

FAIRragro UC onboarding

Introduction of TA measures for new use cases

This document aims to introduce new and prospective use cases (UCs) to the different measures of FAIRragro task areas, with an emphasis on specific information, material, and services that serve as a basis for the exchange between UCs and measures.

For each measure, a summary of activities and the link to the measure web page (with contact information of the team members) are provided. Use the hyperlinks in the table of contents below to browse the measure of interest.

Task Area 2. Community Involvement and Networking.....	1
Measure 2.1: Communication and Dissemination.....	2
Measure 2.2: Community Participation.....	3
Measure 2.3: Use Case Onboarding.....	4
Measure 2.4: Training and Education.....	5
Measure 2.5: Data Steward Service Center (DSSC).....	6
Task Area 3. Standardization, interoperability and quality.....	8
Measure 3.1: Standards for Digital Resources.....	9
Measure 3.2: Standards for Data Management, FAIRness and Discoverability.....	11
Measure 3.3: Measures and Application-data-matrix for Data Quality and Fitness-for-use.....	12
Measure 3.4: Data Quality Annotation, Curation and Feedback/Review.....	13
Measure 3.5: FAIR Workflows and FAIR Digital Objects.....	14
Measure 3.6: Legal Framework and Machine-Actionable Policies.....	16
Task Area 4. Services.....	17
Measure 4.1. Central Services for the FAIRragro community.....	18
Measure 4.2. Network of federated research data infrastructures (RDI).....	19
Measure 4.3: Searchable Inventory of Services and Data.....	20
Measure 4.4: Scientific Workflow Infrastructure (SciWIn).....	21

Task Area 2. Community Involvement and Networking

The agrosystem community is characterized by its wide-ranging presence across various disciplines, such as agronomy, livestock sciences, or plant breeding, as well as within diverse organizations such as universities, non-university research organisations or governmental institutions and farms. It also interfaces with multiple other domains like soil sciences, biodiversity, and environmental sciences. Against this backdrop, the success of FAIRagro as a harmonized data infrastructure for the agrosystem domain relies on the effective integration of its heterogeneous community. This goal forms the core of TA2's community-driven strategy for gathering, engaging, training, educating, and supporting the agrosystem community. We will oversee communication, networking, and information dissemination both within and outside the consortium by organizing multiple roadshows and workshops and maintaining a strong presence of FAIRagro in external events ([Measure 2.1](#)). A Community Advisory Board (CAB) composed of elected domain experts will evaluate progress and provide strategic advice to the consortium ([Measure 2.2](#)). We will consolidate community engagement by soliciting feedback through multiple communication channels ([Measure 2.2](#)) and by onboarding new use cases to broaden the reach of FAIRagro to additional research areas ([Measure 2.3](#)). To reach out to our community, we will develop and provide data management training programs and open educational resources tailored to meet the specific needs of the community ([Measure 2.4](#)). The Data Steward Service Center (DSSC, [Measure 2.5](#)) brings together experts in major areas of agrosystem research data management, such as data management plans, data publication, and licensing. It will provide a permanent support service by processing helpdesk support requests and offering personalized assistance to address the data management challenges faced by the use cases.

Measure 2.1: Communication and Dissemination

Leads: Ulrike Stahl, Oliver Kirchgessner (JKI)

[Staff & contact](#)

The measure aims to strengthen the community by focusing on communication, networking, and information dissemination. We involve FAIRagro partners in various events, to engage directly with the community. We participate in regular roadshow events to promote FAIRagro, gather feedback, and recruit experts and data-related roles, as well as scientific conferences with contributions and information booths. In addition to ensuring FAIRagro's presence at external events, we have been hosting a summit in 2023 for the agrosystem community - two more will follow - to (1) present and refine the objectives of FAIRagro with community feedback and announce a call for new use cases; (2) promote FAIRagro services and provide a feedback summary on the experience acquired with realized use cases; (3) report FAIRagro outcomes and service portfolio, involving the broader community with adjacent disciplines such as forestry, economics, and animal sciences in the onboarding process. Dissemination channels, such as roadshows and workshops, are established to ensure information transfer from all FAIRagro task areas to the community. Furthermore we work closely with NFDI and other consortia in terms of communication networking and exchange.

The FAIRagro portal (fairagro.net) is our information hub: it informs, guides, and interacts with the community through various communication channels and features like search functions, a help desk with a ticket system, and submission forms for prospective use cases. In addition to the FAIRagro portal (fairagro.net) we use Social Media (Mastodon, BlueSky, LinkedIn) to reach the respective audiences and regular Newsletters for the distribution of information.

Besides, we produce FAIRagro information material like flyers and related merchandising products / giveaways like stickers, pens etc. for promotion.

What we need from the UCs:

- Nomination of fix contact persons
- Regular information on planned events, products and success stories
- Regular information on the status of the work within the UCs
- Information on relevant institutions, initiatives, stakeholders, etc. (e.g. research-oriented start-ups, international organisations) relevant for the community
- Support of our FAIRagro Social Media activities, e.g. spreading and likes of FAIRagro social media posts :)
- Ideas and inspiration for topics and activities for the Community Summits

What we offer to the UCs:

- Promotion of UCs' events and products
- Contacts of relevant experts to address specific questions
- Collect and process feedback and usage data for your products on the FAIRagro portal

Measure 2.2: Community Participation

Leads: Anne Sennhenn, James Anderson (ATB)

[Staff & contact](#)

Measure 2.2 aims to drive cultural change towards FAIR research data management practices by systematically collecting structured feedback from the community.

User feedback will be gathered through multiple communication channels, including surveys, interviews, conferences, workshops, social media, and help desk tickets managed by data stewards. We will facilitate all feedback loops between the community, potential users, and FAIRagro developers to ensure that infrastructures and services are developed in alignment with community needs. To achieve this, various participation measures have been established to systematically collect and channel feedback into the respective Task Areas (TAs) and Measures for further development:

- Community Advisory Board (CAB) ([Measure 2.2](#)).
- Use Cases (UC) ([Measure 2.3](#))
- Data Steward Service Center (DSSC) ([Measure 2.5](#))
- Education and training ([Measure 2.4](#)).
- Community summit ([Measure 2.1](#))
- Community survey ([Measure 2.2](#)).
- Website, newsletter, social media ([Measure 2.1](#))

These structured engagement activities ensure that community input is effectively integrated into the continuous improvement of FAIRagro's infrastructures and services.

What we need from the UCs

- Thoughts on meaningful expertise and important institutions, initiatives and stakeholder relevant for the CAB; Ideas and support for the CAB member acquisition
- Pointing out topics, events, etc. where feedback is required, surveys should be conducted

What we offer to the UCs:

- Valuable feedback and evaluation on UC development and onboarding through the CAB and other measures
- Structured feedback (including support for the survey design, implementation and analysis) on UC relevant topics, for UC events etc.

Measure 2.3: Use Case Onboarding

Leads: Anne Sennhenn, James Anderson (ATB)

[Staff & contact](#)

Measure 2.3 focuses on establishing a structured onboarding process within FAIRagro to engage additional relevant institutions, networks, and initiatives. The goal is to expand the community towards neighboring and new research domains and disciplines (from agrosystem science towards agricultural science and beyond).

To achieve this, open calls are announced for the inclusion of new Use Cases. Evaluation criteria are developed to ensure that selected use cases align with strategic considerations and necessary developments.

The onboarding process, as well as the continuous collaboration of Use Cases with individual Task Areas (TAs) and Measures, is actively supported and guided. Progress reports, activities, publications, and outputs will be continuously updated, communicated, and tested in collaboration with the community.

Synergies and resulting working groups play a crucial role in driving the appropriate further development of FAIRagro infrastructures and services. These developments are designed to be closely aligned with research requirements, fostering more efficient, interdisciplinary, and innovative research while also promoting FAIR (Findable, Accessible, Interoperable, and Reusable) and open research data management practices.

Additionally, as part of these efforts, concepts and strategies are being further developed to complement the use case approach. The aim is to integrate community engagement as an active and sustainable component of FAIRagro.

What we need from the UCs

- Inventory of UCs' data management assets and challenges to identify overlaps and gaps
- Shared experience on the process of UC formation and onboarding

What we offer to the UCs:

- Mentoring of new UC managers through their integration into FAIRagro to foster collaboration across UCs, TAs and Measures
- Compiled onboarding documents and information

Measure 2.4: Training and Education

Leads: Sophie Boße, Birte Lindstädt (ZBMED)

[Staff & contact](#)

The objective of measure 2.4 is to enhance data management skills in agrosystem science through tailored training and education. We will develop a training module that covers three aspects of competence development: (1) online and personal training, (2) provision of open educational resources (OERs), and (3) contribution to curricular education. (1) Training content will be developed using existing materials, incorporating discipline-specific aspects and insights from FAIRagro data governance experts, software developers, UC data curators, and data stewards, and tailored to the needs of graduate students, and scientists in the field of agrosystem science alike. Training will be disseminated through collaborations with agricultural universities and other external partners such as NFDI4Biodiversity, NFDI4Earth or de.NBI. Dedicated workshops at scientific conferences, are planned and will also be organised on request by specific working groups or institutions. We will also provide training on FAIRagro tools and services to enable our community to use them. (2) Open educational resources on RDM that are tailored for agricultural sciences will be made available in a variety of formats (e.g., slidedecks, videos, wikis, websites, podcasts, etc.) and will also include materials developed by other consortia or institutions. All training materials will be provided to the community through the FAIRagro Portal and the DALIA knowledge graph. (3) The "agrosystem module" will contribute to the education of PhD-Students through empowering responsible teaching staff and other trainers to integrate RDM into graduate school curricula. Therefore we will provide training materials and train-the-trainer programmes as well as discipline-specific content for our partners' courses. The module will be refined over time, based on experience and community needs, and documented in a handbook.

What we need from the UCs

- Specific training needs in the UC's discipline
- Inventory of data management tools and services used in the UC
- Specific expertise of the UC staff that can be added to our materials and trainings

What we offer to the UCs:

- Tailored training for researchers across FAIRagro disciplines and beyond, also for the respective use case
- Integration of the knowledge drawn from UCs' work into training material for the community

Measure 2.5: Data Steward Service Center (DSSC)

Leads: Marcus Schmidt, Nikolai Svoboda (ZALF)

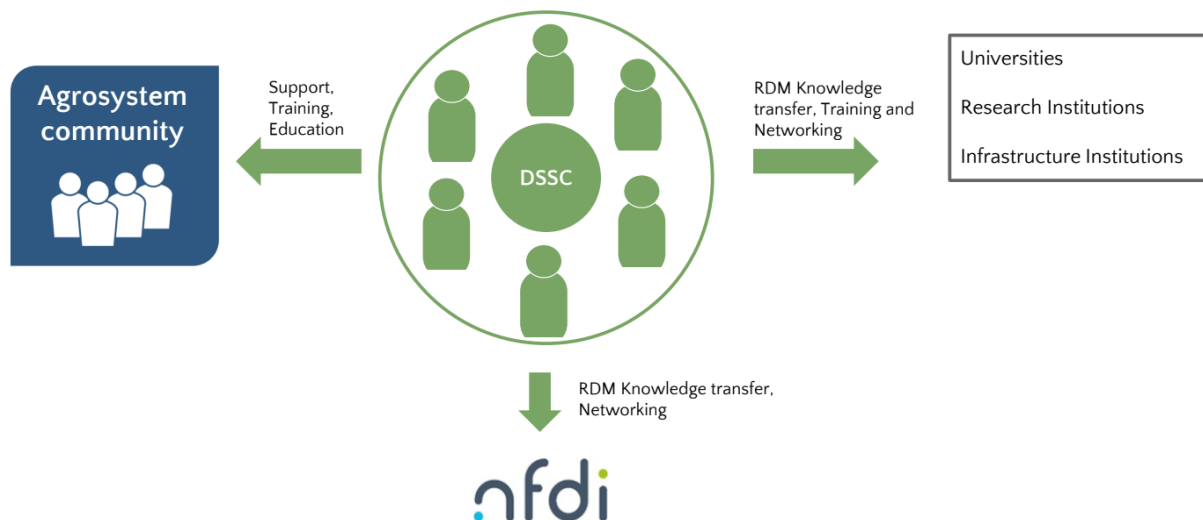
Contact: dataservice@fairagro.net

The Data Steward Support Center (DSSC) plays a pivotal role in harmonizing data management within FAIRagro. Data stewards hold expertise in research data management for agrosystem research, each of them holding core competencies in a specific domain of agrosystem sciences. They will oversee UC data curators and facilitate community engagement. As of the starting phase of the project, the DSSC comprises five data stewards with the following core competencies:

- (1) Plant breeding | Data management plans
- (2) Long-term field experiments | Data publication (PIDs, metadata...)
- (3) Geodata | Data quality
- (4) Field robotics | Big data, data visualization
- (5) Legal & ethical aspects | Copyright, licenses, IT security

The DSSC acts as a knowledge hub and liaison for data management inquiries. It will collaborate with other consortia to establish a cross-NFDI data steward support system.

Data stewards will address user requests submitted via the help desk, if necessary by seeking the involvement of data governance experts and software developers (TA3 & 4). They will also take a supporting role in specific projects of the UCs (e.g., the definition of requirements for data curation, help with locating specific data types, etc.). Additionally, they will actively collaborate with [Measure 2.4](#) in developing education and training material, and with [Measure 3.2](#) to disseminate DMP templates. Data stewards will also have an active role in feedback collection and transmission to [Measure 2.2](#). Overall, the DSSC plays a multifaceted role in training and knowledge dissemination, targeted at both the UCs and the wider community.



What we need from the UCs

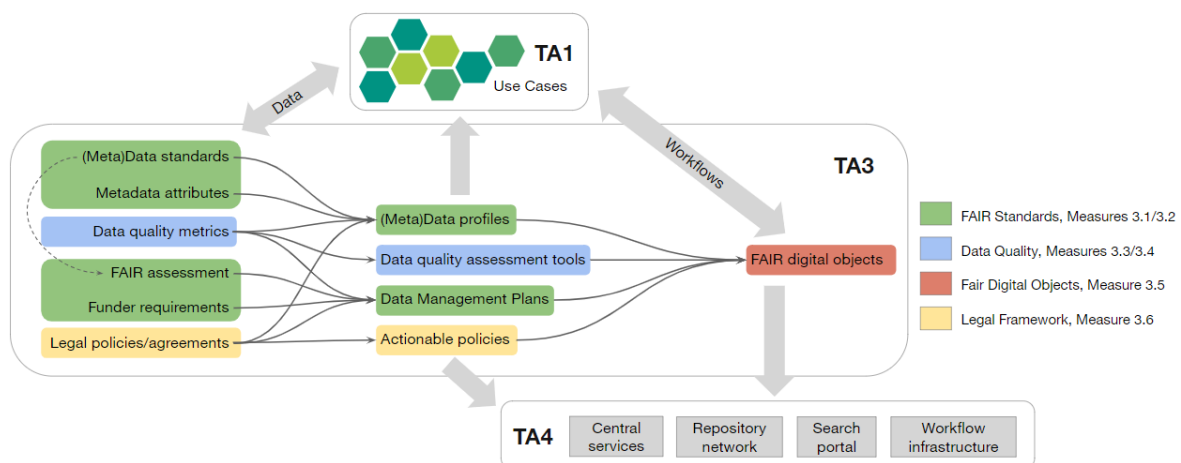
- One primary contact person for regular exchange with the assigned data steward and DSSC coordination
- Feedback on DSSC services and suggestions for continuous improvement

What we offer to the UCs:

- A dedicated data steward as primary contact for “hands-on” support
- Support in making a data management plan and a data policy tailored to the needs of the focal project
- Guidance on locating specific data
- Support in data publication in data repositories and data journals
- Access to our network of experts able to answer very specific questions on research data management (metadata, map services, data standards, etc.)

Task Area 3. Standardization, interoperability and quality

TA3 assembles experts in the domain of information science, data quality, and data governance with the primary objective of facilitating the reuse, quality screening, and annotation of agrosystem research data. To this end, we will provide partner infrastructures with standards and tools to operate together seamlessly within the FAIRagro infrastructure. Our efforts will involve fostering the interoperability among existing data vocabularies and ontologies, establishing metadata models tailored to specific data types, and providing data management plan templates and publication guidelines to enable FAIR-compliant management throughout the data life cycle ([Measure 3.1](#) and [3.2](#)). We will develop a framework for integrated digital representations of complex data objects, that encompass multiple data types with their associated metadata and unique identifiers ([Measure 3.5](#)). These FAIR digital objects will enable interoperability across repositories and workflows while incorporating legal metadata to ensure compliance with specific data protection requirements while maintaining findability ([Measure 3.6](#)). In addition, we will develop data quality metrics based on the specific needs of our use cases ([Measure 3.3](#)) and implement a toolbox for appraising and improving data quality with respect to these metrics ([Measure 3.4](#)). In all TA3 measures, we will maintain a feedback loop with the FAIRagro use cases to provide tools and standards that are tailored to their needs. All the standards, specifications, and tools designed by TA3 will serve as the building blocks for the software services developed by TA4.



Measure 3.1: Standards for Digital Resources

Leads: Daniel Martini (KTBL), Claus Weiland (SGN)

[Staff & contact](#)

Our main goal is to promote awareness and compliance with FAIR principles within the agrosystem community. For this, we will curate an inventory of established (meta)data standards (i.e., structured frameworks that describe the attributes, structure, and context of data), vocabularies (i.e., controlled list of terms used to describe and categorize data consistently), and ontologies (i.e., domain-specific knowledge representations that include vocabularies and the relationships and rules that defined how different terms are connected) for agricultural data, based on the specific requirements of FAIRagro use cases. We will assess these resources for compliance with FAIR principles to identify gaps and extension possibilities. We will use the inventory and FAIRness assessment to produce essential recommendations on annotation and description for different data profiles (e.g., spatial, genotype, phenotype, etc.). These recommendations will be accompanied by concrete agrosystem examples to illustrate the benefit of implementing established data standards, vocabularies, and ontologies. Additionally, we will provide a publication guideline and a metadata model based on standardized terminologies. These tools will be tested for data publication by the use cases and amended based on user feedback. Finally, we will work on integrating and extending existing vocabularies and ontologies. We will use mapping (i.e., convert a standard into another) and alignment (i.e., define relations between representations) to create a common language for data representation in agrosciences, making it easier for different data infrastructures to work together. The metadata model and the network of standards, vocabularies, and ontologies will be used in the design of FAIRagro services aimed at improving data search, integration, and sharing within the agrosystem community ([Measures 4.2](#) and [4.3](#)).

What we need from the UCs

- List of objects covered in UCs' research
- List of attributes/properties/variables of these objects
- Systemic terminologies, taxonomies, glossaries, method description
- Interface/information technology standards used in UCs' research domain

What we offer to the UCs:

- Recommendations on suitable standards and vocabularies covering specific UC needs
- Recommendations to apply these resources for data annotation and description
- Domain-specific publication guidelines

Measure 3.2: Standards for Data Management, FAIRness and Discoverability

Leads: Juliane Fluck, Gabriel Schneider (ZB MED), Daniel Martini (KTBL)

[Staff & contact](#)

In this measure, we aim to create community standards for improving data management practices, data FAIRness, and data visibility. Firstly, we will create templates for data management plans (DMPs) customized to meet the specific requirements of the use cases and the wider community. DMPs are documents that outline how researchers plan to manage data throughout their research project, from initial design to long-term preservation. DMP templates will be made available to researchers for testing and feedback. Secondly, we will extend the FAIRness assessment workflow developed in [Measure 3.1](#) (for standards, vocabularies, and ontologies) to datasets, including specific data quality measures developed in [Measure 3.3](#). Lastly, we intend to extend *schema.org*, a structured data vocabulary that can be read by search engines, to enhance the findability, and hence the visibility, of agrosystem data publications in repositories. This extension should integrate existing standards for biological data and additional definitions from existing vocabularies and ontologies pertaining to agrosience. We will collaborate with domain experts and service providers to ensure that the extension is widely adopted and user-friendly.

What we need from UCs

- Data management requirements of funders
- List of research objects and their associated attributes and specifications (see Measure 3.1)

What we offer to UCs:

- UC-specific data management plan templates
- A *schema.org* extension for Increased findability of datasets

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

Fitness-for-use

Lead: Markus Möller (JKI)

[Staff & contact](#)

Ensuring data quality is crucial for data reusability. Crop-related data varies in quality needs, with challenges like dependencies of time series data on modelling results, uncertainties related to data aggregation methods, the accuracy and completeness of phenotyping data, or the plausibility of long-term agricultural experiments. Despite the stakes, there is a lack of data quality standard criteria for agrosystem research. Using representative agrosystem datasets provided by the UCs, we will identify and formalize relevant data quality metrics, considering fitness for specific uses. Starting from generic quality aspects of agrosystem data, we will identify application-specific data quality standards. This will lead to the creation of data curation guidelines for the data stewards ([Measure 2.5](#)) and the definition of new metadata elements describing data quality that will be incorporated into the FAIRagro search portal ([Measure 4.3](#)).

What we need from UCs

- Exemplary data types (frequently used datasets)
- Data quality criteria (metrics; will be identified in UCxTA workshops & surveys)

What we offer to UCs

- Application-specific data quality curation guidelines
- Toolbox for assessing data quality using the identified metrics

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

Leads: Jan-Henrik Hauernert (UBN), Uwe Rascher (FZJ)

[Staff & contact](#)

Building upon [Measure 3.3](#), which focuses on defining data quality metrics, Measure 3.4 aims to create a toolbox (including algorithms, reference data, and a review system) to assess and enhance data content based on those quality metrics. For example, if "outlier rate" is a defined quality metric, we may offer an algorithm to detect and remove outliers from datasets. We will specifically address agrosystem robotics, precision farming, and in-field plant phenotyping, which requires the analysis of diverse sensor data. Ensuring data quality is essential before combining multi-source data for use in machine learning algorithms. Together with measure 3.1, we will first focus on creating guidelines to assist data contributors in addressing data quality concerns. We will then develop an algorithmic suite to automatically detect, quantify, and resolve quality issues such as outlier removal and data harmonization, and combine datasets overlapping in space and/or time. To appraise the accuracy of these automated quality control processes, we will build a review system that will enable experts to annotate data quality with the help of visualization tools. Last, we will devise methods to visualize harmonized sensor data on mobile devices for real-time decision-making in agrosystems.

What we need from UCs

- Reference data to identify data curation requirements

What we offer to UCs

- Guidelines for data quality curation for sensor data
- Algorithmic suite to perform sensor data quality control (e.g., outlier detection and removal, data harmonization...)
- Visualization tools for on-site decision support (mobile devices)
- Data quality review system in the FAIRagro portal

Measure 3.5: FAIR Workflows and FAIR Digital Objects

People: Claus Weiland (SGN), Daniel Martini (KTBL)

[Staff & contact](#)

Current scientific research is characterized by a so-called “data deluge”, an ever-increasing volume of heterogeneous data sources that becomes too large to be adequately curated by humans. As a result, the digital representations of research objects comprise increasingly scattered and diverse (meta)data linking to disparate repositories and services (Fig. 1a). FAIR digital objects (FDOs) enable interoperability among these components by binding permanent identifiers with a model layer detailing the FDO’s metadata, format, and standards, and a resource layer containing the research object itself (i.e., data, code, or other research output, with associated metadata). In Measure 3.5, we will develop a foundation for integrating specifications and guidelines from data vocabularies and ontologies identified in [Measures 3.1](#) and [3.2](#) into FAIR digital objects. We will design and implement a set of services (e.g., persistent identifier schemes, semantic validation pipelines) to enable the creation and use of such FDOs. Last, we will develop a framework to describe research data workflows as specific FDO types (Research Object Crates; Fig. 1b) that link input and output data with methodologies, specifications, and code, with persistent identifiers. Provenance annotation on data generated through these workflows will be documented as a metadata element. Ultimately, FDOs and RO-crates will be used as a basis for facilitating the integration of research data sources ([Measure 4.2](#)) and data processing workflows ([Measure 4.4](#)) into the FAIRagro service architecture.

What we need from UCs

- FAIR reference datasets and associated research outputs (e.g., method descriptions, code-based data processing workflows)

What we offer to UCs

- A framework to link related research objects (data, code, workflows...) compliant with the FAIR principles
- Basis for the implementation of data processing and analysis workflows into the FAIRagro scientific workflow infrastructure (Measure 4.4)

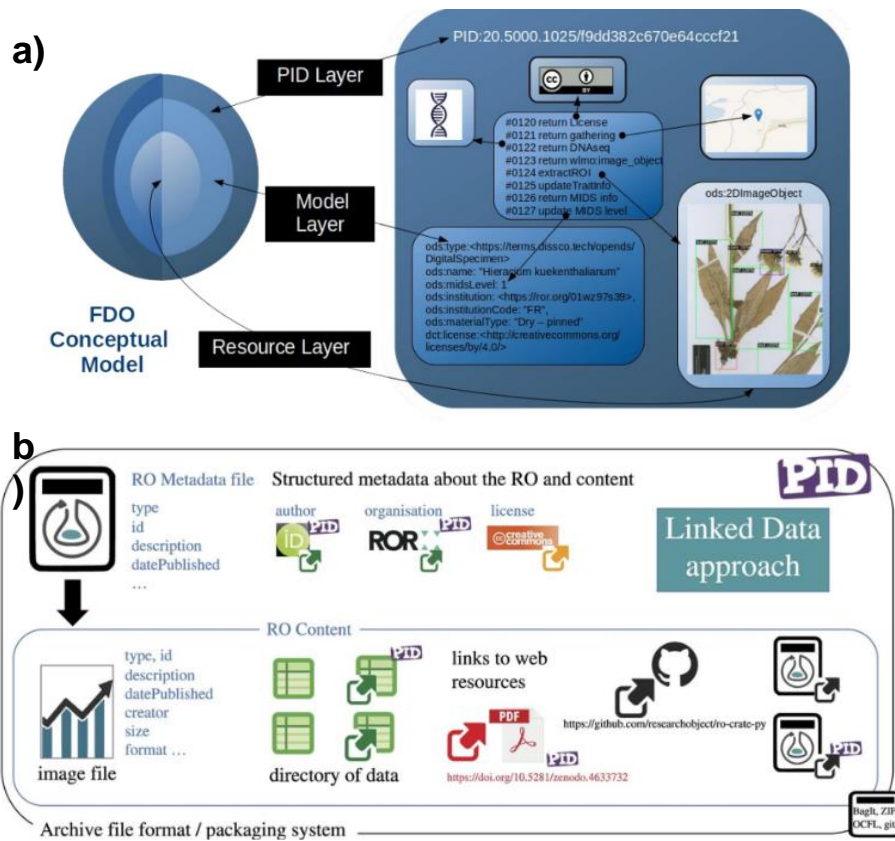


Figure 1. Conceptual representation of (a) a FAIR Digital Object (FDO), (b) a Research Object Crate (RO-Crate; Figure from Soiland-Reyes. 2022. *Data Science* 5:97-138)

Measure 3.6: Legal Framework and Machine-Actionable Policies

Leads: Franziska Boehm, Lea Singson (FIZ), Stephan Lesch (SGN)

[Staff & contact](#)

Measure 3.6 focuses on establishing a legally sound framework for sharing agrosystem data within FAIRagro to enhance data accessibility and compliance. We will conduct an initial assessment to examine current legal challenges faced by FAIRagro partners. The results will be discussed in a workshop where we will initiate a permanent working group to establish a community-driven legal framework. We will develop data policies and legal guidelines taking into account the specific interests of the different use cases. We will define legal metadata standards focused on the FAIR principles that will encompass legal categories relevant to agrosiences (e.g., licensing, trade secrets, data protection law, sensitive data, etc.). These standards will be discussed with the community and refined based on the gathered feedback. Data policies, legal guidelines, and metadata standards will be published on the FAIRagro portal and integrated into the FAIRagro service infrastructure. Ongoing development in data law will be consistently monitored the their implication for FAIRagro analysed. We will implement relevant updates to the legal framework and communicate them in a regular newsletter with the help of [Measure 2.1](#). For handling proprietary data, we will create rules for data usage permission and restrictions and integrate them into machine-actionable policies that will be linked to their associated data through an FDO extension ([Measure 3.5](#)). With the integration of FDOs in the FAIRagro infrastructure ([Measure 4.2](#)), specific data access policies can hence be enforced and used as filters in data searches.

What we need from UCs

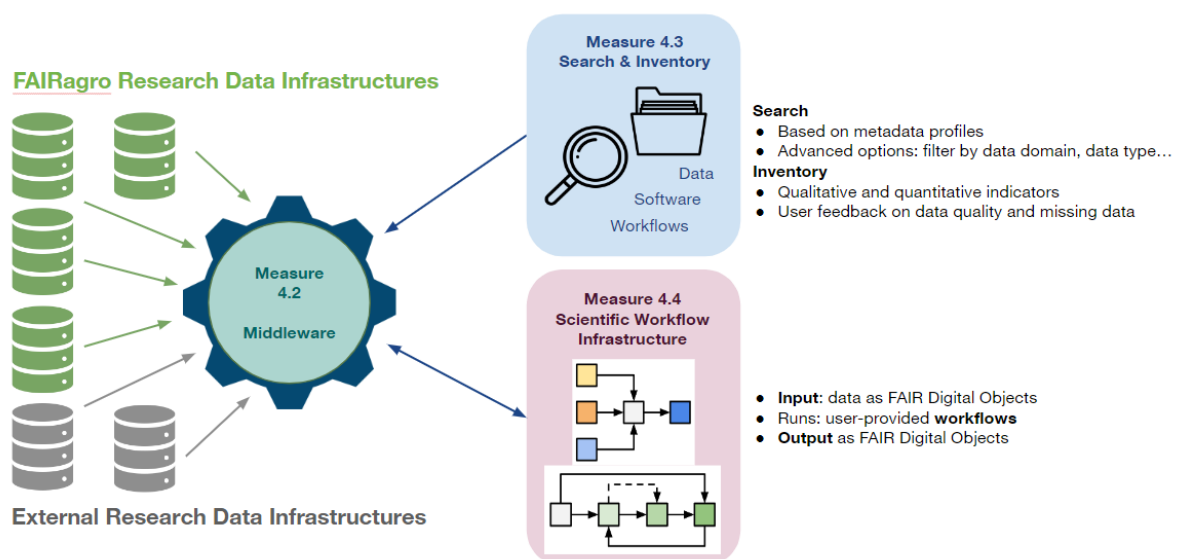
- Domain-relevant legal categories (e.g., sensitive, personal, proprietary data...)

What we offer to UCs

- Legal metadata standards adapted to the UCs' context
- Regular updates on the development of data protection laws via a newsletter
- Integration of legal standards and rules in the FAIRagro infrastructure, allowing filtering search on data usage rights and regulating access

Task Area 4. Services

TA4 has the responsibility of developing infrastructure services for a federated research data management system, aligning with the requirements set forth by the UCs and building upon the standards established in TA3. The primary purpose of this system is to facilitate a research data lifecycle that adheres to FAIR principles for research data management. At the core of this objective is the creation of an interconnected network of research data infrastructures ([Measure 4.2](#)). This network will serve as a central access point to agrosystem data via the FAIRagro search portal, with functionalities for advanced search and data quality appraisal ([Measure 4.3](#)). We will implement a common authentication and authorization infrastructure to ensure compliance with legal constraints on data access, provide. To promote FAIR-compliant practices in the later stages of the research data lifecycle, we will design a framework for hosting and executing reusable data processing and analysis workflows provided by the UCs ([Measure 4.4](#)). Besides, TA4 will oversee the operation of central services to support consortium activities, including a project management platform, a research data management organizer, and a helpdesk system ([Measure 4.1](#)). All our services will be designed with the aim of ensuring technical compatibility for potential adoption in the NFDI's cross-cutting infrastructure.



Measure 4.1. Central Services for the FAIRagro community

Leads: Xenia Specka (ZALF)

[Staff & contact](#)

Measure 4.1 will provide central services to support the activities of the FAIRagro consortium. The central services will comprise four main services:

(1) The FAIRagro Portal will be a central access point for supporting material, data searches, and user requests. We will identify user requirements to develop a concept for the portal. Collaboration with other measures will ensure the integration of features like information dissemination and user interaction ([Measure 2.1](#)), user feedback and performance indicators ([Measure 2.2](#) and [Measure 4.3](#)), submission and onboarding of new use cases ([Measure 2.3](#)), education and training materials ([Measure 2.4](#)), and data search ([Measure 4.3](#)).

(2) A project management platform will be established to facilitate progress control and exchange between members of different institutions within the consortium. The platform will also support open-source software development by including a source code development environment for FAIRagro partners.

(3) A helpdesk with a ticket system and customer relationship management system (CRM) will be set up on the FAIRagro Portal to collect, process, and respond to user inquiries. Data stewards ([Measure 2.5](#)) will manage the helpdesk, with provided templates, documentation, tutorials, and training for efficient user support.

(4) A research data management software (RDMO) instance will be established to help create Data Management Plans (DMPs) for the agrosystem community. The instance will be embedded with data protection measures, and supported by subject-specific questionnaires on data management requirements submitted to the UCs ([Measure 3.2](#)).

What we need from the UCs

- Input on tools and services required for optimal project management

What we offer to the UCs:

- Central access to all FAIRagro internal services
- A helpdesk with a ticket system to collect, process, and respond to user questions/requests
- A project management platform to document and report progress within your UC

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

Leads: Daniel Arend (IPK), Stefan Lesch (SGN), Xenia Specka (ZALF)

[Staff & contact](#)

The primary goal of FAIRagro is to enhance agrosystem research data management, making it faster and easier for scientists to find, access, and share data. In Measure 4.2, we are developing a network that connects various research data sources in agrosystem sciences, such as BonaRes, e!DAL-PGP, DWD Open Data Server, and many others. This network will act as a bridge between different data systems, allowing for smooth information exchange and data access according to the FAIR principles (Findable, Accessible, Interoperable, and Reusable). It will be fitted with a user-friendly interface allowing central and seamless data search across all connected repositories ([Measure 4.3](#)). Data protection will be ensured by secure authentication and authorization mechanisms for remote data access. In the future, we aim to link the FAIRagro network to other NFDI consortia to expand data sharing across different research fields and initiatives.

Table 1. Implementation phases for interlinking the infrastructures with the middleware service

Phase 1 (2023-2024)	Phase 2 (2025-2026)	Phase 3 (2027)
<ul style="list-style-type: none"> • e!DAL-PGP (IPK) • BonaRes (ZALF) • OpenAgrar (JKI) • National Soil & Forest Inventories (Thünen) • PUBLISSO Repository for Life Sciences (ZB MED) 	<ul style="list-style-type: none"> • PhenoRob DB (UBN) • PlabiPD (FZJ) • Edaphobase (SGN) • GBIS/I, LIMS (IPK) • JKI-DataCube (JKI) • SRADI (TUM) • BonaRes Knowledge Library (UFZ) • Open Data Server (DWD) 	<ul style="list-style-type: none"> • Cross-NFDI infrastructures <p><i>Additional infrastructures from agrosystem community</i></p>

What we need from the UCs

- List of repositories and data types of interest for your project

What we offer to the UCs:

- A central access point to agrosystem (meta)data hosted on partner repositories

Measure 4.3: Searchable Inventory of Services and Data

Leads: Juliane Fluck (ZBMED), Björn Usadel (FZJ)

[Staff & contact](#)

While there are many high-quality agronomy-related resources available in Germany, they are currently not searchable from one central place, making data search a daunting task. Measure 4.3 focuses on creating a user-friendly to query the network of interlinked data infrastructure developed in [Measure 4.2](#). With this service, agrosystem scientists will be able to find datasets and infrastructures relevant to their research effortlessly. During the development process, early prototypes will be shared with partner scientists (FAIRagro use cases) for testing. Based on user feedback, updates with more features and content will be provided regularly. The final versions will be operated by ZB MED beyond the project phase, ensuring continued support for the research needs of the wider agrosystem community.

What we need from the UCs

- A list of metadata fields of use in the search process for your focal data types
- Needs in terms of filtering and advanced search functionality

What we offer to the UCs:

- A user-friendly unique interface to search data and seek data access across all FAIRagro repositories

Measure 4.4: Scientific Workflow Infrastructure (SciWIn)

Lead: Harald von Waldow (Thünen), Patrick König (IPK)

[Staff & contact](#)

Reproducibility in computational research is vital for efficient collaboration, verifying results and ensuring transparency. Yet it remains challenging due to complex workflows, inconsistent data management and the reliance on specific software environments.

SciWIn aims to provide tools and infrastructure that extend re-usable data handling to re-usable handling of computational pipelines. We implement SciWIn as **SciWIn-Client** "s4n" https://github.com/fairagro/m4.4_sciwin_client and **SciWIn-Hub** (concept stage).

SciWIn-Client is designed to pick up scientists right at the *in silico* workbench, where iterative and highly interactive processes such as data extraction, cleaning, visualization, exploration, analysis and transformation are carried out. It is a command-line tool designed to easily create, record, annotate and execute computational workflows. What Git does for versioning, **s4n** does for provenance management. From simple one-step calculations to complex multi-branch pipelines, **s4n** records the chain of provenance for data and code artifacts. These records can be re-executed, also on remote computers. The individual artifacts and computational steps form a graph which can be annotated with semantic metadata. **s4n** also supports this annotation. **s4n** can package the resulting workflow as *Workflow RO-Crate* and publish it through [WorkflowHub](#).

SciWIn-Hub is planned as a web-based service that accepts workflows as created by **s4n**. It will be able to merge and partition workflow-graphs. The user can interact with SciWIn-Hub to store, share, visualize, explore, browse, search and download computational workflows.

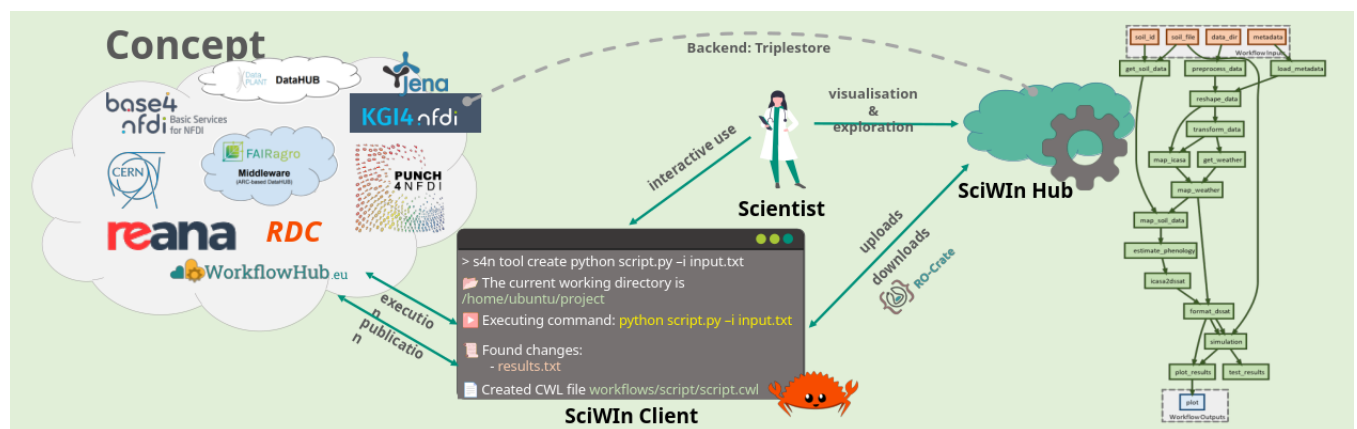


Figure 2. SciWIn concept

What we need from the UCs:

- The need to develop computational workflows, e.g. data download, quality-control, cleaning, transformation, analysis, visualization, ... in an interactive manner, also with custom scripts that are developed throughout the process.
- Scientists that work on the command-line.
- Willingness to try out alpha-software, that might, initially and in the short run, cause more work than it saves.

What we offer to the UCs:

- The chance to be an early adopter of game-changing tooling.
- First-class support and help with workflow development.
- The ability to get specific features into SciWIn.
- Re-usability, organization and management for scripts and data leading to more efficient research in the long run.
- Compute power: SciWIn workflows will be executable on high performance compute instances and clusters.
- Easy publishing of workflows, either alone or as part of paper- or data-publications.