



# TextGrid



## Community-Grid für textbasiert arbeitende Disziplinen



Heike Neuroth, SUB Göttingen



# e-Science

"e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it."  
*(John Taylor, Director General of the UK Research Council)*

- **Infrastruktur** für die Wissenschaft
- Technische **Integration**
- **Vernetzung** von Wissenschaftlern

# ToC

- **D-Grid**
- TextGrid
- Perspektive: Bibliotheken

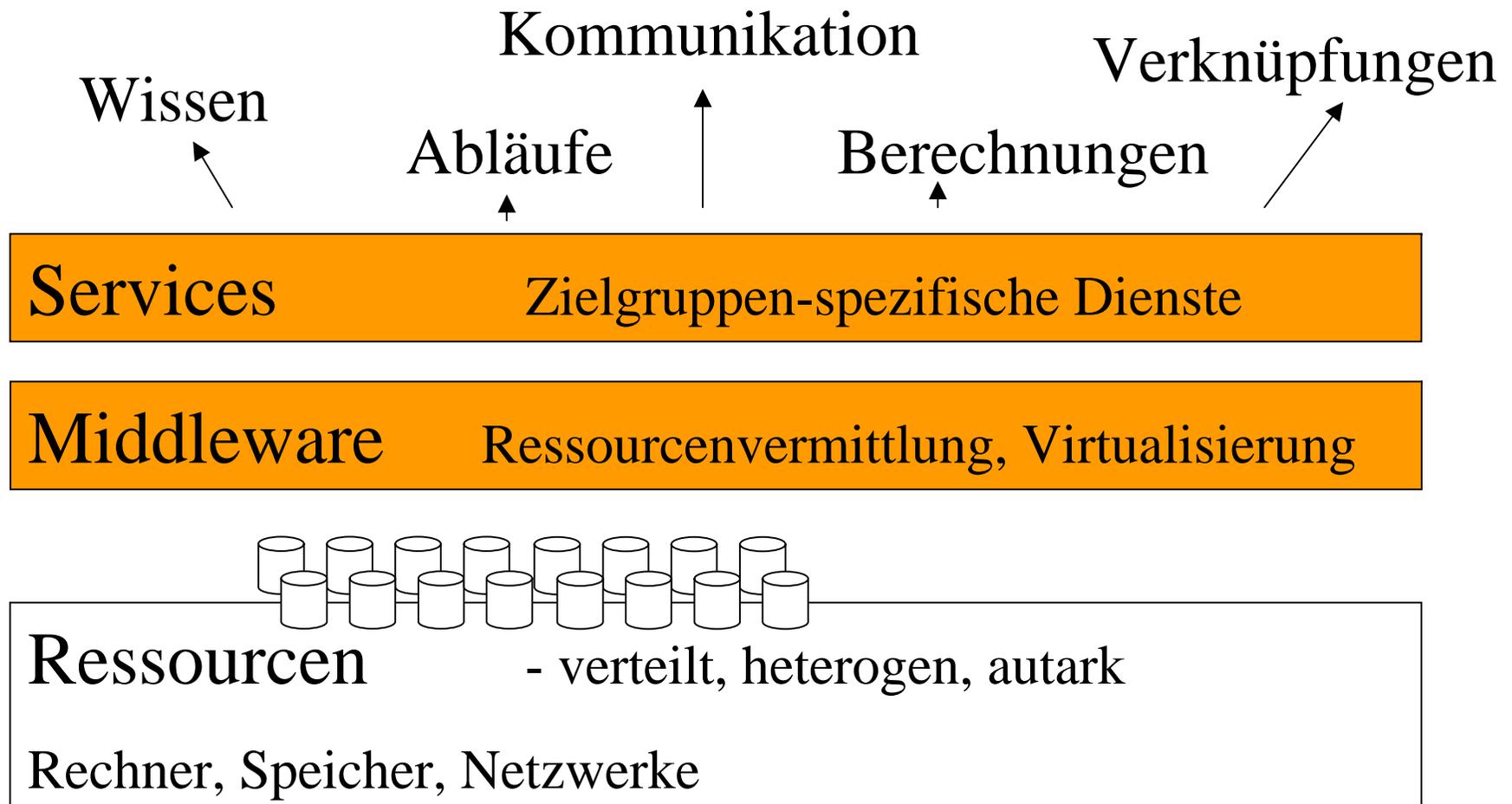
# BMBF Programm

- 2005 - 2009
- 100 Institutionen
- 100 Mio Euro Förderung
- bisher 3 Fokusbereiche:
  - e-Learning (Juni 2004)
  - D-Grid (August 2004)
  - Wissensvernetzung (November 2004)

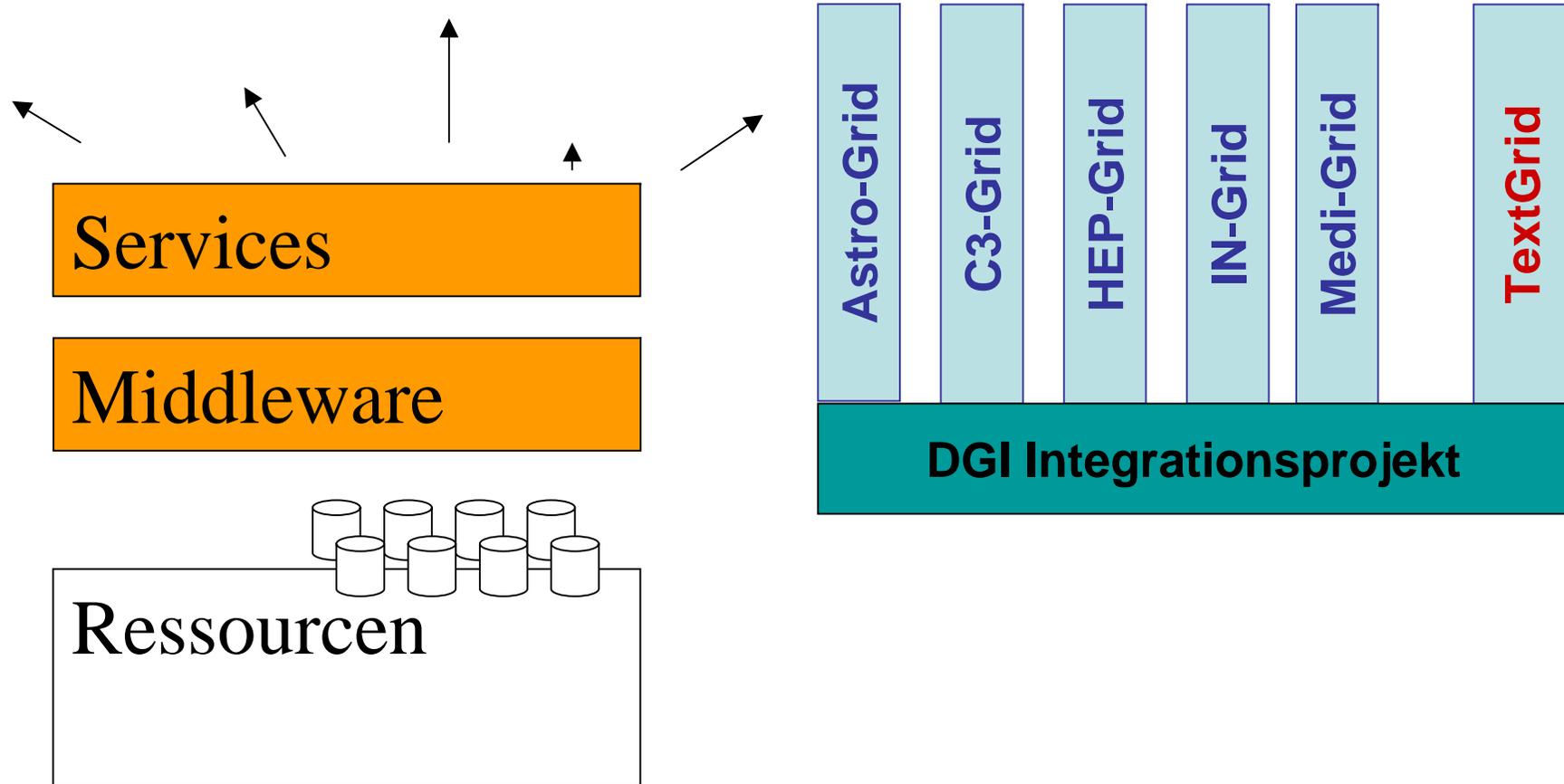


Bundesministerium  
für Bildung  
und Forschung

# Grid Umgebung



# D-Grid



# Chancengleichheit

- e-Science / Grid:  
Naturwissenschaften und Medizin
- in allen Disziplinen:
  - e-Social Sciences
  - e-Teaching
  - e-Humanities
  - ...

# ToC

- D-Grid
- **TextGrid**
- Perspektive: Bibliotheken



**ein Community-Grid  
für die Geisteswissenschaften**

**[www.TextGrid.de](http://www.TextGrid.de)**



# Hintergrund

## **Leitung:**

Niedersächsische Staats- und Universitätsbibliothek Göttingen

## **Fachwissenschaftliche Partner:**

Fachbereich Informatik und Telekommunikation, Fachhochschule  
Worms

Institut für Sprach- und Literaturwissenschaft, TU Darmstadt

Institut für Deutsche Sprache, Mannheim

Kompetenzzentrum für elektronische Erschließungs- und  
Publikationsverfahren in den Geisteswissenschaften, Universität  
Trier

Kompetenzzentrum EDV-Philologie, Universität Würzburg

# Hintergrund

## ***Kleinere und mittlere Unternehmen (KMU):***

DAASI International GmbH, Tübingen

Saphor GmbH, Tübingen

***Laufzeit:*** Februar 2006 bis Januar 2009

***Fördervolumen*** insgesamt: ca. 1,6 Mio Euro

# Aufgaben

- Generische Plattform für philologische Textverarbeitung (linguistisch und literaturwissenschaftlich)
- Service Grid – Storage Grid
- Modularer Aufbau und einfache Erweiterbarkeit, um neue Werkzeuge zu integrieren
- Basiert soweit wie möglich auf offenen Standards
- Open Source Entwicklungen
- Integriert die einzelnen Werkzeuge unter einer Benutzerschnittstelle

# Aufgaben

- Ermöglicht die Definition und Konfiguration von Arbeitsabläufen/Prozessen
- Lokale Arbeitsweise für einige wenige Dienste
- Grid-Integration für (fast) alle Dienste
- Projektteilnehmer stellen exemplarische Text-Korpora zur Verfügung
- Schnittstellen für die schnelle und einfache Integration weiterer Daten/Archive

# Arbeitspakete

- AP 1: Inhaltliche Studie mit Empfehlungen über die Nachnutzbarkeit internationaler Editionstools
- AP 2: Entwicklung Community-spezifischer Werkzeuge
- AP 3: Anbindung der Community-Tools an die Integrations-Plattform
- AP 4: Entwicklung der Community Muster-Applikation
- AP 5: Semantic Web und TextGrid
- AP 6: Projektmanagement und Öffentlichkeitsarbeit

# Beispiele Tools

- *Text Processing*
  - XML-Editor, Metadaten-Annotation, Lemmatisierung, Kollationierung, Streaming-Editor, Tokenizer, Sortieren, Qualitätsanreicherung von OCR-Daten ...
- *Text Retrieval*
  - Query-Interface, Text Retrieval
- *Linking*
  - Link-Editor für Bild-Text, Link-Editor für Text-Text
- *Administrative Dienste*
  - Editor für den technischen Workflow, Editor für den administrativen Workflow

# Bisher

- Uneinheitliche, von einander isolierte Zugänge zu Korpora etc. (z.T nur Intranet)
- Projektbezogene bzw. domain-spezifische Applikationen, Tools etc. (lokale und temporäre Insellösungen)
- Fehlende Verknüpfung mit Erschliessungsmaterialien (z.B. digitale Archive)

# Textwissenschaften

- Erschliessung
- Wissenschaftliche Edition
- Publikation

# Textwissenschaften

- Erschliessung
  - Erfassung von Textkorpora
  - Auszeichnung
  - Metadaten
- Wissenschaftliche Edition
- Publikation

# Textwissenschaften

- Erschliessung
- Wissenschaftliche Edition
  - Sichtung von Autographen und Ausgaben
  - Darstellung von Varianten
  - Einbettung in den Kontext (z.B. Hintergrundmaterial, Rezeptionsgeschichte etc.)
  - Recherchefunktionen
  - Analyse
- Publikation

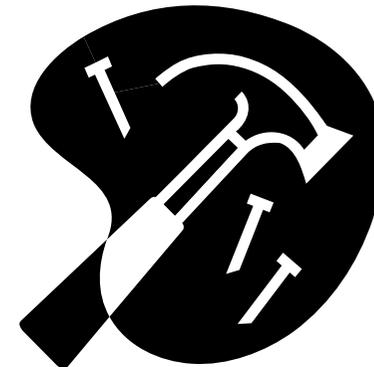
# Textwissenschaften

- Erschliessung
- Wissenschaftliche Edition
- Publikation
  - Textsatz
  - Darstellung in Browser

# TextGrid

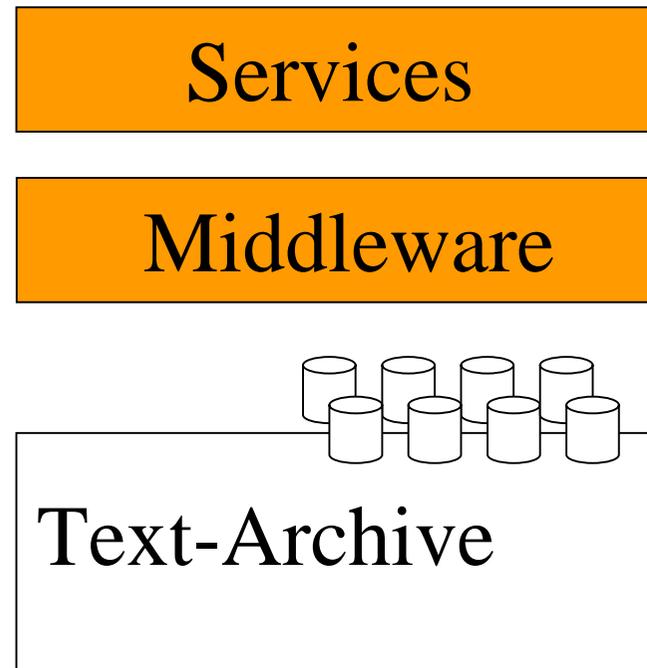
Virtuelle Forschungsbibliothek  
Grid-fähige Workbench für philologische:

- Erfassung und Bearbeitung
- Analyse
- Annotation
- Edition
- Publikation

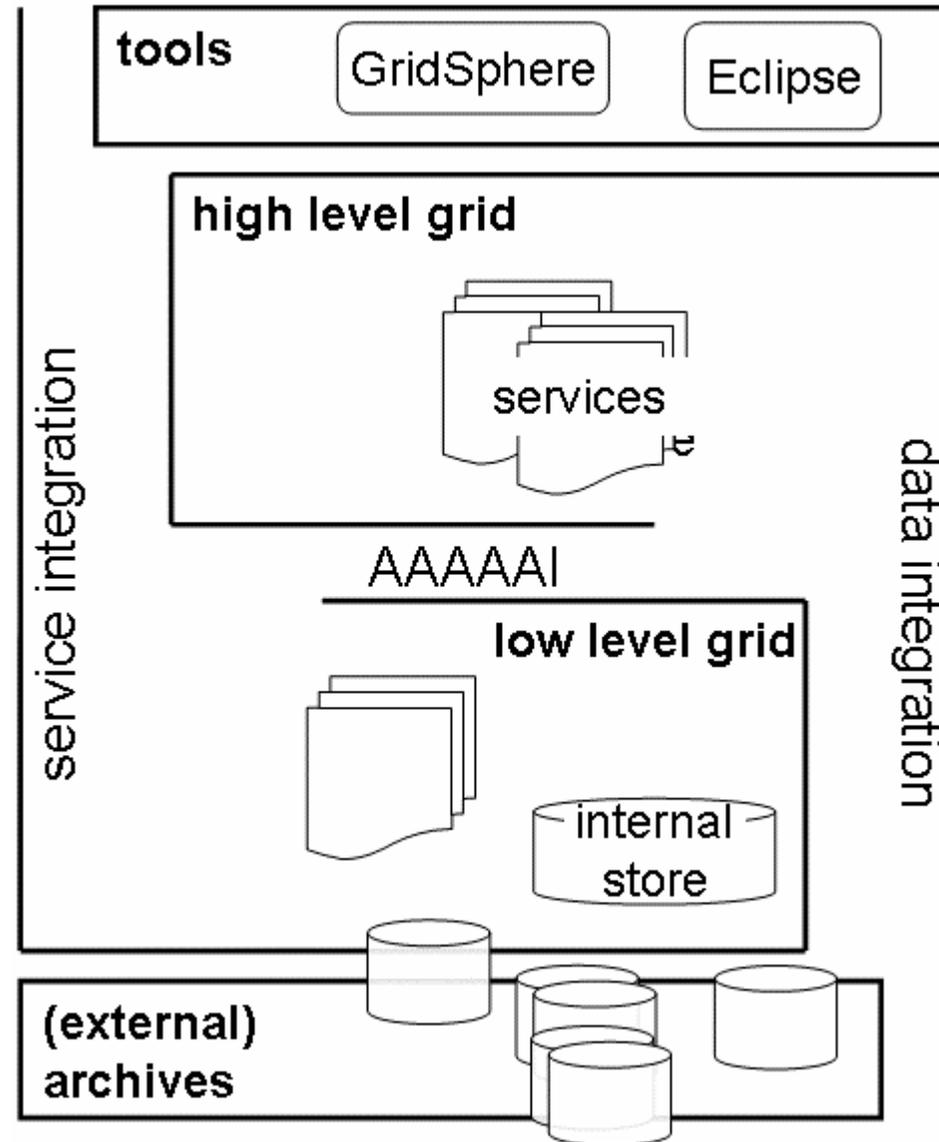


... von Texten

# TextGrid Architektur



# TextGrid Architektur



# Standards + Tools

- TEI  
... Text Encoding Initiative
- SOAP, WSDL etc.  
... W3C
- Globus, WSRF etc.  
... Global Grid Forum
- METS, MARC 21 etc.  
... Metadaten, Erschließung

# ToC

- D-Grid
- TextGrid
- **Perspektive: Bibliotheken**

# e-Science

- **Rechner- und Speicherkapazitäten**

→ virtuelle Wissensumgebungen mit umfassenden Datenbeständen, Visualisierungen und wissenschaftlichen Informationen

- **Technische Integration**

→ Tiefgreifende Verknüpfung vielfältiger Akteure, Dienste, Ressourcen und Methoden

- **Vernetzung von Menschen**

→ Unterstützung von Arbeitsabläufen, Kommunikation und Wissensaustausch

# e-Science

- **Rechner- und Speicherkapazitäten**

→ CERN: Large Hadron Collider Computing Grid Project (LCG), 15 Petabyte Daten pro Jahr, ab 2007

- **Technische Vernetzung**

→ Meteorologie: Messungen an Land, Wasser, im Weltall etc. werden global zusammengeführt

- **Vernetzung von Menschen**

→ AccessGrid - virtuelle Treffpunkte, Kollaboration

→ Collate - Arbeitsabläufe in Videoarchiven

→ eSciDoc - Publikationssystem

# e-Science in Bibliotheken

- Informationsgewinnung
- Informationsverwaltung
- Informationsweiterbearbeitung
- Kommunikation und Zusammenarbeit
- Publizieren

# Aufgaben für Bibliotheken?

- Arbeitsteilige Entwicklung von Diensten, auch mit externen Anbietern (Service Grid)
- Bereitstellung offener Dienste und wiederverwendbarer Ressourcen
- Vernetzung und virtuelle Organisation verteilter Informations-Repositoryen
- Archivierung und Sicherung des Langzeitzugangs (Storage Grid)

# Weitere Informationen

- TextGrid [www.textgrid.de](http://www.textgrid.de)
- D-Grid [www.d-grid.de](http://www.d-grid.de)
- BMBF e-Science Initiative  
<http://www.bmbf.de/de/298.php>
- Digital Library Forum [www.dl-forum.de](http://www.dl-forum.de)
- e-Science Forum  
<http://www.e-science-forum.de/de/60.php>

## Weitere Initiativen

- UK e-Science Programme  
<http://www.rcuk.ac.uk/escience/>
- US Cyberinfrastructure programme  
<http://www.nsf.gov/od/oci/reports/toc.jsp>
- AustrianGrid  
<http://www.austriangrid.at/>
- BIG GRID, the Dutch e-Science Grid  
<http://www.dutchgrid.nl/>
- ...



Vielen Dank für  
Ihre Aufmerksamkeit !

## Fragen, Anmerkungen ... ?



Heike Neuroth, SUB Göttingen  
[neuroth@sub.uni-goettingen.de](mailto:neuroth@sub.uni-goettingen.de)

