eROSA
e-infrastructure Roadmap
for Open Science in Agriculture

D3.7 - Foresight Roadmap Paper

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

DELIVERABLE NUMBER	D3.7
DELIVERABLE TITLE	Foresight Roadmap Paper
RESPONSIBLE AUTHOR	Panagiotis Zervas (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO)

GRANT AGREEMENT N.	730988
PROJECT ACRONYM	eROSA
PROJECT FULL NAME	Towards an e-infrastructure Roadmap for Open Science in Agriculture
STARTING DATE (DUR.)	01/01/2017 (18 months)
ENDING DATE	30/06/2018
PROJECT WEBSITE	erosa.aginfra.eu
COORDINATOR	Odile Hologne
ADDRESS	Route de Saint-Cyr RD 10, Versailles, 78026, France
REPLY TO	odile.hologne@inra.fr
PHONE	+33 1 30 83 33 92
EU PROJECT OFFICER	Mrs. Georgia Tzenou

WORKPACKAGE N. | TITLE	WP3 | Roadmap co-Design & Uptake
WORKPACKAGE LEADER	Agroknow
DELIVERABLE N. | TITLE	D3.7| Foresight Roadmap Paper
RESPONSIBLE AUTHOR	Panagiotis Zervas (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO)
REPLY TO	pzervas@agroknow.com
DOCUMENT URL	http://www.erosa.aginfra.eu/deliverables
DATE OF DELIVERY (CONTRACTUAL)	29 June 2018 (M18)
DATE OF DELIVERY (SUBMITTED)	13 July 2018 (M19)
VERSION | STATUS	V1.0 | Final
NATURE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS PARTNERS	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO), Johannes Keizer (Agroknow)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.5	Table of Contents	16/3/2018	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO), Johannes Keizer (Agroknow)
0.6	Initial content added to all sections	30/4/2018	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO), Johannes Keizer (Agroknow)
0.8	Draft version for read for internal review	24/5/2018	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO), Johannes Keizer (Agroknow)
0.9	Draft version ready for open consultation	8/6/2018	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Odile Hologne (INRA), Sander

			Janssen (ALT-DLO), Johannes Keizer (Agroknow)
1.0	Final version submitted	18/7/2018	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Odile Hologne (INRA), Sander Janssen (ALT-DLO), Johannes Keizer (Agroknow)

PARTICIPANTS		CONTACT
Institut National de la Recherche Agronomique (INRA, France)		Odile Hologne Email: odile.hologne@inra.fr
Stichting Wageningen Research (Alt-DLO, The Netherlands)		Sander Janssen Email: sander.janssen@wur.nl
Agro-Know IKE (Agroknow, Greece)		Nikos Manouselis Email: nikosm@agroknow.com

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This roadmap is one of the outputs of the e-ROSA project "e-infrastructure Roadmap for Open Science in Agriculture" funded by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 730988. The consortium gathers two research performing organizations INRA (FR) and WUR (NL) and Agroknow (GR) a private company. More than 100 stakeholders were involved in the elaboration of this roadmap through the workshops and the project methodology to build a common knowledge, foresight and recommendations to achieve this vision.

The process all along the project was "as open as possible" with workshops, open consultations to discuss, enrich and assess the proposals. The methodology allowed to:

- **Understand the actual landscape using** i) a bibliometric study to identify the academic actors working on "Data AND Agriculture" ii) a bibliographic analysis of e-infrastructure roadmaps as reference for our future work iii) the identification and mapping of different entities (organization, initiatives, facilities, data points etc.) to describe the data ecosystem.
- **Analyse the scientific challenges and researchers needs,** working on use cases in different domains: smart farming, omics/phenotyping, food and nutrition, with interviews and workshop
- **Describe a vision for open science in Agri-food and explain how an e-infrastructure could support this vision**. The writing of this statement was planned from the beginning as a collaborative effort, not limited to the participants of the e-ROSA project. 17 professionals from 3 continents (USA, India, China, Europe) and important organizations (CGIAR, CAAS, ICAR, Bayer, INRA, CABI) participated in the online discussion and editing process.
- **Elaborate recommendations** to achieve the vision addressing different issues, not only technical but also on human and economic aspects: culture change, skills and sustainability of the envisioned e-infrastructure.

Thus, this document is organized in 5 main sections:
- The vision "Open science and Food systems in 2030" as a driver of transformation
- The challenges and the researchers' needs to address
- The analysis of the actual "data ecosystem"
- The requirements for an e-infrastructure and its envisioned architecture
- The recommendations for the implementation

We have used the "Share, connect, collaborate" motto as guide trough this story as it appears to be the most federative challenges to build an e-infrastructure federating the actual scattered data ecosystem in the agri-food sciences:
- Sharing: of the resources of relevance to the scientific process (data, models, papers, etc). Open science is only possible if one is able to share one's research first.
- Connecting: available resources need to be connected to allow integration and tackling large scale and more ambitious questions in science.
- Collaborating: the research community itself needs to collaborate beyond ad-hoc arrangements to create, maintain and supply domain specific resources for open science in a network of regional or domain nodes.

This version will be updated, the readers should refer to the web site http://erosa.aginfra.eu to be informed of the latest developments.

2 THE VISION

In 2030 agri-food systems and businesses will produce healthy nutritious foods for all, through input-efficient methods and an environment that fosters collaborative networks that constantly seek to improve their economic, environmental, technological and social performance for all players.

Further, food systems will respond in an agile way to local, regional, and global needs, and contribute to achieving a wide range of objectives as framed in the Sustainable Development Goals[1]. These objectives include achieving food security, mitigating global warming, ensuring good health and preserving biodiversity.

To be effective, these food systems need to be inclusive, resilient and knowledge intensive. Agri-food systems and businesses create and use knowledge on food production, environmental effects, food processing, distribution, nutritional values, production costs and the likely health benefits. They must be supported by open science-based knowledge systems to stimulate further innovation and accelerate impact.

Delivering on this approach requires a perspective that looks beyond individual value chains, crops, livestock or farm types. In the 'open-science-based food knowledge cloud', researchers should be able to:

1. *openly collaborate* with relevant stakeholders to improve the functioning of the food system by sharing knowledge, including practitioners' know-how.
2. *create novel systems approaches* that harness research to address impacts on the entire food system, that are multi-disciplinary and engage multiple actors and on a range of related criteria.
3. *put in action research* that is fully knowledge-driven, and based on the principle of transparent data and information that allow researchers to rapidly discover, aggregate, integrate, and analyse different types of data and information sources to form a clear picture of the current situation and future needs and develop solutions that are agile and robust.
4. *embrace work approaches* that are most likely to accelerate impact, focusing research on achieving impact in a broad societal context – and where possible, and involving practitioners in participatory research activities.

To realize this **Vision 2030**, the transition of the research approach towards a systemic, integrated, multidisciplinary and global approach needs to be accelerated. As part of this transition, agriculture and food systems research must embrace digitization, transparency and cooperation.

The agri-food sector is very diverse. It relies on complex knowledge – both theoretical and experimental. For policy makers, planners and food producers to progress with the new thinking described here, they will need to take a view that integrates multidisciplinary, multi-scale, multi-actor and geo-location-based approaches. This will be further enhanced by the integration of the range of agricultural data, models and analytical and visualization tools available.

This increase in available tools is driven by the emergence of thinking systems, coupled with the exponential growth of digital resources and global use of the Internet. Other key factors include: automated data collection (robots, Unmanned Aerial Vehicles, connected sensors, etc.); new tools in the "omics" fields and emerging new information and data sources (e.g. Internet of Things, crowdsourcing, text and data mining approaches). These trends are linked to the fact that natural and societal phenomena are increasingly being described by massive data at different scales, from various sources and with different temporal-spatial resolutions. The ability to share, access and integrate heterogeneous data is a

[1] http://www.un.org/sustainabledevelopment/sustainable-development-goals/

key to addressing climate change impacts on food security; to provide healthy and nutritious food for all; for developing sustainable food value chains and supporting local agricultural adaptations and rural development. To best address these challenges in an agile approach, the research and innovation sectors need easy access to digital knowledge, data resources and technologies. This will be achieved through resources such as interoperable data, connected sensors, data sharing and exploration methods, modelling and coding frameworks, intensive simulation environments, or social networks – all of which are fuelled by shared information exchange standards and wide access to high-speed internet connections.

To make this happen, it is critical that the industry and the community of research and development agencies to have a robust and open e-science framework and related facilities or e-infrastructures that operate following common standards. This will enable ethical, responsible and secure sharing of the 'engine' of this new approach – data and information, computing and storage resources, codes and data-mining algorithms, models and ontologies. To make the system work effectively, partners need to have relevant expertise, and access to examples and best practices and an ethical framework that underpins an open science approach for agriculture.
In this light, ensuring sustainable and nutritious regional and global food systems requires our commitment to 'Open Science'. This requires that research and development institutions, and every stakeholder in the innovation process, fully embrace a digital transition for each phase of the knowledge production cycle for innovation. This extends from research planning and design to collection, analysis and simulation of data and the dissemination of knowledge and the underlying data.

This will enable increased collaboration and efficient, ethical and secure data flow between all partners from R&D, farming, supply chains and consumers. The support of policy makers is vital to create an environment for open science that will speed the pace of development and innovation.

Agri-food science and innovation will benefit significantly from such a shared knowledge ecosystem. This shared knowledge will be produced and used by diverse users not only academic researchers but also farmers, the industry, extension services and citizens. A shared global data space will help to build the infrastructures that will open useful information to all these stakeholders and propel the agri-food sector forward.

Embracing digitization, transparency and collaboration for this ambitious endeavour will generate significant benefits:

- **Speeding the pace of knowledge circulation:** As for all sciences, open science for the agri-food development will stimulate the production and circulation of knowledge for farmers or extension services, private companies linked to the agri-food sector and for education. This is also a key issue for developing countries.
- **Increasing spill-over of useful new knowledge to the economy:** Globally, in the agri-food sector, open data helps shape best practices. Transparency around targets, subsidy distribution and pricing, for example, create incentives which affect the behaviour of producers, regulators, researchers and consumers. It also helps public authorities to make more appropriate decisions.
- **Promoting opening access to knowledge – the FAIR principles[2]:** When studying climate change and its impact on agriculture, food security and safety, free access to a worldwide pool of data and practical knowledge is of critical importance to the planet's sustainable development. For this to become a reality, it is critical that all publicly funded research agencies in agriculture, food and environment adopt the 'FAIR' principles. This will ensure that all knowledge produced and

[2] https://www.force11.org/fairprinciples

published is Findable, Accessible, Interoperable and Reusable. FAIR supports open sharing of data and knowledge and open access to publications.

- **Promoting citizens' engagement in science and research to enable mass innovation:** As a science of observation, agricultural research disciplines will strongly benefit from 'citizen science' projects, for example where information is collected and produced from farmers, NGOs, and consumers, to contribute to a better understanding of the ecosystem functioning and to promote common vision of future agro-ecosystems. These common efforts will result in better welfare and production management for farmers and improve food security and safety in a sustainable way.

The following actions will support this transformation:

- Universal knowledge access and promotion of a knowledge sharing culture at community level has to be assured. While different levels of data openness and use need to be acknowledged, the FAIR principles are always an effective guide. Diversity must be recognized and enforced. Data, information and knowledge today exist as an unconnected ecosystem, whose power will be greatly magnified if similar efforts are linked in a common approach, allowing for varying degrees of openness, depending on the context – ranging from sharing the subject matter or title of information, to limited openness and full open access, with embargo where necessary. Regulations will be needed to avoid the establishment of data monopolies and implementation of these regulations needs to be supported.
- The need to enable researchers to be able to easily explore, integrate and simulate their own data and data that have been produced by others – in or outside their community – that are complementary in terms of objects of study, scales and disciplines must be addressed. There is also a need to regulate database management rates whether the data are used by academic scientists, private firms or the non-profit sector. This is most important for potentially valuable data produced by NGOs, who cannot afford the cost of ongoing database management. Only if these links are made, it will be possible to accelerate the production and circulation of knowledge.
- The power of data produced by farmers and by land observation, which includes (but is not limited to) precision agriculture needs to be harnessed. Farms need to become laboratories linked to scientific research. As such, Farmers need to be recognized as partners in participatory research and not only as data providers.
- Data and information management start with research planning, data production – in the lab, on the field or at the observational level. It is important to work closely with equipment manufacturers to establish common principles and standards for data exchange.
- Interoperability across data sources and agreement on the adopting of 'good sense' standards, without reinventing wheels is a "leitmotiv". Standards need to be; open and shared; inter and cross-disciplinary; and co-defined with communities to ensure their adoption.
- It has to be built on existing digital infrastructures within the field and with generic (i.e. technological) infrastructures. Specific infrastructure and services to express communities' needs and requirement need to be developed.
- Distributed efforts and flexible governance for long-term empowerment by and sustainability across the agri-food community is necessary. Appropriate business models for data sharing and related services, especially for 'common goods' such as those supporting semantic interoperability for shared information and data discovery will be developed.
- The scope needs to be beyond a European community, to build a global network. The importance of Big Agricultural Sciences in the Global South must be understood. European Initiatives need to be integrated with those of G20 and G77 countries.
- Machine-readable means for encoding licenses for data and information (to support value-chain legal interoperability) and for encoding provenance of data and information (to enable attribution and quality assessments for any part of the value-chain) will be adopted.

- Approaches to improve the semantic interoperability of data that will facilitate and improve reasoning on data and information will be strongly encouraged.
- Cornerstone is the development of the necessary skills and capacity so that all partners and stakeholders in this endeavour can achieve this **Vision 2030.**

3 GRAND CHALLENGES OF THE AGRI-FOOD COMMUNITY

3.1 THE KEY CHALLENGE

The food system is a provisioning service that uses the natural environment to produce healthy and delicious food. This food system is facing a lot of challenges at the same time, that are increasingly interconnected, which means, for example, that the impact of food supply on health must be linked to the primary agricultural production.

To highlight some of the most important challenges in relation to the food system, the first challenge as often formulated, is “How to feed the 9 Billion in 2050?”. Population is about to grow massively in continents that are most dependent on agriculture, especially Africa. This challenge concerns not only the availability of sufficient food (as a production and distribution), but also the nutritional content of food and the nutritional diversity, as documented in a prominent definition of food security in 2001: Food security [is] a situation that exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life. (FAO. 2002. The State of Food Insecurity in the World 2001. Rome). Even in 2018, more than 700 million people are precarious for food security, a number that has been relatively stable over the past decades. Diseases like armyworm, wheat rusts and zoonosis pose serious threats to maintain crop production.

Climate change and increased climatic variability poses a second important challenge, at the same time as feeding the 9 billion. Due to climate change, extreme events such as changes in the growing season, and changes in the production area directly affect the crop production and animal nutrition. The Food System is also expected to contribute to reducing greenhouse gas emissions, not only in relation to carbon dioxide, but also nitrous oxides, methane and ozone. These emission reductions must not happen only in terms of primary production (especially related to livestock), but also by changes in logistics in terms of resource use. Also relevant is nutritional change, where switching from animal protein to more plant proteins could have an impact on emissions.

At the same time, there is an increase in food-related diseases and health problems (i.e. obesity, malnutrition, cardio-vascular diseases), which is often associated with certain social classes in society, that have less healthy dietary standards. In this context, efforts are being made to preserve the flavours and enjoyment of foods, while reducing salt and fat contents in food. This also requires raising awareness and making healthy choices for all in society across the global. Improving the nutritious content of food also requires changes in breeding and often not only in commodity crops, but also in relation to crops with higher value-added. The lack of consumer confidence in the food chain has increased over the last decades due to food safety issues and food fraud, where these stories have spread very rapidly through the social media.

Finally, all these challenges in the food system have to be achieved within the planetary boundaries (Steffen et al, 2015), as only one planet exists. The food system here intrinsically links to other sectors like energy, water, transport, etc. The Stockholm Resilience Institute has pointed out that genetic diversity, phosphorus and nitrogen are more at risk from a planetary boundaries point of view. Resources such as land or water are limiting factors to increase their use beyond certain limits. To increase agricultural production, more land could be converted into agricultural land, but this is undesirable on a large scale, as it would be at the expense of natural areas and would strengthen deforestation. Similarly, for water, water is also required for other sectors (e.g. cities) and efficiency of water use can be massively improved by using advanced irrigation techniques. Other sectors examine agricultural products as raw materials, for example in bio-plastic or fuel-related sectors.

Some developments could help to find solutions to these challenges. Firstly, an important development is the emergence of a system-perspective where the interconnectedness of challenges is highlighted, instead of being considered as separate independent issues. This food system perspective increases the complexity of the discussions in the short term, but in the long run ensures that responses can be much more targeted and balanced, while considering explicit the trade-offs in the system and the unintended consequences of food system interventions. Secondly, there are significant advances in breeding and genetic techniques, also in non-commodity crops and livestock types, that allow more rapidly and more adaptable development of new breeds fit the local environments: Techniques such as CRISPR-CAS and large-scale phenotyping. Third, there is a rapid digitization of the food system, imprinted as a digital revolution, digital food, digital agriculture & big data.

3.2 THE POLICY FRAMEWORKS

The challenges facing the food system have not been overlooked in the context of developments at international, national or regional level. The most important development is the multi-lateral process of defining the Sustainable Development Goals, a process lead by the UN with strong commitment from all nations around the world. Many of the Sustainable Development Goals have a link to the food system, as visualized in Figure 1. The Sustainable Development Goals are a visionary goal-setting process, where each Member State reports on their achievements and strategy to achieve them. Specifically, regarding the climate change issue, at COP21 in Paris, agreement was reached between nation states to address the implementation of mitigation and adaptation strategies, also for agriculture and nutrition.

In Europe as a complementary action to the SDG processes, the European Commission DG RTD formulated FOOD 2030, a vision on a food systems approach to address the challenges, and its link to knowledge and research. FOOD 2030 promotes a food systems approach, to overcome fragmentation, facilitate collaboration and bridging across disciplines and challenges.

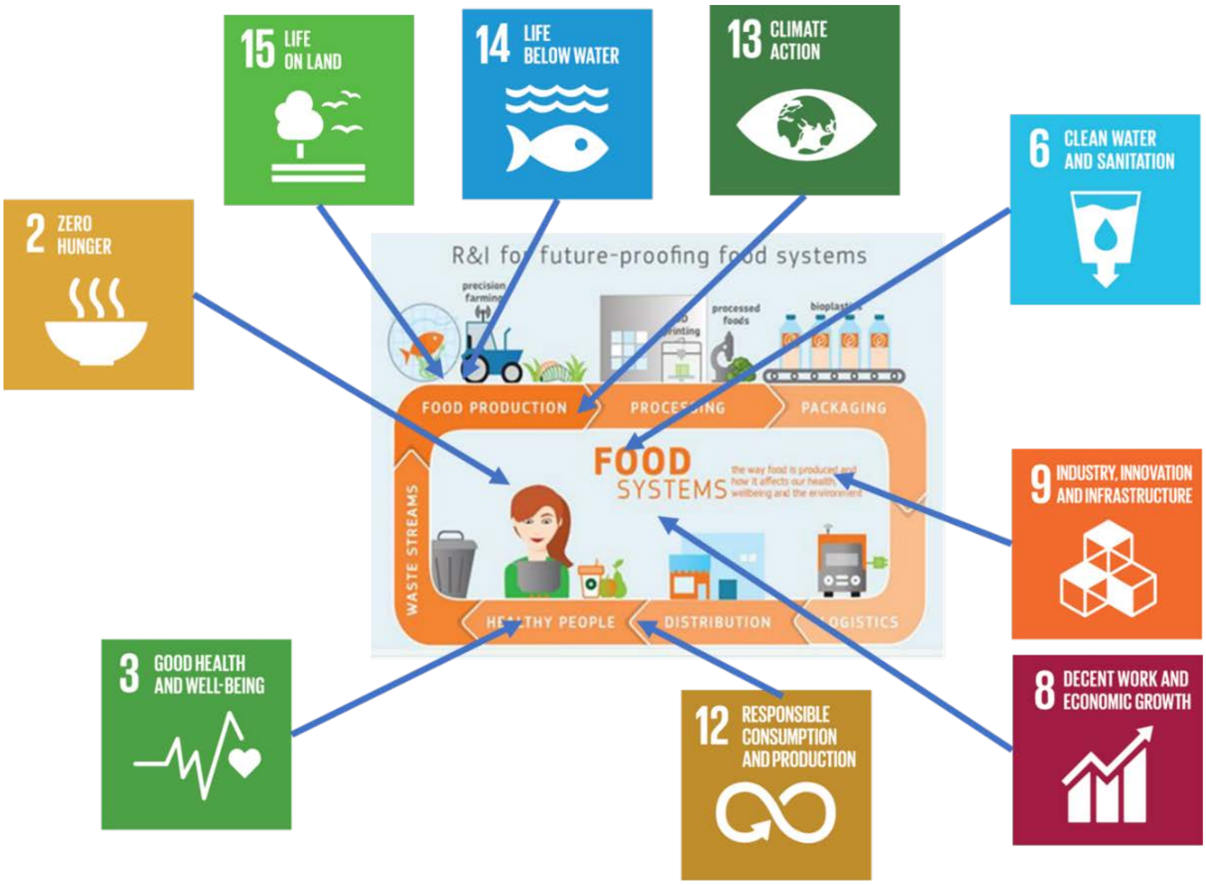

Figure 1: Connecting the SDGs with the different elements of Food Systems as incorporated in DG RTD's FOOD 2030 Vision

In parallel to these challenges driven policy frameworks, two cross-sectoral policy initiatives are relevant, the European Open Science Cloud and the Digital Economy initiatives by DG-Connect, are also closely linked to initiatives at Member State level. The European Open Science Cloud seeks to federate digital infrastructures to support open science practices and foster innovation. The ambition of the Digital Economy initiatives is to increase competitiveness & economic growth in all domains including agriculture, food & nutrition sectors, thanks to digitization and automation.

3.3 SCIENTIFIC CHALLENGES

Science needs to be a crucial provider of solutions and directions for the future development of the food system to be more sustainable, resilient & equitable. The overarching research challenge across domains can be better summarized, as the need to design methods to better target farmers/consumers/value chain actors while improving efficiency, reducing environmental burdens and improving health. More data allows for more accurate understanding of the different components of the food systems and their interactions, while recognizing the trade-offs between these food system components, thus requiring the study of the interactions and a better understanding of the missing elements.

From this overall research challenge a clear need for trans-disciplinary research can be distilled in which stakeholders (e.g. policy officers, business actors, civil society) are directly involved in the set up and execution of the research. Also, studying the food system and better targeting requires a multi-domain inquiry, in which disciplinary limits are less apparent.

The overall scientific challenge and related transformations in science can better be understood by considering more specific issues (like ending hunger, climate action, ensuring food safety, safeguarding genetic & bio-diversity) with associated impacts, societal actors, institutions and the role of science in these. Therefore, the Food System (in figure 3.1) can conceptually be divided in three main categories.

Table 1: Research supporting SDG2 Ending hunger

Impact: SDG goal 2 of Ending Hunger, it is formulated as ‘End hunger, achieve food security and improved nutrition and promote sustainable agriculture’ which describes the vision of the desired impact by 2030.	**Beneficiaries:** Agricultural supply chains, consumers **Users:** SDSN, national governments, agricultural supply chains, national SDG academies, NGO’s and citizen organizations
Role of Science: Research acts as a facilitator, as sounding board, and as a developer of new methodologies. With respect to facilitation, as research organizations and individuals often do not have a specific view or opinion regarding how to reach sustainable development, they can organize the fora to bring the different end user groups together and design jointly the interventions.	**Road to open science**: • to develop new integrated scientific methodologies and implement them with many different data sources and information points • combination of methodology development and application to support these methodologies is extremely challenging

3.3.1 Food, nutrition, health and food safety

In terms of societal benefits, the role of the consumer is crucial. Consumer welfare and the optimization of consumer decision-making for healthy and / or sustainable products / lifestyles are general objectives. This means that we need to understand more about the consumer and his/her choices. Another impact is the efficiency gains in the supply chain, thus improving the supply chain. Improvement of food and health risk assessment, timely production, processing and sales is another impact. Risks should be estimated early, and possibly in a preventive approach instead of a response -to-outbreaks mode. However, if an outbreak occurs, monitoring and detection should be quick and the faster the response, the more targeted it may be, and the less its effect may be. Finally, there is a greater potential to use data-intensive technologies to reduce food waste, as many food is wasted.

Each of these societal impacts links to different research challenges, as research needs to produce knowledge, research products, tools and data to facilitate the achievement of the impacts. For the consumer/nutrition perspective, research needs to connect the intake of food products to the health status of individual consumers, to find out the different impacts. Furthermore, this could be connected to agricultural production and its environmental impacts in a next step. In order to address efficiency gains in the supply chain, research should identify cost-effective ways of improving and tackling efficiency, without increasing transaction costs and overhead costs in the supply chain.

For the impacts around food safety, research challenges are the identification of emerging risks, in a very targeted way. This should be linked to the design of adequate responses at each stage of the supply chain and the management of outbreaks. With regard to food waste and food losses, research should develop a comprehensive data-based understanding of food waste and food loss, wherever it may be, how it can be mitigated, and what effective solutions to tackle it.

Table 2: Food Safety

Impact: Food safety is an integral property of food and agri-products that must be secured for health and wellbeing. The performance of a food supply chain is affected directly and/ or indirectly by many factors such as climate, economy and human behaviour.	**Beneficiaries:** Consumers **Users:** EFSA, National Food Safety Authorities, Testing lab's, supply chain companies
Role of Science: • The opportunity to exploit the new technologies developed in relation to the monitoring and use of citizen science. • New models will be developed that enables integration of complex datasets enabling the prediction of food safety problems at an early stage	**Road to open science**: • Models enabling systemic approach of food safety • Infrastructure (i.e. Web environment) to automatically collect, process, store, compute, visualise and integrate data and information from various origins and natures. • A food safety data platform to collect and connect, compare and share information about food safety generated across Europe.

3.3.2 Gene-based approaches from omics to landscape

The food system uses resources occurring around it to support its functioning and ultimately its performance. These resources can be genetic, biophysical or social, as captured by the landscape as an integrating level at which many interactions occur. The overall societal challenge is to achieve a sustainable use of the resources from the food system, to preserve biodiversity, and there are some associated sub-challenges, which also integrate and combine different resource uses. Firstly, the development of efficient plant and cattle breeding to provide genetic solutions to disruptive food production changes. Such breeding must produce robust and resilient breeds and varieties to external changes and should therefore focus on different aspects than the traditional productivity perspective. Breeding thus must focus on supporting non-intensive (e.g. low input, organic, small scale) types of farming, targeting efficient resource use and stability of production with a high disease and pest resistance. Secondly, invasive species, pests, and diseases are major challenges to food security and new diseases are spreading rapidly, sometimes having an impact on human health. This necessitates the development of new varieties and breeds, but also innovative control mechanisms less dependent on chemical solutions and management interventions that increases the natural control of agro-ecosystems and reduce the chance of emergence of pest and diseases in the first place or the spread of invasive species. Thirdly, the efficiency of resource use and circularity in the food system must be further improved, to ensure an adequate supply of resources in other sectors, which means that the food system has to produce more with less. This also entails the safe-guarding of natural resources at the landscape level, which is linked to a vibrant countryside and urban food landscapes, as well as to the richness of biodiversity.

To achieve these societal impacts, science must better target the needs of stakeholders and lead to the development of participatory techniques for breeding, trade-off analysis, enabling discussion and tailoring. An unresolved scientific challenge is crossing the scales from genes & cells through organisms to population (field, herd) & landscape. When crossing each scale, there are specific scientific questions suggesting that great growth is needed to achieve these multiple benefits and also that they are capable of truly operating on a scale. This links to the point that ecosystem circularity and resilience, up to now, are still poorly understood. Opportunities are well-known, but science is not necessarily strong enough to quantify them or to have a precise understanding of the trade-offs, tipping points and steady states of food-producing ecosystems. Other challenges in relation to breeding are in the speed of extrapolating innovations from lab to fields situations and large-scale phenotyping and characterisation of the environmental components, or GxExM interactions (Gene-Environment-Management interactions).

Table 3: Phenotyping: from omics to landscape

Impact: Genes are preserved in their diversity and managed from the individual level to the landscape scale in order to develop a multifunctional agriculture that is more resilient in the face of perturbations.	**Beneficiaries:** farmers, food systems industries, future populations **Users:** breeders, landscape managers, policymakers at national and international level
Role of Science: To know and preserve the genetic patrimony, to better understand the Genes-Environment-Management interactions from individual to population to landscape scales. To take into account interactions between the agri-ecosystem “actors”: crop, livestock, pests, auxiliaries. To support innovation in terms of biodiversity–based breeding for multifunctional agriculture	**Road to open science**: • enable cross-scale and cross-kingdom (animal, plant, microbial) data integration and analysis facilities and skills • structure scientific communities and users’ collaboration around gene-based and phenome-based e-infrastructures • towards high-throughput analysis of phenomics data • develop participatory supervision of biodiversity

3.3.3 Sustainable farm livelihoods, food security & the environment

Smart farming, food security & the environment concerns itself mostly with the first step in the supply chain of food production up to farm gate. Many challenges come together on the farm as the production unit. From a societal point of view, the farmer and his/her farm is a crucial actor in ensuring sustainability of the food system in the environment, in terms of resource efficient. In all farming systems, the farm as an economic enterprise is directly linked to rural wellbeing and livelihoods, as providing rural employment and incomes. Thus, innovations on the farm level need to incorporate all socio-economic, environmental and productivity aspects at the same time. In industrialized countries, a trend is that agriculture as a sector is becoming increasingly industrialized with high capital costs, scale increases and high input use. This has some undesirable side effects such as environmental pollution, difficulty of transferring family farms to generations, population decline in rural areas and strong dependence on financial institutions. Consequence is a lower societal acceptance of industrialized agriculture. In many developing countries, agriculture is remaining in low productivity on smaller farms with poor soil conditions and expanding land use at the expense of natural areas and inefficient use of fertilizers and plant protection agents. While the variety of farms is increasing with also more commercial farms, there are still many smallholders who produce mainly for food self-sufficiency in food with an incomplete dietary mix. Even in developed countries the economic situation of many farmers is marginal. Agricultural transformation by today's low-productivity farmers into a vibrant, productive agricultural sector that maintains employment in the countryside is essential. Aquaculture and fisheries must play a special role in achieving sustainable use of the oceans and the water system.

With agriculture at a crucial crossroads, science and knowledge production play a role in its transformation. As a scientific challenge, the trade-offs at the level of the agricultural system need to be understood and explained much better, which also translates into the need to define system interventions (in response to climate change) in order to reach the best point in existing trade-offs and to allow win-wins as much as possible. Digital technologies and robotics allow for a much better and more detailed targeting of such a balance between multiple targets. Science needs to develop methods and demonstration to absorb such digital technologies in the primary agricultural production system quickly. This also requires advances in analytical methods to ingest many different data sources and integrate them towards insights for farmers and their advisors, from industrial to subsistence farming. From an organizational and socio-economic point of view, an integration of primary production in value chains is required in way that is fair and sustainable, also from the point of rural employment and vibrant country sides.

Table 4: Global Agricultural monitoring and early warning systems

Impact: Better predictions of famines, drought and agricultural production allow an earlier policy and disaster to alleviate the response.	**Beneficiaries:** farmers, rural population **Users:** GEOGLAM, policy makers at national and international level, FAO, UNWFP, development banks, insurance companies
Role of Science: Innovation in the development and validation of methods and tools required in the fields of data acquisition, data analytics, modelling and decision support integrating agronomic, climate, soil and weather data.	**Road to open science**: Improving the availability of research infrastructures (HPC, storage, grid). Improving the availability and access to data and the capacity to work with Remote Sensing data and other data sources; Development and testing of big data analytics solutions for geospatial data.

3.3.4 Transformation towards open transdisciplinary science

To meet the scientific challenge of providing more targeted and better tailored advice to consumers, farmers and supply chain actors, while turning science into a more open:

- **Share**: sharing of the resources of relevance to the scientific process (data, models, papers, etc);
- **Connect**: available resources need to be connected to allow integration, and tackling large scale and more ambitious questions in science;
- **Collaborate**: the research community itself needs to collaborate beyond ad-hoc arrangements to create, maintain and supply domain specific resources for open science in a network of regional or domain nodes.

4 LANDSCAPE OF THE AGRI-FOOD COMMUNITY ASSETS

4.1 CONNECT: A HETEROGENEOUS AND SCATTERED LANDSCAPE

The agri-food sector relies on a complex science that requires multidisciplinary (from genomics to social sciences), multiscale (from genes to ecosystems), time- and geolocation-based approaches. This implies a significant amount and variety of data and models, which has been increasing exponentially with the adoption of more and more systemic perspectives, the automation of data collection (e.g. thanks to robots, UAV, connected sensors), new engineering tools such as in the omics field, as well as with the development of new types of data sources (e.g. Internet of Things, crowdsourcing, text mining).

As a result, agri-food data are extremely heterogeneous and scattered. Most data are produced, stored and processed in institutional repositories, and to provide FAIR (Findable, Accessible, Interoperable and Reusable) data with a high level of quality and interoperability is a challenge. However, many sources of data are open (as illustrated by the 300 data points on the Map of Agri-Food[3]) and are technologically mature (e.g. data aggregators and repositories), which represents a significant resource that the agri-food research community can leverage on. Initial efforts to enhancing agri-food data discoverability have been made, for example by creating catalogue CIARD RING and the data aggregator AgroLD (see Annex 1 for further information). Improvements are required to ensure a high metadata quality, more significant coverage and interoperability as well as APIs (e.g. BrAPI for plant breeding databases) that support the efficient access to relevant data.

CIARD RING ***(http://ring.ciard.net/)***

The CIARD RING is a federated and curated metadata registry of agri-food datasets and data services. It contains machine-readable metadata profiles for more than 5000 datasets and is fully Linked Data. It is facilitated by GFAR[4] and registered in FAIRsharing.org.

AgroLD ***(http://www.agrold.org)***

AgroLD (Agronomic Linked Data) is a knowledge-based system that exploits the Semantic Web technology and some of the relevant standard domain ontologies, to integrate genome to phenome information on plant species widely studied by the agronomic research community.

BrAPI ***(www.brapi.org)***

The Breeding API, or BrAPI, project is an international initiative to create and maintain a web service specification enabling interoperability and federation of plant databases dedicated to Breeding, including Phenomic, Environment, Genetic variation and association, Genomic Selection. It is being used by IT tools and services from CGIAR and Elixir (INRA, WUR, VIB etc.) and its adoption is growing.

Common standards are required to support interchange of data and achieve interoperability between data storage repositories. In this view, specific resources have been developed to enhance the adoption of standards and good practices in the agri-food domain, such as the Wheat Data Interoperability Guidelines for the wheat research community. Semantic resources in the agri food sector are highly fragmented and unevenly developed among disciplines. Common standards must be enforced and better

[3] http://www.aginfra.eu/discover
[4] http://www.gfar.net/

access is required. More standard semantic resources must be created and shared. This is the focus of several initiatives, including GACS (Global Agricultural Concept Space), the Map of Data Standards, AgroPortal and Crop Ontology, which are under the umbrella of the "Agrisemantics"[5] initiative (see Annex for further information). There is also a need to make semantic technologies and methodologies more accessible to all roles in the agri-food data management workflow.

Wheat Data Interoperability guidelines *(http://datastandards.wheatis.org)*

The Wheat Data Interoperability Guidelines consist of a set of recommendations developed by the Wheat Data Interoperability Working Group[6] for describing and representing data with respect to existing open data standards. They have been endorsed by the International Wheat Initiative WheatIS Expert Working Group (see below) and the Technical Advisory Board of the Research Data Alliance[7].

Agrisemantics Map of data standards *(http://vest.agrisemantics.org)*

The Agrisemantics Map of data standards is a catalogue of data standards of different types (schemas, taxonomies, thesauri, code lists, ontologies etc.) for the agri-food domain, categorized according to subdomain, types of data, format and other criteria.

Global Agricultural Concept Space (GACS) *(http://agrisemantics.org/node/8)*

GACS is a set of concepts commonly used in distributed Knowledge Organization Systems like thesauri, taxonomies, or ontologies. GACS commits to maintain persistent dereferenceable URIs on the long-term so that they can be used as references. In that sense, it is the first step toward a mapping of the data through a community of interconnected, interoperable, semantic assets and services relevant to agriculture and food security. GACS is supported by FAO, USDA/NAL, CABI, INRA, LIRMM and Tor Vergata University.

AgroPortal *(http://agroportal.lirmm.fr)*

The AgroPortal project aims at offering a vocabulary and ontology repository for agronomy and related domains such as nutrition, plant sciences and biodiversity, including services such as ontology browsing, search, recommendation, text annotation and mappings.

Crop Ontology *(http://www.cropontology.org)*

The Crop Ontology is a framework to create standard ontologies for annotating phenotype observations in breeding, agronomy and agro-ecology. A Community of Practice supports the Crop Ontology: http://tinyurl.com/COcommunity.

4.2 SHARE: THE EXISTENCE OF INITIATIVES THAT SUPPORT SHARED DATA HANDLING

Many large-scale initiatives are ongoing at European and international level and seek to embody governance and community-building frameworks, enhance technology adoption or support data discovery & access at a more technical level (see Map of Agri-Food http://map.aginfra.eu/).

[5] http://www.agrisemantics.org

[6] https://www.rd-alliance.org/groups/wheat-data-interoperability-wg

[7] https://rd-alliance.org

In particular, initiatives that provide a strong political support are key in the implementation of more technical initiatives and actions. This is clearly illustrated in the domain of wheat research with the example of the international Wheat Initiative, or AIMS and GODAN networks, which bring together key funders in the field of agri-food to support the open data movement. Indeed, The Wheat Initiative has given a significant boost to the development of a specific data system for wheat called Wheat IS. In parallel, RDA-IGAD supported the establishment of the Wheat Data Interoperability Working Group, which is tightly linked to Wheat IS.

At a more institutional level, political support is required to provide incentives for researchers to share their data and make them FAIR. This is for instance the aim of the CGIAR Big Data Platform as well as CIAT's FAIRness monitoring system (see Annex for further information).

GODAN *(godan.info)*

The GODAN (Global Open Data for Agriculture & Nutrition) initiative launched in 2014 by the USA and UK at its centre supports open access globally to datasets for agriculture and nutrition. The GODAN steering group is now supported by several countries, including Mexico, Germany and major international Institutions such as FAO or the Chinese Academy of Agricultural Science.

The RDA Interest Group for Agricultural Data *(https://www.rd-alliance.org/groups/agriculture-data-interest-group-igad.html)*

The RDA Interest Group for Agricultural Data (RDA-IGAD) is currently the most prominent domain-oriented group (over 170 members) in RDA, working on all issues related to global agriculture data. It is also a forum for sharing experiences and providing visibility to research and work in agricultural data.

AIMS *(aims.fao.org)*

The underlying goal of the AIMS community is to promote and manage information and capacity development activities on Information and Data Access Services, in particular on Access to Global Online Research in Agriculture (AGORA)[8], Agricultural Science and Technology (AGRIS)[9], semantics services[10] including AGROVOC Thesaurus. Additionally, AIMS also trains all interested parties, free of charge, through Open Data Management in Agriculture and Nutrition online course[11], in collaboration with GODAN Action[12].

The Wheat Initiative *(http://www.wheatinitiative.org/)*

Endorsed by the G20 Agriculture Ministries, the Wheat Initiative provides a framework to establish strategic research and organisation priorities for wheat research at the international level in both

[8] http://www.fao.org/agora/it/

[9] http://aims.fao.org/agris-network

[10] http://aims.fao.org/AIMS_semantics

[11] http://www.godan.info/pages/open-data-management-agriculture-and-nutrition

[12] http://www.godan.info/godan-action

developed and developing countries. One of the Wheat Initiative's Expert Groups focuses on the development of a Wheat Information System (Wheat IS)[13].

The CGIAR Big Data Platform *(http://bigdata.cgiar.org/)*

The new CGIAR Platform for Big Data in Agriculture will harness the capabilities of big data to accelerate and enhance the impact of international agricultural research. The Platform focuses on standardizing, opening, sharing, and enabling analysis and visualization of agricultural data. It will provide global leadership in organizing open and FAIR data, convening partners to collaborate on developing innovative ideas, approaches, and tools, and demonstrating the power of big data analytics through inspiring projects**.**

EFITA *(http://www.efita.net)*

The European Federation for Information Technology in Agriculture, Food and Environment is a professional society with member organizations in 15 European countries. Its biennial meetings are important concentration points of discussion on the development of Information Technology as in precision agriculture, sensors, robotics, field data evaluation and semantic interoperability. It has also an important function to bring together the industry sector with academy.

Professional Societies like EFITA have become important over last decade for dialogue between various stakeholders.

Indeed, collaboration is also becoming crucial for private R&D as discovery becomes more and more difficult and cost is rising. There is a strong need for pre-competitive cooperation amongst all players globally. This is for instance the aim of the AgGateway initiative, which supports the implementation of standards for data interoperability across private stakeholders. Individual companies such as Syngenta are taking the lead in opening their data in order to support open, collaborative innovation among private actors but also towards public research and initiatives.

Syngenta - The Good Growth Plan - One planet. Six commitments. *(https://www.syngenta.com/what-we-do/the-good-growth-plan)*

The Good Growth Plan consists in Syngenta's programme to reach six strategic commitments around resource efficiency, soil and biodiversity conservation practices, and rural prosperity including smallholders. It supports a global network of farms where technology adoption and efficiency performance are monitored each year by an independent company. Data is collected and shared in order to monitor the progress towards Syngenta's commitments, but also to inform on innovative and sustainable farming practices and allow other stakeholders and organisations to analyse it, which can feed into Syngenta's work.

AgGateway *(www.aggateway.org)*

AgGateway is a non-profit consortium of over 200 businesses in the agriculture industry, dedicated to promote and enable the industry's transition to digital agriculture, and expanding the use of information to maximize efficiency and productivity. Primarily dedicated to standards implementation,

[13] http://wheatis.org/

AgGateway provides a collaborative environment where companies can meet to identify pre-competitive spaces within which they can cope with pressing interoperability problems faced by the industry.

4.3 COLLABORATE

4.3.1 A broad spectrum of existing knowledge and skills

There is an increasing interest in the issues of Open Science and Big Data in the field of agriculture[14]. In particular, advances in knowledge and technology have been made regarding:

- Information technologies and systems;
- Semantics, with a focus on ontologies;
- Data management and storage issues for farm management;
- Data management and sharing across the food supply chain;
- Crop modelling and other knowledge-based applications;
- Emerging hot topics such as machine learning, Internet of Things and Big Data.

The bibliometric study conducted by the project shows that key European institutions such as INRA and WUR are strongly involved in addressing these issues. It shows some specialisations, for instance, INRA carries out several research activities on knowledge engineering while WUR is more involved in modelling. More broadly, there is already significant collaboration at the European level, especially between the Netherlands, Spain, Denmark, Germany and Greece, thus highlights the leading role of the EU in providing the skills and knowledge to tackle the data challenge.

In addition, there is also a strong leadership of non-European countries and organisations, especially China and the USA, with significant collaboration between these two countries, China has a strong focus on the Internet of Things (e.g. sensors) and the USA on information technology. The international dimension of the open data movement in the agricultural field is illustrated by the involvement of other organisations in India and Australia, as well as the FAO, which plays a central role in collaborating with both European and non-European institutions and in supporting efforts on Open Data, Linked Data and semantics.

As a result, the potential for capacity building and upscaling of the skills and knowledge required to produce and use FAIR data in the agricultural sector is significant at European and international level. Strong support from funders and policy-makers is required to fully take advantage of it.

4.3.2 The development of new services to process and analyse data

Several European and international facilities have been developed (see Map of Agri-Food), providing various data-related services including data discovery & access, data integration, data analysis and visualisation, statistics & modelling. Many of them deliver operational services that are ready for use and openly accessible, and that can address key issues regarding the use of data, including the comparison and interoperability of models, the development of reproducible and reusable workflows, and the creation of Virtual Research Environments that provide tailored resources, solutions and tools to specific research communities.

[14] According to e-ROSA's bibliometric analysis, the number of scientific articles related to these issues has risen from 142 references in 2005 to 642 references in 2015.

AgMIP (*http://www.agmip.org/*)

The Agricultural Model Intercomparison and Improvement Project (AgMIP) is a major international effort linking the climate, crop, and economic modelling communities with cutting-edge information technology to produce improved crop and economic models and the next generation of climate impact projections for the agricultural sector.

MACSUR (*https://macsur.eu/*)

MACSUR is a knowledge hub within FACCE-JPI (Joint Programming Initiative for Agriculture, Climate Change, and Food Security)[15] that gathers the excellence of existing research in livestock, crop, and trade science to describe how climate variability and change will affect regional farming systems and food production in Europe in the near and the far future and the associated risks and opportunities for European food security.

AGINFRA+ (*http://www.plus.aginfra.eu/*)

AGINFRA+ addresses the challenge of supporting user-driven design and prototyping of innovative e-infrastructure services and applications for the agri-food community, building on core European e-infrastructures such as EGI.eu[16], OpenAIRE[17], EUDAT[18] and D4Science[19].

4.3.3 Heterogeneous practices within the research community

Agri-food science relies on a broad variety of disciplines, which implies different practices and a varying degree of maturity in terms of data management skills throughout and within the various research sub-communities. This is for instance illustrated within the wheat community as stated in the survey[20] from the RDA Wheat Data Interoperability WG. It reflects the different research approaches that relate to hypothesis-driven research vs. data-driven research, going from the use of Excel files to artificial intelligence technologies.

Therefore, there is a strong need to build a community of practice through stronger community management in order to promote a data sharing culture. Bringing together the various stakeholders that are involved throughout the data management and research cycle (e.g. researchers, data scientists, data producers and ontology specialists) is key in this process.

More specifically, a strong support to scientists is required in making data FAIR, especially regarding legal and intellectual property issues, data ownership and related licences, as well as the use of developed standards and the provision of sufficient, quality metadata for data interoperability and discovery. Several capacity-building resources and activities are currently under development, for instance the online platform of INRA datapartage.inra.fr, the training for PhD students in WUR and the curriculum developed under the GODAN Action project (see Annex for further information).

[15] https://www.faccejpi.com/

[16] https://www.egi.eu/

[17] https://www.openaire.eu/

[18] http://eudat.eu/

[19] https://www.d4science.org/

[20] http://doi.org/10.5281/zenodo.293759

Datapartage.inra.fr *(www.datapartage.inra.fr)*

Since 2013, INRA has been actively undertaking a series of initiatives and projects to develop its research data management capacity. The datapartage.inra.fr website is part of this effort and aims to:

- inform and provide guidance on INRA's data management and sharing policy;
- share best practices;
- give access to tools and services provided by the Institute to support data management and sharing.

WUR's course on Research Data Management *(https://wgs.crs.wur.nl/courses/details/84/)*

Wageningen University and Research offers an introductory course for PhD candidates and postdocs on research data management (RDM), focusing on data collection and management, WUR's policies and services for RDM, and long-term storage and data publication. The full course content is available at: https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/E8M8GI.

The GODAN Action Curriculum *(http://www.godan.info/sites/default/files/documents/Curriculum_final.pdf)*

GODAN Action[21] is a DFID funded project that brings together agriculture and nutrition specialists and open data experts to support individuals, organisations and communities to engage with open data. The GODAN Action Curriculum was developed as part of the capacity development activities and is to be used in the context of different institutions in agricultural and nutrition knowledge networks and raise awareness on the different types of data formats and uses, and on the importance of reliability, accessibility and transparency.

Despite the complex landscape, consistent collaboration mechanisms have been developed in the last years. Many of the European main players (as Wageningen UR, INRA, CABI, CGIAR and FAO) are in continuous working contacts. International relations, especially with China (CAAS) and Brazil (Embrapa) have been established. The GODAN initiative has enrolled more than 600 partners worldwide on the goal of open data in Agriculture and Nutrition. In the area of access to shared semantic resources the "Agrisemantics" initiative has been setup, bringing shared resources like Agroportal, the GACS and the CIARD Ring into the centre of common efforts.

It is now time to organize these common activities in an institutional framework and to assure sustainable business operations for a commons of shared standards, tools and methodologies. The establishment of GO FAIR implementation networks shows the way to go.

4.4 MATURITY ASSESSMENT OF AGRI-FOOD COMMUNITY DIGITAL ASSETS

As depicted in chapter 4, currently there are several elements of an e-infrastructure (as defined in section 3.4) in place. The following table aims to assess the maturity of these elements based on their technology readiness level across different items and link them with (a) the main challenges of share, connect and collaborate as defined in section 3 and (b) the EOSC initiative.

[21] http://www.godan.info/godan-action

Table 5: Maturity status of elements based on their technology readiness level

Item	Status and needed action[22]	Technology Readiness Level	Refers mainly to	Link to EOSC
Network of trustworthy repositories	There are scattered repositories all over the landscape which need to be organized in the network. There is a huge amount of data Fairization work to do. Good practices have to be promoted	3-9	Share	• Exchange of experiences • Definition of common rules for trustworthiness • Certification mechanism
A registry of trustworthy repositories	CIARD RING as prototype is a powerful tool to organize the network of repositories. But the CIARD RING at the moment is only a prototype with a lack of critical mass. In collaboration with the FairSharing initiative a metadata schema for datasets needs to be released that should be attached to any published dataset	5-6	Connect	Using as much as possible common metadata elements
API registries and PID systems	No domain specific activities should be undertaken, collaboration should be searched for with other EOSC partners	3-6	Connect	Using EOSC resources
Data type registry and services	Need of services to address the issue of DT (cm, C°, F° ...), data typing is key in data sharing and reuse[23].	2	Connect	Services such as http://typeregistry.org/registrar/ should be accessible through EOSC
Metadata Schemes registration	Work is very advanced with the former VEST registry. Coordination and discussion with other similar enterprises in other domain is needed	6	Share	Alignment with the FAIRSharing Registry and other EOSC metadata scheme registries
Ontology and concept repositories	The domain is comparatively advanced having available operative prototypes of Ontology Servers (such as Agroportal) and Concept Server (such as GACS). The existing system are at the level of usable prototypes.	6	Collaborate	Opening Services to EOSC partners as they are not limited to the domain

[22] https://eoscpilot.eu/content/erosa-position-paper

[23] http://hdl.handle.net/11304/c1431d2e-4254-11e5-8a18-f31aa6f4d448

	Investment is necessary to bring to productions level			
AAI mechanisms and services	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners	8	Collaborate	Using EOSC resources
System of license registries	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners	4-5	Collaborate	Using EOSC resources
Ecosystem of tools and operating procedures	Some scattered work has been done but without systematization; It needs to be integrated with the effort to create competence centers	3	Collaborate	Sharing tools and experiences

5 TOWARDS A EUROPEAN E-INFRASTRUCTURE IN AGRICULTURE & FOOD SCIENCE

5.1 THE NEED FOR AN E-INFRASTRUCTURE

In order to efficiently establish a mechanism for "sharing", "connecting" and "collaborating" an e-Infrastructure is needed for the agri-food community. According to the European Commission, an e-infrastructure is usually described as a combination of digital technologies (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organisational structures needed to manage them. They are key in future development of research infrastructures, as activities go increasingly "online" and produce vast amounts of data and are at the heart of the Digital Agenda of the European Commission to support open science and the link between researchers, citizens or private companies.

Thus, an e-Infrastructure has the following characteristics:
1. it is based on and useful for a community
2. it needs governance and a business model for sustainability
3. it needs a technical backbone, that easily can host a variety of resources and services
4. it comprises a variety of resources and services, useful for the community, interoperable, but not necessarily connected or dependent from each other

Consultation of a broad group of stakeholders from the domain of agriculture and food by the e-ROSA project has resulted in a set of requirements. These are based on the observed barriers and challenges, expectations and opportunities, these stakeholders identify from the specific areas of the agri-food domain in which they are active.

5.2 REQUIREMENTS FOR SHARING

Requirements for sharing imply several challenges, among which the one of (i) professionally managing research data, (ii) support for multi-scale, multi-dimensional data, (iii) sharing other research resources alongside with data, (iv) the socio-cultural challenges, (v) the ethics of sharing and (vi) capacity building.

Professional research data management: from the use cases examined by e-ROSA it becomes clear that many research initiatives and communities are struggling with data sharing. First of all, data is hard to discover and access, because of immature data sharing. It still takes huge efforts to get hold of valuable research data, just because it's not easily findable and accessible. To make the tremendous amounts of existing research data available for future data-intensive research, across use cases in agriculture and food, efforts are required in data management in order to make data available in a proper way. In that respect, it is obvious that data managers have an important role, they should move beyond making data available, and also include the broader aspects of FAIRification as well as data governance aspects. In that respect, there's a clear demand for easy to use tools and interoperability platforms that make data sharing practicable for the broader community of researchers in agriculture and food. While some descriptive metadata (e.g. lineage of datasets, documentation of algorithms, models) needs to be manually provided by the author, support for automatic creation and transfer of additional metadata would greatly help researchers to ensure the quality of scientific resources, starting from data collection. A data curator should be as much as possible unburdened from metadata generation tasks, and be better supported by tools embedding standard vocabularies and methodological tips. This would lead to more availability of metadata, of better quality (e.g., up-to-date, detailed or in consistent formats), and consequently an increased potential for the programmatic access of data.

Support for multi-scale, multi-dimensional data: various use cases in the agriculture and food domain report the need to work with multi-scale and multi-dimensional data. Typical for the domain, for example for agricultural monitoring, regional and global impact assessments, climate smart agriculture, is the need to have efficient tools and methods to work with spatio-temporal data, in order to be able to combine different geographical and temporal scales. Another example is the need to be able to work over different resolutions in the use cases of massive scale phenomics in order to better understand gene - environment interactions to be able to anticipate future perturbations. In that case, researchers need to be able to assess the full research domain, "from gene to landscape".

Share all scientific resources, not only data: while there is a focus on e-infrastructures and their capacity to support data sharing, requirements for research infrastructures appear to be a lot broader than the concept of data sharing and the technical and organizational capacities of sharing data. While data is an important fuel for the generation of knowledge within the concept of open science, we need a broader scope with regard to sharing. From the perspective of research, besides data, sharing should also include the sharing of data analytics, models and, at the end, integrated workflows that support the scientific process. As with data, such resources are currently scattered and often not standardized, and sharing them should lead to better reuse of knowledge and methodologies and more harmonized research processes. In short, there is a demand for shared research services, rather than just data services. A good example is the domain of food safety and emerging risks, where currently a lot of models exist, residing in different networks. Similar to data, there's a tendency to use methods that are in scope of a researchers' network, because it is often hard to access and use external ones. Good sharing practices for such resources would support different networks to perform more efficient research.

Address the socio-cultural aspects of sharing: a clear take away, mentioned by many stakeholders in different areas of agriculture and food, is that the social and cultural aspects of sharing knowledge need to be taken into account. A prerequisite for open science to become common practice, is a culture of sharing. This can only be achieved gradually through a combination of advocacy, capacity building and proper guidance and technical support. Moreover, it is clear that such efforts need to be accompanied by mechanisms to reward sharing, which could provide additional incentives for researchers to share their knowledge and speed up the uptake of sharing practices. An example for this is the establishment of open data journals, like the open data journal for agricultural research (ODJAR[24]), that promote publication of (FAIR) data, motivate researchers to document their data and allow data to be cited. Another important issue to deal with in that respect, is that a large part of the research community is still expecting e-infrastructure offerings to be free, or at least affordable. This indirectly links to the demand for viable business models, e.g. in the form of proven business cases, that explain how data value chains can sustainably establish and maintain sharing initiatives. The issue of trust (or veracity) also has to be mentioned here. Many stakeholders indicate that it is often hard to judge the quality and usefulness of datasets that reside outside their own networks. To overcome this barrier for usage of shared data, it is necessary that datasets are well documented and that certification mechanisms for datasets and repositories are in place. Again, recognized, well managed repositories like open data journals could play an important role in building trust and motivating sharing and reuse of scientific resources.

Consider the ethics of sharing: respecting the ethics of sharing, and specifically data sharing is considered to be essential for open science to succeed on the long term. While open science, including scientific work in the public and private sectors, obviously requires the sharing of data, privacy and also competitive stakes should be respected to ensure broad support and uptake. Moreover, many states that the issue of data ownership needs to be addressed in a proper way. There's a specific reference in that case to data ownership of individuals, like farmers or citizens, e.g. in the case of sharing and reusing farm management

[24] www.odjar.org

data or citizen science-based data. Lastly, the issue of having also the less-favoured benefit from open science and e-infrastructures has been frequently mentioned as being desirable. This obviously transcends to a global, rather than a European territory perspective, where for example smallholders and consumers in developing countries should get the opportunity to benefit.

Invest in capacity building: in order to evolve from "classical science" to data science and open science, capacity building is crucial. To reach the vast majority of the research community in agriculture and food, and not only the technical front-runners, and support them to make this transition towards sharing and reuse of shared resources, the sharing of knowledge on open science is required. Practical guidance and support tools must be developed to train researchers to become "open scientists". Examples of topics where capacity building is needed are the FAIR principles, making data FAIR, and the use of tools and methods that allow the reuse of FAIR data. Moreover, there's a need for support on issues such as data stewardship, data curation, data standards, licensing and IPR. As data managers and data stewards will have a central role, training material to educate personnel in these roles is a crucial resource, that needs to be shared across research communities.

5.3 REQUIREMENTS FOR CONNECTING

Better sharing of high quality research data and other resources, like algorithms, models and methods, is an important step towards the envisaged Vision 2030 roadmap for agriculture and food. However, such resources will only be effectively re-used if they can be easily harmonized and connected, not only by technicians, but by the vast majority of European researchers. Only if it becomes possible to easily connect data, methods, models, tools to perform research workflows, e-infrastructures will reveal their full value and only then scientists will be able to conduct open science in a transparent way, maintaining their current research experiences and competences and adding to that the advantages of open science.

Connectivity through standards is required for data intensive research on different levels. It is important to realize that this requires a view on the full data value chain, which is not limited to research, but in many cases includes business and governmental stakeholders. Thus, development of standards and guidance on the use of standards will often need to include the perspective of compliance with legislation and regulation (e.g. INSPIRE) and business standards (e.g. standards and vocabularies in the domains of trade, farm management systems).
There is demand for:

- Mapping the existing landscape of standards, be it data formats, metadata, or other resources standards
- Providing guidance and recommendations on interoperability
- Finding out and bridging the gaps, developing for instance agreed controlled vocabularies

Mapping the existing landscape of standards: there are several recent initiatives to map the landscape of (meta)data standards in the agriculture and food domain including, but not limited to, the RDA Wheat Data Interoperability[25] and Agrisemantics working groups, and the GODAN Action map of agri-food standards[26]. While the data sharing perspective is important, researchers are in the end aiming at performing full scientific workflows, that also incorporate for instance data analytics, models and visualisation components. So again, if it concerns required connectivity, there's a clear call to not only consider the standardisation and semantic connectivity of (meta)data resources, but also to include the standardisation of analytical models and tools. This is particularly relevant since a lot of the current research is performed using models and tools that are in many cases not fully interoperable with current standards (let alone possible future standards). In some cases, they might even be classified as legacy,

[25] 10.12688/f1000research.12234.2

[26] 10.7490/f1000research.1115260.1

and originating from a "pre-standardisation era". One of many examples is the use of numerical models in agricultural monitoring and yield forecasting. These models have often been developed over decades, represent a tremendous amount of knowledge, and have a broad user base. It should be realized that such tools are indispensable in many research areas, and cannot easily be replaced by standard-compliant alternatives. Thus, any open science strategy should incorporate means to adapt (or adapt to) such tools.

Provide guidance and recommendations: Already now, there are a lot of developments in the area of agriculture and food that concern the improvement of interoperability, with the aim to improve connectivity within and over research domains. A lot of standardisation efforts are ongoing, specifically in the area of semantics and the improvement and harmonisation of shared vocabularies and ontologies. Examples are the development of the Global Agricultural Concept Scheme (GACS) and the Crop Ontology. However, many use cases in different areas of the domain report a proliferation of standards, which makes is difficult to choose which standards to use. On top of that, in cases, standards are still developed aiming at specific subdomains of agriculture and food, which implies the risk that currently existing research silo's are maintained instead of being connected. Consequently, there is a strong demand for guidance and for high level agreement on advised standards to use for different types of research applications. The exemplary approach of the RDA Wheat Data Interoperability working group in breaking the boundaries among wheat research community could be generalized and expanded in order to clarify the complex landscape of existing standards and deter communities from duplicating efforts.

Find and bridge the gap: it is generally agreed by stakeholders in the agri-food domain that the identification and rigorous analysis of some strong interdisciplinary use cases would help in further clarify research questions with regard to open science. This could also help to identify in more detail which currently available standards fit such "cross-cutting" research scenarios, which gaps still exist, and how solutions can be developed that have a broad applicability. On the other hand, to ensure short term uptake of open science while advancing on cross-domain interoperability, there's also an immediate demand for guidance and capacity building to support scientists in understanding and effectively using e-infrastructures for research. This might partly be achieved by providing "no-regret scenarios" for open science, that have already proved their value in practice and that are expected to fit into future evolutions of open science. This can be offered through success stories and best practices that are already in place, or by offering learning resources that allow researchers to experiment and work with already available open science services.

5.4 REQUIREMENTS FOR COLLABORATING

Obviously, performing interdisciplinary research in the era of open science supported by e-infrastructures, also implies a different way of collaboration between research communities. E-infrastructures can play an important role in facilitating the collaboration and co-development of (international) research groups. To play such a role, the foreseen e-infrastructure for agriculture and food should (i) allow effective collaboration and co-development, (ii) be user-centered, (iii) build on a sustainable business model.

Allow effective collaboration and co-development: taking aboard some of the previously stated requirements on sharing and connecting, e-infrastructures should allow research groups to establish shared virtual environments, where they can access and work with trusted, high quality datasets, methods and tools. Moreover, they would then implicitly benefit from the high-performance cloud computing and storage facilities that are offered by the underlying e-infrastructure. From a researchers' perspective, they would not be using e-services or e-infrastructures, but rather a virtual research environment that offers transparent access to and seamless use of scientific resources for joint research.

User and research community centered e-infrastructures: a user centered approach build around some strong use cases could help in steering further directions for the technical development of facilities, and for monitoring progress and steering towards effective e-infrastructure support for European open science. It is considered important that environments for open science, while allowing for effective collaboration and co-development, also respect the individual identity and specifics of different scientific domains and research groups. In other words, virtual research environments should not restrict but rather broaden the opportunities of researchers to perform their work. A "system-of-systems" approach, allowing the establishment of regional and domain specific nodes on top of a shared organizational and technical infrastructure for open science, is suggested as a potential model to make this possible. At the same time, this should not impede, but facilitate the cross-nodal connections that are required for interdisciplinary research. Moreover, it should also allow to take aboard individual accomplishments of researchers, and to establish large and small projects to connect to developments in open science. Besides a well thought over, scalable and multi-layered technical architecture, this will require that supporting mechanisms are in place to for example certify open science nodes, and to share resources while respecting IPR, licensing and privacy.

Build on a sustainable business model: it is important to always keep the long-term perspective of open science in mind. Stakeholders feel that a "single project funding perspective" will not be adequate to achieve the longer-term objectives, and a European Research Infrastructure Consortium was mentioned as a possible mechanism. Moreover, there's a call for strategic initiatives that support the roadmap to open science and underpin the importance of open science to advance the agriculture and food domain. One type of such strategic activities is the development of sustainable business models for open science that cover the whole data value chain, take into account viable assets and cover both the public and the private sector. Thus, open science scenarios should not be limited to public sector research but should also include the perspective of private sector research and of funders and funding mechanisms. This also implies that there is an important role for public private partnerships in the co-development and sustainable exploitation of infrastructures for open science. The use cases that were adopted and analysed in the e-ROSA project are a good basis in that respect, as besides broadly covering the agriculture and food domain and a multitude of aspects related to data-intensive science, they are also well connected to stakeholders outside the scientific domain. A second strand of strategic activities concern large scale impact studies, which can underpin these business models as well as the broader societal value of open science. Such studies should clearly outline the Key Performance Indicators (KPI's) that can be achieved through the adoption of open science. They should also, through the monitoring of such indicators, evaluate open science initiatives and seek to advance and show proof of the enabling of innovation through increased sharing in the public, research and private sectors.

5.5 AN E-INFRASTRUCTURE ARCHITECTURE FOR OPEN SCIENCE IN AGRICULTURE & FOOD

As described in the previous sections, the "share, connect, collaborate" paradigm entails many challenges, among which, adequate computational resources, secure data sharing and access, FAIRness of research outputs (data, models. workflows, software, algorithms, metadata, etc.). The concurrent development of Big data approaches (high-throughput techniques, cloud computing, data science) and the Semantic Web offers an unprecedented opportunity to build an infrastructure that tackles these challenges. Beside a necessary technical backbone, our commons (with reference to the Big Data to Knowledge commons[27]) should include resources and services that are Findable, Accessible, Interoperable and Reusable.

The technical backbone consists of:

- Adequate technical resources for storage, transfer, sharing and computing of digital objects
- Effective connections to Laboratory and other research equipment
- Effective connections to Earth Observation facilities, like remote sensing and satellites

[27] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5545976/pdf/nihms891058.pdf

- Facilities that gather data from field operations

FAIR resources and services such as:

- Data channels to production services like gene and protein sequencing, chemical analysis and observation data production.
- Big scientific data workflows (e.g., Galaxy, Taverna)
- Data analysis services
- Semantic support services for annotation and analysis of data (vocabulary and ontology servers)
- Discovery services, like registries for datasets, models, and services. A recommendation system leveraging existing techniques in Big data can be valuable
- Data storage and archival services
- Services to research communities (extension to non-academic players for an open science development)
- Capacity development services with guidelines, courses….

The challenge is to preserve the variability and user centered imposition of front ends and services and to achieve at the same time interoperability between a growing number of resource providers and users in the infrastructure.

It is obvious that this cannot be achieved by one or some centralized processing units, but by an infrastructure that consists of loosely coupled services based on common agreements about policies and semantics. This infrastructure needs to be completely open to accommodate new needs, services, partners or resources.

The common services available now are largely restricted in basic operations, such as (a) Authentication and Authorisation; (b) Document Annotation; (c) Search and Filtering; (d) Publishing and Sharing. Examples of services providing such functionality at large-scale are the EUDAT service collection (B2ACCESS, B2SHARE, B2NOTE, B2SHARE) and the services integrated in the OpenAIRE platform.

At the level of user-targeting tools, there are significant efforts for the establishment of a common, collaborative research environment, e.g. the d4science initiative. Virtual Research Environments provide an important base ground for shifting to truly collaborative and reproducible research. Nevertheless, the overall situation is not satisfactory as it is still enormously difficult to link tools for modelling, analysis, elaboration of data to different datasets, to make them interoperable and to process them simultaneously. Figure 5.1. shows a schematic outline of an infrastructure we propose to overcome these difficulties.

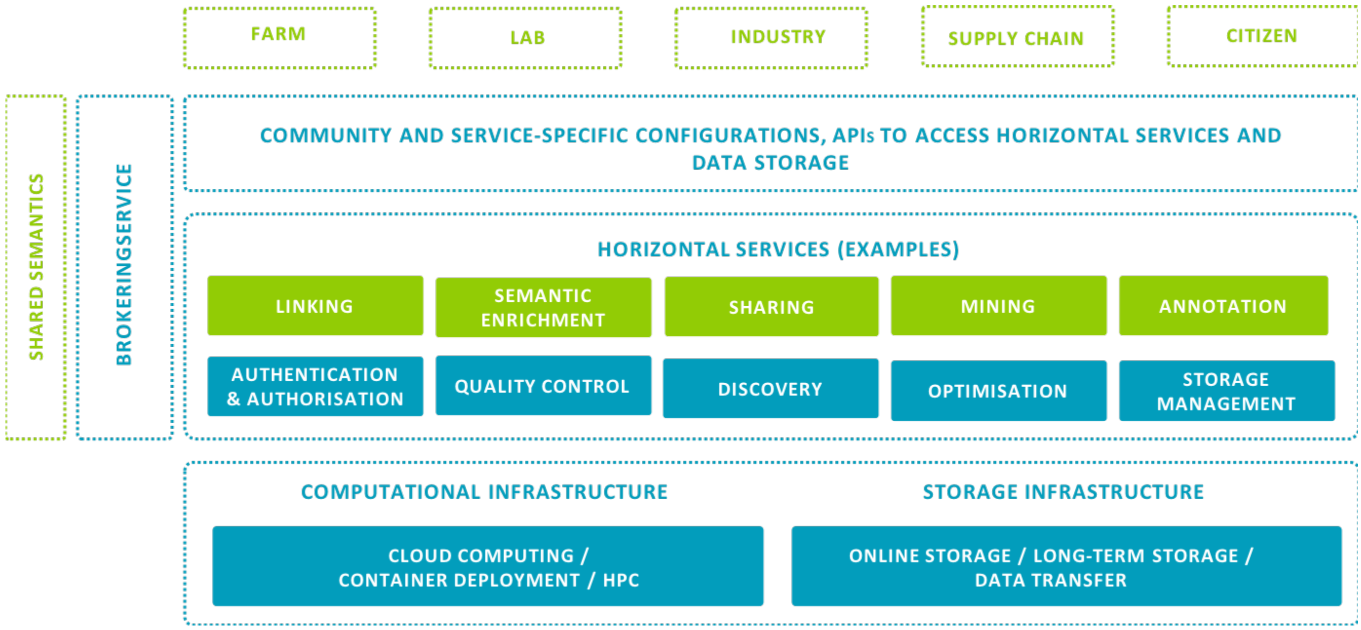

Figure 2: Proposed Future View of an e-Infrastructure for Open Science in Agri-Food

We believe that the next step for building the e-infrastructures of the future is the exploitation of achievements and assets of the different communities via their incorporation in the pool of common, horizontal services to be used across communities and across domains. Some of these horizontal services are not even domain specific (like the computational infrastructure or storage management) and could be delivered on the level of the EOSC.

As different disciplines, however, have different views and requirements, the available horizontal services will be made available to their users via specific instantiations based on *community-specific configurations (encoded in specific APIs)* of a given service.

The invaluable paradigm shift to e-infrastructures will naturally serve as the base ground for the full transition to a shared and collaborative research space. However, there are several aspects that must be considered in order to provide solutions that magnify the benefits and limit the compromises of open science:

- **Findability, Accessibility, Interoperability and Reusability of resources:** scientists must access a shared environment that allows them to find, access, reuse and combine digital objects including datasets, workflows, software, etc. To that aim, all resources and services within the e-infrastructure must comply with the FAIR principles. In particular, we need shared metadata and semantics to standardize, characterize and register the resources and their relations. The metadata that should be associated with each type of resource can be recommended in a guideline in order to harmonize the practices.
- **Resource optimisation**: Despite the importance of establishing a set of widely accessible computational infrastructures, the process of identifying and setting up a computational environment for a given task is not trivial and requires specific expertise not readily available to all researchers and organizations.
- **Semantic interoperability**: A critical aspect for the evolution of e-infrastructures is the ensuring of interoperability between data collections and services. This presupposes – on the one hand – the usage of consistent and standardized formats for describing data, assets and services and – on the other hand – the meticulous mapping between descriptions schemas and formalizations in order to allow cross-community flexibility without violating the principles of connectivity and collaboration.

- **Seamless service integration**: The plethora of relevant services carrying out similar operations should be made opaquely available to the users of the infrastructure, i.e. the horizontal services should be abstracted via a common interface that will unify the different individual services performing the same operation. As an example, different established authentication and authorization mechanisms will be exposed via a common connection point that will communicate with the actual authorization service based on the user's credentials.
- **Rapid integration of future services**: To ensure the long-term sustainability of the envisioned e-infrastructure, it is important to foresee the emergence of new services and workflows that should be incorporated in the ecosystem. It is, thus, critical to automate the creation of interfacing mechanisms for service integration, as this approach will facilitate backward and forward compatibility of the e-infrastructures and will ensure its synchronization with technological and research progress.

From the technical standpoint, our vision focuses on helping all players in the food system to access information wherever and in whichever way it is stored and exposed, and to process and share it in whichever way they want.

As is evident by the analysis of the current technical landscape in the agri-food sector, while data and knowledge sharing are integral to the realization of the societal, economic and environmental goals, the inherent heterogeneity of the latter poses significant technological challenges.

Beyond discrepancies in data representation and standards, this heterogeneity extends to workflows, methods and practices for gathering, producing, processing and sharing information. Each of the scientific disciplines pertaining to agri-science faces different problems, and gives different solutions tailored to the problems.

Towards this technical vision, it is imperative to foster and facilitate the further adoption of semantics, contextualizing and enriching information using rich conceptualizations, expressed in machine-readable standards. Additionally, we emphasize the need for initiatives towards establishing a common ground for semantic descriptions, via the mapping and linking of the various standards and resources of each community.

Such a strong foundation for semantically enriching and linking information will allow the design and implementation of advanced services for transforming and combining data and methods from different disciplines, opening the ground for interdisciplinary, cross-fertilizing advancements. These new services should be usable and accessible from all different frontend tools, frameworks and services used by researchers and stakeholders, via the appropriate API services.

It's worth mentioning that the fostering and support of semantic interoperability and the provision of relevant, reusable services has an impact reaching far exceeding the scope of the envisioned infrastructure. The existence of robust standards as well as readily available tools will encourage the creation and evolution of services by and for industries and businesses, enabling data and service interoperability for the entire agri-food value chain.

Hence, a crucial part of the proposed abstract architecture is the Specification Layer commandeering the communication of existing, emerging and novel solutions and frameworks by enclosing and aligning all the semantics relevant to the infrastructure's content, tools, and protocols used. The layer, by providing the required descriptive and mapping definitions between the available assets and services, forms the foundation for ensuring interoperability and extensibility across the whole infrastructure. Furthermore, it ensures that the services and resources gradually added to the e-infrastructure are constantly and consistently available to front-end services and applications, e.g. via the usage of a brokering mechanism on top of services and resources that utilises their respective descriptions (Nativi et al., 2012, Zhang et al., 2014).

The proposed architecture entails the technological solutions and resources for encapsulating remote and disparate assets at all levels defined in Figure 5.1. More, specifically:

1. The available physical infrastructure is exposed via a Federation Layer which encapsulates information on the capacity, availability and requirements for all the available resources;
2. The different services integrated in the infrastructure are being made opaquely available through an abstraction layer that hides different implementations and parameters under a common interface, following the agreed semantics for describing the services;
3. The services are subsequently exposed via APIs that are produced by API factories based on the service descriptions;
4. A brokering service constitutes the services and their APIs findable and available to application and services that intend to use the underlying infrastructure.
5. Finally, the horizontal services are combined to form more complex and dedicated community-specific services which are used by the end-users of the infrastructure and the platforms built over it.

These main architecture constituents are the necessary foundation towards establishing a large-scale collaborative research environment based on a FAIR data (and more generally scientific assets) ecosystem for the agricultural domain. Respecting and promoting FAIR principles via the appropriate technical solutions is an essential factor for expanding data analysis to cover the whole spectrum of stakeholders and data providers.

6 RECOMMENDATIONS FOR FUTURE DEVELOPMENT OF AN E-INFRASTRUCTURE FOR OPEN SCIENCE IN AGRI-FOOD SCIENCES

6.1 GENERAL PRINCIPLES AND RECOMMENDATIONS

In the further development of an e-Infrastructure for Open Science in Agriculture and Food Sciences, the following principles need to be adhered to:

1. **Ongoing developments are needed on the short, medium and long term:** there is no quick solution available to achieve Open Science for the Food System momentarily, so a long-term commitment is required. As a low hanging fruit, the existing community concentration for infrastructure development for Open Food System Science needs to be maintained and expanded to be propelled into the future.
2. **The developments in governance and business models have to be based on Stakeholder inclusiveness and international collaboration:** A regular gathering can be envisioned on an international basis for the governance of e-infrastructures in the food system, that is open to actors from academia, policy and private sectors. This could be done in connection to Global Open Data for Agriculture & Nutrition, Research Data Alliance, for example it's IGAD group and CGIAR's Big Data Platform.
3. **The community needs sustained common services which require sustainable funding mechanisms:** these services (for semantics, Fairification, computing) will be tailored to the agriculture and food sciences and need constant funding and cannot be dependent only on time bound projects, which implies that a strong integration in the daily routines of research users. These continued funding streams can be based on private sector engagement or common goods funding, but will require further development.

For Open Science to be successful in agriculture and food sciences, it needs to build on and integrate with the European Open Science cloud, the goFAIR movement, and developments in other communities like marine, environment, biodiversity, etc**.** This roadmap serves as input to the formulation of the next framework program of the European Commission, as well as national agendas of member states, the Big Data platform of CGIAR, and other international funders for alignment.

Main innovations are required around 'Connect, Share and Collaborate' to move towards Open Science for the Food System, and Table below gives lessons learned & priorities across aspects of community, governance & business models, and services & technical backbone.

Table 6: Connection between the elements of an e-infrastructure and the objectives of "Connect", "Share" and "Collaborate"

	Community	Governance & Business Models	Services & Technical Backbone
Connect	• Use cases are useful to identify benefits and gaps, and generic research questions need to be synthesized that are integrating and cross cutting across domains, such that solutions developed will have a broad applicability	• Requirements for data sharing across public and private sectors, while respecting privacy and competitive concerns	• Support the use of standards as much as possible, so that data, information and architectures can easily be connected. • Include models and analytical tools as part of open science next to open data

			• Persistent identifiers should become common practice
Share	• Across use cases, efforts are required in data curation and data rescue in order to make data available in a proper way • Beyond data sharing, Open Science should also work on sharing analytics, models and the collaborative scientific process • We must share our knowledge and offer guidance on tools • We should connect classical science and data science	• To improve trust, we require certification of datasets and repositories • We need more incentives, e.g. rewarding open science, and clear institutional/national policies to support individual actions	• We need to develop smarter interoperability platforms, that are easy to use, not challenging to use • Practical guidance should be developed and offered (e.g. for making data FAIR) and tools should be developed to support this
Collaborate	• In terms of user needs, also include funders • Focus on individual accomplishment of researchers, establishing large and small projects to connect to developments in open science	• Establish eROSA as a shared vision based on users' needs and based on the strengths of partners. Infrastructure development needs to follow the user stories. • Advocate for a user centric approach in the development of the European Open Science Cloud. • Single project funding is not adequate. A European Research Infrastructure Consortium seems an option but requires a long process. • Business models need to be elaborate, taking into account relevant viable assets that can survive on the long term. • Large scale impact studies (for high level briefings) are required, that clearly outline Key Performance Indicators (KPI's) that can be achieved through the adoption of open science • Impact studies on the proof of enabling of innovation through increased sharing in the public, research and private sectors	• Infrastructures need to be as invisible as possible, best conceived as a network of roads, where the change from a provincial road to a highway occurs seamlessly. • Make small nodes per research project, region or domain • Establish a marketplace for working across institutes and sharing resources • Include a certification function, so that nodes are open science proof. • Current solutions are not scalable

6.2 CHALLENGE DRIVEN RECOMMENDATIONS

The general principles and recommendations in the previous section can be translated into challenges that bring together several innovations along the axis of community, governance and business models and services and technical infrastructure:

1. **From Open Science to Open Innovation in the Food System**: This action brings together private and public players in food system to discover how e-infrastructures and open Science approaches can be extended and used for open innovation. This requires new collaboration models and building confidence between mixed consortium of public and private partners, but also training of private researchers to work with open science resources. A critical aspect of the technical infrastructure to enable open innovation is the development of a Specification Layer for the creation and preservation of shared and connected semantics for data, models and services relevant to the sector. The establishment of a “common language” for describing, using and contributing services and resources leads to F.A.I.R. Digital Objects (data, code, workflow etc.), enabling rapid development in open innovation throughout the ecosystem.

2. **Food System e-infrastructures for international development**: Challenges in the food system are crossing continental and national boundaries by definition due to global supply chains, and north south interactions. E-infrastructures for food system research thus represent a common public good that can be shared among continents, to also cross the digital divide from potentially stronger and well-equipped institutes to less well-resourced institutes. International collaboration helps to build international commitment to standards and confirm practices internationally, also linked to important global international organizations as CGIAR, CABI, FAO, IICA, and national research infrastructures in other parts of the world (CAAS, Embrapa, USDA, etc.), next to European developments. Such an international collaboration must be facilitated by the following technical developments: 1. Trust and provenance supporting repositories of digital assets; 2. Robust registration mechanisms for adding such repositories to the ecosystem and 3. Trust-enabled repositories for preserving and exposing metadata specifications, ontologies and vocabularies.

3. **Data-Driven sustainability assessment for SDG-achievement**: There are many SDGs related to agriculture and food, so SDGs’ provide a good framework for food system integration, and bring different communities working on food system together. This requires understanding of the maturity of these communities in providing data driven assessment methodologies for implementing and monitoring the achievement of SDGs. As SDGs monitoring is organised at a national level, cross-boundary learning and implementation of data driven solutions is required to achieve alignment in methodologies. Efficient data driven solutions require standardization of important data sets, making them available as FAIR, elaborating standardized workflows allowing for collaborative inspection and discussions among national stakeholders and collaborative reporting and editing environments.

4. **Open Food System Science for Agri-Environmental and Nutrition policy making**: With Open Food System Science becoming more and more established as a common practice in the sciences, other stakeholders can make use of these open science processes and resulting products (data, algorithms). Evidence based policy making are policy processes that aim to use as much as possible available data and knowledge, often produced by science, for reasons of objectivity and transparency. In many policy cases data and knowledge produced through open science could thus play a more prominent role, but this will require joint learning and development across the policy-science interface, requiring targeting training courses for different stakeholders, next to general awareness campaigns. Also, computational infrastructures targeted at open and joint analysis of

policy cases will need to be developed, with strong capabilities in visualization, scenario analysis, and participative discussion and analysis sessions.

5. **Training the "Food System 4.0" scientists for open innovation towards Food 2030**: With the advance of food system thinking, e-infrastructures and open science, a different type of scientist should emerge. This scientist is comfortable to work across disciplinary boundaries, with advanced ICT and data technologies, in an open collaboration with other societal stakeholders. Scientific work will be open to closer scrutiny, while at the same time have greater demands on reproducibility and transparency. Such scientists need to be trained in such aspects, whether it is from use of e-infrastructures API's for their science, to interaction with the general public through social media, to transdisciplinary interaction with many different types of stakeholders. Open learning materials need to be developed and shared across teaching institutions to rapidly reach a large number of scientists. Developing a culture for smart metadata and semantics among scientists and IT people is one of the keys for intensive data reuse, and new knowledge discovery from data integration and reasoning. This requires not only the technical aspects listed here but also the integration of courses on semantics in training programs for students in agronomy, bioinformatics, computer sciences, data & information management and analytics, and the production of methodological guidelines.

6. **Large Scale Public Private Partnership (PPP) on data-driven innovation and research towards Food 2030**: bringing Open Food System Science to scale, requires collaboration across public and private research partners and developing new models for collaboration towards open innovation. This will enable the step from open science to open innovation, in which results, data and insights from science can be brought to value faster, increasing the competitive advantage of the European and global agri-food data economy. This requires the availability of fair access and fair share of data, combined with the availability at scale of powerful data processing services for advanced analytics & artificial intelligence. Such data and processing services together with analytical environments need to be assessed for their technological maturity for broad scale application. Finally, this will require 1. targeted training modules, and 2. the emergence of data scientists with shared working experience both on the private and public research organizations, who can bridge the organizational gaps. Support has to be given to research on fundamental computational problems resulting from the sheer volume of data and the complexity and multidimensionality of processing and analytics applying to agri-food. The sustainability and expansion of an PPP e-infrastructure strongly correlate to solid research results in fields like big data mining, distributed and large-scale analytics, big data visualisation, the application of machine learning in analytics, etc.

REFERENCES

Allemang, D., & Teegarden, B. (2017). A global data ecosystem for agriculture and food. *F1000Research*, *6*.

Allison, L., Asch, M., Gurney, R., Hodgson, S., Uhle, M., Davis, R., … Watson, K. (2016). Belmont Forum EInfrastructure and Data Management Phase 2: Implementation Plan.

Almas, B., Budich, R., Collins, S., Diepenbrook, M., Dillo, I., Genova, F., … Zwölf, C. M. (2017). Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data, (January). Retrieved from https://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf

Bournigal, J. M., Houllier, F., Lecouvey, P., & Pringuet, P. (2015). # AgricultureInnovation2025: 30 projets pour une agriculture compétitive et respectueuse de l'environnement. *# AgricultureInnovation2025, INRA (2015)*.

Candela, L., Castelli, D., & Pagano, P. (2009, July). D4Science: an e-Infrastructure for Supporting Virtual Research Environments. In *IRCDL* (pp. 166-169).

Carolan, L., Smith, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J., & Gerry, W. (2015). How can we improve agriculture, food and nutrition with open data. *London, UK: Open Data Institute*.

Cinquini, L., Nasa, D., Noaa, Deluca, C., Denvil, S., Duffy, D., … Trenham, C. (2015). ESGF Strategic Roadmap.

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. P. (2007). Report of a workshop on "history & theory of infrastructure: lessons for new scientific cyberinfrastructures". *Understanding Infrastructure: Dynamics, Tensions, and Designs*.

ELIXIR_Scientific_Programme 2014-2018. (n.d.) Retrieved from https://www.elixir-europe.org/sites/default/files/documents/elixir_scientific_programme_final.pdf

ELIXIR Excelerate Technical Services Roadmap 2016. (n.d.).

Evaluation of e-Infrastructures and the development of related Key Performance Indicators. (2017). Retrieved from http://e-irg.eu

e-IRG. (2014). e-irg guidelines document: Best Practices for the use of e-Infrastructures by large-sca. Retrieved from http://e-irg.eu

Forum, B. (2015). A Place to Stand: e - Infrastructure s and Data Management for Global Change Research. Belmont Forum e - Infrastructures & Data Management Community Strategy and Implementation Plan. Retrieved from https://www.belmontforum.org/wp-content/uploads/2017/05/A_Place_to_Stand-Belmont_Forum_E Infrastructures__Data_Management_CSIP.pdf

Goldstein, S. (2017). The evolving landscape of Federated Research Data Infrastructures: Final report on the situation in the six Knowledge Exchange partner countries, 39. https://doi.org/10.5281/zenodo.1064730 Grimm, C. (2018). European Data Infrastructure Initial Perspective from GÉANT and PRACE Executive Summary.

Health_RI_vision_Feb-2016_NL. (n.d.).

Holzworth, D. P., Snow, V., Janssen, S., Athanasiadis, I. N., Donatelli, M., Hoogenboom, G., ... & Thorburn, P. (2015). Agricultural production systems modelling and software: current status and future prospects. *Environmental Modelling & Software*, *72*, 276-286.

Janssen, S. J., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W., & Antle, J. M. (2017). Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. *Agricultural systems*, *155*, 200-212.

Jean-Marc Bournigal, M. (n.d.). Octobre 2016 AgGate Portail de données pour l'innovation en agriculture.

Retrieved from http://www.ladocumentationfrancaise.fr/rapports-publics/174000039/index.shtml

Jeffery, K. G., Meghini, C., Concordia, C., Patkos, T., Brasse, V., Ossenbruck, J. V., ... & Marchetti, E. (2017). A Reference Architecture forVirtual Research Environments.

Morrell, S. (2014). e-Infrastructure Roadmap. *Rcuk*, 1–15. Retrieved from http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RoadmapforELC.pdf

Nativi, S., Craglia, M., & Pearlman, J. (2012). The brokering approach for multidisciplinary interoperability: A position paper. IJSDIR, 7, 1-15.

Peters, D. P., Havstad, K. M., Cushing, J., Tweedie, C., Fuentes, O., & Villanueva-Rosales, N. (2014). Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere*, *5*(6), 1-15.

RDA EU call for action: European Data Economy must focus on interoperability. (2018). Retrieved from https://www.rd-alliance.org/rda-eu-call-action-european-data-economy-must-focus-interoperability-rda-europe-statement-european

Scientific and Technical description of LifeWatch ERIC. (2015). Retrieved from https://www.lifewatch.eu/documents/20181/21696/Scientific+and+Technical+description+of+LifeWatch+ERIC/212169f2-3b02-4357-8272-66b55c978fdc

Siné, M., (2016), Techniques, I. (n.d.). L'accès aux données pour la Recherche et l'Innovation en Agriculture Position des Instituts Techniques Agricoles.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., ... & Folke, C. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science*, *347*(6223), 1259855.

Wittenburg, P., Computing & Strawn, G. (2018). Common Patterns in Revolutionary Infrastructures and Data, (February), 1–13.

Zhang, L., Fowley, F., & Pahl, C. (2014). A template description framework for services as a utility for cloud brokerage.

ANNEX 1 - EXAMPLES OF EXISTING INITIATIVES WITHIN THE AGRI-FOOD COMMUNITY

CIARD RING *(http://ring.ciard.net/)*

The CIARD RING (shortened as the RING) is a global cataloguеue of web-based data services and datasets for food and agriculture. It was created by several partners in the former CIARD initiative (primarily the Global Forum on Agricultural Research and Innovation, the Food and Agriculture Organization of the United Nations, the Technical Center for Agricultural and Rural Cooperation, Cornell University, the Chinese Academy of Agricultural Sciences) and is now endorsed by the Global Open Data for Agriculture and Nutrition initiative and maintained by GFAR.

The RING has multiple functions: a) it includes both web-based data services and directly accessible datasets; b) it acts both as a primary catalogue (providers can catalogue individual datasets directly in the RING) and as a federated catalogue (it harvests dataset metadata from other catalogues); c) it provides an end-user search interface and a machine-readable layer.

The datasets are described in detail and categorized according to both content criteria such as thematic coverage, geographic coverage, type of data to which they apply, target audience; and technical criteria such as file format, data format (including specifications, schemas, templates…), data dictionaries and controlled vocabularies used, retrieval protocols etc.

The RING is meant to provide high quality metadata about datasets that can be reused by other more specialized portals or applications, which means that applications have to be able to read all the RING metadata and filter datasets according to their needs. To achieve this, the RING exposes all metadata as an RDF store and through a SPARQL engine, using several existing RDF vocabularies for datasets (DCAT, DataCube, VOID) complemented with some necessary vocabulary extensions. Besides this, it adopts a full Linked Data approach in that all entities have URIs and wherever possible are linked to external URIs in other semantic systems. An example of this are all the categorizations used in the RING to index datasets: for these, the RING hosts local concept schemes, but whenever possible individual concepts are linked to external concepts in some other semantic system (e.g. AGROVOC, Geopolitical Ontology, W3C formats), even though there are no existing LOD-enabled authorities for many of the necessary classifications (type of data covered, data format, file format, protocols etc.).

The RING is FAIR: it is Findable (it's registered in DataHub and has an identifier) and it makes datasets Findable (they are registered in a catalogue and are identified by URIs); it's Accessible (metadata are always accessible through the REST protocol and more precisely through the SPARQL protocol); it's Interoperable (see the description of the Linked Data layer above) and it's Reusable (RING metadata has a license and dataset metadata includes the original dataset license information). There are currently 3201 datasets and 2224 data services registered in the RING.

AgroLD *(http://www.agrold.org)*

AgroLD (Agronomic Linked Data) is a knowledge-based system that exploits the Semantic Web technology and some of the relevant standard domain ontologies, to integrate genome to phenome information on plant species widely studied by the agronomic research community. The current phase of the project covers information on genes, proteins, ontology associations, homology predictions, metabolic pathways and plant traits relevant to the selected species. Currently, AgroLD contains hundreds of millions of triples created by annotating more than 50 datasets coming from 10 data sources such as Gramene.org and

TropGeneDB with 10 ontologies such as Gene Ontology and Plant Trait Ontology. The objective of this effort is to provide the community with a platform for domain specific knowledge, capable of answering complex biological questions such as "find genes/proteins involved in plant disease resistance and high grain yield traits." and in this way facilitating the formulation of new scientific hypotheses.

BrAPI *(www.brapi.org)*

The Breeding API (BrAPI) project has been initiated in 2016 with the help of the Bill & Melinda Gates Foundation and major contributions from the CGIAR, Wageningen University, the French National Institute for Agricultural Research (INRA) and the James Hutton Institute. It targets all data related to plant breeding: genotyping and genetic variability, field and greenhouse phenotyping experiment, plant material traceability and pedigree. It is led by a full-time coordinator funded by the Excellence in Breeding platform of the CGIAR. It is built through a highly collaborative approach based on workshops or hackathon organized twice a year. Thanks to this open approach it has been possible to improve it and to enable a very wide adoption. BrAPI is now a de facto standard for plant phenotyping and serves as a backend for major services like the Elixir Plant Data Lookup service under development.

BrAPI defines web services specification to enable interoperability across tools and databases. Several use cases are already implemented in major systems, among which genetic variation visualization with Flapjack, databases to analysis environment connection, FAIR compliant data discovery, phenotype to genotype linking for genomic selection.

The genetic visualisation was led by the James Hutton Institute based on Germinate and Flapjack. It allows a chromosome wide visualization of nucleic variation among several plant varieties. Thanks to BrAPI, it is possible for Flapjack to pull data from any BrAPI compliant database. This work is being extended to include the traceability of genotyping samples stored in GOBII, a CGIAR database dedicated to genomic and breeding.

BrAPI also allows to connect any database to an R analysis environment, either in command line or through a web interface such as Shiny.

Phenotyping data discovery strongly relies on distributed indexing. Indeed, there is no plan to create a central phenotype file, so it is able to have an easy and accurate view of the multi-file data content and easily grab this data for analysis is vital to avoiding the data silos. INRA and the Elixir plant community thus engaged to build a Findable, Accessible, Reusable, Interoperable (FAIR) compliant system by doing major contributions to the BrAPI and building all the necessary softwares to index and search any BrAPI compliant database.

The adoption and extension of BrAPI for other European infrastructures such as Emphasis is currently very seriously evaluated. This, in addition to its current level of adoption by major actors of the plant science community, makes BrAPI a very strong standard.

Wheat Data Interoperability guidelines *(http://datastandards.wheatis.org)*

The wheat research community, along with data and ontology experts, joined their efforts through the Wheat Data Interoperability Working Group28 to develop the Wheat Data Interoperability Guidelines (WDI-WG)29. The guidelines consist of a set of recommendations for describing and representing data

[28] https://www.rd-alliance.org/groups/wheat-data-interoperability-wg

[29] https://f1000research.com/articles/6-1843/v2

with respect to existing open data standards. From the outset, the objective of the WDI-WG was to deter communities from creating new standards, which would have made the already-complex landscape of existing data standards even more complex. The WDI-WG collected valuable information through two surveys of the wheat research community, comprising responses regarding existing data formats, practices, and the use of ontologies and controlled vocabularies. From these surveys, the WDI-WG then developed a set of specific recommendations, and worked to facilitate data interoperability through the harmonization of data formats, data models and vocabularies usage, thus aiming to address the main interoperability issues. The proposed recommendations have been endorsed by the International Wheat Initiative WheatIS Expert Working Group (see below) and the Technical Advisory Board of the Research Data Alliance[30].

Agrisemantics Map of data standards *(http://vest.agrisemantics.org)*

The Agrisemantics Map of data standards is a catalogue of data standards for the agri-food sector, created under the GODAN Action project and overseen by several organizations (Global Forum on Agricultural Research and Innovation, Food and Agriculture Organization of the United Nations, Open Data Institute, French National Institute for Agricultural Research, Wageningen Environmental Research Alterra, Global Open Data for Agriculture and Nutrition and others) committed to maintain it in the future.

The catalogue builds on two existing efforts in the domain, the VEST Registry of FAO (now superseded by the Map) and the AgroPortal repository of RDF vocabularies, maintained by the University of Montpellier. The Map is synchronized and linked with the AgroPortal, but also allows to manually register and curate records. Therefore, it is both a manually curated catalogue and a federated catalogue.

For the moment, records are federated only from the AgroPortal. Ideally in the future, the catalogue should also federate additional relevant (domain-specific) records from other existing general catalogues of data standards or vocabularies (e.g. FAIRsharing, Linked Open Vocabularies or the Basel BARTOC directory).

On the other hand, there are plans to share the metadata records of the Map with other catalogues, primarily FAIRsharing. In any case, beyond bilateral exchange agreements, records from the Map are available for any other platform through the Linked Data layer of the portal: the RDF model of the Map is based on several existing vocabularies to describe vocabularies and all classifications used are published as SKOS with as many terms as possible linked to existing authoritative lists.

The Map is also synchronized with the CIARD RING catalogue of datasets: datasets in the RING are associated with the data standards applied and the list of standards comes from the Map; conversely, for each data standard in the Map, datasets using that standard are retrieved from the RING.

Within the scope of the Map, “data standards” are intended as “vocabularies” in the broad sense in which vocabularies are defined by W3C (“vocabularies define the concepts and relationships used to describe and represent an area of concern”). They range from description / modelling standards (XML schemas, RDFS schemas, ontologies, application profiles, even UML models) to knowledge organization systems of different types (classifications, thesauri, code lists or even certain types of ISO standards as controlled lists of values).

Regarding the domains covered, since the sub-domains of food and agriculture span across several disciplines (plant sciences, farming systems, natural resources management, forestry, all disciplines

[30] https://rd-alliance.org

involved in the food supply chain, etc.) but are also closely interlinked with neighbouring disciplines (climate, environment, geospatial, biology, etc.), the Map includes standards covering all of these disciplines, categorized against a classification of types of data agreed among different partners.

The Global Agricultural Concept Space (GACS) ***(http://agrisemantics.org/node/8)***

The GACS is a set of reference concepts used as a hub to interconnect semantic resources and datasets. It provides the agricultural stakeholders (research, agencies, companies, etc.) with shared, persistent identifiers and the maintenance of those common concepts. As one of the pillars of Agrisemantics (with AgroPortal and the Map of Data Standards), the GACS aims to leverage the building of a web of interoperable knowledge resources and datasets for agriculture and nutrition.

The GACS addresses two main needs of the community: 1) have a set of commonly-used and persistent concepts (namely URIs) for mapping application and treasure ontologies This reduces the burden of creating many one-to-one alignments between them. It also ensures the long-term maintenance of mappings. 2) have easily actionable lists of terms, organised in a logical way, and offering labels in English and most often in several other languages. Semantic resources developers can rely on such lists when they develop new thesauri and ontologies. To ease their work, GACS concepts are presented and made accessible through complementary views (hierarchy, groups, thematic etc.).

The GACS is based on and serves the semantic web and linked open data approaches and technologies. It is implemented in SKOS, which makes it usable by many stakeholders and actionable in a variety of tools. It is accessible through a browsable web interface and can be downloaded as an open dataset for reuse.
The GACS was created and is supported by several major organizations (Food and Agriculture Organization of the United Nations, the National Agricultural Library of US Department of Agriculture, Centre for Agriculture and Biosciences International, French National Institute for Agricultural Research, University of Tor Vergata, and the LIRMM laboratory of University of Montpellier) committed to maintain it in the future.

AgroPortal ***(http://agroportal.lirmm.fr)***

By reusing the National Center for Biomedical Ontologies (NCBO) BioPortal technology, Univ. of Montpelier (LIRMM) in collaboration with multiple research organizations in agronomy and Stanford, have designed AgroPortal, an ontology repository for the agronomy domain that seeks to provide a one-stop-shop for ontologies & vocabularies in this field, enabling to identify and select an ontology for a specific task as well as offering generic services to exploit them in search, annotation or other scientific data management processes.

The AgroPortal project was originally driven by five use cases which participated in the design and orientation of the project to anchor it in the community and were the original sources of ontologies:

- The Agronomic Linked Data (AgroLD - see above)
- The RDA Wheat Data Interoperability (WDI see below)
- The INRA Linked Open Vocabularies (LovInra), which is an effort to publish vocabularies produced or co-produced by INRA scientists and foster their reuse beyond the original researchers (http://lovinra.inra.fr)
- The Crop Ontology project (see below)
- The Agrisemantics Map of Data Standards (see above)

In addition of these five first driving use cases, other projects or organizations have identified AgroPortal as a relevant application to host, share and serve their ontologies both at the national or international level. For instance, EU projects such as the AnaEE infrastructure, new INRA research units (URGI, MaIAGE, IATE, URFM, MISTEA), IRSTEA projects in agriculture. Or more recently, the new Global Agricultural Concept Space project (GACS - http://www.agrisemantics.org/gacs) inspired from the GACS initiative which resulted in the integration of the Agrovoc thesaurus (FAO), the NAL Thesaurus and the CAB Thesaurus.

AgroPortal is now an advanced prototype featuring ontology hosting, search, versioning, visualization, comment and recommendation; it enables semantic annotation and supports basic alignment features. It is semantic web compliant with all the previously mentioned features available through two endpoints that allow automatic search of the content of the portal: (i) a REST web service API; and (ii) a SPARQL endpoint. The AgroPortal specifically satisfies requirements of the agronomy community in terms of ontology formats (e.g., SKOS vocabularies and trait dictionaries) and supported features (offering detailed metadata and advanced annotation capabilities). The platform currently hosts 90 vocabularies with more than 2/3 of them not present on any similar ontology repository and 11 private ontologies. 80 additional candidate ontologies will be loaded in the future to complement this valuable resource. The platform already has more than 95 registered users and some vocabularies are visited more than 100 times per month. In order to scale up the use and development of AgroPortal, the community needs to develop new semantic resources for semantic-search and retrieval of data, and encourage the agronomy/agriculture communities to embrace the semantic web standards when structuring their knowledge (e.g., phenotypic or functional traits, plant diseases): i.e. SKOS to formalize interoperable vocabularies and thesauri and OWL to develop formal ontologies. We need to support ontology developers and a broader community of experts in producing, releasing, sharing, serving and interlinking their semantic resources. For this, we need to transform the AgroPortal prototype into a widely adopted, long-term supported, robust and curated ontology repository.

Crop Ontology *(http://www.cropontology.org)*

The **Crop Ontology (CO)** is a community-based ontology that to date is an open access collection of 26 species-specific trait ontologies, one for an alga of economic interest and one for woody species. Each concept is fully defined. It proposes the Multicrop Passport Data and a germplasm ontology. Often, in excel files resulting from field observations; the headers of columns where the value of the observation is stored are abbreviated names of the variable decided by the scientist for personal use. Unfortunately, it is difficult to understand from the abbreviated names “what is measured?” “how?” and “using what unit or scale?”, particularly if the protocol or a clear list of variable names with definitions, methods of measurement are not included. This limits the re-use of the datasets by other breeders, for statistical analysis or crop modelling. To help addressing this problem, the Crop Ontology provides community-validated and pre-composed variables useful to guide the creation of fieldbooks and the storage of measurements in databases. A **‘variable'** in the Crop Ontology is equal to '***one trait+one method+one unit or scale***'. As a trait (*plant part+quality*) can be measured through different variables, according to the method or the scale selected by the scientist, CO offers the flexibility to create variables. Such pre-composed variables will accurately and consistently annotate the measurements made in multi-site evaluation programmes. CO is a tool for increasing the quality of data and metadata, their interoperability and the exchange of trait data between crop models. The standard Trait Dictionaries (TDs) are a "must have" for breeding platforms and fieldbooks such as the Integrated Breeding Platform (IBP) or the Next Generation Breeding Databases developed by Boyce Thompson Institute for the CGIAR Roots, Tubers and Bananas Research Programme[31]. Several crop groups interested to develop their species ontology for a project apply the Crop Ontology workflow supported by its guidelines. The Crop Ontology Community

[31] http://www.rtb.cgiar.org/

stimulates and mediates the discussions between breeders, data managers, geneticists, crop modelers to obtain adequate variables and traits description.

The Crop Ontology traits are mapped to the reference species-neutral ontologies of Planteome to provide the semantic links that will support trait data retrieval across species.

The open standard Breeding Application Programme Interface (BrAPI - see above)_has integrated the variable definition and structure for its Observation Variable call. The Crop Ontology Trait Dictionary Template is recommended for phenotypic data annotation by projects like the European Elixir-Excelerate project that developed the **Minimum Information about Plant Phenotyping Experiment[32] (MIAPPE**) and adopt the standard variable structure.

The Crop Ontology is available in the Ontology Look up Service of the European Bioinformatics Institute to support data annotation and in the Agroportal (see above).

The crop Ontology web site is hosted on Google App Engine[33]. The code is being versioned and hosted on GitHub[34].

GODAN *(godan.info)*

Global Open Data for Agriculture and Nutrition (GODAN) is an initiative that seeks to support global efforts to make agricultural and nutritionally relevant data available, accessible, and usable for unrestricted use worldwide. The initiative focuses on building high-level policy as well as public and private institutional support for open data. GODAN encourages collaboration and cooperation among existing agriculture and open data activities, without duplication, and brings together all stakeholders to solve long-standing global problems.

The initiative was launched in 2013, one year after the G8 summit in 2012 where G-8 leaders committed to the New Alliance for Food Security and Nutrition as the next phase of a shared commitment to achieving global food security.

Open access to research, and open publication of data, are vital resources for food security and nutrition, driven by farmers, farmer organizations, researchers, extension experts, policy makers, governments, and other private sector and civil society stakeholders participating in 'innovation systems' and along value chains. According to the Open Data Institute[35], farmers and other stakeholders on the agriculture supply chain can make more informed decisions resulting in improved yields and efficiency – from farm to fork, when they have free access to useful information on agriculture and nutrition.

GODAN and its partners aim to support the open data revolution and hosted the 2016 GODAN summit in New York in September. GODAN has 656 partners from government, international and private organisations around the world. The initiative welcomes all those who share this purpose to join[36] as members and to participate in shaping coordinated activities that can deliver on the potential of open data for agriculture and nutrition.

[32] https://sites.google.com/a/cgxchange.org/cropontologycommunity/home/goog_1591953685

[33] http://www.cropontology-curationtool.org/

[34] https://github.com/bioversity/Crop-Ontology

[35] https://theodi.org/about-the-odi/

[36] http://www.godan.info/partners

The RDA Interest Group for Agricultural Data *(https://www.rd-alliance.org/groups/agriculture-data-interest-group-igad.html)*

Formed in 2013, the RDA Interest Group for Agricultural Data (RDA-IGAD) is to date the most RDA's prominent domain-oriented interest group (over 170 members), working on all issues related to global agriculture data. It represents stakeholders in managing data for agricultural research and innovation, including producing, aggregating and consuming data. RDA-IGAD is also a forum for sharing experiences and providing visibility to research and work in agricultural data.

RDA-IGAD aggregates many working groups (WGs) with their own tangible goals. To date, these are the IGAD-associated WGs:

- The Wheat Data Interoperability WG created in 2014 to build a common interoperability framework for wheat-related data providing guidelines for describing, representing and linking wheat-related data.
- The Rice Data Interoperability WG created in 2017 to build a common interoperability framework for rice-data providing guidelines for describing, representing and linking rice-related data.
- The Agrisemantics WG created in 2016, to gather community-based requirements and use cases for an infrastructure that supports appropriate use of semantics for data interoperability in agriculture.
- The On-Farm Data Sharing WG created in 2017, to develop farmer networks guidelines for data handling, sharing and use.
- The Metrics and Indicators in Agricultural Science WG created in 2017, to create a dialogue between the agricultural research community and metrics/indicators experts to prepare a whitepaper on the use of citation-based metrics, social media/altmetrics, and societal impact in agricultural sciences.
- The Soil Data WG created in 2017, to develop a concept note for an independent technical evaluation of the current soil data exchange models usually referred to as SoilML (ISO 28258, OGC IE, INSPIRE Soil).
- The Capacity Development WG created in 2018 to support the adoption of recommendations and outputs of the WGs under the IGAD umbrella.

AIMS *(aims.fao.org)*

AIMS supports good practices, open standards and technology for open access, open data and open science in the agricultural field. Within the framework of Access to Information and Semantics services offered by the community, AIMS:

- **raises the awareness** of different possibilities to access, retrieve and use information on agriculture, environment, forestry and other related sciences;
- **enhances the capacities** of different parties worldwide who are eager to learn and progress in their profession as much as they can/would like by accessing the outstanding research resources, managing their (open, research) data and knowledge. To this end, AIMS provides:
    - Free Access to Scientific Information Resources in Agriculture (ASIRA)[37] Online Course for Low-Income Countries;
    - Free AGORA: Fundamentals of Information Literacy and Access to Global Online Research in Agriculture[38] Online Course (in English, Spanish and French);

[37] http://aims.fao.org/asira

[38] http://aims.fao.org/online-courses

- Free Spanish Training Courses under the title: Introduction to Research Data Management[39];
- Free Open Data Management in Agriculture and Nutrition online course (see below), in collaboration with GODAN Action[40].

The AIMS Portal also constantly delivers updates on how to access, contribute to and make better use of:

- AGROVOC multilingual Thesaurus[41] and
- AGRIS : the International System for Agricultural Science and Technology[42].

AIMS actively encourages (agricultural, research) communities to practice their data openly, ethically, and according to the FAIR (Findable, Accessible, Interoperable and Re-usable data)[43] principles, which have been actively promoted by different Research Data Alliance - Agricultural Data Interest Group (IGAD - see above) initiatives.

Additionally, AIMS collaborates with the e-Agriculture Community of Practice[44], to facilitate the exchange of information on initiatives deploying ICT in agriculture in order to achieve UN Sustainable Development Goals (SDGs)[45].

The Wheat Initiative *(http://www.wheatinitiative.org/)*

Created in 2011 following endorsement from the G20 Agriculture Ministries, the Wheat Initiative provides a framework to establish strategic research and organisation priorities for wheat research at the international level in both developed and developing countries. The Wheat Initiative fosters communication between the research community, funders and global policy makers, and aims at securing efficient and long-term investments to meet wheat research and development goals. It also initiates and supports activities in order to enhance communication and increase access of all to information, resources and technologies.

The Wheat Initiative actions will lead to the creation of improved wheat varieties and to the dissemination of better agronomic practices worldwide. The combination of new varieties and agronomic practices will in turn enable farmers to improve and stabilize wheat yields in diverse production environments.
One of the Wheat Initiative's Expert Groups focuses on the development of a Wheat Information System (Wheat IS)[46]. The main objective is to provide a single-access web-based system to access the available data resources and bioinformatics tools in order to support the wheat research community.

The CGIAR Big Data Platform *(http://bigdata.cgiar.org/)*

The data revolution offers an unprecedented opportunity to find new ways of reducing hunger and poverty, by encompassing data-driven solutions to ongoing research for development efforts and in so doing, bringing transformational changes by introducing positive disruptive technologies that radically

[39] http://aims.fao.org/spanish-course-research-data-management

[40] http://www.godan.info/godan-action

[41] http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

[42] http://aims.fao.org/agris-network

[43] https://www.force11.org/group/fairgroup/fairprinciples

[44] http://e-agriculture.org/e-agriculture

[45] http://www.fao.org/sustainable-development-goals/en/

[46] http://wheatis.org/

change impact pathways. CGIAR's Big Data Platform for Agriculture envisions generating impactful big data and ICT innovations that revolutionize farming in developing countries. CGIAR is uniquely positioned to be a thought leader on the use of big data and information technology in agriculture to drive equitable rural development. In so doing, CGIAR works with partners to help ensure that the data revolution in agriculture is democratic, and reaches the marginalized rural poor. The Big Data Platform therefore focuses on enhancing CGIAR and partner capacity to deliver big data management, analytics and ICT-focused solutions to CGIAR target geographies and communities. The ultimate goal of the platform is to harness the capabilities of Big Data to accelerate and enhance the impact of international agricultural research. The Platform is beginning to achieve this through ambitious partnerships with initiatives and organizations outside CGIAR, both upstream and downstream, public and private. It focuses on promoting CGIAR-wide collaboration across CRPs and Centers, in addition to developing new partnership models with big data leaders at the global level. As a result, CGIAR and partner capacity will be enhanced, external partnerships will be leveraged, and an institutional culture of collaborative data management and analytics will be established. Important international public goods such as new global and regional datasets will be developed, alongside new methods that support CGIAR and others to use the data revolution as an additional means of delivering on its promise, and addressing hunger and rural poverty.
The Platform has three objectives, sequenced purposefully. As a first and necessary step, CGIAR is working to put its own house in order, with respect to data. A significant step forward is required that values data as a product in itself with global public good potential, and uses the best available means of managing it and making it available to the public following FAIR principles. Objective 1 ("Organize") focuses precisely on the infrastructure, tools and data culture to succeed, in both technical and managerial dimensions. A key early output has been the ability to discover data and publications from across CGIAR (and soon, other entities) via the Global Agricultural Research Data Innovation and Acceleration Network, or GARDIAN[47]. Objective 2 ("Convene") enables the external partnerships needed to deliver on Objective 1, and aims to enable CGIAR and other agricultural research for development entities to make significant advances in their capacity to manage, use and analyze data. New partnership models are being developed with upstream and downstream partners, from public and private sectors, to deepen and widen capacity on big data analytics and use. Finally, Objective 3 ("Inspire") is designed to inspire through the operationalization of CGIAR and partners capacity to innovate around big data. It aims to tackle some of development's most complex problems with new data-driven approaches, and deliver scalable pilots of effective big data for development solutions.

Syngenta - The Good Growth Plan - One planet. Six commitments towards more sustainable agriculture *(https://www.syngenta.com/what-we-do/the-good-growth-plan)*

To feed a fast-growing world population, humanity is dependent on farmers to increase their production in a substantial, fast and sustainable way. The Good Growth Plan informs how our products and services contribute to a more efficient and sustainable farming system. And it makes good business sense – for our customers as well as for us. We're collecting hard evidence to guide further progress and stimulate take-up of best practice. In 2017, we measured big steps forward in crop yields, in efficient use of pesticides, and in reducing carbon footprints.

The Good Growth Plan is central to our strategies for both our Crop Protection and Seeds businesses to ensure their success and long-term viability. It defines six commitments in areas that are material to our business, where improvement is essential to secure the future of agriculture and our world. Each commitment sets hard, stretch targets to be achieved by 2020. We are measuring and reporting our progress against these targets each year, and providing additional progress information online at www.data.syngenta.com. The Plan's principles and priorities are now deeply embedded in the way we do

[47] http://gardian.bigdata.cgiar.org/

business. As it has continued, we have begun to assess not only our progress but also the nature and quality of the value we are adding: the impact on people, communities and the environment. As we build what we learn into our commercial offer, we are also compiling the evidence that it delivers real, measurable value for growers and society at large. The data and insights that we are gaining and sharing have provided the basis for a growing number of partnerships – with governments, academia, NGOs and businesses. These add further value to our efforts and guide the continuing evolution and development of The Good Growth Plan itself.

Supporting the UN Sustainable Development Goals - The SDGs underscore the relevance and significance of our Good Growth Plan. Collectively, the Plan's six commitments contribute towards delivering the SDGs: all six commitments contribute directly to Goal 2 (zero hunger) and Goal 17 (partnerships for sustainability), as well as individually towards a number of other goals.

Our six commitments help farmers meet the challenge of feeding a fast-growing world population sustainably:

- Make crops more efficient - Increase the average productivity of the world's major crops by 20 percent without using more land, water or inputs.
- Rescue more farmland - Improve the fertility of 10 million hectares of farmland on the brink of degradation.
- Help biodiversity flourish - Enhance biodiversity on 5 million hectares of farmland.
- Empower smallholders - Reach 20 million smallholders and enable them to increase productivity by 50 percent.
- Help people stay safe - Train 20 million farm workers on occupational safety, especially in developing countries.
- Look after every worker - Strive for fair occupational conditions throughout our entire supply chain network.[48]

See www.data.syngenta.com and http://www.goodgrowthplan.com/ for more information.

AgGateway *(www.aggateway.org)*

AgGateway is a non-profit consortium of over 200 businesses in the agricultural industry. Its mission is to promote and enable the industry's transition to digital agriculture, and to expand the use of information to maximize efficiency and productivity. It provides a non-competitive collaborative environment, transparent funding and governance models, and anti-trust and intellectual property policies that guide and protect members' contributions and implementations.

AgGateway primarily focuses on implementing existing standards and collaborating with other organizations to extend them when necessary. For the past several years, groups within AgGateway have been addressing interoperability problems in agricultural field operations (such as planting, crop nutrition, irrigation, crop protection, harvest, and post-harvest) and the farmers' business processes those operations are part of, in addition to a decade-long effort to do the same in agricultural supply-chain management. Standards of interest include ISO 11783 used for communication with agricultural machinery, ISO 19156 (Observations and Measurements), AgXML (used for representing grain movement and commercialization operations), and the AgGateway Open Standards (pertaining to supply-chain operations involving manufacturers, distributors, retailers and growers).

[48] https://www.syngenta.com/~/media/Files/S/Syngenta/2018/Progres-Report-2017-EN-The-Good-Growth-Plan.pdf

AgGateway is also investing heavily into developing a distributed system of unique identifiers or "reference data", for the resources used in agriculture, ranging from crop protection products to farmers' machinery, farms and fields, to the properties observed in those fields. This happens in response to the understanding that a lack of common data formats in agriculture (currently being addressed as an open-source project by AgGateway's ADAPT team, www.adaptframework.org) is only part of the interoperability problem in agriculture. The lack of common meanings throughout the industry is at least as important; to that effect, AgGateway is developing semantic resources such as controlled vocabularies, and a distributed system of APIs to distribute reference data and variable definitions.

The AgGateway concept has expanded internationally: the AgGateway Global Network (http://aggatewayglobal.net) has been created as a coordinating organization, and regional AgGateway organizations currently exist, or are being created in Europe, South America, and Japan.

AgMIP (*http://www.agmip.org/*)

The Agricultural Model Intercomparison and Improvement Project (AgMIP) is a major international effort linking the climate, crop, and economic modelling communities with cutting-edge information technology to produce improved crop and economic models and the next generation of climate impact projections for the agricultural sector.

The worldwide agricultural sector faces the significant challenge of increasing production to provide food security for a population projected to rise to 9 billion by mid-century while protecting the environment and the functioning of ecosystems. This challenge is compounded by the need to adapt to climate change by taking advantage of potential benefits and by minimizing the potentially negative impacts to agricultural production. The Agricultural Model Intercomparison and Improvement Project (AgMIP) seeks to improve substantially the characterization of world food security under climate change and to enhance adaptation capacity in both developing and developed countries.

Analyses of the agricultural impacts of climate variability and change require a transdisciplinary effort to consistently link state-of-the-art climate scenarios to crop and economic models. Crop model outputs are aggregated as inputs to regional and global economic models to determine regional vulnerabilities, changes in comparative advantage, price effects, and potential adaptation strategies in the agricultural sector. Climate, crop model, economics, and information technology protocols are presented to guide coordinated AgMIP research activities around the world, along with cross-cutting themes that address aggregation, uncertainty, and the development of Representative Agricultural Pathways (RAPs) to enable testing of climate change adaptations in the context of other global trends. The organization of research activities by geographic region and specific crops is described, along with project milestones.

AgMIP aims to utilize intercomparisons of these various types of methods to improve crop and economic models and ensemble projections and to produce enhanced assessments by the crop and economic modelling communities researching climate change agricultural impacts and adaptation.

MACSUR (*https://macsur.eu/*)

MACSUR is a knowledge hub within FACCE-JPI (Joint Programming Initiative for Agriculture, Climate Change, and Food Security)[49]. The FACCE-JPI Scientific Research Agenda[50] defines five core research themes to address the impacts of climate change on European agriculture. MACSUR gathers the excellence of existing research in livestock, crop, and trade science to describe how climate variability and

[49] https://www.faccejpi.com/

[50] https://www.faccejpi.com/Document-library/Scientific-Research-Agenda

change will affect regional farming systems and food production in Europe in the near and the far future and the associated risks and opportunities for European food security. A knowledge hub is an innovative, tailor-made instrument developed by FACCE-JPI, associating 3 complementary dimensions: networking, research and capacity building.

The knowledge hub consists of 70 institutions from 18 countries. The overarching challenge is to develop a pan-European capability in the development, use and interpretation of models to perform risk assessments of the impacts of climate change on European agriculture. The project focuses on the technical and informational integration of suitable existing models and their application in regional case studies that reflect the European diversity in soil, climate, socio-economy and agricultural systems.

To address this the following challenges must be met:

1. Identify and address a range of issues between models in different themes to enable their closer integration including issues of scale and data processing.
2. Train a new generation of scientists to work across models which contribute to greater integration of models. This challenge can be described as focusing on the creation of integrated modellers as opposed to integrated models.
3. Determine the contribution that can be made to reducing uncertainty over the impacts of climate change on European food security by adopting integrated models of crop production, animal production and trade.

The work in MACSUR is organized in three "Themes" on the modelling of crops, livestock (including permanent grasslands and farms) and the socio-economy. In addition, there are cross-cutting activities for integrating knowledge across disciplines.
MACSUR organizes workshops and major international conferences for knowledge exchange among experts.

MACSUR co-operates closely with other international research networks like AgMIP (see above) and interacts with political stakeholders. Expected outcomes of the project is a procedure for integrating models, assessment of climate change impacts on agricultural production in European case studies, and an assessment of how uncertainty in modelling climate change impacts on European agriculture could be reduced further.

AGINFRA+ Project *(http://www.plus.aginfra.eu/)*

AGINFRA+ aspires to provide a sustainable channel addressing adjacent but not fully connected user communities around Food & Agriculture, by exploiting core European e-infrastructures such as EGI.eu[51], OpenAIRE[52], EUDAT[53] and D4Science[54]. To this end, the project develops, extends and provides the necessary specifications and components for allowing the rapid and intuitive development of variegating data analysis workflows, where the functionalities for data storage and indexing, algorithm execution, results visualization and deployment are provided by specialized services utilizing European large-scale, cloud-based infrastructural assets.

[51] https://www.egi.eu/

[52] https://www.openaire.eu/

[53] http://eudat.eu/

[54] https://www.d4science.org/

Furthermore, AGINFRA+ aspires to establish a framework facilitating the transparent documentation, exploitation and publication of research assets (datasets, mathematical models, software components results and publications), in order to enable their reuse and repurposing from the wider research community. Thus, the vision of AGINFRA+ project is to develop a common technical infrastructure that could initially serve three user communities (namely, Agro-climatic and Economic Modelling, Food Safety & Risk Assessment and Food Security) and it could be evolved to an AGINFRA food cloud demonstrator that will be positioned as the European Open Science Cloud (EOSC)[55] agri-food thematic cloud.

Datapartage.inra.fr *(www.datapartage.inra.fr)*

Datapartage.inra.fr is a website dedicated to (i) communicate INRA's data management and sharing policy and (ii) provide the scientific community with access to data management and sharing related tools and services. Examples of tools and services include:

- Access to reference documents
- Guidance on the legal and ethical aspects of data sharing in accordance with French law.
- Access to the DOI minting service and the institutional repository Data Inra (https://data.inra.fr). Data Inra hosts and provides access to data produced by or in collaboration with INRA.
- Templates and guidance for Data Management Plans (DMP); it includes an access to the DMPOpidor tool (https://dmp.opidor.fr/) which aims to help scientists in drafting data management plans.
- Access to a Q&A service allowing scientists to ask specific questions; the questions and answers are then made available to the visitors of the website

WUR's course on Research Data Management *(https://wgs.crs.wur.nl/courses/details/84/)*

Wageningen University and Research offers an introductory course for PhD candidates and postdocs on research data management (RDM). This course consists of three 3-hour mornings. For: PhD candidates and postdocs. It is offered by Wageningen Graduate Schools and given by WUR Library.

The course consists of lectures and practical assignments that cover various aspects of managing research data: from organising your data files during data collection to publishing your final dataset.
The course is divided into three parts:

- Part 1 focuses on how to set up your data collection system and how to keep it organised and ensure it is understandable.
- Part 2 deals with creating a data management plan (DMP), data storage options and the support and services available at WUR.
- Part 3 goes into the long-term storage and the publication of research data.

The full course content is available at:
https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/E8M8GI.

The GODAN Action Curriculum
(http://www.godan.info/sites/default/files/documents/Curriculum_final.pdf)

GODAN Action[56] is a DFID funded project that brings together agriculture and nutrition specialists and open data experts to support individuals, organisations and communities to deal with open data. The

---

[55] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[56] http://www.godan.info/godan-action

partners in the project all share a belief in the potential power of open data to make a real difference in addressing global food security and nutrition challenges. Our diverse experience working in this sector has taught us that, for this opportunity to be realised, it requires that all stakeholders in the process have, not just the required access to the data, but also the relevant tools and capabilities to use it. Without this some sectors and communities risk missing out - being left behind. GODAN Action therefore seeks to identify and address the barriers that prevent potential data users - from farmers to meteorologists, communities to international NGOs - from using open data effectively. We work to develop their skills, to build shared standards and approaches and to measure the impact of open data on their work. Led by Wageningen UR the GODAN Action project includes CTA, Open Data Institute, Agroknow, Land Portal Foundation, Food and Agriculture Organization of the United Nations (FAO), Institute of Development Studies and AidData.

The GODAN Action Curriculum was developed as part of the capacity development activities and was informed by a consultative workshop on the Open Data and Research Data Management Online Course that took place in Wageningen from 7 to 8 March 2017. From the curriculum a modular online course has been developed titled: “Open Data Management in Agriculture and Nutrition”.

The course is to be used in the context of different institutions in agricultural and nutrition knowledge networks and raise awareness on the different types of data formats and uses, and on the importance of reliability, accessibility and transparency. The curriculum was designed for three target audience groups:

1. Infomediaries, which includes ICT workers, technologist - journalists, communication officers, librarians and extensionists
2. Policy makers, administrators and project managers
3. Researchers and scientists

More specifically, by the end of the course learners will be able to:

- understand the principles and benefits of open data
- understand ethics and responsible use of data
- identify the steps to advocate for open data policies
- understand how and where to find open data
- apply techniques to data analysis and visualisation
- recognise the necessary steps to set up an open data repository
- define the FAIR data principles
- understand the basics of copyright and database rights
- apply open licenses to data

ANNEX 2 - CONTRIBUTIONS TO THE ROADMAP

The strategic goal of e-ROSA project is to provide guidance to European Commission's policies by designing and laying the groundwork for a long-term programme aiming at achieving an e-infrastructure for open science in agriculture and food sciences that would position Europe as a major global player at the forefront of research and innovation in this area. Through a foresight approach, the project has a shared vision of a future sustainable e-infrastructure for open science in agriculture and food sciences and made it operable through pragmatic recommendations that will be reflected in a common roadmap described in this document.

The process involved different stakeholders:

- *The vision* has been discussed in detail with stakeholders from the agri-food community and it has received input from **19 stakeholders** via an open consultation published at the project website
- *The grand challenges of the agri-food community that need to be addressed via a future e-infrastructure for open science in agri-food sciences*. This part is the output of the 2nd project stakeholder workshop held in Wageningen, Netherlands on 27-28/11/2017. The total number of stakeholders that participated to this workshop and provided their feedback/input was **49 stakeholders**
- *The current landscape of the agri-food community assets and maturity assessment* of these assets correspond to what exists and can be utilized for achieving the vision. This part is the output of the 1st project stakeholder workshop held in Montpellier, France on 6-7/7/2017. The total number of stakeholders that participated to this workshop and provided their feedback/input was **51 stakeholders**
- *The technical vision for an e-infrastructure that leads to an abstract future architecture, as well as a set of recommendations* towards achieving the future state of an e-infrastructure for open science in agri-food. This part is the output of the 3rd project stakeholder workshop held in Athens, Greece on 21-22/5/2018. The total number of stakeholders that participated to this workshop and provided their feedback/input was **16 stakeholders**

The eROSA project consortium warmly thank all these contributors!