TextGrid: der GWES im Einsatz

Grid Workflow Workshop 2009
TU Berlin, 5./6. März 2009

Martin Haase
DAASI International GmbH
Martin.Haase@DAASI.de

DAASI International
Directory Applications for Advanced Security and Information Management

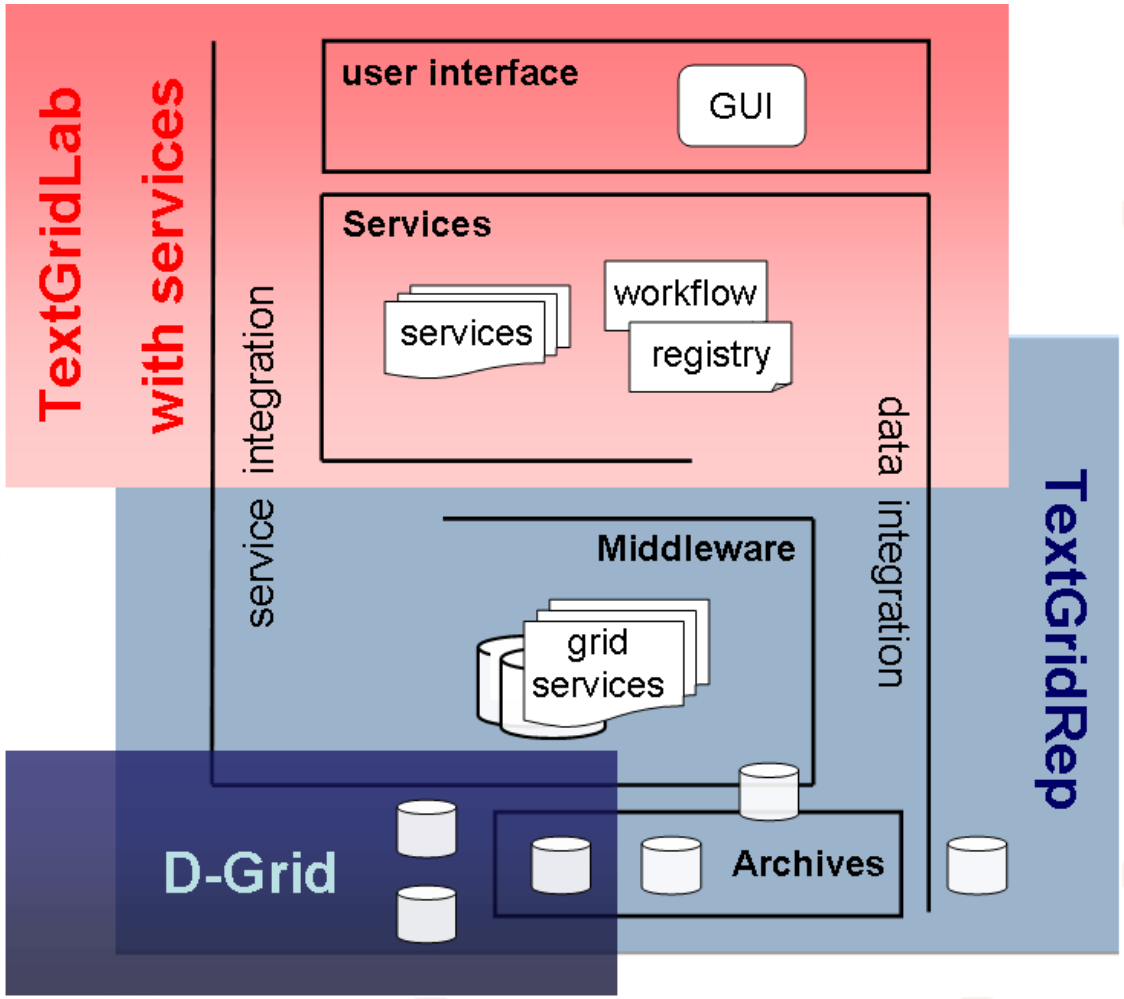
Überblick

- GWES? Grid Workflow Execution Service, die Workflow Engine von Fraunhofer FIRST
- Hintergrund: TextGrid und Anforderungen an Workflow
- Prototyp im Einsatz
- Ausblick: Erweiterungen für Benutzerfreundlichkeit

TextGrid: Projekthintergrund

- D-Grid-Projekt, Laufzeit 10/2005 – 05/2009
- „Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung - ein Community-Grid für die Geisteswissenschaften“
- Partner: TU Darmstadt, SUB Göttingen, IDS Mannheim, Universität Trier, FH Worms, Universität Würzburg, DAASI International GmbH, Saphor GmbH

TextGrid: Architektur



Details zur TextGrid-Architektur

- TextGridRep
  - Storage Grid (GT4/GAT) und angebundene Archive
  - Middleware
    - TG-auth*: Authentifizierung (Shibboleth) und Autorisierung (RBAC)
    - TG-crud: Operationen Create, Read, Update, Delete für Datenobjekte mit Registrierung in sämtlichen TG-Services
    - TG-search: Suche auf Original- und normalisiertem Format (TEI-XML, TextGrid Baseline Encoding)
    - TG-Log: zentrales Logging
- Services: Tokenizer, Lemmatizer, Workflow-Engine, Streaming Editor, Sortierer, ...
- TextGridLab: GUI (Eclipse Rich Client), Frontend zu Services bzw. TextGridRep

Rahmenbedingungen für Workflow

- Operationen auf TextGridObjekten, bestehend aus
  - Nutzdaten (Text, XML, TEI-XML, JPEG, ...)
  - Metadaten
    - Agent (author, contributor, translator, editor, ...)
    - Title, subtitle, date, language, page, volume, ISBN, ...
    - Administratives: project, URI, format
    - Relationen zu anderen Objekten (isVersionOf, hasSchema, ...)
    - Anwendungsspezifisches (custom element)
- Infrastruktur baut auf Web Services auf (SOAP und REST)
- Services akzeptieren URIs oder Nutzdaten direkt
- Services bekommen Session-Token und Log-Token
- CRUD akzeptiert Base64-kodierte Nutzdaten oder MTOM

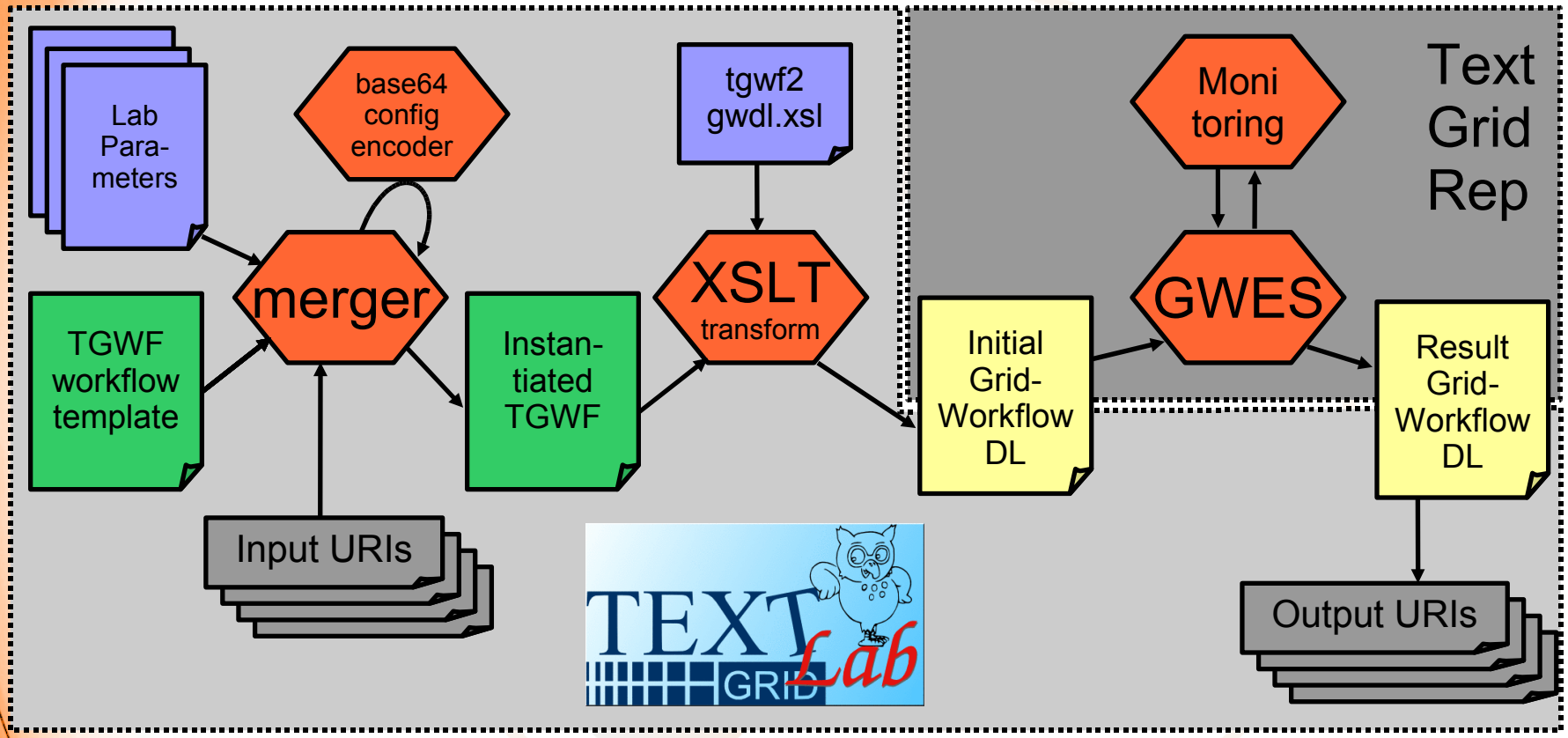
Anforderungen an Workflow

- Batch Processing: viele TextGridObjekte verarbeiten
- Front-End im TextGridLab zu den Services
- TextGridObjekte lesen, verarbeiten, neu erstellen → regelgesteuerte Transformation der Metadaten für zu erstellende Objekte
- Workflow Editor soll den Endbenutzern des TextGridLab abnehmen:
  - Routinearbeit (z.B. kopieren des Session-Tokens, Gerüst zum Einbinden des CRUD, Base64-Kodierung von Konfigurationsdaten)
  - Komplexität von Workflow-Beschreibungssprachen

Prototyp im Einsatz

- Workflow Engine: Grid Workflow Execution Service (GWES) von Fraunhofer FIRST
- Erstellung einer auf die Bedürfnisse von TextGrid abgestimmten Workflow-Beschreibungssprache (TGWF)
- Bearbeitung der TGWF-Dokumente im TextGrid-XMLEditor
- Auswahl von Eingabeobjekten über die vorhandenen Mechanismen im TextGridLab
- Konfigurationsdaten der Services eingebettet in TGWF
- Metadatentransformation mit dem TextGrid-Streaming-Editor, XSLT-Stylesheet eingebettet in TGWF
- Auf Knopfdruck Erzeugung eines Workflow-Dokuments für den GWES mit Einfügung aller Variablen und XSL-Transformation in GridWorkflowDL

Ablaufschema

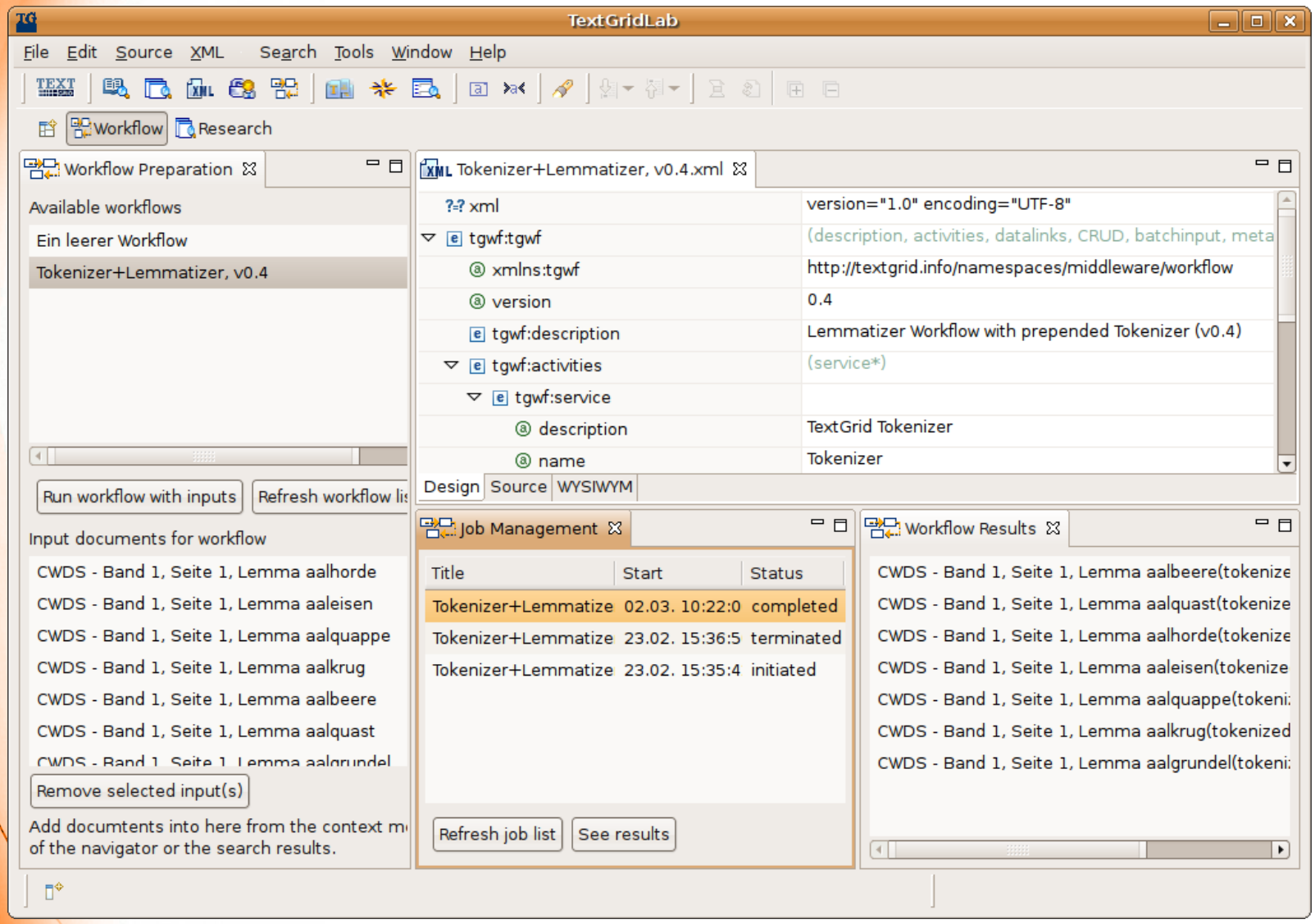


TGWF (vereinfacht)

```
<tgwf>
  <activities>
    <service serviceID="tok" wsdl="..tok.wsdl"/>
    <service serviceID="lem" wsdl="..lem.wsdl"/>
  </activities>
  <datalinks>
    <link from="crudRead" to="tok"/>
    <link from="tok" to="lem"/>
    <link from="lem" to="crudCreate"/>
  </datalinks>
  <CRUD instance="" sessionID="" log=""/>
  <batchinput/>
  <metadatatransformation>
    <xsl:transform>...</xsl:transform>
  </metadatatransformation>
  <inputconstants>
    <activity serviceID="tok">
      <const name="config">...</const>
    </activity>
  </inputconstants>
</tgwf>
```

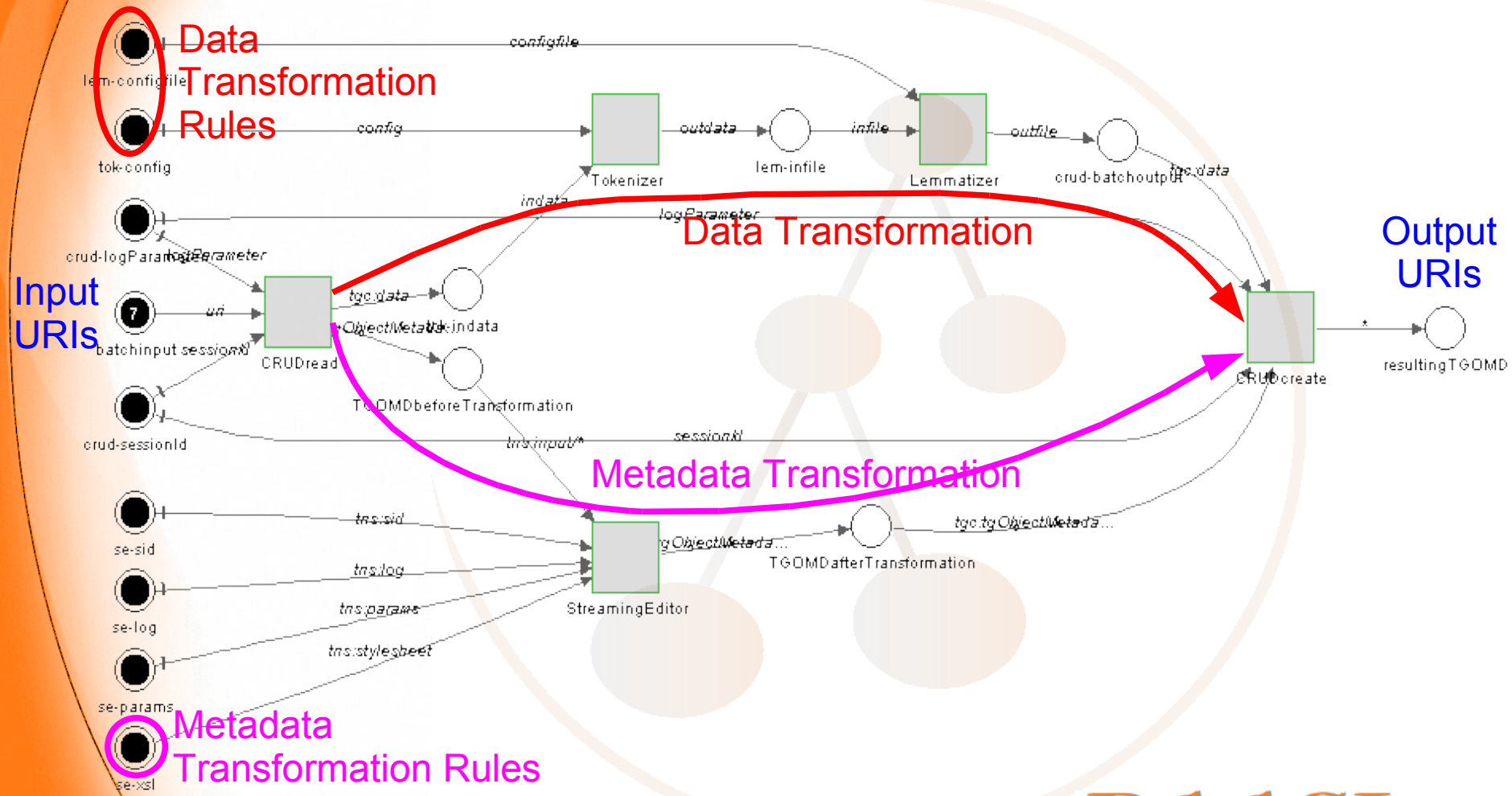

DAASI International
Directory Applications for Advanced Security and Information Management

TextGridLab: Workflow Perspective



Directory Applications
for Advanced Security
and Information Management

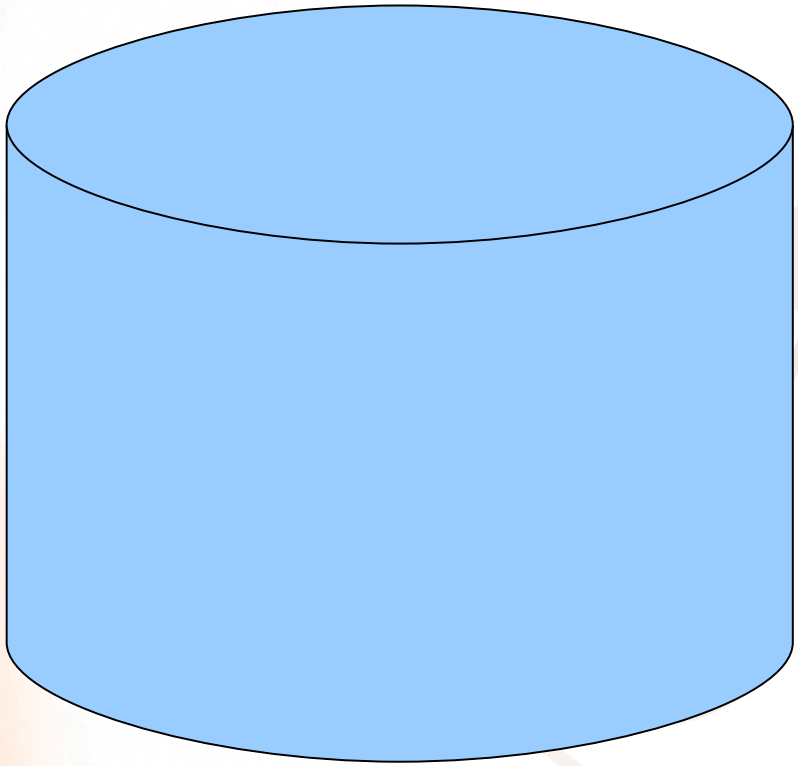
Petrinetz im GWES



Mehr Benutzerfreundlichkeit

- Momentan: Verwendung des TextGridLab-eigenen XMLEditors für das TGWF-Dokument.
- Für unerfahrene Benutzer zu komplex:
  - Informationen über vorhandene Services sind evtl. nicht bekannt bzw. WSDLs nicht vertraut → Service Registry
  - Eingebettete Konfigurationsparameter der Services überfrachten das Dokument → spezieller Editor sowie vorgefertigte Konfigurationsbeispiele in Registry verlinkt
  - XSLT zur Metadatentransformation evtl. nicht vertraut → GUI für die wichtigsten Operationen mit rudimentärem Stylesheet-Generator (abh. v. Metadaten!)
  - Links unübersichtlich im XML → langfristig grafisches Drag&Drop für Services und Links

Erweiterung: Service Registry



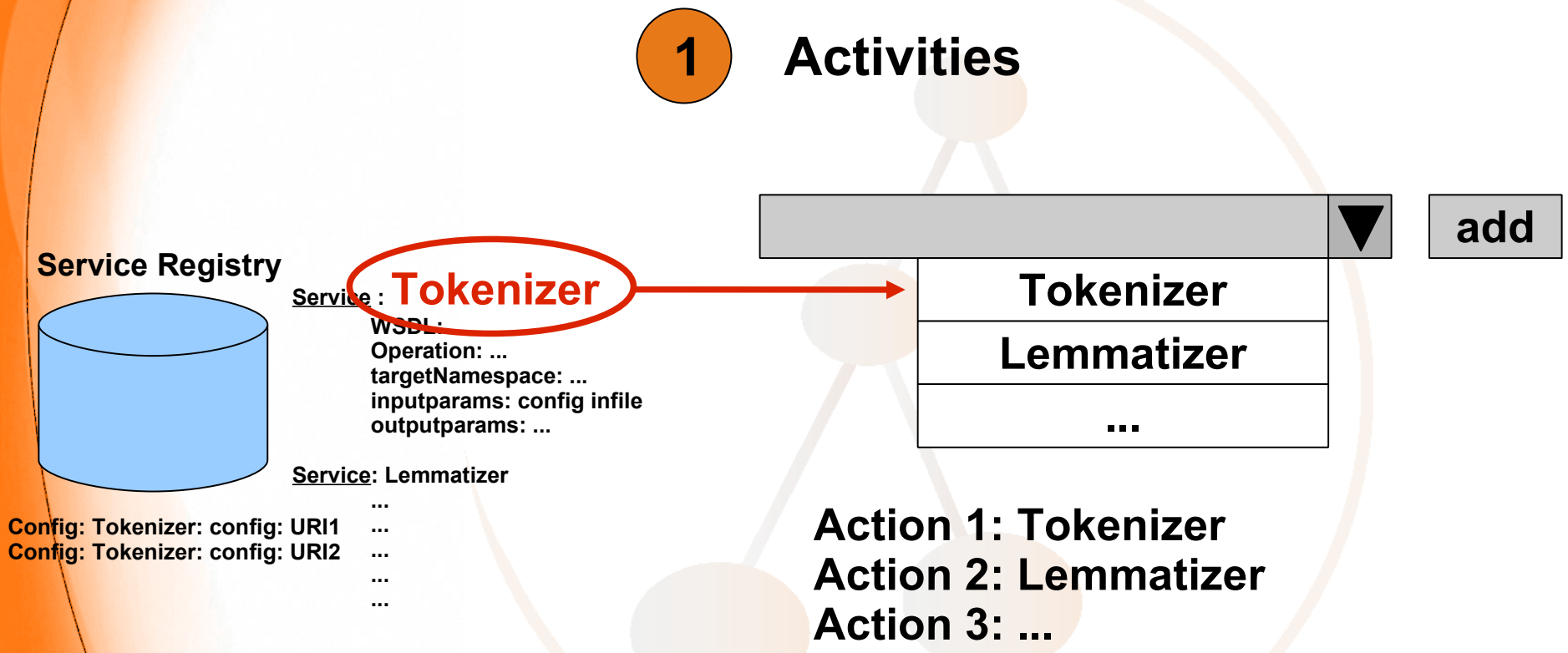
Service : Tokenizer
- WSDL: ...
- Operation: tokenize
- targetNamespace: ...
- inputparams: config, infile
- outputparams: outfile

Service: Lemmatizer
- ...
- ...
- ...
- ...
- ...

Config:Tokenizer: config: URI1
Tokenizer: config: URI2

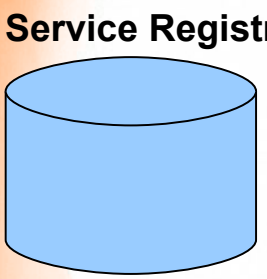
DAASI International
Directory Applications for Advanced Security and Information Management

Erweiterung: Service-Auswahl



Erweiterung: Links einfügen

2 Data links



Service Registry

Service : Tokenizer
WSDL: ...
Operation: ...
targetNamespace: ...

inputparams: config
outputparams: ...

Service: Lemmatizer

Config: Tokenizer: config: URI1 ...
Config: Tokenizer: config: URI2 ...
...
...
...

infile

FROM		TO	
Service	Param.	Service	Param.
CRUD	Input	Tokenizer	Infile
Tokenizer	Outfile	Lemmatizer	Indata
Lemmatizer	Outdata	CRUD	Output

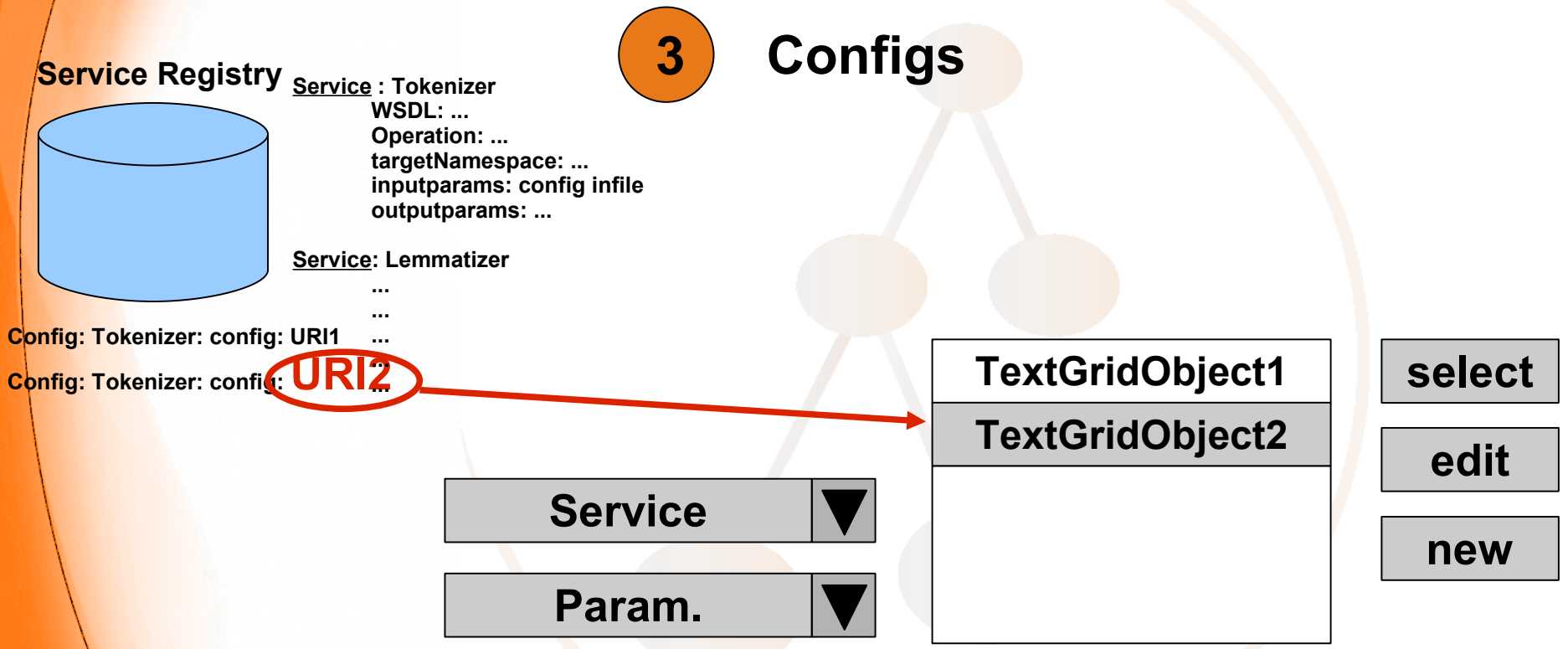
FROM Service ▼ Param. ▼

TO Service ▼ Param. ▼

add new link

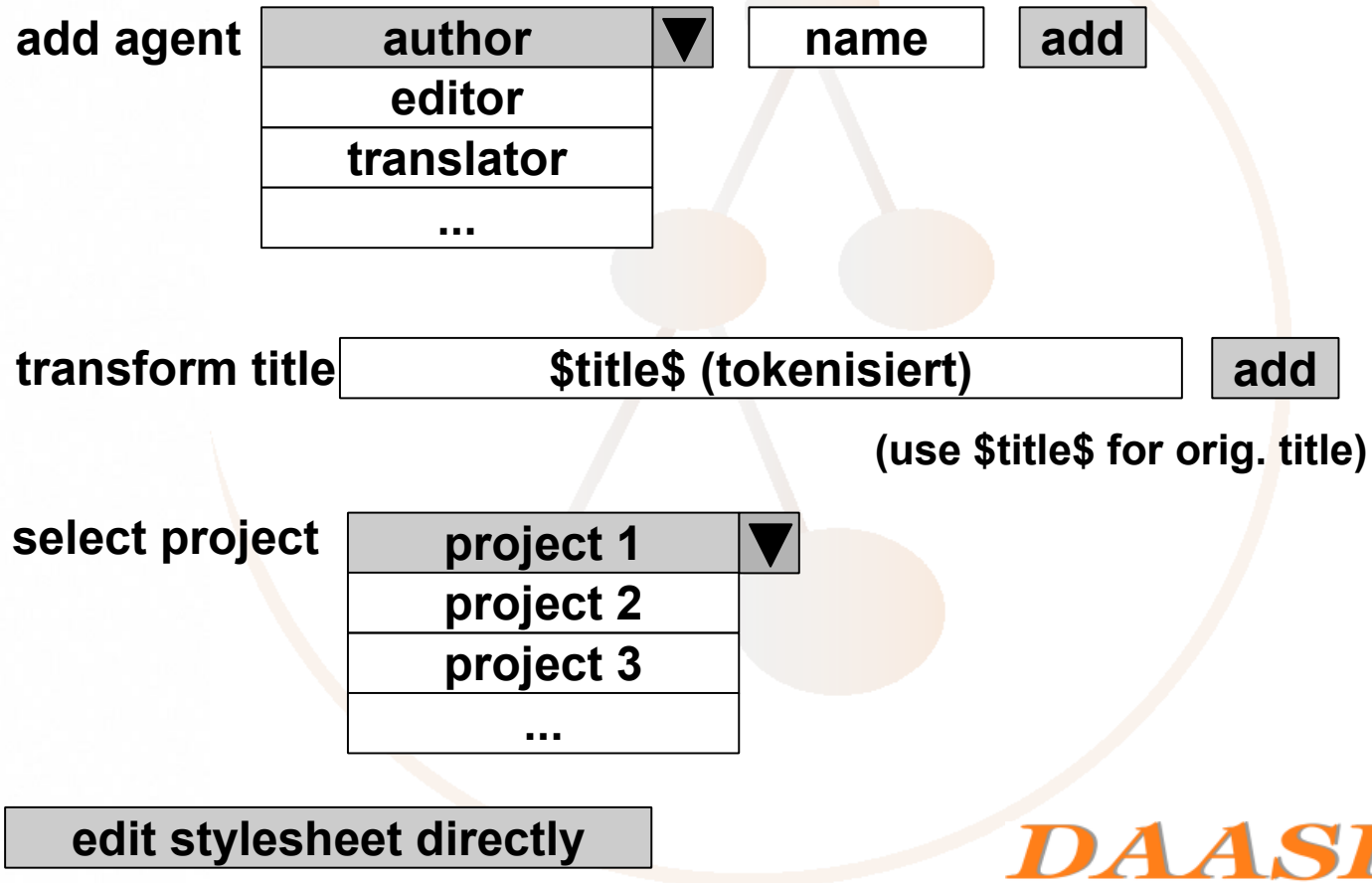
DAASI International
Directory Applications for Advanced Security and Information Management

Erweiterung: Konfigurationsauswahl



Erweiterung: Metadatentransformation

4 Metadata transformation



DAASI International
Directory Applications for Advanced Security and Information Management

Ausblick

- Stand: Prototyp, Community-Tests laufen
- Nutzer brauchen XML- und minimale XSLT-Kenntnisse
- Umsetzung für mehr Benutzerfreundlichkeit bis Projektende angestrebt – auch grafisches Drag&Drop?
- Weitere Services beispielhaft einbinden → Registry

Vielen Dank für Ihre Aufmerksamkeit!

- Fragen?

- DAASI International GmbH
  - www.daasi.de
  - Info@daasi.de