Realising a Science Gateway for the Agri-food: the AGINFRA PLUS Experience

M. Assante*, A. Boizet†, L. Candela*, D. Castelli*, R. Cirillo*, G. Coro* E. Fernández§, M. Filter¶, L. Frosini*, G. Kakaletris‡, P. Katsivelis‖, M.J.R. Knapen**, L. Lelii*, R.M. Lokers**, F. Mangiacrapa*, P. Pagano*, G. Panichi*, L. Penev††, F. Sinibaldi*, P. Zervas‖
*ISTI - National Research Council of Italy, Pisa, Italy – Email: {name.surname}@isti.cnr.it
†French National Institute for Agricultural Research, Paris, France – Email: alice.boizet@inra.fr
§EGI Foundation, Amsterdam. Neederlands – Email: enol.fernandez@egi.eu
¶Federal Institute for Risk Assessment (BfR), Berlin, Germany – Email: matthias.filter@bfr.bund.de
‡University of Athens, Athens, Greece – Email: gkakas@di.uoa.gr
‖Agroknow, Athens, Greece – Email: {katsivelis.panagis, pzervas}@agroknow.com
**Wageningen University & Research, Wageningen, Neederlands – Email: {rob.knapen, rob.lokers}@wur.nl
††Pensoft Publishers, Sofia, Bulgaria – Email: penev@pensoft.net

***Abstract*—The enhancements in IT solutions and the open science movement are injecting changes in the practices dealing with data collection, collation, processing and 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" with datasets, processes, workflows. 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**

I. 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 Agriculture and Food Science research domain [1], [2]. The fragmentation of "resources" of interest 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 in a collaborative way. Such VREs are built by relying on an open and distributed platform (see Sec. II) providing a rich array of services supporting all the phases of an open science research lifecycle from data collection to data analytics and publication.

AGINFRA PLUS is exploiting the VREs approach for three prominent agri-food research communities, namely: (*i*) *agro-climatic and economic 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. II presents the major constituents of the AGINFRA PLUS platform. Sec. III discusses the exploitation scenarios developed by each community and the benefits resulting from the use of the platform. Finally, Sec. IV concludes the paper by reporting some future works.

II. The AGINFRA PLUS Platform

In order to support the AGINFRA PLUS communities, a comprehensive and feature rich platform has been developed and operated. An overall picture of such a platform aiming at offering its facilities by 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

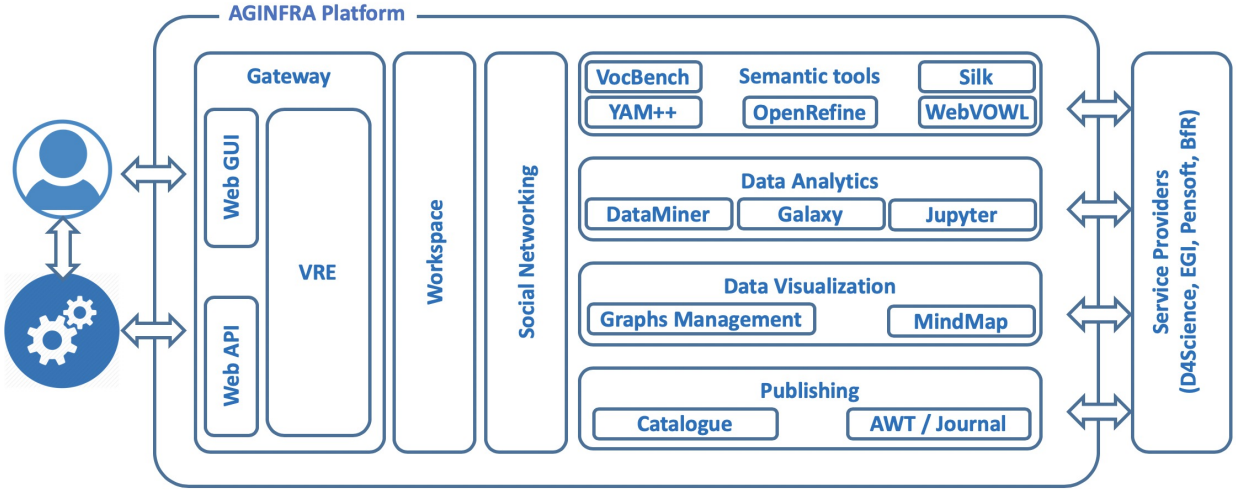

Fig. 1. AGINFRA PLUS Platform Architecture

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 *shared 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 *overall 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.

A. Semantic Data Management Solutions

The AGINFRA PLUS *data & semantics facilities* offer an array of 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 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.

B. Data Analytics Solutions

The AGINFRA PLUS *analytics facilities* offer 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

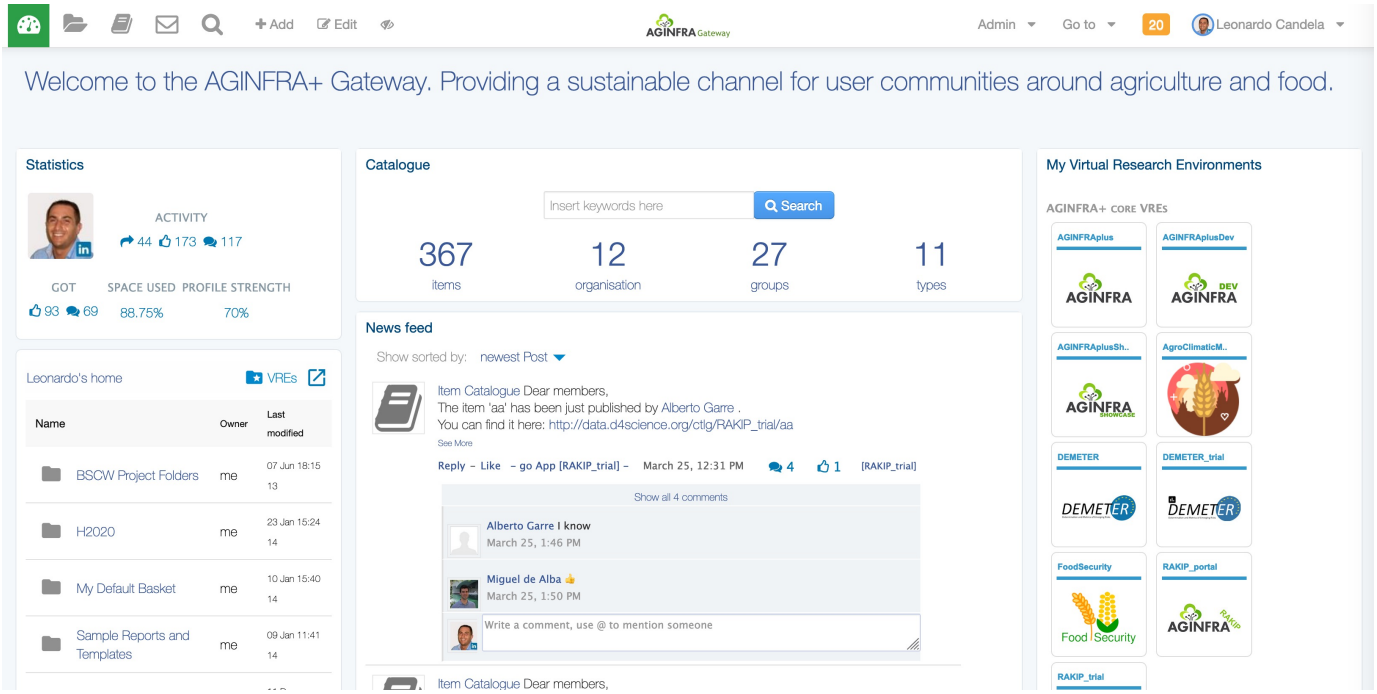

Fig. 2. AGINFRA PLUS Gateway: the Dashboard

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.

C. Data Visualization and Publishing Solutions

The AGINFRA PLUS *data visualization & publishing facilities* provide users with feature-rich and flexible solutions for developing representations (e.g. graphs) out of datasets and publishing "research objects" documenting a research activity and its results. The final goal is to provide the reader with an effective representation of a research activity and its results thus to enable its repeatability. The supported facilities include: (*a*) a *graphs management workbench* for creating several typologies of interactive graphs ranging from generic ones (e.g. Spline, Scatter, Bar, Line, Step, Pie, Doughnut, Polar) to very specific ones (e.g. graphs reporting the height of plants across time with values and images). The platform provide users with facilities to import a dataset of interest, to define how its content has to be used to produce the graph of interest, and to share the produced graphs; (*b*) a *mind map workbench* for managing this typology of diagrams; (*c*) a *network visualisation* for creating visualisations aiming at highlighting the connections among the entities of a connected graph; (*d*) a *catalogue-based publishing platform* to disseminate artefacts according to the FAIR principles [24]. The latter platform [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; (*e*) a *research community dashboard* realising a domain specific access point to search for content of interest. This is based on the OpenAIRE specific service [25] enabling to publish research products and interlink them with the OpenAIRE scholarly communication cloud. (*f*) a *scholarly publishing platform* integrated with Pensoft infrastructure [26] to enable the creation of innovative papers including datasets

and methods hosted by the AGINFRA PLUS platform. 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.

III. Exploitation scenarios

A. Agroclimatic Modeling

The objective is to set-up and evaluate an AGINFRA PLUS VRE for use by agro-climatic researchers to perform crop modelling related work. To guide the selection and development of tools to be included in the VRE, two typical research activities were selected: (*i*) performing crop model simulations at scale; and (*ii*) explorative modelling focussing on crop phenology studies. Based on data availability both have a strong focus on The Netherlands as study area (using AgroDataCube [8] as input), but the approaches can be extended to other regions once sufficient data is collected and a suitable crop model has been added to the VRE.

Initially the well-know WOFOST model [27] has been integrated so that it can be executed as tasks in the AGINFRA PLUS data analytics platform (see Sec. II-B). DataMiner makes algorithms available as Web Processing Service (WPS, a standard from the Open GeoSpatial Consortium, OGC). Using these facilities a 'worker' process has been implemented that can run large batches (1000 - 10000) of crop simulations on the distributed computing infrastructure behind the platform, and a 'scheduler' process that can divide the total workload over all available compute nodes (currently between 6 to 10), and collect all simulations results. The 'workload' for example might consist of running the crop simulations for all crop parcels of one or more years in The Netherlands, about 400,000 crop simulations per year, studying effects of input parameter variations, such as temperature sums or precipitation amounts, which multiplies the total crop simulations needing to be performed.

A second activity that is examined is the use of explorative modelling for the estimation of crop phenology characteristics, using available agronomic data, combined with crop development indicators (e.g. the NDVI vegetation index), derived from remote sensing data. This activity uses the AGINFRA PLUS analytics facilities such as Jupyter Notebooks and RStudio, to experiment with agronomic data analytics. The aim is to test such analytics, providing insight in critical crop development indicators, and to convert these into algorithms deployed as DataMiner processes on the VRE to run them at scale. Results can then be used to more accurately estimate regional crop yields, using long-term agronomic statistics and yield prediction systems.

Early stage evaluation results from piloting both agro-climatic modelling activities indicated that the VRE already is regarded as being well equipped for collaborative research. At that point (about one year ago) there were however reservations concerning the ability of the VRE to support full agro-climatic modelling workflows, due to some limitations regarding the integration of the different processing, analytics and visualisation components available in the VRE. Therefore recent efforts in the use case have been focussing on improving these integration capabilities and on providing better-connected prototypes for both activities, supporting the full research working process. For example by creating a dashboard that visualises crop parcels, the input data for crop model simulations (crop, soil and weather information), and on-the-fly calculated simulation results such as leaf area index and total biomass produced.

The Virtual Research Environment developed for supporting this scenario is available at https://aginfra.d4science.org/web/agroclimaticmodelling.

B. Food Safety Risk Assessment

In the domain of food safety modelling two exploitation 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 [28].

The DEMETER scenario is aiming at developing a working environment supporting the early identification of issues in the food (and feed) chain. This scenario largely build upon the workspace, the data analytics and the catalogue to demonstrate how KNIME-based data mining workflows can be efficiently shared and applied from within the VRE.

The RAKIP scenario aims at providing risk assessors and risk modellers with an environment supporting their efforts to share their knowledge (data, mathematical model, simulation results) 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 upon the FSK-Lab [29], i.e. a community standard to homogenise the representation and packaging of all relevant data, metadata and model scripts in a machine-readable format. This is an extension of the KNIME platform, one of the platforms underlying the data analytics.

The AGINFRA PLUS platform support these scenarios by providing: (*a*) facilities for developing the ontology underlying the FSK-Lab solution for models representation (VocBench, see Sec. II-A); (*b*) two specific processes integrated into the data analytics platform to respectively support the publishing of a model into the catalogue and the execution of any model; (*c*) a catalogue where the models are published according to the community ontology and endowed each with three actionable resources enabling users to respectively download the model, perform a model simulation by using the default parameters, perform a model simulation by tuning the parameters; (*d*) a mind map development and dissemination solution facilitating the communication among the members; (*e*) a journal-based approach for publishing the models. The Food Modeling Journal[1] has been designed and launched to support the needs emerging in this community. It promotes

[1] https://fmj.pensoft.net/

the publishing of Models, Data analytics, Applied study, Data paper, and Software description. Thanks to the integration of the publishing platform into the VREs (see Sec. II-C) it is straightforward to produce papers linking the available artifacts, e.g. the models in their actionable form.

The Virtual Research Environments developed for supporting these scenarios are available at https://aginfra.d4science.org/web/demeter and https://aginfra.d4science.org/web/rakip_portal.

C. Food Security

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 access to phenomics platforms data from the information system OpenSILEX-PHIS [30]; 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[2] 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 at https://aginfra.d4science.org/web/foodsecurity.

[2]https://brapi.org

IV. 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 is bringing 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.

Overall, the AGINFRA PLUS platform is currently serving hundreds of users (more than 340 in Feb. 2019) by 13 active VREs. In the coming months these figures are going to improve because the project will enter into the community validation and uptake phase. In the period Mar. 2018 - Feb. 2019 the users served by this platform and its VREs performed: a total of 24,439 working sessions, with an average of circa 2,036 sessions per month; a total of 1,959 social interactions, with an average of circa 163 interactions per month; a total of 1,842 analytics tasks, with an average of circa 153 tasks per month; a total of 387 items have been published into the catalogue including models, research objects, methods, services, terms, and datasets.

Future developments includes the development of a catalogue supporting semantic queries, the development of tools easing the discovery and access to geospatial datasets, the development of recommender systems, the development of tools supporting the identification of suitable licences for the produced artifacts.

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] J. W. Jones, J. M. Antle, B. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems

science," *Agricultural Systems*, vol. 155, pp. 269 – 288, 2017. [Online]. Available: 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] L. Candela, D. Castelli, and P. Pagano, "Virtual research environments: an overview and a research agenda," *Data Science Journal*, vol. 12, pp. GRDI75–GRDI81, 2013.

[4] M. W. Maier, "Architecting principles for systems-of-systems," *INCOSE International Symposium*, vol. 6, no. 1, pp. 565–573, 1996. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/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] H. Janssen, S. Janssen, M. Knapen, W. Meijninger, Y. v. Randen, I. l. Riviere, and G. Roerink, "AgroDataCube: A big open data collection for agri-food applications," agrodatacube.wur.nl, 2018.

[9] C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques, and J. Keizer, "The AGROVOC linked dataset," *Semantic Web*, vol. 4, no. 3, pp. 341–348, 2013.

[10] German Federal Institute for Risk Assessment, "Foodrisk-labs," https://foodrisklabs.bfr.bund.de/foodrisk-labs/.

[11] M. Assante, L. Candela, D. Castelli, R. Cirilllo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, V. Marioli, P. Pagano, G. Panichi, C. Perciante, and F. Sinibaldi, "The gcube system: Delivering virtual research environments as-a-service," *Future Generation Computer Systems*, vol. 95, no. n.a., pp. 445–453, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X17328364

[12] AGINFRA Consortium, "The AGINFRA gateway," https://aginfra.d4science.org/.

[13] M. Assante, L. Candela, D. Castelli, G. Coro, F. Mangiacrapa, P. Pagano, and P. Costantino, "Enacting open science by gcube," in *Proceedings of the 9th International Workshop on Science Gateways*, 2018.

[14] A. Stellato, S. Rajbhandari, A. Turbati, M. Fiorelli, C. Caracciolo, T. Lorenzetti, J. Keizer, and M. T. Pazienza, "Vocbench: A web application for collaborative development of multilingual thesauri," in *The Semantic Web. Latest Advances and New Domains*, F. Gandon, M. Sabou, H. Sack, C. d'Amato, P. Cudré-Mauroux, and A. Zimmermann, Eds. Cham: Springer International Publishing, 2015, pp. 38–53.

[15] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov, "Silk – a link discovery framework for the web of data," in *Proceedings of the Linked Data on the Web Workshop (LDOW2009), Madrid, Spain, April 20, 2009, CEUR Workshop Proceedings*, 2009. [Online]. Available: http://ceur-ws.org/Vol-538/ldow2009_paper13.pdf

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

[17] S. Lohmann, V. Link, E. Marbach, and S. Negru, "WebVOWL: Web-based visualization of ontologies," in *Knowledge Engineering and Knowledge Management*, P. Lambrix, E. Hyvönen, E. Blomqvist, V. Presutti, G. Qi, U. Sattler, Y. Ding, and C. Ghidini, Eds. Cham: Springer International Publishing, 2015, pp. 154–158.

[18] D. Ngo and Z. Bellahsene, "Overview of yam++—(not) yet another matcher for ontology alignment task," *Journal of Web Semantics*, vol. 41, pp. 30 – 49, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570826816300464

[19] S. Khalifa, Y. Elshater, K. Sundaravarathan, A. Bhat, P. Martin, F. Imam, D. Rope, M. Mcroberts, and C. Statchuk, "The six pillars for building big data analytics ecosystems," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 33:1–33:36, Aug. 2016. [Online]. Available: http://doi.acm.org/10.1145/2963143

[20] G. Coro, G. Panichi, P. Scarponi, and P. Pagano, "Cloud computing in a distributed e-infrastructure using the web processing service standard," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 18, p. e4219, 2017.

[21] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, and B. Wiswedel, "Knime - the konstanz information miner: Version 2.0 and beyond," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 26–31, Nov. 2009. [Online]. Available: http://doi.acm.org/10.1145/1656274.1656280

[22] F. Perez and B. E. Granger, "IPython: A system for interactive scientific computing," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 21–29, 2007.

[23] J. Goecks, A. Nekrutenko, and J. Taylor, "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences," *Genome Biology*, vol. 11, no. 8, p. R86, Aug 2010.

[24] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, p. 160018 EP, 2016. [Online]. Available: http://dx.doi.org/10.1038/sdata.2016.18

[25] P. Príncipe, A. Bardi, A. Vieira, P. Manghi, M. Baglioni, and N. Retberg, "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] L. Penev, "From open access to open science from the viewpoint of a scholarly publisher," *Research Ideas and Outcomes*, vol. 3, p. e12265, 2017. [Online]. Available: https://doi.org/10.3897/rio.3.e12265

[27] A. de Wit, H. Boogaard, D. Fumagalli, S. Janssen, R. Knapen, D. van Kraalingen, I. Supit, R. van der Wijngaart, and K. van Diepen, "25 years of the wofost cropping systems model," *Agricultural Systems*, vol. 168, pp. 154 – 167, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0308521X17310107

[28] L. U. Haberbeck, C. Plaza-Rodríguez, V. Desvignes, P. Dalgaard, M. Sanaa, L. Guillier, M. Nauta, and M. Filter, "Harmonized terms, concepts and metadata for microbiological risk assessment models: The basis for knowledge integration and exchange," *Microbial Risk Analysis*, vol. 10, pp. 3 – 12, 2018, special issue on 10th International Conference on Predictive Modelling in Food: Interdisciplinary Approaches and Decision-Making Tools in Microbial Risk Analysis. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352352218300100

[29] M. de Alba Aparicio, T. Buschhardt, A. Swaid, L. Valentin, O. Mesa-Varona, T. Günther, C. Plaza-Rodriguez, and M. Filter, "Fsk-lab – an open source food safety model integration tool," *Microbial Risk Analysis*, vol. 10, pp. 13 – 19, 2018, special issue on 10th International Conference on Predictive Modelling in Food: Interdisciplinary Approaches and Decision-Making Tools in Microbial Risk Analysis. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352352218300136

[30] P. Neveu, A. Tireau, N. Hilgert, V. Nègre, J. Mineau-Cesari, N. Brichet, R. Chapuis, I. Sanchez, C. Pommier, B. Charnomordic, F. Tardieu, and L. Cabrera-Bosquet, "Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven phenotyping hybrid information system," *New Phytologist*, vol. 221, no. 1, pp. 588–601, 2019. [Online]. Available: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.15385