



EGI-Engage

D2.10 - Market Analysis report of selected sectors

Date	17 September 2016
Activity	NA2.3
Lead Partner	GRNET
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2843

This document describes the market potential of the AgTech sector, the actors and value chains, the results of the data requirements analysis and reports on the competing offerings including technical perspectives and specific recommendations for future business development within EGI.



This material by Parties of the EGI-Engage Consortium is licensed under a [Creative Commons Attribution 4.0 International License](#).

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program

COPYRIGHT NOTICE



This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The EGI-Engage project is co-funded by the European Union Horizon 2020 programme under grant number 654142.

DELIVERY SLIP

	<i>Name</i>	<i>Partner/Activity</i>	<i>Date</i>
From:	Kostas Koumantaros	GRNET/NA2.3	
Moderated by:			
Reviewed by	T. Ferrari B. Kryza	EGI Foundation/NA1	22 Aug 2016 30 Aug 2016
Approved by:	AMB/PMB		07 Sep 2016

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author/Partner</i>
V0.1	31-01-2016	Draft ToC	Kostas Koumantaros (GRNET)
V1.0	15-04-2016	Updated ToC	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET)
V1.3	15-05-2016	Preparation of the Chapter 4 “Market Opportunities for data value chains in the agri-food sector”	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET)
V1.6	27-05-2016	Updates on the Chapter 4	Kostas Kastrantas (GRNET)
V2.0	10-06-2016	Writing the sections methodology and identification of the top challenges in Chapter 5	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET)
V2.1	24-06-2016	Updates in the Chapter 5	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET)
V2.3	04-07-2016	Updating the Chapter 5 in terms of the community challenges validation	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET)
V2.4	14-07-2016	Updates on Chapters 3.2 and 4.3	Kostas Kastrantas (GRNET)
V2.5	22-07-2016	Overall updates based on feedback by Sy Holsinger	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET)
V2.6	26-07-2016	Chapter 3 update and finalization	Panagiotis Zervas (GRNET)

V3.0	03-08-2016	Finalisation for review	Kostas Koumantaros (GRNET) Charalampos Thanopoulos (GRNET), Kostas Kastrantas (GRNET)
V3.1	31-08-2016	Updated version with reviewers' comments incorporated. Updates in chapter 3.	Kostas Koumantaros (GRNET), Panagiotis Zervas (GRNET), Kostas Kastrantas (GRNET)

TERMINOLOGY

A complete project glossary is provided at the following page: <http://www.egi.eu/about/glossary/>



Contents

1	Executive summary.....	6
2	Introduction.....	7
3	Big Data Value Chain.....	8
3.1	Market potential of Big Data	8
3.2	The e-infrastructure perspective	11
3.3	SMEs and public Big Data e-infrastructures.....	11
3.3.1.	A success story from the agri-food sector: Syngenta	12
4	Market opportunities for data value chains in the agri-food sector	14
4.1.	The AgTech Market sector.....	14
4.1.1	How technology could help agri-food	15
4.1.2	Global AgTech Market Analysis.....	16
4.1.3	Sample sources of agricultural data	19
4.1.4	Key market stakeholders.....	21
4.2.	Agri-food big data value chain.....	24
4.2.1.	Analysis of the agri-food value chain	25
4.2.2.	The impact of Big Data in agri-food value chain.....	29
4.3.	E-infrastructures in agri-food sector	32
4.3.1.	Towards an agricultural data e-infrastructure	32
4.3.2.	Case studies on research data e-infrastructures	34
4.4.	Agri-food SMEs case studies.....	37
4.4.1.	The case of The Climate Corporation	37
4.4.2.	The case of AgDNA.....	37
4.4.3.	European case studies	38
5.	Mapping and validating the SMEs requirements	40
5.1.	Methodology.....	40
5.1.1.	Targeted user segment	40
5.1.2.	Questionnaire	41
5.2.	Personas Analysis	43
5.3.	Top challenges identification.....	46
5.3.1.	Data requirements analysis.....	48



5.3.2. Cloud service and Virtual Machine Requirements	49
5.3.3. System requirements.....	51
5.4. Community challenges validation.....	53
5.5. Envisaged solution.....	55
6. Conclusion and Discussion	59
7. Bibliography.....	60
8. Annex.....	61
8.1. Initial requirements elicitation survey.....	61

Table of Figures

Figure 1: The Big Data Value chain.....	10
Figure 2: Annual financing 2010-2015 in billion dollars (AgFunder, 2016)	16
Figure 3: Total Investment in AgTech during 2015 (AgFunder, 2016)	17
Figure 4: Money raised and Deals in different subsectors of AgTech during 2015 (AgFunder, 2016)	18
Figure 5: key market stakeholders on the agri-food sector.....	21
Figure 6: Agri-food value chain (PWC, 2016).....	25
Figure 7: Strategic data related control points in the agricultural value chain (Roland Berger, 2015)	26
Figure 8: An integrated farming system (Monsanto, 2012).....	30
Figure 9: Which of the following describes better the type of your company? (Question 1)	43
Figure 10: What is your role in your organisation (Question 2)	44
Figure 11: What is the status of your application? (Question 3)	45
Figure 12: Do you have a defined business plan for your products? (Question 4)	45
Figure 13: In which agricultural-specific thematic area(s) does your software/ application belong (Question5).....	48
Figure 14: What type of information is needed for your software/ application (Question 6)	49
Figure 15: For your software/ application, which are the typical usage scenarios of the cloud? (Question 9)	51
Figure 16: How often should the required data be updated for your product? (Question 10).....	51
Figure 17: In which format could you use the provided information? (Question 12).....	52
Figure 18: In terms of access to information through API, which of the following is the most preferable? (Question 13).....	52
Figure 19: Which of the following are the top challenges that should be addressed, in order to serve your users in a better way? (Question 11).....	54

1 Executive summary

In order to explore the needs and challenges of the agri-food sector, EGI-Engage aims with this deliverable to investigate the potential of this particular market. This is a user segment completely new to EGI. For that reason, a market analysis of the Agricultural Technology (AgTech) sector was performed, checking the global market data with a focus on startups and SMEs. Furthermore, the AgTech market's stakeholder composition was reviewed focusing on six key segments: research centers, seed and pesticide companies, agricultural equipment manufacturers, ICT and data companies, investors and startups and SMEs.

Following this, an analysis of the agri-food big data value chain was performed targeting e-infrastructures. In terms of e-infrastructures, the overall rationale behind a domain specific infrastructure was explained, presenting key examples of the agri-food use cases targeting research and academia. Additionally, success stories from startups and SMEs in the European and global AgTech scene were analysed. One of the main goals of this deliverable, besides the market analysis, was to collect requirements from business organizations of the agri-food sector. A requirements elicitation process was followed by contacting various stakeholders of the European AgTech community through special events and networks. The requirements and challenges were based on four pillars: the persona that the interviewee represents, business data requirements of the SME or startup that the interviewee represents, cloud and e-infrastructure requirements and top challenges that an AgTech company is facing.

Following the analysis of the top (eleven) validated community challenges, there is a clear need of a federated infrastructure able to manage various similar and unrelated data types, high volume and veracity data that are being syndicated from a wide range of data source. Recommendations were identified aiming at developing customised solutions and enhancing the existing service catalogue of EGI. This includes EGI activating services that are already in place, such as the Open Data Platform, DataHub, ARGO, and AppDB and enhance them with agri-food targeted features in order to address the challenges of the agri-food community. These products could help solve pain points such as improving data access, data combining, and VM management, reducing time consuming activities such as storing, managing and using large amount of data while coupling the ability to custom processing power, storage and memory through the EGI Federated Cloud open standard based interfaces and services.

Provided that the EGI community wants to pursue an expansion to the agri-food sector, it is suggested to follow up and promote the new and existing EGI services through a variety of sectorial events (e.g. European Data Forum) and domain specific networks, such as AGINFRA, and sign bilateral agreements and collaboration with existing domain-specific e-infrastructures.

Overall, the information presented in this report provides a solid foundation for consideration in order to help the EGI community expand the outreach of their services and reach a new market segment.



2 Introduction

The current deliverable aims at investigating the market potential, size, structure, stakeholder composition and segmentation of the agri-food sector. In addition, Big Data value chains were investigated in this sector alongside a global AgTech market analysis.

Since SME engagement is of high importance for the EGI community, a requirements collection and validation process is provided in order to profile new and enhanced EGI services and propose recommendations for big and/or open data services targeting the industry and academia. This activity also focused on developing personas (descriptions of typical users) and scenarios described in detail, and then these assumptions were validated in a series of interviews with potential users from the AgTech communities.

The interviews covered different roles, segments, and activities of the agri-food community and as a result examine which of the community challenges can be addressed by EGI, with existing or new solutions.

3 Big Data Value Chain

This chapter provides the background of this study by subsequently stating the market potential of the Big Data sector and the e-infrastructure perspective, while also presenting relevant multidisciplinary examples of successful stories of SMEs interacting with public Big Data e-infrastructures.

3.1 Market potential of Big Data

Today, we are witnessing a production and collection of massive amount of data that is growing to an unprecedented scale. It is no surprise that by 2020 it is estimated that there will be more than 16 zettabytes (16 Trillion GB) of useful data (Turner et al. 2014). As also indicated in the report of the High-Level Panel of Eminent Persons on the post-2015 Development Agenda, this “data revolution” does not only include the explosion in the volume of data, but highlights an era where every device is online, where sensors are generating continuous streams of data from every point of the world, where the Internet of Things will produce a digital fingerprint of our world.

Big data is a popular phrase used to describe a massive volume of both structured and unstructured data that is so large that it is difficult to process with traditional databases and software techniques (Addison *et al.*, 2015). Good, available and reliable data are able to have important effects on peoples’ lives. In addition, data is helping businesses and individuals for an effective decision making process, which is affecting their well-being. A plethora of mobile devices, different data types and crowd-sourced data, has already affected official data collection processes and the design of similar programmes.

An example from satellite earth observation domain, can prove how data is evolving a particular sector. High-resolution imagery acquisition was by far a costly process. Given the growth of satellite data, a vast amount of high-resolution satellite images is made available in an open manner, reducing the acquisition cost and enabling new services to flourish, such as automated processing. Furthermore, a recent report from the Technical Centre for Agricultural and Rural Cooperation (CTA) indicates that many applications have emerged in this particular sector – enabling innovative business ideas to occur, namely harvest prediction, disaster response to address food security issues, monitoring geographic patterns and disease transmission corridors with geospatial determinants, measuring population density and the spread of new settlements, and mapping and planning transportation infrastructure¹.

According to Doug Laney of META Group (then acquired by Gartner), big data has three (3) dimensions that specify the complexity on managing large sets of data: *volume*, *velocity* and *variety* (Laney, 2001). The big data management in areas, like agri-food sector, takes in consideration the combination of the three properties rather the effectiveness of just one of them (e.g. volume – the amount of data).

¹ Full report can be accessed at: http://publications.cta.int/media/publications/downloads/1937_PDF.pdf



The property *volume* captures the big data that is generated, aggregated and stored in an intense amount. Also, a massive amount of data is analysed by new technologies, like Hadoop (<http://hadoop.apache.org>), which is one of the most popular platform on handling big data from different data sources and executing large scale processing jobs. A typical example of a large amount of data in agri-food sector is the collection and storage of sensing data from remote field sensors, as well in other parts of the production chain of agricultural products (e.g. during the packaging process, selection of the right fruits according to their shape and their colour).

Additionally, the dimension *velocity* focuses on how fast data is processed, covering the aspect that software makes the generated and/or collected data directly available to the end users and analytical tools offer real time processing mechanisms of the data, and direct access to their results (data products). By using such tools, the efficiency of operations is maximized and the need for physical manpower is reducing. There are several companies in the agricultural domain that are dedicated on offering such services that combine the field data with real-time and historic data on soil texture and climate conditions in order to predict the pest infection in the cultivation.

Finally, the third “V”, *variety* refers to the various types of data, combining structured and unstructured data, data from heterogeneous sources, in various formats and different vocabularies that are dedicated to a specific topic (e.g. organic agriculture). An example of such software is able to pull data from different sources, e.g. data from weather observation stations and satellite data and provide better weather forecasting for a particular field location on a particular moment.

The Vs of big data challenge the fundamentals of existing technical approaches and require new forms of data processing to enable enhanced decision-making, insight discovery, and process optimisation. As the big data field matured, other Vs have been added such as *Veracity* (documenting quality and uncertainty), *Value*, etc. (Curry, 2016).

Big data sources provide huge data volumes, which require high storage and processing resources; so using big data requires moving a step further from traditional statistical methods that are unable to handle huge volumes of information into the use of advanced data mining and machine learning algorithms with the required computational efficiency. In some other cases there is a need for analysing and linking heterogeneous data from diverse data sources or extracting information from unstructured data. Estimates of OECD indicate that the large amount of business-generated data could be as high as 80% to 85% unstructured and largely unexploited or underexploited. In the past, extracting value from unstructured data needed a large amount of effort. However, using big data technologies, to link and analyse unexploited data repositories, enables to extract valuable information in an automated and cost-effective way (OECD, 2013). It is evident that the process of creating value from big data is challenging. In order to address this challenge, it is required to develop competences in the three following areas: i) scalable data management (processing, storage and computational resource management), ii) data analysis (harnessing statistics generated) and iii) expertise in the field, i.e. sustainable agricultural development.

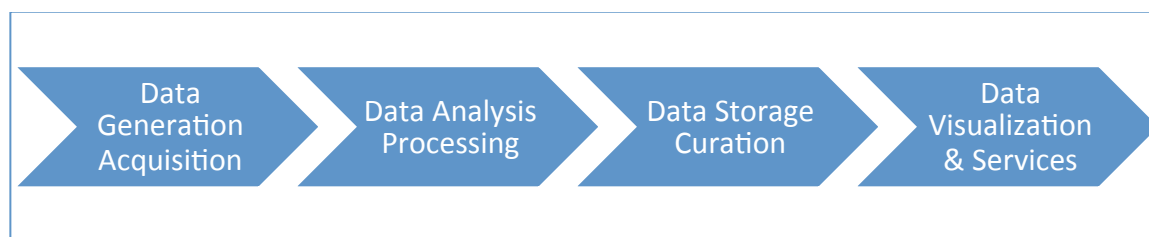


Figure 1: The Big Data Value chain

In order for the Big Data technologies to thrive in a pan-European level, strong stakeholders are needed to participate in every part of the Big Data Value Chain (Figure 1), ranging from data generation and acquisition, through to data processing and analysis, then to curation, usage, service creation and provisioning (Curry, 2016). For a Big Data Ecosystem to flourish, each part of the value chain should grow strong.

Currently, numerous stakeholders from the private sector are active in every part of the Big Data Value Chain, offering combined services and solutions. Some companies focus in providing enormous amount of data either by collecting streaming data from heterogeneous sources or increasing their value using extra processing techniques (including data validation, integration, enrichment and linkage to external resources). Others are focusing in analysing the data for extracting patterns and correlations while others apply these findings in the decision making process of various application domains.

Although there is an increasing amount of companies involved in the big data movement, no coherent data ecosystem exists in European level. On one hand there are software moguls (ATOS, Oracle etc.) that are active in the data business providing services to large corporations and on the other hand there are SMEs offering their services in dispersed sectors (finance, health etc.) or focusing in specific innovative areas (harvesting, visualization etc.). It is evident that, although the production and usage of data is growing, stakeholders from both private and public sector are not fully exploiting the opportunities that could be generated by facilitating their close cooperation throughout the Data Value chain (DG Europe, 2013).

Stakeholders that could be considered as candidates for setting the basis of the Big Data ecosystem fostering new businesses and innovations are:

- Vendors of the ICT industry (Large and SMEs) as providers of data management technology solutions
- Users across different industrial sectors (private and public) that take to their advantage every latest development from the Big Data movement.
- Innovative Big Data Value companies that do not exist yet and will (potentially) emerge from the startup ecosystem.
- Researchers and academics who can provide knowledge and thought leadership

The involvement of numerous stakeholders across different domains is a key element for advancing the Big Data economy in Europe. The participation of SMEs and startups in this ecosystem is also very important for the vitality of the Big Data Value Chain since they are an

essential part of the process to create value based on their specific and strong niche competences at the technical, application and business level.

3.2 The e-infrastructure perspective

An e-infrastructure that supports data-intensive, multi-disciplinary research is needed to facilitate new discoveries and accelerate the pace of science to address 21st century global change challenges. Data discovery, access, sharing and interoperability collectively form core elements of an emerging shared vision of e-infrastructure for scientific discovery. These elements further depend on building relationships among data sets, people, systems, organisations and networks. However, the pace and breadth of change in data and information management across the data lifecycle means that no one country or institution can unilaterally provide the leadership and resources required to use data and information effectively, or to establish and maintain the relationships needed to support a coordinated, global e-infrastructure.

3.3 SMEs and public Big Data e-infrastructures

According to the European Commission², Digital Infrastructures or e-Infrastructures' activities aim at empowering researchers with easy and controlled online access to facilities, resources and collaboration tools, bringing to them the power of ICT for computing, connectivity, data storage and access to virtual research environments. Small and medium-sized enterprises (SMEs) represent 99 percent of all businesses in the European Union. They are estimated to account for over 60 percent of new jobs created in the United States and 60–70 percent of new jobs created across all developed and developing countries, according to Organisation for Economic Co-operation and Development (OECD) estimates. A recent report of Kauffman Foundation³ found that, between 1977 and 2005, startups added on average 3 million jobs in their first year of existence. It is essential to understand how SMEs can create value with open data, and thus benefit global economy. There are two reasons why EGI-Engage should focus on SMEs regarding open data.

First, both startups and SMEs stand among the chief beneficiaries of open data. One of the main difficulties that those business organisations have to address is their inability to access vast amounts of data and to the respective tools that are needed for data processing and analysis. Not to mention the high cost of such services. The evolution of the global open data ecosystem, brings SMEs and startups to the forefront, enabling them not only to access large open datasets, but also to benefit from a wide range of cloud based tools and services that help them take data-driven decisions.

Second, small and medium enterprises could be considered as the main innovation enablers in the open data sector. A recent GOVLAB publication⁴ indicates that: “Many established companies

² <https://ec.europa.eu/digital-single-market/en/digital-infrastructures>

³ <http://www.kauffman.org/what-we-do/research/firm-formation-and-growth-series/the-importance-of-startups-in-job-creation-and-job-destruction>

⁴ <http://www.thegovlab.org/static/files/publications/OpenData-and-SME-Final-Aug2015.pdf>



tend to view open data as a threat— a disruption of existing business models and legacy revenue models”. This way, the approach of big companies seems to be rather sceptical in opening their data. It is obvious that these companies prefer to shield their data raising concerns about open data handling, encryption of public data exchange and data protection. On the other hand, the agility of startups and SMEs allows them to experiment and maximize the potential of open data, following disruptive business models and developing tools that add value to the rich information available.

Moreover, another subdivision of open data – open science data – introduces remarkable growth potential; for startups and SMEs. Science data, which contain various data types, create new opportunities for businesses with a well-specified data expertise (e.g. in health data). Joint efforts for open science data are being driven by platforms such as OpenAIRE⁵, the Open Science Data Cloud⁶ and the Open Science Grid⁷ and other initiatives. The importance of open scientific data, could be highlighted by the increasing number of highly acclaimed donors such as the European Commission and the Gates Foundation, asking open data sharing rights from their beneficiaries. European Commission⁸ claims that “with limited exceptions such as privacy and third party rights-maximising reusability is the best way to maximize scientific innovation and return on investment”. In that context, the number of ventures and businesses that are built on open science data sharing has grown rapidly.

3.3.1. A success story from the agri-food sector: Syngenta

Syngenta⁹ is one of the key players in global agriculture, supplying farmers with agrochemical products and seeds – contributing to a better use of their available resources. In order to continue to advance crop productivity, the company invested more than \$1.4 billion in research and development (R&D) across 150 international sites in 2014¹⁰. Syngenta is well-known in using publicly available data for R&D purposes. Among others, Syngenta’s researchers make use of open land, meteorological and biological data in order to better understand a crop and its characteristics. European Bioinformatics Institute (EMBL-EBI)¹¹, an EGI Council Member, is supported by Syngenta and hosts plenty of the public molecular biology data, which R&D researchers are using. The data covered, include insecticides, fungicides and herbicides, with over 40,000 compound records related to crop protection. As part of the Syngenta’s Good Growth Plan¹², the company decided to publish its own open data. The plan aims at addressing the global

⁵ <http://openaire.eu/>

⁶ <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

⁷ <https://www.opensciencegrid.org/>

⁸ http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

⁹ <http://www.syngenta.com/>

¹⁰ <http://theodi.org/open-enterprise-big-business-case-study-syngenta>

¹¹ <http://www.ebi.ac.uk/>

¹² Syngenta, (2013) Syngenta launches The Good Growth Plan. [Blog] Available at: <http://www.syngenta.com/global/corporate/en/news-center/news-releases/Pages/130919.aspx> [Accessed 2016-06-07]



food security challenge for the world's population, by providing six open datasets including descriptions of productivity, soil, biodiversity and smallholder reach. By releasing these data and making open research a commitment for the future, the company aims at strengthening collaboration between private and public stakeholders for global food security. Taking into account the rise of Internet of Things (IoT) devices, Syngenta will have to manage large quantities of agricultural data, collected by these devices. Syngenta's researchers are provided with a wide range of datasets and should co-operate with organisations that are able to generate, host and analyse agricultural data, in order to create meaningful products and create value for the directly and indirectly involved stakeholders.



4 Market opportunities for data value chains in the agri-food sector

This chapter analyses all relevant market data within the agri-food sector, while including case studies of agri-food SMEs and their interaction with public e-infrastructures.

4.1. The AgTech Market sector

According to FAO¹³, by 2050 the world's population will reach 9.1 billion, 34 percent higher than today. The demand of food is constantly increasing and by 2050 food production must double in order to meet the demand from world's growing population. Based on the declaration of the World Summit on Food Security (2009), food security exists when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their needs and food preferences for an active and healthy life (FAO, 2009). Despite the fact that food availability is the main pillar of food security, mainly through agricultural production, the access to food is raising important concerns for achieving food security in modern societies. Moreover, external factors such as climate change, the agricultural technological advances, the existing high volume of food waste, and the changes of the nutritional habits build the modern global food management framework by directly threatening food security.

Technology is transforming every industry and now agriculture is directed into the spotlight. During the last six years, a growing number of agricultural technology (AgTech) startups and SMEs have applied various innovative technologies in the agri-food sector. Sustainable agricultural technology has the potential to reshape global agriculture. Given that the production of food must be increased during the next forty years, technology will play a significant role in this challenge. Considering that the planet is showing signs of severe environmental stress and natural resources stocks are continuously decreasing, AgTech innovations are essential to overcome these monumental global challenges. As it is stated on Kauffman Foundation white paper¹⁴, "sustainable agricultural technology has the potential to completely reshape global agriculture, dramatically increasing the productivity of the agriculture system while reducing the environmental and social cost of current agricultural production practices". Many observers predict that the growth of Big Data and AgTech will bring positive benefits through enhanced production, resource efficiency, and improved adaptation to climate change.

13

http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

¹⁴<http://www.kauffman.org/what-we-do/research/2014/04/agtech-challenges-and-opportunities-for-sustainable-growth>



The AgTech sector includes a wide variety of subcategories. Namely:

Animal Nutrition & Health	Information Systems
Aquaculture	Integrated Pest Management
Bioenergy	Irrigation Efficiency
Biological Pest Control	Land Management
Biomaterials	Machinery
Bionutrition	Precision Agriculture
Biotechnology	Robotics
Crop Nutrition	Seeds and Genetics
Crop Protection	Soil Amendments
Decision Support Technologies	Soil Health
Feed Efficiency	Sustainable Production Systems
Fertilizer Efficiency	Technology Transfer
Food Ecommerce	Urban Agriculture
Food Traceability and Safety	Water Quality and Preservation
Food Storage and Preservation	Waste Mitigation and Manure Management

4.1.1 How technology could help agri-food

Opportunities in the AgTech sector seem very compelling. It is interesting to watch traditional agricultural giants - that have historically focused their research to improve seeds and fertilizers- now working in close collaboration with Venture Capital (VC) backed technology startups and SMEs. These companies now focus on software, sensors, smart equipment, cloud software, predictive analytics and algorithms as well as robotics and drones. They do so, aiming to collect, analyse and take action on the data that are related to the agri-food sector and focused towards their research interests. A wide variety of mainstream technologies, such as mobile apps, digital mapping, field sensors, big data, cloud-based business systems, smart farming equipment, autonomous aerial and field vehicles are being used to create an interconnected digitalised ecosystem that is tracking product flow from farm to the market. Cloud enabled services and other proven technology capabilities, including data analytics and machine learning can impact yield, productivity and sustainability improvements across farms, fields and factories. Farmers can benefit from similar tools in order to manage more efficiently fields, animals and resources and achieve higher levels of productivity. Digital technologies can provide value to farmers at two levels. First, through on-farm connectivity; farmers can monitor the status of equipment, livestock, and water levels. Second, as the farm is connected to the outside world, farmers can access



advisors, markets, climate information, and best practices. Furthermore, there are various technologic applications from which the farmers can benefit. For instance:

- On-farm 3D printing for cheap, quick access to maintenance equipment
- Ability to stream “how to” videos or video conference with a trusted advisor
- Virtual paddocking capabilities to move livestock automatically from a distance
- Real-time decision support, like taking a photo of a struggling plant, uploading it with diagnostic information, and quickly receiving an analysis that suggests potential causes and optimal next steps
- Unlocking funding sources by leveraging on-farm data to provide transparency about working capital resources.

4.1.2 Global AgTech Market Analysis

Based on CrunchBase¹⁵ data, 2015 was a landmark point for AgTech investments globally. The impressive amount of \$4.6 billion was raised in 2015, which is nearly double the amount of investments’ performance in 2014 in the AgTech sector. Data from KPMG’s Venture Pulse (2015) and AgTech 2015 Investing Report (AgFunder, 2016)¹⁶ show that the global funding of companies supported by Venture Capitals grew 78% in 2014 on 2013 levels and a further 44% in 2015. Compared to this, AgTech investments grew 166% in 2014 and 92% in 2015. If the size of the global food and agriculture market is taken into consideration, which represents about 10% of global GDP - \$7.8 trillion (Plunkett, 2016), the investors are here to stay. This means that the AgTech market hides great potential, since based on 2015 data the investments are less than 0.5% of the entire agri-food market. Moreover, the percentage of total AgTech investment is less than 3.5% of the \$128.5 billion invested in venture-backed companies in 2015, which seems rather small for a sector representing 10% of GDP.

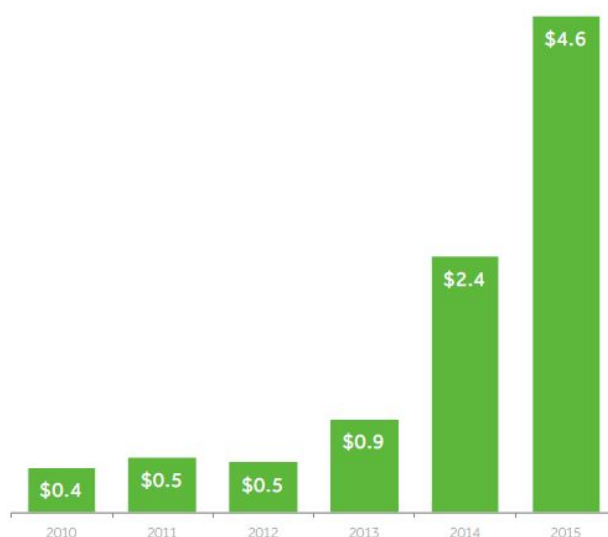


Figure 2: Annual financing 2010-2015 in billion dollars (AgFunder, 2016)

¹⁵ <https://www.crunchbase.com/>

¹⁶ <https://research01.agfunder.com/2015/AgFunder-AgTech-Investing-Report-2015.pdf>

As the figure above indicates, AgTech investment had remarkable growth during the past six years. Since 2015 was the year with the highest amount invested in the sector's history, it should be further examined in order to depict the individual subsectors' dynamic of the global market. From the AgTech subsectors mentioned in section 4.1, the one which gets the lion's share is food ecommerce, raised \$1.65 billion in 2015. The next to follow is irrigation and water technology raising \$673 million. Also, as the third category in the row with an increased portion of the pie is precision agriculture (14%), which includes Drones & Robotics with 8% and Decision Support Technology with 6%.

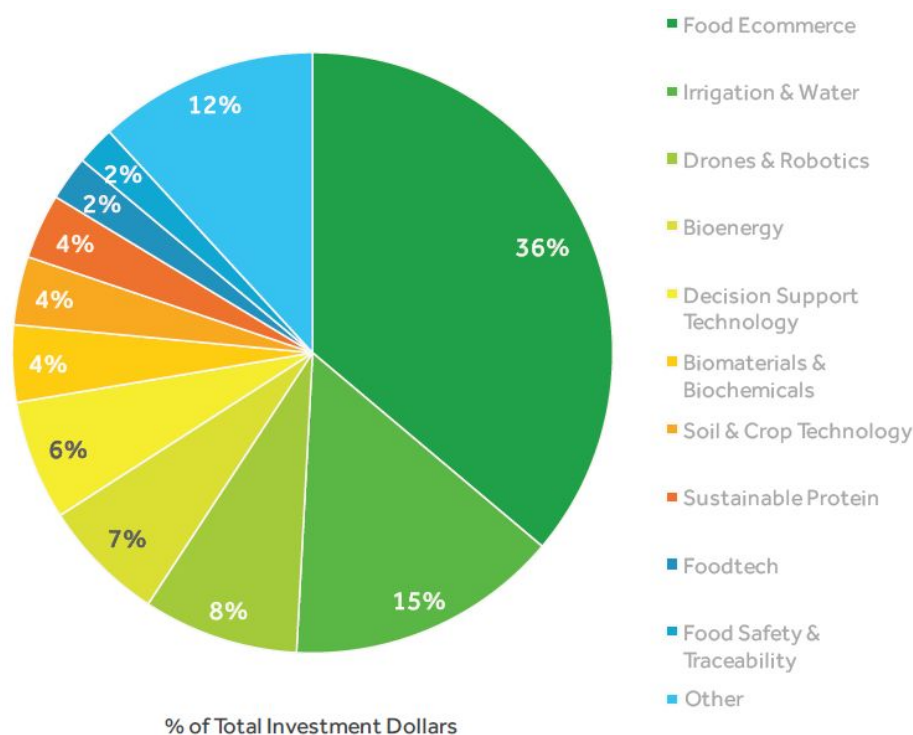


Figure 3: Total Investment in AgTech during 2015 (AgFunder, 2016)

As the total investment in AgTech was analysed in categories, it is important to present the financial data that regard the deals and the money raised in the different subsectors during 2015. A bottom-up analysis of the most invested categories follows based on the data presented in Figure 4, in order to precisely understand the involved AgTech subcategories in each sector.

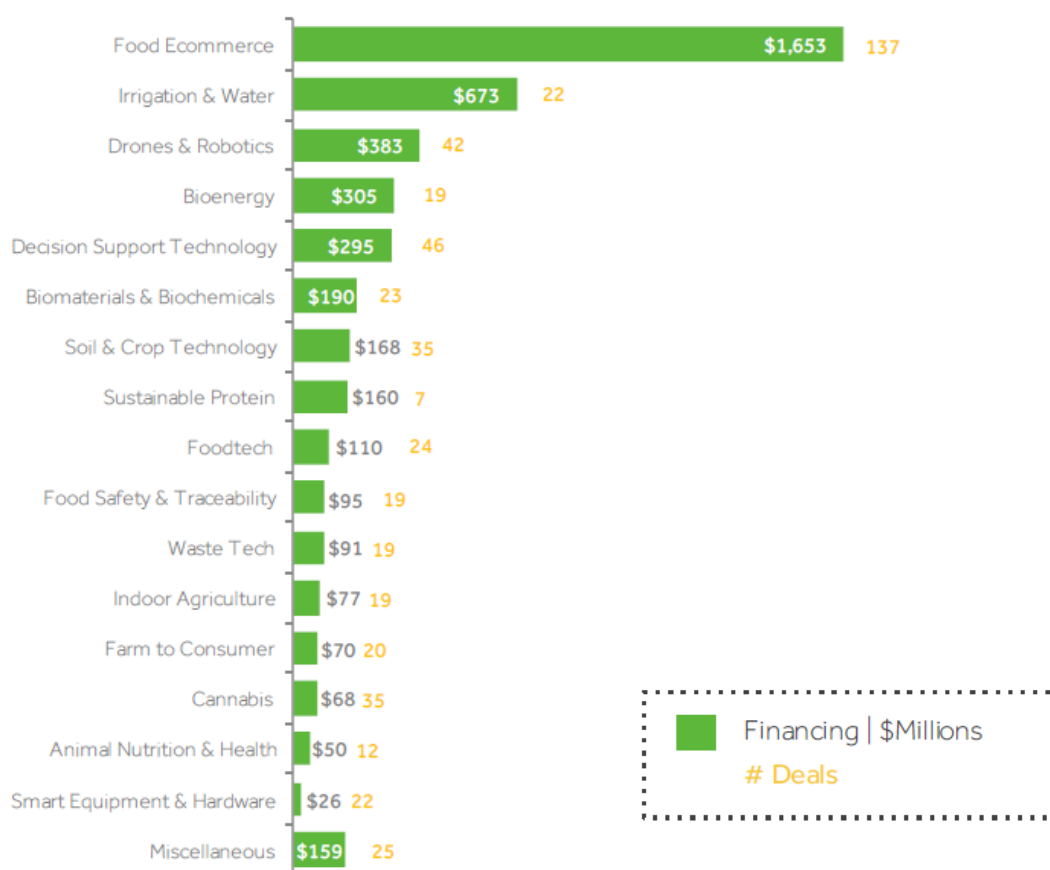


Figure 4: Money raised and Deals in different subsectors of AgTech during 2015 (AgFunder, 2016)

Regarding precision agriculture, which in 2015 attracted \$661 million, it involves a wide spectrum of different sub-sectors. A more granular approach is followed for the grouping of this sector, dividing them by technology (e.g. drones, hardware, sensors, robotics, satellite & imagery and software companies) and application (e.g. irrigation and water and weather related companies). It is worth to mention that 95% of the deals in precision agriculture referred to technology companies, which combine insights from a variety of sources, satellite imagery gathered pace as a decision support tool.

While precision agriculture has captured the AgTech spotlight during the past few years, it seems that high-profile investors (such as Monsanto Growth Ventures¹⁷ and Syngenta Ventures¹⁸) are heavily investing in the soil and crop technology. Within this category biotechnology startups and SMEs producing compounds to protect crops from weeds, fungal diseases and insect pests are included.

The outlier of AgTech with \$1.65 billion invested is food e-commerce. With 137 companies having raised funding around the globe and countries such as China and the United States having

¹⁷ <http://www.monsanto.com/whoweare/pages/monsanto-growth-ventures.aspx>

¹⁸ <http://www3.syngenta.com/global/syngentaventures/en/Pages/home.aspx>

the biggest deals in the sector, e-commerce seems to be the most attractive sub-segment of AgTech for investors.

In addition to the top three most invested categories an extra category – named as the most promising in the AgTech sector – is presented. Based on AgFunder report for 2015, companies manufacturing biological inputs (e.g. biopesticides and biofertilizers) largely dominated 2015's soil and crop technology sub-sector, raising \$120 million. This represents 71% of the sector's \$168 million total, leaving seed tech and agricultural genetics companies to complete the remaining 29%. This abundance of biological solutions coexists with the rise of soil health improvement techniques, sustainable farming methods and low input agriculture. This sub-sector attracted multiple investors including the Bill & Melinda Gates Foundation, The University of Texas Investment Management Company¹⁹, Sequoia Capital, and the corporate venture capitals of agribusiness giants Monsanto and Syngenta. As awareness of the importance of soil health increases, in tandem with global climate change, it is believed that a growing number of environmentally-aware farmers and organic-eating consumers will keep this segment on an upward pace.

4.1.3 Sample sources of agricultural data

Agri-food data was not always accessible through e-infrastructures. Sample sources, which are presented below, relate to big agricultural data and are examples of relevant open and closed datasets and data sources, which are not solely in a digital format.

4.1.3.1 *Agricultural census enumeration areas*

In many countries, cartographic materials and data from the population census are also used for the agricultural census. The sampling frame consists of enumeration areas and aggregated data from the census data collection. As in the population census, random samples of enumeration areas are selected and screened for farms or agricultural holdings for agricultural production surveys.

4.1.3.2 *Farm registers from the agricultural census*

As in the household registers, countries with adequate capacity can use the agricultural census to develop farm registers, which provide a powerful sampling tool because as it allows a choice of many alternative sampling designs. A major weakness is that the registers rapidly become out-of-date; out-of-date population and farm registers erode all of the data quality dimensions because the completeness of coverage changes over time, thus affecting the comparability and accuracy of the resulting estimates.

4.1.3.3 *Area sample frames*

An area sample frame is the land mass of the country or the space within a country containing the populations of interest and is suitable for obtaining information about variables associated with

¹⁹ <http://www.utimco.org>



land such as crops, livestock, forests and water. Both maps and satellite images are used to divide the country into administrative areas such as provinces, districts, etc. Satellite imagery can also be used to subdivide the administrative areas into land-use categories, such as cropland, rangeland, woodlands, urban areas, etc. Sampling units of segments of land with identifiable boundaries can be formed, or each land-use stratum can be divided into square grids with a sample of points becoming the sampling units. During the data collection process, rules of association are used to connect farm holdings or households to the segments or points. Depending on the process used, area frames can be costly and time consuming to construct. However, use of satellite imagery and two-stage sampling of points have reduced the cost and time. An added advantage of an area frame is that the frame does not go out-of-date; it is complete and provides a basis to georeference survey data with the underlying land use. It also provides ground truth useful for classifying satellite imagery by land cover. The primary disadvantage of area frames is that the sampling is based on land use and not on the size and type of agricultural holding.

4.1.3.4 *USDA's Market Analysis & Reporting Services Platform (MARS)*

The USDA's Agricultural Marketing Service supports a dynamically enriched platform called MARS²⁰, collecting and distributing information electronically from remote locations, by combining reports from all commodity areas (livestock, cotton, specialty crops, and dairy) into a single entity. This service improved the transparency, speed, accuracy, and flexibility of US market data, while it allows for better analysis for all agricultural interested parties.

4.1.3.5 *AGRIS*

AGRIS²¹ is a global public domain database with more than 8 million structured bibliographical records on agricultural science and technology that come from a large international network of more than 170 participating institutions from 70 countries.

4.1.3.6 *Group on earth Observations (GEO)*

The Global Earth Observation System of Systems (GEOSS) is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors²². The main goal of the GEOS system is to introduce a global public infrastructure for Earth observations, which will be consisted of a wide and flexible network of different systems and content providers. The key advantage of this system is that it links earth observation systems from 96 different countries and the EU and enables to a wide range of data providers share their data.

²⁰ <https://www.ams.usda.gov/>

²¹ <http://agris.fao.org>

²² Definition from <http://www.earthobservations.org/geoss.php>



4.1.3.7 CIARD RING

CIARD²³ is a global movement for open agricultural knowledge for development, currently consisting of more than 6,200 members; about 440 of which are organisations and public institutions. The network is working on the advocacy on open knowledge for agricultural development, promoting open access to agricultural knowledge. CIARD is working mostly on capacity development and in this context; it produces capacity building material in the form of pathways, webinars, e-discussions, working groups and an advocacy toolkit, all of which aim at facilitating access to agricultural research outcomes so that they become available to all types of stakeholders. CIARD is also responsible for and maintaining the RING, a global directory of web-based services that provide access to any kind of information sources pertaining to agricultural research for development (ARD). CIARD RING provides access to resources such as providers, services and datasets. CIARD RING is one of the core components of the AGINFRA global agri-food research e-infrastructure.

4.1.4 Key market stakeholders

The AgTech market is built upon a diverse ecosystem including traditional agricultural enterprises such as seed and pesticide companies (e.g. Monsanto, Syngenta), agricultural equipment manufacturers (e.g. John Deere) and major research centres (e.g. INRA). Moreover, on the tech side of agri-food sectors, there are many involved investors (VCs, EU funds) and ICT companies that are entering the market. Big multinational tech companies (i.e. IBM, SAP and Oracle) are investing a remarkable amount on bridging the agri-food sector with technology and data science.



Figure 5: key market stakeholders on the agri-food sector

²³ <http://www.ciard.info/>

4.1.4.1 Research Centers

The global agricultural research can be characterised as the key enabler of AgTech and agricultural Big Data applications. Research centers are supported by public or private funding, and based on 2008 estimates²⁴, global agricultural R&D spending accounts for \$40.1 billion. Major global agricultural research organisations such as the CGIAR²⁵ are responsible for the production of large experimental datasets, containing heterogeneous and diverse data types, such as phenotypes, genomics, genetic and germplasm data - to name a few.

4.1.4.2 Seed and Pesticide Companies

The market of pesticide and seed companies is estimated at over \$93 billion (based on 2013 data) and is dominated by the six key players; namely BASF, Bayer, Dow, DuPont, Monsanto and Syngenta. These six companies control 75% of the global agrochemical market. Application and use of nutrients and pesticides are increasingly complex. As producers look for ways to reduce cost and become more efficient, the application of additives is frequently an area for improvement. Chemical companies are playing an increasing role in the research and development of data collection tools and methods to improve application use.

4.1.4.3 Agricultural Original Equipment Manufacturers and Suppliers

Global agricultural equipment market accounts for \$116 billion, based on 2013 estimates of the Association of Equipment Manufacturers²⁶. Data coming off machines is helpful to the farmer using the equipment, but it is also an important resource for manufacturers looking to understand exactly how their equipment is being used, how they can improve feature production, and, perhaps most importantly, how they can create revenue opportunities from product support programs. Likewise, access to machine data helps dealers meet increasing service requirements of very sophisticated equipment. They gain insight into when equipment needs to be serviced, and use that information to inform proactive preventive maintenance visits. Over the last two years, growers have already reduced spending on farm equipment purchases. Along this vein, successful farm equipment manufacturers should emphasise equipment's ability to improve efficiency and demonstrate clear Return on Investment (ROI) on the equipment purchase. In many cases, the manufacturers of traditional farm equipment (e.g. tractors and harvesting combines) are well positioned to expand into data collection technologies since, in many cases, the technology is an extension of the equipment already in use.

4.1.4.4 ICT/ Data Companies

Data companies such as IBM and Fujitsu have also placed big importance in AgTech and Agricultural Big Data. According to the research consultancy IDC²⁷, the global information technology market, encompassing hardware, software, services, and telecommunications, is

²⁴ <https://www.asti.cgiar.org/globaloverview>

²⁵ <http://www.cgiar.org/>

²⁶ <http://www.aem.org>

²⁷ <http://www.idc.com/>



expected to reach \$3.8 trillion in 2016, up from \$3.7 trillion the previous year²⁸. With expertise in high performance supercomputing, computational sciences, analytics and optimisation, ICT and data companies are uniquely able to understand the complexities of agriculture and develop the right weather forecasts, models and simulations that enable farmers and companies to make the right decisions.

4.1.4.5 Investment Funds/ Traders

The global AgTech funding scene counts 499 companies (2015 data) attracting \$4.6 billion of investment across 526 rounds of financing. Since 2013, only 26 funding mechanisms for FoodTech and AgTech existed globally; great traction has been reported with remarkable investors joining forces. Agfunder²⁹ reports 672 unique investors active in the AgTech sector during 2015. Some of the sector's renowned investors are: Bill & Melinda Gates Foundation, The University of Texas Investment Management Company, and Sequoia Capital.

4.1.4.6 Startups and SMEs

The main actors that bridge research and innovation with the agri-food industry are the AgTech startups and SMEs. During the past years, the global AgTech ecosystem has been blooming according to the European Commission³⁰, "SMEs can play a crucial role in developing resource-efficient and cost-effective solutions to secure sufficient supplies of safe, healthy and high quality food and other bio-based products, by developing productive, sustainable and resource-efficient primary production systems, fostering related ecosystem services and the recovery of biological diversity, alongside competitive and low-carbon supply, processing and marketing chains."

In order to present the importance of AgTech and Agricultural Big Data for the "traditional" agricultural market players, a list of services and ventures that are related with AgTech was compiled. It is evident that major market stakeholders from the seed, pesticide and agricultural machinery sectors are investing in the future of digital farming:

Market Sector	Company Name	Big data platform/ service/ venture
Seed and Pesticide Companies	Monsanto	Precision Planting
		Climate Corp.
		640 Labs
		Climate Basic/ Climate Pro/ Field Scripts products
	Syngenta	FarmAssist
		AgriEdge Excelsior
		Water+ Intelligent Irrigation Platform

²⁸ Report available here: <https://www.comptia.org/resources/it-industry-outlook-2016-final>

²⁹ AgTech Funding Report 2015: Year in review, www.agfunder.com

³⁰ http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-sme_en.pdf



	DuPont Pioneer	Encirca Services “whole-farm decision service”
		Encirca Yield Stand
		Encirca Yield Nitrogen Management
	Bayer	Bayer Digital Farming
	BASF	Clearpoint Advanced
	Dow	EXZACT Precision Technology platform
Agricultural Equipment Manufacturers and suppliers	Deere	Greenstar
		RTK satellite navigation
		Crop insurance
		Precision Planting
	CNH	Advanced Farming Systems
	AGCO	VarioDoc
		AgCommand
	Kubota	Kverneland
	CLAAS	Efficient Agriculture Systems

Table 1: Key agricultural market stakeholders and their involvement in AgTech and Agricultural Big Data (adapted from ETC Communiqué #115, 2015)

4.2. Agri-food big data value chain

Based on the CGIAR Strategy and Results Framework 2016-2030³¹, the agri-food system is facing daunting challenges. Poor diets are the #1 cause of ill health globally, with 800 million hungry people, 2 billion malnourished people and 159 million stunted children. The global food system is doing a poor job providing healthy food for all. Too much food is lost or wasted, or used for non-food purposes such as biofuels. Increasing productivity of staple cereals is not going to be enough to tackle this problem. Moreover, the food system is also the primary driver of planetary ill health. Soil degradation on 25% of cropped land, deforestation and loss of biodiversity, water scarcity, pollution of lakes and seas, and a contribution to climate change of about 25% of emissions, jointly make agriculture the key risk to manage to keep humanity within a safe planetary operating space. Employment in the agri-food system, on farms or in the food value chain, is still the only realistic option for the 60% of the next generation of African youth that will not migrate to the cities. These challenges are complex and substantial, and cannot be solved through single technology solutions. Addressing them will require a systems approach and contributions from traditional agriculture sectors as well as from health and environmental sciences, to name just three key silos that have to be overcome. The years to come are crucial to address these challenges and this could not be done unless the agri-food value chain changes.

³¹ <http://www.cgiar.org/our-strategy/>



4.2.1. Analysis of the agri-food value chain

All along the agri-food value chain – producers, processors, distributors and consumer packaged goods brands are adapting and transforming in response to the trends of population growth, urbanisation, resource constraints and technology convergence. All agri-food stakeholders share three main objectives: To improve their yields, to improve their assets productivity and to improve the sustainability of their businesses.

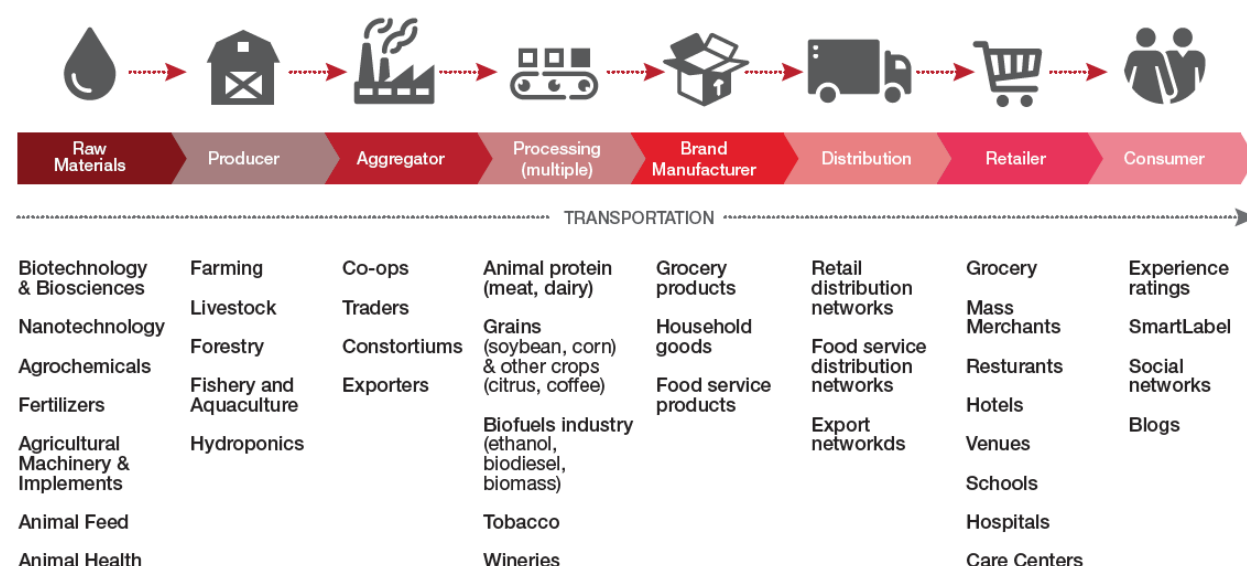


Figure 6: Agri-food value chain (PWC, 2016)

Throughout the value chain, wherever a business stands, it can leverage technology, data, and capability-driven strategies to improve its productivity and efficiently scale business. Along the agricultural value chain, several strategic control points can be identified that are allocated to the steps of different processes; namely input supplying, production, post-harvest, processing and distribution/marketing. As Figure 6 indicates, under each main link of the agri-food value chain, many different subsectors can be found: Raw Materials are the first point of the chain. These are the ingredients for the agri-food production that are exploited by the producers and processors. Raw materials are the main input of the agri-food chain and are strongly connected with machinery, fertilizers and feed suppliers. The next stage of the chain includes production, i.e. farmers, livestock owners, fisheries and aquacultures and greenhouse farmers. One critical point in the chain after the farmer is the aggregator. Aggregators are entities collecting products from farmers (e.g. Cooperatives, traders, exporters) and forwarding them to the parts of the chain that follow. Apart from aggregators, processing and brand manufacturing companies are another critical part of the chain. Given their role, which is to transform one or multiple raw materials to food or other by-products, these two categories contain meat, dairy and crop industries, wineries and biofuels industries as long as fast moving consumer goods (FMCG) companies. After manufacturing stages come distribution and retail.

These two stages of the agri-food value chain are the way to the market for agricultural products. Distributors are responsible for supplying goods to retailers (i.e. groceries, merchants and

restaurants). Retailers are the businesses that sell directly to end customers. Across the aforementioned parts of the value chain in agri-food sector, remarkable amounts of data are being created. An interesting example of consumer-extracted data is the experience ratings and blog or social media posts, where the end customers are sharing their opinion with the community. These data are important for the rest of the parts of the chain because they are valuable insights of what the consumer thinks about their product. Of course, data in itself is not valuable at all. The value is in the analysis done on that data and how the data is turned into information and eventually turning it into knowledge.

As it is displayed below (Figure 7), the agricultural value chain hides many opportunities for innovation applications, which can add value to the processes followed and create new ventures in this sector. Following this approach, the agri-food value chain is analysed based on the volume of the data needed or produced in each phase. Different chain factors are grouped into core categories of input, production, post-harvest, processing and distribution marketing. All of the data produced can be related to the product produced and consequently to the end customers.

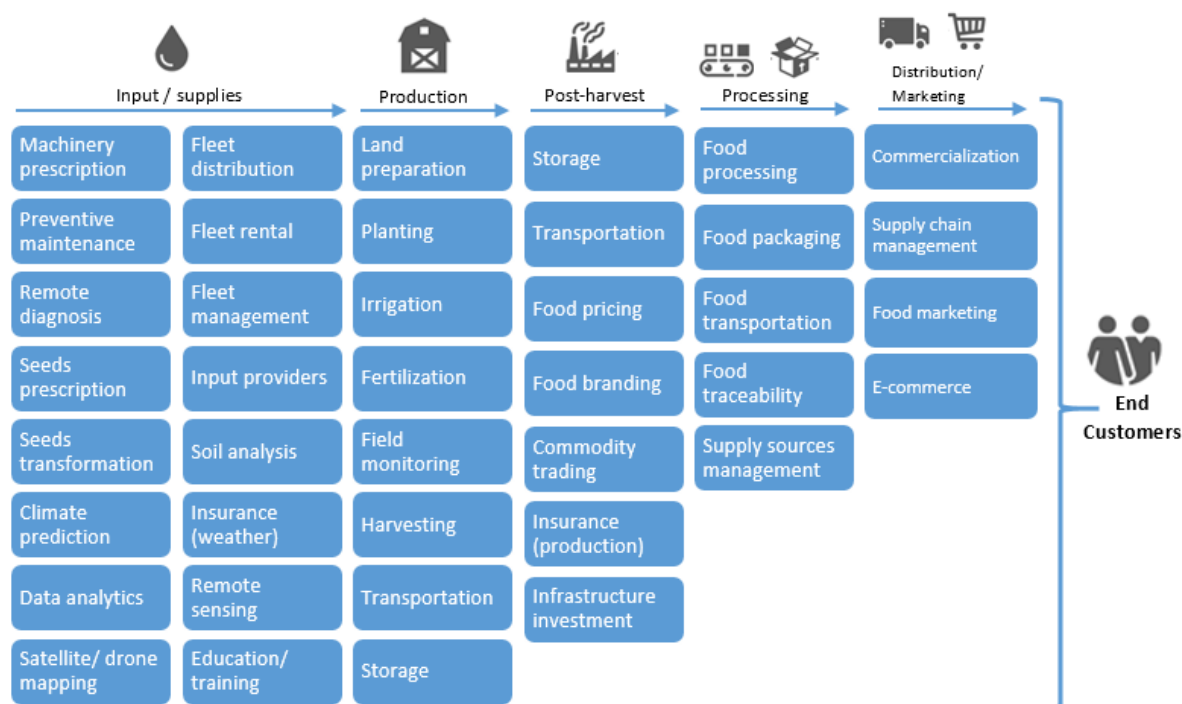


Figure 7: Strategic data related control points in the agricultural value chain (Roland Berger, 2015)

4.2.1.1. Input and supplies related data

As regards the supplies that are needed for the agricultural production, there are many factors that play a significant role at this initial stage of the value chain. Data related to this stage of the production refer to satellite and geospatial data, sensor data, meteorological data and machinery data to name a few:

- Geospatial data refer to data collected from satellites or drones and have explicit geographic positioning information included within it. Geographic Information Systems (GIS) are

incredibly helpful in being able to map and project current and future fluctuations in precipitation, temperature, crop output, and more. By mapping geographic and geologic features of current (and potential) farmland, scientists and farmers can work together to create more effective and efficient farming techniques; this could increase food production in parts of the world that are struggling to produce enough for the people around them. GIS can analyse soil data combined with historical farming practices to determine which could be the best crops to plant, where they should go, and how to maintain soil nutrition levels to best benefit the plants.

- Sensor Data is a broad scheme, which can include a wide range of data coming from various sources. Remote sensing systems can measure solar radiation, sunshine hours, air temperature, dew point, atmospheric pressure, soil moisture, soil pH, and rainfall, which are among the most important data for agriculture. Massive amounts of multitemporal remote sensing data are made available openly by various stakeholders. The opening of the United States Geological Survey's Landsat data archive³², the EU Sentinel mission³³ as well as the EU open data policy enabled the easy access to a record of historical data and related studies on monitoring mainly land cover/land use changes, updating land national cover maps, detect the spatio-temporal dynamics, the evolution of land use change and landscape patterns. On the other side, sensor data are collected by individual farmers, organisations, research institutes and private companies compiling millions of data parcels.
- Meteorological data are of high importance for agricultural production. Agro-meteorological data help to determine the water requirements of different crops growing in the same area, enabling farmers to better plan their growing pattern. For example, using the information obtained from the stations, farmers would be guided on the degree of soil moisture and could decide when their crops would need irrigation, or data on the forecasted timing and amount of impending rain could help determine what measures farmers should take.
- Machinery data relate to data produced from agricultural machinery (e.g. tractors, harvesting equipment). The example of John Deere³⁴ (original equipment manufacturer), which since 2012, has released interconnected agricultural equipment. Machinery is connected not only with each other, but also with owners, operators, dealers and agricultural consultants. On the equipment, sensors are used to help farmers manage their fleet and decrease downtime of their tractors, as well as to save on fuel. The information is combined with historical and real-time weather data, soil conditions, crop features and many other data sets. This type of information helps farmers figure out what crops to plant where and when, when and where to plough, where the best return will be made with the crops and even which path to follow when ploughing.

³² <http://landsat.usgs.gov/>

³³ <https://sentinel.esa.int>

³⁴ <http://www.deere.com>



4.2.1.2. *Production Data*

When it comes to production, farmers and all involved parties, would like to have precise information about the crop, the potential yield and the inputs that are needed (i.e. water, pesticides, and fertilizers) for an efficient management of the farm. The production related data types are mostly sensor-derived data. This means:

- Land and Planting Data are of key importance, because they demonstrate the health of the soil and also are important factors for the crop selection and crop sustainability.
- Irrigation Data display the amount of water needed for farming reasons. By adopting in-farm sensors, farmers are able to combine weather data (e.g. precipitation) with irrigation variables and reduce water usage in farming. For this kind of data, multiple databases are available for global water usage monitoring, supported by FAO (AQUASTAT)³⁵ and the World Bank (water.worldbank.org).

4.2.1.3. *Post-harvest Data*

Post-harvest data are related to the processes that follow the agricultural production stage. As a main aspect of the food supply chain, post-harvest stage includes storage and transportation data, as well as food pricing and commodity data.

- Commodity data: Agricultural commodities are an important variable of farming. Commodity and trading data play a significant role for farmers, because they are indicators of the products' price and consequently of the farmers' potential income. FAOSTAT (faostat.fao.org) is one of the main global databases for agricultural commodity prices, along with the EC Commodity Database³⁶, USDA and the World Bank.
- Transportation data: Post-harvest losses affect farmers' incomes and access to affordable food. Losses from poor storage practices account for an average of 40 percent of total post-harvest loss and based on IBM's data³⁷, 50 percent of food that is ready for harvest never reaches the consumers mouth. By understanding the effect of weather on transportation networks, companies can make better decisions on which routes will be the fastest to transport their food and avoid post-harvest losses.

4.2.1.4. *Processing Data*

Processing includes all actions that are needed to turn an agricultural product into food along with packaging, transportation, logistics and traceability.

- Food traceability data refer to all information collected by the food producer, regard all production and distribution stages and are available to consumers in an open manner. By providing traceability data, consumers and other interested parties are able to determine from where the product originated and its characteristics at any given point in the supply

³⁵ <http://www.fao.org/nr/water/aquastat/main/index.stm>

³⁶ http://ec.europa.eu/agriculture/markets-and-prices/price-monitoring/index_en.htm

³⁷ http://www.research.ibm.com/articles/precision_agriculture.shtml



chain. The new merging technologies in regards to data collection, which enable upstream supply chain partners – specifically the producers – to collect traceability data during every stage of farming operations are getting more attention. These new advancements allow producers to automate and streamline the data during the farming operations. This will minimise data collection and registration inaccuracy allowing for the efficient use of multiple types of information to be captured throughout every stage of farming, transportation and receiving by creating a single point of access to this information.

4.2.1.5. Distribution/ Marketing Data

The final step of the food value chain, before reaching to consumers is Distribution and Marketing. Data in this step regard product commercialisation, management of logistics and supply chain and all market and commerce related data.

- Supply chain data: For a long time, this sector of the food chain has been driven by statistics and quantifiable performance indicators. Given the research and innovation that is being performed in Supply Chain analytics and Big Data, food companies are now able to analyse unstructured data for efficient inventory management, forecasting and transportation logistics. Traditional data monitoring, which would involve sales and order tracking and point of sales data, is nowadays being supplemented with weather, events and news, with the aim being to generate insights in the short term, such as how operations will be affected this week, rather than on a broad, annual timeframe.

4.2.1.6. Consumers Data

Although the need for high frequent real time food price data is undisputed, official food price statistics are typically available monthly, and only at the end of the month or a week after, and rarely with the kind of detail needed for food security early warning, monitoring and policy response. Many policy departments, particularly in developing countries, have adopted big data approaches to have a real time food price monitoring, by ensuring food security and informing their market information and early warning systems. At the same time, many national statistics offices have also started to adopt big data in compiling their official food price statistics, namely, the food and non-alcoholic beverage component of the consumer price index (food CPI), though this has rarely led to more timely statistics. The data types used for computing the food CPI include retail point-of-sale scanner data, data scraped from internet sites, and food price data collected using mobile applications on hand-held devices, such as mobile phones.

4.2.2. The impact of Big Data in agri-food value chain

As it is stated in this report, the AgTech is the key enabler to feed 9 billion people until 2050 and shift societal perceptions of the agricultural industry. A farmer has access to diverse data about his farm and crops. He/she collects information about weather, precipitation, yields and diseases infections, to name a few. He/she knows the effects of these factors but is limited in processing the data collected and take decisions. The agri-food sector is full of examples with large volumes of data that need to be created, stored and visualised in near real-time conditions.



4.2.2.1. Big Data & Precision Agriculture

In recent years, an increasing amount of geospatial data is being generated, from the use of maps via our mobile phones to global positioning system (GPS) driving directions leading to a fundamental shift in the way people are living and are informed. A GIS captures, stores, analyses, manages, and presents geographic location data thereby allowing users to make decisions informed by highly accurate and detailed geographic information. Geospatial information applied to agriculture plays a critical role in providing the right information to key decision-makers about the right practices to improve and to optimise food production. For instance, GIS can analyse soil data combined with historical farming practices to determine the best crops to plant, how they should be planted and how to maintain effective soil nutrition levels to achieve optimum productivity. Precision agriculture (also known as precision farming, satellite farming or site-specific crop management) uses not only GIS and GPS, but also remote sensing and variable-rate technologies to observe, measure and respond to spatial variations in crops, as well as monitor yields. The approach is rapidly changing the way farmers and agri-businesses are looking at crops and relate to the land. Precision data, together with computer-based decision support systems, help optimise production (yield), conserve resources (e.g. water and nutrients) and reduce costs (Venkatalakshmi and Devi, 2014). Figure 8 presents an illustration of this combination, depicting an “integrated farming system” where genetics, physical inputs, sensors and smart machinery are integrated. Through advances in software engineering and environmental testing, farmers are able to create custom field prescriptions for seeds, fertilizers and pest controls. Then, smart machinery will carry out the prescribed treatments and simultaneously collect additional field data that will be provided to the farmer. The same data will allow seed and farm input companies to develop custom products for farmers.

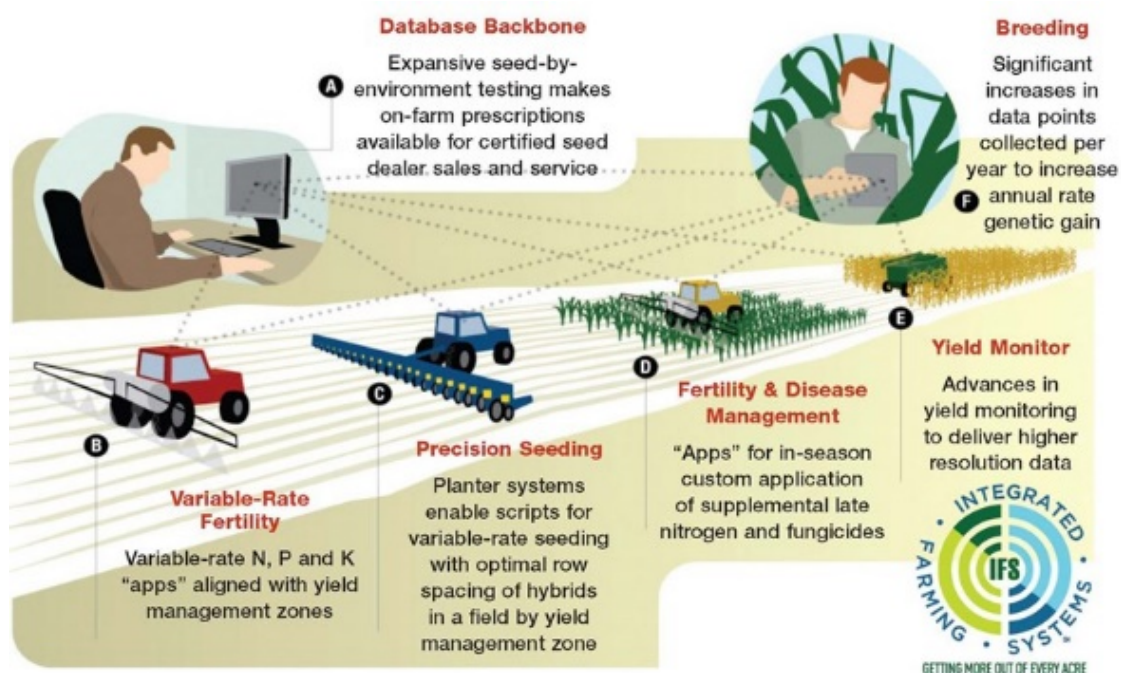


Figure 8: An integrated farming system (Monsanto, 2012)

For many years, precision agriculture was considered irrelevant to small-scale farmers in developing countries because of the coarse resolution and high cost of the images. This has changed. There is now a growing body of research to support the idea that small-scale farmers can benefit from precision agriculture. GPS-equipped sensors on tractors, for example, enable farmers to measure and respond to soil variability across vast tracts of land, and dispense the right amounts of fertilizer and water exactly where it is needed. Multilateral agencies, such as the World Bank and Asian Development Bank, have taken up a more focused approach towards the use of spatial technologies and information for ensuring food security. The UN organisations are promoting Global Geospatial Information Management; the Famine Early Warning Systems Network³⁸, created in 1985 by the US Agency for International Development (USAID), is a leading provider of early warning and analysis on acute food insecurity; the Dutch government, through the Netherlands Space Office, has started a programme called Geospatial for Agriculture and Water. This initiative uses satellite data to improve food security and has projects in Bangladesh, Ethiopia, Indonesia, Kenya, Mali, as well as Uganda, where satellite generated data will be used to improve production and marketing prospects for producers involved in maize, soya bean and sesame value chains. The Copernicus³⁹ Land Monitoring Service also makes use of satellite and in situ data to provide regular geospatial information on the state of global vegetation and water cycle for spatial planning, forest management, water management, food security and agriculture.

Integrated farming systems can take advantage of a wide range of applied big data services, which can create a remarkable value chain in the agri-food sector. The emergence of big data can usher a new era of Precision Agriculture. The accessibility and commercial availability of Unmanned Aerial Systems (commonly referred as “Drones”) increases a farm’s data collecting capabilities. Furthermore, sensors that are monitoring variables such as soil moisture, nutrients and pests enable the collection of large amounts of heterogeneous data. The need that is automatically created regards to cloud-enabled services that are specialised in managing big volumes of data and feeding decision-making applications that can support farmers. Such applications, which can process all factors of the agricultural production, can maximise crop yields, minimise waste and water usage, while ultimately lower production costs and increase product quantity.

4.2.2.2. Big Data & Internet of Things farm applications

In an interconnected and sensor enabled farm, cloud services can become the central hub of this large ecosystem. Through the data stream coming from various devices installed in the farm, as an example, farmers are enabled to evaluate the moisture content of the soil and dispense only the amount of water that is needed based on the needs of every separate plant.

³⁸ <http://www.fews.net/>

³⁹ <http://www.copernicus.eu/main/land-monitoring>



4.3. E-infrastructures in agri-food sector

The variety of stakeholders that are involved in scientific activities that address major societal challenges around agriculture, food and the environment is enormous. They range from researchers working across geographical areas and scientific domains, to policy makers designing development and innovation interventions. These activities have been traditionally informed and powered by a combination of quite heterogeneous data sources and formats, as well as several research infrastructures and facilities, at a local, national, and regional level. In 2010, a SCAR study tried to give an overview of this picture, which has been documented in the report “Survey on research infrastructures in agri-food research”⁴⁰ (2010). As more and more research information and IT systems became available online, the relevance of agricultural knowledge organisation schemes and systems became higher. A recent foresight paper on the topic has been published by the SCAR working group “Agricultural Knowledge and innovation systems”⁴¹. The emergence of the open access and data policies has brought forward new challenges and opportunities (as the 2015 GODAN Discussion Paper⁴² has revealed), which have to be addressed and supported by future e-infrastructure services and environments. In addition to this, Commissioner Moedas pointed out a clear link between societal challenges and openness as a European priority. He has positioned these challenges across the three dimensions of Open Science, Open Innovation and Openness to the World.

“Challenges in areas like energy, health, food and water are global challenges. And Europe should be leading the way in developing global research partnerships to address these challenges...”

Although Europe generates more scientific output than any other region in the world, in some areas we fall behind on the very best science. At the same time, there is a revolution happening in the way science works. Every part of the scientific method is becoming an open, collaborative and participative process.

...I see three strategic priorities: Open Innovation, Open Science, and Openness to the World.”

Carlos Moedas - Commissioner for Research, Science and Innovation, 22 June 2015

4.3.1. Towards an agricultural data e-infrastructure

How an agricultural data e-infrastructure (ADEI) could be best positioned to achieve impact upon all the above dimensions stated above by Commissioner Moedas? The following sections provide a description of how an ADEI fits in these three strategic characteristics.

4.3.1.1. Open Science

The Belmont Forum is a roundtable of the world’s major funding agencies of global environmental change research and international science councils, which collectively work on how they may address the challenges and opportunities associated with global environmental change. In 2013,

⁴⁰ https://ec.europa.eu/research/scar/pdf/final_scar_survey_report_on_infrastructures.pdf

⁴¹ http://ec.europa.eu/research/scar/pdf/akis-3_end_report.pdf

⁴² <http://www.godan.info/launch-of-godan-discussion-paper>



the Belmont Forum initiated the multi-phased E-Infrastructures and Data Management Collaborative Research Action. In August 2015, this initiative published its recommendations on how the Belmont Forum can leverage existing resources and investments to foster global coordination in e-infrastructures and data management, in a report entitled "*A Place to Stand: e-Infrastructures and Data Management for Global Change Research*". The main recommendations of this report included the following:

- Adopt data principles that establish a global, interoperable e-infrastructure with cost-effective solutions to widen access to data and ensure its proper management and long-term preservation.
- Promote effective data planning and stewardship in all research funded by Belmont Forum agencies, to enable harmonisation of the e-infrastructure data layer.
- Determine international and community best practices to inform e-infrastructure policy for all Belmont Forum research, in harmony with evolving research practices and technologies and their interactions, through identification of cross-disciplinary research case studies.

An ADEI should be fully aligned with these recommendations and should have as its strategic impact goal to ensure that Europe brings forward to the Belmont Forum a world-class data infrastructure, demonstrating a number of cost-effective and operational solutions for access, management and preservation of research data that will be based upon (and take advantage of) the core e-infrastructure services at hand.

4.3.1.2. *Open Innovation*

The agricultural business landscape is rapidly changing. Established brands in agriculture such as John Deere, Monsanto, and DuPont are now as much data-technology companies as they are makers of equipment and seeds. Even though agriculture has been slower and more cautious to adopt big data than other industries, Silicon Valley and other investors are taking notice. Startups like Farmers Business Network, which counts Google Ventures as an investor, have made collecting, aggregating, and analysing data from many farms their primary business. Popular, business and tech press constantly highlight the evolution that (big) data brings into the agriculture, food and water business sectors – but also into helping feed 9 billion people. For instance, in the farming sector, data collection, management, aggregation and analytics introduce a wide variety of innovative applications and solutions:

- Sensors can tell how effective certain seed and types of fertilizer are in different sections of a farm
- Software may instruct the farmer to plant one hybrid in one corner and a different seed in another for optimum yield
- Intelligent systems may adjust nitrogen and potassium levels in the soil in different patches
- and this information can also be automatically shared with seed companies to improve hybrids.

This is also creating an investment environment with a tremendous potential for startups and companies that are focusing on data-intensive applications.



An ADEI should aim to take advantage of this investment trend by targeting and involving agriculture and food data-powered companies (and especially startups and SMEs). It should be dedicated on getting such companies involved, and aligning its efforts with the business outreach (through data challenges, hackathons, incubators and accelerators) of European and global networks, such as the Open Data Institute (ODI)⁴³, Big Data Value Association (BDV)⁴⁴ and Global Open Data for Agriculture and Nutrition (GODAN)⁴⁵.

4.3.1.3. Openness to the World

At the 2012 G8 Summit, G8 leaders committed to the *New Alliance for Food Security and Nutrition*, the next phase of a shared commitment to achieving global food security. As part of this commitment, they agreed to “share relevant agricultural data available from G8 countries with African partners and convene an international conference on Open Data for Agriculture, to develop options for the establishment of a global platform to make reliable agricultural and related information available to African farmers, researchers and policymakers, taking into account existing agricultural data systems”. In April 2013, the prestigious G8 International Conference on Open Data for Agriculture took place in Washington DC, announcing the G8 Open Data Action plans. The goal of the EC’s action plan has been “*Open access to publicly funded agriculturally relevant data*” and included flagship initiatives (such as agINFRA⁴⁶, SemaGrow⁴⁷, TRANSPLANT⁴⁸, OpenAIRE⁴⁹).

An ADEI should support the global outreach and collaboration of European agriculture and food data stakeholders with their international counterparts like GODAN, the Research Data Alliance (RDA)⁵⁰ and the Coherence in Information for Agricultural Research and Development (CIARD).

4.3.2. Case studies on research data e-infrastructures

A number of initiatives are already active in a pan-European level for developing core cloud infrastructures or data archiving applications on top of these cloud infrastructures.

4.3.2.1. OpenAIRE

OpenAIRE⁵¹ is an EC FP7-funded initiative, which has built a European scholarly communication infrastructure that enables the aggregation of open access publications and research data catalogues, and importantly, has linked them to funding streams. On top of this graph of information, it provides services.

⁴³ <http://theodi.org/>

⁴⁴ <http://www.bdva.eu/>

⁴⁵ <http://www.godan.info/>

⁴⁶ <http://aginfra.eu/>

⁴⁷ <http://semagrow.eu>

⁴⁸ <http://www.transplantdb.eu/>

⁴⁹ <https://www.openaire.eu/>

⁵⁰ <https://rd-alliance.org/>

⁵¹ <http://www.openaire.eu/>

The origins of OpenAIRE lay in DRIVER (Digital Repository Infrastructure Vision for European Research), which created an e-infrastructure of connected repositories, usually located in research institutions and libraries. OpenAIRE is a network of Open Access repositories, archives, and journals that supports Open Access policies. Importantly, the infrastructure follows a participatory approach with a European footprint employing 32 National Open Access Desks (all member states and 5 associate countries).

Currently, the project is growing its range of publications, and links out to associated data and funding information.

4.3.2.2. EUDAT

The European Data Infrastructure (EUDAT) was launched to target a pan-European solution to the challenge of data proliferation in Europe's scientific and research communities.

EUDAT designs, develops, implements and offers “Common Data Services” as they have been introduced in the “Riding the Wave report”⁵² to all interested researchers and research communities. These common data services share three common characteristics:

- a) They are relevant to several communities
- b) They are available at European level

They are identified by a high degree of openness: **(1)** they follow the Open Access as the default principle; **(2)** they are Independent of specific technologies since these change frequently and **(3)** they are flexible to allow new communities to be integrated (which is not a trivial requirement given the heterogeneity and fragmentation of the data landscape).

However, EUDAT is mainly addressing long-term data preservation/archiving.

4.3.2.3. D4Science

D4Science⁵³ (Distributed collAboratories Infrastructure on Grid ENabled Technology 4 Science) is an initiative co-funded by European Commissions Seventh Framework Programme for Research and Technological Development. Its major outcome is a production e-infrastructure enabling on-demand resource sharing across organisations boundaries. Through its capabilities, this e-infrastructure accelerates multidisciplinary research by overcoming barriers related to heterogeneity (especially related to access to shared content), sustainability and scalability.

In the context of D4Science, content can be of very heterogeneous nature, ranging from textual and multimedia documents to experimental and sensor data, to images and other products generated by elaborating existing data, to compound objects made of heterogeneous parts. The D4Science e-infrastructure also supports the notion of Virtual Research Environments (VREs), i.e. integrated environments providing seamless access to the needed resources as well as facilities for communication, collaboration and any kind of interaction among scientists and researchers.

⁵² <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>

⁵³ <https://www.d4science.org/>



VREs are built by dynamically aggregating the needed constituents, i.e. data collection, services and computing resources, after on-demand hiring them through the e-Infrastructure.

4.3.2.4. *AGINFRA: a data e-infrastructure on agriculture*

One of the pioneering efforts towards the development of an agricultural data e-infrastructure is realised with AGINFRA, an initiative that builds upon the extensive experience and work of its partners, who are key stakeholders in the e-infrastructures ecosystem such as the *OpenAIRE* (for publication and data set aggregation, indexing, mining and disambiguation), *EUDAT* (for cloud-hosted preservation and storage), *EGI* (with cloud compute and high-throughput compute for intensive computational applications), and *D4Science* (for data analytics). It also implements part of a strategic vision shared between stakeholders that are part of a core group of internationally recognised players aiming to put in place a free global data infrastructure for research and innovation in agriculture, food and environmental science. Indicative stakeholders are the National Agronomic Research Institute of France (INRA), the Alterra Institute of the Wageningen University & Research Center (ALTErra), the National Institute for Risk Assessment of Germany (BfR), and the Food and Agriculture Organization (FAO) of the United Nations and the Chinese Academy of Agricultural Sciences.

AGINFRA addresses the challenge of supporting user-driven design and prototyping of innovative e-infrastructure services and applications. It particularly meets the needs of the scientific and technological communities that work on the multi-disciplinary and multi-domain problems related to agriculture and food. It uses, adapts and evolves existing open e-infrastructure resources and services, in order to demonstrate how fast prototyping and development of innovative data- and computing-intensive applications can take place.

AGINFRA is evolving and developing further the resources and services of agricultural data infrastructure projects. The existing core components include the:

- Federated data and software registry of CIARD RING (<http://ring.ciard.info>),
- AGINFRA API gateway for indexing and hosting executable software components for advanced data processing & analysis (<http://ring.ciard.info/views/aginfra>),
- Open source software stack for data analysis, indexing, publication and querying developed by projects such as FP7 SemaGrow and H2020 Big Data Europe (<http://www.big-data-europe.eu>),
- Semantic backbone of the Global Agricultural Concept Scheme (GACS54) that has been based upon the alignment of FAO's AGROVOC with the USDA's National Agricultural Library Thesaurus and CABI's Thesaurus,
- Advanced research data set processing & indexing demonstrators developed within FP7 SemaGrow for specific scientific communities such as Trees4Futures⁵⁵ and AgMIP⁵⁶.

⁵⁴ <http://aims.fao.org/activity/blog/global-agricultural-concept-scheme-gacs-collaborative-integration-three-thesauri>

⁵⁵ <http://www.trees4future.eu>

⁵⁶ <http://www.agmip.org>

4.4. Agri-food SMEs case studies

AgTech evolution is just at the initial stages, but this sector has already proven examples of successful business cases. As it was stated in Chapter 4.1, the main enablers of this positive growth are mostly tech companies (SMEs and startups), which despite their short lifetime, present an impressive investments track record. Two cases (Climate Corporation, AgDNA) are presented as drivers of the AgTech success stories, since both present remarkable growth numbers.

4.4.1. The case of The Climate Corporation

One of the major investments made in AgTech was the recent purchase of The Climate Corporation⁵⁷ by Monsanto⁵⁸ for \$930 million⁵⁹. This acquisition represents the potential of agricultural Big Data for the key players of the agricultural industry. The Climate Corporation's expertise in agriculture analytics and risk-management supports Monsanto's R&D, providing farmers access to a wide variety of factors that affect the success of their crops. This unlocks a new value for the farm through data science. The company's proprietary technology platform combines hyper-local weather monitoring, agronomic data modelling, and high-resolution weather simulations to deliver climate.com, a software-as-a-service (SaaS) solution that helps farmers improve their profits by making better informed operating and financing decisions, and Total Weather Insurance, an insurance offering that pays farmers automatically for bad weather that may impact their profits. The company is also an authorised provider of the U.S. Federal crop insurance program, enabling authorised independent crop insurance agents to provide farmers with the industry's most powerful full-stack risk management solution. In the face of increasingly volatile weather, the global \$3 trillion agriculture industry depends on the company's unique technologies to help stabilise and improve profits and, ultimately, help feed the world.

4.4.2. The case of AgDNA

This Australian based company is providing a platform-as-a-service (PaaS) to large agricultural equipment dealers as a branded solution. This platform enables farmers the ability to make informed decisions about their entire farming operation. AgDNA⁶⁰ has signed data licensing agreements with major tractor manufacturers - John Deere, Case New Holland, and AGCO – and also combines user-submitted data (e.g. smartphones and tablet images), GPS data and other spatial information. They provide a PaaS to large agricultural equipment dealers as a branded solution. So, it has the dealer logo and branding front and center and they in turn sell that as their own data management solution to their grower clients and on a per acre basis. This means the dealer can offer a system that integrates seamlessly with the machinery that's already on the farm

⁵⁷ <https://www.climate.com/>

⁵⁸ <http://www.monsanto.com>

⁵⁹ <http://news.monsanto.com/press-release/corporate/monsanto-acquire-climate-corporation-combination-provide-farmers-broad-suite>

⁶⁰ <https://agdna.com/>



they sell to the growers. AgDNA delivers all the data via the cloud and that is available either through an online browser or a native app in the smartphone or tablet for both Android and Apple devices. The primary method of sales is business-to-business (B2B) and their service already has tens of thousands of registered users in 164 countries encompassing 4.3 million acres of fields under management. AgDNA was founded on the basis of making growers' lives easier and more profitable. AgDNA's founder believes that local dealers are an integral part of the whole farming community and a key stakeholder in the farmers' world. This is the main reason behind their go-to-market strategy. The other real key for big data is that it has to be seamless and fully integrated across the entire farm. Following a holistic approach, AgDNA is looking at the beginning of the season from initial planning and tillage ops, all the way through planting, spraying, crop care, and finally harvesting.

4.4.3. European case studies

As it is stated on previous chapters of this analysis, Europe does not have billion dollar deals and acquisitions to demonstrate. There is a remarkable number of companies active in the agri-food industry, which attract investors' interest. Based on a recent article by Victoria Vzyatysheva in *tech.eu*⁶¹, Europe has interesting AgTech startups and SMEs focused on catering to the needs of the global agriculture industry. Some of these are described below.

4.4.3.1. Vital Fields, Estonia

Vital Fields (<http://vitalfields.com/>) is an Estonian startup that helps farmers manage their fields. But more than just manage them, it aims to offer advice on when to plant, harvest and offer guidance on pest management and use of chemicals. The company provides a field-by-field seven-day weather forecast, which enables plant growth predictions, constant field monitoring and effective disease management. Since 2011, when Vital Fields was founded, it raised €1.2 million through private investors. Currently, Vital Fields claims to be the "fastest growing farm management system in Europe" with over 200.000 hectares under management.

4.4.3.2. Gamaya, Switzerland

Gamaya (<http://www.gamaya.com/>) responds to the need to feed 10 billion people by developing a unique solution for large scale analytics of farmland based on hyperspectral imaging and Big Data technologies. They address specific needs of industrial growers of commodity crops. Gamaya is a Swiss AgTech company – spin-off from the Swiss Federal Institute of Technology (EPFL)⁶², created in 2015. Gamaya addresses commodity crops cultivation challenges by providing an integrated solution for large-scale diagnostics of field crops to boost yield productivity. Their unique solution facilitates an unprecedented level of situational awareness and helps to increase profit margin per hectare of industrial farmers up to 100%. The company's core offering for industrial agricultural producers is early detection of diseases, pests and weeds, detection and diagnostics of stress (mechanical damage, nutrient deficiency, water stress, soil compaction),

⁶¹ <http://tech.eu/features/5480/10-european-agritech-startups/>

⁶² <https://www.ethz.ch/en.html>



monitoring of growth for optimisation of fertilization, as well as prediction of yield. Gamaya is a knowledge company that empowers farmers to know better their land and plant health using unique hyperspectral imaging, data mining and analytics technology. Using state-of-the-art hyperspectral sensors, they acquire information-rich data that they translate into actionable information for agricultural businesses using a simple, scalable and cost effective remote sensing methodology.

4.4.3.3. ECF Farmsystems, Germany

ECF (<http://www.ecf-farmsystems.com/>) designs and builds aquaponic farm systems. These are a solution to efficient food production as the water is used twice: first for fish, then for vegetable production. Furthermore, waste products from fish are used to fertilise the plants and emissions from transportation and cooling chains are avoided. ECF is currently building Europe's largest urban aquaponic farm in Berlin. From 2015, the 1.800m² site will produce approx. 25 tons of fish and 35 tons of high-quality vegetables. Within city boundaries, ECF offers Partnerfarms. In the countryside, ECF offers planning and construction of ECF industrial farms.

4.4.3.4. Farmeron, Croatia

Farmeron (<http://farmeron.com/>) is a Web data service that farmers can use to aggregate the troves of information produced about their animals: diet, health, reproduction, milk production and medicine or drug dosage. Farmers have always had immense amounts of data, but little training or tools to analyse it. Farmeron was founded in 2012 by current CEO Matija Kopic to assist his family in managing their farming operation. Interest peaked in Farmeron's unique approach to holistically managing a farm, not just as a tool, but a complete system. Today, Farmeron is being sold worldwide. By bringing farmers and their production data online, the company is rethinking how agriculture must meet the newest global challenges.

4.4.3.5. CropX, Israel

CropX Ltd. (<http://www.cropx.com>), an agricultural analytics company, develops cloud based software solutions integrated with wireless sensors through an advanced adaptive irrigation software service that delivers crop yield increase, and water and energy cost saving services while conserving the environment. The company also generates irrigation maps and automatically applies the right amount of water to different parts of the same field. It serves farmers worldwide. CropX Ltd. was founded in 2013 and is headquartered in Tel Aviv, Israel with an additional office in San Francisco, California.



5. Mapping and validating the SMEs requirements

This chapter presents the analysis of the requirements that were elicited from key stakeholders of the wider AgTech sector.

5.1. Methodology

In order to collect and assess the SMEs and startups' requirements, the Lean methodology⁶³ was followed for the elicitation of the individual target group's needs. As the first step in requirements collection, information is gathered for subsequent analysis, modelling and validation. The process is called "Requirements Elicitation" and its goal is to identify stakeholders, underlying problems that need to be addressed, and as a result identify system boundaries and goals. Stakeholders are individuals or organisations affected by the system (e.g. software engineers, startups, SMEs) and are divided into separate classes. Goals define the objectives that EGI-Engage has to meet and are broken into high-level and lower-level goals. The focus in requirements elicitation is purely on the problems and needs of the stakeholders and does not look into the solution domain. The process that is followed consists of two phases of requirements elicitation. The initial phase includes a first validation test of the assumptions with a small sample of the intended user segment. After the initial round, the validated challenges are well defined and potential new challenges that are expressed by the sample of the targeted audience are taken into account. Since the challenges list has passed a first validation and review, the updated challenges should be validated by the wider segment of potential users. This process defines the second circle of interviews and challenge validation, where the final requirements and challenges is expressed and verified by the overall AgTech community sample.

5.1.1. Targeted user segment

Based on the key market stakeholders' analysis performed in chapter 4.1.3, the segments that matters the most for EGI are the Startups, SMEs and individual developers. Following the Lean approach, the main actors of the agri-food sector that ideally have enough understanding of the EGI e-infrastructure and are related to data management and technical decision making are the AgTech companies, which play a key role in the sector's development and successfully combine agricultural and ICT expertise.

In order to receive feedback and investigate the challenges that these particular AgTech segments are facing, multiple events were attended (see table below) and important AgTech communities were contacted in order to interact with a wide spectrum of users. The following tables contain the channels where the members of the targeted segments reside:

⁶³ "The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses", Eric Ries, Crown Business (September 13, 2011), ISBN-10: 0307887898, ISBN-13: 978-0307887894

Event name	Event website	Targeted Personas
Linked Data and Analytics for SMEs, Workshop & Hackathon 2015	http://linda-project.eu/workshop2015/	SMEs, Individual developers
Startup Lab Orange Grove	http://athens.startupsafary.com/	Startups
Finish Food Safety Challenge	http://www.finish-project.eu/finish-food-safety-challenge/	Startups, SMEs
BioHorizon Food Brokerage Event 2015	https://www.b2match.eu/foodbrokerage2015	SMEs
EIP Agri Seminar on Data Revolution	https://ec.europa.eu/eip/agriculture/Seminar_Data_Revolution	Startups, SMEs
European Data Forum 2016	http://2016.data-forum.eu/	SMEs, Individual developers

Table 2: Targeted personas in AgTech events for requirements elicitation

Both Table 2 and Table 3 include representative events and groups of the wider AgTech community. Owing to the fact that the targeted user segment of the survey consists of SMEs, startups and individual developers, the events and communities that were selected play a major role in the European startup and SME scene and also attract key players of the AgTech sector.

Targeted community	Targeted Personas
FIWARE Accelerator members	Startups
Smart Agri-food Accelerator members	Startups, SMEs
SME Instrument members	Startups, SMEs
AGINFRA community members	Individual developers, Startups, SMEs

Table 3: Targeted communities for requirements elicitation

5.1.2. Questionnaire

The questionnaire that was initially structured contained a total of 21 questions and aimed to understand the SMEs and startups worldview in terms of their needs in Virtual Machines (VMs), virtualisation technologies and open platforms.

The initial questions, which were part of the survey, can be found in the Annex of this report. After the first round of interviews, and with a clear understanding of the needs of startups, SMEs and individual philosophies, the final questionnaire was reformed. The questionnaire was structured in three sections, which aimed to investigate multiple aspects of the interviewee and her/his involvement, requirements and everyday challenges regarding e-infrastructures in AgTech. The structure of the survey is based on four sections. The first section aimed to build the profile of the interviewee and understand the persona to which she/he belongs. The section that follows regards the company that the interviewee represents and business related data. The third section



maps the requirements of the interviewee and her/his company in terms of cloud services and e-infrastructures. The fourth and final section asks the interviewee to select the top challenges that are being addressed in terms of e-infrastructures management and cloud services.

An overview of the 13 questions that were included in the final version of the survey can be found below – minus the strictly related to the individual questions, e.g. Name, email and company name:

First section: Persona Profile & Information

- Question 1: Which of the following describes better the type of your company? (close-ended question)
- Question 2: What is your role in your organisation/ company? (close-ended question)

Second section: Persona Profile & Information

- Question 3: What is the status of your application? (close-ended question)
- Question 4: Do you have a defined business plan for your application? (close-ended question)
- Question 5: In which agricultural-specific thematic area(s) belong your software/ application? (close-ended question)
- Question 6: What types of information are needed for your software/ application? (close-ended question)

Section three: Requirements and details about cloud services and e-infrastructures

- Question 7: If you are maintaining your own cloud, provide details in terms of number of CPUs, memory and storage. (open-ended question)
- Question 8: If you are using virtualization technologies and open platforms (e.g. Docker) for setting up and managing new VMs images, please give us more details. (open-ended question)
- Question 9: For your software/ application, which are the typical usage scenarios of the cloud? (close-ended question)
- Question 10: How often should the required data be updated for your application? (close-ended question)

Section four: Top challenges

- Question 11: Which of the following are the top challenges that should be addressed, in order to serve your users in a better way? (close-ended question)
- Question 12: In which format could you use the provided information? (close-ended question)
- Question 13: In terms of access to information through API, which of the following is the most preferable? (close-ended question)



5.2. Personas Analysis

Based on the survey outcomes, a persona profile was created in order to depict the characteristics of the profile of each stakeholder representative. Questions 1 and 2 aimed to understand the type of the company and the role of the interviewee in it, whereas questions 3 and 4 reflected the business aspects of the company and the product.

“A persona represents a cluster of users who exhibit similar behavioural patterns in their purchasing decisions, use of technology or products, customer service preferences, lifestyle choices, and the like. Behaviours, attitudes, and motivations are common to a “type” regardless of age, gender, education, and other typical demographics. In fact, personas vastly span demographics.”

Kevin O’Connor, Personas: The Foundation of a Great User Experience, 25 March 2011

The sample consisted of 25 representatives of the agri-food ecosystem. The majority (70%) of the AgTech community sample that was taken into consideration, consisted of SMEs, startup companies and individual consultants. The personas that were identified were mostly technical representatives of the company – including web developers, software engineers and system administrators. The aforementioned types of employees have a clear understanding of the requirements and the issues that a company may face, in terms of e-infrastructures and cloud services.

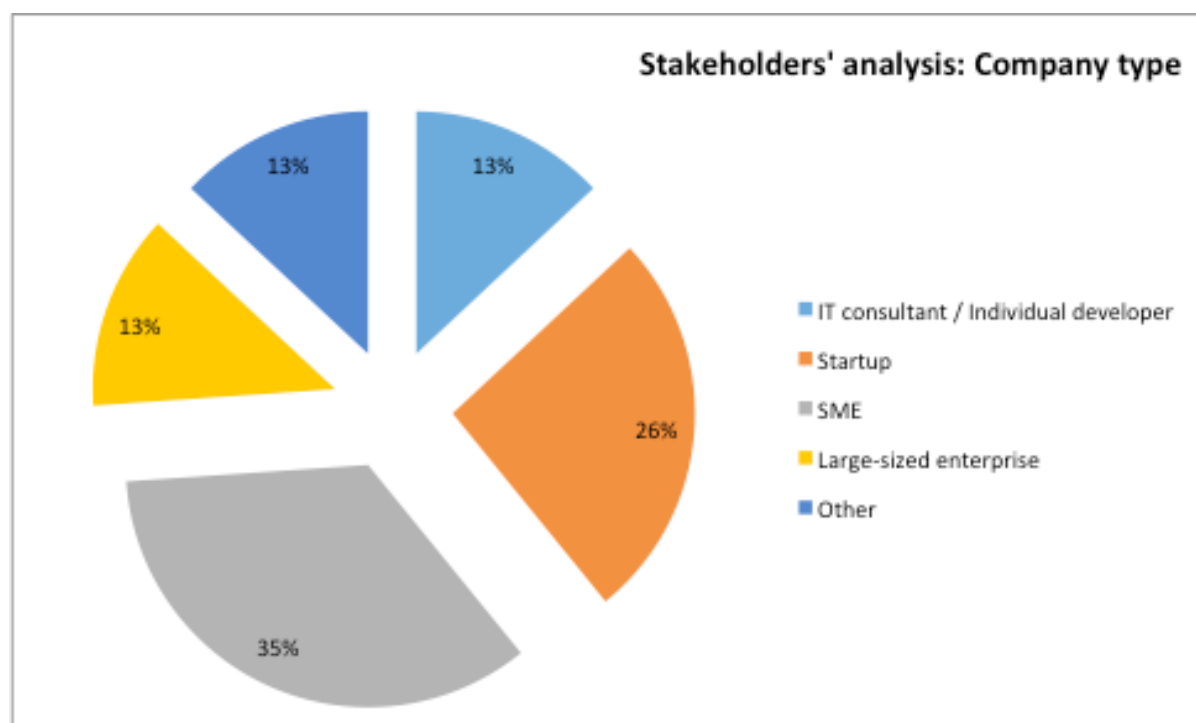


Figure 9: Which of the following describes better the type of your company? (Question 1)

The figures that follow illustrate the profiles of the interviewed stakeholders. The targeted user segment consists of technical representative of AgTech SMEs and startup companies, or individuals who support similar companies as technical consultants. The data collected indicated that the group of people in the selected target segment has the following characteristics:

- Have a technical profession and work either as software engineers, web developers or systems administrators. Also there were cases of people with technical expertise working in the business development department of the company, or as high-level executives (Directors, Chief Operation Officers, and Chief Executive Officers).
- The status of the software product or application that they are working on is by majority an already existing product, which indicates that the company that they are working has a solid business model.
- A large number of people, 36%, is working on a new product concept and this may refer to the startups or the individual developers/consultants that participated in the survey. In that case, the amount of 39%, which indicates the companies that do not have a specified business model, represents mostly this part of the targeted user segment (AgTech startups and individuals).

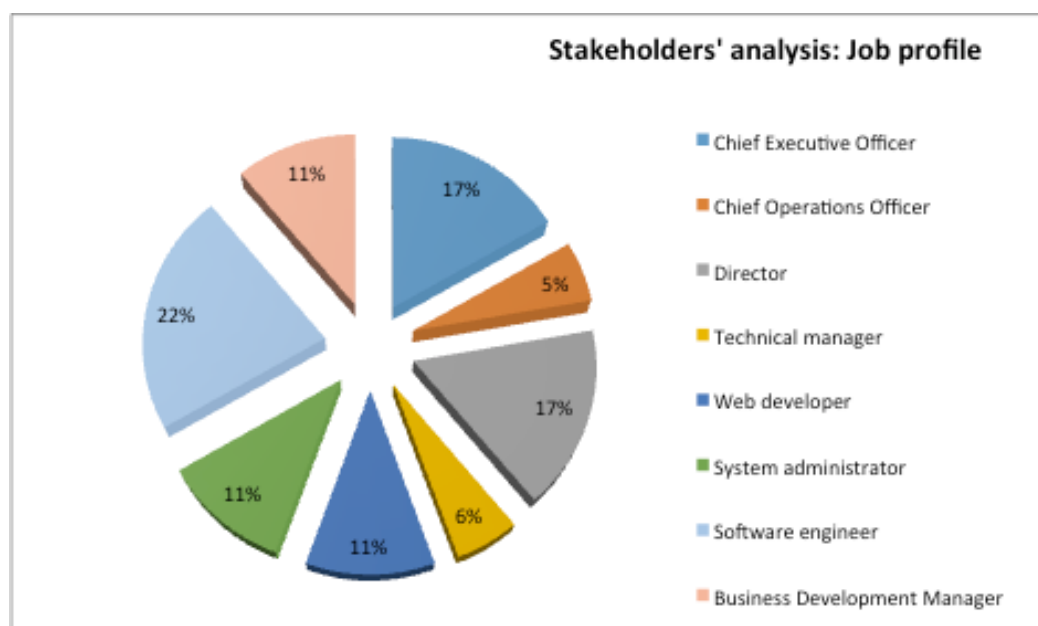


Figure 10: What is your role in your organisation (Question 2)

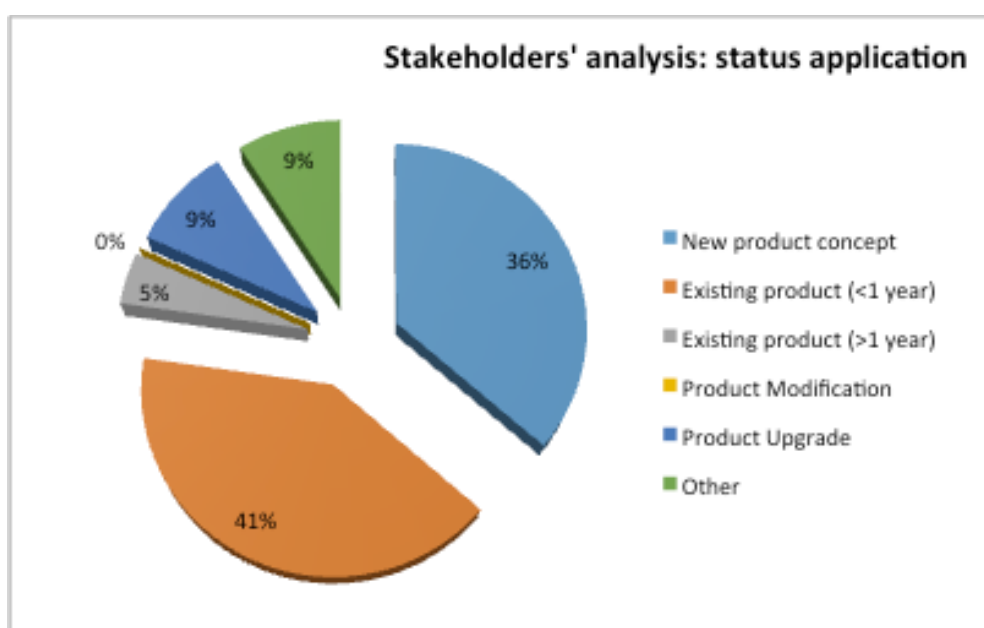


Figure 11: What is the status of your application? (Question 3)

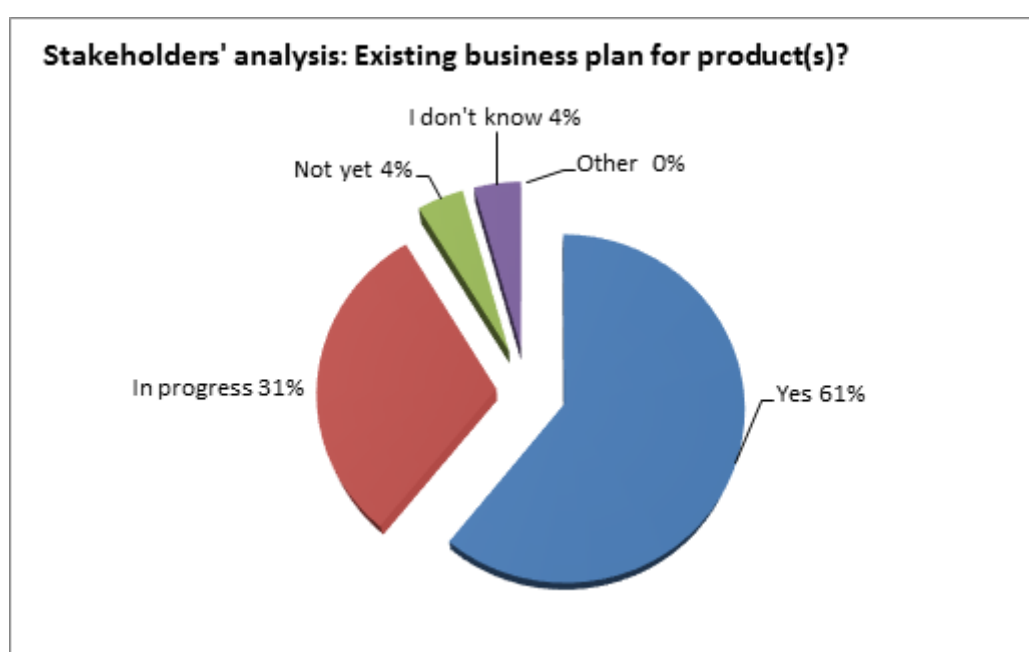


Figure 12: Do you have a defined business plan for your products? (Question 4)

Based on the first four questions of the survey, a respective profile was created for the top 3 personas, namely ICT consultant/individual developer, startups and SMEs:

Persona 1: IT consultant/individual developer

The role of individual developer/IT consultant, regards individuals who work in partnership with their clients advising them how to use information technology in order to meet their business objectives or overcome problems. They work to improve the structure and efficiency of IT systems

in various organisations. IT consultants may be used to provide strategic guidance to clients with regard to technology, IT infrastructures and enabling major business processes through enhancements to IT. They can also be used to provide guidance during selection and procurement as well as providing expert technical assistance, and may be responsible for user training and feedback.

Persona 2: Startups

Representing 25% of the total sample, startup companies are one of the main target personas that were approached. According to Forbes Magazine⁶⁴, a startup is an entrepreneurial venture typically describing newly emerged, fast-growing business. Usually refers to a company, a partnership or an organisation designed to rapidly develop scalable business model. Often, startup companies deploy advanced technologies and are generally involved in the design and implementation of the innovative processes of the development, validation and research for target markets. A confluence of hardware and software technology advances are creating opportunities that need to be addressed by AgTech startups, while inexpensive, but sophisticated hardware sensors have emerged to automate the collection of massive data sets.

Persona 3: Small-Medium sized enterprise (SME)

According to European Legislation⁶⁵, the category of micro, small and medium-sized enterprises (SMEs) is made up of enterprises that employ fewer than 250 persons and which have an annual turnover not exceeding €50 million, and/or an annual balance sheet total not exceeding €43 million.

5.3. Top challenges identification

The Lean methodology requires having a first round of assumptions that will be tested with a sample of the targeted community. In the AgTech case, the initial survey contained 14 hypothetical challenges that the tech personas could be facing. These hypotheses were based on the needs that an SME, startup and/ or an individual consultant may face in terms of serving their users/customers effectively by using cloud-enabled services.

	Which of the following are the top challenges that should be addressed, in order to serve your users in a better way?
1	Difficulty and time consuming to store, manage and use large amount of data for your application.
2	It is hard to find information in machine-readable format, in terms of who does generate & publish such information
3	It is time consuming to search for related information (lots of search pages are needed)
4	Lack of technical expertise or resources in processing the specific formats of data
5	It is hard to combine similar data from various sources
6	Difficulty to get periodically updates and fresh data

⁶⁴ <http://www.forbes.com/sites/natalierobehmed/2013/12/16/what-is-a-startup>

⁶⁵ Article 2, Recommendation 2003/361/EC



7	Difficulty on accessing the public and open data (e.g. behind the firewalls)
8	There is a need for elastic cloud services that allow you to extend your infrastructure by adding custom processing power, storage and memory.
9	It is hard to run VM images with pre-installed software stack (e.g. CMS installation) with only one click
10	You need to distribute VM images with pre-installed software to cloud providers with only one click
11	It is difficult to manage VMs (e.g. upgrade, restart, monitor, back up)
12	VMs are not a good abstraction for your software, you need other higher level services that eases the development and deployment process
13	It is hard to monitor and define alerts for the health of your VMs (e.g. to set up limits for CPU, for memory)
14	It is hard to dynamically adapt your infrastructure to the current load with automatic scaling
15	Other:

Table 4: Hypothetical challenges for targeted segments

The initial list of requirements was tested with a small sample of the total AgTech startup and SME segment. This process helped to understand what assumptions are validated by the potential users and also to investigate if there are any other additional challenges that were not considered during the initial assumption circle.

After the first round of top challenges identification and validation, the final list is comprised by the following:

	Which of the following are the top challenges that should be addressed, in order to serve your users in a better way?
1	Difficulty and time consuming to store, manage and use large amount of data for your application.
2	It is hard to find information in machine-readable format, in terms of who does generate & publish such information
3	It is time consuming to search for related information (lots of search pages are needed)
4	Lack of technical expertise or resources in processing the specific formats of data
5	It is hard to combine similar data from various sources
6	Difficulty to get periodically updates and fresh data
7	There is a need for elastic cloud services that allow you to extend your infrastructure by adding custom processing power, storage and memory.
8	It is hard to run VM images with pre-installed software stack (e.g. CMS installation) with only one click
9	It is difficult to manage VMs (e.g. upgrade, restart, monitor, back up)
10	VMs are not a good abstraction for your software, you need other higher level services that eases the development and deployment process
11	It is hard to dynamically adapt your infrastructure to the current load with automatic scaling
12	Other:

Table 5: Final list of top identified challenges for targeted segments



5.3.1. Data requirements analysis

This section addresses the company and business related data, which are provided by the interviewees. It is essential to point out that the majority of the businesses belong to the wider agri-food sector. In fact, 35% of the sample deals with agricultural sciences software and applications.

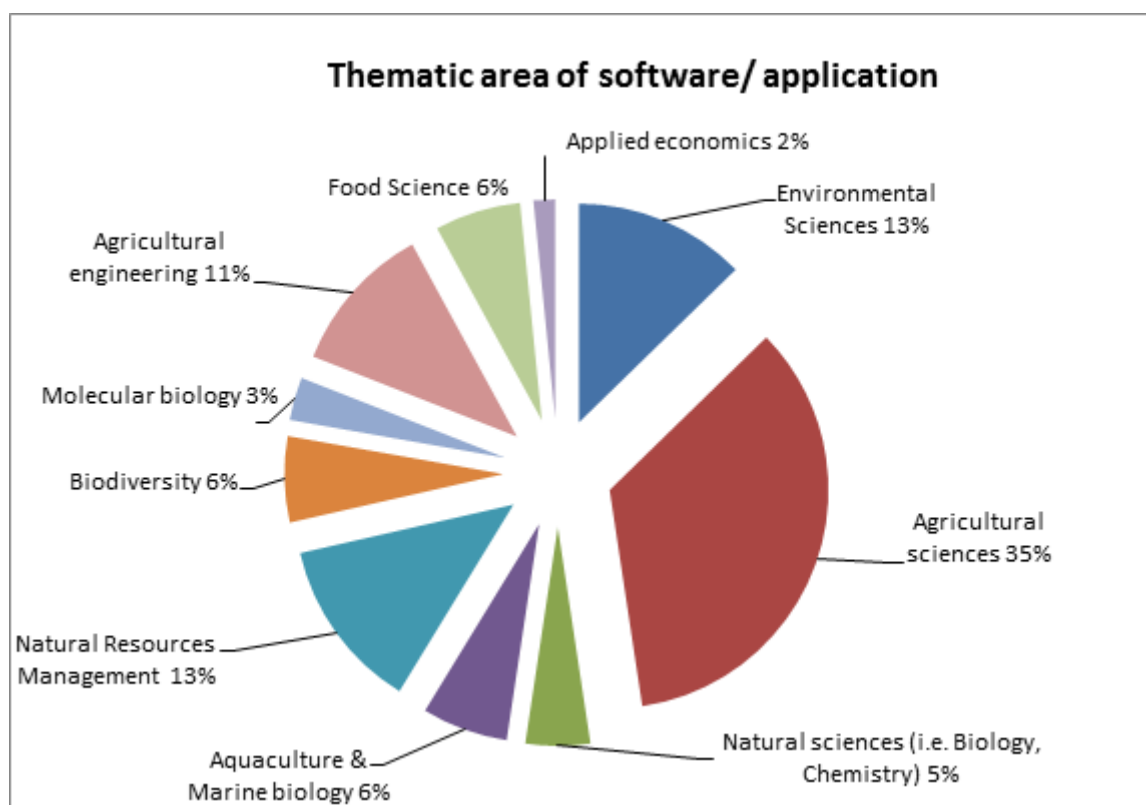


Figure 13: In which agricultural-specific thematic area(s) does your software/ application belong (Question5)

It is evident that the granularity of various data types is illustrated by the answers of the survey participants. Figure 14 reflects the big variety of data types that are needed from AgTech applications. Geospatial data, scientific publications, statistical data, organisational data and multimedia compile the top five data types that are essential for AgTech stakeholders.

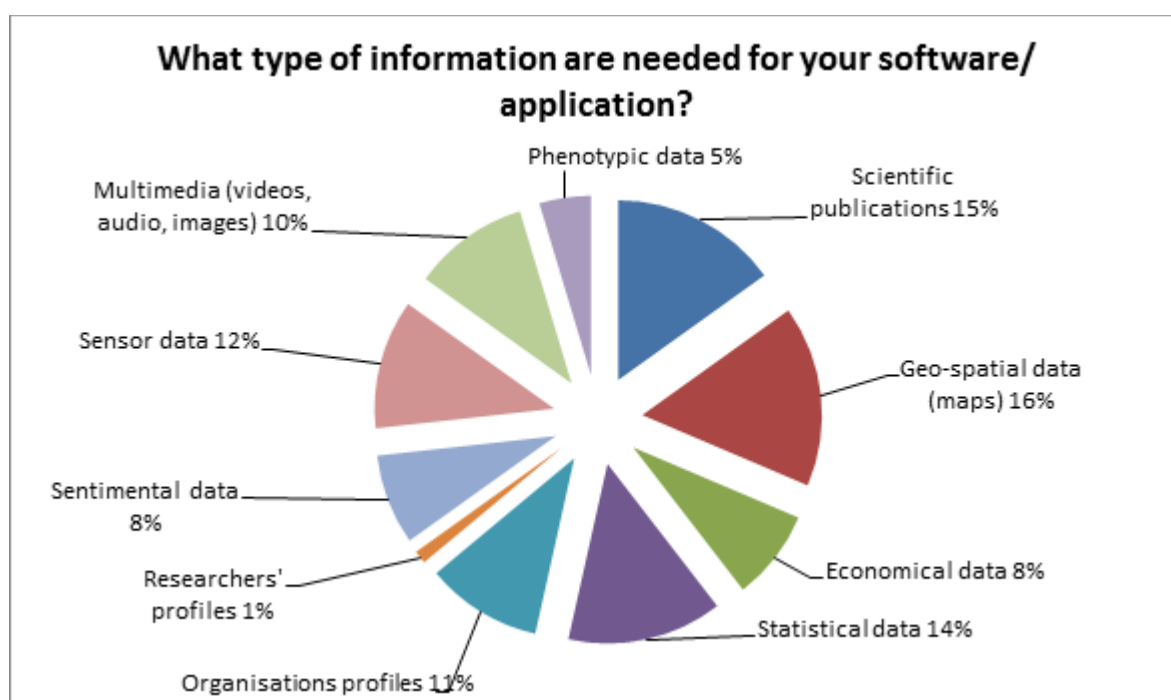


Figure 14: What type of information is needed for your software/ application (Question 6)

5.3.2. Cloud service and Virtual Machine Requirements

In order to capture the needs of the targeted segment, a set of questions for cloud usage and VM management scenarios was introduced to the interviewees.

With regards to the cloud maintenance requirements, the participants were asked to provide more information concerning the setup of their cloud service. It was obvious that the size of the product/application and the type of the company reflected on the number of VMs that a company is making use of.

The interviewees expressed in detail their needs or common scenarios in cloud and VM management. Answers were grouped in 3 segments according to the type of persona that responded. Based on this segmentation, it seems that individual IT consultants are not using specific VM management tools and have limited requirements for cloud hosting services. On the other side, SMEs and startups seem to be avid users of such services, based on the opinions expressed. As illustrated in the table above, startups make use of commercial available cloud services such as Google Cloud and Amazon Web Services, in tandem with various VM management tools. System requirements vary, due to the fact that they depend on the software/application setup. Along similar lines, SMEs are deeply involved in cloud services and VM management tools. Since there are commercially available applications, the SMEs' needs in cloud hosting are larger than startups. Examples of "cloud hosting farms" of over 1500 VMs indicate that the users and the commercial availability of a software application are related to the cloud management requirements of an SME. Furthermore, in terms of VM management tools, apart from AWS, Google cloud and VMware that were already mentioned, a wide variety of tools are being used by

the technical representatives of SMEs (such as: Blueprints, Docker, Ansible, Elasticsearch, Kafka, Zookeeper, Spark, rabbitMQ, redis, etc.).

	IT Consultant/ Individual developer	Startup	SME
Cloud management details	16 cores 32GB RAM 2TB (raid) HDD	96 CPUs 128 GB RAM 3 TB HDD	98 VMs, approx. 200 core, 300GB Ram, 4TB Storage
	CPU>2 cores Memory> 4GB Storage > 50GB	6 CPUs 2core 4GB RAM	Over than 1500 VMs, 2 CPUs with 4gb RAM each and 50 GB storage
		2x3Ghz CPU 6 GB RAM 2TB storage	2 CPU/ 8 GBs/ 500GB, 8CPUs, with 32 Gigs of Memory and up to 800GB storage
		2CPUs 8GB Memory 1TB Storage	
VM management details	Don't use VM management tools	Open Stack, FI LAB VM IMAGES, manual setup, Amazon Web Services, Google cloud	Typical setup via Amazon Machine, Images/ Blueprints, Docker, Ansible, Elasticsearch, Kafka, Zookeeper, Spark, rabbitMQ, redis, Amazon Web Services, VMware

Table 6: Overview of the expressed cloud and VM management details (as presented in questions 7 & 8)

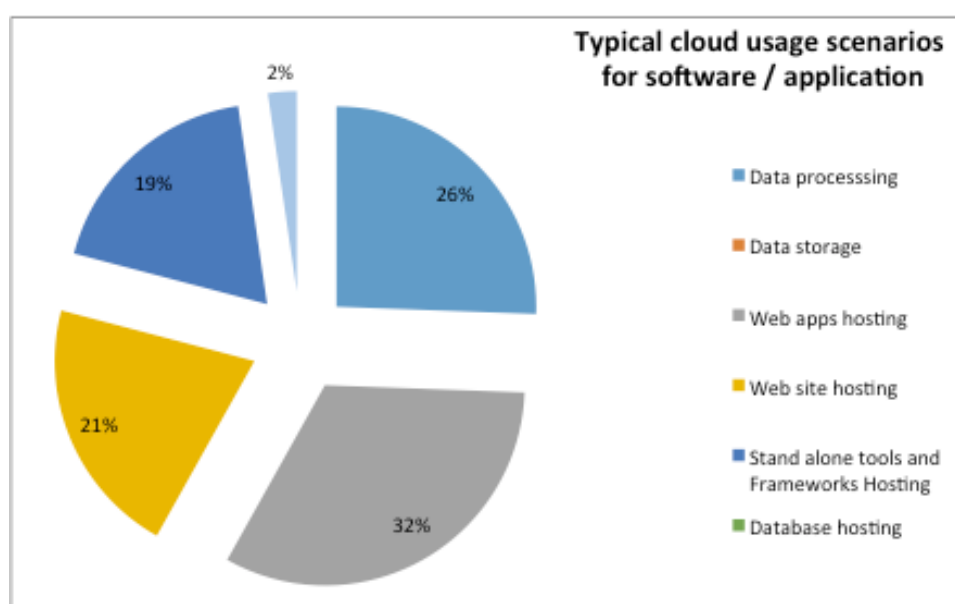


Figure 15: For your software/ application, which are the typical usage scenarios of the cloud? (Question 9)

Two core activities are in the typical cloud usage scenario of software and applications. These are the hosting of web application and the processing of the data created and/or used from these services. Also, another typical usage of the cloud services is for hosting websites, which accounts for 20% of the answers provided. In addition to the typical cloud usage scenarios, interviewees were asked for the optimal frequency for updating their data. Almost 60% of the respondents agreed that data should be updated in less than one-month intervals in order to provide their product with the latest available information.

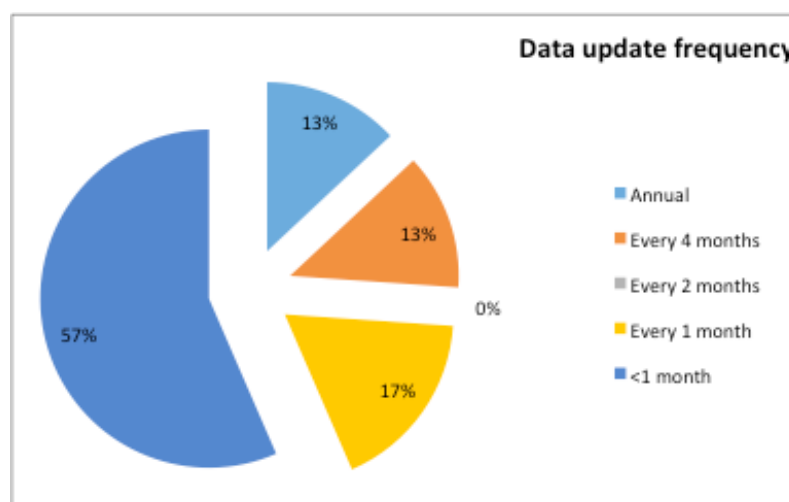


Figure 16: How often should the required data be updated for your product? (Question 10)

5.3.3. System requirements

In regards to the format of the provided information that is needed from the respondents, opinions vary. The use of an API is obvious, since the format that can be used to provide such information is either XML, JSON or CSV. APIs support both XML and JSON formats for data

exchange. An estimated 51% of the respondents prefer XML and JSON format, which - if adding the API percentage - adds up to 74%. This can validate the assumption that APIs can be characterised as the optimal way for data exchange within the AgTech community.

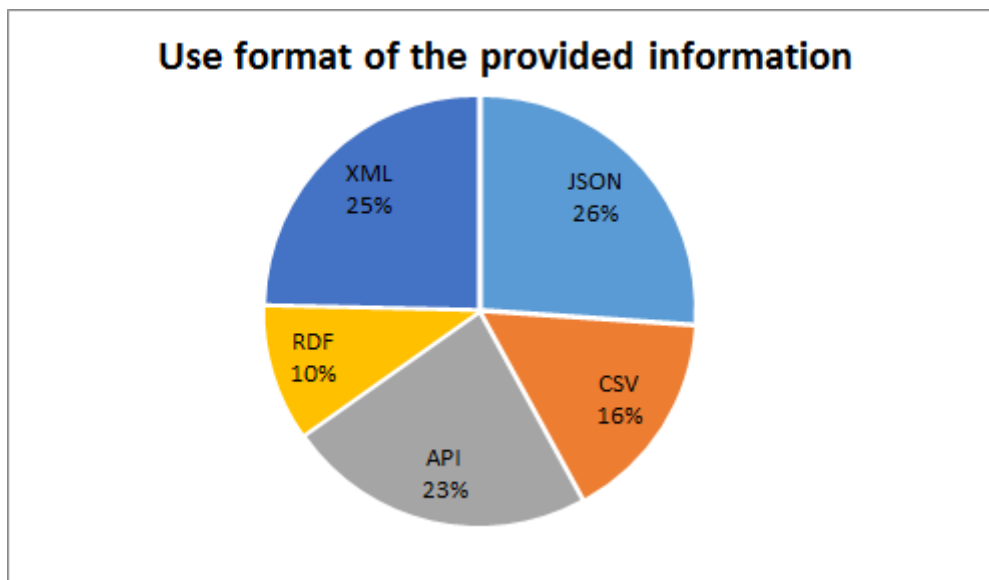


Figure 17: In which format could you use the provided information? (Question 12)

Additionally, it was expressed that if the access to information was through an API, the majority of the users (74%) will require unlimited calls either for single (26%) or unlimited users (48%).

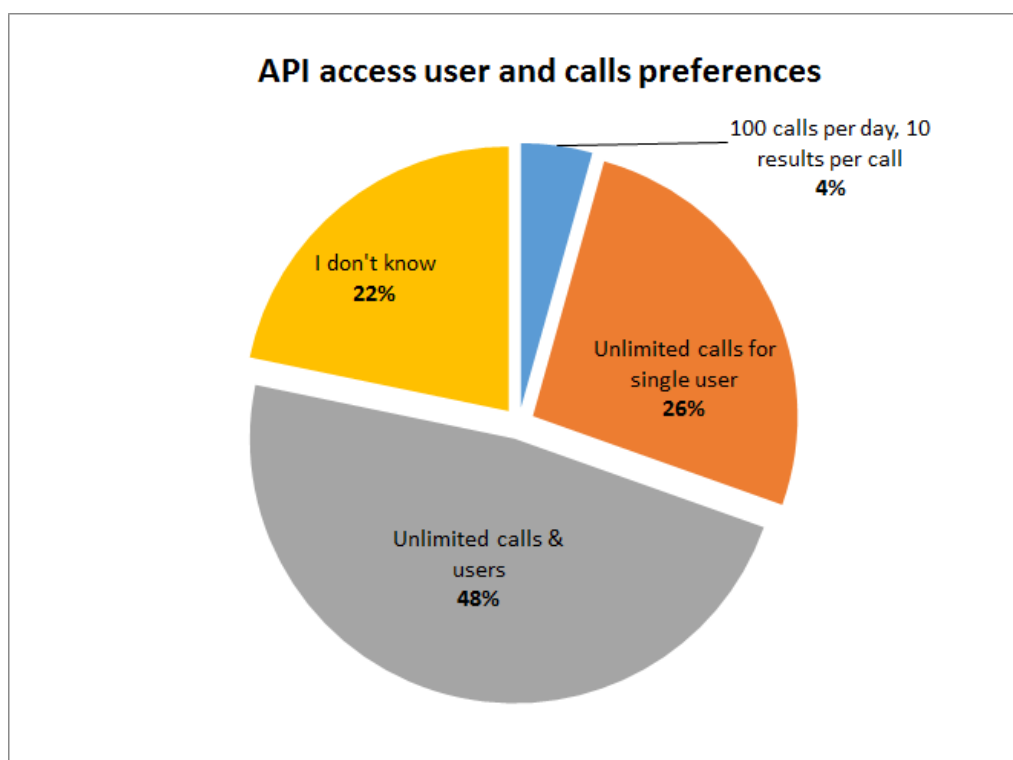


Figure 18: In terms of access to information through API, which of the following is the most preferable? (Question 13)

5.4. Community challenges validation

As it was presented in the previous chapter, the initial version of the questionnaire was slightly changed in order to express in a precise way the challenges that the AgTech community is facing. The initially expressed challenges were reformed so as to meet the requirements that the interviewees expressed and were at the first place validated by the majority of them. The challenges list was finalized to the following:

Challenge 1: Difficulty and time consuming to store, manage and use large amount of data for your application.
Challenge 2: It is hard to find information in machine-readable format, in terms of who does generate & publish such information
Challenge 3: It is time consuming to search for related information (lots of search pages are needed)
Challenge 4: Lack of technical expertise or resources in processing the specific formats of data
Challenge 5: It is hard to combine similar data from various sources
Challenge 6: Difficulty to get periodically updates and fresh data
Challenge 7: Difficulty on accessing the public and open data (e.g. behind the firewalls)
Challenge 8: There is a need for elastic cloud services that allow you to extend your infrastructure by adding custom processing power, storage and memory.
Challenge 9: It is hard to run VM images with pre-installed software stack (e.g. CMS installation) with only one click
Challenge 10: You need to distribute VM images with pre-installed software to cloud providers with only one click
Challenge 11: It is difficult to manage VMs (e.g. upgrade, restart, monitor, back up)

Table 7: Final list of challenges addressed by the AgTech community

The process requires indicating the top three common challenges that the targeted AgTech sector is facing. By addressing the top three problems of the AgTech industry, EGI is capable of building a tailor made solution for the sector's stakeholders, providing them with services and infrastructures that matter the most to them.

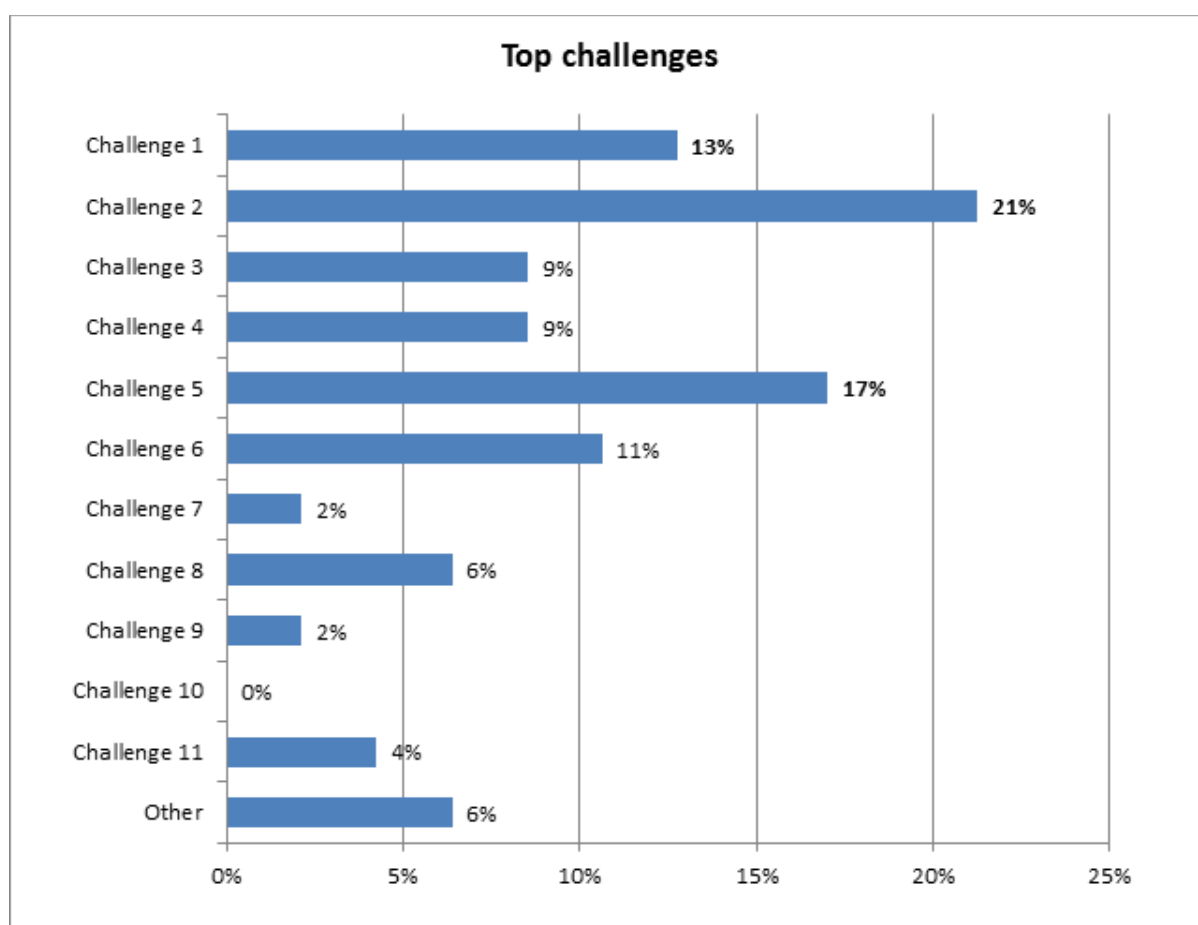


Figure 19: Which of the following are the top challenges that should be addressed, in order to serve your users in a better way? (Question 11)

As demonstrated in Figure 19, the top three challenges that the targeted group of users is facing are:

- It is hard to find information in machine-readable format, in terms of who does generate and publish such information (Challenge 2)
- It is hard to combine similar data from various sources (Challenge 5)
- Difficulty and time consuming to store, manage and use large amount of data for your application. (Challenge 1)

Top challenges indicate that the agri-food community is facing discoverability issues with the accessed data. Respondents point out that it is hard for AgTech startups and SMEs to locate data (or metadata) in machine-readable formats from the respective data publishers and data sources.

Due to the abovementioned challenge, it is apparent that data that are not in the same format - even if it may seem of similar types- cannot be easily combined by the AgTech community. Also, this difficulty seems very challenging from the community aspect, since it refers to the 17% of the total sample. The expressed requirements provide EGI with a specific direction towards agri-food businesses. In the process of building a new product, EGI should aim to address the inability of AgTech stakeholders to locate data in machine-readable format from the numerous data sources

that exist in research and business data environments. Furthermore, the granularity of the data that are made available to AgTech startups, SMEs and individuals is of extreme importance, thus EGI should take into consideration the inability of these actors to store manage and use large amounts of data. Combining the importance of the aforementioned challenges, EGI should aim towards a unique solution for the AgTech sector, taking into consideration the particularities of the high volume and high veracity agri-food data.

5.5. Envisaged solution

After reporting on the top challenges that were addressed by the user communities when dealing with cloud based services and e-infrastructures, this section identifies solutions that could potentially deal with the aforementioned challenges:

Challenge 1: *Difficulty and time consuming to store, manage and use large amount of data for your application*

Cloud and Grid Computing Platforms should be efficiently combined to deal with the specific challenge. Cloud services for providing large amounts of storage and grid services for providing the available compute capabilities for processing and analysing large amounts of data. EGI with its grid and cloud services is well positioned to provide tailored solutions for this specific challenge of the AgTech community. The Open Data Platform of EGI under development is designed to support scientific communities in the dissemination and re-use of open, distributed research datasets. The platform, which is based on the Onedata⁶⁶ technology, enables data management and data access processes - including data replication and 'bring your computation to data' - to existing EGI infrastructure services. EGI could also extend or adapt its Open Data Platform in order to offer Workflow as Service (WFaaS) functionality for generic use cases of the sector's businesses. In that way, EGI enables the agri-food community to make their systems scalable and easy-to-extend, by efficiently responding to continuous workflow requests from users and schedule their executions in the cloud, saving time for these processes.

Challenge 2: *It is hard to find information in machine-readable format, in terms of who does generate & publish such information*

Repository services for storing and retrieving metadata information in proper machine-readable formats (XML, RDF, JSON), is a solution for this challenge. For maximising machine interoperability, the metadata storage process should follow well-known metadata standards (Dublin Core, IEEE LOM, etc.). EGI has the know-how and the expertise to handle large amounts of data and to provide consulting services to the AgTech community on how to build their services by picking and mixing EGI services such as Open Data Platform and Grid and Cloud computing.

⁶⁶ <https://onedata.org>

Challenge 3: *It is time consuming to search for related information (lots of search pages are needed)*

The same solution, which is described for challenge 2, applies also in this challenge. Metadata Repository Services is a common solution for providing efficient discovery services in large amounts of information. EGI currently offers two alternative services that could aid the resolution of this challenge: the first is its Open Data Platform, which is a new service in EGI to support scientific communities in the dissemination and re-use of open, distributed research datasets and the second is EGI Application Database (AppDB ⁶⁷).

Challenge 4: *Lack of technical expertise or resources in processing the specific formats of data*

Cloud computing platforms provide ready SaaS solutions for processing different types of data, without the customer having prior administrative or any other technical skills for installing and executing the software. EGI does not provide SaaS at the moment, but does however offer consulting services and operations services to user communities to build their services on top of EGI solutions. In that way, the agri-food community can benefit from the consulting services of EGI in order to develop technical solution for the process of specific data formats.

Challenge 5: *It is hard to combine similar data from various sources*

Data Federation services for harvesting and aggregating data from different sources could be applied here. These services include implementations of the OAI-PMH protocol, API readers, file dump parsers, etc. However, the selection of such services heavily relies on exposing mechanisms that each data source provider supports. Additionally, an internal workflow mechanism for transforming the structure of the collected data into a single homogeneous structure could be included as an additional solution; although it applies only in cases where data source providers expose their data in different formats. Similarly to Challenge 2, AgTech community could greatly benefit by building its storage services on top of EGIs Open Data Platform.

Challenge 6: *Difficulty to get periodically updates and fresh data*

The same solution described for challenge 5, applies to this challenge. Data Federation Services should constantly update their collected data by frequently harvesting various data providers, thus

⁶⁷ <https://appdb.egi.eu>

updating their internal database with fresh data. Also, EGI's DataHub can provide backbone services for the continuous update of data coming from Resource Centres and Data Providers.

Challenge 7: *Difficulty on accessing the public and open data (e.g. behind the firewalls)*

Open authorisation solutions and standards (such as the OAuth) providing uniform authentication and authorization mechanisms among different resource providers could be a solution for this challenge. EGI AAI Check-In Service supports multiple protocols including SAML, OAuth so it allows uniform access to both EGI Resources and external Data Providers. In other words, EGI Check-In service lowers the barrier of access to both data sources and computing services.

Challenge 8: *There is a need for elastic cloud services that allow you to extend your infrastructure by adding custom processing power, storage and memory*

EGI federated cloud service offers open standard based interfaces and services that could form the ideal solution for dealing with this specific challenge of the AgTech sector. Given the fact that scalability and on-demand resource provisioning is one of the basic characteristics of such architectures, the EGI cloud could provide an extensible infrastructure to the agri-food stakeholders, based on their needs.

Challenge 9: *It is hard to run VM images with pre-installed software stack (e.g. CMS installation) with only one click*

Commercial solutions like Google Cloud or Amazon Web Services provide services for deploying VM images with pre-installed development stacks. EGI can investigate the current state of the art for offering a secure and competitive PaaS solution (e.g. Docker) that could address this challenge.

Challenge 10: *You need to distribute VM images with pre-installed software to cloud providers with only one click*

Bitnami-like solutions allow the deployment of development stacks or software packages for web applications to cloud providers using installers and containers, which could be applied here. EGI currently offers a similar solution through the AppDB, based on VMcaster/VMcatcher, which allows the management and distribution of VM images throughout EGI federated cloud participating providers.

Challenge 11: *It is difficult to manage VMs (e.g. upgrade, restart, monitor, back up)*

VM monitoring services is the proper solution for this kind of challenge. Cloud providers offer such services for their own VM, but there is a range of open source and commercial solutions for single monitoring of VMs from a single or different cloud providers. EGI already has in place a monitoring service called ARGO⁶⁸, which could be easily extended/tailored to monitor community services.

⁶⁸ argo.egi.eu



6. Conclusion and Discussion

The EGI community should take into account the proposed solutions for the top user challenges. As it was stated in chapter 5.4, the top three challenges indicate the need of a federated infrastructure able to manage various similar and unrelated data types, high volume and veracity data that are being syndicated from a wide range of data sources. EGI can activate services that are already in place, such as the Open Data Platform, Data Hub, ARGO, AppDB and enhance them with agri-food targeted features in order to address the challenges of the agri-food community. Furthermore, bilateral agreements and collaboration with existing domain-specific e-infrastructures should be taken into consideration in order to help the EGI community expand the outreach of their services. SMEs and startups are the main actors of the data-driven evolution that is expanding on a global level. Numbers indicate that the AgTech sector will attract additional investments in the coming years, so the need for a robust e-infrastructure and open data ecosystem is evident. Since an increasing amount of the data used and provided by businesses and academia is made available in an open manner, enhanced EGI services for the agri-food community are needed. Provided that EGI envisions a tailor made solution for the AgTech sector, with existing and new EGI services, there are possibilities of pursuing this expansion to agri-food through existing sectorial networks (such as AGINFRA) and promote EGI services through various channels and events (e.g. the European Data Forum).

This report will be analysed by the NA2.3 activity to define an action plan based on the recommendations provided e.g. individual contact points, new or existing dissemination material, events to attend, technical development needed.



7. Bibliography

Addison, C., Boto, I., Mofolo, L. (2015). Data: the next revolution for agriculture in ACP countries. Reader, Brussels Briefing. 40. Available at:

https://brusselsbriefings.files.wordpress.com/2015/01/reader-bb40-data-revolutioneng_v2-low-res-proof2-2.pdf

Curry E. (2016), The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In M. Cavanillas, E. Curry, W. Wahlster (Eds.) New Horizons for a Data Driven-Economy - A Roadmap for Big Data in Europe (pp. 29 - 38), Springer International Publishing.

DG Europe (2013), A European strategy on the data value chain.

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. Technical report, META Group

OECD (2013). *Exploring Data-Driven Innovation as a New Source of Growth*. Available at [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP\(2012\)9/FIN&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP(2012)9/FIN&docLanguage=En)

Turner, V., Gantz, J. F., Reinsel, D., & Minton, S. (2014). *The digital universe of opportunities: rich data and the increasing value of the internet of things*. Rep. from IDC EMC.

UN (2013), A new global partnership: Eradicate poverty and transform economies through sustainable development, The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda, United Nations Publications, 2013 available at: <http://www.post2015hlp.org/wp-content/uploads/2013/05/UN-Report.pdf>

Food and Agriculture Organization of the United Nations, High Level Expert Forum - How to Feed the World in 2050, October 2009, Rome

AgTech Investing Report, Year in Review 2015, February 16, 2016, www.agfunder.com

Venture Pulse Q4 2015, Global Analysis of Venture Funding, January 19 2016, KPMG

Gilpin, Lyndsey. "How Big Data Is Going to Help Feed Nine Billion People by 2050." Tech Republic. Web. 23 May 2016.

Business opportunities in Precision Farming, Roland Berger Strategy Consultants GmbH, July 2015

Improving Agribusiness Performance with Big Data, Oracle Enterprise Architecture White Paper, 2015

Pesce, V., Maru, A., Keizer, J. The CIARD RING, an Infrastructure for Interoperability of Agricultural Research Information Services. Agricultural Information Worldwide, Vol. 4, No 1, 2011 (on-line, accessed 23 November 2012), <http://journals.sfu.ca/iaald/index.php/aginfo/article/view/213/170>.

Data Revolution for Agriculture, Jellema, A., Meijninger, W. and Addison, C., 2015., Open Data and Smallholder Food and Nutritional Security, CTA Working Paper.



8. Annex

8.1. Initial requirements elicitation survey

How data-powered services can support your product

We would like to understand how the cloud-based services hosted and provided by the European Grid Infrastructure (EGI), as well the data services for better accessing the needed data, can support your software / application, in order to better serve your users' communities.

Name: _____

Email: _____

Organization: _____

1. Which of the following describes better the type of your company?		
IT consultant / Individual developer	Startup	Small-sized enterprise
Medium-sized enterprise	Large-sized enterprise	Other:

2. What is your role in your organization / company?		
Chief Executive Officer	Chief Operations Officer	Director
Product manager	Web developer	System Administrator
Software Engineer	Business Development Manager	Other:

3. Size of your tech team (in persons).				
1-3	4-10	11-20	21-50	>51

4. In which thematic area(s) belong your software / application users?				
Agriculture	Education	Health	Energy	Transportation
Finance	Publishing	Media	Other:	

5. In which agricultural-specific thematic area(s) belong your software / application?	
Agricultural Sciences	Environmental Sciences
Natural sciences (i.e. Biology, Chemistry)	Agricultural engineering
Aquaculture & Marine biology	Water Resources Management
Natural Resources Management	Biodiversity
Climate Action, Environment and Resource Efficiency	Organic Agriculture and Sustainable agriculture
Food Science	Applied economics
Water Quality Research	Inland water Research



Hydrogeology	Molecular Biology
Food Science	Development Communication
Applied Economics	Other:

6. What is the status of your application?	
New product concept	Existing product (<1 year)
Existing product (>1 year)	Product Modification
Product Upgrade	Other:

7. Do you have a defined business plan for your application?		
Yes	In progress	Not yet
I don't know	Other:	
8. If you are maintaining your own cloud, provide details in terms of number of CPUs, memory and storage.		

9. How many VMs on average per year are you using (for the last two years)?

10. If you are using virtualization technologies and open platforms (e.g. Docker) for setting up and managing new VMs images, please give us more details.

11. For your software/ application, which are the typical usage scenarios of the cloud?		
Data processing	Data Storage	Web app hosting
Web site hosting	Stand alone tools and Frameworks Hosting	Other:

12. What types of information are needed for your software/application?		
Scientific publications	Geo-spatial data (maps)	Economical data
Statistical data	Organizations profiles	Researchers' profiles
Social media, news & events	Phytochemical data	Molecular data
Sensor data	Multimedia (videos, audio, images)	Phenotypic data
Other:		



13. Which of the following are the top challenges that should be addressed, in order to serve your users in a better way?		
	Difficulty and time consuming to store, manage and use large amount of data for your application.	
	It is hard to find information in machine-readable format, in terms of who does generate & publish such information	
	It is time consuming to search for related information (lots of search pages are needed)	
	Lack of technical expertise or resources in processing the specific formats of data	
	It is hard to combine similar data from various sources	
	Difficulty to get periodically updates and fresh data	
	Difficulty on accessing the public and open data (e.g. behind the firewalls)	
	There is a need for elastic cloud services that allow you to extend your infrastructure by adding custom processing power, storage and memory.	
	It is hard to run VM images with pre-installed software stack (e.g. CMS installation) with only one click	
	You need to distribute VM images with pre-installed software to cloud providers with only one click	
	It is difficult to manage VMs (e.g. upgrade, restart, monitor, back up)	
	VMs are not a good abstraction for your software, you need other higher level services that eases the development and deployment process	
	It is hard to monitor and define alerts for the health of your VMs (e.g. to set up limits for CPU, for memory)	
	It is hard to dynamically adapt your infrastructure to the current load with automatic scaling	
	Other:	
14. How often should be updated the required data for your application?		
Annual	Every 4 months	Every 2 months
Every 1 month	<1 month	

15. Do you consider as useful the human-readable and easy to search interface, in order to find data?		
Yes	No	I don't know

16. What is your need in terms of the information quality?
None (as originally provided by data sources)
Completeness >80%, de-duplicated records. Semantically enriched, tagged with AGROVOC
Other:

17. In which format could you use the provided information?
--



JSON	CSV	API	RDF	XML	Other:
------	-----	-----	-----	-----	--------

18. In terms of access to information through API, which of the following is the most preferable?

100 calls per day, 10 results per call	Unlimited calls for single user
Unlimited calls & users	I don't know
Other:	

19. What should be the customer support guarantee about the services?

Not guaranteed response time	Within 5 working days
Within 2 working days	Within 1 working day
Other:	

20. What should be the Service Level Agreement (SLA)?

None	99%	Other:
------	-----	--------

21. Based on your answers on questions 14-20, what would you expect to pay (as an annual cost)?

No cost at all	< 500 Euros / year	500 - 2000 Euros / year
2000 - 5000 Euros / year	I don't know	Other: