

ABSCHLUSSBERICHT IM PROJEKT “UNTERWEGS”-FORSCHUNGSDATEN 1

Projekt 03F0880 “Unterwegs“-Forschungsdaten

Deutsche Allianz Meeresforschung

Kernbereich „Datenmanagement und Digitalisierung“

Projektlaufzeit: 01.01.2021 – 31.12.2022

Version: 1

Datum der letzten Bearbeitung: 12.07.2023

Inhalt

TEIL I: Kurzbericht.....	3
TEIL II: Eingehende Darstellung.....	6
Arbeitspaket 0: Koordination des Gesamtvorhabens	6
Arbeitspaket 1: Sensor Monitoring / Datenqualitätssicherung und Bereitstellung / Secondary Quality Control / Automatisierung der Prozesse	8
1.1 AG-Bathymetrie.....	8
1.2 AG-Ozeanographie.....	10
Arbeitspaket 2: Datenkuration.....	14
Arbeitspaket 3: Automatisierte Anbindung an internationale Portale	17
Arbeitspaket 4: Datenportal Deutsche Meeresforschung	20
Arbeitspaket 5: Technische und inhaltliche Integration und Betreuung von wissenschaftlichen Geräten und Schiffssensorik für ActionLog und Massendatenspeicher der Deutschen Forschungsflotte	22
Angrenzende Aktivitäten.....	25
Unterstützung der Atmosphärenforschung	25
Zusammenarbeit mit dem NFDI4Earth Pilotprojekt "German Marine Seismic Data"	25
Unterstützung der DAM-Forschungsmissionen.....	26
Fazit und Strategischer Mehrwert.....	27
Referenzen.....	27

TEIL I: Kurzbericht

In diesem Projekt geht es um die systematische Erfassung, Übertragung vom Schiff an Land, Qualitätssicherung und Veröffentlichung von "Unterwegs"-Forschungsdaten nach den FAIR- und "Open"-Prinzipien (Auffindbar, Zugänglich, Interoperabel, Wiederverwendbar). "Open" steht für die offene Zugänglichkeit. Als „Unterwegs“-Forschungsdaten werden jene Daten bezeichnet, die von bordeigenen, meereswissenschaftlich relevanten Sensoren gemessen und nicht im Interesse der Primärforschung an Bord der Forschungsschiffe gesammelt werden. Darüber hinaus werden im Projekt Bathymetrie Daten generell betrachtet, insbesondere für den Schwerpunkt Publikation und Visualisierung. Der Fokus liegt auf den Forschungsschiffen MARIA S. MERIAN, METEOR, POLARSTERN, SONNE und EUGEN SEIBOLD. Zu den Geräten und Sensoren, die behandelt werden, gehört das CTD (Conductivity, Temperature and Depth)-System, Bathymetrie-Signale (inklusive Schallgeschwindigkeitsmessungen), der akustische Strömungsmesser (ADCP), der Thermosalinograph (TSG), die Ferrybox und bio-optische Sensoren.

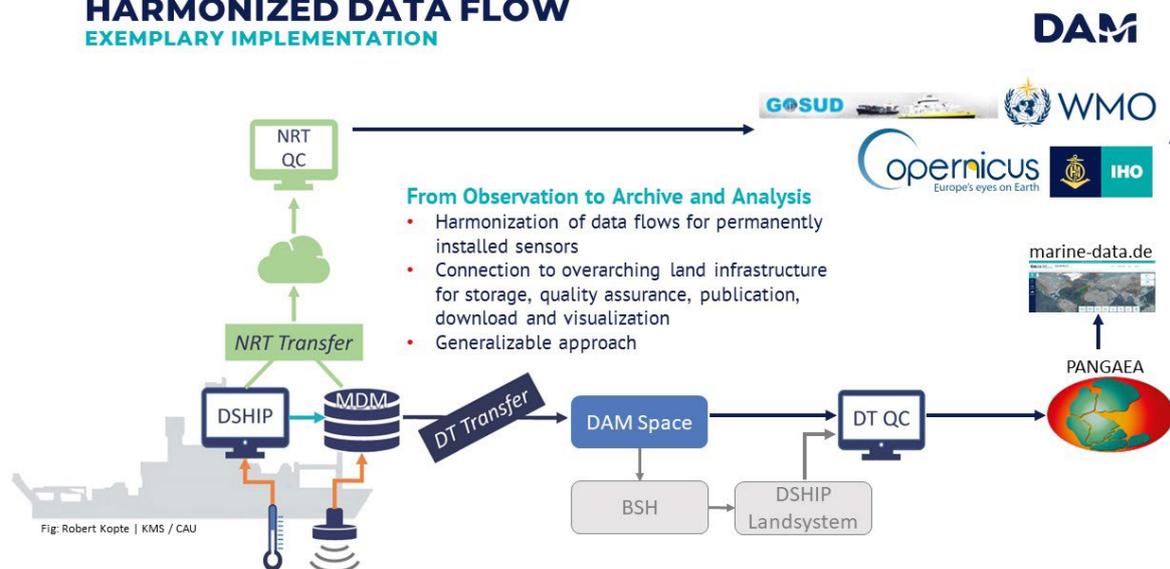
Für Daten, die von Sensoren der o.g. Geräte gesammelt werden, wurden in der Pilotphase (09/2019 – 12/2020) Arbeitsabläufe, Qualitätskontrollen und Datenflüsse vom Moment und Ort der Datenaufnahme bis zur Publikation im Datenarchiv PANGAEA entwickelt. Die Umsetzung sowie die Erstellung von Standards (s.g. „Standard Operation Procedures“ (SOPs)) begann in der Phase I (2021-2022). Zur Veröffentlichung der SOPs wurde eine „Community“ auf Zenodo gegründet. Die SOPs stehen hier zur Verfügung:

<https://zenodo.org/communities/dam/?page=1&size=20>

*HINWEIS zu Zenodo: Zenodo wurde von Forscher*innen entwickelt, um sicherzustellen, dass jeder an der offenen Wissenschaft teilnehmen kann. Das Projekt OpenAIRE, Vorreiter der Open-Access- und Open-Data-Bewegung in Europa, wurde von der Europäischen Kommission beauftragt, ihre im Entstehen begriffene Open-Data-Politik zu unterstützen, indem es ein gemeinsames Repository für von der Kommission finanzierte Forschung bereitstellt. CERN, ein OpenAIRE-Partner und Pionier in Sachen Open Source, Open Access und Open Data, stellte diese Fähigkeit zur Verfügung und Zenodo wurde im Mai 2013 gestartet.*

Der etablierte Datenfluss ist in Abbildung 1 dargestellt.

HARMONIZED DATA FLOW EXEMPLARY IMPLEMENTATION



1

Abbildung 1: Standardisierter Datenfluss vom Sensor bis ins Archiv (PANGAEA) und Datenportal Deutsche Meeresforschung. Erstellt von Robert Kopte/KMS, CAU

Im Dezember 2020 fand eine Expedition auf der MARIA S. MERIAN statt, die besonders der Einführung und dem Test einer neuen Dateninfrastruktur an Bord diente (Becker et. al 2022). Diese neue Dateninfrastruktur ist seit Dezember 2021 auf allen großen deutschen Forschungsschiffen implementiert und im regulären Dauerbetrieb. Die Datenaufnahme und Übertragung an Land wurde auf allen vier großen deutschen Forschungsschiffen (FS) vereinheitlicht. Der Datenfluss ist auf den genannten Forschungsschiffen (Abb. 1) identisch.

Die Wissenschaft bzw. Expeditions-Fahrtleiter:innen werden im Rahmen von Fahrtvorbereitungstreffen der Leitstelle Deutsche Forschungsschiffe durch Projektbeteiligte über alle Neuerungen an Bord sowie den Datenfluss informiert. Da die vereinfachte und vereinheitlichte Übertragung der Daten an das Landsystem ebenfalls die Primärforschungsdaten betrifft, profitiert die Wissenschaft, auch unabhängig von ihrem Bezug zu "Unterwegs"-Daten, von der neuen Infrastruktur. Von einer Implementierung auf der EUGEN SEIBOLD wurde abgesehen, da das System überdimensioniert wäre.

Mit der Anlandung der Daten gelangen die Daten in den „DAM-Space“ (siehe Abbildung 1). Eine Kopie wird am BSH archiviert. Geräteverantwortliche haben im „DAM-Space“ Zugriff auf Daten, die während der Expedition im Massendatenmanagement-System (MDM) auflaufen. "Unterwegs"-Daten werden von Primärforschungsdaten getrennt, qualitätskontrolliert und mit Metadaten angereichert, um die FAIRness der Daten zu maximieren. Die Qualitätskontrolle und

Publikation von "Unterwegs"-Daten wird in Zusammenarbeit mit den Kurator:innen bei PANGAEA und in Rücksprache mit den Expeditions-Fahrtleiter:innen systematisch unternommen. Vom Zeitpunkt des Zugriffs auf "Unterwegs"-Daten im „DAM-Space“ durch die Geräteverantwortlichen bis zur vollständigen Publikation in PANGAEA vergehen durchschnittlich etwa sechs Monate.

Um die Auffindbarkeit und den Zugang zu qualitätskontrollierten und publizierten "Unterwegs"-Forschungsdaten und weiteren bereits vorliegenden Daten der Deutschen Meeresforschung zu optimieren, wurde das Portal Deutsche Meeresforschung weiter überarbeitet und verbessert. Weitere „Nutzerworkshops“ wurden durchgeführt, um die Nutzerfreundlichkeit des Portals zu testen und zu verbessern. Das Portal ist unter <https://marine-data.de/> erreichbar und nutzbar.

Die Anbindung an internationale Portale wurde ausgebaut. Tägliche TSG-Daten werden zu *Global Ocean Surface Underway Data* (GOSUD) und anschließend zur World Meteorological Organization (WMO) weitergeleitet und tragen auf diese Weise zu Wettervorhersagen bei. Bathymetrische Daten, die in PANGAEA publiziert werden, werden systematisch in den Viewer des "Data Centre for Digital Bathymetry" (DCDB) der "International Hydrographic Organization" (IHO) übermittelt und tragen so zum "Nippon Foundation-GEBCO Seabed 2030 Project" bei.

Durch eine enge Verzahnung der MareHub-Aktivitäten mit den Aktivitäten im Pilotprojekt "Unterwegs"-Forschungsdaten leistet die DAM einen maßgeblichen Beitrag zum koordinierten Aufbau einer Forschungsdateninfrastruktur für die Meeresforschung. Die Zusammenarbeit und Abstimmung von Inhalten und Maßnahmen und deren Durchführung mit MareHub-Aktivitäten liefern weiterhin vorbildlich.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Hinweis: Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

TEIL II: Eingehende Darstellung

Zur Bearbeitung der Arbeitspakete und Aufgaben des Pilotprojektes (09/2019-12/2020) wurden vier Arbeitsgruppen (AGs) gebildet. Diese Arbeitsgruppen blieben in der Projektphase I (2021-2022) bestehen. Jede AG war an mehreren Arbeitspaketen beteiligt. Die Abhängigkeiten zwischen den Arbeitsgruppen und Arbeitspaketen, die an verschiedenen Einrichtungen im Verbund bearbeitet werden, sind komplex und im Kapitel 4.9 des Leitanspruchs (Vernetzung der Verbundpartner und Strukturplan) dargelegt (siehe auch Abb. 1). Zur Vereinfachung der Berichterstattung wird im Folgenden zu jedem Arbeitspaket aus der Perspektive der hauptverantwortlichen Arbeitsgruppe berichtet, wenngleich die Zusammenarbeit unter den AGs zu den Arbeitspaketen nötig ist:

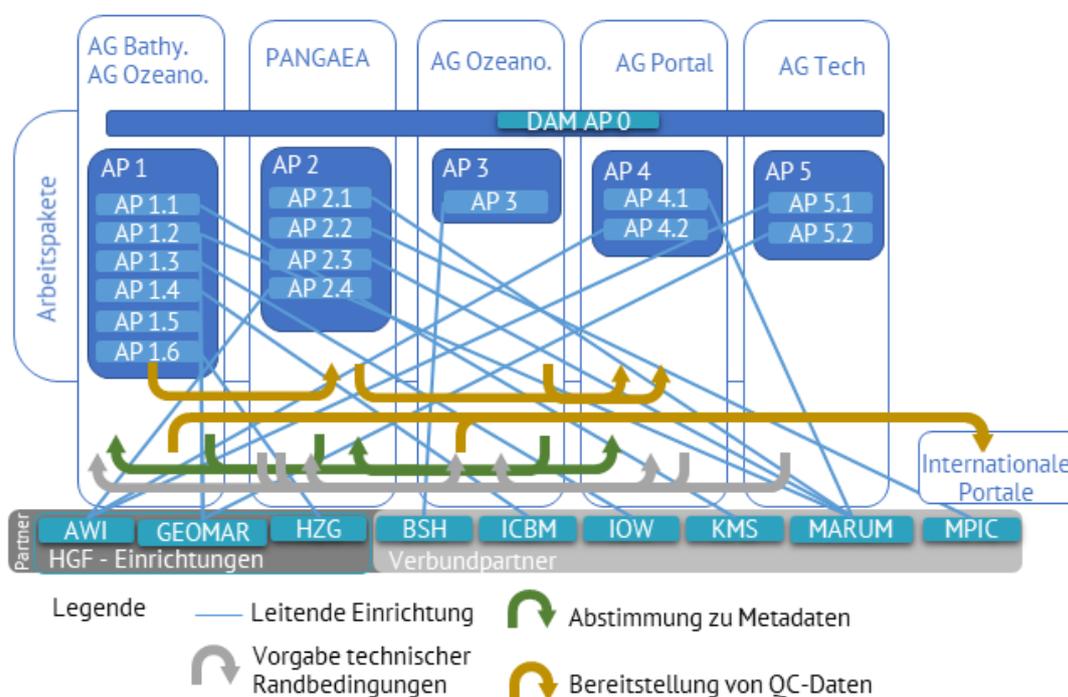


Abbildung 2: Entspricht Abbildung 4 aus dem Leitanspruch zur Darstellung der starken Vernetzung der Partner zur Bearbeitung der Arbeitspakete in Arbeitsgruppen.

Arbeitspaket 0: Koordination des Gesamtvorhabens

Das AP 0 umfasst die Koordination des Gesamtvorhabens, d.h. der koordinierten Arbeitsweise der Arbeitsgruppen untereinander sowie an den Schnittstellen der Arbeitspakete. Am 21. Februar 2021 fand das virtuelle Kick-Off Treffen aller Projektpartner statt. Der Fokus lag auf der zusammenfassenden Präsentation bereits erreichter Ziele aus der Pilotphase (09/2019 -

12/2020) und der anstehenden Teilvorhaben. Die Diskussion zielte auf eine vertiefte Vernetzung und Koordination eines abgestimmten Vorgehens ab. Am 18.01.2022 fand ein Projekttreffen online statt.

Die Organisation, Dokumentation und Protokollierung regelmäßiger virtueller Projekttreffen aller Projektpartner:innen und weiterer Beteiligten wurde über das Management-Tool "Confluence" erreicht, das vom AWI bereitgestellt wird. Die regelmäßigen Treffen (insgesamt 30 Treffen in der Projektlaufzeit) dienen der Koordination der Passfähigkeit der Arbeitspakete untereinander. Jede:r Projektpartner:in berichtete zum aktuellen Stand des jeweiligen Teilprojektes. So konnte die Zusammenarbeit in der stark vernetzten Struktur (siehe Abbildung 2) bei geografisch verteilten Mitarbeiter*innen sichergestellt werden.

An diesen Treffen nahmen regelmäßig Vertreter*innen der MareHub-Initiative der marinen Zentren des Fachbereichs Erde und Umwelt der Helmholtz-Gemeinschaft teil. Ebenfalls nahmen Vertreter*innen des DAM-Projektes "Unterwegs"-Daten an Workshops und Infoveranstaltungen des MareHub und DataHub teil. Steuernde und leitende Akteur*innen der NFDI-Konsortien NFDI4Earth und NFDI4Biodiversity waren über den Inhalt und den Fortschritt des Projektes informiert. Die Kommunikationskanäle, welche die Basis für die Anschlussfähigkeit an die NFDI-Konsortien sind, wurden genutzt. Das Projekt "Unterwegs"-Forschungsdaten und die MareHub-Initiative bilden zusammen die Basis für ein koordiniertes Datenmanagement zur Unterstützung der deutschen Meeresforschung und der internationalen Zusammenarbeit.

Die Daten und die Infrastruktur dieser gemeinsamen Anstrengungen tragen zur NFDI und EOSC bei. Die Beziehungen zwischen diesen Initiativen sowie der European Open Science Cloud sind in der folgenden Infografik dargestellt:

GERMAN MARINE RESEARCH DATA

Joint Management und Infrastruktur

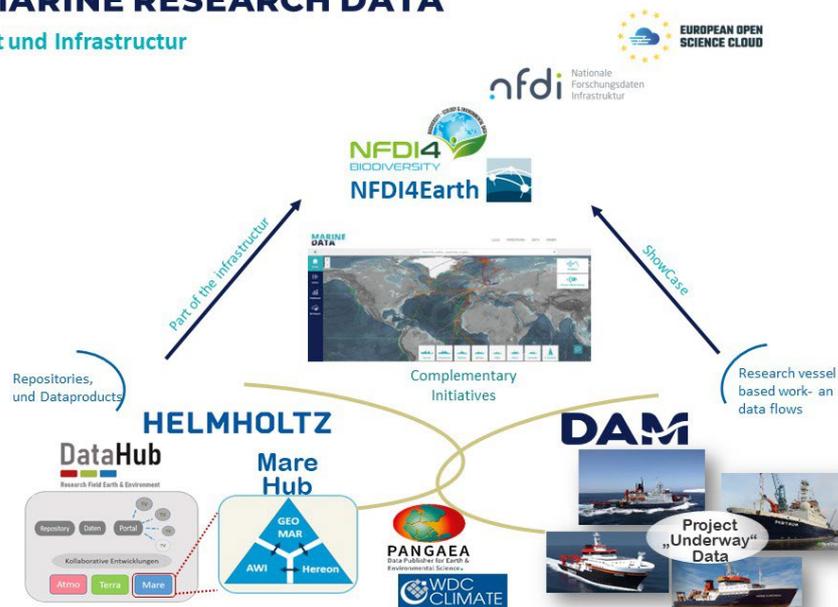


Abbildung 3: Infografik zu den Beziehungen zwischen dem DataHub/ MareHub, dem Projekt "Unterwegs"-Forschungsdaten, der NFDI sowie der European Open Science Cloud.

Das DAM-Projekt "Unterwegs"-Daten und insbesondere die Aktivitäten der AG-Portal (siehe Arbeitspaket 4) wurde im Rahmen einer Vernetzungsaktivität Vertreter*innen der Marinen Dateninfrastruktur Deutschland (MDI.de) präsentiert. Es fanden im Jahr 2021 mehrere Treffen statt. Ziel war es, Kooperationsmöglichkeiten zu erörtern und umzusetzen. S.g. Web Map Services (WMS), z.B. der Marine Dateninfrastruktur Deutschland (MDI.de), können im Portal Deutsche Meeresforschung durch Nutzer*innen implementiert werden.

Zum Projektfortschritt wurde in den DAM-Gremien "offener Vorstand" und "Mitgliederversammlung" sowie in den DAM-Kommunikationskanälen "DAM-Newsletter" und "DAM-Infobrief" berichtet. Das Projekt wurde beim fünften, sechsten und siebten Data Science Symposium vorgestellt, das im Rotationsverfahren von einer der Einrichtungen AWI, GEOMAR und Hereon veranstaltet wird. Das Projekt wurde des Weiteren am 02.02.2022 im Rahmen der Statuskonferenz Deutsche Forschungsschiffe präsentiert.

Im Rahmen von Fahrtvorbereitungstreffen der Leitstelle Deutsche Forschungsschiffe wurden Fahrtleiter*innen anstehender Expeditionen über das Projekt informiert, um Unterstützung gebeten und über Implikationen im Zusammenhang mit der neuen Dateninfrastruktur an Bord der großen deutschen Forschungsschiffe benachrichtigt.

Arbeitspaket 1: Sensor Monitoring / Datenqualitätssicherung und Bereitstellung / Secondary Quality Control / Automatisierung der Prozesse

Das Arbeitspaket 1 wurde federführend von den AGs Bathymetrie und Ozeanographie bearbeitet. Die folgende Berichterstattung erfolgt aus der Perspektive dieser AGs. Die AG-Technik, die AG-Portal und die Aktivitäten innerhalb PANGAEA waren an vielen Meilensteinen beteiligt.

1.1 AG-Bathymetrie

Das in der ersten Pilotphase entwickelte Konzept, (Meta-)Daten von Fächerecholot-Systemen und verschiedenen Bathymetrien-Dateiformaten in einem integrierten und skalierbaren Datenfluss für die Wissenschaft nachnutzbar (FAIR) in PANGAEA zu publizieren, wurde weiter harmonisiert. Gleichzeitig werden hierbei mittels des PANGAEA Bathymetrie Web Map Service (WMS) diese (Meta-)Daten im Portal Deutsche Meeresforschung bereitgestellt. Auf Basis des erprobten PANGAEA-Kurationskonzeptes für Bathymetrie-Daten wurde mit der Übertragung dieses Konzeptes auf Daten von Sedimentecholot-Systemen begonnen (weitere Informationen unter: https://wiki.pangaea.de/wiki/Sediment_echosounder). Die unter

<https://wiki.pangaea.de/wiki/Bathymetry> öffentlich zugänglichen, dokumentierten Standard Operating Procedures für die Aufbereitung und Publikation nach den FAIR Prinzipien von Bathymetrie-Daten in PANGAEA wurden weiter überarbeitet.

Insgesamt wurden im Berichtszeitraum 87 Bathymetrie-Datensätze von insgesamt 57 Expeditionen nach den erarbeiteten Kurationsstandards in PANGAEA publiziert. Diese Datensätze beinhalten 82 Fächerecholot-Rohdaten, davon 14 Fächerecholot-Datensätze der Wassersäule. 49 Datensätze stammen aus dem „historischen“ Datenbestand des BSH und wurden unter direkter Einbeziehung der verantwortlichen Wissenschaftler*innen archiviert. Außerdem wurden 17 Sedimentecholot-Datensätze archiviert. Zusätzlich wurden weitere 30 Datensätze nach den erarbeiteten Kurationsstandards bei PANGAEA archiviert, welche von den verantwortlichen Wissenschaftler*innen publiziert wurden.

Insgesamt wurden im gesamten Projektverlauf weitere 71 (totale Anzahl 170, inklusive Datensätze aus der Anschubphase) großvolumige Fächerecholot-Datensätze (Rohdaten) im „DAM-Space“ (Siehe Abbildung 1) ausgelesen. Dies ermöglicht eine harmonisierte Datenablage (+ Metadaten Extraktion) für PANGAEA und bildet gleichzeitig die Metadaten- und Datengrundlage für die anschließende Visualisierung im Viewer des Portals Deutsche Meeresforschung (WMS). Die Datenaufbereitungs-Software innerhalb des „DAM Space“ wurde weiterentwickelt und ist nun auf weitere „historische“, „Unterwegs“- und „nicht direkt projektbezogene“ Fächerecholot-Datenformate anwendbar. Für die Visualisierung wurde u.a. das PANGAEA-Werkzeug „pangaeapy“, ein Python-basiertes Modul für eine Kommandozeilen-basierte PANGAEA (Meta-)Daten Extraktion, adaptiert und erweitert (Huber et al. 2020).

Gemeinsam mit Projektpartnern aus dem MareHub Projekt wurden neue Datenflüsse konzipiert, welche die in PANGAEA publizierten, prozessierten und qualitätskontrollierten Datensätze in das Portal Deutsche Meeresforschung überführen und im „Viewer“ visualisieren. Es wurden 60 prozessierte Datensätze im „DAM-Space“ aufbereitet, in den WMS eingespeist und somit im Viewer des Portals Deutsche Meeresforschung sichtbar gemacht. Im Berichtszeitraum wurden neben diesen prozessierten Datensätzen alle im „DAM-Space“ aufbereiteten Rohdaten-Publikationen ebenfalls in den WMS integriert und um weitere Metadaten erweitert. Die im Jahr 2020 aufgebaute Datenbank für Vektordaten (Tracklines, d.h. Polylinien und Abdeckungs-Polygone des Fächerecholotes) wurde verwaltet und bildet neben dem GEOSERVER eine technische Grundlage für den bestehenden WMS-Dienst. Über den WMS wird die technische Anschlussfähigkeit an internationale Portale gewährleistet. Das Data Centre for Digital Bathymetry (DCDB) der International Hydrographic Organization (IHO), welches weltweit eines der größten Sammlungen an ozeanischen Tiefenmessungen bereitstellt und zusätzlich die Bathymetrie aus anderen Repositorien über Webdienste zeigt, hat den PANGAEA Bathymetrie WMS bereits integriert. Darüber hinaus werden die Bathymetrie-Datendienste auch von dem EU-Projekt iAtlantic in einem projektspezifischen Datenportal eingebunden. Eine anwenderorientierte Dokumentation über den PANGAEA Bathymetrie WMS wurde erstellt: https://wiki.pangaea.de/wiki/PANGAEA_Bathymetry_Web_Map_Services.

In Zusammenarbeit mit der AG-Technik (AP 5) wurde ein Arbeitsablauf entwickelt, um hochaufgelöste Bathymetrie-Karten und Karten, die die weltweite Fächerecholot-Datenabdeckung zeigen, in DSHIP auf den großen deutschen Forschungsschiffen zur Verfügung zu stellen. Ziel ist es, der Expeditions-Fahrtleitung sowie der Nautik alle nötigen Informationen über kartierte und nicht-kartierte Bereiche des Meeresbodens an Bord bereitzustellen. Mit dieser Information können die Verantwortlichen die Schiffsrouten so wählen, dass sie nicht kartierte Bereiche des Meeresbodens, z.B. auf Transitstrecken, kartieren. Erste Ansätze wurden bereits auf der MSM70, MSM88/2 und der MSM97/2 evaluiert. Die Kartensätze sind erstellt und wurden bereits in der DSHIP Demoversion am GEOMAR erfolgreich getestet. Lösungsvorschläge für die Integration der Karten in DSHIP wurden mit der Firma WERUM diskutiert. Anschließend wurde ein Testdatensatz an Bord der MARIA S. MERIAN für einen zweimonatigen Testlauf bereitgestellt.

Um auch prozessiert-qualitätskontrollierte "Unterwegs"-Bathymetrie-Daten der internationalen Gemeinschaft zur Verfügung zu stellen, wurden Gespräche mit dem "Seabed 2030 Atlantic and Indian Oceans Regional Center" geführt, um 1) Arbeitsabläufe anzugleichen oder gemeinsame Standards zu verwenden und 2), um sicher zu stellen, dass prozessierte Daten möglichst schnell und effektiv für das weltumfassende GEBCO - The General Bathymetric Chart of the Oceans Datenprodukt (eine globale Karte des Meeresbodens) genutzt werden können (<https://www.gebco.net/>).

Es wurde sich auf gemeinsame Prozessierungsschritte für Unterwegs-Bathymetrie-Daten und die anschließende Einbindung in GMRT (Global Multi-Resolution Topography Data Synthesis) geeinigt. GMRT ist eine Zusammenstellung von prozessierten Multibeamdaten heterogener Auflösung, die vom GMRT-Team überprüft, verarbeitet und gerastert werden. GMRT ist in der Ozean-Basiskarte auf Google Earth (seit Juni 2011) und in der GEBCO 2022 Grid enthalten. Darüber hinaus wurden derzeitige und künftige Entwicklungen für die Publikation, generelle Auffindbarkeit von Bathymetrie-Daten sowie dessen Nachnutzung mit verschiedenen nationalen und internationalen Partnern abgestimmt.

1.2 AG-Ozeanographie

In der Arbeitsgruppe Ozeanographie wurden im Berichtszeitraum die Aufgaben für die Geräte ADCP, bio-optische Sensoren, CTD, Ferrybox und TSG auch in Bezug auf einheitliche Formate und die gemeinsame Nutzung von Tools weiter durchgeführt und verbessert. Das betrifft sowohl das Sensor-Monitoring und die Echtzeit (Near Real Time, NRT)-Datenübertragung an internationale Portale als auch die Publikation hochwertiger, qualitätskontrollierter Daten bei PANGAEA für ADCP, bio-optische Sensoren, Ferrybox und TSG. Für die Publikation auf PANGAEA wurden sowohl Standards zur Kennzeichnung der Qualitätslevels durch die Einführung von Flags umgesetzt, als auch ein im Grundgerüst einheitlicher Bericht über die Datenprozessierung ("Data Processing Report") entwickelt.

Zusätzlich zu 13 ADCP-Datensätzen, welche innerhalb der Projektlaufzeit durch wissenschaftliche Gruppen an Bord der MERIAN, METEOR oder SONNE erhoben wurden, konnten mit Hilfe des DAM-Projekts "Unterwegs"-Daten im selben Zeitraum weitere 50 ADCP-Datensätze gewonnen werden. Von diesen 50 zusätzlichen Datensätzen wurden bis zum Ende des Berichtszeitraums 44 hinsichtlich ihrer Qualität evaluiert und bearbeitet. Innerhalb der Projektlaufzeit wurden insgesamt 28 ADCP-Datensätze bei PANGAEA veröffentlicht, weitere 10 der in der Projektlaufzeit erhobenen Datensätze wurden im 1. Quartal 2023 veröffentlicht und 3 weitere sind derzeit in Planung. Im ähnlichen Umfang wurden TSG-Daten erhoben und qualitätskontrolliert veröffentlicht. Hierbei wurden zusätzlich Salzproben erhoben und für die Validierung der Daten gemessen. Insgesamt sind 35 TSG-Datensätze im Berichtszeitraum bei PANGAEA veröffentlicht worden. Des Weiteren sind 2 CTD-Datensätze im Berichtszeitraum veröffentlicht worden, 2 weitere Datensätze sind aktuell in Kuration. Es ist ein Datensatz für bio-optische Sensoren veröffentlicht worden. Für die Ferrybox ist ein erster Datensatz erstellt und zur Veröffentlichung eingereicht worden.

Zusätzlich wurden in der Arbeitsgruppe in 2021 die etablierten Datenflüsse für die Forschungsschiffe MERIAN, METEOR und SONNE mit den Abläufen auf dem Forschungsschiff POLARSTERN harmonisiert. Es wurden die relevanten Abläufe (Erstellung von Stationslisten, Übertragung der Daten) auf die S/Y EUGEN SEIBOLD übertragen. Die Parameter Salzgehalt, Wassertemperatur, Lufttemperatur, Windgeschwindigkeit, Windrichtung, Chlorophyll-a-Fluoreszenz und Sauerstoff-Sättigung werden nun in 5-Minuten-Auflösung per Satellit von S/Y EUGEN SEIBOLD an surveydata.hereon.de gesendet, wo sie öffentlich einsehbar sind.

(Beispiellink:

[http://surveydata.hereon.de/index.cgi?seite=plot_diagram;par_id=666994830;von=:bis=:referenzpunkt=ende;zeitintervall=7D;ymin=:ymax=:menue=1;breite=2232;hoehe=200;yauto=2;fonts=:layout=Default_GD;punktgroesse=-1;qcolors=0\).](http://surveydata.hereon.de/index.cgi?seite=plot_diagram;par_id=666994830;von=:bis=:referenzpunkt=ende;zeitintervall=7D;ymin=:ymax=:menue=1;breite=2232;hoehe=200;yauto=2;fonts=:layout=Default_GD;punktgroesse=-1;qcolors=0).)

Eine weitere NRT-Übertragung erfolgt stündlich via E-Mail an ingest@marine-data.de. Sensoren wurden unter <https://sensor.awi.de> registriert.

Unter der Federführung des IOW wurde ein Fahrtantrag für das mittelgroße Forschungsschiff ELISABETH-MANN-BORGESE (EMB) gestellt, bewilligt und Ende 2022 durchgeführt. Die Fahrt diente der Ermittlung von Arbeitsschritten, die für die Übertragung der auf großen Forschungsschiffen etablierten Abläufe auf mittelgroße Forschungsschiffe nötig sind, um die Phase II des Projektes (2023-2025) vorzubereiten. Wenngleich davon auszugehen ist, dass jedes mittelgroße Schiff teilweise individuelle Anpassungen benötigen wird, sollte die Fahrt mit der EMB als Blaupause für die Übertragung etablierter Abläufe genutzt werden.

Bezüglich schiffsgestützter ADCP-Messungen an Bord der großen deutschen Forschungsschiffe wurden die Methoden zur Durchführung eines standardisierten Arbeitsablaufs weiterentwickelt und implementiert. Der Datentransfer, die Methoden zur Qualitätskontrolle und die

Vorbereitung der Datenprodukte zur Veröffentlichung wurden automatisiert. Die bei PANGAEA veröffentlichten ADCP-Datenprodukte beruhen auf diesem standardisierten Arbeitsablauf. Zur Near-Realtime Übertragung von ADCP-Monitoringdaten und zu der Qualitätskontrolle wurden im „DAM-Space“ neue Applikationen entwickelt. Hinsichtlich der Darstellung komplexer 3D-Strömungsdaten im Datenportal Deutsche Meeresforschung (marine-data.de) wurde ebenfalls ein Arbeitsablauf standardisiert, der die Limitation der Portaltechnik auf 2D-Grafiken kompensiert. Den wissenschaftlichen User*innen stehen jetzt Darstellungen der Vertikalprofile der ADCP-Daten in der Sidebar des Portals zur Verfügung. Die Bilddateien der entsprechenden Plots werden bei der Prozessierung der ADCP-Daten automatisch erzeugt und sind im Portal mit den Datenpunkten des Cruise Tracks verknüpft. In Zusammenarbeit mit der MareHUB AG Ocean Obs. wurden für alle sechs ADCPs der drei LDF-betriebenen Schiffe SOPs erstellt und bei Zenodo mit DOI veröffentlicht. Diese SOPs enthalten alle erforderlichen Meta-Informationen und beschreiben den Gesamtdatenfluss (NRT und DT) vom Sensor bis zum qualitätskontrollierten Datenprodukt (z.B. <https://zenodo.org/record/7525399#.ZDfhEXZBwuV>). Diese SOPs werden zusammen mit dem Processing Report mit jedem Datensatz auf PANGAEA verlinkt.

Der Arbeitsablauf für die Bearbeitung und Darstellung der bio-optischen NRT-Daten und für das Online-Monitoring im DAM-Dashboard ist vollständig implementiert. Die Darstellung der NRT-Daten, auch zur schnellen täglichen visuellen Kontrolle, und Hinterlegung in der ICBM-Datenbank sind finalisiert. Tägliche Kontrollen trugen u.a. dazu bei, Fehlfunktionen der Sensorik zu erfassen und je ggf. auf die Notwendigkeit des Austauschs der Geräte bzw. Sensoren an Bord hinzuweisen. Eine Freigabe der qualitätskontrollierten bio-optischen NRT-Daten an Nutzer*innen ist noch ausstehend. Für die bio-optischen Sensoren liegt im Berichtszeitraum ein qualitätskontrollierter publizierter Datensatz in PANGAEA vor. Weitere Datensätze sind entweder zum Zeitpunkt dieses Berichtes in Evaluation, oder zwar evaluiert, aber noch nicht eingereicht. Das Hauptaugenmerk liegt hierbei zunächst auf Datensätzen, für die eine Referenz-Probennahme oder anderweitige Vergleichsmessungen vorhanden sind, die zur Qualitätskontrolle bzw. Kalibrierung herangezogen werden können. Zur Erhebung solcher Daten wurden im Vorfeld der entsprechenden Expeditionen Absprachen mit den jeweiligen Fahrtteilnehmer*innen getroffen. Weiterhin wurden Expeditionen ohne Freigabe zur Bearbeitung der Daten der bio-optischen Sensoren im Projekt „Unterwegs“-Daten begleitend in der Prozessierung unterstützt. Dies erfolgte u.a. mit internationalen Kolleg*innen, so dass auch hier der DAM-Standard Anwendung fand. Im Rahmen dieser Unterstützung wurde ein weiterer Datensatz bei PANGAEA eingereicht. Es hat sich weiterhin gezeigt, dass das Konzept der klassischen täglichen Referenz-Probennahme mit Auswertung nach der Fahrt, wie sie bei anderen Systemen (z.B. TSG) Anwendung findet, für bio-optische Sensorik nicht praktikabel ist. Unbehandelte Wasserproben verändern ihre optischen Eigenschaften durch Lagerung, und eine Behandlung der Proben (Filtration mit anschließendem Einfrieren) ist vergleichsweise aufwändig und dementsprechend in vielen Fällen nicht von den jeweiligen Expeditionsteilnehmern*innen zu leisten. Dementsprechend bedarf die Qualitätskontrolle der

bio-optischen Sensorik zukünftig Anpassungen an die Ausrichtung und das Arbeitsprogramm der jeweiligen Expedition, um andere Wege zu finden, wie eine Validierung der Daten gewährleistet werden kann. Alle Datenaufbereitungsschritte und dessen Evaluierung werden in einer SOP festgehalten und im ICBM-Wiki hinterlegt. Der Versuch einer automatischen und KI-basierten Qualitätskontrolle bio-optischer Daten war noch nicht erfolgreich. Grund hierfür war unter anderem die erwähnte Schwierigkeit der Referenz-Probennahme zur Validierung der Messdaten, was einen Mangel an KI-Trainingsdaten zur Folge hatte. Die bisherigen Ansätze werden begleitend mit zukünftigen Fahrten fortgeführt.

Für die Datenerhebung mittels CTD wurde ein Software-Framework aufgesetzt, um schiffs- und einsatzbezogene Metadaten in die Prozessierung einfließen zu lassen und in Form einer Nutzeroberfläche Nutzer*innen zur Verfügung zu stellen. Die praktische Erprobung der Software ist abgeschlossen und kann nun implementiert werden. Für die weitere Verarbeitung befindet sich neben einer Oberfläche zur einfachen Visualisierung der CTD-Daten eine Post-Prozessierung zur Validierung der Daten bis hin zur Archivierung (Upload) der Daten in entsprechenden Datenbanken (PANGAEA) in Arbeit. Methodische Grundlage bildet die Operationalisierung einer Standardprozedur (SOP) zur qualitätsgesicherten CTD-Datenerhebung unter Nutzung der auf Forschungsschiffen zur Verfügung stehenden Möglichkeiten (Vergleichsmessungen). Auf der beantragten Erprobungsfahrt mit dem Forschungsschiff ELISABETH MANN BORGESE im November 2022 wurde eine finale Evaluierung durchgeführt. Es liegt nun ein funktionsfähiger Demonstrator vor.

Die Rohdaten der Ferrybox wurden in den von der AG-Technik entwickelten Massendaten-Management (MDM) (siehe Abbildung 1) Datenfluss eingebunden und stehen nun nach der Fahrt in der Cloud-basierten Arbeitsumgebung („DAM-Space“) zur Prozessierung bereit. Zur verbesserten Nachvollziehbarkeit der Datenprozessierung wurden Prozessierungsschritte in den „DAM-Space“ verlagert.

Die Übertragung, der nach internationalen Standards qualitätskontrollierten Near-Real-Time (NRT) Temperatur- und Salzgehaltsdaten (TSG) an internationale Portale sowie die Publikation hochwertiger qualitätskontrollierter Daten in PANGAEA, wurden erfolgreich durchgeführt. Die NRT-Daten werden täglich automatisiert prozessiert, kontrolliert und an GOSUD/CORILIS versendet. Um hochwertige Daten zu erhalten, wurden Salz-Probenahmen an Bord der Schiffe koordiniert und die Messungen der Proben an Mitgliedsinstituten an Land durchgeführt. Die Qualitätskontrolle beinhaltet den Vergleich mit vorhandenen Referenzmessungen und etwaiger sekundärer Datensätze. Das wurde für so viele Seereisen wie möglich durchgeführt. Fertigprozessierte Datensätze von Seereisen wurden auf PANGAEA in definiertem Format und mit einem standardisierten Processing Report veröffentlicht. Zusätzlich wurden SOPs für alle drei von der Leitstelle Deutsche Forschungsschiffe betriebenen Schiffe auf zenodo (siehe <https://zenodo.org/communities/dam/?page=1&size=20>) veröffentlicht, welche die standardmäßige Prozessierung der TSG-Daten beschreiben.

Um einen reibungslosen Übergang der Daten vom Schiff zu PANGAEA zu gewährleisten und gleichzeitig die Daten für das Datenportal "Deutsche Meeresforschung" sowie weitere internationale Projekte und Portale nutzbar zu machen, wurden diese Arbeiten auch weiterhin in der AG-Ozeanographie und das institutionelle und projektbezogene Forschungsdaten-Management am GEOMAR eingebunden und aufeinander abgestimmt. Die gemeinsame Nutzung der institutionellen Forschungsinfrastrukturen wie OSIS und O2A ermöglichen die Etablierung eines nachhaltigen und beständigen Forschungsdaten-Lebenszyklus. Zur Qualitätssicherung der Sensordaten ist für TSG ein Probenkonzept in Betrieb. Für bio-optische Sensoren weicht das Probenkonzept von dem des TSG ab und wird im Rahmen ausgewählter Fahrten in Kooperation und Absprache mit der Wissenschaft an Bord umgesetzt. Es stehen somit nicht für jede Fahrt Referenzproben zur Verfügung. Da die Probennahme zeitaufwendig ist und sich auch nicht mit jeder Expedition-Ausrichtung vereinen lässt, wurden Lösungsansätze geprüft, eine Vergleichbarkeit mit „vorkalibrierter“ Sensorik zur schnellen Messung an Bord umzusetzen. Dies muss zukünftig in Zusammenarbeit mit der Wissenschaft an Bord getestet werden. Für die Ferrybox wird ebenfalls ein Probenkonzept entwickelt, um die Qualität der erhobenen Daten zu validieren. Des Weiteren wird das Sensormanagementsystem (sensor.awi.de) für die Hinterlegung der Sensor-Metadaten genutzt, und diese werden mit den Datenpublikationen verknüpft (via Handle).

Arbeitspaket 2: Datenkuration

Das Arbeitspaket 2 wurde federführend von PANGAEA-Mitarbeiter:innen im Projekt bearbeitet.

Stationslisten:

Die kuratorischen Abläufe zur Archivierung von Stationslisten aus DSHIP ActionLogs wurden kontinuierlich verbessert und an Veränderungen seitens DSHIP angepasst. Je nach Forschungsschiff und unterschiedlichen DSHIP ActionLog Formaten variiert die semi-automatische Bereitstellung sowohl in ihrem Grad der Automatisierung als auch in der Möglichkeit, den Grad der Automatisierung zu erhöhen. Es wird auch in Zukunft die Entwicklung der semi-automatischen Bereitstellung für alle großen Forschungsschiffe vorangetrieben, wie zum Beispiel die Verknüpfung von registry.awi.de (vorher sensor.awi.de) handles mit den Geräteeinsätzen. Somit wird gewährleistet, dass alle wichtigen Metadaten mit den "Unterwegs"-Daten verknüpft und zeitnah archiviert werden können. Die erstellten Stationslisten sind unter <https://pangaea.de/expeditions/> auffindbar.

Um die DSHIP ActionLogs noch effektiver für die Metadatenerfassung zur Datenarchivierung nutzen zu können, wurde eine engere Vernetzung mit AP 5 etabliert.

Standardisierung:

Die Etablierung der Standardisierung der Datenpublikation unter Berücksichtigung der FAIR-Prinzipien wurde weiter vorangetrieben. In diesem Zusammenhang wurde für jedes der auf den

vier Forschungsschiffen betriebenen Geräte (ADCP, Bio-optische Sensoren, CTD, Ferrybox, TSG und Fächerecholot) ein eigenes Standard Operating Procedure (SOP) für die PANGAEA Datenpublikation entwickelt. Die SOPs sind unter https://wiki.pangaea.de/wiki/Best_practice_manuals_and_templates hinterlegt und werden kontinuierlich angepasst.

Die Anwendung von s.g. "Quality Flags" ist für alle ozeanographischen Daten beschlossen worden. Die Nutzung einheitlicher Definitionen von Quality Flags, z.B. die von SeaDataNet (<https://www.seadatanet.org/Standards/Data-Quality-Control>) wird aktiv unterstützt. Eine detaillierte Beschreibung der Datenprozessierung inklusive der Quality Flags wird in "Data Processing Reports" für jeden ozeanographischen "Unterwegs"- Datensatz erstellt. In diesem Zusammenhang wird darauf geachtet, dass die Benennung der angewandten Methodik und der Gerätschaften nach dem internationalen Standard des Natural Environment Research Councils (NERC; <https://vocab.nerc.ac.uk/>) erfolgt. Die oben genannten Schritte dienen der Harmonisierung der Arbeitsabläufe für die Publikation von Daten der DAM-Partnerinstitute. Darüber hinaus werden diese entwickelten Standards der FAIRen Datenarchivierung im Bereich der Bathymetrie und Sedimentecholote auch generell für nicht unmittelbar aus dem Projekt "Unterwegs"-Forschungsdaten stammende Daten angewendet.

Datenarchivierung:

Es werden weiterhin kontinuierlich Datensätze bei PANGAEA archiviert und veröffentlicht. Es werden sowohl Rohdaten als auch prozessierte Daten archiviert, abhängig vom Entwicklungsstand des SOP des jeweiligen Geräts. In diesem Zusammenhang werden auch nicht-öffentliche Datensätze bei PANGAEA publiziert. Das bedeutet, dass lediglich die Metadaten, die einen Datensatz beschreiben (wann, wo und mit welchem Gerät wurden Daten aufgenommen), öffentlich sichtbar sind. Der eigentliche Datensatz ist jedoch mit einem Moratorium belegt und passwortgeschützt. Spätestens zwei Jahre nach der Publikation der Metadaten in PANGAEA, sollten auch die Daten offen zugänglich sein. So sieht es die DAM-Forschungsdaten-Leitlinie vor.

Output:

Während der Projektlaufzeit wurden im DAM "Unterwegs"Projekt 248 Datensätze nach den erarbeiteten Kurationsstandards in PANGAEA archiviert und sind über DOIs auffindbar und zitierbar. Insgesamt wurden 87 Multibeam Echosounder (68 Rohdaten, 5 prozessierte Datensätze, 14 water column datasets), 47 TSG, 32 ADCP, 1 Bio-optische Sensoren und 4 CTD-Datensätze archiviert. Zusätzlich wurden noch 76 qualitätskontrollierte Schiffsrouten (Master Tracks) der großen deutschen Forschungsschiffe in PANGAEA archiviert. Die hier angegebenen Datensätze (siehe Abbildung 4) stammen von insgesamt 107 Expeditionen. Zusätzlich wurden noch 30 Datensätze nach den etablierten Kurationsstandards von nicht unmittelbar aus dem Projekt "Unterwegs"-Forschungsdaten stammenden Daten archiviert.

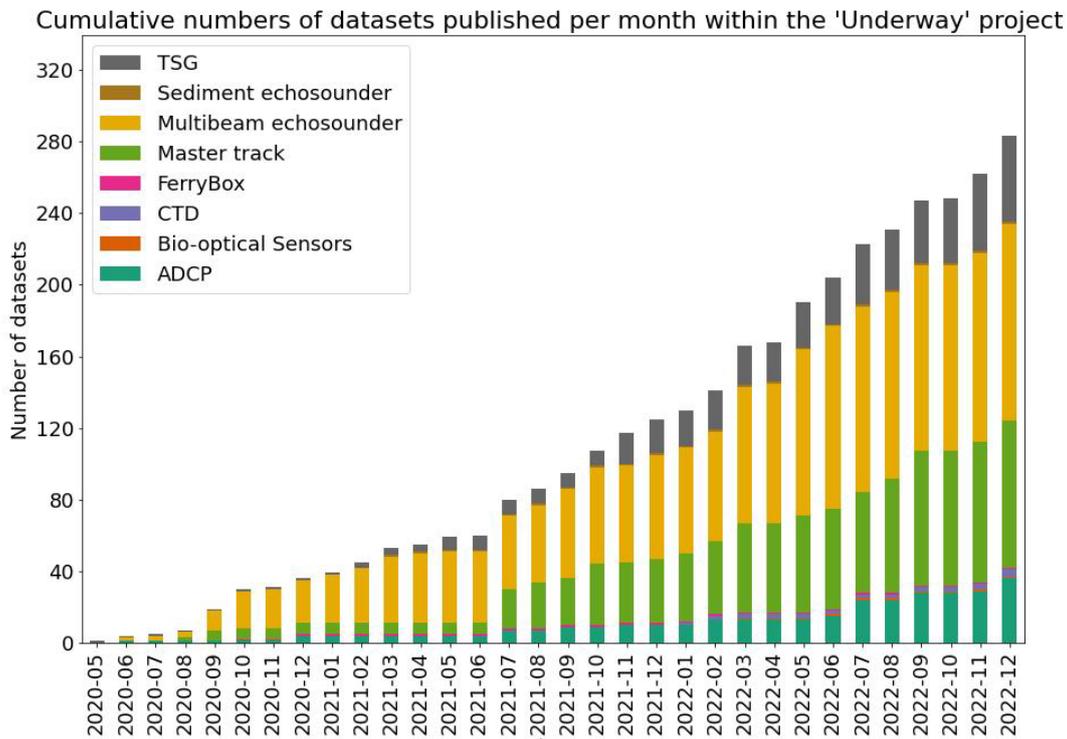


Abbildung 4: Kumulative Zahl der Datenpublikationen in PANGAEA, die aus dem Projekt "Unterwegs"-Daten generiert wurden. Hinweis: Die x-Achse beginnt zeitlich mit der ersten "Unterwegs"-Daten-Publikation in PANGAEA.

Neben der standardisierten Archivierung aktueller Datensätze aus der Bathymetrie sind historische Bathymetrie-Daten aus dem Datenbestand des BSH publiziert worden. Dies umfasst insgesamt 60 Datensätze.

Des Weiteren werden nun auch Daten des Forschungsseglers EUGEN SEIBOLD in PANGAEA aufgenommen und diesbezüglich ein SOP entwickelt und ausgeführt. Da es sich bei den anstehenden Datensätzen um die ersten Datenpublikation der EUGEN SEIBOLD handelt, können die im Rahmen von DAM erarbeiteten Standards und Konventionen für "Unterwegs"-Daten von vornherein berücksichtigt werden.

Verbesserte Sichtbarkeit:

Die Sichtbarkeit der Datensätze im Portal Deutsche Meeresforschung und die damit verbundene technische Verknüpfung zwischen dem Portal Deutsche Meeresforschung und PANGAEA ist verbessert worden. Es wurde eine engere Zusammenarbeit etabliert. Damit eine einfache Unterscheidung im Portal z.B. nach Rohdaten und „prozessierten, qualitätskontrollierten“ Daten möglich wird, wurde neben anderen maschinenlesbaren Metadaten, der Metadaten-Katalog in

PANGAEA um Prozessierungs-Level („processing levels“) erweitert

(https://wiki.pangaea.de/wiki/Processing_levels).

Die Sichtbarkeit FAIRer Datenpublikation in PANGAEA konnte somit weiter erhöht werden.

DAM-Forschungsmissionen:

Im August 2021 startete die erste Forschungsmission „Marine Kohlenstoffspeicher als Weg zur Dekarbonisierung“ (CDRmare), und im Dezember 2021 die zweite Mission zum „Schutz und nachhaltige Nutzung mariner Räume“ (sustainMare). Parallel wurde das Beratungsangebot zum Forschungsdatenmanagement in den Forschungsmissionen ausgebaut.

Ende 2021 wurde ein Arbeitskreis aus Mitgliedern der DAM, PANGAEA und Datenmanager*innen der Forschungsmissionen eingerichtet (regelmäßige Treffen alle 3 Wochen). Dieser Arbeitskreis bildete eine wichtige Grundlage, die übergreifenden Datenmanagement-Ziele der DAM auf die Forschungsmissionen auszuweiten. Es wurde ein koordiniertes Datenmanagement ermöglicht, der Austausch zum Forschungsdatenmanagement sowie die Nutzung gemeinsamer Dateninfrastrukturen (z.B. Marine Data Portal) gefördert. In der Forschungsmission CDRmare wurde in enger Zusammenarbeit mit PANGAEA eine Data Policy und eine Vorlage für einen Datenmanagementplan (DMP) erstellt, welche auch der Forschungsmission sustainMare als Vorlage diente. Die DMP-Vorlage enthielt unter anderem eine Tabelle mit Infos zu Datentypen/-volumen/Repositorien. Diese bildete die Grundlage, um ein Inventar von Datentypen für die Forschungsmissionen zu erstellen, das Datenaufkommen, geplante Repositorien sowie die Zeiträume der Archivierung abschätzen zu können. Die Analyse der DMPs ergab, dass die Archivierung für einen Großteil der Datensätze aus den Forschungsmissionen erst 2023-2024 zu erwarten ist. Insgesamt 7 Datensätze der Forschungsmissionen wurden bis Ende 2022 bei PANGAEA kuratiert und veröffentlicht.

Die Datenmanager*innen/Projektkoordinator*innen der Forschungsmissionen wurden zudem durch individuelle Beratung bzw. verschiedene Weiterbildungsangebote (z.B. Workshops und Infoveranstaltungen) unterstützt. Konkret wurden bis Ende 2022 drei Infoveranstaltungen und 3 Praxisworkshops zum Thema Forschungsdatenmanagement/PANGAEA angeboten.

Arbeitspaket 3: Automatisierte Anbindung an internationale Portale

Zusätzlich zu den Zielen, die das BSH innerhalb dieses Arbeitspaketes erreicht hat, soll die internationale Anbindung bathymetrischer Daten an die International Hydrographic Organisation (IHO) durch die AG-Bathymetrie an dieser Stelle hervorgehoben werden. Die Beschreibung dieses Punktes ist zur Einhaltung der Struktur dieses Berichts unter Arbeitspaket 1, AG-Bathymetrie zu finden. Außerdem wird die tägliche automatische Übertragung der NRT-

Daten des Thermosalinograph (TSG) an das internationale GOSUD Portal ebenfalls schon unter AP1 genannt und hier hervorgehoben.

Im Fokus des BSH-Anteils am Arbeitspaket 3 „Anbindung an internationale Datenportale“ liegen die in PANGAEA veröffentlichten „Unterwegs“-Daten der Geräte Thermosalinograph (TSG), ADCP, Ferrybox und CTD. Die Arbeiten zur Anbindung an das europäische Datenportal Copernicus Marine Service (CMEMS) sind entsprechend dem Meilensteinplan weit vorangeschritten. So wurden die Metadaten beurteilt, stehen vollständig zur Verfügung und erfüllen die Anforderungen an eine Veröffentlichung seitens Copernicus. Ein „Mapping“ der Parameternamen zwischen PANGAEA und Copernicus wurde ebenfalls vorgenommen (Abbildung 5).

Die im Berichtszeitraum veröffentlichten und freigegebenen „Unterwegs“-Daten, welche im Fokus dieses Arbeitspaketes stehen, ordnen sich den Geräten TSG, ADCP und bio-optische Sensoren zu und wurden bezüglich der angewendeten Qualitätskontrolle untersucht. Da diese Daten bereits qualitätskontrolliert sind und das von Copernicus geforderte Flagging-Schema umsetzen, ist eine weitere Qualitätskontrolle für die Veröffentlichung bei Copernicus nicht notwendig.

Durch Beurteilung der in PANGAEA veröffentlichten „Unterwegs“-Daten und deren Bearbeitung im BSH-Databroker hat sich eine Rückkopplung zwischen dem BSH und PANGAEA bzw. den Geräteverantwortlichen aus der AG-Ozeanographie entwickelt. Diese trägt zur Verbesserung der Qualität der veröffentlichten Datensätze bei, indem zum Beispiel vom BSH auf fehlende, falsch zugeordnete oder fehlerhafte Metadaten hingewiesen wird.

Der Databroker des BSH, also die technische Umsetzung der Anbindung an internationale Portale, ist im Berichtszeitraum soweit entwickelt, dass bei PANGAEA neu veröffentlichte oder modifizierte „Unterwegs“-Daten entsprechend des Suchschemas dieses Arbeitspaketes automatisiert gefunden, nach den CMEMS-Vorgaben aufbereitet und in CMEMS veröffentlicht werden. Die Entwicklung geschah in stetigem Austausch mit und durch wertvolle Unterstützung durch PANGAEA. Damit ist sowohl die Schnittstelle zu PANGAEA als auch zu CMEMS im Databroker implementiert. Sie dokumentiert regelmäßig und fortlaufend sowohl Neuerungen/Änderungen der „Unterwegs“-Daten als auch die Weitergabe der aufbereiteten Daten an Copernicus. Des Weiteren bietet sie im Rahmen der PANGAEA-Suche Möglichkeiten zur Erweiterung auf andere Daten-Veröffentlichungen auch außerhalb des Schwerpunkts dieses Arbeitspaketes.



Abbildung 5: Darstellung des Datenflusses von PANGAEA zu Copernicus

Die im Berichtszeitraum bei Copernicus veröffentlichten Datensätze sind dem Gerät ADCP und den bio-optischen Sensoren zuzuordnen. Obwohl auch bereits Datensätze des TSG-Gerätes zur Veröffentlichung bereitstanden, konnten diese im betrachteten Zeitraum noch nicht veröffentlicht werden. Grund hierfür sind bereits in Copernicus aus anderen Projekten als dem Projekt „Unterwegs“-Forschungsdaten vorliegende TSG-Daten deutscher Forschungsschiffe. Diese müssen mit den vorliegenden „Unterwegs“-Forschungsdaten vereinigt werden. Die Umsetzung dieser Vereinigung wurde im Berichtszeitraum durch intensiven Austausch mit der in Copernicus zuständigen Stelle angegangen und entwickelt und steht kurz vor dem Abschluss.

Die in Copernicus veröffentlichten Daten aus dem Projekt werden in sogenannten regionalen „INSITU NRT“-Produkten und zeitlich aggregiert zur Verfügung gestellt. Zur Veröffentlichung der ADCP-„Unterwegs“-Forschungsdaten gab es bei Copernicus folgende Nutzerinfo (22.11.2022): <https://marine.copernicus.eu/user-corner/user-notification-service/insitu-nrt-new-underway-adcp-data-german-research-vessels>

in welcher DAM und PANGAEA explizit genannt werden. In der Nutzerinfo sind ebenfalls die Links zu den regionalen Produkten aufgeführt, unter denen die veröffentlichten Daten zu finden sind. Zur Recherche in den regionalen Produkten ist ein Copernicus-Zugang notwendig.

Insgesamt wurden 13 PANGAEA-Datensätze für die Veröffentlichung verarbeitet; daraus sind insgesamt 31 Copernicus-Dateien entstanden.

Ein Beispiel: Die ADCP-Daten der Meteorfahrt M169, welche im Dezember 2020 stattfand, sind in folgendem PANGAEA-Datensatz zu finden: <https://doi.pangaea.de/10.1594/PANGAEA.933273>.

Hieraus sind für die Copernicus-Veröffentlichung je eine monatliche und eine jährliche Datei erstellt worden: NO_TS_VA_DBBH_2020.nc und NO_TS_VA_DBBH_202012.nc

(Zur Struktur der Dateinamen:

- ‚NO‘ steht für die Produktionseinheit „NorthWestShelf“, welche im BSH angesiedelt ist. Alle Namen der durch das BSH produzierten Copernicus-Dateien beginnen mit „NO“.
- ‚TS‘ bezeichnet den Dateientyp, wobei ‚TS‘ sowohl für Zeitserie als auch für Trajektorie steht;
- ‚VA‘ steht für den Datentyp, also dafür, welches Gerät die Messdaten geliefert hat. In diesem Fall ist es ein „Vessel mounted ADCP“
- ‚DBBH‘ ist das WMO-Rufzeichen der Meteor)

Identische Kopien dieser Dateien liegen in unterschiedlichen regionalen Produkten vor. Die Zuordnung zu den regionalen Produkten ergibt sich aus den in den Dateien vorliegenden geographischen Positionen und den in Copernicus definierten Regionengrenzen.

Weitere Messdaten, welche auf einer anderen Meteorfahrt aufgenommen wurden, würden zu diesen beiden Dateien hinzugefügt werden, sofern sie zeitlich passen. Andernfalls würden neue Dateien mit dem entsprechenden zeitlichen Suffix erzeugt werden.

Der Bezug zur Deutschen Allianz für Meeresforschung ist in den bei Copernicus veröffentlichten „Unterwegs“-Forschungsdaten gegeben durch Angaben in den Metadaten der Dateien. Gleiches gilt für die ursprünglichen Veröffentlichungen in PANGAEA, welche in den Metadaten zitiert werden.

Arbeitspaket 4: Datenportal Deutsche Meeresforschung

Das Arbeitspaket 4 wurde federführend durch die AG-Portal bearbeitet. Kernaufgabe der AG-Portal ist die Entwicklung und der Betrieb des Metadaten Suchportals „Portal Deutsche Meeresforschung“ (Engl. Marine Data Portal, Abbildung 5) inklusive interaktiver Kartenvisualisierungen mit Datenprodukten. Das Portal Deutsche Meeresforschung ist bereits seit Mitte 2020 als Prototyp online und wird bis heute aktiv weiterentwickelt. Das Portal soll der Wissenschaft einen zentralen Einstiegspunkt zum Auffinden von Daten bieten. Zur allgemeinen Verbesserung von Nutzbarkeit und Design des Portals inkl. Visualisierungen wurden in der Projektlaufzeit drei mit insgesamt 35 Teilnehmer*innen aus Wissenschaft und Infrastruktur, s.g. „Usability-Tests“ durchgeführt. Außerdem wurden wichtige Meilensteine, wie Design und Implementierung der Unterstützung für mobile Endgeräte (Responsive Design) sowie grundlegende Optimierungen für Suchmaschinen mit automatischer Erfassung von Nutzungsstatistiken erreicht.

Durch Zu- und Zusammenarbeit mit den anderen Arbeitsgruppen fließen Near-Realtime sowie publizierte Daten der Forschungsschiffe in automatisierte Webdienste, die in einer von der AG-Portal betriebenen Geodateninfrastruktur bereitgestellt werden. Die Webdienste beinhalten Expeditionsrouten der Forschungsschiffe, diverse In-Situ Messungen (z.B. Temperatur, Strömungsdaten, etc.), Rasterdaten (z.B. Bathymetrie, Satellitendaten etc.), Proben sowie Video- und Bilddaten, die ihren Weg in die interaktiven Kartenvisualisierungen finden. Die Kartenansicht dient der Exploration und Auffindbarkeit von Forschungsdaten und wird zusammen mit wichtigen Metadaten und Quellenverweisen für die Zitierbarkeit herausgespielt. Weitere Visualisierungsmöglichkeiten, wie zum Beispiel die Darstellung von Tiefenprofilen und Zeitserien sowie die Abspielmöglichkeit von Videos, sind ebenfalls verfügbar.

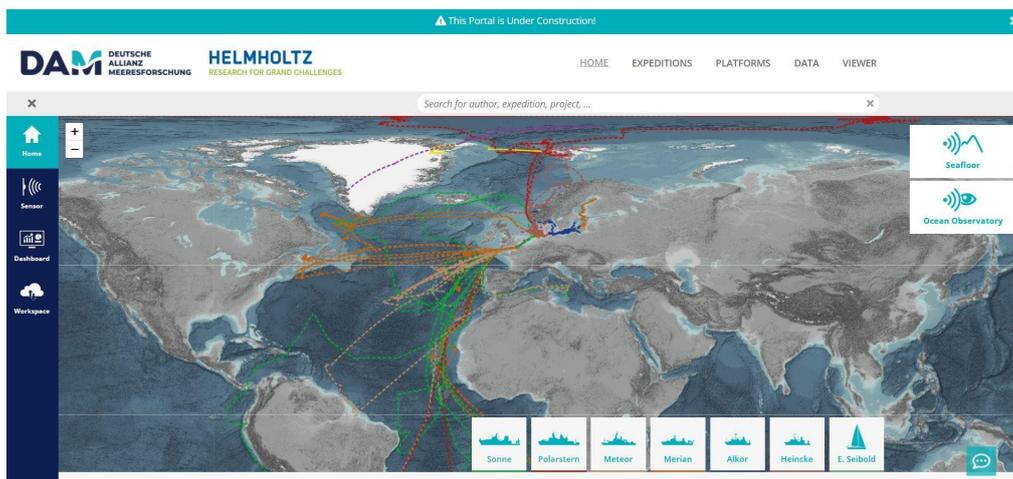


Abbildung 6: Frontend des Datenportals Deutsche Meeresforschung (<https://marine-data.de/>)

Die möglichen Datenflüsse für die Integration der Forschungsdaten wurden in Form von SOPs dokumentiert. Zur Auffindbarkeit der durch die Webdienste bereitgestellten Datenprodukte wurde ein Servicekatalog implementiert, der in Zukunft um eine standardisierte Schnittstelle erweitert wird und somit eine leichte Einbindung in bestehende Portale ermöglicht. Außerdem wurde, nach der Entwicklung eines Design- und Usability Konzepts, die technische Grundlage für einen Login-Bereich implementiert. Dieser wird eine Warenkorbfunktion, Favoritenlisten und konfigurierbare Visualisierungen bieten.

Zur Unterstützung des Verbundprojektes CONMAR der DAM-Forschungsmission sustainMare, wurde ein kartenbasierter Viewer auf Basis föderierter Webdienste erstellt, der Daten zu Munitation in den Meeren visualisiert. Diese Visualisierung dient als Blaupause für zukünftige kuratierte thematische Viewer, mit denen eine interaktive Datenexploration ermöglicht wird (siehe auch Abschnitt *Unterstützung der DAM-Forschungsmissionen*).

Arbeitspaket 5: Technische und inhaltliche Integration und Betreuung von wissenschaftlichen Geräten und Schiffssensorik für ActionLog und Massendatenspeicher der Deutschen Forschungsflotte

Das Ziel der AG-Technik ist, auf technischer Ebene die Datenflüsse an Bord, von Bord an Land und an Land zwischen den beteiligten Infrastrukturen zu optimieren oder wo notwendig neu zu implementieren. Die dafür notwendigen technischen Vorarbeiten und Neuentwicklungen an land- und seegestützten IT-Infrastrukturen sollen zum Teil selber implementiert bzw. als Aufträge an Dritte (z.B. zur Erweiterung bestehender Systeme) vergeben und ihre Durchführung überwacht und koordiniert werden.

Im Jahr 2021 und Anfang des Jahres 2022 wurden folgende Meilensteine durch die AG-Technik erfolgreich abgeschlossen bzw. stehen kurz vor einem Abschluss:

Das Massendatenmanagement-System (MDM) wurde auf den Forschungsschiffen METEOR, MARIA S. MERIAN, SONNE und POLARSTERN wie geplant (siehe Abbildung 7) erfolgreich installiert und in Betrieb genommen. Auf dem MDM werden die aufgezeichneten Forschungsdaten einheitlich an Bord gesammelt und laufend auf geeignete Transportfestplatten kopiert. Diese werden am Ende der Reise vom wissenschaftlich technischen Dienst auf den Schiffen entnommen und per Post zum Bundesamt für Seeschifffahrt und Hydrographie (BSH) geschickt, wo eine Sicherung der Daten vorgenommen wird, ehe die Festplatten an das Alfred-Wegener-Institut (AWI) geschickt werden. Dort werden die Daten in einer Cloud („DAM-Space“) gesichert. Ist dies erfolgt, wird dies an das Schiff rückgemeldet und die Datenkopie an Bord kann gelöscht werden. Die formatierten Transportfestplatten werden vom BSH zurück an die Reederei Briese geschickt, welche sie an die Schiffe weiterleitet.

Neben der Speicherung von „Unterwegs“-Daten bietet der MDM eine auf den großen deutschen Forschungsschiffen standardisierte Möglichkeit, Massendaten durch den Kontext des Ablageordners, mit den korrekten Geräte und Einsatz-Metadaten versehen, abzulegen. In den Fahrtvorbereitungstreffen, wird diese Möglichkeit auch explizit von der AG-Technik beworben und eine Einführung in die Nutzungsweise für die Wissenschaft angeboten.

Die Geräteverantwortlichen der DAM „Unterwegs“-Daten, sowie die Verantwortlichen für die wissenschaftlichen Primärdaten, welche auf dem MDM gespeichert werden, erhalten nach der Speicherung der Daten auf der Cloud eine eindeutige (UUID) URL. Über diese URL können sie auf die Expeditionsdaten nach dem Transfer an Land zugreifen und/oder die Daten weiteren Fahrtteilnehmer*innen zum Download weitergeben.

Alle via MDM an Land gebrachten Daten können im „DAM-Space“ erreicht, bearbeitet und im Datenarchiv PANGAEA publiziert werden. Eine erneute, langwierige Übertragung von Massendaten via Internet ist damit nicht mehr notwendig.

Die MDM-Systeme haben sich an Bord wie auch an Land grundsätzlich sehr bewährt. Sie werden mittlerweile von der Wissenschaft auch für die eigentlichen Forschungsdaten der Expedition als zusätzliche Sicherung und Transportmöglichkeit an Land intensiv genutzt. Die gesammelten Erfahrungen fließen in stetige Verbesserungen der Workflows an Land wie auch an Bord ein und helfen, zielgerichtete Erweiterungen der Software und Workflows anzugehen. Ein Beispiel ist hier die Erweiterung des DSHIP-ActionLogs zur Konfiguration persistenter Identifikatoren (z.B. handle, DOI, URN) für Geräte, die an Bord eingesetzt werden und mit umfangreichen Metadaten (z.B. Kalibrierungen) in der Sensor-Datenbank (REGISTRY) oder anderen passenden Datenbanken registriert sind.

Ein weiteres Teilprojekt, welches erfolgreich bearbeitet wurde, ist die Harmonisierung der in DShip hinterlegten Unterwegs-Parameter-Namen. Im Berichtszeitraum wurde die Harmonisierung erfolgreich auf den Forschungsschiffen MERIAN, METEOR, POLARSTERN, SONNE durchgeführt und abgeschlossen.

In diesem Vorhaben wurde eine auf den Forschungsschiffen einheitliche und gut beschriebene Benennung der "Unterwegs"-Parameter erreicht. Mit dieser Harmonisierung wird eine weitere Steigerung der Metadatenqualität durch eindeutige Referenzierbarkeit bzw. Zuordnung von Messdaten zu Dokumentation der Geräte und Sensoren und deren Ist-Zustand (z.B. Kalibrierung) erzielt. Dies unterstützt z.B. das schiffsübergreifende Prozessieren von Daten. Nun da die Harmonisierung auf den großen deutschen Forschungsschiffen umgesetzt wurde, können auch die Parameternamen auf weiteren Forschungsschiffen im nächsten Schritt harmonisiert werden. Ein erster Schritt in diese Richtung wurde bereits auf der oben beschriebenen Ausfahrt mit der ELISABETH MANN BORGESE unternommen. Dort wurden die Parameter nach dem Vorbild der großen deutschen Forschungsschiffe harmonisiert.

Durch die Standardisierung der Datenflüsse und Metadaten im "DAM-Space" ist eine systematische Publikation der Schiffs-Geräte-Rohdaten mit Metadaten größtenteils automatisierbar. Dies wird bereits vereinzelt genutzt.

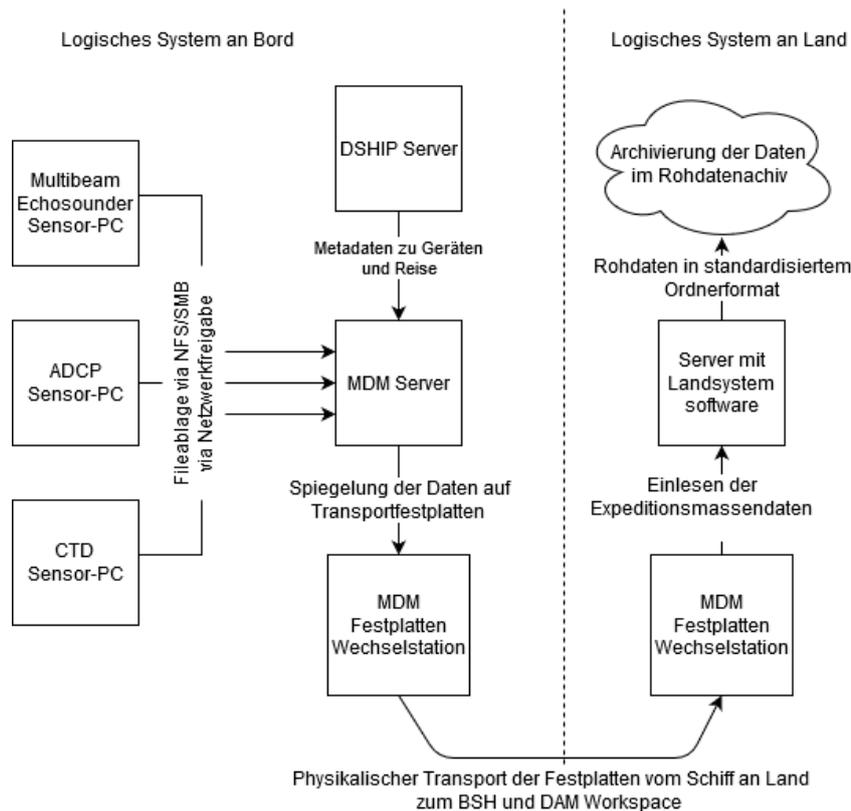


Abbildung 7: Schematische Darstellung des Datenflusses vom Sensor bis ins Archiv.

Neben dem Datentransfer über das MDM wurde auch ein Near-Real-Time (NRT) Datenfluss für unterwegs aufgenommene Messdaten etabliert, welcher wissenschaftliche Daten direkt für das Portal marine-data.de zur Verfügung stellt. Neben ADCP-Daten handelt es sich dabei auch um Positionsdaten der Schiffe und weitere Messdaten wie z.B. Luft-, bzw. Wassertemperatur, Salz oder Chlorophyllgehalt des Meerwassers. Diese Daten werden aus den Datentelegrammen, welche im bordeigenen Datenmanagementsystem genutzt werden, extrahiert und per E-Mail in einem vorgegebenen Rhythmus an das GEOMAR geschickt. Von dort werden die Daten zur weiteren Verarbeitung für die Darstellung im Portal zur Verfügung gestellt. Mit diesem NRT-System wurde eine Basis geschaffen, die kontinuierlich je nach wissenschaftlichem Bedarf erweitert werden kann.

Um eine Event-Erfassung außerhalb des schiffsgebundenen Datenmanagementsystems zu ermöglichen, bestand ein Projekt der AG-Technik darin, ein System zu entwickeln, welches eine mobile Erfassung von "Offline"-Events bei Landexpeditionen oder auf kleineren Forschungsschiffen unabhängig vom Datenmanagement-System an Bord ermöglicht. Zu diesem Zweck wurde die App MobileEventLog entwickelt, die es ermöglicht, offline Events durch die Erfassung von Metadaten zu dokumentieren. So kann zum Beispiel ein Event bei einer Landexkursion gestartet werden und alle Geräteoperationen können über die App geloggt

werden. Zusätzlich zeichnet die App die genauen Positionsdaten des Events auf. Dadurch entsteht eine genaue zeitliche und geografische Auflistung der durchgeführten Geräteeinsätze. Eine Verknüpfung mit der Sensor-Datenbank ist implementiert und ermöglicht, die erfassten Metadaten zu synchronisieren. Damit ist nachhaltig nachvollziehbar, welche wissenschaftlichen Sensoren bzw. Geräte verwendet wurden. Diese App kann auch auf kleinen Forschungsschiffen eingesetzt werden und ist als Android App im Playstore verfügbar.

Angrenzende Aktivitäten

Unterstützung der Atmosphärenforschung

Die Erfassung von oberflächennahen Sensordaten des Seewassers en route ist auf dem Forschungsschiff METEOR durch eine autonome Messbox erweitert worden, die mit Sensoren für die Messung von Partikeln in der Atmosphäre bestückt ist. Eine Auswahl von Parametern dieser kontinuierlichen Luftmessungen auf hoher See wird zusätzlich an das DSHIP-System an Bord gesendet, von diesem mit abgespeichert und zusammen mit den Seewassermessungen nach Expeditionsende im Landsystem beim BSH archiviert. Hier sind die Daten für die Wissenschaft zur ersten Qualitätskontrolle und Analyse abrufbar. Auch diese Daten werden in den Datentransfer- und Publikationsworkflow integriert und als qualitätskontrollierte Datensätze öffentlich zugänglich gemacht. Diese Aktivität belegt, dass die Infrastruktur an Bord der Schiffe und der Datenfluss generisch sind. Partner außerhalb der Meeresforschung (hier Prof. Mira Pöhlker, Leibniz-Institut für Troposphärenforschung - TROPOS) profitieren von der Infrastruktur. Auf diese Weise unterstützen wir die Erdsystemforschung mit "Unterwegs"-Forschungsdaten aus der Meeres- sowie der Atmosphärenforschung.

Zusammenarbeit mit dem NFDI4Earth Pilotprojekt "German Marine Seismic Data"

Zum NFDI4Earth Pilotprojekt German Marine Seismic Data:

Reflexionsseismische Daten sind die wichtigste Informationsquelle für die Struktur des Untergrunds, da sie von allen geophysikalischen Verfahren die höchste Auflösung bieten. Als solche werden sie für eine Vielzahl von akademischen und kommerziellen Zwecken genutzt. Jahrzehntlang waren reflexionsseismische Daten die größten Datensätze in den Geowissenschaften, was zu erheblichen Speicher- und Archivierungsproblemen geführt hat. Bis heute gibt es kein internationales, europäisches oder deutsches Datenportal, das von der gesamten Gemeinschaft genutzt wird, was alle neuen wissenschaftlichen Projekte behindert, die die alten Daten nutzen möchten. Im Rahmen dieses Pilotprojekts werden wir eine

vereinheitlichende Dateninfrastruktur entwickeln, eine Datenarchäologie für bestehende Daten starten und die zukünftige Archivierung von reflexionsseismischen Daten von zukünftigen Forschungsfahrten vorbereiten. Die Initiative dient in erster Linie der deutschen Meeresgeophysik-Gemeinschaft, vertreten durch AGMAR des Fachkollegiums Physik der Erde der DFG. Das Pilotprojekt für reflexionsseismische Daten trägt zu den Bemühungen von NFDI4Earth bei, eine verteilte Infrastruktur für die Datenkuratation durch harmonisierte Datenworkflows mit Verbindungen zu internationalen Datenrepositorien zu etablieren. Schnittstellen und Speicherlösungen für Datenmengen in der Größenordnung von GBs bis TBs werden in vielen weiteren Wissenschaftsbereichen dringend benötigt.

Zur Zusammenarbeit mit dem Projekt “Unterwegs“-Forschungsdaten:

Zur Speicherung reflexionsseismischer Daten an Bord der großen deutschen Forschungsschiffe, zur Übertragung, Archivierung und Publikation an Land wird die Infrastruktur und der Datenfluss genutzt, der im Rahmen des Projektes “Unterwegs“-Forschungsdaten für “Unterwegs“-Forschungsdaten aufgebaut wurde. Dies belegt erneut, dass die Ergebnisse des Projektes auf Expeditionsdaten im Allgemeinen anwendbar sind. Die Visualisierung der Metadaten reflexionsseismischer Daten in einem “Thematischen Viewers” des Portals Deutsche Meeresforschung (siehe auch folgenden Abschnitt) wird in Zukunft zusätzlich ermöglicht. Als Analogie kann der Viewer zu Bathymetrie-Daten herangezogen werden.

Unterstützung der DAM-Forschungsmissionen

Das Portal Deutsche Meeresforschung ermöglicht es den Nutzer*innen, Forschungsdaten aus interdisziplinären Forschungskoperationen und individuellen Forschungsinitiativen zu visualisieren, zu durchsuchen und abzurufen. Neben der Integration einzelner Forschungsdatensätzen fördert es die Aggregation und Visualisierung von kuratierten Datenprodukten und standardisierten webbasierten Datendiensten zur Wiederverwendung über “thematische Viewer”. Die “Thematische Viewer“-Technologie ist generisch im Sinne, dass zusätzliche “Thematische Viewer” bei Bedarf und in enger Zusammenarbeit mit einer wissenschaftlichen Gemeinschaft, die Daten zu einem bestimmten Thema in Web Map Services visualisieren möchte, implementiert werden können. **Zusammen mit dem Verbundprojekt CONMAR der DAM-Forschungsmission sustainMare wurde ein „featured Viewer“ zum Thema “Munition im Meer” im Portal Deutsche Meeresforschung erstellt.** Dieser Viewer visualisiert nahezu jegliche öffentlich zugänglichen Daten bzw. Metadaten zum Thema Munition im Meer und ist damit einzigartig auf dem “Markt” der öffentlich zugänglichen Viewer zu diesem Thema.

Link zum Viewer: <https://marine-data.de/?site=viewer&id=mmdc>

Fazit und Strategischer Mehrwert

Das Projekt "Unterwegs"-Forschungsdaten verfolgt im Rahmen des DAM-Kernbereichs "Datenmanagement und Digitalisierung" fachliche, strukturelle und strategische Ziele. Fachlich – und darauf liegt der Fokus dieses Berichtes – ist das Projekt "Unterwegs"-Forschungsdaten ein konkretes Vorzeigebispiel für die instituts- und flottenübergreifende Bereitstellung der Forschungsdaten deutscher Schiffe nach den FAIR- und Open-Prinzipien.

Strukturell belegt das Projekt, dass der DAM-Kernbereich "Datenmanagement und Digitalisierung" ein geeignetes und erfolgreiches Instrument zur Koordinierung und Weiterentwicklung des institutsübergreifenden Datenmanagements ist. Bestehende (Infra-)Strukturen werden durch die übergreifende Vernetzung von Hochschulen und (Ressort-)Forschungseinrichtungen inhaltlich und technisch ergänzt und verbessert. Der aufgebaute Datenfluss zur systematischen Übertragung von Expeditionsdaten an Land bis zur Archivierung im Repositorium und anschließenden Visualisierung im Portal Deutsche Meeresforschung ist generisch und kann von der Community genutzt werden. Dies zeigt das Beispiel des NFDI4Earth Pilotprojektes German Marine Seismic Data sowie die Visualisierung von WMS zum Thema Munition im Meer gemeinsam mit dem Verbundprojekt CONMAR der DAM-Forschungsmission sustainMare.

Strategisch bildet das Projekt "Unterwegs"-Forschungsdaten einen zentralen Baustein im Datenmanagement-Konzept der deutschen Meeresforschung. Das Projekt trägt zum koordinierten Vorgehen im Bereich Forschungsdatenmanagement, zur Sichtbarkeit sowie zur Sprech- und Handlungsfähigkeit des DAM-Kernbereichs "Datenmanagement und Digitalisierung" bei.

Referenzen

Becker, M., Betz, M., Damaske, D., Kopte, R., Pieck, D., Schlundt, M., & Wiemer, G. (2022). DAM-Underway Acquisition and Transfer, Cruise No. MSM97/2, 12.12.2020 - 24.12.2020, Emden - Emden (Germany). In MARIA S. MERIAN-Berichte (1.0). Begutachtungspanel Forschungsschiffe.

https://doi.org/10.48433/CR_MSM97_2

Kopte, Robert, Betz, Maximilian, & Anselm, Norbert. (2022). SOP 38kHz ADCP aboard of RV Sonne (vessel:sonne:adcp_38khz_650074) (1.1). Zenodo.

<https://doi.org/10.5281/zenodo.7525399>

Robert Huber, Egor Gordeev, Markus Stocker, Aarthi Balamurugan, & Uwe Schindler. (2020, September 3). pangaeapy - a Python module to access and analyse PANGAEA data (Version 1.0.0-beta). Zenodo. <http://doi.org/10.5281/zenodo.4013941>

Schlundt, Michael, & Glemser, Barbara. (2023). SOP of thermosalinograph (TSG) aboard of RV Meteor (1.0). Zenodo. <https://doi.org/10.5281/zenodo.7535546>

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Hinweis: Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.