

Arvid Deppe

FAIR, CARE und mehr

Prinzipien für einen verantwortungsvollen Umgang mit Forschungsdaten

In der Präambel der Forschungsdaten-Leitlinie der Universität Kassel wird der Kern guten Datenmanagements folgendermaßen gefasst:

„Der verantwortungsvolle Umgang mit Forschungsdaten ist für die Nachvollziehbarkeit von Forschungsergebnissen sowie für die Bewahrung und Generierung wissenschaftlicher Erkenntnisse unerlässlich.“¹

Mit ihrem Verweis auf einen verantwortungsvollen Umgang steht die Universität Kassel nicht alleine; 32 von 54 institutionellen Policies weisen eine vergleichbare Formulierung auf.² Wie aber ist ‚verantwortungsvoll‘ in diesem Zusammenhang zu verstehen? Um einer Antwort auf diese Frage näher zu kommen, sollen im Folgenden verschiedene Prinzipien und Maßnahmen in den Blick genommen werden, die jeweils auf unterschiedliche Aspekte der Verantwortung im Kontext des Forschungsdatenmanagements abzielen. Die dabei aufgezeigte Vielschichtigkeit des Verantwortungsbegriffs weist zugleich auf mögliche Anforderungen an Infrastruktureinrichtungen hin.

Als Ausgangspunkt soll die ‚gute wissenschaftliche Praxis‘ dienen, auf welche die Präambel in ihrem zweiten Satz Bezug nimmt. Die DFG schloss 1998 mit der Denkschrift zur Sicherung guter wissenschaftlicher Praxis³ vor dem Hintergrund aktueller Fälle wissenschaftlichen Fehlverhaltens eine

¹ Forschungsdaten-Leitlinie der Universität Kassel, 13. Januar 2017. Online: <<http://uni-kassel.de/go/fdm/policy>>, Stand: 10. Mai 2020.

² Berücksichtigt wurden alle gelisteten institutionellen Policies unter: Forschungsdaten.org – Data Policies, <https://www.forschungsdaten.org/index.php/Data_Policies>, Stand: 2. Juni 2020. In drei der Fälle wurde der Begriff ‚verantwortungsbewusst‘ verwendet.

³ Deutsche Forschungsgemeinschaft (DFG): Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission «Selbstkontrolle in der Wissenschaft», Weinheim 1998.

diesbezügliche Regelungslücke.⁴ Die Aufbewahrung von ‚Primärdaten‘ ist als kleiner Baustein innerhalb dieser Regelungen zu lesen.

Gut 20 Jahre später nehmen Forschungsdaten im Kodex ‚Leitlinien zur Sicherung guter wissenschaftlicher Praxis‘⁵ einen weitaus prominenteren Platz ein. Sie werden eindeutig als wissenschaftliche Leistung gewürdigt [Leitlinie (L) 14] und als integraler Bestandteil des gesamten Forschungsprozesses mitgedacht, etwa bei Fragen der Qualitätssicherung [L7], der Rechtssicherheit [L10], der Standardisierung von Methoden [L11], der Dokumentation [L12] der Veröffentlichung [L13 und L15] und der Archivierung [L17]. Insbesondere in der „Herstellung von öffentlichem Zugang zu Forschungsergebnissen“ [L13] wird deutlich, dass die Bedeutung des richtigen Umgangs mit Daten nicht bei der Vermeidung von Datenfälschung aufhört, sondern mittlerweile auch die Nachnutzung in den Fokus gerückt ist:

„Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit der Forschung und Nutzbarkeit 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.“⁶

Hinter den hier geforderten FAIR-Prinzipien verbirgt sich ein Set von Leitprinzipien und Praktiken, das mittlerweile zur Standard-Anforderung im Datenmanagement geworden ist.

⁴ Vgl. Riescher, Gisela; Haas, Tobias: Verbindlich und kompakt: Der neue DFG-Kodex «Leitlinien zur Sicherung guter wissenschaftlicher Praxis», in: Ordnung der Wissenschaft 01, 2020, S. 33.

⁵ Vgl. Deutsche Forschungsgemeinschaft (DFG): Leitlinien zur Sicherung guter wissenschaftlicher Praxis: Kodex, 2019. Online: <https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf>, Stand: 19. März 2020.

⁶ Ebd., S. 19.

Die FAIR-Prinzipien

Die Forderung, Daten leichter auffindbar und interpretierbar aufzubereiten, war bereits Gegenstand verschiedener Policies⁷ bevor sie 2016 in den FAIR-Prinzipien gleichermaßen griffig akronymisiert wie überzeugend konkretisiert wurde. Diese Prinzipien wurden 2014 von einer Arbeitsgruppe innerhalb der FORCE11 erarbeitet und sind seit ihrer Veröffentlichung⁸ unverändert geblieben. Innerhalb von kurzer Zeit wurden sie breit rezipiert und implementiert und bilden eine wichtige Grundlage etwa für Datenservices, fachspezifische Leitlinien und Forschungsinfrastrukturen. Unter anderem sprachen sich 2016 die G20 für ihre Anwendung in der Forschung aus⁹ und 2017 wurden sie mit dem Open Data Pilot von der Europäischen Kommission für das Förderprogramm Horizon 2020 verbindlich eingeführt.¹⁰ Aktuell stellen sie zudem einen paradigmatischen Kern der designierten European Open Science Cloud (EOSC) dar¹¹.

FAIR steht für eine Reihe von technischen und organisatorischen Prinzipien und Maßnahmen, die sicherstellen sollen, dass Daten durch

⁷ Vgl. European Commission: Turning FAIR into reality, Publications Office of the European Union, 2018. Online: <<http://doi.org/10.2777/1524>>.

⁸ Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u. a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific Data 3 (160018), 12.2016. Online: <<https://doi.org/10.1038/sdata.2016.18>>.

⁹ Vgl. G20 endorse the FAIR principles, Dutch Techcentre for Life Sciences, 13. September 2016, <<https://www.dtls.nl/2016/09/13/g20-endorse-fair-principles/>>, Stand: 19. Mai 2020.

¹⁰ Vgl. European Commission: Guidelines on FAIR Data Management in Horizon 2020, 26 July 2016. Online: <https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>, Stand: 8. April 2020.

¹¹ Vgl. Mons, Barend; Neylon, Cameron; Velterop, Jan u. a.: Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud, in: Information Services & Use 37 (1), 7. März 2017, S. 49–56. Online: <<https://doi.org/10.3233/ISU-170824>>. In Deutschland und darüber hinaus sieht sich allen voran die Bottom-Up-Initiative GO FAIR als Beitrag zu einer EOSC auf Basis der FAIR-Prinzipien (vgl. How does GO FAIR relate to the European Open Science Cloud and the Internet of FAIR Data and Services?, GO FAIR, <<https://www.go-fair.org/faq/go-fair-relation-to-eosc-and-ifds/>>, Stand: 22. Mai 2020).

Weiterverwendung ihren maximalen Wert entfalten können.¹² Hierfür müssen Forschungsdaten auffindbar (findable) und zugänglich (accessible) sein, miteinander verknüpft (interoperable) und nachgenutzt (re-useable) werden können – von Menschen und Maschinen. Die Grundprobleme, die dabei adressiert werden, sind in der Bibliothekspraxis nicht neu:

Auffindbarkeit: Ein Buch, das nicht im Katalog verzeichnet, schlecht verschlagwortet oder falsch im Regal einsortiert ist, ist im Grunde nicht nutzbar. Ähnlich verhält es sich bei Forschungsdaten. Die wichtigsten Maßnahmen, um Daten für Mensch und Maschine auffindbar zu machen, zielen hier auf die Beschreibung der Daten durch Metadaten, die persistente Adressierung und gegenseitige Referenzierung von Daten und Metadaten sowie die Registrierung und Indizierung von Metadaten ab.

Zugänglichkeit: Einrichtung und Etablierung von Selbstbedienungsfunktionalitäten haben in Kassel in den letzten Jahren die Zugänglichkeit zu gedruckten Informationen erhöht. Und auch wenn Bücher sich im Magazin oder an anderen Standorten befinden, stellen standardisierte Verfahren wie interner Leihverkehr, Magazinbestellung oder Fernleihe den Zugang sicher. Bei Daten sind diese standardisierten Wege unter anderem freie und offene Protokolle zum Zugriff auf die Daten, die – wo nötig – auch Rechteverwaltung erlauben. Die Metadaten bleiben verfügbar, auch wenn die Daten selbst es nicht mehr sein sollten.

Interoperabilität: Mensch und Maschine gleichermaßen erkennen zu lassen, um was es sich formal und inhaltlich handelt, ob etwa zwei namensgleiche Autorinnen auch dieselbe Person sind oder ob ein Buch eine Neuauflage, Übersetzung, Fortsetzung oder unabhängig von einem anderen Buch ist, stellt eine bekannte Herausforderung im Bereich der Katalogdaten dar. Wie bei der inhaltlichen und sachlichen Erschließung

¹² „FAIR refers to a set of principles, focused on ensuring that research objects are reusable, and actually will be reused, and so become as valuable as is possible“ (Mons u. a.: Cloudy, 2017, S. 50).

von Büchern sollten bei Daten Normdateien / kontrollierte Vokabulare, Ontologien oder Vergleichbares ebenso wie Referenzierungen und formale Modelle wie RDF zum Einsatz kommen, um sie im Sinne (semi-)automatisierter Optionen, Daten auszutauschen und zu kombinieren, zu beschreiben. Da die Beschreibungsgrundlagen vielfältig und dynamisch sein können, ist es besonders wichtig, auch diese selbst FAIR im besten Sinne des Akronyms zu halten.

Wiederverwendbarkeit: Die Nachnutzung von Daten erfordert in der Regel Informationen über diese Daten und ihren Entstehungskontext. Daher wird im Datenmanagement üblicherweise zu einer möglichst umfassenden Datenbeschreibung – gerne in einer separaten Datei – geraten. Im Sinne der Nachnutzung durch Mensch und Maschine ist es allerdings mindestens ebenso wichtig, nutzungsrelevante Informationen in die Metadaten selbst zu schreiben. Welche Informationen hierunter fallen und wie sie zu erfassen sind, variiert abhängig von Daten und Disziplin. Deshalb sollte bei der Datenbeschreibung, sofern möglich, auf Community-Standards (wie Metadatenschemata und Vokabulare) zurückgegriffen werden. So fragt etwa das Metadatenschema von SowiDataNet | Datorium unter ‚*Methodology*‘ verschiedene Faktoren sozialwissenschaftlicher Forschungsdaten ab und der Crystallographic Information File (CIF) erfasst spezifische mess- und auswertungstechnische Metainformationen zu kristallografischen Daten, um nur zwei Beispiele zu nennen.¹³

Damit eine rechtssichere Nutzung gewährleistet ist, muss die Angabe der Nutzungsbedingungen (Lizenz) ebenfalls in den Metadaten hinterlegt sein.

¹³ Vgl. Jensen, Uwe; Linne, Monika: SowiDataNet – Metadatenschema Version 1.0, in: GESIS Papers, SSOAR – GESIS Leibniz Institute for the Social Sciences, 2017. Online: <<https://doi.org/10.21241/SSOAR.54975>> ; Bruno, Ian J.; Groom, Colin R.: A crystallographic perspective on sharing data and knowledge, in: Journal of Computer-Aided Molecular Design 28 (10), 10.2014, S. 1015–1022. Online: <<https://doi.org/10.1007/s10822-014-9780-9>>.

Die verschiedenen Dimensionen und Maßnahmen hinter den FAIR-Prinzipien legen nahe, dass FAIRness nicht als Schwarz-Weiß-Kategorie, sondern als Kontinuum zu verstehen ist. Um die jeweilige Position auf diesem Kontinuum zu bestimmen, hat 2018 die FAIR Metrics Group, bestehend aus Co-Autor*innen der FAIR-Prinzipien, einen Katalog von 14 beispielhaften Maßstäben erarbeitet, die alle FAIR-Sub-Prinzipien abdecken und so eine (Selbst-)Evaluation der FAIRness erlauben sollen.¹⁴

Wenngleich FAIR nicht automatisch ‚offen‘ bedeutet, fokussieren die FAIR-Prinzipien deutlich auf die potentielle Nachnutzbarkeit von Daten. Hierin steckt ein impliziter Verantwortungsbegriff, der – ganz allgemein gesprochen – darin besteht, den Wert von Daten zu erkennen und zu steigern.¹⁵ Diese Wertsteigerung liegt im abstraktesten Sinne in der Möglichkeit zum Erkenntnisgewinn. Welchem konkreten Zweck, wessen Gunsten oder Ungunsten dieser dienen könnte, ist kein Aspekt der FAIR-Prinzipien; sie fokussieren allein auf den optimalen Übergang der Daten von Datenproduzierenden auf Datennehmende.

Datenschutz und Forschungsethik

Nimmt man dagegen die Verantwortung gegenüber den Datengebenden in den Blick, landet man schnell bei rechtlichen Fragen insbesondere des Datenschutzes sowie bei ethischen Überlegungen.

Im Sinne des Datenschutzes bedeutet verantwortungsvolles Handeln allen voran die technische und organisatorische Umsetzung rechtlicher Regelungen. Diese umfasst zum Beispiel Maßnahmen der Zugriffsregelung oder Anonymisierung, aber auch die Sicherstellung einer informierten Einwilligung. Am Beispiel der neuerlichen Erweiterung der informierten Einwilligung um das Broad-Consent-Konzept, d.h. die Zustimmung zur späteren wissenschaftlichen Datennutzung (Erwägungsgrund 33 der EU-

¹⁴ Vgl. Wilkinson, Mark D.; Sansone, Susanna-Assunta; Schultes, Erik u. a.: A design framework and exemplar metrics for FAIRness, in: *Scientific Data* 5 (180118), 12.2018. Online: <<https://doi.org/10.1038/sdata.2018.118>>.

¹⁵ Vgl. Anm. 12.

DSGVO), wird exemplarisch deutlich, dass der Datenschutz zwar bereits viele ethische Anliegen regelt,¹⁶ seinerseits aber Leerstellen unter anderem für forschungsethische Überlegungen lässt.

Forschungsethik ist als ethischer Rahmen zu verstehen, innerhalb dessen die Gestaltung der Beziehung von Forschenden und Erforschten vor dem Hintergrund von Schaden und Nutzen reflektiert wird. Die zugrundeliegenden Prinzipien liegen vielfach auch in kodifizierter Form vor¹⁷ und adressieren unter anderem die Verantwortung für die Vermeidung psychischen und physischen Schadens der Beforschten sowie für den Schutz vor sozialen und ökonomischen Risiken, die mit dem Bekanntwerden der Daten verbunden wären.¹⁸ Zwar fragen forschungsethische Überlegungen auch nach dem Nutzen der Forschung unter anderem für die Beforschten¹⁹ allerdings primär in Abwägung gegen möglichen Schaden; insofern bedeutet Verantwortung im Kontext von Datenschutz und Forschungsethik mit Blick auf die Beforschten im Wesentlichen Schadensvermeidung.

Anders verhält sich dies bei den CARE-Prinzipien, die den Nutzen der Forschung für die Beforschten nicht lediglich als Abwägungsaspekt, sondern als wesentliches Motiv erfassen. Beforschte stehen dabei weniger als Individuen im Fokus, denn als Kollektiv.

¹⁶ Vgl. Rat für Sozial- und Wirtschaftsdaten (RatSWD): Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften, 2017, S. 14. Online: <<http://doi.org/10.17620/02671.1>>.

¹⁷ Vgl. Übersicht auf: VerbundFDB: Ethik in der Forschung: Ethik-Richtlinien von Berufsfachverbänden, <<https://www.forschungsdaten-bildung.de/ethik#ethik-richtlinien>>, Stand: 18. April 2020.

¹⁸ Vgl. Unger, Hella von; Narimani, Petra; M'Bayo, Rosaline (Hg.): Forschungsethik in der qualitativen Forschung: Reflexivität, Perspektiven, Positionen, Wiesbaden 2014.

¹⁹ Vgl. RatSWD: Forschungsethische Grundsätze, 2017, S. 18.

Die CARE-Prinzipien

Die *CARE Principles for Indigenous Data Governance*²⁰ gehen auf den Workshop ‚Indigenous Data Sovereignty Principles for the Governance of Indigenous Data‘ im Rahmen des 12. RDA Plenary Meetings in Gaborone, Bostwana zurück und wurden 2019 von der Global Indigenous Data Alliance (GIDA) veröffentlicht. In ihnen manifestiert sich der kritische Blick der GIDA auf die Open Data Entwicklung, die nicht hinreichend auf die Rechte und Interessen der indigenen Völker eingehe, sondern die gemeinsame Datennutzung ohne Berücksichtigung der Machtunterschiede und historischen Kontexte befördere. Die GIDA sieht in den CARE-Prinzipien deshalb eine notwendige Ergänzung der FAIR-Prinzipien, um die Daten(nach)nutzung personen- und zweckgerichtet im Sinne der Förderung der Innovation und Selbstbestimmung der indigenen Völker zu lenken.

Als indigene Daten werden dabei unter anderem Informationen über Sprache, Geschichte, Kultur, Lebensweise, Territorium, Technologien und natürliche Ressourcen der jeweiligen Gruppe verstanden. Das Akronym CARE steht für Collective Benefit, Authority to Control, Responsibility und Ethics. Analog zu den FAIR-Prinzipien sind diese Forderungen durch Subprinzipien konkretisiert.

Collective benefit fordert eine strukturelle Gestaltung von Daten-Ökosystemen, die einen Nutzen aus den Daten für indigene Völker erlaubt. Dieser Nutzen umfasst die aktive Förderung indigener Entwicklung und Innovation, informierte Steuerung und Entscheidungsfindung sowie gerechte Verwendung von Ergebnissen zum Vorteil der indigenen Bevölkerungsgruppe.

Authority to control bedeutet, dass das individuelle und kollektive Recht und Interesse indigener Völker an ihren Daten anerkannt werden. Dies umfasst die Einwilligung zur Datenerhebung und Verwendung

²⁰ Vgl. Research Data Alliance International Indigenous Data Sovereignty Interest Group: CARE Principles for Indigenous Data Governance, <<https://www.gida-global.org/care>>, Stand: 27. April 2020.

ebenso wie die Verfügbarkeit der Daten für eigene Zwecke und die Kontrolle über alle weitere Nutzung.

Responsibility meint die Ausrichtung der Datenverwendung auf die Förderung der Selbstbestimmung und den kollektiven Nutzen der indigenen Völker. Hierzu gehören die Wahrung der Würde in der Erhebung und Nutzung, die Förderung der Datenkompetenz und Entwicklung eines indigenen Datenbestandes und entsprechender Infrastruktur auf Basis indigener Kultur und Werte.

Ethics verweist darauf, dass die Rechte und das Wohlergehen der indigenen Völker den aktuellen und zukünftigen Umgang mit den Daten bestimmen müssen. Das heißt, Erhebungen entsprechen in den Rahmenbedingungen den Festlegungen der ‚United Nations Declaration on the Rights of Indigenous Peoples‘ (UNDRIP), berücksichtigen Folgen von Ungleichgewichten in Bezug auf Macht und Ressourcen und akzeptieren die indigene Einschätzung als maßgeblich für die ethische Bewertung von Schaden und Nutzen. Herkunft, Zweck, Einschränkungen und Verpflichtungen müssen in den Metadaten festgehalten werden.

Die CARE-Prinzipien werfen Aspekte der Verantwortung auf, die zum Teil bereits in verschiedenen Disziplinen Gegenstand wissenschaftsprogrammatischer und -theoretischer Diskussionen waren und sind: Ein wichtiger Aspekt ist dabei die Grundannahme, dass Forschungsergebnisse – in diesem Fall Daten – potenziell für positive wie auch negative Zwecke genutzt werden können. Dieser Aspekt etwa ist aus der Dual-Use-Problematik – mit Fokus auf zivile vs. militärische Nutzung – gut bekannt.²¹

Die mit den CARE-Prinzipien einhergehende Verantwortung ist allerdings nicht auf den sehr allgemein formulierten ‚Fortschritt und Wohlstand

²¹ Vgl. Deutsche Forschungsgemeinschaft (DFG): Wissenschaftsfreiheit und Wissenschaftsverantwortung – Empfehlungen zum Umgang mit sicherheitsrelevanter Forschung, 2014. Online: <<https://doi.org/10.1515/jwiet-2016-0112>>.

der Gesellschaft“²², sondern auf ein relativ klar begrenztes – nämlich das jeweils beforschte – Kollektiv bezogen. Insofern die Nutzung der Daten unter anderem deren Vorteil dienen soll, verlangen die CARE-Prinzipien auch, den Boden – intentional nicht unbedingt tatsächlich – voraussetzungs-freier Forschung zu verlassen.²³ Im Sinne der CARE-Prinzipien bedeutet verantwortungsvoller Umgang mit Daten in diesem Sinne nicht nur, Gefahren und Potentiale für Individuen und Gruppen sorgfältig zu eruieren, sondern auch, den Beforschten Kontrolle über die Daten und deren Verwendung zu geben, die Daten zum Vorteil der Beforschten zu nutzen und eine Nutzung durch die Beforschten selbst zu unterstützen.

Obwohl die CARE-Prinzipien auf ein spezielles Forschungsfeld zugeschnitten beziehungsweise aus diesem entstanden sind, werfen Sie auch Aspekte generellerer Natur auf, etwa wie weit die Verantwortung gegenüber Beforschten gehen darf und muss, welches Mitspracherecht Datenlieferant*innen hinsichtlich Bereitstellung, Zweck/Verwendung zukommen sollte und wie mit widersprüchlichen Interessen hinsichtlich von Datennutzung umzugehen ist. Diese Fragen haben ihre Berechtigung nicht allein im Kontext indigener Völker, sondern potentiell jeder beforschten Gruppe – nicht nur aber insbesondere Minderheiten. Auch auf das vielseitige Feld der Citizen Science, auf amtliche Daten sowie andere Daten des öffentlichen Lebens wie sie etwa in den Bereichen Umwelt, Energie, Verkehr, Kriminalität, Verwaltung, Gesundheit oder Bildung vielfach erhoben werden, lassen sich Aspekte der CARE-Prinzipien übertragen.

²² Ebd., S. 6.

²³ In der Wissenschaftstheorie des 20. Jahrhunderts ist die Kontroverse um Wertfreiheit insbesondere in den Sozialwissenschaften immer wieder aufgeflammt. Dabei begegnet die hier geforderte Positionierung beziehungsweise bewusste Parteinahme durchaus als Teil des Selbstverständnisses etwa in Bereichen der Geschlechterforschung oder in der Aktionsethnologie, vgl. beispielsweise Degele, Nina: *Gender, Queer Studies: eine Einführung*, Paderborn 2008, S. 21 ; Elixhauser, Sophie: *Ethik in der angewandten Ethnologie: eine Feldforschung zum Tourismus auf den Philippinen*, Trier 2006, S. 38f.

... und mehr

Eine besondere Herausforderung an einen verantwortungsvollen Umgang mit Daten bringt die Forschung mit Massendaten mit sich. ‚Big Data‘ ermöglicht es, das Verhalten von Menschen, Maschinen und Organisationen aufzuzeichnen, abzuleiten und zu analysieren – die datengesteuerte Entscheidungsfindung ist Teil des täglichen Lebens und bringt daher besondere Verantwortung mit sich. Diese beschränkt sich bei weitem nicht auf die Vermeidung gezielten Missbrauchs von Daten, sondern umfasst auch ethische Fragen, die sich aus der Methodik der Datenwissenschaft (Data Science) ergeben können. Bezugnehmend auf zunehmende gesellschaftliche Bedenken hinsichtlich der Verwendung von Daten und mit dem Ziel, datenwissenschaftliche Techniken, Infrastrukturen und Ansätze für eine verantwortungsvolle Datenwissenschaft zu entwickeln, wurde 2016 das niederländische Konsortium Responsible Data Science (RDS) gegründet. Diese Verantwortung fassen sie in den vier Herausforderungen fairness, accuracy, confidentiality und transparency (FACT), die im Folgenden kurz erläutert werden sollen.²⁴

Fairness: Bestehende Daten, die zum Erlernen eines Modells verwendet werden (Trainingsdaten) können bestehende soziale Verzerrungen widerspiegeln, die der Algorithmus wiederum übernimmt. Das Profiling kann zu einer weiteren Stigmatisierung bestimmter Gruppen führen.

Genauigkeit: Eine große Datenbasis ergibt nicht automatisch genaue Ergebnisse. Datenwissenschaftliche Ansätze sollten nicht nur Ergebnisse

²⁴ Vgl. Responsible Data Science – Fairness – Accuracy – Confidentiality – Transparency, <<https://redasci.org/>>, Stand: 19. Mai 2020 ; Responsible Data Science. Ensuring Fairness, Accuracy, Confidentiality, Transparency (FACT), <<http://aims.fao.org/ru/activity/blog/responsible-data-science-ensuring-fairness-accuracy-confidentially-transparency-fact>>, Stand: 21. Mai 2020 ; Aalst, Wil M. P. van der; Bichler, Martin; Heinzl, Armin: Responsible Data Science, in: Business & Information Systems Engineering 59 (5), 10.2017, S. 311–313. Online: <<https://doi.org/10.1007/s12599-017-0487-z>>.

präsentieren oder Vorhersagen machen, sondern auch Informationen über die Genauigkeit der Ausgabe liefern.

Vertraulichkeit: Datennutzung muss auf sichere und kontrollierte Weise geschehen, um dem Vertrauen der Datengebenden gerecht zu werden.

Transparenz: Die Designentscheidungen, die auf dem Weg von Daten zu Schlussfolgerungen getroffen werden, müssen transparent und verständlich sein, damit der Erkenntnisprozess keine Blackbox bleibt.

Der ethische Verantwortungsbegriff hinter FACT ist auf eine unspezifische Allgemeinheit ausgerichtet, die in dem Fall eine spezifische Gruppe wird, in dem sich Ergebnisse/Aussagen auf sie beziehen oder auswirken²⁵. Insofern veranschaulicht er, wenngleich es sich eher um methodische Fragen als um Forschungsdatenmanagement im engeren Sinne handelt, die besonderen Risiken, die mit der Analyse von ‚Big Data‘ einhergehen, ebenso wie die Tatsache, dass Daten in der Regel auch Teil einer potentiellen größeren Datengesamtheit sind, die bei einer umsichtigen Folgeabschätzung Berücksichtigung finden muss.

Fazit

Bereits dieser kurze Einblick deutet an, wie vielschichtig der Verantwortungsbegriff im Kontext des Forschungsdatenmanagements ist: Vor dem Hintergrund der akademischen Pflicht zur redlichen Forschung muss verantwortungsvoll zunächst bedeuten, die eigenen Daten so aufzubereiten und zu sichern, dass eben diese Redlichkeit nachprüfbar ist.

Mit dem Anspruch der Nachnutzbarkeit kommen weitere Aspekte hinzu; wenn etwa die FAIR-Maßnahmen helfen sollen, Daten für potentielle Weiterverwendung aufzubereiten, verweist dies auf die Verantwortung für Wertehalt und –steigerung der Daten²⁶ gegenüber potentiellen Datennehmern.

²⁵ Van der Aalst et al. nennen exemplarisch „citizens, patients, customers, and employees“ (van der Aalst; Bichler; Heinzl: Responsible Data Science, 2017).

²⁶ Vgl. Anm. 12.

Inhaltlich kann diese Verantwortung unterschiedlich hergeleitet beziehungsweise erfüllt werden, beispielweise aus Forschungsökonomie, Ressourcenschonung, Innovationspotential, Erkenntnisgewinn, einem generellen Open Science Ansatz und so weiter.

In rechtlichen und ethischen Fragen wird deutlich, dass der Umgang mit Daten verschiedene negative Folgen haben kann, für deren Vermeidung der/die Forschende durch entsprechende Maßnahmen Sorge tragen muss. Die CARE-Prinzipien weiten die ethische Verantwortung der Schadensvermeidung in ihrem speziellen Kontext auf die Verpflichtung aus, die Datennutzung auf explizit positive Effekte auszurichten. Mit der Verantwortung bezüglich möglicher Aussagen, die auf Basis großer Datenmengen getroffen werden oder werden können, setzt sich schließlich FACT auseinander.

Diese dargelegten Facetten der Verantwortung ergänzen und überschneiden sich, können einander aber auch einschränken (wie etwa Vertraulichkeit die Transparenz und die Genauigkeit) oder sich widersprechen – insbesondere, wenn verschiedene Interessen einander entgegenstehen. Ein verantwortungsvoller Umgang ist damit immer auch ein Aushandlungsprozess innerhalb des vorgegebenen Rahmens.

Der Anspruch von Forschungsdatenservices muss es sein, für den verantwortungsvollen Umgang mit Daten soweit wie möglich infrastrukturelle Lösungen anzubieten. Hierfür müssen Kompetenzen aufgebaut und – sofern möglich technisch – implementiert werden. So braucht es, um Daten FAIR zu machen, unter anderem Kenntnisse im Bereich semantischer Interoperabilität (Ontologien, Vokabulare, Metdaten), die wiederum in einem Repository technisch eingebunden werden können. Auch beispielsweise Fragen der Nutzungsrechte, der Zweckbindung und des Datenschutzes können und müssen unter anderem durch Lizenz- und Rechtmanagement technisch unterstützt werden.

Der technischen Realisierung sind in der Regel Entscheidungen und Maßnahmen vorgeschaltet, bei denen spezifische Kompetenzen hinzugezogen werden müssen. Dies lässt sich sinnvoll über einen – freilich technisch

gestützten – möglichst standardisierten Workflow einbinden, was umso wichtiger ist, je verteilter die Kompetenzen in der jeweiligen Einrichtung sind. Für den Workflow bietet sich beispielsweise ein Datenmanagementtool wie RDMO an, das die Uni Kassel bereits seit Anfang 2019 betreibt, aber auch die Einbindung ins FIS-System wäre denkbar.

Dass es bei aller technischer Unterstützung auch – wenn nicht gar primär – um den vielbeschworenen ‚Kulturwandel‘ im Umgang mit Daten geht, der auch der Weitergabe der Kompetenzen in Schulung/Lehre und Beratung bedarf, ist evident. Zugleich werden wissenschaftliche Binnendiskurse aber auch neue Aspekte der Verantwortung offenbaren und die sich daraus ergebenden Herausforderungen an die Infrastruktureinrichtungen herantragen. ‚Verantwortungsvoller Umgang‘ bleibt damit aller Voraussicht nach ein dynamischer Auftrag, den wissenschaftliche Einrichtungen als Ganze wahrnehmen und gestalten müssen.