

World Data Center for Climate - Produkte und Services für die Klimaforschung

F. Toussaint, M. Lautenschlager

Einleitung

Seit über drei Jahren betreibt die an das Max-Planck-Institut für Meteorologie angeschlossene Gruppe "Modelle und Daten" das Weltdatenzentrum für Klimadaten (WDC-Climate, siehe [1]). Damit sind rund 200 Terabyte Archivdaten für Forscher in Deutschland, Europa und der übrigen Welt über das Internet abrufbar. Allein 2005 wurden über 600000 Downloads registriert, die insgesamt auf 43 Terabyte Daten zugegriffen.

Schwerpunkt des Datenbestandes bilden Daten aus Klimamodellrechnungen, aber auch zugeordnete Beobachtungsdaten aus verschiedenen Projekten stehen zur Verfügung.

WDC-Climate

Datenzentren weltweit

Das seit 1933 bestehende International Council for Science (ICSU [2]) ist eine der ältesten Wissenschaftsorganisationen weltweit. Aus Anlass des Geophysikalischen Jahres 1957/58 gründete sie den Verbund der Weltdatenzentren [3], dem heute 52 Institute in zwölf Ländern angehören. Sie bieten der Forschungsgemeinschaft wissenschaftliche Daten aus verschiedenen Bereichen der Geowissenschaften.

2003 schlossen sich drei deutsche Weltdatenzentren zum „WDC Cluster Earth System Research“ zusammen. Neben dem WDC-Climate sind es die Datenzentren für Marine Umweltdaten (WDC-MARE an Universität Bremen und Alfred-Wegener-Institut, Bremen) und für Fernerkundungsdaten der Atmosphäre (WDC-RSAT beim Deutschen Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen). Mit dem in Gründung befindlichen WDC-TERRA des Potsdamer Geoforschungszentrums besteht enge Zusammenarbeit. Ziel der Kooperation sind die Sicherstellung von Langzeitdatenhaltung, Erstellung und Betrieb zuverlässiger Systeme für Datenverwaltung und -zugriff, gegenseitige Begutachtung wissenschaftlicher Daten sowie die Entwicklung gemeinsamer Standards und Protokolle.

Aufgaben des Weltklimadatenzentrums

Das WDC-Climate sammelt, archiviert und verteilt Klima(-modell-)datenprodukte um sie der internationalen Forschungsgemeinschaft zur Verfügung zu stellen. Andere Datentypen wie Modellrohdaten oder Satellitenrohdaten werden nicht vorgehalten, da dies die vorhandenen Kapazitäten überstiege bzw. zu den Aufgabenbereichen anderer Institute gehört. Um eine gute Vernetzung solcher Datenzentren der Erdsystem-

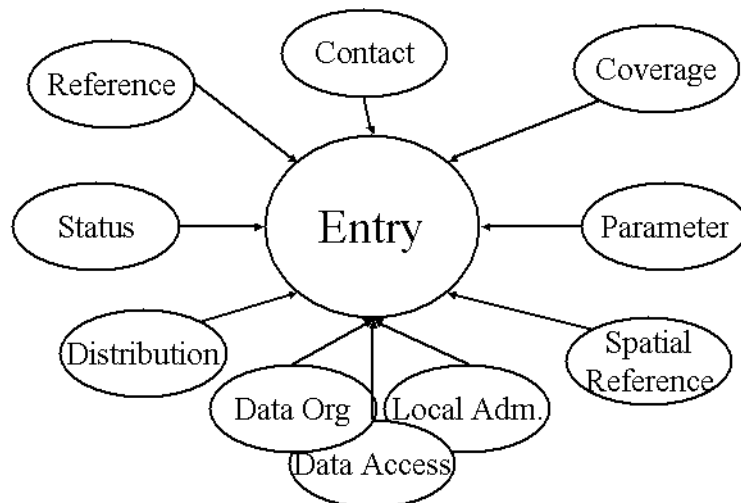


Abb. 1 Tabellengruppen des CERA-2 Datenbankmodells

operationen mit Universitäten und Großforschungseinrichtungen ebenso bedeutsam wie die Zusammenarbeit mit Instituten der Max-Planck-Gesellschaft oder dem Deutschen Wetterdienst.

Technische Grundlagen

Das WDC-Climate nutzt neben eigener Hardware vor allem die technischen Einrichtungen des Deutschen Klimarechenzentrums (DKRZ [4]). Dazu gehören unter anderem mehrere STK Bandsilos von insgesamt 3.4 Petabyte Kapazität. Daten, die längere Zeit nicht nachgefragt werden, können daher auf Band migriert werden. Andererseits sind häufiger verwendete Einträge auf den Festplatten kurzfristig zum Download verfügbar.

STD-DOI:

Hinter STD-DOI oder Scientific and Technical Data - Digital Object Identifier verbirgt sich das Konzept der Veröffentlichung wissenschaftlicher Daten als eigenständige Entitäten analog zu klassischen Publikationen.

Analog zu Veröffentlichungen in wissenschaftlichen Zeitschriften werden publikationsfähige Datenentitäten einem Reviewprozeß und einer Qualitätssicherung unterzogen, mit einem Persistent Identifier, hier DOI, versehen und in das globale Handle-System eingetragen. Das Metadaten-Profil der STD-DOIs enthält alle Einträge für eine elektronische Publikation. Die publizierten Datenentitäten werden zur Zeit im Katalog der Technischen Informationsbibliothek (TIB), Hannover veröffentlicht und sind zusammen mit Zeitschriftenartikeln such- und zugreifbar (Übersicht unter <http://tws.gbv.de> mit Sucheintrag "exk+primaerdaten"). Publiizierte Datenentitäten werden in und analog zu Zeitschriftenartikeln zitiert mit Autor, Erscheinungsjahr und Titel, so dass Datenproduzent und Datenentität stets zusammen genannt werden. Dieses Konzept wurde im Rahmen eines DFG Projektes entwickelt und prototypisch implementiert (Information unter www.std-doi.de).

forschung zu erreichen, besten Kollaborationen mit zahlreichen Instituten verwandter Fachrichtungen wie Erdbeobachtung, Meteorologie, Ozeanografie, Paläoklimatologie und Umweltforschung.

Obwohl das WDC-Climate international ausgerichtet ist und Wissenschaftlern aus aller Welt offen steht, erfolgt der überwiegende Teil der Datenzugriffe aus dem Bereich der deutschen Forschungslandschaft. Dabei sind die Ko-

Für Datenverwaltung, -katalogisierung, -suche, -anfrage und -zugriff kommt das relationale CERA-Datenmodell [5] zur Anwendung. In diesem 1997 [6] in Zusammenarbeit mit dem Potsdam-Institut für Klimafolgenforschung und dem Alfred-Wegener Institut erstellten Datenbank-Entwurf (Lautenschlager, Toussaint, Thiemann, Reincke 1997) werden Daten und Metadaten integriert abgespeichert und dadurch Katalogisierung und Datenzugriff gemeinsam verwaltet.

Das Format der gespeicherten Daten ist überwiegend WMO-GRIB, aber das NetCDF-Format gewinnt an Bedeutung. Einzelne Datensätze werden auch in proprietären Formaten oder als ISO-Klartext gehalten.

Der Datenkatalog

Die Metainformationen aller im WDC-Climate gehaltenen Daten können von Nutzern weltweit abgerufen werden. Dabei ist der Katalogzugang [6] kostenlos und anonym. Erst für den Download der Daten benötigt man einen kennwortgeschützten, persönlichen Zugang. Dabei können neben dem Datensatz selbst auch Zeitraum und geografische Begrenzungen eingegeben werden, um die Datentransfervolumina niedrig zu halten.

The screenshot displays the 'Parameter information for dataset ERA40_PL00_6H_D1' page. The interface is divided into several sections:

- Variable:** Topic: oceans, Parameter: sea/ice, Variable: divergence, Variable description: divergence.
- Code:** Number: 155, Type: ERA40, Acronym: D, Description: divergence.
- Unit:** Acronym: 1/s, Name: 1/second, Description: not filled.
- Aggregation:** Description: 6 hours interval.
- Temporal structure (period 1 of 1):**

	Start	Increment	Date / time information
Year	1957	0	Period start: 01/09/1957 00:00
Month	9	0	Period end: 31/08/2002 18:00
Day	1	0	
Hour	0	6	

On the right side, there is a 'CERA entry information for ERA40_PL00_6H' section with a table of general information:

Property	Value
Acronym	ERA40
Name	ERA40 (1957-2002) (Atmosphere, pressure level)
Type	Experiment
Creation date	2001/02/08
Release date	2001/02/08
Future release date	2006/02/08
Progress	in work
Summary	ERA40 - The new reanalysis project ERA40 will cover the period from mid-1957 to 2002 including the reanalysis of the sea/ice, 1957-2002. The new reanalysis project ERA40 will cover the period from mid-1957 to 2002 including the reanalysis of the sea/ice, 1957-2002.
Project(s)	ERA40
Acronym	ERA40
Name	ERA40 (1957-2002) (Atmosphere, pressure level)
Description	The new reanalysis project ERA40 will cover the period from mid-1957 to 2002 including the reanalysis of the sea/ice, 1957-2002.
Quality	
Accuracy report	not available
Completeness report	complete

Abb. 2 Der Datenkatalog kann anonym und kostenlos im WWW genutzt werden [6].

WMO-GRIB/NetCDF:

WMO-GRIB (GRIdded Binary) and NetCDF (NETwork Common Data Format) sind selbstbeschreibende Datenformate zum Archivieren und Austausch von Datenfeldern.

Beide können zusätzlich zu den Daten auch Informationen über die Datenverteilung (grid) o.ä. enthalten. Das WMO-GRIB ist sehr streng definiert und bietet kaum Freiräume; NetCDF hingegen ist sehr frei und wird von der Nutzergemeinschaft durch "conventions" (Coards, CF) für spezielle Anwendungen eingeschränkt. Für beide Formate gibt es Libraries, die es dem Anwender ermöglichen, Programme zu erstellen, um entsprechend formatierte Files zu schreiben oder zu lesen. Es existieren aber auch fertige Programmpakete. Mehr Informationen gibt es im WEB unter <http://www.mad.zmaw.de/postprocessing-imdi>.

Daten für die Forschung

Klimamodelldaten

Beim überwiegenden Teil der im WDC-Climate vorgehaltenen Daten handelt es sich um Ergebnisse der Modellrechnungen. Dazu gehören vor allem die in Hamburg und an anderen Forschungszentren vorgenommenen Modellierungen der IPCC-Szenarien aber auch Reanalysedaten wie die vom Europäischen Zentrum für mittelfristige Wettervorhersage (ECMWF, ERA-40) und vom National Center for Environmental Prediction (NCEP). Darüber hinaus sind die Ergebnisse paläoklimatischer Rechnungen und verschiedener beobachtungsorientierter Forschungsprojekte verfügbar. Für das IPCC der UN ist das WDC-Climate das Datenverteilzentrum (IPCC-DDC) für Studien zur Anpassung an und Vermeidung von Folgen möglicher Klimaänderungen.

Datenhaltung

In den letzten Jahren hat das Management von Projektdaten an Bedeutung gewonnen, wobei auch die Infrastruktur des WDC-Climate genutzt wird. Dadurch wird es den Projekten ermöglicht, neben der Nutzung von Langzeitarchivierung im WDCC bereitgestellte Daten sowohl im Katalog effizient zu suchen, als auch sofort online auf sie zuzugreifen. Der Zugriff kann mit eigenen Anwendungen erfolgen, die dann direkt über Internet an die Datenbank koppeln. In Absprache mit den Projektleitungen kann dabei sichergestellt werden, dass entsprechend den Regeln guter wissenschaftlicher Praxis alle Daten zehn Jahre nach Projektende verfügbar bleiben.

Zur Unterstützung des Projektdatenmanagements hat das WDCC in enger Kooperation mit dem WDC Cluster Earth System Research und der Technischen Informationsbibliothek in Hannover im Rahmen eines DFG Projektes das STD-DOI Konzept [7] entwickelt und prototypisch implementiert. Die wichtigsten Ziele des neuen Umgangs mit Primärdaten sind also langfristige und allgemein zugängliche Speicherung.

Durchsetzbar ist dies am besten über eine persönliche Motivation der Wissenschaftler. Dies ließe sich nach Meinung der Arbeitsgruppe durch zwei Faktoren befördern: (1) Daten sind nach diesem Konzept nicht mehr ausschließlich Teil einer wissenschaftlichen Veröffentlichung, sondern besitzen eine eigenständige Identität. (2) Damit werden Primärdaten, ähnlich wie Zeitschriftenartikel publizierfähig und zitierbar.

Quellenverzeichnis

Alle URL beziehen sich auf den Januar 2006

Lautenschlager Michael, Frank Toussaint, Hannes Thiemann, Manfred Reinke, Deutsches Klimarechenzentrum, Technical Report No 15, Hamburg, Juli 1998, www.dkrz.de

[1] <http://wdc-climate.de/>

[2] <http://www.icsu.org/>

[3] <http://www.ngdc.noaa.gov/wdc/wdcmain.html>

[4] <http://www.dkrz.de/>

[5] <http://www.mad.zmaw.de/wdc-for-climate/cera-data-model/>

[6] <http://cera.wdc-climate.de/>

[7] <http://www.doi-std.de/>