



NFDI4Ing

Special Interest Group (SIG)

Qualitätssicherung und Metriken für FAIR-Daten

10. Treffen, 08.12.2023

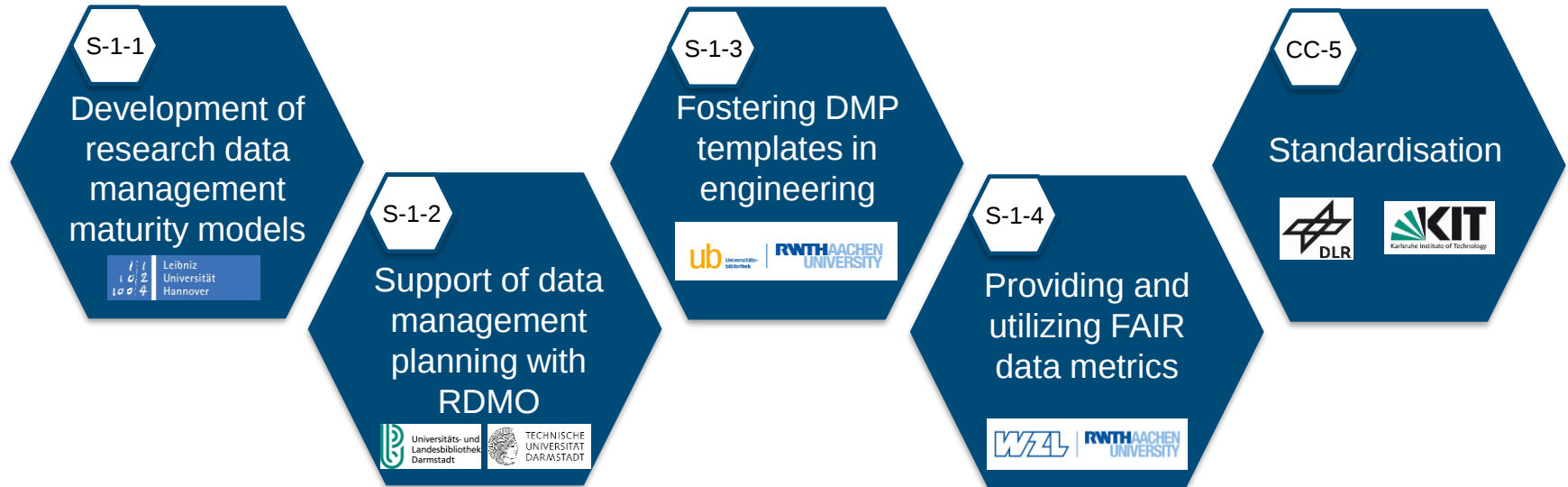
Rückblick & Ausblick

SIG Qualitätssicherung und Metriken für FAIR-Daten

Entwicklung und Austausch verschiedener Produkte, Inhalte und Prozesse zur Qualitätssicherung der Forschungsdaten und Prozesse in den Ingenieurwissenschaften

- Laufzeit: 36 Monate – Start: Jun. 2021 Ende: Jun. 2024
- Meetingfrequenz: Quartalsweise
- ...
- 14.01.2022 (3. Treffen): **Datenmanagementpläne**
- 08.04.2022 (4. Treffen): **Forschungsdatenmanagementpläne und Forschungsprozesse**
- 10.06.2022 (5. Treffen): **FAIR in den Ingenieurwissenschaften**
- 30.09.2022 (6. Treffen): **Forschungsaktivitäten entlang des DLC**
- 13.01.2023 (7. Treffen): **Datenqualitätsmetriken mit der TA GOLO**
- 14.04.2023 (8. Treffen): **Qualitätssicherung in der Forschung**
- 08.09.2023 (9. Treffen): **Qualitätssicherung in der NFDi4ing**
- **08.12.2023 (10. Treffen): *Rückblick & Ausblick***

SIG Qualitätssicherung und Metriken für FAIR-Daten



SIG-Team: TA S-1 und CC-5



Max Leo Wawer
IPeG
Leibniz Universität Hannover



Gerald Jagusch
Universitäts- und Landesbibliothek
TU Darmstadt



Jürgen Windeck
Universitäts- und Landesbibliothek
TU Darmstadt



Michael Selzer
IAM-CMS
Karlsruher Institut für Technologie



Iryna Mozgova
DMB
Universität Paderborn



Mario Moser
WZL
RWTH Aachen University



Daniela Hausen
UB
RWTH Aachen University



Sabine Schönau
UB
RWTH Aachen University



Christian Langenbach
Deutsches Zentrum für Luft- und Raumfahrt

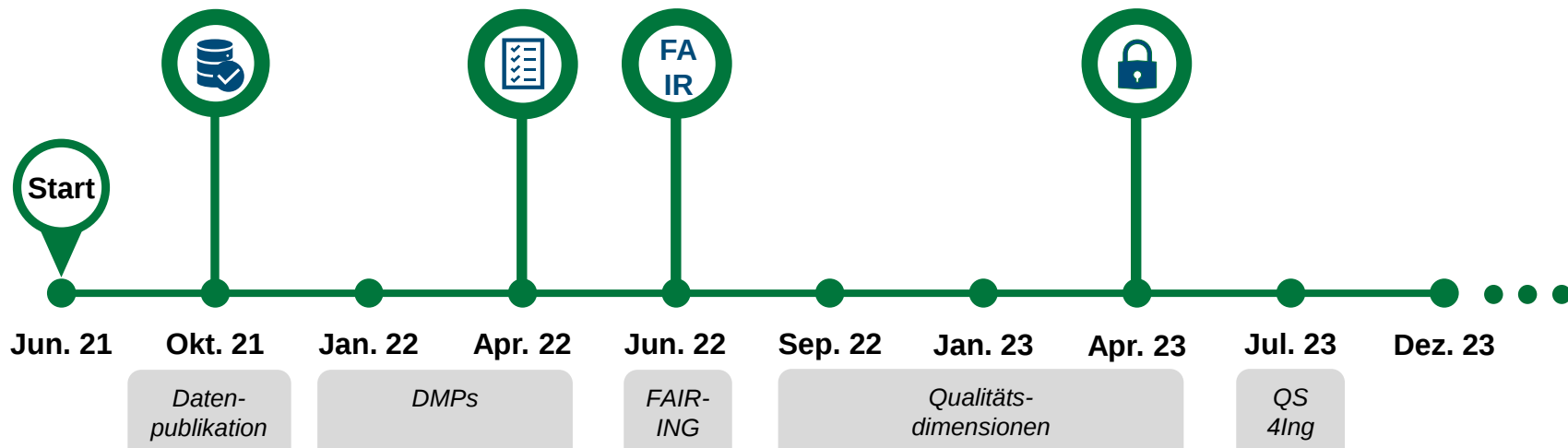


Barbara Steude
Deutsches Zentrum für Luft- und Raumfahrt

Die Rolle der Special Interest Groups (SIGs)



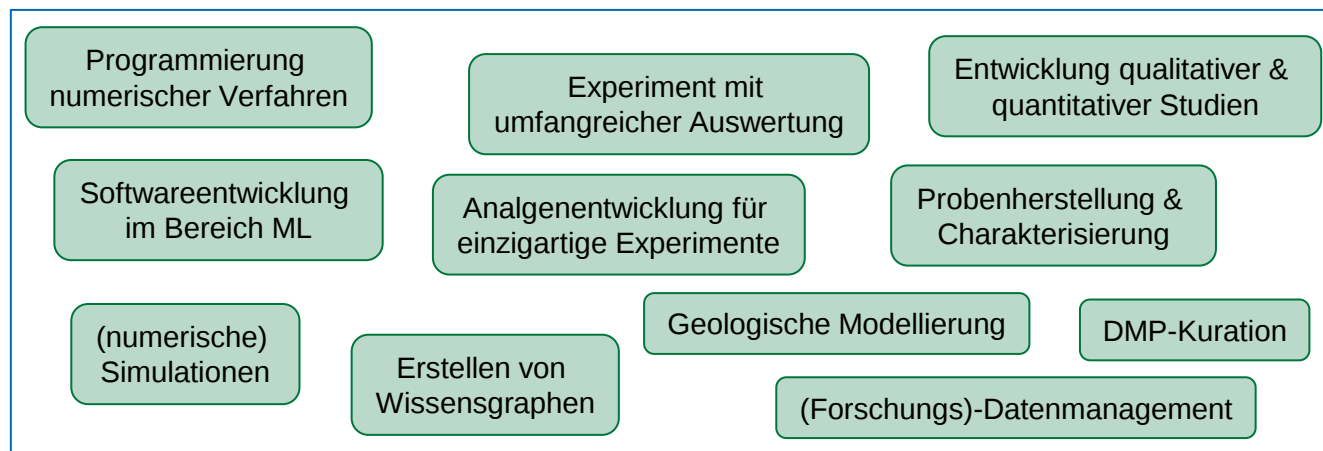
Meilensteine: SIG Qualitätssicherung und Metriken für FAIR-Daten



● Rückblick auf die 1. SIG zum Thema „Kick-Off“ Auswertung der Umfrageergebnisse des Miroboards

Kick-Off – „Get in Touch“

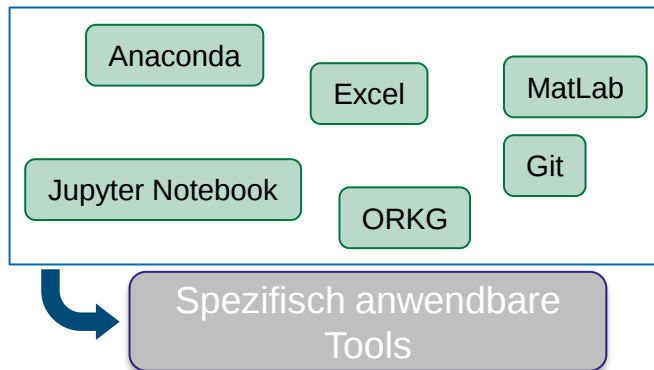
Typische Forschungsaktivitäten der Community



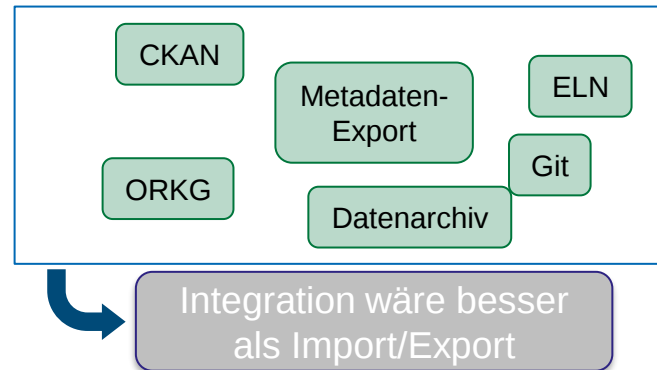
Heterogene Forschungsaktivitäten auch
mit Fokus auf Infrastrukturbereich

Kick-Off – „Get in Touch“

Tools zur Erfassung, Analyse und Weiterverarbeitung der Daten



Anbindung an ein Datenmanagement-Planungstool



● Rückblick auf die 2. SIG zum Thema „Qualitätsgesicherte Publikation von Forschungsdaten“ Auswertung der Umfrageergebnisse des Miroboards

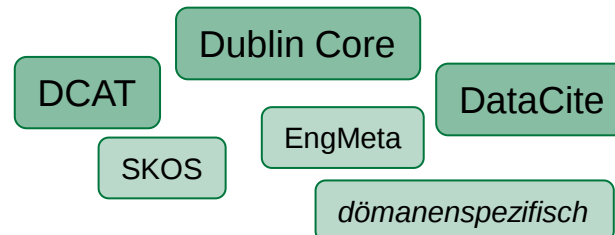
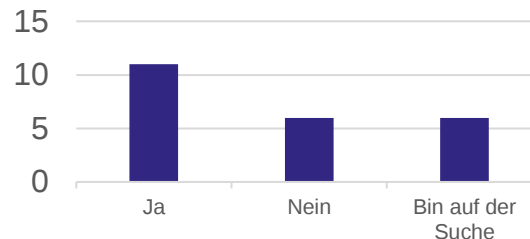
Standards in der Datenpublikation (Mentimeter)

Welche Kriterien zur Datenpublikation müssen erfüllt sein?



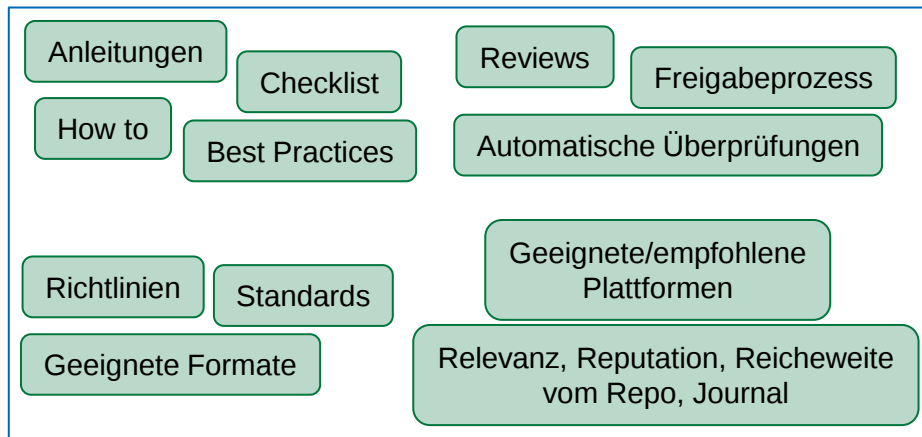
Mindeststandards in der Datenpublikation

Voting (n = 23),
Kennen Sie bereits Mindeststandards?

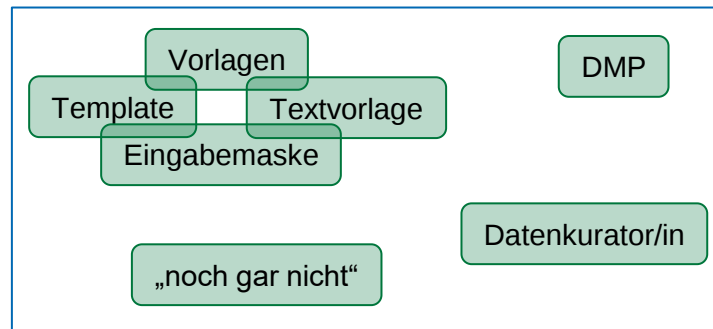


Standards in der Datenpublikation (Mentimeter)

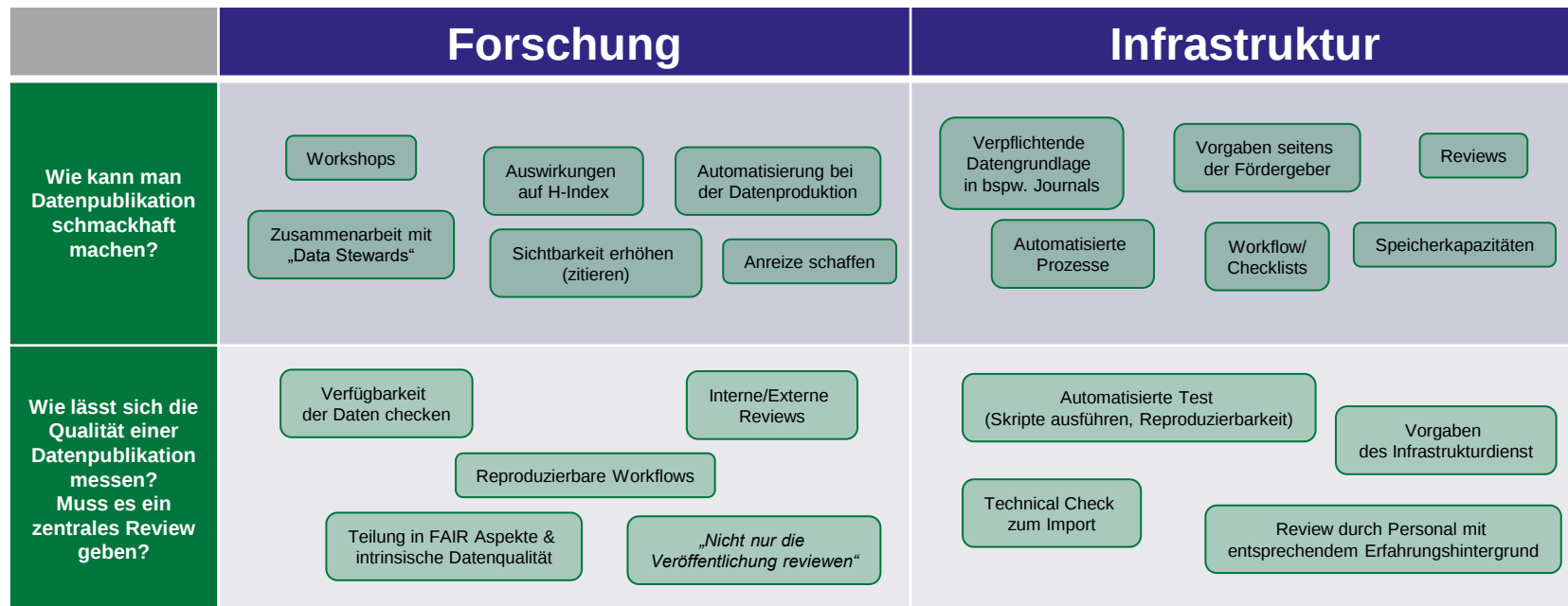
Mechanismen zur Qualitätssicherung von Datenpublikationen



Planung der Datenpublikationen – Hilfsmittel

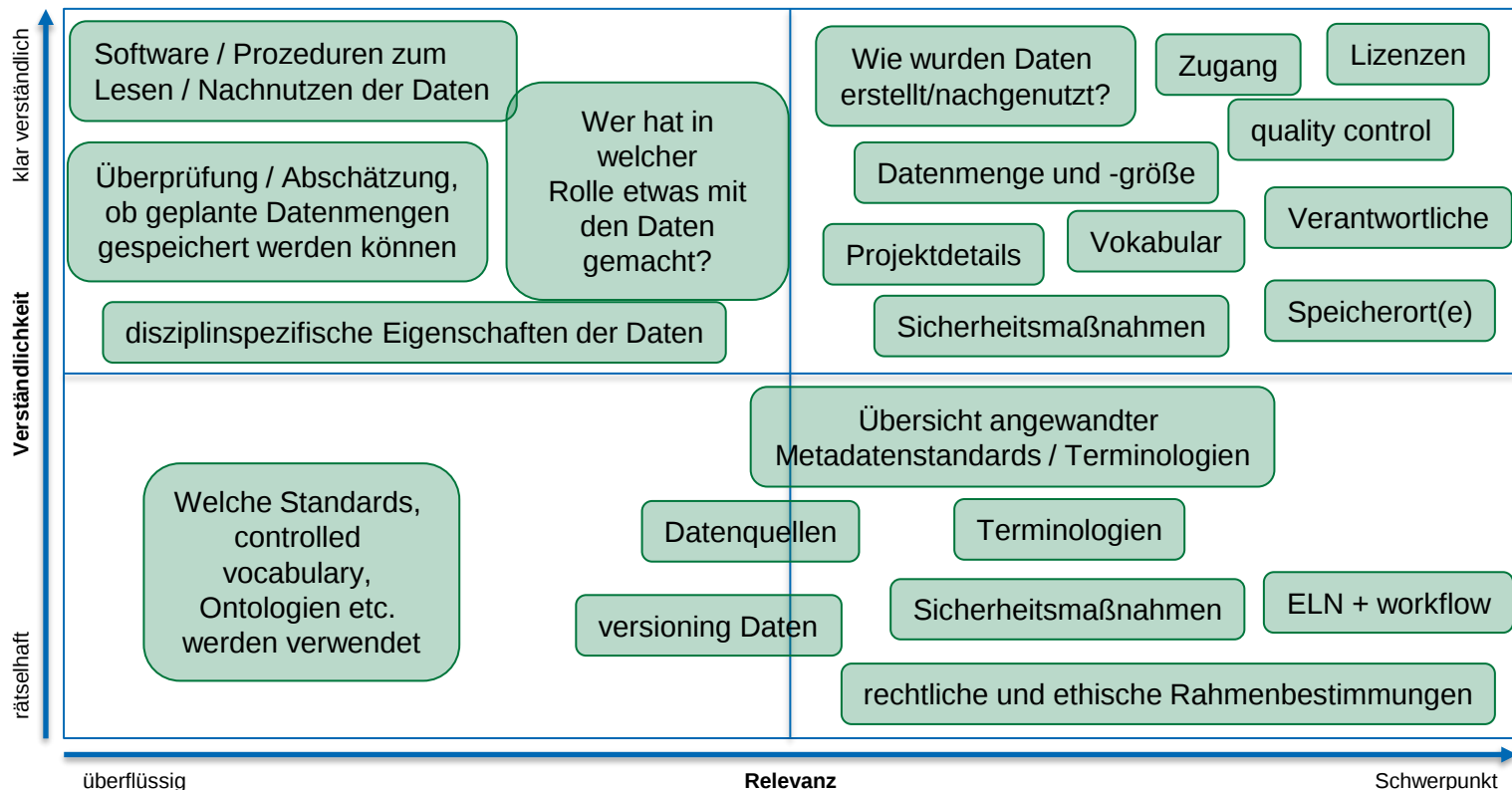


Daten publizieren – Forschung & Infrastruktur



● Rückblick auf die 3. SIG zum Thema „Datenmanagementpläne“ Auswertung der Umfrageergebnisse des Miroboards

Einordnung von DMP-Inhalten



Daten publizieren – Maßnahmen

Welche **Maßnahmen** sind zu ergreifen, um die Daten vor versehentlicher oder absichtlicher **Manipulation**, vor **Verlust** oder unbefugten **Zugriffen** zu **schützen**?

Nutzerver-
waltung

Kontrollierter
Datenzugriff

Nach finaler
Version kann
nur noch
ergänzt werden

Versionisierung

Nach welchen **Kriterien** können **Daten ohne Wissensverlust** wegelassen werden, zur Reduzierung der Datenmenge?

z.B. nur Code &
Inputfile aufbewahren
>> Daten können
(neu) repliziert
werden

Bereinigung
aktualisierter
Datensätze

interpolierbare
Daten können
gekürzt
werden

Erstellen von
Modellen, die die
Daten mit
ausreichender
Genauigkeit abbilden

Gute Dokumentation und
reproduzierbare Datenquelle
anstatt redundante Ablage der
gleichen Starter-Datensätze

Welche **Maßnahmen** können angewandt werden um den **Erstellungsaufwand** zur Archivierung von Forschungsdaten zu **reduzieren**?

automatisierte Beschreibung
mit Metadaten und Ausgabe
an Repository

Data
Curation
Service

kontrollierte
Vokabulare /
Ontologische Begriffe
statt Fließtext

Vorgegebene
Protokolle, die nur
noch ausgefüllt
werden müssen

Services verknüpfen,
die bei Datenerstellung
genutzt werden, um
Metadaten
nachzunutzen

maDMP aus
denen
Informationen
ausgelesen
werden können

Verknüpfung mit alten bzw. analogen Daten: Ist dies aus Ihrer Sicht erforderlich?

ja, da aus
analogen Daten
jederzeit
"modernere,
Digitalisate erstellt
werden können

ja, da Frage nach
Original/Ursprung
sdatei (Kopie von Kopie)
-->
Datenprovenienz

ja, da in
Zukunft neue
Datenformate,
Auflösungen
etc.

manchmal
sind analoge
Daten
einfacher zu
archivieren als
digitale (LZA)

● Rückblick auf die 4. SIG zum Thema „DMPs“ Auswertung der Umfrageergebnisse des Miroboards

Auswertung der Umfrageergebnisse des Miroboards: mit „*must have*“ gevotete Merkmale



NFDI4Ing DMP Template für RDMO

<https://rdmo.nfdi4ing.de>

Inhaltliche Datenbeschreibung

Der vorliegende Fragenkatalog ist die von der DFG herausgegebene Checkliste ergänzt um ingenieurwissenschaftliche Inhalte. Die Fragen aus der Checkliste der DFG müssen bei der Antragstellung und beim Projektabschluss berücksichtigt werden und sollen Ihnen helfen, Ihr Datenmanagement zu strukturieren. Bei Fragen wenden Sie sich bitte an die zuständige Kontaktperson.

Die folgenden Fragen dienen der Beschreibung der Datensätze, die im Projekt erzeugt und/oder verwendet werden. Die Definition dessen, was jeweils ein Datensatz ist, ist eine wichtige konzeptionelle Entscheidung, die für jedes Vorhaben bzw. Projekt individuell getroffen werden muss. Wählen Sie als Datensatz eine logische Einheit Ihrer Forschungsdaten, für die gleiche Angaben bezüglich Methodik, Technik, Zugänglichkeit, usw. gelten.

Die Einteilung in Datensätze hilft, den Wert der Daten hinsichtlich der potentiellen Nachnutzung und einer späteren Archivierung einschätzen zu können. Da alle Fragen für jeden einzelnen Datensatz beantwortet werden müssen, ist eine allzu feingranulare Struktur nicht sinnvoll.

Bitte füllen Sie das Formular für jeden Set aus. Die verschiedenen Sets werden eventuell in späteren Fragen wieder verwendet. Sie können einen neuen Set mit dem grünen Button hinzufügen. Bereits angelegte Sets können mit den Buttons oben rechts bearbeitet oder wieder entfernt werden.

Felddaten Vorlage Rohdaten Feld Set hinzufügen

Um was für einen Datensatz handelt es sich?

- ☐ Tabellen
- ☐ Binäre oder numerische Daten
- ☐ Code
- ☐ Simulationsdaten
- ☐ Algorithmen
- ☐ Messdaten z.B. von Prüfständen
- ☐ Bilddaten
- ☐ Big Data (heterologische Datentypen)
- ☐ Referenzen
- ☐ Textuelle Daten

- Basierend auf der DFG Checkliste <https://www.dfg.de/forschungsdaten/checkliste>
- Angepasste Hilfen und Antwortoptionen für die Ingenieurwissenschaften
- Basierend auf Erhebungen in NFDI4Ing und Ergebnissen der SIG Workshops

Navigation

Wenn Sie die Navigation verwenden, wird Ihre Eingabe gespeichert.

Einträge mit ⌕ können aufgrund Ihrer Eingabe übersprungen werden.

Datennutzung
→ Inhaltliche Datenbeschreibung
Technische Datenbeschreibung
Dokumentation und Datenqualität
Speicherung und technische Sicherung ...
Rechtliche Verpflichtungen und Rahme...
Datenaustausch und dauerhafte Zugän...
Verantwortlichkeiten und Ressourcen

● Rückblick auf die 5. SIG zum Thema „FAIRe Daten“ Auswertung der Umfrageergebnisse des Miroboards (Metriken für FAIR-Daten)

Rückblick: Einstieg in FAIR Prinzipien und Metriken

Tieferer Einstieg in FAIR (Impuls + Quiz)

NATIONALE FORSCHUNGS-
DATENINFRASTRUKTUR FÜR DIE
INGENIEURWISSENSCHAFTEN

NFDi4ing 

Die vier *foundational / core* FAIR Principles

F Findable	A Accessible	I Interoperable	R Reusable
----------------------	------------------------	---------------------------	----------------------

NATIONALE FORSCHUNGS-
DATENINFRASTRUKTUR FÜR DIE
INGENIEURWISSENSCHAFTEN

NFDi4ing 

Wahr oder Falsch?

Es geht noch praktikabler: Die FAIR Prinzipien werden für die Anwendung durch *Metriken* interpretiert und (ggf. Domain-spezifisch) ausgestaltet.

 Ja, auch das ist richtig!

Die FAIR Prinzipien sind kurz und bündig, Domain-unabhängig und high-level, um auf eine möglichst breite Maße an digital objects anwendbar zu sein. Die konkrete Ausgestaltung erfolgt in Form von Metriken.

„Communities have different practices and therefore there should be some flexibility on how the FAIR principles should be implemented within a certain community“ [Behren et al. (2020), S. 5]

Eine exemplarische Metrik werden wir heute noch kennen lernen.

10.02.2023 — Slide 10 — DOI: 10.5281/zenodo.6647047

<https://doi.org/10.5281/zenodo.6647047>

Gastvortrag zur EOSC Metrik

Benefits of this solution

- Unambiguous interpretation of Links
- Explicit pointers to clearly-separated data and metadata

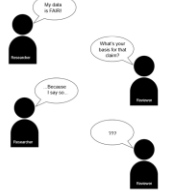
Recommendation for all scenarios: DOI example

doi:10.123/456.78

 Data Cite

Link rel="cite-as" http://doi.org/10.123/456.78

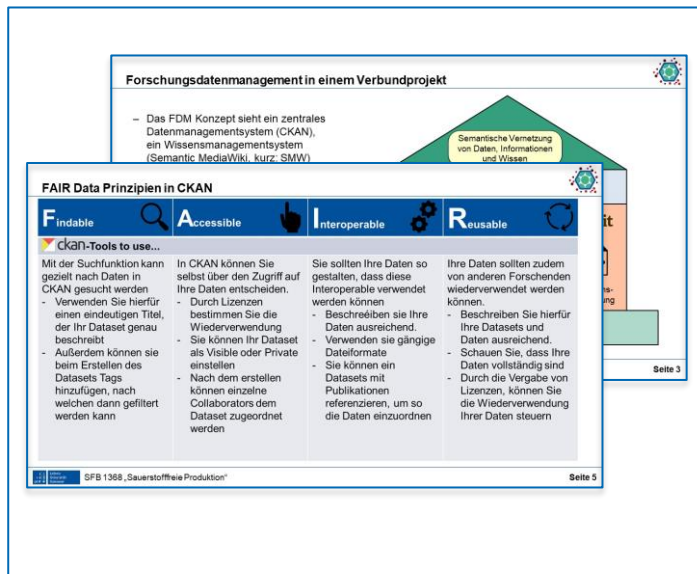
Link rel="described-by" http://data.crosscite.org/10.123/456.78



10.02.2023 — Slide 10 — DOI: 10.5281/zenodo.6647047

Rückblick: FAIR in Forschung

Vortrag zu FAIR in Forschungsprojekt



Forschungsdatenmanagement in einem Verbundprojekt

- Das FDM Konzept sieht ein zentrales Datenmanagementsystem (CKAN), ein Wissensmanagementsystem (Semantic MediaWiki, kurz: SMW)

FAIR Data Prinzipien in CKAN

F indable	A ccessible	I nteroperable	R eusable
ckan-Tools to use... Mit der Suchfunktion kann gezielt nach Daten in CKAN gesucht werden - Verwenden Sie hierfür einen eindeutigen Titel, der Ihr Dataset genau beschreibt - Außerdem können sie beim Erstellen des Datasets Tags hinzufügen, nach welchen dann gefiltert werden kann	In CKAN können Sie selbst über den Zugriff auf Ihre Daten entscheiden. - Durch Lizenzen bestimmen Sie die Wiederverwendung - Sie können Ihr Dataset als Visible oder Private einstellen - Nach dem erstellen können einzelne Collaborators dem Dataset zugeordnet werden	Sie sollten Ihre Daten so gestalten, dass diese Interoperable verwendet werden können - Beschreiben sie Ihre Daten ausreichend. - Verwenden sie gängige Datelformate - Sie können ein Datasets mit Publikationen referenzieren, um so die Daten einzuordnen	Ihre Daten sollten zudem von anderen Forschenden wiederverwendet werden können. - Beschreiben Sie hierfür Ihre Datasets und Daten ausreichend. - Schauen Sie, dass Ihre Daten vollständig sind - Durch die Vergabe von Lizenzen, können Sie die Wiederverwendung Ihrer Daten steuern

Seite 3

Seite 5

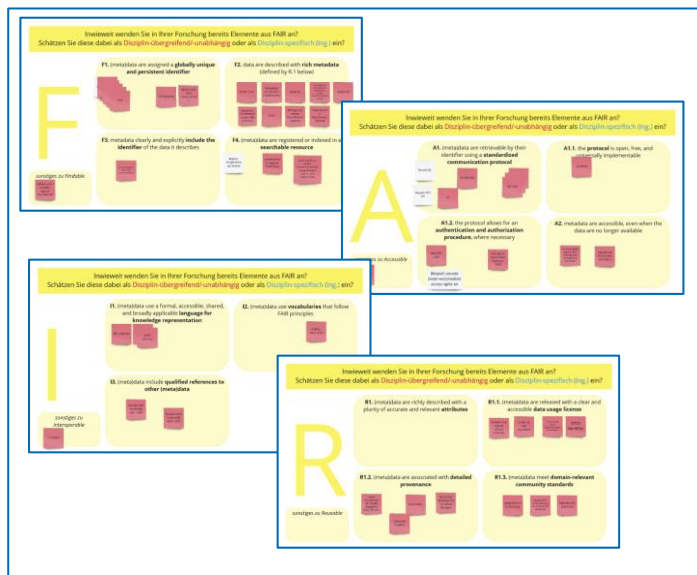
SFB 1368 „Sauerstofffreie Produktion“

Digital Objects in der Forschung

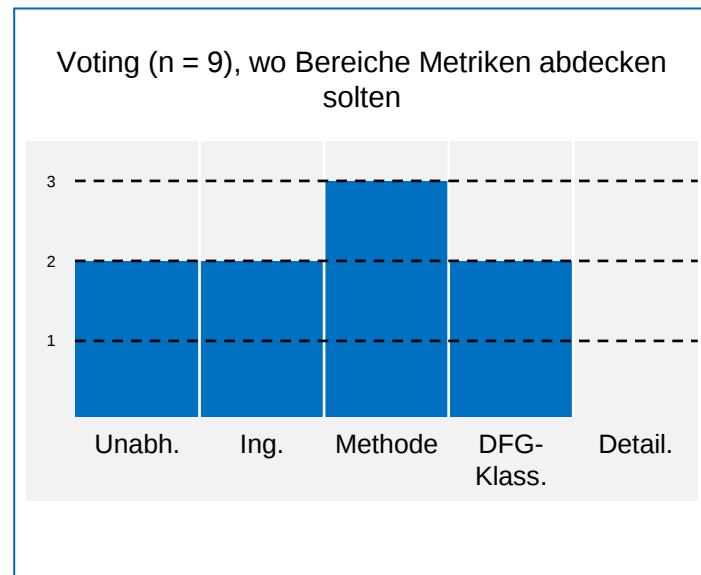
- Daten: Maschinendaten, Fragebögen, Verkehrsdaten uvm.
- Repositorien: GitHub / GitLab, CKAN, Data Collections Explorer uvm.
- Infrastruktur: VIVO, CaosDB
- Publikationen

Rückblick: Miroboard zu FAIR in Ingenieurwissenschaften

FAIR in Ihrer Forschung (allg. / ing.-spez.)



FAIR Metrik für Ingenieurwissenschaften?



Ausblick zu Workshop „FAIR Metriken in den Ingenieurwissenschaften“

Weitere Verwendung

- Inhalte in mehreren Workshops gesammelt:
 - FDM-Netzwerktreffen an RWTH Aachen
 - NFDI4ing-Konferenz 2023
- Workshop-Format dabei leicht weiterentwickelt
 - Frage nach Gewichtung der Kriterien
- Beispiel: Vgl. <https://doi.org/10.5281/zenodo.8382740>

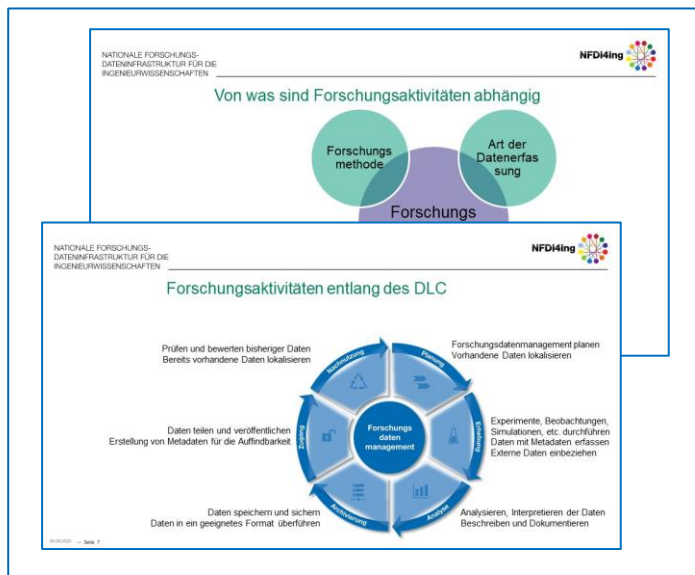
Ziel und Ausblick

- FAIR Metrik (gesamt oder einzelne Elemente) für die Ingenieurwissenschaften entwickeln
- Einzelne Elemente: Einzelne Prinzipien spezifisch für die Ingenieurwissenschaften ausgestalten, z.B. R1. (meta)data
- FAIR Metriken als praktische Anleitung, um FAIRness in den Ingenieurwissenschaften zu erhöhen

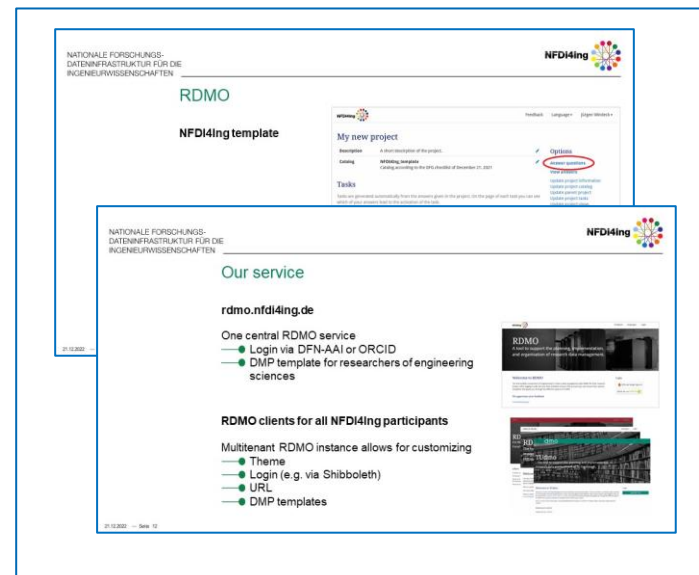
● Rückblick auf die 6. SIG zum Thema „Forschungsaktivitäten entlang des DLC“ Auswertung der Umfrageergebnisse des Miroboards

Rückblick: Einführung in DLC - RDMO

Einführung Datenlebenszyklus



DMP - RDMO



The image shows two screenshots of the NFDI4ing RDMO interface, both featuring the NFDI4ing logo and header.

Top Screenshot: NFDI4ing template

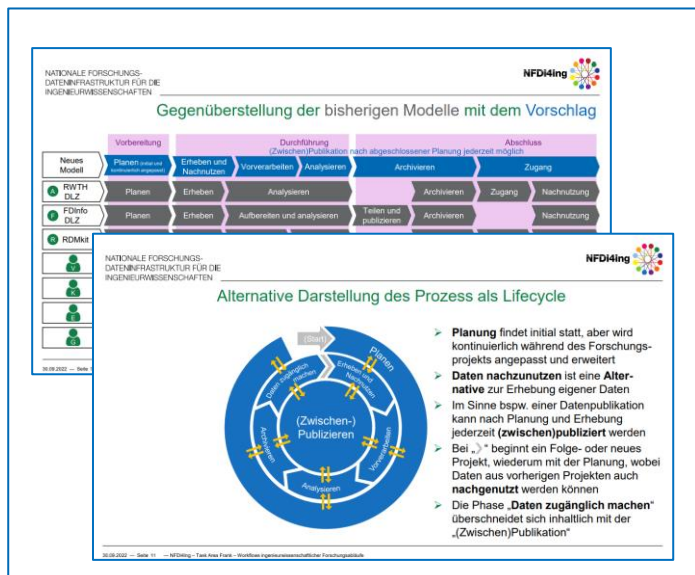
This screenshot shows the "My new project" form. The "Template" dropdown is set to "RDMO template". The "Tasks" section lists several tasks, including "Create a new project" and "Create a new project".

Bottom Screenshot: Our service

This screenshot shows the "Our service" page. It highlights the central RDMO service and lists the login options: "Login via DFN-AAI or ORCID" and "DMP template for researchers of engineering sciences". It also lists the RDMO clients for all NFDI4ing participants: "Theme", "Login (e.g. via Shibboleth)", "URL", and "DMP templates".

Rückblick: Workflow ing.wiss.; Forschungsfelddaten

TA Frank – Workflow ing.wissen.



TA GOLO - Forschungsfelddaten

NATIONALE FORSCHUNGS-
DATENINFRASTRUKTUR FÜR DIE
INGENIEURWISSENSCHAFTEN

G-3: Wiki: Data Quality Metrics

<https://data-quality-metrics.readthedocs.io>

- Metrics for data quality assessment
 - Meaning and interpretation of metrics
 - Range and limits of the metrics
 - Libraries and Repositories for application in Python, C++ and Matlab

NATIONALE FORSCHUNGS-
DATENINFRASTRUKTUR FÜR DIE
INGENIEURWISSENSCHAFTEN

G-2: Survey: "Arbeit mit Forschungsfelddaten"

<https://survey.uni-hannover.de/index.php/659451>

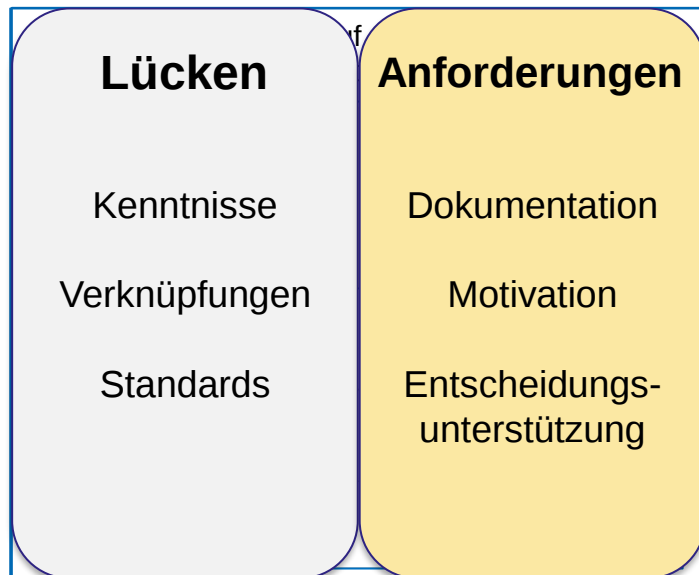
- 35 questions addressing
 - Aspects and dimensions of the field
 - Data quality metrics for sensor data and datasets
 - Knowledge about methods for data collection
 - Interfaces to collect and transfer field data
 - Tools for data management and storage
- Results of first iteration
 - Lack of knowledge about data quality metrics
 - Interest in organization of heterogeneous field data without commercial software

G-2-1: Identification of types of necessary and sufficient information about a technical system
G-2-2: Classification of data collection methods

21.12.2022 — Seite 8 — NFDi4ing 950 "Data & Metadata"

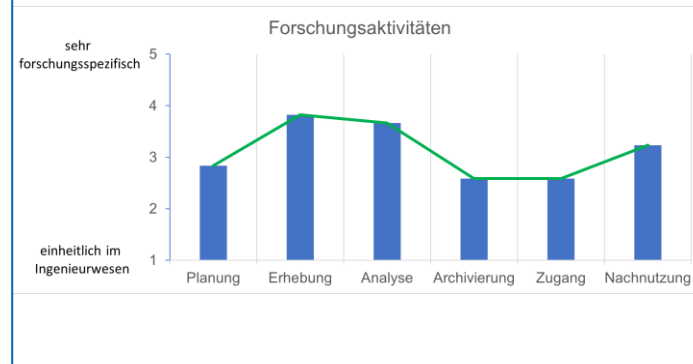
Rückblick: Miroboard - Forschungsaktivitäten entlang des DLC

Forschungsaktivitäten - DLC?

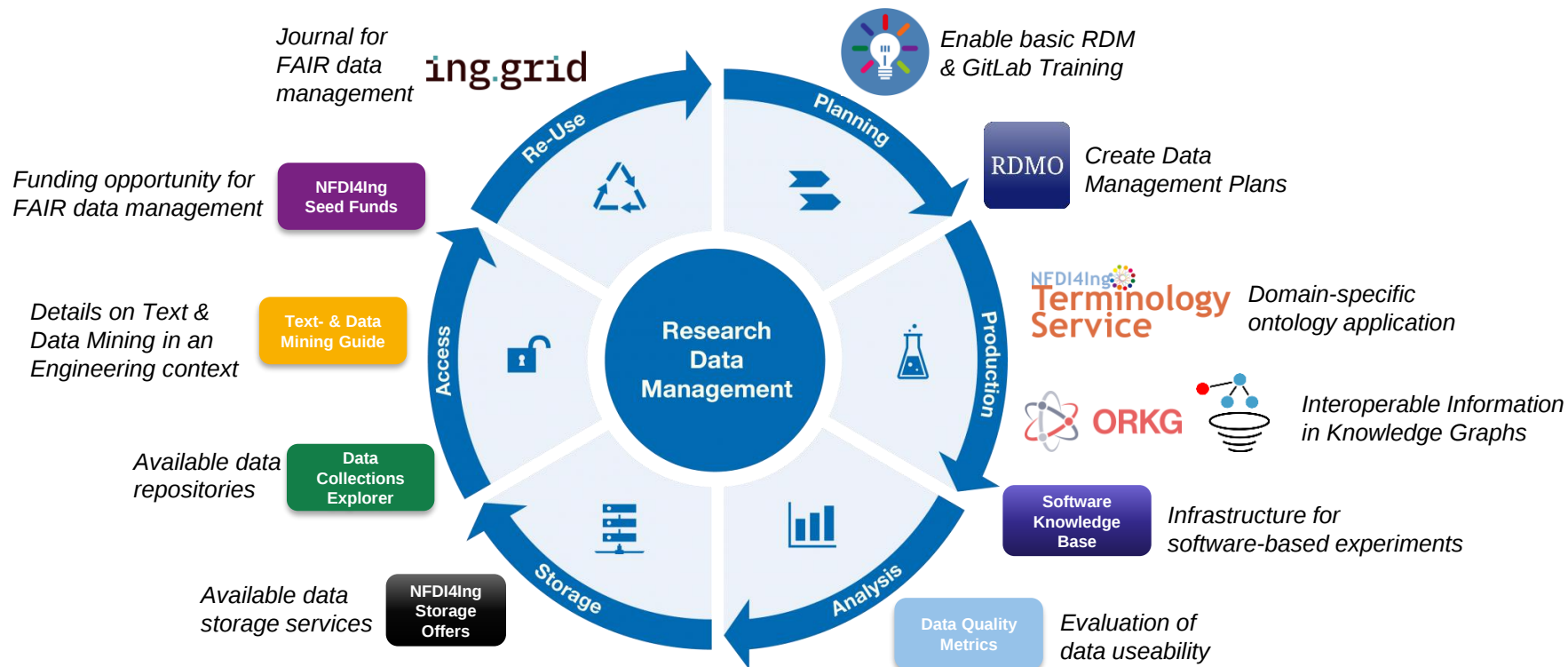


Einheitliche Forschungsaktivitäten?

Voting (n = 18), wie einheitlich die Forschungsaktivitäten in den Phasen des DLC sind



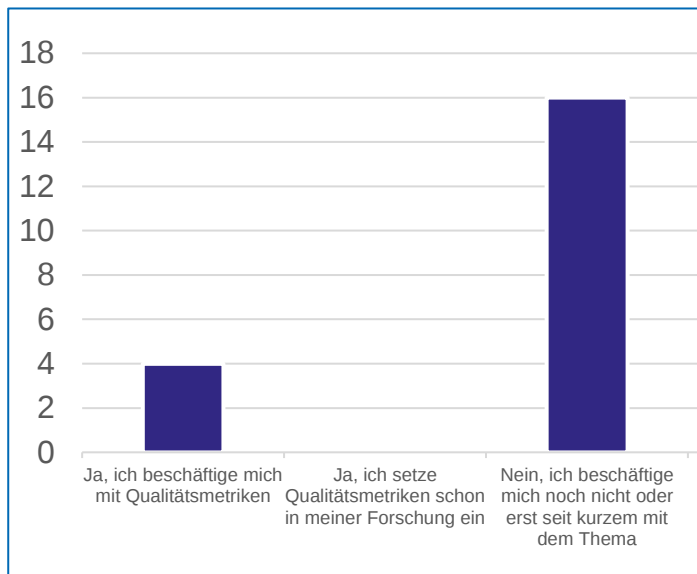
Productive 4Ing Services along the Research-Data-Lifecycle



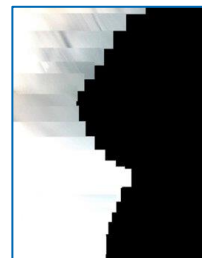
● Rückblick auf die 7. SIG zum Thema „Datenqualitätsmetriken“, Auswertung der Umfrageergebnisse des Miroboards

Rückblick: Einführung in Datenqualität

Umfrage Datenqualitätsmetriken



Einführung



#Copyright by Barbara Steude

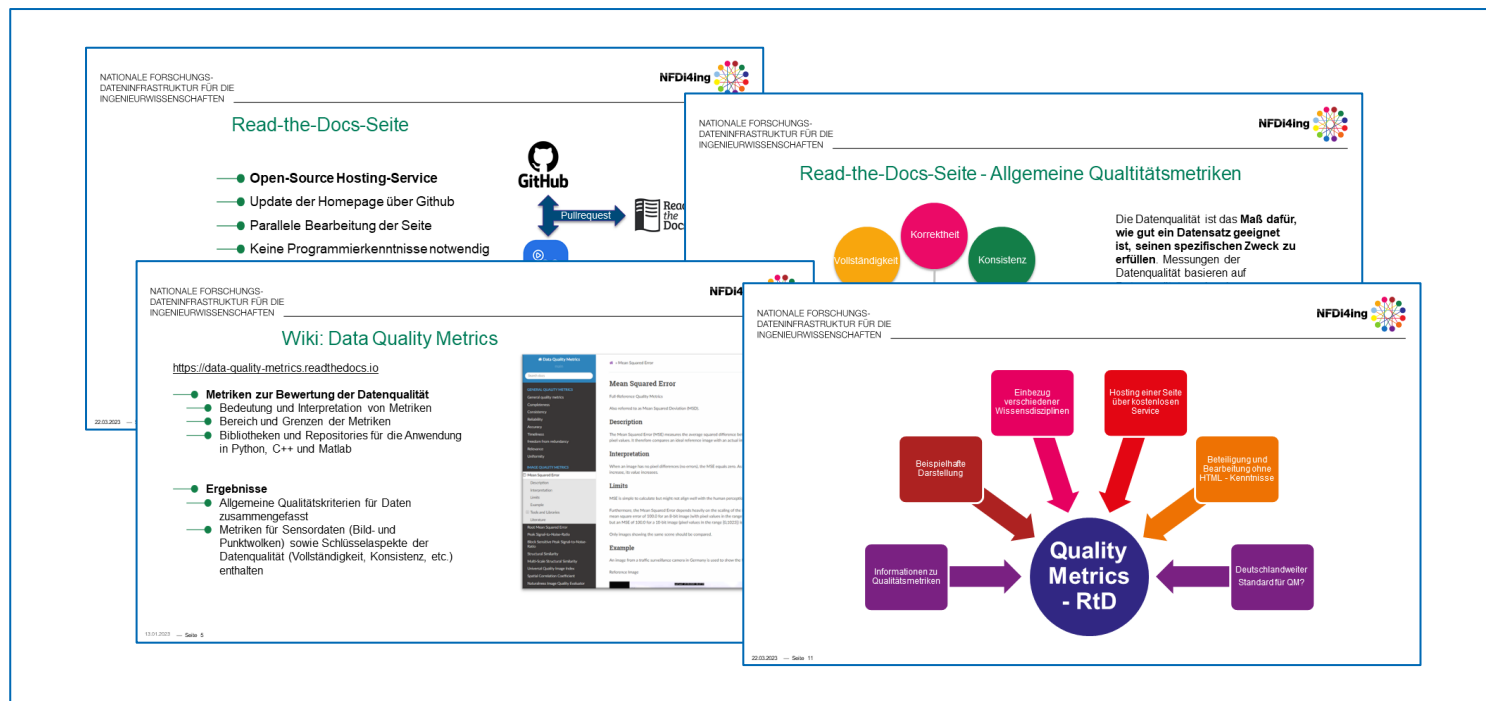


#Glas #Liebe #meins #kaputt
#schwarz #hellgrau #mittelgrau #lila
#weiss #tisch #Plastik #Spiegelung
#durcheinander



It's better to mention the context

Rückblick: TA GOLO – Data Quality Metrics



Rückblick: Miroboard - Datenqualitätsmetriken

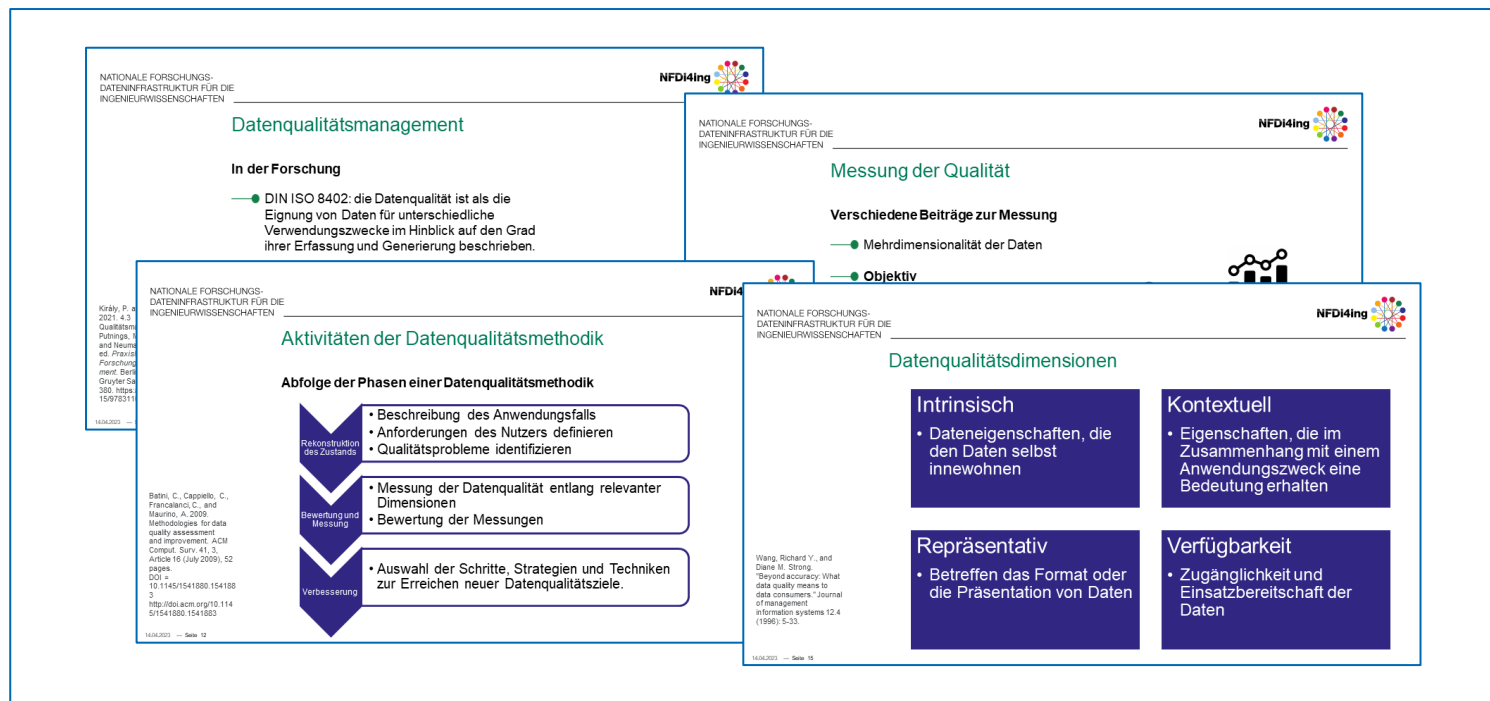
Qualitätsanforderungen in der Forschung



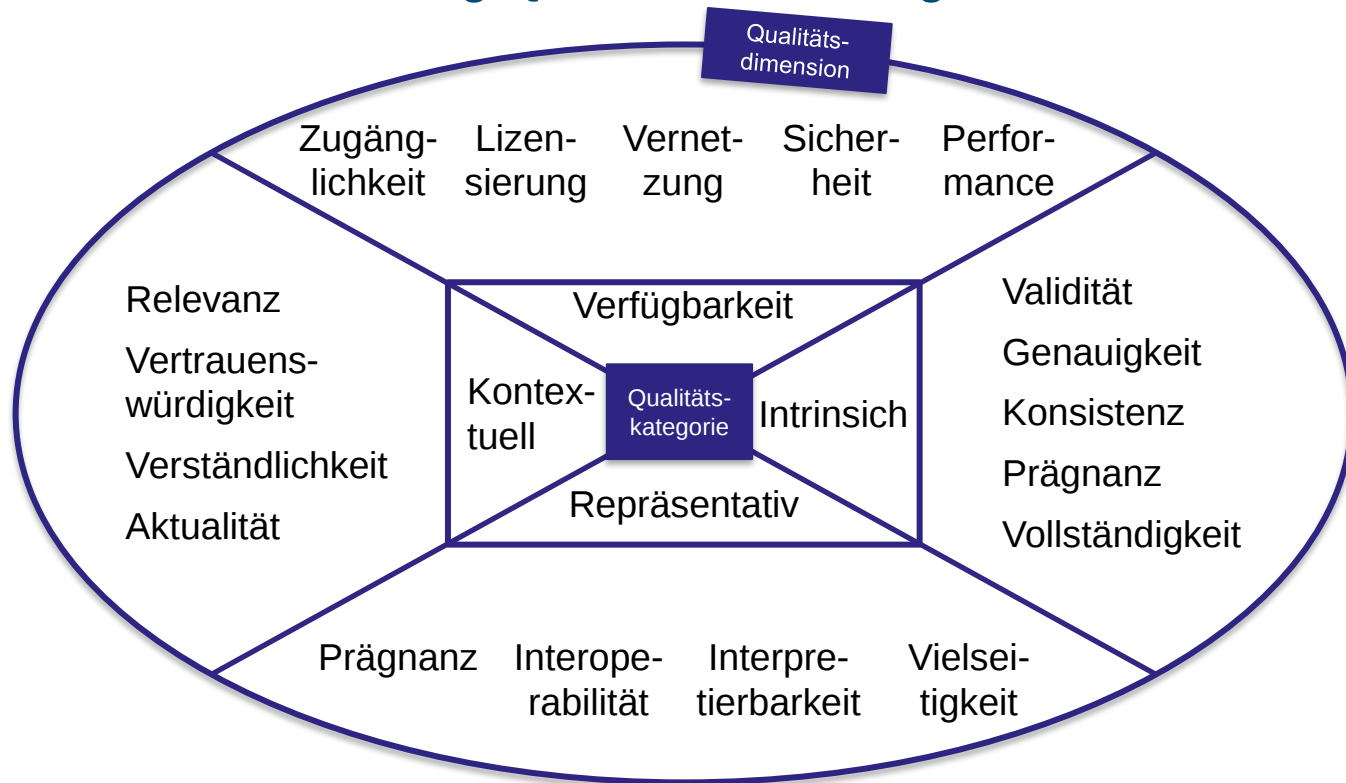
- Abhängigkeit von der Forschungsmethode
- Qualität bezieht sich auf viele Forschungsinhalte
- Bedeutsamkeit der Kriterien abhängig vom Forschungsinhalt
- Vielzahl an Anforderungen
- wenig genannte Metriken

● Rückblick auf die 8. SIG zum Thema „Qualitätssicherung im Forschungsprozess“

8.SIG: Rückblick: Einführung Qualitätssicherung

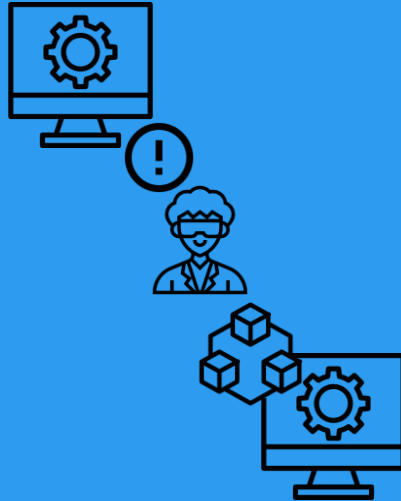


Rückblick: Einführung Qualitätssicherung

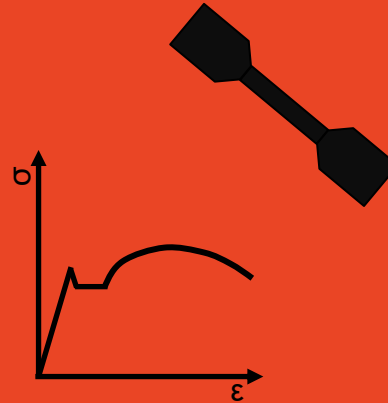


Rückblick: Qualitätssicherung in der Forschung – drei Fallbeispiel

Softwareentwicklung



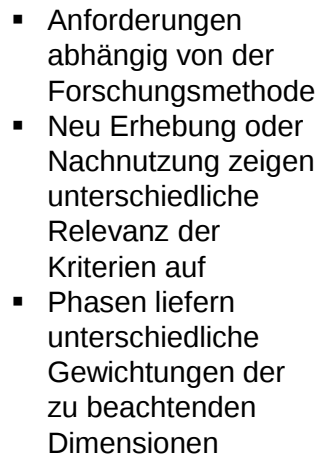
Materialverhalten



Umfrage



Qualitätssicherung in der Forschung (Auszug)



Rückblick: Einführung Qualitätssicherung

Anforderungen

- Ergeben sich auf Basis des eigenen Forschungsgegenstands und der Forschungscommunity (Nachnutzende)

Qualitätsdimensionen

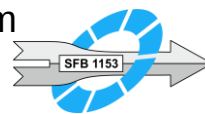
- Lassen sich nicht immer klar Anforderungen zuweisen
- Ermöglichen eine qualitätsbezogene Einordnung der Anforderungen
 - Ableitung von Handlungsmöglichkeiten

Qualitätskategorien

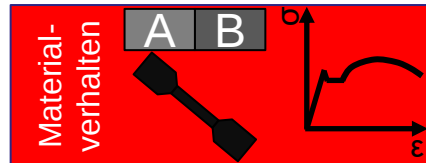
- **Kontextuell** und **Intrinsisch** basieren im Wesentlichen auf Anforderungen, die in der Planung und Erhebung sowie Analyse berücksichtigt werden sollten
- **Repräsentativ** bezieht sich auf die Anreicherung der Daten durch Metadaten (Dokumentation im Rahmen der Archivierung und Zugänglichkeit)
- **Zugänglichkeit** legt die Verbindung zur Forschungsdateninfrastruktur, im Fokus liegen dabei Archivierung und Zugang

Ausblick: Datenqualitätssicherung am Beispiel von Zugversuchen

- Validierung durch Forschende aus einem SFB für die Materialcharakterisierung von multimateriellen Bauteilen



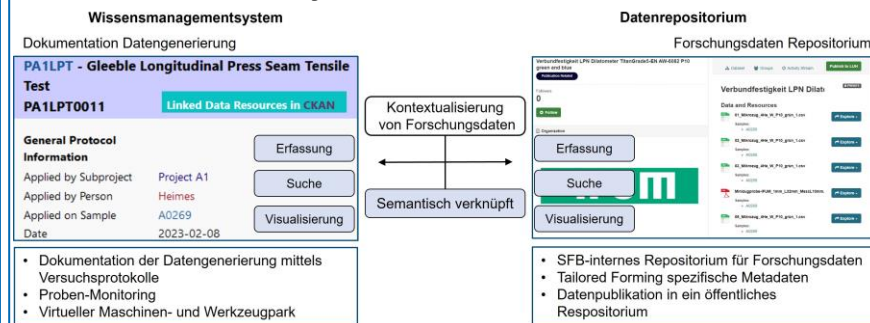
DOI: [10.35199/dfx2023.15](https://doi.org/10.35199/dfx2023.15)



Qualitätssicherung

- Normierte Planung der Versuche
- Vokabular-orientierte Protokolle
- Forschungsspezifische Metadaten
- Kontextualisierung der Forschungsdaten mit Versuchsprotokollen
- Data-Push in ein öffentliches Repo

Internes FDM-System



● Rückblick auf die 9. SIG zum Thema „Qualitätssicherung in der NFDI4Ing – Aktuelle Arbeiten “

Gastbeiträge aus der NFDI4ing

TA DORIS

QUALITY ASPECTS FOR HPMC

High-Performance **Computing & Measurement**

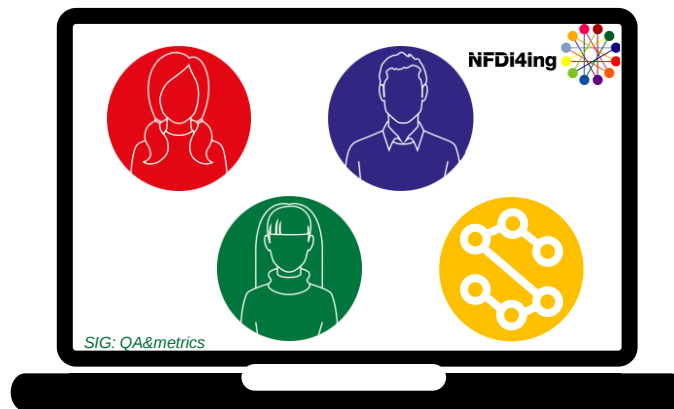
data that are created during a research process or are the result of it

„Quality Assurance“: **Modellsicherheit** / Genauigkeit des simulierten Modells und der Berechnungsmethoden
Validierung (“are we solving the right equations?”)
gegenüber Theorie & vorhandenen (Referenz-)Daten
→ **Stimmt das Modell?**

Verifizierung (“are we solving the equations right?”)

→ **Ist das Modell richtig implementiert? (z.B. im Code)**

► **Qualitätssicherung ergibt sich aus dem Prozess der Generierung**



TA BETTY

QUALITY ASPECTS OF RESEARCH SOFTWARE



PUBLICLY AVAILABLE & PERMISSIVE LICENSE



FUNCTIONAL & ACCURATE



WELL-DOCUMENTED



MODULAR / FLEXIBLE / EFFICIENT / UNDERSTANDABLE



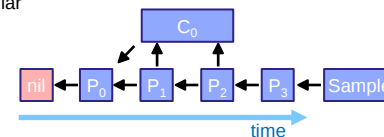
PORTABLE / INTEROPERABLE / REPRODUCIBLE

TA CADEN

QUALITY ASPECTS FOR SEQUENCES OF PROCESSING

„Quality Assurance“: 1. hochstrukturierte Dokumentation von Forschung, und zwar sowohl für Prozesse, als auch Ergebnisse. 2. durch Ursache-Wirkungs-Beziehungen den *Kontext* der Forschung darstellen. 3.

Qualitätsbezogene Tags von Daten mit einem expliziten Vokabular

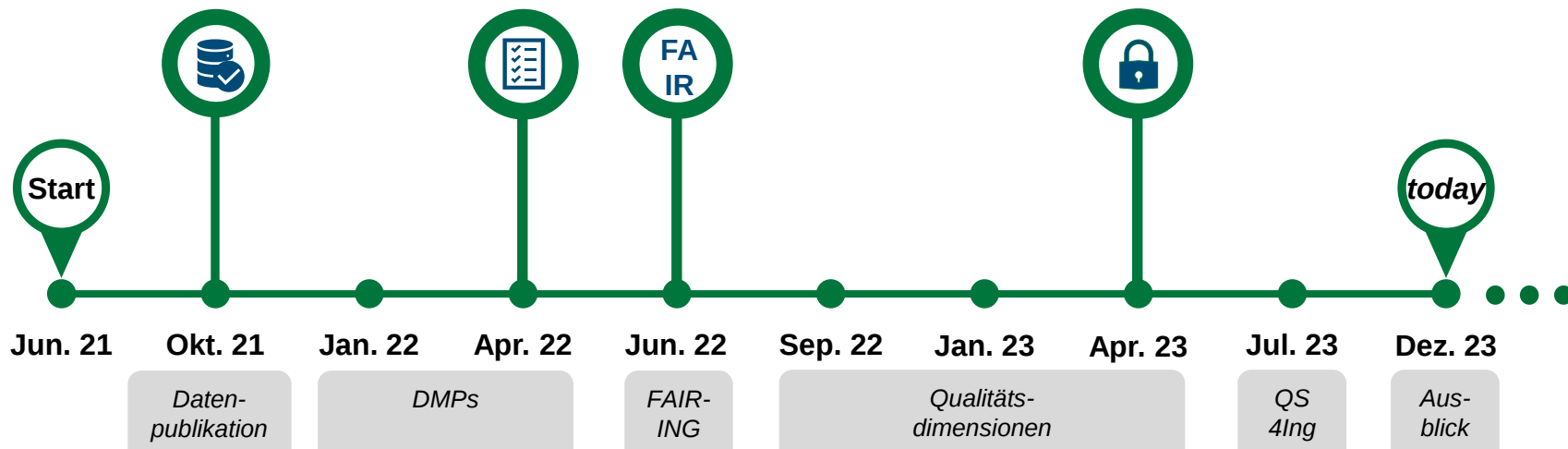


TA S-2

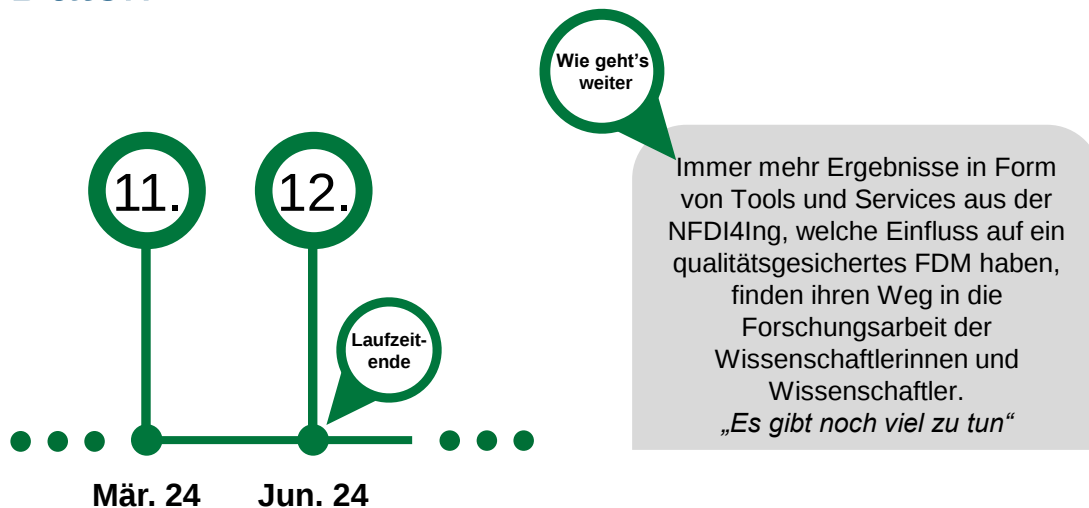
QUALITY ASPECTS FOR CREATED SOFTWARE

„Quality Assurance“: Vorlagen und Best-Practice-Beispiele, die es ermöglichen, **Qualitätsmetriken zu erstellen**, um die Nachhaltigkeit und **Wiederverwendbarkeit** von **Quellcodes** für Forscher greifbar zu machen. Kontinuierliche und **automatische Generierung** von **Qualitätsmetriken**, um Software während der Entwicklung zu verbessern und die **Qualität** bestehenden Codes zu **beurteilen**, mit der zusätzliche **Metadaten** aus Quellcodes zu **extrahieren**

today: Rückblick & Ausblick der SIG QA&metrics



Ausblick: SIG Qualitätssicherung und Metriken für FAIR-Daten



Vielen Dank für die Zusammenarbeit!



Eine besinnliche Weihnachtszeit wünscht
das SIG-Team

- ● contact@nfdi4ing.de
- ● Mailinglist: sig-qa@nfdi4ing.de
- ● Homepage: <https://nfdi4ing.de/community-hub/special-interest-groups-sig/qa-metrics/>

Nächstes SIG-Meeting –
Mär 24