Synthesis of results & contribution to roadmap (M12)

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

DELIVERABLE NUMBER	D1.5
DELIVERABLE TITLE	Synthesis of results & contribution to roadmap (M12)
RESPONSIBLE AUTHOR	INRA

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	Georgia Tzenou

WORKPACKAGE N. | TITLE	WP1 | Ecosystem and Community
WORKPACKAGE LEADER	INRA
DELIVERABLE N. | TITLE	D1.5 | Synthesis of results & contribution to roadmap (M12)
RESPONSIBLE AUTHOR	Madeleine Huber (INRA)
REPLY TO	madeleine.huber@inra.fr
DOCUMENT URL	http://www.erosa.aginfra.eu/sites/erosa_deliverables/D1.5.pdf
DATE OF DELIVERY (CONTRACTUAL)	31 December 2017 (M12)
DATE OF DELIVERY (SUBMITTED)	22 December 2017 (M12)
VERSION | STATUS	V4 | Final
NATURE	R(Report)
DISSEMINATION LEVEL	PU(Public)
AUTHORS (PARTNER)	Sophie AUBIN (INRA), Michaël CHELLE (INRA), Esther DZALE (INRA), Odile HOLOGNE (INRA), Madeleine HUBER (INRA), Sander JANSSEN (WUR), Johannes KEIZER (Agroknow), Nikos MANOUSELIS (Agroknow), Panagiotis ZERVAS (Agroknow)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
1	Vision paper	24/11/2017	Sophie AUBIN (INRA), Michaël CHELLE (INRA), Esther DZALE (INRA), Odile HOLOGNE (INRA), Madeleine HUBER (INRA), Sander JANSSEN (WUR), Johannes KEIZER (Agroknow), Nikos MANOUSELIS (Agroknow), Panagiotis ZERVAS (Agroknow)
2	Elaboration of section on online map	01/12/2017	Sophie AUBIN (INRA), Michaël CHELLE (INRA), Esther DZALE (INRA), Odile HOLOGNE (INRA), Madeleine HUBER (INRA), Sander JANSSEN (WUR), Johannes KEIZER (Agroknow), Nikos MANOUSELIS (Agroknow), Panagiotis ZERVAS (Agroknow)
3	Elaboration of section on bibliometric analysis	14/12/2017	Sophie AUBIN (INRA), Michaël CHELLE (INRA), Esther DZALE (INRA), Odile HOLOGNE (INRA), Madeleine HUBER (INRA), Sander JANSSEN (WUR), Johannes KEIZER (Agroknow), Nikos MANOUSELIS (Agroknow), Panagiotis ZERVAS (Agroknow)
4	Integration of modifications and finalisation	19/12/2017	Sophie AUBIN (INRA), Michaël CHELLE (INRA), Esther DZALE (INRA), Odile HOLOGNE (INRA), Madeleine HUBER (INRA), Sander JANSSEN (WUR), Johannes KEIZER (Agroknow), Nikos MANOUSELIS (Agroknow), Panagiotis ZERVAS (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

EXECUTIVE SUMMARY

This deliverable provides an update on the progress of the activities carried out under Work Package 1 “Ecosystem & Community” and is the continuation of Deliverable 1.4 “Synthesis of results & contribution to roadmap (M6)”. In particular:

It presents the updating process of the bibliometric analysis that supports the scoping of the community of scientific experts working within e-ROSA’s scope and discusses the revised approach and next steps;

It gives an overview of the main characteristics and population process of the e-ROSA online map, which seeks to identify key stakeholders and networks that belong to the e-ROSA Stakeholder Community and that can support and/or benefit from a future e-infrastructure for agri-food science;

It presents the first version of the vision paper that was elaborated by e-ROSA partners as a first input towards the final roadmap document.

This deliverable will be updated one last time at the end of the project (D1.6).

TABLE OF CONTENTS

1 INTRODUCTION

The e-ROSA project seeks to build a shared vision of a future sustainable e-infrastructure for research and education in agri-food in order to promote Open Science in this field and as such contribute to addressing related societal challenges. Its overall objective is to support the co-development of an ambitious, practical roadmap that expresses the needs of related scientific communities and stakeholders at global level and that provides the basis for the design and implementation of such an e-infrastructure in the years to come.

In order to support the elaboration of this roadmap, several activities have been implemented under Work Package 1 “Ecosystem & Community” to map and engage the e-ROSA stakeholder community:

- A bibliometric analysis in order to test a first source of information (i.e. scientific publications) to identify which research organisations work within our scope, on which specific topic(s) and through what kind of collaboration;
- The development of an online map to create an open knowledge base that will allow (i) the discovery by the e-ROSA community of identified key stakeholders and networks that compose the current data and (e-) infrastructures ecosystem in Europe and beyond for agri-food research (including generic e-infrastructures and services), and (ii) the identification by the e-ROSA community itself of new stakeholders, initiatives and infrastructures that can support the implementation of the e-ROSA roadmap;
- The analysis of strategic documents, roadmaps and existing infrastructures to nurture internal reflections on the vision and the roadmap to be developed under e-ROSA.

More detailed information on the overall approach of WP1 is detailed in Deliverable 1.4 “Synthesis of results & contribution to roadmap (M6)”.

This deliverable seeks to provide an update on the progress of those activities and present related outputs. In particular:

- It presents the revised search query that supports the bibliometric analysis and discusses related next steps;
- It gives an overview of the main characteristics of the online map and its population process;
- It presents the first version of the vision paper that was elaborated by e-ROSA partners as a first input towards the final roadmap document.

2 BIBLIOMETRIC ANALYSIS UPDATE

2.1 RATIONALE AND OBJECTIVES

An initial bibliometric study was carried out at the beginning of the e-ROSA project as a first attempt at describing the heterogeneous landscape of agricultural data science. It sought to:

- identify key researchers and research institutions that focus on this issue and analyse to what extend they collaborate;
- link these stakeholders to specific topics that relate to the data science issue in agricultural research in order to better interpret the landscape of identified stakeholders.

Deliverable 1.1 "Bibliometric study results" presents the methodology used and the results of this first study, and it highlights the need to update the study. In addition, the content of this deliverable has been reviewed during the mid-term review of the project, allowing to discuss the need for an update. Thus, the objectives of the updating process are to:

i. include new scientific publications;
ii. refine the scope of the study, and in particular the query applied to the Web of Science (WoS) to extract the study corpus;
iii. refine the expected results of the study, including the relevant format of outcomes for practical use and valorisation under the e-ROSA project:
    - elaborating a relational database that clearly highlights the links between authors, institutions, countries and topics would be of high added value for project partners in order to target key institutions;
    - using an interactive interface to present the data in user-friendly manner could allow for efficient dissemination across the e-ROSA community;
iv. re-evaluate and validate the adopted methodology with project partners according to the study's objectives; and
v. refine the topics of interest in order to fully reflect the data-related issues that the research communities are facing in the agricultural field (e.g. integrate the issue of data production).

2.2 UPDATING OF THE SEARCH QUERY

As mentioned above, one of the main objectives of the analysis update is to refine the search query as the latter has a significant impact on the volumetry and quality of the corpus initially extracted from the WoS. In order for the query to better reflect the scope of e-ROSA, terms from the initial query were re-evaluated and additional terms were identified from various sources of information that directly relate e-ROSA's scope.

The initial query consisted in the crossing of two lists of terms, one focused on agri-food issues and the other on data-related issues. Terms that were considered as too specific (e.g. "RDF", "Hadoop") or too general (e.g. "distributed graph") were taken out of the revised query. Some additional terms were proposed by e-ROSA partners while other terms were identified from specific sources of information, i.e.:

- the terms used in the metadata model used for the online map (see Section 3);
- the terms extracted from e-ROSA deliverables that had been delivered up to now (D1.1 "Bibliometric study results", D1.4 "Synthesis of results & contribution to roadmap (M6)", D2.1 "Identification of Grand Challenges", D3.1 "Community Building and Fine-Mapping

Workshop" and D3.4 "Meetings of the Policy Committee (M7)") thanks to the "Cogito Studio" tool[1];
- the terms extracted from the bibliographic references that supported Task 1.1.2 "Collect & synthesize stakeholder e-infrastructure plans & roadmaps" (see Deliverable D1.4);
- the terms used in two search query developed by INRA, one on food and another on the nexus health-agriculture-nutrition-environment.

As a result, the query was strongly consolidated, especially the list of data-related terms. We now distinguish four different lists of terms (i.e. sub-queries):

1) Data-related terms
2) Food-related terms
3) Agriculture-related terms
4) Terms that reflect both agri-food and data related issues (e.g. *precision farming*, *digital agriculture*).

The final list of terms was validated by e-ROSA partners. Additional tests were carried out by searching specific terms or combinations of terms in the WoS in order to see if they allowed to extract sufficiently relevant publications or if they would bring mainly noise to the final corpus. In the latter case, these terms were removed from the query.

The overall query is the result of the following combination of the four sub-queries listed above.

All queries below were applied to years from **2006 to 2017** and limited to the following document types: **Article OR Book OR Book Chapter OR Proceedings Paper OR Review.**

Type of terms	Sub-query number	Sub-query	Number of references extracted
Data-related terms	#1	TS= ("e-infra*" OR "Virtual Research Environment*" OR "cyberinfrastructur*" OR "science gateway*" OR "virtual lab*" OR "information technolog*" OR "information manage*" OR "data technolog*" OR "data science$" OR "data* manage*" OR "FAIR data*" OR "data principle$" OR "findable data*" OR "accessible data*" OR "interoperable data*" OR "reusable data*" OR "data* standard*" OR "data* harmonization" OR "data description" OR "data* interoperability" OR "metadata" OR "data* format*" OR "data annotation" OR "data* link*" OR "linked data*" OR "data* integration" OR "data* reliability" OR "data* provenance" OR "data quality" OR "data qualification" OR "data* veracity" OR "trusted data*" OR "certified data*" OR "reproducibility of research" OR "reproducibility of science" OR "research reproducibility" OR "science reproducibility" OR "reproducible research" OR "reproducible science" OR "transparency of research" OR "transparency of science" OR "research transparency" OR "science transparency" OR "transparent research" OR "transparent science" OR "data certification" OR "data* curation" OR "data* enrichment" OR "data* retrieval" OR "relevance of	**3,760,716**

[1] http://www.expertsystem.com/products/cogito-studio/

		[the] data" OR "value of [the] data" OR "data fitness for use" OR "fitness for use of [the] data" OR "data use" OR "data re?use" OR "data cycle" OR "data lifecycle" OR "data plan$" OR "data versioning" OR "data* discov*" OR "data* access" OR "data* publication" OR "data paper$" OR "data transfer" OR "data exchange" OR "file transfer" OR "data* stor*" OR "data archiv*" OR "data preservation" OR "data conservation" OR "information extraction" OR "information engineering" OR "information retrieval" OR "knowledge extraction" OR "knowledge engineering" OR "knowledge retrieval" OR "digital" OR "knowledge representation" OR "data visuali$ation" OR "conceptual graph*" OR "distributed graph-processing system*" OR "((knowledge or «graph-based») near/3 reasoning)" OR "Knowledge?base*" OR "open [impact] metrics" OR "natural language process*" OR "semantic*" OR "ontolog*" OR "controlled vocabular*" OR "variable?registries" OR "variable?registry" OR "data production" OR "data acquisition" OR "data collection" OR "bioinformatic*" OR "*omics" OR "phenotypic data" OR "observation* data" OR "field data" OR "socio-economic data" OR "Internet of Things" OR "sensor-based system*" OR "robot*" OR "drone$" OR "geo?locali?ation" OR "geographic information system$" OR "geolocation" OR "GPS" OR "data deposit*" OR "data infrastructure*" OR "data source$" OR "data warehouse$" OR "data repositor*" OR "data registr*" OR "data centre$" OR "data center$" OR "data* catalog*" OR "data* aggregat*" OR "data portal$" OR "data facility" OR "data facilities" OR "computer server$" OR "data processing" OR "data analysis" OR "data exploitation" OR "data treatment" OR "analytical facility" OR "analytical facilities" OR "analytics" OR "data?intensive" OR "data workflow$" OR "workflow interoperability" OR "hardware" OR "computation science*" OR "comput* service$" OR "intensive comput*" OR "high?performance comput*" OR "high?throughput" OR "supercomput*" OR "performance comput*" OR "advanced comput*" OR "quantum computing" OR "distributed comput*" OR "blockchain" OR "grid comput*" OR "grid technolog*" OR "grid connectivity" OR "high?grid" OR "middleware" OR "high?speed" OR "real-time analysis" OR "computer science*" OR "data?mining" OR "text?mining" OR "image analysis" OR "image technolog*" OR "image segmentation" OR "image acquisition" OR "high?resolution" OR "API$" OR "software" OR "script language$" OR "algorithm$" OR "data service$" OR "data provider$" OR "learning agent*" OR "machine learning" OR "algorithm learning" OR "deep learning" OR "artificial intelligence" OR "cloud service*" OR "cloud?based" OR "cloud infrastructure$" OR "cloud comput*" OR "cloud storage" OR "web?based?system*" OR "web service$" OR "online service$" OR "social media" OR "social network$" OR "user	

		interface$" OR "Internet technol*" OR "as?a?service" OR "web data" OR "linked open data*" OR "machine?readable" OR "machine?accessible" OR "big data" OR "data resource$" OR "data object$" OR "data item$" OR "research data*" OR "scientific data" OR "public data*" OR "open data*" OR "open science" OR "open research" OR "e-science" OR "e-research" OR "open access" OR "open research data" OR "free access to data" OR "data?sharing" OR "shar* data" OR "data ownership" OR "data governance" OR "data stewardship" OR "data licens*" OR "access?controlled data*" OR "data security" OR "data privacy" OR "hack*" OR "data skill$" OR "computing skill$" OR "data expert*" OR "IT specialist$" OR "data scientist$" OR "computer scientist$" OR "web developer$" OR "data platform$" OR "data economy" OR "data market" OR "data revolution" OR "digitisation" OR "data consumption" OR "consum* data" OR "data ecosystem" OR "data landscape" OR "data commons" OR "data-driven" OR "data-based") OR TS=((simulat* OR predict* OR numerical OR data) NEAR model*)	
Food-related terms	#2	TS= ("food*" OR "aliment*" OR "beverage$" OR "diet*" OR "meal$" OR "nutrition*" OR "flavour$") NOT (TS="web food*")	**810,345**
Agriculture-related terms	#3	TS= ("agri*" OR "agro*" OR "farm*" OR "land use" OR "landuse" OR "crop$" OR "cropping" OR "cropland$" OR "plant product*" OR "plant protect*" OR "irrigat*" OR "livestock" OR "cattle" OR "breeder$" OR "breeding" OR "feedstuff$" OR "feedlot$" OR "veterinar*" OR "animal health" OR "animal product*" OR "graz*" OR "pasture$" OR "dairy") NOT (TS="wind farm*")	**722,708**
Terms that reflect both **Agri-food and Data** related issues	#4	TS= ("ag tech" OR "agtech" OR "food tech" OR "agri* data" OR "foodtech" OR "precision agriculture" OR "precision farming")	**3,517**
Agri OR **Food**	#5	#3 OR #2	**1,406,148**
(**Agri** OR **Food**) AND **Data**	#6	#5 AND #1	**174,450**
((**Agri** OR **Food**) AND **Data**) OR **AgriFoodand Data**	#7	#6 OR #4	**175,900**

Finally, the sub-corpus covering 10 years from 2008 to 2017 (instead of 12 years) was extracted in order to focus the study on more recent publications. This sub-corpus represents a total of 156,630 references.

2.3 CLEANING OF THE EXTRACTED CORPUS

In order to be relevant, the query results need some cleaning mainly due to the following two reasons:

1) The WoS is queried using the "TS" shortcut, which means that it returns references that contain the terms in either the title, the author keywords, the abstract, or the Keyword+ field. The latter is a list which is added manually by Thomson Reuters editors from the analysis of the citations included in a scientific paper. This feature addresses specific needs that are not relevant for the constitution of a corpus such as ours. As a result, it adds much noise in our corpus.
2) Some terms are ambiguous and can bring references that are out of the scope of our study, for instance "crop" which applies to image editing.

This step of cleaning prevents from asking experts to evaluate obviously out of scope references.

In order to address the first issue mentioned above, the entire corpus will be split into 4 subsets with decreasing confidence levels. We will discard Sub-corpus-4 for any further analysis.

Name	Description
Sub-corpus-1	Terms of "Agri-food", "Data" and "AgriFoodandData" are found in the title or author keywords (but not limited to those fields).
Sub-corpus-2	- Terms of "Agri-food" are in the title or author keywords (but not limited to those fields) while terms of "Data" are found in the abstract (and not in the title nor author keywords); OR - Terms of "Data" are found in the title or author keywords (but not limited to those fields), and terms of "Agri-food" are in the abstract (and not in the title nor author keywords).
Sub-corpus-3	Terms of "Agri-food", "Data" and "AgriFoodandData" are in the abstract and not in the title nor author keywords.
Sub-corpus-4	Terms of "Agri-food", "Data" and "AgriFoodandData" are neither in title, nor author keywords, nor abstract. They are in Keyword Plus only.

For the second phase of cleaning which consists in identifying out of scope references, we will use Luxid Information Analytics on the remaining references (i.e. without Sub-corpus 4). A first attempt at cleaning has been carried out:

- Non-supervised classification has allowed to identify sets of references on animal migration, food marketing, nutritional behaviours, medicine, image editing, breastfeeding, etc., which were discarded.
- WoS Categories[2] were used to filter out relevant references from globally irrelevant clusters. WoS categories of interest are Agriculture, Food Science & Technology, Veterinary Science, Biotechnologies & Applied Microbiology, Remote sensing, Automation & Control Systems.

[2] These are applied to journals, not articles and cannot be used directly for this reason.

2.4 NEXT STEPS

Once the cleaning of the extracted corpus will have been completed, the next steps of the study are the following:

1) Validation of the cleaned corpus by e-ROSA partners in order to ensure that the study corpus in in scope;
2) Revision of the methodology used to analyse the corpus;
3) Analysis of the corpus;
4) Interpretation of results;
5) Dissemination of results.

Regarding the revision of the methodology for the analysis, the revised methodology should allow to understand in more depth the content of the corpus, especially in terms of topics covered. In particular, a refined clustering of the corpus should be conducted (e.g. according to most frequent keywords or expressions and links between them), supporting the creation of categories that reflect a relevant level of representativeness and granularity for each considered topic.

Regarding the interpretation of results, the mid-term review has highlighted the need to extract key findings/messages that describe the community of scientific experts in our field and that can be disseminated to stakeholders that are external to the project (e.g. EU and national representatives and funders).

Regarding the dissemination of the results, e-ROSA partners would like to identify an appropriate interactive visualisation tool in order for the e-ROSA community and other external stakeholders to be able to navigate through the data that has been extracted from the studied corpus. Such an open and interactive dissemination interface would allow to increase the visibility of specific individual members of the community as key scientific experts in e-ROSA's scope and to provide all individual members with knowledge on the community as a whole, thus engaging the community as a whole and fostering collaboration and shared awareness.

3 ONLINE MAP

As explained in the introduction of this deliverable, an online map has been created in order to support the mapping activity carried out under e-ROSA in order to identify key stakeholders and networks that compose the current data and (e-) infrastructures ecosystem in Europe and beyond for agri-food research (including generic e-infrastructures and services). The sections below provide an update on this task, and more specifically on the main characteristics of the online map.

3.1 METADATA MODEL

The eROSA metadata model is used to describe the entities of the eROSA stakeholder community, namely Organizations, Data Points, Initiatives and Facilities. A first version of the metadata model was proposed in D1.4 "Synthesis of results & contribution to roadmap (M6)" and it was refined after discussing and agreeing on the characteristics to highlight for each mapped entity. The mapping task carried out under eROSA aimed at conducting the necessary scoping and assessment of already existing resources and networks that have been developed so far and that can feed into the design of future e-infrastructures and/or services in the context of eROSA. Overarching goals of this task included:

- The identification of key stakeholders and networks that compose the current data and (e-) infrastructures ecosystem in Europe and beyond for agri-food research (including generic e-infrastructures and services);
- The creation of the online knowledge base that will allow the discovery and characterisation of this ecosystem.

The metadata model is presented in Table 1 and is the one currently being implemented in the online eROSA map (http://www.aginfra.eu/).

Table 1: Overarching metadata model for mapped entities

Organizations	Data Points	Initiatives	Facilities
ID	ID	ID	ID
Acronym	Acronym	Acronym	Acronym
Full name	Full name	Full name	Full name
Description	Description	Description	Description
Logo	Logo	Logo	Logo
URL	URL	URL	URL
Geographic coverage	Geographic coverage	Geographic coverage	Geographic coverage
Address	Address	Address	Address
Geographic coordinates	Geographic coordinates	Geographic coordinates	Geographic coordinates
Date added	Date added	Date added	Date added
Scientific discipline(s)	Type	Scientific discipline(s)	Type
	Access policy	Data science category(ies)	Development stage
	Scientific discipline(s)	Related organization(s)	Access policy
	Related organization(s)		Scientific discipline(s)
	Related initiative(s)		Data science category(ies)
	Related Facility(ies)		Related organization(s)
			Related initiative(s)

The different values under each field can be free or restricted depending on the field. Restricted values concern the following fields:

- The list of "Scientific disciplines" has been extracted from the Map of Standards for Agrifood[3];
- The fields "Geographic coverage", "Type of data point", "Type of facility", "Development stage", "Access policy" and "Data science category(ies)" only accept restricted values as listed in Table 2 below

Table 2: Restrained values for Geographic coverage, Type of data point, Development stage and Access policy

Geographic coverage (General)	Type of data point	Development stage	Access policy
National	Set	Project	Open
International	Catalogue	Operational service	Controlled
Other Geographic Region	Repository		Private
	Aggregator		

In the case of "Geographic coverage", each entity can be characterized as:

- "National", and it can be further elaborated by providing one or more countries (derived from ISO3166[4], containing a total of 242 values) that this entity serves/involves.
- "International" addressing entities, which they have a global coverage
- "Other Geographic Region" and it can be further elaborated by using broad geographic regions such as the ones mentioned in Table 3.

Table 3: Restrained values for Geographic coverage (Regions)

Geographic coverage (Regions)[5]		
Africa	Americas	Asia
--Northern Africa	--Latin America and the Caribbean	--Central Asia
--Sub-Saharan Africa	----Caribbean	--Eastern Asia
----Eastern Africa	----Central America	--South-Eastern Asia
----Middle Africa	----South America	--Southern Asia
----Southern Africa	--Northern America	--Western Asia
----Western Africa		

[3] http://vest.agrisemantics.org/
[4] https://www.iso.org/iso-3166-country-codes.html
[5] https://unstats.un.org/unsd/methodology/m49/

Geographic coverage (Regions)	
Europe	Oceania
--Eastern Europe	--Australia and New Zealand
--Northern Europe	--Melanesia
----Channel Islands	--Micronesia
--Southern Europe	--Polynesia
--Western Europe	

Table 4: Restrained values for Agri-food discipline(s), Data topics and Type of facility

Agri-food discipline(s)	Data Topics	Type of facility
Agri-food Economics and Policy	Community building	Farm
Agri-food Education and Extension	Computation	Land
Agricultural equipment	Data analytics	Garden
Agriculture - General	Data discovery & access	Ranch
Animal Production and Health	Data integration	Education center
Engineering, Technology and Research	Data interoperability	Research center/station
Farming Practices and Systems	Data ownership	Greenhouse
Fisheries and Aquaculture	Data production	Virtual Lab
Food distribution	Data publication	e-infrastructure
Food safety and Human nutrition	Data storage	Laboratory
Food Security	Data veracity	Field station
Food technology	Data visualisation	Network of research infrastructures
Forestry	e-infrastructure	
Geographical and Regional Information	Governance	
Government, Administration and Legislation	Information Management	
Natural Resources and Environment	Modelling & simulation	
Plant Production and Protection	Semantics	
Rural and Social Development	Standardization & recommendations	
	Support for decision-making	
	Technology adoption	
	Workflows	

3.2 DISCOVERY SITE

As part of the Ecosystem & Community work package of the eROSA project, an online map of stakeholders and resources related to open science in agriculture has been developed using the metadata model described above. This map is part of the AGINFRA web portal, is hosted at www.aginfra.eu and it aims at supporting the cartography and visualization of the ecosystem. The technologies used are Apache HTML Server, MySQL database and PHP. As for the content management, Drupal CMS was selected.

URL: http://www.aginfra.eu/

Figure 1: eROSA Discovery Page (text search)

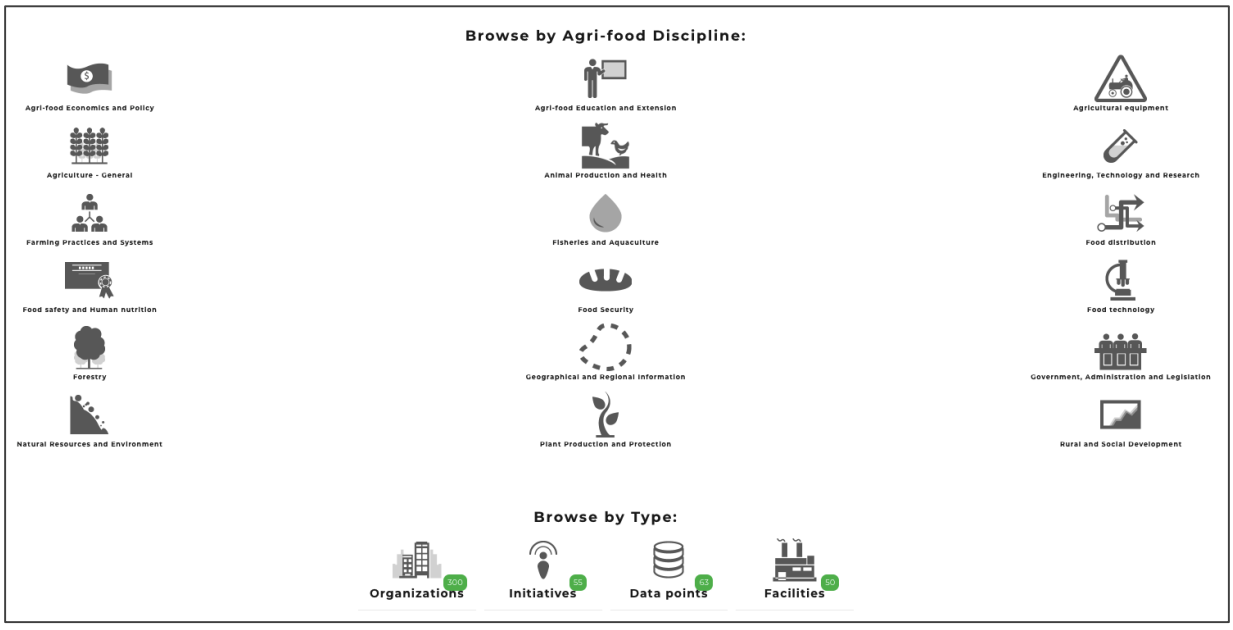


Figure 2: eROSA Discovery Page (browsing)

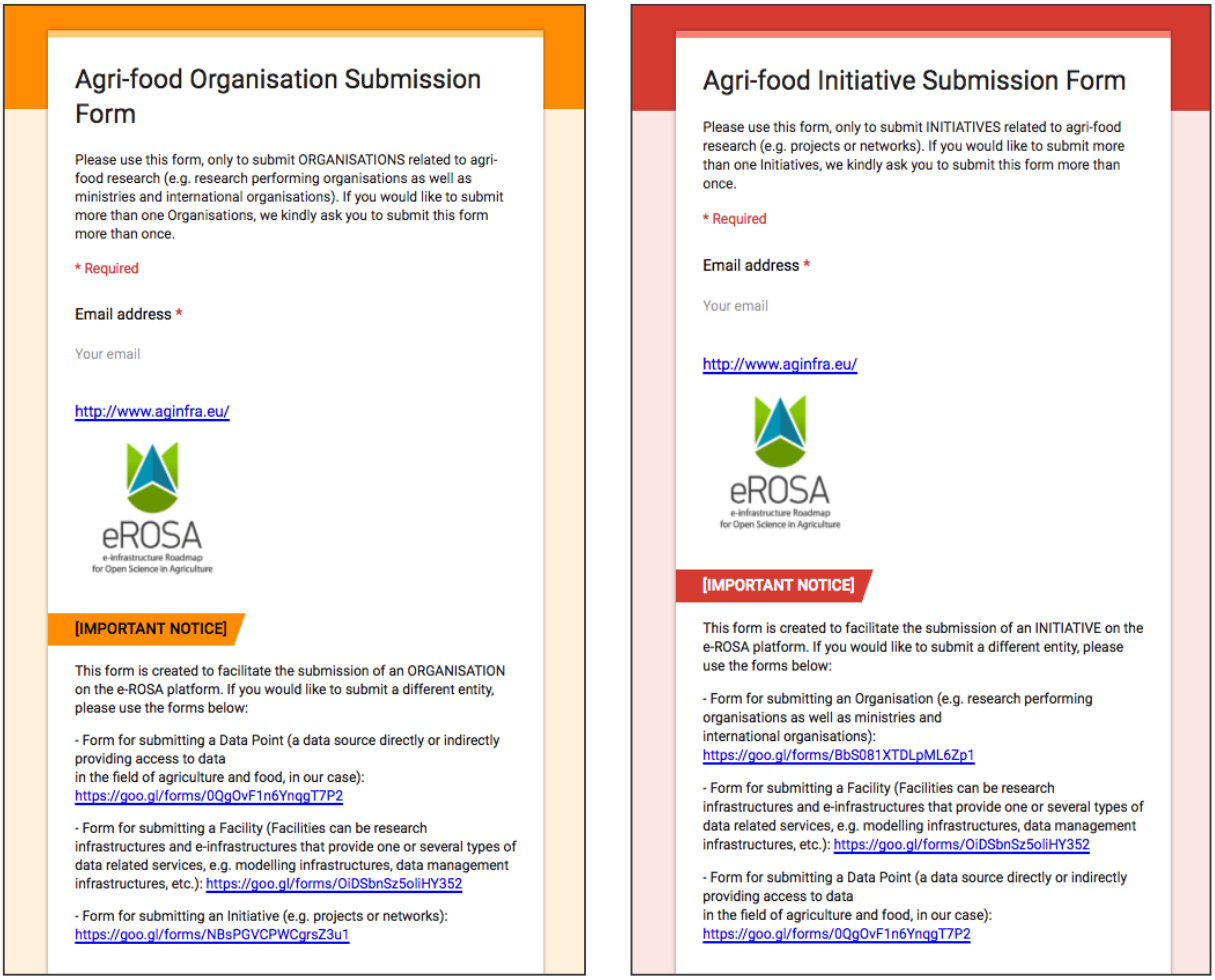
More detailed information about the discovery site can be found in the respective deliverable “D1.2 - Online map of stakeholders & resources (M12)”.

3.3 MAP POPULATION

Related to the initial population of the eROSA online map with entities (i.e. organizations, data points, facilities and initiatives), the eROSA partners carried out a desktop research. In the context of this desktop research, each partner collected input for the different entities of the map providing an initial set of entities that were then populated on the map to provide a concrete basis of organizations, data points, facilities and initiatives.

In addition to that, an open call to the wider community was launched to collect additional input, complementing the existing entities on the eROSA online map. To this end, four different forms were created, making sure that they request sufficient information from the contributors but also making sure that the richness of information requested is kept at a minimum level so that it is easier for the community to provide the input needed. Once the entities are collected from the community, an internal quality assurance process will make sure that all the entities collected are enriched with the full set of data that the metadata model requires. The forms for the open call are available through the following links:

- Organizations: https://goo.gl/forms/67dEMArMoNQZUtlt2
- Initiatives: https://goo.gl/forms/zEYNWqKUEvzQH3Av2
- Data Points: https://goo.gl/forms/JptZeZOUf8qj7vwz2
- Facilities: https://goo.gl/forms/8JkM0ZlTwv7AoJxd2



Agri-food Organisation Submission Form

Please use this form, only to submit ORGANISATIONS related to agri-food research (e.g. research performing organisations as well as ministries and international organisations). If you would like to submit more than one Organisations, we kindly ask you to submit this form more than once.

* Required

Email address *

Your email

http://www.aginfra.eu/

eROSA
e-Infrastructure Roadmap for Open Science in Agriculture

[IMPORTANT NOTICE]

This form is created to facilitate the submission of an ORGANISATION on the e-ROSA platform. If you would like to submit a different entity, please use the forms below:

- Form for submitting a Data Point (a data source directly or indirectly providing access to data in the field of agriculture and food, in our case): https://goo.gl/forms/0QgOvF1n6YnqgT7P2

- Form for submitting a Facility (Facilities can be research infrastructures and e-infrastructures that provide one or several types of data related services, e.g. modelling infrastructures, data management infrastructures, etc.): https://goo.gl/forms/OiDSbnSz5oliHY352

- Form for submitting an Initiative (e.g. projects or networks): https://goo.gl/forms/NBsPGVCPWCgrsZ3u1

Agri-food Initiative Submission Form

Please use this form, only to submit INITIATIVES related to agri-food research (e.g. projects or networks). If you would like to submit more than one Initiatives, we kindly ask you to submit this form more than once.

* Required

Email address *

Your email

http://www.aginfra.eu/

eROSA
e-Infrastructure Roadmap for Open Science in Agriculture

[IMPORTANT NOTICE]

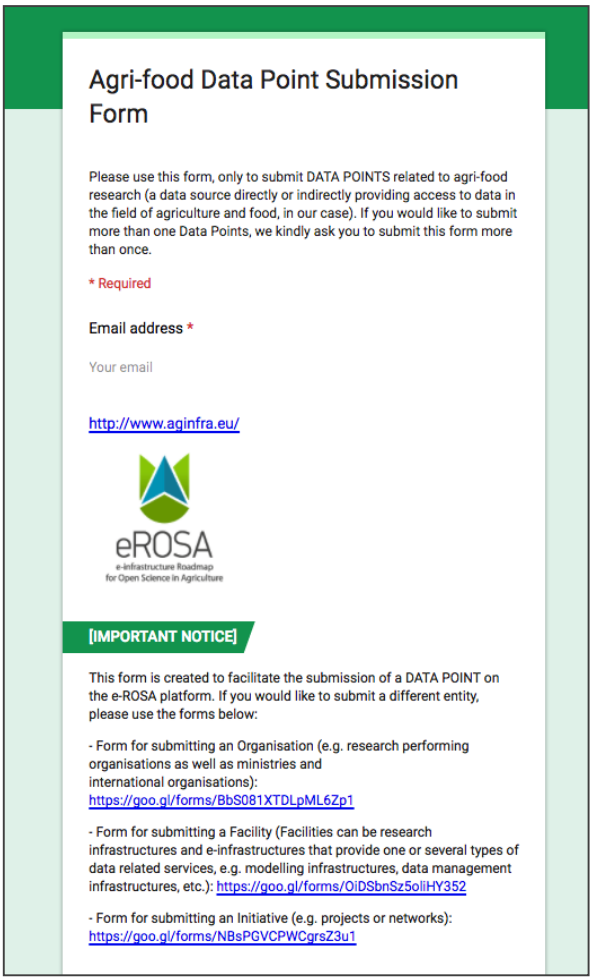
This form is created to facilitate the submission of an INITIATIVE on the e-ROSA platform. If you would like to submit a different entity, please use the forms below:

- Form for submitting an Organisation (e.g. research performing organisations as well as ministries and international organisations): https://goo.gl/forms/BbS081XTDLpML6Zp1

- Form for submitting a Facility (Facilities can be research infrastructures and e-infrastructures that provide one or several types of data related services, e.g. modelling infrastructures, data management infrastructures, etc.): https://goo.gl/forms/OiDSbnSz5oliHY352

- Form for submitting a Data Point (a data source directly or indirectly providing access to data in the field of agriculture and food, in our case): https://goo.gl/forms/0QgOvF1n6YnqgT7P2

Figure 3: Forms for open call for Organizations and Initiatives



Agri-food Data Point Submission Form

Please use this form, only to submit DATA POINTS related to agri-food research (a data source directly or indirectly providing access to data in the field of agriculture and food, in our case). If you would like to submit more than one Data Points, we kindly ask you to submit this form more than once.

* Required

Email address *

Your email

http://www.aginfra.eu/

eROSA
e-Infrastructure Roadmap for Open Science in Agriculture

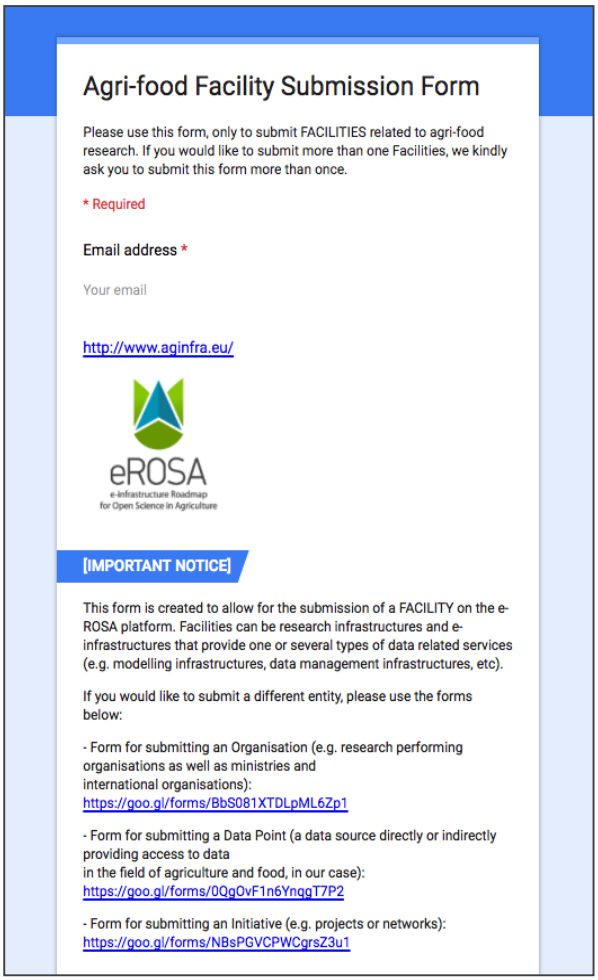
[IMPORTANT NOTICE]

This form is created to facilitate the submission of a DATA POINT on the e-ROSA platform. If you would like to submit a different entity, please use the forms below:

- Form for submitting an Organisation (e.g. research performing organisations as well as ministries and international organisations): https://goo.gl/forms/BbS081XTDLpML6Zp1

- Form for submitting a Facility (Facilities can be research infrastructures and e-infrastructures that provide one or several types of data related services, e.g. modelling infrastructures, data management infrastructures, etc.): https://goo.gl/forms/OiDSbnSz5oliHY352

- Form for submitting an Initiative (e.g. projects or networks): https://goo.gl/forms/NBsPGVCPWCgrsZ3u1



Agri-food Facility Submission Form

Please use this form, only to submit FACILITIES related to agri-food research. If you would like to submit more than one Facilities, we kindly ask you to submit this form more than once.

* Required

Email address *

Your email

http://www.aginfra.eu/

eROSA
e-Infrastructure Roadmap for Open Science in Agriculture

[IMPORTANT NOTICE]

This form is created to allow for the submission of a FACILITY on the e-ROSA platform. Facilities can be research infrastructures and e-infrastructures that provide one or several types of data related services (e.g. modelling infrastructures, data management infrastructures, etc).

If you would like to submit a different entity, please use the forms below:

- Form for submitting an Organisation (e.g. research performing organisations as well as ministries and international organisations): https://goo.gl/forms/BbS081XTDLpML6Zp1

- Form for submitting a Data Point (a data source directly or indirectly providing access to data in the field of agriculture and food, in our case): https://goo.gl/forms/0QgOvF1n6YnqgT7P2

- Form for submitting an Initiative (e.g. projects or networks): https://goo.gl/forms/NBsPGVCPWCgrsZ3u1

Figure 4: Forms for open call for Data Points and Facilities

4 VISION PAPER

As explained in the introduction, the vision paper below consists in the first essential input towards the final e-ROSA roadmap. It was elaborated by e-ROSA partners and discussed during the 2nd e-ROSA Stakeholder Workshop that took place on 27-28 November 2017 in Wageningen. It was also presented to EC representatives and other key players at the 2nd e-ROSA Policy Workshop on 11 December 2017 in Brussels. In short, it seeks to define:

- The envisioned "dream" for 2030;
- The overall challenges in research;
- The rules of engagement;
- And common major priorities.

This first version of the vision paper will be made available to all online. e-ROSA partners will seek feedback from the community on this version, especially from participants of the 1st and 2nd e-ROSA stakeholder workshops.

****

Vision

In 2030 food systems will produce healthy nutritious foods for all, produced through input-efficient methods and supporting a thriving environment. Food Systems will operate as collaborative networks that are constantly seeking to improve their economic, environmental and social performance for all actors of the network, and those food systems can be region specific, in local territories, but also global. The food systems contribute to the achievement of a wide range of objectives as captured by the Sustainable Development Goals[6] such as achieving food security, mitigating global warming, ensuring good health and preserving biodiversity.

These inclusive and resilient food systems are supported by open science based knowledge systems to stimulate further innovations. As such a food system is necessarily knowledge intensive. The food system includes the production of foods, environmental effects, its processing, distribution, its nutritional values and its healthy consequence, requiring to look beyond the individual value chains, individual crops or livestock types, and farm types. In the open science based knowledge system, researchers

1. openly collaborate with different societal stakeholders to further improve the functioning of the food system;
2. deploy a systems approach including the impacts and consequences in the whole food systems in their research, not studying effects and disciplines in isolation;
3. undertake fully data-driven research, in which the data is ubiquitously driving the research, and researcher rapidly plug-n-play different data sources to quantify and estimate effects in a robust way;
4. work impact-based, to place their research in the broader societal context and show what the implications of the research are.

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

Challenges to realise the Vision

Realising the vision by 2030 requires a transition in research to systemic, integrative, multidisciplinary and global approaches. As part of this transition research in agriculture and food needs to embrace *digitalization, transparency and collaboration*.

The agri-food sector relies on a complex science that requires multidisciplinary, multiscale and geolocation-based approaches. This implies a significant amount and variety of agricultural 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, crowd-sourcing, text mining). More generally, natural and societal phenomena are being described by more and more massive data at different scales, from various sources and with different resolutions.

In order to achieve the vision, the ability to share, access and integrate these heterogeneous data is a key issue in order to tackle today's societal challenges, especially in addressing climate change impacts on food security, providing healthy and nutritious food to all, developing sustainable food value chains as well as providing support for local agricultural development. In order to address these growing challenges faced by the global food system, more and more research and innovation is depending on suitable exploitation of data resources and digital technologies: high-technology equipment, high-speed broadband, connected sensors, data sharing and exploration, modeling and coding, intensive simulation, social networks, etc.

The transparency challenge requires the embrace of "Open Science", which engages research institutions in the digital transition for each phase of the knowledge production cycle for innovation (i.e. design of research questions; production, analysis and simulation of data; dissemination of knowledge; knowledge transfer and innovation).

For collaboration it will be required to connect not only researchers amongst themselves, but also researchers with other societal players. The ambition is to achieve an efficient dataflow between Research, Farming and Supply Chain including communication to policy makers to speed up development and innovation.

Therefore, it is necessary to reflect on a common e-science framework and related facilities or e-infrastructures that enable to share and connect data, computing and storage resources, codes and data-mining algorithms, models and ontologies, as well as expertise, efforts and best practices.

Benefits in realising the Vision

Agri-food science and innovation would benefit hugely from a common data ecosystem. Produced and used by diverse stakeholders including academic researchers but also – and of course – farmers, the industry, extension services and – last but not least – citizens, a shared global data space will help build the infrastructures that will propel the agri-food sector forward.

Addressing the challenges of digitalization, transparency and collaboration will generate the following benefits:

- Speeding up the transfer of knowledge: As in all sciences, open science will reduce the time to access to knowledge for farmers or extension services, private companies linked to the agri-food sector and also for education. Access to this knowledge is also a key issue for developing countries.
- Increasing knowledge spill-overs to the economy: Globally, in the agri-food sector, open data helps shape best practices. Transparency around targets, subsidy distribution and pricing, for example, creates incentives which affect the behaviours of producers, regulators and consumers. It also helps the public authorities to make better decisions.
- Addressing global challenges more effectively: When studying climate change and its impact on agriculture and food security, free access to data from all over the world is of great importance. It is therefore strategic for publicly-funded research in agriculture and environment to adopt the "FAIR principles"[7] for data and also open access to publications.
- Promoting citizens' engagement in science and research: As a science of observation, Agricultural sciences should strongly benefit from "citizen science" projects collecting information from farmers which help to better understand ecosystem functioning. In return, farmers and the food processing sector could better manage their production having access to knowledge bases. For the consumers this would result in more food security and safety.

Rules of engagement and Principles

To achieve the vision and overcome the challenges of digitalization, transparency and collaboration, we believe the following rules of engagement and principles will support this transformation:

- Scientific research is enhanced through seamless use not only of research data throughout the scholarly community, but also of data from the farms and the supply chain, as well as environmental NGOs.
- The Food System benefits from easier access to new technological and scientific developments, which will be verified more quickly.
- The Supply Chain has improved information for consumers, retailers and producers.
- We will assure universal data access and promote a data-sharing culture at community level, but we recognise different levels of data openness and we see the FAIR principles as the guide. We recognize and foster diversity. Data, information and knowledge has to be organized in an ecosystem of commons and private. We see the need for regulations to avoid the establishment of data monopolies and we will support the implementation of those regulations.

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

- We emphasize the need to enable researchers to easily explore, integrate and simulate their own data as well as data that have been produced by others within or outside their community and that are complementary in terms of objects of study, scales and disciplines.
- We need to harness the power of data produced by Farmers and by Land observation, which includes Precision Agriculture but is not limited to it. Farms need to become Labs linked to scientific research.
- We strongly believe that data management starts with data production, in the lab, on the field or at the observational level. We aim for a strong collaboration with equipment producers for introducing common data sharing principles and standards.
- We will support interoperability across data sources and agree on standards without reinventing wheels. Standards need to be open and shared. Standards need to be crossdisciplinary and co-defined with communities to ensure their adoption.
- We will build on existing infrastructures within our field as well as generic (i.e. technological) infrastructures, and establish specific infrastructures and services as relevant for agriculture and food science to stimulate collaboration in an open way.
- We will foster distributed efforts and flexible governance for long-term empowerment by and sustainability within the agri-food community. We will help develop appropriate business models for data sharing and related services, especially for our “common goods” such as those supporting semantic interoperability and data discovery.
- We need to develop skills and capacity.
- We want to construct not only a European Community, but an international network. Big Agricultural Sciences are in the Global South, we will link up our initiatives with those of G20 and G77 countries.

5 CONCLUSION

Bibliometric analysis results and the online map need to be considered as tools to engage the community in the development of the roadmap by increasing the visibility of individual members of the community (e.g. as key scientific experts identified in the bibliometric analysis or as key partners of the data & infrastructure ecosystem that can support open agri-food science) and by providing members with knowledge on the community as a whole to foster collaboration and a shared awareness. Thus, increasing the user-friendliness of these knowledge bases and disseminating them broadly and in a targeted manner is key.

Furthermore, the elaboration of the vision paper is a first step towards the roadmap and consists in a powerful tool to gather the community around common priorities and a shared sense of empowerment. The organisation of the 2nd e-ROSA policy workshop and the participation of e-ROSA partners in EOSC-related events (e.g. EOSC Stakeholder Forum and DI4R) have allowed to identify key elements that should be taken into account in the roadmap development (e.g. key priorities, funding instruments and agendas, upcoming key EU-related events). The roadmap should consist in a broad gap analysis that crosses identified challenges and needs with existing service supply both at domain-specific (i.e. services provided within the agri-food community) and generic levels (i.e. services provided by EU generic e-infrastructures). As such, it allows to identify targeted needs for investment and policy support, and specific funding instruments to be used.

More specifically, the content of the roadmap will focus on:

1. Societal and scientific challenges (based on outcomes of the 2nd Stakeholder Workshop)
2. Needs for data & ICT developments (based on outcomes of the 1st and 2nd Stakeholder Workshops)
3. Supply of services (based on outcomes of the 1st Stakeholder Workshop)
4. Urgent needs for development: prioritization of identified gaps (based on outcomes of the 3rd Stakeholder Workshop)
5. Overall implementation agenda: i.e. key milestones, timeline with short-term/long-term objectives, barriers, specific actions to be implemented (based on outcomes of the 3rd Stakeholder Workshop)

In order to prepare a first draft of the roadmap internally, e-ROSA partners will further identify:

- services provided by EU generic e-infrastructures (e.g. by engaging with the EOSC Hub project), and
- the various funding instruments used at national, European and international levels that the agri-food community can rely on in the near future to address urgent needs.

The process of co-design of the roadmap with the community will rely on:

- the co-writing and validation of the roadmap by key community representatives during the 3rd e-ROSA Stakeholder Workshop that will take place in May 2018;
- the endorsement/support of the finalised roadmap by and/or responses from the broader community;
- the discussing of the implementation of the roadmap with key donors and institutions at the end of the project.