

# Text+ Dienste

**Findbarkeit, Darstellung, Pipeline und  
andere Anwendungskontexte**

Stefan Buddenbohm | SUB Göttingen

24. Juni 2024 | Bad Godesberg

# Text+ Dienste

Findbarkeit, Darstellung, Pipeline und  
andere Anwendungskontexte

1. Warum? Ausgangslage, Ziele
2. Wie machen es die anderen? – Ein Blick auf Best Practices
3. Pipeline in Text+
4. Anschlussfähigkeit für andere Communities

# 1. Warum? Ausgangslage, Ziele

1. Warum? Ausgangslage, Ziele
2. Wie machen es die anderen? - Ein Blick auf Best Practices
3. Pipeline in Text+
4. Anschlussfähigkeit für andere Communities

# Warum? Ausgangslage, Ziele

- » Forschungsdatenmanagement ohne Dienste nicht denkbar.
- » Definition in unserem Kontext: Jegliche Ressource, mit deren Hilfe Forschungsdatenmanagement betrieben, Forschungsdaten bearbeitet, beschrieben, archiviert oder bereitgestellt werden.
- » Dies können sein...
  - » Repositories,
  - » Werkzeuge zur Analyse, Prozessierung, Anreicherung,
  - » unterstützende Materialien wie Tutorials, Anleitungen,
  - » Workflows, d.h. Reihungen von Ressourcen, an deren Endpunkt ein Ergebnis steht.

# Warum? Ausgangslage, Ziele

- » Breite, tlw. unstrukturierte Palette an Diensten bei den Daten-, Kompetenz- und Rechenzentren, die zu Text+ beitragen.
- » Zielfelder/Probleme/Herausforderungen:
  - » Was ist ein Text+ Dienst?
  - » Visibility und Uptake in der Community
  - » Angebotsportfolio entwickeln? Kohärenzanspruch der Infrastruktur.
  - » Wie Verknüpfung mit anderen Angeboten?
  - » Einbettung/Nachnutzung vorhandener Ressourcen

# Ausgangslage > Schlussfolgerungen

- » Was ist ein Text+ Dienst? > breite Definition, Ableitung aus dem Datenmodell des SSH Open Marketplace
- » Visibility und Uptake in der Community > Portal nur als ein Informationskanal neben anderen
- » Angebotsportfolio entwickeln? Kohärenzanspruch der Infrastruktur. > Architektur
- » Wie Verknüpfung mit anderen Angeboten? > NFDI-Kontext, bspw. die 4C Registry for Research Tools and Data Services
- » Einbettung/Nachnutzung vorhandener Ressourcen > Marketplace, CI/CD-Pipeline des Text+ Portals

# 2. Wie machen es die anderen?

1. Warum? Ausgangslage, Ziele
2. **Wie machen es die anderen? – Ein Blick auf Best Practices**
3. Pipeline in Text+
4. Anschlussfähigkeit für andere Communities

Software

## Antelope

Permalink  
 Teilen  
 Metadaten  
 Exportieren



Um auf die spezifischen Bedürfnisse der Kulturforschungsgemeinschaften einzugehen, haben die an der TIB ansässigen TA1- und TA5-Teams das Antelope-Projekt (Annotation, Terminology Lookup und Personalisierung) konzipiert und entwickelt.

Entwickler:innen:

Herausgeber:innen:

## Registry für Forschungswerkzeuge und Datendienste

Die Registry ist ein Service, der Metadaten über bestehende Forschungswerkzeuge und Datendienste, die speziell für die Kulturwissenschaften geeignet sind, sammelt und erfasst. Es bietet einen einfachen Überblick und verhindert Doppelentwicklungen.

Suchbegriffe eingeben...

Filter by  
 Schlagwort ▾  
 Akademische Disziplin ▾  
 Schnittstelle ▾  
 Lizenz ▾  
 Betriebssystem ▾  
 Programmiersprache ▾  
 Entwicklungsstand ▾  
 Forschungsaktivität ▾

62 Treffer

- Architektur
- Kunstgeschichte
- Tanzwissenschaft

Zugehörige Fachbereiche:

- Digital Humanities
- Medienwissenschaft
- Theaterwissenschaft

— Identifikation

Name: Antelope

Beschreibung: Um auf die spezifischen Bedürfnisse der Kulturforschungsgemeinschaften einzugehen, haben die an der TIB ansässigen TA1- und TA5-Teams das Antelope-Projekt (Annotation, Terminology Lookup und Personalisierung) konzipiert und entwickelt. Antelope erweitert die Funktionalitäten bestehender Tools



## Tools and Services Catalogue

Home > Tools & Services

DARIAH members and partners contribute to the DARIAH infrastructure with a diverse range of resources. Explore the DARIAH Tools & Services catalogue to discover the richness that DARIAH tools and services have to offer.

Core services

Community Services

Core services are mature services owned by DARIAH-ERIC enabling the infrastructure to carry out its mission.

Hide Sidebar

Number of Tools and Services: 2



### DARIAH Authentication and Authorization (AAI)

A proper and interoperable Authentication and Authorization Infrastructure is the basis for a distributed, federated research infrastructure. Since 2006 DARIAH-DE runs a productive Authentication a...



### DARIAH-DE Helpdesk (DARIAH-EU, CLARIAH-DE, Text+)

A good starting point to receive support for DH-related questions, tools and resources

Text+

### SEARCH

Search..

Search

Reset Filters

### Actors:

DARIAH National Node

CLARIAH-AT

DARIAH-DE

DARIAH-FR

+ More

### Activity:

Academic Publishing

## Planen

Unsere Planungstools unterstützen beim Umgang mit Biodiversitäts-, Ökologie- und Umweltdaten – strukturiert, nachvollziehbar und transparent.

Mehr erfahren [↗](#)

## Entdecken & Erkunden

Unsere Anwendungen zum Entdecken & Erkunden helfen dabei, besser zugängliche Daten aufzubereiten – und sie damit für die langfristige Nachnutzung fit zu machen.

Mehr erfahren [↗](#)

## Harmonisieren

Unsere Harmonisierungstools unterstützen es, Daten zu integrieren, Terminologien zu vereinheitlichen und die Datenherkunft darzustellen. Mehr Informationen nutzen werden.



### Provider

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), Staatliche Naturwissenschaftliche Sammlungen Bayerns/SNSB-IT-Zentrum, Gesellschaft für Biologische Daten e.V.

### Voraussetzungen

Keine

### Anwendungsbereich

Planen, Harmonisieren, Verwenden & Archivieren, Entdecken & Erkunden, Verarbeiten & Visualisieren, Support & Training

## Verarbeiten

Unsere Verarbeitungs- und Visualisierungstools unterstützen es, Daten zu analysieren oder visuell aufzubereiten.

Mehr erfahren [↗](#)

Unser Helpdesk ist unsere Kontaktstelle für Fragen aller Art – auch zur Nutzung unserer Tools und Services. Darüber hinaus bieten wir Trainings an, in denen wir demonstrieren, wie unsere Tools und Services genutzt werden können.

Mehr erfahren [↗](#)

# Schlussfolgerungen

- » Stärken/Chancen:
  - » disziplinäre Expertise, Vertrauen in der Community, Interoperabilität hinsichtlich Technologien und Datenmodellen, Kontextualisierung
- » Schwächen/Risiken:
  - » getrennte Entwicklung von Katalogen; Kuration von Daten durch Nutzer:innen; Lebensdauer/Aktualität/Gültigkeit von Daten

# 3. Pipeline in Text+

1. Warum? Ausgangslage, Ziele
2. Wie machen es die anderen? – Ein Blick auf Best Practices
3. Pipeline in Text+
4. Anschlussfähigkeit für andere Communities



# Pipeline in Text+: Anforderungen

- » Maschinenlesbarkeit, API-Zugänglichkeit
- » hochwertige Erschließung und Präsentation
- » Erschließung in “änderungstoleranter Weise”
- » Zuschnitt auf unsere Community
- » niedrige Nutzungs- und Beteiligungsschwellen
- » gesundes Kosten- vs. Nutzenverhältnis (!)



# Architektur v1.0 3/2024

## Über diese Zeichnung

Die Zeichnung stellt eine funktionale Sicht auf Text+ dar, die wesentliche funktionale Komponenten wie deren Architektur beschreibt. Die Architektur wird basierend auf der Idee, einen neuen Datenraum/Neuen Datenzentrum vor den Fachstellen zu erstellen.

Die Hauptkomponenten der Architektur sind sowohl in der Frontend- als auch in der Backend-Architektur dargestellt. Die Frontend-Architektur zeigt den Aufbau von Text+ von den Fachstellen aus.

Was die Architektur nicht zeigt:

- Sie ist keine umfassende Darstellung der technischen Komponenten.
- Sie bildet nicht den gesamten Aufbau von Text+ ab, sondern nur die wesentlichen Funktionen.

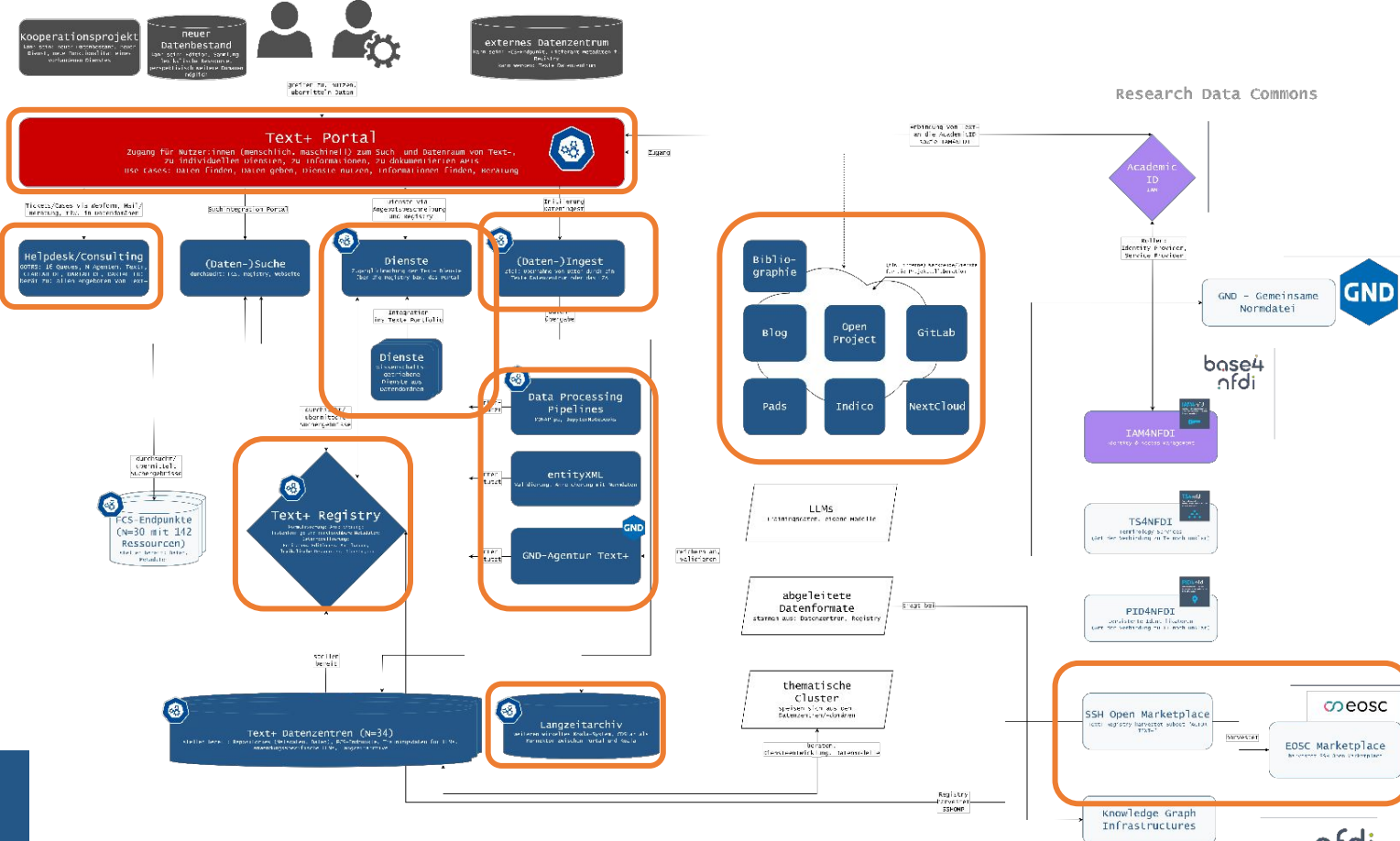
# Dienste in der Text+ Architektur...

User  
Frontend  
Backend

Zugangs- und Präsentationsschicht

Applikationen/ Dienste

Daten

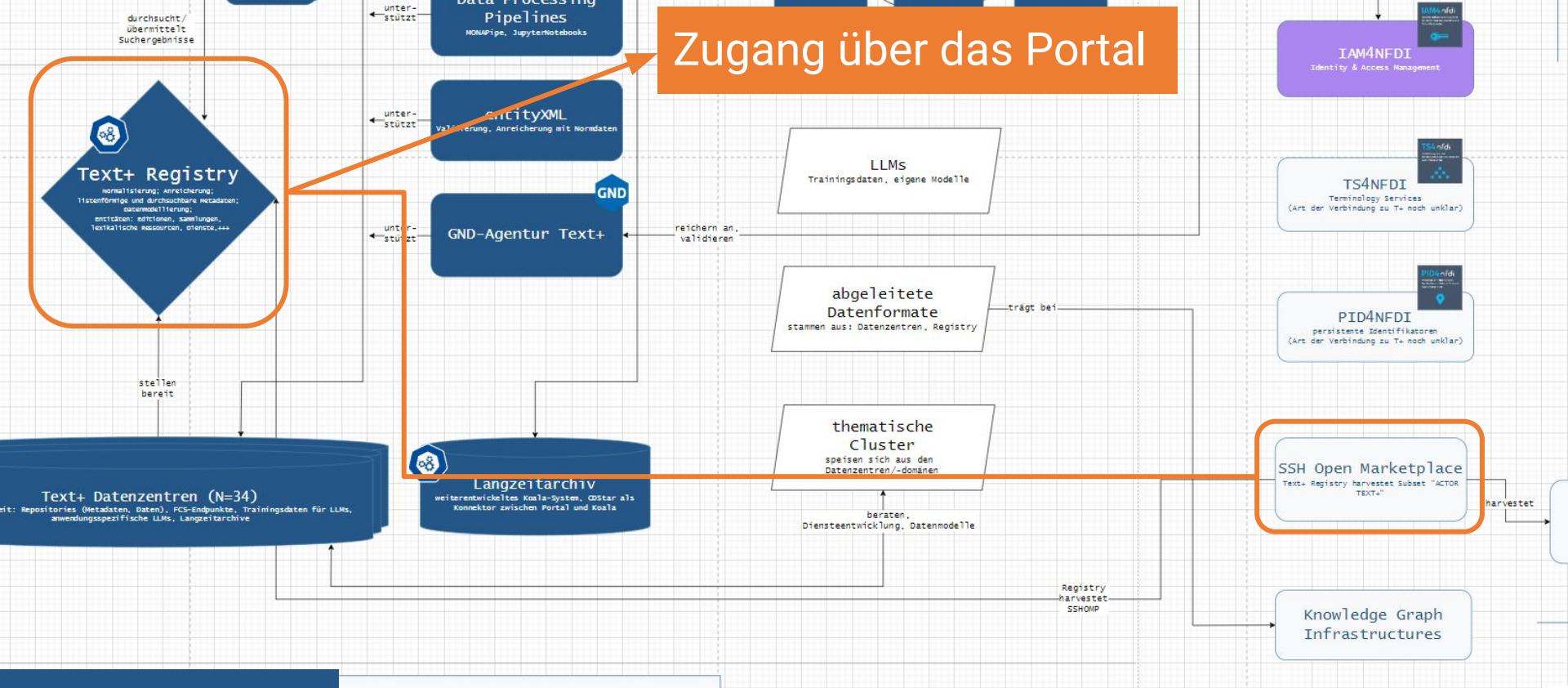


Rechenzentren, Informationsinfrastruktureinrichtungen

# Text+ Architektur

<https://text-plus.org/ueber-uns/architektur/>





# Dienste aus Zugangs-/ Nutzersicht...

- » Dienste als Katalog im Portal
- » Metadatenbereitstellung über die Registry
- » Registry harvestet Dienstbeschreibungen aus dem SSH Open Marketplace

# Pipeline

Autoritative Quelle	 SSH Open Marketplace <small>Social Sciences and Humanities Open Marketplace</small>	 Text+ Registry	 Text+ Portal	weitere Information Provider
<p>Idealerweise stammen die erschließenden Informationen aus dem Entstehungs- oder Bereitstellungskontext der Ressource. Damit kann eine gewisse Qualität und Verlässlichkeit der Metadaten angestrebt werden.</p>	<p>Redaktionelle Prüfung und Anreicherung der Metadaten vor Freischaltung. Kuration durch die Community. Kontextualisierung von Ressourcen. Workflows. Bereitstellung des Bestands für Dritte über eine funktionierende und sauber dokumentierte API.</p>	<p>Text+ Registry als "Katalog" aller in Text+ verzeichneten Ressourcen, seien es Dienste, Forschungsdaten, Sammlungsbeschreibungen oder Datenzentren. Zwischenschritt für Dienste unnötig: wurde gewählt, um den Implementierungsaufwand im Portal zu minimieren.</p>	<p>Harvesting der Entität Dienste über die API der Registry durch HUGO und Deployment in Form der Dienstliste im Portal. Darstellungsweise derzeit als Liste mit wenigen Properties. Nightly Build jeweils eine Stunde nach dem Harvesting der Registry vom SSH Open Marketplace.</p>	<p>Zwei Szenarien: für die Text+ Community relevante Ressourcen im Portal anzeigen.</p> <p>ODER Harvesting der Registry bzw. des SSH Open Marketplace durch andere Information Provider.</p>

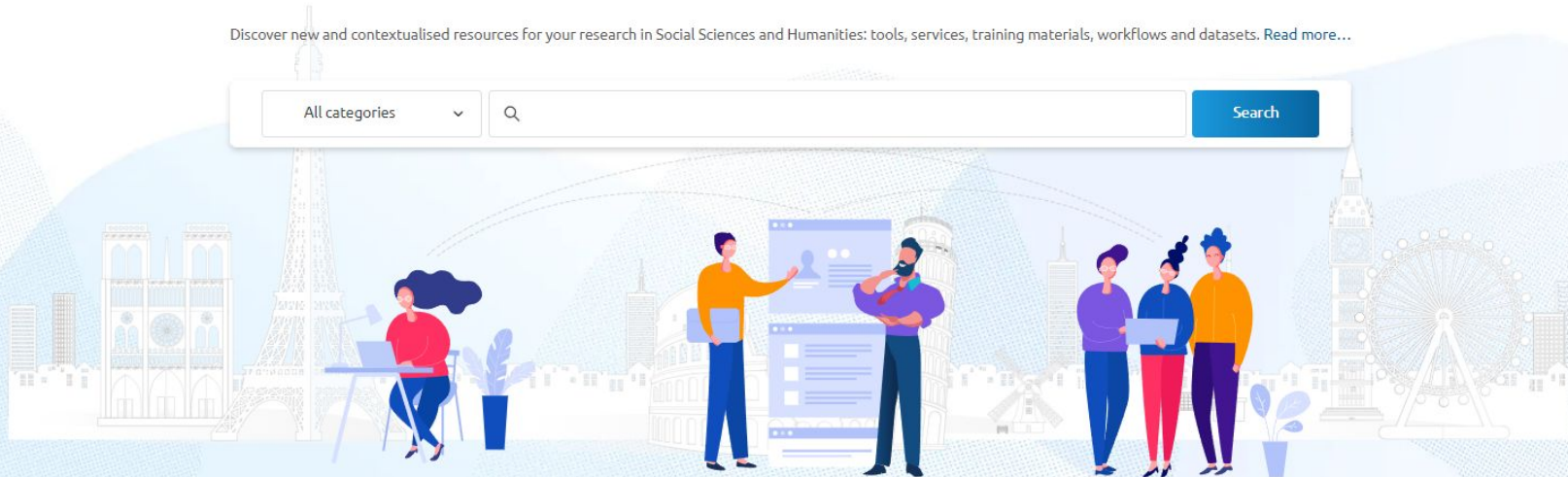




## Social Sciences & Humanities Open Marketplace

Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets. Read more...

All categories ▼



# How does it work in practice?

**3 ERICs service agreement**  
90k€/year for 20 months

**Governing Board members are**  
Mari Kleemola (CESSDA), Dieter van Uytvanck (CLARIN) and Arnaud Roi (DARIAH)

**Service provision is ensured by**  
ACDH-CH (OEAW) and by PSNC.



## Visitors & Contributors

- 300 visits/month (last 6 months)
- 200 registered users
- 130 manually created items & ~1000 enriched (last 6 months)
- 1.5 day hands-on session = 55 resources created or enriched

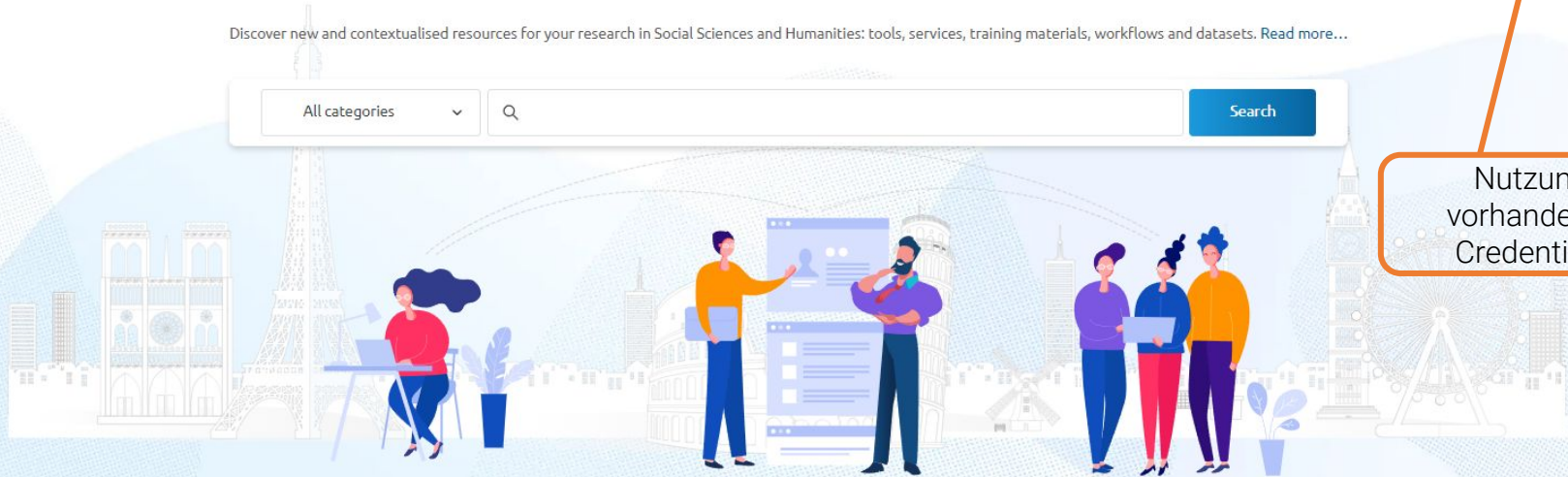
## Editorial Board activities

- 17 members: 8 appointed and funded by the ERICs & 9 volunteers
- Vocabulary curation
- URLs checks
- 8 events (presentations, train-the-trainer, hands-on) over the last 6 months
- 1 new source processed

## Social Sciences & Humanities Open Marketplace

Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets. Read more...

All categories



Nutzung  
vorhandener  
Credentials

Home / Tools & services / MONAPipe



Datenmodell des SSH Open Marketplace

MONAPipe stands for "Modes of Narration and Attribution Pipeline". It provides natural-language-processing tools for German, implemented in Python/spaCy. In addition to spaCy's default components, MONAPipe adds specific custom components and models for Digital Humanities and Computational Literary Studies.

MONAPipe was originally created in the project group [MONA](#) and is now further developed within the [Text+](#) infrastructure.

This resource is supported by Text+. In case of questions you may get in touch with the Text+ helpdesk at [textplus-support@gwdg.de](mailto:textplus-support@gwdg.de).

Edit | History

[Go to Tool or service](#)

Copy to clipboard

### Details

#### ACCESS

License GNU General Public License v3.0 only

#### CATEGORISATION

Activity Publishing Annotating  
Keyword DARIAH Resource Textplus  
Discipline Digital humanities Computational linguistics

#### CONTEXT



## Related items (9)



### TextGrid Repository & Laboratory

TextGrid is a virtual research environment for text-based humanities scholarship. It offers a variety of tools and services for collaboratively creating, analyzing, editing, and publishing texts. The TextGrid environment...

[Read more](#)



### TextGrid Import UI

A Jupyter Notebook-based user interface for TextGrid Import Modelling, which is a command line tool that facilitates the creation of the metadata files required for importing data into TextGrid Rep. The objective is to make the...

[Read more](#)



### The TextGrid Repository Documentation

The TextGrid Repository is a long-term archive for humanities research data. It provides an extensive, searchable, and reusable repository of texts and images. Aligned with the principles of Open Access and FAIR, the TextGrid...

[Read more](#)



### tg-model - TextGrid Import Modeller

What's the aim? This project focuses on attempts for a simple import of text corpora (encoded in XML/TEI) to TextGrid Repository by modeling the required metadata file structure. To generate the TextGrid metadata files, t...

[Read more](#)

[Show more](#)

Kontextualisierung

## Workflow steps (5)

- 1 | **The TextGrid Repository Documentation**
- 2 | **TextGrid Import UI**
- 3 | **tg-model - TextGrid Import Modeller**

[Expand](#) ▾

[Expand](#) ▾

[Expand](#) ▾

This resource is supported by Text+. In case of questions you may get in touch with the Text+ helpdesk at [textplus-support@gwdg.de](mailto:textplus-support@gwdg.de).



e.V. Textplus  
Discipline Digital humanities Linguistics and Literature  
Language English German  
Mode of use Local application  
Resource category Repository

**DURATION**  
Curate merged items TRUE  
**TECHNICAL**  
Technology Readiness Level 8 - system complete and qualified

**CREATOR**  
TextGrid  
[textgrid-support@gwdg.de](mailto:textgrid-support@gwdg.de)  
<https://textgrid.de/en/web/guest>

**CURATOR**  
DARIAH-EU  
[info@dariah.eu](mailto:info@dariah.eu)  
<https://www.dariah.eu/>  
Twitter


**PROVIDER**  
Göttingen State and University Library

**REVIEWER**  
DARIAH-DE  
[info@de.dariah.eu](mailto:info@de.dariah.eu)  
<https://de.dariah.eu/>  
Twitter  
GKFI e.V.  
[kontakt@forschungsinfrastrukturen.de](mailto:kontakt@forschungsinfrastrukturen.de)  
<http://forschungsinfrastrukturen.de/>  
Text+  
[office@text-plus.org](mailto:office@text-plus.org)  
<https://www.text-plus.org/>

Flags/Branding

### Related items (4)

 TextGrid Repository - Supporting the Data Curation Needs of Humanities Researchers  
No description provided  
[Read more](#)

 TextGrid - Creating, archiving, publishing and exploring digital editions and other humanistic research data via a Virtual Research Environment  
No description provided  
[Read more](#)



Support

Suche

Cross-Cutting Topics und Standardisierung

GND-Agentur Text+

## Text- und sprachbasierte Forschungsdaten nutzen und erhalten

Der Verbund Text+ wird text- und sprachbasierte Forschungsdaten langfristig erhalten und ihre breite Nutzung in der Wissenschaft ermöglichen.

Über Text+

Daten suchen

### Arbeitsbereiche in Text+

Die Text+ Infrastruktur ist auf Sprach- und Textdaten ausgerichtet und konzentriert sich zunächst auf digitale Sammlungen, lexikalische Ressourcen und Editionen. Diese sind von hoher Relevanz für alle

**Text+Plus, #05: Von abejejaar bis źurja: Integration niedersorbisch-deutscher Wörterbücher im Kooperationsprojekt INSERT** [↗](#)

Felix Helfer

4. April 2024 09:41



Wer bereits einmal in der östlich gelegenen Lausitz unterwegs war, wird vielleicht eines der mehrsprachigen Schilder bemerkt haben, die dort zu finden sind. Dass die Lausitz ist


Juni 2024 >  
30  
individuelle  
Dienste



## FAIR Check

Findable, accessible, interoperable and reusable (FAIR) - that's how research data should be. Sounds good, but is not so easy? Take our online FAIR check and evaluate your research data. Our questionnaire will help you assess which criteria you meet.

Description

Mehr Infos 




GEMEINSAME  
NORMDATEI

## GND Agency Text+

The GND Agency Text+ is a service that is being set up at the Göttingen State and University Library (SUB Göttingen) as part of the NFDI consortium Text+.

Annotating

Mehr Infos 




## HedgeDoc

Online tool for collaborative text editing to work together on the same texts at the same time.

The HedgeDoc-pad is an open-source-based web editor that allows multiple users to work on a single text simultaneously from different locations.

Collaborating Creating

Mehr Infos 

Text+





Search

Home / API Documentation

## API Documentation

verlässliche und sauber  
dokumentierte API

About the service ▶

SSHOC and the EOSC ▶

About the data population ▶

Who are we? ▶

API Documentation ▶

About the technical aspects ▶

### Table of contents

#### [API documentation](#)

[Basic API calls: exploring the SSH Open Marketplace data](#)[Advanced API calls - editing the SSH Open Marketplace data](#)

## API documentation

The SSH Open Marketplace also includes an Application Programming Interface (API), which is opened for anyone to use. Some of the functionalities are closed, but most are opened, such as searching and retrieving detailed information on all items of the Marketplace.

You can inspect the API Swagger documentation here: <https://marketplace-api.sshopencloud.eu/swagger-ui/index.html?url=/v3/api-docs>

The endpoint of the SSH Open Marketplace API is: <https://marketplace-api.sshopencloud.eu/api/>

The API responses are in JSON and follow the data model that we have created for the Marketplace (see "[Data model](#)" section).

Text+

Home / My account / Items to moderate

## Items to moderate (12)

Editorial Board,  
Qualitätssicherung

### Refine your search

Clear filters

Sort by Last modification

#### CURATION

- Coverage
- Description
- Merged
- Relations
- URL

#### STATUS

- Ingested
- Suggested
- Approved

#### HTML5 based DjVu file viewer

Last updated: March 18, 2024

Category: Tool or service Status: Suggested

[Approve](#) [Reject](#) [Review](#)

#### Documentation and use cases for ALTO XML. Contribute to altoxml/documentation development by creating an account on GitHub

Last updated: March 18, 2024

Category: Tool or service Status: Suggested

[Approve](#) [Reject](#) [Review](#)

#### Detects columns and connects indented lines in HOCR files: nypl-spacetime/hocr-detect-columns

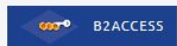
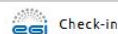
Last updated: March 18, 2024

Category: Tool or service Status: Suggested

[Approve](#) [Reject](#) [Review](#)



## CHOOSE YOUR ACADEMIC/SOCIAL ACCOUNT



or



Search...

- 29 Mayis University
- A\*STAR - Agency for Science, Technology and Research
- A. T. Still University
- AAF Virtual Home
- aai.lab.maeen.sa
- AAI@EduHr Single Sign-On Service
- Aalborg University
- Aalto University
- Aarhus School of Architecture
- Aarhus School of Marine and Technical Engineering
- Aarhus University

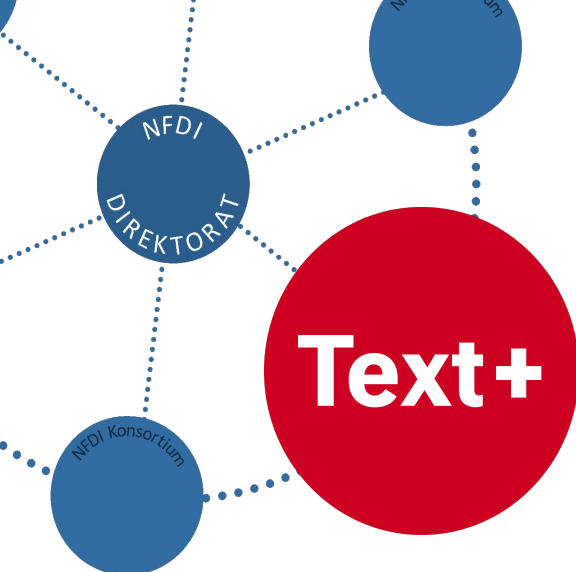
einfache Kuration

# 4. Anschlussfähigkeit für andere Communitys

1. Warum? Ausgangslage, Ziele
2. Wie machen es die anderen? – Ein Blick auf Best Practices
3. Pipeline in Text+
4. Anschlussfähigkeit für andere Communitys

# Anschlussfähigkeit für andere Communitys

- » Maschinenlesbarkeit, API-Zugänglichkeit
- » hochwertige Erschließung und Präsentation
  - » Editorial Board, EOSC, ERICs
- » Zuschnitt auf die Community
  - » Harvesting selektierter Sub-Sets
- » niedrige Nutzungs- und Beteiligungsschwellen
  - » vorhandene Credentials, Support, Editorial Board
- » gesundes Kosten- vs. Nutzenverhältnis
  - » eigentlich keine eigene Infrastruktur notwendig



# Text+ Dienste

**Findbarkeit, Darstellung, Pipeline und andere  
Anwendungskontexte**



Stefan Buddenbohm | SUB Göttingen

24. Juni 2024 | Bad Godesberg

buddenbohm@sub.uni-goettingen.de

<https://marketplace.sshopencloud.eu/about/service>

<https://text-plus.org/themen-dokumentation/bibliographie/>

# Fragen/Themenkomplexe

- » infrastrukturelle vs. konzeptionelle Verantwortlichkeit
- » Dienstedefinition(en)
- » Ownership für Dienste und Beschreibungen
- » Zielgruppe

# Pipeline

## Autoritative Quelle



Idealerweise stammen die erschließenden Informationen aus dem Entstehungs- oder Bereitstellungskontext der Ressource. Damit kann eine gewisse Qualität und Verlässlichkeit der Metadaten angestrebt werden.

Redaktionelle Prüfung und Anreicherung der Metadaten vor Freischaltung. Kuration durch die Community. Kontextualisierung von Ressourcen. Workflows. Bereitstellung des Bestands für Dritte über eine funktionierende und sauber dokumentierte API.

Text+

Registry

Text+ Registry als "Katalog" aller in Text+ verzeichneten Ressourcen, seien es Dienste, Forschungsdaten, Sammlungsbeschreibungen oder Datenzentren. Zwischenschritt für Dienste unnötig: wurde gewählt, um den Implementierungsaufwand im Portal zu minimieren.

Text+

Portal

Harvesting der Entität Dienste über die API der Registry durch HUGO und Deployment in Form der Dienstliste im Portal. Darstellungsweise derzeit als Liste mit wenigen Properties. Nightly Build jeweils eine Stunde nach dem Harvesting der Registry vom SSH Open Marketplace.

weitere  
Information  
Provider

Zwei Szenarien: für die Text+ Community relevante Ressourcen im Portal anzeigen. ODER Harvesting der Registry bzw. des SSH Open Marketplace durch andere Information Provider.

Text+