DEUTSCHE
ALLIANZ
MEERESFORSCHUNG

DAM

Konzeptionierung einer Forschungsdateninfrastruktur durch die Deutsche Allianz Meeresforschung

15.04.2021
DAM-AG „Datenmanagement und Digitalisierung“

1. Motivation

Die deutsche Meeresforschung benötigt eine dezentrale, leistungsstarke und nachhaltige Forschungsdateninfrastruktur, um Information und Wissen zu sichern und Forschungsdaten für Nutzer*innen aus Wissenschaft, Behörden, Wirtschaft und Öffentlichkeit frei zugänglich zu machen und so Innovationen zu fördern. Die Deutsche Allianz Meeresforschung (DAM) will diese digitale Transformation der Meeresforschung koordinieren.

Um eine vollständige und nachvollziehbare Dokumentation der vielfältigen Forschungsdaten zu erreichen, sollten Wissenschaftler*innen in allen Phasen des Datenlebenszyklus (Projektplanung, Datenaufnahme, -aufbereitung, -analyse, -archivierung, -publikation) durch Expert*innen im Datenmanagement unterstützt werden. Die bestehenden Voraussetzungen, Organisationsstrukturen und Infrastrukturen für eine solche Unterstützung sind an den DAM-Mitgliedseinrichtungen sehr heterogen.

Das vorliegende Konzept skizziert den Weg zu einer Homogenisierung des Managements von Forschungsdaten und einer zwischen den DAM-Mitgliedern abgestimmten Forschungsdateninfrastruktur, die integriert ist in die Nationale Forschungsdateninfrastruktur (NFDI). Die NFDI soll die nationalen Datenbestände von Wissenschaft und Forschung systematisch erschließen, nachhaltig sichern und zugänglich machen sowie (inter-)national mit der European Open Science Cloud (EOSC) vernetzen. Sie verfolgt somit kongruente Ziele zur DAM im Kernbereich „Datenmanagement und Digitalisierung“. Die NFDI wird seit dem 01. Oktober 2020 in einem wissenschaftsgetriebenen Prozess als vernetzte Struktur eigeninitiativ agierender Konsortien aufgebaut. Die DAM übernimmt eine koordinierende Funktion für die Meeresforschung im bewilligten NFDI-Konsortium für Biodiversitätsforschung (NFDI4BioDiversity) und plant einen entsprechenden Beitrag zum beantragten NFDI-Konsortium für Erdsystemforschung (NFDI4Earth). Damit unterstützt der DAM-Kernbereich “Datenmanagement und Digitalisierung” den Aufbau einer übergreifenden, dezentralen und nachhaltig betriebenen Forschungsdateninfrastruktur für die nationale Meeresforschung auch in einem übergeordneten Rahmen.

2. Ziele

Die Möglichkeiten für die (Nach-)nutzung qualitätsgesicherter Meeresforschungsdaten sollen bis zur Evaluation der DAM deutlich verbessert werden. Im Fokus stehen dabei der offene Zugang zu qualitätsgesicherten Daten der Meeresforschung nach den FAIR-Prinzipien (auffindbar, zugänglich, interoperabel und wiederverwendbar, aus engl. findable, accessible, interoperable und re-usable). Grundlage dafür bilden bestehende Repositorien und Dateninfrastrukturen, die verstärkt an nationale, europäische und internationale Infrastrukturen angebunden werden sollen. Die Systematik in der Erhebung, Aufbereitung, Auswertung und Archivierung von Meeresforschungsdaten soll grundlegend vereinheitlicht und verbessert werden, damit Daten und Datenprodukte verschiedenen Ursprungs aggregiert und wiederverwendet werden können. Richtlinien, die dies effektiv unterstützen, sollen gemeinsam mit den Mitgliedern erarbeitet und verabschiedet werden. Der Zugang zu Daten und Datenprodukten der Meeresforschung sowie deren Visualisierung soll über das Datenportal Deutsche Meeresforschung (Prototyp: https://marine-data.de) ermöglicht werden. Unterstützt werden diese Entwicklungen durch Methoden der künstlichen Intelligenz (KI). Beispiele wären die automatisierte Datenqualitätskontrolle sowie die Erschließung und Analyse von großen Datenmengen mittels „deep-learning“ Technologien.

Der DAM-Kernbereich "Datenmanagement und Digitalisierung" schafft dabei die Synergien mit den Aktivitäten der Helmholtz-Gemeinschaft (HGF), die im Rahmen der HGF-Digitalisierungsstrategie unternommen werden. Besonders hervorzuheben sind die Maßnahmen für eine gemeinsame Daten- und Informationsinfrastruktur im Forschungsbereich “Erde und Umwelt” der HGF. Dabei werden historisch getrennte Datenrepositorien der atmosphärischen, marinen und terrestrischen Forschung samt ihren Querschnittsfeldern in der Klima- und Biodiversitätsforschung zu einer offenen, vernetzten Informationsinfrastruktur, dem sogenannten DataHub, zusammengeführt. Der DataHub ist ein maßgeblicher Beitrag der HGF zur NFDI. Der marin ausgerichtete MareHub wird daher ein infrastrukturelles Kernelement im Datenmanagementkonzept der DAM bilden. Bei der Konzeption und dem Aufbau einer integrierten und nachhaltigen Dateninfrastruktur für die Meeresforschung werden die existierenden und entstehenden Infrastrukturen innerhalb der Meeres- und Erdsystemforschung bereits gemeinschaftlich zwischen der HGF und der DAM gedacht und umgesetzt.

3. Maßnahmen

Zur Erreichung der Ziele sollen Maßnahmen auf allen Ebenen des Wissenschaftssystems umgesetzt werden. Zusagen an die Politik zu mehr offenen und FAIRen Daten müssen von Senior- wie Juniorwissenschaftler*innen erfüllt werden können. Datenmanager*innen, Kurator*innen und Softwareingenieur*innen sollen sich untereinander sowie mit der Wissenschaft stark vernetzen und eng zusammenarbeiten, um den Umgang mit und den Zugang zu Daten zu vereinheitlichen. Wissenschaftler*innen soll im Umgang mit Daten nach abgestimmten und zum Teil neu zu entwickelnden Standards stärker und systematischer

unterstützt werden, als das bisher möglich ist. Ein Teil der notwendigen Strukturen ist bereits vorhanden; ihre breite Nutzung ist gegenwärtig noch nicht optimal möglich. Bestehende Infrastrukturen müssen in Teilen adaptiert und vernetzt werden, um z.B. Zugang zu dezentralen Datenbeständen zu ermöglichen oder eine Mandantenfähigkeit bestehender Systeme herzustellen. Kernaufgaben in der Entwicklung innovativer Lösungen für Datenflüsse, KI-basierte Qualitätskontrolle und Auswertung sowie die Datenspeicherung müssen erkannt und benannt werden, bevor sie innerhalb der DAM arbeitsteilig umgesetzt werden können.

Im Folgenden wird eine mögliche Entwicklung zu einer gemeinsamen Forschungsdateninfrastruktur unter dem Dach der DAM inhaltlich und zeitlich skizziert (s.a. Abbildung 1).

Datenmanagementkonzept für die Meeresforschungslandschaft

Die Mitgliedseinrichtungen der DAM erarbeiten mit Unterstützung der Geschäftsstelle ein integriertes und verlässliches *Datenmanagementkonzept für die Forschungslandschaft,* um den offenen und einheitlichen Zugang zu Daten nach den FAIR-Prinzipien zu koordinieren. In einem ersten Schritt hin zu einem *Datenmanagementkonzept für die Forschungslandschaft* wurden konkrete Aufgaben identifiziert:

- Die Kartierung existierender Forschungsdatenleitlinien zur Ermittlung von Gemeinsamkeiten und Gegensätzen mit dem Ziel, eine "DAM-Forschungsdatenleitline" zu erstellen.
- Die Erstellung einer Liste qualifizierter Archive, die den Anforderungen der DAM-Forschungsdatenleitlinie entsprechen.
- Die Kartierung existierender und zu integrierender Infrastrukturkomponenten und Dienste.
- Die Erstellung einer Liste mit Aufgaben und Lösungen, die zur Integration der identifizierten Infrastrukturkomponenten und Dienste notwendig sind.
- Die Identifikation notwendiger Entwicklungsmaßnahmen für neue, zukunftsfähige Komponenten.
- Die Identifikation notwendiger Standardisierung von Messgeräten und Sensoren zur verbesserten systematischen Datenerfassung auf deutschen Forschungsschiffen und Observatorien.
- Die Kartierung vorhandene personeller Kompetenzen und deren zeitliche Ressourcen.
- Die Definition konkreter Aufgaben(pakete) und Meilensteine und die Verteilung entsprechend der Kompetenzen innerhalb der Mitgliedseinrichtungen.
- Die Intensivierung der Vernetzung mit der Fachgemeinschaft sowie die Verstärkung von Schnittstellen zu anderen Akteur*innen nationaler und internationaler Datenmanagementaktivitäten.

Bereits vorhandene Kartierungen, Informationen und Maßnahmen der Forschungsgemeinschaft werden berücksichtigt.

Der verlässliche Betrieb der Forschungsdateninfrastruktur muss technisch gewährleistet sein. Hierfür soll das *Datenmanagementkonzept für die Forschungslandschaft* ein Betriebs- und Finanzierungskonzept enthalten, das mit den Leitungen und Verwaltungsabteilungen der Mitgliedseinrichtungen, dem DAM-Vorstand und DAM-Verwaltungsrat abgestimmt ist.

DAM-Forschungsdatenleitlinie

Als Basis für ein gemeinsames Verständnis im Umgang mit Daten der Meeresforschung und Grundstein für ein Datenmanagementkonzept für die Meeresforschung wird eine für die Mitglieder verbindliche DAM-Forschungsdatenleitlinie benötigt, die innerhalb der AG „Datenmanagement und Digitalisierung“ im Einklang mit bestehenden Forschungsdatenleitlinien der Mitgliedseinrichtungen erstellt und in der Mitgliederversammlung beschlossen werden. Eine DAM-Forschungsdatenleitline regelt Grundsätze für den einheitlichen Umgang mit Forschungsdaten. Dies betrifft:

- den offenen Zugang zu Daten und die Einhaltung der FAIR-Prinzipien,
- die Anerkennung von qualitätsgesicherten Datenpublikationen als wissenschaftliche Leistung,
- Kriterien der Qualitätssicherung,
- die Verantwortlichkeiten für die Einhaltung der Kriterien und den sorgfältigen Umgang mit Daten,
- Fristen für die Archivierung von Primärdaten und die Dauer von Moratorien,
- die Dauer der Archivierung,
- die Bereitschaft zur langfristigen Vernetzung der Infrastrukturen.

Die DAM-Forschungsdatenleitlinie berücksichtigt den Kodex zur Sicherung der guten wissenschaftlichen Praxis der Deutschen Forschungsgemeinschaft und konkretisiert dessen fachspezifische Umsetzung. Zur Steigerung der Reproduzierbarkeit und Datenqualität sollen im Rahmen der Umsetzung der FAIR-Prinzipien, die Provenienz der Daten bei der Datenerhebung, -sammlung, -nutzung und -aufbewahrung besonders berücksichtigt werden. Zuwendungsempfänger*innen sollten bestärkt und unterstützt werden, im Projekt gewonnene Daten spätestens nach Projektabschluss frei zugänglich zu machen. Datenpublikationen nach den FAIR-Prinzipien sollten bei der Begutachtung von Förderanträgen als wissenschaftliche Leistung berücksichtigt werden.

Organisationsstruktur und Governance

Ebenfalls als Bestandteil des *Datenmanagementkonzeptes für die Forschungslandschaft*, soll die AG „Datenmanagement und Digitalisierung“ eine Organisations- und Lenkungsstruktur

entwerfen, die von der Mitgliederversammlung beschlossen werden soll. Entscheidungsprozesse und Rollenverteilungen für die kontinuierliche Entwicklung der Infrastruktur und die Umsetzung von Maßnahmen werden darin geregelt. Die Organisationsstruktur soll die Vielzahl an nationalen und internationalen Aktivitäten der DAM-Mitglieder im Bereich Forschungsdatenmanagement integrieren und eine koordinierte und abgestimmte Vorgehensweise im nationalen und internationalen Kontext ermöglichen.

Datenzugriff und Visualisierung: Datenportal Deutsche Meeresforschung

Das Datenportal Deutsche Meeresforschung soll der zentrale Zugriffspunkt für dezentral archivierte Daten, Datenprodukte und Dienste der Meeresforschung werden. Nutzer*innen sollen sich zunächst via text- und kartenbasierter Suche auf dem Portal Deutsche Meeresforschung einen schnellen Überblick über vorhandene, frei verfügbare und qualitätskontrollierte Daten verschaffen und auf sie zugreifen können. Die Suche via API (von englisch application programming interface, wörtlich ‚Anwendungsprogrammierschnittstelle') sollte perspektivisch auch ermöglicht werden. Integrierte Visualisierungswerkzeuge helfen, die Daten zu sichten.
Der Fokus liegt dabei auf zukünftig erhobene Daten und solche, die bereits heute notwendige Standards erfüllen und zitierbar in Repositorien wie PANGAEA oder CERA vorliegen. Öffentlich finanzierte Datenbestände, die mit Inkrafttreten der DAM-Forschungsdatenleitlinie älter als zwei Jahre sind, sollen per Beschluss der DAM-Mitgliederversammlung von Moratorien befreit und unter der jeweils aktuellen Lizenz-Version (CC-BY 4.0 Stand 03/2021) veröffentlicht werden.

Bestehende und bereits publizierte Daten - sogenannte historische Daten - müssen mit Metadaten oder Prozessierungsschritten nachgepflegt oder aufbereitet werden, um sie auffindbar zu machen und sie im Portal Deutsche Meeresforschung visualisieren zu können. Teilweise müssen sie nachträglich mit Lizenzen versehen werden, um die Nachnutzung zu ermöglichen, ohne sich die Erlaubnis der Datenproduzent*innen einzuholen. Daten, wie z.B. Videoaufnahmen auf Magnetbändern oder handschriftlich notierte Daten aus Laborbüchern, sollen digitalisiert werden. Die Digitalisierung solcher Datenbestände soll im Datenmanagementkonzept berücksichtigt werden.

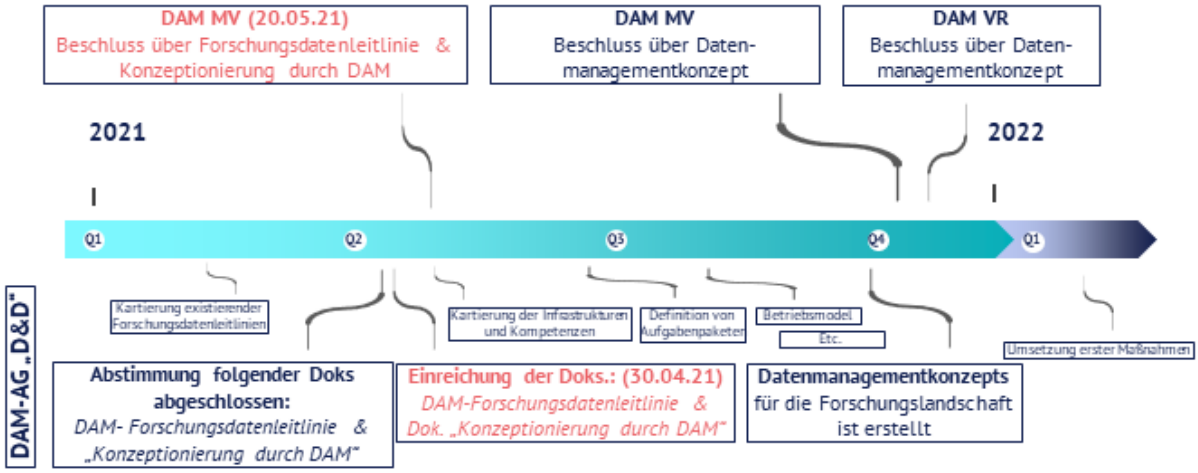

Abbildung 1: Zeitlicher Verlauf für die Erstellung eines mit den Mitgliedern abgestimmten Datenmanagementkonzeptes für die Forschungslandschaft. MV: Mitgliederversammlung, VR: Verwaltungsrat,

Die notwendigen Schritte im Umgang mit Daten, die zur schnellen Auffindbarkeit und zum einfachen Zugriff über das Portal führen, sollen durch die AG „Datenmanagement und Digitalisierung" im Konzept benannt werden. Außerdem sollen Maßnahmen beschrieben werden, die die systematische und enge Zusammenarbeit im gesamten Datenlebenszyklus zwischen Wissenschaftler*innen, Datenmanager*innen und Kurator*innen der Mitgliedseinrichtungen sicherstellen. Aus der Zusammenarbeit mit der Wissenschaft sollen deren Bedarfe an das Datenportal Deutsche Meeresforschung hervorgehen, die im fortlaufenden Prozess implementiert werden müssen.