

# Technische Aspekte von C3Grid

Christian Grimme, Stefan Plantikow

16. Februar 2009



Bundesministerium  
für Bildung  
und Forschung

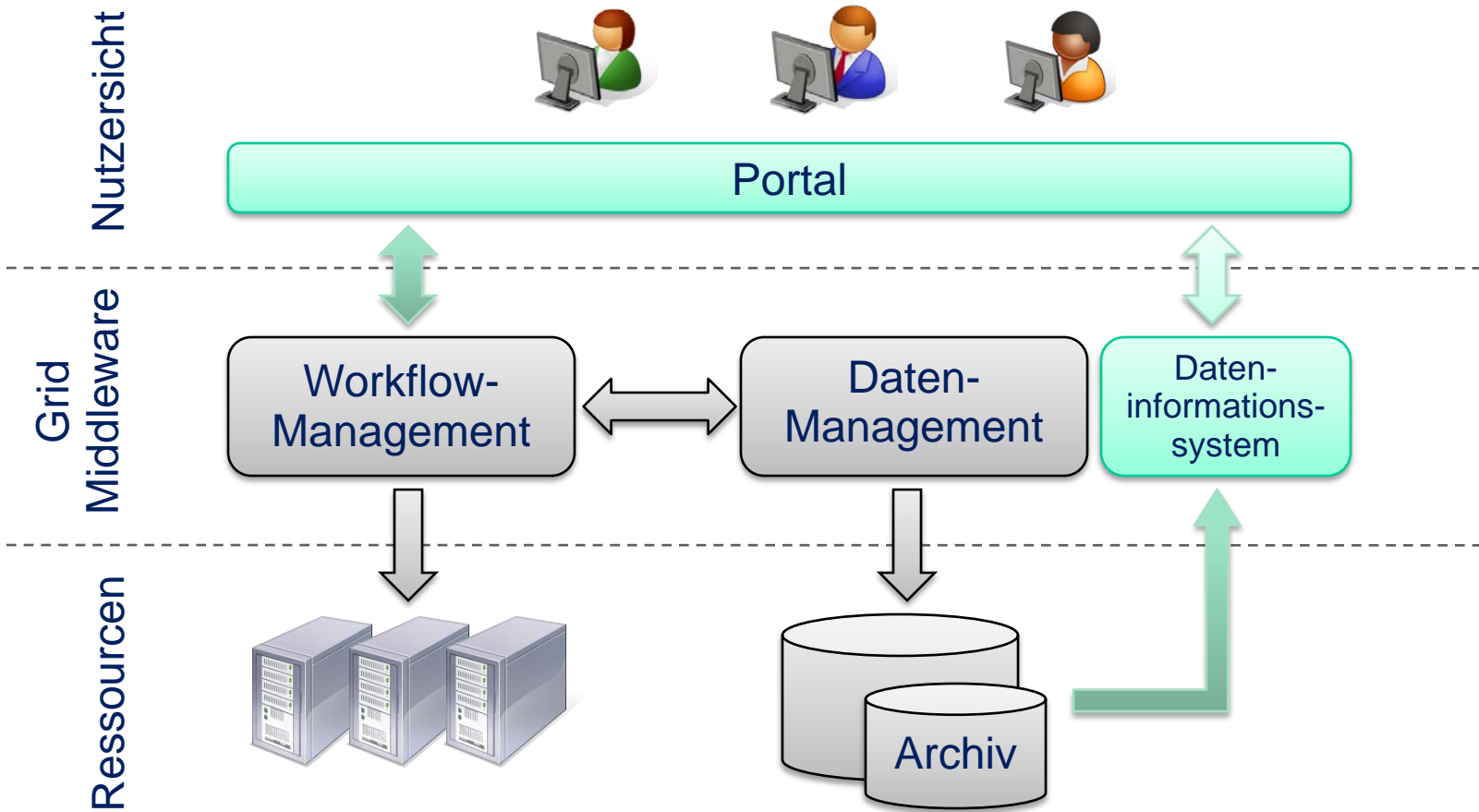
- Idee des Werkzeugs „C3Grid“
  - Ausgangssituation
  - Anforderungen und Konzept
- Technische Infrastruktur
  - Architektur von C3Grid
  - Einige Dienste im Detail
- Teilnahme am C3Grid
  - Datenprovider
  - Rechenprovider
  - Workflowentwickler
- Zusammenfassung

## Technische Sicht auf Workflows im Klimakontext

- Datenbezogen:
  - Datenbeschaffung über Datenbanken bei Daten Providern
  - Datenvorbereitung (Datenselektion, Konvertierung, ...)
  - Manueller Datentransfer und Dateiverwaltung
- Jobbezogen:
  - Auswahl von geeigneten Rechenressourcen; gewöhnlich nur lokale Installation verfügbar
  - Beschaffung von Zugriffsrechten
  - Starten von Skripten über Kommandozeile, einschließlich aufwändiger Parametrisierung

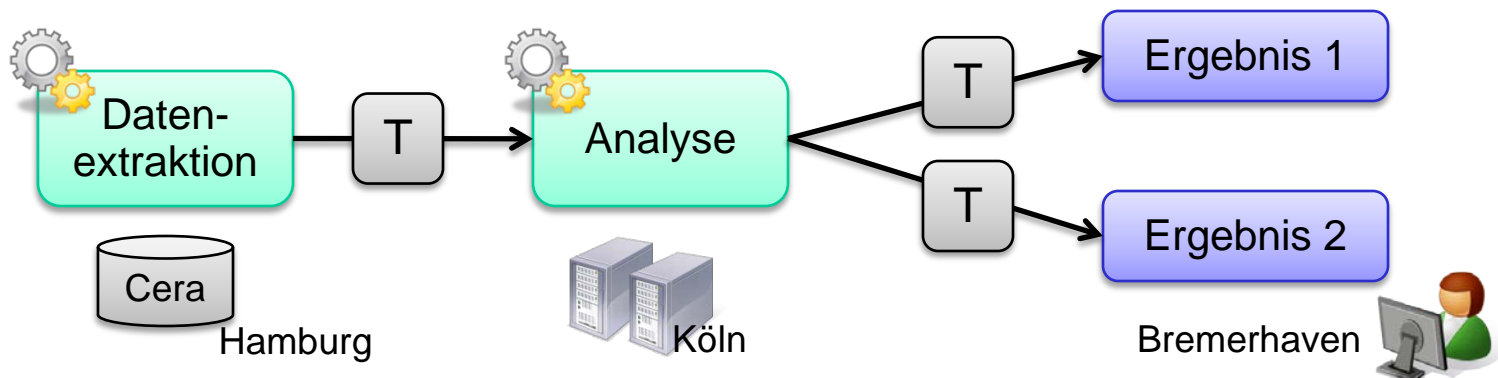
- Verbreiterung der Datenbasis
  - Zentraler Zugriff auf viele Datenarchive
  - Vereinheitlichung der Formate für Analysevorgänge
- Automatisierung der Workflowbearbeitung
  - Formulierung von komplexen Abläufen
  - Nutzung vieler Rechenressourcen
  - Zeitliche Optimierung des Ablaufes
- Übernahme von Datenmanagement-Funktionalität
  - Verwaltung und Transport von Daten
  - Replikation und Archivierung

➔ **Transparenz als wichtigstes Ziel**



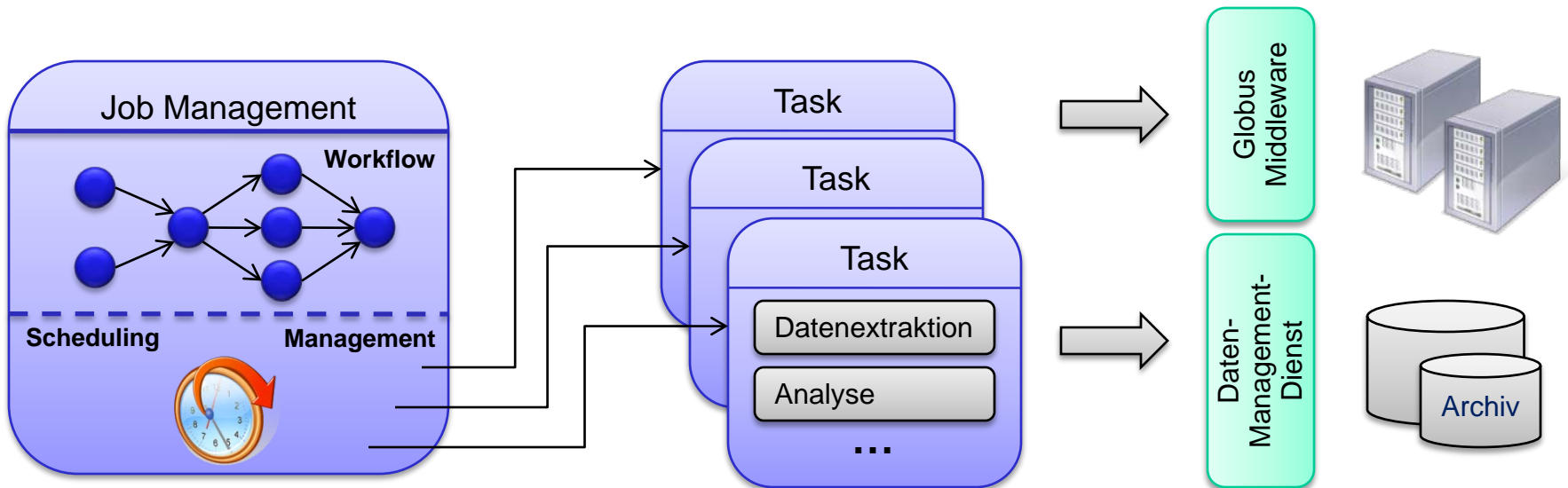
## Workflowmanagement und Scheduling

- Komplexe Abläufe:
  - Datenextraktion, Analyseläufe usw. als atomare Module
  - Darstellung von Abhängigkeiten zwischen Modulen
- Dynamische Planung im Sinne des Nutzers:
  - Einbeziehung verfügbarer Ressourcen
  - Koordination von Daten- und Rechenjobs



### Workflowmanagement-Dienst (WSS)

- Architektur spiegelt allgemeinen Workflow wider
  - Management-Dienst (eine Instanz - ein Workflow)
  - Task-Execution-Dienst (eine Instanz – ein Task)



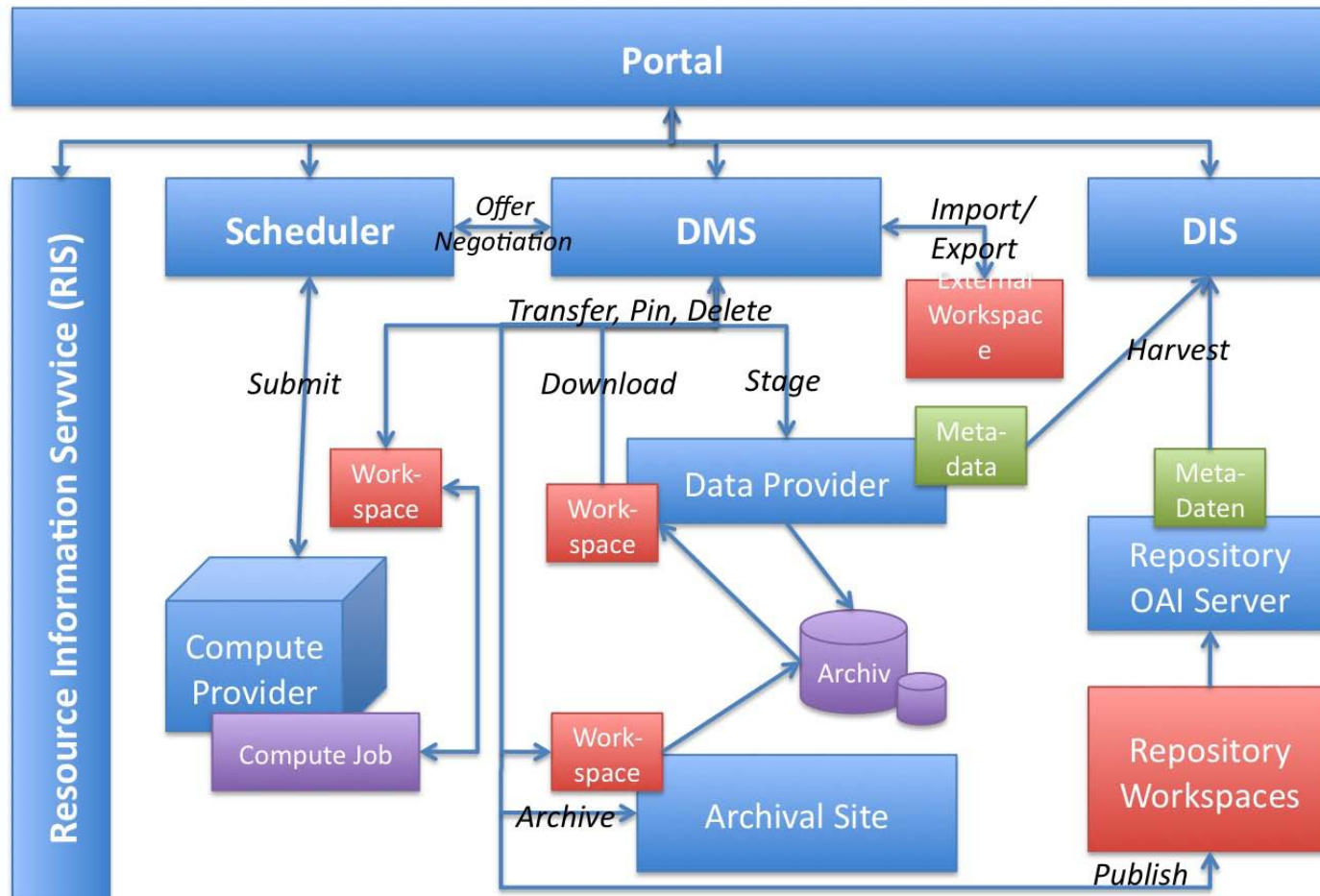
### Datenmanagement-Dienst (DMS)

- Viele verschiedene Aufgaben
  - Datenstaging bei Datenanbietern
  - Transfer von Daten- und Metadatenfiles
  - Publish von Zwischenergebnissen
  - Verwaltung von Festplattenspeicher
  - Koordination und Verhandlung zwischen Workflowmanagement und Datenprovidern

- Datenmanagementsoftware GNDMS
  - Generation N Data Management System
- Viele Dienste
  - Workspace-Manager DSpace
  - **DM-Ausführungs- und Verhandlungsdienst (GORFX)**
    - Transfer
    - Staging
    - Ingest/Publish
  - Dateninformationsdienst DIS (Metadatenharvesting)
  - Ressourceinformationsdienst RIS/MDS
- Viele Komponenten: DMS, Provider ...



C3Grid User  
Community



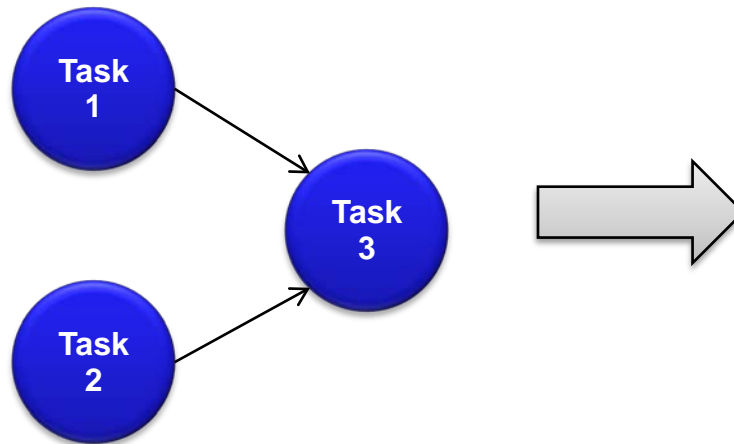
- Was ist nötig um Datenprovider zu werden?
  - Vorbereitung der bereitgestellten Daten und ISO-Metadaten
    - Harvesting durch Ressourceinformationssystem
  - Grid-Zertifikat für eigenen Server und lokale Grid-Nutzer
  - MDS-Eintrag im Ressourceinformationssystem RIS
- Welche Dienste müssen installiert werden?
  - Middleware: Globus-Toolkit 4.0.8
    - Firewallkonfiguration, DMS-Zertifikat autorisieren
  - GNDMS-Software vom ZIB für Globus
    - Install-Skript: Konfiguration von GNDMS als Datenprovider (DSpace + GORFX mit Transfer und Staging)
  - Implementation von min. zwei Shell-Skripten (nächste Folie)

- **Alle Skripte:**
  - Anfrage auf StdIn, Ergebnis auf StdOut
  - Dabei ggf. Dateien im Arbeitsverzeichnis erzeugen
- **Schätz-Skript:**
  - Wie lange wird es vor. dauern, die gegebene Datenanfrage zu bearbeiten?
- **Staging-Skript:**
  - Datenanfrage ausführen
- **Cancel-Skript (optional):**
  - Aufräumen im Fehlerfall

- Datenmanagement betrifft auch Rechenprovider
  - Nutzung des Workspacemanagement-Dienstes DSpace für die Verwaltung von Arbeitsverzeichnissen
- Installation von Workflow/Analyse-Skripten
- Globus Toolkit WS-GRAM zur Aufführung von Skripten durch das Workflowmanagement
- Demnächst: allgemeines Modul-Konzept zum Einbringen eigener Workflow-Skripte

### Entwicklung eigener Workflows

- Workflow Specification Language (WSL) ermöglicht eine XML-basierte Beschreibung
- Beschreibung repräsentiert den Workflowgraphen



```
<Workflow>
  <Task identifier="Task1">
    <jsd1:JobDefinition>
      ...
    </jsdl:JobDefinition>
  </Task>

  <Task identifier="Task2">
    <jsd1:JobDefinition>
      ...
    </jsdl:JobDefinition>
  </Task>

  <Task identifier="Task3">
    <Dependency operator="and">
      <Parent>Task1</Parent>
      <Parent>Task2</Parent>
    </Dependency>
    <jsd1:JobDefinition>
      ...
    </jsdl:JobDefinition>
  </Task>
</Workflow>
```

- Einblick in die eigentlich verborgene Technik
- Anforderungen und Umsetzung in C3Grid
  - Gesamte Architektur
  - Einige Dienste im Detail
- Technische Voraussetzungen zur Teilnahme:
  - Datenprovider
  - Rechenprovider
  - Workflowentwickler