eROSA
e-infrastructure Roadmap
for Open Science in Agriculture

Synthesis of results & contribution to roadmap (M6)

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

DELIVERABLE NUMBER	D1.4
DELIVERABLE TITLE	Synthesis of results & contribution to roadmap (M6)
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	Mrs. Pilar Ocon-Garces

WORKPACKAGE N. | TITLE	WP1 | Ecosystem and Community
WORKPACKAGE LEADER	INRA
DELIVERABLE N. | TITLE	D1.4 | Synthesis of results & contribution to roadmap (M6)
RESPONSIBLE AUTHOR	Madeleine Huber (INRA)
REPLY TO	madeleine.huber@inra.fr
DOCUMENT URL	http://www.erosa.aginfra.eu/sites/erosa_deliverables/D1.4.pdf
DATE OF DELIVERY (CONTRACTUAL)	30 June 2017 (M6)
DATE OF DELIVERY (SUBMITTED)	14 September 2017 (M9)
VERSION | STATUS	V2 | Final
NATURE	R(Report)
DISSEMINATION LEVEL	PU(Public)
AUTHORS (PARTNER)	Esther DZALE (INRA), Madeleine HUBER (INRA), Rob LOKERS (WUR), Panagiotis ZERVAS (Agroknow)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
1	Pre-final version for comment	20/07/2017	Esther DZALE (INRA), Madeleine HUBER (INRA), Panagiotis ZERVAS (Agroknow)
2	Final version	14/09/2017	Esther DZALE (INRA), Madeleine HUBER (INRA), Rob LOKERS (WUR), 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

ACRONYMS LIST

AAI	Authentication & Authorization Infrastructure
e-IRG	e-Infrastructure Reflection Group
EC DG	European Commission Directorate-General
EOSC	European Open Science Cloud
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
ICT	Information and Communication Technologies
RDA	Research Data Alliance
VCC	Virtual Competency Centre
VO	Virtual Organization
VRE	Virtual Research Environment
WP	Work Package

EXECUTIVE SUMMARY

A first set of tasks have been carried out under Work Package 1 “Ecosystem & Community” in order to set the stage for e-ROSA’s community-building and roadmap development activity. These tasks include:

a. The collection and analysis of e-infrastructure plans and roadmaps (Task 1.1.2);

b. The elaboration of an online map of the e-ROSA stakeholder community (Tasks 1.1.3, 1.2 & 1.3).

This deliverable presents the progress achieved so far

- The analysis of existing e-infrastructures and roadmaps (Task 1.1.2) has allowed to provide food for thought on an in-depth characterisation of e-infrastructures that can support the envisioning of future e-infrastructures in the scope of e-ROSA, as well as on the relevant content for the future e-ROSA e-infrastructure roadmap;
- The mapping activity carried out under e-ROSA (Tasks 1.1.3, 1.2 & 1.3) relies on a mapping methodology that has been developed in the first months of the project and pre-validated for the elaboration of this deliverable.

Section 1 describes the overall methodology adopted under WP1 and the objectives of the deliverable.

Section 2 provides food for thought on the strategic envisioning and technical design of a future e-infrastructure for agri-food research.

Section 3 presents the adopted methodology to map relevant stakeholders, initiatives and infrastructures in the scope of e-ROSA in view of assessing the current landscape and initiate a community-building process supporting the e-ROSA roadmap elaboration.

Section 4 provides first elements to consider in the co-elaboration of the e-ROSA roadmap.

The deliverable will be updated twice during the lifetime of the project (Deliverable 1.5 in six months and Deliverable 1.6 at the end of the project) in order to monitor and disseminate further achievements under WP1.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 BACKGROUND

The e-ROSA 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, e-ROSA's main objective is to bring together the relevant scientific communities and stakeholders and engage them in the process of co-elaboration of an ambitious, practical roadmap that provides the basis for the design and implementation of such e-infrastructures in the years to come.

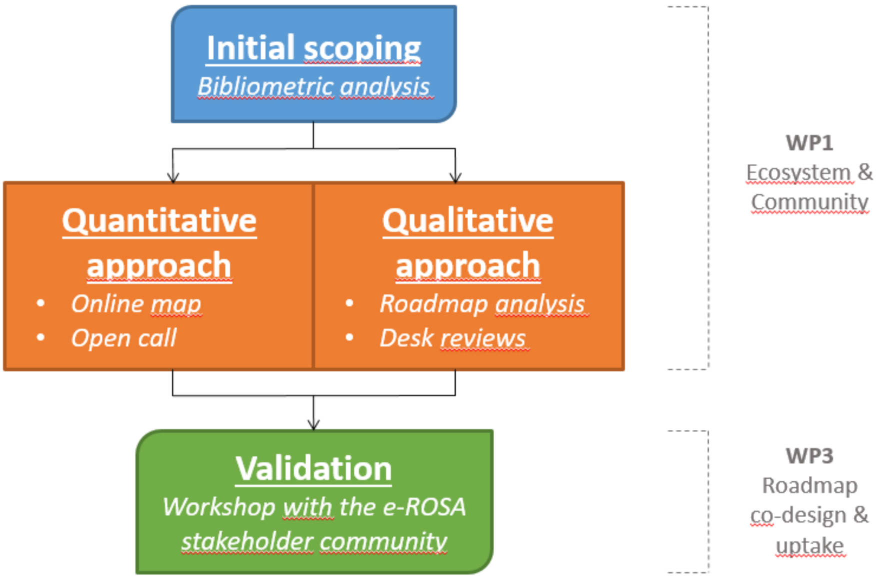
The initial need for scoping and precisely mapping the e-ROSA-related stakeholder community is addressed in the context of Work Package 1 "Ecosystem & Community". The latter aims at:

- providing a detailed, comprehensive overview of the current landscape related to data science in agriculture;
- identifying main stakeholders and existing resources within the e-ROSA scope that can support and benefit from the e-ROSA community-building and roadmap elaboration process;
- identifying and analysing existing e-infrastructures and roadmaps (in the scope of e-ROSA or in other fields) in order to provide food for thought for the envisioning of future e-infrastructures in the scope of e-ROSA.

Figure 1 below indicates the overall method adopted under WP1, which relies on the combination of an initial internal work with a collaborative process that allows the involvement of stakeholders themselves in the mapping exercise:

1) As a first step, we conducted a bibliometric analysis in order to test a first source of information (i.e. scientific publications) to identify who works within our scope on which specific topic(s). This allowed us to understand how the research institutions in our domain tackle the data challenge and to provide a starting basis for the scoping of the e-ROSA community. The first results of this study are available online and will be updated later thanks to a refinement of the methodology used for the bibliometric analysis.
2) The second stage of WP1 relies on an overall methodology that consists in the combination of a more quantitative approach with a more qualitative approach:
    - Quantitative approach: this consists in the elaboration of an online map of e-ROSA stakeholders developed by Agroknow through 1) a first internal inventory 2) an open call to the e-ROSA community to complete the initial inventory;
    - Qualitative approach: this consists in 1) an internal analysis of existing e-infrastructures and roadmaps to identify inspiring examples for the e-infrastructures envisioned by e-ROSA and 2) desk reviews to validate our analysis.
3) The final step for validating the mapped stakeholders under WP1 relies on a collaborative approach as it consists in a workshop with identified stakeholders (under WP3 "Roadmap co-Design & Uptake, 6-7 July 2017): this workshop seeks to promote a community-building process towards the elaboration of a common vision for future e-infrastructures in agri-food research.

Figure 1. Overall approach under Work Package 1 "Ecosystem & Community"



1.2 OBJECTIVE OF THE DELIVERABLE

A first set of tasks have been carried out under Work Package 1 "Ecosystem & Community" in order to set the stage for e-ROSA's community-building and roadmap development activity. These tasks include:

a. The collection and analysis of e-infrastructure plans and roadmaps (Task 1.1.2 "Collect & synthesize stakeholder e-infrastructure plans & roadmaps");

b. The elaboration of an online map of the e-ROSA stakeholder community (Tasks 1.1.3 "Mapping services & knowledge base within AGINFRA portal", 1.2 "Scientific community map" & 1.3 "Research Infrastructure ecosystem map").

This deliverable presents the progress achieved so far under these two tasks:

- The analysis of existing e-infrastructures and roadmaps has allowed to provide food for thought on an in-depth characterisation of e-infrastructures that can support the envisioning of a future e-infrastructures in the context of e-ROSA (see Section 2), as well as on the relevant content for the future e-ROSA e-infrastructure roadmap (see Section 4);
- The mapping activity carried out under e-ROSA relies on a mapping methodology that has been developed in the first months of the project and pre-validated for the elaboration of this deliverable (see Section 3).

The content of this deliverable will be updated twice during the lifetime of the project (D1.5 & 1.6) in order to monitor further achievements under WP1, i.e. regarding:

- the mapping of the e-ROSA stakeholder community including scientific communities, existing generic and domain-specific e-infrastructures and data resources within the scope of agri-food;
- the collection and analysis of further information that can feed into the potential e-infrastructure design and implementation roadmap in line with the EOSC (European Open Science Cloud) vision and implementation.

2 E-INFRASTRUCTURE DEFINITION AND ANALYSIS OF EXISTING E-INFRASTRUCTURES

2.1 DEFINING AN E-INFRASTRUCTURE

2.1.1 General definition approach

The term "e-Infrastructures" came into use with the progress of distributed computing and was rapidly applied to all infrastructures that delivered ICT services.

In 2013, the e-Infrastructure Reflection Group (e-IRG) defined the e-Infrastructure Commons as a "framework for an easy and cost-effective shared use of distributed electronic resources for research and innovation across Europe and beyond".[1] Three different functions of the commons have been identified: (i) A platform for coordination of the services building the Commons; (ii) A set of sustainable and interoperable e-Infrastructure services within the Commons with the possibility for user communities to select the services that fulfil their requirements; (iii) Implementation of innovation projects with the aim for constantly evolving the e-Infrastructures in order to meet the needs of user communities as they evolve, and to take advantage of relevant new technologies when they are mature enough.

The e-IRG identifies three pillars for an e-infrastructure: (i) the technical resources which can be provided, (ii) the community and (iii) the governance. This structured vision is entirely in line with the definition of an e-infrastructure that was elaborated by e-ROSA partners during the Kick-Off meeting of the project (see Section 2.1.2 below) and that highlights the four key components of an e-infrastructure: 1) community, 2) governance and business model, 3) technical backbone, and 4) services.

2.1.2 e-ROSA's e-infrastructure definition

A shared definition of the concept of "e-infrastructure" is key for the project e-ROSA. According to the European Commission, an e-infrastructure is usually described as:

- a combination of digital technologies (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organisational structures needed to manage them.
- They are keys in future development of research infrastructures, as activities go increasingly "online" and produce vast amounts of data and are at the heart of the Digital Agenda of the European Commission to support Open Science and the link between researchers, citizens or private companies.
- They are becoming a part of the research infrastructure.

Whereas these definitions are shared by the members of the e-ROSA project, it is strongly felt that the definition has to be both broadened and specified.

As general considerations the infrastructure should:

- Make sure that connected scientific devices, instruments, software can become discoverable & share the data that they generate & manage;
- Make sure that it provides the storage, computing and connectivity facilities that will make data & device discoverability possible;
- Make sure that it provides the core operational services & adheres to standards that will ensure that different systems may talk to each other & exchange data;

[1] e-IRG Roadmap 2016

- Make sure that the devices, data & services that join are entitled to do so & follow the minimum agreed set of standards required (rules of engagement);
- Make sure that it is up & running (almost) 24/7 providing an agreed quality of service through an agreed SLA.

Infrastructure/s has the following elements:

1) it is based on and useful for a community;
2) it needs governance and a business model for sustainability;
3) it needs a technical backbone, that easily can host a variety of services;
4) it comprises a variety of services, useful for the community, interoperable, but not necessarily connected or dependent from each other.

1. **Community**

Infrastructures are generally community based. A city has an infrastructure, the same goes for a region or a country. Infrastructures are strongly or loosely coupled. Even within an infrastructure there will be networks of institutions with more or less binding collaboration agreements. These collaboration agreements might contain:

- Agreement about the components of the community, science and research, education, private sector, public institutions
- Shared values, i.e. FAIR principles for data
- Governance for data (and broader, including all research objects such as codes, software and workflows): stewardship is often key to use and reuse of data and research objects in general
- Knowledge exchange mechanisms, online and offline
- Structures for maintaining skills, support and capacity development
- Agreements about scaling operations for tasks that could not be executed singly by members of the community
- Agreement about shared services for efficiency gains
- Different categories of stakeholders: Operators of the components of the e-infra - users: scientific users, education, private sector - institutions
- Efficient communication structures and gathering of communities

2. **Governance and Business Model**

Without agreed governance, a clear business model and property rights no infrastructure is sustainable. Governance and business model need to clarify the roles of the different components of the community and the revenue streams that can sustain the infrastructure

- Centralize as much as necessary but decentralize as much as possible with federated structures
- Maximum transparency about rules and operations with a set of agreed charter documents
- Clear licensing schemes that support trusted sharing of data
- Preferring self-sustained operations to operations that are dependent from continuous external input (self-sustained operations are operations that are offered from partners in the network or which create the revenue that they need for functioning)
- Efficient mechanisms to access external funding for investments

3. **Technical Backbone**

Like a traffic infrastructure needs roads, railways, shipyards and car factories as a basic condition, a research infrastructure has similar needs, which can be divided into basic network services and machines to execute infrastructure operations. The most important points are:

- High speed Cloud and Grid connectivity
- Computing centers and distributed computing power with sufficient capacity for analysis, storage and back up
- Effective connections to Laboratory and other research equipment
- Effective connections to Earth Observation facilities, like remote sensing and satellites
- Facilities that gather data from field operations

4. <u>Services</u>

Services have to facilitate a Web environment for the creation, sharing and publication of FAIR agricultural data as well as the tools and methods to process them and transform them into actionable knowledge that will enable to address the big agricultural challenges. Services within an Infrastructure include the entire range of from production to elaboration, storage, analysis and archive of data. They constitute a sustainable virtual working environment consisting of e-/webservices (data, computing, storage, processing, analytics, etc.) that supports full, data-intensive work processes for end-users in a specific domain or specific type of work.

Services within an Infrastructure can be offered by partners as part of their business operations or by facilities maintained by the community with a business and sustainability model for any service to function:

- Data channels to production services like gene and protein sequencing, chemical analysis, observation data production...
- Big scientific data workflows (e.g., Galaxy, Taverna)
- Data analysis Services
- Semantic Support Services for Annotation and Analysis of data (vocabulary and ontology servers)
- Data Discovery services, like registries for data sets
- Data Storage and Archival Services
- Services to research communities (extension to non-academic players for an open science development)
- Capacity Development Services with Guidelines, courses….

2.1.3 Generic and domain-specific e-infrastructures

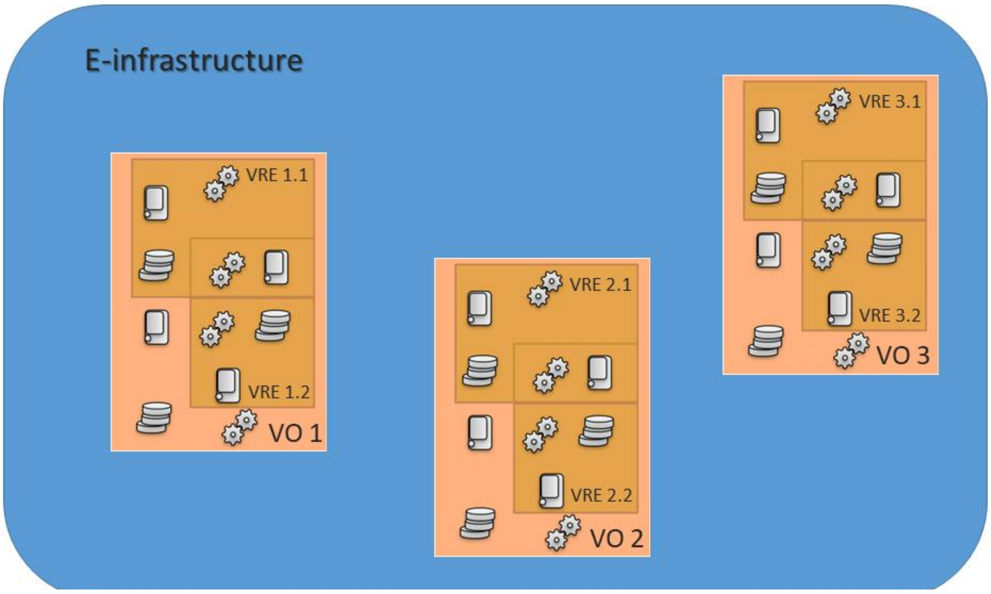
The existing e-infrastructures can be distinguished according to their genericity / specificity. Generic e-infrastructures are well established in general and provide production services which can be used as building blocks for domain-specific e-infrastructures. On their side, domain-specific e-infrastructure are driven by domain-specific use cases and tend to build on existing generic e-infrastructures as much as possible.

Domain-specific e-infrastructures have in common the need for multidisciplinary collaboration at the local, national, regional and international levels to process and analyse large amounts of heterogeneous and distributed information. They usually serve many different communities through Virtual Research Environments (VREs)[2] or Virtual Organizations (VOs) worksites[3]. A VO structures a particular research community with a set of particular requirements. Each VO has a worksite composed of (i) tools/services that are generic to all the VOs such as wikis, blogs and videoconference applications, and (ii) tools/services that are specific to the VO in question. According to the D4Science project, VREs consist in distributed and cloud based environments created dynamically with a set of the e-infrastructure's resources, for a set of

[2] An example of VRE is iMarine: http://www.i-marine.eu/Pages/Home.aspx
[3] An example of VO is See-Grid: http://www.see-grid.org/

users, via interfaces for a limited timeframe at little or no cost for the providers of the e-infrastructure. One or many VREs can be created for a given VO as depicted in Figure 2.

Figure 2. Relationship between VO and VREs



Source: Manzi, A. & Pagano, P. (2011), E-Infrastructure Integration with gCube

2.2 ANALYSIS OF THE STRATEGIC APPROACH OF DOMAIN-SPECIFIC E-INFRASTRUCTURES

In the following section, we have selected existing European domain-specific e-infrastructures that have already matured enough that they can provide food for thought for the envisioning of future e-infrastructures in the field of agri-food in the context of e-ROSA.

Three examples are analysed: ELIXIR, DARIAH and iMarine. The European e-infrastructures ELIXIR and DARIAH, specialised respectively in the field of life science and bioinformatics on the one hand, and Arts and Humanities on the other hand, are both ESFRI (European Strategy Forum on Research Infrastructures) Landmarks, which means that they are already well advanced in their implementation process (see Annex 1). Our third example focuses on iMarine, a European e-infrastructure in the marine area that generates specialised Virtual Research Environments (VRE) based on the generic technology and services provided by the generic VRE-enabling e-infrastructure D4Science.

The table below (Table 1) provides a comparison of these three e-infrastructures in order to identify similarities or differences in their overall approach, including their vision, their definition of an e-infrastructure, their objectives, their organisation and governance, etc.

Table 1. Comparison of the strategic approach of ELIXIR, DARIAH and iMarine

	ELIXIR	DARIAH	iMarine
Vision for the future	Handling massive, complex data in life science Focus on European research	Seize the opportunity of digital advancements to serve research in arts and humanities Focus on European research	Address societal challenges linked to the marine environment Stakeholder-oriented (not only European research, also policy-making, private sector, etc.)
Definition approach for their e-infrastructure	• **Federation/coordination role** • Distributed resources, communities and efforts (i.e. national nodes) • Focus on data (esp. interoperability, integration) • Technical approach (i.e. technical components of the e-infrastructure: Platforms on Data, Compute, Tools, Interoperability – Training as a transversal activity)	• **Teaching/supporting/training role & dissemination of research results** • Distributed resources, decentralisation • Focus on digital sharing of know-how and knowledge (vs. raw data) • Learning approach (Virtual Competency Centres for the e-infrastructure development, sharing of knowledge/methods, sharing of scholarly content, and impact)	• **Service-providing role** • Selection of specific resources: focused on VRE development (multi-layer approach with minimal resources) with a high degree of integration of existing resources • Focus on societal issues, policy-making, decision-making tools • Challenge-driven approach (VREs dedicated to specific use cases)
Strategic objectives	• Sustained, secure and interoperable data • Best practice and standards • Highly skilled workforce • Technical services • Expanding partnerships	• Build a virtual bridge between different resources, initiatives and infrastructures across Europe • Help EU countries establish their own arts and humanities research infrastructures, capacity building for national digital arts and humanities competence • Disseminate digital methods and research activities/knowledge in arts and humanities based on the trusted provision of digital resources	• Accelerate data discovery, exchange, harmonization, and analysis, for a variety of stakeholder communities • Support capacity building in interdisciplinary research communities actively involved in increasing scientific knowledge about resource overexploitation, degraded environment and ecosystem • Providing a more solid ground for informed advice to competent authorities • Enlarge the spectrum of growth opportunities as addressed by the Blue Growth Societal Challenge
Added value of their e-infrastructure	• Transnational coordination • EU representation (at international and national levels) • Formalise and sustain resources in a distributed manner • Support local communities through this	• Enhance knowledge on current state-of-the-art • Integration of research activities and knowledge, embed research activities in the infrastructure: directly supports projects and initiatives, which rely on the DARIAH	• User/community-oriented • User-friendly (VRE interface and seamless integration of resources) • Connection of various e-infrastructure layers (data, compute, etc.) • Challenge-driven, potential for direct impact

	distributed model • High quality, trusted, standardised data (through local expertise) • Critical mass, scaling-up of resources • Changing research habits regarding data stewardship	infrastructure • Transnational coordination of national infrastructures: need for EU approach in such a field, sharing of/open access to knowledge and services • Active dissemination of results and open forum on research activities and digital issues	• Cost-efficient for data resources management
Services provided	**Data/technology/method-focused** services: - Discovery of data - Discovery of services and tools - Standardisation (high quality) and integration of data (for increase of re-use potential) - Data transfer - Curation of data - Training	Services focused on **sharing of knowledge and know-how**: - Technology services: Core platform for core technical services that are created, hosted and managed by DARIAH (e.g. persistent identifiers + shared authentication and authorisation + software to connect national infrastructure to DARIAH) + advanced research services proposed by research communities (not guaranteed by DARIAH) - Knowledge services: reference registries, digital curation, education and training	Services defined according to **challenges/areas** to be addressed / **results** to be achieved: 1) Blue Assessment - services for stock assessment and for the generation of unique identifiers for global stocks; 2) Blue Economy - services supporting the analysis of socio-economic performance in aquaculture; 3) Blue Environment - spatial planning services to identify aquaculture and fisheries infrastructures from satellite imagery; 4) Blue Skills - on-line training services and capacity building on existing training modules for fisheries scientists and other practitioners
Organisation	• Hub = transnational coordination structure • 5 platforms (about 30 to 50 partners each, roadmap and implementation studies per platform): Data, Compute, Tools, Interoperability, Training • Use cases across platforms • National Nodes = collection of research institutes within a country that run and sustain various resources	• Transnational coordination: - Four Virtual Competency Centres (VVC): 1. e-Infrastructure: establish a shared technology platform for arts and humanities research 2. Research and Education Liaison: expose and share research and education work in the digital humanities 3. Scholarly Content Management: facilitate the exposure and sharing of scholarly content 4. Advocacy, Impact and Outreach: interface with key influencers in and for the arts and humanities - Working Groups under each VCC (20 in	Project-based Strong interaction with D4Science

		total) • National coordinators	
Governance	• Consortium Agreement • 20 Members = national governments • Legal entity = EMBL (European Molecular Biology Laboratory) • Decision-making body = ELIXIR Board • National research institutes can apply jointly to become members of the Nodes via a Collaboration Agreement with the Hub	• 17 Members = national governments • Legal entity = European Research Infrastructure Consortium (ERIC) • Transnational coordination: - Decision-making body: General Assembly - Joint Research Committee: Heads of each Virtual Competency Centre	Project-based model
Business model	Initial ESFRI implementation support National contributions for networking and management Project-based funding H2020 (for implementation of ELIXIR and use cases under EXCELERATE)	Initial ESFRI implementation support National contributions (in-cash and in-kind) Project-based funding H2020 (Humanities at Scale)	Project-based (now: BlueBRIDGE to support development of iMarine), funding for the elaboration of 4 VREs
Collaborations	Industry support (focus on SMEs), ESFRI, RDA	ESFRI	D4Science

Source: ELIXIR[4], DARIAH[5], iMarine[6], BlueBRIDGE[7] and D4Science[8] websites; ELIXIR Scientific Programme 2014-2018; DARIAH ERIC: Technical and Scientific Description; BlueBRIDGE Project Proposal Summary (2017); Candela, L., Castelli, D. and Pagano, P. (2009)

[4] https://www.elixir-europe.org/
[5] http://www.dariah.eu/
[6] http://www.i-marine.eu/Pages/Home.aspx
[7] http://www.bluebridge-vres.eu/
[8] https://www.d4science.org/

2.3 E-INFRASTRUCTURE TECHNICAL ARCHITECTURE

In this section, we review the architectures of existing e-infrastructures in order to provide insights to the stakeholders of e-ROSA and inform their choices. To that aim, we need first to identify the components of the e-infrastructures and understand their role. We also need to identify which existing e-infrastructures provide generic and production services that can be reused in domain-specific e-infrastructures such as the one aimed by e-ROSA. ELIXIR and iMarine are two domain-specific e-infrastructures which provide concrete examples of architectures drawn from the needs of the communities they serve and built on generic e-infrastructures.

2.3.1 Technical requirements for e-infrastructures

The main goal of an e-infrastructure is to provide seamless and straightforward access to distributed and heterogeneous compute and/or data resources along with applications (tools and services). To that aim, e-infrastructures seek to fulfil one or several technical requirements, which correspond to layers of services at which the e-infrastructures provide resources. Figure 3 summarizes the different layers that are commonly provided by e-infrastructures and the issues they address.

Figure 3. Common requirements for e-infrastructures

Seamless access and connectivity — Network, authentication and authorization
- Secure data transfer of large files
- Fine grained access control

Data management and sharing — Data infrastructures
- Data discovery
- Data interoperability frameworks
- Data storage

On-demand computational resources — HPC, HTC and cloud computing
- Run thousands of compute-intensive tasks for analyzing large datasets
- Develop integrated applications combining computing and data resources as needed

User interface — Straightforward access to the e-infra resources along with applications
- Service registry
- Portal with different VO's worksites
- VREs
- Workflow tools

The last three service layers are of particular interest for domain-specific e-infrastructures such as the one envisioned in the context of e-ROSA:

- Data management and sharing: there is an increasingly vast amount of distributed data and one added value of an e-infrastructure is (i) to ensure they are correctly and securely handled and stored and (ii) to facilitate their discoverability, access and interoperability accordingly to the FAIR principles[9], making data infrastructures crucial in an e-infrastructure.

[9] FAIR Principles : https://www.force11.org/fairprinciples

- On-demand computational resources: if and as needed, an e-infrastructure is required to offer computational resource to run thousands of compute-intensive tasks for analysing large datasets, and to develop integrated applications that combine both computing and data resources.
- User interface: distributed systems like e-infrastructures tend to be complex and their uptake by the researchers complicated. Thus, there is a strong need to hide the underlying heterogeneity and complexity to the end users and provide them with a straightforward access to the e-infrastructure. We have identified several ways of accessing an e-infrastructure's resources, including a service registry, a web portal with different VO's worksites, VREs and desktop or online workflow tools which use Web services APIs to connect to the e-infrastructure.

Thus, existing generic and domain-specific e-infrastructure can be distinguished according to the layers at which they provide resources (Table 2):

Table 2. Classification of e-infrastructures according to the multiplicity of layers at which they provide services

	Generic e-infrastructures		**Specific e-infrastructures**	
	Name	*Service layer(s) leveraged*	*Name*	*Service layer(s) leveraged*
Focused on one layer	AARC	Network	Copernicus	Data
	PRACE	Computing	GEOSS	Data
	Helix Nebula	Cloud	SeaDataNet	Data
	OpenAire	Data	WheatIS	Data
	Zenodo	Data		
	EUDAT	Data		
Multi-layer	GEANT	Network, Cloud	Elixir	Network, Data and Computing
	EGI	Data, Cloud & computing	iMarine	Network, Data, Computing and User interface
	D4Science	Data, Computing & user interface	Dariah	Network, Data and Computing

2.3.2 Relationship between generic and domain-specific e-infrastructures

As illustrated in the figure above, most of the technical requirements can be fulfilled (with a varying degree of maturity) based on existing generic e-infrastructures such as GEANT, EGI, PRACE, etc. Table 3 recapitulates what resources and services are offered by the most popular generic e-infrastructures and the table in Table 4 shows which components of which existing generic e-infrastructures are leveraged by domain-specific e-infrastructures such as ELIXIR and iMarine. The access policies to generic resources and services need to be clearly identified in order to assess their potential use within a future e-infrastructure for agri-food.

Table 3. Layers of services of existing generic e-infrastructures

Service layers	GÉANT	EGI	PRACE	HELIX NEBULA	EUDAT	ZENODO	OpenAIRE	D4Science
Network	x							
Data		x			x	x	x	x
Computing	x	x	x	x				x
User interface								x

Table 4. Use of generic e-infrastructures by the domain-specific e-infrastructures ELIXIR and iMarine

Requirements/ Service layers	ELIXIR		iMarine	
	Generic e-infrastructure used	*Service layer(s) leveraged*	*Generic e-infrastructure used*	*Service layer(s) leveraged*
Network	GEANT	Network		
Data	EUDAT	Data	D4Science	Data
Computing	EGI	Middleware, computing	D4Science	Computing
	PRACE	Computing	EGI	Computing
User interface			D4Science	VRE management

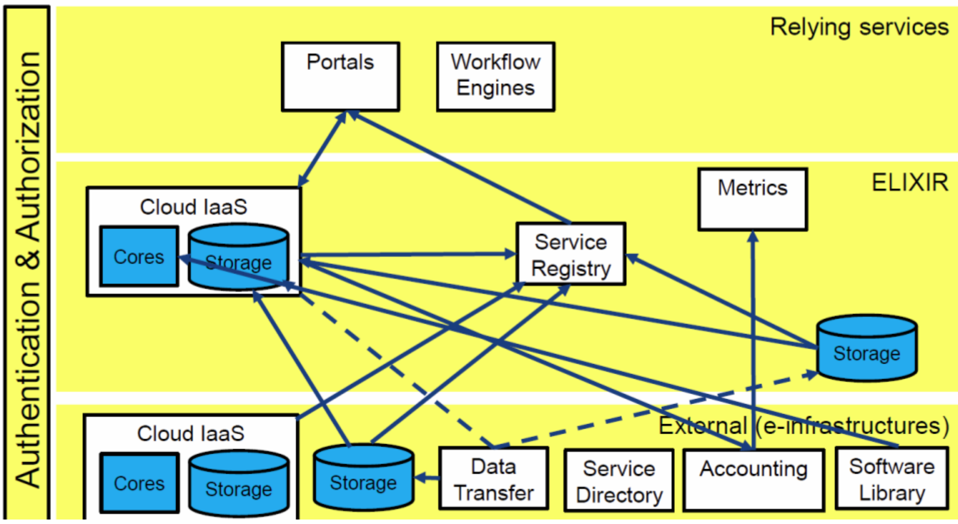
2.4 HIGH LEVEL ARCHITECTURE OF DOMAIN-SPECIFIC E-INFRASTRUCTURES

The following section provides examples of architectures of domain-specific e-infrastructures in order to show the variety of architecture definition approaches and feed into the reflection of the envisioned architecture of the future e-infrastructure for agri-food. The typical architecture of domain-specific e-infrastructure is an N-tier architecture, most of the time a 3-tier architecture including (i) a tier representing consumers, (ii) a mediation or brokering tier, and (iii) a tier representing the providers. The mediation tier has the role to accommodate the needs of resources providers and consumers. The way this mediation tier is implemented (i.e. how loosely or strongly resources are integrated) and the services it offers differ from one e-infrastructure to another. This section provides examples of high level architecture of existing domain-specific e-infrastructures.

<u>Elixir</u>

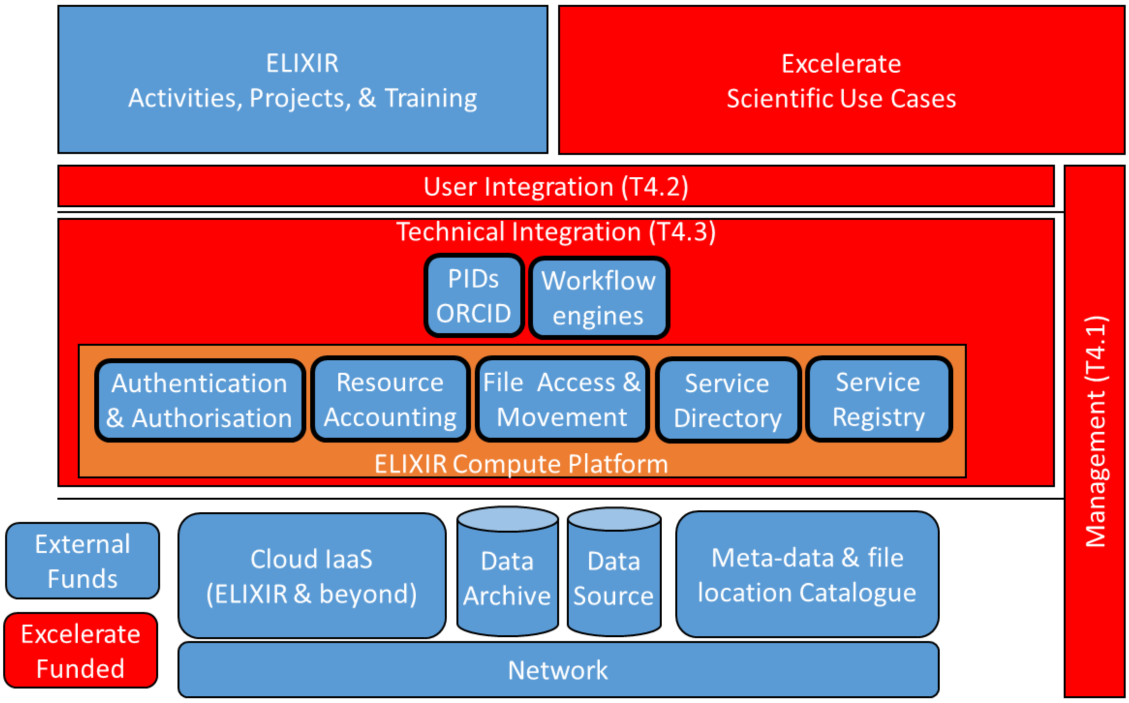
The Elixir e-infrastructure's computing platform (Figure 4 and Figure 5) can be considered as its mediation tier. This computing platform is composed of services operated directly by Elixir's nodes and discoverable by consumers through a service registry. It also reuses external services such as federated identity management infrastructure and trust management fabrics provided by GÉANT.

Figure 4. ELIXIR cloud and compute architecture



Source: Newhouse, S., ELIXIR

Figure 5. ELIXIR architecture



Source: Newhouse, S. and Ruda, M., ELIXIR Cloud Roadmap

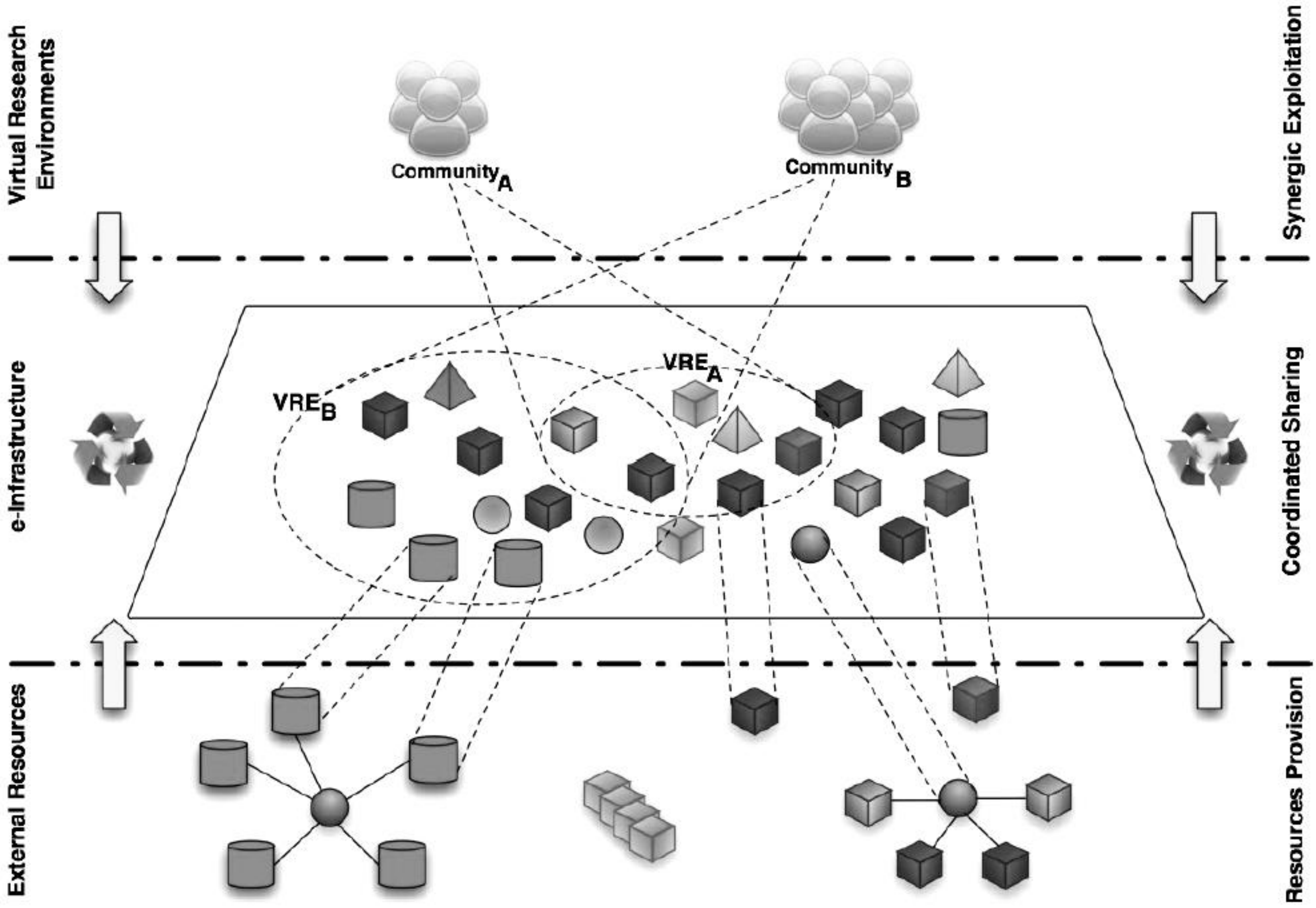
<u>iMarine</u>

The iMarine e-infrastructure aims to establish and operate a Data e-Infrastructure supporting a Community of Practice with an Ecosystem Approach to Fisheries Management[10] and Conservation of Marine Resources. iMarine is based on D4Science which assumes the framework depicted in Figure 6 (Candela, Castelli, & Pagano, 2009). The brokering function of D4Science makes it possible to enrich the resources shared by third parties with functions like retrieval, access, annotation of content and creation of new ones. Furthermore, iMarine provides logistic and technical support for application building, maintenance and monitoring. While relying on different resource providers as Elixir does, iMarine's mediation tier acts also as an operator, providing the possibility to dynamically create Virtual Research

[10] Garcia, S.M. et al. (2003). The ecosystem approach to fisheries.

Environments (VREs) which give access to a selection of infrastructure resources through web-based collaboration and social environments. (Castelli, 2014).

Figure 6. The framework of D4Science e-infrastructure



Source: Candela, Castelli, & Pagano, 2009

GEOSS

The GEOSS (Global Earth Observation System of Systems) offers a portal allowing the user of Earth observations to access, search and use available data, information, tools and services. The GEOSS infrastructure consists of four main elements:

- A web interface (the GEO Portal) allowing the user to access GEOSS and search for information and services.
- An engine (the GEOSS Clearinghouse) connects directly to the various GEOSS components and services, collects and searches their information and distributes data and services via the Portal to the user.
- The GEOSS Components and Services Registry. Each partner contributing components and services to GEOSS provides essential details about the name, contents, and management of their contribution. This assists the Clearinghouse, and ultimately the user, to identify the GEOSS resources that may be of interest.

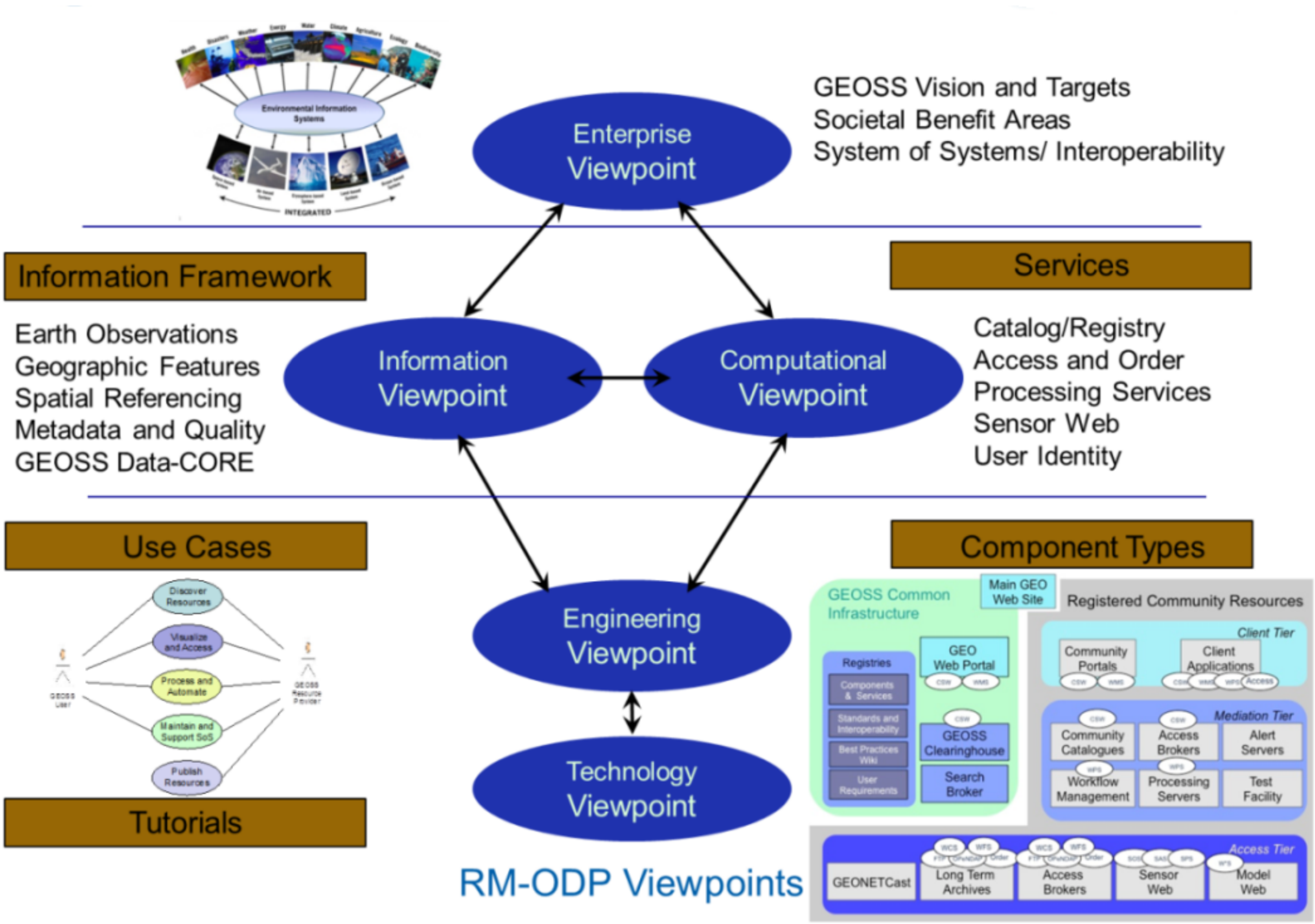
- A Standards and Interoperability Registry, a pillar of the infrastructure. It allows the contributors to GEOSS to configure their systems so that they can interoperate with other systems.[11]

In addition to these four main components, GEOSS provides a Best Practices Wiki allowing the community to propose, discuss and converge upon best practices in all fields of earth observation.

Similarly to the iMarine'e-infrastructure, the mediation tier of GEOSS acts as an "operating system" as it includes resources and services such as processing servers and workflow management.

Figure 7 different viewpoints of the GEOSS e-infrastructure and Figure 8 zooms in the technical architecture of the e-infrastructure.

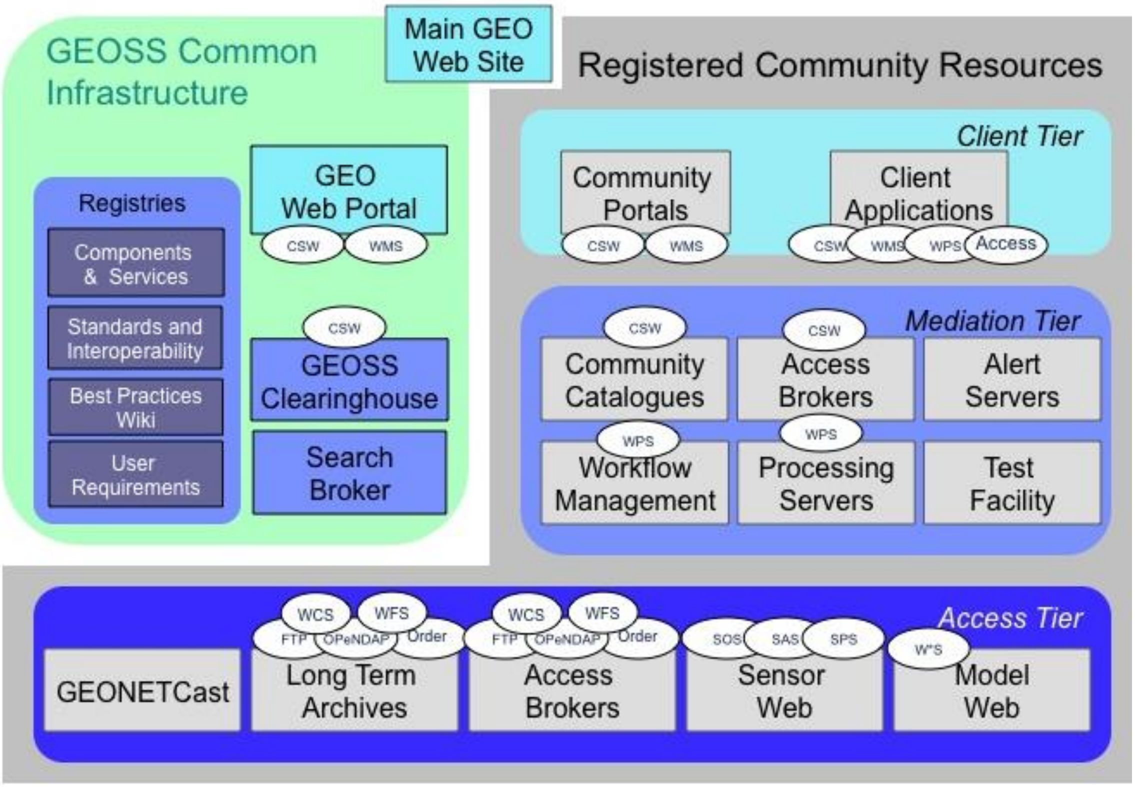
Figure 7. Different viewpoints of the GEOSS e-infrastructure



Source: Percivall, G. and de Lathouwer, B. (2013).

[11] https://www.earthobservations.org/gci_gci.shtml

Figure 8. Technical architecture of the GEOSS e-infrastructure



Source: Percivall, G. and Browdy, S. (2012), GEOSS Design and Interoperability

2.5 KEY LESSONS LEARNT

The comparison conducted in Sections 2.3 and 2.4 allows to identify key aspects to be considered (i.e. typology framework) in the envisioning, design and/or strategic orientation and implementation of any existing or future e-infrastructure in the field of agri-food research:

- Genericity (technology-oriented) vs. domain-specificity
- Single layer (e.g. OpenAIRE) vs. multi-layer (e.g. Virtual Research Environments)
- Architecture definition approach and alignment with EOSC's architecture
- Implementation of the mediation tier of the e-infrastructure: the way this mediation tier is implemented (i.e. how loosely or strongly resources are integrated) and the services it offers differ from one e-infrastructure to another and depends on whether the e-infrastructure as an entity has chosen to act as an operator or not.
- Relationship between the e-infrastructure and research communities: an e-infrastructure dedicated to a large and heterogeneous community such as the one of agri-food can serve different more homogenous sub-communities through VOs. Each VO will then structure the sub-community around a set of generic tools and services such as wikis, blogs and videoconference applications, and specific VREs created on demand to address specific scientific application scenarios.
- Federation of multiple, distributed resources (i.e. inclusive) vs. minimal resources to offer a service: this can be strongly related to the organisation and business model of the e-infrastructure (e.g. overarching hub with national nodes vs. project-based model)
- Coordination vs. integration of services and technologies

- Focus on data stewardship and sharing of know-how (e.g. ELIXIR and DARIAH) vs. on the development of applications directly by the e-infrastructure (e.g. iMarine)
- Oriented towards data challenges (e.g. ELIXIR) vs. societal challenges (iMarine)
- Importance of use cases and relationship with overall strategy (e.g. continuous tailoring of the overall strategy thanks to use cases as in ELIXIR)

Envisioning a future e-infrastructure for agri-food in line with the EOSC vision and architecture (i.e. System-of-systems with facilities *as-a-Service*) is a key priority for e-ROSA partners and stakeholders. This will be further developed throughout the lifetime of the e-ROSA project (e.g. next EOSC Meeting in September 2017 on the EOSC architecture).

3 MAPPING OF EXISTING RESOURCES

3.1 OBJECTIVE

The mapping task carried out under e-ROSA aims 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 e-ROSA. Overarching goals of this task include:

- 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.

3.2 OVERALL APPROACH

3.2.1 Entities to be mapped

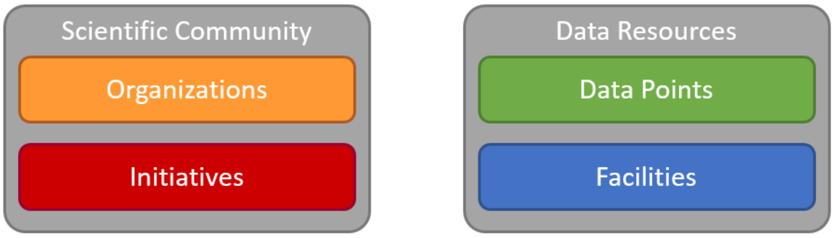
An initial exercise on the definition of the scope of the mapping task was required in order to share a common understanding amongst e-ROSA partners of the stakeholders and resources that need to be mapped for the related work that will be carried out later on under the e-ROSA project (i.e. community-building and roadmap development) and beyond the project (e.g. European Open Science Cloud, GODAN Data Ecosystem Working Group[12], etc.).

Overall, e-ROSA partners seek to provide an in-depth overview of:

- The scientific community: this includes research organisations, projects & networks.
- The data and e-infrastructure resources: this includes data points, infrastructures & related services.

These two categories have each been subdivided into two (Figure 9):

Figure 9. Entities mapped under e-ROSA



Definitions for the four entities of the online map have been elaborated and are detailed below:

1. *Organisations*: the organisations (e.g. research performing organisations as well as ministries and international organisations) within the e-ROSA scope
2. *Initiatives*: Projects, networks and other initiatives within the e-ROSA scope
3. *Data points*: in other words, all forms of data sources directly or indirectly providing access to data in the field of agriculture and food (i.e. data sets, catalogues, repositories, aggregators and portals). The items listed under this entity answer the question “how to get data?”.

[12] The GODAN Data Ecosystem Working Group explores the vision of catalysing the creation a global, common data ecosystem for agriculture and food: http://www.godan.info/working-groups/data-ecosystem-working-group.

4. *Facilities*: research infrastructures and e-infrastructures that provide one or several types of data-related services (e.g. modelling infrastructures, data management infrastructures, etc.). The items listed under this entity answer the question “what can you do with data?”.

3.2.2 Characterisation of mapped entities

An agreement on the initial approach adopted to describe mapped entities was reached:

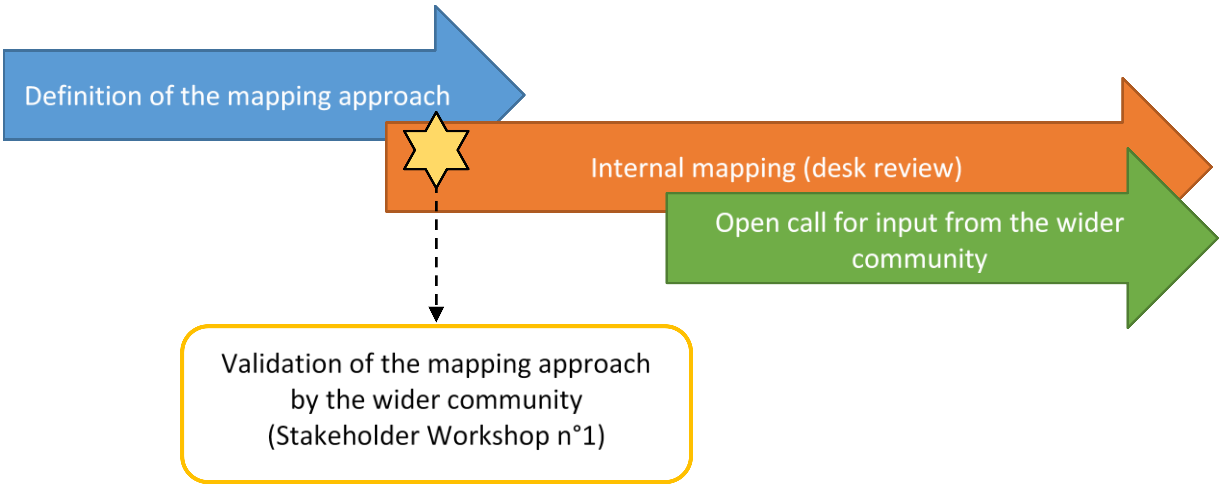
- The **geographic coverage**, the targeted **scientific discipline(s)** and the **impact area(s)** are characterised for all entities. The field “impact area(s)” seeks to reflect the specificity of the mapped item (e.g. more precise agricultural challenge addressed), however it may not be of high added value for the purpose of our mapping exercise and may be removed if deemed relevant.
- **Data science categories** were defined to characterise the items listed under “initiatives” and “facilities”: these allow to describe the data-related challenge(s) and/or service(s) the initiative/facility focuses on (see detailed list of data science categories in Section 4.3).
- The different **types of data points and facilities** have been listed in order to characterise the nature of the items listed under both entities (see detailed lists in Section 4.3).
- The **access policy** approach is described for data points and facilities: listing the detailed access policies and licenses for related items was considered as irrelevant in relationship to the desired level of granularity of the information provided for all entities. Thus, we agreed on describing the broader access policy “approach” adopted: i.e. open, controlled or private.
- The **development stage** is indicated for facilities in order to give a brief assessment of the level of maturity of the considered facility (i.e. project or operational service).
- The **relationships** between the various entities can be characterised: for instance, a link between an organisation and an initiative will be defined if the organisation is a member of the initiative.

The resulting detailed metadata model is described in Section 4.3.

3.2.3 Mapping process

The different steps of the mapping process throughout the lifetime of the e-ROSA project are described in Figure 10 below:

Figure 10. e-ROSA’s mapping process



In particular, the mapping exercise relies on a collaborative process:

- The mapping approach and method developed by e-ROSA partners has been discussed during a meeting with the GODAN Data Ecosystem Working Group in March 2017 as well as with the wider e-ROSA community during the first Stakeholder Workshop in July 2017;
- After an internal desk review to elaborate an initial map of stakeholders and resources, an open call will be launched in order to seek feedback from the e-ROSA community.

3.3 METADATA MODEL

Based on the definition of the entities to be mapped (see Section 4.2.1.), a first metadata model was proposed and refined after discussing and agreeing on the characteristics to highlight for each mapped item (see Section 4.2.2.). The resulting metadata model is presented below (see Table 5) and is subject to further changes if deemed relevant when testing it on first mapped items.

Table 5. Overarching metadata model for mapped entities

Organisations	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
Impact area(s)	Access policy	Data science category(ies)	Development stage
	Scientific discipline(s)	Impact area(s)	Access policy
	Impact area(s)	Related organisation(s)	Scientific discipline(s)
	Related organisation(s)		Data science category(ies)
	Related initiative(s)		Impact area(s)
	Related facility(ies)		Related organisation(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 Agri-food[13];

[13] http://vest.agrisemantics.org/

- The values under "Impact area(s)" should be based on the vocabularies listed under GACS[14]. GACS is a set of curated standards which brings together three major thesauri: AgroVoc[15], CAB[16] and NAL[17].
- The fields "Geographic coverage", "Type of data point", "Type of facility", "Development stage", "Access policy" and "Data science category(ies)" only accept restrained values as listed in Table 6 below.

Table 6. Fields with restrained values

Geographic coverage	Type of data point	Type of facility	Development stage	Access policy
National	Set	Farm	Project	Open
European	Catalogue	Land	Operational service	Controlled
International	Repository	Garden		Private
	Aggregator	Ranch		
		Education center		
		Research center/station		
		Greenhouse		
		Virtual Lab		
		e-infrastructure		
		Laboratory		
		Field station		
		Network of research infrastructures		

Scientific discipline(s)	Data science category(ies)
Agriculture - General	Data production
Animal Production and Health	Data veracity
Economics and Policy	Data discovery & access
Education and Extension	Data storage
Engineering, Technology and Research	Technical data interoperability
Farming Practices and Systems	Semantics
Fisheries and Aquaculture	Data integration
Food safety and Human nutrition	Computation
Food Security	Modelling, statistics & simulation
Forestry	Workflows
General Arts and Humanities	Data visualisation
Geographical and Regional Information	Data ownership
Government, Administration and Legislation	Data publication
Information Management	Support for decision-making
Medicine	Governance
Natural Resources and Environment	
Plant Production and Protection	
Rural and Social Development	

[14] http://browser.agrisemantics.org/gacs/en/index
[15] http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus
[16] http://www.cabi.org/cabthesaurus/
[17] https://agclass.nal.usda.gov/

3.4 NEXT STEPS

As explained above, this section has provided an overview of the developed approach and methodology to map the various stakeholders and resources that are of relevance for future e-infrastructures in the field of agri-food.

Next steps include:

- The launch of the internal mapping procedure, i.e. a desk review carried out by the e-ROSA team to identify the various items under each entity (including “organisations”, “data points”, initiatives” and “facilities”): this task does not aim at exhaustiveness but rather at quality in the relevance/importance of the mapped items and their characterisation.
- The Stakeholder Workshop in July 2017 provides the opportunity to discuss the mapping approach and the metadata model and to prepare the wider community to provide input to the mapping exercise later on through the planned open call.
- The conclusions of the workshop and the mapping test with items that have been included in the knowledge base following our initial metadata model will allow the fine-tuning of the latter.

4 CONTRIBUTION TO ROADMAP ELABORATION

As the overall goal of this report is to provide key elements for the e-ROSA roadmap elaboration, we aim in the following section at proposing a draft table of content that can serve as an initial basis for the future roadmap that will be elaborated in the context of e-ROSA. In order to do so, we have analysed several e-infrastructure roadmaps and related documents, including:

- A Global Data Ecosystem for Agriculture and Food (GODAN, 2016)
- A Place to Stand: e-Infrastructures and Data Management for Global Change Research (Belmont Forum, 2015)
- D4Science: an e-Infrastructure for Supporting Virtual Research Environments
- DARIAH ERIC: Technical and Scientific Description (2011)
- ELIXIR Scientific Programme 2014 - 2018
- ESFRI Strategy Report on Research Infrastructures - Roadmap 2016
- e-IRG Guidelines Document: Evaluation of e-Infrastructures and the development of related Key Performance Indicators (2017)
- Sustaining Scholarly Infrastructures through Collective Action: The lessons that Olson can teach us (Neylon C., 2017)
- Health-RI: The NL Personalised Medicine & Health Research Infrastructure (2016)
- Scientific and Technical description of LifeWatch ERIC
- Research Councils UK: E-infrastructure Roadmap (derived from the roadmap developed by EPSRC)
- GACS Vision paper (draft version)
- INRA: Vers une e-infrastructure pour la recherche agronomique (draft version)

The draft table of content of the future e-ROSA Roadmap could build on the following elements:

1. **<u>The data opportunity in agriculture & food</u>**

This introductory section seeks to justify the need for sustainable, operational e-infrastructures that can support the data challenges in agri-food research. Existing high-level strategies (e.g. GODAN[18], Open Harvest Declaration n°1[19] & n°2[20]) already address the overarching rationale for e-ROSA and future e-infrastructures that can emerge from e-ROSA, and will be further elaborated thanks to the work undertaken under Work Package 2 "Challenges & Ambitions".

2. **<u>The European and international context</u>**

Both data-related and agriculture-related policies, strategies and networks have been established and can support the implementation of the future e-ROSA Roadmap: e.g. the European Policy Food 2030, the European Open Science Cloud and its future Food Cloud Demonstrators, the UN Sustainable Development Goals, the international network GODAN, the Research Data Alliance, etc.

3. **<u>Our vision</u>**

The 2nd Chania Declaration asserts that the federation of multiple, distributed resources can pave the way for new e-infrastructures in the agri-food sector.

An initial overall framework model has been elaborated under e-ROSA (see Figure 11) in order to understand the service-approach of e-infrastructures. Services have been divided into 3 layers (from left to right):

---

18 http://www.godan.info/documents/data-ecosystem-agriculture-and-food
19 http://blog.agroknow.com/wp-content/uploads/2016/05/Chania-Declaration.pdf
20 https://drive.google.com/file/d/0B41Vz7BieQuoY19CYlBISklxbmc/view

1. “Data layer”: the data resources and related services for raw data management
2. “Interoperability layer”: the services allowing access and linking of data from various sources
3. “End-user layer”: the analytical and valorisation services

Figure 11. Three-layer model for e-infrastructure services

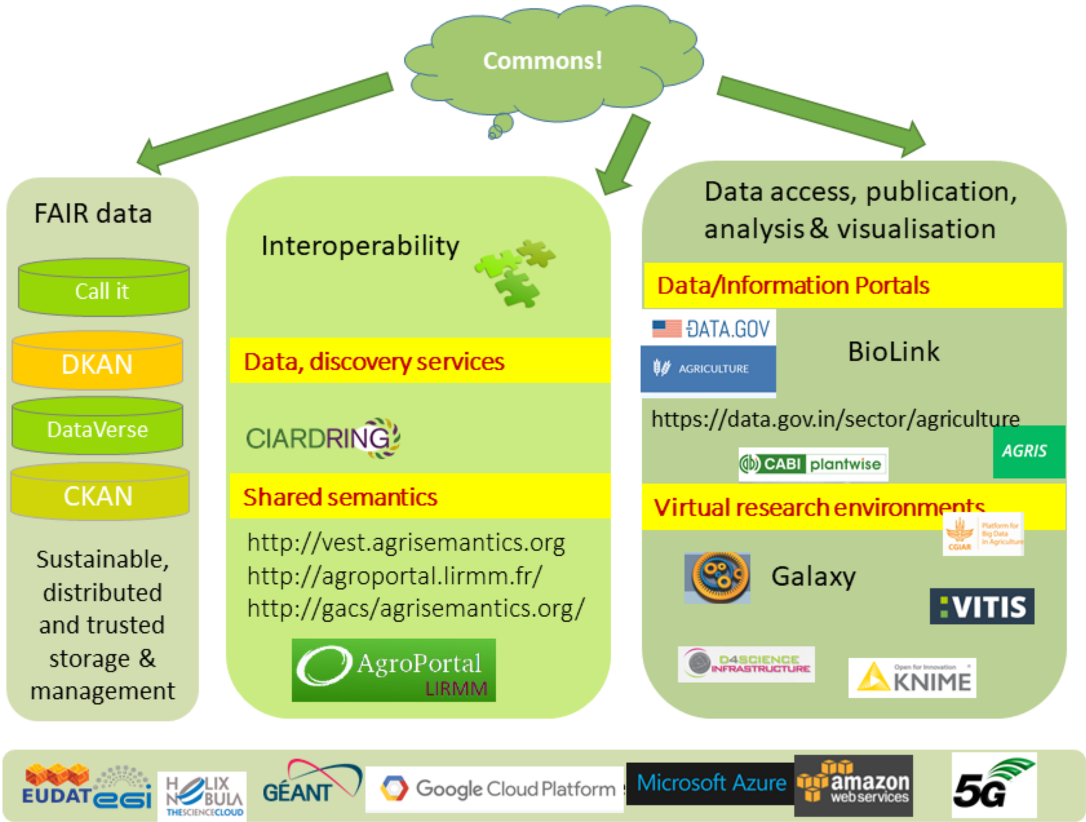
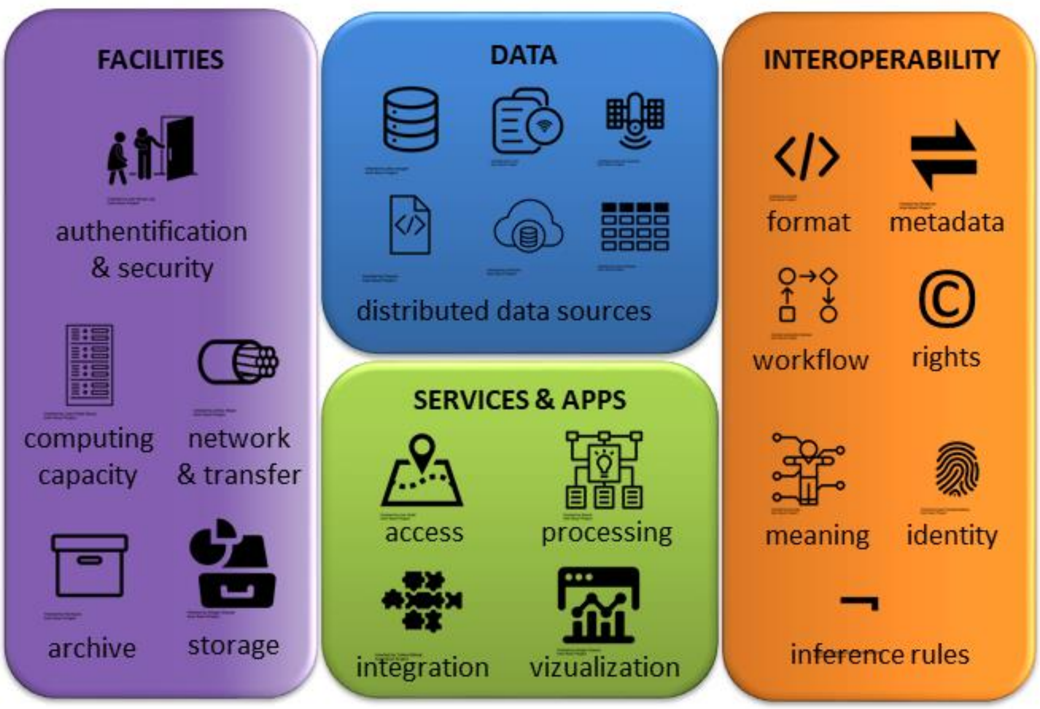


Figure 12 below provides another viewpoint on this model, showing the transversal aspect of facilities and of the interoperability issue:

Figure 12. The e-infrastructure components



4. Needs

In order to achieve the vision, the identification of specific data-related needs is required, as well as the definition of matching objectives in order to address those needs. This task is carried out under WP2 and will especially rely on the input of scientific communities provided during the 2nd e-ROSA Stakeholder Workshop in November 2017.

5. Proposed response

Future (potential) initiatives and perspectives on the evolution of already existing ones should be described in order to demonstrate how we can achieve our vision and address specific needs through practical proposals (i.e. key actions to implement and partnerships to build). The input from the e-ROSA stakeholder community is crucial in order to elaborate such proposals.

Required collaborations (e.g. amongst domain-specific initiatives such as Agrisemantics[21] and GODAN, as well as with generic e-infrastructures such as EUDAT and OpenMinTeD[22]) should be highlighted.

As next steps, an approximate timeline and related funding opportunities (business models) should be agreed on.

6. Long-term perspective and strategic objectives

Long-term objectives should be clearly stated, e.g.:

- **Fostering more efficient research**: this can be achieved through the defragmentation of data infrastructures and service providers, which supports increased data sharing and reuse and hence hinders the duplication of data collection and analysis efforts, and of related research efforts as a whole. In particular, the efficient cataloguing of local dataset repositories can support the potential integration of multiple data sources.

- **Supporting the practical implementation of policy and stakeholder frameworks to meet societal challenges**: The project BlueBRIDGE provides a good example of publications (i) supporting the EU domain-specific strategy[23] and (ii) supporting the EOSC[24].

- **Community-building and stakeholder engagement**: This includes mapping activities and reaching out to the various stakeholder groups, and more importantly future initiatives should provide networking and coordination mechanisms (e.g. the Open Harvest event organised by Agroknow, the Policy Committee meetings organised by e-ROSA to bring together the various EC DGs, etc.). These activities are key to build motivation for:
    - collaboration and sharing; and
    - Innovation, by linking the various stakeholder groups (public, private, citizens).

- **Capacity-building**: The concept of network of Competence Centres across the globe was introduced during the 2nd Open Harvest event 2017. In particular, Competence Centres could provide support for:
    - FAIRifying data: common recommendations, tools and methods, and guidance
    - adopting a use case approach: adapt common methods and tools to the specific environment and conditions of researchers and other stakeholders

[21] http://agrisemantics.org/
[22] http://openminted.eu/
[23] http://www.bluebridge-vres.eu/publications/press-releases/innovative-bluebridge-data-services
[24] http://www.bluebridge-vres.eu/publications/press-releases/how-bluebridge-can-effectively-support-establishment-european-open

- **Fostering distributed efforts and flexible governance**: As adopting a centralised approach would not be relevant, empowerment by national infrastructures supported by related strategies is required in order to ensure sustainability in time. Flexible governance models will allow to move from club-like goods towards public goods. The mapping of resources in local contexts and in relationship to specific challenges is a first step towards a distributed effort.

5 CONCLUSION

This report aims at providing a starting basis for the work that will be carried out under e-ROSA in the upcoming months. In particular, Sections 2 and 3 provide food for thought for the envisioning and design of the future e-infrastructure for agri-food research. Based on this analysis and on the outputs of e-ROSA's first Stakeholder Workshop in July 2017, the project team seeks to co-elaborate a vision paper with the e-ROSA stakeholder community. The first version of this vision paper will be discussed in the 10th Plenary of the Research Data Alliance in September 2017.

In addition, the EOSC meeting on the EOSC architecture in September 2017 will provide an opportunity for the e-ROSA team to further discuss the potential link between the envisioned e-infrastructure for agri-food and the EOSC vision and architecture (i.e. System-of-systems with facilities *as-a-Service*). Key lessons learnt in Section 2.5 will be extended in the updated versions of the deliverable (D1.5 in M12 and D1.6 in M18) thanks to the collection and analysis of further information that can feed into the potential e-infrastructure design and roadmap in line with the EOSC vision and implementation.

The community-building process and the improvement of the knowledge of the current landscape (i.e. existing resources and gaps), future trends, disruptions and needs will be addressed through the co-mapping process via the open call to provide input to the map of stakeholders in the agri-food sector, as well as during the second e-ROSA Stakeholder Workshop in November 2017, which will focus on the identification of scientific and data-related needs, and critical, visionary use cases. The updating of this deliverable (in M12 & M18) will allow to monitor further achievements regarding the mapping of the e-ROSA stakeholder community that can support the roadmap co-elaboration and e-infrastructure design, including scientific communities, existing generic and domain-specific e-infrastructures and initiatives, as well as key data resources within the scope of agri-food.

ANNEX 1 - THE EUROPEAN STRATEGY FORUM ON RESEARCH INFRASTRUCTURES (ESFRI)

ESFRI seeks to facilitate multilateral initiatives leading to the better use and development of research infrastructures, at EU and international level. It identifies Research Infrastructures (RIs) of pan-European interest meeting the long-term needs of Europe's research communities across all scientific areas. As such, it provides to the Council of the European Union a coherent and strategic vision regarding European RIs through the publication of updated ESFRI Roadmaps.

ESFRI acts as an incubator for new RIs or for the upgrading of existing ones through their inclusion in its Roadmap as ESFRI Projects. It supports their implementation within a maximum of one decade, in order to reach sustainability for the long term operation, therefore assuring maximum return on investment in terms of science, knowledge, innovation, training, socio-economic benefits and competitiveness. ESFRI Projects are selected for their high degree of maturity and for their strategic importance within the European Research Infrastructure system. ESFRI Landmarks are successfully implemented ESFRI Projects that are featuring top science services or effectively advancing in their construction. ESFRI facilitates their continuous support for successful completion, operation and upgrade in line with the optimal management and maximum return on investment. ESFRI currently supports 21 ESFRI Projects and 29 ESFRI Landmarks.

In addition, ESFRI provides a mapping of Research Infrastructures open to European scientists at national, European and international levels through its Landscape Analysis. The latter hence gives an overall view of the European RI system and is key to the selection of new ESFRI projects as it allows to better understand the potential impact of proposed RIs.

Over the past decade, ESFRI has improved the efficiency and impact of the European RI system. Most national strategies are now coordinated with that of ESFRI and move towards a sustainable investment for overall competitiveness.

Source: ESFRI Roadmap 2016; https://www.era-learn.eu/alignment/current-approaches/ERALEARN2020_T42_Casestudyno5_JPIOceans_28September2016_Final.pdf

REFERENCES

Literature

Allison, L. and Gurney, R. (2015). A Place to Stand: e-Infrastructures and Data Management for Global Change Research – Belmont Forum e-Infrastructures & Data Management Community Strategy and Implementation Plan. Belmont Forum. Available at: http://www.bfe-inf.org/sites/default/files/A_Place_to_Stand-Belmont_Forum_E-Infrastructures__Data_Management_CSIP.pdf

Allison, L., Asch, M., Gurney, R., Hodgson, S., Uhle, M., Davis, R., Dobson, M., Oakley, K. and Watson, K. (2015). Belmont Forum E-Infrastructure and Data Management – Phase 2: Implementation Plan. Belmont Forum. Available at: http://www.bfe-inf.org/sites/default/files/doc-repository/eI%26DM_Implementation_Plan_15April.pdf

Allemang, D. and Teegarden, B. (2016). A Global Data Ecosystem for Agriculture and Food. GODAN. Available at: http://www.godan.info/documents/data-ecosystem-agriculture-and-food

Arvanitidis, C., Basset, A., González-Aranda, J.M., Hernández, F., Sánchez, B., Van Tienderen, P. and Los, W. (2015). Scientific and Technical description of LifeWatch ERIC. ERIC. Available at: http://lifewatch.eu/-/m/851/Scientific%20and%20Technical%20description%20of%20LifeWatch%20ERIC.pdf

BlueBRIDGE. BlueBRIDGE Project Proposal Summary. Available at: http://www.bluebridge-vres.eu/about-bluebridge

Candela, L., Castelli, D. and Pagano, P. (2009). On-demand virtual research environments and the changing roles of librarians. Library Hi Tech, Vol. 27 No. 2, 2009, pp. 239-251, Emerald Group Publishing Limited, 0737-8831. doi: https://doi.org/10.1108/07378830910968191 . Available at: http://www.nmis.isti.cnr.it/candela/Leonardo_Candela_Website/my_Papers/Entries/2009/3/14_On_Demand_Virtual_Research_Environments_..._files/Library%20hi%20tech.pdf

Candela, L., Castelli, D. and Pagano, P. (2009). D4Science: an e-Infrastructure for Supporting Virtual Research Environments. Available at: https://pdfs.semanticscholar.org/fa5f/3297f4de0a8433b8429ef89a80ce8113a013.pdf

Carolan, L., Smith, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J. and Gerry, W. (2015). How can we improve agriculture, food and nutrition with open data?. Open Data institute. Available at: http://www.godan.info/sites/default/files/old/2015/04/ODI-GODAN-paper-27-05-20152.pdf

DARIAH (2011). DARIAH ERIC: Technical and Scientific Description. Available at: http://www.dariah.eu/fileadmin/Documents/dariah_scientific-technical-description.pdf

ELIXIR (2014). ELIXIR Scientific Programme 2014-2018. Available at: https://www.elixir-europe.org/sites/default/files/documents/elixir_scientific_programme_final.pdf

ESFRI (2016). ESFRI Roadmap 2016 - Strategy Report on Research Infrastructures. Available at: http://www.esfri.eu/roadmap-2016

European Commission (2016). E-infrastructures: Making Europe the best place for Research and Innovation. Available at: https://ec.europa.eu/digital-single-market/en/news/e-infrastructures-making-europe-best-place-research-and-innovation e-IRG (2015). e-IRG Guidelines Document: Best Practices for the use of e-Infrastructures by large-scale research infrastructures. Available at: http://e-irg.eu/catalogue/eirg-1000

e-IRG (2016). e-IRG Roadmap 2016. Available at: http://e-irg.eu/catalogue/eirg-1002

e-IRG (2017). e-IRG Guidelines Document: Evaluation of e-Infrastructures and the development of related Key Performance Indicators. Available at: http://e-irg.eu/catalogue/eirg-1005

GACS (2017). GACS Vision Paper. (draft version)

Garcia, S.M., Zerbi, A., Aliaume, C., Do Chi, T. and Lasserre, G. (2003). The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. FAO Fisheries Technical Paper. No. 443. Rome, FAO. 2003. 71 p. Available at: www.fao.org/3/a-y4773e.pdf

INRA (2017). Vers une e-infrastructure pour la recherche agronomique. (draft version)

Jeffery, K.G., Meghini, C., Concordia, C., Patkos, T., Brasse, V., van Ossenbruck, J., Marketakis, Y., Minadakis, N. and Marchetti, E. (2017). A Reference Architecture for Virtual Research Environments. Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, 13th—15th March 2017. Glückstadt: Verlag Werner Hülsbusch, pp. 76—88. Available at: https://edoc.hu-berlin.de/bitstream/handle/18452/2100/jeffery.pdf?sequence=1&isAllowed=y

Neylon, C. (2017). Sustaining Scholarly Infrastructures through Collective Action: The lessons that Olson can teach us. bioRxiv 116756. doi: https://doi.org/10.1101/116756

Open Harvest (2016). Chania Declaration. Available at: http://blog.agroknow.com/wp-content/uploads/2016/05/Chania-Declaration.pdf

Open Harvest (2017). 2nd Chania Declaration: a Call to Action. Available at: https://drive.google.com/file/d/0B41Vz7BieQuoY19CYlBISklxbmc/view

Research Councils UK. E-infrastructure Roadmap. Available at: http://www.rcuk.ac.uk/documents/documents/roadmapforelc-pdf/

Tillman, L., Morrell, S., Hanlon, T., Emmerson, D., Erat, M., Clark, E. and Erskine, T. (2014). E-Infrastructure Roadmap. EPSRC. Available at: https://www.epsrc.ac.uk/newsevents/pubs/e-infrastructure-roadmap/

Wijmenga, C., Meijer, G., Mons, B., Luijten, P., Dekker, A. and Kok, R.G. (2016). Health-RI: The NL Personalised Medicine & Health Research Infrastructure. Available at: https://www.knaw.nl/shared/resources/adviezen/bestanden/KNAWAgendaHealth_RI.pdf

Candela, L., Castelli, D., & Pagano, P. (2009, March). On-demand virtual research environments and the changing roles of librarians.

Castelli, D. (2014, octobre 3). What do research e-infrastructure operators currently do to facilitate access to and use of their resources by SMEs? Brussels.

Presentations

Bongo, L.A., Borg, M., Cornelis, A. et al. ELIXIR Compute platform [version 1; not peer reviewed]. F1000Research 2017, 6:417 (poster). doi: https://doi.org/10.7490/f1000research.1113867.1

Bongo, L.A., Borg, M., Cornelis, A. et al. The ELIXIR Compute Platform: A Technical Services Roadmap for supporting Life Science Research in Europe. Zenodo. doi: https://doi.org/10.5281/zenodo.60291

Castelli, D., Ellebroek, A., Taconet, M. and Pagano, P. (2012). iMarine. Marine Knowledge All Projects meeting, October 2012. Available at: http://puma.isti.cnr.it/rmydownload.php?filename=cnr.isti/cnr.isti/2012-A3-018/2012-A3-018.pdf

Manzi, A. and Pagano, P. (2011). E-Infrastructure Integration with gCube. EGI User Forum, April 2011. Available at: https://www.slideshare.net/d4science/einfrastructure-integrationwith-gcube

Newhouse, S.. ELIXIR. Available at: https://indico.cern.ch/event/461262/contributions/1132616/attachments/1214189/1772254/ELIXIR-HN-Jan2016b.pdf

Newhouse, S. and Ruda, M.. ELIXIR Cloud Roadmap. Available at: https://indico.egi.eu/indico/event/2452/session/66/contribution/195/material/slides/

Percivall, G. and de Lathouwer, B. (2013). Observation Integration across Disciplines in the GEOSS Architecture Implementation Pilot. IGARSS, 2013. Available at: https://portal.opengeospatial.org/files/?artifact_id=54812

Percivall, G. and Browdy, S. (2012). GEOSS Design and Interoperability. GEO Work Plan Symposium, April 2012. Available at: ftp://ftp.earthobservations.org/201205_Work_Plan_Symposium/Presentations/3_Work%20Plan%20Review/Infrastructure/GEO_WPS1205_Presentation_IN-05.pdf

Consulted websites

AARC: https://aarc-project.eu/

AgriSemantics: http://agrisemantics.org/

AgroVoc: http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

BlueBRIDGE: http://www.bluebridge-vres.eu/

CAB: http://www.cabi.org/cabthesaurus/

Copernicus: http://www.copernicus.eu/

DARIAH: http://www.dariah.eu/

D4Science: https://www.d4science.org/

EGI: https://www.egi.eu/

ELIXIR: https://www.elixir-europe.org/

ESFRI: http://www.esfri.eu/

EUDAT: https://eudat.eu/

e-IRG: http://e-irg.eu/

GACS: http://browser.agrisemantics.org/gacs/en/index

GEANT: https://www.geant.org/

GEOSS: http://www.geoportal.org/

GODAN: http://www.godan.info/

Helix Nebula: http://www.helix-nebula.eu/

Humanities at scale: http://has.dariah.eu/

iMarine: http://www.i-marine.eu/Pages/Home.aspx

NAL: https://agclass.nal.usda.gov/

OpenAIRE: https://www.openaire.eu/

OpenMinTeD: http://openminted.eu/

PRACE: http://www.prace-ri.eu/

Research Data Alliance: https://www.rd-alliance.org/

SeaDataNet: https://www.seadatanet.org/

WheatIS: http://wheatis.org/

Zenodo: https://zenodo.org/