

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Was eine Data Governance für das Forschungsdatenmanagement leisten kann

Das Projekt Krisenmanagement und Resilienz – Corona (KResCo)

Andrea Wuchnerⁱ Johannes Sautterⁱⁱ Joel Kirchner
Aleyna Kilicasla Ekaterina Dobrokhotova
Daniela Beyerⁱⁱⁱ Siegrid Hecht-Veenhuis
Kathrin Tuppi Wolfgang Beinhauer

2022

Zitiervorschlag

Wuchner, Andrea, Johannes Sautter, Joel Kirchner, Aleyna Kilicasla, Ekaterina Dobrokhotova, Daniela Beyer, Siegrid Hecht-Veenhuis, Kathrin Tuppi und Wolfgang Beinhauer. 2022. Was eine Data Governance für das Forschungsdatenmanagement leisten kann. Das Projekt Krisenmanagement und Resilienz – Corona (KResCo). *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2022: S. 1-17. DOI: [10.17192/bfdm.2022.2.8542](https://doi.org/10.17192/bfdm.2022.2.8542).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱORCID: [0000-0002-5549-3434](https://orcid.org/0000-0002-5549-3434)

ⁱⁱORCID: [0000-0002-5934-7979](https://orcid.org/0000-0002-5934-7979)

ⁱⁱⁱORCID: [0000-0002-3089-4856](https://orcid.org/0000-0002-3089-4856)

1 Abstract

In dem Fraunhofer-internen Projekt „Krisenmanagement und Resilienz – Corona“ wurden Maßnahmen gegen die Covid-19-Pandemie in verschiedenen Ländern sowie deren Wirksamkeit untersucht. Hierbei wurden viele Daten erzeugt und nachgenutzt, deren Management in einem eigenen Arbeitspaket gesteuert wurde. Hierzu wurde ein Data-Governance-Ansatz angewendet, im Rahmen dessen verschiedene strategische, rechtliche und operative Bausteine entwickelt und während des Projektverlaufs erprobt wurden. Der Artikel beschreibt diese Bausteine und beurteilt deren Umsetzungsgang im Projekt anhand eines Reifegrad-Modells. Im Ergebnis zeigte sich, dass nicht alle Ziele der Data Governance im Projekt erreicht werden konnten, da keine ausreichende Akzeptanz bei den Forschenden aufgebaut werden konnte, so dass diese die Bausteine nicht im erforderlichen Umfang nutzten.

2 Data Governance und Forschungsdatenmanagement

Forschungsdaten sind Daten, „die im Zuge wissenschaftlicher Vorhaben, z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen“¹. Wissenschaftler:innen stehen vor der immer komplexer werdenden Herausforderung, adäquat mit diesen Daten umzugehen. Eine zentrale Anforderung hierbei ist das Management und die Bereitstellung der Forschungsdaten gemäß den FAIR-Prinzipien². In den letzten Jahren wurden verschiedene Strukturen geschaffen, um Wissenschaftler:innen beim Forschungsdatenmanagement zu unterstützen: An vielen Universitäten wurden Helpdesks und Beratungsangebote eingerichtet. Auf überregionaler Ebene wurden z.B. die Informationsangebote forschungsdaten.org³ und forschungsdaten.info⁴ etabliert und auch Landesinitiativen wie fdm.nrw schaffen Angebote und Synergien, um die Forschenden beim Forschungsdatenmanagement zu unterstützen.⁵ Die Datenbestände von Forschungseinrichtungen werden zum Teil bereits über Forschungsdaten-Repositoryn zur Nachnutzung bereitgestellt⁶, oft flan-

¹Allianz der Wissenschaftsorganisationen, „Forschungsdaten“, zugegriffen 25. März 2022, <https://www.allianzinitiative.de/archiv/forschungsdaten/>.

²Wilkinson, Mark D., Dumontier, Michel, Aalbersberg, IJsbrand Jan, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan-Willem, Bonino da Silva Santos, Luiz, Bourne, Philip, E. et al. „The FAIR Guiding Principles for Scientific Data Management and Stewardship“. *Scientific Data* 3/160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.

³<https://forschungsdaten.org>, zugegriffen 25. März 2022.

⁴<https://forschungsdaten.info>, zugegriffen 25 März 2022.

⁵Vgl. Neuroth, Heike, und Gudrun Oevel. „Aktuelle Entwicklungen und Herausforderungen im Forschungsdatenmanagement“. In *Praxishandbuch Forschungsdatenmanagement*, Hrsg. Markus Putnings, Heike Neuroth, und Janna Neumann, 537–55. Berlin: De Gruyter Saur, 2021. <https://doi.org/10.1515/9783110657807>.

⁶Vgl. Scholze, Frank, Ulrich, Robert, und Hans-Jürgen Goebelbecker. „Wissenschaftlicher Datenmarkt“. In *Praxishandbuch Forschungsdatenmanagement*, Hrsg. Markus Putnings, Heike Neuroth, und Janna Neumann, 165–73. Berlin: De Gruyter Saur, 2021. <https://doi.org/10.1515/9783110657807>.

kiert durch institutionelle Forschungsdaten-Policies. Teilweise werden Forschungsdaten darüber hinaus in fachspezifische Datendienste eingebunden⁷ und es gibt Ansätze zur konsequenten und strategischen Verwertung der eigenen Forschungsdaten in Forschungsorganisationen. Insgesamt schöpfen Forschungsprojekte und Fachcommunities die Potenziale zur Parallel- und Nachnutzung von Forschungsdaten aber noch unzureichend aus.

Ergänzend zum Datenmanagement, das viele Forschende schon betreiben, existiert das aus der Wirtschaft stammende Konzept der Data Governance. Dabei handelt es sich um ein System aus Regeln und Rollen, welches den Umgang mit Daten in einer Organisation so steuern soll, dass die Daten optimal strategisch genutzt werden können. Das Data-Governance-Konzept umfasst vier Ebenen: die Ebene der Systeme, die Ebene des Datenmanagements, die Ebene der Data Governance (im engeren Sinne) sowie die strategische Ebene. Auf der strategischen Ebene werden Ziele in Bezug auf Daten definiert, auf der Data-Governance-Ebene (im engeren Sinne) werden Regeln zu ihrer Erfüllung aufgestellt, auf der Datenmanagement-Ebene findet das operative Datenmanagement statt und auf der Ebene der Systeme werden die Daten technisch bereitgestellt und verwaltet. Der Data-Governance-Ansatz hat folgende Ziele:

- eine konsequente, strategische Verwertung der Daten
- eine Entlastung der Datenerhebenden und -nutzenden, indem Regeln und Entscheidungsstrukturen nicht im Rahmen einzelner Erhebungen/Projekte durch die Forschenden als „Exekutive“ getroffen werden müssen, sondern generell für alle Erhebungen innerhalb einer Datendomäne⁸ einer Organisation durch eine „Legislative“ festgelegt werden⁹
- die Erhöhung der Qualität und Transparenz der Daten.

Eine Organisation kann daher hinsichtlich der Data Governance in die Bereiche Strategie, Data Governance („Legislative“) und Data Management („Exekutive“) unterteilt werden.¹⁰

Der Ansatz der Data Governance bezieht sich klassischerweise auf eine einzelne Organisation;¹¹ für organisationsübergreifende Ökosysteme ist das Konzept noch unausge-

⁷Vgl. Rat für Informationsinfrastrukturen (RfII). *Nutzung und Verwertung von Daten im wissenschaftlichen Raum. Empfehlungen zur Ausgestaltung von Datendiensten an der Schnittstelle zwischen Wissenschaft und Wirtschaft*. Göttingen, 2021. <https://nbn-resolving.org/urn:nbn:de:101:1-2020052673>.

⁸Eine Datendomäne ist eine Gruppierung von Daten, welche ähnliche Eigenschaften aufweisen und für die daher gemeinsame Regeln getroffen werden können.

⁹Vgl. Earley, Susan. *Data management body of knowledge*. Basking Ridge, New Jersey: Technics Publications, 2017.

¹⁰Vgl. Sautter, Johannes, und Andrea Wuchner. „Data Governance als Voraussetzung für Datenexzellenz an Forschungsorganisationen. Poster präsentiert auf der RDA Deutschland Tagung, Potsdam, 25.–27.02.2020.“ Fraunhofer, 2020. <http://publica.fraunhofer.de/documents/N-586577.html>, zugegriffen 25. März 2022.

¹¹Vgl. Kouper, Inna, Raymond, Anjanette H., and Stacey Giroux. „An Exploratory Study of Research Data Governance in the U.S.“. *Open Information Science*, Nr. 4/1 (2020): 122–142. <http://doi.org/10.1515/opsis-2020-0010>.

reift.¹² Es ist jedoch davon auszugehen, dass das Konzept auf institutsübergreifende Projekte oder auch auf Fachdisziplinen bzw. Datendomänen innerhalb einer Fachdisziplin übertragbar ist. So wurde auch in KResCo der Ansatz der Data Governance in einem institutsübergreifenden Projekt angewendet.

3 Data Governance im Projekt KResCo

Das einjährige Fraunhofer-interne Projekt „Krisenmanagement und Resilienz – Corona“ (KResCo) hatte zum Ziel, Maßnahmen zur Covid-19-Pandemie in verschiedenen Ländern sowie die Resilienz von Systemen gegenüber Krisen wissenschaftlich zu untersuchen. An dem Projekt waren fünf Fraunhofer-Institute beteiligt.¹³ Fünf wissenschaftliche Arbeitspakete widmeten sich mit qualitativen und quantitativen Methoden der Sozialforschung spezifischen Fragestellungen rund um die Auswirkungen und Bewältigung der Pandemie. Drei weitere Arbeitspakete widmeten sich dem Projektmanagement, der Entwicklung gemeinsamer Handlungsempfehlungen sowie dem Forschungsdatenmanagement. In den einzelnen Arbeitspaketen kamen Forschende aus unterschiedlichen Institutionen und Disziplinen (z.B. Soziologie, Politikwissenschaft, Informationswissenschaft, Informatik) zusammen.

Im Rahmen des projektinternen Forschungsdatenmanagements wurde ein an das Projekt angepasster Data-Governance-Ansatz erprobt. Es wurden drei Datendomänen identifiziert: qualitative Sozialforschung, quantitative Sozialforschung und Literaturrecherche. Für alle drei Datendomänen wurden verschiedene Bausteine (=Unterstützungsangebote) entwickelt, um die Forschenden zu entlasten und die Datenqualität zu steigern.

4 Die Bausteine

Um den Data-Governance-Ansatz in das Projekt KResCo zu integrieren, wurde ein modularer Ansatz gewählt, der sich aus verschiedenen Bausteinen zusammensetzt. Diese lassen sich den verschiedenen Phasen eines Projekts zu ordnen. Bei der Entwicklung der Bausteine konnte zum Teil auf bereits vorhandene Unterstützungsangebote (z.B. Einwilligungserklärungen oder Metadatendokumentationen) zurückgegriffen werden, die meisten Bausteine wurden aber für KResCo neu entwickelt. Die Bausteine

¹²Vgl. Lis, Dominik, und Boris Otto, „Towards a Taxonomy of Ecosystem Data Governance“. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, 6067. Honolulu: University of Hawai'i at Mānoa, 2021. <https://doi.org/10.24251/HICSS.2021.733>.

¹³Fraunhofer-Institut für naturwissenschaftlich-technische Trendanalysen INT, Fraunhofer-Informationszentrum Raum und Bau IRB, Fraunhofer-Institut für Arbeitsorganisation IAO, Fraunhofer-Institut für System- und Innovationsforschung ISI und Fraunhofer-Institut für Internationales Management und Wissensökonomie IMW.

funktionieren auf zwei Ebenen: Zum einen bietet jeder Baustein für sich eine wertvolle Funktion für das operative Datenmanagement. Zum anderen wirken die Bausteine zusammen und bilden eine Data-Governance-Funktion, die das Ziel hat, operatives Datenmanagement in den forschenden Arbeitspaketen sowie die Forschenden durch Datenverantwortliche, Domänenverantwortliche und Datenkurator:innen zu unterstützen. Die Bausteine können in anderen Projekten Anwendungen finden bzw. adaptiert werden.

4.1 Vor dem Projekt

Bereits in der Antragsphase wurde das Forschungsdatenmanagement auf strategischer Ebene adressiert und strategische Ziele festgelegt, die im Rahmen der Data Governance für das Projekt umgesetzt werden sollten.

4.1.1 Strategische Verankerung im Projektantrag

Die strategische Ausrichtung der Data Governance wurde in Zusammenarbeit mit dem Projektkonsortium im Arbeitspaket „Forschungsdatenmanagement“ festgelegt. Wichtig war hier, die Datenkultur im Projekt positiv zu beeinflussen sowie die Forschenden für das Forschungsdatenmanagement zu sensibilisieren. Die strategischen, bereits im Projektantrag verankerten Ziele für die Data Governance in KResCo waren zum einen die Erzeugung qualitativ hochwertiger Forschungsdaten und zum anderen die Unterstützung und Entlastung der Forschenden beim Forschungsdatenmanagement. Zudem sollten Forschungsdaten aus dem Projekt zur Nachnutzung bereitgestellt werden. Die Erstellung eines Datenmanagementplans war nicht vorgesehen; stattdessen sollte die Ausgestaltung des Datenmanagements über den Data-Governance-Ansatz erfolgen.

4.1.2 Budget für Forschungsdaten und Domänen

Bereits im Projektantrag wurden Ressourcen für das Forschungsdatenmanagement eingeplant. Hierzu zählen zum einen die Ressourcen für das Forschungsdatenmanagement-Kernteam, das das entsprechende Arbeitspaket steuerte, zum anderen Ressourcen für Datenkurator:innen, Domänenverantwortliche und Datenverantwortliche (vgl. Kapitel 3.2.1).

4.2 Zu Beginn des Projekts

Die folgenden Bausteine wurden zu Beginn des Projekts initiiert und fortlaufend während des Projektes aktualisiert.

4.2.1 Aufbauorganisation

Als erster wichtigen Schritt hin zu einer Data Governance wurden drei verschiedene Rollen für das Datenmanagement im Projekt definiert:

- „Datenverantwortliche:r“: Diese Rolle wurde für jedes fachlichen Arbeitspaket implementiert und von Fachwissenschaftler:innen wahrgenommen. Zu den Aufgaben dieser Rolle gehörte die Verantwortung für die Daten im Arbeitspaket sowie die Anwendung der für das Projekt bereitgestellten Bausteine.
- „Domänenverantwortliche:r“: Diese Rolle wurde für jede Datendomäne im Projekt verankert und von Fachwissenschaftler:innen wahrgenommen. Aufgabe der Domänenverantwortlichen war es, in Zusammenarbeit mit dem AP „Forschungsdatenmanagement“ domänenspezifische Standards zu entwickeln.
- „Datenkurator:in“: Über alle Arbeitspakete hinweg wurden zwei Datenkurator:innen installiert. Diese hatten vor allem die Aufgabe, die Qualität der (Meta-)Daten sicherzustellen und die Datenkuration anhand der Datenkurationscheckliste (vgl. Kapitel 3.4.2) durchzuführen.

Alle drei Rollen wurden bei ihrer Tätigkeit durch das Forschungsdatenmanagement-Kernteam eingeführt, begleitet und unterstützt, beispielsweise in gemeinsamen Workshops oder durch unterstützende Materialien.

4.2.2 Memorandum of Understanding zur gemeinsamen Datennutzung im Projekt

Im Projekt KResCo waren fünf verschiedene Fraunhofer-Institute beteiligt. Um den Befürchtungen der Durchführenden zu begegnen, Datensätze könnten durch andere Projektbeteiligte nicht in ihrem Sinne nachgenutzt werden, wurde ein Memorandum of Understanding entworfen. Darin ist die Nutzung der Forschungsdaten zwischen den Projektpartnern geregelt. Die Kernpunkte sind:

- **Gemeinsamer Zugriff:** „Alle Projektbeteiligten erhalten Zugriff und Einsicht in offene, unsensible und sensible Daten. Hochsensible Daten bleiben am jeweiligen Institut. In der Regel sind das Daten, die vertraulichen Personenbezug enthalten.“ (vgl. auch Kapitel 4.3.3)
- **Zweckbindung:** „Geteilte Datensätze dürfen nicht zweckentfremdet werden, d.h. ein Datensatz einer empirischen Erhebung darf zu keinem anderen Zweck verwendet werden als dem Zweck, der als Forschungsabsicht bei der Ersterhebung der Daten von den Forschenden in den Datenschutzhinweisen und in der Einwilligungserklärung gegeben und vom Datengeber bestätigt wurde.“
- **Nennung der Urheber:innen:** „Das Teilen der Datensätze befähigt andere Projektinstitute allein zur Nutzung und zweckmäßigen Verarbeitung unter den gemeinsamen Forschungszielen des Projekts. Die UrheberInnen behalten damit neben anderen UrheberInnen das Recht, bei einer Veröffentlichung als AutorInnen genannt zu werden. Die betrifft bei der Veröffentlichung eines Forschungsdatensatzes alle an der Datenerhebung beteiligten Personen. Bei der Nachnutzung von

Forschungsdaten im Rahmen einer Publikation kann dort ein Data Statement gemacht werden, in dem nachgenutzte Daten und deren UrheberInnen genannt werden.“

- Nutzung außerhalb des Projekts: „Die Projektpartner verpflichten sich hiermit, die geteilten Daten nur unter Einbezug des Datenurhebers außerhalb des Projekts weiter zu nutzen. Eine geplante Nutzung oder Veröffentlichung ist im Konsortium anzuzeigen. Im Falle von personenbezogenen Daten ist die Einwilligungserklärung des Datengebenden zu beachten.“

4.2.3 Überblick der Datensätze und Metadatenheet

Zu Beginn des Projekts wurde zunächst ein Überblick geschaffen, welche Datensätze in welchem Arbeitspaket erhoben bzw. nachgenutzt wurden. Hierzu wurde eine Tabelle angelegt, die von den Mitarbeitenden der Arbeitspakete befüllt wurde. Dabei wurden die Datensätze auch mit Identifiern versehen, welche die Datensätze innerhalb des Projekts eindeutig kennzeichneten. Zum Zwecke der Datendokumentation wurde ein umfangreiches Metadatenheet in tabellarischer Form entwickelt, das von jedem Arbeitspaket befüllt wurde. Wichtig waren dabei Projektmetadaten, Metadaten, die den Forschungsdatensatz betreffen, Metadaten zu den verwendeten Methoden sowie zur Veröffentlichung und zur Archivierung. Um für die unterschiedlichen Phasen des Forschungsprozesses verschiedene Metadaten eintragen bzw. diese aktualisieren zu können, wurden Datenablagepunkte definiert – Zeitpunkte im Projektverlauf wie zum Beispiel Projektantrag, Datenerhebung, Datenanalyse –, an denen das Sheet aktualisiert werden sollte. Die Metadaten wurden durch eindeutige Identifier mit den einzelnen Datensätzen verknüpft.

4.2.4 Interne Kommunikation

Um die Projektbeteiligten zu anstehenden Themen im Projekt zu informieren, wurden regelmäßig über die gesamte Projektlaufzeit insgesamt neun Ausgaben der „Informationen zum Forschungsdatenmanagement“ herausgegeben. Diese bestanden aus einem Foliensatz, der per Mail versandt wurde und in dem jeweils bestimmte Themen des projektinternen Forschungsdatenmanagements aufgegriffen wurden (z.B. die gemeinschaftliche Ablage, die Einwilligungserklärung oder die Kurationscheckliste).

4.2.5 Gemeinsame Datenablage und Verantwortlichkeiten

Da im Projekt fünf verschiedene Fraunhofer-Institute beteiligt waren, musste zunächst eine gemeinsame Ablage für die Forschungsdaten definiert werden, die datenschutzrechtlich unbedenklich war und auf die alle Projektpartner:innen Zugriff hatten. Die Entscheidung fiel auf die Fraunhofer OwnCloud, ein innerhalb der Fraunhofer-Gesellschaft etabliertes Tool. In der OwnCloud wurde pro Arbeitspaket ein Ordner für die

Datensätze angelegt, der wiederum für jeden Datensatz einen Unterordner enthielt. Auf der OwnCloud wurden nur Daten gespeichert, die von den Arbeitspaketen nicht als hochsensibel eingestuft wurden. Die OwnCloud ist eine zentrale Anwendung und kann als App auf dem Rechner installiert werden, so dass die Ordner in die lokale Verzeichnisstruktur eingebunden werden. Hierbei ist eine Synchronisation von lokaler Verzeichnisstruktur und zentralem Server notwendig.

4.3 Vor der Datenerhebung

Zur Vorbereitung der Datenerhebung spielte vor allem das Thema „Rechtsgrundlage“ eine große Rolle.

4.3.1 Vorlagen für Einwilligungserklärungen und Datenschutzinformationen

Zur Unterstützung der Forschenden wurden DSGVO-konforme Einwilligungserklärungen und Datenschutzinformationen zur Verfügung gestellt. Die Einwilligungserklärungen sahen eine Pseudonymisierung der Daten vor, so dass die Daten entsprechend der Empfehlung des RatSWD auf Nachfrage zu Forschungszwecken herausgegeben werden können.¹⁴

4.3.2 Vereinbarung zur Datenübergabe

Da im Projekt auch bereits veröffentlichte Daten nachgenutzt wurden, wurde ein standardisiertes Formular erstellt, in dem die rechtlichen Bedingungen und praktische Aspekte für die Nachnutzung zwischen Datengeber und Datennehmer festgelegt werden konnten. Hierzu gehörten zum Beispiel Angaben zum Datenschutzniveau und zur Speicherung der Daten im Projekt. Das Formular war so ausgestaltet, dass verschiedene Szenarien damit abgedeckt werden konnten und das Formular einheitlich verwendet werden konnte.

4.3.3 Sensibilitätsstufen

Im Projekt wurden vier Sensibilitätsstufen für Daten definiert, die z.B. für die Zugriffsregeln innerhalb der eigenen Organisation herangezogen wurden. Da die betroffenen Personen in einer Einverständniserklärung, die sich am Prinzip „Privacy by Design“

¹⁴Vgl. Rat für Sozial- und Wirtschaftsdaten (RatSWD). „Handreichung Datenschutz“. RatSWD Output, 22. Juni 2020. <https://doi.org/10.17620/02671.50>.

bzw. „Privacy by Default“¹⁵ orientiert, zunächst der geringsten Sensibilitätsstufe zuzustimmen, sind die folgenden Sensibilitätsstufen entsprechend von einer strengen bis hin zu einer offenen Berechtigung angeordnet.

- Sensibilitätsstufe 1: hochsensibel. Hierbei handelt es sich um Daten, die personenbezogene Daten enthalten. Zugriff auf hochsensible Daten hat nur die:der erstellende Wissenschaftler:in und die:der dazugehörige Datenkurator:in oder ein in der Datenschutzzinformation beschriebener Personenkreis.
- Sensibilitätsstufe 2: sensibel. Hierbei handelt es sich um pseudonymisierte Daten. Zugriff auf die sensiblen Daten hat nur die:der erstellende Wissenschaftler:in und die:der dazugehörige Datenkurator:in oder ein in der Datenschutzzinformation beschriebener Personenkreis.
- Sensibilitätsstufe 3: unsensibel. Hierbei handelt es sich um anonymisierte Daten. Es wird mittels eines Walk-Through geprüft, ob der Personenbezug bzw. die Deanonymisierung noch möglich ist. Durch Kategorisierungen werden diese Aspekte eliminiert. Zugriff auf diese Daten hat das gesamte Projektteam.
- Sensibilitätsstufe 4: offen/öffentlich. Diese Daten enthalten keine personenbezogenen oder -beziehbaren Informationen (mehr). Der Zugriff auf diese Daten durch die Öffentlichkeit ist erlaubt, sodass die Daten in ein Repositorium hochgeladen werden können; in der Regel wird eine offene Lizenz vergeben.

4.3.4 Datei- und Ordnerstruktur

Das zentrale Forschungsdatenmanagement-Team stellte jedem Arbeitspaket einen eigenen Ordner für seine Forschungsdaten zur Verfügung. Abbildung 1 visualisiert eine Best-Practice-Untergliederung für Fokusgruppen (a) und Leitfadengespräche (b).

4.3.5 Datenkurationsprofil

Für die Domänen „qualitative Sozialforschung“ und „quantitative Sozialforschung“ wurde jeweils ein Datenkurationsprofil erstellt, in welchem die einzelnen Bearbeitungsschritte und Zwischenprodukte der Daten erfasst wurden. Vorlage hierfür waren die Arbeiten der Purdue University (Indiana).¹⁶ In KResCo wurde das Modell dahingehend adaptiert, dass es nicht pro Datensatz, sondern pro Domäne erstellt wurde. Es diente vornehmlich zur Dokumentation der Datei- und Ordnerstruktur und wurde vor und während der Datenerhebung von den Datenkurator:innen und den Forschenden ausgefüllt. Zum einen konnte hierdurch innerhalb des Projekts ein Standard geschaffen werden, der in weiteren Projekten einem Praxis-Check unterzogen werden kann; auch wurde

¹⁵In Art. 25 DSGVO ist die Rede von „Datenschutz durch Technikgestaltung“ (Privacy by Design) und „Datenschutz durch datenschutzfreundliche Voreinstellungen“ (Privacy by Default). Beide Maßnahmen tragen zum Schutz personenbezogener Daten bei.

¹⁶Vgl. Carlson, Jake. „The Data Curation Profiles Toolkit“. Purdue University, 2010. <https://docs.lib.purdue.edu/dcptoolkit/>, zugegriffen 25. März 2022.

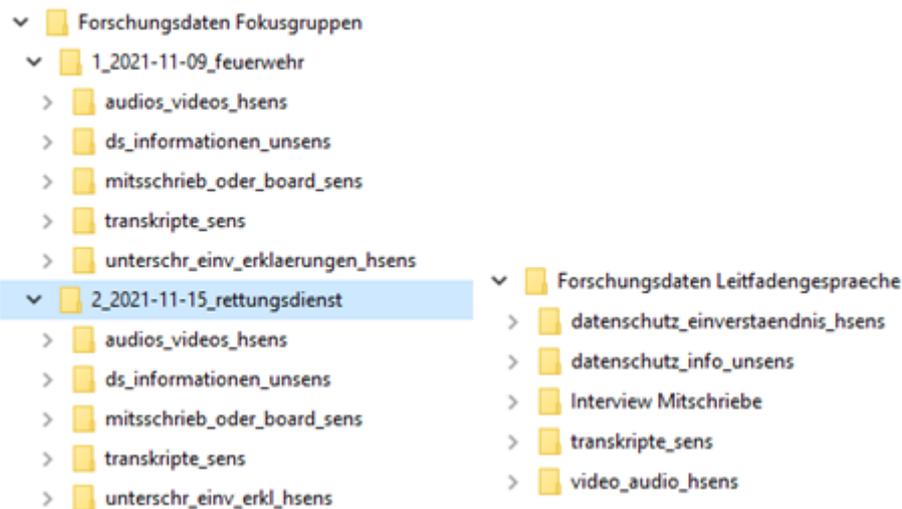


Abbildung 1: Ordnerstruktur der Domäne „qualitative Sozialforschung“ für Fokusgruppen (a) und Leitfadengespräche (b).

Transparenz über den Prozess der Datenbearbeitung hergestellt. Zum anderen diente das Datenkurationsprofil der Datenkuration innerhalb einer Domäne: Es konnte überprüft werden, ob die Daten in den definierten Bearbeitungsschritten bearbeitet wurden.

4.4 Während und nach der Datenerhebung

Die Arbeit während und nach der Datenerhebung fand vor allem innerhalb der Arbeitsspakete auf der operativen Data-Management-Ebene statt, wurde aber von der Data-Governance-Ebene unterstützt und angestoßen.

4.4.1 Identifikation von Samples

Im Projekt wurden sowohl Interviews selbst durchgeführt als auch vorhandene Interviewdaten nachgenutzt. Weil beide Arten von Datensätzen die Grundlage für die nachfolgende Auswertung bildeten, musste eine typenübergreifend einheitliche Nummerierung der Interviews erfolgen. In der ursprünglichen Nummerierung der Interviews waren allerdings zunächst nur die eigenständig durchgeführten Interviews berücksichtigt worden. Im Zuge der Integration der externen Interviews mussten daher alle Identifier neu vergeben werden. Weil Identifier auch für diejenigen Interviews vergeben worden waren, die zwar geplant, aber aus verschiedenen Gründen nicht durchgeführt worden waren, bedurfte es einer weiteren Überarbeitung der Interview-Identifier. Nach weiteren „Refactoring-Prozessen“ wurden letztlich die finalen Interview-Identifier in den

entsprechenden Verzeichnissen angepasst, um zumindest am Ende des Projektes eine einheitliche, transparente und somit übersichtliche Datenstruktur zu gewährleisten. Ein Kernelement des Interviewmanagements war zudem die Sampling-Tabelle, die aufschlüsselt, welche natürliche Person mit welchem Sample verbunden ist, um etwa im Falle eines Widerrufs der Einwilligung tätig zu werden.

4.4.2 Datenkurationscheckliste und FAIR-Checkliste

Um die strategischen Ziele, insbesondere die hohe Datenqualität, sicherzustellen, wurden sowohl eine „Checkliste zur Datenkuration“ als auch eine „FAIR-Checkliste“ erarbeitet, die jeweils pro Arbeitspaket auszufüllen waren. Die Datenkurationscheckliste war in zwei Hauptabschnitte „Metadaten“ und „Datensätze“ gegliedert, die wiederum Checklisten und Informationen sowohl für Forschende als auch für Datenkurator:innen beinhalteten und sich gegenseitig ergänzten. Nachdem die Metadaten und die Datensätze vollständig kuratiert worden waren, erfolgte der Übergang zur FAIR-Checkliste, die sich ausschließlich der Anpassung der Daten und Datensätze an die FAIR-Kriterien widmete. Um die FAIR-Prinzipien in messbare Kennzahlen zu operationalisieren, gibt es verschiedene Modelle. Die in KResCo angewandte FAIR-Checkliste basiert auf dem FAIR Data Maturity Model¹⁷. Die Kriterien sind bei diesem Modell jeweils einem der vier Prinzipien „findable“, „accessible“, „interoperable“ und „reusable“ zugeordnet und jeweils mit einer Prioritätsstufe versehen: essentiell, wichtig oder nützlich. Die in der Checkliste aufgeführten Aspekte für FAIRe Daten konnten sehr einfach mit „erfüllt“ oder „nicht erfüllt“ beantwortet werden. Sie wurden von den Forschenden nach und nach möglichst vollständig ausgefüllt und zur abschließenden Überprüfung an die Datenkurator:innen übergeben.

4.4.3 Kuratierte Methodenliste mit Bezug zu Datensatz und Datendomäne

Für die Domänen „qualitative Sozialforschung“ und „quantitative Sozialforschung“ wurde von den Domänenverantwortlichen jeweils eine kuratierte Methodenliste erstellt. Diese enthielt eine konsolidierte Liste aller Erhebungsmethoden inkl. Literaturverweise, die in der jeweiligen Domäne angewendet werden. Sie enthielt auch Angaben darüber, welche Datensätze mit welcher Methode erstellt wurden. Auf diese Weise wurde Transparenz und Eindeutigkeit hinsichtlich der angewandten Methoden hergestellt.

4.5 In der Projektabschlussphase

Mit dem Abschluss des Projekts mussten Vorkehrungen für den Umgang mit den Forschungsdaten in der Zeit nach dem Projekt getroffen werden.

¹⁷Vgl. RDA FAIR Data Maturity Model Working Group. „FAIR Data Maturity Model. Specification and Guidelines“. Research Data Alliance, 2020. <https://doi.org/10.17620/02671.50>.

4.5.1 Transformationsprozesse

In der Domäne „qualitative Sozialforschung“ erfolgte eine Transformation der Sensibilitätsstufen. So wurde mittels regelbasierter Transkription eine erste Transformation der Sensibilitätsstufe von der rohen Videodatei (hochsensibel) hin zum Transkript (sensibel) erreicht. Anschließend erfolgte sukzessive das Entfernen sämtlicher personenbezogener Daten aus den Dateien, so dass die Daten in die Sensibilitätsstufe „unsensibel“ transformiert wurden. Hierbei nahmen die Forschenden selbst die Transformationsprozesse vor.

4.5.2 Löschkonzept

Für die im Projekt erhobenen Daten bestanden aufgrund verschiedener gesetzlicher Regelungen Lösch- und Archivierungspflichten. Um diese zu erfüllen, wurde basierend auf der DIN-Norm 66398 ein Lösch- und Archivierungskonzept erstellt.¹⁸ Das Löschkonzept enthielt die relevanten Rechtsgrundlagen, Datenarten sowie Lösch- und Archivierungsfristen. Es wurde zudem dargelegt, welche Maßnahmen zu ergreifen sind, um die Daten vollständig von ihrem Speicherort zu entfernen.

4.5.3 Pfadmanagement für dezentrale Archivierung

Laut Art. 17 DSGVO muss sichergestellt sein, dass sich das Recht der Proband:innen auf Widerruf der Einwilligung in die Datenverarbeitung jederzeit umsetzen lässt. Um dies zu garantieren, gab es eine Übersichtstabelle, die auch über das Projektende hinaus verzeichnet, an welchem Speicherort die Forschungsdaten abgelegt sind. Sollte nach Projektende ein Widerruf der Einwilligung erfolgen, können die Daten gemäß DSGVO zuverlässig gelöscht werden.

4.5.4 Datenlogistik

Auf Grundlage der Datenarchitektur, des Pfadmanagements sowie des Löschkonzepts wurde ein Konzept für die Datenlogistik erstellt. Hierbei wurde definiert, welche Datenarten an welchem Zielort final für mindestens 10 Jahre abzulegen sind.¹⁹ Zielorte des Datenlogistikprozesses, der zentral durch das Forschungsdatenmanagement-Team für jedes Arbeitspaket ausgeführt wurde, waren die lokalen Systeme der teilnehmenden Institute, die Fraunhofer OwnCloud sowie öffentliche und zugriffsbeschränkte Ordner auf dem FHG-Content-Server.

¹⁸Die Webseite zur DIN 66398: „Leitlinie Löschkonzept“, zugegriffen 25. März 2022, <https://www.din-66398.de/>.

¹⁹Vgl. Deutsche Forschungsgemeinschaft (DFG). „Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex“. Zenodo, 20. April 2022. <https://doi.org/10.5281/zenodo.6472827>.

4.5.5 Datenveröffentlichung

Die Bereitstellung der in KResCo erhobenen Forschungsdaten wurde bereits im Projektantrag als strategisches Ziel verankert. Die Form der Ergebnispräsentation zielte gerade wegen der Relevanz des Themas „Covid-19“ dabei nicht nur auf die wissenschaftliche Community, sondern auch auf die Presse und interessierte Öffentlichkeit ab und musste daher eine gewisse pragmatische Qualität aufweisen. Derzeit wird an der Bereitstellung maschinenlesbarer Daten über eine API gearbeitet, da auf diese Weise Dritten Zugang zu qualitativ hochwertigen Datenquellen ermöglicht und somit der Erkenntnisgewinn beschleunigt werden kann. Darüber hinaus ist eine interne Archivierung der Forschungsdaten auf Fraunhofer-Servern sowohl für Archiv- als auch für Datenschutzzwecke sowie eine Datenpublikation auf dem Repositorium Fordatis²⁰ der Fraunhofer-Gesellschaft geplant.

5 Evaluation

Im Anschluss an das Projekt fand eine kritische Evaluation der einzelnen Bausteine statt. Hierfür wurden die Bausteine drei Reifegraden zugeordnet:

- Vorlage entwickelt: Dieser Reifegrad wurde erreicht, wenn der Baustein im Projekt entwickelt wurde.
- In Forschungsvorgehen implementiert: Dieser Reifegrad wurde erreicht, wenn der Baustein von Forschenden und Datenkurator:innen verwendet wurde.
- Erfolgreich mit Output: Dieser Reifegrad wurde erreicht, wenn der Baustein im Projekt erfolgreich genutzt wurde und ihr Nutzen erfüllt wurde.

Die Reifegrade 1 und 2 wurden für alle Bausteine erreicht, d.h. es wurden Vorlagen und Konzepte entwickelt und den Forschenden zur Nutzung angeboten. Nicht alle Bausteine wurden jedoch von den Forschenden gut angenommen, weswegen nur einige Reifegrad 3 erreichten.

Insbesondere gestaltete es sich schwierig, wenn Forschende die Bausteine eigenverantwortlich nutzen sollten. Wenn dagegen eine Begleitung und Unterstützung durch das Arbeitspaket „Forschungsdatenmanagement“ erfolgte, etwa beim Ausfüllen des Metadatenheets oder beim Unterzeichnen des Memorandum of Understanding, wirkte sich dies positiv auf die Nutzung der Bausteine aus. In diesem Zusammenhang wurde auch deutlich, dass ein verstärkter Einbezug der Datenkurator:innen sinnvoll ist und zur Entlastung der Forschenden beiträgt. Dies ist z.B. bei der Pflege der Ordner- und Datenstruktur, der Kuration der Daten anhand des Datenkurationsprofils oder auch bei den Transformationsprozessen der Fall. Insbesondere mit den Transformationsprozessen sollten die Forschenden idealerweise nur steuernd zu tun haben. Die Transformation beispielsweise eines hochsensiblen Klarnamenssamplings hin zu einem Samp-

²⁰<https://fordatis.fraunhofer.de>, zugegriffen 25. März 2022.

Phase	Baustein	1. Vorlage entwickelt	2. In Forschungs-vorgehen implementiert	3. Erfolgreich mit Output
Vor dem Projekt	A1 Strategische Verankerung im Projektantrag	X	X	X
	A2 Budget für Forschungsdaten und Domänen	X	X	(X)
Zu Beginn des Projekts	B1 Aufbauorganisation	X	X	
	B2 Memorandum of Understanding	X	X	(X)
	B3 Überblick der Datensätze und Metadaten-sheet	X	X	X
	B4 Interne Kommunikation	X	X	
	B5 Gemeinsame Datenablage	X	X	(X)
Vor der Datenerhebung	C1 Vorlagen für Einwilligungs-Erklärungen und Datenschutz-informationen	X	X	(X)
	C2 Vereinbarung zur Datenübergabe	X	X	
	C3 Sensibilitätsstufen	X	X	(X)
	C4 Datei- und Ordnungsstruktur	X	X	
Während und nach der Datenerhebung	D1 Identifikation von Samples	X	X	
	D2 Datenkurations- und FAIR-Checkliste	X	X	
	D3 Methodenliste	X	X	X
In der Projektabschlussphase	E1 Transformations-prozesse	X	X	
	E5 Löschkonzept	X	X	
	E3 Pfadmanagement für dezentrale Archivierung	X	X	
	E4 Daten-veröffentlichung	X	X	(X)

Tabelle 1: Erreichte Reifegrade der Bausteine.

ling, das nur noch einordnende Kriterien der Personen, jedoch keine Klarnamen mehr enthält, können Datenkurator:innen durchführen. Ebenso kann die Entscheidung, welche der Dateien in die nachfolgenden Datenlogistikprozesse (Archivierung, Veröffentlichung, Löschung) übernommen werden, durch Datenkurator:innen gesteuert werden. Dies hat den Vorteil, dass in dieser Phase Kenntnisse zu Dateiformaten der Langzeitarchivierung der Datenkurator:innen einfließen können. Diese sind bei Forschenden oft nicht vorhanden.

Zum anderen zeigte sich, dass einige Bausteine im Projektverlauf angepasst werden mussten. So wurde z.B. in die Einwilligungserklärung die Pseudonymisierung mit aufgenommen, nachdem sich herausstellte, dass eine Anonymisierung aller Daten technisch nicht zuverlässig zu leisten war. Auch das Metadatenheet musste angepasst werden: Hier zeigte sich, dass die geplante Befüllung zu verschiedenen Zeitpunkten im Verlauf des Forschungsprozesses zu komplex war, so dass die Sheets nur bis zu einem bestimmten Punkt ausgefüllt wurden. Folglich wurden die Datenablagepunkte entfernt und die Befüllung fand gemeinsam durch Wissenschaftler:innen und Mitarbeitende aus dem Arbeitspaket „Forschungsdatenmanagement“ statt. Als entscheidender Aspekt während der Datenerhebung erwies sich zudem die inkrementelle Anpassung der Datenstruktur des Datensatzes, da diese sich im Rahmen eines explorativen Forschungsvorgehens immer wieder verändern kann. Dies sollte proaktiv – und nicht mit einer zu großen zeitlichen Verzögerung erfolgen und als iterativer Prozess verstanden werden. Wird es versäumt, gleich zu Beginn eine Grundstruktur und eine entsprechende Logik des Datenmanagements sowie auch Verantwortlichkeiten im Team sowie bei den Datenkurator:innen zu etablieren, besteht die Gefahr, dass Daten nach individueller und von einer sinnvollen Grundstruktur abweichenden Logik strukturiert werden. Die Folge ist eine unübersichtliche und nicht selbstbeschreibende Datenstruktur, die im schlimmsten Fall zu einem Verlust von Daten führen kann. „Micromanagement“, sehr detail-orientiertes Management der Daten, ist hier hilfreich und sinnvoll, insbesondere bei ersten Datenerhebungen, wie beispielsweise Interviews einer Serie.

6 Fazit und Ausblick

Im Rahmen des Projekts konnten nicht alle angestrebten Ziele vollumfänglich erreicht werden. Insbesondere die Entlastung der Forschenden ist nicht im angestrebten Maße erfolgt. So fand die Nutzung der im Projekt erarbeiteten Baustein durch die Wissenschaftler:innen mit sehr unterschiedlicher Intensität statt und nicht alle Konzepte wurden gleichermaßen gut angenommen. Die Gründe hierfür waren vielfältig: Viele Konzepte des Forschungsdatenmanagements waren den Wissenschaftler:innen nicht vertraut und es war schwierig, Akzeptanz herzustellen, da zunächst vor allem der Arbeitsaufwand für die Nutzung der Konzepte gesehen wurde. Da die Laufzeit des Projekts nur ein Jahr betrug, hatten die Forschenden auch wenig Spielraum, sich mit den

neuen Konzepten zu befassen. Auch waren von vornherein nicht alle Wissenschaftler:innen im Projekt vom Ansatz der Data Governance überzeugt, so dass sich nicht alle Arbeitspakete gleichermaßen beteiligten. Die drei definierten Rollen für das Datenmanagement im Projekt benötigten intensive Betreuung durch das Arbeitspaket „Forschungsdatenmanagement“. Die Datenkurator:innen waren nach Einweisungen in die Bausteine arbeitsfähig; schwieriger gestaltete sich die Arbeit mit den Domänenverantwortlichen und Datenverantwortlichen. Hier war oft die Terminvereinbarung für die einzelnen Gremien bereits eine Herausforderung.

Als Verwertungsziel wurde die Veröffentlichung von im Projekt genutzten und erhobenen Datensätze vorgesehen. Dies gestaltete sich als schwierig. Zum einen wurden im Projekt viele Datensätze aus externen Quellen nachgenutzt und kombiniert, was zu Daten mit unübersichtlicher Rechtsstruktur führte. Eine durch das Arbeitspaket „Forschungsdatenmanagement“ angestrebte Dokumentation der Rechtsstruktur durch die Forschenden erfolgte nicht. Zum anderen waren die Wissenschaftler:innen hinsichtlich der Veröffentlichung ihrer eigenen Daten sehr skeptisch und zurückhaltend und hatten auch Bedenken hinsichtlich der Qualität ihrer Daten. Entsprechend den Empfehlungen des RatSWD²¹ können sogar Daten aus der Domäne „Qualitative Sozialforschung“ grundsätzlich zugangsbeschränkt veröffentlicht werden. Hierbei werden lediglich beschreibende Metadaten veröffentlicht und die eigentlichen Daten sind erst auf Nachfrage und individuelle Genehmigung zugänglich. Dies ist bis jetzt allerdings noch nicht erfolgt. Was die Qualität und Transparenz der Daten angeht, konnte zumindest ein Teilerfolg erzielt werden. Durch verschiedene Bausteine wie die Kurations- und die FAIR-Checkliste sowie die das Datenkurationsprofil und die Methodenliste konnte die Qualität und auch die Transparenz gesteigert werden.

Der Data-Governance-Ansatz und die in KResCo entwickelten Vorlagen können grundsätzlich in weiteren Forschungsprojekte verwendet werden. Sie stehen Fraunhofer intern zur Nachnutzung zur Verfügung, sollten aber vor der Weiterverwendung einem kritischen Usability-Check unterzogen und optimiert werden. Auch empfiehlt es sich, zusätzlich zum Data-Governance-Ansatz einen Datenmanagementplan zu erstellen, in dem wichtige Eckpfeiler des Datenmanagements verbindlich für das ganze Konsortium festgelegt werden. Die Rollen im Projekt sollen noch besser durch Schulung und Briefing unterstützt werden. Das volle Potenzial von Data Governance kann letztlich vor allem auch durch technisch gestützte Workflows und Digitalisierung inklusive einer intelligenten Datenarchitektur realisiert werden.

Das Konzept der Data Governance wurde im Rahmen von KResCo zunächst auf ein einzelnes Projekt ausgerollt. Perspektivisch kann der Ansatz auf weitere Projekte ausgeweitet werden, bis er auf ein Forschungsinstitut, ein Forschungsprogramm, eine Fördereinrichtung und schließlich auf eine Forschungscommunity übertragen werden kann. Hier gilt es, die adäquate Ebene zur Ansiedlung unterschiedlicher Aspekte der Data Governance zu identifizieren, die die Wissenschaftsfreiheit respektiert und gleich-

²¹Vgl. RatSWD. „Handreichung Datenschutz“.

zeitig eine effiziente und möglichst umfassende Nutzung der Daten in Projekten realisiert. Es sollte ferner die strategische Ebene der Aufbauorganisation mit Zuständigkeiten und Mandaten für Datenmanagement gestärkt werden, so dass Entscheidungs- und Steuerungsstrukturen für den Umgang mit Forschungsdaten etabliert werden. Führt man sich die relativ geringe Projektdauer und -größe von KResCo vor Augen, kommt man zu dem Schluss, dass die Projektebene nicht zwingend die richtige Ebene für die Ansiedelung einer Data Governance als steuernde „Legislative“ ist, da sehr viele Parameter definiert und festgelegt werden müssen, die nur für einen begrenzten Rahmen Gültigkeit besitzen. In länger laufenden Großforschungsprojekten wie beispielsweise beim Large Hadron Collider des CERN scheint die Ebene des Projekts dagegen durchaus die adäquate Ebene zu sein.

7 Danksagung

Das Projekt KResCo wurde im Rahmen der internen Programme der Fraunhofer-Gesellschaft, unter Fördernummer Anti-Corona Nr. 840271 gefördert. Ein besonderer Dank geht an Jeannette Anniés (Institut für Arbeitswissenschaft und Technologiemanagement der Universität Stuttgart) für die Arbeiten an den zugrundeliegenden Arbeitshilfen.