

Claudia Engelhardt und Harald Kusch

5.3 Kollaboratives Arbeiten mit Daten

Abstract: Viele Forschungsdaten werden durch kollaborative Initiativen erhoben, integriert, analysiert und der wissenschaftlichen Community zur Verfügung gestellt. Dieser Beitrag beschäftigt sich mit unterschiedlichen Formaten, Herangehensweisen und Spannungsfeldern beim kollaborativen Arbeiten mit Daten und deren Implikationen für das Forschungsdatenmanagement. Es werden zunächst Chancen und Risiken beleuchtet und anschließend verschiedene Ausprägungen, Werkzeugkategorien, organisatorische und regulative Maßnahmen sowie weitere Querschnittsaspekte betrachtet. Das Kapitel Praxistransfer gibt schließlich einen Überblick über unterschiedliche Komplexitätsebenen bei der praktisch-technischen Umsetzung von Dateninfrastrukturen für die kollaborative Forschung und stellt zwei Anwendungsbeispiele aus den Geowissenschaften und der Kardiologischen Grundlagenforschung vor.

Einleitung

Wissenschaft ist seit jeher ein gemeinschaftliches Unterfangen, bei dem Forschende auf den Ergebnissen und Erkenntnissen ihrer Vorgängerinnen und Vorgänger sowie Peers aufbauen. Zunehmend wird Wissenschaft in großen Forschungsteams und -verbänden betrieben, was zum einen mit der Komplexität der Forschungsgegenstände und -fragen zusammenhängt, zum anderen auch oft organisatorische oder finanzielle Gründe hat.¹ Als Vorreiter dieser Entwicklung können die Natur- sowie die Sozialwissenschaften betrachtet werden².

Kollaboratives Arbeiten lässt sich, in einem engeren Sinne, als eine besondere Form der Zusammenarbeit definieren, bei der die Beteiligten gemeinsam und gleichzeitig eine Aufgabe, ein Projekt oder eben Daten bearbeiten. Dies unterscheidet kollaboratives Arbeiten von der „bloßen“ Teamarbeit, bei der zwar ein gemeinsames Ziel verfolgt wird, die zur Erreichung notwendigen Aufgaben aber nicht notwendigerweise zusammen, sondern von Einzelnen und parallel bearbeitet werden.³ Im Alltag werden kollaborative Forschung und das kollaborative Arbeiten an Daten nach dieser engen Definition oft nur einen Teil der Arbeit in einem Forschungspro-

1 Vgl. Schefer 2012, 85.

2 Vgl. Thagard 2002, 242–245.

3 Vgl. Warkentin 2019.

jekt ausmachen. So zeichnet sich kollaborative Forschung nach dem Verständnis des Schweizerischen Nationalfonds (SNF) dadurch aus:

dass die angestrebten Forschungsziele nur erreicht werden können, indem mehrere Gesuchstellende komplementäre Kompetenzen und Kenntnisse in einem neuen, gemeinsamen Forschungsansatz zusammen bringen [sic]. Das gemeinsame Forschungsziel sollte nur durch intensive Zusammenarbeit erreicht werden können.⁴

In ähnlicher Weise wird kollaborative Forschung etwa vom Forschungszentrum Informatik (FZI Karlsruhe) beschrieben:

In kollaborativen Forschungsprojekten arbeitet ein Verbund von Partnerinnen und Partnern an einer definierten Forschungsaufgabe. [...] In der kollaborativen Forschung verläuft der Wissenstransfer nicht ausschließlich von Forschungspartnerinnen und -partner [sic] zu den Auftraggebenden, sondern alle Projektpartnerinnen und -partner unterstützen sich gegenseitig mit ihren Kompetenzen, um ein gemeinsames Forschungsziel zu erreichen.⁵

In großen, kollaborativen Verbundprojekten (Konsortialforschung), wie beispielsweise den Sonderforschungsbereichen (SFB) der Deutschen Forschungsgemeinschaft (DFG), wird zur Erreichung des gemeinsamen Projektziels eher eine Mischform von kollaborativer Arbeit im Sinne der eingangs angeführten engen Definition sowie von Teamarbeit die Regel sein, in der einzelne Teilprojekte zunächst einmal ihr Thema und ihre Daten bearbeiten, die dann zusammengeführt werden. Dies gilt, in kleinerem Maßstab, auch auf der Ebene der Teilprojekte oder für kleinere Projekte. In all diesen diversen Konstellationen ergeben sich spezifische Anforderungen für das Datenmanagement.

1 Chancen, Risiken und Hindernisse beim kollaborativen Arbeiten mit Daten

Kollaboratives Arbeiten kann den Austausch und die Kommunikation innerhalb des Teams fördern. Durch den Ausgleich von Schwächen und die bessere Nutzung von Stärken der Beteiligten sowie die Vermeidung doppelter Arbeiten ermöglicht es eine effektivere Zusammenarbeit.⁶ Wenn Forschende allein Daten erheben, bearbeiten und analysieren, haben sie einerseits die volle Kontrolle über die angewendeten Methoden und deren Dokumentation sowie die Einhaltung der Datensicherheit. Andererseits bleiben eventuell Erkenntnisse in den Daten verborgen, die ohne die hetero-

4 SNF o.J.

5 FZI o.J.

6 Vgl. Warkentin 2019.

genen Blickwinkel eines interdisziplinären Expertenteams nicht zum Vorschein kommen können. Der Einsatz moderner Kommunikationstechnologien und digitaler Werkzeuge erlaubt dabei eine flexible, ortsunabhängige und gleichzeitige Kollaboration auch über institutionelle und räumliche Grenzen hinweg.

Jedoch stehen den Vorteilen des kollaborativen Arbeitens auch eine Reihe potenzieller Hindernisse und Gefahren gegenüber, die überwunden werden müssen, um die Qualität und Nachhaltigkeit der Forschungsergebnisse und -daten zu sichern. Die Bandbreite an Herausforderungen, die sich dabei stellen, wird beispielsweise durch die Ergebnisse einer 2018 durchgeführten, nicht-repräsentativen qualitativen Befragung von Infrastruktur-Teilprojekten von SFBs⁷ illustriert (siehe auch Tab. 1).⁸ SFBs sind große, von der DFG geförderte, interdisziplinäre und kollaborative Forschungsprojekte, im Rahmen derer auch spezielle Infrastruktur-Teilprojekte beantragt werden können, die sich dem Aufbau und Betrieb von Dateninfrastrukturen sowie weiteren Bereichen des Forschungsdatenmanagements (FDM) widmen.

Tab. 1: Herausforderungen in Infrastruktur-Teilprojekten von DFG-Sonderforschungsbereichen (n=20).

Herausforderung	Häufigkeit
Akzeptanz	15
Heterogenität	10
Ressourcenmangel	8
Technische Herausforderungen	7
Informationswissenschaftliche Herausforderungen	7
Rechtliche Aspekte	5
Schwierigkeiten bei der Personalakquise	3
Schwierige Kooperation	2
Sonstiges	4

Am häufigsten (von 15 der 20 teilnehmenden Projekte) wurden Probleme im Zusammenhang mit der Akzeptanz genannt, die vor allem mit einem auf Seiten der Forschenden oftmals nur schwach ausgeprägtem Bewusstsein für die Notwendigkeit und den Mehrwert des Datenmanagements zusammenhängen. Als weitere Aspekte wurden in diesem Zusammenhang zudem der zusätzliche Aufwand, unterschiedliche disziplinäre Wissenschaftskulturen sowie etablierte Forschungsprozesse, die durch die Berücksichtigung des Datenmanagements neu strukturiert werden müssen, angeführt. Die zweithäufigste Herausforderung, von der Hälfte der teilnehmen-

⁷ Vgl. Engelhardt 2020, 22–23.

⁸ Die Befunde bestätigen in ihrer Kernaussage die Ergebnisse einer ähnlichen, 2013 durchgeführten Untersuchung (vgl. Engelhardt 2020, 24–27; Engelhardt 2013).

den Projekte erwähnt, stellt die Heterogenität dar, die sich insbesondere in unterschiedlichen beteiligten Disziplinen, Forschungsmethoden, -prozessen und -daten, aber auch einem inhomogenem Kenntnis- und Kompetenzniveau in Bezug auf den Umgang mit Daten äußert. Die Entwicklung einer Infrastruktur, eines organisatorischen Rahmens und passender Unterstützungsangebote, die dieser Vielfalt gerecht werden, gestaltet sich schwierig. Damit hängen oftmals auch die technischen sowie informationswissenschaftlichen (z. B. die Entwicklung von Metadatenschemata) Herausforderung zusammen, die jeweils sieben Mal genannt wurden. Eine weitere Schwierigkeit, vor die sich acht der an der Befragung teilnehmenden Projekte gestellt sehen, sind mangelnde Ressourcen, vornehmlich in Bezug auf für FDM-Aufgaben zur Verfügung stehende Personalmittel. Weiterhin wurden rechtliche Aspekte (fünf Nennungen), Schwierigkeiten bei der Akquise von qualifiziertem Personal (drei Nennungen) sowie einige weitere, seltener vorkommende Einzelaspekte erwähnt.

Abb. 1 symbolisiert eine Zusammenstellung von fünf potentiellen Spannungsfeldern, die bei der Konzeption von kollaborativen Arbeitsumgebungen und -prozessen berücksichtigt werden sollten: Heterogenität des Konsortiums, Transparenz, Vertrauen, Datenqualität und Datensicherheit.

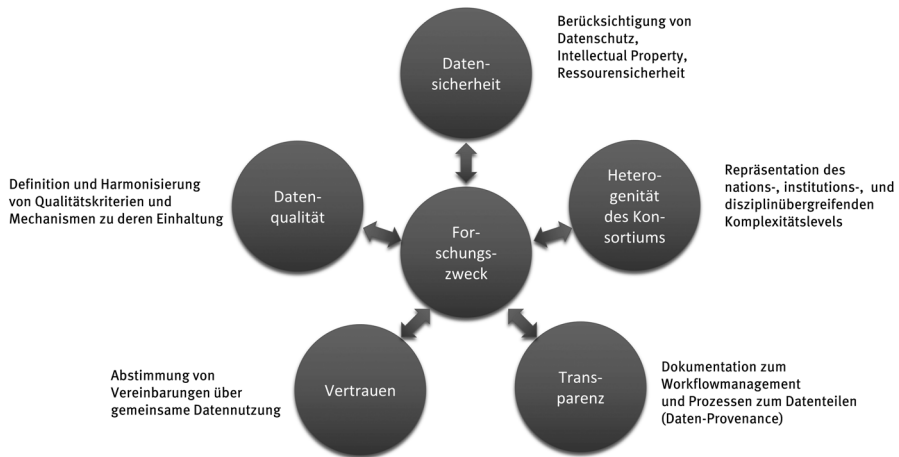


Abb. 1: Potentielle Spannungsfelder in der kollaborativen Forschung. Die Abbildung zeigt eine Zusammenstellung von fünf potentiellen Problemfeldern, die je nach Forschungszweck beim FDM zu berücksichtigen sind, damit kollaborative Arbeit mit Daten effizient und nachhaltig erfolgen kann (modifiziert nach Anhalt-Depies et al. 2019).

Je nach Forschungsszenario und Zweck des Umgangs mit Forschungsdaten variiert die Bedeutung der einzelnen Bereiche. Komplexitätsdimensionen der Heterogenität des Konsortiums existieren schon innerhalb der Fachbereiche, insbesondere aber disziplin- und auch grenzübergreifend. Wie oben bereits angesprochen, sind etwa viele DFG-geförderte SFBs fachübergreifend zusammengestellt, um wissenschaftliche Fragestellungen synergistisch aus verschiedenen Blickwinkeln zu bearbeiten. Daraus ergeben sich vielfältige Unterschiede in der Kultur und den Werkzeugen zur kollaborativen Datenbearbeitung, die in diesen SFBs zumindest anteilig harmonisiert werden müssen. Wenn verschiedene Partner an verteilten Standorten synergistisch forschen wollen, ist ein hohes Maß an Vertrauen in die gemeinsam generierten und zu bearbeitenden Daten erforderlich. Um dieses Vertrauen aufbauen zu können, ist es über die Gewährleistung der Datensicherheit notwendig, Datenerhebungs- und Verarbeitungsverfahren möglichst transparent und reproduzierbar zu gestalten. De Oliveira, de Oliveira und Mattoso beschreiben dazu Kernaspekte, die zur Reproduzierbarkeit von computerwissenschaftlichen Experimenten zu berücksichtigen sind, und formulieren eine entsprechende Terminologie.⁹ Die meisten dieser Aspekte können auch auf andere Wissenschaftsbereiche übertragen werden. Im Kern steht dabei die umfassende Dokumentation aller Daten und datenbezogenen Arbeitsabläufe („Daten-Provenance“). Umfangreiche Provenance-Informationen bilden eine der Grundvoraussetzungen für eine hohe Datenqualität. Schapke et al. adressieren zusätzlich die Aspekte der Kommunikation und Parallelität. Sie führen aus, wie das Kommunikationsmedium und die Art der Kommunikation die Form und Qualität der Zusammenarbeit beeinflussen.¹⁰ Besondere Herausforderungen stellen Aufgaben, die parallel bzw. zeitgleich von vielen verschiedenen Forschenden bearbeitet werden müssen.

Ein wiederkehrendes Problem bei der nachhaltigen Etablierung von in Projektkontexten entwickelten organisatorischen oder technischen Maßnahmen zum transparenten kollaborativen Forschen ist die Kurzfristigkeit der zugrundeliegenden Förderprogramme. Hier ist es empfehlenswert, frühzeitig gemeinsam mit den lokalen Infrastrukturanbietern mittel- bis langfristige Konzepte zu erarbeiten, die den Wissens- und Technologietransfer von Projekt zu Projekt verbessern können.¹¹

Um Hindernisse und Gefahren bei der kollaborativen Bearbeitung zu vermeiden oder zu minimieren, können eine Reihe organisatorischer und technischer Maßnahmen angewandt werden, auf die nachfolgend eingegangen wird.

⁹ Vgl. De Oliveira, de Oliveira und Mattoso 2017.

¹⁰ Vgl. Schapke et al. 2018.

¹¹ Vgl. Fleischer 2020; Stegemann 2020.

2 Formate und digitale Werkzeugkategorien für das kollaborative Arbeiten mit Daten

Kollaborative Forschung kann auf verschiedenen Ebenen organisiert und dadurch unterschiedlich komplex sein. Jedes Organisationsformat hat andere Anforderungen an Prozesse und Werkzeuge, die das gemeinsame Datenerheben und -bearbeiten ermöglichen. Im einfachsten Fall werden schon innerhalb einer Arbeitsgruppe Werkzeuge eingesetzt, die die zeitgleiche digitale Datenerfassung erleichtern und harmonisieren können. Übliche Anwendungskategorien sind hier z. B. „Electronic Data Capture“ (EDC)-Systeme, Datenbanken, elektronische Laborbücher (ELN) oder Plattformen zum gemeinsamen Schreiben wissenschaftlicher Texte.

In der Konsortialforschung nehmen mit steigender Komplexität auch die Anforderungen an solche Systeme stark zu. Eine Kategorisierung der wesentlichen Organisationsaspekte und eingesetzten Werkzeuge wird in den folgenden Abschnitten näher betrachtet.

Spezielle Kollaborationsformate können sich auch aus der grundsätzlichen Art und Lokalisierung der Datenerhebung ergeben. Bei der Feldforschung (z. B. in der Ökologie oder den Geowissenschaften) ist es u. U. notwendig, Werkzeuge einzusetzen, die teilweise ohne Internet („offline“) oder basierend auf besonders einfacher und robuster Hardware funktionieren müssen. Im Labor werden oft Dokumentationsprozesse und Werkzeuge benötigt, die wenig Zeit und Platz beanspruchen, damit sie effizient eingesetzt werden können. Auch kollaborative Datenbearbeitung aus dem Homeoffice oder von sonstigen zugangsbeschränkten Arbeitsplätzen erfordert u. U. spezielle IT-Werkzeuge, um Zugang zu allen wesentlichen Arbeitsumgebungen zu erhalten.

Die folgenden Abschnitte beleuchten wesentliche Querschnittsaspekte, die im kollaborativen Setting von besonderer Bedeutung sind. Praktische Lösungsansätze für das optimierte kollaborative Erheben und Bearbeiten von Forschungsdaten finden sich in der Fachliteratur unter Schlagworten wie „Virtuelle Forschungsumgebungen“ oder „Data Management Framework“. Dabei stehen Themen wie die Beschreibung neuer Werkzeuge und Prozesse, der Datensicherheit und zunehmend auch der Datenqualität im Vordergrund.

2.1 Institutions- und disziplinübergreifende Projekte

Konsortialforschung

Häufig organisieren sich Forschungsinitiativen in standort- und/oder fachübergreifenden Konsortien wie z. B. DFG-geförderten SFBs oder Forschungsgruppen. Um in diesen Konsortien effizient kollaborativ arbeiten zu können, werden abhängig vom

Szenario adäquate technische und organisatorische Arbeitsabläufe genutzt und entwickelt, die i. d. R. auf Vorerfahrungen der beteiligten Konsortialpartner beruhen. Dies stellt die standortübergreifenden Ansätze vor die Herausforderung, dass unterschiedliche, an den jeweiligen Standorten etablierte Vorgehensweisen harmonisiert werden müssen. Üblich ist hierbei, einen Standort (z. B. den Sprechersitz) als Zentrum zu definieren, an dem verschiedene technische Lösungen zusammenlaufen.

Eine dynamische Vielzahl von digitalen, meist web-, teils cloudbasierten Werkzeugen steht für das kollaborative Arbeiten zur Verfügung. Teilweise ermöglichen die Werkzeuge die gleichzeitige Bearbeitung durch viele Anwenderinnen und Anwender.

In vielen Anwendungsszenarien wird es kaum möglich sein, für jedes neue FDM-Werkzeug separate Server-Hardware zu beschaffen und zu betreiben. Hierfür bieten professionelle Rechenzentren die Möglichkeit, Cloud-Infrastruktur zu nutzen. Weiterführende Übersichten über Definitionen zu Begriffen und verschiedene Bereiche des Cloud Computing sind in der Fachliteratur beschrieben.¹²

Wichtige Anwendungskategorien bei kollaborativen FDM-Services sind u. a. Werkzeuge für gemeinsame Datenerhebung,¹³ Datenaufbereitung (z. B. OpenRefine¹⁴), Data Preparation,¹⁵ Datenspeicherung und -verteilung (z. B. ownCloud¹⁶, SharePoint,¹⁷ für eine Übersicht verschiedener Möglichkeiten des File Sharing siehe UK Data Service o. J. b), Softwareentwicklung (GitHub,¹⁸ GitLab¹⁹), Workflowmanagement (z. B. KNIME,²⁰ Taverna²¹), Materialverwaltung (z. B. QUARTZY,²² eLABInventory²³), Literaturverwaltung (z. B. Zotero,²⁴ Mendeley²⁵), Dokumenterstellung

12 Vgl. z. B. Antonopoulos und Gillam 2017; Repschläger, Pannicke und Zarnekow 2010.

13 Vgl. z. B. elektronische Laborbücher s. Adam und Lindstädt 2019; forschungsdaten.org o. J.; BExIS 2 o. J.

14 S. <https://openrefine.org/>. Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

15 Vgl. z. B. talend.com o. J.

16 S. <https://owncloud.org/>.

17 S. <https://www.microsoft.com/de-de/microsoft-365/sharepoint/collaboration>.

18 S. <https://github.com>.

19 S. <https://about.gitlab.com/>.

20 S. <https://www.knime.com/knime-analytics-platform>.

21 S. <https://taverna.incubator.apache.org/>.

22 S. <https://www.quartzy.com/>.

23 S. <https://www.elabinventory.com/>.

24 S. <https://www.zotero.org/>.

25 S. <https://www.mendeley.com/>.

(z. B. ShareLaTeX,²⁶ HedgeDog [ehem. CodiMD]²⁷), virtuelle Meetings (z. B. Jitsi,²⁸ BigBlueButton²⁹) und Datenanalyse (z. B. Jupyter Notebooks,³⁰ Galaxy³¹).

Einhergehend mit der standortübergreifenden Datenverarbeitung entsteht der Bedarf nach Regelwerken („Data Policies“, Datennutzungsvereinbarungen, für Beispiele siehe etwa eResearch Alliance o. J.), die Rechte und Pflichten der beteiligten Konsortialpartner dokumentieren. Damit diese Regeln alle Anwendungsfälle erfassen und von allen Partnern konsentiert werden können, sollten sie gemeinsam erarbeitet und dynamisch angepasst werden. Eine Auswahl verschiedener individueller Setups und Herangehensweisen wurde u. a. in zwei Workshops zu Infrastruktur-Teilprojekten vorgestellt.³²

Citizen Science

Die Beteiligung von Laien-Forschenden bei der Bearbeitung wissenschaftlicher Fragestellungen gewinnt immer mehr an Bedeutung.³³ Hierdurch ergeben sich aber auch eine Reihe von speziellen Herausforderungen z. B. beim Design von Werkzeugen für die Datenerhebung, bei der Publikation von wissenschaftlichen Ergebnissen, die Daten aus Citizen-Science-Ansätzen verwenden oder auch bei der Einordnung von Datensicherheitsaspekten, wie sie z. B. näher von Anhalt-Depies et al. untersucht und beschrieben wurden.³⁴

2.2 Virtuelle Forschungsumgebungen

Virtuelle kollaborative Forschungsumgebungen (VFU) existieren in einer Vielzahl von Ausprägungen unter verschiedenen Begrifflichkeiten. Dabei handelt es sich häufig um spezifische Werkzeug-Setups für bestimmte Anwendungsfälle zur Unterstützung der Kommunikation und Kollaboration von geographisch verteilten Forschenden.³⁵

26 S. <https://www.sharelatex.com/>.

27 S. <https://demo.hedgedoc.org/>.

28 S. <https://jitsi.org/>.

29 S. <https://bigbluebutton.org/>.

30 S. <https://www.jupyter.org/>.

31 S. <https://usegalaxy.org/>.

32 Vgl. Engelhardt 2013; Roertgen et al. 2019; bausteine-fdm.de 2020.

33 Für Definitionen vgl. Eitzel et al. 2017; Rowbotham et al. 2019.

34 Vgl. Anhalt-Depies et al. 2019.

35 Vgl. Ahmed, Poole und Trudeau 2018, 688; Kommission Zukunft der Informationsinfrastruktur 2011, B74–B81; Candela, Castelli und Pagano 2013.

VFUs lassen sich nach funktionalen sowie nach technologischen Gesichtspunkten kategorisieren. Betrachtet man das organisatorische Setup und die wissenschaftlichen Funktionen, lassen sich vier Typen von VFUs mit einem Forschungsfokus identifizieren:³⁶

- *Distributive Research Center (DRS)*: eine Umgebung, in der Expertise sowie personelle und andere Ressourcen rund um ein Forschungsthema zusammengeführt werden und die die Bearbeitung entsprechender Projekte unterstützt;
- *Shared Instruments (SI)*: ermöglicht Remote-Zugang zu Großinstrumenten wie bspw. Teleskopen;
- *Community Data Systems (CDS)*: eine Umgebung, die von einer dispersen Community als Repositorium für die Sammlung von Daten und Informationen genutzt und gepflegt wird;
- *Open Community Contribution Systems (OCCS)*: führt die Einzelbeiträge einer großen Zahl räumlich verstreuter Individuen bei der Bearbeitung eines konkreten gemeinsamen Unterfangens, etwa der Bestimmung von Proteinfaltungen, zusammen.

Basierend auf einer Untersuchung der eingesetzten Informationstechnologien von 164 VFUs bildeten Ahmed, Poole und Trudeau die folgenden fünf Kategorien:³⁷

- *Webpage*: Der Schwerpunkt liegt hier auf Technologien, die die Publikation und Kommunikation unterstützen. VFUs dieser Art dienen primär als Online-Präsenz oder Portal zu weiteren Ressourcen (z. B. Daten, Tools oder Publikationen). Als Beispiel nennen die Autorinnen und Autoren das LIPID MAPS Lipidomics Gateway.³⁸
- *Analytic Community*: Dieser Typ von VFU vereinigt Rechenkapazität, Publikations- und Kommunikationstechnologien. Er umfasst i. d. R. ein Datenrepositorium und stellt Funktionen und Kapazitäten zur Datenanalyse und -annotation zur Verfügung. Genutzt wird er von einer Gemeinschaft von Wissenschaftlerinnen und Wissenschaftlern, die diese Ressourcen für ihre Forschung nutzen und zugleich ein starkes Interesse am Austausch von Ergebnissen und Publikationen sowie der Interaktion mit anderen Nutzerinnen und Nutzern haben. Als Beispiel führen Ahmed, Poole und Trudeau Archaeotools³⁹ an.
- *Moderate Scientific Support/Intense Scientific Support VRE*: Diese beiden sich in ihrer technischen Konfiguration stark gleichenden Typen unterscheiden sich primär in der Anzahl der integrierten Informations- und Kommunikationstechnologien. Der Schwerpunkt liegt bei ihnen auf Instrumenten, Daten und Re-

³⁶ Vgl. Ahmed, Poole und Trudeau 2018, 689–691.

³⁷ Vgl. Ahmed, Poole und Trudeau 2018, 693–698.

³⁸ S. <http://www.lipidmaps.org/>.

³⁹ S. <https://archaeologydataservice.ac.uk/research/archaeotools.xhtml>.

chenressourcen. Zusätzlich werden in geringem Grad Funktionalitäten für Publikation und Kommunikation bereitgestellt. Ein gutes Beispiel für eine Moderate-Scientific-Support-VFU ist laut Ahmed, Poole und Trudeau NEESGrid,⁴⁰ für eine Intense-Scientific-Support-VFU SEEK.⁴¹

- *Archetypical collaboratory*: Diese Kategorie von VFUs integriert sowohl in hohem Maße Instrumente sowie Funktionen zur Datenspeicherung, -archivierung, -analyse, -annotation usw. als auch zahlreiche Publikations- und Kommunikationswerkzeuge und stellt damit eine multidimensionale Umgebung für kollaborative Forschung bereit, während sie zugleich das Community Building um den Forschungsgegenstand herum befördert. Als Beispiel einer VFU dieses Typs geben Ahmed, Poole und Trudeau die Cochrane Collaboration an.⁴²

2.3 Organisatorischer und regulativer Rahmen

Ein organisatorischer und regulativer Rahmen, ein sogenanntes „Data Management Framework“, auf der Ebene einer Institution oder eines großen Verbundprojektes kann dabei helfen, ein einheitliches und effektives Datenmanagement in der gesamten Institution bzw. dem Projekt zu gewährleisten. Er ermöglicht es, einen Überblick über (Teil-)Projekte und in ihnen erhobene Daten zu er- und behalten, die organisatorischen und regulatorischen Rahmenbedingungen für den Umgang mit Forschungsdaten sowie die damit zusammenhängenden Aufgaben und Zuständigkeiten zu definieren und die entsprechenden Informationen und Ressourcen zentral für alle bereitzustellen.⁴³

Um einen umfassenden Rahmen zu erarbeiten, der alle notwendigen Aspekte abbildet, sollten laut ANDS fünf Bereiche berücksichtigt werden:⁴⁴

- Institutionelle Policies und (standardisierte) Verfahren,
- Angebote zur Beratung und Unterstützung,
- IT-Infrastruktur,
- Metadatenmanagement sowie
- Forschungsdatenmanagement.

Idealerweise werden für jedes dieser Felder Rollen und Zuständigkeiten sowie Policies und Standards definiert und die notwendigen Ressourcen (finanziell, personell, sonstige Ausstattung) allokiert. Zudem muss dafür Sorge getragen werden, dass die

⁴⁰ S. <http://www.neesgrid.org/>.

⁴¹ S. <https://seek4science.org/>.

⁴² S. <https://www.cochrane.org/welcome>.

⁴³ Vgl. UK Data Service o. J. c.

⁴⁴ Vgl. ANDS 2018, 2.

Beteiligten die notwendige Qualifikation für ihre Aufgaben besitzen oder sich aneignen (können).⁴⁵

Relevante Dokumente und Informationen können an einer zentralen Stelle für alle zugänglich vorgehalten werden. Zu diesen können, je nach Bedarf, folgende zählen:⁴⁶

- zentrales Datenverzeichnis,
- institutionelle Erklärung oder Policy zum Datenteilen,
- Beispiel-Datenmanagementpläne,
- Informationen zu institutionellen Richtlinien und Regularien in Bezug auf Forschungsdaten,
- Erklärung zu Urheber- und Verwertungsrechten für Forschungsdaten und -ergebnisse,
- Erklärung zur institutionellen Datenmanagement-Infrastruktur und zu Backup-Verfahren,
- IT-Sicherheitsrichtlinie für Datenspeicherung, File Sharing und Datenübertragung,
- Empfehlungen für Standard-Datenformate,
- Policy zum Zurückhalten oder zur Löschung von Daten
- Standards für die Qualitätskontrolle von Datenerhebung und -eingabe,
- Empfehlungen für Digitalisierung und Transkription,
- Empfehlungen für Dateibenennung und Versionskontrolle,
- Projektverzeichnis einschließlich Rollen und Zuständigkeiten,
- Vorlagen für (informierte) Einwilligungserklärungen und Informationen zum Datenteilen im Zusammenhang mit sensiblen Daten,
- Vertraulichkeitsvereinbarungen für den Umgang mit sensiblen Daten,
- Informationen zu ethischen Aspekten und Beispielformulare für eine Begutachtung,
- Richtlinien für Anonymisierung und Pseudonymisierung.

2.4 Datenschutz und Datensicherheit

Die rechtliche Perspektive beim FDM wird zunehmend wissenschaftlich betrachtet.⁴⁷ Kollaboratives Arbeiten mit Forschungsdaten erfordert in besonderer Weise die Berücksichtigung von Aspekten der Datensicherheit bzw. des Datenschutzes. Diese Begriffe werden in unterschiedlichen fachlichen, technischen oder rechtli-

⁴⁵ Vgl. ANDS 2018, 2.

⁴⁶ Vgl. UK Data Service o. J. a.

⁴⁷ Vgl. z. B. Hartmann 2019; DATAJUS o.J.

chen Kontexten heterogen definiert und verwendet.⁴⁸ Durch die zunehmende Anzahl von fach- und standortübergreifenden Initiativen steigt auch die Komplexität von Risikoabschätzungen, um Datenschutz und Datensicherheit praxisnah ausreichend gewährleisten zu können.⁴⁹ Bei länderübergreifenden Kooperationen kommt durch unterschiedliche Gesetzgebungen gegebenenfalls noch eine weitere Komplexitätsebene hinzu.⁵⁰

Zur Abgrenzung der Aspekte, die für den konkreten kollaborativen Anwendungsfall relevant sind, sollten die entstehenden Daten und zugehörigen Arbeitsabläufe inhaltlich, technisch und rechtlich kategorisiert und evaluiert werden. Des Weiteren ist die möglichst konkrete Beschreibung des Zwecks der Umsetzung von Maßnahmen zur Erhöhung der Datensicherheit wesentlich, um Aspekte priorisieren zu können.

Sensible Forschungsdaten

In vielen Forschungskontexten wird mit sensiblen Daten gearbeitet. Die Sensibilität kann dabei durch eine Vielzahl verschiedener Hintergründe wie z. B. Personenbezug, Natur- und Artenschutz, Firmengeheimnisse oder Forschung an Gefahrstoffen entstehen. Besonders häufig fallen personenbezogene Daten an. Der Personenbezug rangiert dabei zwischen einerseits i. d. R. unkritischen wenigen persönlichen Angaben zu beteiligten Forschenden an gemeinsamen Projekten und andererseits höchst sensitiven Informationen zum ökonomischen oder medizinischen Status von Probandinnen bzw. Probanden in Umfragen oder Studien.⁵¹ Eine Vielzahl an Publikationen und technischen Ansätzen befasst sich hierbei mit besonders kritischen Szenarien z. B. in der sozialwissenschaftlichen oder klinischen Forschung.⁵² Aber auch in einfacheren Setups erscheint die Betrachtung von datenschutzbezogenen Fragen sinnvoll⁵³, damit ein abgestimmtes Vorgehen in den Fachcommunities erreicht werden kann.

Intellectual Property/Lizenzen

Insbesondere in standortübergreifenden Forschungskollaborationen sollten abgestimmte Regeln zum geistigen Eigentum (Intellectual Property, IP) und den daraus

⁴⁸ Vgl. z. B. Schmidt und Weichert 2012; GDD o. J.; Solove 2006.

⁴⁹ Vgl. Stiles and Petrilá 2011; Nurmi et al. 2019.

⁵⁰ Vgl. Nurmi et al. 2019, 9; Kalberg 2012; Röttgen 2020; Lauber-Rönsberg 2018.

⁵¹ Vgl. Europäische Kommission o. J.; DSGVO 2016, Art. 4.

⁵² Vgl. Pommerening et al. 2014; Bauer, Eickmeier und Eckard 2018; Eaton und McNett 2020.

⁵³ Vgl. z. B. Schallaböck und Grafenstein 2017.

resultierenden Konsequenzen zum Daten-Besitztum und -Nutzungsrecht erarbeitet werden. Mehrere aktuelle Forschungsprojekte liefern hierbei erste v. a. rechtliche Grundlagen, die berücksichtigt werden können.⁵⁴ Gerade im Hinblick auf Publikationen, denen gemeinsam erarbeitete Forschungsdaten zugrunde liegen, sind auch die Regelungen zur Guten wissenschaftlichen Praxis relevant.⁵⁵ Weiterführende Informationen zu Lizenzmodellen, die die Nachnutzung von Daten regulieren, wurden in verschiedenen Projektkontexten zusammengestellt.⁵⁶

Technische und organisatorische Aspekte

In vielen kollaborativen Arbeitsszenarien sind zunächst die einzelnen beteiligten Standorte für die technische Sicherheit der primär erhobenen Daten verantwortlich. Zusätzlich sollten Konzepte erarbeitet werden, um die auf dieser Basis kollaborativ entstehenden Forschungsdaten vor IT-Sicherheitsproblemen (z. B. Datenverlust oder -beschädigung, Hackerangriffe) bestmöglich zu schützen. Zunehmend wird institutionell per Daten-Policies reguliert, dass erhobene Daten am Erhebungsstandort verbleiben bzw. im Falle der Feldforschung am Standort des erhebenden Projektes zusammengeführt werden. Diese Vorgabe schließt insbesondere die Nutzung einiger Drittanbieter-Services aus, die im privaten Bereich weite Verbreitung finden (z. B. Dropbox, Google Drive, Microsoft OneDrive). Wenn verfügbar, sollten zertifizierte Rechenzentren (z. B. ISO 9001, ISO 27001) mit ihren alternativen Serviceangeboten auch für die kollaborative Datenhaltung priorisiert werden, um Sicherheitsrisiken zu minimieren.

Den oben genannten Herausforderungen bezüglich personenbezogener oder IP-rechtlich geschützter Daten kann durch technisch-organisatorische Lösungsansätze z. T. begegnet werden. Hierzu zählen Mechanismen zu Pseudonymisierung und Anonymisierung, die es z. B. durch Datenaggregation ermöglichen, mit persönlichen Angaben Forschungsfragen aus verschiedenen Fachbereichen zu bearbeiten.⁵⁷ Ein weiterer Ansatzpunkt besteht in der differenzierten Zugangskontrolle zu personenbezogenen Daten. Diese kann physikalisch erfolgen, indem zugangsbeschränkte und kontrollierte Räume zur Dateneinsicht zur Verfügung gestellt werden.⁵⁸ Alternativ können digitale Zugriffsmodelle eingesetzt werden, die im einfacheren Szenario mit Use- & Access-Rollen oder in sensitiveren Bereichen mit der Zwischenschaltung von fachlich versierten Gremien zur Überwachung der Dateneinsicht und -her-

⁵⁴ Vgl. z. B. Kreutzer und Lahmann 2019; Ostendorff und Linke 2019; forschungsdaten.info o. J.; open-access.net o. J.; DATAJUS o. J.

⁵⁵ Vgl. DFG 2019.

⁵⁶ Vgl. z. B. forschungslizenzen.de o. J., radar-service.eu o. J.

⁵⁷ Vgl. Chevrier et al. 2019; Gkoulalas-Divanis und Loukides 2015.

⁵⁸ Vgl. z. B. EUROSTAT 2017; gesis.org o. J.

ausgabe realisiert werden.⁵⁹ Das Vorgehen kann dabei u. U. die Integration von Datentreuhanddiensten erforderlich machen.⁶⁰ Solche Datentreuhandstellen werden z. B. eingesetzt, um im Falle personenbezogener Daten Aufgaben der Pseudonymisierung oder der Anonymisierung wahrzunehmen und die Daten nur pseudonymisiert, anonymisiert oder aggregiert zur Verfügung zu stellen.

2.5 Datenqualität

Der Rat für Informationsinfrastrukturen (RfII) hat eine Vielzahl von Herausforderungen in Bezug auf die Datenqualität identifiziert.⁶¹ An dieser Stelle sollen diejenigen hervorgehoben werden, die im kollaborativen Kontext besondere Relevanz besitzen. Ein Großteil davon steht im Zusammenhang mit der Definition und Anwendung einheitlicher Regeln und Standards für verschiedene im Verlauf des Forschungsprozesses anfallende Aktivitäten. Hier ist vor allem eine unvollständige und uneinheitliche Dokumentation der Daten sowie des Forschungsprozesses und -kontextes zu nennen, die der Nachvollziehbarkeit und Reproduzierbarkeit wissenschaftlicher Ergebnisse entgegensteht. Dies stellt oft schon in der Erhebungsphase ein Problem dar, das – sofern nicht gelöst – im weiteren Verlauf, etwa bei der Archivierung und Publikation oder auch der Zusammenführung von Daten aus unterschiedlichen Quellen, relevant bleibt. Abhilfe können hier einheitliche Richtlinien und Konventionen, z. B. in Bezug auf Metadaten, Dateibenennung und Versionierung, schaffen, die sich – so vorhanden – an in der Fachcommunity etablierten Standards orientieren. Unterstützend kann auch Dokumentationssoftware, z. B. die oben bereits angesprochenen Elektronischen Laborbücher, eingesetzt werden. Für die Qualitätskontrolle und Datenbereinigung sollten außerdem geeignete, einheitliche Kriterien definiert werden, die von allen Beteiligten angewendet werden.

Eine weitere Herausforderung, auf die der RfII hinweist, sind Hard- und Softwareunterschiede, z. T. auch physikalische Umwelteinflüsse, die zu Diskrepanzen in Verarbeitungsprozessen und Ergebnissen und mithin zu Problemen bei der Reproduzierbarkeit führen können. Beim Einsatz von Software können zudem Bedienfehler wie auch eine ungeeignete oder ungenügend dokumentierte Parametrisierung eine Minderung der Datenqualität bedingen. Für viele dieser Probleme gibt es derzeit noch keine ideale Lösung, durch verbesserte Dokumentation können sie jedoch immerhin transparent gemacht werden. Auch der Einsatz quelloffener, frei verfügbarer Software kann in diesem Zusammenhang hilfreich sein.

⁵⁹ Vgl. z. B. Shah, Coathup, Teare et al. 2019; Torcivia-Rodriguez, Hayley-Dingerdissen und Mazumder 2019; Dyke 2020; EGA o. J.; DZHK 2014.

⁶⁰ Vgl. RfII 2020.

⁶¹ Vgl. RfII 2019, 29–57.

Weitere, vom RfII angesprochene relevante Aspekte beziehen sich auf die Klärung der Verfügungsrechte über Daten sowie die Definition und eindeutige Zuweisung von Rollen und Aufgaben, die z. B. bereits in der Planungsphase im Rahmen der Erstellung eines Datenmanagementplans erfolgen kann.

2.6 Softwareentwicklung

Schon lange sind in der Softwareentwicklung digitale Werkzeuge üblich, um effizient und zeitgleich im Team Produkte weiterentwickeln zu können. Einige Open-Source-Projekte werden dabei von einer weltweiten Community von Entwicklerinnen bzw. Entwicklern diskutiert und vorangetrieben.⁶² Hierzu zählen auch viele Werkzeuge, die für das Projektmanagement oder das Arbeiten mit (Forschungs-)Daten genutzt werden können.⁶³

Zunehmend an Bedeutung gewinnen in der Softwareentwicklung die Aspekte der Reproduzierbarkeit⁶⁴ und Nachhaltigkeit („software sustainability“⁶⁵) und der Agilität.⁶⁶ Beide Schwerpunkte adressieren die steigenden Anforderungen, in heterogenen und dynamischen Teamzusammensetzungen auf Bedarfe durch die sich schnell wandelnden wissenschaftlichen Herangehensweisen reagieren zu können und die daraus resultierenden Softwarelösungen langlebig und reproduzierbar zu dokumentieren.

Als Entwicklungsumgebung, Versionierungssystem oder Code-Repository genutzte Applikationen basieren oft auf Lösungen wie Github, Gitlab oder Subversion.⁶⁷ In solchen Softwareprojektmanagement-Umgebungen kann Programmiercode verwaltet und versioniert, aber auch ausführlich dokumentiert und veröffentlicht werden.⁶⁸

Die „Task Group Forschungssoftware“ des Arbeitskreises Open Science in der Helmholtz Gemeinschaft hat kürzlich eine Muster-Richtlinie für nachhaltige Forschungssoftware herausgegeben, die wichtige Aspekte zum FAIRen Umgang mit Softwareentwicklungen zusammengestellt und als Vorlage für andere Institutionen genutzt werden kann.⁶⁹

⁶² Vgl. von Krogh und von Hippel 2003; Kogut und Metiu 2001.

⁶³ Vgl. Patel 2016; Willmes, Kürner und Bareth 2014.

⁶⁴ Vgl. David et al. 2017; Hothorn, Held und Friede 2009; Hothorn und Leisch 2011; Freire, Fuhr und Rauber 2016.

⁶⁵ Vgl. Crouch et al. 2013; Venters et al. 2018; Anzt et al. 2020.

⁶⁶ Vgl. Abrahamsson et al. 2017; Lechler und Yang 2017; it-agile.de o.J.

⁶⁷ S. <https://subversion.apache.org/>.

⁶⁸ Vgl. z. B. Simons 2020; Bender 2018.

⁶⁹ Vgl. Bach et al. 2020.

Weitere Informationen über Netzwerke von Forschungssoftware-Communities können z. B. auf den Webseiten des „EURISE Network“⁷⁰ und der „Research Software Alliance“⁷¹ gefunden werden.

3 Praxistransfer

Bereits durch die „Kommission Zukunft der Informationsinfrastruktur“ wurde erarbeitet, dass Flexibilität ein essentielles Charakteristikum von kollaborativen Forschungsinfrastrukturen für deren Nutzbarkeit zum effizienten Erkenntnisgewinn ist.⁷² In der Folge betonte auch der auf Basis der Kommissionsarbeit begründete Rat für Informationsinfrastrukturen die Bedeutung der Vielfalt bei der Implementierung von Strukturen und Prozessen im FDM.⁷³ Für den Praxistransfer der beschriebenen Aspekte zum kollaborativen Arbeiten mit Daten ergibt sich daraus die Herausforderung, dass konkrete Umsetzungsschritte spezifisch für jedes Forschungsszenario erarbeitet und adaptiert werden müssen.

Eine wesentliche Rolle spielt dabei eine möglichst effiziente Kommunikation zwischen den Beteiligten innerhalb eines Konsortiums einerseits und zwischen Infrastruktur-Entwickelnden bzw. -anbietern und den forschenden Nutzerinnen und Nutzern andererseits. Hilfreich ist darüber hinaus die Kommunikation und Netzbildung zwischen Forschungsverbänden am eigenen und mit anderen Wissenschaftsstandorten.

Für die praktisch-technische Umsetzung einer kollaborativen Forschungsinfrastruktur können vier Komplexitätsebenen betrachtet und genutzt werden, die einen steigenden Grad an Aufwand und benötigter Expertise, aber auch an verbesserter Adaption an spezifische Herausforderungen bedeuten.

Integration von Services, Werkzeugen und anderen Vorlagen, die ad hoc eingesetzt werden können

Insbesondere Webressourcen können häufig mit geringem Aufwand in Forschungsprozesse integriert werden. Beispielsweise können Forschungsdesign, Projektstrukturen, Workflows und Rollenverteilungen in Webportalen wie dem Open Science

⁷⁰ S. <https://technical-reference.readthedocs.io/en/latest/>.

⁷¹ S. <https://www.researchsoft.org/resources/>.

⁷² Vgl. KII 2011.

⁷³ Vgl. RfII 2016.

Framework⁷⁴ oder FAIRDOMHub⁷⁵ gepflegt und veröffentlicht werden. Forschungsdaten können auch in weiteren öffentlichen Repositorien – lokal oder übergreifend – hinterlegt und somit ideal in kollaborativen Umgebungen ausgetauscht sowie eindeutig zitiert werden (z. B. Zenodo, GRO.data⁷⁶). Einige z. T. kommerzielle Anbieter ermöglichen auch den direkten Einsatz („software as a service“) von elektronischen Laborbüchern für die primäre Dokumentation der Forschungsarbeit (z. B. LabArchives,⁷⁷ RSpace⁷⁸). Zur vereinfachten Kommunikation können Videokonferenzsysteme oder Chat-Dienste (z. B. Rocket.Chat⁷⁹) externer Anbieter genutzt werden. Üblicherweise werden solche Dienste auch von den lokalen IT-Serviceanbietern bereitgestellt. Probleme in der Anwendung können sich in dieser Komplexitätsebene z. B. aus den geringen Adaptionmöglichkeiten und der Datenhaltung auf externen Servern ergeben.

Individuelle Nachnutzung bzw. Installation oder Einbindung von virtualisierten Services, Werkzeugen und anderen Vorlagen

Viele FDM-Werkzeuge werden inzwischen als virtualisierte Systeme zur Nachnutzung angeboten. Ein weit verbreitetes Format ist dabei die Bereitstellung als Docker-Container. Die Virtualisierung vereinfacht die initiale Implementierung, da vertiefte Kenntnisse über Abhängigkeiten von Komponenten der Systeme zunächst nicht erforderlich sind und die Werkzeuge für spezifische Adaptionen leichter angepasst werden können. Der Einsatz solcher Umgebungen in kollaborativen Szenarien erfordert ein größeres Maß an IT- und werkzeugspezifischer Expertise, um eine effiziente und sichere Implementierung gewährleisten zu können. I. d. R. wird dafür die Zusammenarbeit mit einem professionellen Rechenzentrum unerlässlich sein. Beispiele für die Verfügbarkeit von virtualisierten Services sind Open-Source-Projekte wie Jupyter Notebooks zur reproduzierbaren Dokumentation von Datenanalyse-Algorithmen, Dataverse als Forschungsdaten-Repositorium,⁸⁰ FEniCS⁸¹ für die wissenschaftliche Modellierung oder das Webportal menoci⁸² für die Repräsentation bio-medizinischer Forschungsdaten.

74 S. <https://osf.io/>.

75 S. <https://fairdomhub.org/>.

76 S. <https://data.goettingen-research-online.de/>.

77 S. <https://www.labarchives.com/>.

78 S. <https://www.researchspace.com/>.

79 S. <https://rocket.chat/>.

80 S. <https://dataverse.org/home>.

81 S. <https://fenicsproject.org/>.

82 S. <https://menoci.io/>.

Individuelle Nachnutzung bzw. Installation oder Einbindung von Services, Werkzeugen und anderen Vorlagen mit Hilfe von Installationsanleitungen der Entwickelnden

Höherer Aufwand ist notwendig, um FDM-Werkzeuge zu implementieren, die bisher nicht in virtualisierter Form zur Verfügung gestellt werden. Insbesondere bei komplexen Systemen, die eine Vielzahl an Abhängigkeiten zwischen benötigten Ressourcen beinhalten, sind vertiefte IT-Kenntnisse und Erfahrungen mit den jeweiligen Systemen erforderlich, um einen störungsfreien Einsatz zu ermöglichen. Beispiele in dieser Kategorie sind das Daten-Portal ckan,⁸³ das System SADE für Digitale Editionen⁸⁴ oder die Datenerfassungsumgebung BExIS.⁸⁵

Neuentwicklung eigener Services, Werkzeuge oder Prozesse, die sich an Erfahrungen von anderen Standorten/Projektszenarien orientieren

Wenn bereits bestehende FDM-Systeme in ihrer Funktionalität nicht genügen, um spezifischen Anforderungen an das kollaborative Forschungsszenario zu begegnen, ist es gegebenenfalls sinnvoll, Eigenentwicklungen zu designen. Üblicherweise orientieren sich diese Neuentwicklungen an Erfahrungen aus ähnlichen Projekthintergründen und bauen auf diesem Vorwissen auf. Für diese Herangehensweise ist ein hoher Aufwand an Ressourcen erforderlich, um das Ziel spezifischer Funktionalität tatsächlich erreichen zu können. Beispiele sind das Drupal-basierte Fachrepositorium publisso⁸⁶ des Informationszentrums Lebenswissenschaften,⁸⁷ die geowissenschaftlichen Projektdatenbanken von SFB/TR 32,⁸⁸ SFB 1211⁸⁹ und SFB/TRR 228⁹⁰ oder EXMARALDA⁹¹ als ein System für das computergestützte Arbeiten mit (vor allem) mündlichen Korpora.

Best-Practice-Beispiele aus DFG-geförderten Sonderforschungsbereichen wurden u. a. in einem Workshop vorgestellt und dokumentiert.⁹² Zwei weitere Anwendungsbeispiele, exemplarisch aus den Geowissenschaften und der Medizin, geben

83 S. <https://ckan.org/>.

84 S. <https://www.bbaw.de/bbaw-digital/telota/forschungsprojekte-und-software/abgeschlossene-projekte/sade>.

85 S. <https://fusion.cs.uni-jena.de/bpp/bexis2-software/>.

86 S. <https://repository.publisso.de/>.

87 Vgl. Arning, Lindstädt und Schmitz 2016.

88 S. <https://www.tr32db.uni-koeln.de/site/index.php>.

89 S. <https://sfb1211.uni-koeln.de/>.

90 S. <https://www.trr228db.uni-koeln.de/site/index.php>.

91 S. <https://exmaralda.org/de/>.

92 Vgl. Roertgen 2020; bausteine-fdm.de 2020; Schwandt 2020.

Einblicke in konkrete Implementierungsschritte für große kollaborative Forschungsverbände.

Implementierungsschritte für große kollaborative Forschungsverbände: Anwendungsbeispiel Geowissenschaften

Von 2007 an wurden im SFB/Transregio 32 („Muster und Strukturen in Boden-Pflanzen-Atmosphären-Systemen: Erfassung, Modellierung und Datenassimilation“) Forschungsfragen mit geowissenschaftlichem Fokus kollaborativ bearbeitet. Die entwickelten Infrastrukturen werden in den nachfolgenden Verbänden SFB 1211 („Evolution der Erde und des Lebens unter extremer Trockenheit“) und SFB/Transregio 228 („Zukunft im ländlichen Afrika: Zukunft-Machen und sozial-ökologische Transformation“) nachgenutzt und erweitert.⁹³ Alle Verbände führen in einer eigenentwickelten Projektdatenbank erhobene Forschungsdaten zusammen, reichern sie dabei mit standardisierten Metadaten⁹⁴ an und ermöglichen die Zitierfähigkeit von Datensätzen über Digital Object Identifier (DOI). Suchfunktionalität steht in Form von Suchmasken und Listen aber auch räumlich (kartenbasiert) zur Verfügung. Die FDM-Systeme wurden in Zusammenarbeit mit dem Regionalen Rechenzentrum (RRZK) der Universität zu Köln aufgebaut und werden dort physisch gehostet, damit auch die langfristige nachhaltige Verfügbarkeit und damit die Wiederverwendbarkeit aller Projektdaten über die Projektförderung hinaus sichergestellt bleibt.⁹⁵

Implementierungsschritte für große kollaborative Forschungsverbände: Anwendungsbeispiel Kardiologische Grundlagenforschung

Im klinischen Sonderforschungsbereich 1002 („Modulatorische Einheiten bei Herzinsuffizienz“) wird seit 2012 eine integrierte und langfristig verfügbare Forschungsdatenplattform⁹⁶ in enger Zusammenarbeit mit den beteiligten Forschenden und auf Basis des Datenmanagementportals menoci⁹⁷ entwickelt. Mit dem Ziel der vereinfachten Nachnutzung („FAIRification“) der im SFB 1002 generierten Forschungsdaten werden diese zentral und digital erfasst, strukturiert und standardisiert. Technische Schwerpunkte bilden dabei die Integration von Community-Standards, die Bereitstellung zitierfähiger Datensätze durch die Verwendung von Persistenten

⁹³ Vgl. Curdt et al. 2019.

⁹⁴ Vgl. Curdt 2014; Curdt 2016.

⁹⁵ Vgl. Curdt und Hoffmeister 2015.

⁹⁶ S. <https://sfb1002.med.uni-goettingen.de/production/>.

⁹⁷ Vgl. Suhr et al. 2020.

Identifikatoren (PID) und ihre verbesserte Integration. Zusätzlich nutzen einige Arbeitsgruppen des Konsortiums für die primäre Labordaten-Dokumentation ein elektronisches Laborbuch und haben hierdurch unterschiedlich weitreichend die papierbasierte Dokumentation abgelöst. Klinische Daten, die im Rahmen dieses Konsortiums erhoben werden, werden in einer distinkten digitalen Infrastruktur verwaltet. Dies erfolgt in Anlehnung an Entwicklungen für das Datenmanagement im Deutschen Zentrum für Herz-Kreislauf-Forschung e.V.⁹⁸ Eine schriftliche Vereinbarung („Data Policy“) über die gemeinsame Datennutzung wurde inzwischen in der zweiten Version erarbeitet und von allen Projektleitenden unterzeichnet. Teile der entwickelten Infrastrukturen und Prozesse werden bereits in weiteren Konsortien nachgenutzt sowie spezifisch erweitert (z. B. SFB 1190 und SFB 1286).

Literatur

Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

- Abrahamsson, Pekka, Outi Salo, Jussi Ronkainen und Juhani Warsta. 2017. „Agile Software Development Methods: Review and Analysis.“ *arXiv* (September): 1709.08439 [cs]. <http://arxiv.org/abs/1709.08439>.
- Adam, Beatrix und Birte Lindstädt. 2019. *ELN-Wegweiser – Elektronische Laborbücher im Kontext von Forschungsdatenmanagement und guter wissenschaftlicher Praxis – ein Wegweiser für die Lebenswissenschaften*. Köln: ZB MED – Informationszentrum Lebenswissenschaften. doi:10.4126/FRL01-006415715.
- Ahmed, Iftekhar, Marshall Poole und Ashley Trudeau. 2018. „A Typology of Virtual Research Environments.“ In: *Hawaii International Conference on System Sciences 2018, Big Island, Hawaii, 2–6 January 2018*, 688–697. Red Hook, NY: Curran Associates, Inc. doi:10.24251/HICSS.2018.087.
- ANDS. 2018. „Creating a Data Management Framework.“ ANDS. 23. März 2018. <https://www.ands.org.au/guides/creating-a-data-management-framework>.
- Anhalt-Depies, Christine, Jennifer L. Stenglein, Benjamin Zuckerberg, Philip A. Townsend und Adena R. Rissman. 2019. „Tradeoffs and Tools for Data Quality, Privacy, Transparency, and Trust in Citizen Science.“ *Biological Conservation* 238 (Oktober): 108195. doi:10.1016/j.biocon.2019.108195.
- Antonopoulos, Nick, und Lee Gillam, Hg. 2017. *Cloud Computing: Principles, Systems and Applications. Computer Communications and Networks*. Cham: Springer International Publishing. doi:10.1007/978-3-319-54645-2.
- Anzt, Hartwig, Felix Bach, Stephan Druskat, Frank Löffler, Axel Loewe, Bernhard Y. Renard, Gunnar Seemann et al. 2020. „An Environment for Sustainable Research Software in Germany and beyond: Current State, Open Challenges, and Call for Action“. *F1000Research* 9 (April): 295. doi:10.12688/f1000research.23224.1.

98 Vgl. DZHK o.J.

- Arning, Ursula, Birte Lindstädt und Jasmin Schmitz. 2016. „PUBLISSO: Das Open-Access-Publikationsportal für die Lebenswissenschaften.“ *GMS Medizin – Bibliothek – Information* 16 (3): Doc15. doi:10.3205/mbi000370.
- Bach, Felix, Oliver Bertuch, Christian Buss, Wolfgang zu Castell, Sabine Celo, Michael Denker, Stefan Dinkelacker et al. 2019. *Muster-Richtlinie Nachhaltige Forschungssoftware an den Helmholtz-Zentren*. Potsdam: Helmholtz Open Science Office. doi:10.2312/os.helmholtz.007.
- Bauer, Christoph, Frank Eickmeier und Michael Eckard. 2018. *E-Health: Datenschutz und Datensicherheit*. Wiesbaden: Springer Fachmedien. doi:10.1007/978-3-658-15091-4.
- Bender, Theresa, Christian R Bauer, Marcel Parciak, Robert Lodahl und Ulrich Sax. 2018. „FAIR Conform ETL Processing in Translational Research.“ In: *63. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMD5)*, DocAbstr. 254. Düsseldorf: German Medical Science GMS Publishing House. doi:10.3205/18gmds095.
- Candela, Leonardo, Donatella Castelli und Pasquale Pagano. 2013. „Virtual Research Environments: An Overview and a Research Agenda.“ *Data Science Journal* 12: GRDI75–81. doi:10.2481/dsj.GRDI-013.
- Chevrier, Raphaël, Vasiliki Foufi, Christophe Gaudet-Blavignac, Arnaud Robert und Christian Lovis. 2019. „Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review.“ *Journal of Medical Internet Research* 21 (5): e13484. doi:10.2196/13484.
- Crouch, Stephen, Neil Chue Hong, Simon Hettrick, Mike Jackson, Aleksandra Pawlik, Shoaib Sufi, Les Carr, David De Roure, Carole Goble und Mark Parsons. 2013. „The Software Sustainability Institute: Changing Research Software Attitudes and Practices.“ *Computing in Science Engineering* 15 (6): 74–80. doi:10.1109/MCSE.2013.133.
- Curd, Constanze. 2014. *TR32DB Metadata Schema for the Description of Research Data in the TR32DB*. O.O: o. V. doi:10.5880/TR32DB.10.
- Curd, Constanze. 2016. „Metadata Management in an Interdisciplinary, Project-Specific Data Repository: A Case Study from Earth Sciences.“ In: *Metadata and Semantics Research*, hg. v. Emmanuel Garoufallou, Imma Subirats Coll, Armando Stellato und Jane Greenberg, 357–368. (Communications in Computer and Information Science.) Cham: Springer International. doi:10.1007/978-3-319-49157-8_31.
- Curd, Constanze und Dirk Hoffmeister. 2015. „Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database.“ *Program: electronic library and information systems* 49 (4): 494–512. doi:10.1108/PROG-02-2015-0016.
- Curd, Constanze, Dirk Hoffmeister, Tanja Kramm, Ulrich Lang und Georg Bareth. 2019. „Etablierung von Forschungsdatenmanagement-Services in geowissenschaftlichen Sonderforschungsbereichen am Beispiel des SFB/Transregio 32, SFB 1211 und SFB/ Transregio 228.“ *Bausteine Forschungsdatenmanagement* 2 (September): 61–67. doi:10.17192/bfdm.2019.2.8103.
- DATAJUS. o. J. „Forschungsprojekt DataJus.“ Document. TU Dresden. <https://tu-dresden.de/gsw/jura/igetem/jfbimd13/forschung/forschungsprojekt-datajus>.
- David, Cédric H., Yolanda Gil, Christopher J. Duffy, Scott D. Peckham und S. Karan Venayagamoorthy. 2017. „An introduction to the special issue on Geoscience Papers of the Future.“ *Earth and Space Science* (Dezember): 441–444. doi:10.1002/2016EA000201.
- DFG. 2019. *Leitlinien zur Sicherung guter wissenschaftlicher Praxis – Kodex*. https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf.
- DSGVO. o. J. „Datenschutz-Grundverordnung (DSGVO) Art. 4.“ <https://dejure.org/gesetze/DSGVO/4.html>.

- Dyke, Stephanie O. M. 2020. „Chapter 2 – Genomic Data Access Policy Models.“ In: *Responsible Genomic Data Sharing*, hg. v. Xiaoqian Jiang und Haixu Tang, 19–32. London: Academic Press. doi:10.1016/B978-0-12-816197-5.00002-4.
- DZHK. 2014. „Verfahrensbeschreibung und Datenschutzkonzept des Zentralen Datenmanagements des DZHK.“ https://dzhk.de/fileadmin/user_upload/Datenschutzkonzept_des_DZHK.pdf.
- Eaton, India und Molly McNett. 2020. „Chapter Six – Protecting the Data: Security and Privacy.“ In: *Data for Nurses*, hg. v. Molly McNett, 87–99. London: Academic Press. doi:10.1016/B978-0-12-816543-0.00006-6.
- EGA. o. J. „home | European Genome-phenome Archive.“ <https://www.ebi.ac.uk/ega/home>.
- Eitzel, M. V., Jessica L. Cappadonna, Chris Santos-Lang, Ruth Ellen Duerr, Arika Virapongse, Sarah Elizabeth West, Christopher Conrad Maximillian Kyba et al. 2017. „Citizen Science Terminology Matters: Exploring Key Terms.“ *Citizen Science: Theory and Practice* 2 (1): 1. doi:10.5334/cstp.96.
- Engelhardt, Claudia. 2013. „Forschungsdatenmanagement in DFG-SFBs.“ *LIBREAS Library Ideas* 23. doi:10.18452/9045.
- Engelhardt, Claudia. 2020. „Forschungsdatenmanagement in DFG-Sonderforschungsbereichen“: *Bausteine Forschungsdatenmanagement* 1 (April): 16–27. doi:10.17192/bfdm.2020.1.8157.
- Europäische Kommission. o. J. „Sensible Daten.“ EU-Kommission – European Commission. https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_de.
- EUROSTAT. 2017. „Micro-Moments Dataset linked micro-aggregated data on ICT usage, innovation and economic performance in enterprises.“ <https://ec.europa.eu/eurostat/documents/203647/6867168/Safe+centre+rules+for+MMD.pdf/>.
- Fleischer, Dirk. 2020. „Was vom Tage übrigblieb‘ – ‚SFB-Legacy‘ von INF-Teilprojekten.“ *Bausteine Forschungsdatenmanagement* 1 (April): 39–44. doi:10.17192/bfdm.2020.1.8090.
- forschungsdaten.info. o. J. „Rechte und Pflichten | Themen | Forschungsdaten und Forschungsdatenmanagement.“ <https://www.forschungsdaten.info/themen/rechte-und-pflichten/>.
- forschungsdaten.org. o. J. „Elektronische Laborbücher.“ https://www.forschungsdaten.org/index.php/Elektronische_Laborbuecher.
- forschungslizenzen.de. o. J. „Forschungslizenzen.“ Forschungslizenzen. <http://forschungslizenzen.de/>.
- Freire, Juliana, Norbert Fuhr und Andreas Rauber. 2016. *Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041)*. Dagstuhl: Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/DAGREP.6.1.108.
- FZI. o. J. „Zusammenarbeiten: FZI Forschungszentrum Informatik.“ <https://www.fzi.de/wir-fuer-sie/zusammenarbeiten/>.
- GDD. o. J. „GDD e. V.“ <https://www.gdd.de/>.
- gesis.org. o. J. „GESIS – Leibniz-Institut für Sozialwissenschaften.“ <https://www.gesis.org/angebot/daten-analysieren/secure-data-center-sdc>.
- Gkoulalas-Divanis, Aris und Grigorios Loukides. 2015. *Medical Data Privacy Handbook*. Cham: Springer International. doi:10.1007/978-3-319-23633-9.
- Hartmann, Thomas. 2019. „Rechtsfragen: Institutioneller Rahmen und Handlungsoptionen für universitäres FDM.“ *Zenodo*. doi:10.5281/zenodo.2654306.
- Hothorn, Torsten, Leonhard Held und Tim Friede. 2009. „Biometrical Journal and Reproducible Research.“ *Biometrical Journal* 51 (4): 553–555. doi:10.1002/bimj.200900154.
- Hothorn, Torsten und Friedrich Leisch. 2011. „Case Studies in Reproducibility.“ *Briefings in Bioinformatics* 12 (3): 288–300. doi:10.1093/bib/bbq084.
- it-agile.de. o. J. „Agiles Arbeiten in verteilten Teams.“ <https://www.it-agile.de/wissen/agile-teams/agiles-arbeiten-in-verteiltten-teams/>.

- Kalberg, Nadine. 2014. *Datenschutz an Hochschulen: eine Analyse der Rechtsgrundlagen und ihrer Umsetzung in integriertem Informationsmanagement und Forschung*. Schriften zum Informations-, Telekommunikations- und Medienrecht 46. Münster: LIT.
- Kogut, Bruce und Anca Metiu. 2001. „Open-Source Software Development and Distributed Innovation.“ *Oxford Review of Economic Policy* 17 (2): 248–264. doi:10.1093/oxrep/17.2.248.
- Kommission Zukunft der Informationsinfrastruktur. 2011. *Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder*. Berlin: Gemeinsame Wissenschaftskonferenz. https://www.hof.uni-halle.de/web/dateien/KII_Gesamtkonzept_2011.pdf.
- Kreutzer, Till und Henning Lahmann. 2019. *Rechtsfragen bei Open Science: Ein Leitfaden*. Hamburg: Hamburg University Press. doi:10.15460/HUP.195.
- Krogh, Georg von und Eric von Hippel. 2003. „Special Issue on Open Source Software Development.“ *Research Policy, Open Source Software Development* 32 (7): 1149–1157. doi:10.1016/S0048-7333(03)00054-4.
- Lauber-Rönsberg, Anne. 2018. „Data Protection Laws, Research Ethics and Social Sciences.“ In: *Research Ethics in the Digital Age: Ethics for the Social Sciences and Humanities in Times of Mediatization and Digitization*, hg. v. Farina Madita Dobrick, Jana Fischer und Lutz M. Hagen, 29–44. Wiesbaden: Springer Fachmedien. doi:10.1007/978-3-658-12909-5_4.
- Lechler, Thomas G. und Siwen Yang. 2017. „Exploring the Role of Project Management in the Development of the Academic Agile Software Discourse: A Bibliometric Analysis.“ *Project Management Journal* 48 (1): 3–18. doi:10.1177/875697281704800101.
- Nurmi, Sanna-Maria, Mari Kangasniemi, Arja Halkoaho und Anna-Maija Pietilä. 2019. „Privacy of Clinical Research Subjects: An Integrative Literature Review.“ *Journal of Empirical Research on Human Research Ethics* 14 (1): 33–48. doi:10.1177/1556264618805643.
- Oliveira, Ary H. M. de, Daniel de Oliveira und Marta Mattoso. 2017. „Clouds and Reproducibility: A Way to Go to Scientific Experiments?“ In *Cloud Computing*, hg. von Nick Antonopoulos und Lee Gillam, 127–151. (Computer Communications and Networks.) Cham: Springer International. doi:10.1007/978-3-319-54645-2_5.
- open-access.net. o. J. „Informationsplattform Open Access: Lizenzen.“ <https://open-access.net/informationen-zu-open-access/rechtsfragen/rechtsfragen-in-deutschland/lizenzen>.
- Ostendorff, Philipp und David Linke. 2019. „Best-Practices im Umgang mit rechtlichen Fragestellungen zum Forschungsdatenmanagement (FDM).“ *Bibliotheksdienst* 53 (10–11): 717–723. doi:10.1515/bd-2019-0098.
- Patel, Dimple. 2016. „Research data management: a conceptual framework.“ *Library Review* 65 (4/5): 226–241. doi:10.1108/LR-01-2016-0001.
- Pommerening, Klaus, Johannes Drepper, Krister Helbing und Thomas Ganslandt. 2014. *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0*. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft. doi:10.32745/9783954662951.
- radar-service.eu. o. J. „Lizenzen für Forschungsdaten | RADAR – Ein Repository für die Wissenschaft.“ <https://www.radar-service.eu/de/lizenzen-fuer-forschungsdaten>.
- Repschläger, Jonas, Danny Pannicke und Rüdiger Zarnekow. 2010. „Cloud Computing: Definitionen, Geschäftsmodelle und Entwicklungspotenziale.“ *HMD Praxis der Wirtschaftsinformatik* 47 (5): 6–15. doi:10.1007/BF03340507.
- Rfll. 2016. *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen: Rat für Informationsinfrastrukturen. <http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>.

- Rfil. 2019. *Herausforderung Datenqualität. Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel*. Zweite Auflage. Göttingen: Rat für Informationsinfrastrukturen. <http://www.rfii.de/?p=4043>.
- Rfil. 2020. *Rfil-Stellungnahme „Datentreuhandstellen gestalten – zu Erfahrungen der Wissenschaft“ April 2020*. Göttingen: Rat für Informationsinfrastrukturen. <http://www.rfii.de/download/rfiistellungnahme-zu-datentreuhandstellen/>.
- Roertgen, Steffen, Harald Kusch, Claudia Engelhardt, Sven Bingert, Valeria Savin, Inga Kraus, Ortrun Brand et al. 2019. „Posters presented at ‚Workshop zu Forschungsdatenmanagement und -infrastruktur in DFG-Sonderforschungsbereichen‘, 26./27. November 2018 in Göttingen.“ *Göttingen Research Online*. <https://data.gro.uni-goettingen.de/citation?persisten-tid=doi:10.25625/22Y1WC>.
- Röttgen, Charlotte. 2020. „Rechtspositionen an Daten: Die Rechtslage im europäischen Rechtsraum.“ In: *Datenrecht in der Digitalisierung*, hg. v. Louisa Specht-Riemenschneider, Nikola Werry und Susanne Werry, 371–407. Berlin: Erich Schmidt Verlag.
- Rowbotham, Samantha, Merryn McKinnon, Joan Leach, Rod Lamberts und Penelope Hawe. 2019. „Does citizen science have the capacity to transform population health science?“ *Critical Public Health* 29 (1): 118–128. doi:10.1080/09581596.2017.1395393.
- Schallaböck, Jan und Max von Grafenstein. 2017. *ORCID aus datenschutzrechtlicher Sicht – ‚Gutachten im Auftrag des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projektes ORCID DE zur Förderung der Open Researcher and Contributor ID in Deutschland‘*. https://gfgz-public.gfz-potsdam.de/rest/items/item_2263903_5/component/file_2265895/content.
- Schapke, Sven-Eric, Jakob Beetz, Markus König, Christian Koch und André Borrmann. 2018. „Collaborative Data Management.“ In: *Building Information Modeling*, hg. v. André Borrmann, Markus König, Christian Koch, und Jakob Beetz, 251–277. Cham: Springer International Publishing. doi:10.1007/978-3-319-92862-3_14.
- Schefer, Maya. 2012. „Der Wissensbegriff am Limit? – Kollaborative Wissensgenerierung im Grossforschungsprojekt ATLAS am CERN.“ In: *MetaATLAS. Studien zur Generierung, Validierung und Kommunikation von Wissen in einer modernen Forschungskollaboration*, hg. v. Gerd Graßhoff und Adrian Wüthrich, 83–108. Bern: Bern Studies in the History and Philosophy of Science.
- Schmidt, Jan-Hinrik und Thilo Weichert, Hg. 2012. *Datenschutz: Grundlagen, Entwicklungen und Kontroversen*. Schriftenreihe, Bd. 1190. Bonn: Bpb, Bundeszentrale für politische Bildung. https://www.bpb.de/system/files/dokument_pdf/1190-Datenschutz-X3.pdf.
- Shah, Nisha, Victoria Coathup, Harriet Teare, Ian Forgie, Giuseppe Nicola Giordano, Tue Haldor Hansen, Lenka Groeneveld et al. 2019. „Sharing Data for Future Research – Engaging Participants’ Views about Data Governance beyond the Original Project: A DIRECT Study.“ *Genetics in Medicine* 21 (5): 1131–1138. doi:10.1038/s41436-018-0299-7.
- Simons, Franz. 2020. *A Robust High-Resolution Hydrodynamic Numerical Model for Surface Water Flow and Transport Processes within a Flexible Software Framework*. doi:10.14279/depositon-ce-9589.
- SNF. o. J. „Was versteht der SNF unter kollaborativer Forschung?“ o. J. <http://www.snf.ch/de/fokus-Forschung/faq/Seiten/faq-foerderinstrument-sinergia-reform-kollaborative-forschung.aspx>.
- Solove, Daniel J. 2006. „A Taxonomy of Privacy.“ *University of Pennsylvania Law Review* 154 (3): 477–564. doi:10.2307/40041279.
- Stegemann, Jessica. 2020. „Was bleibt nach dem Projekt? – Nachhaltigkeitsstrategien für das Forschungsdatenmanagement (FDM) entwickeln.“ *Bausteine Forschungsdatenmanagement* 1 (April): 69–76. doi:10.17192/bfdm.2020.1.8167.
- Stiles, Paul G. und John Pettila. 2011. „Research and confidentiality: Legal issues and risk management strategies.“ *Psychology, Public Policy, and Law* 17 (3): 333–356. doi:10.1037/a0022507.

- Suhr, Markus, Christoph Lehmann, Christian Robert Bauer, Theresa Bender, Cornelius Knopp, Luca Freckmann, Björn Öst Hansen et al. 2020. „menoci: Lightweight Extensible Web Portal enabling FAIR Data Management for Biomedical Research Projects.“ *arXiv* (Februar): 2002.06161 [cs]. <http://arxiv.org/abs/2002.06161>.
- talend.com. o. J. „Data Preparation – kostenloses Datenaufbereitungstool von Talend.“ Talend Real-Time Open Source Data Integration Software. <https://de.talend.com/products/data-preparation/data-preparation-free-desktop/>.
- Torcivia-Rodríguez, John, Hayley Dingerdissen, Ting-Chia Chang und Raja Mazumder. 2019. „A Primer for Access to Repositories of Cancer-Related Genomic Big Data.“ In *Cancer Bioinformatics*, hg. v. Alexander Krasnitz, 1–37. (Methods in Molecular Biology.) New York, NY: Springer. doi:10.1007/978-1-4939-8868-6_1.
- UK Data Service. o. J. a „Data resources library.“ <https://www.ukdataservice.ac.uk/manage-data/collaboration/resources-library.aspx>.
- UK Data Service. o. J. b „File sharing.“ <https://www.ukdataservice.ac.uk/manage-data/collaboration/file-sharing.aspx>.
- UK Data Service. o. J. c „Standard procedures, protocols and policies.“ <https://www.ukdataservice.ac.uk/manage-data/collaboration/coordinate.aspx>.
- Venters, Colin C., Rafael Capilla, Stefanie Betz, Birgit Penzenstadler, Tom Crick, Steve Crouch, Elisa Yumi Nakagawa, Christoph Becker und Carlos Carrillo. 2018. „Software Sustainability: Research and Practice from a Software Architecture Viewpoint.“ *Journal of Systems and Software* 138 (April): 174–88. doi:10.1016/j.jss.2017.12.026.
- Warkentin, Nils. 2019. „Kollaboratives Arbeiten: Tipps für Teams.“ *karrierebibel.de*. 21. Juni 2019. <https://karrierebibel.de/kollaboratives-arbeiten/>.
- Willmes, Christian, Daniel Kürner und Georg Bareth. 2014. „Building Research Data Management Infrastructure Using Open Source Software.“ *Transactions in GIS* 18 (4): 496–509. doi:10.1111/tgis.12060.

