

TEIL B

HINLEITUNG ZUR FRAGESTELLUNG

Teil B dient zur Hinleitung zur Fragestellung. Zunächst wird das Potsdam-Institut für Klimafolgenforschung vorgestellt, in dessen Kontext die Konzeption und Entwicklung der Schnittstelle zum autonomen Datenzugriff durchgeführt wurde (Kap. 6). Danach wird ein Überblick über die vom Institut für eine Verbesserung der Erschließung ausgewählten Datenräume gegeben (Kap. 7). Daran anschließend findet eine Analyse der Ausgangslage und insbesondere der durch diese gegebenen Defizite statt (Kap. 8). Abschließend werden spezifische Voraussetzungen und Herausforderungen für die Entwicklung einer geeigneten Schnittstelle erörtert (Kap. 9) sowie die Aufgabenstellung zusammengefasst (Kap. 10).

6 Das Potsdam-Institut für Klimafolgenforschung (PIK)

„The transformation of the global climate system, set going by the very nature of our technical civilization, represents one of the greatest challenges so far in the history of mankind.“¹⁹⁴

„Das PIK ist von seiner Idee und Struktur her ein Institut ohne Prototyp oder Vorgänger [...]“¹⁹⁵

In Kap. 6 wird das Potsdam-Institut für Klimafolgenforschung (PIK) vorgestellt. Zunächst werden Kontext (Kap. 6.1), Auftrag und Ansatz (Kap. 6.2) sowie die Forschungsstruktur des Institutes (Kap. 6.3) dargestellt. Kap. 6.4 beschreibt die Herausforderungen, die durch die für die Arbeit des Institutes erforderliche komplexe Datenbasis aufgeworfen werden. In Kap. 6.5 werden die verschiedenen Ebenen des von der Scientific Data Management Group des Institutes adressierten wissenschaftlichen Datenmanagements gegeneinander abgegrenzt; Kap. 6.6 skizziert das im Rahmen dieser Arbeit in enger Zusammenarbeit mit der Scientific Data Management Group verfolgte Ziel einer geeigneten Schnittstelle zur autonomen Datenerschließung.

6.1 Kontext

Die Gründung des Potsdam-Instituts für Klimafolgenforschung (Potsdam Institute for Climate Impact Research, PIK) [PIK 2000/01] [PIK 1998/99ab] [PIK 1996/97] [PIK 1994/95] [PIK WWW] ist vor dem Hintergrund eines sich immer deutlicher abzeichnenden menschenverursachten (anthropogenen) Klimawandels¹⁹⁶ zu sehen. Die Emission von sog. Treibhausgasen wie CO₂, CH₄ und N₂O durch Industrie, Verkehr und Haushalte hat die Strahlungsbilanz der Erdatmosphäre verändert und wird in diesem Jahrhundert zu einer Erhöhung der Erdoberflächentemperatur führen¹⁹⁷; überdies findet diese Entwicklung während einer der warmen Phasen der Erdgeschichte statt.

Im Zweijahresbericht 2000/2001 des Institutes wird auf die einzigartige Herausforderung hingewiesen, die durch die von der technologischen Zivilisation verursachte Veränderung des globalen Klimasystems für die Menschheit aufgeworfen wird. Sie ergibt sich aus der Komplexität der erwarteten Rückkopplungen, der gigantischen raumzeitlichen Ausdehnung der Störung, der Unumkehrbarkeit des wahrscheinlichen Schadens für zahllose ökologische und sozioökonomische Systeme sowie den hiermit verbundenen, völlig neuartigen Aspekten internationaler und generationenübergreifender Gerechtigkeit. Als die Problematik, die zunächst nur graduell auf einzelnen wissenschaftlichen Konferenzen adressiert wurde, vor rund 20 Jahren das öffentliche und politische Bewusstsein zu erreichen begann, wurde deutlich, dass weltweit weder die wissenschaftlichen Kapazitäten existierten, um wohlfundierte Antworten für spezifische Teilprobleme zu liefern, noch um gar zu einer integrierten Abschätzung des Problems in seiner Gesamtheit zu gelangen [PIK 2000/01, 13].

¹⁹⁴ Potsdam-Institut für Klimafolgenforschung: *Biennial Report 2000/2001* [PIK 2000/01, 13].

¹⁹⁵ Potsdam-Institut für Klimafolgenforschung: *Biennial Report 1996/1997* [PIK 1996/97, 97].

¹⁹⁶ Einen Überblick über die Dimensionen des Klimawandels gibt der Third Assessment Report (TAR) *Climatic Change 2001* (vgl. [IPCC TAR 2001]) des Intergovernmental Panel on Climate Change (IPCC). Das IPCC wurde 1998 von der World Meteorological Organization (WMO) und dem United Nations Environment Programme (UNEP) ins Leben gerufen und untersucht die wissenschaftlichen Aspekte von Klimasystem und Klimawandel, die Verwundbarkeit sozioökonomischer und natürlicher Systeme durch sowie Auswirkungen und Adaption an einen Klimawandel und schließlich mögliche Optionen zu einer Abmilderung. Wissenschaftler des PIK sind maßgeblich an der Arbeit des IPCC beteiligt [PIK 2000/01, 106]. Vgl. auch die aktuelle BMBF-Studie *Herausforderung Klimawandel* von 2003 [BMBF 2003].

¹⁹⁷ Laut *Biennial Report 2000/2001* des Institutes ist von einer globalen Erwärmung im 21. Jahrhundert um 2 bis 3 °C auszugehen [PIK 2000/01, 13]; Hochrechnungen basierend auf Szenarien des IPCC prognostizieren einen Anstieg der global gemittelten bodennahen Lufttemperatur über den Zeitraum von 1990 bis 2100 um 1,4 bis 5,8 °C [BMBF 2003, 47].

6.2 Auftrag und Ansatz

„Das PIK ist bestrebt, die empirischen Resultate vieler Fachrichtungen und Einrichtungen aufzugreifen, aufzuarbeiten, zusammenzufassen und ihre Einsatzmöglichkeiten zu bewerten. Auf der so entstehenden vielschichtigen und vernetzten Datenbasis kann dann jene Gesamtheit von meteorologischen, hydrologischen, ökologischen, sozioökonomischen und transdisziplinären Modellen errichtet werden, die das wissenschaftliche Stammkapital des PIK bildet und rechnergestützt die Simulation vergangener, gegenwärtiger und zukünftiger Umweltwirklichkeiten ermöglicht.“¹⁹⁸

6.2.1 Kernauftrag des Institutes

Als Reaktion auf diese gravierenden Defizite wurde im Jahr 1992 auf Empfehlung des Wissenschaftsrates das Potsdam-Institut für Klimafolgenforschung (vgl. Abb. 6.1) gegründet. Der Kernauftrag des Institutes besteht im Auffinden von Möglichkeiten, die es erlauben, die Auswirkungen des anthropogenen Klimawandels auf einem verträglichen Niveau zu halten, sowie in der Suche nach geeigneten Maßnahmen zur Anpassung an die unvermeidliche Erwärmung der Erde mit ihren besonders ernsten Konsequenzen für die ärmsten Entwicklungsländer.

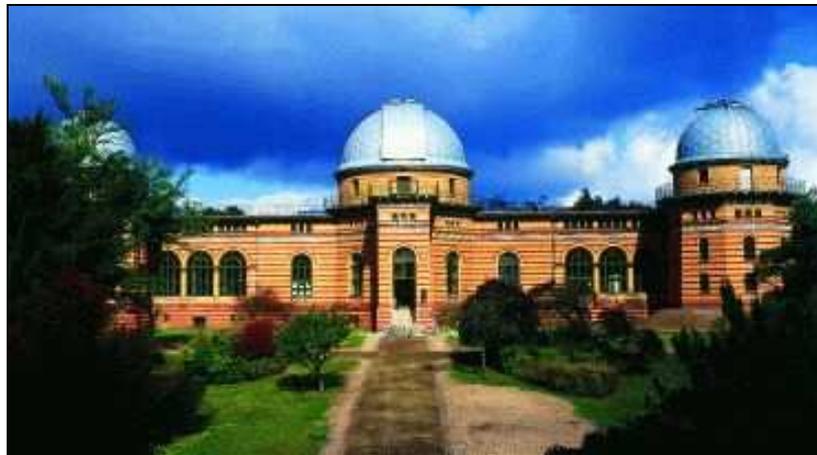


Abb. 6.1 - Das neue Hauptgebäude des Potsdam-Instituts für Klimafolgenforschung im ehemaligen Astrophysikalischen Observatorium des Wissenschaftsparks Albert Einstein in Potsdam¹⁹⁹.

6.2.2 Transdisziplinarität

Seinem komplexen Forschungsgegenstand entsprechend wurde das Institut einem radikal holistischen Ansatz verpflichtet – so sollen einerseits Wissenschaftler aller relevanten Disziplinen (bspw. Klimatologie, Ökologie, Wirtschaftswissenschaften, Systemanalyse etc.) eng zusammenarbeiten (sog. *horizontale Integration*); auf der anderen Seite sollen alle Aspekte des Problems, von seiner Formulierung bis hin zu Vorschlägen für Entscheidungsträger, gleichermaßen berücksichtigt werden (sog. *vertikale Integration*). Zwei fundamentale Einsichten charakterisieren Ausrichtung und Arbeitsweise des Institutes. Zum einen wurde erkannt, dass Klimawandel nur in einem größeren und komplexeren Betrachtungskontext, der die Dynamiken des globalen Wandels berücksichtigt und sowohl die geobiophysischen wie sozioökonomischen Prozesse einbezieht, erfolgreich erforscht werden kann. Entsprechend fand ein Übergang von einer eindimensionalen Klimafolgenforschung hin zur umfassenderen *Erd-systemanalyse* [Schellnhuber, Wenzel 1998] [Schellnhuber 1999] statt. Zum anderen wurde die klassische Struktur von Forschung, die entlang von Abteilungen und hauptsächlich entlang einzelner Wissenschaftsdisziplinen organisiert ist, früh als ungeeignet für den transdisziplinären Forschungsauftrag des PIK erkannt und eine eigene, innovative Struktur zur Organisation des Institutes entwickelt und etabliert (vgl. Kap. 6.3).

¹⁹⁸ Potsdam-Institut für Klimafolgenforschung: *Biennial Report 1996/1997* [PIK 1996/97, 97].

¹⁹⁹ Abbildung übernommen aus [PIK 2000/01, 87].

6.2.3 Die zentrale Rolle von Modellen, Simulation und Daten

Klimafolgenforschung und Erdsystemanalyse, wie sie am PIK durchgeführt werden, sind erst durch den Einsatz modernster Computertechnologie möglich. Es ist offensichtlich, dass die erforderlichen Experimente nicht mit dem System Erde selber - oder auch nur mit Teilsystemen - durchgeführt werden können. So können die potentiellen Auswirkungen eines dramatischen Klimawandels nicht in der Realwelt untersucht werden; zudem sind sowohl die geographischen Räume wie die Zeitskalen, die zu betrachten sind, bei weitem zu groß, um ihre verlässliche Erforschung durch Labor-Experimente zu erlauben.

Daher wird die wissenschaftliche Arbeit des PIK in gleichsam *virtuellen* Laboratorien durch Rechner und *Simulation* durchgeführt; sie basiert entsprechend in hohem Maße auf Modellierung²⁰⁰ und wissenschaftlichem Rechnen (*Scientific Computing*). Der Prozess des Scientific Computing besteht aus der Entwicklung einer abstrakten mathematischen Beschreibung des untersuchten Systems, der Übersetzung dieses mathematischen Modells in eine diskrete, computer-verarbeitbare Beschreibung, ihrer Implementierung und Verteilung auf modernen Rechenanlagen sowie der sorgfältigen Evaluierung und Interpretation der berechneten Ergebnisse. Am PIK wurde seit seiner Gründung ein reichhaltiger Fundus von Simulationsmodellen verschiedener Anwendungsbereiche angepasst und entwickelt, der beständig erweitert und vergrößert wird (vgl. auch Abb. 6.2).

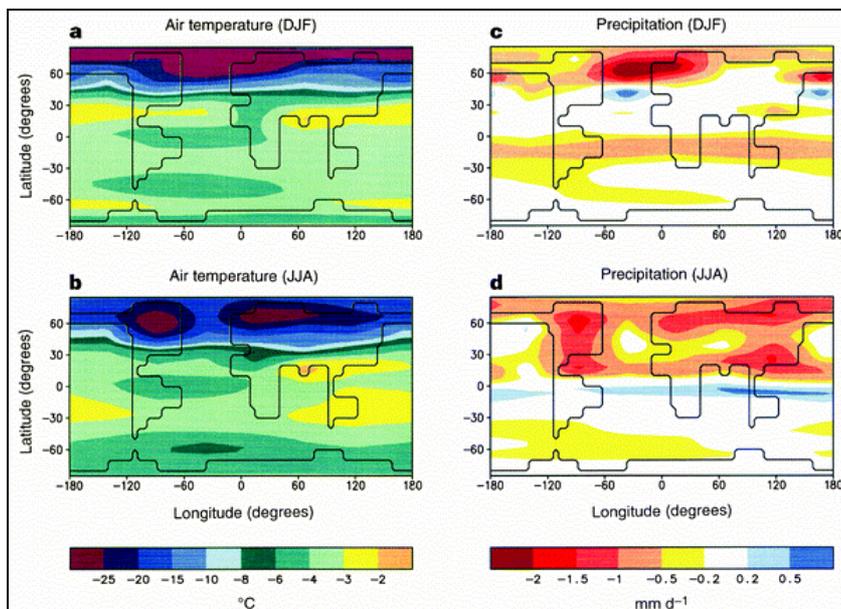


Abb. 6.2 - Visualisierung von Simulationsergebnissen des am PIK entwickelten Klimasystem-Modells CLIMBER²⁰¹: Dargestellt sind die Unterschiede zwischen simuliertem eiszeitlichem Klima vor 21.000 Jahren (Last Glacial Maximum, LGM) und simuliertem gegenwärtigem Klima in Temperatur (a und b) sowie Niederschlag (c und d). Die beiden oberen Bilder (a und c) zeigen die Wintermonate (Dezember, Januar, Februar; DJF), die beiden unteren (b und d) die Sommermonate (Juni, Juli, August; JJA).

Forschungsgegenstand und Methodik des Institutes bedingen eine Nutzung der besten weltweit verfügbaren Datenressourcen zu allen relevanten Teilsystemen des Erdsystems. Diese Datenbasis bildet insbesondere die Voraussetzung für die Entwicklung, den Betrieb und die Einschätzung geeigneter Modelle zur Simulation des Erdsystems. Die Verwendungsbereiche umfassen dabei alle Phasen im Lebenszyklus von Modellen und reichen

²⁰⁰ Die wissenschaftlichen Hintergründe der Modellierung realer Systeme sollen hier nicht vertieft werden. Für eine Einführung in dieses komplexe Forschungsgebiet eignet sich bspw. [Bossel 1989], ein umfassendes Kompendium hierzu bietet [Singh 1987].

²⁰¹ Abbildung übernommen aus [Ganopolski et al. 1998, Fig. 6].

von ihrem Entwurf (Modellentwicklung), ihrer Grobjustierung (Modellkalibrierung), ihrer Feinjustierung (Modellvalidierung), ihrer Versorgung mit Steuerwerten (Modelltriebkräfte) für konkrete Simulationsaufgaben bis hin zur Bewertung modellgenerierter Ergebnisdaten durch Vergleiche mit Referenzdaten und der Abschätzung von Unsicherheiten.

Zur Durchführung komplexer Simulationen des Erdsystems ist die Nutzung von Höchstleistungsrechnern unabdingbar. Seit 1994 stand dem Institut ein paralleler Supercomputer vom Typ IDM RS 6000 SP mit 77 Prozessoren zur Verfügung; im Oktober 2001 wurde der neue parallele High Performance Computer vom Typ IBM SP (240 Prozessoren) eingeweiht, der in der Rangliste der 500 leistungsstärksten Rechner der Welt vom November 2003 den 273. Platz einnimmt²⁰². Die Front-Ends an den Arbeitsplätzen der Wissenschaftler bilden High-End-Workstations und Personal Computer; zur Datenspeicherung steht eine Speicherhierarchie aus Plattenspeichern sowie eine robotergesteuerte Bandkassetten-Bibliothek (Tape Library) mit einer Kapazität von gegenwärtig rund 50 Terabyte zur Verfügung (vgl. Abb. 6.3). Die IT-Infrastruktur des Institutes wird durch ein Computernetzwerk komplettiert, das alle Ressourcen sowohl intern wie mit dem Internet verbindet.



Abb. 6.3 - Automatische Datensicherung im PIK: (a) IBM 3494 Tape Library und Festplatten-Subsysteme; (b) Zugriff auf Bandkassetten durch den Roboter der Tape Library²⁰³.

6.3 Forschungsstruktur

Die Struktur des PIK (vgl. Abb. 6.4) ist entlang von wissenschaftlichen Abteilungen, thematischen Forschungsgebieten (TOPIKs) und Projekten organisiert und wird durch schlanke Verwaltungs- und Serviceeinheiten ergänzt. Die gegenwärtig über 120 wissenschaftlichen Mitarbeiter des Institutes sind fünf wissenschaftlichen Abteilungen (vgl. Kap. 6.3.1) zugeordnet, die für Bereitstellung und Entwicklung der Methoden, Modelle und Datenpools verantwortlich sind, auf denen das interdisziplinäre Arbeitsprogramm des Institutes basiert. Die eigentliche Forschung findet in jeweils auf einige Jahre angelegten, multidisziplinären Forschungsgruppen (Projekten) statt, an denen prinzipiell Wissenschaftler aus mehreren Abteilungen teilnehmen. Diese Projekte werden von den sog. TOPIKs (vgl. Kap. 6.3.2) geführt, die die übergreifenden Forschungsthemen des PIK bilden.

6.3.1 Die fünf Abteilungen

Das PIK gliedert sich in fünf wissenschaftliche Abteilungen (Departments), deren Hauptaufgabe in der Entwicklung, Pflege und Unterstützung der Ressourcen für Projekte und TOPIKs sowie in der Sicherstellung der Qualität der wissenschaftlichen Ergebnisse besteht. Die Abteilungen sind in verschiedenen Wissenschaftsgebieten verankert, den Naturwissenschaften (Abteilungen Climate System sowie Global Change and Natural Systems),

²⁰² <http://www.top500.org/list/2003/11/>

²⁰³ Abbildung übernommen aus [PIK 1998/99a, 26].

im Feld der Sozioökonomie (Abteilung Global Change and Social Systems) sowie in „Strukturwissenschaften“ wie Systemtheorie, Mathematik und Informatik (Abteilungen Integrated System Analysis sowie Data and Computation).

▪ Integrated System Analysis

Da die Erforschung des Globalen Wandels viele Felder von Natur- und Sozialwissenschaften involviert, besitzt die Integration von Forschungsergebnissen unterschiedlichster Disziplinen entscheidende Bedeutung für ein verbessertes Verständnis der möglichen Optionen eines globalen Umweltmanagements. Ziel der Abteilung *Integrated System Analysis*²⁰⁴ ist die Entwicklung von Methoden und Konzepten, um die Forschung der verschiedenen Disziplinen des PIK - und darüber hinaus - zu integrieren. Der Fokus der Forschung liegt dabei weniger auf den Eigenschaften einzelner Subsysteme, sondern primär auf den Eigenschaften der *Interaktionen* zwischen Subsystemen.

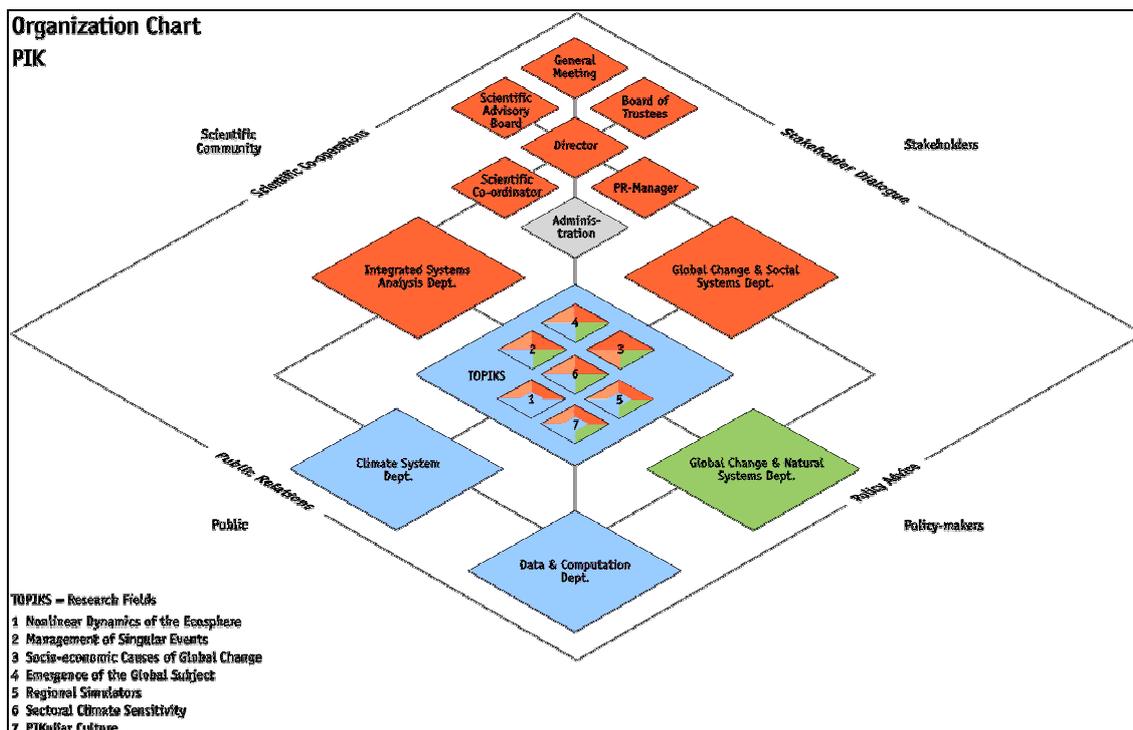


Abb. 6.4 - Die Organisationsstruktur des Potsdam-Instituts für Klimafolgenforschung²⁰⁵.

▪ Global Change and Natural Systems

Die Abteilung *Global Change and Natural Systems*²⁰⁶ untersucht die Sensitivität der verschiedenen Ökosysteme, die die Landoberfläche der Erde bedecken und elementare Ressourcen für Nahrung, Frischwasser etc. bereitstellen, gegenüber Umweltveränderungen. In diesem Kontext werden auch die physikalischen und biogeochemischen Rückkopplungen zwischen Erdsystem und Landoberfläche erforscht.

▪ Climate System

Das Verständnis des Zusammenspiels der verschiedenen Komponenten des Erdsystems bildet einen der Schlüssel für eine umfassende Erdsystemanalyse. Die Abteilung *Climate System*²⁰⁷ untersucht entsprechend das komplexe Verhalten von vergangenem, gegenwärtigem und zukünftigem Klima. Hierunter fallen statistische Analysen von Klimadaten

²⁰⁴ <http://www.pik-potsdam.de/dept/int/>

²⁰⁵ Abbildung übernommen aus [PIK 2000/01, 115].

²⁰⁶ <http://www.pik-potsdam.de/dept/nat/>

²⁰⁷ <http://www.pik-potsdam.de/dept/cli/>

über Vergangenheit und Gegenwart, die Konstruktion zukünftiger Klimaszenarien, die Entwicklung eines Klimasystem-Modells sog. mittlerer Komplexität²⁰⁸ sowie die Ozean-Modellierung.

▪ **Global Change and Social Systems**

Die Abteilung *Global Change and Social Systems*²⁰⁹ analysiert sozioökonomische Veränderungen, die für einen verantwortlichen Umgang mit den globalen Umweltveränderungen erforderlich sind. Sie erforscht Möglichkeiten für einen sog. Übergang zur Nachhaltigkeit (sustainability transition) sowie die Hindernisse, die diesem entgegenstehen, anhand der Bereiche Energie, Lebensstile und politisches Management.

▪ **Data and Computation**

Die Abteilung *Data and Computation (D&C)*²¹⁰ verwaltet die informationstechnologische Infrastruktur des Institutes und stellt seine hohen technologischen Standards sicher. Sie ist verantwortlich für wissenschaftliches Daten- und Metadaten-Management und betreibt eine Unterabteilung für wissenschaftliches Rechnen (Scientific Computing). Aufgrund der zentralen Bedeutung von Modellierung und wissenschaftlichem Rechnen für die Erdsystemanalyse bildet D&C das Rückgrat der am PIK betriebenen Forschung.

6.3.2 Von Kernprojekten zu TOPIKs

Von 1994 bis 2000 wurde die Forschungsarbeit des PIK in neun transdisziplinären sog. Kernprojekten (Core Projects) organisiert, in denen das Erdsystem aus globalen, regionalen und sektoralen Perspektiven erforscht wurde. Auf Empfehlung des Wissenschaftsrats ging das Institut nach Abschluss der Kernprojekte zu einem neuen Forschungsprogramm über und definierte sieben thematische Forschungsgebiete, die als TOPIKs bezeichnet werden und seit 2001 das Rahmenwerk für die Durchführung multidisziplinärer Projekte bilden. Die TOPIKs adressieren eine Vielzahl breit angelegter thematischer Felder:

- ▶ die nichtlineare Dynamik der Ökosphäre;
- ▶ den Umgang mit Extremereignissen wie Sturm-, Dürre- oder Flutkatastrophen;
- ▶ die sozioökonomischen Triebkräfte eines globalen Umweltwandels;
- ▶ eine globale vernetzte Zivilisation;
- ▶ die Entwicklung regionaler Simulatoren;
- ▶ die Verwundbarkeit ökonomischer Sektoren durch den Klimawandel.

Ergänzt werden diese Gebiete durch ein siebtes TOPIK, das

- ▶ die Definition übergreifender Prioritäten und Qualitätsstandards sowie einen interdisziplinären Dialog zur Reflektion über den philosophischen Hintergrund der am PIK betriebenen Forschung

verfolgt. Ende 2003 fand die auf drei Jahre angelegte Forschung anhand der ersten Runde der innerhalb dieser TOPIKs durchgeführten Projekte ihren Abschluss; das PIK definierte den nächsten Satz von dreijährigen TOPIK-Projekten, die 2004 begonnen haben.

6.4 Herausforderung Datenbasis

Das PIK betritt von seiner Konzeption her wissenschaftliches Neuland; die dort betriebene transdisziplinäre Erdsystemanalyse ist eine neue und sich dynamisch entwickelnde Wissenschaft. Die beteiligten Wissenschaftler können daher nicht vollständig vorab definieren, welche Daten genau in Zukunft benötigt werden; die hierfür benötigte Datenbasis verändert sich vielmehr im Forschungsprozess. Der gesamte gegenwärtig und zukünftig einzubezie-

²⁰⁸ Das Klimasystem-Modell CLIMBER (für CLIMate and BiospheRE) (vgl. bspw. [Ganopolski et al. 1998] [Ganopolski, Rahmstorf 2001] sowie Abb. 6.2).

²⁰⁹ <http://www.pik-potsdam.de/dept/soc/>

²¹⁰ <http://www.pik-potsdam.de/dept/dc/>

hende Datenraum ist damit nicht vollständig beschreibbar; sein Umfang und seine Komplexität führen zu besonderen Herausforderungen, die in diesem Kapitel skizziert werden.

6.4.1 Notwendigkeit der Kompilierung

Die für die Arbeit des Institutes erforderliche umfassende Datenbasis kann aufgrund der Vielzahl der einzubeziehenden Bereiche von keiner einzelnen existierenden Einrichtung bereitgestellt werden. Die Datenbasis existiert damit nicht in einem Zustand, der ihre sofortige Nutzung erlaubt, sondern muss zunächst durch die Einbeziehung der komplexen und heterogenen Datenressourcen einer Vielzahl nationaler und internationaler Einrichtungen angenähert werden.

Bundesanstalt für Gewässerkunde (BFG) Koblenz http://www.bafg.de	
Carbon Dioxide Information Center (CDIAC) Oak Ridge, Tennessee, USA http://cdiac.esd.ornl.gov/	
Climatic Research Unit (CRU) University of East Anglia, Norwich, UK www.cru.uea.ac.uk/	
Deutscher Wetterdienst (DWD) Offenbach http://www.dwd.de/	
International Energy Agency (IEA) Paris http://www.iea.org/	
National Aeronautics and Space Administration (NASA) Washington, DC http://www.nasa.gov/	
National Center for Atmospheric Research (NCAR) Boulder, Colorado http://www.ncar.ucar.edu/ncar/	
National Climatic Data Center (NCDC) Asheville, North Carolina http://www.ncdc.noaa.gov/oa/ncdc.html	
Organisation for Economic Co-operation and Development (OECD) Paris http://www.oecd.org/home/	
United Nations (UNO) http://www.un.org/	
World Meteorology Organization (WMO) Genf http://www.wmo.ch/	
World Bank (Weltbank) Washington, DC http://www.worldbank.org/	

Tab. 6.1 - Datengeber des PIK (Auswahl).

Das PIK verwendet entsprechend Daten unterschiedlicher Institutionen, die als *Datengeber* des Institutes fungieren. Tab. 6.1 gibt einen Überblick über einige der weltweit verteilten Datengeber des Institutes. So dokumentierte allein die *meteorologische* Datenbasis des

PIK bereits im Jahr 1999 mehr als 2300 Erhebungsstationen aus allen fünf Kontinenten und integrierte unter anderem Messdaten von den nationalen meteorologischen Diensten in Australien, China, Deutschland, England, Finnland, Frankreich, Griechenland, Hongkong, Israel, Island, Japan, Südkorea, Malaysia, Neuseeland, Niederlande, Österreich, Polen, Schweiz, Schweden, Singapur, Türkei, UdSSR und den USA [Österle et al. 1999].

6.4.2 Immanente Komplexität und Heterogenität

Die für die Erdsystemanalyse relevante Datenbasis ist durch eine *hochgradige Komplexität und Heterogenität* gekennzeichnet, die entsprechend der Vielfalt der einzubeziehenden Aufgabenbereiche, Wissenschaftsdisziplinen und Forschungsfragen in vielfältigen Ausprägungen auftritt. Die Daten variieren dabei u.a. in folgenden Aspekten:

- **Heterogene Themengebiete**

Die Daten beziehen sich nicht auf ein einziges, eng abgegrenztes Themengebiet, sondern auf eine Vielzahl unterschiedlicher naturwissenschaftlicher und sozioökonomischer Sektoren und Untergebiete.

- **Heterogene Quellen**

Die Daten stammen weder aus einer Hand noch aus einem einzigen Erzeugungsprozess, sondern von einer Vielzahl nationaler und internationaler Datengeber und sowohl aus Messungen, der Erhebung empirischer Daten oder Simulationen.

- **Heterogene Raumbezüge**

Die überwiegende Mehrheit der Daten steht in Bezug zu bestimmten geographischen Positionen, Orten, natürlich gegebenen oder virtuell definierten Gebieten. Die Daten weisen dabei keinen einheitlichen Raumbezug auf, sondern liegen auf diversen Skalen, für diverse Prozesse und in diversen Auflösungen vor.

- **Heterogene Zeitbezüge**

Für die überwiegende Mehrheit der Daten ist zusätzlich ein Zeitbezug gegeben. Die Zeitbezüge finden sich ebenfalls in vielfältigen Ausprägungen bezüglich Perioden, Ereignissen oder Auflösungen.

- **Heterogene Beschreibungsformate**

Der Vielfalt der Daten entspricht eine große Zahl heterogener Beschreibungsformate; sie werden beispielsweise anhand von Punkten, Gittern oder Vektoren beschrieben sowie in einer Vielzahl proprietärer Formate abgespeichert.

- **Heterogene Datenorganisation**

Für die logische Organisation der Daten finden vielfältige Formen Anwendung; so können Daten in einzelnen Datenbanken, in Dateien oder Dateisystemen organisiert sein.

- **Heterogene Speichermedien**

Die Heterogenität der relevanten Daten spiegelt sich auch in der Vielfalt der zu ihrer persistenten Speicherung verwendeten Medien. So finden im PIK als Speichermedien u.a. Festplatten, CD-ROMs sowie - für Datensicherung wie für die Vorhaltung von Massendaten und seltener genutzten Daten - ein automatisches Bandarchivierungssystem mit Zugriff über Bandroboter (vgl. auch Abb. 6.3) Verwendung.

6.4.3 Notwendigkeit der lokalen Bereitstellung

Die dem PIK zur Verfügung gestellten Datenressourcen werden im Institut zusammengeführt und dort für die wissenschaftliche Arbeit vorgehalten. Für die lokale Bereitstellung der Daten sind insbesondere Zeitgründe, Aufwandsreduzierung und Qualitätssicherung ausschlaggebend.

▪ Zeitgründe

Auch im Zeitalter von Internet und World Wide Web ist der Zugriff auf externe wissenschaftliche Daten ein zeitaufwendiger Prozess. Angeforderte Daten können in einigen Fällen beim Datengeber online abgerufen werden; oft werden sie allerdings erst mit einer Verzögerung, die einige Tage umfassen kann, bereitgestellt. Die lokale Bereitstellung einmal erhaltener Daten stellt ihre schnelle Verfügbarkeit sicher.

▪ Aufwandsreduzierung

Daten zu einzelnen wissenschaftlichen Themengebieten müssen jeweils aus mehreren Quellen zusammengeführt und zunächst in einen gemeinsam bearbeitbaren Zustand gebracht werden. Die lokale Vorhaltung aus unterschiedlichen Quellen zusammengeführter Daten reduziert diesen Aufwand auf jeweils einmalige Prozesse.

▪ Qualitätssicherung

Obwohl die jeweiligen Ausgangsdaten von den Datengebern geprüft werden, stellen sich bei ihrer Zusammenführung Fehler in den Daten heraus [Österle et al. 1999]. Das Institut benötigt jedoch eine hochqualitative Datenbasis und führt entsprechende Qualitätssicherungen bspw. durch die Anwendung komplexer statistischer Prüfmethoden durch. Die lokale Vorhaltung qualitätsgesicherter Daten erlaubt die zentrale Bereitstellung hochqualitativer Datensammlungen für die Wissenschaftler des Institutes.

6.4.4 Notwendigkeit der Spezialisierung

Die erforderlichen Prozesse von Zusammenführung, Aufbereitung und Qualitätssicherung wissenschaftlicher Daten setzen jeweils disziplinspezifisches Expertenwissen voraus. Aufgrund der großen Vielfalt der benötigten Daten werden diese Prozesse im Institut für spezifische thematische Gebiete jeweils von Fachwissenschaftlern einzelner Disziplinen durchgeführt.

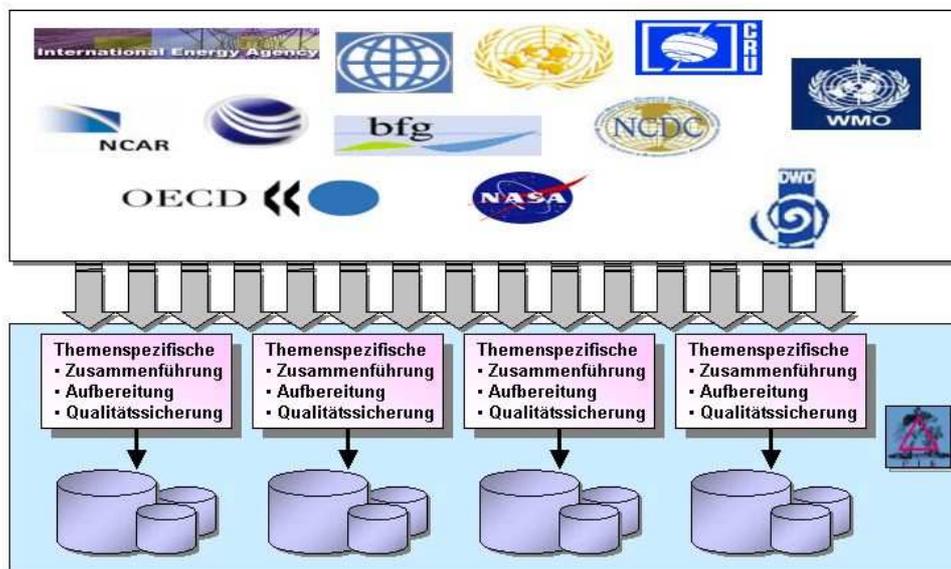


Abb. 6.5 - Datenbereitstellung im PIK.

6.4.5 Entstehen getrennter Datenräume

Da einerseits eine schnelle Verfügbarkeit geeigneter Daten für die darauf basierende Forschungstätigkeit gefordert ist und andererseits Spezialisten aus den jeweiligen Fachdisziplinen zur Zusammenführung und Qualitätssicherung diesbezüglicher externer Daten unabdingbar sind, wurde vor diesem Hintergrund die Definition eines einheitlichen Datenbankmodells auch nur für thematische Teilausschnitte als aufwendiger, langwieriger und fehlerträchtiger Prozess eingestuft. Entsprechend wurde ab 1994 mit dem Aufbau jeweils

spezifischer Datenbanken in der Kompetenz von Spezialisten unterschiedlicher Abteilungen oder Projekte des Institutes begonnen. Abb. 6.5 veranschaulicht die schematische Infrastruktur des PIK für die Bereitstellung zusammengeführter und qualitätsgesicherter Daten aus externen Quellen.

6.5 Wissenschaftliches Datenmanagement

Die beständig ansteigende Vielschichtigkeit und Komplexität der Datenbasis des PIK erfordert besondere Strategien, um ihre effiziente Nutzung durch die individuellen Wissenschaftler zu ermöglichen. Diese Aufgabe wird von der *Scientific Data Management Group* (SDM)²¹¹ der Abteilung Data and Computation adressiert, in der auch der Autor dieser Arbeit tätig ist. Da es als wenig aussichtsreich erscheint, für die veränderliche und daher nicht final beschreibbare Datenbasis des Institutes ein einzelnes, universelles Datenmodell zu entwickeln, verfolgt das PIK einen kombinierten dreifachen Ansatz zum wissenschaftlichen Datenmanagement:

- ▶ *Verteiltes Datenmanagement*, organisiert von den einzelnen wissenschaftlichen Arbeitsgruppen (scientific working units) des Institutes, die innerhalb von Projekten und / oder Abteilungen verankert sein können;
- ▶ *Zentrales Datenmanagement* für die Bereitstellung von Datenräumen von übergreifendem Interesse für die Wissenschaftler des Institutes, organisiert von der Scientific Data Management Group; sowie
- ▶ *Zentrales Metadatenmanagement* für die Bereitstellung von notwendigen Informationen über die heterogenen Datenräume des Institutes, ebenfalls organisiert von der Scientific Data Management Group.

6.6 Ziel – Eine geeignete Schnittstelle zur autonomen Datenerschließung

Um eine komfortable Erschließung der für die Forschungsarbeit besonders relevanten multidisziplinären Metadaten- und Datenräume des PIK zu ermöglichen, beschloss die Scientific Data Management Group Ende 1996 die Entwicklung einer geeigneten *graphisch-interaktiven netzwerkfähigen Zugriffsschnittstelle*. Diese Schnittstelle soll individuellen Wissenschaftlern eine intuitive Navigation über wesentlichen, komplexen und heterogenen Datenräumen des Institutes erlauben und sie bei einer *autonomen und flexiblen Selektion von Daten* aus diesen unterstützen. Die Konzeption und Realisierung dieser Schnittstelle fand durch den Autor im Rahmen der vorliegenden Dissertation statt und wurde in enger Zusammenarbeit mit der Scientific Data Management Group des PIK durchgeführt.

Ziel der Schnittstellenentwicklung war ein Beitrag zur Verringerung existierender Defizite bei der autonomen und individuellen Datenversorgung der Wissenschaftler im Kontext von Erdsystemanalyse und PIK. Dazu wurde eine spürbare *reale* Verbesserung der gegebenen Ausgangssituation angestrebt. Es war eine Softwarelösung zu entwickeln, die sich als geeignet für eine autonome Datenversorgung in der wissenschaftlichen Praxis des Institutes erweist und insbesondere von dessen Wissenschaftlern auch *tatsächlich eingesetzt* wird. Die Evidenz der gewählten Lösung wird damit an ihrer *Akzeptanz* durch die heterogene Wissenschaftlergemeinschaft des Institutes zu messen sein.

²¹¹ http://www.pik-potsdam.de/dept/dc2/dc_sys_sdm.html