FAIR Data Infrastructure for Agrosystems

FAIRagro

Proposal with revised Work Programme

(public version)

Version 2.2 (from July, 21st 2023)

Table of contents

Acronyms and abbreviations

A.I.	Artificial Intelligence
AAAS	American Association for the Advancement of Science
AAI	Authentication and Authorisation Infrastructure
AARC BPA	Blueprint Architecture of Authentication and Authorisation for Research and Collaboration
AGMEMOD	Agricultural Member State Modelling (network)
AgMIP	Agricultural Model Intercomparison and Improvement Project
Agrontology	Web Ontology Language vocabulary providing a set of domain properties to the AGROVOC thesaurus
AGROVOC	Multilingual Thesaurus (FAO)
API	Application Programming Interface
ATB	Leibniz Institute for Agricultural Engineering and Bioeconomy
BLU	Bielefeld University
BMBF	Federal Ministry of Education and Research
BMEL	Federal Ministry of Food and Agriculture
BonaRes	BMBF project “Soil as a sustainable resource for the bioeconomy”
BrAPI	Breeding API
BSA	Directorate General of the Bavarian State Archives
BUC	Bridging Use Cases
BWI	German National Forest Inventory (Bundeswaldinventur)
BZE Wald	German National Forest Soil Inventory (Bodenzustandserhebung Wald)
BZE-LW	German National Agricultural Soil Inventory (Bodenzustandserhebung Landwirtschaft)
CA	Consortium Agreement
CAB	Community Advisory Board
CAPRI	Common Agricultural Policy Regionalised Impact Modelling System
chap.	Chapter
CI	Continuous Integration
CKAN	Comprehensive Knowledge Archive Network
CODATA	Committee on Data
CRM	Customer-Relationship-Management-System
DaCuRa	Data Cube and Risk Assessment
DAFA	German Agricultural Research Alliance
DAKIS	Digital Agricultural Knowledge and Information System (project)
DataCube	Array Database Management System by JKI
DBFZ	German Biomass Research Center gGmbH
DBG	The German Soil Science Society
DCAT	Data Catalog Vocabulary
de.NBI	German Network for Bioinformatics Infrastructure
DFG	German Research Foundation
DFN	German National Research and Education Network
DiSSCo	European Distributed System of Scientific Collections
DivSeek	Community driven initiative for breeders and researchers to mobilize a vast range of plant genetic variation
DMP	Data Management Plan

DO	Digital Objects
DOI	Digital Object Identifier
DOIP	Digital Object Interface Protocol
DPG	German Society for Plant Protection and Plant Health
DQ	Data quality
DSS	Decision Support System
DSSAT	Decision Support System for Agrotechnology Transfer
DSSC	Data Steward Service Center
Dublin Core	Metadata Standard
DWD	Deutscher Wetterdienst (Germany's National Meteorological Service)
e!DAL	Electronic Data Archive Library
ECPGR	European Cooperative Programme for Plant Genetic Resources
Edaphobase	Database for distribution and ecology of soil organisms
EGU	European Geoscience Union
EJP	European Joint Programme
ELIXIR	A distributed infrastructure for life-science information
ELSA	NFDI Sektion “Ethical, Legal & Social Aspects”
EMPHASIS	European Infrastructure for Multi-scale Plant Phenomics and Simulation
ENVRI	Environmental Research Infrastructure
EOSC	European Open Science Cloud
EPPO	European and Mediterranean Plant Protection Organization
ESDAC	European Soil Data Centre
ESFRI	European Strategy Forum on Research Infrastructures
EUC	Extension Use Cases
EURISCO	European Search Catalogue for Plant Genetic Resources
FACCE JPI	Joint Programming Initiative on Agriculture, Food Security and Climate Change
FAIR	Findable, Accessible, Interoperable and Reusable
FAIR-DO	FAIR Digital Objects
FAO	Food and Agricultural Organization
FAQ	Frequently Asked Questions
FDMentor	BMBF project “Reusable Strategies and Tools for Research Data Management”
FDO	FAIR Digital Object
FDO-TSIG	Technical Specification & Implementation Group in the FAIR Digital Object Forum
FID Forstland	Fachinformationsdienst für die Forstwissenschaften und nachhaltige Landnutzung
FIZ	Leibniz Institute for Information Infrastructure
FLOPO	Flora Phenotype Ontology
FTE	Full-time equivalent
FZJ	Forschungszentrum Jülich
GAIA-X	Federated Data Infrastructure for Europe
GAUG	Georg-August University Göttingen
GBIF	Global Biodiversity Information Facility
GBIS/I	Information system of the IPK ex situ collection for crop and wild relatives that provides passport data and primary evaluation data
GBV	Gemeinsamer Bibliotheksverbund
GCBN	German Crop BioGreenformatics Network

GDPR	General Data Protection Regulation
GeoNode	Web-based application and platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructures (SDI)
GeRDI	Generic Research Data Infrastructure
GFBio	German Federation for Biological Data
GHGA	German Human Genome-Phenome Archive
GIL	Gesellschaft für Informatik in der Land-, Forst- und Ernährungswirtschaft e.V.
GitLab	Open source end-to-end software development platform
GLASSNET	Global to Local Analysis of Systems Sustainability
GLTEN	Global Long Term Experimental Network
GODAN	Global Open Data for Agriculture and Nutrition
GO FAIR	Bottom-up, stakeholder-driven and self-governed initiative
GPW	Gesellschaft für Pflanzenbauwissenschaften e.V.
TUM-GZW	Graduate Centre Weihenstephan of Technical University of Munich
HEF	Hans Eisenmann-Forum for Agricultural Sciences
HeFDI	Hessische Forschungsdateninfrastrukturen
HSWT	Weihenstephan-Triesdorf University of Applied Sciences
IBSNAT	International Benchmark Sites Network for Agrotechnology Transfer
ICASA	International Consortium for Agricultural Systems Application
IGAD	Interest Group Agricultural Data (of the Research Data Alliance, RDA)
IGSN	International Geo Sample Number (PID)
INKA BB	Innovationsnetzwerk Klimaanpassung Brandenburg Berlin (project)
INSPIRE	Infrastructure for Spatial Information in Europe
IOSDV	International Organic Nitrogen Long-term Fertilisation Experiment
IPBES	Intergovernmental Platform on Biodiversity and Ecosystem Services
IPK	Leibniz Institute of Plant Genetics and Crop Plant Research
IPM	Integrated Pest Management
IPPN	International Plant Phenotyping Network
ISIP	Informationssystem für die integrierte Pflanzenproduktion
ISMC	International Soil Modeling Consortium
ISO	International Organization for Standardization
IUSS	International Union of Soil Science
JKI	Federal Research Centre for Cultivated Plants - Julius Kühn-Institute
KIT	Karlsruhe Institute of Technology
KTBL	Kuratorium für Technik und Bauwesen in der Landwirtschaft
LAILAPS-QSM	Library for semantic query extension in the crop domain
LfL	Bayerische Landesanstalt für Landwirtschaft
LIMS	Laboratory Information Management System
LIVIVO	Search Portal for Life Sciences (by ZB MED)
LoC	Letter of Commitment
LTE	Long-term field experiments
MACSUR	Modelling European Agriculture with Climate Change for Food Security
MCPD	Multi-Crop Passport Descriptors
MIAPPE	Minimal Information About a Plant Phenotyping Experiment
NFDBBB	Netzwerk Forschungsdaten Berlin-Brandenburg

NFDI-RDC	NFDI Research Data Commons
NUC	New Agrosystem Use Cases
OA	Open Access
OBO	Open Biological and Biomedical Ontology
ODRL	Open Digital Rights Language
OER	Open Educational Resources
OGC	Open Geospatial Consortium
OLS	Ontology Lookup Service
OpenAgrar	Open-access repository for the publication of documents, publication references and research data of research institutions (by JKI)
ORCID	Open Researcher and Contributor ID
OSGeo	The Open Source Geospatial Foundation
P&D	Pest and Disease
PGP	Plant Genomics and Phenomics Research Data Repository
PhenoRob	Robotics and Phenotyping for Sustainable Crop Production
PID	Persistent identifier
PlabiPD	Database focused on crop plants
PROV	Provenance Model of the W3C consortium
PUBLISSO	Open Access Publishing Platform for Life Sciences (by ZB MED)
REST	Representational State Transfer
RDA	Research Data Alliance
RDC	Research Data Commons
RDI	Research Data Infrastructure
RDM	Research Data Management
RDMO	Research Data Management Organiser
RDMO4Life	RDM tool for the entire life cycle of data
RfII	German Council for Scientific Information Infrastructures
RO-Crate	Research Object Crate
ROR	Research Organization Registry (PID)
SC	Steering Committee
SCAI	Fraunhofer Scientific Computing and Algorithms Institute
SciWin	Scientific workflow infrastructure
SDC	Secure Data Cluster at the Thünen Institute
SEAMLESS	System for Agricultural Modelling- Linking European Science and Society
SEMAF	Flexible Semantic Mapping Framework
SGN	Senckenberg Gesellschaft für Naturforschung
SNP	Single Nucleotide Polymorphisms
SRADI	Smart Rural Area Data Infrastructure
SSS	EGU Devision Soil Science System
SYNOPS	Synoptic evaluation of plant protection products
TA	Task Area
TDM	Text and Data Mining
TDWG	Taxonomic Databases Working Group
Thünen	Federal Research Institute for Rural Areas, Forestry and Fisheries - Johann Heinrich von Thünen Institute
TISDAR	Thünen Institute Spatial Data Repository

TKFDM	Thüringer Kompetenznetzwerk Forschungsdatenmanagement
TO	Plant Trait Ontology
TTT	Office Think Tank Digitalisation at Thünen Institute
TUM	Technical University of Munich
UAV	Unmanned Air Vehicle
UBN	University of Bonn
UC	Use Case
UFZ	Helmholtz-Zentrum für Umweltforschung
UHOH	University of Hohenheim
VoID	Vocabulary of Interlinked Datasets
VZG	Head Office of the GBV Common Library Network
W3C	Word Wide Web Consortium
WG	Working Group
XML	Extensible Markup Language
ZALF	Leibniz Centre for Agricultural Landscape Research
ZB MED	Information Centre for Life Sciences
ZEPP	Zentralstelle der Länder für EDV-gestützte Entscheidungshilfen und Programme im Pflanzenschutz

1 General Information

1.1 Name of the consortium in English and German

FAIR Data Infrastructure for Agrosystems

FAIRe Dateninfrastruktur für die Agrosystemforschung

1.2 Summary of the proposal in English and German

Agriculture is facing increasing challenges, such as growing food demand but stagnating productivity, climate change, biodiversity loss and degradation of natural resources. Integrated research and data approaches across disciplines and scales are required to provide sustainable solutions. Existing agricultural research data infrastructures (RDIs) are heterogeneous and barely interlinked and lack uniform standards for research data management (RDM). Data is often stored locally or in non-accessible repositories due to the lack of expertise of data creators. However, a functioning RDM is an important prerequisite for integrated, cross-disciplinary research for future agriculture in line with the United Nations Sustainable Development Goals.

FAIRagro is a community-driven initiative of NFDI4Agri and focuses on the agrosystem domain integrating important disciplines and scales needed to develop sustainable crop production and agroecosystems. In this proposal, we comprehensively address the reviewers' comments on the previous NFDI4Agri proposal. FAIRagro focuses on the well-organised domain of agrosystem research to enable researchers a FAIR and quality-assured RDM to generate, publish and access relevant data, innovative RDM services and modern data science methods to support and advance agrosystem research. Based on an open call, the agrosystem community provided use cases to address current challenges for RDM. Six flagship use cases were selected to engage different user groups in the areas of crop breeding and phenotyping, crop nutrient- and pest management and digital farming. They address key research areas and will help scaling up the implementation and use of standards and services for key RDIs - piloting the FAIRification of agrosystem data.

We will establish the FAIRagro Portal as the central access point to our services and create an interoperable and scalable RDI by connecting available repositories to make research data FAIR. We will facilitate combined data analyses with a computational environment and predefined workflows. We will establish a multilevel support system by setting up a Data Steward Service Center, provide guidelines and information material, and focus on knowledge transfer for agrosystem researchers. FAIRagro will address quality and legal security challenges beyond the FAIR principles. Data quality will be ensured through the development of subject-specific quality metrics and curation systems. Privacy policies will be developed to ensure a balance between the interests of data providers and users - including approaches to handling sensitive data. Our aims will be accomplished by close collaboration with other NFDI consortia and the inter/national community of agrosystem data users and providers. We will advance NFDI ambitions with expertise and approaches for legal data security challenges, stakeholder

involvement and RDIs for highly interdisciplinary and multi-scale research in a field of high societal relevance.

Zusammenfassung

Die Landwirtschaft steht vor gewaltigen Herausforderungen. Einer steigenden Nachfrage stehen eine stagnierende Produktivität, der Klimawandel, Verlust der Biodiversität und die Degradation natürlicher Resourcen gegenüber. Integrierte Forschungsansätze über fachliche Disziplinen, sowie räumliche und zeitliche Skalen hinweg sind erforderlich, um nachhaltige Lösungen zu entwickeln. Bestehende agrarwissenschaftliche Forschungsdateninfrastrukturen (FDI) sind heterogen, kaum vernetzt und es fehlen einheitliche Standards für ein gemeinsames Forschungsdatenmanagement (FDM). Forschungsdaten werden oft aufgrund fehlender Kenntnisse der Datenerheber:innen in unzugänglichen Repositorien abgelegt. Ein übergreifendes FDM ist jedoch eine wichtige Voraussetzung für eine integrierte, interdisziplinäre Forschung, die eine zukünftige Landwirtschaft im Einklang mit den Nachhaltigkeitszielen der Vereinten Nationen ermöglicht.

FAIRagro ist eine Community-getriebene Initiative der NFDI4Agri mit Fokus auf die Agrosystemforschung, die wichtige Disziplinen und Skalen für nachhaltige Pflanzenproduktions- und Agroökosysteme integriert. Aufbauend auf den Vorschlägen der Gutachter konzentriert sich FAIRagro auf die Agrosystemforschung. Ziel ist es, Forscher:innen ein FAIRes und qualitätsgesichertes FDM für die Erzeugung und Veröffentlichung sowie den Zugriff auf Forschungsdaten zu ermöglichen, innovative und benutzerfreundliche RDM-Dienste bereitzustellen und moderne *Data-Science*-Methoden für das Voranbringen der Agrosystemforschung zu schaffen. Es wurden sechs Flagship Use Cases aus der FAIRagro-Community ausgewählt, um verschiedene Nutzergruppen in den Bereichen Pflanzenphänotypisierung, Nährstoff- und Pflanzenschutzmanagement und digitale Landwirtschaft einzubinden. Diese adressieren zentrale FDM-Herausforderungen und tragen zur Umsetzung von Standards und Services für ausgewählte FDIs bei.

Mit dem FAIRagro-Portal werden wir einen zentralen Zugangspunkt zu unseren Services etablieren. Durch die Vernetzung existierender, disziplinärer Repositorien werden Forschungsdaten auffindbar und zugänglich. Die Bereitstellung einer Analyseplattform mit vordefinierten Workflows ermöglicht reproduzierbare, integrierte Datenanalysen. Wir werden ein FAIRes FDM durch ein mehrstufiges Support-System mit Hilfe eines Data Steward Service Center‘s und durch die Bereitstellung von Leitfäden und Informationsmaterialien aufbauen. FAIRagro wird Herausforderungen zur Datenqualität und zur Rechtssicherheit der Bereitstellung und Nutzung sensibler Daten adressieren. Diese Ziele werden wir durch enge Zusammenarbeit mit anderen NFDI-Konsortien und der inter/nationalen Community der Datennutzer:innen und –anbieter:innen erreichen. Wir werden die Ambitionen der NFDI mit Fachwissen und Ansätzen für rechtliche Herausforderungen, zur Einbeziehung von Stakeholdern und einer FDI für hochgradig interdisziplinäre und multiskalige Forschung in einem Bereich von hoher gesellschaftlicher Relevanz voranbringen.

1.3 Applicant institution

Applicant institution	Location
ZALF: Leibniz Centre for Agricultural Landscape Research	Eberswalder Straße 84 15374 Müncheberg

1.4 Spokesperson

Spokesperson	Institution, location
Prof. Dr. Frank Ewert	Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84 15374 Müncheberg

1.5 Co-applicant institutions

Co-applicant institutions	Location
FIZ: Leibniz Institute for Information Infrastructure	Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen
FZJ: Forschungszentrum Jülich	Wilhelm-Johnen-Straße 52428 Jülich
IPK: Leibniz Institute of Plant Genetics and Crop Plant Research	Corrensstraße 3 06466 Seeland OT Gatersleben
JKI: Federal Research Centre for Cultivated Plants	Erwin-Baur-Straße 27 06484 Quedlinburg
KTBL: Kuratorium für Technik und Bauwesen in der Landwirtschaft	Bartningstraße 49 64289 Darmstadt
SGN: Senckenberg – Leibniz Institution for Biodiversity and Earth System Research	Senckenberganlage 25 60325 Frankfurt
Thünen: Thünen Institute	Bundesallee 50 38116 Braunschweig
TUM: Technical University Munich	Arcisstraße 21 80333 München
UBN: University of Bonn	Regina-Pacis-Weg 3 53113 Bonn
ZB MED: Information Centre for Life Sciences	Gleueler Straße 60 50931 Köln

1.6 Co-spokespersons

Co-spokespersons	Institution, location	Task area
Prof. Dr. Senthold Asseng	Technical University Munich Liesel-Beckmann-Straße 2, 85354 Freising	TA1
Dr. Til Feike	Federal Research Centre for Cultivated Plants, Erwin-Baur-Straße 27, 06484 Quedlinburg	TA1
Prof. Dr. Jochen C. Reif	Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstraße 3, 06466 Seeland OT Gatersleben	TA1
Birte Lindstädt	Information Centre for Life Sciences (ZB MED), Gleueler Straße 60, 50931 Köln	TA2
Dr. Ulrike Stahl	Federal Research Centre for Cultivated Plants, Erwin-Baur-Straße 27, 06484 Quedlinburg	TA2
Prof. Dr. Franziska Boehm	Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz- Platz 1, 76344 Eggenstein-Leopoldshafen	TA3
Prof. Dr. Jan-H. Haunert	University of Bonn Regina-Pacis-Weg 3, 53113 Bonn	TA3
Daniel Martini	Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL) Bartningstraße 49, 64289 Darmstadt	TA3
Dr. Claus Weiland	Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Senckenberganlage 25, 60325 Frankfurt	TA3
Prof. Dr. Juliane Fluck	ZB MED: Information Centre for Life Sciences, Gleueler Straße 60, 50931 Köln; UBN: University of Bonn Regina-Pacis-Weg 3, 53113 Bonn	TA4
Florian Hoedt	Thünen Institute Bundesallee 50, 38116 Braunschweig	TA4
Dr. Matthias Lange	Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstraße 3, 06466 Seeland OT Gatersleben	TA4
Prof. Dr. Björn Usadel	Forschungszentrum Jülich, Wilhelm-Johnen-Straße, 52428 Jülich	TA4
Dr. Xenia Specka	Leibniz Centre for Agricultural Landscape Research, Eberswalder Straße 84, 15374 Müncheberg	TA5

1.7 Participants

Acronym	Participating institutions	Location
ATB	Leibniz Institute for Agricultural Engineering and Bioeconomy	Max-Eyth-Allee 100 14469 Potsdam
BLU	Bielefeld University	Universitätsstr. 25 33615 Bielefeld
BSA	Directorate General of the Bavarian State Archives	Schönfeldstr. 5 80539 München
DBFZ	German Biomass Research Center gGmbH	Torgauer Str. 116 04347 Leipzig
DBG	The German Soil Science Society	Am Bärenberg 70 37077 Göttingen
DPG	German Society for Plant Protection and Plant Health	Messeweg 11-12 38104 Braunschweig
DWD	Deutscher Wetterdienst	Frankfurter Straße 135 63067 Offenbach
GAUG	Georg-August University Göttingen	Wilhelmsplatz 1 37073 Göttingen
GPW	Gesellschaft für Pflanzenbauwissenschaften	Herrenhäuser Str. 2 30419 Hannover
HSWT	Weihenstephan-Triesdorf University of Applied Sciences	Am Hofgarten 4 85354 Freising
LfL	Bayerische Landesanstalt für Landwirtschaft	Lange Point 12 85354 Freising
TTT	Office Think Tank Digitalisation at Thünen Institute	Bundesallee 50 38116 Braunschweig
UFZ	Helmholtz Centre for Environmental Research	Permoserstr. 15 04318 Leipzig
UHOH	University of Hohenheim	Schloss Hohenheim 1 70599 Stuttgart
VZG	GBV Head Office	Platz der Göttinger Sieben 1 37073 Göttingen
ZEPP	Zentralstelle der Länder für EDV-gestützte Entscheidungshilfen und Programme im Pflanzenschutz	Rüdesheimer Str. 60-68 55545 Bad Kreuznach

Acronym	Contributions of participating institutions
ATB	In TA2, ATB leads Measures 2.2 and 2.3 and thus contributes significantly to community engagement. In TA1 ATB will contribute by providing data, especially for Measure 1.6.
BLU	Operates one of seven de.NBI cloud sites of the German Network for Bioinformatics (de.NBI); will support FAIRagro in establishing services in the de.NBI cloud environment (Measure 4.4); and provides computing and storage resources according to the general rules of the de.NBI cloud.
BSA	Serves as a facilitator and interface to NFDI4BioDiversity and NFDI4Earth with respect to long-term archiving.
DBFZ	Provides bioeconomy data and collaborates on metadata and interfaces.
DBG	Serves as a facilitator and multiplier for the German soil science community and is involved in TA2.
DPG	Serves as a facilitator and multiplier to the plant protection and plant health community and is involved in TA2.
DWD	Public authority that provides data infrastructure and services (TA4, Measure 4.2) and climate data relevant for multiple use cases (Measures 1.1-1.6).
GAUG	Is involved in UC2 (TA1, Measure 1.2), provides expertise on RDM of large area land use data for data aggregation; supports the data steward network (TA2, Measure 2.5).
GPW	Serves as a facilitator and multiplier for the German crop production community and is involved in TA2.
HSWT	Provides sensor data and supports prototype development in the field within Measure 1.6.
LfL	Provides applied expert knowledge and requirements for prototype development in the field within use Measure 1.6.
TTT	Point of contact for several expert networks with a focus on RDM, A.I., digitalisation and RDM-Infrastructures from research institutions subordinated to the Federal Ministry of Food and Agriculture (BMEL), involved in TA2.
UFZ	Is involved in a use case (TA1, Measure 1.2), provides the BonaRes Knowledge Library as an information infrastructure, and serves as a facilitator.
UHOH	Contributes data from wheat population and will apply for an extension of the UC1 (Measure 1.1) for additional crops at a later stage.
VZG	Is involved in a use case (TA1, Measure 1.3), provides infrastructure and serves as a facilitator.
ZEPP	Is involved in a use case (TA1, Measure 1.3), provides infrastructure and serves as a facilitator.

Participating individuals

Participating individuals	Institution, location
Prof. Dr. Anna Maria Häring	German Agricultural Research Alliance (DAFA), Bundesallee 50, 38116 Braunschweig
Dr. Manfred Röhrig	Informationssystem für die integrierte Pflanzenproduktion (ISIP e.V.); Rüdesheimer Str. 60-68, 55545 Bad Kreuznach

- Contribution of Prof. Dr. Anna Maria Häring: as the deputy spokeswoman of the executive board of DAFA, she provides interfaces to a majority of publicly funded institutes in agricultural research and serves as a facilitator and multiplier.
- Contribution of Dr. Manfred Röhrig: is involved in a use case (TA1, Measure 1.3), provides infrastructure and serves as a facilitator.

1.8 Names and numbers of the DFG review boards (DFG-Fachkollegien) that reflect the subject orientation of the proposed consortium

The FAIRagro addresses the DFG research areas Soil Sciences (207-01), Plant Breeding and Plant Pathology (207-02), Plant Cultivation, Plant Nutrition, Agricultural Technology (207-03), Ecology of Land Use (207-04), Agricultural Economics, Agricultural Policy, Agricultural Sociology (207-05).

2 Scope and Objectives

2.1 Research domains or research methods addressed by the consortium and specific aim(s)

Challenges for agriculture and agricultural research

Agriculture is facing tremendous challenges of great societal relevance, ranging from climate change and increasing extreme events with drought and heat spells, increasing natural resource scarcity and degradation, accelerated biodiversity loss to growing food insecurity, malnutrition and rural poverty, increasingly associated with conflict and mass migration. In the face of these challenges, the transformation of agriculture to meet the United Nations Sustainable Development Goals has been stressed in various high level policy documents, such as the Farm-to-Fork-Strategy of the European Green Deal (European Commission, 2019) (European Commission, 2019) and the final report of the National Commission on the Future of Agriculture (Zukunftskommission Landwirtschaft Geschäftsstelle, 2021). Finding socially acceptable and effective solutions to these challenges will require integrated systems approaches that explicitly consider trade-offs and synergies between goals, stakeholders, regions and sectors.

Science is essential in underpinning the transformation of agriculture and developing viable solutions for sustainable agricultural and food systems (von Braun *et al.*, 2021a; von Braun *et al.*, 2021b). However, agriculture sciences cover a wide range of disciplines (from genetics and molecular biology to crop, soil, animal and forest sciences, agro- and geoecology, agrotechnology, market economics and social and cultural sciences) and span spatial and temporal scales and levels of organisation. Despite much progress in certain domains, the overall level of research integration is not yet sufficient to address many of the problems facing agriculture. A national research data infrastructure (RDI) cannot overcome such constraints, but it will be an essential component to facilitate and advance integrated research approaches across disciplines and scales. This was also stated in the evaluation of an earlier proposal of the broader agricultural community to NFDI (Research Proposal NFDI4Agri, 2020) and combined with the suggestion to better focus efforts on research data management (RDM) in agriculture around a research area nucleus with a well-developed and integrated research community.

The agrosystem research domain

In Europe and Germany, approximately 50 percent of land is used for agriculture, largely dominated by crop production for food and feed, fibre and fuel, and shaping rural landscapes. At the same time, public policy tries to reflect the need to improve the resilience and sustainability of crop production reducing the negative impacts on ecosystems and the environment. However, specific measures to meet targets such as the reduction of pesticide and fertiliser use, increasing the area of organic farming, promoting soil carbon sequestration, and enhancing biodiversity and ecosystem service delivery in agricultural landscapes still need to be developed and implemented. The development of viable solutions for these multiple targets utilising new and emerging digital and other technologies requires an agrosystem approach that integrates a broad range of relevant disciplines (Figure 1), from genetics and breeding to crop, soil and geo-sciences, and accounts for interactions with other domains in the agricultural sciences and beyond (e.g., animal sciences, molecular biology, ecology, agricultural economics, social and cultural sciences, nutrition sciences and marine sciences). An agrosystem approach also requires integration of research methods from the respective disciplines including monitoring and experimentation, data analysis and modelling, as well as synthesis and visualisation.

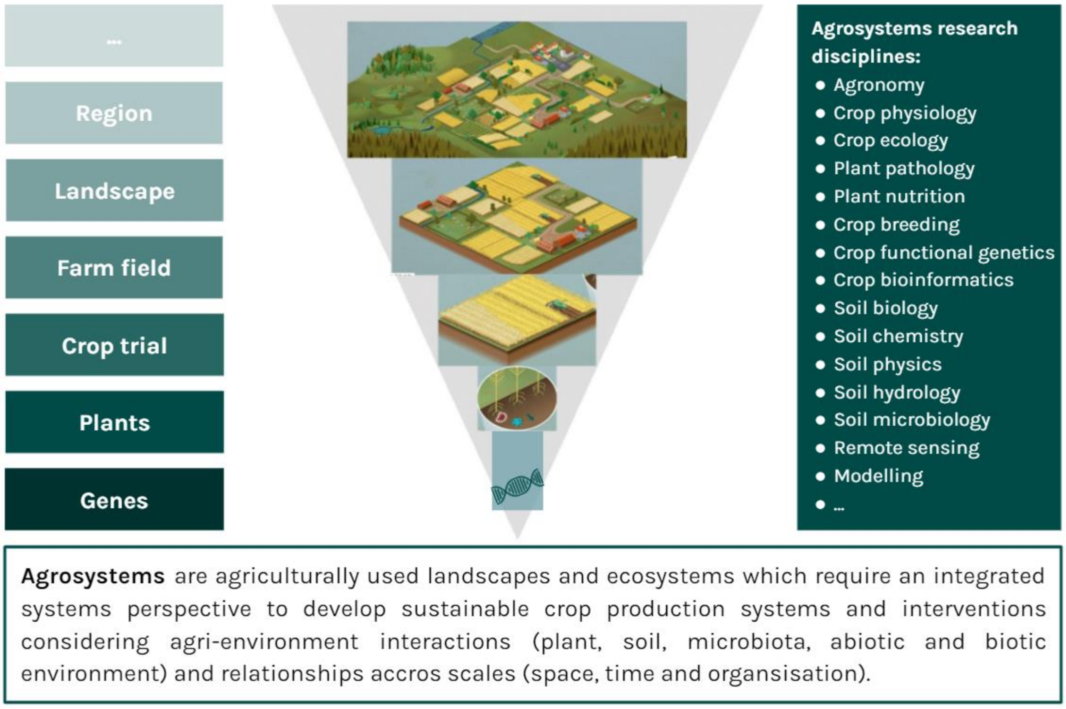

Figure 1: Agrosystem domain with respect to relevant scales and selected relevant disciplines.

The level of integration of agrosystem research is relatively high as achieved and evidenced by performing simulation modelling, integrating monitoring activities and developing data infrastructures [e.g., DAKIS (DAKIS, 2021), PhenoRob (PhenoRob, 2021b) and patchCROP (patchCROP, 2021) projects]. The relatively high level of integration within agrosystem modelling at the international, European and national levels, is evident from large research programmes with thousands of participating scientists, such as the Agricultural Model Intercomparison and Improvement Project (AgMIP, 2021b), the European Knowledge Hub on Modelling Agriculture with Climate Change for Food Security (MACSUR, 2021) and the BonaRes - Soil as a sustainable resource for the bioeconomy (BonaRes, 2021a), with a newly established network for data from long-term field experiments (BonaRes, 2021c). FAIRagro partners play leading roles in each of these national and international initiatives.

FAIRagro aims

To address the complex and multiple problems facing agriculture, FAIRagro aims to advance agrosystem research by strengthening linkages across disciplines, scales and methods through collaborative RDM. By integrating the interdisciplinary agrosystem scientific community into FAIRagro, we endeavor to create the basis for a broad recognition and acceptance by the wider agricultural science community of the benefits and need for FAIR RDM approaches. In developing the FAIRagro proposal, we have explicitly considered suggestions and comments by the reviewers of the last NFDI4Agri proposal in 2020:

- Focus on a research area nucleus, i.e., the agrosystem research domain that is central to the agricultural sciences, and includes a good level of integration of the scientific community from different disciplines and has strong inter/national networks and visibility;
- Strong involvement of researchers as data analysts and users of the proposed RDM measures, infrastructures, services and tools along six clearly formulated flagship use cases;
- Joint formulation of demands by researchers on RDM and demonstration of the added value of the implemented RDM measures for integrated agrosystem research activities;
- Strengthened role of data stewards on the interface between users/researchers and infrastructure developers organised along specific RDM challenges;
- Clear vision and procedure to further build the FAIRagro community and engage with the larger agricultural science community;
- Adapted work program, project structure and consolidated measures to respond to the needs addressed in the use cases;
- Concrete measures of interactions defined with other NFDI-consortia and national/international programs, particularly on topics where FAIRagro can make a unique contribution to integrated systems research in terms of legal data security issues, interdisciplinary and cross-scale approaches and involvments of stakeholders (farmers, policy-makers).

2.2 Objectives and measuring success

Key objectives of FAIRagro are centred on developing RDM services with and for the agrosystem scientific community to enable more robust science and promote FAIR RDM within the wider agricultural community inline with the NFDI and international community (Figure 2). Strong emphasis is placed on users, namely, data analysts and scientists of the agrosystem domain, data creators and infrastructure providers. Key research areas are considered in the form of use cases (see TA1, chap 5.1) that cover a representative portfolio of research questions, methods, disciplines and scales. Topically, the use cases range from more basic research on the use of new digital technologies for sustainable crop rpoduction [e.g., the DFG Cluster of Excellence PhenoRob (PhenoRob, 2021a)] to applied research topics including the link to farmers for improved agrosystem management. Data stewards ensure that the RDM infrastructure development activities match the use cases demands and ensure that the services and tools are developed iteratively to meet the needs of the research community for successful adoption. The FAIRagro Portal is central to the interaction with the agrosystem community and will eventually provide the requested services to internal and external users.

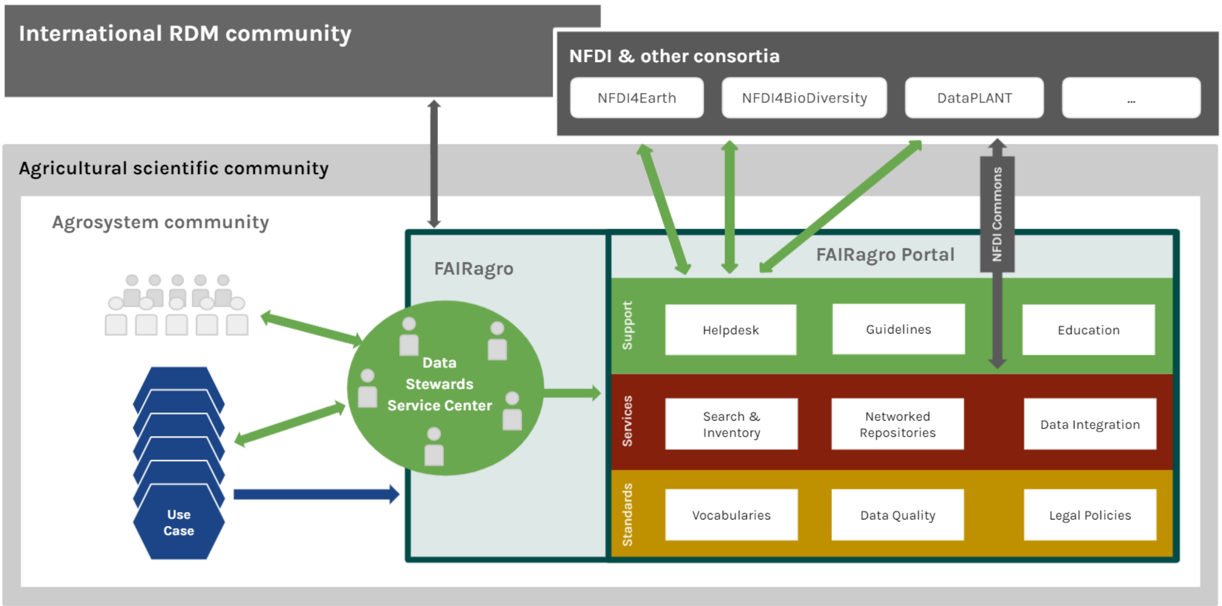

Figure 2: FAIRagro consortium with provided services and its embedding in scientific communities, the NFDI and international community.

The FAIRagro key objectives and measures of success are summarised in Table 1. Our methods and measures of their success are described in the following sections. The measures of success will be regularly evaluated, and the results will be communicated in public reports (see TA5, chap 5.5). Importantly, the cultural change towards FAIR and open RDM within the agrosystem research community and the broader agricultural sciences community will be monitored and communicated within larger forums, such as NFDI (see TA2 and TA5).

FAIRagro will develop RDM approaches in close interaction with NFDI. There is close collaboration with NFDI4Earth, NFDI4BioDiversity and DataPLANT due to the complementarity of the research questions and the associated demand for the sharing of approaches, data and services. Collaborations on specific topics have been established with other consortia, e.g., NFDI4Health, NFDI4Objects, NFDI4DataScience, NFDI4Culture, NFDI4Memory, NFDIxCS, MaRDI, and NFDI4Chem. Finally, research and RDM approaches developed by FAIRagro to enable multidimensional integration will also be of value for the NFDI, particularly in terms of multi-scale and multi-actor approaches and related legal aspects of RDM.

Table 1: FAIRagro key objectives (marked in bold) and measures of success (below).

No.	**FAIRagro key objectives** and measures of success
1	**Enabling researchers, students, and infrastructure providers to achieve an open and sustainable research data life cycle and foster a cultural change towards FAIR and collaborative RDM.** - Establish the Data Steward Service Center - Establish a helpdesk, including a ticket support system - Provide RDM education via lectures, courses, train-the-trainer events, workshops, and summer schools - Provide education & training material; establish education programs at graduate schools - Promote and implement FAIR RDM in academic societies - Establish a community advisory board - Promote regular attendance at FAIRagro plenaries and community workshops - Present appropriate responses to new calls for use cases - Develop and use self-evaluation criteria for effectiveness of measures
2	**Continuously identify RDM challenges in the agrosystem community and provide solutions via continuously developed RDM services to meet the needs of our community and develop FAIRagro to NFDI4Agri.** - Call for new use cases and number of submitted use case proposals - Integrate new use cases in FAIRagro and successful completed use cases - Establish a community advisory board - Launch community summits and workshops - Provide community surveys on community needs and feedback of our service - Extend the FAIRagro community to include other research disciplines in a structured onboarding process for use cases - Ensure that FAIRagro covers additional agricultural research disciplines over time
3	**Establish the FAIRagro Portal as the central access point for RDM in the agrosystem sciences to enable knowledge exchange, technology transfer and participatory processes to promote dialogue with our community.** - Promote frequent use of the FAIRagro Portal and the help desk by the community - Set up the knowledge base in the FAIRagro Portal - Publish information material, guidelines, and best-practice recommendations in the knowledge base of the FAIRagro Portal
4	**Improve the findability of published research data and existing data repositories and provide discipline-specific interoperable infrastructures to enable standardised data exchange that will support the development of new, integrated research approaches and thus greatly enhance agrosystem research.** - Connect the available disciplinary repositories with an interoperable network via the middleware under consideration of the NFDI-RDC - Implement a user-friendly, machine-actionable search service in the FAIRagro Portal - Enable semantic queries within the connected, federated, disciplinary infrastructures - Provide the number of connected repositories, data retrieval requests, transferred data volume, performed search queries, findable services and findable datasets
5	**Establish a set of standards, guidelines and best-practices as guidance for researchers to promote and enable FAIR RDM and infrastructure providers.** - Developed and publish a minimum set of metadata schema for FAIRagro

No.	**FAIRagro key objectives** and measures of success
	- Publish a domain-specific extension for schema.org - Map disciplinary domain standards, controlled vocabularies and ontologies developed to satisfy national and international requirements for the long-term integration of agricultural research data into a single common NFDI
6	**Ensure research data quality via domain-specific measures of quality control and establish a quality feedback and curation system to facilitate the reuse of research data by agricultural scientists in, for example, model simulations.** - Develop and publish quality metrics considering agricultural applications - Present information on data quality from the metadata of FAIRagro infrastructures - Provide usage statistics of provided services
7	**Provide legal certainty to infrastructure providers, data providers and users for the publication and reuse of sensitive data.** - Develop and publish guidelines, policies and ethical standards with respect to the publication and reuse of sensitive data - Provide the number of processed legal support requests and supported cases
8	**Enable reproducible research results, the deployment and publishing of models and data analysis workflows, and contribute to the reproducibility of research results.** - Develop and launch a scientific workflow infrastructure (SciWIn) - Present the number of analysis runs, published models and deployed workflows container of SciWIn - Provide usage statistics on computation infrastructure
9	**Ensure the interoperability of an agrosystem research data infrastructure within related domains in the agricultural sciences and within NFDI (e.g., NFDI4Earth, NFDI4BioDiversity and DataPLANT) as well as with the overarching NFDI and international initiatives (e.g., EOSC, GAIA-X, RDA), contribute to cross-cutting topics within the NFDI and develop synergies with other consortia.** - Perform data exchange realised with infrastructures from other NFDIs - Participate in the NFDI Senate - Contribute to cross-cutting topics and to the usage of NFDI commons - Encourage close cooperation with other NFDI consortia
10	**Contribute to and lead European and international developments of RDI in agrosystem and agricultural research.** - Promote the international visibility of FAIRagro - Provide a voice for the German agrosystem and agricultural RDM communities in international initiatives - Contribute to Research Data Alliance (RDA) and international standardization - Maintain leading roles in European efforts such as the upcoming European Partnerships “Agriculture of Data”, “Agroecology Living Labs” and the Science-Policy Hub MACSUR

3 Consortium

3.1 Composition of the consortium and its embedding in the community of interest

Agrosystem research comprises many disciplines, and the FAIRagro consortium represents a large part of the agrosystem research community in Germany, including typical data providers and (re-)users, and providers of important agricultural RDI. FAIRagro partners also represent different types of institutes, universities and universities of applied sciences, federal and state research institutes, infrastructure facilities, and professional associations within the agrosystem domain (Figure 3).

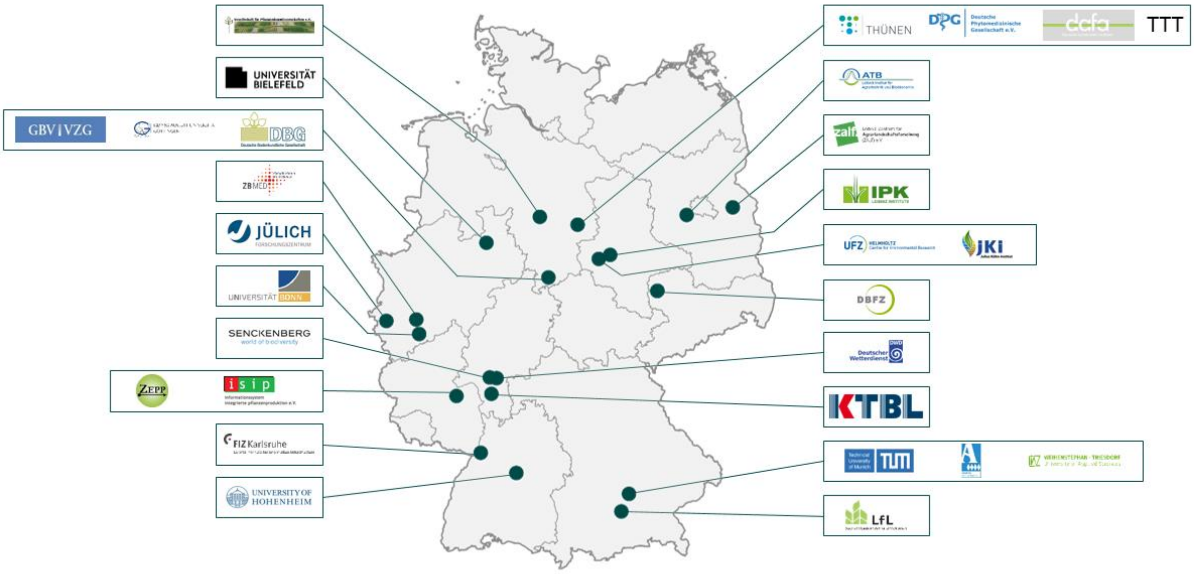

Figure 3: Map of (co-)applicants (11) and participants (18) in the FAIRagro consortium (see chap. 1), represented by the logo of their institution. Note that several locations comprise more than one partner.

FAIRagro partners include internationally well-recognised researchers in the field of plant phenotyping and breeding, agronomy, crop and soil sciences, and remote sensing and modelling. Due to the broad range of scientific and RDM expertise contained in these institutes, the applicant and most co-applicants fulfil dual roles as infrastructure providers, creators and users of agrosystem data. The participants support the (co-)applicants with their domain-specific expertise, with data infrastructure, and/or as multipliers in the agrosystem community.

Developing the NFDI4Agri nucleus FAIRagro

The FAIRagro consortium was formed in early 2021 as an initiative of the NFDI for Agricultural Sciences (NFDI4Agri) community with the goal to develop a FAIR RDM for a specific agricultural research domain as nucleus: the agrosystem research domain. This development anticipates a central recommendation by the evaluation panel reviewing NFDI4Agri in 2020.

Forming NFDI4Agri: In 2018, before the NFDI call for proposals, data scientists and providers of several agricultural RDIs and services in Germany were actively engaged in networks with sound insights into the current research data landscape and the corresponding infrastructure facilities. These networks covered a representative part of the agricultural sciences community and formed the core from which the NFDI4Agri consortium was developed. In a community workshop in June 2019 attended by many representatives from different agricultural research institutions, it was decided to establish the NFDI4Agri consortium with ZALF as the lead and applicant institution (Figure 4). More than 30 partners formed the consortium to develop a joint proposal submitted in 2019. The proposal was not recommended for funding because of concerns by the reviewers related to: (a) insufficient identification of community-relevant measures, (b) lack of involvement of research areas such as the animal and livestock, and forest sciences, and (c) weak governance structure and maturity of the consortium.

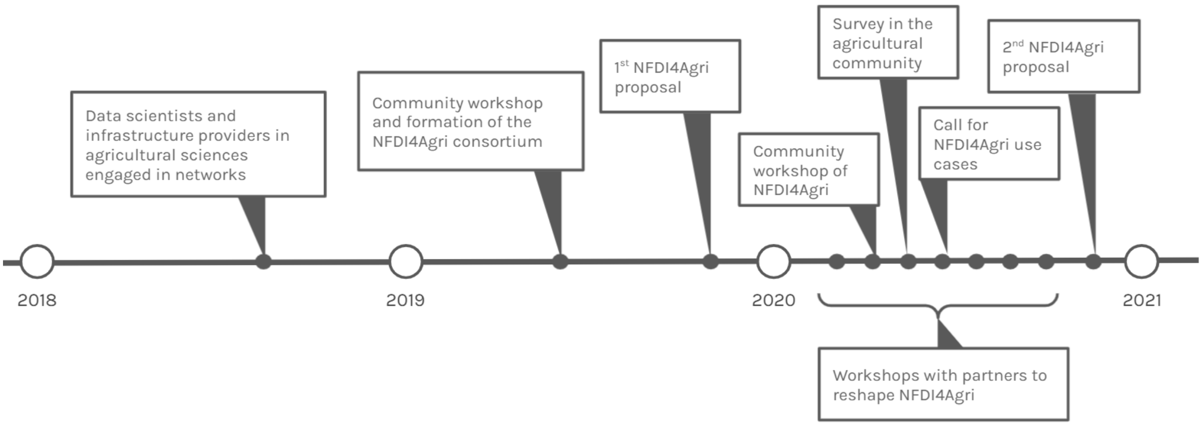

Figure 4: Major milestones and activities of the NFDI4Agri consortium between 2018 and 2021.

Reshaping NFDI4Agri in 2020: The community development continued and because of the large interest and demand expressed by the community it was decided to develop a new proposal. The proposal consortium involved all co-applicants of NFDI4Agri, and new partners were integrated covering additional research domains from animal and forestry, e.g., the Technical University Dresden and University of Göttingen. Frank Ewert, the Scientific Director of ZALF, was appointed as the new spokesperson of NFDI4Agri. The governance structure was revised to describe a clearly defined decision process and conflict management. In June and July 2020, a workshop and a community survey were performed to identify the needs of the broad agricultural community. At the community workshop of NFDI4Agri with more than 120 participants, an open use case call was launched to identify specific community needs and RDM challenges (Figure 4). More than 37 use cases have been submitted, among which 30 were selected to be addressed by NFDI4Agri in the proposal submitted to the second NFDI call in 2020. The international reviewing panel acknowledged the relevance of the agricultural sciences for the NFDI as a whole, the improved representative composition of the consortium and its expertise. It positively assessed the organisational changes and the adapted governance structure. However, particular concerns were expressed about the low level of integration across disciplines of agricultural sciences and the large number of use cases. Instead, a focus on a smaller but better-integrated research domain and flagship use cases was strongly suggested. It was also recommended to extend the tasks of the planned data stewards to a stronger integration of the different disciplines. These concerns guided the decision of the NFDI Expert Group to not recommend funding of the second NFDI4Agri proposal.

Formation of FAIRagro in 2021: Because of the persistent large demand and interest by the community as well as the high relevance for a RDI in agricultural sciences, the NFDI4Agri steering committee decided to develop a third proposal with substantial revision of the focus, work direction and structure, following the reviewers comments. In close exchange between community members and leading German research institutes, the agrosystem research domain was selected as a nucleus area. The agrosystem research community already has achieved a relatively high level of integrated research with good but still fragmented RDM solutions. Improved RDM solutions would greatly advance integrated research of this community. At a workshop in June 2021 with more than 60 participants, the new focus of the FAIRagro consortium as an initiative of NFDI4Agri was presented to the community. As a result the community supported this development and also agreed on some changes regarding the composition of the FAIRagro consortium. A number of NFDI4Agri partners with expertise in agrosystem research and RDM remained in FAIRagro while some other partners related to research domains outside the agrosystem nucleus, such as the animal and forestry sciences, have not been further included in FAIRagro at this stage but remain in NFDI4Agri. Also, new partners were involved in FAIRagro to better balance the consortium's expertise through stronger involvement of researchers and data analysts from the agrosystem domain.

Our ambition: From FAIRagro to NFDI4Agri to NFDI

We initiated new activities with the agrosystem community with the objective of identifying the community's specific needs. First, we established a task force in the agrosystem domain focusing on sustainable crop production and agroecosystems (Figure 5, see chap. 4.1).

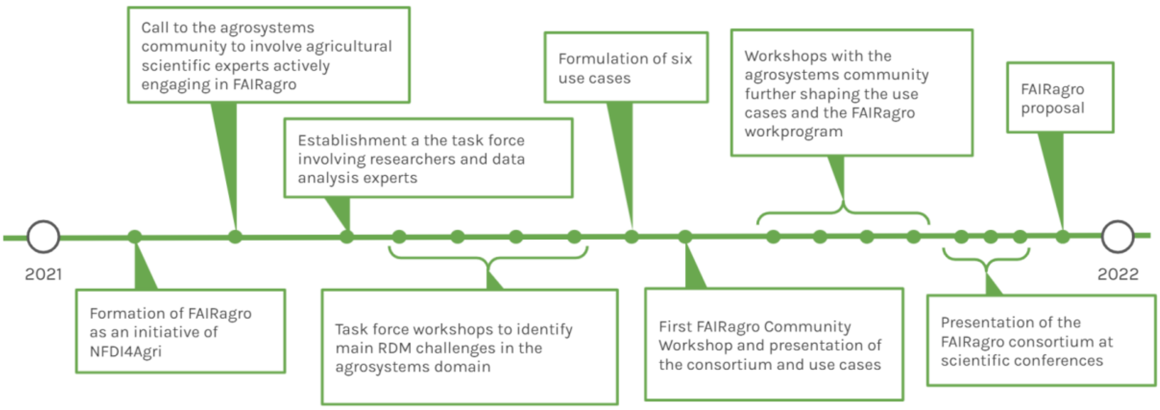

Figure 5: Activities of the FAIRagro consortium in preparation for this proposal in 2021.

In four workshops, the task force identified the main RDM-related challenges of agrosystem research, which resulted in the formulation of six flagship use cases. These use cases were discussed and further shaped in several workshops with broad participation of the community. Criteria for selecting the use case were a representative coverage of relevant topics in the agrosystem domain and important RDM challenges for which researchers and data scientists shall jointly develop solutions. Several members of the agrosystem task force have been integrated as co-spokespersons into FAIRagro to better represent the needs of researchers in the consortium.

Involving the community: We will follow different approaches to engage with the agrosystem community. As part of our governance (see chap. 3.5) we will establish a **Community Advisory Board** (CAB) that includes members representing the agrosystem community but also some representatives of other agricultural domains and of the data sciences. We have already been active in presenting the consortium at scientific conferences related to the agrosystem domain, such as the German Plant Protection Conference (Ewert *et al.*, 2021c), the annual meeting of the Society of Crop Science [GPW, Ewert *et al.* (2021b)] and the DAFA (Ewert *et al.*, 2021a). Furthermore, we introduced the ambitions of the FAIRagro consortium and the procedure of onboarding of further use cases (TA2, Measure 2.3) in the FAIRagro Community Workshop in June 2021 to further advance community building and to obtain a more specific overview on FAIRagro needs and community contributions. The current flagship use cases identified by the task force demonstrate the potential of implementing and automating FAIR standards to substantially increase the research output covering a wide range of agrosystem relevant levels of organization (genes, plants, crop trial, field, and region). These use cases are integrated into a separate task area into the work program of FAIRagro (TA1, chap. 5.1). This onboarding process has been further structured to support further onboarding activities and the development of FAIRagro (TA2, Measure 2.3) to address new research areas and communities. We will initiate three further calls for use cases in 2024, 2025 and 2026 that will be evaluated by the Steering Committee of FAIRagro based on recommendations of the CAB (Figure 6).

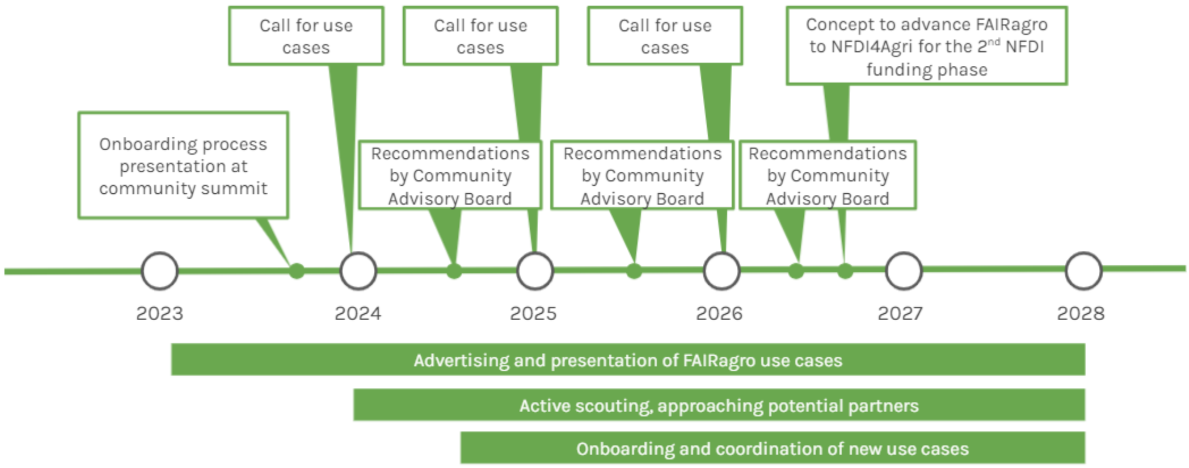

Figure 6: FAIRagro procedure and calls for onboarding of new use cases. The first set of flagship use cases was selected in 2021 and will start in the early project phase. Further use cases will be onboarded in three phases to address further challenges in the agrosystem research or advance exististing use cases.

Newly submitted use cases can be developed as additional agrosystem research use cases (NUC - New Agrosystem Use Cases, Figure 7) or as extensions of existing use cases (EUC -

Extension Use Case), such as with extended content-related or new methodological links to FAIRagro services. In an advanced later stage of the project, the focus will be on bridging use cases (BUC - Bridging Use Case) that allow for extension of the FAIRagro community to other domains of agricultural sciences considered in NFDI4Agri. An example could be the Leibniz Innovation Farm (ATB Potsdam, 2021) which combines research on animals and crops and biorefinery for a biobased circular economy bridging several domains of the agricultural sciences.

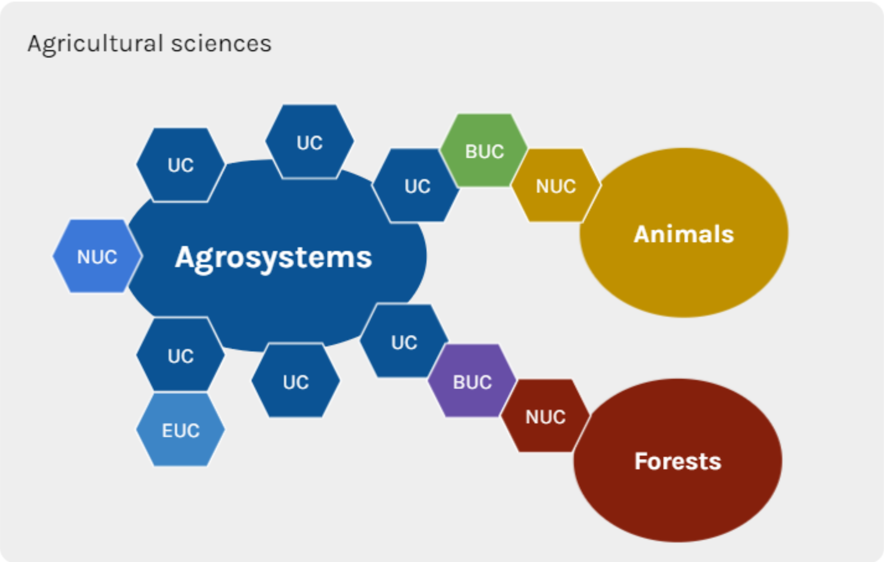

Figure 7: Process of further developing FAIRagro to include additional disciplines and domains. The present set of flagship use cases (UCs) of the agrosystem nucleus can be extended through onboarding of new use cases, either as extension of existing use cases (EUCs) or new agrosystem use cases (NUCs). Bridging use cases (BUC) can be proposed to expand FAIRagro and integrate further agricultural domains, such as animal or forest sciences.

To coordinate and structure the onboarding process, we developed an established list of use case criteria that consists of various elements, such as the added value compared to existing use cases, connection points to existing UCs, potential integration of new domains and communities, feasibility and capacity of the use cases and strategic eligibility of the use cases.

Attracting the next generation of researchers: FAIRagro involves many including academic societies, who will continuously act as multiplicators in their respective research communities and domains. We will network with academic societies and use surveys, roadshows, workshops and conferences to actively engage the FAIRagro consortium in the communication and capacity building. Furthermore, we will establish a training and education program in universities to develop RDM and data science competencies for students and young researchers. For this, we already cooperate with different partners, such as NFDI4BioDiversity and NFDI4Earth (see chap. 3.2 and Measure 5.3) or FID ForstLand (whose proposal is currently reviewed in the DFG funding program *Specialised Information Services* (DFG, 2021). FAIRagro and FID ForstLand agreed to collaborate on a joint training concept.

Reshaping the work program: In FAIRagro, we introduce new concepts for engaging and involving our community and making FAIRagro even more open to the community. We reshaped our work program accordingly and formulated tangible milestones and deliverables substantiating the five-year implementation plan. A risk assessment and appropriate contingency measures are provided (see chap. 3.4) and our services (see chap. 4.4) will be iteratively developed with early but regular feedback from the community to align the service according to their needs.

Composition, expertise and roles of the FAIRagro consortium

The following sections summarise the characteristics and strength of the applicant and co-applicant institutions and their involvement and planned role in the consortium.

The **Leibniz Centre for Agricultural Landscape Research (ZALF)** is a German research centre with more than 400 employees, approximately half of whom are scientists. ZALF's mission is to provide society with a knowledge base and solutions towards landscape-scale sustainable land management through inter- and transdisciplinary research, i.e., soil science, agronomy, plant sciences, microbiology, hydrology and socioeconomic research. ZALF maintains a unique infrastructure for RDM and the modelling of agricultural systems. ZALF develops and manages (together with UFZ) the BonaRes Centre for Soil Research, which is the coordinating centre of the large national research programme ”‘BonaRes – Soil as a natural resource for the bioeconomy”. ZALF is a well-known partner in national and European research projects with extensive experience in coordinating European projects. ZALF, as an applicant, is experienced in coordinating large agricultural research networks and consortia, e.g., INKA-BB (INKA-BB, 2021) and DAKIS (DAKIS, 2021). It is an experienced player in RDM through its operation of the BonaRes Repository and passes this knowledge on to the FAIRagro data stewards (Measure 2.5). The BonaRes Repository provides soil and agricultural research data with a special focus on soil profiles and long-term experiments (LTE) that will contribute to FAIRagro. ZALF offers regular training on all areas of RDM in agricultural sciences with the aim of data publication. Furthermore, ZALF is experienced in networking for data acquisition and provision. ZALF provides DOI for complex and related research data and develops mechanisms for long-term preservation (archiving) of research data together with ZB MED. The FAIRagro Spokesperson **Frank Ewert** has an international track record as a leading scientist in agrosystem research and modelling with a particular focus on research related to climate change impact assessment, food security and sustainable crop production. He has been one of the few highly cited researchers in agricultural sciences in Germany (Clarivate Statistics, ISI Web of Science) in 2016, 2017, 2018, 2019 and 2020, and has been nominated for 2021. He is also one of the top1000 world's most influential climate scientists (Rank 83) according to Reuters' Hot List. Frank Ewert is PI in the DFG PhenoRob cluster of excellence at the University of Bonn and a leading advisor for national and international committees and organizations (e.g., Chair of the Scientific Advisory Board of FACCE JPI, member of the Scientific Group of the UN Secretary-General’s 2021 Food Systems Summit). **Xenia Specka** has a background in computer science, environmental informatics and agronomy, and has participated in several AgMIP and MACSUR model comparison studies. She heads the ZALF “Data Infrastructures” working group, is a key person in developing and implementing its data strategy and leads the development of the BonaRes Repository. Together with Frank Ewert, she coordinated the proposal for FAIRagro and will lead the FAIRagro Secretariat (Measure 1.1) as the project administrative coordinator. ZALF will lead TA5, and contribute to TA1 (Measures 1.2 and 1.4), TA2 (Measure 2.5), TA3 (Measure 3.3) and TA4 (Measures 4.1 and 4.2). Furthermore, leading scientists from ZALF, **Heidi Webber** who has a background in agricultural engineering and modelling and was recently awarded by the female professorship program of the Leibniz Association (Professorinnenprogramm) and **Gunnar Lischeid**, who is a joint Professor in Landscape Hydrology with the University of Potsdam has a background in data science and agricultural landscape water regimes, were involved in the development of two FAIRagro use cases.

Leibniz Institute for Information Infrastructure - FIZ Karlsruhe (FIZ) provides research infrastructure for academia and business and conducts research in computer science and on legal questions regarding the entire set of content, technologies, methods, processes, and associated services that enable the generation, distribution, and preservation of knowledge. FIZ research areas include intellectual property rights in distributed information infrastructures specializing in copyright law, licensing, data privacy law, and IT security. FIZ provides legal support for RDM, developing legal best practices and recommendations, and is involved in other initiatives, e.g. NFDI4Culture, NFDI4Chem, NFDI4DataScience, the NFDI-section ELSA and FAIR Data Spaces in EU GAIA-X. FIZ analyses the legal requirements for the FAIR principles, Open Access (OA), and upcoming regulations for new tools and technologies like A.I. or text and data mining (TDM). Co-spokesperson **Franziska Boehm** is the Vice President of Intellectual Property Science and Professor for IP-Rights at Karlsruhe Institute for Technology (KIT). FIZ leads Measure 3.6 (legal framework, policies, guidelines and legal metadata standards) and contributes to the Data Steward Service Center (Measure 2.5).

Forschungszentrum Jülich (FZJ) plays a leading role in the field of plant phenotyping, develops new sensors and measurement concepts and integrates them into semi-automated and fully-automated systems. FZJ coordinates the European Phenotyping infrastructure EMPHASIS-PREP and is a member of ELIXIR-de, and the MIAPPE steering committee. It is developing BrAPI adapters and connectors for plant phenotyping installations and data integration solutions for EOSC. FZJ contributes to PlabiPD, which hosts data on functional annotation of plant genomes using the MapMan Annotation to FAIRagro and curates genomic data on German crop plants. As a co-lead of TA4, FZJ will take responsibility for the implementation of the central search service (Measure 4.3), with a focus on providing the infrastructure registry. **Uwe Rascher** is a leading scientist in remote sensing, from the field to satellite data and contributes to the use case 5 in Measure 1.5. **Björn Usadel**, a highly cited researcher in plant science, is a co-lead of TA4. He also acts as a bridge to the related

DataPLANT consortium, where he leads Taks Area 1 “Standardization, Quality, and Interoperability”.

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) combines conservation, research and exploitation of the genetic diversity of cultivated plants in the Federal Ex situ Gene Bank through innovative research approaches in the fields of molecular genetics, genome research, molecular plant biology, systems biology, bioinformatics and modelling. IPK has expertise in RDM training courses in the de.NBI network and is a board member of the ELIXIR plant community for common standards and RDM practices for plant science of MIAPPE steering committee. IPK is also a member of the BrAPI advisory board and is mandated by the ECPGR Steering Committee to manage the European Search Catalogue for Plant Genetic Resources [EURISCO (Weise *et al.*, 2017)], which offers services and guidance in assigning PIDs for plant genetic resources (Weise *et al.*, 2020), and membership in the DataCite and ORCID-DE consortium. The Bioinformatics and Information Technology research group of the IPK offers its experience in the field of plant-related research data standardization and its connection to relevant communities. IPK contributes: (1) GBIS/I, the information system of the IPK ex situ collection for crop and wild relatives, and (2) e!DAL-PGP, a plant genomics and phenomics (PGP) research data repository. IPK will support the outreach and coordination activities to plant RDM programs at the European and national levels, including the coordination of FAIRagro activities with the ELIXIR plant community, DivSeek, ECPGR, and de.NBI. IPK brings profound knowledge and expertise in the field of specific RDM to the FAIRagro data steward network (Measure 2.5). Senior Scientist **Matthias Lange** contributes to the above mentioned FAIRagro activities by leading the team responsible for scientific RDM at IPK and leading TA4 of FAIRagro. IPK contributes to TA3 Measures 3.5 and 3.6. Furthermore, **Jochen Reif** leads the Department of Breeding Research at IPK, co-leads TA1, and demonstrates his scientific excellence in breeding methods to FAIRagro by leading the implementation of a use case (UC1, Measure 1.1).

Federal Research Centre for Cultivated Plants (JKI): One of the main strengths of JKI is its long-term, continuous work on research issues that are highly relevant to making agricultural data available in a structured and user-friendly way. The JKI RDM covers different agricultural areas including plant genetics, plant breeding research, crop production, plant nutrition, soil science, plant protection, spatiotemporal information, and time series of weather and satellite imagery. Furthermore, JKI brings in the infrastructure OpenAgrar (OpenAgrar, 2021), a collective repository of research institutions that are affiliated with the Federal Ministry of Food and Agriculture (BMEL) in Germany. Co-spokesperson **Ulrike Stahl** from JKI leads TA2 and Measure 2.1 and coordinates direct interaction with the community. She has a background in biology with a strong emphasis on data driven research. For the past five years she has coordinated the RDM of the JKI with a focus on open data, data publication, data policies, data management plans, and implementing the FAIR principles in data infrastructures. Furthermore, she is the spokesperson of the working group “Research Data Management of the research institutions subordinated to the Federal Ministry of Food and Agriculture (BMEL)", thus bringing together a living network of experts in RDM in agricultural science. Co-spokesperson and TA1 co-lead **Til Feike** is a research group leader at JKI. As an expert in data-driven and model-based analysis in agrosystem research, he leads the implementation of a use case (UC3, Measure 1.3). **Markus Möller** from JKI brings geodata expertise to the Data Steward Service Center (Measure 2.5). He is involved in the further development of the JKI geodata infrastructure (JKI-GDI), specifically in incorporating internal high-speed computing capacities such as the JKI Data Cube (Gomes *et al.*, 2020b) and external cloud-based capacities such as CODE-DE (Storch *et al.*, 2019). Applying OGC-compliant standards (Baumann *et al.*, 2021) ensures that these components are interoperable.The JKI RDM group, which runs the JKI Geonode (Geonode, 2021) platform and supports the curation of provenance records associated with the JKI geodata infrastructure, is responsible for establishing and extending data quality metrics (Measure 3.3).

Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL) is a registered association with more than 400 members from across agriculture and industry, public administration, and consulting sectors. Its main mandate is knowledge transfer and providing planning data, facts and information for agricultural decision-making on all levels from production planning to policy advice. KTBL promotes future-oriented agriculture that not only supports farmers' living and economic outcomes, but also conserves the environment and climate and supports societal goals such as sustainability or animal welfare. Leveraging semantic technologies to create an interoperable and coherent data space that enables experts to tackle current societal and environmental challenges in agriculture has become a focal topic at KTBL. **Daniel Martini** is the deputy head of the databases and knowledge technologies team and is responsible for research and development in that area. His team has been involved in a number of projects, including e.g., iGreen, SimLearn (funded by the German Ministry for Research and Education), agriXchange, FIspace (funded by the EU), PAM and GeoBox (funded by the German Ministry for Agriculture). The team cooperates closely with the AGROVOC core team at the FAO, and supports conceptual, editorial and technical work on this thesaurus and linked dataset and is well-connected and involves with initiatives such as GODAN, the RDA, and the AEF on issues of information management in agriculture. KTBL is the lead of TA3 and Measure 3.1 and contributes its knowledge on metadata, data standards, and ontologies as well as its experience in the use of semantic technologies and mapping and alignment tools.

The **Senckenberg - Leibniz Institution for Biodiversity and Earth System Research (SGN)** is Germany's largest natural history research facility and comprises seven research institutes and three museums. SGN has experience in developing and operating international RDIs with partly sensitive data and has collaborations and data exchanges with many partners. SGN hosts and contributes substantial online resources in the field of agrobiodiversity - via the soil organism database Edaphobase. SGN specializes in harmonizing and linking heterogeneous data sources and types within data warehouses and in machine-learning-based trait extraction and subsequent semantic annotation of plant characteristics. **Claus Weiland** is the leader of the development team that maintains Senckenberg's biodiversity databases and collection management system. He leads the work package "Technical architecture & service provision" of the H2020 project DiSSCo Prepare. Weiland is also a member of DiSSCo's Technical Team responsible for the technical design of the overall data infrastructure and allied elements for the ESFRI project Distributed System of Scientific Collections. He is a co-chair of the Technical Specification & Implementation Group (FDO-TSIG) of the FDO-Forum, which aims to establish an ecosystem of FAIR Data and Services. He also co-leads the task area 2 "Interoperate" in the NFDI Consortium Earth System Sciences (NFDI4Earth). He is co-lead of TA3 and the lead of Measure 3.5 and will focus on developing the FAIR Digital Object (FDO) Architecture of FAIRagro. As co-lead of Measure 4.2, SGN will bring in expertise implementing the middleware, i.e., the Authentication and Authorisation Infrastructure (AAI).

The **Thünen Institute (Thünen)** is a German Federal Research Institute under the auspices of the German Ministry of Food and Agriculture. As part of its federal monitoring programmes. Thünen is responsible for conducting the National Soil Inventories and the National Forest Inventory and reporting the results to the Federal Ministry of Food and Agriculture. These data, as well as the expertise on data processing and usage, will be contributed to FAIRagro and connected to the FAIRagro network of infrastructures as part of Measure 4.2. Thünen is represented by **Florian Hoedt,** who is the team leader of the RDM service unit and part of the center for information management. Moreover, Thünen is involved in an international community of developers through its contribution to the Open-Source Geospatial Foundation (OSGeo) project GeoNode. As the lead of Measure 4.4, Thünen brings its expertise on standards for managing geospatial data and its geoinformatics skills and is responsible for the development and prototype deployment of scientific workflow infrastructure (SciWIn).

The **Technical University of Munich (TUM)** is one of the leading universities in Europe and one of the top three universities of excellence in Germany. It shows outstanding performance in research and teaching, interdisciplinarity, and talent promotion with consistent excellent results in international and national rankings. NFDI has high priority at TUM, which shows among others through its current participation in different NFDI projects (FAIRmat, MaRDI, GHGA, NFDI4Microbiota, NFDI4Cat, DAPHNE4NFDI, DataPLANT, NFDI4Ing, NFDI4Earth, and FAIRagro). Several TUM professors play important active roles as co-spokesperson. To establish productive synergies and exchange between the different consortia, the Munich Data Science Institute at TUM is committed to establish a network that connects all the NFDI-activities within TUM. The Central Institute Hans Eisenmann Forum for Agricultural Sciences (HEF) links all chairs and institutions of the TUM with a focus on agricultural science. It supports research cooperations with other agricultural science institutions at the Weihenstephan research campus and beyond. Its director **Senthold Asseng**, chair of Digital Agriculture, is an internationally, highly recognized researcher in crop modelling. He brings experience in successfully co-coordinating international research consortia, such as AgMIP, with over 1000 members. He is recognized as a Fellow of the American Association for the Advancement of Science (AAAS), was awarded Highly Cited Researcher by Web of Science in 2019, and is in the top1000 s of the world's top climate scientists (#192) according to Thomson Reuters. His contributions to FAIRagro are twofold: 1) through his long-term experience in research coordination, he will greatly contribute in his role as TA1 lead to support the coordination between Measure 1.6 and all Measures of TA1; and 2) moreover, he will provide significant expertise in research and data requirements for modelling atmosphere-plant-soil-systems within Measure 1.6. Furthermore, the Graduate Centre Weihenstephan of TUM (TUM-GZW) supports training and education activities of FAIRagro (TA2). It offers infrastructure for conducting courses for young scientists and doctoral researchers and supports the development of educational modules for graduate schools.

The **University of Bonn (UBN)** is one of Germany's Universities of Excellence. Researchers within its agricultural faculty in nutritional and food sciences, geodesy and geoinformation, foods and resource economics, agricultural engineering, and animal science jointly work towards solving global challenges, such as food security and resource conservation. Structural elements such as regional networks and off-campus research facilities support research, teaching, and knowledge transfer. In particular, the agricultural faculty comprises facilities and 181 hectares of arable land at Campus Klein-Altendorf, where research on crop cultivation and breeding is conducted under natural conditions. The cluster of excellence, "PhenoRob - Robotics and Phenotyping for Sustainable Crop Production", is a collaboration between UBN and FZJ and contributes its RDI to FAIRagro. PhenoRob aims to develop new methods and technologies that can be used to observe, analyse, understand, and manage plants better. As the only cluster of excellence in agriculture in Germany, PhenoRob aims to reduce the environmental footprint of crop production, maintain the quality of soil and arable land, and analyse the best routes to improve the adoption of technology. **Jan-Henrik Haunert** is a professor for geoinformation and vice dean for research at the agricultural faculty of UBN. He co-leads TA3, leads Measure 3.4, contributes to the Data Steward Service Center (Measure 2.5), and provides scientific background knowledge on algorithms and data structures for geoinformation and information visualization. He is an associate member of the PhenoRob cluster of excellence and is responsible for RDM. His group has developed the PhenoRob RDI for storing large volumes of heterogeneous data that are collected on Campus Klein-Altendorf with mobile sensor systems, including field robots and drones. **Juliane Fluck**, who has a background in biology and immunology, started working in the emerging field of bioinformatics at the Fraunhofer Institute for Algorithms and Scientific Computing SCAI. She has been a professor of information management at the agricultural faculty at the University of Bonn and has headed a research group for knowledge management at ZB MED since 2018. She is an associate member in the cluster of excellence PhenoRob and supports at UBN the RDM of PhenoRob and contributes to the Data Steward Service Center (Measure 2.5).

The expertise of the **Information Centre for Life Sciences (ZB MED)** lies in providing information and data for research, teaching and practical applications in the life sciences. Additionally, ZB MED conducts projects in the field of information science and provides infrastructure, services, training, and advice for RDM. Notably ZB MED provides publication and search infrastructure (e.g., PUBLISSO, LIVIVO). ZB MED is currently involved in three NFDIs, NFDI4Health, NFDI4Microbiota (both as the applicant institution) and NFDI4DataScience. **Juliane Fluck** heads a research group for knowledge management at ZB MED. Key topics of her research include text and data mining, semantic data integration and the electronic availability and interoperability of data and knowledge. She coordinates the national infrastructure of personalised health data, NFDI4Health, which is one of the nine national RDIs that started in October 2020. Juliane Fluck co-leads TA4 and Measure 4.3, leads Measure 3.2, and participates in Measure 3.1. In these measures, she will develop and host the central search service, contribute to the development of publication guidelines and the metadata model, and will provide a terminology service for FAIRagro ontologies and terminologies. **Birte Lindstädt** has a background in geography and information science and heads the unit of RDM services at ZB MED. Her responsibilities include developing technical information infrastructure, consulting and training. She coordinates the NFDI activities of ZB MED and is involved in standards for publication, metadata, training and data stewardship pilots NFDI4Health. Thus, she is already a member of the NFDI section training and education. Birte Lindstädt co-leads TA2, heads Measure 2.4, and is responsible for training, open educational resources (OER), and educational programs.

Roles of the Participants

Currently, FAIRagro integrates 18 participants across a broad range of expertise, roles, responsibilities, and competencies related to agrosystem RDM (see chap. 1). Many participants, e.g., the GPW, DBG, DAFA, and DPG, continued their commitment to NFDI4Agri also to FAIRagro. New participants were asked to join the task force crop production and contribute their expertise or infrastructures either to one of the six FAIRagro use cases (TA1) or to the other task areas of FAIRagro (TA2, TA3 and TA4).

3.2 The consortium within the NFDI

Links to other NFDI consortia

FAIRagro addresses common NFDI aspects and activities for NFDI-wide cooperation are considered in all task areas. We dedicated Measure 5.3 in TA5 to coordinate and ensure our contributions with other NFDIs, NFDI sections and the directorate. We already initiated many collaborations with other consortia (see below and Measure 5.3), and members of FAIRagro started to join the NFDI association. Furthermore, Measure 5.2 will work on a long-term operational model for FAIRagro embedded in the NFDI. The FAIRagro consortium fully supports the idea of the Research Data Commons [RDC; Bierwirth *et al.* (2020); Glöckner *et al.* (2020)]

and will cross-link FAIRagro infrastructures and align the development of its services with this concept.

FAIRagro will become the NFDI node for agrosystem research data and link to the wider community of agricultural sciences. It will cross-link FAIRagro infrastructures and services in terms of the NFDI-RDCs. FAIRagro members participate in other consortia (NFDI4Earth, DataPLANT, NFDI4BioDiversity, NFDI4Health, NFDI4Objects, NFDI4DataScience, NFDI4Culture, NFDI4Memory, NFDIxCS, MaRDI, and NFDI4Chem), allowing for direct exchange of information and the development of work in consortia. FAIRagro agreed with NFDI4Earth, NFDI4BioDiversity, DataPLANT, NFDI4Microbiota, NFDI4Health, NFDI4Objects, and NFDI4DataScience on cross-consortia activities and developments to demonstrate the benefit for interdisciplinary research enabled by the NFDI.

FAIRagro will ensure NFDIs representation and visibility at European and international level for RDM in agrosystem research. FAIRagro will make a particular contribution to NFDI with potential benefits for other consortia for legal and sensitive data issues, for RDIs supporting integrated systems research across disciplines and scales, and for links to stakeholders (farmers, policy makers) for data collection management and use.

Cross-consortia activities in detail

With **NFDI4Earth**, we will work on interlinking existing help desks to facilitate cross-disciplinary support requests, which will be an added value for the NFDI. We will demonstrate exemplary networking of help desks from FAIRagro and NFDI4Earth and develop a generic concept that can be adopted by other NFDI consortia. Close cooperation is planned with NFDI4Earth's task area *2interoperate* on the alignment and mapping of semantic resources. With respect to education and training, we will integrate training options in cooperation with NFDI4BioDiversity (U Bremen Research Alliance) and NFDI4Earth (EduHub of NFDI4Earth) in the training program of FAIRagro (Measure 2.4). FAIRagro will contribute to the education of data stewards with partners from NFDI4Earth by participation in one site of the planned NFDI4Earth Academy. With **NFDI4BioDiversity**, there are some overlaps in the investigated objects, e.g., plants and fungi, and there are repositories that should be made accessible and usable by both consortia. Specifically, collaborations in the areas of object identification, phenotypic trait collection and genotyping data management have been pre-identified between both consortia. Shared membership of some partners in both consortia guarantees an efficient interaction.

With **DataPLANT**, we will collaborate in the creation and harmonization of ontologies in the plant sector and the exploitation of relevant databases for data exchange. FAIRagro will comply with the NFDI-RDC in setting up the network of interlinked RDIs (Measure 4.2) and the implementation of the SciWIn (Measure 4.4). We collaborate with NFDI4BioDiversity and DataPLANT in this regard. **NFDI4Objects** and FAIRagro already share some research topics that are covered by both initiatives, e.g., related to soil science and near-surface geophysics. Consequently, both initiatives have agreed to participate in joint meetings to ensure an exchange of experience in RDM and for mutual learning. NFDI4Objects will participate at the FAIRagro community summits as well as in the FAIRagro data stewards' network meetings.

In Measure 4.4 we will collaborate with **NFDIxCS** on the methods and implementations of container technologies to make them interoperable within the NFDI. Furthermore, it has been agreed with NFDIxCS to cooperate across disciplines on IT-related topics by inviting each other to community events, by dovetailing in the area of user help desks and by exchanging concepts and methods for implementing data protection and access to sensitive data.

FAIRagro will regularly exchange with NFDI4Objects and NFDI4Earth on how to optimise, develop and synergise community support through the helpdesk and will coordinate with NFDI4Biodiversity and NFDI4Earth with regard to curricula and training scheme development.

FAIRagro's contributions to cross-cutting topics of the NFDI

Several NFDI consortia are working on shared, cross-domain topics. In August 2019, eleven consortia met in Berlin and agreed on a common position on the importance and handling of the cross-cutting topics (Glöckner *et al.*, 2019). These ideas have been further developed, resulting in the Berlin-Leipzig-Declaration (Bierwirth *et al.*, 2020). In October 2021, the NFDI directorate formed four NFDI sections to address cross-cutting topics across the consortia boundaries. We will offer the following contributions in the cross-cutting sections and collaborate therein with other consortia, presenting an added value for the whole NFDI:

Section 1: (Meta)data, Terminologies and Provenance

- Contribute to the sharing and reuse of semantic artefacts such as controlled vocabularies, thesauri and ontologies and the creation of cross-functional mappings and alignments between these semantic artefacts;
- Contribute to the FAIR Digital Object approach (Measure 3.5) and elaborate a framework to enable both humans and machines to classify the permitted and prohibited actions over resources of FAIRagro (machine actionability). There, we will closely cooperate with NFDI4Earth's task area *2interoperate*;
- Contribute to enabling and facilitating semantic interoperability of earth observation data cubes (Giuliani *et al.*, 2019), e.g. FAIRagro's JKI Data Cube and NFDI4Earth's Climate-Biosphere Data Cube;

Section 2: Common Infrastructures

- Contribute to interlinking existing help desks to facilitate cross-disciplinary support requests for the whole NFDI; demonstrate exemplary networking of help desks from NFDI4Earth and NFDI4BioDiversity and develop a generic concept that can be adopted by other NFDI consortia;
- Develop the concept of SciWIn (Measure 4.4) jointly together under the umbrella of the NFDI-RDC and collaborate with NFDI4BioDiversity and DataPLANT as well as the prototypic container deployment to the de.NBI cloud node at BLU;

- Implement NFDI recommendations for an AAI system to integrate the FAIRagro infrastructures into the NFDI but also into international networks;
- Contribute to the development of performance indicators to measure success and impact of RDM services; develop joint concepts for service and product performance with respect to the FAIR principles that are developed;

Section 3: Training and Education

- Support the exchange of data stewards from different disciplines to advance the networking of data stewards from different consortia (e.g. DataPLANT, NFDI4Earth or NFDI4Health), facilitate experience exchange and derive common guidelines and a cross-NFDI strategy for a networked data steward support structure (Measure 2.5);
- Contribute to the education of data stewards with partners from NFDI4Earth by participating at events of the planned NFDI4Earth Academy (Measure 2.5).
- Engage with the NFDI Directorate to establish a unified community support structure of the NFDI - including the aspects of a data steward network and interlinked help desks - as a relevant cross-cutting topic to be dealt with in a further NFDI section;
- Provide a training module “agrosystem” as domain specific content to generic material created by the section (Measure 2.4);

Section 4: Ethical, Legal & Social Aspects

- Develop guidelines addressing data licensing issues and data policies in relation to agricultural data (Measure 3.6);
- Propose legal metadata standards for improved reusability and implement them with other NFDI consortia for fostering interdisciplinary research (Measure 3.6);
- Contribute legal expertise for best practices in handling sensitive data concerning business secrets and interests, together with NFDI4Chem, NFDI-Matwerk or NFDI4Ing (Measure 3.6);
- Contribute legal expertise for best practices in handling data with restricted access to public authorities, together with, for example, NFDI4Earth, NFDI4Biodiversity or KonsortSWD (Measure 3.6);
- Initiate and figure out legal standards for international research projects with partners from other consortia and the NFDI Directorate (Measure 3.6);
- Discuss convenient channels and formats for transporting legal information to RDM staff, to researchers and research institutions with the vision of data rights literacy, together with partners engaged in training and education (Measure 3.6);
- Campaign the requirements for a FAIR RDM in ongoing regulations on national and EU levels such as Data Act in the EU, and further reform packages in copyright and data protection law (Measure 3.6);
- Implement technical solutions to grant access to restricted and sensitive data (Measures 3.6 and 4.2).

3.3 International networking

There are a number of international initiatives and various networks in which consortium partners have a leading role or are active. FAIRagro will internationally promote best practices and solutions coming from FAIRagro using its networks. Furthermore, FAIRagro benefits from these connections by establishing accepted standards and the implementation of international directives. We dedicated Measure 5.3 to ensure and promote the international visibility and awareness of FAIRagro and to support the development of novel approaches for RDM by connecting and actively participating in international initiatives. FAIRagro will provide a network of international contact points within the NFDI to support the international positioning of agrosystem-related RDM interests.

Research Data Alliance (RDA): Many partners of FAIRagro are active in various RDA working and interest groups, e.g., Data Citation WG, Biodiversity Data Integration IG and Interest Group Agricultural Data (IGAD). The key results of RDA groups and Global Open Data for Agriculture and Nutrition (GODAN) will be considered and implemented in consortia with a very strong mutual interest in the respective developments.

European Open Science Cloud (EOSC): FAIRagro is already a stakeholder of the EOSC, with the most prominent activity being the building of an RDI and an umbrella for the activities in the different research communities and EU member states. Different members of FAIRagro participate in EOSC workgroups, e.g., in SEMAF (Broeder *et al.*, 2021). FAIRagro will contribute to the development of FAIR Digital Objects (Measure 3.5). We will ensure the interoperability of the FAIRagro services with the functionalities of the EOSC. With our activities, FAIRagro will ensure the information flow and coordination of activities between the agrosystem domain in Germany and Europe.

GAIA-X: In 2020, ZALF, together with Thünen, assumed the leadership of the working group Science/Economy - Data Exchange for Research Purposes of the GAIA-X domain Agriculture, ensuring close cooperation with GAIA-X on the European level and identifying possible ways to align FAIRagro and GAIA-X services.

International modelling networks: ZALF and Thünen Institut (along with individual researchers from the FAIRagro consortium) have been leading partners in the European ***MACSUR*** Knowledge Hub. FAIRagro will build on this network and connect the data provision and preparation with the user-side demands from the modelling community. With MACSUR SciPol (coordinated by ZALF), FAIRagro will have the opportunity to directly implement and validate a workflow on direct access data and automatic processing of data to be directly used in simulation models. With the strong connections of the spokesperson Frank Ewert and co-spokesperson Senthold Asseng to ***AgMIP*** FAIRagro will build on the experience of the international agricultural modelling community and on standards such as ***IBSNAT*** (Uehara and Tsuji, 1998) and ***SEAMLESS*** (Ewert *et al.*, 2009; Janssen *et al.*, 2009a; Van Ittersum *et al.*, 2008). Presently, links are also being established to ***GLASSNET*** (GlassNet, 2021) and

FAIRagro partners are also engaged and will advance links to ***ISMC*** (ISMC, 2021). These connections to networks and communities from data users will help FAIRagro develop its services along the demands (and keeping the different standards in mind) of the international modelling community, as well as general research interests. The Thünen Institute leads the ***AGMEMOD*** consortium (AGMEMOD, 2021), in which one of the key activities is the collection, harmonization, compilation and distribution of the AGMEMOD datasets across different EU modelling teams.

Agrosystem networks: FAIRagro partners cooperate closely with ***GLTEN*** (GLTEN, 2021) and IOSDV. These networks will support our ambitions of harmonizing the publication of long-term experiments and will be used to implement our guidelines and recommendations on an international level.

Networks in plant sciences: Partners of FAIRagro play an active role in steering committees for data standardization and exchange initiatives, e.g., ***Breeding API (BrAPI)***, the *Global Biodiversity Information Facility* ***(GBIF)*** and ***MIAPPE***, and will consequently foster the integration of international standards into FAIRagro's RDM and policies. ***ELIXIR*** is the pan-European infrastructure for biological information supporting research in life science and its transfer to the environment, bioindustries and society. FAIRagro partners BLU, IPK, and FZJ are members of ***ELIXIR-DE***. We will coordinate existing infrastructure efforts such as ***ELIXIR-AAI*** to align with our services.

Networks in soil sciences: Partners of FAIRagro are members of the EJP Soil Programme that have positions and hold positions on the Ethics Board and are National Communication Representative and Programme Manager, IUSS and EGU Division ***Soil Science Systems*** *(SSS)*. Thünen advises the BMEL on the establishment of a soil database and on the ***FAO Global/European Soil Partnership***, within which it is involved in various working groups. This arrangement ensures close connections among soil research projects on an international level and the promotion of the development of FAIRagro in these communities.

FAIRagro members are furthermore active in international initiatives and efforts, such as ***CODATA, OGC, DataCite, IPPN, EMPHASIS-PREP, ENVRI-FAIR, TDWG, OBO or IPBES***. KTBL and ZALF belong to the editorial board of the internationally accepted standard ***AGROVOC*** Multilingual Thesaurus (FAO). With this broad and international spectrum of network activities and involvement in different communities FAIRagro is in an excellent position to learn from existing initiatives and endeavours, and to contribute to the international science and research community. FAIRagro will use this good position to connect communities and to start the exchange of ideas and solutions by hosting workshops, presenting results in these networks and inviting international colleagues to participate in the discussions and work of FAIRagro (Measure 5.3).

3.4 Organisational structure and viability

The organizational structure and governance of FAIRagro (Figure 8) was designed to ensure that tasks can be realized in an efficient but also agile manner. With this structure, we will be able to initially identify, respond and continuously adjust to the needs of the agrosystem community. For the further implementation and communication of the governance structure, we will benefit from activities such as workshops in our community and the initial FAIRagro Portal (FAIRagro, 2021).

Governance and Decision Making

The **FAIRagro spokesperson** is the overall leader of the consortium and represents the consortium internally and externally. He chairs the Steering Committee (SC) and the FAIRagro Plenary. He represents the interests of FAIRagro in the NFDI, coordinates and establishes a regular exchange between the NFDI bodies (Directorate, Scientific Senate, and Consortia Assembly) and initiates cooperation with other consortia to transfer experiences and results from the agrosystem community into the overall NFDI context, and vice versa. If required, the spokesperson will be deputized by the project administrative coordinator (see FAIRagro Secretariat).

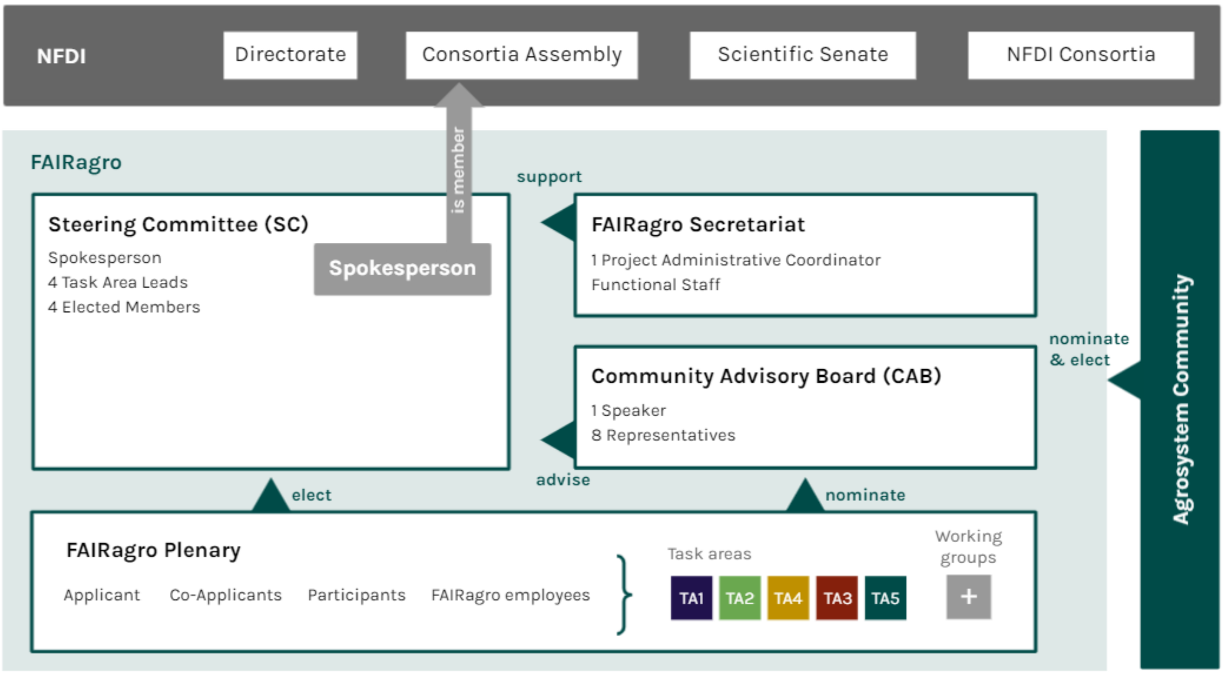

Figure 8: Governance and organizational structure of FAIRagro.

The **Steering Committee (SC)** consists of the spokesperson (TA5 lead), four task area leads (TA1 - TA4) and four elected members of the FAIRagro Plenary, one representing the groups of co-applicants, two representing participants and one representing young researchers. The SC is responsible for the implementation of the general strategy and decisions into the task areas and supervises the progress of the project, supported by the FAIRagro Secretariat. The SC is further responsible for decisions regarding scientific research directions, adaptation of strategies, budget, e.g., (re-)allocations of the funding and allocation of the centrally managed funds (see Financial Management below), organizational structure, e.g., integration of new participants into FAIRagro, establishment of FAIRagro working groups (see below) and onboarding of new use cases. Based on contributions from the task areas and supported by the FAIRagro Secretariat, the SC prepares periodic reports, e.g., the consortia progress reports and the final report. The SC can nominate members for the Community Advisory Board. The SC will hold monthly meetings (video/telephone conferences) and meet in face-to-face meetings once a year. Decisions are made by majority voting, including proxy voting. In case of a tie, the spokesperson has the tie breaking vote. In the case of decisions violating DFG (funding) rules, he will postpone the decision and consult DFG.

The **Community Advisory Board (CAB)** consists of one speaker and eight community representatives, all external to FAIRagro, who are renowned scientists and data experts covering the wide range of the agrosystem community representing different national and international perspectives and stakeholders, e.g., young and senior researchers, universities, infrastructure providers, farms and agricultural cooperatives, state and public interest groups. The composition of the CAB and its interaction with the disciplines, will be described in the CAB guideline (Measure 2.2). The CAB will be headed by a speaker and will provide strategic input on an annual basis. The CAB reviews the status of FAIRagro, formulates recommendations to the SC about use case proposals, the development and strategic operation of FAIRagro, and supports international interactions with relevant infrastructures and initiatives. The CAB candidates will be nominated by the FAIRagro Plenary and by the wider agrosystem community at the annual community summit (Measure 2.1). The nominees are elected by the community representatives attending the community summit (Measure 2.2) for terms of two years.

The **FAIRagro Plenary** encompasses representatives of all participating institutions of the FAIRagro consortium, including the spokesperson, all co-spokespersons, participants and all FAIRagro-employees. The plenary will meet once a year in a meeting organized by the FAIRagro Secretariat (TA5). At this meeting, the project status and progress will be presented and discussed. Workshops complementing the annual plenary meeting can be organized by individual TAs to facilitate TA and measure-overarching networking, technical discussion or presentation of research results. Four consortium representatives are nominated and elected (re-election is possible) for terms of two years from the plenary to represent the broad interests of the consortium in the SC, representing the group of co-applicants (one), the group of participants (two), and the group of young scientists (one) in FAIRagro. Furthermore, the FAIRagro Plenary can nominate members for the CAB.

The **FAIRagro Secretariat** (TA5), located at ZALF, supports the work of the spokesperson and the work of all of the abovementioned boards. The secretariat is responsible for the project and administrative management of the consortium. The *project administrative coordinator* is the spokesperson's deputy and heading the FAIRagro Secretariat. The *project operational manager* is responsible for the day-to-day management of FAIRagro, including tasks that monitor the project's progress (milestones and deliverable monitoring), reporting and DFG/NFDI administrative tasks. He/she will perform and document risk assessments, assisted by the task area leads. The *administrative manager* is responsible (1) for the creation and fulfilment of the cooperation agreement with the partners and for financial management and financial control, including the monitoring of the proper use of funds, and (2) for coordinating the transformation of FAIRagro into a sustainable association. One office assistant will support the general daily business of the secretariat and the organization of Plenary and SC meetings.

In addition to this organizational structure and the TAs, we will implement the concept of **FAIRagro working groups (WGs)** that focus on cross-cutting topics and specific tasks with a common interest for FAIRagro. New working groups can be proposed by the FAIRagro Plenary to the SC, which will decide on the adoption of a WG based on feasibility and relevance. WGs will be supported by the FAIRagro Secretariat to facilitate their work and ensure proper communication.

We will complement this governance model with the **FAIRagro consortium agreement (CA)** (Measure 5.1), which will include the organizational structure and associated regulations and their compliance with DFG regulations. The organizational model will be regularly reviewed and adapted to changing requirements, e.g., for the long-term establishment and operation of FAIRagro and its embedding in the NFDI.

Financial Management of FAIRagro Project Funds

The FAIRagro Secretariat at ZALF will manage the finances centrally in accordance with the DFG rules. Co-applicants and leading participants with an active contribution to specific measures and use cases (see FAIRagro work program in chap. 5 and funding tables in chap. 6) will receive funding through respective contracts from the beginning. FAIRagro is an open consortium, and it is actively planned to grow during the funding period. Therefore, the consortium agreed to dedicate a share of 15% of the overall funding to centrally managed flexible funds, managed in TA2, to ensure the integration of additional partners with new ideas and contributions. These funds will mainly be used for the onboarding of new use cases (see chap. 3.1) to allow the integration of new competences of the agrosystem domain and to extend to other domains of the agricultural sciences, but also for adjustments to new developments and community needs. The onboarding process is described above (see chap 3.1) and in the corresponding measure (Measure 2.3).

Risk assessment, contingency approach

The FAIRagro project faces several risks related to its long-term success and envisioned task. These risks and related mitigation approaches are listed below:

- **Problems and delays in hiring staff.** FAIRagro requires a fast startup to fulfil the envisioned tasks, and we require highly qualified and motivated staff to initiate the formation of core elements, e.g., the Data Steward Service Center in TA2. A general overarching risk is the recruitment and retention of qualified staff with adequate skills. Delays that potentially arise from this approach can potentially be addressed through recruiting via partners from the university sector, through cooperation with the NFDI initiative for joint recruitment and by (co-)application(s) and participants' own contributions. We will benefit from the training and education measures in TA2 and their close links to potential new staff.

- **Insufficient capacity of data stewards to handle community support requests.** Depending on the acceptance of service provided by the data stewards to the agrosystem community, there is a risk of the data stewards being swamped by too many support requests that cannot all be processed. As we designed the data steward network to be scalable and coordinated by the Data Steward Services Center (Measure 2.5), new data stewards can be easily integrated into this support network of FAIRagro and better balance the support requests. In fact, FAIRagro has already extended the number of data stewards to five as compared to three in earlier NFDI4Agri proposal also following reviewers' comments.
- **Lack of community acceptance:** FAIRagro and every other NFDI consortium rely on cultural changes in their community toward RDM. All consortia must ensure that they address their community needs and appeal to their communities. FAIRagro strives to foster this cultural change by active interaction with the community via new use cases, workshop and training activities and continuous networking with agrosystem scientific societies. Over the last years of preparing for FAIRagro (and NFDI4Agri) the number of workshops and community activities have steadily increased.

  The funding of participating RDIs that are discontinued. There is a risk if (co-)applicants(s) or participants cannot continue to operate core RDIs relevant for FAIRagro as specified in the inkind contributions. Approaches, such as the transfer of operation to the consortium or the early migration to alternative, publicly funded RDIs (Zielinski *et al.*, 2019), are defined in an exit strategy as part of the operating model described in chap. 3.5.
- **Technical limitations can delay the implementation of the FAIRagro network of federated infrastructures.** Potentially, insufficient quality or availability of technical documentation on an RDI can delay the implementation of interfaces to partner RDIs in Measure 4.2 and its access in Measure 4.3. Ideally, we will follow a Brownfield approach (Axehill *et al.*, 2021) and attempt to update every RDI to improve their interoperability and allow integration into the middleware. If this step is not possible, it will be necessary to follow a Greenfield approach (Axehill *et al.*, 2021) by extracting the data and metadata and establishing a new, more comprehensive version of the infrastructure.
- **Need to increase the computing service capacities**. For our data integration services, e.g., operated on the SciWIn platform in Measure 4.4, there could potentially arise the need to increase the server capacity of the RDIs in line with the increase in required data access bandwidths, i.e., compute and network capacity of the RDI hosts, which cannot

yet be fully anticipated. These risks can potentially be addressed by increasing the inkind contribution of the respective institutions or reallocating the budget for infrastructure adjustments.

FAIRagro is well aware of the possible problems that accompany such an ambitious endeavour. In addition to the above mentioned risks, we are reckoning on many yet unforeseen problems that could arise in the process. The large assets of FAIRagro in developing solutions in case of upcoming challenges are the international experienced members in FAIRagro, the strong connections to large research institutes in the domain and the well-established communication channels to other NFDI initiatives.

Conflict management

If problems arise related to undesirable developments in the work program or in achieving the specific objectives, the task area leads should inform the spokesperson. Supported by the SC and the FAIRagro Secretariat, the spokesperson develops and proposes mitigation strategies that must be endorsed by the FAIRagro SC and that can lead to changes in the work program. In case of conflict situations in a task area, the respective task area leads should report them to the spokesperson. Major conflicts will be addressed by the SC. The spokesperson supported by the FAIRagro secretariat will collect the positions of the conflicting parties and report them to the SC. The SC will be responsible for resolving conflict situations reported to them with the possibility of advice by the CAB. The members of the SC vote on positions that pertain to the conflict. In the case of a tie, the vote of the spokesperson decides the matter. If conflicts cannot be resolved by the SC and the CAB, external neutral bodies such as the central ombudsman committee of the Leibniz Association or the NFDI Directorate will be contacted.

3.5 Operating model

The FAIRagro consortium fully adheres to the FAIR principles extended by free open data and the free open code principle. We follow a no-user-fee policy for the provided FAIRagro services.

Contributions and long-term commitments

FAIRagro builds on a federated infrastructure of disciplinary repositories provided by (co-)applicants and participants (see chap 4.4). The (co-)applicants and participants commit (a) their institutional infrastructures (see chap. 6) in the long-term and (b) to the long-term provision of central FAIRagro services and infrastructures. ZALF will operate the project collaboration platform with an integrated helpdesk and customer-relationship-management system (CRM), one RDMO (research data management organizer) instance and a central platform for source code development. JKI will host the FAIRagro Portal. ZB MED will host the central search service. A SciWIn pilot will be rolled out at de.NBI operated by Bielefeld University (BLU). Additionally, all software that will be developed by members of FAIRagro will be published as open source and made available on appropriate platforms, e.g., the central project management platform, including the source code development environment platform hosted by ZALF. Guidelines and training materials for the consortium and community will be developed, and thourough training courses and education programs will be established. In addition to scientific publications, FAIRagro will publish all information in public repositories with persistent identifiers to ensure the long-term availability of those resources for NFDI. We commit to research software sustainability, by publishing developed software in FAIRagro as open source on suitable platforms in agreement with the NFDI open source strategy. Furthermore, the applicant and co-applicants will provide substantial support in terms of inkind staff contributions, the contribution of facilities and additional expertise and support services (see chap. 6.4).

For the success of FAIRagro, it is important to ensure a sustainable provision of the distributed and available RDI. To address sustainability within FAIRagro, FAIRagro will transform into a non-profit association (within the NFDI e.V. and in consultation with FAIRagro members and the community), providing the basic service of long-term availability of data. Support from the BMEL has been indicated. The establishment of the association which is independent from the current FAIRagro partner institutions is part of FAIRagro's risk management strategy. To prevent lost and unrecoverable service infrastructure and to prepare for the unlikely case of transition into a a different NFDI structure the association will serve as a platform to enable service transition or data migration in part or as a whole to alternative service hosts or repositories. Following exemplar cases of early migration to alternative, publicly funded RDIs (Zielinski *et al.*, 2019), FAIRagro will develop a migration plan with alternative service hosting bodies or data repositories (Measure 5.2). Furthermore, for overlapping domains, we will closely collaborate with other NFDI consortia to develop data migration paths. We will actively develop and contribute to a cross-NFDI approach to exit strategies at the repository and service levels. As a second measure to sustain the achievements of FAIRagro, we will explore different possibilities to sustain the services developed and established beyond the funding phase (Measure 5.2).

4 Research Data Management Strategy

4.1 State of the art and needs analysis

Status quo of RDM in agrosystem science

Agrosystem research aims to understand and develop sustainable crop production systems reducing the impacts on ecosystems and the environment and to develop related management interventions considering relationships across scales. It requires a systems approach integrating effects of environmental factors (climate, soil), management activities and crop genetic characteristics on crop growth and yield, yield quality and stability, natural resource use and resource use efficiency, agro-biodiversity and agroecosystem and landscape sustainability for the range of relevant crops and cropping systems in a specific region considering different levels of organisation from genetics to farmers fields to regions (and eventually to globe; see chap. 2.1). To address the key challenges of sustainable development, such as sufficient and secure food production, resilience to climate change, reduced environmental impact, and improved biodiversity and ecosystem and nature protection, agrosystem research and related approaches such as sustainable intensification and agroecology urgently need to develop evidence-based solutions (Hertel *et al.*, 2021; Wezel *et al.*, 2020; HLPE, 2019; FAO, 2018; Pretty, 2018; Foley *et al.*, 2011)

Example 1: The **breeding** of new, adapted crop varieties is an integral part of productive and stable agricultural production (Reynolds *et al.*, 2021) particularly under climate change and increasing climate variability with extreme heat and drought events. Therefore current breeding practices require genotype and phenotype data to identify resistant and yield stable varieties. However, phenotype data are often lacking or are poorly annotated, especially from field experiments. The selection process, on which breeding of new varieties is based, currently contemplates solely genotype or phenotype data. Environmental and management information are often not considered due to a lack of available data. This results in a gap between selection success in experimental environments and on farms. The integration of existing data into a FAIR RDI is therefore an important basis for innovations, which are urgently needed to further increase the performance of plant breeding (Tian *et al.*, 2021; Watt *et al.*, 2020).

Example 2: Research on **nitrogen dynamics and carbon sequestration** in soils (Minasny *et al.*, 2017) as a building block of humankind's response to climate change. In Europe's new research program Horizon Europe, an entire mission will be dedicated to soils, i.e. the EU Mission: A Soil Deal for Europe. The change in carbon content in the soil is a long-lasting process that can be measured over decades. Time series and legacy data from agricultural long-term experiments (LTE) between 20 and 180 years, are relevant to address such research questions. In Europe alone, 435 such LTEs have been identified by October 2021 (BonaRes, 2021c). However, research data on LTE are currently difficult to find, with little to no access and lack standardized descriptions. Often, information about and data from LTEs are only found in scientific publications, and important links to respective soil, climate or environmental data are missing.

Example 3: Integrated **pest management** (IPM) is a central part of integrated crop management. IPM aims to minimize the use of pesticides and their harmful effects on the environment (Barzman *et al.*, 2015). Despite increasing efforts from policy, science and extension regarding the promotion of IPM, its resounding success has thus far not been achieved (Hossard *et al.*, 2014). One major reason lies in the lack of findability, standardization and integration of IPM-related data, models and respective decision support (Lamichhane *et al.*, 2015). IPM considers the reduction of yield instabilities caused by **pests and diseases** (P&D). P&D research builds on field surveys [e.g. Lechenet *et al.* (2017); Hossard *et al.* (2014)] and experiments [e.g., Feike *et al.* (2021); Schwarz *et al.* (2018)]. Integrating these separate data sources is crucial for comprehensive and evidence-based IPM-related research. However, there are considerable challenges presently, including variations in experimental design (e.g., regarding control treatments) and disease assessment procedures (i.e., timing, scale, sample size), which hinder comparability of results (Willocquet *et al.*, 2021). The interplay of continuous crop genetic adaptation, agronomic management changes, climatic changes, landscape level effects and P&D evolution is highly complex and requires a solid data basis that can be utilized effectively through data integration, analysis and modelling by the research community.

Example 4: The **modelling** of complex effects of climate and environment and management interventions on agrosystems considering interactions among processes in agrosystems and the impact of agricultural production on climate, environment and biodiversity. Advanced modelling approaches are necessary to understand and describe relevant agrosystem processes. Today's modelling approaches are multidisciplinary and cross-scale, integrating soil, weather, agricultural management and increasingly also ecological data, and market or farm economic data. However, model simulations are often highly uncertain due to the lack of experimental data for model development, parameterization and validation. While many experiments have quantified the response of relevant parameters (e.g., contents of soil organic matter or plant nitrogen) to soil types and agricultural management (Denef *et al.*, 2007; Pulleman *et al.*, 2005), the data is frequently neither findable, nor available in standardized form for model use. Agronomic data are required for model applications, but essential soil, weather, management, market or farm economic data are often sporadically published and difficult to find. Efforts are underway to analyse and categorize data for use in agrosystem modelling (Kersebaum *et al.*, 2015). Furthermore, valuable research data from LTE are needed, e.g., for model calibration and validation, but not published FAIR (see above). Through increasingly widespread international model comparisons (AGMIP, 2021a; MACSUR, 2021), a contribution is made to bring the agrosystem's modelling community closer together and to improve modelling and data standards (Rosenzweig *et al.*, 2018; Ewert *et al.*, 2015; Asseng *et al.*, 2014; Asseng *et al.*, 2013; Rosenzweig *et al.*, 2013). An integrative RDM between experimentalists and modellers would increase the reusability of data.

Main RDM challenges in agrosystem research

Collaborative, interdisciplinary and sustainable agrosystem research data analytics and reuse as requested by the scientific community and envisioned by FAIRagro is so far not possible. This is because of the different maturity levels of the involved disciplines, the use of heterogeneous RDM standards (see chap. 4.2), and the insufficient alignment of the FAIR principles to the RDM life cycle (see chap. 4.3). The publication of standardized and harmonized agrosystem research data, described with rich metadata, including data quality and provenance information, is still lacking. Most domain-related journals do not yet require authors to publish research data in FAIR repositories.

Hence, publishing data is currently hampered by

(1) Incomplete knowledge about data publication and available repositories

(2) Uncertainty about the legal implications of data publication and reuse of sensitive data,

while reuse is constrained by

(3) High diversity of data formats, scales, resolution and units

(4) Lack of information about data quality and plausibility.

(1) Available research data repositories in the agrosystem research domain: The use of domain-specific infrastructures (e.g., the BonaRes Repository for soil sciences (BonaRes Repository, 2021; Grosse *et al.*, 2020b) or institutional repositories [e.g., GBIS/I, Oppermann *et al.* (2015)], e!DAL-PGP (Arend *et al.*, 2016) as well as general publication services OpenAgrar (OpenAgrar, 2021; Oeltjen *et al.*, 2019), and the PUBLISSO Repository for Life Science (PUBLISSO, 2021) have become more common in the last decade (ELIXIR, 2021; Nature Scientific Data, 2021; Rigden and Fernandez, 2019; He and Han, 2017). Usually these RDIs support a specific metadata schema (see chap 4.2) but lack interoperability with respect to information exchange or machine-readable access to research data, which hampers interdisciplinary data analyses. RESTful services (Garriga *et al.*, 2016) are commonly used but implement heterogeneous data structures, endpoint authentication and authorization schemes, rendering the subsequent use of the data considerably more difficult.

(2) Sensitive data: Agrosystem research data are also frequently collected on private farms, and are often linked to personal data or property- or business-related sensitive data, which raises questions regarding legal aspects and data protection. Additionally, the ownership (data rights) of research data is often unclear, posing a major challenge in making these data publicly available and reusable (Hartmann, 2021). Furthermore, sensitive data often have inconsistent legal rules for reuse and specialized technologies are required to encode sensitive or restricted data in repositories (Hallinan, 2020a). These are to provide open-access information on their existence on the one hand but also restricting their access on the other hand (Hallinan, 2020b). A secure data environment allowing for the (semi-)automated authentication of data users requesting such data as well as subsequent user-specific data-release authorization and licensing is still missing.

(3) Data types and formats: RDM in agrosystem sciences faces considerable challenges as the broad and heterogeneous landscape of scientific disciplines included as well as handling the huge variety of data types involved represent a major and ongoing task. The spectrum of data considered ranges from temporal and spatial data as well as from field book records of field experiments to large-scale calculations made by model ensembles on high-performance computers (Venkatesan *et al.*, 2018). Moreover, research data can originate from manifold sources, such as from landscape monitoring, field and laboratory experiments, sequencing, phenotyping and spectrophotometric measurements, soil mappings, near and far remote sensings, economic census data and model simulations.

(4) Data quality (DQ) and plausibility: The quality of published research data often cannot be assessed by reusers. This concerns not only legacy data with no or poor documentation (Möller *et al.*, 2012; Carré *et al.*, 2007) but also sensor data including spatial information with mobile location-aware technologies, for which no standardized DQ protocols exist so far (Mehdipoor *et al.*, 2015). Data reusers have to perform extensive content tests [e.g., plausibility and consistency] to be able to determine the suitability along their research question. Furthermore, in a data-rich, network-based and heterogeneous environment, DQ standards and descriptions of the history of data products or web resources [provenance; Di *et al.* (2013)] are of high relevance but are currently missing. Standards and tools for context-based curation and the visibility of DQ are needed, which refer to both methodological aspects of data acquisition and production, as well as to additional information about data suitability and fitness for further use (German Council for Scientific Information Infrastructures (RfII), 2020). Legacy data, e.g., from agricultural long-term experiments, measurements and observations [e.g., Kaspar *et al.* (2015)] or from more recent mappings often come with no or varying indication of DQ (Senaratne *et al.*, 2016; Mehdipoor *et al.*, 2015; Carré *et al.*, 2007). Although the importance of DQ is well known [e.g., German Council for Scientific Information Infrastructures (RfII) (2020)], only a few infrastructures in the agrosystem domain have implemented structured or standardized DQ-assurance [e.g. IBSNAT/ICASA (White *et al.*, 2013; Uehara and Tsuji, 1998)]. The International Organization for Standardization (ISO) published the norm ISO 19157 (2013), which helps data producers objectively describe and determine the quality of geographic information using statistics rather than subjective judgement.

Conclusion: FAIRagro will address these challenges has set several objectives (see chap. 2.2) to substantially improve this situation through a coordinated effort by supporting and guiding the community towards FAIR RDM, initiating cultural change, and providing technical solutions to facilitate RDM in agrosystem sciences.

Needs analysis of the agrosystem research community

The RDM challenges in agrosystem research have also been perceived and communicated by the community on many occasions, such as the NFDI4Agri Community Workshop in 2020 with >120 participants (see chap 3.1, Figure 4). The NFDI4Agri Community Survey in 2020 (Senft,

2021) likewise confirms this situation. The survey was performed to identify the needs and challenges of the agricultural research community and included a total of 54 questions on the background of the respondents, their RDM practices and their RDM needs. In total, we received 196 (partially) completed questionnaires from data providers, data users, and infrastructure and information service providers. According to the survey, only 13-26% of data providers indicated that they publish data or information about data on a regular basis. Reasons for not publishing included concerns related to data privacy (62%), business interests (49%), and research ethics (15%). In particular, 92% of data providers who do not make their data freely available estimated that these data have a high to medium reuse potential. When data are published, only approximately 50% of data providers add metadata and 75% do not know which metadata standard they use. The situation is similar for the use of controlled vocabularies and ontologies. A total of 85-97% of the answers indicate that the respondents do not know common ontologies in agricultural sciences. Most respondents regarded an overview of all domain-specific services and associated metrics on quality, availability, and data rather positively. Overall, the current situation is characterized by uncertainty and lack of knowledge regarding the publication of agricultural research data, the resulting legal implications and an impeded findability and reusability of these data, thus contradicting the FAIR data principles.

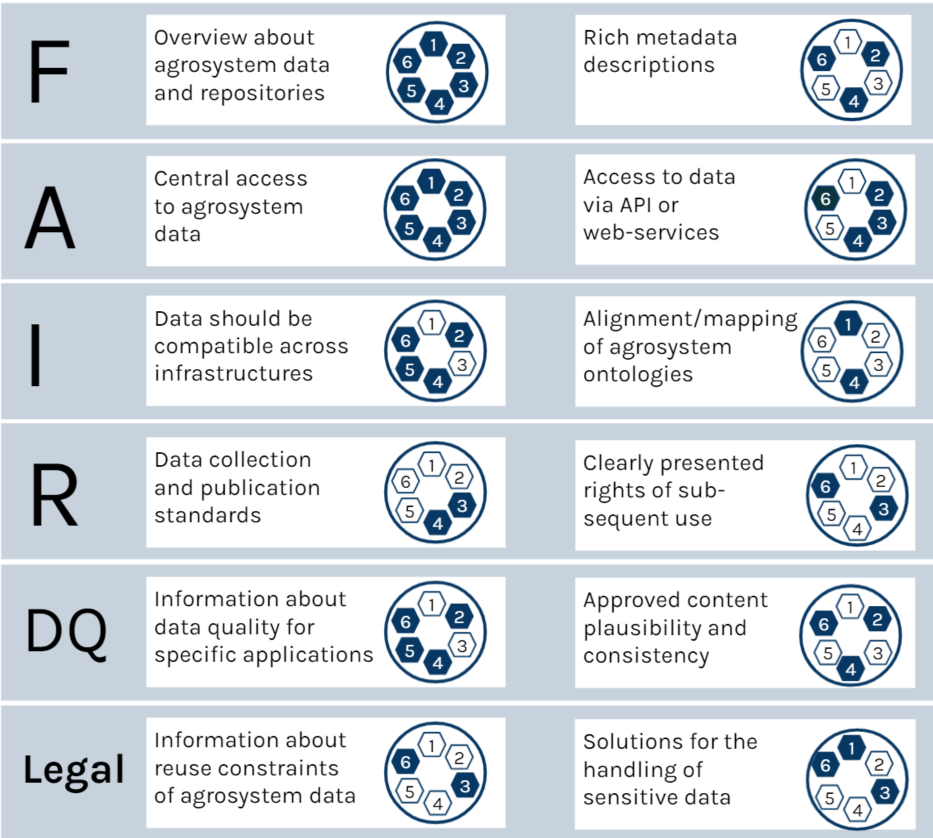

Figure 9: RDM challenges in agrosystem research as described by representatives of the use cases with respect to FAIR (Findable, Accessible, Interoperable and Reusable), data quality (DQ) and legal challenges. Numbers in circles refer to flagship use cases (see chap 5.1, TA1) and blue hexagons represent the use cases which expressed the respective challenge.

For the preparation of the FAIRagro proposal, we initiated new activities with the agrosystem community (see chap. 3.1, Figure 5) with the objective of identifying the community's specific data needs. With the establishment of the task force involving researchers and data analysis experts from the agrosystem domain, we identified the main RDM-related challenges of agrosystem research, which resulted in the formulation of six flagship use cases. These use cases were discussed and further shaped in several workshops with broad participation of the community. They all address relevant RDM challenges (Figure 9) for which researchers and data scientists will jointly develop solutions.

RDM strategy of the FAIRagro consortium

FAIRagro aims to support scientists in all steps of the RDM data life cycle, e.g. with the Data Steward Service Center (Measure 2.5) and its wider network, facilitating data workflow and capture of metadata. In a community-driven process, FAIRagro will link existing RDM services in the agrosystem sciences into a federated network that will be embedded into the NFDI also considering international initiatives. These efforts will be aligned with community needs as addressed in the use cases (TA1), supplemented by the consolidation and application of existing standards and policies, ensuring a mature level of implementing FAIR and data quality as an integral part of the RDM life cycle. Comprehensive community outreach activities will ensure an increased and sustainable awareness of RDM for scientists, fostering cultural change. Due to the quickly changing nature of research methodologies, core elements of the proposed RDM strategy must be adaptive and constantly evolving. The FAIRagro RDM strategy consists of the following eight components:

(1) A network of **data stewards to provide on-demand support** and consultation for scientists during all data life cycles in RDM. This includes assistance with research data annotation and processes, as well as the documentation of requirements and RDM needs in the community. The provision of dedicated personnel is one of the main pillars of our RDM strategy supplementing the technical services and standardization processes of the consortium (TA2).

(2) A **platform for knowledge exchange** - the FAIRagro Portal**,** as a gateway to engage with the research community and as a central access point to the knowledge hub and technical services provided by the consortium. It will bring together user support and helpdesk, education and learning materials, guidelines and documentation, federated data infrastructures and computing services provided by FAIRagro and the community (TA2).

(3) A comprehensive **training and education program** for the dissemination of best practices and guidelines for RDM, individually tailored for separate target groups, e.g., students or senior researchers (TA2).

(4) Coordination of a user-need driven **standardization process** for metadata and data quality annotation to develop standards for all annotation, quality and curation efforts within FAIRagro (TA3). We will not develop new standards, but rather build on existing internationally recognized standards (see chap. 4.2 and 4.3) and consolidate and harmonize them.

(5) Development of **legal guidelines, policies, model agreements and legal metadata standards to facilitate reusability of sensitive data**. Provision of a legal framework for agrosystem data to reduce uncertainty in publishing and reusing sensitive data (TA3).

(6) **Facilitate findability, access and data analytics** across disciplines by interlinking available data infrastructures into a federated network of interoperable RDI under the consideration of the NFDI-RDC. Develop search functions to improve the findability of research data and data infrastructures, provide relevance ranking for search results, and provide computing infrastructure services to enable reusable data integration featuring cloud-based analysis workflows (TA4).

(7) Address challenges for the application of simulation models by **standardizing and harmonizing data** formats where applicable and supporting FAIR publication of LTE and other data, and provide services to scale and aggregate data to the required resolution (TA1).

(8) In the first funding phase, FAIRagro will provide and consolidate FAIR RDM services for the agrosystem community. We will adapt to new user requirements as formulated by additional use cases addressing further challenges in agrosystem research. At the end of the first funding phase, we strive to successively **expand the FAIRagro nucleus by bridging other disciplines and domains of the agricultural sciences towards an integrated NFDI community for agricultural sciences.**

The proposed strategy is driven by user needs including main components and measures that were identified to be crucial for FAIR data management in agrosystem disciplines, which is an important prerequisite to be accepted by the community. We will perform regular surveys within the whole community, engaging constantly with roadshows and workshops the research institutions to align our services continuously to the needs of the community. We communicate directly via our data stewards with scientists, and provide hands-on training and live demos to ensure effective knowledge transfer and awareness of new services of the FAIRagro infrastructure. We will follow the NFDI open-source strategy and commit to research software sustainability by publishing developed software in FAIRagro as open source on suitable platforms.

4.2 Metadata standards

This diversity of agrosystem disciplines and the data types processed therein are also reflected in the heterogeneity of the metadata standards used. Some of these standards are very generic (e.g. *DataCite* (DataCite Metadata Working Group, 2017), Dublin Core (2020), DCAT (2020), W3C Prov (2021), and W3C VoID (2021), while others specify metadata for particular domains. In plant sciences, the most important metadata standards comprise the Multi-Crop Passport Descriptors (MCPD), which describe the provenance of genetic resource accessions (Alercia *et al.*, 2015), and the ICASA standards (Bostick *et al.*, 2004), which were revised in 2013 (White *et al.*, 2013) and further developed into MIAPPE, (Cwiek-Kupczynska *et al.*, 2016), which is currently implemented in BrAPI (BrAPI, 2020; Selby *et al.*, 2019). The geospatial community has established a number of metadata standards in recent years, e.g., ISO 19115 Geographic information — Metadata, which has been extended by the European INSPIRE directive and provides additional implementation rules (European Commission Joint Research Centre, 2007), or the BonaRes Metadata Schema (Specka *et al.*, 2019).

Provenance information, such as detailed descriptions of thetype and origin of data, is essential to evaluate data quality and to improvise reproducibility, interpretability and fitness-for-use evaluation of research data. Observational data require a description of the sensors used [e.g., calibration information, accuracy, units of measure]. For data from simulation models, a detailed description of the input datasets, parameter settings and scenario descriptions are required using common vocabularies. However, such provenance metadata are usually not available in metadata or only available as textual descriptions. **The use of common vocabularies, persistent identifiers for data, contributors, institutions, instrumentation and sensors are required to establish provenance information.**

Analogous to metadata standards, there are a large number of semantic resources that are commonly used in agrosystem sciences, e.g., AGROVOC (FAO, 2021), Agrontology of the FAO (FAO, 2019) and SEAMLESS-IF ontology (Janssen *et al.*, 2009b), general ontologies of the OBO Foundry (Agronomy Ontology, Environment Ontology, Plant Ontology, Plant Trait Ontology, Experimental Conditions Ontology, Crop Ontology (Plant Ontology, 2019; Arnaud *et al.*, 2012; Shrestha *et al.*, 2012; Jalswal *et al.*, 2005)], and genomics and proteomics ontologies such as the Gene Ontology or Protein Ontology (Harris *et al.*, 2004).

FAIRagro strategy on metadata standards

Our main goal is to develop common metadata standard profiles and principles for agrosystem sciences by consolidating and harmonizing currently used standards. For a basic level of comparability of research data and interoperability of RDIs, we will consolidate applied metadata standards across disciplines to improve the findability and reusability of research data and support data integration. We will build on experiences for mapping disciplinary metadata standards (Specka *et al.*, 2019) and identify the minimum metadata elements required to describe research data (TA3). We will develop an inventory of domain-specific metadata schemes, exchange protocols, controlled vocabularies, and ontologies, including an assessment concerning the FAIR principles from which minimal and FAIR metadata profiles for FAIRagro will be extracted. **TA3 (Measure 3.1) will address harmonizing metadata, vocabularies, the use of persistent identifiers and vocabularies and the structured description of provenance and data quality information and align these efforts with developments of the NFDI and international initiatives.**

Moreover, we strive to enable the integrative research and interoperability of ontology-based services. The meanings (semantics) of different data types need to be specified to integrate data from various infrastructures. Therefore, we will use existing ontologies [e.g., AGROVOC (FAO, 2021), Environment Ontology (Buttigieg *et al.*, 2016; Buttigieg *et al.*, 2013), crop phenotypes (Plant Trait Ontology (Arnaud *et al.*, 2012) or Crop Ontology (Shrestha *et al.*, 2012)] and extend them based on requirements formulated in the agrosystem community so that these standards will become an integral part of the RDM workflows in agrosystem research. Publication guidelines (Measure 3.2), trainings (Measure 2.4) and the data stewards (Measure 2.5) will support researchers in the application of standards and workflows, starting with the six flagship use cases (TA1).

To improve our digital resources' findability, we will establish **a domain-specific extension of schema.org** enabling structured data on web pages, such as for publication landing pages in repositories (Measure 3.2). As a result, research data made available within the FAIRagro network will be indexed and findable by major search engines, e.g., Google Dataset Search (Google Dataset Search, 2021) or OpenAIRE (OpenAIRE, 2021).

4.3 Implementation of the FAIR principles and data quality assurance

The FAIR principles (FORCE11, 2019; Wilkinson *et al.*, 2016) promote data to be **F**inable, **A**ccessible, **I**nteroperable and **R**eusable. FAIR data management is a key aspect addressed by the NDFI as a whole. Data and metadata standards, formats and RDIs used in agrosystem sciences exhibit varying degrees of conformance to the FAIR principles. This is due to a lack of data and metadata standards, insufficient data quality, and lack of scientific recognition for data management and low acceptance of RDM services. In FAIRagro, we address these challenges to foster acceptance by scientists of FAIR data management for the whole research data lifecycle. For validation, FAIRagro will build on the work of the FAIR metrics group (FAIR metrics, 2021) which provides templates to measure the FAIRness of digital resources (Wilkinson *et al.*, 2018), which we will apply to vocabularies, ontologies, metadata standards and data alike.

FAIRagro strategy for addressing the FAIR principles

Findable: Data in FAIRagro repositories (see chap. 4.4, Table 2) will be described with rich machine-readable metadata linked with domain-specific and cross-domain vocabularies (Measure 3.1). We will provide guidelines for the usage of persistent unique identifiers for data, contributors, or institutions [Measure 3.2, addressing F1 and F3 of FAIR principles (GO FAIR, 2021a)]. Data infrastructures within our consortium will provide DOI for datasets using services such as DataCite. By establishing and validating a domain-specific extension of schema.org (Measure 3.2), we facilitate automated indexing and registering by search engines (F4). In addition, the FAIRagro search service (see chap. 4.4 and Measure 4.3) will ensure findability of all datasets from a single-entry point.

Accessible: Resolving the URL provided by the DOI of datasets, all data will be retrievable. Data infrastructures in FAIRagro will provide standardized APIs (Measure 4.2) to facilitate access to metadata and data. Furthermore, we will implement the NFDI-AAI service (Measure 4.2) to control and restrict access to sensitive data if needed, as required by the GDPR. One of the fundamental concepts of FAIRagro is the FAIR Digital Object approach (Measure 3.5): by introducing virtual data objects (the FDO) that encapsulate persistent identifiers (PIDs), metadata detailing access and the data itself, all derived data can be linked. Based on these links, seamless unified access to all information can be provided (Islam *et al.*, 2020). We will use PIDs as long-lasting references to unambiguously identify any kind of digital object and to connect compound data objects, using different PID schemes for different entities, e.g., DOI for datasets, ROR for organizations or ORCID for persons (Hardisty *et al.*, 2021).

Interoperable: A key foundation for data integration and semantic interoperability is the use of domain standards, controlled vocabularies and ontologies. We will evaluate existing ontologies and make them interoperable by developing mapping and alignment methods (Measures 3.1 and 3.2), thereby setting up qualified references between them and the datasets using them (principle I3). We will employ Linked Data and Semantic Web standards (W3C, 2021) to express data semantics and relations and to use inference on these data.

Reusable: FAIRagro metadata will be released with a clear and accessible data usage licence supported by legal policies and guidelines (R1.1, Measure 3.6). Legal information will be consolidated to develop methods for dealing with sensitive data, such as farm data, which will be transferred into the community and made available through metadata (Measures 3.1 and 3.6). This legal information will be included in the machine-actionable FDO to enable transport to and processing by heterogeneous information systems depending on permission, prohibition, and obligation statements for content usage, which makes them widely reusable by both humans and machines (Measures 3.5 and 3.6). The provision of provenance information by the use of accepted vocabularies and ontologies will represent the focus of Measure 3.1 (R1.2). In Measure 3.5, we will create approaches for provenance annotations during the data life cycle leveraging the joint RDA/TDWG Attribution WG Metadata recommendation (Thessen *et al.*, 2019), which uses W3C's PROV data model. FAIRagro will also establish a service for integrated data analysis (Measure 4.4), lowering the barriers for scientists by allowing them to contribute, collaborate and access the research data capacity to improve transparency and efficiency. Together with the joint NFDI-RDC infrastructure, made up of NFDI4Earth and NFDI4BioDiversity (Glöckner *et al.*, 2020), several tasks will focus on the provision of FAIR data containers of harmonized and interoperable data collections and homogeneous, integrated, FAIR and programmatic access to services from FAIRagro (SciWIn, Measure 4.4).

Data quality assurance

Although the importance of data quality (DQ) is well known [e.g., German Council for Scientific Information Infrastructures (RfII) (2020)], it is not addressed by the FAIR principles (R1, 2021b). Provenance information is needed that describes the data origins, their derivation history (lineage) and qualities in a scientific workflow or web context, and the link to the responsible data provider considering privacy and data protection rights. FAIRagro will supplement the FAIR requirements with DQ characteristics (Measure 3.3), including legal aspects (Measure 3.6). We will implement objective DQ criteria across infrastructures by considering accepted standards as addressed in TA3. Under Measure 3.3, existing fitness-for-use metrics will be evaluated and adapted to the agrosystem data. With the implementation of an application-data-matrix, previous uses of data will be documented, facilitating further reuse. Best practice guidelines and algorithms for automatic data curation will be developed in Measure 3.4 and published in an online framework. The development of a data integration platform (SciWIn, Measure 4.4) will allow for the integration and analysis of heterogeneous data from distributed infrastructures and deliver an open and well-defined infrastructure in which workflows such as integration algorithms or environmental analyses can be stored, modified and shared for the transparent reusability of datasets.

4.4 Services provided by the consortium

FAIRagro will provide services comprising information and support, including guidelines, training, consulting and educational services, as well as infrastructure services, based on the community needs formulated by the use cases (Figure 10). The FAIRagro services will be embedded in an NFDI-wide service portfolio to facilitate RDM for researchers across domains. FAIRagro services are built on a flexible, scalable infrastructure of existing components provided by consortium partners.

The **FAIRagro Portal** (Measures 2.1 and 4.1) will become the central entry point for all FAIRagro services. It will provide access to first- and second-level RDM support and to nontechnical infrastructure services of the consortium. One of the main pillars of the services, as developed by FAIRagro, will be the nontechnical offers to **support, guide and train the community.** With these services, we aim to foster a cultural change and to enable FAIR data management in agrosystem research. First-level support will be mainly addressed by the provision of **information and training materials and guidelines** (Measures 2.1, 2.4, 2.5, 3.2 and 3.6). Cornerstones of the second level support are the **helpdesk and the data stewards,** organized and networked within the Data Steward Service Center (Measure 2.5). Data stewards will support and be in contact with the community, network external data stewards and ensure excellent RDM. Furthermore, FAIRagro will organize roadshows and provide RDM **education and training** (Measure 2.4), including aspects and requirements for data management plans

(DMP). Training concepts and all contents will be provided as open educational resources (OERs). The education and training services will thus support agrosystem scientists in using the provided technical services (see next paragraph) and will raise awareness of FAIR RDM.

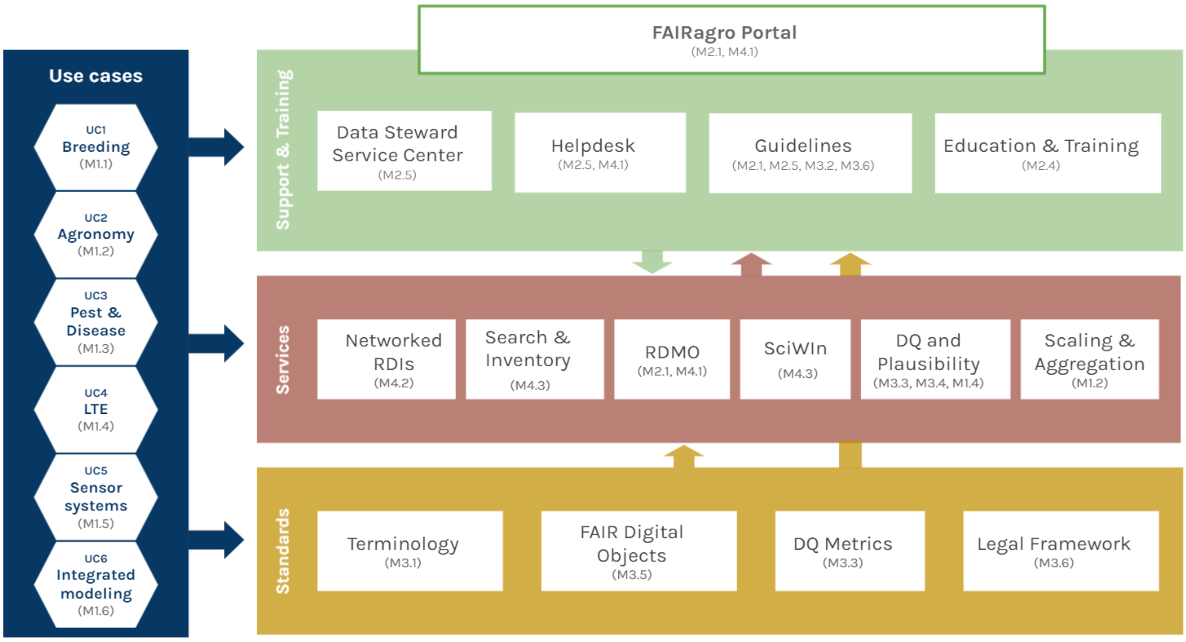

Figure 10: Overview of FAIRagro services. The use cases reflect specific community needs on RDM, which were the basis for the development of the work program. The respective measures providing the service are referred to by their number, e.g. M1.1 or M2.3. The colour schema indicates the task areas (see chap. 5). Use cases are integrated in TA1 (blue), community activities are addressed in TA2 (green), data and infrastructure services in TA4 (red) and standards by TA3 (yellow).

FAIRagro's support and training will be accompanied by the provision of **technical, infrastructure-based services**. One key service here is the **networking of federated RDIs** (Measure 4.2), which will enable discovery, search and reuse of data across distributed infrastructures. Based on the use cases, the community workshops and survey, we identified relevant RDIs for the agrosystem community (Table 2), which will be interlinked. These infrastructures will be provided by FAIRagro partners to the consortium for the long term (see chap. 3.6 and chap. 6). The **search and inventory** services (Measure 4.3), building on the networked RDIs, will provide user-friendly access to all resources available in the FAIRagro repositories. The inventory service will comprise a searchable overview of available infrastructures and repositories, including performance information about these services. This inventory will increase the visibility of each service and make them known beyond their immediate area. The data search service will be based on rich metadata descriptions of agrosystem data and enable, e.g., geospatial selections and semantic search options for field trial information, genotypic or phenotypic features or disease information. In its advanced version, it will include ranking functions, and qualitative, quantitative and FAIR metrics criteria and allow curation of data.

FAIRagro will operate one instance of the **research data management organizer (RDMO)** tool as service for the community (Measures 2.1 and 4.1). The development of subject-specific templates for the creation of DMPs customized to the needs as described in the use cases and further communities are addressed in the work program (Measure 3.2). The service will be accompanied by consulting activities of the data stewards. The **Scientific Workflow Infrastructure** (SciWIn, Measure 4.4) will connect storage and computational instances, versioning and publishing of scientific workflows, such as data scaling and aggregation algorithms (Measure 1.2) or complex data analysis pipelines in a user-friendly, efficient, secure and scalable way. A particular focus in the consortium will be on the provision of services to the agrosystem modelling community to address their articulated needs (see Measures 1.1, 1.2, 1.4 and 1.6). Based on standardized data quality metrics (Measures 3.1 and 3.3), we will develop a **data quality and plausibility** service to facilitate and improve data assessment and reusability for model applications (Measures 1.4, 3.3 and 3.4). Furthermore, a **framework for data scaling and aggregation** will be provided enabling modellers to scale and aggregate data in the required resolution for their specific simulations (Measure 1.2).

Table 2: Overview of repositories from FAIRagro (co-)applicants and participants that provide relevant data for our use cases and that will be interlinked into the network of federated RDIs (Measure 4.2). The first column includes a general description of the infrastructure. The second column indicates the general data domain that is addressed by this RDI. The third column shows the results of the needs analysis and lists UCs that indicated a need for the RDI’s data.

Name and description of the repository	**Data domain**	**UCs**
The **BonaRes Knowledge Library** of UFZ (BonaRes, 2021b) is a structured database for scientific publications on processes in soils and related soil properties providing valuable information for model simulations.	Publications focusing on soils	UC2
The **BonaRes Repository** (BonaRes Repository, 2021; Grosse *et al.*, 2020a) hosted at ZALF provides access to German national and international soil and agricultural research data, soil profile data and time-series from agricultural long-term experiments (LTE).	Soil, Agriculture & LTE	UC2 UC4 UC6
e!DAL-PGP (Arend *et al.*, 2020; Arend *et al.*, 2016), provided by IPK, is a plant genomics and phenomics research data repository (PGP). It publishes large volumes of primary data files in a FAIR manner. PGP is accepted as a data publication repository by Nature publishing group and Giga Science, ELIXIR, de.NBI.	Plant Genomics & Phenomics	UC1
The European data warehouse **Edaphobase** (Hausen *et al.*, 2017; Burkhardt *et al.*, 2014) combines data from diverse sources (literature, museum collections, research data) on soil animals and their sites of occurrence. The data query portal of Edaphobase is open access and provides basic data-analysis tools.	Soil Biodiversity	UC1
GBIS/I is an information system of the IPK ex situ collection for crop and wild relatives. It provides passport data and primary evaluation data (Oppermann *et al.*, 2015).	Plant Genetic Resources	UC1

Name and description of the repository	Data domain	UCs
As part of the JKI Spatial Data Infrastructure, **JKI-DataCube** is an Array Database Management System (Gomes *et al.*, 2020a), which makes multi-dimensional time series of gridded geodata (e.g., weather and phenological data or satellite imagery) available by using standardized web services (Baumann *et al.*, 2021).	Geodata	UC1 UC6
LIMS is the IPK institutional laboratory information management system. It is used in general for documenting and recording all kinds of experimental research data in the lab or in the field like sequence data, phenotypic images or sensor data (Ghaffar *et al.*, 2020; Arend *et al.*, 2014).	Plant Genetic Resources	UC1
The **National Agricultural Soil Inventory (BZW-LW)** of Thünen Institute offers a consistent and representative inventory of chemical and physical soil properties in the top 100 cm in agricultural soils. The main purpose is to understand the influence of climate, land use, agricultural management and site properties on soil organic carbon stocks. Additional parameters such as texture, pH-value, bulk density or parent rock are provided.	Agricultural soil	UC1 UC2 UC4 UC6
The **National Forest Soil Inventory** (BZE Wald) of Thünen focuses on status and changes of soil and forests condition, impacts of environmental factors on soil and crown condition like atmospheric deposition, climate, growth, vegetation, tree nutrition and biodiversity, and sustainable forest management.	Forest soil	UC1 UC2 UC4 UC6
OpenAgrar (Oeltjen *et al.*, 2019) is the collective repository of research institutions affiliated with the BMEL in Germany. It is an open access repository which publishes, stores, archives and distributes documents, publication references and research data.	Publication	UC1 UC3 UC4 UC6
The **Open Data Server** of the DWD (DWD, 2021) provides spatial weather and climate information including data model forecasts, radar data, current measurements and observations and a large amount of separate types of climate data.	Weather, Climate	UC2 UC6
PlabiPD of FZJ contains plant genomic, transcriptomic and ontology data for crops relevant to the German landscape. It provides careful annotation of plant genomes using manual derived annotation and curation data (PlabiPD, 2021).	Plant Genomics	UC6
PhenoRob is an infrastructure of UBN, providing agricultural field data spanning the scale from single plots to the region, including field phenotyping and satellite data. Data are linked to mechanistic and structural plant models as well as economic and ecological landscape models.	Plant Phenomics	UC5 UC6
The **PUBLISSO Repository** for Life Sciences is an infrastructure for secondary publication of articles, books, proceedings and research data provided by ZB MED (PUBLISSO, 2021).	Life Sciences	UC4
The **SRADI** infrastructure of TUM is an interdisciplinary data platform for the comprehensive structuring and sharing of research data in the field of life sciences at TUM.	Life Sciences, Agronomy	UC6

5 Work Programme

The FAIRagro consortium is structured in five task areas (Figure 11).

TA	Measures	TA Leads
TA 1	**Use Cases - Implementation** Measure 1.1: UC1 - Exploiting genotype×location×year×management interactions for crop production Measure 1.2: UC2 - Assessing tradeoffs for optimal crop nitrogen management * Measure 1.3: UC3 - Streamlining pest and disease data to advance integrated pest management Measure 1.4: UC4 - Learning from incomplete data * Measure 1.5: UC5 - Non-invasive phenotyping with autonomous robots Measure 1.6: UC6 - Automated data flows for crop simulation models *	**TA Leads** S. Asseng (TUM) T. Feike (JKI) J. C. Reif (IPK)
TA 2	**Community involvement and networking** Measure 2.1: Communication and Dissemination * Measure 2.2: Community Participation Measure 2.3: Use Case Onboarding * Measure 2.4: Training and Education * Measure 2.5: Data Steward Service Centre (DSSC) *	**TA Leads** U. Stahl (JKI) B. Lindstädt (ZB MED)
TA 3	**Standardization, Interoperability and Quality** Measure 3.1: Standards for Digital Resources * Measure 3.2: Standards for Data Management, FAIRness and Discoverability * Measure 3.3: Measures and Application-data-matrix for Data Quality and Fitness-for-use * Measure 3.4: Data Quality Annotation, Curation and Review * Measure 3.5: FAIR Workflows and FAIR Digital Objects * Measure 3.6: Legal Framework and Machine-Actionable Policies *	**TA Leads** D. Martini (KTBL) C. Weiland (SGN) J.-H. Haunert (UBN)
TA 4	**Infrastructure Services** Measure 4.1: Central Services for the FAIRagro Community * Measure 4.2: Network of interoperable research data infrastructures * Measure 4.3: Searchable Inventory of Services and Data * Measure 4.4: Scientific Workflow Infrastructure (SciWIn) *	**TA Leads** M. Lange (IPK) J. Fluck (ZB MED) B. Usadel (FZJ)
TA 5	**Management and Coordination** Measure 5.1: Project Management, Governance and Financial Controlling Measure 5.2: Sustainability and Business Model * Measure 5.3: Cross-NFDI and international networking *	**TA Leads** F. Ewert (ZALF) X. Specka (ZALF)

Figure 11: FAIRagro task areas (TA) and measures as well as task area leads with their affiliation. All measures marked with an asterisk are relevant to other NFDI consortia.

5.1 Task Area 1: Use Cases – Implementation

Lead: TUM, JKI, IPK

Agrosystem research is facing manifold RDM challenges, e.g., high diversity of data formats, scales and resolution, lack of standardization in addressing these diverse datasets, and having incomplete RDM knowledge, which are described in more detail in chap. 4.1. Improvements in RDM are needed towards FAIR standards to foster agrosystem research through more effective research data utilization, allowing a vastly improved understanding of the permanent, dynamic, concurrent and interrelated changes of the genotype × environment × management system. This also entails the use of modern A.I. tools for data analysis, including machine learning, neural networks, optimization, data assimilation and simulation and standardizing common challenges across many domains related to missing data and data quality.

The objective of TA1 is to address challenges of the agrosystem sciences formulated in six specific use cases (UCs), on which the concepts and services developed in TAs 2-4 for data, metadata harmonisation, standardization, data findability and data access will be based. The outputs of the TA1 actions are therefore an important part for training and education to be aligned with TA2. The UCs in TA1 cover a wide range of agrosystem research questions outlined above considering the scales *genes, plants, crop trial, field*, and *region* (Figure 14).

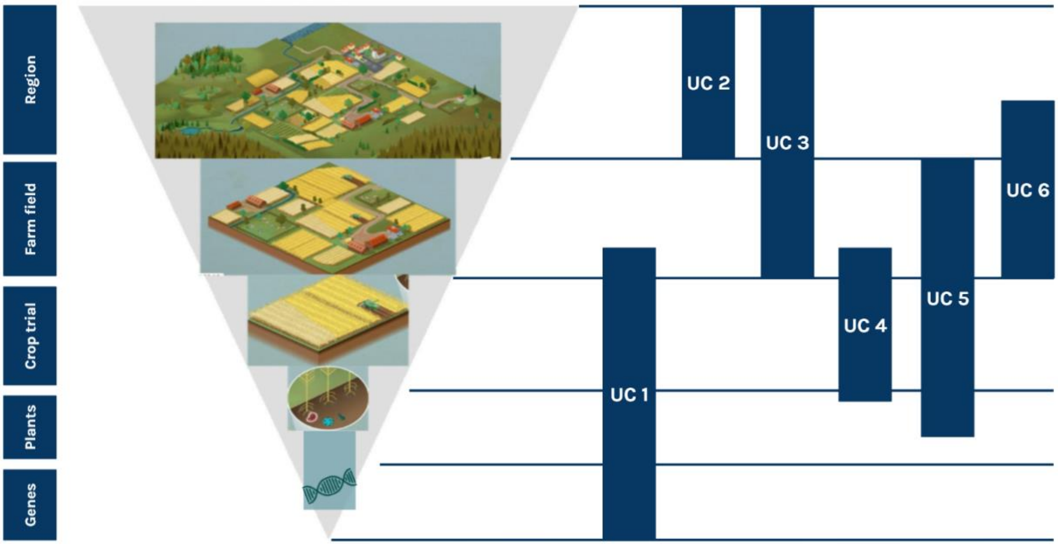

Figure 14: Flagship use cases (UC) shown with their considered levels of organisation. The UCs are described in Measures 1.1 – 1.6 of TA1.

The UCs demonstrate the potential of the implementation of FAIR standards and their automation to substantially increase the research output across a wide range of research activities in agrosystem science. They will address domain- and case-specific challenges that are generalized by the FAIRagro consortium to benefit the entire agrosystem community. TA1 contributes to the definition and harmonisation of standards, implementing data quality, plausibility and scaling services, and supporting the FAIR publication of legacy data. TA1 members are nationally and internationally well-established agrosystem scientists bringing in diverse challenges of different research areas, disciplines and scales.

Measure 1.1: UC1 - Exploiting genotype × location × year × management interactions for sustainable crop production

Contributors: IPK (lead), UHOH

Links Measures 3.1, 3.2, 3.5, 4.2 and 4.4

Background

In current breeding practice, disease-resistant and yield-stable varieties are selected for plant production, which show a high average yield and excellent quality characteristics in multienvironmental field trials. In this process, the focus lies entirely on the scale “genetics”, and trial environments are subject to local production techniques with regard to fertilization and plant protection when varieties are approved. Environmental and management parameters are considered noise terms in statistical models. Only in later stages, such as in regional variety trials or detailed experiments at breeding companies, follows an evaluation of specific variety suitability for individual environments and recommendations for optimal management. However, this "one-size-fits-all" strategy is not suitable for developing locally adapted varieties that consider "genotype × location × year × management" interactions to meet the needs of future sustainable crop production. A consequence is a gap between selection success in experimental environments and on farms. One opportunity to overcome this bottleneck is to deepen our understanding of genotype × location × year × management interactions and develop prediction models that integrate data from different scales, i.e., parameters describing the environment (e.g., soil properties, precipitation, temperature, and plant available water), crop management (e.g., fertilization and pesticide management), and genetics. The integrated use of data from genotypes, environments and crop management is currently hampered by a lack of availability of comprehensive curated data.

Objectives

The main objective in this use case is to build up the required data management processes and prototype analysis workflow that enable knowledge-based prediction models considering genotype × location × year × management interactions for crops. Data on environmental parameters, weather data, trial design, genetics of varieties and important agronomic traits will be curated (Action 1), harmonized, stored and made available in a FAIRagro infrastructure (Action 2). The data will be provided by public real-world laboratories focusing on experimental field stations. The expected results will enable a comprehensive and continuous use of data sources needed for exploiting genotype × location × year × management interactions for sustainable crop production (Action 3).

Action 1: Develop data curation pipelines for genomic and phenotypic data

Existing data currently available in in-house repositories of the participating partners are transformed to follow metadata standards defined in the FAIR-DO approach in Measures 3.1, 3.2 and 3.5. Field phenotyping data for wheat and the trait flowering time are mapped to standard vocabulary and scoring schema defined in TA3. Genotype sample names are harmonized, and the plant material provenance is annotated using persistent plant material identifiers and passport data as defined in standards, such as MCPD and MIAPPE (M1.1.1; D1.1.1). The proposed initial work in wheat will be extended to other crops during the course of the project in close interactions with the Institute of Plant Breeding, Seed Science and Population Genetics, UHOH.

Action 2: Establish the interoperability between datasets generated in different projects

First, Single Nucleotide Polymorphism (SNP) matrices from different genotyping platforms are integrated, and prototypes for SNP imputation are developed (M1.1.2). To facilitate further adoption of the developed curation pipeline, the strategy facilitates an integrated analysis of field phenotyping data using wheat and the trait flowering time as a model crop and model trait. Phenotyping data are scattered and less interoperable in IPK's data infrastructure and need to be exposed and integrated in a FAIR manner in Measure 5.2. The field data will be linked to weather and soil data, and a designed curation pipeline for genomic and phenotypic data will be published as a report (D1.1.2).

Action 3: Analysis and interpretation of model results

The genotype × location × year × management prediction model will generate new data that will be analysed in a software workflow (Measure 3.5) using the FAIRagro SciWIn environment (Measure 4.4). Prediction models and resulting data will be made findable, accessible, reusable and published following FAIRagro standards (M1.1.3, D1.1.3).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M1.1.1	Prototype of curation pipeline for genomic and phenotypic data	Q1 2024
	D1.1.1	Report on designed curation pipeline for genomic and phenotypic data	Q2 2024
2	M1.1.2	Establish the interoperability between genomic and phenotypic datasets collected for wheat	Q1 2025
	D1.1.2	Data have been published following the FAIR criterion	Q2 2025
3	M1.1.3	Genomic, phenotypic data is integrated with information on environmental information	Q2 2026
	D1.1.3	Report on software workflow of genotype × location × year × management prediction model	Q3 2026

Measure 1.2: UC2 - Assessing tradeoffs for optimal crop nitrogen management

Contributors: ZALF (lead), Thünen, UFZ, DWD, GAUG

Links to Measures 3.1, 4.2, and 4.4

Background

Minimizing nitrogen losses to the environment from arable crop production has benefits for biodiversity preservation, water quality, climate change mitigation and resource use efficiency. With improved knowledge of the spatial and temporal heterogeneity in crop-soil nitrogen pathways, simulation models driven by inter- and intraseasonal weather projections can evaluate the probability of optimizing tradeoffs between yield levels and nitrogen losses for various crop rotations and nitrogen management regimes. These models can also inform about nitrogen and carbon interactions to better understand options for carbon sequestration. Despite this potential, the use of process-based models to optimize nitrogen use has been limited due to a number of data-related challenges. First, model simulations are often highly uncertain due to the lack of experimental data for model development and parameterization. While many experiments have quantified the response of soil organic matter and nitrogen to soil types and agricultural management (e.g., Denef *et al.* (2007); Pulleman *et al.* (2005), the data are often not findable or available in standardised form for model use. Furthermore, these studies rarely report sufficient information or data required for simulations, such as soil, weather, management, market or farm economic data. A second key challenge relates to the scaling of input and simulation data to consistent levels (pedon to field to farm to region). The results of aggregation can be heavily influenced by the choice of land use mask, years considered in production area weighting and the aggregation method (Porwollik *et al.*, 2017). Overcoming these data-related challenges results in more robust projections with reduced uncertainties to inform optimal nitrogen management for sustainable crop and soil (carbon) management.

Objectives

In this context, the first objective is to facilitate the translation of published research results into FAIR datasets for use in model calibration to improve the quality of crop-soil nitrogen model simulations (Action 1). The second objective is to reduce model simulation uncertainties associated with data aggregation and scaling (Action 2).

Action 1: Develop a processing pipeline to create model calibration datasets

This UC will extend and add functionality to the Bonares Knowledge Library of UFZ (BonaRes, 2021b), a web-based interface compiling published knowledge on soil processes and properties with detailed metadata on site and boundary conditions, as well as main soil variables. In FAIRagro, together with Measure 3.1, we will pilot an extension of the library to extract values for the main variables, making use of current text digitization software to export the values for use in model calibration (M1.2.1). Importantly, we will link the library and the DWD's Open Data Server to the FAIRagro network of interlinked infrastructures (Measure 4.2, M1.2.2, D1.2.1) such that the newly digitised data can be easily combined with georeferenced weather and soil datasets needed for simulations. The variables and associated variables compiled in the extended library will be harmonized to be consistent with the ICASA standards (AGMIP, 2021a) (Measure 3.1), facilitating its usability in model simulations. Based on this experience, we support the development of the metadata model and the publication guideline in Measure 3.1. Through this last task, the UC will contribute to the FAIRagro key objective 4.

Action 2: Create a workflow for data scaling and aggregation

The core of this activity is to design and implement a workflow for data scaling and aggregation integrated in the SciWIn infrastructure (Measure 4.4, M1.2.3, D1.2.2). With this workflow, we enable researchers to use published data within the FAIRagro network and scale or aggregate these data by specifying the required temporal and spatial scales or any preference for land use masks and land uses. Finally, aggregated data will be provided at the desired resolution in ICASA standards (Measure 3.1). First, algorithms will be developed for each specific data type, e.g., spatial averaging (e.g., weather or climate), median (e.g., phenology) or the value corresponding to the largest area (e.g., soil) as weighted or masked by an underlying land use or production area dataset, as appropriate. Furthermore, the algorithms will offer analytics to quantify the uncertainty resulting from a land use mask choice and aggregation method selection. In the next step, the data aggregation algorithms will be deployed to SciWIn (M1.2.4, Measure 4.4) to set up a workflow to be ready-to-use by researchers.

Milestones and Deliverables

Action	No.	Description	Due end of
1	M1.2.1	Pilot implementation for extracting and exporting publication results	Q3 2024
	M1.2.2	Bonares Knowledge Library links with spatially congruent data in FAIRagro network	Q4 2025
	D1.2.1	Report on developments to Bonares Knowledge Library	Q4 2025
2	M1.2.3	Algorithms and procedures for data aggregation developed	Q1 2027
	M1.2.4	Workflow for data scaling and aggregation added to SciWIn	Q2 2027
	D1.2.2	Report describing the scaling procedure	Q2 2027

Measure 1.3: UC3 - Streamlining pest and disease data to advance integrated pest management

Contributors: JKI (lead), ZEPP, ISIP

Links to Measures 2.1, 3.1, 3.6 and 4.3

Background

Crop protection aims to minimize yield losses due to pests and diseases (P&D). However, there is increasing scientific and public concern regarding the use of pesticides. Accordingly, the farm-to-fork strategy of the EU targets a 50% reduction in the use of pesticides by 2030. Integrated pest management (IPM) aims at minimizing the use of pesticides and related environmental impacts by utilizing versatile crop management options, including decision support systems (DSS). IPM may thus help to minimize related trade-offs. Despite increasing efforts from policy, science and extension regarding the promotion of IPM, its resounding success has so far not been achieved. One major reason lies in the lack of findability, standardization, accessibility and integration of IPM-related data, models and respective decision support. There are several major challenges regarding RDM of P&D data, which are mainly data from yield-loss trials, epidemiological experiments, and P&D infestation data. First, comparison and integration of data is challenged by differences in experimental design (e.g., regarding control treatments) and disease assessment procedures (i.e., timing, scale, sample size). Second, information on the existence and potential accessibility of specific P&D data in Germany is insufficient. Third, different types of models for IPM-related decision support exist, building on the above-described data. However, there is a lack of integrated decision support for plant protection that considers the potential yield loss and environmental risk of pesticide application. The future integration of different types of models is therefore of vital importance to advance IPM-related decision support and make IPM work. Finally, the interplay of continuous crop genetic adaptation, agronomic management changes, climatic changes, landscape level effects and P&D evolution is highly complex and requires a solid database that can be utilized effectively through data integration, analysis and modelling by the research community.

Objectives

To overcome the above-described RDM-related limitations and challenges in current P&D research in Germany, this UC has three main objectives and related actions.

Action 1: Develop guidelines for standardization of yield loss trials

First, this action aims to develop guidelines to realize harmonization and standardization of yield loss trial data in close cooperation with Measure 3.1. This would foster comparability of studies, allow meta-analysis for same crops and pathogens and improve the (re)usability of results for P&D model development. We will utilize existing protocols (e.g., from EPPO) and standards (e.g., from IBSNAT or ICASA) and map to existing ontologies. Relevant stakeholders and data holders will be involved through consultation and workshops (together with Measure 2.1; M1.3.1, D1.3.1).

Action 2: Establish an inventory for and improve the accessibility of IPM-related data

Additionally, UC3 aims to establish an inventory of IPM-related data, including experiments, surveys, and auxiliary data of different spatial scales (i.e., landscape structures, land use and management data). This aims at using existing datasets and paves the way for effective future data storage. This work will also support Measure 3.1 in developing publication guidelines and corresponding metadata for IPM-related data. In addition, crucial questions of intellectual property rights and data access options will be dealt together with Measure 3.6. The inventory will be made available via OpenAgrar and the FAIRagro Portal (Measures 2.1 and 4.3; D1.3.2).

Action 3: Integration of P&D models and crop yield models

Finally, a blueprint for the integration of P&D models and crop models shall be developed that demonstrates the effective utilization of the above-described strongly improved P&D data infrastructure (D1.3.3). The workflow is executed for a specific crop-pathogen system (e.g., Septoria tritici in wheat) and will serve as a blueprint for future IPM-related DSS. Therefore, a dynamic P&D model (ZEPP) and a crop yield model shall be linked in the SYNOPS GIS platform. This demonstrates the potential for a holistic assessment of plant protection and minimization of trade-offs between environmental risks (SYNOPS) and potential crop loss due to P&D when deciding for alternative plant protection strategies at the field, farm and landscape levels. The detailed workflow of model linkage, including retrieval and integration of various empirical data for model calibration and evaluation, as well as RDM-related challenges, will be recorded and made available to the research community as part of the elaborate description of the blueprint (D1.3.3).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M1.3.1	1st workshop with IPM data stakeholders conducted	Q4 2023
	D1.3.1	FAIRagro guidelines for yield loss trials published	Q1 2025
2	M1.3.2	Develop IPM data-inventory	Q4 2025
	D1.3.2	Inventory published	Q4 2026
3	M1.3.3	Link P&D and crop models	Q2 2027
	D1.3.3	Report on workflow of model integration published	Q3 2027

Measure 1.4: UC4 - Learning from incomplete data

Contributors: ZALF (lead), ZB MED, UBN, TUM

Links to Measures 2.4, 2.5, 3.1, 3.3, and 4.3

Background

Data from long-term field agricultural field experiments (LTE) are of high interest in agrosystem sciences, especially for use in model validation and simulations. Although numerous LTEs with different factors are conducted both nationally and internationally (BonaRes, 2021c; Grosse *et al.*, 2021; Grosse *et al.*, 2020a), their data have rarely been published and thus are not available for application in crop models. LTE studies often aim to identify the respective major constraints of crop yield in a particular setting, e.g., the effects of pests and diseases, water or nutrient deficiency, or detrimental climate conditions and try to understand long-term effects on soil carbon to support measures for enhancing soil carbon sequestration. However, usually only one or two factors are considered, providing only limited information about the respective setting, which can be a major obstacle in terms of using LTE data for more generic meta studies or the application of biophysical or machine learning yield models. Furthermore, the use of LTE data for crop model calibration and validation requires knowledge about the data's quality and plausibility, which is often missing in published datasets.

Objectives

This UC aims to improve availability and to complement LTE data, which are of special interest for crop modellers or meta-studies. One activity in this use case will be the support of LTE experimentalists in publishing their datasets in existing FAIRagro infrastructures (Action 1). Furthermore, to make LTE data applicable for generic studies, additional contextual information will be collected complementing LTE data. An iterative approach will be developed, automated as far as possible and published as a guideline on how to complement existing LTE datasets for further use (Action 2). Additionally, a service will be developed to analyse and assess data quality and plausibility (Action 3), which is needed to reuse LTE data in model applications.

Action 1: Support and standardize new LTE data publication

To date, only a minor fraction of comprehensive LTE experiments have been published. Based on the previous work of ZALF in BonaRes (Grosse, 2019), a metadata model meeting FAIR criteria providing a comprehensive description of LTE data will be developed and published in cooperation with Measure 3.1. To facilitate reuse of LTE data, a data scheme will be developed together with Measure 3.1, allowing the harmonized management of these datasets (M1.4.1). Supported by the Data Steward Service Center (Measure 2.5), further LTE starting with experiments from our Partners UBN and TUM will be published in a FAIRagro infrastructure applying the harmonized data scheme and newly developed metadata model (M1.4.2). An LTE data publication guideline will be developed and published (Measure 3.1; D1.4.1), facilitating further publication of LTE data in FAIRagro infrastructures. Researchers will be supported by

FAIRagro data stewards (Measure 2.5) to fill in the rich metadata, which will greatly enhance the findability of these data in the FAIRagro Portal.

Action 2: Data curation supplementing already published incomplete LTE data
Published LTE datasets often do not contain extensive metadata descriptions and context information that would be required to use the data in simulation models. In this activity, we will curate relevant German LTE metadata that have high relevance for simulations (e.g., in the BonaRes Repository). These metadata sets will be a) transformed into the newly developed metadata model developed in Action 1 together with Measure 3.1 and b) checked for completeness supported by the Data Steward Service Center (DSSC, Measure 2.5). For the update of the metadata information, the web forms of the search service (Measure 4.3) will be used. Additionally, this work will be documented in such a way (D1.4.2) that it could be used as a basis for developing and training automated processes (not part of FAIRagro).

Action 3: Implement a service to assess data plausibility
To use LTE data in crop models, data quality and plausibility must be identified. Together with TA3 (Measure 3.3) and based on preliminary work of BonaRes, a community service to assess data plausibility is further developed (D1.4.3) and made accessible via the FAIRagro Portal. This service comprises the development of an expert-based plausibility database complemented by a statistical model-based approach based on a compilation of numerous biophysical and statistical crop yield models, accounting for numerous interactions between different effects (D1.4.4). Together with Measure 3.3, data flow processes are developed and applied to feedback into the original datasets by enriching their metadata.

Milestones and Deliverables

Action	No.	Description	Due end of
1	M1.4.1	LTE metadata model and harmonized metadata scheme published	Q2 2025
	M1.4.2	Data from two LTE published	Q2 2026
	D1.4.1	LTE data publication guideline published	Q4 2026
2	D1.4.2	LTE data curation guideline published	Q4 2026
3	D1.4.3	Plausibility database including model results established	Q3 2027
	D1.4.4	Data quality and plausibility service established	Q4 2027

Measure 1.5: UC5 - Noninvasive phenotyping with autonomous robots

Contributors: FZJ (lead), UBN

Links to Measures 2.5, 3.1, 3.2, 3.4, and 4.3

Background

Sustainable crop production requires knowledge on the structural and functional status of the crops in the field to plan and execute precise field interventions. For instance, rather than frequent pesticide or fertilizer application, a plant-specific demand should be determined, and plant-level interventions should be planned and executed. An important agricultural research question is how such an increase in precision can be achieved. Drones and field robots in combination with modern sensor systems and cameras are promising tools to provide on demand real-time data on crop traits; thus, plant phenotyping has been adopted in plant breeding and modern agriculture. To quantify complex plant traits such as stress resilience or disease resistance, combinations of multiple noninvasive phenotyping with sensors and advanced data analytics methods and machine-learning algorithms are nowadays state of the art (Watt *et al.*, 2020). While significant progress has been made in the past years on sensor technology, machine learning, and autonomous robotics, we are still in the early stages of realizing the full potential of these data. One possible analysis extension is the fusion and combined analysis of repeatedly acquired sensor data with environmental data by combining information across years, including information on previous field interventions. Another relevant factor is the integration of genetic and biochemical characteristics of different crop types and breeding lines. Furthermore, it is still an open research question whether and how such autonomous multisensor robots can be developed and employed in the field and how data streams from the combination of sensors and algorithms can be merged to identify relevant plant traits or detect stress situations at an early stage.

Objectives

The main objective of this UC is to investigate and showcase the potential of a FAIR data infrastructure for the development of multimodal data analytics methods and machine-learning algorithms for in-field plant phenotyping and agricultural robotics. In particular, this UC aims to define requirements for data curation services. This includes automated tools for quality assurance and data harmonization that increase the usefulness of data for combined analyses. This UC will also explore ways to visualize field environment data stored in the data infrastructure to researchers to assist them in operating multisensor systems or controlling agricultural robots (e.g., deciding which field operations a robot shall perform). By conducting a pilot study, this UC will act as an early adopter of the services developed in FAIRagro and, in particular, in Measures 3.4. Furthermore, we are working closely with the data stewards from TA2 to incorporate the specific phenotype data requirements for use in machine learning from this use case into FAIRagro standards and guidelines (Measure 3.1) and to inform the data stewards (Measure 2.5) in how to curate additional datasets accordingly. It remains a scientific challenge to align the requirements of machine learning experts with the potential of field phenotyping. While field phenotyping provides a huge amount of quantitative crop data, essential information may still be missing to make these data usable for machine learning approaches.

Action 1: Development of RDM requirements for field phenotypic data

With this action, we will define the needed interfaces between the two disciplines of phenotyping with machine learning. We will build on the knowledge of various questionnaires and networking activities, which have already identified the gaps and overlaps and will test these theoretical concepts with the real-world data of the Excellence cluster PhenoRob. Using specific PhenoRob datasets, we will evaluate in which form plant-phenotyping data are best suited for use in multimodal data analytics methods and machine-learning algorithms. We will scientifically evaluate which data preparation steps are usually undertaken before the phenotypic data could be fed into an algorithm by combining the knowledge from the PhenoRob cluster with an extensive literature review (D1.5.1). In a next step, with the support of data stewards (Measure 2.5), we will compare these findings with the requirements of the other UCs and adapt them accordingly. Finally, this action, together with Measure 3.4, will develop a recommendation, which of these steps can be automated and which can thus be moved to the service infrastructure of FAIRagro. This action will focus on PhenoRob data but will at the same time support the data stewards in adapting the standards and guidelines accordingly (together with Measures 3.1). Furthermore, in such a way, the data stewards are informed how to curate further datasets accordingly.

Action 2: Generation and release of a benchmark dataset of complete field phenotyping data

After the general description of the requirements (Action 1), we will generate a benchmark dataset of complete field phenotyping data, which cover several vegetation periods and various spatial and temporal resolutions and addresses a wide range of relevant plant traits from major Central European crops (D1.5.2). We will use the data from the Central Experiment of the Excellence Cluster PhenoRob and select an unprecedented dataset that shall be suitable for a wide range of machine learning approaches. These data will be published within the PhenoRob infrastructure and will be made findable within the FAIRagro search service (Measure 4.3). This dataset will be discussed within FAIRagroo and beyond, and it will be determined which elements are of general interest and thus can serve as a blueprint for a universal data concept for field phenotypic data. Here, again, close cooperation with the data stewards is envisioned to identify the common schemes across user groups and data concepts.

Action 3: Pilot study with data curation and visualization services

The most relevant data of the benchmark dataset (Action 2) will be further structured based on the knowledge of Actions 1 and 2, and the most relevant field phenotypic data will be provided and visualized in the field using the software that is developed in Measure 3.4. We will test these novel visualization technologies to provide 3D point clouds of crops that were collected in the previous vegetation period for a visual comparison of plant structure across years. Such a real-world pilot study (M1.5.1, M1.5.2) is generated with FAIRagro visualization services for mobile devices (Measure 3.4). The usability will be evaluated by experts in the field, and the feedback will be used to improve the visualization and data organization (D1.5.3). Additionally, these data will serve as test cases for automated data curation in this real-world setting. Existing gaps will be identified and documented to help to improve the data organization, automated data curation and data visualization of FAIRagro.

Milestones and Deliverables

Action	No.	Description	Due end of
1	D1.5.1	Review paper describing the state-of-the-art	Q1 2024
2	D1.5.2	Benchmark dataset completed and made available to FAIRagro, accompanied by a data article	Q4 2024
3	M1.5.1	Pilot study (first year) with PhenoRob members with the goal to provide real-time data in the field	Q3 2025
	M1.5.2	Pilot study (second year) with PhenoRob members with the goal to test data access and visualization of field data under field conditions	Q3 2026
	D1.5.3	Article describing the outcomes from the two pilot study on the data access and visualization in the field	Q4 2026

Measure 1.6: UC6 - Automated data flows for crop simulation models

Contributors: TUM, LfL, HSWT, BSA, ATB

Links to Measures 1.2, 1.3, 1.4, 3.1, 3.3, 3.4, 3.5, 3.6, 4.2 and 5.3

Background

Crop simulation models have become important tools in agricultural research and crop systems analysis (Chenu *et al.*, 2017). They play an increasing role in research on decision-making for automation in crop management from planting to harvesting. These crop models require data from diverse sources with different accessibility, size, aggregation, units, quality, temporal and spatial scales and formats to operate, including field records, soil surveys, weather stations, climate change scenarios, on-time field sensors, remote data from drones and satellites for data assimilation and seasonal and market weather forecasts. Some of the required data are generated based on qualitative information from different sources and converted into quantities, such as cultivar parameters. One important aspect hereby refers to the accessibility of data sources: some of these data are publicly available (e.g., weather data), while some are collected in ongoing research and are not published yet (e.g., some field sensor data), and others have restricted access rights depending on regulations. The integration of all model input data requires experts in a range of disciplines, such as agronomy, soil science, crop science, breeding, biogeochemical, hydrological, ecological, pathology, agricultural economy, meteorology, climate science and informatics for locating, accessing, and transferring these data, converting file formats, scales and units, quality checks and filling missing information. The crop models also generate large amounts of data that need to be quality checked, documented, made available to derive decisions for robots and drones conducting future field operations, prepared for long-term archiving, and made accessible to the research community in agriculture and other fields (e.g., earth systems-, climate impact science) for other studies.

Objectives

Measure 1.6 will identify data requirements and define a generic framework for a seamless integration of data with crop models in close collaboration with Measures 1.2 and 1.3. We will outline and develop a prototype for a seamless workflow to apply crop model inputs, crop model simulations and crop model outputs for parameterization of the DSSAT-Potato model (Šťastná *et al.*, 2010) as part of research for automation of a potato growth simulation from planting to harvest. The expected results will enable a comprehensive and continuous use of data sources (according to the respective data access rights; Measures 3.6 and 4.2) needed for operating the DSSAT-Potato model and crop models in general. The developed framework and prototype will guide the development of seamless infrastructures to integrate a range of data and simulation models in the agricultural research communities and hence in FAIRagro as a whole. This UC will integrate the FAIRagro data principles (Measures 3.1 and 3.5) into a scientific workflow infrastructure (Measure 4.4) and consider existing data infrastructure [e.g., SRADI (Smart Rural Area Data Infrastructure) at TUM] to enhance the quality of the existing data and enable interoperability with other data infrastructures.

Action 1: Development of workflows for crop model applications as digital objects SciWIn
We will integrate and enhance existing RDIs (e.g., SRADI at TUM) and data formats (e.g., XML-client of BSA) within the network of interlinked infrastructures (Measure 4.2) and the scientific workflow infrastructure (SciWIn, Measure 4.4, M1.6.1, D1.6.1), as well as combine manually available data with already integrated datasets and new possible automated data flows developed at ATB according to the common (meta) data standards for crop models (e.g., ICASA, OGC-Standards) elaborated by Measures 3.1 and 3.5.

Action 2: Automating data processing and storage for crop models
We will develop digital objects to automatically identify, download and process data inputs and outputs (M1.6.2) in accordance with corresponding data access rights policies (Measures 3.6 and 4.2). Inputs thereby comprise static and dynamic, real-time data, which cover among others sensor data from UAV-based 3D images or tractor-operated cameras on crops and ambient conditions (e.g., disease pressure (also in collaboration with Measure 1.3) supported by ATB, TUM and HSWT. Routines will be developed closely linked with SciWIn (Measure 4.4, D1.6.1, D1.6.2) for the identification, selection, conversion and correction in case of faulty data, for the example of a DSSAT-Potato model application (link to Measure 3.4). Newly created data (in particular outputs) will be made available in the long term in collaboration with BSA based on enhancements to existing archiving solutions (link to Action 1). BSA's archiving competence and its participation in other NFDI consortia guarantees knowledge exchange and added value to FAIRagro and to the whole NFDI, which will be realized via the data stewards (Measure 2.5) and communication events (Measure 2.1) and work on cross-cutting topics (Measure 5.3).

Action 3: Automated plausibility checks for inputs and outputs of crop models
We will develop digital objects for automated procedures for plausibility checks of inputs and outputs of a potato crop model. Specific crop modelling knowledge by TUM and LfL will be provided to extend the plausibility database developed in Measure 1.4 (M1.6.3, D1.6.3).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M1.6.1	Workflow designed for crop model applications as digital objects in SciWIn of of FAIRagro	Q4 2023
1 & 2	D1.6.1	Workflow as a digital object for a seamless and automated integration, processing and evaluation of in- and output data for crop models	Q1 2024
2	M1.6.2	Designed framework for automated data processing for crop models	Q4 2024
	D1.6.2	Software tool for automated data search, -transfer and -conversion for crop simulation model input	Q1 2025
3	M1.6.3	Cropping system specific knowledge is integrated into UC4's services to conduct plausibility checks for in- and outputs for crop modelling	Q3 2025
	D1.6.3	Protocol and software tools for plausibility checks for crop model in- and output data	Q4 2025

5.2 Task Area 2: Community Involvement and Networking

Lead: JKI, ZB MED

TA2 stands for FAIRagro’s community-driven approach of gathering, engaging, training and educating, and supporting our community (Figure 15). A special characteristic of the agrosystem research community is its distribution over many disciplines, institutions and organisations. Agrosystem research also includes different stakeholders (e.g., consultants, breeders, farmers, policy makers), and is linked to other domains such as plant, environmental, biodiversity and earth sciences. FAIRagro and the NFDI will only be successful through a sufficient integration of the respective communities and activities. Measure 2.1 meets this challenge by gathering our community. Here, we build on work that has already been done at the genesis of the FAIRagro community, e.g., three community workshops (see chap. 4.1), and consolidate our communication, networking, and dissemination concept. In addition, Measure 2.2 and Measure 2.3 engage the community by creating opportunities to participate directly and indirectly in FAIRagro's developments. TA2 manages the CAB (see chap. 3.4) and interacts with the community by bundling ideas, needs, support requests, and feedback and providing training. Therefore, TA2 actions, the development of TA3 and the services from TA4 will be regularly evaluated by the community. Ultimately, after an initial phase in Measure 2.3, the community can expand FAIRagro by suggesting ideas for UCs that cover other agrosystem research areas or bridge into new domains of the agricultural sciences. By onboarding new UCs and participants, FAIRagro will progress towards the strategic aim of one integrated NFDI4Agri community.

Figure 15: TA2's main tasks and interaction with the FAIRagro community.

In addition to receiving the input of the (wider) FAIRagro community, it is also important to reach out to our own community. We dedicate two measures to this task. Measure 2.4 provides training and open educational resources as training material, and content for educational programs. All training and teaching content is merged in a module called “agrosystem science”, which can be added to general courses. The Data Steward Service Center (DSSC, Measure 2.5) consolidates all RDM network activity exchanges in the NFDI environment, and channels support requests to the appropriate data stewards. These data stewards are in direct contact with the community, train further data stewards, and ensure excellent RDM.

Measure 2.1: Communication and Dissemination

Contributors: JKI (lead), DPG, DBG, DAFA, GPW, TTT

Links to all measures in the work program

Objectives

The overall goal of this measure is to connect and consolidate our community by focusing on two tasks: (i) communication and networking and (ii) information. By conducting roadshows (on-site marketing events), regular workshops, seminars, and talks at conferences of scientific societies, FAIRagro will interact directly with its community. These interactions will foster exchanges about their specific RDM requirements and spread information about our services, events, and opportunities to participate in FAIRagro. In addition to gaining presence at external events, FAIRagro will host three summits to bring the agrosystem community and adjacent disciplines together and enable networking and participation structures. In addition to direct and active communication and networking, a central access point to all content and services provided by FAIRagro — the FAIRagro Portal — is needed. Dissemination channels must be established to ensure information transfer from all TAs to our users. A community manager will be installed at JKI to manage this service.

Action 1: Network management and expansion

Through the preparatory work of NFDI4Agri and FAIRagro (e.g., community workshops, survey), our consortium has already reached out to a large network of experts and stakeholders from various disciplines of agrosystem and agricultural science. The existing management of these contacts will be bundled and professionalized with a system CRM (provided technically by Measure 4.1). To expand our network, we will develop a flexible roadshow concept tailored to different types of stakeholders (M2.1.1). The main goals of these regular on-site marketing events are: (i) FAIRagro promotion (i.e., aims, services and support, events, and participation opportunities), (ii) on-site collection of needs and feedback (together with Measure 2.2), and (iii) recruitment of experts, data-driven working groups, repository hosts, and data stewards. We will recruit institutions, organisations, and expert clusters that are not yet involved with FAIRagro or established multiplicator groups such as country associations (e.g., fdm.nrw, HeFDi, NFDBBB, TKFDM) based on an inventory (M2.1.1). The regular roadshows that will be held throughout Germany (D2.1.1) will be accompanied by at least one data steward (Measure 2.5). Additionally, we will conduct talks, sessions, and workshops at the annual conferences of scientific societies such as DBG, DPG, GPW, and DAFA. Through networks of RDM experts organised under the roof of TTT (D2.1.2), we will promote surveys conceptualised in Measure 2.2 via their member network. Together, we will develop sustainable concepts and strategies (e.g., permanent working groups or sessions) to spur cultural change.

Action 2: Organisation of FAIRagro community summits

We will organise three FAIRagro community summits at the end of the first, third, and fifth year after project start (M2.1.2). The target group for the first two summits is the entire agroscience community. The focus of each of the summits will be linked to the respective development status of FAIRagro. The goal of the first summit is to present and refine the objectives of FAIRagro with the help of the community, publish a position paper, constitute the CAB (Measure 2.2) and announce the first call for new UCs (Measure 2.3). During the first summit, data stewards will introduce themselves to the community in M2.5.1. The second summit aims to promote established FAIRagro services, realized/onboarded UCs, lessons learned (feedback summary), evaluation/continuing of actions, readjustment of set up aims/tasks for the final period based on needs, and the election of the CAB (second round). The third summit will report FAIRagro outcomes/service portfolios, include and involve the broader NFDI4Agri community in adjacent disciplines such as forestry, economics, animal sciences in the "onboarding" process, and establish permanent networks within the community as prepared and designed in Action 1.

Action 3: Editorial office and operation of the FAIRagro Portal

The FAIRagro Portal, which will be technically implemented in Measure 4.1, will be the unified access point for all FAIRagro services and content in a corporate design and will be built to i) inform, ii) guide and iii) interact with the community. In the first phase, we conceptualize the portal with respect to its fundamental structure and usability (M2.1.3) from a user point of view to create a user-friendly, intuitive website. In the second phase, we set up an information workflow with all other TAs to centralise the information pipeline to the FAIRagro Portal editor. To inform the community, the FAIRagro Portal editor will identify and utilize different communication channels, such as classic media (website), social media (Twitter and similar), press (interviews, TV, radio), or direct inquiry to different stakeholders (universities, societies, etc.), by curating user-friendly content and continuously promoting activities (e.g., an event calendar), and knowledge (e.g., guidelines, open educational resources, best practices, catalogues, positions) in an appropriate manner (e.g., via videos, blogs, newsletters).

To guide the community and provide first-level support, the FAIRagro Portal will include an overarching search function of the portal's content with features such as multifaceted and semantic search, an autocomplete function, and FAQs. To interact with the community, the FAIRagro Portal will have features such as a help desk (operated by Measure 2.5) with a connected ticket system and integrated CRM (implemented by Measure 4.1), a submission form for new UCs and participants (Measure 2.3) and tools such as RDMO for DMPs, and computing services provided by FAIRagro (Measure 4.1).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M2.1.1	Roadshow concept and inventory list of target institutions	Q1 2024
	D2.1.1	Roadshows conducted	Q2/Q4 2024 Q2/Q4 2025 Q2/Q4 2026 Q2/Q4 2027
	D2.1.2	Talks, workshops on conferences of societies and expert networks conducted	Q4 2024 Q4 2025 Q4 2026 Q4 2027
2	M2.1.2	FAIRagro community summits conducted	Q1 2024 Q1 2026 Q4 2027
	D2.1.3	FAIRagro principles published as position paper in a journal	Q4 2023
3	M2.1.3	FAIRagro Portal concept (fundamental design, structure and functions)	Q4 2023
	D2.1.4	Operating the FAIRagro Portal	Q1 2024

Measure 2.2: Community Participation

Contributors: ATB (lead), all (co-)applicants and participants, FAIRagro community

Links to all other TAs and measures.

Objectives

The goal of this measure is to provide opportunities to openly and flexibly adapt FAIRagro according to the needs and feedback of our community. One cornerstone towards this goal is giving the community a representative and guiding role in FAIRagro by creating a CAB. The CAB should be endorsed by the broader community and consist of internationally recognized experts from academia and economics. The CAB's tasks, composition, and implementation according to FAIRagro governance are described in chap. 3.4. This measure is responsible for organising the constitution of the CAB, which will be supported by the FAIRagro Secretariat (Measure 5.1). In addition, it is important to know the wishes and needs of our wider scientific community to orient the strategic direction of the project. It is also important to understand how the project and our developed services are received and can be improved. Here, we ensure a structured and, above all, transparent information flow from the community to FAIRagro and vice versa.

Action 1: Establishment of the Community Advisory Board (CAB)

Elections of the nine members of the CAB (as embedded in the governance and organizational structure of FAIRagro; Figure 8) will be organized at each of the three community summits (Measure 2.1). Guidelines will be developed to govern the composition of the CAB and ensure broad, relevant, and balanced coverage of the different disciplines. Since disciplines may be added during the project, this guideline will be adapted to the future direction of FAIRagro. Before the summits, FAIRagro Plenary will invite and nominate potential candidates (e.g., experts or representatives of professional scientific societies) from different disciplines within our active community network (Measure 2.1). Nominations can also come directly from the community at the community summits. Potential candidates then stand for election at the community summits and are elected by summit attendees for a two-year term in accordance with the CAB guideline (M2.2.1). The CAB will stay informed of new developments of FAIRagro and provide strategic advice regarding future implementation and the continuation phase after five years. CAB meetings occur once a year, with the first meeting to take place in the second year (M2.2.2). The results of these meetings will be compiled in an evaluation report and catalogue of recommendations. During the legislative period, members of the CAB will prepare a short interim report on the perception of FAIRagro within their domain and suggestions for improvements or developments on an annual basis (D2.2.1). These tasks are supported by the FAIRagro Secretariat (Measure 5.1).

Action 2: Structured collection of needs and feedback

Conducting an actionable needs and feedback assessment entails evaluation, forwarding, and presentations. We gather feedback and information to assess needs by using the full range of possible communication channels, including attending conferences and workshops (all TAs, Measure 2.1). We will extract, process, and aggregate information on relevant needs (Measure 2.1). We are also in close contact with the DSSC (Measure 2.5), which provides hands-on insights from the community (e.g., help desk tickets). Further systematic online surveys and interviews will be conducted during the project based on content from the TAs (M2.2.3).

To correctly measure the applicability and adoption rate of FAIRagro services and infrastructures, user feedback and performance indicators will be integrated into the FAIRagro Portal and all suitable web-based products (Measure 2.1, TA4) collected and analysed. Furthermore, we will include specified questions in each general survey, offer opportunities events to provide feedback at FAIRagro, such as roadshows, hackathons, working groups, and training units (e.g., Measures 2.1 and 2.4), and collect feedback from social media channels (Measure 2.1). After processing, needs and feedback will be raised with those responsible, (i.e., from the user to the developer). Developers will provide a statement (to the user) of all the possible solutions. All feedback, needs, questions, and suggestions processed in this measure will be regularly updated, publicly presented according to the GDPR, and summarized in a final FAIRagro implementation report via the FAIRagro Portal (D2.2.2; D2.2.3).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M2.2.1	CAB election at FAIRagro community summits in accordance with the actual CAB guideline	Q4 2023 Q4 2025 Q4 2027
	M2.2.2	CAB meetings	Q2 2024 Q2 2025 Q2 2026 Q2 2027
	D2.2.1	CAB evaluation and feedback reports (alternating main and interim reports)	Q2/Q4 2024 Q2/Q4 2025 Q2/Q4 2025 Q2/Q4 2026 Q2/Q4 2027
2	M2.2.3	Community surveys	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q4 2027
	D2.2.2	Yearly needs and feedback report (including onboarding gap update and onboarding report, see Measure 2.3)	Q2 2024 Q2 2025 Q2 2026 Q2 2027
	D2.2.3	Final implementation report of all Measure 2.2 and 2.3 actions	Q4 2027

Measure 2.3: Use Case Onboarding

Contributors: ATB (lead), ZALF

Links to all other TAs and measures

Objectives

To make FAIRagro sustainably open, we will create a structured and community-driven onboarding process targeting previously unaddressed research domains and disciplines. In doing so, FAIRagro can successively expand existing UCs (EUC, e.g., via content-related or methodological connecting points or interfaces to FAIRagro services) or with new UCs, opening up a new branch of agrosystem research (NUC, see chap 3.1 and Figure 7). *Bridge UCs* (BUC, see chap. 3.1 and Figure 7) are a special focus, especially at an advanced stage of FAIRagro developments. Bridge UCs not only have links to existing achievements and services of FAIRagro, but also the potential to expand the FAIRagro community to include other disciplines and domains of agricultural sciences (e.g., animal via projects such as the Leibniz Innovation Farm, leibniz-innohof.de) towards one integrated NFDI4Agri community (see chap. 3.1).

Action 1: Establishment, consolidation and implementation of the onboarding process

In the first phase, we will elaborate upon the onboarding process and design a viable concept of implementation (M2.3.1). Since the onboarding process should be backed by the agrosystem community, specification and orientation will be presented at the first FAIRagro community summit (Measure 2.1) and then further discussed and evaluated to develop recommendations for the SC at annual CAB Meetings (Measure 2.2). A standardized submission platform on the

FAIRagro Portal will be developed together with Measure 2.1 and TA4. The most up-to-date information on the needs and possible points of connection of new UCs and a submission template will be published and communicated, thereby building the basis for the open call for UCs (D2.3.1). In addition to the needs of FAIRagro itself, the results of the needs and feedback analysis (Measure 2.2) will be included. A list of criteria for new UCs (e.g., added value, interfaces, connection of new discipline and communities, feasibility/capacity, strategic fit, etc.) will be established, thus building the basis for the subsequent selection process (M2.3.1). The open call will be announced after one year to allow FAIRagro to strengthen primary planned developments first. It will be a continuous and dynamic call with a duration of 2.5 years and a total funding period of new UC of 3 years (M2.3.2).

Action 2: Managing the onboarding process

In addition to the submission platform on FAIRagro Portal, we will also recruit and engage potential new participants based on actual trends, needs, and feedback within FAIRagro and its community (Measure 2.1). Three review and onboarding rounds (after 1.5, 2.5 and 3.5 years) will be organised. The CAB will be involved in the review to provide recommendations for the use case selection (during annual CAB meetings, organized in Measure 2.2) and the SC will decide which use cases will be selected and integrated into FAIRagro (D2.3.2). Once selected, Measure 2.3 takes on a mentoring role and helps the new UC managers integrate into the consortium, and other experts will be involved as needed (M2.3.2). The annual onboarding process will include UC development reports of incorporated UCs, which will be published on the FAIRagro Portal and discussed at the CAB meetings (D2.3.2).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M2.3.1	Criteria catalog for and direction of onboarding process	Q4 2023
1	D2.3.1	Constantly updated information on submission platform formulating needs and gaps and points of connection for new UCs	Q1 2024 Q1 2025 Q1 2026
1	M2.3.2	Dynamic call for proposals	Q1 2024 Q1 2025 Q1 2026
2	D2.3.2	Onboarding evaluation and UC development report	Q2 2024 Q2 2025 Q2 2026 Q2 2027
2	M2.3.3	Onboarding of new UCs and Participants	Q3 2024 Q3 2025 Q3 2026

Measure 2.4: Training and Education

Contributors: ZB MED (lead), TUM-GZW, DAFA

Links to all other measures

Objectives

Handling research data in agrosystem science in accordance with state-of-the-art standards needs to be implemented in a sustainable and comprehensive manner. The NFDI4Agri Survey 2020 already highlighted this and concluded a need for support, as most information service providers identified a need for training concepts and teaching materials tailored to the specific requirements of agricultural science (Senft, 2021). Therefore, the establishment of FAIRagro must be accompanied by a strong effort to strengthen the data skills of the respective scientific community. The main objective will be to develop a training and teaching module “agrosystem science” tailored to the following points: (i) online and personal training, (ii) provision of training material as open educational resources (OERs), and (iii) contribution to the education of data stewards/data managers.

Action 1: Online and personal training

As a basis **for online and personal training** for the agrosystem community, we will look at existing materials from projects on a general level, e.g., FDMentor (Biernacka *et al.*, 2020), but especially upon NFDI section training and education. Since FAIRagro is already involved in this section, the knowledge base and general training (planned to be built up at the end of 2022) will be an important source and collaboration opportunity. The content of the training resources is elaborated in TA3 and TA4, but especially in the six UCs from contributions of the data curators (TA1) and data stewards (Measure 2.5) involved. The UCs focus on topics such as data curation, data quality aspects, publication guidelines, data harmonization, and data integration. The implementation and solutions provided for these topics from TA3 and TA4 together with TA1 are provided by the data stewards and which will be transferred into data workshop concepts for the use cases and community training (see Measure 2.5). The content will be appropriately edited and tailored to the specific didactical needs of undergraduate students, graduate students, and scientists. The output of the action will be a training concept (D2.4.1).

For the dissemination of training, we will follow four possible concepts:

(1) Integrate existing training options in cooperation with external partners, such as NFDI4BioDiversity (U Bremen Research Alliance), NFDI4Earth (EduHub of NFDI4Earth) and de.NBI, into the training program. One focus of de.NBI is conducting trainings for cloud technologies and containering (bioinformatics) and special aspects such as data management plans (German Crop BioGreenformatics Network, 2021), including agrosystem-specific templates for DMPs (see Measure 3.2). Three trainings each (one per year from 2024 to 2026) will be offered to the FAIRagro community by instructors from de.NBI and FAIRagro, together with NFDI4BioDiversity and NFDI4Earth.

(2) Two trainings per year will be integrated as workshops and seminars at conferences of professional societies (participants of FAIRagro) together with Measure 2.1.

(3) To strengthen cooperation with (applied) universities who have faculty in agriculture and related fields, we will establish training courses with the installed RDM units at the universities (three per year), which usually offer training on a generic level and provide local services. FAIRagro will add the content of the “agrosystem module” to that training. This concept has already been suggested and proven within NFDI4Health.

(4) To cover additional needs for training as voiced from the community via the help desk, we plan to add the capacity for two additional trainings in the years 2025 and 2026 (M2.4.1). Beyond this, the role of FAIRagro is to activate the community to take part training through the actions of Measure 2.1, including the FAIRagro Portal.

Different teams from FAIRagro will be constituted from the pool of data stewards and representatives from TA1, TA3 and TA4, and the training will be customized to the needs of the respective participants. The training coordinator at ZB MED is central to this effort. Feedback on training and further needs are identified with Measure 2.2 via feedback questionnaires as supplements to the single training courses or in the context of user surveys.

Action 2: Provision of training material as open educational resources
The training materials will be produced in formats such as videos, social media, wikis, websites, and podcasts on specific topics. All of the training resources will be made available as open educational resources (OERs). Beyond the training materials based on the outputs of the TAs and especially TA1 by contribution of the data stewards, further existing OER materials relevant for the FAIRagro community, such as material developed by or together with partners NFDI4Earth or NFDI4BioDiversity (D2.4.2), will be collected and processed. All materials will be provided in a structured manner on the FAIRagro Portal in cooperation with Measure 2.1 and within the knowledge base of the NFDI section training and education (M2.4.2).

Action 3: Contribution to education of data stewards
FAIRagro will contribute to the education of data stewards via the “agrosystem module”, which is conceptualized in action 1 and tailored to educational programs. The module is developed as a part of curricula (1) with partners from NFDI4Earth on one of the three initially planned sites of the NFDI4Earth Academy, (2) in the NFDI section T*raining and Education* for the planned certificate course where subject-specific modules are added, (3) to a FAIRagro train-the-trainer course on RDM. The generic parts of this module can pull from the well-established train-the-trainer concept of the project FD Mentor (Biernacka *et al.*, 2020) (4) as part of education programs at graduate schools (M2.4.3, D2.4.3).

As a dissemination point, the TUM-GZW offers infrastructure for FAIRagro courses to young scientists and doctoral researchers which facilitates further development and consolidation. The dissemination will be expanded to other graduate schools compiled in M2.1 (once per year starting in the third year) with support of DAFA. The “agrosystem module” for training and education will be further developed and conceptualized by the training coordinator, data stewards, and other persons from the consortium involved in training activities (see action 1) after a minimum of three years of experience working with the use cases (TA1), the help desk, and with the broader community. The data stewards will create a handbook for the FAIRagro training and education module (D2.4.4).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M2.4.1	Trainings conducted	Q4 2024 Q4 2025 Q4 2026 Q4 2027
	D2.4.1	Training and dissemination concept	Q4 2023
2	M2.4.2	Collection and processing of OER on the FAIRagro Portal	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q4 2027
	D2.4.2	Concept for OER on the FARIagro Portal	Q4 2023
3	M2.4.3	Participation on educational programs	Q4 2024 Q4 2025 Q4 2026 Q4 2027
	D2.4.3	Concept for training modules “agrosystem” for educational programs and train-the-trainer workshops	Q4 2024
	D2.4.4	Handbook on training content for education programs	Q2 2026

Measure 2.5: Data Steward Service Center (DSSC)

Contributors: ZALF (lead), FIZ, JKI, UBN, IPK, GAUG

Links to all other measures

Objectives

The DSSC is the central institution within FAIRagro to harmonize data management and being anchored in the scientific community. The center organizes the continuous exchange of RDM knowledge, and experience with other institutions, e.g. via training (with Measure 2.4), channels user requests from the community, and transfers knowledge from the TAs to the data stewards. FAIRagro data stewards are experts in the field of RDM for agrosystem research supervising and training data curators of TA1 on a close-knit basis. Data stewards have core competencies in RDM (e.g., cross-scale from genes, phenomics, and management to region: sensitive data, remote sensing, time series, plant, soil and related FAIRagro data). Knowledge and expertise is pooled to provide the full range of expertise to the community. This understanding of the different FAIRagro domains and the integrative work of the data stewards will foster the coalescence of the community.

Action 1: Establishment and operation of the DSSC

The DSSC will be an essential building block to weave the community and FAIRagro as one. It will be the central contact partner for all questions regarding data management, and will continue to expand this function. After the establishment and familiarization phase (D2.5.1), the coordinator of the center will interact with other consortia (namely NFDI4Earth, NFDI4Health and NFDI4BioDiversity) and networks, pass on state of the art RDM knowledge to the pool of data stewards and co-works on a cross-NFDI strategy for a networked data steward support structure. Support requests via the help desk (Measure 4.1) were channeled and meetings among the data stewards are organized and enriched with external experts (e.g. representatives of other consortia). The data steward network will be expanded by connecting more institutions such as universities (e.g. GAUG) and associating additional data stewards (M2.5.2). Feedback from data stewards regarding FAIRagro products will be aggregated and passed on to Measure 2.2.

Action 2: Professionalization of the data stewards and training of further multipliers.

An overview of specific RDM expertise needed for FAIRagro will be developed and published together with all FAIRagro TAs. Data stewards will acquire general (e.g., using DMPs, data publication, suitable repositories) and FAIRagro specific RDM knowledge during the start-up phase and pass it to the "agrosystem module", e.g. through the planned handbook, and train-the-trainer workshops (Measure 2.4). They co-work next to institutional data stewards, participate in the planned NFDI4Earth Academy, become experts and develop core competencies that will be distributed within the FAIRagro data stewards network (D2.5.2). Core competencies are (i) a confident understanding of agrosystem data domain and scales (“Genes”, “Plants”, “Crop trial”, “Field”, and “Region”) and multiple data formats (e.g., farm and management data, time series, long term field experiments, geospatial data such as remote sensing), (ii) knowledge about the use of the services developed in TA4, (iii) competence in legal matters and data policy relating to the publication and use of sensitive data and (iv) training skills (e.g. train the FAIRagro trainer, guide FAIRagro metadata, publication along the FAIR principles, recommended legal standards, policies and guidelines (Measure 3.6, M2.5.3). By being located at the respective institutions, the data stewards are always up to date and well trained. Data stewards will be installed at the following institutions (compare section 3.1 of this proposal) with outstanding RDM expertise on hand: **IPK** (training courses, guidance for long term data deposition, utilizing DMPs, implementing MIAPPE, data stewardship activities), **ZALF** (regular community trainings on data publication, standardized metadata, DOI, long term preservation, data acquisition and FAIR provision), **FIZ** (intellectual property rights in distributed information infrastructures, copyright law, licencing, data privacy law and IT security, legal support), **JKI** (support researchers on curation of geodata within spatial data infrastructures (SDI) considering quality aspects and provenance, enabling SDI interoperability), **UBN** (management of large volumes of heterogeneous data originating from field robots and drones (Farming 4.0), teaching and knowledge transfer, information visualization).

Action 3: Supporting the FAIRagro community and networking beyond FAIRagro

Data Stewards are the link between the community including the FAIRagro use cases (TA1) and the developers of the FAIRagro products (TA4). They will have a comprehensive standing of data issues, process support requests from the help desk (2nd level support; Measure 4.1) and establish contact to the developers of TA4 if necessary (3rd level support). Besides the personal support of the community, easy-to-understand materials provide low-threshold and personal-extensive access to FAIRagro-relevant information (1st level support). If necessary, they will intensively take care of a special project (e.g. supporting data curation pipeline in Measure 1.1, LTE data curation in Measure 1.4 and the definition of requirements for data curation services in Measure 1.5) in order to implement a professionalized RDM within a short period of time. Integrated into their hosting institutions, they work together with Measure 2.4 in developing necessary documents and training materials. Existing DMP templates will be adapted by Measure 3.2 to the needs of FAIRagro and disseminated by the data stewards. They support the use of the FAIRagro RDMO tool and collect feedback on FAIRagro during all contacts with the community (e.g. roadshows and the “data stewards’ network corner” on community summits (M2.5.1) in Measure 2.1). Data stewards perform the role of trainers for the community. In particular, they train the data curators working in the use cases, but they are also responsible for the FAIRagro participants and the wider community; they regularly meet (M2.5.4), pass on the acquired knowledge (D2.5.3), foster RDM competence and thus promote the coalescence of the broader NFDI4Agri and adjacent communities.

Milestones and Deliverables

Action	No.	Description	Due end of
1	M2.5.1	DSSC presented on FAIRagro community summit	Q1 2023
	D2.5.1	DSSC operative	Q2 2023
	M2.5.2	DSSC networking event	Q4 2024 Q4 2026
2	D2.5.2	Data stewards operational; launch of FAIRagro helpdesk	Q4 2023
	M2.5.3	Regular input to agrosystem modules and training material	Q2 2024 Q2 2025 Q2 2026
3	D2.5.3	FAIRagro specific RDM and core competencies published	Q1 2026
	M2.5.4	FAIRagro data stewards network meeting	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q3 2027

5.3 Task Area 3: Standardization, Interoperability and Quality

Lead: KTBL, UBN, SGN

The aim of TA3 is to facilitate reuse, quality screening and annotation of research data. We will consolidate the FAIR data designs of the involved research infrastructures (see chap. 4.4) by guiding stakeholders towards FAIR standards, improving the interoperability of involved vocabularies and ontologies, establishing metadata models and providing publication guidelines for the various types of FAIRagro data to ensure findability and reusability (Measure 3.1). We will extend *schema.org* and data management plans (Measure 3.2). The task area partners will establish data quality metrics based on an application matrix (Measure 3.3) and implement a toolbox (including an algorithmic suite and a review system) for rating and improving the data content with respect to the formalized quality metrics (Measure 3.4). We will provide an abstraction layer of FAIR Digital Objects (FDO) to enhance the interoperability of services and workflows (Measure 3.5) along the data life cycle and develop a legal framework and compile on this basis all outputs of TA3 into actionable policies and legal metadata standards for improved reusability (Measure 3.6). The planned combined measures in TA3 address current gaps in (meta)data management and handling (Measures 3.1, 3.2), data quality (Measures 3.3, 3.4) and data governance (Measures 3.5, 3.6). In all TA3 measures, we work in a feedback loop with the use case experts (TA1) to provide appropriate standards for their requirements and to obtain rapid feedback on feasibility. We support TA2 in the development of training materials and are supported in the organization of events, dissemination and through the work of the data stewards. All developed standards, specifications and tools from TA3 determine the implementation of the services in TA4.

Measure 3.1: Standards for Digital Resources

Contributors: KTBL (lead), JKI, ZB MED

Links to Measures 1.1-1.6, 3.2-3.6, 4.2, and 4.3

Objectives

Measure 3.1 addresses the heterogeneity of data and metadata standards in the agrosystem research data landscape (see chap. 4.2). One of the main objectives of Measure 3.1 is to improve awareness and compliance with FAIR, which will be addressed by Actions 1 and 2. This encompasses the extensibility and interoperability within the domain by clarifying the need for a formal language for knowledge representation conforming to principle *I1* of FAIR (GO FAIR, 2021a) following the broader approaches within the NFDI of using semantic web technologies and methods. Reuse of existing vocabularies and ontologies is often poor, partly due to lack of awareness but also due to knowledge gaps on reuse, embedding and modification design patterns. Surmounting these obstacles is a major objective addressed by Action 2. From a practical point of view, integration and extension mechanisms have to be demonstrated on core standards, ontologies and vocabularies that have proven to be relevant and adequate. This supplies the proof-of-concept that is necessary to take along the agrosystem research community towards FAIRer data handling. Finally, work from all other measures in this TA and requirements from TA4 will be picked up and integrated into this overall system of networked standards and ontologies.

Action 1: Assessing standards, ontologies and vocabularies

We will compile an inventory of mature standards, ontologies and vocabularies (M3.1.1, addressing D3.1.1), that can serve to describe digital agricultural resources needed within the initial UCs (see chap. 4.1) and infrastructures to be integrated as a contribution to the domain-specific resource sets of the NFDI as a whole. To supplement the inventory and create FAIR awareness, a community workshop will be conducted (M3.1.3). Inventory entries will be assessed with regard to compliance with the FAIR principles (M3.1.2). Based on that, we will provide recommendations for (re)use, highlighting gaps and identifying development needs. Inventorization and FAIR assessment will accompany the integration of use cases into infrastructure and be conducted throughout the entire project.

Action 2: Providing best practices and guidance for reuse and interoperability

Relying on existing minimum (meta)data profiles like DCAT-AP and settled standards for describing e.g. cardinalities of data fields like SHACL, we will provide extensions to these profiles covering domain specific requirements (e.g., spatial, genotype, phenotype, M3.1.4) documented in D3.1.2. The purpose of these subsets is to provide a low entry barrier to the integration of existing infrastructures and facilitate data annotation for researchers. Building upon this, we will deliver reuse, cross-embedding and modification design patterns, illustrating the reuse of existing standards, vocabularies and ontologies based on agrosystem domain examples (D3.1.3). We will provide guidance on how to tackle special issues and requirements, an example thereof being that many agrosystem data objects are changing in space and time (M3.1.3, D3.1.1) - as illustrated, for example, in UCs 1 and 6 (Measures 1.1 and 1.6).

Action 3: Enhance the data findability and semantic-based information retrieval by setting up a publication guideline for data publication.

The publication guideline will include a metadata model that consists of the preservation metadata (Provenance, Reference, Context, and Access Rights) as well as rich representation information, which maps the data object into the corresponding agronomic domain (Systems, 2012). This representation metadata includes minimal spatial, genotype and phenotype information based on standardized terminology that in part already exists, e.g., through standards such as MIAPPE. We will provide a guideline that will be published (M3.1.4), tested for data publications from the different use cases and approved by all FAIRagro partners (D3.1.3, D3.1.7). Updates for further UCs and data types will continuously be included. The metadata model will serve as a specification for metadata exchange for the search service, the infrastructures to be integrated and the middleware (Measures 4.2 and 4.3). Such a harmonized agro metadata model will empower similar data searches within the different infrastructures and allow for central searches in the search service developed in Measure 4.3.

Action 4: Integrating standards, ontologies and vocabularies

RDM complying with various data standards imply handling heterogeneous semantics and ontologies, which necessitates effective mappings and the provision of alignments (D3.1.5) but also offers guidance on integrating legacy data representations and delivering implementation hints. In addition, we will provide extensions eventually necessary for data quality annotation [e.g., Kahn *et al.* (2016)] and representation of legal policies (D3.1.6). All standards for digital resources produced in TA3 will be gathered, including minimal information standards, domain ontologies, metadata schemas, specifications and other data standards and will be registered in already existing searchable resources (for example: fairsharing.org, agroportal.lirmm.org or bioportal.org but also github.com) as demanded by FAIR principle F4 (F4, 2021) and implied by fair principle I2 (I2, 2021). They will be cross-linked from FAIRagro Portal and integrated into the OLS-based terminology service hosted by ZB MED and the search service provided by Measure 4.3 (M3.1.5).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M3.1.1	Inventory of relevant standards describing digital resources	Q4 2023
	M3.1.2	FAIR assessment of inventory of standards methodology including recommendations for improving FAIRness	Q1 2024
	M3.1.3	Community Workshop on inventory and assessment	Q4 2024
	D3.1.1	Compilation of relevant standards for RDM in agrosystem sciences including their FAIR assessment and improvements to make them FAIR	Q2 2024
2	M3.1.5	Compilation of minimal and FAIR metadata profiles available	Q4 2025
	D3.1.2	Documentation of identified minimal information standards as a basis for more complex domain models	Q3 2025
	D3.1.4	Reuse and integration design pattern guidelines and examples and guidelines for semantic compliance and on addressing agronomy specific requirements	Q2 2026
3	M3.1.4	Metadata set for publication standards is defined	Q3 2025
	D3.1.3	Publication guidelines for spatial, genotype and phenotype	Q4 2025
	D3.1.7	Best practice examples for data publication are described in at least three examples from the use cases in TA1	Q3 2027
4	M3.1.6	Standards resources registered and integrated into existing infrastructures and cross-linked in the FAIRagro portal	Q4 2026
	D3.1.5	Ontology and vocabulary mappings and alignments provision	Q4 2026
	D3.1.6	Recommendations for technical representation of data quality and legal rights	Q2 2027

Measure 3.2: Standards for Data Management, FAIRness and Discoverability

Contributors: ZB MED (lead), KTBL, SGN, IPK

Links to Measures 1.1-1.6, 3.1, 3.3, 3.4, 4.1, and 4.2

Objectives

To support RDM in the FAIRagro community from the beginning, domain-specific templates for DMPs are needed (Action 1). In order to enhance discoverability of digital resources, we will develop domain-specific extensions of schema.org in Action 2.

Action 1: Provide specific templates for data management plans

We will use existing DMP templates already developed in related agronomic science projects (e.g., by the European projects CAPITALISE, AGENT or CONVERGE) and will customize them to the special needs of the UCs (TA1) and further community needs. Those templates (D3.2.1) will be released within the public FAIRagro RDMO instance provided by Measure 4.1.

Action 2: Improve the findability of our digital resources by establishing a domain-specific extension of schema.org

This extension enables structured data on web pages, such as publication landing pages in repositories. It helps in implementing search tools in TA4, allows the major search engines that support *schema.org* to deliver good results to a much broader user base and enables integration of existing *schema.org* features, such as user ratings as required by the review system designed in Measure 3.4. In addition to the *schema.org* standard recommendations and domain specification in the biological domain (Garcia *et al.*, 2019), we will use existing vocabularies and ontology definitions. To agree on a domain-specific extension of schema.org (D3.2.3) and to promote its adoption, we will involve a variety of domain experts, expert associations and service providers to work collaboratively, via two AgriHackathons (M3.2.2), on definitions. This *schema.org* extension will be complemented with validation mechanisms (D3.2.4).

Milestones and Deliverables

Action	No.	Description	Due end of
1	D3.2.1	DMP templates for the use cases are developed and proved by the use cases	Q4 2026
2	M3.2.2	AgriHackathons on schema.org extension is organized	Q2 2024 Q2 2026
	D3.2.3	Domain specific extension of schema.org is agreed	Q4 2025
	D3.2.4	Validation mechanisms for the domain specific extension of schema.org	Q2 2027

Measure 3.3: Measures and Application-data-matrix for Data Quality and Fitness-for-use

Contributors: JKI, ZALF

Links to Measures 1.1-1.6, 3.1, 3.4, 3.6, and 4.3

Objectives

Approved data quality is a prerequisite for the reusability of data. Crop-relevant data are characterized by heterogeneity with distinct quality requirements and open issues. This includes, for instance, dependencies of spatial, spectral and temporal resolutions of time series data on modelling results (Measures 1.1 and 1.5), uncertainties related to data aggregation levels and their scale-specific representativity (Measure 1.2), the accuracy and completeness of phenotyping and legacy data (Measures 1.1 and 1.3), and the plausibility of data from long term agricultural field experiments (LTE; Measure 1.4). In summary, data quality review, curation and documentation are needed; however, community-driven standard criteria are still lacking. Based on existing and representative agrosystem datasets related to FAIRagro UCs (TA1), relevant data quality metrics with special considerations of data-fitness-for-use aspects will be identified, formalized and transferred to application-specific queries. In doing so, the results of Measures 3.4 and 1.4 will be included, which focus on specific data quality aspects related to agrosystem robotics (e.g., outlier rate) and LTE data plausibility. The work leads to an exemplary data-for-fitness-curation primers and a curation geodata/LTE data checklist for the data stewards (Measure 2.5). The outcomes support the definition of additional FAIRagro quality metadata descriptors (Measure 3.1) and are incorporated into the FAIRagro inventory and search service (Measure 4.3).

Action 1: Identification of relevant agrosystem data quality aspects

For better transparency, exemplary data use profiles are documented, commented and visualized. In addition, relevant data quality aspects required for typical applications in agrosystems related to LTE and geodata based on UCs (TA1) and community workshops (M3.3.1) will be identified and documented (M3.3.2).

Action 2: Data-for-fitness-use formalization

We take an important step from rather generic data quality standards towards an application-specific formalization of the fitness-for-use of data in a multidimensional application-data-matrix framework. A selected set of data-fitness elements in the FAIRness measurement templates as well as a proposal of new metadata elements for data fitness classifications will be developed. For typical data from long-term experiments and for geodata we will construct a visualization and documentation system for the relationships between the quality of the input data and the expected quality of the outcome of an analysis (D3.3.2).

Action 3: Definition of application-specific queries

As soon as exemplary data quality metrics are published, we will determine typical application-specific queries that will be documented (M3.3.4), presented and discussed with the community (M3.3.5), and operationalized (M3.3.6). The formal technical representation and accompanying samples, e.g., for queries, will be provided through Measure 3.1, which will facilitate the findability of dataset fit for specific applications. Such queries can then be issued either by data users or through search engine-like interfaces, as developed in TA4.

Milestones and Deliverables

Action	No.	Description	Due end of
1	M3.3.1	One community workshops to identify users' application-specific data quality needs	Q1 2024
	M3.3.2	Frequently used datasets and user needs for specific applications	Q3 2024
2	M3.3.3	Exemplary elements of data-fitness for FAIRness measurement templates	Q3 2025
	D3.3.1:	Conceptual design of application-specific data quality matrix for agrosystem data	Q3 2025
	D3.3.2	Exemplary framework for documentation of relationships between data quality of input data and expected quality of outcomes	Q3 2026
3	M3.3.4	Concept of typical application-related data quality queries for implementation in search engines	Q2 2026
	M3.3.5	Community workshop to present and discuss data quality metrics	Q4 2026
	M3.3.6	Concept for operational query and application of data quality metrics along exemplary data types	Q3 2027
	D3.3.3	Joint publication on data-for-fitness curation primers	Q1 2027
	D3.3.4	Data-for-fitness curation checklist to support data stewards	Q4 2027

Measure 3.4: Data Quality Annotation, Curation and Feedback/Review

Contributors: UBN (lead), FZJ

Links to Measures 1.5, 3.1, 3.3, and 4.3

Objectives

In extension of Measure 3.3, which focuses on formalizing criteria and metrics of data quality, Measure 3.4 focuses on developing a toolbox (including an algorithmic suite, reference datasets and a review system) for rating and improving the data content with respect to the formalized quality metrics. For example, while Measure 3.3 may define the “outlier rate” as an important quality metric, Measure 3.4 may yield an algorithm for outlier detection that could be used to quantify the outlier rate of a given dataset or allow it to remove them. Measure 3.4 will specifically address current research on agrosystem robotics, precision farming, and in-field plant phenotyping conducted in UC5 (Measure 1.5), which requires combined analyses of data acquired with a large heterogeneity of sensors (e.g., multi- and hyperspectral cameras or 3-d laser scanners) and sensor platforms (e.g., field robots and UAVs). Information on data quality is essential before combining multi-source data and, in particular, for successfully using multi-modal data analytics and machine learning methods. Therefore, by providing algorithms for rating and improving the data quality, Measure 3.4 will act as a prerequisite for artificial intelligence in crop science. Since not all quality issues can be automatically resolved, the algorithms will be complemented with interfaces that allow humans to review datasets based on appropriate visualizations. Moreover, Measure 3.4 will provide plausibility tests and visualizations of datasets and their quality to support on-site decision-making, especially addressing the needs of breeders or researchers who work with robots and mobile sensor systems in the field. The development of tools in Measure 3.4 will include their conceptual design, prototypical implementations and tests with UC5 participants. Based on these tests, we will decide on the final version of the toolbox, which we will implement in a way that benefits UC5 and facilitates reuse by other researchers. This means in particular that all code will be documented and released as open source.

Action 1: Development of guidelines

In close cooperation with Measure 3.1, we will develop guidelines to help data contributors resolve ambiguities with respect to the aggregation, processing, and quality level required for data they load into the data infrastructure (D3.4.1). In particular, we will address the needs of researchers in agrosystems robotics and sensing, who are usually experts with respect to (geo)spatial data and sensors but often have a lack of knowledge on concepts of plant phenotyping and the related standards (e.g., the MIAPPE standard).

Action 2: Algorithmic suite and reference data for data curation

We will develop an algorithmic suite for data curation enabling the automatic detection, quantification, and resolution of certain quality issues (D3.4.2, M3.4.4). This relates, in particular, to the detection and removal of outliers with machine learning methods, plausibility checks developed in close collaboration with UCs, and processes of data harmonization (e.g., resampling to ensure that different datasets have a comparable spatial, spectral, and temporal resolution). Furthermore, the algorithmic suite will contain methods for data conflation, meaning combining multiple datasets that overlap in space and/or time to a single dataset.

Action 3: Review system

Because not all quality issues can be automatically resolved, we will develop a review system that will allow experts to provide feedback on the data quality by annotating the data with the help of visualization tools (M3.4.5). In our measure within TA3, we will focus on developing such visualization tools as well as appropriate representations of quality-related data annotations, which will become accessible and editable through the user interface within the search service implemented in Measure 4.3. Datasets that have passed defined data curation procedures (including automated steps and reviewing by experts) will be annotated with quality certificates. Multiple datasets that have been harmonized with each other will be provided in the form of sets of harmonized data layers, as required, e.g., by UC5.

Action 4: Visualization for on-site decision support

We will develop methods to visualize the harmonized data layers on mobile devices for on-site decision support, e.g., to provide humans in control of agrosystems robots with an overview of the situation (M3.4.3). To visualize georeferenced information, we will use the Geospatial viewer for CKAN (Comprehensive Knowledge Archive Network) resources, which supports the JavaScript libraries OpenLayers and Leaflet for geovisualization. Similarly, to visualize abstract (i.e., non-spatial) information such as phylogenetic trees, we will implement visualization components based on existing JavaScript libraries, such as the JavaScript InfoVis Toolkit. As an open challenge, we will look at ways to visualize spatial and abstract information on crops and information on data quality in an integrated manner (i.e., rather than in separate windows).

Milestones and Deliverables

Action	No.	Description	Due end of
1	D3.4.1	Workshop with experts in mobile sensing, machine learning, and plant phenotyping to identify aggregation, processing, and data quality levels that are optimal for multi-modal data analytics and machine-learning methods for plant phenotyping tasks	Q2 2023
2	D3.4.2	Tests with prototypical implementation of algorithmic suite for data curation and members from UC5 completed and results condensed into requirements for the final version of the algorithmic suite	Q4 2024
	M3.4.4	Toolbox for data curation fully operable and released as open source accompanied with documentations and video tutorials	Q4 2026
3	M3.4.5	Review system fully operable as part of the search platform developed in TA4	Q4 2027
4	M3.4.3	Visualization components for spatial and non-spatial information and optimized for mobile devices in the field environment; scientific publication	Q2 2025

Measure 3.5: FAIR Workflows and FAIR Digital Objects

Contributors: SGN (lead), JKI, KTBL

Links to Measures 1.1-1.6, 3.1, 3.2, 4.2, 4.3, 4.4, and 5.3

Objectives

Measure 3.5 will develop the conceptual basis to embed specifications and guidelines from vocabularies and ontologies identified by Measures 3.1 and 3.2 into FAIR workflows (Goble *et al.*, 2020) and digital object data models, specifically FAIR Digital Objects [FDO; De Smedt *et al.* (2020); Islam *et al.* (2020)]. These are integrated into the central service architecture for software development and database integration in TA4, with the aim of facilitating and automating data integration and RDI federation (Measure 4.2) and supporting the development of data acquisition tools ensuring FAIR principles in newly generated data (Measure 4.4). Levering established approaches such as the extended FAIRification Framework of GO FAIR (GO FAIR, 2021b), we will mainly expand on the principles *F2* to enable automatic detection and classification of relevant services (Measure 4.3), *I3* to semantically cross-link resources and open those up by mapping and alignment (provided by Measure 3.1) and *R1* (accurate and relevant attributes) to enable automatic inference and decision on the usefulness of data. We will also provide design blueprints for the bidirectional interfaces to the search service (Measure 4.3) and for the mapping of heterogeneous information systems (Measure 4.2), with a particular focus on rich attributes for legacy data sources (Measure 3.6).

Action 1: Design machine-actionable data operation layer based on FAIR Digital Objects

We will implement a data abstraction and interoperability layer (Lannom *et al.*, 2020; Schwardmann, 2020) for FAIRagro's infrastructure in accordance with recommendations of RDA's Data Fabric IG (RDA, 2016). The data will be linked through a FAIR Digital Object infrastructure (DiSSCoTech, 2020) in which data collections and aggregations such as Data Cubes (Mahecha *et al.*, 2020) are the principal object types. For this, it is planned to use the FDO middleware Cordra (Cordra, 2021), which we will expand by combining a semantic data model with tools for data validation (Gayo *et al.*, 2017) and machine learning (Grieb *et al.*, 2021). The data architecture will be detailed in a comprehensive report (D3.5.1) and elaborated within the scope of Actions 2 and 3.

Action 2: Implement a FDO service ecosystem for the FAIRification of infrastructure

Accordingly, we will design an FDO ecosystem (Collins *et al.*, 2018) including persistent identifier schemes [e.g., IGSN (2021)], data type registries and semantic validation pipelines for FDO to enable integration of agronomic data into the data streams of hyperinfrastructures such as NFDI and EOSC. The consistent improvement of interoperability within FAIRagro's service architecture will be addressed by employing open specifications involving standards for Application Programming Interfaces [APIs, e.g., OpenAPI (2021); SmartAPI (2021)], protocols [e.g. Digital Object Interface Protocol/DOIP 202 (DONA, 2021), M3.5.1], identifiers (Hardisty *et al.*, 2021), containers [e.g., FDO; Islam *et al.* (2020)], PhenoPackets (2021). All of the work conducted on the API specifications, guidelines and tool development will be provided as a continuously updated document (D3.5.1).

Action 3: Supporting FAIR workflows and models

Develop a framework to support FAIR usage of research data and workflows leveraging community-developed schemas, standards (RO-Crate (2021) / Common Workflow Language (CWL, 2021), and topical registries (WorkflowHub (2021), D3.5.2). Workflows are considered research products of their own kind and will be in this respect represented by a specific FDO type encapsulating and linking to methodology, input/output data and tool specifications to enable model comparison, sharing and publishing (with PIDs). This will be investigated in a FDO hackathon (M3.5.2). In addition, we will adapt W3C's PROV data model (Zhang *et al.*, 2020; Thessen *et al.*, 2019) for our UCs to support provenance annotation including (D3.5.2) descriptions of workflow that led to the data in concern (why and how given data has been obtained by a model or particular processing).

Milestones and Deliverables

Action	No.	Description	Due end of
1	D3.5.1	Report detailing domain model specifications and core digital object model and the associated ecosystem of PIDs, protocols and other key data specification	Q4 2024
2	M3.5.1	First conceptual version of Protocol and API specifications as required by Measure 4.3	Q4 2023
3	M3.5.2	FDO hackathon with user and developer community	Q4 2025
	D3.5.2	FAIR workflow descriptions compiled based on UCs (TA1) supplemented by rich attributes with particular focus on provenance data enrichment	Q2 2026

Measure 3.6: Legal Framework and Machine-Actionable Policies

Contributors: FIZ (lead), SGN, KTBL

Links to Measures 1.3, 1.6, 3.5, and 5.3

Objectives

Measure 3.6 develops legal guidelines, policies, model agreements and legal metadata standards so that we can provide agrosystem data at the FAIRagro infrastructures more findable, accessible and reusable (see Measure 3.5). As researchers, research institutions and agro-repositories are dealing with diverse research methods and diverse sources of data (see in particular Measures 1.3 and 1.6), the legal framework for agrosystem data is inconsistent, and significant legal uncertainty reduces data sharing in the community and retains different kinds of restrictions. Special emphasis is placed on incorporating a data steward within the DSSC (Measure 2.5) to handle questions of legal requirements and licences for research data within FAIRagro. We support the researchers, research institutions and FAIRagro infrastructures, providing a legally sound and FAIR framework with the following actions:

Action 1: Establishing legal guidelines, policies and model agreements

We establish legal support and compliance with legal guidelines and policies. In the first step, an assessment reviews the status quo of dealing with the legal challenges at the FAIRagro partners, such as their conditions and terms of use (M3.6.1). The findings will be presented and discussed in a workshop (M3.6.2) that serves as a kick-off for bringing together already engaged and interested partners in FAIRagro and NFDI (i.e., in cooperation with NFDI4Earth, see Measure 5.3). The permanent working group (M3.6.3) initiated herein is led by Measure 3.6 and strengthens a community-driven legal framework. We develop chapters for data policies and legal guidelines (D3.6.2) that are legally reliable, taking into account the specific interests of the agrosystems research stakeholders. For typical research settings in FAIRagro, we prepare model agreements that are focused on the FAIR principles and assign the legal metadata standards (see D3.6.3). Coordinated by the DSSC (see Measure 2.5), the recommended policies, guidelines and model agreements are published on the FAIRagro Portal and introduced to the community at conferences and institute meetings (M3.6.4).

Action 2: Developing legal metadata standards and implementing them in FAIRagro services

We identify the legal categories relevant for agrosystems data such as copyright (including licensing), trade secrets, data protection law, good scientific practice and other legal backgrounds for data regarded as (legally) sensitive. These legal categories will be applied to the agrosystems datasets, clustering them into a classification (first version, D3.6.2). With the legal classification for agrosystems datasets, we develop legal metadata standards that explain the metadata standards' impact on FAIRness and their technical implementation at FAIRagro infrastructures. The impacts and consequences of the legal metadata standards proposed are discussed at workshops with the FAIRagro community, with metadata specialists and with legal experts (M3.6.5). Workshop topics might focus on consequences of different copyright statuses (data copyright protected or public domain), licensing models and specific licensing recommendations, data protection law statuses (data protected or not) or international outreaches. Integrating the feedback from the workshops, we complete and finalise the legal metadata standards (D3.6.3), including an attached manual. Then, we support the technical implementation of the legal metadata standards in the FAIRagro infrastructures.

Action 3: Releasing updates about data regulations

We monitor the ongoing initiatives in data law and regulation. The copyright and the data protection law have been subject to reforms at the EU level with key aspects for data-driven research (European Parliament and of the Council, 2019, 2016). Recently, in 2021, the EU commission presented a bundle of legislation with a data act. For the first time, this legislation explicitly addresses research data of public authorities on the different German state levels that are important data providers in FAIRagro. We analyse the implications of these developments for FAIRagro and release legal updates at the FAIRagro Portal and in newsletters facilitated through TA2 (D3.6.4).

Action 4: Leverage FAIR Digital Objects for machine-actionable policies

Some UCs (Measures 1.1, 1.3 and 1.6) require handling data that are not fully open access and have separate kinds of restrictions on the data. Based on the W3C ODRL (Open Digital Rights Language) Information Model for fundamental content usage involving permissions, prohibitions, and obligations, we will operationalize the legal policy framework and encapsulate this set of policies, rules and actions. Building on common key principles of FDO and FAIR - data and metadata are retrievable by their identifier using a standard communication protocol (A1) - machine-actionable policies provide direct persistent linking to their associated data that should subsequently entail implementation of an FDO extension (D3.6.9). A process will be set up to describe restriction and licensing of data access, application for, and permission to access restricted data. This process will be designed not only to provide services for FAIRagro but also to be suitable and expandable for the whole NFDI. Implementation of the needed services will take part in Measure 3.5 for the FDO and in Measure 4.2 for the middleware.

Milestones and Deliverables

Action	No.	Description	Due end of
1	M3.6.1	Assessment of legal status quo at FAIRagro partners	Q2 2023
	M3.6.2	Workshop for community-feedback to the assessment	Q4 2023
	M3.6.3	Starting a permanent working group for the legal framework	Q4 2023
	D3.6.1	Chapters for data policies and legal guidelines	Q1 2025
	M3.6.4	Dissemination of the recommended policies, guidelines and model agreements	Q4 2026
2	D3.6.2	Clusters of agrosystems data in accordance with relevant legal categories	Q3 2023
	M3.6.5	Workshop to the legal metadata standards	Q2 2024
	D3.6.3	Legal metadata standards	Q2 2025
3	D3.6.4	Starting a newsletter and updates about data regulations	Q3 2023
4	D3.6.5	Implementation of machine-actionable policy framework involving the ODRL information model as a FDO type	Q1 2027

5.1 Task Area 4: Services

Lead: IPK, FZJ, ZB MED

This TA will implement the actions of sharing, finding, integration and publication of the widely accepted research data life cycle model (Latif *et al.*, 2019; Griffin *et al.*, 2017) in a sustainable and FAIR manner. Based on the agrosystem domain, the demands formulated by the UCs (TA1), and the standards defined in TA3, TA4 implements the necessary components and infrastructure services of the federated RDM along a FAIR enabled research data lifecycle, which will be made accessible via the FAIRagro Portal.

TA4 is responsible for the provision and technical operation of the central FAIRago services of the consortium (see Measure 4.1). To address key objectives with respect to the UC requirements, TA4 implements different RDM services in Measures 4.2-4.4. In Measure 4.2, the interlinking of existing data repositories will be realized by implementing a middleware for interoperable data and meta-data access across infrastructures, including a common authentication and authorisation infrastructure (AAI), to assure there are legal constraints on data access. Measure 4.3 provides a searchable inventory of services and data to improve findability. This also includes a service for transparent disclosure of harmonised metrics of technical service quality to ensure optimal reusability of data. A framework for reusable data integration workflows to feature use case specific, cloud enabled analysis workflows will be implemented in Measure 4.4. Along with these technical tasks, TA4 will support the Data Steward Service Center (Measure 2.5) with technical expertise in agrosystem data infrastructures. To implement the *NFDI and FAIRagro Open Source* strategy, software developed in TA4 and other TAs will be made available in a central FAIRagro source code repository provided in Measure 4.1. Furthermore, TA4 is responsible for ensuring the technical compatibility and potential adoption of NFDI’s cross-cutting infrastructure services, as coordinated in NFDI section *Common Infrastructures* (see chap. 3.2 and Measure 5.3).

Measure 4.1: Central Services for the FAIRagro Community

Contributors: ZALF (lead), JKI

Links to Measures 3.2, 4.2, 5.1 and 5.3

Objectives

Measure 4.1 is responsible for the technical backend and operation of central services to support the FAIRagro community and consortium activities. The technical implementation and operation of the FAIRagro Portal will be realized in Action 1, whereas the content and redaction is provided by TA2, which implies close cooperation with the measures for conceptualization, development and operation. Additional central services include a project management platform for collaboration and software development (Action 2), a helpdesk system for user support (Action 3) and a RDMO instance (Action 4).

Action 1: Technical implementation of the FAIRagro Portal

Together with Measure 2.1 and other TAs, we will identify the technical requirements and develop a concept for the FAIRagro Portal (M4.1.1) based on the user requirements outlined in Measure 2.1. The following measures will be involved in the technical conceptualization of the portal: Measure 2.1 for the general concept of information dissemination and interaction of users with the portal; Measure 2.2 and Measure 4.3 for the integration of user feedback and performance indicators; Measure 2.3 for onboarding and technical platform for the submission of new UCs; and Measure 2.4 for the integration of education and training materials. Based on user needs and concepts (Measure 2.1), we will evaluate appropriate software solutions (e.g., Django, Joomla, M4.1.2) and develop a strategy (D4.1.1) to implement and operate the FAIRagro Portal (D4.1.2). The FAIRagro Portal will be a central access point and organize different requests (e.g., searches for education material, user support requests, and searches for relevant data) to different web services developed in the consortium, e.g., the searchable infrastructure and data inventors (Measure 4.3).

Action 2: Project management platform for collaboration and software development

A project management platform will be established, not only to facilitate the control of the project progress (Measure 5.1), but also to provide an exchange platform within the consortium and to facilitate collaboration between partners from different institutes. Appropriate ticket systems such as Redmine, which has been in use at ZALF for 5 years, or the integrated issue management of the project management platform *OpenProject* (2021), as NFDI's widely accepted project management system, will be evaluated (M4.1.4). The exchange platform will also support distributed software development and integration tests within the consortium and the community to produce qualitative *Open Source* software, which can be used and extended by other persons. The management platform will therefore also comprise a source code development environment (e.g., GitLab, D4.1.3) for FAIRagro partners and the agrosystem community facilitating continuous integration (CI), task management and automatic unit and functional tests. In this cross-cutting action, we will collaborate with other software development measures in TA1, TA3 and TA4 to promote and enhance the existing guidelines and best practices in software management, (e.g. Anzt *et al.* (2020); Lamprecht *et al.* (2020), to support research software engineering in the agrosystem sciences and beyond (D4.1.4), therefore contributing to the general Open Source Strategy in the NFDI (Measure 5.3).

Action 3: Helpdesk for User Support

This action sets up a ticket system accessible via the FAIRagro Portal to support the multi-level user support structure of FAIRagro addressed in Measure 2.5 (Data Steward Service Center). The ticket system will be combined with a CRM and integrated with the project management platform (see Action 2) to realize a fully functional helpdesk system for managing, processing and responding to user requests (M4.1.5). Answer templates, documentation, tutorials and training will be provided and continuously updated for the data stewards, who mainly run the helpdesk (D4.1.6).

Action 4: RDMO instance to support DMPs in agrosystem sciences

This action will set up an instance of RDMO as a community service to facilitate the creation of DMPs within the agricultural community. The RDMO instance will be implemented and hosted at ZALF (D4.1.7). Data protection will be ensured by embedding the consortium-wide AAI service (Measure 4.2) into RDMO. The creation of subject-specific questionnaires for the agroscience community focused on the UCs is an action of Measure 3.2. The tailored questionnaires will then be part of consulting for the data stewards (Measure 2.5) and content for training and education (Measure 2.4).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M4.1.1	Technical Requirements of the FAIRagro Portal	Q2 2023
	M4.1.2	Evaluation and Selection of software solution for the FAIRagro Portal	Q4 2023
	D4.1.1	Implementation strategy for the FAIRagro Portal	Q4 2023
	D4.1.2	Updated versions of the FAIRagro Portal according to the implementation strategy (D4.1.1)	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q4 2027
2	M4.1.4	Evaluation and selection of project management systems	Q3 2023
	D4.1.3	Project management platform operational	Q3 2023
	D4.1.4	Report on best-practices on software management in agrosystem sciences	Q2 2026
3	M4.1.5	Ticket system and CRM operational	Q3 2023
	D4.1.6	Documentation and manuals for the use of the helpdesk	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q4 2027
4	D4.1.7	RDMO instance operational	Q3 2023

Measure 4.2: Network of federated research data infrastructures (RDI)

Contributors: IPK (lead), ZALF, SGN, JKI, Thünen, ZB MED, UBN, TUM, FZJ, DWD

Links to Measures 1.1-1.6, 3.1, 3.2, 4.3 and 5.3

Objectives

This measure creates a network of interoperable RDIs that will facilitate the exchange of information and provide access to data to advance the integration of FAIRagro into hyperinfrastructures such as the “One NFDI” (NFDI-RDC) or EOSC. Currently, existing RDIs in agrosystem sciences use different standards, e.g., for metadata annotation or machine-actionable accessibility. We will develop a middleware service based on standards contributed by TA3. This comprises harmonized interfaces to RDI to enable standardized metadata exchange, persistent data access and a continuous recording of technical quality metrics (Action 1). FAIRagro will align the development of middleware implementation with established approaches, e.g., the three-tier architecture pattern (Eckerson, 1995), which provides a structure comprising a presentation tier, a logic tier and a database tier. The database tier connects to particular RDIs and maps the internal data structures and interfaces to the FAIRagro standards developed by TA3. The logic tier implements common tasks, e.g., ensure authentication and authorization and manage the data streams and API calls. The presentation layer is the homogeneous access interface for all client services and delivers well-defined data formats and query interfaces as online services. This general concept will be aligned with the NFDI-RDC, national [Generic Research Data Infrastructure - GeRDI (Latif *et al.*, 2019)], international developments [EOSC (EOSC, 2020), ELIXIR (ELIXIR, 2021), GAIA-X (GAIA-X, 2021)] and new recommendations developed in the RDA. The search and inventory service (Measure 4.3) will be based on the middleware service. Additionally, the middleware will implement a cross-NFDI compliant mechanism for authentication and authorization (Action 3) to support secure and remote data access as required for SciWIn (Measure 4.4). The middleware will be connected to the core RDIs of the FAIRagro network (Table 2 in chap. 4.4, Action 2) and extended to the RDIs of neighbouring consortia (e.g., NFDI4Earth) to provide data across NFDI (Action 4).

Action 1: Design and implementation of the middleware service

In close cooperation with TA3, which develops concepts for harmonizing metadata schemas, data publication and technical service quality standards, we will design the middleware service according to FAIRness and Openness. Initially, we will converge the requirements - guided by common standards as evaluated, compiled and adopted in Measure 3.1 and Measure 3.2 - and design the middleware architecture (M4.2.1). It will be implemented iteratively, starting with a prototype implementation enabling harmonized metadata exchanges and persistent data access, to ensure an early proof-of-concept and to validate the middleware service (D4.2.1). In addition, the functionality is successively extended to provide interfaces for direct data access, as needed by Measure 4.4.

Action 2: Connecting the middleware service to FAIRagro RDIs

In a stepwise process, the middleware service will be connected to the core RDIs (see chap. 4.4, Table 2), which are provided by FAIRagro partners and whose data are required by the UCs (TA1). We will start to connect the first set of infrastructures (D4.2.1, see Table 3 Phase 1) to validate the concept of the middleware and provide a basis for the search and inventory service (Measure 4.3). In an iterative approach, additional infrastructures (D4.2.2, see Table 3 Phase 2) will be integrated to complement the network of interoperable RDI to cover the major domains of the agrosystem community as defined in TA1 and chap 2.1.

After connecting the first core RDIs, a workshop in the agrosystem community and with other NFDI consortia will be organized to promote the FAIRagro repository network. This is necessary to obtain feedback on this service and to support the additional infrastructure of the community to be included in the FAIRagro network of interlinked infrastructures in the third implementation phase (M4.2.3).

Table 3: Implementation phases for interlinking the infrastructures with the middleware service

Phase 1 (2023-2024)	**Phase 2** (2025-2026)	**Phase 3** (2027)
● e!DAL-PGP (IPK) ● BonaRes (ZALF) ● OpenAgrar (JKI) ● National Soil & Forest Inventories (Thünen) ● PUBLISSO Repository for Life Sciences (ZB MED)	● PhenoRob DB (UBN) ● PlabiPD (FZJ) ● Edaphobase (SGN) ● GBIS/I, LIMS (IPK) ● JKI-DataCube (JKI) ● SRADI (TUM) ● BonaRes Knowledge Library (UFZ) ● Open Data Server (DWD)	● Cross-NFDI infrastructures *Additional infrastructures from agrosystem community*

Action 3: Integration of a cross-NFDI authentication and authorization infrastructure

Due to the technical heterogeneity of access control of the FAIRagro repository network, and to enable the application of the specifications of Measure 3.6, it is indispensable to establish uniform single-sign-on AAI users to facilitate accessibility across infrastructures and communities. This infrastructure initiative will be aligned with the national undertakings of the NFDI for providing basic services such as the DFN-AAI, and will consider internationally established concepts such as the ELIXIR AAI or AARC BPA to integrate the most suitable AAI system to connect the FAIRagro infrastructures into the NFDI and international networks (Measure 5.3, D4.2.3). Additionally, API endpoints for the data access process, and an interpreter for the communication protocol via ODRL from Measure 3.6, will be developed.

Action 4: Extension of the middleware service to interlink cross-NFDI infrastructures

This action will integrate FAIRagro infrastructures into the NFDI, facilitating cross-disciplinary FAIR data sharing and establishing an interoperable infrastructure (ONE NFDI). Coordinated with Measure 1.3, we will organize workshops together with NFDI4Earth, NFDI4BioDiversity, DataPLANT and other consortia and work on guidelines for harmonized APIs in an effort towards Research Data Commons aligned with international approaches (M4.2.4). In the third implementation phase (see Table 3, Phase 3), we will extend the middleware to interlink cross-NFDI infrastructures (D4.2.4).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M4.2.1	Initial version of specifications for middleware interfaces available	Q2 2024
	M4.2.2	Prototype implementation of middleware finished	Q4 2024
2	D4.2.1	First set of infrastructures connected to set up the initial network of interlinked infrastructures	Q4 2024
	M4.2.3	Workshop in the agrosystem community to support further expansion of the FAIRagro network	Q3 2025 Q3 2026
	D4.2.2	Expansion of the FAIRagro network by interlinking additional infrastructures in the second implementation cycle	Q3 2026
3	D4.2.3	Integration of cross-NFDI compatible AAI finished	Q1 2026
4	M4.2.4	Definition of recommendations and guidelines to implement interoperable network of RDIs	Q4 2026
	D4.2.4	Cross-NFDI infrastructures connected	Q4 2027

Measure 4.3: Searchable Inventory of Services and Data

Contributors: ZB MED (lead), FZJ, UBN

Links to Measures 3.1, 3.2, 3.3, 3.4 and 4.2

Objectives

Although a multitude of high-quality agronomy-related infrastructures are available in Germany and are supported by the FAIRagro consortium (see Chap. 4.1 and 4.4), there is no central search service and infrastructure registry to easily navigate to an RDI or to find datasets relevant to the agrosystem domain. In addition, currently, there is no orchestrated network of interlinked data among agrosystem RDIs and related resources such as weather information or soil data, and knowledge about these resources vary widely across agricultural disciplines. Measure 4.3 examines the extent to which searches are possible only on (a) the basis of extensive metadata shared by infrastructures about the datasets and (b) core datasets providing environmental information.

For the search and inventory services, early prototypes will be available for testing and usability (Actions 1 and 2). Organized by Measure 2.2, all partners will closely work together with TA1 to gather user feedback based on the UCs for the developed services. Based on user feedback, the services will be subsequently improved. After the prototype release, we regularly release updates with extended functionality and content (Actions 3-4). Furthermore, together with the TA2 concept of thematic workshops and roadshows (Measure 2.3), we will present the services to the FAIRagro community to advertise our services, train potential users and gather further feedback. The final versions, advanced further through user feedback, will be operated by ZB MED beyond the project phase.

Action 1: Inventory of infrastructures

Measure 4.3 will develop an inventory of infrastructures for genotypic and phenotypic crop data. Due to its holistic approach, it will also encompass crop seeds and seedlings, field trial data, soil data, environmental data, field management data, plant phenotyping and crop disease data, as well as FAIRagro software components and available workflows. This inventory will increase the visibility of each service and make them known beyond their immediate area. Measure 4.3 will add search capabilities for infrastructure and services based on metadata describing those services (M4.3.1, D4.3.1).

Action 2: Central Search Service for agrosystem data

The second service provided by Measure 4.3 is a central search service for FAIRagro data. The search will allow users to find datasets based on their metadata information and will list and rank the results in a simple user-friendly way. In the first version of the search platform, searches will be based only on the preservation information in the currently exposed metadata or on text searches in titles and abstracts (M4.3.1 and D4.3.2). The results will reference the underlying database and identifiers, as well as the underlying licence and data access conditions. The search service allows searching across different research foci in agronomy-related research. In particular, searches with geospatial and localization information are of utmost importance for the agrosystem domain. Therefore, the search service will support the representation of geographic objects (e.g., by using CKAN in combination with the geospatial viewer for CKAN resources, which supports the JavaScript libraries OpenLayers and Leaflet for geovisualization) and will be based on software components allowing such interactive, refinable searches (e.g., interactive maps, faceted clustering). The metadata information for datasets will be harvested or uploaded from the various infrastructures via interfaces provided by the middleware (Measure 4.2) and incorporated within the metadata repository of the search infrastructure. In addition to the API, a web-based user interface with templates, and the opportunity for batch uploads, will allow small repositories to integrate into the search service and enable findability of these data by referring to the small repository.

Action 3: Advancing search functionality for agricultural data

In an advanced version, we intend to integrate German weather records [e.g., (Fronzek *et al.*, 2018)] and soil information [e.g., (Kristensen *et al.*, 2019)] from published datasets within the search to enable search and result visualizations depending on local events or environmental information such as weather or soil. In addition to geospatial searches, and based on the metadata models provided by Measure 3.2, semantic searches will also be possible from various entry points, e.g., starting with a search of field trial information, genotypic or phenotypic features or disease information (M4.3.3). A first release of the advanced version will be ready in Q2 2027 (D4.3.3). In the advanced version of the inventory service, it will also add (1) information on the quality and availability of services based on objective metrics [e.g., service downtime and/or “currently service unavailable” and response times] and (2) service summaries [e.g., number of records, size of data, and special visualization capabilities] that can be filtered on. Additionally, simple quantitative indicators related to the number of datasets or sizes will be added to allow users to prioritize infrastructures.

Action 4: Integration of user relevance measures

Additional features to be integrated are new ranking functions and user-related functions for both services. The ranking functions for the central search platform will take also qualitative, quantitative and FAIR metrics criteria (Measure 3.3, 3.4) into account and will be evaluated by test users. Furthermore, we will evaluate existing recommender systems and will build an interface to one, if the functionalities meet FAIRagro requirements. Such services can draw on stored search histories and navigation behavior to make other dataset suggestions (e.g. Esch *et al.*, 2015). All functions will be tested in early prototype service versions and integrated into the final service version released in Q1 2028 (D4.3.4).

Milestones and Deliverables

Action	No.	Description	Due end of
1	D4.3.1	Availability of a first basic inventory service	Q1 2024
1, 2	M4.3.1	First FAIRagro inventory and search service version publicly available	Q2 2025
2	D4.3.2	Prototype search service for datasets with basic visualization features for spatial data is publicly available	Q2 2026
3	M4.3.3	Advanced search functionalities are included into the services	Q4 2026
	D4.3.3	Advanced versions of FAIRagro inventory and search service are released and public	Q2 2027
4	M4.3.4	Final FAIRagro inventory and search service available	Q1 2028
1-4	D4.3.4	Final and improved inventory service including result visualization and search service reflecting latest middleware components	Q1 2028

Measure 4.4: Scientific Workflow Infrastructure (SciWIn)

Contributors: Thünen (lead), IPK, BLU

Links to Measures 1.2, 1.5, 1.6, 2.3, 2.4, 3.5, 4.2 and 5.3

Objectives

Storing analysis or data integration workflows in a FAIR manner is a cumbersome process involving the use of different services, stakeholders and platforms (Yenni *et al.*, 2019). This is inefficient and hinders reproducible science. Furthermore, algorithms running on these integrated datasets use libraries that will obtain updates and fixes, which could change the research outcomes and render the code unreproducible. Thus, researchers are unable to publish and collaborate on their work done in a FAIR way. The concept of data integration is partly described in the RDC mediation layer and covered by semantic tools (Glöckner *et al.*, 2020), and a concept for integrated data and process storage is part of the DataPLANTs ARC model (Krantz *et al.*, 2021), but a fully integrated infrastructure is provided by neither of those consortia. Therefore, Measure 4.4 will extend the two architectural designs and provide a workflow infrastructure that applies the FAIR DO concepts (Measure 3.5), the service middleware components (Measure 4.2) and its own workflow hub as an easy-to-use interface to work on and create new FAIR DO outputs with automatically annotated provenance graphs.

As illustrated in Figure 13, SciWIn consists of five components. The *AAI* (A) authorizes the user (D4.1.3). The *Workflow Hub* (B) will enable the creation of *Workflow Objects* (E) based upon the *Workflow RO-Crate* schema (Soiland-Reyes *et al.*, 2021) and provides storage for those objects. The *Compute Instances* (C) will provide computing and runtime capabilities, as well as metainformation about these parameters. *Storage Instance*s (D) will hold RDC-compliant *FAIR DO*s, and provide APIs to load those into the *Compute Instances*.

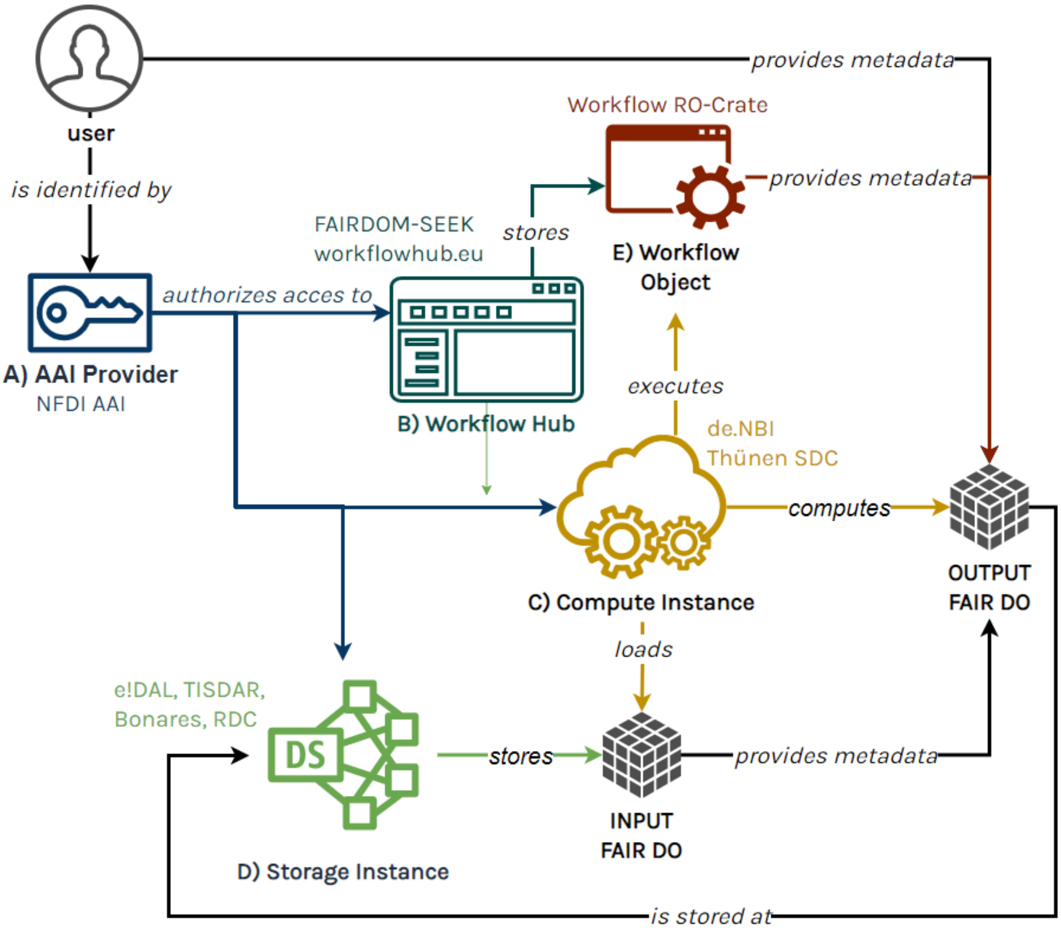

Figure 16: Overview of Scientific Workflow Infrastructure components and an example workflow process. For acronyms and abbreviations see pages IV – VIII at the beginning of this proposal.

To use the SciWIn, an authorized scientist can choose an already provided workflow or write a new workflow to apply to a *FAIR DO* within a *Compute Instance*. This will create a new *FAIR DO* as output, with prefilled provenance graph information and metadata that will be fetched from the user, the input FAIR DO and the *Workflow Object*. This infrastructure will allow the connection of storage, computational instances and versioning and publishing of scientific workflows such as data integration algorithms or complex data analysis pipelines in a user-friendly, efficient, secure, scalable way. To achieve this goal, the following actions will be performed:

Action 1: Develop a joint concept for SciWIn

SciWIn will be part of the NFDI cross-cutting topic “RDC implementations”, and therefore, its concept will be developed as a joint effort between FAIRagro, FAIR-DS, Dataplant and the NFDI section, RDC [as part of cross-cutting topics (Ebert *et al.*, 2021)]. Coordinated in Measure 5.3, the SciWIn working group brings these stakeholders together and will be initiated by a kickoff meeting (M4.4.1). Thus, the fundamental principles, architectures and interfaces are described, and a coordinated concept will be created and published (D4.4.1), which will incorporate the RDC ideas and DataPLANTs ARC model.

Action 2: Evaluate compute, storage and hub instances for SciWIn

Based on D4.4.1, a set of compute and storage instances will be evaluated for their fitness for use as SciWIn components. As *Compute Instances*, the Thünen Secure-Data-Cluster (SDC) and the de.NBI cloud node (BLU, Alexander Sczyrba) will be evaluated as an on-premise and cloud service solution, respectively (M4.4.2). As storage instances e!DAL-PGP, the BonaRes Repository, TISDAR and the RDC will be evaluated (M4.4.3). As a *Workflow Object* model and storage solution *Workflow RO-Crate* and *FAIRDOM-SEEK* (2021) will be evaluated (M4.4.4). Based on these findings, a development roadmap for the SciWIn components will be published (D4.4.2).

Action 3: Provide SciWIn as proof-of-concept

According to the development roadmap (D4.4.2), the development of the SciWIn will be initiated. As the first step, the pilot components for storage and computational instances will be prepared to work within the SciWIn (M4.4.5). After these services are prepared, the development of the Workflow Hub will start, and lead to pilot deployments in the cloud and on-premise (D4.4.3). The UCs (Measures 1.1, 1.2, 1.6) will provide example workflows that will be ported to the SciWIn and can be used for test cases.

Action 4: Foster community engagement through training material, workshops and public courses on reproducible science with SciWIn

Together with TA2, we will provide training material and courses that will lead to a workshop on reproducible science using SciWIn (D4.4.4). The workshop will teach the attendees how to use SciWIn and evaluate its current status. This includes a stress test of the running infrastructure and a review of further potential deployment options on the computing resources available in NFDI and ELIXIR at that time. As part of the workshop proceedings, this information will be published to show where the platform performs well and where there are shortcomings (D4.4.5).

Milestones and Deliverables

Action	No.	Description	Due end of
1	M4.4.1	Kick-Off meeting with RDC working group	Q1 2024
	D4.4.1	Joint concept of SciWIn as part of the RDC semantic toolset	Q3 2024
2	M4.4.2	Compute instances are evaluated	Q1 2025
	M4.4.3	Storage instances are evaluated	Q2 2025
	M4.4.4	Workflow hub and models are evaluated	Q3 2025
	D4.4.2	SciWIn development roadmap	Q4 2025
3	M4.4.5	SciWIn components are ready to use	Q2 2026
	D4.4.3	On-premise and cloud pilots of SciWIn	Q4 2026
4	D4.4.4	Workshop on reproducible science with SciWIn	Q2 2027
	D4.4.5	SciWIn evaluation report	Q4 2027

5.4 Task Area 5: Management and Coordination

Lead: ZALF

TA5 will establish a sustainable and transparent governance and operational structure for FAIRagro and is responsible for the coordination and management of the consortium. TA5 will initiate and coordinate actions to sustain the achievements of FAIRagro. Initially, TA5 will establish all functional bodies of the project’s governance (see chap 3.4, Figure 8) embedded in the NFDI organisation structure. It will organise the day-to-day business of FAIRagro, manage the consortium’s finances, and ensure compliance with funding conditions. TA5 is in charge of organising periodic reportsand FAIRagro plenary events (Measure 5.1). Measure 5.2 describes the actions taken to develop FAIRagro in a sustainable, long-lasting institution beyond the NFDI funding phase. Measure 5.3 will coordinate communication and relations between FAIRagro, the NFDI and relevant national and international RDM initiatives.

Measure 5.1: Project Management, Governance and Financial Controlling

Contributors: ZALF (lead)

Links to Measure 2.2

Objectives

We will establish the **FAIRagro Secretariat** at ZALF to guarantee a smooth operation of the FAIRagro consortium and provide support to achieve the consortium goals. The secretariat will set up governance mechanisms and define the interaction within and between the FAIRagro functional bodies, members and the NFDI (Action 1). The secretariat will administer FAIRagro operations to support common governance, manage day-to-day business, control finances and control project progress (Action 2). The FAIRagro Secretariat will be in charge of coordinating and controlling the finances within the consortium - implementing DFG rules and procedures in FAIRagro (Action 3).

Action 1: Set-up and support of the project’s governance

In preparation of this proposal, the FAIRagro consortium drafted and agreed on an organisational structure (see chap. 3.4) that will be established in the early project phase by organising kick-off meetings for the project’s bodies (M5.1.1, M5.1.2). The process for establishing the Community Advisory Board (CAB) will be organized by TA2 (Measure 2.2), as this will be an element of TA2's community activities. In accordance with DFG rules, this governance structure will be further detailed in the FAIRagro consortium agreement (CA, Mittelweiterleitungs- und Kooperationsvertrag), which must be acknowledged by the applicant, all co-applicants and participants (D5.1.1). The CA includes definitions of the members’ roles, rights, duties and responsibilities. The FAIRagro Secretariat supports functional bodies by organising regular face-to-face meetings of the SC (D5.1.2) and the FAIRagro Plenary (D5.1.3) and regular telephone/video conferences of the SC.

Action 2: Establishing and running the FAIRagro Secretariat

The FAIRagro Secretariat (M5.1.3) manages the FAIRagro consortium in its day-to-day business, including the monitoring of project progress. It will be operated at ZALF, which has committed to supporting the Secretariat by providing office space. The project administrative coordinator heads the FAIRagro Secretariat and deputizes the spokesperson. The project operational manager is responsible for the day-to-day management of FAIRagro and supports the co-spokespersons and functional bodies of the initiative. The project progress will be overseen – with the help of the task area leads – by the preparation of annual reports (D5.1.4) that will integrate information on the progress of use cases, status on success measures (see chap. 2.2), performance indicators and finances. The secretariat provides on-demand communication for conflict management.

Action 3: Financial controlling

The FAIRagro Secretariat will be managing and monitoring the use of funds, including the financial transactions of the grants from the DFG to the partners according to the work program in this proposal and the rules as specified in the DFG's contract. Since funds will be provided to the applicant institution (ZALF), they will be distributed to the co-applicant institutions based on their cost statements. ZALF will establish the financial control of the project (D5.1.5) to facilitate fund distributions to the co-applicants and serve as a central contact point for financial issues. ZALF will also apply for all funds intended for participants and forward them accordingly. ZALF will organise the CA (D5.1.1), which describes the distribution of funds between co-applicants and participants. When the funds are forwarded, the approval conditions are also forwarded to the co-applicants, who thus assume responsibility for the proper use of their funds. The FAIRagro Secretariat will work directly with the administrative staff at each co-applicant institution to organise money transfers and obtain financial numbers on the funding spent. Regular financial reports (D5.1.4) will be created, including justification of resources and amendments.

Milestones and Deliverables

Action	No.	Description	Due end of
1	D5.1.1	FAIRagro Consortium Agreement according to DFG rules	Q2 2023
	M5.1.1	Kick-off meeting for SC	Q3 2023
	M5.1.2	Governance structure established	Q2 2023
	D5.1.2	Annual meeting of the SC	Q2 2024 Q2 2025 Q2 2026 Q2 2027
	D5.1.3	Annual FAIRagro Plenary meeting	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q4 2027

Action	No.	Description	Due end of
2	M5.1.3	FAIRagro Secretariat established and running	Q2 2023
	D5.1.4	FAIRagro annual reports	Q1 2024 Q1 2025 Q1 2026 Q1 2027 Q1 2028
3	D5.1.5	Financial controlling and money transfer established (Month 3)	Q2 2023

Measure 5.2: Sustainability and Business Model

Contributors: ZALF

Links to all other measures

Objectives

FAIRagro is an essential prerequisite in the development of an NFDI for the agricultural sciences. To establish lasting effects, a change of the culture within research communities is needed. Therefore, FAIRagro needs to be sustainable and self-sufficient beyond the funding period. FAIRagro must develop and implement a suitable business model to become sustainable in the long-term in order to preserve the developments made within the funding period.

Action 1: Establishing a FAIRagro as a long-lasting institution

To ensure the long term existence of the FAIRagro infrastructures and the progress made within the FAIRagro project for the agrosystem and agricultural community, the participating institutions will establish a non-profit association (gemeinnütziger Verein) (M5.2.1). The association will serve as a platform for further activities and to secure a basic infrastructure (see chap 4.4). The members of the FAIRagro association will provide infrastructures, and FAIRagro will explore possibilities to sustainably finance the services after the funding period (see Actions 3 and 4). The association will become a member of the NFDI e.V. and serve as the voice of the agrosystem community within NFDI (see Measure 5.3). It will be open to all interested actors in the agrosystem domain (and later to actors from other domains and disciplines of the agricultural sciences), who are willing to support the establishment of an NFDI and are supportive of raising awareness of the importance of a professional research data culture in the field of agrosystem research. The FAIRagro association will be a platform to host all activities (see Actions 2-4) within FAIRagro and will be a platform to open new research areas in the agricultural domain (D5.2.1).

Action 2: Cultural change in Agrosystem Research Data Management (RDM)

FAIRagro provides user-friendly RDM infrastructures, meaningful support services and educational activities to advance cultural change in the agrosystem community. To secure this development beyond the funding period, FAIRagro will, together with institutional stakeholders such as academic societies, explore the possibilities of establishing communication formats and address change processes. Building on the work of TA2, joint activities will be established to discuss the demands in the communities and offer training and support for institutional stakeholders adopting FAIR and open data principles. Part of this action is to actively exchange with other NFDI consortia (see Measure 5.3) about experiences and best practices for cultural change and how to support stakeholders in this change process (D5.2.3) by developing a scheme for support and advisory actions (D5.2.2)

Action 3: Securing cultural change via training and services

FAIRagro will explore the different possibilities of turning FAIRagro services into business models to maintain and sustain the FAIRagro project for the long term. Here, we focus on two main services:

(1) To ensure a lasting change initiated by FAIRagro, the project will establish several training courses and modules addressing researchers (see Measure 2.4). This training curriculum will be developed to continue further beyond the project funding period. Offering capacity building, continuous training and train-the-trainer courses are important measures to guarantee changes in the research culture. Training could be a possible option for a self-funding service. FAIRagro will evaluate the possibilities of establishing training measurements as a self-sufficient action (M5.2.2) and start this service as a spin off (D5.2.4).

(2) The service provided by the data stewards (Measure 2.5) is a cornerstone for establishing and maintaining a new research culture of FAIR research data. FAIRagro will explore whether this service could be established as a spin-off to support research projects without an institutionally funded RDM service. Pursuing this, the service and the demands and a market analysis (M5.2.3) will be evaluated, and the data steward service will be established as a self-sufficient data steward consultation service (D5.2.5).

Action 4: Towards dual-usability data

Data - research and commercial data - are important resources in the field of agriculture. In the current situation, collections of data are key for large-scale simulations and research. Simultaneously, to the scientific data collections, commercial data are gathered in the agricultural sector every day. Currently, commercial data are not freely accessible for farmers due to the lack of standards, capabilities and infrastructure and not usable for research due to legal uncertainties. This deadlock situation offers an opportunity for research to help farmers and to open commercial data for sciences. FAIRagro will explore the possibilities for a dual-usability model for commercial data. Building on the experiences from the work in the GAIA-X initiative, FAIRagro aims to make scientific, commercial and administrative data available and usable (M5.2.4). In the second step, FAIRagro will explore ways of data fiduciaries and examine whether this fiduciary service could turn into a spin off and a commercial activity for financing the infrastructure of FAIRagro (D5.2.6). The aim of these considerations is to make all kinds of different agricultural data available for research accessible.

Milestones and Deliverables

Action	No.	Description	Due end of
1	M5.2.1	Establishing a non-profit association	Q1 2027
	D5.2.1	FAIRagro association becomes a base for prospective activities	Q4 2027
2	D5.2.2	Support and consultation scheme developed	Q1 2026
	D5.2.3	Establishing best practice modells	Q4 2026
3	M5.2.2	Market analysis for training measurements	Q4 2026
	M5.2.3	Evaluation and market analysis for data stewards consultation service	Q4 2026
	D5.2.4	Establishing training service as a spin off	Q1 2028
	D5.2.5	Establishing data steward consultations service as a spin off	Q1 2028
4	M5.2.4	Analysis completed & concept developed	Q4 2025
	D5.2.6	Business plan for a fiduciary service & establishing a spin off	Q2 2027

Measure 5.3: Cross-NFDI and international networking

Contributors: ZALF (lead)

Links to Measures 2.4, 2.5, 3.6, 4.2, 4.4

Objectives

The overall goal of this measure is to coordinate communication, relations and collaborations between FAIRagro, other NFDI consortia, the NFDI and relevant international RDM initiatives (see chap. 3.2 and 3.3). We will contribute to a common vision for and integrate FAIRagro in the NFDI to facilitate cross-disciplinary FAIR data sharing. We will cross-link FAIRagro infrastructures and services in terms of the NFDI-RDCs and actively contribute to cross-cutting topics (Ebert *et al.*, 2021; Bierwirth *et al.*, 2020; Glöckner *et al.*, 2019) by actively engaging in NFDI sections *Training & Education, Ethical, Legal & Social Aspects, Common Infrastructures*, and (*Meta)data, Terminology and Provenance* and directly collaborating with other NFDI consortia.

Action 1: Contributing to NFDI cross-cutting topics and cooperation with other NFDI consortia

We will organize and participate in workshops focusing on cross-cutting topics to establish an interoperable RDM infrastructure in Germany in the vision of One NFDI (D5.3.1).

Section Training & Education:

Education and training for agrosystem researchers is a key objective in FAIRagro addressed in Measure 2.4. As this is a cross-cutting topic already acknowledged in the NFDI by the initiation of the *Training & Education* section, we will actively work in the section we are already involved in. We will contribute domain-specific perspective and content of the agrosystem domain to the material created in this NFDI section). Furthermore, training options in cooperation with NFDI4BioDiversity (U Bremen Research Alliance), NFDI4Earth (EduHub of NFDI4Earth) and de.NBI are integrated into the training program of FAIRagro (Measure 2.4). FAIRagro will contribute to the education of data stewards with partners from NFDI4Earth by participation in the planned **NFDI4Earth Academy** (M 5.3.2).

Concepts for providing direct support by establishing data stewards can be found in different consortia, e.g., DataPLANT, NFDI4Earth or NFDI4Health. We will organize workshops and support the exchange of data stewards from different disciplines to advance the networking of data stewards, facilitate experience exchange and to derive common guidelines and a **cross-NFDI strategy for a networked data steward support** structure (together with Measure 2.5). Essential components for supporting the community in RDM issues are help desks, which are currently being established by many NFDI consortia for their respective disciplines. We will work on **interlinking existing help desks to facilitate cross-disciplinary support requests**, which will be an added value for the NFDI. We will demonstrate exemplary networking of help desks from our neighbouring consortia NFDI4Earth and NFDI4BioDiversity and develop a generic concept (D5.3.2) that can be adopted by other NFDI consortia. We will engage with the NFDI Directorate to establish the vision of a **unified community support structure of the NFDI** - including the aspects of a data steward network and interlinked help desks - as a relevant cross-cutting topic to be addressed in a later NFDI section.

Section (Meta)data, Terminologies and Provenance:
Of particular importance for cross-domain collaboration is the sharing and reuse of semantic artefacts such as controlled vocabularies, thesauri and ontologies plus the creation of **cross-functional mappings and alignments between these semantic artefacts** (Measure 3.1). In this context, close cooperation is planned with NFDI4Earth's task area *2interoperate* as well as further collaborative efforts within the NFDI section *(Meta)Data, Terminologies and Provenance* with respect to the FAIR Digital Object approach (Measure 3.5).

Section Ethical, Legal & Social Aspects:
In agrosystem research, uncertainty in the handling and reuse of sensitive data is one of the key challenges addressed in FAIRagro (Measure 3.6). Therefore, we will also be active in the NFDI section *Ethical, Legal & Social Aspects*. We will develop guidelines in collaboration with NFDIxCS addressing data licensing issues and data policies in relation to agricultural data and implement technical solutions to grant access to restricted data (Measure 4.2). The developed concepts and operationalizations will contribute to the whole NFDI.

Section Common Infrastructures:
FAIRagro also supports and adheres to the NFDI-RDCs (Bierwirth *et al.*, 2020; Glöckner *et al.*, 2020) and will therefore actively engage in the NFDI section *Common Infrastructures*. FAIRagro will comply with the NFDI-RDC in setting up the network of interlinked research infrastructures (Measure 4.2) and the implementation of SciWIn (Measure 4.4). We will **align our concept for interlinking disciplinary research infrastructures with RDC**s and connect infrastructures from the agrosystem research domain with other domains (Measure 4.2). Additionally, we will implement **NFDI recommendations for an AAI** system to integrate the FAIRagro infrastructures into the NFDI but also into international networks. The concept of SciWIn (Measure 4.4) will be jointly developed under the umbrella of the NFDI-RDC with partners from FAIR-Data Sharing, DataPLANT and the NFDI section *Common Infrastructures*. We will bring

together different stakeholders by organising a workshop to develop a concept describing principles, architectures and interfaces of a data integration platform incorporating the RDC ideas and DataPLANTs ARC model. Furthermore, together with NFDIxCS, we will develop methods and implement container technologies to make them interoperable (Measure 4.4).

Action 2: Promote and integrate FAIRagro in international initiatives

The focus of this action is to ensure and promote international awareness of FAIRagro and to consider/adhere to novel developments of RDM by connecting and actively participating in international initiatives. We will monitor the concepts of other RDM stakeholders and communicate new developments to the FAIRagro TAs or Measures. Additionally, we will liaise with research institutes, libraries, governmental institutions, IT centers and related networks interested in FAIR RDM in agrosystem sciences.

Here, we will crosslink with TA3 concerning the international embedding and development of regulations and standards. **We will contribute to improved harmonisation of RDM and standards on the international level by the adopting the FAIR standards.** FAIRagro will actively participate in and initiate **RDA** work groups on cross-cutting RDM topics in collaboration with other NFDI initiatives. We will provide an overview of relevant networks (D5.3.4) and initiatives and will use existing collaborations and memberships to connect and promote FAIRagro ideas to the international community (D5.3.5).

The **European Open Science Cloud (EOSC)** is the most prominent activity building an RDI and an umbrella for the activities in the different research communities and EU member states. FAIRagro is already a stakeholder and will be connected to EOSC as an observer (M5.3.2), as well as by its different members participating in the workgroups of EOSC. With these activities, FAIRagro will ensure the information flow and coordination of activities between the agrosystem domain in Germany and Europe. Furthermore, FAIRagro participates in the development of FAIR Digital Objects (see Measure 3.5) as a possibility to transfer FAIRagro data into EOSC data spaces. The **GAIA-X** initiative is an important approach for defining data standards in industries and research. ZALF - like other members of FAIRagro - is already an active member of the GAIA-X agriculture domain. FAIRagro will build on the existing connections and closely monitor the developments within GAIA-X (see Measure 5.2, M5.3.3).

In addition to the connections to other modelling networks, **GLASSNET** is of special interest for FAIRagro, as GLASSNET will bring data producers, data curators and data users in the field of sustainable land use together by connecting existing networks and communities. FAIRagro will become a member of GLASSNET and will use GLASSNET as a hub to connect FAIRagro to the international community. FAIRagro will connect its services to the European Soil Data Centre (**ESDAC**) and the upcoming **EU Soil Repository** (within the Horizon Europe Soil Mission) (M5.3.4). FAIRagro will become - such as ZALF - a registered stakeholder at the European Strategy Forum on Research Infrastructures (ESFRI). The project will use the **ESFRI**

platform to connect to other RIs in the agrosystem domain and represent the positions and demands of the German community (D5.3.3).

Milestones and Deliverables

Action	No.	Description	Due end of
1	D5.3.1	Establish and participate in NFDI networks and workshops to continuously work, update and expand on cross-cutting topics; results will be regularly reported.	Q4 2023 Q4 2024 Q4 2025 Q4 2026 Q4 2027
	M5.3.1	Contribute agrosystem-domain specific content to the NFDI section *Training & Education*	Q4 2025
	D5.3.2	Concept for interlinking help desks developed	Q4 2026
2	M5.3.2	FAIRagro becomes an observer in EOSC	Q4 2025
	M5.3.3	Interlink FAIRagro with GAIA-X	Q4 2025
	M5.3.4	Connecting FAIRagro to the EU Soil Repository	Q4 2025
	D5.3.3	FAIRagro acts as a proxy for the German community within ESFRI	Q4 2025
	D5.3.4	Overview of relevant RDM networks and initiatives in which FAIRagro members are active, that will be published on the FAIRagro Portal	Q4 2024
	D5.3.5	Connecting FAIRagro community with existing international networks and stakeholders	Q4 2025

6 Appendix

6.1 Bibliography and list of references

A1, 2021. FAIR principles: A1 - Meta)data are retrievable by their identifier using a standardised communication protocol, https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/

AGMEMOD, 2021. Agricultural Member State Modelling, https://agmemod.eu/

AGMIP, 2021a. AgMIP Data Interoperability Group (AgDIG). https://agmip.org/agdig-data-interoperability-group/

AgMIP, 2021b. AgMIP Website, https://agmip.org/

Alercia, A., Diulgheroff, S., Mackay, M., 2015. FAO/Bioversity Multi-Crop Passport Descriptors V.2.1, p. 11 p., https://www.bioversityinternational.org/index.php?id=244&tx_news_pi1%5Bnews%5D=7639&cHash=6e87580466bd8236a004bb0a458177b3

Anzt, H., Bach, F., Druskat, S., Loffler, F., Loewe, A., Renard, B.Y., Seemann, G., Struck, A., Achhammer, E., Aggarwal, P., Appel, F., Bader, M., Brusch, L., Busse, C., Chourdakis, G., Dabrowski, P.W., Ebert, P., Flemisch, B., Friedl, S., Fritzsch, B., Funk, M.D., Gast, V., Goth, F., Grad, J.N., Hegewald, J., Hermann, S., Hohmann, F., Janosch, S., Kutra, D., Linxweiler, J., Muth, T., Peters-Kottig, W., Rack, F., Raters, F.H.C., Rave, S., Reina, G., Reissig, M., Ropinski, T., Schaarschmidt, J., Seibold, H., Thiele, J.P., Uekermann, B., Unger, S., Weeber, R., 2020. An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action. F1000Res 9, 295.10.12688/f1000research.23224.2

Arend, D., Colmsee, C., Knüpffer, H., Oppermann, M., Scholz, U., Schüler, D., Weise, S., Lange, M., 2014. Data Management Experiences and Best Practices from the Perspective of a Plant Research Institute. 8574, 41-49.10.1007/978-3-319-08590-6_4

Arend, D., Junker, A., Scholz, U., Schuler, D., Wylie, J., Lange, M., 2016. PGP repository: a plant phenomics and genomics data publication infrastructure. Database (Oxford) 2016.https://doi.org/10.1093/database/baw033

Arend, D., Konig, P., Junker, A., Scholz, U., Lange, M., 2020. The on-premise data sharing infrastructure e!DAL: Foster FAIR data for faster data acquisition. GigaScience 9.10.1093/gigascience/giaa107

Arnaud, E., Cooper, L., Shrestha, R., Menda, N., Nelson, R., Matteis, L., Skofic, M., Bastow, R., Jaiswal, P., Mueller, L., McLaren, G., 2012. Towards a Reference Plant Trait Ontology for Modeling Knowledge of Plant Traits and Phenotypes, International Conference on Knowledge Engineering and Ontology Development, pp. 220-225.10.5220/0004138302200225.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2014. Rising temperatures reduce global wheat production. Nature Climate Change 5, 143-147.10.1038/nclimate2470

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nature Climate Change 3, 827-832.10.1038/nclimate1916

ATB Potsdam, 2021. Leibniz-Innovationshof entsteht als „Schaufenster der Bioökonomie" mit rund 20 Partnern in Groß Kreutz, Brandenburg, In: Öffentlichkeitsarbeit, P.-u. (Ed.), https://www.atb-potsdam.de/de/aktuelles-und-presse/pressemitteilungen/pressemitteilungen-detailseite/leibniz-innovationshof-entsteht-als-schaufenster-der-biooekonomie-mit-rund-20-partnern-in-gross-kreutz-brandenburg

Axehill, J.W., Herzog, E., Tingström, J., Bengtsson, M., 2021. From Brownfield to Greenfield Development – Understanding and Managing the Transition. INCOSE International Symposium 31, 832-847.10.1002/j.2334-5837.2021.00873.x

Barzman, M., Bàrberi, P., Birch, A.N.E., Boonekamp, P., Dachbrodt-Saaydeh, S., Graf, B., Hommel, B., Jensen, J.E., Kiss, J., Kudsk, P., Lamichhane, J.R., Messéan, A., Moonen, A.-C., Ratnadass, A., Ricci, P., Sarah, J.-L., Sattin, M., 2015. Eight principles of integrated pest management. Agronomy for Sustainable Development 35, 1199-1215.10.1007/s13593-015-0327-9

Baumann, P., Misev, D., Merticariu, V., Huu, B.P., 2021. Array databases: concepts, standards, implementations. Journal of Big Data 8.10.1186/s40537-020-00399-2

Biernacka, K., Buchholz, P., Danker, S.A., Dolzycka, D., Engelhardt, C., Helbig, K., Jacob, J., Neumann, J., Odebrecht, C., Wiljes, C., Wuttke, U., 2020. Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement.10.5281/zenodo.4322849

Bierwirth, M., Glöckner, F.O., Grimm, C., Schimmler, S., Boehm, F., Busse, C., Degkwitz, A., Koepler, O., Neuroth, H., 2020. Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung.10.5281/zenodo.3895209.

BonaRes, 2021a. BonaRes Project Website, https://www.bonares.de

BonaRes, 2021b. Knowledge Library. https://klibrary.bonares.de

BonaRes, 2021c. Overview of Long Term Experiments, https://lte.bonares.de/

BonaRes Repository, 2021. https://maps.bonares.de/mapapps/resources/apps/bonares

Bostick, W.M., Koo, J., Walen, V.K., Jones, J.W., Hoogenboom, G., 2004. A Web-Based Data Exchange System for Crop Model Applications. Agronomy Journal 96, 853.10.2134/agronj2004.0853

BrAPI, 2020. Breeding API (V1.3), https://brapi.docs.apiary.io

Broeder, D., Budroni, P., Degl'Innocenti, E., Le Franc, Y., Hugo, W., Jeffery, K., Weiland, C., Wittenburg, P., Zwolf, C.M., 2021. SEMAF: A Proposal for a Flexible Semantic Mapping Framework, Zenode.10.5281/zenodo.4651421.

Burkhardt, U., Russell, D.J., Decker, P., Döhler, M., Höfer, H., Lesch, S., Rick, S., Römbke, J., Trog, C., Vorwald, J., Wurst, E., Xylander, W.E.R., 2014. The Edaphobase project of GBIF-Germany—A new online soil-zoological data warehouse. Applied Soil Ecology 83, 3-12.10.1016/j.apsoil.2014.03.021

Buttigieg, P.L., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E., Consortium, E., 2013. The environment ontology: contextualising biological and biomedical entities. Journal of Biomedical Semantics 4, 43.10.1186/2041-1480-4-43

Buttigieg, P.L., Pafilis, E., Lewis, S.E., Schildhauer, M.P., Walls, R.L., Mungall, C.J., 2016. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. Journal of Biomedical Semantics 7, 57.10.1186/s13326-016-0097-6

Carré, F., McBratney, A.B., Minasny, B., 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141, 1-14.10.1016/j.geoderma.2007.01.018

Chen, J., Scholz, U., Zhou, R., Lange, M., 2018. LAILAPS-QSM: A RESTful API and JAVA library for semantic query suggestions. PLoS Comput Biol 14, e1006058.10.1371/journal.pcbi.1006058

Chenu, K., Porter, J.R., Martre, P., Basso, B., Chapman, S.C., Ewert, F., Bindi, M., Asseng, S., 2017. Contribution of Crop Models to Adaptation in Wheat. Trends Plant Sci 22, 472-490.10.1016/j.tplants.2017.02.003

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., Mietchen, D., Petrauskaitė, R., Wittenburg, P., 2018. Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. Luxembourg : Publications Office of the European Union.doi:https://doi.org/10.2777/54599

Cordra, 2021. Highly configurable software for managing digital objects at scale. https://www.cordra.org/

Cwiek-Kupczynska, H., Altmann, T., Arend, D., Arnaud, E., Chen, D., Cornut, G., Fiorani, F., Frohmberg, W., Junker, A., Klukas, C., Lange, M., Mazurek, C., Nafissi, A., Neveu, P., van Oeveren, J., Pommier, C., Poorter, H., Rocca-Serra, P., Sansone, S.A., Scholz, U., van Schriek, M., Seren, U., Usadel, B., Weise, S., Kersey, P., Krajewski, P., 2016. Measures for interoperability of phenotypic data: minimum information requirements and formatting. Plant Methods 12, 44.10.1186/s13007-016-0144-4

CWL, 2021. Common Workflow Language. https://www.commonwl.org/

DAKIS, 2021. DAKIS Website, https://www.agrarsysteme-der-zukunft.de/konsortien/dakis

DataCite Metadata Working Group, 2017. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1. DataCite e.V.http://doi.org/10.5438/0014

DCAT, 2020. https://www.w3.org/TR/vocab-dcat-2/

De Smedt, K., Koureas, D., Wittenburg, P., 2020. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications 8, 21, https://www.mdpi.com/2304-6775/8/2/21

Denef, K., Zotarelli, L., Boddey, R.M., Six, J., 2007. Microaggregate-associated carbon as a diagnostic fraction for management-induced changes in soil organic carbon in two Oxisols. Soil Biology and Biochemistry 39, 1165-1172.10.1016/j.soilbio.2006.12.024

DFG, 2021. DFG Funding Programm for Specialised Information Services, https://www.dfg.de/en/research_funding/programmes/infrastructure/lis/funding_opportunities/specialised_info_services/index.html

Di, L., Yue, P., Ramapriyan, H.K., King, R.L., 2013. Geoscience Data Provenance: An Overview. IEEE Transactions on Geoscience and Remote Sensing 51, 5065-5072.10.1109/tgrs.2013.2242478

DiSSCoTech, 2020. Fundamentals of Digital Specimen Architecture, Technical posts about the design of the DiSSCo infrastructure, https://dissco.tech/2020/03/06/fundamentals-of-digital-specimen-architecture-dsarch/

DONA, 2021. Digital Object Interface Protocol Version 1.0. https://www.dona.net/doipv1doc

Dublin Core, 2020. https://www.dublincore.org/

DWD, 2021. https://www.dwd.de/EN/ourservices/opendata/opendata.html

Ebert, B., Fluck, J., Glöckner, F.O., Koepler, O., Miller, B., Schmitt, R., Schrade, T., Stegle, O., Steinbeck, C., von Suchodoletz, D., Wagemann, K., Knebes, J., Kraft, S., Seitz-Moskaliuk, H., Sure-Vetter, Y., Wössner, E., 2021. NFDI Cross-cutting Topics Workshop Report.10.5281/zenodo.4593770

Eckerson, W.W., 1995. Three Tier Client/Server Architectures: Achieving Scalability, Performance, and Efficiency in Client Server Applications. Open Information Systems 3, 46-50,

ELIXIR, 2021. Data platform, https://elixir-europe.org/platforms/data

EOSC, 2020. https://www.eosc-portal.eu/

Esch, M., Chen, J., Colmsee, C., Klapperstuck, M., Grafahrend-Belau, E., Scholz, U., Lange, M., 2015. LAILAPS: the plant science search engine. Plant Cell Physiol 56, e8.10.1093/pcp/pcu185

European Commission, 2019. The European Green Deal, Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52019DC0640

European Commission Joint Research Centre, 2007. INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, p. 75, http://inspire.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf

European Parliament and of the Council, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

European Parliament and of the Council, 2019. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, https://eur-lex.europa.eu/eli/dir/2019/790/oj

Ewert, F., Asseng, S., Böhm, F., Feike, T., Fluck, J., Haunert, J.-H., Hoedt, F., Hoffmann, C., Lange, M., Lindstädt, B., Martini, D., Reif, J., Xenia, S., Stahl, U., Usadel, B., Weiland, C., 2021a. FAIRagro eine Initiative der NFDI4Agri zur Entwicklung einer FAIRen Dateninfrastrukturen für die Agrosystemforschung, Mitgliederversammlung der DAFA,

Ewert, F., Asseng, S., Böhm, F., Feike, T., Fluck, J., Haunert, J.-H., Hoedt, F., Hoffmann, C., Lange, M., Lindstädt, B., Martini, D., Reif, J., Xenia, S., Stahl, U., Usadel, B., Weiland, C., 2021b. FAIRe Dateninfrastruktur für die Agrosystemforschung (FAIRagro).10.5281/zenodo.5584803.

Ewert, F., Asseng, S., Böhm, F., Feike, T., Fluck, J., Haunert, J.-H., Hoedt, F., Hoffmann, C., Lange, M., Lindstädt, B., Martini, D., Reif, J., Xenia, S., Stahl, U., Usadel, B., Weiland, C., Deutsche, P., Online, V., 2021c. FAIRe Dateninfrastruktur für die Agrosystemforschung (FAIRagro), In: Deutsche Pflanzenschutztagung, G.P.i.V.f.u.W. (Ed.), 62. Deutsche Pflanzenschutztagung : Gesunde Pflanzen in Verantwortung für unsere Welt ; 21. - 23. September 2021, -Kurzfassungen der Vorträge und Poster-. Julius-Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen, Berlin ; Quedlinburg, pp. 393-394.

Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M.A., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Gaiser, T., Palosuo, T., Tao, F., Nendel, C., Roggero, P.P., Bartošová, L., Asseng, S., 2015. Crop modelling for integrated assessment of risk to food production from climate change. Environmental Modelling & Software 72, 287-303.10.1016/j.envsoft.2014.12.003

Ewert, F., van Ittersum, M.K., Bezlepkina, I., Therond, O., Andersen, E., Belhouchette, H., Bockstaller, C., Brouwer, F., Heckelei, T., Janssen, S., Knapen, R., Kuiper, M., Louhichi, K., Olsson, J.A., Turpin, N., Wery, J., Wien, J.E., Wolf, J., 2009. A methodology for enhanced flexibility of integrated assessment in agriculture. Environmental Science & Policy 12, 546-561.10.1016/j.envsci.2009.02.005

F1, 2021. FAIR principles: F1, https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/

F2, 2021. FAIR principles: F2, https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/

F3, 2021. FAIR principles: F3, https://www.go-fair.org/fair-principles/f3-metadata-clearly-explicitly-include-identifier-data-describe/

F4, 2021. FAIR principles: F4, https://www.go-fair.org/fair-principles/f4-metadata-registered-indexed-searchable-resource/

FAIR metrics, 2021. http://fairmetrics.org

FAIRagro, 2021. www.fairagro.net

FAIRDOM-SEEK, 2021. FAIRDOM-SEEK Website, https://github.com/seek4science/seek

FAO, 2018. The 10 elements of agroecology: Guiding the transition to sustainable food and agricultural systems, http://www.fao.org/3/

i9037en/i9037en.pdf

FAO, 2019. Agrontology, http://aims.fao.org/agrovoc/agrontology

FAO, 2021. AGROVOC. FAO, http://aims.fao.org/vest-registry/vocabularies/agrovoc

Feike, T., Riedesel Freiherr zu Eisenbach, L., Lieb, R., Gabriel, D., Sabboura, D., Rahman Shawon, A., Wetzel, M., Klocke, B., Krengel-Horney, S., Schwarz, J., 2021. Einfluss von Pflanzenschutzstrategie und Bodenbearbeitung auf den CO2-Fußabdruck von Weizen, Themenheft Dauerfeldversuche: Grundlage für die Landwirtschaft der Zukunft – 25 Jahre Dahnsdorf.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockstrom, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P., 2011. Solutions for a cultivated planet. Nature 478, 337-342.10.1038/nature10452

FORCE11, 2019. The FAIR data principles, https://www.force11.org/group/fairgroup/fairprinciples

Fronzek, S., Webber, H., Rötter, R., Ruane, A., Ewert, F., 2018. A daily time-step observed and scenario climate dataset on a European grid for crop modelling applications Version 2 https://dx.doi.org/10.4228/ZALF.DK.94

GAIA-X, 2021. https://www.data-infrastructure.eu

Garcia, L., Giraldo, O., Garcia, A., Rebholz-Schuhmann, D., 2019. Biotea-2-Bioschemas, facilitating structured markup for semantically annotated scholarly publications. Genomics Inform 17, e14.10.5808/GI.2019.17.2.e14

Garriga, M., Mateos, C., Flores, A., Cechich, A., Zunino, A., 2016. RESTful service composition at a glance: A survey. Journal of Network and Computer Applications 60, 32-53.10.1016/j.jnca.2015.11.020

Gayo, J.E.L., Prud'hommeaux, E., Boneva, I., Kontokostas, D., 2017. Validating RDF Data. Synthesis Lectures on the Semantic Web: Theory and Technology 7, 1-328.10.2200/s00786ed1v01y201707wbe016

Geonode, 2021. Geonode Website, https://geonode.org

German Council for Scientific Information Infrastructures (RfII), 2020. The Data Quality Challenge - Recommendations for Sustainable Research in the Digital Turn, Göttingen,

German Crop BioGreenformatics Network, 2021. Website of the GCBN, https://www.denbi.de/training-archive-sorted-according-by-de-nbi-units/german-crop-biogreenformatics-network-gcbn

Ghaffar, M., Schuler, D., Konig, P., Arend, D., Junker, A., Scholz, U., Lange, M., 2020. Programmatic Access to FAIRified Digital Plant Genetic Resources. J Integr Bioinform 16.10.1515/jib-2019-0060

Giuliani, G., Masó, J., Mazzetti, P., Nativi, S., Zabala, A., 2019. Paving the Way to Increased Interoperability of Earth Observations Data Cubes. Data 4, 113.10.3390/data4030113

GlassNet, 2021. GlassNet Website, https://mygeohub.org/groups/glassnet

Glöckner, F.O., Diepenbroek, M., Felden, J., Güntsch, A., Stoye, J., Overmann, J., Wimmers, K., Kostadinov, I., Yahyapour, R., Müller, W., Scholz, U., Triebel, D., Frenzel, M., Gemeinholzer, B., Goesmann, A., König-Ries, B., Bonn, A., Seeger, B., 2020. NFDI4BioDiversity - A Consortium for the National Research Data Infrastructure (NFDI).10.5281/zenodo.3943645.

Glöckner, F.O., Diepenbroek, M., Felden, J., Overmann, J., Bonn, A., Gemeinholzer, B., Güntsch, A., König-Ries, B., Seeger, B., Pollex-Krüger, A., Fluck, J., Pigeot, I., Toralf, K., Mühlhaus, T., Wolf, C., Heinrich, U., Steinbeck, C., Koepler, O., Stegle, O., Weimann, J., Schörner-Sadenius, T., Gutt, C., Stahl, F., Wagemann, K., Schrade, T., Schmitt, R., Eberl, C., Gauterin, F., Schultz, M., Bernard, L., 2019. Berlin Declaration on NFDI Cross-Cutting Topics, Zenodo.10.5281/zenodo.3457213.

GLTEN, 2021. Global Long-Term Agricultural Experiment Network, https://www.rothamsted.ac.uk/glten-global-long-term-agricultural-experiment-network

GO FAIR, 2021a. FAIR principles https://go-fair.org/fair-principles/

GO FAIR, 2021b. FAIRification Process. https://www.go-fair.org/fair-principles/fairification-process/

Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M.R., Peters, K., Schober, D., 2020. FAIR Computational Workflows. Data Intelligence 2, 108-121.10.1162/dint_a_00033

Gomes, V., Queiroz, G., Ferreira, K., 2020a. An Overview of Platforms for Big Earth Observation Data Management and Analysis. Remote Sensing 12.10.3390/rs12081253

Gomes, V., Queiroz, G., Ferreira, K., 2020b. An Overview of Platforms for Big Earth Observation Data Management and Analysis. Remote Sensing 12, 1253.10.3390/rs12081253

Google Dataset Search, 2021. https://datasetsearch.research.google.com/

Grieb, J., Weiland, C., Hardisty, A., Addink, W., Islam, S., Younis, S., Schmidt, M., 2021. Machine Learning as a Service for DiSSCo's Digital Specimen Architecture. Biodiversity Information Science and Standards 5.10.3897/biss.5.75634

Griffin, P.C., Khadake, J., LeMay, K.S., Lewis, S.E., Orchard, S., Pask, A., Pope, B., Roessner, U., Russell, K., Seemann, T., Treloar, A., Tyagi, S., Christiansen, J.H., Dayalan, S., Gladman, S., Hangartner, S.B., Hayden, H.L., Ho, W.W.H., Keeble-Gagnere, G., Korhonen, P.K., Neish, P., Prestes, P.R., Richardson, M.F., Watson-Haigh, N.S., Wyres, K.L., Young, N.D., Schneider, M.V., 2017. Best practice data life cycle approaches for the life sciences. F1000Res 6, 1618.10.12688/f1000research.12344.2

Grosse, M., Ahlborn, M.C., Hierold, W., 2021. Metadata of agricultural long-term experiments in Europe exclusive of Germany. Data Brief 38, 107322.10.1016/j.dib.2021.107322

Grosse, M., Heinrich, U., Hierold, W., 2019. Fact Sheet for the Description of Long-Term Field Experiments.https://doi.org/10.20387/bonares-r56g-fgrw

Grosse, M., Hierold, W., Ahlborn, M.C., Piepho, H.-P., Helming, K., 2020a. Long-term field experiments in Germany: classification and spatial representation. Soil 6, 579-596.10.5194/soil-6-579-2020

Grosse, M., Hoffmann, C., Specka, X., Svoboda, N., 2020b. Managing long-term experiment data: a repository for soil and agricultural research. 167-182.10.1016/b978-0-12-818186-7.00010-2

Hallinan, D., 2020a. Article 89, In: Boehm, F., Cole, M. (Eds.), GDPR Commentary. Elgar, Forthcoming 2020.

Hallinan, D., 2020b. Broad consent under the GDPR: an optimistic perspective on a bright future. Life Sciences, Society and Policy 16, 1.10.1186/s40504-019-0096-3

Hardisty, A., Addink, W., Glöckler, F., Güntsch, A., Islam, S., Weiland, C., 2021. A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Research Ideas and Outcomes 7.10.3897/rio.7.e67379

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R., Gene Ontology, C.,

2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32, D258-D261.10.1093/nar/gkh036

Hartmann, T., 2021. Forschungsdaten in den Naturwissenschaften: Eine urheberrechtliche Bestandsaufnahme mit ihren Implikationen für universitäres FDM, E-Science-Tage 2021. Mohammadianbisheh, https://e-science-tage.de/index.php/en/programm_2021?field_event_category_target_id=12

Hausen, J., Scholz-Starke, B., Burkhardt, U., Lesch, S., Rick, S., Russell, D., Ross-Nickoll, M., Ottermanns, R., 2017. Edaphostat: interactive ecological analysis of soil organism occurrences and preferences from the Edaphobase data warehouse. Database (Oxford) 2017.10.1093/database/bax080

He, L., Han, Z.B., 2017. Do usage counts of scientific data make sense? An investigation of the Dryad repository. Library Hi Tech 35, 332-342.10.1108/lht-12-2016-0158

Hertel, T., Elouafi, I., Tanticharoen, M., Ewert, F., 2021. Diversification for enhanced food systems resilience. Nature Food,

HLPE, 2019. Agroecological and other innovative approaches for sustainable agriculture and food systems that enhance food security and nutrition, In: Security, H.L.P.o.E.o.F.S.a.N.o.t.W.F. (Ed.), http://www.fao.org/cfs/cfs-hlpe/en/

Hossard, L., Philibert, A., Bertrand, M., Colnenne-David, C., Debaeke, P., Munier-Jolain, N., Jeuffroy, M.H., Richard, G., Makowski, D., 2014. Effects of halving pesticide use on wheat production. Sci Rep 4, 4405.10.1038/srep04405

I1, 2021. FAIR principles: I1, https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/

I2, 2021. FAIR principles: I2, https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/

I3, 2021. FAIR principles: I3, https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/

IGSN, 2021. a globally unique and persistent identifier for material samples. https://www.igsn.org/

INKA-BB, 2021. Innovation Network of Climate Change Adaptation Brandenburg Berlin, http://www.inka-bb.de/

Islam, S., Hardisty, A., Addink, W., Weiland, C., Glöckler, F., 2020. Incorporating RDA Outputs in the Design of a European Research Infrastructure for Natural Science Collections. Data Science Journal 19.10.5334/dsj-2020-050

ISMC, 2021. The International Soil Modeling Consortium, https://soil-modeling.org/

ISO 19157, 2013. ISO 19157:2013 Geographic information — Data quality, https://www.iso.org/standard/32575.html

Jalswal, P., Avraham, S., Ilic, K., Kellogg, E.A., McCouch, S., Pujar, A., Reiser, L., Rhee, S.Y., Sachs, M.M., Schaeffer, M., Stein, L., Stevens, P., Vincent, L., Ware, D., Zapata, F., 2005. Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. Comparative and Functional Genomics 6, 388-397.10.1002/cfg.496

Janssen, S., Andersen, E., Athanasiadis, I.N., van Ittersum, M.K., 2009a. A database for integrated assessment of European agricultural systems. Environmental Science & Policy 12, 573-587.10.1016/j.envsci.2009.01.007

Janssen, S., Ewert, F., Li, H., Athanasiadis, I.N., Wien, J.J.F., Thérond, O., Knapen, M.J.R., Bezlepkina, I., Alkan-Olsson, J., Rizzoli, A.E., Belhouchette, H., Svensson, M., van Ittersum, M.K., 2009b. Defining assessment projects and scenarios for policy support: Use of ontology in Integrated Assessment and Modelling. Environmental Modelling & Software 24, 1491-1500.10.1016/j.envsoft.2009.04.009

Kahn, M.G., Callahan, T.J., Barnard, J., Bauck, A.E., Brown, J., Davidson, B.N., Estiri, H., Goerg, C., Holve, E., Johnson, S.G., Liaw, S.T., Hamilton-Lopez, M., Meeker, D., Ong, T.C., Ryan, P., Shang, N., Weiskopf, N.G., Weng, C., Zozus, M.N., Schilling, L., 2016. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC) 4, 1244.10.13063/2327-9214.1244

Kaspar, F., Zimmermann, K., Polte-Rudolf, C., 2015. An overview of the phenological observation network and the phenological database of Germany's national meteorological service (Deutscher Wetterdienst). Advances in Science and Research 11, 93-99.10.5194/asr-11-93-2014

Kersebaum, K.C., Boote, K.J., Jorgenson, J.S., Nendel, C., Bindi, M., Frühauf, C., Gaiser, T., Hoogenboom, G., Kollas, C., Olesen, J.E., Rötter, R.P., Ruget, F., Thorburn, P.J., Trnka, M., Wegehenkel, M., 2015. Analysis and classification of data sets for calibration and validation of agro-ecosystem models. Environmental Modelling & Software 72, 402-417.10.1016/j.envsoft.2015.05.009

Krantz, M., Zimmer, D., Adler, S.O., Kitashova, A., Klipp, E., Muhlhaus, T., Nagele, T., 2021. Data Management and Modeling in Plant Biology. Front Plant Sci 12, 717958.10.3389/fpls.2021.717958

Kristensen, J.A., Balstrøm, T., Jones, R.J.A., Jones, A., Montanarella, L., Panagos, P., Breuning-Madsen, H., 2019. Development of a harmonised soil profile analytical database for Europe: a resource for supporting regional soil management. Soil 5, 289-301.10.5194/soil-5-289-2019

Lamichhane, J.R., Arendse, W., Dachbrodt-Saaydeh, S., Kudsk, P., Roman, J.C., van Bijsterveldt-Gels, J.E.M., Wick, M., Messéan, A., 2015. Challenges and opportunities for integrated pest management in Europe: A telling example of minor uses. Crop Protection 74, 42-47.10.1016/j.cropro.2015.04.005

Lamprecht, A.-L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P.A., McQuilton, P., Valencia, A., Harrow, J., Psomopoulos, F., Gelpi, J.L., Chue Hong, N., Goble, C., Capella-Gutierrez, S., Groth, P., Groth, P., Dumontier, M., 2020. Towards FAIR principles for research software. Data Science 3, 37-59.10.3233/ds-190026

Lannom, L., Koureas, D., Hardisty, A.R., 2020. FAIR Data and Services in Biodiversity Science and Geoscience. Data Intelligence 2, 122-130.10.1162/dint_a_00034

Latif, A., Limani, F., Tochtermann, K., 2019. A Generic Research Data Infrastructure for Long Tail Research Data Management. Data Science Journal 18.10.5334/dsj-2019-017

Lechenet, M., Dessaint, F., Py, G., Makowski, D., Munier-Jolain, N., 2017. Reducing pesticide use while preserving crop productivity and profitability on arable farms. Nat Plants 3, 17008.10.1038/nplants.2017.8

LIVIVO, 2021. https://www.livivo.de/

MACSUR, 2021. MACSUR Website, https://macsur.eu/

Mahecha, M.D., Gans, F., Brandt, G., Christiansen, R., Cornell, S.E., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., Donges, J.F., Dorigo, W., Estupinan-Suarez, L.M., Gutierrez-Velez, V.H., Gutwin, M., Jung, M., Londoño, M.C., Miralles, D.G., Papastefanou, P., Reichstein, M., 2020. Earth system data cubes unravel global multivariate dynamics. Earth System Dynamics 11, 201-234.10.5194/esd-11-201-2020

Mehdipoor, H., Zurita-Milla, R., Rosemartin, A., Gerst, K.L., Weltzin, J.F., 2015. Developing a Workflow to Identify Inconsistencies in Volunteered Geographic Information: A Phenological Case Study. PLoS One 10, e0140811.10.1371/journal.pone.0140811

Metrics, F., 2021. FAIRMetrics Github, https://github.com/FAIRMetrics/Metrics

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59-86.10.1016/j.geoderma.2017.01.002

Möller, M., Koschitzki, T., Hartmann, K.-J., Jahn, R., 2012. Plausibility test of conceptual soil maps using relief parameters. Catena 88, 57-67.10.1016/j.catena.2011.08.002

Nature Scientific Data, 2021. Springer Nature Limited, https://www.nature.com/sdata/

Oeltjen, W., Neumann, K., Stahl, U., Stephan, R., 2019. MyCoRe macht Forschungsdaten FAIR. Bibliothek Forschung und Praxis 43, 82.https://doi.org/10.1515/bfp-2019-2013

OpenAgrar, 2021. https://www.openagrar.de/content/index.xml

OpenAIRE, 2021. OpenAIRE. https://www.openaire.eu/

OpenAPI, 2021. https://www.openapis.org/

OpenProject, 2021. https://www.openproject.org/de/

Oppermann, M., Weise, S., Dittmann, C., Knupffer, H., 2015. GBIS: the information system of the German Genebank. Database (Oxford) 2015, bav021.10.1093/database/bav021

patchCROP, 2021. patchCROP Website, https://comm.zalf.de/sites/patchcrop/SitePages/Homepage.aspx

Phenopackets, 2021. Open and Computable Bioinformation. http://phenopackets.org/

PhenoRob, 2021a. http://www.phenorob.de

PhenoRob, 2021b. PhenoRob Website, https://www.phenorob.de/

PlabiPD, 2021. https://plabipd.de/

Plant Ontology, 2019. https://www.ebi.ac.uk/ols/ontologies/po

Porwollik, V., Müller, C., Elliott, J., Chryssanthacopoulos, J., Iizumi, T., Ray, D.K., Ruane, A.C., Arneth, A., Balkovič, J., Ciais, P., Deryng, D., Folberth, C., Izaurralde, R.C., Jones, C.D., Khabarov, N., Lawrence, P.J., Liu, W., Pugh, T.A.M., Reddy, A., Sakurai, G., Schmid, E., Wang, X., de Wit, A., Wu, X., 2017. Spatial and temporal uncertainty of crop yield aggregations. European Journal of Agronomy 88, 10-21.10.1016/j.eja.2016.08.006

Pretty, J., 2018. Intensification for redesigned and sustainable agricultural systems. Science 362.10.1126/science.aav0294

PUBLISSO, 2021. https://www.publisso.de/en/publishing/repositories/repository-for-life-sciences/

Pulleman, M.M., Six, J., van Breemen, N., Jongmans, A.G., 2005. Soil organic matter distribution and microaggregate characteristics as affected by agricultural management and earthworm activity. European Journal of Soil Science 56, 453-467.10.1111/j.1365-2389.2004.00696.x

R1, 2021a. FAIR principles: R1, https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/

R1, 2021b. FAIR principles: R1 - (Meta)data meet domain-relevant community standards. https://www.go-fair.org/fair-principles/r1-3-metadata-meet-domain-relevant-community-standards/

R1.1, 2021. FAIR principles: R1.1, https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/

R1.2, 2021. FAIR principles: R1.2, https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/

RDA, 2016. Global Digital Object Cloud (DOC) - A Guiding Vision. https://www.rd-alliance.org/group/data-fabric-ig/wiki/global-digital-object-cloud

Reynolds, M., Atkin, O.K., Bennett, M., Cooper, M., Dodd, I.C., Foulkes, M.J., Frohberg, C., Hammer, G., Henderson, I.R., Huang, B., Korzun, V., McCouch, S.R., Messina, C.D., Pogson, B.J., Slafer, G.A., Taylor, N.L., Wittich, P.E., 2021. Addressing Research Bottlenecks to Crop Productivity. Trends Plant Sci 26, 607-630.10.1016/j.tplants.2021.03.011

Rigden, D.J., Fernandez, X.M., 2019. The 26th annual Nucleic Acids Research database issue and Molecular Biology Database Collection. Nucleic Acids Res 47, D1-D7.10.1093/nar/gky1267

RO-Crate, 2021. Research Object Crate, https://www.researchobject.org/ro-crate/

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The Agricultural Model Intercomparison and

Improvement Project (AgMIP): Protocols and pilot studies. Agricultural and Forest Meteorology 170, 166-182.10.1016/j.agrformet.2012.09.011

Rosenzweig, C., Ruane, A.C., Antle, J., Elliott, J., Ashfaq, M., Chatta, A.A., Ewert, F., Folberth, C., Hathie, I., Havlik, P., Hoogenboom, G., Lotze-Campen, H., MacCarthy, D.S., Mason-D'Croz, D., Contreras, E.M., Muller, C., Perez-Dominguez, I., Phillips, M., Porter, C., Raymundo, R.M., Sands, R.D., Schleussner, C.F., Valdivia, R.O., Valin, H., Wiebe, K., 2018. Coordinating AgMIP data and models across global and regional scales for 1.5 degrees C and 2.0 degrees C assessments. Philos Trans A Math Phys Eng Sci 376.10.1098/rsta.2016.0455

Schwardmann, U., 2020. Digital Objects – FAIR Digital Objects: Which Services Are Required? Data Science Journal 19.10.5334/dsj-2020-015

Schwarz, J., Klocke, B., Wagner, C., Krengel, S., 2018. Untersuchungen zum notwendigen Maß bei der Anwendung von Pflanzenschutzmitteln in Winterweizen in den Jahren 2004 bis 2016. Gesunde Pflanzen 70, 119-127.10.1007/s10343-018-0422-3

Selby, P., Abbeloos, R., Backlund, J.E., Basterrechea Salido, M., Bauchet, G., Benites-Alfaro, O.E., Birkett, C., Calaminos, V.C., Carceller, P., Cornut, G., Vasques Costa, B., Edwards, J.D., Finkers, R., Yanxin Gao, S., Ghaffar, M., Glaser, P., Guignon, V., Hok, P., Kilian, A., König, P., Lagare, J.E.B., Lange, M., Laporte, M.-A., Larmande, P., LeBauer, D.S., Lyon, D.A., Marshall, D.S., Matthews, D., Milne, I., Mistry, N., Morales, N., Mueller, L.A., Neveu, P., Papoutsoglou, E., Pearce, B., Perez-Masias, I., Pommier, C., Ramírez-González, R.H., Rathore, A., Raquel, A.M., Raubach, S., Rife, T., Robbins, K., Rouard, M., Sarma, C., Scholz, U., Sempéré, G., Shaw, P.D., Simon, R., Soldevilla, N., Stephen, G., Sun, Q., Tovar, C., Uszynski, G., Verouden, M., 2019. BrAPI—an application programming interface for plant breeding applications. Bioinformatics.10.1093/bioinformatics/btz190

Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C., Haklay, M., 2016. A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science 31, 139-167.10.1080/13658816.2016.1189556

Senft, M.S., Ulrike; Svoboda, Nikolai, 2021. Dataset: survey about research data management in agricultural sciences in Germany. OpenAgrar Repository.10.5073/20211013-105447

Shrestha, R., Matteis, L., Skofic, M., Portugal, A., McLaren, G., Hyman, G., Arnaud, E., 2012. Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. Frontiers in Physiology 3, 10.10.3389/fphys.2012.00326

SmartAPI, 2021. https://smart-api.info/

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L.J., Coppens, F., Fernández, J.M., Daniel Garijo, B.G., Marco La Rosa, S.L., Eoghan Ó Carragáin, Portier, M., Trisovic, A., Community, R.-C., Groth, P., Goble, C., 2021. Packaging research artefacts with RO-Crate. Zenodo.10.5281/zenodo.5146228

Specka, X., Gärtner, P., Hoffmann, C., Svoboda, N., Stecker, M., Einspanier, U., Senkler, K., Muqit Zoarder, M.A., Heinrich, U., 2019. The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes. Computers & Geosciences.https://doi.org/10.1016/j.cageo.2019.07.005

Šťastná, M., Toman, F., Dufková, J., 2010. Usage of SUBSTOR model in potato yield prediction. Agricultural Water Management 97, 286-290.10.1016/j.agwat.2009.09.015

Storch, T., Reck, C., Holzwarth, S., Wiegers, B., Mandery, N., Raape, U., Strobl, C., Volkmann, R., Böttcher, M., Hirner, A., Senft, J., Plesia, N., Kukuk, T., Meissl, S., Felske, J.-R., Heege, T., Keuck, V., Schmidt, M., Staudenrausch, H., 2019. Insights into CODE-DE – Germany's Copernicus data and exploitation platform. Big Earth Data 3, 338-361.10.1080/20964471.2019.1692297

Systems, C.-C.C.f.S.D., 2012. Reference Model for an Open Archival Information System (OAIS), http://public.ccsds.org/publications/archive/650x0b1.pdf

Thessen, A.E., Woodburn, M., Koureas, D., Paul, D., Conlon, M., Shorthouse, D.P., Ramdeen, S., 2019. Proper Attribution for Curation and Maintenance of Research

Collections: Metadata Recommendations of the RDA/TDWG Working Group. Data Science Journal 18, 54.10.5334/dsj-2019-054

Tian, Z., Wang, J.W., Li, J., Han, B., 2021. Designing future crops: challenges and strategies for sustainable agriculture. Plant J 105, 1165-1178.10.1111/tpj.15107

TISDAR, 2021. Thünen Institute Spatial Data-Repository (TISDAR) https://www.thuenen.de/de/infrastruktur/thuenen-atlas-und-geoinformation/tisdar/

Uehara, G., Tsuji, G.Y., 1998. Overview of IBSNAT. 7, 1-7.10.1007/978-94-017-3624-4_1

Van Ittersum, M., Ewert, F., Wéry, J., Heckelei, T., Belhouchette, H., Bergez, J.-E., J.-E., Hengsdijk, H., Janssen, S., Russell, G., Therond, O., 2008. Integrated assessment of agricultural systems - Modular System for Agricultural and Environmental Modelling (SEAMLESS-IF). Italian Journal of Agronomy 3, 311-312, https://hal.inrae.fr/hal-02656832

Venkatesan, A., Tagny Ngompe, G., Hassouni, N.E., Chentli, I., Guignon, V., Jonquet, C., Ruiz, M., Larmande, P., 2018. Agronomic Linked Data (AgroLD): A knowledge-based system to enable integrative biology in agronomy. PLoS One 13, e0198270.10.1371/journal.pone.0198270

von Braun, J., Afsana, K., Fresco, L.O., Hassan, M., 2021a. Food systems: seven priorities to end hunger and protect the planet. Nature 597, 28-30.10.1038/d41586-021-02331-x

von Braun, J., Afsana, K., Fresco, L.O., Hassan, M., Torero, M., 2021b. Food system concepts and definitions for science and political action. Nature Food 2, 748-750.10.1038/s43016-021-00361-2

W3C, 2021. https://www.w3.org/standards/semanticweb

W3C Prov, 2021. https://www.w3.org/TR/prov-overview/

W3C VoID, 2021. https://www.w3.org/TR/void/

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., Schurr, U., 2020. Phenotyping: New Windows into the Plant for Breeders. Annu Rev Plant Biol 71, 689-712.10.1146/annurev-arplant-042916-041124

Weise, S., Lohwasser, U., Oppermann, M., 2020. Document or Lose It-On the Importance of Information Management for Genetic Resources Conservation in Genebanks. Plants (Basel) 9.10.3390/plants9081050

Weise, S., Oppermann, M., Maggioni, L., van Hintum, T., Knupffer, H., 2017. EURISCO: The European search catalogue for plant genetic resources. Nucleic Acids Res 45, D1003-D1008.10.1093/nar/gkw755

Wezel, A., Herren, B.G., Kerr, R.B., Barrios, E., Gonçalves, A.L.R., Sinclair, F., 2020. Agroecological principles and elements and their implications for transitioning to sustainable food systems. A review. Agronomy for Sustainable Development 40.10.1007/s13593-020-00646-z

White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., Koo, J., Kim, S., Porter, C.H., Wilkens, P.W., Hoogenboom, G., 2013. Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. Computers and Electronics in Agriculture 96, 1-12.10.1016/j.compag.2013.04.003

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ’t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018, https://doi.org/10.1038/sdata.2016.18

Wilkinson, M.D., Sansone, S.-A., Schultes, E., Doorn, P., Bonino da Silva Santos, L.O., Dumontier, M., 2018. A design framework and exemplar metrics for FAIRness. Scientific Data 5, 180118.10.1038/sdata.2018.118

Willocquet, L., Meza, W.R., Dumont, B., Klocke, B., Feike, T., Kersebaum, K.C., Meriggi, P., Rossi, V., Ficke, A., Djurle, A., Savary, S., 2021. An outlook on wheat health in Europe from a network of field experiments. Crop Protection 139, 105335.10.1016/j.cropro.2020.105335

WorkflowHub, 2021. Scientific computational workflows. https://workflowhub.eu/

Yenni, G.M., Christensen, E.M., Bledsoe, E.K., Supp, S.R., Diaz, R.M., White, E.P., Ernest, S.K.M., 2019. Developing a modern data workflow for regularly updated data. PLoS Biol 17, e3000125.10.1371/journal.pbio.3000125

Zhang, M., Jiang, L., Zhao, J., Yue, P., Zhang, X., 2020. Coupling OGC WPS and W3C PROV for provenance-aware geoprocessing workflows. Computers & Geosciences 138, 104419.10.1016/j.cageo.2020.104419

Zielinski, T., Hay, J., Millar, A., 2019. The grant is dead, long live the data - migration as a pragmatic exit strategy for research data preservation [version 2; peer review: 2 approved]. Wellcome Open Research 4.10.12688/wellcomeopenres.15341.2

Zukunftskommission Landwirtschaft Geschäftsstelle, 2021. Zukunf Landwirtschaft. Eine gesamtgesellschafliche Aufgabe. Empfehlungen der Zukunfskommission Landwirtschaft, https://www.bundesregierung.de/resource/blob/997532/1939908/39f791fc6a9ff293ae520e2fea41baa7/2021-07-06-zukunftskommission-landwirtschaft-data.pdf