

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Forschungsdatenmanagement im SFB 754 “Klima – Biogeochemische Wechselwirkungen im Tropischen Ozean”

Hela Mehrtensⁱ Pina Springerⁱⁱ

2019

Zitiervorschlag

Mehrtens, Hela und Pina Springer. 2019. Forschungsdatenmanagement im SFB 754 “Klima – Biogeochemische Wechselwirkungen im Tropischen Ozean”. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2019: S. 84-88. DOI: [10.17192/bfdm.2019.2.8073](https://doi.org/10.17192/bfdm.2019.2.8073).

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

ⁱGEOMAR – Helmholtz-Zentrum für Ozeanforschung Kiel. ORCID: [0000-0002-4526-2472](https://orcid.org/0000-0002-4526-2472)

ⁱⁱGEOMAR – Helmholtz-Zentrum für Ozeanforschung Kiel. ORCID: [0000-0003-1071-3234](https://orcid.org/0000-0003-1071-3234)

1 Abstract

Das Datenmanagement des SFB 754 gehört zum Teilprojekt Z1 (Central activities of the Collaborative Research Centre 754). Das GEOMAR Datenmanagement-Team bietet Systeme und Services zur Beschreibung und zum Austausch von Daten innerhalb des Projektes an (OSIS) und unterstützt bei der Publikation von Daten bei PANGAEA zur langfristigen Verfügbarkeit und Nachnutzung nach Projektende. Dazu gibt es Schulungen und Einzelberatungen zum Datenmanagement, auch wird die Darstellung des SFB 754 nach außen über die technische Pflege der Webseiten (Inter- und Intranet) unterstützt. Die Sichtbarkeit des Datenmanagements zu erreichen sowie die Definition von dessen Aufgaben waren am Anfang eine Herausforderung. Mittlerweile ist durch verschiedene Maßnahmen jedoch die Vernetzung sehr gut. Da der nachhaltige Umgang mit Daten immer in Konkurrenz zu anderen wissenschaftlichen Arbeiten steht, ist ein stetiges 'ins Bewusstsein rufen' notwendig, das mit Serviceangeboten und dem Angebot technischer Lösungen für das Forschungsdatenmanagement gekoppelt wird. In den letzten 10 Jahren ist dadurch schon ein Bewusstseinswandel erreicht worden.

2 Struktur

Im Sonderforschungsbereich 754 (Klima - Biogeochemische Wechselwirkungen im Tropischen Ozean)¹ werden durch Schiffsexpeditionen, Experimente und Modellsimulationen die biogeochemischen Wechselwirkungen im tropischen Ozean untersucht. Die Forschungsdaten setzen sich aus ozeanographischen Messdaten, Laboranalysen und Modellsimulationen in diversen Formaten und Größen zusammen.

Das Datenmanagement im SFB ist in jeder der drei Projektphasen von 2008 bis 2019 ein fester Bestandteil. Es ist im Teilprojekt Z1 (Central activities of the Collaborative Research Centre 754) verortet. Die im SFB 754 angestellte Datenmanagerin arbeitet integriert im fünfköpfigen Datenmanagement-Team des GEOMAR Daten- und Rechenzentrums. Es wurde 2008 als Pool von drei Vollzeitstellen für Datenmanagement der beteiligten Großprojekte und einer GEOMAR-finanzierten Vollzeitstelle mit dem Ziel gegründet, eine nachhaltige Organisation und Infrastruktur für das Management der Forschungsdaten am GEOMAR aufzubauen. Dadurch konnten nicht nur spezielle Systeme für den SFB in dessen Förderzeitraum aufgebaut und betrieben, sondern eine zentralisierte und nachhaltig erhaltbare Infrastruktur geschaffen werden, die auch für andere zukünftige Projekte und die Meeresforschung in Kiel generell nutzbar ist.

¹<http://www.sfb754.de>

3 Ziel

Das Ziel ist es, eine dauerhafte Einrichtung für Forschungsdatenmanagement in den Kieler Meereswissenschaften zu etablieren. Das Datenmanagement berücksichtigt den gesamten Datenlebenszyklus von der Planung, Erhebung und Beschreibung über die Auswertung bis hin zur Publikation, Archivierung und Nachnutzung der Forschungsdaten. Das Datenmanagement-Team unterstützt die Forschenden in allen Phasen, um einen transparenten und nachvollziehbaren Umgang mit den Daten in Bezug auf folgende Fragen zu erreichen: Welche Daten sind wo und wann erhoben worden? Welche Methoden wurden verwendet? Welche Personen sind verantwortlich? Zu welchem Zeitpunkt sind die Daten wo und unter welcher Lizenz langfristig verfügbar und auffindbar?

4 Methoden

Das GEOMAR Datenmanagement-Team bietet mehrere Systeme und Services zur Speicherung, Beschreibung und zum Austausch innerhalb des Projektes an und unterstützt bei der Publikation der Forschungsdaten zur langfristigen Verfügbarkeit und Nachnutzung, auch über das Projektende hinaus. Dazu gibt es Schulungen, Einzelberatungen sowie Tutorials zum Datenmanagement. Desweiteren wird die Darstellung des SFB 754 nach außen über die technische Wartung der Webseiten (Inter- und Intranet) unterstützt.

Das Datenportal für den SFB basiert auf der Portalsoftware Liferay als Hauptkomponente, in die eine selbst entwickelte Webanwendung für den internen Austausch von Daten, das Ocean Science Information System (OSIS)², integriert ist. Automatisch generierte Publikationslisten sind aus dem mit OSIS verknüpften institutionellen Repositoryum OceanRep³ verfügbar. Ferner werden Dokumentenaustausch, Wiki und Blog Funktionalitäten der Portal-Software genutzt.

Die Webanwendung OSIS basiert auf Grails und PostgreSQL. Da die Daten der Meeresforschung hauptsächlich auf Expeditionen, bei Modellsimulationen und Experimenten erhoben werden, sind diese drei Klassen zur Beschreibung verfügbar. Zu jeder Klasse werden spezifische Informationen (z.B. Abfahrt, Ankunft, Fahrtleitung, Gebiet einer Expedition, Modelltyp, Laufzeit, Thema einer numerischen Simulation) und allgemeine Informationen wie Titel, Beschreibung, Projekt und verantwortliche Person zusammengeführt. OSIS zeigt die Metadaten öffentlich an und dient damit zur Information innerhalb und außerhalb des Projekts, es stellt diese zugleich über maschinell lesbare Webschnittstellen (JSON) anderen Systemen, z.B. der GEOMAR Webseite, zur Verfügung. Zugehörige Fahrtberichte, Publikationen, Datenrepositorien und Informationen über Proben werden hier ebenfalls verlinkt.

²<https://portal.geomar.de/osis>

³<https://oceanrep.geomar.de>

Zu den jeweiligen Klassen ist der Upload von Forschungsdaten in diversen Fileformaten und Bearbeitungsstufen mit definierten Zugriffsrechten möglich. Im Projektzeitraum können daher von Rohdaten bis zu analysierten Daten mehrere Versionen zwischen den beteiligten Nutzern ausgetauscht werden. OSIS wird auch als "Data Management Record" in Form eines ständig aktualisierten Datenmanagementplans eingesetzt. Über "Data Deliverables" werden Listen mit Informationen darüber generiert, welche Daten zu erwarten sind, zu welchem Zeitpunkt diese verfügbar sein werden und welche Person dafür verantwortlich ist. Entsprechende Erinnerungen zum Upload und zur Veröffentlichung der Daten werden per E-Mail verschickt.

Weitere im Datenmanagement eingesetzte Technologien sind Versionierungssysteme (Subversion, Git) für Modellcode und sehr große Datenmengen. Ein THREDDS-Server mit OPeNDAP-Schnittstelle ermöglicht den Austausch/Zugriff und die Veröffentlichung vor allem von Modelldaten und gegitterten Daten im NetCDF Format.

Für die Veröffentlichung der prozessierten Daten nutzt der SFB 754 vor allem das Welt Datenzentrum PANGAEA (Data Publisher for Earth and Environmental Science)⁴, das grundsätzlich die langfristige Verfügbarkeit, die Zitierbarkeit und die Nachnutzbarkeit der Daten ermöglicht und so die Zusammenführung mit anderen Datensätzen besser ermöglicht als ein Inhouse-System. Es können aber auch Daten bei ENA⁵ oder anderen disziplinspezifischen Repositorien veröffentlicht werden. Alle veröffentlichten Daten sind sowohl durch die Verlinkung in OSIS als auch in OceanRep⁶ in Verbindung mit wissenschaftlichen Publikationen auffindbar. Diese Repositorien sind auch von der Webseite des SFB 754 aus erreichbar. Unveröffentlichte Daten, z.B. Rohdaten, sind auch weiterhin über OSIS auffindbar.

5 Einbindung

Die Sichtbarkeit des Datenmanagements in den Projekten und am GEOMAR zu erreichen sowie die Festlegung von dessen Aufgaben waren am Anfang eine Herausforderung. Es sollte sowohl ein sehr generisches und dadurch nachhaltiges System entstehen, das für verschiedene, auch künftige Projekte zur Verfügung steht, als auch ein schnell und einfach nutzbares Produkt. OSIS war zunächst als Übergangslösung geplant und sollte durch ein workflowbasiertes System abgelöst werden. Es zeigte sich jedoch bald deutlich, dass zu strenge Vorgaben an Dateiformate und Inhalte den Austausch von Daten erschweren. Daher hat sich das GEOMAR Datenmanagement-Team darauf konzentriert, seinen Ansatz möglichst niedrigschwellig zu halten und auf die tatsächliche Nutzung zu reagieren indem agil neue Features in OSIS implementiert

⁴<https://www.pangaea.de/?q=project%3Alabel%3ASFB754>

⁵<https://www.ebi.ac.uk/ena>

⁶<https://oceanrep.geomar.de/view/project/SFB754.html>

werden. Durch diese Vorgehensweise ist ein Angebot entstanden, das von den Forschenden immer weiter in neuen Projekten und in den Forschungseinheiten genutzt wird.

Durch die regelmäßige Teilnahme an Mitgliederversammlungen und allen Retreats einschließlich der Young Scientist Retreats sowie eine enge Zusammenarbeit mit der Leitung des SFB ist die Vernetzung inzwischen gut. Da der nachhaltige Umgang mit Forschungsdaten gefühlt immer in Konkurrenz zu anderen wissenschaftlichen Arbeiten steht, ist er stetig ins Bewusstsein zu bringen. Dies wird mit Serviceangeboten gekoppelt. In den letzten 10 Jahren ist dadurch schon ein gewisser Bewusstseinswandel erreicht worden und der Mehrwert des vermeintlichen Aufwands wird zunehmend erkannt.

6 Lessons learnt

Um das Datenmanagement in einem Verbundprojekt wie dem SFB 754 verlässlich zu gestalten, ist die verbindliche Festlegung einer Richtlinie (Data Policy)⁷ sehr wichtig. In diesem Dokument sind der Informationsaustausch, der Zugang zu den Daten, die Qualitätskontrolle und die Fristen für Moratorien und Veröffentlichungen geregelt. Für die letzte Phase wurde die Data Policy aktualisiert und Regelungen zum Umgang mit Daten bei Verlassen der Institutionen ergänzt.

Das GEOMAR Datenmanagement-Team hat außerdem die Erfahrung gemacht, dass oft der Wunsch nach der Vorgabe von strukturierten Dateien für den Datenaustausch (Templates) geäußert wird, sie dann aber trotz Bereitstellung nicht genutzt werden. Daher wird inzwischen eher auf gute Beispiele von veröffentlichten Datensätzen hingewiesen, wenn nach dem 'richtigen' Format gefragt wird.

Nach den bisherigen Erfahrungen ist außerdem die Nutzung vorhandener, nachhaltiger Forschungsdateninfrastrukturen wie unsere Anbindung an das Weltdatenzentrum PANGAEA empfehlenswert. Auch die starke Integration des Datenmanagements in einer institutionellen Infrastruktur wie dem Daten- und Rechenzentrum hat sich als essentiell erwiesen. Die Aktivitäten des Datenmanagements müssen in die vorhandene IT Landschaft passen und sind für einen nachhaltigen Betrieb auf Support angewiesen.

Sehr hilfreich war auch die Vernetzung mit anderen Personen aus dem Bereich Datenmanagement auf Konferenzen und Workshops.

Es hat sich gezeigt, dass die Einbindung des Projekt-Datenmanagements in ein dauerhaftes Team sehr nützlich ist, da in vielen Bereichen die Zeitspanne von der Datenerhebung bis zur Datenpublikationen eher zehn als drei Jahre beträgt und damit über den geförderten Projektzeitraum weit hinaus reichen kann.

⁷<https://oceanrep.geomar.de/47369/>