TextGrid - A Virtual Research Environment for the Humanities

Dr. Heike Neuroth
Director Research & Development Department, SUB Göttingen
eHumanities Consultant, Max Planck Digital Library

Washington DC, April 2011
Virtual Research Environment (VRE)

“The term VRE is now best thought of as shorthand for the tools and technologies needed by researchers to do their research, interact with other researchers (who may come from different disciplines, institutions or even countries) and to make use of resources and technical infrastructures available both locally and nationally.

The term VRE also incorporates the context in which those tools and technologies are used.”

JISC VRE Collaborative Landscape Study
http://www.jisc.ac.uk
TextGrid – www.TextGrid.de

• ... supports access to and exchange of research data in the arts and humanities
• ... comprises tools, infrastructure and resources
• ... offers flexible collaborative structures, particularly for (small) research groups

➢ Enables collaborative research in a distributed, secured environment and the collective use of tools, data and digital methods

Funded by the Federal Ministry of Education and Research Germany since 2006 (about 4 Mio Euro)
Two Components

- **TextGrid Laboratory**
  - A single point of entry to the virtual research environment
  - Integrated access to tools and services
  - Continuously advanced, enhanced with new features and released regularly for public beta testing (http://www.textgrid.de/en/beta.html)

- **TextGrid Repository**
  - *Humanities data centre* for research data curation, embedded in a grid infrastructure
  - Ensures long-term availability, access and re-use of research data
## TextGridLab: Tools und Services
(v1.0 – June 2011)

### Tools in the TextGridLab:

<table>
<thead>
<tr>
<th>Tool</th>
<th>Available in current Beta, full functionality and documentation in v1.0 (June 2011)</th>
<th>In development, full functionality in v1.0 (June 2011)</th>
<th>Additional tools and services with constricted support, already available in current Beta version</th>
</tr>
</thead>
<tbody>
<tr>
<td>XML Editor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text-Image-Link Editor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dictionary Search Tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workflow Tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text Publisher Web</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemmatizer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project and User Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Projektbrowser / Navigator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata Editor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregation Composer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upload Tool</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Further services:

- Streaming Editor
- Service Registry
- Tokenizer
- Sort Tool
TextGridRep: Grid Storage

- 275 TB Grid Storage, 250 TB Tape Storage
- First instance: bitstream preservation - redundant storage and tape single (B+C), PID, LTP metadata, ...
- End of 2011: Higher-Value LTP Services (WissGrid)
- Storage security levels
  - A) single storage (disk)
  - B) redundant storage (disk)
  - C) tape single
  - D) tape redundant
  - E) distributed (multiple sites) + redundant (max. security)
TextGridRep: Data Curation Framework

- Repository (3 Variants)
- LTP Guidelines
- Virtual Research Environment
- Metadata Extraction
- Format Validation
- Conversion
- Provenance Service
- Integrity Service
- Replication

- Community =
- WissGrid =
- D-Grid / Infrastructure =
TextGrid - Architecture
(v1.0 – will be released in June 2011)

Eclipse Frontend

TextGridLab

Ingest

TextGridRep

rights management

TG-auth*

search index

TG-publish

+ metadata validation /QA
+ Persistent Identifier
+ (LTP MD)
+ (LTP services)

Portal(s)

search index

Ingest

TG-publish

+ (metadata validation)
+ (persistent identifier)
+ (LTP MD + services)

LTP

Grid Storage

dynamic

static

• open interface
• specific portals

• huge amounts of data
• individually adaptable

open interface

specific portals
Research Infrastructure for the Humanities

Astronomy
Climate Research
High-Energy Physics
Medicine
Humanities

1. Sustainability
2. Blueprints for new VREs
3. Research Data Curation

Support, Monitoring, Accounting, Billing, etc.
Summary & Outlook

• TextGrid is a “development project”, not so easy for scholars to invest time and effort in IT issues

• It is very important that TextGrid is totally community-driven, but it needs time to agree on a common culture between scholars and IT experts

• Sustainability of VREs after project end, funding is not ensured, but:
  ➢ TextGrid is part of DARIAH – EU/ESFRI programme (www.DARIAH.eu), which will start with the construction phase now, and of the German Grid Initiative: some kind of sustainability!

• June 2011: TextGrid - first release (version 1.0)
** Thank You **

Questions, Comments ?