

**2. Workshop: Präsentation der Anbieter**

**Angebote der deutschen  
Akademien der Wissenschaften**

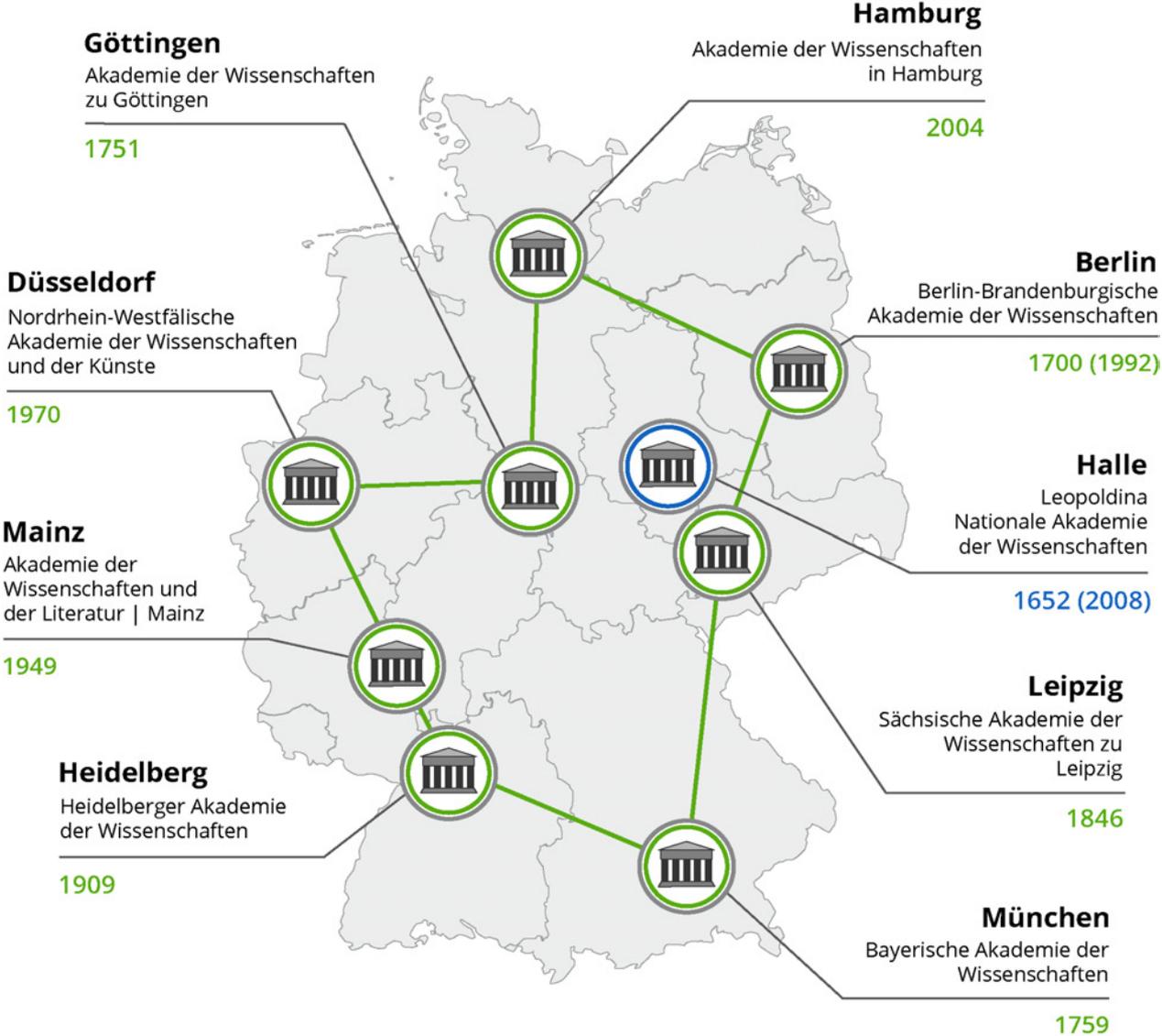
**Berlin | 15.06.2018**

**Für die Akademien: Torsten Schrade (AWLM)**

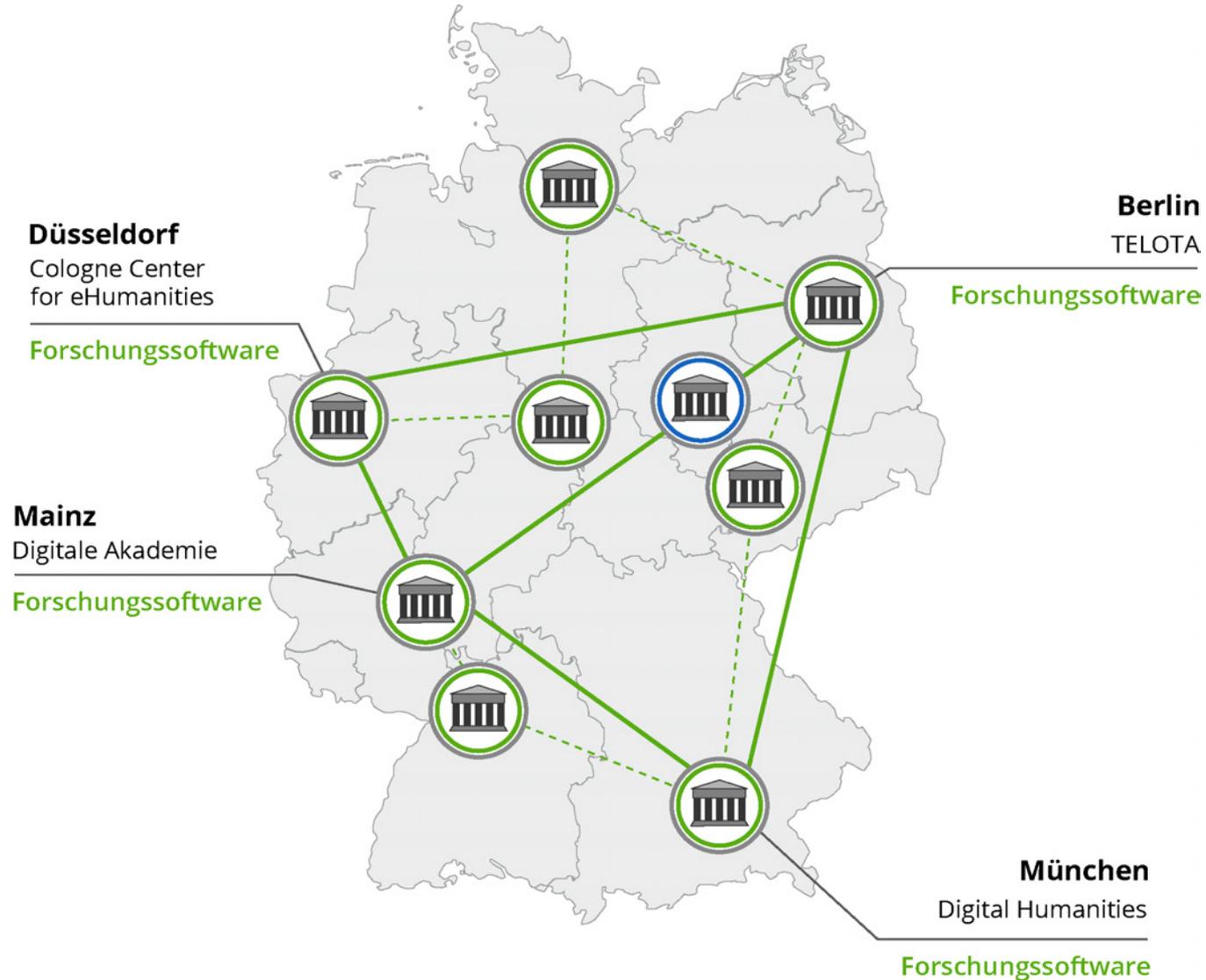
## Gliederung

1. Die Akademien als wissenschaftsgeleitete Forschungsdateninfrastruktur für die geisteswissenschaftliche Grundlagenforschung
2. Kompetenzschwerpunkte der Akademien als Anbietende in einer nationalen Forschungsdateninfrastruktur
3. Relevanz der Akademieangebote für die Fachdisziplinen und Fachcommunities
4. Beispiele für konkrete Angebote der Akademien im Rahmen einer nationalen Forschungsdateninfrastruktur

# Die Akademien als wissenschaftsgeleitete Forschungsdateninfrastruktur



# Entwicklung geisteswissenschaftlicher Forschungssoftware und Datendienste

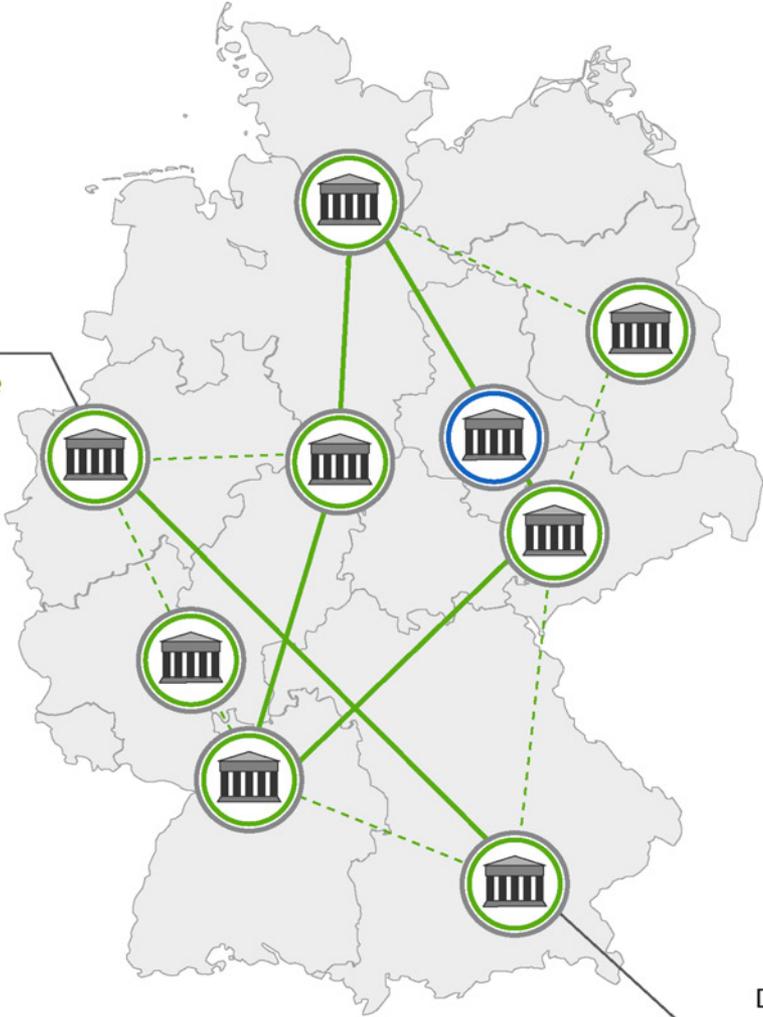


# Daten- und Rechenzentren für die geisteswissenschaftliche Grundlagenforschung



**Düsseldorf**  
Cologne Center  
for eHumanities

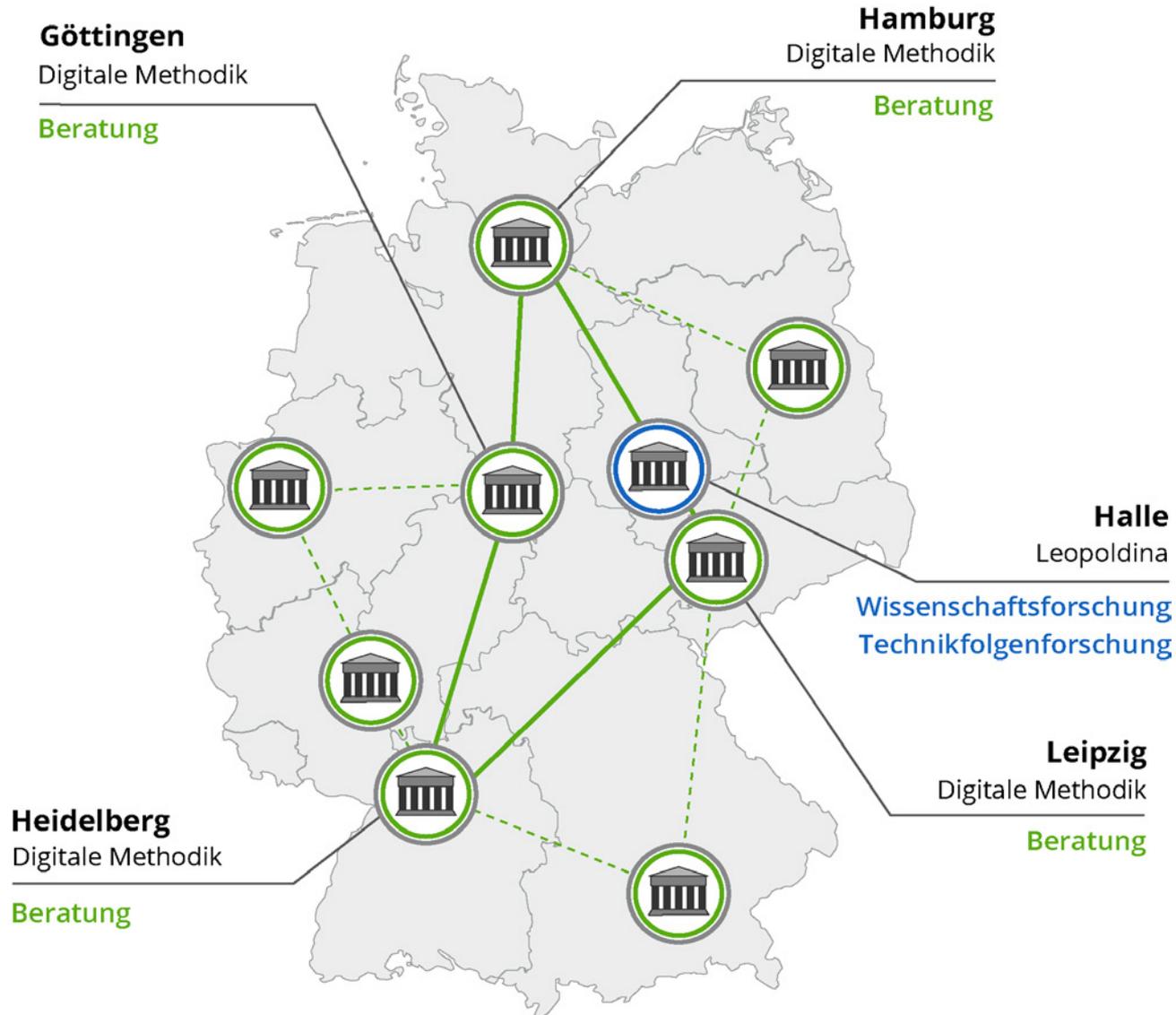
Data Center for the  
Humanities



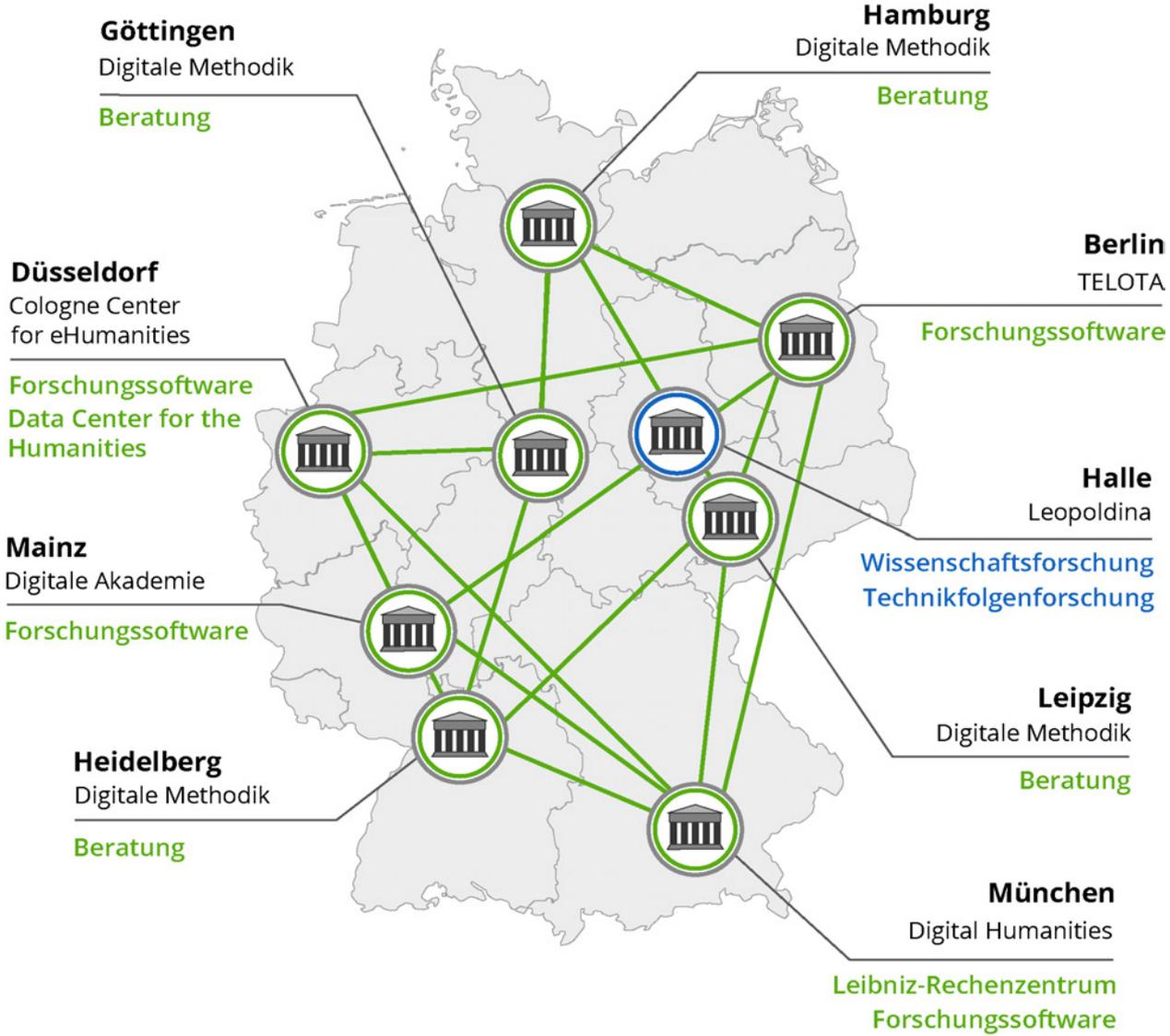
**München**  
Digital Humanities

Leibniz-Rechenzentrum

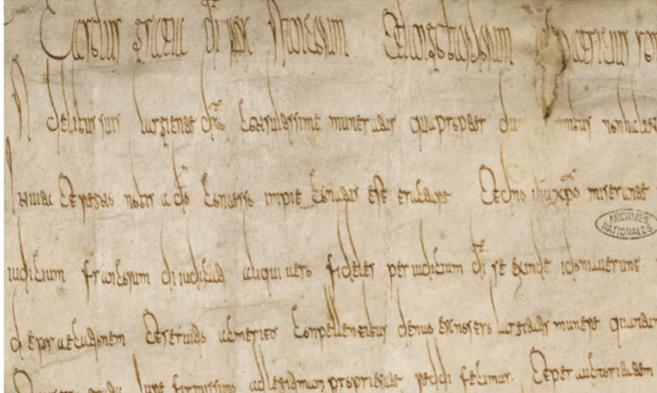
# Beratung, Wissenstransfer, Technikfolgenforschung



# Kompetenznetzwerk für Software, Infrastruktur und Digitale Methodik



# Kompetenzen in der Verarbeitung zahlreicher Forschungsdatentypen



**Geschichte: Urkunde Karls des Großen von 797**



**Kunstgeschichte: Glasmalerei**



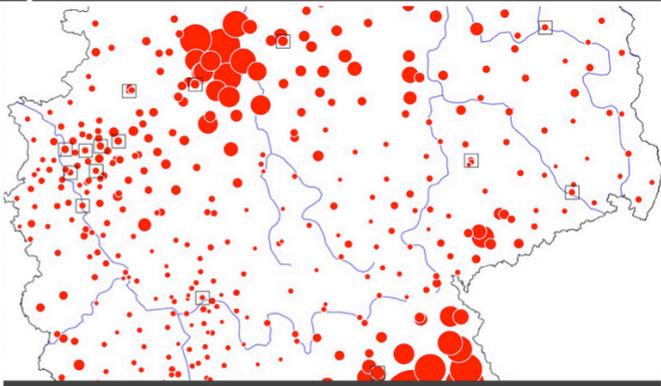
**Epigraphik: Inschrift auf dem Mainzer Domportal**



**Theologie: Religiöse Streitschriften**

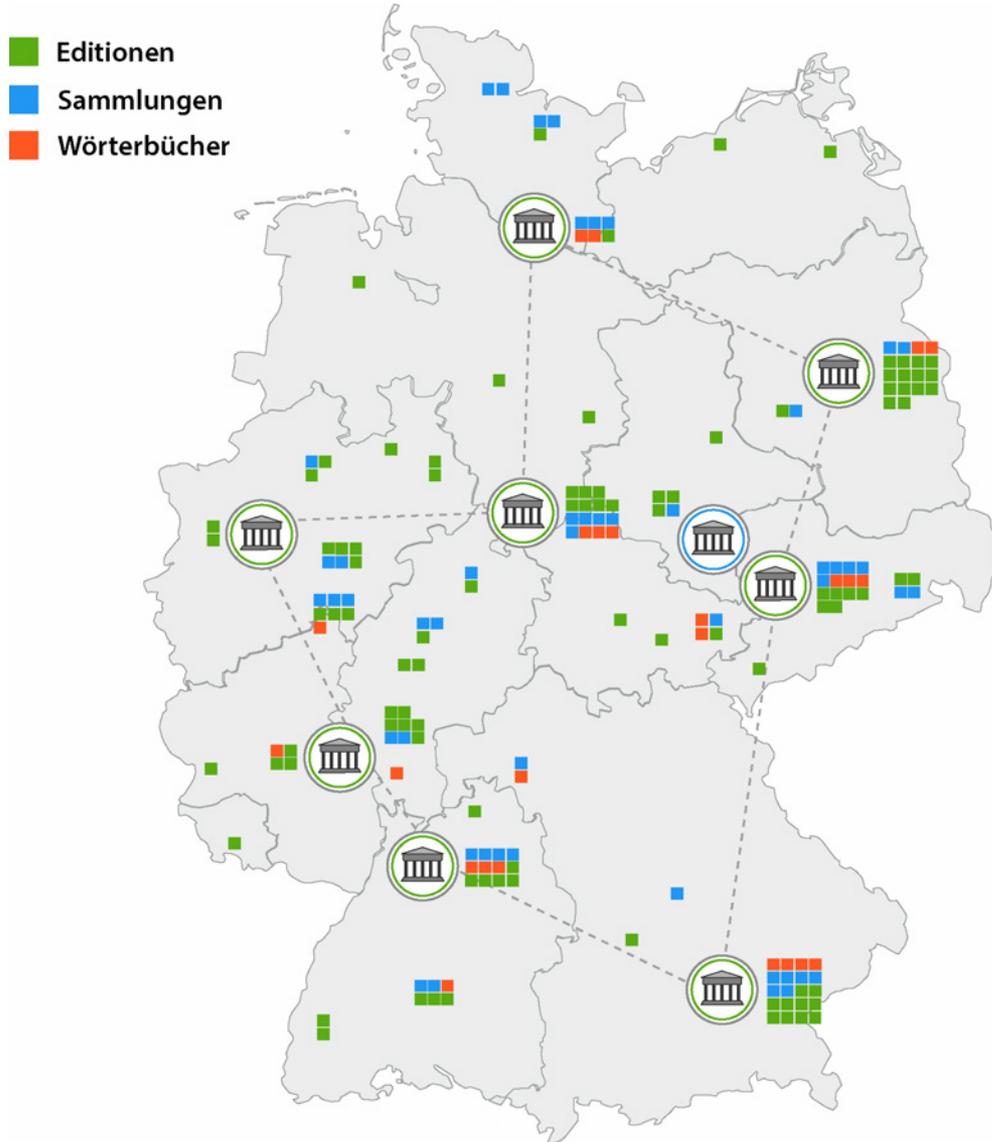


**Musik: Autograph aus Glucks Oper Alceste**



**Linguistik: Kartierte Verteilung des Namens 'Meier'**

# Schwerpunkte der Akademien: Editionen, Sammlungen, Wörterbücher

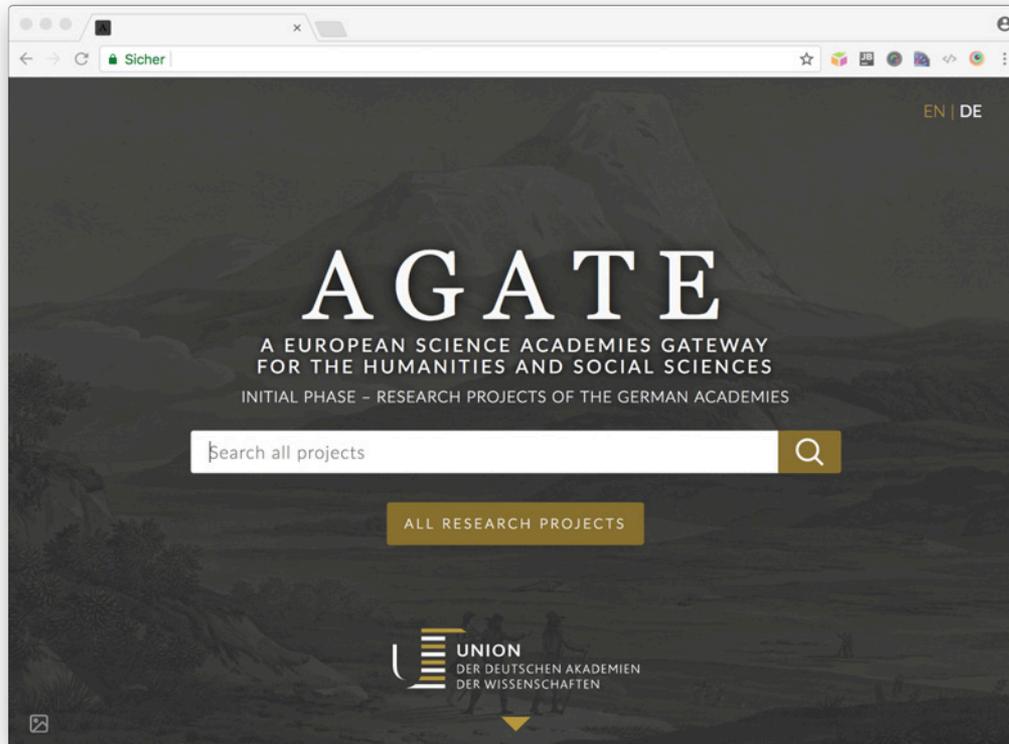


**Kartierung der (digitalen)  
Editionen, Sammlungen und  
Wörterbuchprojekte des  
Akademienprogramms in  
Deutschland** (Quelle: AGATE)

## Zwischenfazit

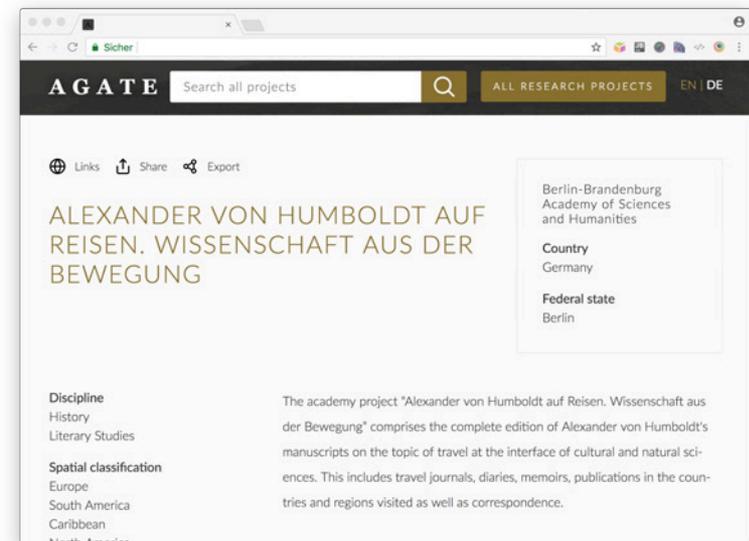
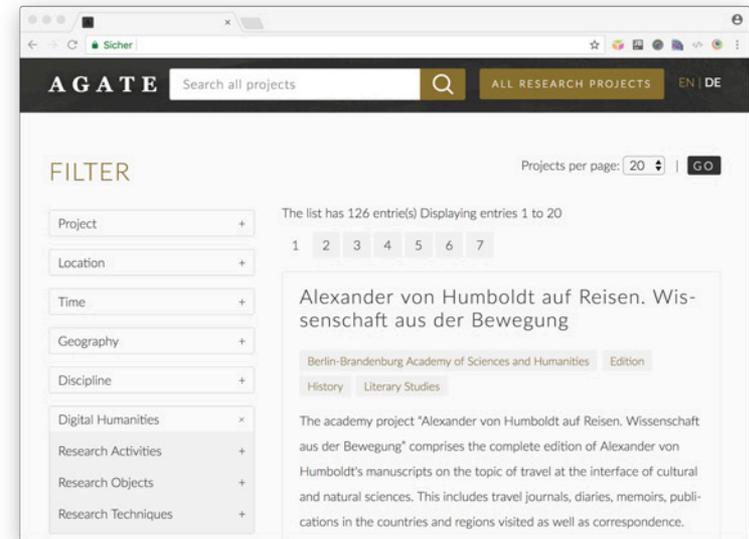
- **Akademien bilden bereits ein wissenschaftsgeleitetes Kompetenznetzwerk aus „Knoten“ (Akademien) und verbindenden gemeinsamen „Kanten“ (Akademievorhaben)**
- **300-jährige Erfahrung in der geisteswissenschaftlichen Grundlagenforschung**
- **Jahrzehntelange Erfahrung in den Bereichen Forschungssoftware-Entwicklung, Forschungsdateninfrastrukturen, Beratung und Technikfolgenforschung für die Geisteswissenschaften**
- **Besondere Kompetenzschwerpunkte im Bereich von digitalen Editionen, Sammlungen und Wörterbüchern**
- **Die Akademienforschung und die dort entwickelten, nachhaltigen Softwarelösungen, Infrastrukturen und (Daten-)dienste besitzen hohe Relevanz für die darauf aufbauenden Fachdisziplinen**

# Forschungsinformationssystem für die europäische Akademieforschung

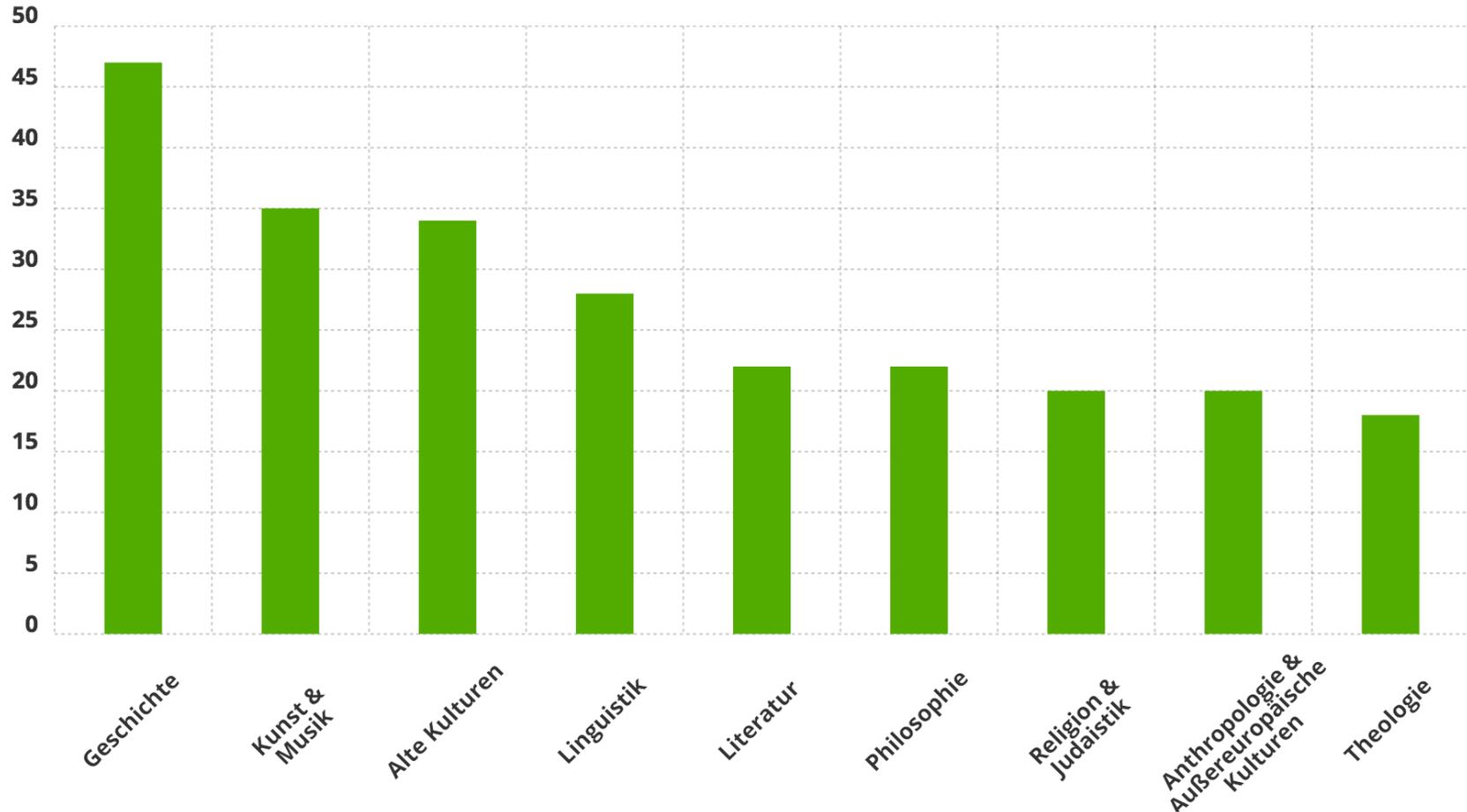


<https://agate.academy/>

**Forschungsinformationssystem enthält alle Akademievorhaben seit 1979; Klassifizierungen und Links zu Digitalen Ressourcen & Werkzeugen**



# Transfer von Wissen, Tools, Diensten und Infrastruktur in die Fachdisziplinen



**Verortung der Akademieforschung in ausgewählten Fachdisziplinen nach DFG-Systematik**  
(Quelle: AGATE)

# Relevanz und Transfer: Forschungsobjekte im Akademienprogramm



Video	Standards	Methods	Research Results	Language	Text Bearing Objects	Metadata
3D						
Sound	Sheet Music	Tools				
Multi media	Software	File	Persons	Research	Manuscript	Text
Projects	Map					
Com puters	Link	Bibliographic Listings	Data			
Infra structure		Research Process		Literature		
Inter action	VREs	Images				
Named Entities						

**Klassifikation von Forschungsgegenständen im Akademienprogramm nach TaDiRAH (Quelle: AGATE)**

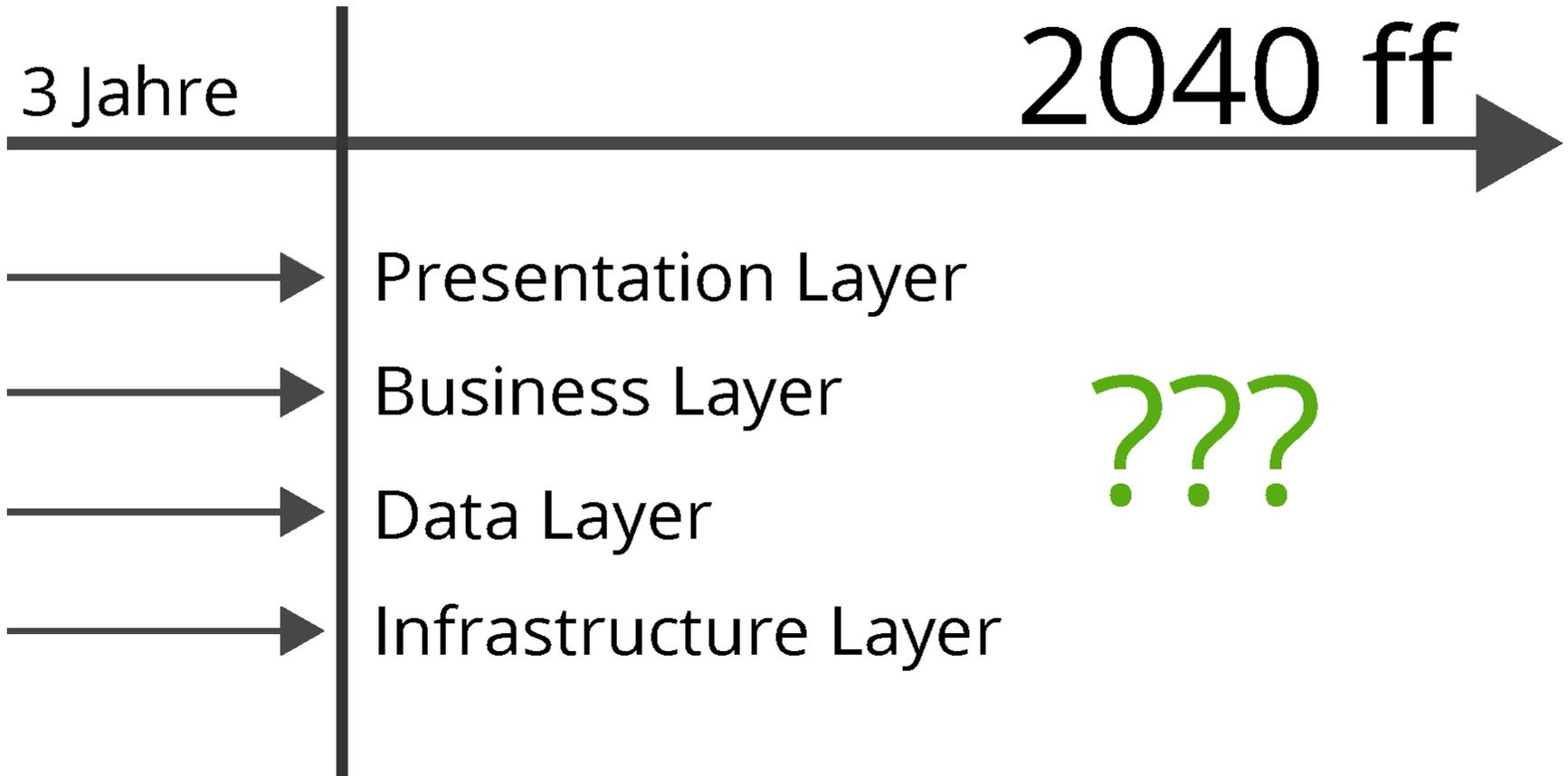
# Relevanz und Transfer: Digitale Methodik im Akademienprogramm



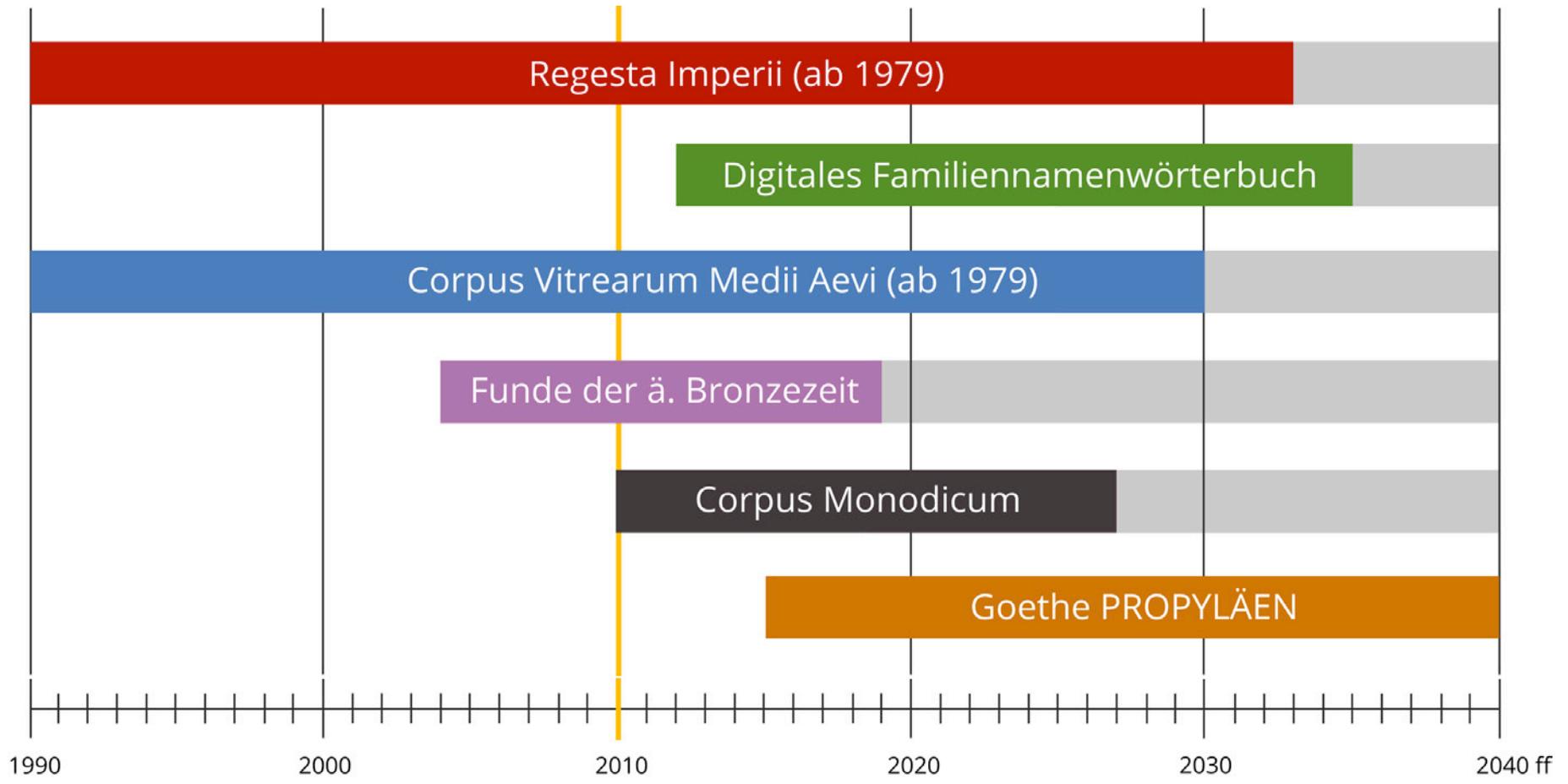
Modeling	Data Recognition	Teaching Learning	Collaboration	Translation	Programming	Relational Analysis	Discovering
Spatial Analysis	Cleanup	Stylistic Analysis					
Project Management							
Network Analysis	Theorizing	Conversion	Organizing		Editing	Writing	Contextualizing
Community Building	Designing	Identifying	Commenting				
Web development		Imaging	Structural Analysis	Content Analysis		Gathering	
Give Overview	Sharing	Preservation	Transcription		Archiving		
Assessing		Communicating	Visualization				

**Klassifikation digitaler Forschungsmethoden im Akademienprogramm nach TaDiRAH (Quelle: AGATE)**

# Akademien: Nachhaltige Entwicklung von Software, Infrastruktur, Diensten



# Akademien: Nachhaltige Entwicklung von Software, Infrastruktur, Diensten



**Software ++ Infrastruktur ++ Dienste**



# Beispiele aus dem Portfolio der Akademien: Forschungssoftware

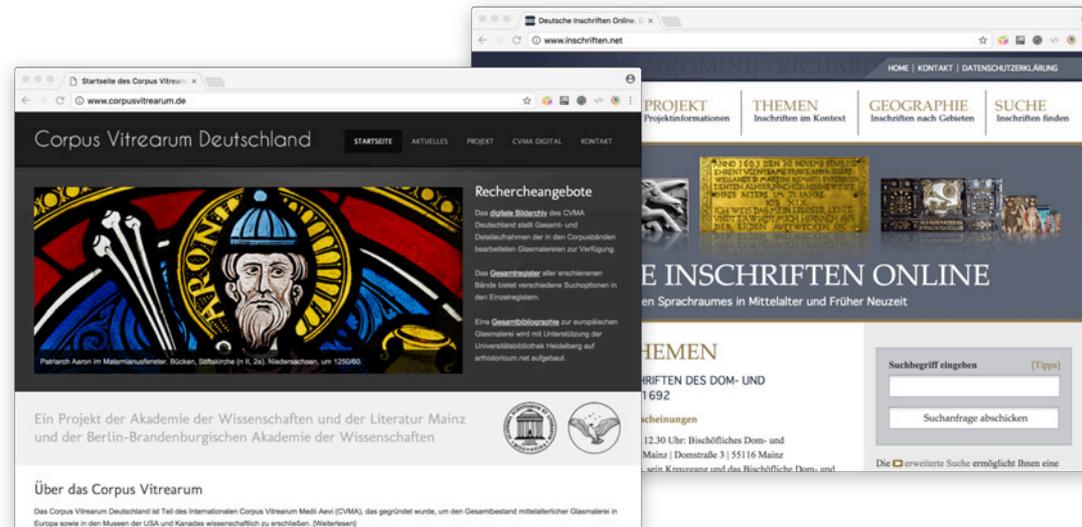
## ediarum

- **Digitale Arbeitsumgebung für Editionen; Konsequenter Einsatz von Standards (DTABf)**
- **Umfasst alle Schichten einer Digitalen Edition**
- **Einsatz in 25 Projekten**
- <https://goo.gl/mwLbXB>



## Cultural Heritage Framework

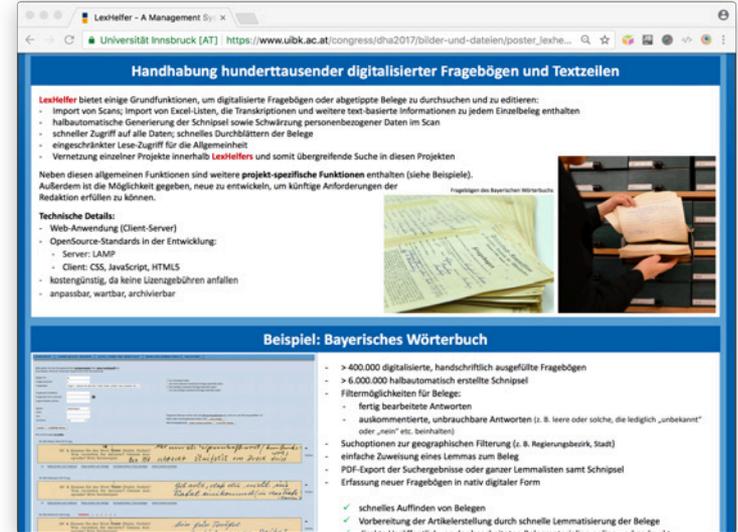
- **Webbasiertes Framework für Sammlungen und Editionen**
- **Einsatz in 10 Akademievorhaben und 4 Drittmittelprojekten**
- **6 beteiligte Akademien**
- <https://goo.gl/EQH6fx>



# Beispiele aus dem Portfolio der Akademien: Forschungssoftware

## LexHelfer

- **Webbasiertes Wörterbuchredaktionssystem mit Vernetzungsfunktionen**
- **Schnelle Lemmatisierung der Belege, Veröffentlichung der Belegmaterialien online und gedruckt**
- **Beispiel Bayerisches Wörterbuch: 6 Mio Belege, 400.000 digitalisierte Fragebögen**
- <https://goo.gl/QzCpJr>



**Handhabung hunderttausender digitalisierter Fragebögen und Textzellen**

LexHelfer bietet einige Grundfunktionen, um digitalisierte Fragebögen oder abgetippte Belege zu durchsuchen und zu editieren:

- Import von Scans; Import von Excel-Listen, die Transkriptionen und weitere text-basierte Informationen zu jedem Einzelbeleg enthalten
- halbautomatische Generierung der Schnipsel sowie Schwärzung personenbezogener Daten im Scan
- schneller Zugriff auf alle Daten; schnelles Durchblättern der Belege
- eingeschränkter Les-Zugriff für die Allgemeinheit
- Vernetzung einzelner Projekte innerhalb LexHelfers und somit übergreifende Suche in diesen Projekten

Neben diesen allgemeinen Funktionen sind weitere **projekt-spezifische Funktionen** enthalten (siehe Beispiele). Außerdem ist die Möglichkeit gegeben, neue zu entwickeln, um künftige Anforderungen der Redaktion erfüllen zu können.

**Technische Details:**

- Web-Anwendung (Client-Server)
- OpenSource-Standards in der Entwicklung:
  - Server: LAMP
  - Client: CSS, JavaScript, HTML5
- kostengünstig, da keine Lizenzgebühren anfallen
- anpassbar, wartbar, archivierbar

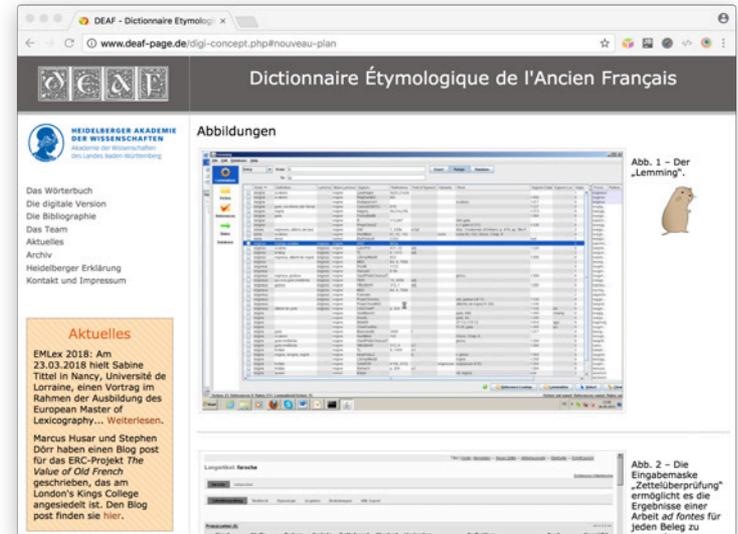
**Beispiel: Bayerisches Wörterbuch**

- > 400.000 digitalisierte, handschriftlich ausgefüllte Fragebögen
- > 6.000.000 halbautomatisch erstellte Schnipsel
- Filtermöglichkeiten für Belege:
  - fertig bearbeitete Antworten
  - auskommentierte, unbrauchbare Antworten (z. B. leere oder solche, die lediglich „unbekannt“ oder „nie“ etc. beinhalten)
- Suchoptionen zur geographischen Filterung (z. B. Regierungsbezirk, Stadt)
- einfache Zuweisung eines Lemmas zum Beleg
- PDF-Export der Suchergebnisse oder ganzer Lemmalisten samt Schnipsel
- Erfassung neuer Fragebögen in nativ digitaler Form

✓ schnelles Auffinden von Belegen  
✓ Vorbereitung der Artikelstellung durch schnelle Lemmatisierung der Belege

## DEAF

- **Quelloffenes Wörterbuchredaktionssystem**
- **Automatisiert mehrere Redaktionsschritte, z.B. Anreicherung mit bibliographischen Daten**
- **Einsatz auch in französischen, schweizer, britischen und italienischen Projekten**
- <https://goo.gl/bshJzP>



**DEAF - Dictionnaire Étymologique**

**Dictionnaire Étymologique de l'Ancien Français**

**Abbildungen**

Das Wörterbuch  
Die digitale Version  
Die Bibliographie  
Das Team  
Aktuelles  
Archiv  
Heidelberger Erklärung  
Kontakt und Impressum

**Aktuelles**

EMLex 2018: Am 23.03.2018 hielt Sabine Tittel in Nancy, Université de Lorraine, einen Vortrag im Rahmen der Ausbildung des European Master of Lexicography... Weiterlesen.

Marcus Husar und Stephen Dorr haben einen Blog post für das EIC-Projekt *The Value of Old French* geschrieben, das am London's Kings College angesiedelt ist. Den Blog post finden sie [hier](#).

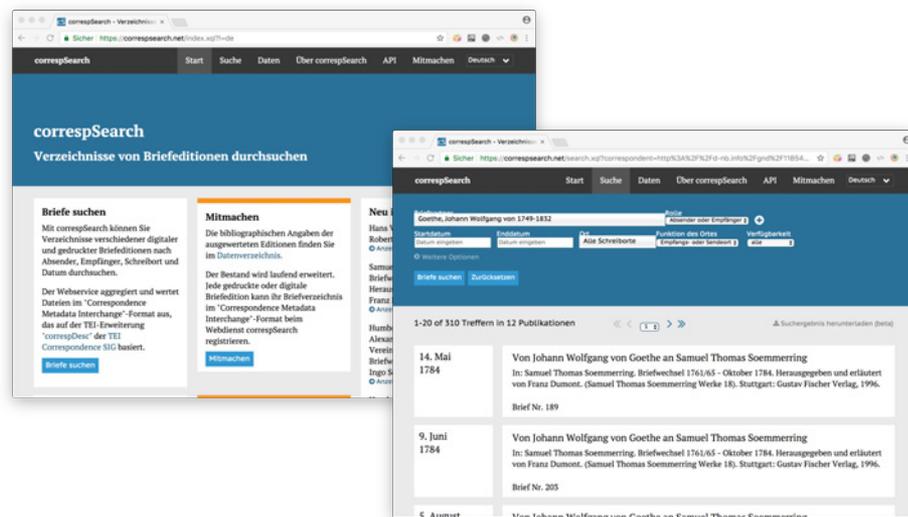
Abb. 1 - Der „Lemming“.

Abb. 2 - Die Eingabemaske „Zettelüberprüfung“ ermöglicht es die Ergebnisse einer Arbeit ad fontes für jeden Beleg zu überprüfen.

# Beispiele aus dem Portfolio der Akademien: Such-/Metadatendienste

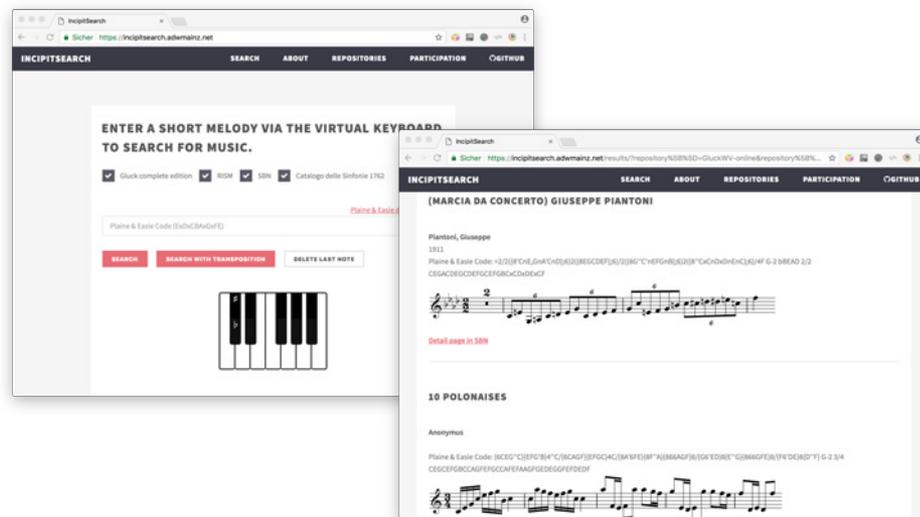
## correspSearch

- **Metasuche für Briefkorrespondenzen**
- **Briefmetadaten von 37.467 Briefen aus 124 gedruckten und/oder digitalen Editionen**
- **Standards und offene Schnittstellen**
- <https://goo.gl/TRQwmK>



## IncipitSearch

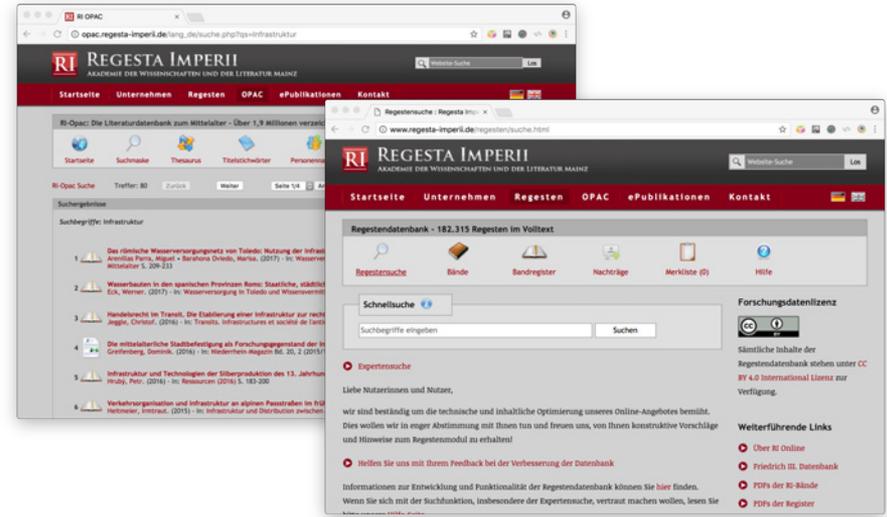
- **Metasuche für notierte Musik**
- **946.514 Notenincipits aus freien Quellen (RISM, SBN, Werkverzeichnisse, gedruckte Kataloge)**
- **Basierend auf LOD-Technologien; einfaches Schema, offen für die Einbeziehung weiterer Verzeichnisse**
- <https://goo.gl/CcxsTr>



# Beispiele aus dem Portfolio der Akademien: Fachinformationssysteme

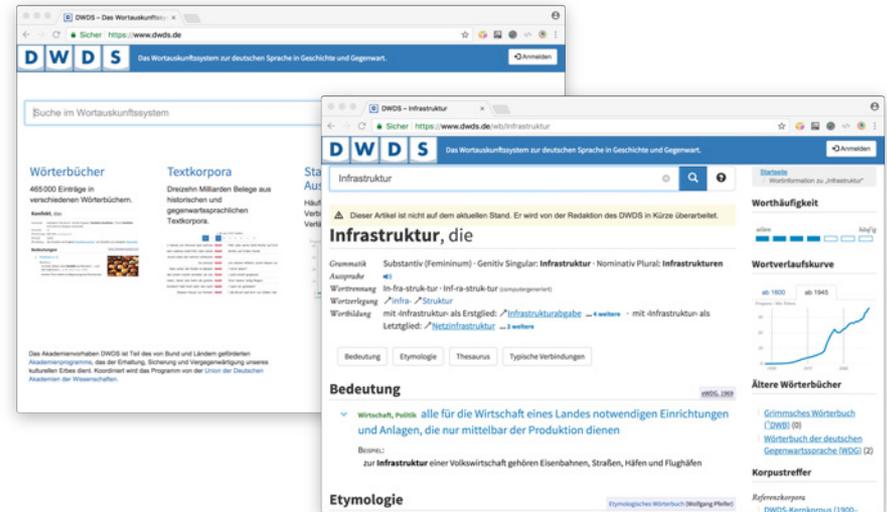
## Regesta Imperii Online

- Fachinformationssystem für die Mittelalterforschung (800 bis 1500)
- Literaturinformationssystem (OPAC) mit 2,1 Mio. Titeln; Quelldatenbank mit 180.000 Regesten
- CC-lizenzierte Forschungsdaten; 430.322 Nutzer, 1.6 Mio. Aufrufe
- <https://goo.gl/ofaev1>



## DWDS

- Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart
- 465.000 Wörterbucheinträge, 13 Milliarden Belege
- Häufigkeiten, typische Verbindungen und zeitliche Verläufe über 400 Jahre
- <https://goo.gl/rw49Vt>

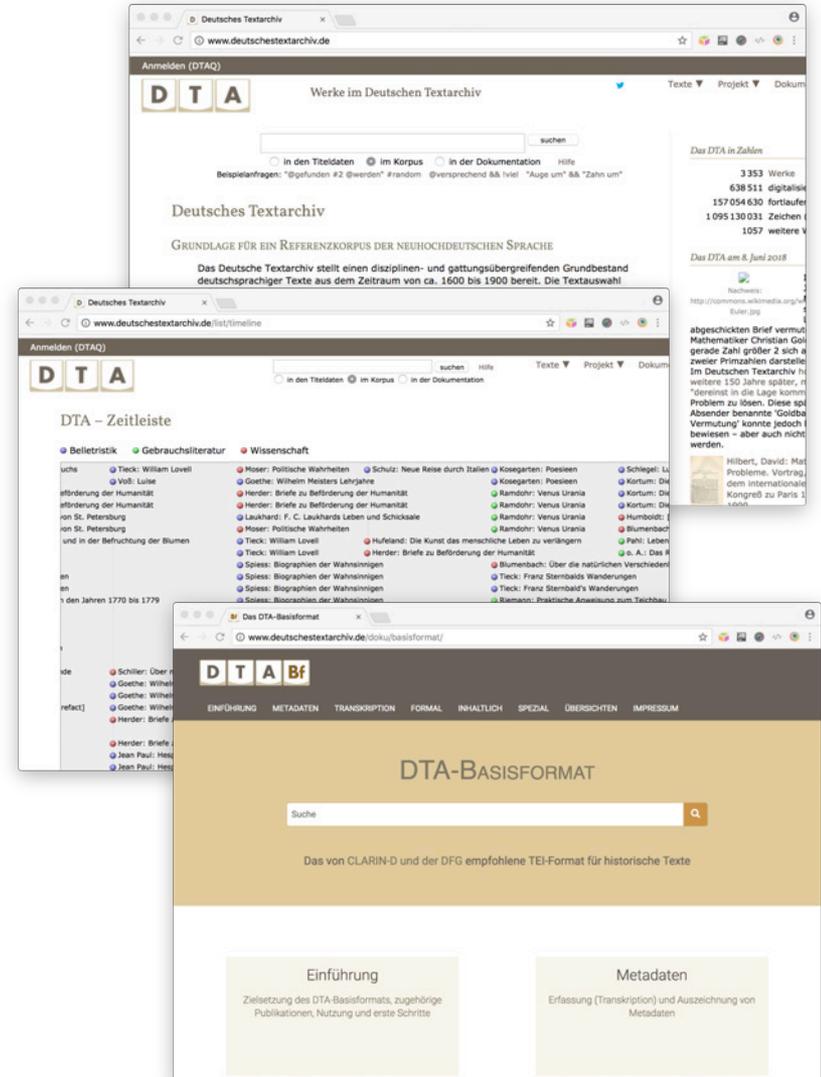


# Beispiele aus dem Portfolio der Akademien: Standardisierung

## Deutsches Textarchiv

- Das DTA stellt einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900 bereit
- 3 353 Werke, 638 511 digitalisierte Seiten, 157 054 630 fortlaufende Wortformen bzw. 1 Mrd. Zeichen
- Freie Nutzbarkeit der Forschungsdaten (CC-Lizenz), offene Schnittstellen und Webservices (DTA::CAB Webservice zur fehlertoleranten Analyse historischer Texte)
- Standardisierung: DTA-Basisformat für einheitliche Erschließungstiefe und Vermeidung von Ambiguitäten => von der DFG empfohlene TEI/P5 XML Kodierung

• <https://goo.gl/6uPFs7>



The image displays three screenshots of the German Text Archive (DTA) website and its services:

- Top Screenshot:** The main DTA homepage. It features a search bar, navigation tabs (DTA), and a sidebar with statistics: 3353 Werke, 638 511 digitalisierte Seiten, 157 054 630 fortlaufende Wortformen, and 1 057 weitere Zeichen. It also mentions the DTA was established on June 8, 2018.
- Middle Screenshot:** The 'DTA - Zeitleiste' (Timeline) page, showing a list of works categorized by discipline (e.g., Belletristik, Wissenschaft) and time period (e.g., 1770 bis 1779).
- Bottom Screenshot:** The 'DTA-Basisformat' page, which is the recommended TEI/P5 XML format for historical texts. It includes a search bar and sections for 'Einführung' (Introduction) and 'Metadaten' (Metadata).

## Fachtagungen, Symposien, Workshops

### Veranstaltungen der AG-eHumanities zu zentralen DH-Themen (seit 2007):

- **Geisteswissenschaftliche Forschungsdaten (2017)**
- **Nachnutzung und Nachnutzbarkeit (2016)**
- **Historische Semantik und Semantic Web (2015)**
- u.w.: <https://goo.gl/jx1RD6>

## Mitarbeit in Gremien

### DHd AGs:

- **AG DH-RSE**
- **AG Datenzentren**
- **AG Graphentechnologien**
- **u.w.**

## Studium, Ausbildung, Lehre

### Mitwirkung in DH-Studiengängen

- **Berlin**
- **Köln**
- **Mainz**
- **u.w.**

# Portfolio der Akademien: Alle Beispiele im Überblick

## Forschungsinformationssysteme

**AGATE (Europäische Akademienforschung)**  
**DEV.AGATE (Tools, Software; im Aufbau)**  
**LOD.AGATE (Linked Open Data; im Aufbau)**

## Forschungssoftware

**ediarum (Editionsumgebung)**  
**Cultural Heritage Framework (Sammlungen)**  
**LexHelfer (Wörterbuchredaktion)**  
**DEAF Redaktionssystem**

## Suchdienste & Webservices

**correspSearch (Briefeditionen vernetzen)**  
**IncipitSearch (Suche auf notierter Musik)**  
**Datumserkenner / DTA::CAB**  
**XTriples (TEI/XML → RDF/LOD)**

## Fachinformationssysteme

**RI-OPAC (Mittelalterliche Geschichte)**  
**DWDS (Sprachwissenschaften)**  
**Maya-Textdatenbank (Alte Kulturen)**

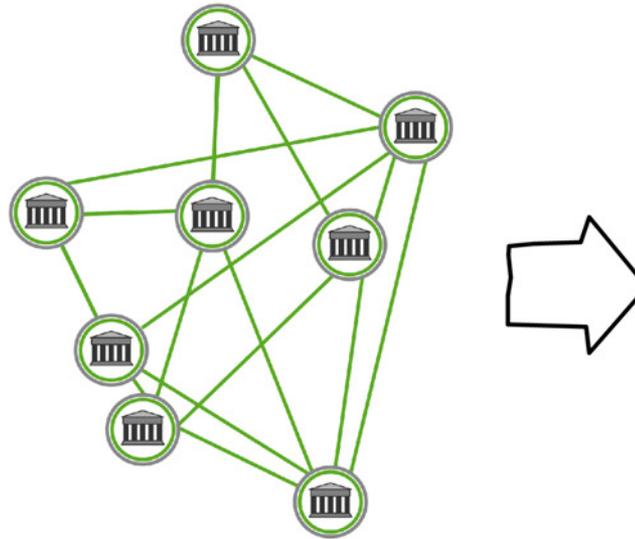
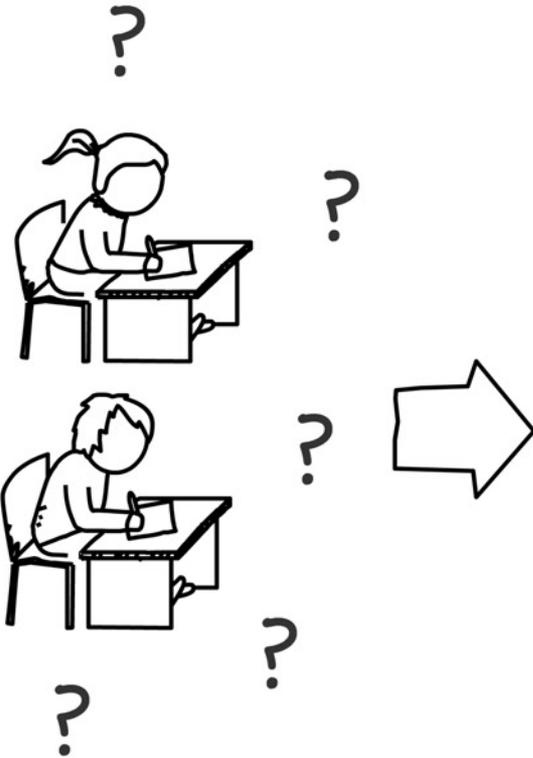
## Standardisierung & Digitale Methodik

**DTABf (DFG-empfohlene TEI-Kodierung)**  
**correspDesc (Metadaten für Briefeditionen)**  
**Digitale Editorik, Netzwerkforschung**  
**Graphentechnologien, LOD**

## Kompetenzaufbau & Wissenstransfer

**Lehre in DH-Studiengängen**  
**Veranstaltungen der AG eHumanities**  
**Sommerschulen, Tagungen, Symposien**  
**Technikfolgenforschung**

# Konkreter Beispielprozess: Erstellung einer Digitalen Edition



Nachhaltige Bereitstellung
Linked Data, Nachnutzung
Schnittstellen, Normdaten
Präsentationsschicht
Texttechnologien, Datendienste
Datenstandards, Modellierung
Editionsumgebung
Systemarchitektur

**1** Wie erstellt man eine Digitale Edition?

**2** Kompetenznetzwerk der Akademienunion

**3** "Schichten" der Digitalen Edition

## Zusammenfassung

- Die Akademien sind sehr daran interessiert und in der Lage, ihre Forschungssoftware, Informationssysteme, Such- und Datendienste, Beratungs- und Wissensvermittlungsangebote im Rahmen einer nationalen Forschungsdateninfrastruktur zu öffnen und wissenschaftsgeleitet weiterzuentwickeln
- Schwerpunkte der Akademien liegen auf der Standardisierung und der Entwicklung nachhaltiger digitaler Lösungen, insbesondere im Bereich von Editions-, Sammlungs-/Erschließungs- und Wörterbuchprojekten
- Die Akademien können im Rahmen einer nationalen Forschungsdateninfrastruktur dauerhaft Verantwortung für die von ihnen betreuten Materialcluster übernehmen
- Akademien bieten institutionelle Konstanz und bringen dauerhafte personelle Kompetenzen in die Forschungs- und Entwicklungsprozesse ein

**2. Workshop: Präsentation der Anbieter**

**Danke für die Aufmerksamkeit**

**Berlin | 15.06.2018**