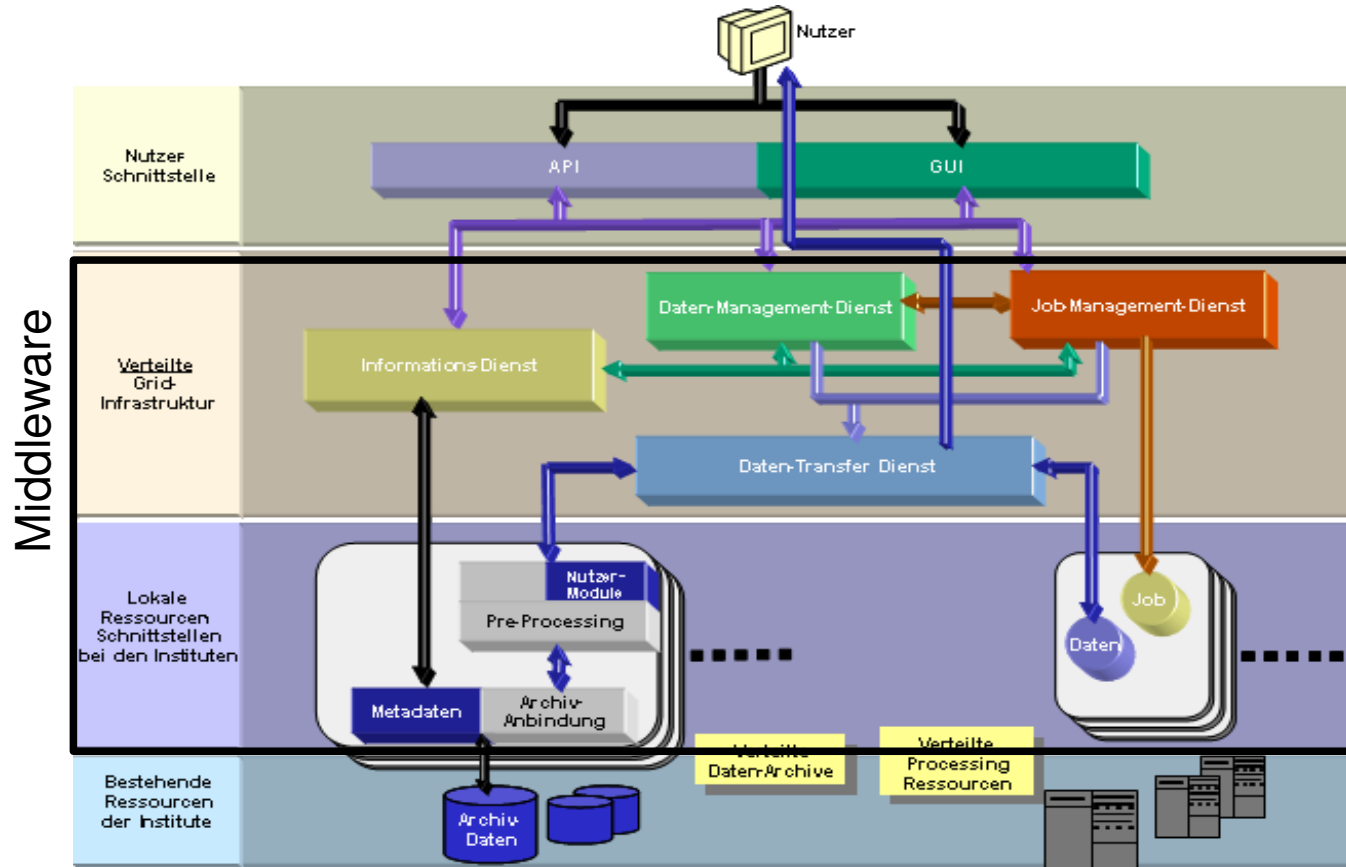
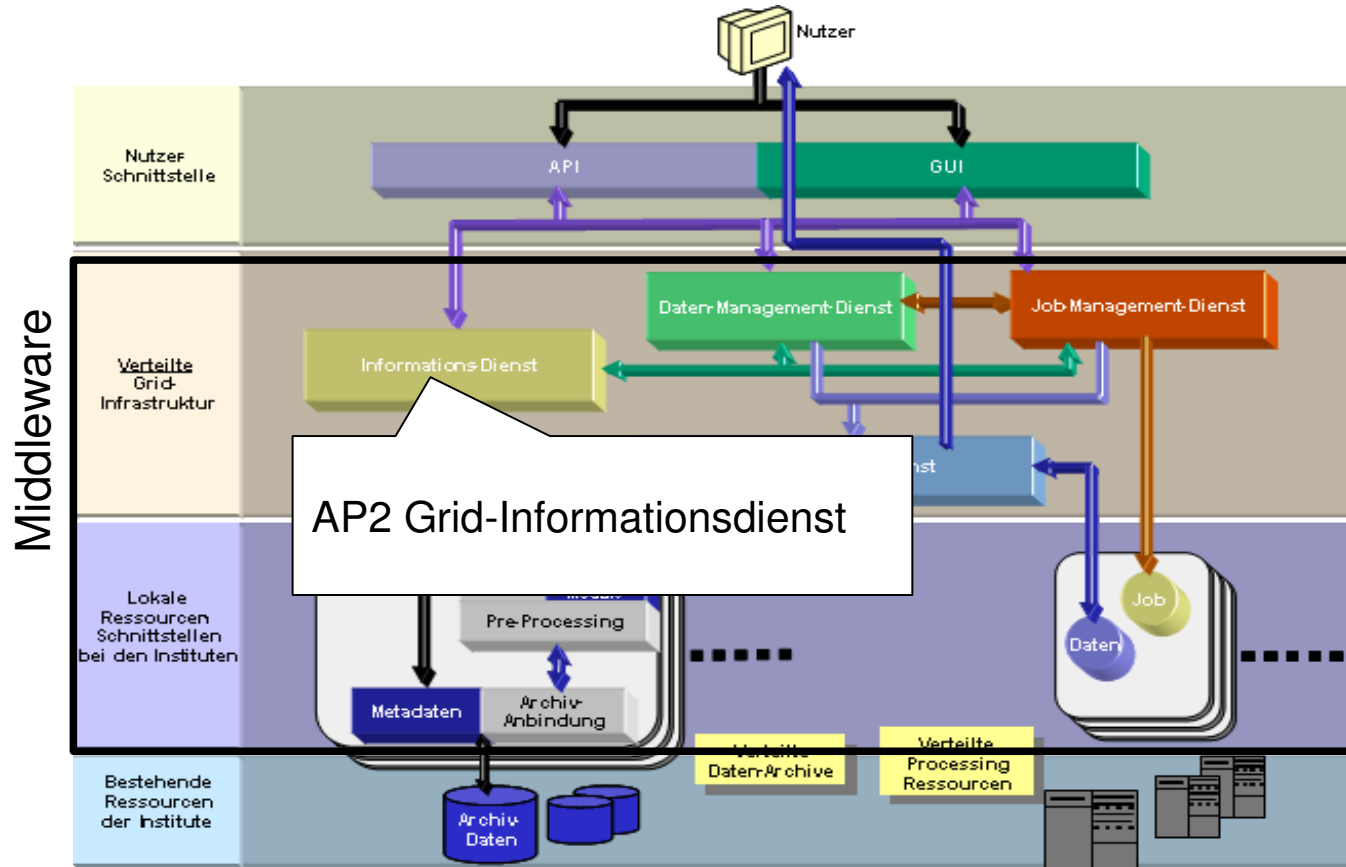


C3-Grid Middleware

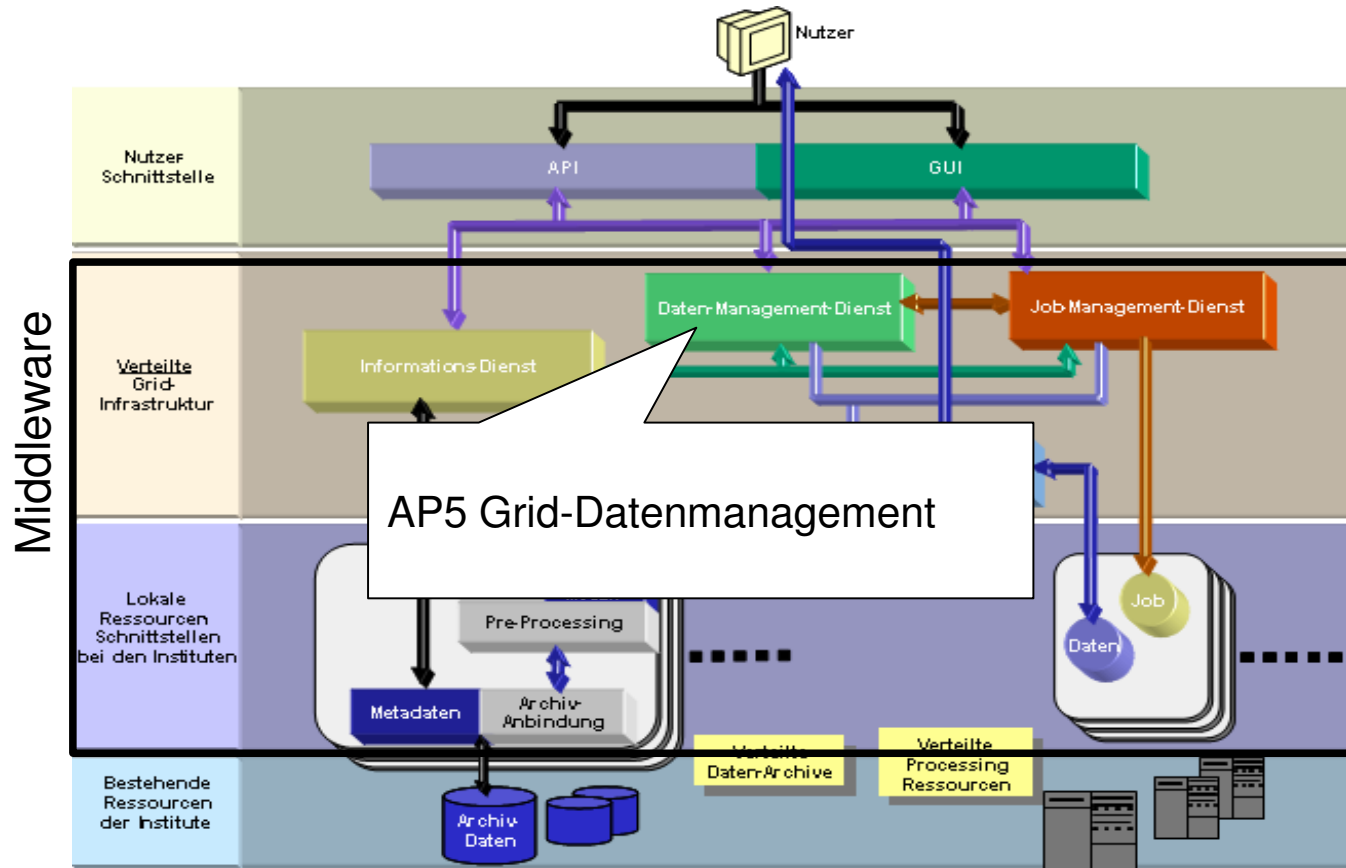
Anforderungen und Architekturskizze

- Ziele
 - Grid-weiter Zugriff auf Klima- und Erdsystemdaten aller teilnehmender Institute
 - Kein Zentralarchiv
 - Einfacher, transparenten Zugriff auf heterogene Datenquellen (Datenbanken, Archivsysteme,...)
 - Einheitliches Metadatenformat um Daten zu beschreiben
- Grobe Anforderungen an die Middleware
 - Datenreduktion, Pre-Processing vor Ort
 - Ausnutzung von Datenlokalität, Replikation, Caching
 - Job-Management
 - Ressourcen- und Metadaten-Informationssdienst

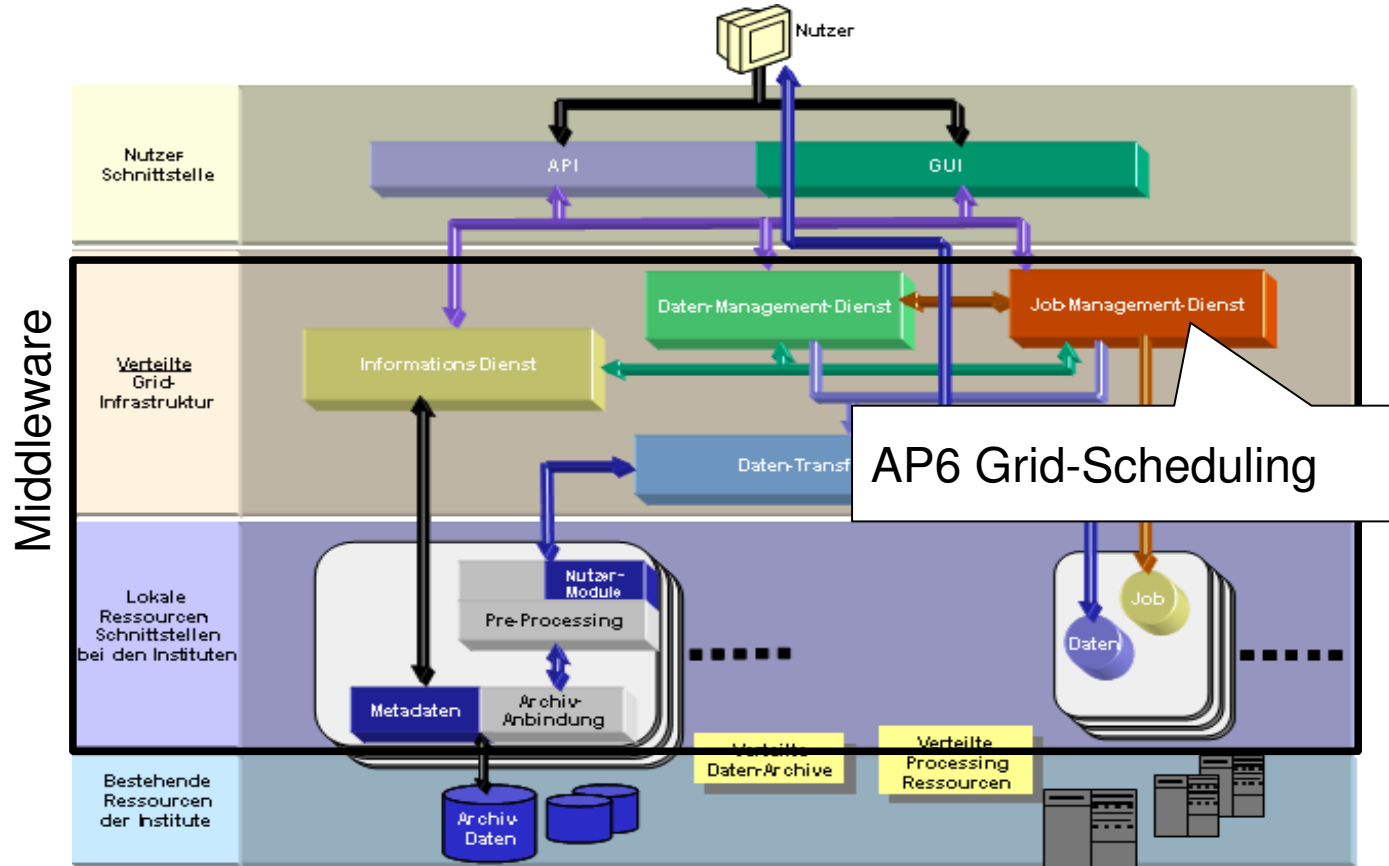




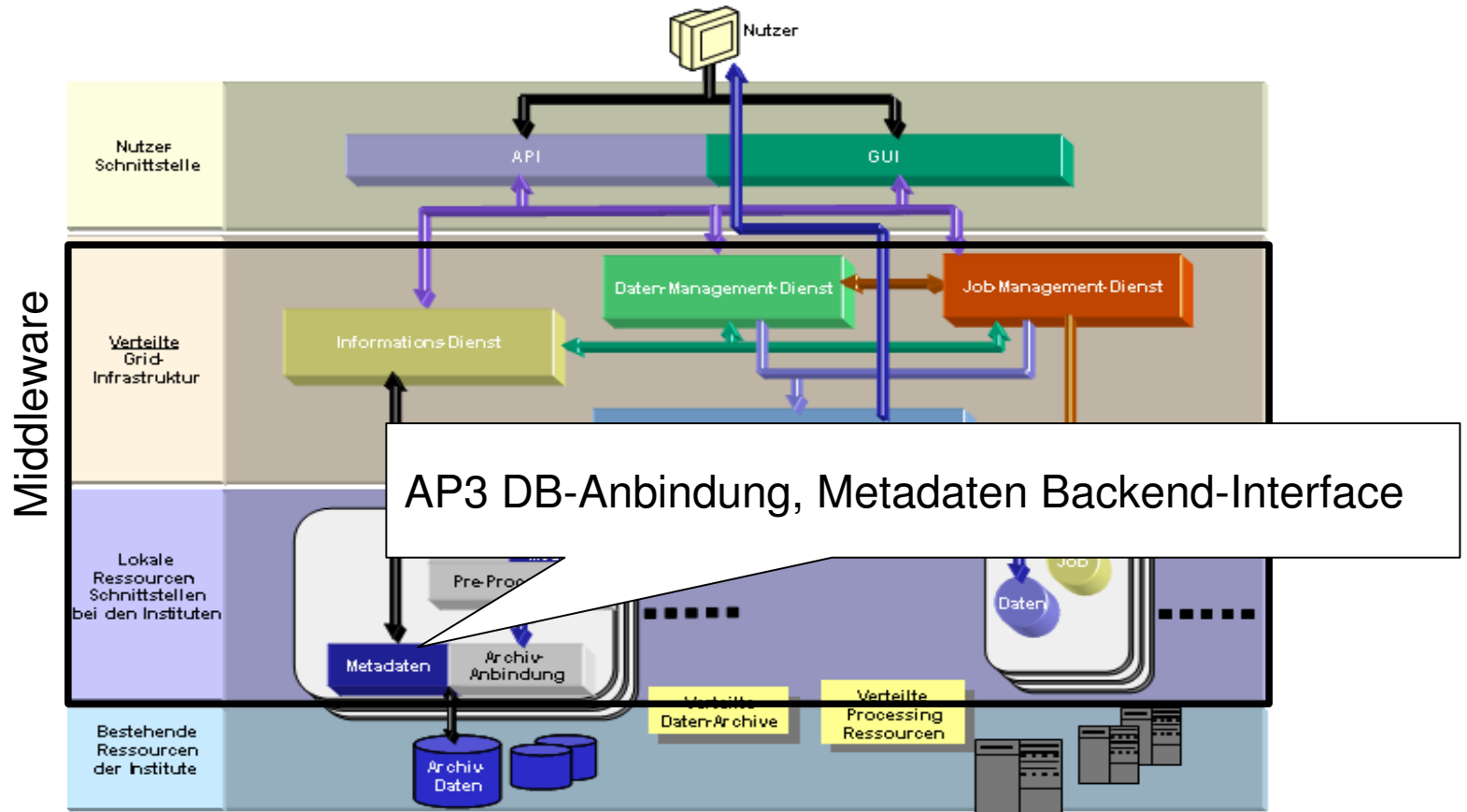
- Ressourcen-Informationdienst
 - Netzwerk, Speicher, Software, ...
 - (Umsetzung mit MDS von Globus Toolkit 4)
- Daten-Informationdienst
 - Suchdienst für Metadaten
 - Auch Suche nach geographischen Bereichen
 - Metadaten von primären und für das Grid erzeugten Daten
 - Standardisiertes Metadatenformat (ISO19xxx)
 - (Dazu spezielle C3-Arbeitsgruppe)
 - Fehlertoleranz durch verteilten Katalog
 - Pflege automatisch erzeugter Metadaten



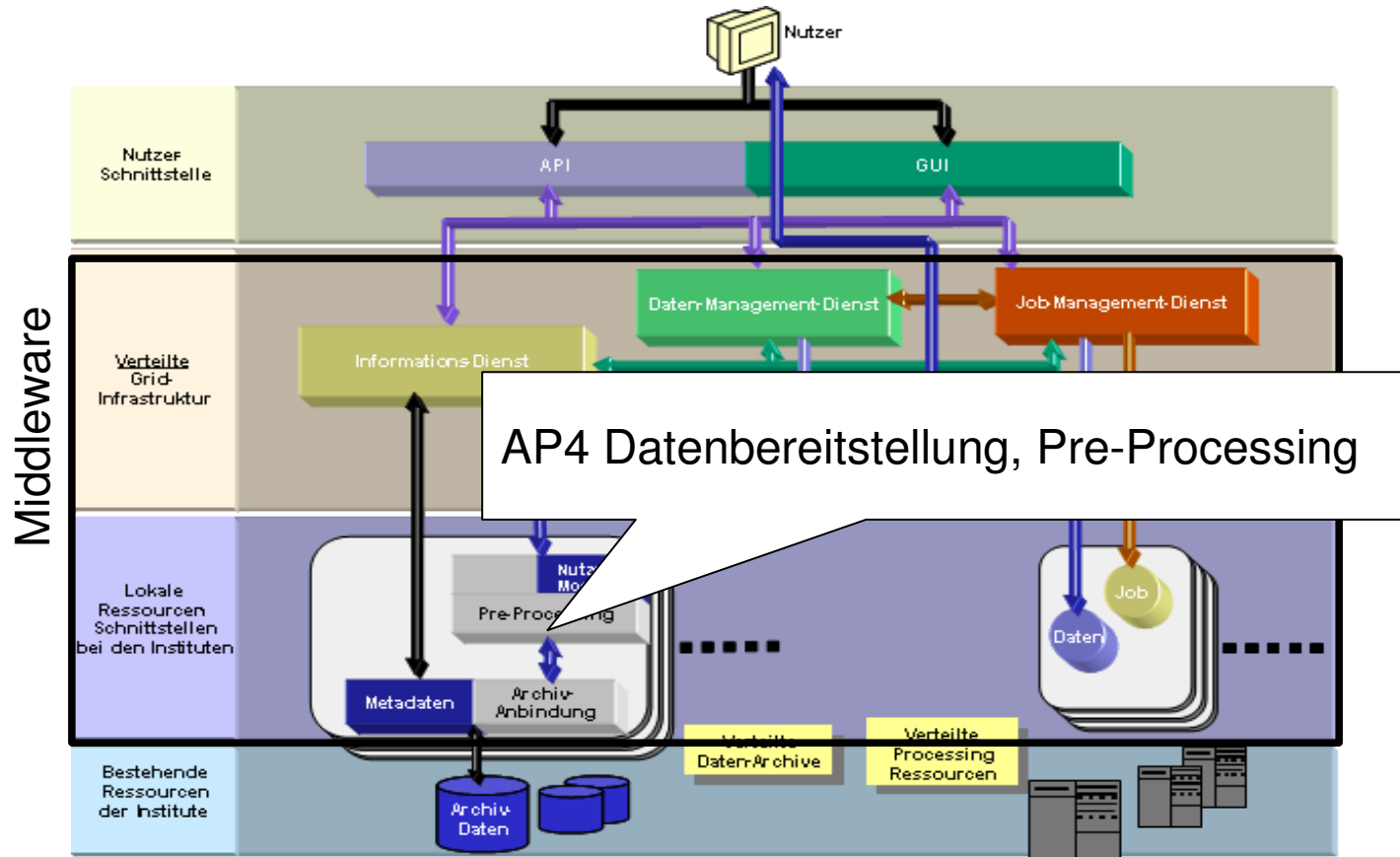
- Datenmanagement-Service
 - Bereitstellung von Daten aus den Repositories (mit AP4)
 - Verwalten von Replikaten gleicher Daten
 - Berücksichtigung von Kenngrößen
 - Netzwerkbandbreite, -topologie
 - Zeit/Datenvolumen für das Staging aus Archiven
 - Bereitstellung der Daten auf Anforderung des Schedulers (AP6)
 - Reservierung von Speicher
 - für einen Zeitraum A bis B auf Resource R
 - Bereitstellung von Daten
 - für einen Zeitraum A bis B auf Resource R



- Workflow Scheduler
 - Zuordnung von Aufgaben und Daten zu Rechnersystemen
 - Nutzen von Information über
 - Bereitstellungszeit, Compute-Ressourcen, Netzwerkbandbreiten, Datengrößen und -orten, Speichergrößen, ... (mit AP2)
 - Ziel: Reduktion des Transfervolumen durch Repository-nahe Datenreduktion (z.B. CERA-Daten im Petabyte-Bereich)
 - Datentransfer in Kooperation mit AP5



- Metadaten Backend-Interface
 - Einheitliche Schnittstelle zu lokalen Metadaten-Informationendiensten
 - Bietet Service an Daten-Informationdienst
 - Generierung von Metadaten aus Faktendaten



- Datenbereitstellung
 - Bereitstellung der Daten im vereinbarten Format
 - Bereitstellung zugehöriger Metadaten
 - Extraktion von Primärdaten aus Archiven
 - Reduktion der Datenmenge vor Ort
 - Pre-Processing mit bereits bestehenden Tools
 - Aggregation in der Datenbank

- Für GAT relevante Middleware?
 - Workflow Scheduler
 - Datenmanagement (größtenteils Eigenentwicklung, ggf. SRB auf Instituts-Ebene)
 - Preprocessing bei den Datenanbietern
- Welche GAT-Applikationen im C3-Grid?
 - Keine Neuentwicklung von Processing-Applikationen geplant
-> Jobs nicht „grid-aware“
- Java oder C++