FAIRagro

Letter of intent - FAIRagro

1   Binding letter of intent as advance notification of a full proposal

x	Binding letter of intent (required as advance notification for proposals in 2021)

2   Formal details

Planned name of the consortium

FAIR Data Infrastructure for Agrosystems

Acronym of the planned consortium

FAIRagro

Applicant institution

Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, 15374 Müncheberg

Prof. Dr. Frank Ewert

Spokesperson

Prof. Dr. Frank Ewert, frank.ewert@zalf.de, Leibniz Centre for Agricultural Landscape Research (ZALF)

Co-spokesperson

Dr. Xenia Specka, xenia.specka@zalf.de, Leibniz Centre for Agricultural Landscape Research (ZALF)

Co-applicant institution

Federal Research Centre for Cultivated Plants (JKI), Erwin-Baur-Straße 27, 06484 Quedlinburg, Prof. Dr. Frank Ordon

Co-spokesperson:

- Dr. Til Feike, til.feike@julius-kuehn.de, Federal Research Centre for Cultivated Plants (JKI)
- Dr. Ulrike Stahl, ulrike.stahl@julius-kuehn.de, Federal Research Centre for Cultivated Plants (JKI)

Deutsche Forschungsgemeinschaft
Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany
Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

DFG

Co-applicant institution

Information Centre for Life Sciences (ZB MED), Gleueler Straße 60, 50931 Köln, Prof. Dr. Dietrich Rebholz-Schuhmann

Co-spokesperson:

- Prof. Dr. Juliane Fluck, fluck@zbmed.de, Information Centre for Life Sciences (ZB MED)

- Birte Lindstädt, lindstaedt@zbmed.de, Information Centre for Life Sciences (ZB MED)

Co-applicant institution

Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL), Bartningstraße 49, 64289 Darmstadt, Prof. Dr. Eberhard Hartung

Co-spokesperson:

- Daniel Martini, d.martini@ktbl.de, Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL)

Co-applicant institution

Senckenberg – Leibniz Institution for Biodiversity and Earth System Research (SGN), Senckenberganlage 25, 60325 Frankfurt, Prof. Dr. Klement Tockner

Co-spokesperson:

- Dr. Claus Weiland, claus.weiland@senckenberg.de, Senckenberg – Leibniz Institution for Biodiversity and Earth System Research (SGN)

Co-applicant institution

Forschungszentrum Jülich (FZJ), Wilhelm-Johnen-Straße, 52428 Jülich, Prof. Dr.-Ing. Wolfgang Marquardt

Co-spokesperson:

- Prof. Dr. Björn Usadel, b.usadel@fz-juelich.de, Forschungszentrum Jülich (FZJ)

Co-applicant institution

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Corrensstraße 3, 06466 Seeland, Prof. Dr. Andreas Graner

Co-spokesperson:

FAIRagro

- Prof. Dr. Jochen C. Reif, reif@ipk-gatersleben.de, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
- Dr. Matthias Lange, lange@ipk-gatersleben.de, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

Co-applicant institution

Leibniz Institute for Information Infrastructure (FIZ Karlsruhe), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Sabine Brünger-Weilandt

Co-spokesperson:

- Prof. Dr. Franziska Boehm, franziska.boehm@fiz-karlsruhe.de, Leibniz Institute for Information Infrastructure (FIZ Karlsruhe)

Co-applicant institution

University of Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Prof. Dr. Dr. h. c. Michael Hoch

Co-spokesperson:

- Prof. Dr. Jan-Henrik Haunert, haunert@igg.uni-bonn.de, University of Bonn

Co-applicant institution

Thünen Institute (Thünen), Bundesallee 50, 38116 Braunschweig, Prof. Dr. Folkhard Isermeyer

Co-spokesperson:

- Florian Hoedt, florian.hoedt@thuenen.de, Thünen Institute

Co-applicant institution

World Agricultural Systems Center of Technical University Munich, Liesel-Beckmann-Straße 2, D-85354 Freising, Prof. Dr. Senthold Asseng

Co-spokesperson:

- Prof. Dr. Senthold Asseng, senthold.asseng@tum.de, World Agricultural Systems Center of Technical University Munich

Participants

- Dr. Andreas Donaubauer, Technical University Munich (TUM)
- Dr. Einar Eberhardt, Federal Institute for Geosciences and Natural Resources (BGR)
- Prof. Dr. Karl-Heinz Feger, The German Soil Science Society (DBG)

DFG

- Dr. Falko Feldmann, German Society for Plant Protection and Plant Health (DPG)
- David Gackstetter, World Agricultural Systems Center of Technical University Munich
- Dr. Markus Gandorfer, Bayerische Landesanstalt für Landwirtschaft (LfL)
- Dr. Christoph Germeier, Federal Research Centre for Cultivated Plants (JKI)
- Dr. Aida González-Mellado, Thünen Institute
- Dr. Carsten Hoffmann, Leibniz Centre for Agricultural Landscape Research (ZALF)
- Dr. Frank Kaspar, Deutscher Wetterdienst (DWD), National Climate Monitoring
- Dr. Benno Kleinhenz, Zentralstelle der Länder für EDV-gestützte Entscheidungshilfen und Programme im Pflanzenschutz (ZEPP)
- Prof. Dr. Ingrid Kögel-Knabner, Technical University Munich (TUM)
- Prof. Dr. Gunnar Lischeid, Leibniz Centre for Agricultural Landscape Research (ZALF)
- Dr. Markus Möller, Federal Research Centre for Cultivated Plants (JKI)
- Kathleen Neumann, GBV Head Office (VZG)
- Prof. Dr. Patrick Noack, Weihenstephan-Triesdorf University of Applied Sciences (HSWT)
- Dr. Daniela Röder, Graduate Centre Weihenstephan (GWZ), Technical University Munich (TUM)
- Dr. Manfred Röhrig, Informationssystem für die integrierte Pflanzenproduktion (ISIP e.V.)
- Dr. Markus Schmalzl, Staatliche Archive Bayern
- Prof. Dr. Karl Schmid, University of Hohenheim
- Prof. Dr. Alexander Sczyrba, Bielefeld University
- Dr. Marco Selig, Deutsches Biomasseforschungszentrum (gGmbH, DBFZ)
- Dr. Matthias Senft, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB)
- Prof. Dr. Steffen Siebert, Georg-August University Göttingen
- Dr. Esther Sundermann, Geschäftsstelle Think Tank Digitalisierung, Thünen Institute
- Dr. Nikolai Svoboda, Leibniz Centre for Agricultural Landscape Research (ZALF)
- Prof. Dr. Hans-Jörg Vogel, Helmholtz Centre for Environmental Research (UFZ)
- Dr. Heidi Webber, Leibniz Centre for Agricultural Landscape Research (ZALF)
- Prof. Dr. Cornelia Weltzien, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB)
- Prof. Dr. Hubert Wiggering, Deutsche Agrarforschungsallianz (DAFA)
- Prof. Dr. Burkhard Wrenger, OWL University of Applied Sciences and Arts (THOWL)

3   Objectives, work programme and research environment

3.1   Research area of the proposed consortium (according to the DFG classification system: https://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/index.jsp)

207 (Agriculture, Forestry and Veterinary Medicine)

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

Research on agrosystems develops strategies for sustainable primary production while maintaining and enhancing ecosystem services and agro-biodiversity. Relevant system levels range from the soil microbiome to landscape scales. Agrosystems research covers a broad field of different disciplines, such as crop production, soil science and plant nutrition, phytopathology and plant protection, applied genetics and plant breeding, but also agroecosystems research, as well as agricultural-, landscape- and geo-ecology. New technologically oriented disciplines in the field of digitalization (robotics, sensor technology, artificial intelligence and modeling) are also receiving increasing attention. Different types of data from diverse sources are generated, collected, structured, and analyzed in the various disciplines. These include crop sequencing and phenotyping data, landscape monitoring and field trial data, remote sensing data, and modeling data with different spatial and temporal resolutions. Increasing use of sensor technologies in the field and in controlled environments, as well as digitization and automation in agriculture (Agriculture 4.0), produce a large amount of valuable high-resolution data. Connectivity and availability of these highly diverse data is not yet ensured by data (infrastructure) providers.

Currently, research data in agrosystems research are predominantly located in poorly networked, disciplinary and institutional infrastructures or are partly unstructured and without metadata in heterogeneous storage systems. Valuable research data, e.g. from long-term studies, are often difficult to access and inadequately described. The handling of sensitive data plays a special role in agrosystems research. Legal rules for the reuse of these data are still inconsistent today and make their access difficult. FAIRagro aims at making agrosystem research data findable, accessible, interoperable, reusable (FAIR) and, moreover, open for research. FAIRagro meets the needs, respects the feedback of the agrosystem research community and sets up a flexible, interoperable and scalable data infrastructure connecting existing repositories thus making publicly funded and yet isolated research data interdisciplinarily and sustainably available. Educational responsibilities are assumed by knowledge transfer to the next generation of agronomists. Research data quality must be assured by domain-specific measures of data quality control and curation systems. Privacy and ethical standards will be developed, striking a fair balance between the interests of authors and

users of research data, including concepts for handling sensitive data. The FAIRagro work plan will be realized in five task areas:

Task area 1: Management and Coordination: TA1 is responsible for consortium coordination and support. It will develop a cooperation agreement and integrate the consortial governance within the overarching NFDI governance. TA1 is responsible for allocating the funds to the co-applicants and the proper use of the funds including 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.

Task area 2: Community involvement and networking: TA2 involves the agricultural community by i) active communication to raise awareness and acceptance of the FAIRagro activities, ii) structured participation process to increase the network by addressing needs, collecting feedback and incorporating further expertises and data infrastructures iii) implementing and disseminating concepts and services with respect to training and education. TA2 will reach scientists via professional societies and graduate centres and assist them directly with the installation of data stewards. Farmers and consultants will be reached by the participation of established central services assisting them. They will be involved in community boards to advise FAIRagro and provide expertise to the cross cutting topic "education".

Task area 3: Standardization, Interoperability and Quality: TA3 develops standards for digital resources toward the principles of FAIR data, focusing in particular on interoperability and reusability as often neglected or underrepresented criteria in the crop domain. It addresses several standards, such as for metadata, (data) publication and data quality. It will draft FAIR Digital Objects (FAIR-DO), including their packaging mechanisms, to generate processing-ready data packets from services and datasets in repositories. By conceptualizing data-curation processes (including automated workflows for data harmonization, validation and assurance of data quality), TA3 lays important foundations for the software services implemented in TA4. TA3 develops practical guidance on agricultural data (personal data and business related data) and ensures that data center policies are mutually compatible and legally sound.

Task area 4: Infrastructure Services: TA4 implements necessary components and infrastructure services of federated research data management (RDM) along the research data lifecycle and is the interface to NFDI cross-cutting technical services. This encompasses a service registry, a search portal, middleware for data access, services for data collection, data structuring, planning and publication, APIs, integrated curated data containers and analysis workflows. Along these technical tasks, concepts for ensuring service quality, technical support and metrics for operational monitoring will be developed and transferred to a long-term operating model.

Task area 5: Use Cases - Realization, Implementation: TA5 will implement FAIRagro standards and services in use cases originating from the agrosystem community. Each use case addresses specific challenges in the agrosystem sciences and will serve as an example for a specific challenge in making agrosystem research data FAIR. Based on use case requirements and challenges, TA5 will develop and apply practical RDM solutions.

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 FAIRagro is to make existing, distributed and disciplinary infrastructures interoperable and connect agrosystem research data within a superordinate exploration platform. We propose to start with well-established repositories and information systems provided by the FAIRagro co-applicant institutions. Covering a wide range of agrosystem research disciplines, they are predestined as pilot infrastructures to build interdisciplinary components upon and to harmonize and provide heterogeneous data within the project. Examples are GBIS/I, as an information system of the IPK ex-situ collection for crop and wild relatives, providing passport data and primary evaluation data; FLOPO from SGN, as a knowledge base of morphological traits and phenotype data text-mined from floras, monographies and thematic databases; the BonaRes Repository from ZALF, as the repository for national soil-agricultural research data; PlabiPD from FZJ with data from plant genomic, transcriptomic and ontology for crops; and the National Soil & Forest Inventories (Thünen), which include federal inventories with datasets for agricultural and forest areas (soil parameters, forest status and forest production potential). The cluster of excellence "PhenoRob - Robotics and Phenotyping for Sustainable Crop Production", a collaboration between UB and FZJ, is contributed to FAIRagro. PhenoRob aims at developing methods and new technologies that can be used to observe, analyze, better understand and treat plants.

Additionally to domain-specific services, general life-science related infrastructures like LIVIVO, a search engine for literature and information, and SemLookP, a repository for ontologies and terminologies, from ZB MED will be interconnected. As building blocks for future infrastructures state-of-the-art technologies will be used: e!DAL as an open infrastructure to store on-site research files and integrate them into data publishing; Research Data Management Organizer (RDMO) for data management planning; The OSGeo based Thünen Institute Scientific Data Repository (TISDAR) as an open-source application stack to provide OGC conformant data repositories with focus on spatial data for researchers. Additionally, educational and consulting services will be integrated, i.e. Data Privacy Law Support (FIZ Karlsruhe) and Awareness Building for Long-Term Archiving (ZB MED). These together provide a baseline service and data

infrastructure which covers the wide range of research fields of the consortium. From there we will work towards a FAIRification of those services and infrastructures.

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

Common NFDI aspects and NFDI-wide cooperation are addressed in all task areas of FAIRagro. FAIRagro supports the idea of the Research Data Commons (RDC) as the NFDI-wide agreed and developed set of basic and common NFDI standards, technological approaches and infrastructure components and will streamline its design and developments according to this concept.

The already established multicloud method for cross-domain data exchange (NFDI4Biodiversity, NFDI4Earth, DeNBI, FAIR-DS (GAIA-X)) is also supported by FAIRagro. With NFDI4BioDiversity, there are some overlaps in the investigated objects and there are repositories which should be made accessible and usable by both consortia. Specifically, collaborations in the areas of object identification, phenotypic trait collection and genotyping data management will be sought. We also plan to adopt BioCASe tools of NFDI4BioDiversity to facilitate the mapping of disciplinary infrastructures of FAIRagro. One use case of NFDI4Biodiversity has already been jointly formulated. With NFDI4Earth, we plan to cooperate closely in the area of spatial data as spatial reference is an essential characteristic of agrosystem data. Mutual complementarity with both NFDI4Biodiversity and NFDI4Earth is desired. Strong linkages are (a) in the area of infrastructure (provide data on both sides, create and use data spaces, common data pools, semantic containers) from which different communities can benefit and (b) in training and education activities, which is planned to be jointly realized and oriented to the RDC. We will jointly initiate further use cases, launch the process and establish it in the first phase of the project, e.g., "Challenges of Climate Change" together with NFDI4BioDiversity, NFDI4Earth and FAIRagro.

With DataPLANT, we plan to collaborate in the creation and harmonization of ontologies in plant science and the exploitation of relevant databases for data exchange. The consortia differ in terms of their audiences (basic research on plants mostly under laboratory or greenhouse conditions versus more applied research under real field conditions on plants in agricultural systems). Additional to the joint work and interfaces addressed as cross cutting topics (see below) FAIRagro consortium members do simultaneously participate in other NFDI consortia, thus acting as direct interfaces.

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.

FAIRagro has identified the following cross-cutting topics (see Berlin Declaration on NFDI Cross-Cutting Topics, Leipzig-Berlin-Erklärung) relevant for the consortium that need to be designed and developed by several or all NFDI consortia but are particularly important with regard to a cooperative approach to the other planned life science consortia.

1. Collaborative governance and general framework:
   a. Common vision of the NFDI, long term foresight and common strategic planning
   b. Governance & sustainability
   c. Cultural change aka reputation, publication/funding policies and credit systems
   d. Policy advice, consultation and outreach with respect to NFDI
   e. International visibility and networking of NFDI
   f. Human resource management, recruitment, development

2. Community (User) involvement:
   a. User driven adaptive development of NFDI (cross domain use cases)
   b. Training, undergraduate and graduate education, professional development
   c. Stimulating a cultural change of users and providers towards FAIR and open research data

3. Technical infrastructure and concepts:
   a. Research data commons (RDC) - incl. cloud infrastructures, computing power, Identity and Access Management (IAM) and interfaces between data providers and users across domains
   b. Standardization/harmonisation (terminologies, terminology management and services; (Meta)data harmonisation, findability; Provenance; Interoperability across domains; Common data and metadata standards and encodings (e.g. NFDI core metadata); Persistent unique identifier systems; Infrastructures
   c. Quality management and assurance- incl. certification of services (data, software and service quality criteria, evaluation and qualification criteria, qualification and/or certification processes for NFDI service offerings)
   d. Scalability of tools and services

4. Legal and ethical aspects
   a. Legal aspects (licensing, intellectual property rights and data protection & privacy)

    b.   Commercial use of data (commercialisation of data)

    c.   Ethical aspects of sharing research data and research software

    d.   Ethical-legal aspects of managing person-related data for research

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

Most of these cross-cutting topics are dealt with under the aspects of agrosystem sciences in the working areas of FAIRagro. The generic solutions can be used by other consortia or integrated into overarching issues. Above listed cross-cutting topics are integrated to the work program of the FAIRagro Task areas:

- TA1 - Management and Coordination: 1a, 1b, 1d, 1e
- TA2 - Community involvement and networking: 1c, 1d, 2a, 2b, 2c, 3b, 4a, 4c
- TA3 - Standardization, Interoperability and Quality: 1c, 3a, 3b, 3c, 3d, 4a, 4b, 4c, 4d
- TA4 - Infrastructure Services: 1c, 2a, 2b, 2c, 3a, 3b, 3c, 3d

Thus, FAIRagro will contribute to the cross-cutting topics of legal and ethical aspects of data usage (4a, 4b, 4c). The specifics of agrosystem data, often collected on and/or pertains to private land, particularly farms, and farm animals are also relevant to 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. 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 FAIRagro therefore provides benefits to the NFDI as a whole.

On the technical side of cross-cutting topics FAIRagro will support the RDC (3a) by implementing its IAM for agrosystem researchers. Software components for data integration will enhance the RDCs semantic data layer and support scientists to analyse and gain information from the RDC and therefore foster the adoption rate of RDC usage. These will cover FAIR containerized workflows which can be shared between scientists inside the RDC to provide reproducible science components. Additionally FAIRagro will serve as a knowledge hub for finding, using, accessing and analyzing agrosystem related data for different disciplines and applications by ensuring interoperability to related standards (metadata, ontologies, vocabularies, APIs).