

ROSA position paper

In 2030 food systems will produce healthy nutritious foods, produced through a input-efficient methods and supporting a thriving environment. Food Systems will operate as collaborative networks, that are constantly improving to improve their economic, environmental and social performance for all actors in the network, and those food systems can be region specific, in local territories, but also global. The food systems contributes to the achievement of a wide range of objectives as captured by the SDG's and contributes to mitigating climate change impacts, and adapting to negative consequences of climate change.

Agri-food science are complex from a data science point of view, with different disciplines (from genomics to social sciences), different scales (from genes to ecosystems) and geolocalisation. The ability to integrate these heterogeneous data is a key issue to tackle new societal challenges (e.g. food security in the face of climate change, sustainable food value chains, digital agriculture and food technology). In addition, the automation of data collection, new techniques in “omics” as well as the development of new types of data sources (e.g. Internet of Things, crowd-sourcing, text mining) has allowed to collect an exponentially increasing amount of data. Thus, there is a need for an e-infrastructure that connects data, infrastructures, resources and people and that allows to share efforts and expertise, and support innovation. In particular, such an e-infrastructure can develop Virtual Research Environments (VREs) that provide tailored solutions for specific communities.

Our vision is an ecosystem of people, data and devices. Stakeholders are not only academic researchers. The farm has to become part of the research environment as the supply chain needs to feed back to the lab and to the farm. Last but not least citizens will participate in the collection of data and information.

The paradigm change in science from “mostly data production” to “more and more data analysis” will need new skill sets. The combination of data from lab, farm, supply chain, satellites and citizens will create new insights and new possibilities for data analysis. Data Scientists or knowledge engineers, are needed not only in the lab, but also within the industry and in the service of decision makers.

Facets of a data ecosystem and implementation status in the AgriFood domain

Item	Status and needed action
Network of trustworthy repositories	There are scattered repositories all over the landscape which need to be organized in the network
A registry of trustworthy	CIARD RING as prototype is powerful tool to organize the

repositories	network of repositories
Repository API/PID Registration	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners
DOI Registration	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners
Metadata Schemes registration	Work is very advanced with the former VEST registry. Coordination and discussion with other similar enterprises in other domains is needed
Concept server and registration system	The domain is comparatively advanced having available operative prototypes of Ontology Servers (Agroportal) and Concept Servers (GACS)
System of Authorisation records	No domain specific activities should be undertaken, collaboration should be seared with other EOSC partners
System of licence registries	No domain specific activities should be undertaken, collaboration should be seached with other EOSC partners
Ecosystem of tools and operating procedures	Some scattered work has been done but without systematization; it needs to be integrated with the effort to create competence centers

In implementing this vision we are not at the beginning. As a community of interest involved in projects as agInfra, semagrow, openminted, aginfra+ we have :

- introduced agricultural scientific communities to the vision of open and participatory data intensive science
- Served as a proof of concept for a European thematic hub for data powered research in agriculture and food
- Demonstrated how grid and cloud based services can be used by scientists and information managers
- Provided a set of domain specific recommendations for the open publication of research outcomes

Particularly we have created infrastructure elements for the community

- The Agrisemantics framework, with Agroportal, the GACS and the Vest Registry as efficient storage, development and access points for common semantics
- Use case driven e-Infrastructure demonstrations for key thematic areas(a) Water safety, Food microbial biodiversity, Gene regulation networks in cultivated plants, phenotypes in Wheat

The next steps are going to the establishment of an agINFRA cloud within EOSC in 2020-2022 to target a sustainable infrastructure,

- Building on top of the AgInfra+ infrastructure
- Implementing the eROSA roadmaps
- Run pilots in larger scale based on selected uses cases across the food system
- Showcase the operational mode of EOSC (eg. thematic EOSC marketplace for cloud&data services
- Focus on the H2020 DT-SFS-26-2019: Food cloud demonstrators

Author(s)

Johannes Keizer, eROSA Project