

# Forschungsdatenmanagement in der Ressortforschung

## Ein FAIRagro-Workshop zu Brennpunkten und Lösungsansätzen

in Kooperation mit dem Bayerischen Staatsministerium für Ernährung, Landwirtschaft, Forsten und Tourismus

**21. + 22. Oktober 2024, München**

Lucia Vedder<sup>1</sup>, Florian Beyer<sup>2</sup>, Sophie Boße<sup>3</sup>, Elena Rey Mazón<sup>4</sup>, Wahib Sahwan<sup>5</sup>, Marcus Schmidt<sup>5</sup>,  
Lea Sophie Singson<sup>6</sup>, Ulrike Stahl<sup>2</sup> im Namen des FAIRagro Konsortiums

<sup>1</sup>Rheinische Friedrich-Wilhelms-Universität Bonn, <sup>2</sup>Julius-Kühn-Institut, <sup>3</sup>ZB MED Informationszentrum Lebenswissenschaften, <sup>4</sup>Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), <sup>5</sup>Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) e.V., <sup>6</sup>Leibniz-Institut für Informationsinfrastruktur (FIZ Karlsruhe)

# Vorträge (21.10.2024)

1. FAIRagro – FAIRes Datenmanagement in der Agrosystemforschung  
(Ulrike Stahl)
2. Grundprinzipien des professionellen Forschungsdatenmanagements  
(Sophie Boße)
3. Datenmanagementpläne als Grundlage eines erfolgreichen Forschungsdatenmanagements  
(Wahib Sahwan, Elena Rey Mazón)
4. Umgang mit Projektdaten: Von der Erhebung über die Organisation bis hin zur Veröffentlichung  
(Lucia Vedder, Florian Beyer)
5. Q&A zu rechtlichen und ethischen Stolpersteinen in der Agrosystemforschung  
(Lea Sophie Singson)
6. Wo bekomme ich Unterstützung? Die FDM-Helpdesks und Services von FAIRagro, NFDI4Biodiversity und Co.  
(Marcus Schmidt)

# Hands-on Sessions (22.10.2024)

## Legal Lounge

7. (Don't) take it personally: Datenschutz am Beispiel von „DigiMilch“  
(Lea Sophie Singson, Sophie Boße)
8. To license or not to license: Q&A rund um Rechte an Daten  
(Lea Sophie Singson, Sophie Boße)

## Einblicke ins Datenlabor

9. Big Data-Verarbeitung im Behördenalltag  
(Lucia Vedder, Florian Beyer)

## Von A wie Auffinden bis V wie Veröffentlichen

10. Von A wie Auffinden bis V wie Veröffentlichen  
(Marcus Schmidt, Wahib Sahwan, Elena Rey Mazón)

# FAIRagro

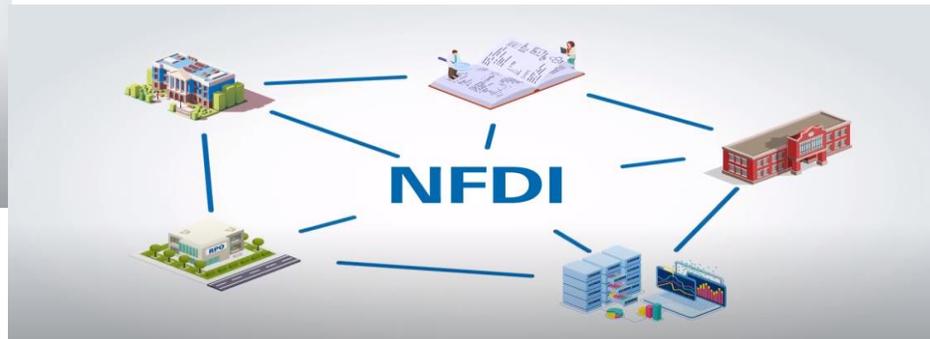
## FAIRes Datenmanagement in der Agrosystemforschung

Ulrike Stahl, Julius Kühn-Institut (JKI) im Namen des FAIRagro Konsortium

[www.fairagro.net](http://www.fairagro.net)



# FAIRagro – ein Konsortium der Nationalen Forschungsdateninfrastruktur (NFDI)

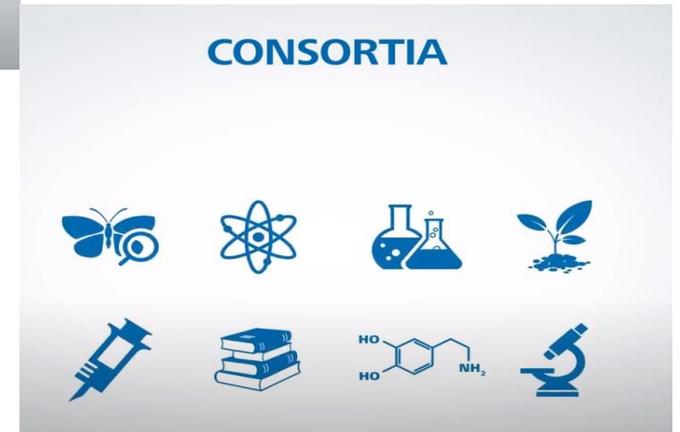


## Problematik:

- dezentrale und temporäre Speicherung von Forschungsdaten
- nicht standardisierte Metadaten
- unterschiedliche Qualität

## Ziele der (NFDI):

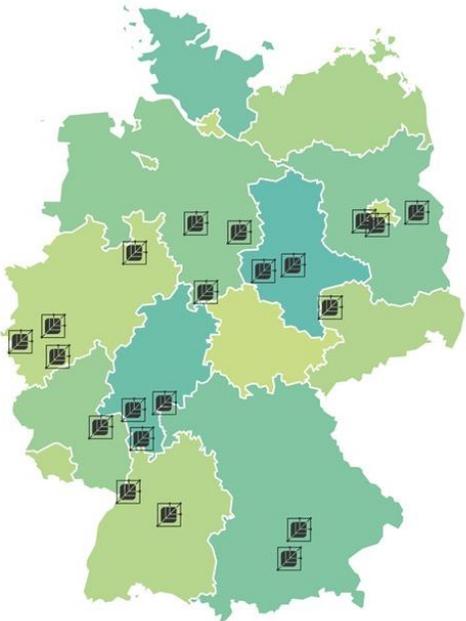
- **Datenbestände** öffentlich finanzierter Forschung systematisch zu **erschließen**
  - nachhaltig zu **sichern** sowie **auffindbar** und **zugänglich** zu machen
  - **Standards** im Datenmanagement entwickeln und etablieren
  - nationale und internationale **Vernetzung** der Daten gewährleisten
- **neue wissenschaftliche Erkenntnisse gewinnen**
- **innovative Forschung ermöglichen**
- **internationale Anschlussfähigkeit garantieren**



# FAIRagro – ein Konsortium von 26



# FAIRagro – Partner, Disziplinen, Struktur



## 30 Partner

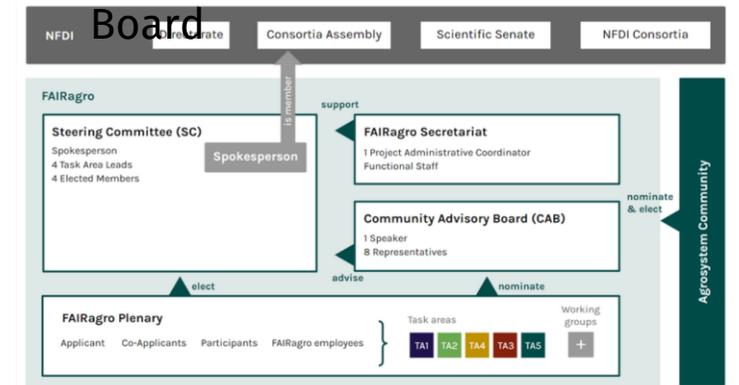
- Universitäten
- Fachhochschulen
- Forschungsinstitute
- Landesforschungseinrichtungen
- Bundesforschungseinrichtungen
- Infrastrukturdienstleister
- Fachgesellschaften

## Disziplinen

- Pflanzen- und Gemüswissenschaften
- Bodenwissenschaften
- Agrar- und Umweltwissenschaften
- Sozioökonomie

## Struktur

- 38 Mitarbeitende
- 9 Mitglieder im Community Advisory Board



# Was sind Herausforderungen der Agrosystemforschung?

## Herausforderungen

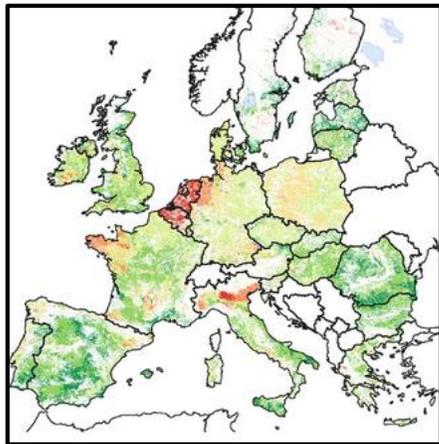
- Ernährung sichern
- Klimawandel
- Biodiversitätsverlust
- Rückgang der natürlichen Ressourcen
- Umweltverschmutzung

## Ziel für 2030 (EU F2F Strategie)

- Verringerung des Einsatzes von chemischen Pestiziden um 50
- Reduzierung der Nährstoffverluste um > 50 %
- Reduzierung des Düngemiteleinsatzes um > 20 %
- Ausweitung des ökologischen Landbaus auf > 25 % der landwirtschaftlichen Nutzfläche der EU...

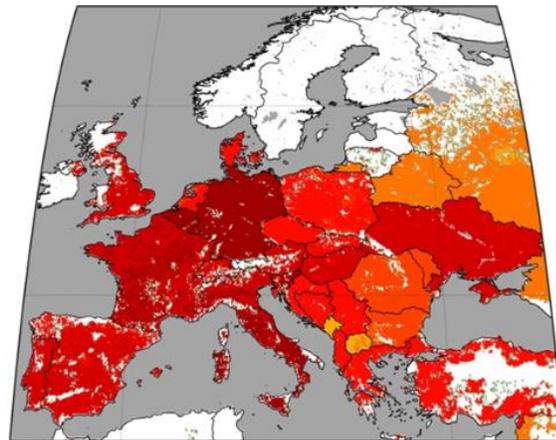


## Stickstoffüberschüsse



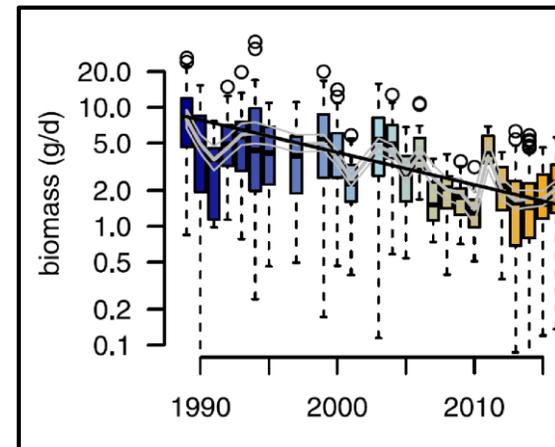
De Vries (2021)

## Ausbringungsmengen von Glyphosat



Maggi et al. (2019)

## Rückgang der Insektenbiomasse



Hellmann et al., (2017)

## Verlust von Ökosystemdienstleistungen

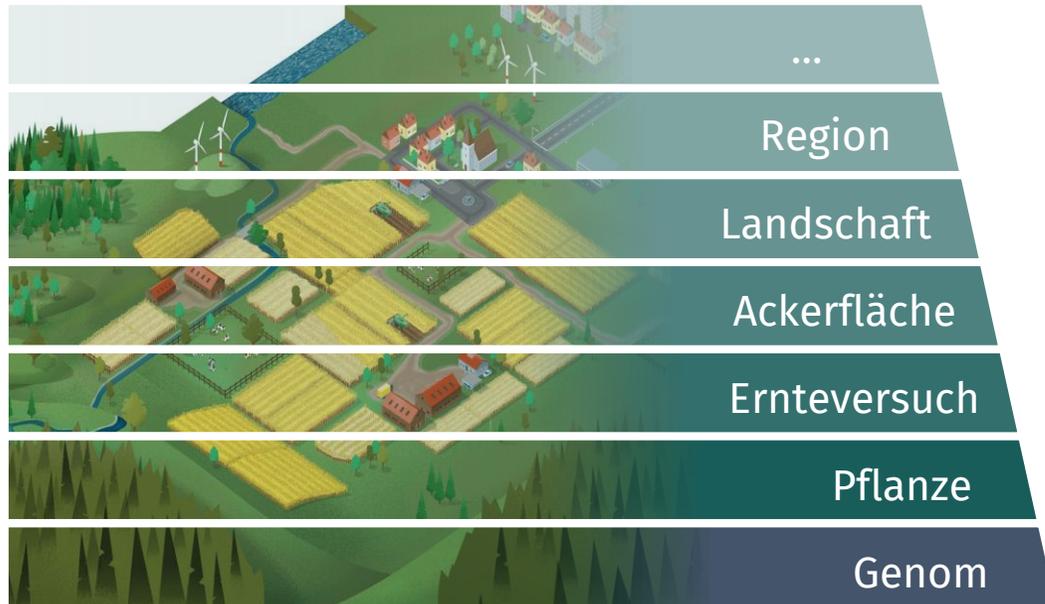


J. Holland (Unsplash)

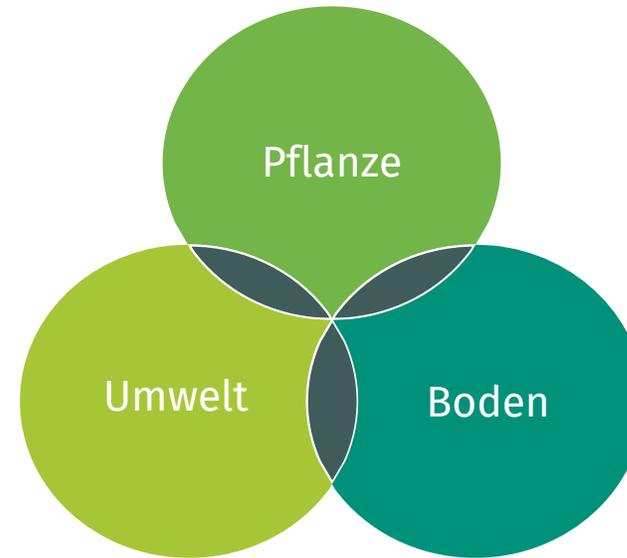


# Was sind die Herausforderung im Datenmanagement ?

## Skalen



## Themen und Disziplinen



## Datenkategorien



- heterogene isolierte Dateninfrastrukturen
- verschiedene Standards
- unterschiedliche Datenqualität
- differenzierte Rechtslage



# Was sind die Fragen und Bedarfe der Forschenden ?

Wir brauchen einen Kulturwandel

Wir brauchen einheitliche Standards

Wir brauchen Rechtssicherheit für sensible Daten

Soll ich meine Daten publizieren?

Wo und Wie kann ich meine Daten publizieren?

Ist der Datensatz vollständig?

Darf ich diese Daten veröffentlichen, z.B. Interview mit Landwirt

Agrosystem Community



Ist die Luftaufnahme eingeschränkt zu nutzen?

Wo und wie finde ich Bodendaten?

Welche Methode wurde benutzt um den Boden PH-Wert zu messen?

Kann ich den Daten vertrauen?

Sind diese Ertragsdaten plausibel?

Wir brauchen eine vernetzte Dateninfrastruktur für übergreifende Forschung

Wir brauchen Qualitätsmetriken

Wir brauchen nachvollziehbare Analyseworkflows

# Was ist die Antwort von FAIRagro ?

Unterstützen, Anleiten, Lehren, Ausbilden, Sensibilisieren

Data stewards  
Helpdesk  
Rechtssupport

Material für Ausbildung,  
Trainings

Roadshows,  
Konferenzen,  
Workshops

Infrastrukturservices anbieten

Suchportal von  
vernetzten  
Dateninfrastrukturen &  
Inventurkataloge

Infrastrukturen um  
Analyseworkflows zu  
dokumentieren

Datenqualitäts- und  
Plausibilitätsmetriken

Use Cases etablieren

6 Flagship Use Cases

1. Use Case Call  
Juli 2024

2. Use Case Call  
Anfang 2025

www.fairagro.net

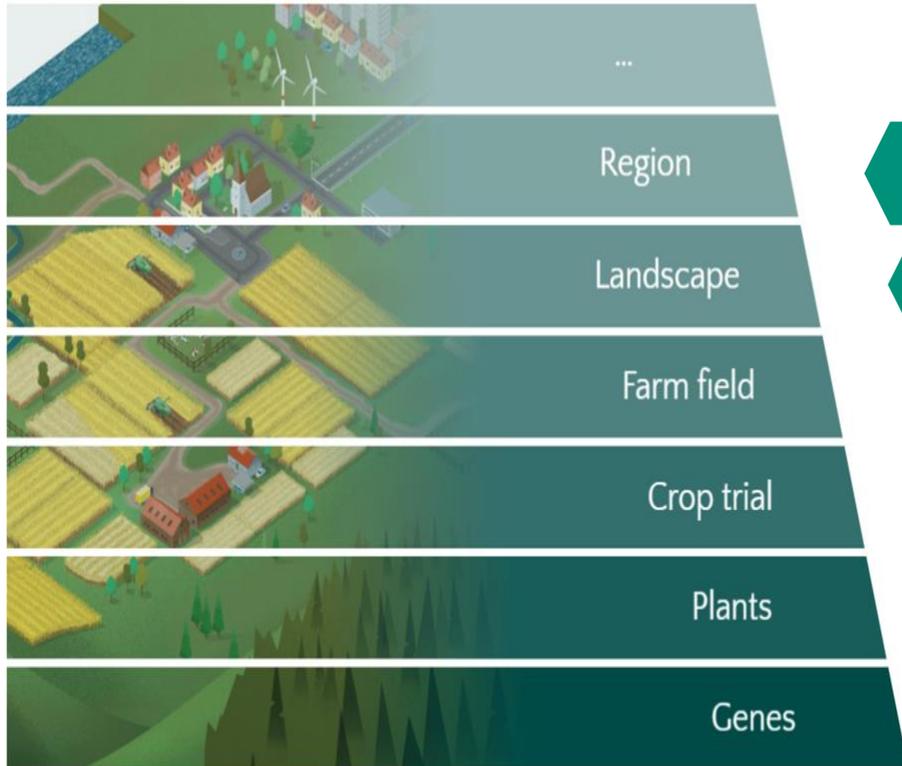


Agrosystem  
community



## Use Cases reflektieren relevante Probleme Agrosystemforschung

### Scales



### 6 Flagship Use Cases

- UC 6 Integrated modeling
- UC 2 Agronomy Long-term experiment
- UC 4 s
- UC 3 Pest & Diseases
- UC 5 Sensor systems
- UC 1 Breeding

### 5 neue Use Cases 1. Call Juni 2024

Fördervolumen:  
UC Pilots: max. 45.000€  
UC Projects: max. 180.000€

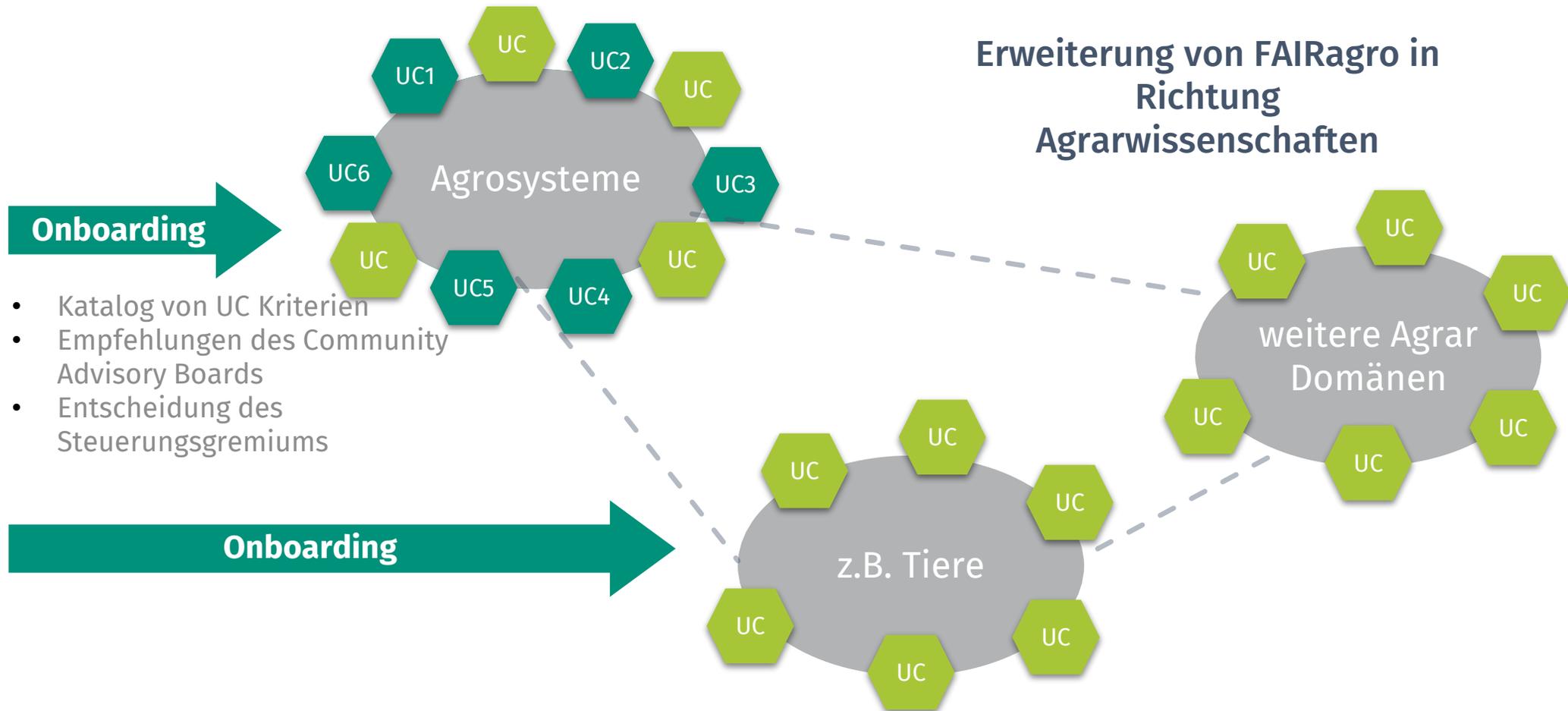
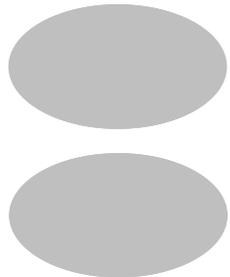
- Show case Tools for Extreme Events
- Pilot 1 Socio-economic Information
- Pilot 2 Synchronization Horticulture Data
- Pilot 3 Arable weed data
- Pilot 4 BrAPI Integration

### neue Use Cases 2. Call Anfang 2025

- ??
- ??
- ??
- ??
- ??

# Weiterentwicklung von FAIRagro

potentielle Use Cases und Domänen



# Das FAIRragro Portal – alles auf einen Blick



## Blog

Wussten Sie schon?

An dieser Stelle veröffentlichen wir in unregelmäßiger Erscheinungsfolge Texte aus dem FAIRragro-Universum. Dinge, die Sie schon immer einmal wissen wollten, Einblicke ins Tagesgeschäft, Antworten auf brennende und nicht so brennende Fragen, Berichte von vor Ort, alles rund um die Welt von FAIRragro.

### Datendokumentation und Wissenschafts-Karma

10. Oktober 2024

So machst du deine Forschung fit für andere. Ein Beitrag von Janes Uhlott und Sophie Boße – Teil 1. Daten gibt es viele – aber welche haben die Qualität, die ich brauche? Die Vorteile von Datenqualitätsinformationen sind enorm. Die Daten sind zuverlässig und die Nutzenden können aufgrund der umfangreichen Informationen für sich selbst bewerten, ob... [weiterlesen](#)

**Kein Urheberrecht, kein Problem? Welche Folgen falsches Zitieren haben kann**



## Toolbox

Wenn Sie mit Agrarforschungsdaten arbeiten, mehr darüber erfahren wollen, diese suchen, veröffentlichen wollen oder sie managen, sind Sie an dieser Stelle richtig. Wir haben für Sie Übersichten und Checklisten zusammengestellt, die Sie beim Umgang mit solchen Daten und Ihren Besonderheiten unterstützen. Unsere Toolbox ist aktuell und wird ständig erweitert. Wenn Sie Ideen und Vorschläge für weitere Beiträge haben, wenden Sie sich gern an uns über unser [Feedback](#).

### Fakten über Fakten – Datentypen in der Agrosystem-Forschung

Welche Datentypen existieren in der Agrosystemforschung? Was kann ich bei Erhebung und Publikation – auch in sozialer Medien – beachten? Hier geben wir einen kurzen Überblick über relevante Datentypen und deren Besonderheiten im Forschungsdatenmanagement. Mit Hinweisen zu unseren Experten im DSSC und weiteren Ressourcen.

EMBL-EMBL  
Erschienen: 9. Juli 2024

### Wo finde ich ein passendes Repositorium?

Sie suchen nach einem geeigneten Repositorium zur Veröffentlichung Ihrer Daten? Sie sind auf der Suche nach gut beschriebenen Daten für Ihre Hochkultur? Dann sind Sie hier richtig – wir helfen Ihnen durch den Durchgang über verschiedene Informationskanäle. Hier finden Sie eine aktuelle Liste mit dem Experten unseres Helpdesks für die besten Tipps.

EMBL-EMBL  
Erschienen: 11. Mai 2024

### Wo finde ich Lern- und Lehrmaterialien zum Thema FDM?

Sie möchten mehr über den Umgang mit Forschungsdaten lernen? Sie benötigen Lehrmaterialien (open educational resources), um eigene Trainings zu entwickeln? Dann sind Sie hier richtig – wir haben für Sie eine Sammlung für Lernende erstellt. Bildungsressourcen zum FDM in den Agrarwissenschaften sind in unserem Dürrenatlas zusammengestellt.

EMBL-EMBL  
Erschienen: 18. Oktober 2024

# www.fairragro.net



## Was ist FAIRragro?

Das FAIRragro-Konsortium mit mehr als 25 Partnern baut ein FAIRes Forschungsdatenmanagement (Findable, Accessible, Interoperable, Reusable) für die Community der Agrosystemforschung auf. Wir entwickeln die passenden Tools und Workflows und schaffen so die Grundlagen für eine nachhaltige Pflanzenproduktion – jetzt und in Zukunft.

## Aktuelles

### Use Case 3: Status und Ausblick

17. Oktober 2024

Der FAIRragro Use Case 3 zielt ab auf eine integrierte Entscheidungshilfe für den Pflanzenschutz, die den potenziellen Ertragsverlust und das Umweltrisiko des... [weiterlesen](#)

### FAIRragro auf dem International Symposium on Integrative Bioinformatics

14. Oktober 2024

Auf dem International Symposium on Integrative Bioinformatics vom 12. bis 13.9.2024 (<https://www.imbio.de/ib/2024/>) an der Universität Zürich präsentierte Jorge

### Datendokumentation und Wissenschafts-Karma

10. Oktober 2024

So machst du deine Forschung fit für andere. Ein Beitrag von Janes Uhlott und Sophie Boße – Teil 1. Daten gibt es... [weiterlesen](#)

## Unsere Use Cases

Über Detail der Aufbau einer Interoperable und stabilen Forschungsdateninfrastruktur in der Agrosystemforschung und darüber hinaus. Sie machen ein Forschungsdaten-FAIR und ermöglichen einen Wissensgewinn.

Um die sich durch agrarwissenschaftliche Entwicklungen ergeben haben wir zunächst sechs Use Cases entwickelt, die auf die vorwiegend aktuellen Herausforderungen und Bedürfnisse der Agrarwissenschaftler\*innen und -ökologinnen\*innen abzielen. Die ersten drei Use Cases sind im September 2024 geplant, die weiteren drei im ersten Halbjahr 2025.

Auf dieser Seite hier finden Sie die Beschreibung der momentan sechs entwickelten Use Cases und Ihre Aufgaben. Wir haben für Sie auch aktuelle Berichte zum Status Quo der einzelnen Use Cases zusammengestellt.



**Use Case 1:** Nutzung der Wechselwirkungen zwischen Genotyp, Standort, Jahr und Bewirtschaftung für eine nachhaltige Pflanzenproduktion  
**Use Case 2:** Bewertung von Trade-offs für ein optimales Sozietätsmanagement bei Pflanzen  
**Use Case 3:** Optimierung von Schädlings- und Krankheitsrisiken zur Förderung des integrierten Pflanzenschutzes

## Helpdesk

### User Data Steward Service Center (DSSC)

Das DSSC ist ein zentraler Service Center für die FAIRragro-Community.

Unsere Team im Data Stewards bietet Ihnen als Teil der Agrarforschungs-Community Unterstützung bei allen Fragen zum Forschungsdatenmanagement. Wir sind in allen Phasen des Datenlebenszyklus – von der Datenreife über die Erstellung eines Datenmanagementplans, bis hin zur Datenveröffentlichung – für Sie da.

**FREE** Dieses FAIRragro Angebot steht der gesamten Agrarforschungs-Community zur Verfügung und ist selbstverständlich kostenfrei.

Kontaktieren Sie uns gern, wenn Sie Unterstützung bei diesen Themen benötigen.

[dssc@fairragro.net](mailto:dssc@fairragro.net)

For Name \*  
Ihre E-Mail-Adresse \*  
Ihre Anfrage \*

## Informieren

Erfahren Sie mehr über die vielfältigen Aktivitäten und neuesten Entwicklungen von FAIRragro.

Abonnieren Sie gerne unseren Newsletter über diese Mailingliste

<https://www.listenvon.de/brympa/info/fairragro-announce>

FAIRragro Newsletter

## Vielen Dank



# Grundprinzipien des professionellen Forschungsdatenmanagements

München, 21.10.24

Sophie Boße (ZB MED Informationszentrum Lebenswissenschaften) im Namen des FAIRagro Konsortiums



project number 501899475

In cooperation with



# Was sind eigentlich Forschungsdaten?

Eine (?) Definition

# Was sind Forschungsdaten?

"Zu Forschungsdaten zählen u.a. **Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben**, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. **Methodische Testverfahren, wie Fragebögen, Software und Simulationen** können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden." [[DFG: Leitlinien zum Umgang mit Forschungsdaten.](#)]



Created by Dewi Novita Sari  
from Noun Project

# Definition Forschungsdaten

Es gibt keinen Konsens, was Forschungsdaten sind

→ Unterschiedliche Definitionen, abhängig von

- ◆ Disziplin
- ◆ Förderer
- ◆ teilweise von Forschungsprojekt zu  
Forschungsprojekt

→ Heute: **digitale** Forschungsdaten



Created by Dewi Novita Sari  
from Noun Project

# Was bedeutet Forschungsdatenmanagement?

Ein Gedankenexperiment

# Gedankenexperiment

Stellen sie sich vor, Lena ruft an!



Created by popcornarts  
from Noun Project



Created by Amethyst Studio  
from Noun Project

# Gedankenexperiment

Stellen sie sich vor, Lena ruft an!

- Wer schafft es innerhalb von 2 Minuten die gewünschten Daten für Lena rauszusuchen?
  - Wird Lena in der Lage sein die Daten zu öffnen?
  - Wird sie in der Lage sein, die Daten zu verstehen?
- ◆ und falls nicht, verstehen sie die Daten noch?



Created by Amethyst Studio  
from Noun Project

# Was es braucht, damit Lena die Daten nutzen kann

Backup



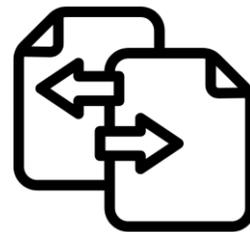
Created by Rafia Rauf  
from Noun Project

Daten-  
organisation



Created by Rafia Rauf  
from Noun Project

Inter-  
operabilität



Created by Rafia Rauf  
from Noun Project

Metadaten



Created by Rafia Rauf  
from Noun Project

Doku-  
mentation



Created by Rafia Rauf  
from Noun Project

# Was macht Daten FAIR?

Prinzipien als Leitfaden fürs FDM

# Definition der “FAIR data” Prinzipien

“**FAIR data**” Prinzipien = ein prägnantes und messbares Set von Grundsätzen, die als Leitfaden dienen können, um die Wiederverwendbarkeit von Datensätzen zu verbessern

**FAIR** steht für:

- **F**indability (Auffindbarkeit)
- **A**ccessibility (Zugänglichkeit)
- **I**nteroperability (Interoperabilität)
- **R**eusability (Wiederverwendbarkeit)



# Findability (Auffindbarkeit)

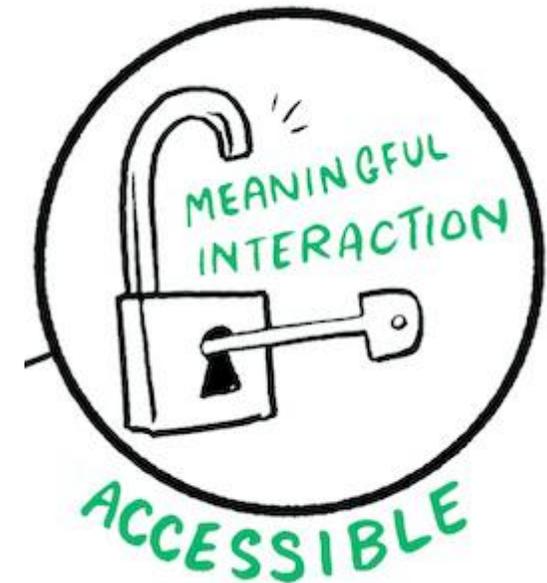
- Daten werden mit umfangreichen **Metadaten** beschrieben
- (Meta-)Daten sind in einer **durchsuchbaren Ressource** (z. B. einem Datenrepositorium) registriert oder indexiert
- (Meta-)Daten erhalten einen weltweit eindeutigen und **persistenten Identifier**



# Accessibility (Zugänglichkeit)

- (Meta-)Daten sind anhand ihrer Identifier über ein **standardisiertes Kommunikationsprotokoll** (z.B. http(s)) abrufbar
- Das Protokoll ist **offen, frei und universell implementierbar**
- Das Protokoll ermöglicht ein **Authentifizierungs- und Autorisierungsverfahren**, falls erforderlich
- Die **Metadaten** sind zugänglich, auch wenn die Daten nicht mehr verfügbar sind

**FAIR** ≠ **FOIR** (O = Open)



# Interoperability (Interoperabilität)

**Dateninteroperabilität** = Fähigkeit eines Datensatzes, mit anderen Datensätzen oder Systemen ohne besonderen Aufwand für den Nutzenden zusammenzuarbeiten

- (Meta-)Daten verwenden eine formale, zugängliche, gemeinsame und breit anwendbare **Sprache für die Wissensdarstellung**
- (Meta)daten verwenden **Vokabulare**, die den FAIR Prinzipien entsprechen
- (Meta)daten enthalten **qualifizierte Verweise** auf andere (Meta)daten
- **offene Dateiformate** und **Open Source Software** werden verwendet



# Reusability (Nachnutzbarkeit)

- Meta(daten) werden mit einer Vielzahl genauer und relevanter **Attribute** reichhaltig beschrieben
- (Meta)Daten sind mit einer detaillierten **Provenienz** verbunden
- (Meta)daten werden mit einer klaren und zugänglichen **Datennutzungslizenz** freigegeben



## Ein gutes Beispiel:

Papoutsoglou, E. A. (2022) „A case-study for improved reusability of plant phenotyping data with MIAPPE“. Zenodo. doi: [10.5281/zenodo.5572772](https://doi.org/10.5281/zenodo.5572772).

# Weiterführende Materialien 1/2

## Lernmaterialien

- [FAIR in \(biological\) practice](#) [The Carpentries Incubator]
- [FAIR sharing and access](#) [MANTRA, The University of Edinburgh]
- [How to be FAIR with your data](#) [Engelhardt *et al.* 2022]
- To assess your knowledge of the FAIR Principles: [FAIR-AWARE](#) [FAIRsFAIR]

## Wie mache ich meine Daten FAIR?

- [FAIR cookbook](#) [ELIXIR]
- [PARTHENOS Guidelines to FAIRify data management and make data reusable](#)
- [Preparing data for sharing: The FAIR principles](#)
- [Top 10 FAIR Data & Software Things](#)

# Weiterführende Materialien 2/2

## Wie bewerte ich die FAIRness eines Datensatzes?

- [How FAIR are your data?](#) [Jones and Grootveld 2017]
- Self-Assessment Tool to Improve the FAIRness of Your Dataset ([SATIFYD](#)) [DANS]
- [FAIR data maturity model indicators](#) [[Bahim et al. 2020](#), Table 1]
- [FAIRevaluatorHelper](#) [Fraunhofer FIT]

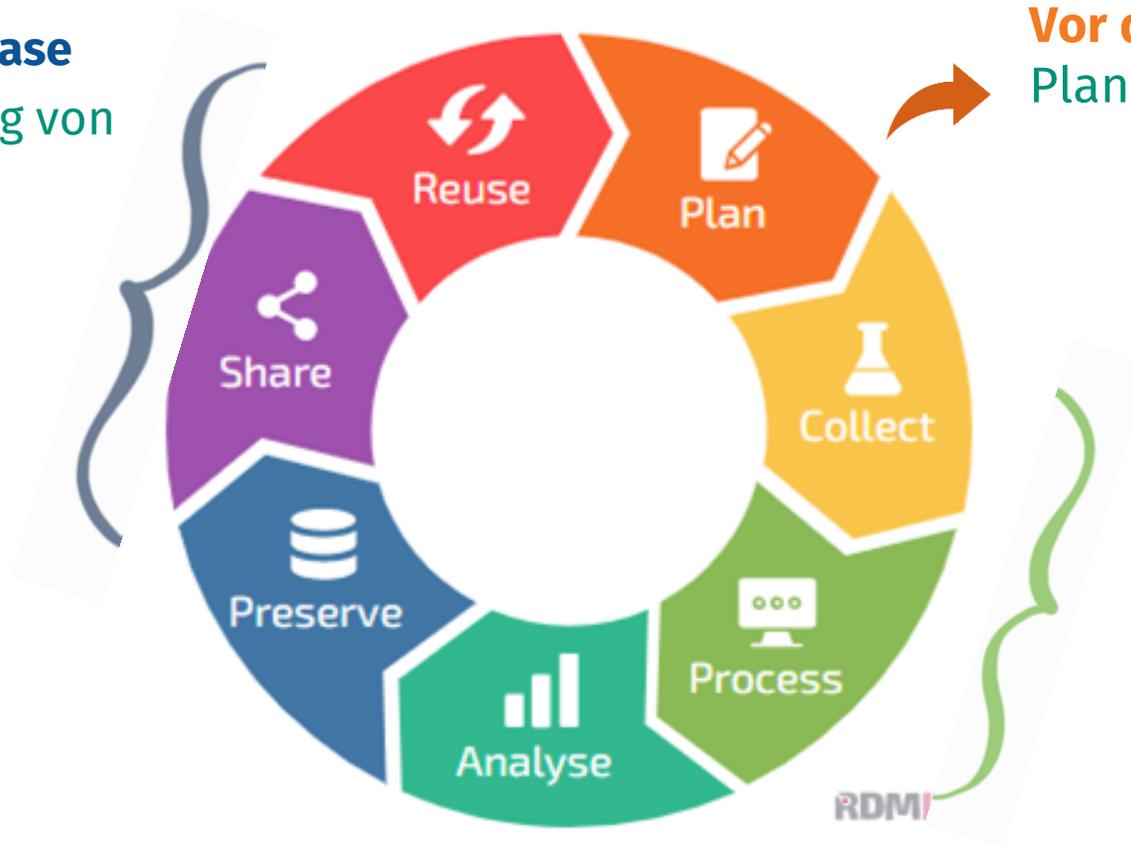
# Wie erreicht man das?

FDM im gesamten Forschungsprojekt

# Forschungsdaten-Lebenszyklus

## Postproduktive Phase

- Veröffentlichung von Daten
- Repositorien
- Lizenzen
- Metadaten
- Rechtliche Aspekte
- Archivierung



## Vor der Umsetzung

Planung aller weiteren Schritte → DMP

## Aktive Projektphase

- Identifizierung der vorhandenen Daten
- Speicherung und Backup
- Datenorganisation
- Datendokumentation
- Sicherheit und Zugang

# Warum ist FDM wichtig?

Auch wenn Lena gar nicht anruft

# Warum FDM wichtig ist - für die Wissenschaft

- Teil der guten wissenschaftlichen Praxis
  - ◆ 11 der 19 Leitlinien der DFG beschreiben (auch) Umgang mit Forschungsdaten
- Nachnutzbare Forschungsdaten sparen Geld, Zeit und Ressourcen
- Gewährleistet Transparenz, Überprüfbarkeit und Reproduzierbarkeit



**Guidelines for Safeguarding  
Good Research Practice**

Code of Conduct

**DFG**

# Warum FDM wichtig ist - für Sie

- Berechtigung zur Finanzierung
- Spart Zeit, Geld und Ressourcen
- Erleichtert die Einarbeitung neuer Mitarbeitender
- Verbessert die Teamarbeit und die Kollaboration
- Hilft, den Überblick über das Projekt zu behalten
- Zitationen, Zitationen, Zitationen....



Created by Amethyst Studio  
from Noun Project

# Weiterführende Ressourcen zu FDM

## Generische Infos zu FDM

- Informationsplattform  
Forschungsdaten -  
[forschungsdaten.info](https://forschungsdaten.info)
- Wiki zu digitalen  
Forschungsdaten -  
[forschungsdaten.org](https://forschungsdaten.org)
- [RDM Lifecycle Checklist](#) vom  
Open Science Framework (OSF)

## Fachspezifische Informationen über FDM

- [RDM Support Pack](#) für Agrarwissenschaften von CGIAR
- [RDMkit](#) für Lebenswissenschaften von elixir europe
- [Sammlung](#) disziplinspezifischer Lehr- und  
Lernressourcen
- FAIRagro [Helpdesk](#)
- FAIRagro [Toolbox](#) - Data-Factsheets, Sammlung von  
Repositorien

# Fragen?



Created by Icons Field  
from Noun Project

# Datenmanagementpläne als Grundlage eines erfolgreichen Forschungsdatenmanagements

**Klug planen, besser verwalten: Die Macht von Datenmanagementplänen in der  
Forschung**

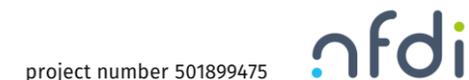
München, 21. Oktober 2024

Wahib Sahwan und Elena Rey Mazón im Namen des FAIRagro Konsortiums

Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) & Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)



In cooperation with



# Inhalt

## 1. *Forschungsdatenmanagement*

- a) Warum Forschungsdatenmanagement?

## 2. *Datenmanagementpläne*

- a) Was ist ein Datenmanagementplan eigentlich?
- b) Warum ist ein Datenmanagementplan wichtig?
- c) Wie kann man das Datenmanagement planen?
- d) Hauptkomponenten eines Datenmanagementplans

## 3. *Entwicklung eines Datenmanagementplans mit Datenmanagementplan-Tools*

- a) Wie erstelle ich einen Datenmanagementplan?
- b) Verfügbare Datenmanagementplan-Tools (DMP-Tools)

# 1. Forschungsdatenmanagement

# Warum Forschungsdatenmanagement (FDM)?

- **FDM** gewährleistet, dass die Forschung zuverlässig, reproduzierbar, ethisch vertretbar und wirkungsvoll ist.
- Eine ordnungsgemäße Datenverwaltung:
  - schützt die Integrität der Forschung,
  - maximiert ihren Nutzen,
  - gewährleistet die Einhaltung rechtlicher und finanzieller Anforderungen
  - und fördert die Zusammenarbeit und den Wissensaustausch.



Data Integrity by bsd studio from [Noun Project](#) (CC BY 3.0)

# Forschungsdaten-Lebenszyklus

- Der Forschungsdaten-Lebenszyklus ist ein **Modell** für das Verständnis der Schritte des Prozesses der wissenschaftlichen Entdeckung.
- Wird je nach Projekt und Datenlage umgesetzt (**angepasst**).
- Idealerweise sollten die **FAIR-Prinzipien** im gesamten Zyklus berücksichtigt werden.



RDM Cycle - RDM Kit - [https://rdmkit.elixir-europe.org/media\\_kit.html](https://rdmkit.elixir-europe.org/media_kit.html)

## 2. Datenmanagementpläne

# Was ist ein Datenmanagementplan eigentlich?

- Ist der **Startpunkt** des Datenlebenszyklus.
- Lebendiges und ständig aktualisiertes Arbeitsdokument.
- **Beschreibt** wie Daten gehandhabt, organisiert und strukturiert werden.
  - von Erhebung & Analyse bis zur Dokumentation am Ende des Projektes und darüber hinaus.



plan by DHAVID TAH HILLAH SAPUTRA from [Noun Project](#) (CC BY 3.0)

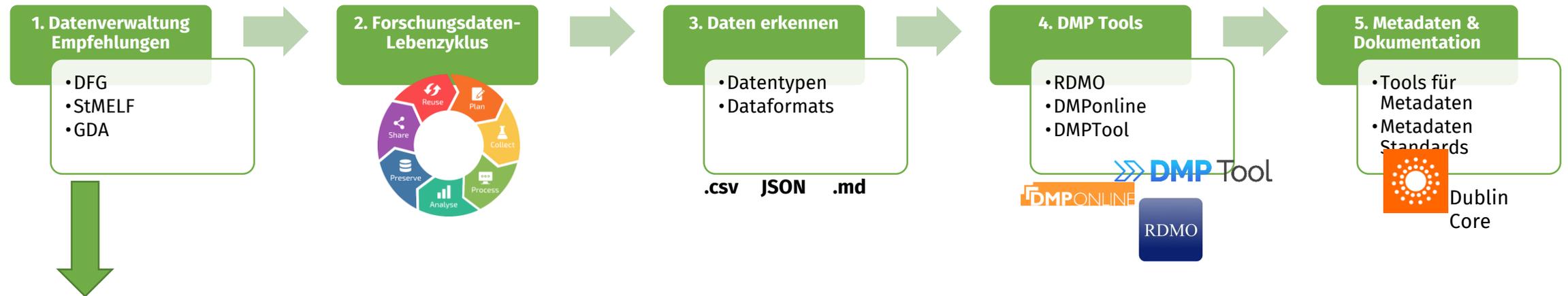
# Warum ist ein Datenmanagementplan wichtig?



important by Rikas Dzihab from [Noun Project](#) (CC BY 3.0)

- Spart Zeit und Ressourcen
- Gute wissenschaftliche Praxis
- Um die Anforderungen einiger Behörden zu erfüllen (Förderstellen)
- Verbesserung der Datensicherheit (weniger Verlust)

# Wie kann man das Datenmanagement planen?



- Deutsche Forschungsgemeinschaft (DFG) – [www.dfg.de](http://www.dfg.de)
  - Fachspezifische Empfehlungen zum Umgang mit Forschungsdaten
- Bayerisches Staatsministerium für Ernährung, Landwirtschaft, Forsten und Tourismus (StMELF)
- Generaldirektion der Bayerischen Staats-archiven (GDA)
  - „Data Governance Act“ (DGA)

# Hauptkomponenten eines Datenmanagementplans

- Datenbeschreibung (Sammlung & Dokumentation)
  - Welche Art von Daten werden erhoben?
    - Quantitative, Qualitative, Geo, Umfragen, ...
    - Numerische, Textuelle, Bilder oder Medien, Code, ...
    - Datenvolumen und Größe
  - Wie werden sie dokumentiert?
    - Digitaler form, Papier, Online-Formulare, ...
  - Wie werden die Daten beschrieben? (Metadaten)
    - Administrative, strukturelle und technische Ebene
    - Metadaten Standards
    - Codebooks oder Data Dictionary?

# Hauptkomponenten eines Datenmanagementplans

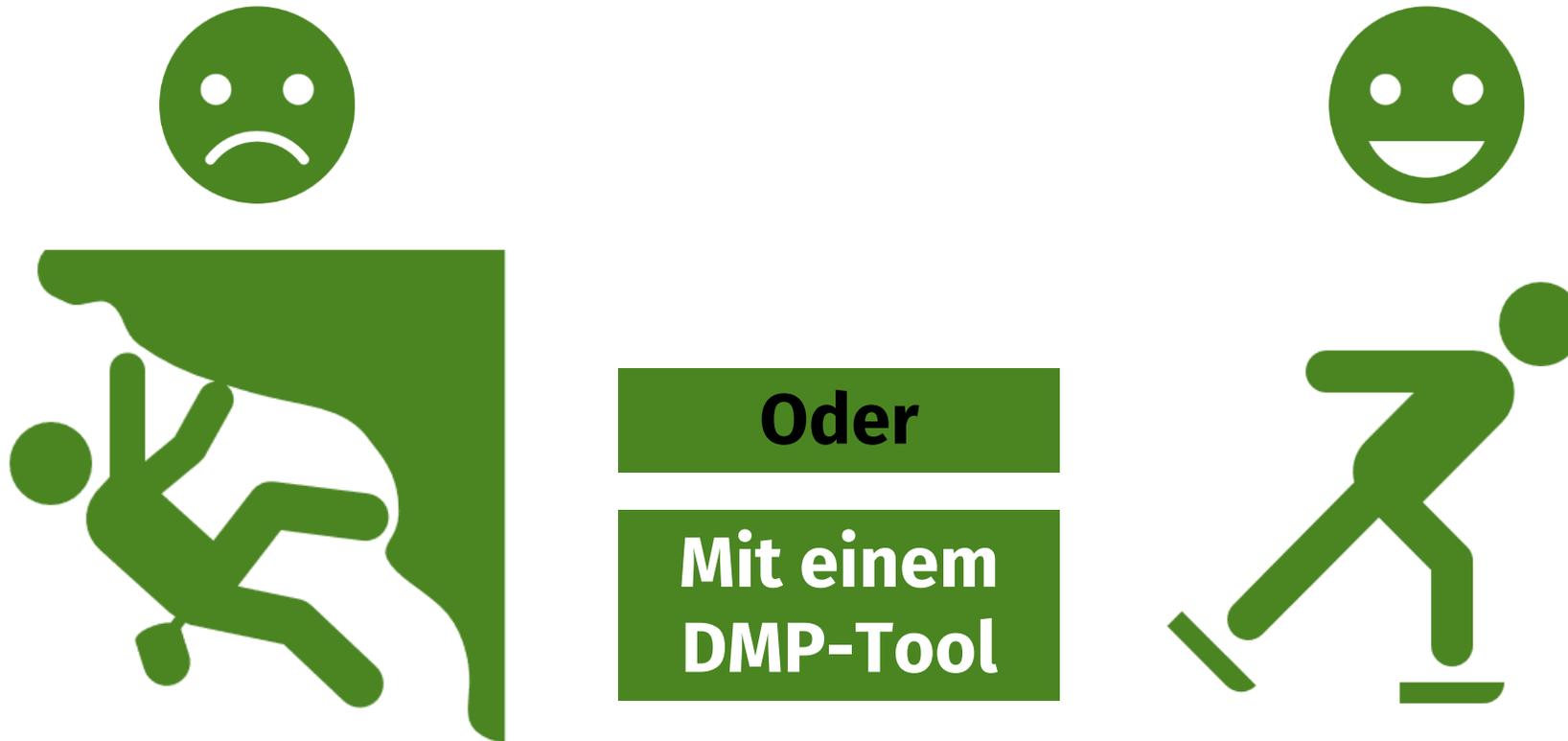
- **Speicherung und Datensicherheit**
  - Wo werden die Daten während und nach der Durchführung des Projekts gespeichert?
  - Sicherheitsmaßnahmen, Backup-Strategien und Zugangskontrolle.
- **Data Sharing und Zugänglichkeit von Daten**
  - Werden die Daten weitergegeben?
  - Wann und wie werden sie anderen zur Verfügung gestellt?
  - Welche Repositorien werden verwendet?

# Hauptkomponenten eines Datenmanagementplans

- **Rechtliche und ethische Überlegungen**
  - Ethische oder rechtliche Aspekte im Zusammenhang mit der gemeinsamen Nutzung sensibler Daten.
  - Datenschutz, geistiges Eigentum und Einhaltung der rechtlichen Rahmenbedingungen
- **Archivierung und Langzeitarchivierung**
  - Wie werden die Daten nach Abschluss des Projekts zur weiteren Verwendung gespeichert?
  - Strategien für die langfristige Archivierung, einschließlich Formate und Speicherorte für die Archivierung der Daten.

### 3. Entwicklung eines Datenmanagementplans mit Datenmanagementplan-Tools

# Wie erstelle ich einen Datenmanagementplan?



# Research Data Management Organiser (RDMO)

## Was ist RDMO?

- Ein webbasiertes Werkzeug zur Unterstützung des Forschungsdatenmanagements (FDM).
- Entwickelt im Rahmen eines DFG-Projekts.

<https://rdmo.forschungsdaten.info/>

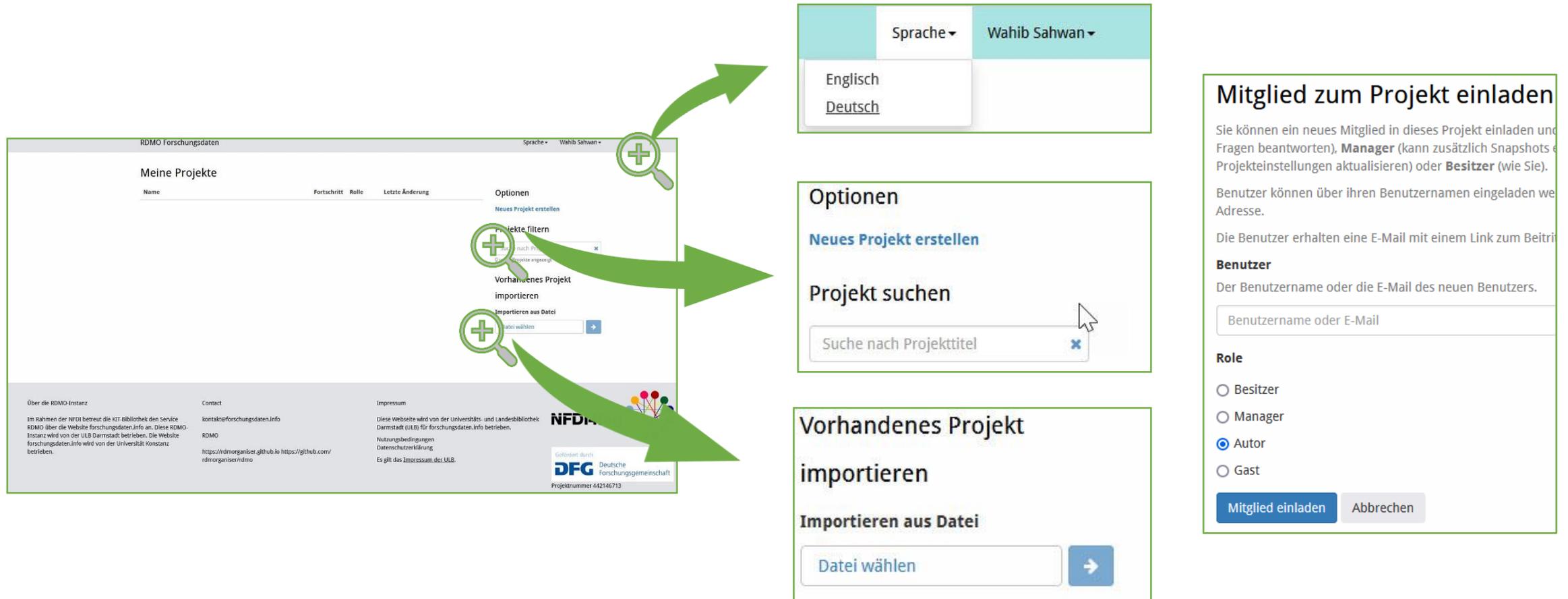
## Wie funktioniert RDMO?

- Strukturierte Fragebögen
- Teamarbeit und Kollaboration
- Zentrale Verwaltung von Informationen.
- Kontinuierliche Updatefähigkeit.
- Übersicht über verschiedene Projekte.



SIGN IN with ORCID 

# Research Data Management Organiser (RDMO)



The image shows a screenshot of the RDMO (Research Data Management Organiser) web interface. The main area is titled "Meine Projekte" and contains a table with columns for "Name", "Fortschritt", "Rolle", and "Letzte Änderung". To the right of the table are several options: "Neues Projekt erstellen", "Projekte filtern", "Vorhandenes Projekt importieren", and "Importieren aus Datei". Three green callout boxes with arrows point to specific features:

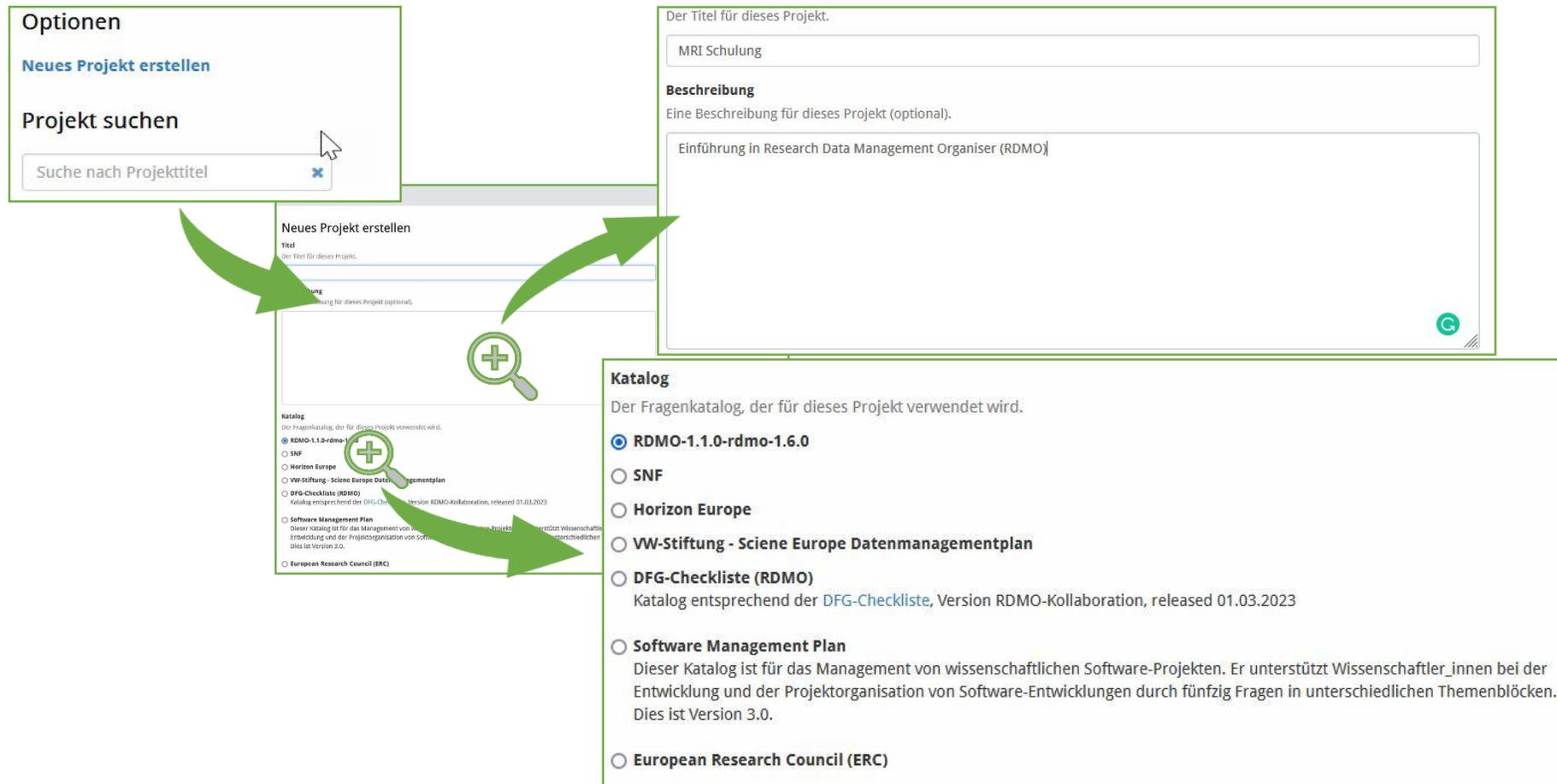
- Top Callout:** A language dropdown menu showing "Sprache" with a dropdown arrow, currently set to "Wahib Sahwan". The menu is open, showing "Englisch" and "Deutsch".
- Middle Callout:** A section titled "Optionen" with a blue button "Neues Projekt erstellen" and a search input field labeled "Projekt suchen" with the placeholder text "Suche nach Projekttitel".
- Bottom Callout:** A section titled "Vorhandenes Projekt importieren" with a blue button "Importieren aus Datei" and a file selection input field labeled "Datei wählen".

On the right side of the interface, there is a form titled "Mitglied zum Projekt einladen". The text explains that users can invite new members and assign roles. The form includes:

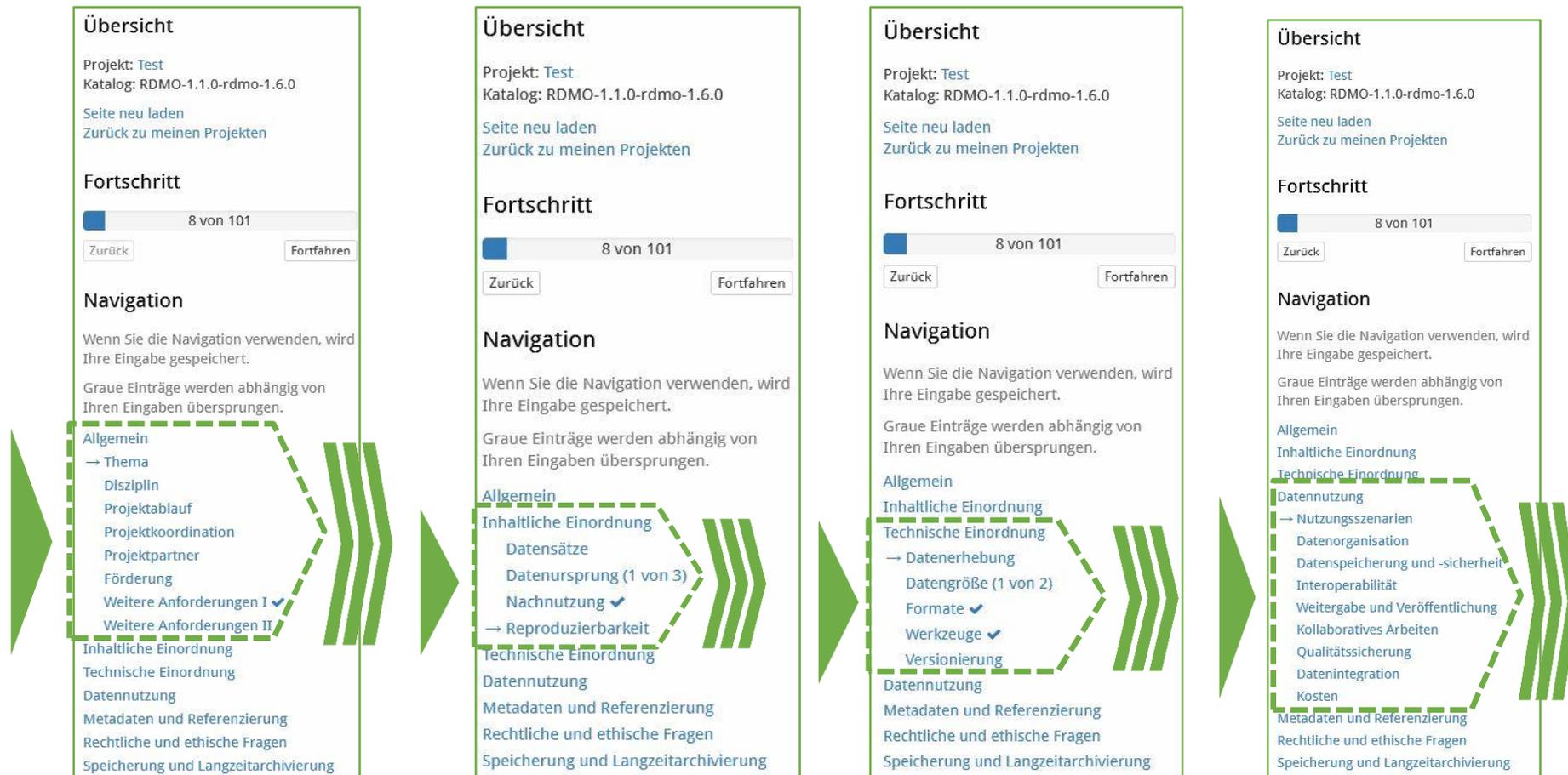
- A text input field for "Benutzername oder E-Mail".
- A section for "Role" with radio buttons for "Besitzer", "Manager", "Autor" (selected), and "Gast".
- Two buttons: "Mitglied einladen" (blue) and "Abbrechen" (grey).

At the bottom of the main interface, there is a footer with information about the RDMO instance, contact details, and logos for NFDI and DFG.

# Research Data Management Organiser (RDMO)



# Research Data Management Organiser (RDMO)



# RDMO: Allgemeine Fragen

**Übersicht**

Projekt: Test  
 Katalog: RDMO-1.1.0-rdmo-1.6.0  
 Seite neu laden  
 Zurück zu meinen Projekten

**Fortschritt**

8 von 101  
 Zurück Fortfahren

**Navigation**

Wenn Sie die Navigation verwenden, wird Ihre Eingabe gespeichert.  
 Graue Einträge werden abhängig von Ihren Eingaben übersprungen.

- Allgemein
- Thema
- Disziplin
- Projekttablauf
- Projektkoordination
- Projektpartner
- Förderung
- Weitere Anforderungen I
- Weitere Anforderungen II
- Inhaltliche Einordnung
- Technische Einordnung
- Datennutzung
- Metadaten und Referenzierung
- Rechtliche und ethische Fragen
- Speicherung und Langzeitarchivierung

- **Thema** (die primäre Forschungsfrage des Projektes)
- **Disziplin** (nach Fachsystematik der DFG)
- **Projekttablauf** (Beginn und Ende des Projekts)
- **Projektkoordination** (Personen oder Institutionen)
- **Projektpartner** (Personen oder Institutionen)
- **Förderung**
- **Weitere Anforderungen**

**Weitere Anforderungen I**

Gibt es von weiteren Seiten (z. B. von der Fachcommunity) Anforderungen an das Datenmanagement, die beachtet werden müssen?

Beispiele für fachspezifische Empfehlungen und Richtlinien sind:

- Richtlinien zum Umgang mit Forschungsdaten in der Biodiversitätsforschung
- Empfehlungen zur Bereitstellung und Nutzung quantitativer Daten in der Bildungsforschung
- Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft
- Empfehlungen zu datentechnischen Standards und Tools sowie zu rechtlichen Fragen bei der Erhebung von Sprachkorpora.

Weitere fachspezifische Empfehlungen und Richtlinien zum Umgang mit Forschungsdaten werden beispielsweise von der Deutschen Forschungsgemeinschaft (DFG) bereitgestellt.

Ja  
 Nein  
 Noch zu klären

# RDMO: Inhaltliche Einordnung

**Übersicht**

Projekt: Test  
 Katalog: RDMO-1.1.0-rdmo-1.6.0  
 Seite neu laden  
 Zurück zu meinen Projekten

**Fortschritt**

8 von 101  
 Zurück Fortfahren

**Navigation**

Wenn Sie die Navigation verwenden, wird Ihre Eingabe gespeichert.  
 Graue Einträge werden abhängig von Ihren Eingaben übersprungen.

**Allgemein**

- Inhaltliche Einordnung
  - Datensätze
  - Datenursprung (1 von 3)
  - Nachnutzung ✓
  - Reproduzierbarkeit
- Technische Einordnung
- Datennutzung
- Metadaten und Referenzierung
- Rechtliche und ethische Fragen
- Speicherung und Langzeitarchivierung

- **Datensätze des Projekts:**
  - Sind sie nachgenutzt (aus einem anderen Projekt)?
  - Werden sie neuerzeugt?
- **Die im Projekt erzeugten Daten:**
  - können nachgenutzt werden?
  - Von wem (Personen, Gruppen oder Institutionen)?
- **Einschätzung der Reproduzierbarkeit und der späteren Archivierung:**
  - Ließe er sich, wenn er verloren ginge, erneut erstellen oder erheben?
  - Sind die Kosten für die Archivierung mehr/weniger als die Kosten für die Reproduzierung?

# RDMO: Technische Einordnung

**Übersicht**

Projekt: Test  
 Katalog: RDMO-1.1.0-rdmo-1.6.0  
 Seite neu laden  
 Zurück zu meinen Projekten

**Fortschritt**

8 von 101

Zurück Fortfahren

**Navigation**

Wenn Sie die Navigation verwenden, wird Ihre Eingabe gespeichert.  
 Graue Einträge werden abhängig von Ihren Eingaben übersprungen.

Allgemein  
 Inhaltliche Einordnung  
**Technische Einordnung**  
 → Datenerhebung  
 Datengröße (1 von 2)  
 Formate ✓  
 Werkzeuge ✓  
 Versionierung  
 Datennutzung  
 Metadaten und Referenzierung  
 Rechtliche und ethische Fragen  
 Speicherung und Langzeitarchivierung

- **Zeitangaben zur:**
  - Datenerhebung / -erstellung
  - Datenbereinigung / -aufbereitung
  - Datenanalyse
  
- **Tatsächliche oder erwartete Größe des Datensatzes**
  
- **Dateiformatauswahl:**
  - **Zu beachten:** die Konsequenzen für die kollaborative Nutzung, die Langzeitarchivierung sowie die Nachnutzung.
  - **Zu empfehlen:** möglichst standardisierte, allgemeine Formate.
  
- **Werkzeuge**
  - Instrumente, Software, Technologien oder Verfahren
  
- **Versionierung:**
  - Werden verschiedene Versionen des Datensatzes erzeugt bzw. wie?

# RDMO: Datennutzung



**Übersicht**

Projekt: Test  
Katalog: RDMO-1.1.0-rdmo-1.6.0

Seite neu laden  
Zurück zu meinen Projekten

**Fortschritt**

8 von 101

Zurück Fortfahren

**Navigation**

Wenn Sie die Navigation verwenden, wird Ihre Eingabe gespeichert.

Graue Einträge werden abhängig von Ihren Eingaben übersprungen.

Allgemein  
Inhaltliche Einordnung  
**Technische Einordnung**  
Datennutzung  
→ Nutzungsszenarien  
Datenorganisation  
Datenspeicherung und -sicherheit  
Interoperabilität  
Weitergabe und Veröffentlichung  
Kollaboratives Arbeiten  
Qualitätssicherung  
Datenintegration  
Kosten  
Metadaten und Referenzierung  
Rechtliche und ethische Fragen  
Speicherung und Langzeitarchivierung

- **Nutzungsszenarien:**
  - Welche Ressourcen (Technische, IT-Ressourcen/Expertise) für die geplante Nutzung der Daten sind nötig?
- **Datenorganisation:**
  - Wo werden die Daten gespeichert?
  - Gibt es Richtlinien dafür bzw. welche?
- **Datenspeicherung und -sicherheit:**
  - Wer darf auf den Datensatz zugreifen?
  - Sollten Backups erstellt werden, bzw. wer erstellt die Backups?
- **Interoperabilität:**
  - Sind die Daten für den Datenaustausch und die Nachnutzung geeignet?

# Verfügbare DMP-Tools

 **DMP** Tool

<https://dmptool.org/>

 **DMP** ONLINE

<https://dmponline.dcc.ac.uk/>

**TUB** -DMP

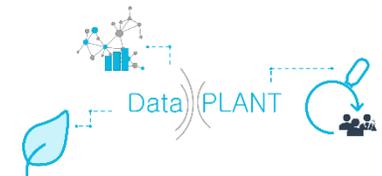
<https://www.tu.berlin/en/ub/szf/tips-tools/planning/tool-tub-dmp>



<https://rdmorganiser.github.io/>

 **DSW**

<https://ds-wizard.org/>



<https://www.nfdi4plants.de/dataplan/>

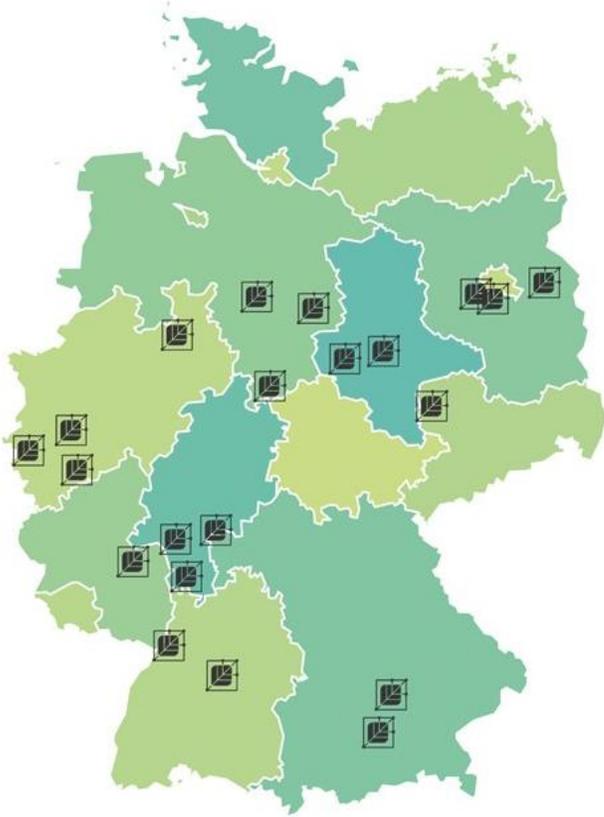
# Gibt es noch Fragen?



questions by Llisole from [Noun Project](#) (CC BY 3.0)

# Vielen Dank für Ihre Aufmerksamkeit!

## ... bei weiteren Fragen kontaktieren Sie gerne unseren FAIRagro Helpdesk!



[fairagro.net/helpdesk](https://fairagro.net/helpdesk)



# Umgang mit Projektdaten: Von der Erhebung über die Organisation bis hin zur Veröffentlichung

FDM Workshop zu Brennpunkten und Lösungsansätzen in der Ressortforschung

21. Oktober 2024

Lucia Vedder & Florian Beyer im Namen des FAIRagro Konsortiums

Rheinische Friedrich-Wilhelms-Universität Bonn,  
Julius-Kühn-Institut

# Inhalt

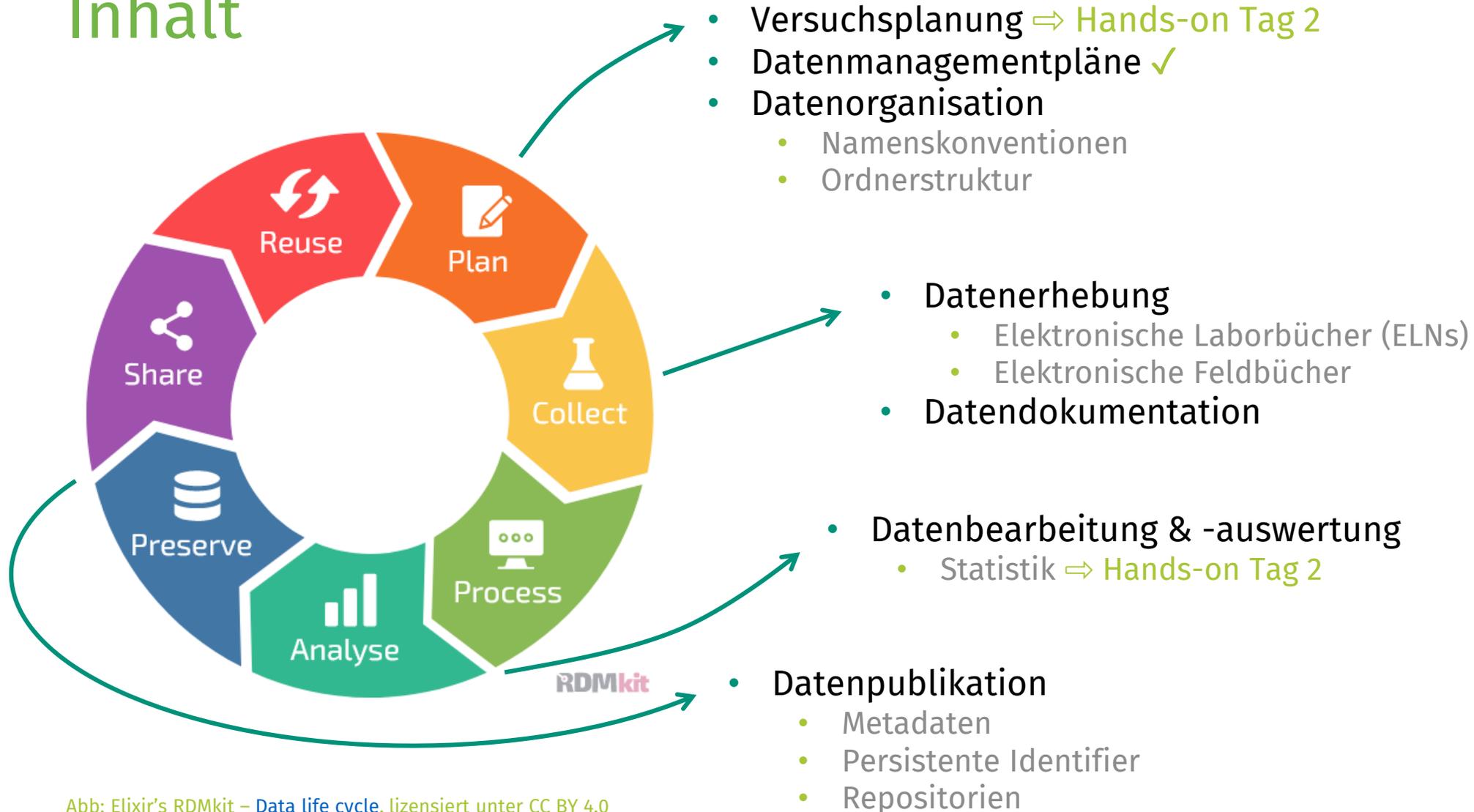


Abb: Elixir's RDMkit – [Data life cycle](#), lizenziert unter CC BY 4.0

# Datenorganisation

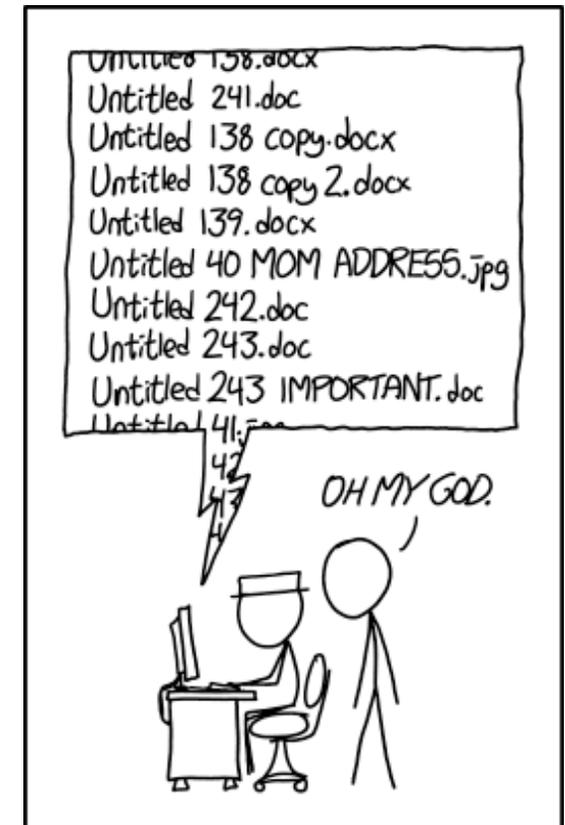


... Namenskonventionen  
... Ordnerstruktur

Abb: Elixir's RDMkit – [Data life cycle](#), lizenziert unter CC BY 4.0

# Namenskonventionen

- Dateinamen sollten es ermöglichen, eine **Verbindung** zu einer bestimmten **Datensammlung** herzustellen
- Innerhalb einer Forschungsgruppe sollte gelten:
  1. Eine Konvention zur Benennung von Dateien und Ordnern erstellen/auswählen
  2. Konvention dokumentieren
  3. Dokumentation allen Mitgliedern der Forschungsgruppe zur Verfügung stellen
  4. Konsistent bleiben!



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Quelle: Verändert nach Bobrov *et al.* 2021 und Bres *et al.* 2022 durch ZB MED 2023 und FAIRagro 2024; „Documents“ Comic von xkcd ist lizenziert unter [CC BY NC](https://creativecommons.org/licenses/by/4.0/)

# Empfehlungen für Namenskonventionen

## Technisch

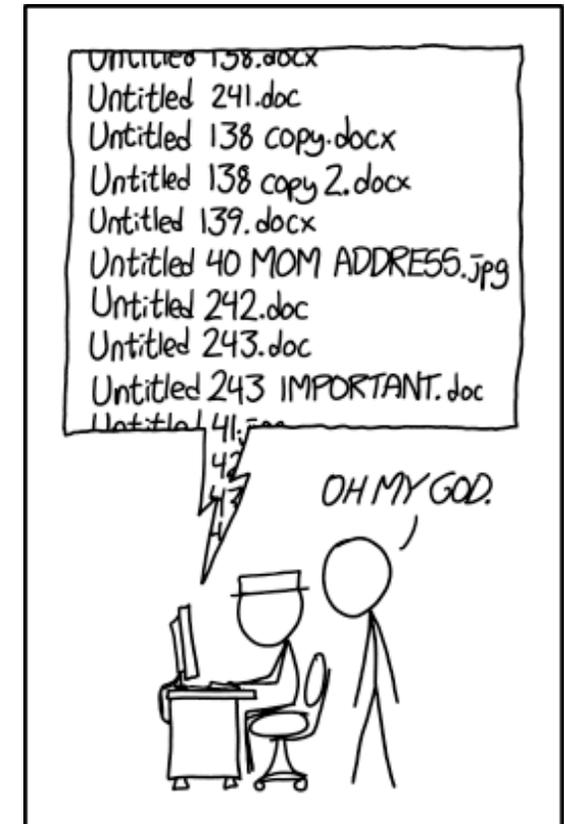
- Keine **Sonderzeichen**
- Keine **Umlaute** o.ä. (ü, ö, ä, ß, é, ...)
- „\_“ oder „-“ statt **Leerzeichen**
- Keine **Großbuchstaben**
- **Max. 256 Zeichen** (inkl. Dateipfad)
  - Dateinamen auf  $\leq 32$  **Zeichen** beschränken

## Inhaltsbezogen

- **Alphabetisch-sortierbare** Namen bevorzugen
- Namen bevorzugen, die den **Inhalt** wiedergeben und **eindeutig** sind
- Standardisierte **Zeit-/Datumsangaben** nutzen
- Autor/Organisation angeben
- Kategorien nutzen
- Versionierung: **Vorangestellte Nullen** nutzen
  - Dateimenge 1 – 99: 01 ... 10 ... 99
  - Dateimenge 1 – 999: 001 ... 010 ... 100 ... 999

# Beispiele für Dateinamen

- **Gute Struktur: JJJ-MM-TT\_XY\_Projekt\_Experiment.format**
  - XY: Initialen des Autors/Erstellers/Bearbeiters/...
  - Projekt: Abkürzung für das Projekt
  - Experiment: Abkürzung/Bezeichnung des Experiments
- **Gute Namen:**
  - 2016-01-04\_ProjectA\_Ex1Test1\_SmithE\_v1.0.xlsx
  - 2000\_USNM\_379221\_01.tiff
  - Projekt001\_Probe045\_MassSpec\_20200824.csv
- **Schlechte Namen:**
  - Test data 2016.xlsx
  - Meeting notes Jan 17.doc
  - Final FINAL last version.docx



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Quelle: Verändert nach Bobrov *et al.* 2021 und Bres *et al.* 2022 durch ZB MED 2023 und FAIRagro 2024; „Documents“ Comic von xkcd ist lizenziert unter [CC BY NC](https://creativecommons.org/licenses/by/4.0/)

# Tools zum gleichzeitigen Umbenennen von Dateien

## Multiple

- [Adobe Bridge](#)
- [jExifToolGUI](#)

## Linux

- [Gnome Commander](#)
- [GPRename](#)

## Mac

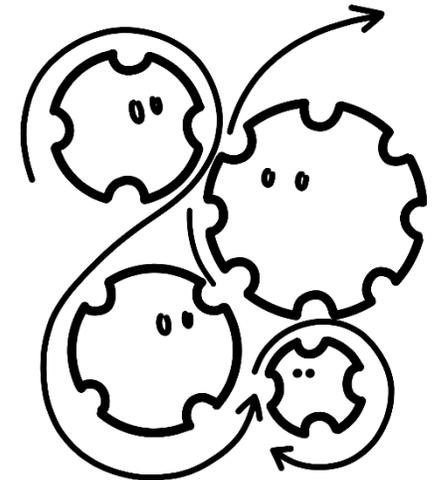
- [ExifRenamer](#)
- [NameChanger](#)
- [Renamer 6](#)

## Unix

- `mv` command

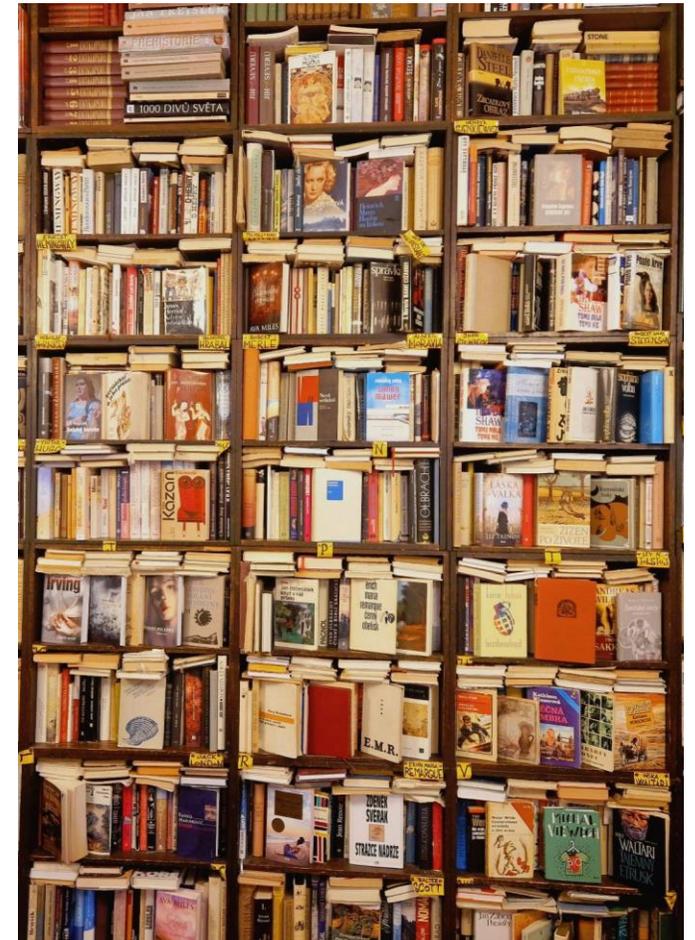
## Windows

- [Advanced Renamer](#)
- [Altap Salamander](#)
- [Ant Renamer](#)
- [Bulk Rename Utility](#)
- [ExifToolGUI](#)
- [Rename-IT](#)
- [Total Commander](#)
- [WildRename](#)



# Gründe für eine systematische Ordnerstruktur

- Zeitersparnis bei der Suche nach Dateien
- Verbessertes Arbeitsablauf während des Projekts
- Erleichterung von Kollaborationen

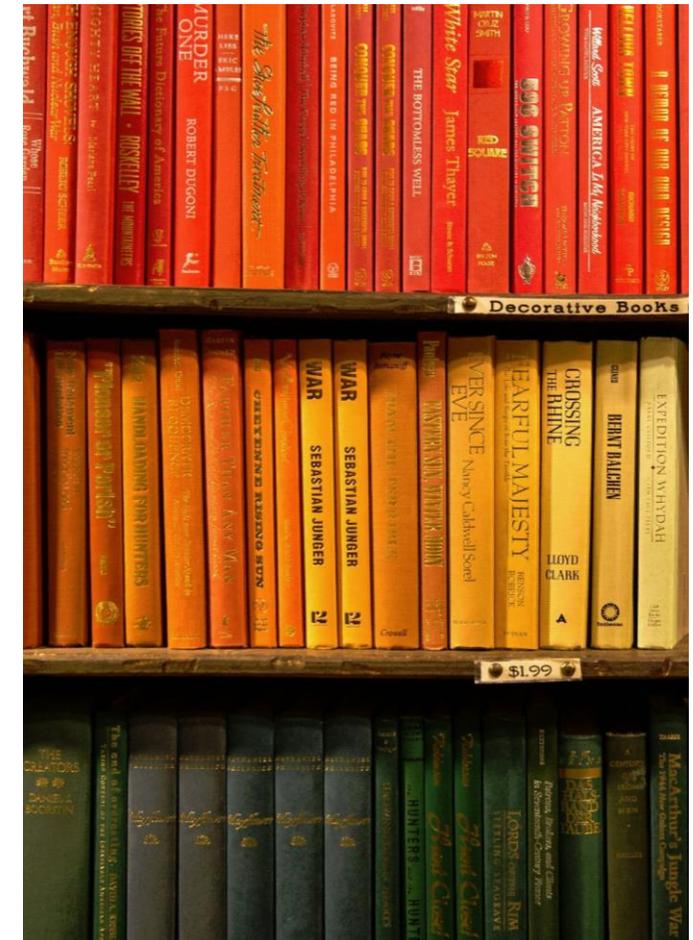


Quelle: Verändert nach Goldberg 2020 durch ZB MED 2024 und FAIRagro 2024; Foto von [Vladalek](#) auf [Unsplash](#)

# Gründe für eine systematische Ordnerstruktur

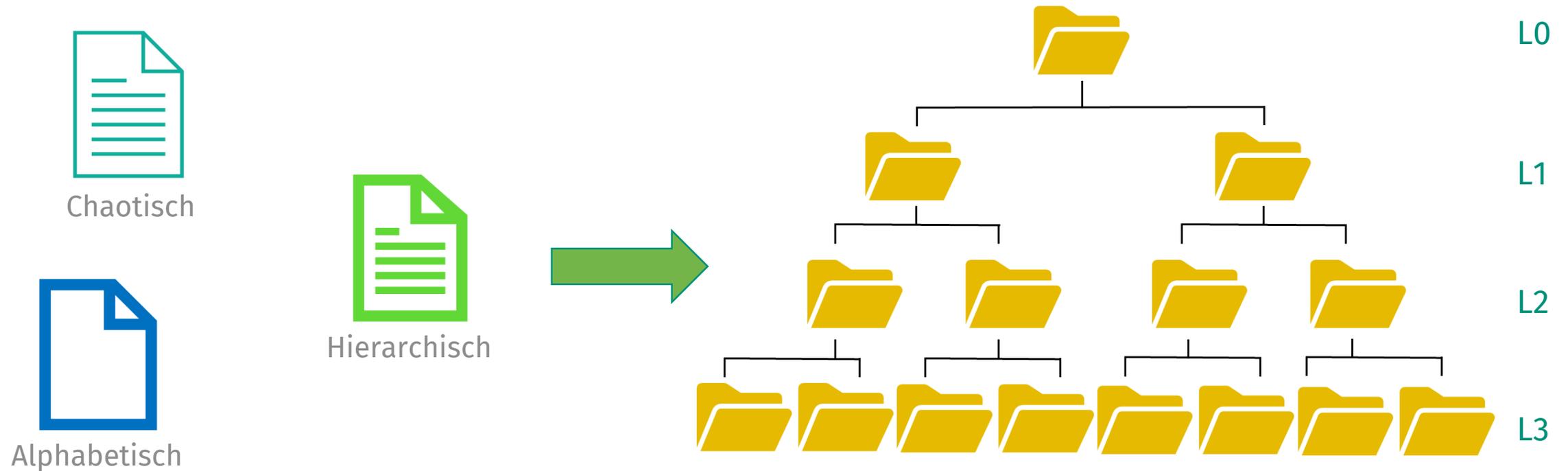
- Zeitersparnis bei der Suche nach Dateien
- Verbessertes Arbeitsablauf während des Projekts
- Erleichterung von Kollaborationen

Stellen Sie sich vor: Jemand (vielleicht Ihr zukünftiges Ich) sieht sich Ihre Dateien und Ordner an und versteht sofort im Detail, was Sie gemacht haben und warum...



Quelle: Verändert nach Goldberg 2020 durch ZB MED 2024 und FAIRagro 2024; Foto von [Jason Leung](#) on [Unsplash](#)

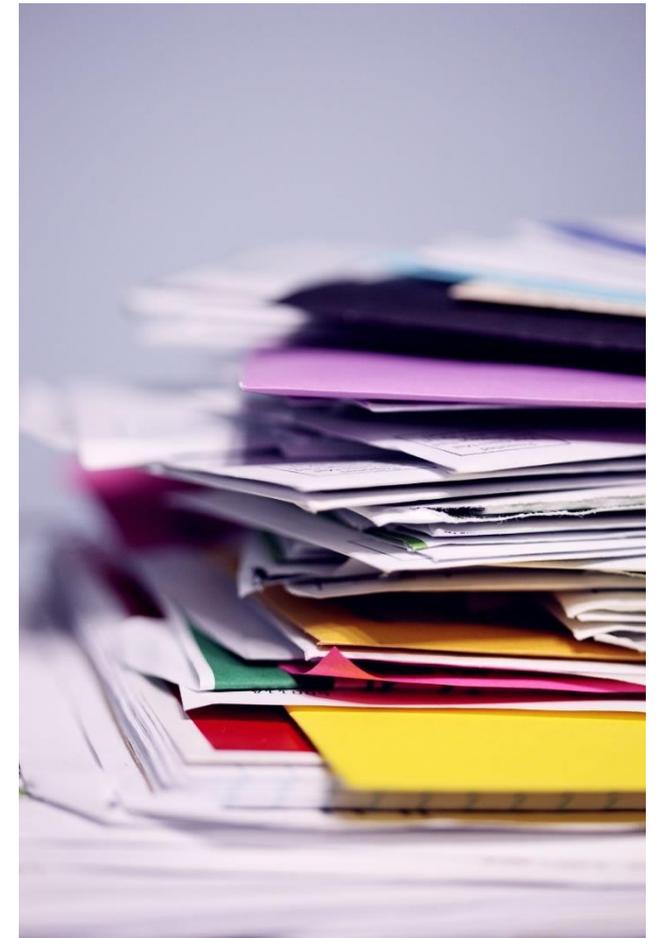
# Empfehlungen für eine Ordnerstruktur



Quelle: Verändert nach Ireneusz Kleppert 2024 durch ZB MED 2024 und FAIRagro 2024

# Empfehlungen für eine Ordnerstruktur

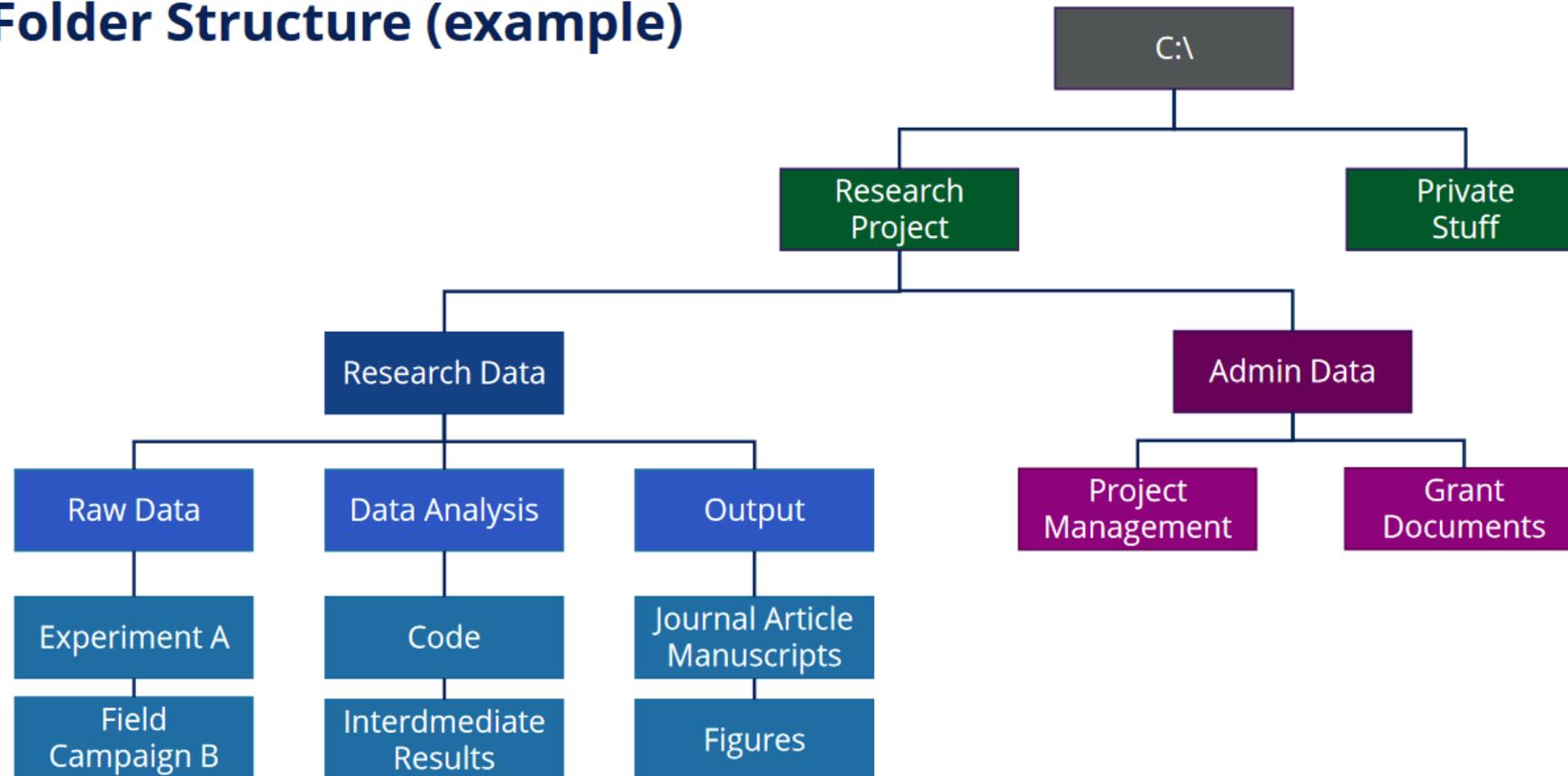
- Überlappende Kategorien vermeiden
  - Keine Kopien von Dateien in verschiedenen Ordnern!
- Namenskonventionen anwenden
- Maximal 4 Level
- Maximal 10 Elemente je Ordner
- Generische Ordner vermeiden
  
- Ein System für die gesamte Gruppe nutzen
  - Zeit zur Planung investieren!



Quelle: Verändert nach Bobrov *et al.* 2021 and Bres *et al.* 2022 durch ZB MED 2023 und FAIRagro 2024; Foto von [Alexander Grey](#) auf [Unsplash](#)

# Beispiel einer Ordnerstruktur

## Folder Structure (example)



Quelle: Briney *et al.* 2020

# Datenerhebung

... Elektronische Laborbücher (ELNs)  
... Elektronische Feldbücher



Abb: Elixir's RDMkit – [Data life cycle](#), lizenziert unter CC BY 4.0

# Elektronische Laborbücher (ELNs): Definition

Elektronische Laborbücher („Electronic Lab Notebooks“, ELNs) = Eine spezielle Software für:

- die strukturierte Beschreibung eines Experiments, d.h. das Führen von Aufzeichnungen über
  - die Vorgehensweise
  - die verwendeten Materialien und Instrumente
  - die während des Experiments gemachten Beobachtungen (inkl. Fotos)
  - die Metadaten/Beschreibung der im Rahmen des Experiments erzeugten Daten und ggf. der Dateien (oder den Links zu ihnen) und der damit verbundenen Schlussfolgerungen
- die langfristige Aufbewahrung und der Zugang zu all diesen Informationen unter Einhaltung der Kriterien für die Langzeitarchivierung



Created by Dewi Novita Sari  
from Noun Project

Quelle: [ELN- Finder \(20.0624\)](#); „[definition](#)“ von Dewi Novita Sari ist lizenziert unter [CC BY 3.0](#)

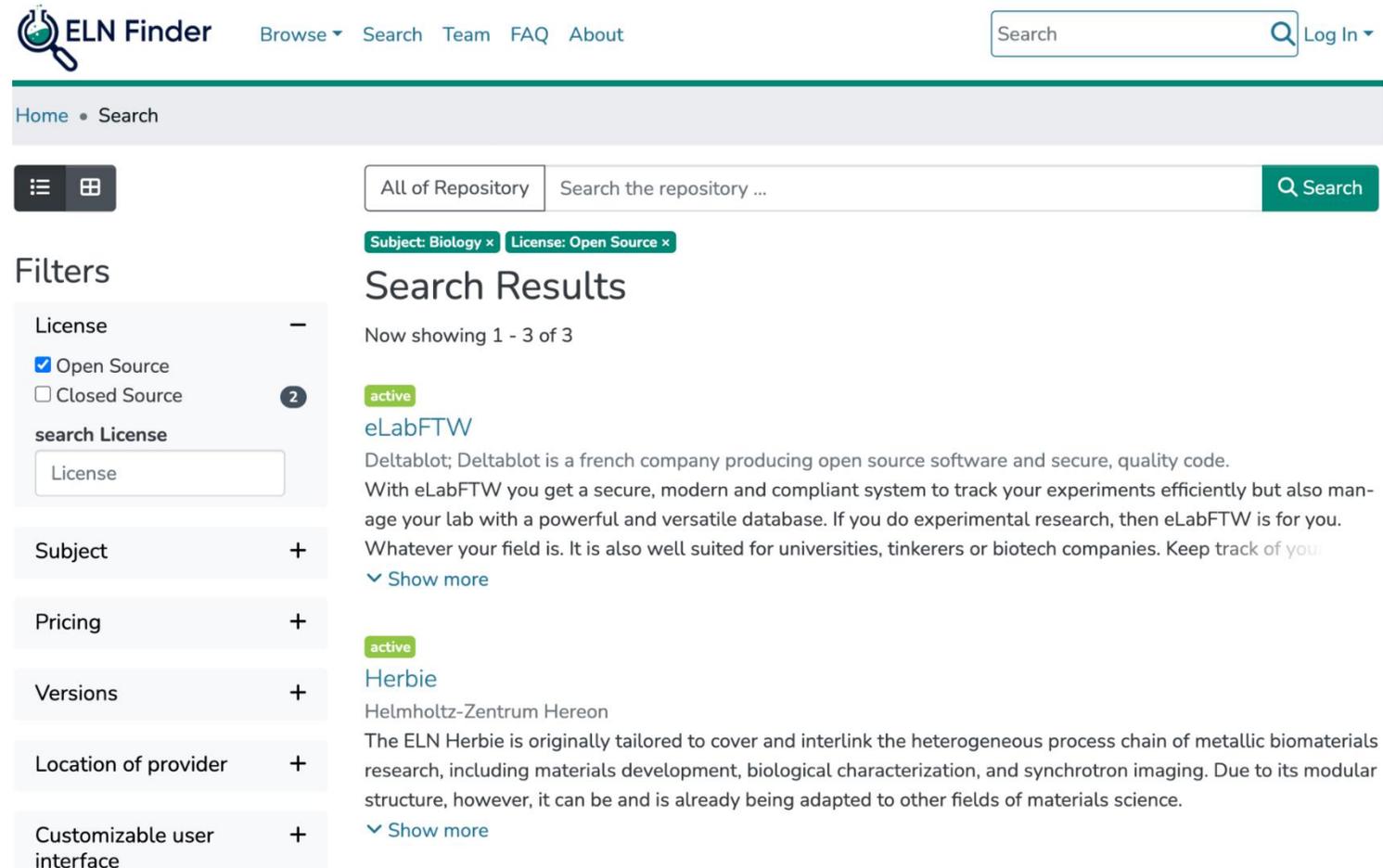
# Ein passendes ELN finden

## ZB MED ELN Guide ([de/en](#))

- Zielgruppe
  - Wissenschaftler
  - Wissenschaftsinfrastruktur Teams
- Ziele
  - Auswahl eines ELNs
  - Nützliche Referenzen

## ELN Finder

- Interaktives Tool zum Filtern von ELNs, basierend auf 40 Kriterien



The screenshot shows the ELN Finder website interface. At the top, there is a navigation bar with the ELN Finder logo, a search bar, and links for Browse, Search, Team, FAQ, and About. Below the navigation bar, there is a search bar with the text "Search the repository ..." and a "Search" button. The main content area is divided into two columns. The left column contains a "Filters" section with the following options: License (Open Source checked, Closed Source unchecked), Subject, Pricing, Versions, Location of provider, and Customizable user interface. The right column contains a "Search Results" section with the text "Now showing 1 - 3 of 3". The first result is "eLabFTW" (active), described as a French company producing open source software. The second result is "Herbie" (active), described as an ELN originally tailored for metallic biomaterials research.

Quelle: Verändert nach Vandendorpe 2021, Vandendorpe *et al.* n.d. durch ZB MED 2023 und FAIRagro 2024; Screenshot von [ZB MED et al. 2021](#), Screenshot vom [ELN Finder](#)

# Elektronische Feldbücher

- Anwendungen (Apps) zur digitalen Datenerhebung (z.B. von Phänotypisierungsdaten) im Feld
- Digitale Formulare können offline ausgefüllt werden
- Datentransfer in eine Cloud möglich
- Beispiele: [Bonitur App \(JKI\)](#), [ODK](#), [KDSmart](#), [Field Book](#)

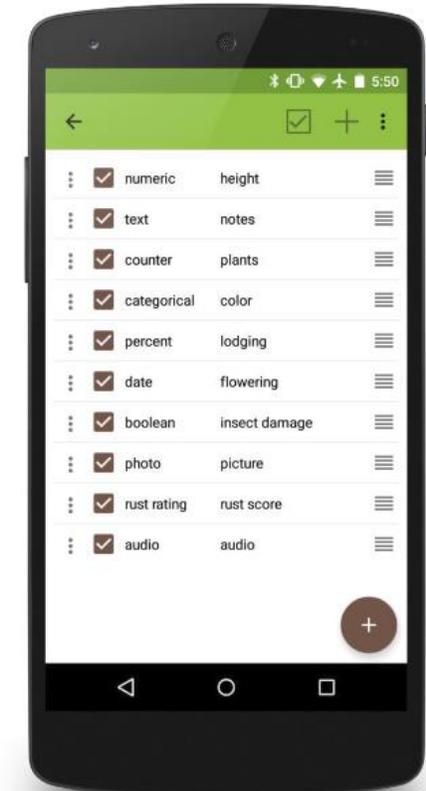
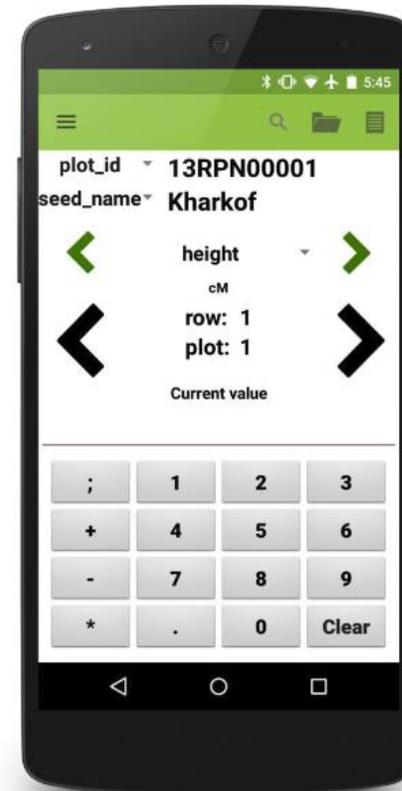


Abbildung: [Field Book Manual](#)

# Datendokumentation



Abb: Elixir's RDMkit – [Data life cycle](#), lizenziert unter CC BY 4.0

# Gründe für eine Datendokumentation

Für die Wiederverwendung von Forschungsdaten müssen diese **sinnvoll und so detailliert wie möglich beschrieben** werden

- für andere Forscher
- für andere Teammitglieder
- für das zukünftige Ich

Es ist daher immer am besten zu dokumentieren, **wie, wann und zu welchem Zweck** die Daten erstellt wurden, **sobald sie entstanden sind**.

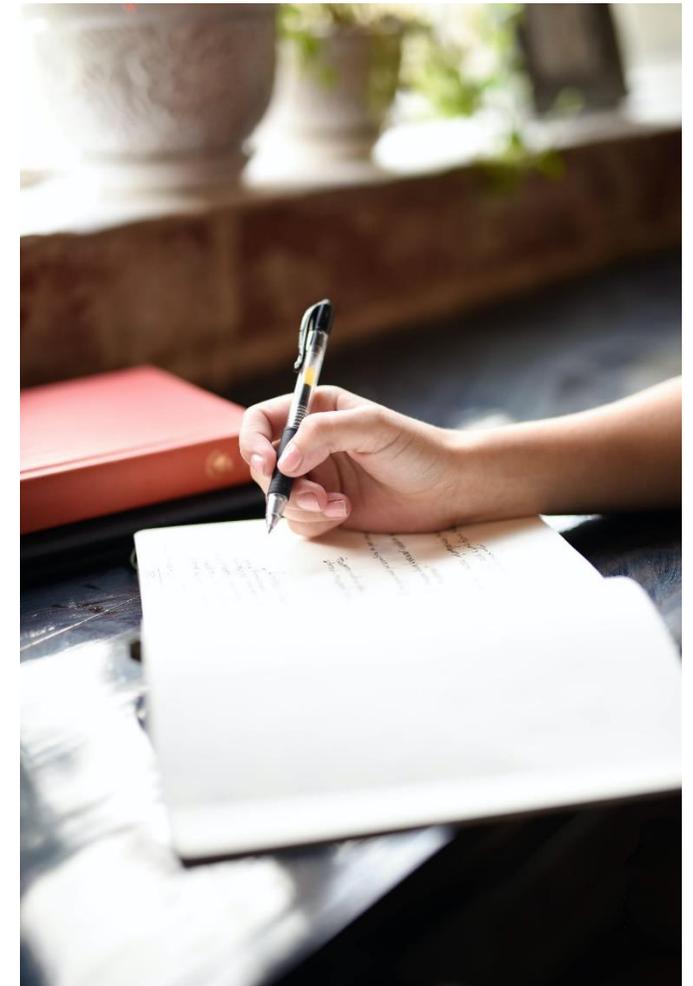
Eine gute Dokumentation macht Daten

- **verständlich**
- **wiederverwendbar** und
- garantiert die **Reproduzierbarkeit** von Forschung.

Quelle: [forschungsdaten.info](https://www.researchdata.info/) und verändert nach Engelhardt *et al.* 2022 und KU Leuven 2022 durch ZB MED 2023 und FAIRagro 2024

# Datendokumentationstypen

- Daten Dictionary
- Codebook
- (Elektronische) Labor- /Feldbücher
- Kommentierter Code
- README Datei
- Standard Operating Procedures (SOPs)
- Laborprotokolle
- User Guides / Nutzungsrichtlinien
- ...



Quelle: Verändert nach KU Leuven 2022 durch ZB MED 2023 und FAIRagro 2024; Foto von [Hannah Olinger](#) auf [Unsplash](#)

# Inhalte einer Datendokumentation

- Beschreibung des Forschungsprojekts
- Ziele des Projekts
- Hypothesen
- Informationen zur Datenerhebung
- Maßnahmen zur Datenbereinigung und Schritte der Datenverarbeitung
- Struktur der Daten und Beziehung der Dateien zueinander
- Erläuterung von Variablen, Bezeichnungen und Codes
- Unterschiede zwischen verschiedenen Versionen
- Informationen über den Zugang und die Nutzungsbedingungen



# Datenpublikation

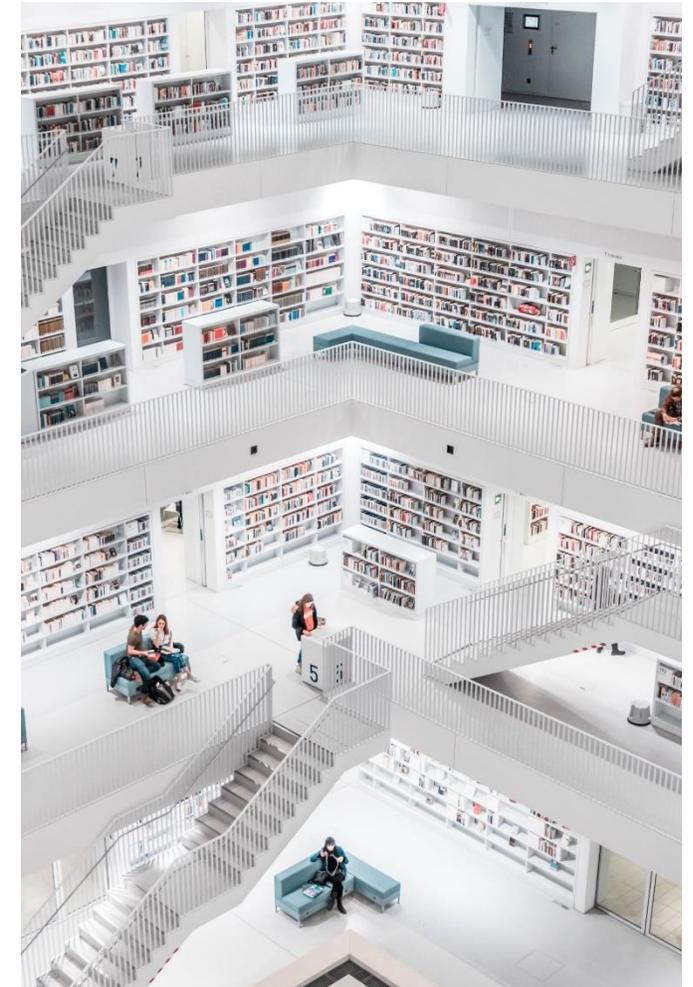


- ... wiss. Publikation
- ... Daten-Repos
- ... Metadaten(standards)
- ... PIDs

Abb: Elixir's RDMkit – [Data life cycle](#), lizenziert unter CC BY 4.0

# Vorteile von Datenpublikationen

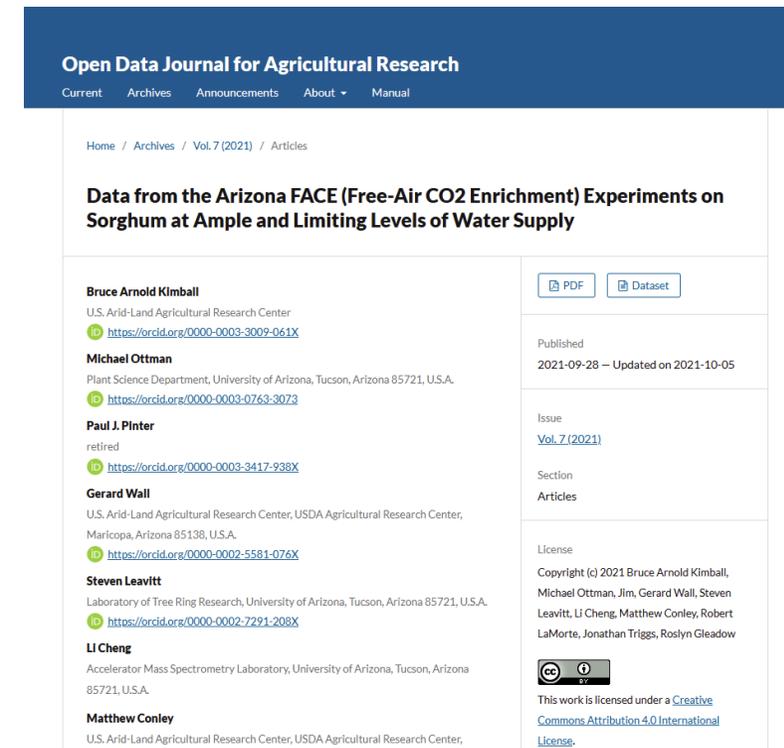
- Daten langfristig **verfügbar, wiederverwendbar** und **zitierfähig** machen
- Daten zu einem **unabhängigen wissenschaftlichen Ergebnis** machen
- Möglichkeiten für Meta-Analysen schaffen
- Erhöhung der **Sichtbarkeit, Transparenz** und **Verantwortlichkeit**
- ...
- **Credits** als Wissenschaftler:in / Autor:in



Quelle: Verändert nach Rathmann *et al.* 2021 durch ZB MED 2023 und FAIRagro 2024; Foto von [Sara Kurfuß](#) auf [Unsplash](#)

# Potentielle Datenpublikationen

- Anhang an Artikel (peer-reviewed) : Data supplements (Bsp.: [nature](#))
- Daten-Journals / Datenartikel ([Liste](#))
  - Bsp.: [Open Data Journal for Agricultural Research](#), [Earth System Science Data](#), [Scientific data \(Nature\)](#)
- Eigenständige Datenpublikation
  - Disziplinspezifische Daten-Repositoryn
  - Interdisziplinäre Repositoryn
  - Institutionseigene Repositoryn



Open Data Journal for Agricultural Research

Current Archives Announcements About Manual

Home / Archives / Vol.7 (2021) / Articles

### Data from the Arizona FACE (Free-Air CO<sub>2</sub> Enrichment) Experiments on Sorghum at Ample and Limiting Levels of Water Supply

[PDF](#) [Dataset](#)

**Bruce Arnold Kimball**  
U.S. Arid-Land Agricultural Research Center  
<https://orcid.org/0000-0003-3009-061X>

**Michael Ottman**  
Plant Science Department, University of Arizona, Tucson, Arizona 85721, U.S.A.  
<https://orcid.org/0000-0003-0763-3073>

**Paul J. Pinter**  
retired  
<https://orcid.org/0000-0003-3417-938X>

**Gerard Wall**  
U.S. Arid-Land Agricultural Research Center, USDA Agricultural Research Center, Maricopa, Arizona 85138, U.S.A.  
<https://orcid.org/0000-0002-5581-076X>

**Steven Leavitt**  
Laboratory of Tree Ring Research, University of Arizona, Tucson, Arizona 85721, U.S.A.  
<https://orcid.org/0000-0002-7291-208X>

**Li Cheng**  
Accelerator Mass Spectrometry Laboratory, University of Arizona, Tucson, Arizona 85721, U.S.A.

**Matthew Conley**  
U.S. Arid-Land Agricultural Research Center, USDA Agricultural Research Center,

Published  
2021-09-28 — Updated on 2021-10-05

Issue  
Vol. 7 (2021)

Section  
Articles

License  
Copyright (c) 2021 Bruce Arnold Kimball, Michael Ottman, Jim, Gerard Wall, Steven Leavitt, Li Cheng, Matthew Conley, Robert LaMorte, Jonathan Triggs, Roslyn Gleadow

  
This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Screenshot from [Open Data Journal for Agricultural Research](#)

# Repositorien: Definition

## Daten-Repositorium:

Ort, an dem digitale (und physische) Objekte gespeichert und dokumentiert werden und der die separate Veröffentlichung und Archivierung dieser Objekte ermöglicht.

Der Zugang zu den Daten kann entweder offen oder auf eine Gruppe von Nutzern beschränkt sein.

Automatisches Handling der Standards

... Metadaten

... PIDs



Created by Dewi Novita Sari  
from Noun Project

Quelle: Verändert nach Engelhardt *et al.* 2022 und Rathmann *et al.* 2021 durch ZB MED 2022 und FAIRagro 2024; „[definition](#)“ von Dewi Novita Sari ist lizenziert unter [CC BY 3.0](#)

# Repositorien: Klassifikation

	Institutionell	Interdisziplinär	Disziplin-spezifisch
<b>Umfang</b>	Alle eigenen Fachbereiche der Institution	Interdisziplinäre und/oder heterogene Datensätze	Sichtbarkeit innerhalb der Forschungscommunity
<b>Service</b>	Forschungsdaten-management (FDM) Service	Datenkuratierung und Qualitätskontrolle nicht gewährleistet	FDM Expertise und spezialisierte Tools
<b>Beschränkungen</b>	Uneingeschränkter Zugang für Mitarbeiter	Ungeeignet für sensible Daten, die möglicherweise in ihrem Ursprungsland verbleiben müssen	Daten von Wissenschaftlern aus verschiedenen Einrichtungen
<b>Beispiele</b>	<a href="#">Apollo</a> , <a href="#">ETH Zurich's Research Collection</a> , <a href="#">DepositOnce</a> , ...	<a href="#">DRYAD</a> , <a href="#">Figshare</a> , <a href="#">RADAR</a> , <a href="#">TIB AV-Portal</a> , <a href="#">Zenodo</a> , ...	Siehe folgende Folien ( <a href="#">PANGAEA</a> )

Quelle: Verändert nach Cozatl et al. 2021 und Engelhardt et al. 2022 durch ZB MED 2023 und FAIRagro 2024

# Repositorien für die Agrarforschung

- [BonaRes](#) Repository (betrieben am ZALF)
  - Datensätze zu Landwirtschaft, Boden und Landschaftsökologie
  - Möglichkeit, die Daten in Bezug auf ihre geografische Lage zu veröffentlichen
- [e!DAL-PGP](#) Repository (betrieben vom IPK Gatersleben)
  - Daten zu Phänotypisierung und Genotypisierung in der Pflanzenforschung
- [PUBLISSO](#) Repository für Lebenswissenschaften (FRL) (betrieben von ZB MED)
  - Sekundärveröffentlichung von Artikeln, Büchern, Proceedings und Forschungsdaten im Bereich der Biowissenschaften
- [OpenAgrar](#) (betrieben vom BMEL)
  - Repository der Bundesbehörden, in dem Textpublikationen und Forschungsdaten aus dem Bereich der Agrarwissenschaften veröffentlicht
- ...

# Repositorien für die Agrarforschung

Empfohlene Repositorien des FAIRagro Helpdesks für die Agrosystemforschung							
Uneingeschränkter Zugriff							
Service Name	Link	Domäne	Kurzbeschreibung (deu)	Art der Ressourcen	Besonderheiten bei der Datensuche / Publikation	Lizenzen	Zugriff (siehe Besonderheiten)
BonaRes Repository	<a href="https://maps.bonares.de/mapaps/resources/apps/bonares/index.html?lang=en">https://maps.bonares.de/mapaps/resources/apps/bonares/index.html?lang=en</a>	Boden- und Agrarforschungsdaten	Publikation von (inter-)nationalen Daten zu Boden, Agrarforschung und Ökologie	CSV, Excel, File-Geodatabase, TXT, Datensets	Spezifische Beschreibung einzelner Tabellenspalten, Embargo möglich	CC-BY	uneingeschränkt
GFZ Data Services	<a href="https://bib.telegrafenberg.de/dataservices/">https://bib.telegrafenberg.de/dataservices/</a>	Geowissenschaften	Offenes Repository für Daten mit Geobezug, betrieben vom Deutschen GeoForschungsZentrum Potsdam	Vorwiegend Datensätze und Software	Suche über Verortung möglich / komfortable Metadaten-Eingabe	CC0, CC-BY, CC-BY-SA	uneingeschränkt
PANGAEA	<a href="https://www.pangaea.de/">https://www.pangaea.de/</a>	Geosphäre, Biosphäre, Cryosphäre, Atmosphäre und mehr	Von AWI und MARUM betriebenes Repo für georeferenzierte Erdwissenschaftsdaten, CORE TRUST-zertifiziert	Datensätze (georeferenziert)	-	CC-BY, möglicherweise weitere	uneingeschränkt
EDI Data Portal	<a href="https://edirepository.org/">https://edirepository.org/</a>	Umweltwissenschaften	Repository der Umweltwissenschaften, gefördert von der U.S. National Science Foundation umgesetzt durch zahlreiche Partner, CORE TRUST-zertifiziert	Datensätze	Embargo möglich, Suche über Koordinaten möglich	CC0, CC-BY	uneingeschränkt
EDAL-PGP	<a href="https://edal-pgp.ipk-gatersleben.de/">https://edal-pgp.ipk-gatersleben.de/</a>	Pflanzenforschung	Infrastruktur zur Erhaltung von Daten aus der Pflanzenforschung	Gesamtheit aus der Phänotypisierung und Mikroskopie von Pflanzen, unfertige Datensätze, Genotypisierungsdaten, Visualisierungen von morphologischen Pflanzenmodellen, Daten aus der	Interner Begutachtungsprozess / umfangreiche Suche incl. Facettes/Filter	Verschiedene Lizenzen, u.a. CC-BY oder „Alle Rechte vorbehalten“.	uneingeschränkt
EnsemblPlants	<a href="https://plants.ensembl.org/index.html">https://plants.ensembl.org/index.html</a>	Genetik, Pflanzenwissenschaften	Der pflanzenwissenschaftliche Ensembl, dem europäischen NCBI				uneingeschränkt
NCBI	<a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a>	Genetik, Lebenswissenschaften	Sehr umfangreiche Genetik (RNA, DNA), Taxonomie und Tool				uneingeschränkt
Zenodo	<a href="https://zenodo.org/">https://zenodo.org/</a>	Alle	Großes Repo ohne Einschränkung zur Veröffentlichung aller Art von Forschungsdaten	Datensätze, Software, Präsentationen, Poster, Video/Audio, Online-Kurse			uneingeschränkt

Weitere Infos in unserer **FAIRagro Toolbox!**



FAIRagro Toolbox – Wo finde ich ein passendes Repository?, Tabelle: [Zenodo](#)

# Metadaten: Definition

Metadaten = „Daten über Daten“ → Abgrenzung zu Primärdaten

- **Standardisierte** Informationen
- **Strukturierte** Informationen
- Teil der **Dokumentation**:
  - beschreiben, erklären, verorten die Daten
  - erleichtern das Auffinden, Nutzen, Verwalten von Informationsressourcen
- **Menschen-** und **maschinenlesbar**



Created by Dewi Novita Sari  
from Noun Project

Quelle: Verändert nach Bres *et al.* 2022, Rfll and Voigt *et al.* 2022 durch ZB MED 2023 und FAIRagro 2024; „[definition](#)“ von Dewi Novita Sari ist lizenziert unter [CC BY 3.0](#)

# Metadaten: Beispiele

- **Bibliographische Metadaten:**
  - Titel
  - Beschreibung
  - Autor
  - Publikationsdatum
  - ...
- **Administrative Metadaten:**
  - Dateiformat
  - Dateiname
  - Lizenz
  - Zugriffsrechte
  - Eindeutige ID (z.B. DOI)
  - ...
- **Disziplinspezifische Metadaten:**
  - Nutzpflanze
  - Behandlung
    - Abiotisch (z.B. Licht, Temperatur, Boden, ...)
    - Biotisch (z.B. Tiere, Pilze, ...)
    - Bewirtschaftung (z.B. Anwendung von Chemikalien, Rotation, Dichte, ...)
  - Georeferenz
  - Sensor Informationen
    - Sensor Typ (z.B. RGB, MS, HS)
    - Wellenlänge
  - ...

Quelle: FAIRagro 2024 und verändert nach Engelhardt et al. 2022 durch ZB MED 2023 und FAIRagro 2024

# Metadatenstandards

- Einigung auf eine standardisierte Definition von Begriffen, die zu einem **gemeinsamen Verständnis** führt
- Förderung der **Integration** in der Community, die dem gewählten Standard folgt
- Daten können leichter **kombiniert** und **ausgetauscht** werden
- Verbesserung der Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit von Daten (FAIR Prinzipien)
  
- Beispiele:
  - [Dublin Core](#) (generisch, weit verbreitet)
  - [ISO 19115](#) (internationaler Standard für geographische Informationen)
  - [MIAPPE](#) (Standard für Pflanzenphänotypisierung)

Quelle: Verändert nach KU Leuven 2022 durch ZB MED 2023 und FAIRagro 2024; ZB MED 2024

# General Metadata Standard

**Dublin Core (DC)** = domain-agnostic, basic, widely used metadata standard.

- **Project level:** applicable to all datasets
- **Data level:** there are discipline-specific standards to branch into
- **DC Simple:** 15 metadata elements
- **DC Metadata Terms:** 15 metadata elements + several dozen properties and classes

nr.	Dublin Core element
1	Titel
2	Subject
3	Description
4	Type
5	Source
6	Relation
7	Coverage
8	Creator
9	Publisher
10	Contributor
11	Rights
12	Date
13	Format
14	Identifier
15	Language

# Persistente Identifier: Definition

Persistent Identifier (PIDs) = weltweit eindeutige, umsetzbare und maschinell auflösbare Zeichenketten aus (alphanumerischen) Zeichen, die:

- als dauerhafter Verweis auf ein digitales Objekt dienen
- zu einer zentralen Landing Page aufgelöst werden
- von vertrauenswürdigen Organisationen verwaltet werden

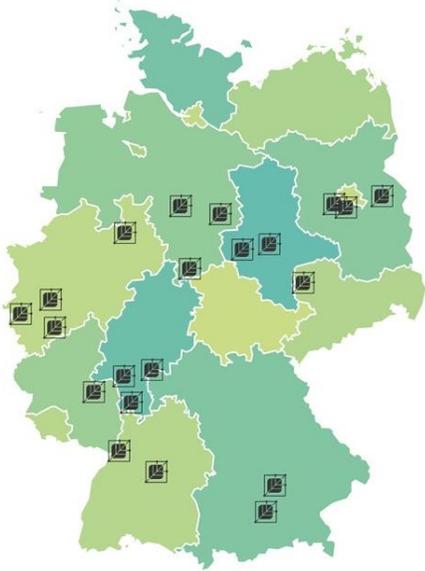
Beispiele:

- [DOI](#) – Digital Object Identifier
- [ORCID](#) – Persönliche ID in der Wissenschaft (Open Researcher and Contributor ID)
- [ROR](#) – ID für Wissenschaftsinstitutionen (Research Organisation Registry)



Created by Dewi Novita Sari  
from Noun Project

Quelle: ZB MED 2023; „[definition](#)“ von Dewi Novita Sari ist lizenziert unter [CC BY 3.0](#)



Vielen Dank für die Aufmerksamkeit!

... bei weiteren Fragen kontaktieren Sie gerne unseren FAIRagro Helpdesk!



[fairagro.net/helpdesk](https://fairagro.net/helpdesk)



# Rechtliche Stolpersteine im Forschungsdatenmanagement

Sie haben gefragt – wir antworten!

Lea Sophie Singson, FIZ Karlsruhe

**München, 21.10.2024**

# Was haben Rechtsthemen mit dem Forschungsdatenmanagement zu tun?



Quelle: [The Turing Way Community, & Scriberia. \(2023\).](#)



[ELIXIR \(2021\) Research Data Management Kit. A deliverable from the EU-funded ELIXIR-CONVERGE project \(grant agreement 871075\), licensed under CC BY 4.0](#)

# Fragen über Fragen...das beschäftigt Sie:

Es gibt verschiedene Definitionen "personenbezogener Information". Die Bandbreite reicht von der Angabe bloßer Koordinaten, die gemeinsam mit anderen Daten personenbezogen werden können, bis hin zu unmittelbar bezogenen Daten mit Namen, Adressen etc.. Wo ist die Grenze zwischen Daten, die als Open Data frei zugänglich gemacht werden können und personenbezogenen Daten, die dem Datenschutz unterliegen? Wer entscheidet, welche Daten in welcher z.B. aggregierter Form unter welcher Lizenz veröffentlicht werden dürfen?

Open Access versus Datenschutz:  
was muss man beim  
Veröffentlichen von Daten  
beachten?

Es wird zunehmend schwieriger Daten zur Biodiversität (Pflanzen-, Tierarten) auf den Flächen der Landwirte zu erheben. Der Datenschutz wird hier immer umfangreicher und damit die Verträge mit den Landwirten. Die Landwirte wollen diese komplexen Vereinbarungen nicht mehr unterschreiben. Auf der anderen Seite steht die Forderung nach open access und Zeitreihen zur Entwicklung der Biodiversität.

Zwiespalt Datenspeicherung im Rahmen der Projektlaufzeit vs. anschließende wissenschaftliche Publikation, die nicht unbedingt im Zusammenhang mit der Projektlaufzeit gegeben ist, sondern meist unabhängig davon ist. Nachvollziehbare Forschungsergebnisse bedingen die dauerhafte Archivierung von Erhebungsdaten/Umfragedaten. Könnte im Konflikt zur Projektlaufzeit stehen.

Wem "gehören" die Daten/Abbildungen die in Zeitschriften o.ä. veröffentlicht sind und wie sieht es mit einer weiteren Nutzung aus?

Der Datenschutz hat sich enorm verändert. Was ist bei "alten" Daten zu beachten?

Handhabung/Vorgehen bei eigener Erhebung von Daten durch Befragungen inkl. der Themen wie lange diese aufzubewahren sind (auch hinsichtlich nach wissenschaftlicher Publikation) vs. Vorgabe bei der Erhebung der Daten

Rechtliche Grundlagen für die Verwendung und Veröffentlichung von Daten.

Wer (Hochschule, Profs., WiMis...) hat unter welchen Bedingungen welche Rechte an Forschungsdaten?

# Fragenkategorien

Datenrechte

Rechtliche Grundlagen für die Verwendung und Veröffentlichung von Daten.

Wem "gehören" die Daten/Abbildungen die in Zeitschriften o.ä. veröffentlicht sind und wie sieht es mit einer weiteren Nutzung aus?

Wer (Hochschule, Profs., WiMis...) hat unter welchen Bedingungen welche Rechte an Forschungsdaten?

Es wird zunehmend schwieriger Daten zur Biodiversität (Pflanzen-, Tierarten) auf den Flächen der Landwirte zu erheben. Der Datenschutz wird hier immer umfangreicher und damit die Verträge mit den Landwirten. Die Landwirte wollen diese komplexen Vereinbarungen nicht mehr unterschreiben. Auf der anderen Seite steht die Forderung nach open access und Zeitreihen zur Entwicklung der Biodiversität.

Handhabung/Vorgehen bei eigener Erhebung von Daten durch Befragungen inkl. der Themen wie lange diese aufzubewahren sind (auch hinsichtlich nach wissenschaftlicher Publikation) vs. Vorgabe bei der Erhebung der Daten

Es gibt verschiedene Definitionen "personenbezogener Information". Die Bandbreite reicht von der Angabe bloßer Koordinaten, die gemeinsam mit anderen Daten personenbezogen werden können, bis hin zu unmittelbar bezogenen Daten mit Namen, Adressen etc.. Wo ist die Grenze zwischen Daten, die als Open Data frei zugänglich gemacht werden können und personenbezogenen Daten, die dem Datenschutz unterliegen? Wer entscheidet, welche Daten in welcher z.B. aggregierter Form unter welcher Lizenz veröffentlicht werden dürfen?

Zwiespalt Datenspeicherung im Rahmen der Projektlaufzeit vs. anschließende wissenschaftliche Publikation, die nicht unbedingt im Zusammenhang mit der Projektlaufzeit gegeben ist, sondern meist unabhängig davon ist. Nachvollziehbare Forschungsergebnisse bedingen die dauerhafte Archivierung von Erhebungsdaten/Umfragedaten. Könnte im Konflikt zur Projektlaufzeit stehen.

Open Access versus Datenschutz: was muss man beim Veröffentlichen von Daten beachten?

Der Datenschutz hat sich enorm verändert. Was ist bei "alten" Daten zu beachten?

Datenschutz

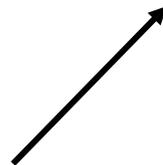
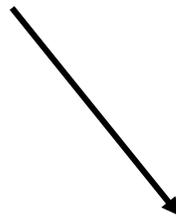
# Fragenkategorien

Datenrechte

Rechtliche Grundlagen für die Verwendung und Veröffentlichung von Daten.

Wem "gehören" die Daten/Abbildungen die in Zeitschriften o.ä. veröffentlicht sind und wie sieht es mit einer weiteren Nutzung aus?

Wer (Hochschule, Profs., WiMis...) hat unter welchen Bedingungen welche Rechte an Forschungsdaten?



**Session 2 „To license or not to license“**  
**Q&A rund ums Thema Datenrechte**

Es wird zunehmend schwieriger Daten zur Biodiversität (Pflanzen-, Tierarten) auf den Flächen der Landwirte zu erheben. Der Datenschutz wird hier immer umfangreicher und damit die Verträge mit den Landwirten. Die Landwirte wollen diese komplexen Vereinbarungen nicht mehr unterschreiben. Auf der anderen Seite steht die Forderung nach open access und Zeitreihen zur Entwicklung der Biodiversität.

Handhabung/Vorgehen bei eigener Erhebung von Daten durch Befragungen inkl. der Themen wie lange diese aufzubewahren sind (auch hinsichtlich nach wissenschaftlicher Publikation) vs. Vorgabe bei der Erhebung der Daten

Es gibt verschiedene Definitionen "personenbezogener Information". Die Bandbreite reicht von der Angabe bloßer Koordinaten, die gemeinsam mit anderen Daten personenbezogen werden können, bis hin zu unmittelbar bezogenen Daten mit Namen, Adressen etc.. Wo ist die Grenze zwischen Daten, die als Open Data frei zugänglich gemacht werden können und personenbezogenen Daten, die dem Datenschutz unterliegen? Wer entscheidet, welche Daten in welcher z.B. aggregierter Form unter welcher Lizenz veröffentlicht werden dürfen?

Zwiespalt Datenspeicherung im Rahmen der Projektlaufzeit vs. anschließende wissenschaftliche Publikation, die nicht unbedingt im Zusammenhang mit der Projektlaufzeit gegeben ist, sondern meist unabhängig davon ist. Nachvollziehbare Forschungsergebnisse bedingen die dauerhafte Archivierung von Erhebungsdaten/Umfragedaten. Könnte im Konflikt zur Projektlaufzeit stehen.

Open Access versus Datenschutz: was muss man beim Veröffentlichen von Daten beachten?

Der Datenschutz hat sich enorm verändert. Was ist bei "alten" Daten zu beachten?

Datenschutz

# Fragenkategorien

Datenrechte

Rechtliche Grundlagen für die Verwendung und Veröffentlichung von Daten.

Wem "gehören" die Daten/Abbildungen die in Zeitschriften o.ä. veröffentlicht sind und wie sieht es mit einer weiteren Nutzung aus?

Wer (Hochschule, Profs., WiMis...) hat unter welchen Bedingungen welche Rechte an Forschungsdaten?

**Session 1:  
(Don't) take it personally:  
Datenschutz am Beispiel DigiMilch**

Es wird zunehmend schwieriger Daten zur Biodiversität (Pflanzen-, Tierarten) auf den Flächen der Landwirte zu erheben. Der Datenschutz wird hier immer umfangreicher und damit die Verträge mit den Landwirten. Die Landwirte wollen diese komplexen Vereinbarungen nicht mehr unterschreiben. Auf der anderen Seite steht die Forderung nach open access und Zeitreihen zur Entwicklung der Biodiversität.

Handhabung/Vorgehen bei eigener Erhebung von Daten durch Befragungen inkl. der Themen wie lange diese aufzubewahren sind (auch hinsichtlich nach wissenschaftlicher Publikation) vs. Vorgabe bei der Erhebung der Daten

Es gibt verschiedene Definitionen "personenbezogener Information". Die Bandbreite reicht von der Angabe bloßer Koordinaten, die gemeinsam mit anderen Daten personenbezogen werden können, bis hin zu unmittelbar bezogenen Daten mit Namen, Adressen etc.. Wo ist die Grenze zwischen Daten, die als Open Data frei zugänglich gemacht werden können und personenbezogenen Daten, die dem Datenschutz unterliegen? Wer entscheidet, welche Daten in welcher z.B. aggregierter Form unter welcher Lizenz veröffentlicht werden dürfen?

Zwiespalt Datenspeicherung im Rahmen der Projektlaufzeit vs. anschließende wissenschaftliche Publikation, die nicht unbedingt im Zusammenhang mit der Projektlaufzeit gegeben ist, sondern meist unabhängig davon ist. Nachvollziehbare Forschungsergebnisse bedingen die dauerhafte Archivierung von Erhebungsdaten/Umfragedaten. Könnte im Konflikt zur Projektlaufzeit stehen.

Open Access versus Datenschutz: was muss man beim Veröffentlichen von Daten beachten?

Der Datenschutz hat sich enorm verändert. Was ist bei "alten" Daten zu beachten?

Datenschutz

# Fragenkategorien

Datenrechte

Rechtliche Grundlagen für die Verwendung und Veröffentlichung von Daten.

Wem "gehören" die Daten/Abbildungen die in Zeitschriften o.ä. veröffentlicht sind und wie sieht es mit einer weiteren Nutzung aus?

Wer (Hochschule, Profs., WiMis...) hat unter welchen Bedingungen welche Rechte an Forschungsdaten?

Es wird zunehmend schwieriger Daten zur Biodiversität (Pflanzen-, Tierarten) auf den Flächen der Landwirte zu erheben. Der Datenschutz wird hier immer umfangreicher und damit die Verträge mit den Landwirten. Die Landwirte wollen diese komplexen Vereinbarungen nicht mehr unterschreiben. Auf der anderen Seite steht die Forderung nach open access und Zeitreihen zur Entwicklung der Biodiversität.

Handhabung/Vorgehen bei eigener Erhebung von Daten durch Befragungen inkl. der Themen wie lange diese aufzubewahren sind (auch hinsichtlich nach wissenschaftlicher Publikation) vs. Vorgabe bei der Erhebung der Daten

Es gibt verschiedene Definitionen "personenbezogener Information". Die Bandbreite reicht von der Angabe bloßer Koordinaten, die gemeinsam mit anderen Daten personenbezogen werden können, bis hin zu unmittelbar bezogenen Daten mit Namen, Adressen etc.. Wo ist die Grenze zwischen Daten, die als Open Data frei zugänglich gemacht werden können und personenbezogenen Daten, die dem Datenschutz unterliegen? Wer entscheidet, welche Daten in welcher z.B. aggregierter Form unter welcher Lizenz veröffentlicht werden dürfen?

Zwiespalt Datenspeicherung im Rahmen der Projektlaufzeit vs. anschließende wissenschaftliche Publikation, die nicht unbedingt im Zusammenhang mit der Projektlaufzeit gegeben ist, sondern meist unabhängig davon ist. Nachvollziehbare Forschungsergebnisse bedingen die dauerhafte Archivierung von Erhebungsdaten/Umfragedaten. Könnte im Konflikt zur Projektlaufzeit stehen.

Open Access versus Datenschutz: was muss man beim Veröffentlichen von Daten beachten?

Der Datenschutz hat sich enorm verändert. Was ist bei "alten" Daten zu beachten?

Datenschutz

## Worum geht's heute:

**Der Datenschutz hat sich enorm verändert. Was ist bei "alten" Daten zu beachten?**

**Es gibt verschiedene Definitionen "personenbezogener Information". Die Bandbreite reicht von der Angabe bloßer Koordinaten, die gemeinsam mit anderen Daten personenbezogen werden können, bis hin zu unmittelbar bezogenen Daten mit Namen, Adressen etc.. Wo ist die Grenze zwischen Daten, die als Open Data frei zugänglich gemacht werden können und personenbezogenen Daten, die dem Datenschutz unterliegen? Wer entscheidet, welche Daten in welcher z.B. aggregierter Form unter welcher Lizenz veröffentlicht werden dürfen?**

# Die DSGVO: Wann sie gilt – und wann nicht



Mariia Shalabaieva for [Unsplash](#)

# Die DSGVO: Geltungsbereich

## Sachlicher Anwendungsbereich

- Wann ist die DSGVO zu beachten?

## Persönlicher Anwendungsbereich

- Wer muss die DSGVO beachten?

## Räumlicher Anwendungsbereich

- Wo muss die DSGVO beachtet werden?

## Zeitlicher Anwendungsbereich

- Wann muss die DSGVO beachtet werden?

# Die DSGVO: Geltungsbereich

## Sachlicher Anwendungsbereich

- Wann ist die DSGVO zu beachten?

## Persönlicher Anwendungsbereich

- Wer muss die DSGVO beachten?

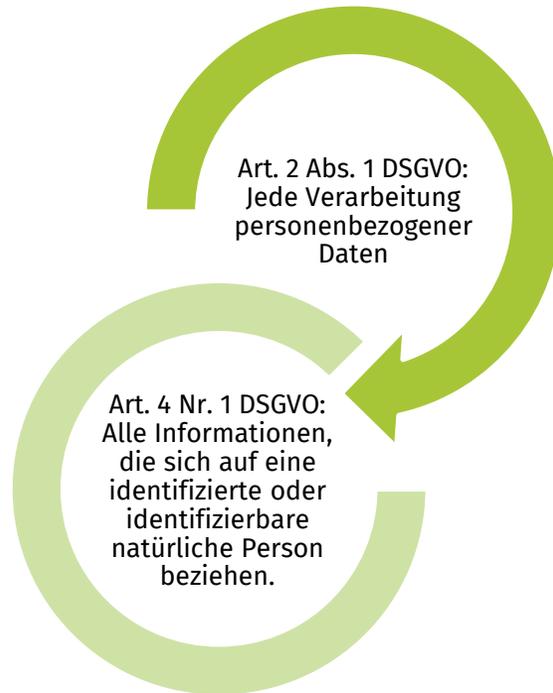
## Räumlicher Anwendungsbereich

- Wo muss die DSGVO beachtet werden?

## Zeitlicher Anwendungsbereich

- Wann muss die DSGVO beachtet werden?

# Sachlicher Anwendungsbereich



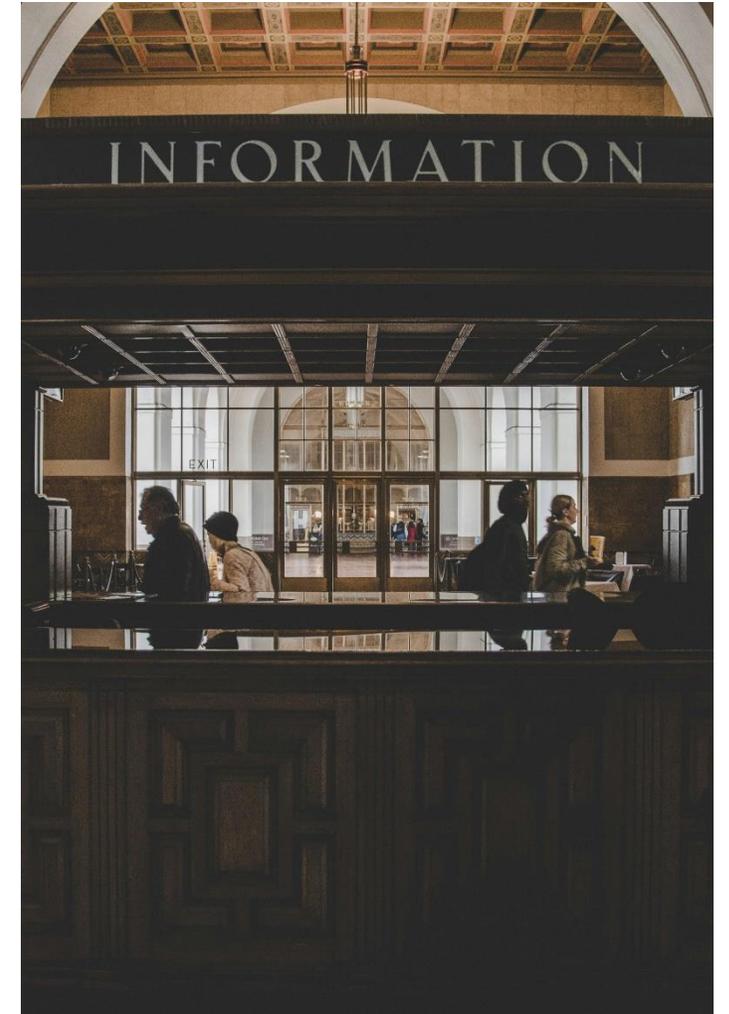
➔ DSGVO ist nicht  
anzuwenden bei:  
anonymen Daten, Daten  
verstorbener Personen

↓

Gibt es mehrere  
Definitionen von  
personenbezogenen Daten?

↓

Verwirrung häufig bei der Variante der  
**identifizierbaren** Information



Philip Strong for [Unsplash](#)

# Sachlicher Anwendungsbereich

Wann sind Personen durch Informationen identifizierbar:

Informationen an sich führen nicht zur Identifikation, aber in Verbindung mit anderen Informationen, z.B. durch **Zuordnung zu einer Kennung** wie Name, Standortdaten oder zu anderen besonderen Merkmalen, die Ausdruck der Identität der Person sind, ist eine Identifikation möglich.

Wie beurteilt man, ob Informationen zur Identifikation führen könnten?

- Es werden alle Mittel berücksichtigt, die vom Verantwortlichen oder anderen Personen nach allgemeinem Ermessen **wahrscheinlich** genutzt werden, um Personen direkt oder indirekt zu identifizieren.
- Faktoren dabei sind: Kosten der Identifizierung und dafür erforderlicher Zeitaufwand und deren Verhältnis zum Nutzen einer potenziellen Identifikation -> Aufwand vs. Informationswert
- Es wird für jeden Verantwortlichen am Maßstab seines mobilisierbaren Zusatzwissens gemessen -> **welche technischen Möglichkeiten sind während der vorgesehenen Speicherdauer zu erwarten**

Auch subjektive Informationen, wie Werturteile, Einschätzungen und Meinungsäußerungen

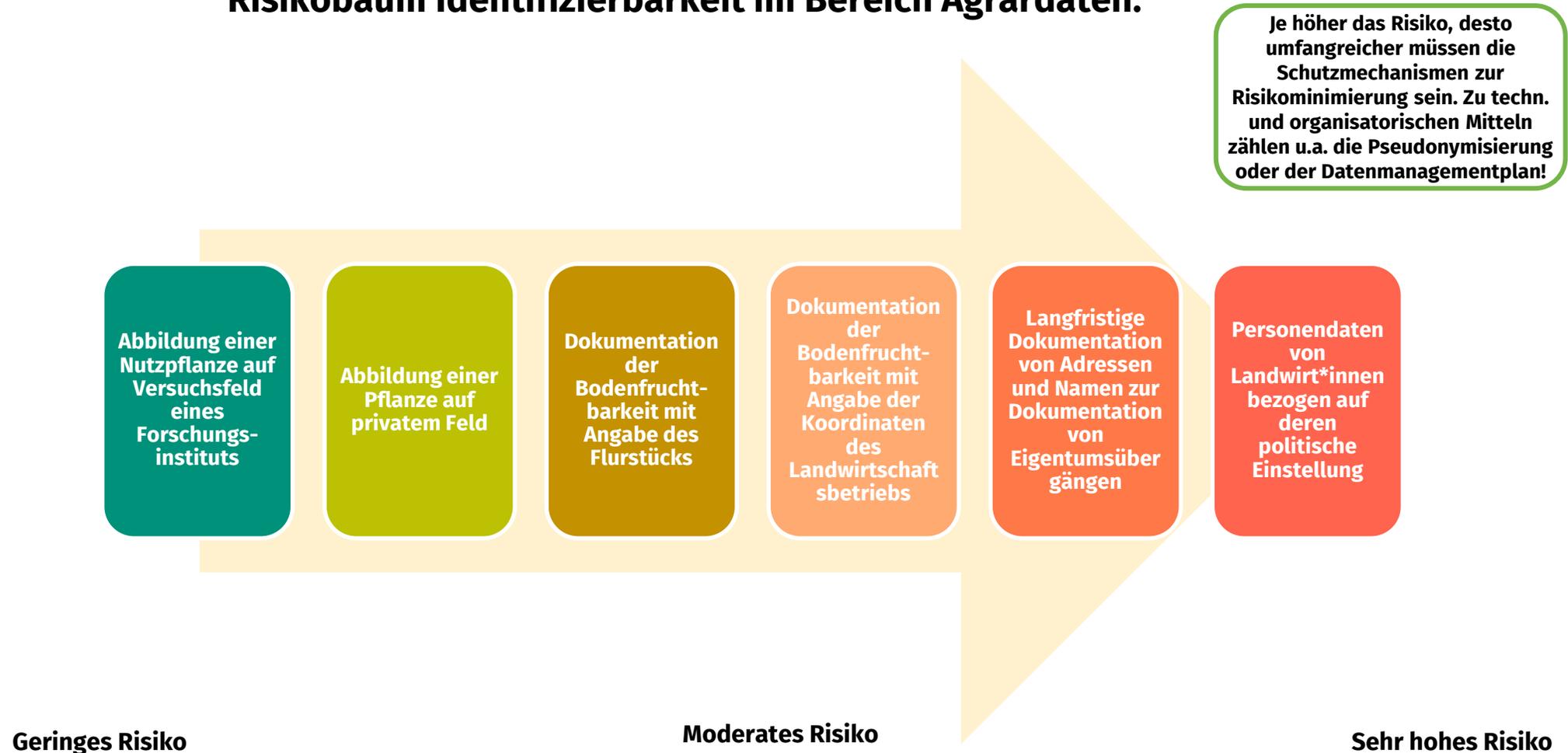


ELIXIR (2021) Research Data Management Kit\_A deliverable from the EU-funded ELIXIR-CONVERGE project (grant agreement 871075), licensed under CC BY 4.0

**Wer beurteilt das?**  
**Final Gerichte; aber: Datenschutzentscheidung stets Risikoabschätzung -> Das DSSC kann helfen, diese Risikoabschätzung zu treffen!**

# Identifizierbarkeit von Daten

## Risikobaum Identifizierbarkeit im Bereich Agrardaten:



## Zeitlicher Anwendungsbereich:

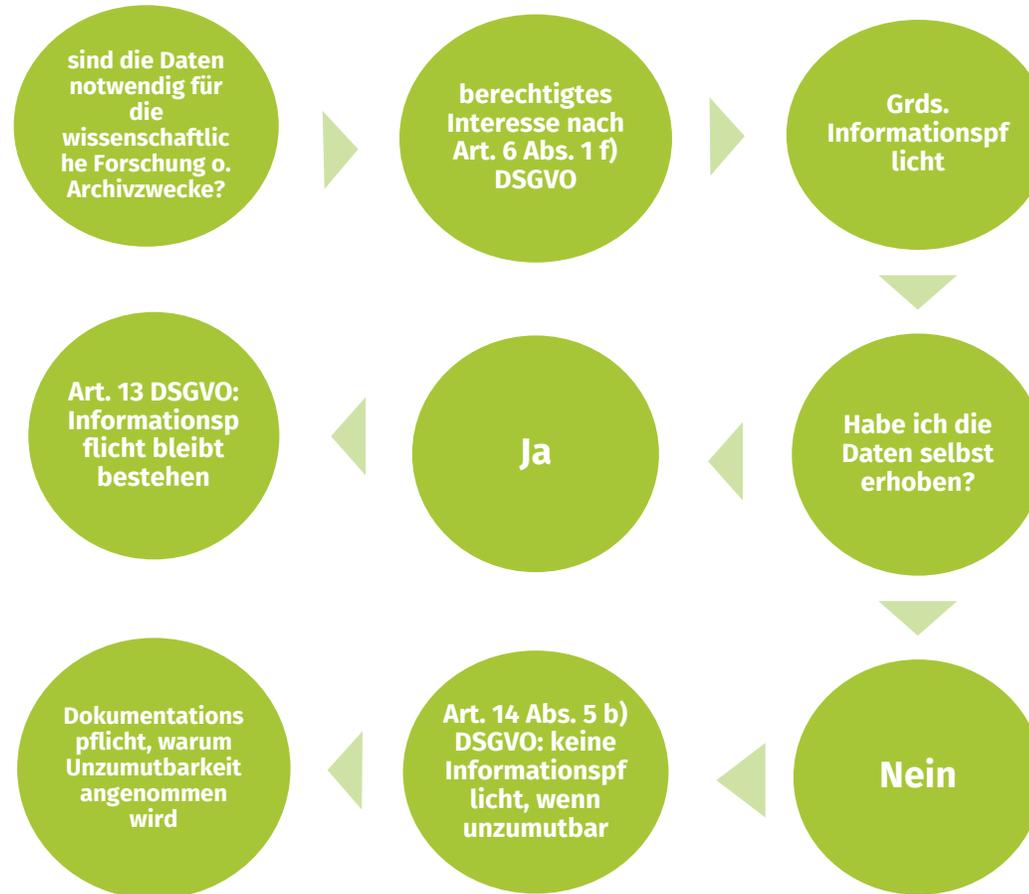
Grundsatz: DSGVO gilt für alle Datenverarbeitungen ohne Unterscheidungen nach Alter

Szenario 1:  
Einwilligung  
existiert nach §  
4a BDSG

bleibt vollumfänglich wirksam, weil:  
Hauptsache, es liegt eine Einwilligung vor,  
die die Anforderungen der DSGVO erfüllt  
= freiwillig, informiert und  
unmissverständlich

# Zeitlicher Anwendungsbereich:

Szenario 2: Es existiert keine Einwilligung





FAIRagro

Funded by  
**DFG**  
Deutsche  
Forschungsgemeinschaft  
German Research Foundation

project number 501899475

In cooperation with



Leibniz-Zentrum für Agrarlandschaftsforschung ZALF

**Marcus Schmidt**

[marcus.schmidt@zalf.de](mailto:marcus.schmidt@zalf.de)



im Namen des FAIRagro Konsortiums

## Wo bekomme ich Unterstützung? Die FDM-Helpdesks und Services von FAIRagro, NFDI4Biodiversity und Co.



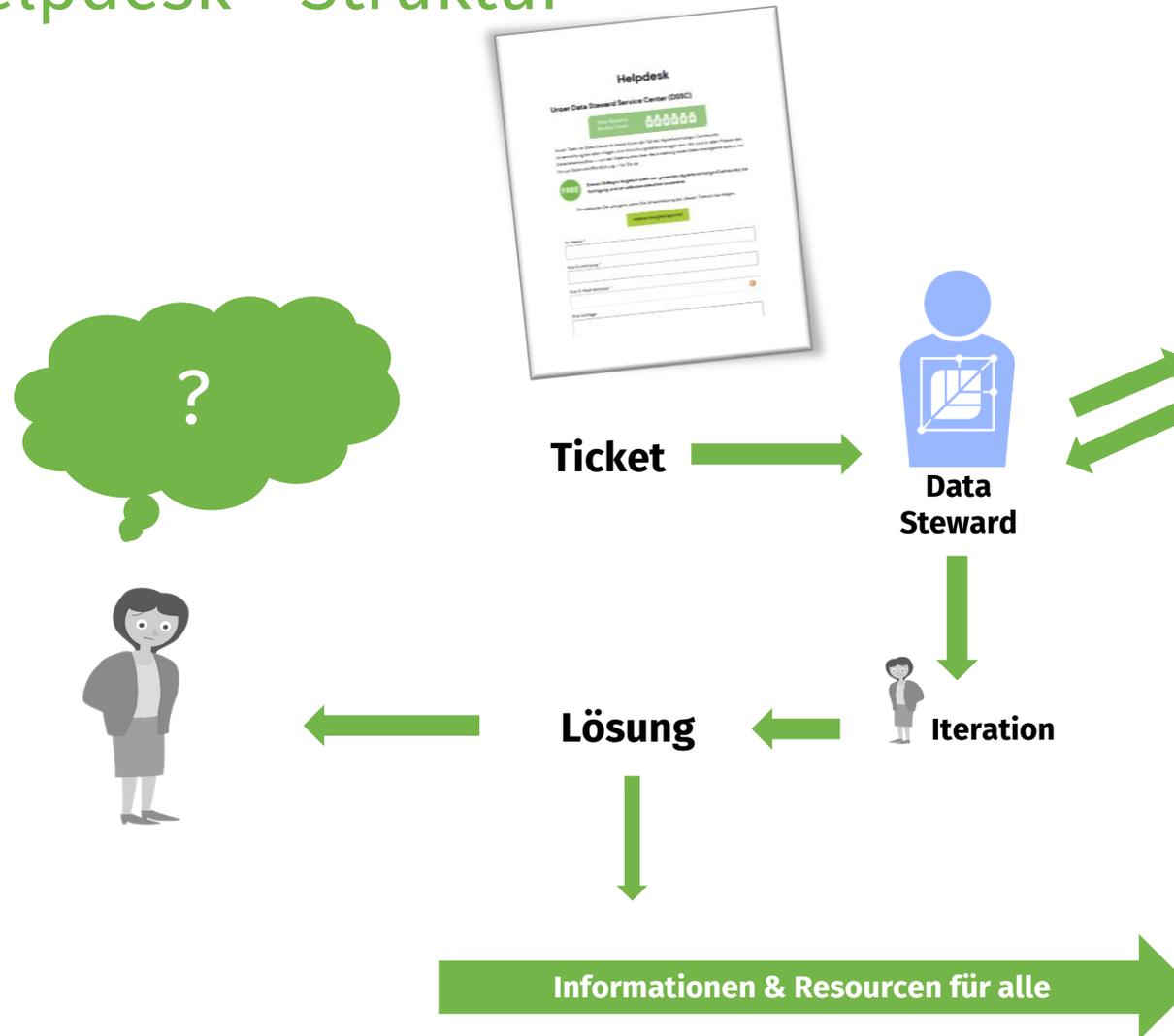
Agrosystem Community



Kostenfrei und offen für Alle  
in der Agrarforschung!



# FAIRragro Helpdesk - Struktur



**NFDI & andere Helpdesks  
FAIRragro ExpertInnen und Entwickler**



# FAIRagro Helpdesk – ExpertInnen und Themen



Elena

**Plant Breeding**  
Data management  
plan (DMP)  
MIAPPE  
...



Lucia

**Farming 4.0**  
Field robotics  
Visualization  
...



Florian

**Big Geodata**  
Curation and metadata  
Web services and SDI  
Quality and provenance  
...

**Coordination**  
NFDI networking  
Data Science  
...



Marcus



**Data Publication**  
Long term  
experiments  
DOI  
Metadata



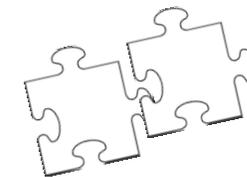
Wahib



**Law & Ethics**  
Intellectual  
property  
Copyright and  
licensing  
Data privacy  
...



Lea



Sophie

Link to our  
FAIRagro Training  
Measure

## Nehmen Sie gern Kontakt auf wenn Sie:

- **frei verfügbare Daten** zu Boden, Feldexperimenten, Phänologie o.ä. suchen.
- Ihre **Daten veröffentlichen** möchten, zum Beispiel für ein Paper oder eine Dissertation und auf der Suche nach einem **passendes Repository** sind.
- ein Projekt planen und Beratung zum Aufsetzen einer **Data Policy** benötigen.
- nach passenden **Tools für die Auswertung** bestimmter Daten suchen.
- sich ein **maßgeschneidertes Training** oder didaktischen Support zum Forschungsdatenmanagement für ihre Arbeitsgruppe wünschen.
- ... oder **sonstige Fragen** zum Forschungsdatenmanagement haben.



**Wir freuen uns von Ihnen zu hören!**

[dataservice@fairagro.net](mailto:dataservice@fairagro.net)



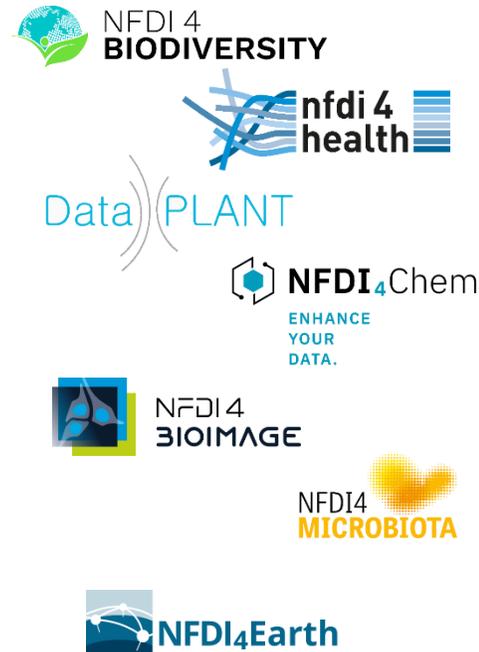
# Und wenn wir es mal nicht wissen? – Unser Netzwerk

## Wissensaustausch und internationale Kooperation



Agroforst-Projekt Ackerbaum

## Fachnahe NFDI Helpdesks



## Institutionelle Netzwerke & Landesinitiativen

- Universitäten und Hochschulen

UNIKASSEL  
VERSITÄT



- Landesinitiativen



- Forschungseinrichtungen



# Und wenn wir es mal nicht wissen? – Unser Netzwerk



Konsortium	Domäne / Disziplinen
<b>NFDI4Earth</b>	Erdsystemwissenschaften
<b>NFDI4Bioimage</b>	Bilddaten aus Biowissenschaften/Biologie
<b>NFDI4Health</b>	Medizin, Öffentliche Gesundheit, Epidemiologie
<b>NFDI4Microbiota</b>	Mikrobiologie
<b>DataPLANT</b>	Pflanzenwissenschaften
<b>FAIRagro</b>	Agrarsystemforschung
<b>NFDI4Biodiversity</b>	Biodiversitätswissenschaft: Umwelt, Ökologie, Biologie





## So erreichen Sie uns:

FAIRagro Helpdesk

[dataservice@fairagro.net](mailto:dataservice@fairagro.net)

[training@fairagro.net](mailto:training@fairagro.net)

<https://www.fairagro.net/helpdesk>

Oder gern auch: [marcus.schmidt@zalf.de](mailto:marcus.schmidt@zalf.de)



# ~~Don't~~ take it personally

## Datenschutz am Beispiel DigiMilch

[Sophie Boße](#) und [Lea Sophie Singson](#) im Namen des FAIRagro Konsortiums

# Code of Conduct

Für ein konstruktives Zusammenarbeiten möchten wir diese Leitlinien befolgen:

- Verwenden Sie eine einladende und integrative Sprache
- Respektieren Sie unterschiedliche Standpunkte und Erfahrungen
- Nehmen Sie konstruktive Kritik an
- Konzentrieren Sie sich darauf, was das Beste für die Community ist
- Zeigen Sie Höflichkeit und Respekt gegenüber anderen Mitgliedern der Community

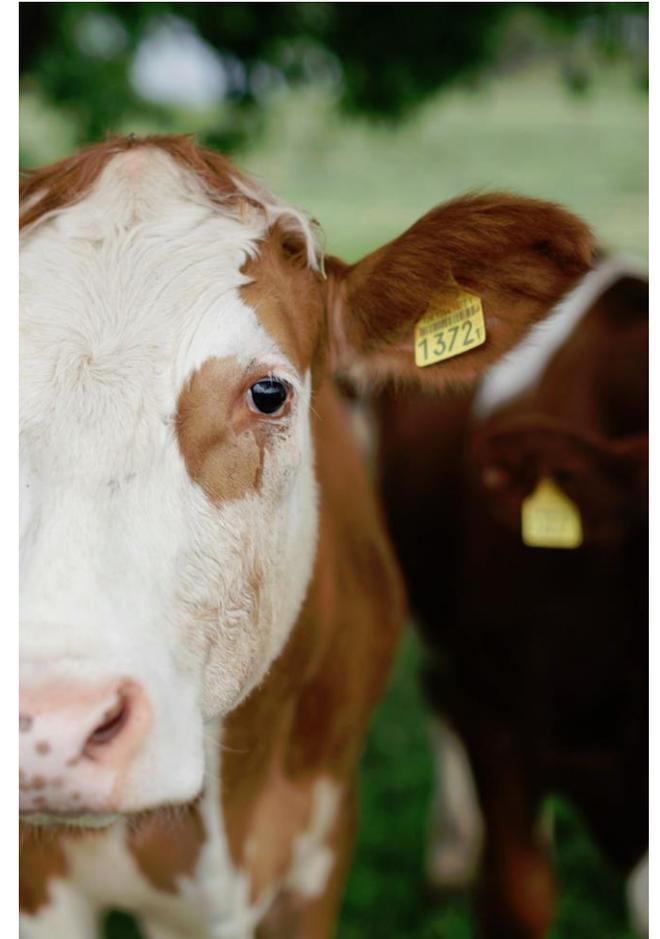
# Das Projekt DigiMilch

Experimentierfeld "Digitalisierung in der Prozesskette Milch"

**Ziel:** Überprüfung der Eignung digitaler Lösungen für die Prozesskette Milcherzeugung in familiengeführten Praxisbetrieben

**Methoden:** 5 Demonstrationsprojekte werden in 5-12 Betrieben, sowie 2 LfL-Versuchsstationen erprobt, bewertet und demonstriert

- Wirtschaftsdüngermanagement
- sensorgestützte Ertragsermittlung
- Fütterungsmanagement
- vernetzte Stalltechnik
- vernetzte, tierindividuelle Sensorsysteme



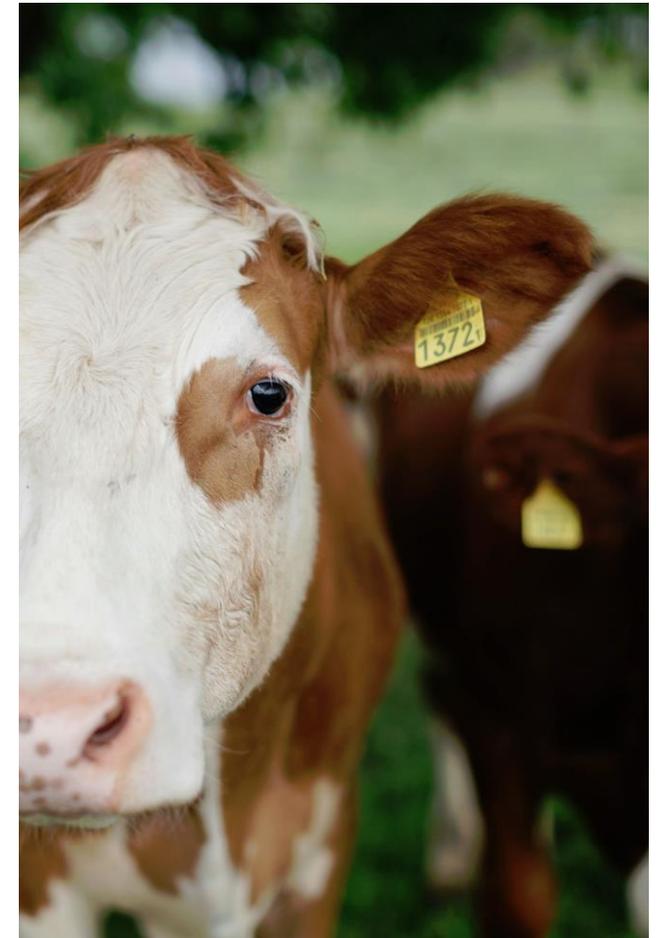
# Das Projekt DigiMilch

## Daten:

- Betriebsdaten (zu Betrieb, einzelnen Tieren, Erträge, Futter)
- Aus Software exportiert und manuell erhoben
- In Datenbank oder in SharePoint gespeichert

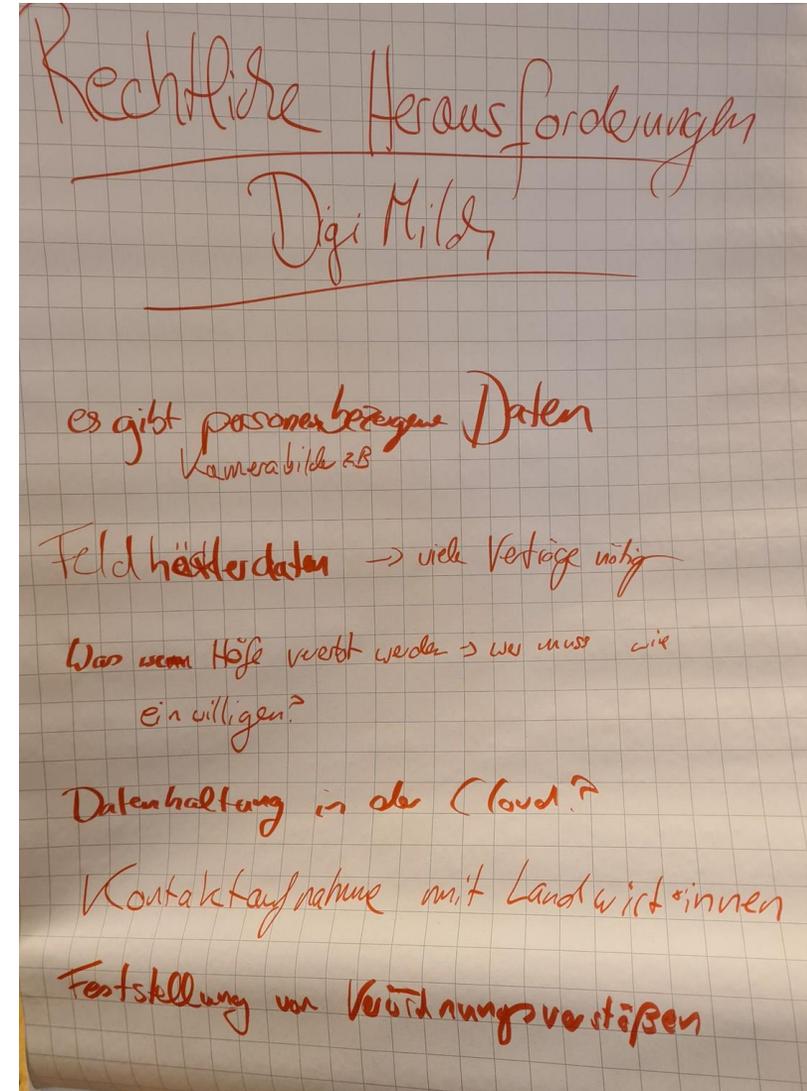
## Betriebe:

- Familienbetriebe



# Brainstorming

Welche rechtlichen  
Probleme/Herausforderungen  
fallen euch im Projekt  
DigiMilch auf?



# Datenschutz - Fragen aus dem Projekt

1. Sind die erhobenen Daten datenschutzrechtlich relevant?
2. Gibt es “Teil-anonymisierte” Daten?
3. Wie wirkt sich die Methode der Erhebung datenschutzrechtlich aus?
4. Was passiert mit den Daten nach Ende der Projektlaufzeit?
5. Was wenn Landwirt\*innen die Verträge nicht unterschreiben wollen?



Created by M Yudi Maulana  
from Noun Project

Sind die erhobenen Daten  
datenschutzrechtlich relevant?

# Datenschutzrechtliche Relevanz der Daten

## **Erinnerung Definition personenbezogene Daten:**

Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen!

→ Daten zum Betrieb selbst: möglich



Created by Cindy Van Heerden  
from Noun Project

# Datenschutz

Forschende müssen sich folgende Fragen stellen:

- Welche Daten verarbeite ich?
  - Was soll mit den Daten geschehen?
  - Zu welchem Zweck verarbeite ich die Daten?
  - Bin ich die einzige, die mit den Daten in Berührung kommt?
  - Findet im Verarbeitungsprozess eine Pseudonymisierung oder eine Anonymisierung statt?
- Bei Interviewdaten: Name der Landwirt\*innen
  - Hinsichtlich der Felder: Standortdaten, wie PLZ, Flurstück, Landkreis
  - Forschende\*r muss die Daten ggf. abfragen, aufschreiben, speichern, mit Dritten teilen?
  - Zwecke könnten sein: Nachvollziehbarkeit ihrer Thesen aus dem Experiment, Organisation, Nachverfolgbarkeit bei evtl. Rückfragen

# Datenschutz

Wie sieht es hier aus:

Landwirt\*innen



Created by Justin Blake  
from Noun Project

→ wann ist eine Person identifizierbar?

wenn sie direkt oder indirekt mittels einer Kennung, wie z.B. ihrem Namen zu einer Kennnummer, **Standortdaten** oder anderen Merkmalen, die zur Identifikation der Person führen können, zugeordnet werden kann

→ da DigiMilch aus Deutschland (=EU-Mitgliedstaat) kommt, muss die DSGVO im Hinblick auf die Person der Landwirt\*innen aller Standorte beachten, wenn deren Namen erhoben werden

# Standortdaten - Datenschutz relevant?

**Standortdaten** haben dann **Personenbezug**, wenn durch sie **natürliche Personen** identifiziert werden können.

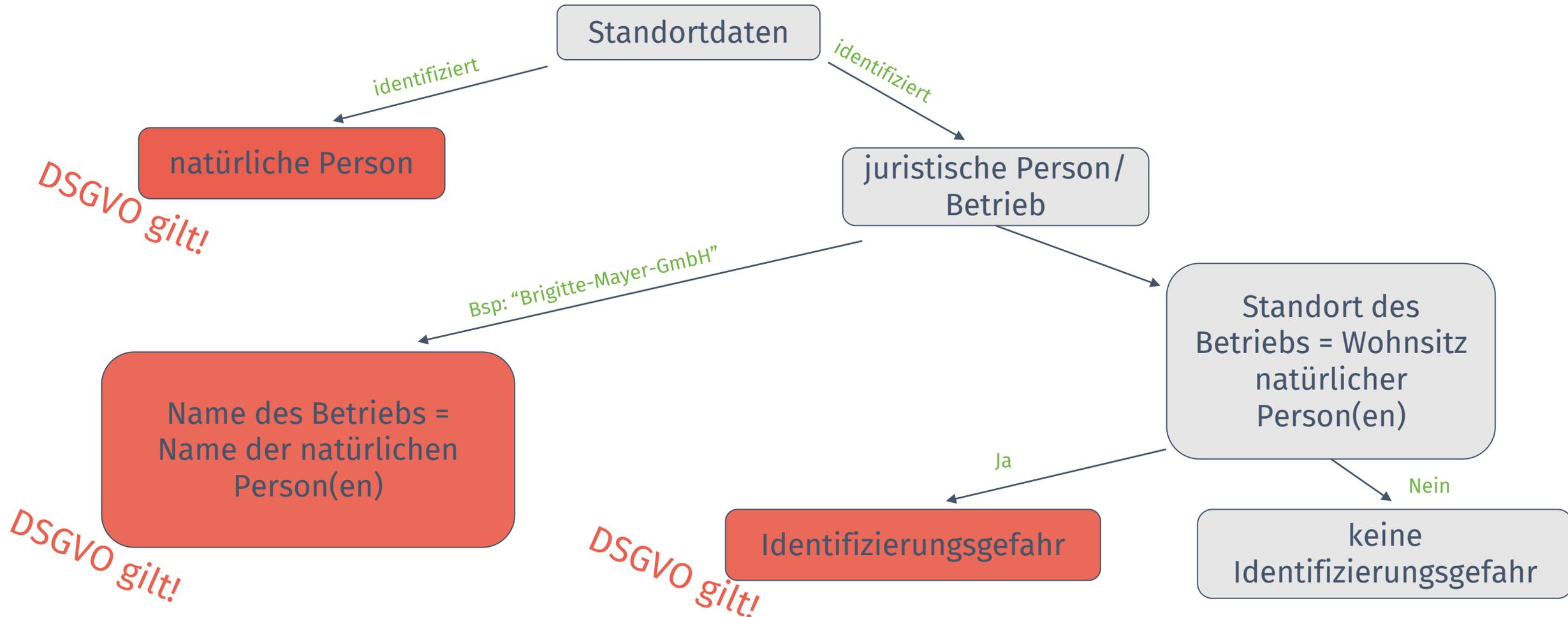
Betrieblichen Standortdaten: häufig steht hinter landwirtschaftlichen Feldern keine einzelne Person, sondern ein Betrieb. -> Wann besteht da Personenbezug?



Created by M Yudi Maulana  
from Noun Project

- DSGVO schützt nur natürliche Personen, **keine juristischen Personen** oder sonstige **Betriebe**
- Ausnahme 1: wenn Name der jur. Person eine oder mehrere natürliche Personen bestimmt (EuGH C-92, 93/09), dann Grundrechte auf Privatsphäre und Datenschutz aus Art. 7 und 8 GRCh., also Schutzanspruch nach DSGVO
  - ◆ z.B. **“Brigitte-Mayer-GmbH”**
- Ausnahme 2: Befinden sich auf dem Standort neben dem betriebl. Feld auch Wohnsitze von natürlichen Personen und besteht Identifizierungsgefahr?

# Standortdaten - Datenschutz relevant?



# Gibt es “Teil-anonymisierte” Daten?

Und wie funktioniert das mit der Anonymisierung überhaupt?

# Was ist Anonymisierung - und was nicht?

## Pseudonymisierung

- Direkte o. indirekte Identifizierungsmerkmale werden getrennt, gesondert aufbewahrt und durch Kennzeichen ersetzt
- Re-identifikation möglich



DSGVO anwendbar, Einwilligung notwendig!



## Anonymisierung

Veränderung von Daten, Einzelangaben über persönliche oder sachliche Verhältnisse **nicht mehr/nur mit unverhältnismäßigem Aufwand** an Zeit, Kosten und Arbeitskraft zuordnungsbar

DSGVO nicht anzuwenden, keine Einwilligung notwendig!



# Anforderungen der DSGVO an die Anonymisierung

Hinweis: der Anonymisierungsvorgang ist eine Verarbeitung personenbezogener Daten, die eine Ermächtigungsgrundlage nach Art. 6, 9 DSGVO benötigt

**Absolute Anonymisierung:** Personenbezug praktisch für jedermann unmöglich

**Faktische Anonymisierung:** Re-Identifizierung scheidet aufgrund Unverhältnismäßigkeit des Aufwands aus

-> Die Grenze der Identifizierbarkeit liegt in der Unmöglichkeit, der Unverhältnismäßigkeit oder im Rechtsverstoß

-> Hinsichtlich techn. Re-Identifizierungsmöglichkeit: es kommt auf die im Zeitpunkt der Verarbeitung verfügbaren Technologien und technologischen Entwicklung im Zeitraum der Speicherdauer an.

# Umgang mit personenbezogenen Daten

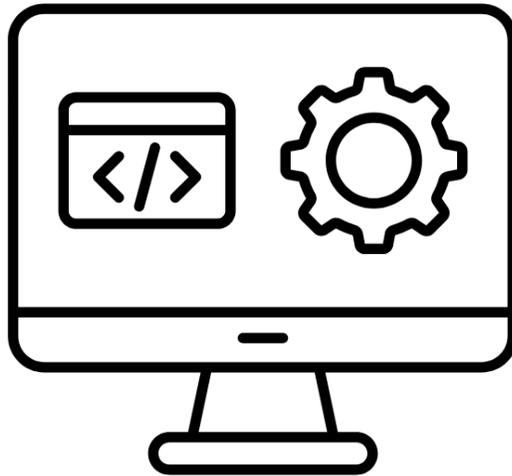
Für DigiMilch machen die Vorteile einer Anonymisierung in diesem Stadium nur Sinn, wenn die Daten für sie bereits anonym sind.

- bei Standortdaten: nicht möglich
- Interviewdaten: anonyme online- Umfrage: sinnvoll für Experiment?

# Wie wirkt sich die Methode der Erhebung datenschutzrechtlich aus?

Erhebungsmedium: kommerzielle Software

# Erhebungsmedium



Created by Articon  
from Noun Project

Wenn kommerzielle Software zur Erhebung genutzt wird, muss geklärt werden:

## Ist die Software mit der DSGVO vereinbar?

- “Red-Flags”  :
  - Werden Daten in Drittländer außerhalb der EU transferiert?
  - Werden erhobene Daten zum Training der Software verwendet?
  - Bestehen Datenschutzbeschwerden gegen die Software?
- Gibt es einen Leitfaden der leitenden Einrichtung über die Verwendung?

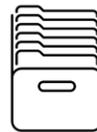
# Was passiert mit den Daten nach Ende der Projektlaufzeit?

Wie lange muss ich archivieren?

# Archivierung

## Archivierung nach DSGVO:

- Grds. Aufbewahrung so lange, wie für Verarbeitungszwecke notwendig.
- Dabei Beschränkung auf unbedingt erforderliches Mindestmaß
- obj. Verpflichtung des Verantwortlichen, fortlaufend zu überprüfen, ob Speicherung noch notwendig.
- bei Forschungsdaten keine feste Maximaldauer



Created by Kusuma Potter from Noun Project



## Archivierung nach Vorgaben der Fördermittelgeber/ GWP:

- Leitlinie 17 Erl.:  
Aufbewahrung i.d.R. 10 Jahre in Einrichtung oder in Repo
- ggf. Konkretisierung in Konsortialverträgen



Created by Kusuma Potter from Noun Project

Was wenn Landwirt\*innen die  
Verträge nicht unterschreiben wollen?

Eure Erfahrungen

# To license or not to license

## Q&A rund um Rechte an Daten

[Sophie Boße](#) und [Lea Sophie Singson](#) im Namen des FAIRagro Konsortiums

# Datenrechte - Ihre Fragen

1. Wem "gehören" die Daten/Abbildungen die in Zeitschriften o.ä. veröffentlicht sind und wie sieht es mit einer weiteren Nutzung aus?
2. Open Access versus Datenschutz: was muss man beim Veröffentlichen von Daten beachten?
3. Wer (Hochschule, Profs., WiMis...) hat unter welchen Bedingungen welche Rechte an Forschungsdaten?



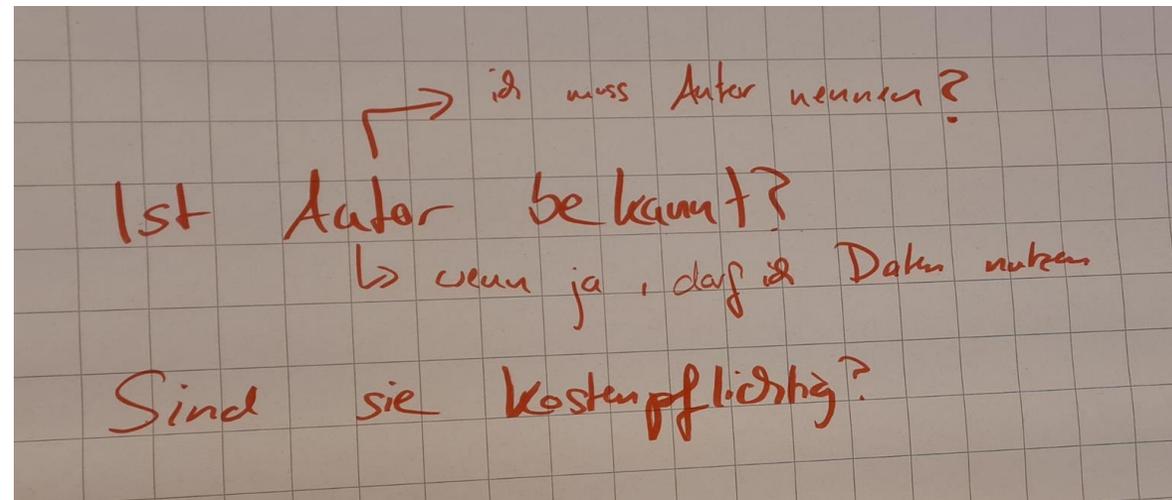
Created by M Yudi Maulana  
from Noun Project

# Datennachnutzung

“Wie sieht es mit einer weiteren Nutzung aus?”

# Brainstorming

Was müssen Sie bei der Datennachnutzung beachten?



# Datennachnutzung

- Meistens haben veröffentlichte Daten eine **Lizenz** angehängt.
- Diese bestimmt, in welchem Umfang und zu welchen Bedingungen ich die Daten **nachnutzen darf**.

# Was ist eine Lizenz?

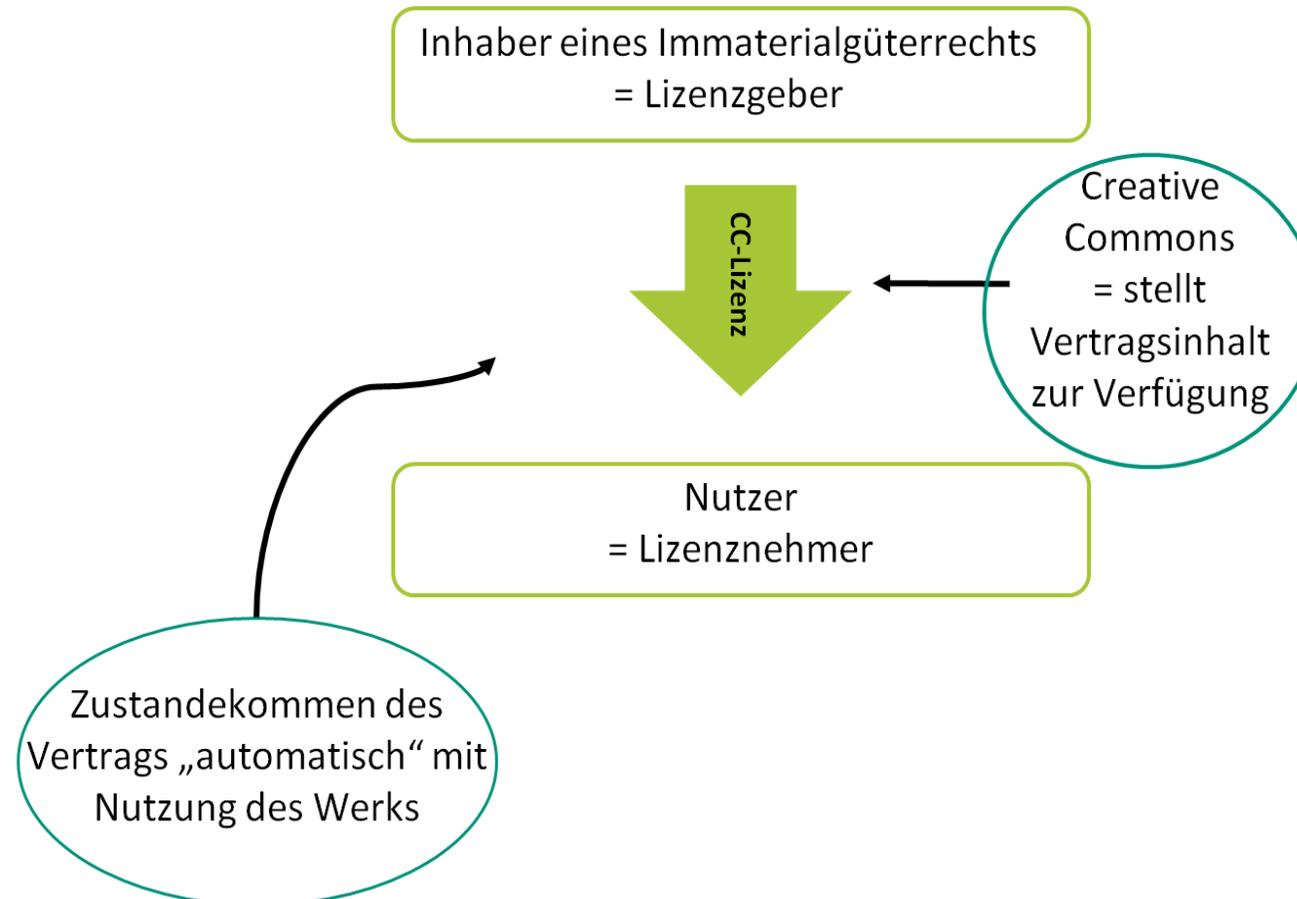
**Lizenz** = ein vertraglich vereinbartes Nutzungsrecht. Damit erlaubt der Rechteinhaber seinem Vertragspartner, ein Werk auf verschiedene Arten zu nutzen (z.B. zu kopieren, zu speichern oder digital zugänglich zu machen)

➔ nicht lizenzierte Werke sind nicht frei verwendbar!

**kommerzielle Lizenz** = Vergabe der Lizenz gegen Zahlung einer Lizenzgebühr

**freie Lizenz** = Gestattung der unentgeltlichen Nutzung des Werkes

# Was ist eine Lizenz?



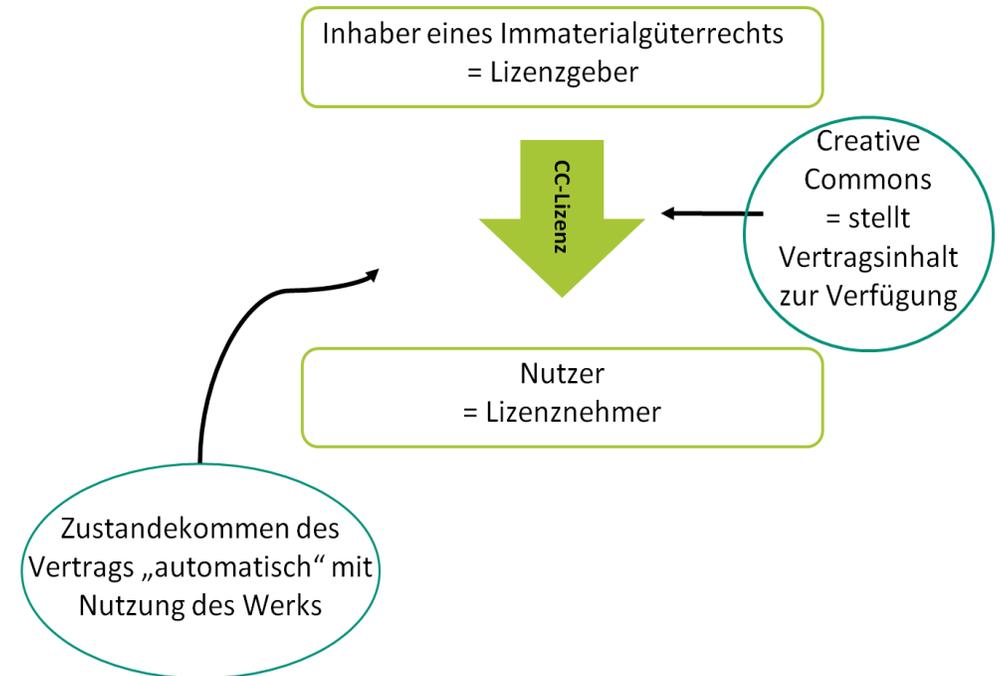
# Urheberrecht

Was besagt diese?:

“Creative Commons-Lizenzen sind bewährte Standardlizenzverträge, die eine weitgehende Nutzungsfreigabe von **urheberrechtlich geschütztem Material** ermöglichen.[...]“ (creativecommons.net/FAQ Punkt 2.1)“

Es sollten nur Inhalte und Materialien unter eine CC-Lizenz gestellt werden, die auch Schutz nach den Vorschriften des Urheberrechts genießen. Für das Teilen gemeinfreier Inhalte kann die Public Domain Mark gewählt werden. Sie ist keine Lizenz, sondern ein Hinweis auf die Gemeinfreiheit. [...].“

Da gemeinfreie Inhalt per Gesetz ohne jegliche Beschränkung oder Pflicht genutzt werden können, dürften die Lizenzbestimmungen angesichts dieser Formulierung hierfür gar nicht gelten.



# Creative Commons-Lizenzen

## CC- Lizenzen

	Lizenzbedingung	Weitergabe
	 gemeinfrei	Materialien dürfen frei genutzt werden ohne weitere Bedingungen oder Angaben
	 Namensnennung	Materialien dürfen geteilt, verändert und kommerziell genutzt werden
	 Namensnennung + Weitergabe unter gleichen Bedingungen	Materialien dürfen geteilt, verändert und kommerziell genutzt werden
	 Namensnennung + Nichtkommerziell	Materialien dürfen geteilt und verändert werden
	 Namensnennung + Nichtkommerziell + Weitergabe unter gleichen Bedingungen	Materialien dürfen geteilt und verändert werden
	 Namensnennung + keine Bearbeitung	Materialien dürfen nur geteilt werden
	 Namensnennung + Nichtkommerziell + keine Bearbeitung	Materialien dürfen nur geteilt werden

Die Abbildung basiert auf den Werken:

1. Erläuterung der einzelnen CC-Lizenzen von Christine Ruthenfranz, Ruhr-Uni Bochum eScouts OER CC-BY-SA 4.0 [https://moodle.ruhr-uni-bochum.de/m/pluginfile.php/417553/mod\\_gaze/content/45/CC-Lizenzen-Tabelle.jpg](https://moodle.ruhr-uni-bochum.de/m/pluginfile.php/417553/mod_gaze/content/45/CC-Lizenzen-Tabelle.jpg)
2. Creative Commons license spectrum von Shaddim, original CC license symbols von Creative Commons, CC BY 4.0, via Wikimedia Commons [https://upload.wikimedia.org/wikipedia/commons/e/e1/Creative\\_commons\\_license\\_spectrum.svg](https://upload.wikimedia.org/wikipedia/commons/e/e1/Creative_commons_license_spectrum.svg)



# CC-Lizenzen - Kombinationsmöglichkeiten

- Wenn mehrere Datensätze kombiniert werden sollen, müssen die Lizenzen der einzelnen Datensätze kombiniert werden können
- der [Lizenz Mixer](#) hilft

## Creative Commons Mixer

**Input**  
List of all used licenses in your composed work.

-   ZERO
-   BY
-    BY SA
-    BY NC
-    BY ND
-     BY NC SA
-     BY NC ND



**Output**  
List of all allowed licenses for the composed work.

   BY SA

**Will it blend?**

This is no legal advice. Its a prototype of a 1day hackathon. We invite everyone to develop this prototype together with us into a service the community can rely on for the future: [info@edu-sharing.net](mailto:info@edu-sharing.net)

# Was denken Sie?

Was passiert, wenn der Datensatz mit einer **CC0 Lizenz** oder einem **Public Domain Mark** versehen ist?

→ entbindet mich das von der Pflicht, die Autorin zu zitieren?

**NEIN! -> gute wissenschaftliche Praxis!**

# Gute wissenschaftliche Praxis

## Leitlinie 7: Phasenübergreifende Qualitätssicherung

- ▶ Die Wissenschaftlerinnen und Wissenschaftler führen jeden Teilschritt im Forschungsprozess *lege artis* durch. Wenn wissenschaftliche Erkenntnisse öffentlich zugänglich gemacht werden (im engeren Sinne in Form von Publikationen, aber auch im weiteren Sinne über andere Kommunikationswege), werden stets die angewandten Mechanismen der Qualitätssicherung dargelegt. Dies gilt insbesondere, wenn neue Methoden entwickelt werden.

# Gute wissenschaftliche Praxis

## Leitlinie 7: Phasenübergreifende Qualitätssicherung

- ▶ Die Wissenschaftlerinnen und Wissenschaftler führen jeden Teilschritt im Forschungsprozess *lege artis* durch. Wenn wissenschaftliche Erkenntnisse öffentlich zugänglich werden, werden stets die Herkunft von im Forschungsprozess verwendeten Daten, Organismen, Materialien und Software wird kenntlich gemacht und die Nachnutzung belegt; die Originalquellen werden zitiert. Art und Umfang von im Forschungsprozess entstehenden Forschungsdaten werden beschrieben. Der Umgang mit ihnen wird, entsprechend den Vorgaben im betroffenen Fach, ausgestaltet. Der Quellcode von öffentlich zugänglicher Software muss persistent, kationen, aber au werden stets die gelegt. Dies gilt i

Die Herkunft von im Forschungsprozess verwendeten Daten, Organismen, Materialien und Software wird kenntlich gemacht und die Nachnutzung belegt; die Originalquellen werden zitiert. Art und Umfang von im Forschungsprozess entstehenden Forschungsdaten werden beschrieben. Der Umgang mit ihnen wird, entsprechend den Vorgaben im betroffenen Fach, ausgestaltet. Der Quellcode von öffentlich zugänglicher Software muss persistent,

# Datennachnutzung

Darf ich aus einer Publikation Abbildungen kopieren und verwenden?

- Meistens auch über Lizenz abgedeckt
- Wenn nicht: Grundsatz “alle Rechte vorbehalten”, wenn urheberrechtlich geschützt
- Nutzbarkeit über Zitatrecht nach § 51 UrhG nur unter bestimmten Voraussetzungen

# Daten“eigentum”

Wem “*gehören*” die Daten?

# Kontext Dateneigentum

- deutsches Recht kennt nur ein Eigentum an Sachen (körperlichen Gegenständen)
- aber auch geistiges Schaffen wird anerkannt und geschützt
- Gesetzgeber hat (in engen Grenzen) eigentumsähnliche Schutzmechanismen geschaffen
- **“Geistiges Eigentum”**
- **Urheberrecht, Markenrecht, Patentrecht**

# Kontext Dateneigentum

ABER:

- Forschungsdaten sind nicht immer “geistiges Eigentum”
- häufig nicht durch ein genanntes Gesetz geschützt

Wann haben meine Daten urheberrechtlichen Schutz?

# Urheberrecht

## **nationales Recht: UrhG**

was regelt das UrhG?

- persönliche geistige Schöpfungen der Literatur, Wissenschaft und Kunst (§ 1)

Warum spielt es hier eine Rolle?

- legt fest, welche Regeln für die Nutzung und den Umgang mit den Ergebnissen eines Schaffensprozesses gelten
- ist Voraussetzung für die Vergabe bestimmter Lizenzen: z.B. der CreativeCommons-Lizenzen

persönliche geistige Schöpfungen = Werke

Voraussetzungen der Einordnung als Werk:

- Wahrnehmbare Formgestalt
- persönliches Schaffen
- geistiger Gehalt
- Eigenpersönliche Prägung

# Urheberrecht

## Wahrnehmbare Formgestalt

nicht bloße Ideen.  
Mittelbare  
Wahrnehmbarkeit (über  
techn. Hilfsmittel)  
ausreichend

## Persönliches Schaffen

keine von Maschinen  
oder Tieren  
geschaffenen Dinge  
  
-> KI-generierte Inhalte  
sind keine Werke iSd  
UrhG

## Geistiger Gehalt

Erzeugung einer  
Gedanken- oder  
Gefühlswelt, die in  
irgendeiner Weise  
anregend auf den  
Betrachter wirkt

## Eigenpersönliche Prägung

Mindestmaß an  
Originalität, sog.  
“hinreichend  
schöpferischer  
Eigentümlichkeitsgrad”  
BGH GRUR 1988, 533 (535)  
  
oder auch:  
**Schöpfungshöhe**

# Kontext Dateneigentum

## Urheberrechtlich geschützte Daten:

- Entscheidungen über die Verwertung und Ausübung der Urheberrechte obliegen der Urheberin §§ 11 UrhG
- also derjenigen Person, die das Werk geschaffen/die Publikation verfasst/die Grafik geschaffen hat

## Nicht urheberrechtlich geschützte Daten:

- Vorgaben der GWP

# Gute wissenschaftliche Praxis

## Leitlinie 10: Rechtliche und ethische Rahmenbedingungen, Nutzungsrechte

- ▶ Wissenschaftlerinnen und Wissenschaftler gehen mit der verfassungsrechtlich gewährten Forschungsfreiheit verantwortungsvoll um. Sie berücksichtigen Rechte und Pflichten, insbesondere solche, die aus gesetzlichen Vorgaben, aber auch aus Verträgen mit Dritten resultieren, und holen, sofern erforderlich, Genehmigungen und Ethikvoten ein und legen diese vor. Im Hinblick auf Forschungsvorhaben sollten eine gründliche Abschätzung der Forschungsfolgen und die Beurteilung der jeweiligen ethischen Aspekte erfolgen. Zu den rechtlichen Rahmenbedingungen eines Forschungsvorhabens zählen auch dokumentierte Vereinbarungen über die Nutzungsrechte an aus ihm hervorgehenden Forschungsdaten und Forschungsergebnissen.

# Gute wissenschaftliche Praxis

## Leitlinie 10: Rechtliche und ethische Rahmenbedingungen, Nutzungsrechte

- ▶ Wissenschaftlerinnen und Wissenschaftler sind zu rechtlich gewährten Forschungsrücksichtigen Rechten und Pflichten verpflichtet, aber auch auf diese vor. Im Hinblick auf Forschungsschätzung der Forschungsfolgen und ethischen Aspekte erfolgen. Zu den Nutzungsrechten an aus ihm Forschungsergebnissen.

Wissenschaftlerinnen und Wissenschaftler treffen, sofern möglich und zumutbar, zu einem frühestmöglichen Zeitpunkt im Forschungsvorhaben dokumentierte Vereinbarungen über die Nutzungsrechte. Dokumentierte Vereinbarungen bieten sich insbesondere an, wenn an einem Forschungsvorhaben mehrere akademische und/oder nicht akademische Einrichtungen beteiligt sind oder wenn absehbar ist, dass eine Wissenschaftlerin oder ein Wissenschaftler die Forschungseinrichtung wechseln wird und die von ihr/von ihm generierten Daten weiterhin für (eigene) Forschungszwecke verwenden möchte. **Die Nutzung steht insbesondere der Wissenschaftlerin und dem Wissenschaftler zu, die/der sie erhebt.** Im Rahmen eines laufenden Forschungsprojekts entscheiden auch die Nutzungsberechtigten (insbesondere nach Maßgabe datenschutzrechtlicher Bestimmungen), ob Dritte Zugang zu den Daten erhalten sollen.

# Datenveröffentlichung

Was muss man beim Veröffentlichen von Daten beachten?

# Urheberrecht

## Wann hat eine Arbeit die nötige Originalität/Schöpfungshöhe?

Originalität braucht folgendes:

- es muss ein **Teil des Schaffenden** erkennbar sein durch Auswahl, Anordnung und Kombination in einer originellen Weise
- der Inhalt des Schaffens darf **nicht im Wesentlichen durch die enthaltenen Informationen bestimmt** werden, sodass die Arbeit allein durch technische Funktionen gekennzeichnet **ist**

Bsp: Anfertigung und Zusammentragen von Bodenmessungen; Niederschrift eines Untersuchungsprotokolls (-)

# Wie lizenziere ich meine Daten?

- Nutze standardisierte Lizenzverträge
- Prüfe ob eine bestimmte Lizenz empfohlen oder verpflichtend für dich ist
  - ◆ z.B vonseiten des Förderers, eines Repos oder deiner Institution
- Leitfrage: Was sollen andere mit meinen Daten tun dürfen? (Und gilt das für alle Bestandteile, wie Metadaten, Volltexte, Bilder etc. ?)
- weit verbreitete Lizenzen für Forschungsdaten: Creative Commons Lizenz



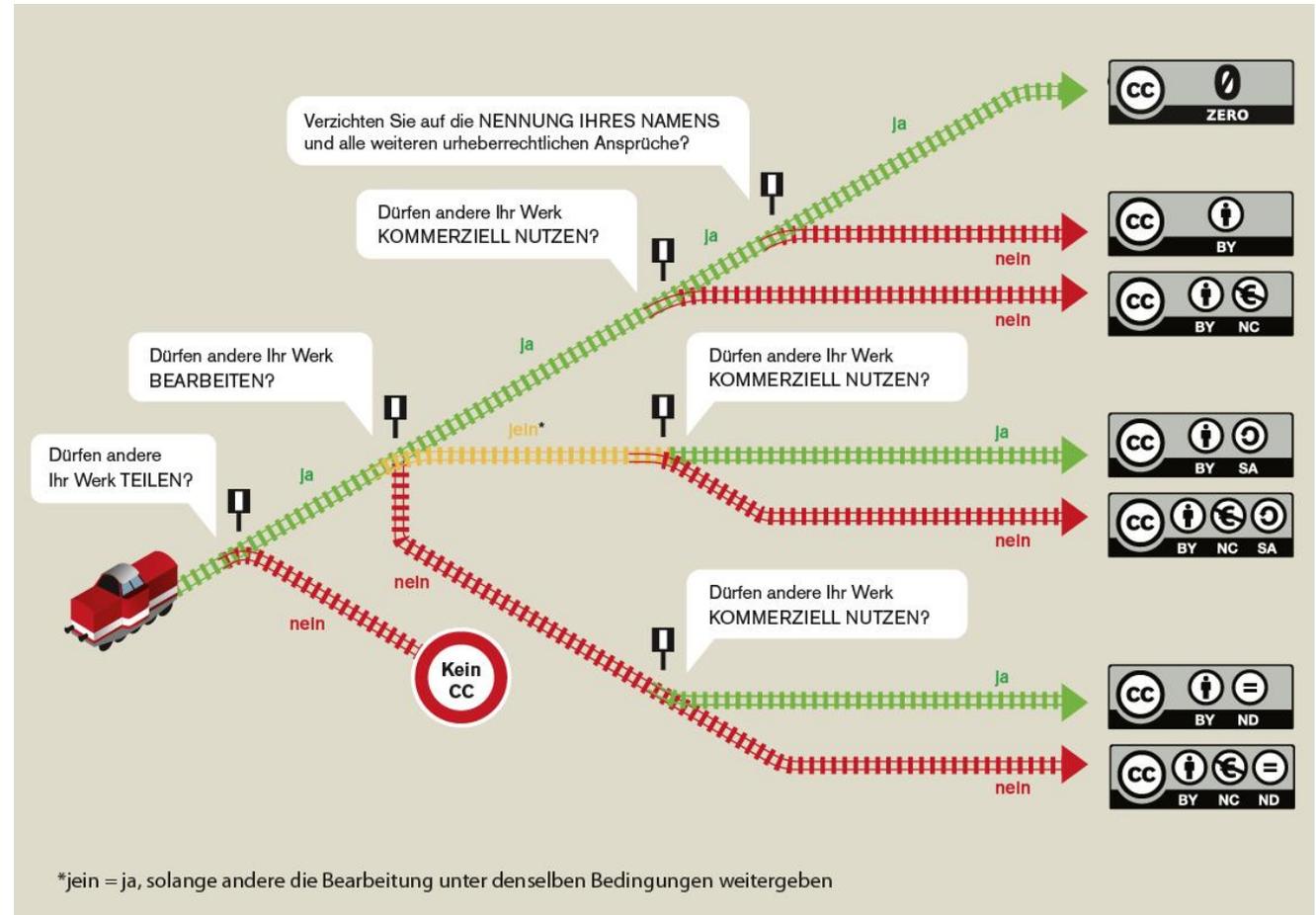
# CC-Lizenzen: richtige Lizenz auswählen

## Eine Lizenz auswählen

- [License Chooser](#) [Creative Commons]
- [Overview CC Licences](#) [NFDI4Health]

## Andere

- [A Practical Guide to Using Creative Commons Licences](#) [Dr. Till Kreuzer]
- [Copyright in Academic Work](#) [BMBF]
- [Creative Commons Mixer](#) [edu-sharing]
- [FAQs](#) [Creative Commons]



# Nutzungsrecht

Wer hat unter welchen Bedingungen welche Rechte an Forschungsdaten?

# Nutzungsrechte

- Nutzungsrechte regeln, **wer** mit einem Datensatz **was tun darf**.
- **ausschließliches Nutzungsrecht** = außer dem Begünstigten darf kein anderer die Daten nutzen; auch nicht die Urheberin selbst

## Fun Fact

Lizenzen sind vorformulierte  
Nutzungsverträge, in denen  
Nutzungsrechte geregelt sind

# Big Data-Verarbeitung im Behördenalltag

## Cloud-Lösungen

Lucia Vedder & Florian Beyer im Namen des FAIRagro Konsortiums

Rheinische Friedrich-Wilhelms-Universität Bonn,  
Julius-Kühn-Institut

Funded by  
**DFG** Deutsche  
Forschungsgemeinschaft  
German Research Foundation

project number 501899475

In cooperation with

**nfdi**

## thematische Rückmeldungen...

Wir haben in einem Projekt genomische Daten erzeugt, mit denen wir die Widerstandsfähigkeit von Eschen gegenüber dem Eschentriebsterben vorhersagen möchten.

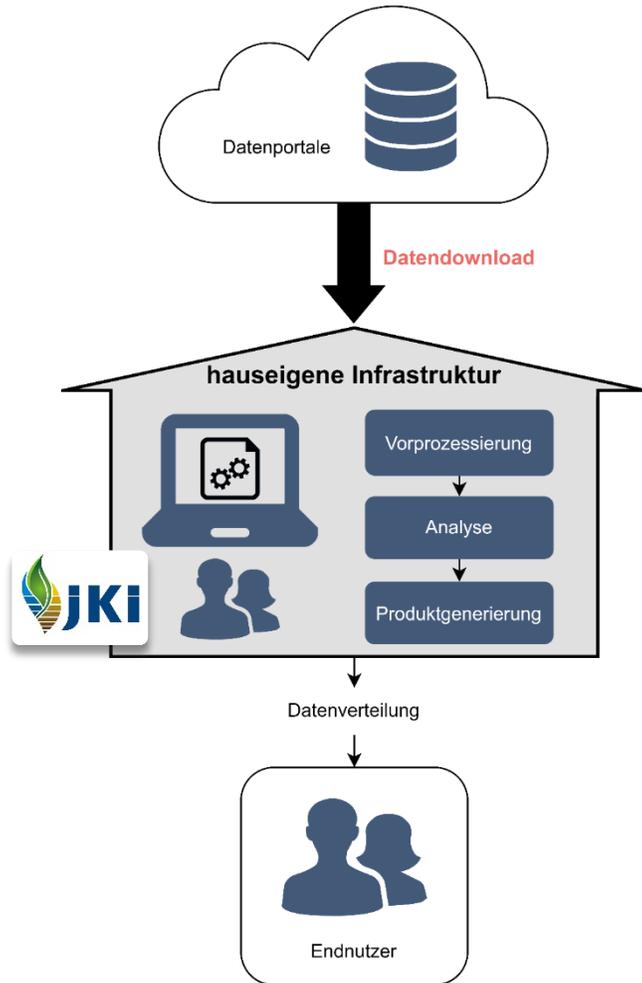
**Unsere bestehende IT-Struktur lässt eine Auswertung der Daten nur schwer zu.**

Ich habe 25 große Datensätze als Dreiecksmatrix mit Werten aus [0;1]. Die Dimension der Matrix beträgt pro Datensatz ~ 30.000 x 30.000 (mit entsprechend 30.000 \* 30.000 / 2 = 4,5 \* 10<sup>8</sup> Einträgen).

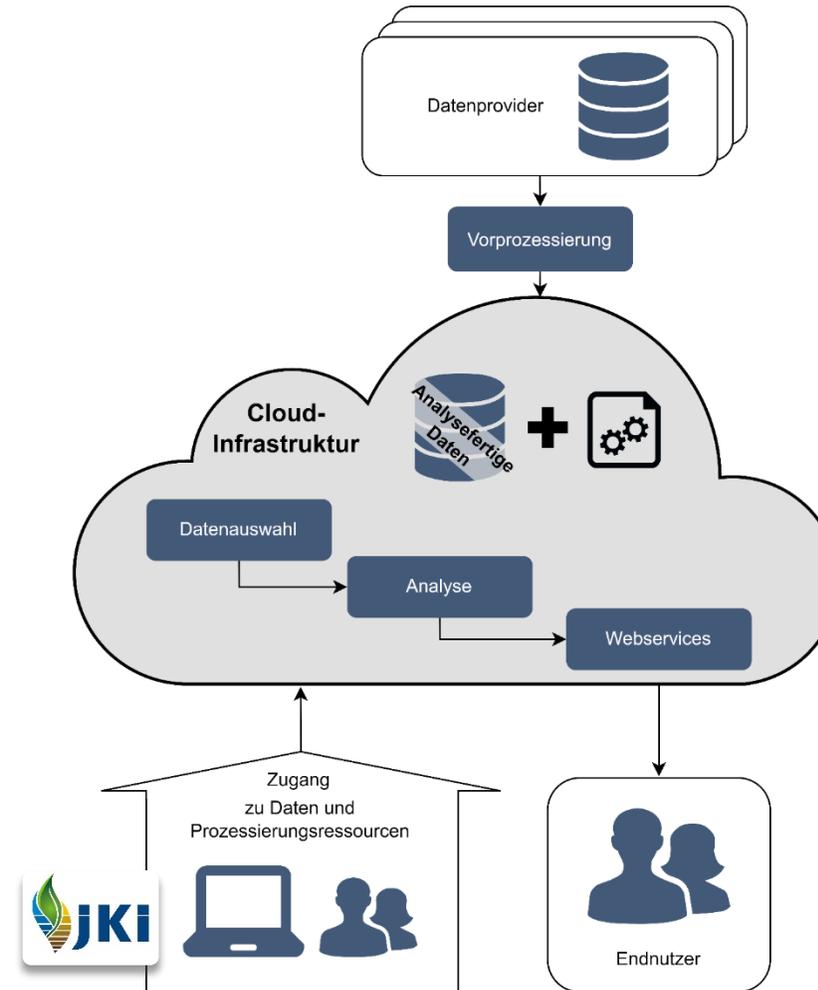
Diese Matrix soll (mit Spalten- und Zeilennamen) in eine Form mit 3 Spalten (Spaltenname, Zeilenname, Zelleneintrag) und 4,5 \* 10<sup>8</sup> Zeilen gebracht werden, um später (nach "grammar of graphics") eine Grafik daraus zu erstellen.

Ich habe auch noch andere Beispiele **für große Datensätze**, falls gewollt.

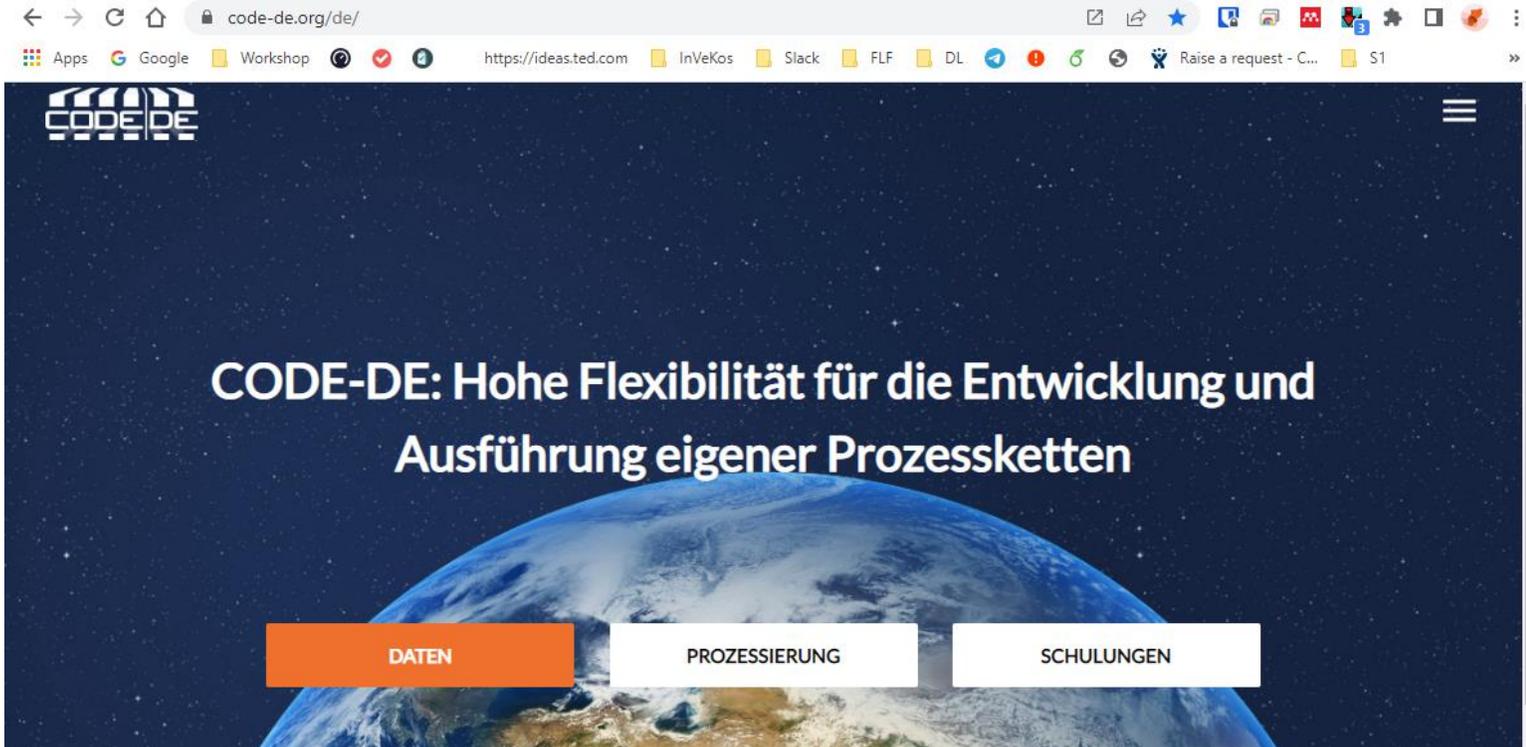
## Gestern...



## Heute (eher Morgen?..)



# CODE-DE – Cloudressourcen für deutsche Behörden



Screenshot (2023-11-10) [code-de.org](https://code-de.org)

- Einfacher **Zugang zu Copernicusdaten** und **skalierbare Cloudressourcen**
- Skalierbare Prozessierungsleistung und Breitbandanschluss
- Benutzerfreundliche Datenanalysetools
- **BSI-zertifiziertes** in Frankfurt am Main (ISO 27001 & C5)
- Hohe Flexibilität für die Entwicklung und Ausführung eigener Prozessketten
- Teil der Deutschen Geoinformationstrategie

[code-de.org](https://code-de.org)

Förderer



Bundesministerium  
für Digitales  
und Verkehr

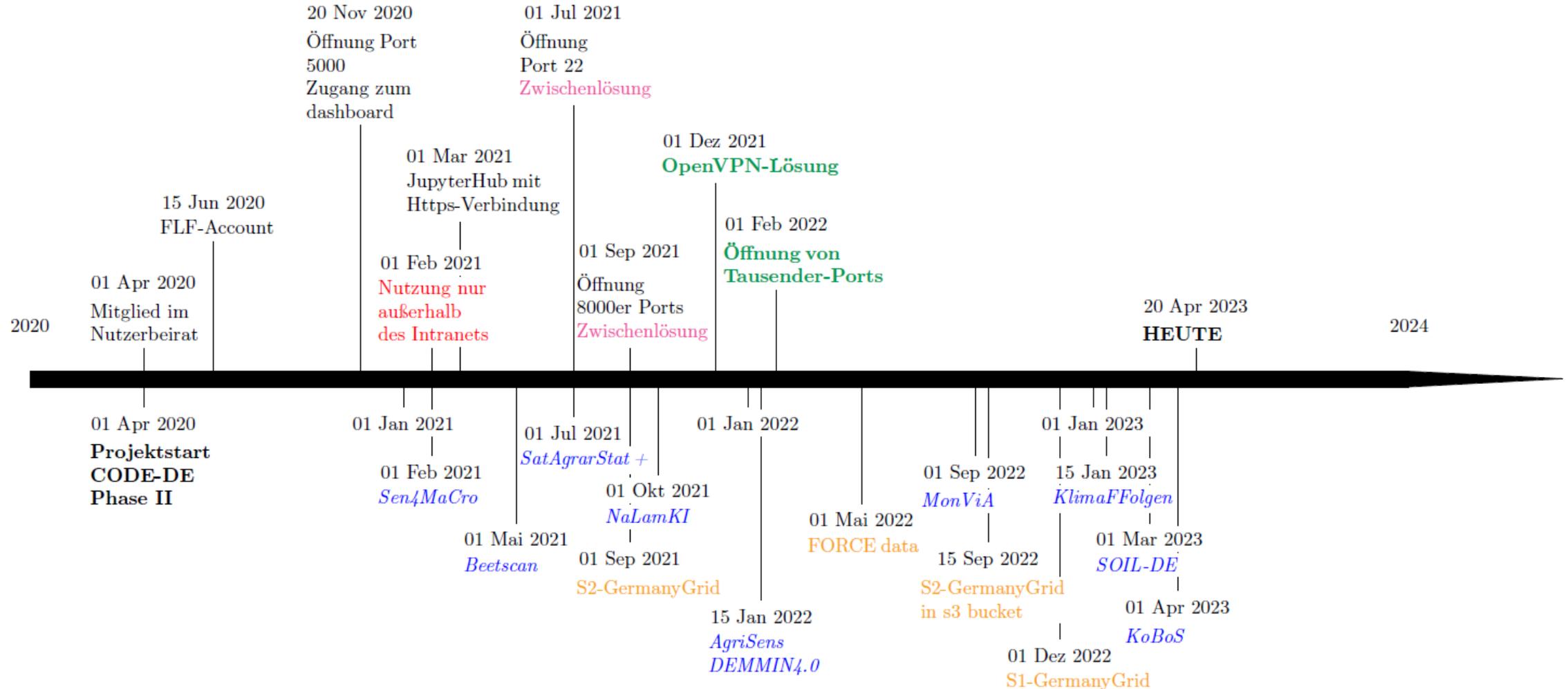
Management



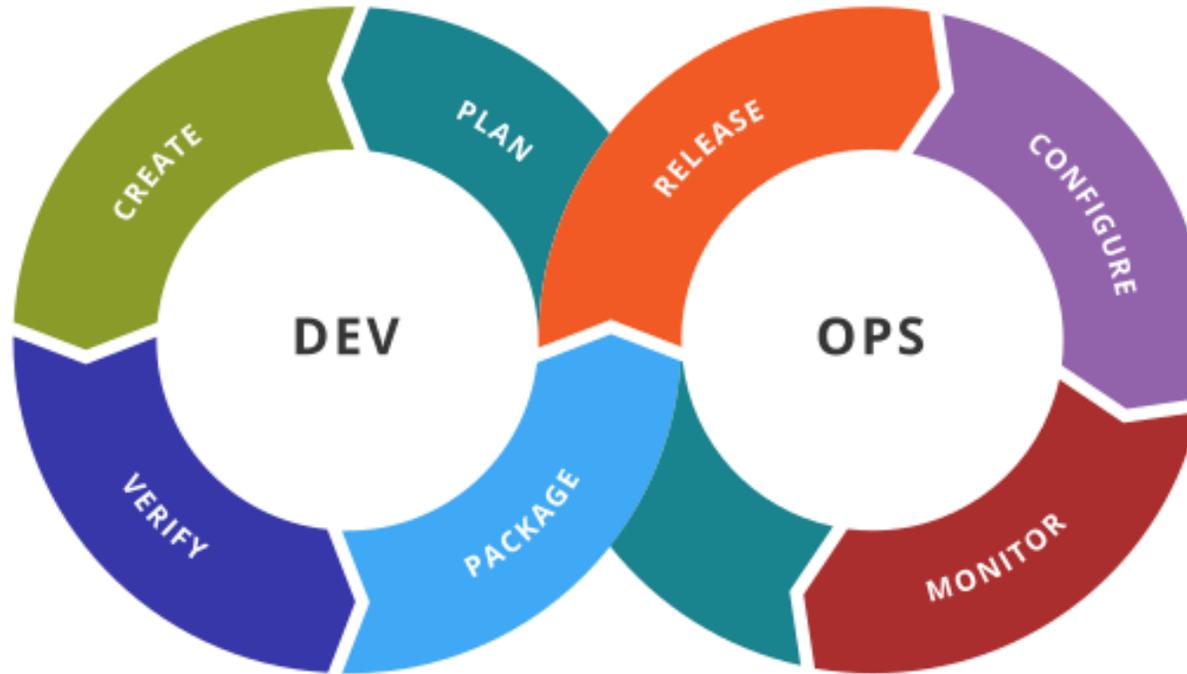
Technische  
Betreuung



# CODE-DE am JKI – Eine Rückschau



# DevOps-Prinzipien

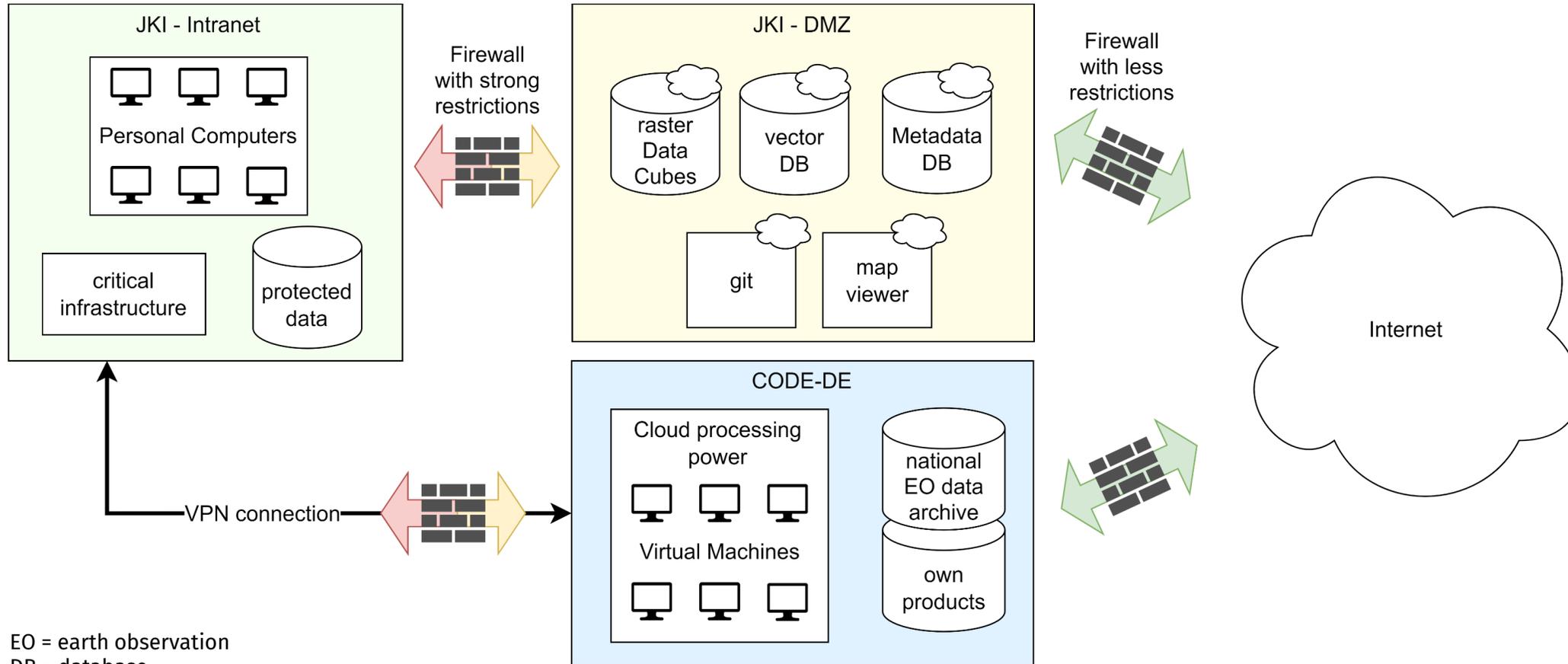


Wissenschaftler:innen

IT-Service (DK – Digitalisierung und Künstl. Intelligenz)

- Bottom-up
- Schaffung von agilen Gruppen aus DEV und OPS die sich regelmäßig treffen

# Geodateninfrastruktur des JKI



EO = earth observation  
DB = database  
DMZ = demilitarized zone

# Clouddienste (für Behörden/Wissenschaft)

## Allgemein:

- Google Cloud, Amazon AWS, ...
- [Microsoft Azure](#)
- [Deutsche Verwaltungswolke](#)
- [DFN-Clouddienste](#) (buchbare Clouds: [Hier](#))

## Geodaten (Fernerkundung):

- [EO lab](#) (für die Wissenschaft)
- [Code-DE](#) (für Behörden)

## Biowissenschaften:

- [de.NBI](#) (Bioinformatik)

## Cloudressourcen für bayrische Landesbehörden?

- private Anbieter existieren
- [Ankündigung](#) vom Land Bayern (2023)

# Von A wie Auffinden bis V wie Veröffentlichen

Marcus Schmidt, Wahib Sahwan, Elena Rey Mazón  
im Namen des FAIRagro Konsortiums

Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) e.V.  
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)

Funded by  
**DFG** Deutsche  
Forschungsgemeinschaft  
German Research Foundation

project number 501899475

In cooperation with

**nfdi**

## AGENDA

1. *Datenmanagementpläne*
2. *Datenhaltung während der Projektlaufzeit*
3. *Daten veröffentlichen im BonaRes Repository*

## Datenmanagementpläne

- DMP-Tools Pflanzen:

- NFDI Dataplan -  
<https://www.nfdi4plants.de/dataplan/>
- RDMO Instanzen -  
<https://rdmorganiser.github.io>

- DMP-Tools generisch:

- DMP Online -  
<https://dmponline.dcc.ac.uk/>
- Data Stewardship Wizard -  
<https://ds-wizard.org/>

Allgemeine DMP-Informationen von einem anderen NFDI-Konsortium -  
<https://knowledgebase.nfdi4microbiota.de/RDM-Plan/01-dmp.html#introduction>

# Datenhaltung während der Projektlaufzeit

# DATENMANAGEMENTPLAN (DMP)

DATEN ERHEBEN



DATEN VERÖFFENTLICHEN  
IM REPOSITORIUM MIT DOI

Mglk. 1

„Legt eure Daten an Ort xy ab.“



- Chaos vorprogrammiert

Mglk. 2

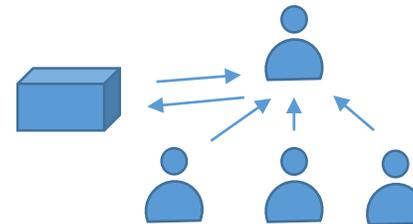
„Schickt eure Daten an Person xy.“



- Was wenn die Person geht?

Mglk. 3

KURATIERTE ABLAGE



- Regelt in Policy
- Einheitliche Metadaten
- Versionierung

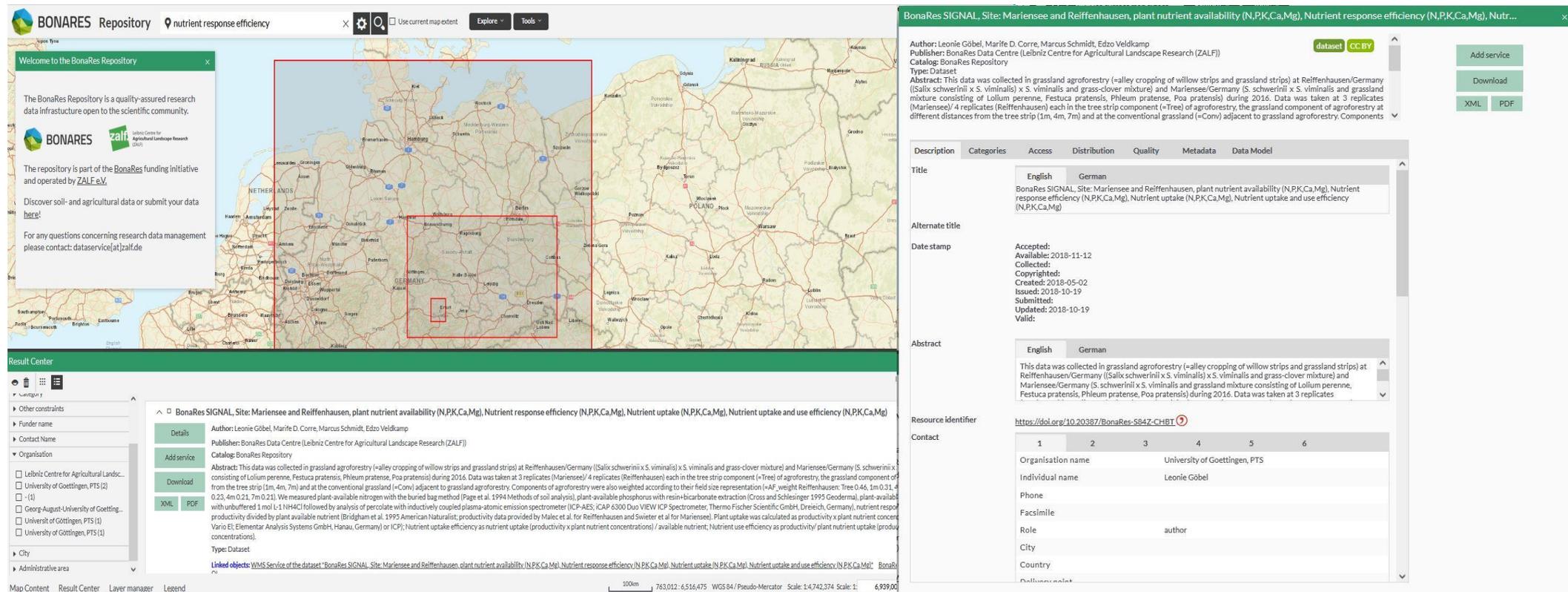


BEREIT ZUR VERÖFFENTLICHUNG

PS: Systeme, die Projektlaufzeit und Veröffentlichung kombinieren existieren und sind sicher die Zukunft, sie sind aber noch lange nicht überall vorhanden

# Daten veröffentlichen im BonaRes Repository

# Dataportal: <https://maps.bonares.de/mapapps/resources/apps/bonares/index.html?lang=en> Suchen, Filtern, Karte und Download



**BonaRes SIGNAL, Site: Mariensee and Reiffenhausen, plant nutrient availability (N,P,K,Ca,Mg), Nutrient response efficiency (N,P,K,Ca,Mg), Nutr...**

Authors: Leonie Göbel, Marife D. Corre, Marcus Schmidt, Edzo Veldkamp  
 Publisher: BonaRes Data Centre (Leibniz Centre for Agricultural Landscape Research (ZALF))  
 Catalog: BonaRes Repository  
 Type: Dataset

**Abstract:** This data was collected in grassland agroforestry (=alley cropping of willow strips and grassland strips) at Reiffenhausen/Germany (Salix schwerinii x S. viminalis) and Mariensee/Germany (S. schwerinii x S. viminalis and grass-clover mixture) during 2016. Data was taken at 3 replicates (Mariensee)/ 4 replicates (Reiffenhausen) of agroforestry, the grassland component of agroforestry at different distances from the tree strip (1m, 4m, 7m) and at the conventional grassland (=Corn) adjacent to grassland agroforestry. Components

**Description** | Categories | Access | Distribution | Quality | Metadata | Data Model

**Title**  
 English: BonaRes SIGNAL, Site: Mariensee and Reiffenhausen, plant nutrient availability (N,P,K,Ca,Mg), Nutrient response efficiency (N,P,K,Ca,Mg), Nutrient uptake (N,P,K,Ca,Mg), Nutrient uptake and use efficiency (N,P,K,Ca,Mg)  
 German: BonaRes SIGNAL, Site: Mariensee und Reiffenhausen, pflanzliche Nährstoffverfügbarkeit (N,P,K,Ca,Mg), Nährstoffeffizienz (N,P,K,Ca,Mg), Nährstoffaufnahme (N,P,K,Ca,Mg), Nährstoffaufnahme und -nutzungseffizienz (N,P,K,Ca,Mg)

**Alternate title**

**Date stamp**  
 Accepted: 2018-11-12  
 Available: 2018-11-12  
 Collected: 2018-05-02  
 Copyrighted: 2018-05-02  
 Created: 2018-10-19  
 Issued: 2018-10-19  
 Submitted: 2018-10-19  
 Updated: 2018-10-19  
 Valid: 2018-10-19

**Abstract**  
 English: This data was collected in grassland agroforestry (=alley cropping of willow strips and grassland strips) at Reiffenhausen/Germany (Salix schwerinii x S. viminalis) and Mariensee/Germany (S. schwerinii x S. viminalis and grass-clover mixture) during 2016. Data was taken at 3 replicates (Mariensee)/ 4 replicates (Reiffenhausen) of agroforestry, the grassland component of agroforestry at different distances from the tree strip (1m, 4m, 7m) and at the conventional grassland (=Corn) adjacent to grassland agroforestry. Components  
 German: Diese Daten wurden in Grünland-Agroforst (Allee-Anlage von Weidenstreifen und Grünlandstreifen) bei Reiffenhausen/Deutschland (Salix schwerinii x S. viminalis) und Mariensee/Deutschland (S. schwerinii x S. viminalis und Gras-Klee-Mischung) im Jahr 2016 gesammelt. Die Daten wurden an 3 Replikaten (Mariensee) / 4 Replikaten (Reiffenhausen) von Agroforst, dem Grünlandbestandteil der Agroforst, in unterschiedlichen Abständen vom Baumstreifen (1 m, 4 m, 7 m) und an dem konventionellen Grünland (=Mais) angrenzend an Grünland-Agroforst, gesammelt. Die Komponenten

**Resource identifier**  
<https://doi.org/10.20387/BonaRes-S84Z-CHBT>

**Contact**

1	2	3	4	5	6
Organisation name	University of Goettingen, PTS				
Individual name	Leonie Göbel				
Phone					
Facsimile					
Role	author				
City					
Country					

