

# NLP-Infrastruktur für die Klassische Philologie

Kolloquium »Phänomenologie der Digital Humanities«

FU Berlin, 11.1.2024

**Dr. Andrea Beyer & Konstantin Schulz**  
Humboldt-Universität zu Berlin



— Digital Research for All —  
**DDAIDALOS**

Gefördert durch  
**DFG**

Deutsche  
Forschungsgemeinschaft



# Folien

<https://kurzelinks.de/infra-klassphil>

Weitere Materialien in der  
[Zenodo-Community](#)  
[Daidalos!](#)

# Auf dem Weg zu einer NLP-Infrastruktur

- # Daidalos-Projekt
- # Ausgangslage
- # User Stories durch Forschungsstandems
- # Interaktive Lern- und Forschungsinfrastruktur



Forschungsfelder der Klassischen Philologie (Hose 2023)

# Daidalos-Projekt

Entwicklung einer Infrastruktur zum Einsatz von Natural Language Processing für Forschende der Klassischen Philologie

# Projektdaten

- **Interdisziplinär:** Klassische Philologie, Linguistik, Rechenzentrum
- Ziel: **philologische Forschung** digital unterstützen
- Fokus: maschinelle **Sprachverarbeitung** für **literarische** Texte
- **Zielgruppe:** Personen mit (literatur-)wissenschaftlichem Interesse an lateinischen & griechischen Texten
- **Stellen:** je 1x 100% WiMi / Research Software Engineer + 1 sHk
- Dauer: **3 Jahre**, 2023-2026
- Mittelgeber: **DFG**, e-Research-Technologien (Forschungsinfrastruktur)
- Ausrichtung: explorativ mit **Option auf Verlängerung** um 3 Jahre



# Ziel von Daidalos

Aufbau einer **Forschungs- und Lerninfrastruktur** (Prototyp), die Forschende der Klassischen Philologie sowie angrenzender Disziplinen befähigen soll,

1. ein spezifisches, beliebig großes **Textkorpus** zusammenzustellen, zu analysieren und zu adaptieren,
2. die **Analyseergebnisse** zu visualisieren, zu speichern, zu teilen und zu exportieren,
3. über Ergebnisse und Methodik zu **reflektieren**, Forschungsfragen zu verfeinern oder aus neuen Perspektiven zu betrachten sowie
4. die eigene **Data Literacy** mithilfe didaktischer Elemente auszubauen

# Projektziele nach Domäne

- Institut für Klassische Philologie
  - **Community**arbeit
  - Praktische Anwendung in der Literaturwissenschaft
- Institut für deutsche Sprache und Linguistik
  - **Sprachtechnologien**
  - Konzeptuelle Vernetzung der Arbeitsschritte
- CMS
  - Interoperabilität & Nachhaltigkeit
  - **Konfigurationen**, Vorschläge



— Digital Research for All —  
**DDAIDALOS**

Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft



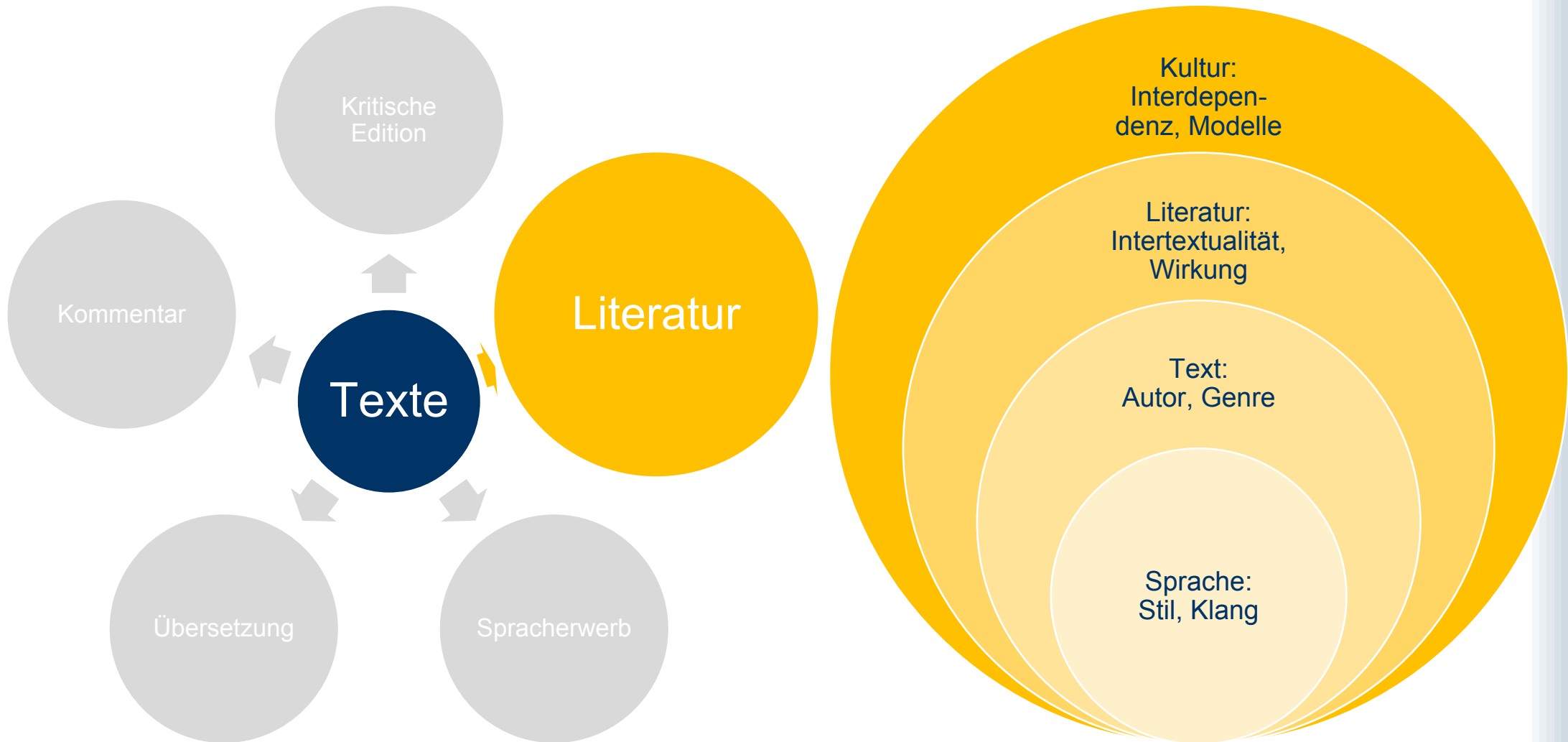
# Forschungsfelder von Daidalos



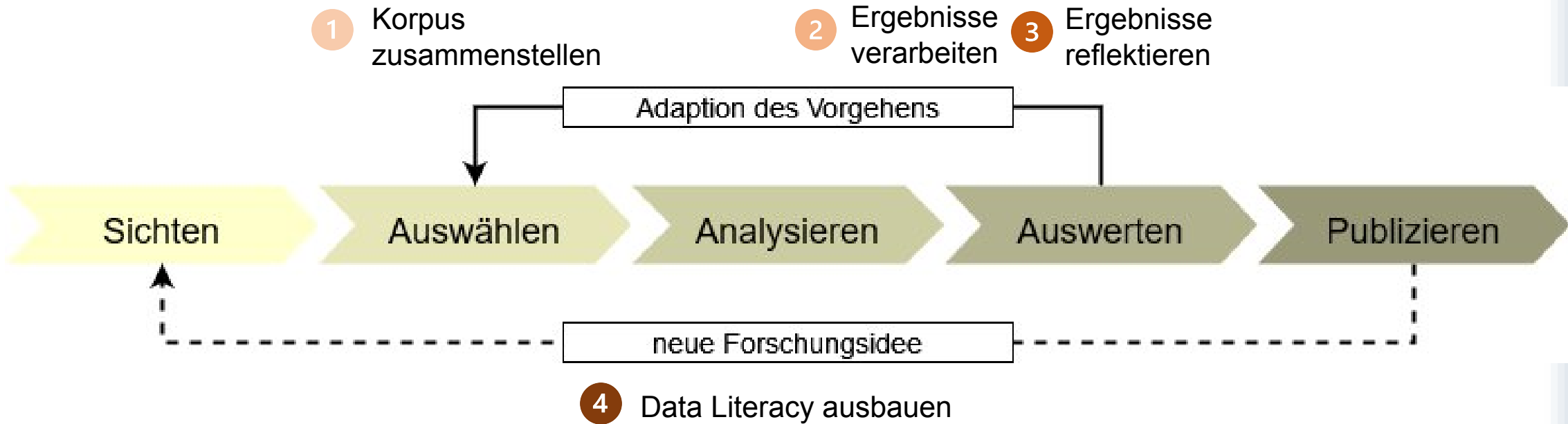
— Digital Research for All —  
**DAIDALOS**



Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft



# Workflow in der Forschungscommunity



# Anforderungen an das Projekt

---

- **Werkzeuge** für gängige **Textanalyse**methoden (z. B. POS-Tagging, Word Embeddings, Sentimentanalyse, NER, Topic Modelling) und **Ergebnisvisualisierungen** nachnutzen und methodisch sinnvoll verzahnen
- schneller Einstieg **ohne Programmierkenntnisse** mit Option zur **Weiterbildung** und zur **Speicherung** von Zwischenergebnissen
- **Workshopserie** und **Lernbausteine** zur Einführung in NLP-Verfahren und in die Funktionsweise von Daidalos



— Digital Research for All —  
**DAIDALOS**

Gefördert durch  
**DFG**

Deutsche  
Forschungsgemeinschaft

# Herausforderungen

- unterschiedliche **Fach- und Kommunikationstraditionen**
- (Ir-)Relevanz **technischer Details** für literaturwissenschaftliche Anwendung
- Notwendigkeit von **Sprachkenntnissen** (Latein/Altgriechisch) für linguistische & technische Konzeption
- **Urheberrecht** & digitale Verfügbarkeit benötigter Texte
  - mögliche Alternativen
    - Digitalisierung älterer Ressourcen (z.B. Lexika)
    - eigene Kuratierung durch manuelle Annotation
- Modellierung **komplexer Phänomene** wie Intertextualität
- Evaluation: kaum vorhandene Testdatensätze & **Benchmarks**



— Digital Research for All —  
**DDAIDALOS**

Gefördert durch  
**DFG**

Deutsche  
Forschungsgemeinschaft

# Interdisziplinäre Zusammenarbeit als Lösungsansatz

- **technische** Grundlagen
  - Software as a Service
  - Neuronale Netzwerke & Statistische Sprachmodelle
  - Participatory Design
  - No-Code-Plattformen
- **linguistische** Grundlagen
  - distributionelle Semantik
  - Abhängigkeitsgrammatik
- **praktische** Umsetzung
  - Planung: Workshops & Interviews, Forschungsstandems
  - Analyse: flexible Algorithmen mit variabler Textgrundlage
  - Ergebnisse: interaktive Visualisierung
  - Evaluation: Testdatensätze & Nutzungsstudien, OA-Publikation



— Digital Research for All —  
**DDAIDALOS**

Gefördert durch  
**DFG**

Deutsche  
Forschungsgemeinschaft

# Ausgangslage

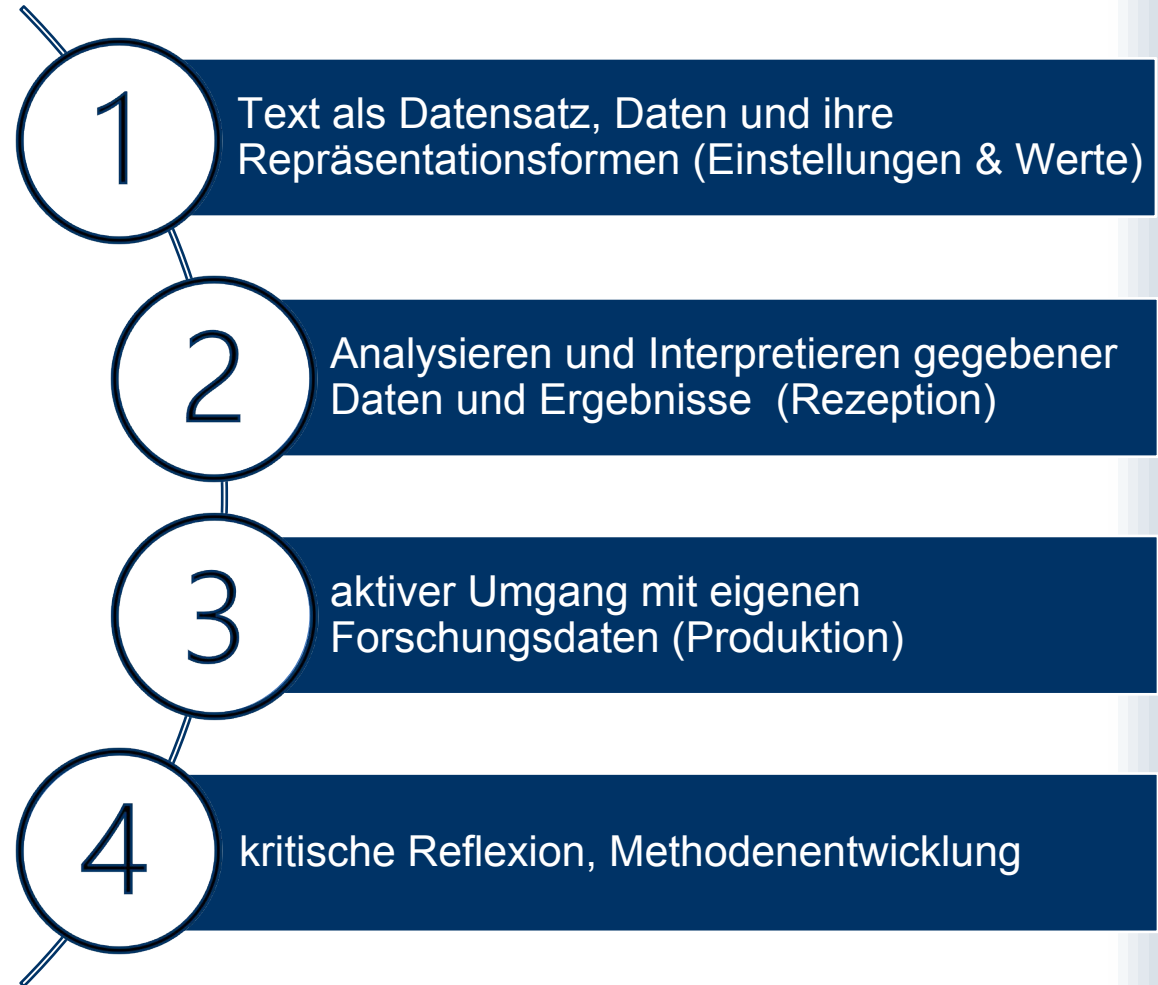
Forschungscommunity verstehen

„Wie soll ich ein  
Forschungsvorhaben entwickeln,  
bei dem der Einsatz von NLP-  
Methoden sinnvoll wäre, wenn ich  
diese Methoden gar nicht kenne?“

# 1. Fachspezifische Data Literacy

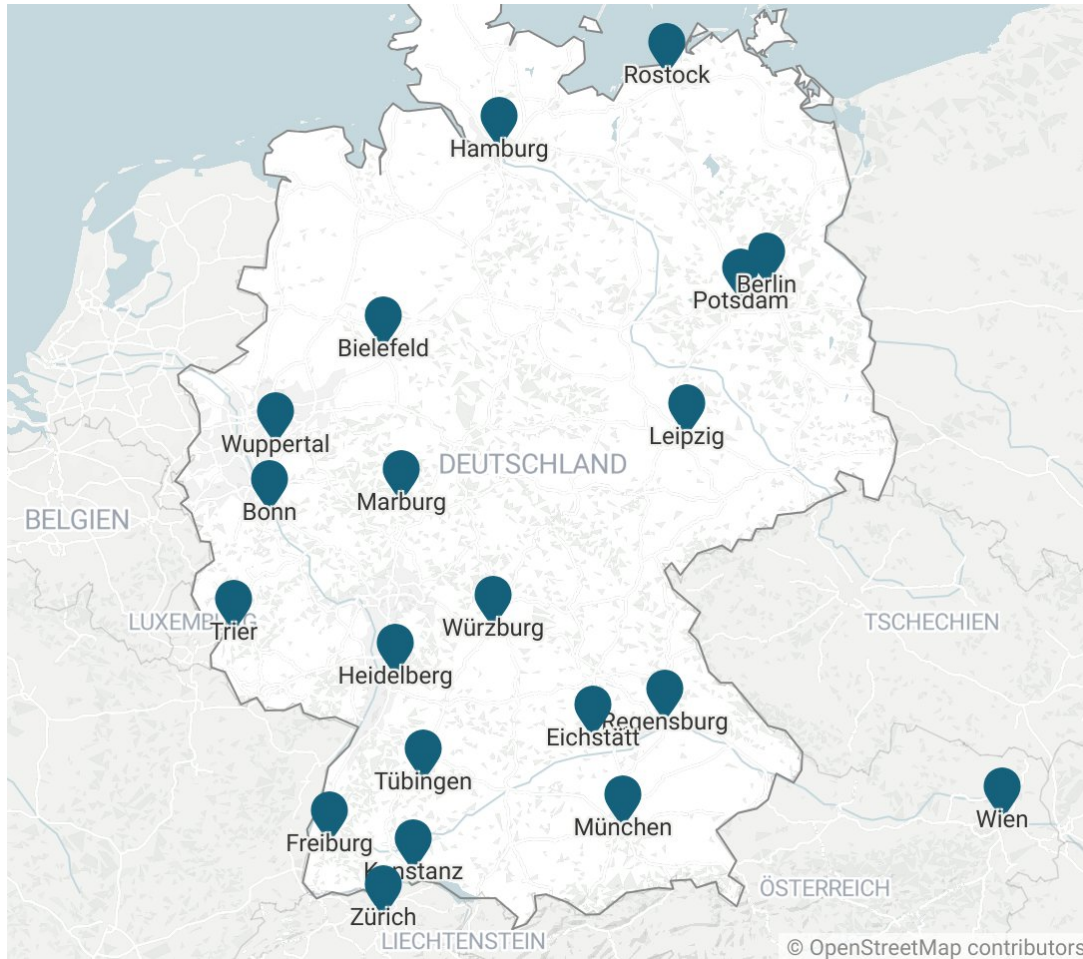
## entwickeln

- Forschungskompetenz um fachspezifische Data Literacy erweitern
- Data Literacy: Fähigkeit, Daten einerseits reflektiert zu sammeln, zu verwalten und zu bewerten, andererseits Daten selbst zu gewinnen, zu verarbeiten und zu visualisieren (Schüller 2019)





# 2. Community of Practice aufbauen



- Basis: Arbeitskreis BridgeClassics (Standorte s. Abb.)
- Ziel: situiertes Lernen mithilfe von User Stories (Lave & Wenger, 1991)
- Methode:
  - Forschungsstandems aus erfahrenen und eher unerfahrenen Usern von NLP-Methoden
  - Workshopserie & Colloquia

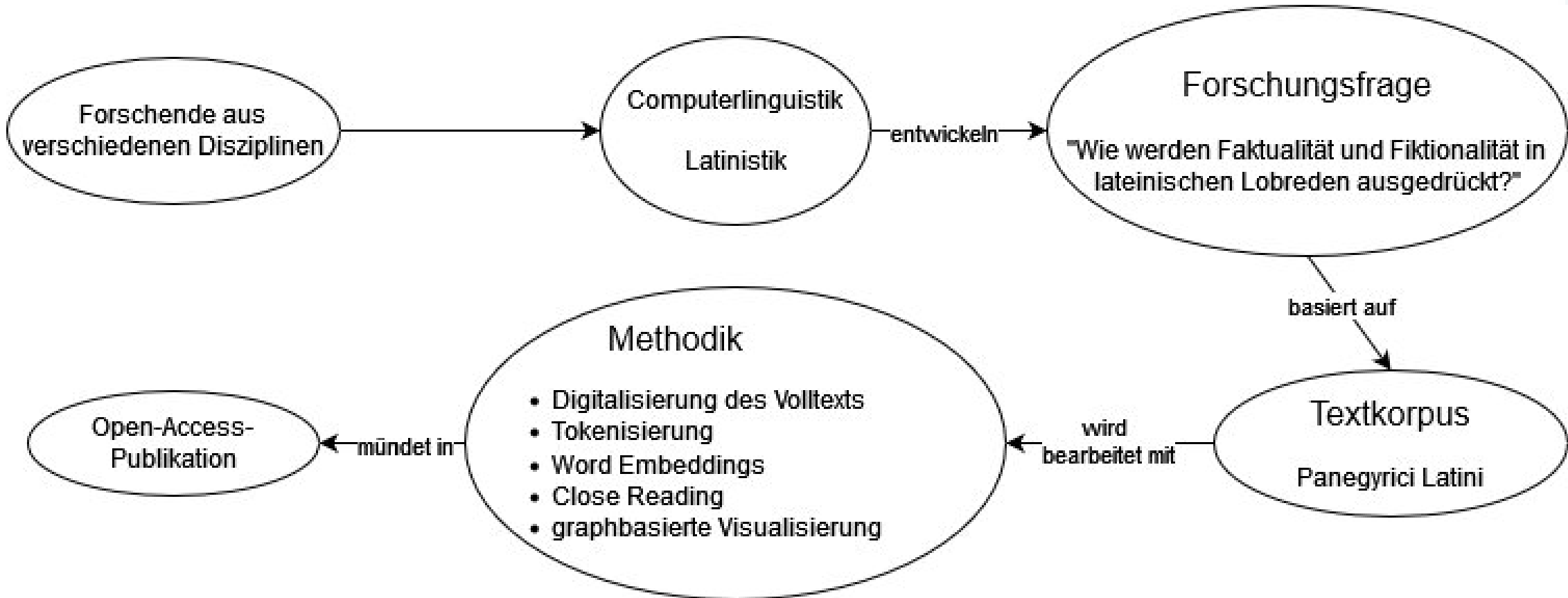


— Digital Research for All —  
**draidalos**




Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft

# 3. Forschungsstandems etablieren



# Ziel der Communityarbeit

- Daidalos-Infrastruktur an authentischen **Forschungsfragen** entwickeln & User **Interface** an Bedarfen ausrichten
- **Interesse** für die Forschungsarbeit mit DH-Methoden wecken
- Konzept einer fachspezifischen **Data Literacy** entwickeln
- NLP-Methoden auch **ohne Code** zugänglich machen
- **Fachkultur** öffnen  für kollaboratives Arbeiten und OA-Publikationen
- und **User** gewinnen

# User Stories durch Forschungstandems

Lösungen gemeinsam entwickeln

# Ablauf Forschungsstandem

---

1. **Kontaktaufnahme**, Basisinfos zu NLP & Ablauf bereitstellen
2. **Erstgespräch** (online) zur Erarbeitung einer Forschungsfrage
3. Forschende: Bereitstellung von **Material** (von Hand ausgewertete Beispielstellen, Korpus & Referenzkorpus)
4. Daidalos: fragenspezifische **Anwendung** von NLP-Methoden, Ergebnisse zurück an Forschende
5. Zweitgespräch (online) zur **Auswertung** der Ergebnisse, ggf. Adaption der Forschungsfrage und Überarbeitung des Inputs
6. **Wiederholung** 3-5, bis Ergebnisse publikationsreif
7. Kooperative Arbeit an einer **OA-Publikation**

# Beispiel einer User Story

---

Ausgangsfrage:

Wie findet man, was nicht explizit im Text genannt wird?

DFG-Heisenberg-Projekt „Formen des ‚Vergessens‘ in der römischen Literatur“, Prof. Dr. Verena Schulz, KU Eichstätt-Ingolstadt

# Konkretisierung der Frage

- Hintergrund: Geschichtsschreiber **erwähnen** bestimmte Ereignisse **nicht**, obwohl sie diese aufgrund ihrer Relevanz erwähnen sollten, wie Parallelquellen zeigen.
- Beispiel: Bei Cassius Dio kommt die **Konferenz von Luca 56 v. Chr.** nicht vor, Cicero lässt bspw. Crassus weg.
- Narrativer Zusammenhang: Gibt es eine **kanonische** Art (Ort, Personal, Thema), diese Konferenz zu **erwähnen**? Welche **Kontexte** sprechen für eine Erwähnung, welche dagegen?
- NLP-Methoden: **NER** und **Lemmatisierung** für lat. & griech. Texte, **Graphdatenbank** für Korpusuche

# Eingereichte Materialien: Konferenz von Luca

- Cassius Dio: Fehlen der Konferenz von Luca insb. bei Cass. Dio 39,24-27; teilweise aber bis Cass. Dio 39,36 als zum Kontext gehörig angegeben
- Parallelstellen mit Erwähnung der Konferenz von Luca: Cic. fam. 1,9,8-9 / Suet. Iul. 24,1 / Plut. Caes. 21 / Plut. Pomp. 51 / Plut. Crass. 14-15 / Plut. Cat. min. 41, 1-2 / App. civ. 2,17 / Vell. 2,46,1-2
- Weitere Informationen: Textstellen mit Übersetzungen, eingesehene Literatur, verwendete Übersetzungen



— Digital Research for All —  
**DAIDALOS**



Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft



# Nachnutzung von Ressourcen

- [Perseids CTS](#): Zitations- & Abrufprotokoll für digitale antike Texte
- [spaCy](#): NLP-Plattform (Modelle, Anwendung)
- [Flair NLP](#): NLP-Framework (Modelle, Anwendung)
- [Huggingface](#): NLP-Plattform, v.a. für Transformer (Modelle, Anwendung)
- [graphANNIS](#): Graphdatenbank für Korpusssuche



— Digital Research for All —  
**daidalos**

Gefördert durch  
**DFG**

Deutsche  
Forschungsgemeinschaft

# Angewandte NLP-Methoden



— Digital Research for All —  
**daidalos**



Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft

## 1. Automatische **Annotation**

- **Tokenisierung:** Wortgrenzen
- **Segmentierung:** Satzgrenzen
- **Lemmatisierung:** Wörterbuchgrundform für jedes Wort
- **Named Entity Recognition:** Personen- und Ortsnamen

## 2. **Verarbeitung** der Annotationen für Anwendungszwecke

- **Korpussuche** mit domänenspezifischer Abfragesprache (ANNIS Query Language)
- **Ziel:** Textstellen mit namentlicher Erwähnung der 3 Zielpersonen oder des Zielorts finden

## 3. **Ausgabe** der gefundenen Textpassagen

## Suet. Iul. 24

sed cum **Lucius Domitius PERSON** consulatus candidatus palam minaretur consulem se effecturum quod praetor nequisset adempturumque ei exercitus, **Crassum PERSON** **Pompeiumque PERSON** in urbem prouinciae suae **Lucam LOC** extractos conpulit, ut detrudendi **Domitii PERSON** causa consulatum alterum peterent, perfecitque tutrumque, ut in quinquennium sibi imperium prorogaretur. qua fiducia ad legiones, quas a re publica acceperat, alias priuato sumptu addidit, unam etiam ex **Transalpinis LOC** conscriptam, uocabulo quoque **Gallico LOC** — **Alauda LOC** enim appellabatur—, quam disciplina cultuque Romano institutam et ornatam postea uniuersam ciuitate donauit. nec deinde ulla belli occasione, ne iniusti quidem ac periculosi abstinuit, tam foederatis quam infestis ac feris gentibus ultro laccessitis, adeo ut senatus quondam legatos ad explorandum statum Galliarum mittendos decreuerit ac nonnulli dedendum eum hostibus censuerint. sed prospere de cedentibus rebus et saepius et plurium quam quisquam umquam dierum supplicationes impetrauit.

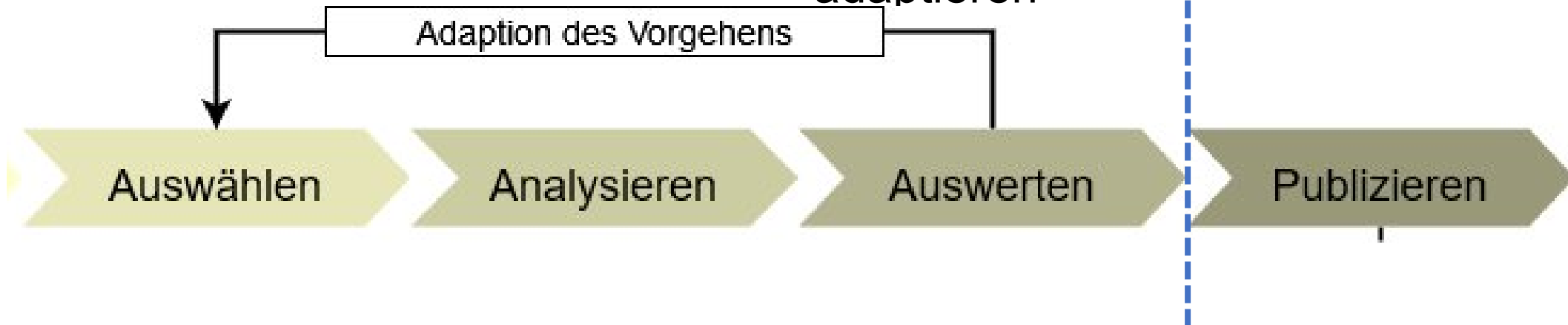
ὁ δὲ Καίσαρ PER ἔν τε Κελτοῖς LOC καὶ Βρεττανοῖς MISC πολλὰ καὶ λαμπρὰ εἰργασμένος, ὅσα μοι περὶ Κελτῶν MISC λέγοντι εἴρηται, πλούτου γέμων ἐς τὴν ὁμορον τῆ Ἰταλίας LOC Γαλατίαν LOC, τὴν ἀμφὶ τὸν Ἑριδανὸν LOC ποταμὸν, ἦκεν, ἐκ συνεχῶς πολέμου τὸν στρατὸν ἀναπαύσων ἐπ' ὀλίγον. ὅθεν αὐτῷ περιπέμποντι ἐς Ῥώμην LOC πολλὰ πολλοῖς χρήματα αἶ τε ἐτήσιοι ἀρχαὶ παρὰ μέρος ἀπήντων καὶ οἱ ἄλλως ἐπιφανεῖς ὅσοι τε ἐς ἡγεμονίας ἐθνῶν ἢ στρατοπέδων ἐξήεσαν, ὡς ἑκατὸν μὲν ποτε καὶ εἴκοσι ῥάβδους ἀμφ' αὐτὸν γενέσθαι, βουλευτὰς δὲ πλείους διακοσίων, τοὺς μὲν ἀμειβομένους ὑπὲρ τῶν ἤδη γεγονότων, τοὺς δὲ χρηματιουμένους, τοὺς δ' ἄλλο τι τοιοῦτότροπον αὐτοῖς ἐξεργασομένους. πάντα γὰρ ἤδη διὰ τούτου ἐπράσσετο στρατιᾶς τε πολλῆς οὐνεκα καὶ δυνάμεως χρημάτων καὶ σπουδῆς ἐς ἅπαντας φιλανθρώπου. ἀφίκοντο δ' αὐτῷ καὶ Πομπήϊος PER καὶ Κράσσοι PER, οἱ κοινῶν τῆς δυναστείας. καὶ αὐτοῖς βουλευομένοις ἔδοξε Πομπήϊον PER μὲν καὶ Κράσσον PER αὐθις ὑπατεῦσαι, Καίσαρι PER δ' ἐς τὴν ἡγεμονίαν ἣν εἶχεν ἐθνῶν, ἄλλην ἐπιψηφισθῆναι πενταετίαν. ὣδε μὲν ἀπ' ἀλλήλων διεκρίθησαν, Πομπηίῳ PER δ' ἐς τὴν ὑπατείαν ἀντιπαρήγγελλε Δομίτιος PER Αἰνόβαρβος· καὶ τῆς κυρίας ἡμέρας ἄμφω κατήεσαν ἔτι νυκτὸς ἐς τὸ πεδῖον ἐς τὴν χειροτονίαν. τῶν δ' ἀμφ' αὐτοῦς ἔριδες ἦσαν καὶ συνεπλέκοντο, μέχρι τις τὸν Δομιτίου PER δαδούχον ἐπάταξε ξίφει. καὶ φυγὴ μετὰ τοῦτο ἦν, Δομίτιός PER τε αὐτὸς ἐς τὴν οἰκίαν διεσώζετο μόλις, καὶ Πομπηίου PER τὴν ἐσθῆτά τινες ἠμαγμένην ἔφερον οἴκαδε. παρὰ τοσοῦτον ἐκάτερος ἦλθε κινδύνου.

# Erste Ergebnisse (Proof of Concept)

Textstelle	gefunden durch Luca	gefunden durch Personennamen	falscher Treffer
Cic. fam. 1,9,9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Suet. Iul. 24,1	<input checked="" type="checkbox"/>	nur Pompeius & Crassus	
Plut. Caes. 21,2	<input checked="" type="checkbox"/>	nur Pompeius & Crassus	
Plut. Caes. 21,3		<input checked="" type="checkbox"/>	
Plut. Caes. 21,4		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Plut. Pomp. 51,3	<input checked="" type="checkbox"/>	nur Pompeius & Crassus	
Plut. Crass. 14,1		<input checked="" type="checkbox"/>	
Plut. Crass. 14,5	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Plut. Cat. min. 41,1		<input checked="" type="checkbox"/>	
Cass. Dio 39,24-36			
Vell. 2,46,1-2		<input checked="" type="checkbox"/>	

# Weiteres Vorgehen

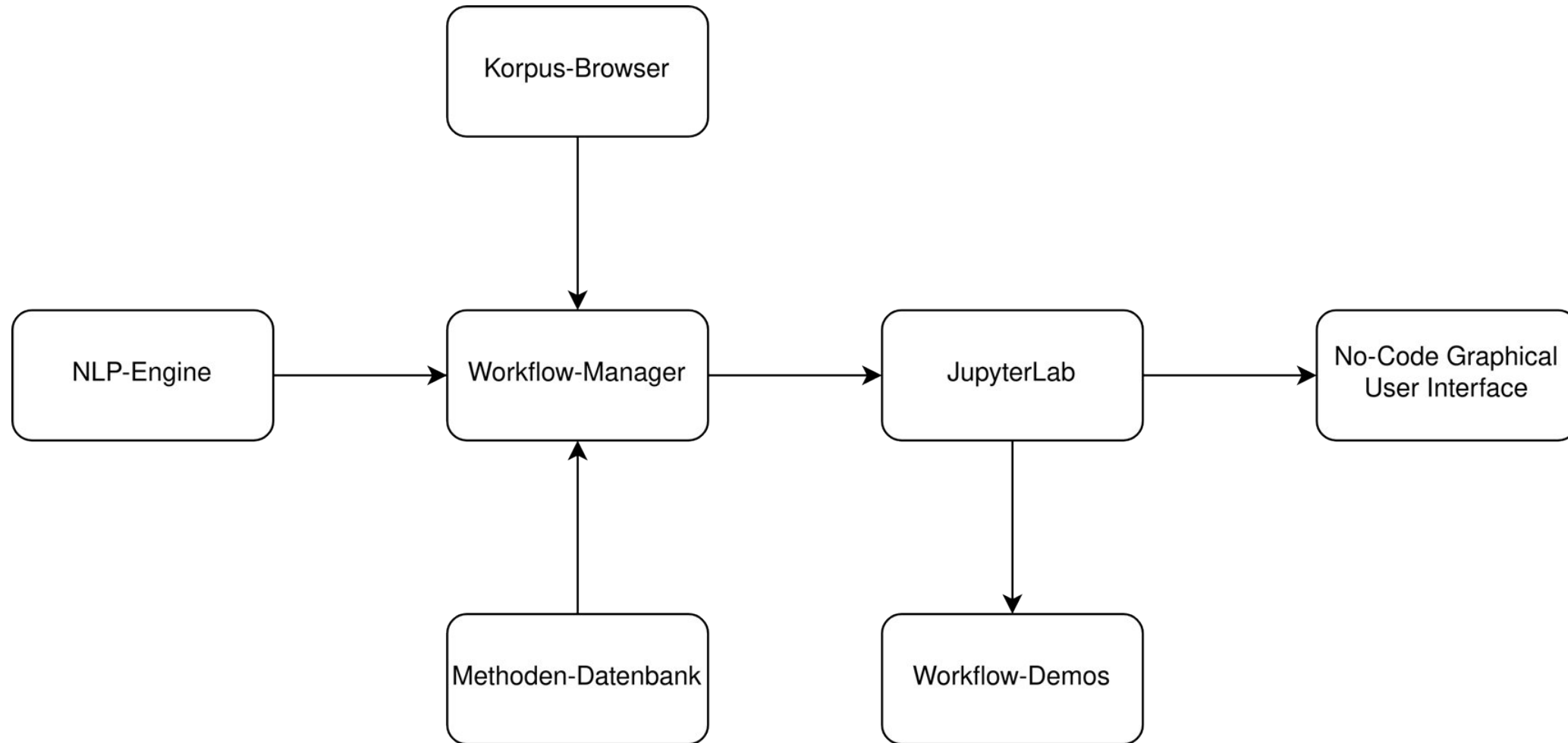
Ergebnisse  
gemeinsam  
reflektieren, ggf.  
Frage  
adaptieren



OA-Publikation  
als Teil eines  
Sammelbandes

# Interaktive Lern- und Forschungsinfrastruktur

# Softwarekomponenten der Infrastruktur





# Level 1: Grundverständnis

The screenshot shows the daidalos-platform.org interface. It features three main sections:

- Forschungsfrage**: A text input field containing the question "Wie findet man, was nicht explizit im Text genannt wird?".
- Textkorpora**: A dropdown menu labeled "Textkorpora auswählen" with a list of options: "Perseus", "LASLA", and "Corpus Corporum".
- Nächste Schritte**: A workflow diagram with three steps: **A** Tokenisierung, **B** Lemmatisierung, and **C** Topic Modelling, connected by horizontal lines.

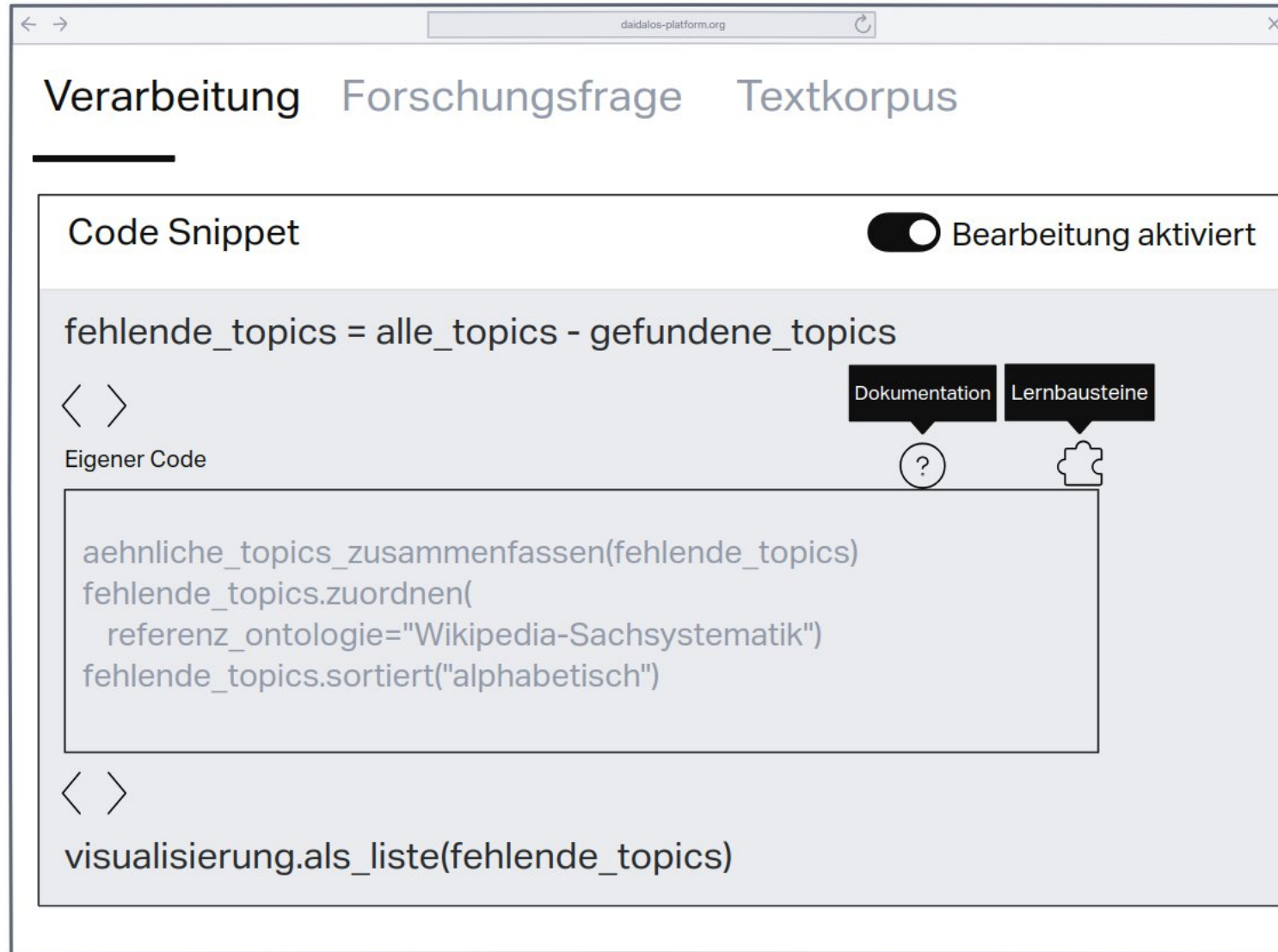
# Level 2: Interpretation

The screenshot shows a web browser window with the URL `daidalos-platform.org`. The page has a navigation bar with three items: `Verarbeitung`, `Forschungsfrage`, and `Textkorpus`. Below the navigation bar is a section titled `Code Snippet` with a scroll indicator. The code snippet contains the following text:

```
fehlende_topics = alle_topics - gefundene_topics
visualisierung.als_liste(fehlende_topics)
```

Below the code, there is a dark button with a white icon of a list and the text `fehlende Topics anzeigen`. Below the code snippet is a box containing the text: `Übung: Welche Themen sind im folgenden Text enthalten?`. To the right of this box is a quiz interface. The quiz question is: `Welche Topics sind in folgendem Text enthalten?`. Below the question is the text `Odi et amo` and a small black box with the text `H-P`. Below the text are four buttons: `Liebe` (green), `Politik` (grey), `Religion` (grey), and `Hass` (green). Below the buttons is the text `Correct!` and a progress indicator showing a green bar, a yellow star, and the text `1/1`.

# Level 3: Manipulation



The screenshot shows a web browser window with the URL `daidalos-platform.org`. The page has a navigation bar with three items: **Verarbeitung**, **Forschungsfrage**, and **Textkorpus**. Below the navigation bar, there is a section titled **Code Snippet** with a toggle switch labeled **Bearbeitung aktiviert** (checked). The main area contains a code editor with the following code:

```
fehlende_topics = alle_topics - gefundene_topics
```

Below the code editor, there are two callout boxes: **Dokumentation** (with a question mark icon) and **Lernbausteine** (with a puzzle piece icon). The code editor also has a label **Eigener Code** and navigation arrows. The code in the editor is:

```
aehnliche_topics_zusammenfassen(fehlende_topics)
fehlende_topics.zuordnen(
  referenz_ontologie="Wikipedia-Sachsystematik")
fehlende_topics.sortiert("alphabetisch")
```

Below the code editor, there are more navigation arrows and the following code:

```
visualisierung.als_liste(fehlende_topics)
```

Und noch etwas Werbung in eigener Sache:  
**Kick-Off-Workshop** am 14.6.2024, HU Berlin  
Schwerpunkt NER & Word Embeddings  
Theorie & Hands-On  
Kontakt: [daidalos-projekt@hu-berlin.de](mailto:daidalos-projekt@hu-berlin.de)

# Literatur (I)

- Bamman, David, and Patrick J Burns. “Latin BERT: A Contextual Language Model for Classical Philology.” arXiv Preprint arXiv:2009.10053, 2020, 1–14.
- Beersmans, Marijke, Evelien de Graaf, Tim Van de Cruys, and Margherita Fantoli. “Training and Evaluation of Named Entity Recognition Models for Classical Latin,” 1–12, 2023. <https://zenodo.org/doi/10.5281/zenodo.8337363>.
- Bjerva, Johannes, and Raf Praet. “Word Embeddings Pointing the Way for Late Antiquity.” In Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), 53–57. Beijing, China: Association for Computational Linguistics, 2015. <https://doi.org/10.18653/v1/W15-3708>.
- Burns, Patrick J, James A Brofos, Kyle Li, Pramit Chaudhuri, and Joseph P Dexter. “Profiling of Intertextuality in Latin Literature Using Word Embeddings,” 4900–4907, 2021. <https://aclanthology.org/2021.naacl-main.389.pdf>.
- Ehrmann, Maud, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello, and Antoine Doucet. “Named Entity Recognition and Classification in Historical Documents: A Survey.” ACM Computing Surveys, 2021. <https://dl.acm.org/doi/pdf/10.1145/3604931>.
- Forstall, Christopher W, Walter J Scheirer, Christopher W Forstall, and Walter J Scheirer. What Is Quantitative Intertextuality? Springer, 2019.
- Hose, Martin. Formen Und Funktionen Griechisch-Römischer Literatur : Aufsätze Zur Literaturgeschichte Und Literaturgeschichtsschreibung. Edited by Annamaria Peri and Tobias Thum. Stuttgart: Franz Steiner Verlag, 2023. <https://dnb.info/1273972600/04>.
- Krahn, Kevin, Derrick Tate, and Andrew C Lamicela. “Sentence Embedding Models for Ancient Greek Using Multilingual Knowledge Distillation.” arXiv Preprint arXiv:2308.13116, 2023. <https://arxiv.org/pdf/2308.13116.pdf>.
- Lave, Jean, and Etienne Wenger. Situated Learning: Legitimate Peripheral Participation. Cambridge university press, 1991.
- Mehler, Alexander, Bernhard Jussen, Tim Geelhaar, Alexander Henlein, Giuseppe Abrami, Daniel Baumartz, Tolga Uslu, and Wahed Hemati. “The Frankfurt Latin Lexicon: From Morphological Expansion and Word Embeddings to SemioGraphs.” arXiv:2005.10790 [Cs], May 21, 2020. <http://arxiv.org/abs/2005.10790>.



# Literatur (II)

- Perrone, Valerio, Simon Hengchen, Marco Palma, Alessandro Vatri, Jim Q. Smith, and Barbara McGillivray. “Lexical Semantic Change for Ancient Greek and Latin.” In *Computational Approaches to Semantic Change*, edited by Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, and Simon Hengchen, 287–310. *Language Variation* 6. Language Science Press, 2021. <https://doi.org/10.5281/zenodo.5040241>.
- Schüller, Katharina. “Ein Framework Für Data Literacy.” *AStA Wirtschafts-Und Sozialstatistisches Archiv* 13, no. 3–4 (2019): 297–317.
- Singh, Pranaydeep, Gorik Rutten, and Els Lefever. “A Pilot Study for BERT Language Modelling and Morphological Analysis for Ancient and Medieval Greek.” In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 128–37. Punta Cana, Dominican Republic (online): Association for Computational Linguistics, 2021. <https://doi.org/10.18653/v1/2021.latechclfl-1.15>.
- Sprugnoli, Rachele, Giovanni Moretti, and Marco Passarotti. “Building and Comparing Lemma Embeddings for Latin. Classical Latin versus Thomas Aquinas.” *IJCoL. Italian Journal of Computational Linguistics* 6, no. 6–1 (June 1, 2020): 29–45. <https://doi.org/10.4000/ijcol.624>.
- Sprugnoli, Rachele, Marco Passarotti, and Giovanni Moretti. “Vir Is to Moderatus as Mulier Is to Intemperans. Lemma Embeddings for Latin,” 2019. [https://www.researchgate.net/profile/Rachele-Sprugnoli/publication/336798734\\_Vir\\_is\\_to\\_Moderatus\\_as\\_Mulier\\_is\\_to\\_Intemperans\\_Lemma\\_Embeddings\\_for\\_Latin/links/5db2a47e92851c577ec259b4/Vir-is-to-Moderatus-as-Mulier-is-to-Intemperans-Lemma-Embeddings-for-Latin.pdf](https://www.researchgate.net/profile/Rachele-Sprugnoli/publication/336798734_Vir_is_to_Moderatus_as_Mulier_is_to_Intemperans_Lemma_Embeddings_for_Latin/links/5db2a47e92851c577ec259b4/Vir-is-to-Moderatus-as-Mulier-is-to-Intemperans-Lemma-Embeddings-for-Latin.pdf).
- Stopponi, Silvia, Nilo Pedrazzini, Saskia Peels-Matthey, Barbara McGillivray, and Malvina Nissim. “Evaluation of Distributional Semantic Models of Ancient Greek: Ancient Language Processing.” *Proceedings of the Ancient Language Processing Workshop*, September 8, 2023, 49–58.
- Yousef, Tariq, Chiara Palladino, and Stefan Jänicke. “Transformer-Based Named Entity Recognition for Ancient Greek.” *Zenodo*, 2023. <https://doi.org/10.5281/zenodo.8107629>.
- Yousef, Tariq, Chiara Palladino, and Farnoosh Shamsian. “Classical Philology in the Time of AI: Exploring the Potential of Parallel Corpora in Ancient Languages,” 2023. [https://www.researchgate.net/profile/Chiara-Palladino/publication/373638720\\_Classical\\_Philology\\_in\\_the\\_Time\\_of\\_AI\\_Exploring\\_the\\_Potential\\_of\\_Parallel\\_Corpora\\_in\\_Ancient\\_Languages/links/64f49e0bfa851147de0fa850/Classical-Philology-in-the-Time-of-AI-Exploring-the-Potential-of-Parallel-Corpora-in-Ancient-Languages.pdf](https://www.researchgate.net/profile/Chiara-Palladino/publication/373638720_Classical_Philology_in_the_Time_of_AI_Exploring_the_Potential_of_Parallel_Corpora_in_Ancient_Languages/links/64f49e0bfa851147de0fa850/Classical-Philology-in-the-Time-of-AI-Exploring-the-Potential-of-Parallel-Corpora-in-Ancient-Languages.pdf).



— Digital Research for All —  
**daidalos**

Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft