eROSA
e-infrastructure Roadmap
for Open Science in Agriculture

Challenges & Solutions Envisioning Workshop

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

DELIVERABLE NUMBER	D3.2
DELIVERABLE TITLE	**Challenges & Solutions Envisioning Workshop**
RESPONSIBLE AUTHOR	Sander Janssen, Wageningen UR

GRANT AGREEMENT N.	730988
PROJECT ACRONYM	e-ROSA
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	WP2 | Challenges & Ambitions
WORKPACKAGE LEADER	WUR
DELIVERABLE N. | TITLE	D3.2 | Challenges & Solutions Envisioning Workshop
RESPONSIBLE AUTHOR	Sander Janssen, Wageningen UR
REPLY TO	Sander.janssen@wur.nl
DOCUMENT URL	http://www.erosa.aginfra.eu/deliverables
DATE OF DELIVERY (CONTRACTUAL)	31 November 2017 (M11)
DATE OF DELIVERY (SUBMITTED)	30 January 2018 (M13)
VERSION | STATUS	V1.1 | Final
NATURE	O(Other)
DISSEMINATION LEVEL	PU(Public)
AUTHORS (PARTNER)	Sander Janssen (WUR), Rob Lokers (WUR), Jandirk Bulens (WUR), Michael Chelle (INRA)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.9	Draft	12 January 2018	Sander Janssen (WUR), Rob Lokers (WUR), Jandirk Bulens (WUR), Michael Chelle (INRA)
1	Final version (after internal review)	25 January 2018	Sander Janssen (WUR), Rob Lokers (WUR), Jandirk Bulens (WUR), Rob Knapen (WUR), Michael Chelle (INRA)
1.1	Modification in Section 1.2.2	30 January 2018	Sander Janssen (WUR), Rob Lokers (WUR), Jandirk Bulens (WUR), Rob Knapen (WUR), Michael Chelle (INRA)

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

EXECUTIVE SUMMARY

This document presents the report of the 2nd e-ROSA Stakeholder Workshop "Challenges & Solutions towards an e-Infrastructure for Open Science in Agriculture" that was held on 27-28 November 2017 in Wageningen. The goal of this workshop was to look at the societal & scientific challenges in agricultural & food systems, that could be facilitated through the development of a global e-infrastructure in line with the European Open Science Cloud's agenda. 50 Participants with a good mix of European partners, International organizational, and knowledge partners from other continents gathered to discuss the necessary developments for open science from a food systems perspective.

As an overall conclusion, the overarching research challenge for open science on food systems can best be summarized as the need to design methods for better targeting of farmers, consumers, value chain actors and simultaneously improving efficiency, lowering environmental burdens, improving health. More data allows for more precise understanding of the different components of the food systems and their interactions, while at the same time recognizing the trade-offs that lie between these food system components, thus requiring the study of the interactions and a sharper 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.

To realise this overall scientific challenge, participants identified items in three pathways, that are starting points for the e-ROSA roadmap:

- Share: 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;
- 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.

With these outcomes, the participants achieved a joint answer to the following objectives:

- Identify research challenges that benefit from an open science e-infrastructure in agri-food;
- Identify common challenges in ICT & data that could be tackled with an e-infrastructure approach;
- Engage a broad community of scientists with a diverse background to ensemble transformative use cases.

Plenary sessions allowed to discuss pressing scientific challenges such as better linking food consumption and production, digitizing agriculture and improving nutrition. They provided examples of data-driven research and the use of ICT and data infrastructure to solve related questions.

The workshop was organised around three main break-out groups: 1) Smart farming, food security & the environment; 2) Gene-based approaches from omics to landscape; and 3) Food Safety, Nutrition & Health. In each break out group specific use cases were used to initiate more in-depth discussions. A first working session focused on the scientific & societal challenges enabled with e-infrastructures through the following questions: (i) what societal challenges can benefit from open science in agriculture and how do they benefit; and (ii) what scientific challenge have to be overcome to achieve these benefits. A second working session identified expectations & obstacles for using state-of-the-art data & ICT to solve the next generation scientific challenges, keeping in mind the need for upscaling of data & ICT opportunities as well as the domain-specificity vs. genericity of discussed issues.

The workshop in itself corresponds to Deliverable 3.2 under Work Package 3 “Roadmap co-Design & Uptake”. e-ROSA Stakeholder Workshops consist in a collaborative mechanism that allows to bring together the e-ROSA Stakeholder Community in view of envisioning the future e-infrastructure for Open Science in agri-food and co-elaborating the common roadmap that will support the implementation of this e-infrastructure.

CONTENTS

TABLE OF FIGURES

1 MEETING NOTES 2ND E-ROSA WORKSHOP: CHALLENGES & SOLUTIONS TOWARDS AN E-INFRASTRUCTURE FOR OPEN SCIENCE IN AGRI FOOD

1.1 KEY NOTES

1.1.1 Scientific and data challenges for addressing SDG 2: the missing middle (Linda Veldhuizen, Wageningen UR and SDSN)

Linda Veldhuizen is part of the Sustainable Development Networks (SDSN), which is a expert and academic network involved in the formulation of the SDG's. SDSN is a global network of 700 partners worldwide with a secretariat in Paris. Linda coordinates on behalf of WUR a network of partners for SDG 2. There are some specific challenges in SDG 2:

- Food consumption: undernourishment is 11% and increasing again.
- Stunting in children: 23%, but also overweight is increasing also in the developing world
- Food production: Agricultural productivity, still large yields gaps around the world, What is sustainable agriculture?

Interesting part is what is sitting in-between the food consumption & production. For example, increases in production just don't lead to sustainable consumption in the same geography. This is called the missing middle. There are quite some examples of these missing middles concepts. In government, the Ministry of agriculture might not be in contact to the Ministry of Health. Similarly, in companies, using sustainable farmed sources for fast food does not lead to a healthy consumption. In research, also these sort of gaps introduced, and we have large differences in the things we study. Some researchers are working on diet advice, while their colleagues are working on lowest level of detail getting more nutritional values in the foods they are producing. There are certainly methods in research to bridge these gaps, through participatory backcasting activities. Linda mentioned projects starting using new methodologies to identify this missing middle and overcome some of the barriers. Such emerging communities as around SDSN need support from e-infrastructures, to facilitate their communications and to find out what is going.

1.1.2 Research Challenges in supporting digitization of agriculture (Brian King, CGIAR Big Data Platform)

Brian King, from the CGIAR Big Data platform, introduced the Big Data platform, which is just starting from September 2017, when it had its kick-off. Brian King introduces the topic of scientific challenges with the example of an Indicator Framework developed at Michigan State University, that looks at trade-offs across disciplinary indicators in scales. This could then also link to the data source organization, doing that it in the same way as the indicators, to a data infrastructure.
CGIAR is a networked organizations, with different capabilities in IT, with different science focus, and different levels of compliance required in its different geo-graphies. There is the danger of thinking of everything in terms of a definite platform to rule them all. This will be unlikely to happen. There is definitely a trade off between tweaking for a specific purpose and the breadth of the utility of the platform. We should be aware of the broader applications of platforms, but accept that these have a specific purpose still. Brian mentions the development of the CGIAR CERES platform, that serves as an aggregator across CGIAR data sources and centres served.
The connections across the domain are as important as the data in the domain. There might be network effects here, it might be possible to make connections across the networks through the use of data infrastructures, by doing more network anlysis. Showing people who is working on what, that might be relevant to you, so that it can be connected easily.

1.1.3 A national agenda for precision agriculture, and its link to data/ICT infrastructures (Frans Lips, Ministry of Agriculture, Nature and Food Quality, the Netherlands)

Frans Lips is working as a policy advisor at the Ministry, working on challenges in relation to the use of data and ICT (especially geo-data) in agriculture practice. The Netherlands recognises 7 to 8 different sectors in agriculture (open cultures, closed horticulture, Dairy, intensive animal husbandries, fisheries). The Netherlands is a large agricultural producer with 100B Euros (8% of GDP) with half of the Dutch soil under agricultural. There are some trends happening: ageing of the sector, decreasing number of farmers, pressures of space. There also a lot of pressures: dependencies on imports, environmental pressures (with excess of manure), long complex supply chains, soil compaction (due to low country), climate change, low profit margins, …. Frans himself beliefs in technology development, and solving problems through technology.
He is heading a thematic development on precision farming, that should benefit the farmer through the better use of data, sensors and ICT, and impact on people, planet & profit. There is not a clear research agenda for precision farming in the Netherlands, but the High Tech to Feed the World seems to be the most relevant program, all the projects are on the Topsector AgroFood website.
Frans himself was involved in the Open data policies of the Ministry of Agriculture, Nature and Food Quality. Lately the soil data (as managed by Wageningen UR) was published through an effort of the Ministry. Frans also worked on the availability of the Satellite data, through the Dutch National Satellite database as pre-decessor for the Copernicus program. There is also the AgroDataCube, as an open & big data processor.
The Ministry lead a process to find the state-of-play for the precision farming in the Netherlands. The ministry was getting impatient with the progress with precision farming, and Wageningen carried out an exploration of the state of play, and this lead to the identification of a lot of of barriers (e.g. too short time frame of projets, lack of investment by farmers, technologically complex).

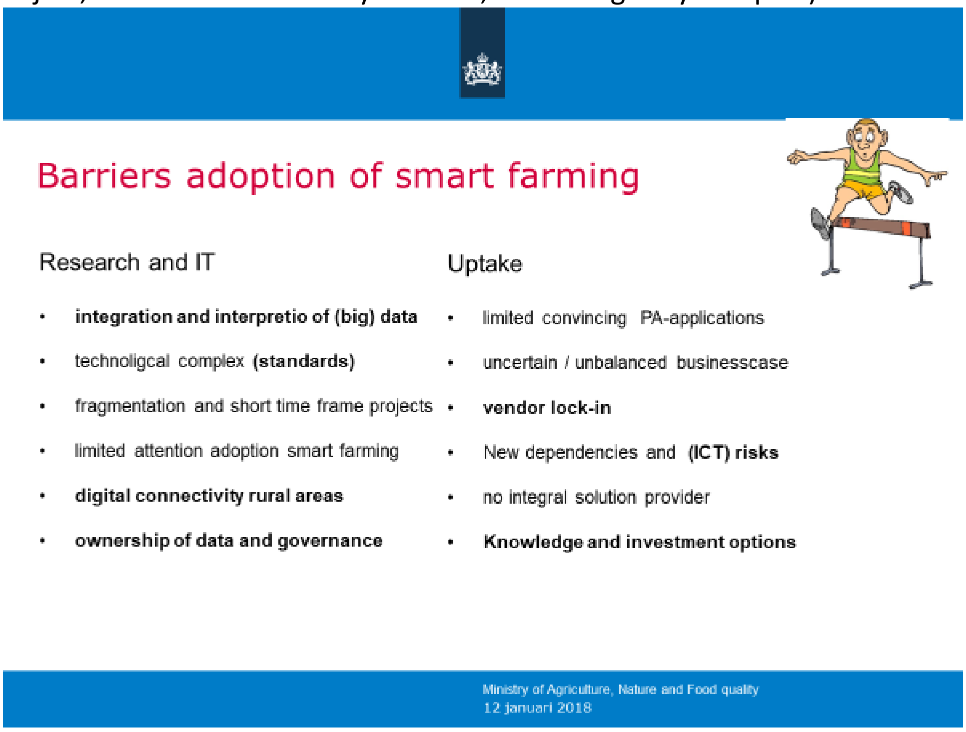

Figure 1. Barriers to adoption of Smart Farming as presented by Frans Lips from the Ministry of Agriculture, Nature and Food Quality in the Netherlands

1.1.4 New applications with satellite information and remote sensing (Rolf de By, ITC Twente)

There are many spatial unknowns in small holder farming, e.g. cropland acreage, acreage per crop, yield per farm management unit, cropping system and calendars. Agriculture in industrialized countries is data rich, agriculture in smallholder farming is data poor, we don't really know what is going on, and don't have the validation data. The STARS project worked on exploitation of high resolution images for smallholder agriculture, in three zones in the world (West Africa (grass roots), East Africa (government), and Bangladesh (small enterprises)). There were a lot of trials there. The objective was to produce Global Public Goods, that could be taken up by the community in the long term. The STARS landscape study identified 10 opportunities for applying remote sensing & drone techniques for further development and investment. The project also delivered open data and open processing software that is shared on the project website. The team build new algorithms for processing the large amount of data and building up the library of tools for using and processing the RS data library. Three lessons learned from this presentation: 1. How to set up these infrastructures and the useful tools developed within a good institutional framework?; 2. How do we keep the models, data and algorithms alive in good software implementations that can work in 10 years?; 3. How to motivate students/researchers to document the meta-data and publish the data? There are now reward systems.

1.1.5 Food, nutrition and health research infrastructures (Pieter van 't Veer, WUR)

Pieter is working from the perspective of the consumer, and tells his students that what is on their plate is influencing what is going on around them on the food system. He is working with others on setting up a research infrastructures, on health and food. There is a link to also policy and industry. There are global trends on digitization, openness, and increased engagement with the consumer. There should occur a transition from fragmented disciplinary sectoral to more coherent interdisciplinary efficient research. With national governments and EC working to establish an ESFRI. Within the currently ongoing ongoing project the research community is working on impact cases along a few use cases: care homes for the elderly, personalized nutrition, etc.
Next to FAIR also the Global Data Protection Rules are important to ensure privacy. The building blocks are now Food Consumer and Health, with cross cutting data, tools and services. A conceptual framework has been made to harmonise the database, and tools. Different organizations are already involved in a platform, but there are now setting up MOU's for sharing more of the data.

1.1.6 Global Global Rust Network and its data based approaches (Jens Grobach Hansen, Aarhus University)

The Borlaug Global Rust Network is a truly global network, that has been working together for many years on tackling wheat rust. A wheat rust strand just turned up in Europe again in 2015. Denmark (Aarhus) hosts the Global Rust Reference Centre. The lab trains people and researchers and is the reference lab for analysing samples on wheat rust. The lab also has many facilities for virtual data sharing and open science, with the Wheat Rust Toolbox. Over 12 years through different projects, different tools have been build, becoming more and more complex as the time goes on. At the moment, it has three layers, there are the databases, then there is a business logic layer, that represents the toolbox which has been made FAIR. A lot of it is about data provision and archiving. There are many stakeholders also in national governments working together to analyse the data with the tool box, and sharing the data. Through maps it is visualized which strain is where, and how they are spreading, for also comparative analyses. Although Aarhus University is serving the maps, it is not very visible in the websites and tools, as the user does not care about this.
Jens gives the example of the return of Yellow Rust in Italy in 2016. They coordinated with Nature and FAO to share the news on the same day to create maximum awareness.

The Rust Network in continuously making its data FAIR, directly at the point when the data is uploaded. Jens advocates for a network of networks or systems of systems perspective. He believes that data that is lost is unFAIR.

1.1.7 Embrapa's approach to open agricultural science (Patrícia Rocha Bello Bertin, Embrapa)

Embrapa is the national agriculture research institute in Brazil, it is linked to many other institutions within Brazil. Within Embrapa there are now many developments towards open science. At the moment, there is still a lot of work needed to build the next generation of data sharing, on cultural sharing, addressing tension between societal goals and business interest, etc. There is a large movement on open government and open government data, there is a large motivation to comply with the law, next to supporting innovative business and supporting innovative research. The government is now looking at opening databases that are relevant to research. If the database is not classified as sensitive, it should be made open. Embrapa has to deal with these requirements. There is a project, lead by Patrizia to lead to more shared data sources, with lot of steps. There are over a 100 people working on finding the relevant data sources, by March 2018, so that these can published online, starting by July 2018. Guiding question: from all data produced by your unit, which datasets possess greater potential for openness? Considering the current data management, also capabilities in terms of IT and the potential relevance to others in society. Data that has the potential to be made open needs often data rescue to improve it's annotation. To get open data in science we need cross research funding agencies and government to align their policies on open data sharing. It needs an institutional program.

1.1.8 Towards a domain-specific e-infrastructure: the example of DARIAH

DARIAH has been set up as a European Research infrastructure, already since 2012. It is working on another field of science, the humanities, and provides some digital services to support its researchers. It is one of the first European infrastructures, and getting on the EU infrastructure roadmap took about 7 to 8 years. Efforts were made to bring on board research institutions from different member states, where also member states governments were involved to provide their commitment. With this set up it is important to remember that each member state will implement this differently, and that there is no magical bullet here. The research infrastructure thus needs to be flexible to deal with the differences between member states. Central to such a long process are highly committed individuals at some core partners that can steer this process, and have the patient to wait for the final outcomes. When something is recognized as a research infrastructure, it is a mechanism to launch new calls according to the research priorities of the community. The running of the network itself has some small funding attached, but is mainly dependent on in-kind contributions of the participating institutions to take the lead of the secretariat. There are different bodies overseeing the science and the policy governance of the research infrastructure. Two crucial elements to the success of DARIAH have been it's role to provide digital services to researchers that are not ordinarily provided by commercial or normal university services (thus specialized to the community) and its role to provide training and capacity building to young researchers that are then more likely to remain part of the research community.

1.2 BREAK OUT SESSIONS

Each of the break-outs discussed the same questions, in two rounds. Each round had some overlapping and some different participants in the room to allow for a good mix of discussions. The questions for discussion where:

- What societal challenges can benefit from open science in agriculture? How do these benefit?
- What scientific challenge have to be overcome to achieve these benefits?
- Expectations & obstacles for using state-of-the-art data & ICT to solve the next generation scientific questions/challenges?
- How can we bring data & ICT opportunities to scale?

- Specify what is specific & generic to the break out topic: What expectation & obstacle do you expect that is only related to your topic? What are expectations & obstacles that are generic across topics?

The set up of the break outs was such that they cover different parts of the Food System approach as conceptualized by DG-RTD in its FOOD2030 strategy (Figure 2 below), one topic covering the consumer, health, nutrition, logistics and packaging aspects called Food Safety, Nutrition & Health, and a second topic covering the production, processing and waste streams aspects titled 'Smart farming, food security & the environment'. Finally in a third session the enabling conditions for system wide innovation at a higher and lower aggregation level were considered with the title: `Gene-based approaches from omics to landscape.'

Figure 2 The Food System approach as presented by DG RTD as part of the FOOD 2030 strategy

1.2.1 Smart farming, food security & the environment

The session started with two short presentations on the topic of Smart farming, food security & the environment.

Rob Lokers presented some examples from the field of crop yield forecasting, focussing on the European MARSOP initiative and future challenges in the face of Open Science. Societal challenges are evident and lie in a multitude of fields, e.g. food security, disaster risk reduction and sustainable and climate-smart agriculture. Improved exploitation of e-infrastructure should particularly lead to more efficient and transparent workflows from data acquisition to data publication and to better semantic linkage between the agronomic and related domains to enhance interdisciplinary solutions. For the research community the main challenges, besides the evident one of improving methods and models, lie in changing the attitude towards sharing of resources and improving on data curation and data stewardship.

Jandirk Bulens presented work on the setup of services for smallholder farmers in developing countries, specifically referring to a case study in Ethiopia.

The session was performed two times with two different groups of participants, specifically aiming at discovering the main societal and scientific challenges relevant for the topic.

The following challenges were identified by the two workshop groups:

Societal challenges

- While e-infrastructures and virtual research might initiate disruptive changes in food production, this should be accomplished without damage to the less favoured, specifically smallholders and consumers. It should also ensure a fair & sustainable position for (smallholder) farmers, e.g. regarding data ownership, data protection.
- Open science should adopt an inclusive approach, and not be isolated in a "scientific cocoon". Particularly the support and the use of knowledge from local communities was mentioned.
- Open sciences should include thinking about opportunities for new business models – for example "agriculture as a service"
- Open science should (also) focus on non-intensive farming (smallholder, organic etc.)
- Approaches should balance between a supply and (especially qualitative) demand driven approach, e.g. include health and nutrition as an important driver for research.
- Open Science should take into account and ensure responsible ownership of data, especially where it concerns the less powerful.

Scientific challenges

- Open science should aim at improving the data value chain in general but should specifically focus on the local scale:
- By using more timely and more localized data and knowledge
- To be able to serve local stakeholders and provide more precise and localized advice
- Development of Open Science as a working field should also take aboard e-capacity building, and not only for scientists but also for intermediaries, NGO's, farmers etc.
- Open Science should focus on opening up and sharing data and sharing of e-infrastructure (hardware, software, data repositories etc.) for the benefit of society as a whole.

Building on the results of the challenges session, participants discussed the current obstacles for Open Science with regard to the topic smart farming, food security & the environment. Also the main expectations were discussed. Remarkably the discussion revealed a lot of obstacles and expectations that are more generally linked to Open Science and only a small part of the findings were specifically related to the topic.

The following obstacles and expectation were identified by the participants:

Obstacles

- There's still a large cultural issue in the research community blocking the full implementation of Open Science, specifically a lack of culture for sharing and re-use of data, methods, knowledge.
- There's a big gap between current scientific working practice and Open Science (reg. ICT's, capacity, IPR, licensing models etc.)
- There is for most researchers still a lack of incentives to practice OS, besides the scientific mechanisms to reward "openness", this includes a lack of understanding of viable business models.
- There is a lack of advocacy and education for Open Science
- There's still an issue of trust around new technologies, like big data analytics, (e.g. privacy & commercial issues) with scientists and society in general, that creates a negative atmosphere.
- There is a lot of uncertainty around ownership and IPR of data and knowledge that is openly shared.
- Lack of information on provenance, traceability, transparency on currently available resources for OS, results in lack of trust and hinders the broad uptake for OS.

- There is not enough knowledge on standards & interoperability. If and how standards can be applied to specific resources is often unclear, e.g. because providers are not clear on it, because (use of) standards are not properly documented or because tools and guidance are not available.

Expectations

- Open science and e-infrastructures will not only support agricultural production but also benefit the environment and strengthen livelihoods.
- Open science will also come with the development of more respect for and protection of privacy (e.g. of farmers)
- With open science we will get a better grip on data sharing and data protection
- OS will bring better valorisation opportunities (economic-monetizing, scientific-citation etc.)
- Open science will lead to a culture of sharing data, methods and knowledge and to more effective and collaborative research
- OS will make it easier to work on broader, cross-domain and cross-community use cases
- A well-developed OS ecosystem will also lead to better access to better data and data integration tools that are easier to use.
- OS will come with associated activities to improve capacity to work with e-infrastructures
- OS will lead to “reverse science”, using data analytics as the input for new research

1.2.2 Gene-based approaches from omics to landscape

The session started with two short presentations on the topic of Gene-based approaches from omics to landscape.

Michaël Chelle presented the topic of phenotyping, taking many examples from the plant science. Phenotyping becoming phenomics faces several challenges regarding. The development of high-throughput measurements raises questions about data storage, reuse, integration, as well as the urgent needs in terms of meta-data and semantic level to well characterize. He illustrated the statements with the experience of the EU project Emphasis. As for the first break-out session, the development of interconnected e-infrastructures should particularly lead to more efficient and transparent workflows from data acquisition to data publication and to easy the required cross-disciplinary and integrative approaches.

Elizabeth Arnaud (CGIAR) presented the topic of plant biodiversity-based research. Contrarily to molecular sciences and bio-informatics that enable to generate and provide ready access to huge amounts of biological datasets, researchers working on biodiversity are facing increasing complexity, cost and uncertainty with regard to access to, use and exchange of biological material and information. Bioversity International’s research projects on identifying proper varieties fitting the farmers’ needs through participatory projects are based on the hypothesis that if farmers have better information and access to a wide range of species and varieties, they are more able to choose what best suits their conditions and cope with unpredictable weather.
However, due to the scattered status of the data on agrobiodiversity, on useful plants and ecosystem services as well as many gaps in knowledge, a farmer, policymaker or scientist cannot currently draw on evidence to guide the selection of species that can be planted together on-farm and maintained by farmers to restore soils or mitigate drought risk while providing nutritious foods and reliable yields. Therefore, extracting and harmonizing functional trait data with ontologies is a basic service that will support efficiently decision-making tools and can be initiated by a multidisciplinary agricultural research community.

The following challenges were identified by the two workshop groups (in bold, the cross-boundaries items):

Societal	Scientific
• Developing efficient plant and cattle breeding to provide genetic solutions to the disruptive changes in food production • Breeding to support non-intensive farming (smallholder, organic etc.) • Speed-up the control of new invasive species (pests) • Providing genetic solutions adapted to the end-user needs (farmers, consumer, etc) • Helping the development of plant participatory breedings	• Helping the up-scaling: from omics to population • For plant breeding, easy the extrapolation of results from lab to field(S) • Improving the characterisation of the environment components of phenotyping systems. • Develop model-assisted breeding • Providing an alternative to GMOs? • **Opening and sharing data** • **Sharing of e-infrastructure (hardware, software, data repositories etc.)**

Obstacles	Expectations
• Available skills to take profit of the open-science approach • Shared and adopted international standards • Starting from problems: having an actual and efficient user involvement • **Integrate a large diversity (type of data, cultural differences between omics and higher-scales communities, IT skills,…** • Having actual interoperable systems • Involvement of private companies (which business model, which IP?) • Available innovation platforms • **Different levels of progress between the plant, microbiome, and animal communities** • Knowledge gap between current scientific working practice and Open Science (reg. ICT's, capacity, IPR, licensing models etc.)	• Better understanding of positive and negative impacts of openness and sharing • Easier to work on broader, cross-domain and cross-community use cases • E-infrastructures to not only favour data exchanges and analysis, but also models and training • The FAIRification should be transparent • Better valorisation opportunities (monetizing, citation etc.) • Higher virtualisation of the IT system: web services, cloud => interoperability, scaling up, traceability, security, etc • **Demonstrating cases of linked data use and analytics.**

1.2.3 Food Safety, Nutrition & Health

The break out session was introduced by two presentations, one on Food Safety by Matthias Filter from BFR in Germany and another one from Jan Top from WUR in the Netherlands.
In his presentation, Matthias Filter focused on getting knowledge and insights out to different audiences. Scientists in food safety very much talk to scientists, while consumers and other societal actors more and more start to have an interest. These could be involved, but we need different modes of communication of making them understand also aspects like risks and uncertainty (or even probability), without immediately being scared of food safety risks. Next, he focused on the aspect of

knowledge integration. There is quite a lot of data out there, tools and models, there is a lack of integration across the tools. We don't have any standards and meta-data to describe the data and tools. Data standards is a foundation for knowledge interoperability. So, this needs urgent attention to link the different aspects.
In food safety it is crucial to consider the role of the risk manager, who makes decisions about the risks and estimates their impacts. His or her work needs to be better supported by the tools we develop as a community of open scientists, so we need to be aware of his actions and needs. This links to another point that is the process of a food outbreak investigation. There are steps followed, and open science has the potential of making this easier, but this needs to be thought through and well understood.

In his presentation, Jan Top took the perspective of focusing on Sustainable food behavior by posing a couple of questions and raising some thoughts. First of all, a food system approach needs to think and work along the value chain. The supply chain is a type of network, with the consumer at the end, which is often not well incorporated. So, how can we incorporate the role of the consumer? For example what if everyone is eating meat? What does this mean in the protein-transition? What are the environmental aspects?
Agriculture is not only producing food anymore, but also fuel, and other materials for the the bio-economy. This will have an impact on food availability and composition, although we don't really understand which impacts. We need to define and compute the scenarios in the bio-economy.
There is also a link to city planning and food system organization. We have this vision of sustainable cities, and we now need to understand what this means. Do we need to produce foods close to the city? Can we have a more realistic estimation on the shelf life on fresh products?
Traceability and transparency in the supply chain are considered important aspects, but also themselves have problems: What does the consumer really need/know want to know? Supplying detailed information on each and every ingredient is costly for the information. In some cases there might be win-wins for the business supplying this type of information and for the consumer accessing the information. There might be trade offs here, potentially the consumer will be happy enough if he or she knows that he can access the information that is around?
Finally there is a real challenge in linking the health and food side aspects, and again with the food production. As an example, there is a Dutch program on Personalized Nutrition & health: what are the heuristics around the rules of eating healthy food? How do cultural aspects link in? This sort of analysis needs to be related to scientific evidence in the context. What is the relationship with the individual? A complicating factor is that product data is unreliable and not accurate.

The discussion of the participants over the two break-out sessions in the workshop raised answers to almost all of the questions posed, although not necessarily in good order. In terms of societal impacts benefits, the role of the consumer was stressed. Consumer well being and optimizing consumer decision space for healthy and/or sustainable products/lifestyles are overall goals. This means that we need to understand more about the consumer and his/her choices. Another impact is the tackling in-efficiencies in the supply chain, thereby improving the supply chain. Third impact is to help achieve better risk assessment of food and heath issues, early on the production, processing and sales. Risks should be estimated early on, and potentially in a preventive methods instead of a response -to-outbreaks mode. However, if an outbreak occurs, then the tracking and tracing needs to be fast, and the faster the response is, the more targeted it can be, and the smaller its impact. Finally, there is a large potential to use data-intensive technologies to lower food waste, as still a lot of food is being 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 steps. In tackling in-efficiencies in the

supply chain, research needs to identify cost-effective ways of improving and tackling in-efficiencies, while not increasing the transaction costs and overhead in the supply chain.
For the impacts around food safety, research challenges are the identification of emerging risks, in a very targeted way. This needs to connect to designing adequate responses at each step of the supply chain and of managing outbreaks. With respect to food waste & losses, research needs to build a comprehensive data based understanding of food waste and loss, on where it occurs, how it can mitigated and what are efficient solutions to tackle it.
To enable the work on these societal and research challenges, advanced data, ICT solutions foundational to open science could play a role. However, in the present there are a few obstacles that hinder the wide spread adoption of such solutions and Open Science. The first obstacle is the differences purchasing power in the supply chain for accessing the data. There are some strong powerfull players, like large multi-national cooperations and supermarkets, who can buy access to relevant data, that is not available to publicly funded research. It is therefore sometimes hard for public research to keep up with such organizations. A second obstacle is the lack of public research infrastructures that work fully along the long supply chain. Often the research infrastructures only cover part of the supply chain, for example production or processing, but not from start to finish. A third obstacle is the perception (and reality) that the access or ownership of data equates to power and thus as has financial benefits. This obstructs the sharing of data to other partners in the supply chain, or to research organizations. A final obstacle is the lack of models for benefit sharing across the supply chain, so that all players experience equal benefits of the available research.

In terms of expectations and obstacles with respect to open science for the topic `Food Safety, Nutrition and Health‘, the following elements were defined:

Obstacles	**Expectations**
• Purchasing power in the value chain buys data access • Data = power = money • Lack of mechanisms of benefit sharing across the supply chain • Lack of public infrastructures that work along the supply chain • Legal validity and governance issues • Dissemination of scientific outcomes: raising sensitivity around risks and benefits • Lack of standardized vocabularies, lack of standardization. • Weaknesses in data curation and data rescue	• Better understanding of positive and negative impacts of openness and sharing • Urgently need data sharing arrangements • Need for a broader innovation approach than the current step in the supply chain • Demonstrating cases of linked data use and analytics. • Collaborative models with the different actors in the supply chain

1.3 SYNTHESIS

Finally, through rounds of synthesis (in sub-groups, individual and over groups), overlapping requirements and desires for e-infrastructure development for open science were generated based on the use case analysis. With respect to the overall synthesis, the overarching research challenge across domains can best be summarized, as the need to design methods for better targeting of farmers/consumers/value chain actors and simultaneously improving efficiency, lowering environmental

burdens, improving health. More data allows for more precise understanding of the different components of the food systems and their interactions, while at the same time recoginizing the trade-offs that lie between these food system components, thus requiring the study of the interactions and a sharper 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, in studying the food system and the better targeting requires multi-domain research, in which disciplinary boundaries are less prominent.

To realise this overall scientific challenge, participants identified items in three pathways, that are starting points for the e-ROSA roadmap:

- Share: 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 one researchers first;
- 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.

1.3.1 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 need to develop smarter interoperability platforms, that are easy to use, not challenging to use.
- To improve trust we require certification of datasets and repositories
- Practical guidance should be developed and offered (e.g. for making data FAIR) and tools should be developed to support this.
- We need more incentives, e.g. rewarding Open Science
- We must share our knowledge and offer guidance on tools
- We should connect classical science and data science

1.3.2 Connect

- Support the use of standards as much as possible, so that data, information and architectures can easily be connected.
    - At the moment there is a proliferation of standards, so it is difficult to choose which standard to use.
    - It is important to achieve a high level agreement on the type of standards required:
        - Data format (domain/data specific)
        - Meta data structure/schema, giving profiles of meta data standards
        - (Controlled) Vocabularies
    - Recommendations on which standards to use for what would be beneficial for a large number of researchers
    - Persistent identifiers should become common practice, but also here recommendations would be very helpful.

- Use cases are useful to identify benefits and gaps, and generic research questions need to be synthezid that are integrating and cross cutting across domains, such that solutions developed will have a broad applicability.
- Success stories and best practices are required with a domain-specific focus such that researchers know what to do.
  - Demonstrate added value of following the best practices
  - Creating learning resources for capacity building, so that researchers can discover themselves how to work with the offered services.
- Include models and analytical tools as part of open science next to open data.

1.3.3 Collaborate

- Open science should not only work for the public sector, but also in the private sector for researchers and analysts. This implies a role for public private partnerships in the development of infrastructures for open science.
  - Requirements for data sharing across public and private sectors, while respecting privacy and competitive concerns.
- Establish e-ROSA as a shared vision based on users needs and based on the strengths of partners. Infrastructure development needs to follow the user stories.
- In terms of user needs, don't just focus on researchers themselves, but:
  - Include funders
  - Focus on individual accomplishment of researchers, establishing large and small projects to connect to developments in open science..
- Advocate for an user centric approach in the development of the European Open Science Cloud.
- System-of-systems thinking is important, as there should not be one large organizational and technical infrastructure for open science, but regional and domain nodes are required.
  - Infrastructures needs 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 market place for working across institutes and sharing resources
  - Include a certification function, so that nodes are open science proof.
  - Current solutions are not scalable
- Data managers have an important role, and should move beyond making data available, and also include elements of data clean-up, rules-of-the-game, and FAIRification. Training materials for data managers are thus crucial resources, that need to be shared across partners.
- Open Science needs to be achieved with a long term perspective in mind.
  - Single project funding is not adequate. An European Research Infrastucture Consortium seems an option, but requires a long process.
  - Business models need to be elaborate, taking into account relevant viable assets that can survice 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

1.4 FINAL REFLECTIONS BY PARTICIPANTS

As a conclusion each participant gave a reflection on the important elements for the roadmap:

- The discussion is quite high level, make small project proposals or user stories
- Is e-ROSA going for an ERIC? Maybe there are different research fields that should tru to go there.

- Using innovation approaches to unlock the supply driven. It has a danger of being supply drien.
- Focus on capacity building of the researchers
- Data driven approach is from the push side, apply a user centred approach.
- Private sector research should be more connected to this process.
- Inventory of agricultural linking projects, and doing also a citizen science linkage.
- focus on social aspect of buiding the infrastructure -- NO professorware
- Linking to knowledge sharing (as done in GODAN Action), linking to other sciences.
- Reliability is an important to work together, next to transparency, incorporate this in the Vision
- Traceability as an important keyword for the vision and define the scope more scharply what is in and what is out.
- Train the policy and government officials, they have to support it. and sometimes they don't understand us.
- Capacity building need on the research managers, librarians,
- We should move beyond data, to tools and analytics → we need more than sharing of data
- What can we learn of existing research infrastructures?
- Show the value proposition behind it → changing the culture: we need strong statements.
- The first workshop was more technical, now we moved from the societal challenges. We were more focused on the scientific/societal change. We explored more the governance, user-need, and human issue. There is a link between the national strategy and international strategy. This also underlined the importance of the roadmap exercise→ thinking about the long term sustainability of the EOSC.
- The problem is very large. We have an elephant in the room, and only seeing a part of the elephant. We might seem to have the discussions over and over, but we are building an environment of trust. We are getting more people engaged.

2 AGENDA OF THE MEETING

12.00- 13.00: Walk in lunch & registration

13.00: Start of the workshop

13.00-14.00: Workshop objectives

- The e-ROSA process: towards an open e-science infrastructure for agriculture & the FOOD2030 (Odile Hologne, INRA)
- The roadmap & vision for an open e-science infrastructure for agriculture as a work in progress (Johannes Keizer, e-ROSA)
- Workshop objectives, approach and set up (Sander Janssen, Wageningen UR)

14.00-15.00 Key notes (15 min per presentation)

- Scientific and data challenges for addressing SDG 2: the missing middle (Linda Veldhuizen, Wageningen UR and SDSN)
- Research Challenges in supporting digitization of agriculture (Brian King, CGIAR Big Data Platform)
- A national agenda for precision agriculture, and its link to data/ICT infrastructures (Frans Lips, Ministry of Agriculture, Nature and Food Quality, the Netherlands, tbc)
- Global Research Platforms as a G20 initiative (Stefan Lange, Research Director, Thunen Institute)

15.00-15.30: coffee break

15.30-17.00 working session e-ROSA impact cases: scientific & societal challenges enabled with e-infrastructures

- World cafe with standing tables per e-ROSA storyline:
  - What societal challenges can benefit from open science in agriculture? How do these benefit?
  - What scientific challenge have to be overcome to achieve these benefits?
- 3 Topic tables, group circulates with 30 minutes each round. Starting with a short presentation (5 min) of the e-ROSA storylines:
  - Smart farming, food security & the environment
    - Rob Lokers: Crop Yield Forecasting
    - Jan-dirk Bulens: Services for smallholder farmers
  - Gene-based approaches from omics to landscape
    - Michael Chelle: Genomics
    - Elizabeth Arnaud: biodiversity data
  - Food Safety, Nutrition & Health
    - Matthias Filter: Food safety
    - Jan Top/Nicole Koenderink: food and nutrition
- Plenary feedback per topic

17.30 End of day

18.00: Drinks and dinner at Diels Restaurant (dinner from 19.00 at http://www.dielsrestobar.nl/contact/)

Second half day: Explore and discover

9.00-9.15: Looking back at day 1, lessons learned and next steps

9.15-10.15: Inspirational talks

- New applications with satellite information and remote sensing (Rolf de By, ITC Twente)
- Food, nutrition and health research infrastructures (Pieter van 't Veer, WUR)
- Borlaug Global Wheat Rust network and its data based approaches (Jens Grønbech Hansen, Aarhus University)

10.15-10.45: coffee break

10.45-12.00: Break out workshops: Expectations & obstacles for using state-of-the-art data & ICT to solve the next generation scientific questions/challenges

- Introductory presentation:
  - Embrapa's approach to open agricultural science (Patrícia Rocha Bello Bertin, Embrapa)
  - Towards a domain-specific e-infrastructure: the example of DARIAH
- Topics for break outs:
  - Smart farming, food security & the environment (Rob Lokers as facilitator)
  - Gene-based approaches from omics to landscape (Michael Chelle as facilitator)
  - Food Safety, Nutrition & Health (Sander Janssen as facilitator)
- Discussion between participants:
  - Expectations & obstacles for using state-of-the-art data & ICT to solve the next generation scientific questions/challenges?
  - How can we bring data & ICT opportunities to scale?
  - Specify what is specific & generic to the break out topic: What expectation & obstacle do you expect that is only related to your topic? What are expectations & obstacles that are generic across topics?

12.00-13.00: Lunch

Third half day: Syntheses to common challenges

13.00 - 13.30: Plenary feedback

13.30-15.30: Synthesis working session towards the Roadmap

- Feedback from first workshop (Odile Hologne, 10 min)
- Identifying cross-cutting issues to elaborate, buzz groups: participants discuss in groups of 2 or 3 individuals on cross cutting issues, and write them down on post-its (10 min), Organise post-its (20 mins)
- 2 rounds of 15 minutes World cafe on 4 priority issues (4 topic leaders: Rob Lokers, Michael Chelle, Sander Janssen, Nikos Manouselis)

15.30-16.00: coffee break

16.00-17.00: Closing plenary

- Feedback from Synthesis working session
- Individual feedback from participants: crowd sourcing with your 3 top priority for the e-ROSA vision & roadmap paper & lessons learned during the workshop
- Closing words by the e-ROSA coordinator Odile Hologne

3 PARTICIPANT LIST

	Name	Affiliation
1	Erick Antezana	Bayer Crop Science
2	Elizabeth Arnaud	CGIAR
3	Ioannis Athanasiadis	Wageningen UR
4	Sophie Aubin	INRA
5	André Bannink	WUR
6	Patricia Bertin	EMBRAPA
7	Hugo Besemer	FAO
8	Filipa Borrego	INOV INESC Inovação
9	Yamine Bouzembrak	Wageningen UR
10	Bert Bredeweg	University of Amsterdam
11	Christopher Brewster	TNO
12	Jandirk Bulens	Wageningen UR
13	Grigorios Chatzikostas	BioSense Institute
14	Michaël Chelle	INRA
15	Rolf de By	ITC Twente
16	Esther Dzalé Yeumo	INRA
17	Sonigitu Asibong Ekpe	Michael Okpara University of Agriculture
18	Matthias Filter	German Federal Institute for Risk Assessment
19	Richard Finkers	Wageningen UR
20	Sophie Fortuno	CIRAD
21	Anand Gavai	Wageningen UR
22	Ken Giller	Wageningen UR
23	Jens Grønbech Hansen	Aarhus University
24	Odile Hologne	INRA
25	Madeleine Huber	INRA
26	Sander Janssen	Wageningen UR
27	Clement Jonquet	University of Montpellier
28	Paul Kasoma	Yitedev
29	Asanee Kawtrakul	Kasetsart University
30	Johannes Keizer	e-ROSA
31	Brian King	CGIAR
32	Takuji Kiura	NARO
33	Rob Knapen	Wageningen UR

34	Frans Lips	Dutch Ministry of Agriculture, Nature and Food Quality
35	Cheng Liu	Wageningen UR
36	Rob Lokers	Wageningen UR
37	Nikos Manouselis	Agroknow
38	Hans Marvin	Wageningen UR
39	Pascal Neveu	INRA
40	Pasquale Pagano	CNR - ISTI
41	Martin Parr	GODAN Secretariat
42	Ilse Rasmussen	ICROFS
43	Laurent Romary	DARIAH (remote participation)
44	Alejandro Salazar Romero	Wageningen UR
45	Babis Thanopoulos	Agroknow
46	Jan Top	Wageningen UR
47	Pieter van't Veer	Wageningen UR
48	Linda Veldhuizen	Wageningen UR
49	Zhengcong Wang	Wageningen UR