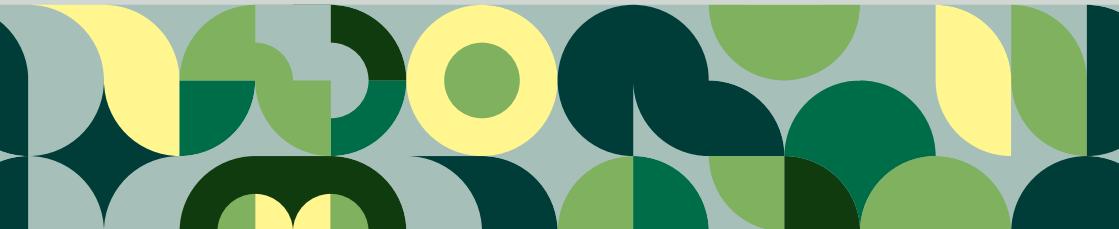


*Pilotprojekt*  
»Neue Potenziale  
für die digitale Lexikographie  
des Deutschen«



# Neue Potenziale für die digitale Lexikographie des Deutschen

„Neue Potenziale für die digitale Lexikographie des Deutschen“ (PDL) ist der Titel eines Pilotprojekts der Bayerischen Akademie der Wissenschaften. Während seiner zweijährigen Laufzeit sollen Grundlagen, Techniken und Strukturen für die Zusammenführung von lexikographischen Daten zur deutschen Sprache unter Berücksichtigung ihrer regionalen, historischen und sozialen Facetten erforscht werden.

Zentrales Ziel ist der Aufbau einer Forschungsdatenbank, um die weitgehend verstreuten und heterogenen Datenbestände systematisch zu integrieren und für fortgeschrittene linguistische Analysen und moderne algorithmische Verfahren verfügbar zu machen. Ebenso stellt die Entwicklung moderner digitaler Tools einen Schwerpunkt dar.

Das Pilotprojekt wird in Kooperation mit mehreren Wissenschaftsakademien der Akademienunion und in enger Zusammenarbeit mit dem Leibniz-Rechenzentrum der BAdW realisiert. Ein wissenschaftlicher Beirat und die Expertise der Digital Humanities vor Ort unterstützen die Entwicklung.

Bereits in der Pilotphase wird eine umfangreiche Vernetzung mit weiteren Akademien, Universitäten und Forschungseinrichtungen angestrebt, um dem Charakter eines föderal organisierten Kompetenzzentrums zu entsprechen.

Fortschritte und Ergebnisse werden auf der Webseite des Projektes unter [pdl.badw.de](http://pdl.badw.de) veröffentlicht.

## Aufgaben, Themen

## und Forschungsbereiche

## des Pilotprojekts

### Vernetzung

Förderung von Austausch und Kooperation der beteiligten Akademien

### Definition Datenformate

Standardisierte Verarbeitung von lexikographischen Daten des Deutschen auf Basis existierender Standards

### Definition Austauschregeln

Erarbeitung klarer Regeln für die Datenweitergabe und -nutzung zur Stärkung der vernetzten Forschung

### Aufbau Datenbank

Performante, verteilte, skalierbare und sichere Speicherung von Wörterbüchern und Korpora

### Aufbau Lexikographie-Portal

Bereitstellung von Daten und Tools für die wissenschaftliche Forschungsgemeinschaft

### Wörterbuchkonsistenzprüfung

Algorithmen zur automatisierten Prüfung auf Fehler und heterogene redaktionelle Praktiken

### KI-gestützte Workflows

Annotations-Copilots und Einsatz von Sprachmodellen zur Vereinfachung und Harmonisierung der Lemma-Erstellung

### LLMs für Low-Resource-Sprachen

Transfer Learning und Modelltraining für Dialekte und Sprachvarietäten mit wenig Trainingsdaten

### Vektor-/Embeddings-Tools

Systematische und standardisierte Berechnung von Embeddings und Speicherung in Vektordatenbanken

### Visualisierungen/Wissensgraphen

Prägnante Kommunikation und Analyse von Wort- und Textzusammenhängen

Ideen

Kontakte

Vernetzung

Sie interessieren sich für die Arbeit des Pilotprojekts?

Nehmen Sie gerne Kontakt mit uns auf:

### **Ansprechpartner**

Wolfgang Huang

wolfgang.huang@adl.badw.de

+49 89 23031-1375

Dr. Fabian Simonjetz

fabian.simonjetz@adl.badw.de

+49 89 23031-1376

### **Webseite**

pdl.badw.de



**Neue Potenziale  
der digitalen Lexikographie  
des Deutschen**

**BAdW**

BAYERISCHE  
AKADEMIE  
DER  
WISSENSCHAFTEN