

# Bausteine Forschungsdatenmanagement

## Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern

### Disziplinübergreifendes Modell zur Ausbildung von Forschungsdatenmanagement- und Data Science-Kompetenzen: „Data Train – Training in Research Data Management and Data Science“

Tanja Hörner<sup>i</sup>

Frank Oliver Glöckner<sup>ii</sup>

Rolf Drechsler<sup>iii</sup>

Iris Pigeot<sup>iv</sup>

2021

#### Zitiervorschlag

Hörner, Tanja, Frank Oliver Glöckner, Rolf Drechsler und Iris Pigeot. 2021. Disziplinübergreifendes Modell zur Ausbildung von Forschungsdatenmanagement- und Data Science-Kompetenzen: „Data Train – Training in Research Data Management and Data Science“. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 3/2021: S. 56-69. DOI: [10.17192/bfdm.2021.3.8343](https://doi.org/10.17192/bfdm.2021.3.8343).

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

<sup>i</sup>Universität Bremen; Bremen Research Alliance. ORCID: [0000-0003-3280-6941](https://orcid.org/0000-0003-3280-6941).

<sup>ii</sup>Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar und Meeresforschung, Bremerhaven; Universität Bremen, Fachbereich 5: Geowissenschaften, Zentrum für Marine Umweltwissenschaften – MARUM. ORCID: [0000-0001-8528-9023](https://orcid.org/0000-0001-8528-9023).

<sup>iii</sup>Deutsches Forschungszentrum für Künstliche Intelligenz, AG Cyber Physical Systems, Bremen; Universität Bremen, Fachbereich 3: Mathematik und Informatik, Data Science Center. ORCID: [0000-0002-9872-1740](https://orcid.org/0000-0002-9872-1740).

<sup>iv</sup>Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen; Universität Bremen, Fachbereich 3: Mathematik und Informatik. ORCID: [0000-0001-7483-0726](https://orcid.org/0000-0001-7483-0726).

## 1 Zusammenfassung

Die datengetriebene Forschung wird immer bedeutsamer, um die Forschungsfragen unserer Zeit zu beantworten. Dies lässt sich seit Längerem beobachten und wird aktuell durch die COVID-19-Pandemie unter Beweis gestellt. In der Wissenschaft und Wirtschaft besteht jedoch ein starkes Defizit an qualifizierten Personen in (Forschungs-) Datenmanagement, Data Science und innovativen Technologien. Diesem Bedarf folgend hat die U Bremen Research Alliance<sup>1</sup> mit Unterstützung des Landes Bremen für Promovierende der Mitgliedseinrichtungen eine institutions- und disziplinübergreifende Ausbildung „Data Train – Training in Research Data Management and Data Science“<sup>2</sup> aufgebaut. Data Train zielt darauf ab, Kompetenzen des wissenschaftlichen Nachwuchses in Data Literacy, Datenmanagement und Data Science zu stärken und dabei die Promovierenden interdisziplinär und überinstitutionell zu vernetzen. An der Entwicklung beteiligen sich vier in Bremen vertretene Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI): NFDI4Health, NFDI4BioDiversity, KonsortSWD und NFDI4Ing. Dieses disziplinübergreifende Ausbildungsmodell, das die gesamte Datenwertschöpfungskette in der Wissenschaft abdeckt und von der auch Wirtschaftsunternehmen profitieren, liefert einen wichtigen Baustein zur Datenkompetenz-Ausbildung weit über den Standort Bremen hinaus.

## 2 Von der digitalen Transformation zum zentralen Ausbildungsprogramm für Promovierende in der U Bremen Research Alliance

Datengetriebene Forschung für die Gesellschaft und Ökonomie sind die Motoren des Fortschritts: Zielgerichtetes Analysieren großer, heterogener Datenmengen, darunter Verfahren Künstlicher Intelligenz (KI) wie Maschinelles Lernen, führt zu einer Prozessoptimierung und Steigerung des Erkenntnisgewinns. Voraussetzung für komplexe und disziplinübergreifende Datenanalysen sind jedoch zugängliche, qualitativ hochwertige und gut dokumentierte Daten. Speziell in der Forschung erfordert dies ein konsequentes Forschungsdatenmanagement (FDM), d. h. a) die detaillierte Beschreibung der Daten mit Metadaten sowie aller Vorgänge ihrer Erhebung (Provenance), b) die Sicherung der Datenqualität und c) eine Datenablage gemäß den „FAIR-Prinzipien“<sup>3</sup> (nach denen Forschungsdaten für Dritte jederzeit auffindbar (*findable*), zugänglich

<sup>1</sup>Der U Bremen Research Alliance e. V. ist ein kollaborativer Zusammenschluss der Universität Bremen mit zwölf der von Bund und Ländern getragenen, außeruniversitären Forschungseinrichtungen sowie weiteren von der Freien Hansestadt Bremen finanzierten Instituten und bildet ein Zentrum exzellenter Wissenschaft (<https://www.uni-bremen.de/research-alliance>).

<sup>2</sup><https://uni-bremen.de/data-train>.

<sup>3</sup>Mark D. Wilkinson et al., „The FAIR guiding principles for scientific data management and stewardship“, *Scientific Data* 3, Nr. 1 (2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

(*accessible*), interoperabel (*interoperable*) und wiederverwendbar (*reusable*) sein müssen) unter Einhaltung der rechtlichen und ethischen Rahmenbedingungen sowie eine Open Data-Kultur.<sup>4</sup> FDM und Data Science (DS) bilden demnach die Basis für eine transparente, nachhaltige sowie zukunftsweisende Forschung und Ökonomie mit hohem Innovationspotenzial. Daher sollten beide Komponenten in den Strukturen etabliert und auch bei Ausbildungskonzepten mitgedacht werden (siehe auch Steinmann und Drechsler, diese Ausgabe).<sup>5</sup>

Der Bedarf an Datenkompetenz ist hoch: IBM und das Business-Higher Education Forum in Zusammenarbeit mit Burning Glass prognostizierten 2017 für den Arbeitsmarkt der USA einen Zuwachs von 365.000 auf 2.720.000 Stellenangebote mit Daten- und Datenanalytik-Kompetenz bis 2020.<sup>6</sup> Bis 2023 werden hierzulande mittlerweile 700.000 Spezialist:innen für „komplexe Datenanalysen“ und „innovative Technologien“ gesucht. Demnach müssten 140.000 Personen jährlich Daten- und Technologiekompetenzen erlernen, wie der Stifterverband für die Deutsche Wissenschaft in Kooperation mit McKinsey & Company betont.<sup>7</sup> Der Rat für Informationsinfrastrukturen veröffentlichte eine Untersuchung, gemäß derer die Qualität und damit auch die Qualitätskontrolle wissenschaftlicher Daten verbessert werden muss,<sup>8</sup> und fordert unter anderem die Etablierung von Berufsbildungs- und Ausbildungsangeboten<sup>9</sup>. Institutsübergreifende Lösungsansätze wie z. B. gemeinsame Angebote und Austausch zu Best Practice-Beispielen empfiehlt die Allianz der deutschen Wissenschaftsorganisationen und gibt Leitgedanken für die Umsetzung an die Hand.<sup>10</sup> Als Zeichen des Aufbruchs hat Anfang dieses Jahres der Stifterverband im Einklang mit der Datenstrate-

<sup>4</sup>Brian A. Nosek et al., „Promoting an open research culture“, *Science* 348, Nr. 6242 (2015): 1422-1425, <https://doi.org/10.1126/science.aab2374>.

<sup>5</sup>Lena Steinmann und Rolf Drechsler, „Verzahnung von Data Stewardship und Data Science – Wege und Perspektiven“, *Bausteine Forschungsdatenmanagement*, Nr. 3 (2021), <http://doi.org/10.17192/bfdm.2021.3.8342>.

<sup>6</sup>Will Markow et al., „The quant crunch. How the demand for data science skills is disrupting the job market“, *Burning Glass Technologies* (2017), <https://www.ibm.com/downloads/cas/3RL3VXGA> (25.04.2021).

<sup>7</sup>Volker Meyer-Guckel et al., *Future Skills: Welche Kompetenzen in Deutschland fehlen* (Essen: Stifterverband für die Deutsche Wissenschaft, 2018), <https://www.stifterverband.org/medien/future-skills-welche-kompetenzen-in-deutschland-fehlen> (25.04.2021).

<sup>8</sup>Rat für Informationsinfrastrukturen (RfII), *Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel* (Göttingen: Rat für Informationsinfrastrukturen, 2019), <urn:nbn:de:101:1-2019112011541657732737>.

<sup>9</sup>Rat für Informationsinfrastrukturen (RfII), *Digitale Kompetenzen - dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft* (Göttingen: Rat für Informationsinfrastrukturen, 2019), <http://www.rfii.de/?p=3883> (25.04.2021).

<sup>10</sup>Allianz der deutschen Wissenschaftsorganisationen, *Wege zur digitalen Qualifikation: ein Diskussionspapier* (Allianz der deutschen Wissenschaftsorganisationen, 2020), <https://doi.org/10.2312/allianzoa.038>.

gie der Bundesregierung<sup>11</sup> und der Berliner Erklärung zur Digitalen Gesellschaft<sup>12</sup> eine Data-Literacy-Charta publiziert, die die für alle Menschen erforderlichen Datenkompetenzen des digitalen Zeitalters umfasst und diese „als unverzichtbaren Bestandteil der Allgemeinbildung“ deklariert.<sup>13</sup>

Um Kompetenzen in Data Literacy, FDM und DS zu stärken, baut die U Bremen Research Alliance mit Unterstützung des Landes Bremen seit Dezember 2019 die Ausbildung „Data Train – Training in Research Data Management and Data Science“ für Promovierende der Mitgliedsinstitutionen (vgl. [Abbildung 1](#)) als freiwilliges Zusatzangebot auf. Data Train ist eine instituts- und disziplinübergreifende Ausbildung, die zusätzlich zum Erwerb der Domänenexpertise angeboten wird. Die Zielgruppe (Promovierende) wird die erlangten Datenkompetenzen sowohl in die Forschung als auch in die Wirtschaftslandschaft einbringen. Damit findet auch die eingangs erwähnte Bewusstseinschärfung („Open Data-Kultur“) in beiden Sektoren statt.

An der Entwicklung beteiligen sich vier in Bremen vertretene NFDI-Konsortien (vgl. [Abbildung 1](#)), die seit Oktober 2020 mit dem Aufbau der nationalen Forschungsdateninfrastruktur begonnen haben: NFDI4Health<sup>14</sup>, NFDI4BioDiversity<sup>15</sup>, KonsortSWD<sup>16</sup> und NFDI4Ing<sup>17</sup> (siehe Sonderband von „Bausteine Forschungsdatenmanagement“ zur Nationalen Forschungsdateninfrastruktur (2021)). Alle neun geförderten Konsortien der ersten Ausschreibungsrunde im Jahr 2019 definierten das Schaffen von Qualifizierungsangeboten zur Verbesserung der Datenkompetenz als eines ihrer Querschnittsthemen.<sup>18</sup> Der Aufbau von Data Train erfolgt in enger Zusammenarbeit mit dem Data Science Center der Universität Bremen<sup>19</sup> (siehe Steinmann und Drechsler, diese Ausgabe); neben der gemeinsamen Entwicklungsarbeit stellt das im Jahr 2020 gegründete Data Science Center die technische Infrastruktur für die operative Durchführung der Lehrveranstaltungen zur Verfügung.

<sup>11</sup>Bundesregierung, *Datenstrategie der Bundesregierung. Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum. Kabinetttfassung* (Berlin: Presse- und Informationsamt der Bundesregierung, Januar 2021), <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632> (25.04.2021).

<sup>12</sup>Europäische Union, *Berliner Erklärung zur Digitalen Gesellschaft und wertebasierten digitalen Verwaltung* (Berlin: Bundesministerium des Inneren, für Bau und Heimat, Dezember 2020), [https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2020/eu-rp/gemeinsame-erklarungen/berliner-erklarung-digitale-gesellschaft.pdf?\\_\\_blob=publicationFile&v=10](https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2020/eu-rp/gemeinsame-erklarungen/berliner-erklarung-digitale-gesellschaft.pdf?__blob=publicationFile&v=10) (25.04.2021).

<sup>13</sup>Katharina Schüller, Henning Koch und Florian Rampelt, *Data Literacy Charta, Version 1.1* (Essen: Stifterverband für die deutsche Wissenschaft, Januar 2021), [https://www.stifterverband.org/sites/default/files/data-literacy-charta\\_v1\\_2.pdf](https://www.stifterverband.org/sites/default/files/data-literacy-charta_v1_2.pdf) (25.04.2021).

<sup>14</sup><https://www.nfdi4health.de>.

<sup>15</sup><https://www.nfdi4biodiversity.org>.

<sup>16</sup><https://www.konsortswd.de>.

<sup>17</sup><https://nfdi4ing.de>.

<sup>18</sup>Maik Bierwirth et al., *Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung* (Juni 2020), <http://doi.org/10.5281/zenodo.3895209>.

<sup>19</sup><https://www.dsc-ub.de>.



Abbildung 1: Mitgliedsinstitutionen der U Bremen Research Alliance zusammen mit den vier NFDI-Konsortien der ersten Förderphase, an denen Bremer Institutionen beteiligt sind und die an der Entwicklung von Data Train mitwirken.

## 2.1 Die Entwicklungsphase

Die Entwicklung des Grundkonzepts und des Curriculums wird von der Arbeitsgruppe „Forschungsdaten“<sup>20</sup> der U Bremen Research Alliance begleitet, in der alle Mitgliedsinstitutionen durch Repräsentant:innen vertreten sind. Nach einer intensiven externen Recherche zu bestehenden Ausbildungskonzepten<sup>21</sup> und Curricula<sup>22</sup> wurde eine qualitative Erfassung des Status und der Ausbildungsbedarfe in der U Bremen Research Alliance durchgeführt. In über 70 leitfadengestützten Interviews wurden Kompetenzbedarfe in FDM und DS bei Angehörigen verschiedener Statusgruppen und Sektoren (Management, Infrastruktur und Forschung) erfasst. Fokussiert wurde auf (Senior-) Wissenschaftler:innen aller Wissenschaftsbereiche, Repräsentant:innen der in Bremen ansässigen Forschungsdatenrepositorien (PANGAEA<sup>23</sup>, Qualiservice<sup>24</sup>) und Koordinator:innen interner und externer Graduiertenprogramme, die Qualifizierungsbedarfe aus ihrer täglichen Arbeit ableiten können. Der qualitative Ansatz wird durch eine quantitative Erhebung ergänzt. Zum einen bilden die Informationen aus den Befragungen die Basis für die Entwicklung des zielgruppenspezifischen Data Train-Konzepts und des Curriculums. Zum anderen sind sie die Grundlage für zukünftige themenbezogene Aktivitäten (z. B. Unterstützungsangebote, Kooperationsvorhaben, etc.) in der U Bremen Research Alliance.

## 3 Das Data Train-Modell

Das Ausbildungsprogramm Data Train wird instituts- und disziplinübergreifend für Promovierende angeboten und zielt daher auf grundlegende Kompetenzen in FDM und DS ab. Data Train setzt sich aus vier Ausbildungskomponenten zusammen:

1. Lehrveranstaltungen (Curriculum, angeordnet in „Tracks“, siehe [Abschnitt 3.1](#)),
2. eingeladene Vorträge aus Wirtschaft und Wissenschaft („Data Stories“),
3. Zusammenstellung von existierenden digitalen Aus- bzw. Weiterbildungsmaterialien („Digital Toolkit“),
4. Plattform für ungezwungene Treffen zur Vernetzung und zum Austausch („Hacky Hours“).

<sup>20</sup><https://www.uni-bremen.de/research-alliance/forschungsdaten/ag-forschungsdaten>.

<sup>21</sup>Unter anderem Andreas Grillenberger und Ralf Romeike, *Vorstudie. Hochschulübergreifende Konzepte zum Erwerb von 21st century skills am Beispiel von Data Literacy* (Berlin, 2019), <http://doi.org/10.5281/zenodo.2633091>.

<sup>22</sup>Unter anderem Angebote und Studiengänge (FDM und DS) der Universität Bielefeld, der Georg-August-Universität Göttingen, der Technischen Universität Dortmund, der Ludwig-Maximilians-Universität München, des Hasso-Plattner-Instituts, der City University of New York, des Massachusetts Institute of Technology und der Fraunhofer-Gesellschaft.

<sup>23</sup><https://www.pangaea.de/>.

<sup>24</sup><https://www.qualiservice.org/>.

Das Curriculum (1) ist die am stärksten gewichtete und umfassendste Komponente. Die weiteren drei Komponenten (2) – (4) sollen zusätzlich zu den Lehrveranstaltungen angeboten werden und befinden sich derzeit noch in der Entwicklung. Zur Vermittlung der Basiskompetenzen wird Expertise aus der gesamten Allianz zentral gebündelt und gemeinschaftlich weitergegeben. Lehrende aus den verschiedenen Institutionen, den NFDI-Konsortien und Wirtschaftsunternehmen, die ebenfalls die Notwendigkeit des Aufbaus von (Personal-)Ressourcen erkannt haben, tragen zum Curriculum bei. Dies optimiert Ressourcen und fördert zudem die Vernetzung der Promovierenden über Institutionen und Disziplingrenzen hinweg. Besonders für Promovierende ist die interdisziplinäre Vernetzung in einer frühen Karrierephase vorteilhaft. Über die Grundlagen hinaus gilt es natürlich, disziplinspezifische Qualifikationen zu beachten, die aufbauend auf den vermittelten Basiskompetenzen in den Wissenschaftsfeldern und NFDI-Konsortien ausgebildet werden könnten. Dieses Modell einer zentralen, institutsübergreifenden Basisausbildung, die durch fachspezifische Inhalte dezentral erweitert wird, kann als Best Practice für weitere Standorte dienen.

### 3.1 Das Curriculum

Aus der Recherche und den Befragungen abgeleitete **Anforderungen (A)**:

Das Curriculum soll beide Kompetenzfelder, FDM und DS, basal abdecken (A1). Es wird parallel zur Promotion und strukturierten Promotionsprogrammen auf einer freiwilligen Basis angeboten. Die limitierten zeitlichen Ressourcen der Promovierenden (A2), ihre stark variierende thematische Ausrichtung (A3) und die unterschiedlichen Vorkenntnisse (A4) müssen berücksichtigt werden. Die Dynamik der Themenfelder verlangt darüber hinaus ein flexibles Grundgerüst an Kursen (A5).

Strategie zur **Umsetzung** der Anforderungen:

Inhaltlich bildet das Curriculum die gesamte Datenwertschöpfungskette, d. h. alle Schritte des Datenlebenszyklus unter Einbeziehung der Arbeitsschritte von Data Science-Anwendungen, wie in Stodden (2020)<sup>25</sup> gezeigt, aus der Sicht Forschender ab (A1, vgl. [Abbildung 2](#)). Dadurch werden Synergie- und Innovationspotenziale beider Kompetenzfelder bei der Gestaltung des Curriculums genutzt. Die einzelnen Kurse orientieren sich strukturell an der Datenwertschöpfungskette, können jedoch auch für sich alleine stehen (A5). Dies erlaubt unkomplizierte Modifikationen des Curriculums und verschafft den Promovierenden die Möglichkeit, Kurse auszuwählen (A2, A3). Außerdem werden die Lehrveranstaltungen ihrem Schwierigkeitsgrad und dem Kompetenzbereich entsprechend eingeteilt (A3, A4).

<sup>25</sup>Victoria Stodden, „The data science life cycle: a disciplined approach to advancing data science as a science“, *Communications of the Association of Computing Machinery* 63, Nr. 7 (2020): 58–66, <https://doi.org/10.1145/3360646>.



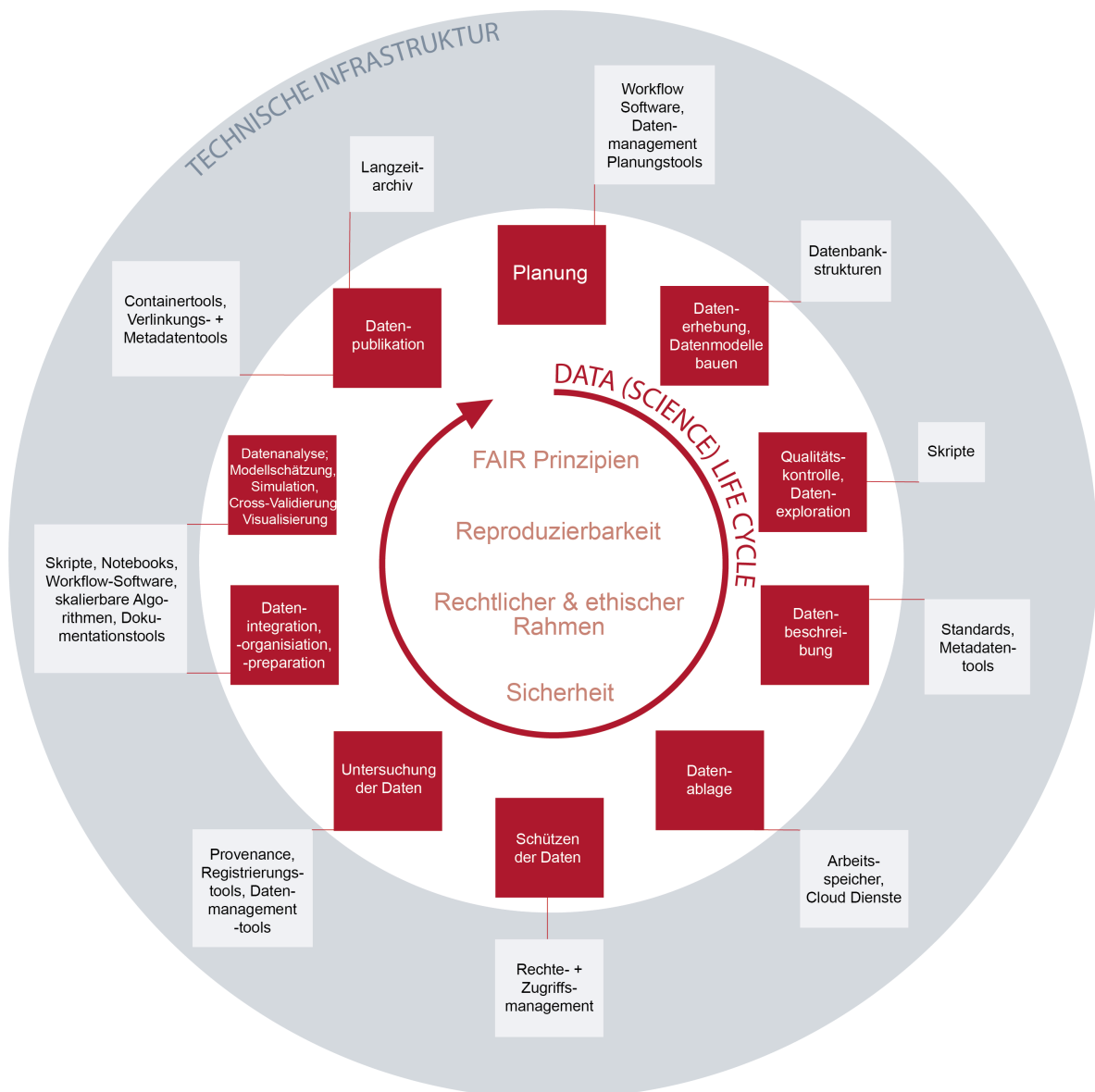


Abbildung 2: Schematische Darstellung des Datenlebenszyklus, modifiziert nach GFBio e. V.<sup>26</sup> und Stodden (2020)<sup>27</sup>. Der innere Kreis (weiß hinterlegt) zeigt die einzelnen Schritte des Datenlebenszyklus unter Einbeziehung der Arbeitsschritte eines Data Scientists. Der äußere Kreis (grau hinterlegt) zeigt die für FDM und DS notwendige, technische Infrastruktur. Aus den einzelnen Arbeitsschritten und der zu bedienenden technischen Infrastruktur lassen sich Kompetenzansprüche synergetisch ableiten.

<sup>26</sup><https://www.gfbio.org/training/materials/data-lifecycle>.

<sup>27</sup>Victoria Stodden, „The data science life cycle“.



Es gibt dementsprechend drei Ausbildungspfade, bestehend aus einzelnen Kursen, (im Folgenden bezeichnet als „Tracks“), die Promovierende beschreiten können: einen Pfad für Einsteiger:innen („Starter Track“) mit Überblicksvorlesungen zu übergeordneten Themen beider Kompetenzfelder und zwei Pfade („Operator Tracks“) mit Hands-On-Workshops, thematisch unterteilt in den „Data Steward Track“ und den „Data Scientist Track“ (siehe [Abbildung 3](#)).

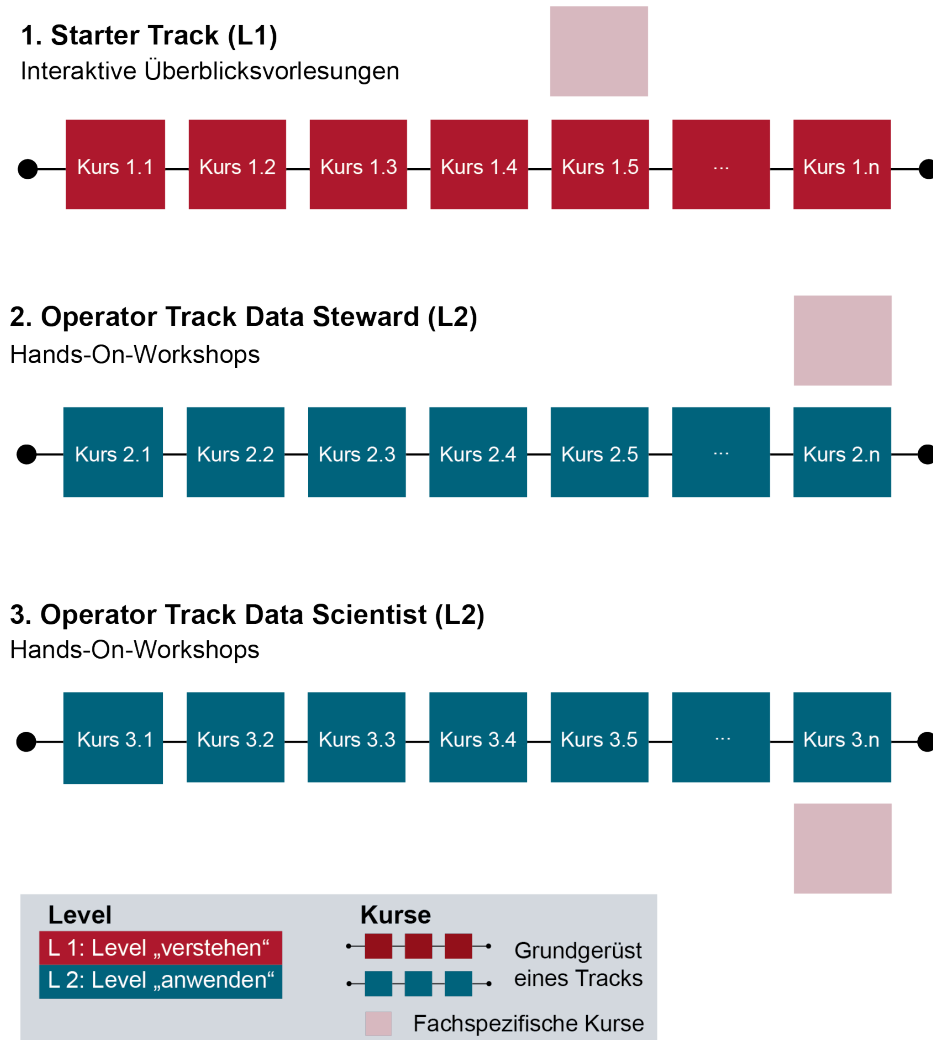


Abbildung 3: Schematische Darstellung der Einteilung der Lehrveranstaltungen. Dargestellt sind die drei Tracks des Programms. Kästchen symbolisieren einzelne Kurse. Alle Kurse, die zu einem Track gehören und damit das Grundgerüst des Tracks bilden, sind durch eine schwarze Linie verbunden. Überblicksvorlesungen des Starter Tracks (Level 1) sind rot, Hands-On-Workshops der Operator Tracks (Level 2) sind blau. Hellrote Kästchen stehen für fachspezifische Kurse, die aufbauend auf den im Rahmen von Data Train vermittelten Grundlagen in den wissenschaftlichen Domänen oder den NFDI-Konsortien etabliert werden sollen.

### 3.1.1 Definition und Selektion der Kompetenzen

Der Starter Track soll die Promovierenden an übergeordnete Themen beider Kompetenzfelder (FDM und DS) heranführen und Basiswissen überblicksartig vermitteln. Er dient als Orientierungshilfe für die spätere Auswahl der zeitintensiveren Workshops der Operator Tracks (durchschnittliche Dauer: 2-3 Tage) und findet daher als erstes in jedem Durchlauf statt. Grundlagen folgender Themenfelder werden in der Pilotphase getestet und wurden im Rahmen des ersten Starter Tracks vermittelt: Data Science, Big Data, Statistik, Informatik (Programmiersprachen, Kryptographie, Privacy und Systemsicherheit), FDM (FAIR-Prinzipien, Datenschutz und Urheberrecht, Management von schützenswerten Daten (Industriedaten und personenbezogene Daten), Management von qualitativen Daten).

Für die Konzeptionierung der Operator Tracks ist es erforderlich, die Kernkompetenzen beider Bereiche anhand der Datenwertschöpfungskette zu definieren und Schnittmengenpotentiale synergetisch zu nutzen (Abbildung 2). Die Operator Tracks finden daher nacheinander im Anschluss an den Starter Track statt. Dabei werden, wie bereits erwähnt, sowohl die Anforderungen an einen Data Steward als auch den Data Scientist berücksichtigt.

Die Aufgabenfelder eines **Data Stewards** unterscheiden sich je nach Disziplin, Anwendungsgebiet innerhalb der Wissenschaft und lokalen Rahmenbedingungen.<sup>28</sup> Einigkeit herrscht darüber, dass Data Stewards die FAIR-Prinzipien umsetzen sollen.<sup>29</sup> Die FAIR „Datenkultur“ muss nach Scholten et al. (2019) in verschiedenen Sektoren der Wissenschaft umgesetzt werden: A) Infrastruktur, B) Management (übergeordnet Politik), C) Forschung.<sup>30</sup> Diesem Ansatz folgend konzentriert sich die Data Train-Ausbildung zielgruppenspezifisch auf den Sektor Forschung und die essentiellen Kompetenzen, die Forschende für den FAIRen Umgang mit Daten benötigen: Programmierfähigkeiten (pilotiert: R, Python, MATLAB), Workflows im FDM (pilotiert: Datenmanagementpläne, Reproduzierbarkeit, Datenaufbereitung/-prozessierung), innovative Software- und IT-Anwendungen (pilotiert: Datenbanken, Datenbereitstellung bzw. -extraktion aus Online-Plattformen) – speziell auch für kollaboratives Arbeiten (pilotiert: Git/GitHub) (siehe auch Abbildung 4a)). Das Training im Rahmen des Data Steward Tracks ist nicht mit einer Ausbildung für professionelle Data Stewards gleichzusetzen. Die Ausbildung von professionellen, fachspezifischen Data Stewards, die beratend und koordinierend

<sup>28</sup>Laura Rothfritz, „Data Stewardship als Boundary-Work“ (Masterarbeit, Humboldt-Universität zu Berlin, 2019), <http://dx.doi.org/10.18452/20636>.

<sup>29</sup>Z. B. Angus Whyte et al., *Skills and capability framework. Deliverable 7.3.* (EOSCpilot, 2018), <https://eoscipilot.eu/sites/default/files/eoscipilot-d7.3.pdf> (25.04.2021): „As the FAIR principles articulate, stewardship implies long-term and sustainable care across multiple lifecycles. This makes stewardship a collective endeavour, involving at least the individual researcher, colleagues in the study (the team), their host organisation (or and others providing services) and the research domain(s) or communities that care about the data“ (S. 29).

<sup>30</sup>Salome Scholtens et al., *Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise* (2019), <http://doi.org/10.5281/zenodo.3474789>.

in Wissenschaftsinstitutionen tätig sind, sollte darüber hinaus verfolgt werden.<sup>31</sup>

**Data Scientists** benötigen Kompetenzen in Statistik, Mathematik, Informatik sowie die jeweilige Domänenexpertise (Abbildung 4b). Die thematische Ausgestaltung innerhalb dieser übergeordneten Kompetenzen basiert auf der Bedarfsanalyse und orientiert sich an existierenden Curricula<sup>32</sup>, Artikeln<sup>33</sup> und Positionspapieren.<sup>34</sup>

Das Einbeziehen der Bedarfsformulierungen in die Auswahl der Lehrinhalte führt dazu, dass diese in gewisser Weise spezifisch für das in Bremen vorliegende Forschungs- und Anforderungsspektrum ist. Es werden daher nicht nur rein technische bzw. formale Veranstaltungen angeboten (pilotiert: Quantitative Analysemethoden, Machine Learning (ML)-Methoden, Deep Learning-Methoden, Datenvisualisierung). So wird als Teil des Curriculums „Critical Thinking“ vermittelt (z. B. Evaluierungsmethoden für KI- / ML-Algorithmen und philosophische Fragestellungen in Bezug auf Visualisierungen, siehe auch Abbildung 4), das als notwendig erachtet wird, um angemessen und verantwortlich über die Aufbereitung, Weitergabe und Verwendung von Daten entscheiden zu können. Über Disziplinen hinweg wird so eine gemeinsame Sprache etabliert, die sich möglicher Grenzen und Schwierigkeiten bewusst ist. Dies führt zu mehr informierter Akzeptanz und Offenheit gegenüber neuartigen Arbeitsweisen, Forschungsansätzen und Technologien.<sup>35,36</sup> Thematisiert werden außerdem die ethischen und rechtlichen Rahmenbedingungen, die die Arbeiten eines Data Stewards und Data Scientists flankieren.<sup>37,38</sup>

<sup>31</sup>Mijke Jetten et al., *Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship* (2021), <http://doi.org/10.5281/zenodo.4486423>.

<sup>32</sup>Unter anderem Angebote und Studiengänge (FDM und DS) der Universität Bielefeld, der Georg-August-Universität Göttingen, der Technischen Universität Dortmund, der Ludwig-Maximilians-Universität München, des Hasso-Plattner-Instituts, der City University of New York, des Massachusetts Institute of Technology und der Fraunhofer-Gesellschaft.

<sup>33</sup>Miguel A. Hernán, John Hsu und Brian Healy, „A second chance to get causal Inference right: A classification of data science tasks“, *Chance* 32, Nr. 1 (2019): 42-49, <https://doi.org/10.1080/09332480.2019.1579578>; Cord Wiljes und Philipp Cimiano, „Teaching research data management for students“, *Data Science Journal* 18, Nr. 38 (2019): 1-9, <https://doi.org/10.5334/dsj-2019-038>.

<sup>34</sup>Unter anderem Yuri Demchenko, Adam Belloum und Tomasz Wiktorski, *EDISON Data Science Framework: Part 1. Data science competence framework (CF-DS)* (EDISON, 2017), [https://edison-project.eu/sites/edison-project.eu/files/filefield\\_paths/edison\\_cf-ds-release2-v08\\_0.pdf](https://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-release2-v08_0.pdf) (25.04.2021).

<sup>35</sup>Norman Sieroka, Vivianne I. Otto und Gerd Folkers, „Critical thinking in education and research – why and how?“, *Angewandte Chemie* 57 (2018): 16574-16575, <http://doi.org/10.1002/anie.201810397>.

<sup>36</sup>Peter Schneider et al., „Rethinking drug design in the artificial intelligence era“, *Nature Reviews Drug Discovery* 19 (2020): 353–364, <http://doi.org/10.1038/s41573-019-0050-3>.

<sup>37</sup>Unter anderem Jenna Burrell, „How the machine ‚thinks‘: Understanding opacity in machine learning algorithms“, *Big Data & Society* 3, Nr. 1 (2016) : 1-12, <https://doi.org/10.1177/2053951715622512>.

<sup>38</sup>Brent D. Mittelstadt et al., „The ethics of algorithms: Mapping the debate“, *Big Data & Society* 3, Nr. 2 (2016): 1-21, <https://doi.org/10.1177/2053951716679679>.

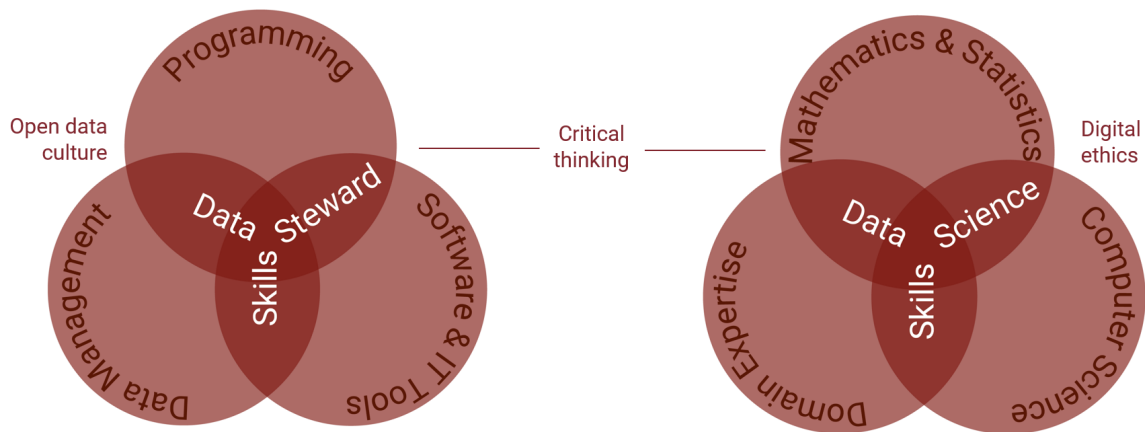


Abbildung 4: a) Data Steward-Kompetenzen, b) Data Science-Kompetenzen.

## 4 Die Pilotphase

Das Curriculum befindet sich derzeit in der Pilotphase. In dieser Phase werden die Thementauswahl und -auslegung sowie die aufeinander abgestimmte thematische Struktur optimiert.

In einem Seminar (Wintersemester 20/21) für Studierende aller Fachbereiche der Universität Bremen wurden grundlegende Themen aus beiden Kompetenzfeldern in einem ersten Test gelehrt. Ziel des Seminars war es, einen interdisziplinären Überblick über aktuelle Themen aus FDM und DS zu vermitteln. Dabei wurden weniger technische und methodische Fähigkeiten vermittelt, sondern vielmehr die wesentlichen Anwendungsfelder und grundlegenden Ideen möglichst interdisziplinär erarbeitet. Fachexpert:innen, hauptsächlich aus der U Bremen Research Alliance, aber auch eines Wirtschaftsunternehmens aus Bremen und des Instituts für Softwaretechnologie – Deutsches Zentrum für Luft- und Raumfahrt (DLR) in Braunschweig beteiligten sich an dem Seminar. Das pandemiebedingte Online-Format erlaubte es, interessierte Personen aller Statusgruppen als Zuschauer:innen an dem Seminar teilnehmen zu lassen. Im Anschluss an das Seminar laufen im Jahr 2021 zum ersten Mal alle drei Tracks nacheinander ab: Starter Track – Operator Track Data Steward – Operator Track Data Scientist. Die Starter Track-Kurse setzen keine Vorkenntnisse voraus. Sofern die Veranstaltungen rein virtuell oder hybrid stattfinden und somit hohe Teilnehmendenzahlen zulassen, werden die Veranstaltungen für alle Statusgruppen und Externe geöffnet. Zwar werden auch in den Hands-On-Workshops Grundlagen vermittelt und angewandt, dennoch setzen sie teils Vorkenntnisse voraus, die in den Kursbeschreibungen dargelegt werden. Für die praktischen Workshops sind die maximalen Teilnehmendenzahlen meist auf 20-30 Personen limitiert, damit eine angemessene Betreuung der einzelnen Teilnehmenden gewährleistet ist. Bei großer Nachfrage werden zunächst Promovierende aus der U Bremen Research Alliance, anschließend andere Statusgruppen oder Externe berücksichtigt.

Alle Veranstaltungen werden durch die Teilnehmenden evaluiert. In Online-Umfragen werden u. a. die Inhalte, der Mehrwert durch den disziplinübergreifenden Austausch, die Relevanz des Themas und das Anforderungslevel bewertet. Die Evaluierung der Veranstaltungen stellt einen Meilenstein in der Pilotphase dar. Sie ist essentiell für die Optimierung der abgestimmten thematischen Struktur und für die bedarfsgerechte Themenausrichtung. Sobald eine erste Konsolidierung erreicht und die Relevanz des Programms durch die verschiedenen Akteure bestätigt ist, werden externe Gutachter:innen für den Evaluierungsprozess hinzugezogen. Intern wird das Programm durch ein Leitungsgremium wissenschaftlich und inhaltlich begleitet. Zudem wurde eine interdisziplinäre und institutsübergreifende Arbeitsgruppe von Fachexpert:innen eingerichtet, in der das Konzept und speziell das Curriculum regelmäßig auf dem Prüfstand stehen.

## 5 Diskussion

An der Ausbildung beteiligen sich hochschul- oder außeruniversitär affilierte Lehrende aller Wissenschaftsbereiche aus Überzeugung, dass ein solches Ausbildungsprogramm dringend erforderlich ist. Dieses Programm zeigt darüber hinaus, dass bereits in der Ausbildung Expertise disziplinübergreifend zusammengeführt werden kann. Das damit geschaffene überinstitutionelle Netzwerk steigert die Wertschöpfung und wird im digitalen Zeitalter immer gefragter.<sup>39</sup>

Das Data Train-Modell kann ggf. für andere Standorte als Orientierung dienen und somit einen wichtigen Baustein zur Steigerung der Datenkompetenz beitragen. Dank digitaler Lehrkonzepte und innovativer Informations- und Kommunikationstechnologien kann ein solches Konzept auch hochschulübergreifend über größere Entfernungen hinweg implementiert werden. Das Data Train-Konzept wird bei Bedarf weiteren Standorten im Rahmen des übergeordneten NFDI-Vorhabens zur Verfügung gestellt. Perspektivisch wird angestrebt, Lehrmaterialien mit der FDM-Fachcommunity zu teilen, wobei zu beachten ist, dass die Entscheidung darüber bei den Lehrenden liegt. Herauszustellen ist, dass die Umsetzung der FAIR-Prinzipien in der Hand eines bzw. einer jeden einzelnen liegt. Durch Data Train werden notwendige Data Steward- und Data Scientist- Kompetenzen ausgebildet<sup>40</sup> und damit das Bewusstsein für FDM und die „Open Data-Kultur“ auf Ebene der Forschenden in der U Bremen Research Alliance geschärft. Schon jetzt zeichnet sich statusgruppenunabhängig ein großes Interesse an den Angeboten des Ausbildungsprogramms ab, so dass zu überlegen ist, eine derart flexible Ausbildung auf Nachfrage für weitere Zielgruppen des Wissenschaftsbereichs zu öffnen.

Langfristig sollte die Ausbildung von Datenkompetenzen im Hochschulbereich verpflichtend in Promotionsprogramme genauso wie in Curricula von Studiengängen auf-

<sup>39</sup>Allianz der deutschen Wissenschaftsorganisationen, *Wege zur digitalen Qualifikation*.

<sup>40</sup>Europäische Kommission, *Turning FAIR into reality – Final report and action plan from the European Commission* (Brüssel: Europäische Kommission, 2018), <https://doi.org/10.2777/1524>.

genommen werden. In der U Bremen Research Alliance ist bereits jetzt zu beobachten, dass auf Fachbereichsebene die Ausbildung von FDM-Qualifikationen verpflichtend in Promotionsvereinbarungen und Curricula aufgenommen wird. Dies ist ein erster wichtiger Schritt, der den Ausbau von Qualifikationsangeboten nach sich ziehen kann. Mit einer solchen Ausbildung werden Nachwuchswissenschaftler:innen bzw. BerufseinsteigerInnen befähigt, innovative Ansätze und Technologien im Bereich Forschungsdatenmanagement und Data Science zu nutzen. Laut einer Studie des Stifterverbands arbeitet heute jedes vierte Unternehmen mit Hochschulen zusammen und in fünf Jahren jedes dritte (37%).<sup>41</sup> Perspektivisch werden Business Analytics und Big Data Systeme somit stärker in den Vordergrund der Ausbildung rücken.

---

<sup>41</sup>Volker Meyer-Guckel et al., *Future Skills: Welche Kompetenzen in Deutschland fehlen* (Essen: Stifterverband für die Deutsche Wissenschaft, 2018), <https://www.stifterverband.org/medien/future-skills-welche-kompetenzen-in-deutschland-fehlen> (25.04.2021).