



# International Symposium on Grid Computing 2008

7 ~ 11 April 2008, Academia Sinica, Taipei, Taiwan

## On the Road towards Arts and Humanities e-Infrastructure in Germany



*Excellent Information Services for Excellent Research*

Dr. Heike Neuroth  
Göttingen State & University Library  
Max Planck Digital Library Berlin  
[neuroth@sub.uni-goettingen.de](mailto:neuroth@sub.uni-goettingen.de)

# ToC

---

- **TextGrid** - A Community Grid for the Arts and Humanities  
[Federal Ministry of Education and Research, BMBF]



- **eHum** - National eHumanities Infrastructure project  
[German Research Foundation, DFG]



# TextGrid Goals

---

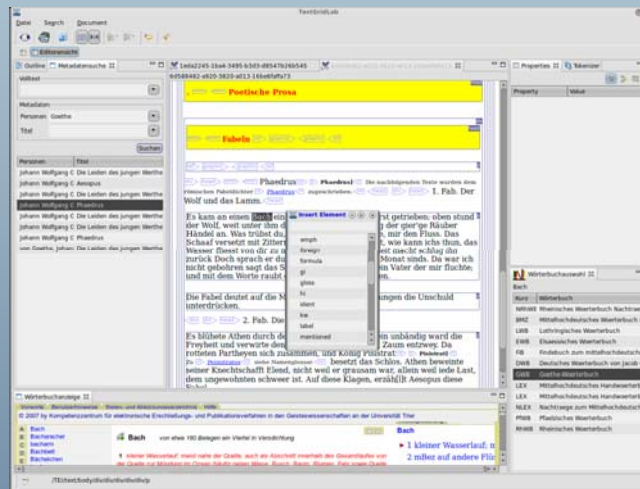
- Establishment of a “**Virtual Research Infrastructure**”
- TextGrid: a **generic platform** for scientific text data processing
- **Service Grid** - Toolbox for collaborative work
- **Data Grid** - Virtual archive(s) for data curation



# TextGridLab

- Eclipse-based interactive client integrating all tools and services in a unified interface
- TextGridLab: scholarly workbench for creating and processing text resources

- Intuitively usable by way of dynamic views and context menus

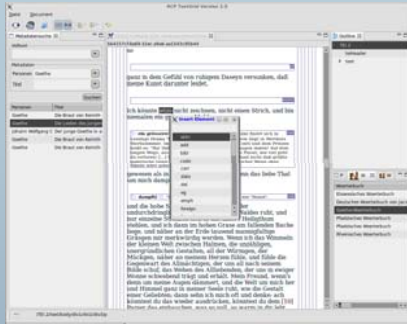


- Unified help and reference system increases usability

- Open standards and interoperability

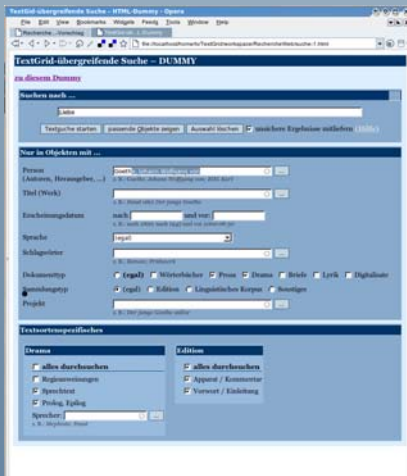
- Additional tools and services can be integrated at any time

# TextGridLab



## • XML Editor

- WYSIWYM mode (What you see is what you mean) on top of the Eclipse source code editor

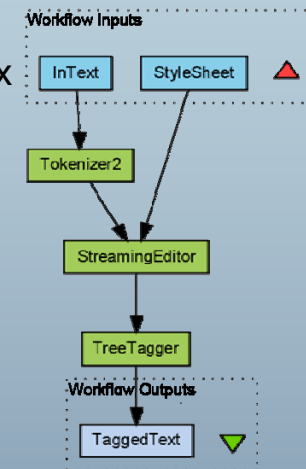


## • Searching

- Search across the entire TextGrid data pool
- Full text search and metadata-restricted search
- Available as Web interface and TextGridLab module

## • Workflow Editor

- Tool for defining complex workflows
- At this stage realised as external tool Taverna
- Later on: integration in TextGridLab



## • Project Management

- Management of users, roles and access rights for projects
- Based on Role-Based Access Control (RBAC, an ANSI standard)
- Shibboleth-enabled



# TextGrid Tasks

---

- Roadmap for community building (difficult!)
- (Business/Organizational) models for sustainability
- Research infrastructure providing new functionality:
  - new forms of collaboration and cooperation
  - new, reliable standards; easy exchange of data and research results; new forms of linking up researchers, services, and content
  - new impulses on the conduct of research
  - long-term preservation of primary and secondary data

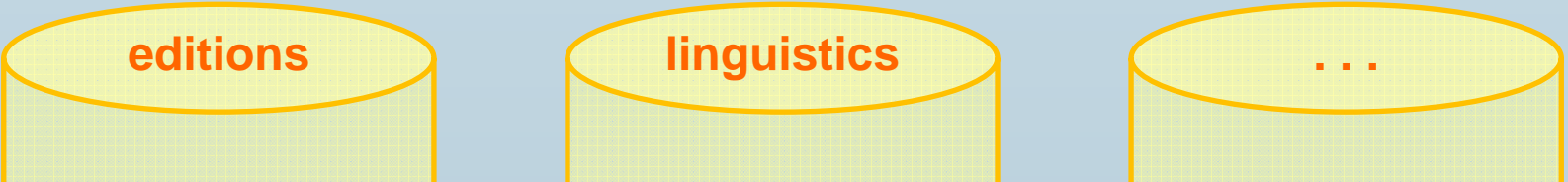
→ **TextGrid: contributes to the preservation and the keeping alive of our Cultural Heritage**

---

# TextGrid User Groups

---

- **End users** who make use of TextGrid, or its web interface, for personal purposes like searching, reading or copying reliable texts of reference quality
- **Editors**, i.e. philologists or linguists providing text content and using TextGrid as a toolbox and a repository for long-term data preservation
- **Tool developers** who contribute to TextGrid by developing new functionalities (e.g. text mining tools)
- **Institutional content providers** like libraries, archives, or academies who provide content on a large scale utilising the TextGrid data format, repository, and web services



**TextGridLab**

XML-editor

search

workflow editor

link editor

bibliography tool

metadata annotation

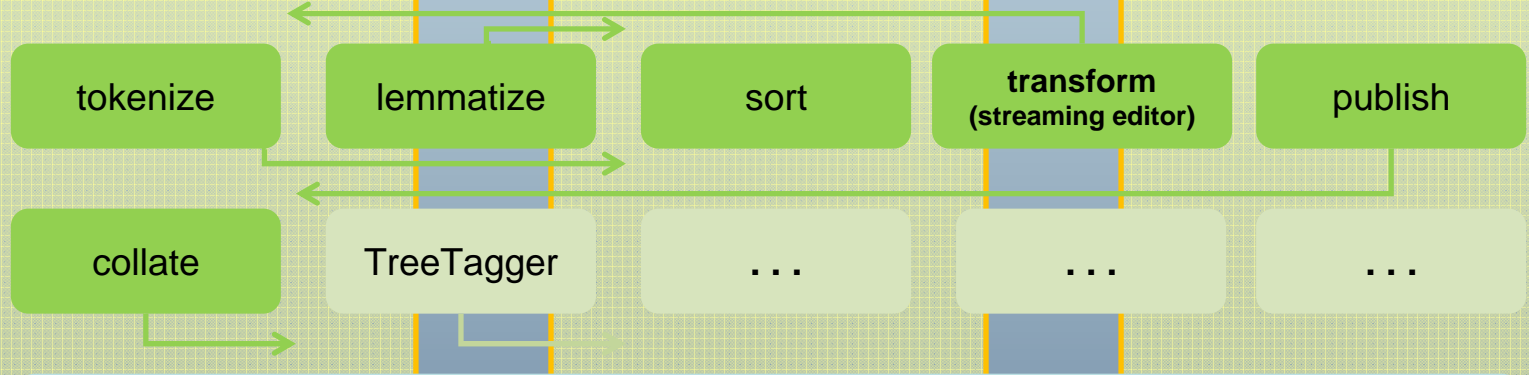
file /rights management

tool interface

...

...

(additional plug-ins)



**TextGrid-Middleware**  
 (file-services, text-retrieval, property management, rights management, adaptors, etc.)

**Grid (including TextGridRep)**  
 archives, long-term repositories, databases

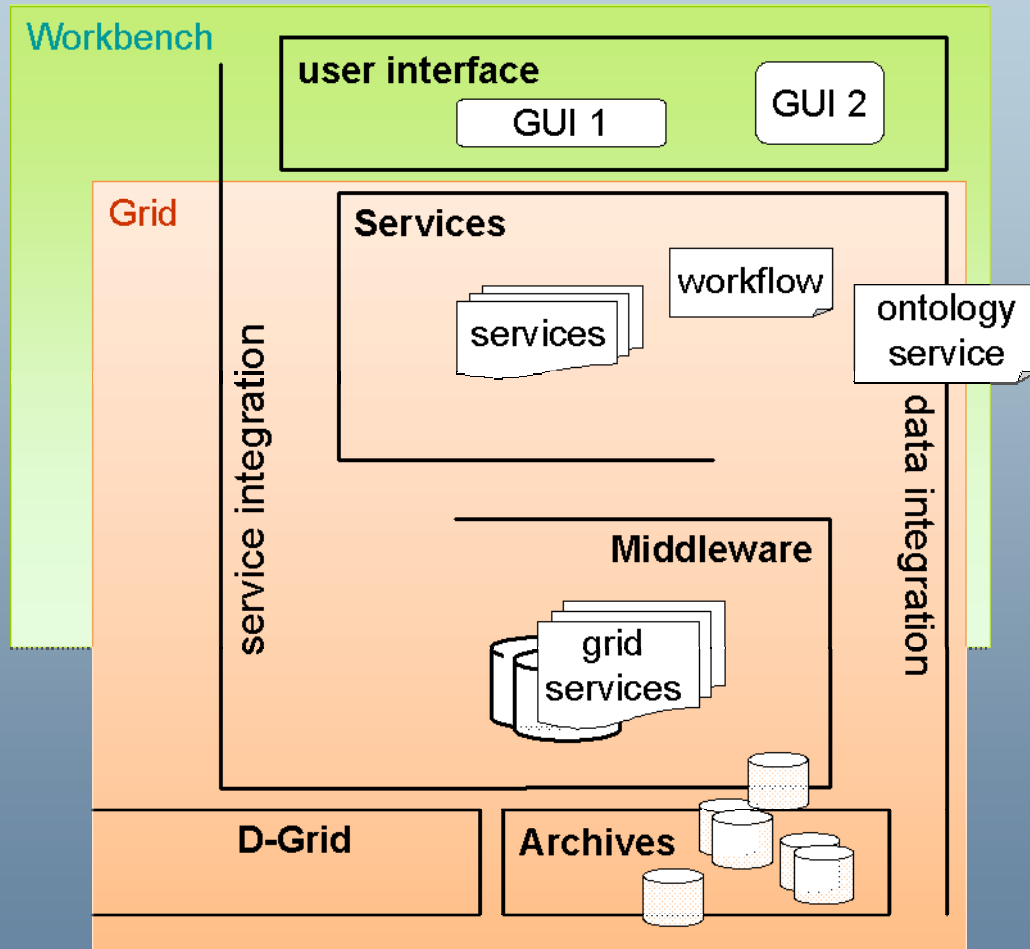
catalogues,  
 external  
 archives, other  
 data sources

# Technical Challenges - Services

---

- Service Oriented Architecture (communication between processes, registry of services, orchestration of workflows, common security features ...)
- Proving encapsuled tools, e.g.
  - tokenizer, lemmatizer, dictionary services, streaming editor, interactive editors, ...
  - workflow enacter
  - middleware interaction, e.g., registering and retrieving resources
- ...

# Middleware/Architecture



- **Globus-based** grid infrastructure with a specific middleware layer and a service layer of specialised functionalities for textual processing
- Service layer is conceived as an **open web service** environment
- **Eclipse-based** interactive client as workbench
- General public can access published contents using a **web interface**
- Openness and interoperability through usage of standards (TEI and the XML family, RDF, SOAP, WSDL 2.0, GSI, WSRF, SAML, LDAP, and BPEL)

# National eHumanities Infrastructure

---

- **Developing Germany's eInfrastructure for the humanities**
- **Conceptualizing in 2008**
  - Analysis of over 30 initiatives, organizations, projects (e.g. Digital Humanities, European Resource observatory for the Humanities and Social Sciences, JISC e-Infrastructure-Programme 2006-2009 ...)
- **Goals and main objectives**
  - to explore the current provision of Germany's eInfrastructure for the humanities
  - to help define its future development

# National Project: eHum

---

- **Overall objectives**

- to create a **strategy** for the development of a research infrastructure within the humanities in Germany
- to develop a **roadmap** for the building of collaboration and partnerships between institutions and organizations for a German competence centre for e-Humanities

- **Specific objectives**

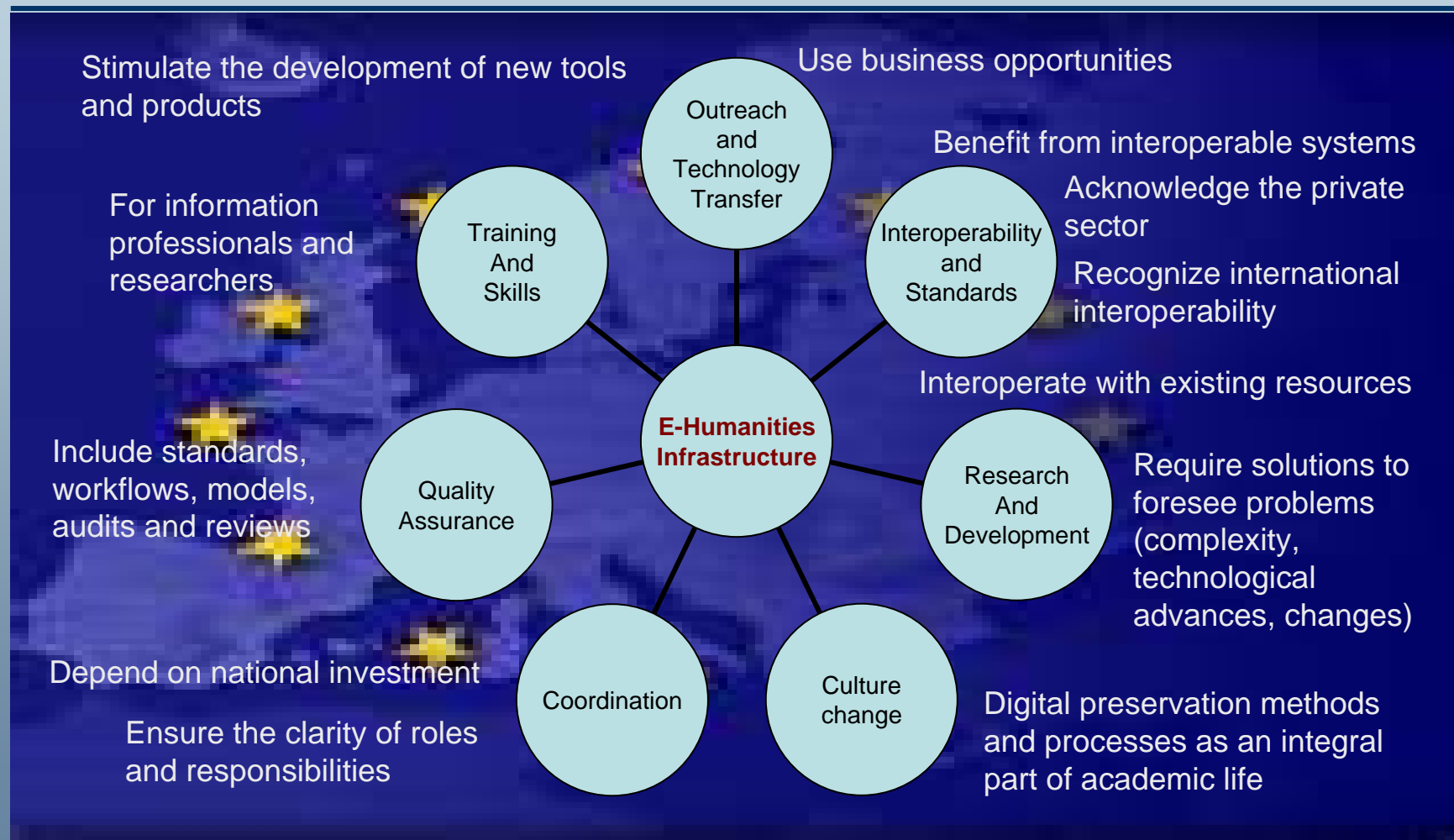
- to analyze national and international merges in the field of e-Humanities
- to gather information about current activities especially in Europe and USA
- to create overviews of such organizations to promote collaboration and synergies at German level

# Vision for a National eHumanities Infrastructure

---

- **Germany's infrastructure should provide researchers**
    - access to systems, services, networks and resources
    - discovery of new resources
    - trustworthiness of integrity, authenticity and quality of resources, services
    - accessibility of output (now and in future)
    - access to multiple data sources
    - training, education
  - **Germany's infrastructure should enable researchers**
    - to collaborate and communicate
    - to exploit the power of information technologies
    - to support innovation and experimentation
    - to share research outputs
    - to engage with industry
    - to explore new research terrain
    - to develop new research methodologies (e.g. based on text mining)
  - **Germany's infrastructure must enable knowledge transfer**
    - to track outputs
    - to protect individuals work and privacy
    - to protect intellectual property and rights management
    - to preserve, curate research (data, publication)
-

# Principal Themes





# International Symposium on Grid Computing 2008

7 ~ 11 April 2008, Academia Sinica, Taipei, Taiwan

**Thank you for your attention!**  
**Questions, comments?**



**Dr. Heike Neuroth**  
**Göttingen State & University Library**  
**Max Planck Digital Library Berlin**  
**[neuroth@sub.uni-goettingen.de](mailto:neuroth@sub.uni-goettingen.de)**