

# PANGAEA

**PaleoNetwork for Geological and Environmental Data**



*Antrag gestellt bei der Phillip Morris Stiftung  
1996*



Hannes Grobe  
Alfred Wegener Institute for Polar and Marine Research,  
Bremerhaven, Germany

## Kurzfassung

Die Erde speichert ihre Klimageschichte in geologischen Archiven (z.B. Sedimente, Korallen, Bäume). Wesentliche Aufgabe der Klimaforschung ist die Entschlüsselung dieser Archive. Die Geowissenschaften leisten damit den historischen Beitrag zur aktuellen Klimadiskussion.

Die in den einzelnen Disziplinen anfallenden Daten, wie auch die umfassenden und vernetzten Fragestellungen fordern ein integrierendes Datenmanagement. Die moderne Rechnertechnologie in Verbindung mit der globalen Vernetzung durch das Internet erlaubt heute den Aufbau eines entsprechenden Systems.

Vorgestellt wird ein Informationssystem für die Paläoklima- und Umweltforschung. Das flexible Datenmodell erlaubt die Erfassung nahezu aller in der naturwissenschaftlichen Grundlagenforschung anfallenden geographisch einzuordnenden Daten. Das Netzwerkkonzept ermöglicht einen allgemeinen Zugriff bei gleichzeitig hoher Funktionalität. Trotz der dem System und den Daten innewohnenden Komplexität wurde größter Wert auf Akzeptanz und damit Nutzerfreundlichkeit gelegt. Erreicht wurde dies durch eine Abstraktion des Datenmodells, das sich hinter einer einfachen graphischen Nutzeroberfläche verbirgt.

Das System für marine Daten (SEPAN) ist aktiviert und läuft zur Zeit zwischen fünf Instituten in Deutschland. Weitere Institute werden angeschlossen, Daten aus den einzelnen Arbeitsgruppen werden eingelesen. Das System beinhaltet in seinem gegenwärtigen Stand 60 000 geologische Stationen und 40 000 Datensätze.

# PANGAEA (PaleoNetwork for Geological and Environmental Data)

## Entwicklungsbeschreibung

### 1 Motivation, Stellung im Rahmen der Klimaforschung

Diskussionen über mögliche, durch den Menschen verursachte Änderungen des Weltklimas, vorwiegend bedingt durch eine Verstärkung des natürlichen Treibhauseffektes, sind allgegenwärtig. Klimarelevante Fragestellungen haben daher einen zunehmend größeren Anteil an der naturwissenschaftlichen Grundlagenforschung. In internationalen Programmen werden die einzelnen Bereiche der Erde (Bio-, Atmo-, Hydro-, Kryo- und Geosphäre) interdisziplinär angegangen, um so das komplexe Klimasystem besser verstehen zu lernen. Ein wesentlicher Schlüssel, sowohl für das Verständnis des gegenwärtigen Klimas, als auch für Prognosen zukünftiger Entwicklungen, liegt in der Erfassung und Quantifizierung vergangener Klimaabläufe. Aus ihrer Rekonstruktion und der daran anschließenden Interpretation erfährt die Wissenschaft etwas über das Verhalten der verschiedenen oben genannten Bereiche der Erde während natürlicher Klimavariationen.

Klimadatenreihen decken, im Gegensatz zu Wetterdaten, langfristige Zeiträume ab. Somit ist die Zeit die Achse, auf der sich alle in der Klimaforschung gewonnenen Meßwerte abbilden. Die untersuchten Zeiträume umfassen mehrere Größenordnungen von Jahrzehnten über geschichtliche Epochen bis hin zu geologischen Zeiträumen. Zur Rekonstruktion der Klimaentwicklung in den einzelnen, unterschiedlich langen Zeitreihen bemüht die Wissenschaft Archive unterschiedlichster Art (Tab. 2).

Seit einigen Jahrzehnten werden Wetterdaten hochauflösend erfaßt und ermöglichen damit für den kurzen Zeitraum der jüngsten Geschichte auch eine detaillierte Rekonstruktion der Klimaentwicklung. Deutlich schwächer wird die Quelle der für das Klimaverständnis direkt nutzbaren, vom Menschen niedergeschriebenen, historischen Daten, wenn man sich in der geschichtlichen Zeit zurückbewegt. Hier liefern bereits Archive, die die Natur zur Verfügung stellt (Bäume, Korallen), vollständigere Datenreihen. In vorgeschichtlicher Zeit verschiebt sich die Auswahl an Informationsquellen noch weiter zugunsten natürlicher Klimaaufzeichnungen (Gletscher, Seen). Schließlich bieten ausschließlich erdgeschichtliche Archive geologische Abfolgen, in denen die Klimaentwicklung in geologischen Zeiträumen dokumentiert ist (Eisschilde, Ozeansedimente).

Eine wesentliche Aufgabe der Klimaforschung ist es, unter Verwendung der verfügbaren Archive Klimaveränderungen in historischen und geologischen Zeiträumen und die daran gekoppelten Änderungen in der Umwelt zu rekonstruieren. Es ist offensichtlich, daß das zentrale Werkzeug für diese umfassende Aufgabe nur ein Informationssystem sein kann, daß Daten aus allen genannten wissenschaftlichen Bereichen, Archiven und Zeiträumen von den aktuellen Umweltdaten bis zu geologischen Zeitreihen in ihrer Vielfalt in einem einheitlichen Format abbildet und allgemein verfügbar hält. Eine Interpretation dieser komplexen Datensätze ist eine Herausforderung. Sie ist aber überhaupt nur möglich, wenn ein entsprechendes System mit auf die wissenschaftlichen Fragestellungen abgestimmten Zugriffswerkzeugen die Auswertung unterstützt.

## 2 Fragestellung und wissenschaftlicher Nutzen

Der Weg vom Meßpunkt oder Probenmaterial zu einem validierten Wert, der in numerischer Form vergangene Klima- oder Umweltverhältnisse widerspiegelt, ist langwierig, vielschichtig und kostenintensiv. In den verschiedenen naturwissenschaftlichen Disziplinen ist die Vielfalt der Parameter, wie auch ihre informationstechnische Organisation unterschiedlich. Sie reicht von einer zentralen, durchorganisierten Erfassung, wie sie z.B. für meteorologische Daten zwangsläufig als Ausgangsprodukt für den täglichen Wetterbericht notwendig ist, bis zu individuellen Insellösungen auf den Arbeitsplatzrechnern einzelner Wissenschaftler. Die Ordnung war bisher im wesentlichen abhängig von der Fragestellung, die hinter der Datenerfassung steht.

Zum einen wird durch die stetig steigende Datenfülle in allen Disziplinen den Wissenschaftlern zunehmend die Notwendigkeit für ein integrierendes Datenmanagement bewußt, zum anderen fordern auch die neuen Fragestellungen im Rahmen der Klimaforschung eine bessere Datenorganisation und -verfügbarkeit. Bisher ist offensichtlich, daß umfangreiche analytische Datensätze der wissenschaftlichen Gemeinschaft verloren gehen, da sie nicht zentral verfügbar gemacht werden. Im günstigsten Fall werden die Daten in einer Publikation in Tabellenform abgedruckt; nur ein geringer Prozentsatz von Daten wird auf allgemein zugänglichen Speichern abgelegt, jedoch auch dort nur in tabellarischer Form.

In einer modernen Forschungslandschaft sollte jeder Wissenschaftler problemlos auf die von ihm benötigten Datensätze zugreifen können. Die informationstechnische Infrastruktur erlaubt dies seit einigen Jahren; somit bestehen, neben dem dringenden Bedarf, auch die technischen Möglichkeiten. Es ist daher jetzt an der Zeit entsprechende Systeme zu entwickeln, zu

implementieren und zu etablieren. Das vorgestellte System soll im Rahmen der Klima- und Umweltforschung exemplarisches leisten.

### 3 PANGAEA - Initiative und Konzept

Vor dem Hintergrund oben angesprochener Bedürfnisse und Notwendigkeiten entstand 1993 ein Projekt zur Entwicklung eines Informationssystems für Paläoklimadaten. Es wurde durch an der Paläoklimaforschung beteiligte Wissenschaftler in Deutschland initiiert. Eine Anschubfinanzierung bewilligte auf Antrag der Bundesminister für Bildung, Wissenschaft, Forschung und Technologie (BMBF) für drei Jahre. Auf einer nach der Hälfte des Projektes geforderten Zwischenbegutachtung stimmte die Kommission aus Gutachtern und zukünftigen Nutzern uneingeschränkt für seine Fortsetzung. Das System wurde von einer Gruppe von Wissenschaftlern und Informatikern am Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI), Bremerhaven, in Zusammenarbeit mit anderen, an der Paläoklimaforschung beteiligten Instituten (Tab. 1) entwickelt. Das Gesamtsystem wurde, in Anlehnung an seine Struktur, nach dem Superkontinent benannt, in dem alle Kontinente vor 200 Millionen Jahren vereint waren: PANGAEA (PaleoNetwork for Geological and Environmental Data); das erste Teilsystem für marine Daten erhielt die Bezeichnung SEPAN (Sediment and Paleoclimate Data Network), das zweite System LATIN (Lake-Terra-Ice Data Network) wird Daten aus terrestrischen Klimaarchiven (Eis, Sedimente, Seen) verarbeiten, ein drittes System ARON (Atmospheric Research and Observation Data Network) ist für Daten aus der Atmosphärenforschung vorgesehen.

Für viele an der Klimaforschung beteiligte Wissenschaftler sind Computer nicht das wichtigste analytische Arbeitsgerät. Ein Rechner wird vorwiegend benutzt um Daten auszuwerten und zu publizieren. Die Motivation zur Nutzung komplexer Software, zusätzlich zu den unbedingt notwendigen Programmen, ist gering, das notwendige Wissen um den informationstechnischen Hintergrund fehlt. Bei der Entwicklung eines Informationssystems, das den Wünschen des Nutzers folgt, flexibel und erweiterbar ist und intuitiv und unkompliziert bedient werden kann, ist daher Nutzerfreundlichkeit unabdingbare Voraussetzung für seine Akzeptanz. Eine Vielzahl erfolgloser Versuche, entsprechende Systeme aufzubauen, zeigt die Schwierigkeit, hier das richtige Konzept zu finden.

Entscheidend für die Nutzbarkeit eines Informationssystems ist die Abbildung der heterogenen wissenschaftlichen Umgebung im Datenmodell, in den Nutzerschnittstellen und in den allgemeinen Werkzeugen. Diese Teile müssen so weit generalisiert werden, daß sich

alle Beteiligten im System wiederfinden, gleichzeitig das System aber noch die spezifischen Anforderungen der einzelnen Disziplinen erkennen läßt. Die Abstraktion der Umgebung, in der wissenschaftliche Informationen abgebildet werden, bei gleichzeitiger Umsetzung in ein von jedem nutzbares System, ist der Schlüssel zur Akzeptanz. Schnittstellen zur Anbindung individueller Erweiterungen können trotzdem die Nutzung fachspezifischer Software erlauben.

Das Innovative des vorgestellten Systems ist die Universalität des Datenmodells, die Erweiterbarkeit für beliebige Parameter und die Anwenderfreundlichkeit trotz der dem System wie auch den Daten innewohnenden Komplexität. Bei der Konzeption des vorgestellten Systems wurde, neben einer ausgefeilten Datenhaltung, größter Wert auf Nutzbarkeit und Akzeptanz gelegt. Die allgemeine Verfügbarkeit über das Internet entspricht dem Stand des heute technisch Möglichen. Damit stellt das System der Forschung eine bisher einmalige Arbeitsplattform zur Verfügung.

## 4 Datenmodell

Da es auch mit modernster Informationstechnologie nicht möglich und auch nicht sinnvoll ist alle Daten zur Klimaforschung in einem System zu halten, finden sich im PANGAEA-Netzwerk verschiedene Teilsysteme, die einen bestimmten, wissenschaftlich abgrenzbaren Bereich beinhalten (Abb. 1). Jedes einzelne Teilsystem ist im Programmcode identisch um den Aufwand für Pflege und Management möglichst gering zu halten. Voraussetzung für die Nutzbarkeit eines entsprechenden Systems 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 und gegen räumliche oder zeitliche Dimensionen darstellbar sind. Dies trifft für den weitaus größten Teil aller Untersuchungen zu; auch sind die Positionen von Proben- oder Meßpunkten die wichtigsten Metadaten (=beschreibende Daten, s. Tab. 3) zu den eigentlichen analytischen oder gemessenen Daten.

Somit steht in dem hierarchisch gegliederten Modell der wissenschaftliche Bereich, aus dem die Daten stammen, an oberster Stelle. Beim Einstieg in das System entscheidet der Nutzer, mit welchem Bereich er arbeiten möchte; in diesem Teilsystem kann er sich frei bewegen. In dem diesem Teilsystem zugrunde liegenden Datenmodell (Abb. 2) entspricht die Hierarchie dem Weg, den ein Wissenschaftler in der Regel gehen muß, um einen spezifischen Meßwert zu produzieren oder zu erhalten.

Die Hierarchie des Modells bildet sich durch mehrere Ebene ab. Verschiedene Projekte oder Institute (PROJECT) sind mit ihren Arbeiten an einem Forschungsbereich beteiligt. Jedes Institut untersucht ausgewählte Gebiete während bestimmter Unternehmungen (CAMPAIGN). Diese Ebene, die auch Länder, Profile oder Expeditionen enthalten kann, erlaubt eine zusätzliche Gliederung einer Vielzahl von Proben- bzw. Meßpunkten, um so eine bessere Übersicht und damit später Abfragemöglichkeit zu gewährleisten. Der geographische Ort einer Untersuchung findet sich mit allen weiteren begleitenden Metainformationen auf der nächsten Ebene (SITE). Die optionale Ebene STATION ermöglicht eine Gruppierung von Beprobungspunkten wie sie in einigen Schemata üblich ist. Zu jeder Position lassen sich beliebig viele Meßreihen speichern (DATA). Auf dieser untersten Ebene sind die eigentlichen Daten mit den für ihr Verständnis notwendigen Metainformationen zu finden. Die Ebene CURATOR erlaubt optional die kuratorielle Verwaltung beliebigen wissenschaftlichen Probenmaterials in Archiven. Proben lassen sich mit Informationen und Daten in anderen Ebenen verknüpfen.

Das zentrale Modell wird von randlichen Tabellen begleitet (Abb. 2), die die einzelnen Ebenen mit standardisierten Einträgen füllen und beliebig erweiterbar sind. Die Tabellen 'Staff' und 'Institution' beinhalten z.B. Namen und Adressen der am System beteiligten WissenschaftlerInnen. Unter 'Reference' sind die Zitate zu Veröffentlichungen über Daten oder Expeditionen zu finden. Die Tabelle 'Parameter' beinhaltet die Namens- und Definitionsliste der speicherbaren analytischen Daten (Tab. 4). Parameter sind in Gruppen zusammengefaßt, analytische Daten werden in mehrere Kategorien untergliedert (Tab. 3).

## 5 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 nochmals gespeichert sind, um so ein möglichst schnelles Arbeiten innerhalb der Metadaten zu gewährleisten (Abb. 3). Nur bei einer Anfrage nach analytischen Daten werden diese über das Internet vom zentralen System angefordert. Die Software für Client und Nebenserver wurde unter 4th-Dimension (ACI) geschrieben und läuft unter MSWindows und MacintoshOS. Sie erlaubt auch das Erstellen einer graphischen Oberfläche (Abb. 4), wie sie für ein nutzerfreundliches System unabdingbar ist.

Das Informationssystem erlaubt Einbindung und Aufruf allgemeiner und fachspezifischer Anwendungen. Für die Darstellung wissenschaftlicher Daten wurden zwei graphische Werkzeuge entwickelt. Das Programm *PanMap* erlaubt das Zeichnen von Proben-/Meßpunkten mit Metadaten oder analytischen Daten in geographische Karten (Abb. 5, 6), das Programm *PanPlot* ermöglicht die Darstellung von Parametern gegen die Tiefe oder Zeit in verschiedenen Darstellungsformen (Abb. 7). Für die Digitalisierung von Karten zur Weiterverwendung in PanMap steht ein Digitalisierwerkzeug zur Verfügung. Alle Programme sind einzeln nutzbar, oder, nach einer entsprechenden Datenbankabfrage, aus dem System heraus aufrufbar und automatisch auf den extrahierten Datensatz anwendbar. Es existiert bereits ein Sammlung weiterer fachspezifischer Werkzeuge (Diffraktogrammauswertung, Korngrößenberechnung, Zugriff auf World Ocean Atlas) auf die hier nicht weiter eingegangen wird; das Funktionsprinzip ist jedoch ähnlich. Zum Export von Datensätze steht auf jeder Ebene ein Abfrage- und Konfigurationswerkzeug zur Verfügung, das nach einem einheitlichen Schema eine Verknüpfung beliebiger Felder mit variablen Bedingungen im gewünschten Format erlaubt.

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 wichtige Zuordnung der Metainformationen zu den einzelnen Datensätzen erledigt eine komfortable Importroutine.

Der Export von Daten erfolgt auf den einzelnen Ebenen in konfigurierbaren Texttabellen. Suchwerkzeuge, die auf allen Ebenen die gleiche Funktionalität aufweisen, erlauben die Extraktion beliebiger Teilmengen. Ein umfassendes Suchwerkzeug auf der DATA-Ebene erlaubt komplexe Abfragen auch nach klar definierten Teilmengen größerer Datensätze (z.B. 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. Die unveröffentlichten Daten eines jeden Anwenders sind daher geschützt, Zugriffsrechte können individuell vom Datenproduzenten vergeben werden.

## 6 Das System im Betrieb

Implementierung und Test des Informationssystems erfolgte mit Klimadaten aus marinen Sedimenten. Der erste Schritt, das System mit Daten zu füllen, war die Übernahme der Metadaten des in Deutschland vorhandenen, marin-geologischen Probenmaterials. Es folgten Daten aus den großen Kernlager der Vereinigten Staaten, deren Probenmaterial international Verwendung findet und aus dem zahlreiche Datensätze in der Literatur vorliegen. Auch die Metainformationen zu den Bohrungen des internationalen 'Deep Sea Drilling Project' (DSDP) und seinem Nachfolger, dem 'Ocean Drilling Project' (ODP) liegen im System vor. Aus anderen Bereichen geologischer Klimaarchive wurden die wichtigsten Eiskerne und ein Teil der von deutschen Gruppen erbohrten Seesedimente importiert. Ozeanographisch Daten liegen von etwa 4000 Stationen vor. Bisher sind im System Metainformationen zu 70 000 Proben- und Meßstationen gespeichert. Die Anzahl analytischer Datensätze (bisher ca. 50.000) wird sich nach Anschluß weiterer, am Netzwerk beteiligter Institute mit dem Import der dort bereits vorliegenden Daten um ein Vielfaches erhöhen.

Das System wird, neben der routinemäßigen Datenhaltung für die beteiligten Institute, eingesetzt für das Datenmanagement des Baikal Drilling Project, das Cape Roberts Projekt zu Tiefbohrungen in der Antarktis, für verschiedene Sonderforschungsbereiche, ein DFG Schwerpunktprogramm, für Kooperationsprojekte zwischen Deutschland, Skandinavien und Rußland im Nordpolarmeer und für die vom BMBF geförderte Paläoklimaforschung sowie für die deutsche Polarforschung. Für das Institut für Ostseeforschung wurden alle geologische Daten der Ostsee übernommen, die während der DDR produziert wurden. Das PANGAEA-system ist am Projektdatenmanagement für EU-Projekte beteiligt und ist eingebunden in das neue Abkommen zur Kooperation zwischen holländischen und deutschen Forschungseinrichtungen im Land Bremen.

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

Nicht unerwähnt bleiben soll die kuratorielle Bedeutung des vorgestellten Informationssystems für die allgemeine Verfügbarkeit von Probenmaterial. So lagern

weltweit in den Archiven sehr große Probenmengen aus den verschiedenen Bereichen natürlicher Klimaarchive, 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 ihrem Arbeitsgebiet vorhandenen Proben verschaffen können.

## 6 Ausblick

Alle bisher noch in einem Teilsystem (SEPAN) gehaltenen Daten aus verschiedenen natürlichen Klimaarchiven werden zukünftig auf entsprechende identische Teilsystem verteilt. Mit Anwachsen der Datenmenge würde für besonders umfangreiche Projekte das gesamte System einschließlich aller Metainformationen an anderen Instituten gespiegelt. Eine entsprechende Verwaltung der dann verteilt gehaltenen analytischen Daten übernehme ein Navigation-Server. Die Spiegelung von Daten würde auch notwendig, falls das System über interkontinentale Entfernungen betrieben werden sollte, da das zunehmend belastete Internet hier keine ausreichenden Direktzugriffszeiten mehr zuließe. Es besteht bekundetes Interesse seitens mehrerer amerikanischer Institute, dieses System zu nutzen, da auch in den U.S.A. nichts vergleichbares verfügbar ist.

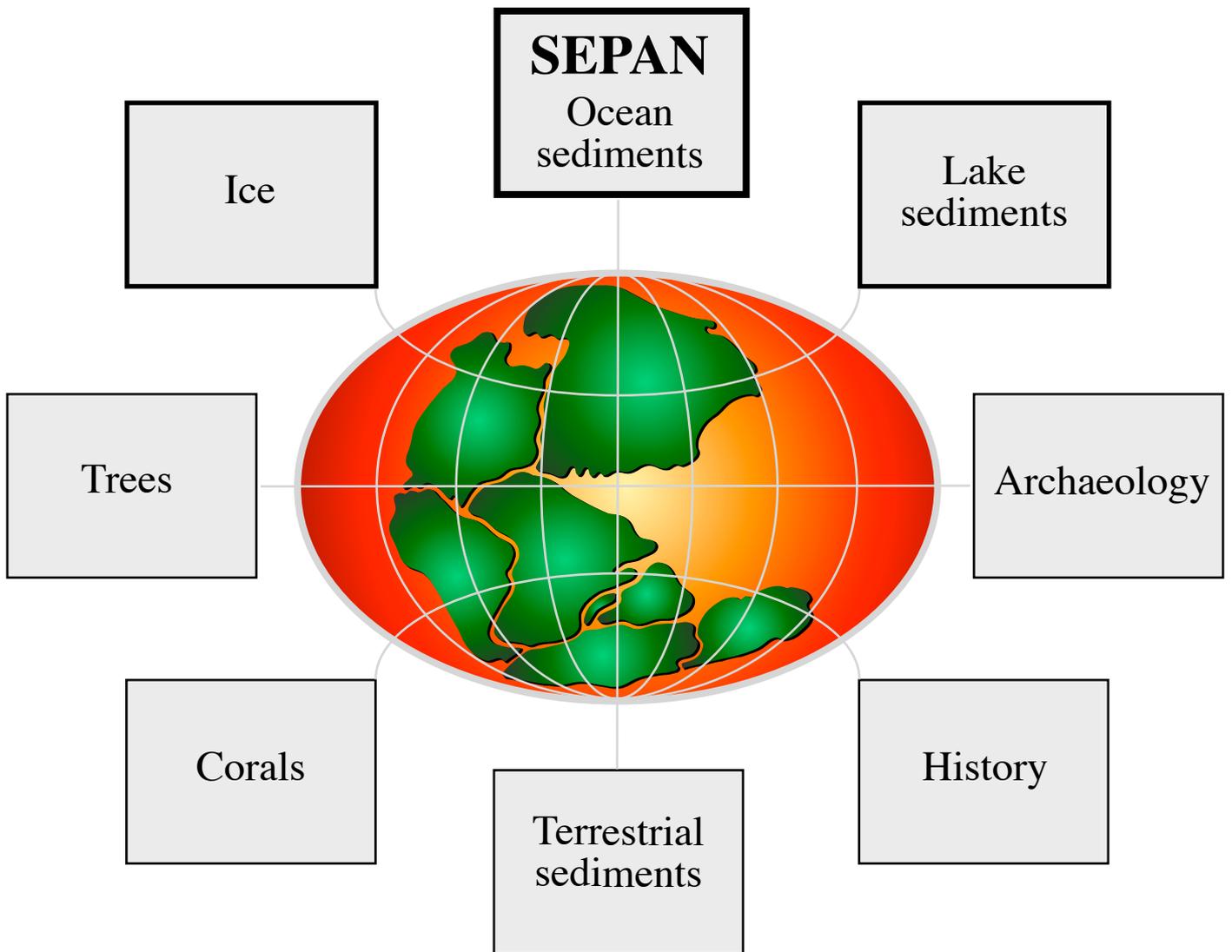
Das Prinzip eines Client/Server-Systems fordert, daß jeder Nutzer auf seinem Arbeitsplatzrechner die systemspezifische Client-Software betreibt. Die von der Wissenschaft geforderte Funktionalität des Informationssystems erlaubt bisher keinen allgemeinen Zugriff über eine World Wide Web (WWW)-Schnittstelle. Eine erste experimentelle Version wurde unter Java realisiert, die es erlaubt, über einen WWW-Browser auf die Metadaten von PANGAEA zuzugreifen, zur Zeit erfolgt eine Erweiterung auch auf die analytischen Daten. Zukünftige Programmversionen sollen durch optionale Kompilierung die volle Funktionalität der Client-Software im Web abbilden können.

Ein Sammelpunkt für Klima- und Umweltdaten sind die *World Data Center* (WDC); verantwortlich für Paläoklimadaten ist zum Beispiel das WDC-A for Paleoclimatology in Boulder, Colorado. Hier werden Daten, meist verknüpft mit einem Literaturzitat, in einem Filesystem abgelegt und sind über das Internet abrufbar. Datenbankfunktionalität im Sinne

des hier vorgestellten Systems ist nicht verfügbar. Auch lagern hier lediglich einige hundert Datensätze, da die Motivation der Wissenschaftler für die Unterstützung eines solchen Systems gering ist. Mit dem WDC-A besteht ein reger Informationsaustausch, auch da dies ein zukünftiger Knoten in einem globalen Datennetzwerk werden könnte. Weiterhin existieren einige kleine Datenbanken als individuelle Insellösungen, die jedoch weder allgemein verfügbar sind, noch ohne detaillierte Vorkenntnisse genutzt werden können. Zumindest liegen hier Daten bereits in einer konsistenten Form vor. Grundsätzlich besteht Zusammenarbeit und Informationsaustausch mit vielen Einrichtungen, die in die Klimaforschung eingebunden sind, insbesondere mit dem Deutschen Klimarechenzentrum (DKRZ) in Hamburg, das Klimadaten aus Modellen speichert und der Datenbank für terrestrische Daten (PKDB) in Hohenheim.

Ein besonderer Schwerpunkt soll zukünftig auf das Datenmanagement für EU-Projekte gelegt werden. Da die EU für internationale Projekte einen entsprechenden Teil in den Anträgen vorsieht, entwickelt sich zur Zeit ein Markt, in dem Institutionen diese Arbeiten als Dienstleister für Projekte anbieten. Ein erster Schritt, der hier unternommen wurde, ist die Veröffentlichung der aus einem Projekt resultierenden Daten in elektronischer Form, zumeist auf CD-ROM. Damit besteht zwar nicht mehr die Gefahr eines Totalverlustes der Daten und die Daten liegen innerhalb des Projektes konsistent vor, sie sind jedoch nicht jederzeit und allgemein verfügbar und liegen, in Relation zu anderen Projekten, nicht in einer konsistenten Form vor. Erst durch die in PANGAEA betriebene Art der Datenhaltung erhalten wissenschaftliche Daten jedoch den für überregionale und umfassende Auswertungen notwendigen 'Mehrwert'.

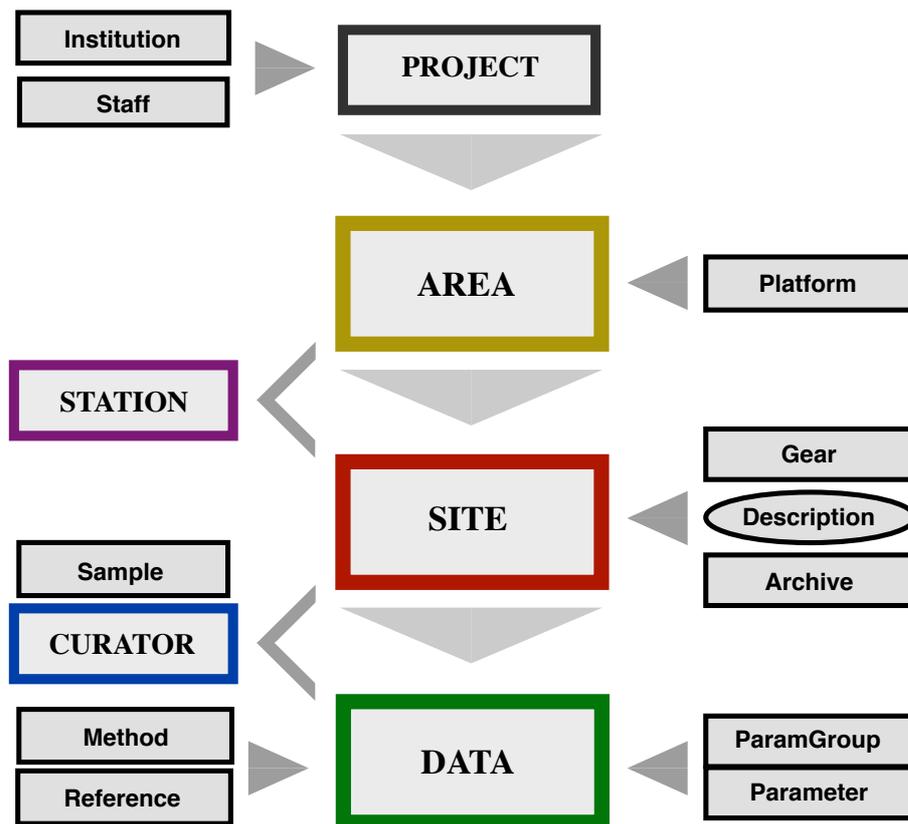
Der Aufbau und die Stabilisierung eines umfassenden Informationssystems für einen Teilbereich der naturwissenschaftlichen Grundlagenforschung ist sicherlich ein langwieriger Prozeß. Die Implementierung des vorgestellten PANGAEA-Systems in der wissenschaftlichen Welt könnte nicht nur der initiale Schritt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



**Structure of the information system PANGAEA  
(PaleoNetwork for Geological and Environmental Data)**

A network of single cloned systems is used to store paleoclimatic data from different historical or geological archives. Each system has the same internal structure and data model, thus only one system has to be managed. All subsystems will be linked through a common tool, which allows access to all data archives.

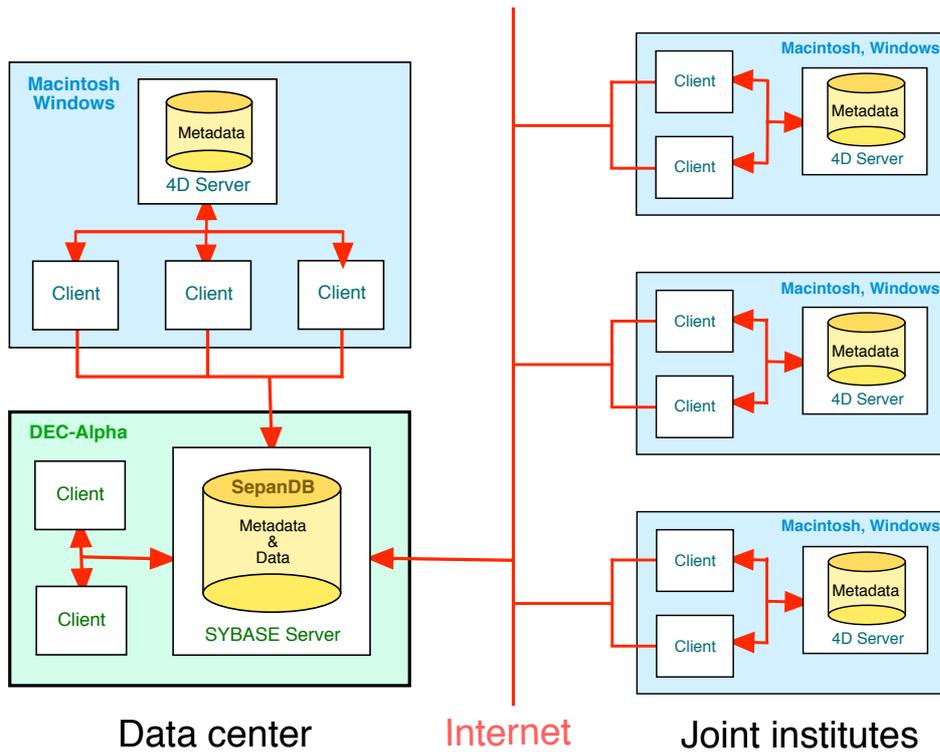
A prototype of the data model is realized and filled with paleoclimatic data from marine sediments (SEPAN, Sediment and Paleoclimate Data Network, 60 000 sites, 20 000 data sets). The flexibility of the model is tested by storing data from oceans, ice and lake sediments in a cloned system.



### Data model of a clone (e.g. SEPAN)

The same flexible and simple relational data model is used for all clones within the PANGAEA network. The model follows the way, data are produced. Different institutions/projects (PROJECT) carry out investigations in specific areas or on expeditions (AREA). Within that area at different locations (STATION) a number of samples (SITE) may be taken. At distinct points a core or profile is sampled/measured for different investigations (CURATOR). From each sample analytical data (DATA) are produced.

Lists including standardized metainformation are connected to the main data fields (e.g. gear, platform, method). The parameter list is gathered up into parameter groups. Data types are defined as primary, secondary, and tertiary data. New parameters can be defined by the user at any time. All analytical data sets are linked to the related metadata (e.g. latitude, longitude, depth, expedition, reference).



## Network concept

SEPAN uses client/server technology through the Internet via on-line connection to the data center. The database software (SYBASE) runs on a DEC Alpha under OSF. The user friendly client software for access to the server (4th Dimension) is designed for Macintosh and Windows platforms. To improve the access speed, all information down to the data series description (metainformation) is mirrored on local 4D-servers, which are also running on PC platforms. The mutual update of metainformation is made in the background. Only on request of analytical data the local server has direct access to the main server at the data center.

GeoB1115-3.4 **GEOB** Data points 151

**Calcium carbonate** %

Group Chemistry, sediment  
 Type numeric Status published  
 Method **Element analysis LECO**  
 Input date Oct 2 1996 4:35PM Last update

**Principal investigator**  
 Torsten Bickert

**Reference**  
 Bickert, T (1992) Rekonstruktion der spätquartären Bodenwasserzirkulation im östlichen Südatlantik über stabile Isotope benthischer Foraminiferen, Berichte, Fachbereich Geowissenschaften, Universität Bremen, 27, 205 p

Depth	Calcium carbonate (70)
0.030	94.30
0.080	94.10
0.130	94.80
0.180	93.80
0.230	94.80
0.280	95.10
0.330	93.60
0.380	93.70
0.430	91.70
0.480	93.40
0.530	91.30
0.580	92.00
0.630	93.40
0.680	93.30
0.730	92.70
0.780	93.00
0.830	91.20

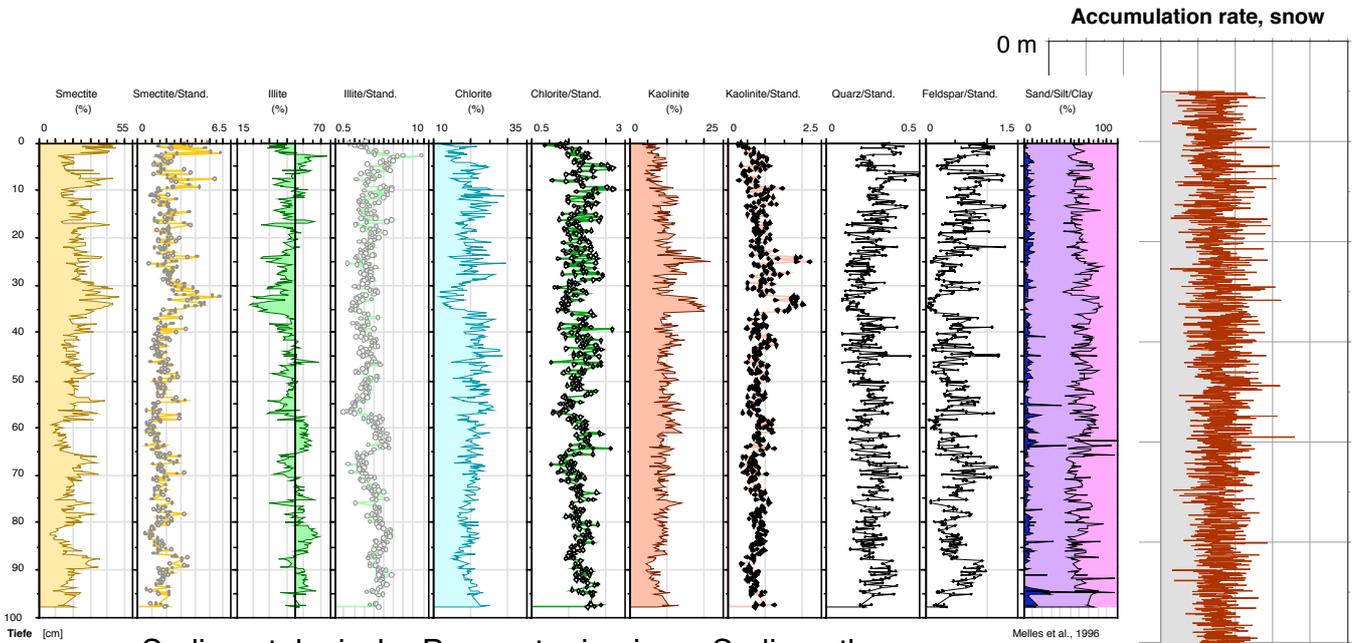
Status selection

Comment  
 no data between 3.83 and 6.53 m (Turbidite!)

<< < > >> Delete Cancel Save

## DATA Window

The data window, which opens when choosing DATA from the data model, allows access to the analytical data. The window provides all meta-information (reference, method, comment) necessary for the understanding of the data. The principal investigator is able to give access rights for unpublished data to selected users. Data with the status 'published' can be accessed by everyone, only the reference has to be cited. The data can be edited and a status can be given to each value to mark suspect data. The parameter name is related to a list, where each parameter is well defined. Icons allow the export of data and call plot software for graphical presentation.



Sedimentologische Parameter in einem Sedimentkern

Schneeakkumulationsraten im GRIP-Eiskern  
(Greenland Ice Core Project)

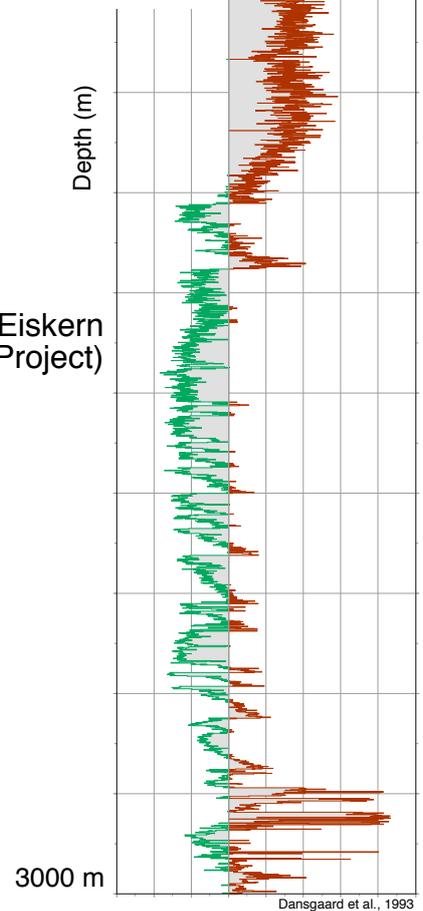


Abb. 7: Graphisches Werkzeug PanPlot

PanPlot erlaubt die Darstellung von bis zu 60 Parametern gegen die Tiefe oder Zeit in einem Profil/Kern oder die Darstellung von einem Parameter in entsprechend vielen Profilen. Graphische Attribute der einzelnen Datensätze können individuell eingestellt werden, die Graphiken können in andere Programme überführt werden. Das Programm wird, nach einer entsprechenden Abfrage, von der Datenbank aufgerufen, die Daten werden übergeben und dargestellt. Die Darstellung einer Vielzahl von Parametern ist eine häufige Anwendung aller beteiligten Disziplinen. Eine Software, die entsprechendes schnell und unkompliziert leistet, ist am Markt nicht verfügbar.

Mitarbeiter der Projektgruppe "Paläoklimadatenbank"  
am Alfred-Wegener-Institut für Polar- und Meeresforschung, 27568 Bremerhaven  
Dr. Michael Diepenbroek Geologe, Informatiker  
Dr. Hannes Grobe Geologe, Projektleiter  
Dr. Manfred Reinke Biologe, verantw. für Datenbanken am AWI  
Dr. Reiner Schlitzer Physiker, modellierender Ozeanograph  
Dr. Rainer Sieger Mathematiker  
Uwe Siems Informatiker

*e-mail: sepan@awi-bremerhaven.de*

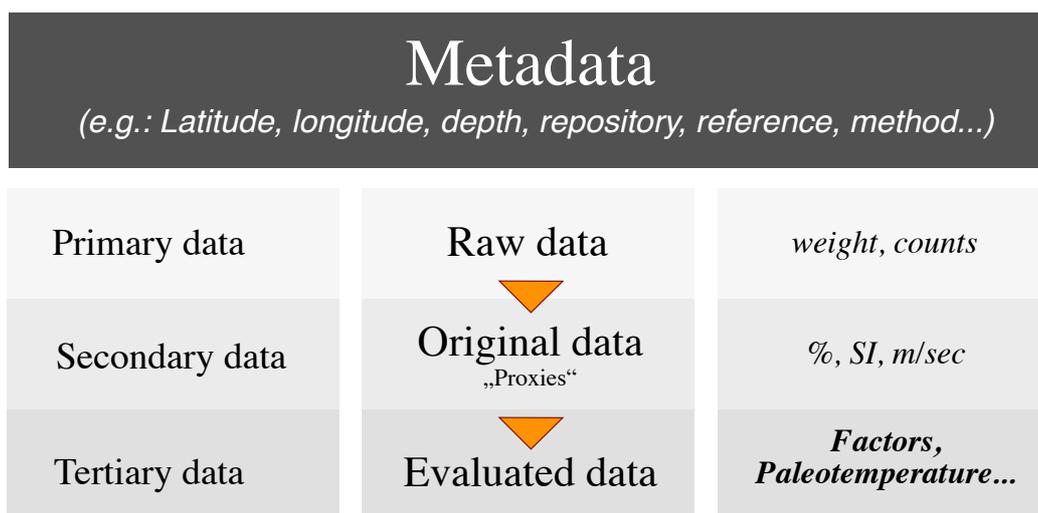
Tab. 1: Liste der bisher in das SEPAN-Netzwerk eingebundenen Institute und Projekte. Fettgedruckt sind Einrichtungen, zwischen denen bereits eine Netzwerkverbindung besteht, kursiv dargestellte Institute werden in nächster Zukunft eingebunden, weitere Projekte und Institute sind in der Diskussion. SEPAN, als das marine Teilsystem von PANGAEA, beinhaltet auch die Metadaten zu meeresgeologischen Proben der wichtigsten Institute aus den U.S.A. und Kanada. (Quelle: Core Curator Index, CCI). Aufgeführt ist die Anzahl der aus den einzelnen Instituten gespeicherten Probenstationen.

Institut/Projekt	Stationen	Quelle	Teilnahme
Akademie der Wissenschaften, Universität Heidelberg			geplant
<b>Alfred-Wegener-Institut, Bremerhaven</b>	1870	AWI	vernetzt
<b>Alfred-Wegener-Institut, Potsdam</b>	226	AWI	vernetzt
Antarctic Research Facility, Florida State University, Tallahassee	2272	CCI	geplant
Atlantic Geoscience Center, Bedford Inst. of Oceanography, Canada		AGC	
Australian Geological Survey Organisation, Canberra	51	AGSO	
Baikal Drilling Project	1		geplant
Biogeochemical Ocean Flux Study, British Oceanographic Data Center	46	BOFS	
British Antarctic Survey			diskutiert
<i>Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover</i>	1361	BGR	zugesagt
Cape Roberts Project			geplant
Cooperation with Russian Institutes	391	AWI	
Deep Sea Drilling Project	620	DSDP	
Department of Geology, Oregon State University	2667	CCI	
<b>Fachbereich Geowissenschaften, Universität Bremen</b>	778	FGB	vernetzt
Geochemical Ocean Section Studies			
<i>Geoforschungszentrum Potsdam</i>			zugesagt
<b>Geologisch-Paläontologisches Institut, Universität Kiel</b>	2446	GIK	vernetzt
Geologisch-Paläontologisches Institut, Universität Göttingen			diskutiert
Geologisch-Paläontologisches Institut, Universität Tübingen			beantragt
<b>GEOMAR-Forschungszentrum für marine Geowissenschaften, Kiel</b>			vernetzt
Institute für Biogeochemie und Meereschemie, Universität Hamburg			diskutiert
<i>Institut für Ostseeforschung, Warnemünde</i>			zugesagt
Institut für Petrographie und Geochemie, Universität Karlsruhe	121	IPGK	
Lamont-Doherty Earth Observatory, Columbia University	7878	CCI	
Ocean Drilling Program, College Station, USA	323	ODP	
Rosenstiel School of Marine & Atmospheric Sciences, Univ. of Miami, USA	1453	CCI	
Scripps Institution of Oceanography, USA	3455	CCI	
Southampton Oceanographic Center, GB			geplant
Woods Hole Oceanographic Institution, Dept. of Geology & Geophysics, USA	2451	CCI	
World Data Center-A for Paleoclimatology, Boulder, Colorado, USA			Kooperat.
DFG/SPP Wandel der Geo-Biosphäre in den letzten 15 000 Jahren			geplant

Klima-Archiv	Zeitraum (Jahre)	Auflösung
Historische Aufzeichnungen	2000	Stunden/Tage
Archäologie	20 000	10 Jahre
Gletscher	10 000	Jahre
Baumringe	12 000	Jahreszeiten
Pollen	100 000	1000 Jahre
Korallen	100 000	Jahreszeiten
Paläoböden	100 000	100 Jahre
polare Eiskerne	300 000	Jahre
See-Sedimente	1 Million	Jahre
Löss	1 Million	10 Jahre
Ozean-Sedimente	100 Millionen	100 Jahre

Tab. 2: Historische und natürliche Klimaarchive

Klimaveränderungen verlaufen naturgemäß in langfristigen bis geologischen Zeiträumen. Für das Verständnis des Klimageschehens auf der Erde ist es daher unabdingbar das Klima der Vergangenheit aus den verfügbaren Archiven zu rekonstruieren. Aufgeführt sind die wichtigsten Archive mit ihrer zeitlichen Reichweite und Auflösung. Die Konzeption des vorgestellten Informationssystems erlaubt die übersichtliche und konsistente Aufnahme aller aus diesen Archiven stammenden Daten, auch mit begleitenden aktuellen Klimadaten.



Tab. 3: Datentypen

In Informationssystemen werden grundsätzlich zwei Datentypen unterschieden: Metadaten und Meßdaten (analytische Daten). Die Meßdaten werden wiederum untergliedert in primäre Daten (Rohdaten), sekundäre Daten, die aus den Rohdaten evaluierte Werte darstellen und tertiäre Daten. Letztere sind besonders für die Paläoklimaforschung wichtig, da sie die, aus den Sekundärdaten (Proxies) mit Hilfe bestimmter Algorithmen berechneten eigentlichen Klimadaten darstellen (z.B. Paläotemperaturen des Meerwassers).

Parameter group

- Chemistry, sediment
- Chemistry/Physics, water (pore, ocean)
- Components, biogeous
- Components, biogeous (smear slides)
- Components, terrigenous
- Components, terrigenous (smear slides)
- General data
- Grain Size
- Ice/Snow
- Isotopes, stable
- Isotopes, unstable (radionuclides)
- Lake Sediments
- Magnetics
- Micropal., Coccoliths
- Micropal., Diatoms
- Micropal., Dinoflagellates
- Micropal., Foraminifers, benthic
- Micropal., Foraminifers, planktonic
- Micropal., Ostracods
- Micropal., Pollen
- Micropal., Pteropods
- Micropal., Radiolaria
- Micropal., Silicoflagellates
- Mineralogy, bulk
- Mineralogy, clay
- Mineralogy, heavy minerals
- Paleoclimate data
- Physics/Logging

Name of parameter	Abbreviation	Unit
Aluminium	Al	%
Aluminumoxide	Al2O3	%
AntimonySb	ppm	
Arsenic As	ppm	
Barium Ba	ppm	
CadmiumCd	ppm	
Cadmium/Calcium ratio	Cd/Ca	*E-3
Calcium carbonate	CaCO3	%
Calcium, bulk	Ca	%
Calcium, lithogen	Ca lith.	%
Calcium, marine	Ca mar.	%
Calciumoxide	CaO	%
Carbon, carbonate	C(carb)	%
Carbon, total	C	%
Carbon, total (carbonate free)	TOC	%
Carbon, total organic	TOC	%
Carbon/Nitrogen ratio	C/N	
Cerium Ce	ppm	
ChlorineCl	%	
ChromiumCr	ppm	
CO2/AlkCO2/Alk		%
Cobalt Co	ppm	
Copper Cu	ppm	
Dysprosium	Dy	ppm
Erbium Er	ppm	
EuropiumEu	ppm	
Fluorine, bulk	F	ppm
Gadolinium	Gd	ppm
GalliumGa	ppm	
HafniumHf	ppm	
HolmiumHo	ppm	
Hydrogen index	HI	mgHC/gTOC
Hydrogen index, S2	HI S2	mgHC/gS
Hydrogen index, Tmax	HI Tmax	°C
Iron Fe	%	
IronoxideFe2O3	wt%	
Lanthanum	La	ppm
Lead Pb	ppm	
Lithium Li	ppm	
LutetiumLu	ppm	
Magnesium	Mg	%
Magnesiumoxide	MgO	%
Manganese	Mn	%
Manganesoxide	MnO	%
MolybdenMo	ppm	
Neodymium	Nd	ppm
Nickel Ni	ppm	
NiobiumNb	ppm	
Nitrogen N	%	
Nitrogen (carbonate free)	N	%
Nitrogen, inorganic	N inorg	%
Nitrogen, organic	N org	%
Opal Opal	%	
Opal in clay	Opal	%/clay
Opal in sand	Opal	%/sand
Opal in silt	Opal	%/silt
OsmiumOs	ppb	
Oxygene index	OI	mgCO2/gC
PhosphorP	ppm	
Phosphoroxide	P2O5	%
PotassiumK	%	
Potassiumoxide	K2O	%
Praseodymium	Pr	ppm
RubidiumRb	ppm	
SamariumSm	ppm	
ScandiumSc	ppm	
Silica Si	%	
Silicaoxide	SiO2	%
Sodium Na	ppm	
Sodiumoxide	Na2O	%
StrontiumSr	ppm	
Sulphur, calcinated (sediment)	Sigal P	%
Sulphur, Total	S	%
TantalumTa	ppm	
TelluriumTe	ppm	
ThalliumTl	ppm	
ThoriumTh	ppm	
ThuliumTm	ppm	
TinSn	ppm	
Titan Ti	%	
TitanoxideTiO2	%	
Uranium U	ppm	
VanadiumV	ppm	
WolframW	ppm	

Tab. 4: Parameterlisten

Eine Stärke des PANGAEA-Informationssystems ist seine beliebige Erweiterbarkeit für Parameter analytischer Daten. In einer sich ständig verändernden Forschungslandschaft mit immer neuen Fragestellungen war dies eine der wichtigsten Anforderungen an das System. Die Parameter sind der Übersichtlichkeit halber in Parametergruppen gegliedert (linke Tabelle). Die rechte Tabelle zeigt eine Auswahl von Parametern der Gruppe *Chemistry, sediment*. Bisher sind ca. 1600 verschiedene Parameter mit den Datentypen *text, picture und numeric* definiert. Die Erkennung von Parametern beim Datenimport erfolgt über IDs.