

Abschlussbericht



Gruppe „Modelle und Daten“ (M&D)

Autoren:

Dr. Michael Lautenschlager
Prof. Dr. Jochem Marotzke
Dr. Stephanie Legutke
Dr. Frank Toussaint
Hannes Thiemann
Heinke Höck

BMBF- Projekt des Max-Planck-Instituts für Meteorologie
Förderzeitraum: 01.01.2000 bis 31.12.2009, Förderkennzeichen: 01LG0601B

Inhaltsverzeichnis

I.	Kurzfassung Abschlussbericht	3
1.	Aufgabenstellung	3
2.	Voraussetzung für das Vorhaben	3
3.	Planung und Ablauf des Vorhabens	4
4.	Wissenschaftlicher und technischer Stand	5
5.	Zusammenarbeit mit anderen Stellen	6
II.	Abschlussbericht Gruppe „Modelle und Daten“	7
1.	Verwendung der Zuwendung	7
1.1.	Ziele – Veränderung und Verwirklichung	8
1.2.	Arbeit des Wissenschaftlichen Lenkungsausschusses	10
1.3.	Öffnung des Deutschen Klimarechenzentrums für die deutsche Klimaforschungsgemeinschaft	14
1.4.	Gemeinschaftsmodelle und Modellierungsumgebung	18
1.5.	Konsortialrechnungen	26
1.6.	CERA Klimadatenbank und ICSU World Data Center Climate	33
2.	Notwendigkeit und Angemessenheit der Arbeit	51
3.	Verwertung des Ergebnisses	51
4.	Bekannte Entwicklungen auf dem Gebiet	51
5.	Veröffentlichungen	51

I. Kurzfassung Abschlussbericht „Modelle und Daten“

1. Aufgabenstellung

Im ersten Projektantrag zur Finanzierung der Gruppe "Modelle und Daten" (M&D) als Service-Einrichtung der deutschen Klimaforschung waren die Aufgaben definiert als Bereitstellung von anwendungsbereiten Klimamodellen und Hilfe beim Einsatz solcher Modelle, Bereitstellung von beobachteten und simulierten Klimadaten zum Zwecke der Klimadiagnostik, und die Durchführung von aufwendigen, in der Klimaforschungsgemeinschaft abgestimmten Rechnungen (Konsortialrechnungen) mit Klimamodellen.

Begleitend zum Projekt „Modelle und Daten“ wurde ein Wissenschaftlicher Lenkungsausschuss (WLA) vom BMBF ins Leben gerufen. Die strategische Ausrichtung und damit Aufgabenstellung und Priorisierung der Aktivitäten von M&D wurden im WLA über die Projektlaufzeit hinweg betrachtet und an die Bedürfnisse der nationalen Klima-Community angepasst. Ziele des WLA waren es, die Klima-Community an Entscheidungsprozessen zur thematischen Ausrichtung von M&D sowie zur Nutzung des DKRZ zu beteiligen und eigene Entscheidungen transparent zu machen, um so die Akzeptanz von M&D und DKRZ als nationale Service-Einrichtungen der Klimaforschung zu stärken.

Der WLA konkretisierte die Aufgaben aus dem Erstantrag. Über die zehn jährige Projektlaufzeit führte die Aktivität des WLA zur deutlichen Verschiebung von Arbeitsschwerpunkten:

- Die direkte Unterstützung bei Entwicklung und Pflege von Klimamodell-Code wurde verlagert in die Entwicklung und Implementierung einer Infrastruktur zum effizienten Betrieb komplexer Klimamodelle auf den Rechnersystemen des DKRZ.
- Die parallel zur Infrastrukturentwicklung erforderliche Optimierung des Modellcodes selbst für neue Rechnerarchitekturen wurde als Arbeitsschwerpunkt des DKRZ realisiert.
- Konsortialrechnungen als zentrales Steuerungselement zur Nutzung des DKRZ wurden in der Klima-Community etabliert.
- Wissenschaftliches Datenmanagement zur effizienten Langzeitarchivierung vollständig dokumentierter Klima(modell)daten und deren Integration in internationale Datenföderationen wurde als Alleinstellungsmerkmal von M&D und DKRZ ausgebaut.

Dieses in der Projektlaufzeit von M&D und in der Klima-Community abgestimmte Aufgabenspektrum hatte auch nach Projektende und Reintegration von M&D ins DKRZ Bestand.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Bis ins Jahr 1999 wurde der oben erwähnte Aufgabenbereich von der „Modellbetreuungsgruppe“ am DKRZ wahrgenommen. Die Aufgaben wurden vom Gutachterausschuss des BMBF (Bundesministerium für Bildung und Forschung) im

Jahre 1996 befürwortet. Allerdings empfand das BMBF die bisherige Finanzierungsform aus haushaltstechnischen Gründen als langfristig nicht haltbar. An Stelle der Integration der „Modellbetreuungsgruppe“ in das DKRZ (Deutsches Klimarechenzentrum GmbH), das von der Max-Planck-Gesellschaft (MPG), der Universität Hamburg, dem GKSS Forschungszentrum und dem Alfred-Wegener-Institut (AWI) betrieben wird, wurde eine Anbindung an eine institutionell geförderte Forschungseinrichtung angeregt. Die Max-Planck-Gesellschaft hat diese Aufgabe übernommen.

Obwohl die erste Förderungsperiode nur einen Zeitraum von drei Jahren umfasste, sollte M&D nach Vorstellung der DKRZ Gesellschafter eine Dauereinrichtung sein. Das BMBF begrenzte seine grundsätzliche Finanzierungsbereitschaft auf zehn Jahre.

3. Planung und Ablauf des Vorhabens

Das Max-Planck-Institut für Meteorologie (MPI-M) betreute die Gruppe im operationellen Sinn. Strategische Entscheidungen wurden von einem neu etablierten Steuerungsausschuss, dem wissenschaftlichen Lenkungsausschuss (WLA), gefällt. Dies betraf insbesondere die Auswahl der von der Gruppe unterstützten Modelle, der betreuten Datenbestände, sowie Art, Umfang und Priorität von Konsortialrechnungen.

Mit dem Wechsel in der Verantwortung wurde zugleich eine stärkere Berücksichtigung der Interessen der Klimaforschungsgemeinschaft in Deutschland bei den Aktivitäten von M&D angestrebt. Der WLA als Steuerungs- und Koordinierungsgremium besaß am Ende des Projektes große Akzeptanz in der Klima-Community.

Die wesentlichen Ergebnisse des Projektes „Modelle und Daten“ lassen sich in sechs Punkten zusammenfassen:

1. Unter Mitwirkung des WLA wurden M&D und DKRZ als nationale Service-Einrichtungen für die Klimaforschung etabliert bzw. geöffnet.
2. Zur effektiven Nutzung der begrenzten Personalressourcen von M&D wurde in den Fachdisziplinen der Erdsystemforschung ein Abstimmungsprozess zur Identifizierung einzelner, durch M&D zu betreuender Gemeinschaftsmodelle durchgeführt. Dieser Abstimmungsprozess wurde vom WLA geleitet.
3. Die Optimierung der Nutzung des HLRE (Höchstleistungsrechner für die Erdsystemforschung) wurde von DKRZ und M&D gemeinsam vorangetrieben. Während sich das DKRZ auf die Optimierung der Modellcodes fokussierte, konzentrierten sich die Bemühungen von M&D auf die Entwicklung und Optimierung einer Modellierungsumgebung für die Gemeinschaftsmodelle auf dem HLRE.
4. Das Instrument der Konsortialrechnungen zur effektiven Nutzung des HLRE wurde von WLA, M&D und DKRZ entwickelt und von M&D bei den Rechnungen für den 4. Sachstandsbericht (AR4) des IPCC (Intergovernmental Panel for Climate Change (<http://www.ipcc.ch>)) erstmalig umgesetzt. Die Konsortialrechnungen als Instrument zur

Nutzung des Rechners durch „Flaggschiff“-Projekte ist von den Gesellschaftern und Nutzern des DKRZ akzeptiert und unterstreicht die Bedeutung des DKRZ als Höchstleistungsrechenzentrum.

5. Das DKRZ, die DKRZ „Users Group“ und M&D haben ein neues Datenmanagementkonzept entwickelt, das eine flexible, wenn auch kontingentierte, Nutzung des zur Verfügung stehenden Speicherplatzes zum Ziel hat. Langzeitarchivierung ist darin auf Daten beschränkt, die für zukünftige Arbeiten relevant sind.

6. Das WDCC (ICSU World Data Center Climate), getragen von M&D/MPI-M und DKRZ, ist als Langzeitdatenarchiv für Ergebnisse aus Konsortialrechnungen und Wissenschaftsprojekten der Klimaforschung etabliert. Der internationale, interdisziplinäre Zugriff auf diese Daten wird über die Web-Services des WDCC abgewickelt. Das Datenvolumen im Archiv wuchs während der Projektlaufzeit um den Faktor 100, die Datenzugriffe wuchsen um den Faktor 500.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Die Abteilung Modellbetreuung am DKRZ, aus der die Gruppe M&D hervorgegangen ist, hatte bereits neben der Betreuung der Klimamodelle das wissenschaftliche Datenmanagement als weitere Aufgabe aufgegriffen. Zum Zeitpunkt des Übergangs vom DKRZ zum MPI-M wurden im Wesentlichen zwei Aufgabenbereiche bearbeitet:

- a) Die Modellbetreuung konzentrierte sich zu dem Zeitpunkt auf Codepflege und Anwendungsunterstützung von Modellen der Klimaforschung. Dies beinhaltete auch die Optimierung des Modellcodes und die Kopplung von Komponentenmodellen (Ozean, Meereis, Atmosphäre) des Erdsystems.
- b) Im Laufe der 10jährigen Zugehörigkeit zum DKRZ hatte sich allmählich das wissenschaftliche Datenmanagement als zweiter Aufgabenbereich der Gruppe herausgebildet. Zur Bereitstellung von Klimamodelldaten und verwandten Beobachtungen wurde eine erste Implementierung der Klimadatenbank am DKRZ entwickelt und in Betrieb genommen.

Im Laufe der Zeit wuchs die Wahrnehmung und Akzeptanz der DKRZ Abteilung durch die Klimaforschungsgemeinschaft. Es wurden Lücken im Aufgabenspektrum identifiziert, und angeregt, die Aufgaben der Abteilung um neue Anforderungen aus der deutschen Klimaforschungsgemeinschaft zu erweitern. Es gab allerdings zum Zeitpunkt des Übergangs von M&D ans MPI-M keine Regelung, wie an die Gruppe herangetragene Wünsche kanalisiert und integriert werden könnten. Insbesondere wurde die Auswahl der von M&D betreuten Klimamodelle diskutiert.

5. Zusammenarbeit mit anderen Stellen

Die finanzielle Ausstattung von M&D beinhaltete als Sachausstattung Mittel für Arbeitsplatzrechner und einen Abteilungsserver für Entwicklungsarbeiten. Implementierungen für die Klimaforschungsgemeinschaft wurden auf den Compute- und Daten-Servern des DKRZ vorgenommen. Damit war das DKRZ der wesentliche Kooperationspartner im Bereich Höchstleistungsrechnen und Datenspeicherung.

Im wissenschaftlichen Umfeld erstreckte sich die Zusammenarbeit auf alle nationalen Einrichtungen der Klimaforschung.

II. Abschlussbericht M&D (Modelle und Daten)

1. Verwendung der Zuwendung

Der vorangestellte tabellarische, historische Überblick über die Entwicklung der Gruppe in den letzten 20 Jahren dient der Einordnung der im folgenden dargestellten Ergebnisse und strategischen Ausrichtung.

- **1991 – 1999: Abteilung ‚Modellbetreuung‘ am DKRZ**
 - Codepflege und Kopplung von Klimamodellen
 - Nutzerunterstützung
 - Start wissenschaftliches Datenmanagement
- **2000-2009: Gruppe ‚Modelle und Daten‘ (M&D) am MPI-M**
 - 1999/2000: Überleitung der DKRZ-Abteilung in ein BMBF finanziertes Projekt am MPI-M unter Aufsicht des WLA (Wissenschaftlicher Lenkungsausschuss)
 - 2001: Definition des CERA¹⁻² Datenmodells (M&D / AWI / PIK)
 - 2003: Anerkennung der CERA-DB als ICSU WDC Climate (WDCC)
 - 2004/2005: Erweiterung der im EU FP5 PRISM Projekt entwickelten Modellierungsumgebung um Datenaspekte zur ‚Integrierenden Modell und Daten Infrastruktur‘ (IMDI)
 - 2004: Durchführung erster Konsortialrechnungen (IPCC AR4) und deren Regionalisierung mit dem regionalen Klimamodell CCLM
 - 2005: Erweiterung von M&D durch die BMBF geförderte ‚Service Gruppe Anpassung‘ (SGA) zur Unterstützung des Forschungsschwerpunkts ‚Anpassung an Klimaänderungen‘ (zu SGA vgl. parallelen Abschlussbericht)
 - 2006: Neues Archivierungskonzept am DKRZ mit WDC Climate als Langzeitdatenarchiv
- **01.01.2010: Überleitung M&D in die DKRZ-Abteilung ‚Datenmanagement‘**

Die Wandlung des Namens, von ‚Modellbetreuung‘, über ‚Modelle und Daten‘ zu ‚Datenmanagement‘, spiegelt den Wandel der Aufgaben und Arbeitsschwerpunkte über die Jahre wider.

Die Entwicklung von M&D ist in den öffentlich zugänglichen Protokollen des WLA im Einzelnen dokumentiert. Hier sind die Diskussionen, Beschlüsse und strategischen Entscheidungen niedergelegt. Die Protokolle sind öffentlich und finden sich aktuell auf den Web-Seiten von M&D (<http://www.mad.zmaw.de/scient-steering-board/archiv/>).

¹ CERA: Climate and Environmental Retrieval and Archiving

Ziele – Veränderung und Verwirklichung

Ausgangspunkt der Gruppe „Modelle und Daten“ (M&D) im Jahre 2000 war die Abteilung „Modellbetreuung“ am DKRZ. Diese war 1991 gegründet worden als Bindeglied zwischen einerseits wissenschaftlicher Anwendung und andererseits Implementierung und Betrieb auf den Rechnersystemen des DKRZ. Das Ziel war, die Wissenschaftler von Routineaufgaben der Klimamodellierung zu entlasten. Neben der Pflege von ausgewählten Klimamodellen wuchs das wissenschaftliche Datenmanagement allmählich zu einem weiteren Aufgabenbereich der Abteilung heran.

Im Übergang der DKRZ-Abteilung in die zeitlich befristete Projektfinanzierung von M&D durch das BMBF waren die Arbeiten der DKRZ-Abteilung Modellbetreuung anerkannt, aber es konnten viele Wünsche aus der Klimaforschungsgemeinschaft (im Folgenden auch Klima-Community oder nur Community) nicht erfüllt werden. Das lag nicht nur an der begrenzten Anzahl von Personen in der Abteilung, sondern insbesondere auch an fehlenden Verfahren zum transparenten Umgang mit begrenzten Ressourcen. Der Wunsch hingegen war, diesen Service der Modellbetreuung allgemein in der Klima-Community nutzen und gestalten zu können. Diesem Wunsch sollte in der Etablierung der Gruppe Modelle und Daten Rechnung getragen werden. Um die Öffnung von DKRZ und M&D gegenüber der nationalen Klimaforschung zu sichern und um die strategische Ausrichtung von M&D an sich wandelnde Nutzerwünsche über eine Projektlaufzeit von 10 Jahren zu gestalten, wurde der Wissenschaftliche Lenkungsausschuss (WLA) ins Leben gerufen.

Der WLA setzte sich paritätisch zusammen aus vier Mitgliedern, die durch die Gesellschafter des DKRZ und vier Mitgliedern, die durch die Klimaforschungsgemeinschaft (vertreten durch das BMBF² und die DFG³) bestimmt wurden. Erklärtes Ziel des WLA war und blieb es, DKRZ und M&D als nationale Serviceeinrichtungen für die deutsche Klimaforschung zu öffnen bzw. zu etablieren. Der WLA war von Anfang an um Transparenz und Einbindung der Klima-Community in Entscheidungsprozesse bemüht. Instrumente waren die öffentlichen Sitzungsprotokolle des WLA, die Einladung an die DKRZ Users Group als ständiger Gast an den Sitzungen teilzunehmen und die Organisation von öffentlichen Workshops zur Abstimmung und Entscheidungsfindung innerhalb der Klima-Community

In den Diskussionen des WLA wurden M&D und DKRZ nie getrennt betrachtet, sondern als nur administrativ getrennte Teile einer Serviceeinheit betrachtet, und in diesem Sinne wurde auch die strategische Aufgabenteilung aufeinander abgestimmt. Diese Betrachtung von M&D und DKRZ als komplementäre Teile einer Serviceaktivität unterstützte am Ende der Projektlaufzeit die Reintegration von M&D als zusätzliche Abteilung „Datenmanagement“ ins DKRZ.

Der WLA konkretisierte die pauschale Aufgabenstellung von M&D und passte sie im Laufe der Projektlaufzeit an die aktuellen Gegebenheiten und abgestimmten Wünsche aus

² Bundesministerium für Bildung und Forschung

³ Deutsche Forschungsgemeinschaft

der Community an. Im Erstantrag wurde die Aufgaben von M&D gefasst als Bereitstellung von anwendungsbereiten Klimamodellen und Hilfe beim Einsatz solcher Modelle, die Bereitstellung von beobachteten und simulierten Klimadaten zum Zwecke der Klimadiagnostik, und die Durchführung von aufwendigen, in der Klimaforschungsgemeinschaft abgestimmten Rechnungen (Konsortialrechnungen) mit Klimamodellen. Diese Aufgabenstellung war die Fortschreibung der Aktivitäten der Abteilung Modellbetreuung erweitert um den Aspekt Konsortialrechnungen als weitere Serviceaktivität, die motiviert war durch Standardrechnungen mit Klimamodellen für IPCC Sachstandsberichte. Im Laufe der 10-jährigen Projektlaufzeit wurden die Aufgaben zusammen mit der Klima-Community diskutiert, konkretisiert und angepasst. Die Arbeitsschwerpunkte von M&D (und z.T. DKRZ) am Projektende lassen sich in den folgenden Punkten zusammenfassen:

- Die direkte Unterstützung bei Entwicklung und Pflege von Klimamodell-Code wurde verlagert in die Entwicklung und Implementierung einer Infrastruktur zum effizienten Betrieb komplexer Klimamodelle auf den Rechnersystemen des DKRZ. (Im Detail ausgeführt im Abschnitt 1.4)
- Die parallel zur Infrastrukturentwicklung erforderliche Optimierung des Modellcodes selbst für neue Rechnerarchitekturen wurde als Arbeitsschwerpunkt des DKRZ realisiert. (In diesem Bericht nicht ausgeführt, da nicht Aufgabe von M&D)
- Konsortialrechnungen als zentrales Steuerungselement zur Nutzung des DKRZ wurden in der Klima-Community etabliert. (Im Detail ausgeführt in Abschnitt 1.5)
- Wissenschaftliches Datenmanagement zur effizienten Langzeitarchivierung vollständig dokumentierter Klima(modell)daten und deren Integration in internationale Datenföderationen wurde als Alleinstellungsmerkmal von M&D und DKRZ konsolidiert. (Im Detail ausgeführt im Abschnitt 1.6)

Dieses von WLA und Klima-Community entwickelte Aufgabenspektrum wurde zusammen mit M&D ins DKRZ übernommen und fortgeführt. Die transparente Gestaltung von Entscheidungsprozessen und die Einbindung der Klima-Community in diese Entscheidungsprozesse trugen entscheidend zur großen Akzeptanz von M&D und DKRZ in der Klima-Community am Ende der Projektlaufzeit bei.

Zusammenfassung:

Im Laufe seiner zehn jährigen Tätigkeit etablierte bzw. stärkte der WLA die Akzeptanz von M&D und DKRZ als nationale Serviceeinrichtungen der Klimaforschung. M&D wurde mit Ablauf der Projektförderung mit seinem in der Klima-Community abgestimmten Service-Angebot ins DKRZ integriert. Die entwickelten Methoden zur Kommunikation und Integration von Nutzerinteressen sind in der Klima-Community akzeptiert und werden über die Projektlaufzeit hinweg fortgeführt.

In den folgenden Abschnitten sollen exemplarisch Diskussions- und Entscheidungsprozesse dargestellt werden, die in den einzelnen Bereichen hinter den Veränderungen und Umsetzungen der Ziele stehen. Beginnend mit den Entscheidungen im WLA über die Öffnung des DKRZ (nicht Bestandteil der Projektförderung, aber Bestandteil der Arbeit des WLA) hin zu Gemeinschaftsmodellen, Konsortialrechnungen und Klimadatenbank.

Arbeit des Wissenschaftlichen Lenkungsausschusses

Die strategische Ausrichtung von M&D wurde im wissenschaftlichen Lenkungsausschuss (WLA) diskutiert und festgelegt.

Auf der konstituierenden Sitzung des WLA wurde Einigkeit hergestellt, dass das übergreifende Ziel des WLA die Öffnung der beiden Einrichtungen (M&D und DKRZ) für die Klimaforschungsgemeinschaft (im Folgenden auch Klima-Community oder nur Community) und für wissenschaftlich bedeutende, interdisziplinäre Vorhaben sei. Hierzu gehörte auch, die hierfür vorhandenen Einrichtungen und Möglichkeiten bekannt zu machen. Der WLA war von Beginn an um Transparenz gegenüber der Community bemüht. Ein wesentliches Instrument der Kommunikation war und ist die Veröffentlichung der Protokolle der WLA-Sitzungen (<http://www.dkrz.de/about/Organisation/wla>). Weitere Instrumente für Einbindung und Transparenz waren die Einladung an die DKRZ Users Group als ständiger Gast an den WLA Sitzungen teilzunehmen sowie die Ausrichtung von Workshops zu Themen, die breitere Diskussion und Abstimmung innerhalb der Community erforderten.

Die weitere Beratung auf der konstituierenden Sitzung führte zu der Feststellung, dass der WLA grundsätzlich nur Empfehlungen bezüglich der DKRZ-Aktivitäten aussprechen könne. Letztendlich entscheide die Gesellschafterversammlung des DKRZ für den Bereich, der den Gesellschafteranteil betrifft. Dem WLA wurde die Entscheidung für die Verwendung der Ressourcen im BMBF-Anteil übertragen. Im Jahre 2000 war dies die Hälfte der am DKRZ zur Verfügung stehenden Rechenzeit. Eine Kontingentierung von Speicherplatz war zu diesem Zeitpunkt noch nicht erforderlich. Die Gleichverteilung der Ressourcen zwischen den Gesellschaftern des DKRZ und dem BMBF wurde über die Projektlaufzeit nicht verändert.

Zu Beginn waren die zentralen Diskussionspunkte im WLA die von M&D zu unterstützenden numerischen Modelle (Fortführung der Aktivitäten der Abteilung Modellbetreuung am DKRZ), die Verteilung der Rechenzeit aus dem BMBF Anteil und die effektive Nutzung des Klimarechners im Sinne der Community. Wissenschaftliches Datenmanagement gehörte anfangs nicht zu den Themen des WLA.

Mit Blick auf die bevorstehende Inbetriebnahme des HLRE-I (Höchstleistungsrechner für die Erdsystemforschung) am DKRZ in 2002 und den zu erwartenden großen Datenmengen wurden neben der Modellunterstützung als neue Schwerpunkte der Aktivitäten von M&D das wissenschaftliche Datenmanagement und die Entwicklung

einer Modellierungsumgebung zur effizienten Durchführung großer Modellierungsprojekte von allgemeiner Bedeutung diskutiert und beschlossen. Damit wurde die strategische Ausrichtung von M&D an Anforderungen der Klima-Community vorgenommen, die aus dem Betrieb des HLRE-I erwachsen.

Nach Inbetriebnahme des HLRE-I wurde die Frage der Auslastung und der zweckgemäßen Verwendung des Höchstleistungsrechners als ständiges Thema auf den WLA Sitzungen diskutiert. Ein Teil des Höchstleistungsrechners sollte für sehr große Rechenprojekte reserviert werden, die nur auf dieser Klasse von Rechnern durchführbar sind. Als weiteres Kriterium für derartige Rechnungen wurde ein übergreifendes Interesse in der Community festgelegt. Dieser Typ von Rechnungen wurde „Konsortialrechnungen“ genannt im Hinblick darauf, dass ein Forschungskonsortium dahinter stehen sollte. Allerdings waren die Auswahlkriterien noch nicht genau spezifiziert. Es wurde vereinbart, in der Regel bis zu 25% des gesamten Rechners für Konsortialrechnungen zu reservieren. Dieser Ressourcenanteil sollte zu gleichen Teilen den Anteilen der Gesellschafter des DKRZ und dem Anteil des BMBF entnommen werden. Die Definition der Konsortialrechnungen sollte auf Workshops innerhalb der interessierten Community diskutiert und abgestimmt werden. Wichtig an dieser Stelle war der Interessenausgleich zwischen dem Konflikt große „Flaggschiff“-Projekte auf dem HLRE-I durchzuführen aber den Rechner nicht als Werkzeug für den Rest der Klima-Community zu blockieren. Verfahren und Vorgehen zur Entscheidungsfindung, Interessenabwägung und transparenter Darstellung nach außen wurden im weiteren Verlauf des Projektes M&D im WLA entwickelt.

Die Diskussion über Gemeinschaftsmodelle und deren Unterstützung durch M&D wurde auf den WLA Sitzungen kontinuierlich geführt, ohne allerdings zu einem akzeptierten Ergebnis zu kommen. Die Abbildung der Wünsche aus Community und WLA auf die begrenzten personellen Ressourcen bei M&D erwies sich als schwierig. Anfang 2003 wurde vom MPI-M eine Trennung von Modellentwicklung und den Aufgaben von M&D vorgeschlagen. Die Aufgabe von M&D sollte sich danach auf die Kopplung, Portierung und Unterstützung der Nutzer bei der Anwendung von Klimamodellen konzentrieren. Dieser Vorschlag des MPI-M war motiviert durch eine Erfolg versprechende Entwicklung einer Modellierungsumgebung in einem EU-Projekt, den stagnierenden Einigungsprozess im Bereich Gemeinschaftsmodelle und Unterstützung durch M&D, die uneinheitliche Meinung im WLA zum Thema Betreuung von Gemeinschaftsmodellen und nicht zuletzt durch die begrenzten Personalressourcen bei M&D, die direkte, individuelle Modellunterstützung in der gewünschten Breite und Vielzahl nicht zuließen.

Die Einordnung der Klimaszenarien-Rechnungen für den 4. Sachstandsbericht des IPCC (IPCC-AR4) und deren Regionalisierung als Konsortialrechnungen war im WLA unumstritten. Diskussionen entzündeten sich allerdings an der Frage, was eine Konsortialrechnung im Allgemeinen sei und wie viel Rechenzeit dafür reserviert werden solle.

Für die IPCC-AR4 Konsortialrechnungen wurde letztlich der Beschluss gefasst, bis zu 1/3 des dem BMBF zustehenden Rechenzeitanteils für die IPCC Konsortialrechnungen zunächst für das Jahr 2004 zur Verfügung zu stellen. Gleichzeitig

werden die Gesellschafter gebeten, aus ihrem Anteil ein gleich großes Kontingent an Rechenzeit für die IPCC-Konsortialrechnungen für den AR4 bereitzustellen. Die Rechnungen wurden wie beschlossen und gewünscht durchgeführt. Die IPCC-AR4 Rechnungen das erste vom WLA begleitete Konsortialrechnungsprojekt. Zu diesem Zeitpunkt gab es noch kein transparentes, allgemein bekanntes Antragsverfahren für neue Konsortialrechnungen.

Zur Vorbereitung der Nachfolgebeschaffung für den HLRE-I am DKRZ und zur Klärung der Zukunft von M&D wurde auf Wunsch des BMBF eine Begutachtung von DKRZ und M&D am 07. Juli 2004 in Hamburg durchgeführt. Das positive Ergebnis bestätigte den eingeschlagenen Weg von M&D und die Arbeit des WLA. Neben der Zusammenführung von M&D und DKRZ und der Institutionalisierung des DKRZ wurde für M&D angeregt sich weiterhin im Aufgabenfeld Entwicklung einer Infrastruktur für Erdsystemmodellierung zu engagieren. Die Entwicklung der Modellkomponenten selbst sollte bei den wissenschaftlichen Einrichtungen verbleiben. M&D sollte für die Integration der Modelle in die Modellierungsumgebung zuständig sein

In 2005 traten erstmals Engpässe auf dem HLRE-I auf. Die Rechenleistung reichte nicht mehr für alle Anträge (Überbuchung bis zu 100%), ebenso wenig entsprachen die Kapazitäten zur Datenspeicherung den Anforderungen. Transparenz der Rechnernutzung und Akzeptanz der Aufteilung der Ressourcen war eine der Forderungen aus der Begutachtung von DKRZ und M&D. Es wurde eine Strategie zur Kürzung von Rechenzeitanträgen ausgearbeitet und kommuniziert.

Anfang 2005 begannen die Planungen zur Ersetzung des HLRE-I. Eine Neubeschaffung war geplant für 2007 mit einer Erhöhung der Rechenleistung um den Faktor 20-40. Im Hinblick auf den nächsten Rechner sollten die Kosten für den zukünftigen Datenanfall stärker berücksichtigt werden. Erste Überlegungen zur Änderung des Archivierungskonzepts am DKRZ wurden mit der DKRZ ‚Users Group‘ diskutiert. Ein neues Konzept wurde von der DKRZ ‚Users Group‘ in Abstimmung mit M&D und DKRZ im WLA vorgestellt. Der WLA zusammen mit ‚Users Group‘, M&D und DKRZ entwickelten den Ansatz weiter zu einem nachhaltigen Konzept zur Datenhaltung am DKRZ. Dieses Konzept wurde mit Inbetriebnahme des HLRE-II umgesetzt.

Gegen Mitte der Projektlaufzeit, als die Diskussionen um das M&D Aufgabenspektrum einen stabilen Zustand im WLA und in der Klima-Community erreicht hatten, wurde von M&D ein Leistungskatalog vorgelegt, der die aktuellen Dienste der Gruppe beschreibt (<http://www.mad.zmaw.de/about-us/leistungskatalog/>). Schwerpunkte des Leistungskatalogs waren die Weiterentwicklung und Nutzung der Modellierungsumgebung, die Durchführung von Konsortialrechnungen und Entwicklung der Dateninfrastruktur im Rahmen der Klimadatenbank. Der WLA begrüßte den Leistungskatalog. Er wurde im Sinne der Transparenz auf dem M&D Webserver veröffentlicht und behielt seine Gültigkeit bis zum Ende des Projektes.

Der WLA beteiligte sich aktiv an der Zusammenführung der Community Interessen bei der Definition des HLRE-II. Vor Inbetriebnahme wies er als neue strategische

Komponente auf die Notwendigkeit verstärkter personelle Unterstützung bei der Nutzung des neuen Systems hin. Im Hinblick auf die bevorstehende Beschaffung des HLRE-II als massiv paralleles System wurde vom WLA die folgende Stellungnahme abgegeben:

„Der WLA hält eine Stärkung der Unterstützung und Ausbildung in Hinblick auf die effiziente Nutzung moderner Hochleistungsrechner im Bereich der Erdsystemmodellierung für dringend geboten.

Dies umfasst insbesondere

- eine personelle Verstärkung der fachspezifischen Nutzerunterstützung durch die Abteilung Anwendungssoftware des DKRZ,
- ein deutlich erweitertes Schulungsangebot im Bereich moderner Programmiermethoden, insbesondere der Parallelisierung, unter besonderer Berücksichtigung der Bedürfnisse der Nutzer des DKRZ,
- eine Verankerung entsprechender Ausbildung und Forschung, auch hinsichtlich fortschrittlicher Methoden der numerischen Erdsystemmodellierung, an den Universitäten, insbesondere durch zügige Besetzung des Lehrstuhls "Wissenschaftliches Rechnen in den Geowissenschaften" an der Fakultät für Mathematik, Informatik, und Naturwissenschaften der Universität Hamburg.“

Dieser Beschluss stand im Einklang mit den Forderungen, die von der DKRZ „Users Group“ vorgetragen wurden. Die hier formulierte Ausrichtung auf verstärkte personelle Unterstützung im Gleichklang mit der Installation neuer Rechnerarchitekturen besitzt nach wie vor Gültigkeit. Moderne Rechensysteme besitzen einen Komplexitätsgrad, dass spezielle Kenntnisse erforderlich sind, um sie effektiv nutzen zu können. Diese Entwicklung begann mit der Inbetriebnahme des HLRE-II und ist nicht abgeschlossen.

Zusammenfassung:

Das Gremium WLA koordinierte erfolgreich Community-Interessen und integrierte sie in die nationalen Serviceeinrichtungen DKRZ und M&D. Der WLA ist in der Klima-Community als Institution anerkannt.

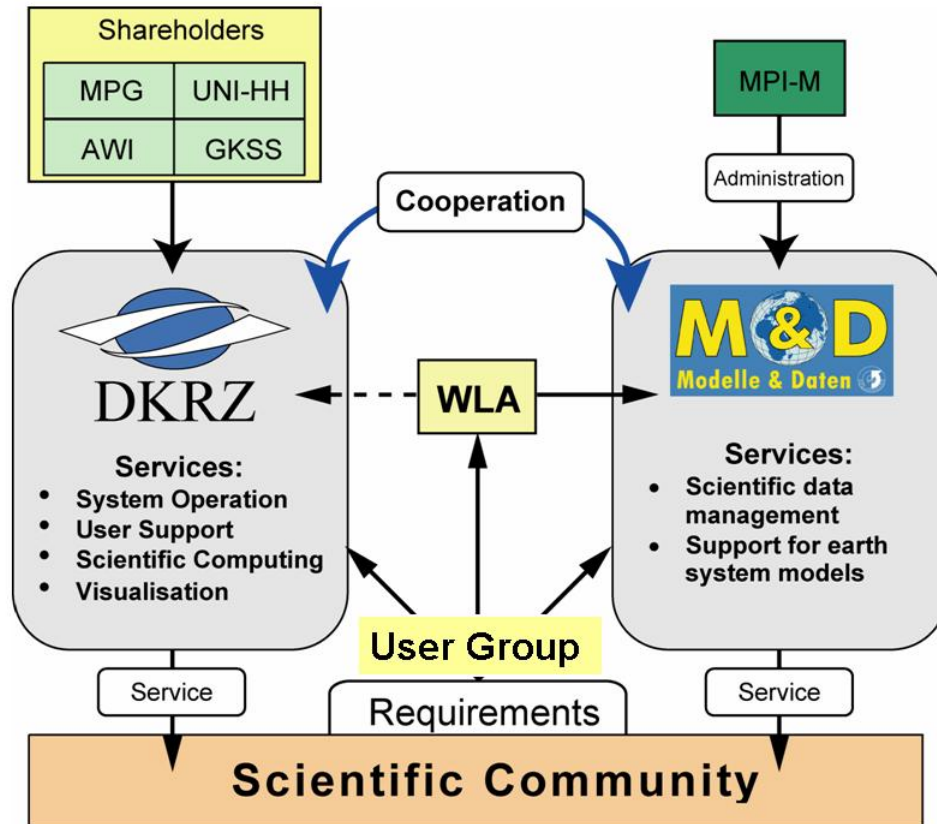
Öffnung des Deutschen Klimarechenzentrums für die Klimaforschungsgemeinschaft

Neben der Integration der Anforderungen aus der Klima-Community in die strategische Ausrichtung von M&D fungierte der WLA als Beratungsgremium für das DKRZ, um eine stärkere Öffnung des DKRZ für die nationale Klimaforschung zu erreichen.

Die Gründung des WLA erfolgte parallel zur bereits angelaufenen Beschaffung des HLRE-I am DKRZ. Der WLA begleitete die Beschaffung und drängte im Hinblick auf den damals für 2007 anstehenden 4. Sachstandsbericht des IPCC auf zügige Abwicklung der Beschaffung und Installation. Anfang 2002 wurde der HLRE-I (NEC-SX6 mit 28 TB Gesamtplattenplatz) installiert. Damit stand dieser Höchstleistungsrechner für die Arbeiten zum IPCC-AR4 zur Verfügung. Sie wurden als Konsortialrechnungen im Auftrag und als Service für die deutsche Klimaforschungsgemeinschaft durchgeführt.

Ein zweiter von der deutschen Klimaforschungsgemeinschaft als wichtig erachteter Service des DKRZ ist die Archivierung und Dokumentation der Ergebnisse von wichtigen Modellrechnungen, sowie ihre zeitnahe Verfügbarkeit für die Klimaforscher. Diese Anforderung wurde mit der Entwicklung einer Lösung für ein automatisches Füllen der Klimadatenbank am DKRZ umgesetzt. Mit dem automatischen Füllen kann eine deutlich größere Datenfüllrate erreicht werden, als mit dem bis dahin manuell kontrollierten Prozess, so dass trotz gesteigerter Rechenleistung und der damit verbundenen Steigerung der Datenproduktion eine zeitnahe Archivierung und Verteilung der Modellergebnisse für die Nutzer des HLRE-I realisiert werden konnte.

Während bei den Aufgabenschwerpunkten im Modellbereich von M&D eine Verlagerung zur Entwicklung einer Modellierungsumgebung zur effektiven Durchführung von Klimamodellrechnungen auf dem HLRE-I stattfand, gewann der Bereich wissenschaftliches Rechnen am DKRZ stärker an Bedeutung. Die Entwicklung der Klimamodelle wurde bei den wissenschaftlichen Instituten belassen, aber die technische Anpassung des Modell-Codes auf die NEC-SX6 Architektur (Parallelisierung, Optimierung) zur Effizienzsteigerung wurde vom DKRZ übernommen.

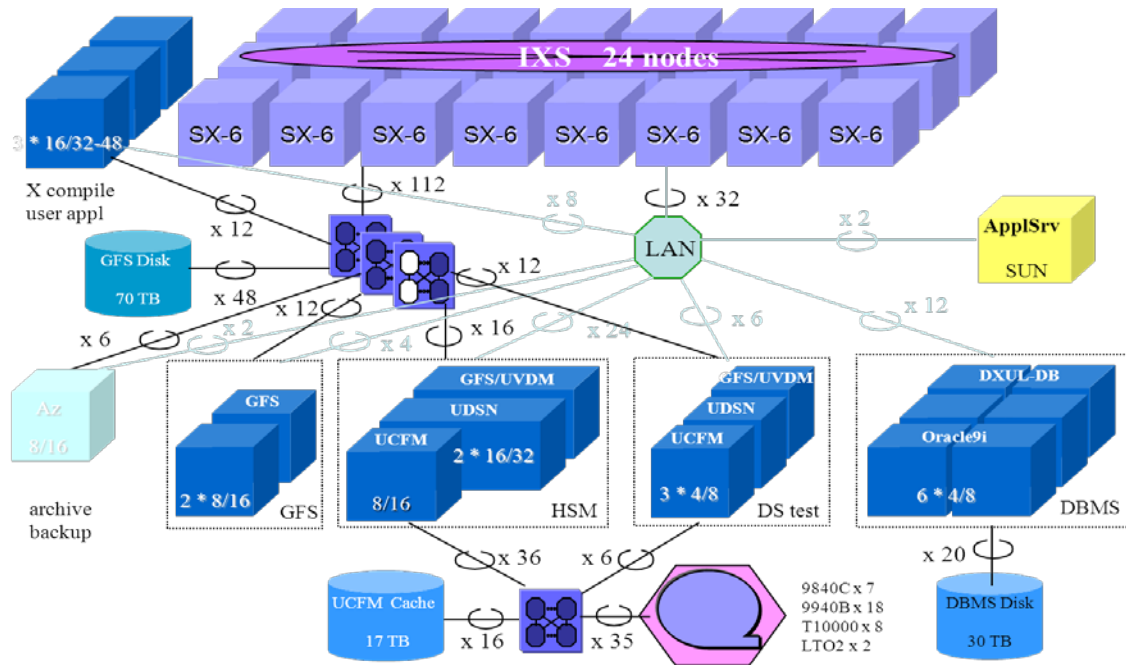


Die **Abbildung** zeigt das Kooperationsdiagramm von DKRZ und M&D. In dem Punkt „Support for earth system models“ bei M&D sind die Modellierungsumgebung und die Konsortialrechnungen enthalten. Im „Scientific Computing“ am DKRZ ist die Optimierung der Modell-Codes eingeordnet.

Mit der Inbetriebnahme des HLRE-I wurde auch das Vorgehen bei der Vergabe von Rechenzeit diskutiert. Dem WLA als Vertreter der Interessen der deutschen Klimaforschungsgemeinschaft am DKRZ war an einer bestimmungsgemäßen und ausgewogenen Nutzung des Rechners gelegen. Darum enthält ein Projektantrag auf Rechenzeit aus dem BMBF-Kontingent neben dem benötigten DKRZ-Ressourcenbedarf auch Informationen, die die wissenschaftliche und technische Beurteilung des Antrags ermöglichen. Dazu gehören eine Kurzbeschreibung des Projektes, die Kompetenz des Antragstellers, die Angabe der benötigten Ressourcen und eine Liste der projektrelevanten Veröffentlichungen des Antragstellers. Die Eignung des verwendeten Programms für den HLRE-I muss nachgewiesen werden. Damit wurde eine Objektivierung und transparentere Gestaltung der Rechenzeitvergabe (Antragwesen und Begutachtung) unterstützt.

Nutzern wie Betreibern des HLRE-I war eine möglichst hohe Auslastung des HLRE-I wichtig. Deshalb war bei der vom BMBF in Auftrag gegebenen Begutachtung von M&D und DKRZ am 7. Juli 2004 angeregt worden, dass der WLA neben den BMBF Ressourcen auch die Nutzung der Ressourcen durch die Gesellschafter beobachten solle.

Die Nutzung des HLRE-I soll auf Aufgaben beschränkt werden, die eine effektive Auslastung der vorhandenen Infrastruktur ermöglichen. Kleine Jobs sollen auf ein (Linux-)Cluster verlagert werden. Über die Standzeit des HLRE-I betrug die Auslastung meist 95% und darüber, nie aber unter 90%. Als ein wichtigstes Instrument zur Steuerung der Auslastung erwiesen sich die Konsortialrechnungen.



Die **Abbildung** zeigt den HLRE-I (NEC-SX6) am DKRZ in seinem Vollausbau in 2007. Abgebildet sind der Compute-Server (SX-6), der zentrale Plattenplatz (GFS Disk, 70 TB) und, in der unteren Reihe, die Datenserver (HSM für das Bandarchiv und DBMS für den Datenbankserver) mit ihren Cache-Platten. Der Würfel „ApplSrv“ bezeichnet den Applikationsserver der Datenbankumgebung, der den Datenbankzugriff über Internet und Web-Browser realisierte.

Ab 2005 zeichneten sich Engpässe in der Speicherkapazität des DKRZ ab. Die im Betriebskonzept des HLRE-I vorgesehene doppelte Datenhaltung musste mit Blick auf die Betriebskosten stark eingeschränkt werden. Im Hinblick auf einen HLRE-II mit deutlich mehr Rechenleistung und Datenproduktion wurde ein neues Datenhaltungskonzept am DKRZ erforderlich. Ein Vorschlag hierzu wurde dem WLA Anfang 2006 von DKRZ ‚Users Group‘, DKRZ und M&D vorgelegt. Der Vorschlag wurde diskutiert und befürwortet. Das Konzept sieht als zentralen Punkt vor, dass nur noch ein kleiner Teil der auf dem HLRE erzeugten Daten in die Langzeitarchivierung überführt wird, alle anderen Daten werden im Kontext der DKRZ-Rechenzeitprojekte gespeichert, mit einem Zeitstempel versehen und nach Ablauf der DKRZ-Projekte gelöscht. Überführung in die Langzeitarchivierung verlangt eine wissenschaftliche Entscheidung für die Speicherung und die vollständige Dokumentation im CERA-

Datenkatalog des WDCC. Das Konzept wird in Abschnitt 1.6.7 im Rahmen der Datenservices erläutert. Mit der Umsetzung wurde noch während der Projektlaufzeit von M&D begonnen.

Die Vorbereitungen für die Beschaffung des HLRE-II als Nachfolgesystem des HLRE-I wurden in kontinuierlicher Abstimmung mit dem WLA und der DKRZ „Users Group“ vom DKRZ durchgeführt. Es gab keine Kontroversen und der WLA unterstützte die Bemühungen des DKRZ während der gesamten Beschaffung. Es wurde die Entscheidung getroffen Compute- und Daten-Server getrennt voneinander auszuschreiben, um für den Datenservice eine flexible, modulare Lösung zu finden, die möglichst über die aktuelle Beschaffung hinaus tragfähig bleibt.

Ende 2007 wurde der Kaufvertrag für den HLRE-II zwischen IBM und DKRZ unterzeichnet. Die technischen Eckdaten des neuen Rechnersystems sind: 240 Rechenknoten mit je 16 Dual-Core CPUs vom Typ IBM Power 6, insgesamt also 7680 Rechenkerne. Jeder dieser Kerne hat eine Spitzenleistung von knapp 19 GFlop/s; damit erreicht das System insgesamt eine Spitzenleistung von mehr als 140 TFlops/s. Aufgrund der von IBM durchgeführten Anwendungsbenchmarks erwartete das DKRZ eine Steigerung des Gesamtdurchsatzes gegenüber dem NEC Rechner um das 50- bis 60-fache. Diese Steigerung wurde hauptsächlich dadurch erreicht, dass das neue System über 40-mal so viele Rechenkerne verfügt wie der HLRE-I. Im September 2008 wurde über die HLRE-II-Komponenten Netzwerk, Archiv-Hardware, Stromversorgung und HSM-System entschieden.

Zusammenfassung:

Über die Projektlaufzeit von „Modelle und Daten“ wurde die Öffnung des DKRZ für die Community umgesetzt. Ein transparentes und mit der Klima-Community abgestimmtes Verfahren zur Vergabe der Ressourcen des HLRE am DKRZ wurde entwickelt und implementiert. Es wurde dabei parallel auf die Optimierung der Auslastung des HLRE geachtet. Im Laufe der Projektlaufzeit hatte sich eine Auslastung von 95% als Standard etabliert.

Gemeinschaftsmodelle und Modellierungsumgebung

Seit der Gründung der Abteilung ‚Modellbetreuung‘ am DKRZ, und danach für M&D, gehörte die Unterstützung der Modellanwender auf den Rechnern des DKRZ zu den Hauptaufgaben der Gruppe. In diesen 20 Jahren stieg die Komplexität der Anwendungen enorm an. Dies lässt sich nicht nur am Umfang der Modell Quellcodes ablesen, sei es an der Anzahl der Zeilen, der Unterprogramme, oder der eingebundenen Bibliotheken, sondern auch die dargestellten Klimasysteme haben durch die Einbindung von immer mehr Teilsystemen an Realitätsnähe gewonnen. Während in den 90er Jahren das Hauptgewicht bei den Modell Anwendungen auf Modellen der allgemeinen atmosphärischen oder ozeanischen Zirkulation (als ‚standalone‘ Modelle oder gekoppelt als Klimamodelle) lag, rückte mit der Zeit die Anwendung von Erdsystemmodellen, die auch Darstellungen der Landoberfläche, mariner bio-geo-chemischer Kreisläufe, den hydrologischen Kreislauf, und die Chemie der Atmosphäre in steigender Komplexität umfassten, in den Vordergrund. Auch die in den Teilsystemen dargestellten Prozesse, wie zum Beispiel Bildung von Wolken oder Meereis, entwickelten sich von einfachen Parametrisierungen zu prognostischen Berechnungen basierend auf den zugrundeliegenden physikalischen Gesetzen.

Nicht zuletzt verlangt die Entwicklung der Rechnerarchitekturen zunehmende Informatik-Expertise von Modell Anwendern und -entwicklern. Die hohe Rechenzeitanforderung der Klimamodelle verlangt, dass sie auf möglichst vielen Rechnerarchitekturen, wie einfachen Arbeitsplatzrechnern, Clustern, Vektormaschinen, oder massiv parallelen Architekturen effektiv einsatzfähig sind. Nur so können durch Ausnutzung aller zur Verfügung stehenden Systeme Modellentwicklung und Durchführung von langen Simulationen mit möglichst hoher Auflösung des Rechengitters in akzeptabler Zeit zum Abschluss gebracht werden. Zeitdruck entsteht dabei durch die immer mehr an Bedeutung gewinnenden internationalen Projekte wie z.B. die Erstellung der Zustandsberichte des IPCC und internationale Modellvergleiche.


Ein weiterer Aspekt, der die Anforderungen an die Gruppe ‚M&D‘ steigen ließ, war die Zunahme der Diversität der allein in der deutschen Klimaforschungsgemeinschaft verwendeten Modelle. So gab es zum Beispiel nicht nur mehrere in Deutschland, sondern auch mehrere außerhalb Deutschlands entwickelte Ozeanmodelle, die intensiv genutzt wurden. Diese zunehmende Modellvielfalt war eine Herausforderung, der sich M&D und der WLA stellen musste im Hinblick auf Verteilung begrenzter Ressourcen.

Es wurde gesehen, dass eine Beschränkung der von M&D zu betreuenden Modelle stattfinden musste, und zwar mit für die Nutzergemeinde möglichst transparenten Kriterien. Die ausgewählten Modelle sollten ‚Gemeinschaftsmodelle‘ sein, in dem Sinne, dass sowohl die Modelle als auch die Nutzerunterstützung durch M&D der ganzen deutschen Forschungsgemeinschaft zur Verfügung steht.

Zur Frage der Gemeinschaftsmodelle, die durch M&D unterstützt werden sollten, wurden von M&D und WLA am 09. und 10. Januar 2001 in Bremerhaven ein offener Workshop ausgerichtet. Zentrale Frage an die Communities der Erdsystemforschung war, ob es

bereits Modelle gibt, die de facto als Gemeinschaftsmodelle angesehen werden können, oder ob sich die Entwicklung solcher Modelle abzeichne. Eingeladen waren alle bekannten Einrichtungen, die sich mit Klimaforschung und insbesondere mit numerischer Modellierung befassen. Es kamen mehr als 100 Teilnehmer zum Workshop.

Organisatorisch wurden die numerischen Modelle entsprechend der verschiedenen Klimateilsysteme, die sie darstellten, in entsprechenden Klassen zugeordnet (siehe Abbildung). Für jede Modellklasse gab es im Plenum einführende Übersichtsvorträge von Experten und nachfolgend, in teilweise parallelen Sitzungen, Diskussionen über die Rolle der einzelnen Komponenten in der Klimaforschung.



Das Gemeinschaftsmodell für die Deutsche Klimaforschung
1. Workshop: 9./10. Januar 2001 in Bremerhaven

Im Rahmen der Breakout-Sessions des Workshops wurden zu den einzelnen Themenbereichen zahlreiche Kurzvorträge gehalten, die verschiedene Aspekte der Komponenten eines Gemeinschaftsmodells behandelten und bereits existierende Modelle dazu vorstellten.

Diese Seite stellt eine **Liste von Links** zusammen, die zu den einzelnen Themen relevant sind. Die Verweise basieren auf den bisher eingegangenen Vorschlägen der Vortragenden in den entsprechenden Sessions.

Session 1: <u>Regional</u> Atmosphere Models	Session 2: <u>Global</u> Atmosphere Models
Session 3: <u>Global</u> Atmosphere Chemistry Models	Session 4: <u>Bio-Geo-Chemistry Models,</u> <u>ECO-System Models and</u> <u>Land Surface Models</u>
Session 5: <u>Regional</u> Ocean Models	Session 6: <u>Global</u> Ocean Models
Session 7: <u>Coupled Models –</u> <u>Model Coupling</u>	

Die Ergebnisse dieser Diskussionen wurden im Plenum zusammengeführt.

Session 1) Regionale Atmosphärenmodelle

Die Diskussion darüber, ob regionale Atmosphärenmodelle existierten, die als Gemeinschaftsmodell in Frage kommen könnten, konnte während des Workshops nicht abgeschlossen werden. Es war aber eine gewisse Konvergenz in den Meinungen zu erkennen. Ein Angebot des DWD (Deutscher Wetterdienst) lag vor, sein Regionalmodell LM (Lokalmodell) als Gemeinschaftsmodell zur Verfügung zu stellen und auch zukünftig zu unterstützen. Nach ausführlicher Diskussion wurde das LM als potentiell Gemeinschaftsmodell anerkannt, und M&D wurde mit der Sammlung von Informationen zur Beurteilung des LM beauftragt. Die GKSS machte eine Zusage zur langfristigen Unterstützung des Modells, sollte es den Status eines Gemeinschaftsmodells erhalten. Nachfolgend wurden sowohl das LM als auch das am MPI-M entwickelte regionale Atmosphärenmodell REMO als Gemeinschaftsmodelle betrachtet.

Einige Jahre später, auf einer WLA Sitzung im Januar 2005, wurde nach kurzer Diskussion das CCLM (COSMO in Climate Mode) als Gemeinschaftsmodell zur Modellierung des regionalen Klimas der Atmosphäre angenommen. Es handelt sich bei dem CCLM um eine Modellvariante des Lokalmodells (LM) des DWD, die speziell für Klimastudien konfiguriert ist. LM und CCLM werden seitdem vom DWD und der CCLM Community in enger Zusammenarbeit weiterentwickelt, und von M&D unterstützt. Mit dem CCLM wurden später von M&D Regionalisierungsexperimente zu den globalen Rechnungen für den 4. Sachstandsbericht des IPCC durchgeführt (s.u.).

Session 2) Globale Atmosphärenmodelle

Eindeutig war das Ergebnis der Diskussion auf dem Workshop in Bezug auf potentielle globale Atmosphärenmodelle. Sämtliche Entwicklungsstufen von ECHAM, ein Modell des Max-Planck-Instituts für Meteorologie (MPI-M) in Hamburg, wurden sowohl in Deutschland als auch international von einer großen Forschungsgemeinschaft benutzt. In Deutschland gab und gibt es kein Modell, das auch als Gemeinschaftsmodell in Frage käme. Zum Zeitpunkt des Workshops war ECHAM5 die aktuelle Version. Langfristig war am MPI-M aber der Übergang zu einem Gitterpunktmodell vorgesehen. Hierfür kam das vom DWD betriebene GME-Vorhersagemodell in Frage; andere Modelle sollten aber vor einer Entscheidung betrachtet werden. Das MPI-M wurde um eine Zusage zur Unterstützung von ECHAM als Gemeinschaftsmodell und um ein Langfristkonzept zur Weiterentwicklung gebeten. Die Einigung auf ECHAM als Gemeinschaftsmodell war einhellig.

Session 3) Globale Modelle der atmosphärischen Chemie

Im Januar 2001 existierten zwei globale Modelle der atmosphärischen Chemie, die in der deutschen Klimaforschungsgemeinschaft von übergeordneter Bedeutung waren. Dabei handelte es sich zum einen um das am MPI-M favorisierte Modell ‚MOZART‘, das am NCAR (Boulder, USA) entwickelt wurde, zum anderen um CHEM, eine Entwicklung des Max-Planck-Instituts für Chemie in Mainz. Die Entscheidung, ob eines dieser Modelle den Status eines Gemeinschaftsmodells erhalten sollte, konnte über viele Jahre hinweg nicht getroffen werden.

Erst in 2007 wurde die Diskussion über Chemiemodelle wieder aufgenommen. Das am Max-Planck-Institut für Chemie entwickelte Modellsystem EMAC (ECHAM/MESSy Atmospheric Chemistry, Nachfolger von CHEM) wurde als Kandidat für Gemeinschaftsmodelle vorgestellt. Es sollte in die zwischenzeitlich von M&D entwickelte Modellierungsumgebung IMDI (Integrating Model and Data Infrastructure) zur Unterstützung und Verwendung von Gemeinschaftsmodellen eingefügt werden.

Die Integration von EMAC in die IMDI-Umgebung war zum Ende der Projektlaufzeit von M&D technisch abgeschlossen. Aber es blieben Fragen nach Weiterentwicklung und Pflege von EMAC in IMDI offen. Divergierende Modellentwicklungen an den beiden Max-Planck Instituten standen hier einer alle Seiten zufriedenstellenden Lösung im Weg.

Session 4) Modelle der marinen Bio-Geo-Chemie, ozeanischer Ökosysteme, und der Landoberfläche

Im Hinblick auf Modelle der ozeanischen Ökosysteme wurde in Frage gestellt, ob das Konzept eines Gemeinschaftsmodells anwendbar sei. Es wurde von den anwesenden Vertretern dieser Modelle eine große Modellvielfalt befürwortet. Es konnte aber eine Liste verschiedener Modelle aufgestellt werden, für die eine Unterstützung durch M&D begrüßt worden wäre.

Im Hinblick auf globale Modelle der Landoberfläche wurde das LPJ (Lund-PIK-MPI Jena) Modell als Kandidat für ein Gemeinschaftsmodell diskutiert. Als globales Modell mariner bio-geo-chemischer Kreisläufe kam nach Meinung der Workshop-Teilnehmer nur das am MPI-M entwickelte HAMOCC Modell in Frage. Es sollten aber noch Art und Umfang der zukünftigen Unterstützung von HAMOCC als Gemeinschaftsmodell geklärt werden. Nachfolgend wurde dann im WLA beschlossen, dass das LPJ-Modell als Gemeinschaftsmodell der Landbiosphäre behandelt und von M&D betreut werden solle.

Session 5) Regionale Ozeanmodelle

Grundsätzlich können die meisten Ozeanmodelle sowohl als globale als auch als regionale Modelle angewendet werden. Bei Ozeanmodellen die explizit als ‚regional‘ bezeichnet werden, handelt es sich meistens um Modelle, deren Formulierung speziell für bestimmte Regionen optimiert ist. Ebenso wie bei den Modellen der marinen Ökosysteme, bei denen es sich auch um spezialisierte Regionalmodelle handelt, wurde deshalb die Notwendigkeit auf eine Einigung bei den einzelnen Betreibern eher reserviert gesehen. Es wurde keine Empfehlung abgegeben.

Session 6) Globale Ozeanmodelle

In dieser Modellklasse bewarben sich mehrere Modelle um die Anerkennung als Gemeinschaftsmodell. Darunter waren das am Institut für Meereskunde der Universität Kiel betriebene, am GFDL (Princeton, USA) entwickelte Modell ‚MOM‘, das am MPI-M entwickelte ‚HOPE-C‘, und das am IPSL (Paris, Frankreich) entwickelte ‚OPA‘ Modell. Die Unterschiede und Potentiale der drei Modelle wurden kontrovers diskutiert. Eine Empfehlung wurde auf dem Workshop nicht abgegeben. Die Diskussion um den Gemeinschaftsmodell-Status eines globalen Ozeanmodells konnte auch in den folgenden

Jahren vom WLA nicht entschieden werden. Sowohl MOM als auch OPA und HOPE standen weiterhin zur Diskussion.

Session 7) Gekoppelte Modelle – Modellkopplung

Die Komponenten eines Erdsystemmodells (Atmosphäre, Ozean, Landbiosphäre, Stoffkreisläufe, ...) wurden oft an verschiedenen Instituten, oder durch verschiedene Wissenschaftler entwickelt. Die Kopplung solcher unabhängig voneinander programmierter Komponenten bedeutet daher für die Wissenschaftler, die häufig nicht Experten für alle Komponentenmodelle sind, eine große technische Herausforderung, die nicht ihre unmittelbaren Forschungsziele betrifft. Eine Software, die die flexible Kopplung von Komponenten erlaubt, und einfach in der Anwendung ist, wird daher als große Arbeitserleichterung gesehen, die den Wissenschaftlern erlaubt, sich eher auf ihre Forschungsziele zu konzentrieren. Der Bedarf an einem Koppler als Teil einer gemeinsam benutzten Infrastruktur wurde deshalb allgemein anerkannt. Zum Zeitpunkt des Workshops existierten als Alternativen die von der GMD entwickelte Koppler-Bibliothek MPCCI (am AWI installiert), die am CERFACS entwickelte Kopplungssoftware OASIS, oder ein am PIK in der Entwicklung befindlicher Koppler. Es wurde befürwortet, dass M&D zusammen mit der GMD, dem AWI und dem PIK die Möglichkeit einer Zusammenführung der Koppler evaluieren sollte. Das EU-Projekt PRISM (2001 – 2004) stellte letztlich die Weichen in Richtung Weiterentwicklung des am CERFACS in Frankreich entwickelten OASIS-Kopplers. M&D installierte im Folgenden die OASIS-Software auf den Rechnern des DKRZ, baute die Schnittstellen zu OASIS in mehrere Komponentenmodelle ein, darunter die Komponenten des am MPI-M benutzten Erdsystemmodells, testete die gekoppelten Modelle und leistete Anwenderunterstützung.

Die Diskussion über die Definition von Gemeinschaftsmodellen gestaltete sich auch deshalb so schwierig, weil der Umfang der gewünschten Unterstützung bei der Modellanwendung, abhängig von den Forschergruppen oder den speziellen Modellen, sehr unterschiedlich ausfiel. Auch entwickelte sich die gewünschte Unterstützung abhängig von der geleisteten Zuarbeit, die von der in M&D vorhandenen Kompetenz und Erfahrung abhing. Im Verlauf des Jahres 2001 betreute M&D die Gemeinschaftsmodelle ECHAM5 und REMO und zusätzlich sollte die Betreuung der Modelle LPJ für die Landbiosphäre und das LM Modell des DWD erfolgen. Für andere Modelle sollte M&D Unterstützung bei der Kopplung leisten und den Nutzern Modelle und Dokumentationen zugänglich machen. Als vorrangige Aufgabe für M&D wurde die Unterstützung des OASIS-Kopplers gesehen. Die Betreuung der beiden Modelle LPJ und LM durch M&D erwies sich zum Beispiel im weiteren Verlauf als schwierig, da keine engen Kontakte zwischen den Modellentwicklern und M&D bestanden. Bei anderen Modellen, die kompetent unterstützt werden konnten, stiegen die Anforderungen proportional zu der geleisteten Unterstützung. Die Frage nach den durch M&D zu betreuenden Modellen und wie die Betreuung aussehen sollte, begleitete die Gruppe in den nächsten Jahren, auch als Thema auf den WLA Sitzungen. Der WLA beauftragte in 2001 zur Entscheidungsfindung eine Gruppe aus seinen Reihen, einen Kriterienkatalog zu erarbeiten, der bei der Auswahl von Gemeinschaftsmodellen herangezogen werden kann.

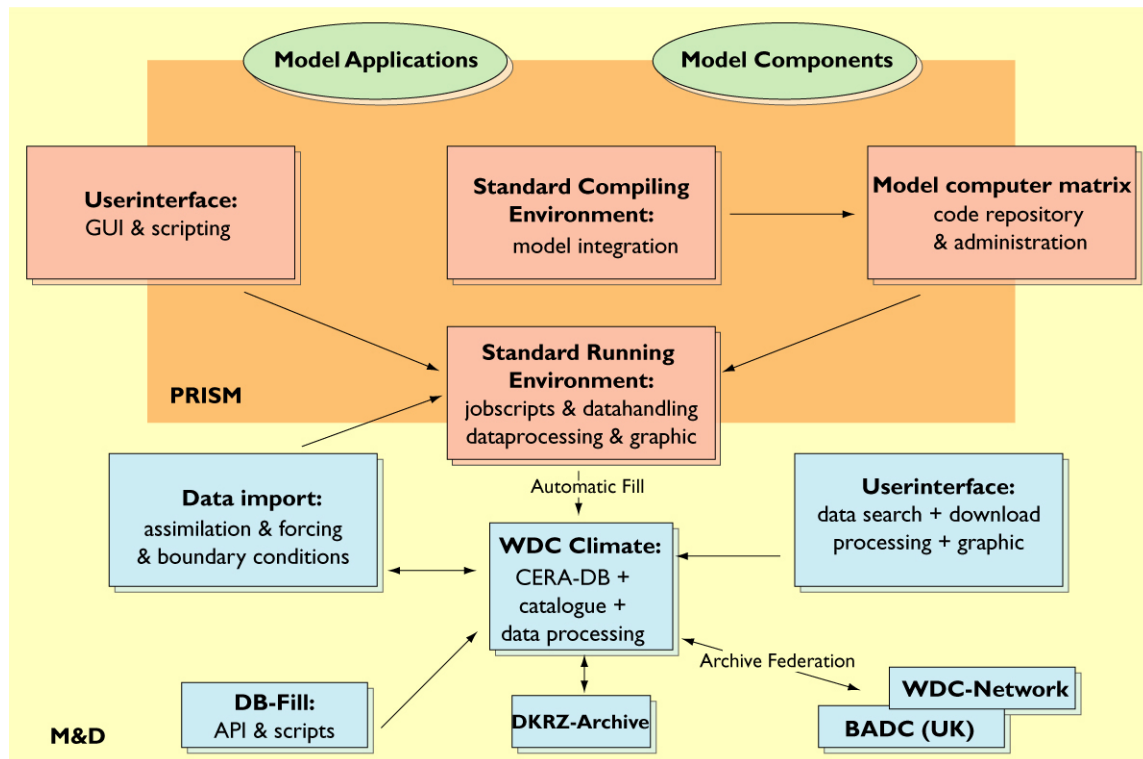
Ein weiterer Workshop "Gemeinschaftsmodelle für die Deutsche Klimaforschung" im Februar 2003 in Hamburg brachte keine größere Klarheit über Kriterien für die Auswahl von Modellen. Auch eine Liste mit präzisen Kriterien für Art und Umfang der Unterstützung von Gemeinschaftsmodellen durch M&D konnte nicht formuliert werden. Auf dem Workshop wurde aber übereinstimmend festgestellt, dass es Aufgabe des WLA sei, hier möglichst bald eine Definition der Kriterien und Prozeduren für Modellunterstützung durch M&D vorzunehmen, und diese der Community bekannt zu machen. Ein weiterer Workshop zu diesem Thema fand während der Projektlaufzeit von M&D nicht statt.

Parallel zur Diskussion um Gemeinschaftsmodelle der deutschen Klimaforschung und deren Betreuung durch M&D startete im Dezember 2001 das ‚Program for Integrated Earth System Modelling‘ (PRISM). Das Projekt wurde von der Europäischen Union im Rahmen des Fifth Framework Programme (EU FP5) mit 5 Mio. Euro für 3 Jahre finanziert. Es vereinigte 17 europäische Klimaforschungszentren und 4 Unternehmen aus der Computerindustrie in dem Vorhaben, eine Infrastruktur für die Entwicklung und Einführung eines europäischen Klimamodellnetzwerks zu schaffen. M&D war u. A. mit der Leitung des Arbeitspakets ‚Model Assemblage‘ beteiligt, in dem eine Modellierungsumgebung für state-of-the-art Erdsystemmodelle entwickelt wurde, mit dem Ziel die Konfiguration, Durchführung, und Analyse von numerischen Experimenten des Erdsystems zu vereinheitlichen und zu vereinfachen. Zu den Ergebnissen des Projekts gehörte die Spezifikation einer Reihe von Standards für Erdsystemforschung. Speziell im ‚Model Assemblage‘ Arbeitspaket wurden die PRISM *Standard Compile Environment* (SCE) und PRISM *Standard running Environment* (SRE) entwickelt, die flexibel genug waren, Experimente mit den Modellen der beteiligten Institute auf mehreren Rechnerarchitekturen aufzusetzen und durchzuführen. M&D beteiligte sich außerdem an der Entwicklung einer graphischen Benutzeroberfläche für die Konfiguration und Überwachung von Experimenten mit numerischen gekoppelten Modellen und eines Systems zur graphischen Darstellung der Ergebnisse.

Die Ergebnisse des PRISM-Projektes waren Ausgangspunkt für eine neuen strategische Ausrichtung im Modellbereich von M&D.. Die Diskussion im WLA bewegte sich weg von der direkten Modellunterstützung hin zur Implementierung einer Modellierungsumgebung am DKRZ, um komplexe, gekoppelte Klimamodelle effizient auf sich verändernden, komplizierte Rechnerarchitekturen ausführen zu können. Parallel dazu gewannen auch Anpassung und Optimierung des Modell-Codes selbst größere Aufmerksamkeit im WLA und wurden als Arbeitsschwerpunkt im DKRZ aufgegriffen.

Als Ergebnis der Diskussion über Modellunterstützung und der Aktivitäten von M&D im PRISM-Projekt wurde ab 2004 die Entwicklung von IMDI (Integrating Model and Data Infrastructure) in Abstimmung mit dem WLA begonnen. Die Modellierungsumgebung IMDI integrierte die Entwicklungen aus PRISM für Compiler- und Laufzeit-Umgebung und erweiterte sie um Datenverarbeitung und Archivierung. Im Rahmen von IMDI wurde das Konzept des automatisierten Füllens der CERA Datenbank und des DKRZ Datenarchivs umgesetzt, um die erwarteten großen Datenmengen aus den Konsortialrechnungen zeitnah bearbeiten zu können. Ziel von IMDI war, die Ressourcen des DKRZ für

Modellrechnungen und Datenbearbeitung optimal zu nutzen und die Ergebnisse der Rechnungen möglichst unmittelbar nach Abschluss der Rechnungen für Qualitätskontrolle und Nutzung zugänglich zu machen.



Die **Abbildung** zeigt IMDI wie sie 2004/2005 zur Integration von Gemeinschaftsmodellen und zur Durchführung der Konsortialrechnungen von M&D implementiert war. Die roten Rechtecke bezeichnen Komponenten, deren Wurzeln im PRISM-Projekt liegen; blaue Rechtecke enthalten Entwicklungen von M&D im Rahmen des wissenschaftlichen Datenmanagements.

Im WLA wurde beschlossen, dass Modellunterstützung und die Durchführung der Konsortialrechnungen nur noch im Rahmen von IMDI stattfinden sollen. Alle für Konsortialrechnungen verwendeten Gemeinschaftsmodelle mussten demnach in IMDI integriert werden. Damit ergab sich mit der Integrierbarkeit in IMDI, d.h. der Einhaltung einiger Standards, auf die sich die PRISM Partner verständigt hatten, automatisch ein Kriterium für potentielle Gemeinschaftsmodelle. Es hatte sich also mit der Entwicklung von IMDI und der Durchführung der Konsortialrechnungen ab Mitte 2004 im Modellbereich von M&D eine Verschiebung des Aufgabenspektrums weg von der direkten Unterstützung von Modellen und hin zu Entwicklung und Betrieb einer Modellierungsumgebung am DKRZ vollzogen.

Zusammenfassung:

WLA und M&D wirkten aktiv in Richtung Entwicklung, Definition und Unterstützung von Gemeinschaftsmodellen für die deutsche Klimaforschung. Ein für die Klima-Community transparente Festlegung eines numerischen Modells als Gemeinschaftsmodell blieb über die gesamte Projektlaufzeit hinweg schwierig. Komplizierter werdende Rechnerarchitekturen und komplexere Klimamodelle erforderten Optimierung der Modellcodes selbst und in der Ausführung der Klimamodelle. Unter der strategischen Leitung des WLA wurden diese Aspekte vom DKRZ (Codeoptimierung) und M&D (Modellierumgebung) aufgegriffen und umgesetzt.

Konsortialrechnungen

Das Konzept der Konsortialrechnungen und der damit beabsichtigten effektiven Nutzung der DKRZ Infrastruktur wurde von der Community akzeptiert und gewann immer breiteren Raum in den Diskussionen im WLA. Eng damit verbunden war und ist das Archivkonzept am DKRZ und Anforderungen, die Ergebnisse aus Konsortialrechnungen schnell, effektiv und qualitätsgeprüft interdisziplinär zur Verfügung zu stellen.

Kriterien für Konsortialrechnungen wurden im Rahmen des WLA über einen längeren Zeitraum diskutiert und entwickelt. Auf der WLA-Sitzung im Januar 2006 wurde schließlich ein Katalog vorgelegt und genehmigt (URL:

<http://www.mad.zmaw.de/service-support/consortium-model-runs/kriterien/>):

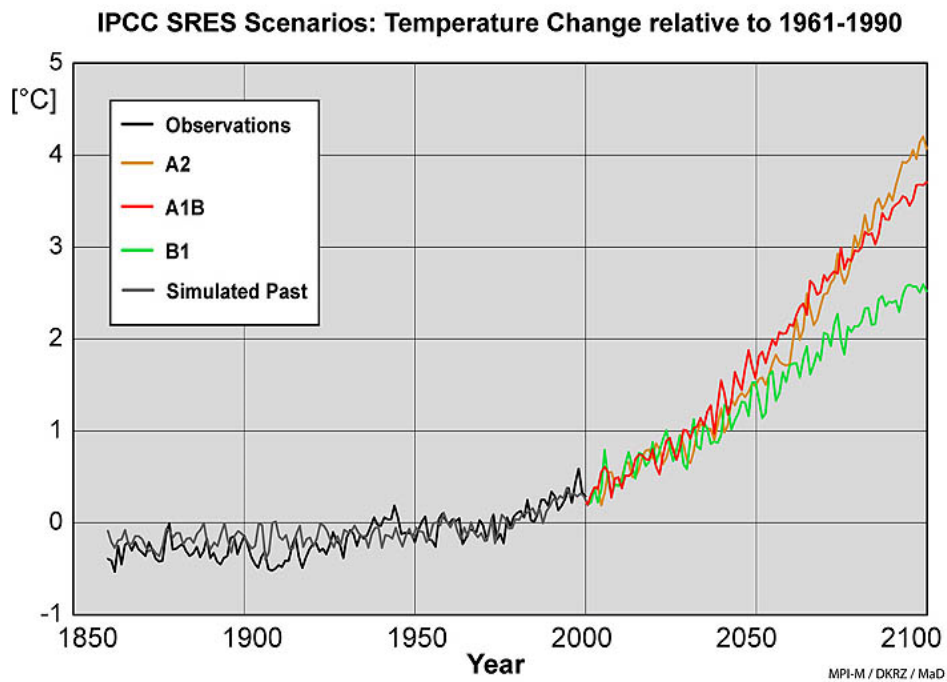
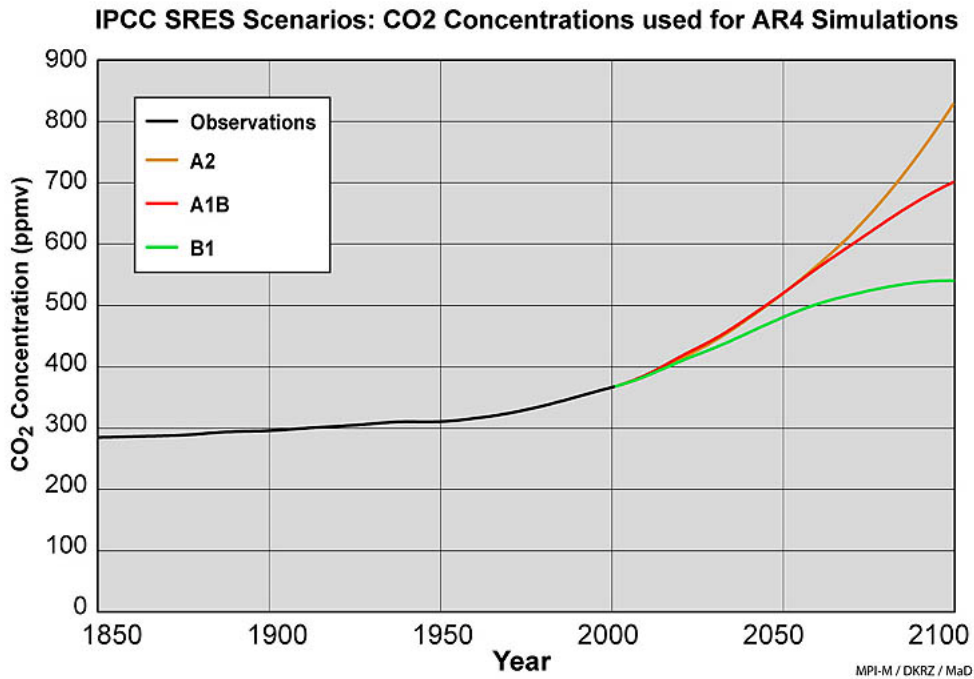
„Als Kriterien für eine Konsortialrechnung sind die folgenden Punkte zu erfüllen:

1. Die Rechnungen müssen Grundlage für umfassende nationale bzw. internationale Forschungsvorhaben sein. Zur Abstimmung der Rechnungen muss Konsens in der jeweiligen Forschergemeinschaft hergestellt werden (z.B. durch einen Workshop).
2. Die Antragsteller müssen dem WLA gegenüber deutlich machen, dass ein "breiter" Konsens für die Rechnung besteht, dass die Rechnungen eine nationale Aufgabe darstellen und der Umfang der Rechnungen die Nutzung des HLRE rechtfertigt.
3. Kriterien an Programme/Modelle:
 - o Die Programme/Modelle müssen in der Regel bei M&D in einer lauffähigen Version vorliegen und getestet sein.
 - o Die Programme/Modelle müssen validiert sein
 - o Die Programme müssen dokumentiert sein
 - o Die Programme müssen für die vorgesehenen Rechner optimiert sein
 - o Es muss ein verantwortlicher Ansprechpartner beim Programmierer benannt sein
4. Der Ressourcenbedarf der Programme muss die Nutzung des HLRE rechtfertigen. Die Rechnungen werden von M&D, DKRZ und dem federführenden Institut in der zur Verfügung stehenden Modell- und Datenstruktur (IMDI) mit den Ressourcen des DKRZ durchgeführt.
5. Die Daten sind allgemein und sofort verfügbar und gemäß den (M&D/WDCC-) Kriterien für die Datenpublikation zu zitieren.
6. Über die Ressourcenzuteilung entscheidet der WLA unter Berücksichtigung der Bedürfnisse des DKRZ-Normalbetriebs
7. Antrag, Durchführung und Ergebnisse werden im WEB dokumentiert“

Die ersten Modellrechnungen, die unter der Bezeichnung Konsortialrechnungen durchgeführt wurden, waren globale Rechnungen zu Projektionen der künftigen Klimaentwicklung wie sie zur Erstellung einer Datenbasis für den 4. Sachstandsbericht des IPCC (IPCC-AR4) vereinbart waren. Diese Rechnungen wurden von 17 Modellierungsgruppen weltweit durchgeführt. Das IPCC Referenzdatenarchiv ist Teil des ICSU World Data Center Climate (WDCC) am DKRZ

(http://www.mad.zmaw.de/IPCC_DDC/html/SRES_AR4/index.html).

Der deutsche Beitrag wurde auf einem vom WLA organisierten Abstimmungstreffen im Mai 2004 in Bonn beim BMBF diskutiert. Zusammenfassend wurde beschlossen, dass die Rechnungen des ECHAM5/MPI-OM Modells in der vorgestellten Weise als Konsortialrechnungen auf dem HLRE-I des DKRZ durchgeführt werden sollten. Dem Vorhaben wurde vom WLA 1/3 des BMBF-Kontingents zugestanden; die Gesellschafter des DKRZ steuerten einen vergleichbaren Anteil bei. Die vereinbarten Rechnungen wurden ab August 2004 bis Sommer 2005 von MPI-M und M&D durchgeführt. Die Daten wurden, wie international vereinbart, im abgestimmten Format zum PCMDI (Programm for Climate Model and Data Intercomparison, Lawrence Livermore, USA) transferiert. Die Atmosphärendaten des ECHAM5 (ca. 115 TByte) stehen außerdem seit April 2005 der internationalen Forschergemeinschaft im WDC-Climate zur Verfügung (<http://www.mad.zmaw.de/service-support/consortium-model-runs/ipcc-experiments/>). In den ersten sechs Monaten des Jahres 2005 verzeichnete das WDCC bereits 51.000 Datenzugriffe. Auf einem Workshop im September 2005 in Hamburg wurden diese Daten der Öffentlichkeit übergeben. Hintergrundinformationen (<http://www.dkrz.de/Klimaforschung/konsortial/ipcc-ar4>) und Dokumentation (<http://www.mpimet.mpg.de/fileadmin/grafik/presse/Klimaprojektionen2006.pdf>) sind über die Web-Server von DKRZ und MPI-M verfügbar.



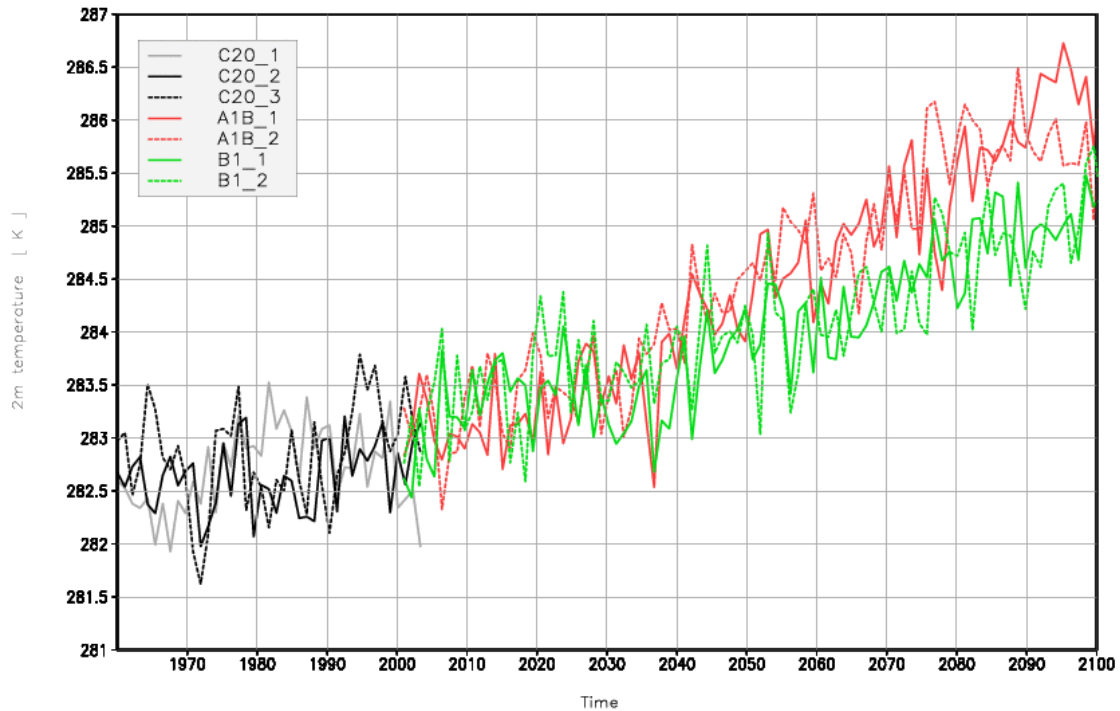
Die **Abbildungen** zeigen die Treibhausgas-Konzentrationen (obere Graphik) und die aus den IPCC-AR4 Konsortialrechnungen mit ECHAM5/MPI-OM resultierende globale Temperaturänderung (untere Graphik).

Der zweite Block von Konsortialrechnungen umfasste die Regionalisierung der globalen Szenarien-Rechnungen, die Grundlage des 4. Sachstandsberichts des IPCC sind. Ein vom WLA organisiertes Abstimmungstreffen dazu fand im November 2004 in Bonn beim BMBF statt. Es war zu diesem Zeitpunkt bereits abzusehen, dass diese Rechnungen keinen Eingang mehr in den IPCC-AR4 finden würden. Die Ergebnisse sollten aber als Datengrundlage für die neuen BMBF Förderschwerpunkte „klimazwei“ (Klimaschutz und Anpassung an die Klimafolgen) und später „KLIMZUG“ (Klimawandel in Regionen) dienen. Auf dem Workshop wurden zwei Schlüsselszenarien für die Berechnungen ausgewählt.

Die folgende Diskussion im WLA ergab dann zwar die Anerkennung als Konsortialrechnung und die Auswahl des CCLM als Gemeinschaftsmodell, aber die Bewertung war nicht so hoch wie für die globalen Rechnungen. Man einigte sich schließlich auf eine Kontingent-Zuweisung von 1/4 des HLRE-I für ein Jahr. Die Rechnungen wurden August 2005 gestartet und in Kooperation mit der BTU Cottbus, der CCLM Community und M&D durchgeführt. Sie wurden Ende 2006 beendet. Die Datenaufbereitung und Restrechnungen außerhalb des Konsortialkontingents dauerten bis Frühjahr 2007.

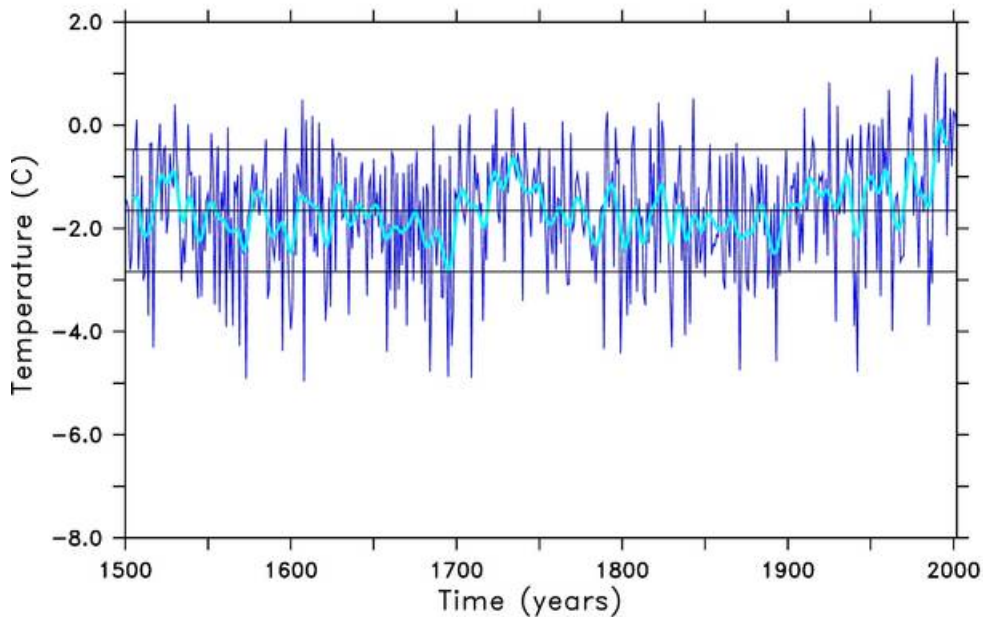
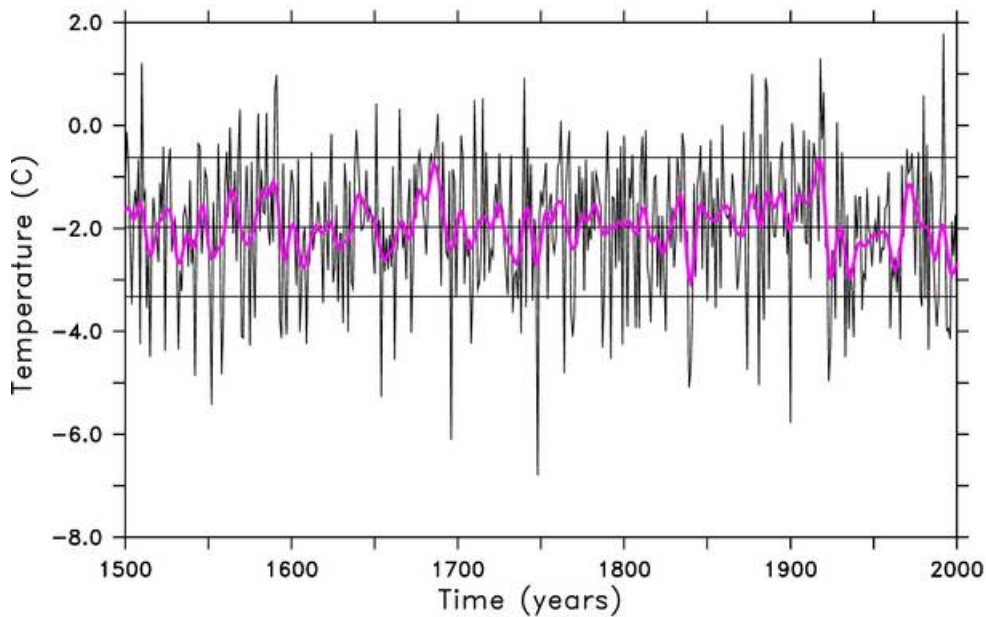
Die geplante interdisziplinäre Nutzung der regionalen Klimamodelldaten stellte neue Anforderungen an Qualitätskontrolle und Datenaufbereitung. Im Gegensatz zu den globalen Klimarechnungen, bei denen die Datenfreigabe wie üblich zunächst für die Modellierungs-Community erfolgte, und erst später nach einer Endkontrolle für die Allgemeinheit zugänglich wurden, musste hier die Endkontrolle unmittelbar nach dem Modelllauf und vor der Freigabe der Daten durchgeführt werden, da diese Daten unmittelbar in die interdisziplinäre Nutzung gehen sollten. Weiterhin war ein zusätzlicher Datenstrom angefragt, in dem die Modellergebnisse auf ein reguläres Gitter interpoliert wurden. Dies stellte neue Anforderungen an das Management der Rechnungen und Datenströme, sowie an die Entwicklung von IMDI als Arbeitsumgebung.

Die Dokumentation der Rechnungen (<http://www.mad.zmaw.de/service-support/consortium-model-runs/clm-experiments/>) und der Ergebnisse (http://www.mad.zmaw.de/fileadmin/extern/documents/reports/MaD_TechRep3_CLM_1_.pdf) findet sich dem Web-Server von M&D. Diese Ergebnisse wurden von der Service Gruppe Anpassung (SGA, <http://www.mad.zmaw.de/projects-at-md/sg-adaptation/>) bei M&D im Laufe der folgenden Jahre in die interdisziplinären Anwendungen vermittelt. Die Nutzer der Daten wurden durch SGA unterstützt. Die Ergebnisse wurden auf dem SGA Workshop am 07.12.2007 in Hamburg (<http://www.mad.zmaw.de/projects-at-md/sg-adaptation/workshops/clm-workshop-2007/>) vorgestellt und am 11.12.2008 nach Abschluss der Qualitätskontrollen auf dem Workshop in Cottbus (<http://www.mad.zmaw.de/projects-at-md/sg-adaptation/workshops/clm-forum-2008/>) als fertiges Produkt der Öffentlichkeit übergeben.



Die **Abbildung** zeigt das Jahresmittel der 2m-Temperatur, berechnet mit CCLM über Europa, von 3 Realisierungen des Gegenwartsclimas und je 2 Realisierungen für die Zukunftsszenarien A1B und B1.

2006 starteten, parallel zu den CCLM Konsortialrechnungen, die Rechnungen des Millennium Konsortialprojekts. Hierbei handelte es sich um Simulationen des Klimas des letzten Jahrtausends unter Verwendung aller zur Verfügung stehenden Beobachtungsdaten, wie Treibhausgase oder anthropogener und vulkanischer Aerosole. Es wurde nur eine grob auflösende Version des ECHAM5/MPI-OM Klimamodells verwendet, allerdings mit aktiver Landoberfläche und Stoffkreisläufen im Ozean zur Berechnung des Kohlenstoffkreislaufs. Dadurch war der Ressourcenbedarf deutlich kleiner als für die CCLM Rechnungen. Insofern gab es keine Ressourcenkonflikte mit den verzögerten CCLM Rechnungen. Die Millennium-Rechnungen dauerten bis Ende 2008 und waren das letzte Konsortialprojekt auf dem HLRE-I. Dokumentation des Millennium-Projekts und Datenzugriff finden sich auf dem Web-Server von M&D (<http://www.mad.zmaw.de/service-support/consortium-model-runs/millennium-experiments/>)



Die **Abbildungen** zeigen europäische Winter (DJF) Oberflächentemperaturen aus dem Millenniumskontrollexperiment (oben) und der Rekonstruktion von Luterbacher (2004). Dünne Linien: jährliche Werte, dicke Linien: 11-jähriges laufendes Mittel. Das Kontrollexperiment berücksichtigt nicht den Anstieg der Treibhausgaskonzentrationen im letzten Jahrhundert, und kann deshalb nicht den beobachteten Temperaturanstieg reproduzieren.

Auf dem neuen Rechner, dem HLRE-II (IBM Power6 Cluster), wurden ab August 2009 zwei neue Konsortialrechnungen initiiert. Das Experiment STORM (High-resolution community climate change simulations) verwendet das Klimamodell ECHAM6/MPIOM in einer Version mit sehr hochauflösendem Ozean. STORM betrachtet insbesondere kleinräumige Phänomene im Ozean. Das zweite Konsortialexperiment sind die globalen und regionalen Klimamodellrechnungen als Beitrag zum nächsten IPCC Assessment Report (IPCC-AR5). Beide Vorhaben erreichten 2009 nicht mehr die operationelle Phase. Bei Storm dauerte die Vorbereitungsphase länger als erwartet und der Zeitplan für die Erstellung des neuen IPCC Report wurde kurzfristig um ein Jahr verschoben.

Zusammenfassung:

Konsortialrechnungen als „Flaggschiff“-Projekte zeigen die Leistungsfähigkeit, aber auch die Leistungsgrenzen des HLRE am DKRZ. Konsortialrechnungen stehen häufig unmittelbar nach ihrer Fertigstellung unterschiedlichsten Klassen von (interdisziplinären) Anwendungen zur Verfügung. Dieser Ansatz stellt neue Herausforderungen an Datenaufbereitung und Qualitätskontrolle. Datenaufbereitung und Erstellung von Datenprodukten müssen synchron zur Modellrechnung durchgeführt werden. Vor der Datenfreigabe sind Dokumentation, geeignete Speicherung und Qualitätsprüfung der Ergebnisse erforderlich. Insbesondere die Qualitätsprüfung großer Mengen von Modelldaten wurde von M&D entwickelt und im Rahmen der CCLM-Konsortialrechnungen erprobt.

CERA Klimadatenbank und ICSU World Data Center Climate

1.1.1. Ausgangssituation Ende der 90er Jahre

Die Entwicklung der Klimadatenbank begann bereits in der Abteilung Modellbetreuung am DKRZ im Rahmen eines BMBF-geförderten Projekts. Hier wurde auch die erste Version des CERA-Datenmodells (Climate and Environmental Data Retrieval and Archiving) entwickelt und in einem relationalen Datenbankmanagementsystem (RDBMS) implementiert. Durch die Anbindung an dieses professionelle System konnten die Inhalte schon bald als Metadatenkataloge über das Internet verfügbar gemacht werden.

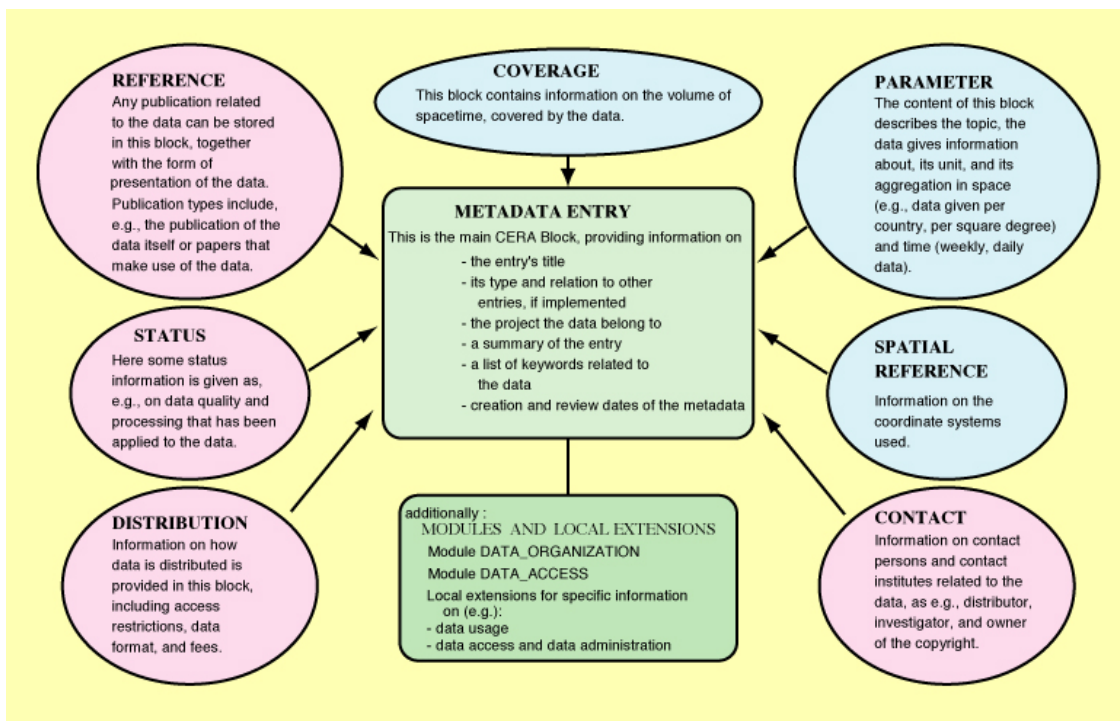
Aufgrund der schnell steigenden Anforderungen hinsichtlich Flexibilität und Skalierbarkeit zeigte sich Ende der 1990er Jahre, dass der Pflegeaufwand des primären Modells CERA-1 relativ hoch sein würde. Neue Projekte erforderten oftmals spezifische Datenstrukturen, die im bestehenden Entwurf nur schwer abzubilden waren. Man entschloss sich daher zu einer Weiterentwicklung des bestehenden Datenmodells.

1.1.2. Das Datenmodell CERA-2

Die Entwicklung der zweiten Generation des CERA-Datenmodells (CERA-2) wurde noch im DKRZ-Umfeld begonnen und im Jahr 2001 abgeschlossen. Es handelte sich um eine gemeinsame Aktivität von AWI (Alfred-Wegener-Institut, Bremerhaven), PIK (Potsdam-Institut für Klimafolgenforschung) und M&D.

In Erweiterung der ersten Version wurde hier ein modularer Architekturentwurf des Datenmodells entwickelt. Diese Modularisierung erlaubt die flexible Erweiterung des Datenmodells und so die problemlose Einbindung neuer, bisher nicht berücksichtigter Datentypen und Datenhaltungsaspekte. So war zum Beispiel die Integration von Beobachtungsdaten in CERA-1 praktisch unmöglich, in der aktuellen Version CERA-2 stellt diese kein Problem dar.

Der Entwurf von CERA-2 erwies sich für die gesamte Laufzeit von M&D als tragfähig. Größere Änderungen waren in dieser Zeit nicht notwendig. Der modulare Aufbau ermöglichte es, kleinere Ergänzungen jederzeit problemlos vorzunehmen, wann immer sich durch Projekte oder Nutzer neue Anforderungen stellten.



Die **Abbildung** zeigt den modularen Aufbau des Datenmodells CERA-2 (03.2001). Das komplette Modell ist in dem DKRZ Technical Report No. 15 (http://dx.doi.org/10.2312/WDCC/DKRZ_Report_No15) dokumentiert.

Um eine verteilte Datenhaltung besonders zu unterstützen, sieht CERA-2 über seinen modularen Aufbau hinaus vor, einzelne strukturelle Gruppen lediglich bei Bedarf zu nutzen. Dadurch war in Gemeinschaften unterschiedlicher Nutzergruppen ein gemeinsamer Kernbereich von Metadaten festgelegt. Darüber hinaus gab es periphere Bereiche, die bedarfsbezogen hinzuoptiert werden konnten.

Das CERA-Datenmodell erfüllt den ISO19115-Standard für Metadaten. Daher kann der Katalog von internationalen Datenportalen angesprochen werden, die diesen Standard einhalten. Das Füllen des Kataloges erfolgt in der Regel über XML-Dateien, bei einer anschließenden Kontrolle können die Einträge noch manuell korrigiert werden. Die automatisierte Eingabe in die DB ist im Internet dokumentiert (<http://input.wdc-climate.de>), der Datenexport kann ebenfalls in XML erfolgen und daher leicht in jedes benötigte Format umgewandelt werden.

1.1.3. Weitere Anforderungen der Nutzer

Das neu entwickelte, hoch flexible Datenmodell CERA-2 musste nach seiner Implementierung den Nutzerbedürfnissen angepasst werden. Hierzu gehörte sowohl die detaillierte Ausgestaltung der Modellstrukturen als auch die Festlegung der Zugangsformen über die Nutzeroberfläche.

Um diese Fragen nutzernah zu beantworten, wurde ähnlich dem Community-Modell-Workshop am 26. und 27. März 2002 ein Abstimmungsworkshop zur Datenhaltung am DKRZ und zur Unterstützung im Datenmanagement durch M&D organisiert. Zentraler Punkt waren dabei die Vorstellung und Diskussion der CERA-Klimadatenbank sowie die Interessenlagen der verschiedenen Nutzer im Bereich der deutschen Klimaforschung.

In der Abschlussdiskussion wurden anzustrebende Ziele definiert:

- Aufbau geographisch verteilter Archive mit Informationen über die verfügbaren Daten und deren Zugang,
- Erweiterung des bestehenden Werkzeuge zur Bearbeitung von Klimamodelldaten in der CERA-Datenbankumgebung,
- zentrale Kapazitäten am DKRZ zum Post-Processing von Klimamodelldaten und zu deren Reduzierung im Umfang vor dem Transport über das Internet,
- direkter Zugriff auf Beobachtungsdaten des DWD.

Von Seiten M&D wurde bereits 2000 an den WLA der Wunsch herangetragen, Daten des DWD, insbesondere Stationsdaten im Rahmen der M&D Serviceleistungen in Hamburg zu spiegeln und der Klimaforschungsgemeinschaft zugänglich zu machen. Hierzu war vielfältiger Bedarf vorhanden. Die Möglichkeit einer Zugriffskontrolle als Grundforderung des DWD bestand zwar bei M&D und wurde auch bereits für den Zugriff auf ECMWF Re-Analysedaten genutzt, aber eine Spiegelung der DWD-Daten konnte nicht realisiert werden. Der WLA befürwortete diesen Wunsch und bat den BMBF, diese Aktivität zu unterstützen und mit dem DWD über eine entsprechende Rahmenregelung für die deutschen Wissenschaftseinrichtungen zu sprechen. Diese Anforderung konnte bis Projektende nicht umgesetzt werden. Der DWD stellte jedoch zum Ende der Projektlaufzeit ausgewählte Daten über ein eigenes Web-Portal zur Verfügung.

Weitere grundlegende Forderungen, so zum Beispiel vermehrte Föderation von Datenarchiven mit zentralem Zugriffsportal oder Server-site-Daten-Processing und Zugang zu den Beobachtungen des DWD begleiteten die Diskussion des Datenmanagements bei M&D über die gesamte Laufzeit des Projektes. In diesen Punkten konnten in unterschiedlichem Umfang Fortschritte erzielt werden.

1.1.4. Die Anerkennung als Weltdatenzentrum

Um der deutschen Klimaforschung einschließlich ihrer Datenproduktion und -haltung international eine bessere Sichtbarkeit zu geben, wurde Mitte 2001 ein Antrag auf Anerkennung der CERA-2 Klimadatenbank als ICSU⁴ World Data Center gestellt und von DKRZ und MPI-M unterstützt.

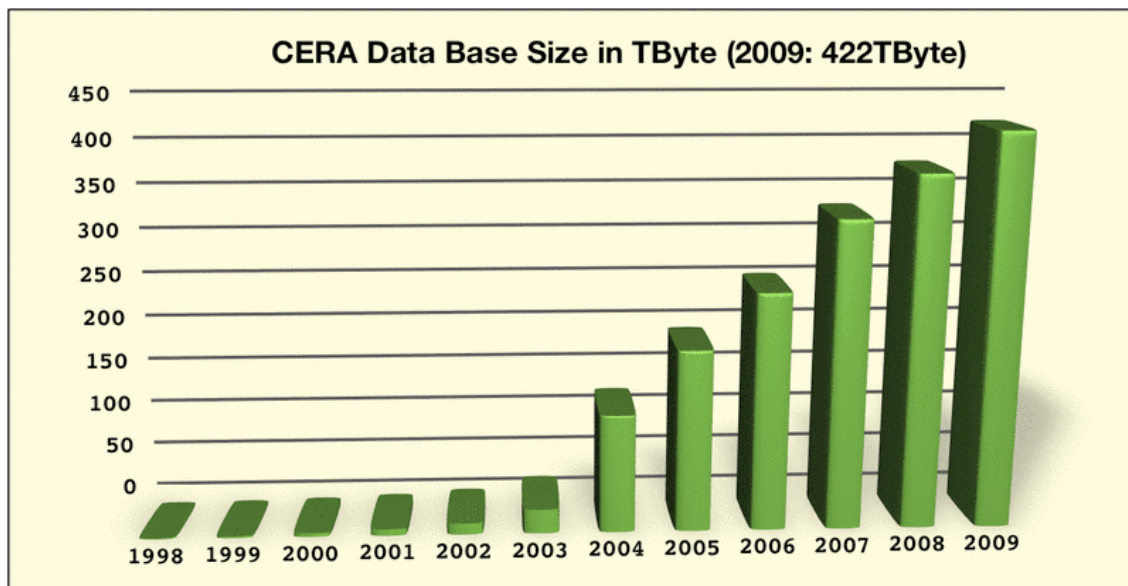
⁴ ICSU: International Council for Science

Die Begutachtung durch das ICSU World Data Center Panel und die Deutsche Forschungsgemeinschaft (DFG) fand im September 2002 statt. Die Befürwortung der CERA Datenbank als ICSU World Data Center Climate (WDCC) erfolgte im November 2002. Im Februar 2003 schließlich traf die formale Anerkennung durch das ICSU WDC Panel im DKRZ ein. Mit dieser Anerkennung als World Data Center Climate wurde die CERA-Datenbank als wissenschaftliches Langzeitdatenarchiv für Klimamodell­daten und verwandte Beobachtungen international erheblich besser wahrgenommen.

1.1.5. Entwicklung der Datenvolumina

Zum Zeitpunkt der Überleitung von M&D zum MPI-M enthielt die CERA-Datenbank am 08. Januar 2000 zunächst 4.4 Terabyte Klimadaten, die in 95 Experimenten und 9897 verschiedenen Datensätzen gespeichert waren.

Die Entwicklung der CERA Datenbank des WDCC ist in der folgenden Abbildung dargestellt.



Die **Abbildung** zeigt die Entwicklung des Datenvolumens im WDCC des DKRZ bis Ende 2009.

Die Entwicklung des WDCC während der zehnjährigen Projektlaufzeit zeigte einen Datenzuwachs von einem Faktor 100 von etwas über 4 TB Ende 1999 bis auf gut 400 TB Ende 2009.

Die 422 TB WDCC Datenvolumen Ende 2009 teilten sich auf in:

- Datenexperimente (typisch einzelnes Modellexperiment): 1.396
- Zeitserien einzelner Variablen (hier Datensätze): 167.278
- Zahl der Datenentitäten für Zugriff (z.B. ein globales Feld): 8.582.418.447

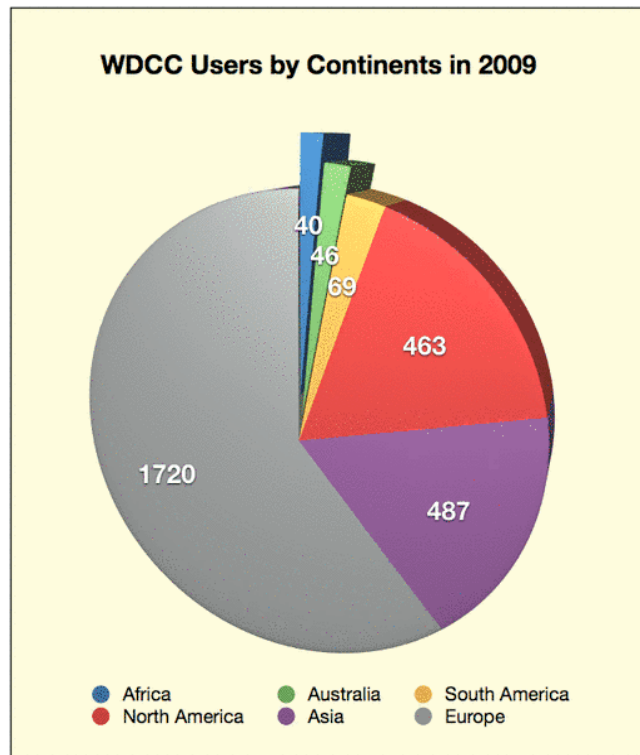
Ein deutlicher Sprung in der Zuwachsrates zeigt sich 2004. Ursache hierfür ist der Beginn der Konsortialrechnungen, die im Rahmen der Integrated Model and Data Infrastructure (IMDI) zusammen mit dem automatisierten Füllen der Datenbank gestartet wurden. Die Jahre 2004 und 2005 waren dominiert von den Daten aus den globalen Berechnungen für den vierten Assessmentreport des International Panel on Climate Change (IPCC-AR4-Rechnungen), der Tochterorganisation von WMO und UNEP.

Die Zuwächse der Jahre 2006 und ein Teil von 2007 spiegeln die Regionalisierung der Klimaszenarien durch CCLM und REMO wider. Die Zuwächse 2007 und 2008 wurden verursacht durch das Millennium-Experiment und Langzeitarchivierung aus wissenschaftlichen Projekten. Der Zuwachs 2009 zeigt den Eintrag aus internationalen Projekten wie z.B. ENSEMBLES und COPS.

1.1.6. Die Entwicklung der Nutzergemeinde

In der CERA-Datenbank waren zum dem Zeitpunkt des Übergangs der Gruppe M&D zum MPI-M 105 Benutzer eingetragen. Die Nutzungsstatistik der CERA-Datenbank verzeichnete im Januar 850 Daten-Downloads mit einem Datenvolumen von 60 GByte.

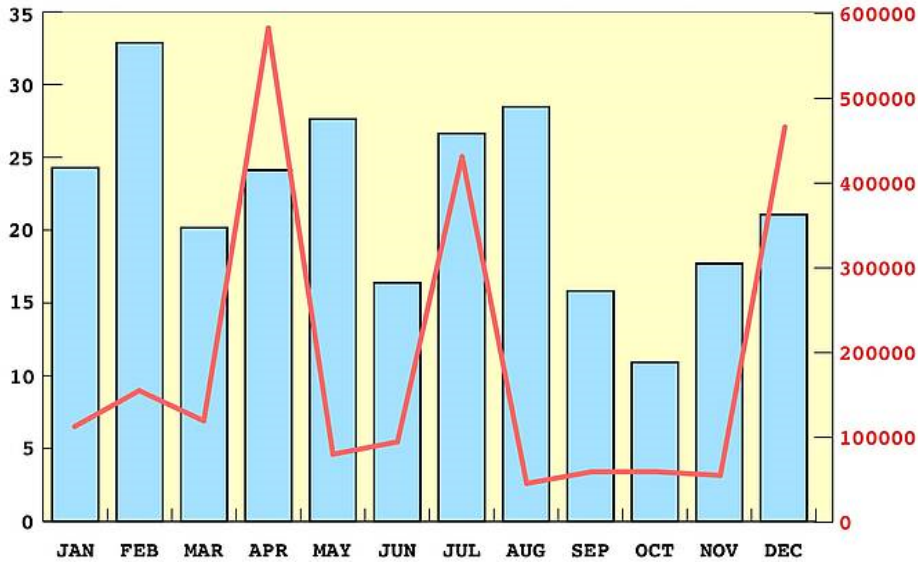
Dreiviertel der Datenzugriffe kamen von außerhalb Hamburgs insbesondere auch von internationalen Forschungseinrichtungen. Schwerpunkte der Datenzugriffe waren IPCC-Modelldaten und die Re-Analysen von ECMWF (European Centre for Medium-Range Weather Forecasts) und NCEP (U.S. National Center for Environmental Prediction). Gezählt wurden und werden in der Nutzerstatistik nur Nutzer und Zugriffe, die Klimadaten aus dem WDCC heruntergeladen haben. Zugriffe auf den Web-basierten Datenkatalog erforderten keine Nutzererkennung und wurden nicht erfasst.



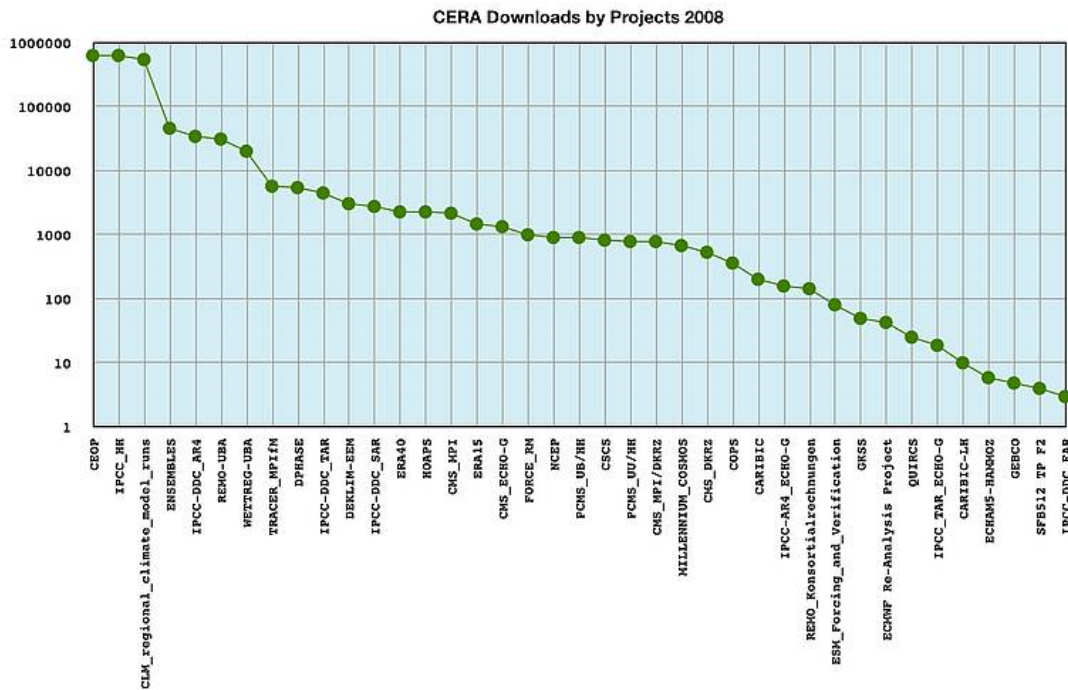
Die **Abbildung** zeigt die geographische Verteilung der der Datennutzer des WDCC im Jahre 2009.

Die Anzahl der WDCC-Nutzer von ca. 2800 im Jahre 2009 ist zu vergleichen mit den 100 Nutzern zu Beginn des Jahres 2000. 2009 kamen ca. 60% der Nutzer aus Europa. Nordamerika und Asien stellten jeweils einen Anteil von ca. 17% und die übrigen Kontinente, Australien, Südamerika und Afrika, stellten zusammen einen Anteil von 6%.

Cera Download 2008 in terabyte (left, blue) and counts (right, red) Metadata requests are not counted



Die **Abbildung** zeigt die Datenzugriffe als Monatssummen des Jahres 2008. Blaue Balken und blaue Legende gibt die Monatssumme in TByte, rote Kurve und rote Legende die Monatssumme der Einzelzugriffe auf die Datenbank des WDCC.



Die **Abbildung** zeigt die Datenzugriffe in 2008 klassifiziert nach Datenprojekten im WDCC in logarithmischer Darstellung.

Die beiden oberen Graphiken zeigen die Zugriffsstatistik des WDCC. Ausgewählt wurde das Jahr 2008, da 2009 nicht repräsentativ war. 2009 fand am DKRZ der Wechsel vom HLRE-I auf den HLRE-II statt. Dies war für die Datenbank nicht nur mit einem Hardware-Wechsel verbunden, sondern auch die Technologie zur Datenspeicherung wurde geändert und auf die Erfordernisse der zukünftigen CMIP5/IPCC-AR5 internationalen Datenföderation umgestellt. Damit verbunden waren Zeiten, an denen die Datenbank nicht zur Verfügung stand.

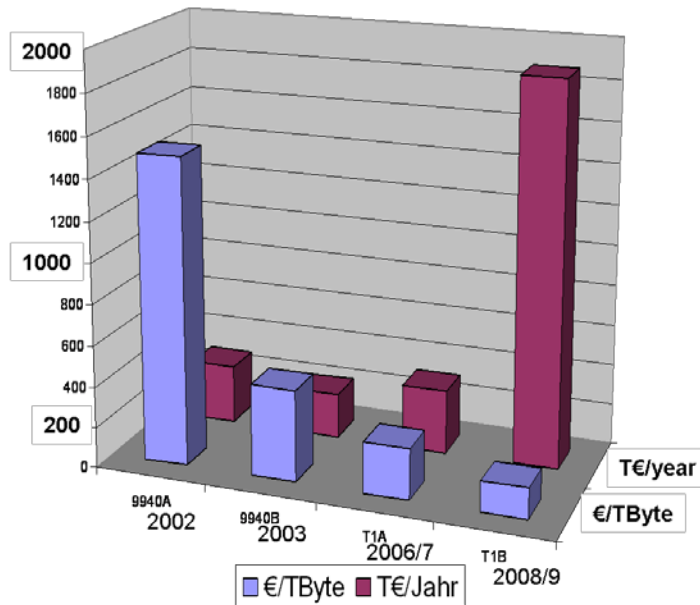
Im Januar 2000 wurden im WDCC 850 Datenzugriffe mit einem Gesamtvolumen von 60 GByte gezählt. Das ist zu vergleichen mit den Monatssummen der Einzelzugriffe und des Datenvolumens der Statistik 2008 (Januar 2008: 24 TB und 94000 Einzelzugriffe).

In der logarithmischen Darstellung der Datenzugriffe pro Datenprojekt des WDCC im Jahre 2008 waren die beiden Konsortialrechnungsprojekte des IPCC-AR4 („IPCC_HH“ und „CLM regional climate model runs“) auf Platz zwei und drei vertreten. Das Millenniums-Experiment („MILLENNIUM_COSMOS“), das sich 2008 noch im Aufbau befand, war auf einem Mittelplatz zu finden. Die Plätze eins und vier der Datenprojekt-Statistik wurden von Zugriffen auf zwei internationale Projekte belegt, nämlich CEOP und ENSEMBLES. In dieser Darstellung ist zu berücksichtigen, dass nur die Zahl der einzelnen Zugriffe dargestellt ist. Insoweit macht sie keine Aussage über die heruntergeladenen Datenvolumina.

1.1.7. Das neue Datenarchivierungskonzept

Die Diskussion zu einem neuen Archivkonzept am DKRZ begann 2005, als erste Engpässe im Massenspeicher des HLRE-I auftraten. Es war abzusehen, dass mit Archivengpässen zukünftig verstärkt zu rechnen sein würde, und diese ein ernstes Problem für den Nachfolgerechner darstellen könnten, wenn das Archivierungskonzept am DKRZ nicht angepasst würde. Die Meinungsbildung in Community und WLA führte zur Übereinkunft, ein zukünftiges Archivkonzept so zu gestalten, dass die Ressourcenrelation zwischen Compute- und Daten-Service weiterhin ein deutliches Gewicht auf dem Compute-Service zeigt. Damit könnten nicht mehr alle auf dem Compute-Server produzierten Daten langfristig gespeichert werden. Nach einer ersten Diskussion der Problematik im WLA wurden DKRZ, DKRZ Users Group und M&D gebeten, ein entsprechendes Archivkonzept auszuarbeiten.

Development of media costs vs. total costs per year



Die **Abbildung** zeigt die prognostizierte Kostenentwicklung der Speichermedien pro TByte und gesamt am DKRZ für 2002, 2003, 2006/7 und 2008/9 (HLRE-II) bei Speicherung aller produzierten HLRE Daten.

Das Archivkonzept für den HLRE-II enthält drei Datenklassen:

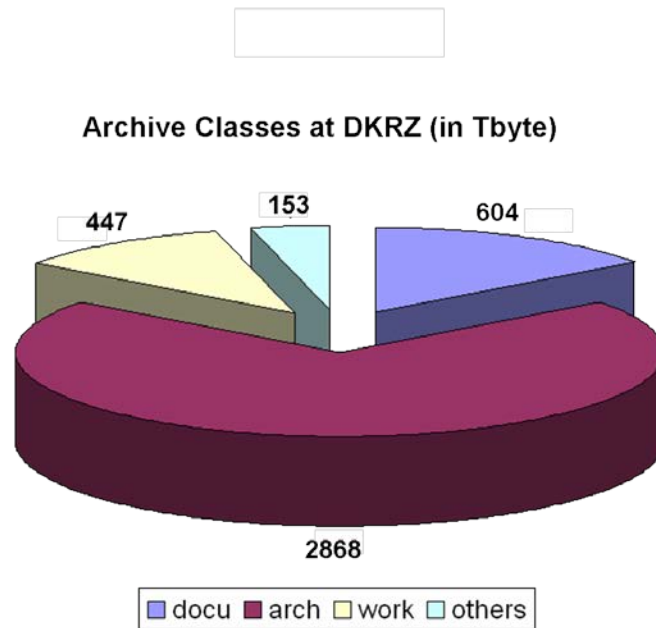
- Arbeitsdaten sind Zwischenergebnisse aus Modellrechnungen, die temporär während des aktuellen Modelllaufs gespeichert und danach gelöscht werden.
- Projektdaten sind Daten, die einem Rechenprojekt am DKRZ zugeordnet sind. Entsprechend dem Projektantrag wird neben Rechenzeit auch Speicherplatz (Festplatten- und Magnetband-Kapazität) beantragt. Das bewilligte Speicherkontingent wird vom Projektverantwortlichen verwaltet und steht dem Projekt während seiner Laufzeit zur Verfügung. Nach Ende des DKRZ-Projektes müssen Ergebnisse in die Langzeitarchivierung überführt werden. Alle nicht überführten Daten werden gelöscht. Für die Langzeitarchivierung besteht eine strenge Kontingentierung. Ein wissenschaftliches Auswahlverfahren befindet sich aktuell in der Entwicklung.

- Die Daten der Langzeitarchivierung werden im WDCC gespeichert und dokumentiert. Ohne Metadateneintrag werden Daten nicht für die Langzeitarchivierung angenommen. Tendenziell wird damit am DKRZ ein vollständig dokumentiertes, suchbares Langzeitdatenarchiv aufgebaut. Entsprechend den Regeln der guten wissenschaftlichen Praxis erhalten die Daten eine Sicherheitskopie und die Datenintegrität wird im Rahmen der Möglichkeiten des DKRZ für 10 Jahre zugesichert. Ein Löschdatum ist für diese Daten nicht vorgesehen. Aktuell befinden sich aktive Daten im WDCC die fast 20 Jahre alt sind.

Die beiden Aspekte der Langzeitarchivierung, Wissensspeicherung und Verifizierbarkeit wissenschaftlicher Ergebnisse, stellen insbesondere bei interdisziplinärem Arbeiten höhere Anforderungen an Datenschutz, Qualitätssicherung und Benutzerfreundlichkeit als es bei den beiden vorhergehenden Kategorien der Fall ist. Vor allem weil die Daten nicht ohne weiteres reproduzierbar sind und Dritte vielfach nicht das zu ihrer Handhabung und Einschätzung nötige Hintergrundwissen besitzen. Daraus folgen Anforderungen an Beschreibung und Katalogisierung.

Dieses Konzept wurde mit den DKRZ-Nutzern im Rahmen der Users Group abgestimmt. Mit Installation des HLRE-II im Jahre 2009 wurde mit der Umsetzung begonnen. Nicht abgeschlossen ist die Entwicklung eines abgestimmten, transparenten Verfahrens zur Auswahl der Projektdaten, die in die Langzeitarchivierung überführt werden sollen.

Es besteht noch Bedarf an einem Geschäftsmodell, das es erlaubt, auch Daten aus Drittmittelprojekten, die nicht auf dem HLRE-II gerechnet wurden, auf Selbstkostenbasis in die Langzeitarchivierung des DKRZ zu übernehmen. Wissenschaftlich ist dieses Vorgehen sinnvoll (siehe dazu auch WDCC Zugriffsstatistik 2008 für CEOP und ENSEMBLES). Es wird aktiv nachgefragt und kann von der Infrastruktur des DKRZ geleistet werden. Die Menge der von extern zu importierenden Daten aus Forschungsprojekten ist klein gegenüber der auf dem HLRE-II produzierten und archivierten Datenmenge.



Die **Abbildung** zeigt die Mengenverteilung der Daten des DKRZ-Bandarchivs nach Klassen im Jahre 2007. Der Bereich „docu“ in der Graphik steht für Langzeitarchivierung im Konzept, die Bereiche „arch“ und „work“ in der Graphik bezeichnen die Ebene der Projektdaten im Konzept.

Neu am Datenspeicherkonzept des DKRZ sind vor allem drei Dinge: Die Aufteilung zwischen Projekt- und Langzeitdaten, das Löschen nach Überschreitung des Verfallsdatums, wenn nicht aktiv widersprochen wird (ARCH-Ebene), und die Dokumentationsanforderungen an die langfristig gespeicherten Daten (DOCU-Ebene). Letztere geht auf die Erfahrung zurück, dass nicht dokumentierte Daten außerhalb der Projektumgebung in der Regel ihren Informationswert verlieren.

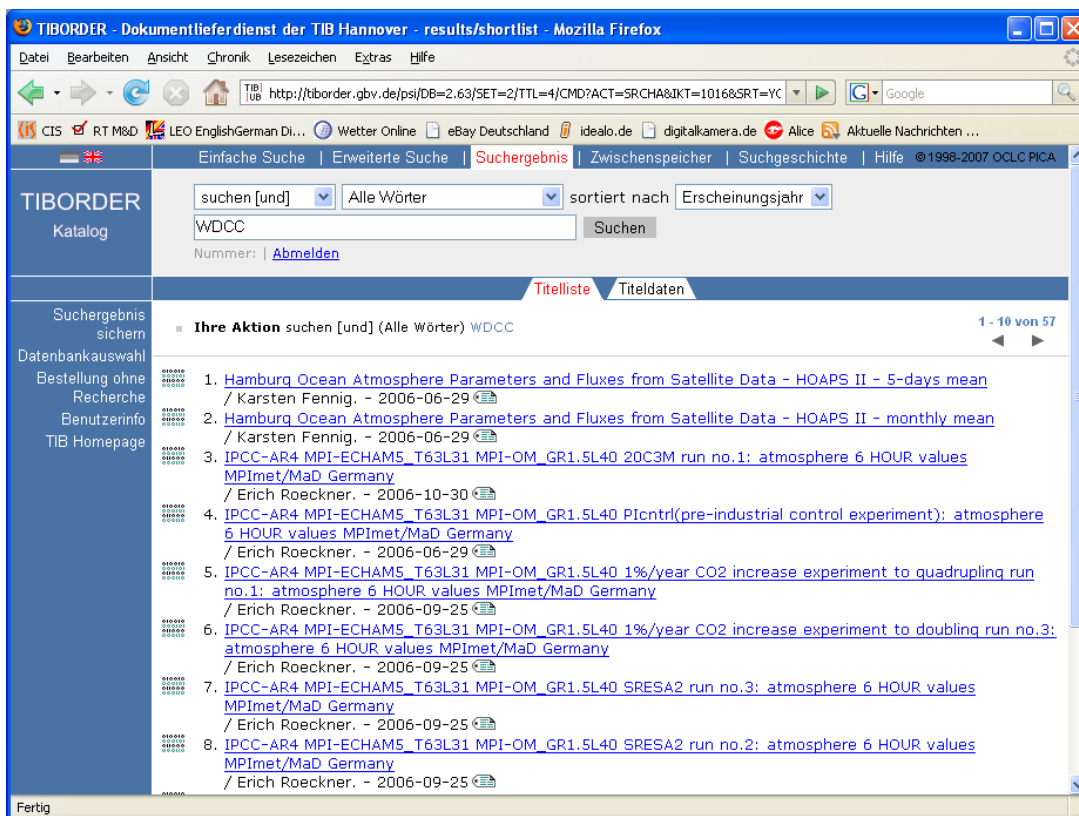
Die Katalogisierung und die Dokumentation erfolgen im Rahmen des am WDC Climate bereits bestehenden, im Internet publizierten Datenkataloges und Metadatenmodells CERA-2. Es enthält Funktionalitäten für Durchsicht, Suche, Identifizierung und Download der Klimadaten des Datenbanksystems sowie der außerhalb der Datenbank in einfachen Dateien gespeicherten Daten.

Die zur Nutzung erforderlichen Metadaten sind im Datenmodell CERA-2 enthalten, so dass nach Größen wie Variablen, raumzeitlicher Auflösung, Zuständigkeiten oder Datenqualität problemlos gesucht werden kann und eine Beurteilung der Daten möglich ist, ohne diese physikalisch herunterzuladen. Dieser Aspekt ist wichtig für Datenspeicherung auf Magnetbändern deren Zugriffszeiten um ein Vielfaches größer sind als Zugriffszeiten für Festplatten.

1.1.8. Die Einführung von persistenten Kennungen für Klimadaten (Digital Object Identifier, DOI)

Schon bald nach Beginn des Projektes hat für die Gruppe M&D das Management von Daten aus wissenschaftlichen Projekten an Bedeutung gewonnen, wobei auch die Infrastruktur des WDC-Climate genutzt wurde. Dadurch wurde 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 mindestens zehn Jahre nach Projektende verfügbar bleiben.

Zur Unterstützung des Wissenschaftlichen Datenmanagements hat das WDCC in enger Kooperation mit dem WDC Cluster Earth System Research gearbeitet. Zu ihm gehören neben dem WDCC auch das WDC-MARE am Bremer MARUM, das WDC-RSAT bei der DLR in Oberpfaffenhofen und das Geoforschungszentrum in Potsdam. Außerdem war die Technische Informationsbibliothek in Hannover (TIB) an dieser Kooperation beteiligt. Im Rahmen eines DFG-Projektes wurde das STD-DOI-Konzept entwickelt (www.std-doi.de) und prototypisch implementiert. Dabei handelt es sich um eine Erweiterung der Vorgehensweise, wie sie bei normalen wissenschaftlichen Publikationen üblich ist. Jede Publikation erhält eine einmalige Kennung, mit der sie in Zitaten dauerhaft identifizierbar und zugreifbar ist.



Die **Abbildung** zeigt die Darstellung von Klimadaten des WDC im DOI-Katalog der Technischen Informationsbibliothek, Hannover

Das Konzept STD-DOI integriert die Primärdatenpublikation in allgemeinere Publikationssysteme, wie sie in Textpublikationen verwendet werden. Zu den am WDC Climate gehaltenen Daten gelangt man im Katalog der TIB (TIBORDER, URL: <http://tiborder.gbv.de/psi/DB=2.63/LNG=DU/>) mit dem Schlüsselwort WDC. Über den DOI ist ein direkter Zugriff auf die referenzierten Daten möglich. Alle über STD-DOI registrierten Daten besitzen eine Zitiervorschrift zur Verwendung in wissenschaftlichen Publikationen und sie sind nicht mehr änderbar, um eine Verifikation wissenschaftlicher Ergebnisse auch nach deren Veröffentlichung zu garantieren.

Die wichtigsten Ziele dieses neuen Umgangs mit Primärdaten sind also langfristige und allgemein zugängliche Speicherung. Dies wird idealerweise über eine entsprechende persönliche Motivation der Wissenschaftler erreicht, die sich am besten durch zwei Faktoren befördern lässt:

- (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.

das STD-DOI Konzept wurde im Rahmen von DFG-Projekten ab 2003 umgesetzt mit folgenden Schwerpunkten:

- a) Etablierung der Publikation wissenschaftlicher Daten als neue wissenschaftliche Veröffentlichungsform.

Die Publikation und Registrierung wissenschaftlicher Daten wurde als eigenständiger, fächerübergreifender Wissenschaftsservice bei der TIB etabliert.

Als Pilotanwendung dienten dabei die Geowissenschaften/Erdsystemforschung, da hier sowohl die Probleme der Datenvielfalt und -menge auftreten als auch verschiedenste Formen der Datenerhebung vorkommen (u.a. Experimente, Feldstudien, Simulationen, Monitoring). Außerdem war hier bereits mit den WDCs eine Infrastruktur zur Langzeitarchivierung vorhanden.

- b) Übertragung des Verfahrens auf andere Fachgebiete.

Aufbauend auf den oben genannten Pilotanwendungen wurde ein auf andere Fachgebiete übertragbares Modell für die Registrierung von Primärdaten entwickelt.



Die **Abbildung** zeigt die Beteiligung von Wissenschaftlern, WDCC und TIB an den verschiedenen Schritten der Qualitätssicherung und Zuweisung einer permanenten Kennung für die wissenschaftliche Datenpublikation.

c) Etablierung der TIB als globale Registrierungsagentur für wissenschaftliche Daten. Mit der Bereitstellung einer Metadatenbank an der TIB für registrierte und damit qualitätsgesicherte und langzeitarchivierte wissenschaftliche Daten wurde die Sichtbarkeit und Nachnutzbarkeit der Daten zusätzlich erhöht. Andere Informationsanbieter erhielten die Möglichkeit, die Metadaten in ihre eigenen Portale einzubinden, wofür am WDCC generische Internet-Schnittstellen geschaffen wurden.

d) Integration der Datenpublikation in den Routinebetrieb der beteiligten Datenzentren. Durch Vergleich und Abstimmung zwischen den beteiligten Datenzentren, Wissenschaftsverlagen sowie weiteren, externen Datenzentren wurden zwischen allen Projektpartnern möglichst einheitliche generische Abläufe abgestimmt, die auch auf weitere naturwissenschaftliche Bereiche übertragbar waren.

e) Entwicklung von Kostenmodellen für Publikationsagenten und Registrierungsagentur.

Vorhandene prototypische Kostenansätze wurden systematisiert. Langzeitarchivierung wurde sowohl bei den Fördereinrichtungen als auch bei den Wissenschaftsorganisationen als Kostenfaktor eingeführt.

Im Jahre 2009 wurde in einem DFG-Projekt zur Publikation von Umweltdaten damit begonnen, ein Verfahren zur Veröffentlichung von Beobachtungsdaten zu definieren (<http://www.wikidora.com/wikidora/>), zu dokumentieren und zu etablieren. Hierzu gehört die Entwicklung schematisierter Arbeitsabläufe zur Automatisierung des Prozesses ebenso wie die Sicherstellung der Nachnutzbarkeit von Ergebnissen auch in anderen Umweltwissenschaften. Schließlich wird die Integration des Verfahrens in den Regelbetrieb des WDCC erfolgen. Die Arbeiten dauerten gegen Projektende von M&D an. Die bisher in diesem Bereich angefallenen Ergebnisse mündeten auf internationaler Ebene in das Projekt DataCite (www.datacite.org).

Dieser Service der Publikation wissenschaftlicher Daten als eigenständige, zitierfähige Datenentitäten fand auch Eingang in das Management der Klimamodelldaten, die für den nächsten IPCC Statusbericht (IPCC-AR5) produziert werden. Als dritte Stufe im Qualitätssicherungsprozess ist die STD-DOI Datenpublikation vorgesehen. Auf die verwendeten Daten kann dann präzise in den IPCC-relevanten Veröffentlichungen verwiesen werden und die in den Publikationen dargestellten Ergebnisse bleiben über Jahrzehnte nachvollziehbar, da auf die Daten über die permanente und eindeutige Kennung DOI zuverlässig auf Langzeit archivierte Daten zugegriffen werden kann.

Die bei M&D entwickelte Publikation von wissenschaftlichen Daten wurde als Service in die Infrastruktur des WDCC übernommen und steht auch nach der Projektlaufzeit allen WDCC Nutzern weiterhin zur Verfügung.

1.1.9. Der Übergang auf die neuen Hardwarestrukturen am DKRZ gegen Ende der Projektlaufzeit

Im Zuge der Vorbereitung der Beschaffung im Rahmen von HLRE-II wurde auch die technische Implementierung der CERA Datenbank neu überdacht. CERA besteht aus technischer Sicht im Wesentlichen aus drei Komponenten

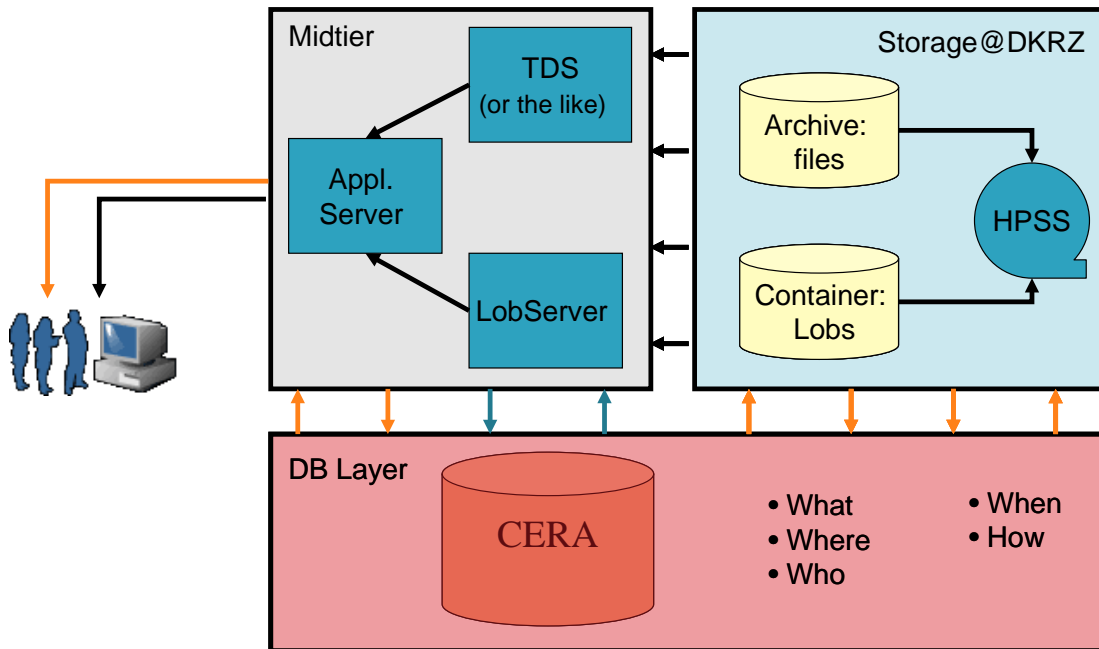
- Metadatenbank
- Benutzerportal
- Datenspeicher mit Anbindung an DKRZ Bandarchiv.

Die Konzeption und Implementierung der Metadatenbank und des Benutzerportals hatten sich bewährt, eine grundlegende Neugestaltung dieser Komponenten war daher nicht erforderlich.

Angesichts der großen in CERA verwalteten Datenmengen steht, nicht zuletzt aus Kostengründen, stets die Anbindung der Datenbank an das Bandspeicherarchiv des DKRZ im Mittelpunkt der Konzeption. Zu Beginn der Projektlaufzeit von M&D war entschieden worden, ein spezifisches Produkt des am DKRZ verwendeten hierarchischen Speichermanagementsystems (HSM) zu verwenden. Dieses Produkt war im Jahre 2007 abgekündigt worden, der Support lief mit der Außerbetriebnahme von HLRE-I aus. Dadurch bedingt musste für HLRE-II eine neue Lösung gefunden werden.

Ausschlaggebend für die später gefundene Lösung war die Anforderung, die Anbindung von CERA an das HSM künftig unabhängig von dem verwendeten HSM-System zu gestalten. Für diese Funktionalität gab es kein fertig vorgefertigtes Produkt, daher fiel die Entscheidung zugunsten einer Eigenentwicklung. Diese Eigenentwicklung wurde von M&D gemeinsam mit dem DKRZ konzipiert und bis zum Ende des Projektzeitraumes in Grundzügen implementiert. Angesichts der Außerbetriebnahme der im Rahmen von HLRE-I beschafften Hardware war jedoch zunächst die Übertragung der im Bandarchiv abgelegten Daten in das neu konzipierte System vordringlich. Diese Übertragung war zum Jahresende 2009 zum allergrößten Teil abgeschlossen. Die zu erstellenden Softwarekomponenten zur flexiblen Anbindung der Klimadatenbank an HSM-Systeme wurden nach Übergang von M&D ins DKRZ weiter entwickelt.

Einhergehend mit der Neugestaltung der Datenspeicheranbindung wurden in CERA zudem die Voraussetzungen verbessert um über das CERA Portal größere Datenbestände außerhalb der direkten Datenbankumgebung zu adressieren. Dieser transparente Zugriff auf Daten-Files auf Betriebssystemebene ist vor allem im Hinblick auf die rapide steigenden Datenmengen (1 PB und mehr) im Kontext IPCC-AR5 wichtig.



Die **Abbildung** zeigt die Metadaten- und Datenzugriffe auf das CERA Archiv. Neben den intern in CERA gespeicherten Daten werden auch Datenzugriffe auf außerhalb des CERA Datenbanksystems gehaltene Daten möglich. Die CERA Metadatenbank hält die Zugriffszeiger auf die Daten, durch systemimmanente Eigenschaften werden Zugriffe in angemessener Form protokolliert und erlauben somit auch künftig die Anpassung von CERA an die Benutzeranforderungen.

Zusammenfassung:

Der Datenteil von M&D hat sich mit seinen Entwicklungen auf dem Gebiet des wissenschaftlichen Datenmanagements zu einem anerkannten Partner in der Erdsystemforschung entwickelt und bietet Services für Speicherung und Dokumentation von sowie Zugriff auf Daten aus der Erdsystemforschung. Das Datenmodell CERA-2 wird bis heute operationell verwendet und erlaubt eine einfache Integration neuer Datentypen und Inhalte. Es ist vollständig im Hinblick auf Abbildung der Metadaten auf internationale Standards wie ISO 19115, Dublin Core, WMO-WIS und INSPIRE. Die Positionierung der CERA Datenbank als wissenschaftliches Langzeitdatenarchiv wurde mit der Anerkennung als ICSU World Data Center Climate (WDCC) im Jahre 2002 honoriert. Das Datenvolumen des WDCC wuchs während der Projektlaufzeit um den Faktor 100 auf 420 TB, die Zahl der eingetragenen Nutzer wuchs um den Faktor 30 und betrug zum Ende des Projektes 2800. Vergleichbare Zuwachsraten sind bei den Datenzugriffen (Faktor 100) und der verteilten Datenmenge (Faktor 400) zu erkennen. Neben der Etablierung der Archivfunktion wurde ein Konzept zur Publikation wissenschaftlicher Daten entwickelt und umgesetzt. Wissenschaftliche Daten werden so zitierfähig als Teil wissenschaftlicher Veröffentlichungen.

Gegen Ende der Projektlaufzeit wurde zusammen mit DKRZ und DKRZ Nutzern ein neues Archivkonzept entwickelt, das darauf zielt, die Kostenrelation zwischen Compute-

und Daten-Service zu balancieren bei einer möglichst flexiblen und effektiven Nutzung des vorhandenen Speicherplatzes. Damit verbunden ist eine Kontingentierung des Speicherplatzes und eine wissenschaftlich sinnvolle Auswahl der Daten für die Langzeitarchivierung im Rahmen des WDCC. Das Archivkonzept setzte die Regeln der guten wissenschaftlichen Praxis um und sichert Speicherung von mindestens 10 Jahren zu. Das Langzeitdatenarchiv entwickelt sich in diesem Konzept zu einem vollständig dokumentierten Archiv, aus dem die Daten auch noch nach Jahrzehnten abgerufen und zusammen mit den vorhandenen Meta-Informationen verwendet werden können.

2. Notwendigkeit und Angemessenheit der Arbeit

Die Durchführung des Vorhabens entsprach der Arbeits-, Zeit- und Ausgabenplanung. Inhaltliche Änderungen und/oder Konkretisierungen wurden in Abstimmung mit dem wissenschaftlichen Lenkungsausschuss vorgenommen.

3. Verwertung des Ergebnisses

Die Ergebnisse aus dem Projekt stehen der Klimaforschungsgemeinschaft nach Überleitung von M&D ins DKRZ im Rahmen der DKRZ Dienste weiterhin zur Verfügung.

4. Bekannte Entwicklungen auf dem Gebiet des Vorhabens bei anderen Stellen

Vergleichbare Entwicklungen für die deutsche Klimaforschung bei anderen Stellen sind nicht bekannt.

5. Veröffentlichungen

Auf eine detaillierte Auflistung der Einzelveröffentlichungen wird an dieser Stelle verzichtet auf Grund des Umfangs. Auf zentrale Dokumente für die Arbeit von M&D wird im Text zusammen mit der Quellenangabe direkt verwiesen.

Die Liste der Vorträge und Poster ab 2002 findet sich auf dem Web-Server von M&D unter

<http://www.mad.zmaw.de/service-support/lectures-posters/>

Die Liste der Veröffentlichungen ab 2003 findet auf dem M&D Web-Server unter:

<http://www.mad.zmaw.de/service-support/publications/>

Die Liste der Technical Reports von Modellbetreuung und M&D seit 1991 findet sich auf dem Web-Server von M&D unter:

<http://www.mad.zmaw.de/service-support/documents/>

Alle Dokumente stehen als PDF Dateien zum Herunterladen bereit.