

PANGAEA

Bremen/D

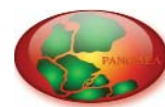


Begehung am 08.04.2013
aktualisiert 18.11.2014

Aufgaben	
Hauptaufgaben:	Langzeitarchivierung und Publikation von Erd- und Umweltdaten.
weitere Aufgaben:	Bereitstellung von Dateninfrastrukturen (u. a. Portalen) u. Forschungswerkzeugen; Projektdatenmanagement
nicht Teil der Aufgaben:	wissenschaftliche Auswertung von Daten
relevante Fachrichtung:	Erd- und Umweltwissenschaften, Geowissenschaften, Biologie; Daten zur Erforschung von Wasser, Sediment, Eis und Atmosphäre

Allgemeine Angaben	
Rechtsform:	integrierte Abteilungen/Arbeitsgruppen innerhalb des Alfred-Wegener-Institut für Polar- und Meeresforschung (Forschungsinstitut der Helmholtz-Gemeinschaft) und MARUM. Center for Marine Environmental Sciences (Universität Bremen)
Gründung/Geschichte:	1993 erstes operationales System 1995 Beginn von Datenmanagement in Projekten 2001 Datenzentrum bei International Council for Science (ICSU) 2007 Kollaboration mit Wissenschaftsverlagen; Member of the ICSU World Data System (World Data Center PANGAEA); World Radiation Monitoring Center (WRMC)
Anzahl Mitarbeiter:	30, davon arbeiten 7-8 als Datenmanager extern in geographisch verteilten Projekten Anzahl FTE (Planstellen) 20 % (z. T. auch unbefristet aus Drittmitteln finanziert) Anzahl Hilfskraftstellen (Planstellen) unbekannt Anzahl FTE (befristet, Drittmittel) 80 % Anzahl Hilfskraftstellen (befristet, Drittmittel) unbekannt Anzahl WissMA 20 Anzahl IT 2 Uni Bremen + 2 AWI Bremerhaven
Zertifizierung:	Zertifizierung und Akkreditierung durch ICSU World Data System (WDS) Zertifizierung und Akkreditierung durch WMO als „Data Collection and Processing Center“ im Rahmen des „WMO Information System (WIS)“
Referenzmodell:	kein Referenzmodell

Datenbestand	
Verpflichtung Daten zu sichern?	keine generelle Verpflichtung; aber z. B. Doktoranden von MARUM müssen Daten dort abgeben
Findet gezielte Datenakquise statt?	Ja, vor allem in den ersten Jahren, etwa seit 2000 müssen Datengeber nicht mehr aktiv gesucht werden



Anzahl an Projektarchiven:	ca. 335.000 Projektdatensätze mit insgesamt ca. 8.8 Milliarden Einzelinformationen (data items) (Stand November 2014)
Anzahl an „grauer Literatur“:	nicht relevant
Anzahl an Doktorarbeiten:	nicht relevant
Anzahl an retrodigitalisierten Zeitschriften:	nicht relevant
Anzahl an Einträgen Nachweiskatalog:	nicht relevant
Gesamtgröße Archiv:	ca. 500TB (Stand November 2014)
Datensatzgröße:	Einzeldateien überwiegend kleiner als 10 MB
Datenzuwachs jährlich:	10.000 – 12.000 Datensätze
Altdaten vorhanden:	Ja

Datengenerierung

Technische Unterstützung bei Datenerhebung (z. B. durch Tools, Software, etc.):	Nein
Fachliche Unterstützung bei Datenerhebung:	Ja durch Beratung beim Daten-/Projektmanagement

Datenupload

Online-Upload möglich?	Ja, über Ticket-System JIRA
Auswahlkriterien für Aufnahme von Daten:	nur fachliche Relevanz
Werden angebotene Daten abgelehnt? (Gründe)	Daten werden nicht angenommen, wenn die Metadaten nicht ausreichend sind, um eine sinnvolle Nachnutzung zu ermöglichen
Vertragsverhältnis Datengeber – Datenarchiv:	Wenn die Daten aus einem finanziertem Projekt stammen, wird ein Archivierungs- und Publikationsvertrag abgeschlossen.
Software/Collection Registry:	Ticketsystem (JIRA) für Upload und RDBM in 4 th Dimension

Kuratierung von Daten

Arbeitsschritte zur Datenarchivierung:	Workflow ist vergleichbar mit Peer-Review-Prozesses bei normalen Publikationen 1. Einreichung von Daten mit Metadaten (Datengeber) 2. Kontrolle technische Qualität und Vollständigkeit der Metadaten (Datenkurator) 3. Aufbereitung, Import in System und Archivierung (Datenkurator)
Werkzeuge für Datenarchivierung:	Eigenentwicklung Ja, verschiedene Tools zur Vereinheitlichung und Aufbereitung von Daten, vor allem Tabellen und strukturierte Text-Dateien verfügbare Software Für Verwaltung von Metadaten und Import von Dateien kommt ein Client-Sever-System von 4 th Dimension zum Einsatz. (kommerziell, OpenSource)
Versionskontrolle:	Ja, in den Metadaten kann auf frühere/spätere Versionen eines Datensatzes hingewiesen werden.



Rollenkonzept:	Archive Standard-Nutzer; Primärforscher/Autor; Nutzer mit expliziten Zugriffsrechten; Datenkurator; Admin
Qualitätssicherung:	Review
Schutzmechanismen für bestimmte Informationen:	Access Rights können für jeden Datensatz individuell gesetzt werden; Datengeber kann Zugriffsrechte in bestimmten Fällen einschränken personenbezogene Metadaten/Dateien Metadata is licensed under CC (by) raumbezogene Metadaten/Dateien Metadata is licensed under CC (by)

Metadaten & Interoperabilität

Mindestanforderungen an Daten:	Mindestangaben für Datensätze müssen vorliegen; Vorgaben zu Datenformaten (vor allem Tabellen und strukturierte Text-Dateien)
Verwendung von Standards:	Ja, s. u.
Metadatenmodell:	eigenes XML-Format (one for all schema) aus dem verschiedene Austauschformate mittels XSLT erzeugt werden können: ISO 19115 DIF Darwin Core STD-DOI (Data Cite) Dublin Core eigene Anpassungen an Metadaten? Ja
vorhandene Schnittstellen:	Webservices (SOAP/WSDL) WFS (OGC) OGC catalogue Service OAI-PMH DiGIR
Sichtbarkeit der Metadaten in anderen Portalen/Aggregatoren:	Ja, z. B. bei Google, mehrere Fachportalen, bei Verlagen, bei Bibliotheken etc.
Sichtbarkeit der archivierten Dateien in anderen Portalen/Aggregatoren:	Ja, v. a. bei Verlagen und Fachportalen
System für Persistente Identifikatoren:	DOI (pro Datensatz) Anzahl PIDs insgesamt 682.902 DOIs

Nicht-Technische Dienstleistungen

Bereitstellung Guidelines, Ratgeber etc.:	Keine, es gibt aber eine kurze Übersicht für empfohlene Abgabeformate.
Antragsberatung:	Ja
Schulungen/Workshops:	unregelmäßig Workshops für Datenkuratoren (primär interne Mitarbeiter), Schulungen zur Nutzung des PAN-GAEA-Systems
Weiterentwicklung Standards:	unbekannt
Support Datenmanagement:	Ja



Finanzen: Einnahmen	
Jahresbudget:	ca. 1 Mio Jahresumsatz nur PANGAEA davon speziell für das Datenzentrum s. o.
Basisfinanzierung:	< 20 % Helmholtz-Gemeinschaft Zeitraum der vertraglich Finanzierung unbekannt nachhaltige Finanzierung? unbekannt
Drittmittel:	Seit 1995 in über 140 Projekten involviert, darunter auch Langzeitstudien und internationale Projekte
restliche Finanzierung:	unbekannt
Gebühren für das Archivieren von Daten:	Nutzung im Prinzip kostenfrei, aber ein Unkostenbeitrag von 500 € wird begrüßt/empfohlen
Gebühren für die Nutzung von Daten:	keine

Finanzen: Ausgaben		
Kostenstruktur nach Kostenträger:	Personalkosten	90 %
	Planstellen	unbekannt
	Drittmittelstellen	unbekannt
	Hardware/Software	10 %
	Reisekosten, Gebäudeinstandhaltung	unbekannt
Kostenstruktur nach Kostenstellen:	Daten akquirieren und aufbereiten	ein Drittel
	Daten archivieren	ein Drittel
	Daten zugänglich machen	ein Drittel
	Softwareentwicklung	unbekannt
	Kosten für PIDs	keine generelle Verpflichtung; aber z. B. Doktoranden von MARUM müssen Daten dort abgeben
Höhe Durchschnitts-Kosten pro 1 GB Daten:	pro Datensatz: halber Tag; ca. 500 €	

Negativa	
nicht funktionierende/ optimierbare Workflows oder Prozesse:	unbekannt
nicht gelöste Probleme:	unbekannt