

TextGrid & Linked Open Data

Wolfgang Pempe, SUB Goettingen

SPONSORED BY THE



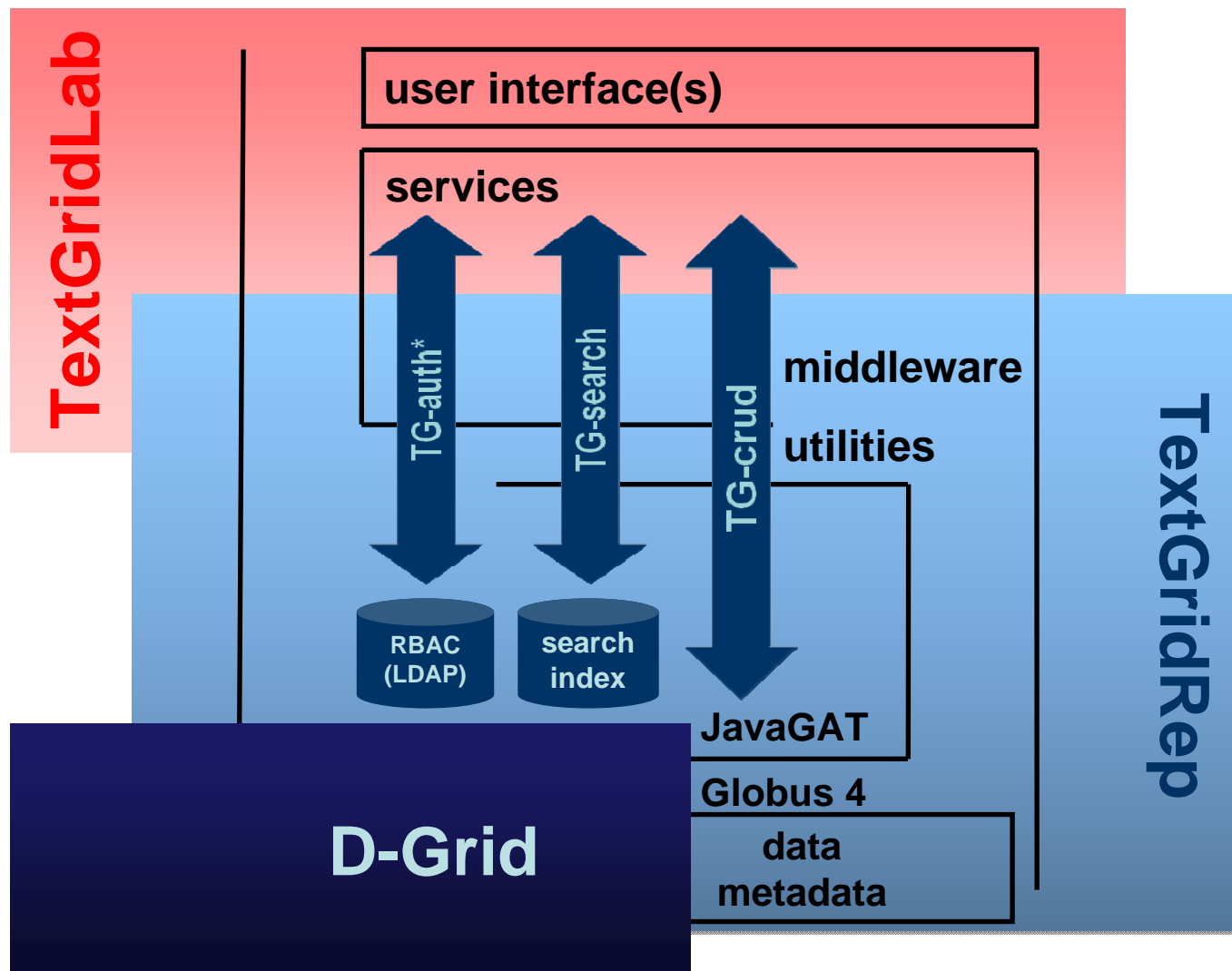
Federal Ministry
of Education
and Research



Questions

- How to provide human and machine readable access to the contents of the TextGrid Repository?
- Link up with the Linked Open Data Cloud?
 - How – technically?
 - How to provide meaningful annotations to existing resources?
- **some ideas**

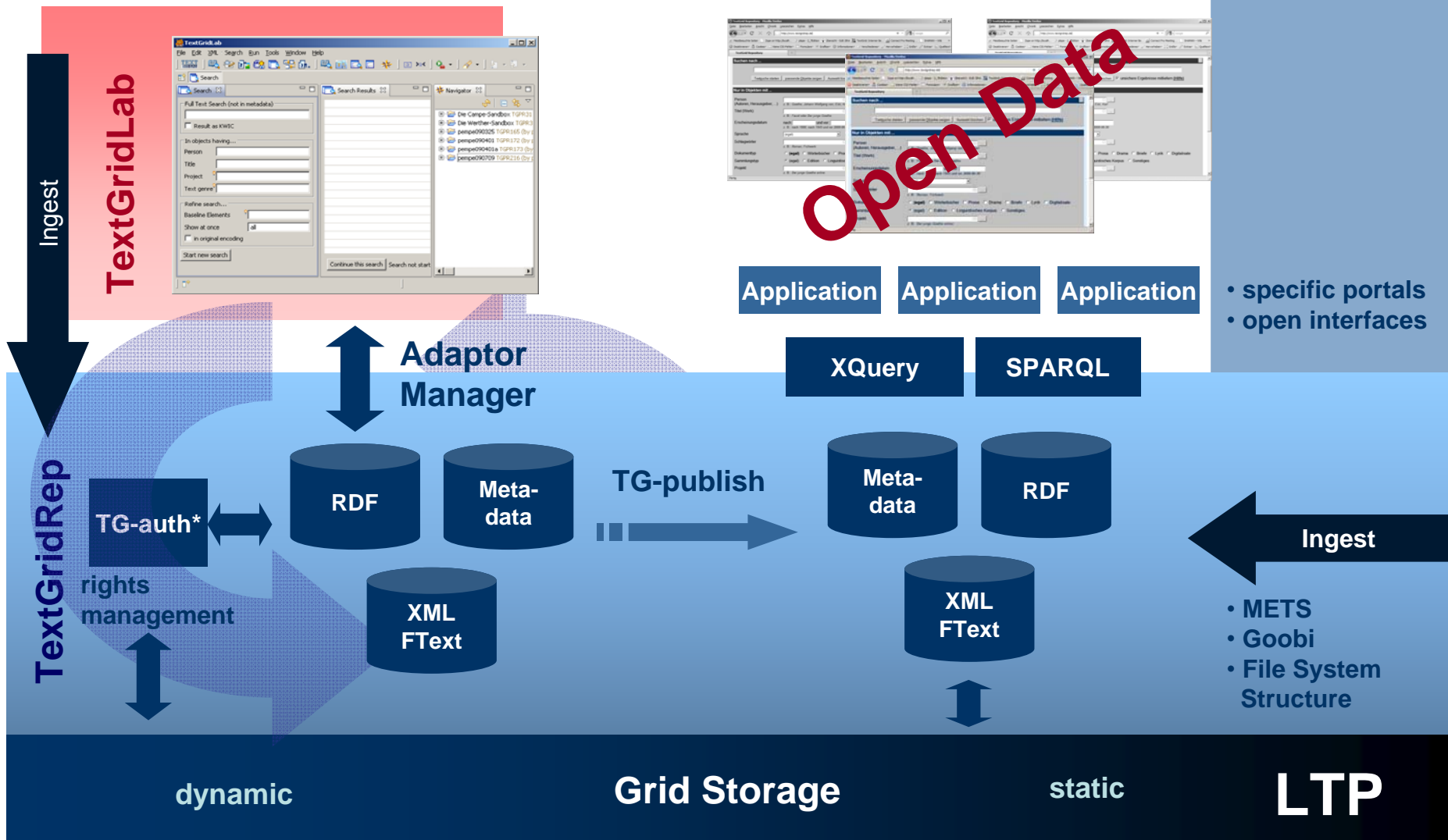
The TextGrid Architecture



TextGrid Infrastructure

Eclipse Frontend

Portal(s)



Semantic Framework – Components

Current Status

- RDF Triple Store
 - sesame 2.4.0 (SPARQL 1.1)
 - *for administrative purposes only*
- Controlled Vocabularies
 - Getty Thesaurus of Geographic Names
 - PND (Personennamendatei, LD-Service of the DNB) – as test
 - *Metadata Editor only*
- NER Service (GATE)
 - via Workflow Manager
- Resolving Service
 - wrapper on top of several infrastructure services
 - various representations of a TextGrid object
 - *experimental*

Linked Data (1/2)

Tim Berners-Lee - Design Issues, Linked Data:

(<http://www.w3.org/DesignIssues/LinkedData.html>)

1. Use URIs as names for things
 - ✓ the TG pattern: `textgrid:h5fs.0` (= name?)
2. Use HTTP URIs so that people can look up those names.

✗ not really, but several RESTful services:

<http://textgridlab.org/1.0/tgsearch/info/textgrid:h5fs.0>

<http://textgridlab.org/1.0/tgcrud/rest/textgrid:h5fs.0/data>

experimental:

<http://134.76.20.86:9090/multiresolver/textgrid:h5fs.0>

Linked Data (2/2)

3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

✓ - but useful?

```
<rdf:Description rdf:about="textgrid:h5fs.0">  
  <inProject xmlns="http://textgrid.info/relations#" rdf:resource="textgrid:project:TGPR-d6b7f..." />  
</rdf:Description>
```

4. Include links to other URIs. So that they can discover more things.

✓ / ✗ only in some cases:

```
<edition>  
  <isEditionOf>textgrid:h5ft</isEditionOf>  
<agent role="author" id="pnd:118729241">Lovecraft,  
Howard Phillips</agent>  
<formOfNotation>Latin</formOfNotation>  
</edition>
```

So what to do next?

- Use (HTTP) URIs:
 - Implement a simple RESTful interface as single entry point to the TextGrid Repository (wrapper) – include links in RDF-output
 - Multiple representations of one object - according to type of interaction and particular object type
 - Splash pages (→ [Example](#))
- Provide useful Information, links to other objects
 - Scholars don't want to type in additional metadata / relations
 - Deeper and **seamless integration** with the TextGrid Lab and tools, **automatic triple extraction** (XML-Editor, Link-Editors, NER, ...)
 - Model (almost) everything in RDF, **RDFize Metadata?**

... and even more

- Provide a virtual endpoint for addressing “names”, <http://textgridlab.org/getStuffAbout/Fontane>
 - trigger metadata-search for “fontane” ([Example](#))
 - gather all related information (relations, metadata)
 - list all related objects (links)
 - model everything in RDF (optional: HTML / JSON ...)
- Separate, project-specific RDF repository instances
- Public SPARQL interface
- More Controlled Vocabularies: CERL Thesaurus, PDR (BBAW, DARIAH)
- NE-Management: CONE (MPDL, DARIAH)
- Additional services for information extraction
 - WisNetGrid services? (e.g. Ontology Workbench)
 - More GATE? (<http://gate.ac.uk>)



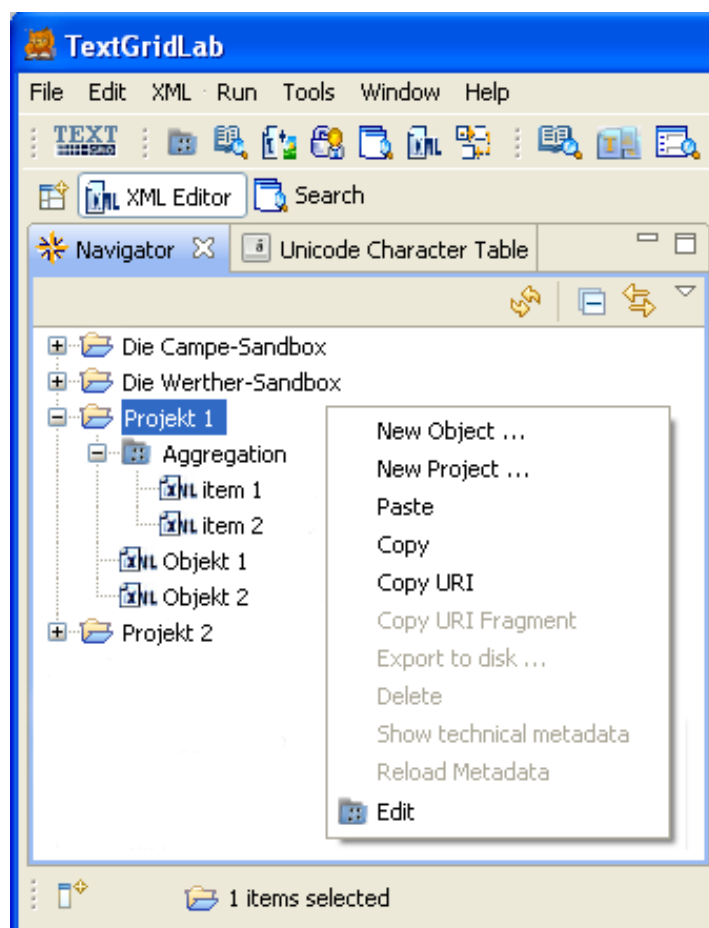
**Thank you
for your attention.**

—

**What do you think?
Any ideas,
suggestions?**

Logical vs. Physical Structure

- Generic hierarchy: Project – Aggregation – Object
(*Project* comparable with *Context* in eSciDoc or *Bucket* in S3)



- Revisions / Versions:
only the latest revision of
an object is displayed

But ...

- Physically, the world is flat
- Everything is an object
 - Project (-File)
 - Aggregation (-File)
 - “Simple” Object (-File)

Object Semantics

- Semantic hierarchy on the object level:
item – edition – work (mandatory for publication)
- Containers for bibliographic metadata

