

# LAUDATIO-Workshop

Entwicklersession zu Gemeinsamkeiten  
in Forschungsdatenrepositorien

Technische (Weiter-)Entwicklung LAUDATIO-Repository für  
historische Textkorpora  
8. Oktober 2014, Berlin

\$whoami

Dennis Zielke LAUDATIO-Repository Entwickler

Source code auf Github: <https://github.com/DZielke/laudatio>

Weitere Projekte:

<http://repositoryranking.org/>

 <http://de.linkedin.com/pub/dennis-zielke/>

<http://www.laudatio-repository.org>

# Komplexe Datenstruktur

Deutsche historische Texte und ihre linguistischen Annotationen einschließlich aller Dialekte vom 9. bis 19. Jahrhundert

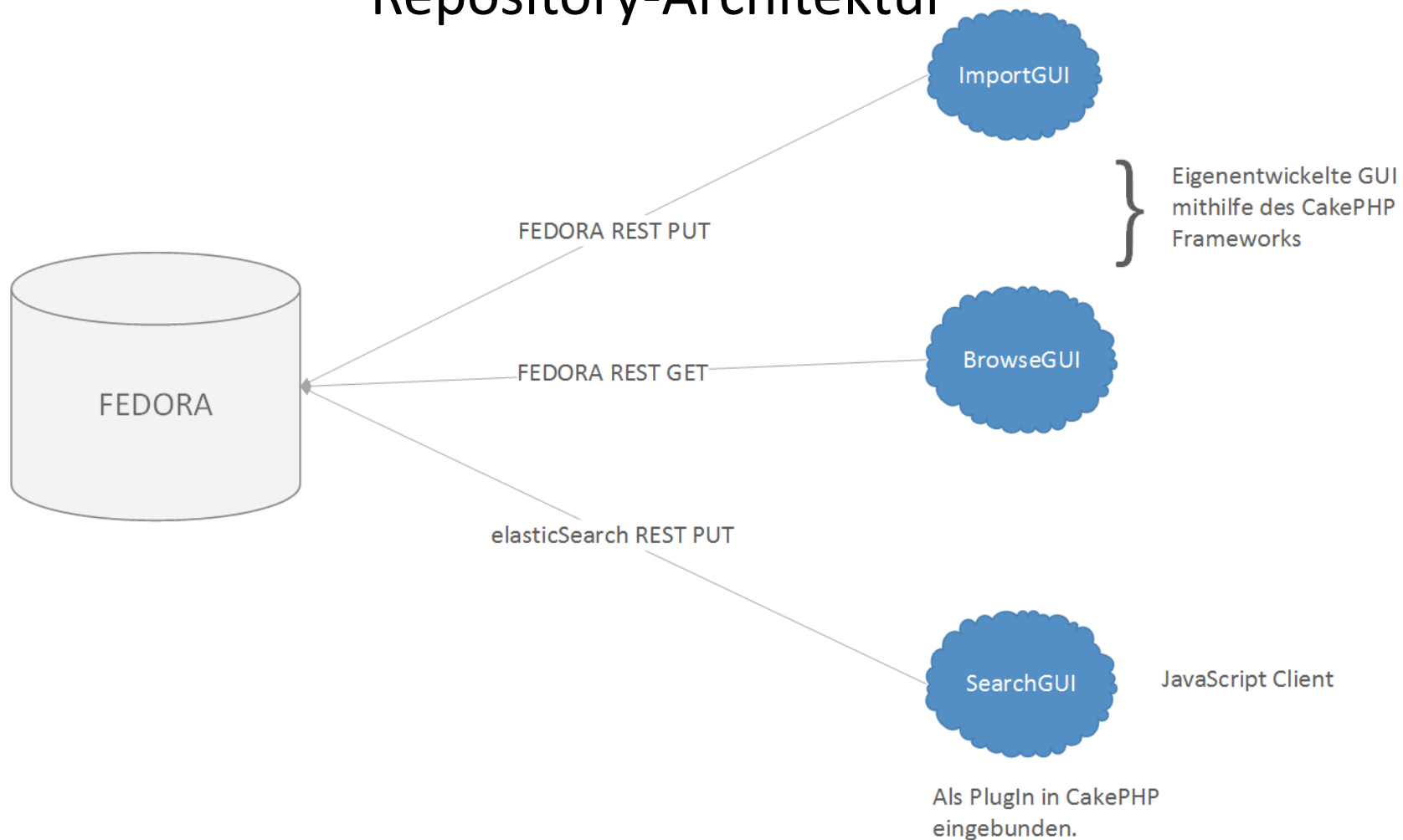
- Metadaten zum Korpus
- Dokumente
- Annotationen
- Bearbeitungsschritte

# TEI XML P5

```

- <teiCorpus>
+ <teiHeader type="CorpusHeader"></teiHeader>
- <teiCorpus>
+ <teiHeader type="DocumentHeader" style="Herbology"></teiHeader>
+ <teiHeader type="DocumentHeader" style="Herbology"></teiHeader>
+ <teiHeader type="DocumentHeader" style="Herbology"></teiHeader>
- <teiHeader type="DocumentHeader" style="Herbology">
- <fileDesc xml:id="V4-1543-NewKraeuterbuch">
- <titleStmt>
  <title>New Kreüterbuch</title>
- <author>
  - <persName>
    <forename>Leonhart</forename>
    <surname>Fuchs</surname>
  </persName>
  </author>
- <editor>
  - <persName>
    <forename>NA</forename>
    <surname>NA</surname>
  </persName>
  
```

# Repository-Architektur



## Open-Source-Technologien

- CakePHP 2.4-Framework
- Fedora 3.6 für die Datenhaltung und Speicherung
  - Fedora REST-Schnittstelle zum Datenaustausch
- ElasticSearch als Suchengine
  - REST-Schnittstelle zum Datenaustausch
- Externer PID-Webservice zur Verwaltung der Persistent Identifier (handle)
- Third party Open Source libraries auf Github
  - <http://tinyurl.com/lf26u97>
- Flat-Design (HTML5, CSS3) work in progress

## Entwickeln/Testen/Prüfen

- Technische Infrastruktur:
  - 2 Virtuelle Maschinen
  - Ubuntu 14.04
  - jeweils 2 CPUs und 12 GB RAM Arbeitsspeicher (skalierbar)
  - Ausfallsicherheit durch Monitoring mithilfe von Nagios am CMS
- Versionsverwaltungssystem: Git
  - verschiedene nützliche Tools z. B. Git Bash und GitWeb

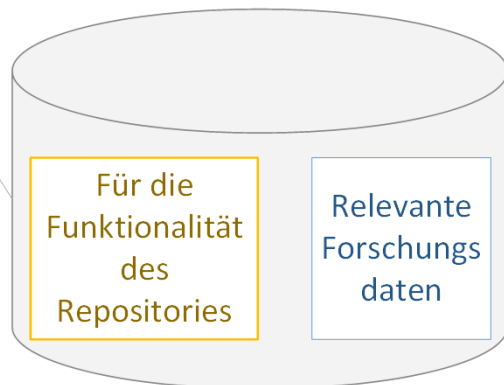
## Entwickeln/Testen/Prüfen

- Entwicklungsumgebung: PhpStorm 7.1.3
- Ticketsystem: Redmine
- ES-HEAD-PlugIn: Testen von Mapping, Facetten, Nested Facets
- FedoraAdmin Client: Testen und Prüfen der Datastreams in Fedora

# Zentrale Datenspeicherung in Fedora

## Schema Konfiguration:

- TEI-Metadatenchema (RelaxNG):
  - teiODD\_Corpus
  - tei\_ODD\_Document
  - tei\_ODD\_Preparation
- IndexMapping } Implementierung eines
- ViewMapping } JSON-Editors



## TEI-Metadatenheader (Metadaten + Annotationen + dazugehöriger Dokumente):

- Jede Version eines TEI-Metadatenheaders:
  - TEI-header\_versionX\_SchemaX
- Pro Version ein Persistent Identifier:
  - TEI-header\_versionX\_SchemaX\_2014-09-30T09:38:04:263Z": "0000-0000-2C89-5
- Zsfg der verschiedenen TEI-header Versionen:
  - TEI-header\_version1\_Schema7; TEI-header\_version1\_Schema6
- CC-Lizenz + Direktlink zu creativecommons.org
- Verschiedene Repräsentationen (Entwicklungsprozeß) als .zip (z.B. TEI XML, PDF, EXMARaLDA, EXCEL, relANNIS)

ID	Label	MIME Type
DC	Dublin Core Record for this object	text/xml
teiODD_Corpus	teiODD_Corpus.rng	application/xml
teiODD_Document	teiODD_Document.rng	application/xml
teiODD_Preparation	teiODD_Preparation.rng	application/xml
indexMapping	Index Mapping	text/json
viewMapping	View Mapping	text/json

ID	Label	MIME Type
TEI-header_version1_Schema6	TEI Header	application/xml
teiVersion	Tei Header Versions	text/plain
license		text/xml
handlePIDs	Handle PIDs	text/plain
TEI-header_version2_Schema7	Tei-Header_S7	application/xml
relANNIS	relANNIS.zip	application/zip



# Nachnutzungsmöglichkeiten

## Module:

- Anbindung an externen PID-Webservice
- Datenhaltung im Fedora-Repository
- GUI-Client für die Suche
  
- Erweiterbares IndexMapping für ElasticSearch und automatische Schema-Validierung
  - nutzbar für weitere Formate
  
- Einbindung Frontend-Editor für Redakteure
  
- Adminbereich
  - Nutzermanagement
  - Authentifizierung LDAP, HTTP-Login
  - Konfiguration

## Ausblick

- Anbindung an das HU-Medienrepositorium via API-Upload Plugin
- Implementierung eines Metadateneditors
- Migration von Fedora 3.6 zu Fedora 4
- Unterstützung weiterer Metadatenformate, je nach Bedarf der jeweiligen Disziplin
- Bereitstellung einer OAI-PMH Schnittstelle zur Anbindung an externe Datenaggregatoren

# Vielen Dank!

E-Mail: [zielkede@cms.hu-berlin.de](mailto:zielkede@cms.hu-berlin.de)