AGINFRA

D3.1- OPEN SCIENCE DATA ANALYTICS TECHNOLOGIES

DELIVERABLE NUMBER	D3.1
DELIVERABLE TITLE	Open Science Data Analytics Technologies
RESPONSIBLE AUTHORS	Leonardo Candela (CNR), Roberto Cirillo (CNR), Gianpaolo Coro (CNR), Lucio Lelii (CNR), Pasquale Pagano (CNR), Giancarlo Panichi (CNR), Paolo Scarponi (CNR), Fabio Sinibaldi (CNR)

GRANT AGREEMENT N.	731001
PROJECT ACRONYM	AGINFRA PLUS
PROJECT FULL NAME	Accelerating user-driven e-infrastructure innovation in Food & Agriculture
STARTING DATE (DUR.)	01/01/2017 (36 months)
ENDING DATE	31/12/2019
PROJECT WEBSITE	plus.aginfra.eu
COORDINATOR	Nikos Manouselis
ADDRESS	110 Pentelis Str., Marousi GR15126, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Mrs. Georgia Tzenou
WORKPACKAGE N. | TITLE	WP3 | Data Analytics & Processing Layer
WORKPACKAGE LEADER	Consiglio Nazionale Delle Richerche(CNR)
DELIVERABLE N. | TITLE	D3.1 | Open Science Data Analytics Technologies
RESPONSIBLE AUTHOR	Leonardo Candela
REPLY TO	leonardo.candela@isti.cnr.it
DOCUMENT URL	http://www.plus.aginfra.eu/sites/plus_deliverables/D3.1.pdf
DATE OF DELIVERY (CONTRACTUAL)	31December 2017
DATE OF DELIVERY (SUBMITTED)	22December 2017
VERSION | STATUS	1.0 | Final
NATURE	Demonstrator
DISSEMINATION LEVEL	Public
AUTHORS (PARTNER)	Leonardo Candela (CNR), Roberto Cirillo (CNR), Gianpaolo Coro (CNR), Lucio Lelii (CNR), Pasquale Pagano (CNR), Giancarlo Panichi (CNR), Paolo Scarponi (CNR), Fabio Sinibaldi (CNR)
CONTRIBUTORS	-
REVIEWERS	Antonis Koukourikos (Agroknow)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.5	Pre-final version (wiki page) sent to internal review	8-12-2017	Leonardo Candela (CNR)
0.9	Reviewer commentsincorporated (wikipage)	18-12-2017	Leonardo Candela (CNR)
1.0	Final Version	21-12-2017	Leonardo Candela (CNR)

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

EXECUTIVE SUMMARY

Deliverable D3.1 "Open Science Data Analytics Technologies" is a deliverable of type **Demonstrator** meaning that it manifests in artefacts (software releases) other than reports. In particular, the deliverable is about the software realising the Data Analytics & Processing Layer of the AGINFRA+.

This software is part of a large software system named gCube (www.gcube-system.org). The gCube system offers a large array of services supporting the entire lifecycle underlying a research activity (data management and collation, analytics, collaboration, sharing) and the possibility to combine these services in Virtual Research Environments[1].

In the context of AGINFRA PLUS the following gCube components have been primarily exploited, consolidated and enhanced to serve the analytics needs arising in the context of the project use cases.

DataMiner, i.e. a service enacting its users to perform data analytics tasks by relying on an array of analytics methods and a distributed and heterogeneous computing infrastructure. This service is available by a web-based GUI as well as via a web-based API based on the OGC WPS standard.

SAI (Statistical Algorithm Importer), i.e. a service enacting its users to make available their own analytics methods via the DataMiner service.

In addition to that, the entire analytics solution made available for AGINFRA PLUS cases counts on (i) a *shared workspace* realising a cloud-based file manager for managing content of interest and sharing this content with co-workers, (ii) a *social networking* area enabling users to post messages and have discussions, (iii) a flexible catalogue enabling to publish and discover items of interest including "research objects" resulting from an analytics task.

This technology is deployed in its latest version in every Virtual Research Environment supporting AGINFRA PLUS cases[2].

The major enhancements to the technology pertaining to AGINFRA PLUS have been included in three gCube major releases[3] 4.7 (October 2017), 4.8 (November 2017), and 4.9 (under production).In particular, with these releases a new"black-box" oriented approach (https://wiki.gcube-system.org/gcube/Statistical_Algorithms_Importer:_Java_Project#Black_Box_Integration)has been envisaged and implemented to enact analytics method owners and developers to easily integrate theirsolutions into the DataMinerservice. Among the supported black-box typologies there is that for KNIME workflows, i.e. analytics methods implemented by a KNIME workflow. KNIME is among the key technologies supporting the Food Safety Risk Assessment cases. In order to enact the execution of KNIME-based black-boxes, the distributed computing part of the data analytics platform has been extended to integrate the KNIME execution engine. Other cases are counting on the same mechanism to integrate entire applications (WOFOST[4]) as well as Python-based methods.

The deliverable has been created on the public space of the AGINFRA+ Wiki and is accessible through the following link:

https://support.d4science.org/projects/aginfraplus_wiki/wiki/D31_-_Open_Science_Data_Analytics_Technologies

---

[1]Candela, L., Castelli, D. & Pagano, P., (2013). Virtual Research Environments: An Overview and a Research Agenda. Data Science Journal. 12, pp.GRDI75–GRDI81. DOI: http://doi.org/10.2481/dsj.GRDI-013

[2]https://aginfra.d4science.org

[3]https://www.gcube-system.org/software-releases

[4]https://www.wur.nl/en/Expertise-Services/Research-Institutes/Environmental-Research/Facilities-Products/Software-and-models/WOFOST.htm