

Ist 'FAIR Data' das neue 'Open Data'?

Ein Konferenzbericht zum „FAIR Convergence Symposium“,
27. November bis 4. Dezember 2020

Julian Dederke

Das Management, die Archivierung und das Publizieren von Daten sind aus dem Alltag von Forschenden und Bibliotheken an Forschungseinrichtungen nicht mehr wegzudenken. Dabei müssen Prinzipien des Aufbewahrens, Kuratierens und Zugänglichmachens mit den mittlerweile allgegenwärtigen Anforderungen an offene Wissenschaft (Open Science) und offene Daten (Open Data) in Einklang gebracht werden. Hinzu kommt, dass sich zunehmend nicht mehr nur das Prinzip offener Daten, sondern vielmehr FAIRer Daten durchsetzt. FAIR bezieht sich dabei auf

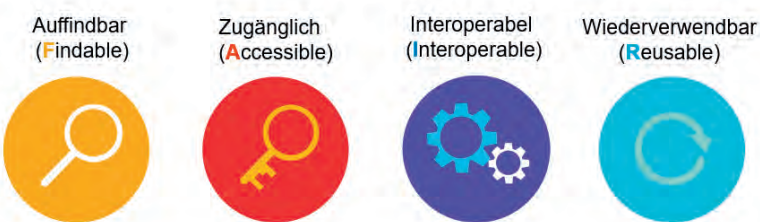


Abbildung 1: Die FAIR-Prinzipien

Grafik: Paulina Halina Sieminska / Bearbeitung: Dr. Ilona Lang / CC BY-SA 4.0

die Eigenschaften *findable* (auffindbar), *accessible* (zugänglich), *interoperable* (interoperabel), und *reusable* (wiederverwendbar),¹ die als Gütekriterien für Forschungsdaten im 21. Jahrhundert herhalten.²

Um auffindbar (*Findable*) zu sein, müssen Datensätze u.a. durch einen einzigartigen, persistenten Identifikator global abrufbar sein (z.B. durch einen DOI), und an einem öffentlich zugänglichen und durchsuchbaren Ort hinterlegt sein (i.d.R. in einem digitalen Repository). Um zugänglich (*Accessible*) zu sein, müssen Datensätze u.a. sowohl von Menschen als auch Maschinen über einen offenen, frei verfügbaren Kommunikationsweg oder ein -protokoll erreichbar sein. Um

interoperabel (*Interoperable*) zu sein, müssen Datensätze außerdem durch eine formale, zugängliche, verallgemeinerbare Sprache beschrieben sein, die standardisiertes Vokabular nutzt, und die den Datensatz mit anderen, verwandten Informationen verknüpft. Um wiederverwendbar (*Reusable*) zu sein, müssen Datensätze mit möglichst vielen beschreibenden Attributen und Informationen angereichert werden, mit einer klar definierten und zugänglichen Lizenz versehen sein, und den formalen Standards einer bestimmten wissenschaftlichen Community entsprechen.

Wie sich hier bereits zeigt, sind die *FAIR-Prinzipien* deutlich detaillierter und an mehr Voraussetzungen geknüpft als das Prinzip *Open Data*.³ Nicht ohne Grund also erwartete mich vom 27. November bis 4. Dezember 2020 ein umfangreiches Programm beim *FAIR Convergence Symposium*.⁴ Die Online-Konferenz wurde organisiert vom CODATA-Komitee des Internationalen Wissenschaftsrats⁵ und der GO FAIR-Initiative.⁶ Für mich zeichneten drei übergreifende Themen die mit virtuellen Panels, aufgezeichneten Vorträgen und interaktiven Online-Workshops gespickte Konferenz aus.

Die Internationalität der Anliegen rund um offene und FAIRe Wissenschaft

Das erste übergreifende Thema, *Internationalität und Globalität*, war nicht nur daran erkennbar, dass Vortragende und Zuhörer/-innen von überall auf dem Globus kamen, sondern auch in dem Anspruch, international oder sogar global einheitliche Standards anzustreben. Nur wenn international vernetzte Wissenschaft sich auf gemeinsame Standards einigt, können diese sich langfristig durchsetzen. Dementsprechend oft tauchte das Wort „Konvergenz“ aus dem Titel des Symposiums auch in Vorträgen und Diskussio-

1 <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

2 Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9, <https://www.doi.org/10.1038/sdata.2016.18>.

3 <https://opendefinition.org/>

4 Eine Übersicht der Beiträge und aufgezeichneten Vorträge findet sich auf der Konferenzwebseite <https://conference.codata.org/FAIRconvergence2020/programme/>

5 <https://codata.org/>

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

nen auf. Sprecherinnen und Gäste von UNESCO und Internationalem Wissenschaftsrat, Datenmanager, Wissenschaftlerinnen und Praktiker aus der ganzen Welt brachten ihr Know-how zum Thema FAIRe Daten und Wissen zu ihren Potenzialen und Hürden mit in den virtuellen Konferenzspace. Schnell wurde klar, dass sich durch den Anspruch, internationale Standards zu etablieren, eine Menge regulative Hürden und Fragen ergeben: Wie einigt man sich auf die beste Praxis? Woher kommt eigentlich die Autorität etablierter Standards innerhalb von Disziplinen? Wer sorgt dafür, dass Wissenschaftler/-innen diese Standards einhalten? Spätestens als in den Titeln von Panels 2 und 3 der Begriff „Globale Konvergenz“ auftauchte, stand für mich fest: Hier liegen große Aufgaben vor der GO-FAIR-Gemeinde, Bibliotheken, Forschungsinstitutionen, -dienstleister/-innen und allen anderen Beteiligten!

Die COVID-19-Pandemie im Kontext von offenen und FAIRen Daten

Thematisch eng mit dem Motiv der Internationalität verknüpft war als zweites Thema *COVID-19*. Das lag nicht nur wegen des globalen Ausmaßes der Pandemie nahe, sondern auch dadurch, dass die Prinzipien von Open und FAIR Data eine nie dagewesene Geschwindigkeit der Erforschung von COVID-19 ermöglichten. Die „Coronafizierung“ der Welt und die abertausenden Studien, die seit Anfang 2020 dazu erschienen sind, stellen gewissermaßen den Effekt von offener Wissenschaft und offenen Daten exemplarisch zur Schau. So unterstrich etwa der Vortrag von Prof. Flahault, Direktor des Institute of Global Health an der Universität Genf, dass die schnelle Verfügbarkeit, die mit offenen COVID-19-Daten einhergeht,

mittlerweile eine 7-Tages-Prognose der COVID-Fälle für alle Länder mit zuverlässiger Datenlage zulässt. Der gewählte Präsident des Internationalen Wissenschaftsrates, Prof. Peter Gluckman, betonte, dass die Prozesse der Risikobewertung durch Regierungen gegenwärtig oft schwach ausgeprägt seien. Es bedürfe sowohl besserer Daten als auch Infrastrukturen, die hohen und einheitlichen Standards genügen müssten – ein eindrücklicher Hinweis darauf, dass die Umsetzung der FAIR-Prinzipien neben der Bedeutung für die Forschung auch unmittelbare Auswirkungen auf unser gesellschaftliches Zusammenleben haben dürfte. Zugleich schätzte einer der Konferenzorganisatoren, dass FAIRe, maschinenlesbare Daten in vielen Ländern eine um drei Monate schnellere Antwort auf die COVID-Krise ermöglicht hätten. Dies verdeutlicht exemplarisch die großen Unterschiede, die zwischen den Prinzipien Open und FAIR Data liegen: Während die freie Verfügbarkeit von Daten zumindest im Falle von COVID bereits gut gewährleistet ist, erfordern die FAIR-Prinzipien z.B. explizit auch unmittelbare Lesbarkeit von Daten und Metadaten durch Mensch und Maschine. FAIRe Daten und FAIRes Datenmanagement zu gewährleisten, ist an deutlich mehr Voraussetzungen geknüpft als Daten „nur“ offen zugänglich zu machen.

Die Rolle von Data Stewards: Kompetenzen, Rollenmodelle, Finanzierung und Schulung

Als drittes großes Thema erwies sich der Trend zu Data Stewardship. Data Stewardship beschreibt eine neue Rolle, die es an Forschungsinstitutionen zukünftig auszufüllen gilt, um Anforderungen an FAIRe Daten gerecht zu werden. Betont wurde in diesem Zusammenhang, dass Nachwuchsforscher/-innen heutzuta-

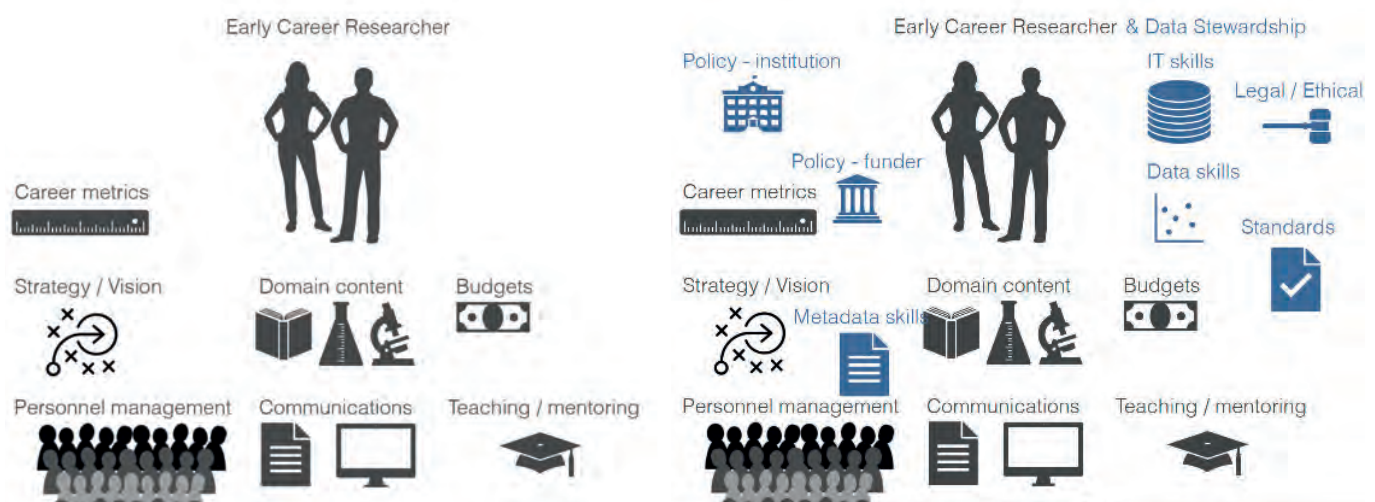


Abbildung 2: Steigende Anforderungen an Nachwuchswissenschaftler/-innen CC-BY 4.0 International, Erik Schultes <https://osf.io/yhu85/>

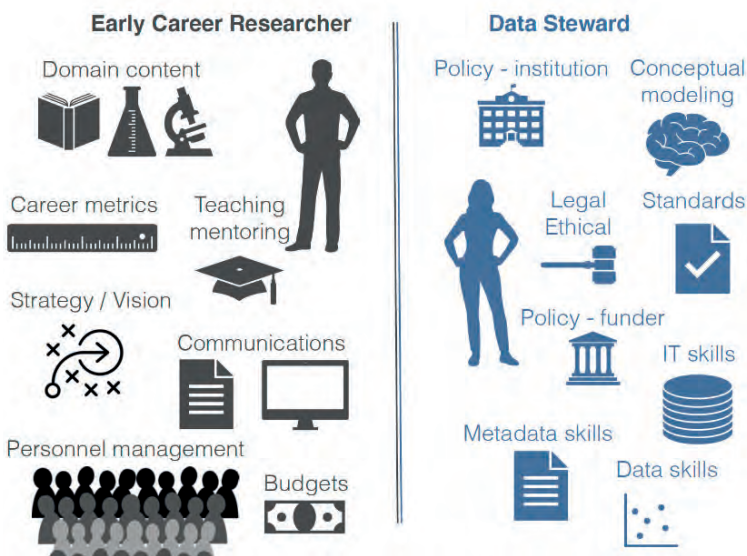


Abbildung 3: Data Stewardship als eigenständige Funktionsbeschreibung
CC-BY 4.0 International, Erik Schultes <https://osf.io/yhu85/>

ge mit zahlreichen zusätzlichen Aufgaben konfrontiert sind, die über ihre eigentliche Tätigkeit als Forscher/-in hinausgehen. So kämen etwa zu den unmittelbar forschungsbezogenen Aufgaben in den letzten Jahren immer mehr Anforderungen in den Bereichen IT, Policy-Kenntnisse, rechtliche und ethische Bedingungen und Datenmanagement hinzu. Im Panel „Data stewardship competencies, skills and training“ wurde deshalb auch dafür plädiert, die Data Stewardship-Rolle als eigenständige Funktion zu etablieren. Eine Illustration von Erik Schultes (International Science Coordinator beim GO FAIR International Support and Coordination Office) machte dies besonders anschaulich (Abbildungen 2 und 3).

Der Trend zu Data Stewardship schlug sich nieder in umfangreicher Diskussion zu Rollenbeschreibung und Ausbildungscurriculum sowie zur Frage, wo an einer Institution die Funktion etabliert werden soll. Zur Diskussion stand etwa die Ansiedlung an zentralen Institutionen wie Bibliotheken, Archiven oder Informatikdiensten, oder aber dezentral an Instituten und Departementen. Für zentrale Institutionen sprachen dabei u.a. die leichtere Vereinheitlichung von Standards, Rollenbeschreibungen und Curricula. Für eine dezentrale Ansiedlung sprach u.a. die Nähe zu den Wissenschaftler/-innen in den Instituten und Teams vor Ort, und die Rolle von disziplinspezifischem Wissen. Als gemeinsamer Nenner erschien bei allen Ansätzen und Vortragenden die Überzeugung, dass eine konsequente Umsetzung der FAIR-Prinzipien nur ge-

linge, wenn diese Aufgabe explizit durch qualifiziertes Fachpersonal unterstützt würde.

Modellcharakter für den Umgang mit neuen Anforderungen nahm das an der TU Delft etablierte Data Stewardship-Programm ein.⁷ Ähnlich befassen sich mittlerweile zahlreiche Universitäten mit der Zielstellung, Data Stewards an ihren Institutionen zu etablieren. So berichteten etwa Vertreterinnen der Allianz führender Technischer Universitäten in Deutschland (TU9)⁸ von einem gemeinsamen Arbeitspapier zur Koordination der Rolle und Struktur von Data-Stewardship-Programmen, das mit der Leitung der jeweiligen Universitätsbibliotheken besprochen werden sollte. Neben eindrucksvollen Beispielen, wie das Konzept Data Stewardship mitunter bereits eingeführt wird, bleibt für mich auch die Gewissheit, dass das Thema diejenigen, die sich mit dem Management, Archivieren und Publizieren von Daten auseinandersetzen, weiter beschäftigen wird.

Fazit

Ist also, zusammengefasst, *FAIR Data* das neue *Open Data*? Ein Trend, der einen anderen ersetzt? Um den Versuch einer kurzen Antwort zu unternehmen: Die FAIR-Prinzipien sind insgesamt an mehr Voraussetzungen geknüpft und haben an Bedeutung gewonnen. Das Prinzip der offenen Daten erscheint nicht mehr als ausreichend und die Gütekriterien für Forschungsdaten und Wissenschaft der Zukunft sind nun vielmehr assoziiert mit den Lettern F-A-I-R. ■



Dr. Julian Dederke

ETH-Bibliothek Zürich
Forschungsdatenmanagement und
digitaler Datenerhalt
julian.dederke@library.ethz.ch

⁷ <https://zenodo.org/communities/datastewards>

⁸ <https://www.tu9-universities.de/>