Smart Rural Areas Data Infrastructure (SRADI) – an information logistics framework for digital agriculture based on open standards

David Gackstetter[1], Mandana Moshrefzadeh[1,2], Thomas Machl[1,2] and Thomas H. Kolbe[1,2]

Abstract: Agricultural research is embedded in a complex system of dynamically changing and interacting elements and subsystems. This includes stakeholders and data from various disciplines and institutions, and consequently a highly distributed nature of data resources. To solve the challenges coming along within this ecosystem, we introduce an interdisciplinary cooperation and information logistics framework named "Smart Rural Area Data Infrastructure" (SRADI) for research topics in agricultural sciences. This framework interconnects resources between different stakeholders and platforms, while supplying a basis for integrating both pre-existing, historical and real-time information for physical things. By this, SRADI provides both a quantitative and qualitative enhancement of data foundations for agricultural research.

Keywords: agricultural landscape, distributed digital twin, data infrastructure, catalog, IoT

1 Introduction

Agricultural landscapes represent a highly complex system with dynamically changing elements and interacting subsystems. Single components are managed and owned by a multitude of stakeholders from various fields and disciplines. In order to efficiently and sustainably understand, manage, and operate such a complex system, it is critical to understand the interconnections between the relevant disciplines from different viewpoints of the involved stakeholders. Within this system, actors supply comprehensive specialized information, which is typically available in specialized information systems and spread over various organizations. Consequently, a primary challenge consists in the demand-oriented availability and the harmonization of the information across various stakeholders, to the extent that it helps to establish a common and cross-scale understanding of the landscape and its components.

[1] Technische Universität München, Hans Eisenmann-Forum für Agrarwissenschaften, Liesel-Beckmann-Str. 2, 85354, Freising, david.gackstetter@tum.de
[2] Technische Universität München, Lehrstuhl für Geoinformatik, Arcisstr. 21, 80333, München, mandana.moshrefzadeh@tum.de, thomas.machl@tum.de, thomas.kolbe@tum.de

2 Smart Rural Data Infrastructure (SRADI)

In order to address the above-mentioned requirements within the domain of agricultural research, we have developed a concept named “Smart Rural Area Data Infrastructure – SRADI” [Mo20]. This concept takes into account not only the technical aspects, but also the organisational level and human interactions within the whole system, by considering on the one hand roles, interests, and requirements of stakeholders, and on the other hand the existing and required systems and services. However, SRADI is not a single platform, but rather an infrastructure of various and distributed platforms and resources – an information logistics framework. A conceptual representation of SRADI is shown in Fig. 1. The SRADI concept is structured into three main pillars: i) the actors or stakeholders from various disciplines, which have different interests and roles in the whole systems, as well as owing or maintaining several resources and information; ii) the applications, which form an interface between actors and the digital replica of the agricultural landscape; and iii) the so called “distributed Digital Twin (DT)”. The SRADI distributed DT of the agricultural landscape handles the distributed information resources across different stakeholders and platforms, while providing a basis for integrating both already existing and real-time information for physical things such as landscape objects. The distributed digital twin of SRADI itself consists of four more components: a) Virtual Landscape Model, which is a digital spatio-temporal representation of the agricultural environment; b) Analytic Toolkit, which is in fact a warehouse of existing modelling, analysis and simulation tools, linked to the rest of the SRADI components using standardized service interfaces; c) Sensors and Internet of Things (IoT) comprising all available in-situ and remote sensors.

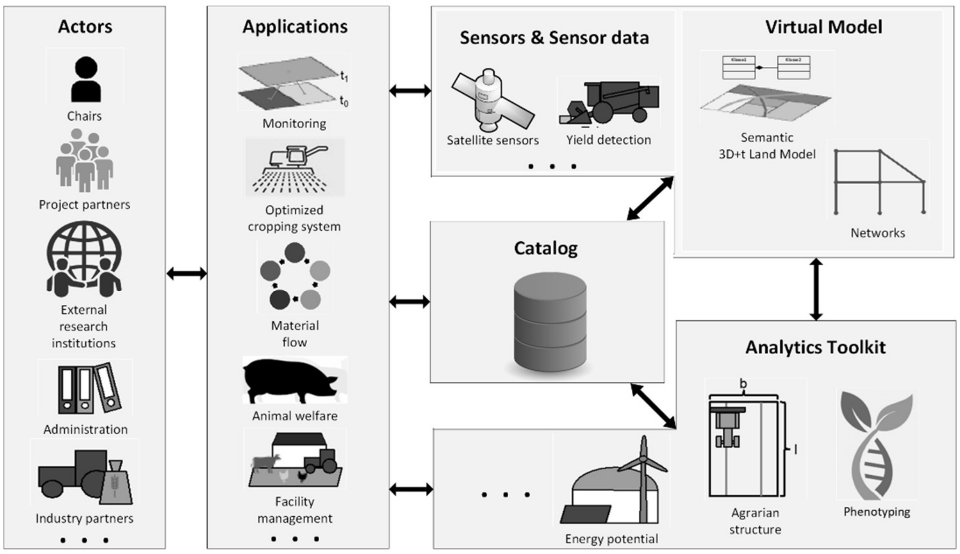

Fig. 1.: The Smart Rural Area Data Infrastructure – SRADI Concept

This also includes data from third party providers, which are used for enriching the DT with real time information; and d) Catalog that provides on the one hand the basis for linking diverse types of resources to each other and on the other hand represents the complex interconnections of resources as well as their owners and organizations. One important aspect of SRADI is interoperability. The components, data models and software are designed and chosen based on their usage of international standards such as the ones from Open Geospatial Consortium (OGC) and International Organization for Standardization (ISO).

3 Handling information resources of the distributed digital twin

The catalog is a key component of the distributed DT of SRADI. Its particular features distinguish it from the other typical catalogs, in for example spatial data infrastructures (SDIs) or governmental open data platforms. The distributed DT catalog specifically addresses these key challenges:

- **Lack of transparency:** different stakeholders have different information about the same real-world objects. They typically do not know what data is available from other sources.
- **Data integration in a multidisciplinary and distributed environment:** even if different stakeholders know about the data of others, the way they perceive information related to the same real-world objects, is not necessarily the same. This makes integration of information difficult and causes problems when linking the data of multiple actors.
- **Digital representation of real-world objects:** in order to integrate various information which is related to the same real-world objects it is required to have a unique and sustained digital representation of those real-world objects. However, a big challenge consists in determining the level of details to digitally represent these real-world objects, taking into account that each should have a unique stable identifier.
- **Data persistence**: it is important to guarantee the uniqueness and persistence of the information, which includes not only the digital representation for real-world objects in the catalog, but also entries that reference historical and real-time information and resources.
- **Interoperability and vendor neutral**: in order to serve different stakeholders from various disciplines and domains it is essential to provide a solution that is interoperable and platform agnostic.

Given these challenges, we propose to set up a catalog as a registry for resources of all types and for their providers (stakeholders). In contrast to the existing catalogs, which are used in for example open data platforms, this catalog is not limited only to managing datasets and services carried out or collected during different spatio-temporal ranges, but in addition handles functions at the level of individual physical things (like an individual agricultural parcel or machine) as well as information related to projects, methods, etc. Hence, the catalog of the distributed DT covers diverse types of resources including methods, projects, software, applications, devices, landscape objects and even animals. By explicitly defining the type of each resource, we add specific semantics related to the digital twin of the agricultural ecosystem. This improves structuring and retrieval of information provided by diverse organizations.

Another important characteristic of this catalog, which differentiates it from other existing catalogs, is its role in facilitating the SRADI distributed DT. In order to achieve the full benefits of the distributed DT in our approach, for each individual landscape object we define a digital geo-replica as a single entry using data provided by the government (geo-base data). All these objects have stable identifiers using official geo-data, thus, each of them behaves as a basis for integrating and relating other information resources to each other. The use of official identifiers ensures that established relationships will remain valid during the life cycle of the real-world objects. Such an approach helps us to be able to not only better understand and maintain the life cycle of the registered resources as single DTs but also discover the effects of distributed DTs on each other through the geo-temporal relationship between individual landscape objects. In addition, the catalog establishes semantic relations between these diverse distributed resources. The semantic interconnections between datasets help, on the one hand, to link related information resources to each other, and on the other hand, to give meaning to those interconnections by providing specific types of relationships such as "depends_on", "child_of" or "links_to". Moreover, the semantic relationships are also used to show how organisations are connected. It has to be noted that information resources in the catalog are directly managed by their own organizations, and the catalog manages the interactions between the organizations and stakeholders that are involved in the whole process.[2]

The catalog and its metadata model are designed taking into account the key aspects of the well-known metadata model standards such as Dublin Core, ISO 15119 and DCAT version 2. However, in contrast to existing catalogs, which are used in for example open data platforms, this catalog is not limited only to managing datasets and services, but it additionally handles functions at the level of individual physical things (like an individual agricultural parcel or machine). The catalog covers diverse types of resources from methods, projects, devices to the raw data, which are carried out or collected during different spatio-temporal ranges and owned or maintained by different stakeholders. Moreover, the management of these information resources in the catalog is strictly administered by their own organizations and hence, the catalog manages stakeholders and organizations that are involved in the whole process.

[2] The key characteristics and their implementations of the distributed DT catalog are shown in this video: https://youtu.be/YuEAyZ6Vm8Y

4 Integrating real-time information via remote access

Another important component of the SRADI distributed digital twin refers to the integration of real-time information on landscape objects. Critical factors in agricultural research can vary within a short time; delays of adequate reaction to these changes can potentially cause detrimental impacts. From a real-time connection to sensors in the physical landscape, researchers obtain temporally high-resolution, quantitative information about actual, prevalent conditions. Such information substantially improves the basis towards more efficient and effective planning.

In agricultural research, proprietary and non-interoperable systems represent a major barrier to adopting sensor technology. Facing this situation, our institute pursues developing a collaborative, high-performance, and affordable technological concept for sensor network infrastructures that is exclusively based on open and standardized technologies. The developed sensor data infrastructure is generally open to any wireless technology (Fig. 2). In addition to the existing solutions around mobile radio and WLAN, we decided to offer a radio network solution based on the LPWAN standard LoRaWAN, due to its power efficiency, long transmission range, communication on license-free frequency bands and low-cost hardware equipment [Lo15]. An open-source implementation for LoRaWAN from The Things Industry – The Things Network Stack – can here be used as suitable network server. The data management system builds upon OGC Sensor Web Enablement (SWE) standards, to ensure interoperability between different platforms and data sources. From the scope of the SWE standards, we chose the SensorThings API as the core element

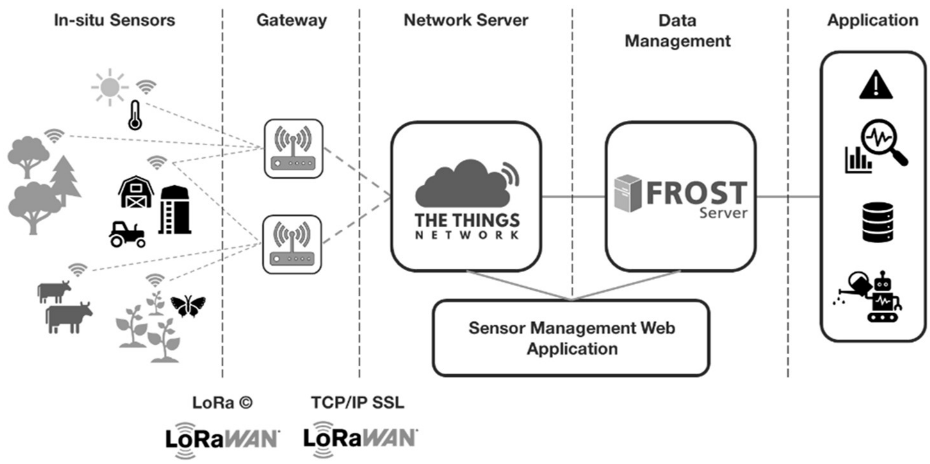

Fig. 2: End-to-end sensor data infrastructure with open-source implementations for network (TheThingsNetwork) and data management servers (FRaunhofer Opensource SensorThings-Server = FROST-Server) using the example of LoRaWAN-radio technology

for managing data streams [Li16]. This standard provides an open and consistent framework for connecting IoT devices, data and web applications. Accordingly, the proposed infrastructure integrates applications that implement this API, but can also be networked with other types of interfaces through existing adaptation tools. A crucial aspect also concerns the management of user access and identity. To ensure full trustworthiness, security and transparency for data management, all components are reliably secured by open-source, state-of-the-art security technologies.

5 Conclusions

In this paper we described our concept of SRADI as an information logistics framework for the digital transformation of agricultural research. The concept has been developed in the context of agricultural science and landscape research and is based on the requirements of a multidisciplinary agricultural research centre. This paper's main contributions are: 1) Development of a metadata model that provides support to the management of distributed and diverse types of information resources, which is the basis for the catalog. The catalog serves as a platform for a transparent exchange and integration of information. In order to ensure the interoperability, the metadata model is mapped and hence compliant with the DCAT version 2. 2) An Internet of Things (IoT) infrastructure that links real-time information from different real-life sources with a secure and user-friendly end-to-end data management system. 3) An implementation of the distributed DT, which relies on widely used open and international standards and open-source software, and a demonstration of the DT's capabilities and benefits for agricultural research. To further enlarge the information logistics framework, our future research will focus on integrating simulation models and analysis tools into the SRADI–distributed DT.

References

[Mo20] Moshrefzadeh, M. et.al.: Towards a Distributed Digital Twin of the Agricultural Landscape. Journal of Digital Landscape Architecture (5), 2020, doi:10.14627/537690019.

[Li16] Liang, S., et al.: OGC SensorThings API, Open Geospatial Consortium, 2016, [Online] Available at: https://www.opengeospatial.org/standards/sensorthings. Accessed: 05.09.2020.

[Lo15] LoRa Alliance: A technical overview of LoRa® and LoRaWAN™, LoRa Alliance, [Online] Available at: //lora-alliance.org/sites/default/files/2018-04/what-is-lorawan.pdf. Accessed: 12.09.2020.