

## Vorhabensbeschreibung zur

# Aufnahme des Datenbestandes der PKDB in das Informationssystem PANGAEA

## I. Ziele

### Gesamtziel des Vorhabens

1985 wurde auf Anregung des BMBF mit der Entwicklung einer Paläoklima-Datenbank (PKDB) am Botanischen Institut der Universität Hohenheim unter der Leitung von Prof. Burkhard Frenzel begonnen. Das Projekt bestand aus einem technischen Teil zum Aufbau und Betrieb einer EDV-Infrastruktur und aus einem wissenschaftlichen Teil. Das System wurde technisch bis 1998 betrieben. Der wissenschaftliche Teil soll fortgeführt werden und hat zum Ziel die umfangreiche internationale Literatur zur Klimageschichte des Quartärs zu sichten und zu bewerten. Das Klima soll in seiner zeitlichen Variabilität erfaßt werden um es in quantifizierter Form auf regionaler Basis darstellen zu können.

Von AWI und MARUM wird seit 1996 in Kooperation das Informationssystem PANGAEA - Network for Geological and Environmental Data betrieben. Das Datenmodell erlaubt die Aufnahme beliebiger geocodierbarer Daten aus Klima-, Umwelt- und Meeresforschung, die Netzwerkstruktur ermöglicht einen schreibenden Zugriff auch durch externe Institute, die Daten sind im Internet allgemein verfügbar. Das System wird von ca. 20 Projekten für die Archivierung und Publikation der Projektdaten genutzt.

In Absprache zwischen Zuwendungsgeber und den Betreibern von PKDB und PANGAEA wurde es als sinnvoll angesehen nur ein System technisch zu betreiben. Ziel dieses Projektes ist die Integration des PKDB-Datenbestandes in das Informationssystem PANGAEA, auch mit dem Hintergrund, Befunde des marinen, glazialen und terrestrischen Milieus technisch zusammenzuführen und damit besser wissenschaftlich verknüpfen und auswerten zu können.

### Bezug zu den förderpolitischen Zielen

Prof. Frenzel hat die PKDB im Rahmen des Klimaforschungsprogramms der Bundesregierung unter Beteiligung von Projektgruppen der terrestrischen Paläoklimatologie aufgebaut. Das Vorgängersystem zu PANGAEA wurde im Umfeld der marinen Paläoklimatologie entwickelt, ist jedoch nach der Umstellung auf beliebige geocodierbare Daten auch für andere Forschungsbereiche einsetzbar. Das System wird auch in Projekten der DFG und der EU für das Datenmanagement eingesetzt und leistet damit einen umfassenden Beitrag zur Sicherung wissenschaftlicher Datenbestände, zu deren allgemeiner Verfügbarkeit und zur Unterstützung der projektübergreifenden Auswertung umfangreicher Datensätze.

Im Rahmen der Programminitiative 'Geotechnologien' von BMBF und DFG vom 23.7.1999 ist dieses Projekt ein Beitrag zu den Schlüsselthemen 'Globale Klimaänderungen - Ursache und Auswirkungen' und 'Informationssysteme im Erdmanagement'.

## II. Stand der Technik, bisherige Arbeiten

### Technik, Betrieb und Nutzung von PANGAEA

1993 wurde von den an der marinen PAGES-Forschung beteiligten Wissenschaftlern in Deutschland ein Projekt zur Entwicklung eines Informationssystems für Daten aus marinen Sedimenten initiiert. Es wurde mit finanzieller Unterstützung durch den Bundesminister für Bildung und Forschung (BMBF) von einer Gruppe von Wissenschaftlern und Informatikern am Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI), Bremerhaven, in Zusammenarbeit mit der Universität Bremen entwickelt. Nach der Erweiterung auf die Erfassung beliebiger in Raum und Zeit geocodierbarer Daten wurde das System in Anlehnung an seine Struktur und seine Aufgaben nach dem Superkontinent benannt, in dem alle Kontinente vor 200 Millionen Jahren vereint waren: PANGAEA - Network for Geological and Environmental Data (<http://www.pangaea.de>).

Das System besitzt ein vollständig normalisiertes Datenmodell, das auf Grund seiner wissenschaftlichen wie auch seiner informationstechnischen Struktur universell verwendbar ist. Trotz der dem System wie auch den Daten innewohnenden Komplexität wurde bei der Konzeption, neben einer ausgefeilten Datenhaltung, größter Wert auf Nutzbarkeit und Akzeptanz gelegt. Die allgemeine Verfügbarkeit über das Internet entspricht dem Stand des heute technisch Möglichen. Das PANGAEA-Netzwerk nutzt Client/Server-Technologie, das zentrale Datenmodell ist in einer relationalen Datenbank abgebildet. Der Datenimport erfolgt über systemeigene Client-Software, der Zugriff für Suche und Export im Internet wurde über einen Web-Client realisiert. Der Web-Client ist durch seine generische Struktur universell verwendbar und kann an sich verändernde Datenbankstrukturen angepaßt werden. Die Geocodierung von Daten ermöglicht Abfragen und Verknüpfungen beliebiger Teildatensätze in Raum und Zeit. Ein Passwortschutz erlaubt den internationalen Austausch unpublizierter Daten im Internet z. B. in laufenden Projekten. Einfach zu benutzende graphische Werkzeuge ermöglichen die Darstellung wissenschaftlicher Daten.

Der technische Betrieb des Systems wird durch das Rechenzentrum des AWI sichergestellt. Er umfaßt Datensicherung, Aktualisierung von Hardware und Software, Systemmanagement und Internetanbindung. Stand der Technik: Hardware SUN E10000 (22 Prozessoren, 16 GB RAM, 250 GB Festplattenkapazität), Datenbankmanagementsoftware: SYBASE, Client-Software für Import: 4th Dimension (ACI), Middleware für Import/Export in C, WWW-Client Software in JAVA, Internet-Verbindung: ATM 34 MBit.

Der Aufwand für den wissenschaftlichen Betrieb ist durch die beteiligten Projekte zu erbringen. Dieser umfaßt die ständig notwendigen Anpassungsarbeiten aller Softwarekomponenten an neue Technologien und wissenschaftliche Fragestellungen, Projektbetreuung, Installationen und Publikation von Daten im Internet.

PANGAEA wird seit 4 Jahren auf Tagungen und Workshops im nationalen wie internationalen Rahmen vorgestellt. Das System wird für das Projektdatenmanagement in folgenden größeren Projekten eingesetzt:

ADEPD (EU) Atlantic Data Base for Exchange Processes at the Deep Sea Floor  
APIS (international) Antarctic Pack Ice Seals  
BAH (MARUM) Erfassung der Zeitreihen und Projektdaten der BAH (Biologische Anstalt Helgoland)  
BASYS (EU) Baltic Sea System Study  
CRP (international) Cape Roberts Project  
IMAGES (international) International Marine Global Change Study  
MADAM (BMBF) Mangrove Dynamics and Management  
QUEEN (ESF/EU) Quaternary Environment of the Eurasian North  
SFB 261 (DFG) Der Südatlantik im Spätquartär  
SINOPS (EU) Silicon Cycling in the World Ocean  
SFB 313 (DFG) Umweltveränderungen im Nordatlantik  
GeoBio15k (DFG) Wandel der Geo-Biosphäre in den letzten 15 000 Jahren  
EPICA (international) European Project for Ice Coring in Antarctica

ausgewählte Publikationen:

- Diepenbroek, M., Grobe, H., and Reinke, M., 1997, PANGAEA - ein Informationssystem für die Meeresforschung: Deutsche Gesellschaft für Meeresforschung Mitteilungen, v. 1-2, p. 94-96.
- Diepenbroek, M., Fütterer, D., Grobe, H., Miller, H., Reinke, M., and Sieger, R., 1998, PANGAEA information system for glaciological data management: Annals of Glaciology, v. 27, p. 655-660.
- Diepenbroek, M., Grobe, H., Reinke, M., Schlitzer, R., and Sieger, R., 1999, Data management of proxy parameters with PANGAEA, in Fischer, G., and Wefer, G., eds., Proxies in Paleoceanography, Springer Verlag, Heidelberg, p. 715-727.

## Gewinnung terrestrischer Paläoklimadaten aus der Literatur

Die internationale Literatur der historischen, philologischen, geo- und biowissenschaftlichen Disziplinen, die wenigstens teilweise die Umwelt des Menschen berühren, ist reich an Angaben, die für eine Klimarekonstruktion der Vergangenheit genutzt werden können. Zur Schaffung einer ausreichenden Datengrundlage werden in der Arbeitsgruppe um Prof. Frenzel quantitative Paläoklimadaten aus der Literatur erfaßt und qualitative paläoökologische Daten klimatologisch quantifiziert. Die Ausarbeitung einer zuverlässigen Datierung ist hierbei ein wesentlicher und arbeitsaufwendiger Bestandteil der Arbeit. Seit Bestand der PKDB wurde in langjähriger Literaturarbeit ein umfangreicher Datensatz geschaffen, der sicherlich in der terrestrischen Paläoklimaforschung als einmalig bezeichnet werden kann. Hervorgehoben sei hierbei insbesondere die Auswertung von Literatur in 18 Sprachen. Auf der Tagung zur Haltung von Paläoklimadaten im Weltzentrum für Paläoklimatologie, Boulder (WDC) 1997 wurde der PKDB bescheinigt, daß sie die einzige Bank ist, bei der eine große Zahl verschiedener terrestrischer Daten eng verknüpft sind.

Die gewonnenen Daten wurden in Karten niedergelegt, so daß eine räumliche Klimaanalyse der Vergangenheit für verschiedene Zeiträume möglich wurde. Vier verschiedene Zeiträume wurden für die Nordhalbkugel kartographisch dargestellt, drei weitere des spätglazialen Klimawandels wurden für Nordeurasien erarbeitet (Frenzel, Pécsi & Velichko, 1992; Frenzel, 1994). Die Arbeiten sollen fortgeführt werden; ein Paläoklima-Atlas für die Südhemisphäre befindet sich in Vorbereitung.

Der Bestand der PKDB umfaßt nach Abschätzung der am AWI vorliegenden Texttabellen etwa 350 000 bibliographische Titel, davon sind 200 000 Titel mit Schlüsselworten und 173 000 mit Zusammenfassungen versehen. Ein wesentlicher Teil dieses Bestandes wurde aus bestehenden Literaturdatenbanken nach paläoklimatologischen Schlagworten extrahiert. 18 000 Publikationen liegen in Kopie im Hohenheimer Archiv vor und wurden von Prof. Frenzel kritisch bewertet; 1000 Exzerpte wurden in die PKDB übernommen. Die eigentlichen Faktendaten liegen in 68 000 Datensätzen zu 25 unterschiedlichen Datengruppen vor. Der Umfang der an das AWI gelieferten Tabellen beläuft sich auf ca. 600 MB. Die PKDB läuft mit der Datenbankmanagementsoftware EMPRESS unter AIX. Seit 1994 war die PKDB mit eigenen Seiten im WorldWideWeb präsent, seit Mitte 1997 konnte in den Bibliographietabellen und Paläodatenbeständen der PKDB recherchiert werden. Die PKDB ist weiterhin im Internet verfügbar (<http://www.pkdb.uni-hohenheim.de/>).

Besonderes Interesse der PKDB galt den folgenden Witterungs- und Klimaparametern: Verbreitung/Typologie des Dauerfrostes; Höhenlage der orographischen und klimatischen Schneegrenze; Gleichgewichtslinie von Gletschern und Eiskappen; Eishaushalt; eustatische und isostatische Meeresspiegeländerungen; Seespiegelschwankungen der Kontinente; Wasserhaushalt; Temperaturen; Oberflächenalbedo; Frequenzanalyse/Geschwindigkeit von Klimaschwankungen; Windrichtungen; Bodenfeuchte; Typologie/Verbreitung fossiler Böden; Aërosolgehalt der Atmosphäre.

## Validierung der Datierungen

Alle Befunde wurden bei der Erfassung hinsichtlich ihrer Datierungsqualität überprüft und bewertet. Die Datierungen basieren bei schriftlichen Quellen auf dem jeweils angewandten Kalender, bei naturwissenschaftlichen Daten aber auf relativgeologischen oder absoluten, bzw. physikalischen Datierungsverfahren. Die zeitliche Auflösung umfaßt Durchschnittswerte eines Jahres, bzw. einiger Jahre, einzelner Jahreszeiten oder Monate und Extremereignisse sehr kurzer Dauer. In vielen Fällen sind die zeitlichen Abstände zwischen den stratigraphischen Fixpunkten jedoch deutlich größer.

Die Altersdaten müssen bei der Überführung von der PKDB nach PANGAEA einer erneuten Kontrolle unterzogen werden, da sich die Methoden der Altersnormierung geändert haben und die unterschiedlichen Altersskalen der PKDB auf die einheitliche Zeitskala in PANGAEA für eine sinnvolle Nutzung angepaßt werden müssen. Auch muß dargestellt werden, inwiefern paläoökologischer Inhalt und Datum zusammenpassen und welcher Altersbereich mit Fehlergrenzen der einzelnen Alter anzugeben ist; weiterhin ist die Methode und das zur Datierung verwendete Material zu nennen. Die Datierungen und Altersmodelle werden, wie auch alle anderen Daten, mit den zum Verständnis notwendigen Metainformationen in PANGAEA abgelegt. Dies erlaubt dem Nutzer später die Daten z.B. vor einer Verwendung in Zeitscheibenrekonstruktionen oder in Modellrechnungen kritisch überprüfen und bewerten zu können.

Die Aufnahme der terrestrischen Paläoklimadaten aus der PKDB darf keine rein technische Übernahme bereits publizierter Daten sein, sondern wird zum Teil eine wissenschaftliche Arbeit darstellen, die nur von einem Wissenschaftler mit dem notwendigen Sachverstand für geologische Fragestellungen ausgeführt werden kann. Auch bedarf diese Person des wissenschaftlichen Rückhaltes der an der Paläoklimaforschung beteiligten Gruppen. Unsicher ist die Aufwandsabschätzung für die Validierung und Anpassung der Datierungen, da es keine Möglichkeit gibt im Vorfeld den Umfang und die Vielfalt an in der PKDB vorhandenen Datierungen zu erfassen.

### III. Arbeitsplan

#### Aufgaben bei der Datenüberführung

Beide Datenbanken sind relational, arbeiten aber, geschichtlich bedingt, mit unterschiedlichen Betriebssystemen, verschiedener Datenbankmanagementsoftware und basieren auf unterschiedlichen Datenmodellen. Es gibt daher keine technische Lösung, die Daten als Ganzes von einem System direkt in das andere zu überführen. Der ehemalige Systemmanager der PKDB hat auf Anforderung von Prof. Frenzel alle Daten als Tabellen im ASCII-Format extrahiert und der PANGAEA-Gruppe zur Verfügung gestellt. Die bereits erfolgte Sichtung zeigte nur eine zuverlässige Möglichkeit des Datentransfers auf: Die Daten werden nach Publikationen geordnet aus den Tabellen herausgezogen, in das PANGAEA-Importformat gebracht und dort eingelesen. Gleichzeitig soll eine erneute Überprüfung in Verbindung mit der Publikation selbst und durch Zugriff auf die Originaldaten in der PKDB über das Internet erfolgen.

Umfangreiche Teile des Bestandes sind Datensätze, die aus anderen Datensammlungen im Internet übernommen wurden. Hierbei handelt es sich im wesentlichen um Beobachtungen aus dem weltweiten Netz von Klimastationen (*The Global Historical Climatology Network*, Quelle: NOAA) und um das *Eurasian Glacier Inventory* (Quelle: National Snow and Ice Data Center). Weitere kleinere Datenbanken aus anderen Quellen (World Data Center) sind die *Global Pollen Data Base*, *Glacier Mass Balance Database*, *Tree Ring Database* und *Sea Levels during the past 35 000 years*. Diese Daten werden nicht übernommen, da sie im Internet ständig an ihrer Originalquelle verfügbar sind und nur in Ausnahmefällen dupliziert werden sollten. Falls ein Projekt später bei der Auswertung der eigenen Daten mit PANGAEA bestimmte Datensätze benötigt, können diese relativ schnell von der Originalquelle in das System übernommen werden; hierzu ist allerdings vorher die Zustimmung des Besitzers/Datenzentrums einzuholen.

Paläoklimadaten aus terrestrischen Archiven werden, im Gegensatz zur Auswertung mariner Archive, aus deutlich mehr Klimaindikatoren abgeleitet. Hierin liegt zwar die Stärke und die Sicherheit terrestrischer Paläodaten, weil so eine gegenseitige unabhängige Kontrolle der auf verschiedenen Wegen gewonnenen Proxydaten erfolgen kann, sie sind jedoch auch eine Größenordnung komplexer. Die wesentliche Arbeit im Datentransfer von der PKDB nach PANGAEA besteht darin, die terrestrischen Daten so umzuformen, daß sie in PANGAEA geocodiert abgelegt werden können um später auf beliebige Abfragen hin für Auswertungen zur Verfügung zu stehen. Die Datenablage soll der Organisation und vor allem den Abfragemöglichkeiten in der PKDB weitgehend entsprechen bzw. diese noch erweitern. Das universelle Datenmodell und die vierdimensionale Geocodierung (Raum/Zeit) stellen hierfür die notwendigen Infrastrukturen in PANGAEA zur Verfügung. Bei der Überführung müssen insbesondere die unterschiedlichen Strukturen der Daten und die Alterssynchronisierung von Befunden aus verschiedenen Paläo-Archiven beachtet und untereinander abgestimmt werden.

## Datenfluß PKDB -> PANGAEA

Die Daten werden nach Zitaten geordnet aus den vorhandenen PKDB-Texttabellen herausgezogen. Hierfür steht ein Werkzeug zur Verfügung, das von der PANGAEA-Gruppe für die universelle Datenkonversion entwickelt wurde (PanTool). Es ermöglicht auch die Handhabung umfangreicher durch Relationen verbundener Texttabellen. Die individuelle Umformung in das PANGAEA-Importformat erfolgt über einen kommerziellen Editor. Vor dem Import der Faktendaten werden die Metadaten (Zitate, Besitzer, Methoden, Parameter) gesammelt in Blöcken aufgenommen. Beim Import wird die Zuordnung dieser Metadaten zu den einzelnen Faktendatensätzen durch die Importroutine erledigt; die Zuordnung der Daten zu den Parametern erfolgt über eine Identifikationsnummer.

Da PANGAEA als System zur Archivierung, Auswertung und Verteilung von Faktendaten konzipiert wurde, werden die Datensätze mit den zugehörigen Zitaten übernommen, die aus den Arbeiten der Datenkuratoren in Hohenheim resultieren und den eigentlichen Wert der Sammlung darstellen. Zitate, die keine Daten beinhalten, werden in einer lokalen Literaturdatenbank abgelegt, die weiterhin für Literaturarbeit zur Verfügung steht. Der Aufbau einer im Internet verfügbaren Literaturdatenbank erscheint angesichts bereits zahlreicher vorhandener Systeme und sich im Internet etablierender Bibliotheksverbände nur dann sinnvoll, wenn die Zitate durch eine Geocodierung erweitert werden. Dies würde eine Suche nach Publikationen zu einem ausgewählten geographischen Gebiet ermöglichen. Die Geocodierung von Tausenden von Publikationen würde jedoch ein eigenes Projekt erfordern.

## IV. Verwertungsplan

Die Aufnahme des Datenbestandes der PKDB in PANGAEA wird die Daten in neuer Qualität und im Zusammenhang mit einem weiteren umfangreichen Bestand an Paläoklimadaten aus anderen Archiven (Seen, Eis, Ozeane, Korallen) der wissenschaftlichen Gemeinschaft zur Verfügung stellen. Diese Zusammenführung ist auch als weiterer Schritt im Aufbau eines Systems zu sehen, mit dem sich verschiedene Klimaproxyparameter aus unterschiedlichen Paläo-Archiven auf einer gemeinsamen Zeitachse zusammenführen und auswerten lassen. Zwangsläufig ist das System auch ein Datenarchiv, daß seinen Bestand der internationalen Gemeinschaft über das Internet zur Verfügung stellt.

Die Umformung der Daten wird in direkter Zusammenarbeit mit der Arbeitsgruppe um Prof. Frenzel erfolgen. Die Beratung durch die Datenkuratoren in Hohenheim ist Voraussetzung, um die umfangreiche Datensammlung in Einzelfiles zu zerlegen und in das neue Format überführen zu können. Hier ist sowohl das in Hohenheim vorhandene technisch-organisatorische Wissen um die Arbeit mit den terrestrischen Daten und der PKDB, als auch der wissenschaftliche Sachverstand unabdingbarer Bestandteil dieses Projektes. Im Antrag sind für diese Aufgaben daher Mittel für Werkverträge vorgesehen.

Ein intensiver Informationsaustausch ist ebenfalls mit den Datenkuratoren des DFG-Schwerpunktes *'Wandel der Geo-Biosphäre in den letzten 15 000 Jahren'* sichergestellt (Dr. Wunderlich, Dr. Sagemann, Koordinator: Prof. Andres). Der DFG-Schwerpunkt verwendet ebenfalls PANGAEA zur Archivierung seiner Daten; somit bietet sich hier gleichzeitig die Möglichkeit Daten aus zwei eng verwandten Forschungsprogrammen in einem System zusammenzuführen und diese auch für zukünftige Projekte gesammelt zur Verfügung zu haben.

Der dritte Kooperationspartner ist das HGF-Projekt *'Klimavariationen in historischen Zeiten'* (KIHZ) mit der Koordinationsstelle am GFZ (Prof. Negendank). HGF-Einrichtungen in Zusammenarbeit mit Universitäten arbeiten an dem Aufbau eines Multiproxyparameter-Netzwerkes mit dem Ziel Paläoklimadaten des Holozäns auf einer gemeinsamen 'synthetischen' Zeitachse abzubilden, auszuwerten und für Modelle verfügbar zu machen. Auch KIHZ beabsichtigt zukünftig hierfür PANGAEA zu verwenden.