



FAIRagro: Ein Konsortium in der Nationalen Forschungsdateninfrastruktur (NFDI) für Forschungsdaten in der Agrosystemforschung

Herausforderungen und Lösungsansätze für den Aufbau einer FAIRen Forschungsdateninfrastruktur

Xenia Specka¹ · Daniel Martini² · Claus Weiland³ · Daniel Arend⁴ · Senthold Asseng⁵ · Franziska Boehm⁶ · Til Feike⁷ · Juliane Fluck⁸ · David Gackstetter⁵ · Aida Gonzales-Mellado⁹ · Thomas Hartmann⁶ · Jan-Henrik Haurert¹⁰ · Florian Hoedt⁹ · Carsten Hoffmann¹ · Patrick König⁴ · Matthias Lange⁴ · Stephan Lesch³ · Birte Lindstädt⁸ · Gunnar Lischeid¹ · Markus Möller⁷ · Uwe Rascher¹¹ · Jochen Christoph Reif⁴ · Markus Schmalzl¹² · Matthias Senft⁷ · Ulrike Stahl⁷ · Nikolai Svoboda¹ · Björn Usadel¹¹ · Heidi Webber¹ · Frank Ewert¹

Angenommen: 10. Dezember 2022 / Online publiziert: 24. Januar 2023
© Der/die Autor(en) 2023

Zusammenfassung

FAIRagro ist ein Konsortium in der Nationalen Forschungsdateninfrastruktur (NFDI) in Deutschland um Forschungsdaten der Agrosystemforschung FAIR – d. h. auffindbar (F), zugänglich (A), interoperabel (I) und für andere Forschende domänenübergreifend nachnutzbar (R) zu machen. In der deutschen Forschungslandschaft rund um nachhaltige Agrosysteme werden heterogene Forschungsdaten erhoben und nur zum Teil in existierenden Forschungsdatenrepositorien veröffentlicht. Das Spektrum der Datenformate erstreckt sich beispielsweise von Laborergebnissen, über Satellitenbilder bis hin zu qualitativen Interviews mit Landwirt:innen. Um diese Daten zukünftig für Forschende verschiedener Disziplinen besser auffindbar und nachnutzbar zu machen, wird FAIRagro eine Forschungsdateninfrastruktur (FDI) für die Agrosystemforschung einrichten, in der disziplinäre Dateninfrastrukturen miteinander verknüpft werden. Spezifische Herausforderungen im Forschungsdatenmanagement (FDM) fachlicher Disziplinen wie Pflanzenzüchtung, integrierter Pflanzenschutz oder Agrarrobotik werden als Use Cases in FAIRagro adressiert und für diese Lösungen entwickelt. Darüber hinaus wird FAIRagro ein Netzwerk aus direkten Ansprechpersonen für Fragen zum Forschungsdatenmanagement in der Agrosystem-Community bereitstellen. In Übereinstimmung mit den Zielsetzungen der NFDI und der European Open Science Cloud ist FAIRagro aktiv an der konzeptionellen Implementierung eines interoperablen Datenraums beteiligt.

Aktuelle Herausforderungen in der Agrosystemforschung

Herausforderungen für die Landwirtschaft und die Agrarforschung

Die Landwirtschaft steht vor enormen Herausforderungen von großer gesellschaftlicher Relevanz, wie z. B. dem Kli-

mawandel mit Zunahme von Extremereignissen wie Dürre- und Hitzeperioden, der zunehmenden Verknappung und Degradierung natürlicher Ressourcen, dem beschleunigten Verlust der biologischen Vielfalt bis hin zur wachsenden Ernährungsunsicherheit und zunehmenden Herausforderungen im ländlichen Raum. Angesichts dieser Herausforderungen wurde die Umgestaltung der Landwirtschaft zur Erreichung der UN-Nachhaltigkeitsziele [1] in verschiedenen hochrangigen politischen Dokumenten hervorgehoben, z. B. in der Farm-to-Fork-Strategie des Europäischen Green Deal [2] und im Abschlussbericht der Nationalen Kommission zur Zukunft der Landwirtschaft [3]. Um gesellschaftlich tragfähige und wirksame Lösungen für diese Herausforde-

✉ Daniel Martini
d.martini@ktbl.de

Ausführliche Informationen zu den Autoren befinden sich auf der letzten Seite dieses Artikels.

lungen zu finden, sind integrierte Systemansätze erforderlich, die Kompromisse und Synergien zwischen Zielen, Interessengruppen, Regionen und Sektoren ausdrücklich berücksichtigen.

Agrosystemforschung

In Europa und Deutschland werden etwa 50% der Fläche für den Anbau von Pflanzen zur Erzeugung von Nahrungs- und Futtermitteln sowie Rohstoffen landwirtschaftlich genutzt. Gleichzeitig wird die Notwendigkeit, die Widerstandsfähigkeit und Nachhaltigkeit der Pflanzenproduktion zu verbessern und die negativen Auswirkungen auf Ökosysteme und Umwelt zu verringern, auch von Öffentlichkeit und Politik widerspiegelt. Spezifische Maßnahmen zur Verringerung des Pestizid- und Düngemittleinsatzes, Ausbau des ökologischen Landbaus, Kohlenstoffsequestrierung im Boden, Verbesserung der biologischen Vielfalt und Ökosystemleistungen in Agrarlandschaften müssen jedoch weiterentwickelt und umgesetzt werden. Die Entwicklung praktikabler Lösungen für diese Ziele erfordert einen agrosystemischen Ansatz, der eine Vielzahl an Disziplinen (z.B. Genetik, Züchtung, Pflanzen-, Boden- und Geowissenschaften) umfasst, und der Wechselwirkungen mit anderen Bereichen der Agrarwissenschaften (z.B. Tierwissenschaften, Molekularbiologie, Ökologie, Agrarökonomie, Sozial- und Kulturwissenschaften und Ernährungswissenschaften) und darüber hinaus berücksichtigt (Abb. 1).

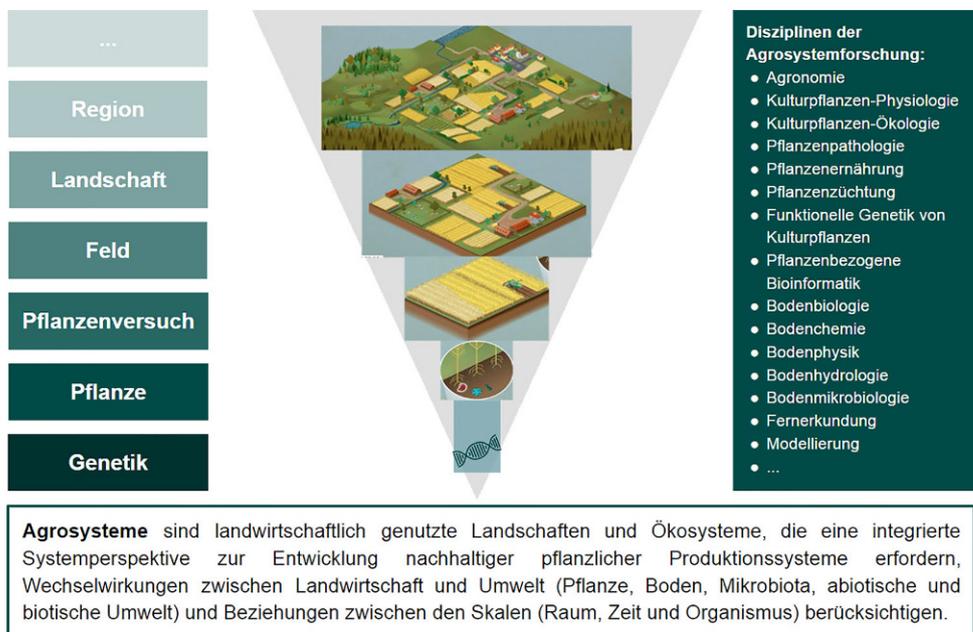
Zentrale Herausforderungen

In der Agrosystemforschung leiten sich eine Reihe zentraler Anforderungen für das Forschungsdatenmanagement sowie der Datenbereitstellung und -nutzung ab, die in den folgenden drei Beispielen kurz dargestellt werden.

Beispiel 1 „Pflanzenzüchtung“: Die Züchtung angepasster Pflanzensorten nutzt sowohl Daten zur Merkmalsausprägung in Pflanzenbeständen (Phänotyp) als auch Daten zur zugrunde liegenden Genetik (Genotyp), um resistente und ertragsstabile Sorten zu ermitteln, die unter den Bedingungen des Klimawandels und der zunehmenden Klimavariabilität mit extremen Hitze- und Dürre-Ereignissen besser umgehen können. Genotypische Informationen und insbesondere phänotypische Daten aus Feldversuchen sind häufig nur unzureichend beschrieben und/oder öffentlich nicht verfügbar. Dies hemmt die Durchführung von vertieften Studien zu Wechselwirkungen zwischen Genotypen und Umwelten und führt letztendlich zu einer Diskrepanz zwischen dem Selektionserfolg in experimentellen Umgebungen und landwirtschaftlichen Betrieben.

Beispiel 2 „Bodenkunde“: Die Forschung zur Stickstoffdynamik und Kohlenstoffbindung in Böden [4] ist ein wichtiges Element der menschlichen Reaktion auf den Klimawandel. Die Veränderung des Kohlenstoffgehalts im Boden ist ein lang anhaltender Prozess, der nur über Jahrzehnte hinweg gemessen werden kann. Zeitreihen und Altdaten aus landwirtschaftlichen Langzeitfeldversuchen (LTE) zwischen 20 und 180 Jahren sind für die Beantwortung solcher Forschungsfragen von enormer Bedeutung. Allerdings sind Forschungsdaten aus LTE derzeit schwer zu finden, da sie kaum veröffentlicht sind und standardisierte Beschrei-

Abb. 1 Agrosystemforschung und die von ihr umfassenden Skalen und Disziplinen



bungen fehlen. Oftmals finden sich Informationen und Daten über LTEs nur in wissenschaftlichen Publikationen, und wichtige Zusammenhänge zu entsprechenden Boden-, Klima- oder Umweltdaten fehlen.

Beispiel 3 „Integrierter Pflanzenschutz“: Die Farm-to-Fork-Strategie [2] der Europäischen Kommission zielt darauf ab, den Einsatz und das Risiko von Pflanzenschutzmitteln bis 2030 um 50 % zu reduzieren. Dazu ist eine umfassende und wirksame Umsetzung des integrierten Pflanzenschutzes (IPM) von entscheidender Bedeutung, um massive Ertragseinbußen aufgrund von Schädlingen und Krankheiten zu vermeiden und die Ernährungssicherheit und eine hohe Ressourcennutzungseffizienz zu gewährleisten. Ein verbessertes Forschungsdatenmanagement (FDM) in dieser Disziplin ist dringend erforderlich, um eine effektivere Forschung und Entwicklung für den integrierten Pflanzenschutz zu ermöglichen. Die größten Herausforderungen im FDM hierbei sind (1) Nichteinheitliche Versuchsdesigns bei Ertragsverlustversuchen zum IPM sowie in den Verfahren zur Bewertung von Krankheiten, (2) unzureichende Informationen über das Vorhandensein und die potenzielle Zugänglichkeit von spezifischen IPM-Daten in Deutschland und (3) fehlende integrierte Entscheidungshilfen für den Pflanzenschutz, die potenzielle Ertragsverluste und das Umweltrisiko des Pestizid-Einsatzes berücksichtigen.

Der agrosystemische Forschungsansatz erfordert darüber hinaus die Verwendung von Daten aus angrenzenden Fachgebieten wie den Geowissenschaften oder der Meteorologie. Die Wiederverwendbarkeit von Forschungsdaten über Fachgebietsgrenzen hinweg ist daher eine der Kernanforderungen aus der Agrosystemforschung an eine Nationale Forschungsdateninfrastruktur.

Datenpublikation und -nutzung in der Agrosystemforschung heute

Die von der Wissenschaft geforderte kollaborative, interdisziplinäre und nachhaltige Analyse und Wiederverwendung von Daten aus der Agrosystemforschung ist bisher nicht möglich. Obwohl mittlerweile verschiedene Repositorien zur Datenveröffentlichung existieren, ist eine repositorienübergreifende, systematische Suche kaum möglich. Forschungsdaten sind nicht einheitlich beschrieben, sodass zusätzliche Aufwände notwendig sind, um überhaupt vorhandene Inhalte und Strukturen zu erfassen. Schlussendlich sind die meisten Datensätze auch nicht einfach integrier- und zusammenführbar, da Konventionen zur eindeutigen Bezeichnung von Datenfelder und -einträgen fehlen oder nicht berücksichtigt werden. Die Veröffentlichung von standardisierten und harmonisierten Forschungsdaten aus der Agrosystemforschung, die mit umfangreichen Metadaten zu Forschungshintergrund, Datenqualität und -herkunft

beschrieben werden, steht noch aus. Bei der Veröffentlichung von Daten herrscht bei Forschenden derzeit große Unsicherheit über verfügbare Datenrepositorien sowie über Anforderungen an die Datenpublikation [5]. Darüber hinaus wird die Nachnutzung von Forschungsdaten durch die Vielfalt existierender Datenformate, räumlicher und zeitlicher Auflösungen und Einheiten sowie einen Mangel an Informationen über die Qualität und Plausibilität der Daten behindert.

Repositorien in der Agrosystemforschung

In den letzten zehn Jahren hat die Nutzung von domänenspezifischen Infrastrukturen (z. B. das BonaRes Repository für Boden- und Agrarforschungsdaten [6, 7]), institutionellen Repositorien (z. B. GBIS/I [8], e!DAL-PGP [9]) und allgemeinen Publikationsdiensten (z. B. OpenAgrar [10, 11], PUBLISSO Repository der Lebenswissenschaften [12]) an Bedeutung gewonnen [13–16]. Obwohl verschiedene qualitativ hochwertige agrarbezogene Datenrepositorien zur Verfügung stehen, gibt es keinen zentralen Suchdienst und keine Übersicht an verfügbaren Repositorien, um einfach relevante Datensätze aus der Agrosystemforschung zu finden. Die Forschungsdateninfrastrukturen unterstützen meist ein bestimmtes Metadatenchema, allerdings wird die Durchführung von interdisziplinären Analysen durch fehlende Interoperabilität zwischen den Infrastrukturen bzgl. des Informationsaustauschs sowie des maschinenlesbaren Zugangs zu den (Meta-)Daten erschwert. RESTful Services [17] werden häufig verwendet, implementieren aber heterogene Datenstrukturen und Endpunkt-Authentifizierungs- und Autorisierungsschemata, was die spätere Nutzung der Daten erheblich erschwert. Eine komplette Liste aller einzubindenden Repositorien ist auf der FAIRagro Webseite [18] verfügbar.

Metadaten, Ontologien und Standards

Die Vielfalt der Teildisziplinen, Betrachtungsebenen und verarbeiteten Datentypen in der Agrosystemforschung spiegelt sich in der Heterogenität der verwendeten Daten- und Metadatenstandards wider. Mittlerweile kommen zwar auch anerkannte Metadatenstandards auf Basis generischer und erweiterbarer Modelle zum Einsatz (z. B. DataCite [19], Dublin Core [20], DCAT [21], W3C PROV [22] und W3C VoID [23]). Es wurden jedoch auch domänenspezifische Standards geschaffen, die nur unzureichend in Informationsmodelle dieser allgemeineren Standards integrierbar sind.

In den Pflanzenwissenschaften existieren verschiedene Metadatenstandards z. B. die Multi-Crop Passport Descriptors (MCPD), die die Herkunft von genetischen Ressourcen beschreiben [24], und die vom International Consortium for

Agricultural Systems Applications (ICASA) entwickelten Standards [25], die überarbeitet [26] und zu MIAPPE (Minimum Information About Plant Phenotyping Experiments) weiterentwickelt wurden [27]. Die Geodatengemeinschaft hat in den letzten Jahren eine Reihe von Metadatenstandards etabliert, z. B. ISO 19115 Geographic information – Metadata, der durch die europäische INSPIRE-Richtlinie (Infrastructure for Spatial Information in Europe) erweitert wurde und zusätzliche Implementierungsregeln enthält, oder das BonaRes-Metadaten-Schema [28].

Herkunftsinformationen sind für die Bewertung der Datenqualität und die Verbesserung der Reproduzierbarkeit, Interpretierbarkeit und Bewertung der Gebrauchstauglichkeit von Forschungsdaten unerlässlich. Beobachtungsdaten erfordern z. B. eine Beschreibung der verwendeten Sensoren, für Daten aus Simulationsmodellen sind Informationen zu Eingabedaten, Einstellungen und betrachteten Szenarien erforderlich. Solche Metadaten werden bislang oft nicht bereitgestellt oder sind nur als textuelle Beschreibungen verfügbar. Die Verwendung gemeinsamer kontrollierter Vokabulare, dauerhafter Identifikatoren für Daten, Mitwirkende, Institutionen, Instrumente und Sensoren ist erforderlich, um Provenance-Informationen zu erstellen.

Semantische Ressourcen, die mit oben genannten generischen Standards kompatibel sind und erlauben, Daten z. B. über einheitliche URIs (Uniform Resource Identifier) zu vernetzen, werden in den Agrosystemwissenschaften teilweise bereits verwendet. So enthält der AGROVOC-Thesaurus der FAO (Food and Agriculture Organization of the United Nations) aktuell mehr als 40.000 Konzepte, die unter Nutzung des Simple Knowledge Organization System (SKOS) des W3C beschrieben sind und über verschiedene Dienste bereitgestellt werden [29]. Die zugehörige Agrontology [30] modelliert Zusammenhänge wie Ursache-Wirkungs-Beziehungen oder Pflanze-Produkt-Beziehungen. In Kombination mit entsprechenden Metadatenannotationen können so auch gezielte Abfragen zu möglichen Zusammenhängen zwischen Datensätzen durchgeführt werden. Weitere zum Einsatz kommende Ontologien sind beispielsweise die der OBO Foundry (Agronomy Ontology, Environment Ontology, Plant Ontology, Plant Trait Ontology, Experimental Conditions Ontology, Crop Ontology [31–34]) sowie Genomik- und Proteomik-Ontologien wie die Gene Ontology oder Protein Ontology [35].

Datenqualität und Plausibilität

Eine große Herausforderung zur Nachnutzung von Daten stellt aktuell der Mangel an Informationen bzgl. der Qualität veröffentlichter Daten dar. Wissenschaftler müssen oft umfangreiche inhaltliche Prüfungen der Daten vornehmen, um eine Aussage über die Eignung veröffentlichter Daten in ihrem Forschungskontext treffen zu können. Es fehlen

Standards zur Beschreibung der Datenqualität unter Beachtung der Anforderungen der Fachdisziplinen, sowie Informationen zur Historie und Provenienz von Daten und Web-Ressourcen [36]. Es werden Standards und Werkzeuge zur kontextbasierten Kuratation zur Beschreibung der Datenqualität benötigt, die sich sowohl auf methodische Aspekte der Datenerfassung und -erzeugung als auch auf zusätzliche Informationen über die Eignung und Tauglichkeit der Daten für die weitere Nutzung beziehen [37].

FAIRagro – kurz und knapp

FAIRagro ist ein Konsortium in der Nationalen Forschungsdateninfrastruktur (NFDI) für den Bereich der Agrosystemforschung, um institutionenübergreifende, interdisziplinäre Forschung durch eine verbesserte Forschungsdateninfrastruktur zu erleichtern.

Nationale Forschungsdateninfrastruktur (NFDI)

Die GWK (Gemeinsame Wissenschaftskonferenz) hat am 26. November 2018 den Aufbau und die Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) beschlossen, die von Bund und Ländern in der Zeit von 2019–2028 jährlich mit 90 Mio. € unterstützt wird. Die NFDI bringt verschiedene Akteure durch den Aufbau von Konsortien unterschiedlicher wissenschaftlicher Disziplinen zusammen, um langfristig Datendienste für die gesamte Forschungsgemeinschaften bereitzustellen. Das Ziel der NFDI in Deutschland ist eine nachhaltige, qualitative und systematische Sicherung, Erschließung und Nutzbarmachung von Forschungsdaten für die gesamte Breite der wissenschaftlichen Disziplinen. Als Basis dienen dabei die FAIR-Prinzipien [38], die in vier übergreifenden Kategorien (die insgesamt 15 Prinzipien umfassen) Maßnahmen beschreiben, um Daten auffindbar (Findable), zugänglich (Accessible), interoperabel (Interoperable) und wiederverwendbar (Reusable) zu machen. Schlüsselaspekt der FAIR-Prinzipien ist es, Datensätze durch Annotation und Anreicherung mit spezifischen Metadaten sowohl besser für die Beurteilung durch Menschen als auch für die (semi-)autonome Prozessierung durch Software-Agenten (sogenannte Machine-Actionability) zu erschließen [39]. Damit stellen sie die Leitlinie für das Ziel der NFDI dar, eine Forschungsdateninfrastruktur zu etablieren, die Forschungsdaten domänenübergreifend in bestehende internationale Initiativen und Datenföderationen wie der EOSC (European Open Science Cloud [40, 41]) integriert.

Ziele

FAIRagro hat das Ziel, eine stärkere Vernetzung der verschiedenen Disziplinen, Skalen und Methoden in der Agrosystemforschung durch ein kollaboratives Forschungsdatenmanagement voranzutreiben. Dabei werden verschiedene FDM Services unter Berücksichtigung aktueller Entwicklung in der NFDI und der internationalen Gemeinschaft für die Agrosystemforschung entwickelt. FAIRagro hat u. a. folgende Ziele:

- Einrichtung des FAIRagro-Portals als zentralen Zugangspunkt für das FDM in der Agrosystemforschung, um Wissensaustausch, Technologietransfer und partizipative Prozesse zur Förderung des Dialogs mit unserer Community zu ermöglichen.
- Verbesserung der Auffindbarkeit veröffentlichter Forschungsdaten und bestehender Datenrepositorien sowie Bereitstellung disziplinspezifischer interoperabler Infrastrukturen für einen standardisierten Datenaustausch.
- Etablierung von Standards, Leitlinien und Handlungsempfehlungen für ein FAIRes FDM für Forscher und Betreiber von Forschungsdateninfrastrukturen.
- Sicherstellung der Qualität von Forschungsdaten durch domänenspezifische Maßnahmen der Qualitätskontrolle
- Ermöglichung reproduzierbarer Forschungsergebnisse, z. B. durch Veröffentlichung von Modellen und Datenanalyseworkflows

Partnereinrichtungen

Das FAIRagro-Konsortium repräsentiert einen großen Teil der Agrosystemforschungsgemeinschaft in Deutschland, einschließlich typischer Datenanbieter und (Nach-)Nutzer sowie Anbieter wichtiger agrarwissenschaftlicher Forschungsdateninfrastrukturen. Die FAIRagro-Partner repräsentieren ebenfalls Forschungseinrichtungen, Universitäten und Fachhochschulen, Bundes- und Landesforschungsinstitute, Infrastruktureinrichtungen und Berufsverbände im Bereich der Agrosystemforschung (Abb. 2).

Einbindung der Agrosystem-Community

In Zusammenarbeit mit Agrosystemforscher:innen und Datenanalyseexpert:innen wurden aktuelle Herausforderungen im FDM identifiziert, die in Form von sechs Use Cases (UC) in FAIRagro adressiert werden. Diese Use Cases umfassen verschiedene Forschungsfragen, Methoden, Disziplinen und Skalen der Agrosystemforschung (Abb. 3).

- *UC 1 – Nutzung der Wechselwirkungen zwischen Genotyp, Standort, Jahr und Bewirtschaftung für eine nachhaltige Pflanzenproduktion:* Herausforderungen in der Züchtung von Nutzpflanzen; Etablierung von Datenmanagementprozessen, um Interaktionen zwischen Genotyp, Standort, Jahr und dem Management zu berücksichtigen;
- *UC 2 – Bewertung von Trade-offs für ein optimales Stickstoffmanagement bei Pflanzen:* Herausforderungen pro-

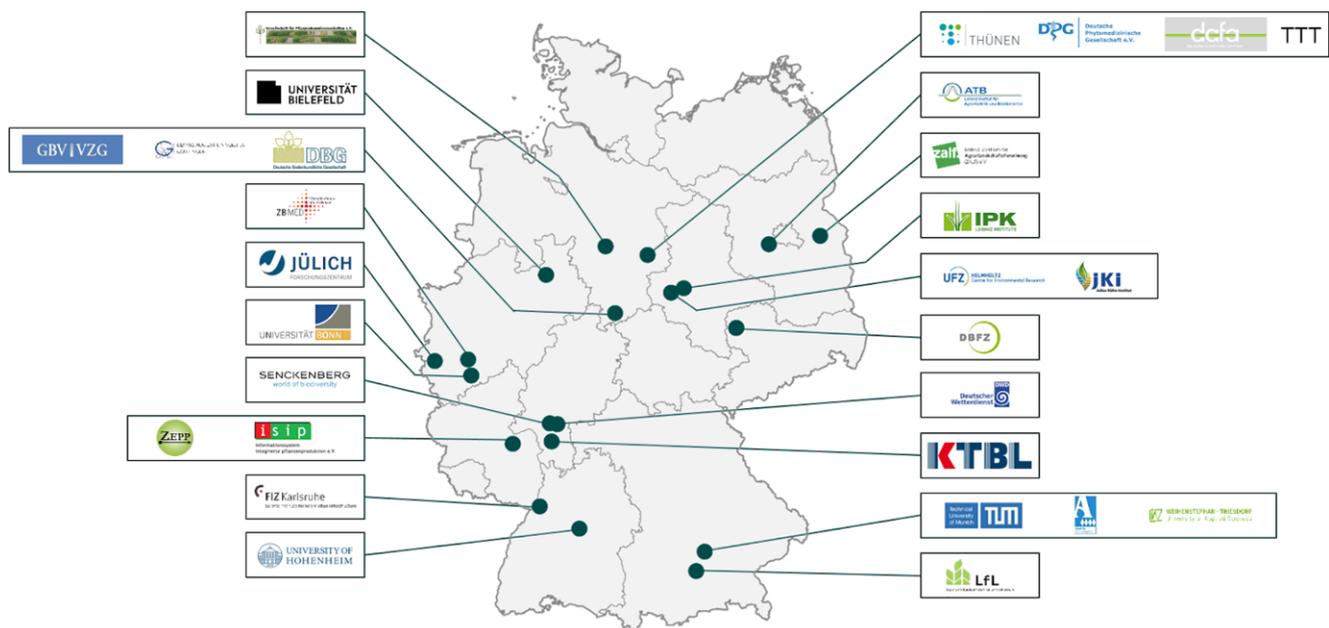


Abb. 2 Karte der Partnereinrichtungen von FAIRagro, die durch ihre jeweiligen Logos repräsentiert sind

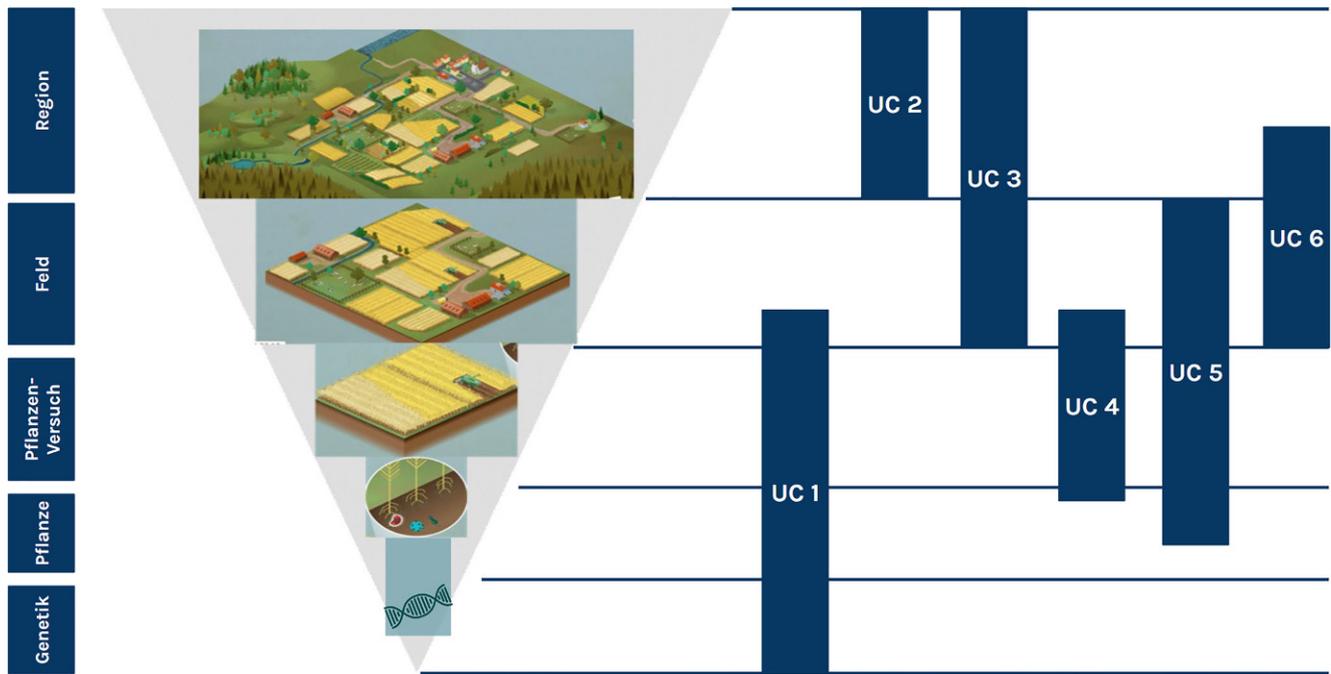


Abb. 3 FAIRagro Use Cases und ihre Einordnung in Skalenebenen

zessbasierter Pflanzenmodellierung vor dem Hintergrund der Optimierung des Stickstoffesatzes;

- UC 3 – *Optimierung von Schädlings- und Krankheitsdaten zur Förderung des integrierten Pflanzenschutzes*: datenspezifische Herausforderungen beim Management von Schädlingen und Krankheiten;
- UC 4 – *Lernen aus unvollständigen Daten*: Herausforderungen im Umgang mit unvollständigen Daten aus Langzeitfeldexperimenten;
- UC 5 – *Nichtinvasive Phänotypisierung mit autonomen Robotern*: Potenzial von multimodalen Datenanalysemethoden und Algorithmen des maschinellen Lernens für die Phänotypisierung von Pflanzen im Feld;
- UC 6 – *Automatisierte Datenflüsse für Pflanzenmodelle*: Datenprobleme bei der Kalibrierung und Anwendung von Ertragsmodellen;

Zur Bearbeitung dieser Herausforderungen wurden entsprechende Partner in FAIRagro involviert, um gemeinsam im Konsortium Lösungen zu entwickeln. Für eine weitere Einbindung der Community ist ein Onboarding-Prozess zur Integration weiterer Use Cases während der Laufzeit von FAIRagro geplant. Damit werden sowohl weitere Herausforderungen in der Agrosystem-Community adressiert als auch weitere Partner aus den Agrarwissenschaften integriert.

Geplante Lösungen und Services von FAIRagro

FAIRagro wird technische Infrastrukturdienste, aber auch Informations- und Beratungsangebote für ein verbessertes FDM in der Agrosystemforschung bereitstellen. Die FAIRagro-Dienste werden in ein NFDI-weites Serviceportfolio eingebettet, um Forschern aller Fachdisziplinen den Zugriff auf Forschungsdaten der Agrosystemforschung zu ermöglichen.

Vernetzung von Repositorien und Suchportal

Ein zentraler Service von FAIRagro wird die Vernetzung der existierenden, verteilten Repositorien umfassen, um die Auffindbarkeit und Wiederverwendung der Daten über verteilte Infrastrukturen hinweg zu ermöglichen. Dafür wird eine Middleware-Infrastruktur entwickelt, die harmonisierte, maschinenlesbare Schnittstellen für den (Meta-)Datenaustausch und einen persistenten Datenzugriff ermöglicht. FAIRagro wird sich bei der Entwicklung der Middleware-Implementierung an etablierten Ansätzen orientieren, z. B. an dem dreistufigen Architekturmuster [42], das eine Struktur mit einer Präsentations-, einer Logik- und einer Datenbankebene vorsieht. Dieses allgemeine Konzept wird mit den NFDI-Research Data Commons (NFDI-RDC) [43], nationalen (Generic Research Data Infrastructure – GeRDI [44]), internationalen Entwicklungen (EOSC [45], ELIXIR [13], GAIA-X [46]) und neuen, in der RDA entwickelten Empfehlungen abgestimmt und unter Beachtung größtmög-

licher Kompatibilität zu etablierten Schnittstellen der jeweiligen Infrastrukturen implementiert. Hierbei wird ein weiterer Schwerpunkt die Integration einer infrastrukturübergreifenden Authentifizierungs- und Autorisierungsinfrastruktur (AAI) sein, um den geschützten Zugang zu Daten und Diensten zu gewährleisten.

Darauf aufbauend werden Inventar- und Suchdienste entwickelt, um einen nutzerfreundlichen Zugang zu den verfügbaren Daten in den verteilten Infrastrukturen zu bieten. Der Inventarisierungsdienst wird einen durchsuchbaren Überblick zu technischen und qualitativen Leistungsindikatoren der verfügbaren Infrastrukturen und Repositories umfassen. Die Suchfunktion wird Metadaten zur Beschreibung dieser Dienste ausnutzen. Darüber hinaus wird ein Datensuchdienst aufgebaut und im zentralen FAIRagro-Portal integriert, der sich auf umfangreiche Metadatenbeschreibungen von Agrosystemdaten stützt und z. B. georäumliche Selektionen und semantische Suchoptionen für Feldversuchsinformationen, genotypische oder phänotypische Merkmale oder Krankheitsinformationen ermöglicht. Weiterhin ist geplant, eine Relevanzbewertung und entsprechende Sortierung gemäß Nutzungs- und Nutzerprofilen zu ermöglichen sowie qualitative und quantitative Aussagen zu FAIR-Konformität und Gebrauchstauglichkeit für bestimmte Anwendungsfälle über verfügbare Forschungsdaten zu integrieren.

Integration von Metadaten, Vokabularen und Ontologien

Das Bewusstsein, dass die konkreten Handlungsempfehlungen hinter den FAIR-Prinzipien Grundvoraussetzungen für verteilte Datendienste und -Infrastrukturen sind, um das Teilen und die Nutzung von Forschungsdaten auf machbare Art und Weise zu ermöglichen, ist seit der Publikation derselben im Jahr 2016 bei den relevanten Akteuren gestiegen. Nichtsdestotrotz muss in der Agrosystemforschung nach wie vor mit Datenquellen umgegangen werden, die den FAIR-Prinzipien noch nicht in allen Punkten genügen und deren Einbindung bei Nutzung und Verarbeitung aufwendig ist. Im Rahmen des Projekts soll daher auch nochmals zur Bewusstseinsbildung beigetragen werden, indem Fragestellungen, mit denen sich Datenbereitsteller konfrontiert sehen, anhand von typischen Beispiele für die Disziplin der Agrosystemforschung adressiert werden, z. B.: Welche Identifier-Systeme genügen den in den FAIR-Prinzipien formulierten Ansprüchen an Persistenz, Dereferenzierbarkeit usw.? Was ist mit „qualifizierten Referenzen zu weiteren Daten“ gemeint? Was verbirgt sich hinter den in I1 geforderten „Wissensrepräsentationssprachen“? Dabei soll ein Inventar von domänenspezifischen Metadatenschemata, Austauschprotokollen, kontrollierten Vokabularen und Ontologien zusammengestellt werden, einschließlich einer Bewertung hinsichtlich der FAIR-Prinzipien.

Eine wichtige Aufgabe ist die Senkung von Einstiegsbarrieren. Eines der Hauptziele ist daher die Zusammenstellung von Mindestmetadatenelementen und die Entwicklung gemeinsamer Metadatenprofile als konsolidierte und harmonisierte Teilmengen derzeit verwendeter Standards. Diese, sowie begleitende Grundsätze und Handlungshinweise, sollen Datenbereitstellern und Entwicklern Orientierung bieten.

Integrative Forschung zu Agrarsystemen erfordert ein hohes Maß an Datenintegration. Interoperabilität soll dabei auf Basis ontologiebasierter Dienste erreicht werden. Die Semantik der verschiedenen Datentypen, die Dienste bereitstellen, wird dabei beschrieben und neben den eigentlichen Daten und den Endpunkten der Dienste bereitgestellt. Dabei werden bestehende generische Ontologien wie die W3C DCAT-Ontologie oder W3C PROV-O für Provenance, aber auch vorhandene domänenspezifische Ontologien (z. B. AGROVOC [29], Environment Ontology [47, 48], Pflanzenphänotypen für Modellorganismen (Plant Trait Ontology [32]) und Wild Types (Flora Phenotype Ontology [49]) oder Crop Ontology [33]) wiederverwendet und bei Bedarf auf der Grundlage der in der Agrarsystem-Community formulierten Anforderungen erweitert, sodass grundsätzliche Vorgehensweisen und *Best Practices* der Datenbeschreibung sowie Standards zu einem integralen Bestandteil der FDM-Workflows in der Agrosystemforschung werden. Außerdem werden Alignments zwischen Vokabularen und Ontologien bereitgestellt. Auf diesem Weg sollen Daten aus verschiedenen Infrastrukturen gemeinsam und übergreifend nutzbar werden.

Neben Ontologien mit einem hohen Grad an Formalisierung wurden insbesondere im Umfeld von Webcrawlern und Suchmaschinen leichtgewichtige Metadatenschemas und -standards zur Annotation von Ressourcen, wie z. B. schema.org entwickelt, die Indexierung vereinfachen und Auffindbarkeit unterstützen sollen. Durch eine breitere Nutzung solcher Standards in den Repositorien soll die Aufnahme und Verarbeitbarkeit von Daten in Suchmaschinen verbessert werden. Um diese Ansätze zu unterstützen, werden entsprechende Leitlinien zur Nutzung sowie gegebenenfalls Erweiterungen für schema.org sowie die Spezifikationen der fachspezifischeren Initiative bioschemas.org erarbeitet.

Beschreibung von Datenqualität und -plausibilität

Geprüfte Datenqualität ist eine Voraussetzung für die wissenschaftliche Nachnutzbarkeit von Daten. Für die Agrosystemforschung relevante Daten sind durch ihre Heterogenität und unterschiedlichen Qualitäten gekennzeichnet. Informationen über die Datenqualität sind unerlässlich für die Kombination von Daten aus verschiedenen Quellen (z. B. räumlichen, spektralen und zeitlichen Daten), insbesondere

für den erfolgreichen Einsatz multimodaler Datenanalysen und maschineller Lernverfahren. Unsicherheiten in Bezug auf die vorliegende Aggregationsebene und deren skalenspezifische Genauigkeit, die Vollständigkeit und Attributierung von Phänotypisierungs- und Altdaten ebenso wie die Plausibilität von Daten aus landwirtschaftlichen Langzeitfeldexperimenten verhindern die Nachnutzung.

Datenqualität muss standardisiert überprüft, kuratiert und derart dokumentiert werden, dass diese Informationen für Nachnutzer sinnvoll bereitstehen. Es fehlen jedoch Kriterien, die helfen, Datenqualität umfassend zu bewerten und Daten entsprechend den Anforderungen der jeweiligen Fragestellung zu selektieren. Aufbauend auf bestehenden typischen Datensätzen aus der Agrosystemforschung werden Datenqualitätsmetriken unter besonderer Berücksichtigung des „Data-Fitness-for-Use“-Ansatzes identifiziert, strukturiert und zugänglich gemacht. Da nicht alle Qualitätsfragen automatisiert gelöst werden können, werden die Algorithmen durch Schnittstellen ergänzt, die es Menschen ermöglichen, Datensätze anhand geeigneter Visualisierungen zu überprüfen. Am Ende wird es eine nutzerfreundliche Anwendung geben, mit der Autoren vor der eigentlichen Nachnutzung von Daten deren Eignung bewerten können. Data Stewards und Datenkuratoren werden mit einer Kurationscheckliste bei der Datenveröffentlichung unterstützt. Datensätze, die die festgelegten Datenkuratierungsverfahren (einschließlich automatisierter Schritte und Überprüfung durch Experten) durchlaufen haben, werden mit entsprechenden Qualitätsnachweisen versehen. Jegliche ergänzenden Informationen zur Datenqualität und „Data-Fitness-for-Use“ werden als „FAIRagro-Qualitäts-Metadaten-Deskriptoren“ in den Metadaten ergänzt und damit suchbar.

FAIR Digital Objects

Die Umsetzung der FAIR-Prinzipien, insbesondere die eigenständige Exploration und Prozessierung von Daten durch Software-Agenten, erfordert über den Aufbereitungsprozess der Datensätze (entsprechend der GO-FAIR-Initiative: FAIRification [50]) hinaus deren Integration in ein unterstützendes Infrastrukturnetzwerk, das unter anderem Services für Identifikatoren zur dauerhaften und persistenten Referenzierung der Daten (Persistent Identifiers/PIDs), Registrierungsdatenbanken für verwendete Datentypen und Terminologiedienste für eingebundene semantische Artefakte wie kontrollierte Vokabulare, Thesauri, Ontologien sowie entsprechende Terminologie Mappings bereitstellt [51].

Aufbauend auf Ergebnissen der Research Data Alliance [52] wird FAIRagro Dateninteroperabilität und Nachnutzbarkeit auf der Basis sogenannter FAIR Digital Objects (FDOs) realisieren [53, 54]. FAIR Digital Objects stellen digitale Container dar, die basierend auf einem grundlegenden Modellkonzept (Abb. 4) verteilte Ressourcen (Primärdaten, Lizenzen, Workflows, Medien wie Bilder) in einen neuen komplexen Datentyp integrieren. Ermöglicht werden hierdurch die Kapselung von Datenressourcen, Abbildung von Relationen zwischen diesen Daten sowie Operationen auf das digitale Objekt. So können beispielsweise im Rahmen von Hochdurchsatz-Phänotypisierung in der Pflanzenzüchtung [55] unterschiedliche materielle und digitale Samples wie erfasste Bilddaten, Genomdaten und Spektralmessungen der biochemischen Eigenschaften der Pflanzen und Messungen von Umweltparametern (Boden, Feuchtigkeit, Stickstoffgehalt) zu einem neuen digitalen Objekt gebunden werden, um die quantitative Ausprägung von Merkmalen (Quantitative Trait) basierend auf Genom-Umwelt-Interaktion zu repräsentieren.

Im Rahmen konkreter Implementationen entwickelt eine Fach-Community spezifische Versionen des konzeptuellen FDO-Modells, sogenannte FDO-Typen, wobei die Ressourcen-Schicht jeweils entsprechend den fachspezifischen Anforderungen modelliert wird. Die Schichten für Resolving, Bindung und Kapselung sowie die Semantik für automatisierte Operationen folgen einem von der FDO-Community [56] standardisierten Aufbau, der die Interoperabilität von Instanzen unterschiedlicher FDO-Typen in einem geteilten Datenraum ermöglicht. Die mit dem Ansatz verbundene Vision ist die Transformation der gegenwärtig durch Datensilos geprägten Forschungsdatenlandschaft hin zu einer domänenübergreifenden, durch weitgehende Einbindung von Software-Agenten für Discovery und Analyse charakterisierten „Integrated Virtual Data Collection“ [57].

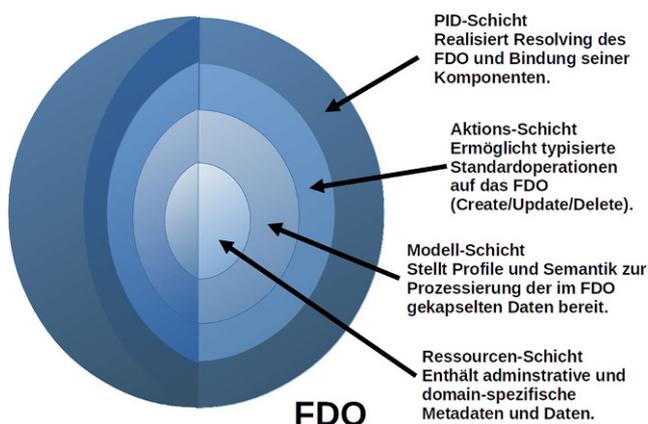


Abb. 4 Fundamentales Modellkonzept bestehend aus PID-Schicht, Aktions-Schicht, Modell-Schicht und Ressourcen-Schicht. (FDO FAIR Digital Objects)

Community, Data Stewards, Training und Lehre

FAIRagro hat bereits im Vorfeld der Konzeption und Entwicklung der beschriebenen Infrastruktur die Bedarfe der Community eingeholt, um zielgerichtet planen zu können. Dazu wurde eine großangelegte Umfrage im Jahr 2020 durchgeführt [5, 58]. Ergänzt wird die Bedarfserhebung laufend durch weitere Umfragen [59] und Workshops, auf denen Vorschläge direkt mit der Community diskutiert werden [60].

Unterstützung bei der Verwendung der FAIRagro-Infrastruktur (z. B. bei der Veröffentlichung von Daten) erhalten Wissenschaftler von FAIRagro Data Stewards. Diese bringen profundes Wissen über die entwickelten Services mit und stehen bereit, auf Anfragen zu reagieren. Neben grundlegendem technischem Support sind die Data Stewards Ansprechpartner in allen Fragen zum Management agrarwissenschaftlicher Daten. Sie sind auf relevante disziplinspezifische Themen spezialisiert (Datenveröffentlichung, Rechtsfragen, Dauerfeldversuche, Farming 4.0, Pflanzenzüchtung, Geodaten etc.) und an den entsprechenden Einrichtungen lokalisiert. Organisiert sind die FAIRagro Data Stewards im *Data Steward Service Center*, welches unter anderem Aufgaben des Netzwerkes innerhalb der Community und zur NFDI übernimmt.

Erster Anlaufpunkt für Suchende wird das FAIRagro-Portal sein, in welchem nicht nur Handouts und Dokumentationen bereitgestellt werden (First-Level-Support), sondern auch der Kontakt zum Data Steward Service Center (Second-Level-Support) hergestellt wird. Bei tiefergehenden Fragen zur Nutzung der FAIRagro-Infrastruktur und Services wird der Kontakt zu den jeweiligen Entwicklern hergestellt (Third-Level-Support).

In FAIRagro wird ein umfassendes Trainings- und Ausbildungsprogramm etabliert, um Wissenschaftler aller Karrierestufen mit Wissen zum FDM zu versorgen und einen kulturellen Wandel hin zu mehr Datenveröffentlichung und Nutzung einzuleiten [61]. Alle entstehenden Trainingsmaterialien und Schulungskonzepte werden als Open Educational Resources (OER) zur Verfügung gestellt. Trainings aber auch Erfahrungen im agrarwissenschaftlichen FDM werden im Bereich der NFDI abgestimmt und geteilt. Bereits jetzt ist ein entsprechender Austausch und ein Netzwerk mit NFDI4Earth [62], NFDI4Biodiversity [63] und NFDI4Health [64] formalisiert und wird durch Veranstaltungen wie Workshops gelebt.

Ausblick

Im November 2022 wurde auf einer Pressekonferenz der GWK bestätigt [65], dass FAIRagro als eines von acht weiteren Konsortien in der dritten und letzten Förderrun-

de in die NFDI aufgenommen wird. Damit werden insgesamt 27 Konsortien aus unterschiedlichen wissenschaftlichen Disziplinen im Rahmen der NFDI gefördert. FAIRagro wird am 1. März 2023 mit einer vorgesehenen Förderdauer von fünf Jahren starten.

Aktuell starten bereits erste Aktivitäten zur Abstimmung mit anderen NFDI-Konsortien mit dem Ziel, vorhandene Entwicklungen nachzunutzen und gemeinsame Arbeiten zu initiieren. Dies betrifft zum aktuellen Zeitpunkt insbesondere Konsortien, die in thematisch eng verwandten Bereichen angesiedelt sind wie DataPLANT [66], NFDI4Earth [62] und NFDI4BioDiversity [63]. Gemeinsam mit NFDI4Earth, der NFDI für Erdsystemwissenschaften, wird eine konzeptionelle Implementierung für die Interoperabilität von Geodaten auf Basis von FAIR Digital Objects erfolgen. Gemeinsam mit NFDI4Earth, DataPLANT und NFDI4BioDiversity werden Möglichkeiten zum Aufbau eines disziplinübergreifenden User-Support-Netzwerks und der Vernetzung von fachspezifischen Data Stewards diskutiert. Außerdem wird sich FAIRagro an einigen der bereits etablierten NFDI-Sektionen [67] zur Bearbeitung übergreifender Fragestellung beteiligen und zu den Arbeitsgruppen beitragen.

Mit der Aufnahme von FAIRagro in die NFDI wird die zentrale Bedeutung der Digitalisierung für die Agrosystemforschung und daraus folgend die grundlegende Bedeutung des Forschungsdatenmanagements als Voraussetzung für hochwertige Forschung in Bereichen wie Pflanzenzüchtung, integriertem Pflanzenschutz, Roboterisierung, Precision Farming, Nachhaltigkeit, Klimaeinfluss und -anpassung, Ernährungssicherheit und Pflanzenwachstumsmodellierung deutlich. Durch die Beteiligung am Aufbau eines interoperablen Datenraums mit der Perspektive der Integration in den größeren Kontext der EOSC werden Agrosystemdaten darüber hinaus effektiver zu interdisziplinären Lösungen zu gesellschaftlichen Herausforderungen wie Klimawandel, Nachhaltigkeit, Biodiversität, Agrar-Ernährung-Gesundheit und Welternährung für die Erreichung der UN-Nachhaltigkeitsziele beitragen können.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- UN (2022) 17 Nachhaltigkeitsziele. <https://sdgs.un.org/goals>. Zugegriffen: 17. Jan. 2023
- European Commission (2019) The European green deal
- Zukunftskommission Landwirtschaft Geschäftsstelle (2021) Zukunft Landwirtschaft. Eine gesamtgesellschaftliche Aufgabe. Empfehlungen der Zukunftskommission Landwirtschaft. <https://www.bundesregierung.de/resource/blob/997532/1939908/39f791fc6a9ff293ae520e2fea41baa7/2021-07-06-zukunftskommission-landwirtschaft-data.pdf>. Zugegriffen: 17. Jan. 2023
- Minasny B, Malone BP, McBratney AB, Angers DA, Arrouays D, Chambers A, Chaplot V, Chen Z-S, Cheng K, Das BS, Field DJ, Gimona A, Hedley CB, Hong SY, Mandal B, Marchant BP, Martin M, McConkey BG, Mulder VL, O'Rourke S, Richer-de-Forges AC, 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. <https://doi.org/10.1016/j.geoderma.2017.01.002>
- Senft M, Stahl U, Svoboda N (2022) Research data management in agricultural sciences in Germany: We are not yet where we want to be. *PLoS One* 17(9):e274677. <https://doi.org/10.1371/journal.pone.0274677>
- (2022) BonaRes Repository. <https://maps.bonares.de/mapapps/resources/apps/bonares>. Zugegriffen: 17. Jan. 2023
- Grosse M, Hoffmann C, Specka X, Svoboda N (2020) Managing long-term experiment data: a repository for soil and agricultural research. In: Long-term farming systems research, S 167–182 <https://doi.org/10.1016/b978-0-12-818186-7.00010-2>
- Oppermann M, Weise S, Dittmann C, Knupffer H (2015) GBIS: the information system of the German Genebank. Database (Oxford). <https://doi.org/10.1093/database/bav021>
- 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). <https://doi.org/10.1093/database/baw033>
- (2022) OpenAgrar. <https://www.openagrar.de/content/index.xml>. Zugegriffen: 17. Jan. 2023
- Oeltjen W, Neumann K, Stahl U, Stephan R (2019) MyCoRe macht Forschungsdaten FAIR. *Bibl Forsch Prax* 43(1):82. <https://doi.org/10.1515/bfp-2019-2013>
- PUBLISSO (2022) <https://www.publisso.de/en/publishing/repositories/repository-for-life-sciences/>. Zugegriffen: 17. Jan. 2023
- ELIXIR (2021) Data platform. <https://elixir-europe.org/platforms/data>. Zugegriffen: 17. Jan. 2023
- (2021) Nature Scientific Data. <https://www.nature.com/sdata/>. Zugegriffen: 17. Jan. 2023
- Rigden DJ, Fernandez XM (2019) The 26th annual nucleic acids research database issue and molecular biology database collection. *Nucleic Acids Res* 47(D1):D1–D7. <https://doi.org/10.1093/nar/gky1267>
- He L, Han ZB (2017) Do usage counts of scientific data make sense? An investigation of the Dryad repository. *Libr Hi Tech* 35(2):332–342. <https://doi.org/10.1108/lht-12-2016-0158>
- Garriga M, Mateos C, Flores A, Cechich A, Zunino A (2016) RESTful service composition at a glance: A survey. *J Netw Comput Appl* 60:32–53. <https://doi.org/10.1016/j.jnca.2015.11.020>
- FAIRagro (2022) Forschungsdateninfrastrukturen von FAIRagro. <https://www.fairagro.net/index.php/de/ueber-uns/unsere-forschungsdateninfrastrukturen>. Zugegriffen: 17. Jan. 2023
- DataCite Metadata Working Group (2017) DataCite metadata schema documentation for the publication and citation of research data. Version 4.1. DataCite e. V. <https://doi.org/10.5438/0014>
- Dublin Core (2020) Specifications. <https://www.dublincore.org/specifications/>. Zugegriffen: 17. Jan. 2023
- DCAT (2020) <https://www.w3.org/TR/vocab-dcat-2/>. Zugegriffen: 17. Jan. 2023
- W3C Prov (2021) <https://www.w3.org/TR/prov-overview/>. Zugegriffen: 17. Jan. 2023
- W3C VoID (2021) <https://www.w3.org/TR/void/>. Zugegriffen: 17. Jan. 2023
- Alercia A, Diulgheroff S, Mackay M (2015) FAO/Biodiversity multi-crop passport descriptors V.2.1. https://www.biodiversityinternational.org/index.php?id=244&tx_news_pi1%5Bnews%5D=7639&cHash=6e87580466bd8236a004bb0a458177b3. Zugegriffen: 17. Jan. 2023
- Bostick WM, Koo J, Walen VK, Jones JW, Hoogenboom G (2004) A web-based data exchange system for crop model applications. *Agron J* 96(3):853. <https://doi.org/10.2134/agronj2004.0853>
- White JW, Hunt LA, Boote KJ, Jones JW, Koo J, Kim S, Porter CH, Wilkens PW, Hoogenboom G (2013) Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Comput Electron Agric* 96:1–12. <https://doi.org/10.1016/j.compag.2013.04.003>
- Cwiek-Kupczynska H, Altmann T, Arend D, Arnaud E, Chen D, Cornut G, Fiorani F, Frohberg W, Junker A, Klukas C, Lange M, Mazurek C, Nafissi A, Neveu P, van Oeveren J, Pommier C, Poorter H, Rocca-Serra P, Sansone SA, Scholz U, van Schriek M, Sereen 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. <https://doi.org/10.1186/s13007-016-0144-4>
- Specka X, Gärtner P, Hoffmann C, Svoboda N, Stecker M, Einspanier U, Senkler K, Muqit Zoarder MA, Heinrich U (2019) The BonaRes metadata schema for geospatial soil-agricultural research data—Merging INSPIRE and DataCite metadata schemes. *Comput Geosci*. <https://doi.org/10.1016/j.cageo.2019.07.005>
- FAO (2021) AGROVOC. <https://www.fao.org/agrovoc>. Zugegriffen: 17. Jan. 2023
- FAO (2019) Agrontology. <https://www.fao.org/agrovoc/agrontology>. Zugegriffen: 17. Jan. 2023
- (2019) Plant ontology. <https://www.ebi.ac.uk/ols/ontologies/po>. Zugegriffen: 17. Jan. 2023
- 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. In: International Conference on Knowledge Engineering and Ontology Development
- 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. *Front Physiol* 3:10. <https://doi.org/10.3389/fphys.2012.00326>
- Jaiswal P, Avraham S, Ilic K, Kellogg EA, McCouch S, Pujar A, Reiser L, Rhee SY, Sachs MM, 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. *Comp Funct Genomics* 6(7–8):388–397. <https://doi.org/10.1002/cfg.496>
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kib

- be W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R (2004) The Gene Ontology (GO) database and informatics resource. *Nucl Acids Res* 32:D258–D261. <https://doi.org/10.1093/nar/gkh036>
36. Di L, Yue P, Ramapriyan HK, King RL (2013) Geoscience data provenance: an overview. *IEEE Trans Geosci Remote Sens* 51(11):5065–5072. <https://doi.org/10.1109/tgrs.2013.2242478>
37. German Council for Scientific Information Infrastructures (2020) The data quality challenge: Recommendations for sustainable research in the digital turn. <https://rfii.de/?p=4203>. Zugegriffen: 17. Jan. 2023
38. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, 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. *Sci Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>
39. Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, Courtot M, Crosas M, Dumontier M, Evelo CT, Goble C, Guizzardi G, Hansen KK, Hasnain A, Hettne K, Heringa J, Hooft RWW, Imming M, Jeffery KG, Kaliyaperumal R, Kersloot MG, Kirkpatrick CR, Kuhn T, Labastida I, Magagna B, McQuilton P, Meyers N, Montesanti A, van Reizen M, Rocca-Serra P, Pergl R, Sansone S-A, da Silva SLOB, Schneider J, Strawn G, Thompson M, Waagmeester A, Weigel T, Wilkinson MD, Willighagen EL, Wittenburg P, Roos M, Mons B, Schultes E (2020) FAIR principles: Interpretations and implementation considerations. *Data Intell* 2(1–2):10–29. https://doi.org/10.1162/dint_r_00024
40. European Open Science Cloud (2017) EOSC declaration. https://eosc-portal.eu/sites/default/files/eosc_declaration.pdf. Zugegriffen: 17. Jan. 2023
41. Directorate-General for Research and Innovation (European Commission), EOSC Executive Board, Corcho O, Eriksson M, Kurowski K, Ojstersek M, Choirat C, Sanden M, Coppens F (2021) EOSC interoperability framework : report from the EOSC executive board working groups FAIR and architecture <https://doi.org/10.2777/620649>
42. Eckerson WW (1995) Three tier client/server architectures: achieving scalability, performance, and efficiency in client server applications. *Open Inf Syst* 3:46–50
43. Glöckner FO, 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) <https://doi.org/10.5281/zenodo.3943645>
44. Latif A, Limani F, Tochtermann K (2019) A generic research data infrastructure for long tail research data management. *Data Sci J*. <https://doi.org/10.5334/dsj-2019-017>
45. EOSC (2020) <https://www.eosc-portal.eu/>. Zugegriffen: 17. Jan. 2023
46. (2021) GAIA-X. <https://www.data-infrastructure.eu>. Zugegriffen: 17. Jan. 2023
47. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, Consortium E (2013) The environment ontology: contextualising biological and biomedical entities. *J Biomed Semant* 4(1):43. <https://doi.org/10.1186/2041-1480-4-43>
48. Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, Mungall CJ (2016) The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperability. *J Biomed Semant* 7(1):57. <https://doi.org/10.1186/s13326-016-0097-6>
49. Hoehndorf R, Alshahrani M, Gkoutos GV, Gosline G, Groom Q, Hamann T, Kattge J, de Oliveira SM, Schmidt M, Sierra S, Smets E, Vos RA, Weiland C (2016) The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants. *J Biomed Semant* 7(1):65. <https://doi.org/10.1186/s13326-016-0107-8>
50. GO FAIR (2021) FAIRification process. <https://www.go-fair.org/fair-principles/fairification-process/>. Zugegriffen: 17. Jan. 2023
51. Collins S, Genova F, Harrower N, Hodson S, Jones S, Laaksonen L, Mietchen D, Petruskaitė R, Wittenburg P (2018) Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. Publications Office of the European Union, Luxembourg <https://doi.org/10.2777/54599>
52. RDA (2022) Research data alliance. <https://www.rd-alliance.org/>. Zugegriffen 17. Jan. 2023
53. 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 Sci J*. <https://doi.org/10.5334/dsj-2020-050>
54. De Smedt K, Koureas D, Wittenburg P (2020) FAIR digital objects for science: from data pieces to actionable knowledge units. *Publications* 8(2):21
55. Danilevicz MF, Tay Fernandez CG, Marsh JI, Bayer PE, Edwards D (2021) High-throughput genotyping technologies in plant taxonomy. *Methods Mol Biol* 2222:149–166. https://doi.org/10.1007/978-1-0716-0997-2_9
56. FAIRDO (2022) FAIR DO community. <https://fairdo.org/>. Zugegriffen: 17. Jan. 2023
57. Wittenburg P, Strawn G (2021) Revolutions take time. *Information* 12(11):472. <https://doi.org/10.3390/info12110472>
58. Senft M, Stahl U, Svoboda N (2021) Dataset: survey about research data management in agricultural sciences in Germany. *OpenAgrar repository*. <https://doi.org/10.5073/20211013-105447>
59. FAIRagro (2021) 2. FAIRagro Umfrage. <https://fairagro.net/index.php/de/informieren/veranstaltungen-2/110-2-umfrage-2021>. Zugegriffen: 17. Jan. 2023
60. FAIRagro (2021) Workshops <https://fairagro.net/index.php/en/component/tags/tag/workshop>. Zugegriffen: 17. Jan. 2023
61. Robson SG, Baum MA, Beaudry JL, Beitner J, Brohmer H, Chin JM, Jasko K, Kouros CD, Laukkonen RE, Moreau D, Searston RA, Slagter HA, Steffens NK, Tangen JM, Thomas A (2021) Promoting open science: a holistic approach to changing behaviour. *Collabra: Psychol*. <https://doi.org/10.1525/collabra.30137>
62. NFDI4Earth (2022) Website NFDI4Earth. <https://www.nfdi4earth.de>. Zugegriffen: 17. Jan. 2023
63. NFDI4BioDiversity (2022) Website NFDI4BioDiversity. <http://www.nfdi4biodiversity.de>. Zugegriffen: 17. Jan. 2023
64. NFDI4Health (2022) Website NFDI4Health. <https://www.nfdi4health.de/en/>. Zugegriffen: 17. Jan. 2023
65. GWK (2022) Nationale Forschungsdateninfrastruktur (NFDI) mit der Aufnahme von acht Konsortien in die Förderung komplett
66. DataPLANT (2022) DataPLANT. <https://nfdi4plants.de>. Zugegriffen: 17. Jan. 2023
67. NFDI Sektionen (2022) NFDI Website, Sektionen. <https://www.nfdi.de/sektionen/>. Zugegriffen: 17. Jan. 2023

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.

Affiliations

Xenia Specka¹ · Daniel Martini² · Claus Weiland³ · Daniel Arend⁴ · Senthold Asseng⁵ · Franziska Boehm⁶ · Til Feike⁷ · Juliane Fluck⁸ · David Gackstetter⁵ · Aida Gonzales-Mellado⁹ · Thomas Hartmann⁶ · Jan-Henrik Haurert¹⁰ · Florian Hoedt⁹ · Carsten Hoffmann¹ · Patrick König⁴ · Matthias Lange⁴ · Stephan Lesch³ · Birte Lindstädt⁸ · Gunnar Lischeid¹ · Markus Möller⁷ · Uwe Rascher¹¹ · Jochen Christoph Reif⁴ · Markus Schmalz¹² · Matthias Senft⁷ · Ulrike Stahl⁷ · Nikolai Svoboda¹ · Björn Usadel¹¹ · Heidi Webber¹ · Frank Ewert¹

¹ Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Deutschland

² Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL), Bartningstr. 49, 64289 Darmstadt, Deutschland

³ Senckenberg – Leibniz Institution for Biodiversity and Earth System Research (SGN), Senckenberganlage 25, 60325 Frankfurt, Deutschland

⁴ Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstr. 3, 06466 Seeland OT Gatersleben, Deutschland

⁵ Hans-Eisenmann-Forum für Agrarwissenschaften der TU München, Liesel-Beckmann-Str. 2, 85354 Freising, Deutschland

⁶ Leibniz-Institut für Informationsinfrastruktur (FIZ Karlsruhe), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Deutschland

⁷ Julius Kühn-Institut – Bundesforschungsinstitut für Kulturpflanzen, Erwin-Baur-Str. 27, 06484 Quedlinburg, Deutschland

⁸ Informationszentrum Lebenswissenschaften (ZB MED), Gleueler Str. 60, 50931 Köln, Deutschland

⁹ Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei, Johann Heinrich von Thünen-Institut, Bundesallee 50, 38116 Braunschweig, Deutschland

¹⁰ Rheinische Friedrich-Wilhelms-Universität Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Deutschland

¹¹ Forschungszentrum Jülich, Wilhelm-Johnen-Str., 52428 Jülich, Deutschland

¹² Generaldirektion der Staatlichen Archive Bayerns, Schönfeldstr. 5, 80539 München, Deutschland