D8.4 – AGINFRA PLUS Sustainability Roadmap

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

DELIVERABLE NUMBER	D8.4
DELIVERABLE TITLE	AGINFRA PLUS Sustainability Roadmap
RESPONSIBLE AUTHOR	Nikos Manouselis & Irene Kouriantaki (Agroknow)

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	Mr. Christophe Doin
WORKPACKAGE N. | TITLE	WP8 | Dissemination & Exploitation
WORKPACKAGE LEADER	Agroknow
DELIVERABLE N. | TITLE	D8.4 | AGINFRA PLUS Sustainability Roadmap
RESPONSIBLE AUTHOR	Nikos Manouselis (Agroknow)
REPLY TO	nikosm@agroknow.com
DOCUMENT URL	http://www.plus.aginfra.eu/
DATE OF DELIVERY (CONTRACTUAL)	31 December 2019 (M36)
DATE OF DELIVERY (SUBMITTED)	5 February 2019 (M26), 31 January 2020 (M37, Updated version)
VERSION | STATUS	V2.0 | Final
NATURE	Report (R)
DISSEMINATION LEVEL	Public (PU)
AUTHORS (PARTNERS)	Nikos Manouselis & Irene Kouriantaki (Agroknow)
CONTRIBUTORS	Leonardo Candela (CNR), Enol Fernández (EGI), Teodor Georgiev (Pensoft), Lyubomir Penev (Pensoft), Panagis Katsivelis (Agroknow), Thodoris Kontogiannis (Agroknow), Eva Bozou (Agroknow), Nikos Manouselis (Agroknow)
REVIEWER	Gergely Sipos (EGI)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Table of Contents	07-12-2018	Panagiotis Zervas (Agroknow)
0.2	Chapters 1	10-12-2018	Panagiotis Zervas (Agroknow), Rob Lokers (ALTERRA), Gergely Sipos (EGI), Matthias Filter (BfR)
0.3	Chapters 2	14-12-2018	Panagiotis Zervas (Agroknow), Gergely Sipos (EGI), Matthias Filter (BfR), Rob Lokers (ALTERRA), Lyubomir Penev (PENSOFT), Leonardo Candela (CNR)
0.4	Chapters 3	17-12-2018	Panagiotis Zervas (Agroknow), Rob Lokers (ALTERRA), Leonardo Candela (CNR), Gergely Sipos (EGI), Teodor Georgiev (PENSOFT)
0.5	Chapters 4	08-01-2019	Panagiotis Zervas (Agroknow), Rob Lokers (ALTERRA), Matthias Filter (BfR), Lyubomir Penev (PENSOFT), Leonardo Candela (CNR), Teodor Georgiev (PENSOFT)
0.7	Chapters 5	15-01-2019	Panagiotis Zervas (Agroknow), Rob Lokers (ALTERRA), Matthias Filter (BfR), Gergely Sipos (EGI), Alice Boizet (INRA), Enol Fernández (EGI)
1.0	Chapters update, Formatting update	05-02-2019	Panagiotis Zervas (Agroknow)
1.5	Feedback from partners	17-12-2019	Leonardo Candela (CNR), Enol Fernández (EGI), Teodor Georgiev (PENSOFT), Lyubomir Penev (PENSOFT)
2.0	Updating version	31-01-2020	Nikos Manouselis (Agroknow) Panagis Katsivelis (Agroknow), Thodoris Kontogiannis (Agroknow), Irene Kouriantaki (Agroknow), Eva Bozou (Agroknow)

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

Table of Contents

INTRODUCTION 12
BACKGROUND: Addressing Review Comments 14
1. AGINFRA PLUS TECHNOLOGY OVERVIEW 15
1.1 AGINFRA PLUS Technology Outcomes 15
1.1.1 AGINFRA data infrastructure & VRE 15
1.2 AGINFRA PLUS software tools & online environments 21
1.2.1 Ontological Engineering tools 21
1.2.2 Data analytics tools and Open Science helpers 23
1.2.3 Visualisation tools 27
2. THE NEED 29
2.1 The Business Opportunity 29
2.1.1 Business needs 29
2.1.2 Vision 29
2.1.3 Brand 29
2.1.4 The Solution: AGINFRA offerings 30
2.2 Market Introduction 31
2.2.1 Market Overview 31
2.2.2 Focusing on three market segments: 32
3. OUR POSITIONING 33
3.1 Market Analysis & Positioning 33
3.1.1 Market Description of Public Sector Funded Scientists 33
3.1.2 The key characteristics of this market 33
3.1.3 The key attributes of the publicly funded scientist persona 34
3.1.4 Market maturity 35
3.2 Market Description of Food Industry Scientists 35
3.2.1 The key characteristics of this market 35
3.2.1 The key attributes of this persona 35
3.2.3 Market readiness 35
3.3 Market description of the agrochem industry scientists 36
3.3.1 The key attributes of this persona 36

3.3.2 Market readiness 37
3.4 Market Description of agtech & foodtech SMEs powered by data science 37
3.4.1 The market 37
3.4.2 The key characteristics of this market 38
3.4.3 The key attributes of the data science startups persona 38
3.5 Relevant AGINFRA offerings 38
3.6 Competition 40
3.6.1 D4Science-powered VREs 40
D4Science-powered VREs 41
3.6.2 EGI-powered Jupyter Notebooks 42
EGI-powered Jupyter Notebooks 42
3.6.3 EGI-powered Galaxy Workflows 43
3.6.4 Data Harvesting Workflows & Data Linking Services 44
AGINFRA Data Harvesting & Linking Services 45
3.6.5 ARPHA-powered Data Journals 46
ARPHA Platform 47
3.7 Marketing & Promotion Strategy 48
3.7.1 Branding of AGINFRA PLUS bundle of offerings 48
3.7.2 How to create user awareness & new marketing leads 48
4. OUR DELIVERY 49
4.1 Management & Operations 49
4.1.1. Operations of D4Science-powered VREs 49
4.1.2 The Team 49
4.1.3 Structure and Management Style 49
4.1.4 Staff Recruiting 50
4.1.5 Equipment and Facilities 50
4.1.6 Key Partnerships 50
4.1.7 IPR Management 50
4.1.8 Roadmap after AGINFRA PLUS 50
4.1.9 Risk Assessment 51
4.2 Operations of EGI-powered Jupyter Notebooks and Galaxy Workflows 51
4.2.1 The Team 52

4.2.2 Structure and Management Style 52
4.2.3 Staff Recruiting 52
4.2.4 Equipment and Facilities 52
4.2.5 Key Partnerships 53
4.2.6 IPR Management 53
4.2.7 Roadmap after AGINFRA PLUS 53
4.2.8. Risk Assessment 54
4.3. Operations of Data Harvesting & Data Linking Services 54
4.3.1 The Team 54
4.3.2 Structure and Management Style 55
4.3.3 Staff Recruiting 55
4.3.4 Equipment and Facilities 55
4.3.5 Key Partnerships 55
4.3.6 IPR Management 56
4.3.7 Roadmap after AGINFRA PLUS 56
4.3.8 Risk Assessment 56
4.4 Operations of ARPHA-powered Data Journals 57
4.4.1 The Team 59
4.4.2 Structure and Management Style 59
4.4.3 Staff Recruiting 61
4.4.4 Equipment and Facilities 61
4.4.5 Key Partnerships 62
4.4.6 IPR Management 62
4.4.7 Roadmap after AGINFRA PLUS 63
4.4.8 Risk Assessment 63
5. OUR FINANCIALS 65
5.1 Financial Projections of D4Science-powered VREs 65
5.1.1 Business Model(s) 65
5.1.2 Financial Forecasts 66
5.1.3 Funding Needs and Opportunities 67
5.2 Financial Projections of EGI-powered Jupyter Notebooks and Galaxy Workflows 67
5.2.1 Business Model(s) 67

5.2.2 Financial Forecasts 68
5.2.3 Funding Needs and Opportunities 69
5.3 Financial Projections of Agroknow's Data Consulting Unit 69
5.3.1 Business Model(s) 69
5.3.2 Financial Forecasts 70
5.3.3 Funding Needs and Opportunities 72
5.4 Financial Projections of ARPHA-powered Data Journals 72
5.4.1 Business Model(s) 72
5.4.2 Financial Forecasts 72
5.4.3 Funding Needs and Opportunities 74
ANNEX A: Initial exploitation table of technology outcomes, IPR & exploitation paths (as per DoA). 75
ANNEX B: Questionnaires used to collect partner input 81
ANNEX C: Excel spreadsheet used to estimate financial projection data 83

LIST OF TABLES

Table 1: Mid-term review recommendation and consortium actions 14
Table 2:The top three stakeholders in terms of total EU funding contribution and the number of agri-food R&D projects 33
Table 3: The percentage of the R&D investment that was put in place by each of the key stakeholders. 36
Table 4: An overview of the AGINFRA PLUS offerings 39
Table 5: An outline of the features that are important for D4Science-powered VREs audience 41
Table 6: An outline of the features that are important for EGI-powered Jupyter Notebooks audience 42
Table 7: An outline of the features that are important for EGI-powered Galaxy Workflows audience 43
Table 8: An outline of the features that are important for agri-food researchers audience 45
Table 9: An outline of the features that are important for ARPHA-powered Data Journals audience. 47
Table 10: The possible risks for operating D4Science-powered VREs 51
Table 11: The possible risks for operating EGI-powered Jupyter Notebooks and Galaxy Workflows 54
Table 12: The main risks in successfully spinning off the service line of Agroknow as a new business 56
Table 13: The possible risks and their mitigation measures for Roadmap after AGINFRA PLUS 63
Table 14: Pricing model for annual subscriptions 65
Table 15: Projection of the revenue of CNR from AGINFRA-related users 66
Table 16: Projection of the annual costs foreseen for offering VREs to the targeted audiences. 66
Table 17: The annual profitability forecast for the VRE services 67
Table 18: Prices for typical requests of the service 68
Table 19: Projection of the sales revenue of the EGI Jupyter Notebooks and Galaxy 68
Table 20: Projection of the annual costs foreseen in terms of personnel and other expenses 68
Table 21: The sales and cost projections to calculate the annual profitability forecast for the Galaxy and Jupyter services 69
Table 22: Pricing model for the private consulting contracts 70
Table 23: Projection of the sales revenue of the data consulting spin of and an estimation of future sales until 2023. 70
Table 24: Projection of the annual costs foreseen in terms of personnel and other expenses for offering and evolving the data consulting services to the targeted audiences. 71
Table 25: The sales and cost projections to calculate the annual profitability forecast for the data consulting business 71
Table 26: projection of the sales revenue according to the number of articles that we expect to process and publish in our Data Journals 72
Table 27: Projection of the annual costs foreseen in terms of personnel and other expenses for supporting and maintaining the Data Journals 73

Table 28: The sales and cost projections to calculate the annual profitability forecast for the Data Journals revenue line of Pensoft 73

LIST OF FIGURES

Figure 1: The Global AgTech Landscape (source: Seana Day, 2019 Agtech Landscape) 37
Figure 2: European and Global annual AgTech Investments (source: AgFunder, 2018) 38
Figure 3: Management structure of Pensoft..... 60
Figure 4: Workflow for creation and dissemination of press releases...... 61
Figure 5: General Costs excel sheet 84
Figure 6: Pesonnel Costs excel sheet..... 84
Figure 7: Profitability excel sheet 85
Figure 8: Sales - Subscription excel sheet 85
Figure 9: Sales - Funding 86

INTRODUCTION

In the AGINFRA PLUS Sustainability Roadmap, we describe the way in which the technology partners of the consortium are envisioning how the key software infrastructures, services and tools of the project are going to be sustained after the project ends. This deliverable first provides an overview of the key software outcomes and highlights the ones selected for further exploitation. Then, it focuses on the targeted stakeholders and market segments, the business models chosen per targeted segment, as well as the operational and financial resources required.

After the project's Mid-Term Review meeting, an early version of this deliverable was prepared and submitted according to the reviewers' request, to reflect our thinking and approach in terms of sustainability at that point of time. In this new version of the deliverable, we significantly revise, extend and complete our original thinking. It has become clearer to us that in an era of radical digital transformations in the food and agriculture sector, the way in which project outcomes may be positioned and offered to potential users has changed. This creates numerous opportunities for the partners that are interested to continue this investment after the end of the project.

The five (5) key technology offerings that we have decided to continue investing into are the following:

1. **D4Science-powered VREs (CNR)**: existing or new Virtual Research Environments (VREs) that specific communities or organisations may continue using or creating, over the D4Science e-infrastructure.
2. **EGI-powered Jupyter Notebooks (EGI)**: existing or new scientific applications that will be using Jupyter Notebooks for interactive data science and scientific computing, over the EGI e-infrastructure.
3. **EGI-powered Galaxy Workflows (EGI)**: existing or new scientific workflows that will be executed over Galaxy, over the EGI e-infrastructure.
4. **Data Harvesting & Data Linking Services (Agroknow)**: existing or new software applications that will be using the semantic indexing and data lookup services of Agroknow (or request a white-label set up of their own similar service).
5. **ARPHA-powered Data Journals (Pensoft)**: existing or new scientific data journals that will be powered by ARPHA.

In terms of large-scale computing facilities and cloud storage resources, the above technology offerings have been powered by the two European e-infrastructures that have been serving as the backbone of the project:

- **D4Science e-infrastructure (CNR & UoA):** the Data Infrastructure that connects more than 7.000 scientists in over 50 countries, hosts more than 100 VREs for a variety of communities and that promises 99,8% service availability.
- **EGI e-infrastructure (EGI):** the federation of cloud providers and data centres, spread across Europe and worldwide, that delivers advanced computing services to scientists, projects and research infrastructures.

The deliverable has modified and extended its initial scope of describing an "AGINFRA Association", as the consortium has decided not to move forward with such a legal entity. Not all partners were equally interested in a joint commitment about the way to continue to support and evolve the various project outcomes. It was therefore decided that a select number of partners responsible for the above key AGINFRA PLUS technology offerings (namely, Agroknow, CNR, Pensoft, and EGI) develop jointly a plan about the future of these offerings. To this end, the original title of this deliverable "AGINFRA Association Blueprint" has been revised to "AGINFRA PLUS Sustainability Roadmap".

This Roadmap includes the following parts:

- An overview of the main technology outcomes of AGINFRA PLUS, also comparing their IP, TRL and licensing to the ones originally foreseen in the DoA.
- An introduction of the need that we see in the market & technology outcomes that we have decided to offer: what is the situation in the current market, with the food and agriculture sector being under a huge transformation; who are the potential users/customers that we may address; which are the specific market segments that we believe have potential for our partners and offerings.
- An overview of the market segments that we target and the offerings that are relevant: what are the specific audiences and their needs; which are the AGINFRA PLUS offerings that we see as relevant for these audiences; what is the competitive advantage that we have against the competition; what is the business model that we have decided to adopt in each case.
- A description of the strategic, operational and financial requirements to continue supporting each offering: which are the labour and computing resources needed; what are the corresponding costs; how do the financial forecasts look for each offering.

BACKGROUND: Addressing Review Comments

This section describes the main comments that reviewers have made during the Mid-Term Review meeting in terms of AGINFRA PLUS sustainability. We are reflecting upon the way that this feedback was taken into consideration when developing the revised Sustainability Roadmap.

Table 1: Mid-term review recommendation and consortium actions

No	Mid-term review recommendation	Consortium action
1	The consortium should provide: ● Background IP used to develop the project's deliverables. ● Foreground IP developed in the project. ● Relevant TRL information for all components that are advanced in the project in terms of technology readiness	The revised version of the Mid-Term Progress Report included in initial outline of this information. In this version of D8.4, a full chapter is dedicated to present this information for all technology outcomes produced by the project.
2	An early version of deliverable D8.4, shall be delivered at M24. This version should provide details on: (i) IPR management and the IP shares of the partners in the foreground generated in the project (ii) Plans for engaging SMEs in the project and for pursuing relevant exploitation opportunities (iii) A detailed sustainability plan for the project's results (VRE, journals), including relevant financial analysis and plans for attracting funding and generating revenue.	An early version of D8.4 was delivered at M24, mainly focusing on scenarios to further support and sustain VREs. In this version, we have significantly revised the original plan in order to cover all key technology offerings that the partners have decided that they wish to further exploit.

1. AGINFRA PLUS TECHNOLOGY OVERVIEW

1.1 AGINFRA PLUS Technology Outcomes

In this section, a complete list of the technology outcomes of AGINFRA PLUS is presented. We reflect on what has been foreseen originally in the DoA versus what we have actually delivered, as well as the Technology Readiness Level (TRL) and corresponding IP ownership details for each component or service. From this extensive list, four of our technology partners have decided that they wish to continue sustaining a number of core technology offerings on their own - independently from each other or together in software/service bundles.

1.1.1 AGINFRA data infrastructure & VRE

1.1.1.1 AGINFRA D4Science-powered VREs (ex- AGINFRA Virtual Organisation as per DoA)

As part of AGINFRA PLUS activities, a series of Virtual Research Environments (VREs) has been provisioned to target research communities, over the D4Science e-infrastructure. A VRE serves as an integrated environment providing researchers with seamless access to data, software, services and computing resources needed to perform research activities independently of organisational and technical barriers. In the case of AGINFRA PLUS, a VRE user can potentially acquire access to three kinds of tools and services: (i) Data and Semantics tools, (ii) Data Visualization tools and services, (iii) Data processing tools and services. This set is typically complemented with a number of collaborative features, such as social networking and shared folder (workspace) capabilities. Moreover, every VRE is equipped with a highly customisable resource catalogue enacting the VRE designated community to define the typologies of items worth sharing by publishing them into such a catalogue.

Background IP used to develop project deliverables (on M1)	**TRL before project (on M1)**	**Foreground IP developed within AGINFRA PLUS (M1-M36)**	**TRL at the end of project (on M36)**	**Foreground IPR owner**	**Licensing or Protection Action (if applicable)**
D4Science: a comprehensive and open infrastructure offering mechanisms for creating and configuring VREs by combining gCube services with gCube friendly services	TRL 5 – technology validated in relevant environment	gCube-friendly services extended with new data & semantics, visualization and processing tools and services	TRL 9 –actual system proven in operational environment	CNR, EGI	European Union Public Licence (EUPL v.1.1) license

An as-a-Service delivery model has been decided for all AGINFRA PLUS VREs. This model makes it possible (i) for the service provider, to leverage economies of scale to keep developments and operational costs low; (ii) for the service consumer, to acquire the services and the capacity needed in an elastic way. The presence of services delivered with the as-a-Service delivery model is potentially reducing the effort and costs needed to implement research supporting environments yet it is not nullifying neither the efforts nor the costs. Depending on the typologies of services that are made available by service providers, scientists might be requested to implement and operate on their own what is missing to get the research supporting environment they need.

CNR has been responsible for the integration, evolution and customisation of the AGINFRA PLUS provisioned VREs. Moreover, CNR has deployed and operated the gateway hosting all the VREs co-developed during the AGINFRA PLUS by the technical partners and the scientific communities. The D4Science offering of services and computing capacity is complemented by EGI both in terms of services to include into VREs (e.g. Jupyter and Galaxy cf. Sec. 1.2.2) and computing capacity exploitable by the data analytics solution.

Location of this software outcome:

- The AGINFRA PLUS Gateway of relevant Virtual Research Environments can be accessed here: https://aginfra.d4science.org
- The D4Science e-infrastructure is powered by gCube, an open-source software toolkit used for building and operating Hybrid Data Infrastructures, which is available on GitHub: https://github.com/gcube-team/gcube-releases.

1.1.1.2 AGINFRA Data Registry

During the project duration, a large amount of research assets has been collected, processed, aggregated and made openly available through the AGINFRA PLUS e-infrastructure. The so-called AGINFRA Data Registry supports the following typologies of data, gathered from 9 federated data registries online, including the AGINFRA-powered VREs:

- Publications
- Datasets
- Research Objects
- Projects & Initiatives
- Semantic Resources
- Models & Algorithms
- Organizations
- Data Sources
- Software & Services

The federated registry was assembled and populated using a novel data harvesting workflow-based approach. It has also been enhanced with a new semantic querying engine (mentioned in 1.1.5) over the collected metadata records. To further develop its visibility and exploitation options, a synchronization with the CIARD RING registry is already planned.

Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Foreground IPR owner	Licensing or Licensing or Protection Action (if applicable)
The agINFRA registry of data and software, imprinted as the agINFRA API Gateway in CIARD RING	TRL 5 – technology validated in relevant environment	An enriched registry of data collected, processed and made available as a service	TRL 8 –system complete and qualified	Agroknow	Open Data Commons Open Database License

Location of this software outcome:

- The AGINFRA Data Registry is available online and can be invoked according the following documentation: https://api.agroknow.com/search-api/swagger-ui.html#/. All registry metadata records are available under an Open Data Commons Open Database License.
- Its source code is available on GitHub: https://github.com/AGINFRA-PLUS/SearchAPI

1.1.1.3 AGINFRA Registry of Semantic Resources

Along the project duration, a registry of ontologies and vocabularies was stored and managed using the Ontological Engineering Technologies of the proposed e-infrastructure (as described in D2.2). To populate the registry, a strategic synergy with Agroportal was achieved to facilitate the pulling of content that was relevant to the engaged use-case communities. At the same time, new *data standards* (such as the FSK-ML) and *domain-specific ontologies* (such as community-generated GACS subsets) have been defined and stored with the use of the online tools.

Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Agroknow	Licensing or Protection Action (if applicable)
Did not exist before M1	TRL1 – basic principles observed	An API-accessible registry of ontologies and metadata descriptions	TRL 6 – technology demonstrated in relevant environment	Agroknow	Apache v2.0

The registry was later made accessible through a dedicated API-based service, that vertically exposed the wealth of information gathered from all AGINFRA PLUS Virtual Research Environments. This service now is now part of the D4Science infrastructure, but its contents were also later harvested by

the AGINFRA Data Harvesting Workflow, to further build upon the typology of Semantic Resources that would be used for annotation and schema mapping purposes.

A communication channel with the Agroportal team has been established for future integrations with the Registry, so as its contents can eventually become part of the VEST Registry / AgroPortal map of standards[1].

Location of this software outcome:

- The AGINFRA Registry of Semantic Resources can be found as part of the infrastructure currently hosting the identified Ontological Engineering tool, VocBench: https://vocbench1.d4science.org/api/docs/. All registry records are available under an Open Data Commons Open Database License.
- Its source code is openly available on GitHub: https://github.com/AGINFRA-PLUS/semantic-resources-api.
- The data within the registry are also available for exploitation through the AGINFRA Semantic API (link).

1.1.1.4 Metadata of all indexed data sources

During the successful run of the e-ROSA project, the landscape of the agri-food open science resources and stakeholders had been developed (described here). Part of the mapping activities focused on imprinting the relevant indices, repositories, catalogues, aggregators and databases that exist, along with information about the assets they store and their respective access details. Agroknow has been responsible for collecting, enriching and publishing metadata for these data sources, even after the duration of e-ROSA.

In the context of AGINFRA PLUS activities, the so-called Data Ecosystem Map has been extended, both in content, but also with the introduction of an API-based web service that exposes it to external users. At the same time, the Agroknow Data Platform has collected its content and made it available through its respective web service.

Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Foreground IPR owner	Licensing or Protection Action (if applicable)
The Data Ecosystem Map generated from eROSA project activities	TRL 6 – technology demonstrated in relevant environment	A API-accessible registry of global data sources, along with information on the data records they provide	TRL 8 –system complete and qualified	Agroknow	Open Data Commons Open Database License

[1] http://aims.fao.org/vest-registry

Location of this software outcome:

- The online Data Ecosystem Map can be found on: http://map.aginfra.eu/.
- The data sources available on the map can be accessed openly via the respective API: http://map.aginfra.eu/api/sources, but also through the AGINFRA Search API: https://api.agroknow.com/search-api/swagger-ui.html#/. All metadata records are available under an Open Data Commons Open Database License.

1.1.1.5 Data Harvesting Workflows & Data Linking Services

The AGINFRA Data Harvesting Workflows is a software component that performs the tasks of scanning one or more federated data registries and collecting new metadata records that they include. The component execution can be configured to occur periodically, thus enabling the option to be offered as a service. Agroknow has been responsible for it as part of WP2 activities and has tested it with nine major federated data sources:

1. AGRIS,
2. AGINFRA PLUS VREs,
3. Agroportal,
4. USDA,
5. CGIAR GARDIAN,
6. OpenAIRE,
7. ZENODO,
8. eROSA,
9. Unpaywall.

The AGINFRA Data Linking Services is a family of services that enable users to search and submit data assets to the AGINFRA Data Registry, but also gain access to the AGINFRA Registry of Semantic Resources and use it as a pool for data annotation. More specifically, the Data Linking Services include:

1. AGINFRA Search API, which serves as the primary gateway to the AGINRA Data Registry, offering semantic discovery capabilities;
2. AGINFRA Data Integration API, which provides all CRUD operations to the AGINFRA Data Registry;
3. AGINFRA Semantic API, which provides text analytics and semantic annotation capabilities.

Background IP used to develop project deliverables (on M1)	**TRL before project (on M1)**	**Foreground IP developed within AGINFRA PLUS (M1-M36)**	**TRL at the end of project (on M36)**	**Foreground IPR owner**	**Licensing or Protection Action (if applicable)**
agINFRA Harvester in the scope of the agINFRA FP7 project.	TRL 6 – technology demonstrated in relevant environment	Re-engineering to enable the generalization of the service for data sources beyond publications	TRL 8 –system complete and qualified	Agroknow	Apache v2.0

Location of this software outcome:

- An online interface, where users can configure and invoke the AGINFRA Data Harvesting Workflow can be found here: http://52.214.72.17:8888/admin/
- The AGINFRA Data Harvesting Workflow software code can be found on GitHub: https://github.com/AGINFRA-PLUS/AGINFRA_PLUS_Data_Harvesting_Workflows

1.1.6 AGINFRA web portal with search engine and map of ecosystem

The AGINFRA web portal is a family of websites and front-end tools that promote access to the AGINFRA services and data ecosystem. These include:

1. The AGINFRA Map of the Data Ecosystem, visualizing and providing details of all organizations, initiatives, data sources and facilities that comprise the digital landscape of agri-food sciences today.
2. The AGINFRA Semantic Search engine[2], which has been prototyped as part of AGINFRA PLUS project activities (described in D2.3), providing access to more than 7,7 million open data assets.
3. Information and links to the AGINFRA PLUS e-infrastructure, such as links to D4Science-powered VREs, supported communities' Data Journals, and consortium outputs in general.

Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Foreground IPR owner	Licensing or Protection Action (if applicable)
Did not exist on M1	Did not exist on M1	AGINFRA web portal including: the e-ROSA website, the AGINFRA Map of Data Ecosystem and the AGINFRA project website	TRL 9 –actual system proven in operational environment (for web portal and data ecosystem map) TRL 7 – system prototype demonstration in operational environment (for AGINFRA Search Engine)	Agroknow	Apache v2.0

Location of this software outcome:

- The AGINFRA web portal can be accessed through: https://aginfra.eu.

[2] https://plus.aginfra.eu/semantic-search

- The source code of the Semantic Search engine can be found on GitHub: https://github.com/AGINFRA-PLUS/semantic-search.

1.2 AGINFRA PLUS software tools & online environments

1.2.1 Ontological Engineering tools

WP2 activities resulted in two different layers of the proposed AGINFRA PLUS e-infrastructure: The Ontological Engineering layer and the Data Linking services layer. The first one comprises of all the individual user-facing software tools that enable users to perform the tasks of ontology management and vocabulary management.

The identified tools have been selected as per the needs of the AGINFRA PLUS use-case communities and have been integrated as part of the Virtual Research Environments (VREs) powered by the project as part of the D4Science infrastructure. Any given community that uses the AGINFRA PLUS e-infrastructure can be given access to the given tools, per VRE.

The only exception is the Data Integration Tool, which was a new tool that was prototyped outside of the D4Science infrastructure.

Tool	Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Foreground IPR owner	License
VocBench	The development of VocBench3 is managed by the Publications Office of the European Union.	TRL 9 – actual system proven in operational environment	Extension of the authentication layer of VocBench that allowed its integration with D4Science	TRL 9 –actual system proven in operational environment	Agroknow	BSD 3-Clause "New" or "Revised" license
YAM++ Matcher	The YAM++ Large Scale Library developed by Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier	TRL 6 – technology demonstrated in relevant environment	Development of a wrapper interface for the YAM++ library and its adaption over the D4Science infrastructure	TRL 7 – system prototype demonstration in operational environment	Agroknow	GNU General Public License

OpenRefine + RDF Extension	OpenRefine was created by Metaweb Technologies, Inc. The RDF extension has been developed by the Digital Enterprise Research Institute.	TRL 9 – actual system proven in operational environment	OpenRefine and RDF extension were installed and integrated into the D4Science infrastructure	TRL 9 –actual system proven in operational environment	Agroknow	BSD 3-Clause "New" or "Revised" License
WebVOWL	WebVOWL was developed in the context of the Visual Data Web project	TRL 9 – actual system proven in operational environment	WebVOWL was installed and integrated as part into the D4Science infrastructure	TRL 9 –actual system proven in operational environment	Agroknow	MIT License
SKOS Play!	SKOS Play! Is developed by Sparna	TRL 9 – actual system proven in operational environment	SKOS Play! was installed and integrated into the D4Science infrastructure	TRL 9 –actual system proven in operational environment	Agroknow	GNU Lesser General Public License v3.0
Data Integration Tool	Did not exist before M1	-	Tool was prototyped with the contribution of INRA	TRL 3 - characteristic proof-of concept	Agroknow	Apache v2.0

Location of this software outcome:

- The AGINFRA PLUS Ontological Engineering Tools are currently accessible online under the VREs powered by the AGINFRA PLUS Gateway: https://aginfra.d4science.org/group/aginfra-gateway. Individually, each tool can be accessed on the following URLs:
  - VocBench: https://aginfra.d4science.org/group/aginfraplus/vocbench;
  - YAM++ Matcher: https://aginfra.d4science.org/group/aginfraplus/ontology-matching;
  - OpenRefine: https://aginfra.d4science.org/group/aginfraplus/openrefine;
  - WebVOWL: https://aginfra.d4science.org/group/aginfraplus/ontology-visualization;
  - SKOS Play!: https://aginfra.d4science.org/group/agroclimaticmodelling/vocabulary-visualization.
  - Data Integration Tool: https://github.com/AGINFRA-PLUS/data-integration-tool

1.2.2 Data analytics tools and Open Science helpers

Under the data analytics umbrella developed in AGINFRA PLUS there are a lot of services and facilities that once exploited together make it possible to implement open science practices. These tools and services include:

- The Data Analytics platform, i.e. the web-based workbench enacting its users to execute analytics tasks by using the available analytics methods integrated;
- The SAI, i.e. an integration environment enacting its users to transform an implementation of an analytics method into a method offered by the data analytics platform;
- Analytics method executor helpers, i.e. an URI resolver and a portlet enacting to identify and execute a single method of the analytics platform;
- A Catalogue service, enacting its users to publish any artifact worth being published for discovery purposes including methods;
- REST-based services, namely a REST-based service for accessing workspace content and a REST-based service for interfacing with the catalogue content;
- Integrations with Galaxy and JupyterHub services, powered by EGI;
- Integrations with KNIME-based workflows for food predictive and microbial modelling;
- Crop growth model configurations mechanism for model simulations on D4Science VREs.

Tool	Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Foreground IPR owner	Licensing or Protection Action (if applicable)
gCube Data Analytics	gCube Data Analytics	TRL 8 – system complete and qualified	The DataMiner framework deployed in the AGINFRA PLUS VREs was modified to accommodate a new algorithm integration approach, new logging and accessing mechanisms. Furthermore it was extended to be interoperable with other systems, namely KNIME and Jupyter	TRL 9 – actual system proven in operational environment	CNR	European Union Public License

gCube SAI	gCube SAI	TRL 8 – system complete and qualified	The gCube SAI has been extended with several features, e.g. new parameter typologies,new algorithm typologies	TRL 9 – actual system proven in operational environment	CNR	European Union Public License
gCube URI Resolver	gCube URI Resolver	TRL 8 – system complete and qualified	Extended the resolver to deal with two specific applications: analytics executor and KNIME	TRL 9 – actual system proven in operational environment	CNR	European Union Public License
gCube Analytics Executor	gCube	n.a.	A new component enacting the execution of a single method of the analytics platform	TRL 8 – system complete and qualified	CNR	European Union Public License
gCube SHub	gCube Home Library	TRL 8 – system complete and qualified	A new REST based service for interfacing with the workspace	TRL 8 – system complete and qualified	CNR	European Union Public License
gCube Catalogue	gCube Catalogue	TRL 8 – system complete and qualified	Extended to support geospatial coverage, item rating.	TRL 8 – system complete and qualified	CNR	European Union Public License
gCube gCat	gCube Catalogue WS	TRL 7 – System prototype demonstration in operational environment	A new REST based service for interfacing with the catalogue	TRL 8 – system complete and qualified	CNR	European Union Public License
gCube gFeed	gCube	n.a.	A new REST based service	TRL 7 – System prototype	CNR	European Union Public

			for publishing into the catalogue analytics methods	demonstration in operational environment		License
FSK-Lab	FSK-Lab	TRL 8 – system complete and qualified	The KNIME extensions were made interoperable with the AGINFRA PLUS Analytics framework	TRL 9 – actual system proven in operational environment	BfR	GNU General Public License v3.0
PMM-Lab	KNIME extension for predictive microbial modelling	TRL 8 – system complete and qualified	Integrated KNIME extension	TRL 8 – system complete and qualified	BfR	GNU General Public License v3.0
Food Chain-Lab	KNIME extension for tracing analysis and supply chain visualization	TRL 8 – system complete and qualified	Integrated KNIME extension	TRL 8 – system complete and qualified	BfR	GNU General Public License v3.0
JupyterHub as a Service	JupyterHub	TRL 8 – system complete and qualified	The integration components allow the interoperation of Jupyter Notebooks with the AGINFRA PLUS framework	TRL 8 – system complete and qualified	EGI	BSD license
Galaxy as a Service	Galaxy	TRL 8 – system complete and qualified	The integration components allow the interoperation of Galaxy with the AGINFRA	TRL 8 – system complete and qualified	EGI	Academic Free License version 3.0

			PLUS framework			
WOFOST CLOUD	Standalone and desktop-dependent WOFOST model - no cloud version existing before M1	-	Extended for distributed simulations and job management based on the D4Science infrastructure	TRL 7 – System prototype demonstration in operational environment	DLO	European Union Public License v1.2

Location of this software outcome:

- The tools of this area are all made available by AGINFRA PLUS Virtual Research Environments, e.g. in the AGINFRA PLUS VRE:
  - Data analytics: https://aginfra.d4science.org/group/aginfraplus/data-miner
  - Jupyter: https://aginfra.d4science.org/group/aginfraplus/jupyter-egi
  - Galaxy: https://aginfra.d4science.org/group/aginfraplus/galaxy-egi
  - Catalogue: https://aginfra.d4science.org/group/aginfraplus/catalogue
- The software code can be found on GitHub and Zenodo:
  - gCube software: https://github.com/gcube-team/gcube-releases, https://zenodo.org/communities/gcube-system
  - Galaxy and Jupyter extensions: https://github.com/AGINFRA-PLUS
  - FSK-Lab: https://github.com/SiLeBAT/FSK-Lab
  - PMM-Lab: https://github.com/SiLeBAT/PMM-Lab
  - FoodChain-Lab: https://github.com/SiLeBAT/BfROpenLab
  - WOFOST CLOUD: https://github.com/AGINFRA-PLUS/wofost-cloud-poc

1.2.3 Visualisation tools

As part of WP4 activities, the following tools were made available under the AGINFRA PLUS VREs:

- Geoanalytics Framework: for managing and visualising various formats of geospatial data. Project management for fusing and sharing data and execution of custom analytics algorithms.
- Visualisation Framework: includes the backend service and the presentation applications / portlets
- AgroDataCube Dashboard: Application for the visualisation of parcels in the Netherlands and visualisation of a number of information about them using charts.

Tool	Background IP used to develop project deliverables (on M1)	TRL before project (on M1)	Foreground IP developed within AGINFRA PLUS (M1-M36)	TRL at the end of project (on M36)	Foreground IPR owner	Licensing or Protection Action (if applicable)
Geoanalytics Framework	This work was already available with shared IPR among UoA and CITE S.A. and was available under EUPL V1.1	TRL 6 – technology demonstrated in relevant environment	UoA extended the framework to support additional data formats, improve the UI for the end-users to provide better management of the data, concerning their style and presentation, to support additional CRSs and improve performance and metadata publication for each dataset.	TRL 7 – system prototype demonstration in operational environment	UoA	EUPL v1.1
Visualisation framework	-	TRL 1 –basic principles observed	UoA developed the backend services, a number of visualisation libraries and the front-end applications. The front-end applications are available in both as standard React web application and as a JSR 286 portlet.	TRL 6 – technology demonstrated in relevant environment	UoA	EUPL v1.1
Agrodatacube Dashboard	-	TRL 1 –basic principles	UoA exploited the Visualisation	TRL 7 – system	UoA	EUPL v1.1

Application		observed	Framework and developed a web application / portlet for presenting a number of charts with information about parcels in the Netherlands.	prototype demonstration in operational environment		
Network Graph Application	-	TRL 1 –basic principles observed	UoA developed a web application for visualising changes in the structure of a network (graph) during time	TRL 5 – technology validated in relevant environment	UoA	EUPL v1.1
MindMap portlet	An open source project was available under AGPL V3 (link)	TRL 8 – system complete and qualified	UoA extended this open source software and developed a native JSR 286 portlet, adding integration with the Agrinfra+ workspace for storing and retrieving the created mind maps and supporting URLs in node's text.	TRL 8 – system complete and qualified	UoA	AGPL V3

Location of this software outcome:

All tools can be accessed through the AGINFRA PLUS gateway and its VREs: https://aginfra.d4science.org/

- Users can register to the AGINFRADev VRE and have access to all applications: https://aginfra.d4science.org/group/aginfraplusdev/data-visualisation
- https://aginfra.d4science.org/group/aginfraplusdev/geospatial

The source code of the tools is available on Github:

- Visualisation Framework: https://github.com/madgik/VisualisationFramework
- Geoanalytics Framework: https://github.com/gcube-team/gcube-releases/tree/master/geoanalytics
- MindMap Portlet: https://github.com/madgik/mind-map-portlet

2. THE NEED

2.1 The Business Opportunity

2.1.1 Business needs

Due to the emergence of big data technologies and the open science paradigm, the way that research is taking place today around the world is under huge transformation. New methods and tools (such as computationally-powered research) arise; new research professions (such as data managers and scientists) are created; new requirements (on making publicly funded research outcomes openly accessible) are being posed by the funding agencies; search engines like Google Scholar (http://scholar.google.com) create huge catalogues of research publications and make them instantly discoverable; data processing, analysis and visualisation tools are made available either as open source software that everyone can use or as commercial off-the-shelf tools. This fact is expected to change the way in which research is taking place also in the agri-food sector, which becomes a very large potential market for ICT solution providers.

2.1.2 Vision

As it was stated in Agroknow's contribution at the GODAN discussion paper "A Global Data Ecosystem for Agriculture and Food" (2016)[3], a global data space for agriculture and food would propel the whole industry forward[4]. Information would become available to all actors producing innovation. Analytical and decision-making tools could incorporate a greater abundance of data sources. A digital economy would arise with online services and applications that use machine readable, interoperable and often publicly shared data. The necessary infrastructure components, including technology, people, policy and business ones, could seamlessly integrate and work together.

Three years later, the vision is still the same. AGINFRA PLUS has showcased the way in which digital infrastructures and computational resources may boost the potential of data-intensive applications that scientists use. In our continuously interconnected society, our digital tools, services and infrastructures aim to empower the scientists that work on addressing some of the grandest challenges that our planet is facing.

2.1.3 Brand

Our intention is not to build a single AGINFRA brand that will package together the different partner offerings and promote them as a whole. We rather embrace the notion of contributing to an open data and service ecosystem, where different tools and services can be deployed, interconnect, exchange data, request or provide services, and serve in different ways scientists and research staff in food and agriculture. This is fully in line with the EOSC vision but also with the vision that the food and agriculture stakeholders have outlined in documents such as the Chania Declarations 2016[5] and 2017[6]. To this end, the different AGINFRA PLUS offerings will maintain their own identities, brands and positioning in the market. Putting them together in action in the context of AGINFRA PLUS is helping us to demonstrate the value of combining infrastructural components with various software tools and services in order to further empower scientists.

[3] https://www.godan.info/documents/global-data-ecosystem-agriculture-and-food
[4] https://docs.google.com/document/d/1QZZJ2gxgVfrJ8czogexsT5ea-a0tYkKQ-SUUXeAJXrY/edit
[5] http://blog.agroknow.com/wp-content/uploads/2016/05/Chania-Declaration.pdf
[6] https://drive.google.com/file/d/0B41Vz7BieQuoY19CYlBISklxbmc/view

2.1.4 The Solution: AGINFRA offerings

The key technology outcomes of the project that partners have decided to continue sustaining and evolving after its end, are the following:

- **D4Science-powered VREs (CNR)**: D4Science Virtual Research Environments (VREs)[7] are web-based working environments that are conceived to serve the needs of their designated communities. Each VRE is developed by promoting (a) the re-use of domain-specific data and services by integrating them into a unifying and augmented resource space that preserve provenance and attribution, (b) the co-creation and co-development of the working environment of interest for the community of practice, and (c) the use of state of the art solutions for collaboration, communication and open science. D4Science has been developed and offered to a number of scientific communities during the past years. It is a collaborative effort by a number of institutions (including CNR & UoA) and it will continue to evolve after AGINFRA PLUS ends. The scientific communities that co-created VREs powered by D4Science within AGINFRA PLUS, should expect that this service should continue to be offered after the project ends. Moreover, new VREs replicating AGINFRA PLUS VREs from the functional perspective can be created and deployed to serve the needs of new communities and cases. All in all, D4Science VREs represent a sort of "super" service providing their users with an aggregative and user friendly working environment where the services listed below can be packaged together with (i) a *shared workspace* to store and organise any version of a research artefact; (ii) a *social networking area* to have discussions on any topic (including working version and released artefacts) and be informed on happenings; (iii) a *data analytics platform* to execute processing tasks either natively provided by VRE users or borrowed from other VREs to be applied to VRE users' cases and datasets; and (iv) a *catalogue-based publishing platform* to make the existence of a certain artefact public and disseminated. In the respective parts of this deliverable we describe how this is relevant and will be done for food and agriculture stakeholders.
- **EGI-powered Jupyter Notebooks (EGI)**: the Jupyter Notebook is an open source web application that supports interactive data science and scientific computing across several programming languages. Jupyter works as an interactive computational environment in which users can combine code execution, rich text, mathematics, plots and rich media. During AGINFRA PLUS, it has been integrated as a tool within the D4Science VREs that enables the creation of notebooks that access and combine programmatically the different features of the D4Science VREs in single notebook document. The AGINFRA PLUS development serves as basis to the EGI Notebooks service that can be offered as a standalone application. In (section 3.5) we describe how this is relevant and will be done for food and agriculture stakeholders.
- **EGI-powered Galaxy Workflows (EGI)**: Galaxy is an open, web-based platform for data intensive biomedical research. It allows bioinformatics and life science researchers to run complex workflows and share data. During AGINFRA PLUS, it has been integrated as a tool within the D4Science VREs, but can also be offered as a standalone application - accessed via the EGI Applications on Demand service. In (section 3.5) we describe how this is relevant and will be done for food and agriculture stakeholders.
- **Data Harvesting & Data Linking Services (Agroknow)**: a number of publication metadata harvesting and indexing workflows have originally been developed within FP7 agINFRA. They have then been enriched within document mining and semantic annotation capabilities within H2020 OpenMintTeD. During AGINFRA PLUS, these services have been extended to include harvesting, crawling and indexing of other scientific data types (such as models and data sets). Additionally, a new set of semantic APIs has been set up, to allow other platforms search into an aggregate of almost 8 million of scientific resources in food and agriculture, use the AGINFRA registry of semantic resources to analyze and annotate their own data assets, but

[7] Assante et al. (2019) Enacting open science by D4Science. Future Generation Computer Systems 10(1016), 555–563 (2019), http://www.sciencedirect.com/science/article/pii/S0167739X1831464X

also to provide their own data and metadata schemas to the AGINFRA PLUS registry. This API instance has been made accessible from within the D4Science VREs but can also be offered in a SaaS model by Agroknow. In (section 3.5) we describe how this is relevant and will be done for food and agriculture stakeholders.

- **ARPHA-powered Data Journals (Pensoft)**: ARPHA is a full-featured open access publishing platform for journals, books and data, with managed hosting services. Offered as Software as a Service (SaaS), ARPHA allows users to build their own publishing solution to manage and host journals, books, conference abstracts, proceedings and institutional documents. During AGINFRA PLUS, it has been integrated as a tool within the D4Science VREs but can also be offered in a SaaS model by Pensoft. In (section 3.5) we describe how this is relevant and will be done for food and agriculture stakeholders.

2.2 Market Introduction

2.2.1 Market Overview

Agri-food research is taking place in a complex ecosystem that includes a variety of stakeholders, including (among others) research and academic organisations, funding agencies and donors, policy and decision makers, regulatory and certification bodies, farmers and producers, food and cosmetic industries, as well as processing and retail businesses. In this ecosystem, several specific markets exist that are self-referenced and that have their own challenges and problems: for example, the agricultural research organisations, the food safety community, the organic agriculture community and so it goes. Our aim is to focus on these markets and their segments, work on revealing their specific problems and needs, and to provide a customized set of solutions that is particularly fit to each segment.

The user problems that we focus on are related to research information: that is, all types of information that is being produced during, after and about the research that people are carrying out in this sector. This information goes beyond bibliographic information (publications or grey literature) and includes: primary research data (e.g. measurements), secondary research data (e.g. statistically processed data), visualisations of data (e.g. diagrams and figures), research methods (e.g. models, algorithms) and instruments; or even information on research projects, organisations, teams, individual researchers, funding opportunities, event opportunities etc. Challenges are not generic but rather depend on the application context: for example, viticulture researchers have difficulties in documenting and managing observations about grape varieties during their phenotypic and genetic research; food defense and disease prevention researchers are interested in discovering and combining relevant data on reported foodborne illnesses and food product recalls; etc.

From the variety of stakeholders in this ecosystem we primarily focus on the people carrying out the actual research and communicating it to the outside world. The main economic benefit will come from the fact that they will adopt a solution that will help them manage their complete research and data workflows and then allow them to create different dissemination channels for their scientific outcomes. Thus, they will improve the way that they manage all their research information, reduce the time that they need with current tools, but also avoid new investment: each new science project can build on top of the pool of research knowledge that other scientists have produced and that can be discovered online. An important selling point to our approach is the fact that much of this information already exists in various online sources, such as institutional information systems, publication catalogs from publishers, theme-specific bibliographic aggregations and search engines, model repositories, open source data science pipelines and software stacks etc.

Our total addressable market (TAM) is the agri-food research one: all organisations and people funding, producing or consuming research activities and outcomes related to agriculture and food. Still, since our solutions are operating online, the currently served available market (SAM) includes the existing users of research information systems – and more specifically, online ones.

There is an existing market base of over 10 million users of social research platforms and bibliography/reference management toolkits such as ResearchGate (>4M users), Academia.edu (>10M users) and Mendeley (>2,5M users). This is also a growing market: in the OECD area, around 4.3 million persons were employed as researchers in 2011 and this indicator has steadily increased over the last two decades (http://stats.oecd.org).

As far as the public sector is concerned, our aim is to prioritise regions/countries that are making a high investment in research: these are mainly Europe, where the H2020 SC2 Work Programme invests hundreds of millions every year on food system research; the US, where the Department of Agriculture (USDA) spends almost $2.5 billion for R&D every year; and Canada, where the Department of Agriculture and Agri-Food Canada (AAFC) invests over CA $500 millions every year in research activities, while provincial governments deploy own programs.

In the private sector, investments to R&D and innovation are also very high. For instance, agrochemical companies, like Bayer and Syngenta, invest more than 10% of their annual revenue in R&D - which corresponds to amounts between 1 billion to 1.5 billion euros. This is why a significant market for our technology outcomes includes also the scientists working in the private sector.

More specifically, the market sectors that we are aiming to include:

- **The public sector:** this includes research institutions and centers, universities, scientific associations and groups, and other organisations that belong to the public sector and are funded either by governmental funds or by donors.
- **The private sector:** this includes the food industry stakeholders (that is, food producers, processors, and distributors), the agrochemical industry stakeholders (researching products for crop protection and crop yield improvement), and small and medium enterprises (SMEs) of the agtech and foodtech sector that are developing products powered by science.

2.2.2 Focusing on three market segments:

In order to exploit the AGINFRA PLUS offerings described above, the responsible partners have decided to focus on the following three market segments in which they will try to position the relevant offerings:

- The scientists that are working in public sector universities & research centers in the European Union and associated countries[8]. More specifically, the ones working in institutions eligible to receive both EC and national research funding, therefore being able to take advantage of European digital infrastructures.
- The scientists working in the food industry, especially in departments and groups that are critical to business operations and innovation. More specifically, the ones working in large food companies (over 1BEuros revenue) that have very large and distributed scientific teams working in Food Safety, Quality Assurance, Regulatory, Product R&D and Innovation departments.
- The scientists working in data science powered startups and SMEs in ag/foodtech. More specifically, the ones responsible for computationally demanding tasks that innovative companies put in place (such as predictive analytics, DNA sequence mining, satellite image processing, sensor streams analysis, etc).

[8] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/3cpart/h2020-hi-list-ac_en.pdf

3. OUR POSITIONING

3.1 Market Analysis & Positioning

3.1.1 Market Description of Public Sector Funded Scientists

Publicly funded research is one of the main segments that the AGINFRA PLUS will target. According to the United States Department of Agriculture, North America and Europe are the top regions among high-income countries with the largest amount of public agricultural R&D investment (almost 14 billion dollars). Both regions are being examined as the two main sub-segments of the market of publicly funded science in food and agriculture. There are common priorities that can be observed in both sub-segments of the market focusing on data and digital technology adoption. Harnessing technological innovation is one of the top priorities of U.S. government's agricultural research interests (Report to the President of the United States from the Task Force on Agriculture and Rural Prosperity, 2017) , whereas for European Union competitiveness and growth for the Food and Natural Resources sectors through research innovation and technology is on the top of the list based on the Horizon Europe research and innovation funding programme for 2021 – 2027.

3.1.2 The key characteristics of this market

We will focus first on the European market, especially the part funded by the EC but also from national funds. We use data from the market research that Agroknow has done during 2018-2019 on major funding programmes, key institutions, outlook of H2020 and Horizon Europe funding on relevant topics. Give an estimate of the number of projects that were running and that will be running after 2020, to size the market targeted.

The importance of the two market sub-segments required a market analysis to be conducted in order to highlight key stakeholders (and potential solution adopters) as long as the size of each segment. The consortium foresees agri-food research projects as an individual entity which can support and sustain the AGINFRA PLUS solutions.

Table 2:The top three stakeholders in terms of total EU funding contribution and the number of agri-food R&D projects

Public R&D Organization	Number of Agri-food R&D projects	Total EU funding contribution (in million euros)
Stichting Wageningen Research	113	63.0M
Institut National de la Recherce Agronomique	96	49.4M
Wageningen University	58	30.5M

Agri-food research and innovation in Europe is being funded through a) European Commission's funding programmes - such as Horizon 2020 and Horizon Europe, b) Regional or interregional research

funds (e.g. Interreg and Prima) and c) national research funding schemes. Projects that are funded mostly through category a) are considered as a market worth exploring for the AGINFRA PLUS solutions. Especially in the context of Horizon 2020, the total investment for food and agriculture accounted for almost 4 billion euros (2014-2020) and covers 754 individual research projects across Europe. Based on data published by the European Commission for Horizon 2020, the top three stakeholders in terms of total EU funding contribution and the number of agri-food R&D projects are presented in Table 2.

Besides the top three organizations mentioned above, from the list of 1.274 Horizon 2020 research organizations participating entities, as the key players of this market are also considered the following, according to the market value and size of funding for agri-food R&D they attract: University of Copenhagen, Aarhus University, Netherlands Organisation for Applied Scientific Research (TNO), Flanders Research Institute for agriculture, fisheries and food (ILVO), TEAGASC- The Agriculture and Food Development Authority of the Republic of Ireland, Ghent University, Landbrug & Fødevarer F.m.b.A. (SEGES PS).

With regards to the North America subsegment, the potential solution adopters can be all agricultural research organizations that are granted federal and governmental funding. For 2017, the United States Department of Agriculture (USDA) Research & Development funding accounted for 2.3 billion dollars. USDA conducts intramural research at federal facilities with government-employed scientists and supports external research at universities and other facilities through competitive grants and formula-based funding. The USDA agencies which are considered as an integral part of the North America market sub-segment are Agricultural Research Service (ARS) – consisting of 90 laboratories nationwide with about 6600 employees and supporting over 300 active research projects (ARS data, 2017). Another key institution of USDA is the National Agricultural Library, one of the department's primary information repositories for food, agriculture, and natural resource sciences undertaking key research data management activities across the US. Moreover, National Institute of Food Agriculture (NIFA) provides federal funding for research conducted in partnership with the State Agricultural Experiment Stations, the State Cooperative Extension System and with land grant universities, colleges, and other research and education institutions. NIFA has over 150 agriculture research programs that have multiple projects underway[9].

Key players among the North America publicly funded research are a) for the United States: University of California-Davis, Cornell University, University of Massachusetts-Amherst, University of Florida, Harvard University, Michigan State University, Purdue University and b) for Canada: Agriculture and Agri-Food Canada (AAFC), University of Guelph, University of Saskatchewan, University of Alberta and University of Columbia. (AIC, An Overview of the Canadian Agricultural Innovation System. Ottawa, ON, 2017)

3.1.3 The key attributes of the publicly funded scientist persona

To briefly describe the buyer personas in this market, from the analysis that we did in 2018. Refer to the research staff that are the users but also the various economic buyers / decision makers that we have identified (project coordinators, lab directors, etc).

In this market segment the main persona which is involved in decision making is not the actual beneficiary of the service/ product. Most of the times the beneficiary should influence the economic buyer in order to decide about the purchase of a service, technology or solution. Project coordinators play the key decision-making role in publicly funded research segment. Project coordinators are characterized by scientific excellence, long year research experience and their academic role accomplishments. They are usually professors, associate professors and distinguished scientists. Apart from key decision makers, they are responsible for the sustainability of their teams and/or their laboratories, the financial viability of their department and the sustainability of their scientific projects.

[9]National Institute of Food and Agriculture Grants https://nifa.usda.gov/grants

3.1.4 Market maturity

The driving force for the adoption of the AGINFRA PLUS services is the constant digitalization of the agricultural Research & Development sector along with the importance that is put by the funding organizations on research e-infrastructures. Particularly for Europe, owing to the research data open access policy, it is requested for the publicly funded scientific projects to share, communicate and make available their research data and scientific models, as part of their projects' outcomes. At national level, in European countries, the e-Infrastructure policies are implemented by the responsible national authorities and research institutes. In that case they could act as promoters of the AGINFRA PLUS services to the national projects. From the global point of view, the importance of keeping data F.A.I.R. (Findable, Accessible, Interoperable and Re-usable) is being highlighted by multiple funding schemes. Metadata management, proper documentation and standards adoption are therefore fundamental to every agriculture and food research project, regardless of the country it takes place.

3.2 Market Description of Food Industry Scientists

This market segment refers to the scientists employed in the food and beverage industry and mainly possessing key roles within Quality Assurance, Scientific & Regulatory Affairs, Quality Control, Research & Development and Innovation departments.

3.2.1 The key characteristics of this market

The types of companies sought in this market segment are mainly billion-revenue conglomerates that are characterized by complex, global supply chains and are manufacturing, distributing and/or selling food and beverage products to consumers. According to Forbes Global 2000[10] and Fortune 500[11] lists, the top 25 companies in the food and beverage sector globally generated $1.1 trillion in revenue during 2018, up from $749 billion in 2017.

3.2.1 The key attributes of this persona

The targeted customer in this market segment is a food scientist, highly educated (possessing PhD title most of the times), who oversees Quality, Regulatory, Food Safety, R&D, Innovation functions across the business unit s/he is in charge of. The professionals under this persona are responsible for the reputation of the brand and the organization they represent and are very focused to highlight, protect and sustain their internal processes. At the same time, they need the food safety & quality community recognition and even though most of them are conservative -especially in terms of changing habits- they are very ambitious and want to be appreciated as innovators and experts not only among their organizations but also in the wider scientific community. In the corporate ladder, the best representatives of this persona are professionals with the roles of Vice President, President and in general C-level executives focusing on Quality, Food Safety, Regulatory Affairs, R&D and Innovation. They manage multinational teams led by Heads or Directors of each region who are responsible for the production of safe food and the prevention of unfortunate situations.

3.2.3 Market readiness

The global food industry market can be characterized as one of the major segments for the adoption of AGINFRA PLUS solutions. Recent studies show that key industry players (such as Walmart and

[10] https://www.forbes.com/global2000/

[11] https://fortune.com/fortune500/

Nestle, as they mentioned in a recent article in Forbes magazine[12]) have already invested in setting up digital infrastructures, putting technology solutions at the forefront of innovation in the food sector. To a great extent, most of the multinational food industry stakeholders are monitoring quality aspects of their global supply chains, sharing data and interact with each other using digital facilities and enterprise software systems. With that in mind, the adoption of the proposed solutions would be essential, provided that the value proposition is clear to the economic buyers of such organizations.

3.3 Market description of the agrochem industry scientists

A sector within the agri-food industry that includes heavy use of computational resources and digital research infrastructures is the one of agrochem research. Agrochemical companies are an integral part of the agricultural production, as they conduct R&D, produce and supply a variety of inputs for farming (from seeds to fertilizers and herbicides). The key stakeholders of this market are the four biggest agrochem corporations: Bayer (DE), Syngenta (CH), DowDupont (USA) and BASF (DE). Annual sales revenue of the companies mentioned above accounts for over 35 billion dollars, according to their annual reports' data for FY 2016. The table below highlight the percentage of the R&D investment that was put in place by each of the key stakeholders.

Table 3: The percentage of the R&D investment that was put in place by each of the key stakeholders.

Company Name	Headquarters Location	Annual Sales	R&D Investment
Bayer Crop Science	Germany	€ 9.9 B	€ 1.2 B (12%)
Syngenta	Switzerland	$ 12.8 B	$1.3 B (10%)
DowDupont	USA	$ 16.2 B	$ 1 B (6%)
BASF	Germany	€ 5.6 B	€ 483 M (9%)

3.3.1 The key attributes of this persona

The targeted customer in this market segment is a researcher/ scientist, highly educated (possessing PhD title most of the times), who is employed in the R&D, Innovation department of large multinational agrochemical companies. The professionals under this persona are responsible for conducting research and developing new active substances for plant protection and crop protection products, fertilizers and seeds. Most of the scientists of this group have a background in agronomy, plant pathology, chemistry, biology, statistics or mathematics and are dealing with large complex datasets. Moreover, they make heavy use of computational resources in order to deploy and run statistical models that will help them perform simulations on different environmental conditions. The outcomes of their research are most of the times proprietary information that belongs to the company and the research center that they collaborate with, so they are really concerned about the anonymization and security of their datasets and models. Data and informational resources can be shared among the different regions or different R&D centers of a company but seldom with third parties.

[12] https://www.forbes.com/sites/rogeraitken/2017/08/22/ibm-forges-blockchain-collaboration-with-nestle-walmart-for-global-food-safety/#192320403d36

3.3.2 Market readiness

The agrochemical industry can be characterized as a high potential market for the AGINFRA plus solutions, due to the increasing amount of R&D investment that are taking place year by year. Digital transformation that is happening in most multinational industries requires faster decisions and accurate use of scientific models. This segment has already in place tools and services (i.e. WEKA) which are domain-agnostic and are being deployed for specific agricultural or environmental research applications. The most important criterion that will make the agrochem industry to incorporate AGINFRA plus services to their daily operations is the reliability and sustainability of the provided services.

3.4 Market Description of agtech & foodtech SMEs powered by data science

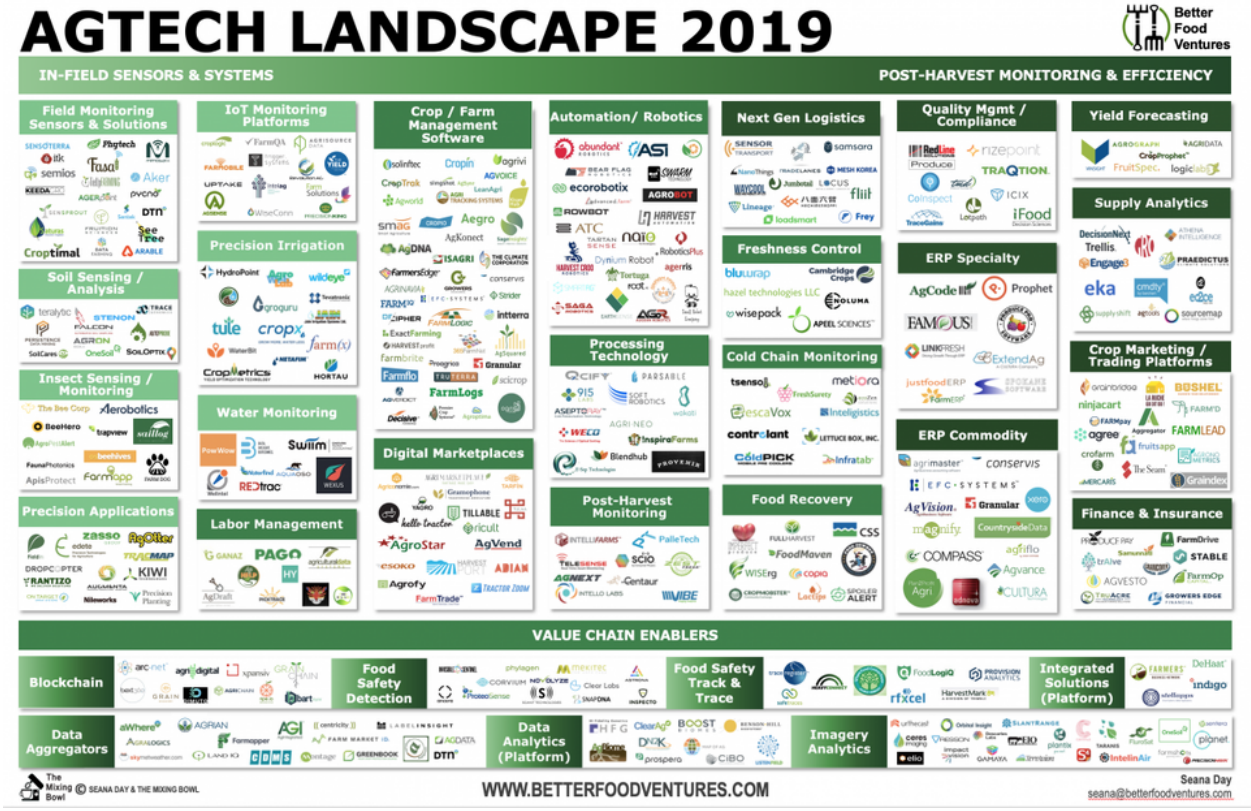

Figure 1: The Global AgTech Landscape (source: Seana Day, 2019 Agtech Landscape)

3.4.1 The market

The AgTech ecosystem has flourished during the last decade. Digital technology and data science have set the ground for a very promising future for food and agriculture ranging from predictive analytics to DNA sequencing mining, satellite image processing and sensor measurements analysis. From 2012 where the AgTech global ecosystem started to get significant funding traction (2.6 billion dollars), the annual global investment skyrocketed to 16,9 billion dollars in 2018, with 10% of the investments

concerning European companies. AgFunder[13] reports that the global agri-food market is estimated at 7.8 trillion dollars, while Food & Agriculture Technology and Products Market size is estimated at 494.9 billion dollars for 2018.

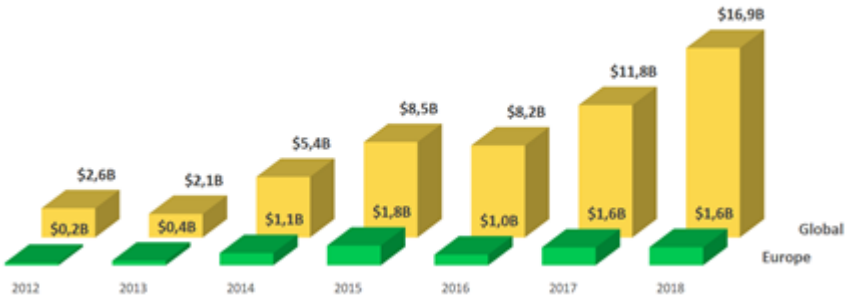

Figure 2: European and Global annual AgTech Investments (source: AgFunder, 2018)

The complexity of the agri-food supply chain requires demanding statistical analysis and computational power that will support all stakeholders (startups, SMEs, scaleups) and allow them to grow through data and scientific models.

3.4.2 The key characteristics of this market

According to AgFunder, the investments that took place since 2012, concerned over 1600 startups globally. And this is the tip of the iceberg, since the AgTech and FoodTech growth rate is exponential, the number of startups is constantly increasing. The type of companies that are a good fit for the AGINFRA PLUS offerings are the ones that have data processing and algorithm optimization needs through which the use of a powerful e-infrastructure will help them to advance their analytical and predictive capabilities.

3.4.3 The key attributes of the data science startups persona

The persona that is the key decision maker in this market sub-segment is most of the times the founder/ CEO or CTO of the company. The majority of these persona representatives are characterised by a computer science or data science background with former business or research experience in this field. They can understand data and technology very well and can translate their data challenges into actionable requirements from a technology solution provider as AGINFRA PLUS.

3.5 Relevant AGINFRA offerings

Table 4 below provides an overview of the AGINFRA PLUS offerings that we have decided to offer to these market segments and the respective partner that has agreed to take the responsibility to position each offering.

[13] AgFunder AgriFood Tech Investing Report - 2018 https://agfunder.com/research/agrifood-tech-investing-report-2018/

Table 4: An overview of the AGINFRA PLUS offerings

AGINFRA PLUS offering	To whom it could be relevant	Why it may add value	Who is the partner to position and offer it
D4Science-powered VREs	Could be relevant to all scientists, and more specifically to any community of practice willing to co-create its own Virtual Research Environment that by aggregating into a unifying whole the data and services of interest promise to deliver a state-of-the--art solution to collaborative (research) data[14] management challenges.	The value that scientists would get is (a) that they will spend their effort entirely on the research questions they are posing rather than distracting part of the effort for the development and deployment of the working environment supporting their research activity and (b) that scientists and researchers are provided with state-of-the-art solutions for implementing their scientific activity, solutions natively integrated with the rest of the open science ecosystem (e.g. through EOSC).	**CNR** (with EGI, Agroknow, and Pensoft for specific services)
EGI-powered Jupyter Notebooks	Could be relevant to all scientists, and more specifically to those researchers willing to share their ideas in the form of interactive notebooks that combine text, code and images.	The value that scientists would get is that they can interact programmatically with the rest of the AGINFRA PLUS offerings using their language of choice using Jupyter, which is becoming data scientists' computational notebook of choice.	**EGI**
EGI-powered Galaxy Workflows	Could be relevant to all scientists, and more specifically to those researchers that need an easy to use web based workflow management tool.	The value that scientists would get is that they are able to create and share reproducible workflows using that interact with the rest of the VREs capabilities.	**EGI**

[14] Research data is used here with the largest intent possible thus to include papers, data, software and any digital artifact stemming from a research activity.

Data Harvesting & Data Linking Services	Could be relevant to all scientists, especially multi-national projects that are continuing work performed in previous initiatives or other groups.	The value that scientists would get is that they can discover, connect and combine research outcomes of various formats and types produced from other groups and institutions , if they are published as open science.	**Agroknow**
ARPHA-powered Data Journals	Could be relevant to any scientist working in food science and viticulture. Scientists may be engaged with the public or private sector.	The value that scientists would get is that it is now possible to publish non-conventional but valuable outputs of the research cycle: models, data and software, Also, scientists can use a streamlined workflow for conversion of metadata describing a model (FKSX) into manuscripts in the ARPHA Writing Tool, that will be further edited, completed and submitted for peer review and publication.	**Pensoft**

3.6 Competition

3.6.1 D4Science-powered VREs

Virtual research environments are also known as Science Gateways in North America or Virtual Laboratories in Australia[15,16]. The great diversity of the resulting environments and application domains makes the comparison among tools and services supporting their development a challenging tasks. For instance, Shahand et al.[17] have identified eleven frameworks explicitly exploited to develop Science Gateway including Apache Airavata, Catania SG Gateway, Globus, HUBzero(+Pegasus), ICAT Job Portal, and WS-PGRADE/gUSE. These tools are very diverse from each other, in the majority of cases they are technologies borrowed from other cases to develop VREs rather than technologies specifically conceived to develop VREs. The recently launched inventory of gateways and available

[15] Candela et al. (2013) Virtual research environments: An overview and a research agenda. Data Science Journal 12, GRDI75–GRDI81 (2013). DOI 10.2481/dsj.GRDI-013

[16] Barker et al. (2019) The global impact of science gateways, virtual research environments and virtual laboratories. Future Generation Computer Systems, 95, 240-248 (2019). DOI 10.1016/j.future.2018.12.026

[17] Shahand et al. (2015) Science gateway canvas: A business reference model for science gateways, in: SCREAM '15 Proceedings of the 1st Workshop on The Science of Cyber-infrastructure: Research, Experience, Applications and Models, 2015. doi:10.1145/2753524.2753527.

software to create and support gateway services and functionality lists over 500 exemplars worldwide.[18]

The main competitors that offer similar services already in this market, are the following:

- HUBzero[19]: a platform for building websites (also known as science gateways) aiming at providing analytical tools and facilities to publish data, share resources, collaborate and build communities in a single web-based ecosystem.
- Open Science Framework (OSF)[20]: a web-based service enabling users to keep files, data, and protocols pertaining to any user defined project in a single, shared place.

The table that follows provides an outline of the features that we believe are important for this audience. We compare the AGINFRA PLUS offering vs the other main players in this space.

Table 5: An outline of the features that are important for D4Science-powered VREs audience

Name	Open Science-friendliness and flexibility	Services covering the whole research lifecycle	API for programmatic access and (re-)use	VRE specification-driven deployment	Service integration options
D4Science-powered VREs	Yes	Yes	Yes	Yes	Yes
HUBzero	Yes	Partially	Yes	No	Partially
OSF	Yes	Partially	Yes	No	No

AGINFRA PLUS unique value proposition offering against the competition is the richness, flexibility and open science-friendliness of the existing and potential VREs. Its competitive advantages include:

- A unifying web-based working environment where all the constituents looks like being part of a coherent whole;
- A rich array of off-the-shelf services covering all phases of a research lifecycle from data collection and collation through analytics up to publishing. Such services are actually platforms that VRE designated communities can easily customize and extend to integrate their community / domain-specific assets. E.g. the data analytics solution provides its users with facilities for easily integrating almost any legacy / existing analytics method implementation without requesting specific programming models;
- A set of REST-based APIs (whenever possible based on standards) enacting VRE users to securely interact with the specific VRE in a programmatic manner, by using clients other than the web browser. E.g. it is possible to manage every VRE workspace content by FUSE clients, it is possible to manage analytics methods processes by OGC WPS clients, it is possible to manage every VRE catalogue content via a simple REST client;

[18] https://catalog.sciencegateways.org/

[19] McLennan & Kennell (2010) HUBzero: A platform for dissemination and collaboration in computational science and engineer- ing, Computing in Science & Engineering 12 (2) (2010) 48–53. doi:10.1109/MCSE.2010.41

[20] https://osf.io

- A unique mechanism for VRE creation based on specification wizard enacting the designated communities to characterize the planned VRE from the functional perspective by selecting those available from a list. The actual deployment of a VRE is completely automatic and transparent.
- Several options to integrate existing services and applications into the unifying web-based working environments implemented by every VRE from several perspectives including authentication and authorization, content management, GUI;

3.6.2 EGI-powered Jupyter Notebooks

Project Jupyter[21] develops the Jupyter Notebooks as an Open Source software for interactive computing across dozens of programming languages. The software can be run locally at the user's computer but a server-based version is also available. The EGI-powered Jupyter Notebooks offers a fully-managed server so users do not need to care about the setup and maintenance of the software. Other competitors in the market offer similar solutions, including:

- Microsoft Azure Notebooks[22] offers a fully-managed notebooks service with limited CPU and RAM. It offers a git-based workspace and allows permanent installation of packages for the languages supported. There are no sharing capabilities.
- Google Collaboratory[23] offers a Google-drive backed notebooks service, which makes it easy to share and collaborate on single documents. There is limited computing capacity available for each notebook and there is no permanent installation packages (need to be re- installed on every session). Although based in Jupyter, the service presents a heavily customised interface that might be confusing for existing Jupyter users.
- Binder[24] is a notebooks service that allows to recreate computation environments as interactive notebook without any pre-requirement (e.g. creation of an account). While fully customisable it does not offer any kind of permanent storage for the notebooks so it's mostly useful for demos and testing environments..

The table that follows provides an outline of the features that we believe are important for this audience. We compare the AGINFRA PLUS offering vs the other main players in this space.

Table 6: An outline of the features that are important for EGI-powered Jupyter Notebooks audience

Name	Fully Managed Service	Permanent installation of libraries	Shared notebooks	New kernels	Custom computing backend
EGI-powered Jupyter Notebooks	Yes	Yes	Yes, via shared workspace	Yes	Yes
Microsoft Azure Notebooks	Yes	Yes	No	No	Partially
Google	Yes	No	Yes	No	Partially

[21] Project Jupyter: https://jupyter.org/
[22] https://notebooks.azure.com/
[23] https://colab.research.google.com/
[24] https://mybinder.org/

Collaboratory					
Binder	Yes	Partially	No	Yes	No

The AGINFRA PLUS offering unique value proposition against the competition is the integrated environment with the rest of the VRE with highly customisable options but still using the familiar Jupyter Notebooks experience. Its main competitive advantages are listed below:

- Full Jupyter Notebooks experience with the modern JupyterLab interface.
- Customizable kernels on demand. Besides the usual python, R and Julia languages, the AGINFRA PLUS offering can be customised to include any extra kernel needed to execute the users notebooks.
- Permanent storage and access to the workspace files. Users' files remain in the system after session ends and can be easily moved from/to the VRE workspace for sharing.
- Custom installation of libraries. The service allows permanent installation of libraries used for the notebooks, users can find their libraries without the need of reinstalling them for every session.
- User-defined computing capacity. Users can decide the computing capacity to run their notebooks. New hardware (e.g. GPUs) can be easily added to the offer as needed.

3.6.3 EGI-powered Galaxy Workflows

EGI offers a Galaxy-based service for workflow management and execution integrated with the VRE features (no need to relogin and interaction with the DataMiner analytics platform). There are several open Galaxy services targeting generic users that offer limited capacity and little customisation options to users. Other Galaxy services are usually heavily customised for specific communities and are not available for direct use by the AGINFRA PLUS users. The main generic Galaxy servers open to any user are:

- Use Galaxy[25], and
- Use Galaxy Europe[26]

Both offer a similar setup for users to discover Galaxy features with limited capacity.

The table that follows provides an outline of the features that we believe are important for this audience. We compare the AGINFRA PLUS offering vs the other main players in this space.

Table 7: An outline of the features that are important for EGI-powered Galaxy Workflows audience

Name	Galaxy web interface	VRE integration	Customisable
EGI-powered Galaxy workflows	Yes	Yes	Yes
Use Galaxy instances	Yes	No	No

The AGINFRA PLUS Galaxy offers a tailored solution to meet the needs of the VRE users, its main competitive advantages are:

[25] https://usegalaxy.org/
[26] https://usegalaxy.eu/

- Integration with the VRE: users do not need to log into another external service and all of the DataMiner algorithms are available as native tools in the Galaxy environment so users can easily re-use existing analytics tools into their workflows.
- Customisable computing backend, the computing resources allocated to users can be expanded as needed to cover the workflow execution demands
- Customisable tools and configuration: the users can request any Galaxy modification or installation of custom tools for incorporating them into workflows.

3.6.4 Data Harvesting Workflows & Data Linking Services

The main competitors to AGINFRA Data Harvesting & Linking Services are mainly providers of data management solutions (Software as a Service - SaaS) and data delivery services (Data as a Service - DaaS) over a variety of different domains and clients. Naturally, the competition landscape could feature numerous candidates, but almost none focuses purely on agri-food researchers as a target audience. Therefore, to perform a competition analysis, a number of popular generic cases has been picked, usually offering domain-agnostic solutions that inherently need additional adaptation to fit potential scientific use-cases in the agri-food domain.

The identified competitors are the following:

- The **PoolParty Semantic Suite**[27] is a set of tools for knowledge management, data analytics and content organisation, based on principles introduced by the Semantic Web movement. PoolParty offers out-of-the-box tools for ontology and vocabulary engineering, as well as concept tagging and term extraction capabilities. By using the suite, users may connect their data sources, add context to their data by linking it to standard ontologies and vocabularies and making it semantically discoverable.

  It is important to note that PoolParty is not particularly directed towards researchers or the agri-food sector. The suite becomes only as context-aware as the level of the knowledge management performed by its users allows it. As a result, one disadvantage of the suite is the requirement from its users to adapt to knowledge engineering routines that might burden their daily activity. Arguably this is the reason why PoolParty has yet to become a mainstream product to non-expert users who are not familiar with the concepts introduced by the Semantic Web.
- **HorizonScan**[28] is a software and data platform that provides food industry professionals and scientists relevant food safety information for their direct or indirect supply chain. As its name suggests, HorizonScan scans more than 100 food safety agencies and other sources daily, collecting data on emerging food safety issues, providing its customers with the appropriate notifications and alerts. Its business model relies heavily on the collection of relevant and high-quality agri-food data from trusted sources globally.

  Because of its commercial nature, HorizonScan is strictly focused on a fixed amount of data types that are tailored and offered to the end-users based on their preferences. Although data is generally searchable and linked to specific taxonomic classifications (such as hazards or products), no clear links to well-known thesauri (such as FoodOn) seem to be present, thus not tapping to any additional semantic capabilities and opportunities for integration with other systems.
- **LexisNexis Data as a Service**[29] is an API-based platform that collects, enriches and delivers the relevant, archival and current data, that organizations require for a broad range of data

[27] https://www.poolparty.biz/product-overview
[28] https://www.foodchainid.com/technical-services/horizonscan/
[29] https://internationalsales.lexisnexis.com/nexis-daas

research initiatives. Offering a range of easy-to-integrate APIs, it covers data from more than 80.000 sources, to support a variety of different use-cases.

Although its promising numbers, LexisNexis DaaS does not seem to provide enough support for research data types, mostly dealing in historical data records for machine learning algorithms training[30]. Furthermore, there is no indication that the gathered data are mapped to any specific domain vocabularies or ontologies. Last but not least, an option for deeper discovery of the federated data sources seems to be outside of the provided offerings.

The table that follows provides an outline of the features that we believe are important for this audience. We compare the AGINFRA PLUS offering vs the other main players in this space.

Table 8: An outline of the features that are important for agri-food researchers audience

Name	Data harvesting from external sources	Data linking capabilities to agri-food semantic resources	Support for multiple research data types	Semantic discovery of federated data
AGINFRA Data Harvesting & Linking Services	Yes	Yes	Yes	Yes
PoolParty	No	Potentially	Yes	Yes
HorizonScan	Yes	No	Partially	Partially
LexisNexis DaaS	Yes	No	Partially	No

Compared to the rest of the competition, the AGINFRA Data Harvesting & Linking Services can offer a unique value proposition:

1. It is fully domain-related. All gathered metadata records refer to agri-food research assets that have been published to widely acknowledged sources and tagged with the appropriate descriptors.
2. Data sources are trusted, hand-picked and curated to fit into the AGINFRA Map of the Data Ecosystem.
3. AGINFRA PLUS helped us develop powerful workflow-based data harvesting capabilities. Based on a state-of-the-art data pipelining mechanism, it can be adapted to draw data from virtually any open repository and integrate it seamlessly to its back end. From the identified competition, seemingly no candidate features such ease of integration of new data sources.
4. It features a flexible metadata schema which can be adapted to numerous data typologies, thus resolving the need for time-consuming knowledge organization and metadata management.
5. All gathered metadata records are linked to standard ontologies and vocabularies (such as GACS or AGROVOC) and can be thus easily integrated with other similarly tagged assets or be re-used by relevant systems. If no semantic links exist, Agroknow's Workflows & Services can perform data linking on the spot, thus resolving interoperability issues automatically.

[30] https://internationalsales.lexisnexis.com/nexis-daas/academic-research

6. AGINFRA Data Harvesting & Linking Services are agnostic to any knowledge management tooling in-place. In the context of AGINFRA PLUS, tools in the Ontological Engineering layer were functioning independently from the Platform itself, but it would frequently harvest their results as newly-added semantic resources. This approach leverages the highly flexible Data Harvesting Workflows, while still allowing the users to work with the tools that they feel comfortable with on the knowledge management perspective of the data.
7. All data are easily discoverable through the appropriate web services. To get results, users need to perform queries on the underlying back-end, instead of obtaining access to chunks of stacked, uncategorized records provided by typical web services.
8. Users may perform faceted search and data aggregations on the spot. Any user may instantly know important statistics about the data they are interested in obtaining, with the execution of a simple query.
9. Semantic search is offered out-of-the-box on top of the federated metadata, with options for configuring the semantic expansion it performs, thus making it applicable in a variety of different use-cases. With this embedded feature in place, data discoverability is heightened, while other novel features (such as query disambiguation) greatly boost the user experience.
10. They can be offered as an API- and subscription-based service that does not require any pre-existing or new software installations in the consumer's desktop. As it primarily delivers data (not functionality), it can be easily integrated with any external tool or service with minimum programmatic efforts.

3.6.5 ARPHA-powered Data Journals

Top competitors to ARPHA are those publishing workflows that use manuscript authoring tools as gates for writing, editing and submission of manuscripts for further review and publication. To the best of our knowledge, there are two such tools on the market: Overleaf[31] and Authorea[32]. Both started as small independent companies but are currently in possession of large groups, namely Digital Science (Overleaf) and Atypon/Wiley (Authorea).

The main competitors that offer similar services already in this market, are the following:

- Overleaf is based on an online collaborative solution of the LaTeX text processing tool. Originally, LaTeX has been created to serve mostly scientists from physical and mathematical sciences. The tool has rich editing functionalities for formulas and math symbols. A big advantage of the system is that the PDF is generated directly from the tool in a format which can be modified to a ready-to-publish state. Overleaf offers a large number of pre-defined templates that serve various purposes - articles, preprints, institutional documents, theses and so on. A disadvantage of the system is that it is not integrated with journal submission systems through XML-delivery of content but rather with delivery in file formats. Another significant disadvantage is the slightly over-complicated editing tool which requires a basic knowledge in coding. Currently, overleaf is part of the portfolio of Digital Science company[33].
- Authorea is a collaborative online authoring tool of many strengths (a concise description of services at: https://www.authorea.com/product). Initially, it has been designed as an independent writing tool that provides templates for various journals and still can be used as such. Recently, Authorea became a part of the Atypon journal submission system.

The table below provides an outline of the features that we believe are important for this audience.

[31] https://www.overleaf.com/
[32] https://www.authorea.com/
[33] https://www.digital-science.com/

Table 9: An outline of the features that are important for ARPHA-powered Data Journals audience.

Name	Writing environment in HTML	Full-text XML integration with journal submission systems	Conversion of structured XML files into manuscripts
ARPHA Platform	Yes	Yes	Yes
Overleaf	No	No	No
Authorea	Yes	Partially	No

The unique value proposition of the AGINFRA PLUS offering against the competition (see also: Filter M et al. 2019)[34] are:

1. The ARPHA Writing Tool provides a collaborative manuscript authoring environment and a set of pre-defined, but flexible article templates covering most types of research outcomes.
2. Within the ARPHA Writing Tool, the co-authors may work collaboratively on a manuscript, but can also invite external contributors, such as mentors, pre-submission reviewers, linguistic and copy editors or just colleagues, who may correct and comment on the manuscript before submission. These external contributors are not included amongst the co-authors of the manuscript.
3. A rich set of functionalities of the ARPHA Writing Tool allows for search and import of literature and data references, cross-referencing of in-text citations, import of tables, upload of images and multimedia, building plates of images and many more.
4. An automated technical validation step will save authors' and editors' time by checking the manuscript for consistency, in addition to a human-provided, pre-submission, technical validation by the FMJ's Editorial Office.
5. On choice of the authors, pre-submission external peer-review(s) can still be performed during the authoring process in the ARPHA Writing Tool. Pre-submission reviews can be submitted together with the manuscript to speed up post-submission evaluation and publication.
6. The collaborative peer-review process provides an easy communication environment through change tracking, comments and replies and automated, but customisable email and social network notifications.
7. For the convenience of editors, peer reviews in ARPHA are automatically consolidated into a single online file that makes the editorial process straightforward, easy and pleasant.
8. Published papers can be commented on via both the inbuilt ARPHA commenting tool and an integrated hypothes.is plugin and can also be a subject of an open, post-publication peer review.
9. Authors can convert published papers back into editing mode in the ARPHA Writing Tool at the click of a button. The manuscript can then be revised and re-published in a new version under different DOI, linked to previous versions via CrossMark, realising in this way the concept of a "living article".
10. Last but not least, the users of food model repositories may automatically convert **FSK-ML (Food Safety Knowledge Markup Language)** metadata describing their models into

[34] Filter M, Candela L, Guillier L, Nauta M, Georgiev T, Stoev P, Penev L (2019) Open Science meets Food Modelling: Introducing the Food Modelling Journal (FMJ). Food Modelling Journal 1: e46561. https://doi.org/10.3897/fmj.1.46561

manuscripts in the ARPHA Writing Tool, which can then be further elaborated, submitted, peer reviewed and published in the Food Modelling Journal

3.7 Marketing & Promotion Strategy

3.7.1 Branding of AGINFRA PLUS bundle of offerings

For public sector scientists, especially in Europe where funding primarily comes from EC projects, partners will offer these services as part of the EOSC marketplace and service registry. The AGINFRA brand can also be used to jointly promote software bundles and supporting services to scientists across Europe and the world. The D4Science VREs are also a recognisable brand that will be offered as a distinct service to new scientific communities, bundling together the services of other partners that are already included and interoperable (if the potential users wish to include/activate them). Standalone services like ARPHA will also be offered directly from the organisations responsible.

3.7.2 How to create user awareness & new marketing leads

The AGINFRA brand will be used as a quality mark that the services have already been tested by large numbers of food and agriculture scientists. Agroknow will continue to maintain and operate the AGINFRA.eu web site through its Data Consulting Unit (to be described later on) and its thought leadership marketing activities. The existing channels of AGINFRA PLUS will be used to communicate news about the evolution of the services, their adoption from new users, and new feature releases.

Apart from that, each one of the partners will continue to offer the corresponding technology offerings that they are responsible for in their own channels and marketing actions. For instance, Pensoft will use the two Data Journals as case studies that will attract more scientists from the food and agriculture communities to become editors and authors; EGI will further promote, support and engage contributors to the Galaxy and Jupyter Notebook groups; etc.

4. OUR DELIVERY

4.1 Management & Operations

4.1.1. Operations of D4Science-powered VREs

D4Science-powered VREs (CNR) are existing or new Virtual Research Environments (VREs) that specific communities or organisations may continue using, over the D4Science e-infrastructure.

During the AGINFRA PLUS project lifetime a total of 16 Virtual Research Environments have been defined, deployed and operated. Existing VREs are diverse. Some VREs have been created to support the project activities (e.g. AGINFRA PLUS to enact intra-project communication and collaboration, AGINFRA PLUSDev to offer a sandbox to develop, test and run new features) while others have been conceived to serve specific communities and scenarios (e.g. AgroClimaticModeling, RAKIP, FoodSecurity). Among the VREs serving specific communities there are those specifically associated with the three agri-food use cases targeted in the project as well as those serving other communities, communities expressing their interest during the AGINFRA PLUS operation (e.g. ORIONKnowedgeHub, FoodBorneOutbreak, EMPHASIS, NitrogenScrum). Moreover, the AGINFRA4BioCos VRE has been created to provide the BioCos SME team with a dedicated environment to develop and test their analytics workflow into the AGINFRA platform and the FMJ_lab has been created to support the Food Modeling Journal community in creating and accessing the innovative papers supported by the journal. Apart from the VREs supporting the project activities, the communities exploiting the rest of the VREs already expressed their willingness to continue using and developing the so-far created environments as well as creating new ones. Existing VREs will be maintained for two years after the project ends and this period will be used to set up specific service level agreements. This approach proved to be successfully in other contexts having the exploitation of D4Science-powered VRE as a topic, e.g. in the context of the BlueBRIDGE project an SLA for the maintenance of 9 VREs was signed between FAO and D4Science.
Moreover, the achievements and offering of AGINFRA PLUS VREs is raising interest in other communities and contexts through D4Science, e.g. the BlueCloud community is requesting VREs equipped with both the usual analytics and Jupyter, FAO and BlueCloud community are interested in the offering of VREs for supporting the production of innovative open access publications.

4.1.2 The Team

The D4Science team (https://www.d4science.org/team) comprise several members with diverse profiles ranging from managerial profiles to development and operation profiles. A key role in VRE management is played by the Research, Development and Operation team called to deal with the management of all activities of research, design, development, testing, release and operation of infrastructure services. In particular, the VRE Manager is called to coordinate the tasks governing VRE development yet the delivery and operations of VREs is really a collaborative activity involving the team as a whole.

4.1.3 Structure and Management Style

The D4Science team is responsible for the development and operation of the D4Science infrastructure.
The interaction with potential communities of practices willing to use D4Science powered VREs follows a typical customer relationship management procedure including requirements gathering and analysis, piloting and co-design activities and transition to production with establishments of SLA between the service provider (i.e. D4Science) and the specific customer. When there is the need to

include a third-party service into a VRE, there are two options (a) D4Science established an Operation Level Agreement with the specific service provider, (b) the SLA between the potential customer and D4Science is transformed into an SLA between the customer and all the service providers.

4.1.4 Staff Recruiting

The D4Science team is staffed with CNR-ISTI members with some IT collaborators from an SME. New hires are planned in the coming years to reinforce the D4Science Research, Development and Operation team for the expansion of the services.

4.1.5 Equipment and Facilities

D4Science services can run on resources provided by several providers. In the current settings, D4Science services are operated by counting of resources from four sites:

- Pisa, operated by the Networked Multimedia Information Systems Laboratory of the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" of the National Research Council of Italy;
- Athens, operated by the Management of Data, Information, and Knowledge Group of the Department of Informatics & Telecommunications of the University of Athens and by the Communication & Information Technologies Experts S.A.;
- GARR, the Italian NREN operated by a non-profit association whose main objective is to provide high-performance connectivity and to develop innovative services for the daily activities of researchers, professors and students as well as for international collaboration.
- The FedCloud provided by EGI with the support of CESGA, GeoGRID, IISAS-FedCloud, RECAS-BARI and UPV-GRyCAP.

4.1.6 Key Partnerships

D4Science is an infrastructure mainly developed with the support received by the European Commission via several projects including DILIGENT, D4Science, D4ScienceII. During the years the infrastructure has been exploited to support several communities: currently there are 19 gateways and more than 200 Virtual Research Environments. Some of these environments are serving research infrastructures like SoBigData, RISIS, Parthenos.

Some of the D4Science services and VREs are exploited by initiatives contributing to the development of the European Open Science Cloud, e.g. EOSC-Pillar and EOSCSecretariat.

Among the customers of D4Science services there are institutions like the FAO, the Aquatic Sciences and Fisheries Abstracts (ASFA), the Istituto Nazionale di Geofisica e Vulcanologia.

4.1.7 IPR Management

The software underlying D4Science services is mainly based on the gCube system, an open source software released with EUPL license. Such software has been largely developed by CNR-ISTI members.

4.1.8 Roadmap after AGINFRA PLUS

There are three key milestones to achieve after the end of AGINFRA PLUS for the services:

- **Milestone A [deadline: 12/2021]:** establish a series of SLAs with the designated communities of existing VREs to guarantee their operation for the future;
- **Milestone B [deadline: 12/2020]:** establish a collaboration agreement with EGI to add to the D4Science service portfolio Virtual Research Environments with Jupyter and/or Galaxy;

- **Milestone C [deadline: 06/2020]:** establish a collaboration agreement with Pensoft and BfR for the sustainability of the Virtual Research Environment specifically conceived to support the FMJ Journal;

4.1.9 Risk Assessment

Table 10: The possible risks for operating D4Science-powered VREs

Risk	Probability	Impact	Contingency Plans
Technology obsolescence	Low	High	Regular review of technology development plans and proactively adapt to use new technologies
Capacity of the services is not enough to meet users demand	Medium	Medium	Regular review of capacity plans and proactively adapt to the needs
Low uptake from communities due to lack of fit for their needs	Low	High	Early engagement with agri-food users to understand and adapt to their requirements

4.2 Operations of EGI-powered Jupyter Notebooks and Galaxy Workflows

Jupyter Notebooks and **Galaxy Workflows** are already integrated in the D4Science VREs, but they can also be offered to scientists as stand-alone applications. They allow for the execution of various computationally demanding tasks over the EGI e-infrastructure.

The services were extended to fit the needs of the AGINFRA PLUS users with:

- Custom kernels and libraries for the Notebooks to cover the requirements of the different VRE communities including data science, visualization and earth observation tools.
- Custom selection of execution environment (up to 4CPUs and 8GB RAM per user)
- Custom libraries for interaction with the DataMiner algorithms and the integration within Galaxy workflows
- Access to the workspace files from the Notebooks environment
- Notebooks visualization engine for facilitating the quick sharing of research ideas

Notebooks is a service that has raised interest from several communities during the different dissemination events and has been requested by several communities since its publication in the EGI marketplace during 2019. EGI expects the demand to grow in the next years, both as a stand-alone service as part of a D4Science VRE offering for communities in the agri-food sector and beyond.

4.2.1 The Team

EGI services are developed and delivered by a team that includes the following roles:

- Service Owner: in charge of the definition of the services that acts as the primary contact point for the service.
- Service Developer: experienced software engineer with a background in Python development and responsible for developing any extensions to the existing Jupyter and Galaxy Open Source projects to meet the needs of AGINFRA PLUS or other communities.
- Service Operator: experienced engineer managing the deployment and operation of the service in production and ensuring the continuity, availability of the service and planning the capacity needs to serve the users of the service.
- Customer Relationship Manager: acts as the primary contact point for a specific customer and maintains the relationship with that customer by regular communication.

4.2.2 Structure and Management Style

EGI Foundation coordinates the EGI Federation and defines and manages the operations of the EGI services. Within EGI Foundation, the Services, Solutions and Support Department defines technical roadmap of the services, manages and oversees service delivery, and engages with the customers. Roles defined above for the team are allocated to the Services, Solutions and Support Department.

EGI Foundation follows an Integrated Management System that is certified ISO 9001:2015 (Quality Management System) and ISO/IEC 20000-1:2011 (IT Service Management System). The EGI Foundation has also most of their staff members that are certified at the expert level for the FitSM standard (http://fitsm.itemo.org/) and has contributed to the development of professional IT service management skills in the Research e-Infrastructure community by delivering a large number of trainings.

New customers are onboarded following a CRM procedure that includes requirements gathering and analysis, piloting and co-design activities and transition to production with establishments of SLA between EGI Foundation and the customers. EGI engages with federation members via OLAs that support the operation of the service towards the customer.

4.2.3 Staff Recruiting

All the roles defined above are staffed with the current team of EGI Foundation. The team is capable of handling the Notebooks and Galaxy services for existing customers. New hires are planned in 2020 for the Services, Solutions and Support Department team that will assist with the Service Developer and Customer Relationship Management roles for the expansion of the services in the future if needed.

4.2.4 Equipment and Facilities

The service can run on any cloud provider supported by Kubernetes. EGI has an OLA with two of the federation members to support the operation of the service: INFN-CATANIA-STACK in Catania (Italy) and CESGA in Santiago de Compostela (Spain). Jointly these providers bring 38 virtual CPUs, 86 GB of RAM and 1 TB of storage, capable of handling approximately 80 concurrent users to the service. The OLAs are renewed yearly with the possibility to expand them following the user demand. Peaks of demand can be covered by engaging with new providers of the federation or with commercial providers.

4.2.5 Key Partnerships

The EGI federation is the largest distributed computing infrastructure in the world and brings together hundreds of data centres worldwide and also includes the largest community cloud federation in Europe with tends of cloud providers across most of the European countries offering IaaS cloud and storage services. EGI expanded the federation of its facilities with other non-European digital infrastructures in North America, South America, Africa-Arabia and the Asia-Pacific region, as such EGI fully realise the "Open to the World" vision.

In the context of EOSC, the EGI Foundation is one of the main contributors to the design and definition of the EOSC architecture and the federated service management framework. EGI.eu is the EOSC-hub project coordinator, and responsible for operating and provisioning a pan-European research cloud infrastructure to EOSC.

The development of Jupyter Notebooks and Galaxy services build on partnerships and collaboration of the EGI Foundation with the Open Source communities that deliver the software and EU projects and initiatives that contribute to those communities (OpenDreamKit[35], ELIXIR[36], PaNOSC[37] and others).

4.2.6 IPR Management

The services build on top of the existing software from Project Jupyter[38] (Modified BSD License) and Galaxy Project[39] (Academic Free License version 3.0). EGI extensions to the service follow an MIT or Apache 2.0 License and are available at the following GitHub repositories:

- https://github.com/enolfc/d4science-galaxy-authn
- https://github.com/enolfc/galaxy-dataminer
- https://github.com/enolfc/galaxy-k8s-d4science
- https://github.com/enolfc/d4scienceauth
- https://github.com/egi-foundation/notebooks

The services are delivered to AGINFRA PLUS partners via an SLA established with CNR for supporting the use of EGI services within the D4Science platform. This agreement is renewed on a yearly basis (next renewal expected on January 1st, 2021) and is provided without any charge. In the case of need, the SLA can be expanded at any time and include both pay-per-use or sponsored access (free at the point of use and costs covered by other EGI sources of funds).

4.2.7 Roadmap after AGINFRA PLUS

These are the 3 key milestones to achieve after the end of AGINFRA PLUS for the services:

- **Milestone A [deadline: 06/2020]**: Enable reproducible execution of notebooks using Binder. Prototypes already available and current EGI resources sufficient to deliver the milestone.
- **Milestone B [deadline: 12/2020]**: Expand notebooks execution capabilities within the D4Science VREs (e.g. provide easy way to turn Notebooks into DataMiner algorithms). Resources ensured by SoBigData++ project (4 years, starting January 2020).
- **Milestone C [deadline: 06/2021]**: Introduction of Galaxy as a generally available service for EGI customers, including those from the agri-food community not already engaged via VREs. Resources ensured by EGI Foundation

[35] https://opendreamkit.org/
[36] https://elixir-europe.org/
[37] https://panosc.eu/
[38] https://jupyter.org/
[39] https://galaxyproject.org/

4.2.8. Risk Assessment

Table 11: The possible risks for operating EGI-powered Jupyter Notebooks and Galaxy Workflows

Risk	Probability	Impact	Contingency Plans
Capacity of the services is not enough to meet users demand	Medium	Medium	Regular review of capacity plans and proactively adapt to the needs
Low uptake from communities due to lack of fit for their needs	Low	High	Early engagement with agri-food users to understand and adapt to their requirements

4.3. Operations of Data Harvesting & Data Linking Services

For Agroknow, there are two types of operations related to the offering of data harvesting technology and services to the identified market segments:

One is related to the delivery of services through its Data Platform in the form of SaaS - what could also be described as Data (and analytics) as a Service (DaaS). This means that Agroknow's team is continuously discovering, scanning, collecting, translating, indexing, enriching and making available to third parties a large variety of scientific data. It includes both data types and sources covered during AGINFRA PLUS (such as scientific publications, open data sets, scientific models, etc.). But also other data types and sources of relevance to the private sector scientists (such as legislation, lab tests, open data from governmental authorities and agencies, open data from customs and other inspection authorities, etc.). Potential customers may request an instance or a combination of data that is indexed by the platform. They may also request an additional data source that is relevant to the ones covered by the platform.

The other is related to the development and delivery of customised versions of the data harvesting and linking services. It typically concerns the development of a data discovery, scanning, collection etc workflow that addresses data types other than the ones that the SaaS part of the business is serving. Or the deployment of a custom-made instance of the Data Platform to serve the needs of a particular scientific organisation or community.

In terms of supporting the AGINFRA PLUS demonstrators and partners, Agroknow has decided to maintain for free the AGINFRA Search API currently powering both the AGINFRA PLUS VREs and the front-end search engine of the AGINFRA web site. Depending on the needs of its clients or similar projects, an update or expansion of the data coverage of the search API will be made. For the time being, it is envisaged as a large-scale demonstrator of the capabilities of the data platform.

4.3.1 The Team

We have envisaged the team responsible for the sustainability of the project outcomes as a dedicated Data Consulting Unit, which will be also spun out as a new company. The roles foreseen for this team are the following:

- Managing Director of the Consulting Unit: senior engineer with expertise in similar projects and consulting projects. Required from Year 1 (2020).
- Back-end Software Engineer: experienced engineer with expertise in data management and processing technologies, to develop, deploy and maintain custom APIs and platform instances. Required from Year 2 (2021).
- Front-end Software engineer: mid- or junior-level software engineer that may support and contribute to the implementation of custom software projects. Required from Year 1 (2020) and expected to grow to 3 full time positions in Year 3 (2022).
- Head of Consulting Business Development: senior commercial executive responsible for approaching, recruiting and acquiring new projects. Required from Year 3 (2022).
- Project Manager: software engineer or business analyst responsible for coordinating communication with clients and managing the project implementation and delivery phases. Required from Year 3 (2022).

4.3.2 Structure and Management Style

Our previous experience has shown that such a Unit should not operate within Agroknow, as its consulting mindset and mode of operation can be quite confusing and distractful in a small team. We have decided to follow a different, more agile approach that is compatible to the new commercial strategy of Agroknow. More specifically, the Data Consulting Unit is going to be spun off as a separate company. A dedicated team will be allocated to it, led by a senior software engineer of Agroknow. It will operate as a trusted partner and with support/mentorship from the mother company, but its activities will be differentiated. Existing clients and projects that Agroknow currently serves will be transferred to this new company. During the first years of operation (2020-2021), the mother company will be sharing some key operation resources to help the new company set up. As the Consulting Unit, it will acquire its own resources and people in key functions such as Business Development, Marketing and Finance.

4.3.3 Staff Recruiting

The Data Consulting Unit will be staffed with an existing senior software engineer as its Director and will proceed to the hire of a junior Front-End Developer. According to the progress of sales and revenue, new hires will come to fill in the rest of the foreseen positions.

4.3.4 Equipment and Facilities

The Data Consulting Unit will be renting its own office space and lease or buy its initial digital and furniture equipment.

4.3.5 Key Partnerships

The Data Consulting spin off will capitalise on the strategic partnerships and collaborations that Agroknow has developed among the years. These include: partnerships with INRA and WUR that have flourished through initiatives such as eROSA and AGINFRA PLUS; international networks such as GODAN[40], GWPP and RDA IGAD[41]; and international collaborations with global leaders such as the University of Guelph[42] in Canada, Michigan State University[43] and the University of Minnesota[44] in the

[40] https://www.godan.info/

[41] https://www.rd-alliance.org/groups/agriculture-data-interest-group-igad.html

[42] https://www.uoguelph.ca/

[43] https://www.canr.msu.edu/

[44] https://ageconsearch.umn.edu/pages/?page=about

US, and the Agricultural Information Institute[45] of the Chinese Academy of Agricultural Sciences (CAAS).

4.3.6 IPR Management

In terms of software technology developed by Agroknow within AGINFRA PLUS, the company owns the IP of software code that it has been developed during the project. As described in the corresponding chapter, the majority of this code is also made available as open source with an appropriate software license. The operational instances of these services (such as the AGINFRA web portal and the Semantic Search service and respective API) are also offered for free and will continue to be supported by the Data Consulting spin off for at least 2 years after the end of AGINFRA PLUS. Parts of this code have also been included in the commercial Data Platform of Agroknow, powering services such as its FOODAKAI product.

A number of open source tools have also been customised, extended or integrated in the VREs. The IP of these tools belongs to their original IP owner.

4.3.7 Roadmap after AGINFRA PLUS

In order to be able to deploy the Data Consulting Unit as a separate company and team, the following key milestones will need to be achieved:

- **Milestone A [deadline: 2/2020]**: the existing contracts, projects and technology that are relevant to the new entity are selected and an agreement is drafted between Agroknow and the team to be spinned off.
- **Milestone B [deadline: 6/2020]**: the name, logo, web site and identity of the new entity are being prepared and the spin off is publicly announced.
- **Milestone C [deadline: 12/2020]**: all relevant new business is consistently channeled to the new entity so that it acquires the necessary critical mass of jobs and revenue so that it may continue to operate without support from Agroknow.

4.3.8 Risk Assessment

The table that follows highlights the main risks in successfully spinning off this service line of Agroknow as a new business.

Table 12: The main risks in successfully spinning off the service line of Agroknow as a new business

Risk	Probability	Impact	Contingency Plans
Lack of new sales and revenue for the new business unit	Low	High	Existing contracts and projects will be transferred to the new team, to bootstrap its establishment and financial viability. All relevant new business is already earmarked for this Unit.
Agreement between	Low	High	A baseline of the

[45] http://aii.caas.cn/

Agroknow and the new stalls due to negotiation challenges			agreement has already been agreed, as the new team will be taking over a line of business that Agroknow has decided to discontinue as part of its main activities (digital science data consulting).
Existing partners and clients do not wish to collaborate with a startup business	Medium	Medium	Senior Agroknow staff will serve at the advisory board of the new entity, to ensure continuity and support the team in making its first steps.

4.4 Operations of ARPHA-powered Data Journals

Pensoft's ARPHA Platform[46] emerged as the first-ever publishing platform, supporting the full life cycle of a manuscript, from authoring through submission, peer review, publication and dissemination, within a single, fully Web- and XML-based, online collaborative environment. The acronym ARPHA stands for "Authoring, Reviewing, Publishing, Hosting and Archiving" - all in one place, for the first time.

The platform promotes Open Science by increasing the use, transparency and efficiency throughout the research cycle. Further, ARPHA increases the proportion of structured text and data within a journal article's content, so as to allow for both human use and machine readability. ARPHA is offered as either a Pensoft co-publishing or white-label solution to users that want to publish their own data journals. Offered as Software as a Service (SaaS), ARPHA currently supports more than 50 open access journals[47].

'Open access' as a publishing model revolutionised the way scientists communicate their results and transformed the research process, making it collaborative rather than a primarily competitive endeavour. When compared, the outdated data-disseminating structures lacked reproducibility and reusability of research. The shift towards a free and immediate online availability of data — and the rights to use it — should not bypass the agricultural and food sciences. It is necessary to encourage the agricultural scientists to come together and work on assuring and improving food supply, quality and safety in our globalised and rapidly changing world. This is especially valid when considering that the agricultural sector is one of the largest contributors to greenhouse gas emissions and conversely, climate change affects agriculture throughout the world.

Sustainability in agriculture and food studies is now dependent on data-driven innovations. In order to discuss potential solutions and identify the opportunities emerging from cross-domain interactions amongst agriculture experts, hydrologists, dairy experts, aquaculture experts, and big data analytics experts, it is necessary to create an environment with free access to scientific knowledge, information and data.

[46] https://arphahub.com

[47] https://arphahub.com/about/browse_journals/

To provide knowledge sharing and knowledge transfer, as an European research hub and thematic aggregator AGINFRA PLUS catalogues and makes discoverable publications, data sets and software services developed by Horizon 2020 research projects on topics related to agriculture, food and the environment. AGINFRA PLUS aims to design and develop a scientific data infrastructure that will facilitate the development of policies and services that promote the sharing of data among agricultural scientists in a manner that develops trust within and amongst their communities.

In the spirit of Open Access in science, AGINFRA PLUS fosters two peer-reviewed journals: Food Modeling Journal[48] and Viticulture Data Journal[49] and The first one covers a wide range of topics, from genetic research, food safety of viticultural products to climate change adaptation of grapevine varieties through grape specific research. The latter is focused on mathematical models, datasets and software solutions in the area of food science. Research papers address food and feed security and safety, the competitiveness of the European agri-food industry and the sustainability of food production, processing and consumption.

To support its authors, AGINFRA PLUS and its partners have provided them with access to ARPHA's writing tool. It is recognised for its data publishing, machine–readability and dissemination of content solutions, and provides several innovative and unique features.

The platform implements various technologically advanced tools and workflows to ensure open data publishing and the right direction towards open science workflows.

One of its innovations lies in **creating linked content** - embedding machine-readable database records in each publication that are linked to the world-wide network of linked knowledge hubs.

That includes the Refindit reference tool, updated to work with the specialised agricultural information management database AGRIS.

Sector-oriented templates are additionally created within the ARPHA writing tool for the needs of open agricultural and food research.

All data published in Viticulture Data Journal and Food Modeling Journal can be downloaded in tabular format straight from the article text and re-used by anyone, provided that the original source is cited.

Upon publication, the primary agricultural data (for example, models, model validation studies, software, data analytics pipelines and visualisation methods) are also **automatically exported into machine-readable data**.

ARPHA provides solutions for the publishing challenges faced by agriculture and food sciences and enables a smooth transition from the conventional, document-based workflows to text and data mining, aggregation and re-use. During the AGINFRA PLUS project, ARPHA has been extended to be used from the AGINFRA PLUS Virtual Research Environment (VRE)[50], which would allow the authors to use the VRE as an additional gate to the AWT and the journal, as well as to benefit from the integration of AWT with several other services offered by the AGINFRA PLUS platform (Ballis et al.

[48] https://fmj.pensoft.net/
[49] https://vdj.pensoft.net
[50] https://aginfra.d4science.org/explore

2018[51]). The AGINFRA PLUS platform has been designed as a Gateway[52] providing online access through a one-stop endpoint to services (for list and description of services see: Ballis et al. 2018, Filter et al. 2019[53]), aiming at the integration of the traditional narrative of research articles with their underlying data, software code and workflows.

We expect a significant increase in the demand of open data and open science publishing in the coming years, due to several factors, mainly to the political decisions at EU level that led to the European Science Policy Platform[54]. The political decisions have recently resulted in the launch of Plan S by a coalition of several EU national science funding organisations (cOAlition S[55]), envisaging a full transformation to open access to publicly funded research from 2021 to 2024.

4.4.1 The Team

The team needed to maintain and develop the journals should include the following roles:

1. Technical supervisor
2. Software developer
3. Editorial supervisor
4. Managing editor
5. Copy editor
6. Typesetter

The level of involvement of the technical team will be decreasing (technical supervisor and software developer) in the course of advancement of the journals, while the involvement of the editorial staff will be increasing.

4.4.2 Structure and Management Style

Pensoft Publishers Ltd. (Pensoft) is a limited liability company (LTD) registered in Bulgaria. All activities are performed in the office of the company at: 12, Professor Georgi Zlatarski Street, 1700 Sofia, Bulgaria. The company's business includes advanced open access publishing, production, printing and marketing of academic journals and books, research and technological development of innovative software tools and platforms, dissemination and communication of scientific results, and project development. Currently Pensoft employs 30 people, organised in four departments: 'Publishing', 'Information Technologies (IT)', 'Projects and RTD', and 'Business Development and Management' (Fig 1.).

The **'Publishing'** department handles the overall editorial management and production of currently 18 open access academic journals. The department also deals with publishing of scientific books and e-books.

[51] Ballis A, Boizet A, Candela L, Castelli D, Fernandez E, Filter M, Gunther T, Kakaletris G, Karampiperis P, Katris D, Knapen MJ, Lokers RM, Penev L, Sipos G, Zervas P (2018) Serving Scientists in Agri-Food Area by Virtual Research Environments. 2018 IEEE 14th International Conference on e-Science (e-Science) https://doi.org/10.1109/escience.2018.00124

[52] http://aginfra.d4science.org/

[53] Filter M, Candela L, Guillier L, Nauta M, Georgiev T, Stoev P, Penev L (2019) Open Science meets Food Modelling: Introducing the Food Modelling Journal (FMJ). Food Modelling Journal 1 https://doi.org/10.3897/fmj.1.46561

[54] https://fmj.pensoft.net/article/46561/element/4/5004895//

[55] https://www.coalition-s.org/

The **'Information Technologies'** department deals with the technical implementation of innovative methods and technologies for publishing, and the system administration and maintenance of the entire e-infrastructure of the company. It also provides technical support to Pensoft's clients.

The **'Projects and RTD'** department coordinates research and innovations development and participation in various research projects, mainly funded by the previous FP6, FP7 and the current H2020 programs. Within EU projects, Pensoft specialises in dissemination, communication and publication, as well as in designing and maintaining websites and communication platforms, and providing analytical online tools.

The **"Business Development"** department has an overarching function at Pensoft, facilitating the work of and setting the strategies for development in all other departments. The main objective of this department is to ensure the right placement and promotion of the company's products on the targeted markets.

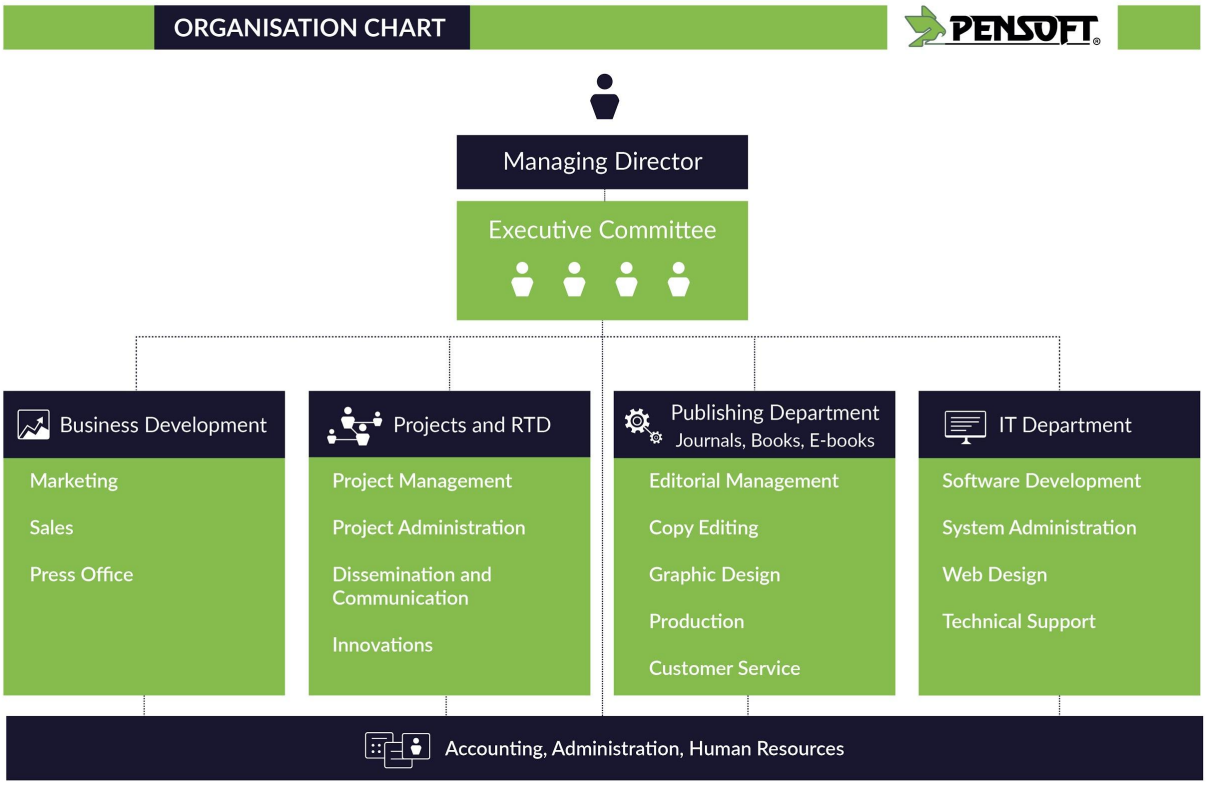

Figure 3: Management structure of Pensoft.

The marketing strategy, dissemination and PR are vital parts of the successful development of a new journal. Pensoft's press office offers a full-cycle of promotion for important scientific results. Press services start from support to authors in "translating" their results to a language that would provoke the interest of the general public, towards dissemination of press materials and consecutive liaison with journalists to ensure a wide and high-quality coverage as illustrated below.

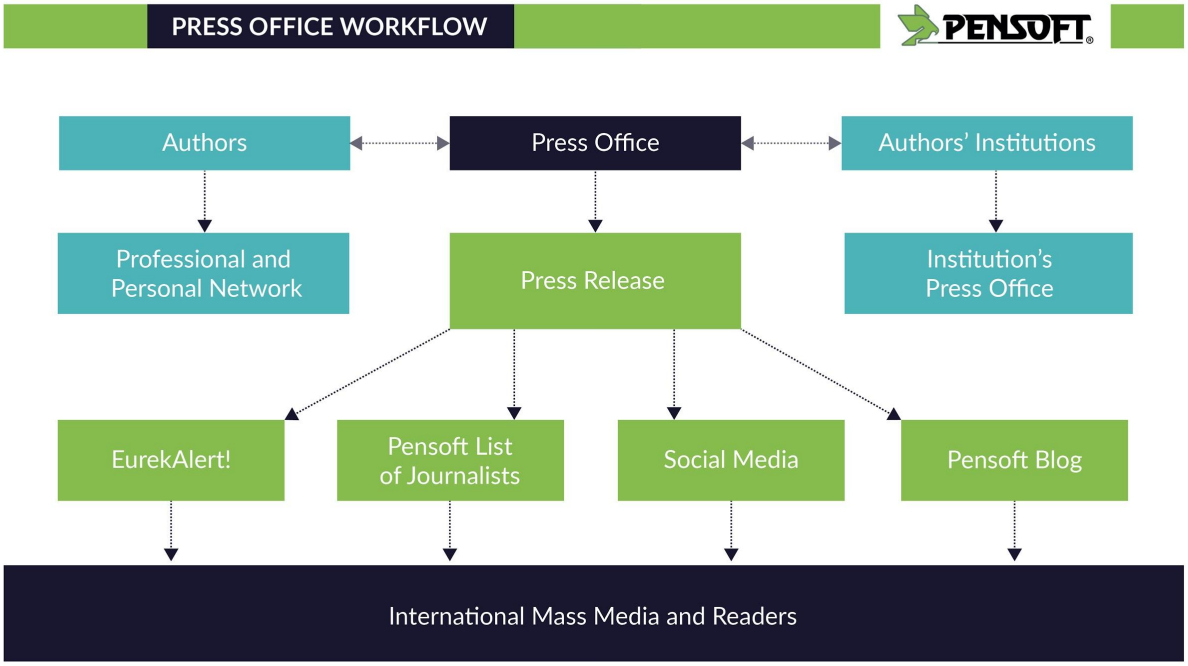

Figure 4: Workflow for creation and dissemination of press releases.

The figure illustrates the multi-channel dissemination approach of science news releases. Press materials are prepared by Pensoft's press office and then shared with the authors and their institutions for comments and ideas for further elaboration. When the campaign is launched the press office posts the press release via EurekAlert!, the biggest science news distributor worldwide, alongside a social media campaign and a blog post.

Press releases are also personally forwarded to our growing list of journalists who have indicated an interest to receive news from Pensoft's journals. At the next level of dissemination news stories are picked up by the mass media worldwide and featured in leading titles including BBC, CNN, DPA, National Geographic, Discovery, New York Times, Washington Post, The Guardian, Daily Mail, and more.

Pensoft is also actively using and developing various social media accounts (Facebook, Twitter, Google+) for its journals, benefitting from the wide outreach and interactivity offered by this communication environment.

4.4.3 Staff Recruiting

All roles listed above are available and currently occupied at Pensoft and will take responsibility for the journals as part of their everyday duties.

4.4.4 Equipment and Facilities

Pensoft has in its possession all equipment, facilities and software needed to successfully maintain and develop the new journals. These include own servers (both in-house file server and servers hosted in a data hosting centre), own submission, peer review, editorial, production, publication and

dissemination management system (ARPHA), typesetting and copy-editing software (InDesign, eXstyles).

4.4.5 Key Partnerships

FMJ has been launched after intensive preliminary consultations with various communities participating in the AGINFRA PLUS project and beyond and we expect to enjoy their support and encouragement after the expiration of the project:

- Agroknow, Athens, Greece
- Bundesinstitut für Risikobewertung, Berlin, Germany
- Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - Consiglio Nazionale delle Ricerche, Pisa, Italy
- Montpellier SupAgro – INRA (French National Institute for Agricultural Research Director of Mistea Laboratory, Montpellier, France

For FMJ, of special importance is the support of the Risk Assessment Modelling and Knowledge Integration Platform (RAKIP) community. RAKIP was established by three European institutions specialising in food safety risk assessment - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES)56, Bundestinstitut für Risikobewerting (BfR)57 and National Food Institute (DTU Food)58 - to provide an infrastructure for hosting, saving and preservation of "a variety of scientific knowledge (e.g. scientific publications, experimental data and mathematical models) and resources (databases and software tools for model generation and application), based on harmonized data formats and consistent rules for knowledge annotation" (FoodRisk Labs 201659). Now FMJ is expected to support this and many other community activities in the domain of food modelling by adding a scholarly publishing infrastructure that will facilitate the exchange of digital assets at high, community- and industry-accepted standards.

Of vital importance for the development of VDJ, besides the above mentioned AGINFRA PLUS partnering organisations, was the partnership with the Laboratory of Viticulture, Agricultural University of Athens, Athens, Greece. Also, support was provided by the staff of the Department of Sustainable Crop Production Fruit Culture and Viticulture Section Università Cattolica del Sacro Cuore, Piacenza, Italy, the Institute of Agrifood Research and Technology (IRTA), Barcelona, Spain and the University of Ljubljana, Ljubljana, Slovenia. We plan to extend the collaboration with currently ongoing and future international projects and communities in viticulture (for example) and sincerely hope that the viticultural research community will embrace the opportunity to publish their unconventional, though valuable, research outcomes in the Journal.

4.4.6 IPR Management

The code of ARPHA is property of Pensoft and the company will take the financial risk and rights for running the journals after the project expiration. The formats, mappings and plugins developed in the course of AGINFRA PLUS are free to use.

All articles published in the journals will be available in golden open access under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Data and software published in then journals will be available under three publishing licenses on choice:

[56] http://www.anses.fr/fr

[57] https://www.bfr.bund.de/de/start.html

[58] http://www.food.dtu.dk/english

[59] https://foodrisklabs.bfr.bund.de/rakip-web-portal/. Accessed on: 2019-2-05.

The default data publishing license used by the journals is the Open Data Commons Attribution License (ODC-By), which is a license agreement intended to allow users to freely share, modify, and use the published data(base), provided that the data creators are attributed (cited or acknowledged). As an alternative, the other licenses or waivers, namely the Creative Commons CC0 waiver (also cited as "CC-Zero" or "CC-zero") and the Open Data Commons Public Domain Dedication and Licence (PDDL), are also STRONGLY encouraged for use in Pensoft journals. According to the CC0 waiver, "the person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighbouring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission."

The journals will be free to publish for all AGINFRA PLUS partners for at least 2 (two) years after the expiration of the project.

4.4.7 Roadmap after AGINFRA PLUS

Here are the three key milestones in the development of the journals after the expiration of the project.

- **Milestone A [deadline: 8/2020]**: Results of the promotional campaign reviewed; financial and other resources ensured by Pensoft.
- **Milestone B [deadline: 12/2021]**: The journals become self-sustainable; financial and other resources ensured by Pensoft.
- **Milestone C [deadline: 06/2023]**: The journals bring profit which allows their further development and expansion on the market; financial and other resources ensured by Pensoft.

4.4.8 Risk Assessment

The possible risks and their mitigation measures are listed in the table below.

Table 13: The possible risks and their mitigation measures for Roadmap after AGINFRA PLUS

Risk	Probability	Impact	Contingency Plans
The mission, focus and scope of the data journal not sufficiently clear to the potential users	Low	High	Publication of clear and concise guidelines on the journals' website; descriptions of the non-conventional article types (such as data papers, software or model descriptions); explanatory leaflets; blogs.
The lack of impact factor	Medium	High	Publication of high-quality papers is a condition for successful application for indexing in Web of Science and Scopus; the uniqueness

			and innovative character of the journals will attract articles which otherwise won't be published in high-impact journals.
Low uptake by the community	Medium	High	Active and continuous promotional campaigns across the potential user communities

5. OUR FINANCIALS

5.1 Financial Projections of D4Science-powered VREs

CNR is focusing here on the financial scenarios related to the support and delivery of VRE opportunities resulting from AGINFRA PLUS and addressed to the food and agriculture communities. This is considered to be one part of its D4Science-related activities, as it is serving more communities and domains.

5.1.1 Business Model(s)

5.1.1.1 Revenue Streams

Two revenue streams are foreseen for CNR for this type of VREs:

1. Annual subscriptions to set up, support and maintain agri-food community VREs. Different service levels are expected.
2. Donor funding, mostly from D4Science in order to continue supporting and evolving the infrastructure for these particular communities.

5.1.1.2 Pricing Model

The pricing model for annual subscriptions is as it follows. It distinguishes among different types of service levels:

- Annual prices are different for existing vs. new VREs.
- Prices are different depending on the sophistication and requirements of the VRE, distinguishing between a Simple VRE service, a Community VRE service and a GPU-powered VRE service.
- The donor funding is expected as an annual flat amount.

Table 14: Pricing model for annual subscriptions

Pricing Table (per service, per year)	
A. Existing Community VRE as a Service (annual subscription)	**25.000 €**
B. New Community VRE as a Service (annual subscription)	**35.000 €**
C. Simple VRE as a Service (annual subscription)	**15.000 €**
D. GPU powered VRE as a Service (annual subscription)	**50.000 €**

5.1.2 Financial Forecasts

5.1.2.1 Revenue forecast

Using the provided spreadsheet (see Annex C), we have calculated a projection of the revenue of CNR from AGINFRA-related users.

Table 15: Projection of the revenue of CNR from AGINFRA-related users

Revenue stream	2020	2021	2022	2023
Annual subscriptions of various types	34.000€	96.000€	171.000€	271.000€
D4Science funding	15.000€	15.000€	15.000€	15.000€

5.1.2.2 Cost Analysis

Using the provided spreadsheet, we have calculated a projection of the annual costs foreseen in terms of personnel (calculating a percentage of the D4Science personnel to be allocated to these communities) and other expenses that are the most important cost centers for offering VREs to the targeted audiences.

Table 16: Projection of the annual costs foreseen for offering VREs to the targeted audiences.

Annual Costs	2020	2021	2022	2023
Personnel Costs	27.000€	26.000€	32.000€	32.000€
Cost of Software Delivery & Maintenance	8.000€	23.000€	23.000€	8.000€
Operational & Administrative Expenses	1.000€	1.000€	1.000€	1.000€

5.1.2.3 Profitability forecast

The table that follows uses the sales and cost projections presented earlier, to calculate the annual profitability forecast for the VRE services (minus some small adjustments for depreciation and other types of expenses).

Table 17: The annual profitability forecast for the VRE services

Profitability metrics	2020	2021	2022	2023
Annual Net Profit (before tax)	-2.000€	31.000€	100.000€	230.000€
% of Profitability	-3.6%	28%	53.45%	80.2 %

5.1.3 Funding Needs and Opportunities

The presented estimations foresee that the evolution and further extension of the D4Science VREs for the targeted communities will be mainly sustained by its annual subscriptions. A small loss is expected during the first year, that is going to be quickly recovered after the 2nd year. Although no further R&D financing is foreseen in this estimation, it is expected that additional extensions and integration capabilities will be developed through the participation in new EU-funded projects with users from the food and agriculture communities.

5.2 Financial Projections of EGI-powered Jupyter Notebooks and Galaxy Workflows

5.2.1 Business Model(s)

5.2.1.1 Revenue Streams

EGI services are offered using three different access methods:

- Wide access: free of charge with a limited set of resources (storage for notebooks/data and compute power) for each user
- Policy-based: by joining projects or initiatives that fund the service operation, those users related to the project are able to access the service free of charge
- Market-Driven: by establishing a contract between the users and EGI to pay for the service. While different mechanisms can be used for these contracts, an annual subscription is the preferred way in this case.

Revenue streams are expected for the Policy-based and Market-Driven access methods.

5.2.1.2 Pricing Model

Prices are determined on a case-by-case basis, with the main variable cost being the number of concurrent users to access the service on a given period of time. Wide-access offers limited capacity for free and policy-based and market-driven access offer custom and advanced features (e.g. more capacity) for a cost, following a freemium model. The table below summarizes prices for typical requests of the service:

Table 18: Prices for typical requests of the service

Pricing Table (per service, per year)	
A. Jupyter/Galaxy as a Service wide access (annual subscription)	**0 €**
B. Custom Jupyter/Galaxy as a Service (annual subscription, up to 25 concurrent users)	**20,000 €**
C.Custom Jupyter/Galaxy as a Service with GPUs (annual subscription, up to 25 concurrent users)	**40,000 €**

5.2.2 Financial Forecasts

5.2.2.1 Revenue forecast

Using the provided spreadsheet, we have calculated a projection of the sales revenue of the EGI Jupyter Notebooks and Galaxy, based on the current projects and an estimation of sales to market-driven customers until 2023.

Table 19: Projection of the sales revenue of the EGI Jupyter Notebooks and Galaxy

Revenue stream	2020	2021	2022	2023
Policy-access (EU funded projects)	80,000€	120,000€	120,000€	150,000€
Market-driven access	5,000€	10,000€	60,000€	70,000€

5.2.2.2 Cost Analysis

Using the provided spreadsheet, we have calculated a projection of the annual costs foreseen in terms of personnel and other expenses that are the most important cost centers for offering and evolving the services to the targeted audiences.

Table 20: Projection of the annual costs foreseen in terms of personnel and other expenses

Annual Costs	2020	2021	2022	2023
Personnel Costs	41,.000€	48,000€	68,000€	68,000€
Cost of Software Delivery & Maintenance	21,000€	31,000€	31,000€	41,000€

Operational & Administrative Expenses	9,000€	9,000€	9,000€	9,000€

5.2.2.3 Profitability forecast

The table that follows uses the sales and cost projections presented earlier, to calculate the annual profitability forecast for the Galaxy and Jupyter services (minus some small adjustments for depreciation and other types of expenses).

Table 21: The sales and cost projections to calculate the annual profitability forecast for the Galaxy and Jupyter services

Profitability metrics	2020	2021	2022	2023
Annual Net Profit (before tax)	-21,000€	4,000€.	33,000€	53,000€
% of Profitability	-24.99%	2.89%	18.11%	23.91%

5.2.3 Funding Needs and Opportunities

The presented estimations foresee that the evolution and further extension of the Galaxy and Jupyter services will be sustained by a mixed revenue scheme of R&D funding and access fees. Participation in EU-funded projects will ensure that this community project continues to develop through its community engagement and enlargement.

5.3 Financial Projections of Agroknow's Data Consulting Unit

The Data Consulting activities will be spun off as a new company that will be operating independently (but in close collaboration) with Agroknow.

5.3.1 Business Model(s)

5.3.1.1 Revenue Streams

The Data Consulting spin-off will be operating as a service business, implementing one-off or support contracts for its clients. It is expected to have two main revenue streams:

1. Private contracts that are assigned by an organisation for the implementation of an one-off task or the annual support of an existing system.
2. Participation in EU-funded or other donor-funded projects, in which a consortium undertakes the implementation of a research and/or innovation action.

5.3.1.2 Pricing Model

The initial pricing model for the private consulting contracts of the new company is as it follows. It distinguishes among different types of contracts (support vs one-off, public sector vs private) and it is based on the existing clients and contracts that the mother company is serving.

Table 22: Pricing model for the private consulting contracts

Pricing Table (per service, per year)	
A. Support contract w/ Public Sector (MSU, etc)	**20.000 €**
B. Support contract w/ Private Sector (Syngenta, etc)	**25.000 €**
C. One-off consulting contract w/ Public Sector	**10.000 €**
D. One-off consulting contract w/ Private Sector	**10.000 €**
E. Other service contracts & clients	**10.000 €**

In terms of donor-funded projects, the pricing is based on the actual hourly rate of the personnel expected to be engaged in each project. For instance, for EU-funded projects this rate is typically multiplied with a flat overheads factor (25%) and then the resulting amount is reimbursed to the organisation.

5.3.2 Financial Forecasts

5.3.2.1 Revenue forecast

Using the provided spreadsheet, we have calculated a projection of the sales revenue of the data consulting spin off, based on the current contracts of Agroknow and an estimation of future sales until 2023.

Table 23: Projection of the sales revenue of the data consulting spin of and an estimation of future sales until 2023.

Revenue stream	**2020**	**2021**	**2022**	**2023**
Consulting contracts of various types	41.000€	94.000€	158.000€	234.000€

Donor-funded projects (EU, Gates, other)	100.000€	150.000€	250.000€	350.000€

5.3.2.2 Cost Analysis

Using the provided spreadsheet, we have calculated a projection of the annual costs foreseen in terms of personnel and other expenses that are the most important cost centers for offering and evolving the data consulting services to the targeted audiences.

Table 24: Projection of the annual costs foreseen in terms of personnel and other expenses for offering and evolving the data consulting services to the targeted audiences.

Annual Costs	2020	2021	2022	2023
Personnel Costs	65.000€	120.000€	230.000€	390.000€
Cost of Software Delivery & Maintenance	15.000€	20.000€	20.000€	25.000€
Marketing & Sales Expenses	10.000€	25.000€	25.000€	40.000€
Operational & Administrative Expenses	15.000€	30.000€	30.000€	50.000€

5.3.2.3 Profitability forecast

The table that follows uses the sales and cost projections presented earlier, to calculate the annual profitability forecast for the data consulting business (minus some small adjustments for depreciation and other purposes).

Table 25: The sales and cost projections to calculate the annual profitability forecast for the data consulting business

Profitability metrics	2020	2021	2022	2023
Annual Net Profit (before tax)	19.000€	27.000€	36.000€	52.000€

% of Profitability	13.6%	11%	8.7%	8.9 %

5.3.3 Funding Needs and Opportunities

The presented estimations foresee that the Data Consulting spin off can be a profitable (therefore, sustainable) from its first year of independent operation. It will be based on Agroknow's existing contracts and clients to bootstrap the business; therefore, we do not expect that any additional capital investment will be required. A significant part of its revenue is going to be through the participation of donor-funded projects that European or international consortiums bid for. This is a successful part of the mother company which we will spin off, directing all new business opportunities to the new company. This includes existing H2020 opportunities, Gates-funded and DFID-funded opportunities, as well as forthcoming Horizon Europe projects.

5.4 Financial Projections of ARPHA-powered Data Journals

The Data Journals that are powered by the ARPHA platform are part of the regular publishing activities and services of Pensoft.

5.4.1 Business Model(s)

5.4.1.1 Revenue Streams

The standard business model that Pensoft follows for all its journals is also adopted for the Data Journals. Revenue is being generated by charging article processing fees to the publishing authors.

5.4.1.2 Pricing Model

The foreseen article processing charges are 450 euros per article published.

5.4.2 Financial Forecasts

5.4.2.1 Revenue forecast

Using the provided spreadsheet, we have calculated a projection of the sales revenue according to the number of articles that we expect to process and publish in our Data Journals.

Table 26: projection of the sales revenue according to the number of articles that we expect to process and publish in our Data Journals

Revenue stream	2020	2021	2022	2023
# of data papers processed per year	120	140	150	160

Revenue from article processing fees	54.000€	63.000€	68.000€	72.000€

5.4.2.2 Cost Analysis

Using the provided spreadsheet, we have calculated a projection of the annual costs foreseen in terms of personnel and other expenses that are the most important cost centers for supporting and maintaining the Data Journals. Costs are expected to decrease, as the provided services will become more routine in the portfolio of Pensoft.

Table 27: Projection of the annual costs foreseen in terms of personnel and other expenses for supporting and maintaining the Data Journals

Annual Costs	**2020**	**2021**	**2022**	**2023**
Personnel Costs	54.000€	41.000€	38.000€	40.000€
Cost of Software Delivery & Maintenance	2.000€	2.000€	2.000€	2.000€
Marketing & Sales Expenses	2.000€	1.000€	2.000€	1.000€
Operational & Administrative Expenses	2.000€	2.000€	2.000€	2.000€

5.4.2.3 Profitability forecast

The table that follows uses the sales and cost projections presented earlier, to calculate the annual profitability forecast for the Data Journals revenue line of Pensoft.

Table 28: The sales and cost projections to calculate the annual profitability forecast for the Data Journals revenue line of Pensoft

Profitability metrics	**2020**	**2021**	**2022**	**2023**
Annual Net Profit (before tax)	-14.000€	8.000€	14.000€	17.000€

% of Profitability	-25%	13,10%	20,6%	23,6%

5.4.3 Funding Needs and Opportunities

The presented estimations foresee that the Data Journals of Pensoft can become a profitable line of business after one year, as soon as more than 120 articles are being processed per year (i.e. 10 per month). Profitability is expected to be over 20% within 3 years of operation. No additional funds will be required, as the loss of the first year will be covered by the company itself.

ANNEX A: Initial exploitation table of technology outcomes, IPR & exploitation paths (as per DoA)

Initial version of the GA table (2.2.1.2 Exploitation of project results) with the exploitation path of each key technology outcome. An analysis of how this has evolved until M36 is included in a separate chapter at the beginning of this deliverable.

Product / Service	Partners	Protection or licencing actions	Targeted stakeholders / customers	Initially considered exploitation paths
AGINFRA data infrastructure & VRE				
AGINFRA Virtual Organisation (VO): federated grid & cloud resources to support agri-food research	EGI, CNR	Open software license such as Mozilla MPL53 or GNU GPL54 (equivalent to CCBY55 & more appropriate for software)	Research infrastructure stakeholders	EGI Competence Center to sustain & further develop Future in kind contributions in cloud & grid resources as part of EOSC Future public/private donor financing
AGINFRA federated registry of data & software (CIARD RING)	Agroknow	Open software license such as Mozilla MPL or GNU GPL	Technology providers Data providers Research infrastructure stakeholders	Free service for all users sustained by Agroknow (in collaboration with FAO) Future public/private donor financing
AGINFRA federated registry of metadata schemas, ontologies, and classifications (VEST registry)	Agroknow	Open software license such as Mozilla MPL or GNU GPL	Technology providers Data providers	Free service for all users sustained by Agroknow (in collaboration with FAO) Planned integration with Agro LD portal of INRA, in the context of

				harmonised AgriSemantics service Future public/private donor financing
Datasets and dataset metadata produced as part of the AGINFRA registries	Agroknow, INRA	Open Data License (ODL) or CC-BY-ND	Data providers and consumers Technology & application providers Targeted end users	Data publicly available for reuse and exploitation by third parties Application development (third party VREs). End user services exploiting AGINFRA data.
AGINFRA data aggregation, indexing, processing components & workflows	Agroknow	Open software license such as Mozilla MPL or GNU GPL	Data providers	Free service for all users sustained by dedicated not-for-profit legal entity (where key stakeholders will be members). Membership fee for legal entity sustainability and portal hosting expenses. In kind contributions for provision of services and basic operation needs. Future public/private donor financing for further evolution and extension.

AGINFRA Open Science Data analytics components	CNR	Open software license such as Mozilla MPL or GNU GPL	Data providers Technology providers	Free service for all users sustained by CNR Following evolution and support of gCube Data Analytics of D4Science Future public/private donor financing for further evolution and extension within AGINFRA
AGINFRA Web portal: Providing discovery & data publication support to u	Agroknow	Open software license such as Mozilla MPL or GNU GPL	Data provider Technology providers Targeted beneficiaries/ users	Free service for all users sustained by dedicated not-for-profit legal entity (where key stakeholders will be members) Membership fee for legal entity sustainability and portal hosting expenses In kind contributions for provision of services and basic operation needs Future public/private donor financing for further evolution and extension
AGINFRA PLUS Software tools & online environments				
Agroportal of metadata	INRA	Open software license such as Mozilla MPL or GNU GPL	Technology providers Data providers	Free service for all users sustained by INRA

Schemas, ontologies, and classifications				Planned integration with VEST registry, in the context of harmonised AgriSemantics service Future public/private donor financing
PMF-ML: The PMFML specification can be used to exchange mathematical models in a software independent format	BfR	Open software license such as Mozilla MPL or GNU GPL	Data providers and consumers Technology & application providers	Free service for all users sustained by BfR Further evolution with BfR and German Government funding Future public/private donor financing
FoodProcess-Lab: KNIME extension for food processing simulations	BfR	Open software license such as Mozilla MPL or GNU GPL	Data providers and consumers Technology & application providers Targeted beneficiaries/ users	Free service for all users sustained by BfR Further evolution with BfR and German Government funding Future public/private donor financing
FoodChain-Lab: KNIME extension for tracing analysis and supply chain visualization	BfR	Open software license such as Mozilla MPL or GNU GPL	Data providers and consumers Technology & application providers Targeted beneficiaries/ users	Free service for all users sustained by BfR Further evolution with BfR and German Government funding Future public/private donor financing
ARPHA-based Publishing Frontend	PENSOFT	Licensed for use during project lifetime	Data providers and consumers Technology & application providers Targeted beneficiaries/ users	Appropriate licensing scheme to be defined by PENSOFT for future users Enhancement specific to PENSOFT publications and other services

				financed from own resources Public/private donor financing for future development & evolution
Refindit Agri-food Content & Data Harvesting and Aggregation Technologies	PENSOFT	Licensed for use during project lifetime	Data providers and consumers Technology & application providers	Appropriate licensing scheme to be defined by PENSOFT for future users Enhancement specific to PENSOFT publications and other services financed from own resources Public/private donor financing for future development & evolution
Advanced data fusion algorithms & components developed during AGINFRA PLUS & ported to VRE (e.g. user modelling, data mining, visualisation widgets, etc.)	Agroknow, UoA & third parties*	For AGINFRA PLUS funded software: open license such as Mozilla MPL or GNU GPL For third-party: appropriately selected by third party	Targeted beneficiaries/ users Technology providers Data providers	For AGINFRA PLUS partner contributions: software available as open source, accessible through the AGINFRA Github repository and software registry For third party contributions: to be defined by third party, licensed for use by AGINFRA PLUS partners & scientific communities for a given time period after end of project
Mobile & online apps to be developed by	Various third parties*	Will be appropriately selected by third party	Targeted beneficiaries/ users	Will be appropriately selected by third party

AgriTech startups on top of AGINFRA (during hackathons, challenges, incubation etc.)				Licensed for use by AGINFRA PLUS partners & scientific communities for a given time period after end of project

ANNEX B: Questionnaires used to collect partner input

Potential audiences/users of AGINFRA PLUS outcomes:

A. The scientists that are working in public sector universities & research centers in the European Union and associated countries.
B. The scientists working in the food industry, especially in departments and groups that are critical to business operations and innovation.
C. The scientists working in data science powered startups and SMEs in ag/foodtech.

For each one of the AGINFRA PLUS digital infrastructures/services, the corresponding tech partners need to provide a short description and reply to a few sustainability-related questions

1. Positioning Questions

- What is the title/brand of your service?
- Which are the competing services and who is providing them?
- Which are the top-5 characteristics of your service and which ones are offered by your competitors?
- From the audiences targeted by AGINFRA who are potential users that would pay for your service?
- How will you communicate and promote your service to these audiences? Which physical and digital channels you plan to use and how?
- Which are the strategic and commercial partnerships that you need to establish in order to reach them?

2. Business Model Questions

- Who owns the IPR of your software code and/or your service?
- In which way do you expect to charge your customers for your service? (subscription, one-off contract, per working hour/day, other)
- What is the price that they should pay?
- How much did you get paid for this particular service during 2019?
- How much revenue do you expect/want to have from this service for 2020 & 2021?

3. Operations Question

- What is the structure & the roles of the team managing the service?
- What is the structure & the roles of the team operating and supporting the service technically?
- What was the cost of hosting facilities and equipment (both in terms of technology and premises) for the provision of the service?
- What is the hiring plan for more people that are needed in 2020 and 2021?

- What is the estimated costs of hosting facilities and equipment needed in 2020 and 2021?
- What is the overall cost of maintaining and providing the service for 2020 and 2021 to serve the targeted audiences?

ANNEX C: Excel spreadsheet used to estimate financial projection data

All key technical partners, EGI, CNR, Pensoft and Agroknow had to define their financial viability scenarios that they envisage to continue offering and evolving their software infrastructures. These scenarios were accompanied with spreadsheet templates (https://bit.ly/2uLldXI) concerning their Business Plans for 2020-2023.

Particularly, CNR had to complete two spreadsheets for both D4Science and the VREs, Pensoft one spreadsheet for ARPHA-powered Data Journals, EGI one for EGI e-infrastucture and Agroknow one for Data Platform.

In these Excel spreadsheets they first had to complete the estimated costs to continue offering the service during 2020-2023, including the people/roles and their salaries. Then, to complete the other types of costs, such as hosting, marketing and business development ones. After the cost requirements were in place, they could then complete the two revenue/sales sheets, depending on whether they intended to pay for these costs from users/clients (eg via subscriptions), EU funding, or both.

Some special screenshots of the excel template spreadsheet that has been used to calculate foreseen costs, estimate potential revenues, and assess the profitability of the various AGINFRA PLUS tools & services follows.

AGINFRAplus SOFTWARE XNAME HEREX OTHER EXPENSES PLAN 2020-2023

Please complete only the cells in light blue (expense category) and orange (annual cost per category)

	ANNUAL COST ESTIMATES (IN € 000's)				
Cost of Software Delivery & Maintenance	2020	2021	2022	2023	Total
Travel costs for software delivery/maintenance/evolution	0	0	0	0	0
Subcontractor costs for software delivery/maintenance/evolution	0	0	0	0	0
Software hosting (cloud, web site, other)	0	0	0	0	0
Other software maintenance costs	0	0	0	0	0
	0	0	0	0	0
Marketing & Sales Expenses	2020	2021	2022	2023	Total
Organisation of marketing events	0	0	0	0	0
Sales & marketing trips	0	0	0	0	0
Conference registrations	0	0	0	0	0
Commercial fairs participation	0	0	0	0	0
	0	0	0	0	0
Operational & Administrative Expenses	2020	2021	2022	2023	Total
Administrative supplies/costs (phone, internet etc)	0	0	0	0	0
Non-delivery equipment costs (laptops, printers, etc)	0	0	0	0	0
Office rental & furniture	0	0	0	0	0
...other management & administration costs...	0	0	0	0	0

Profitability | Sales - Subscription | Sales - Funding | Personnel Costs | **General Costs**

Figure 5: General Costs excel sheet

AGINFRAplus SOFTWARE XNAME HEREX PERSONNEL PLAN 2020-2023

Please complete only the cells in light blue (names of roles needed) and orange (# of people, annual salary costs)

TITLES OF ROLES NEEDED	# OF PEOPLE REQUIRED				ANNUAL SALARIES PER ROLE (IN € 000's)				TOTAL PERSONNEL COSTS (IN € 000's)				
	2020	2021	###	###	2020	2021	2022	2023	2020	2021	2022	2023	Total
SOFTWARE DELIVERY / EVOLUTION													
Director of Product	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
Head of Innovation	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Innovation Front End	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Project Manager	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Junior R&D Associate	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Data Services Leader	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
Dev Ops	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Food Domain Specialist	0.0	0.0	1.0	1.0	0	0	0	0	0	0	0	0	0
Product Front-end	2.0	2.0	1.0	1.0	0	0	0	0	0	0	0	0	0
Customer Success	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
UI/UX Specialist	0.0	0.0	0.0	1.0	0	0	0	0	0	0	0	0	0
	5.0	5.0	5.0	6.0					0	0	0	0	0
SOFTWARE MARKETING & SALES													
Head of Marketing	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Marketing Associate	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Head of Accounts & Sales	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
EU Projects / Proposals Consultant	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Account Associate	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
	2.0	2.0	2.0	2.0					0	0	0	0	0
Management/Admin													
Head of Finance	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
Accounting Associate	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0
Head of Administration	1.0	1.0	1.0	1.0	0	0	0	0	0	0	0	0	0
Legal Manager	0.0	0.0	0.0	0.0	0	0	0	0	0	0	0	0	0

Profitability | Sales - Subscription | Sales - Funding | **Personnel Costs** | General Costs

Figure 6: Pesonnel Costs excel sheet

AGINFRAplus SOFTWARE XNAME HEREX BUSINESS PLAN 2020-2023										
All values automatically calculated, please complete all other sheets										
	2020		2021		2022		2023		TOTAL 4Y	
REVENUE	300		665		655		758		2,378	
COST OF GOODS SOLD (excl. labor)	0		0		0		0		0	
GROSS PROFIT/MARGIN	300	100.0%	665	100.0%	655	100.0%	758	100.0%	2,378	######
Direct Labor (actual prod)	0	0%	0	0%	0	0.0%	0	0.0%	0	0%
CONTRIBUTION MARGIN	300	100.0%	665	100.0%	655	100.0%	758	100.0%	2,378	######
Operating expenses:										
Business Dev labor	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!
Management/Admin labor (actual+bonus)	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!
Other marketing & sales expenses	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!
Other operating expenses	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!	0	#DIV/0!
	0		0		0		0		0	
NET OPERATING INCOME	300		665		655		758		2,378	
Other expenses:										
Depreciation	0		0		0		0		0	
…other…	0		0		0		0		0	
	0		0		0		0		0	
PRETAX PROFIT	300		665		655		758		2,378	
as a % of Revenue	100.00%		100.00%		100.00%		100.00%		100.00%	

Figure 7: Profitability excel sheet

AGINFRAplus SOFTWARE XNAME HEREX SUBSCRIPTIONS REVENUE PLAN 2020-2023												
Please complete only the cells in light blue (names of services) and orange (prices, # of items sold)												
								TOTAL SALES PROJECTION (in € 000's)				
Step I: Specifying Pricing Model & Price								2020	2021	2022	2023	Total
								0	0	0	0	0
Pricing Table (per service, per year)												
A. Software A as a Service (annual subscription)	0 €											
C. One-off software C use price (per piece/time)	0 €											
Step II: Estimating Number of Sales Per Year		2020	2021	2022	2023	TOTAL		2020	2021	2022	2023	Total
Γ. Product Revenue												
		# of new sales per year						Accumulative revenue per year (in € 000's)				
A. Software A as a Service (annual subscription)		0	0	0	0	0		0	0	0	0	0
C. One-off software C use price (per piece/time)		0	0	0	0	0		0	0	0	0	0
												0
		0	0	0	0			0	0	0	0	0
TOTAL		0	0	0	0			0	0	0	0	0

Figure 8: Sales - Subscription excel sheet

AGINFRAplus SOFTWARE XNAME HEREX EU OR DONOR FUNDING PLAN 2020-2023

Please complete only the cells in light blue (names of funders) and orange (total funding acquired per year € 000's)

		TOTAL FUNDING PROJECTION (in € 000's)				
		2020	2021	2022	2023	Total
		0	0	0	0	0
Step I: Estimating Funding Attracted Per Year		2020	2021	2022	2023	
1	EU or national R&D Funding (H2020 etc)	0	0	0	0	0
2	Donor-funded projects (Gates, DFID, etc)	0	0	0	0	0
		0	**0**	**0**	**0**	0

Figure 9: Sales - Funding