Entwickeln und Implementieren von Datenmanagementplänen für ingenieurswissenschaftliche Projekte

Soo-Yon Kim, Steffen Hillemacher, Sabine Schönau und Daniela Hausen

03. März 2022

NFDi4ing

What is a Data Management Plan (DMP)?

A DMP is a formal and at the same time living document.

"In a DMP a researcher or research team describes **what** data goes into a project (reuse) and comes out of it (potential reuse), **How** the team takes care of the data, and **Who** is allowed to do **What** with the data **When**."

(source: https://www.openaire.eu/blogs/why-is-this-a-good-data-management-plan)

Benefits of a DMP

Supports research

Supports FAIR

Saves time in the long term

Offers continuity

Helps consider issues related to confidentiality, ethics, security and copyright

Elements of a DMP

Legal requirements

Intellectual property rights

Budget

Data organisation

Data description

Ethics and privacy

Existing data

Archiving and preservation

Format

Audience

Metadata

(source: Creative Commons Attribution-ShareAlike 4.0 International License, CESSDA ERIC)

Access and sharing

Storage and backup

Quality Assurance

Responsibility

Security

Selection and retention periods

Requirements of Research Funders

DFG

1.) Data description
2.) Documentation and data quality
3.) Storage and technical archiving the project
4.) Legal obligations and conditions
5.) Data exchange and long-term data accessibility
6.) Responsibilies and resources
(source: https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html)

Passed in June 2021

BMBF

"Independent of the economic prospects of success the scientific and/or technical prospects for success prospects of success should be presented (with time horizon) -including how the planned results can be used in other ways (e.g. for public tasks, databases, networks, transfer offices, etc.). At this cooperation with other institutions, companies, networks, research with other institutions, companies, networks, research etc. should also be included."
(source: Bundesministerium für Bildung und Forschung: *Richtlinien für Zuwendungsanträge auf Ausgabenbasis -AZA*.2013. Online verfügbar: http://foerderportal.bund.de/easy/easy_index.php?auswahl=easy_formulare&formularschrank=bmbf#t1)
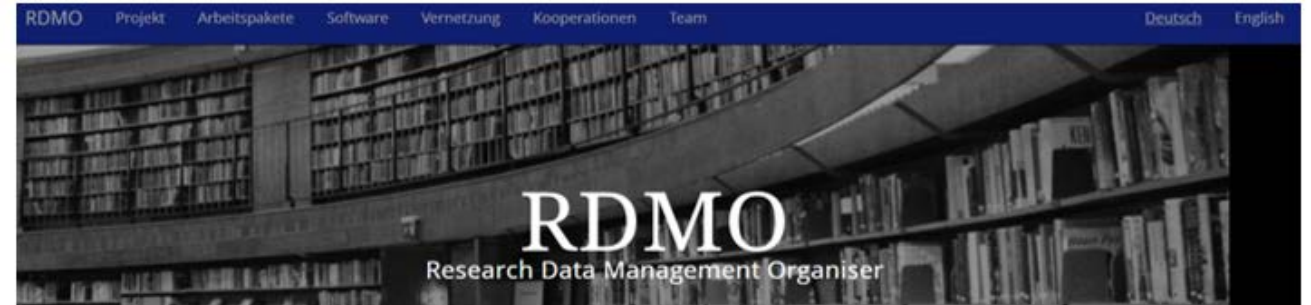
EU

Mandatory – 6 months, middle and end of project
- Data summary
- **FAIR data**
- Other research outputs
- Allocation of resources
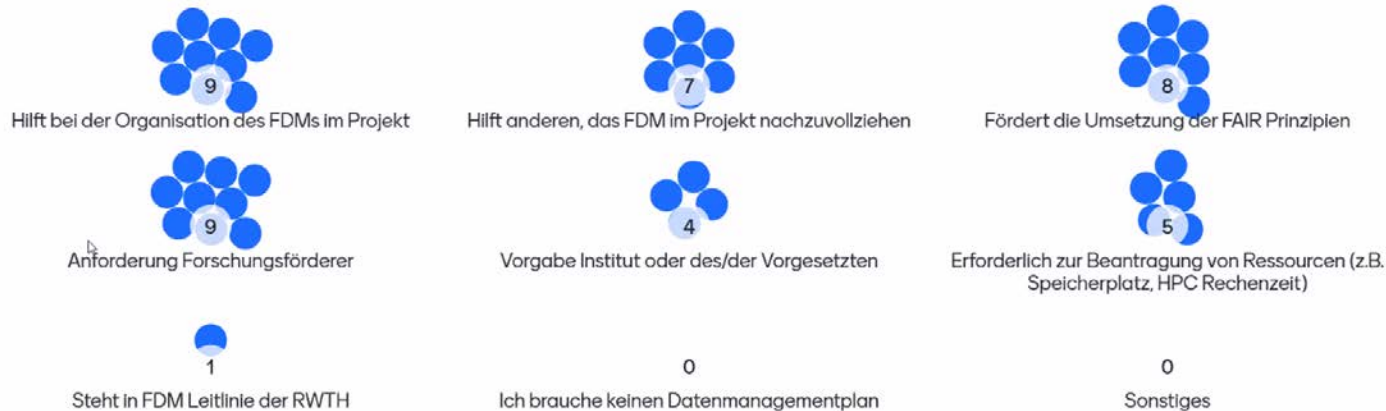- Data security
- Ethics
- Other issues
Template: https://bit.ly/34bCx9N

Research Data Management Organiser (RDMO)

- Tool for creating data management plans
- Support of structured planning of your research data management
- Can be used for collaborative working

Was motiviert Sie, einen Datenmanagementplan zu erstellen?

- Hilft bei der Organisation des FDMs im Projekt — 9
- Hilft anderen, das FDM im Projekt nachzuvollziehen — 7
- Fördert die Umsetzung der FAIR Prinzipien — 8
- Anforderung Forschungsförderer — 9
- Vorgabe Institut oder des/der Vorgesetzten — 4
- Erforderlich zur Beantragung von Ressourcen (z.B. Speicherplatz, HPC Rechenzeit) — 5
- Steht in FDM Leitlinie der RWTH — 1
- Ich brauche keinen Datenmanagementplan — 0
- Sonstiges — 0

Research Data Management Organiser (RDMO)

Folgen Sie bei der Datendokumentation spezifischen Standards, Vorgaben oder Vereinbarungen?

Abgestimmte oder standardisierte Metadatenschemata und Vokabulare erhöhen die Interoperabilität Ihrer Daten mit anderen. Auch wenn es für viele Forschungsdaten noch keine Standards gibt, kann es sinnvoll sein, sich an einem anderen Standard zu orientieren.
Einen Überblick über Metadatenschemata diverser Fächerkulturen bietet z. B. www.fairsharing.org. Im Rahmen des Projektes DIPL-ING wurde das Metadatenschema EngMeta zur Beschreibung von Forschungsdaten aus den (computergestützten) Ingenieurwissenschaften entwickelt.
Wenn Sie keinen Vereinbarungen oder Standards folgen, können sie dieses Feld auch unbeantwortet lassen.

Bitte nutzen Sie für die Eintraege jeweils eine eigene Zeile. Sie können weitere Eintraege mit dem grünen Button hinzufügen und mit dem blauen Kreuz (×) entfernen.

EINTRAG HINZUFÜGEN

Wo werden die Metadaten erfasst und wie wird die Verbindung zu den Daten sichergestellt?

☐ In der Datei selbst:

☐ Text-Datei:

☐ Tabelle:

☐ Datenbank:

☐ In zugehörigen Publikationen (im finalen DMP Zitat/DOI ergänzen):

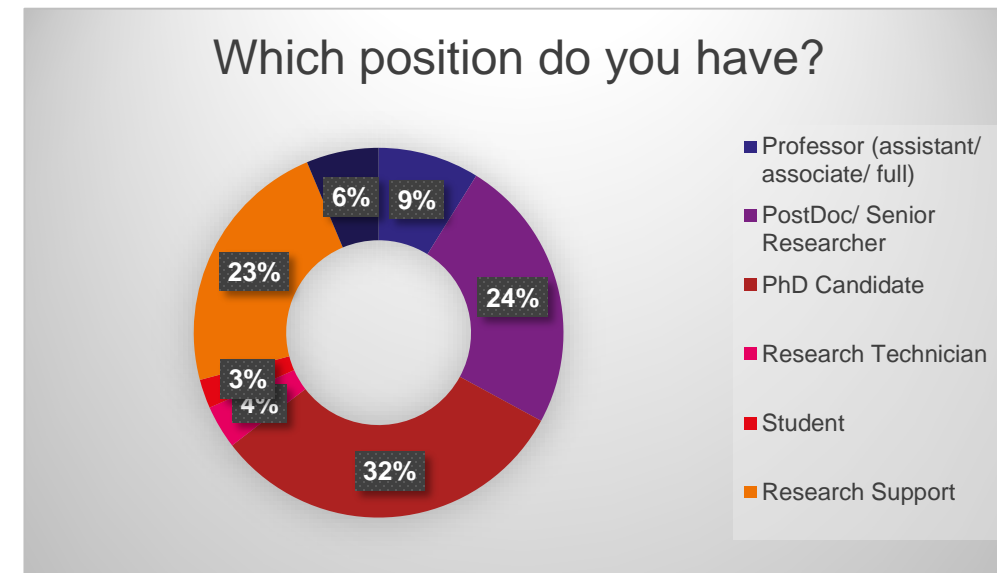☐ Sonstiges:

ZURÜCK   ÜBERSPRINGEN   SICHERN   SICHERN UND FORTFAHREN

- Webseite: rdmorganiser.github.io
- Code: github.com/rdmorganiser/rdmo
- Fragenkatalog: github.com/rdmorganiser/rdmo-catalog
- Dokumentation: rdmo.readthedocs.io
- Mailingliste: rdmo@listserv.dfn.de
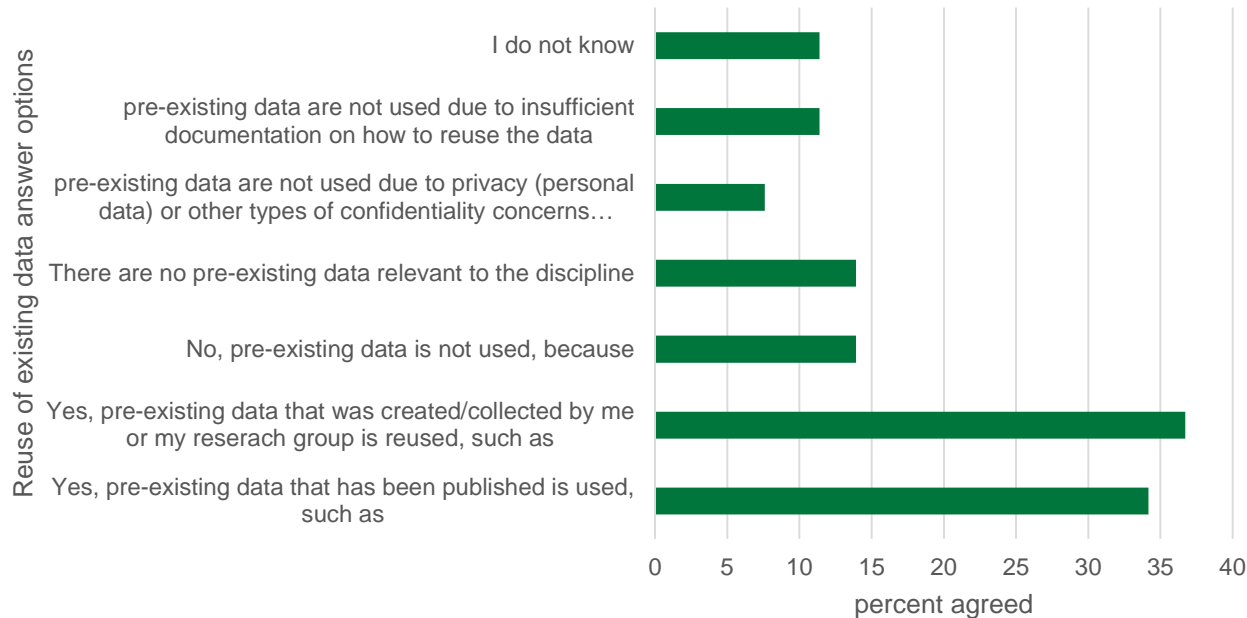- Twitter: @rdmorganiser
- Slack: rdmo.slack.com

Online survey on discipline-specific aspects in DMP

- Survey from 11th Oct. 2021 – 14th Jan. 2022
- 358 responses -> 79 from engineering
- 21 questions about
  - Demographics
  - Data Description and Collection
  - Data Documentation and Quality
  - Active Data Storage and Back-up
  - Data Archiving, Publishing and Sharing
  - Guidelines, Principles and Best Practices

Which position do you have?

- Professor (assistant/ associate/ full)
- PostDoc/ Senior Researcher
- PhD Candidate
- Research Technician
- Student
- Research Support

9% 24% 32% 4% 3% 23% 6%

Data Description and Collection

Do you or the researchers you support ever re-use existing data?

Reuse of existing data answer options (y-axis), percent agreed (x-axis)

- I do not know
- pre-existing data are not used due to insufficient documentation on how to reuse the data
- pre-existing data are not used due to privacy (personal data) or other types of confidentiality concerns…
- There are no pre-existing data relevant to the discipline
- No, pre-existing data is not used, because
- Yes, pre-existing data that was created/collected by me or my reserach group is reused, such as
- Yes, pre-existing data that has been published is used, such as

Reuse of published data

- weather data
- 3d-Models
- Voxeldata
- some flight controller safety scenarios
- KITTI Dataset
- Open Street Map Data, Remote Sensing Data
- lab experiments
- results of lab experiments
- models like FEM outputs or model parameters for standard models
- Comparison with current studies
- Materials Project, Atomwork Adv., MAGNDATA, ICSD
- Databases such as OQMD, ICSD, materials project, atomwork adv.

Data Archiving, Publishing and Sharing

Do you or the researchers you support publicly/openly share data/code for the long-term (at least 10 years) after the project?
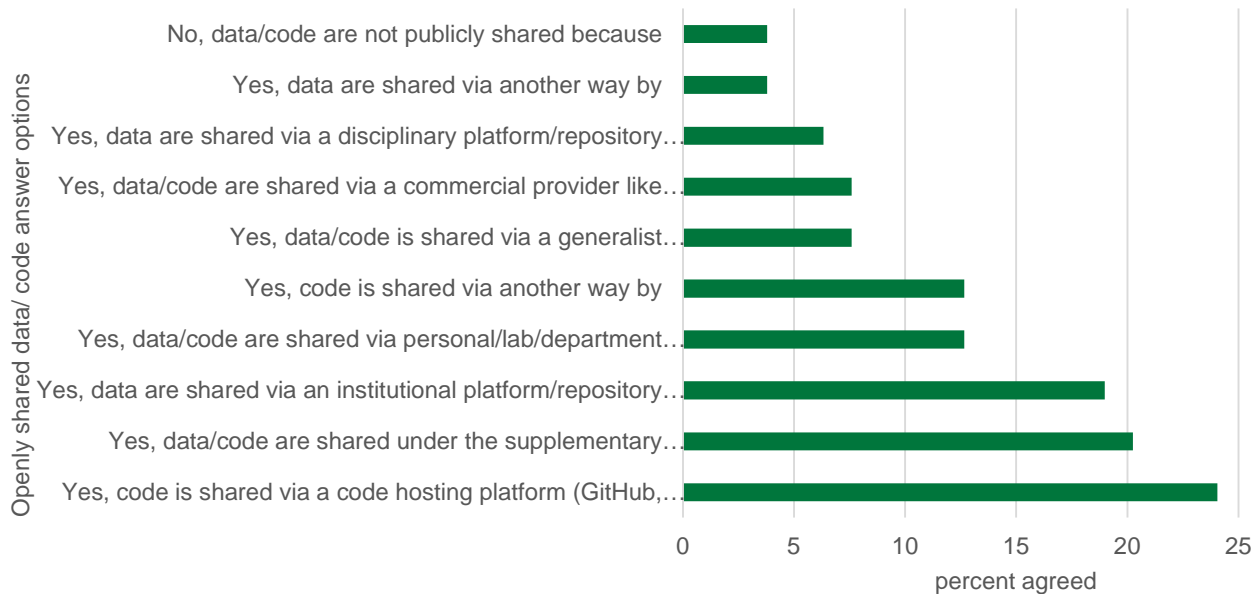
Discipline-specific repositories
- NOMAD
- NOMAD Oasis
- Openagrar
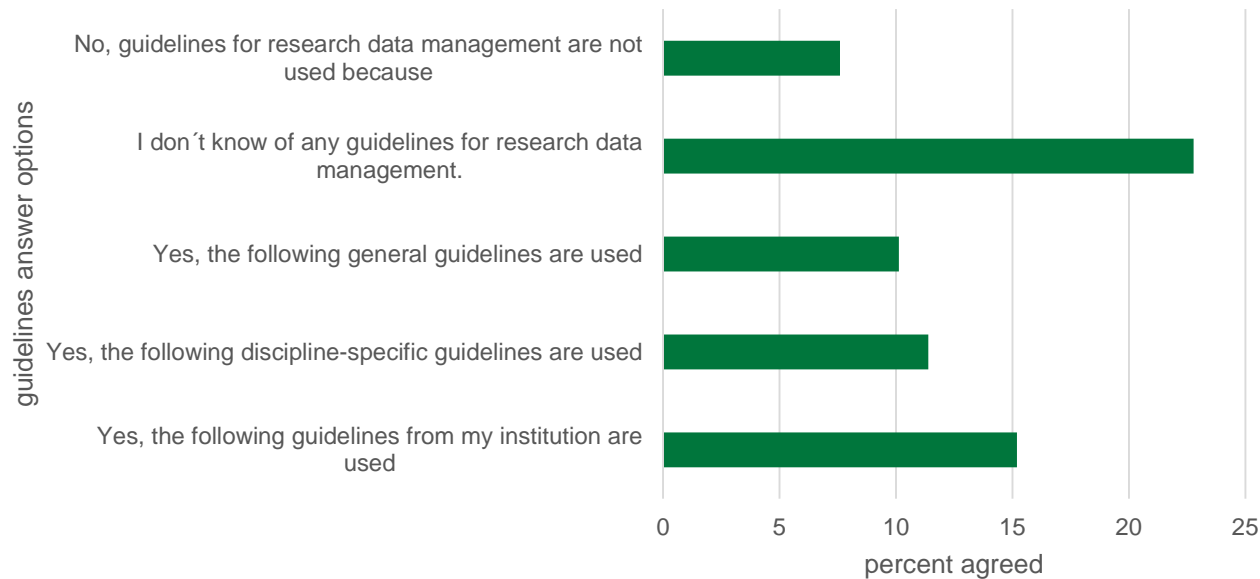- National Center for Biotechnology Information (NCBI)

Code hosting platform
- GitHub
- GitLab (hosted by RWTH or self-hosted)

General repositories
- Zenodo
- OSF

Guidelines, Principles and Best Practices

Do you or the researchers you support make use of any guidelines for research data management?

(horizontal bar chart)

guidelines answer options (y-axis):
- No, guidelines for research data management are not used because — ~7.5
- I don´t know of any guidelines for research data management. — ~23
- Yes, the following general guidelines are used — ~10
- Yes, the following discipline-specific guidelines are used — ~11.5
- Yes, the following guidelines from my institution are used — ~15

x-axis: percent agreed (0, 5, 10, 15, 20, 25)

Discipline-specific guidelines
- No engineering-specific guideline

Institutional guidelines
- RDM guideline
- Shared best practice examples from colleagues
- Internal guideline

General guidelines
- Funders guidelines
  - DFG
  - EU
- FAIR
- Institutional guidelines

Cluster of Excellence "Internet of Production"
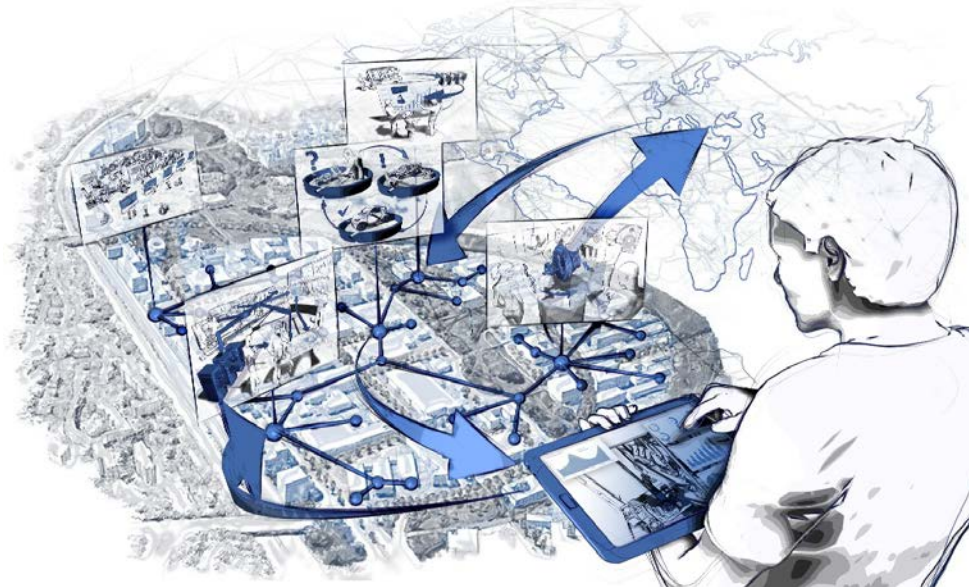
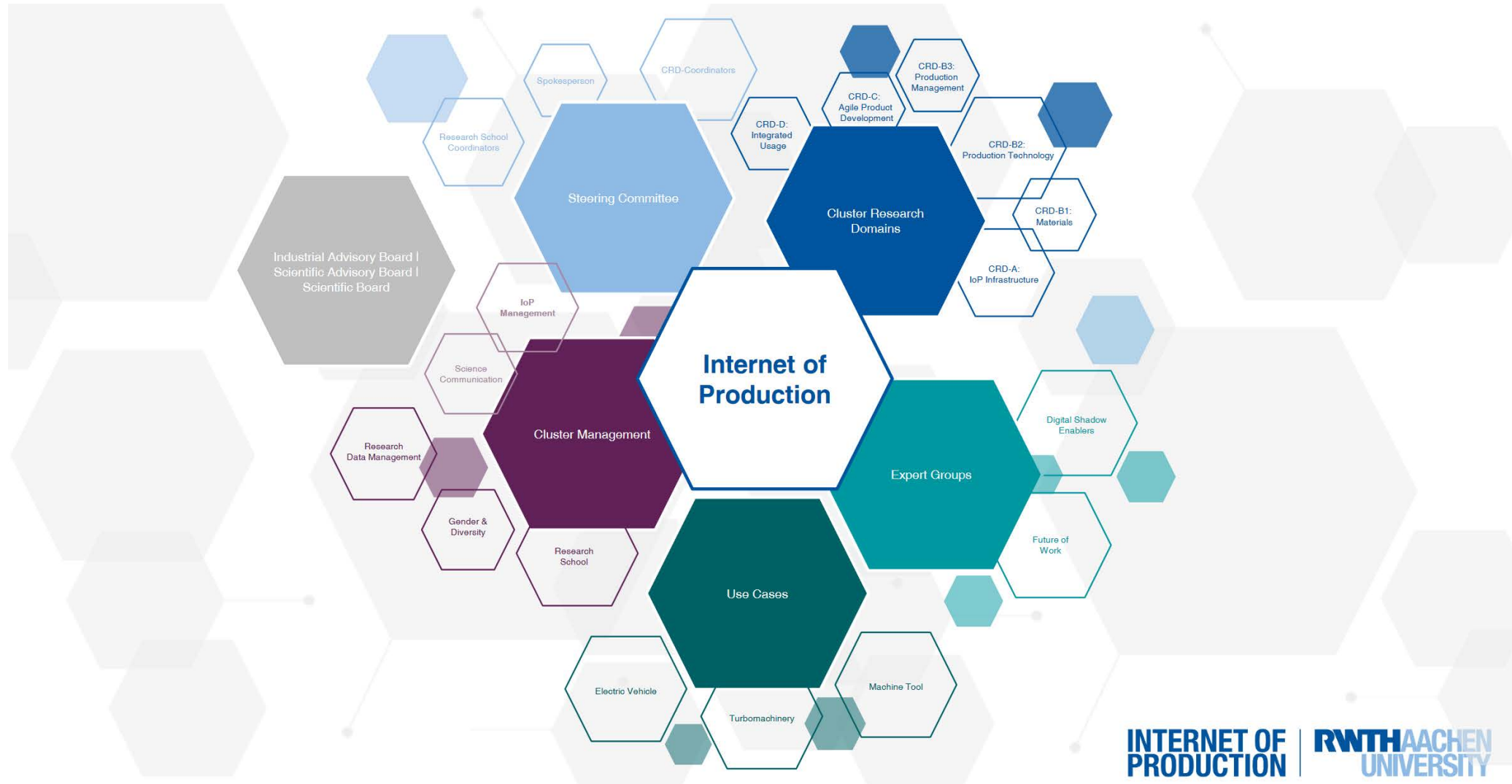**General information**

- CoE at RWTH since 2019
- Successor of CoE "Integrative Production Technology for High-Wage Countries"
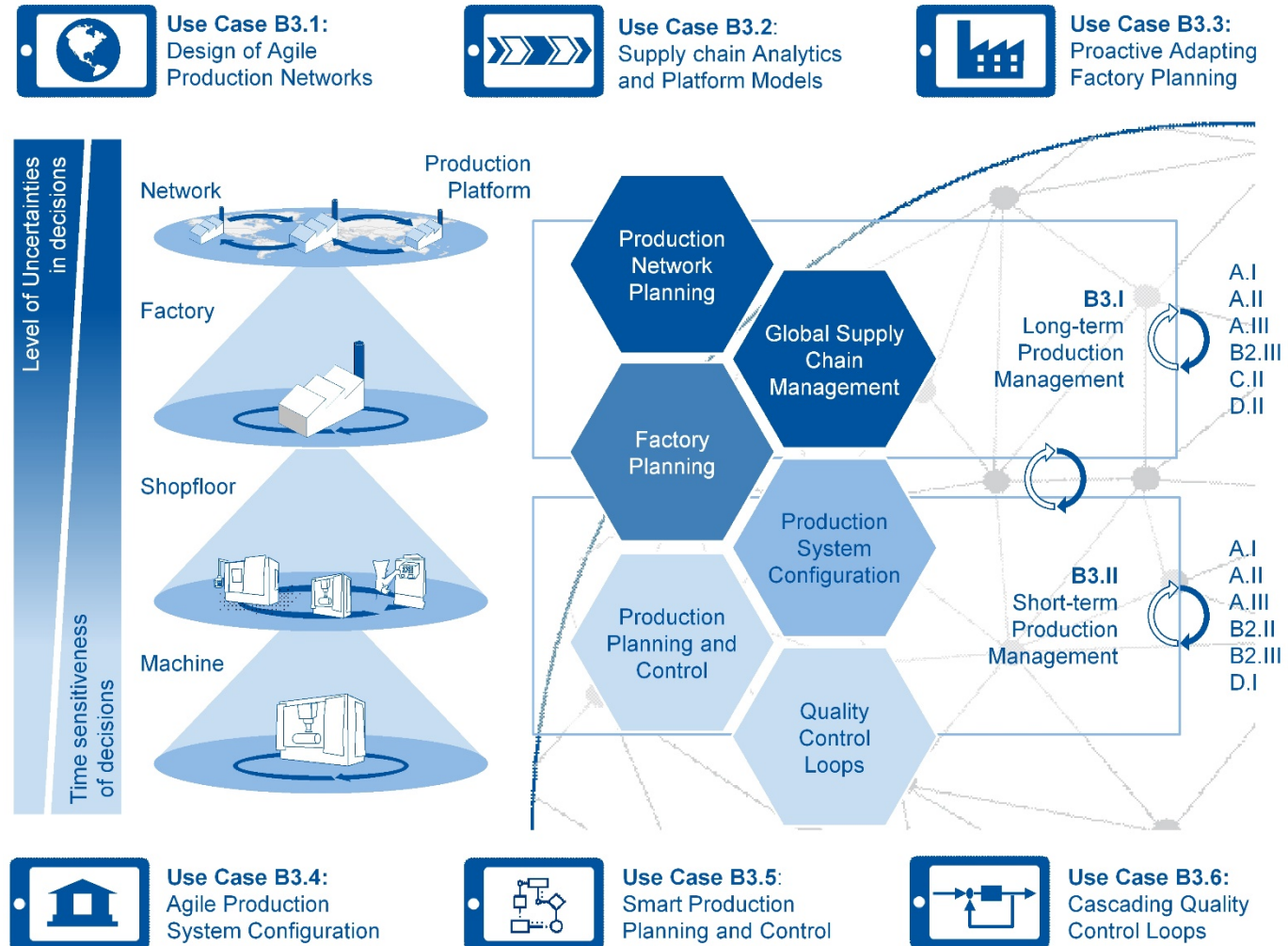- Data-driven industrial production

Scientific core concept

- Digital shadow
  - As opposed to digital twin
  - Abstracted models
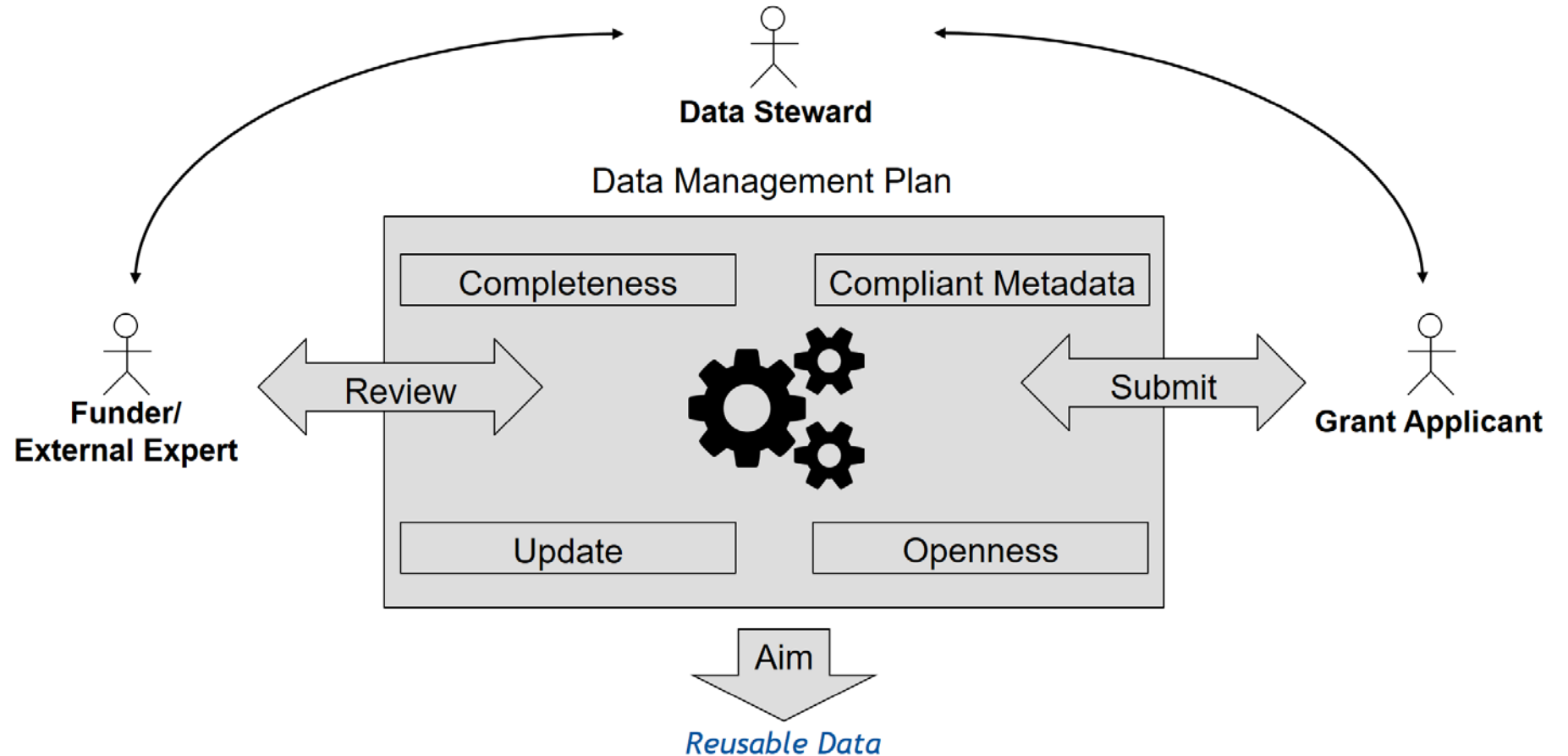 - Real-time
  - Cross-domain, interdisciplinary
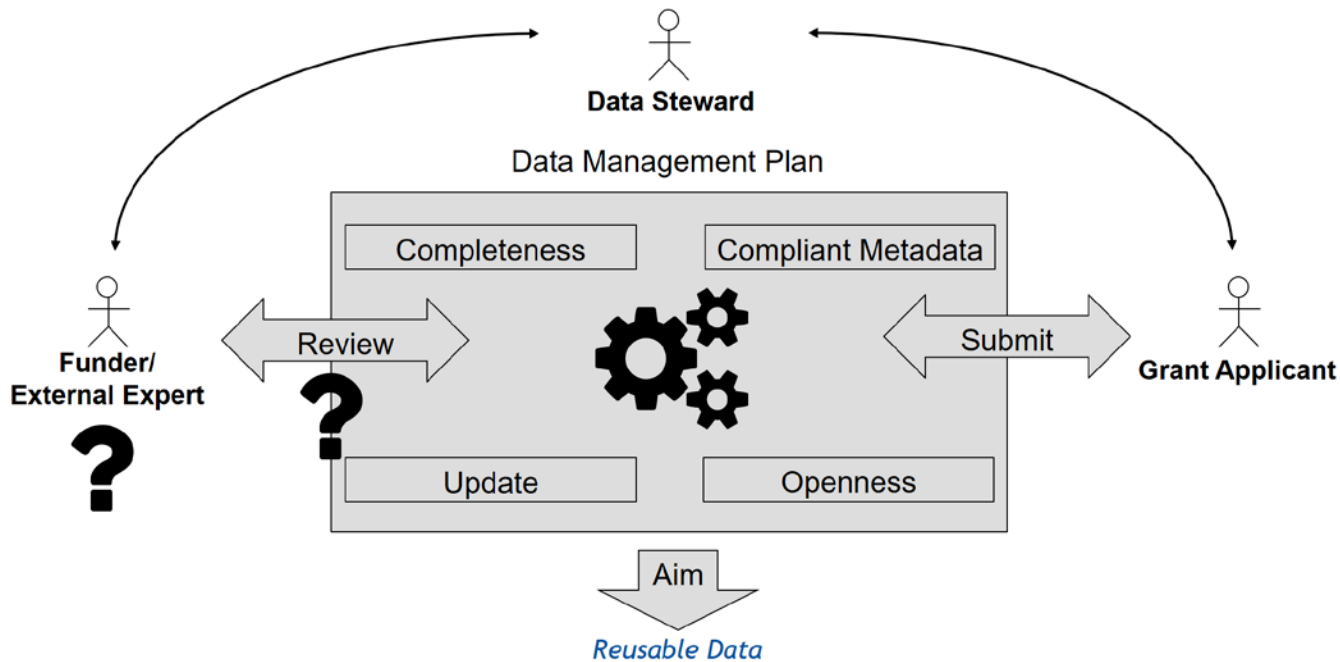
Structure of the CoE

Structure of the CoE – CRD-B3

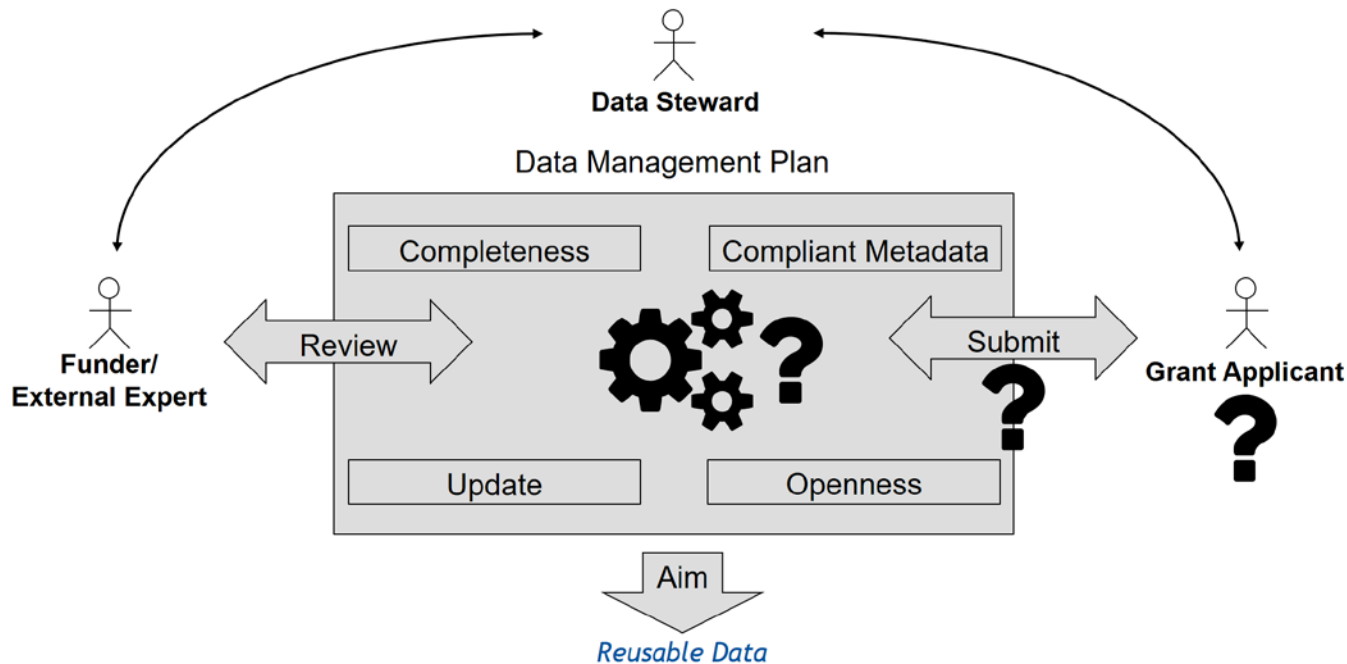# Data Management Plan Grant Application Context

Reviewer

Reviewer's role/required capability
- Raising awareness
- Ensuring data quality

Possible reviewers
- Funders?
  - E.g. DFG, …
- Cluster management?
- Project leads?
- Other researchers?
  - E.g. dual control principle
- Data stewards?
- No one?
- …

Integrating DMP work into project work

- Researchers are supposed to fill in the DMPs
- Hurdles impacting actionability:
  - Unclear instructions/guidelines
  - Lack of knowledge
  - Additional work, need for adaptations
- Idea: Identify points to combine DMP work with everyday project work

Examples
- Obligatory along with "Laufzettel"/annual reports
- Include in onboarding for new employees
- …

Resources, Methods, Tools

Education material

Trainings

Data Stewards

File naming convention

GitLab

Metadata convention

Persistent Identifiers

Sciebo

Data Management Consultants

…