

How FAR is FAIR?

Der öffentliche Zugang zu geisteswissenschaftlichen Forschungsdaten als gute wissenschaftliche Praxis und die Rolle des Forschungsdatenmanagements

Dr. Ulrike Wuttke, Dr. Jochen Klar

DH-Kolloquium, BBAW, 2.8.2019

DOI: <http://doi.org/10.5281/zenodo.3365979>

FH Potsdam, FB Informationswissenschaften, RDMO

Twitter: [@uwuttke](https://twitter.com/uwuttke)

ORCID: <https://orcid.org/0000-0002-8217-4025>

Berater und Softwareentwickler, RDMO

Twitter: [@jochenklar](https://twitter.com/jochenklar)

ORCID: <https://orcid.org/0000-0002-5883-4273>

FAIR?

EVIL;

Prolog

EU Recommendation (2018)

“Member States should ensure that, as a result of these policies or action plans:

- **data management planning becomes a standard scientific practice early in the research process** when data is generated or collected, including through the requirement of data management plans,
- research data that results from publicly funded research becomes and stays findable, accessible, interoperable and re-usable (**‘FAIR principles’**) within a secure and trusted environment, through digital infrastructures”

Neuer DFG Kodex “Leitlinien zur Sicherung guter wissenschaftlicher Praxis”



- Inkrafttreten 01.08.2019
- Übergangsregelung bis 31.07.2021
- Neuer Kodex ist durchdrungen von Open Science-Paradigmen:
 - Offener Zugang zu Forschungsergebnissen
 - Vielfalt von Publikationsorganen und -formen
 - FAIR-Prinzipien
 - Langzeitarchivierung

Leitlinie 13: Herstellung von öffentlichem Zugang zu Forschungsergebnissen

“Grundsätzlich bringen Wissenschaftlerinnen und Wissenschaftler **alle Ergebnisse** in den wissenschaftlichen Diskurs ein. [...] Wissenschaftlerinnen und Wissenschaftler entscheiden in eigener Verantwortung – unter Berücksichtigung der Gepflogenheiten des betroffenen Fachgebiets –, ob, wie und wo sie ihre Ergebnisse öffentlich zugänglich machen. Ist eine Entscheidung, Ergebnisse öffentlich zugänglich zu machen, erfolgt, **beschreiben Wissenschaftlerinnen und Wissenschaftler diese vollständig und nachvollziehbar**. **Dazu gehört es auch, soweit dies möglich und zumutbar ist, die den Ergebnissen zugrunde liegenden Forschungsdaten, Materialien und Informationen, die angewandten Methoden sowie die eingesetzte Software verfügbar zu machen und Arbeitsabläufe** umfänglich darzulegen. **Selbst programmierte Software** wird unter Angabe des Quellcodes öffentlich zugänglich gemacht.”

https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf

Leitlinie 13: Herstellung von öffentlichem Zugang zu Forschungsergebnissen

“Erläuterungen: Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit der Forschung und Nachnutzbarkeit hinterlegen Wissenschaftlerinnen und Wissenschaftler, wann immer möglich, **die der Publikation zugrunde liegenden Forschungsdaten und zentralen Materialien** – den **FAIR-Prinzipien** („**Findable, Accessible, Interoperable, Re-Usable**“) folgend – zugänglich in anerkannten Archiven und Repositorien.”

Leitlinie 17: Archivierung

„Wissenschaftlerinnen und Wissenschaftler sichern öffentlich zugänglich gemachte Forschungsdaten beziehungsweise Forschungsergebnisse sowie die ihnen zugrunde liegenden, zentralen Materialien und gegebenenfalls die eingesetzte Forschungssoftware, gemessen an den Standards des betroffenen Fachgebiets, **in adäquater Weise** und bewahren sie für einen **angemessenen Zeitraum** auf. Sofern nachvollziehbare Gründe dafür existieren, bestimmte Daten nicht aufzubewahren, legen die Wissenschaftlerinnen und Wissenschaftler dies dar. **Hochschulen und außerhochschulische Forschungseinrichtungen stellen sicher, dass die erforderliche Infrastruktur vorhanden ist, die die Archivierung ermöglicht.**“

Forschungsdaten in den Geisteswissenschaften und die FAIR Prinzipien

Definition: Geisteswiss. Forschungsdaten

- Keine allgemeingültige Definition, in den Geisteswissenschaften:
 - große **Heterogenität**
 - Daten in **analoger und digitaler** Form (digitalisiert und “born digital”)
 - **Unterschiedliche mediale Formen**, bilden i.d.R. den Inhalt oder einen Teil eines **Dokumentes** (=Träger von Informationen von kulturellem, künstlerischem, geistigem, medialem, sozialem, soziologischem, historischem, politischem oder religiösem Interesse)
 - Erhoben in einer Form, die eine geisteswissenschaftliche **Interpretierbarkeit** gewährleistet, d.h. Dokumentation des Entstehungs- und Erhebungskontextes, um die wissenschaftliche Überprüfbarkeit (**Validität**) und Verlässlichkeit (**Reliabilität**) zu gewährleisten (Dokumentation verwendeter Normdaten etc.)

Beispiele: geisteswiss. Forschungsdaten

- Primäre Quellen
- Sekundäre Quellen
- Digitale Werkzeuge
- Annotationen
- Umfragen
- Sprachkorpora
- Wissenschaftliche Sammlungen
- Bibliographische Daten
- etc.

Von der Fragilität digitaler Daten...

Bis zu 90% der digitalen
Forschungsdaten gehen innerhalb eines
kurzen Zeitraums verloren (!)

Stefan Winkler-Nees (DFG) 2011 im Vorwort des Handbuch
Forschungsdatenmanagement,

hg. von Büttner et al.

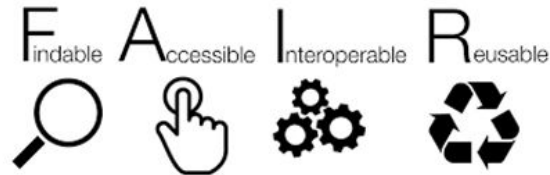
Forschungsdatenmanagement, Open, FAIR

Drei Konzepte bzgl. Forschungsdaten mit unterschiedlichem Fokus

Research Data Management = Stewardship of data from the point of conception onwards, research data management planning, no assumptions about access

Open Data = Data sharing, focus on access and reuse

FAIR Data = Data sharing, focus on access and reuse, FAIR Data Principles



FAIR Data = Open Data?

Open Data = “practice of making underlying research data publicly available, accessible and reusable with minimal restrictions” “as open as possible as closed as necessary”, Tim Berners-Lee’s [5-stars of Linked Open Data](#) (LOD)

FAIR Data =

Findable (Auffindbar),

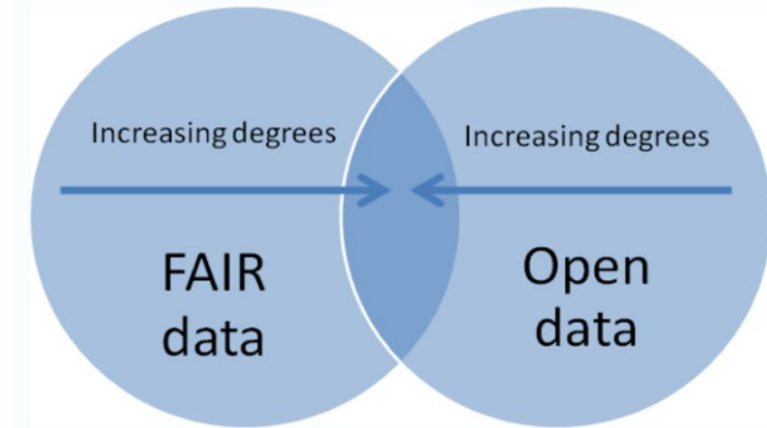
Accessible (Zugänglich),

Interoperable (Interoperabel),

Reusable (Nachnutzbar)

“to both human- driven and machine-driven

activities”, machine-actionability is very important, spectrum with varying degrees of FAIRness



FAIR Data Principles (1)

Findable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the [FAIRification process](#).

- [F1. \(Meta\)data are assigned a **globally unique and persistent identifier**](#)
- [F2. Data are described with **rich metadata** \(defined by R1 below\)](#)
- [F3. Metadata clearly and explicitly include **the identifier of the data** they describe](#)
- [F4. \(Meta\)data are **registered or indexed** in a searchable resource](#)

Accessible: Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

- [A1. \(Meta\)data are **retrievable** by their identifier using a standardised communications protocol](#)
 - [A1.1 The protocol is **open, free, and universally implementable**](#)
 - [A1.2 The protocol allows for an **authentication and authorisation procedure**, where necessary](#)
- [A2. Metadata are **accessible**, even when the data are no longer available](#)

FAIR Data Principles (2)

Interoperable. The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- [I1. \(Meta\)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.](#)
- [I2. \(Meta\)data use vocabularies that follow FAIR principles](#)
- [I3. \(Meta\)data include qualified references to other \(meta\)data](#)

Reusable: The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- [R1. Meta\(data\) are richly described with a plurality of accurate and relevant attributes](#)
 - [R1.1. \(Meta\)data are released with a clear and accessible data usage license](#)
 - [R1.2. \(Meta\)data are associated with detailed provenance](#)
 - [R1.3. \(Meta\)data meet domain-relevant community standards](#)

The principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).



Created by Frühstück
from Noun Project

Beispiel: FAIRe Musikdaten

Findable: URIs (to every note, composer, in principle everything that conveys meaning)

Accessible: HTTP

Interoperable: Shared terms

Reusable: Open licenses

Status Quo:

“Repositories and digital libraries mainly offer digitisations of scores and lyrics (77%), published as PDF and/or JPG (40%)”

“The more the scale of repositories increases, the less structured formats for representing symbolic notation seem to be used and the less depth of analysis is provided”

Beispiel und Zitat adaptiert von Albert Mereño-Peñuela et al.: Linking Midi Data for FAIR Music, Presentation DARIAH Annual Meeting 2019

Warsaw 2019. Link: <http://bit.ly/fair-midi>, Artificial Intelligence and Music DARIAH WG: <https://ai-music.org/>

Zwischenfazit

- Nachnutzbare Daten im Sinne der FAIR-Prinzipien sind mehr als nur online
- Häufige Barrieren und Stolpersteine auf dem Weg zu FAIRen Daten:
 - Fehlende Dokumentation (Herkunft, verwendete Variablen, Methodologie)
 - Unbrauchbare Formate (PDF)
 - Fehlende Angaben über Nutzungsrechte, inkompatible Lizenzen
 - “Trennung” von Metadaten und Forschungsobjekten
 - etc.

Forschungsdatenmanagement

Definition: Forschungsdatenmanagement

“Forschungsdatenmanagement umfasst den **Prozess**, diese Daten im Verlauf ihres gesamten **Lebenszyklus**‘, von der Planung über die Erzeugung, Auswahl, Auswertung und über die Speicherung bis hin zur Aufbereitung für die Nachnutzung, zu verwalten. Aufgrund der Heterogenität von Daten ist es notwendig, **disziplinspezifische**, gegebenenfalls auch **datenspezifische** Lösungen zu entwickeln.”

Gute Gründe für Forschungsdatenmanagement

- Effektiver Mitteleinsatz (kein Datenverlust am Projektende, keine Doppelungen)
- Zunehmender Druck Forschungsergebnisse (inkl. Daten) einer breiten Öffentlichkeit zugänglich zu machen
- Forschungsdaten sind wertvoll
- Digitale Daten sind fragil
- Ausschöpfung der Potentiale der digitalen Transformation der Wissenschaft, die weltweites Data Sharing überhaupt erst möglich macht
- (Big) Data erlaubt ganz andere Analysen
- Reproduzierbarkeit von Forschung (*reproducibility crisis*)
- etc.

Vgl. AG Forschungsdaten der Schwerpunktinitiative “Digitale Information” der Allianz der deutschen Wissenschaftsorganisationen, Forschungsdatenmanagement: Eine Handreichung, 2018, S. 5-6. Online: <http://doi.org/10.2312/allianzaoa.029> (CC BY 4.0)

Last und Lust des Forschungsdatenmanagements

- **Gute Datenqualität ist Ergebnis guten Datenmanagements**
- Gutes Datenmanagement fängt vor dem Projektstart an und geht darüber hinaus (Lebenszyklus von Forschungsdaten)
- **Gute Datenmanagementstrategie hat interne Nutzen für Wissenschaftler*innen, Projekte und Institutionen**
 - Vereinfacht Arbeit mit Daten während Projekt und danach (Effektivität, Wissenserhalt)
 - Kompass für Steuerung von Forschungsprozessen und Verwaltung der Ergebnisse (FDM = digitales Projektmanagement)
 - Sicherung der Nachvollziehbarkeit und Reproduzierbarkeit der Bearbeitung, Analyse und Ergebnisse (GwP!)
 - Vorbereitung auf Datenpublikation
 - Wissenschaftliche Anerkennung und Reputation

Der Datenmanagementplan = Kompass für FDM

In einem DMP behandelte Aspekte:

- Administrative Angaben
- Projektrollen und -aufgaben
- Datenschutz und andere rechtliche Aspekte (Einverständniserklärung, Lizenzen, IPR etc.)
- Angaben zur Wahl der Formate, Metadatenschemata, Namensgebung, Datenstruktur - und -organisation
- Dokumentation
- Aufwand und Finanzierung des Forschungsdatenmanagements
- etc.

Der Datenmanagementplan = Kompass für FDM

- DMP = **Dokument**, das den beabsichtigten Umgang mit Forschungsdaten beschreibt
- **Keine formale allgemeingültige Vorgaben**: ein DMP kann wenige Absätze bis zu mehreren Seiten lang sein
- **Bestandteile**: je nach Projektgröße und Datenvielfalt verschieden, zu beachten sind Vorgaben und Empfehlungen Dritter (Förderer, Institutionen etc.)
- Ideal: **Skizze** beim Antrag, **Aktualisierung** im Verlauf des Projektes

Hilfestellung und Unterstützung DMP

- **Checklisten**

- z. B. Leitfaden WissGrid:

http://www.forschungsdaten.org/images/b/b0/Leitfaden_Data-Management-WissGrid.pdf

- DMP-Checklist des DCC:

<http://www.dcc.ac.uk/resources/data-management-plans/checklist>

- **Muster**

- https://www.cms.hu-berlin.de/de/dl/dataman/arbeiten/dmp_erstellen

- **Werkzeuge**

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

- DMPTool (DCC): <https://dmponline.dcc.ac.uk/>

RDMO



Research Data Management Organiser

Realistischere Sicht auf DMP

- Datenmanagementpläne werden gesehen:
 - als Auflage im Antragsprozess
 - als statisches Dokument
- DMPTool/DMPonline konzentrieren sich auf:
 - Auffinden der verschiedenen Vorlagen der Förderer
 - Produzieren von Text-Dokumenten
- DMPTool/DMPonline sind zentralisierte Webapplikationen:
 - Weitergabe von potentiell sensiblen Informationen
 - Eingeschränkte Anpassbarkeit (institutionell, fachspezifisch)
- Generische Tools lassen sich schlecht mit lokalen Infrastrukturen verknüpfen



RDMO - Vom Plan zum Organisier

- Unterstützung des Datenmanagements über den gesamten Projektverlauf
- Hilfsmittel für Forschende und alle anderen im Forschungsdatenmanagement involvierten Akteure
- Wissensdatenbank der relevanten Informationen für ein nachhaltiges Datenmanagement
 - Strukturiertes Interview
 - Rollenmodell für den Zugriff zu Informationen
 - Ausgabe eines textuellen dmp
 - Schnittstellen zu anderen Tools → *actionable DMP*



RDMO - Lokal statt zentral

- Lokale Installation durch
 - Bibliotheken
 - Forschungsinstitute
 - Infrastrukturen
 - Kollaborationen
- Inhaltliche Anpassung
 - Institutionelles Umfeld
 - Disziplinspezifische Aspekte
- Lokale Anpassungen für eigenes *Corporate Design*
- Reibungslose Installation, Nutzung von Standardtechnologien, Open Source



Projekt RDMO



Harry Enke (AIP)



Jochen Klar (selbst.)



Claudia Kramer (KIT)



Jens Ludwig (SPK)



Olaf Michaelis (AIP)



Heike Neuroth (FHP)



Janine Straka (FHP)



Robert Ulrich (KIT)



Kerstin Wedlich (KIT)



Ulrike Wuttke (FHP)



Leibniz-Institut für
Astrophysik Potsdam



Fachhochschule Potsdam
University of
Applied Sciences



Projekt RDMO

Phase 1 (November 2015 - April 2017)

- Konzeption eines generischen Fragenkatalogs
- Entwicklung der Software
- Community und Nutzertests

Phase 2 (Dezember 2017 - Herbst 2020)

- Weiterentwicklung der Software
- Integration in die Infrastruktur
- Etablierung in der Community
- Nachhaltigkeit / Verstetigung

Gefördert durch



Software

- Realisierung einer Webanwendung auf Basis von:
 - Python und dem Django Framework
 - JavaScript und AngularJS 1
 - Bootstrap 3
 - Pandoc
- Nutzung einer relationalen Datenbank (MySQL, PostgreSQL, Oracle, SQLite)
- Authentifizierungsschnittstellen (LDAP, Shibboleth, OAuth2)
- Open Source seit dem *ersten Tag*, Apache 2.0 Lizenz
- Code auf GitHub unter <https://github.com/rdmorganiser/rdmo>
- DOI mit Zenodo <http://doi.org/10.5281/zenodo.596581>

RDMO Community

RDMO Instanzen

- ~ 10 produktiv / ~ 20 Test/Evaluation

Zusammenarbeit

- Webseite: rdmorganiser.github.io
- GitHub: github.com/rdmorganiser
- Dokumentation: rdmo.readthedocs.io
- Mailingliste: rdmo@listserv.dfn.de
- Slack: rdmo.slack.com
- Twitter: [@rdmorganiser](https://twitter.com/rdmorganiser)
- Tutorials auf forschungsdaten.org



Startseite



RDMO Demo Sprache ▾ Anmelden

RDMO

Ein Werkzeug zur Planung, Umsetzung und der Verwaltung des Forschungsdatenmanagements.

Nasssammlung (Christopher Bulle) / CC BY 2.0

Willkommen bei RDMO

Das Ziel des RDMO Projekts ist es, eine Webapplikation bereitzustellen, die die strukturierte Planung, Umsetzung und Verwaltung der Daten in einem wissenschaftlichen Projekt unterstützt. Zusätzlich sollen die gesammelten Informationen in textueller Form für Anforderungen von Förderern oder für Berichte ausgebbar sein.

Diese Webseite ist ein Prototyp zur Veranschaulichung der bereits implementierten Funktionen.

Wenn Sie mehr über das Projekt erfahren wollen, besuchen Sie [rdmorganiser.github.io](https://github.com/rdmorganiser/rdmorganiser).

Anmelden

Benutzername











Passwort

RDMO Test Instanz: rdmo.aip.de

Projektübersicht

RDMO Demo Management Admin Sprache Jochen Klar

Meine Projekte

Name	Rolle	Erstellt	
Ein neues Forschungsdatenmanagementprojekt	Besitzer	9. Dezember 2016 22:02	 
Oberdeutscher Spiegel Historiae Digital	Besitzer	13. November 2018 17:38	 
Research Data Management Organiser (RDMO)	Besitzer	16. November 2016 15:08	 
Test	Besitzer	30. November 2018 14:53	 
test jf project	Gast	7. August 2017 15:32	 

Meine Daten

Name
Jochen Klar
ORCID ID
0000-0002-5883-4273

Meine Gruppen

Optionen

[Neues Projekt erstellen](#)

Import

Strukturiertes Interview

RDMO Demo Management Admin Sprache Jochen Klar

Fragebogen für Projekt *Research Data Management Organiser (RDMO)*

Allgemein / Thema

Wie lautet die primäre Forschungsfrage des Projektes?

RDMO II baut auf den Ergebnissen des Vorgängerprojektes FDMP-Werkzeug (RDMO) auf, das mit dem Research Data Management Organiser (RDMO) ein Werkzeug zur Unterstützung des Forschungsdatenmanagements entwickelt hat. RDMO wird es Projekten ermöglichen, das Datenmanagement über den gesamten Projektverlauf - d.h. von der Antragstellung bis zum Ende - zu organisieren und nach Bedarf anzupassen. Zum Ende des Projektes FDMP-Werkzeug wurde eine Version veröffentlicht, die hierfür bereits ein Minimal-Set an Funktionalitäten enthält, das vor allem die Planung des Datenmanagements vor dem Projektstart unterstützt.

Bitte geben Sie einige Schlagworte zur Forschungsfrage an.

Bitte nutzen Sie für die Schlagwörter jeweils eine eigene Zeile. Sie können weitere Schlagwörter mit dem grünen Knopf hinzufügen und mit dem roten Kreuz (x) entfernen.

- Forschungsdatenmanagement
- Datenmanagement
- Datenmanagementplan
- active DMP

Fortschritt

Progress bar

Navigation

Bitte beachten sie das durch die Benutzung der Navigation ungesicherte Eingaben verworfen werden.

[Zur vorherigen Frage springen](#)
[Zur nächsten Frage springen](#)
[Zurück zum Projekt](#)

Übersicht

- Allgemein
- Thema
- Disziplin
- Projektablauf
- Projektkoordination
- Projektpartner

Antworten für *Research Data Management Organiser (RDMO)*

Im Folgenden haben wir die von Ihnen eingegebenen Informationen über das Projekt noch einmal zusammengefasst.

Allgemein

Thema

Wie lautet die primäre Forschungsfrage des Projektes?

RDMO II baut auf den Ergebnissen des Vorgängerprojektes FDMP-Werkzeug (RDMO) auf, das mit dem Research Data Management Organiser (RDMO) ein Werkzeug zur Unterstützung des Forschungsdatenmanagements entwickelt hat. RDMO wird es Projekten ermöglichen, das Datenmanagement über den gesamten Projektverlauf - d.h. von der Antragstellung bis zum Ende - zu organisieren und nach Bedarf anzupassen. Zum Ende des Projektes FDMP-Werkzeug wurde eine Version veröffentlicht, die hierfür bereits ein Minimal-Set an Funktionalitäten enthält, das vor allem die Planung des Datenmanagements vor dem Projektstart unterstützt. Ziel von RDMO II ist es, das Werkzeug bis zum Erreichen des vollständig notwendigen Funktionsumfangs weiterzuentwickeln und bis zur Produktionsreife zu bringen, wobei der Fokus im beantragten Nachfolgeprojekt auf der Organisation und Umsetzung des Datenmanagements (v.a. der Aufgabenverwaltung) während des Projekts liegen wird. Darüber hinaus gehört zu den Zielen, die Integration in die lokale Infrastruktur in unterschiedlichen fachlichen und institutionellen Kontexten zu vereinfachen, die Anbindung an überregionale Infrastrukturelemente (z.B. re3data) zu ermöglichen sowie das Tool in der Community zu etablieren. Ein wichtiger Aspekt unter dem Gesichtspunkt der Nachhaltigkeit ist die Etablierung einer Community, die das Tool betreibt und auch nach Ende des Projektes RDMO II technisch wie auch konzeptionell (Anpassung des Fragenkataloges an unterschiedliche disziplinspezifische und institutionelle Kontexte) weiterentwickelt. Daraus ergeben sich eine Reihe von Teilforschungs- (& Entwicklungs)fragen: Welche Features und Funktionalitäten sind (sowohl insgesamt als auch für die einzelnen Stakeholder) notwendig, um das Planung, Organisation und Anpassung des Datenmanagements im gesamten Projektverlauf optimal zu unterstützen und wie lassen sich diese konzeptionell und technisch umsetzen? Welche Anforderungen in Bezug auf die Implementierung / Integration in lokale sowie

Snapshots

→ Aktuell
Zwischenbericht 2018
DFG Antrag

Optionen

[Zurück zum Projekt](#)

Export

[PDF](#)
[Rich Text Format](#)
[Open Office](#)
[Microsoft Office](#)
[HTML](#)
[Markdown](#)
[mediawiki](#)
[LaTeX](#)

Horizon 2020

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

Dataset Anforderungen_FDMP: Als Referenz bezüglich der Anforderungen der verschiedenen Stakeholder an das Tool, insbesondere solcher, die im FDMP-Projekt noch nicht oder nicht vollständig umgesetzt werden konnten.

Dataset Anforderungen_RDMO: Um die inhaltlichen und technischen Anforderungen der verschiedenen Stakeholder an das Tool zu dokumentieren, strukturieren und analysieren.

Dataset Sourcecode and Dokumentation: Entwicklung des RDMO

What types and formats of data will the project generate/collect?

Dataset Anforderungen_FDMP: docx und pdf XML, HTML text format for readme, protocols etc.

Dataset Anforderungen_RDMO: docx und pdf XML, HTML text format for readme, protocols etc.

Dataset Sourcecode and Dokumentation: text formats for readme, protocols, source code

Will you re-use any existing data and how?

Dataset Anforderungen_FDMP: Nachgenutzt

Dataset Anforderungen_RDMO: Erzeugt

Dataset Sourcecode and Dokumentation: Nachgenutzt

What is the origin of the data?

Dataset Anforderungen_FDMP: Die Daten wurden im Vorgängerprojekt "FDMP-Werkzeug" von Claudia Engelhardt und Jochen Klar erhoben bzw. gesammelt.

Snapshots

[→ Aktuell](#)

[Zwischenbericht 2018](#)

[DFG Antrag](#)

Optionen

[Zurück zum Projekt](#)

Export

[PDF](#)

[Rich Text Format](#)

[Open Office](#)

[Microsoft Office](#)

[HTML](#)

[Markdown](#)




[mediawiki](#)





[LaTeX](#)




Management




RDMO Demo Management Admin Sprache Jochen Klar





Fragen




Abschnitt Allgemein `rdmo/general`   




Fragenset Thema `rdmo/general/topic-research_question`    





Frage Wie lautet die primäre Forschungsfrage des Projektes?
`rdmo/general/topic-research_question/title` → `project/research_question/title`   

Frage Bitte geben Sie einige Schlagworte zur Forschungsfrage an.
`rdmo/general/topic-research_question/keywords` → `project/research_question/keywords`   

Fragenset Disziplin `rdmo/general/topic-research_field`    

Frage Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen?
`rdmo/general/topic-research_field/research_field` → `project/research_field/title`   

Frage Welche Personen oder Institutionen sind verantwortlich für die Projektkoordination?
`rdmo/general/topic-research_field/name` → `project/coordination/name`   

Fragenset Projektablauf `rdmo/general/project-schedule-schedule`    

Katalog

RDMO

Filter

Optionen

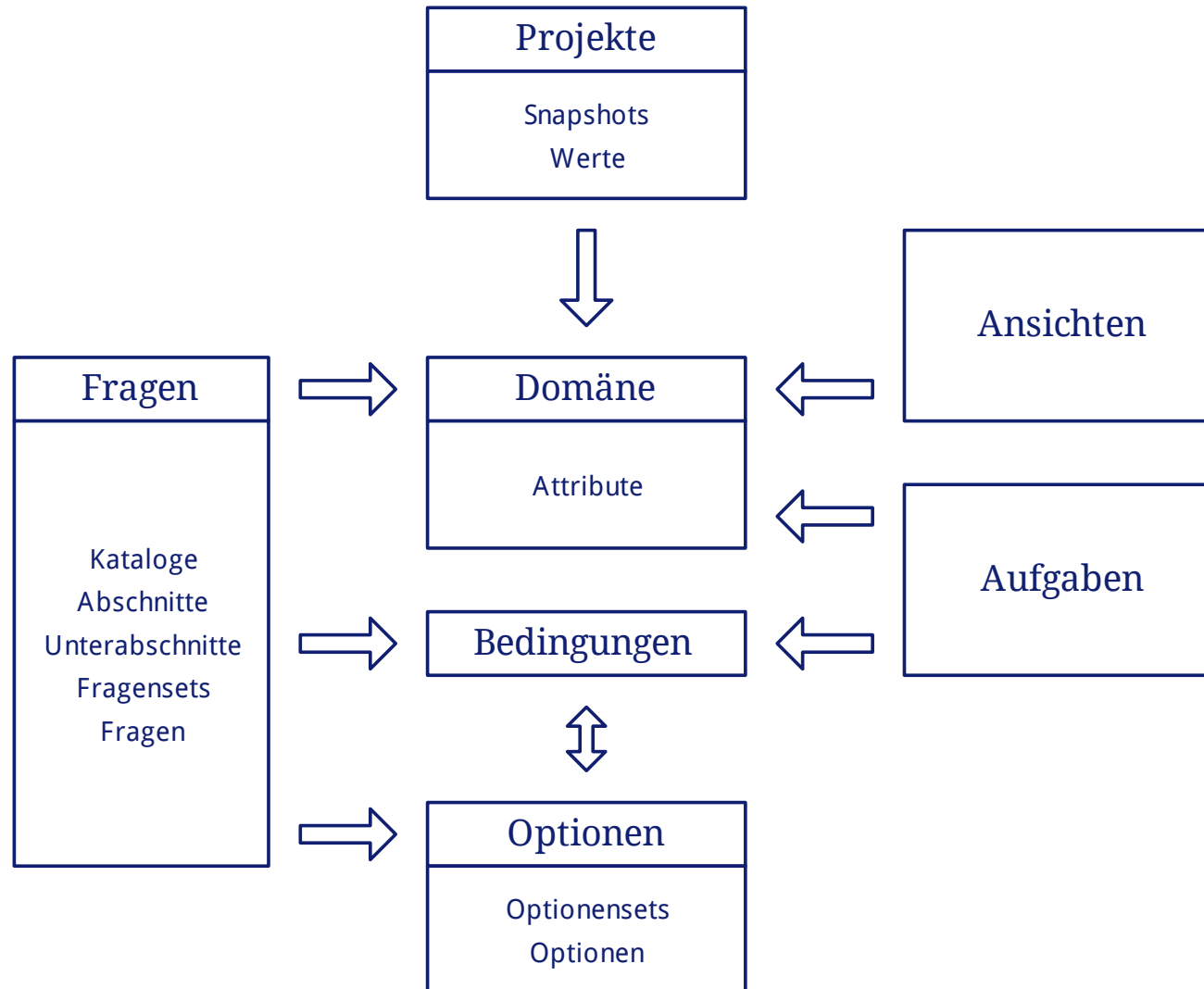
- Katalogeigenschaften bearbeiten
- Katalog entfernen
- Neuen Katalog erstellen
- Neuen Abschnitt erstellen
- Neues Fragenset erstellen
- Neue Frage erstellen

Export

- PDF
- Rich Text Format
- Open Office

14 / 19

RDMO Architektur



Domänenmodell

- Internes Vokabular für die RDMO gespeicherten Eingaben
 - `project`
 - `project/dataset`
 - `project/dataset/creator`
- Verbindet Fragen, Antworten (Werte), Ansichten, Aufgaben, API etc.
- Motiviert aus [Kerndatensatz Forschung](#), [CERIF](#), [CASRAI](#), [FOAF](#)
- Gewachsen mit der Erstellung neuer Fragenkataloge für RDMO
- Derzeitige Überarbeitung
- Zusammenarbeit mit [Research Data Alliance DMP Common Standards WG](#)

XML Export/Import Format

```
<rdmo xmlns:dc="http://purl.org/dc/elements/1.1/">
  <catalog dc:uri="https://rdmorganiser.github.io/terms/questions/rdmo">
    <dc:comment/>
    <order>1</order>
    <title lang="en">RDMO</title>
    <title lang="de">RDMO</title>
  </catalog>
  <section dc:uri="https://rdmorganiser.github.io/terms/questions/rdmo/general">
    <dc:comment/>
    <order>0</order>
    <title lang="en">General</title>
    <title lang="de">Allgemein</title>
  </section>
  ...
</rdmo>
```

- URI zur Identifikation einzelner Elemente
- GitHub Repository zum Austausch: <https://github.com/rdmorganiser/rdmo-catalog>

Programmierbare JSON API

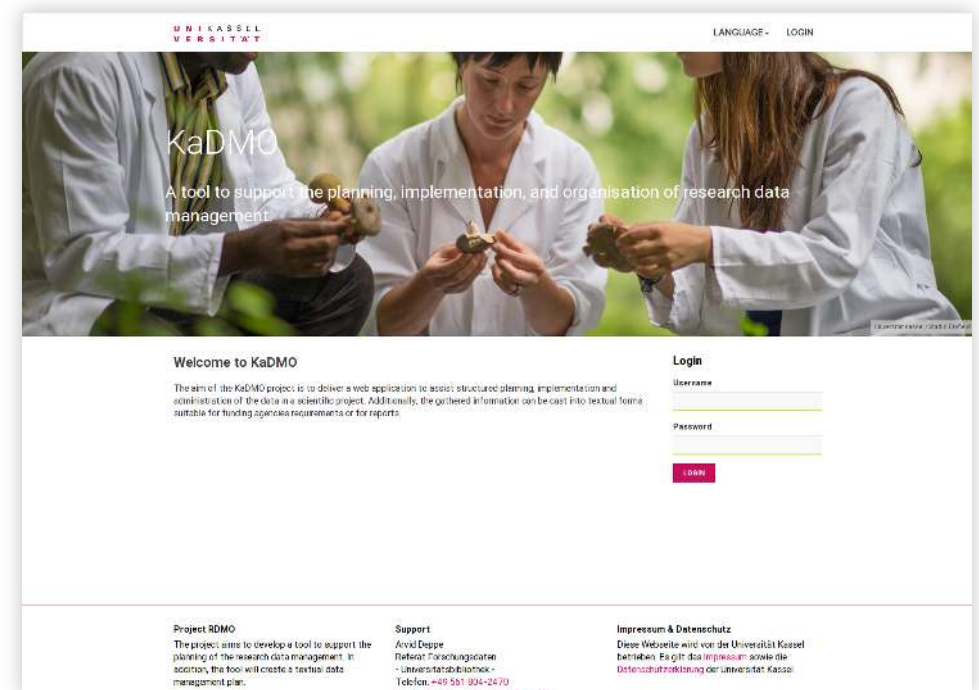
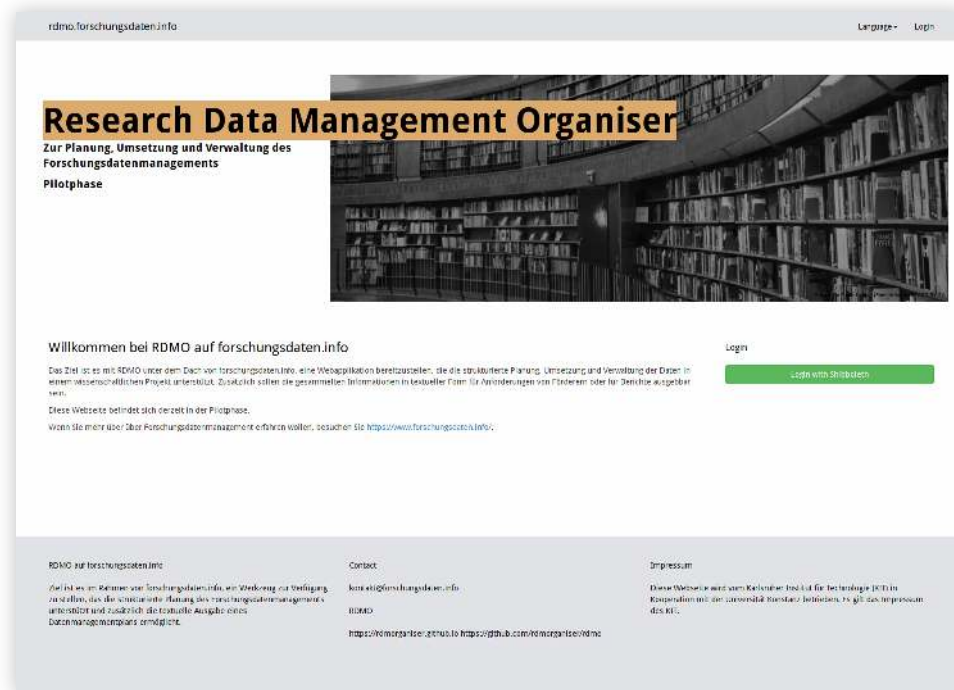
```
curl -X GET -H 'Authorization: Token oojoh3phaighaebiNeiyeeCeiY3Peuv2eitoojoh' \  
https://rdmo.aip.de/api/v1/projects/values/?attribute__path=project/dataset/size/volume
```

```
[  
  {  
    "id":10061,  
    "project":"https://rdmo.aip.de/api/v1/projects/projects/69/",  
    "attribute":"https://rdmo.aip.de/api/v1/domain/attributes/262/",  
    "set_index":0,  
    "collection_index":0,  
    "text":"","  
    "option":null,  
    "created":"2017-05-29T14:50:20.009917Z",  
    "updated":"2017-05-29T14:50:20.009924Z"  
  },  
  ...  
]
```

- Integration in lokale Workflows (Repositorien, FIS, Speicherinfrastruktur, etc.)

Themes

- Lokale Anpassungen für eigenes *Look and Feel*
- Kein Fork des **rdmo** Repositories nötig
- Jedes HTML-Template und jede CSS Datei kann überschrieben werden



RDMO Use Case

Folien von Robin Rothe



RDMO4Life im Projekt EmiMin

Vorstellung des Projektes „Emissionsminderung Nutztierhaltung“



KTBL



<https://www.ktbl.de/themen/emimin/>



C | A | U

Christian-Albrechts-Universität zu Kiel
Agrar- und
Ernährungswissenschaftliche Fakultät
ILV – Institut für
Landwirtschaftliche Verfahrenstechnik



**UNIVERSITÄT
HOHENHEIM**



Vorstellung des Projektes „Emissionsminderung Nutztierhaltung“

- ▶ Förderprogramm des Zweckvermögens des Bundes bei der Landwirtschaftlichen Rentenbank
- ▶ Projektdauer: 60 Monate (1. Phase 36 Monate; 2. Phase 24 Monate)
- ▶ Planzeitraum: 01.07.2018 bis 30.06.2023

Ziele

- Entwicklung und Validierung von Messmethoden mit künstlichen Tracergasen für die Untersuchungen in freigelüfteten Ställen und Ausläufen
- Untersuchung der Wirksamkeit von Maßnahmen und Maßnahmenkombinationen im Hinblick auf die Reduktion von Ammoniak-Emissionen, Geruch und Methan
- Ableitung von Emissionsfaktoren und Minderungsgraden für Ammoniak, Geruch und Methan auf Grundlage der Analyse der Projektdaten sowie von Berechnungsmodellen
- Aufbau einer Forschungsdatenbank zur Bereitstellung von Messdaten, Berechnungsergebnissen und Berechnungsmodellen sowie deren Metadaten
- Aufbau eines projektbegleitenden Datenmanagementplans zur Evaluation und Planung zukünftiger Vorhaben in der Agrartechnik



Verbundvorhaben
Emissionsminderung
Nutztierhaltung (EmiMin)

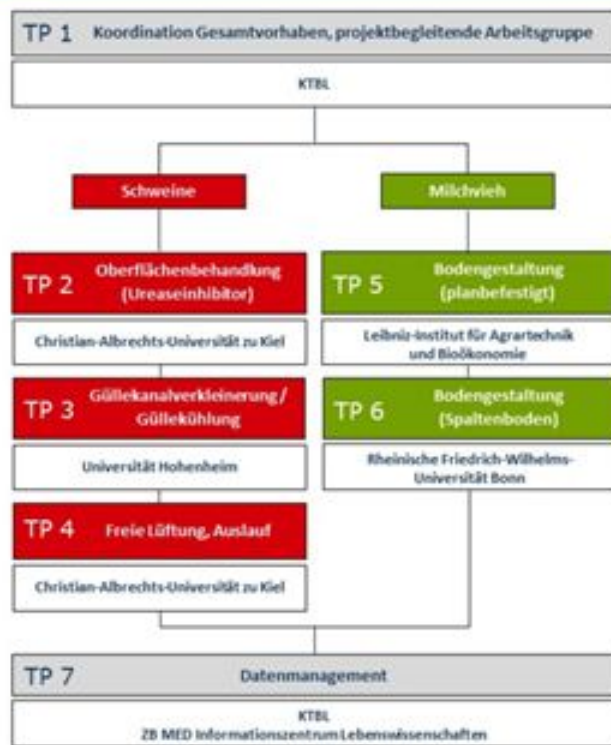


Vorstellung des Projektes „Emissionsminderung Nutztierhaltung“



„Die erzeugten Forschungsdaten werden von Projektbeginn an in einen Managementprozess einbezogen, so dass die publizierbaren Emissionsdaten im Fachrepositorium Lebenswissenschaften (FRL) veröffentlicht werden. Hierfür wird ein dynamischer, projektbegleitender Datenmanagementplan erstellt und der Veröffentlichungsprozess inklusive der notwendigen Datenaufbereitung durchgeführt. Das FRL wird um die zusätzlich erforderlichen Strukturen zur Datenhaltung und Datenbeschreibung erweitert.“

Quelle: EmiMin (2018). *Gemeinsame Vorhabensbeschreibung*. S. 4. Nicht veröffentlicht



TP = Teilprojekt



Christian-Albrechts-Universität zu Kiel



- ▶ Dynamischer, projektbegleitender Datenmanagementplan DMP (gesamt und für Teilprojekte)
- ▶ Überblick zur Datenverfügbarkeit im Verbundprojekt
- ▶ Anpassung des Tools an fachliche, projektspezifische Erfordernisse

- ▶ Workshops mit Verbundpartnern
- ▶ Abstimmung zum Forschungsdatenmanagement und DMP
- ▶ Abstimmung zu projektspezifischen Metadaten (für Projektdatenbank und mögliche Publikation)
- ▶ Abstimmung über zu publizierende Forschungsdaten

- ▶ Fachrepositorium Lebenswissenschaften: Erweiterung des Metadatenschemas (techn. Umsetzung durch hbz Hochschulbibliothekszentrum NRW)
- ▶ Publikation der Forschungsdaten

Aufbau von RDMO4Life

- ▶ Projektbegleitendes Datenmanagement
- ▶ Erfassung der Anforderungen aus der agrartechnischen Forschung anhand eines Anwendungsfall
- ▶ Anpassung des generischen Tools an Projektspezifika
- ▶ 1. Ziel: Bereitstellung eines an die Agrartechnik angepassten RDMO
- ▶ 2. Ziel: Erweiterung von RDMO4Life an die Agrarwissenschaften insgesamt

Meine Projekte

Name	Rolle
Christian Albrechts Universität zu Kiel, Institut für Landwirtschaftliche Verfahrenstechnik	Besitzer
Leibniz Institut für Agrartechnik und Bioökonomie e.V. (ATB), Abteilung Technik in der Tierhaltung	Besitzer
Projektkoordination	Besitzer
Universität Bonn, Institut für Landtechnik, Verfahrenstechnik in der tierischen Erzeugung	Besitzer
Universität Hohenheim, Institut für Agrartechnik, Fachgebiet Verfahrenstechnik der Tierhaltungssysteme	Besitzer

Eine Anwendung von



Bisheriger Ablauf

- ▶ Betrachtung des generischen RDMO-DMP
- ▶ Aufbau von 2 Fragenkatalogen mit ersten Spezifika inkl. Aufteilung von Fragen
- ▶ Befragung der Projektverantwortlichen vor Ort
- ▶ Besichtigung der Forschungsstätten
- ▶ Einarbeitung der Antworten in RDMO4Life inkl. Anpassung der Kataloge
- ▶ Überlegungen zu weiteren Fragenkomplexen

Katalog

EmiMin V1.0

EmiMin V1.0

EmiMin Leitung V1.0

RDMO

Unterabschnitt Weitere Beteiligte `emimin/general/additional_parties`

Fragenset `emimin/general/additional_parties/facilities_quantity` → `facilities_quantity`

Frage `emimin/general/additional_parties/facilities_quantity/facilities_quantity` → `facilities_quantity/quantity`

Wieviele Betriebe werden im Projekt von der Einrichtung betreut?

Frage `emimin/general/additional_parties/facilities_quantity/facilities_name` → `facilities_quantity/name`

Welche Betriebe mit welcher Versuchsanordnung sind angebunden?

Frage `emimin/general/additional_parties/facilities_quantity/species` → `facilities_quantity/species`

Welche Tierart wird in den Betrieben gehalten?

Frage `emimin/general/additional_parties/facilities_quantity/contact_person` → `facilities_quantity/facility-contact`

Wer ist dort verantwortlich für das Projekt (Kontaktperson)?

Frage `emimin/general/additional_parties/facilities_quantity/responsible_facility_communication` → `facilities_quantity/project-contact`

Wer ist verantwortlich für die Organisation der Kommunikation mit den Betrieben?



(Zwischen -) Fazit

- ▶ RDMO bietet durch Open Source & Community-Driven-Development gute Voraussetzungen
- ▶ Eine ergebnisoffene Herangehensweise macht Arbeit, aber es lohnt sich
- ▶ RDMO entfaltet das wirkliche Potenzial erst bei flexibler Gestaltung
- ▶ (Drittmittel-) Projekte sind sehr wichtig, um Ressourcen zu schaffen
- ▶ DMP sind ein sehr gutes Mittel, um transdisziplinär „ins Gespräch zu kommen“
- ▶ Infrastruktur und Forschung müssen sich gegenseitig verstehen lernen
- ▶ Komfortzonen sollten (müssen) verlassen werden



Danke! Fragen?

Robin Rothe

Fachrepositorium Lebenswissenschaften

Forschungsdatenmanagement

Gleueler Straße 60

50931 Köln

rothe@zbmed.de



RDMO und FAIR

FAIR Data & Geisteswissenschaften & RDMO (1)

- **F**indable = DESCRIBE, e.g.
 - Persistent Identifiers
 - Rich Metadata
 - (Meta)data are registered or indexed in a searchable source
- **A**ccessible = OPEN, e.g.
 - (Meta)data are retrievable by their identifier using a standardised open communications protocol
 - Trusted Repositories
 - (Meta)data are accessible, even if data are not (anymore)
- **I**nteroperable = LINK, e.g.
 - Machine actionable APIs
 - FAIR (Meta)data vocabularies
- **R**eusable = PUBLISH, e.g.
 - Clear and accessible data usage licence
 - Rich documentation (provenance, etc.)
 - Meta(data) meet domain relevant community standards (reusable formats etc.)



<https://www.go-fair.org/fair-principles/>, EPFL Library Research Data Management FAST Guides:

https://researchdata.epfl.ch/wp-content/uploads/EPFL_Library_RDM_FastGuide_All.pdf, PARTHENOS Guidelines to FAIRify data management:

<https://doi.org/10.5281/zenodo.2668479>

FAIR Data & Geisteswissenschaften & RDMO (2)

- **F**indable = DESCRIBE, z. B.
 - Will persistent identifiers (PIDs) be used for this data set?
 - Which system of persistent identifiers shall be used?
 - Which standards, ontologies, classifications etc. are used to describe the data and context information?
- **A**ccessible = OPEN, z. B.
 - Is the repository or data centre chosen certified (e.g. Data Seal of Approval, nestor Seal or ISO 16363)?
 - If yes, under which terms of use or license will the dataset be published or shared?
- **I**nteroperable = LINK, z. B.
 - Is this dataset interoperable, i.e. allowing data exchange and re-use between researchers, institutions, organisations, countries etc.?
 - Are metadata and context information being checked for correctness and completeness?
 - Which file formats are used?
- **R**eusable = PUBLISH, z. B.
 - Where will the data (including metadata, documentation and, if applicable, relevant code) be stored or archived after the end of the project?
 - Which measures of quality assurance are taken for this dataset?

Vorteile des Forschungsdatenmanagements mit RDMO

- Datenhoheit durch eigene RDMO-Instanz
- Aktives Forschungsdatenmanagement
- Individuell anpassbar (Institution, Projekt, Disziplin)
- Standardisierte Vokabulare für das Forschungsdatenmanagement
- Zukunft: Unterstützung *actionable data management plan*
- Zukunft: Interoperabilität mit Nachweissystemen wie re3data und DataCite Metadata Schema
- Zukunft: RDMO-Metadatenschema

(Inter-)nationale Perspektive und Ausblick

GO FAIR

“GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the FAIR data principles, making data Findable, Accessible, Interoperable and Reusable.”

<https://www.go-fair.org/>



Current Implementation Networks

The GO FAIR Implementation Networks operate within diverse domains ranging from life science disciplines to social sciences and humanities. GO FAIR supports convergence of their efforts to implement a variety of elements of the Internet of FAIR Data and Services.

Learn more about current Implementation Networks



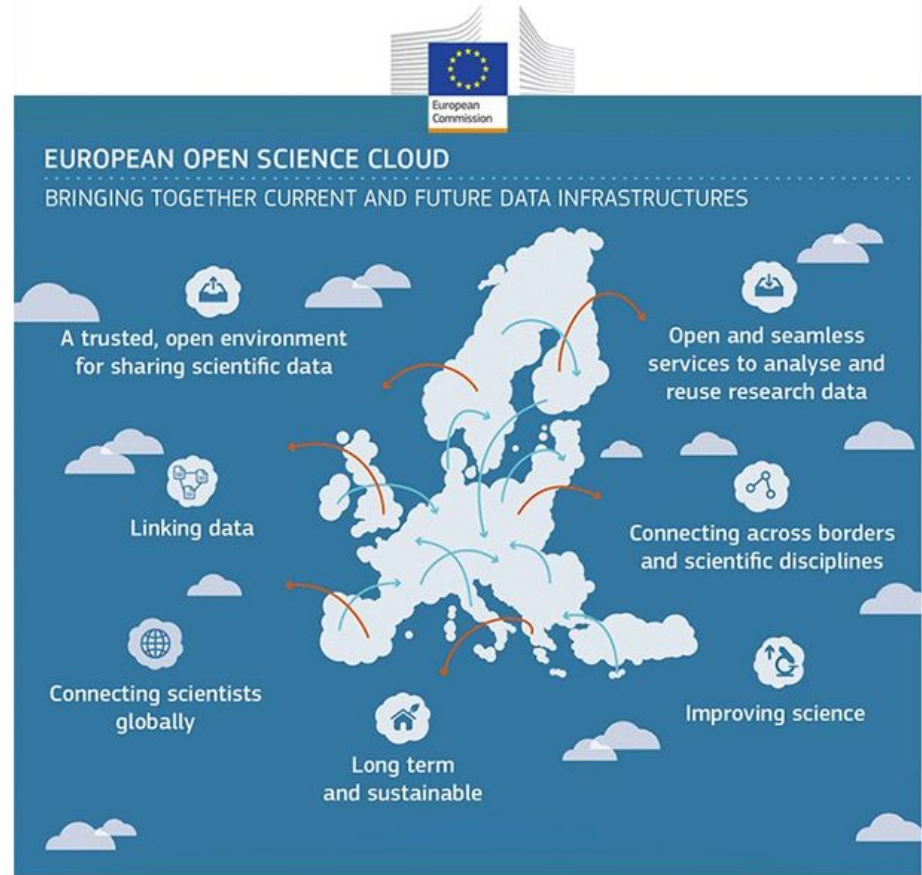
NFDI

- 2014 Digitale Agenda (Internationalisierungsstrategie für die Wissenschaft)
- Status Quo: kein flächendeckendes Angebot an Informationsinfrastrukturen für das Forschungsdatenmanagement und keine nationale Strategie
- 2019 Erste Ausschreibung für den Aufbau einer Nationalen Forschungsdateninfrastruktur (Finanzierung Bund und Länder im Zeitraum von 2019 bis 2028 mit bis zu 90 Millionen Euro pro Jahr)
- “In der NFDI sollen Datenbestände in einem aus der Wissenschaft getriebenen Prozess systematisch erschlossen, langfristig gesichert und entlang der **FAIR-Prinzipien** über Disziplinen- und Ländergrenzen hinaus zugänglich gemacht werden.”

EOSC

“a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines”

<https://www.eosc-portal.eu/about/eosc>



Source Picture: <https://library.ktu.edu/news/launch-of-the-european-open-science-cloud-eosc/>

Ausblick

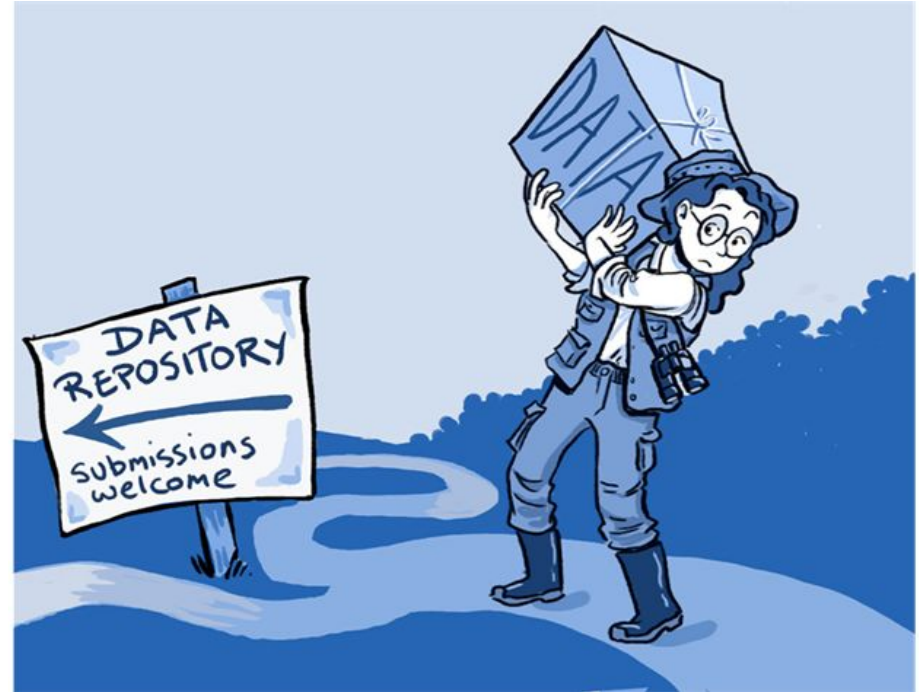


Picture: Michal Klajban [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

https://upload.wikimedia.org/wikipedia/commons/7/7c/A_pier_at_a_campsite_during_sunset%2C_Sidney_Spit_%28part_of_Gulf_Islands_National_Park_Reserve%29%2C_Sidney_Island%2C_British_Columbia%2C_Canada_20.jpg

How FAR is FAIR?

- Gutes Datenmanagement = Gute wissenschaftliche Praxis
- Grundkenntnisse FDM als Teil der Forschungsqualifikationen
- DMP = Systematisches digitales Forschungsmanagement bzw. “Datenperspektive”
- Ein DMP ist ein AKTIONSPLAN für FAIRe Daten und die Sicherung der eigenen Forschungsintegrität und Transparenz = Aktives Forschungsdatenmanagement während des gesamten Lebenszyklus
- Tools wie RDMO zur Unterstützung des **aktiven Datenmanagements**



“Das grundsätzliche logische Problem, dass nämlich a priori nicht entschieden werden kann, welche aus Forschungsprozessen stammende Daten für zukünftige Forschung relevant sein können und daher langfristig verfügbar sein sollen, kann keine Infrastruktur lösen. Hier muss jedes Forschungsprojekt jeweils das Optimum aus vertretbarem Aufwand und zukunftsöffener Angebotsleistung zu finden versuchen.”

Quelle: Michael Müller, Georg Schelbert, Nachbetrachtung zu einem Workshop zu Forschungsdaten in den Geisteswissenschaften, insbesondere bild- und objektbezogenen Wissenschaften, an der FU Berlin, 03.05.2019

[https://blog.arthistoricum.net/beitrag/2019/05/03/nachbetrachtung-zu-einem-workshop-zu-forschungsdaten-in-den-geisteswissenschaften-insbesondere-bild/](https://blog.arthistoricum.net/beitrag/2019/05/03/nachbetrachtung-zu-einem-workshop-zu-forschungsdaten-in-den-geisteswissenschaften-insbesondere-bild-)

Open und FAIRe Forschungsdaten sind in aller Munde....:

- “Um welche Daten welchen Typs aus welchem (institutionellen/fachlichen) Kontext geht es?
- Welche technischen Infrastrukturen werden für die Veröffentlichung genutzt?
- Wer sind die (potentiell) Nutzenden dieser Daten bzw. welche Nutzungszwecke sind zu erwarten?
- Welche Anreize für die Veröffentlichung gibt es ggf. über die Nutzung durch Dritte hinaus?
- Wie werden die Daten sichtbar und durchsuchbar gemacht? Wie sind sie mit anderen digitalen Ressourcen verknüpft?
- Welche Lizenzen werden genutzt?
- Welche Maßnahmen zur Qualitätssicherung werden angewendet?”

Wie FAIR sind Ihre Daten?

Kontakt

Ulrike Wuttke

Twitter: [@uwuttke](https://twitter.com/uwuttke)

ORCID: <https://orcid.org/0000-0002-8217-4025>

Jochen Klar

Twitter: [@jochenklar](https://twitter.com/jochenklar)

ORCID: <https://orcid.org/0000-0002-5883-4273>

- [RDMO Webseite](#)
- [RDMO Testinstanz](#)