



e-Infrastructures for Cultural Heritage

Brussels Conference 29 October 2010



TextGrid - A Virtual Research Environment for the Humanities



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

Bruxelles, October 2010

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 at a glance

- The joint project TextGrid aims to support ***access to and exchange of data*** in the arts and humanities by means of modern information technology (the grid)
- TextGrid comprises ***tools, infrastructure and resources***. It offers flexible collaborative structures, particularly for research groups.
- TextGrid enables collaborative research in a distributed, secure, flexible and extensible environment and the ***collective*** utilization of tools, data and methods.



Virtual Research Environment TextGrid



- **TextGrid Laboratory**

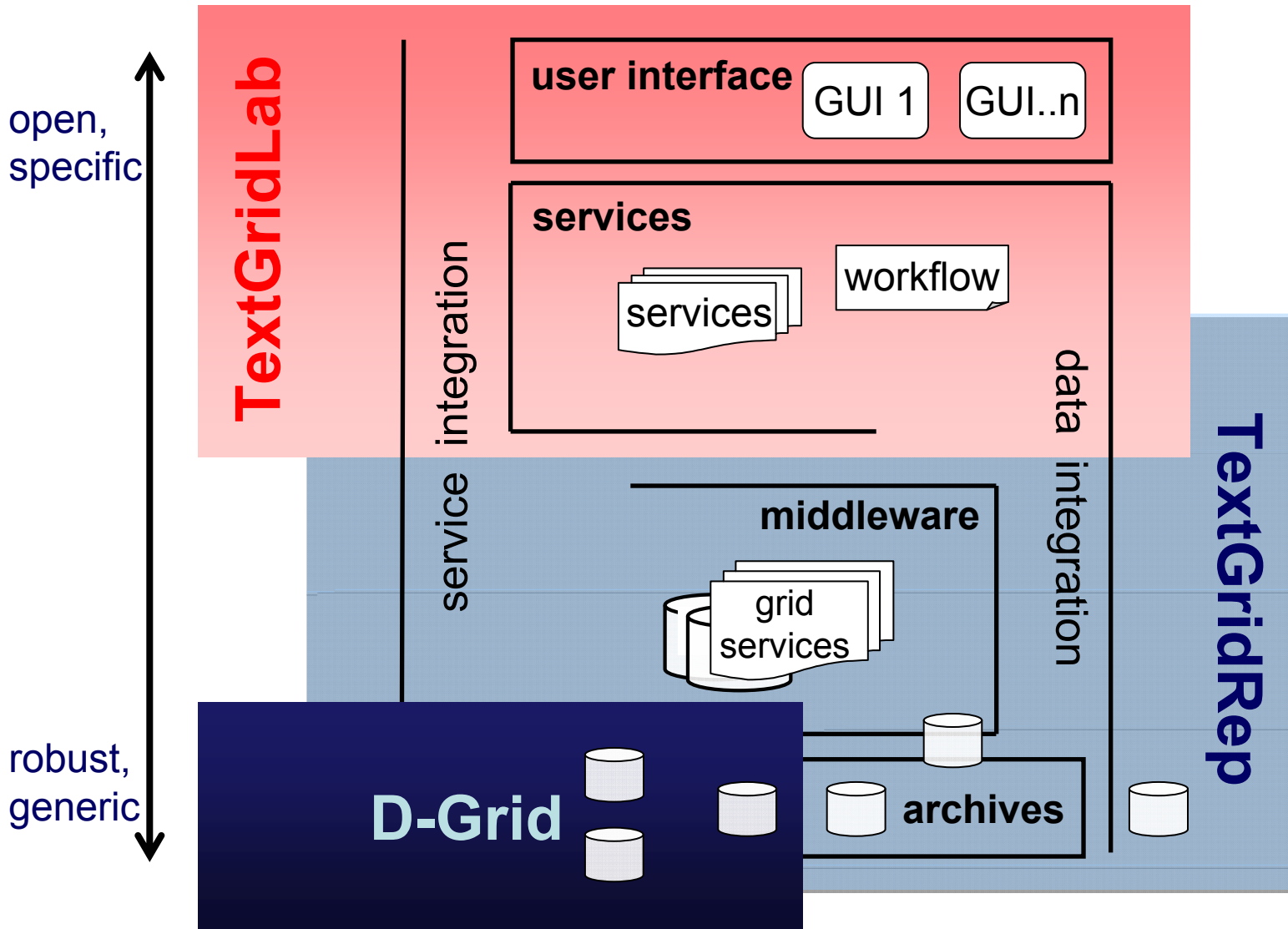
- A single point of entry to the virtual research environment
- Integrated access to tools and services via user-friendly software
- Continuously advanced, enhanced with new features and released regularly for public beta testing

- **TextGrid Repository**

- Long-term archive for research data in the humanities, embedded in a grid infrastructure
- Ensures long-term availability and access to its research data as well as interoperability



The Developer's View



TextGridLab: Tools und Services

(v1.0 – June 2011)

Tools in the TextGridLab:



XML Editor



Project and User Management



Text-Image-Link Editor



Projektbrowser / Navigator



Dictionary Search Tool



Search Tool



Workflow Tool



Metadata Editor



Text Publisher Web



Aggregation Composer



Lemmatizer



Upload Tool

Further services:



Streaming Editor



Service Registry



Tokenizer



Sort Tool

Tool	Available in current Beta, full functionality and documentation in v1.0 (June 2011)
Tool	In development, full functionality in v1.0 (June 2011)
Tool	Additional tools and services with constricted support, already available in current Beta version

TextGridLab: Tools und Services

(v1.0 – June 2011)

Tools in the TextGridLab:



XML Editor



Pro



More than 450 testers

Manny research groups that want to use TextGrid:
Blumenbach-Online, Archaeo18, Fontane-Edition, etc.

Fur



S

Service Registry



Tokenizer



Sort Tool

Tool

available in current
a, full
functionality and
documentation in
) (June 2011)

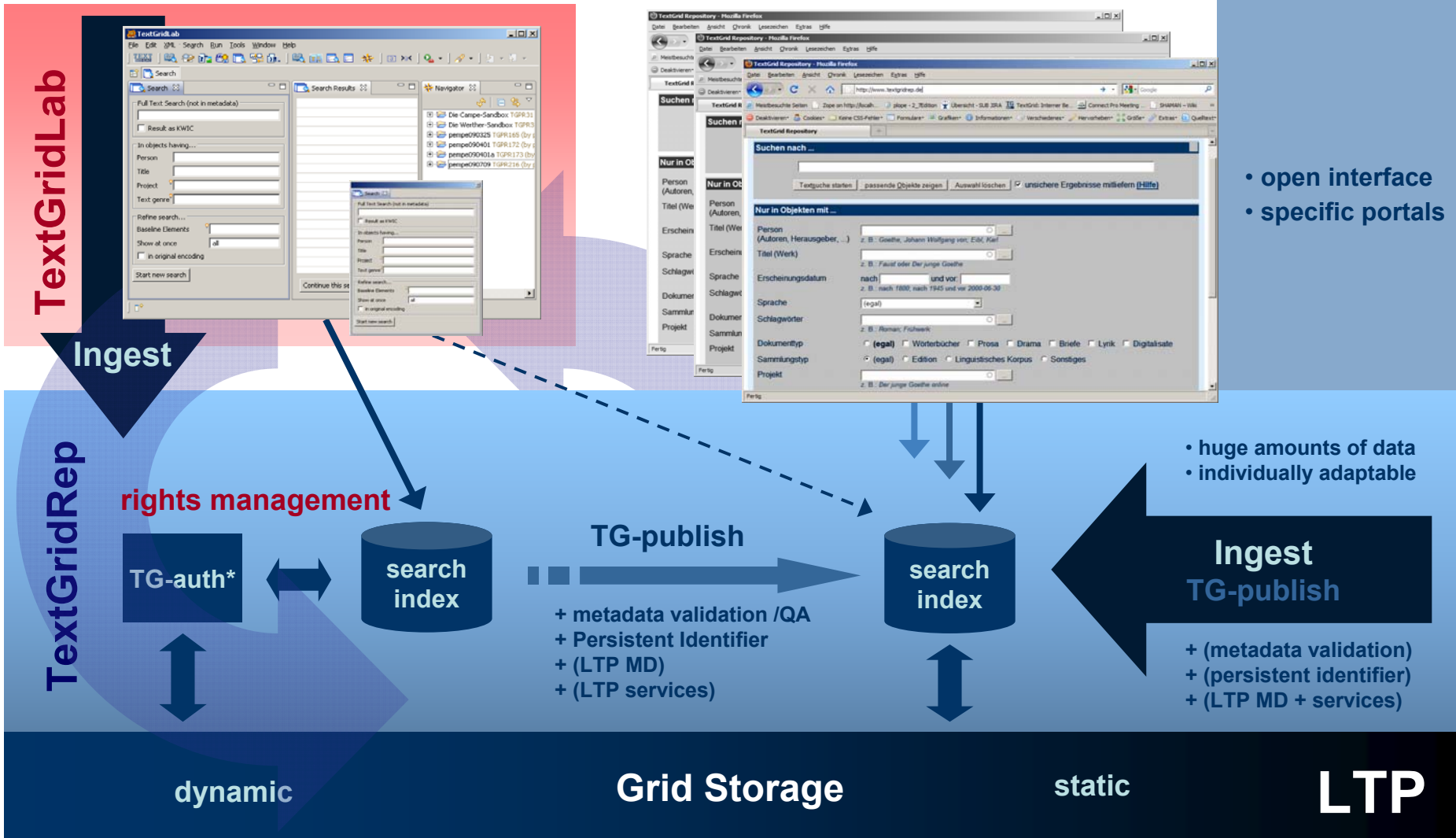
development,
functionality in
) (June 2011)

Additional tools
and services with
constricted
support, already
available in current
Beta version

TextGridRep: Overview

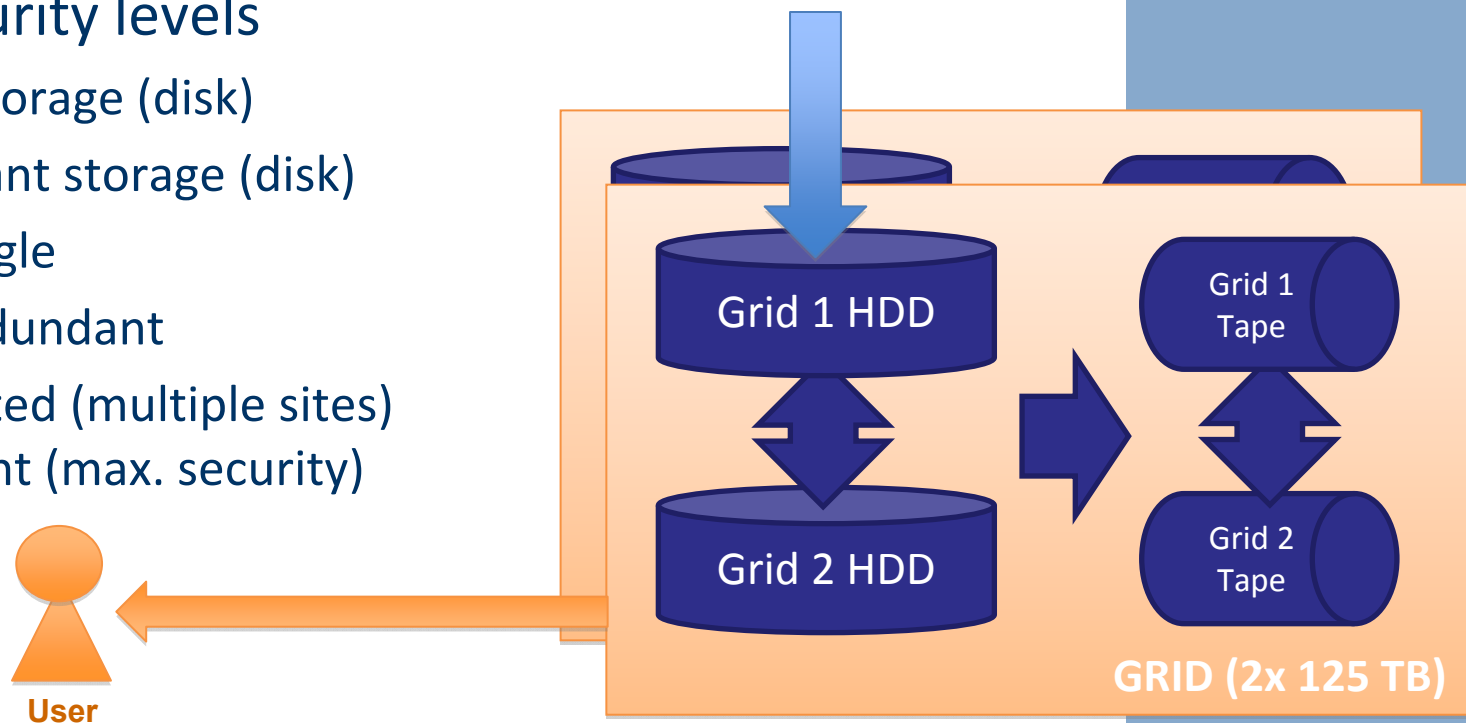
Eclipse Frontend

Portal(s)

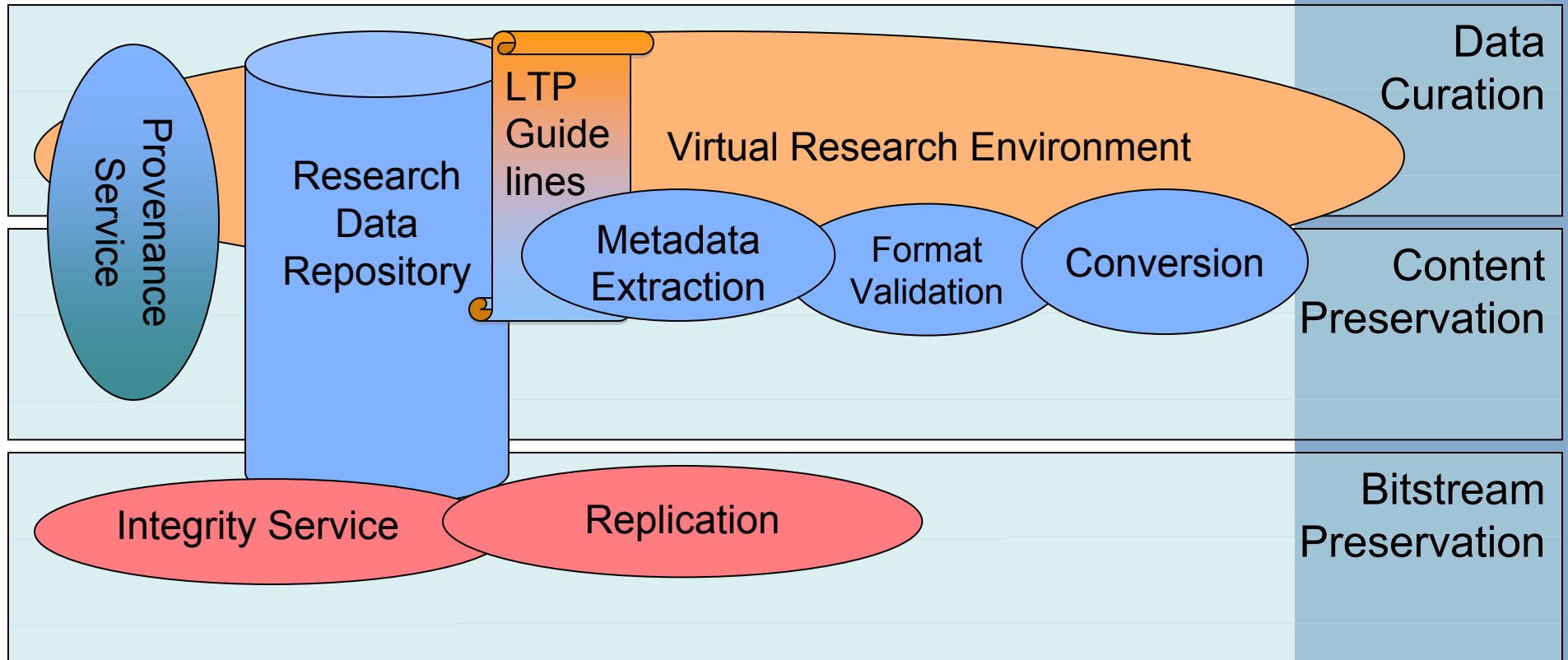


TextGridRep: Grid Storage

- 275 TB Grid Storage, 250 TB Tape Storage
- First instance: bitstream preservation - redundant storage and tape single (B+C), PID, LTP metadata, ...
- In 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: WissGrid Service Framework



WissGrid =

Community =

D-Grid / Infrastructure Provider =



TextGridRep: LTP Levels / SLA

	Bitstream Preservation for max. 10 years (→ DFG-Guidelines)	Bitstream Preservation (Long-Term)	Additional Higher-Value LTP Services and SLAs
Redundant Storage and Tape single	€	€€	€€€
Redundant Storage and Tape redundant	€€	€€€	€€€€
+ distributed (multiple sites)	€€€	€€€€	€€€€€

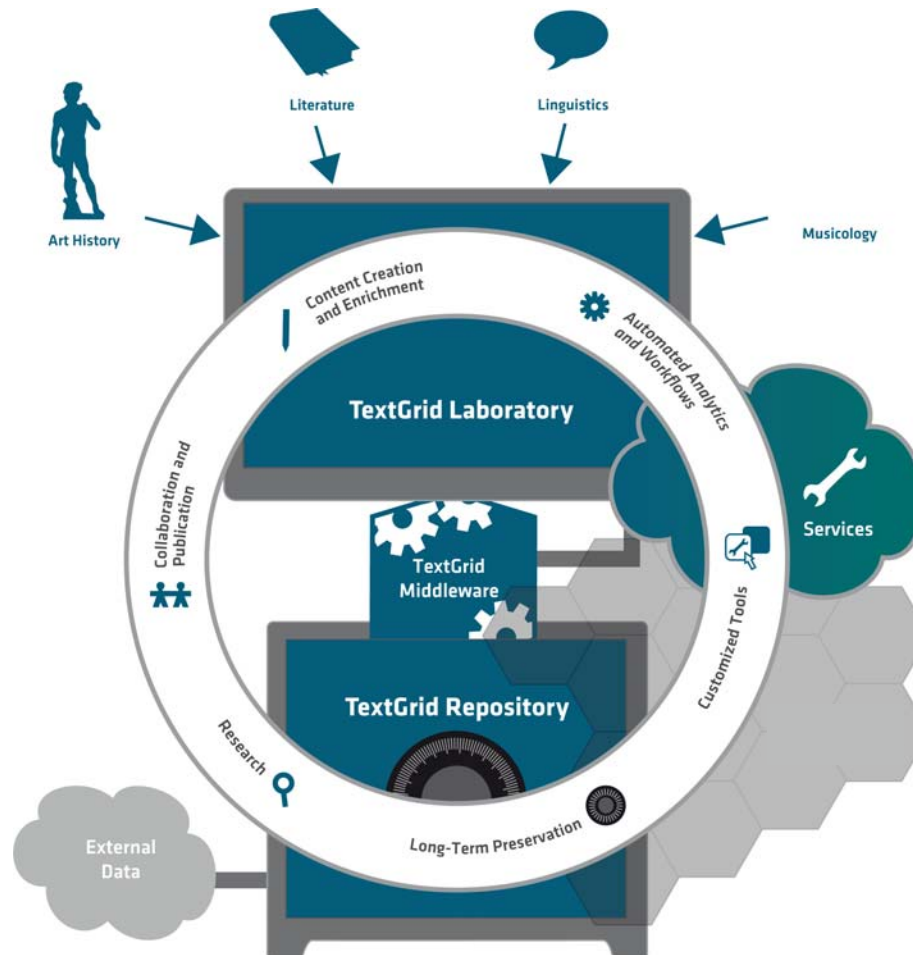
Metadata and Persistent Identifier

- **structural markup** tagging of content (paragraphs, page breaks ...)
 - **bibliographical metadata** description of TextGrid objects (title, creator ...), using controlled vocabulary (Getty Thesaurus, GND, ...)
 - **relations** (relations between objects, persons, things, etc.)
 - **technical metadata** → management of data, persons, projects, ...
- **Implementation of a new metadata schema in TG v1.0**

PIDs for TextGrid

- **handle-service of the GWDG:** <http://handle.gwdg.de/PIDservice/>
(in collaboration with and for the MPG)
- reliable and institutionally supported
- GWDG – Partner in *EPIC - European Persistent Identifier Consortium*
- <http://www.pidconsortium.eu/>
- DOI: commercial service, additional costs
URN: problems with scalability and granularity

Infrastructure for the eHumanities



Astronomy
Climate Research
High-Energy Physics
Medicine
Humanities

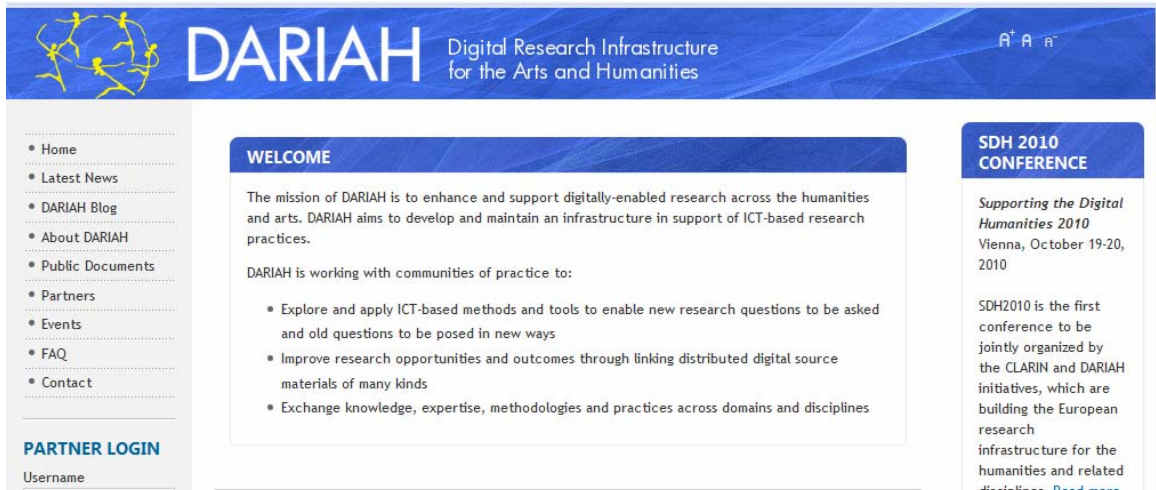
1. Sustainability
2. Blueprints for new VREs
3. Research Data Curation



Support, Monitoring, Accounting, Billing, etc.

Summary & Outlook

- TextGrid is a developing project, not so easy for humanities scholars to invest time and effort in IT issues
- Need time to agree on a common culture between scholars and IT experts (discussion, terminology, etc.)
- Sustainability of VREs after project funding is not ensured (but for World Data Centers also not)
- TextGrid is part of DARIAH – ESFRI project, which will start with the construction phase next year, and German Grid Initiative: ***some kind of sustainability!***



The screenshot shows the DARIAH website header with the logo and tagline "Digital Research Infrastructure for the Arts and Humanities". The main content area is divided into three columns:

- Left Column:** A navigation menu with links for Home, Latest News, DARIAH Blog, About DARIAH, Public Documents, Partners, Events, FAQ, and Contact. Below the menu is a "PARTNER LOGIN" section with a "Username" input field.
- Middle Column:** A "WELCOME" section. It states the mission of DARIAH and lists three key activities: exploring ICT-based methods, improving research opportunities through linking digital sources, and exchanging knowledge across domains.
- Right Column:** A section for the "SDH 2010 CONFERENCE" in Vienna, October 19-20, 2010. It describes the conference as the first jointly organized by CLARIN and DARIAH, aimed at building European research infrastructure.



TEXT
GRID

**** Thank You ****

Questions, Comments ?



**e-Infrastructures
for Cultural Heritage**

Brussels Conference 29 October 2010