B-1   Progress Report Template Part 1, for publication

1   General Information

- Name of the consortium

DataPLANT: Data in PLANT research

- Research domains or research methods addressed by the consortium

DataPLANT caters to basic plant sciences. This multidisciplinary domain served by DataPLANT encompasses plant genetics, physiology, biochemistry, cell and developmental biology, as well as systems biology and plant biotechnology. It aims to shed light on fundamental processes such as plant growth and development, stress tolerance, nutrient use, interactions and adaptations to the environment or unravelling biosynthetic pathways to name but a few. DataPLANT is including support for modern omics-based plant sciences including genomics, metabolomics, proteomics and phenomics methods. Interpreting large and complex datasets as those encountered in modern plant sciences might require expertise in biology, chemistry, physics, mathematics, and computer science which is also supported by DataPLANT by providing multi actor collaboration platforms.

- URL of the consortium website and repositories used for publishing output

Website: https://nfdi4plants.org/
Github (Software development platform): https://github.com/nfdi4plants
DataHUB (ARC platform): https://git.nfdi4plants.org/
ARChive (ARC publishing platform): https://archive.nfdi4plants.org/

2   Summary

DataPLANT developed from a nascent research data management (RDM) landscape for fundamental plant research. Focusing on efficient resource allocation, we established four expertise-rich activity clusters fostering rapid community outreach. Collaboration with key stakeholders and major research initiatives drove progress and expanded partnerships. Our iterative, community-driven approach is rooted in existing frameworks and standards to address plant research effectively and sustainably. To address diverse segments of the plant research community, we strategically engaged with both large and smaller research groups. Active participation in conferences and events, together with consortia such as NFDI4Biodiversity and FAIRagro, helped shape collaborations within the NFDI and beyond. Hackathons and surveys further stimulated creative problem solving and insights. Our online presence, helpdesk and

designated Data Stewards offer vital resources and personalised support. By embedding experts in labs, the Data Steward model ensures accessible assistance. We are committed to open science, enabling researchers to share FAIR Digital Objects. The model's success is evident in collaborations with key research projects, including Excellence Strategy programmes and Collaborative Research Centres/Transregios, as well as other larger scale research units, underlining our far-reaching impact within the plant research community.

RDM is central to DataPLANT's activities, requiring global networking and compatibility with international standards. They release open-source software and training resources via GitHub, encouraging user engagement. DataPLANT adheres to international policies and standards, actively contributing to schemas such as ISA and adapting Git-LFS for improved user experience. DataPLANT's tools facilitate effective RDM, exemplified by the Annotated Research Context (ARC), a DataPLANT profile implementation of the RO-Crate format leveraging the internationally recognised ISA standard. DataPLANT integrates the Common Workflow Language (CWL). DataPLANT collaborates closely with the Galaxy community, to use ARCs directly in Galaxy and by enriching training materials thus benefiting multi-omics and molecular disciplines. DataPLANT forges strong partnerships with the European Life Science Infrastructure for Biological Information (ELIXIR) and the ELIXIR Plant Sciences Community, collaborating on joint projects and hackathons. This collaboration led to the integration of the phenotyping standard MIAPPE into ARC-ISA, simplifying user access. Finally, DataPLANT collaborates and works together with relevant plant ontology stakeholders and the BreedingAPI.

DataPLANT actively collaborates with NFDI consortia to optimise resource utilisation and tap into specific expertise. Examples such as supporting the Base4NFDI Initiative and the LTA, IAM and further working groups underscore our commitment. We consistently engage in similar initiatives to foster shared support concepts and drive innovation. By closely partnering with other NFDI consortia, DataPLANT seeks to create synergies and cultivate a culture of collaborative innovation within the NFDI community. Alongside NFDI-wide meetings, we maintain regular interactions with NFDI4Biodiversity, FAIRagro, and NFDI4BIOIMAGE. These engagements range from joint events to informal exchanges, pooling resources for mutual topics or offering assistance to new NFDI consortia. DataPLANT has also established connections with other consortia such as NFDI4Microbiota, NFDI4Chem, NFDI4Health, and MatWerk. These collaborations address various themes, including bridging ontology gaps. Furthermore, the ARC, DataPLANT's FDO, is being explored for integration in diverse methodologies, such as imaging. A project to combine these domains was accepted at this year's Biohackathon Germany. Similarly, we're exploring direct collaboration with FAIRagro for joint helpdesks, ARC use, and shared topics.

DataPLANT aligns with the OneNFDI vision of an open and unified data ecosystem beyond its core community. The initial step involves forging agreements and cultivating closer collaboration within the life sciences domain. This collective vision is materialising through advancements in Base4NFDI. DataPLANT will continue to prioritise tools and services for its community while integrating core NFDI developments to avoid redundancy. Initiatives such as IAM and the helpdesk are underway, with potential expansion into infrastructure building blocks such as versioning and repository services. Identifying shared standards and services is crucial. We anticipate deeper integration of various consortia in the upcoming funding round, already starting to coordinate and collaborate more rigorously. Preserving the grassroots ethos of the NFDI within larger structures is essential, and support in promoting this spirit of coordination and collaboration would be highly valuable. In summary, DataPLANT prioritises compatibility and interoperability, contributing to international standards such as ISA, CWL, MIAPPE, and relevant ontologies and frameworks.

In terms of tools and resources, DataPLANT has developed the Swate Workflow Annotation Tool for Excel (Swate) which streamlines metadata annotation using controlled vocabularies and ontologies, ensuring ISA compliance. Swate provides curated templates to assist users in recording metadata, promoting robust data annotation and sharing. The Swate OBO Updater (Swobup) automates ontology updates, ensuring integrity and versioning. This approach avoids technological barriers and embraces the principles of open software and open science principles. The DataPLANT consortium's core objective is to provide a easy-to-use gateway and data platform and has programmed the ARCitect which is a graphical user interface tool to generate, modify and work with ARCs. The DataPLANT DataHUB facilitates the management of data throughout its lifecycle, from annotation to publication of results. It serves as a central science gateway, acting as a starting point for various analysis workflows across scientific disciplines. Workflow descriptions are stored in the ARC, making them accessible to processing frameworks such as Galaxy or nf-core Nextflow pipelines. The recently implemented ARChive simplifies data publication, providing guidance and support to researchers. Additionally, DataPLANT offers the DataPLAN wizard-based service for generating data management plans, aligning with funder requirements and promoting sustainable and reusable metadata standards.

3 Composition of the consortium

- Applicant institution - Spokesperson

Applicant institution	Location	Duration
Albert-Ludwigs University of Freiburg (UFR) Head: Prof. Dr. Kerstin Krieglstein	Fahnenbergplatz, 79104 Freiburg	10/2020 – 09/2023

- Spokesperson

Spokesperson	Institution, location	Task area(s)	Duration
Dr. Dirk von Suchodoletz https://orcid.org/0000-0002-4382-5104	Computer Center, UFR	Task Area IV	10/2020 – 09/2023

- Co-applicant institutions - Co-spokespersons

Co-applicant institutions	Location	Duration
Jülich Research Center (FZJ) Head: Prof. Dr. Astrid Lambrecht	Wilhelm-Johnen-Straße 52428 Jülich	10/2020 – 09/2023
Eberhard Karls University Tübingen (EKUT) Head: Prof. Dr. Karla Pollmann	Geschwister-Scholl-Platz 72074 Tübingen	10/2006 – 09/2023
University of Kaiserslautern-Landau (RPTU) Head: Prof. Dr. Arnd Poetzsch-Heffter Prof. Dr. Gabriele E. Schaumann	Erwin-Schrödinger-Straße 52 67663 Kaiserslautern	10/2020 – 09/2023

- Co-spokespersons

Co-spokespersons	Institution, location	Task area(s)	Duration
Prof. Dr. Björn Usadel https://orcid.org/0000-0003-0921-8041	IBG-4 Bioinformatics, FZJ	Task Area I	10/2020 – 09/2023
Dr. Jens Krüger https://orcid.org/0000-0002-2636-3163	High Performance and Cloud Computing Group, IT Center, EKUT	Task Area II	10/2020 – 09/2023
Prof. Dr. Timo Mühlhaus https://orcid.org/0000-0003-3925-6778	Computational Systems Biology, RPTU	Task Area III	10/2020 – 09/2023

- Participants

Participating individuals	Institution, location	Duration
Dr. Björn Grüning	Albert-Ludwigs-Universität Freiburg, Freiburg	10/2020 – 10/2023
Prof. Dr. Gerhard Schneider	Albert-Ludwigs-Universität Freiburg, Freiburg	10/2020 – 03/2022
Prof. Dr. Ralf Reski	Albert-Ludwigs-Universität Freiburg, Freiburg	10/2020 – 10/2023
Prof. Dr. Rolf Backofen	Albert-Ludwigs-Universität Freiburg, Freiburg	10/2020 – 10/2023
Dr. Marianne Dörr	Eberhard Karls Universität Tübingen, Tübingen	10/2020 – 10/2023
Dr. Olaf Brandt	Eberhard Karls Universität Tübingen, Tübingen	10/2020 – 10/2023
Dr. Sven Nahnsen	Eberhard Karls Universität Tübingen, Tübingen	10/2020 – 10/2023
Prof. Dr. Thomas Walter	Eberhard Karls Universität Tübingen, Tübingen	10/2020 – 10/2023
Prof. Dr. Eric Kemen	Eberhard-Karls-Universität Tübingen, Tübingen	10/2020 – 10/2023

Prof. Dr. Andreas P. M. Weber	Heinrich-Heine-Universität Düsseldorf, Düsseldorf	10/2020 – 10/2023
Dr. Dominik Brilhaus	Heinrich-Heine-Universität Düsseldorf, Düsseldorf	09/2021 – 10/2023
Dr. Daniel Lang	Helmholtz Zentrum München - Deutsches Forschungszentrum für Gesundheit und Umwelt, München	10/2020 – 2021
Prof. Dr. Klaus F.X. Mayer	Helmholtz Zentrum München - Deutsches Forschungszentrum für Gesundheit und Umwelt, München	10/2020 – 10/2023
Prof. Dr. Klaus Rechert	Hochschule Kehl, Kehl	10/2020 – 10/2023
Prof. Dr. Dr. hc. Edda Klipp	Humboldt-Universität zu Berlin, Berlin	10/2020 – 10/2023
Prof. Dr. Dario Leister	Ludwig-Maximilians-Universität München, München	10/2020 – 10/2023
Prof. Dr. Alisdair Fernie	Max-Planck-Institut für molekulare Pflanzenphysiologie (MPI-MP), Golm	10/2020 – 10/2023
Prof. Dr. Christoph Garth	Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern	10/2020 – 10/2023
Prof. Dr. Heike Leitte	Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern	10/2020 – 10/2023
Dr. Inga Scheler	Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern	10/2020 – 10/2023
Prof. Dr. Stefan Deßloch	Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Kaiserslautern	10/2020 – 10/2023
Prof. Dr. Ute Kraemer	Ruhr-Universität Bochum, Bochum	10/2020 – 10/2023
Prof. Dr. Andrea Bräutigam	Universität Bielefeld, Bielefeld	10/2020 – 10/2023
Prof. Dr. Karl Schmid	Universität Hohenheim, Hohenheim	10/2020 – 10/2023
Jun. Prof. Dr. Sandra Schmöckel	Universität Hohenheim, Hohenheim	10/2020 – 10/2023
Prof. Dr. Waltraud Schulze	Universität Hohenheim, Hohenheim	10/2022 – 10/2023

Dr. Anja Oberländer	Universität Konstanz, Konstanz	10/2020 – 10/2023
Dr. Petra Hätscher	Universität Konstanz, Konstanz	10/2020 – early 2022
Matthias Landwehr	Universität Konstanz, Konstanz	early 2022 – 10/2023

Participating institutions	Location	Duration
Heinrich-Heine-Universität Düsseldorf, Düsseldorf	Düsseldorf	10/2020 – 10/2023
Helmholtz Zentrum München - Deutsches Forschungszentrum für Gesundheit und Umwelt, München	München	10/2020 – 10/2023
Humboldt-Universität zu Berlin, Berlin	Berlin	10/2020 – 10/2023
Ludwig-Maximilians-Universität München, München	München	10/2020 – 10/2023
Max-Planck-Institut für molekulare Pflanzenphysiologie (MPI-MP), Golm	Golm	10/2020 – 10/2023
Ruhr-Universität Bochum, Bochum	Bochum	10/2020 – 10/2023
Universität Bielefeld, Bielefeld	Bielefeld	10/2020 – 10/2023
Universität Hohenheim, Hohenheim	Hohenheim	10/2020 – 10/2023
Universität Konstanz, Konstanz	Konstanz	10/2020 – 10/2023
Hochschule Kehl, Kehl	Kehl	03/2020 – 10/2022