eROSA

e-infrastructure Roadmap
for Open Science in Agriculture

D1.2 - Online map of stakeholders & resources (M12)

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

DELIVERABLE NUMBER	D1.2
DELIVERABLE TITLE	Online map of stakeholders and resources
RESPONSIBLE AUTHOR	Panagiotis Zervas (Agroknow)

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

WORKPACKAGE N. | TITLE	WP1 | Ecosystem & Community
WORKPACKAGE LEADER	INRA
DELIVERABLE N. | TITLE	D1.2 | Online map of stakeholders and resources
RESPONSIBLE AUTHOR	Panagiotis Zervas, Agroknow
REPLY TO	pzervas@agroknow.com
DOCUMENT URL	http://www.erosa.aginfra.eu/deliverables
DATE OF DELIVERY (CONTRACTUAL)	31 December 2017 (M12)
DATE OF DELIVERY (SUBMITTED)	22 December 2017 (M12)
VERSION | STATUS	V1.0 | Final
NATURE	Websites, patents filling, etc.
DISSEMINATION LEVEL	PU(Public)
AUTHORS (PARTNER)	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Kostas Kastrantas (Agroknow)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Table of contents, Chapters	31/10/2017	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Kostas Kastrantas (Agroknow)
0.5	Content updates	10/11/2017	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Kostas Kastrantas (Agroknow)
0.8	Corrections-revisions based on technical developments	30/11/2017	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Kostas Kastrantas (Agroknow)
1.0	Updates- Finalization	11/12/2017	Panagiotis Zervas (Agroknow), Nikos Manouselis (Agroknow), Pythagoras Karampiperis (Agroknow), Kostas Kastrantas (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

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. This map is part of the AGINFRA web portal (as it has been foreseen in the eROSA DoA), it is hosted at www.aginfra.eu and it aims at supporting the cartography and visualization of the data ecosystem in agri-food. The technologies used are Apache HTML Server, MySQL database and PHP. As for the content management, the Drupal CMS has been selected.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The eROSA project seeks to build a shared vision of future sustainable e-infrastructures for research and education in agriculture in order to promote Open Science in this field and as such contribute to addressing related societal challenges. In order to achieve this goal, eROSA's main objective is to bring together the relevant scientific communities and stakeholders and engage them in the process of co-development of an ambitious, practical roadmap that provides the basis for the design and implementation of such e-infrastructures in the years to come.

In this context, the eROSA online map will support:

- The documentation of the scientific stakeholders by identifying different institutions, initiatives and groups – with a particular emphasis on stakeholders working on information technologies and computing, and linked to agriculture (e.g. involved in data interoperability, big data in agriculture and related topics).
- The documentation of the research infrastructure facilities and e-infrastructure resources and services that have been deployed for agriculture, at a national, regional or international level (also covering generic e-infrastructures that could be of relevance to agriculture).

With this document, we present the system used for documenting scientific stakeholders and research infrastructure facilities and e-infrastructure resources in an online map. We also present the mapping and visualisation solutions selected as well as an analysis of the entities documented for the 1st release of the online map.

2 THE ONLINE MAP OF STAKEHOLDERS AND RESOURCES

In the following chapter, the functionalities of the online map are described along with the capabilities of the discovery mechanism that allows for searching and browsing through the entities documented on the e-ROSA online map.

2.1 DISCOVERY

URL: http://www.aginfra.eu

By clicking on the magnifying glass on the top right of the eROSA online map, the user is redirected to the page shown in Figure 3. In this page, the visitor can:

1. Search through the use of free text, by typing in the text box in the top of the page,
2. Browse by Agri-Food Discipline by selecting the corresponding icon from the eighteen (18) available disciplines,
3. Browse by Type (i.e. entity), by selecting one of the four icons that represent organizations, data points, facilities and initiatives

Figure 1: eROSA Discovery Page (text search)

As it can be seen in Figure 4, the icons that represent the different Agri-Food disciplines are carefully selected so that they represent the disciplines in a clear and easy to understand way. Also, in the bottom of the page, the different icons that indicate the entities of the map also show the count of each type of entity that is published on the map.

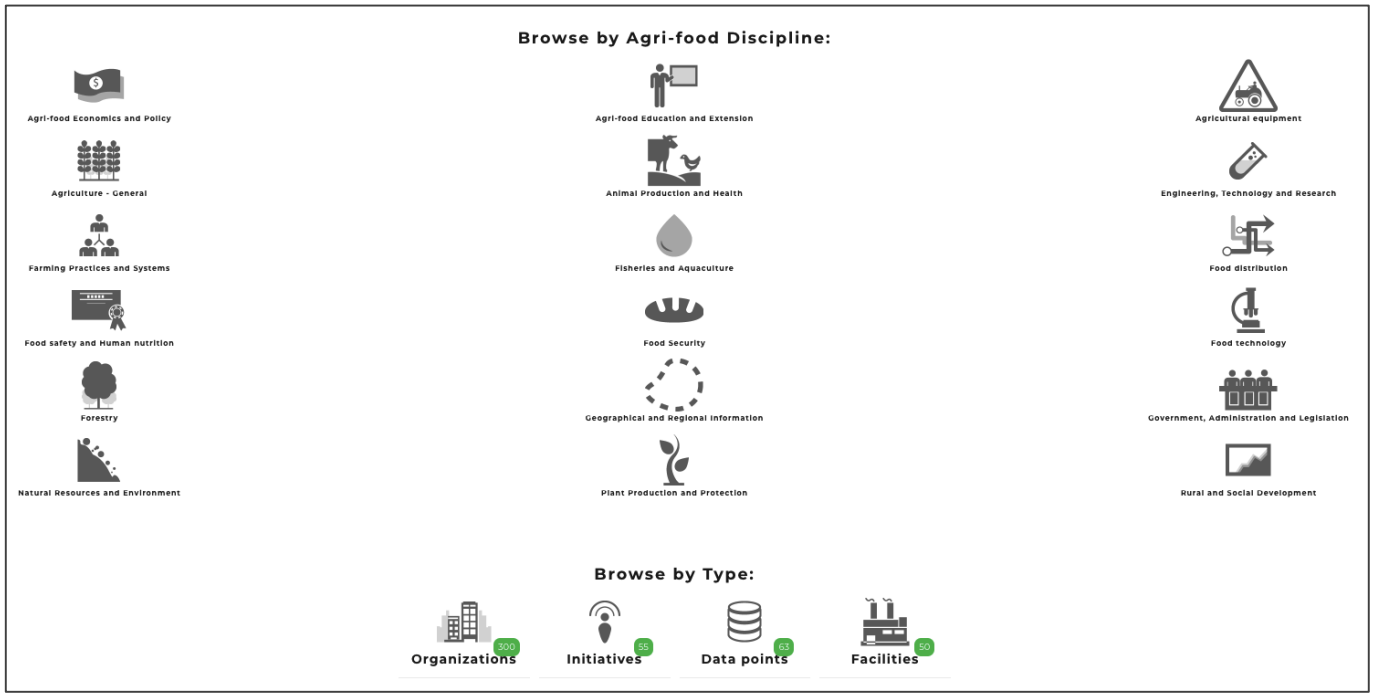

Figure 2: eROSA Discovery Page (browsing)

When the visitor clicks on one of the icons, or inputs a search term in the search box shown in Figure 3, the results' page comes up, as it can be seen in Figure 5, showing the results of the search.

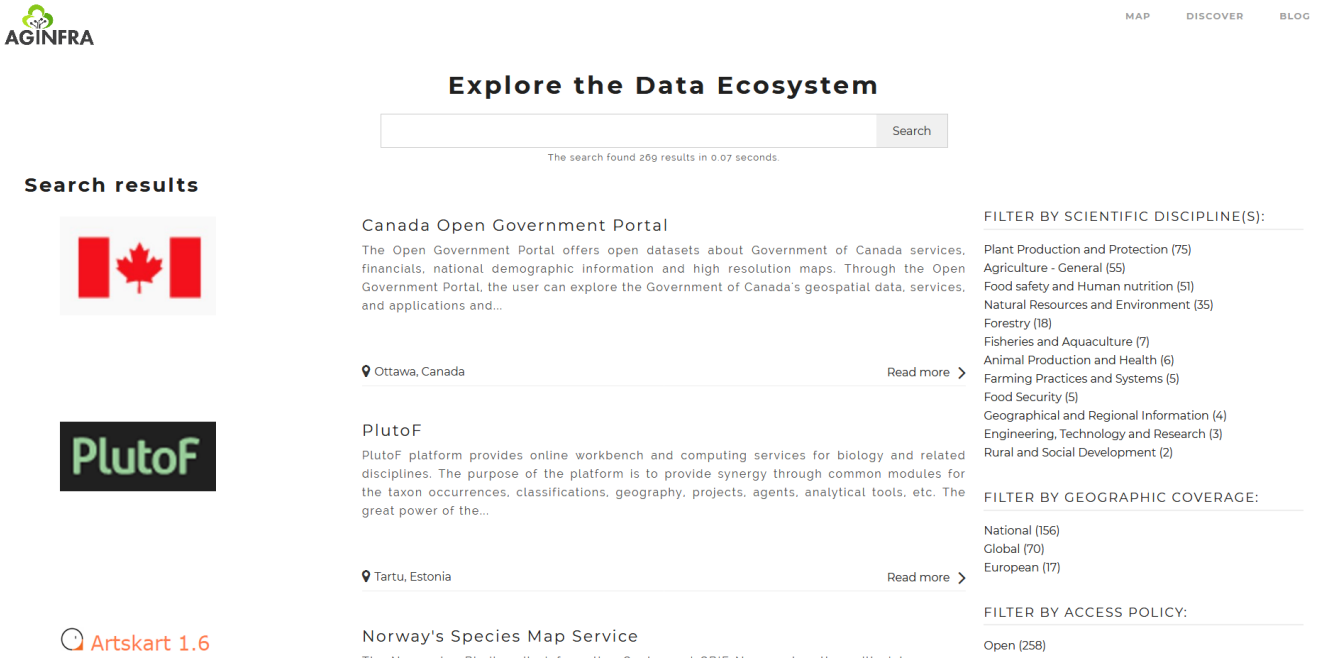

Figure 3: eROSA Search Results

The results' page shows for each search result, a logo, its title, its location and a part of each description so that the visitor can quickly assess if this result is of interest to him/her. By default, the results of each search come up in sets of ten per page. These results can be filtered further by the visitor, based on various search facets. Once the visitor selects a specific result, they are transferred in a page like the one in Figure 6, where the information for this specific entity is depicted.

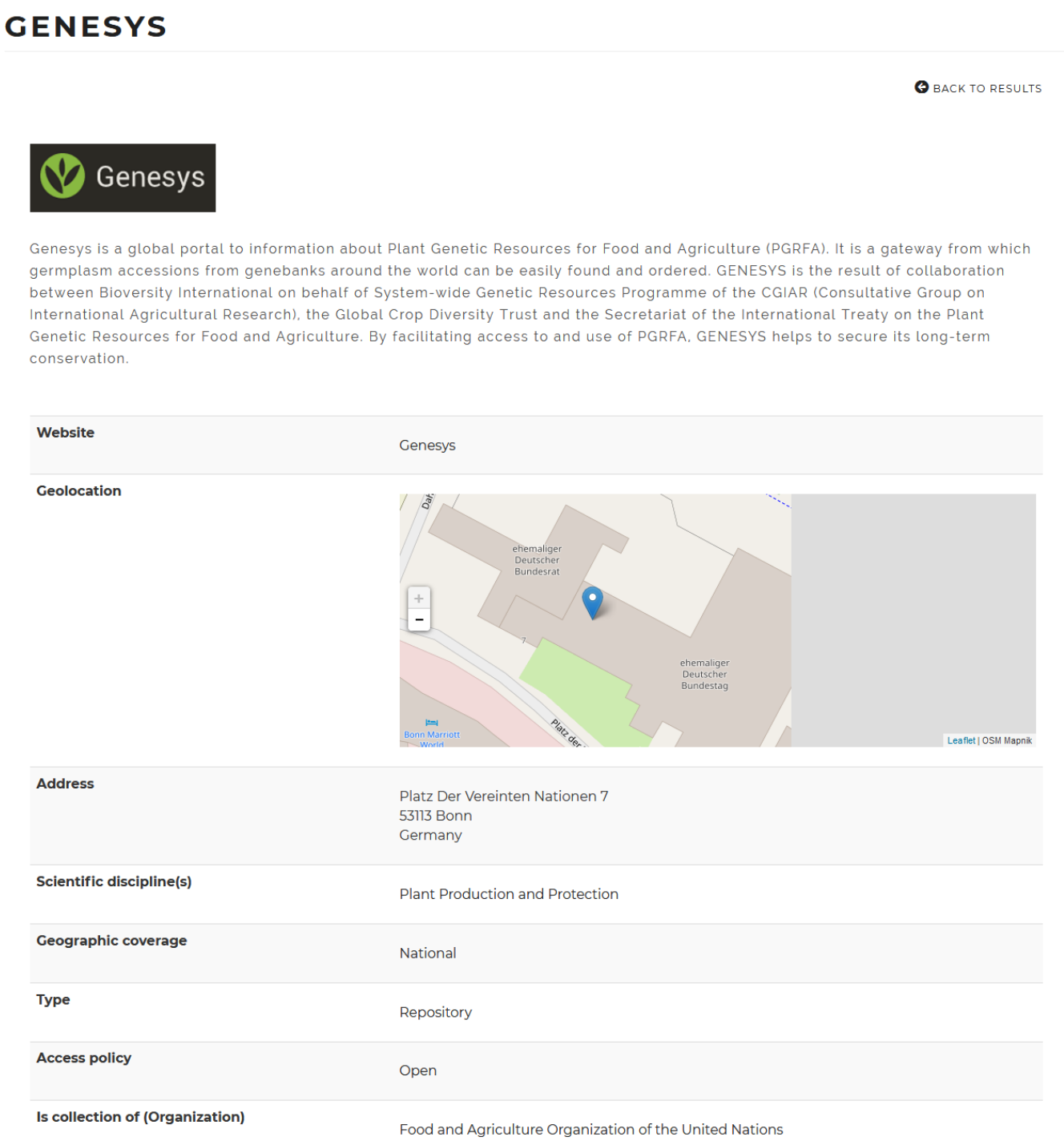

Figure 4: Entity page on eROSA online map

2.2 GEOGRAPHIC MAP

The eROSA online map (http://www.aginfra.eu/map), presents in a user-friendly manner, the eROSA stakeholder communities. Detailed information of the mapping activity that was carried out to populate the online map, can be found in "D1.5 – Synthesis of results & contribution to roadmap (M12)". As soon as the map loads, it shows by default the layer containing the Data Points uploaded on eROSA. On the map, single Data Points are represented by a red circle with transparent fill, whereas aggregations of up to ten Data Points in a specific area (which is also affected by the map zoom level), are represented with a yellow circle with yellow fill. Aggregations of more than ten Data Points are shown with an orange

circle with a light orange fill whereas aggregations of more than 60 Data Points are shown with a red circle with a red fill.

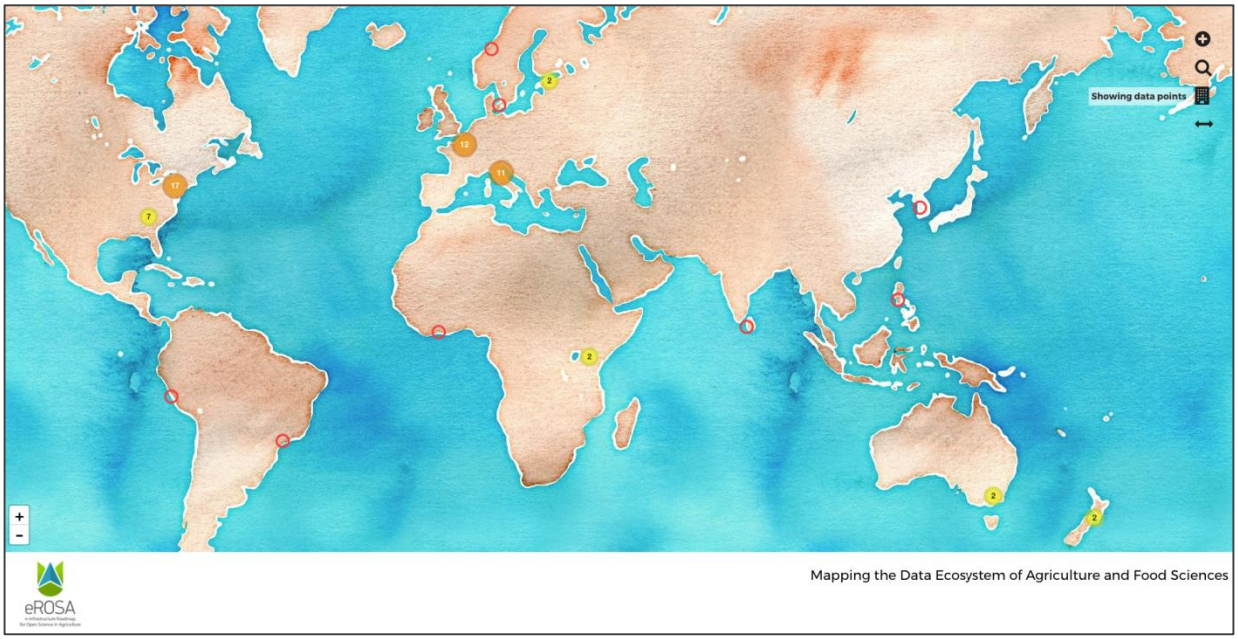

Figure 5: eROSA Online Map

As the map is zoomed in, the depiction of Data Points (or any other entity, that is organizations, facilities and initiatives) for a specific area, changes and the points are re-grouped based on how much the visitor zooms in or out. As an example, in the following table, the same selection of Data Points is shown, in two cases, with a zoom factor of "1" (left) and then with a zoom factor of "2" (right). In the case of the higher zoom (right) it can be seen that the groupings of "12" and "11" Data Points on the left were changed to "11" and "9" as some Data Points belonged to areas that are clearly set apart from the previous grouping now that the map is zoomed in.

Figure 6: Different grouping of Data Points based on map zoom

The visitor can click on the aggregations of entities (Data Points in our example) and the map will automatically zoom in the selected group, showing the entities in a greater zoom level. On the top right of the map, the user can find the map controls from where it is possible to:

- Propose a new entity for inclusion on the eROSA online map (organization, data point, facility or initiative),
- Open the discovery page through which the visitor can search through the entities of the eROSA online map by facets
- Change the type of entity depicted on the map (available views: data points, organizations/initiatives, facilities),
- Show the connections that exist among different entities

The next section will elaborate on the search capabilities that the discovery page mentioned above offers to the visitor of the eROSA online map.

3 ANALYSIS OF THE CONTENT OF THE 1ST RELEASE

In the following paragraphs, we perform an analysis and presentation on the main characteristics of the content that is contained in the 1st release of the online map. This analysis will serve as input to the enrichment of the content for the next release of the online map and will also help to better target the campaigns that will invite relevant stakeholders to review the coverage of the online map and add or revise information, by adding their institution, team or project to it.

3.1 ORGANIZATIONS

For the 1st release of the map, three hundred organisations (300) were identified, originating from a comprehensive online search for relevant stakeholders. In the word cloud below, the most common terms used in the description of the organisations that were retrieved are summarized, providing an indication of the scope of the organisations analysed. We see that the main terms coming up in the word cloud, include “research”, “agricultural/agriculture”, “development”, “institute” and “food”.

Figure 7: Word cloud derived from the Organisations’ descriptions

3.1.1 Organizations per Country

Looking at these organisations, we see that the most represented country is India with a share of 14% in relation to the total number of organisations. China (9.9%) and United States (7%) follow, with Egypt (5.7%), Brazil (4.5%) and Russia (4.1%) following. Completing the top ten with the most represented countries, Peru (2.9%), Canada (2.6%), Italy (2.6%) and Great Britain (2.6%) follow.

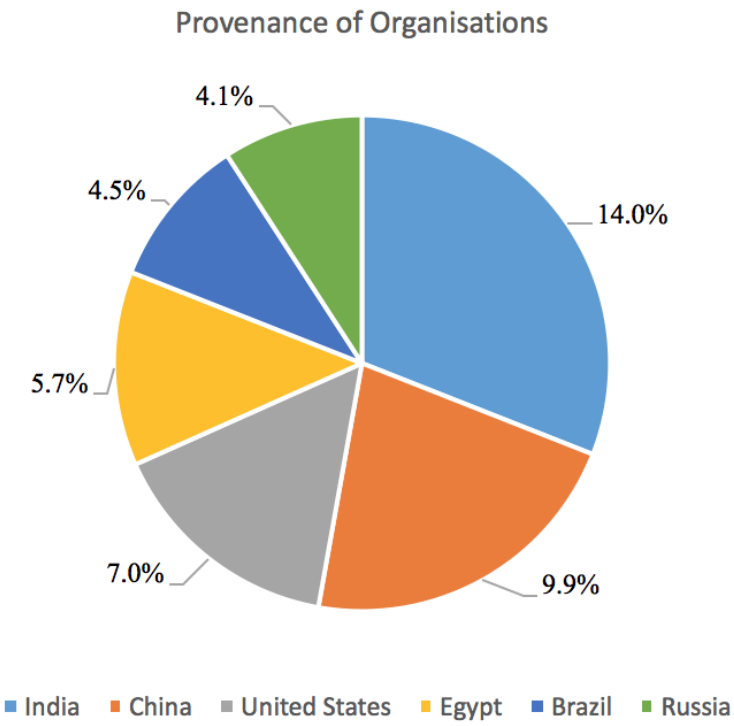

Figure 8: Countries with highest number of Organisations identified

In total, three hundred (300) organisations come from eighty-three (83) distinct countries, which shows that the sample of organisations uploaded on the online map, spreads across the globe, offering a representative sample of organisations for the 1st release of the map.

3.1.2 Organizations per Continent

Looking at the distribution of Organisations across the five continents, we see that, apart from Oceania, all the other continents are equally represented.

Continent	Count	Percentage (%)
Africa	54	18%
America	71	23.7%
Asia	95	31.3%
Europe	77	25.7%
Oceania	4	1.3%
TOTAL	**300**	**100%**

Table 1: Organisations per Continent

This analysis shows that organisations are spread around the different continents with Asia to be the most represented continent. It is expected that from the open call continents with lower number of organizations such as Oceania will be targeted.

3.1.3 Organizations per Scientific Discipline

Looking at the primary scientific discipline of each organisation inserted onto the eROSA online map, the main scientific disciplines represented so far are "*Education and Extension*" (16.9%), "*Engineering, Technology and Research*" (14%), "*Plant Production and Protection*" (14%) and "*Agriculture – General*" (12.7%).

In addition to the previous ones, "Animal Production and Health" (8.9%), "*Natural Resources and Environment*" (5.7%), "*Food safety and Human nutrition*" (5.1%), "*Economics and Policy*" (3.8%) and "*Rural and Social Development*" (3.8%) come off as disciplines that are quite frequent in the organisations represented on the online map thus far.

Scientific Discipline	Count	Percentage (%)
Education and Extension	49	15.6%
Engineering, Technology and Research	41	13.1%
Plant Production and Protection	41	13.1%
Agriculture - General	37	11.8%
Animal Production and Health	26	8.3%
Natural Resources and Environment	19	6.1%
Food safety and Human nutrition	16	5.1%
Economics and Policy	12	3.8%
Rural and Social Development	12	3.8%
Forestry	10	3.2%
Government, Administration and Legislation	9	2.9%
Fisheries and Aquaculture	6	1.9%
Information Management	6	1.9%
Medicine	6	1.9%
Farming Practices and Systems	5	1.6%
Food Security	5	1.6%
TOTAL	**300**	**100%**

Table 2: Organisations per Scientific Discipline

3.2 INITIATIVES

Looking at a word cloud created by the description of the different initiatives identified thus far, we see that the main terms that stand out are “data”, “European”, “open”, “research” and “agriculture”. Other terms that are also of interest although not that common as the previous ones, include “innovation”, “science” and “food”.

Figure 9: Word cloud derived from the Initiatives’ descriptions

3.2.1 Initiative Coverage

Out of the thirty-seven (37) initiatives represented in the first version of the online map, the majority of them (57%) represent initiatives that are covering Europe through their activities whereas there is also a significant number of initiatives that act on an international level (32%). Last but not least, a small subset (11%) represents initiatives that act locally (National initiatives).

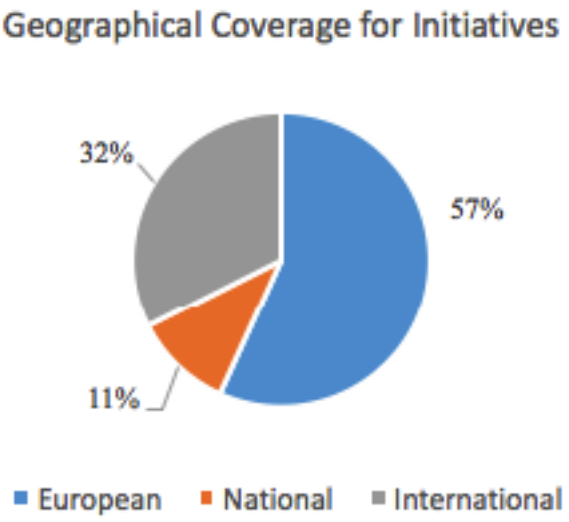

Figure 10: Geographical Coverage for Initiatives

3.2.2 Initiatives per Scientific Discipline

Looking at the primary scientific discipline of each initiative inserted onto the eROSA online map, the main scientific disciplines represented so far are "*Agriculture – General*" (41.7%), "*Information Management*" (25%) and "*Engineering, Technology and Research*" (19.4).

Scientific Discipline	Count	Percentage (%)
Agriculture - General	15	40.5%
Information Management	9	24.3%
Engineering, Technology and Research	7	18.9%
Food security	3	8.1%
Food safety and Human nutrition	2	5.4%
Farming Practices and Systems	1	2.7%
TOTAL	**37**	**100%**

Table 3: Initiatives per Scientific Discipline

3.2.3 Initiatives per Data Science Category

Looking at the Data Science categories, three of them have certainly stood out in regards to the initiatives that are presently described on the eROSA online map. More specifically, "*Governance*" (28.9%), "*Technology adoption*" (15.6%) and "*Data discovery & access*" (13.3%) are the main Data Science categories retrieved. Despite the fact that the remaining categories are lower in numbers, we see that all of them are represented which indicates that the sample of initiatives currently on the online map is quite diverse and well selected.

Data Science Category	Count	Percentage (%)
Governance	13	28.9%
Technology adoption	7	15.6%
Data discovery & access	6	13.3%
Community building	3	6.7%
Modelling, statistics & simulation	3	6.7%
Technical data interoperability	3	6.7%
Semantics	2	4.4%
e-infrastructure	2	4.4%
Data ownership	2	4.4%
Computation	1	2.2%
Workflows	1	2.2%
Support for decision-making	1	2.2%
Data publication	1	2.2%

Table 4: Initiatives per Data Science Category

3.3 DATA POINTS

Looking at a word cloud created by the description of the different data points identified thus far, we see that the main terms that stand out are “data”, “information”, “research” and “food”. Other terms that are also of interest although not that common as the previous ones, include “database”, “resources” and “development”.

Figure 11: Word cloud derived from the Data Points’ descriptions

3.3.1 Data Point Type

Related to the Type of the data points published on the eROSA online map, their majority (87.2%) is either a “*Repository*” (47.8%) or a “*Catalogue*” (39.4%). A limited number of “*Aggregators*” of data points were identified (8.4%) whereas only 12 (4.4%) data points are offered as a “*Set*”.

Data Point Type	Count	Percentage (%)
Repository	128	47.6%
Catalogue	106	39.4%
Aggregator	23	8.6%
Set	12	4.5%
TOTAL	**269**	**100%**

Table 5: Data Points per Type

3.3.2 Data Point Coverage

Looking at the geographical coverage of the data points on the eROSA online map, we see that 58% of the data points contain data of a national scope whereas 36% of them relate to data that are international. Only 6% of the data points offer data that cover the European region as a whole.

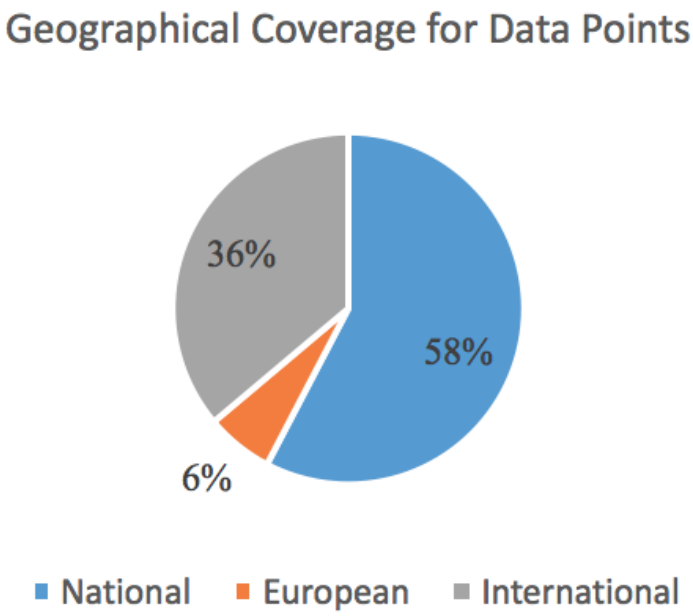

Figure 12: Geographical Coverage for Data Points

3.3.3 Data Point Access

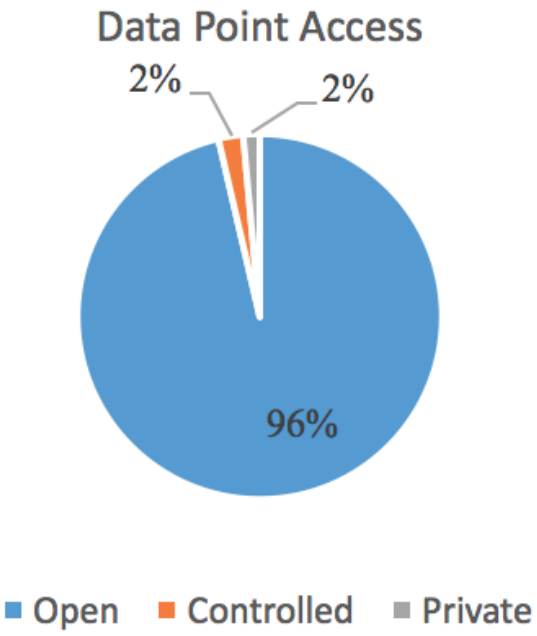

Figure 13: Access for Data Points

The high majority of the data points have an open access policy. However, there are also cases with data points that adopt other access policies such as controlled or private.

3.3.4 Data Points per Scientific Discipline

Examining the Data Points, it is clear that the sample uploaded on the eROSA online map is quite diverse. The most common category of Data Point is the ones related to "*Plant Production and Protection*" (27.7%), followed by "*Agriculture-General*" (20.8%) and "*Food Safety and Human Nutrition*" (18.6%). "*Natural Resources and Environment*" (12.8%) is the fourth most represented category.

Following the aforementioned, "*Forestry*" (6.6%), "*Animal Production and Health*" (2.6%) and "*Fisheries and Aquaculture*" (2.6%) are also found in a significant number of Data Points.

Scientific Discipline	Count	Percentage (%)
Plant Production and Protection	76	27.7%
Agriculture - General	57	20.8%
Food safety and Human nutrition	51	18.6%
Natural Resources and Environment	35	12.8%
Forestry	18	6.6%
Animal Production and Health	7	2.6%
Fisheries and Aquaculture	7	2.6%
Farming Practices and Systems	5	1.8%
Food Security	3	1.1%
Engineering, Technology and Research	3	1.1%
Economics and Policy	2	0.7%
Geographical and Regional Information	2	0.7%
Rural and Social Development	2	0.7%
Medicine	1	0.4%
TOTAL	**269**	**100%**

Table 6: Data Points per Scientific Discipline

3.4 FACILITIES

Looking at a word cloud created by the description of the different facilities identified thus far, we see that the main terms that stand out are “data”, “climate”, “infrastructure”, “research” and “access”. Other terms that are also of interest although not that common as the previous ones, include “services”, “scientific” and “information”.

Figure 14: Word cloud derived from the Facilities’ descriptions

3.4.1 Facility Coverage

Out of the 31 facilities represented on the eROSA online map, the majority of them (55%) concern European facilities whereas a large number relates to international facilities (39%). Only a small percentage of them (6%) relates to facilities that operate on a national level.

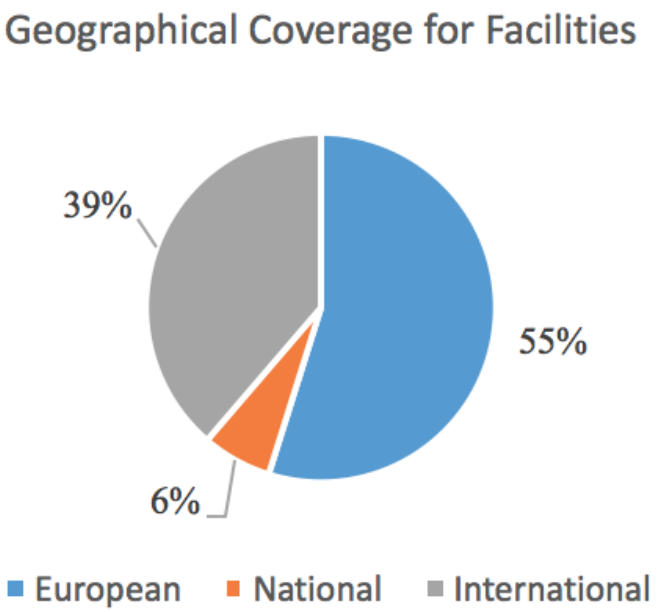

Figure 15: Geographical Coverage for Facilities

3.4.2 Development Stage of Facilities

When it comes to the development stage of the facilities on the eROSA online map, most of them (81%) represent an operational service whereas only 19% of them represent a facility that is being developed in the context of a project.

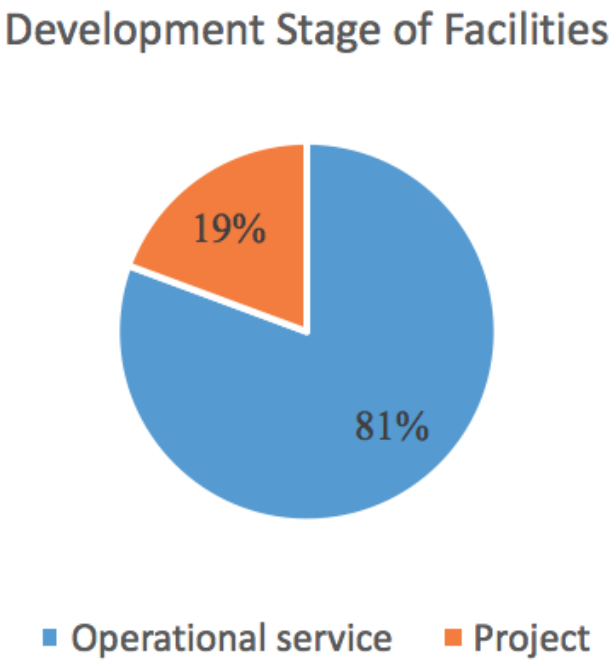

Figure 16: Development Stage of Facilities

3.4.3 Facility Access Policy

As far as the access policy is concerned, most of the facilities identified are open (74%) whereas 19% of them offer controlled access and 7% of them are private.

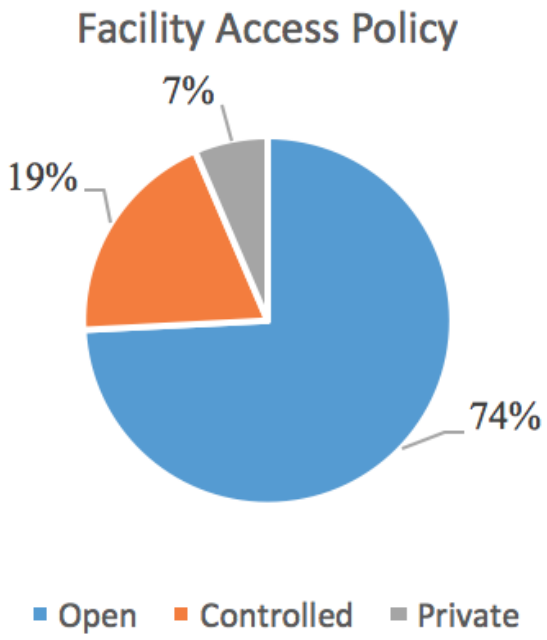

Figure 17: Development Stage of Facilities

3.4.4 Facilities per Scientific Discipline

The majority of facilities uploaded on the eROSA online map, are split among the disciplines of "*Agriculture*" (29%), "*Engineering, Technology & Research*" (29%) and "*Information Management*" (22.7%). Another 16.1% of them are related to "*Natural Resources and Environment*" and only one (1) facility is related to "*Plant Production and Protection*".

Scientific Discipline	Count	Percentage (%)
Agriculture - General	9	29%
Engineering, Technology & Research	9	29%
Information Management	7	22.7%
Natural Resources and Environment	5	16.1%
Plant Production and Protection	1	3.2%
Total	**31**	100%

Table 7: Facilities per Scientific Discipline

3.4.5 Facilities per Data Science Category

In regards to the Data Science categories of the facilities currently on the eROSA online map, their majority (41.9%) is related to "*Data discovery & access*" and 25.9% are related to "*e-Infrastructures*". "*Semantics*" (6.5%) and "*Data Integration*" (6.5%) are the third most "popular" Data Science categories

whereas the remaining are represented by one (1) case (3.2%), namely "*Technical data interoperability*", "*Community building*", "*Data production*", "*Data visualisation*", "*Modelling, statistics & simulation*" and "*Data publication*".

Data Science Category	Count	Percentage (%)
Data discovery & access	13	41.9%
e-infrastructure	8	25.8%
Semantics	2	6.5%
Data integration	2	6.5%
Technical data interoperability	1	3.2%
Community building	1	3.2%
Data production	1	3.2%
Data visualisation	1	3.2%
Modelling, statistics & simulation	1	3.2%
Data publication	1	3.2%
Total	**31**	**100%**

Table 8: Facilities per Data Science Category

3.5 SCIENTIFIC DISCIPLINES FOR ALL ENTITIES

Looking at the Scientific Disciplines across all the different entities, it is evident that most of the entities uploaded on the eROSA online map are related to either "*Agriculture-General*" (20.4%) or to "*Plant Production and Protection*" (18%). Other scientific disciplines that were also represented in the sample of entities were "*Natural Resources and Environment*" (12.1%), "*Food safety and Human nutrition*" (10.3%), "*Engineering, Technology and Research*" (8%) and "*Education and Extension*" (7.9%).

Scientific Discipline	Count	Percentage (%)
Agriculture - General	137	20.4%
Plant Production and Protection	121	18.0%
Natural Resources and Environment	81	12.1%
Food safety and Human nutrition	69	10.3%
Engineering, Technology and Research	54	8.0%
Education and Extension	53	7.9%
Animal Production and Health	35	5.2%
Forestry	28	4.2%
Information Management	15	2.2%
Economics and Policy	14	2.1%
Rural and Social Development	14	2.1%
Fisheries and Aquaculture	13	1.9%

Farming Practices and Systems	11	1.6%
Food Security	10	1.5%

Table 9: Scientific Disciplines for all entities

4 CONCLUSIONS AND NEXT STEPS

Looking at the analysis of the entities collected thus far onto the eROSA online map, the following conclusions can be made:

- Related to the organizations identified so far, the sample of 300 organizations is well-balanced and representative on the level of continents and countries. Looking also at the main disciplines that these organizations represent, it seems that all disciplines currently in the eROSA data model are represented in the selected sample. Through the Open Call (as described in "D1.5 – Synthesis of results & contribution to roadmap (M12)") it is expected to extend the number of organizations targeting organizations from all continents and countries.
- As far as the initiatives are concerned, most of them are targeted at a European or International level which showcases the fact that major initiatives with high impact are selected in this version of the map, including all the major ongoing initiatives on open science in agriculture. Through the Open Call it is expected to increase also the number of initiatives covering also national initiatives.
- In relation to the data points, the majority of the selected ones concern either aggregators or repositories, which are more technologically mature and offer a wide range of data. In addition to that, most of the identified data points are offered only on a national level which enhances the necessity of their presence on the eROSA online map, to allow for their visibility and potential reuse on applications and services. Through the Open Call it is expected to increase also the number of data points targeting data points with European or International coverage.
- On the same topic, the large majority of the data points identified are open (96%), which allows for their use and reuse and they also cover all the agricultural disciplines which makes the map relevant to a wider range of interested stakeholders of the eROSA communities.
- As far as the facilities identified are concerned, their majority are offering services on a European or International level, a fact that enhances their importance for the community as a whole. Also, their majority (81%) represents operational services that are ready for use while at the same time, they are open (74%) to anyone that wishes to use them. On top of that, the facilities identified are a representative sample in terms of the data science categories defined by the eROSA data model. Through the Open Call it is expected to increase the number of facilities targeting all data science categories defined by the eROSA data model; as presented in D1.4/D1.5 "Synthesis of results & contribution to roadmap (M6/M12)".
- Looking at the scientific disciplines across the different entities, it is evident that all of them are sufficiently represented in the sample of almost 650 entities identified so far on the eROSA online map.