

Empfehlungen für Data Stewardship an akademischen Forschungsinstitutionen

Ergebnisse des
Projektes DataStew

Seidlmayer

Hoffmann

Dierkes

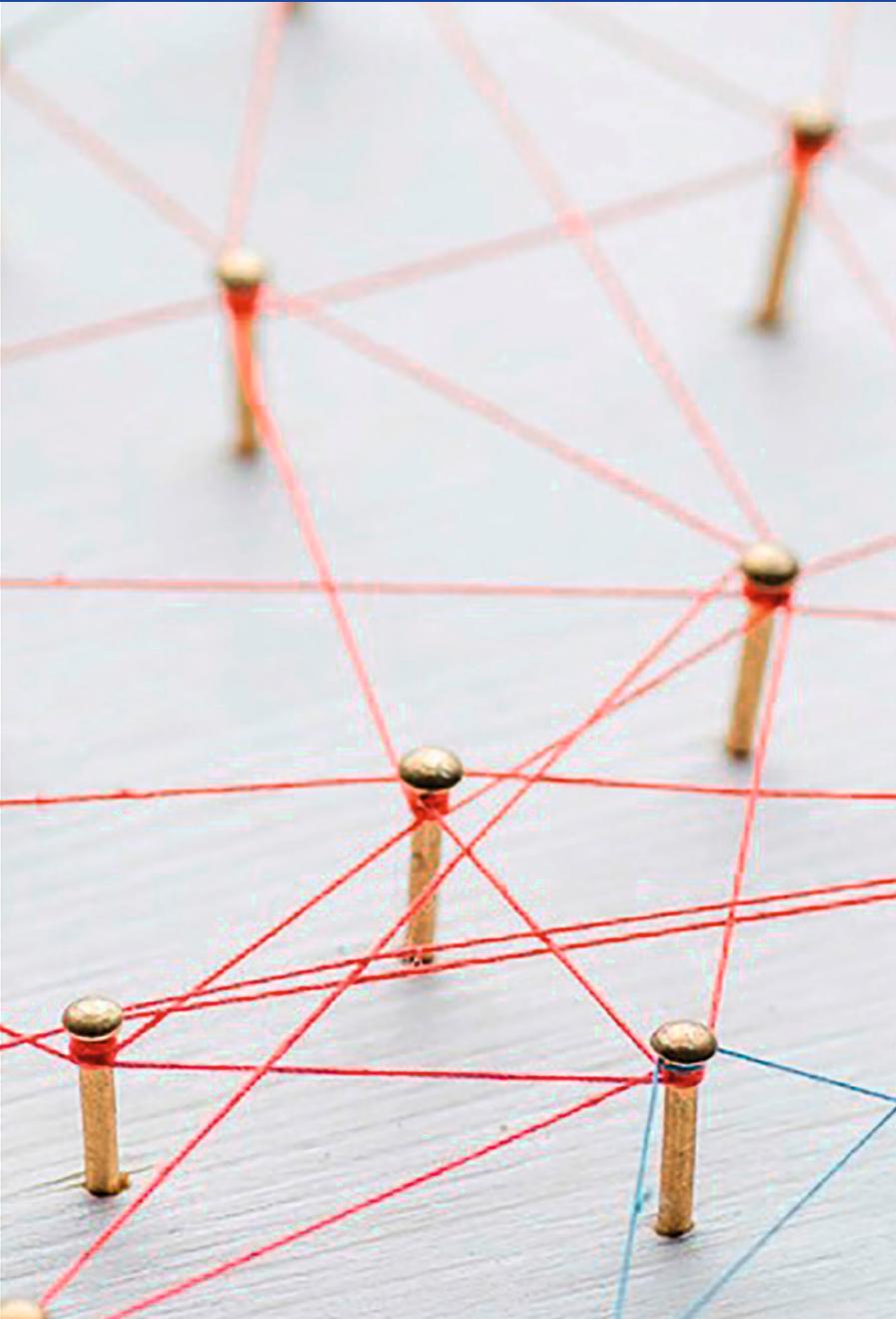
Lindstädt

Depping

Förstner

Band 91

b  i  t  verlag





b i t verlag

Band 91

b.i.t.online innovativ

2023

b.i.t.verlag gmbh, Wiesbaden

Empfehlungen für Data Stewardship an akademischen Forschungsinstitutionen

Ergebnisse des Projektes DataStew

Eva Seidlmayer, ZB MED – Informationszentrum Lebenswissenschaften

Fabian Hoffmann, Universitäts- und Stadtbibliothek Köln, Universität zu Köln

Jens Dierkes, Universitäts- und Stadtbibliothek Köln, Universität zu Köln

Birte Lindstädt, ZB MED – Informationszentrum Lebenswissenschaften

Ralf Depping, Universitäts- und Stadtbibliothek Köln, Universität zu Köln

Konrad U. Förstner, ZB MED – Informationszentrum Lebenswissenschaften

b.i.t.online innovativ

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

ISBN 978-3-9824425-8-7

978-3-9824425-8-7

ISSN 1615-1577

Neuaufgabe 2023

© b.i.t.verlag gmbh, Wiesbaden, 2023 Wiesbaden

Alle Rechte vorbehalten, insbesondere die des Nachdrucks und der Übersetzung.

Ohne Genehmigung des Verlages ist es nicht gestattet, dieses Werk oder Teile daraus in
einem fotomechanischen oder sonstigen Reproduktionsverfahren oder unter Verwendung
elektronischer Systeme zu verarbeiten, zu vervielfältigen und zu verbreiten.

Printed in Germany

A network diagram is shown on a light-colored wooden surface. It consists of several brass pins arranged in a grid-like pattern, connected by thin red strings. The strings form a complex web of interconnected lines, representing a network structure. The background is slightly blurred, focusing attention on the network.

Forschung unterstützen:
**Empfehlungen für Data Stewardship an
akademischen Forschungsinstitutionen**

Ergebnisse des Projektes DataStew
(Stand: 30.06.2023)

Projektbearbeitung und Berichtserstellung:

Eva Seidlmayer, ZB MED – Informationszentrum Lebenswissenschaften

Fabian Hoffmann, Universitäts- und Stadtbibliothek Köln, Universität zu Köln

Jens Dierkes, Universitäts- und Stadtbibliothek Köln, Universität zu Köln **Birte**

Lindstädt, ZB MED – Informationszentrum Lebenswissenschaften

Ralf Depping, Universitäts- und Stadtbibliothek Köln, Universität zu Köln

Konrad U. Förstner, ZB MED – Informationszentrum Lebenswissenschaften und Technische Hochschule Köln

**Technology
Arts Sciences
TH Köln**



DOI (Digitalausgabe): **Version 1:** DOI: 10.4126/FRL01-006441397

Lizenz:



Diese Arbeit ist lizenziert unter [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Von dieser Lizenz ausgenommen sind die verwendeten Logos.

Veröffentlichungsdatum: Version 1 am 28. April 2023
Version 2 am 30. Juni 2023

„In der Betreuung von Anwendungen bzw. Geräten sowie in der Forschungsdokumentation ist der Bedarf an spezialisierter Unterstützung in den letzten Jahren in dramatischem Umfang gewachsen. Nachhaltige personelle Lösungen fehlen jedoch [...]. Eine Prüfung, wo diese Aufgaben letztlich sinnvoll verortet werden, wie auch eine Professionalisierung der Aufgabenwahrnehmung sind aus Sicht des RfII dringend erforderlich.“
(RfII 2019: 23)



Gefördert durch das BMBF unter dem Kennzeichen 16DWWQP05A.

Inhaltsverzeichnis

Abkürzungsverzeichnis	7
o. Kurzfassung	8
1. Einleitung	9
1.1. Problembeschreibung	9
1.2. Das Projektvorhaben	11
1.3. Begriffsbestimmung und Abgrenzung zu ähnlichen Berufsbildern	12
2. Konzeptionelle Entwürfe Data Stewardship	16
2.1. Was sind die Aufgaben von DS?	16
2.2. Was sind die notwendigen Kompetenzen von Data Stewards?	25
2.3. Institutionelle Verankerung von Data Stewards	29
2.4. Schärfung der Forschungsfragen	31
3. Ausbildung von Data Stewards	33
3.1. Studiengänge und Weiterbildungen in Deutschland	34
3.2. Von anderen lernen: Ein Blick zu den europäischen Nachbarn	39
4. Bedarfe in der Praxis: Untersuchung von Stellenausschreibungen	42
4.1. In den Anzeigen genannte Aufgaben	44
4.2. In den Anzeigen geforderte Kenntnisse	48
4.3. In den Anzeigen gesuchte soziale Fähigkeiten	50
4.4. Vorläufige Antworten auf die Forschungsfragen	51
5. Die Expertise nutzen: Interviews mit Expert:innen	53
5.1. Aufgaben von Data Stewards	54
5.2. Institutionelle Verortung	54
5.3. Verständnis Data Stewardship: Wissenschaft oder Service?	56
5.4. Basiswissen von Data Stewards	58
5.5. Wann sind fachwissenschaftliche Kenntnisse obligatorisch?	60
5.6. Spezialisierung auf Fachwissenschaft	60
5.7. Vorläufige Antworten auf die Forschungsfragen	61

6. Fokusgruppendifkussionen	63
6.1. Vorbereitung: Thesen und Abstimmung	63
6.2. Fokusgruppendifkussionen	65
6.3. Nachbereitung	70
6.4. Vorläufige Antworten auf die Forschungsfragen	71
7. Profilbildung mit Kompetenzen und Ausbildung	72
7.1. Profil 1: Data Stewards als Generalist:innen	73
7.2. Profil 2: FDM-beratende Data Stewards	74
7.3. Profil 3: Disziplinär-betreuende Data Stewards	75
7.4. Profil 4: Data Stewards als Koordinator:innen	76
7.5. Profil 5: Informationsinfrastrukturnahe Data Stewards	77
7.6. Übersicht Kompetenzprofile	78
8. Antworten auf Forschungsfragen und Empfehlungen	80
8.1. Beantwortung der Forschungsfragen	80
8.2. Empfehlungen für Akteure im Data Stewardship	84
8.3. Empfehlungen für Forschungseinrichtungen	84
8.4. Empfehlungen für Politik und Fördereinrichtungen	87
8.5. Empfehlungen für Ausbildungseinrichtungen	88
Literaturverzeichnis	90
Webseitenverzeichnis	96
Anhang: DataStew - Interviewleitfaden	98
Fragen an Data Stewards	98
Interviewfragen an Forschende	100
Interviewfragen an Personen, die in der Politikberatung beschäftigt sind	102
Interviewfragen an Personen, die in der Aus- und Weiterbildung beschäftigt sind	104
Interviewfragen an Personen, die in der Infrastruktur beschäftigt sind	106

Abbildungsverzeichnis

Abbildung 1: Beispiel für einen Arbeitsbereich auf dem Mural Board.	66
Abbildung 2: Durchschnittlicher Zustimmungswert zu den umformulierten Thesen aus Tab. 7.	70
Abbildung 3: Profil Data Stewards als Generalist:innen.	73
Abbildung 4: Profil FDM-beratende Data Stewards.	74
Abbildung 5: Profil disziplinar-betreuende Data Stewards.	75
Abbildung 6: Profil Data Stewards als Koordinator:innen.	76
Abbildung 7: Profil Informationsinfrastrukturnahe Data Stewards.	77
Abbildung 8: Aufgabenbereiche von Data Stewards.	81
Abbildung 9: Mögliche Umsetzungen von Data Stewardship an einer Institution.	82
Abbildung 10: Basiswissen im Data Stewardship.	83
Abbildung 11: Notwendige disziplinäre Kenntnisse im Data Stewardship.	83
Abbildung 12: Entscheidungsbaum für die Umsetzung von Data Stewardship in dem jeweiligen Kontext.	86
Abbildung 13: Beispiel für die Umsetzung von Data Stewardship an einer großen Einrichtung (Modell G).	87

Tabellenverzeichnis

Tabelle 1: Rollen, Tätigkeitsbereiche, Aufgaben von Data Stewards.	22
Tabelle 2: Kompetenzstufen von Data Stewards.	27
Tabelle 3: Kompetenzen von Data Stewards nach Gruber et al. (2021: 23–26).	27
Tabelle 4: Kompetenzen von Data Stewards.	28
Tabelle 5: Übersicht von Ausbildungs- und Weiterbildungsangeboten im Kontext von Data Stewardship.	35
Tabelle 6: Stellenbezeichnungen, Zahl der Treffer und nötige Abschlüsse.	43
Tabelle 7: Themenbereiche und Ausgangsversion der Thesen zum DSS.	64
Tabelle 8: Überarbeitete Version der kontrovers bewerten Thesen zum DSS.	69
Tabelle 9: Fachkompetenzen, die in den Stellenanzeigen gefordert waren.	78

Abkürzungsverzeichnis

CARE	Prinzipien zur Sicherstellung von Collective benefit, Authority to control, Responsibility, Ethics (Schulze 2020)
DFG	Deutsche Forschungsgemeinschaft
DS	Data Steward
DSS	Data Stewardship
EOSC	European Open Science Cloud
FAIR	Prinzipien zur Ermöglichung von Findability, Accessibility, Interoperability, Reusability von Forschungsdaten
FID	Fachinformationsdienst
HRK	Hochschulrektorenkonferenz
NFDI	Nationale Forschungsdateninfrastruktur
RfII	Rat für Informationsinfrastruktur
SFB	Sonderforschungsbereich
SHG	Stakeholdergruppe

0. Kurzfassung

Im Sinne von Open Science sollen Forschungsdaten an Hochschulen und Forschungseinrichtungen nachvollziehbar und nachnutzbar abgelegt werden. In den letzten Jahren hat sich für diese Aufgaben ein neues Berufsfeld etabliert, das häufig Data Stewardship genannt wird. Was genau jedoch die Aufgaben und Kompetenzen von Data Stewards sind und wie sich das Data Stewardship an den jeweiligen Einrichtungen idealerweise umsetzen lässt, ist nicht geklärt.

Der vorliegende Bericht dokumentiert die Untersuchung des Verbundprojekts „DataStew“ von Universitäts- und Stadtbibliothek Köln und ZB MED – Informationszentrum Lebenswissenschaften. In diesem Projekt wurde die derzeitige akademische Forschungslandschaft mit ihren unterschiedlichen Stakeholdern zum Konzept Data Stewardship untersucht. Das Vorhaben umfasst sowohl die Auswertung des vorliegenden Wissensstandes anhand von Literatur zum Thema und die Analyse von realen Stellenausschreibungen als auch qualitative Interviews sowie Gruppendiskussionen mit Expert:innen. Die Ergebnisse der Analysen, Interviews und Gespräche stellen die Grundlage der in diesem Bericht vorgestellten Handlungsempfehlungen dar.

Zentrales Ergebnis der Studie ist, dass eine allgemeingültige Realisierung von Data Stewardship sowie der damit verbundenen Aspekte nicht formuliert werden kann und es stark von den jeweiligen lokalen Bedingungen einer Einrichtung abhängt, innerhalb derer es umgesetzt wird. In der Untersuchung wird daher ein Perspektivwechsel vorgeschlagen, der anhand der Kapazitäten und Bedarfe der Einrichtung unterschiedliche Modelle von Data Stewardship beschreibt. In einem Baukastensystem, für das im Projektbericht Instrumente bereitgestellt werden, können spezifische Teams von Data Stewards zusammengestellt werden.

Neben konkreten Vorschlägen und Entscheidungshilfen für Einrichtungen, die Data Stewardship umsetzen wollen, werden im letzten Teil des Ergebnisberichts auch Handlungsempfehlungen für politische Entscheidungsträger und Förderinstitutionen sowie für Träger von Ausbildungen zur Qualifizierung von Data Stewards formuliert.

1. Einleitung

1.1. Problembeschreibung

Wissenschaftliches Arbeiten hat sich gegenüber den letzten Jahrzehnten stark verändert. Viele IT-gestützte Hilfsmittel sind entwickelt worden, gleichzeitig setzt sich zunehmend die Idee einer offenen Wissenschaft mit allgemein zugänglichen und nachnutzbaren Forschungsergebnissen durch. Dies gilt nicht nur für Literatur, sondern auch für die Publikation von offenen Daten (Rothfritz et al. 2021: 262) sowie weiteren Forschungsprodukten wie Software und Modellen. Die Veröffentlichung im Sinne der FAIR-Prinzipien wird dabei zunehmend auch von Verlagen und der Forschungsförderung eingefordert und zur Bedingung der Förderung gemacht (Schmitz 2021). Die dafür notwendige zentrale öffentliche Infrastruktur hat lange Zeit gefehlt, wird nun aber nachhaltig aufgebaut, etwa in der Nationalen Forschungsdateninfrastruktur (NFDI) auf der deutschen oder in der European Open Science Cloud (EOSC) (Rothfritz 2019: 14, Wittenburg & Beck 2021: 26) auf europäischer Ebene. Die neuen Werkzeuge erlauben dabei einerseits die Erstellung immer größerer Mengen von Daten und zugehöriger Publikationen, erfordern aber andererseits Kenntnisse in der Handhabung und Beschreibung der erzeugten Informationen sowie ihrer dauerhaften Speicherung.

Um den Paradigmenwechsel zu offener Wissenschaft und guter wissenschaftlicher Praxis angemessen umsetzen zu können, sind unter anderem Kenntnisse über vorhandene Publikationsplattformen, Repositorien und Urheberrechte notwendig. Ein professioneller und zeitgemäßer Umgang mit den Forschungsergebnissen bedarf darüber hinaus nicht nur spezifischer Kenntnisse zu Tools und Open Science, sondern ist auch zeitintensiv. Dass trotz der zur Unterstützung des Datenmanagements zunehmend aufkommenden Tools (z.B. „Data Stewardship Wizard“, „Research Data Management Organizer“ (RDMO), „Data Steward Tool“) (Dierkes 2021: 311, Wegner et al. 2022), diese Aufgaben alleine durch elektronische Werkzeuge und Programme übernommen werden können, erscheint wegen des komplexen, nicht linearen Forschungsprozesses und der hohen Wandlungsgeschwindigkeit der Bedarfe und politischen Vorgaben derzeit nicht realistisch.

Data Stewards: Daher werden Wissenschaftler:innen im Forschungsdatenmanagement zunehmend von sogenannten Data Stewards (DS) unterstützt. Die Übernahme des Data Stewardship (DSS) durch spezialisiertes Personal ist im Sinne des effizienten Zeiteinsatzes der Wissenschaftler:innen sowie der Optimierung der Datenqualität und damit der Nachnutzbarkeit der Daten auch finanziell ein „exzellentes Investment“ (Mons 2020). Die Zusammenarbeit zwischen Wissenschaftler:innen, die ihre Daten am besten kennen, und den DS ist besonders wichtig, nicht zuletzt, da Wissenschaftler:innen häufig kurzfristige Verträge haben und oft nur für die Dauer eines Projektes an einer Einrichtung bleiben (Borgmann 2018: 385; Kvale 2022). Während die Forschenden sich durch hohe örtliche Mobilität auszeichnen, bleiben die Daten meist an den entsprechenden Einrichtungen oder zentralen Repositorien und müssen entsprechend der FAIR-Prinzipien dauerhaft gepflegt werden.

Ausbildungsgänge: Ausgebildete Fachkräfte in ausreichender Zahl, um die skizzierten Aufgaben wahrnehmen zu können, fehlen jedoch. Barend Mons schätzt beispielsweise, dass allein für die EOSC in den nächsten Jahren rund 500 000 DS gebraucht werden, wenn man je 20 Datenerzeugende eine:n DS annimmt (Mons 2016). Ebenso fehlen laut Vertreter:innen der NFDI bisher auch genügend Ausbildungsgänge (Neuroth & Oevel 2021: 552–553) mit entsprechenden Spezialisierungen, die diese Personen grundständig oder mittels einer Weiterbildung qualifizieren könnten. Auch der Rat für Informationsinfrastruktur (RfII) sah einen „dramatischen Bedarf“, für den bisher „professionelles“ Personal fehlte:

In der Betreuung von Anwendungen bzw. Geräten sowie in der Forschungsdokumentation ist der Bedarf an spezialisierter Unterstützung in den letzten Jahren in dramatischem Umfang gewachsen. Nachhaltige personelle Lösungen fehlen jedoch [...]. Eine Prüfung, wo diese Aufgaben letztlich sinnvoll verortet werden, wie auch eine Professionalisierung der Aufgabenwahrnehmung, sind aus Sicht des RfII dringend erforderlich.
(RfII 2019: 23)

Der Mangel an Ausbildungsgängen, die den enormen Bedarf an Aus- und Weiterbildung für DS – nicht nur in Deutschland – befriedigen könnten, wird vom RfII ebenfalls beklagt (RfII 2019: 27). Zwar haben sich in den letzten Jahren hierzulande einige generische Ausbildungsinitiativen, namentlich an der TH Köln gemeinsam mit fdm.NRW und ZB MED, der Universität Bremen, der Humboldt-Universität zu Berlin und der Fachhochschule Potsdam, sowie disziplinäre Ausbildungsinitiativen ergeben (vgl. Kapitel 3 „Studiengänge“), doch sind diese noch nicht aufeinander abgestimmt, wie es von RfII und der Hochschulrektorenkonferenz HRK empfohlen wird (RfII 2019: 27, Meyer-Doerpinghaus/Neuroth 2015). Eine solche Abstimmung findet momentan innerhalb der Sektion „Training & Education“ der NFDI statt (Herres-Pawlis et al. 2022).

Begriffsverwendung: Der Umstand, dass das DSS, seine Aufgaben, die Ausbildung, Zielgruppen und Kompetenzen so schwierig zu fassen sind, spiegelt sich auch in der uneinheitlichen Begriffsbildung wider, die – wie sich im Verlauf der Untersuchung immer wieder zeigen wird – die Auseinandersetzung erschwert. In der englischsprachigen Forschung hat sich der Begriff des „Data Stewardship“ durchgesetzt (Curdt et al. 2021: 71–72) – nicht zuletzt durch die initiale Publikation „The FAIR Guiding Principles for scientific data management and stewardship“, die die FAIR-Prinzipien als erste prägnant formulierte (Wilkinson et al. 2016). Auch in den international agierenden Netzwerken GO FAIR (Pasquale et al. 2019; Helling et al. 2022) und der Research Data Alliance (RDA; z.B. Ayres et al. 2022) wird der Begriff für Aufgaben des Datenmanagements und des Supports des Datenlebenszyklus verwendet.

¹ Landesinitiative Forschungsdatenmanagement des Landes Nordrhein-Westfalen, <https://www.fdm.nrw>
[Für alle Weblinks in diesem Bericht gilt das Aufrufdatum 02.02.2023.]

Da sich die Verwendung des Begriffs DSS auch im deutschen Sprachraum häuft (Rothfritz 2019: 10, Böhme 2022: 2), wird diese Untersuchung die Bezeichnung ebenfalls verwenden. *Data Stewardship* (DSS) ist das übergreifende Konzept der langfristigen und nachhaltigen Sorge um Forschungsdaten. *Data Stewards* (DS) sind demnach Personen mit Tätigkeitsprofilen, die bestimmte Aufgaben des FDM abdecken und eine spezifische institutionelle Verortung aufweisen.

1.2. *Das Projektvorhaben*

Die Auseinandersetzung mit dem Komplex DSS ist nicht neu. Es gibt bereits vielfältige Forschung zu diesem Thema. Häufig bleiben die Analysen und Initiativen dabei in einem regional abgegrenzten oder institutionellen Rahmen. Damit wird eine wertvolle Grundlage gelegt. Konsolidierungsarbeit ist notwendig, um eine einheitliche Begriffsbildung sowie eine einheitliche Ausbildung und Vernetzung voranzubringen. Das Forschungsprojekt DataStew soll zu dieser Konsolidierung über die lokalen und institutionellen Eigenheiten hinweg beitragen.

Der vorliegende Projektbericht dokumentiert das Projekt DataStew, das innerhalb von zehn Monaten von Februar bis November 2022 durch die USB Köln und ZB MED bearbeitet wurde. Um diesem komplexen Feld in der begrenzten Projektlaufzeit genügend Raum geben zu können, beschränkt sich DataStew auf DSS im wissenschaftlichen Umfeld. Eine Untersuchung von DSS in der Privatwirtschaft muss an anderer Stelle geleistet werden. Mit den Bedarfen und Anforderungen in Unternehmen und der Wirtschaft hat sich beispielsweise eine Studie des Business Application Research Center beschäftigt und dabei 200 Unternehmen aus Deutschland, Österreich und der Schweiz befragt (BARC 2018). Einen anderen Weg geht die Studie „Final report to the National Forum for Research Data Management“ (Wildgaard et al. 2020); sie bezieht neben akademischen auch die nicht-akademischen Einrichtungen ein.

In DataStew werden aktuelle Erkenntnisse aus sechs relevanten Themenfeldern (vgl. Abschnitt 2.4) des DSS analysiert und in einem iterativen Prozess ausgebaut. Die fünf methodischen Schritte beginnen mit einer *Literaturrecherche* und beziehen dann eine Analyse von Stellenanzeigen ein. Ziel der Literaturrecherche war es, durch bereits in der Forschung vorliegende Systematisierungsversuche unterschiedliche Aspekte des DSS weiter einzugrenzen und die in der Projektskizze formulierten Leitfragen als Forschungsfragen zu schärfen. Die Leitfragen, die zum Ausgangspunkt für die Analyse werden, sind:

- 1. Was sind die Funktionen/Aufgaben eines:r DS?**
- 2. Was sind die notwendige Ausbildung bzw. die Kompetenzen von DS?**
- 3. Wo sind DS idealtypischerweise institutionell verankert (z.B. SFB/ Exzellenzcluster, NFDI-Konsortium, FDM-Stelle einer Hochschule)?
Zusätzlich zu diesen thematischen Feldern werden im Anschluss die bestehenden Studiengänge und Weiterbildungen in Deutschland und im Ausland vorgestellt.**

Durch die anschließende *Analyse der Stellenanzeigen* wurden die tatsächlichen Bedarfe der Institutionen und die Anforderungen an DS und verwandte Berufe noch deutlicher. Die hier gewonnenen Erkenntnisse dienten dann zur Erstellung eines qualitativen Interview-Leitfadens. Anhand dieses Leitfadens wurden neun *Interviews mit Expert:innen* der Interessengruppen Wissenschaftler:innen, Politikberater:innen, Mitarbeitende der Infrastruktur, DS sowie Ausbilder:innen durchgeführt. Auf Basis der durch die neun Expert:innen-Interviews aktualisierten Zwischenergebnisse konnten 22 Thesen zum DSS erstellt werden. Diese Thesen, die von Expert:innen in einer Umfrage kontrovers bewertet wurden, wurden anschließend in zwei Fokusgruppendifkussionen mit Repräsentanten der Interessensgruppen diskutiert.

Im Folgenden erfolgt zunächst eine Abgrenzung zu anderen Tätigkeitsfeldern im Umgang mit Daten, um sich dem unscharfen Feld weiter anzunähern. In einem nächsten Schritt werden die verschiedenen vorliegenden Konzepte dargestellt, um zuletzt auf vorliegende Studien- und Ausbildungsgänge im In- und Ausland zu verweisen.

1.3. *Begriffsbestimmung und Abgrenzung zu ähnlichen Berufsbildern*

*„Data Stewardship, Noun, da-ta stew-ard-ship | \ dā-tə'stū-ərd-ship
1: the roles, functions and responsibilities of a data steward
2: the systematic, sustainable and responsible management of data for public benefit“ (Verhulst et al. 2020:2)*

„A Data Steward is an expert on the preparation and treatment of data including data selection, storage, preservation, annotation provenance and other metadata maintenance, and dissemination. Data librarians are professional library staff who are experts on RDM, using research data as a resource or supporting researchers dealing with data (description, archiving and dissemination).“ (European Commission 2021: 21)

Der *Begriff „Stewardship“* stammt originär aus der Managementtheorie und bedeutet eine „aktive Verantwortungsübernahme“ oder wohlwollende „Verwaltung“. Er steht im Gegensatz zur Inhaberschaft („ownership“) oder der Urheberschaft der Daten (Rothfritz 2019: 25). Zunehmend setzt sich der Begriff auch im Deutschen gegen die häufig verwendeten Begriffe „Forschungsdatenmanagement“ und „Data Librarianship“ durch (Rothfritz 2019: 10, Böhme 2022: 2), bleibt dabei aber unklar umrissen. Scholtens et al. (2019: 11) stellen eine Liste mit immerhin acht Definitionen des DSS zusammen. Die Begriffe Data Librarian, Data Manager und Data Curator werden dabei z.T. synonym verwendet (Steinke et al. 2022: 3). Der Wettbewerb pas

sender Begriffe für neu entstehende Sachverhalte bis hin zur Herausbildung eines oder einiger akzeptierter Termini betrifft nicht nur das Feld des DSS (Neuroth & Oevel 2021: 553), sondern ist typisch für den Prozess der Sprachentwicklung in Bezug auf sich verändernde oder neu entstehende Sachverhalte („Begriffsgeschichte“) (Koselleck 1989: 358–359). Die Entwicklung von Begriffen ist immer eine Transferleistung von Vergangenheits-Erfahrungen und Zukunft-Erwartung auf reale Geschehnisse und wurde von Reinhart Koselleck eingehend untersucht. Mit Koselleck kann die Begriffsbildung, die hier skizziert wird, als „Erfahrungsstiftungsbegriff“ verstanden werden, der zwar an überkommenen Erfahrungen und Worten klebt, aber gegenüber zukünftigen Erfahrungen offen ist („Öffnung zur Zukunft“) (Koselleck 2010: 68).² Für Begriffe, die viele Erwartungen in sich tragen, ist es demnach typisch, dass sie „strittig“ sind. Diese „Streitigkeit“ um den passenden Begriff lässt sich auch beim DS oder DSS beobachten.

Auch zum DSS *verwandte Berufsbezeichnungen* wie Data Manager, Data Curator, Data Scientist oder Data Analyst haben keinen klar umrissenen Gegenstand. Zur unscharfen Bestimmung trägt des Weiteren bei, dass durch die fehlende spezifische Ausbildung Personen mit unterschiedlichen Hintergründen diese Tätigkeiten ausüben und unsere Erfahrung damit entsprechend prägen. Nach Steinke et al. (2022:3) lassen sich inhaltlich immerhin Data Scientists bzw. Industrial Data Stewards klar von den DS abgrenzen, die direkt mit den Daten mit dem Ziel einer Wissensgewinnung arbeiten. Diese Abgrenzung nach außen und die Einordnung in das Umfeld werden im Folgenden weitergeführt. Welche Schwierigkeiten eine fehlende etablierte Bezeichnung darüber hinaus mit sich bringt, wird sich auch im Verlauf der Untersuchung, etwa bei der Auswahl relevanter Stellenanzeigen oder der Studiengänge, immer wieder zeigen.

Die Begriffsgeschichte des Ausdrucks DSS und auch wie er sich seit den 2000er Jahren in der englischen Forschung gegen Begriffe wie „data governance“, „data administration“, „digital curation“, „scientific stewardship“ oder „digital stewardship“ durchsetzte, zeichnet Rothfritz (2019: 25–29) in ihrer Arbeit „Data Stewardship als Boundary Work“ nach. Einen Überblick über die frühe Entwicklung des Kompetenzbereichs noch unter dem Begriff „Data Librarianship“ geben Pampel, Bertelmann und Hobohm (2010). Die Bedeutung des Begriffs DSS entwickelte sich demzufolge von einer reinen Infrastrukturaufgabe der Langzeitarchivierung hin zu einem aktiven Datenmanagement im Einklang mit den FAIR-Prinzipien (Wilkinson et al. 2016, Rothfritz 2019: 35). Eine ähnliche Entwicklungsgeschichte – nur mit Bezug auf den im Deutschen präsenten Begriff Forschungsdatenmanagement – hin zu einer stärkeren Orientierung an den FAIR-Prinzipien stellt auch die Landesinitiative „Forschungsdatenmanagement in Brandenburg“ fest, ohne den Begriff DSS dabei explizit zu verwenden (Wuttke et al. 2021: 13).

² Koselleck, der sich vornehmlich mit der frühen Neuzeit auseinandersetzte, führt als Beispiele für diesen Begriffstyp Ausdrücke wie „Ökonomie“, „Staat“ und „Verfassung“ an.

In einer *Abgrenzung des DSS von anderen Tätigkeiten* und im Hinblick auf seine institutionelle Verortung stellen Scholtens et al. fest, dass innerhalb des Feldes der professionellen Betreuung von Daten der:die DS in der Institution eingebunden ist und dort eine Funktion in Kooperation mit anderen Stellen einnimmt: „Although data stewards perform a key function in data stewardship, data stewards bear only part of the responsibility for data stewardship in the institute or the project“ (Scholtens et al. 2019: 14). Auch der EOSC-Report zu „Digital skills for FAIR and Open science“ beschreibt ein Umfeld, in das die DS eingebunden sind und mit denen sie im engen Austausch stehen und vermitteln (European Commission 2021). Die Gruppe („role“)³ der „Data Stewards/Data Librarians“ ist dabei eine der zehn Gruppen relevanter Akteure. Die anderen sind: Researcher, EOSC Enabler, Data Scientist/Data Analyst, Research Software Engineer, Data Research Infrastructure Support Professional, EOSC Educator, Data Curator, Citizen, Policy Maker.

Dass auch hier die Aufgaben, die den jeweiligen Akteuren zugeschrieben werden, unscharf abgegrenzt sind, wird in Abschnitt 2.1 thematisiert. Die Einordnung wird zeigen, dass es besonders mit den Data Research Infrastructure Support Professionals, Research Software Engineers oder Wissenschaftler:innen Überschneidungen in der Betreuung von Informationsinfrastruktur bzw. der Datenauswertung gibt. Zur Gruppe der Policy Makers gibt es ebenso verwandte Punkte.

Auch Whyte et al. siedeln das DSS in einen Kontext verwandter Tätigkeiten mit gewissen Schnittmengen an (Whyte et al. 2018: 12). Das DSS wird im Grenzbereich zur domänenspezifischen Forschung, zu Data Science & Analytics, zu Data Management & Curation sowie zu Data Service Engineering von Informationsmitteln verortet. Bei Whyte et al. erscheint DSS zusammengesetzt aus diesen Teilbereichen, ohne dabei in die Tiefe des jeweiligen Bereiches zu gehen.

Dieses Zusammenkommen unterschiedlicher Bereiche sieht auch fdm.nrw, wenn sie die DS als „fachspezifisch arbeitende Expert:innen [sic] mit einer *Schnittstellenfunktion zwischen Wissenschaft und Infrastruktur*“ beschreiben.⁴ Diese Beobachtung der Schnittstellenfunktion oder des:r Übersetzer:in zwischen den verschiedenen Stakeholdergruppen wird von vielen Forschungsarbeiten geteilt (z.B. Shutsko & Lindstädt 2020: 4; Dutch Techcentre for Life Sciences 2022). Um diese Eigenschaft noch weiter zu untersuchen, aktualisiert Rothfritz das Konzept der „boundary work“ aus der Wissenschaftssoziologie für das DSS. Dieses Konzept wurde zuerst von dem amerikanischen Soziologen Thomas F. Gieryn entwickelt, um Tätigkeiten an der Grenze von wissenschaftlichem und nicht-wissenschaftlichem Arbeiten besser fassen zu können. Vor diesem Hintergrund entwickelt Rothfritz die Tätigkeit des:der DS als „Boundary Spanner“. Der Begriff bedeutet, dass diese Personen hauptsächlich Übersetzungs- und Koordinationstätigkeiten zwischen den beteiligten Gruppen, wie Forschenden und Infrastrukturvertreter:innen, übernehmen. Sie sind für die erfolgreiche Zusammenarbeit zwischen den verschiedenen Gruppen verantwortlich.

³ Der Begriff der Rolle taucht in der Auseinandersetzung mit den Berufs- und Aufgabenfeldern in der Datenverarbeitung sehr häufig auf, um unterschiedliche Gruppen oder Typen voneinander abzugrenzen. Um die begriffliche Trennung zu den etwas später auch als „Rolle“ (nämlich „Rollen“ innerhalb der von der EOSC-Arbeitsgruppe bestimmten „Rolle der DS“) deutlich zu machen, sprechen wir im Folgenden mit Bezug auf die von der EOSC-Arbeitsgruppe separierten „Rollen“ von „Gruppen“.

⁴ <https://web.archive.org/web/20220529114600/https://www.fdm.nrw/index.php/fdm-nrw/data-stewardship/>

Folglich wird das DSS im Sinne einer „langfristigen und nachhaltigen Sorge um Forschungsdaten“ beschrieben, die an der Schnittstelle zwischen Fachwissenschaft und Data Science, Softwareentwicklung und Beratung, Metadaten-Strukturen und Open Science angesiedelt ist sowie zwischen den unterschiedlichen Fachkulturen übersetzt. Im Folgenden wird diese erste Qualifizierung des DSS inhaltlich weiter ausdifferenziert. Die Ausdifferenzierung konzentriert sich auf die Themen Funktionen/Aufgaben, Kompetenzen und notwendige Vorbildung sowie die Verankerung innerhalb der jeweiligen Einrichtung.

2. Konzeptionelle Entwürfe Data Stewardship

In verschiedenen Studien liegen bereits Systematisierungsversuche vor, um das entstehende Arbeitsumfeld des DSS besser zu verstehen. Auswertung und Darstellung dieser Konzeptionen orientieren sich an den folgenden oben skizzierten Leitfragen, um sie anhand der vorliegenden Literatur weiter auszudifferenzieren:

4. Was sind die Aufgaben eines:r DS? (vgl. Abschnitt 2.1)
5. Was sind die notwendige Ausbildung bzw. die Kompetenzen von DS (vgl. Abschnitt 2.2)?
6. Wo sind DS idealtypischer Weise institutionell verankert (z.B. SFB/Exzellenzcluster, NFDI-Konsortium, FDM-Stelle einer Hochschule) (vgl. Abschnitt 2.3)?

2.1. Was sind die Aufgaben von DS?

„[S]cientific data stewardship is considered and addressed from two distinct perspectives: One is Scientific Data-Stewardship: that is, a systematic, scientific way of stewarding data. The other is Scientific-Data Stewardship: that is, the stewardship of scientific data. Stewardship of scientific data requires not only ensuring the integrity and accessibility of data but also its meaning and usefulness. Ensuring and improving the content and value of the scientific data is a big part of Scientific-Data Stewardship.“ (Peng et al. 2018: 7)

Übereinstimmend beschreiben die Forschungsarbeiten zum Thema Data Stewardship (DSS) die Aufgaben als breites und unklar abgegrenztes Feld im Grenzbereich zwischen Wissenschaft, Beratung und Infrastruktur. Ebenso besteht weitgehend Einigkeit darüber, dass eine einzelne Person lediglich Teile dieser Aufgaben leisten kann, oder, wie später zu sehen ist, an andere delegieren kann (vgl. DS-Profil: DS als Generalist:innen). Es gibt folglich in der Regel DS mit verschiedenen Schwerpunktsetzungen. Damit unterliegen die verschiedenen Aufgaben im DSS einer gemeinsamen Verantwortung („collective responsibility“), in der jede:r DS einen Teil zum Ganzen beiträgt („shared responsibility“) (Scholtes et al. 2019).

Vor diesem Hintergrund werden in der Forschung immer wieder Vorschläge gemacht, Rollen zu den **Spezialisierungen** im DSS zu definieren. Diese Konzepte von Rollentypen sind schwierig, weil die Autor:innen in der Regel gleichzeitig keine klare Abgrenzung der Rollen untereinander finden und die Grenzen des DSS somit nicht nur nach außen, sondern auch nach innen verschwimmen (z.B. Verheul et al. 2019, Shutsko & Lindstädt 2020: 4, European Commission 2021: 16, Curdt et al. 2021: 79). Auch

aufgrund der hohen Wandlungsgeschwindigkeit sind die Konzeptionen typischer Rollen in der Betreuung von Daten eher als paradigmatisch oder repräsentativ zu verstehen (European Commission 2021: 16).

Beim DSS handelt es sich also um ein Tätigkeitsspektrum, für das idealtypische Rollen formuliert werden können, die von der:dem einzelnen DS aber in der Regel nicht in dieser idealtypischen Form ausgeübt werden. Ob das Bild von Rollen, die in der Forschung häufig benutzt wird, überhaupt eine produktive Metapher ist oder nicht eher jenes eines Spektrums mit unterschiedlichen Dimensionen, wie es etwa Whyte et al. vorschlagen, sinnvoll ist, wird sich zeigen.⁵

Obwohl hier nur ein Ausschnitt aus der umfassenden Forschung ausgewertet werden kann, geben die einbezogenen Studien dennoch einen guten Überblick über die Konzepte, wobei sich einige wichtige Linien abzeichnen. Ausgehend von der Definition der European Commission von DS für die EOSC erfolgt eine Annäherung an die verschiedenen Aufgabengebieten.

Die Autor:innen der European Commission 2021 haben, wie oben skizziert, Data Stewards/Data Librarians in einem Feld von anderen Berufen, die mit Daten zu tun haben, angesiedelt. Innerhalb dieser Abgrenzung werden den DS folgende Fähigkeiten mit Bezug zur EOSC zugeschrieben:

- 
- **„Deep understanding of FAIR principles to ensure that research data from various domains is aligned with FAIR and CARE (Collective benefit, Authority to control, Responsibility, Ethics) principles.**
 - **Ability to use EOSC-Core and EOSC-Exchange services for data publication and preservation and to facilitate the continued development of an infrastructure and library services to support data discovery, curation, preservation and sharing according to those principles.**
 - **Ability to validate the fulfillment of open science principles in EOSC-Core and EOSC-Exchange services related to data.**
 - **Ability to advise faculty and students on RDM according to the FAIR and CARE principles, including the discovery and reuse of existing datasets, through the EOSC services and ecosystem.“** (European Commission 2021: 21)

In diesem Verständnis von DSS konzentrieren sich DS also auf die Kenntnisse und Umsetzung von FAIR- und CARE-Prinzipien in Bezug auf spezifische Dateninfrastruktur, namentlich Betreuung von Publikations-Services, Discovery-Services, der Durchführung von Langzeitarchivierung, Datenaustausch sowie die Schulung und Beratung. Dabei können auch verwandte Rollen bzw. ihre Aufgaben unter die Typen-Charakterisierung der:er DS fallen (European Commission 2021: 21).

⁵ Hier sei bereits angemerkt, dass die Rollen-Metapher in den Diskussionen mit Expert:innen eher zu Missverständnissen führt. Die Rede von einer „Rolle“ transportiert Vorstellungen der Soziologie, die auf einen Status und eine feste Haltung innerhalb eines sozialen Gefüges verweisen. Dies zeichnet sich eben nicht durch das volatile bewegliche „Dazwischen“ von DS aus, die in ihrer spezifischen Aufgabenstellung auch zwischen mehreren Aufgabenschwerpunkten arbeiten können.

Innerhalb der Gruppe DS werden in der Forschungsliteratur einzelne Aufgaben bestimmt. Die Aufgabenfelder werden dabei oftmals mit Rollen gleichgesetzt (z.B. „Trainer:in“, „Analyst“). Dies erschwert eine klare Unterscheidung von Tätigkeiten und Rollen in der folgenden Darstellung. Bevor im Folgenden insgesamt sieben Cluster von Tätigkeiten identifiziert werden können, wird zunächst ein alternativer Systematisierungsansatz entlang des Ablaufs eines Forschungsprozesses vorgestellt.

Peng et al. (2018) und die Initiative EOSCpilot schlagen jeweils methodische prozessorientierte Modelle vor (Whyte et al. 2018). Peng et al. siedeln die inhaltlichen Aufgaben entlang eines Projekt-Management-Modells „Plan–Do–Check–Act“ (PDCA) an. Da hierbei zugleich der iterative Charakter des DSS betont wird (Peng et al. 2018: 13), ist es oft schwierig, den einzelnen Phasen konkrete Tätigkeiten (z.B. Beratung zu F.A.I.R.) oder eben Rollen zuzuordnen.

Die Initiative EOSC pilot bestimmt in ihrem Report „Skills and Capability Framework“ (Whyte et al. 2018: 15) Aufgaben im DSS als folgende Cluster: (1) Plan and Design, (2) Capture and Process, (3) Integrate and Analyse, (4) Appraise and Preserve, (5) Publish and Release, (6) Expose and Discover, (7) Govern and Assess, (8) Scope and Resource, (9) Advise and Enable.

Die Tätigkeiten werden dabei in einem dreidimensionalen Spektrum positioniert, das durch die Achsen Kompetenzen („competences“, „expertise“), Verantwortung („responsibility“) und Verortung innerhalb einer Organisation („organisation“) definiert ist (Abb. 2 in Whyte et al. 2018:13). Dieses Spektrum-Modell, das komplexere Charakterisierungen zulässt, scheint tragfähiger zu sein als die Darstellung des DSS in Rollen.

Ein Vorteil dieser methodischen Modelle von Peng et al. und Whyte et al. ist die Orientierung am Datenlebenszyklus. Dies steht im Einklang mit der breit geteilten Auffassung, nach der das DSS alle Aufgaben und Verantwortlichkeiten umfasst, die sich auf die Pflege der Daten in den verschiedenen Phasen des gesamten Forschungskreislaufs beziehen.

Neben dieser am Projektablauf orientierten Herangehensweise zur Beschreibung der Tätigkeiten im DSS beschäftigt sich der weit größere Teil der Literatur mit der Darstellung von Aufgabengebieten. Hier können sieben Felder abgesteckt werden. Ein Tätigkeitsfeld kommt in allen gesichteten Systematisierungsversuchen vor und adressiert damit offenbar Kernaufgaben der Tätigkeit als DS: der Bereich der Beratung und Schulung.

Das Feld der *Beratung und Schulung zu FAIR und Open Science*⁶ sowie zu anderen Themen, das auch in dem Papier der EOSC-Arbeitsgruppe als eine zentrale Aufgabe beschrieben wird („Ability to advise faculty and students on RDM“; European Commission 2021: 21), wird von mehreren Autor:innen gleichermaßen als Tätigkeitsfeld oder als eine Rolle von DS verstanden (Verheul et al. 2019; Shutsko & Lindstädt).

⁶ Vgl. <https://www.go-fair.org/resources/faq/ask-question-difference-fair-data-open-data/>.

Unter der Überschrift „generic and advisory task area“ (Verheul et al. 2019: 8) wird diese Tätigkeit von Verheul et al. auf einer generischen Ebene angesiedelt, die von unterschiedlichen Projekten angefragt werden kann. Curdt et al., die die Ergebnisse eines DINI/Nestor-Workshops mit Expert:innen aus dem November 2020 dokumentieren, sehen in diesem Aufgabenfeld dagegen zwei unterschiedliche Tätigkeitsbereiche „Schulung“ und „Beratung/Recherche“ (Curdt et al. 2021: 78). Die Rollen, die Curdt et al. mit Bezug zu diesen Tätigkeiten sehen, nennen sie „Trainer:in“ und „Berater:in/Vermittler:in/Schnittstelle/Übersetzer:in“.

Auch bei Wildgaard et al. wird das Tätigkeitsfeld Beratung und Schulung angeführt, aber ebenso wenig klar abgegrenzt wie bei Verheul et al. und Shutsko & Lindstädt. Dagegen fällt es hier zusammen mit der Aufgabe der Entwicklung von Leitlinien unter die Rolle des „Developer“ (Wildgaard et al. 2020: 27).

Bei Scholtens et al. fehlt dieser explizite Fokus auf Beratung und Schulung (Scholtens et al. 2019). Neben den drei Rollen Policy, Research, Infrastructure wird jedoch die Kompetenz „Service“ genannt, die eine adäquate Unterstützung („adequate support on research data management“) garantieren soll (Scholtens et al. 2019: 15). Unter dieser Unterstützungsarbeit lassen sich auch Beratung und Schulungen fassen.

Ein weiteres Tätigkeitsfeld, das von fast allen Autor:innen als eigenes Aufgabenfeld oder Rolle abgegrenzt wird, betrifft die *Entwicklung von Leitlinien und Standards*. Damit sind sowohl Metadatenstandards als auch die strategische Entwicklung von Workflows/SOP (Standard Operating Procedures) auf institutioneller Ebene gemeint. Bei Verheul et al. wird dieses Aufgabenfeld „Policy, strategy and coordination task area“ genannt (Verheul et al. 2019: 8), bei Scholtens et al. schlicht „data steward – Policy“ (Scholtens et al. 2019:15).

Analog dazu beschreiben Curdt et al. einen Aufgabenbereich „Materialerstellung“ (Curdt et al. 2021: 78). Die Bezeichnung ist etwas irreführend, da sie zunächst wie die reine Dokumentation von Leitlinien oder die zielgruppengerechte Kommunikation von Darreichungen verstanden werden kann. Tatsächlich soll darunter aber auch die Formulierung von „disziplinspezifische[n] Regeln und Leitlinien“, die Erarbeitung von „Handreichungen bspw. für die fachspezifische Implementierung der FAIR-Prinzipien“ sowie die „Sammlung von Best Practices“ fallen. Zusätzlich definieren Curdt et al. als Ergebnis ihres Workshops die Aufgabe der „Qualitätssicherung“, die auch eine Entwicklung von Indikatoren für FAIR-Prinzipien enthält, was sich dem hier aufgemachten Themenfeld der Entwicklung von Leitlinien und Standards zuordnen lässt (Curdt et al. 2021: 78). Bei Curdt et al. umfasst das Aufgabenfeld „Qualitätssicherung“ zudem noch die praktische Prüfung von Daten auf FAIR-Prinzipien.

Bei Wildgard et al. 2020 fällt dem „Developer“ neben der oben beschriebenen Aufgabe der Beratung und Schulung eine ganz ähnliche Aufgabe zu (Wildgaard et al. 2020: 27). Der Developer ist ein:e DS, der:die mit der Entwicklung von „procedures and guidelines“ eine ganz ähnliche Aufgabe hat. Anders sieht es die EOSC-Arbeitsgruppe, die die Aufgabe der Entwicklung von Policies der von den DS/Data Librarians unterschiedenen Gruppe der „Policy Makers“ zuordnet (European Commission 2021: 22).

Etwas kontroverser wird in der Literatur zum DSS das Aufgabenfeld einer *direkten Betreuung von Wissenschaftler:innen und ihrer Daten bzw. der verwendeten Informationsmittel* (Repositorien, Datenbanken, Server, Cloud-Infrastruktur, Suchmaschinen, etc.) diskutiert. Die durchlässige Abgrenzung der Tätigkeiten des DSS sowohl zur Wissenschaft als auch zur Infrastruktur zeigt sich in der häufigen Zusammenlegung der Bereiche in der Forschungsliteratur. Hier wird einerseits von DS erwartet, dass sie eng mit den Wissenschaftler:innen zusammenarbeiten und dabei gegebenenfalls auch Aufgaben der Datenanalyse übernehmen. Andererseits wird mit dieser Aufgabe häufig die Pflege und die Weiterentwicklung der Infrastruktur verbunden.

Diese im Grunde recht unterschiedlichen Aufgaben werden häufig als kombiniertes Aufgabenfeld entwickelt: Verheul et al. sehen hier in der „embedded and operational task area“ etwa eine umfassende technische Unterstützung der Arbeitsgruppe, auch bei Fragen der Datenverarbeitung und -extraktion oder bei der Nutzung von Dateninfrastrukturen (Verheul et al. 2019). Bei Wildgaard 2020 wird diese Tätigkeit unter dem Begriff „Analyst“ gefasst (Wildgaard et al. 2020: 27). Die EOSC-Arbeitsgruppe beschreibt diese Tätigkeit einerseits als Fähigkeit der DS/Data Librarians, Repositorien und Publikations-Services zu nutzen und diese Infrastruktur weiterzuentwickeln (European Commission 2021: 21). Andererseits schreibt die Arbeitsgruppe diese Aufgaben zugleich aber auch den eigens darauf spezialisierten Gruppen der Data Analyst:innen, Data Research Infrastructure Support Professionals, Research Software Engineers oder Wissenschaftler:innen zu (European Commission 2021: 18–20). Hier zeigt sich wiederum deutlich die Durchlässigkeit der Grenzen des DSS zu Wissenschaft und Infrastruktur. Die Erwartung, dass DS sich auch an der Verbesserung der Dateninfrastruktur beteiligen, erfordert dabei mindestens grundlegende Programmierkenntnisse (Wildgaard et al. 2020: 15, 27; Verheul et al. 2019).

Anders als bei den zuvor genannten Konzepten werden bei Scholtens et al. 2019 die beiden Aspekte der Arbeit mit den Forschenden sowie der Betreuung der Infrastrukturen klar auseinandergelassen. Die dort vorgeschlagene Konzeption unterscheidet die Rollen „data steward – research“ und „data steward infrastructure“. In der Rolle „research“ arbeitet der:die DS direkt im Projekt eng mit den Wissenschaftler:innen zusammen und unterstützt sie u.a. bei der Umsetzung von Leitlinien und Datenmanagementplänen (Scholtens et al. 2019: 16). Eine Unterstützung bei der Auswertung der Daten wird nicht genannt und ist offenbar nicht vorgesehen. In der Rolle der „infrastructure“ hat der:die DS eine vermittelnde Rolle zwischen den Wissenschaftler:innen und der IT-Entwicklungsabteilung (Scholtens et al. 2019: 16).

Etwas aus der Reihe fällt das Tätigkeitsfeld der eigenen Publikation von Forschungsergebnissen auf Basis der Daten. Bei Verhulst et al. wird dies unter dem Titel „Dissemination and Communication of findings“ angeführt (Verhulst et al. 2020: 13). Dies aber wohl auch, weil diese Studie das Feld DSS nicht ausschließlich für öffentliche bzw. wissenschaftliche Institute konzeptionell zu fassen sucht. Somit wird die wissenschaftsnahe Arbeit der Publikation von Ergebnissen als explizites Tätigkeitsfeld auch den DS zugewiesen. Bei den anderen vorliegenden Studien fällt dieses Aufgaben

feld vermutlich zum Teil unter die „Betreuung von Wissenschaftler:innen“, weil sich daraus ggf. auch eine Publikationstätigkeit und Ko-Autor:innenschaft ergeben kann.

Als weiteren eigenen Bereich konzipieren Wildgaard et al. den Tätigkeitsschwerpunkt der *rechtlichen Beratung, der IT-Sicherheit und des Datenschutzes*. Auch dieser Typ von DS soll als domänenspezifisch ausgebildete Fachwissenschaftler:in eng im Team mitarbeiten. Die Rolle wird dort „Administrator“ genannt (Wildgaard et al. 2020:27). Der Aspekt der IT-Sicherheit und des Datenschutzes wird auch von Verhulst et al. gesehen und unter der Tätigkeit „Data audit, ethics, and assessment of value and risk“ subsumiert (Verhulst et al. 2020: 11–12).

Eine weitere Rolle, die nur von Wildgaard et al. spezifisch abgegrenzt wird, ist die des „Agent of Change“, der das *Change Management von neuen Informationsmitteln und Standards* begleitet und zwischen Wissenschaftler:innen und Infrastruktur vermittelt (Wildgaard et al. 2020: 27). In diesem Sinne übersetzt der:die DS spezialisierte Richtlinien nutzerfreundlich. Er:sie behält auch ethische Fragen im Blick. Vereinbar mit diesem Aspekt ist der von Peng et al. konzipierte Aufgabenbereich des „Act/Improve“, bei dem die genutzten Workflows und Vorgaben evaluiert und ggf. angepasst werden (Peng et al. 2019: 23).

Die EOSC-Arbeitsgruppe definiert noch ein weiteres Tätigkeitsfeld: die *Umsetzung von FAIR- und CARE-Prinzipien sowie die Qualitätssicherung* von Daten im Hinblick auf diese Maximen (European Commission 2021: 21). Hier lässt sich das von Curdt et al. formulierte Aufgabenfeld der „Qualitätssicherung“ anschließen, das sowohl die operative Prüfung als auch die Entwicklung von Indikatoren umfasst (Curdt et al., 2021: 78). Die Aspekte der ethischen Reflexion auf Daten und ihre Verarbeitung, wie sie durch die CARE-Prinzipien festgehalten werden, sprechen auch Verhulst et al. zusammen mit dem Thema Datenschutz an (Verhulst et al. 2020: 11–12). Verhulst et al. sehen auch die Notwendigkeit einer Förderung und Umsetzung von Standards für die nachhaltige Nutzung von Daten (Verhulst et al. 2020: 13). Die Begrifflichkeit FAIR-Data fällt hier zwar nicht, doch ist implizit mindestens innerhalb einer Einrichtung die nachhaltige Ablage und Aufbereitung von Daten gemeint und unter der Rolle „Nurture data collaboratives to sustainability“ angesprochen. Mit Blick auf die ethischen Aspekte können hier auch die Anteile der Datenethik der von Wildgaard et al. formulierten Rolle des „Agent of Change“ zugeordnet werden (Wildgaard et al. 2020: 27).

Curdt et al. beschreiben mit der *„Vernetzung“* eine Aufgabe von DS, die sonst nur von Verhulst et al. genannt wird. Bei Verhulst et al. zerfällt der Bereich in die zwei Rollen „partnership and community engagement“ und „internal coordination and staff engagement“. Der erste Aspekt richtet sich nach außen, auf andere Institute und Akteure. Der zweite Aspekt zielt nach innen auf die Zusammenarbeit mit institutsinternen relevanten Fachbereichen, der IT oder der Verwaltung (Verhulst et al. 2020: 10–11). Curdt et al. nennen die Rolle, die dem Aufgabenbereich zugeordnet werden kann, „Netzwerker:in“ (Curdt et al. 2021: 78).

Tabelle 1: Rollen, Tätigkeitsbereiche, Aufgaben von Data Stewards.

Inhaltlich bestimmte Rollen/ Tätigkeitsbereiche	Aufgaben	Vertreten von
„embedded and operational DS“	Aufgaben umfassen technische Unterstützung spezifischer Arbeitsgruppen wie z.B. die Datenverarbeitung und -extraktion sowie die Beratung und Lösungsfindung zur Verbesserung der bestehenden Dateninfrastruktur und deren Nutzung	Verheul et al. 2019
„generic and advisory DS“	Beratungen zu FDM für mehrere Projektgruppen	Verheul et al. 2019
„policy, strategy and coordination DS“	strategisch koordinierend, Entwicklung von Leitlinien	Verheul et al. 2019
„validate the fulfillment of open science principles related to data“	Umsetzung FAIR- und CARE-Prinzipien sowie Qualitätssicherung von Daten	European Commission 2021
„advise faculty and students on RDM“	Schulung und Beratung	European Commission 2021
„Use of services for data publication and preservation and to facilitate the continued development of an infrastructure“	Nutzung und Weiterentwicklung von IT-Systemen	European Commission 2021
„Administrator“	Fachwissenschaftler:in, Kommunikation mit Projektmitarbeiter:innen, Expert:in in Data Privacy und IT-Sicherheit	Wildgaard et al. 2020
„Analyst“	Datenanalyse, Informationsmittel-Betreuer*in (Repositorien, Server, Cloud-Infrastruktur, Suchmaschinen etc.)	Wildgaard et al. 2020
„Developer“	Entwickler:in von Guidelines, Berater:in für FAIR	Wildgaard et al. 2020
„Agent of Change“	Projektmanager:in, Schnittstelle zu Kundinn:en, Berater:in zu ethischen Aspekten	Wildgaard et al. 2020
„Schulung + Training“	allgemeine und themenspezifische Workshops, zielgruppenspezifische Schulungen z.B. für Promovierende, Vermittlung von Data Literacy, aber auch externe Trainings „einkaufen“ (z.B. Kurse in Data/Software Carpentry)	Curdt et al. 2021
„Beratung/Recherche“	kontinuierliche Beratung z.B. in Form von Sprechstunden, ethische und rechtliche Beratung, Beratung zu Software, Metadatenstandards und geeignetem Repositorium/Datenjournal, Beratung bei DMP-/Data-Policy-Erstellung, Begleitung über den Datenlebenszyklus hinweg	Curdt et al. 2021
„Materialerstellung“	disziplinspezifische Regeln und Leitlinien formulieren, Handreichungen bspw. für die fachspezifische Implementierung der FAIR-Prinzipien erarbeiten, Sammlung von Best Practices (z.B. in Form eines Wikis)	Curdt et al. 2021

2. Konzeptionelle Entwürfe Data Stewardship

Tabelle 1 Fortsetzung: Rollen, Tätigkeitsbereiche, Aufgaben von Data Stewards.

Inhaltlich bestimmte Rollen/ Tätigkeitsbereiche	Aufgaben	Vertreten von
„Vernetzung“	Öffentlichkeitsarbeit und Community-Aufbau (Webseite, Veranstaltungen), Anreize schaffen, Informationsbrücke zwischen Forschenden und verschiedenen Infrastruktureinheiten bzw. Bereichen der Institution oder externen Datenmanagement-/Fach-Expert:innen etablieren, disziplinübergreifenden Erfahrungsaustausch von Forschenden anregen (z.B. in Form von fachübergreifenden Veranstaltungen)	Curdt et al. 2021
„Qualitätssicherung“	Prüfung nach FAIR-Prinzipien, (disziplinbezogene) Indikatoren (z.B. Key Performance Indicators) mitentwickeln	Curdt et al. 2021
„Partnership and Community engagement“	Identifying partners, key questions and engaging with the community.	Verhulst et al. 2020
„Internal coordination and staff engagement“	Coordinating and engaging with company staff.	Verhulst et al. 2020
„Data audit, ethics, and assessment of value and risk“	Conducting data audits, promoting data ethics, and assessing value and risk.	Verhulst et al. 2020
„Dissemination and Communication of findings“	Disseminating and communicating findings to outside parties.	Verhulst et al. 2020
„Nurture data collaboratives to sustainability“	Nurturing data collaboratives to ensure their long-term sustainability.	Verhulst et al. 2020
„Plan/Define“		Peng et al. 2018
„Do/Create“		Peng et al. 2018
„Check/Assess“		Peng et al. 2018
„Act/Improve“		Peng et al. 2018
„Plan and design“		Whyte et al. 2018
„Capture and process“		Whyte et al. 2018
„Integrate and analyse“		Whyte et al. 2018
„Appraise and preserve“		Whyte et al. 2018
„Publish and release“		Whyte et al. 2018
„Expose and discover“		Whyte et al. 2018
„Govern and assess“		Whyte et al. 2018
„Scope and resource“		Whyte et al. 2018

Tabelle 1 Fortsetzung: Rollen, Tätigkeitsbereiche, Aufgaben von Data Stewards.

Inhaltlich bestimmte Rollen/ Tätigkeitsbereiche	Aufgaben	Vertreten von
„Advise and enable“		Whyte et al. 2018
„Data steward Policy“	„translates general policies, procedures, regulations and principles into guidelines“	Scholtens et al. 2019
„Data steward Research“	„supporting the researcher; implementation domain specific procedures to data management plans“	Scholtens et al. 2019
„Data steward Infrastructure“	„facilitates implementation of the IT infrastructure, can give access to data and software and may also perform hands-on work in the project“	Scholtens et al. 2019

Zwischenfazit: Der Tätigkeitsbereich des DSS ist bisher unklar definiert. Gründe hierfür liegen darin, dass die Datenwissenschaft und ihre Werkzeuge und Standards einem starken Wandel unterliegen. Zudem gibt es eine Reihe von verwandten Arbeitsbereichen, mit denen DS interagieren und zu denen die Aufgaben nicht klar abgegrenzt werden können bzw. durch die Zusammenarbeit verschwimmen. Zusätzlich können spezifische Realisierungen des DSS an einer Einrichtung Gegenbeispiele für Ausnahmefälle liefern, die das Bemühen um eine systematische Bestimmung der Tätigkeiten zusätzlich erschweren.

Die Auswertung eines Ausschnitts von wissenschaftlichen Entwürfen zu idealtypischen Tätigkeitsfeldern von DS hat gezeigt, dass in dem Versuch einer Struktur vor allem inhaltliche Felder definiert werden, aus denen häufig sogenannte „Rollen“ des DSS abgeleitet werden. Etwas weniger häufig werden auch methodische Phasen entlang des Datenlebenszyklus entworfen.

Die wesentlichen inhaltlichen Bereiche, die von fast allen Konzeptionen in der einen oder anderen Weise genannt wurden, sind Beratung und Schulung zu FAIR, Entwicklung von Leitlinien und Standards sowie direkte Betreuung von Wissenschaftler:innen und ihrer Daten bzw. der verwendeten Informationsmittel. Etwas weniger häufig wurden die folgenden Aufgabengebiete in der ausgewerteten Literatur gefunden: ethische Beratung, IT-Sicherheit und Datenschutz, Umsetzung von FAIR- und CARE-Prinzipien inklusive der Qualitätssicherung im Hinblick darauf, Change Management für neue Informationsmittel und Standards sowie Vernetzung mit relevanten Akteuren innerhalb und außerhalb der eigenen Einrichtung. Zum Teil werden diese aber sicherlich von einigen Konzeptionen mitgedacht, aber weniger prominent in den anderen genannten Kategorien mit abgehandelt.

Die starke Ausdifferenzierung der Aufgabenbereiche geht einerseits mit der unklaren Abgrenzung einzelner Kompetenzbereiche einher, andererseits mit der einhellig vertretenen Meinung, dass nicht alle der beschriebenen Tätigkeiten von einer:m einzelnen DS ausgeübt werden können. Das Idealbild von DSS, das in der Literatur

gezeichnet wird, ist ein Team von unterschiedlich spezialisierten DS, die sich gegenseitig ergänzen.

Eine weitere Fragestellung ergibt sich hinsichtlich der Unterstützung und Zusammenarbeit mit der Wissenschaft bzw. der Infrastruktur. Die betrachtete Literatur bleibt bei der Behandlung dieser Fragestellung recht unscharf. Hier zeigt sich aber deutlich die Schnittstellenfunktion der DS zwischen Wissenschaft und der klassischen Forschungsinfrastruktur bzw. den Services zur Nutzung dieser Infrastruktur. Dabei wird die Abgrenzung zu genuinen Aufgaben der Forschung und genuinen Aufgaben der Infrastruktur, wie Pflege, Aufbau, Betreuung von Nutzenden, notwendig. In Bezug auf die Infrastruktur ist fraglich, inwieweit es als DS ausreicht, diese Infrastruktur zu nutzen, zu verstehen und zu erklären, oder ob es auch notwendig ist, diese Infrastruktur selbstständig anzupassen bzw. zu optimieren und dabei auch die Software weiterzuentwickeln oder zumindest bestehende Tools anzupassen.

Auf der anderen Seite steht die Nähe zur wissenschaftlichen Arbeit, bei der sich die Frage stellt, inwieweit DS selbst in die fachspezifische Arbeit eingebunden sind und dann auch als Ko-Autor:innen etwa auf wissenschaftlichen Publikationen mit angeführt werden. Inwieweit domänenspezifisches Fachwissen für DS obligatorisch ist, wird unten unter der Leitfrage 3 „Was sind die notwendigen Ausbildungen bzw. Die Kompetenzen von DS?“ erörtert. Andererseits ist für DS das Feld des DSS auch selbst Gegenstand von Forschung (Peng et al. 2018: 7).

In diesem Sinne lässt sich die ursprüngliche Frage „Was sind die Funktionen/Aufgaben eines:r DS?“ um die Frage nach dem Verständnis des DSS – „Wissenschaft oder Service?“ – erweitern. Im Abschnitt 2.4 werden die differenzierten Forschungsfragen ausgeführt.

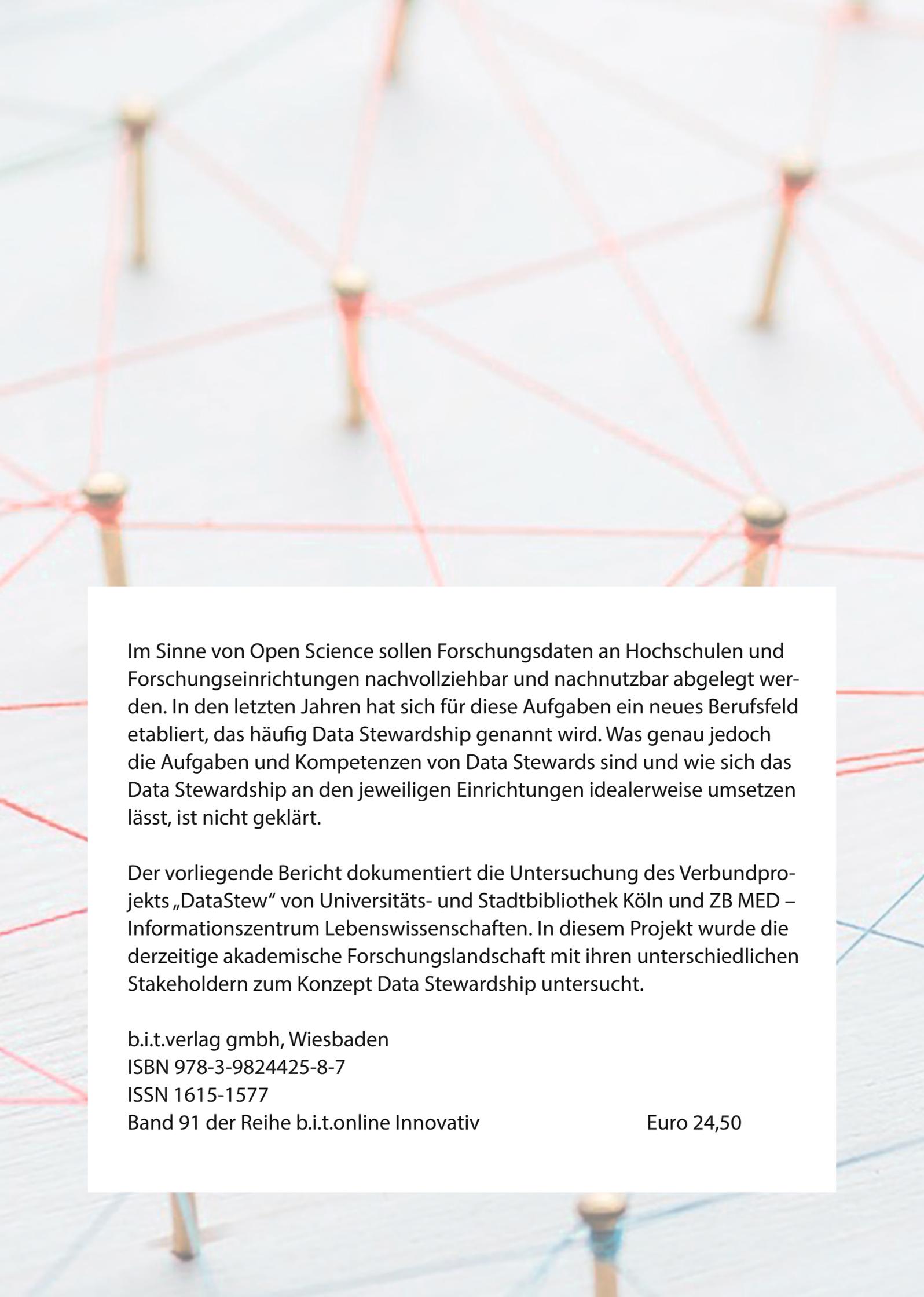
2.2. *Was sind die notwendigen Kompetenzen von Data Stewards?*

Die Frage nach notwendigen Kenntnissen für ein erfolgreiches Arbeiten im DSS zielt auf unterschiedliche Kompetenzstufen im Sinne einer Befähigungstiefe bzw. eines Verantwortungsgrades. In diesem Abschnitt werden zunächst mögliche Kompetenzstufen, die in der Literatur unterschieden werden, vorgestellt und dann konkret auf einige damit verbundene Kompetenzen eingegangen. Dabei ist die Frage nach konkreter Verantwortung eng mit der nach den jeweiligen Aufgaben verbunden, die eine DS ausübt. Im Abschnitt 2.1 wurde bereits deutlich, dass DS in der Regel nur einen Ausschnitt an Tätigkeiten aus dem Spektrum übernehmen.

Kompetenzstufen: Neben den oben genannten Modellen, die auf unterschiedliche Tätigkeitsbereiche fokussieren, weisen Curdt et al., Peng et al. und Whyte et al. systematische Unterscheidungen auf, die mit dem Maß der Expertise zusammenhängen. Mit den Kompetenzen sind Fähigkeiten gemeint, keine abgeschlossenen Examina oder Qualifikationsstufen (BA, MA, Promotion). Gleichzeitig gehen mit der Tiefe der Fähigkeiten auch Unterschiede im Verantwortungsbereich einher. Curdt et al. füh-

ren eine systematische Unterscheidung der Rollen ein, in der eine:r „Expert:in“ die/der „Allrounder:in“ gegenübergestellt wird (Curdt et al., 2021: 78). Peng et al. werden noch spezifischer in ihrer Rollenaufteilung, die einer Kompetenzsteigerung folgt: Sie differenzieren zwischen einem „Point-of-Contact“, der lediglich Anfragen an andere Kompetenzteams weiterleitet, einem „Specialist“, einem „Subject matter expert“ und dem „Steward“. Dabei verfügt der:die Steward über „Extensive knowledge and expertise in a specific domain and general knowledge in other relevant domains, e.g., science/business and technology“ und übertrifft damit noch den Domänenexperten („Subject matter expert“) der Stufe darunter (Peng et al. 2018: 24). Wie wichtig die Spezialisierung auf die Bedarfe der Fachwissenschaft ist, die in der höchsten Kompetenzstufe bei Peng et al. liegen, stellen auch Shutsko und Lindstädt heraus, die am Aufbau der NFDI4Health beteiligt sind (Shutsko & Lindstädt 2020). So gibt es für den Bereich der medizinischen Daten viele domänenspezifische Anforderungen, die personenbezogene Gesundheitsdaten mit sich bringen (Shutsko & Lindstädt, 2020: 1). Auch Whyte et al. führen neben der Dimension der Kompetenz im Sinne einer Fähigkeit eine Dimension der Verantwortung ein (Whyte et al. 2018: 13). Auf der untersten Kompetenzstufe steht hier das Verstehen („comprehend“), gefolgt vom Anwenden („apply“) und schließlich das Bewerten oder Evaluieren und das Synthetisieren von Wissen („evaluate/synthesize“) auf der obersten Stufe. Die jeweiligen Kompetenzstufen dürfen die darunterliegenden Stufen umfassen und auf ihnen aufbauen. Die mit den Kompetenzstufen korrespondierende Verantwortungsdimension beinhaltet bei Whyte et al. die Abstufungen Unterstützung („support“), Ermessensspielraum („discretion“), erheblicher Ermessensspielraum („substantial discretion“), die Festsetzung von Zielvorgaben („objective setting“) sowie zuletzt die volle Rechenschaftspflicht („full accountability“) (Whyte et al. 2018: 13).

Kompetenzen: Die Verantwortungsstufen hängen auch mit konkreten Aufgaben, für die unterschiedliche Expertise nötig ist, zusammen. Whyte et al. unterscheiden in diesem Sinne die unterschiedlichen Services in Kernkompetenzen („Core training services“), Generische Kompetenzen („generic competences“) und Thematische Kompetenzen („thematic competences“). Die Kernkompetenzen umfassen für sie die Fähigkeit, mit Metadaten umzugehen, die generischen Kompetenzen beinhalten, Beratung durchzuführen und Kenntnisse im Umgang mit Daten im Sinne der FAIR-Prinzipien zu haben. Die höchste Kompetenzstufe bezieht sich auf domänenspezifische Kompetenzen (Whyte et al. 2018: 34).



Im Sinne von Open Science sollen Forschungsdaten an Hochschulen und Forschungseinrichtungen nachvollziehbar und nachnutzbar abgelegt werden. In den letzten Jahren hat sich für diese Aufgaben ein neues Berufsfeld etabliert, das häufig Data Stewardship genannt wird. Was genau jedoch die Aufgaben und Kompetenzen von Data Stewards sind und wie sich das Data Stewardship an den jeweiligen Einrichtungen idealerweise umsetzen lässt, ist nicht geklärt.

Der vorliegende Bericht dokumentiert die Untersuchung des Verbundprojekts „DataStew“ von Universitäts- und Stadtbibliothek Köln und ZB MED – Informationszentrum Lebenswissenschaften. In diesem Projekt wurde die derzeitige akademische Forschungslandschaft mit ihren unterschiedlichen Stakeholdern zum Konzept Data Stewardship untersucht.

b.i.t.verlag gmbh, Wiesbaden

ISBN 978-3-9824425-8-7

ISSN 1615-1577

Band 91 der Reihe b.i.t.online Innovativ

Euro 24,50