

Zusammenfassung zum Schlussbericht zur

Übernahme des Datenbestandes der PKDB in das
Informationssystem PANGAEA

Förderkennzeichen: 01LK0001

Laufzeit: 01.07.2000 bis 30.06.2001

Im Institut für Botanik an der Universität Hohenheim wurde unter Leitung von Prof. Dr. Burkhard Frenzel in den 90er Jahren mit der PKDB (Paläoklimadatenbank des Quartär) eine Datenbank für terrestrische Paläoklimadaten entwickelt und im weiteren Verlauf mit dem Bestand des Publikationsarchivs von Prof. Frenzel und mit am Institut für Botanik erarbeiteten Paläoumwelt und -klimadaten Daten gefüllt. Im Rahmen dieses Projektes wurde der Datenbestand der PKDB in das von AWI (Stiftung Alfred-Wegener-Institut für Polar- und Meeresforschung) und MARUM (Zentrum für Marine Umweltwissenschaften) betriebene Informationssystem PANGAEA (Network for Geological and Environmental Sciences) übernommen und im Internet allgemein verfügbar gemacht (<http://www.pangaea.de/Projects/PKDB/>). Zusätzlich wurden ältere Datensätze zu Pollenprofilen und eine Baumringchronologie archiviert. Aus dem 'Atlas of Paleoclimates and Paleoenvironments' wurden 35 Karten digital bereitgestellt; auch das Literaturarchiv zur terrestrischen Paläoklimatologie (27 000 Zitate) von Prof. Frenzel steht als Texttabelle zur allgemeinen Verfügung. Eine langfristige Verfügbarkeit der Daten über den PANGAEA-Webserver ist durch die betreibenden Institute sichergestellt.

Schlussbericht zur

Übernahme des Datenbestandes der PKDB in das Informationssystem PANGAEA

Förderkennzeichen: 01LK0001

Laufzeit: 01.07.2000 bis 30.06.2001

Aufgabenstellung

An der Universität Hohenheim wurde unter Leitung von Prof. Dr. Burkhard Frenzel in den 90er Jahren mit der PKDB (Paläoklimadatenbank des Quartär) eine Datenbank für terrestrische Paläoklimadaten entwickelt und im weiteren Verlauf mit dem Bestand des Publikationsarchivs von Prof. Frenzel und mit am Institut für Botanik erarbeiteten Daten gefüllt. Im Rahmen dieses Projektes sollte der Datenbestand der PKDB in das von AWI (Stiftung Alfred-Wegener-Institut für Polar- und Meeresforschung) und MARUM (Zentrum für Marine Umweltwissenschaften) betriebene Informationssystem PANGAEA (Network for Geological and Environmental Data, <http://www.pangaea.de>) übernommen und im Internet allgemein verfügbar gemacht werden. Planung und Arbeiten zu diesem Vorhaben wurden von der Gruppe am AWI und MARUM ausgeführt, die das Informationssystem betreiben.

Durchführung

Der Datenbestand wurde durch den ehemaligen Systemmanager aus der PKDB-Datenbank an der Universität Hohenheim in Texttabellen konvertiert. Die Tabellen im Gesamtumfang von 600 MB wurden per Filetransfer an das AWI übertragen. Es wurde ein Werkzeug entwickelt, um in den Tabellen navigieren zu können. Die Paläoklimadatensätze des PKDB-Projektes wurden den Quellen/Publikationen und den Datenerzeugern zugeordnet, die relevanten Publikationen in PANGAEA übernommen, die Datensätze in das PANGAEA-Importformat konvertiert, eingelesen und mit den Zitaten verknüpft. Dieser Datenfluss war der wesentliche Teil des Projektes, die technischen Schritte sollen hier nicht im einzelnen ausgeführt werden, da sie im wesentlichen komplexe Konvertierungen und logische Verknüpfungen umfassen. Es sei lediglich erwähnt, dass bei allen Arbeiten eine Qualitätskontrolle durchgeführt wurde, um jederzeit sicherstellen zu können, dass die Daten präzise gespiegelt und ohne Informationsverlust in ihre neue Form überführt werden.

Die Datensätze sind über den PANGAEA-Webserver unter **<http://www.pangaea.de/Projects/PKDB/>** allgemein verfügbar.

Die PANGAEA-Suchmaschine PangaVista erlaubt weiterhin eine individuelle Suche im Gesamtdatenbestand. Die beliebig wählbaren Stichworte müssen lediglich in der Metadatenbeschreibung zu einem Datensatz vorkommen (z.B. Autorennamen, Gebietsbezeichnungen, Probenkennungen oder Parameter). Eine Suche kann durch die Vorgabe von Koordinaten geographisch eingegrenzt werden.

PANGAEA stellt weiterhin mit dem Advanced Retrieval Tool (ART) einen auf Internet-Browsern nutzbaren Client zur Verfügung, der als Java-Applet eine individuelle Abfrage und Konfiguration von Exportdatensätzen ermöglicht. Die Nutzung von ART erfordert allerdings Kenntnisse über das Datenmodell, die Fähigkeit eine Abfrage mit boolescher Logik zu formulieren und ein grundsätzliches Verständnis der gewünschten Daten. Ein einfaches Tutorial bietet Unterstützung.

Alle Daten wurden in Raum und Zeit geocodiert eingelesen. Die Codierung umfaßt als die drei Raumkoordinaten die geographische Position und die Elevation, optional die relative Entfernung auf einem Profil. Da es sich um geologische Daten handelt, konnte als zeitliche Codierung das jeweilig Alter des Datenpunktes in Jahren vor heute (kyr BP) angegeben werden. Sind die Daten einem Zeitraum zugeordnet, wurde zusätzlich ein Minimal- und ein Maximalalter angegeben. Im Rahmen der Auswertung von in der Literatur beschriebenen Klimainformationen wurden im PKDB-Projekt relative Klimaänderungsparameter für Niederschlag, Temperatur, Windgeschwindigkeit und Verdunstung als eigenständige Parameter definiert. Diese bilden sich in PANGAEA technisch als nominale numerische Parameter mit einer entsprechenden Definition ab, wissenschaftlich stellen sie eine Art Klimaproxy-Datengruppe dar.

Das Verzeichnis des Literatur- und Sonderdruckarchivs von Prof. Frenzel (ca. 27 000 Zitate) wurde in eine vereinfachte Tabellenform überführt und über die PKDB-Webseite verfügbar gemacht. Das Tabellenformat mit Tabulator-getrennten Spalten im ASCII-Format ermöglicht dem Nutzer die Zitate in ein Literaturverwaltungsprogramm seiner Wahl einzulesen.

Wichtiger Bestandteil in der PKDB sind Rekonstruktionen von heutigen und letzteiszeitlichen Schneegrenzen aus den Arbeiten in Tibet von Prof. Frenzel in den

Jahren 1989, 1992 und 1996. Diese wurden mit einem neuen Parameter (Snow line elevation) in der Einheit 'm über dem Meeresspiegel' und entsprechender Geocodierung der räumlichen Verbreitung und der zeitlichen Reichweite eingelesen. Dieser Datenbestand wurde zusätzlich von Herrn Frenzel überarbeitet, auf den aktuellen Stand gebracht und mit den entsprechenden Zitaten versehen.

Als ein Standardwerk der terrestrischen Paläoklimatologie wurde es als sinnvoll angesehen den 'Atlas of Paleoclimates and Paleoenvironments' von Frenzel et al. (1992) zu digitalisieren und im Internet zur Verfügung zu stellen. 35 Karten im Format A2 wurden in unterschiedlicher Auflösung für Bildschirm und Druck auf den PKDB-Webseiten zum Herunterladen bereitgestellt.

Die Metadaten zu den importierten PKDB-Datensätzen werden nach Abschluß dieses Projektes an das Global Change Master Directory geliefert (GCMD, <http://gcmd.nasa.gov/>). Damit stehen die Daten auch über das zentrale Verzeichnis von Datensätzen aus der internationalen Global Change-Forschung zur Verfügung.

Der Betrieb des PKDB-Systems an der Universität Hohenheim wurde nach der Überführung der Daten eingestellt. Auf den Webseiten des Institutes für Botantik in Hohenheim verweist ein Link auf die neuen Zugriffsseiten auf dem PANGAEA-Webserver.

Die Bestandsaufnahme des Dateninhaltes zeigte, daß neben eigenen Daten des PKDB-Systems ein wesentlicher Bestandteil Klimadatensammlungen aus anderen internationalen Datenzentren waren. Diese wurden, wie im Antrag formuliert, nicht übernommen.

Es handelt sich um die folgenden Archive (Anzahl der Lokalitäten in Klammern):

- (1) Calcium Carbonate Database (311)
<ftp://ftp.ngdc.noaa.gov/paleo/paleocean/carbonate/>
- (2) Relative Sea Level Database (392)
(Betrieb wurde eingestellt)
- (3) Global Seawater Oxygen-18 Database (419)
<http://www.giss.nasa.gov/data/o18data/>
(diese Datenbank wurde mit Zustimmung der Betreiber übernommen)
- (4) The Oxford Lake-Level Data Bank (536)
<ftp://ftp.ngdc.noaa.gov/paleo/lakelevels/oxford/>
- (5) Global Pollen Database (799)
<http://www.ngdc.noaa.gov/paleo/gpd.html>

- (6) Paleoceanographic Database (983)
<http://www.ngdc.noaa.gov/paleo/paleocean.html>
- (7) Paleobiology Database (1045)
<http://flatpebble.nceas.ucsb.edu/public/index.html>
- (8) Tree Ring Database (1256)
<http://www.ngdc.noaa.gov/paleo/treering.html>
- (9) Global Volcanism Program, List of Holocene Volcanoes (1509)
<http://www.volcano.si.edu/gvp/index.htm>
- (10) The Global Historical Climatology Network (10249)
<http://cdiac.esd.ornl.gov/ghcn/ghcn.html>
- (11) Eurasian Glacier Inventory (49148)
http://nsidc.org/NOAA/wgms_inventory/

Grundsätzlich könnten die aufgeführten Archive in PANGAEA übernommen werden; das vollständig normalisierte Datenmodell und die technische Kapazität des Systems würden dies erlauben. Eine Spiegelung von größeren externen Datenbanken sollte jedoch nur nach Absprache mit den Betreibern erfolgen. Mittelfristig ist eine Spiegelung verschiedener Datenbanken des WDC for Paleoclimatology geplant, ein allgemeines Portal für Paläodaten im Internet, initiiert vom WDC-MARE (World Data Center for Marine Environmental Sciences) und IGBP/PAGES, ist in Vorbereitung. Eine einmalige Spiegelung macht nur für publizierte und damit statische Datensätze Sinn, da ein automatischer laufender Spiegelungsprozess zwischen technisch unterschiedlichen Systemen einen unverhältnismäßig hohen Aufwand durch die präzise Definition der Schnittstellen und deren Implementierung bedeuten würde.

Ein Teil der ursprünglich erwarteten Daten aus Hohenheimer Projekten war nicht in der PKDB verfügbar, sollte jedoch im Rahmen dieses Projektes erfaßt werden, um zumindest wesentliche Teile der wertvollen Pollenprofile und Baumringchronologien aus den Arbeiten der Paläobotanik langfristig zu archivieren und allgemein verfügbar zu machen. In Diskussionen um zukünftige Projekte und Informationsinfrastrukturen im europäischen Rahmen hatte sich herausgestellt, daß im internationalen Vergleich ein Defizit an Daten aus deutschen Projekten besteht und ein verbesserter Austausch von Partnern in anderen Ländern dringend gewünscht wird. Insgesamt 6 Pollenprofile, sowie ein Oberflächendatensatz aus der Arbeitsgruppe von Prof. Frenzel wurden zusammengestellt, in elektronische Form überführt und in PANGAEA erfaßt. Im Rahmen der angestrebten Archivierung der Hohenheimer Baumringdaten wurde ein Datensatz der Main-Chronologie als Demonstrationsfile geliefert und über PANGAEA publiziert.

Ausblick

Grundsätzlich hat die Datenübernahme aus der PKDB zu einem verbesserten Zugriff auf Paläodaten aus deutschen Forschungsprojekten über das Internet beigetragen. Zukünftige Projekte werden hiervon profitieren. So wird zur Zeit auf europäischer Ebene von einer Gruppe um das PEP III Programm (Pole-Equator-Pole Transect) ein System zur Archivierung, Verteilung und Auswertung von Paläoklimadaten konzipiert. Als technische Infrastruktur zur Erfassung und Konsistensierung der Daten wird PANGAEA verwendet. Ein entsprechender EU-Antrag mit dem Akronym MUPPETS wurde gestellt. Bestandteil von MUPPETS wird auch die Spiegelung der europäischen Pollendatenbank (EPD) sein, die, auch in Verbindung mit den Pollendaten aus Hohenheim, Grundlage für ein zukünftiges europäisches Paläodatenarchiv bilden werden.

PANGAEA wird seit der Aufnahme des Routinebetriebes 1995 von nationalen und internationalen Projekten für die Archivierung und Publikation von geocodierbaren Daten verwendet. Mit der Übernahme der PKDB-Daten wurde dieser Bestand erweitert und bildet im Zusammenhang mit den Daten aus anderen Projekten den Grundstock für ein Datenarchiv, welches die Voraussetzung für eine umfassende Diskussion und Modellierung von Klimaveränderungen bilden soll. Wie im laufenden Deutschen Klimaforschungsprogramm DEKLIM angestrebt, sollen aktuelle Messdaten und Ergebnisse zur Klimageschichte der Erde jetzt verstärkt zusammengeführt und als Grundlage für die Vorhersage zukünftiger Klimaentwicklungen in Rechenmodelle einfließen. Relativ zu den ständig erzeugten Daten und den daraus resultierenden veröffentlichten Ergebnissen muß der Bestand an allgemein verfügbaren Daten jedoch immer noch als sehr gering eingestuft werden. Zu einem ausgewählten Wissenschaftsbereich wie z.B. der Paläoklimaforschung liegt nach unseren Abschätzungen der Anteil langfristig archivierter und verfügbarer Daten vom Gesamtbestand laufend publizierter Ergebnisse bei unter einem Prozent. Eine wesentliche Voraussetzung für die vom entsprechenden BMBF-Arbeitskreis für DEKLIM vorgeschlagene 'konsistente Interpretation von (Paläo-)Daten in Atmosphäre und Ozean' ist somit nicht gegeben. Ein Problem bleibt hier, wie auch in anderen Wissenschaftsbereichen, die fehlende Akzeptanz und die Etablierung des Datenmanagements als essentieller Bestandteil einer modernen Wissenschaft.