

UBremen
Research
Alliance

Etablierung eines kooperativen Forschungsdatenmanagements in der U Bremen Research Alliance

– Whitepaper –





In der **U Bremen Research Alliance** kooperieren die Universität Bremen und zwölf Institute der von Bund und Ländern finanzierten außeruniversitären Forschung mit Sitz im Bundesland Bremen. Die Zusammenarbeit erstreckt sich über die vier Wissenschaftsschwerpunkte „Meeres-, Polar- und Klimaforschung“, „Materialwissenschaften und ihre Technologien“, „Gesundheitswissenschaften“ und „Minds, Media, Machines“. Die U Bremen Research Alliance will hierbei beste Bedingungen bieten, um frei, innovativ, kreativ, nachhaltig und kooperativ zu forschen. Dabei repräsentiert sie den Kernbereich der Bremer Wissenschaft.

Von der Tiefsee bis ins Weltall

Redaktionsgruppe

- Prof. Dr. Iris Pigeot – Universität Bremen und Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS
- Prof. Dr. Frank Oliver Glöckner – Universität Bremen und Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)
- Prof. Dr. Rolf Drechsler – Universität Bremen und Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Bremen
- Dr. Tanja Hörner – Universität Bremen und U Bremen Research Alliance
- Derk H. Schönfeld – Universität Bremen und U Bremen Research Alliance
- Dr. Lena Steinmann – Universität Bremen
- Dr. Björn Oliver Schmidt – Universität Bremen

Wir danken allen Mitwirkenden, insbesondere

- Prof. Dr. Antje Boetius – Universität Bremen und Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)
- Prof. Dr. Andreas Breiter – Universität Bremen und Institut für Informationsmanagement Bremen GmbH
- Dr.-Ing. Florian Cordes – Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Bremen
- Dr. Nicolas Dittert – Leibniz-Zentrum für Marine Tropenforschung (ZMT)
- Prof. Dr.-Ing. Rainer Fechte-Heinen – Universität Bremen und Leibniz-Institut für Werkstofforientierte Technologien – IWT
- Prof. Dr. Jutta Günther – Universität Bremen
- Prof. Dr.-Ing. Horst Hahn – Jacobs University Bremen und Fraunhofer-Institut für Digitale Medizin MEVIS
- Dr. Jan-Ocko Heuer – Universität Bremen
- Prof. Dr. Betina Hollstein – Universität Bremen
- Dr. Elisabeth Huber – Universität Bremen
- Prof. Dr. Frank Kirchner – Universität Bremen und Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Bremen
- Prof. Dr. Christoph Lüth – Universität Bremen und Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Bremen
- Prof. Dr.-Ing. Lutz Mädler – Universität Bremen und Leibniz-Institut für Werkstofforientierte Technologien – IWT
- Dr. Guido Prause – Fraunhofer-Institut für Digitale Medizin MEVIS
- Dr.-Ing. Norbert Riefler – Leibniz-Institut für Werkstofforientierte Technologien – IWT
- Prof. Dr. Norman Sieroka – Universität Bremen

1. Inhaltsverzeichnis

0.	Kurzfassung	6
1.	Präambel	7
2.	Forschungsdatenmanagement: Schlüsselkomponente für eine kollaborative, datengetriebene Wissenschaft	8
3.	Status quo Forschungsdatenmanagement in der U Bremen Research Alliance	9
3.1.	Data Train – Training in Research Data Management and Data Science	10
3.2.	Infrastrukturen für Forschungsdaten in der U Bremen Research Alliance	11
3.3.	Herausforderungen im Forschungsdatenmanagement	12
4.	Notwendige Handlungsfelder	14
4.1.	Aufbau einer föderalen technischen Infrastruktur für Forschungsdatenmanagement in der U Bremen Research Alliance	14
4.2.	Aufbau des Kompetenznetzwerks für Forschungsdatenmanagement	15
4.3.	Anerkennung und Bewusstsein für Forschungsdatenmanagement vermitteln und steigern	16
5.	Mehrwert für den Wissenschaftsstandort Bremen	18
A.	Referenzen	20
B.	Websites	22
B.1.	Mitglieder	22
B.2.	Weitere	23
C.	Glossar	24

Anhang A: Kurzprofile aus der U Bremen Research Alliance	25
Kurzprofil 1: NFDI-Konsortien der ersten Förderphase mit Bremer Beteiligung	25
Kurzprofil 2: Forschungsdatenmanagement in der Meeresforschung	25
Kurzprofil 3: Forschungsdatenmanagement in der Sozial- und Wirtschaftsforschung ...	26
Kurzprofil 4: Forschungsdatenmanagement in den Gesundheitswissenschaften	27
Kurzprofil 5: Forschungsdatenmanagement in Materialforschung und Technik	27
Kurzprofil 6: Data Science an der Universität Bremen	28
Anhang B: Use Cases aus der U Bremen Research Alliance	29
Use Case 1: Das „Observation to Archive and Analysis (O2A)“-Framework für das FAIRe Forschungsdatenmanagement am AWI	29
Use Case 2: Forschungsdatenmanagement und Data Science am Leibniz-BIPS	31
Use Case 3: Langzeitarchivierung und Nachnutzung von qualitativen sozialwissen- schaftlichen Forschungsdaten im Forschungsdatenzentrum Qualiservice	32
Use Case 4: Robotik und KI am DFKI in Bremen	33
Use Case 5: Die Leitlinien zum Umgang mit Forschungsdaten der Universität Bremen	35
Use Case 6: Implementierung eines Forschungsdatenmanagement-Systems am Leibniz-IWT	36
Use Case 7: Zusammenführung von Sensor- und Messsystemen auf Hardware- und Datenebene am Fraunhofer IFAM	38
Herausgeber	39

0. Kurzfassung

Die U Bremen Research Alliance ist Kernbereich der Bremer Wissenschaft. Sie steht für Kooperation auf den Ebenen Forschung, wissenschaftlicher Nachwuchs, Rekrutierung und Infrastruktur. Um kooperative zukunftsorientierte Wissenschaft zu stärken, verfolgt sie gemeinschaftlich die Schaffung eines abgestimmten Raumes für Innovation, Infrastruktur und Forschung. Die Bereiche Forschungsdatenmanagement, Data Science und Digitalisierung bilden einen Schwerpunkt der Arbeit der Mitglieder der U Bremen Research Alliance. Beispiele dieser Entwicklung sind der Bremer Leibniz-WissenschaftsCampus „Digital Public Health“, die Gründung des Data Science Centers (DSC) an der Universität Bremen, der Aufbau der interdisziplinär und interorganisational getragenen Doktorand*innenausbildung „Data Train – Training in Research Data Management and Data Science“ und die aktive Beteiligung der Mitglieder an vier der neun im Juni 2020 ausgewählten Vorhaben zum Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI). Bremen ist damit im bundesweiten Vergleich einer der erfolgreichsten NFDI-Standorte. Dieses Whitepaper skizziert den aktuellen Status sowie Handlungsfelder für die disziplinübergreifende **Etablierung eines kooperativen Forschungsdatenmanagements in der U Bremen Research Alliance.**

1. Präambel

Der Wissenschaftsstandort Bremen steht für die erfolgreiche Kooperation einer Forschungsuniversität mit einer Vielzahl von außeruniversitären Forschungseinrichtungen. Zur Intensivierung und zur strategischen Stärkung dieses Kooperationsnetzwerks arbeiten die Universität Bremen und die zwölf von Bund und Ländern finanzierten außeruniversitären Forschungsinstitute aller deutschen Wissenschaftsorganisationen im Bundesland Bremen in der U Bremen Research Alliance zusammen. Damit steht die Allianz als fraktales Abbild der deutschen Forschungslandschaft exemplarisch für einen Kooperationsverbund zwischen Universitäten und außeruniversitären Forschungseinrichtungen. Das gelingt der U Bremen Research Alliance durch eine multilaterale strategische Ausrichtung und Stärkung der Forschungsk Kooperation. Kooperative Nachwuchsförderung, gemeinsame Personalentwicklungsprogramme, übergreifend genutzte Forschungsinfrastrukturen und abgestimmte Forschungsagenden sorgen für Skaleneffekte und tragen zu mehr internationaler Sichtbarkeit bei. Der Wissenschaftsstandort Bremen will beste Bedingungen bieten, um frei, innovativ, kreativ, nachhaltig und kooperativ zu forschen.

Aktuell manifestiert sich der digitale Wandel überall in der Gesellschaft und Wissenschaft. Die Bedeutung und Notwendigkeit einer „digitalen Transformation“ wurde durch die Corona-Pandemie eindrücklich belegt. Durch die Zunahme an „Big Data“-Anwendungen, das „Internet of Things“ und durch immer neue innovative Technologien eröffnen sich bisher unbeschränkte Wege. Dabei fallen immense und in vielerlei Hinsicht heterogene Datenmengen an, die die Grundlage für die interdisziplinäre Schlüsseldisziplin „Data Science“ bilden. Durch die Anwendung zukunftsweisender Data Science-Verfahren wie maschinelles Lernen können auch komplexe und hochdimensionale Datenbestände wertschöpfend genutzt und so ein überproportionaler Anstieg des Erkenntnisgewinns erzeugt werden. Ein Fundament dafür bilden ein abgestimmtes Forschungsdatenmanagement sowie eine „Open Data-Kultur“ (Nosek et al., 2015). Begründet durch die gute wissenschaftliche Praxis (DFG, 2019) und im Einklang mit dem Aufbau der Nationalen Forschungsdateninfrastruktur (NFDI) gilt es, „Datenbestände von Wissenschaft und Forschung systematisch zu erschließen, nachhaltig zu sichern und zugänglich zu machen sowie (inter-)national zu vernetzen“ (vgl. DFG und NFDI).^{1,2}

→ Der Wissenschaftsstandort Bremen will **beste Bedingungen bieten**, um frei, innovativ, kreativ, nachhaltig und kooperativ zu forschen.

¹ www.dfg.de/foerderung/programme/nfdi

² www.nfdi.de

Dieses als „FAIR Data“ bezeichnete Entwicklungsziel (Wilkinson et al., 2016) verfolgt die U Bremen Research Alliance durch die Schaffung eines kooperativen Forschungsdatenmanagements. Dieses Whitepaper skizziert den Status, die aktuellen Herausforderungen und zeigt Handlungsfelder auf, die die Wissenschaftslandschaft Bremen national und international vernetzter und wettbewerbsfähiger machen.

2. Forschungsdatenmanagement: Schlüsselkomponente für eine kollaborative, datengetriebene Wissenschaft

→ Forschungsdatenmanagement ist die Voraussetzung für eine deutliche Steigerung des Erkenntnisgewinns in der Wissenschaft.

Intelligentes Forschungsdatenmanagement ist die Voraussetzung für die disziplinübergreifende Nutzung von Forschungsdaten und daraus folgend eine deutliche Steigerung des Erkenntnisgewinns. Dies entspricht den aktuellen Forderungen der führenden Wissenschaftsorganisationen: des Wissenschaftsrats (WR, 2011 a-c), der Allianz der Wissenschaftsorganisationen (AdW, 2018) und des Rats für Informationsinfrastrukturen (RfII, 2016). Grundvoraussetzungen hierfür sind a) die detaillierte Dokumentation der Daten selbst sowie aller Vorgänge ihrer Erhebung (Provenance), b) die Sicherung der Datenqualität und c) eine Datenablage gemäß den „FAIR-Prinzipien“: Forschungsdaten müssen für Dritte jederzeit auffindbar (findable), zugänglich (accessible), interoperabel (interoperable) und wiederverwendbar (reusable) sein (Wilkinson et al., 2016). Hierbei gilt es, neben technischen Lösungen, insbesondere bei personenbezogenen Daten, ethische und juristische Aspekte wie Anonymisierung und die informierte Einwilligung zum Data Sharing zu beachten, die in der wissenschaftlichen Gemeinschaft zurzeit umfassend debattiert werden. In dem Zusammenhang ist ein kritisches Bewusstsein („Critical Thinking“) sehr wichtig, um angemessen und verantwortlich über die Aufbereitung, Weitergabe und Verwendung von Daten entscheiden zu können.³ Dadurch wird über Disziplinen hinweg eine gemeinsame Sprache etabliert, die sich möglicher Grenzen und Schwierigkeiten bewusst ist. Diese Sprache soll zu mehr wohlinformierter Akzeptanz und Einsatzbereitschaft gegenüber neuen Herangehensweisen und Technologien führen (vgl. Sieroka et al., 2018. Schneider et al., 2020). Ein dahingehender Kulturwandel wird angemahnt und ist dringend notwendig.

³ „Critical Thinking“: Gemeint ist die kritische Reflexion der eigenen Arbeit, aber auch das Entwickeln von Empathie gegenüber anderen Denkweisen oder -mustern (beispielsweise anderer Disziplinen).

3. Status quo Forschungsdatenmanagement in der U Bremen Research Alliance

Zur nachhaltigen Schaffung eines kooperativen Forschungsdatenmanagements, das als „Blaupause“ für andere Einrichtungen und Organisationen dienen mag, soll die Zusammenarbeit zwischen den Einrichtungen der U Bremen Research Alliance weiter gestärkt werden. In der Allianz können wir im Forschungsdatenmanagement auf einer hervorragenden Entwicklung der letzten Jahre aufbauen. So hat die U Bremen Research Alliance mit Unterstützung des Landes Bremen die Entwicklung der **Doktorand*innenausbildung „Data Train – Training in Research Data Management and Data Science“** (siehe Kapitel 3.1) begonnen.⁴ Über die Mitglieder der Allianz werden darüber hinaus viele institutionelle Aktivitäten zusammengeführt.⁵ Für eine Darstellung der Aktivitäten in den verschiedenen Wissenschaftsbereichen sei auf Anhang A verwiesen. Herauszuheben ist die Beteiligung von Einrichtungen der Allianz in zahlreichen **Konsortien der Nationalen Forschungsdateninfrastruktur**: NFDI4BioDiversity (Sprecher: Prof. Dr. Frank Oliver Glöckner, Universität Bremen; Teilnehmer: Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI), Leibniz Zentrum für Marine Tropenforschung (ZMT)), NFDI4Health (Co-Sprecherin: Prof. Dr. Iris Pigeot, Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS; Mitantagsteller: Universität Bremen, Fraunhofer Institut für digitale Medizin MEVIS), KonsortSWD (Co-Sprecherin: Prof. Dr. Betina Hollstein, Universität Bremen) und NFDI4Ing (Teilnehmer: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)). **Bremen ist damit im bundesweiten Vergleich einer der erfolgreichsten NFDI-Standorte.** Auch an der Antragstellung für NFDI-Konsortien in der zweiten Antragsrunde waren mehrere Mitgliedseinrichtungen der Allianz beteiligt. Eine Entscheidung wird im Sommer 2021 erwartet.

Die Universität Bremen verfügt über eine zentrale Beratungsstelle zum Forschungsdatenmanagement (Referat Forschung und wissenschaftlicher Nachwuchs), die beispielsweise Beratung beim Erstellen eines Datenmanagementplans für Forschungsanträge anbietet. Im Jahr 2018 sind an der Universität Bremen zudem mit dem Lenkungskreis „Forschungsdatenmanagement“ ein zentrales Gremium geschaffen und in einem partizipativen Prozess Empfehlungen zum Forschungsdatenmanagement erarbeitet worden. Diese Empfehlungen und die bei der Erstellung gesammelten

⁴ www.uni-bremen.de/data-train

⁵ www.uni-bremen.de/research-alliance/forschungsdaten

→ In der Allianz können wir im Forschungsdatenmanagement auf einer hervorragenden Entwicklung der letzten Jahre aufbauen.

→ Bremen ist damit im bundesweiten Vergleich einer der erfolgreichsten NFDI-Standorte.

Erfahrungen können Grundlage sein, um ein praktikables und effizientes Modell für kooperatives Forschungsdatenmanagement in der Allianz zu entwickeln.

Darüber hinaus wurde an der Universität Bremen das interdisziplinäre **Data Science Center (DSC)** eingerichtet, das die Anwendung von innovativen Datenanalyseverfahren insbesondere der Künstlichen Intelligenz in allen Disziplinen vorantreibt, um einen maximalen Erkenntnisgewinn aus Forschungsdaten zu ermöglichen (siehe Anhang A).

Die Mitglieder der U Bremen Research Alliance beweisen somit bereits ihre Fähigkeiten zur Erarbeitung von Leitlinien für ein intelligentes Forschungsdatenmanagement sowie ihre Verbundstärke. Damit untermauern sie optimale Voraussetzungen für die kooperative Umsetzung von institutionenübergreifendem Forschungsdatenmanagement.

3.1. Data Train – Training in Research Data Management and Data Science

→ Bis 2023 werden rund 455.000 Expert*innen für komplexe Datenanalysen zusätzlich benötigt.

Bereits 2018 stellten der Stifterverband für die Deutsche Wissenschaft und McKinsey im Diskussionspapier „Future Skills“ heraus, dass der Bedarf an Technologie-Spezialist*innen enorm hoch ist. Bis 2023 würden rund 700.000 Personen mehr als heute benötigt – hiervon alleine 455.000 für komplexe Datenanalysen (Kirchherr et al., 2018). Zudem veröffentlichte der Rat für Informationsinfrastrukturen im Jahr 2019 eine Untersuchung, die besagt, dass die Qualität und damit auch die Qualitätskontrolle wissenschaftlicher Daten verbessert werden muss (RfII, 2019). Als Zeichen des Aufbruchs hat der Stifterverband für die Deutsche Wissenschaft im Jahr 2021 im Einklang mit der Datenstrategie der Bundesregierung und der Berliner Erklärung zur Digitalen Gesellschaft eine Data-Literacy-Charta publiziert. Sie stellt fest: Die Datenkompetenzen sind unverzichtbarer Bestandteil der Allgemeinbildung. Auf diesen Bedarf – in Wissenschaft, Wirtschaft und Gesellschaft – hat die U Bremen Research Alliance mit Unterstützung des Landes Bremen bereits 2019 reagiert. Um Kompetenzen in Data Literacy, Datenmanagement und Data Science zu stärken, wird für Doktorand*innen der U Bremen Research Alliance derzeit eine interdisziplinäre Ausbildung „Data Train – Training in Research Data Management and Data Science“ aufgebaut.



Basierend auf internen Befragungen im Rahmen einer Analyse des „Status quo Forschungsdatenmanagement und Data Science der U Bremen Research Alliance“ und einer ausführlichen externen Recherche ist ein disziplinübergreifendes Curriculum entwickelt worden. In diesem Curriculum wird Expertise aus der gesamten Allianz gebündelt und gemeinschaftlich weitergegeben. Es beteiligen sich zudem allianzexterne Wissenschaftseinrichtungen. Auch Wirtschaftsunternehmen aus der Region, beispielsweise Airbus Deutschland, haben die Notwendigkeit des Aufbaus von (Personal-)Ressourcen erkannt und sind vertreten.

3.2. Infrastrukturen für Forschungsdaten in der U Bremen Research Alliance

Eine Kerninfrastruktur des Landes Bremen bildet das gemeinschaftlich vom **MARUM – Zentrum für Marine Umweltwissenschaften der Universität Bremen** und dem **Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)** betriebene Informationssystem **PANGAEA**. Als international anerkanntes und zertifiziertes Repositorium für Daten aus der Erd- und Umweltforschung ist es nicht nur eine wichtige Anlaufstelle für viele Forschende aus der ganzen Welt, sondern auch ein zentrales Archiv für Forschungsdaten in der U Bremen Research Alliance und für die **Deutsche Allianz Meeresforschung (DAM)**. Jeder publizierte Datensatz erhält ein vollständiges Zitat einschließlich eines persistenten Identifikators (DOI), um ihn weltweit eindeutig auffindbar und referenzierbar zu machen.

Das **MARUM GeoB Core Repository** und das **Bremen Core Repository (BCR)** vergeben global eindeutige, persistente Identifikatoren für das langfristig archivierte und kuratierte Probenmaterial geowissenschaftlicher Expeditionen. Als Mitglied und sogenannter „Allocating Agent“ des internationalen Vereins IGSN e.V. fördert das MARUM damit die Zitierbarkeit, Identifizierbarkeit und Auffindbarkeit von physikalischen, wissenschaftlichen Proben.

Der in Bremen angesiedelte gemeinnützige Verein **GFBio – Gesellschaft für Biologische Daten e. V.** ist deutschlandweit aktiv mit dem Ziel, Wissenschaft und Forschung im Bereich des Datenmanagements speziell in der Biodiversitätsforschung, der Ökologie und

der Umweltforschung zu fördern. Zu seinem Portfolio gehören die Erfassung und der Zugang zu Daten über ein einheitliches Portal sowie Information und Training von Forschenden zum Datenmanagement in Projekten und die Erstellung von individuellen Datenmanagementplänen für Projekte und Anträge. Der GFBio bildet zudem die organisatorische Kernstruktur der NFDI4BioDiversity.

Das vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierte Forschungsdatenzentrum **Qualiservice** an der Universität Bremen archiviert qualitative sozialwissenschaftliche Forschungsdaten und stellt sie für die wissenschaftliche Nachnutzung zur Verfügung. Mit dem Fokus auf sensiblen personenbezogenen Daten deckt Qualiservice das gesamte Spektrum qualitativer sozialwissenschaftlicher Daten ab (Interviewtranskripte, Audio-, Videodaten etc.). Zusammen mit der Bibliothek der Humboldt-Universität zu Berlin verantwortet Qualiservice den Fachinformationsdienst Sozial- und Kulturanthropologie (FID SKA). Neben der Datenkuration bietet Qualiservice ein umfangreiches Informations- und Beratungsangebot. Durch die Vergabe von persistenten Identifikatoren (DOI) werden Datensätze international such- und auffindbar. Qualiservice nutzt das Metadatenschema der Data Documentation Initiative (DDI) und die innovative PANGAEA-Technologie.

Die vom renommierten Institute for Artificial Intelligence (IAI) betriebene webbasierte Wissensdatenbank **openEASE** stellt semantisch annotierte Aktivitätsdaten aus Robotik-Experimenten zur Verfügung mit dem Ziel, die Interoperabilität in der Robotik zu verbessern und Barrieren für die Roboterprogrammierung zu verringern. OpenEASE kann weltweit von Forscher*innen, aber auch direkt von autonomen Robotern genutzt werden. So können Roboter unmittelbar voneinander lernen und Wissenschaftler*innen einzelne Abläufe und Entscheidungen von Maschinen untersuchen.

Als Unterstützung des Forschungsdatenmanagements bietet das **Green IT Housing Center**, betrieben von der Universität Bremen, eine zentrale Plattform für (Arbeits-)Speichermöglichkeiten, während die Administration der Server dezentral in den Fachbereichen teils auf Arbeitsebene vorgenommen wird.

3.3. Herausforderungen im Forschungsdatenmanagement

Im Rahmen einer Befragung zum Status quo innerhalb der U Bremen Research Alliance sind Handlungsbedarfe ersichtlich geworden: Bislang werden häufig fachspezifische Teillösungen auf Projekt- oder

Arbeitsgruppenebene realisiert. Diese decken den Datenlebenszyklus (vgl. Abb. 1) oftmals lediglich partiell ab. Ein qualitätsgesicherter Umgang mit Daten sowie eine – für die Nachnutzung essenziell notwendige – standardisierte Erfassung von Metadaten und die Kontextualisierung der Forschungsdaten erfordern jedoch ein kontinuierliches sowie ganzheitliches Daten- und Probenmanagement über das Ende des jeweiligen Forschungsvorhabens hinaus. Bei der derzeit zunehmenden Menge an erzeugten Daten und der folglich wachsenden Datenanreicherung steht zudem häufig die angestrebte Langlebigkeit der Daten den finanziellen, infrastrukturellen und personellen Anforderungen und Möglichkeiten entgegen. Durch die projektbezogene Förderung ist es in der Regel nicht möglich, Forschungsdateninfrastrukturen nachhaltig aufzubauen und zu pflegen oder Personal dauerhaft zu beschäftigen. Insbesondere Vorhaben, bei denen Partner aus Wissenschaft und Wirtschaft (international) zusammenarbeiten, verlangen juristische Expertise, um die rechtlichen Rahmenbedingungen adäquat zu gestalten – sei es beim Umgang mit sensiblen, personenbezogenen und schützenswerten Daten (vgl. RatSWD, 2020), Lizenzen oder beim Einsatz biologischer Objekte im Forschungskontext (vgl. Buck und Hamilton, 2011).

→ Der Datenlebenszyklus (vgl. Abb. 1) wird oftmals lediglich partiell abgedeckt.

→ Durch projektbezogene Förderungen ist es in der Regel nicht möglich, Forschungsdateninfrastrukturen nachhaltig aufzubauen.



Abbildung 1: Datenlebenszyklus, modifiziert aus <http://www.gfbio.org/training/materials/data-lifecycle> und nach Stodden (2020). Blaue Kästchen zeigen die einzelnen Schritte des Datenlebenszyklus, die außen angrenzenden Kreise zeigen die notwendige technische Infrastruktur.

4. Notwendige Handlungsfelder

In Abbildung 1 ist der ideale Datenlebenszyklus unter Einbeziehung der infrastrukturellen beziehungsweise technischen Voraussetzungen (siehe blaue Kreise am äußeren Rand) dargestellt. Diesen gilt es in der U Bremen Research Alliance zu etablieren. Die Verbesserung der Situation im Einklang mit der Umsetzung des Datenlebenszyklus nach den FAIR-Prinzipien verlangt umfangreiche Organisationsentwicklungsprozesse in den Wissenschaftseinrichtungen. Diese reichen von der Schaffung eines kulturellen Wandels im Sinne einer Open Data-Kultur und der Bewusstseins-schärfung bereits beim wissenschaftlichen Nachwuchs über die Schaffung des Zugangs zu essenzieller, benutzerfreundlicher Ausstattung (Hard- / Software wie Plattformen / Cloud-Dienste) bis hin zu Wertschätzungsmechanismen wie Anerkennung der erbrachten Datenleistung ähnlich eines wissenschaftlichen Artikels.

Dezentrale (partielle) Lösungen, wie sie momentan an der Universität Bremen und an vielen außeruniversitären Forschungsinstituten in der U Bremen Research Alliance vorliegen, müssen zusammengeführt und zentral ausgebaut werden.

Im Einzelnen sind folgende Maßnahmen notwendig:

4.1. Aufbau einer föderalen technischen Infrastruktur für Forschungsdatenmanagement in der U Bremen Research Alliance

Gemeinsames föderiertes Identitäts- und Rechtemanagement zwischen den außeruniversitären Instituten und der Universität Bremen

Es ist notwendig, einen zentralisierten Zugriff auf alle Cloud-Dienste und Online Tools in der Allianz, also ein allianzweites Single Sign On, zu etablieren – ähnlich wie das Education Roaming (eduroam), das einen Internetzugang auf dem gesamten Campus mit einer einmaligen Anmeldung ermöglicht. Die Anbindung an das NFDI und EOSC single-sign-on System (AAI) muss von Anfang an mitgedacht werden.

Bereitstellung von umfassenden kollaborativen Online Tools und innovativen Technologien als essenzielle Bausteine gemeinsamer Infrastruktur

Online Tools und innovative Technologien wie Cloud-Dienste ermöglichen konnektives und damit kooperatives Arbeiten und bieten flexible Rahmenbedingungen. Sie sind wesentlicher Bestandteil des digitalen Zeitalters. Entsprechende Informations- und

Kommunikationstechnologien müssen allianzweit zur Verfügung gestellt werden. Diese sollen gleichzeitig als Baustein der NFDI-übergreifenden Research Data Commons- und EOSC-Aktivitäten dienen.

Persistente Referenzierbarkeit von Informationen

Persistente Identifikatoren für Informationen, die global eindeutig sind und zugleich Interoperabilität (durch Metadaten-Verknüpfung) ermöglichen, müssen technisch etabliert werden. Dies verlangt Fachpersonal, das die Identifikatoren in die jeweiligen Systeme einführt und in bestehende Workflows langfristig integriert, sowie national wie international mit der wissenschaftlichen und technischen Community im Austausch steht. Mit PANGAEA wird dies durch die Einführung von persistenten Identifikatoren (DOI) für jeden publizierten Datensatz bereits seit mehr als zwei Jahrzehnten praktiziert.

Langzeitarchivierung

Für die Sicherstellung der Langzeitarchivierung müssen Repositorien wie PANGAEA und Qualiservice einen kontinuierlichen und langfristigen Betrieb gewährleisten und dementsprechend zukunftsorientiert (hinsichtlich Kapazität, Technologie, Referenzierbarkeit, Kuratierung und Personal) ausgebaut und gefördert werden.

4.2. Aufbau des Kompetenznetzwerks für Forschungsdatenmanagement

Zentrale Unterstützung zur Entwicklung und Umsetzung von individuellen Datenmanagementplänen

Die Verarbeitung von heterogenen Daten benötigt möglichst flexible und komplexe Lösungen in Bezug auf das Datenmanagement. Maßgeschneiderte Datenmanagementpläne sind deshalb für ein erfolgreiches, effizientes und nachhaltiges Forschen entscheidend. Der gemeinsame Aufbau eines Datenmanagement-Tools, basierend auf den Erfahrungen des GFBio, würde einen klaren Mehrwert für Bremen schaffen.⁶

Aufbau eines „Data Stewardship Netzwerks“ zur Unterstützung des fachspezifischen Umgangs mit Forschungsdaten

Für einen nachhaltigen Umgang mit Forschungsdaten ist eine umfassende Unterstützung der Universität Bremen und der außeruniversitären Institute der U Bremen Research Alliance durch Data Stewards erforderlich. Diese sollen Forschenden und im Rahmen des Transfers Wirtschaftsunternehmen bei der Auswahl geeigneter Repositorien und Standards bis hin zu Lizenz- und Rechtsfragen Auskunft geben können. Die Data Stewards wirken dabei als

→ Für einen nachhaltigen Umgang mit Forschungsdaten ist eine umfassende Unterstützung der Universität Bremen und der außeruniversitären Institute der U Bremen Research Alliance durch Data Stewards erforderlich.

⁶ www.gfbio.org/plan

→ Die Umsetzung der FAIR-Prinzipien verlangt umfangreiche Organisationsentwicklungsprozesse und einen kulturellen Wandel hin zu einer Open Data-Kultur.

„Front Office“ nah an der Forschung und der Praxis, während die NFDI-Infrastruktur im „Back Office“ die Daten entgegennimmt und nach den FAIR-Prinzipien aufbereitet und zugänglich macht.

Zentrale Unterstützung bei Fragen zur Lizenzierung und Publikationsstrategie

Bei schützenswerten Daten und Forschungssoftware müssen neben einer detaillierten technischen Begutachtung und Qualitätsprüfung die Rahmenbedingungen der Forschungsarbeit in Betracht gezogen werden, bevor Entscheidungen bezüglich der Lizenzierung getroffen und Strategien zur Bereitstellung der Daten und Software entworfen werden können. Zentrale Services (Fachpersonal und/oder Auftragsvergabe) entwerfen generische Konzepte zu Lizenzmodellen, auf die fallbezogen zurückgegriffen werden kann, um sie projektspezifisch anzupassen.

Zentrale Rechts- und Ethikberatung zum Thema Datenmanagement für Bremen

Die interdisziplinäre Nutzung heterogener Daten tangiert rechtliche und ethische Aspekte, die zwingend erkannt und eingehalten werden müssen. Mit den Rechtswissenschaften und der Ethikkommission der Universität Bremen sollen gemeinsame Strategien zu den rechtlichen und ethischen Aspekten des Forschungsdatenmanagements erarbeitet und pragmatisch implementiert werden.

Ausarbeitung von einheitlichen Forschungsdatenleitlinien

Eine Einigung aller Mitglieder der U Bremen Research Alliance auf einheitliche Leitlinien zum nachhaltigen Umgang mit wissenschaftlichen Daten wird angestrebt. Hierbei kann auf bestehende Ansätze, Ausarbeitungen und Erfahrungen in den Wissenschaftseinrichtungen, die eigenständig oder über die deutschen Wissenschaftsorganisationen generiert worden sind, zurückgegriffen werden. Diese Forschungsdatenleitlinien sollen daher in einer gemeinsamen Arbeitsgruppe der U Bremen Research Alliance erstellt werden.

4.3. Anerkennung und Bewusstsein für Forschungsdatenmanagement vermitteln und steigern

Um den angemahnten Kulturwandel in der Wissenschaft zu schaffen, muss für Themen des Forschungsdatenmanagements sensibilisiert beziehungsweise ein Bewusstsein für die Thematik geschaffen werden. Dies gelingt zum Beispiel durch die Etablierung einer Ausbildung für alle Statusgruppen und durch die Erweiterung der begonnenen Ausbildung Data Train (siehe Kapitel 3.1) auf neue Zielgruppen.

Die erfolgreiche Umsetzung der FAIR-Prinzipien wird aber auch von der Anerkennung aller getragen. Zur Steigerung der Anerkennung sollte der FAIRe Umgang mit Daten wie die Publikation eines wissenschaftlichen Artikels gewichtet werden. Dafür ist die Qualitätssicherung und Referenzierbarkeit der kuratierten und publizierten Datensätze essenziell.

Darüber hinaus ist es notwendig, über geschaffene Leitlinien und Standards sowie Angebote – sei es im Bereich der technischen Infrastruktur, Beratung oder Dienstleistung – zu informieren und an sie heranzuführen. Damit wird von dem Bewusstsein und der Anerkennung die Brücke zur Etablierung der Open Data-Kultur geschlagen.

Qualifizierung aller Statusgruppen

Angebote zur Qualifizierung aller Statusgruppen und damit zur breiten Generierung von grundlegenden Schlüsselkompetenzen in Data Literacy und Forschungsdatenmanagement müssen zentral vorgehalten werden. Das spart Ressourcen und entfacht zugleich interdisziplinären Austausch und Vernetzung. Beides ist für das übergeordnete Ziel eines abgestimmten und kooperativen Forschungsdatenraumes essenziell. Darauf aufbauende spezifische Herangehensweisen müssen in den Fachdisziplinen vermittelt werden. Geplant ist, das Konzept, das bereits für die Doktorand*innenausbildung Data Train der U Bremen Research Alliance umgesetzt wird, um weitere Zielgruppen ((wiss.) Fachpersonal, Wissenschaftsmanagement, Masterstudierende) zu erweitern. Damit würde die Komplexität des Programms gesteigert werden, sodass dies bei der Koordination und Organisation des Programms berücksichtigt werden muss. Ein Ausbau der technischen Infrastruktur (d. h. eine professionelle Lernplattform inkl. Evaluationstools und (digitales) Equipment für die Lehre) ist daher notwendig.

Transferarbeit

Die Fähigkeit, die Aufbereitung, Weitergabe und Verwendung von Daten kritisch zu reflektieren, ist als Kernkompetenz in Ausbildungsangeboten im Wissenschaftsbereich bereits verankert – so beispielsweise in der Doktorand*innenausbildung Data Train oder auch allgemeiner in den interdisziplinären Ausbildungskonzepten der Arbeitsgruppe „Theoretische Philosophie“ der Universität Bremen. Dies muss weiter ausgebaut werden. Kritisches Reflektieren sollte der Öffentlichkeit auch in Form von Veranstaltungs- und Weiterbildungsprogrammen vermittelt werden. An dieser Stelle kann die Wissenschaft einen wichtigen Beitrag für

→ **Bewusstsein und Anerkennung schlagen die Brücke zur Etablierung einer Open Data-Kultur.**

die Gesellschaft leisten und mit ihr in einen fruchtbaren Austausch treten. Denn Fragen zum Umgang mit Daten stellen sich an vielen Stellen des Alltags und beschäftigen unsere Gesellschaft in vielerlei Hinsicht, sei es in Bezug auf die Einhaltung der Privatsphäre beim Umgang mit Daten oder auf mögliche Einschränkungen der eigenen Handlungsspielräume.⁷

Anerkennung von Ausbildungsleistung

Ziel von Data Train – Training in Research Data Management and Data Science ist es, instituts- und fachbereichsübergreifend Expertisen in einem gemeinsamen, abgestimmten Curriculum zusammenzutragen und Kompetenzen weiterzugeben. Das Einbringen von Expertise, Kapazität und Lehrleistung seitens der Dozent*innen in das gemeinschaftliche Vorhaben muss wertgeschätzt und honoriert werden. Nur so kann das Ausbildungsmodell nachhaltig funktionieren.

5. Mehrwert für den Wissenschaftsstandort Bremen

Im Falle der erfolgreichen Realisierung eines kooperativen Forschungsdatenmanagements in der U Bremen Research Alliance sind die Mehrwerte überproportional und entfalten Wirkungen weit über die Landesgrenzen hinaus. Wie aufgezeigt können die U Bremen Research Alliance und der Standort Bremen mit komplementären Investitionen zu einem interdisziplinären Kompetenzzentrum in diesem Bereich ausgebaut werden und so eine Vorreiterrolle einnehmen, mit engen Verbindungen zu regionalen Wirtschaftsunternehmen. Es ist daher dringend angezeigt, die Umsetzung der FAIR-Prinzipien in Bremen voranzubringen und die dafür notwendigen Strukturen zu etablieren. Es bestehen bereits weitreichende Erfahrungen in der Kooperation zwischen der Universität Bremen und den von Bund und Ländern finanzierten Einrichtungen, die Skaleneffekte bei der Etablierung von Servicestrukturen realisierbar machen. Zentral gesteuerte, standardisierte Arbeitsabläufe und Zugänge bei gleichzeitiger Berücksichtigung fachspezifischer methodologischer Gütekriterien erhöhen die Qualität der Forschung. Darüber hinaus wirken sie sich positiv und unterstützend für die Forschenden aus und haben somit das große Potenzial, Effizienz und Qualität der Forschung zu steigern. Konkrete Beispiele zum gelebten Forschungsdatenmanagement, wie sie in der Allianz bereits vorhanden sind, sind

→ Die U Bremen Research Alliance und der Wissenschaftsstandort Bremen können im engen Verbund mit der Wirtschaft eine Vorreiterrolle einnehmen.

dabei der Schlüssel zum Mehrwert und zum Kulturwandel (siehe Use Cases aus der U Bremen Research Alliance in Anhang B).

Die effektive, kooperative sowie internationale (Nach-)Nutzung von Forschungsdaten wird die wissenschaftliche Leistung und die Sichtbarkeit der Mitgliedseinrichtungen der Allianz und damit Bremens als Wissenschaftsstandort erhöhen. Dies schafft nicht nur eine „bessere Wissenschaft“ in der Allianz, sondern macht Bremen zu einem (inter-)national beachteten Protagonisten. Dringend benötigte Effekte bei der Rekrutierung von Fachpersonal sowie der Anwerbung von exzellenten Wissenschaftler*innen und technischem Fachpersonal sind zu erwarten. Eine stärkere interdisziplinäre und datenorientierte Ausbildung der Absolvent*innen in Bremen würde den (regionalen) Arbeitsmarkt in den Bereichen Forschungsdatenmanagement und Data Science mit hoch qualifiziertem Personal versorgen.

→ Die effektive, kooperative sowie internationale (Nach-)Nutzung von Forschungsdaten wird die wissenschaftliche Leistung und die Sichtbarkeit der Mitgliedseinrichtungen der Allianz und damit Bremens als Wissenschaftsstandort erhöhen.

⁷ Siehe blogs.ethz.ch/RETHINK/2020/09/28/philosophy-of-ai-and-the-role-of-digital-design/

Allianz deutscher Wissenschaftsorganisationen (2018): Nationale Forschungsdateninfrastruktur (NFDI), Diskussionspapier zur Ausarbeitung einer Bund-Länder-Vereinbarung zum Aufbau einer Nationalen Forschungsdateninfrastruktur. <https://www.allianzinitiative.de/2018/08/08/diskussionspapier-zum-aufbau-einer-nationalen-forschungsdateninfrastruktur-nfdi/> (abgerufen am 20.04.2021)

Buck M, Hamilton C (2011): The nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. *Review of european community & international environmental law* 20: 47–61. doi:10.1111/j.1467-9388.2011.00703.x

Bundesregierung (2021): Datenstrategie der Bundesregierung. Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum. Kabinettdfassung, Bundeskanzleramt, Berlin. <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632> (abgerufen am 20.04.2021)

Deutsche EU-Ratspräsidentschaft (2020): Berliner Erklärung zur Digitalen Gesellschaft und wertebasierten digitalen Verwaltung. Berlin. <https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2020/eu-rp/gemeinsame-erklarungen/berliner-erklarung-digitale-gesellschaft.html> (abgerufen am 20.04.2021)

Deutsche Forschungsgemeinschaft (2019): Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. doi:10.5281/zenodo.3923602

Kirchherr J et al. (2018): Future Skills: Welche Kompetenzen in Deutschland fehlen. www.stifterverband.org/medien/future-skills-welche-kompetenzen-in-deutschland-fehlen (abgerufen am 20.04.2021)

Nosek BA et al. (2015): Promoting an open research culture. *Science* 348: 1422-1425. doi: 10.1126/science.aab2374

Rat für Informationsinfrastrukturen (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen. <http://www.rfii.de/?p=1998>

Rat für Informationsinfrastrukturen (2019): Herausforderungen Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel. Göttingen. <http://www.rfii.de/?p=4043>

Rat für Sozial- und Wirtschaftsdaten (2020): Handreichung Datenschutz. 2., vollständig überarbeitete Auflage. RatSWD Output 8 (6). Berlin. doi:10.17620/02671.50

Sieroka N et al. (2018): Critical thinking in education and research – why and how? (guest editorial). *Angewandte Chemie* 57: 16574-16575. doi:10.1002/anie.201810397

Stodden V (2020): The data science life cycle: a disciplined approach to advancing data science as a science. *Communications of the association for computing machinery* 63 (7): 58-66. doi:10.1145/3360646

Schneider P et al. (2020): Rethinking drug design in the artificial intelligence era. *Nature reviews drug discovery* 19: 353-364. doi:10.1038/s41573-019-0050-3

Schüller K et al. (2021): Data Literacy Charta, Version 1.1., Stifterverband. Berlin. https://www.stifterverband.org/sites/default/files/data-literacy-charta_v1_2.pdf (abgerufen am 20.04.2021)

Wilkinson M et al. (2016): The FAIR guiding principles for scientific data management and stewardship. *Scientific data* 3: 160018. doi:10.1038/sdata.2016.18

Wissenschaftsrat (2011 a): Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Berlin. Drs. 10465-11. <https://www.wissenschaftsrat.de/download/archiv/10465-11.html> (abgerufen am 20.04.2021)

Wissenschaftsrat (2011 b): Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen. Berlin. Drs. 10464-11. <https://www.wissenschaftsrat.de/download/archiv/10464-11.html> (abgerufen am 20.04.2021)

Wissenschaftsrat (2011 c): Empfehlungen zur Zukunft des bibliothekarischen Verbundsystems in Deutschland. Berlin. Drs. 10463-11. <https://www.wissenschaftsrat.de/download/archiv/10463-11.html> (abgerufen am 20.04.2021)

B. Websites

22

U Bremen Research Alliance: www.uni-bremen.de/research-alliance.html

B.1. Mitglieder:

Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI): www.awi.de

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI): www.dfki-bremen.de

Deutsches Schiffahrtsmuseum Leibniz-Institut für Maritime Geschichte (DSM):
www.dsm.museum

Deutsches Zentrum für Luft- und Raumfahrt (DLR) – Institut für Raumfahrtsysteme:
www.dlr.de/irs

Deutsches Zentrum für Luft- und Raumfahrt (DLR) – Institut für den Schutz maritimer
Infrastrukturen: www.dlr.de/mi

Fraunhofer-Institut für Digitale Medizin MEVIS: www.mevis.fraunhofer.de

Fraunhofer-Institut für Fertigungstechnik und Angewandte Materialforschung IFAM:
www.ifam.fraunhofer.de

Fraunhofer-Institut für Windenergiesysteme IWES: www.iwes.fraunhofer.de

Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS:
www.leibniz-bips.de

Leibniz-Institut für Werkstofforientierte Technologien – IWT: www.iwt-bremen.de

Leibniz-Zentrum für Marine Tropenforschung (ZMT): www.leibniz-zmt.de

Max-Planck-Institut für Marine Mikrobiologie (MPIMM): www.mpi-bremen.de

Universität Bremen: www.uni-bremen.de

23

B.2. Weitere

Bremen Core Repository: www.marum.de/Forschung/IODP-Bohrkernlager.html

Data Science Center: www.dsc-ub.de

Deutsche Allianz Meeresforschung: www.allianz-meeresforschung.de

EOSC: www.eosc-portal.eu

GFBio e.V.: www.gfbio.org

Institute for Artificial Intelligence: www.ai.uni-bremen.de

KonsortSWD: www.ratswd.de/konsortswd

Leibniz-WissenschaftsCampus „Digital Public Health“: www.lsc-digital-public-health.de

MARUM GeoB Core Repository: www.marum.de/Infrastruktur/MARUM-GeoB-Kernlager.html

NFDI: www.nfdi.de

NFDI4BioDiversity: www.nfdi4biodiversity.org

NFDI4Health: www.nfdi4health.de

NFDI4Ing: www.nfdi4ing.de

NFDI-Förderprogramm der DFG: www.dfg.de/foerderung/programme/nfdi

openEASE: www.open-ease.org

PANGAEA: www.pangaea.de

Qualiservice: www.qualiservice.org/de

AAI	Authentication and Authorization Infrastructure
DFG	Deutsche Forschungsgemeinschaft
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
GFBio	GFBio – Gesellschaft für Biologische Daten e. V.
MARUM	MARUM – Zentrum für Marine Umweltwissenschaften der Universität Bremen
NFDI	Nationale Forschungsdateninfrastruktur
openEASE	Web-based knowledge service providing robot and human activity data
PANGAEA	Data Publisher for Earth & Environmental Sciences
Qualiservice	Forschungsdatenzentrum für qualitative sozialwissenschaftliche Forschungsdaten (Universität Bremen)
RatSWD	Rat für Sozial- und Wirtschaftsdaten
RfII	Rat für Informationsinfrastrukturen
WR	Wissenschaftsrat

Kurzprofil 1: NFDI-Konsortien der ersten Förderphase mit Bremer Beteiligung

Im **NFDI4BioDiversity**-Konsortium kooperieren Partnereinrichtungen aus den Bereichen Biodiversität, Ökologie und Umweltdaten. Hinter Biodiversität verbirgt sich dabei mehr als „nur“ die Vielfalt der Arten. Sie umfasst auch die genetische Vielfalt, die funktionelle Vielfalt, die Interaktionen und die Vielfalt ganzer Ökosysteme. Das Spektrum der Datenproduzenten reicht dabei von den Bürgerwissenschaftler*innen über Fachgesellschaften, Verbände und Behörden bis zu den Wissenschaftler*innen. Ziel dieses NFDI-Konsortiums ist, die Bandbreite der Biodiversitätsdaten für die anstehenden Entscheidungen in Politik und Gesellschaft verfügbar zu machen.

Im Rahmen von **NFDI4Health** sollen hochwertige Daten aus den Bereichen Epidemiologie, Gesundheitswesen und klinischer Forschung unter Wahrung des Datenschutzes nach den FAIR-Prinzipien international zugänglich und verknüpfbar gemacht werden. Dies eröffnet der Forschung ganz neue Optionen. NFDI4Health umfasst eine vollständige Abdeckung großer epidemiologischer Studien. Ein multidisziplinäres Team aus Wissenschaftler*innen soll in Deutschland fokussiert die Verschmelzung dieser personenbezogenen Gesundheitsdaten fördern.

Die mit **KonsortSWD** verbundenen Disziplinen – Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften – adressieren mit ihrer Forschung die großen Herausforderungen unserer Gesellschaft. Die Datengrundlage sowie Datenquellen sind dabei hoch divers. Wie bei den gesundheitlichen Daten gilt es hier, ethische und rechtliche Aspekte zu berücksichtigen. KonsortSWD ermöglicht eine FAIRe Datengrundlage, die multidisziplinäre Zusammenarbeit zulässt, und schafft zusätzlich Unterstützungsangebote. Beteiligte Einrichtungen (u. a. Qualiservice aus Bremen) bringen dabei Expertise und Erfahrungen im Betrieb von nutzungsorientierten Forschungsdateninfrastrukturen in die NFDI ein.

Das **NFDI4Ing**-Konsortium verbindet die Fachdisziplinen der Ingenieurwissenschaften. Die ingenieurwissenschaftlichen Daten werden FAIR aufbereitet. Dabei soll ein innovativer methoden- und benutzer*innen-orientierter Ansatz individuelle Szenarien beim Umgang mit Daten berücksichtigen und somit alle Datentypen durch Workflows abdecken.

Kurzprofil 2: Forschungsdatenmanagement in der Meeresforschung

Der Standort Bremen zeichnet sich durch seine starke Beteiligung in der Meeresforschung aus: Fünf Forschungsinstitute und zwei Hochschulen setzen in diesem Bereich Forschungsschwerpunkte. Als Mitglieder der **Deutschen Allianz Meeresforschung (DAM)** verfolgen die Meeresforschungseinrichtungen vier Kernbereiche – Forschungsmissionen, Infrastrukturen, Datenmanagement und Digitalisierung sowie Transfer – und tragen neben der Weiterentwicklung und Sichtbarkeit der deutschen Meeresforschung maßgeblich zur einheitlichen, föderalen

Datenstrategie bei. Das „**Observation to Archive and Analysis (O2A)**“-Framework für das FAIRe Forschungsdatenmanagement am **Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)** besteht aus aufeinander abgestimmten und miteinander kommunizierenden Softwarekomponenten, die einen durchgängigen Fluss der Daten von Geräten und Sensoren über deren Echtzeitmonitoring bis zur Analyse und Datenpublikation erlauben. Ursprünglich für den reibungslosen Datenfluss auf und von den Forschungsschiffen gedacht, erlaubt seine generische Architektur einen breiten Einsatz überall dort, wo Daten und Metadaten konsistent erfasst und zur Verfügung gestellt werden sollen. Aktuell werden O2A-Komponenten auf den Forschungsschiffen der DAM ausgerollt, um einen einheitlichen FAIRen Fluss der Daten und deren nachhaltige Verfügbarkeit zu gewährleisten.

Das **Leibniz-Zentrum für Marine Tropenforschung (ZMT)**, ebenfalls Mitglied der DAM, hat mit dem Leibniz-Sondertatbestand **DigiZ** die Weichen für eine Forschungsdateninfrastruktur gestellt und damit den zukunftsorientierten Ausbau von Datenwissenschaften zum Ziel. Die typischerweise heterogenen Forschungsdaten aus tropischen Küstenregionen sollen FAIR aufbereitet werden, um sie so dann mit modernen Data-Science-Methoden auszuwerten. Damit können sozial- und meereswissenschaftliche Daten systematisch verknüpft werden, um so neue Erkenntnisse mit transdisziplinärer Relevanz zu gewinnen. Gemeinsam mit dem Forschungsdatenzentrum Qualiservice entwickelt das ZMT Workflows für den Langzeiterhalt qualitativer sozialwissenschaftlicher Forschungsdaten.

Kurzprofil 3: Forschungsdatenmanagement in der Sozial- und Wirtschaftsforschung

Da unsere Gesellschaft mit allen Fachdisziplinen der Wissenschaft Berührungspunkte findet, gestaltet sich die Forschung hier multidisziplinär. Die Digitalisierung eröffnet dabei neue Möglichkeiten zur Datenerhebung und erweitert somit das ohnehin schon breite Spektrum an Methoden der Sozial- und Wirtschaftswissenschaften – sowohl bei der Datenerhebung als auch bei der Datenauswertung. Dementsprechend vielfältig sind Datenquellen und -typen. Im Rahmen des **Sonderforschungsbereichs (SFB) 1342 „Globale Entwicklungsdynamiken der Sozialpolitik“** der Universität Bremen arbeiten Institute der Sozial-, Rechts-, Politikwissenschaften, Geografie und Informatik zusammen, um globale Daten zu sozialpolitischen Fragestellungen zusammenzutragen und auszuwerten. Quantitative Daten werden in einer dafür neu entwickelten Datenbank „**WeSIS**“ archiviert. Qualitative Daten, die eine gesonderte Handhabung verlangen, werden im dafür spezialisierten Forschungsdatenzentrum **Qualiservice** (s. u.) archiviert. Die kuratierten Daten sollen langfristig (unter Wahrung der Datenschutzbestimmungen) zugänglich sein. Im Rahmen des **Forschungsinstituts Gesellschaftlicher Zusammenhalt (FGZ)**, das in Bremen und neun anderen Bundesländern angesiedelt ist, werden interdisziplinäre quantitative und qualitative Forschungsdaten zu Kulturen und Strukturen des Zusammenhalts generiert. Das in Bremen angesiedelte **Datenzentrum Zusammenhalt (DZZ)** widmet sich der Sammlung, Weitergabe und wissenschaftlichen Nutzung der quantitativen Forschungsdaten, während das Forschungsdatenzentrum Qualiservice die Kuratierung und Archivierung der qualitativen Forschungsdaten übernimmt.

Kurzprofil 4: Forschungsdatenmanagement in den Gesundheitswissenschaften

Auch in den Gesundheitswissenschaften ermöglicht die Digitalisierung den Einsatz innovativer Technologien und demzufolge neuartige, digitale Forschungsmethoden und Forschungsfragen. Diesen widmet sich der Bremer **Leibniz-WissenschaftsCampus „Digital Public Health“**, in dem die **Universität Bremen**, das **Fraunhofer-Institut für Digitale Medizin MEVIS** und das **Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS** kollaborativ forschen.

Das Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS entwickelt innovative Methoden für die Erhebung von Gesundheitsdaten und deren statistische Analyse. Dabei nutzt es vielfältige Datenquellen zur Erforschung bedeutsamer Gesundheitsstörungen sowie zur Entwicklung und Evaluation von präventiven Maßnahmen. Es leistet darüber hinaus einen maßgeblichen Beitrag zum Aufbau von überregionalen und internationalen Forschungsinfrastrukturen, beispielsweise im Rahmen von NFDI4Health. Das BIPS bietet für hausinterne Forschungsprojekte ein zentralisiertes Datenmanagement über den gesamten Datenlebenszyklus. Dies beinhaltet Beratung sowie Unterstützung bei der Datenerhebung und -auswertung und eine qualitätsgesicherte, standardisierte Datenaufbereitung und -verwaltung. Die gebündelte und kontinuierlich weiterentwickelte Kompetenz führt zu einer herausragenden Expertise. Zur Wahrung der rechtlichen Aspekte, die insbesondere bei personenbezogenen Daten beim Aufbau einer nationalen Forschungsdateninfrastruktur beachtet werden müssen, beteiligt sich das **Institut für Informations-, Gesundheits- und Medizinrecht (Universität Bremen, IGMR)** an dem NFDI-Vorhaben. Auch das **Kompetenzzentrum für Klinische Studien (Universität Bremen, KKSB)** setzt bei der Durchführung klinischer Studien auf ein zentrales Datenmanagement und bietet Beratung sowie Unterstützung an. Darüber hinaus verfügen die Statistiker*innen über eine methodische Expertise, die sie zukunftsorientiert weiterentwickeln. Ziel des KKSB ist es, die klinische Forschung in der Region auf einem hohen wissenschaftlichen und organisatorischen Niveau voranzutreiben. Das Fraunhofer MEVIS entwickelt Softwaresysteme für die bild- und datengestützte Früherkennung, Diagnose und Therapie sowie zur Unterstützung der qualitätsgesicherten Durchführung und Auswertung klinischer Studien. Gemeinsam mit dem BIPS ist Fraunhofer MEVIS an Deutschlands größter Gesundheitsstudie „NAKO“¹ beteiligt und verantwortet unter anderem die technische Infrastruktur und das Bilddatenmanagement.

Kurzprofil 5: Forschungsdatenmanagement in Materialforschung und Technik

Im **Leibniz-Institut für Werkstofforientierte Technologien – IWT** werden im interdisziplinären Austausch zwischen den Fachgebieten Fertigungs-, Verfahrens- und Werkstofftechnik neue Materialien entwickelt und erforscht. Eine kleine Auswahl der am Institut untersuchten Materialien steht in den beiden BMBF-Projekten **MatDigi** und **NanoS-QM** im Mittelpunkt hinsichtlich der mit ihnen verknüpften Daten. Im Verbund mit den Projektpartnern (Bundesanstalt für Materialforschung und -prüfung (BAM), Fraunhofer-Institut für Werkstoffmechanik IWM, Karlsruher Institut für Technologie (KIT), Max-Planck-Institut für Eisenforschung (MPIE)) werden bei **MatDigi** verschiedenste Materialien und deren Eigenschaften maschinenlesbar digitalisiert. Dazu gehören sowohl physikalische Eigenschaften als auch produktions- und

¹ www.nako.de

anwendungsbezogene Parameter. Ziel ist, eine gut nutzbare Plattform zu etablieren, die den Anforderungen aller Projektpartner genügt und in der Daten sowohl eingepflegt als auch abgerufen werden können. Die Plattform soll nicht nur Forschungseinrichtungen, sondern auch Technologieunternehmen unterstützen und liefert damit einen gesamt-ökonomischen Mehrwert.

Im Projekt **NanoS-QM** arbeitet das IWT mit vier Leibniz-Instituten (FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Leibniz-Institut für Arbeitsforschung an der TU Dortmund (IfADo), INM – Leibniz-Institut für Neue Materialien, IUF – Leibniz-Institut für umweltmedizinische Forschung) zusammen an der Entwicklung von Kurationskriterien und Qualitätsstandards für Forschungsdaten mit Fokus auf Sicherheitsaspekten von Nanomaterialien. Zusammen mit Expert*innen aus anderen Fachgebieten werden Daten zu Materialien samt ihrer toxikologischen Wirkung erfasst. Anhand laborrelevanter Workflows wird ein Metadatenschema strukturiert entworfen, mit dem eine klare und verständliche Nachnutzung möglich wird. Zudem sollen diese Daten so verlässlich und aussagefähig sein, dass sie bei regulatorischen Prozessen wie Grenzwert-Setzungen als Entscheidungsgrundlage dienen können (vgl. auch Use Case 6).

Das **Fraunhofer-Institut für Fertigungstechnik und Angewandte Materialforschung IFAM** leistet Spitzenforschung mit starker Anwendungsorientierung auf den Gebieten Klebtechnik, Oberflächen, Formgebung und Funktionswerkstoffe. Im Zuge dieser Forschung werden Datenwertschöpfungsketten unter anderem zur Materialherstellung, -verarbeitung und -prüfung untersucht und implementiert mit dem Ziel, die Daten, soweit möglich, nach den FAIR-Prinzipien zu prozessieren (vgl. Use Case 7).

Im Rahmen des **SFB 1232 „Von farbigen Zuständen zu evolutionären Kunstwerkstoffen“** haben Materialwissenschaftler*innen und Informatiker*innen eine Projektdatenbank aufgebaut, die es erlaubt, große Mengen heterogener Daten integrativ zu archivieren und gleichzeitig das Probenmaterial, an dem Messungen durchgeführt worden sind, zu verwalten. Dabei werden Proben direkt mit den Analyseergebnissen verknüpft. Neue Technologien ermöglichen immer schnellere, simultan durchgeführte Messungen. Die rasante Datenproduktion stellt große Anforderungen an das Proben- und Datenmanagement. Auch in der Robotik müssen Datenströme (meist Multisensordaten) möglichst in Echtzeit prozessiert, harmonisiert und zum Monitoring dargestellt werden. Dafür haben Institute, Projekte oder Arbeitsgruppen eigene Tools und Prozeduren erstellt, die perspektivisch automatisiert werden sollen.

Kurzprofil 6: Data Science an der Universität Bremen

Das **Data Science Center (DSC)** ist ein interdisziplinäres Institut, das die Data-Science-Kompetenzen der Universität Bremen bündelt, koordiniert und weiterentwickelt. Dabei beruht es auf den drei Säulen Forschung, Qualifizierung und Dienste. Die Aufgaben des DSC bestehen darin, datengetriebene Forschung voranzutreiben, Data-Science-Kompetenzen zu stärken und die interdisziplinäre Zusammenarbeit zu fördern. Dabei bringt das DSC Wissenschaftler*innen aus allen Disziplinen zusammen, um die Entwicklung von innovativen Forschungsfragen an der Schnittstelle von Theorie, Methodik und Anwendung zu ermöglichen. So können aus der Praxis

heraus Impulse für die Grundlagenforschung gegeben, aber auch bereits etablierte Methoden und Technologien auf neue Anwendungskontexte übertragen werden. Darüber hinaus bietet das DSC IT-Ressourcen zur Durchführung von rechenintensiven Datenanalysen sowie Unterstützung bei der Entwicklung und Implementierung von Data-Science-Methoden wie maschinellem Lernen. Das übergeordnete Ziel des DSC ist es, die Universität Bremen als Impulsgeberin und den Standort Bremen als Vorreiter im Bereich Data Science zu positionieren.

Weitere Leuchttürme im Bereich Data Science sind das renommierte **IAI – Institute for Artificial Intelligence**, das mit dem **openEASE**-Projekt die Open Data-Kultur lebt, und das **ZeTeM – Zentrum für Technomathematik**, welches moderne mathematische Methoden in Industrie und Wissenschaft transferiert.

Anhang B: Use Cases aus der U Bremen Research Alliance

Use Case 1: Das „Observation to Archive and Analysis (O2A)“-Framework für das FAIRe Forschungsdatenmanagement am AWI

Autor*innen: Roland Koppe, Peter Gerchow, Tilman Dinter, Janine Felden, Frank Oliver Glöckner

Das **Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)** untersucht mit innovativer Forschung, einer ausgezeichneten wissenschaftlichen Infrastruktur und langjähriger Expertise praktisch alle Bereiche des Erdsystems – von der Atmosphäre bis zum Grund der Meere. Das Klimageschehen der Erde zu verstehen, ist dabei ein zentrales Anliegen der Wissenschaftler*innen. Wer das Klima der Zukunft verstehen will, muss gegenwärtige Veränderungen analysieren, die Klimageschichte des Planeten kennen und kurzfristige Schwankungen von langfristigen Trends unterscheiden können. Die Forscher*innen am AWI betreiben eine Vielzahl von Observatorien, die Messdaten über lange Zeiträume erheben. Sie erforschen mit ihren Schiffen und Flugzeugen sowie einer Vielzahl an Messinstrumenten die Atmosphäre, das Eis, das Meer und die Küste.

Das AWI versteht sich dabei als Produzent und Lieferant von qualitätsgesicherten Forschungsdaten für die Forschung, Gesellschaft und Politik. Um diesem Auftrag nachzukommen, wurde und wird das **„Observation to Archive and Analysis (O2A)“-System** mit den folgenden Zielen (weiter-)entwickelt: 1.) Durchgängiges qualitätsgesichertes und FAIRes Datenmanagement über den gesamten Datenlebenszyklus hinweg. 2.) Umfangreiche Beschreibung aller Daten mit Metadaten. 3.) Zentrale Datenhaltung und geregelter Zugriff für die wissenschaftlichen Teams. 4.) Nachverfolgbarkeit aller Transformations- und Prozessierungsschritte (Provenance). 5.) Zeitnahe Veröffentlichung der Daten in ausgewiesenen Langzeitrepositorien. 6.) Zitierbarkeit und Nachverfolgbarkeit der Nachnutzung aller Forschungsdaten in klassischen sowie Daten-Publikationen.

O2A besteht aus aufeinander abgestimmten und miteinander kommunizierenden Softwarekomponenten, die einen durchgängigen Fluss der Daten von den Geräten (Sensoren) über das Echtzeitmonitoring bis zur Analyse und Datenpublikation erlauben (siehe nachstehende Abbildung).

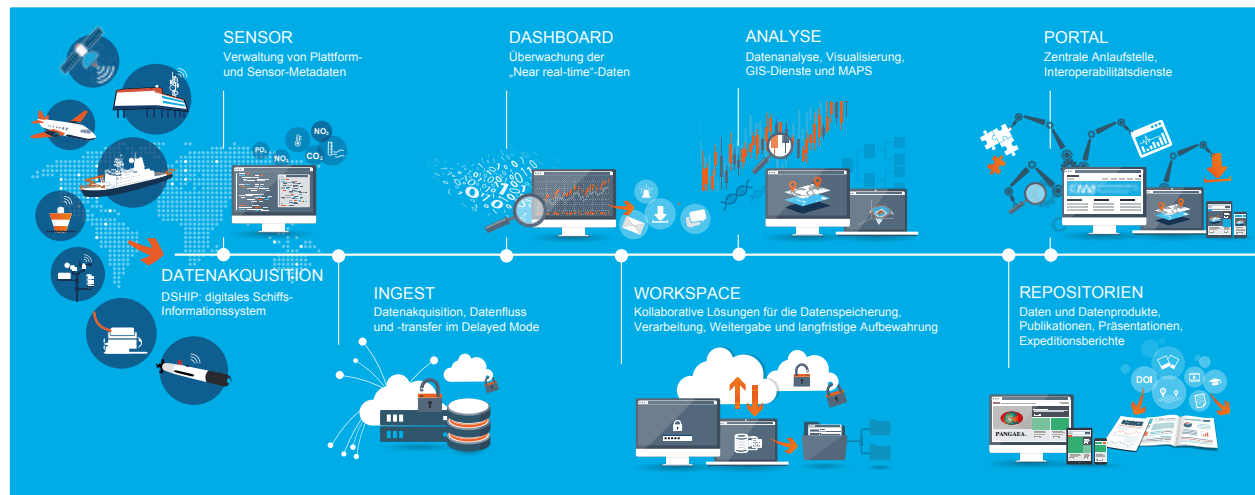


Abbildung: Übersicht über die O2A-Komponenten

©Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)

Beschreibung der O2A-Komponenten

Geräte (Sensoren): Verwaltung von gerätebezogenen Informationen und Versionierung. Aufzeichnung und Rückverfolgung der Gerätehistorie und Ereignisse, Erfassung umfangreicher Metadaten für jeden Datensatz.

Ingest: (fast) Echtzeit-Transport von Daten aus Geräten und Sensoren auf Schiffen, in Flugzeugen, über Stationen, Bojen, Glider etc. in Speichersysteme auf dem Schiff und an Land. Zusätzlich können Daten von Bändern und Festplatten nahtlos eingespielt werden.

Dashboard: Überwachung der (fast) Echtzeit-Daten mittels einer flexibel von den Nutzer*innen konfigurierbaren Oberfläche.

Workspace und Analysefunktionen: zentrale Speicherung der Daten aus den Geräten sowie zusätzlicher durch die Nutzer*innen hochgeladener Daten im Sinne eines gemeinsamen Datenpools, cloud-basierte flexible Arbeitsumgebung für die Analyse der Daten mittels virtueller Maschinen (R, Matlab etc.) und Jupyterhub nach dem Prinzip des „function shipping“ statt „data shipping“.

Repositorien: Den Leitlinien des AWI folgend müssen Roh- und Primärdaten zeitnah in ausgewiesenen Langzeitrepositorien abgelegt und zugänglich gemacht werden. PANGAEA (Data Publisher for Earth & Environmental Science) und ePIC (electronic Publication Information Center) erlauben die Archivierung und Publikation von Daten, Datenprodukten, Publikationen, Präsentationen und Expeditionsberichten. Alle in PANGAEA abgelegten und

publizierten Daten werden durch Kurator*innen nach internationalen Standards aufbereitet und erhalten eine DOI.

Portal: pangaea.de und marine-data.de sind integrative Anlaufstellen zum Auffinden und Wiederverwenden der AWI-Forschungsdaten, -Expeditionen und -Datenprodukte. Der Zugriff auf die Daten ist mittels Weboberfläche und programmatisch möglich.

Use Case 2: Forschungsdatenmanagement und Data Science am Leibniz-BIPS

Autor*innen: Achim Reineke, Iris Pigeot

Das **Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS** entwickelt wirksame Strategien zur Prävention chronischer, nichtübertragbarer Erkrankungen. Dabei ist es vorrangiges Ziel, die Weichen für einen gesunden Lebenslauf frühzeitig zu stellen. Im Mittelpunkt der Forschung stehen deshalb Ansatzpunkte von krankheitsübergreifender Relevanz wie Lebensstil und Lebensumwelt, biologische und soziale Faktoren sowie Früherkennung und Therapiesicherheit. Das Forschungsspektrum reicht von der Methodenentwicklung über die Erkennung von Krankheitsursachen bis hin zur Interventions- und Implementationsforschung. Forschungsgrundlage sind umfangreiche Primär- und Sekundärdatensammlungen zu Gesundheits- beziehungsweise Krankheitsstatus sowie möglichen Einflussfaktoren. Diese Daten sind als personenbezogene Daten besonders zu schützen. Die Wahrung der Regeln des Datenschutzes ist damit oberstes Gebot für die Forschung am BIPS. Dies spiegelt sich auch in den verschiedenen Datenaufbereitungs- und Datenmanagementprozessen sowie einer strikten Reglementierung der Zugriffsberechtigungen wider. Die Verantwortlichkeiten für die Datenerhebung, -aufbereitung und -auswertung für alle Studien des BIPS liegen zentral bei eigens dafür zuständigen Fachgruppen. Diese verfügen über entsprechend qualifiziertes Personal und bestehen auch über die jeweiligen Projektlaufzeiten hinaus.

Erhebung, Aufbereitung, Analyse und Archivierung von Forschungsdaten: Dazu sind am BIPS verschiedene strukturierte Datenmanagementprozesse etabliert worden. Diese Prozesse umfassen den kompletten Lebenszyklus der Daten von ihrer Entstehung bis zur abschließenden Archivierung und werden durch dafür erstellte Werkzeuge und Datenbanken unterstützt. Dazu gehören eine Projektdatenbank mit Basisinformationen zu den jeweiligen Studien (Metadaten) und weitere Datenbanken mit detaillierteren Informationen über verwendete Fragebögen oder sonstige Erhebungsinstrumente bis hin zur Beschreibung der einzelnen Variablen und ihrer möglichen Ausprägungen. Die Werkzeuge und Prozesse werden regelmäßig überprüft und gegebenenfalls angepasst.

Datenmanagementprozesse: Die ersten Datenmanagementprozesse starten während der Studienplanung und schaffen die Grundlage für ein qualitativ hochwertiges Forschungsdatenmanagement. In dieser Phase werden die Datenmodelle anhand der Angaben zu den in einer Studie zu erhebenden beziehungsweise genutzten Informationen erstellt und das gesamte Management der Daten geplant. Die Planungen werden dokumentiert oder in Datenmanagementplänen festgehalten. Für die Speicherung und Verwaltung der Metadaten zu den einzelnen

Variablen wird eine eigens dafür erstellte Datenbank verwendet. Diese stellt einerseits eine standardisierte Ablage der Metadaten sicher und ist andererseits die Quelle für die in den folgenden Studienphasen benötigten Informationen. So können in der Phase der Datenerhebung Plausibilitätsprüfungen basierend auf den hinterlegten Angaben zu gültigen Wertebereichen durchgeführt oder die für die abschließende Auswertung benötigten Codebooks erstellt werden. Nach dem Abschluss einer Studie werden die Daten und verwendeten Auswertungsprogramme schließlich nach einem definierten Prozedere archiviert, um auf diese Weise die Nachvollziehbarkeit der Analysen und die Nachnutzbarkeit der Daten sicherzustellen.

Qualitätsgesichertes Datenmanagement: Die Maßnahmen haben insgesamt zum Ziel, ein qualitätsgesichertes Datenmanagement über den gesamten Datenlebenszyklus hinweg und damit letztendlich die Qualität und Zuverlässigkeit der Studienergebnisse selbst sicherzustellen. Sie ermöglichen zudem die Nutzbarkeit der Daten über das Projektende hinaus. In Zukunft sollen die Prozesse um Komponenten zur vereinfachten Bereitstellung der Daten für die externe Nutzung erweitert werden.

Datenauswertung: Um eine qualitätsgesicherte Datenauswertung mit modernsten statistischen Methoden sicherzustellen, sind am BIPS drei Fach- und zwei Forschungsgruppen mit spezifischer statistischer Expertise etabliert. Diese beraten die Forschenden in den Fachabteilungen des BIPS, unterstützen bei der statistischen Planung und Auswertung und entwickeln statistische Methoden, die auf die wissenschaftlichen Fragestellungen des BIPS zugeschnitten sind.

Use Case 3: Langzeitarchivierung und Nachnutzung von qualitativen sozialwissenschaftlichen Forschungsdaten im Forschungsdatenzentrum Qualiservice

Autor*innen: Elisabeth Huber, Jan-Ocko Heuer, Kati Mozygemba, Betina Hollstein

Das am **SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik der Universität Bremen** angesiedelte Forschungsdatenzentrum **Qualiservice** archiviert qualitative sozialwissenschaftliche Forschungsdaten aus unterschiedlichen Disziplinen und stellt sie für die wissenschaftliche Nachnutzung zur Verfügung. Qualitative Forschungsmaterialien dokumentieren beispielsweise Interviews und Beobachtungen. Sie können in Schriftform als Protokoll oder Transkript vorliegen, aber auch als Audio-, Bild- oder Videodatei. Diese Dokumentationen enthalten in der Regel sensible, personenbezogene Daten und stellen in datenschutzrechtlicher und forschungsethischer Hinsicht eine besondere Herausforderung dar. Um die persönlichen Daten der Forschungsbeteiligten zu schützen, hat Qualiservice sichere Arbeitsabläufe und ein Safe Center eingerichtet. Im Safe Center führt Qualiservice Maßnahmen zum Schutz der Daten durch und schließt die Vorbereitungen für die Sekundärnutzung ab. Ein Arbeitsplatz für Gastwissenschaftler*innen ermöglicht die Forschung direkt vor Ort. Stärker anonymisierte Datensätze dürfen die Sekundärnutzer*innen auch extern beforschen. Darüber hinaus sind qualitative Daten in der Regel nur wenig standardisiert und bedürfen einer sorgfältigen Dokumentation, Beschreibung und Kontextualisierung, um sinnverstehende Nachnutzungen zu ermöglichen. Die von Qualiservice angebotenen Leistungen beinhalten die persönliche und studienspezifische Beratung, die Kuration und Aufbereitung der Daten für die Langzeitarchivierung sowie

die Bereitstellung archivierter Forschungsdaten und relevanter Kontextinformationen für Forschung und Lehre. Das Angebot richtet sich an Projekte, (inter-)nationale Forschungsverbünde sowie inner- und außeruniversitäre Forschungsinstitute. Qualiservice adaptiert seine Workflows für spezifische Bereiche wie die ethnografische oder die Raumforschung. Im Juni 2019 ist Qualiservice vom RatSWD akkreditiert worden und orientiert sich an dessen Kriterien zur Qualitätssicherung. Im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI) koordiniert Qualiservice die Vernetzung und Weiterentwicklung wissenschaftlicher Infrastrukturen für die Archivierung und Sekundärnutzung qualitativer Daten aus der Sozialforschung in Deutschland. Es werden Konzepte entwickelt, die zur Abstimmung von Workflows in der interdisziplinären Datengenerierung und verteilten Archivierung qualitativer Daten beitragen.

Beratung beim Forschungsdatenmanagement: Qualiservice berät Forschende bei der Archivierung und Nachnutzung von qualitativen sozialwissenschaftlichen Forschungsdaten und stellt dafür Services und Tools (z. B. das Qualiservice Anonymisierungstool QualiAnon) zur Verfügung. Die Leistungen beziehen sich auf das Forschungsdatenmanagement sowie auf die Kuration von Forschungsdaten in der Langzeitarchivierung und tragen damit zur Datenqualität bei. Qualiservice berät Forschende zur Beantragung benötigter Ressourcen für das Data Sharing. Umfassende Beratungen werden zur fortlaufenden Dokumentation der Forschungsprozesse, zur informierten Einwilligung, Anonymisierung und Kontextualisierung der Forschungsdaten angeboten.

Datenkuration und Archivierung: Die Langzeitarchivierung und Veröffentlichung von Forschungsdaten und Metadaten wird gemeinsam mit PANGAEA und der Staats- und Universitätsbibliothek Bremen (SuUB) sichergestellt. Durch international anschlussfähige Metadaten werden Datensätze such- und auffindbar. Persistente Identifikatoren (DOI) sorgen für eine nachhaltige Zitierfähigkeit von Daten und Studienkontexten. Datenübergabevereinbarungen können studienspezifisch angepasst werden und gewährleisten den Schutz der Primärforschenden und Forschungsbeteiligten.

Nachnutzung von Forschungsdaten: Der Studienreport, der mit den Metadaten veröffentlicht wird, dient als wichtige Informationsquelle für die Nachnutzung von Forschungsdaten. Qualiservice unterstützt Sekundärforschende bei der Auswahl und Zusammenstellung geeigneter Daten und Kontextmaterialien für Forschung und Lehre. Nachnutzungsvereinbarungen stellen den forschungsethisches und datenschutzrechtlich angemessenen Umgang mit Forschungsdaten in der Sekundäranalyse sicher.

Use Case 4: Robotik und KI am DFKI in Bremen

Autoren: Florian Cordes, Christoph Lüth

Das **Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)** ist in Bremen mit zwei Forschungsbereichen vertreten. Das **Robotics Innovation Center (RIC)** und **Cyber-Physical Systems (CPS)** erforschen und entwickeln Künstliche Intelligenz für Systeme, die mit der Umwelt interagieren. Dieses sind zum einen mobile Robotersysteme, die an Land, zu Wasser, in der Luft und im Weltraum für komplexe Aufgaben eingesetzt werden, und zum anderen

intelligente und vernetzte Systeme, wie sie beispielsweise in Handys, Autos oder Flugzeugen zum Einsatz kommen.

Das DFKI untersucht und entwickelt Verfahren, um die korrekte Funktion von Systemen sicherzustellen (funktionale Sicherheit), sowohl auf der Ebene der Hardware (Schaltkreisentwurf) als auch auf der Ebene der Software. Die hierbei entwickelten Verfahren erzeugen einerseits Daten, deren Archivierung, Versionierung und Zugriff Gegenstand der Entwicklung eines kooperativen Forschungsdatenmanagements sind. Darüber hinaus können diese Verfahren maßgeblich dazu beitragen, die korrekte Funktionalität des Datenmanagements der Initiative sicherzustellen, beispielsweise durch formale Modellierung der Daten und der sie verarbeitenden Algorithmen.

In der Robotik sind Experimente auf Subsystem- und Gesamtsystemebene der Kern der Sicherstellung von Funktion und Robustheit der entwickelten Systeme und Lösungen. Hierbei wird im Allgemeinen eine große Menge heterogener Daten erzeugt, die in geeigneter Form für einen Langzeitzugriff und Interoperabilität gespeichert werden müssen. Hierbei geht es im Wesentlichen um Sensor-Aktorsysteme, die durch eine Interaktion mit der Umwelt ein Feedback erfahren und darauf aufbauend Handlungen ausführen. Ein Roboter besteht im Allgemeinen aus verschiedenen Gliedmaßen, die wiederum aus einzelnen Aktuatoren bestehen. Ein einzelner Roboter kann bereits das Gesamtsystem für eine Aufgabe darstellen, ebenso können mehrere Roboter zusammen ein System aus Systemen („System of Systems“, SoS) bilden, um komplexere Aufgaben durchzuführen. Auf allen Ebenen solcher komplexer Systeme fallen Daten an, die für den Betrieb wie auch die fortlaufende Entwicklung entscheidend sind.

Neben Experimentergebnissen und Betriebsdaten fallen Daten während des Designs eines Roboters an. Dies umfasst CAD-Daten der mechanischen Konstruktion und von Elektronik, die eigens für die Roboter erstellt werden. Aus den konstruktiven Daten werden zum einen Fertigungszeichnungen abgeleitet; gleichzeitig werden die Daten weitergereicht, um Simulationsmodelle und kinematische Modelle für die Gesamtsystemkontrolle zu erstellen. Hierfür stehen am DFKI RIC verschiedene Workflows bereit, um die einmal erzeugten Daten für die verschiedenen Arbeitsschritte aufzubereiten.

Beispiel Subsystemebene: Auf der Subsystemebene „Aktuator“ werden zur Entwicklungszeit Experimente durchgeführt, die beispielsweise zur Systemidentifikation, Aufnahme von Leistungsdaten und Regleroptimierung genutzt werden. Am DFKI werden hierauf aufbauend sogenannte Lebensläufe von Aktuatoren geführt, die jederzeit einen Vergleich mit vergangenen Leistungsdaten erlauben. Auf diese Weise können, um nur ein Beispiel zu nennen, Abnutzungserscheinungen anhand abweichender Stromkennlinien identifiziert werden. Hierzu wird ein Aktuator entworfen, der basierend auf den Normkennwerten zur Lebensdauer sein Systemmodell und seine lokalen Reglerparameter anpassen kann, um Abnutzung wie zunehmendem Lagerspiel selbstständig entgegenwirken zu können.

Beispiel Gesamtsystemebene: In Feldversuchen werden die robotischen Systeme des DFKI in realistischen beziehungsweise analogen Umgebungen erprobt. Im Falle des Rovers SherpaTT mit artikulierteem Fahrwerk waren dies jeweils vier Wochen in den Wüsten Utahs

und Marokkos in mars-analogen Umgebungen. Der jüngste Feldversuch wurde im März / April 2021 mit einer Dauer von fünf Wochen in einer Sandgrube in Norddeutschland durchgeführt. Neben wichtigen Daten aus dem Subsystemverhalten in natürlichen Umgebungen ergeben sich große Mengen an Daten aus der Umgebungsperzeption. Dies umfasst Laser-scanner- und Kameradaten wie auch verarbeitete Daten, beispielsweise Punktwolken, digitale Höhenkarten (Digital Elevation Map, DEM) oder Befahrbarkeitskarten. Im Rahmen des Projekts OG10-ADE (Autonomous DEcision Making in Very Long Traverses) erstellt das DFKI derzeit ein sogenanntes Validation Toolset. Hier können die Felddaten einerseits auf Metaebene analysiert und andererseits verschiedene Algorithmen auf den aufgezeichneten Sensorstreams gegeneinander evaluiert werden. Ein Nutzungsbeispiel wären verschiedene Verfahren zum Erzeugen einer DEM auf Basis eines in realer Umgebung erzeugten Laserscannerstreams. Mittelfristig müssen hierfür entsprechende Forschungsdatenmanagement-Strukturen genutzt werden, um diese Verfahren auch über das DFKI hinaus auf Basis der FAIR-Prinzipien nutzbar zu machen.

Repräsentation als Digitaler Zwilling: Im gerade gestarteten NFDI4Ing (NFDI für Ingenieurwissenschaften) ist das DFKI in der Task Area Golo – Felddaten und Verteilte Systeme beteiligt. Hier wird das Prinzip des Digitalen Zwillings genutzt, um Systeme und deren Daten zu organisieren, zu strukturieren und verfügbar zu machen. Das DFKI bringt die oben genannten robotischen Use Cases mit ein, während ein Verkehrsmanagementsystem als zweiter Use Case vorgesehen ist. Die Daten aus der Repräsentation werden genutzt, um (i) den Betrieb zu verbessern, (ii) Fehler und Probleme zu erkennen und vorausszusehen („Predictive Maintenance“) und (iii) Generationen von Systemen schrittweise verbessern zu können, indem auf die Felddaten der vorherigen Generation zugegriffen werden kann. Neben Sensordaten werden beispielsweise auch Prozesse der Datenaufnahme, Skripte und Software zur Analyse über den Digitalen Zwilling gespeichert.

Gesundheitsdaten: Bei zahlreichen Aktivitäten im Bereich der Gesundheit (wie in den Projekten KI-SIGS, ASCAPE, CrowdHEALTH) fallen Gesundheitsdaten an, aus denen mit Methoden der KI neue Erkenntnisse gewonnen werden sollen. So wird im Projekt KI-SIGS versucht, aus den von auf der Intensivstation behandelten Patient*innen abgeleiteten Daten Anzeichen für eine kardiopulmonale Dekompensation zu erkennen. Im Projekt ASCAPE wird indes der Zusammenhang zwischen Lebensqualität und Gesundheit von Krebspatient*innen und den angewandten Behandlungen untersucht. In beiden Projekten ist das DFKI an der zentralen Datenverarbeitungsplattform maßgeblich beteiligt. Diese Daten unterliegen besonders strengen Vertraulichkeitsregeln und können ein weiterer Use Case für die hier vorgeschlagene Initiative sein.

Use Case 5: Die Leitlinien zum Umgang mit Forschungsdaten der Universität Bremen

Autor: Björn Oliver Schmidt

Speicherung und Archivierung von Forschungsdaten finden bereits in der 2017 verabschiedeten Abfassung der **Universität Bremen** zur guten wissenschaftlichen Praxis Erwähnung.

Hier heißt es, dass Forschungsdaten für mindestens zehn Jahre auf gesicherten Datenträgern gespeichert werden sollen. Die Rahmenbedingungen in Form von nationalen und internationalen Vorgaben und Empfehlungen haben sich seitdem weiterentwickelt. Das Management von Forschungsdaten hat in den letzten Jahren deutlich an Bedeutung gewonnen. Vor diesem Hintergrund hat es sich die Universität Bremen im Rahmen des „Lenkungsraums Forschungsdatenmanagement“ zur Aufgabe gemacht, Leitlinien zu erarbeiten, die Rahmenbedingungen für ihre Mitglieder formulieren.

Eine Herausforderung hat zunächst darin bestanden, eine Klärung zu erwirken, welche Personen und Statusgruppen hiermit angesprochen werden. Werden Studierende mit einbezogen oder beziehen sich die Empfehlungen lediglich auf forschende Personen an der Universität Bremen? Man hat sich letztlich mit dem Bremer Hochschulgesetz beholfen und alle Mitglieder der Universität angesprochen, wodurch auch Studierende einbezogen werden.

Da es sich um Empfehlungen handelt, die keinen verbindlichen Charakter haben, ist es umso mehr von Bedeutung gewesen, einen Diskussionsprozess über die komplette Universität und alle Fachbereiche hinweg zu führen, sodass eine höhere Akzeptanz erreicht werden kann. Hierdurch ist der Prozess deutlich verlängert und es sind immer wieder neue Versionen des Papiers vorgelegt und diskutiert worden. Letztlich hat dies aber dazu geführt, dass eine höhere Qualität erreicht und ein Konsens über den Inhalt erzielt worden ist.

Die Diversität der Fachbereiche und Forschungsthemen, die an der Universität Bremen vertreten sind, haben zur Folge gehabt, dass die Empfehlungen sehr allgemein gehalten sind. Wie Forschungsdaten im Detail, beispielsweise vor dem Hintergrund der FAIR-Prinzipien, zu behandeln und zu verarbeiten sind, muss von den wissenschaftlichen Fachgesellschaften oder den Fach- und Forschungsbereichen selbst erarbeitet werden. Detaillierte und auf Fächer abgestimmte Empfehlungen und / oder Vorgaben können von der Universität nicht geleistet werden.

Dies hat immer wieder zu Kritik an den Empfehlungen selbst geführt. Diese seien zu allgemein formuliert und böten letztlich keinen konkreten Nutzen für die Praxis. Hiermit sollen aber lediglich Rahmenbedingungen geschaffen werden, die dann von den Bereichen mit weitergehenden Inhalten gefüllt werden müssen. Darüber hinaus soll das Papier dafür sorgen, dass in den Bereichen über das Thema diskutiert wird und es zu einer allgemeinen Sensibilisierung für das Thema Forschungsdatenmanagement kommt, sollte dies noch nicht geschehen sein. Eine Konkretisierung der Leitlinien erfolgt nun (auch) durch das vorliegende Whitepaper.

Use Case 6: Implementierung eines Forschungsdatenmanagement-Systems am Leibniz-IWT

Autoren: Norbert Riefler, Lutz Mädler

Ausgangsüberlegungen: Das **Leibniz-Institut für Werkstofforientierte Technologien – IWT** vereint drei Fachgebiete der Produktionstechnik (Fertigungs-, Verfahrens- und Werkstofftechnik), deren Forschungsdaten sehr heterogen sind bezüglich Inhalt, Struktur und Größe der

Daten. Beispielsweise können experimentelle Daten einfache Wertetabellen im unteren kByte-Bereich aufweisen, wohingegen High-Speed-Kamera-Filme viele GByte groß sein können. Daten aus Simulationen besitzen ebenfalls nicht nur stark unterschiedliche Größen, sondern sind auch inhaltlich sehr verschieden (z. B. elektromagnetische Streulichtsimulationen versus strukturmechanische Festigkeitsberechnungen). Dennoch sollen alle Daten, die in irgendeiner Form veröffentlicht werden (Paper, Vortrag, Dissertation etc.), in ihrer Entstehungsgeschichte möglichst vollständig verfügbar und direkt aus den Rohdaten nachvollziehbar sein zur Erfüllung der guten wissenschaftlichen Praxis und den FAIR-Prinzipien einerseits sowie als interne Dokumentation für eine Nachnutzung andererseits. Für diese Ausgangslage ist eine pragmatische Lösung gesucht worden.

Umsetzung: Aufgrund des hohen Aufwandes bei der Kuratierung heterogener Daten, was für eine strukturierte Datenablage in einer Datenbank problematisch ist, ist eine All-in-One-Lösung als unrealistisch verworfen worden. Stattdessen sollte das Forschungsdatenmanagement-System (FDM) aus bereits verfügbaren Komponenten bestehen und zu einem integrierten, kollaborativen System ausgebaut werden. Zwei Hauptkomponenten stehen dabei im Zentrum:

1. ein Electronic Lab Notebook (ELN), das aufgrund der Heterogenität der Daten generisch nutzbar sein muss und mit dem die Wissenschaftler*innen ihre tägliche Laborpraxis dokumentieren sollen. Darüber hinaus sollen Daten aus Messgeräten sowie Parameter aus den Versuchen interaktiv oder automatisiert über Application Programming Interfaces (APIs) hochgeladen werden können. Die Daten können mit Zeitstempel versehen werden und die Experimente lassen sich in Form eines PDFs mit QR-Code für den Link auf die Daten im ELN zusammenfassen, womit ein Probenmanagement möglich wird.

2. eine strukturierte Datenablage nach streng vorgegebenem Muster bezüglich Verzeichnisnamen, in dem sämtliche in einer Publikation dargestellten Daten nachvollziehbar in einem Ordner abgelegt werden. Die Daten müssen mit einem Mindestmaß an Metadaten versehen werden (DataCite-Metadaten-Schema). Alle Messungen, Materialproben oder auch Simulationen werden im ELN dokumentiert und mit einem eindeutigen Hash-Wert (QR-Code) versehen.

Diese Kopplung eines gewöhnlichen Speichers (NAS) mit einem generischen ELN stellt den ersten Schritt eines den FAIR-Prinzipien genügenden FDMs dar. Seit 2021 müssen Mitarbeiter*innen aus einer der drei Hauptabteilungen (IWT-VT) dieses System verpflichtend nutzen, während das System in den anderen Hauptabteilungen schrittweise eingeführt wird. Dazu ist eine Datenmanagement-Guideline verfasst worden, in der alle erforderlichen Aktionen festgelegt worden sind, etwa eine auf die konventionelle Speicherung auf Festplatten zugeschnittene Namenskonvention für Dateien und Verzeichnisse. In Planung befindet sich bereits eine Datenbank mit webbasiertem Frontend, in dem alle Veröffentlichungen verzeichnet und mit den eingepflegten Metadaten versehen sind, um damit neben der internen Nutzung auch einen Zugang zu den Daten von außen zu ermöglichen.

Use Case 7: Zusammenführung von Sensor- und Messsystemen auf Hardware- und Datenebene am Fraunhofer IFAM

Autoren: Michael Hoffmann, Sascha Thinius

Im Fokus der derzeitigen Bemühungen am **Fraunhofer-Institut für Fertigungstechnik und Angewandte Materialforschung IFAM** stehen das Zusammenführen von Sensor- und Messsystemen auf Hardware- und Datenebene, von Modellierung und Experiment sowie Einbettung in Industrie-4.0-Produktions- und -Datenarchitekturen über digitale Werkzeuge und Workflows innerhalb der **Materials4Mobility**- und **DMD4Future**-Projekte.

Einer der Kernpunkte der Datenprozessierung ist ihre Speicherung. Soll ein Datensatz mit mehreren Instanzen geteilt werden, ist die Gewährleistung der Interoperabilität erstrangig. Ermöglicht wird dies durch den Einsatz von Standards oder semantischen Technologien.

Derzeit wird beispielsweise die Verwendung des DICONDE-Datenformats als generisches Format zur langfristigen verfahrens- und systemübergreifenden Speicherung und Dokumentation von Informationen erprobt. Der DICONDE-Standard kommt im Fraunhofer IFAM bei der Prüfung von Materialien, etwa bei der Röntgenfotoelektronenspektroskopie, zum Einsatz.

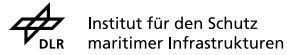
Im OntoTRANS-Projekt kommen aktuelle semantische Technologien multifunktional zum Einsatz. Ziel ist dabei die Etablierung von semantischer Interoperabilität bei der Entwicklung eines Produktes oder einer Dienstleistung. Zum Einsatz kommen dabei Triplestore- oder Resource-Description-Framework (RDF)-Modelle zur Speicherung der Daten und zum Abruf von Daten in Triples durch semantische Abfragen. Diese Triples kodieren eine Aussage über semantische Daten in Form von Subjekt-Prädikat-Objekt-Ausdrücken. Eine derartige Form des Forschungsdatenmanagements ermöglicht eine Einbettung und Kommunikation von Material- und Prozessdaten in unterschiedliche Nutzungs- und Anwendungsumgebungen vom Shopfloor bis zu Produktionsmanagementsystemen und ist übertragbar auf andere Forschungsbereiche.

Herausgeber

U Bremen Research Alliance e. V.; vertreten durch den Vorstand
Prof. Dr.-Ing. Bernd Scholz-Reiter (Vorsitzender)
Prof. Dr. Iris Pigeot (stellvertretende Vorsitzende)
Frauke Meyer
Oliver A. Dreher



Mitglieder der U Bremen Research Alliance:



www.uni-bremen.de/research-alliance

