Position paper - NFDI4Agri

Research Data Infrastructure for Agricultural and Soil Sciences

Priorities of Agricultural and Soil Sciences

Agriculture is central to all people. One focus of agricultural research is on ensuring food security through resource-efficient, sustainable food production (**bioeconomy**). As agricultural land is finite and competition with feed, bioenergy and raw material production becomes more important, **food production** per unit area must continue to **increase in the future**. Agricultural research develops and investigates strategies to increase production while preserving the ecosystem services of the soil as well as the accompanying landscape compartments water, biodiversity and climate in the long term („**sustainable land-use intensification**"). Agricultural science covers a **broad and heterogeneous field of research disciplines relevant for the 'NFDI4Agri' consortium** such as soil science, plant and animal production, agricultural economics and -sociology, and plant physiology[1]. Current research questions, among many others, deal with the impact of land use change and intensification on soil functions, plant development and production, the development of resistant seeds, the accurate application of fertilizers and pesticides (precision farming), the protection of biodiversity in agricultural landscapes, the development of adaptation strategies to climate change, studies of nutrient cycles, and improvement of animal welfare.

Research data diversity in Agricultural and Soil Sciences

High diversity of data is generated in agricultural research in terms of content and, additionally, in structure due to the wide range of research questions among the various disciplines. In agricultural field measurements, data is collected to describe the physical condition and development of plants as a function of soil, management and climate. Agricultural sciences have to deal with diverse data sources originating from laboratory analysis, landscape monitoring, sequencing, phenotyping, spectrophotometer measurements, remote sensing, economic market data or modelling results. Agricultural data is usually both, spatial and temporal referenced as **geodata** (soil maps) and **time series** (yield data). In this context, agricultural long-term experiments, some of which have been carried out for more than 100 years, are of particular importance.

In addition to research data, agricultural sciences also produce **sensitive information** about landowners' areas. Thus, aspects of **data protection** must be considered. As a result of increasing use of sensor technology and the upcoming **digitization** in agriculture (**Farming 4.0**), the production of spatially and/or temporally high-resolution data (**big data**) increases considerably. Furthermore, data analyses in the different disciplines are highly diverse, so that a wealth of different methods is available to deal with generated research data. The high diversity of agricultural data, different use conditions and applied standards represent a particular challenge for a research data infrastructure.

[1] categories as provided by the DFG

Current situation and future requirements

The national agricultural research disciplines face a **heterogeneous** landscape of data infrastructures that vary in their way of supporting common standards in the field of research data management (RDM). However, the integration of these infrastructures as national nodes for existing physical and informal exchange platforms is still lacking. On an international level, there are many efforts in networking of soil and agricultural **data and knowledge**. We emphasize Global Soil Partnership (GSP) of FAO, Global Open Data for Agriculture & Nutrition (GODAN), World Soil Information (ISRIC), IGAD Interest Group Agricultural Data within the Research Data Alliance (RDA), the Global Long-term Field Experiment Network (GLTEN), Agricultural Model Intercomparison and Improvement Project (AGMIP), the European life-sciences Infrastructure for biological Information (ELIXIR) and Modelling Agriculture with Climate Change for Food Security (MACSUR).

Only few infrastructures exist within the various disciplines to make agricultural research data publicly available and reusable. The repository of the **BonaRes Data Centre** located at ZALF Müncheberg and part of the BonaRes initiative 'Soil as a sustainable resource for the bioeconomy' (www.bonares.de) - is one example for a spatial data infrastructure specialized in the provision of open soil and soil-related agricultural research data. Furthermore, the IPK Genebank informations system GBIS is a safeguard of plant biodiversity and is an important source of information for plant researchers as well as crop breeders. It provides information to crop plants and their wild relatives that is conserved in IPK Genbank, Gatersleben. This includes conservation, characterization and evaluation data of the collection. Complementing this resource, a comprehensive Laboratory Information Management System records and documents data from research studies that uses Genbank material, such as phenotyping and genotyping data.

Today, there is no organized, interoperable **infrastructure**, which faces the challenges and demands of agricultural sciences and which overarches disciplines. However, there is an increasing trend for inter-/transdisciplinary **data analyses.** Developments of NFDI4Agri build on these demands and will be science-driven while taking into account the specific requirements of different disciplines.

The NFDI4Agri approach

The NFDI4Agri approach requires potential users from the research community and RDM experts from all disciplines of agricultural research to implement common **international standards** to realize **FAIR** (Findable, Accessible, Interoperable, Reusable) and **Linked Open Data principles:** Controlled vocabularies (AGROVOC), certificates (DINI, CoreTrustSeal), licences (Creative Commons), and interfaces (OGC) in the community. These requirements for the provision of research data are indispensable for the scientific reputation and a publicly financed infrastructure for agricultural research data.

Our mission is to meet the needs of the agricultural research community and connect agricultural disciplinary repositories and, hence, make publicly funded and yet isolated research data inter- and transdisciplinary available and thus reusable.

Our challenges are to ...

- ➔ **set-up a flexible, interoperable and scalable data infrastructure** by connecting existing and supporting the creation of new discipline specific infrastructures,
- ➔ **respond properly to future developments** by identifying needs and technological trends of the agricultural and soil community,
- ➔ **integrate legacy research data into the infrastructure and provide it for free reuse.** To this end, concepts are being developed to identify such data treasures, define responsibilities, prepare them and make them available,
- ➔ **assure data quality** by domain specific measures of quality control and set-up a quality feedback and curation system. The evaluation of the data fitness-of-use is an important data quality feature ensuring data reusability,
- ➔ **take educational responsibilities** into account by transferring our knowledge to the next generation of scientists (agricultural informatics), and to
- ➔ **create legal certainty** both on the side of the providers (authors) of agricultural and soil data and on the side of the reusers,
- ➔ **develop privacy and ethical standards** constituting a fair balance between the interests of authors (including their institutions) of agricultural and soil data and of those in the international research communities who wish to reuse the data.

NFDI4Agri and its integration into NFDI and NFDI4Life

The NFDI4Agri consortium defines its **role in the overall NFDI initiative** as part of the NFDI4Life umbrella consortium. Although NFDI4Agri focuses primarily on soil and agricultural sciences, other sectors must also be considered like meteorology, geomorphology, biodiversity. The concept of mutual understanding is also important for the interaction between different NFDI consortia.

Versions

This document is subject to constant changes due to additions by partners and new developments. We keep versions after important changes in order to be able to track the progress.

- **Version 0.6** sent to consortional partners (5.3.2019)
- **Version 0.65** includes the commentary of the partners (8.3.2019)