

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

Aufgaben- und Kompetenzspektrum der Data Stewards an der RWTH Aachen University – allgemein und am konkreten Beispiel des Sonderforschungsbereichs 1382

Lukas C. Bossertⁱ Daniela A. Hausenⁱⁱ
Ute Trautwein-Brunsⁱⁱⁱ

2021

Zitiervorschlag

Bossert, Lukas C., Daniela A. Hausen und Ute Trautwein-Bruns. 2021. Aufgaben- und Kompetenzspektrum der Data Stewards an der RWTH Aachen University – allgemein und am konkreten Beispiel des Sonderforschungsbereichs 1382. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 3/2021: S. 92-105. DOI: [10.17192/bfdm.2021.3.8352](https://doi.org/10.17192/bfdm.2021.3.8352).

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

ⁱIT Center RWTH Aachen University. ORCID: [0000-0003-3076-3968](https://orcid.org/0000-0003-3076-3968)

ⁱⁱUniversitätsbibliothek RWTH Aachen University. ORCID: [0000-0001-9083-0670](https://orcid.org/0000-0001-9083-0670)

ⁱⁱⁱUniversitätsbibliothek RWTH Aachen University. ORCID: [0000-0003-0531-0182](https://orcid.org/0000-0003-0531-0182)

1 Zusammenfassung

Das Themenspektrum und die Aufgabengebiete von Data Stewards sind mitunter sehr unterschiedlich. Je nach zugeordneter Institution bzw. Fachgebiet unterstützen sie Forschende bei verschiedenen und klar formulierten Problemstellungen. Am Beispiel des Data Stewards des Sonderforschungsbereichs 1382 „Darm-Leber Achse“ (SFB 1382) wird gezeigt, wie eine Unterstützung im Forschungsdatenmanagement konkret aussehen kann. Als allgemeine Ergänzung der Aufgabengebiete werden weitere Aspekte der Tätigkeitsbereiche anderer Data Stewards der RWTH Aachen University aufgezeigt. Neben dem Aufgabenspektrum wird auch kurz das Kompetenzspektrum beleuchtet, so dass ein Eindruck über die generellen, heterogenen Anforderungen an Data Stewards und deren Einsatzgebiete vermittelt wird.

Dieser Artikel führt den Beitrag *Data Stewards at RWTH Aachen University – New structures and tasks*¹ fort, welcher auf dem DINI/Nestor-Workshop „Data Stewardship im Forschungsdatenmanagement“ (16.11.2020) gehalten wurde, und baut auf der Publikation *Data Stewards an der RWTH Aachen University – Aufbau eines flexiblen Netzwerks*² auf.

2 Data Stewards als Forschungsunterstützung

Forschungsdatenmanagement (FDM) hat viele Facetten und dient im Wesentlichen dem Ziel, Forschungsprozesse effektiv zu gestalten und qualitativ hochwertige Daten zu produzieren, die den FAIR-Prinzipien³ entsprechen. Um Forschungsdaten im gesamten Datenlebenszyklus von der Planung bis zur Nachnutzung zu pflegen, bietet FDM verschiedene Werkzeuge und Konzepte. Dabei handelt es sich z. B. um Datenmanagementpläne (DMP), Forschungsdatenpolicies, Elektronische Laborbücher (ELN), Clouddienste zum Teilen und Repositorien zum Veröffentlichen der Daten. Diese Werkzeuge und Konzepte können generisch und/oder fachspezifisch sein und befinden sich größtenteils in einem dynamischen Entwicklungsprozess. Für einzelne Forschende bedeutet es einen hohen Zeitaufwand, diese Entwicklungen zu überblicken, zu verfolgen und für sich anzupassen. Obwohl das Bewusstsein für notwendige Veränderungen/Optimierungen der eigenen FDM-Praxis großteils vorhanden ist, werden selbst kleine Schritte aus Ressourcenmangel oder fehlenden Fähigkeiten nicht umgesetzt.

¹Hausen, Daniela, Lukas Bossert. 2020. „Data Stewards at RWTH Aachen University – New structures and tasks“. https://kups.ub.uni-koeln.de/29550/9/01_DINI_WS_Data%20Stewardship_Hausen_Bossert.pdf (letzter Zugriff: 2021-02-22).

²Hausen, Daniela, Jessica Rosenberg, Ute Trautwein-Bruns, und Annett Schwarz. 2020. „Data Stewards an Der RWTH Aachen University – Aufbau Eines Flexiblen Netzwerks“. *Bausteine Forschungsdatenmanagement*, Nr. 2 (November). German:20-28. <https://doi.org/10.17192/bfdm.2020.2.8278>.

³Wilkinson, Marc, Michel Dumontier, Ijsbrand Aalbersberg, et al. 2016. „The FAIR Guiding Principles for scientific data management and stewardship“. *Sci Data* Vol. 3, <https://doi.org/10.1038/sdata.2016.18>.

Dieses ist u. a. daran erkennbar, dass bei (öffentlich) geteilten Daten häufig Dokumentation, Metadaten etc. nur rudimentär bis gar nicht vorhanden sind und die Daten nicht maschinenlesbar veröffentlicht werden.⁴

Das FDM-Team der RWTH Aachen University, welches in der zentralen Infrastruktur der RWTH Aachen University verankert ist, verbreitet über Angebote wie Beratung, Weiterbildung sowie Veranstaltungen zur Awareness und Akzeptanz die Werkzeuge und Konzepte. Allerdings werden hierbei eher generelle Aspekte adressiert oder nur einzelne Werkzeuge und/oder Konzepte fokussiert. Um diese Werkzeuge und Konzepte verstärkt in die Wissenschaft zu bringen und dort praxisnah umzusetzen und anzupassen, setzt die RWTH Aachen University Data Stewards dezentral in den Exzellenzclustern (ExC), Sonderforschungsbereichen (SFB), Graduiertenkollegs (GRK) und Sonderschwerpunktprogrammen der Deutschen Forschungsgemeinschaft (DFG) – als sogenannte „embedded“ Data Steward ein.⁵ Die Finanzierung stammt sowohl aus Projekt- als auch aus Haushaltsmitteln der RWTH, so dass Ende 2020 sieben Data Stewards an der RWTH beschäftigt waren. Darüber hinaus sind dem zentralen FDM-Team mindestens zwölf weitere FDM-verantwortliche Personen mit Data Steward-Aufgaben⁶ bekannt.

Data Stewards sind wissenschaftliche Mitarbeitende mit möglichst fachnahe wissenschaftlichen Hintergrund. Sie sind somit häufig Seiteneinsteigende in das FDM und arbeiten projektspezifisch und wissenschaftsnah. Als Ansprechpersonen dienen die Data Stewards sowohl den Forschenden wie dem zentralen FDM-Team. Wie diese Arbeit beispielhaft aussehen kann, ist im [Abschnitt 3](#) beschrieben.

Für das FDM an der RWTH übernehmen die Data Stewards eine tragende Rolle. Sie sind Teil eines Netzwerks, das hilft, FDM in die Breite zu adressieren und an der RWTH als Ganzes voranzubringen. Zur Unterstützung der Data Stewards hat das zentrale FDM-Team die folgenden Angebote geschaffen:

- Offenes FDM-Netzwerktreffen
- Tandem-Programm (FDM-Experte und Data Steward)
- Mailingliste
- Workshops zur Reflektion und Weiterbildung

Mit den zentralen Angeboten soll zum einen ein schneller Einstieg der Data Stewards in das Thema FDM sowie das Learning-on-the-job vereinfacht und verkürzt werden. Zum anderen soll ein Wissenstransfer, Wissenserhalt und die Nachnutzbarkeit des Wissens generiert und weiterentwickelt werden. Das monatliche *offene FDM-Netzwerktreffen* stellt dafür eine niederschwellige Gelegenheit dar, um einen Austausch unter Data Stewards und/oder Forschenden zu fördern. Hier werden unterschiedliche Themen

⁴Mons, Barend. 2020. „Invest 5% of research funds in ensuring data are reusable“. *Nature*, Vol. 578, 491.

⁵Verheul, Ingeborg, Imming, Melanie, Ringerma, Jacquelij, Mordant, Annemie, Ploeg, Jan-Lucas van der und Martine Pronk. 2019. „Data Stewardship on the map: A study of tasks and roles in Dutch research institutes“. *Zenodo*. <http://doi.org/10.5281/zenodo.2669150>.

⁶Hausen, Daniela, Jessica Rosenberg, Ute Trautwein-Bruns, und Annett Schwarz. „Data Stewards an der RWTH Aachen University“.

des FDM sowie auch deren Herausforderungen diskutiert. Es zeigt sich, dass es allgemeine sowie überwiegend disziplin-spezifische Herausforderungen gibt und dass sich diese Bandbreite auch in den entsprechenden Aufgaben widerspiegelt.⁷

Im nächsten Kapitel schildert Lukas Bossert als Data Steward seine Herausforderungen, Aufgaben und Kompetenzen am Beispiel des SFB 1382.

3 Forschungsdatenmanagement im SFB 1382 „Darm-Leber Achse“

Im DFG-Antrag des Sonderforschungsbereichs 1382 „Darm-Leber Achse“ wurden Arbeitspakete für das FDM formuliert, für deren Durchführung eine halbe Stelle für einen Data Steward eingeplant waren.⁸ Wenige Monate nach dem Start des SFB wurde die Stelle im Oktober 2019 besetzt. Zu den dezidiert im Antrag genannten Aufgaben wurde schnell klar, dass das Aufgabenspektrum eines Data Stewards innerhalb des SFB weit mehr umfassen wird, als dies zunächst angedacht war.

Der SFB erforscht die Mikrobiota, Stoffwechselwege und immunologischen Kreisläufe, die die Darm-Leber Achse bilden. Seit Juli 2019 beteiligen sich knapp 30 Principal Investigators in 18 Teilprojekten, die sich über die Standorte Aachen, Berlin, Düsseldorf, Frankfurt/Main, Leipzig und Jena verteilen. Zusammen mit den Promovierenden, Post-Docs und dem Laborpersonal umfasst der SFB über 70 Personen.

Im Folgenden wird ein Überblick über die sehr unterschiedlichen Bereiche und Aufgaben gegeben, wie die Forschenden des SFB als auch der SFB an sich bzgl. des Datenmanagements unterstützt werden.

Das FDM ist im Teilprojekt Q4 („Central tasks of the Collaborative Research Centre“) angesiedelt⁹, für das der Data Steward die Koordination und Entwicklung aller Datenmanagement-Aktivitäten innerhalb des SFB verantwortet. Zudem fungiert er als Bindeglied zwischen dem zentralen FDM-Team der RWTH und dem SFB. Zu diesen Funktionen wurden im Antrag die Aufgaben konkretisiert: *Die Forschenden sollen bei der Erstellung eines Datenmanagementplans (DMP) unterstützt werden, wobei der Data Steward den Eingabeprozess begleitet und das sukzessive Ausfüllen von allen Teilprojekten im Blick behält. Darüber hinaus soll die Einführung eines Elektronischen Laborbuchs koordiniert werden.* Es zeigte sich jedoch schnell, dass neben diesen zwei im Antrag genannten Aufgabenbereiche weitere Aspekte und Themen auf den Data Steward zukamen. Eine Auswahl wird in den Unterkapiteln beschrieben.

⁷Hausen, Daniela, Jessica Rosenberg, Ute Trautwein-Bruns, und Annett Schwarz. „Data Stewards an der RWTH Aachen University“.

⁸Weitere Stellenanteile sind bei einem zweiten SFB und beim zentralen FDM-Team (über das IT Center) der RWTH Aachen University angesiedelt.

⁹<https://www.crc1382.org/Q04> (letzter Zugriff: 2021-02-19).

3.1 Datenmanagementplan

DMPs stellen eine Möglichkeit der zentralen Dokumentation des FDM dar und helfen dem Data Steward zu Beginn einen Überblick über die einzelnen Teilprojekte zu gewinnen, im Laufe des Projektes den Fortschritt zu beobachten und bei der laufenden Aktualisierung des DMPs zu unterstützen.

An der RWTH wird bereits seit 2017 der Research Data Management Organizer¹⁰ (RDMO) für die Erstellung und das Ausfüllen der DMP-Vorlagen eingesetzt. Daher wurde auch in diesem Fall auf das RDMO-Tool zurückgegriffen, um in enger Zusammenarbeit mit dem zentralen FDM-Team von den Erfahrungen zu profitieren und die Umsetzung zu beschleunigen. Es zeigte sich schnell, dass bestehende DMP-Vorlagen für die Forschenden im SFB 1382 zum einen zu umfangreich und zum anderen zu unspezifisch waren. Häufig berücksichtigen bestehende Fragenkataloge nicht die IT-Infrastruktur vor Ort und lassen wichtige Themenkomplexe des medizinischen Bereichs außen vor, wie beispielsweise die Art der Labordokumentation. Eine entsprechende Anpassung musste vorgenommen werden, um die Forschenden effektiv bei der Erstellung von DMPs zu unterstützen. Auch hier profitierte der Data Steward von den Erfahrungen des zentralen FDM-Teams, das sich zum Beispiel bei der Research Data Alliance in der Arbeitsgruppe *Discipline-specific Guidance for DMPs*¹¹ stärker mit dem Thema auseinandersetzt und dadurch wichtige Impulse liefern konnte.

In mehreren Iterationen wurde eine Vorlage erarbeitet und in die lokale RDMO-Instanz eingespielt. Anschließend wurde für jedes Teilprojekt ein DMP in RDMO angelegt, zu dem der jeweilige Principal Investigator eingeladen wurde. In individuellen Videokonferenzen mit den Forschenden der SFB-Teilprojekte ist der Data Steward die Vorlage durchgegangen und es wurden gemeinsam die im Teilprojekt entstehenden und zu erwarteten Daten und Datentypen eruiert. Das Ausfüllen der DMP-Vorlage erfolgte anschließend in Verantwortung der Principal Investigators und deren Projektmitarbeitenden. In regelmäßigen Abständen beobachtet der Data Steward den Fortschritt und hilft bei Fragen. Die größte Herausforderung für die Forschenden ist die Strukturierung der individuellen Daten in Datensätze. Wann fasst man Daten unter welchen Gesichtspunkten als zusammengehörend auf? Da dies von Projekt zu Projekt unterschiedlich ist, berät der Data Steward die Forschenden bei der Entwicklung von Strukturen.

3.2 Elektronisches Laborbuch

Neben einem DMP stellt ein elektronisches Laborbuch (ELN) für die Forschenden ein weiteres wichtiges Hilfsmittel bei der zunehmenden Digitalisierung dar, die sich auch auf den Laborbereich und die Dokumentation von Versuchen auswirkt. Der Umstieg von einer analogen zu einer digitalen Dokumentation ist zunächst für den Forscher

¹⁰<https://rdmorganiser.github.io/> (letzter Zugriff: 2021-02-28) bzw. <https://rdmo.itc.rwth-aachen.de> (letzter Zugriff: 2021-04-29).

¹¹<https://www.rd-alliance.org/groups/discipline-specific-guidance-data-management-plans-wg> (letzter Zugriff: 2021-02-28).

weit aufwändiger als ein reines Eintragen der Versuchsaufbauten oder -ergebnisse in ein Textdokument. Aufwändiger deswegen, weil bei einem elektronischen Laborbuch Eintragungen standardisiert und mitunter formalisierter als bei einem persönlich, physisch geführten Laborbuch sind. Die Dokumentation wird umfangreicher, weil bei einem Versuch Verknüpfungen zu anderen Experimenten, Datensätzen oder Auswertungsprotokollen gezogen werden können. Dennoch ist dieser Prozess gewinnbringend, weil gerade unterschiedliche Komponenten eines Versuchs gebündelt und maschinell prozessierbar gemacht werden. Dabei gilt es stets verschiedene Vorgaben der Guten Wissenschaftlichen Praxis¹² einzuhalten, wie bspw. die Dokumentationspflicht der Autorenschaft einer Eintragung oder Veränderung im ELN oder die nach Möglichkeit Zugänglichmachung von Datensätzen.

Schnell wurde klar, dass die Einführung eines ELNs, das nur von den Mitgliedern des SFB 1382 jeweils separat genutzt wird, im Forscheralltag nicht realistisch ist, da in den jeweiligen Organisationseinheiten meist mehrere Personen gleichzeitig an einem Versuch beteiligt sind und gemeinsam das ELN nutzen und darauf zugreifen müssen. Die Einführung eines ELN ist daher nur in Abstimmung mit der gesamten medizinischen Fakultät sinnvoll, was Mehraufwand bedeutet. Der Data Steward übernimmt in diesem Prozess die Koordination der Termine und Aufrechthaltung der Kommunikation innerhalb der dafür eingerichteten Arbeitsgruppe, die sich aus unterschiedlichen Personen verschiedener Abteilungen und Positionen zusammensetzt – vom Prodekan, über Institutsdirektor*innen, Mitarbeitende der IT-Abteilung des Universitätsklinikums bis hin zum/zur medizinisch-technischen Assistent*in.

Die Erstellung, Durchführung und Auswertung einer Bedarfsanalyse innerhalb der medizinischen Fakultät, um im Vorfeld auf benötigte Funktionalitäten Rücksicht nehmen zu können, markierte den Startpunkt des gesamten Evaluationsprozesses.

Seitdem wird in iterativen Schritten nach einem konkreten Produkt gesucht, wobei verschiedene Hürden hinsichtlich des Datenschutzes und der IT-Infrastruktur genommen werden müssen. Die Kommunikation mit den Herstellern der ELNs, die Organisation von Vorfürhungen und das Einrichten von Testgruppen wurde vom Data Steward übernommen, der in diesem Einführungsprozess die Rolle und Aufgaben eines Projektmanagers innehat.

Seitens des zentralen FDM-Team wurde hierzu bisher nur eine geringe Expertise aufgebaut, da das Thema ELN nur unter generischen Gesichtspunkten betrachtet wurde. ELNs sind allerdings häufig fachspezifisch ausgerichtet, da verschiedene Fächer unterschiedliche Anforderungen haben. Die hier gewonnenen Erfahrungen des Data Stewards werden an das zentrale FDM-Team reflektiert und auch im offenen FDM-Netzwerktreffen präsentiert.

¹²Bspw. <https://wissenschaftliche-integritaet.de/kodex/dokumentation/> (letzter Zugriff: 2021-02-19).

3.3 Datenorganisation und -bereinigung

Wie bereits erwähnt spielt die Entwicklung von Best Practices im Sinne von Handlungsempfehlungen für individuelle Bereiche, eine entscheidende Rolle,¹³ daher findet die Arbeit als Data Steward auch ganz konkret in einzelnen Teilprojekten und auf Dateiebene statt. Beispielsweise wird ein Projekt in der Datenerfassung und Dokumentation unterstützt, in dem Patienten zu verschiedenen Zeitpunkten Blut entnommen und mit unterschiedlichen Parametern ausgewertet werden. Die Datenerfassung und Dokumentation erfolgten bis dato in einer Excel-Datei. In engem Austausch mit den Projektbeteiligten wird nun die Dokumentation verbessert und gleichzeitig der Weg für die spätere Datenpublikation bereitet. Hierbei wird unter anderem eine sinnvolle und einheitliche Spaltenbenennung sowie kontrolliertes Vokabular eingeführt. Zudem wird eine Dokumentation der Metadaten vorgenommen, um ein Mapping nach einer standardisierten Ontologie zu ermöglichen. [Abbildung 1](#) und [2](#) zeigen die Datenorganisation im Tabellenblatt vor und nach der gemeinsamen Betrachtung. In [Abbildung 3](#) sind die Datenfelder dokumentiert, sodass auch für Außenstehende die Datenprovenienz und Dateneingabe leicht nachvollziehbar sind und sich Nachnutzende schnell(er) einen Überblick über die Daten verschaffen können.

	A	B	C	D	E	F	G	H	I	J
1	FACS Sample ID	Zellzahl	Sample handling	Pat-ID	Lab-ID	Biobank-ID (UKP)	Aufnahme	Geburtsdatum	Alter	Geschlecht
2	S1	2,247*10 ⁶	Frozen	166	LV2002010621	UKP5021575246F	04.05.2020	07.01.1964	56	M
3	S2	2,1*10 ⁶	FRESH	167	LV2002029589	UKP7531655608C	1.02.2020	22.1.1937	83	M
4	S3	1,575*10 ⁶	Frozen	168	LV2002042498	UKP4734173975D	04.09.2020	2.04.1945	75	W
5	S4	2,975*10 ⁶	liquid	169	LV2002511086	UKP0744831512E	15.08.2020	13.1.1964	56	M
6	S5	2,0*10 ⁶	Froen	170	LV2002051688	UKP3746829544V	2.05.2020	2.05.1992	28	M
7	S6	2,331*10 ⁶	Frozen	171	LV2002514028	UKP742105612D9	06.09.2020	17.12.1966	54	
8	S7		frsh	172	LV2002513829	UKP65337288310	23.09.2020	05.02.1976	44	M
9	S8	2,7*10 ⁶	Frozen	173	LV2002515631	UKP25476702710	1.05.2020	01.06.1949	71	M
10	S9	1,425*10 ⁶	Frozen	174	LV2002515278	UKP1284651396F	27.09.2020	9.8.1940	80	M
11	S10	2,175*10 ⁶	Frozen	175	LV2002515818	UKP06046254078	10.10.2020	1.6.1964	56	W
12	S11	3,1*10 ⁶	Frozen	176	LV2002518518	UKP3406107799F	6/21/2020	6/25/1973	47	M
13	S12	1,95*10 ⁶	Frozn	177	LV2002514298	UKP253722024B9	10/21/2020	10/1/1937	83	M
14	S13	1,4*10 ⁶	Frozen	178	LV2002512694	UKP753968652DB	5/4/2020	12/4/1996	24	W
15	S14	1,55*10 ⁶	Frozen	179	LV2002515817	UKP2527316861B	10/24/2020	1/15/1953	67	
16	S15	3,925*10 ⁶	Frozen	180	LV2002515839	UKP219245104BE	10/23/2020	5/16/1944	76	M
17	S16	4,9*10 ⁶	deep frozen	181	LV2002524576	UKP125283874C9	8/2/2020	2/1/1958	62	M
18	S17	1,4*10 ⁶	Frozen	182	LV2002525650	UKP1616518832D	10/28/2020	12/7/1970	50	M
19	S18	2,7*10 ⁶	Frozen	183	LV2002575948	UKP78152624745	9/28/2020	4/5/1951	69	
20	S19	2,05*10 ⁶	Fresh	184	LV2002515775	UKP468484497FE	10/8/2020	10/12/1969	51	M
21	S20	0,775*10 ⁶	Frozen	185	LV2002516928	UKP593262851D3	11/29/2020	11/17/1969	51	W

Abbildung 1: Ursprüngliche Excel-Datei zur Versuchsdokumentation (Daten sind authentisch nachgebildet, aber nicht original). In Spalte C wird ersichtlich, dass kein kontrolliertes Vokabular verwendet wurde, sodass mehr als nur die zwei zulässigen Werte (fresh, frozen) eingetragen sind. Problematisch ist bspw. die Eingabe des Datums (Spalte G und H), da sie nicht standardisiert und uneinheitlich ist.

¹³Ziel im SFB ist es, aus diesen Best Practices ein „Data Handling Protocol“ (FDM-Policy) zu erarbeiten. Damit wird bewusst der Ansatz gewählt, gemeinsam mit den Forschern und Forscherinnen die

	A	B	C	D	E	F	G	H	I	J	K	L	
1	UUID	ID	patient_ID	patient_ID_Biobank	patient_date_of_rec	patient_date_of_bir	patient_age	patient_sex	patient_ARDS	patient_days_of_f	sample_ID	lab	sample_cell_count
2	3qgqk	1	166	UKP5021575246F	2020-05-04	1964-01-07	56	male	true	29	LV2002010621		2.25
3	vpdn6	2	167	UKP7531655608C	2020-02-01	1937-01-22	83	male	false		LV2002029589		2.10
4	8bfor	3	168	UKP4734173975D	2020-09-04	1945-04-02	75	female	false		LV2002042498		1.58
5	jfzcu	4	169	UKP0744831512E	2020-08-15	1964-01-13	56	male	true	43	LV2002511086		2.98
6	w3tcl	5	170	UKP3746829544V	2020-05-02	1992-05-02	28	male	true		LV2002051688		2.00
7	jpnlw	6	171	UKP742105612D9	2020-09-06	1966-12-17	54	female	false	13	LV2002514028		2.33
8	biz5a	7	172	UKP65337288310	2020-09-23	1976-02-05	44	male	false		LV2002513827		
9	ooh5h	8	173	UKP25476702710	2020-05-01	1949-06-01	71	male	false	11	LV2002515631		2.70
10	mk12i	9	174	UKP1284651396F	2020-09-27	1940-08-09	80	male	false	10	LV2002515278		1.43
11	f9pzb	10	175	UKP06046254078	2020-10-10	1964-06-01	56	female	true	8	LV2002515818		2.18
12	yl8ym	11	176	UKP3406107799F	2020-06-21	1973-06-25	47	male	true	34	LV2002518518		3.10
13	6dqld	12	177	UKP253722024B9	2020-10-21	1937-10-01	83	male	false	5	LV2002514298		1.95
14	guvbx	13	178	UKP753968652DB	2020-05-04	1996-12-04	24	female	true	9	LV2002512694		1.40
15	7ul1o	14	179	UKP2527316861B	2020-10-24	1953-01-15	67		true	22	LV2002515817		1.55
16	6vh0b	15	180	UKP219245104BE	2020-10-23	1944-05-16	76	male	true	11	LV2002515839		3.93
17	Scazn	16	181	UKP125283874C9	2020-08-02	1958-02-01	62	male	true		LV2002524576		4.90
18	71ljq	17	182	UKP1616518832D	2020-10-28	1970-12-07	50	male	true	11	LV2002525650		1.40
19	tisd7	18	183	UKP78152624745	2020-09-28	1951-04-05	69		true		LV2002575948		2.70
20	ppmm4	19	184	UKP468484497FE	2020-10-08	1969-10-12	51	male	true		LV2002515775		2.05
21	8dwba	20	185	UKP593262851D3	2020-11-29	1969-11-17	51	female	false		LV2002516928		0.78

Abbildung 2: Excel-Datei nach Besprechung und Anwendung von Standardisierungen und Vereinheitlichungen (Daten sind authentisch nachgebildet, aber nicht original). Um Datensätze eindeutig zu identifizieren wurde ein Identifier (Spalte A) eingeführt, der bisherige Identifier (Spalte B) wurde vereinfacht. Datumsangaben wurden nach der ISO 8601 vereinheitlicht. Wenn es Angaben im Sinne von „ja/nein“ / „wahr/falsch“ gibt, wurde dies mit Boolean-Werten „true/false“ umgesetzt. In Spalte L wurde die Zellzahl einheitlich im englischen Format ohne wissenschaftlichen Exponenten umgesetzt.

	A	B	C	D	E	F	G
1	Field	Original fieldname	description	category	control	values	origin
2	UUID		a computer generated code (length of 5) for identifying a data set	all	false	alphanumeric	externally generated (e.g. https://
3	ID	FACS Sample ID	id for handling the samples, human readable	all	false	integer	Laboratory of the CABS-Group
4	patient_ID	Pat-ID	ID of patient in the system of the Biobank	patient	false	integer	Biobank
5	patient_ID_Biobank	Biobank-ID (UKP)	ID for the database of the Biobank	patient	false	alphanumeric (15)	Biobank
6	patient_date_of_record	Aufnahme	Date when patient has been enrolled in study	patient	false	date (YYYY-MM-DD)	Laboratory of the CABS-Group
7	patient_date_of_birth	Geburtsstag	Date of birth of patient	patient	false	date (YYYY-MM-DD)	Biobank
8	patient_age	Alter	Sex of birth of patient	patient	false	integer	Biobank
9	patient_sex	Geschlecht	Check whether patient had Acute Respiratory Distress Syndrome (ARDS)	patient	true	male, female	Biobank
10	patient_ARDS	Group	Amount of days patient had fever when enrolled in study	patient	true	true, false	Laboratory of the CABS-Group
11	patient_days_of_fever	FeverDays	ID of sample	patient	false	integer	Laboratory of the CABS-Group
12	sample_ID_lab	Lab-ID	Amount of cells in one sample (value times 10 ⁶)	sample	false	alphanumeric	Laboratory of the CABS-Group
13	sample_cell_count	Zellzahl	Condition of sample when arrived in laboratory	sample	false	float	Laboratory of the CABS-Group
14	sample_handling	Sample handling	Date when blood sample has been taken from patient, day 1	sample	true	Fresh, Frozen	Laboratory of the CABS-Group
15	sample_slot_A	Tag 1	Date when blood sample has been taken from patient, between day 5 and 7	sample	false	date (YYYY-MM-DD)	Laboratory of the CABS-Group
16	sample_slot_B	Tag 5-7	Date when blood sample has been taken from patient, between day 12 and 14	sample	false	date (YYYY-MM-DD)	Laboratory of the CABS-Group
17	sample_slot_C	Tag 12-14		sample	false	date (YYYY-MM-DD)	Laboratory of the CABS-Group

Abbildung 3: Metadatenschema zur Tabelle in Abb. 2 enthält neben Feldbezeichnung und Beschreibung auch die Angabe der zulässigen Werte und Provenienz.

Die konkrete Zusammenarbeit mit den Forschenden an ihren Datensätzen war Inhalt einer Ausgabe des Newsletters (siehe unten), sodass alle Mitglieder des SFB von den Diskussionen und Empfehlungen profitieren.¹⁴

konkreten Handlungsabläufe zu besprechen, zu dokumentieren und soweit aufzuarbeiten, dass sie für andere Teilprojekte übertragbar sind.

¹⁴Vgl. <https://www.crc1382.org/4956/> (letzter Zugriff: 2021-04-29).

3.4 Newsletter und Webseite

Die Aufgaben als Data Steward des SFB 1382 sind vielfältiger als die Verwaltung von DMPs oder die Projektorganisation des ELN-Einführungsprozesses. Dabei sind die Grenzen zwischen der Hilfe zu FDM-Prozessen und der Hilfe bei einer allgemeinen Verbesserung von IT-Prozessen fließend. Das Erbringen der Hilfeleistung erfordert einen hohen Kommunikationsaufwand, damit sie vielen Personen des SFB und weiteren medizinischen Einrichtungen zugutekommen kann.

Neben der technischen Betreuung der SFB-eigenen Webseite¹⁵ verantwortet der Data Steward die Unterseite zum FDM, auf der in regelmäßigen Abständen in kurzen Newslettern über das FDM an sich und innerhalb des SFB informiert wird (vgl. [Abbildung 4](#)).¹⁶ Es ist eine bewusste Entscheidung, mit den Newslettern nicht nur die Mitglieder des SFB mit kleinen Tipps, Tricks und Hinweisen zu einem (guten) Datenmanagement zu bewegen, sondern mit den öffentlich zugänglichen Newslettern alle Interessierten gleichermaßen zu informieren. Der Newsletter wird nicht verschickt, sondern auf der Webseite des SFB publiziert und ist über eine eindeutige URL referenzierbar. Dies ermöglicht neue Mitglieder auf einen gemeinsamen Informationsstand zu bringen oder bei wiederkehrenden Problemen oder Hilfeanfragen auf eine bestimmte Ausgabe des Newsletters zu verweisen. Auf neue Ausgaben des Newsletters wird per Email oder per Ankündigung auf der Webseite des SFB hingewiesen. Mit der Sichtbarkeit des Newsletters über die Grenzen des SFB hinweg, soll zu einer regen Vernetzung und zum Informationsaustausch mit anderen medizinischen Einrichtungen angeregt werden.

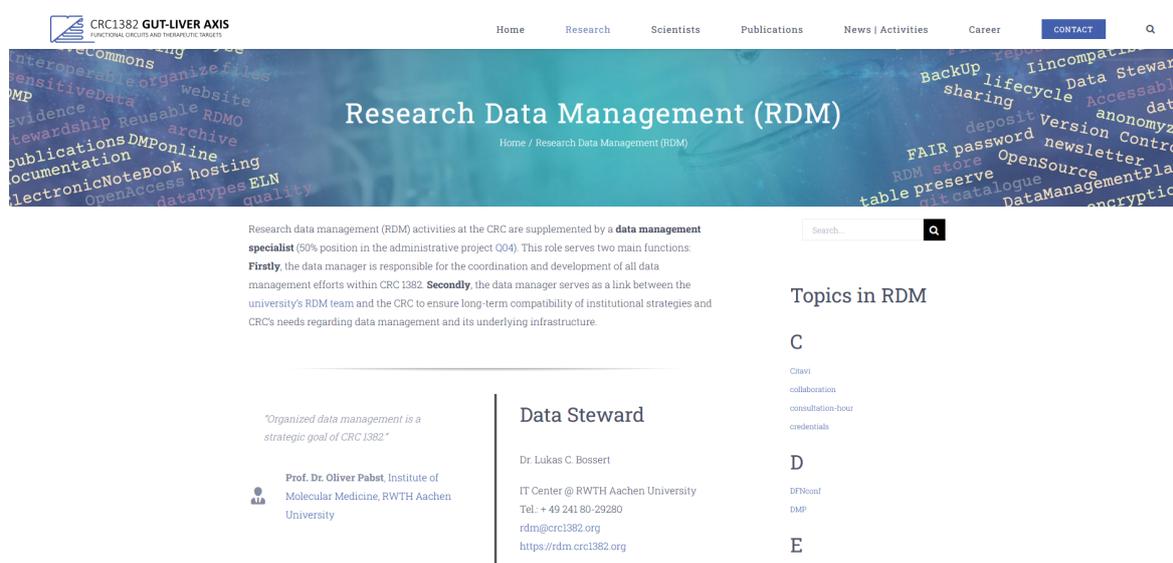


Abbildung 4: Bildschirmfoto der FDM-Unterseite vom 2021-02-19.

¹⁵<https://www.crc1382.org> (letzter Zugriff: 2021-02-19).

¹⁶<https://www.crc1382.org/rdm> (letzter Zugriff: 2021-02-19).

Auch diese Arbeit fließt an das zentrale FDM-Team zurück, so dass die Themen zum Beispiel von weiteren Data Stewards für andere Fachbereiche adaptiert werden können.

3.5 Auswertung der Publikationen

Forschungoutput lässt sich unter anderem an der Anzahl der Publikationen messen. Damit die Datenbasis aktuell und zuverlässig bleibt, kümmert sich der Data Steward um die Pflege der Publikationen des SFB in der Publikationsdatenbank der Hochschule *RWTHpublications*. Neue Publikationen werden eingetragen und mit dem SFB verknüpft.¹⁷ Damit können die Daten für verschiedene Zwecke nachgenutzt werden: Zunächst werden in einem automatisierten Prozess die Publikationsdaten auf der Webseite des SFB angezeigt, sodass sie auch dort durchsuchbar sind.¹⁸ Hierbei profitierte der Data Steward vom engen Austausch mit den Kolleginnen und Kollegen des zentralen FDM-Teams aus der Hochschulbibliothek, das mit Hinweisen zu standardisierten Publikationsmetadatenschemata und deren Verarbeitung weiterhelfen konnte.

Der Vorteil dieser Vorgehensweise ist, dass Publikationsdaten nur an einer Stelle gesammelt werden müssen. Mit *RWTHpublications* hat die RWTH eine Hochschulbibliographie, in der ohnehin alle Publikationen von Forschenden erfasst und kuratiert werden. Der Data Steward braucht sich somit nur um die stetige Aktualisierung der zentralen Publikationsdatenbank kümmern und kann qualitätsgeprüfte Daten für die lokalen Anwendungen nachnutzen.

Im Zuge des technischen Replikationsprozesses der Publikationsdaten werden diese zugleich für ein tagesaktuelles Reporting verwendet. Es werden nicht nur Statistiken erstellt, aus denen hervorgeht, in welchen Zeitschriften wie oft publiziert wurde, oder die Gesamtzahl der Publikationen innerhalb eines Jahres, sondern es wird auch der Grad der Vernetzung des SFB gemessen (vgl. [Abbildung 5](#)). Dies erfolgt über die Auswertung der Co-Autorenschaften von PIs über die Publikationen. Diese Netzwerke¹⁹ und die damit verbundenen Analysen können nicht nur auf Ebene der Forschenden oder deren Institute gemacht werden, sondern auch auf Ebene der einzelnen Teilprojekte. Dabei werden Netzwerkstrukturen (wie bspw. eine Gruppenerfassung) innerhalb des SFB sichtbar. Dieses Reporting ist nicht nur geeignet, um die Zusammenarbeit innerhalb des SFB zu „messen“, sondern auch um schnell korrekte und differenzierte Daten für einen Statusbericht oder für einen Folgeantrag zu liefern.

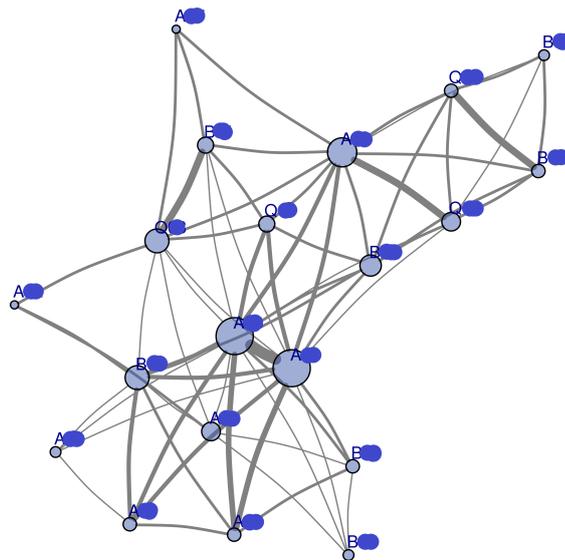
Für diese Arbeit helfen dem Data Steward Kenntnisse in verschiedenen Programmier-

¹⁷<https://publications.rwth-aachen.de/record/783071> (letzter Zugriff: 2021-02-19).

¹⁸<https://www.crc1382.org/publications> (letzter Zugriff: 2021-02-19). Die Vorgehensweise des Migrationsprozesses und v. a. die Aufbereitung der Publikationsdaten wurde publiziert in: Lukas C. Bossert 2020. „Mit »biber –tool« Bibliografieinträge bearbeiten“. *Die TEXnische Komödie*. Vol. 2, 24–33, <https://doi.org/10.5281/zenodo.3784922>.

¹⁹Vgl. Perianes-Rodríguez, A., Olmeda-Gómez, C. & Moya-Anegón, F. Detecting, identifying and visualizing research groups in co-authorship networks. *Scientometrics* 82, 307–319 (2010). <https://doi.org/10.1007/s11192-009-0040-z>.

und Skriptsprachen wie Python, LaTeX, R oder awk. Die technische Auswertung der Publikationsdaten ist in einem RWTH-offenen GitLab-Repository dokumentiert und bereits von vornherein kompatibel programmiert. Die Ergebnisse wurden im Netzwerktreffen vorgestellt, sodass auch andere Data Stewards die Skripte für die Auswertung der Publikationen für ihr Institut/SFB/EXC nachnutzen und zur gemeinsamen Weiterentwicklung beitragen können.²⁰



graph density: 0.352631578947368

Abbildung 5: Netzwerkanalyse der Teilprojekte des SFB 1382 auf Grundlage der Co-Autorenschaften von Publikationen (Schwärzung der konkreten Teilprojektnummern). Die Angabe „graph density“ liefert einen Wert zwischen 0 und 1, was den Wert der Vernetzung wiedergibt. Je höher der Wert, desto mehr Verknüpfungen zwischen den Knoten gibt es; das heißt, je höher der Wert desto intensiver wird über Projektgrenzen hinweg kooperiert und schlussendlich publiziert.

²⁰Der Zugriff auf das Repository (<https://git.rwth-aachen.de/rdm-bundle> [letzter Zugriff: 2021-05-18]) wird auf Anfrage gewährt.

3.6 Unterschiedliches Aufgaben- und Kompetenzspektrum

Wie oben ausgeführt gehören zu den Aufgaben von Lukas Bossert als Data Stewards im SFB 1382 neben den beiden Schwerpunkten DMP und ELN auch das Erstellen und die Pflege eines Newsletters und einer Webseite ebenso wie die Auswertung der Publikationen und die Datenorganisation. In Teilen unterschiedlich hierzu ist das Aufgabenspektrum des Data Stewards des Exzellenzclusters „Exzellenzcluster *Fuel Science Center (FSC)* – Adaptive Umwandlungssysteme für erneuerbare Energie- und Kohlenstoffquellen“ (Exzellenzcluster EXC 2186)²¹. Hier stehen unter anderem die Erarbeitung einer FDM-Policy sowie eines Konzepts zur Archivierung der Daten im Mittelpunkt. In einem längerfristigen Ziel soll eine web-basierte Integrationsplattform eingeführt werden, bei der die verschiedenen Bausteine und Tools implementiert sind.²²

In einem internen Workshop der RWTH Aachen University mit den Data Stewards kristallisierten sich als gemeinsamer Nenner zum einen die Einführung IT-unterstützter Prozesse als Gemeinsamkeit, auch wenn die Softwarelösungen differieren. In einem weiteren Exzellenzcluster wird die vorhandene Softwarelösung um Datenmanagement-Aspekte wie z. B. die Einführung einer Ontologie erweitert. Zum anderen zeigte sich als überschneidende Aufgabe die Ausarbeitung einer projektspezifischen DMP-Vorlage sowie die Unterstützung der Forschenden bei der Befüllung der DMP-Vorlage. Die meisten Data Stewards nutzen das DMP-Konzept, um mit überschaubarem Aufwand den Weg zu strukturiertem und koordiniertem FDM in ihrem Verantwortungsbereich einzuschlagen, und arbeiten an einer angepassten DMP-Vorlage. Erste Forschende haben die jeweilige disziplinär und institutionell angepasste DMP-Vorlage für sich genutzt und erste DMPs erstellt. Durch die Anpassung der Vorlage sind Forschende eher bereit einen DMP zu befüllen, allerdings gilt es abzuwarten, ob dieses auch dauerhaft funktioniert oder ob weitere Unterstützung von Seiten des Data Stewards erforderlich ist. Neben den Aufgaben wurde in dem internen Workshop der Bereich Kompetenzen und Erwartungen thematisiert. Bei den Kompetenzen waren sich die Data Stewards einig – Geduld, Kommunikationskompetenz, Vernetzung und Programmierkenntnisse. Bei der Frage nach den Erwartungen an die Position Data Steward variierten die Antworten hingegen. Generelle Erwartungen wie „Sichtbarkeit“, „Sandwichposition händeln“, „Bewusstsein schaffen“, „dem wissenschaftlichen Mitarbeitenden das FDM leichter machen“ und „Einfach nutzbare Tools zur Verfügung stellen“ wurden genannt. Disziplinspezifische Erwartungen zielten sowohl auf die (Meta-) Daten-Perspektive „Metadatenschemata für Datentypen entwickeln“ und „Alle Datentypen der Projekte zu kennen“ wie auch auf disziplinspezifische Lösungen „Fachliche Unterstützung z. B. bei Fragen zu Git oder auch in der Verwendung anderer Lösungen“.

Auch wenn die konkreten Aufgaben der Data Stewards an der RWTH Aachen University entsprechend der individuellen Kompetenzen und der an sie herangetragenen Erwartungen variieren, lassen sie sich zu folgenden Punkten zusammenfassen:

²¹<https://www.fuelcenter.rwth-aachen.de> (letzter Zugriff: 2020-09-17).

²²Hausen, Daniela, Jessica Rosenberg, Ute Trautwein-Bruns, und Annett Schwarz. „Data Stewards an Der RWTH Aachen University“.

- FDM-Ansprechpartner vor Ort (z. B. fachspezifischer Beratung und Schulung, individuelle Datenorganisation)
- Einführung und Verbesserung IT-gestützter Prozesse (z. B. ELN, Integrationsplattform)
- Vereinheitlichen und Standardisieren (z. B. Metadaten, Ontologien)
- Anleiten und Informieren (z. B. DMP-Vorlage, Policy, Archivierungskonzept, FDM-Newsletter)
- Entlastung der Forschenden von organisatorischem Overhead (z. B. Bibliographie)

4 Fazit

Es zeigt sich, dass die Aufgaben der Data Stewards an der RWTH Aachen University durch die Anpassung an die Anforderungen des Projekts und der jeweiligen Forschenden zu einem gewissen Grad variabel sind. Im Fokus der Aufgaben stehen jeweils der DMP sowie die Betreuung der Umstellung verschiedener manueller Prozesse auf IT-basierte Prozesse. Auch zeigt sich an den Erwartungen, dass ein wissenschaftsnaher Hintergrund für Data Stewards von Vorteil – aber kein Ausschlusskriterium²³ – ist. Es wird eine gezielte disziplinspezifische Unterstützung gewährleistet und disziplinspezifische Lösungen entwickelt, dessen Spektrum in hohem Maß von den konkreten Aufgaben und Wünschen seitens der Forschenden bestimmt wird.

Neben den Aufgaben und Erwartungen spielen auch die Kompetenzen der einzelnen Data Steward eine bedeutende Rolle. Das Beispiel des Data Stewards des SFB 1382 verdeutlicht sehr stark, dass Kommunikationskompetenz und die Kompetenz zum kollaborativen Arbeiten hervorstechen.

Das zentrale FDM-Team unterstützt Data Stewards in den Herausforderungen, indem es vernetzend agiert und generische Aspekte einbringt, um gemeinsam effektive Lösungen zu finden. Gleichzeitig schließen Data Stewards die fachspezifischen Lücken, die ein zentrales FDM-Team nicht abdecken kann. Durch Rückkopplung von fachspezifischen Umsetzungen können diese in die Beratung und Weiterbildung einfließen und so das Angebot an fachspezifischen Inhalten und Best Practices verbessern. Durch das Angebot des offenen FDM-Netzwerktreffens, das als Austausch- und Diskussionsplattform sehr gut angenommen wird, profitieren Data Stewards und FDM-Verantwortliche gegenseitig von ihren Erfahrungen. Erarbeitete Lösungskonzepte können durch andere Data Stewards oder FDM-Verantwortliche auf die individuellen Herausforderungen adaptiert und damit die Lösungen nachhaltiger und in der Breite nutzbar gemacht werden.

Für die RWTH Aachen University lässt sich ein erstes positives Zwischenergebnis ziehen. Forschende bekommen über geteilte Best Practices eine praxisnähere und dis-

²³Dies zeigt das Beispiel des Data Stewards des SFB 1382: <https://blog.rwth-aachen.de/forschungsdaten/2020/05/12/interview-mit-data-steward-lukas-c-bossert/> (letzter Zugriff: 2021-03-1).

ziplinspezifischere Beratung und Weiterbildung im FDM. Außerdem baut die stetige Präsenz des Data Stewards Hürden bei den Forschenden ab, sich mit dem Thema FDM auseinanderzusetzen. Data Stewards wiederum werden von zentraler Seite mit nötigen Informationen versorgt und dienen so als Multiplikator*innen.

Im Bestreben dauerhafte FDM-Strukturen im Einklang mit den FAIR-Prinzipien umzusetzen verfolgt die RWTH Aachen University den Ansatz wissenschaftsnaher, „embedded“ Data Stewards. Diese unterstützen die Forschenden konkret vor Ort und sind ein wichtiges Bindeglied im FDM-Netzwerk. Dabei sind die Arbeiten als Data Steward vielfältig und umfangreich, was sich am Beispiel des Data Stewards des SFB 1382 zeigt. Das Aufgabenspektrum ist weit mehr als eine „reine Datenverarbeitung“. Dies liegt nicht nur daran, dass das Thema Datenmanagement umfassend betrachtet werden kann oder soll, sondern auch daran, dass in Forschungsverbänden oft eine IT-affine Ansprechperson für die Forschenden ebenso fehlt wie eine Person, die sich um forschungsrelevanten Overhead kümmern kann, ohne dass es zulasten des eigenen Forschungsoutputs oder der eigenen Promotion geht. Eine Lücke, zu deren Schließung ein Data Steward maßgeblich beitragen kann.

Acknowledgements

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Sonderforschungsbereich 1382 „Darm-Leber Achse“ (DFG-Projektnr. 403224013).