Dresden University of Technology

Data Warehouse, Data Peers, Data Grid, …:

von Buzzwords zu substantiellen Forschungsfragen?

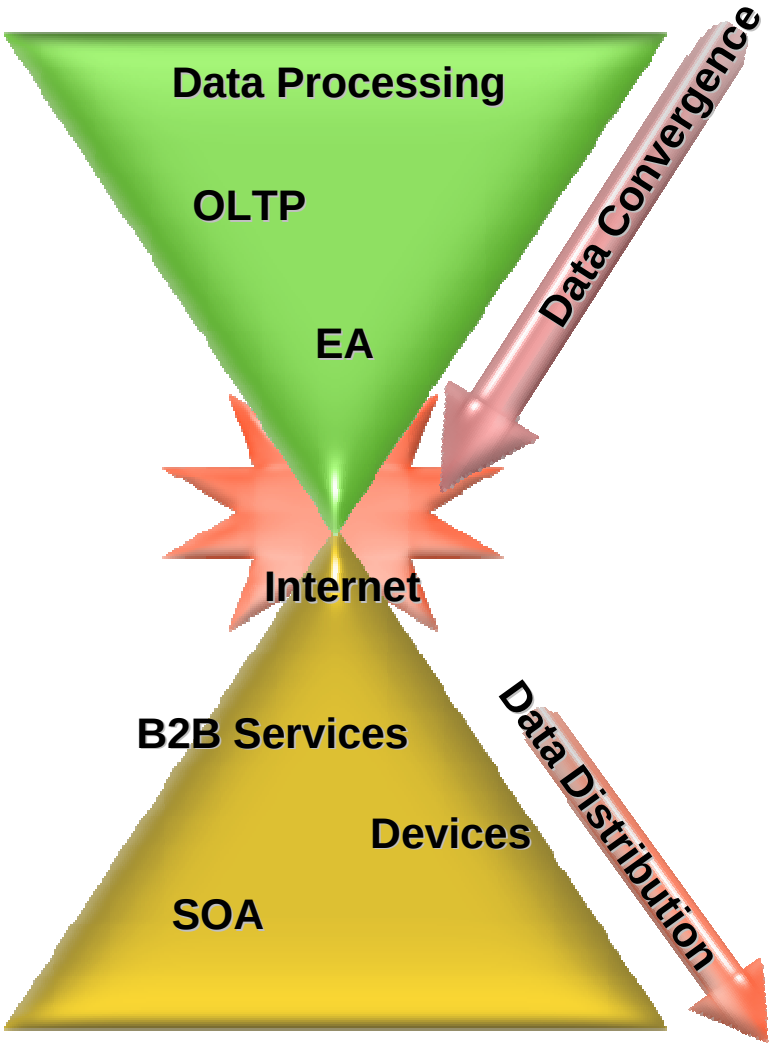
Wolfgang Lehner
Dresden University of Technology
Database Technology Group

Disclaimer

Dresden University of Technology

- **Ziel**
  - "Reverse" Talk (H. Garcia-Molina, SIGMOD2005)
- **… zur Diskussion stehende Fragestellungen**
  - Verlieren wir uns in dem Nachjagen von Buzzwords?
    oder
    Verpassen wir echte Research-Opportunities?
  - Was ist die Kernidee / Leistung aktuell diskutierter Informationsinfrastrukturen?
  - Stellen der berühmten Frage: "Are we polishing a round ball?"
    d.h.
    Wie hoch ist der Anteil an "Tooling", wo ist die „Novelty"?
- **… Warum der Titel?**
  - DWH-Infrastrukturen wurden vor 10 Jahren skeptisch hinterfragt...
  - "Grid-Service Technologie kommt in den Mainstream"
    (J.L. Encarnacao, Feldafingerkreis 2005)

… noch ein Schritt fundamentaler!



- **“The Data Monolith”**

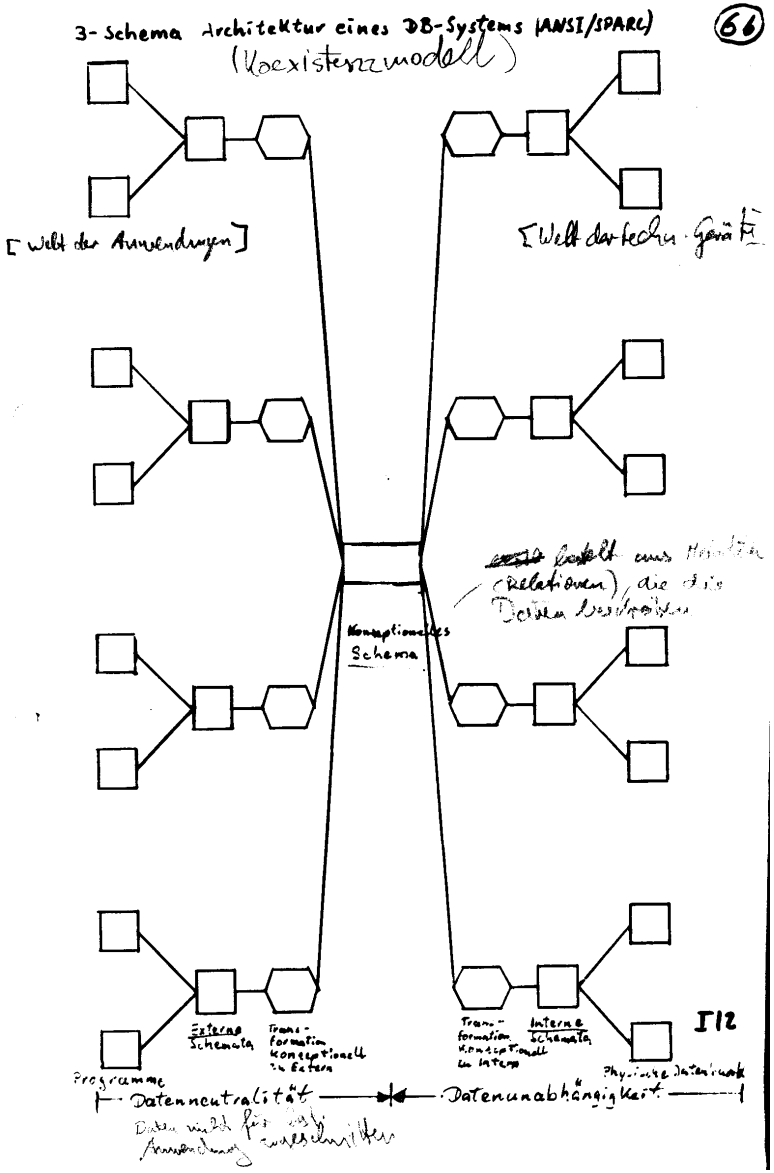
  (David Cambell)

  - ursprünglicher Anspruch des konzeptionellen Datenbankentwurfs: → Datenkonsolidierung auf Schemaebene
  - /Grid/SOA/B2B/… impliziert eine Datenverteilung
  - … die Datenbankwelt hat nicht (ausreichend) reagiert

- **… Konsequenz?**

… und noch einen Schritt abstrakter!

- **Anspruch der 3-Schema-Schichtenarchitektur nach ANSI/Sparc**
  - zentral: Forderung eines gemeinsamen Verständnisses (sowohl einer Sprache als auch einer Grammatik)
  - Transformationen zwischen den “Welten”



Vergleich: Data-Warehouse – Data-Grid

Dresden University of Technology

- **Data-Warehouse**
  - getrieben aus der Anwendung
  - große Nachfrage aus der Industrie
  - großer Impuls für die Wissenschaft
  - wichtigstes Ziel: Konsolidierung
    → pragmatische Lösung der "Integration per Kopie"
  - wichtige Nebeneffekte: Effiziente Umsetzung
    → Vielzahl von Lösungen

- **Data-Grid**
  - getrieben aus Standardisierungsgremien heraus
    (gibt es Anwendungen?)
    (was sind Anwendungsklassen)
  - Nachfrage aus dem Bereich des HPC/Metacomputing
  - aktuelle Arbeiten Fokussierung auf technische Integration
    → XML, WebServices and friends …
  - „re-inventing the wheel"

Foster (2003): Datenbanken werden nur für Verzeichnisdienste genutzt

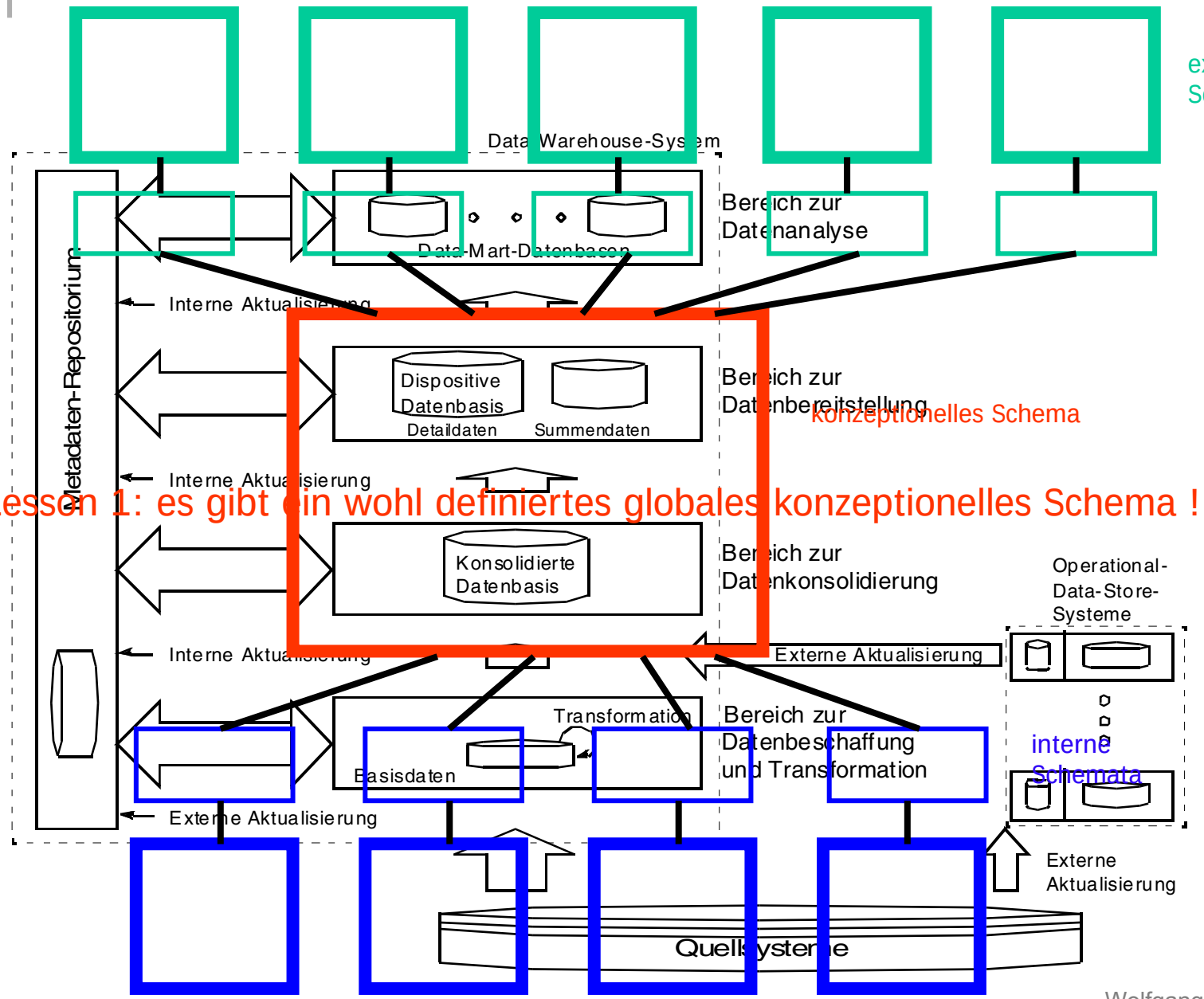
Data-Warehouse-Infrastruktur

Historisches

Dresden University of Technology

- **Definition (Bill Inmon)**
  - A data warehouse is a subject-oriented, integrated, time-varying, non-volatile collection of data in support of the management's decision-making process.
- **Definition (GI-AK Konzepte und Techniken des DWH)**
  - "Data Warehouse"
    - Physische Datenbank als integrierte Sicht auf (beliebige) Daten
    - Der Auswertungsaspekt (analyse-orientiertes Schema) steht im Mittelpunkt
    - Häufig, aber nicht notwendigerweise, findet eine Historisierung der Daten statt
  - "Data-Warehouse-System"
    - Informationssystem, bestehend aus einer Vielzahl von Datenbanken und Systemkomponenten, die obige Aufgaben erfüllen

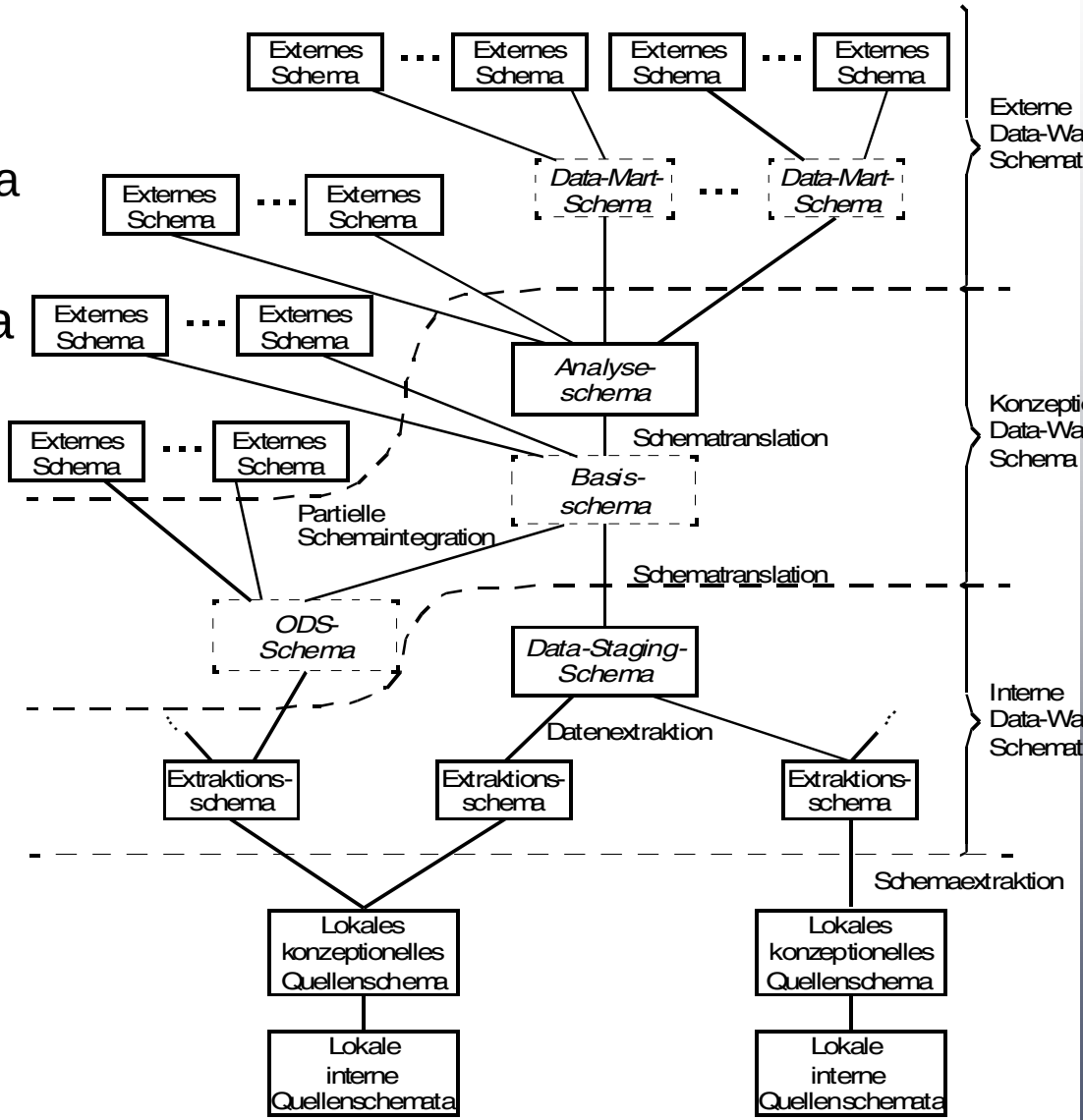
Bestandteile eines Data-Warehouse-Systems



Lesson 1: es gibt ein wohl definiertes globales konzeptionelles Schema !!!

Architektur eines Data-Warehouses

- Interne Schemata
  - Extraktionsschemata
  - Data-Staging-Schemata der Basisdaten
- Konzeptionelle Schemata
  - (Basisschema der konsolidierten DB)
  - Analyseschema der dispositiven DB
  - ODS-Schema
- Externe Schemata
  - Data-Mart-Schemata



Lesson 1: es gibt ein wohl definiertes globales konzeptionelles Schema !!!

Bewertung

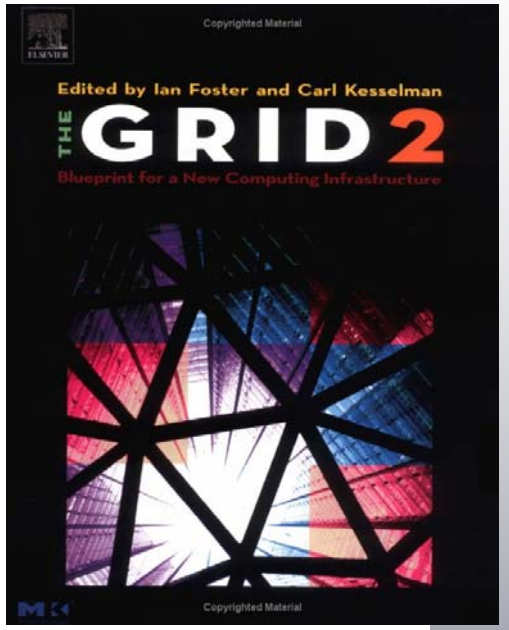
Dresden University of Technology

- **Überbrückung der semantischen Heterogenität**
  - Definition eines Ziel-Modells und Ziel-Schemas (Rolle von Daten: Stamm- und Bewegungsdaten)
  - Einführung von ETL
  - physische Datenintegration (Integration per Datenkopie)
  - tiefes Verständnis der unterschiedlichen Arten an Daten
    - Stammdaten / Dimensionsdaten /…
    - Bewegungsdaten / Faktdaten / …
  - Entwicklung geeigneter Modelle (multidimensionales Modell)
- **Systemtechnische Umsetzung**
  - Vielzahl technischer Entwicklung
  - massiver Einfluss auf wissenschaftliche Arbeiten

Lesson 2: es gibt ein grundlegendes Weltverständnis (Modell) !!!

Data-Grid-Infrastruktur

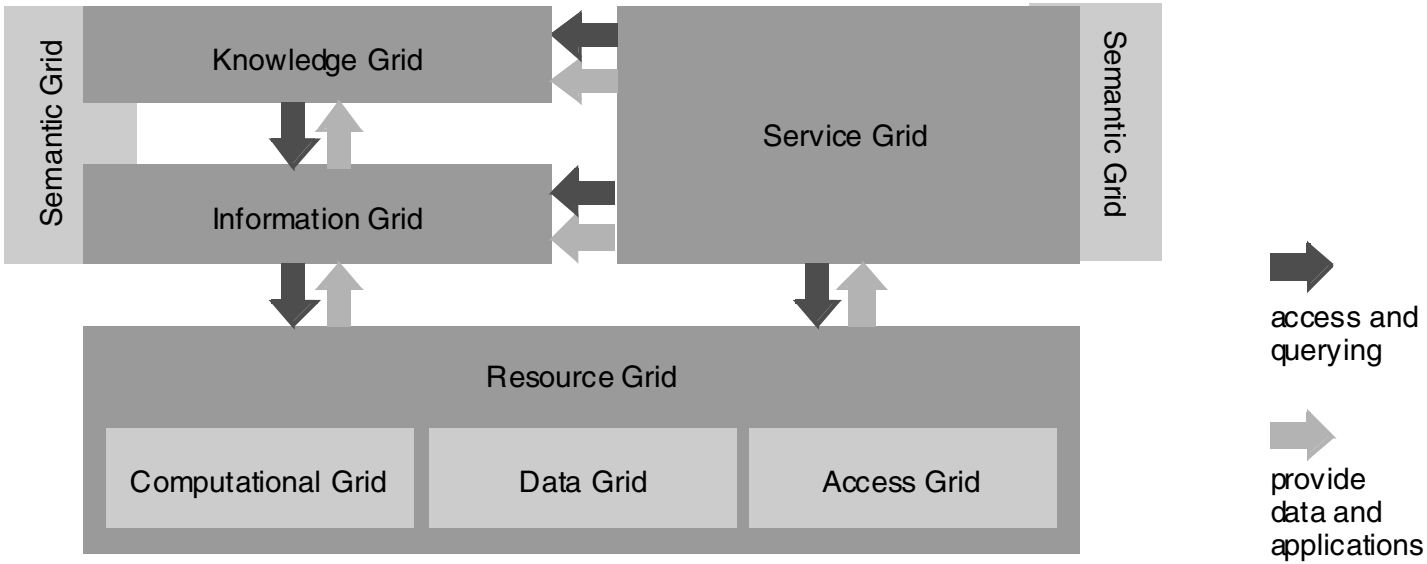
Grid-Computing: Taxonomy / Historie



- **Definition (I. Foster/C. Kesselmann, 1999)**
  - Gemeinsame Nutzung vernetzter Ressourcen
  - einfach und einheitliche Nutzung von Ressourcen
- **Virtualisierung**
  - IT-Leistungen sind verfügbar wie Strom, Wasser im Haushalt
    - Abrechnung, Sicherheit und Qualität, Verfügbarkeit
    - Ort der Leistungserbringung ist irrelevant
- **Anwendungen (??)**
  - getrieben aus dem Bereich HPC (MetaComputing)
  - Propagierung der Idee in vielfältige Anwendungsgebiete

Grid-Taxonomy

- Definition: Grid (Schlesinger, 2004)
  - Ein Grid ist eine skalierbare Hardware- und Softwareinfrastruktur, die durch Virtualisierung einen zuverlässigen, konsistenten, kostengünstigen, koordinierten und transparenten Zugriff auf verteilte, nicht zentral kontrollierte Ressourcen bietet, um Probleme in dynamischen, viele Institutionen umfassenden virtuellen Organisationen unter Einbeziehung von Qualitätsmerkmalen zu lösen.



Der Grid-Hype

Plattform vereint Semantic-Webservices mit verteilter Rechenleistung

EU-Projekt macht das Grid praxistauglich

Potsdam (rr) – 2006 wird in dem EU-Projekt Adaptive Services Grid (ASG) der Prototyp einer Softwareplattform fertig sein, der zwei Zukunftstechnologien verknüpft: das Semantic Web und das Grid Computing. Profit schlagen daraus alle wissensintensiven Geschäftsprozesse.

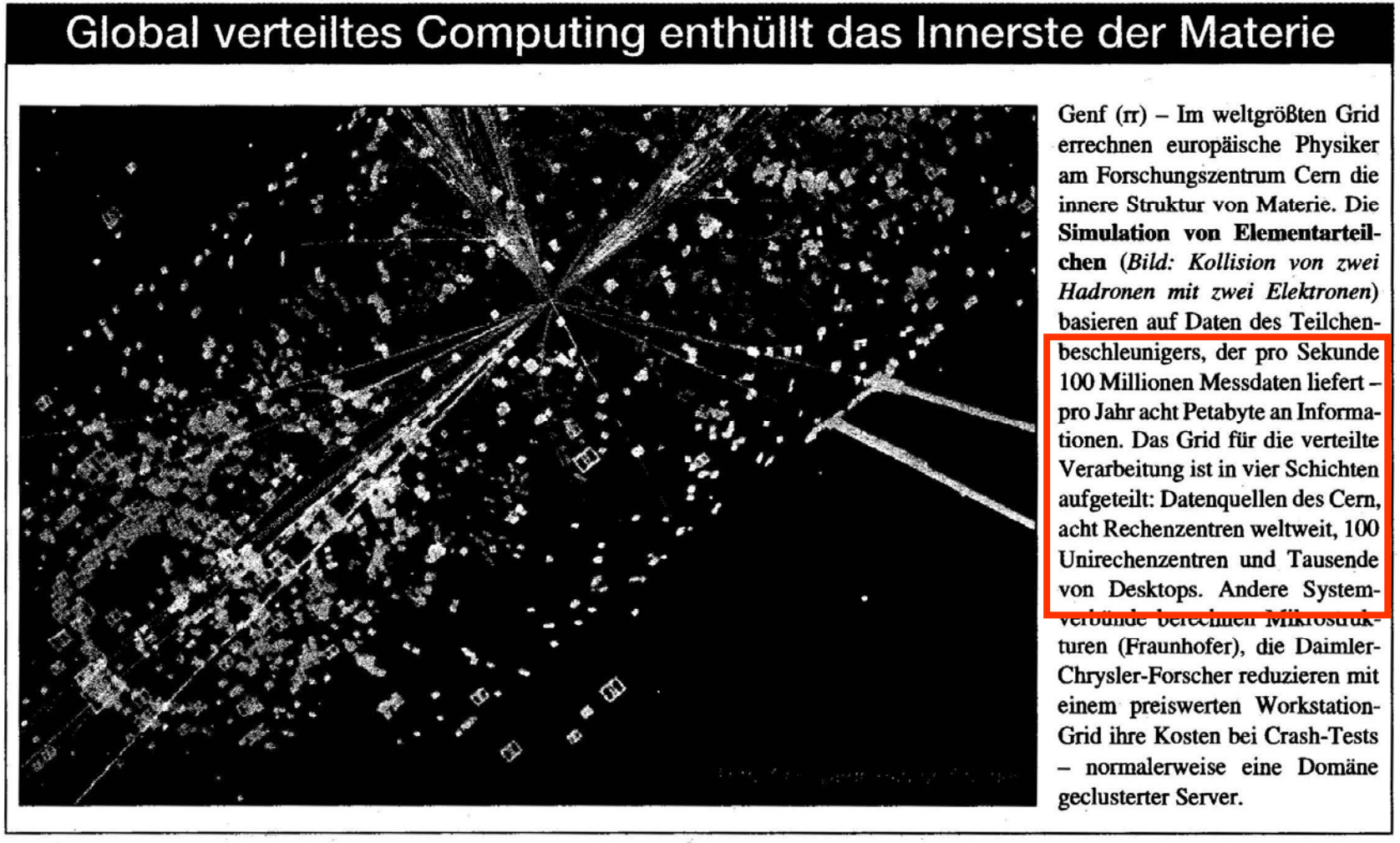
Der Ansatz des 11,5 Millionen Euro schweren Projekts ist laut Koordinator Professor Mathias Weske vom Hasso-Plattner-Institut (HPI) in Potsdam praxisorientiert: „Im Semantic Web wird ein Inhalt mit einer präzise definierten Bedeutung versehen, die ein Computer versteht und interpretieren kann – der Mensch kann effizienter mit dem Rechner kommunizieren." Und seine Anfrage wird als Service formuliert und in einem verteilten Rechnernetz beantwortet. „Die EU hat über Jahre die Grid-Infrastruktur stark gefördert", berichtet HPI-Professor Andreas Polze, der die zentrale Infrastruktur verantwortet. „Jetzt rückt die vertikale Integration in den Vordergrund."

21 Partner aus sechs EU-Ländern und Australien arbeiten an der ASG-Plattform, die über das Internet Softwarefunktionen miteinander verknüpft und weltweit Rechenleistung nutzbar macht. Polze liefert ein Beispiel: „Sollen in der Geometrie Grafen transformiert werden, will niemand über Webservices-Schnittstellen reden. Mit der Semantic-Web-Technik wird die domänenspezifische Anfrage formuliert." Diese werde in der zweiten Abstraktionsebene als Workflow reformuliert „und dann in der dritten im Grid abgearbeitet".

Die Sicherheit der Daten im Grid Computing ist nach Polzes Einschätzung eines der schwierigsten Themen. Die Industriepartner haben sich ihre Gedanken gemacht: „Siemens und Daimler-Chrysler inszenieren in ihrem Intranet ein Application Service Provisioning – sie konsolidieren ihre Serverlandschaften und setzen dafür die ASG-Technologie ein." Die drei beteiligten Telecomunternehmen aus Norwegen, Polen und Frankreich leiden laut Polze weniger unter Berührungsängsten: „Sie denken über internetweite Applikationen nach, weil sie das jetzt schon machen und ihren Sicherheitsverfahren vertrauen."

Der Mittelstand kann, so der Middleware-Spezialist, zweifach partizipieren: „In das Konsortium dürfen mittelfristig weitere Partner einsteigen oder sie kollaborieren als Partner und wir machen unsere ASG-Ressourcen verfügbar."

Der Grid-Hype (2)



Global verteiltes Computing enthüllt das Innerste der Materie

Genf (rr) – Im weltgrößten Grid errechnen europäische Physiker am Forschungszentrum Cern die innere Struktur von Materie. Die **Simulation von Elementarteilchen** (*Bild: Kollision von zwei Hadronen mit zwei Elektronen*) basieren auf Daten des Teilchenbeschleunigers, der pro Sekunde 100 Millionen Messdaten liefert – pro Jahr acht Petabyte an Informationen. Das Grid für die verteilte Verarbeitung ist in vier Schichten aufgeteilt: Datenquellen des Cern, acht Rechenzentren weltweit, 100 Unirechenzentren und Tausende von Desktops. Andere Systemverbünde berechnen Mikrostrukturen (Fraunhofer), die Daimler-Chrysler-Forscher reduzieren mit einem preiswerten Workstation-Grid ihre Kosten bei Crash-Tests – normalerweise eine Domäne geclusterter Server.

Data Grid: just storage and processing of massive data volume ???

Der Grid-Hype (3)

Deutsche Unternehmen liegen bei Basisstandards und Forschung vorne

Fehlendes Verständnis bremst Firmen-Grids

Brüssel (mr) – Die Akzeptanz von Grid-Infrastrukturen in den Unternehmen steigt, das offenbart der Grid-Index 2005. Aber häufig fehlt das Wissen, wie die verteilte Rechenleistung intern einzusetzen ist.

In nur neun Monaten ist das Grid-Barometer des Marktforschungsunternehmens Quocirca in Europa von 3,1 auf 4,4 Punkte gestiegen *(siehe Grafik)*. Es gibt den Grad des Verständnisses und der Nutzung der Technologie in Unternehmen wieder.

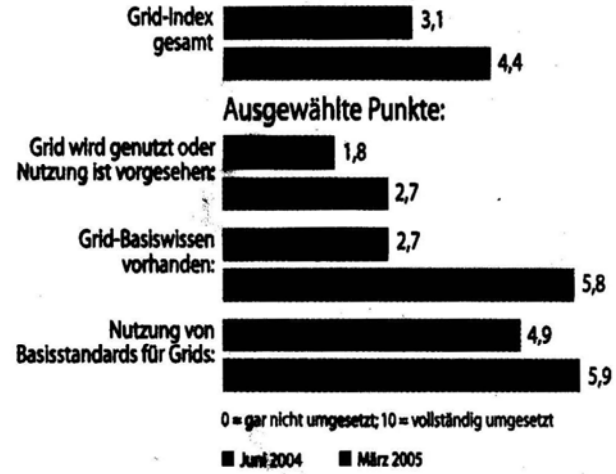
Trotz dieser positiven Tendenz existiert noch ein deutliches Wahrnehmungsdefizit. Selbstkritisch räumt Bernd Kosch, Direktor des Normierungsgremiums Enterprise Grid Alliance (EGA) und Vizepräsident von Fujitsu Siemens Computers (FSC) ein, dass es sogar derzeit an einer grundlegenden Definition des Begriffs Grid im Business-Umfeld mangele. Bislang habe hier jeder Hersteller sein eigenes Süppchen gekocht, inzwischen seien aber deutliche Tendenzen einer Zusammenarbeit erkennbar. So diskutiert das EGA-Board seit zwei Wochen ein erstes Referenzmodell, das im Sommer veröffentlicht werden soll.

Kosch sieht wegen der Wissenslücke die IT-Branche in der Bringschuld. „Die Anwender kommen nicht zu uns, wir müssen aktiv auf sie zugehen und ihnen die konkreten Vorteile der Technologie erklären." Erkennen sie erst einmal die Vorzüge, dann setzen sie relativ schnell Grids ein. Dies bestätigt Gianluca Moretto, CIO des italienischen Sozialversicherers Fondazione Enasarco: „Wir sparen nicht nur im laufenden Betrieb rund 30 Prozent der Hardwarekosten durch eine bessere Auslastung unserer Maschinen. Wichtiger war für uns, dass wir in Zukunft unsere Rechnerlandschaft preiswert auch mit heterogenen Servern erweitern können." Diese Rechner will Moretto über lediglich eine Verwaltungskonsole administrieren. Im Gegensatz zu einem Cluster braucht das Grid keine homogene Struktur. Dadurch verringere sich zudem der Personalaufwand.

Ausdrücklich gelobt werden deutsche Firmen in dem von Oracle gesponsorten Grid-Index: Weltweit sind sie führend bei der Nutzung von Basisstandards wie IP (Internet Protocol) oder Webservices und könnten Grid-Technik schnell umsetzen. Auch bei der Forschung spielen sie vorne mit, wie Wolfgang Bosch, EU-Referatsleiter Grid-Research betont – aber bei der Nutzung halten sie sich stärker zurück als die skandinavischen Länder. **Seite 8**

Basiswissen ist vorhanden

Die Entwicklung des europäischen Grid-Index von Juni 2004 bis März 2005 auf einer Skala von 0 bis 10



Brüssel (mr) – In den vergangenen neun Monaten haben das **Verständnis und die Nutzung** der Grid-Technologie deutlich zugenommen, wie der von dem Marktforschungsunternehmen Quocirca erstellte Grid-Index zeigt. Weltweit wurden dabei über 1350 IT-Leiter von mittelgroßen und großen Unternehmen befragt. Besonders stark gewachsen ist demnach im Vergleich zum Vorjahr das Basiswissen über die Technik. Dieses korreliere aber häufig nicht mit den spezifischen Anforderungen der Unternehmen, sondern eher mit denen im wissenschaftlichen Umfeld, wie Andrew Sutherland, Technologiechef bei Oracle Europa, erläutert. Damit erklärt er auch das prozentual geringere Wachstum bei der Nutzung der grundlegenden Standards auf dem alten Kontinent.

Quelle: Quocirca **COMPUTER ZEITUNG 15/2005**

Der Grid-Hype (4)

Dresden University of Technology

Antidot gegen "Grid-Hype"

- Grid-Technologie eröffnet keine „grundsätzlich" neuen Möglichkeiten
- Verteiltes Rechnen, Meta-Computing, Föderierung von Datenarchiven, sichere Kommunikation etc. gab es auch vorher

Aber:

- Grid-Technologie zielt auf Vereinfachung:
  - Standard-Services statt vielfache Eigenentwicklungen
  - Flexibilität statt „festverdrahteter" Lösungen
- Viele Anwendungen erst dadurch ökonomisch sinnvoll

Data on the Grid (Patrick Dantressangle, 2005)

Dresden University of Technology

- Resource sharing Grid
  - Data forms an important part of the Grid
- Data on Grids
  - May be geographically distributed
  - Storage technology and formats are not homogeneous
  - Need to dynamically bind to data sources on demand
- True virtualised data resources
  - Discoverable through published metadata
  - Robust against variations in standards
  - Easy to aggregate, federate and manage

Data on the Grid

Data on the Grid

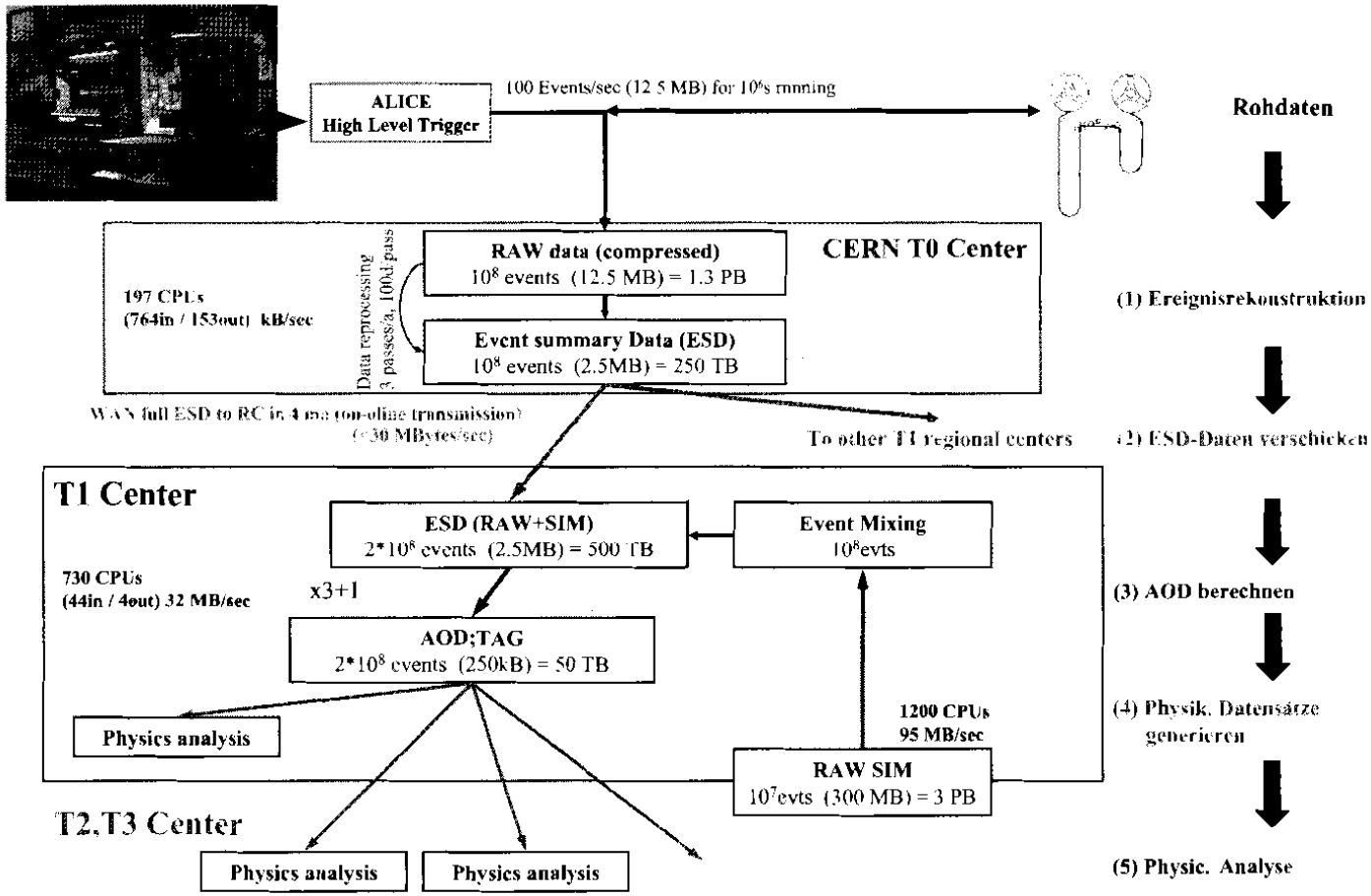


Abb. 3: Die Offline-Datenhierarchie am Beispiel des ALICE-Experiments (AOD – Analysis Object Data – »reduzierte« Datensätze für die Analyse). Es wurden hier Prozessoren mit einer Leistung von 4000 SPECint2000 angenommen.

Definition “Data-Grid”: Versuch 1

Google

Web Bilder Groups Verzeichnis News Froogle Neu!

"data grid" Suche Erweiterte Suche Einstellungen

Suche: Das Web Seiten auf Deutsch Seiten aus Deutschland

Web Ergebnisse **1** - **10** von ungefähr **208.000** für **"data grid"**. (**0,05** Sekunden)

The DataGrid Project - [Diese Seite übersetzen]
The DataGrid Project. DataGrid is a project funded by European Union. The objective is to build the next generation computing infrastructure providing ...
eu-datagrid.web.cern.ch/eu-datagrid/ - 10k - Im Cache - Ähnliche Seiten

Particle Physics **Data Grid** - [Diese Seite übersetzen]
... The Particle Physics **Data Grid** Collaboratory Pilot (PPDG) is developing and ... keenly aware of the need for **Data Grid** services to enable the worldwide ...
www.ppdg.net/ - 19k - Im Cache - Ähnliche Seiten

www.cacr.caltech.edu/ppdg/ - [Diese Seite übersetzen]
1k - Im Cache - Ähnliche Seiten

Avaki : Home - [Diese Seite übersetzen]
Avaki Corporation provides software that helps corporate, R&D, and IT leaders in a variety of industries improve access and integration between company ...
www.avaki.com/ - 32k - Im Cache - Ähnliche Seiten

Avaki : Products : Avaki Ell Plus - [Diese Seite übersetzen]
Avaki Ell Plus. Avaki is enterprise information integration (Ell) software that streamlines integration of data from many distributed sources while ...
www.avaki.com/products/ - 24k - Im Cache - Ähnliche Seiten

Globus: The **Data Grid** - [Diese Seite übersetzen]
... The Globus **Data Grid** Effort. "Access to distributed data is typically as ... The Globus Project's **data grid** effort attempts to identify, prototype, ...
www.globus.org/datagrid/ - 15k - Im Cache - Ähnliche Seiten

Data Grid Control Component by Jpowered - [Diese Seite übersetzen]
The java **Data Grid** Control applet enables the display of data in rows & columns in java & web applications. Powerful features include Fast Sorting, ...
www.jpowered.com/data_grid_control/ - 20k - Im Cache - Ähnliche Seiten

DIDC **Data Grid** work - [Diese Seite übersetzen]
... to develop tools which allow applications to efficiently access the "**Data Grid**". ... The Combustion Cooridor Project; The Particle Phyics **Data Grid** ...
www-didc.lbl.gov/data_grid/ - 3k - 11. Apr. 2005 - Im Cache - Ähnliche Seiten

NERC **Data Grid** - [Diese Seite übersetzen]
... Try the NDG Discovery Service NDG portal. Version 0.1 NEW. NERC DataGrid Version 0.1 Release of Tools, Source Code and Documentation. Overview ...
ndg.badc.rl.ac.uk/ - 11k - Im Cache - Ähnliche Seiten

Anzeigen

ActiveX **Grid** Controls
Grid-Controls namhafter Hersteller! Testversionen, Infos, Beispiel-Code
www.zoschke.com

AxpDataGrid for ASP.NET
HTML Edit, Paging, Sorting, Scroll SQL Server, Access, Oracle
www.axezz.com

itGrid - a superior **grid**
Small, very fast, reliable and easy to use. Download A Free Demo Now.
www.it-partners.com

Unique Web **Data Grid**
Discover Power & High Performance ASP, ASP.NET, JSP, Java. Free Tria
www.uolweb.com

Grid Komponenten
NET, ActiveX, DLL, VCL, MFC Vergleichen, evaluaieren, prüfen
www.componentsource.com

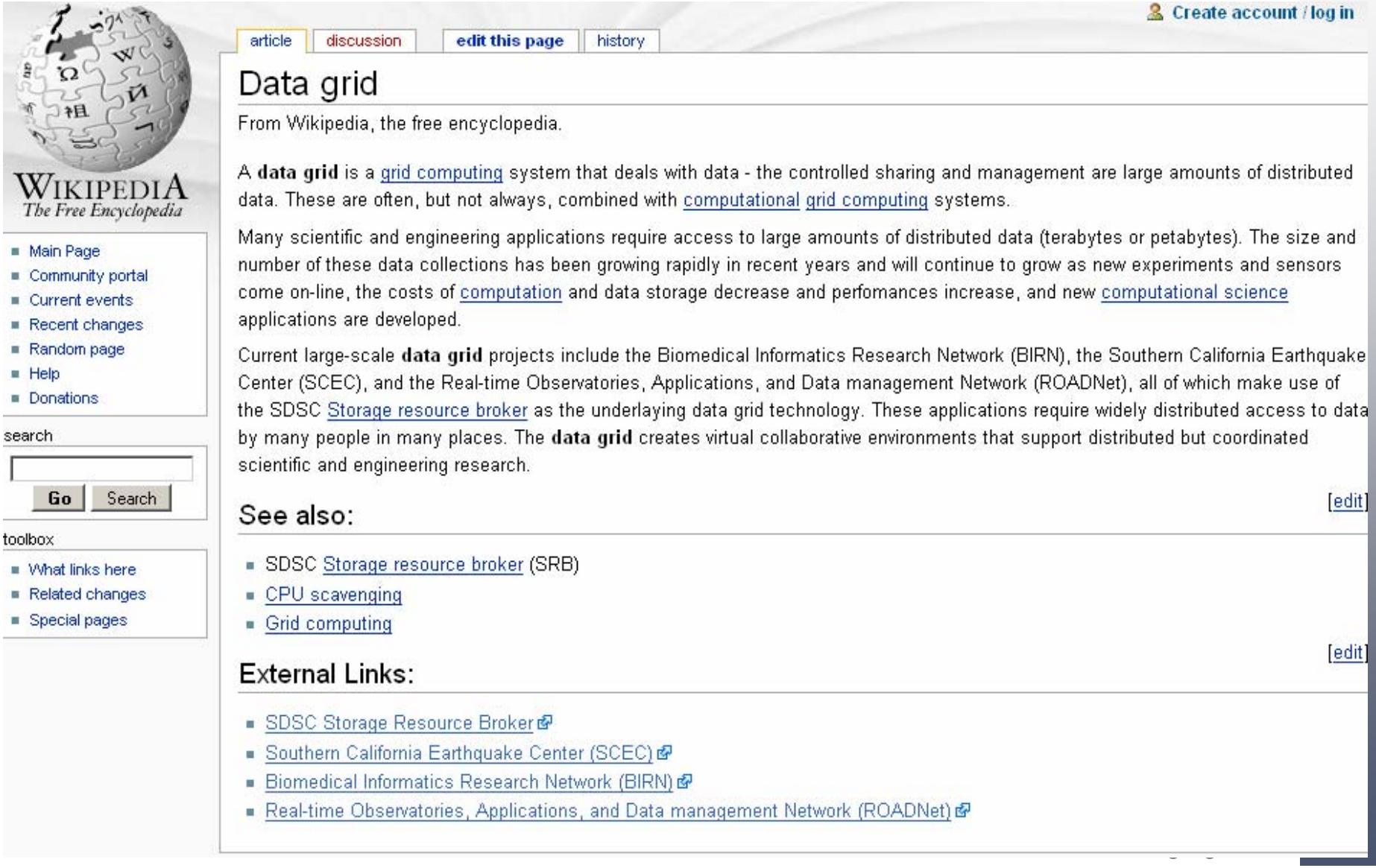
DbNetGrid (ASP.NET/ASP)
Link, sort, search, edit, print, copy, export and much more
www.dbnetgrid.com

Definition “Data-Grid”: Versuch 1

- **… try Google**
  - 208.000 entries
  - the top hits
    - Data Grid at Cern (EU project)
    - Particle Physics
    - Avaki
    - Globus
  - … seems to be the storage structure of high performance computing
    - high data volume
    - high data rate

Definition “Data-Grid”: Versuch 2

- … Wikipedia



WIKIPEDIA
The Free Encyclopedia

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help
- Donations

search

Go | Search

toolbox

- What links here
- Related changes
- Special pages

Create account / log in

article | discussion | edit this page | history

Data grid

From Wikipedia, the free encyclopedia.

A **data grid** is a grid computing system that deals with data - the controlled sharing and management are large amounts of distributed data. These are often, but not always, combined with computational grid computing systems.

Many scientific and engineering applications require access to large amounts of distributed data (terabytes or petabytes). The size and number of these data collections has been growing rapidly in recent years and will continue to grow as new experiments and sensors come on-line, the costs of computation and data storage decrease and perfomances increase, and new computational science applications are developed.

Current large-scale **data grid** projects include the Biomedical Informatics Research Network (BIRN), the Southern California Earthquake Center (SCEC), and the Real-time Observatories, Applications, and Data management Network (ROADNet), all of which make use of the SDSC Storage resource broker as the underlaying data grid technology. These applications require widely distributed access to data by many people in many places. The **data grid** creates virtual collaborative environments that support distributed but coordinated scientific and engineering research.

See also: [edit]

- SDSC Storage resource broker (SRB)
- CPU scavenging
- Grid computing

External Links: [edit]

- SDSC Storage Resource Broker
- Southern California Earthquake Center (SCEC)
- Biomedical Informatics Research Network (BIRN)
- Real-time Observatories, Applications, and Data management Network (ROADNet)

Definition “Data-Grid”: Versuch 3

- **The Globus Data Grid Effort**
  - *"Access to distributed data is typically as important as access to distributed computational resources."*
  - Distributed scientific and engineering applications often require access to large amounts of data (terabytes or petabytes)
  - Future applications: access to widely distributed data (For example, access in many places by many people, virtual collaborative environments, etc.)
  - The Globus Project's data grid effort attempts to identify, prototype, and evaluate the key technologies required to support data grids for scientific and engineering collaborations.
  - The Globus Project focuses on challenges associated with collaborative science and engineering spanning multiple organizations. It is our intent to offer a generic data grid infrastructure in the form of core data transfer services and generic data management libraries. These fundamental building blocks can then be used in a variety of interesting ways to build systems and applications for specific end users.

The Globus Data Grid Effort: Software

Dresden University of Technology

- **Main Strategy**
  - The Globus Project is currently engaged in defining and developing the following core capabilities which we believe will be necessary in order to build data grids for scientific collaborations.
- **GridFTP**
  - A high-performance, secure, robust data transfer mechanism
- **Globus Replica Catalog**
  - A mechanism for maintaining a catalog of dataset replicas.
- **Globus Replica Management**
  - A mechanism that ties together the Replica Catalog and GridFTP technologies, allowing applications to create and manage replicas of large datasets.
- **Experimental Projects**
  - Earth Systems Grid
  - European DataGrid
  - GriPhyN
  - Network for Earthquake Engineering Simulation
  - Particle Physics Data Grid

Definition “Data-Grid”: Versuch 4

Dresden University of Technology

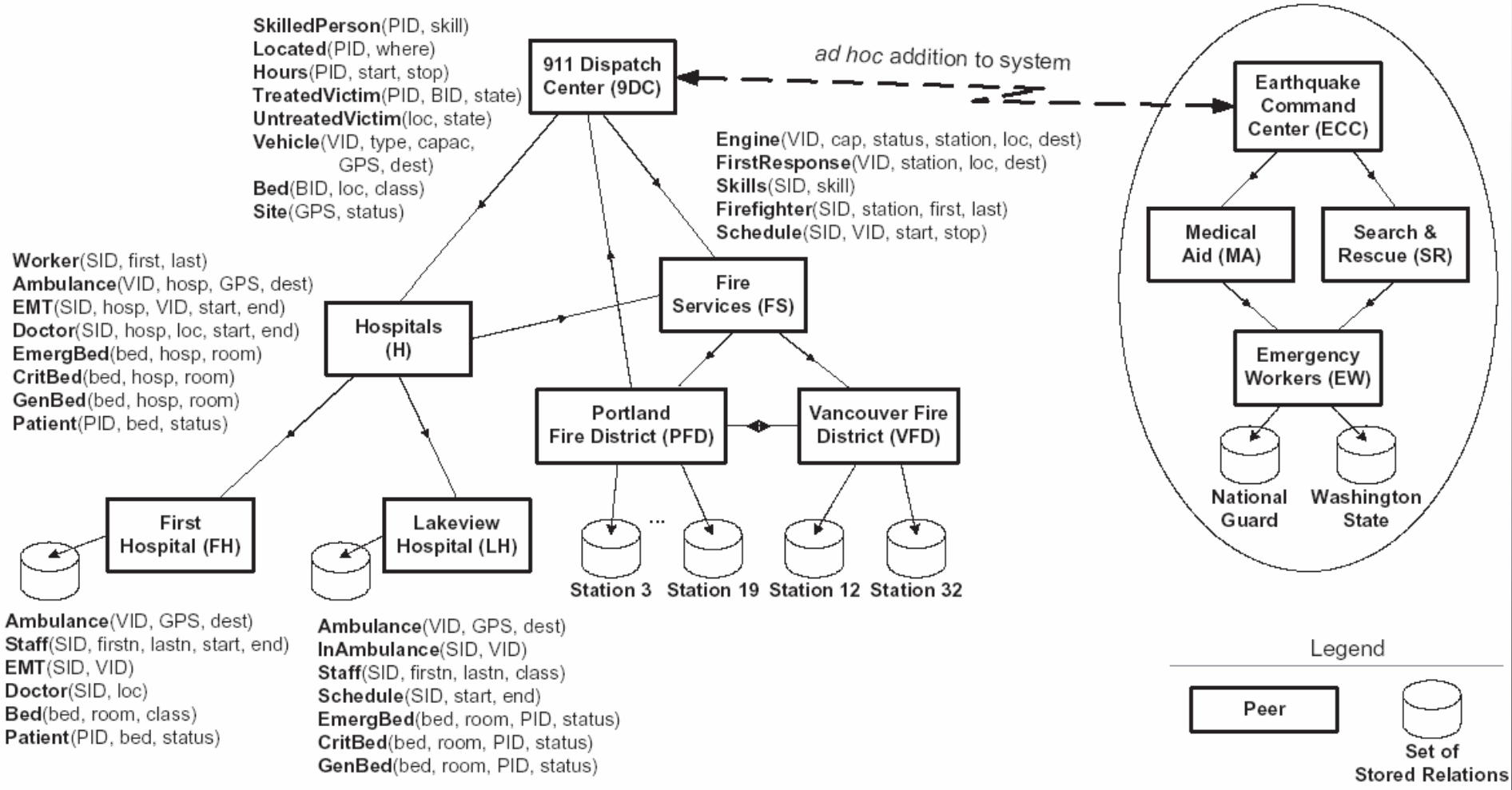
- **… Literatur (IBM Systems Journal)**

Towards an information infrastructure for the grid

by S. Bourbonnais, V. M. Gogate, L. M. Haas, R. W. Horman, S. Malaika, I. Narang, V. Raman

- **Eigenschaften**
  - **virtualized** – allowing a collection of distributed information resources to be shared and managed as if they were a single information store
  - **autonomic** – ensuring that the interconnected information systems can be managed effectively and efficiently through self-management
  - **open** – utilizing open interfaces and agreed-upon standards to enable highly interoperable systems and processes
- **Anwendungsszenarien: Patient Health Information System**
  - Patientendatenbanken verteilt – Überblick wäre für Therapie sinnvoll
  - Computer-gestützte Diagnose – Lokationstransparenz für Patient und Arzt (→ eDiaMoND: national scale database of mammograms to support the UK Breast Screening Program)

Beispiel: Katastrophenmanagement - Piazza (Alon Halevy)

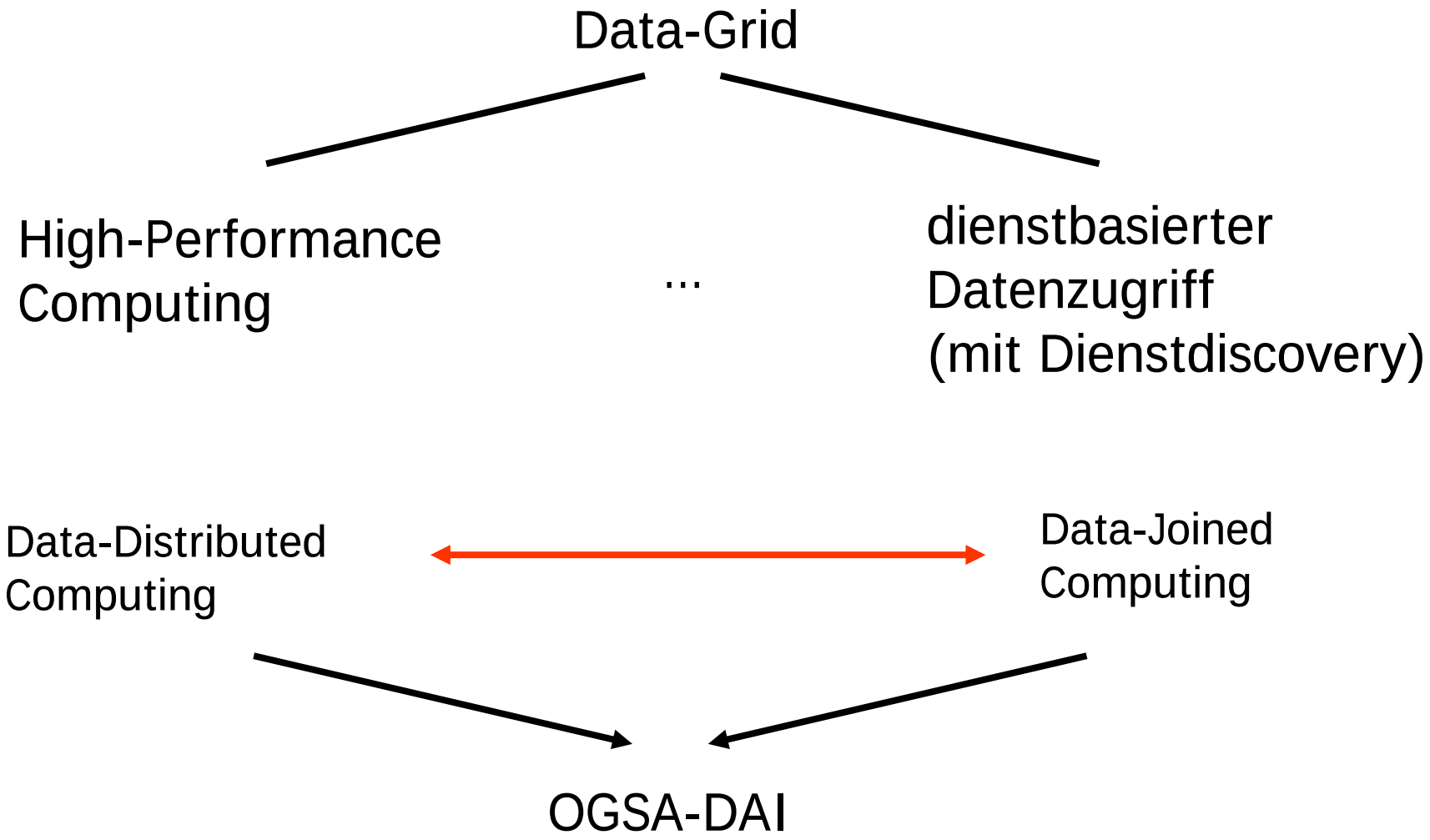


Nebenbemerkung: "Die Rolle von P2P ..."

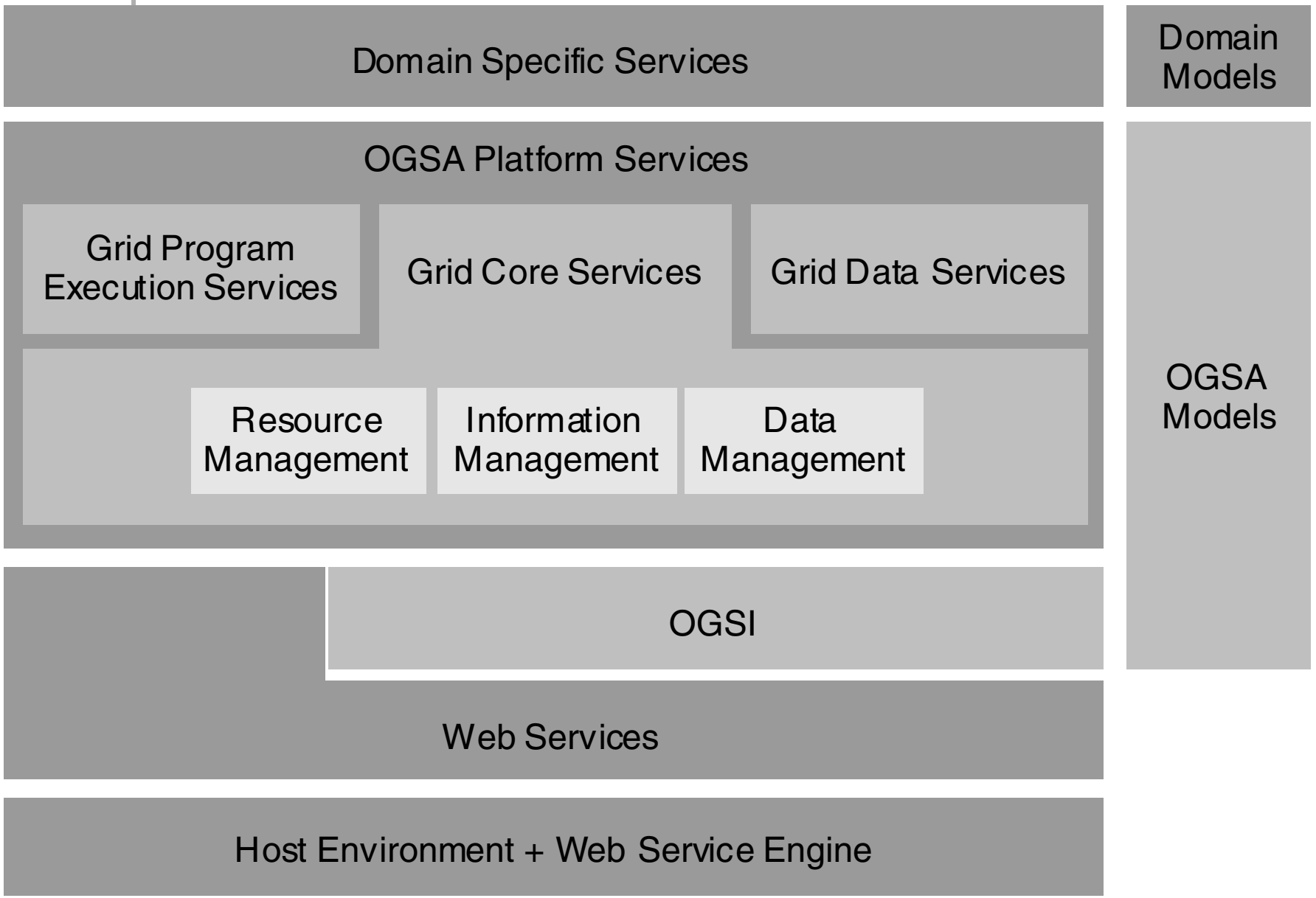
Dresden University of Technology

- **P2P-Systeme**
  - pre-alpha-Version eines Data Grid-Szenarios
  - keine höherwertigen Diensteleistungen
  - verteilt, unstrukturierte Nutzdaten
- **Eigenschaften**
  - Benutzer kennen nur das lokale Schema
  - Gesamter Datenbestand ist erreichbar (transitive Hülle der von Schemamappings)
  - Hinzufügen von neuen Schemata erfolgt inkrementell und transparent
  - Globales Schema wird durch automatisches Schema-Mapping gefunden
  - keine Updates
- **Anmerkung**
  - vielleicht auch technische Realisierung der Verwaltung eines gemeinsamen Datenbestands
  - (k)eine (legale) Anwendung (Andrew Tanenbaum, 2005)

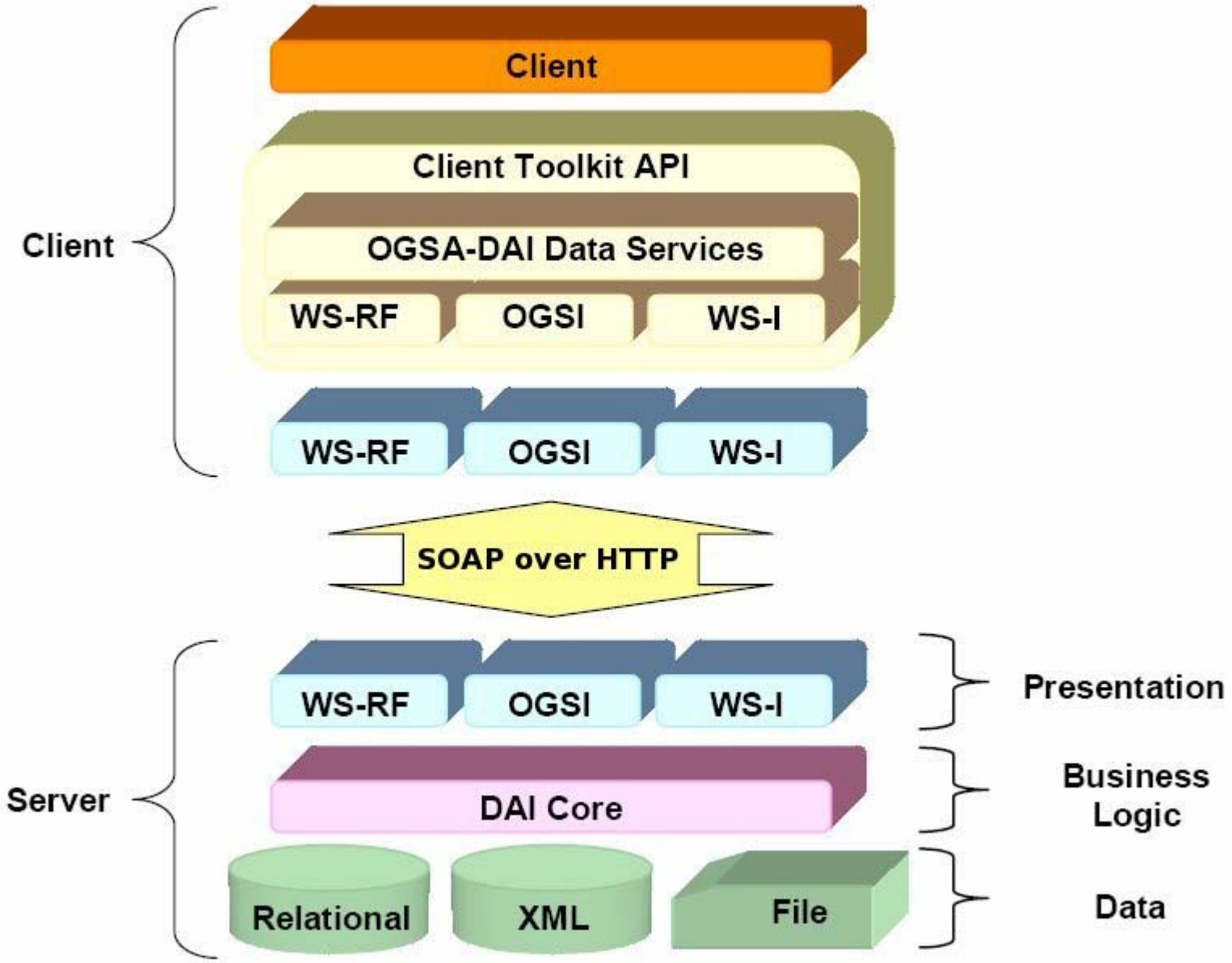
Mögliche Klassifikation



Open Grid Service Architecture



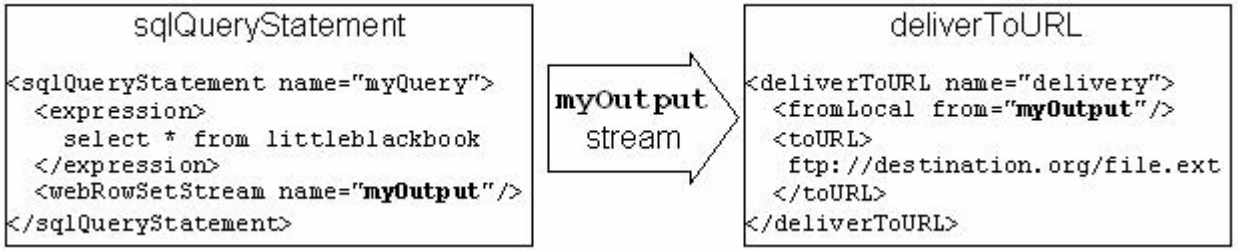
Überblick: OGSA-DAI Architektur



Interaktion mit Daten-Ressourcen

Dresden University of Technology

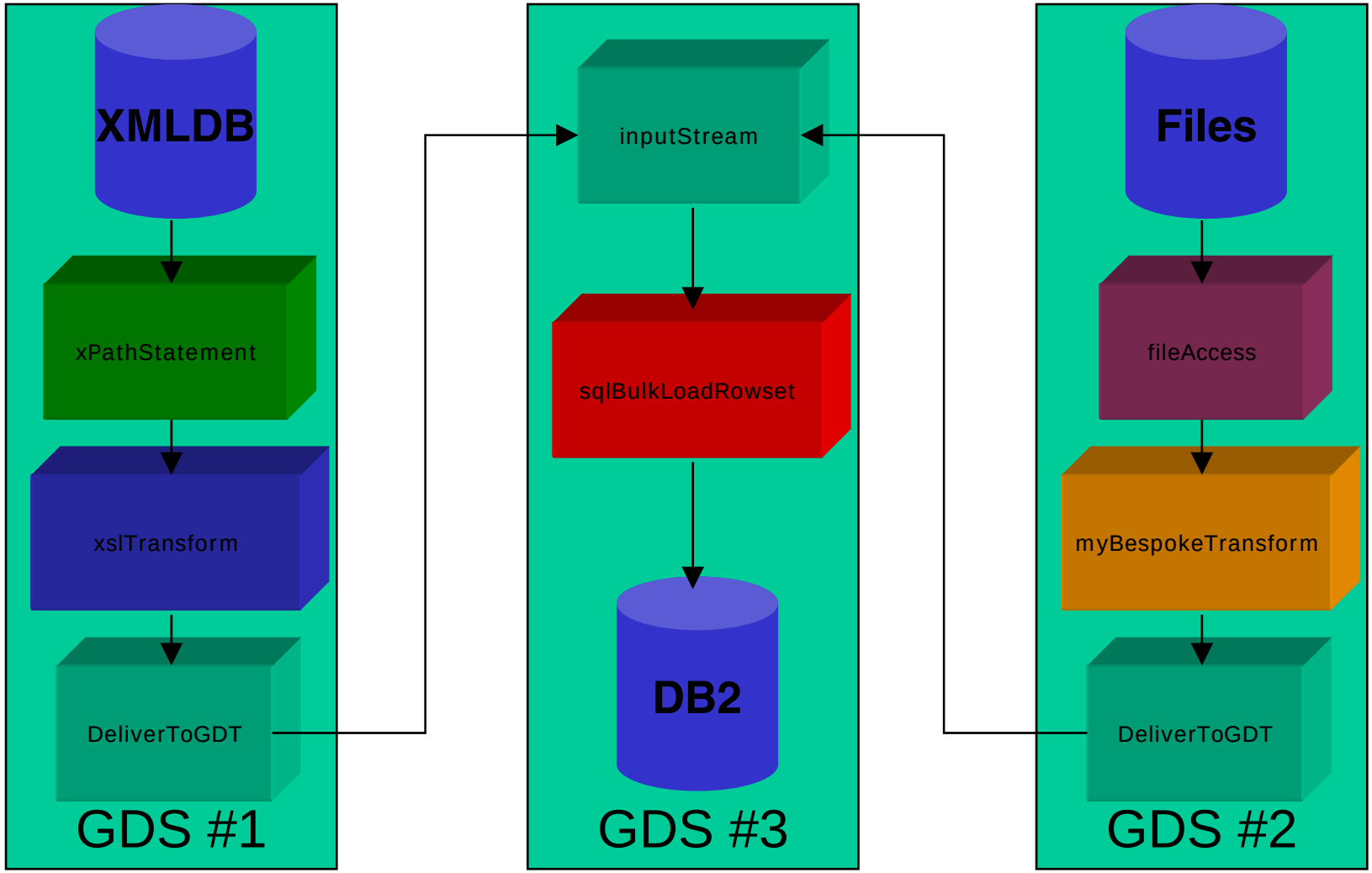
- **Activities**
  - elementare Blöcke für “Perform”-Dokumente
  - Definition von Strömen zwischen Aktivitäten
  - Vielzahl möglicher Aktivitäten
    - Relational Activities
    - XML Activities
    - Delivery Activities
    - Transformation Activities
    - …



- **Perform Documents**
  - Definition von “Anfragen” (Manipulation, Transformtion, …) durch einen Client
- **Response Documents**
  - Rückgabewert (auch Nutzdaten)

Beispiel eines Grid Data-Service-Flows

(Patrick Dantressangle, 2005)

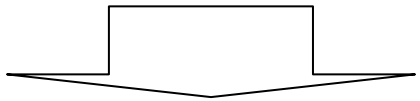


Interaktion mit Daten-Ressourcen

Dresden University of Technology

"Perform"-Dokument

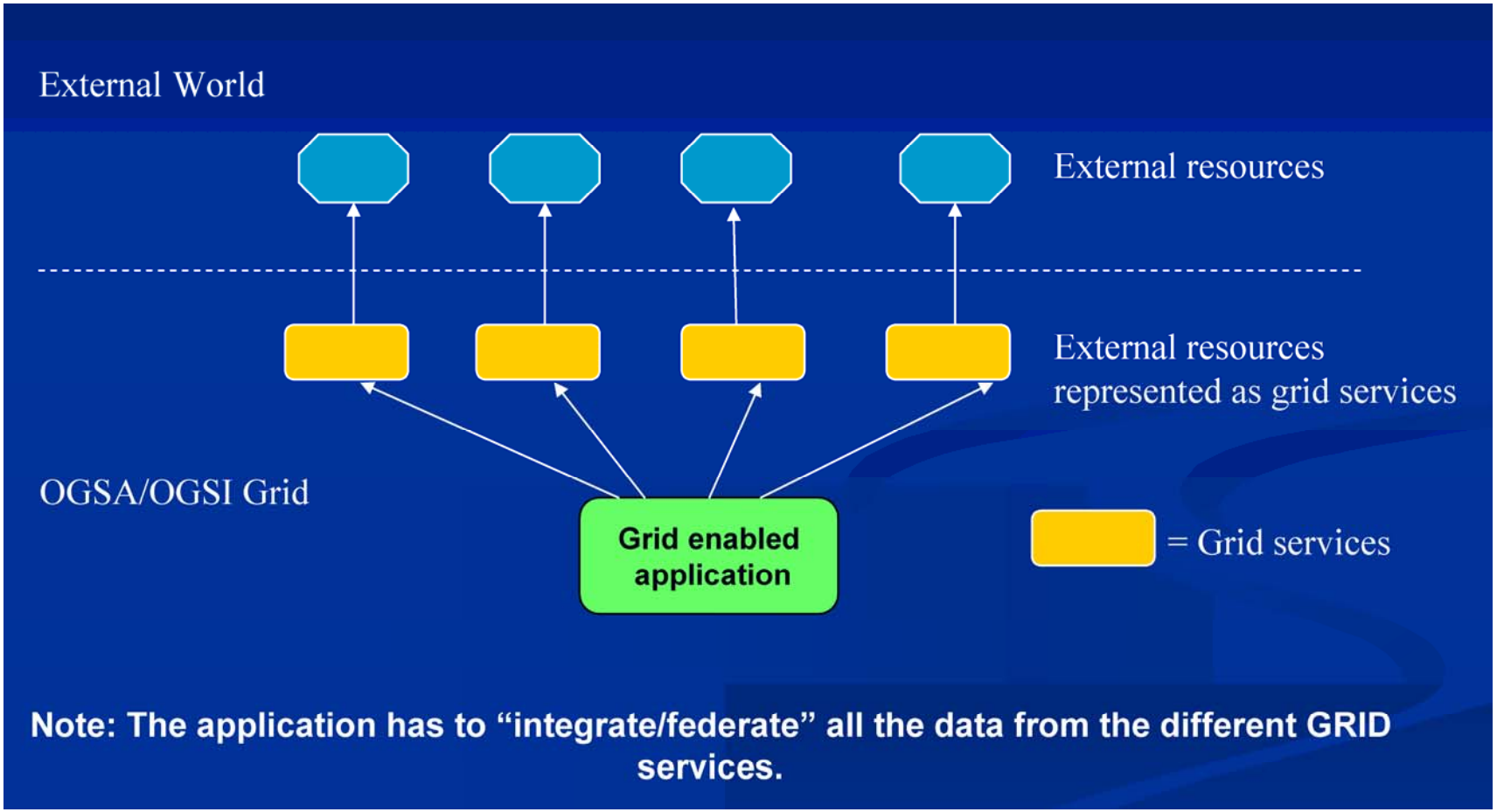
```
<?xml version="1.0" encoding="UTF-8"?>
<perform
  xmlns="http://ogsadai.org.uk/namespaces/2005/03/types">
  <documentation>
    Perform a simple SELECT statement.
  </documentation>
  <sqlQueryStatement name="myQuery">
    <expression>
      select * from littleblackbook where id=10
    </expression>
    <webRowSetStream name="myQueryOutput"/>
  </sqlQueryStatement>
</perform>
```



"Return"-Dokument

```
<?xml version="1.0" encoding="UTF-8"?>
<response xmlns="http://ogsadai.org.uk/namespaces/2005/03/types">
  <request xmlns:NS1="http://ogsadai.org.uk/namespaces/2005/03/types"
    NS1:status="COMPLETED"/>
  <result xmlns:NS1="http://ogsadai.org.uk/namespaces/2005/03/types"
    NS1:name="myQuery" NS1:status="COMPLETED"/>
  <result xmlns:NS1="http://ogsadai.org.uk/namespaces/2005/03/types"
    NS1:name="myQueryOutput" NS1:status="COMPLETED">
    <![CDATA[<?xml version="1.0" encoding="UTF-8"?>
      <webRowSet schemaLocation="http://java.sun.com/xml/ns/jdbc
        http://java.sun.com/xml/ns/jdbc/webrowset.xsd">
      ...
      <currentRow>
         <columnValue>10</columnValue>
         <columnValue>John Smith</columnValue>
         <columnValue>123 Some Lane, AnyTown</columnValue>
         <columnValue>0131-555-1234</columnValue>
      </currentRow>
      ...
  </result>
</response>
```

„Grid-enabled“ Anwendungen

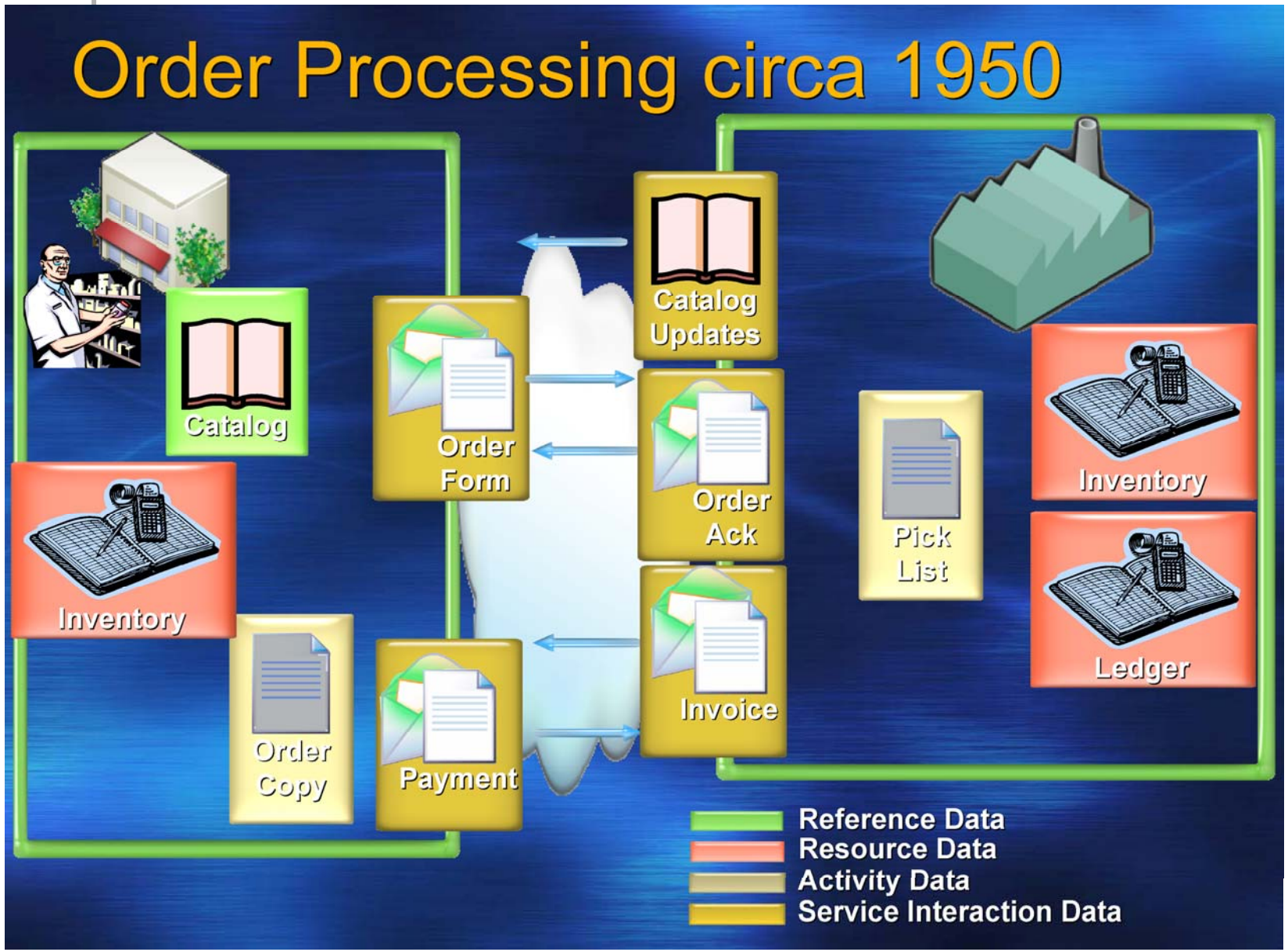


Lesson 1: Es gibt kein globales konzeptionelles Schema !!!

OGSA-DAI ist “XML-basiertes JDBC mit Ergebnisweiterleitung”

Frage 2: Gibt es ein gemeinsames Weltverständnis???

Typen von Daten in einem Data-Grid / Information Infrastructure for the Grid



Ein erster Ansatz … (David Campbell, 2005)

Dresden University of Technology

- **Reference Data**
  - wird benutzt zur Erzeugung von Interaktionen
  - interpretierbar von allen Parteien (jeder kann den Katalog lesen)
  - Container hat einen stabilen Bezeichner (Katalog, Stand März 2005)
  - Referenzierung aus einem Request heraus (Artikel 4711, Seite 64, März 2005 Katalog)
  - Unbeschränkte Möglichkeit der Replikation/des Cachings
  - keine 100%-ige Konsistenz erforderlich (Bestellung am 5.April aus dem März-Katalog)
- **Resource Data**
  - sehr lange Lebensdauer (SKU's, Customers, Accounts…)
  - Format/Schema und Aktualisierungen sind privat für jeden Dienst
  - Konkurrierender Zugriff (Lagerbestand kann von vielen Aufträgen (TAs) geändert werden)
  - klassische TA-Systeme, stabile, wohl bekannte Lokation
- **Activity Data Defined**
  - Lebenszeit ist an die Aktivität/Prozessschritt gebunden (Auftrag, Buchung, …)
  - Format/Schema ist privat für den Dienst
- **Service Interaction Data**
  - Payload von Aufträgen/Service Calls
  - Format/Schema kann von allen Parteien interpretiert werden
  - Nachrichten können verloren gehen (Wiederholtes Versenden eines exaktes Replikas)

Lesson 2: Es gibt (noch) kein globales Weltverständnis !!!

Ein erster Ansatz … (David Campbell, 2005)

- **Klassifikation**
  - Daten
  - Eigenschaften

Kind of Data	Private to Service?	Needs En-capsulation?	XML?	Con-currency	Object Style	Concurrency Approach
Service Interaction	No	No: Has Open Schema	Yes	Read Only	None	None
Reference	No	No: Has Open Schema	Yes	Read Only	None	None
Activity	Yes	Yes	Maybe	Very Low	Object Persistance	Optimistic
Resource	Yes	Yes	Maybe	May Be High	COM+ Transactional	Pessimistic (Locking)

Lesson Learned

Dresden University of Technology

- **Data-Grid definiert sich über eine Vielzahl von Basiseigenschaften und Mehrwertdiensten**
  - Transparenz (immer gewünscht...?)
  - Autorisierung, Vertrauenswürdigkeit („trust management“)
  - Verfügbarkeit (QoS - Datengüte und Datenqualität)
  - Abrechnung, Verantwortlichkeit
  - Anfragemöglichkeit, Aktualisierungen (Transaktionen)
  - dynamisches Verhalten (Hinzunahme/Wegnahme von Datenquellen)
- **Aktueller Stand (meine Meinung)**
  - „re-inventing the wheel …“ → Föderierte DBs, …
  - technische Heterogenität ist/wird gelöst
    - OGSA und Globus Toolkit
    - CIM, ARMS, …
  - semantische Heterogenität
    Definition und Standardisierung von Interface-Semantiken ist völlig offen
    - Modell-Management (Bernstein)
    - Description Logic

Data Warehouse, Data Peers, Data Grid, …:

von Buzzwords zu substantiellen Forschungsfragen?

- **… getreu' dem Motto eines "Reverse Talks"**
  - bitte keine Fragen, sondern Antworten!