

Die Arbeit im C3Grid ist eng in den internationalen Kontext der Klimaforschung eingebunden. Der nächste Klimareport des Intergovernmental Panel on Climate Change (IPCC) soll 2013 erscheinen – es laufen bereits jetzt die Vorbereitungen auf Hochtouren. Auch die deutsche Klima-Community wird wieder ihren Beitrag zu diesem internationalen Projekt leisten: Im Zuge der interministeriellen Zusammenarbeit wurde vom BMBF und BMU eine deutsche IPCC-Koordinierungsstelle eingerichtet, die die Beteiligung an IPCC-Berichten intensivieren und damit die deutsche Klimaforschung international noch sichtbarer positionieren wird.

Die Sachstandsberichte (assessment reports, AR) spiegeln den erreichten Stand der Klimaforschung wider, indem sie eine Zusammenfassung aller verfügbaren Informationen versuchen. Die Schwerpunkte liegen dabei im Bereich der Grundlagen (physical science basis), der Impact- und Adaptationsforschung, der Vermeidung von Klimaänderung sowie in der Bestandsaufnahme zur Treibhausgasemission. Für ersteren werden zahlreiche hochkomplexe Simulationsrechnungen durchgeführt, die die Basis für weitere Untersuchungen bilden. Um diese Daten einem breiten Interessentenkreis zur Verfügung zu stellen, ist eine effiziente Infrastruktur notwendig.

IPCC-Datenföderation

Für den Bereich der globalen Modellierung sind schon verbindliche Vereinbarungen getroffen worden, die auf eine verteilte Datenhaltung der Simulationsergebnisse zielen. Weltweit verteilte Datenknoten speichern die Ergebnisse der Simulationsrechnungen, die im Rahmen des IPCC AR5 gerechnet werden. In den sogenannten Gateways werden diese Daten dann gesammelt bzw. synchronisiert und Wissenschaftlern zugänglich gemacht. Ausgezeichnete Gateways bilden den IPCC Kerndaten-Pool, dem Deutschland durch das Welt Datenzentrum *WDC Climate* angehört. Hier wird ein definiertes Subset an Daten (core data) von allen Gateways und Datenknoten gesammelt und zentral verfügbar gemacht. Der Bereich *core data* umfasst ca. 10% aller auf den Datenknoten gespeicherten Daten von insgesamt über 5 PByte. Über die Gateways erhalten die Wissenschaftler einen browser-basierten Zugang zu den verteilten Datenbeständen. Schnittstellen zur Einbindung der Daten in spezifische Prozessierungswerkzeuge sind vorgesehen.

International climate model data federation (IPCC AR5)

C3Grid

Das Welt Datenzentrum *WDC Climate* bildet mit seinem Datenbestand die Schnittmenge zwischen C3Grid und der IPCC Datenföderation. Die im Rahmen des IPCC AR5 bereitgestellten Daten werden auch im C3Grid bekannt gemacht, indem die Metadaten dem dort gültigen Metadatenprofil angepasst und dann publiziert werden. Da C3Grid neben den Simulationsdaten aber auch weitere Mess- und Beobachtungsdaten zur Verfügung stellt, kann der C3Grid-Nutzer durch Verknüpfung unterschiedlicher Datenquellen neue Erkenntnisse für seine wissenschaftliche Arbeit gewinnen.

Außerdem wird im C3Grid auch weiterhin die gewohnte Funktionalität angeboten, wie das Pre-Processing beim Datenprovider, um Ausschneideoperationen in den Datensätzen gleich lokal auszuführen und damit den zu transportierenden Datenstrom zu minimieren.

Technische Umsetzung

Die C3Grid-Infrastruktur ist als service-orientierten Architektur realisiert. Hier setzen verschiedene, auf international standardisierten Technologien basierende Mehrwertdienste wie Portal, Workflow-Management und Datenlebenszyklusverwaltung um. Die Flexibilität des Systems stellt sicher, dass eine Adaptation an neue Anwendungsgebiete wie die Mitarbeit am IPCC AR5 mit relativ geringem Aufwand geleistet werden kann und bietet dabei eine einfach zu bedienende Nutzerschnittstelle für das wissenschaftliche Arbeiten.

Portal

Der Nutzer bekommt durch das Portal Zugang zum C3Grid. Die integrierte, Google-inspirierte Freitextsuchfunktion nutzt unterliegend das Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), um die Bestände der einzelnen Datenanbieter zu indizieren und stellt zudem umfangreiche Eingabemasken für die feingranulare Suche bereit. Zudem steuert der Nutzer über das Portal die Auswahl und Durchführung von wissenschaftlichen Workflows, die komplexe Prozessketten für die Analyse von Klimadaten abbilden. Schließlich ermöglicht das Portal den bequemen Download von Klimadatensätzen auf den eigenen Arbeitsplatz

Workflow-Management

Das Workflow-Management analysiert die Abhängigkeiten zwischen einzelnen Tasks und nutzt die im Grid zur Verfügung stehenden Ressourcen zur koordinierten Ausführung. Dies schließt die intelligente Auswahl geeigneter Ressourcen für die Ausführung, die rechtzeitige Bereitstellung von Eingabedaten sowie den Transport von Ausgabedaten (als Eingabe für abhängige Tasks oder als Rückgabe an den Nutzer) ein. Zugleich sorgt eine Schedulingstrategie für einen möglichst zeitoptimalen Ablauf, um Wartezeiten für die Nutzer zu minimieren und die vorhandenen Kapazitäten auf Rechen- und Daten- und Netzwerkressourcen so effizient wie möglich auszulasten.

Datenlebenszyklusverwaltung

Die Datenlebenszyklusverwaltung koordiniert in Zusammenarbeit mit dem Scheduler den Zugriff auf Klimadaten der Datenprovider im gesamten Grid-System. Sie initiiert Anfragen zum Datenstaging, transferiert Daten zwischen Compute- und Datenprovidern und bietet Funktionalität für den Datenexport aus dem Grid. Dabei werden Informationen über verfügbare Replikate als auch Ausführungszeitschätzungen der Datenprovider berücksichtigt und Garantien für die gewünschte Dienstgüte gegeben. Zudem stellt es den Virtual Workspace bereit, der eine verteilte, für den Nutzer transparente Umgebung für die Ablage von Arbeitsdaten bietet.

