

Letter of Intent

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

x	Binding letter of intent (required as advance notification for renewal proposals in 2024)
---	---

2 Formal details

- Name of the consortium

Data in PLANT research

- Acronym of the consortium

DataPLANT

- Applicant institution

Albert-Ludwigs-University of Freiburg (UFR), Fahnbergplatz, 79104 Freiburg

Head: Prof. Dr. Kerstin Krieglstein

- Spokesperson

Dr. Dirk von Suchodoletz, dirk.von.suchodoletz@rz.uni-freiburg.de, Computer Center, University of Freiburg

- Co-applicant institution

University of Kaiserslautern-Landau (RPTU), Erwin-Schrödinger-Straße 52, 67663 Kaiserslautern

Head: Prof. Dr. Arnd Poetzsch-Heffter and Prof. Dr. Gabriele E. Schaumann

Forschungszentrum Jülich (FZJ), Wilhelm-Johnen-Straße, 52428 Jülich

Head: Prof. Dr. Astrid Lambrecht

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstraße 3, 06466 Seeland

Head: Prof. Dr. Nicolaus von Wirén

- Co-spokesperson

Prof. Dr. Timo Mühlhaus, muehlhaus@bio.uni-kl.de, Computational Systems Biology, University of Kaiserslautern-Landau

Prof. Dr. Björn Usadel, b.usadel@fz-juelich.de, IBG-4 Bioinformatics, Forschungszentrum Jülich

Dr. Uwe Scholz, scholz@ipk-gatersleben.de, Bioinformatik und Informationstechnologie, Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung

- Participant institution

Bielefeld University

Eberhard Karls University Tübingen

Georg-August University of Göttingen

Gesellschaft für Biologische Daten, GFBio e.V. Bremen

Heinrich-Heine-University Düsseldorf

Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt

Humboldt University of Berlin

Ludwig-Maximilians University of München

Martin Luther University Halle-Wittenberg

Max-Planck-Institute of Molecular Plant Physiology

Philipps-University of Marburg

Ruhr-University Bochum

Technical University of Munich

University of Cologne

University of Duisburg-Essen

University of Hohenheim

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



University of Konstanz

University of Potsdam

University of Rostock

- Participant individual

Olaf Brandt, University Library, Eberhard Karls University Tübingen

Prof. Dr. Rolf Backofen, Bioinformatics, Albert-Ludwigs University of Freiburg

Prof. Dr. Andrea Bräutigam, Computational Biology, Bielefeld University

Dr. Dominik Brillhaus, CEPLAS, Heinrich-Heine-University Düsseldorf

Dr. Magdalene Cyra, fdm.nrw, University Library, University of Duisburg-Essen

Prof. Dr. Stefan Deßloch, Heterogeneous Information Systems, University of
Kaiserslautern-Landau

**Dr. Barbara Ebert, Executive Secretary, Gesellschaft für Biologische Daten, GFBio
e.V. Bremen**

Prof. Dr. Alisdair Fernie, Central Metabolism, Max-Planck-Institute of Molecular Plant
Physiology

Prof. Dr. Christoph Garth, Scientific Visualization Lab, University of Kaiserslautern-Landau

Dr. Björn Grüning, Bioinformatics, Albert-Ludwigs University of Freiburg

Prof. Dr. Eric Kemen, Center for Plant Molecular Biology, Eberhard Karls University
Tübingen

**Oliver Kohl-Frey, Communication, Information, Media Centre, University of
Konstanz**

Prof Dr. Nadia Kamal, Computational Plant Biology, TU München

Prof. Dr. Maria von Korff Schmising, Heinrich-Heine-University Düsseldorf

Prof. Dr. Ute Krämer, Molecular Genetics and Physiology of plants, Ruhr-University
Bochum

Dr. Jens Krüger, High Performance and Cloud Computing Group, IT Center, Eberhard Karls University Tübingen

Matthias Landwehr, Open Science at the Communication, Information, Media Centre University of Konstanz

Prof. Dr. Dario Leister, Plant Molecular Biology/Botany, Ludwig-Maximilians University of München

Prof. Dr. Heike Leitte , Visual Information Analysis, University of Kaiserslautern-Landau

Prof. Dr. Michael Lenhard, Genetics, University of Potsdam

Dr. Rosa Lozano-Durán, Center for Plant Molecular Biology, Eberhard Karls University Tübingen

Prof. Dr. Klaus Mayer, Plant Genome and Systems Biology, HMGU München

Prof. Dr. Isabel Monte, Center for Plant Molecular Biology, Eberhard Karls University Tübingen

Prof. Dr. Thomas Nägele, Plant Evolutionary Cell Biology, Ludwig-Maximilians University of München

Prof. Dr. Sven Nahnsen, Center for Quantitative Biology, Eberhard Karls University Tübingen

Prof. Dr. Marcel Quint, Institute of Agricultural und Nutritional Science, Martin Luther University Halle-Wittenberg

Prof. Dr. Klaus Rechert, Informatik, Hochschule Kehl

Prof. Dr. Ralf Reski, Plant Biotechnology, Albert-Ludwigs University of Freiburg

Jun.-Prof. Dr. Andreas Richter, Plant metabolic Physiology, University of Rostock

Dr. Inga Scheler, Computer Center, University of Kaiserslautern-Landau

Prof. Dr. Christian Schmitz-Linneweber, Molecular Genetics, Humboldt University of Berlin

Prof. Dr. Waltraut Schulze, Plant Systems Biology, University of Hohenheim

Prof. Dr. Nadine Töpfer, Institute for Plant Sciences, University of Cologne

Prof. Dr. Haim Treves, Plantmetabolism, University of Kaiserslautern-Landau

Prof. Dr. Jan de Vries, Applied Bioinformatics, Georg-August University of Göttingen

Prof. Dr. Thomas Walter, IT Center, Eberhard Karls University Tübingen

Prof. Dr. Andreas P. M. Weber, Plant Biochemistry, Heinrich-Heine-University Düsseldorf

apl. Prof. Dr. Stefanie Weidtkamp-Peters, Center for Advanced Imaging, Heinrich-Heine-University Düsseldorf

Prof. Dr. Felix Wilmund, Molecular Plant Physiology, Philipps-University of Marburg

3 Objectives, work programme and research environment in the second funding period

- [Research area of the proposed consortium \(according to the DFG classification system\)](#)

Plant Science (202 (2.12) with topics –01 –04, –05, –06, and –07)

- **Concise summary of the consortium's main objectives and task areas**

Plant research aims to understand the processes governing plant life using interdisciplinary research employing sophisticated high throughput technologies such as mass spectrometry, next-generation sequencing, spectroscopy, imaging, and phenotyping as well dedicated low throughput technologies. Collaborative work and the integration of diverse data from multiple sources necessitate continuous, harmonized data contextualization. DataPLANT's annotation principles scale with these requirements, prioritizing the researchers without burdening them with redundant concepts. **DataPLANT's overarching goal is to provide robust, easily accessible and user-oriented research data management (RDM) practices, tools, and infrastructure to support collaborative plant biology research.** By using, improving and establishing common standards, software, and infrastructure, DataPLANT ensures data availability, quality, and interoperability. Continuous improvement and adaptation of these tools and practices is necessary to meet the evolving community needs. Recognizing the socio-technical challenges of RDM, DataPLANT seeks to embed RDM as a core component of the plant community fostering open development and governance models, embracing the "open-source spirit", treating well-annotated research data as shared community resources. Committed to a participatory model, DataPLANT spurs active participation to ensure that its tools and infrastructure align with the community's needs.

DataPLANT is driven by the following principles: **(1) Facilitating FAIRness:** Meet the FAIR (Findable, Accessible, Interoperable, Reusable) requirements of institutions and funding agencies, promoting a continuous FAIR-by-design RDM approach. **(2) Valorising Data and Service Interoperation:** Enable researchers to incorporate and benefit from others' FAIR Digital Objects (FDOs) and allow them to effectively utilize available analysis resources and RDM services. **(3) Fostering Community Participation:** Ensure that scientists can benefit from and contribute to the (inter)national open research data environment. **(4) Knowledge Transfer:** Provide opportunities for researchers to educate (themselves) about RDM technologies, services, and infrastructures; measures for **(5) Ensuring Long-Term Sustainability** of RDM practices and infrastructure. Adhering to these principles, DataPLANT enhances the efficiency, reproducibility, and impact of plant research. Data PLANT is structured into four closely integrated **task areas**, organized using a dynamic approach following directions of the community, and addressing the written and oral interim review. These **task areas** act as competence clusters, bundling specific expertise and resources to address community requirements within the broader NFDI framework.

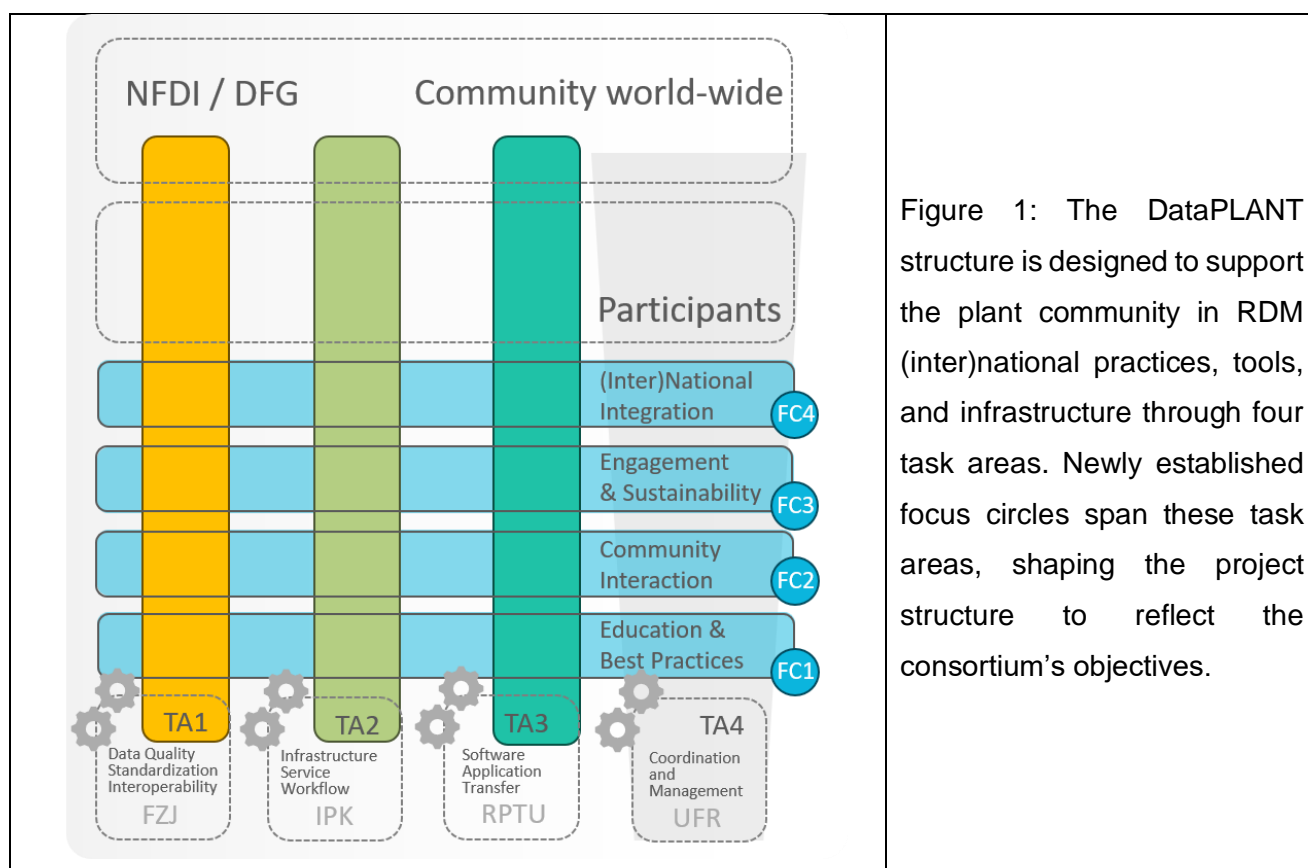


Figure 1: The DataPLANT structure is designed to support the plant community in RDM (inter)national practices, tools, and infrastructure through four task areas. Newly established focus circles span these task areas, shaping the project structure to reflect the consortium's objectives.

To ensure adherence to the community, the NFDI goals and the reviewer comments, we will establish **focus circles** that span the task areas, facilitating the implementation of overarching objectives. These focus circles comprise **Education & Best Practices (FC1)**, **Community Interaction (FC2)**, **Open Engagement & Sustainability (FC3)**, and **(Inter)National Integration (FC4)**. They will enhance communication and management, ensuring that goals are competently supported by experts from each task area working together (Fig. 1).

Task Area 1 (Data Quality, Standardization, Interoperability) will provide assistance including AI approaches for collecting vocabulary/ontologies encouraging participation (FC3). TA1 will improve ontologies embedded in international consortia, required to enhance data and workflow interoperability and machine actionability for AI-readiness. This will bridge user documentation, technical endpoints, workflow engines, and domain requirements (FC4). To ensure (meta)data quality, TA1 will implement automatic validation processes which can be complemented by a community-driven data publication review system (FC2+4). Additionally, TA1 will promote RDM practices through data stewards fostering a culture of traceability within the community (FC1).

Task Area 2 (Infrastructure, Service, and Workflow Platform) will advance the DataPLANT DataHUB and services for a deep integration with national and international infrastructures, such as the NFDI research data commons, OpenAIRE, EOSC and EBI improving recognition by journals, NFDI, the EU, and international RDM initiatives and stakeholders (FC4). TA2 supports efficient data management and analysis, capable of running actionable FDOs within defined

workflow environments (FC1+2). TA2 will use consortia cloud environments and the NFDI FAIR data spaces as an infrastructure-as-a-service distribution and maintenance model (FC3).

Task Area 3 (Software, Application, and Transfer) will foster an open-source spirit, promoting participation and skill development in research data projects and software. TA3 will cultivate community support and engagement, providing personalized assistance with tailored digital assistance (FC3). TA3 will mature and consolidate the plant FDO support software stack to enhance interoperability with other NFDI and international RDM services (FC4). TA3 will offer **open educational resources** and implement a "train the trainer" model supporting our embedded data steward approach. Enhanced documentation, tutorials, and teaching materials will be developed (FC1), enabling active community participation through transparent, community-driven open feedback and development (FC1). Additionally, TA3 will establish a support structure for embedded and geo-distributed data stewards (FC2).

Task Area 4 (Coordination and Management) will focus on effectively managing the consortium and safeguarding the adherence to community focus aims. Information on the project and results of the task areas will be disseminated coordinately by TA4, making sure that all interested parties inside and outside the consortium are well-informed. TA4 will establish bidirectional advocacy structures and low-barrier participation mechanisms for the entire community and beyond (FC3). This task area will also build special interest groups for the life sciences (LS), moderate legal issues and DFG representation (FC2), and bring together state-level initiatives, compute centers, faculty IT, research institutions, libraries, and NFDI liaison and consolidation (FC4).

- [Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives](#)

DataPLANT's technology simplifies FAIR requirements for researchers, promoting a FAIR-by-design approach named Annotated Research Context (ARC). By integrating standards like Investigation Study Assay (ISA) and Common Workflow Language (CWL) into the ARC-Plant-Research-FDO container, an implementation of the RO-Crate format based on (Bio)schema.org, DataPLANT ensures seamless data management to publication and will continue to contribute to international frameworks and improve tools to make them more user-friendly and exploring AI-based support for existing tools and services. Responding to community needs, DataPLANT will include lab process documentation, including digital lab books. DataPLANT aims to enhance its tools and services, such as SWATE (for annotation) and ARCItext (for building ARC-FDOs), to create well-integrated components (e.g. through ARCFs) accessible beyond DataPLANT. This will be achieved through open, transparent, and sustainable cloud-oriented software practices, avoiding lock-ins. All services and platforms are designed to be modular and reusable, transferable to other sites. The DataHUB (hosting interactive version-controlled ARC-FDOs) and the ARChive (for long-term access) already run on the de.NBI cloud and at other participating institutions. We will upgrade the DataHUB's search capabilities and extend its current data publication functionality to a peer review platform and improve automatic quality control and testing. DataPLANT's primary objective is to provide an area for plant research data and workflows with robust tool support, enabling export to various community repositories. Plans include developing meta- and data converters to facilitate easy transfer to technique-specific international repositories (e.g., hosted by the EBI) and common NFDI infrastructure in a FAIR manner. With the envisioned tooling support for FDO actionability, researchers will be able to effectively utilize available analysis resources, DataHUB FDOs, and RDM services by seamlessly integrating them into the Galaxy workflow engine, Nextflow, and all CWL-compatible platforms. DataPLANT ensures plant researchers can benefit from and contribute to the (inter)national open research data environment by integrating into additional community and NFDI AAI beyond existing ELIXIR AAI and ORCID services. While currently reliant on some publicly available commercial services, DataPLANT aims to replace these with common NFDI infrastructure, starting with the Base4NFDI initiative on a shared software repository. Helpdesk and consultation services will be jointly managed under the "biodata" label for life sciences consortia. We are committed to the long-term sustainability of RDM practices and infrastructure, ensuring the availability of a fully functional long-term archiving and access (LTA) solution. This may be partially provided by DataPLANT or integrated into the NFDI LTA stack. Long-term stable infrastructures are essential for data publication, ideally through community repositories or extended local institutional systems for fundamental plant data.

- [Interfaces to other NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration](#)

DataPLANT maintains strong links with other NFDI consortia, ensuring a unified approach to advancing RDM in and beyond the LS. This includes jointly organized hackathons for the LS and coordinated participation in key conferences. For the second funding period, agreements are being prepared for a special NFDI “biodata” interest group to harmonize RDM strategies involving biological data, including cross-community approaches for joint tool development and consolidating community resources. Open sequence data remains a key issue for the biodata group, which includes NFDI4Biodiversity, NFDI4Microbiota, DataPLANT, FAIRagro and the methods oriented NFDI4BIOIMAGE. de.NBI resources provide the bases to form a larger LS service ecosystem. Therefore, all DataPLANT cloud-oriented and modular services can be hosted by any institution, consortium and with a basic set of open-source services. To expand service offerings and integrate into the evolving NFDI infrastructure, a dedicated focus circle has been formed (FC4). Other areas of cooperation are joint helpdesk and consultation services, providing comprehensive user support in areas where one consortium is strongly represented but others aren't, as well as shared knowledge bases, training materials, roadshows for life sciences communities, or a joint news desk. Common interests include the application of Bioschema.org, and collaborations on ontology development in the metadata section (TS4NFDI). We plan joint interdisciplinary research showcases using FDOs, advancing ARC standardization to include NFDI4Biodiversity, NFDI4BIOIMAGE and FAIRagro requirements, and integrating tooling reciprocally. There is a concerted effort with NFDI4BIOIMAGE to establish an online training platform. To strengthen links, the NFDI4BIOIMAGE speaker and NFDI4Biodiversity Executive secretary have joined DataPLANT as participants and a co-spokesperson of NFDI4Biodiversity will become co-spokesperson in DataPLANT. These collaborations provide the foundation for the shared geo-distributed data steward concept operated by DataPLANT in exchange with LS consortia.

At the NFDI level, DataPLANT is engaged in proposing and participating in the development of BASE4 infrastructure, integrating e.g. IAM identifiers, LTA strategies, terminology services data management plans. The participation in NFDI sections is used to exchange concepts, evolve ideas and plan joint undertakings e.g. on the use of AI in RDM. Expertise like legal advice is sourced from the NFDI ELSA section; EduTrain offers synergies for qualification modules. Strategic collaboration within the governance and sustainability working group focuses on proposing concepts for NFDI sustainability and provides coordinated policy advice on RDM. The overarching goal is to advance the “OneNFDI” vision and embed NFDI tools and services across various research institutions.

4 International and national networking

DataPLANT pursues national and international collaborations. Besides the interactions with other NFDIs mentioned above, DataPLANT strongly interacts with its stakeholders, i.e. plant oriented clusters, actors and local research management structures. Hence, DataPLANT interacts with the Plant Excellence Cluster CEPLAS, the SFBs 1535 (MibiNet), 1644 (Plant phenotypic plasticity), 1664 (Plant Proteoform Diversity), TRR 356 (Plant Microbe), TRR341 (Plant Ecological Genetics), and TRR175 (The Green HUB), MAdLand (DFG priority programme 2237). These build on ARCs and the DataHUB and are offering joint training. Besides, DataPLANT presents at large plant community meetings at e.g. the Plant Status Seminar, Breeder's days organized by the GFPI, and German Plant Science conferences, International Conference of the German Society for Plant Sciences. In addition, a strong interaction takes place and is evolving with the computing and data management offices/centres at the Universities of Düsseldorf, Cologne, Freiburg, Tübingen and Kaiserslautern-Landau and with RDM state initiatives, e.g. in NRW and BW to develop joint infrastructure and RDM policy advice. These interactions will be strengthened and broadened. Internationally, DataPLANT is well-connected, especially through its associations via the European ELIXIR infrastructure and in particular ELIXIR's Plant Science Community. DataPLANT regularly participates, discusses and plans new transEuropean ideas in e.g. international Hackathons and working groups. Collaboration is pursued with transnational projects on plant sciences shaping standards in the MIAPPE (Minimal standards on a plant phenotyping experiment) and e.g. FAIDARE as well as with the international BrAPI (Breeding API) consortium to assess relevant information electronically. Seeking standardization and interoperability, DataPLANT works closely together with the ISA tool, the RO-Crate (for ARC-FDO) and the common workflow communities through mature specifications/implementations and commits. These cooperations will only increase in the future. Finally, DataPLANT taps into the international eLabFTW laboratory notebook community to follow demands from its community beside NFDI interactions.

Since ontologies are very important, DataPLANT collaborates closely with plant ontology providers such as the AgroPortal and Planteome. In these collaborations joint work is done on ontologies and DataPLANT adds its own userbase requirements. To also link to data management portals, DataPLANT is active in EOSC in regular ELIXIR EOSC focus groups and was involved in an EOSC Life project linking different European resources. Furthermore, DataPLANT has contributed to Galaxy, which has added 50.000 registered users during DataPLANT's lifetime. Here, ARC-FDO and Zenodo/Invenio support was integrated into Galaxy. Its users are now able to export complete data analysis with its complete provenance information as RO-Crate into Zenodo/Invenio and save the DOI back to Galaxy.