

FAIR Data Infrastructure for Agrosystems (FAIRagro)

*Sprecher: Prof. Dr. Frank Ewert, Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF)
fairagro@zalf.de*

Forschungsfeld und Forschungsdaten: Eine nachhaltig funktionierende Landwirtschaft ist gesellschaftlich unverzichtbar. Sie ist wesentlich für die Gewährleistung der Ernährungssicherheit und beeinflusst maßgeblich die Ökosystemleistungen und Agro-Biodiversität komplexer Landschaftssysteme. Besondere Anforderungen ergeben sich daher an eine nachhaltige Gestaltung der Primärproduktion für Nahrungs- und Futtermittel, Bioenergie- und Rohstoffgewinnung. Die Forschung zu Agrosystemen entwickelt Strategien für eine nachhaltige Primärproduktion unter Erhaltung und Verbesserung von Ökosystemleistungen und Agro-Biodiversität. Relevante Systemebenen erstrecken sich vom Bodenmikrobiom bis hin zu Landschaftsräumen. Die Agrosystemforschung umfasst dabei ein breites Feld verschiedener Fachdisziplinen wie dem Acker- und Pflanzenbau, der Bodenkunde und Pflanzenernährung, Phytopathologie und Pflanzenschutz, der angewandten Genetik und Pflanzenzüchtung, aber auch der Agroökosystemforschung, sowie der Agrar-, Landschafts- und Geoökologie. Auch neue technologisch orientierte Disziplinen im Bereich der Digitalisierung (Robotik, Sensorik, Künstlichen Intelligenz und Modellierung) erfahren zunehmende Aufmerksamkeit. In den verschiedenen Fachdisziplinen werden unterschiedliche Arten von Daten aus diversen Quellen erzeugt, gesammelt, strukturiert und analysiert. Diese umfassen Sequenzierungs- und Phänotypisierungsdaten von Nutzpflanzen, Daten aus Landschaftsmonitoring und Feldversuchen, Fernerkundungsdaten, oder auch Modellierungsdaten mit unterschiedlicher räumlicher und zeitlicher Auflösung. Durch den zunehmenden Einsatz von Sensortechnologien auf dem Feld und in kontrollierten Umgebungen sowie durch die zunehmende Digitalisierung und Automatisierung in der Landwirtschaft (Landwirtschaft 4.0) werden verstärkt hochaufgelöste Daten erzeugt. Bis heute fehlt eine Vernetzung dieser hochgradig diversen Daten und deren bedarfs- und nutzergerechte Verfügbarmachung über das eigene Fachgebiet hinaus.

Anforderungen und Ziele an das Forschungsdatenmanagement (FDM): Aktuell befinden sich die Forschungsdaten der Agrosystemforschung überwiegend in wenig vernetzten, disziplinären und institutionellen Datenbanken oder liegen teilweise unstrukturiert und ohne Metadaten in heterogenen Speichersystemen. Wertvolle Forschungsdaten, z.B. aus Langzeitstudien, sind oftmals schwer zugänglich und unzureichend beschrieben. Der Umgang mit sensiblen Daten spielt in der Agrosystemforschung eine besondere Rolle. Rechtliche Regeln zur Nachnutzung dieser Daten sind bis heute uneinheitlich und erschweren deren Zugang. Diese aktuellen Hindernisse sowie zukünftige konkrete Bedarfe an das FDM von Seiten der Community müssen quantifiziert, regelmäßig identifiziert und fortlaufend angepasst werden, um gezielte Maßnahmen bedarfsgerecht abzuleiten und fortzuschreiben. Eine Forschungsdateninfrastruktur (FDI) mit standardisierten Regeln zur Daten-Interoperabilität und Nutzungsrechten ist daher eine Voraussetzung für interdisziplinäre Agrosystemforschung.

FAIRagro hat das Ziel eine flexible, interoperable und skalierbare FDI für die Agrosystemforschung zu entwickeln, existierende disziplinäre Repositorien miteinander zu vernetzen und fehlende fachspezifische Repositorien unter Berücksichtigung eines gemeinsamen Infrastrukturkonzepts "Research Data Commons" aufzubauen. Weiterhin wird FAIRagro ein webbasiertes zentrales Zugangportal für eine interaktive, Repositorien-übergreifende Suche und Schnittstellen für einen maschinenlesbaren Zugriff auf Forschungsdaten für diverse Nutzergruppen bereitstellen. Damit werden Forschungsdaten der Agrosystemforschung zukünftig leicht auffindbar, zugänglich, interoperabel und langfristig nachnutzbar und die FAIR-Prinzipien in enger Zusammenarbeit mit weiteren NFDI-Konsortien konsequent umgesetzt. Forschungsdaten, die aus Mitteln der öffentlichen Hand finanziert wurden, sollen unter Wahrung der Erstverwertungsrechte und unter Berücksichtigung datenschutzrechtlicher Aspekte (DSGVO), zur freien, ggf. auch wirtschaftlichen, Nachnutzung (DNG) bereitgestellt werden. Darüber hinaus stellt FAIRagro die Qualität von Forschungsdaten durch Entwicklung und Etablierung von domänenspezifischen Maßnahmen der Qualitätskontrolle und Einrichtung eines Qualitäts-Feedback- & Kurationsystems sicher. Es werden Datenschutz- und ethische Standards, Richtlinien für den Umgang mit sensiblen Daten und technische Lösungen zur Umsetzung dieser Richtlinien entwickelt und verbreitet. Um die Wissensvermittlung an Nachwuchswissenschaftlern zu einem effektiven Forschungsdatenmanagement vorzubereiten und damit einen kulturellen Wandel herbeizuführen, werden im Rahmen der konsortialen Arbeit verschiedene Trainings- und Schulungsmodule entwickelt und angeboten.

Erfahrungen im Konsortium und Vernetzung: Das fachlich breit aufgestellte Konsortium umfasst derzeit mehr als 30 Partner mit langjährigen Erfahrungen im Forschungsdatenmanagement und mit exzellenter fachlicher Expertise in den relevanten Disziplinen und Verbindung zur landwirtschaftlichen Praxis sowie einer sehr guten Einbindung in nationale und internationale Forschungsnetzwerke und disziplinäre Fachgesellschaften und Forschungsallianzen. Die am Konsortium beteiligten Institutionen koordinieren bzw. repräsentieren zahlreiche wichtige Pilotinitiativen und Infrastrukturen des Managements von Forschungsdaten für die agrarwissenschaftliche Gemeinschaft, z. B. das Exzellenzcluster PhenoRob (Universität Bonn, FZ Jülich), TERENO, EMPHASIS, IPPN (FZ Jülich), BonaRes Repository (ZALF), Edaphobase (SGN), PUBLISSO (ZB MED), TISDAR (Thünen), OpenAgrar (JKI), ELIXIR, de.NBI und GBIS (IPK) u.v.m. Weiterhin sind die beteiligten Institute in relevanten internationalen agrarwissenschaftlichen Gremien aktiv, z. B. MIAPPE, Breeding API, ECPGR, ORCID-DE, DivSeek, IP-BES, GFBio, DiSSCo Prepare, EUdaphobase, FAO-AGROVOC, RDA-IGAD, GAIA-X Domäne Agrar und EJP-SOIL. Im Bereich der Agrosystemmodellierung nehmen Vertreter des Konsortiums international führende Rollen wahr, wie z.B. in AgMIP (HEF, ZALF), MACSUR (ZALF, Thünen) und der Wheat-Initiative (HEF, ZALF, JKI).

Schnittstellen zur NFDI: Das FAIRagro-Konsortium hat das Ziel, die wissenschaftliche Gemeinschaft der Agrosystemforschung als Teil der gesamten NFDI zu integrieren und einen benutzerfreundlichen Datenaustausch zwischen verschiedenen Disziplinen zu fördern. Die Aktivitäten des interdisziplinären Austauschs spiegeln sich in der Beteiligung der Mitglieder von FAIRagro an diesen NFDI-Konsortien wider: NFDI4BioDiversity, NFDI4Earth, DataPlant, NFDI4Health, NFDI4Life Umbrella, und NFDI4Culture. FAIRagro kooperiert mit anderen

Konsortien bzgl. der Vernetzung disziplinärer NFDIs und ist aktiv an der Bearbeitung gemeinsam relevanter Querschnittsthemen beteiligt, z.B. Umgang mit sensiblen Daten und gemeinsame Metadatenstandards (siehe *'Leipzig-Berlin Erklärung zu NFDI Querschnittsthemen und Infrastrukturentwicklung'* <http://doi.org/10.5281/zenodo.3895208>). Zu NFDI4BioDiversity gibt es neben komplementären Bereichen auch Überschneidungen bei den Forschungsbereichen Pflanzen und Pilze. Entsprechende Daten und Repositorien werden sowohl von NFDI4BioDiversity als auch von FAIRagro zugänglich und nutzbar gemacht. Mit dem DataPLANT-Konsortium ist eine enge Zusammenarbeit bei der Schaffung und Harmonisierung von Ontologien und Standards im Pflanzenforschungsbereich vorgesehen. Kooperationen mit NFDI4Earth und NFDI4Health betreffen vor allem den Umgang und Austausch von Geodaten bzw. sensiblen Personendaten. FAIRagro und KonsortSWD werden gemeinsame Ansätze zur Verknüpfung der Anwendungspotenziale von Daten zur Landnutzung und Nahrungsmittelproduktion, soziodemographischen Daten sowie Wirtschafts- und Unternehmensdaten untersuchen.

Querschnittsthemen: Die FAIRagro betrachtet die Bereitstellung und Nutzbarmachung von Forschungsdaten bei den Partnern, die Integration der vorhandenen Infrastrukturen in die NFDI und die Langzeitarchivierung als ein wichtiges Querschnittsthema für die gesamte NFDI. Darüber hinaus werden die folgenden NFDI-übergreifende Themengebiete für das Konsortium bedient: Ausbildung und Lehre zum FDM, Datenqualität, einheitliche Metadatenstandards, Interoperabilität von Services, rechtliche und ethische Aspekte, Datensicherheit und -eigentumsrechte, Umgang mit unvollständigen Daten in der Modellierung sowie Reputation und Zitation. FAIRagro wird einen Beitrag zum Thema "Datenschutz und Datenpolitik" leisten. Landwirtschaftliche Daten sind sowohl ökologischer als auch sozio-ökonomischer Natur und werden meist auf privatem Land bzw. auf landwirtschaftlichen Betrieben erhoben. Diese Daten sind oft räumlich explizit, und die Identität des dazugehörigen Betriebes kann anhand von Hintergrundinformationen, die für die Wiederverwendbarkeit erforderlich sind, identifiziert werden. Die gesetzlichen Anforderungen an den Datenschutz sind bei der Veröffentlichung und dem Zugriff solcher Daten daher von wesentlicher Bedeutung. Es muss eine Balance zwischen der freien Nachnutzung, Autorenrechten und dem berechtigten Interesse von Landwirt*innen als Eigentümer der beforschten Produktionsmittel am Datenschutz geschaffen werden. FAIRagro wird die Rahmenbedingungen für den nötigen Schutz landwirtschaftlicher Daten mit räumlichem Bezug analysieren und Richtlinien sowie technische Lösungen für den offenen Zugang zu Daten, die nicht zuverlässig anonymisiert werden können, entwickeln. Durch die Berücksichtigung dieser rechtlichen und ethischen Aspekte wird die FAIRagro für die gesamte NFDI von hohem Nutzen sein.

Erwartung an die NFDI-Konferenz: Wir wollen mit anderen Konsortien weiter vernetzen, um bestehende und neue Querschnittsthemen zu fokussieren. Neben einem strategischen Austausch erhoffen wir uns eine inhaltliche Weiterentwicklung eigener Ideen und Konzepte, mit dem Ziel, die Interoperabilität und Nutzbarkeit von Daten zu verbessern. Die Quervernetzung mit anderen Forschungsbereichen soll entwickelt werden, um damit nicht nur Daten- sondern auch Denksilos zu vermeiden. Durch frühzeitige Absprachen sollen Parallelentwicklungen vermieden und Synergien genutzt werden.

Vorgesehene Mitglieder des Konsortiums (Co-Sprecherinnen/Co-Sprecher und die weiteren, beteiligten Institutionen):

Co-Sprecher/in	Zugehörige Institution
Prof. Dr. Senthoid Asseng Direktor Hans Eisenmann-Forum senthoid.asseng@tum.de	Technische Universität München, Hans Eisenmann-Forum, Lehrstuhl für „Digital Agriculture“ Liesel-Beckmann-Str. 2 85354 Freising
Prof. Dr. Franziska Boehm Bereichsleiterin Immaterialgüterrechte franziska.boehm@fiz-karlsruhe.de	FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen
Prof. Dr. Juliane Fluck Leitung Bereich Wissensmanagement fluck@zbmed.de	Deutsche Zentralbibliothek für Medizin (ZB MED) Informationszentrum Lebenswissenschaften Gleueler Straße 60 50931 Köln
Prof. Dr. Jan-Henrik Haurert Professor für Geoinformation haurert@igg.uni-bonn.de	Rheinische Friedrich-Wilhelms-Universität Bonn Institut für Geodäsie und Geoinformation Meckenheimer Allee 172 D-53115 Bonn
Florian Hoedt Leiter Geoinformation, Zentrum für Informationsmanagement florian.hoedt@thuenen.de	Johann Heinrich von Thünen-Institut, Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei Bundesallee 50 38116 Braunschweig
Dr. Carsten Hoffmann Wissenschaftlicher Mitarbeiter AG Forschungsdatenmanagement hoffmann@zalf.de	Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) Eberswalder Straße 84 15374 Müncheberg
Dr. Matthias Lange Koordination wissenschaftliche Datenhaltung lange@ipk-gatersleben.de	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Corrensstraße 3 06466 Seeland OT Gatersleben
Birte Lindstädt Leitung Abt. Forschungsdatenmanagement lindstaedt@zbmed.de	Deutsche Zentralbibliothek für Medizin (ZB MED) - Informationszentrum Lebenswissenschaften Gleueler Straße 60 50931 Köln
Daniel Martini Stellvertretender Teamleiter Datenbanken und Wissenstechnologien d.martini@ktbl.de	Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL) Bartningstraße 49 64289 Darmstadt

Prof. Dr. Uwe Rascher Leiter Forschungsbereich 'Shoot Dynamics' (Forschungszentrum Jülich), Professor für 'Quantitative Physiologie der Nutzpflanzen' (Universität Bonn) u.rascher@fz-juelich.de	Forschungszentrum Jülich Wilhelm-Johnen-Straße 52428 Jülich
Dr. David J. Russell Leiter Sektion Mesofauna, wiss. Leiter der Görlitzer Biodiversitätsinformatik david.russell@senckenberg.de	Senckenberg Museum für Naturkunde Görlitz Am Museum 1 02826 Görlitz
Prof. Dr. Ulrich Schurr Institutsdirektor u.schurr@fz-juelich.de	Forschungszentrum Jülich Wilhelm-Johnen-Straße 52428 Jülich
Dr. Matthias Senft Referent für Forschungsdatenmanagement, Qualitätssicherung und wissenschaftliches Berichtswesen msenft@atb-potsdam.de	Leibniz-Institut für Agrartechnik und Bioökonomie e.V. (ATB) Max-Eyth-Allee 100 14469 Potsdam
Dr. Xenia Specka Leiterin AG Dateninfrastrukturen specka@zalf.de	Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) Eberswalder Straße 84 15374 Müncheberg
Dr. Ulrike Stahl Koordination Forschungsdatenmanagement ulrike.stahl@julius-kuehn.de	Julius Kühn-Institut (JKI), Bundesforschungsinstitut für Kulturpflanzen Erwin-Baur-Straße 27 06484 Quedlinburg
Dr. Nikolai Svoboda Leiter AG Forschungsdatenmanagement svoboda@zalf.de	Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) Eberswalder Straße 84 15374 Müncheberg
Prof. Dr. Björn Usadel Direktor Institut für Bioinformatik, IBG-4 und Leiter Institut für Biological Data Science (HHU) b.usadel@fz-juelich.de	Forschungszentrum Jülich Wilhelm-Johnen-Straße 52428 Jülich
Prof. Dr. Jan Vanderborght Direktor Institut Agropäre, IBG-3 j.vanderborght@fz-juelich.de	Forschungszentrum Jülich Wilhelm-Johnen-Straße 52428 Jülich
Dr. Claus Weiland WP Lead Distributed System of Scientific Collections (DiSSCo) claus.weiland@senckenberg.de	Senckenberg Gesellschaft für Naturforschung (SGN) Senckenberganlage 25 60325 Frankfurt