



## Collaborative Climate Community Data and Processing Grid (C3-Grid)

# Portal Generation 0: Spezifikation und Dokumentation

---

Arbeitspaket:	AP1 - Nutzerschnittstelle
Autoren:	B. Fritsch, B. Bräuer
Version:	Generation 0 (Stand: Okt.2006)
Veröffentlichungsdatum:	Februar 2007
AP-Koordination:	AWI
Partner:	Uni Bremen, DKRZ, M&D, MPI/IFM- Geomar, FUB, CEI, ZIB
Ansprechpartner:	B. Bräuer
Email:	benny.braeuer@awi.de

---

## **Inhalt**

1. Einleitung
2. Funktionalität
  - 2.1. Suchfunktion
  - 2.2. Beispiel für Datensuche
  - 2.3. Demoworkflow
3. Implementierung der C3-spezifischen Portlets
4. Referenzen

# 1. Einleitung

Für die Anwendung und Verbreitung des C3-Grids ist es notwendig, die technischen Details der Grid-Technologie für den Nutzer möglichst transparent zu machen. Daher kommt dem Portal eine hohe Bedeutung zu.

Der Nutzer des C3-Grids erhält durch das Portal Zugriff auf die Ressourcen (Daten und Rechenkapazität) des Grids. Im Portal initiiert er Suchanfragen über die verteilten Datenbestände und schickt Processing-Jobs ab. Im Folgenden sollen die wesentlichen Eigenschaften des Portals in der Generation 0 vorgestellt werden.

Eine erste Testversion des Portals konnte im Mai 2006 online geschaltet werden, so dass Nutzer aus dem Projekt sich bereits anmelden und testen können. In einem speziellen projektinternen Meeting wurde die Testversion den Nutzern vorgestellt und die weitere Entwicklung an ihre Anforderungen durch die Workflows angepasst. Im Oktober 2006 erfolgte die Vorstellung des C3-Grids auf der 7. Deutschen Klimatagung in München. Dokumentiert wird in diesem Papier dieser Implementationsstand.

# 2. Funktionalität

Als Basis für die Implementierung des Portals wurde GridSphere gewählt, das auch vom DGI unterstützt wird.

Abbildung 1 stellt die Architektur dar. Darin ist bereits die für spätere Generationen geplante Einbindung von Shibboleth enthalten. Als Portalserver kommen 2 Server (je 2 Opteron248, 2.2 GHz) zum Einsatz (1 Produktionssystem, 1 Testsystem am AWI).

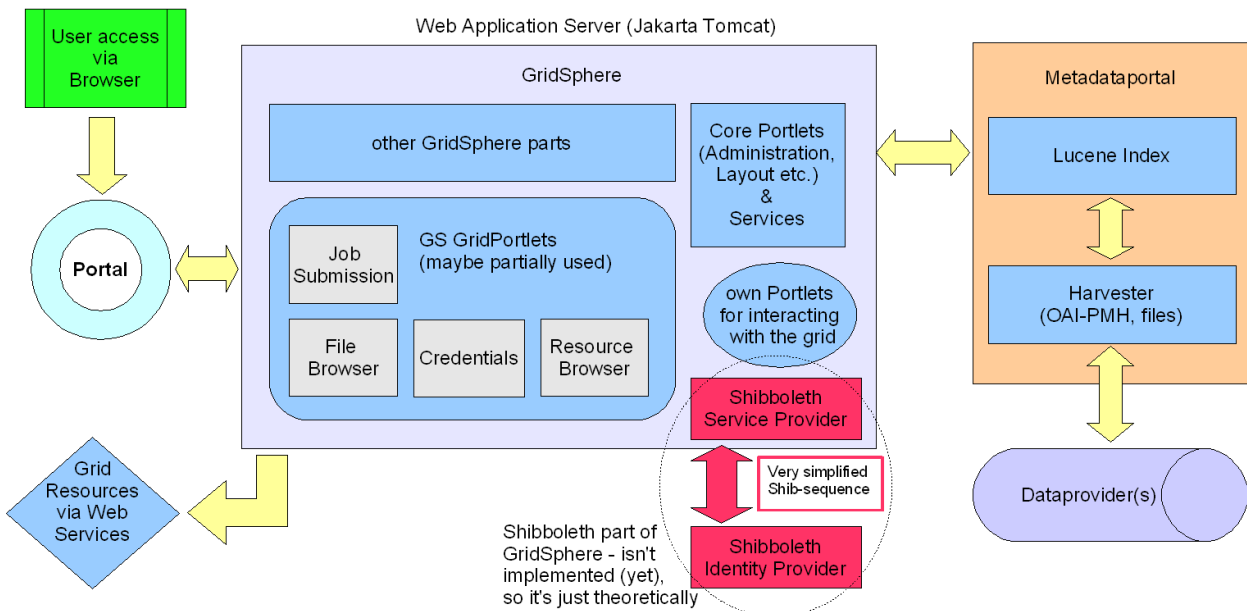


Abbildung 1: Architekturbild für das Portal im C3-Grid Generation 0

Das Portal bietet in der Generation 0 folgende grundlegende Funktionen an:

- Einfache google-ähnliche Suche
- Erweiterte Suche
- Ein Beispielworkflow Stormtrackanalyse

## 2.1. Suchfunktion

Für die Suche greift das Portal auf einen zentralen Metadatenkatalog zu. Der globale Dateninformationsdienst DIS besitzt in einem zentralen Katalog die Information über alle Daten im Grid. Die entsprechenden Metadaten werden über das Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) von den einzelnen Datenanbietern bezogen. Um den Umfang dieses Zentralkatalogs zu begrenzen und die Suche performant zu gestalten, werden dort nur solche Metadaten gespeichert, die für die Suche unbedingt notwendig sind (discovery metadata). Für das weitere Prozessieren der Datensätze werden aber zusätzliche Informationen benötigt (use metadata), die dann lokal bei den Datenanbietern vorliegen.

Für die ersten Tests wurden die Metadaten der vom WDC Climate zur Verfügung gestellten IPCC-Rechnungen im zentralen Metadatenkatalog indiziert, so dass sie damit im C3-Grid suchbar sind. Für die Indizierung kommt Apache Lucene zum Einsatz. Entsprechende Vorarbeiten wurden an der Universität Bremen geleistet. Damit wird ein Volltext-Index erstellt, der Grundlage für eine einfache Suche („google-like“) ist. Daneben existiert die Option einer erweiterten Suche, deren Parameter direkt an die Anforderungen der Community für typische Workflows angepasst sind. Weitere technische Details des Dateninformationsdienstes DIS sind [Lang2006,1] zu finden

## 2.2. Beispiel für Datensuche

Beispielhaft soll die Suche nach einer Variablen in der Testversion des Portals demonstriert werden. Der Nutzer sucht auf der Einstiegsseite (siehe Abbildung 2), nach allen vorhandenen IPCC-Rechnungen, die von einem bestimmten Datenprovider – hier WDC Climate - angeboten werden und möchte dort nach bestimmten Variablen suchen.

The screenshot shows the 'Advanced Search' interface. The main heading is 'Einstieg ins C3-Grid'. Below it, there's a sub-heading '→ über die Suche'. The search criteria are: 'Suche nach' (ipcc), 'Suche in' (WDC Climate selected), 'Ergebnisse pro Seite' (10), and 'Spezielle Einstellungen' (Suche nach Variablen selected). A 'Weiter' button is present. The second section is '→ über den Workflow (folgt)' with options: Basisworkflow, Stormtrack, and Kampagnenbegleitung, followed by another 'Weiter' button.

Abbildung 2: Einstiegsseite für die Suche im Portal

Auf der folgenden Seite erhält er dann eine Übersicht über die vorhandenen Variablen, aus der er dann auswählen kann (Abbildung 3). Die Variablennamen stammen aus der Climate Forecast (CF) Convention. Da diese Liste sehr lang ist, ist das Blättern darin recht umständlich. Der Nutzer kann daher Wortbestandteile eingeben (hier „sea“). Danach reduziert sich die Auswahlliste auf diejenigen Variablen, die den Wortbestandteil enthalten.

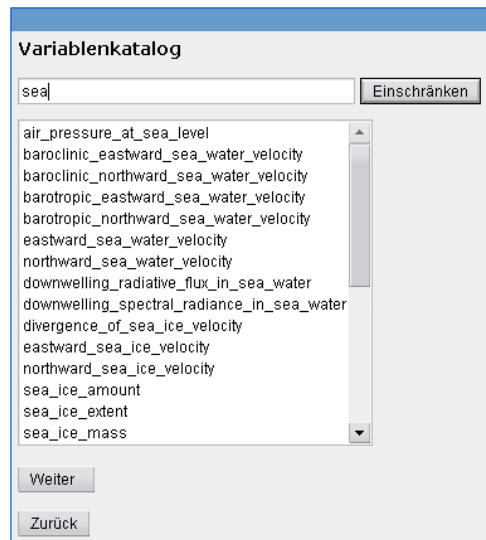


Abbildung 3: Variablenliste

Danach erfolgt die Auswahl in Raum und Zeit (Abbildung 4). Dabei können die Regionen einerseits durch die Angabe von Längen- und Breitengraden vom Nutzer definiert werden. Außerdem wurden häufig verwendete Regionen (z.B. die einzelnen Kontinente oder Ozeane) bereits im Portal vordefiniert und erlauben so einen besonders einfachen Zugriff. Die vertikalen Grenzen des Gebiets können danach spezifiziert werden. Da im C3-Grid sowohl atmosphärische als auch ozeanografische Datensätze zugreifbar sind, kann die Höhe wahlweise in m oder in hPa angegeben werden.

Durch die Eingabe von Start- und Enddatum werden nur solche Datensätze adressiert, die den angegebenen Zeitraum vollständig umfassen. Mit diesen eingegebenen Parametern wird dann die Suche gestartet.

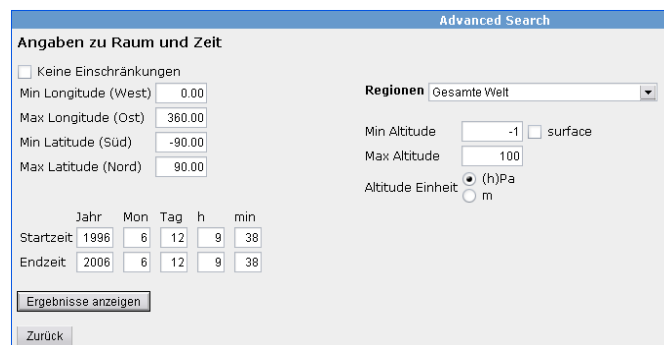


Abbildung 4: Auswahl von regionalen und/oder zeitlichen Ausschnitten

Nachdem das Portal den Data Information Service DIS kontaktiert hat, werden die gefundenen Datensätze angezeigt (Abbildung 5). Durch den mitgelieferten Abstract bekommt der Nutzer weitere Detailinformationen, die bei der endgültigen Auswahl des konkreten Datensatzes helfen können.

The screenshot shows a search results page titled 'Ergebnis' with 'Found 36 result(s)'. Below the search bar, there are four search results, each with a radio button, a title, a 'Details anzeigen' link, and an abstract. The results are:

- IPCC\_HH:EHS-T63L31\_OM-GR1.5L40\_20C\_3\_OC (World Data Center for Climate)**: Abstract describes a scenario run using ECHAM5.2.02a coupled to MPI-OM Vers. 1.0 GR1.5L40, starting 1860 from pre-industrial control run (CTL). Monthly means are available, too. Ocean data are available. Related experiments: - EHS-T63L31\_OM-GR1.5L40\_CTL - EHS-T63L31\_OM-GR1.5L40\_A1B\_1,2,3 - EHS-T63L31\_OM-GR1.5L40\_B1\_1,2,3 - EHS-T63L31\_OM-GR1.5L40\_A2\_1,2,3 - EHS-T63L31\_OM-GR1.5L40\_CMIP\_1,2,3 - EHS-T63L31\_OM-GR1.5L40\_20C3M\_1,2,3 - EHS-T63L31\_OM-GR1.5L40\_2C02. The output from the model run: hurrikan.dkrz.de:/ut/m/m214002/EXP500/run524 Please note: experiment\_name/acronym was renamed (27-JUN-2005, 20C\_2 changed to 20C\_3)
- IPCC\_HH:EHS-T63L31\_OM-GR1.5L40\_CTL\_6H (World Data Center for Climate)**: Abstract describes 6 hourly values of pre-industrial climate simulation. Here, for the year 1860, concentrations of well mixed greenhouse gases (CO2, CH4, N2O) were prescribed. Ocean data (MPI-OM) are available within this experiment, they are stored in EXTRA FORMAT. Datasets with monthly mean values are also available. As the boundary conditions are not time dependent, the time access is arbitrary. Technical data to this experiment: The experiment is using ECHAM5.2.02 coupled to MPI-OM Vers.1.0 GR1.5L40 and was run on a NEC-SX6(hurrikan). The output from the model run: hurrikan.dkrz.de:/ut/3/m214002/EXP500/run520
- IPCC\_HH:EHS-T63L31\_OM-GR1.5L40\_CTL\_MM (World Data Center for Climate)**: Abstract describes monthly average values of pre-industrial climate simulation. Here, for the year 1860, concentrations of well mixed greenhouse gases (CO2, CH4, N2O) were prescribed. Ocean data (MPI-OM) are available within this experiment, they are stored in EXTRA FORMAT. Datasets with higher resolution (6 hour mean) are also available. As the boundary conditions are not time dependent, the time access is arbitrary. Technical data to this experiment: The experiment is using ECHAM5.2.02 coupled to MPI-OM Vers.1.0 GR1.5L40 and was run on a NEC-SX6(hurrikan). The output from the model run: hurrikan.dkrz.de:/ut/3/m214002/EXP500/run520
- IPCC\_HH:EHS-T63L31\_OM-GR1.5L40\_20C\_1\_6H (World Data Center for Climate)**: Abstract describes 6 hourly values of a 20th century simulation (including year 2000) with observed anthropogenic forcings(CO2, CH4, N2O, CFCs, O3 and sulfate) initialized in year 2190 of the preindustrial control run. This is followed by a commitment experiment for the 21th century (years 2001-2100) with all concentrations fixed at their levels of the year 2000. Data Sets with monthly mean values are also available. Technical data to this experiment: The experiment is using ECHAM5.2.02a coupled to MPI-OM Vers. 1.0 GR1.5L40 The output from the model run: hurrikan.dkrz.de:/ut/k/k204076/EXP000/run009 Please note: experiment\_name/acronym was renamed (27-JUN-2005, 20C\_0 changed to 20C\_1)

Abbildung 5: Liste der gefundenen Modell-Läufe

### 2.3. Demoworkflow

Neben der Datenrecherche ist die Abarbeitung von typischen Workflows im Gridumfeld wesentliches Ziel von C3-Grid. Im Portal gibt es dazu einen alternativen Einstiegspunkt über die Workflows. Als Prototyp wurde die Analyse von Stormtracks gridifiziert. Das Tool wurde von den Kollegen der FUB zur Verfügung gestellt. In Generation 0 ist dieser Workflow zwar noch statisch, es gibt genau einen Datenprovider und genau einen Compute-Ressourcen-Provider. Der Scheduler muss daher den Job nur an die entsprechenden, fest einprogrammierten Provider durchreichen. Trotzdem kann damit bereits die Arbeitsweise im Grid demonstriert werden.

Der Workflow stellt an seine Inputdaten einige einschränkende Anforderungen (min/max Höhe 500 hPa, Datensatz muss die Variable „geopotential height“ enthalten, 6-Stunden-Mittel). Diese Einschränkungen sind für die Demonstration fest im Portal verankert und schränken damit die möglichen Datensätze ein. Dadurch ist gewährleistet, dass im Grid nur mit sinnvollen Daten gearbeitet wird. Der Nutzer wählt danach den von ihm gewünschten Datensatz aus und schickt den Job ab. Das Portal kontaktiert den Scheduler, der die Processierungsanfrage dann an den Datenmanagementsdienst (DMS) und die Provider von Compute-Ressourcen weiterleitet.

## 3. Implementierung der C3-spezifischen Portlets

Das Portal weist in dieser Generation zwei Schnittstellen auf – eine zum DIS sowie eine weitere zum Scheduler. Da GridSphere vollständig die Java Specification Request 168 erfüllt, folglich komplett in Java implementiert ist, wurden die Klassen des DIS direkt in das Portal eingebunden. Gegenüber der ebenfalls möglichen Web Service-Lösung kann so eine Schicht gespart werden, was sich positiv auf die Leistung der Suche auswirkt.

Die Schnittstelle zum Scheduler ist ein klassischer Axis Web Service. Das Portal als Client baut sich anhand von Benutzereingaben, Informationen aus den Metadaten sowie einem in Generation 0 fest kodierten Scriptpfad (+ Aufrufparameter) ein Job Submission Description Language (JSDL) - Objekt zusammen, welches dem Scheduler übergeben und von diesem ausgewertet wird.

## 4. Referenzen

[Langh2006,1] T, Langhammer, F. Schintke, Collaborative Climate Community Data and Processing Grid (C3Grid), T2.1: Grid Information Service Architecture and Specification, 2006