



NFDI4ing

Special Interest Group (SIG) „Qualitätssicherung und Metriken
für FAIR-Daten“

FAIR in den Ingenieurwissenschaften

5. Treffen, 10.06.2022

● Die FAIR Prinzipien: Quiz und Übersicht zum Einstieg

Mario Moser
Werkzeugmaschinenlabor WZL der RWTH Aachen
m.moser@wzl.rwth-aachen.de

Die vier *foundational / core* FAIR Principles

F

Findable



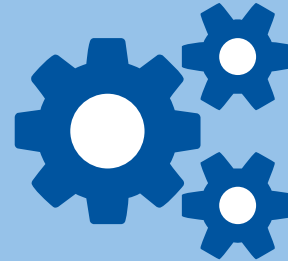
A

Accessible



I

Interoperable



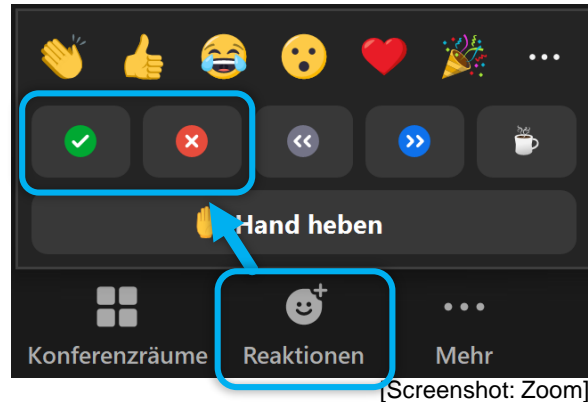
R

Reusable



Wahr oder Falsch? Quiz zur Entwicklung, den Zielen und dem Selbstverständnis von FAIR

Bitte benutzen Sie die „Reaktionen“ in Zoom, um Ihre Antwort zu nennen:



Die Aussage ist wahr.



Die Aussage ist falsch.

Wahr oder Falsch?

*Zum Warm werden: „FAIR“ steht hier für „Findable,
Accessible, Interoperable, Reusable“*

Wahr oder Falsch?

*Zum Warm werden: „FAIR“ steht hier für „Findable,
Accessible, Interoperable, Reusable“*



Stimmt genau!

Wahr oder Falsch?

So wurde es: Die FAIR Prinzipien wurden 2014 bei
einem Workshop in Leiden (Niederlande)
entwickelt und 2016 im Nature Magazin
veröffentlicht

Wahr oder Falsch?

So wurde es: Die FAIR Prinzipien wurden 2014 bei einem Workshop in Leiden (Niederlande) entwickelt und 2016 im Nature Magazin veröffentlicht



Die Aussage ist korrekt!

Das Paper zum Nachlesen:

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>



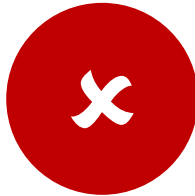
[<https://www.nature.com/articles/sdata201618>]

Wahr oder Falsch?

Darum geht's: Die FAIR Prinzipien beziehen sich ausschließlich auf erhobene Daten (DLC Phase 2) sowie nachgenutzte Daten (DLC Phase 6) wie Messwerte oder Simulationsdaten

Wahr oder Falsch?

Darum geht's: Die FAIR Prinzipien beziehen sich ausschließlich auf erhobene Daten (DLC Phase 2) sowie nachgenutzte Daten (DLC Phase 6) wie Messwerte oder Simulationsdaten



Das stimmt so nicht.

Die FAIR Prinzipien sind für ‚scholar digital objects‘, was neben Daten im engeren Sinne auch Repositorien, Workflows, Algorithmen, Vokabulare, Infrastrukturen etc. umfasst:

„the principles apply not only to ‘data’ in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects – from data to analytical pipelines – benefit [...]“ [Wilkinson et al. (2016), S. 1]

Beachte: Im Kontext von FAIR findet sich häufig die Bezeichnung „Daten“ synonym für Digital Objects.

Wahr oder Falsch?

Dafür ist es da: Die Ziele der FAIR Prinzipien sind „transparency, reproducibility, and reusability“ zu erhöhen

Wahr oder Falsch?

Dafür ist es da: Die Ziele der FAIR Prinzipien sind „transparency, reproducibility, and reusability“ zu erhöhen



Das ist richtig!

Ziel ist die langfristige Verbesserung der Wiederverwendbarkeit von Forschungsdaten. Dazu sind die Transparenz, die Reproduzierbarkeit sowie die Wiederverwendbarkeit genannt.

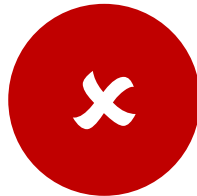
Das Zitat stammt aus Wilkinson et al. (2016, S. 1).

Wahr oder Falsch?

Pflicht: Die FAIR Prinzipien sind als Standard entworfen, den es unbedingt einzuhalten gilt.

Wahr oder Falsch?

Pflicht: Die FAIR Prinzipien sind als Standard entworfen, den es unbedingt einzuhalten gilt.



Das ist leider (glücklicherweise) nicht korrekt!

Die FAIR Prinzipien verstehen sich als „Guideline“, nicht jedoch als Standard, Spezifikation oder Implementierung.

[Wilkinson et al. (2016, S. 1), Mons et al. (2017, S. 51)]

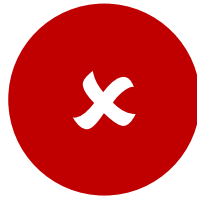
„FAIR principles are aspirational in nature, focusing on a long-term view of improving the potential for reuse of research data. As such, they should not be interpreted as strict rules.“ [Bahim et al. (2020), S. 5]

Wahr oder Falsch?

Ganz oder gar nicht: Alle FAIR Prinzipien müssen voll erfüllt werden, um FAIR zu sein.

Wahr oder Falsch?

Ganz oder gar nicht: Alle FAIR Prinzipien müssen voll erfüllt werden, um FAIR zu sein.



Das stimmt nicht.

Die FAIR Prinzipien blicken nicht binär / schwarz-weiß auf die digitalen Objekte, sondern als kontinuierliche Verbesserung.

So spricht bspw. die EOSC (2021, S. 19) von „binary judgements avoided“

Wahr oder Falsch?

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

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“ [Bahim et al. (2020), S. 5]

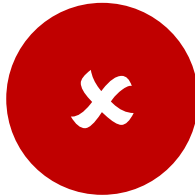
Eine exemplarische Metrik werden wir heute noch kennen lernen.

Wahr oder Falsch?

*Perfect FDM: Wenn ich FAIR komplett einhalte,
habe ich perfektes FDM.*

Wahr oder Falsch?

*Perfect FDM: Wenn ich FAIR komplett einhalte,
habe ich perfektes FDM.*



Leider nein! (Auch wenn es schön wäre.)

Die FAIR Prinzipien dienen zwar zur Erhöhung der Wiederverwendbarkeit von Forschungsdaten und zum ‚guten‘ Datenmanagement, doch decken sie nicht alle Aspekte des FDMs ab. FAIR berücksichtigt bspw. nicht die Planung des Forschungsvorhabens oder die Datensicherungsstrategie oder Aspekte der Datenqualität.

FAIR kann nur ein Bestandteil der Quality Assurance (QA) sein.

Wahr oder Falsch?

Nicht in Stein gemeißelt: Die FAIR Metriken können sich zukünftig ändern und angepasst werden.

Wahr oder Falsch?

Nicht in Stein gemeißelt: Die FAIR Metriken können sich zukünftig ändern und angepasst werden.



Die Aussage ist korrekt!

Zukünftige Entwicklungen und Erfordernisse können zu einer Anpassung der FAIR Metriken führen.

„The Principles define characteristics that contemporary data resources, tools, vocabularies and infrastructures should exhibit“ [Wilkinson et al. (2016), S. 4]

„As technology as well as the community requirements will evolve over time, it is crucial to revisit the FAIR principles and the corresponding indicators to better interpret and implement them.“ [Devaraju et al. (2021), S. 12]

Wahr oder Falsch?

Von Beginn an: Die FAIR Prinzipien sollten bereits
frühzeitig berücksichtigt werden und nicht erst
gegen Ende des Forschungsprojekts

Wahr oder Falsch?

Von Beginn an: Die FAIR Prinzipien sollten bereits frühzeitig berücksichtigt werden und nicht erst gegen Ende des Forschungsprojekts



Das ist richtig (und wichtig)!

Die eigenen digital objects sollten schon früh FAIR gestaltet werden bzw. die FAIRness kontinuierlich gesteigert werden.

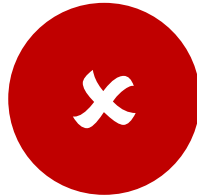
FAIR Aspekte, die bereits die Datenerhebung betreffen, lassen sich nur schwerlich oder gar nicht im Nachhinein bei der Publikation anpassen.

Wahr oder Falsch?

Lesbarkeit: Im Zeitalter der Digitalisierung fokussiert sich FAIR ausschließlich auf die Maschinenlesbarkeit von Daten.

Wahr oder Falsch?

Lesbarkeit: Im Zeitalter der Digitalisierung fokussiert sich FAIR ausschließlich auf die Maschinenlesbarkeit von Daten.



Nein, nicht ausschließlich.

FAIR zielt darauf ab, dass Daten sowohl für Menschen als auch Maschinen auffindbar, zugänglich, interpretierbar und verarbeitbar sind.

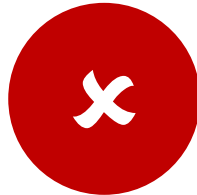
"improve knowledge discovery through assisting both humans, and their computational agents"; "The emphasis placed on FAIRness being applied to both human-driven and machine-driven activities" [Wilkinson et al. (2016), S. 3]

Wahr oder Falsch?

*Faire Offenheit: FAIRe Daten sind auch immer
offene Daten (open data).*

Wahr oder Falsch?

*Faire Offenheit: FAIRe Daten sind auch immer
offene Daten (open data).*



Nein, das ist ein häufiges Missverständnis.

Daten können auch dann FAIR sein, wenn sie vor unautorisiertem Zugriff geschützt werden.

„openess is not a requirement of FAIRness“

[Wilkinson et al (2018), S. 3]

„The ‘A’ in FAIR stands for ‘Accessible under well defined conditions’.“

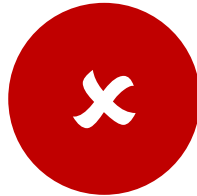
[Mons et al. (2017), S. 51]

Wahr oder Falsch?

Benchmarking: Primäres Ziel der FAIR Prinzipien
ist es, mit anderen vergleichen zu können.

Wahr oder Falsch?

Benchmarking: Primäres Ziel der FAIR Prinzipien
ist es, mit anderen vergleichen zu können.



Nein, nicht unbedingt.

Die FAIR Prinzipien dienen dazu, ein eigenes ‚gutes‘ Datenmanagement zu betreiben und zu verbessern. Der Vergleich mit anderen ist zwar denkbar, aber nicht das primäre Ziel.

Dies setzt zudem voraus, dass die gleichen Metriken zwischen den zu vergleichenden Digital Objects verwendet worden sind.

Literatur

- Wilkinson, M.; Dumontier, M.; Aalbersberg, I. J.; *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3: 160018
<https://doi.org/10.1038/sdata.2016.18>
- Mons, B.; Neylon, C.; Velterop, J.; *et al.* (2017). Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* 37: 1, S. 49–56
<https://doi.org/10.3233/ISU-170824>
- Bahim, C.; Casorrán-Amilburu, C.; Dekkers, M.; *et al.* (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, 19: 41, S. 1–7
<https://doi.org/10.5334/dsj-2020-041>
- Devaraju, A.; Mokrane, M.; Cepinskas, L.; *et al.* (2021). From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. *Data Science Journal*, 20: 4
<https://doi.org/10.5334/dsj-2021-004>
- European Commission, Directorate-General for Research and Innovation; Aronsen, J.; Beyan, O.; Harrower, N.; *et al.* (2021). Recommendations on FAIR metrics for EOSC. Jones, S. (Editor); Genova, F. (Editor), Publications Office
<https://doi.org/10.2777/70791>

Die vier *foundational / core* FAIR Principles und ihre zugehörigen 15 FAIR *guiding* Principles

F

Findable



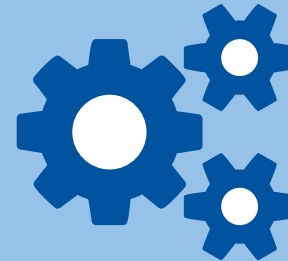
A

Accessible



I

Interoperable



R

Reusable



Die 15 FAIR *Guiding* Prinzipien: Findable

F

Findable



F1. (meta)data are assigned a **globally unique and persistent identifier**

→ Auffindbarkeit anhand von Identifikatoren

F2. data are described with **rich metadata**
(defined by R1)

→ Auffindbarkeit von Daten anhand von Metadaten

F3. metadata clearly and explicitly **include the identifier** of the data it describes

→ Link von Metadaten zu Daten über Identifikator

F4. (meta)data are registered or indexed in a **searchable resource**

→ In durchsuchbarer Ressource

[Wilkinson et al., 2016, p. 4]

Die 15 FAIR *Guiding* Prinzipien: Accessible

A

Accessible



A1. (meta)data are retrievable by their identifier using a **standardized communications protocol**

A1.1. the **protocol** is open, free, and universally implementable

→ Daten über offenes, freies und universelles Kommunikationsprotokoll zugänglich

A1.2. the protocol allows for an **authentication and authorization** procedure, where necessary

→ Authentifizierung und Authorisierung wird angeboten und unterstützt

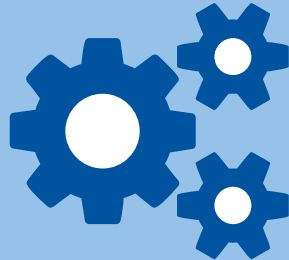
A2. metadata are accessible, even when the data are **no longer available**

→ Metadaten bleiben verfügbar

Die 15 FAIR *Guiding* Prinzipien: Interoperable



Interoperable



11. (meta)data use a formal, accessible, shared, and broadly applicable **language for knowledge representation**

→ Wissen wird zugänglich und strukturiert repräsentiert

12. (meta)data use **vocabularies** that follow FAIR principles

→ Begriffe werden in FAIRen Vokabularen erklärt

13. (meta)data include **qualified references to other (meta)data**

→ Daten sind untereinander durch Verlinkungen vernetzt

Die 15 FAIR *Guiding* Prinzipien: Reusable

R

Reusable



R1. meta(data) are richly described with a plurality of accurate and relevant **attributes**

→ Beschreibung durch Attribute

R1.1. (meta)data are released with a clear and accessible **data usage license**

→ Lizenzangabe (mensch- und maschinenlesbar)

R1.2. (meta)data are associated with **detailed provenance**

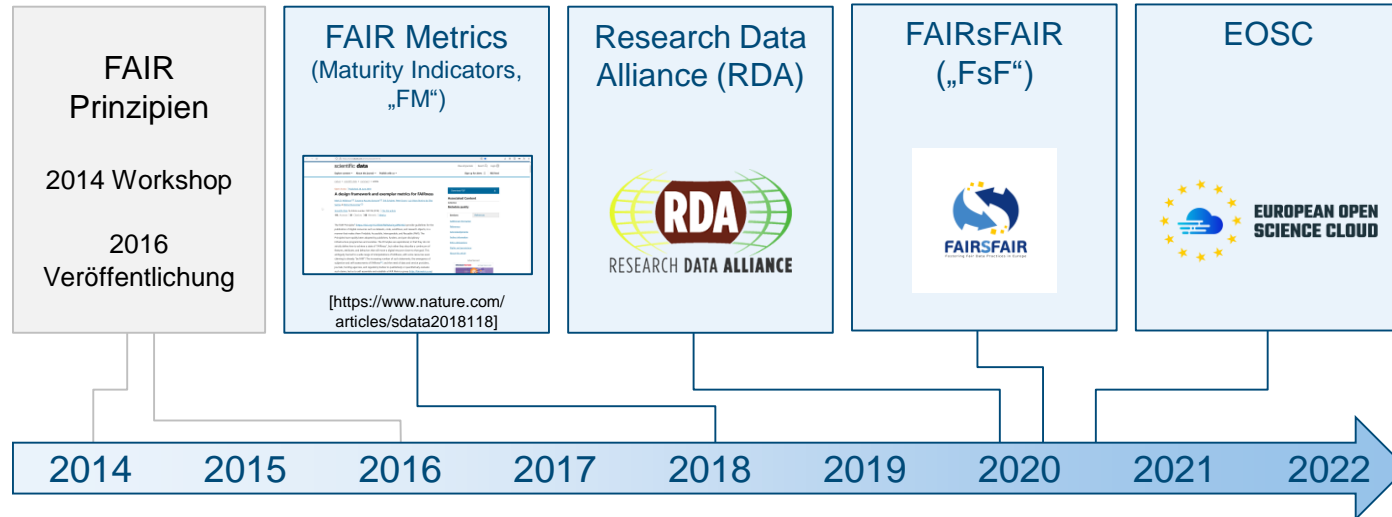
→ Herkunft und Verarbeitungsschritte der Daten sind nachvollziehbar (auch: Data Lineage)

R1.3. (meta)data meet **domain-relevant community standards**

→ Community-Standards werden angewendet [Wilkinson et al., 2016, p. 4]

Umsetzung der FAIR Prinzipien in Metriken

- Interpretationsspielraum der Prinzipien, Präzision von Begriffen
- Anpassungsmöglichkeit für spezielle Communities, kontinuierlicher Austausch
- Auswahl gängiger Metriken:



Vielen Dank!

Mario Moser

Werkzeugmaschinenlabor WZL der RWTH Aachen

m.moser@wzl.rwth-aachen.de