

DAM KERNBEREICH DATENMANAGEMENT UND DIGITALISIERUNG

**ZUKÜNFTIGE HANDLUNGSFELDER ZUR OPTIMIERUNG DES DATENMANAGEMENTS UND
DER DIGITALISIERUNG IN DER DEUTSCHEN MEERESFORSCHUNG**

Stand September 2024

DAM-AG „Datenmanagement und Digitalisierung“

Entwurf

Inhalt

| | | |
|-----|-------------------------------------------------------------------------------------------------------|----|
| 1. | Vorwort..... | 3 |
| 2. | Hintergrund..... | 3 |
| 2 | Ausgangslage | 4 |
| 3 | Handlungsfelder..... | 7 |
| 3.1 | FAIRe Daten relevanter Forschungsinfrastrukturen durch den systematischen Ausbau der Datenflüsse..... | 8 |
| 3.2 | Stärkung der internationalen Anbindung..... | 9 |
| 3.3 | Verbesserung der Echtzeitdatenübertragung..... | 10 |
| 3.4 | Mariner „Interims-Datenraum“ für Projekte und Institute..... | 11 |
| 3.5 | Stärkere Nutzung von PANGAEA als zertifiziertes Repozitorium für FAIRe Meeresforschungsdaten | 12 |
| 3.6 | Förderung von Datenwissenschaften und der Anwendung moderner Analysemethoden..... | 13 |
| 4 | Referenzen | 15 |

1. Vorwort

Das vorliegende Dokument enthält Vorschläge für zukünftige Handlungsfelder im Bereich Datenmanagement und Digitalisierung, die im Zeitraum 2026–2030 von der deutschen Meeresforschung bearbeitet werden könnten. Voraussetzung dafür ist eine positive Evaluation der DAM sowie das fortwährende Engagement eines zukünftigen Vereins der deutschen Meeresforschung für das Thema „Datenmanagement und Digitalisierung“.

Das Dokument ist als ein weiterer Baustein auf dem Weg zu einer umfassenden Datenstrategie für die deutsche Meeresforschung zu verstehen. Es ergänzt und baut auf der DAM-Forschungsdatenleitlinie sowie dem Datenmanagementkonzept für die deutsche Meeresforschung auf.

In Kürze wird ein Prozessvorschlag folgen, der zur Definition von „essentiellen Meeresforschungsdaten“ aus Sicht der deutschen Meeresforschung führen soll. Ziel wird sein, die hier vorgeschlagenen Handlungsfelder mit dem Fokus auf essentiellen Meeresforschungsdaten zu präzisieren.

2. Hintergrund

Die systematische Erfassung, Speicherung, Verwaltung und Nutzung von Daten schafft die Grundlage für neue wissenschaftliche Erkenntnisse und technologische Entwicklungen, insbesondere bei interdisziplinären Fragestellungen. Standardisierte Datenformate und Protokolle ermöglichen die Integration von Daten aus heterogenen Quellen, was zu neuen Erkenntnissen und innovativen Lösungen für komplexe Herausforderungen führen kann.

Die Deutsche Allianz Meeresforschung (DAM) möchte durch die Bereitstellung umfassend beschriebener Meeres-Forschungsdaten einen wesentlichen Beitrag zu den Zielen der Nationalen Forschungsdateninfrastruktur (NFDI) leisten. Ziel ist es, ein tieferes Verständnis und eine effektivere Beobachtung der Meeresumwelt zu erreichen, indem die Anwendung moderner Analysemethoden wie Maschinelles Lernen und Künstliche Intelligenz (KI; engl. Artificial Intelligence, AI) ermöglicht werden. Die deutsche Meeresforschung verfügt bereits über viele Komponenten eines kollaborativen "Datenökosystems", um Forschungsdaten für Nutzer:innen aus Wissenschaft, Behörden, Wirtschaft und Öffentlichkeit frei zugänglich zu machen.

2 Ausgangslage

Die Deutsche Allianz Meeresforschung (DAM) koordiniert Forschungsdatenmanagement im Bereich Meeresforschung seit ihrer Gründung im Jahr 2019. Sie berücksichtigt dabei bestehende und sich in Entwicklung befindende Forschungsdateninfrastrukturen, um so Innovation und Datenwissenschaften zu fördern. Forschungsdatenmanagement wird als wesentlicher Bestandteil qualitätsorientierter und anschlussfähiger Forschung verstanden. Deshalb muss der Umgang mit Forschungsdaten sorgfältig geplant, dokumentiert und beschrieben werden. Eine Nachnutzungsmöglichkeit der Forschungsdaten muss, wann immer möglich, angestrebt werden. Forschungsdatenmanagement kann zudem zur systematischen Anwendung moderner Analysemethoden dienen, indem Daten "KI-ready" zur Verfügung gestellt werden.

Die Koordinationstätigkeit der DAM konzentriert sich bislang auf die umfassende Förderung der Bereitstellung möglichst FAIRer Daten der deutschen Meeresforschung. Zu diesem Zweck hat die DAM eine wichtige formale Basis geschaffen, indem eine DAM-Forschungsdatenleitlinie [2] erstellt und verabschiedet wurde. Forschungsdatenleitlinien dienen dem Kodex zur Sicherung der guten wissenschaftlichen Praxis der Deutschen Forschungsgemeinschaft [3] und konkretisieren dessen fachspezifische Umsetzung.

Durch die Verabschiedung der DAM-Forschungsdatenleitlinie haben sich die Mitgliedseinrichtungen (z.T. wiederholt) zur grundsätzlichen Bereitstellung FAIRer und offener Daten bekannt. Die Einhaltung der Forschungsdatenleitlinie und die Unterstützung der Wissenschaft entlang des Datenlebenszyklus liegen in der Verantwortung jeder DAM-Mitgliedseinrichtung bzw. ihren Angestellten. In einigen Mitgliedseinrichtungen sind institutionelle Forschungsdatenleitlinien erstellt worden, die spezifischer als die DAM-Forschungsdatenleitlinie sein können. Beispiele für spezifische Leitlinien sind die Forschungsdaten-Leitlinie des AWI, [4] GEOMAR [5] und des MARUM [6].

Auf infrastruktureller Ebene stehen zur Umsetzung der Forschungsdatenleitlinie Werkzeuge bereit, die von einzelnen DAM-Mitgliedseinrichtungen betrieben und zur gemeinsamen Nutzung durch alle Mitglieder angeboten werden. Die nötigen personellen Unterstützungsmöglichkeiten für die Wissenschaft zur Nutzung vorhandener Infrastrukturen sind heterogen. Die Aufgabe der DAM besteht darin, vorhandene Dateninfrastrukturen unter den DAM-Mitgliedern bekannt zu machen, die gemeinsame Nutzung zu fördern und die Weiterentwicklung unter Berücksichtigung der Bedarfe aller zu koordinieren.

Zusätzlich zur Forschungsdatenleitline wurde im Jahr 2022 ein erstes Datenmanagementkonzept für die marine Forschungslandschaft [7] erstellt und verabschiedet. Eines der zentralen Ziele des Konzeptes bestand darin, eine Vereinbarung zur Verknüpfung und Öffnung bewährter Forschungsdateninfrastrukturen einzelner DAM-Mitgliedseinrichtungen für die gesamte deutsche Meeresforschung zu treffen. Das vorliegende Dokument soll Handlungsfelder aufzeigen, die zur weiteren Optimierung des Datenmanagements und der Digitalisierung in der Meeresforschung beitragen. Dabei wird an den bereits eingeschlagenen Weg zu einer Harmonisierung des Managements von Forschungsdaten und einer zwischen den DAM-Mitgliedern abgestimmten Vorgehensweise angeknüpft.

Datenökosystem der DAM

Abbildung 1 zeigt das Datenökosystem der deutschen Meeresforschung, in dem Akteure Daten produzieren, anbieten, finden, nachnutzen, verarbeiten, archivieren und publizieren. Das Datenökosystem basiert auf zwei Initiativen: dem DataHub des Helmholtz-Forschungsbereichs „Erde und Umwelt“ und dem DAM-Kernbereich „Datenmanagement und Digitalisierung“. Beide arbeiten komplementär und abgestimmt zusammen.

Der DataHub harmonisiert das Datenmanagement über alle Helmholtz-Zentren des Fachbereichs „Erde und Umwelt“ hinweg und ermöglicht die Umsetzung der FAIR-Prinzipien. Bestehende Infrastrukturen werden zu einem verteilten Datenraum zusammengeführt.



Abbildung 1: Schematische Darstellung des Datenökosystems der Deutschen Meeresforschung

Der DAM-Kernbereich koordiniert die Nutzung der durch den DataHub bereitgestellten Infrastrukturen, insbesondere (aber nicht ausschließlich) für Metadaten und Daten im Zusammenhang mit Forschungsschiffen und DAM-Forschungsmissionen. Das Marine Data Portal (<https://marine-data.de/>) spielt eine zentrale Rolle bei der Bereitstellung von Metadaten und Daten der deutschen Meeresforschung. Es bietet einen zentralen Zugriff auf verteilte Repositorien und ermöglicht benutzerfreundlichen Zugang zu Daten. Das Portal fördert zudem die internationale Sichtbarkeit der Forschungsergebnisse durch die Visualisierung und Bündelung von Daten. Die FAIR-Bereitstellung erfordert enge Zusammenarbeit zwischen Wissenschaftler:innen, Datenmanager:innenn und Data Stewards. Sie definieren gemeinsam Standards und Datenflüsse, die zu Datenprodukten führen. Unterschiedliche Datentypen erfordern spezifische Konzepte für standardisierte Datenflüsse von der Beobachtung bis zum Archiv.

Das Portal wird bereits genutzt, um expeditionsbezogene Daten deutscher Forschungsschiffe sowie Daten von Langzeitobservatorien und themenspezifische Datenkomplikationen zu visualisieren. Weitere Infrastrukturkomponenten wie das Registry-System (AWI) und das Ocean Science Information System (OSIS, GEOMAR) unterstützen die Dokumentation der Forschungsdaten.

Diese Aktivitäten tragen zur Nationalen Forschungsdateninfrastruktur (NFDI) bei. Die DAM übernimmt eine koordinierende Funktion für die Meeresforschung in den NFDI-Konsortien für Biodiversitätsforschung (NFDI4Biodiversity) [8] und für Erdsystemforschung (NFDI4Earth) [9]. Über die Beteiligung an den Querstrukturen der NFDI [10], wie der Sektion Common Infrastructures oder Metadata, werden Standards und Architekturen mit der gesamten Gemeinschaft der Forschenden abgestimmt und geteilt.

Was in diesem „Datenökosystem“ bisher (in Teilen) fehlt, ist:

- A. die Ausweitung der Koordinationstätigkeiten im Bereich Datenmanagement auf weitere Forschungsinfrastrukturen (z.B. seegehende Großgeräte, Glider, Langzeitobservatorien)
- B. die Stärkung der internationalen Anbindung,
- C. der Ausbau von Infrastrukturen und Diensten für Echtzeit-Datenaustausch und Echtzeit-Datenvisualisierung an Bord und an Land. Die virtuelle Teilnahme von

(internationalen) Wissenschaftler:innen an Forschungsfahrten soll ermöglicht werden.

- D. die Möglichkeit zur Nutzung eines „Interims-Datenraums“, in dem die Wissenschaft während der Projektlaufzeit souverän Daten teilen, mit Metadaten anreichern, ggf. analysieren und Dienste erstellen kann bevor daraus FAIRer Datenpublikationen generiert werden. Es bestehen institutionelle sowie Open Source Infrastrukturlösungen für einen solchen „Interims-Datenraum“, in welchem die Wissenschaft ihre Datenmanagementaufgaben entlang des Datenlebenszyklus erfüllen kann. Die Notwendigkeit eines solchen „Interims-Datenraums“ nimmt mit zunehmender Digitalisierung und Interdisziplinarität der Forschung zu. Woran es mangelt, sind Datenmanager:innen bzw. Data Stewards innerhalb der Forschungsprojekte (oder Institute), die die Wissenschaft bei der Anwendung vorhandener Lösungen unterstützen und leiten, um Datenmanagement während der Projektlaufzeit für projektbezogene Zwecke so zu gestalten, dass bis zum Projektende wesentliche Datenmanagementaufgaben für FAIRer Daten bereits erfüllt sind. Ein Positivbeispiel für die Nutzung eines solchen „Interims-Datenraums“ ist das Projekt CONMAR der DAM-Forschungsmission sustainMare,
- E. ein Institutionen übergreifendes Konzept zur standardisierten und systematischen Bereitstellung von „long-tail“ Forschungsdaten nach den FAIR und Open Data Prinzipien, die „KI-ready“ sind,
- F. der Beleg für den Mehrwert vielfältiger und vielzähliger, FAIRer und offener Daten durch deren Wiederverwendung und Analyse unter Zuhilfenahme moderner Analysemethoden,

3 Handlungsfelder

Im Folgenden werden Handlungsfelder vorgestellt, die im Zeitraum 2026 bis 2030 und aufbauend auf dem etablierten Datenökosystem von der deutschen Meeresforschung bearbeitet werden könnten. Sie leiten sich von den (in Teilen) fehlenden Komponenten ab. Die Handlungsfelder umfassen den systematischen Ausbau von Datenflüssen, die Stärkung der internationalen Zusammenarbeit, die Verbesserung der Echtzeitdatenübertragung, die Einrichtung von Interims-Datenräumen, die verstärkte Nutzung von PANGAEA für die Publikation FAIRer Daten sowie die Förderung von Datenwissenschaften und modernen Analysemethoden. Diese Handlungsfelder sollen die Forschung unterstützen und den wissenschaftlichen Erkenntnisgewinn durch Datenmanagement und Digitalisierung optimieren.

3.1 FAIR Daten relevanter Forschungsinfrastrukturen durch den systematischen Ausbau der Datenflüsse

Hinweis: Als relevante Forschungsinfrastrukturen werden hier jene Infrastrukturen bezeichnet, die von einer breiten Meeresforschungsgemeinschaft genutzt werden oder ihnen Daten für ihre Forschung liefern. Darunter zählen, aber nicht ausschließlich Forschungsschiffe, seegehende Großgeräte, Langzeitobservatorien und Glider

Das Projekt „Unterwegs“-Forschungsdaten hat in enger Zusammenarbeit mit der DataHub-Initiative gezeigt, wie systematische und homogene Datenflüsse von der Beobachtung auf See bis zur Visualisierung von Daten im Datenportal Deutsche Meeresforschung für fest installierte Sensoren an Bord der großen deutschen Forschungsschiffe aufgebaut und betrieben werden können (z.B. [11]). Die Ausweitung dieser Aktivität auf mitgebrachte Sensoren sowie mittelgroße Schiffe ist Teil der aktuellen Projektphase (Stand September 2024). Erkenntnisse, Erfahrung und Infrastrukturkomponenten aus dem Projekt und der DataHub-Initiative können mit Unterstützung der wissenschaftlichen Gemeinschaft auf weitere Forschungsinfrastrukturen (z.B. seegehende Großgeräte, Langzeitobservatorien) übertragen werden und zur Optimierung vorhandener Datenflüsse beitragen. Damit soll das Ziel der benutzerfreundlichen Bereitstellungen FAIRer Daten deutscher Forschungsschiffe und weiterer Forschungsinfrastrukturen über das Datenportal Deutsche Meeresforschung erreicht werden. Die vorhandene Dateninfrastruktur wurde bereits so konzipiert, dass sie skaliert. In anderen Worten: Beim Aufbau der technischen Infrastruktur wurde bereits berücksichtigt, dass zukünftig größere Datenvolumina zu verarbeiten sein werden. Die Vorgehensweise zur Etablierung der Datenflüsse kann in zwei Schritten erfolgen:

1. Forschende und Forschungsinfrastrukturbetreibende definieren Bedarfe und Rahmenbedingungen für das Datenmanagement. Die Definition der Bedarfe und Rahmenbedingungen umfassen Aspekte wie die Rücksprache zu den tatsächlich zu behandelnden (Meta-)Daten, deren Qualitätskontrolle, Archivierung, Lizenzierung, Visualisierung, Komponenten zur Echtzeit-Übertragung etc., die wie bereits erwähnt mit den jeweiligen Gemeinschaften en Detail besprochen werden müssen.
2. Unter Nutzung möglichst vorhandener Infrastrukturen werden Datenflüsse so weit wie möglich automatisiert. Zunächst werden diese aber mit pragmatischen Ansätzen aufgebaut und anschließend sukzessive optimiert.

Die wesentlichen Voraussetzungen für die Umsetzbarkeit der Maßnahme sind die DataHub-Initiative, das Personal im Projekt „Unterwegs“-Forschungsdaten, das sich z.T.

für die hier genannte Maßnahme einbringen könnte, sowie die Bereitschaft, PANGAEA noch stärker als bisher für die Publikation FAIRer Daten zu nutzen.

3.2 Stärkung der internationalen Anbindung

Die deutsche Meeresforschung agiert in einem globalen Kontext und muss ihre Aktivitäten stärker mit internationalen Vertragsorganisationen, Regelwerken und Koordinierungsmechanismen wie der Weltorganisation für Meteorologie (WMO), der Zwischenstaatlichen Ozeanographischen Kommission (IOC) der UNESCO und dem Sekretariat der Klimarahmenkonvention der Vereinten Nationen (UNFCCC) abstimmen, um die Wirksamkeit ihrer Maßnahmen weiter zu erhöhen.

Neben den hochrangigen Leitungsgremien dieser Organisationen gibt es auf technischer Ebene verschiedene internationale Koordinierungsmechanismen, mit denen sich die DAM zusätzlich zu bestehenden Netzwerken wie dem Europäischen Meeresbeobachtungs- und Datennetzwerk (EMODnet), dem Copernicus-Programm und SeaDataNet befassen könnte. Dazu zählen das Global Ocean Observing System (GOOS), dessen Lenkungsausschuss und Beobachtungskoordinierungsgruppe, das Globale Klimabeobachtungssystem (GCOS) und der Ausschuss für Physik und Klima der Ozeanbeobachtung.

GOOS, ein internationales Programm zur Koordination der Beobachtung, Analyse und Vorhersage von Ozeanentwicklungen, wird von der IOC der UNESCO geleitet. Es sammelt über das Metadatenportal OCEANOBS der WMO/IOC verlässliche Daten zu physikalischen, chemischen und biologischen Eigenschaften der Ozeane, die für Klimaforschung, Küstenschutz und Wettervorhersagen entscheidend sind. Diese Daten stammen aus In-situ-Messungen (wie Bojen und Schiffen) und Satellitenbeobachtungen und werden integriert und analysiert, um ein umfassendes Bild der ozeanischen Bedingungen zu liefern.

Im Ozean-Daten- und Informationssystem (ODIS)-Katalog des International Oceanographic Data and Information Exchange (IODE) der IOC sind auch die Datenportale des Bundesamtes für Seeschifffahrt und Hydrographie (BSH), der Wasserstraßen- und Schifffahrtsverwaltung (WSV) sowie das Marine Data Portal der DAM aufgelistet.

Die Hauptnutzer der GOOS-Daten sind Wissenschaftler:innen, Regierungen und die Industrie, die diese Informationen für Forschung, politische Entscheidungen und wirtschaftliche Aktivitäten nutzen.

Bisher übermittelt die deutsche Meeresforschung ihre Daten an GOOS teils unkoordiniert und auf Initiative einzelner Wissenschaftler:innen. Dennoch gibt es nationale Bestrebungen, diese Prozesse zu optimieren. Institutionen wie das BSH und Netzwerke wie die DAM arbeiten zunehmend an einer besseren Datenintegration und Koordination auf nationaler Ebene. Es wird daher vorgeschlagen, diesen Prozess weiter zu stärken, um GOOS-kompatible Daten zentral zu erfassen und zu übermitteln.

Der s.g. GOOS-National Focal Point Deutschlands ist am BSH angesiedelt. Es wird vorgeschlagen, GOOS-kompatible Metadaten aus PANGAEA regelmäßig über den BSH-Databroker an GOOS bzw. Copernicus zu liefern. Der BSH-Databroker erfüllt bereits heute die Funktion „Unterwegs“-Forschungsdaten aus PANGAEA an Copernicus zu liefern. Der Databroker überführt PANGAEA-Datenmodelle in das Copernicus-Datenmodell, um sie für die entsprechenden Systeme maschinenlesbar bereitzustellen. Diese Funktion könnte auf weitere Daten und „Empfänger“ ausgeweitet werden.

3.3 Verbesserung der Echtzeitdatenübertragung

Für die virtuelle Teilnahme an Expeditionen bedarf es der Datenübertragung von Bord an Land in (nahezu) Echtzeit. Die Echtzeitdatenübertragung ermöglicht die intelligente Verschneidung von Datensätzen aus unterschiedlichsten Quellen, um quasi live eine möglichst vierdimensionale Betrachtung der marinen Umwelt auch an Bord und an Land zu erhalten, die dann dynamisch in Entscheidungsprozesse an Bord einfließen können.

Hier wird vorgeschlagen, die Möglichkeiten zur Echtzeitdatenübertragung von relevanten Forschungsinfrastrukturen - zunächst von deutschen Forschungsschiffen - zum Land und vice versa erheblich zu steigern.

Durch (ggf. zunächst projektweise) Finanzierung des satellitenbasierten Datenübertragungssystems Starlink, das ausreichende Bandbreite im Internetzugang auf deutschen Forschungsschiffen bieten würde, könnte die Echtzeitdatenübertragung und Visualisierung von Forschungsdaten nutzerfreundlich implementiert werden. Daten- und Technikexpert:innen haben im Rahmen der Schiffsausfahrt EMB 304 bereits belegt, dass eine Echtzeitdatenübertragung ans Land, inklusive der Visualisierung der Forschungsdaten, für den Zweck einer virtuellen Cruise-Teilnahme bereits heute von Bord des Forschungsschiffs Elisabeth Mann Borgese (EMB) möglich ist. Das notwendige technische Fachwissen für die prototypische Umsetzung kann von Meereswissenschaftler:innen nicht erwartet werden. Die nutzerfreundliche und standardmäßige Implementierung dieser Technologie kann nur bei ausreichender Internetbandbreite umgesetzt werden.

3.4 Mariner „Interims-Datenraum“ für Projekte und Institute

„Interims-Datenräume“ erlauben das Teilen von und gemeinsame Arbeiten mit Forschungsdaten in geschützten Bereichen (Arbeitsgruppen, Projekte, Verbünde, etc.). Der Bedarf an "Interims-Datenräumen" steigt zunehmend u.a. mit steigender Anzahl an Projektpartnern und Interdisziplinarität der Projekte. Mit dem *Ocean Science Information System (OSIS)* des GEOMAR steht der Gemeinschaft ein auf Expeditionsdaten spezialisierter „Interims-Datenraum“ im Rahmen von DAM-Missionen zur Verfügung, der das Teilen von Expeditions-, Experiment- und Simulations(meta)daten in Gruppen erlaubt. Zudem stehen über OSIS Expeditionsmetadaten der deutschen Forschungsflotte für alle offen zur Nachnutzung zur Verfügung.

Für den Forschungsverbund CONMAR (CONcept for conventional MArine munition Remediation in the German North and Baltic Sea), aus der DAM Forschungsmision sustainMare, wurde ein „Interims-Datenraum“ mit der am Thünen-Institut mit entwickelten Open Source Software GeoNode (<https://geonode.org/>) aufgebaut. Hier werden Daten unabhängig von einzelnen Ereignissen, wie Expeditionen, zur gemeinsamen Nutzung bereitgestellt. Die Daten können hochgeladen und durch einen integrierten GeoServer als OGC-Dienste mit den Projektpartnern geteilt werden. Das Projekt CONMAR hat damit eine Blaupause für einen „Interims-Datenraum“ mit GeoNode erarbeitet, der ergänzend zu OSIS auch die Arbeit mit Daten und Datenprodukten erlaubt, um Projektpartner beim Datenmanagement zu unterstützen. Zudem fungiert GeoNode am Thünen-Institut als institutionelles Datenmanagementsystem.

In ihren spezifischen Ausprägungen und Nutzungsszenarien ergänzen sich OSIS und GeoNode hervorragend, weshalb eine Kopplung dieser Systeme angestrebt wird. Beispielsweise kann OSIS über Standardschnittstellen mit GeoNode derart verknüpft werden, dass über den Datenfluss Registry -> DSHIP -> OSIS -> GeoNode Schiffsexpeditionen der deutschen Forschungsflotte mit den Datenprodukten der in GeoNode verfügbaren Projektdaten gemeinsam als Dienste verfügbar gemacht werden. DSHIP ist ein Datenbanksystem, das an Metadaten und Daten bestimmter Sensoren an Bord der Schiffe speichert. Diese zukünftige nutzerfreundliche Verknüpfung von OSIS und GeoNode wird dann für die gesamte deutsche Meeresforschung als Bestandteil des marinen Datenökosystems nutzbar sein und kann ein kollaboratives Qualitätsmanagement ermöglichen.

Die notwendige Expertise zur Unterstützung der Wissenschaft über einen gemeinsamen „Interims-Datenraum“ entlang des gesamten Datenlebenszyklus ist in der Gemeinschaft vorhanden. Hier wird vorgeschlagen und angeboten, die Blaupause aus dem Projekt

CONMAR mit OSIS zu koppeln und für Projekte oder institutionelle Nutzung in der DAM auszuweiten, um die Wissenschaft möglichst optimal während der Projektlaufzeit bei ihren Datenmanagementaufgaben zu unterstützen. Die Voraussetzung dafür ist, dass Mitarbeiter:innen mit der entsprechenden Kompetenz und Fähigkeit Forschungsprojekte (oder gar Institute) begleiten.

3.5 Stärkere Nutzung von PANGAEA als zertifiziertes Repository für FAIRer Meeresforschungsdaten

Die DAM-Forschungsdatenleitlinie empfiehlt die Publikation von Meeresforschungsdaten in qualifizierten, zertifizierten Repositorien. Die Repositorien PANGAEA und das World Data Center for Climate (WDCC) können wesentliche Anteile an Meeresforschungsdaten inhaltlich abdecken. Die Herausforderung für PANGAEA (sowie vermutlich für das WDCC) besteht darin, dass die personellen Kapazitäten nicht ausreichen, um die große Zahl an Daten der deutschen Meeresforschung in angemessener Zeit zu archivieren und zu publizieren. Um dieser Herausforderung zu begegnen, hat PANGAEA das sogenannte *Frontend-Backend-Prinzip* eingeführt. Das Prinzip besteht darin, dass regelmäßige Datenlieferanten (z.B. Forschungseinrichtungen oder Projekte) Personal einstellen, das von und mit PANGAEA Kurator:innen darin trainiert wird, die PANGAEA Infrastruktur für das Datenmanagement auf institutioneller oder Projektebene zu nutzen. So wird der kostspielige Aufbau und Betrieb eines institutionellen Repositoriums für Einrichtungen obsolet, die bisher kein Repository betreiben, aber die Notwendigkeit dafür sehen. Die Heterogenität der Metadaten, die mit einer Vielzahl institutionell betriebener Repositorien (bzw. Archiven) einhergeht, wird zugleich minimiert. Zur Konsolidierung von Repositorien wird daher vorgeschlagen, die Ausweitung des *Frontend-Backend-Prinzip* von PANGAEA innerhalb der DAM anzubieten. Bestehendes und ggf. zukünftiges Datenmanagementpersonal könnte mit Zustimmung der PANGAEA-betreibenden Einrichtungen (AWI, MARUM) darin trainiert werden, die PANGAEA-Infrastruktur für das institutionelle oder projektbezogene Datenmanagement zu nutzen.

Langjährig betriebene Repositorien, die weiterhin betrieben werden sollen, sollten zur Einhaltung der FAIR-Prinzipien ertüchtigt werden und - wie bereits im Datenportal Deutsche Meeresforschung umgesetzt - geharvestet werden. Es gilt jedoch, dass jedes Repository und die dazugehörige Einrichtung differenziert betrachtet werden müssen. Die Ertüchtigung zur Einhaltung der FAIR-Prinzipien, wie es im zertifizierten Repository PANGAEA möglich ist, kann mehr oder weniger aufwändig sein. Unter Umständen ist der Betrieb eines institutionellen Repositoriums als "Interims-Datenraum" sinnvoll, währenddessen die Publikation und Langzeitarchivierung wie oben beschrieben systematisch durch die Nutzung der PANGAEA-Infrastruktur erfolgen

kann. Inwiefern ein ähnliches Modell mit dem WDCC etabliert werden könnte, bleibt zu prüfen.

3.6 Förderung von Datenwissenschaften und der Anwendung moderner Analysemethoden

Unterschiedliche Institutionen, Forschungseinrichtungen oder Initiativen bieten Programme zur Förderung der Anwendung und Entwicklung von KI in der Wissenschaft. Als Beispiel sei das Helmholtz AI Voucher Programm (<https://voucher-system.helmholtz.ai/>) genannt. Es bietet Forschenden Zugang zu KI-Expertise, Infrastruktur und Ressourcen durch sogenannte „Vouchers“. Diese können für die Entwicklung neuer KI-Algorithmen, die Analyse großer Datensätze oder die Optimierung bestehender Methoden genutzt werden.

Das Programm richtet sich an Wissenschaftler:innen aus verschiedenen Disziplinen, von Gesundheitswissenschaften über Umweltforschung bis hin zu Materialwissenschaften und Ingenieurwesen. Helmholtz AI fördert mit weiteren Maßnahmen wie „AI Consultants“ (Beratung rund um AI) und AI Projektförderung (jährlicher Aufruf im Volumen von bis zu 400 k€) die Zusammenarbeit zwischen Forschungszentren und erleichtert den Austausch von Wissen und Technologien. Zusätzlich bietet Helmholtz AI Schulungen und Workshops an, um Forscher:innen im Umgang mit KI-Tools und -Techniken zu schulen.

Insgesamt zielt Helmholtz AI darauf ab, die Anwendung von KI in der Wissenschaft zu beschleunigen und innovative Forschungsprojekte zu unterstützen.

Ein weiteres Beispiel auf institutioneller Ebene ist die Data Science Unit (DSU) am GEOMAR. Die DSU besteht aus einem zentralen Team von Informatiker:innen und Geowissenschaftler:innen mit vielfältigen wissenschaftlichen Hintergründen, die durch Data Science (DS) Forschende in den GEOMAR Forschungsbereichen als assoziierte Mitglieder ergänzt werden. Durch die vielfältige kreative Zusammenarbeit herrscht ein "Startup-Spirit", der weitere assoziierte Mitglieder anlockt. Klar formuliertes Ziel der DSU ist die Unterstützung der Wissenschaftler:innen am GEOMAR. Dazu vernetzt die DSU DS-bezogene Themen und Aktivitäten im Zentrum und ist Ansprechpartner für alle DS-bezogenen Anfragen von und nach innen und außen, wie zum Beispiel zu Helmholtz AI.

Die DSU fördert den Austausch und die Erweiterung von DS-Fachwissen mit Forschenden, die in allen GEOMAR-Forschungsbereichen mit unterschiedlichen Datentypen arbeiten. Sie unterstützt die Forschenden bei der Entwicklung von DS-Methoden zur effektiven Erforschung der komplexen Meeresumwelt und des Systems

Erde. Darüber hinaus entwickelt die DSU Data Science Methoden für die Meeresforschung und wendet diese gemeinsam mit den Forschenden an heterogenen Daten des GEOMAR und seiner Partnerinstitute an. Die DSU bietet den Forschenden am GEOMAR Support und Training zu Data Science Methoden und fungiert als Partner bei der Beantragung von Drittmitteln für Data Science Projekte, sowie bei der Publikation von wissenschaftlichen Ergebnissen, die mithilfe von Data Science Methoden erforscht werden.

Hier wird vorgeschlagen, die beiden obigen Ansätze zu kombinieren und eine DAM-Data Science Unit einzurichten. Durch die Stärkung dieser Einheit mit weiterem, virtuell gepooltem Personal, das sowohl neu als auch bestehend und dezentral organisiert sein könnte, könnte die DSU am GEOMAR ihre Services für die gesamte deutsche Meeresforschung anbieten. Dieser Vorschlag versteht sich als Kooperationsangebot an interessierte Einrichtungen, die frei entscheiden können, in welchem Umfang sie sich beteiligen möchten. Die „DAM-Data Science Unit“ würde somit als zentrale Anlaufstelle fungieren, die den Austausch von Data Science-Expertise und -Ressourcen fördert, ohne dass dabei eine Verpflichtung zur Teilnahme entsteht. Die Handlungsfelder 3.1, 3.3 und 3.5 bieten in diesem Zusammenhang besonders gute Anknüpfungspunkte, um den Mehrwert durch Datenmanagement zu unterstreichen und neue wissenschaftliche Erkenntnisse durch die Anwendung moderner Analysemethoden zu fördern.

Der Vorschlag soll Forschungseinrichtungen unterstützen, die Data Science und KI verstärkt nutzen möchten, indem er eine offene, flexible Zusammenarbeit ermöglicht, die den individuellen Bedürfnissen und Kapazitäten der Einrichtungen gerecht wird.

4 Referenzen

- [1] Wilkinson, M.D., Dumontier, M., Jan Aalbersberg, I. et al. Addendum: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 6, 6 (2019). <https://doi.org/10.1038/s41597-019-0009-6>
- [2] Wiemer, G., Agostinho Nogueira, M., Bauer, K., Becker, M., Brix, H., Droste, C., Feenders, C., Feistel, S., Freiherr von Lukas, U., Frickenhaus, S., Gerdes, J., Glöckner, F. O., Hemmelskamp-Pfeiffer, B., Jurasiczki, G., Kern, S., Kleeberg, U., Kösters, F., Liebers-Helbig, D., Lorenz, S., ... Voss, D. (2021). DAM-Forschungsdatenleitlinie. [Zenodo](#). <https://doi.org/10.5281/zenodo.8341227>
- [3] Deutsche Forschungsgemeinschaft. (2019). Guidelines for Safeguarding Good Research Practice. Code of Conduct. <https://doi.org/10.5281/zenodo.3923602>
- [4] AWI Directorate (2020): Forschungsdaten-Leitlinie des Alfred-Wegener-Instituts, Helmholtz-Zentrum für Polar- und Meeresforschung <https://epic.awi.de/id/eprint/52227/>
- [5] GEOMAR Direktorium, (2022) Forschungsdaten-Leitlinie. GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel. [doi: 10.3289/data_policy_V1_2022](https://doi.org/10.3289/data_policy_V1_2022).
- [6] Glöckner, Frank Oliver, Kucera, Michal, Pälike, Heiko, Zabel, Matthias, & Schulz, Michael. (2021). Research Data Policy at MARUM – Center for Marine Environmental Sciences of the University of Bremen (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5643724>
- [7] Wiemer, G., Agostinho Nogueira, M., Bauer, K., Becker, M., Binder, K., Brix, H., Droste, C., Feenders, C., Feistel, S., Freiherr von Lukas, U., Frickenhaus, S., Gerdes, J., Glöckner, F. O., Hemmelskamp-Pfeiffer, B., Jurasiczki, G., Kern, S., Kleeberg, U., Kösters, F., Liebers-Helbig, D., ... Voss, D. (2022). Forschungsdatenmanagementkonzept für die deutsche Meeresforschung (Version 1). [Zenodo](#). <https://doi.org/10.5281/zenodo.8232832>
- [8] Weber, J., Ebert, B., Diepenbroek, M., Kostadinov, I., & Glöckner, F. O. (2021). NFDI4BioDiversity - NFDI-Konsortium für Biodiversitäts-, Ökologische und Umweltdaten. Bausteine Forschungsdatenmanagement, (2), 98–109. <https://doi.org/10.17192/bfdm.2021.2.8334>

[9] Bernard, Lars, Braesicke, Peter, Bertelmann, Roland, Frickenhaus, Stephan, Gödde, Hildegard, Keßler, Carsten, Lorenz, Sören, Mahecha, Miguel, Marschall, Horst, Hezel, Dominik, Nagel, Wolfgang E., Reichstein, Markus, Sester, Monika, Thiemann, Hannes, Weiland, Claus, Wytzisk-Arens, Andreas, & NFDI Consortium Earth System Sciences. (2021). NFDI Consortium Earth System Sciences - Proposal 2020 revised (Revision 1 from 2021-10-01). [Zenodo. https://doi.org/10.5281/zenodo.5718944](https://doi.org/10.5281/zenodo.5718944)

[10] Ebert, Barbara, Fluck, Juliane, Glöckner, Frank Oliver, Koepler, Oliver, Miller, Bernhard, Schmitt, Robert, Schrade, Torsten, Stegle, Oliver, Steinbeck, Christoph, von Suchodoletz, Dirk, Wagemann, Kurt, Knebes, Jennifer, Kraft, Sophie, Seitz-Moskaliuk, Hendrik, Sure-Vetter, York, & Wössner, Elena. (2021). NFDI Cross-cutting Topics Workshop Report. [Zenodo. https://doi.org/10.5281/zenodo.4593770](https://doi.org/10.5281/zenodo.4593770)

[11] Kopte R, Becker M, Fischer T, Brandt P, Krahmann G, Betz M, Faber C, Winter C, Karstensen J and Wiemer G (2024) FAIR ADCP data with OSADCP: a workflow to process ocean current data from vessel-mounted ADCPs. *Front. Mar. Sci.* 11:1425086. doi: [10.3389/fmars.2024.1425086](https://doi.org/10.3389/fmars.2024.1425086)