

PANGAEA - ein Informationssystem für die Meeresforschung

Michael Diepenbroek, Hannes Grobe & Manfred Reinke
Alfred-Wegener-Institut für Polar- und Meeresforschung, 27568 Bremerhaven
info@pangaea.de, www.pangaea.de

Die Vielfalt und der Umfang der in den einzelnen Disziplinen der Meeresforschung anfallenden Daten, wie auch die überregionalen und vernetzten Fragestellungen fordern zunehmend ein integrierendes Datenmanagement. Die moderne Rechner- und Softwaretechnologie in Verbindung mit der globalen Vernetzung durch das Internet erlaubt heute den Aufbau eines entsprechenden Systems.

Vor dem Hintergrund oben angesprochener Bedürfnisse wurde 1993 durch an der Meeresforschung beteiligte Wissenschaftler in Deutschland die Entwicklung eines Informationssystems initiiert (PANGAEA: PaleoNetwork for Geological and Environmental Data). Eine Anschubfinanzierung bewilligte der Bundesminister für Bildung, Wissenschaft, Forschung und Technologie (BMBF). Das System wurde von einer Gruppe von Wissenschaftlern und Informatikern am Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI) in Bremerhaven entwickelt.

Das vorgestellte Informationssystem ermöglicht die Erfassung von Raum/Zeit orientierten Daten aus der naturwissenschaftlichen Grundlagenforschung. Das Netzwerkkonzept sieht einen allgemeinen Zugriff bei gleichzeitig hoher Funktionalität vor. Bei der Entwicklung wurde größter Wert auf Nutzerfreundlichkeit gelegt. Erreicht wurde dies durch eine Abstraktion des Datenmodells, das sich hinter einer einfachen graphischen Nutzeroberfläche verbirgt. Das Innovative ist die Universalität des Datenmodells, die Erweiterbarkeit für beliebige Parameter und die Anwenderfreundlichkeit trotz der dem System wie auch den Daten innewohnenden Komplexität. Die allgemeine Verfügbarkeit über das Internet entspricht dem Stand des heute technisch Möglichen. Das System beinhaltet in seinem gegenwärtigen Stand über 60 000 geologische und hydrographische Stationen aus allen Weltmeeren und ca. 20 000 Datensätze.

Datenmodell

Voraussetzung für die Nutzbarkeit eines vielfältig nutzbaren Informationssystems ist ein einfaches, flexibles Datenmodell, unter dem sich alle wissenschaftlichen Teildisziplinen wiederfinden. Es muß jede Form von Daten aufnehmen können, die an Lokalitäten auf der Erde gebunden sind. Die Raum/Zeitkoordinaten von Proben- oder Meßpunkten sind somit auch die wichtigsten Metadaten zu den eigentlichen Analyse- oder Meßdaten.

Die Hierarchie des Modells bildet sich durch mehrere Ebene ab, die in ihrer Reihenfolge den Weg widerspiegeln, der begangen werden muß, um einen analytischen Wert zu produzieren bzw. zu erhalten (Abb. 1). Verschiedene Projekte oder Institute (PROJECT) sind mit ihren Arbeiten an einem Forschungsbereich beteiligt. Jedes Institut untersucht ausgewählte Gebiete und/oder führt Expeditionen durch (CAMPAIGN). Diese Ebene erlaubt damit eine grundsätzliche Gliederung einer Vielzahl von Proben- bzw. Meßpunkten, um eine bessere Übersicht zu gewährleisten. Der geographische Ort einer Untersuchung findet sich mit allen weiteren begleitenden Metainformationen auf der nächsten Ebene (EVENT). Die optionale Ebene SITE ermöglicht eine Gruppierung von Beprobungspunkten wie sie in einigen Schemata üblich ist. Zu jeder Position lassen sich beliebig viele Meßreihen speichern (DATA), die mit den für ihr Verständnis notwendigen Metainformationen verknüpft sind. Die Ebene SAMPLE erlaubt optional die kuratorielle Verwaltung beliebigen wissenschaftlichen Probenmaterials in Archiven.

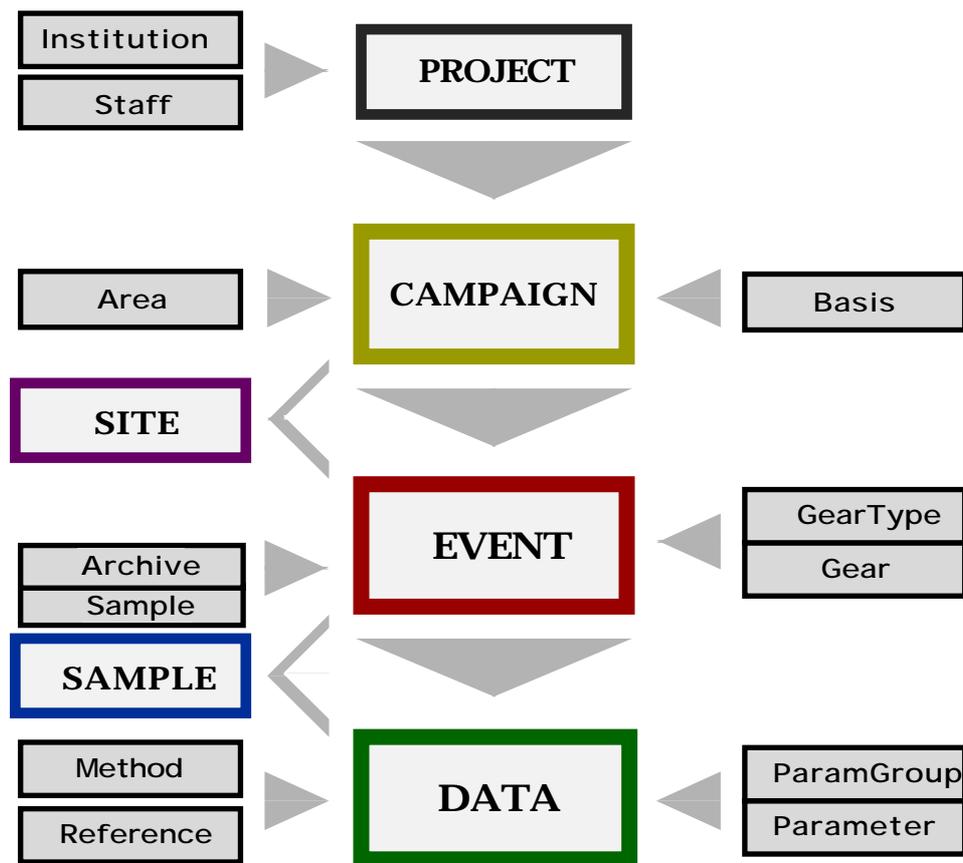


Abb. 1 Datenmodell eines PANGAEA-Systems zur Erfassung, Archivierung und Auswertung von Metainformationen und analytischen Daten aus der Meeresforschung

Das zentrale Modell wird von randlichen Tabellen begleitet, die einzelne Ebenen mit standardisierten Einträgen füllen und beliebig erweiterbar sind. Die Tabellen 'Staff' und 'Institution' beinhalten z.B. Namen und Adressen der am System beteiligten Wissenschaftler. Unter 'Reference' sind die Zitate zu Veröffentlichungen über Daten oder Expeditionen zu finden. 'Parameter' beinhaltet die Namens- und Definitionsliste der speicherbaren analytischen Daten, Parameter sind in Gruppen zusammengefaßt. Alle Tabellen sind jederzeit durch neue Definitionen erweiterbar.

Systembeschreibung

Das PANGAEA-Netzwerk nutzt die Client/Server-Technologie. Das zentrale Datenmodell ist in einer relationalen Datenbank abgebildet. Die Datenbankmanagementsoftware (SYBASE) läuft auf einem DEC-Alpha-Rechner mit 4 Prozessoren und 2 GB Hauptspeicher. Jedes, der an PANGAEA beteiligten Institute betreibt einen Nebenserver (PC) auf dem alle Metadaten gespiegelt sind, um so ein möglichst schnelles Arbeiten innerhalb der Metadaten zu gewährleisten (Abb. 2). Bei Anfragen nach analytischen Daten werden diese über das Internet vom zentralen System angefordert. Die Software für Client und Nebenserver wurde unter 4th-Dimension (ACI) geschrieben. Sie erlaubt auch das Erstellen einer graphischen Oberfläche, wie sie für ein nutzerfreundliches System unabdingbar ist. Die Client-Software ist für MSWindows und MacintoshOS-Systeme verfügbar, der WWW Client wurde als Java-Applet realisiert.

Auf jeder Ebene steht ein Abfragewerkzeug zur Verfügung um Datensätze zu suchen und zu extrahieren; es erlaubt nach einem einheitlichen Schema eine Verknüpfung beliebiger Felder mit variablen Bedingungen. Für die Visualisierung wissenschaftlicher Daten wurden graphische Werkzeuge entwickelt. Das Programm 'PanMap' erlaubt das Zeichnen von Proben-/Meßpunkten in Verbindung mit Metadaten oder analytischen Daten in geographischen Karten, das Programm 'PanPlot' ermöglicht die Darstellung von Parametern gegen die Tiefe/Höhe oder Zeit. Für die Darstellung umfangreicher flächiger Datensätze besteht eine Anbindung an 'Ocean Data View' (ODV). Alle Programme sind sowohl einzeln nutzbar als auch nach einer entsprechenden Datenbankabfrage auf den extrahierten Datensatz anwendbar.

Der Import von Daten erfolgt über standardisierte Tabellen, die jeder Nutzer selbst anlegen kann. Sie entsprechen im wesentlichen dem Format, mit dem Wissenschaftler in Tabellenkalkulationsprogrammen ihre Daten verarbeiten. Die Zuordnung der die Datensätze beschreibenden Metainformationen erledigt eine komfortable Importroutine.

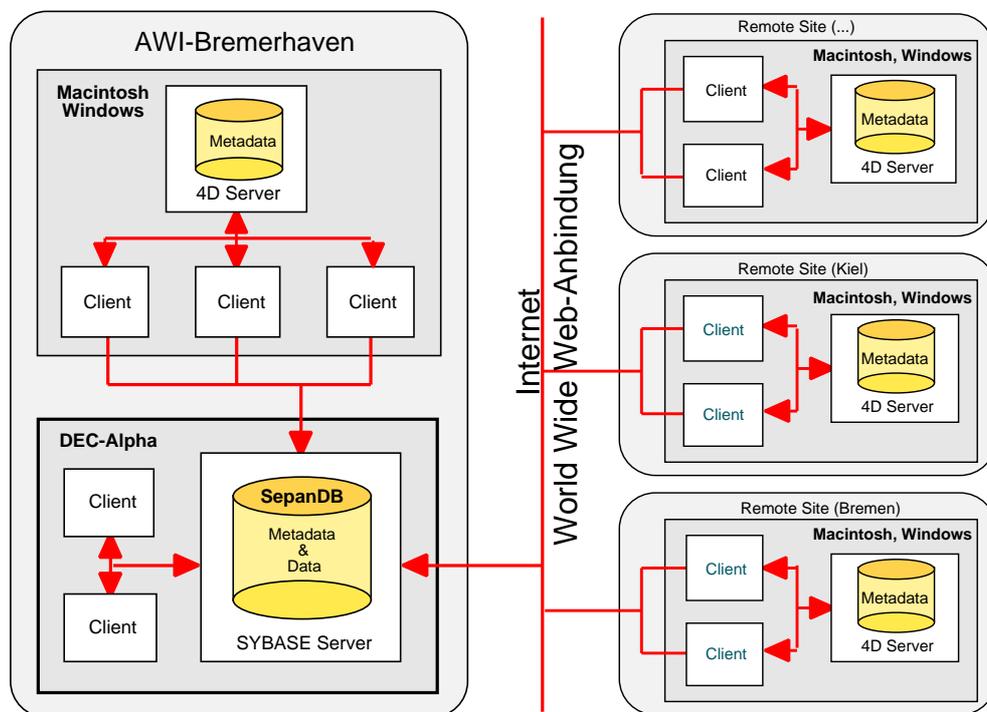


Abb.2 Das Netzwerk eines PANGAEA-Systems erlaubt die Verteilung und den Zugriff von Daten aus der Meeresforschung durch jedes eingebundene Institut

Der Export erfolgt auf den einzelnen Ebenen in konfigurierbaren Texttabellen. Suchwerkzeuge, die auf allen Ebenen die gleiche Funktionalität aufweisen, erlauben die Extraktion beliebiger Teilmengen von Metadaten. Ein umfassendes Suchwerkzeug auf der DATA-Ebene erlaubt komplexe Abfragen auch nach klar definierten Teilmengen größerer analytischer Datensätze (Raum-/Zeitscheiben) und den Export in Verknüpfung mit den gewünschten Metainformationen.

In der Diskussion mit den zukünftigen Nutzern wurde deutlich, daß ein Informationssystem, das nicht ausschließlich der Archivierung von Daten dient, sondern mit dem wissenschaftlich gearbeitet werden soll, auch unpublizierte Daten aufnehmen muß. Es war somit notwendig, einen Datenschutz zu integrieren. Unveröffentlichte Daten sind primär geschützt, Zugriffsrechte können individuell vom Datenproduzenten vergeben werden.

Nicht unerwähnt bleiben soll die kuratorielle Bedeutung des vorgestellten Informationssystems für die allgemeine Verfügbarkeit von Probenmaterial. So lagert weltweit in den Archiven eine Vielfalt von biologischen, hydrographischen und geologischen Proben aus den Ozeanen, die zum Teil mit erheblichem technischem und finanziellem Aufwand gewonnen wurden. Neue, in den letzten Jahrzehnten entwickelte Technologien und Plattformen haben dazu beigetragen, daß das verfügbare Probenmaterial nicht nur an Umfang, sondern auch an Qualität gewonnen hat. Die Archive werden lokal verwaltet und erlauben selten einem Außenstehenden einen Überblick über verfügbares Material. Das vorgestellte System wird auch als Werkzeug gesehen, mit dem sich einzelne Wissenschaftler einen Überblick über die in ihrer Disziplin oder in ihrem Arbeitsgebiet vorhandenen Proben verschaffen können.

Ausblick

Zur Rekonstruktion von Umweltveränderungen in kurzfristigen wie auch geologischen Zeiträumen muß auf gründlich validierte und konsistente Datensätze zugegriffen werden können. Die Modellierung klimarelevanter Prozesse ist nur im überregionalen, häufig nur im globalen Rahmen nachvollziehbar. Die deskriptiv und analytisch arbeitenden Disziplinen liefern die aus dem Probenmaterial rekonstruierten, häufig global verteilten Ausgangsdaten. Sie dienen sowohl als Eckwerte für Modellläufe, als auch zur Verifizierung von Modellergebnissen. In diesem Rahmen kann das System eine zentrale Rolle in der Versorgung der modellierenden Klima- und Umweltforschung mit Daten übernehmen. Voraussetzung hierfür ist allerdings eine deutlich größere inhaltliche Datenbasis.

Der Aufbau und die Stabilisierung eines Informationssystems für Daten aus der Meeresforschung ist sicherlich ein langwieriger Prozeß. Die Implementierung des vorgestellten PANGAEA-Systems in der wissenschaftlichen Welt könnte nicht nur der initiale Schritt für eine fortschrittlichere und effizientere Arbeit in den Naturwissenschaften sein, sondern auch neue methodische Ansätze bei der Interpretation von Daten erlauben.

*Diepenbroek, M, Grobe, H, & Reinke, M (1997)
PANGAEA - Ein Informationssystem für die Meeresforschung,
Deutsche Gesellschaft für Meeresforschung Mitteilungen, 1-2, 94-96 p*