

# Letter of intent - NFDI4Agri

## 1 Binding letter of intent as advance notification or non-binding letter of intent

<input checked="" type="checkbox"/>	Binding letter of intent (required as advance notification for proposals in 2019)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2020)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2021)

## 2 Formal details

### Planned name of the consortium

NFDI for Agricultural Sciences

### Acronym of the planned consortium

NFDI4Agri

### Applicant institution

Leibniz Centre for Agricultural Landscape Research (ZALF)

Prof. Dr. Frank Ewert

### Spokesperson

Dr. Uwe Heinrich, uheinrich@zalf.de, Leibniz Centre for Agricultural Landscape Research (ZALF)

## 3 Objectives, work programme and research environment

### 3.1 Research area of the proposed consortium

Agricultural sciences comprise a broad and heterogeneous field of research disciplines such as soil science, plant production and livestock farming, applied genetics and physiology, agricultural economics and sociology. The proposed consortium corresponds to the DFG Research Area 23: Agriculture, Forestry and Veterinary Medicine.

### 3.2 Concise summary of the planned consortium's main objectives and task areas

The NFDI4Agri consortium aims at making agricultural research data findable, accessible, interoperable, reusable and, moreover, open, thus thoroughly implementing the FAIR principles. NFDI4Agri will meet the needs of the agricultural research community and set-up a flexible, interoperable and scalable data infrastructure by connecting existing disciplinary repositories and, hence, making publicly funded and yet isolated research data interdisciplinarily and sustainably long-term available. Educational responsibilities are assumed by transferring our knowledge to the next generation of agronomists. Research data quality must be assured by domain-specific measures of data quality control and curation systems. The evaluation of the data fitness-for-use is an important quality feature to attract future data reuse. Privacy and ethical standards will be developed, constituting a fair balance between while ensuring the interests of authors and users of agricultural data, including concepts for handling sensitive data. These overarching goals will be implemented within six task areas.

### **Task area 1: Coordination**

*Spokesperson: Uwe Heinrich (ZALF)*

TA1 is responsible for the coordination of the entire consortium. This initially includes the coordination and submission of the application. During the funding period, the coordination team will be responsible for developing a cooperation agreement and a governance structure. TA1 is responsible for allocating the funds to the co-applicants and the proper use of the funds. This also includes monitoring the implementation of the developed use cases. The spokesperson represents the consortium externally and is responsible in particular for cooperation with the other consortia and NFDI bodies. The tasks of the coordination team also include the support of the web presence for the external presentation of the consortium. In order to support cooperation within the consortium, a personalized work area is set up for the members.

### **Task area 2: Agricultural Community**

*Co-Spokespersons: Ulrike Stahl (JKI), Ulrich Meyer-Doeringhaus (ULB), Jens Freitag (IPK), Nikolai Svoboda (ZALF), Birte Lindstädt (ZB MED), Matthias Senft (ATB), Ulrich Schurr (FZJ)*

TA2 focuses on the involvement of the agricultural community, i.e., by raising awareness of the activities of NFDI4Agri, getting feedback of developed and planned actions, implementing and disseminating developed concepts or services. The main target group is both established and young scientists, which will be addressed via professional societies and agricultural faculties. TA2 aims at the development of procedures to involve other stakeholders, i.e., farmers and consultants. Special measures of dissemination and outreach in the sense of “train-the-trainer” will be developed. Additional activities will include further development of curricula of faculties, preparation of training material and e-learning tools, the creation of a model for dissemination at universities through a system of data stewardship with different roles and task areas, offering workshops to the whole community, and organization of hackathons and summer schools. Community feedback on NFDI4Agri developments will be collected by conducting surveys with questions about further needs for and usefulness of already implemented services, and by collecting research questions together with the stored data to identify and predict future needs.

### **Task area 3: Data Fitness for Use in Agricultural Research**

*Co-Spokespersons: Jan-Henrik Hauer (UB), Uwe Rascher (FZJ), Carsten Hoffmann (ZALF)*

TA3 focuses on the definition of domain-specific criteria of data quality and quality levels as well as the identification of procedures for quality assurance (QA) to facilitate information. This may address the data's fitness for use, e.g. for machine learning approaches in deep phenotyping of plants and farm animals, or for food and nutrition simulations. The current research data infrastructures do not satisfactorily support searching for data that meet task-specific requirements. TA3 will establish a discussion forum with agronomists, machine learning-, and modelling experts, who produce and reuse agricultural research data, to identify common tasks and requirements, and develop appropriate data quality criteria. Domain specific and transferable QA methods and recommendations will be defined to deal with common practices of data curation and as guidelines for future agricultural research projects (e.g. a web-based quality feedback and review system with appropriate visualizations of the data).

### **Task area 4: Standards and Interoperability - Making Agronomic Data FAIR**

*Co-Spokespersons: Claus Weiland (SGN), Daniel Martini (KTBL), Juliane Fluck (ZB MED), Norbert Reinsch (FBN), Ralf Kunkel (FZJ)*

TA4 aims at identifying, developing and providing standards and interfaces for the development of an interoperable agricultural research data infrastructure. Key foundations for data integration, interoperability and reuse are open and accepted standards, persistent unique identifiers and domain-specific thesauri. Building from existing standards, TA4 will establish new interoperability layers for extended data exchange between agronomic data providers. The

application of modern high-throughput methods, e.g. sequencing and genotyping, results in large amounts of diverse data. For the project as a whole, the planned work sequence and its completeness must be documented and visible to all cooperation partners. Distributed ledgers may offer good technical possibilities meeting these requirements by capturing metadata in real time for defined analysis steps. Domain-specific and standardized methods, interfaces, metadata schemas and ontologies must be mapped and/or aligned. Workflows and technical instructions will be developed that enable collaborative data curation and annotation. The compliance of annotation pipelines will be evaluated by a GO FAIR implementation matrix.

#### **Task area 5: Agronomic Data and Infrastructure Services**

*Co-Spokespersons: Matthias Lange (IPK), Xenia Specka (ZALF), Juliane Fluck (ZB MED), Birte Lindstädt (ZB MED), Ulrike Stahl(JKI), Björn Usadel (FZJ), Jan Vanderborght (FZJ)*

TA5 focuses on the provision of long-term preserved research data in an interoperable and scalable infrastructures landscape of (i) repositories and (ii) infrastructure services. Repositories are used for the persistent deposition and publication of research data and documents. Infrastructure services include elements necessary for the interoperable and sustainable operation of repositories under FAIR criteria, also covering technical interfaces, documentation, consulting and operational monitoring. Actions will be derived to upgrade repositories on a technical and conceptual level. Later, the step-by-step establishment of a common access point to the distributed infrastructures will be designed and implemented. This will be done by aligning with the use cases to enable early feedback from the community. Along these technical tasks, joint concepts for ensuring service quality, technical support desks and metrics for operational monitoring will be developed and transferred to a concept for a long-term operating model.

#### **Task area 6: Legal, Ethical and Organizational Aspects**

*Co-Spokespersons: Franziska Boehm (FIZ), Peter Kostädt (UP), David J. Russell (SGN)*

Agricultural data is often collected on and/or pertains to private land, particularly farms, and farm animals. It is often spatially explicit, identifiable from contextual information that must be provided for reusability. Privacy requirements are essential for data publication and access, ensuring open data access and intellectual property rights of data providers. TA6 analyses the privacy framework for agricultural data with spatial reference and develops guidelines for data access that cannot be reliably anonymized. It moderates a participatory process to develop best practice guidelines for researchers, data centers/repositories and institutional data contacts, taking into account research ethics, legal advice, organizational aspects, and the experience and best practices in other disciplines, particularly social and economic sciences.

### **3.3 Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives**

One main objective of NFDI4Agri is to make existing, distributed and disciplinary infrastructures for agricultural research data interoperable and connect them into a superordinate exploration platform. Concepts and methods will be developed by analysing current capabilities and implementations of relevant disciplinary infrastructures of consortium partners, e.g. the open access publishing platforms for life sciences ([PUBLISSO](#)), the [BonaRes repository](#) for soil- and agricultural research data, the open access repository [OpenAgrar](#), the German Genebank ([GBIS](#)), the electronic data archive library for plant genomics and phenomics ([e!DAL-PGP](#)), the distributed TERENO data infrastructure [TEODOOR](#) providing longterm data on the ecological, social and economic impact of global change at regional level, the European Search Catalogue for Plant Genetic Resources ([EURISCO](#)), the soil-zoological database [Edaphobase](#) and [plaBiPD](#), a database that hosts data functional annotation of important plant genomes. These infrastructures will be pioneering for method development and for testing its practical

implementation. If practicability is assured, further existing infrastructures within the consortium (e.g. the [Research Centre for Remote Sensing in Agriculture](#) (FLF), the [GeoPortal.JKI](#), the [Plant Health knowledge portal](#) providing information on Germany's phytosanitary measures, the German Plant Phenotyping Network ([DPPN](#)), EMPHASIS which provides a link to the European Open Science Cloud (EOSC) within the [EOSC-Life](#) project, the [ALPS online database](#) for non-chemical plant protection measures information, or [FarmPheno](#) database genotype-phenotype-distinction of farm animal biology,) can adopt new requirements to become part of the distributed, but interoperable network. Currently, many gene banks (e.g. [GBIS](#), the [German Gene Bank Grape](#) or European Vitis Database ([EU-VITIS](#)), Vitis International Variety Catalogue ([VIVC](#)), the International Data Base for Beet ([IDBB](#)), European *Avena* Database ([EADB](#)) or the [German Gene Bank Fruit](#) are integrated in decentralized gene bank networks. Concepts and methods for the realization of such networks will be used to make other disciplinary infrastructures of the consortium interoperable.

The free online tool [CADIMA](#) supports systematic reviews and maps by increasing the efficiency of the evidence synthesis process and facilitating the reporting of activities to maximise methodological rigour. The [SYNOPS](#) web tool enables the assessment and comparison of the potential environmental risk of crop protection application scenarios under realistic conditions. The [EMRA](#) system helps farmers and consultants to manage extreme weather events in Germany. The Jülich SuperComputing Centre (JSC) provides one of the most powerful computing infrastructures that can be used by scientists and modellers to dynamically solve highly complex problems with simulations, calculations and comprehensive analyses of big data.

Developed documents and training material can be published on existing community-specific publication platforms, e.g. [PUBLISSO Life Science Repository](#) or [OpenAgrar](#). The creation of data management plans will be supported by the existing research data management organiser for life sciences ([RDMO4Life](#)). Standardized exchange formats and interfaces for (meta-)data exchange must be defined based on existing standards, e.g. [OGC CSW](#), [OAI-PMH](#), or interfaces, e.g. the Breeding API ([BrAPI](#)). The international ontology standard for plant phenotyping data MIAPPE needs to be developed within a German background (MIAPPE-Germany) based on the European interactions between EMPHASIS and ELIXIR. Services for alignment and mapping of existing disciplinary and widely used ontologies, e.g. [AGROVOC](#), Provenance Ontology ([PROV-O](#)), the Biological Collections Ontology ([BCO](#)), or [Crop Ontology](#), [ATOL](#), will be used to ensure data interoperability within NFDI4Agri. The semantic lookup platform SemLookP for ontologies and terminologies is a central element of semantic data integration and will be further extended to connect different, disciplinary knowledge domains.

### **3.4 Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration**

NFDI4Agri is currently in contact with the following consortia with regard to cooperation opportunities:

- NFDI4Health
- DataPlant
- NFDI4Microbiota
- NGFDI4Life Umbrella
- NFDI4BioDiversity
- NFDI4Objects
- NFDIKonsortSWD
- NFDI4Earth

As part of the life sciences, the agricultural sciences naturally have close links to NFDI4Microbiota, NFDI4Health, NFDI4BioDiversity and NFDI4Life Umbrella, but also to the geosciences and the social sciences. DataPlant considers fundamental plant research, so that a cross-link at an early stage is important. Explicit spatial reference is an essential characteristic of agricultural science research data, so that close cooperation with NFDI4Earth is necessary in this respect. Social science data are also used and generated in agricultural research, so that cooperation with NFDIKonsortSWD will be highly appropriate as well.

There is a clear intention to cooperate in all consortia mentioned, whereby the interfaces and the topic areas still have to be clarified in detail.

## **4 Cross-cutting topics**

### **4.1 Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.**

NFDI4Agri has identified the following cross-cutting topics:

1. Education & training
2. Data quality
3. Metadata
4. Standards
5. Services
6. Interfaces
7. Interoperability
8. Knowledge transfer
9. Data security and data ownership
10. Long-term archiving
11. Reputation & citation
12. Legacy data
13. Legal and ethical aspects

### **4.2 Please indicate which of these cross-cutting topics your consortium could contribute to and how.**

Most of the above-mentioned cross-cutting topics are dealt with under the aspects of agricultural sciences in the working areas of NFDI4Agri. The generic solutions can be used by other consortia or integrated into overarching solutions. The following connection exists between the cross-cutting topics and the task areas: TA2 (1, 4, 8 and 11), TA3 (2, 3, 4 and 11), TA4 (3, 4, 6, 7 and 11), TA5 (4, 5, 8 and 11), TA6 (4, 9, 11 and 13).

An overarching contribution can be made, in particular, to the cross-cutting topics 9 “Data security and data ownership” and 13 “Legal and ethical aspects”: The specifics of agricultural data, often collected on and/or pertains to private land, particularly farms, and farm animals are also relevant to some data from earth, environmental, and biodiversity sciences. The resulting privacy requirements are essential, in particular for data publication and accessibility. Furthermore ethical and organizational challenges have to be taken into account. Some of these challenges are similar to those relating to geodata, environmental information and agricultural statistics from government agencies. However, the legal exemptions that govern those do not usually apply to research data. Legal, ethical and practical guidance geared towards the social sciences and economics, in turn, does not usually take the specifics of agricultural data into account. Addressing these issues within NFDI4Agri therefore provides benefits to the NFDI as a whole.