



Fachhochschule Graubünden
University of Applied Sciences

Churer Schriften zur Informationswissenschaft

Herausgegeben von
Wolfgang Semar

Arbeitsbereich
Informationswissenschaft

Schrift 126

**Unterrichtskonzept eines Forschungs-
datenmanagement-Kurses für Mediziner
im Masterstudium an der Universität
Bern**

Mirjam Nydegger

Chur 2020

Churer Schriften zur Informationswissenschaft

Herausgegeben von Wolfgang Semar

Schrift 126

Unterrichtskonzept eines Forschungs- datenmanagement-Kurses für Mediziner im Masterstudium der Universität Bern

Mirjam Nydegger

Diese Publikation entstand im Rahmen einer Thesis zum Bachelor of Science FHGR in Information Science.

Referent: Gerhard Bissels

Korreferentin: Dr. Michelle Schaffer

Verlag: Fachhochschule Graubünden

ISSN: 1660-945X

Ort, Datum: Chur, Januar 2021

Kurzfassung

Das Forschungsdatenmanagement nimmt im wissenschaftlichen Prozess eine wichtige Stellung ein. In diesem Zusammenhang verlangen Forschungsförderer, wie der Schweizerische Nationalfonds, bei Gesuchseinreichung einen Datenmanagementplan. Auch für Studierende ist die Ausbildung von solchen Kompetenzen wichtig. Darum ist das Ziel dieser Arbeit das Erstellen eines Unterrichtskonzeptes mit Zielgruppe Medizinstudierende im Masterstudium der Universität Bern. Mittels Literaturrecherche und eines Experteninterviews wurden Anforderungen an den Kurs zusammengetragen, zu Lernzielen formuliert und anschliessend in der Sachanalyse behandelt. Zum Abschluss der Arbeit wurden Ideen zur Umsetzung skizziert. Die Auswertung der Anforderungen hat ergeben, dass Studierende der Medizin Grundlagen sowie disziplinspezifische Inhalte bezüglich des Forschungsdatenmanagements brauchen. Grund dafür ist die Arbeit mit gesundheitsbezogenen Daten, die besondere Anforderungen an das klinische Forschungsdatenmanagement und insbesondere an den Datenschutz, -sicherheit und die Nachnutzung stellen. Es konnten fünf Hauptthemen – Einführung; rechtliche Aspekte & Ethik; Datenmanagement & -organisation; Preservation; sowie Publikation, Re-Use & SNF – identifiziert werden, die jeweils mehrere Unterthemen beinhalten. Für die Umsetzung des Kurses gibt es verschiedene denkbare Methoden und Medien, die endgültige Umsetzung ist noch nicht geklärt.

Schlagwörter: Klinisches Forschungsdatenmanagement, Masterstudierende, Unterrichtskonzept

Abstract

Research data management plays an important role in the scientific process. In this context, research funders, such as the Swiss National Science Foundation, require a data management plan when submitting an application. The development of such skills is also important for students. Therefore, the aim of this thesis is to develop a teaching concept with a target group of medical students in the Master's program of the University of Bern. By means of literature research and an expert interview, course requirements were compiled, formulated as learning objectives and then treated in a factual analysis. At the end of the thesis, ideas for implementation were outlined. The evaluation of the requirements showed that medical students need basic knowledge as well as discipline-specific content regarding research data management. The reason for this is the work with health-related data, which places special demands on clinical research data management and in particular on data protection, security and re-use. Five main topics -

Introduction; Legal Aspects & Ethics; Data Management & Organization; Preservation; and Publication, Re-use & SNSF - could be identified, each of which contains several sub-themes. There are various possible methods and media for the implementation of the course, the final implementation is not yet clear.

Keywords: Clinical research data management, Graduates, Teaching concept

Vorwort

An dieser Stelle bedankt sich die Autorin bei den Betreuern, Herrn Gerhard Bissels und Frau Dr. Michelle Schaffer, für die geschätzte und hilfreiche Unterstützung bei dieser Arbeit. Weiter bedankt sich die Autorin bei ihren Eltern dafür, dass sie ihr ständig zur Seite stehen – während und neben dem Studium. Ein weiterer Dank gebührt der Alwina-Buchner-Stiftung für die finanzielle Unterstützung während des Bachelorstudiums.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problem-, Fragestellungen & Relevanz	2
1.2	Methodische Vorgehensweise & Gliederung	3
2	Stand der Forschung & wichtige verwendete Quellen	5
3	Analyse der Anforderungen – Inhalte ermitteln	9
3.1	Anforderungen aus der Literatur	9
3.2	Experteninterview	15
3.3	Fazit	18
4	Kurskonzept	23
4.1	Curriculum Medizin & Kompetenzen	23
4.2	Lernziele	26
4.3	Sachanalyse	29
4.3.1	Einführung	29
4.3.2	Rechtliche Aspekte & Ethik	36
4.3.3	Datenmanagement & -organisation	39
4.3.4	Preservation	54
4.3.5	Publikation, Re-Use & SNF	58
4.4	Überblick über die Vermittlung von Forschungsdatenmanagement-Kompetenzen an anderen akademischen Institutionen	64
4.4.1	Grundlagen und Forschungsdatenmanagement mit breitem Ansatz	65
4.4.2	Disziplinspezifisches Forschungsdatenmanagement	66
4.4.3	Fazit	67
4.5	Umsetzung – Methoden und Medien	67
5	Fazit	73
5.1	Beantwortung der Fragestellungen & Diskussion der Ergebnisse	73
5.2	Offene Fragen und Ausblick	76
6	Quellenverzeichnis	77
7	Anhänge	85
7.1	Ergebnisprotokoll Interview	85
7.2	Anhang: Tabelle – Aufstellung Vermittlung FDM	90

Abbildungsverzeichnis

Abbildung 1: Datenlebenszyklus, Quelle: Forschungsdaten.info, 2020h	34
Abbildung 2: eigene Abbildung, angelehnt an informatikZentrale (2010); Erklärung einer einfachen Datenbank	50

Tabellenverzeichnis

Tabelle 1: Inhalte Kurskonzept	22
Tabelle 2: Kompetenzen	25
Tabelle 3: Empfehlung Dateiformate	43

Abkürzungsverzeichnis

BASEC	Business Administration System for Ethics Committees
BORIS	Bern Open Repository and Information System
CCDM	Certified Clinical Data Manager
CD	Compact Disc
CDISC	Clinical Data Interchange Standards Consortium
CDM	Clinical Data Management
CDMS	Clinical Data Management System
CER	Comparative Effectiveness Research
CRDM	Clinical Research Data Management
CRF	Case Report Form
CSV	Comma Separated Value
CTU	Clinical Trial Unit
DCC	Digital Curation Centre
DLCM	Data Life-Cycle Management
DMP	Datenmanagementplan
DOI	Digital Object Identifier
DSG	Datenschutzgesetz (Schweiz)
DSGVO	Datenschutzgrundverordnung
DVD	Digital Versatile Disc
ECRIN	European Clinical Research Infrastructure Network
EDÖB	Eidgenössische Datenschutz- und Öffentlichkeitsbeauftragter
FAIR	Findable, Accessible, Interoperable and Reusable
FDM	Forschungsdatenmanagement
FFV1	FFmpeg Videocodec 1
GCDMP	Good Clinical Data Management Practices
GCP	Good Clinical Practice, Gute Klinische Praxis
GDPR	General Data Protection Regulation, DSGVO auf Deutsch
HFG	Humanforschungsgesetz
HTML	Hypertext Markup Language
JPEG	Joint Photographic Experts Group
KlinV	Verordnung über klinische Versuche in der Humanforschung
LZA	Langzeitarchivierung
MeSH	Medical Subject Heading
MPEG	Moving Picture Experts Group
MRT	Magnetresonanztomographie, MRI auf Englisch

NC	Non-Commercial
ND	No Derivatives
ORCID	Open Research Contributor Identification Initiative
PDF	Portable Document Format
PROFILES	Principal Relevant Objectives and Framework for Integrated Learning and Education in Switzerland
RCT	Randomized Controlled Trial
RDM	Research Data Management
REDCap	Research Electronic Data Capture
SA	Share alike
SCDM	Society for Clinical Data Management
SD	Secure Digital
SIARD	Software Independent Archival of Relational Databases
SNCTP	Swiss National Clinical Trials Portal
SNF	Schweizerischer Nationalfonds
SNSF	Swiss National Science Foundation
SOP	Standard Operation Procedures
SQL	Structured Query Language
TIFF	Tagged Image File Format
URL	Uniform Resource Locator
USB	Universal Serial Bus
WHO	World Health Organization
WLAN	Wireless Local Area Network
XML	Extended Markup Language

1 Einleitung

«Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process (Wilkinson et al., 2016, S. 1).»

Scaramozzino et al. (2012, S. 361) erklären weiter, dass Forschungsdaten einen intrinsischen Wert besitzen und deshalb das Teilen und Weitergeben solcher Daten nur logisch ist. Das sieht auch der Schweizerische Nationalfonds (SNF) so und fordert: «Forschungsdaten sollen offen und für alle zugänglich sein – für die Wissenschaft wie für die Gesellschaft (Schweizerischer Nationalfonds, o. J.-b).» Mit diesem Grundsatz verlangt der SNF seit Oktober 2017 nicht nur einen Data Management Plan (bei Einreichung des Forschungsgesuches), sondern auch, dass produzierte Forschungsdaten nach Abschluss der Arbeiten öffentlich zugänglich sind (Schweizerischer Nationalfonds, o. J.-a, o. J.-b). Hier kommt das Research Data Management (RDM) – oder Forschungsdatenmanagement (FDM) zu Deutsch – zum Zug: Es hilft Daten zu organisieren, strukturieren, speichern und zu pflegen (University of Oxford, 2013a). Solide Kenntnisse und Fähigkeiten im Verwalten von Forschungsdaten sind fundamental und zentral für jegliche Forschungsdisziplinen und bilden die Grundlage für herausragende Forschung. So sind Kurse für Forscher¹ zur Förderung und Ausbildung dieser Fähigkeiten an vielen Universitäten etabliert, solche für Studierende jedoch sehr selten (Wiljes & Cimiano, 2019, S. 1).

Der Umgang mit Daten ist aber nicht nur in der Forschung wichtig: Patientinnen von praktizierenden Ärzten nehmen zur Vergrößerung der Evidenzbasis ebenfalls an klinischen Studien teil (Novustat, 2020). Auch in diesem Szenario sind die Medizinerinnen mit Forschungsdaten und deren Handhabung konfrontiert. Laut Kuhn et al. (2018, S. 368) verändert aber auch die digitale Transformation den Beruf des Arztes je länger wie mehr: Ärztinnen stehen heute viel grössere Datenmengen zur Verfügung, die zur Entscheidungsfindung berücksichtigt werden müssen. Und weil die endgültige Therapieentscheidung noch immer beim Mensch liegt, benötigt der Mediziner Kompetenzen, «[...] um einerseits die Daten und andererseits die darauf beruhenden Entscheidungsfindungsprozesse verstehen und richtig einschätzen zu können (Kuhn et al., 2018, S. 368).» Es ist wichtig, diese Kompetenzen auszubilden, denn auf der

¹ Aus Gründen der Lesbarkeit wird jeweils alternierend – wenn eine geschlechtsneutrale Schreibweise nicht möglich ist – die männliche und weibliche Form verwendet.

Konsumebene sind die Digital Natives bestens mit digitalen Medien vertraut, berufsspezifische digitale Handlungskompetenzen sind aber kaum ausgeprägt (Kuhn et al., 2018, S. 368). Holles & Schmidt (2018) erklären treffend, dass das Forschungsdatenmanagement von Forschungsförderern zwar verlangt wird und somit einen Grund liefert, Kompetenzen in diesem Bereich auszubilden, dass aber ein bottom-up Ansatz zu effektiverem und effizienterem Fortschritt führt.

1.1 Problem-, Fragestellungen & Relevanz

Es stellt sich folgende Problemstellung: angehende Medizinerinnen der Universität Bern sollen – aus den zuvor aufgeführten Gründen – im Management von Daten geschult werden. Es wird aber bis dato noch kein Kurs für die Zielgruppe «Masterstudierende» angeboten. Mit der Erarbeitung eines Kurskonzeptes soll der Grundstein für die Aufnahme eines fakultativen Kurses für Masterstudierende der Medizin im Abschlussjahr zum Thema «Forschungsdatenmanagement» gelegt werden. Ziel dieser Arbeit ist, auf der Basis von Anforderungen, ein Kurskonzept zu entwickeln, das Inhalte und eine mögliche Umsetzung berücksichtigt. Die Inhalte sollen den angehenden Medizinerinnen Wissen und Werkzeuge mitgeben, um Forschungsdaten adäquat zu handhaben und sie im Umgang mit Daten zu sensibilisieren. Daraus ergeben sich folgende Fragestellungen:

Für das Kapitel Anforderungen (und die Sachanalyse):

«Welche Inhalte sind wichtig und relevant, um das Thema Forschungsdatenmanagement in einem Kurs mit Zielgruppe Masterstudierende im Medizinstudium zu vermitteln?»

Für die Kompetenzen und Lernziele:

«Wozu werden Inhalte zum klinischen Forschungsdatenmanagement vermittelt?»

Das heisst, es wird definiert, welche Qualifikationen und Kompetenzen die Studierenden durch den Kurs erlangen und welche Ziele erreicht werden sollen.

Für das Kapitel zu Kursen an anderen akademischen Institutionen:

«Wie werden den Studierenden Kompetenzen im Forschungsdatenmanagement an anderen akademischen Institutionen vermittelt?»

Für die Umsetzung:

«Wie und womit soll das Thema klinisches Forschungsdatenmanagement vermittelt werden?»

Das Kurskonzept dient der Weiterverwendung für Informationsspezialisten sowie medizinische Bibliothekarinnen, die das Konzept an einen Lehrplan oder ein Schulungsangebot adaptieren möchten. Das Konzept wird in erster Linie für die medizinische Bibliothek der Universität Bern erstellt.

Zur Eingrenzung der Thematik legt das Kurskonzept den Fokus auf die Zielgruppe Medizinstudierende, daher auch die Eingrenzung auf das klinische Forschungsdatenmanagement. Forschungsdaten der Medizin haben spezifische zusätzliche Anforderungen (Anonymisierung, sensible Daten etc.) an das Datenmanagement. Darum ist es nicht möglich einen gängigen Forschungsdatenmanagement-Kurs eins zu eins zu übernehmen. Weiter bildet das Konzept nur Inhalte ab, die von Mitarbeitern einer Bibliothek unterrichtet werden können. Das heisst Themen, wie Ethik und Gesetzgebung, die nicht zu den Kernkompetenzen von Bibliothekarinnen gehören, werden nur beleuchtet, beziehungsweise auf die Kernpunkte reduziert.

1.2 Methodische Vorgehensweise & Gliederung

Als Methode zur Erarbeitung des Konzeptes wurde die Literaturrecherche gewählt. Zur Ergänzung wird ein Interview mit Muriel Helmers, Head of Data Management der Clinical Trial Unit (CTU) der Universität Bern geführt. Muriel Helmers arbeitet in der Abteilung für klinische Studien der medizinischen Fakultät des Universitätsspitals Bern und der Universität Bern und ist dort für die Qualität der Forschungsdaten und Aufsetzung der Datenbanken zuständig. Das Interview fand am 11. Juni 2020 via Zoom-Meeting statt. Ihre Inputs bezüglich spezieller Bedürfnisse des Datenmanagements von medizinischen Forschungsdaten waren sehr wertvoll für die vorliegende Arbeit.

Das Konzept ist im Hauptteil grob in zwei Teile aufgeteilt: Anforderungen an Inhalte und Kurskonzept, wobei bei letzterem der Fokus auf den Inhalten, sprich der Sachanalyse, liegt. Das heisst, in einem ersten Schritt wird erfasst, welches Wissen die Zielgruppe über das Forschungsdatenmanagement braucht. Danach werden daraus die Lernziele abgeleitet und auf dieser Basis wird die Sachanalyse zur Vermittlung der Lerninhalte entwickelt. Im Kapitel über die Anforderungen wird analysiert, welche Themen im Hinblick auf die Zielgruppe generell im Forschungsdatenmanagement und im Speziellen bei klinischen Daten Wichtigkeit und Relevanz besitzen. Dazu werden Publikationen, wie Case Studies, Bedarfsanalysen und Umfragen hinzugezogen. Zum Thema Forschungsdatenmanagement gibt es sehr viele Artikel. Um die Auswahl und Schwerpunkte der zu behandelnden Themen nachvollziehen zu können und nicht willkürlich festzulegen, wird eine systematische Recherche gemacht. Diese in dieser

Recherche gefundenen Quellen dienen als Ausgangslage und Basis für die Analyse der Anforderungen. Zusätzlich wird mittels Rückwärtszitationssuche (Pearl Growing) die Literatur ergänzt. Diese Analysen werden dann im eigentlichen Konzept zu Lernzielen operationalisiert. In der Sachanalyse werden die wichtigsten Themen und Aspekte, sowie die spezifischen Anforderungen an das klinische Forschungsdaten-management aufbereitet. Nach der Sachanalyse folgt ein kurzes Kapitel zu Forschungsdatenmanagement-Kursen an anderen akademischen Institutionen, um einordnen zu können, was andere Hochschulen und Universitäten in diesen Bereich leisten. Zum Abschluss des Konzeptes wird die Umsetzung – konkreter die Methoden und Medien – angeschnitten. Dieser Teil der Arbeit soll nur beschreibend sein und mögliche Lösungen zur Umsetzung anbieten.

2 Stand der Forschung & wichtige verwendete Quellen

Die vorliegende Arbeit beschäftigt sich mit interdisziplinärer Thematik. Im Informationswissenschaftlichen Bereich ist das Datenmanagement zentral. Daher ist der Forschungsstand zum Thema Forschungsdaten (=Research Data) und Forschungsdatenmanagement (=Research Data Management) insbesondere für die Inhalte von Bedeutung. Es gilt die Websites Dataone.org (DataONE, o. J.-b), Research Data Oxford (University of Oxford, o. J.), forschungsdaten.info (Forschungsdaten.info, 2020o) und MANTRA (The University of Edinburgh, 2020a) zu nennen. Weiter beschäftigt sich das Konzept auch mit der Medizin: Die Yale University stellt Wissen, im Speziellen für das Forschungsdatenmanagement in der Medizin, zur Verfügung (Yale Library, 2020g). Weitere Quellen sind der Buchteil «Research Data Stewardship for Healthcare Professionals» von Jansen et al. (2019) und das Buch «Practical Guide to Clinical Data Management» von Prokscha (2007). Laut Read (2019, S. 89) wurde das Forschungsdatenmanagement für klinische Daten in der Ausbildung und Lehre bisher kaum ins Visier genommen:

«[...] clinical research community [...] is one that has not received much attention. Clinical research data management (CRDM), an integral part of the clinical research process, differs from the broader concept of research data management because it involves rigorous procedures for the standardized collection and careful management of patient data to protect patient privacy and ensure quality and accuracy in medical care (Read, 2019, S. 90).»

Read erläutert weiter, dass zum klinischen Forschungsdatenmanagement aber bereits etablierte Richtlinien bestehen. Mit diesen Richtlinien sind die Good Clinical Data Management Practices (GCDMP) der Society for Clinical Data Management (SCDM) gemeint. Neben diesen Richtlinien, die sich mit klinischen Daten auseinandersetzen, gibt es noch weitere nationale und internationale Leitfäden und Prinzipien, wie die «Gute Klinische Praxis» oder die «FAIR-Prinzipien». Letztere beziehen sich auf Forschungsdaten im Allgemeinen und wurden von Wilkinson et al. (2016) kommentiert.

Ein Hinweis darauf, dass das Forschungsdatenmanagement – zumindest in der Literatur – diskutiert wird, liefern Perrier et al. (2017): Sie haben mit dem Ziel Umfang, Themen und methodischen Charakter zu identifizieren 301 Artikel zum Thema Forschungsdatenmanagement in akademischen Institutionen untersucht. Die ersten Publikationen stammen aus den 90er-Jahren, wobei aber ab 2010 die meisten Artikel (85%) veröffentlicht wurden. Es wurden elf Studientypen identifiziert, davon sind fast 80% Querschnittsstudien, Interviews und Case Studies. Der Medizin und der

Informationswissenschaft konnten je 31 Artikel zugeordnet werden – damit sind diese beiden Disziplinen am meisten vertreten. Die Autoren dieses Scoping Reviews konnten ein Muster erkennen und die Artikel in fünf Gruppen (Stakeholder, Daten, Bibliothek, Tools und Publikation) einteilen, die wiederum Subgruppen enthalten. Diese Gruppen zeigen, mit welchem Thema sich die Artikel beschäftigen. Es wurde herausgefunden, dass das Thema Datenqualität besonders beschäftigt (Perrier et al., 2017, S. 5, 9 & 11).

Ein Angebot sollte auf die Nachfrage abgestimmt sein, darum muss das Konzept auf Anforderungen und Bedürfnisse eingehen: Goben und Griffin (2019) haben über 50 Bedarfsanalysen zum Thema Forschungsdatenmanagement zusammengefasst und dabei institutsübergreifende Trends, Themen und Probleme erfasst. Wiljes & Cimiano (2019) erklären in ihrem Artikel «Teaching Research Data Management for Students», dass das Forschungsdatenmanagement in der akademischen Ausbildung ein relativ neues Forschungsfeld darstellt. Weiter befürworten sie, dass das Forschungsdatenmanagement bereits früh in der akademischen Ausbildung gelehrt werden sollte. Darum wird an der Universität Bielefeld seit 2013 ein Kurs zum Forschungsdatenmanagement angeboten. Die Evaluierung des Kurses hat unter anderem ergeben, dass eine praktische Herangehensweise besser geeignet ist als abstrakte Wissensvermittlung. Es wurde zudem festgestellt, dass Kurse für Forscherinnen an vielen Universitäten fester Bestandteil, Angebote für Studierende aber sehr selten sind. Das haben auch Kuhn et al. (2018) erkannt und haben für die Universität Mainz einen Kurs zum Thema «Data Literacy» entwickelt, der seit dem Herbstsemester 2017 läuft. Grund des Problems sehen sie darin, dass digitale Kompetenzen weder im Masterplan Medizinstudium 2020 noch im Nationalen Kompetenzbasierten Lernzielkatalog Eingang finden. Auch in der Schweiz sind digitale Kompetenzen nicht wirklich in den Lernzielen – beispielsweise in «Principal Relevant Objectives and Framework for Integrated Learning and Education in Switzerland» (PROFILES von Michaud et al., 2017) Lernziele 6.3 bis 6.5 – verankert, sondern nur weitläufig umschrieben.

Ein wichtiger Teil des Konzeptes machen die Lernziele aus, die es zu formulieren gilt. Eine Anlaufstelle ist das Werk von Robert F. Mager (1962) «Preparing instructional objectives». Weiter können die PROFILES hinzugezogen werden, die von Michaud et al. (2017) verfasst wurden.

Zum Abschluss und allgemeinen Verständnis: Data Literacy (Datenkompetenz) ist nicht gleichgestellt mit dem (Forschungs-)Datenmanagement. Laut Ridsdale et al. (o. J., S. 8) bezeichnet Datenkompetenz «[...] the ability to collect, manage, evaluate, and apply data, in a critical manner.» Ridsdale et al. (o. J.) haben fünf Kompetenzfelder identifiziert:

Konzeptrahmen (Einführung in Daten); Datenerhebung und -sammlung (Recherche, Qualität der Quellen und Daten); Datenmanagement (Organisation, Metadaten, Re-Use, Sicherheit, Preservation); Datenevaluierung (Analyse, Interpretation, Visualisierung, Entscheidungsfindung); Datenanwendung (Ethik, Zitation, Sharing). Das Datenmanagement ist also Teil der Datenkompetenz, behandelt aber die Themen Analyse, Visualisierung, Datensammlung, -recherche, -erhebung und -evaluierung nicht oder nur teilweise.

3 Analyse der Anforderungen – Inhalte ermitteln

In diesem Kapitel werden die Anforderungen an einen Forschungsdatenmanagement-Kurs ermittelt und dargelegt. Es wird versucht folgende Frage zu beantworten: «Welche Inhalte sind wichtig und relevant, um das Thema Forschungsdatenmanagement in einem Kurs mit Zielgruppe Masterstudierende im Medizinstudium zu vermitteln?». Es soll aufgezeigt werden, welche Themen wichtig sind und abgedeckt werden müssen, damit Medizinstudierende im Master das nötige Rüstzeug haben, um mit Forschungsdaten umgehen zu können. Anhand dieser Anforderungen werden die Inhalte ermittelt und so die Lernziele abgeleitet. Für das Zusammentragen der Anforderungen wurden über 25 Quellen, bestehend aus Artikeln und Websites zu Rate gezogen und untersucht. Zu Ergänzung wurde ein Interview mit Muriel Helmers, Head of Data Management bei der CTU an der Universität Bern, geführt, um die Eigenheiten von klinischen Daten herauszuarbeiten und zu ermitteln, in welchen Bereichen des Forschungsdatenmanagements auf spezifische Anforderungen der klinischen Daten hingewiesen werden muss.

3.1 Anforderungen aus der Literatur

Wie im Forschungsstand schon aufgezeigt, existieren bereits einige Publikationen zum Forschungsdatenmanagement. Es gibt aber nur wenig Literatur zu Bedarfsanalysen und keine einzige zu Bedürfnissen und Anforderungen an das klinische Forschungsdatenmanagement von Medizinstudierenden. Aus diesem Umstand wurde die Suche auf weitere Publikationstypen und Suchbegriffe ausgedehnt, so, dass die Anforderungen umfassend erläutert werden können. Wie Perrier et al. (2017) konnte die Autorin bei der Literaturrecherche ebenfalls erkennen, dass sich die Publikationen in die im Forschungsstand kurz vorgestellten Gruppen (Stakeholder, Daten, Bibliothek, Tools und Publikation) einteilen lassen.

Auch Goben & Griffin (2019, S. 913) haben erkannt, dass bereits unzählige Case Studies an Universitäten gemacht wurden, um die Bedürfnisse an das FDM zu untersuchen. Weiter stellten sie aber auch Forschungslücken fest: Über die Bedürfnisse und Ansprüche von Mitarbeiterinnen und Postdoktoranden, die nicht einer Fakultät zugeordnet sind, sowie Master- (Graduates) und Bachelorstudierende (Undergraduates) wurde am wenigsten geforscht. Gerade die letzten beiden Personengruppen haben aber am meisten Kontakt mit Forschungsdaten (Goben & Griffin, 2019, S. 913) und sind für das vorliegende Konzept zentral. Carlson et al. (2011, S. 8) bestärken diesen Punkt: «Graduate students actively generate and curate data in support of their own research.

Many also oversee the management of data generated by the entire research group.» Es kann aber davon ausgegangen werden, dass die Anforderungen (von Studierenden und Forschern) grundsätzlich ähnlich sind, die Studierenden aber einen etwas bereiteren Ansatz benötigen, als die Forschenden, die in eng gefassten Bereichen forschen, schon sehr erfahren im Umgang mit der Forschung sind und genau wissen, was ihre Informationsbedürfnisse sind (Wiljes & Cimiano, 2019, S. 8). Das heisst, Studierende benötigen neben dem Fokus auf die Studienrichtung – im Fall des vorliegenden Konzeptes die Medizin – erst eine Grundlage, auf der dann aufgebaut werden kann.

Laut Carlson (2014, S. 64) können – von einer höheren Ebene aus betrachtet – zwei verschiedene Arten des Datenmanagements unterschieden werden: das Datenmanagement und die Data Curation. Das Datenmanagement bezieht sich hierbei auf die aktive Phase des Data Life Cycles, wenn Forschende Daten erheben, generieren und analysieren. Während dieser Phase sind das Entwickeln eines Datenmanagementplans (DMP), die Ausbildung von Studierenden und Forscherinnen in den Bereichen Dokumentation und Organisation, sowie die Handhabung von FDM-Tools und Ressourcen zentral. Das Ziel ist die Daten auf eine sichere Art und Weise zu speichern (Carlson, 2014). Data Curation wird wie folgt definiert:

«Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities across the sciences, social sciences, and the humanities. Data curation activities enable data discovery and retrieval, maintain data quality, add value, and provide for re-use over time. This new field includes representation, archiving, authentication, management, preservation, retrieval, and use (Center for Informatics Research in Science and Scholarship & University of Illinois, o. J.).»

In dieser Phase sind die Forschungsdaten oft einem grösseren Publikum und Forschern ausserhalb der ursprünglichen Forschungsgruppe (öffentlich) zugänglich. Die Daten selber werden nicht weiterentwickelt, aber angereichert, um den Wert derselben zu steigern. Wie im Zitat oben schon vermerkt, gehört das Anreichern mit Metadaten – um die Daten besser zugänglich und verständlicher zu machen, sowie in Kontext zu stellen – zu den Aufgaben des Kuratierens von Daten. Dazu gehören auch das Sicherstellen der Auffindbarkeit und Zitierfähigkeit durch das Zuweisen von eindeutigen und permanenten Identifikatoren, wie beispielsweise dem Digital Object Identifier (DOI). Darüber hinaus wird in dieser Phase der langfristige Zugriff durch die Entwicklung von Richtlinien und Plattformen gesichert. Die Massnahmen, die der Phase der Data Curation zugeschrieben werden, können aber auch bei der aktiven Verwaltung der Daten ergriffen werden

(Carlson, 2014). Henderson & Knott (2015, S. 54) finden es wichtig und lohnenswert, wenn Data Curation dazu beiträgt Daten wertvoller, besser zugänglich und wiederverwendbar zu machen – beispielsweise durch die Aktualisierung der Dateiformate – priorisieren diese Massnahmen aber nicht, weil der Aufbau dieser Massnahmen viel Aufwand in Personal und Technologie benötigt. Hier gilt es noch anzumerken, dass die Universität Bern einerseits bereits über Infrastruktur und Know-How verfügt (beispielsweise CTU der Universität Bern) und das andererseits für die Ausbildung von Kompetenzen im Forschungsdatenmanagement kein valides Argument darstellt. Ebenfalls nicht zu den Top-Prioritäten gehören nach der Meinung von Henderson & Knott (2015, S.55) Open Access, Literaturverwaltung/ Zitationsregeln und Wissen über Copyright. Henderson & Knott (2015, S. 54) begrüßen aber Kurse im FDM für Studierende und insbesondere das Aufsetzen eines Datamanagementplans, um sie auf zukünftige Forschungsprojekte und damit einhergehende Anträge für Fördergelder vorzubereiten. Ausserdem hilft das FDM beim Verwalten der eigenen Recherchen für Projekte und Arbeiten (Henderson & Knott, 2015).

Goben & Griffin (2019) haben über 50 Fallstudien aus den Jahren 2007 bis 2017 von Bibliotheken zur Bedarfsanalyse von FDM in Bildungsinstitutionen untersucht. Zu den Teilnehmenden der einzelnen Bedarfsanalysen zählten Fakultäten und Mitarbeiter derselben, Master- und Bachelorstudierende, Postdoktorandinnen und Mitarbeiter der Administration. Goben & Griffin (2019, S. 913) haben erkannt, dass die Speicherung (Storage), Nachnutzung (Sharing) und Fragen, die sich um den langfristigen Zugang zu Daten drehen – wie Dateiformate, Backup und Preservation – die Themen sind, die die meisten Sorgen bereiten. Insbesondere Bedenken in Hinblick auf das Sharing wurden oft genannt. Mittleres Interesse wurde beim Datenmanagementplan (DMP), der Datensicherheit, Datenorganisation, dem Datenschutz und der für die Datenverwaltung verantwortlichen Partei bekundet. Das Bedürfnis «Datenmanagementplan» wurde in fast 50% der Artikel genannt und ist auch nach über einem Jahrzehnt immer noch oder wieder vermehrt aktuell. Wenig Interesse erzeugten die Themen Datenlebenszyklus, Daten finden und Policy/Richtlinien. Gerade zum letzten Aufzählungspunkt gilt es anzumerken, dass dieses Thema erst ab 2015 auftauchte: zuvor wurde Data Policy nicht als relevant aufgefasst (Goben & Griffin, 2019). Darum und weil die Anforderungen an Fördermittel des Bundes gestiegen sind, sollten Data Policy und der DMP nicht leichtfertig abgetan werden. Holles & Schmidt (2018) haben in ihrem Forschungsdatenmanagement-Kurs für Masterstudierende acht Bereiche identifiziert und ausgearbeitet: Datenmanagementplan; Archivierung und Preservation; Sharing und Reuse; datenrechtliche und ethische Belangen; Dokumentation und Metadaten; Speicherung (Storage), Back-up und

Sicherheit; Datenorganisation sowie Datentypen und -formate. Die Befragung der teilnehmenden Studierenden hat ergeben, dass das Thema Ethik vertieft behandelt werden darf. Laut Holles & Schmidt (2018) ist vorgesehen, Datenmanipulation und Ethik im Zusammenhang mit dem Sharing in die nächsten Kurse mitaufzunehmen. Die bestehenden acht Bereiche wurden von den Studierenden bestätigt.

Aus weiteren Quellen konnten ähnliche Resultate entnommen werden, wobei einige Publikationen das Thema «Data Literacy» behandelt haben (Calzada Prado & Marzal, 2013; Carlson et al., 2011, 2015; Federer et al., 2016). Weil aber das Datenmanagement Teil von Data Literacy ist (Ridsdale et al., o. J.), konnten die Ergebnisse aus diesen Artikeln auch – zumindest teilweise – hinzugezogen werden. Obwohl die verschiedenen Umfragen, Case Studies und Berichte zu Kursen unterschiedliche Zielgruppen behandeln/ adressierten – beispielsweise Forscherinnen medizinischer Zentren, Bibliothekare, die in FDM geschult werden, sowie Studierende und Mitarbeiterinnen von Fakultäten aus Medizin und anderen Forschungsgebieten – und die Inhalte oder zumindest die Begriffe dazu sich differenzieren, konnten trotzdem Themen ausfindig gemacht werden, die immer wieder genannt werden. Bereiche, die vermehrt auftauchen, sind:

- Einführung in Daten (Begrifflichkeiten), Datenbanken, Datenformate und -typen, Datenlebenszyklus, sowie Grundlagen des Forschens (Carlson et al., 2011, 2015; Piorun et al., 2012; Read et al., 2019; Whitmire, 2015; Wiljes & Cimiano, 2019)
- Datenmanagement und Organisation, inkl. DMP und File-Naming (Carlson et al., 2011, 2015; Federer et al., 2016; Piorun et al., 2012; Read, 2019; Read et al., 2019; Whitmire, 2015; Wiljes & Cimiano, 2019)
- Metadaten, Standards und Datenqualität (Calzada Prado & Marzal, 2013; Carlson et al., 2011, 2015; Federer et al., 2016; Piorun et al., 2012; Read, 2019; Read et al., 2019; Whitmire, 2015; Wiljes & Cimiano, 2019)
- Preservation, Storage/ Speicherung, Kuration inkl. Back-up und Archivierung (Carlson et al., 2011, 2015; Federer et al., 2016; Piorun et al., 2012; Read, 2019; Read et al., 2019; Whitmire, 2015; Wiljes & Cimiano, 2019)
- Sicherheit und Schutz inkl. Datenschutz, sensible Daten, Ethik und Richtlinien (Carlson et al., 2011, 2015; Piorun et al., 2012; Wiljes & Cimiano, 2019)
- Publikation, Sharing, Re-Use inkl. Copyright und Lizenzen (Carlson et al., 2011, 2015; Federer et al., 2016; Piorun et al., 2012; Whitmire, 2015; Wiljes & Cimiano, 2019)

Weitere Themen, die nicht so häufig auftreten, sind: Visualisierung, Analyse, Interpretation, Evaluierung, Datenerhebung/ -recherche und -sammlung, Open Science/ Open Access, Data Mining sowie Verantwortlichkeiten und Rollen (Calzada Prado & Marzal, 2013; Carlson et al., 2011, 2015; Federer et al., 2016; Read, 2019; Surkis et al., 2017; Whitmire, 2015; Wiljes & Cimiano, 2019). Die Visualisierung, Evaluierung und Interpretation von Daten sowie die Datenerhebung/ -recherche und -sammlung kamen nur in Publikationen vor, die explizit von Data Literacy, aber nicht von FDM gehandelt haben.

Bis hierhin wurden Themen und Bereiche zusammengetragen, die für das FDM allgemein Bedeutung und Relevanz haben. Im Folgenden sollen nun Anforderungen an das FDM im Speziellen für medizinische, beziehungsweise klinische Daten aufgezeigt werden. Es gilt, die unterschiedlichen Bedürfnisse an das FDM von verschiedenen Forschungsfeldern zu beachten: die Medizin hat andere/ weitere Bedürfnisse als beispielweise die Kunstwissenschaft. Akers & Doty (2013) stellen zu diesem Thema ihre Umfrage im Artikel «Disciplinary Differences in Faculty Research Data Management Practices and Perspectives» vor. Es wurden 330 Forschende befragt. Um einen effektiveren und verständlicheren Ansatz bieten zu können, ist es wichtig, die Gemeinsamkeiten und Unterschiede von verschiedenen Forschungsdisziplinen herauszuarbeiten (Akers & Doty, 2013, S. 17). So kann der Unterricht im FDM auf die Zielgruppe zugeschnitten werden.

«Different disciplines vary widely in their research funding, technical infrastructures, collaboration networks, source materials, subject populations, methodologies, ethical considerations and types of research outputs. Therefore, to be most effective, data curation requires discipline-specific approaches (Akers & Doty, 2013, S. 14).»

Akers & Doty (2013) haben die Befragten in die Kategorien Kunst und Geisteswissenschaften, Sozialwissenschaften, Medizin sowie Grundlagenforschung eingeteilt. Dadurch war es ihnen möglich, wichtige disziplinäre Unterschiede im Datenmanagement bezüglich Massnahmen, Einstellung und Haltung sowie Interesse in Unterstützungsdiensten der Bibliothek aufzudecken. Im Kontext dieses Artikels gehört Forschung, die in einem klinischen Umfeld oder anderweitig «in der Natur angewandt» wurde, zur Medizin (Akers & Doty, 2013, S. 7). Die Resultate zeigen, dass Forscherinnen aus der Medizin und der Sozialwissenschaft am wenigsten häufig Daten mit anderen Personen ausserhalb der Projektgruppe teilen, weil diese persönliche oder vertrauliche Information enthalten. Ausserdem verlangen die Daten einen sicheren und

eingeschränkter Zugang. Die Umfrage hat ausserdem ergeben, dass über 60% der befragten Personen nicht mit der Dokumentation und/oder dem Konzept Metadaten vertraut sind. Zwischen den verschiedenen Disziplinen gab es keine markanten Unterschiede (Akers & Doty, 2013, S. 12). Das lässt darauf hindeuten, dass die Dokumentation und Metadaten Bereiche sind, die in der Ausbildung noch gefördert werden müssen.

Im Zusammenhang mit disziplinären Unterschieden weisen Henderson & Knott (2015, S. 49) auf die Website datacurationprofiles.org hin. Dort werden für definierte Kriterien/Themen die spezifischen Anforderungen der verschiedenen Forschungsgebiete dargelegt. Hier gilt es noch anzumerken, dass das Digital Curation Centre (DCC) ebenfalls solche Profile verschriftlicht hat (Digital Curation Centre, 2020b). Auf beiden Websites gibt es für klinische beziehungsweise medizinische Daten kaum relevantes Material. Nur auf der Digital Curation Centre Website ist für die klinische Medizin ein spezifischer Metadatenstandard aufgelistet (Digital Curation Centre, 2020a), aber bereits dokumentierte Eigenheiten der klinischen Forschungsdaten sind leider keine vorhanden.

Zozus et al. (2017) haben eine Umfrage mit über 230 Clinical Research Data Managers durchgeführt, um die Kompetenzen, die es in diesem Beruf braucht, einschätzen und verfeinern zu können. Die Umfrage wurde auf Grundlage der bereits bestehenden Good Clinical Data Management Practices (GCDMP) und der Prüfung zum Certified Clinical Data Manager (CCDM) erstellt. Die Resultate aus dieser Umfrage werden zur Revision der Guidelines und der Prüfung genutzt, werden aber auch als Quelle zur Entwicklung und Auffrischung von Bildungsprogrammen zur Verfügung gestellt. Hinter den GCDMP und der Prüfung zum CCDM steht die Society for Clinical Data Management (SCDM). Die Befragten haben 91 verschiedene Kompetenzen beziehungsweise Aufgabe bestätigt und zehn neue Kompetenzen sowie acht neue mögliche Bereiche im Grundlagenwissen identifiziert. Das Grundlagenwissen, wie beispielsweise Kenntnisse in medizinischen Fachbegriffen sowie von allgemeinen Regeln des Forschens, Konzepten von Datenlebenszyklen und medizinischen (Metadaten-)Standards, bilden zusammen mit Praxisnormen, wie GCDMP oder dem Humanforschungsgesetz (HFG) – das in der Schweiz Gültigkeit besitzt – die Basis für die Kompetenzen. Die 91 Kompetenzen wiederum können in acht Kategorien eingeteilt werden: Design; Ausbildung/Training; Datenverarbeitung; Programmierung; Erprobung/Testing; Personalmanagement; Koordination & Projektmanagement sowie Überprüfung/Review. Der Vergleich von Datenmanagern, die sich an unterschiedlichen Stellen der Karriere befinden, zeigt Kompetenzverschiebungen auf. Zu Beginn der Karriere sind Kompetenzen und Aufgaben in den Kategorien Datenverarbeitung und Erprobung/Testing häufiger als bei

fortgeschrittener Karriere. Dafür werden mit der Zeit Kompetenzen in Ausbildung/Training, Personalmanagement, Koordination & Projektmanagement sowie Überprüfung/Review immer wichtiger (Zozus et al., 2017, S. 738–743). Zozus et al. (2017, S. 738) haben drei wichtige Anwendungszusammenhänge definiert: der Studientyp für welchen die Forschungsdaten verwaltet werden; das Umfeld, in dem die Daten gemanagt werden und die Datentypen, die verwaltet werden. Zu den Studientypen zählen unter anderem randomisierte kontrollierte Studien (=randomized controlled trial, RCT), Register und Comparative Effectiveness Research (CER). Unter Datentypen werden beispielsweise klinische Daten aus dem Prüfungsbogen (=Case Report Form, CRF), Daten von medizinischen oder privaten Geräten, öffentliche Daten, wie die Todesrate, sowie klinische Bilder verstanden. Diese Anwendungszusammenhänge stellen nochmals bestimmte Anforderungen an das Forschungsdatenmanagement, wobei aber keine signifikanten Differenzen zwischen den Antworten aus der Industrie und der Akademie bezüglich Forschungsumfeld gefunden wurden. Um eine angemessene Datenqualität und -konsistenz gewährleisten zu können, egal, ob in der Industrie oder der Akademie, soll man sich auf allgemeine Regeln, wie der Nachvollzieh- und Reproduzierbarkeit, des Forschens stützen (Zozus et al., 2017, S. 744). Weil diese Umfrage zwar ganz gezielt das klinische Forschungsdatenmanagement abdeckt, jedoch die Zielgruppe «Expertinnen» befragt, können die erhobenen Anforderungen beziehungsweise Kompetenzen und Grundlagenwissen nur bedingt und ansatzweise für das vorliegende Konzept hinzugezogen werden. Darum werden im Folgenden nur die meistgenannten Kompetenzen der Kategorien Design, Datenverarbeitung und Erprobung/ Testing aufgezählt. Diese Kategorien wurden ausgewählt, weil Berufseinsteiger vor allem mit diesen Aufgaben zu tun haben und somit Studierende am ehesten damit konfrontiert werden. Zu den meist genannten Kompetenzen und Aufgaben zählen die Erstellung und Überarbeitung eines Case Report Form (CRF), das Schreiben eines Managementplans, das Ausarbeiten und Testen von Edit Checks (Validierung der Daten), das Identifizieren von Diskrepanzen und das Eingeben von Daten, sowie deren Abstimmung, Verfolgung/ Überprüfung und Codierung (Zozus et al., 2017, S. 740).

3.2 Experteninterview

Muriel Helmers, Head of Data Management, arbeitet bei der Clinical Trial Unit (CTU) an der Universität Bern und konnte wichtige Hinweise zu den Eigenheiten und Besonderheiten von klinischen Daten liefern (Ergebnisprotokoll siehe Anhang 7.1). Sie ist verantwortlich für das Aufsetzen der Studiendatenbanken. Ausserdem führt die CTU

Beratungen zur Datensammlung und Erstellung von Datenbanken unter Einhaltung der gesetzlichen Rahmenbedingung durch.

Das besondere an klinischen Daten ist laut Muriel Helmers (persönliche Kommunikation, 11. Juni 2020), dass das Umfeld sehr reguliert ist, «weil mit Daten von Menschen gearbeitet wird.» Daraus resultiert, dass insbesondere der Datenschutz und die Datensicherheit grosse Themen sind. Aus diesem Grund wird im Umfeld der Studie auch mit codierten Daten gearbeitet, das heisst, die klinischen Daten werden so codiert, dass nicht erkennbar ist, um welchen Patienten es sich bei einem Datensatz handelt. Genauer bedeutet das, dass die Forschungsdaten nicht mit den persönlichen Daten zusammen aufbewahrt werden. Eine Software generiert einen Code, der das Verbindungsglied zwischen Forschungsdaten (Datensatz) und den persönlichen Daten darstellt. Ohne diesen Code können die Forschungsdaten nicht dem Patienten zugeordnet werden. Weiter ist definiert, dass die Datenbanken Login, Passwort und Rollenmodelle brauchen, um die Forschungsdaten von unberechtigten Zugriffen zu schützen. «Es dürfen [...] nur die Personen Zugriff auf die Datenbank haben, die die Daten für ihre Arbeit benötigen, alle anderen haben keinen Zugriff (M. Helmers, persönliche Kommunikation, 11. Juni 2020).»

Begründet ist der Schutz der personenbezogenen Daten auf Datenschutzgesetzen (DSGVO, DSG), der ärztlichen Schweigepflicht und dem Humanforschungsgesetz (HFG):

«Im Datenschutzgesetz ist [...] festgelegt, was gesammelt werden darf. Ausserdem geht es darum, dass niemand benachteiligt wird, bspw. Versicherungen, die den Patienten nicht mehr annehmen, oder Patienten, die keinen Job mehr erhalten (M. Helmers, persönliche Kommunikation, 11. Juni 2020).»

Und weiter: «Diese Daten dürfen nicht an die Öffentlichkeit.» Laut Artikel 4 des HFG haben Interesse, Gesundheit und Wohlergehen des einzelnen Menschen Vorrang gegenüber den Interessen der Wissenschaft und der Gesellschaft. Ausserdem muss laut Muriel Helmers die zuständige Ethikkommission jedes Studienprojekt bewilligen. Es dürfen nur die Daten gesammelt werden, die auch tatsächlich gebraucht werden, zudem muss die Datensammlung begründet sein. Im Studienprotokoll wird festgehalten, wann welche Variablen gesammelt werden. Überdies wird eine Risikoanalyse gemacht. Muriel Helmers sagt dazu weiter: «Der Benefit [der Studie] soll grösser sein als das Risiko.» Die Gesetzeslage sieht vor, dass die Patienten hinreichend über das Forschungsprojekt informiert werden und der Teilnahme schriftlich einwilligen, ansonsten ist ein Einbezug dieser Person ausgeschlossen (HFG Art. 16). Muriel Helmers hat erklärt, dass eine

solche Einwilligung «Consent» genannt wird. Es gibt zwei Formen: den general Consent – der wird bei Spitaleintritt unterschrieben und betrifft Routedaten – und den informed Consent. Der informed Consent muss bei klinischen Studien eingeholt werden.

Weitere Bereiche, die im klinischen Forschungsdatenmanagement besonders berücksichtigt werden sollten/müssen, sind: Standards/Dateiformate, Sharing, das Handling unterschiedlicher Datentypen, Langzeitaufbewahrung und Open Science/ SNF.

Nach Abschluss einer Studie werden die Forschungsdaten (Datensätze in einer Datenbank) und die Metadaten in verschiedenen Dateiformaten – beispielsweise Comma Separated Value (CSV) und Extended Markup Language (XML) – abgespeichert, damit die Daten später wieder verwendet werden können. So wird das zumindest in der Clinical Trail Unit der Universität Bern gemacht. Die Auswahl der Dateiformate ist wichtig für die weitere Verwendung der Daten, das Sharing und somit auch für das Thema Open Science. Muriel Helmers hat auf das Clinical Data Interchange Standards Consortium (CDISC) hingewiesen. Diese Organisation entwickelt Standards zum Austausch von Daten aus klinischen Studien. CDISC-Standards sind vor allem in der Industrie schon etabliert, in der akademischen Forschung aber kaum in Gebrauch – laut Muriel Helmers zumindest in der Schweiz – weil teilweise das Knowhow in den Studienteams fehlt und der Aufwand der Einführung gross ist.

Beim Sharing ist der Datenschutz beziehungsweise die De-Identifizierung zentral. Bevor Daten der Öffentlichkeit oder anderen Forschern zugänglich gemacht werden, müssen die Forschungsdaten anonymisiert werden, so dass «die Kombination von Feldern keinen Aufschluss auf die Person gibt.» Muriel Helmers führt weiter aus:

«Für die De-Identifizierung eines bestehenden Datensatzes ist noch ein Mensch nötig, der das Wissen hat, um zu entscheiden, was alles anonymisiert werden muss – mehrere nicht identifizierende Datenpunkte können in Kombination wieder identifizierend sein, zum Beispiel bei seltenen Krankheiten oder einer kleinen Population, wie Personen mit einem Alter über 90.»

Die Anonymisierung der personenbezogenen Forschungsdaten ist auch bei anderen Datenformaten wichtig, stellt aber eine andere/ weitere Herausforderung dar. Muriel Helmers meint zur De-Identifizierung von Bildern:

«Bilder sind ein Problem. In den Metadaten sind oft Patientendaten hinterlegt, aber auch auf dem Bild drauf sind personenbezogene Daten. Darum werden die Bilder oft im Zentrum [Forschungszentrum oder CTU] gehalten und nur die Werte – strukturierte Daten – kommen in die Datenbank. Die Bilder sind nur für das Studienteam einsehbar, werden aber nicht weitergeben.»

Für die Anonymisierung von Bildern wird daher auch eine spezielle Software benötigt.

Bei der Langzeitarchivierung – aber auch bei der Datensammlung – muss laut Muriel Helmers darauf geachtet werden, wo die Forschungsdaten gespeichert werden. Die Daten sollen in der Schweiz bleiben, also ist bei der Verwendung von Clouds Vorsicht geboten, weil die Server für ebendiese Speicher teilweise im Ausland liegen und so möglicherweise ungenügend geschützt sind.

Der letzte Bereich, der im Interview diskutiert wurde, ist Open Science und der SNF. Auf die Frage, was sie von der Forderung des SNF hält, Forschungsdaten öffentlich zugänglich zu machen, hat Muriel Helmers gesagt:

«Es wurden oft Studien mit schlechten Ergebnissen/ Werten verschwiegen, weil niemand auf einer schlechten Publikation sein wollte/will. Die klinische Forschung ist aber teuer, daher kommt auch der Sharing-Gedanke. Es ist aber auch ethisch konnotiert: Wenn Menschen schon an einer Studie teilgenommen haben und ein gewisses Risiko auf sich genommen haben, sollten die Daten zur Verfügung stehen, damit nicht etwas später wieder dieselben Fehler gemacht werden und die Daten auch für andere Studien benützt werden können.»

Daraus ergibt sich dann auch die Notwendigkeit die personenbezogenen Daten zu anonymisieren. Es zeigt sich also, dass vor allem der Schutz der personenbezogenen Daten einen hohen Stellenwert im Management der klinischen Forschungsdaten einnimmt.

3.3 Fazit

Für die Erhebung und Analyse der Anforderungen und somit der Ermittlung der Inhalte für einen Forschungsdatenmanagement-Kurs mit Zielgruppe Masterstudierende der Medizin wurden nicht nur Bedarfsanalyse, sondern auch Case Studies und Umfragen hinzugezogen. Die aus der Literatur ermittelten Inhalte haben teilweise ein anderes Abstraktionsniveau und werden je nach Autor und Meinung unterschiedlich priorisiert. Ausserdem werden verschiedene Begriffe verwendet, obwohl das gleiche gemeint ist. Aber die Analyse hat gezeigt, dass es trotz dieser Differenzen grundsätzlich immer um dieselben Themen geht.

Die Analyse hat zudem gezeigt, dass die Abgrenzung zwischen Data Literacy und dem (Forschungs-)Datenmanagement nicht ganz einfach ist. Wenn die fünf Bereiche des Data Literacy vor Augen geführt werden, erkennt man, dass das Forschungsdatenmanagement – soll es umfassend und verständlich vermittelt werden – nicht nur auf das

Wissen aus dem «Datenmanagement», sondern auch aus «Konzeptrahmen» und «Datenanwendung» angewiesen ist. Die beiden Bereiche «Datenerhebung und -sammlung» und «Datenevaluierung» sind aber hingegen für ein sauberes Datenmanagement eher zweitrangig. Es kann nicht abgestritten werden, dass sie von den anderen Bereichen beeinflusst werden. Die Erhebung und Sammlung aber differenziert sich insofern von den anderen relevanten Bereichen, dass dort die Daten noch nicht im Besitz der Forscherinnen sind. Die Evaluierung hebt sich insofern ab, dass dort den Daten Bedeutung verliehen wird. Laut der Website forschungsdaten.info (2020o) zielt das Forschungsdatenmanagement darauf ab, Daten zugänglich, nachnutzbar und -prüfbar zu halten. Die Erhebung, Sammlung, Analyse und Interpretation zählen also nicht per se zu den Aufgaben des Forschungsdatenmanagement. Darum werden für die Erarbeitung des Konzeptes die potenziellen Inhalte, die der Datenerhebung, -sammlung und -evaluierung zugeordnet werden können, wie Recherche, Quellenauswahl, Visualisierung und Entscheidungsfindung, aussen vorgelassen.

Es hat sich ergeben, dass das Management von klinischen Forschungsdaten nicht grundlegend anders ist, als das von anderen Disziplinen. Speziell an den klinischen Daten ist, dass es sich um sensible Daten der menschlichen Gesundheit handelt, die durch gesetzliche Grundlagen und ethische Richtlinien besonders geschützt sind. So resultieren andere/ weitere erforderliche Massnahmen und Entscheidungen, die im Zusammenhang mit dem Forschungsdatenmanagement bedacht und ergriffen werden müssen. Dass das Management von klinischen Daten auf derselben Basis beruht, wie die anderen Disziplinen, hat das Experteninterview bestätigt: Das Interview hat keine neuen Inhalte per se geliefert. Jedoch konnten dank Muriel Helmers die Anforderungen beziehungsweise Eigenheiten von klinischen Daten verdeutlicht und aufgezeigt werden, in welchen Bereichen des FDM besonders auf diese Eigenheiten geachtet werden muss.

Jetzt bleibt nur noch die Fragestellung «Welche Inhalte sind wichtig und relevant, um das Thema Forschungsdatenmanagement in einem Kurs mit Zielgruppe Masterstudierende im Medizinstudium zu vermitteln?» zu beantworten. Aus der Literaturanalyse und dem Experteninterview ergeben sich folgende Themen, die für ein Unterrichtskonzept wichtig und relevant sind:

Themen	Detailliertere Inhalte:
Einführung	
Motivation	Gründe und Vorteile der Datenverwaltung; Forschungsförderer als Stakeholder; Besonderheit klinischer Daten
Begrifflichkeiten	Daten; Forschungsdaten; (klinisches) Forschungsdatenmanagement
Datenlebenszyklus / Stages of Data	Modell; Aufgaben des Forschungsdatenmanagements
Rechtliche Aspekte & Ethik	
Datenschutz	Consent; Umgang sensible Daten; Gesetze & Richtlinien; minimale Anforderungen
Ethik	Ethik im Zusammenhang mit dem Datenschutz; Link Swissethics
Datenmanagement & -organisation	
DMP	Definition; Zweck; Grobe Inhalte; Tools; SNF
Datentypen, Datenformate & Dateiformate	Primär- vs. Sekundärdaten; Kategorien von Datentypen und -formaten; Geeignete Dateiformate
Dateiorganisation	Gründe/ Nutzen; Ordnerstruktur
Dateinamen und Versionierung	Gründe / Nutzen; Best Practice
Dokumentation, Metadaten & -standards	Gründe / Nutzen; Verweise Data-Curation Profiles; Definition Metadaten; Zweck von Standards; CDISC & weitere (allgemeine) Standards; Exkurs Kontrolliertes Vokabular

Datenbanken & Zugriffsrechte	Grundlegende Begriffe inkl. Primärschlüssel; CDMS; KlinV und Zugriffskontrolle
Datenqualität	Unterschied Qualitätssicherung und -kontrolle; Massnahmen; Kriterien zur Bewertung der Qualität
Datenverarbeitung	Überblick Aufgaben der Datenverarbeitung; Einführung Diskrepanzen & Edit Checks
Verantwortlichkeiten	Verantwortlichkeit von Forschenden und Forschungsinstitution bezüglich Forschungsdatenmanagement
Preservation	
Datenspeicherung & -sicherheit	Backup (3-2-1 Regel); Speichermedien und -orte; Sicherheitsrisiken
Literaturverwaltung	Nutzen; Tools
Langzeitarchivierung	Definition; Emulation & Migration, Verweise Dateiformate, Metadaten, Datensicherheit und -schutz; Bewertung und Auswahl
Publikation, Sharing & SNF	
Urheberrecht & Lizenzen	Urheberrecht, Urheber, Creative Commons
Sharing, Nachnutzung und Zitation	Gründe & Nutzen von Publikation/Sharing; Voraussetzung/ Massnahmen vor Publikation; Verweise Datenschutz, -sicherheit und Dokumentation/ Metadaten; Persistente Identifikatoren (ORCID, DOI, URL); Hinweis Literaturverwaltungsprogramm und Zitierstile
Repositorien, Open Access & Open Data	Zweck und Ziel von Open Access und Open Data; Repository; Verweis Zenodo, BORIS und re3data.org

Erwartungen SNF & FAIR-Prinzipien	Erwartungen SNF; FAIR-Prinzipien; Verweis DMP und Open Science
-----------------------------------	--

Tabelle 1: Inhalte Kurskonzept

Aus der Analyse habe sich fünf Hauptbereiche herauskristallisiert, die jeweils mehrere Unterthemen beinhalten (siehe Tabelle 1). In den Unterthemen werden mehrere Inhalte behandelt. Diese detaillierten Inhalte stammen teilweise ebenfalls aus der Analyse, teilweise wurden sie von der Autorin bestimmt, je nachdem was als wichtig und relevant für das jeweilige Thema eingestuft wurde. Die Kompetenzen und Lernziele werden anhand dieses Gerüsts ausgearbeitet.

4 Kurskonzept

In diesem Kapitel werden nun die Lernziele formuliert und die Inhalte erarbeitet. Bevor die Sachanalyse vorgenommen wird, wird Bezug auf das Curriculum der Medizin genommen und die Kompetenzen ausformuliert. Nach der Erarbeitung der Inhalte folgen noch didaktische Überlegungen und Vorschläge zum Format sowie ein Überblick über die Vermittlung von Forschungsdatenmanagement-Kompetenzen an anderen Universitäten und Hochschulen.

4.1 Curriculum Medizin & Kompetenzen

Mit diesem und nachfolgendem Kapitel (Lernziele) wird versucht die Frage «Wozu werden Inhalte zum klinischen Forschungsdatenmanagement vermittelt?» zu beantworten. Es geht darum aufzuzeigen, welche Qualifikationen und Ziele angestrebt werden.

Momentan sind digitale Kompetenzen noch nicht im übergeordneten Lehrplan der Medizin PROFILES verankert. Das ist ein möglicher Grund, wieso es erst wenige Angebote für Studierende gibt/ im Curriculum sind. Die Universität Bern hat in ihrer Digitalisierungsstrategie zur Umsetzung der Vision fünf Stossrichtungen präsentiert. Eine davon ist «Digital Literacy in der Lehre». Durch das Erlernen von digitalen Kompetenzen sollen Studierende, wie auch Dozierende, «für die Herausforderungen der digitalen Transformation befähigt werden (Universität Bern, 2019a, S. 5).» In den Zielen (Ziel 1) wird diese Stossrichtung konkretisiert (Universität Bern, 2019a, S. 7):

«Die Universität Bern stellt sicher, dass alle Studierenden während ihres Studiums digitale Kompetenzen erlernen. Digital Literacy umfassen algorithmisches Denken, Fähigkeiten zur Datenanalyse und ein generelles Verständnis der Chancen und Risiken der Digitalisierung. Dazu werden einerseits disziplinübergreifende Lerneinheiten zu technischen und gesellschaftlichen Aspekten [...] angeboten, an denen Studierende aus allen Fachrichtungen teilnehmen können. Andererseits schaffen die Fakultäten disziplinspezifische Kursinhalte und integrieren diese in ihre Studiengänge.»

Ziel 15 besagt weiter (Universität Bern, 2019a, S. 10):

«Die Universität Bern schafft fächerübergreifende Kurzweiterbildungen zu Digitalthemen und Digital Literacy. Längerfristig werden diese Weiterbildungskurse in die modularisierte Struktur der Weiterbildung [...] eingebettet. [...] Darüber hinaus soll auf diese Art der Weiterbildung auch die interdisziplinäre Zusammenarbeit in

Forschung und Lehre gefördert werden. Ein Pool von Digital Literacy Kursen wird geschaffen, welche die Weiterbildungsprogramme einbinden können.»

Und Ziel 18 erklärt (Universität Bern, 2019a, S. 11):

«Die Universität Bern fördert Open Science (u.a. Research Data Management and Sharing, Open Access, Zugang zur Forschungsinformationen) und erleichtert somit die Schaffung, die Verbreitung und den Zugang zu digitalem Wissen.»

Aus diesen Passagen geht hervor, dass ein Forschungsdatenmanagement-Kurs für Studierende legitimiert, relevant und angebracht ist, denn Studierende sind nicht nur Konsumenten, sondern auch Produzenten von Daten (Carlson et al., 2011, S. 22). Weiter sind Kompetenzen im Forschungsdatenmanagement wichtig für die Studien- sowie für die Berufslaufbahn (Schmidt & Holles, 2018). Adamick et al. (2012, S. 187) erklären, dass ein Forschungsdatenmanagement-Kurs einen bereiten Überblick, sowie Disziplinen spezifische Inhalte bieten soll, damit die Studierenden effektiver mit dem Thema Research Data Management vertraut gemacht werden können. Aus diesem Grund soll kein gewöhnlicher Forschungsdatenmanagement-Kurs angeboten werden, sondern – im vorliegenden Fall – einer, der auf klinische Daten zugeschnitten ist. Die Besonderheiten, die untermauern, wieso ein auf die Medizin zugeschnittener Forschungsdatenmanagement-Kurs notwendig ist, werden im Kapitel 4.3.1.1 behandelt.

Eine Lehrveranstaltung zum klinischen Forschungsdatenmanagement könnte im sechsten Studienjahr eingebettet und so im Curriculum verankert werden, um den Studierenden vor Beginn der Dissertation das nötige Rüstzeug mitzugeben. Aber auch für aus der Masterarbeit abgeleitete Publikationen sind Kompetenzen im Forschungsdatenmanagement relevant und hilfreich.

Folgenden Kompetenzen (Tabelle 2) verhalten sich ergänzend zu den im Kapitel 4.2 formulierten Lernzielen. Diese Kompetenzen sollen auf einem höheren Abstraktionsniveau aufzeigen, welche Qualifikationen die Studierenden aus dem Kurs mitnehmen sollen. Jedem Oberthema sind entsprechende Kompetenzen zugeordnet.

Themen	Kompetenzen
Einführung	Die Studierenden müssen die Motivation hinter dem klinischen Forschungsdatenmanagement verstehen, die wichtigsten Begriffe verstehen und anwenden können, erkennen, wieso das Datenmanagement eine Notwendigkeit darstellt, die Besonderheit von klinischen Daten hervorheben und auf das

	Datenmanagement adaptieren können und sich der verschiedenen möglichen Arten und Stadien von Daten anhand des Datenlebenszyklus bewusst sein.
Rechtliche Aspekte & Ethik	Die Studierenden müssen wissen, wie der datenschutzkonforme Umgang mit klinischen Forschungsdaten gehandhabt wird und welche Gesetze, Richtlinien und Standards bezüglich Datenschutz und Ethik im klinischen Forschungsdatenmanagement Relevanz besitzen. Die Studierenden können dieses Wissen auf das persönliche Datenmanagement und das Forschungsdatenmanagement, das im Zusammenhang mit Arbeiten aus dem Studium steht, adaptieren.
Datenmanagement & -organisation	Die Studierenden müssen die Bedeutung des klinischen Forschungsdatenmanagements und dessen Auswirkungen auf die Nutzbarkeit erkennen sowie die verschiedenen Elemente auf die eigene Forschungsarbeit adaptieren und anwenden können.
Preservation	Die Studierenden müssen sich der Notwendigkeit bewusst sein, die ermittelten oder erzeugten (Forschungs-)Daten, die Metadaten, sowie weitere Daten datenschutzkonform und sicher zu speichern, um diese (Forschungs-)Daten so zu verwalten, dass sie anschließend nachgenutzt werden können (angelehnt an Calzada Prado & Marzal, 2013, S. 131).
Publikation, Nachnutzung & SNF	Die Studierenden müssen die Bedeutung der Nachnutzung von Forschungsdaten erkennen, die rechtlichen Aspekte mit klinischen Forschungsdaten im Bezug auf ihre Publikation verstehen und Forschungsdaten so zur Nachnutzung bereitstellen, dass die Anforderungen des SNF, sowie der FAIR-Prinzipien erfüllt sind.

Tabelle 2: Kompetenzen

4.2 Lernziele

In diesem Kapitel werden die Lernziele erarbeitet. Grundlage dafür sind die Anforderungen aus Kapitel 3 und insbesondere die Zusammenfassung aus Kapitel 3.3. Die Kompetenzen aus Kapitel 4.1 stehen ergänzend dazu. Die Lernziele wurden gemäss Instruktionen des Buches «Preparing Instructional Objectives» von Robert F. Mager (1962) formuliert. Die Ziele beziehen sich auf das Können, das die Studierenden nach dem Absolvieren des Kurses aufweisen sollen.

Die Kursziele lauten:

Dieser Kurs soll die Studierenden dazu befähigen, das Forschungsdatenmanagement bewältig- und einordbar wahrzunehmen, sowie angehen zu können.

Die Studierenden können mit Hilfe von geeigneten Referenzen/ Quellen, Tools und unter Berücksichtigung von Datenschutz und Ethik klinische Forschungsdaten so verwalten, dass die wissenschaftliche Arbeit, Resultate und Forschungsdaten nachvollziehbar und FAIR sind.

Die Ziele für die fünf Themenbereiche enthalten mehrere Lernziele. Einige Lernziele können auch mehreren Themenbereichen zugeordnet werden, jedes Ziel wird aber nur einmal aufgeführt.

Einführung

Lernziel 1: Die Studierenden können fünf Gründe zur Datenverwaltung aufzählen.

Lernziel 2: Die Studierenden können erklären, wieso Forschungsförderer – wie der SNF – Open Research Data gutheissen und welche Auswirkungen die Erwartungen von solchen Förderern auf die Forschungsarbeit haben.

Lernziel 3: Die Studierenden können die Besonderheit von klinischen Daten im Gegensatz zu Daten anderer Forschungsdisziplinen erklären und können zwei Themenbereiche aufzählen, die im Speziellen von den Besonderheiten dieser Daten betroffen sind.

Lernziel 4: Die Studierenden können erklären, was Forschungsdaten sind und im Kontext der klinischen Medizin drei konkrete Beispiele von Forschungsdaten aufzählen.

Lernziel 5: Die Studierenden können mit Hilfe des Datenlebenszyklus der Website forschungsdaten.info je zwei Aufgaben des

Forschungsdatenmanagements pro Stadium aufzählen, die auch für die eigene Forschung relevant sind.

Rechtliche Aspekte & Ethik

Lernziel 6: Die Studierenden können aus einem Datenset sensible (personenbezogene) Daten identifizieren und können zwei Massnahmen nennen, um die Daten datenschutzkonform für die Datenerhebung, -nutzung und -verarbeitung zu handhaben.

Lernziel 7: Die Studierenden können zwei Gesetze oder Richtlinien aufzählen, die im Zusammenhang mit dem Datenschutz wichtig für das klinische Forschungsdatenmanagement sind.

Lernziel 8: Die Studierenden können erklären, was der informed Consent bezweckt.

Lernziel 9: Die Studierenden können zwei minimale Anforderungen an den Datenschutz nennen, die im klinischen Forschungsdatenmanagement eingehalten werden müssen.

Datenmanagement & -organisation

Lernziel 10: Die Studierenden können für ein eigenes Forschungsprojekt einen Datenmanagementplan verfassen, der SNF-konform ist.

Lernziel 11: Die Studierenden können anhand eines Beispiels den Unterschied zwischen Datentypen, Datenformaten und Dateiformaten erklären.

Lernziel 12: Die Studierenden können für die Datenformate Text/Dokument, Tabelle, Bild, Video und relationale Datenbank je ein geeignetes Dateiformat nennen, das auch für die Langzeitarchivierung der Daten akzeptabel ist.

Lernziel 13: Die Studierenden können fünf konkrete Best Practices-Regeln zur Benennung und Versionierung von Dateien und Ordnern nennen.

Lernziel 14: Die Studierenden können fünf konkrete Elemente der Datendokumentation aufzählen und erklären, wieso die Datendokumentation und Metadaten wichtig für die Nachnutzung der Daten sind.

Lernziel 15: Die Studierenden können erklären, was der Zweck von Standards ist.

- Lernziel 16: Die Studierenden können erklären, was ein Primärschlüssel und dessen Funktion ist.
- Lernziel 17: Die Studierenden können den Zweck von Logins, Rollen und Audit Trails erklären.
- Lernziel 18: Die Studierenden können den Zweck der Qualitätssicherung und -kontrolle erklären, drei Massnahmen nennen, die dazu ergriffen werden können und erklären, anhand welcher Kriterien bewertet wird, wie es um die Qualität der Daten steht.
- Lernziel 19: Die Studierenden können Diskrepanzen in einem Beispieldatensatz identifizieren und die Massnahme «Edit Check» in diesem Zusammenhang erklären.
- Lernziel 20: Die Studierenden können erklären, welche Verantwortung der Forscher und welche die Forschungseinrichtung im Forschungsdatenmanagement trägt.

Preservation

- Lernziel 21: Die Studierenden können die 3-2-1 Regel erklären und für ihre eigenen Daten angemessen anwenden.
- Lernziel 22: Die Studierenden können ihre gesammelten Referenzen unter Zuhilfenahme eines Literaturverwaltungsprogrammes korrekt zitieren und adäquat verwalten.
- Lernziel 23: Die Studierenden können die beiden Strategien Emulation und Migration im Zusammenhang mit der Langzeitarchivierung erklären.
- Lernziel 24: Die Studierenden können die eigenen Forschungsdaten bewerten und auswählen, welche Daten archiviert werden sollten.

Publikation, Re-Use & SNF

- Lernziel 25: Die Studierenden können das Urheberrecht anhand eines Beispiels erklären und eine geeignete Creative Commons Lizenz für die Veröffentlichung von eigenen (Forschungs-)Daten auswählen.
- Lernziel 26: Die Studierenden können fünf Voraussetzung nennen, damit der Publikation von Forschungsdaten nichts im Wege steht.
- Lernziel 27: Die Studierenden können erklären, was persistente Identifikatoren (beispielsweise DOI) sind und welchen Zweck sie erfüllen.

Lernziel 28: Die Studierenden können ein eigenes ORCID-Profil anlegen.

Lernziel 29: Die Studierenden können erklären, welches Ziel die Open Access und Open Data-Bewegung verfolgt und können unter Zuhilfenahme von geeigneten Mitteln (wie SNF Richtlinien, Policies, Websites etc.) ein oder mehrere mögliche Repositorien für die Publikation von Forschungsdaten eines bestimmten Projektes ausfindig machen.

Lernziel 30: Die Studierenden können die FAIR-Prinzipien erklären.

Im Folgenden Kapitel werden anhand dieser Lernziele die Inhalte erarbeitet.

Zusammenfassend und als Antwort auf die Fragestellung «Wozu werden Inhalte zum klinischen Forschungsdatenmanagement vermittelt?» kann gesagt werden, dass die Studierenden nach dem Absolvieren dieses Kurses die nötigen Kompetenzen und Qualifikationen besitzen sollen, um die eigene Forschung, die im Rahmen des Studiums anfällt adäquat verwalten zu können. Von einem höheren Standpunkt aus betrachtet, sollen bestimmte digitale Kompetenzen gefordert und gefördert werden. Ausserdem geht es darum, die Angst vor dem Thema zu nehmen und den Studierenden nützliche Tools und Hilfsmittel auf den Weg zu geben, die auch für die berufliche Laufbahn von Nutzen sein können/ werden.

4.3 Sachanalyse

In diesem Kapitel folgt nun die Erarbeitung der Inhalte anhand der formulierten Lernziele und zusammengetragenen Anforderungen. Dieses Kapitel ist in die fünf Themenbereiche und dazugehörigen Unterthemen, wie in Tabelle 1 ab Seite 15 dargestellt, unterteilt.

4.3.1 Einführung

Die Einführung behandelt die Unterthemen Motivation, Begrifflichkeiten und Datenlebenszyklus. Es werden insbesondere die Vorteile von, Gründe für und Aufgaben des Forschungsdatenmanagements aufgezeigt.

4.3.1.1 Motivation

(Digitale) Daten stellen für die Wissenschaft eine wichtige Ressource dar. Mit der digitalen Transformation hat sich der Umfang und die Arbeitsweise mit den Daten verändert: Es stehen immer mehr Daten zur Verfügung, um eine Forschungsfrage beantworten zu können, somit nimmt auch der Bedarf an Daten zu. Dadurch können neue Methoden entwickelt werden, aber vor allem verändert sich der Umgang mit diesen

Daten. Übergreifende Kooperationsstrukturen dank entsprechender Infrastrukturen und internationale Zusammenarbeit haben Einfluss auf die Arbeitsweisen. Dieser veränderte Umgang mit Forschungsdaten hat alle Disziplinen erfasst und verlangt professionelle Kompetenzen. Das gestiegene Interesse an Forschungsdaten und dem Forschungsdatenmanagement und somit deren Bedeutung zeigt sich auch in veränderten Förderrichtlinien (Cremer et al., 2015, S. 14).

«Angetrieben durch weitgreifende Veränderungen der wissenschaftlichen Landschaft (Open Access, Data Intensive Science) wird – neben der herkömmlichen Veröffentlichung von Forschungsergebnissen – vermehrt ein direkter und dauerhafter Zugang zu den Primärdaten wissenschaftlicher Aktivität, d.h. den Forschungsdaten gefordert (Schneider, 2017).»

Gefordert wird dieser Zugang unter anderem auch vom SNF:

«Der SNF erachtet den offenen Zugriff auf Forschungsdaten als einen wesentlichen Beitrag zur Wirkung, Transparenz und Reproduzierbarkeit wissenschaftlicher Forschung. Forschungsdaten sollten aus Sicht des SNF nicht nur sorgfältig aufbereitet und archiviert, sondern auch so weit wie möglich zugänglich gemacht werden (Schweizerischer Nationalfonds, o. J.-b).»

Diese Grundsatzerklärung zeigt, dass der SNF «Open Research Data» gutheisst und unterstützt. Darum verlangt der SNF seit Oktober 2017 auch bei Einreichung eines Antrags für Fördermittel einen Datenmanagementplan (Schweizerischer Nationalfonds, o. J.-b). Ausserdem hebt diese Erklärung auch schon einige Vorteile des Forschungsdatenmanagements hervor. Weitere Vorteile und Gründe aus Perspektive einer Forscherin sind:

Das Forschungsdatenmanagement

- hilft sich und die eigenen Forschung zu organisieren (Liguori, 2016).
- verbessert die Versionskontrolle von Daten (Liguori, 2016).
- hilft Datenverluste vorzubeugen (Liguori, 2016).
- hält die Datenqualität hoch (Liguori, 2016).
- hält die Daten sicher und geschützt (The University of Edinburgh, 2020c).
- hilft die Reproduzierbarkeit zu gewährleisten (Liguori, 2016).
- verifiziert die Integrität der Daten und somit der Forschungsergebnisse (Yale Library, 2020g).
- trägt zur Auffindbarkeit der Daten bei (Yale Library, 2020g).

- macht durch die Dokumentation Daten verständlicher und somit wiederverwendbar (Liguori, 2016).
- ermöglicht kollaborative Arbeitsformen (The University of Edinburgh, 2020c).
- unterstützt die Nachnutzung/ gemeinsame Nutzung (Sharing) und fördert so die Glaubwürdigkeit und Anerkennung eines Forschers (Liguori, 2016).

Aus Sicht eines Studierenden können Kompetenzen im Forschungsdatenmanagement zudem einen Vermarktungspunkt bei der Stellensuche darstellen, insbesondere, wenn eine Karriere ausserhalb des akademischen Bereichs (beispielsweise Industrie) angestrebt wird (Goben & Griffin, 2019, S. 912). Kenntnisse im Datenmanagement können aber auch während dem Studium von Vorteil sein: bei Projekten, Arbeiten und Seminaren.

Das Forschungsdatenmanagement ist also nicht nur persönlich und für die eigene Forschung nützlich und von Vorteil, sondern unterstützt auch andere, die Interesse an den Daten haben und wird darüber hinaus von Forschungsförderern verlangt (The University of Edinburgh, 2020c), zumindest die Ausarbeitung eines DMP. Es ist nun klar, wieso das Forschungsdatenmanagement im Allgemeinen eine wichtige Aufgabe im Forschungsprozess ist. Aber was ist das Besondere an klinischen Daten? Das soll im Folgenden geklärt werden.

«When new drugs or devices are tested in humans, the data generated by, and related to, these trials are known as clinical data. [...] It is these data that will eventually make a new product useful, and marketable, in disease therapy. The management of clinical data, from its collection to its extraction for analysis, has become a critical element in the steps to prepare a regulatory submission and to obtain approval to market a treatment. As its importance has grown, clinical data management (CDM) has changed from an essentially clerical task in the late 1970s and early 1980s to the highly computerized specialty it is today (Prokscha, 2007, Preface).»

Es spielen datenschutzrechtliche und ethische Belangen mit hinein: Laut Muriel Helmers (M. Helmers, persönliche Kommunikation, 11. Juni 2020) sind Datensicherheit, -schutz und Ethik besonders wichtig, weil mit Daten von Menschen gearbeitet wird. Sollten personenbezogene Daten aus klinischen Studien enthüllt werden, kann das für den Probanden zu Konsequenzen, wie Verlegenheit, Verweigerung des Versicherungsschutzes oder Diskriminierung am Arbeitsplatz, führen (Society for Clinical Data Management, 2013, S. 14, Data Privacy).

Es kann deshalb nicht ein beliebiger Kurs zum Forschungsdatenmanagement zur Schulung dieser Kompetenzen genommen werden. Es ist nötig, dass die Inhalte auf die Besonderheiten von klinischen Daten adaptiert werden.

4.3.1.2 Begrifflichkeiten

«Data and information are closely related concepts. Data provide the foundation from which knowledge is derived, whereby data become information in the process of being analysed (The University of Edinburgh, 2020c).»

Daten beschreiben eine Folge von Zeichen und können quantitativ, qualitativ, numerisch, alphanumerisch, textlich oder (audio-)visuell sein. Durch Kontext erlangen die Zeichen (Daten) Bedeutung und werden so zu Information (Czernik, 2016; The University of Edinburgh, 2020c).

Laut der Website forschungsdaten.info (2020k) gibt es bisher keine einheitliche Definition zum Begriff Forschungsdaten. Die Definitionen differenzieren sich von Fach zu Fach. Folgende Definition der Deutschen Forschungsgesellschaft (2015, S. 1) wird auch auf der ebengenannten Website publiziert und stellt das Dilemma treffend dar:

«Forschungsdaten sind eine wesentliche Grundlage für das wissenschaftliche Arbeiten. Die Vielfalt solcher Daten entspricht der Vielfalt unterschiedlicher wissenschaftlicher Disziplinen, Erkenntnisinteressen und Forschungsverfahren. Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.»

Das bedeutet der Kontext (Wissenschaftsdisziplin) definiert, was konkrete Forschungsdaten sind (Schneider, 2017). Schneider (2017) erklärt weiter, dass jegliche digitale vorliegende Datentypen im Kontext der wissenschaftlichen Forschung als Forschungsdaten gilt. Das sehen Kindling & Schirmbacher (2012, S. 130) sehr ähnlich:

«Unter digitalen Forschungsdaten verstehen wir [...] alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind. Der Forschungsprozess umfasst dabei den gesamten Kreislauf von der Forschungsdatengenerierung [...] über die Bearbeitung und Analyse bis hin zur Publikation und Archivierung von Forschungsdaten. Digitale Forschungsdaten entstehen in allen Wissenschaftsdisziplinen [...].»

Es wird hier von digitalen Daten gesprochen, weil im Forschungsdatenmanagement fast immer mit solchen gearbeitet wird (Forschungsdaten.info, 2020k). Vereinfacht gesagt, sind Forschungsdaten (digitale) Daten, die beim wissenschaftlichen Arbeiten produziert werden und als Basis für die aus dem Forschungsprozess resultierten Ergebnissen stehen (Forschungsdaten.info, 2020k). Als Beispiele für (klinische) Forschungsdaten können

- Quellen, wie Notizen, Bilder (beispielsweise Magnetresonanztomographie (MRT) von verschiedenen Patienten), Texte und audiovisuelle Aufnahmen, Niederschriften von Interviews;
- Daten aus Beobachtungen, Experimente & Simulationen, wie Datenbestände (Tabellen), mit Echtzeit-Daten oder Laborwerten, wie beispielsweise Blutwerten, Gewicht und Dosierung, statistische Aufzeichnungen;
- Methodische Testverfahren, wie Software und Fragebögen;
- Referenzen, wie bereits publizierte Datenbestände;
- etc.

genannt werden (Universität Bern, o. J.; Yale Library, 2020g).

Nicht zu den Forschungsdaten gehören Zusammenfassung, Publikationen oder Entwürfe von Büchern oder Artikeln, Peer-Review-Berichte und Kommunikation zwischen und mit Kollegen (Yale Library, 2020g).

Das Forschungsdatenmanagement wird von der Universität Edinburgh (2020a) wie folgt definiert:

«Research data management refers to the activity of working with research data throughout the research process, from data collection, to data storage and backup, through to data sharing at the end of a research project.»

Die Website [forschungsdaten.info](https://www.forschungsdaten.info) (2020l) definiert den Begriff wie folgt:

«Forschungsdatenmanagement (FDM) umfasst die Prozesse der Transformation, Selektion und Speicherung von Forschungsdaten mit dem gemeinsamen Ziel, diese langfristig und personenunabhängig zugänglich, nachnutzbar und nachprüfbar zu halten.»

Es geht also um die Verwaltung der Forschungsdaten, die während der wissenschaftlichen Arbeit involviert sind. Um den Umfang des Forschungsdatenmanagements zu erfassen, kann laut Surkis und Read (2015, S. 154) der Datenlebenszyklus beigezogen werden (siehe Kapitel 4.3.1.3). So lautet auch die Definition auf der Website [forschungsdaten-bildung.de](https://www.forschungsdaten-bildung.de): «Forschungsdaten bezeichnen

den systematischen und planvollen Umgang mit Forschungsdaten über den gesamten Lebenszyklus der Daten hinweg (DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation, o. J.).» Beim klinischen Forschungsdatenmanagement geht es demnach um die Verwaltung von Daten aus beispielsweise klinischen Studien oder Registern.

4.3.1.3 Datenlebenszyklus

Der Datenlebenszyklus bietet sich als Orientierungshilfe an, um den Umfang des Forschungsdatenmanagements erfassen zu können (Forschungsdaten.info, 2020I). Es gibt mehr als ein Model des Datenlebenszyklus. Carlson (2014, S. 83) weist auf diese verschiedenen Modelle hin und erklärt, dass es wichtig ist zu verstehen, wie die Forschung der Zielgruppe durchgeführt wird, um ein aussagekräftiges und repräsentatives Modell bezüglich deren Forschungspraktiken und Anforderungen zu erstellen: «Simply adopting a life cycle model produced by another organization without first acquiring this understanding is no likely to be an effective approach (Carlson, 2014, S. 83–84).» Weil das Ziel dieser Arbeit nicht die Entwicklung eines *Datenlebenszyklus* für die Medizin beinhaltet, wird auf ein einfaches Modell der Website [forschungsdaten.info](https://www.forschungsdaten.info) zurückgegriffen. Für einen Grundkurs sollte diese Grafik genug ausführlich sein. Ein komplexeres, ausführlicheres Modell des Datenlebenszyklus kann auf der Website <https://www.dcc.ac.uk/about/digital-curation> studiert werden.

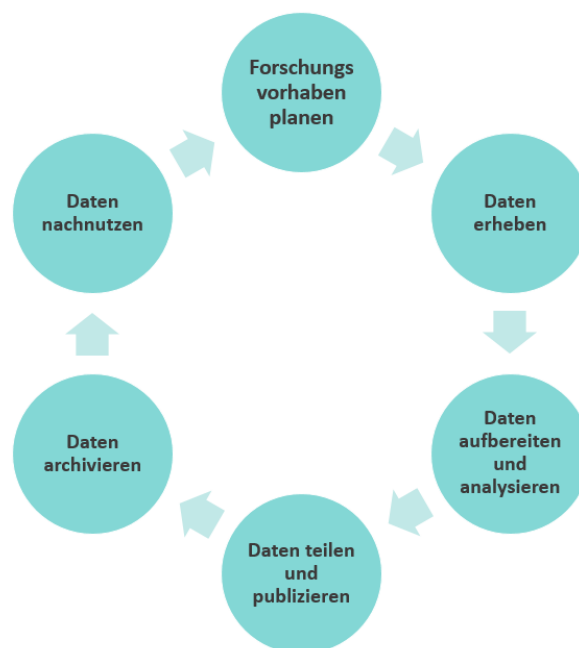


Abbildung 1: Datenlebenszyklus, Quelle: Forschungsdaten.info, 2020h

Der Datenlebenszyklus beschreibt, wie der Name schon sagt, das «Leben» der Daten. Laut Carlson (2014, S. 63) durchlaufen Daten, von ihrer Entstehung oder Generierung bis zur Archivierung und Nachnutzung, einen Zyklus, in dem die Forschungsdaten mehrere Transformationen bezüglich Format, Gebrauch und möglicherweise Zweck durchmachen. Ausserdem kann die Verwendung eines solchen Modells helfen, die Stadien in Kontext zu setzen. Wie in Abbildung 1 ersichtlich, besteht dieser Datenlebenszyklus aus sechs Stadien. Für jedes Stadium fallen Aufgaben an – je nach Forschungsdisziplin und -vorhaben können die unterschiedlich umfangreich ausfallen (Forschungsdaten.info, 2020h):

Beim Planen des Forschungsvorhabens werden das Untersuchungsdesign erstellt, bereits bestehende Daten ermittelt und das Datenmanagement mittels DPM, sowie Vorbereitungen für die (Nach-)Nutzung (beispielsweise Consent) geplant. Der nächste Schritt beschreibt die Datensammlung mittels definierter Methoden, beispielsweise Beobachtung, Experiment, Umfrage, Messungen etc. In diesem Stadium wird beim Probanden auch die Zustimmung für die Datennutzung eingeholt. Bei der Aufbereitung und Analyse der Daten, wird eingeben, digitalisiert, transkribiert, übersetzt, gesichert, verwaltet, geprüft, validiert, bereinigt, anonymisiert, beschrieben und interpretiert. Ausserdem werden in diesem Stadium die Ergebnisse erarbeitet und die Erhaltung der Daten vorbereitet. Weiter geht es mit dem Publizieren: Das Urheberrecht muss festgelegt werden, ausserdem werden die Daten geteilt und promotet sowie der Zugang dazu definiert (wer hat alles Zugang? Braucht es ein Login? Etc.). Der zweitletzte Punkt beschreibt die Aufgaben Daten in/ auf geeignete Formen und Medien (Datenträger) migrieren; Daten sichern (Backups erstellen) und mittels Metadaten dokumentieren, sowie für den langfristigen Zugang archivieren. Die Nachnutzung schliesslich beschreibt die Aufgaben Prüfen der bisherigen Ergebnisse; weitere Untersuchungen unter Verwendung der Daten durchführen und bisherige Forschungen kritisch beurteilen. Zudem wird in diesem Stadium mit den Daten gelernt und gelehrt (Forschungsdaten.info, 2020h). Mit diesem letzten Schritt schliesst sich der Kreis und die Daten können in neuen Forschungsvorhaben wiederverwendet werden. Jetzt bleibt noch die Frage offen, welche dieser Aufgaben in das Forschungsdatenmanagement fallen. Die Antwort: Grundsätzlich können Massnahmen in allen sechs Stadien ergriffen werden, um Forschungsdaten FAIR (siehe Kapitel 4.3.5.4) zu halten (Forschungsdaten.info, 2020l). Aber nicht jede Aufgabe gehört explizit zum Forschungsdatenmanagement, wie in Kapitel 3.3 bereits erwähnt und in der Definition des Begriffs (Kapitel 4.3.1.2) ausgeführt, besteht das Ziel darin, die Daten zugänglich, nachnutzbar und nachprüfbar zu halten

(Forschungsdaten.info, 2020). So gehören also alle Aufgaben, die diesem Zweck dienen zum Forschungsdatenmanagement.

4.3.2 Rechtliche Aspekte & Ethik

Dieses Kapitel behandelt die Unterthemen Datenschutz und Ethik. Es werden insbesondere die rechtlichen Aspekte im Hinblick auf klinische Daten, sowie Massnahmen zur datenschutzkonformen Verwaltung dieser Daten aufgezeigt. Nicht diskutiert wird das Urheberrecht: das wird im Kapitel 4.3.5.1 nachgeholt.

4.3.2.1 Datenschutz

«Der Umgang mit besonderen Arten personenbezogener Daten (wie Gesundheitsdaten) erfordert ein erhöhtes Maß an Sensibilität für den Datenschutz (Datenschutz.org, 2018a).» Man spricht daher im Zusammenhang mit medizinischen Daten auch von sensiblen Daten (Datenschutz.org, 2018b). Besonders schützenswert sind diese Daten, weil es sich um Personendaten über die Gesundheit und Intimsphäre handelt (Datenschutzgesetz Art. 3 lit. c Ziff. 2) und die Veröffentlichung zu Verlegenheit und Benachteiligung führen kann (M. Helmers, persönliche Kommunikation, 11. Juni 2020; Society for Clinical Data Management, 2013, S. 14, Data Privacy). Laut Humanforschungsgesetz (HFG Art. 3 lit. f) werden solche Daten gesundheitsbezogene Personendaten genannt. Laut DSGVO Art. 3 lit. a und DSGVO Art. 4 Abs. 1 sind personenbezogene Daten Informationen/ Angaben, die sich auf eine bestimmte oder bestimmbare (identifizierte oder identifizierbare) Person beziehen. Als Beispiele solcher Daten können Name, Geburtsdatum, Adresse, Versicherungsnummer und Telefonnummer genannt werden.

«According to the principle of accountability in the GDPR [General Data Protection Regulation], it is the institute's responsibility to ensure that the fundamental principles relating to processing of personal data are respected, as well as the ability to demonstrate compliance. [...] Possible consequences of not adhering to these principles include reputation damage, liability, and losing or having to refund a research grant (Jansen et al., 2019, S. 39).»

Zu Deutsch ist die GDPR die Datenschutzgrundverordnung (DSGVO). Das Schweizer Pendant ist das Bundesgesetz über den Datenschutz (DSG). Es sind aber noch andere Richtlinien, Standards und Gesetze von Relevanz, wenn es um den Datenschutz beim klinischen Forschungsdatenmanagement geht: das Recht auf informationelle Selbstbestimmung, ärztliche Schweigepflicht, das Humanforschungsgesetz (HFG), die

Gute Klinische Praxis² (GCP), die Good Clinical Data Management Practices³ (GCDMP) und die European Clinical Research Infrastructure Network⁴ (ECRIN) (Datenschutz.org, 2018a; M. Helmers, persönliche Kommunikation, 11. Juni 2020). Anmerkung: Die Richtlinien und Standards beschäftigen sich nicht nur mit dem Datenschutz, sondern auch mit der ganzen Reichweite des Forschungsdatenmanagements.

Das Zitat von Jansen et al. (2019) zeigt die möglichen Konsequenzen auf, wenn die Gesetze nicht eingehalten werden. Um die Datenerhebung, -nutzung und -verarbeitung gesetzeskonform zu handhaben, benötigt es daher Instrumente wie eine Einwilligung (Consent) des Patienten, Schutz der Daten vor unbefugten Dritten und Anonymisierung/Pseudonymisierung der Daten (Datenschutz.org, 2018b, 2018a; M. Helmers, persönliche Kommunikation, 11. Juni 2020). Es gibt einen general Consent (Generalkonsent) und einen informed Consent: der general Consent wird bei der Aufnahme im Spital eingesetzt (Routinedaten), der informed Consent kommt bei Forschungsprojekten zum Zug (M. Helmers, persönliche Kommunikation, 11. Juni 2020). Mit dieser schriftlichen Einwilligung gibt der Patient/ Proband das Einverständnis zur Teilnahme an der Studie und allenfalls zur Weiterverwendung der erhobenen Daten (HFG Art. 16 & 17). Der Schutz vor unbefugtem Zugriff wird in Kapitel 4.3.3.6 – Datenbanken & Zugriffsrecht – sowie Kapitel 4.3.4.1 – Datenspeicherung und -sicherheit – genauer erklärt. Um personenbezogene Daten datenschutzkonform zu verarbeiten und weiterzugeben, werden die Daten anonymisiert oder pseudonymisiert. Bei der Anonymisierung werden die Daten so de-identifiziert, dass kein Rückschluss mehr – oder dann nur mit enormen Aufwand – auf eine konkrete Person erlauben (Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB), o. J.-a; Jansen et al., 2019, S. 45). Zur Pseudonymisierung erklärt der Eidgenössische Datenschutz- und Öffentlichkeitsbeauftragte (EDÖB):

«Wenn wegen der Art oder des Zwecks der Forschung nicht mit anonymisierten Daten gearbeitet werden kann, weil es beispielsweise wichtig ist, die betroffenen Personen regelmässig kontaktieren zu können, müssen die Forscherinnen und Forscher die Daten kodieren oder verschlüsseln (pseudonymisierte Daten). So können sie anstelle des Namens und der anderen identifizierenden Daten einen

² Einsehbar unter folgendem Link: <https://www.gesetze-im-internet.de/gcp-v/BJNR208100004.html>

³ Die GCDMP können kostenlos (unter Angabe einer E-Mail-Adresse) auf der Website <https://scdm.org/gcdmp/> gedownloadet werden.

⁴ Dieser Standard setzt sich laut Muriel Helmers (persönliche Kommunikation, 11. Juni 2020) immer wie mehr durch. Er kann unter folgendem Link eingesehen werden: <https://www.ecri.org/sites/default/files/Data%20centre%20certification/Standards%20v4%20201804.pdf>

Code/Referenznummer verwenden, sodass es nicht mehr möglich ist, ohne die Referenznummer bestimmte Daten mit einer bestimmten Person in Beziehung zu setzen (Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB), o. J.-a).»

Dieser Code/ Identifikator/ Referenznummer wird von einem Computer automatisch generiert und – weil er personenbezogene Daten und die restlichen Daten in Beziehung setzen kann – sollte an einem sicheren Ort aufbewahrt werden – beispielsweise einer Stelle, die nicht an der Studie beteiligt ist (Europäische Patientenakademie, o. J.; M. Helmers, persönliche Kommunikation, 11. Juni 2020). Konkrete Beispiele zur Anonymisierung von Daten sind beispielsweise Löschen von Einträgen oder Ersetzen von Einträgen mit Platzhaltern oder Merkmalen mit ähnlicher Bedeutung. Bei audiovisuellen Medien können Massnahmen, wie Stimmen nachsprechen, verzerren oder auspiepsen, Bilder schwärzen oder Effekte einbauen, verwendet werden (Lindenmann et al., 2019a, S. 1–2)⁵. Es gilt zu beachten, dass personenbezogene Daten auch in den Metadaten hinterlegt werden können, ausserdem stellen alle nicht strukturierte Daten, wie Bilder und Videos eine besondere Herausforderung in der Archivierung und Anonymisierung dar. Bei letzterem wird dann auch spezielle Software benötigt (M. Helmers, persönliche Kommunikation, 11. Juni 2020).

Laut den GCDMP (Society for Clinical Data Management, 2013, S. 2–3, Data Privacy) sind die minimalen Anforderungen, die eingehalten werden sollten:

- Alle involvierten Personen/ Forscherinnen sollten in Sachen Datenschutz geschult sein.
- Alle Instrumente, die zur Datensammlung benutzt werden und personenbezogene Daten enthalten könnten, sollten mit dem minimal erforderlichen Identifikator/ Code versehen sein, das gilt für Case Report Forms (CRF, = Erhebungsbögen in klinischen Studien), Datenbanken und Spezifikationen zum Datentransfer.
- Personenbezogenen Daten sollten nicht identifizierbar sein und nur über den Identifikator/ Code/ Referenznummer einer Person zugeordnet werden können.
- Alle Forschungsprozesse sollten den aktuellen Datenschutzgesetzen und Richtlinien entsprechen.

⁵ Auf https://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/service/open_science/forschungsdatenmanagement/index_ger.html#pane826303 unter «Sensible Daten» finden sich weitere informative Handouts und Fact Sheets zum Thema personenbezogene Daten.

Um das Einhalten dieser minimalen Anforderungen zu erleichtern empfiehlt die Society for Clinical Data Management (2013, S. 4, Data Privacy) das Benützen von Standard Operation Procedures (SOP). Auf Deutsch sind das standardisierte Vorgehen. SOPs können mit Checklisten verglichen werden: Sie helfen einen Prozess gemäss bestimmten Regeln einzuhalten, so, dass kein Schritt vergessen wird und alles gesetzeskonform und/oder gemäss vorgegebenen Richtlinien abgewickelt wird.

4.3.2.2 Ethik

Im Kontext des Forschungsdatenmanagements sind Datenschutz und Ethik eng miteinander verflochten: Die ärztliche Schweigepflicht ist eine ethische Grundregel, entspringt nicht dem Datenschutz (Datenschutz.org, 2018a), ist aber trotzdem mit ihm verwoben, wie im vorherigen Kapitel ausgeführt.

Laut Artikel 4 des Humanforschungsgesetz haben Interesse, Gesundheit und Wohlergehen des einzelnen Menschen Vorrang gegenüber den Interessen der Wissenschaft und der Gesellschaft. Ausserdem muss laut HFG Art. 45 die Durchführung eines Forschungsprojektes – und somit die Erhebung, Verarbeitung und Weitergabe von Daten – von der zuständigen Ethikkommission bewilligt werden. Die Gesetzeslage sieht vor, dass die Patientinnen hinreichend über das Forschungsprojekt informiert werden und der Teilnahme schriftlich einwilligen, ansonsten ist ein Einbezug dieser Person ausgeschlossen (HFG Art. 16). Hier kommt also der zuvor bereits angesprochene Consent⁶ zum Zug. Sobald ein Gesuch bewilligt wurde, wird die Studie im Forschungsregister Swiss National Clinical Trials Portal (SNCTP) angezeigt, wobei die Daten aus der Gesuchseinreichungsplattform «BASEC» (Business Administration System for Ethics Committees) und dem Register der World Health Organization (WHO) stammen (Kofam Koordinationsstelle Forschung am Menschen, o. J.).

Weil Datenschutz und Ethik sehr wichtige Themen im klinischen Forschungsdatenmanagement sind, werden in den folgenden Kapiteln immer mal wieder Verweise auf diese beiden Unterkapitel platziert werden.

4.3.3 Datenmanagement & -organisation

Dieses Kapitel behandelt die Unterthemen Datenmanagementplan, Datentypen, Datenformate & Dateiformate, Dateinamen & Versionierung, Dateioorganisation,

⁶ Auf der Website des Schweizerischen Verbandes der Forschungsethikkommissionen (<https://swissethics.ch/>) sind Vorlagen und Checklisten zu Studieninformationen und -einwilligungen verfügbar. Ausserdem sind weitere nützliche Links und Informationen zum Einreichungsprozess einer Studie hinterlegt.

Dokumentation, Metadaten & -standards, Datenbanken & Zugriffsrechte, Datenqualität, Datenverarbeitung sowie Verantwortlichkeit.

4.3.3.1 Datenmanagement Plan (DMP)

Der Data Management Plan, auch Forschungsdatenmanagementplan oder Data Handling Plan, gibt Auskunft über den Umgang mit den Forschungsdaten in einem Projekt (Kindling & Pampel, 2016; Prokscha, 2007, S. 3; Yale Library, 2020f). Datenmanagementpläne müssen Forschungsförderern (beispielsweise die Europäische Kommission⁷) beim Einreichen eines Gesuches vorgelegt werden (Yale Library, 2020f). Das ist auch beim SNF so. Der SNF definiert den DMP wie folgt:

«Der Data Management Plan (DMP) dient der Planung des Lebenszyklus von Forschungsdaten. Er ist langfristig angelegt und legt dar, wie die Daten produziert, erhoben, dokumentiert, veröffentlicht und archiviert werden sollten (Schweizerischer Nationalfonds, o. J.-a).»

Laut GCDMP (Society for Clinical Data Management, 2013, S. 3, Data Management Plan) liegt der Zweck eines DMP darin, allen Stakeholdern das nötige Wissen zu vermitteln, um eine qualitativ hochstehende Datenbank zu erstellen und zu verwalten. Ein DMP ist vom Prüfplan/Studienprotokoll zu unterscheiden: Obwohl ein DMP Teile des Studienprotokolls enthalten kann – zum Beispiel die Messwerte, die erhoben werden – bezieht sich das Studienprotokoll auf das gesamte Forschungsvorhaben der klinischen Studie. Der DMP fokussiert sich aber «nur» auf die Daten, die in diesem Forschungsvorhaben anfallen und wie sie verwaltet werden (Society for Clinical Data Management, 2013, S. 1 & 4, Data Management Plan).

Der Inhalt eines DMP kann sich je nach Projekt, Forschungsdisziplin und deren spezifischen Praktiken und Standards unterscheiden (Schweizerischer Nationalfonds, o. J.-a; Yale Library, 2020f).

Ein Datenmanagementplan enthält in der Regel folgende Elemente (Forschungsdaten.info, 2020s; Prokscha, 2007, S. 4; Yale Library, 2020f):

- Überblick: Ziele, Projektverantwortliche, Rollen, Laufzeit, Partner und Studiendesign etc.

⁷ Im Rahmen des Forschungsprogramms «Horizon 2020» werden Datenmanagementpläne gefordert. Weitere Information zu Open Access und Tools sowie eine Vorlage zum DMP sind unter folgendem Link einzusehen: <https://www.horizont2020.de/einstieg-open-access.htm>

- Datenbestand: Angaben zu bereits bestehenden Daten (Herkunft, Qualität), Verantwortlichkeiten und Rollen im Datenmanagement
- Angaben zu Datentypen, -formaten und -menge, sowie zur Dokumentation und zur Art der Datengenerierung (inkl. Case Report Form)
- Angaben zur Datenverwaltung, -speicherung, sowie Versionierung, Synchronisation und Kollaboration
- Angaben zu Richtlinien von Förderorganisationen, Zugänglichkeit und Nachnutzung der Daten inkl. Datenschutz, Sicherheit, Urheberrecht
- Angaben zu Metadaten, zur Datenauswahl sowie -übergabe ins Archiv

Diese Inhalte des DMP werden in den nachfolgenden Kapiteln und Unterkapiteln behandelt.

Bei Einreichung des Gesuches beim SNF müssen Angaben zum Datenmanagement zu vier Bereichen gemacht werden: 1. Datenerhebung und -dokumentation, 2. Ethische, rechtliche und Sicherheitsfragen, 3. Datenspeicherung und -erhalt und 4. Austausch und Weiterverwendung der Daten (Schweizerischer Nationalfonds, o. J.-a). «Der DMP ist ein integraler Teil des Forschungsgesuchs. Das Gesuch kann erst dann eingereicht werden, wenn auch der DMP ausgefüllt wurde (Schweizerischer Nationalfonds, o. J.-a).» Während dem Forschungsprojekt kann der DMP angepasst werden, um der Entwicklung des Projektes zu entsprechen. Die finale Version wird zusammen mit den Forschungsergebnissen und -daten publiziert, um die Sichtbarkeit der Resultate zu erhöhen. Ausserdem sind für andere Forscher die Datensätze leichter zugänglich und weiterverwendbar (Schweizerischer Nationalfonds, o. J.-a). Davon abgesehen, dass ein DMP von Forschungsförderern verlangt wird, bietet ein DMP weitere Vorteile: Anzahl Zitationen werden erhöht; Datenverluste werden minimiert; beim Personalwechsel ist das Wissensmanagement verbessert und das Risiko, dass die Daten aufgrund veralteter Dateiformate nicht mehr gelesen werden können, ist verringert (Forschungsdaten.info, 2020s).

Der SNF stellt für das Ausfüllen des DMP einen Hilfstext⁸ zur Verfügung. Ausserdem verweist der SNF auf seiner Website auf das Digital Curation Centre⁹ (DCC) und das «Swiss DLCM»-Projekt¹⁰, die Beispiele, Vorlagen und Checklisten zum DMP und dem

⁸ Der Hilfstext ist unter http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_de.pdf verfügbar.

⁹ Beispiele für Datenmanagementpläne sind unter <https://www.dcc.ac.uk/dmps> verfügbar. Das DCC bietet im Weiteren auch Checklisten und Best Practices zum DMP an.

¹⁰ DLCM steht für Data Life-Cycle Management. Insbesondere die «Data Management Checklist» gilt es zu erwähnen: Sie bietet eine Aufstellung mit allen Fragen, die man sich zum FDM stellen sollte. Diese und weitere Checklisten, Vorlagen und nützliche Links zum Thema DMP sind unter <https://www.dlcm.ch/resources/dlcm-dmp> verfügbar.

Forschungsdatenmanagement im Allgemeinen anbieten (Schweizerischer Nationalfonds, o. J.-b, o. J.-a). Die Universität Bern stellt für verschiedene Forschungsdisziplinen Beispiele von DMP auf der Website der Universitätsbibliothek zur Verfügung, darunter auch ein konkretes Beispiel für die Medizin¹¹ (Universität Bern, o. J.). Für das Ausfüllen eines DMP stehen Tools zur Verfügung: Das «DMPTool»¹² wird von verschiedenen Quellen empfohlen (DataONE, o. J.-a; Forschungsdaten.info, 2020s; Yale Library, 2020f), ist kostenlos, open-source und online verfügbar. Das Tool wird vom University of California Curation Center betrieben und bietet Vorlagen für die wichtigsten US-Förderorganisationen, sowie das DCC Template (Forschungsdaten.info, 2020s; Yale Library, 2020f). Das DLCM Projekt betreibt den «DMP Canvas Generator»¹³, der SNF-konform ist und der Struktur des vom SNF vorgegebenen Formulars entspricht (DLCM, o. J.).

4.3.3.2 Datentypen, Datenformate und Dateiformate

Datentypen können grob in zwei Typen unterschieden werden: Primär- und Sekundärdaten. Daten, die während der Forschung generiert oder produziert werden, nennt man Primärdaten. Sekundärdaten sind Daten, die bereits bestehen und nachgenutzt werden können. Beispiele für Letztere sind Daten, wie Anzahl Erkrankte, die von statistischen Ämtern erhoben werden. Die Daten können aber auch von vorherigen Forschungsprojekten stammen (Lindenmann et al., 2019b). Weiter können Datentypen in Daten aus Beobachtungen, aus Experimenten, aus Simulationen oder abgeleitet (beispielsweise durch Data-Mining) sowie referenziert (beispielsweise Volkszählung) aus anderen Quellen eingeteilt werden (Oregon State University - Libraries, 2019a).

Daten können in insgesamt fünf Formate eingeteilt werden (The University of Edinburgh, 2020c):

- Text, wie Umfragen, Fragebögen, Protokolle, Notizen aus Beobachtungen etc.
- Numerisch, wie Messungen/ Werte, Antworten aus Umfragen (Skalen),
- Multimedia, wie Bilder, audiovisuelle Dateien
- Code, wie Algorithmen, Quellcode
- Syntax, wie Software-spezifischer Code zur statistischen Analyse

¹¹ Direkter Link zum DMP Beispiel Medizin:

https://www.unibe.ch/unibe/portal/content/e809/e962/e963/e6304/e583799/e573822/e613300/e611703/e587963/pane721254/e721257/2018_SNSF_Example_DMP_Medicine_ger.pdf

¹² Das Tool ist unter <https://dmptool.org/> verfügbar.

¹³ Der DMP Canvas Generator ist unter <https://dmp.vital-it.ch/> verfügbar. Ein Login ist erforderlich. Weitere Tools können auf der Website www.forschungsdaten.info unter der Rubrik Datenmanagementplan eingesehen werden.

Und dann gibt es noch die Dateiformate. Um die Dateien und somit die Daten auch in Zukunft zugänglich und nachnutzbar zu halten, gibt es Formate, die besser geeignet sind als anderen. Formate sind eher zugänglich, wenn

- das Format nicht proprietär (herstellergebunden) ist,
- es sich um offene und dokumentierte Standards handelt,
- das Format gängig ist,
- das Format keine Kompression nutzt,
- zukunftsfähig ist (verbreitet: grosse Anzahl Softwares, die das Format lesen können) und
- standardisierte Zeichenkodierung verwendet wird (Oregon State University - Libraries, 2019a; Universität Bern, o. J.).

Folgende Dateiformate können daher empfohlen werden (NYU Libraries, 2020a; Schweizerisches Bundesarchiv, 2020, S. 4; Universität Bern, o. J.):

Text, Dokumente	XML, HTML, .TXT, PDF/A (gilt insbesondere für Microsoft Office Dokumente)
Tabellen	CSV (comma separated value), XML
Relationale Datenbanken	XML, SQL, SIARD
Strukturierte (Text-)Daten	XML
Bilder	TIFF, JPEG2000
Audio	WAV, MP3 (verlustbehaftet, daher nicht immer zu empfehlen)
Audiovisuell, Video	FFV1, MPEG-4,

Tabelle 3: Empfehlung Dateiformate

Die Verwendung dieser empfohlenen Formate begünstigt nicht nur die Zugänglichkeit und Nachnutzbarkeit, sondern unterstützt auch die Interoperabilität von Systemen. Damit ist gemeint, dass unterschiedliche Systeme nahtlos zusammenarbeiten. Damit das funktioniert, braucht es etablierte Standards (Society for Clinical Data Management, 2013, S. 2, Data Management Standards in Clinical Research): Diese werden im Kapitel 4.3.3.5 zusammen mit den Metadaten besprochen. Für die Archivierung nicht geeignete

(nicht archivwürdige) Dateiformate sind (Forschungsdaten.info, 2020p): .doc (Word); .ppt (PowerPoint) und .xls (Excel). Grund dafür ist, dass diese Formate proprietär sind und darüber hinauslaufend aktualisiert werden und dementsprechend alte Dokumente von den neuen Programmversionen nicht mehr oder nur ungenügend dargestellt werden – beispielweise Zeichen werden nicht mehr erkannt. Wenn ein solcher Fall vorliegt, müssen die Dokumente in ein anderes Dateiformat umgewandelt werden. Man spricht auch von konvertieren. «Eine verlustfreie Konvertierung ist in der Regel vorzuziehen. Liegt die Priorität allerdings bei geringeren Dateigrößen, müssen oft Informationsverluste in Kauf genommen werden (Forschungsdaten.info, 2020p).» Informationsverluste entstehen, wenn ein verlustbehaftetes Format gewählt wird, wie beispielsweise MP3 oder JPEG. Konkret bedeutet das, dass bei diesen Dateien die Qualität (Ton oder Bild) schlechter, dafür weniger Speicherplatz benötigt wird (Forschungsdaten.info, 2020p).

4.3.3.3 Dateiorganisation

Die Organisation von Dateien und Ordnern ist wichtig, weil die Dateien und Ordner so einfacher abrufbar und durchsuchbar sind. Das bedeutet auch, dass andere Personen die Daten leichter finden (The University of Edinburgh, 2020b).

«Um spätere Fehler, Verwechslungen und langes Suchen zu vermeiden, ist es sinnvoll, bereits zu Beginn des Projektes Zeit in eine systematische Organisation der Dateien und Ordner zu investieren. Dies ist insbesondere dann wichtig, wenn mit anderen Forschungsgruppen zusammengearbeitet wird. Dabei sollten sich alle, an einem Projekt beteiligten Personen, mit einem Schema einverstanden erklären und dieses auch einhalten (Universität Bern, o. J.).»

Dieses Schema, also das Organisations- und Benennungssystem wird in einer Konvention festgehalten (Universität Bern, o. J.). Grundsätzlich sollten folgende Punkte berücksichtigt werden, damit die Organisation nutzbringend ist (Forschungsdaten.info, 2020g; Universität Bern, o. J.):

- Daten/ Dateien in Ordnern ablegen
- Ordner eindeutig und unmissverständlich benennen. Die Benennung nach Mitarbeiter ist nicht empfehlenswert, weil nicht ersichtlich ist, was sich im Ordner befindet.
- Verwandte Dateien in Ordnern gruppieren.
- Hierarchische Ordnerstruktur verwenden, aber nicht der Gefahr laufen, die Struktur zu verschachteln.

- Nach Projektende überprüfen, welche Daten/ Dateien noch benötigt und welche gelöscht werden können.

Wie die Benennung der Dateien inklusive Versionskontrolle am besten gemacht wird, wird im nächsten Kapitel genauer erläutert.

4.3.3.4 Dateinamen und Versionierung

«Naming records consistently, logically and in a predictable way will distinguish similar records from one another at a glance, and by doing so will facilitate the storage and retrieval of records, which will enable users to browse file names more effectively and efficiently. Naming records according to agreed conventions should also make file naming easier for colleagues because they will not have to 're-think' the process each time (Gryzbowski, 2007).»

Eine Benennungskonvention, die von Beginn an befolgt wird, kann also wertvolle Zeit sparen (The University of Edinburgh, 2020b). Auch die GCDMP (Society for Clinical Data Management, 2013, S. 10, External Data Transfers) begrüßen eine solche Konvention, gerade, wenn die Dateien auch versendet werden. Dann braucht es Information, die die Daten mit ihrer Quelle und dem entsprechenden Projekt verknüpfen. Dazu gehört auch eine versionskontrollierte Benennungskonvention.

Grundsätzlich sollten die Dateinamen eindeutig und leicht verständlich sein (Universität Bern, o. J.). Des Weiteren sollte die Namensgebung – von der Verzeichnisstruktur, über die Ordner bis zu den einzelnen Dateien – konsistent und deskriptiv (beschreibend) sein. Das heisst, es sollten immer die gleichen Elemente und entsprechenden Standards (beispielsweise YYYYMMDD) verwendet werden (University of Oregon, 2020b). Best Practices sind (Gryzbowski, 2007; The University of Edinburgh, 2020b; Universität Bern, o. J.; University of Oregon, 2020b; Yale Library, 2020c, 2020d):

- Möglichst kurze Namen verwenden und die Dateierweiterung (beispielsweise .jpg; .txt) miteinbeziehen. Ordner- und Dateinamen sollten weniger als 32 Zeichen beinhalten. Lange Namen führen zu langen Pfaden und somit langen URLs. Das erhöht das Risiko eines Fehlers (Error).
- Keine Sonderzeichen (: , ; ? ! & * % \$ £ (] { @ # ^ ~ ' < > / \) verwenden. Sonderzeichen sind oft für das Betriebssystem reserviert.
- Unterstriche anstelle von Leerzeichen oder Punkten verwenden, denn auch diese beiden sind für das Betriebssystem reserviert.
- Wenn eine fortlaufende Nummerierung verwendet wird, sollten Nullen als Platzhalter gebraucht werden, um die Lesbarkeit bei mehrstelliger Version zu erhöhen. Für die

Nummerierung 1-10 sollte die Schreibweise 01-10 und bei Zahlen von 1-100 «001-010-100» verwendet werden.

- Die Elemente, die zur Erstellung eines Dateinamens verwendet werden, sollten auf einer logischen Reihenfolge aufgrund der Abrufmethode basieren. Wenn also Daten zeitspezifisch oder wiederkehrend sind, sollte das Datum/ die Zeit als erstes im Dateinamen verwendet werden, bei Dateien, die an ein bestimmtes Ereignis gebunden sind, sollte der Name dieses Ereignisses zuvorderst stehen. Ein Dateiname kann folgenden Elemente¹⁴ enthalten:
 - Version (z. B. v1 oder Entwurf/ Final oder Details zu Änderungen: ausgeschnitten, formatiert etc.)
 - Datum (z. B. für 3. Juli 2020: 20200703 oder 2020_07_03)
 - Name des Erstellers (z. B. für Max Muster: mm oder mmuster)
 - Inhaltsbeschreibung (z. B. für Pressekonferenz; pk)
 - Vorgesehenes Publikum/ Zielgruppe (z. B. für den Kundensupport: sup)
 - Arbeitsgruppe (z. B. für die Abteilung Clinical Trials: ct)
 - Freigabetermin (Release) oder Veröffentlichungsdatum (z. B. für den 20. Juli 2020, 11:45 Uhr: 20200720_1145 oder pub2020_07_20)
 - Projektnummer oder -name (z. B. P420 oder P_BT)
 - Abteilungsnummer (z. B. Abt7)
 - Kondition (z.B. für die Temperatur 50° Celsius: 50C)
 - Serie (z. B. Serie1)

Falls viele Dateien schon benannt sind, aber keiner Konvention folgen, kann ein Renamer¹⁵ verwendet werden. Mit diesem Tool können mehrere Dateien oder Ordner auf einmal umbenannt werden (University of Oregon, 2020b).

Eine Benennungskonvention ist laut der Website [forschungsdaten.info](https://www.researchdata.info) (2020g) das Herzstück einer funktionierenden Versionskontrolle. Mit der Versionskontrolle ist es möglich, den Überblick über Änderungen nicht zu verlieren (NYU Libraries, 2020b). Es wird empfohlen, eine finale Datei entsprechend zu kennzeichnen. Weiter empfiehlt sich das Erstellen von «Meilenstein-Versionen», die in regelmässigen Abständen an einem bestimmten Ort abgespeichert werden und nicht mehr verändert werden können/ dürfen. Das senkt das Risiko von Datenverlust durch das parallele oder kollaborative Arbeiten

¹⁴ Anschauliche Beispiele zur richtigen Verwendung von Elementen können auf der Website der University of Oregon (2020b) angesehen werden.

¹⁵ Weitere Information ist auf der jeweiligen Website erhältlich. Für Windows: <https://www.advancedrenamer.com/> Für Mac: <https://renamer.com/>

mehrerer Forscherinnen. Das Führen einer Versionstabelle ist eine ergänzende Massnahme, um Änderungen und neue Benennungen nachvollziehen zu können. Falls mit grösseren Mengen von Daten gearbeitet wird, ist es ratsam, anstatt einer Tabelle, eine Versionsverwaltungssoftware¹⁶ zu benützen (Forschungsdaten.info, 2020g).

4.3.3.5 Dokumentation, Metadaten & -standards

Auch die Dokumentation von Forschungsdaten leistet einen Beitrag zu deren Nachnutzung, Reproduzierbarkeit, Transparenz und Nachvollziehbarkeit (Forschungsdaten.info, 2020f; Universität Bern, o. J.):

«Je besser ein Datensatz dokumentiert ist, desto wahrscheinlicher ist es, dass dieser von Dritten genutzt und zitiert wird und die Datenerzeugerin oder der Datenerzeuger dadurch „credit“ in Form einer wissenschaftlichen Zitation erhält (Forschungsdaten.info, 2020f).»

Eine Datendokumentation nützt also nicht nur den potentiellen Nachutzern, sondern birgt auch Vorteile für die Datenproduzentinnen und Autoren der Publikation (Forschungsdaten.info, 2020f).

«Um das Bereitstellen von Forschungsdaten und ihre Nachnutzung zu ermöglichen, sind Metadaten und eine Datendokumentation essentiell, die den Kontext der Forschungsdaten beschreiben sowie die Werkzeuge, mit denen sie erzeugt, gespeichert, bearbeitet und analysiert wurden (Kindling & Schirnbacher, 2012, S. 130).»

Dieses Zitat erklärt schon etwas genauer, was die Dokumentation von Forschungsdaten beinhaltet. Grundsätzlich geht es bei der Dokumentation darum, die Geschichte der Forschungsdaten (Veränderungen, Konvertierung in andere Dateiformate, Umstrukturierung der Datensätze oder Anonymisierung personenbezogener Daten etc.) festzuhalten. Wenn vom Forschungsförderer oder der Forschungseinrichtung keine Vorgaben bestehen, kann man sich am Datenlebenszyklus (siehe Kapitel 4.3.1.3) orientieren (Forschungsdaten.info, 2020f, 2020d). Für ausführliche und langfristige Dokumentationen steht – von der Purdue Universität – ein Toolkit¹⁷ zur Verfügung, um ein Data-Curation Profile anzulegen. Ein Data-Curation-Profile gibt Auskunft über den

¹⁶ Eine Auflistung von open-source Versionsverwaltungssoftware kann hier https://de.wikipedia.org/wiki/Kategorie:Freie_Versionsverwaltungssoftware angesehen werden.

¹⁷ Das Toolkit bietet eine Vorlage, sowie Anleitung und ist abrufbar unter <https://docs.lib.purdue.edu/dcptoolkit/>. Beispiele für Data-Curation-Profiles können hier: <https://docs.lib.purdue.edu/dcp/> angesehen werden. Für die klinische Medizin ist leider kein Beispiel vorhanden, es gibt aber eins zum Thema «Abwehrsystem menschlicher Zellen».

Datensatz, -lebenszyklus, alle relevanten Massnahmen, die im Zusammenhang mit dem Management, Organisation, Sharing und Preservation getätigt wurden, sowie offenen Fragen und Probleme hinsichtlich dieser Massnahmen (Forschungsdaten.info, 2020d). Eine grundlegende Dokumentation kann folgende Elemente beinhalten (Forschungsdaten.info, 2020f; Jansen et al., 2019, S. 46–47; Universität Bern, o. J.; Yale Library, 2020a):

- Kontext des Forschungsvorhabens: Projektname/ -nummer und -ziele, Autorinnen/Personen, Hypothesen, Fragen
- Erhebungsmethode: Methode, Erhebungsort und -zeitraum/ -datum, Instrumente, Hard- und Software, sekundäre Quellen, etc.
- Struktur der Daten: Organisation (beispielsweise Tabelle, Graph, Klassen etc.), Beziehung untereinander
- Erklärungen: Glossar/ Wörterbuch, Abkürzungen, Codes und Codebuch
- Qualitätsmassnahmen: Bereinigung, Gewichtung, Prüfung, Information zu fehlenden Werten
- Versionen: Unterschiede der verschiedenen Versionen
- Zugang: Nutzungsbedingungen, Datenschutz

Diese Punkte liefern also Kontext, damit die Forschungsdaten und die Ergebnisse verstanden werden können (Oregon State University - Libraries, 2019b). Man spricht in diesem Zusammenhang auch von Metadaten. Metadaten sind einfach ausgedrückt «Daten über Daten» (Assfalg, 2014, S. 159; Yale Library, 2020a) und bezeichnen strukturierte, maschinenlesbare Information über Objekte (Lindenmann et al., 2019b, S. 2; Universität Bern, o. J.). Metadaten geben Auskunft über beispielsweise Verfasserin/Urheberin, Sprache, Datum, Art der Publikation, Titel und Schlagwörter, sowie Identifikator einer Ressource. Im klinischen Forschungsdatenmanagement können spezifische Metadaten beispielsweise Information über klinische Proben oder verwendete Medikamente, sowie verwendete Instrumente und Methoden sein. (Forschungsdaten.info, 2020i; Yale Library, 2020a).

Aus Nutzersicht tragen die Metadaten einen wichtigen Teil zum Finden, Durchsuchen, Benutzen und Zitieren der Forschungsdaten bei. Um diese Aufgaben zu erleichtern und der Interoperabilität gerecht zu werden, werden Metadatenstandards eingesetzt. Durch Metadatenstandards ist es möglich, Felder immer gleich zu interpretieren und somit die Verwertbarkeit zu automatisieren (Forschungsdaten.info, 2020j; Jansen et al., 2019, S. 47). Das heisst, diese Standards geben vor, welche Felder (beispielsweise Titel, Autor, Datum etc.) verwendet werden und mit welchen Inhalten diese Felder gefüllt werden

dürfen. Durch den Gebrauch von maschinenlesbarer Sprache, können diese (Meta-)Daten dann automatisch ausgelesen werden. Es gibt viele disziplinspezifische Metadatenstandards (Forschungsdaten.info, 2020i). Ein allgemeiner, bekannter Standard ist der von DataCite¹⁸. Dieser Standard ist open-source verfügbar (Oregon State University - Libraries, 2019b). Das Clinical Data Interchange Standards Consortium (CDISC) entwickelt Standards¹⁹ zur Unterstützung der Erfassung und des Austauschs von klinischen Forschungsdaten sowie Metadaten. Diese von CDISC entwickelten Standards sind plattformunabhängig und sollen die Interoperabilität zwischen Systemen ermöglichen (Clinical Data Interchange Standards Consortium, o. J.; Society for Clinical Data Management, 2013, S. 5, Data Management Standards in Clinical Research). Laut Muriel Helmers (M. Helmers, persönliche Kommunikation, 11. Juni 2020) werden diese Standards in der Industrie eingesetzt, sind aber in der Einführung aufwändig. Ein anderes Beispiel für einen Metadatenstandard ist der Dublin Core, der vom Repository «Bern Open Repository and Information System» (BORIS) verwendet wird (Universität Bern, o. J.). Weitere Standards können unter der Website <https://fairsharing.org/standards/> gesucht und gefunden werden. FAIRsharing.org bietet neben Standards auch Einträge zu Richtlinien und Datenbanken.

Zum Abschluss dieses Kapitels soll noch kurz auf kontrolliertes Vokabular eingegangen werden: «Controlled vocabularies [...] are [...] lists of predefined, authorized terms (Yale Library, 2020e).» Mit der Verwendung von kontrolliertem Vokabular ist das Auffinden von Publikationen einfacher, da nur autorisierte Begriffe verwendet werden dürfen. So fällt ein grosser Teil aller möglichen Bezeichnungen eines Begriffes weg und die Suche wird einfacher. Medical Subject Heading (MeSH) ist ein bekanntes kontrolliertes Vokabular, das für die Datenbank PubMed verwendet wird (Yale Library, 2020e).

4.3.3.6 Datenbanken & Zugriffsrechte

Früher oder später werden in klinischen Studien die (Forschungs-)Daten in einer Datenbank eingegeben und verwaltet (Prokscha, 2007, S. 19). Darum folgt hier eine kurze Einführung in das Thema Datenbanken.

«A database is simply a structured set of data. This may be an Excel spreadsheet, a Microsoft Access application, an indexed proprietary data format, a collection of

¹⁸ Weitere Informationen können hier: <https://schema.datacite.org/> eingesehen werden.

¹⁹ Weiter Angaben und Details zu den Standards sind unter <https://www.cdisc.org/standards> zu finden.

SAS tables, or a set of tables built in one of the traditional relational applications such as Oracle® (Prokscha, 2007, S. 19).»

ID	Temperatur [C°]	Puls [b/min]	Blutdruck [mmHg]
1	36.6	70	120/80
2	36.5	68	118/78
3	36.6	77	129/84

Primärschlüssel (rot markiert)
 Spalten = Attribute (grün markiert)
 Spaltenüberschrift = Feldnamen = Bezeichnung für Datenfelder/ Attribute (violett markiert)
 Zeilen = Datensätze = Tupel (blau markiert)
 Zellen = Datenfelder = Daten/ Werte (orange markiert)

Abbildung 2: eigene Abbildung, angelehnt an informatikZentrale (2010); Erklärung einer einfachen Datenbank

Man unterscheidet die Begriffe Datenbank (zu verwaltende Datenmenge), Datenbankmanagementsystem (Software, mit der verwaltet wird) und Datenbanksystem (Softwaresystem, bestehend aus Datenbank und Datenbankmanagementsystem). Das relationale Datenbankmodell ist das verbreitetste: Dort werden die Daten in einer Tabellenstruktur abgelegt (informatikZentrale, 2010). Das oben aufgeführte Zitat von Prokscha beschreibt also ein relationales Datenbankmodell. Abbildung 2 zeigt eine einfache Darstellung einer Datenbank mit Tabellenstruktur. Spalten werden Attribute genannt, eine Zeile entspricht einem Datensatz (auch Tupel genannt). In einer Zeile befinden sich Werte. Jeder Datensatz/ Tupel muss eindeutig bestimm- und von den anderen unterscheidbar sein. Dazu wird jedem Datensatz ein einzigartiger Wert, den sogenannten Primärschlüssel, zugeordnet (informatikZentrale, 2010). Im Beispiel unten sind die Nummern eins bis drei, es könnten aber beispielsweise auch alphanummerische Codes sein.

Zur Verwaltung der klinischen Forschungsdaten werden Clinical Data Management Systems (CDMS) genutzt. Beispiele dafür, die auch die CTU Bern betreut, sind REDCap²⁰ und SecuTrial (CTU Universität Bern, 2018). Research Electronic Data Capture (REDCap) wurde von der Vanderbilt University, Nashville, entwickelt, ist kostenlos und Good Clinical Practice-konform. Zurzeit wird es in über 130 Ländern von

²⁰ Hier geht's zur Homepage von REDCap: <https://projectredcap.org/>

über 4'300 Institutionen verwendet und verzeichnet 1,3 Millionen Nutzer. Mit REDCap ist es möglich Studiendatenbanken, electronic Case Report Forms und Surveys etc. zu erstellen (REDCap, o. J.; Universität Bern, 2017). Ausserdem ist REDCap – im Gegensatz zu einer Excel-Tabelle oder Access – konform mit dem Humanforschungsgesetz und der Verordnung über klinische Versuche in der Humanforschung (KlinV) (CTU Universität Bern, 2018). Artikel 18 Absatz 1 lit. a bis c der KlinV besagen:

«1 Wer im Rahmen eines klinischen Versuchs gesundheitsbezogene Personendaten aufbewahrt, muss deren Schutz durch geeignete betriebliche und organisatorische Massnahmen sicherstellen, namentlich:

- a. den Umgang mit den gesundheitsbezogenen Personendaten auf diejenigen Personen beschränken, die diese Daten zur Erfüllung ihrer Aufgaben benötigen;
- b. die unbefugte oder versehentliche Offenlegung, Veränderung, Löschung und Kopie der gesundheitsbezogenen Personendaten verhindern;
- c. alle zur Gewährleistung der Rückverfolgbarkeit massgeblichen Bearbeitungsvorgänge dokumentieren.»

Daraus ergibt sich die Notwendigkeit von Zugriffsbeschränkungen. Für die Einhaltung von Absatz a ist ein personalisiertes Login (mit Passwort) und für Absatz b eine Zugriffskontrolle in Form von Rollen erforderlich (CTU Universität Bern, 2018, S. 20). Diese Massnahmen ermöglichen unbefugte Zugriffe zu verhindern, indem beispielsweise Rollen ohne Bearbeitungsrechte erstellt werden oder bestimmte Daten nur mit einer höheren Rolle eingesehen werden können etc. Den verschiedenen Rollen werden jeweils andere Berechtigungen zugewiesen, so kann verhindert werden, dass die Daten nicht von Unbefugten manipuliert werden können. Die Notwendigkeit von Rollen, Logins und Zugangskontrollen werden auch in Prokscha (2007, S. 133–138) besprochen. Zur Gewährleistung der Rückverfolgbarkeit müssen Audit Trails eingesetzt werden. Audit Trails zeichnen vorgenommene Änderungen an den klinischen Daten auf. Diese Änderungen müssen zu jeder Zeit abgerufen werden können (CTU Universität Bern, 2018, S. 20; Prokscha, 2007, S. 49). Diese Kontrollmechanismen hängen auch mit den rechtlichen Aspekten aus Kapitel 4.3.2 zusammen.

4.3.3.7 Datenqualität

«High quality clinical research data provide the basis for conclusions regarding the safety and efficacy of a medical treatment (Society for Clinical Data Management, 2013, S. 1, Assuring Data Quality).» Letztendlich fällt also die Qualität der Daten auf die Qualität der Forschungsergebnisse zurück. Darum ist es laut der Website der Yale Library (2020b)

auch wichtig, den Entstehungszusammenhang zu kennen und sich bei Recherchen zu fragen, wer das Forschungsprojekt geführt oder finanziert hat, ob die Daten komplett sind, oder, ob (wichtige) Daten weggelassen wurden und, ob diese Daten genutzt (Lizenz, Zitation) werden dürfen. Die Begriffe «Qualitätssicherung» und «Qualitätskontrolle» werden oft als Synonyme verwendet, lassen sich aber laut der Website der American Society for Quality unterscheiden: Unter der Qualitätssicherung werden alle planerischen und strategischen Massnahmen zusammengefasst, die Qualitätskontrolle hingegen schliesst alle operativen Massnahmen ein (American Society for Quality, o. J.). Konkrete Beispiele für die Qualitätssicherung sind (DataONE, o. J.-c; Yale Library, 2020b):

- Ressourcen bereits im Fördergesuch berücksichtigen
- Vollständige Dokumentation sicherstellen
- Manuelle Überprüfung der Daten (bis zu 10%)
- Terminologie/ kontrolliertes Vokabular verwenden
- Standards definieren (Metadaten, Datenformat, Codes)
- Verantwortliche Person bestimmen

Massnahmen, die bei der Qualitätskontrolle ergriffen werden können sind (DataONE, o. J.-c; Yale Library, 2020b):

- Daten doppelt eingeben, falls verschiedenen Personen mit der Eingabe der Daten betraut sind und das System die Überprüfung und Fehleridentifizierung vornimmt
- Standardisierte Datenformate verwenden, wie YYYYMMDD etc.
- Mit Hilfe von Diagrammen oder statistischen Mitteln Ausreisser/ Extremwerte identifizieren, um mögliche Fehler auszuschliessen

Um die Qualität von Daten zu bewerten, können folgende Kriterien hinzugezogen werden (Yale Library, 2020b):

- Vollständigkeit: Datensätze sollten so vollständig wie möglich sein. Ziel ist es keine leeren Werte zu haben.
- Eindeutigkeit: Jeder Datensatz sollte einzigartig sein, das heisst, Duplikate müssen eliminiert werden/ sein.
- Aktualität: Die Daten sollten zeitnah und bedeutsam für die Gegenwart sein.
- Gültigkeit: Die Daten sollten gültig sein. Damit ist gemeint, dass die Daten nicht unmöglich sind, beispielsweise Blutwerte liegen in einem Bereich, die gar nicht möglich sein können.

- Genauigkeit: Die Daten sollten möglichst genau der Realität entsprechen, z. B. ein Sprint nicht manuell mit der Stoppuhr messen, sondern mit Lichtschranken.
- Konsistenz: Die Daten sind widerspruchsfrei, das heisst, Daten in einer Spalte werden immer auf dieselbe Weise oder nach demselben Standard erfasst oder berechnet.

Neben diesen aufgeführten Kriterien gibt es noch «Authentizität», «Integrität» und «Vertraulichkeit». Diese Begriffe werden in Kapitel 4.3.4.1 «Datenspeicherung & -sicherheit» behandelt.

4.3.3.8 Datenverarbeitung

Laut DSGVO Art. 4 Ziff. 2 beschreibt die Datenverarbeitung Aktivitäten, «wie das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung [...]» von personenbezogenen Daten. Zozus et al. (2017, S. 740) schreiben im Kontext des klinischen Forschungsdatenmanagements folgende Aufgaben der Datenverarbeitung zu: Daten sammeln, eingeben, verknüpfen, abstimmen, transformieren, codieren, randomisieren, exportieren und archivieren. Eine weitere Aktivität ist das Identifizieren von Diskrepanzen. Weil diese Aktivität von Berufseinsteigerinnen im klinischen Forschungsdatenmanagement besonders häufig durchgeführt wird (Zozus et al., 2017, S. 743), folgt eine kurze Vertiefung dieser Aufgabe.

Die GCDMP (Society for Clinical Data Management, 2013, S. 10, Glossary) definieren den Begriff «Diskrepanz» wie folgt: «Inconsistency in two or more data points collected in a clinical trial that must be addressed [...]» Ein konkretes Beispiel dafür: Wie im Kapitel Datenqualität erwähnt, werden – zur Kontrolle derselben – Daten doppelt eingegeben. Wenn also bei der zweiten Eingabe ein oder mehrere Werte nicht mit der ersten Eingabe übereinstimmen, wird eine Warnung ausgegeben (Society for Clinical Data Management, 2013, S. 7, Data Entry Processes). Eine Diskrepanz kann aber auch bedeuten, dass ein Wert ganz fehlt, beispielsweise das Geburtsdatum oder die Dosis eines Medikamentes (Prokscha, 2007, S. 62). Es ist auch möglich, dass eine dritte Person die Sichtung manuell macht und so die Inkonsistenzen identifiziert (Society for Clinical Data Management, 2013, S. 7, Data Entry Processes). Mittels Edit Checks können die Diskrepanzen automatisch identifiziert werden (Society for Clinical Data Management, 2013, S. 10, Design and Development of Data Collection Instruments). Die GCDMP

(Society for Clinical Data Management, 2013, S. 11, Design and Development of Data Collection Instruments) beschreiben Edit Checks wie folgt:

«Edit checks are intended to ensure data integrity and improve data quality by bringing attention to data that are out of the expected range, inconsistent, illogical or discrepant. When data meet the predefined criteria of an edit check, a flag or warning notifies CDM personnel that the data point should be carefully examined to ensure the accuracy of the data point.»

Die Identifizierung dieser Diskrepanzen wird dem Prozess der Datenverarbeitung zugesprochen (vgl. Zozus et al., 2017, S. 740), trägt aber wesentlich zur Datenqualität bei. Die Edit Checks sind ein wichtiges Werkzeug, um Inkonsistenzen aufzudecken und sind ein Erfolgsfaktor, um die Qualität der Daten während der Datenverarbeitung zu gewährleisten, sowie Studien schneller zum Abschluss zu führen (Prokscha, 2007, S. 62 & 73).

4.3.3.9 Verantwortlichkeit

Zum Abschluss des Kapitels Datenmanagement und -organisation wird noch geklärt, wer die Verantwortung für das Management trägt. Jansen et al. (2019, S. 39) schreiben dazu:

«As a clinical researcher, you are the principal data steward. In practice, this means that you are responsible for the complete scientific process: from study design to data collection, analysis, storage, and sharing. Protecting the privacy of study subjects is also your responsibility.»

Weiter wird erklärt, dass die formale Verantwortlichkeit aber beim Forschungsinstitut liegt. Das hiesst, die Forschungseinrichtung muss garantieren, dass die DSGVO eingehalten wird und Fachwissen (einschliesslich entsprechender Richtlinien für die Forschenden) bezüglich dem Management von Forschungsdaten vorliegt (Jansen et al., 2019, S. 39).

4.3.4 Preservation

Dieses Kapitel behandelt die Unterthemen Datenspeicherung & -sicherheit, Literaturverwaltung und Langzeitarchivierung. Insbesondere die Datensicherheit inklusive Back-up sind zentral.

4.3.4.1 Datenspeicherung & -sicherheit

Forschungsdaten sind in verschiedenen Dateiformaten, auf diversen unterschiedlichen Medien und Trägermedien oder Speicherorten zu finden. Die Speichermedien und -orte

weisen Vor- und Nachteile auf und können sogar Sicherheitsrisiken darstellen (Forschungsdaten.info, 2020m). Wo Forschungsdaten gespeichert werden, sowie welche Sicherheitsrisiken bestehen, soll im Folgenden geklärt werden. Geeignete Dateiformate wurde bereits in Kapitel 4.3.3.2 behandelt und werden in Kapitel 4.3.4.3 «Langzeitarchivierung» nochmals aufgegriffen.

Um Daten vor Verlusten zu schützen, ist es zwingend notwendig – in regelmässigen Abständen – Backups (=Sicherheitskopien) zu machen. Dazu wird die 3-2-1 Regel empfohlen: 3 Kopien – das Original und zwei Sicherheitskopien – aufbewahren, die zwei Sicherheitskopien sollten auf 2 unterschiedlichen Medien (beispielsweise Cloud und externe Festplatte), und 1 dieser beiden Sicherheitskopien an einem externen Speicherort (eben beispielsweise die Cloud) abgelegt werden (Universität Bern, o. J.; University of Oregon, 2020a). Folgende Liste gibt einen Überblick über häufig eingesetzte Speichermedien und deren Lebensdauer (Forschungsdaten.info, 2020m):

- Festplatte Computer, Laptop: Lebensdauer von durchschnittlich fünf bis zehn Jahren, können aber auch bis zu 30 Jahre halten.
- Externe Festplatte: weil externe Festplatten eher weniger häufig genutzt werden, ist die Lebensdauer etwas höher, als die der fest verbauten. Eine externe Festplatte sollte problemlos zehn Jahre halten.
- USB-Stick (USB steht für Universal Serial Bus) und Secure Digital-Karte (SD-Karte): Halten ebenfalls circa zehn Jahre.
- Compact Disc (CD), Digital Versatile Disc (DVD), Blu-ray Disc: Eine CD hält ebenfalls bis zu zehn Jahre, eine DVD bis zu 30 Jahren und eine gepresste DVD oder Blu-ray Disc bis zu 80 Jahren.

Die Lebensdauer kann je nach Nutzung, Qualität des Mediums und äusseren Einflüssen schwanken (Forschungsdaten.info, 2020m). Die genannten Jahre sind also als Richtwerte zu verstehen. Speicherorte sind Festplatten (verbaut im PC / Laptop), Laufwerke, Netzwerke und Clouds. Wie die Website [forschungsdaten.info](https://www.forschungsdaten.info) (2020m) weiter ausführt, stellen diese Speicherorte Sicherheitsrisiken dar, vor allem der unberechtigte Zugriff über Inter- oder Intranet und der Verlust von Daten aufgrund defekter Speicher stellen ein grosses Risiko dar. Als Prävention ist es ratsam, keine offenen Wireless Local Area Network (WLAN) zu verwenden und auf keine unbekanntem Links, Websites oder Anhänge in E-Mails zu klicken. Ausserdem kann ein Antivirus-Programm empfohlen werden. Damit hängt auch zusammen, dass die verwendete Software immer aktuell ist – sprich, alle Updates sind installiert (Forschungsdaten.info, 2020m). In Kapitel 4.3.3.6 wurden die Zugriffsrechte – Login und Rollen – bereits

behandelt, die spielen in der Datensicherheit ebenfalls eine wichtige Rolle: Diese Sicherheitsvorkehrungen im Zusammenhang mit der Zugriffsbeschränkung stellen die Authentizität – Daten sind echt und können dem Ursprung zugeordnet werden; Vertraulichkeit – keine unbefugter Zugriff; und Integrität – damit ist gemeint, dass die Daten nicht unerkannt verändert wurden – der Forschungsdaten sicher (Prokscha, 2007, S. 133; Weis, 2018). In den GCDMP (Society for Clinical Data Management, 2013, S. 2, Data Storage) wird ergänzt, falls Forschungsdaten auf Papier vorliegen sollten, müssen diese auch an sicheren Orten gelagert werden, bei denen der unbefugte Zugriff verwehrt bleibt.

Falls Clouds zum Speichern der Daten verwendet werden, muss unbedingt darauf geachtet werden, in welchem Land die Server des Providers stehen. Mängel bestehen dort oft bei der Verschlüsselung und dem Datenschutz. Gerade im Hinblick auf personenbezogene Daten muss sichergestellt werden, dass die datenschutzrechtlichen Bestimmungen eingehalten werden (Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB), o. J.-b; Forschungsdaten.info, 2020m).

Auch die sichere Speicherung und Verwaltung von bibliografischen Daten ist wichtig. Das Thema wird im folgenden Kapitel behandelt.

4.3.4.2 Literaturverwaltung

Czeschik (o. J.) erklärt, dass ein Literaturverwaltungsprogramm für das wissenschaftliche Arbeiten eine enorme Hilfe darstellt. Ein Literaturverwaltungsprogramm, oder auch Referenzmanager genannt, speichert Referenzen und die dazugehörigen Artikel. Es hilft zudem diese Referenzen zu organisieren, zu durchsuchen, sowie Referenzen und Literaturverzeichnisse (in verschiedenen Zitierstilen) in Dokumenten anzulegen. Die Einträge im Programm müssen aber nicht manuell eingegeben werden: Der Import von Referenzen und dazugehörendem Artikel kann direkt von der jeweiligen Website aus veranlasst werden – es stehen Plugins für verschiedene Browser zur Verfügung. Ausserdem machen manche Literaturverwaltungsprogramme das kollaborative Arbeiten leichter in dem der Zugriff auf das Programm freigegeben werden kann. Auch für das zitieren in Dokumenten (wie Word oder LibreOffice) stehen Plugins zur Verfügung Beispiele für Literaturverwaltungsprogramme²¹ sind: Zotero; Mendeley; Citavi; Endnote und RefWorks (Czeschik, o. J.). Extra-Tipp: Das kontinuierliche Dokumentieren der

²¹ Eine kompakte Übersicht dieser Programme kann unter <https://www.lehmans.ch/page/literaturverwaltungsprogramm> eingesehen werden.

verwendeten Literatur erleichtert das korrekte Zitieren ungemein. Mehr dazu folgt im Kapitel 4.3.5.2.

4.3.4.3 Langzeitarchivierung

«Langzeitarchivierung (LZA) von Daten soll deren langfristige Nutzbarkeit über einen nicht definierten Zeitraum hinweg sicherstellen (Forschungsdaten.info, 2020q).» Weil die Zukunft insbesondere in technischer Hinsicht ungewiss ist, muss die Nutzbarkeit dieser Daten regelmässig überprüft werden: Dateiformate können obsolet werden und sind somit nicht mehr nutzbar. Mit der Langzeitarchivierung wird zudem sichergestellt, dass die Daten authentisch, integer, zugänglich und verständlich sind. Dazu gibt es zwei verschiedene Ansätze: die Emulation und die Migration. Bei der Emulation wird alte Software auf einem neuen Gerät simuliert, damit proprietäre Dateiformate/ Software (beispielsweise Microsoft Office-Palette oder Adobe Photoshop) trotzdem noch eingesehen werden können – diese Methode ist aber sehr aufwändig. Einfacher geht es, wenn Daten migriert werden können. Die Migration beschreibt das Transferieren von Daten auf einen anderen Träger oder Speicherort. Das bedingt, dass die Dateiformate nicht proprietär sind (Forschungsdaten.info, 2020q). Wie aus Kapitel 4.3.3.2 hervorgeht, sind offene, nicht proprietäre Dateiformate zu präferieren. Das zeigt sich also auch in der Langzeitarchivierung. Für beide vorgestellten Strategien – Emulation und Migration – gilt, dass die Forschungsdaten mit Metadaten, sprich Kontext und Bedeutung angereichert werden, damit sie für spätere Nutzer verständlich sind (Forschungsdaten.info, 2020q). Bei der Langzeitarchivierung müssen selbstverständlich auch die in den Kapiteln Datensicherheit und Datenschutz besprochenen Themen berücksichtigt werden. Als Langzeitarchiv können Repositorien (auch Dokumentenserver genannt) dienen. Dazu mehr im Kapitel 4.3.5.3.

Mit der Archivierung von Daten geht die Frage «Welche Daten sollen aufbewahrt werden?» einher. Nach Abschluss eines Forschungsprojektes müssen die gesammelten Daten bewertet werden, um zu entscheiden, welche Forschungsdaten behalten werden. Das ist notwendig, weil die technische Infrastruktur knapp ist/wird und die Bewirtschaftung dieser Daten finanziell nicht lukrativ ist: Es können schlichtweg nicht alle Daten aufbewahrt werden (Forschungsdaten.info, 2020c). Grob betrachtet, müssen fünf Punkte geklärt werden, um zu bestimmen, ob Daten aufbewahrt werden sollen oder nicht (Forschungsdaten.info, 2020c):

1. Der Wiederverwendungszweck sollte gegeben sein. Mit den Daten sollte der Forschungsprozess nachvollzogen werden können.

2. Gibt es Daten, die aufbewahrt werden müssen? Die Gute Klinische Praxis beispielsweise hält in § 13 Absatz 10 fest, dass alle wesentlichen Unterlagen noch zehn Jahre nach Beendigung des Projektes aufbewahrt werden müssen.
3. Es muss geklärt werden, ob Interessenskonflikte bestehen: Bei klinischen Forschungsdaten bedeutet das, dass der Datenschutz in Bezug auf personenbezogene Daten eingehalten werden muss – die Daten müssen anonymisiert werden.
4. Daten, die einen langfristigen Wert aufweisen, sind besser zu bewerten.
5. Die Kosten müssen abgewogen werden. Nicht nur der Speicherplatz kostet Geld, sondern auch die Bewirtschaftung (Kontrolle Nutzbarkeit, Dateiformate konvertieren etc.).

Weiterführende und vertiefte Information zur Auswahl und Bewertung kann unter <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/auswaehlen-und-bewerten/> eingesehen werden.

4.3.5 Publikation, Re-Use & SNF

Dieses Kapitel behandelt die Unterthemen Urheberrecht & Lizenzen, Sharing, Nachnutzung & Zitation, Repositorien, Open Access & Open Data, sowie die FAIR-Prinzipien und den SNF.

4.3.5.1 Urheberrecht & Lizenzen

Die Veröffentlichung von Forschungsdaten verlangt eine Lizenzierung dieser Daten (Digital Curation Centre, 2020c), denn nur der Eigentümer hat das Recht die Daten zu kopieren, zu verändern oder zu verbreiten (University of Oxford, 2014). Die beiden häufigsten Rechte sind das Urheberrecht und das Leistungsschutzrecht von Datenbankherstellern (University of Oxford, 2014): Letzteres kann bei Repositorien angewendet werden, schliesst das Urheberrecht nicht aus und erlaubt dem Ersteller das alleinige Recht der Verbreitung und Vervielfältigung der Datenbank (Forschungsdaten.info, 2020n). «Der urheberrechtliche Schutz entsteht mit der Schöpfung des Werkes automatisch, eine Kennzeichnung oder gar Anmeldung wie bei Patenten ist nicht erforderlich (Forschungsdaten.info, 2020b).» Und weiter: «Das Urheberrechtsgesetz schützt die Urheber von Werken der Literatur und Kunst. Geschützt ist die Art und Weise, in welcher eine Idee zum Ausdruck gelangt, nicht aber die Idee oder das Konzept selbst (Eidgenössisches Institut für Geistiges Eigentum, o. J.).» Das heisst nur die Darstellung der Forschungsdaten ist urheberrechtlich geschützt, das

gleiche gilt für die daraus produzierten Ergebnisse: die Resultate sind nicht geschützt, nur die Form, also der Text oder das Bild (Kuschel, 2018). Ein publizierter Artikel zu einer klinischen Studie darf also verwendet werden, aber nicht mit denselben Worten, wie im Original-Artikel. Das Eidgenössische Institut für Geistiges Eigentum (o. J.) führt weiter aus, dass das Urheberrecht noch bis 70 Jahre nach dem Tod des Urhebers gilt. Wenn also die Forschungsergebnisse und dazugehörenden -daten weiterverwendet werden dürfen/ sollen, ist es ratsam sich mit Creative Commons Lizenzen auseinanderzusetzen. Diese regeln, zu welchen Konditionen die Werke (beispielsweise Artikel, Foto, Buch) weiterverbreitet und gebraucht werden dürfen (Jansen et al., 2019, S. 43). Ausserdem sorgen diese Lizenzen dafür, dass der Urheber Anerkennung für seine Arbeit erhält (Creative Commons, 2017). Es bestehend folgende Lizenzen²² (Creative Commons, 2017; DataONE, o. J.-d):

- CC BY – Namensnennung: Diese Lizenz erlaubt das Werk zu vervielfältigen, zu kopieren und zu verändern – sogar kommerziell. Das Original muss aber anerkannt/referenziert werden.
- CC BY-SA – Namensnennung – Weitergabe unter gleichen Bedingungen: Auch diese Werke dürfen (zu kommerziellen Zwecken) vervielfältigt, kopiert und verändert werden. Bedingung ist, dass auf das Original verwiesen wird und das neue Werk zu den gleichen Konditionen (also CC BY-SA) lizenziert wird. Diese Lizenz wird auch copyleft genannt.
- CC BY-ND – Namensnennung – Keine Bearbeitung: Werke dürfen auch kommerziell verwendet werden unter der Bedingung, dass der Urheber genannt wird. Die Lizenz untersagt aber das Verändern des Werkes.
- CC BY-NC – Namensnennung – Nicht kommerziell: Werke dürfen adaptiert, kopiert, verändert und veröffentlicht werden, es muss auf den Urheber verwiesen werden und das Werk darf nicht zu kommerziellen Zwecken verwendet werden.
- CC BY-NC-SA – Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen: Zusätzlich zur vorherigen Lizenz muss das neue Werke unter der gleichen Lizenz laufen, wie das Original.
- CC BY-NC-ND – Namensnennung – Nicht kommerziell – Keine Bearbeitung: Die Lizenz erlaubt nur den Download und Weitergabe des Werkes, es darf jedoch nicht

²² Weitere Information zu Lizenzen sind verfügbar unter: <https://www.forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/freie-lizenzen-vs-gemeinfrei/> und <https://www.forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/creative-commons-lizenzen/>.

verändert oder adaptiert werden. Ausserdem muss der Urheber referenziert und das Original darf nicht zu kommerziellen Zwecken verwendet werden.

Erläuterung zu den Akronymen: SA steht für Share Alike, NC für Non-Commercial und ND für No Derivatives. Für Forschungsdaten ist die CC BY Lizenz geeignet (Forschungsdaten.info, 2020a).

Es muss berücksichtigt werden, dass Forschungsdaten, die in einem Arbeitsverhältnis an einer Hochschule produziert werden, nicht nur dem Urheber, also der Forscherin, Assistenten oder einem Mitglied des Forschungsteams, gehören – das gilt auch für Studierende. Die Hochschule hat ein Nutzungsrecht an den Daten und kann die Forschungsdaten kopieren und die Nachnutzung für andere erlauben (Kuschel, 2018). Auch Sponsorinnen und Forschungsförderer können Nutzungsrechte besitzen (DataONE, o. J.-d).

Weiterführende Information zum Urheberrecht von Forschungsdaten, sowie zum Leistungsschutzrecht des Datenbankherstellers sind unter <https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/> verfügbar.

4.3.5.2 Sharing, Nachnutzung & Zitation

«Research data remain a valuable resource, even after the life of the research project for which they were collected and used. Sharing the research outcomes enables future researchers to open up new lines of enquiry without the duplication of effort involved in collecting the data again (University of Oxford, 2013b).»

Darüber hinaus verlangen Forschungsförderer die Veröffentlichung von Forschungsdaten (siehe dazu Kapitel 4.3.1.1 und 4.3.3.1). Weitere Gründe, die für die Veröffentlichung von Forschungsdaten sprechen, sind die Stärkung der Integrität, sowie Förderung von Open Science, Kooperationen und Weiterentwicklung und Nachnutzung von Daten und Erkenntnissen. Forschungsdaten können über Repositorien oder Datenjournale veröffentlicht werden. Ein Datenjournal²³ enthält lediglich eine Beschreibung der Forschungsdaten und referenziert auf ein Repository, das die Datensätze beinhaltet (Forschungsdaten.info, 2020e; Universität Bern, o. J.).

Bei der Publikation von klinischen (Forschungs-)Daten müssen einige Dinge beachtet werden: Wie in Kapitel 4.3.2.1 besprochen, müssen die Einwilligungen der

²³ Beispiele für Datenjournale der Medizin sind Open Health Data: <https://openhealthdata.metajnl.com/> und GigaScience: <https://academic.oup.com/gigascience>. Weitere Datenjournale können unter <https://www.cms.hu-berlin.de/de/dl/dataman/teilen/dokumentation/datajournal> eingesehen werden.

Studienteilnehmerinnen vorliegen und die Daten müssen anonymisiert werden, bevor sie veröffentlicht werden dürfen – sollten die Daten trotzdem nicht veröffentlicht werden dürfen, beispielsweise aus Gründen des Datenschutzes, muss das im DMP vermerkt werden (Schweizerischer Nationalfonds, o. J.-b). Ausserdem muss ein Repository oder Datenjournal ausgewählt werden, das zur Fachdisziplin gehört – wo Repositorien der Medizin gefunden werden können, wird im nächsten Kapitel erläutert. Das Repository sollte langfristig verfügbar sein, adäquaten Datenschutz und -sicherheit bieten und eine Infrastruktur aufweisen, die es erlaubt, die Forschungsdaten zu finden. Es sollten überdies die verwendeten Metadatenstandards, Lizenzen und Datenüberlassungs- und Datennutzungsvereinbarungen überprüft werden (Forschungsdaten.info, 2020e). Des Weiteren müssen alle bis jetzt besprochenen Schritte wie Dokumentation der Daten durch Metadaten und Auswahl der Dateiformate etc. bedacht und ausgeführt sein, damit der Nachnutzung der Forschungsdaten nichts im Wege steht.

«Gutes Forschungsdatenmanagement zeichnet sich unter anderem dadurch aus, dass wissenschaftliche Datensätze für Dritte leicht nachnutzbar sind. Hierbei geht es vor allem um einheitliche Standards und Formate und das Versehen der Datensätze mit genügend Metadaten. Dies allein reicht jedoch nicht aus, denn vor der Nachnutzung muss ein potentieller Interessent die Daten auch finden können (Forschungsdaten.info, 2020j).»

Die Lösung dafür heisst persistente (dauerhafte) Identifikatoren: Persistente Identifikatoren sind digital und werden einem digitalen Objekt (beispielsweise Datensatz, Artikel, Buch) zugeordnet. Ein Beispiel dafür sind Uniform Resource Locators (URL). URLs sind für digitale Objekte, wie Artikel oder Datensätze, aber nicht wirklich geeignet, weil sie «nicht auf einen bestimmten Inhalt, sondern auf einen ‚Ort‘ im Internet» verweisen (Forschungsdaten.info, 2020j). Das führt zu Unzuverlässigkeit, weil das Objekt an mehreren Orten abgelegt ist oder das Objekt an einen anderen Ort verschoben werden kann und der URL somit ungültig ist. Darum hat sich der Standard Digital Object Identifier (DOI) etabliert. Mit dem DOI²⁴ können Objekte eindeutig und dauerhaft lokalisiert werden (Forschungsdaten.info, 2020j).

Es gibt auch ein Pendant für Personen: Open Research Contributor Identification Initiative (ORCID). Mit diesem Identifikator können Personen – unabhängig von

²⁴ Die Organisation DataCite vergibt DOIs für Datensätze (es gibt aber noch andere Institutionen). Weitere Information sind unter <https://datacite.org/index.html> einsehbar.

Schreibvarianten oder Forschungsinstitution – eindeutig identifiziert werden. Die eigene ORCID kann man unter <https://orcid.org/> durch eine Registrierung kostenlos erstellen (Forschungsdaten.info, 2020j; Universität Bern, 2019b).

DOIs helfen – durch das exakte Identifizieren von Objekten – auch beim Zitieren von Daten. Die Organisation DataCite (o. J.) erklärt weiter, dass bei der Nutzung von Publikationen mindestens folgende Elemente miteinander verbunden werden müssen: Autorin; Publikationsjahr; Titel; Verleger/ Herausgeberin; Identifikator; und optional Version und Ressourcen Typ. Das bedeutet aber auch, dass diese Metadaten vom Autor der Publikation/ des Datensatzes erfasst und zur Verfügung gestellt werden müssen. Zum Abschluss sei noch erwähnt, dass es viele verschiedene Zitationsstile gibt und je nach Forschungsdisziplin andere Standards verwendet werden. Für die Medizin empfiehlt Nagel (2019) den Zitationsstil «Vancouver». In Kapitel 4.3.4.2 wurden die Literaturverwaltungsprogramme bereits behandelt, es empfiehlt sich ein solches zu verwenden, um den gesamten Prozess des Zitierens zu vereinfachen.

4.3.5.3 Repositorien, Open Access & Open Data

«Eine Open-Access-Publikation erlaubt den Nutzerinnen und Nutzern einer wissenschaftlichen Publikation, das publizierte Dokument kostenlos zu lesen, herunterzuladen, zu speichern, es zu verlinken und zu drucken. Es darf also entgeltfrei genutzt werden. Open Access schließt auch die Zugänglichmachung von wissenschaftlichen Primär- und Metadaten, Quellentexten und digitalen Reproduktionen ein (Forschungsdaten.info, 2020t).»

Open Access steht also für den freien, kostenlosen Zugang zu Daten und Information (können gedruckt oder online sein), es müssen «nur» die Lizenzbedingungen (siehe Kapitel 4.3.5.1) eingehalten werden. Leider werden Forschungsdaten noch relativ selten veröffentlicht, weil sich solche Daten nicht wirklich für den Druck eignen (Forschungsdaten.info, 2020t). Ein anderer Grund sind die Forschenden, die diesen Prozess blockieren, in dem sie aus Angst vor Kontrollverlust ihre Forschungsdaten nicht publizieren. Konkreter befürchten Forscherinnen, dass ihre Daten missbräuchlich verwendet, fehlinterpretiert oder, dass sie sogar für ihre eigenen Interpretationen kritisiert werden (Fecher & Friesike, 2014).

Neben Open Access gibt es noch weitere Bewegungen, wie Open Source, Open Government und Open Data. Open Access fokussiert sich im Gegensatz zu Open Data auf ausformulierte Texte und Artikel. Forschungsdaten passen demnach besser zu Open Data. Das Ziel ist bei Open Access und Open Data aber dasselbe: die freie Nutzbarkeit von Daten soll vorangetrieben werden (Forschungsdaten.info, 2020t). Zusammengefasst

kann gesagt werden, dass Open Access und Open Data die Nachnutzbarkeit von (Forschungs-)Daten fördern.

Im Kapitel zum Thema Sharing und Nachnutzung wurde schon erwähnt, dass Repositorien als Archiv und Veröffentlichungsplattform dienen können. Ein Repository besteht aus einer Software und einer Datenbank, ist also ein Speicherort für Publikationen oder Forschungsdaten, je nachdem auf welche Art von Daten das Repository ausgelegt ist. Damit Nutzer Publikationen oder Datensätze in einem Repository finden, nutzen und zitieren können, benötigt es Metadaten (inkl. Angaben zur Lizenz), eine Suchfunktion und persistente Identifikatoren (Forschungsdaten.info, 2020r). Es wurde auch schon thematisiert, dass bei der Auswahl eines geeigneten Repositoriums einige Dinge beachtet werden müssen. Ein Aspekt, der noch nicht aufgegriffen wurde, ist die Anforderung des SNF oder eines anderen Forschungsförderers: Der SNF verlangt für die Archivierung und Veröffentlichung von Forschungsdaten ein Datenarchiv, das nicht kommerziell ist (Schweizerischer Nationalfonds, o. J.-a). Ein solches Repository ist zum Beispiel Zenodo. Zenodo ist ein interdisziplinäres Repository und entspricht den Anforderungen des SNF, sowie den FAIR-Prinzipien (werden im nächsten Kapitel behandelt). Es enthält Artikel, Software, Datensätze, Videos und weitere Arten von Publikationen, betrieben wird es vom CERN (CERN, o. J.; Schweizerischer Nationalfonds, o. J.-a). Ein Überblick über Repositorien für Forschungsdaten bietet der Dienst «Registry of Research Data Repository²⁵» (Forschungsdaten.info, 2020r; Schweizerischer Nationalfonds, o. J.-a). Für die Medizin sind derzeit über 580 Einträge verfügbar. Ein Beispiel für den Teilbereich «Public Health, Health Service Research, Social Medicine» mit Filter «Switzerland» ist das «Sicas Medical Image Repository²⁶». Es enthält medizinische Forschungsdaten. Abschliessend soll noch das Repository BORIS erwähnt werden: BORIS wird von der Universität Bern betrieben und erfüllt laut eigenen Angaben (Universität Bern, o. J.) die Anforderungen des SNF. Für grosse Datensätze ist dieses Repository aber nicht geeignet. Zur Zeit wird ein Forschungsdatenrepository (BORIS Research Data) aufgebaut (Universität Bern, o. J.).

4.3.5.4 Erwartungen SNF & FAIR-Prinzipien

Wie schon einige Male ausgeführt, sind Forschungsförderer ein zentraler Grund, wieso Forschungsdaten veröffentlicht und gemanagt werden. Der SNF erwartet, dass alle

²⁵ Link dazu: <https://www.re3data.org/>

²⁶ Die Website ist unter <https://www.smir.ch/> verfügbar.

Forschungsdaten (inkl. Metadaten) archiviert, in einem Repository veröffentlicht – falls keine triftigen Gründe dagegen vorliegen – und allgemein zugänglich gemacht werden, so, dass die Daten nachgenutzt werden können (Schweizerischer Nationalfonds, o. J.-b). Ein wichtiges Instrument dafür ist der Datenmanagementplan (siehe Kapitel 4.3.3.1). Als Best Practice für die Planung und den Austausch der Daten verweist der SNF auf die FAIR-Prinzipien, die im besten Fall bereits für den DMP berücksichtigt werden (Schweizerischer Nationalfonds, o. J.-a): «Der SNF erwartet von den Forschenden, dass sie ihre Daten im Sinne der FAIR Daten Prinzipien in öffentlich zugänglichen, digitalen Datenbanken veröffentlichen (Schweizerischer Nationalfonds, o. J.-a).» FAIR ist ein Akronym und steht für Findable (auffindbar), Accessible (zugänglich), Interoperable (interoperabel) und Reusable (wiederverwendbar). Genauer bedeutet das (Wilkinson et al., 2016, S. 4):

- Findable: Die Daten sind mit Metadaten angereichert, die (Meta-)Daten sind in einer durchsuchbaren Ressource gespeichert und haben einen eindeutigen Identifikator zugewiesen, der ebenfalls in den Metadaten vermerkt ist.
- Accessible: Die Übermittlung der (Meta-)Daten erfolgt über ein offenes Standard-Kommunikationsprotokoll (beispielsweise https). Die Metadaten müssen auch abrufbar sein, wenn die zugrundeliegenden Daten nicht mehr verfügbar sind.
- Interoperable: (Meta-)Daten sollen mit anderen (Meta-)Daten verknüpft und mittels Zuhilfenahme von Standards und kontrolliertem Vokabular erstellt werden.
- Reusable: Die Metadaten müssen alle nötigen Angaben (beispielsweise Autor, Jahr, DOI) beinhalten, damit die Daten korrekt zitiert werden können, ausserdem muss eine Lizenz mit den Bedingungen der Nachnutzung und der Entstehungszusammenhang hinterlegt sein.

Ziel dieser FAIR-Prinzipien ist, dass sowohl Mensch, wie auch Maschine die (Meta-)Daten (aus)lesen können und alle benötigte Information hinterlegt ist, damit die Forschungsdaten gefunden, interpretiert und genutzt werden können (Schweizerischer Nationalfonds, o. J.-a; Wilkinson et al., 2016). Das ist wichtig, weil Forschungsdaten eine wichtige Ressource darstellen. Somit ist der Kreis (und Datenlebenszyklus) zum ersten Kapitel der Sachanalyse geschlossen.

4.4 Überblick über die Vermittlung von Forschungsdatenmanagement-Kompetenzen an anderen akademischen Institutionen

In diesem Kapitel wird der Frage «Wie werden den Studierenden Kompetenzen im Forschungsdatenmanagement an anderen akademischen Institutionen vermittelt?»

nachgegangen. Es wird ein Blick über den Tellerrand geworfen und aufgezeigt, wie anderer akademischen Institutionen Forschungsdatenmanagement lehren und vermitteln. Für diesen groben Überblick wurden über 20 verschiedene Universitäten und Hochschulen aus der Schweiz, Deutschland, dem Vereinigten Königreich und den Vereinigten Staaten von Amerika untersucht. Es wurden nur die Universitäten und Hochschulen berücksichtigt, die im Zuge für das Kapitel 3 (Anforderungen) und Kapitel 4.3 (Sachanalyse) durchgeführte Recherche gefunden wurden. Es wurden explizit nicht nur akademische Institutionen mit Inhalten zum medizinischen oder klinischen Forschungsdatenmanagement ausgewählt, weil erhofft wird, dass so mehr Daten zur Auswertung zur Verfügung stehen, um allgemeingültigere Aussagen treffen zu können. Die Aufstellung der Universitäten und Hochschulen inklusive URL auf die jeweiligen Websites und untersuchter Kriterien ist im Anhang 7.2 zu finden.

4.4.1 Grundlagen und Forschungsdatenmanagement mit breitem Ansatz

Alle untersuchten akademischen Institutionen bieten Inhalte zum Forschungsdatenmanagement an. Die Inhalte sind auf einer separaten Website extra für das FDM (siehe Universität Basel, Mainz und Minnesota), auf einer Unterseite der Hochschul-Website (siehe Universität Fribourg, Edinburgh und Purdue University) oder über die Website der Universitätsbibliothek (siehe Universität Basel, Zürich und New York University) einsehbar. Der grösste Teil der Inhalte sind nicht auf bestimmte Forschungsdisziplinen ausgelegt, sind also allgemein gehalten. Das könnte auch damit zu tun haben, dass die Inhalte auf den Websites nicht – oder nur selten – in die Tiefe gehen. Die Ausnahme bilden die Angebote, für die extra eine separate Website zum Thema FDM erstellt wurde. Aber viele der Inhalte, die über die Websites der Hochschulen oder Bibliotheken verfügbar sind, sind auf das Wichtigste reduziert, so dass sie in kompakter Art und Weise präsentiert werden können. Ein anderer Grund ist sicher, dass ein erheblicher Mehraufwand entsteht, wenn für jeden Studiengang oder für jedes Forschungsfeld, das die Hochschule oder Universität bedient, die Eigenheiten und speziellen verfügbaren Tools und Links zu hilfreichen Websites herausgearbeitet werden müssen. Die Ausnahme bestätigt die Regel: Von den Hochschulen und Universitäten, die das FDM grundsätzlich allgemein halten, gibt es aber auch solche, die die Inhalte mit breitem Ansatz mittels Links zu weiterführender Information einer oder mehreren Forschungsdisziplinen anreichern. Beispiele dafür sind die John Hopkins Sheridan Libraries und die Universität Fribourg.

Wie schon erwähnt, sind die Inhalte auf einer Website aufbereitet. Die meisten vermitteln die Inhalte über Text, manche betten kurze Videos ein, beispielsweise die Universitäten

Basel und Pittsburgh. Einige der Hochschulen und Universitäten bieten Trainings und Kurse im FDM an. Manche Trainings sind auf Forschende ausgerichtet (ETH Zürich, EPFL und Universität Fribourg), es gibt aber auch einige, die Forscher und Studierende berücksichtigen (Universität Basel, University of Cambridge, MIT Libraries, John Hopkins Sheridan Libraries, University of Edinburgh bzw. MANTRA und Universität Bielefeld). Die Universität Bielefeld bietet den Kurs «Research Data Management» als obligatorisches Wahlpflichtfach für die drei Masterstudiengänge «Intelligente Systeme», «Naturwissenschaftliche Informatik» und «Data Science» an (Universität Bielefeld, o. J.; Wiljes & Cimiano, 2019, S. 2). Die Universität Basel bietet über ihre Website zwar «nur» Inhalte zu allgemeinem Forschungsdatenmanagement an, es werden aber über die Universitätsbibliothek auch im Speziellen Trainings und Events für Mediziner beispielsweise «Endnote für Mediziner» angeboten, dort dürfen sich auch Studierende einschreiben. Momentan werden die Events, die durchgeführt werden – wegen Covid-19 – online angeboten. Bei manchen akademischen Institutionen (z.B. Luzern, Universität Zürich, University of Pittsburgh und NYU Libraries) ist nicht bekannt, wer die Zielgruppe für die Trainings ist. Als einzige Universität von allen, die einen bereiten Ansatz beim FDM wählen, bietet Minnesota einen Kurs nur für Studierende an. Die Kursmaterialien und Inhalte der University of Minnesota sind auf einer separaten Website verfügbar, der Kurs ist im Modus flipped classroom konzipiert. Der Kurs ist aber nicht in einem Curriculum verankert. Dann gibt es noch akademische Institutionen (Purdue University, Queens University Libraries, University of Oxford und University of Oregon Libraries/Oregon State University Library), die keine Trainings oder Kurse anbieten. Einige bieten aber Services im Zusammenhang mit dem FDM an, wie Beratungsgespräche oder Hilfe beim Erstellen eines DMP. Ein Spezialfall stellt die Universität Mainz dar: Auf der Website werden ähnlich, wie bei der Universität Basel die Inhalte allgemein gehalten, im Curriculum der Medizin ist aber ein Kurs zu Digitalen Kompetenzen verankert (Johannes Gutenberg-Universität Mainz, 2017). Es wird also nicht das FDM speziell gelehrt, sondern digitale Kompetenzen vermittelt, die auch für das FDM wichtig sind.

4.4.2 Disziplinspezifisches Forschungsdatenmanagement

Nur gerade zwei der 21 untersuchten akademischen Institutionen weisen auf der jeweiligen Website disziplinspezifische Inhalte im FDM auf, die sich nicht nur auf Links oder Hinweise beschränken. Das sind die Yale Library (Harvey Cushing/John Hay Whitney Medical Library) und die Bodleian Libraries. Beide Bibliotheken sind auf die Medizin ausgerichtet und bieten dementsprechend Inhalte zu diesem Forschungsfeld. Die Yale Library verweist auf verschiedene Kurse (bzw. Folien), die über GitHub zur

Verfügung stehen, ansonsten können die wichtigsten Dinge im Zusammenhang mit dem FDM auf der Website nachgelesen werden. Die Bodleian Libraries handhabt es ähnlich: Die Inhalte sind auf der Website aufbereitet und sind mit vereinzelt Videos angereichert. Es werden ausserdem Video-Trainings angeboten, die gedownloadet werden können. Ansonsten verweisen die Bodleian Libraries auf externe (allgemeine) Kurse, die auf der Website <https://researchdata.ox.ac.uk/> gelistet sind.

4.4.3 Fazit

Jetzt wird noch die Frage «Wie werden den Studierenden Kompetenzen im Forschungsdatenmanagement an anderen akademischen Institutionen vermittelt?» versucht zu beantworten: Dieser grobe Überblick über einige ausgewählte akademischen Institutionen hat gezeigt, dass zwar alle über eine Website – manche ausführlicher als andere – und zum grossen Teil mittels Trainings oder Events Inhalte zum FDM anbieten, dass aber kaum Kurse angeboten werden, die auf Studierende und Forschungsdisziplin zugeschnitten sind. Viele Angebote von Trainings sind auf Forscher oder Universitätsangehörige ausgelegt und sollen eine breite Basis, also unterschiedliche Disziplinen bedienen. Es wurde ausserdem festgestellt, dass keine einzige der untersuchten Institutionen einen Kurs oder ein Seminar für Studierende anbietet, der das klinische Forschungsdatenmanagement ganzheitlich abdeckt.

Es gilt zu bedenken, dass dieser Überblick nur oberflächlich aufzeigt, was in Sachen FDM an akademischen Institutionen geleistet wird und kein abschliessendes Bild zur Frage vermitteln kann. Damit aussagekräftigere Ergebnisse präsentiert werden können, müssten die ausgewählten Universitäten und Hochschulen vertiefter, umfassender und systematischer untersucht werden. Ein Problem in diesen Zusammenhang ist auch, dass viele Curricula nicht öffentlich zugänglich sind und deshalb die Beantwortung erschwert, wenn nicht sogar verunmöglicht. Das haben auch Piorun et al. (2012) in ihren Artikel «Teaching Research Data Management: An Undergraduate/Graduate Curriculum» festgestellt.

4.5 Umsetzung – Methoden und Medien

In diesem Kapitel wird die Frage «Wie und womit soll das Thema klinisches Forschungsdatenmanagement vermittelt werden?» behandelt. Dieses Kapitel soll aber lediglich Ideen liefern, wie ein Kurs im klinischen Forschungsdatenmanagement umgesetzt werden könnte und ist daher nur deskriptiv. Es besteht ausserdem kein Anspruch auf Vollständigkeit. Die Vermittlung von Kursinhalten ist abhängig vom

Dozenten/ Lehrenden und der jeweiligen Klasse und kann daher nicht detailliert und umfassend besprochen werden.

Bevor Methoden und Medien behandelt werden, sollten die allgemeinen Unterrichtsvoraussetzungen angesprochen werden: Es ist grundsätzlich von Vorteil, Grundlagen im wissenschaftlichen Arbeiten zu besitzen, um dem Kurs folgen und die Inhalte gänzlich verstehen zu können. Da der Kurs auf Masterstudierende ausgelegt ist, sollte diese Voraussetzung aber kein Problem darstellen.

Carlson et al. (2011, S. 4) haben erkannt, «[...] it is not simply enough to teach students about handling incoming data, they must also know, and practice, how to develop and manage their own data with an eye toward the next scientist down the line.» Für die Durchführung eines Forschungsdatenmanagement-Kurses sollte also bedacht werden, dass dieser nicht als reiner Frontalunterricht aufgebaut ist, sondern mindestens praktische Übungen beinhaltet. Ein Beispiel dafür kann das Erstellen eines DMP (manuell oder mit dem DMP Canvas Generator) für ein eigenes Forschungsprojekt, das Zusammenstellen eines Data Curation Profiles, das Anonymisieren eines Datensatzes, das Erstellen eines eigenen ORCID Profils, die Auswahl eines geeigneten Repositoriums gemäss einer Case Study oder die Entwicklung einer Namenskonvention für die eigenen Dateien sein. Die Ergebnisse könnten zudem vor der Gruppe präsentiert und diskutiert werden – mit einer Präsentation würde eine weitere digitale Kompetenz (Evaluierung) trainiert werden (siehe dazu auch die Erläuterungen zu Data Literacy im Forschungsstand). Kjelvik & Schultheis (2019, S. 2) ergänzen zur Praxisorientierung: «We predict that the strongest learning experiences surrounding data literacy will arise when students have the opportunity to engage in quantitative reasoning and data science while exploring data sets from scientific research.» Weiter wird erklärt, dass oft – um im Unterricht ein bestimmtes Problem zu adressieren und verdeutlichen – Datensätze zusammengestellt werden, dass es aber solchen unauthentischen Daten an Komplexität und Kontext mangelt (Kjelvik & Schultheis, 2019). Die Verwendung von authentischen Daten im Unterricht bringt also einen Mehrwert. Fraglich bleibt, ob gerade im klinischen Forschungsdatenmanagement authentische Daten, die noch nicht anonymisiert wurden etc. überhaupt verfügbar sind. Auch Piorun et al. (2012, S. 50) befürworten, den Unterrichtsstoff möglichst realitätsnah zu vermitteln: Sie setzen Case Studies ein, die Probleme aus dem Forschungsdatenmanagement widerspiegeln. Ein Tipp der Carlson et al. (2015, S. 15) geben, um die Thematik für Studierenden greifbarer zu machen, ist das Referenzieren auf die eigene persönliche Datenverwaltung der Studierenden, wie Fotos oder Dokumente. Wiljes & Cimiano (2019, S. 3) indes starten jede Lektion mit einer

Mitteilung aus den News, die in Zusammenhang mit dem Forschungsdatenmanagement stehen, um die Relevanz aufzuzeigen und die Motivation hochzuhalten.

Holles & Schmidt (2018) erklären, dass Seminare, die von Bibliothekarinnen durchgeführt werden oft keine aussagekräftigen Beispiele beinhalten, weil die Instrukturen zu wenig Erfahrung in der jeweiligen Forschungsdisziplin ausweisen würden. Es wird weiter ausgeführt, dass ein Kurs Theorie und Praxis abdecken sollte. Idealerweise wird der Kurs von einer Informationsspezialistin/ eines Bibliothekars und von einem Mitglied der entsprechenden Fakultät unterrichtet, um möglichst realitätsnahe Beispiele und Sachverhalte mit der Theorie zu verbinden (Holles & Schmidt, 2018). Surkis et al. (2017, S. 185) haben in ihrem Kurs sogar sieben Lehrende – aus der medizinischen Fakultät und der Universitätsbibliothek – für insgesamt acht Unterrichtsblöcke eingesetzt. Hier gilt es noch anzumerken, dass dieser Kurs für Forschende der Medizin entwickelt wurde und dementsprechend sehr spezialisierte Inhalte vermittelt wurden. In einem Kurs für Studierende, der neben den Eigenheiten der klinischen Forschungsdaten, vor allem Grundlagen vermitteln soll, muss also nicht auf eine solche Vielzahl an Spezialistinnen zurückgegriffen werden. Schmidt & Holles (2018) argumentieren, dass das Forschungsdatenmanagement interdisziplinär ist und dementsprechend nicht nur von Angehörigen einer bestimmten Disziplin gelehrt werden sollten. Vielmehr braucht es mehr als eine Perspektive, um das FDM angemessen behandeln zu können. So haben Sie auch einen Trend in Richtung «als Team unterrichten» festgestellt (Schmidt & Holles, 2018).

Eine Möglichkeit einen Kurs abzuhalten, wäre für den theoretischen Teil einen Impulsvortrag zu geben und die Theorie mittels Übungen (bei denen auch Tools verwendet werden) praktisch zu vertiefen. Um der Digitalisierungsstrategie der Universität Bern (2019a, siehe dazu auch Kapitel 4.1) besser zu entsprechen, könnten anstatt eines Impulsvortrages jeweils Videos verwendet werden, die die wichtigsten Inhalte eines Themenblockes vermitteln. Vorteil für die (vielbeschäftigten) Studierenden wäre, dass die Videos zeit- und ortsunabhängig studiert werden können. Die University of Minnesota hat einen solchen Kurs²⁷ entwickelt. Dieser Kurs ist nach der Methode flipped classroom bzw. inverted classroom konzipiert. Das hiesst, die Grundlagen werden von den Studierenden mittels der Videos (online) selbständig erarbeitet, im Präsenzunterricht (offline) werden die gelernten Inhalte gemeinsam vertieft und praktisch erprobt (Carlson et al., 2015, S. 15). Diese Unterrichtsmethode könnte auch eine Lösung

²⁷ Das Kursmaterial kann hier: <https://sites.google.com/a/umn.edu/data-management-workshop-series-instructor-site/> eingesehen werden. Die Videos sind auf YouTube frei zugänglich.

für das vorliegende Konzept sein, wobei für die Erstellung der Lernvideos noch einiges an Arbeit investiert werden müsste. Eine andere Möglichkeit wäre, bereits bestehende Videos zum «allgemeinen» Forschungsdatenmanagement für die Vorbereitung zu nutzen und die Eigenheiten des klinischen Forschungsdatenmanagement in der Gruppe (mit kurzem Input des Instructors) zu erarbeiten. Gerade im Hinblick auf die Corona-Krise wäre es durchaus denkbar, den ganzen Unterricht in ein Webinar zu verlagern. Da würde aber der Austausch unter den Studierenden und zwischen Studierenden und Instruktor erschwert werden. Aus eigener Erfahrung kann die Autorin sagen, dass das eine grosse Einbusse bezüglich Erkenntnisgewinn, Engagement und Gruppenzugehörigkeit darstellen würde.

Wiljes & Cimiano (2019, S. 2 & 5) empfehlen, maximal drei Sitzungen à je zwei Lektionen und für die Teilnahme Kreditpunkte zu vergeben. Der Leistungsnachweis besteht aus einem DMP, der benotet wird. Whitmire (2015) spricht in seiner Publikation «Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lesson Learned» von einem elfwöchigen Kurs, der zwei Kreditpunkten entspricht. Um den Kurs attraktiver zu machen, ist sicherlich empfohlen Kreditpunkte für die Teilnahme und das Einreichen einer Arbeit zu vergeben. Das trifft insbesondere zu, falls der Kurs über mehrere Wochen in Blöcken (modularer Aufbau) abgehalten wird. Grundsätzlich könnte der Inhalt auch an einem ganzen Nachmittag vermittelt werden. Henderson & Knott (2015, S. 51) sind folgender Meinung: «Even five minutes as part of an orientation session can be a start.» Je nachdem, welches Format gewählt wird, müsste dann der Inhalt aus der Sachanalyse etwas «zusammengestaucht» oder angereichert werden. Zur Strukturierung der Inhalte kann gesagt werden, dass zum Beispiel der Datenlebenszyklus als roter Faden und Leitlinie dienen könnte, um den Kreislauf der Daten zu verdeutlichen. Ausserdem wäre es denkbar, die FAIR-Prinzipien anstatt am Ende des Kurses bereits zu Beginn einzuführen.

Es gilt noch zu erwähnen, dass die Kurse evaluiert und mit den Feedbacks der Studierenden ständig weiterentwickelt und verbessert werden sollten, um den Bedürfnissen der Studierenden zu entsprechen (Wiljes & Cimiano, 2019, S. 3).

Zusammengefasst lässt sich die Frage «Wie und womit soll das Thema klinisches Forschungsdatenmanagement vermittelt werden?» wie folgt beantworten: Ein praxisorientierter Ansatz, zum Beispiel in Form eines Seminars, ist sicherlich zielführender und lehrreicher für die Studierenden. Um das zu erreichen, können Übungen und Praxisbeispiele in den Unterricht eingebaut und Tools benützt werden. Grundsätzlich zeigt der Trend, dass der Unterricht in Teams von Spezialistinnen

abgehalten wird. Auf Stufe Grundlagen und Einführung ist es aber durchaus möglich – wenn genug Erfahrung / Fachwissen auf Seiten der Instruktorin vorhanden ist – den Kurs ohne weitere Experten durchzuführen. Es wäre natürlich denkbar und eine Abwechslung, wenn beispielsweise für Ethik, Datenschutz oder Datenbanken Experten eingeladen werden würden, die Fragen zu diesen Themen auch vertieft und umfassend beantworten können. Es gibt diverse Möglichkeiten wie und womit die Thematik «klinisches Forschungsdatenmanagement» vermittelt werden kann: ob online, offline oder eine Kombination aus beidem, ob modularer Aufbau oder ein bis zwei Halbtage, grundsätzlich ist alles denkbar.

5 Fazit

Über die ganze Arbeit hinweg wurden verschiedenste Fragen rund um die Konzeption eines Forschungsdatenmanagement-Kurses beantwortet: Das «Wofür?», das Auskunft zum Zweck und der Verwendungssituation gibt, das «Für wen?», das aufzeigt, wer die Zielgruppe ist, das «Wozu?», das beantwortet, welche Lernziele und Qualifikationen angestrebt werden, das «Was?», das klärt, welcher Bedarf und welche Inhalte vermittelt werden sollen, das «Wie?» und «Womit?», das angesprochen wurde und mögliche Lösungen zur Umsetzung präsentiert und schliesslich das «Wer?», das über die Adressaten der vorliegenden Arbeit informiert. Die wichtigsten Erkenntnisse dieser Fragen werden im Folgenden aufgegriffen und diskutiert.

5.1 Beantwortung der Fragestellungen & Diskussion der Ergebnisse

Die vorliegende Arbeit beinhaltet ein Unterrichtskonzept zum Thema Forschungsdatenmanagement, das auf die Zielgruppe Medizinstudierende im Masterstudium zugeschnitten ist. Adressiert werden Informationsspezialistinnen und (medizinische) Bibliothekare, die das Konzept für einen Kurs benützen und adaptieren können. Ziel ist es, digitale Kompetenzen zu fördern und den Umgang mit Daten bereits im Studium zu fördern, um den zukünftigen Forschern und Ärztinnen das nötige Rüstzeug mitzugeben, um mit Forschungsdaten angemessen umgehen zu können.

Für das Kapitel 3 «Anforderungen» und somit auch für das Kapitel 4.1 «Sachanalyse» wurde folgende Fragestellung formuliert: «Welche Inhalte sind wichtig und relevant, um das Thema Forschungsdatenmanagement in einem Kurs mit Zielgruppe Masterstudierende im Medizinstudium zu vermitteln?» Die Analyse der recherchierten Literatur hat ergeben, dass sich die Inhalte, die Studierenden gelehrt werden, grundsätzlich nicht von denen unterscheiden, die Forscherinnen vermittelt werden, dass aber die Studierenden einen breiteren Ansatz brauchen. Damit ist gemeint, dass die Studierenden noch kein fundiertes Grundwissen besitzen und deshalb Grundlagen vermittelt werden müssen. Es ist aber trotzdem wichtig, disziplinspezifische Vertiefungen vorzunehmen, um die Relevanz verdeutlichen und die Motivation hochhalten zu können. Die Meinung, welche Inhalte am wichtigsten sind, gehen auseinander, es konnte aber festgestellt werden, dass immer wieder die gleichen Themen genannt wurden. So konnten fünf Hauptbereiche – Einführung, rechtliche Aspekte & Ethik, Datenmanagement & -organisation, Preservation, sowie Publikation, Re-Use & SNF – identifiziert werden. Es wurde ausserdem aufgezeigt, dass die Unterscheidung von «allgemeinem» FDM und disziplinspezifischem FDM wichtig ist, weil – wie im vorliegenden Fall des klinischen

Forschungsdatenmanagements – vor allem andere/ weitere Anforderungen gelten, die sich auf diverse Bereiche des Forschungsdatenmanagements auswirken. Grund dafür sind gesundheitsbezogene Daten, aus denen ebendiese speziellen Anforderungen resultieren, die das FDM adäquat adressieren und bewältigen muss. Wie Muriel Helmers im Interview ausführte, sind das im Falle des klinischen Forschungsdatenmanagements insbesondere die Bereiche Datenschutz, Ethik, Datensicherheit, Nachnutzung und Archivierung.

Aus den Anforderungen wurden die Kompetenzen und Lernziele abgeleitet. «Wozu werden Inhalte zum klinischen Forschungsdatenmanagement vermittelt?» ist die Leitfrage dafür. Zusammenfassend und als Antwort auf diese Fragestellung kann gesagt werden, dass die Studierenden nach dem Absolvieren dieses Kurses die nötigen Kompetenzen und Qualifikationen besitzen sollen, um die eigene Forschung, die im Rahmen des Studiums anfällt adäquat verwalten zu können. Von einem höheren Standpunkt aus betrachtet, sollen bestimmte digitale Kompetenzen gefordert und gefördert werden. Ausserdem geht es darum, die Angst vor dem Thema zu nehmen und den Studierenden nützliche Tools und Hilfsmittel auf den Weg zu geben, die auch für die berufliche Laufbahn von Nutzen sein können/ werden.

Basierend auf den Lernzielen, Kompetenzen und anhand der zusammengetragenen Anforderungen wurde dann die Sachanalyse geschrieben. Sie behandelt das Thema klinisches Forschungsdatenmanagement im Detail, dabei wurde als Methode die Literaturrecherche verwendet. Die Aufarbeitung hat gezeigt, dass die Inhalte miteinander vernetzt sind und zum Teil Einfluss aufeinander ausüben. So hat die Auswahl eines geeigneten Dateiformates nicht nur Einfluss auf das aktive Datenmanagement, sondern auch auf die Langzeitarchivierung und die Nachnutzung der Daten. Zur Sachanalyse muss noch angemerkt werden, dass die Themen, die im Kapitel «Anforderungen» aus der Literatur erarbeitet wurden, häufig abstrakt und wenig konkret sind. Zum Beispiel wurde die Archivierung von Daten genannt, aber nicht genau, was dort behandelt werden soll – geeignete Dateiformate, Aufbewahrungsdauer, Bewertung etc. So musste die Autorin Wohl oder Übel teilweise selbst über die genauen Inhalte bestimmen. Ausserdem führte die grosse Menge an möglicher Literatur, die hinzugezogen werden konnte und das dementsprechende umfangreiche Thema an sich, dazu, dass die Autorin Inhalte nicht besprechen konnte. Als Beispiel können die Kosten angeführt werden, die auch im Forschungsdatenmanagement in Betracht gezogen werden müssten. Es war eine Herausforderung nur das Wichtigste und für die Studierenden Relevanteste herauszupicken. Es gibt sicher noch Themen, die für Medizinstudierende von Bedeutung sind, die im vorliegenden Konzept aber keinen Eingang gefunden haben. Ausserdem

wurden im Konzept ausschliesslich digitale Daten behandelt, analoge Forschungsdaten wurden nicht adressiert. Ein weiterer Kritikpunkt an der Sachanalyse stellt das Verhältnis Theorie Praxis dar: Im Konzept wurde die Theorie aufgearbeitet, aber die Praxis ist wohl noch etwas zu kurz gekommen. Auch im Hinblick auf den medizinischen Teil könnte noch mehr herausgeholt werden. Aber der Autorin fehlt die nötige (Praxis-)Erfahrung in diesem Bereich.

Nach der Sachanalyse wurde ein kurzer Blick über den Tellerrand geworfen, um zu sehen, wie andere Hochschulen und Universitäten das Forschungsdatenmanagement lehren. Folgende Fragestellung wurde dazu formuliert: «Wie werden den Studierenden Kompetenzen im Forschungsdatenmanagement an anderen akademischen Institutionen vermittelt?» Die Antwort dazu kann nicht abschliessend gegeben werden, weil die Ergebnisse nicht aussagekräftig genug sind. Im Rückblick wäre es besser gewesen, sich nur auf das Angebot von medizinischen Fakultäten oder Bibliotheken zu konzentrieren. Ausserdem hat die mangelnde Transparenz bezüglich Curricula dazu geführt, dass die Beantwortung der Forschungsfrage noch schwieriger wurde. Es kann lediglich mitgenommen werden, dass ein grosser Teil der akademischen Institutionen das Forschungsdatenmanagement genug wichtig empfinden, um einerseits grundlegende Inhalte über eine Website öffentlich zugänglich zu machen und andererseits mittels Trainings, Events, Kursen und/oder Services (von der Bibliothek offeriert) vertieft Unterstützung anbieten. Es scheint aber dann doch nicht so wichtig zu sein – oder die Wichtigkeit wurde noch nicht realisiert – um einen Forschungsdatenmanagement-Kurs im Curriculum zu verankern.

Zum Abschluss dieser Arbeit wurden Ideen zur Umsetzung eines Forschungsdatenmanagement-Kurses zusammengetragen. Die Fragestellung dafür lautete: «Wie und womit soll das Thema klinisches Forschungsdatenmanagement vermittelt werden?» Das Einflechten von Übungen und Praxisbeispielen ist sicher ein guter Anfang, um das Thema für die Studierenden fassbarer zu machen. Aus diesem Grund könnte auch in Betracht gezogen werden, einen Kurs als Team mit einer Expertin für klinischen Datenmanagement abzuhalten. Oder für Vertiefungen von anderen Inhalten, wie Ethik oder Datenbanken entsprechende Spezialisten hinzuzuziehen. Das ist aber nicht zwingend notwendig. In diesem Kapitel werden diverse Möglichkeiten präsentiert, wie und womit die Thematik «klinisches Forschungsdatenmanagement» vermittelt werden kann: ob online, offline oder eine Kombination aus beidem, ob modularer Aufbau oder ein bis zwei Halbtage, grundsätzlich ist alles denkbar. Zur Umsetzung muss aber noch vertieft abgeklärt werden, was möglich ist, je nachdem kann

der Aufwand ziemlich hoch ausfallen, sollten beispielsweise Videos erstellt werden wollen.

Abschliessend gilt noch zu erwähnen, dass es viele mögliche Ansätze, Methoden etc. gibt, um das Forschungsdatenmanagement zu vermitteln – der richtige Weg hängt mitunter aber auch von den jeweiligen Teilnehmenden/ der Klasse ab. Es ist deshalb sicherlich zielführend, wenn nach einem Kurs Feedback von den Teilnehmerinnen eingeholt wird und für weitere Durchführungen Wünsche zu Inhalten oder Unterrichtsmethoden berücksichtigt werden.

5.2 Offene Fragen und Ausblick

Mit der vorliegenden Arbeit haben sich einige Fragen beantwortet, manche bleiben offen und weitere haben sich aufgetan: Noch offen, beziehungsweise nicht abschliessend geklärt ist die Frage bezüglich der konkreten Umsetzung (als Team; online, offline oder Kombination; Übungen etc.) dieses Kurses. Ebenfalls noch offen bleibt die Frage, wie andere akademischen Institutionen das Forschungsdatenmanagement vermitteln. Weiter könnte abgeklärt werden, ob Interesse besteht, weitere Themen dazu zu nehmen oder sogar den Kurs von Forschungsdatenmanagement auf Data Literacy auszudehnen. Es wäre ausserdem interessant zu untersuchen, ob die Anforderungen an einen solche Kurs anders sind, wenn die Perspektive – beispielsweise aus Sicht des Arbeitsmarktes oder nach dem Informationssuchverhalten – geändert wird. Generell müsste noch mehr bezüglich Bedürfnisse an einen Kurs im klinischen Forschungsdatenmanagement auf Stufe der Studierenden geforscht werden. Konkrete Ergebnisse dazu liegen noch keine vor.

6 Quellenverzeichnis

Adamick, J., Reznik-Zellen, R., & Sheridan, M. (2012). Data Management Training for Graduate Students at a Large Research University. *Journal of eScience Librarianship*, 1, 180–188. <https://doi.org/10.7191/jeslib.2012.1022>

Akers, K., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5–26. <https://doi.org/10.2218/ijdc.v8i2.263>

American Society for Quality. (o. J.). *Quality Glossary*. American Society for Quality. Abgerufen 5. Juli 2020, von <https://asq.org/quality-resources/quality-glossary>

Assfalg, R. (2014). Metadaten. In R. Kuhlen, W. Semar, & D. Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation* (6. Ausgabe, S. 159–171). De Gruyter Saur.

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, 63(2), 123–134. <https://doi.org/10.1515/libri-2013-0010>

Carlson, J. (2014). The Use of Life Cycle Models in Developing and Supporting Data Service. In Purdue University (Hrsg.), *Research Data Management—Practical Strategies for Information Professionals* (S. 63–86). <https://books.google.ch/books?hl=de&lr=&id=qZStAQAAQBAJ&oi=fnd&pg=PA63&dq=#v=onepage&q&f=false>

Carlson, J., Fosmire, M., Miller, C., & Nelson, M. S. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Libraries Faculty and Staff Scholarship and Research*, 1–30.

Carlson, J., Sapp Nelson, M., Johnston, L., & Koshoffer, A. (2015). Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates. *Bulletin of the American Society for Information Science and Technology*, 41(6), 14–17. <https://doi.org/10.1002/bult.2015.1720410608>

Center for Informatics Research in Science and Scholarship, & University of Illinois. (o. J.). *Data Curation Education Program (DCEP)*. CIRSS. Abgerufen 16. Juni 2020, von <http://cirss.ischool.illinois.edu/Project/project-details.php?id=19>

CERN. (o. J.). *Zenodo*. Zenodo. Abgerufen 13. April 2020, von <https://zenodo.org/>

Clinical Data Interchange Standards Consortium. (o. J.). *About CDISC*. Clinical Data Interchange Standards Consortium. Abgerufen 4. Juli 2020, von <https://www.cdisc.org/about>

Creative Commons. (2017, November 7). *Mehr über die Lizenzen—Creative Commons*. Creative Commons. <https://creativecommons.org/licenses/?lang=de>

Cremer, F., Engelhardt, C., & Neuroth, H. (2015). Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale. *Bibliothek Forschung und Praxis*, 39(1), 13–31. <https://doi.org/10.1515/bfp-2015-0006>

CTU Universität Bern. (2018). *GCP & HRA compliant Data Management*.

Czernik, A. (2016, August 5). *Definition und Unterscheidung der Begriffe Daten, Informationen & Wissen*. Datenschutzbeauftragter Info. <https://www.datenschutzbeauftragter-info.de/definition-und-unterscheidung-der-begriffe-daten-informationen-wissen/>

Czeschik, C. (o. J.). *Literaturverwaltung leicht gemacht*. Lehmanns Media. Abgerufen 9. Juli 2020, von <https://www.lehmanns.ch/page/literaturverwaltungsprogramm>

DataCite. (o. J.). *Why is it so important to cite data?* DataCite. Abgerufen 10. Juli 2020, von <https://datacite.org/cite-your-data.html>

- DataONE. (o. J.-a). *Data Management Planning*. DataONE. Abgerufen 2. Juli 2020, von <https://www.dataone.org/data-management-planning>
- DataONE. (o. J.-b). *DataONE*. Abgerufen 12. April 2020, von <https://www.dataone.org/>
- DataONE. (o. J.-c). *Lesson 5: Data Quality Control and Assurance*. Abgerufen 5. Juli 2020, von https://www.dataone.org/sites/all/documents/education-modules/handouts/L05_DataQualityAssurance_Handout.pdf
- DataONE. (o. J.-d). *Lesson 10: Legal & Policy Issues*. Abgerufen 10. Juli 2020, von https://www.dataone.org/sites/all/documents/education-modules/handouts/L10_LegalPolicy_Handout.pdf
- Datenschutz.org. (2018a, Juli 13). Datenschutz in der Pflege: Sensibler Umgang mit Patientendaten wichtig! *Datenschutz*. <https://www.datenschutz.org/pflege/>
- Datenschutz.org. (2018b, Juli 13). Patientendaten: Besonderer Datenschutz bei Gesundheitsdaten. *Datenschutz*. <https://www.datenschutz.org/patientendaten/>
- Deutsche Forschungsgemeinschaft. (2015). *Leitlinien zum Umgang mit Forschungsdaten*. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf
- Digital Curation Centre. (2020a). *Clinical Medicine*. DCC. <https://www.dcc.ac.uk/resources/metadata-standards/disciplinary/clinical-medicine>
- Digital Curation Centre. (2020b). *DCC*. DCC. <https://www.dcc.ac.uk/>
- Digital Curation Centre. (2020c). *How to License Research Data*. DCC. <https://www.dcc.ac.uk/guidance/how-guides/license-research-data#x1-4000>
- DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation. (o. J.). *Forschungsdatenmanagement: Was ist Datenmanagement?* VerbundFDB. Abgerufen 29. Juni 2020, von <https://www.forschungsdaten-bildung.de/datenmanagement?la=de>
- DLCM. (o. J.). *Data Management Plan*. DLCM. Abgerufen 2. Juli 2020, von <https://www.dlcm.ch/resources/dlcm-dmp>
- Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB). (o. J.-a). *Datenschutz und Forschung im Allgemeinen*. Datenschutz und Forschung im Allgemeinen. Abgerufen 1. Juli 2020, von <https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/statistik--register-und-forschung/forschung/datenschutz-und-forschung-im-allgemeinen.html>
- Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB). (o. J.-b). *Erläuterungen zu Cloud Computing*. Erläuterungen zu Cloud Computing. Abgerufen 9. Juli 2020, von https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/Internet_und_Computer/cloud-computing/erlaeuterungen-zu-cloud-computing.html
- Eidgenössisches Institut für Geistiges Eigentum. (o. J.). *Urheberrecht – was ist das?* IGE IPI. Abgerufen 10. Juli 2020, von <https://www.ige.ch/de/etwas-schuetzen/urheberrecht/grundlegendes.html>
- Europäische Patientenakademie. (o. J.). *Anonyme Codierungssysteme*. Europäische Patientenakademie. Abgerufen 1. Juli 2020, von <https://www.eupati.eu/de/glossary/anonyme-codierungssysteme/>
- Fecher, B., & Friesike, S. (2014, Januar 27). Ein Einblick in ein laufendes Forschungsprojekt: Die Systematic Review und das Teilen von Forschungsdaten. *Digital Society Blog*. <https://www.hiig.de/ein-einblick-in-ein-laufendes-forschungsprojekt-die-systematic-review-und-das-teilen-von-forschungsdaten/>
- Federer, L. M., Lu, Y.-L., & Joubert, D. J. (2016). Data literacy training needs of biomedical researchers. *Journal of the Medical Library Association : JMLA*, 104(1), 52–57. <https://doi.org/10.3163/1536-5050.104.1.008>

- Forschungsdaten.info. (2020a, Mai 12). *Creative Commons Lizenzen*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/creative-commons-lizenzen/>
- Forschungsdaten.info. (2020b, Mai 12). *Freie Lizenzen vs. Gemeinfrei*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/freie-lizenzen-vs-gemeinfrei/>
- Forschungsdaten.info. (2020c, Juni 18). *Auswählen und Bewerten*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/auswaehlen-und-bewerten/>
- Forschungsdaten.info. (2020d, Juni 18). *Das Data-Curation-Profile*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/organisieren-und-aufbereiten/data-curation-profile/>
- Forschungsdaten.info. (2020e, Juni 18). *Daten publizieren*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/>
- Forschungsdaten.info. (2020f, Juni 18). *Datendokumentation*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/beschreiben-und-dokumentieren/datendokumentation/>
- Forschungsdaten.info. (2020g, Juni 18). *Datenorganisation*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/organisieren-und-aufbereiten/datenorganisation/>
- Forschungsdaten.info. (2020h, Juni 18). *Der Datenlebenszyklus*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>
- Forschungsdaten.info. (2020i, Juni 18). *Metadaten und Metadatenstandards*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/beschreiben-und-dokumentieren/metadaten-und-metadatenstandards/>
- Forschungsdaten.info. (2020j, Juni 18). *Persistente Identifikatoren*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/persistente-identifikatoren/>
- Forschungsdaten.info. (2020k, Juni 18). *Was sind Forschungsdaten?* Forschungsdaten.info. <https://www.forschungsdaten.info/themen/informieren-und-planen/was-sind-forschungsdaten/>
- Forschungsdaten.info. (2020l, Juni 18). *Wie beginnt Forschungsdatenmanagement (FDM)?* Forschungsdaten.info. <https://www.forschungsdaten.info/themen/informieren-und-planen/wie-beginnt-forschungsdatenmanagement/>
- Forschungsdaten.info. (2020m, Juni 19). *Lebensdauer von Datenträgern*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/lebensdauer-von-datentraegern/>
- Forschungsdaten.info. (2020n, Juni 22). *Urheberrecht*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>
- Forschungsdaten.info. (2020o, Juni 23). *Forschung und Daten managen*. Forschungsdaten.info. <https://www.forschungsdaten.info/>
- Forschungsdaten.info. (2020p, Juni 24). *Formate erhalten*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/formate-erhalten/>
- Forschungsdaten.info. (2020q, Juni 24). *Langzeitarchivierung*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/langzeitarchivierung/>
- Forschungsdaten.info. (2020r, Juni 24). *Repositorien*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/repositorien/>
- Forschungsdaten.info. (2020s, Juli 1). *Der Datenmanagementplan*. Forschungsdaten.info. <https://www.forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/>

Forschungsdaten.info. (2020t, Juli 2). *Open Data, Open Access und Nachnutzung*.

Forschungsdaten.info. <https://www.forschungsdaten.info/themen/finden-und-nachnutzen/open-data-open-access-und-nachnutzung/>

Goben, A., & Griffin, T. (2019). In Aggregate: Trends, Needs, and Opportunities from Research Data Management Surveys. *College & Research Libraries*, 80(7), 903–924.

<https://doi.org/10.5860/crl.80.7.903>

Gryzbowski, A. (2007, Juli). *Naming conventions*. The University of Edinburgh.

<https://www.ed.ac.uk/records-management/guidance/records/practical-guidance/naming-conventions>

Henderson, M. E., & Knott, T. L. (2015). Starting a Research Data Management Program Based in a University Library. *Medical reference services quarterly*, 34(1), 47–59.

<https://doi.org/10.1080/02763869.2015.986783>

Holles, J. H., & Schmidt, L. (2018, Juni 23). Implementing a Graduate Class in Research Data Management for Science and Engineering Students. *Association for Engineering Education - Engineering Library Division Papers; Atlanta*.

<http://search.proquest.com/lisa/docview/2315576383/71998CC707C54CEFPQ/2>

informatikZentrale. (2010, August 9). *Datenbanken: Grundlagen*. informatikZentrale.

<https://www.informatikzentrale.de/datenbanken-grundlagen.html>

Jansen, P., van den Berg, L., van Overveld, P., & Boiten, J.-W. (2019). Research Data Stewardship for Healthcare Professionals. In P. Kubben, M. Dumontier, & A. Dekker (Hrsg.), *Fundamentals of Clinical Data Science* (S. 37–53). Springer. <http://www.ncbi.nlm.nih.gov/books/NBK543528/>

Johannes Gutenberg-Universität Mainz. (2017, Juni 2). *Universitätsmedizin Mainz erweitert digitales Lehrangebot als Curriculum 4.0*. Uni Mainz. https://www.uni-mainz.de/presse/aktuell/2566_DEU_HTML.php

Kindling, M., & Pampel, H. (2016, Februar 26). *Werkzeuge des Forschungsdatenmanagements*. handbuch.io.

https://handbuch.tib.eu/w/Werkzeuge_des_Forschungsdatenmanagements#Strategien_der_Zug.C3.A4ngleichmachung_und_Nachnutzung

Kindling, M., & Schirnbacher, P. (2012). Die digitale Forschungswelt als Gegenstand der Forschung / Research on Digital Research / Recherche dans la domaine de la recherche numérique. *Information - Wissenschaft & Praxis*, 64(2–3), 127–136. <https://doi.org/10.1515/iwp-2013-0017>

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting Messy with Authentic Data: Exploring the Potential of Using Data from Scientific Research to Support Student Data Literacy. *CBE—Life Sciences Education*, 18(2), es2. <https://doi.org/10.1187/cbe.18-02-0023>

Kofam Koordinationsstelle Forschung am Menschen. (o. J.). *Nach klinischen Versuchen suchen*. Kofam Koordinationsstelle Forschung am Menschen. Abgerufen 1. Juli 2020, von <https://www.kofam.ch/de/studienportal/nach-klinischen-versuchen-suchen/>

Kuhn, S., Kadioglu, D., Deutsch, K., & Michl, S. (2018). Data Literacy in der Medizin. *Der Onkologe*, 24(5), 368–377. <https://doi.org/10.1007/s00761-018-0344-9>

Kuschel, L. (2018, September 12). *Wem «gehören» Forschungsdaten?* Forschung & Lehre. <https://www.forschung-und-lehre.de/wem-gehoeeren-forschungsdaten-1013/>

Liguori, S. (2016). *Lesson 1: Data Management*. https://www.dataone.org/sites/all/documents/education-modules/handouts/L01_DataManagement_Handout.pdf

Lindenmann, I., Besmer, C., Morger, J., Keller, A., Bellanger, S., & Kaufmann, D. (2019a). *Handout: Empfehlungen und hilfreiche Links zum Umgang mit sensiblen Daten* (Universität Basel & Universität

Bern, Hrsg.).

https://www.unibe.ch/unibe/portal/content/e809/e962/e963/e6304/e583799/e573822/e583783/pane826303/e826311/DOK_Handout-FDM-sensitiveDaten_20199793_ili_ger.pdf

Lindenmann, I., Besmer, C., Morger, J., Keller, A., Bellanger, S., & Kaufmann, D. (2019b). *Handout: Forschungsdatenmanagement: Das Wichtigste in Kürze* (Universität Basel & Universität Bern, Hrsg.). https://researchdata.unibas.ch/fileadmin/user_upload/researchdata/Documents/FS-FDM_20190202.pdf

Mager, R. F. (1962). *Preparing instructional objectives*. Fearon Publishers / Lear Siegler, Inc., Education Division. <http://hdl.handle.net/2027/mdp.39015020958941>

Michaud, P. A., Juncker-Kupper, P., & and members of the Profiles working group. (2017). *PROFILES; Principal Relevant Objectives and Framework for Integrated Learning and Education in Switzerland*. Joint Commission of the Swiss Medical Schools. <https://www.profilesmed.ch/>

Nagel, A. (2019, November). Welcher Zitierstil für welches Studienfach? *CONTEXTA*. <https://www.contexta.de/welcher-zitierstil/>

Novostat. (2020, Januar 13). Evidenzbasierte Medizin—Möglichkeiten und Herausforderungen aus Sicht der Statistik. *Statistik Service*. <https://novostat.com/statistik-blog/evidenzbasierte-medizin.html>

NYU Libraries. (2020a, Juni 22). *Research Guides: Data Management Planning: File Format Selection*. NYU Libraries. https://guides.nyu.edu/data_management/file-formats

NYU Libraries. (2020b, Juni 22). *Research Guides: Data Management Planning: Version Control*. NYU Libraries. https://guides.nyu.edu/data_management/version-control

Oregon State University - Libraries. (2019a, September 3). *Research Data Services: Data Types & File Formats*. Oregon State University. <https://guides.library.oregonstate.edu/research-data-services/data-management-types-formats>

Oregon State University - Libraries. (2019b, September 3). *Research Data Services: Metadata/Documentation*. Oregon State University. <https://guides.library.oregonstate.edu/research-data-services/data-management-metadata>

Perrier, L., Blondal, E., Ayala, A. P., Dearborn, D., Kenny, T., Lightfoot, D., Reka, R., Thuna, M., Trimble, L., & MacDonald, H. (2017). Research data management in academic institutions: A scoping review. *PLoS ONE*, *12*(5). <https://doi.org/10.1371/journal.pone.0178261>

Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E., Colombo, P., & LaPelle, N. R. (2012). Teaching Research Data Management: An Undergraduate/Graduate Curriculum. *Journal of eScience Librarianship*, *1*(1), 46–50. <https://doi.org/10.7191/jeslib.2012.1003>

Prokscha, S. (2007). *Practical guide to clinical data management* (2nd ed). CRC/Taylor & Francis.

Read, K. B. (2019). Adapting data management education to support clinical research projects in an academic medical center. *Journal of the Medical Library Association*, *107*(1), 89–97. <https://doi.org/10.5195/jmla.2019.580>

Read, K. B., Koos, J., Miller, R. S., Miller, C. F., Phillips, G. A., Scheinfeld, L., & Surkis, A. (2019). A model for initiating research data management services at academic libraries. *Journal of the Medical Library Association*, *107*(3), 432–441. <https://doi.org/10.5195/jmla.2019.545>

REDCap. (o. J.). *About*. Abgerufen 4. Juli 2020, von <https://projectredcap.org/about/>

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (o. J.). *Strategies and Best Practices for Data Literacy Education* (S. 1–75). Dalhousie University.

Scaramozzino, J. M., Ramírez, M. L., & McGaughey, K. J. (2012). A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University. *College & Research Libraries*, 73(4), 349–365. <https://doi.org/10.5860/crl-255>

Schmidt, L., & Holles, J. H. (2018, Juni 23). Teaching Research Data Management: It Takes a Team to Do It Right! *Association for Engineering Education - Engineering Library Division Papers; Atlanta*. <http://search.proquest.com/lisa/docview/2315582539/71998CC707C54CEFPQ/3>

Schneider, R. (2017). Modul-1-1: Grundlagen des Forschungsdatenmanagements. In E. Mastandrea, N. Prongué, & N. Stettler (Hrsg.), *Kursbuch Forschungsdatenmanagement*. HTW Chur – HEG Genève. https://campus.hesge.ch/researchdatamanagement/?page_id=843

Schweizerischer Nationalfonds. (o. J.-a). *Data Management Plan (DMP)—Leitlinien für Forschende*. Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung. Abgerufen 13. April 2020, von http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx

Schweizerischer Nationalfonds. (o. J.-b). *Open Research Data*. Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung. Abgerufen 13. April 2020, von http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx

Schweizerisches Bundesarchiv. (2020). *STANDARDS FÜR DIE ARCHIVIERUNG DIGITALER UNTERLAGEN - Archivtaugliche Dateiformate*. <https://www.bar.admin.ch/bar/de/home/archivierung/ablieferung/digitale-unterlagen.html>

Society for Clinical Data Management. (2013). *Good Clinical Data Management Practices*. <https://scdm.org/wp-content/uploads/2019/05/Full-GCDMP-Oct-2013.pdf>

Surkis, A., LaPolla, F. W. Z., Contaxis, N., & Read, K. B. (2017). Data Day to Day: Building a community of expertise to address data skills gaps in an academic medical center. *Journal of the Medical Library Association*, 105(2), 185–191. <https://doi.org/10.5195/JMLA.2017.35>

Surkis, A., & Read, K. (2015). Research data management. *Journal of the Medical Library Association : JMLA*, 103(3), 154–156. <https://doi.org/10.3163/1536-5050.103.3.011>

The University of Edinburgh. (2020a, Juni). *MANTRA Research Data Management Training*. MANTRA Research Data Management Training. <https://mantra.edina.ac.uk/>

The University of Edinburgh. (2020b, Juni). *Organising data*. MANTRA Research Data Management Training. <https://mantra.edina.ac.uk/organisingdata/>

The University of Edinburgh. (2020c, Juni). *Research data in context*. MANTRA Research Data Management Training. <https://mantra.edina.ac.uk/researchdataexplained/>

Universität Bern. (o. J.). *Forschungsdatenmanagement*. Universitätsbibliothek. Abgerufen 11. Juli 2020, von https://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/service/open_science/forschungsdatenmanagement/index_ger.html

Universität Bern. (2017, Dezember 15). *REDCap*. Department for BioMedical Research (DBMR). https://www.dbmr.unibe.ch/dienstleistungen/medizinale_softwares_floating_licenses/redcap/index_ger.html

Universität Bern. (2019a). *Digitalisierungsstrategie*. <https://www.digitalisierung.unibe.ch>

Universität Bern. (2019b, August 27). *Identifikatoren*. Universitätsbibliothek. https://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/service/open_science/identifikatoren/index_ger.html#pane583367

Universität Bielefeld. (o. J.). *Studieninformation: 39-M-Inf-RDM Research Data Management*. Universität Bielefeld - Studieninformation. Abgerufen 12. Juli 2020, von <https://ekvv.uni-bielefeld.de/sinfo/publ/modul/79251504#kontexte>

University of Oregon. (2020a, Juni 15). *Research Data Management: Backup & Storage*. UO Libraries. [//researchguides.uoregon.edu/data-management/backup](https://researchguides.uoregon.edu/data-management/backup)

University of Oregon. (2020b, Juni 15). *Research Data Management: File Naming & Tracking Changes in Files*. UO Libraries. [//researchguides.uoregon.edu/data-management/filenaming](https://researchguides.uoregon.edu/data-management/filenaming)

University of Oxford. (o. J.). *Research Data Oxford*. Research Data Oxford. Abgerufen 19. Juli 2020, von <https://researchdata.ox.ac.uk/>

University of Oxford. (2013a, Dezember 9). About RDM. *Research Data Oxford*. <http://researchdata.ox.ac.uk/home/introduction-to-rdm/>

University of Oxford. (2013b, Dezember 9). Sharing data. *Research Data Oxford*. <https://researchdata.ox.ac.uk/home/sharing-your-data/>

University of Oxford. (2014, Januar 27). Rights and licensing. *Research Data Oxford*. <https://researchdata.ox.ac.uk/home/sharing-your-data/rights-licensing/>

Weis, E. (2018, Juli 11). *Vertraulichkeit, Integrität und Verfügbarkeit—Schutzziele der Informationssicherheit*. BRANDMAUER IT Security Blog. <https://www.brandmauer.de/blog/it-security/schutzziele-der-informationssicherheit>

Whitmire, A. L. (2015). Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned. *Journal of Librarianship and Scholarly Communication*, 3(2), eP1246. <https://doi.org/10.7710/2162-3309.1246>

Wiljes, C., & Cimiano, P. (2019). Teaching Research Data Management for Students. *Data Science Journal*, 18(1), 1–9. <https://doi.org/10.5334/dsj-2019-038>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L. B. da S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1–9. <https://doi.org/10.1038/sdata.2016.18>

Yale Library. (2020a, März 13). *Research Data Management for the Health Sciences: Data Documentation*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/documentation

Yale Library. (2020b, März 13). *Research Data Management for the Health Sciences: Data Quality*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/quality

Yale Library. (2020c, März 13). *Research Data Management for the Health Sciences: File Naming*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/file_names

Yale Library. (2020d, März 13). *Research Data Management for the Health Sciences: File Versioning*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/versioning

Yale Library. (2020e, März 13). *Research Data Management for the Health Sciences: Metadata and Ontology Schemas*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/metadata_schema

Yale Library. (2020f, März 13). *Research Guides: Research Data Management for the Health Sciences: Data Management Plans*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/dmps

Yale Library. (2020g, März 13). *Yale University Library Research Guides: Research Data Management for the Health Sciences: Home*. Yale Library - Harvey Cushing/John Hay Whitney Medical Library. https://guides.library.yale.edu/rdm_healthsci/home

Zozus, M. N., Lazarov, A., Smith, L. R., Breen, T. E., Krikorian, S. L., Zbyszewski, P. S., Knoll, S. K., Jendrasek, D. A., Perrin, D. C., Zambas, D. N., Williams, T. B., & Pieper, C. F. (2017). Analysis of professional competencies for the clinical research data management profession: Implications for training and professional certification. *Journal of the American Medical Informatics Association : JAMIA*, 24(4), 737–745. <https://doi.org/10.1093/jamia/ocw179>

7 Anhänge

7.1 Ergebnisprotokoll Interview

Thema: Klinisches Forschungsdatenmanagement

Expertin: Muriel Helmers Walker, Head of Data Management, CTU

Interviewerin: Mirjam Nydegger

11. Juni 2020, 14:00 Uhr, Zoom Meeting

Head of Data Management beim CTU: Was ist deine Zuständigkeit /Verantwortlichkeit?

Verantwortlich fürs Aufsetzen der Studiendatenbanken

Data Management heisst beim CTU die Erstellung des Gerüsts für die Daten. Es wird aber nicht mit den Daten gearbeitet. Beratung der Kunden in der HFG Kompatibilität, Good Clinical Practice (GCP) Compliance für die Datensammlung und Erstellen der Datenbank für die Kunden.

Stammen die Daten alle aus klinischen Studien?

Daten stammen nicht nur aus klinischen Studien sondern auch aus Registern etc.

Was für Daten werden bei klinischen Studien produziert? Bspw. Bilder, Video, Excel etc.

Bilder und Video nimmt das CTU nicht, weil es da andere Infrastrukturen braucht. Datenbanken der CTU sind auf strukturierte Daten ausgelegt. Daten werden häufig direkt eingeben, es gibt auch Fälle, bei denen die Daten aus einem anderen System importiert werden.

Dateiformate, was ist speziell? Gibt es proprietäre Formate, die besonders sind? (Hinweis Interoperabilität der Daten)

Meistens werden CSV-Dateien importiert. Am Schluss einer Studie werden die Daten und Metadaten in verschiedenen Dateiformaten (bspw. csv, xml) abgespeichert, damit die Daten später wiederverwendet werden könnten. Ausserdem gibt es auch den Standard CDISC, der wird in der akademischen Forschung noch wenig benützt, weil teilweise das Know-How in den Studienteams fehlt und der Aufwand gross ist. In der Industrie ist dieser Standard bereits etabliert. Die anderen CTU's haben den CDISC auch noch nicht in Verwendung.

Was ist besonders an medizinischen/ klinischen Forschungsdaten im Gegensatz zu anderen Forschungsdaten? Was hat die Pseudoanonymisierung damit zu tun?

Das Umfeld ist sehr reguliert, weil mit Daten von Menschen gearbeitet wird. Datensicherheit und -schutz sind grosse Themen, darum wird mit codierten Daten gearbeitet. Identifizierende Daten sind nur erlaubt, wenn es einen Grund gibt. Forschungsdaten und persönliche Daten werden bei codierten Daten separat aufbewahrt. Codierung und Pseudoanonymisierung ist nicht dasselbe. Bei der Pseudoanonymisierung wird ein Pseudonym erstellt. Aus bspw. dem Jahrgang wird ein Code berechnet, von dem aus nicht auf den Patienten geschlossen werden kann. Aber: Würde sich eine Person doppelt einschreiben, wird auch zweimal derselbe Code berechnet. So können Dubletten identifiziert werden. Häufig werden beim CTU Codes benutzt, die nicht auf eine Person schliessen lassen.

Die Ethikkommission muss jedes Studienprojekt bewilligen. Es wird geschaut, ob die Daten gesammelt werden, die gebraucht werden. Es darf nicht grundlos Daten gesammelt werden. Es gibt ein Studienprotokoll, in dem wird definiert, welche Variablen wann gesammelt werden. Es wird auch eine Risikoanalyse gemacht. Der Benefit soll grösser sein als das Risiko. Es darf genau das gesammelt werden, was angegeben wurde. Sonst muss die Ethikkommission die Abweichung wieder bewilligen.

Es gibt zwei verschiedene Consent: general consent, wird bei Spitaleintritt unterschrieben (Routinedaten). Bei klinischen Studien braucht es einen informed consent. Der muss bei allen Studien eingeholt werden.

Klinische Daten sind besonders schützenswert. Auf welcher gesetzlichen Grundlage? Recht auf Selbstbestimmung? Ärztliche Schweigepflicht?

Im Datenschutzgesetz ist auch festgelegt, was gesammelt werden darf. Ausserdem geht es darum, dass niemand benachteiligt wird, bspw. Versicherungen, die den Patienten nicht mehr annehmen, oder Patienten, die keinen Job mehr erhalten. Diese persönlichen Daten dürfen nicht in die Öffentlichkeit. Dies ist alles zusätzlich zur ärztlichen Schweigepflicht. Es dürfen auch nur die Personen Zugriff auf die Datenbank haben, die die Daten für ihre Arbeit benötigen, alle anderen haben keinen Zugriff. Es ist auch definiert, dass Datenbanken Login, Passwort und Rollenmodelle brauchen, damit keine Person Zugriff hat, die diese nicht für diese Arbeit benötigen.

Wenn Daten weitergeben werden, braucht es noch eine weitere Unterschrift vom Patienten. Ist das zusätzlich? Wie läuft das ab?

Es kommt auf die Studie drauf an. Im informed consent kann schon informiert und geregelt werden, ob und wie die Daten weiterverwendet werden dürfen. Daten werden beim Data-Sharing anonymisiert. Es wird deidentifiziert, damit die Kombination von Feldern keinen Aufschluss auf die Person gibt. Das ist zum Beispiel bei seltenen Krankheiten wichtig.

Beim Gebrauch der Daten im Umfeld der Studie werden die Daten «nur» codiert. Der SNF pusht und fördert Data-Sharing. Forschungsergebnisse sollen anderen zur Verfügung gestellt werden, ausser es gibt einen guten Grund.

Welche Bereiche / Themen beim FDM sind besonders betroffen? Datenschutz, Datensicherheit und Sharing haben wir schon besprochen. Gibt es noch weitere Bereiche, wie Langzeitspeicherung, wo gut darauf geachtet werden muss, damit alles konform ist?

Bei der Datensammlung wie bei der Langzeitaufbewahrung muss darauf geachtet werden, wo Daten gespeichert werden. Falls eine Cloud benutzt wird: Wo der Server steht ist wichtig. Kann sichergestellt werden, dass die Daten in der Schweiz bleiben? Etc. Das sind Fragen, die bedacht werden müssen. Anmerkung Muriel Helmer: Siehe auch https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/Internet_und_Computer/cloud-computing/erlaeuterungen-zu-cloud-computing.html, Kapitel «Risiken bei der Nutzung von Cloud Computing»

Anonymisierung und Pseudonymisierung: scheint gerade bei medizinischen Forschungsdaten besonders wichtig zu sein.

- **Wie wird identifiziert, welche Daten anonymisiert werden müssen?**
Ethikkommission sagt, es darf Geschlecht und Geburtsjahr gesammelt werden, mehr nicht.
- **Das wird von der Ethikkommission vorgegeben?**
Ja.
- **Spielt da die DSGVO auch rein? Oder ist die nicht so relevant für medizinische Daten?**
Doch, ist relevant. Beim Sammeln der Daten ist es auch wichtig, dass das konform ist mit diesen Gesetzen. Der Eidgenössische Datenschutz- und Öffentlichkeitsbeauftragte (EDÖB) ist zuständig bei Datenbearbeitungen durch Bundesorgane und Private. Datenbearbeitungen durch kantonale und kommunale Organe fallen unter kantonales Recht.
- **Wer macht die Codierung/ Anonymisierung, Ärzte, Forscher?**

Das macht die Software, es wird automatisch ein Code generiert. Für die De-Identifizierung eines bestehenden Datensatzes ist noch ein Mensch nötig, der das Wissen hat, um zu entscheiden, was alles anonymisiert werden muss – mehrere nicht identifizierende Datenpunkte können in Kombination wieder identifizierend sein, zum Beispiel bei seltenen Krankheiten oder einer kleinen Population, wie Personen mit einem Alter über 90). Die De-identifizierung wird vor dem Data Sharing gemacht. Nach dem Humanforschungsgesetz ist eine Studie nur anonym, wenn Daten nicht codiert sind, also, wenn bspw. Menschen auf der Strasse angesprochen werden.

Wo muss anonymisiert werden? (Was ist mit Computertomografie (CT)/MRI, MRT vom Kopf?) -> kann das rekonstruiert werden?

Bilder sind ein Problem. In den Metadaten sind oft Patientendaten hinterlegt, aber auch auf dem Bild drauf sind personenbezogene Daten. Darum werden die Bilder oft im Zentrum gehalten und nur die Werte – strukturierte Daten – kommen in die Datenbank. Die Bilder sind nur für das Studienteam einsehbar, werden aber nicht weitergeben. Für die Anonymisierung von Bildern benötigt man spezielle Software, die die Daten sicher entfernt.

Reicht das, wenn man die Zustimmungen der Ethikkommission und der Patienten hat oder sind noch weitere Massnahmen nötig, um die Daten zu verwenden / speichern?

Nein, für die Studie dürfen die Daten dann benützt werden. Ist auch im Consent festgehalten.

Kann der Patient seine Zustimmung zur Verwendung der Daten auch zurückziehen? (Vertrag?)

Nach Abbruch werden keine neuen Daten mehr gesammelt. Nach Humanforschungsgesetz dürfen die Daten vor Abbruch verwendet werden. Es wird aber so gemacht, dass nach Abschluss der Studie der Patient nicht mehr eine Nummer hat (codiert) sondern ein anderes Zeichen, so kann die Verlinkung Datensatz und Person nicht mehr nachvollzogen werden.

Nach Gesetzgebung der EU müssen die Daten gelöscht werden, sobald der Patient seine Teilnahme an der Studie abbricht.

Der SNF fordert die Forschungsdaten zu veröffentlichen, damit die Ergebnisse reproduzierbar sind und bleiben. Ist das ein Problem?

Es wurden oft Studien mit schlechten Ergebnissen/ Werten verschwiegen, weil niemand auf einer schlechten Publikation sein wollte/will. Die klinische Forschung ist aber teuer,

daher kommt auch der Sharing-Gedanke. Es ist aber auch ethisch konnotiert: Wenn Menschen schon an einer Studie teilgenommen haben und ein gewisses Risiko auf sich genommen haben, sollten die Daten zur Verfügung stehen, damit nicht etwas später wieder dieselben Fehler gemacht werden und die Daten auch für andere Studien benützt werden können. Darum ist auch wichtig, dass die Daten gut anonymisiert werden.

Best Practices? Dokumente oder Richtlinien vorhanden im Zusammenhang mit dem Datenmanagement?

HFG

GCP, Humanforschungsgesetz, ECRIN (Hinweis Link). Dieser Standard setzt sich immer wie mehr durch.

GCDMP auch empfehlenswert (Hinweis Link).

REDCap als Tool zur Forschungsdatenverwaltung, auch als Repository?

Neben REDCap wird auch SecuTrial gebraucht. Diese werden aber nicht als Repository benützt. Das Inselspital (Abteilung IDSC) betreibt eins. Beim CTU gibt es nur die Dateninseln der einzelnen Studien, für die der Consent der Patienten vorhanden ist. Weitergabe und Anonymisierung wird nur auf Anfrage und mit Einwilligung des Sponsors gemacht, weil die Daten dem Sponsor gehören.

7.2 Anhang: Tabelle – Aufstellung Vermittlung FDM

Institution	Im Curriculum?	Disziplin o. Allg.	Zielgruppe	Angebot Kurs/ Training/ Workshop?	Vermittlungsmethode auf Website	Weiteres
<u>EPFL Lausanne</u>	k. A.	Allg.	Forscher	Ja	Text	
<u>ETH Zürich</u>	k. A.	Gemischt	Forscher	Ja	Text	
<u>John Hopkins Sheridan Libraries</u>	k. A.	Allg. mit Verweisen auf klinischen FDM	Studierende & Forscher	Nein	Text	Verweis Coursera (klinisches FDM)
<u>MIT Libraries</u>	k. A.	Allg.	Studierende & Forscher	Ja	Text	
<u>New York University Libraries</u>	k. A.	Allg.	k. A.	Ja (online)	Text	
<u>Oregon State University UO Libraries</u>	k. A.	Allg.	k. A.	Nein	Text	<u>Verweise von und auf UO Libraries</u>
<u>Oxford LibGuides</u>	k. A.	Medicine	k. A.	Nein	Text	Verweis auf University of Oxford
<u>Purdue University</u>	k. A.	Allg.	k. A.	Nein	Text	

<u>Queens University Library</u>	k. A.	Allg.	k. A.	Nein	Text	
<u>Universität Basel</u>	k. A.	Allg.	eher Forscher	Ja	Video & Text	Cheatsheet auf Website und Verweis auf swiss clinical trial organisation
<u>Universität Bielefeld</u>	k. A.	Allg.	Studierende	Ja	Text, Verweis auf Seminar	Kurs für Studierende, Forscher und Arbeitsgruppen
<u>Universität Fribourg</u>	k. A.	Gemischt	Forscher	Ja	Text	
<u>Universität Mainz</u>	<u>Ja</u>	Medizin	Studierende	Ja	Text, extra Website	Kursthema ist nicht FDM, sondern Data Literacy
<u>Universität Zürich</u>	k. A.	Allg.	k. A.	Ja	Text	bieten auch Weiterbildungen an
<u>University of Cambridge</u>	k. A.	Gemischt	Studierende & Forscher	Ja	Text	Data Champion
<u>University of Edinburgh</u>	k. A.	Allg.	Allgemeinheit	Ja	Interaktive Website (MANTRA)	Verweis auf MANTRA und Coursera
<u>University of Minnesota</u>	Nein	Allg.	Studierende	Ja (Kurs im Flipped Classroom)	Text (Website enthält ganzes Kursmaterial)	Verweise auf Purdue University, Cornell University und Oregon State University

<u>University of Oxford</u>	k. A.	Gemischt	k. A.	Nein	Text	Verweise auf externe Trainings
<u>University of Pittsburgh</u>	k. A.	Gemischt	k. A.	Ja	Video & Text	
<u>Yale Library</u>	k. A.	Health Science	k. A.	Ja	Text	Verweis DataOne
<u>ZHB Luzern</u>	k. A.	Allg.	k. A.	Ja	Text	

Legende:

k. A. = keine
Angaben

Allg. = Allgemein

Gemischt = teilweise disziplinspezifische
Hinweise

Bisher erschienene Schriften

Ergebnisse von Forschungsprojekten erscheinen jeweils in Form von Arbeitsberichten in Reihen.
Sonstige Publikationen erscheinen in Form von alleinstehenden Schriften.

Derzeit gibt es in den Churer Schriften zur Informationswissenschaft folgende Reihen:
Reihe Berufsmarktforschung

Weitere Publikationen

Churer Schriften zur Informationswissenschaft – Schrift 112

Herausgegeben von Wolfgang Semar

Jin Chei

Digitalisierungslandschaft in Schweizer Archiven

Explorative Studie bezüglich Situation, Herausforderungen und gewünschter Unterstützung

Chur, 2020

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 113

Herausgegeben von Wolfgang Semar

Raphael Jung

Agile Marketing für Digital Marketing Agenturen

Analyse von Projektmanagement Methoden und Definition von konkreten Handlungsempfehlungen

Chur, 2020

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 114

Herausgegeben von Wolfgang Semar

Livia Mosberger

Einflüsse auf das Vertrauen und die Nutzerakzeptanz von Voice Commerce in der Schweiz

Chur, 2020

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 115

Herausgegeben von Wolfgang Semar

Franziska Corradini

Sustainable Development Goals in Bibliotheken

Einfluss der SDGs auf die deutschsprachige Bibliothekswelt und SDG Vermittlungsmöglichkeiten in

Bibliotheken

Chur, 2020

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 116

Herausgegeben von Wolfgang Semar

Marlan Röthlisberger

Digital Nudging Decoy Effect and Social Norms Nudge in E-commerce

Testing the effectiveness of the decoy effect and social norms nudge in the context of an e-

commerce flower store

Chur, 2020

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 117

Herausgegeben von Wolfgang Semar

Mahmoud Hemila

Nutzung von Sprachlern-Abteilungen in öffentlichen Bibliotheken

Chur, 2020

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 118
Herausgegeben von Wolfgang Semar
Rebekka Hirsbrunner
Conversation Design für textbasierte Conversational Agents
Umsetzung eines Dialogkonzepts am Beispiel eines Studienberatungs-Chatbox einer Hochschule
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 119
Herausgegeben von Wolfgang Semar
Nicole Zimmermann
Elektronische Patientenaufklärung in Schweizer Spitälern
Eine Machbarkeitsstudie
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 120
Herausgegeben von Wolfgang Semar
Sarah Stalder
Wie können Barrieren im organisationalen Wissenstransfer abgebaut werden?
Barrieren bei der Teilung und Nutzung von Wissen und Lösungsansätze für die
Unternehmenspraxis
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 121
Herausgegeben von Wolfgang Semar
Saskia Müller
Mit humanoiden Robotern Bewerbungsgespräche trainieren
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 122
Herausgegeben von Wolfgang Semar
Christian Arthur Müller
Archivöffnung: Zugänglichkeit der SRG-Sendungsarchive am Beispiel von SRF
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 123
Herausgegeben von Wolfgang Semar
Susanne Grieder
Archive: Infrastruktur- und Bestandesnutzung durch Menschen mit Sehbehinderung oder Blindheit
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 124
Herausgegeben von Wolfgang Semar
Sophia Zimmerer
Digital Nudging im Pre-Purchase Kontext
Einfluss des Social Norm Nudge im Social-Media-Advertising während der Need Recognition
Phase
Chur, 2020
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 125
Herausgegeben von Wolfgang Semar
Nadine Christinger
Medienpädagogik in Schulbibliotheken
Zukünftige Rolle von Schul- und Gemeindebibliotheken im Bereich der Medienpädagogik am
Beispiel des Kantons St. Gallen
Chur, 2020
ISSN 1660-945X

Über die Informationswissenschaft der Fachhochschule Graubünden

Die Informationswissenschaft ist in der Schweiz noch ein relativ junger Lehr- und Forschungsbereich. International weist diese Disziplin aber vor allem im anglo-amerikanischen Bereich eine jahrzehntelange Tradition auf. Die klassischen Bezeichnungen dort sind Information Science, Library Science oder Information Studies. Die Grundfragestellung der Informationswissenschaft liegt in der Betrachtung der Rolle und des Umgangs mit Information in allen ihren Ausprägungen und Medien sowohl in Wirtschaft und Gesellschaft. Die Informationswissenschaft wird in Chur integriert betrachtet.

Diese Sicht umfasst nicht nur die Teildisziplinen Bibliothekswissenschaft, Archivwissenschaft und Dokumentationswissenschaft. Auch neue Entwicklungen im Bereich Medienwirtschaft, Informations- und Wissensmanagement und Big Data werden gezielt aufgegriffen und im Lehr- und Forschungsprogramm berücksichtigt.

Der Studiengang Informationswissenschaft wird seit 1998 als Vollzeitstudiengang in Chur angeboten und seit 2002 als Teilzeit-Studiengang in Zürich. Seit 2010 rundet der Master of Science in Business Administration das Lehrangebot ab.

Der Arbeitsbereich Informationswissenschaft vereinigt Cluster von Forschungs-, Entwicklungs- und Dienstleistungspotenzialen in unterschiedlichen Kompetenzzentren:

- Information Management & Competitive Intelligence
- Collaborative Knowledge Management
- Information and Data Management
- Records Management
- Library Consulting
- Information Laboratory
- Digital Education

Diese Kompetenzzentren werden im Swiss Institute for Information Research zusammengefasst.

Impressum

Impressum

FHGR - Fachhochschule
Graubünden
Information Science
Pulvermühlestrasse 57
CH-7000 Chur

www.informationsscience.ch

www.fhgr.ch

ISSN 1660-945X

Institutsleitung

Prof. Dr. Ingo Barkow

Telefon: +41 81 286 24 61

Email: ingo.barkow@fhgr.ch

Sekretariat

Telefon: +41 81 286 24 24

Fax: +41 81 286 24 00

Email: clarita.decurtins@fhgr.ch