networking is a means to have informed dialogues; (*b*) the *easiness to share results of any form* within and across the boundaries of their communities and the platform itself, e.g. the catalogue is a valuable service for disseminating research artefacts and enable users to access them, the integration with the OpenAIRE dashboard and the scholarly communication platform reduces the gaps with the scholarly communication domain; (*c*) the attention dedicated to *ease the flowing of existing artefacts into the platform* thus to reduce fragmentation and facilitate their reuse, e.g. the plethora of programming languages and approaches supported by the analytics facilities make it possible to easily integrate almost any existing analytics method, the array of solutions for ontology management facilitate their reuse.

ACKNOWLEDGMENT

This work has received funding from the European Union's Horizon 2020 research and innovation programme under AGINFRA PLUS project (grant agreement No. 731001).

References

1. Jones JW, Antle JM, Basso B, et al. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems* 2017; 155: 269 - 288. doi: https://doi.org/10.1016/j.agsy.2016.09.021
2. e-ROSA Consortium . A Roadmap for a pan-European e-Infrastructure for Open Science in Agricultural and Food Sciences. e-ROSA Roadmap; 2018.
3. Candela L, Castelli D, Pagano P. Virtual Research Environments: an Overview and a Research Agenda. *Data Science Journal* 2013; 12: GRDI75-GRDI81. doi: 10.2481/dsj.GRDI-013
4. Maier MW. Architecting Principles for Systems-of-Systems. *INCOSE International Symposium* 1996; 6(1): 565-573. doi: 10.1002/j.2334-5837.1996.tb02054.x
5. D4Science Consortium . D4Science: an e-Infrastructure supporting Virtual Research Environments. www.d4science.org; .
6. EGI Foundation . EGI e-Infrastructure. www.egi.eu; .
7. OpenAIRE Consortium . OpenAIRE: the European Scholarly Communication Data Infrastructure. www.openaire.eu; .
8. Janssen H, Janssen S, Knapen M, et al. AgroDataCube: A Big Open Data collection for Agri-Food Applications. agrodatacube.wur.nl; 2018

9. Caracciolo C, Stellato A, Morshed A, et al. The AGROVOC Linked Dataset. *Semantic Web* 2013; 4(3): 341-348. doi: 10.3233/SW-130106

10. German Federal Institute for Risk Assessment . FoodRisk-Labs. https://foodrisklabs.bfr.bund.de/foodrisk-labs/; .

11. Assante M, Candela L, Castelli D, et al. The gCube system: Delivering Virtual Research Environments as-a-Service. *Future Generation Computer Systems* 2019; 95(n.a.): 445-453. doi: 10.1016/j.future.2018.10.035

12. AGINFRA Consortium . The AGINFRA Gateway. https://aginfra.d4science.org/; .

13. Assante M, Candela L, Castelli D, et al. Enacting open science by D4Science. *Future Generation Computer Systems* 2019; 101: 555 - 563. doi: 10.1016/j.future.2019.05.063

14. Stellato A, Rajbhandari S, Turbati A, et al. VocBench: A Web Application for Collaborative Development of Multilingual Thesauri. In: Gandon F, Sabou M, Sack H, d'Amato C, Cudré-Mauroux P, Zimmermann A., eds. *The Semantic Web. Latest Advances and New Domains*Springer International Publishing; 2015; Cham: 38–53.

15. Volz J, Bizer C, Gaedke M, Kobilarov G. Silk – A Link Discovery Framework for the Web of Data. In: ; 2009.

16. Verborgh R, De Wilde M. *Using OpenRefine*. Packt Publishing . 2013.

17. Lohmann S, Link V, Marbach E, Negru S. WebVOWL: Web-based Visualization of Ontologies. In: Lambrix P, Hyvönen E, Blomqvist E, et al., eds. *Knowledge Engineering and Knowledge Management*Springer International Publishing; 2015; Cham: 154–158.

18. Ngo D, Bellahsene Z. Overview of YAM++—(not) Yet Another Matcher for ontology alignment task. *Journal of Web Semantics* 2016; 41: 30 - 49. doi: https://doi.org/10.1016/j.websem.2016.09.002

19. Khalifa S, Elshater Y, Sundaravarathan K, et al. The Six Pillars for Building Big Data Analytics Ecosystems. *ACM Comput. Surv.* 2016; 49(2): 33:1–33:36. doi: 10.1145/2963143

20. Coro G, Panichi G, Scarponi P, Pagano P. Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurrency and Computation: Practice and Experience* 2017; 29(18): e4219. doi: 10.1002/cpe.4219

21. Berthold MR, Cebron N, Dill F, et al. KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. *SIGKDD Explor. Newsl.* 2009; 11(1): 26–31. doi: 10.1145/1656274.1656280

22. Perez F, Granger BE. IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering* 2007; 9(3): 21-29. doi: 10.1109/MCSE.2007.53

23. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* 2010; 11(8): R86. doi: 10.1186/gb-2010-11-8-r86

24. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016; 3: 160018 EP. doi: 10.1038/sdata.2016.18

25. Príncipe P, Bardi A, Vieira A, Manghi P, Baglioni M, Retberg N. OpenAIRE Dashboard for Research Communities: Enabling Open Science publishing for Research Communities and Research Infrastructures. Poster presented at the Open Science Conference 2019, Berlin, Germany, 19-20 March 2019; 2019.

26. Penev L. From Open Access to Open Science from the viewpoint of a scholarly publisher. *Research Ideas and Outcomes* 2017; 3: e12265. doi: 10.3897/rio.3.e12265

27. Wit dA, Boogaard H, Fumagalli D, et al. 25 years of the WOFOST cropping systems model. *Agricultural Systems* 2019; 168: 154 - 167. doi: https://doi.org/10.1016/j.agsy.2018.06.018

28. Knapen MJR, Lokers RM, Candela L, Janssen S. AGINFRA PLUS: Running Crop Simulations on the D4Science Distributed e-Infrastructure. In: Athanasiadis IN, Frysinger SP, Schimak G, Knibbe WJ., eds. *Environmental Software Systems. Data Science in Action - 13th IFIP WG 5.11 International Symposium, ISESS 2020, Wageningen, The Netherlands, February 5-7, 2020, Proceedings*. 554 of *IFIP Advances in Information and Communication Technology*. Springer; 2020: 81–89

29. Haberbeck LU, Plaza-Rodríguez C, Desvignes V, et al. Harmonized terms, concepts and metadata for microbiological risk assessment models: The basis for knowledge integration and exchange. *Microbial Risk Analysis* 2018; 10: 3 - 12. Special issue on 10th International Conference on Predictive Modelling in Food: Interdisciplinary Approaches and Decision-Making Tools in Microbial Risk Analysisdoi: https://doi.org/10.1016/j.mran.2018.06.001

30. Alba Aparicio dM, Buschhardt T, Swaid A, et al. FSK-Lab – An open source food safety model integration tool. *Microbial Risk Analysis* 2018; 10: 13 - 19. Special issue on 10th International Conference on Predictive Modelling in Food: Interdisciplinary Approaches and Decision-Making Tools in Microbial Risk Analysisdoi: https://doi.org/10.1016/j.mran.2018.09.001

31. Neveu P, Tireau A, Hilgert N, et al. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Phytologist* 2019; 221(1): 588-601. doi: 10.1111/nph.15385

32. Tennant JP, Crane H, Crick T, et al. Ten Hot Topics around Scholarly Publishing. *Publications* 2019; 7(2). doi: 10.3390/publications7020034

33. Gabriel A, Capone R. Executable Paper Grand Challenge Workshop. *Procedia Computer Science* 2011; 4: 577 - 578. Proceedings of the International Conference on Computational Science, ICCS 2011doi: https://doi.org/10.1016/j.procs.2011.04.060

34. Bardi A, Manghi P. Enhanced publications: Data models and information systems. *LIBER Quarterly* 2014; 23(4): 240–273. doi: http://doi.org/10.18352/lq.8445

35. Mietchen D, Mounce R, Penev L. Publishing the research process. *Research Ideas and Outcomes* 2015; 1: e7547. doi: 10.3897/rio.1.e7547

36. Filter M, Candela L, Guillier L, et al. Open Science meets Food Modelling: Introducing the Food Modelling Journal (FMJ). *Food Modelling Journal* 2019; 1: e46561. doi: 10.3897/fmj.1.46561

37. Desvignes V, Buschhardt T, Guillier L, Sanaa M. Quantitative microbial risk assessment for Salmonella in eggs. *Food Modelling Journal* 2019; 1: e39643. doi: 10.3897/fmj.1.39643

38. Manouselis N (Ed.) . Digital Science Recommendations for Food & Agriculture. Discussion paper, AGINFRA PLUS Consortium; 2020

39. Ind N, Coates N. The meanings of co-creation. *European Business Review* 2013; 25(1): 86-95. doi: https://doi.org/10.1108/09555341311287754

40. Markus ML, Keil M. If We Build It, They Will Come: Designing Information Systems that People Want to Use. *Sloan Management Review* 1994; 35(4).

41. Nelson WA, Bueno KA, Huffstutler S. If You Build It, They Will Come. But How Will They Use It?. *Journal of Research on Computing in Education* 1999; 32(2): 270-286. doi: 10.1080/08886504.1999.10782278

42. Finholt TA, Birnholtz JP. *If We Build It, Will They Come? The Cultural Challenges of Cyberinfrastructure Development*: 89–101; Dordrecht: Springer Netherlands . 2006

□