

25 Nov. 2014

TextGrid-Nutzertreffen
TextGrid und Personendaten-Repository



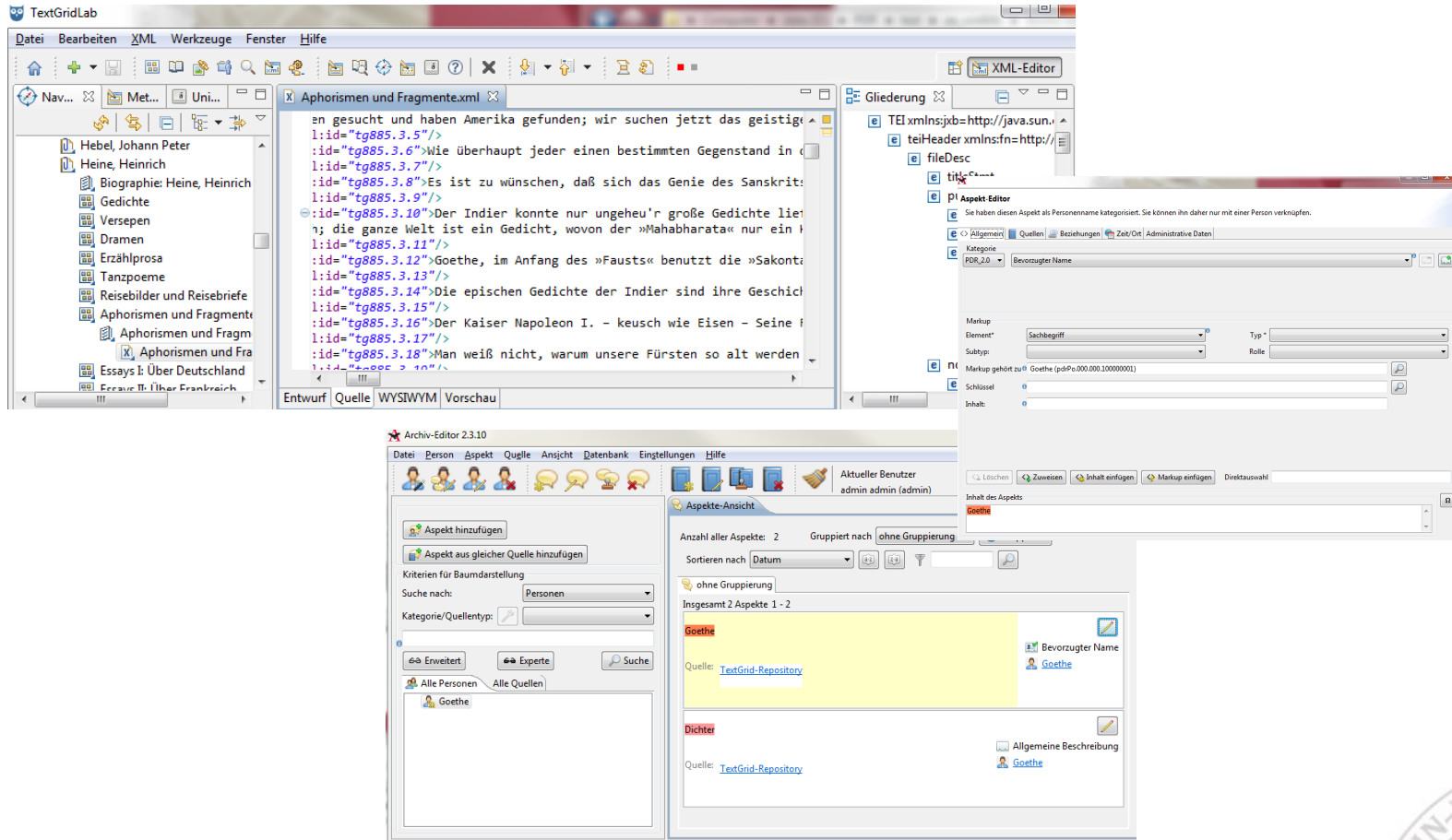
BERLIN-BRANDENBURGISCHE
AKADEMIE DER WISSENSCHAFTEN



TextGrid und Personendaten-Repository. Ideen zur Verbindung von Text- und Personendaten

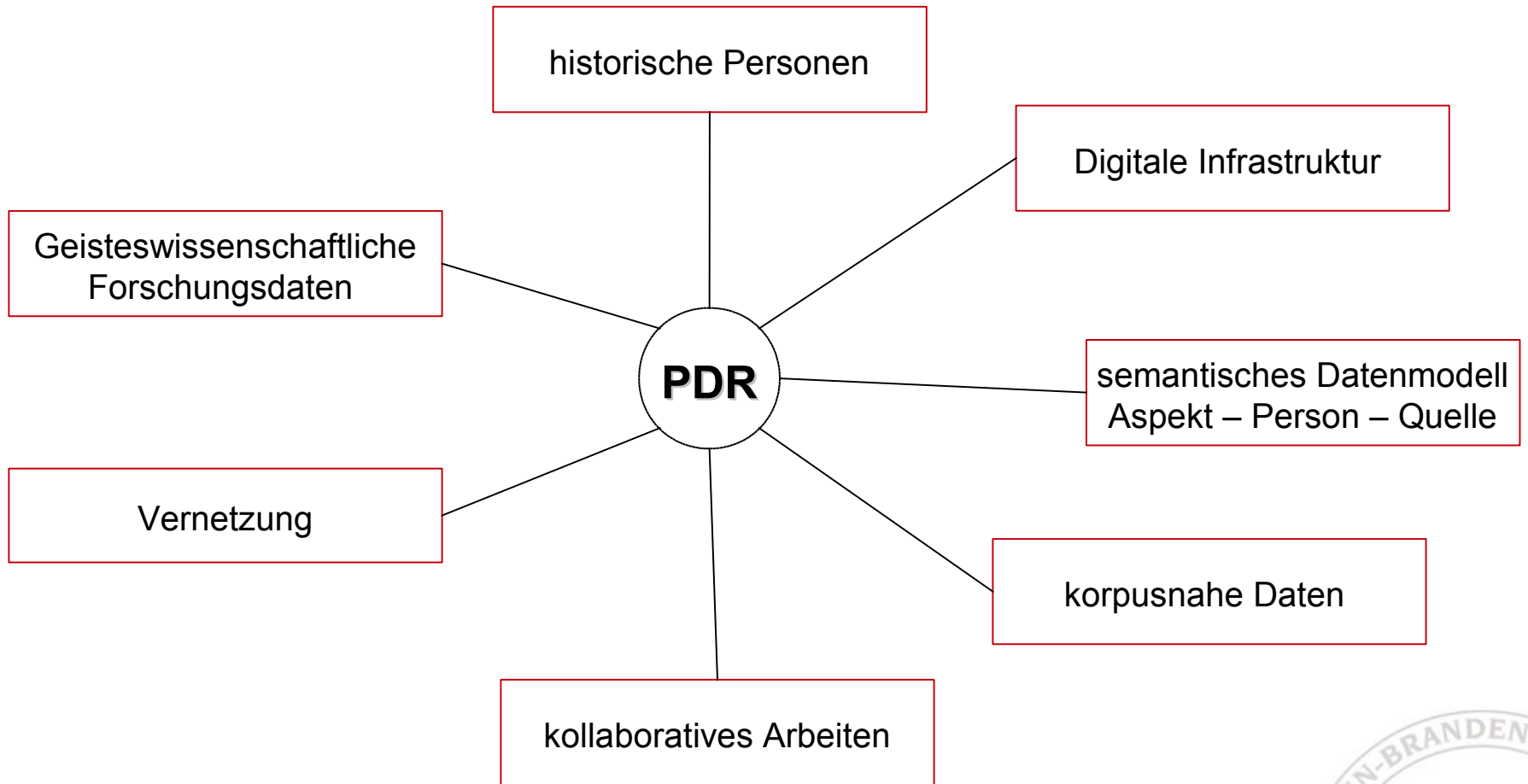
- I Einleitung: Bedarf
- II Das Personendaten-Repository
- III Interaktionen zwischen TextGrid und PDR
- IV Beispiele
- V Weitere Kombinationen
- VI Zusammenfassung

- Editionsprojekte:
reichhaltiger Bestand an Personendaten
- Projekt „Personendaten-Repository“ (BBAW):
digitaler Lösungsansatz für die Arbeit mit
Personendaten
- Welche Möglichkeiten gibt es, die Vorteile von
TextGrid und PDR zu verbinden?



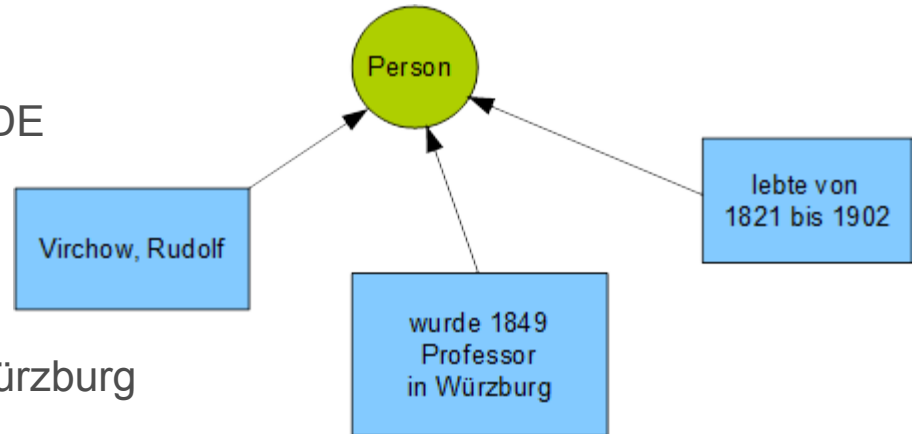
The screenshot displays the TextGridLab interface. The main window shows an XML editor with a document titled 'Aphorismen und Fragmente.xml'. The XML content includes several entries with IDs like 'tg885.3.5' through 'tg885.3.18'. A sidebar on the left shows a navigation tree with categories like 'Hebel, Johann Peter', 'Heine, Heinrich', and 'Aphorismen und Fragmente'. A 'Gliederung' (Outline) pane on the right shows the XML structure. A 'P1 Aspekt Editor' dialog is open, showing fields for 'Markup', 'Element', 'Subtyp', 'Rolle', and 'Inhalt'. Below this, the 'Archiv-Editor 2.3.10' is visible, showing a search interface for 'Aspekt-Ansicht' with filters for 'Personen' and 'Goethe'. The search results show two aspects for 'Goethe': 'Goethe' (Source: TextGrid-Repository) and 'Dichter' (Source: TextGrid-Repository).





Aspekt 1

- Quelltext = Virchow, Rudolf
- Semantische Kategorie = Normname_DE
- Vorname = Rudolf (Markup)
- Nachname = Virchow (Markup)

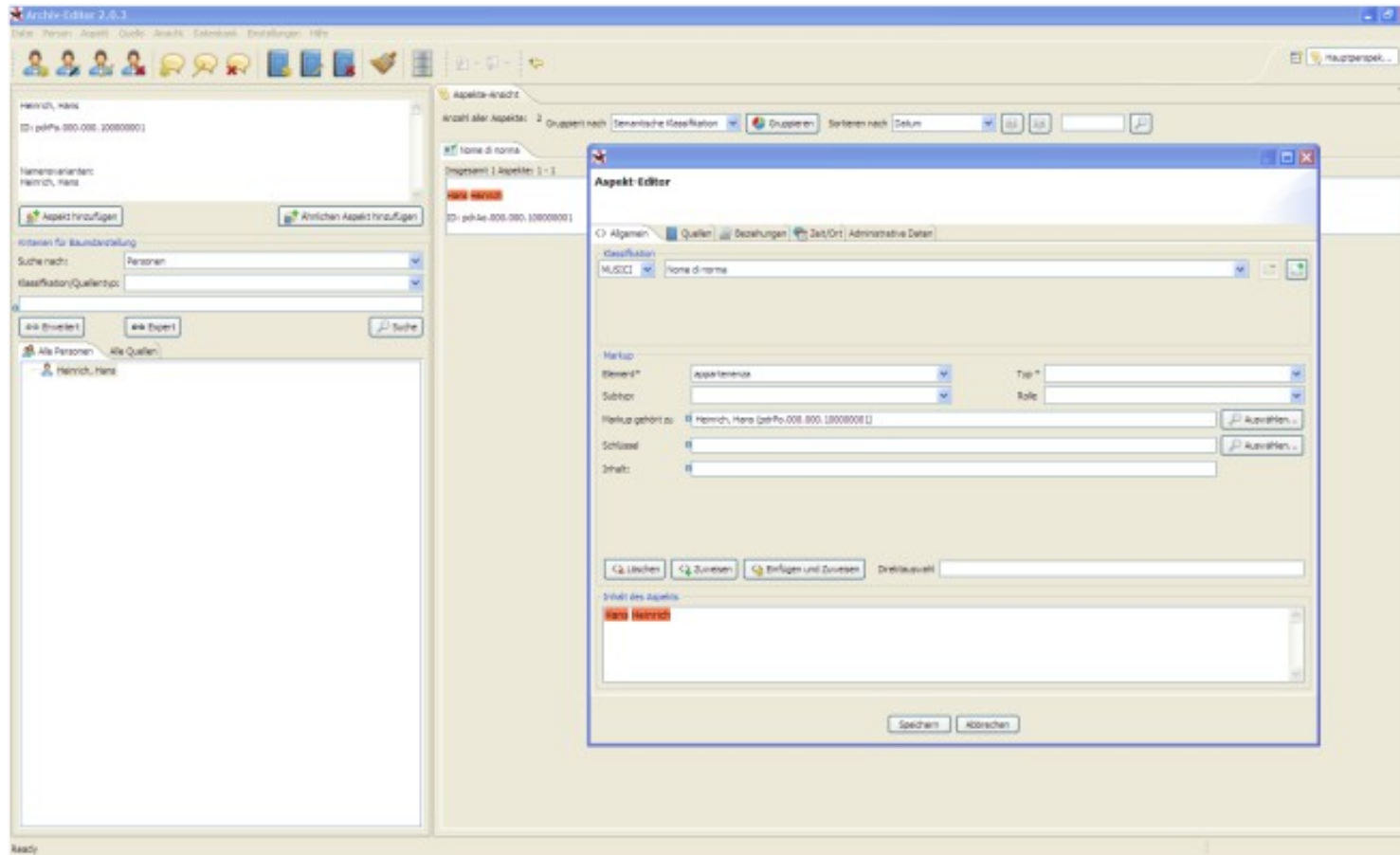


Aspekt 2

- Quelltext = wurde 1849 Professor in Würzburg
- Semantische Kategorie = Beruf
- Zeit = 1849 (Markup)
- Ort = Würzburg (Markup)

Aspekt 3

- Quelltext = lebte von 1821 bis 1902
- Semantische Kategorie = Lebensdaten
- Geburt = 1821 (Markup)
- Tod = 1902 (Markup)



PDR & Archiv-Editor bieten

- ein flexible & erweiterbares Datenmodell
- Integration mit GND, VIAF...
- Verwaltung & Referenzierung von Personen mit & **ohne** GND
- Verwaltung projektspezifischer Personeninformationen
- Webfrontend, Webservices
- Archiv-Editor als Eclipse RCP 3.x / 4.x

(a) Parallelbetrieb von TG und AE

- Technisch: zwei unabhängige Programme
- Aufwand: keine Anpassungen erforderlich

(b) Integration der Anwenderprogramme (Desktop)

- Technisch: beide Eclipse RCP
- Aufwand: abhängig von Bedarf und Umfang

(c) Integration der Repositorien (Web)

- Technisch: Webservices, Webfrontend
- Aufwand: abhängig von Bedarf und Umfang

Bedarf: Personendatenblatt

- Prosopografische Daten zu Autoren verwalten
- Personen in Texten auszeichnen
- Biografische Daten zu Personen anlegen
- Glossare zu Texten erstellen
- Beziehungen zwischen Personen beschreiben
- noch mehr?

Archiv-Editor in TextGrid-Lab

```

<lb xml:id="tg885.3.9"/>
<p xml:id="tg885.3.10">Der Indier konnte nur ungeheu'r große Gedichte liefern, weil er nichts aus de
ungsmensch; die ganze Welt ist ein Gedicht, wovon der »Mahabharata« nur ein Kapitel. – Vergleich der ir
<lb xml:id="tg885.3.11"/>
<p xml:id="tg885.3.12"><tei:persName ref="pdrPo.001.002.12473489">Goethe</tei:persName>, im Anfang c
<lb xml:id="tg885.3.13"/>
<p xml:id="tg885.3.14">Die epischen Gedichte der

```

pdprod.bbaw.de/pit/2-2/getPerson.php?personId=pdrPo.001.002.000000252

Robert Brown

- Robert Brown

Biographische Details

curriculum

- 1793-1800 Chirurg in der schottischen Armee. Anschließend bis 1805 Teilne
- 1808 Bibliothekar
- 1847-1849 Präsident der Linnean-Societät. Seit 1810 Aufseher der Banks'scl


memberships


- Korrespondierendes Mitglied der Königlich-Preußischen Akademie der Wis
- Auswärtiges Mitglied der Königlich-Preußischen Akademie der Wissensch


disciplines


- Botanik

Insgesamt 2 Aspekte 2 - 2

Goethe 

Quelle: [TextGrid-Repository](#)  **Bevorzugter Name**

Dichter 

Quelle: [TextGrid-Repository](#)  **Allgemeine Beschreibung**



(a) Einfache technische Umsetzung

- Integration als Erweiterung (Plugin) in TextGrid Lab
- diese holt Personendaten über Webschnittstelle (z.B. PIT)
- so können Personen, die im PDR liegen, in TextGrid ausgezeichnet werden
- Erfassung neuer Personen in parallel (nicht integriert) laufendem Archiv-Editor
- nach Synchronisierung sind diese dann über Webschnittstelle verfügbar

Vorteile:

- kontrolliertes Vokabular des Archiv-Editors für Personendaten
- projektweite oder auch allgemeine Personen-Identifikatoren in TextGrid Lab

(b) Tiefe Integration

- automatische Identifikation von Personen in Texten (Concordancer)
- Erweiterungspunkte in Archiv-Editor
 - Abfrage von Personen
 - Abfrage von Personendatenblättern
- Plugin für TextGrid-Lab: Annotieren von Personen
- Erweiterung des Text-Editors in TextGrid-Lab
- Plugin im Archiv-Editor: Registererstellung
- Installation des Archiv-Editors ins TextGrid-Lab

Vorteile:

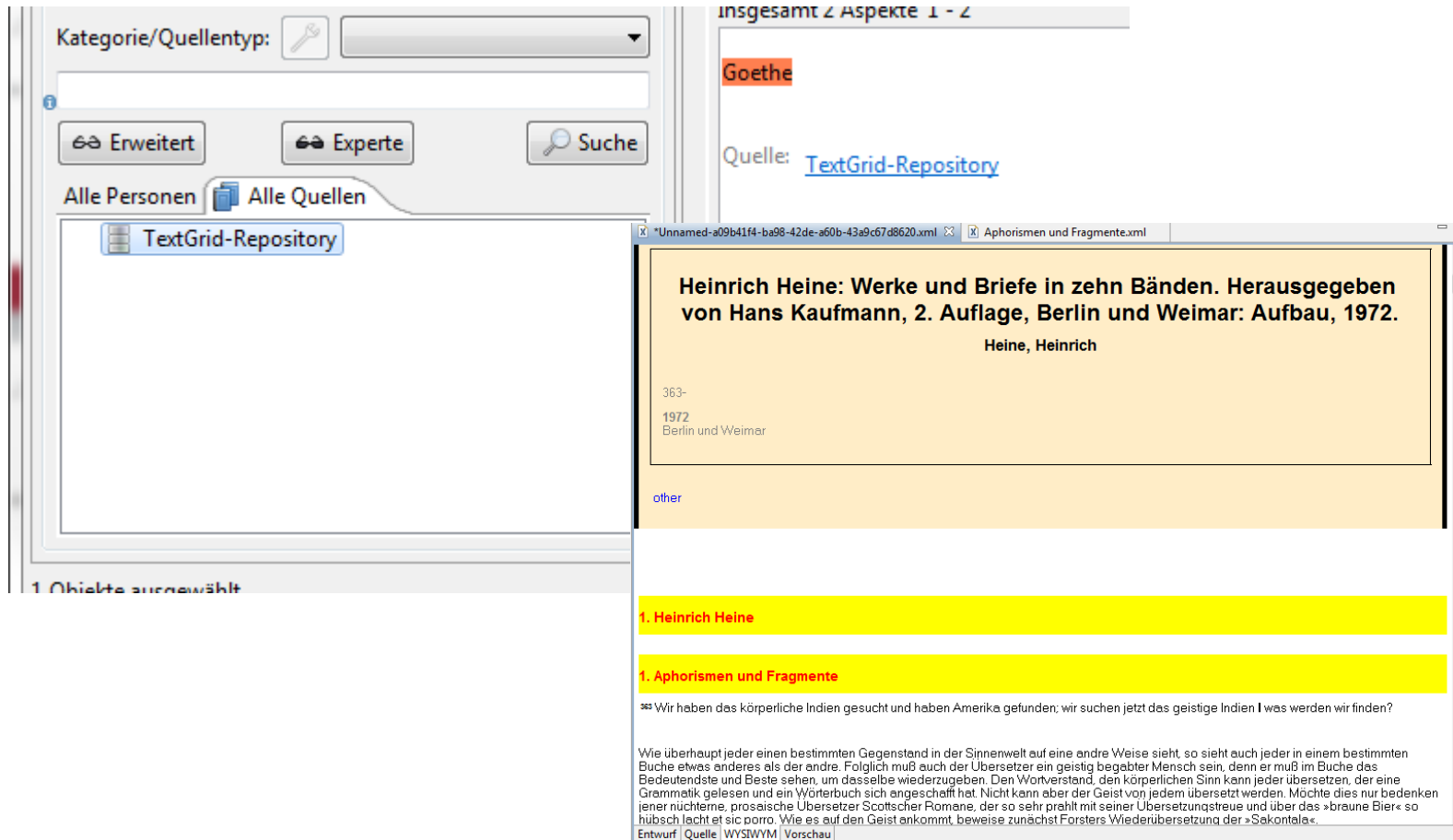
- mächtiges integriertes Werkzeug für komplexe Editionsprojekte

IV Beispiel 1: Personendatenblatt



Bedarf: Textquellen

- Textquellen liegen in TextGrid
- Texte sollen als Quellen referenziert werden
- Import biografischer Daten aus Texten
- Verwendung von Textschnipseln als Aspekte im Archiv-Editor
- Anzeige kompletter biografischer Quellentexte



Kategorie/Quellentyp:

Erweitert Experte Suche

Alle Personen **Alle Quellen**

TextGrid-Repository

insgesamt 2 Aspekte 1 - 2

Goethe

Quelle: [TextGrid-Repository](#)

Heinrich Heine: Werke und Briefe in zehn Bänden. Herausgegeben von Hans Kaufmann, 2. Auflage, Berlin und Weimar: Aufbau, 1972.

Heine, Heinrich

363-
1972
Berlin und Weimar

other

1. Heinrich Heine

1. Aphorismen und Fragmente

Wir haben das körperliche Indien gesucht und haben Amerika gefunden; wir suchen jetzt das geistige Indien I was werden wir finden?

Wie überhaupt jeder einen bestimmten Gegenstand in der Sinnenwelt auf eine andre Weise sieht, so sieht auch jeder in einem bestimmten Buche etwas anderes als der andre. Folglich muß auch der Übersetzer ein geistig begabter Mensch sein, denn er muß im Buche das Bedeutendste und Beste sehen, um dasselbe wiederzugeben. Den Wortverstand, den körperlichen Sinn kann jeder übersetzen, der eine Grammatik gelesen und ein Wörterbuch sich angeschafft hat. Nicht kann aber der Geist von jedem übersetzt werden. Möchte dies nur bedenken jener nüchterne, prosaische Übersetzer Scottscher Romane, der so sehr prahlt mit seiner Übersetzungstreue und über das »braune Bier« so hübsch lacht et sic porro. Wie es auf den Geist ankommt, beweise zunächst Forsters Wiederübersetzung der »Sakontala«.

Entwurf | Quelle | WYSIWYM | Vorschau

(a) Einfache technische Umsetzung

- Auszeichnung von Quellenobjekten im Archiv-Editor mit TextGrid PID
- Texte können per Referenzierung über URL geholt werden
- GUI ruft URL des Textes im TextGrid Rep über Browser auf

Vorteil:

- Integrität von Einzelinformationen (PDR) und Informationsquelle (TG)

(b) Tiefe Integration

- Identifikations- und Referenzierungssystem entwickeln
- Erweiterung im Archiv-Editor:
 - Import von Text-Objekten aus TextGrid Rep oder -Lab
 - Einlesen von Metadaten aus TextGrid Rep
 - Workflow zum Erstellen von Aspekten aus TextGrid-Texten und Textauszügen, dabei Verarbeitung von TEI-Markup aus TextGrid mithilfe eines konfigurierbaren Prozessors mit entsprechenden Mappings
 - Anzeige von biographischen Quellentexten
- Erweiterung in TextGrid Lab:
 - Erstellen von PDR-Personenobjekten (oder Vorstufe) aus Texten

(a) **Concordancer** zur Personen-Identifizierung als Plugin

- Identifizierung von Personen und Personennamen in Texten
- automatische Auszeichnung von Personennamen mit GND

(b) Nutzung von **Ortsnamen-** und **Datumserkennern** als Plugins

- Webservices des PDR für die Erkennung von Ortsnamen
- ebenso für Datumsangaben („Ostern 1859“ → ISO 1859-04-22)
- diese sind ggf. leicht in TextGrid Lab integrierbar

(c) Verknüpfungen der **Repositorien** von PDR und TextGrid

- bisherige Beispiele: Synergien zwischen Clients
- Zielsetzung hier: Synergien auf Ebene der Webdarstellung
- realisierbar z.B. über PIT (allgemeine Middleware des PDR)
 - Anzeige von Personendaten im TextGrid Rep (wie Register)
 - Import von Textquellen in die PDR-Anzeige
- oder auch über IDI
- Nutzung des BEACON-Service

Interaktion	Ziel	In TG-Lab	AE	Aufwand
Parallelbetrieb	beides nutzen	-	-	0
Personendatenblatt (a)	Personendaten in PDR	Plugin, Zugriff PIT	-	niedrig
Personendatenblatt (b)	AE in TG-Lab	P-Erkenner, Markup	Schnittstellen, Integration	hoch
Textquellen (a)	Text in TG referenzieren	-	Plugin, Zugriff auf TG	mittel
Textquellen (b)	Personendaten importieren	Plugin Person erstellen	Textimporte, Referenzierung	hoch
Concordancer	P. identifizieren	Plugin, WS	-	niedrig
Rep. Web	Web integrieren	Webfrontend	Webfrontend	mittel

Es bestehen vielfältige Verknüpfungsmöglichkeiten!

- teilweise sind diese bereits praktizierbar
- an vielen Stellen wäre zusätzliche Implementierung notwendig
- Aufwand ist abhängig vom gewünschten Grad der Integration

Was tun mit den schönen Ideen?

- Chance: Forschungsprojekt, das konkreten Bedarf äußert
- synergetisches Potenzial ausloten und ggf. realisieren
- stets: Aufwand nicht unterschätzen, wenn es nachhaltig werden soll

25 Nov. 2014

TextGrid-Nutzertreffen

TextGrid und Personendaten-Repository

Vielen Dank!

Kontakt:

Christoph Plutte <plutte@bbaw.de>

Torsten Roeder <torsten.roeder@uni-wuerzburg.de>