2018 IEEE 14th International Conference on e-Science

Serving Scientists in Agri-food area by Virtual Research Environments

A. Ballis*, A. Boizet†, L. Candela‡, D. Castelli‡, E. Fernández§, M. Filter¶, T. Günther¶, G. Kakaletris*, P. Karampiperis‖, D. Katris*, M.J.R. Knapen**, R.M. Lokers**, L. Penev††, G. Sipos§, P. Zervas‖
*University of Athens, Athens, Greece – Email: {aballis, gkakas, dkatris}@di.uoa.gr
†French National Institute for Agricultural Research, Paris, France – Email: alice.boizet@inra.fr
‡ISTI - National Research Council of Italy, Pisa, Italy – Email: {candela, castelli}@isti.cnr.it
§EGI Foundation, Amsterdam. Neederlands – Email: {enol.fernandez, gergely.sipos}@egi.eu
¶Federal Institute for Risk Assessment (BfR), Berlin, Germany – Email: {matthias.filter, taras.guenther}@bfr.bund.de
‖Agroknow, Athens, Greece – Email: {pythk, pzervas}@agroknow.com
**Wageningen University & Research, Wageningen, Neederlands – Email: {rob.knapen, rob.lokers}@wur.nl
††Pensoft Publishers, Sofia, Bulgaria – Email: penev@pensoft.net

***Abstract*—Agri-food research calls for changes in the practices dealing with data collection, collation, processing and analytics, and publishing thus to fully benefit from and contribute to the Open Science movement. One of the major issues faced by the agri-food researchers is the fragmentation of the "assets" that can be exploited when performing research tasks, e.g. data of interest are heterogeneous and scattered across several repositories, the tools exploited by modellers are diverse and often rely on local computing environments, the publishing practices are various and rarely aim at making available the "whole story" with datasets, processes, workflows. This paper presents the AGINFRA+ endeavour to overcome these limitations by providing researchers in three designated communities with Virtual Research Environments facilitating the access to and use of the "assets" of interest and promote collaboration.**

***Index Terms*—Virtual Research Environment, e-Infrastructure, Agro-climatic modelling, Food Safety, Food Security**

I. Introduction

The developments in information and communication technology including big data availability and management, web and cloud technologies, as well as open science related practices are not yet fully embraced by Agricultural and Food Science [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, computing capacity availability varies a lot across teams.

The AGINFRA+ project has been set up to develop an innovative approach aiming at overcoming the limitations stemming from the above settings by leveraging on existing e-Infrastructures and services. In particular, AGINFRA+ promotes the exploitation of *Virtual Research Environments* (VREs) [3] to provide designated communities with 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 e-Infrastructure developed with the system of systems [4] paradigm. Such an e-Infrastructure aggregates resources from "generalist" service providers (e.g. D4Science [5], EGI [6]) as well as from community-specific ones (e.g. AgroDataCube [7], AGROVOC [8], RAKIP model repository [9]) to build a unifying space where the aggregated resources can be exploited via VREs [10].

AGINFRA+ 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 [11] to support phenomics researchers to select plant species and varieties which are the most adapted to specific environments and to global changes.

The AGINFRA+ VREs [12] are offering a large array of facilities as-a-Service by an integrated environment including: (*i*) a *data & semantics facilities* enacting the creation and organisation of semantically rich metadata (by VocBench [13], Silk [14], WebVOWL [15]); (*ii*) a *shared workspace* to store, organise and share any version of a research artefact [16]; (*iii*) a *social networking area* to have discussions on any topic and be informed on happenings, e.g. the availability of a research outcome [16]; (*iv*) a *data analytics platform* to execute analytics tasks either provided by the user or provided by others [17]. It is endowed with importing and sharing facilities for analytics methods implemented in forms including R, Java, Phyton, and KNIME [18] (largely used by the food safety community). The platform enacts tasks execution by a distributed and hybrid computing infrastructure including EGI resources; (*v*) a *catalogue-based publishing platform* to make the existence of a certain artefact "public" according to FAIR principles [16]; (*vi*) a Jupyter *notebook based environment* for documenting and recording analytics processes [19]; (*vii*) a *scholarly publishing platform* integrated with Pensoft infrastructure to enact the creation of innovative papers including datasets and methods hosted by AGINFRA; (*viii*) a *platform for data visualisation* enabling to create smart

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00124

graphs and share them. An overall picture of the system architecture is given in Fig. 1.

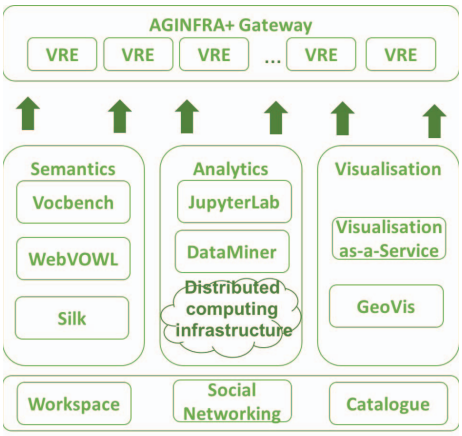

Fig. 1. AGINFRA+ system architecture

Fig. 2 showcases the home page of the AGINFRA+ gateway where users can have access to the facilities and working environment developed.

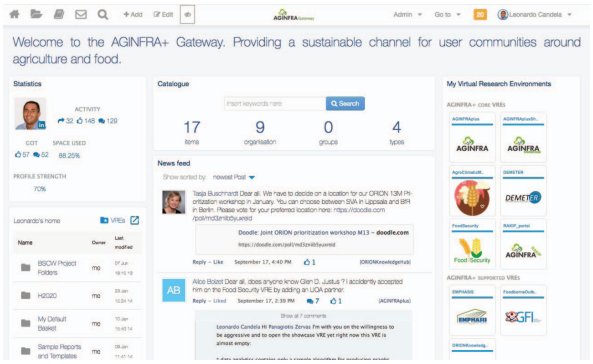

Fig. 2. AGINFRA+ gateway home

Fig. 3 showcases the home page a virtual research environment.

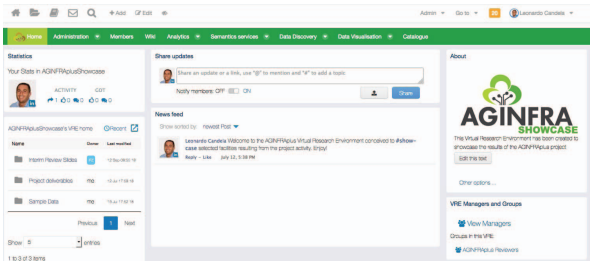

Fig. 3. AGINFRA Showcase VRE Home

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] Wageningen Environmental Research, "AgroDataCube: A big open data collection for agri-food applications," agrodatacube.wur.nl.

[8] 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.

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

[10] M. Assante, L. Candela, D. Castelli, G. Coro, L. Lelii, and P. Pagano, "Virtual research environments as-a-service by gcube," *PeerJ Preprints*, 2016.

[11] F. Fiorani and U. Schurr, "Future scenarios for plant phenotyping," *Annual Review of Plant Biology*, vol. 64, no. 1, pp. 267–291, 2013, pMID: 23451789. [Online]. Available: https://doi.org/10.1146/annurev-arplant-050312-120137

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

[13] 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.

[14] 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

[15] 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.

[16] 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.

[17] 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.

[18] 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

[19] 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.