

Nachhaltigkeit digitaler Ressourcen aus institutioneller Sicht

25 Jahre TELOTA – demnächst mit KI?

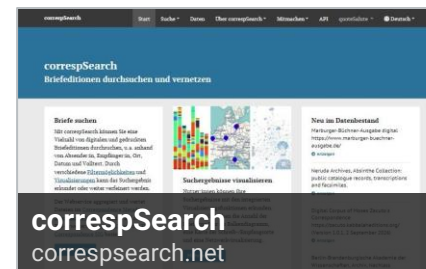
Alexander Czmieł

TELOTA – The Electronic Life of the Academy
Berlin-Brandenburgische Akademie der Wissenschaften
czmieł@bbaw.de

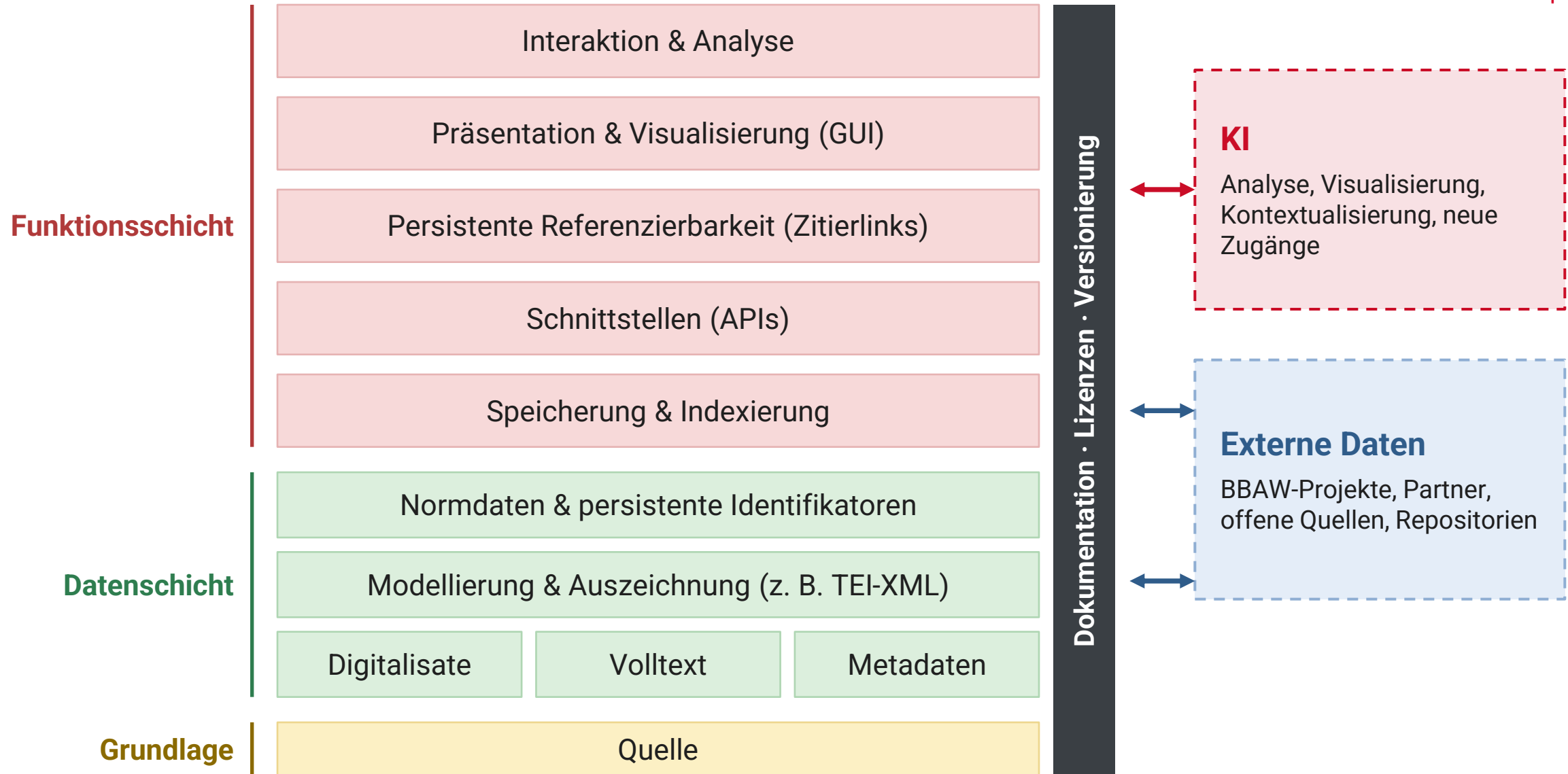
Nachhaltigkeit und generative KI
Göttingen, 16. Oktober 2026

25 Jahre digitale Ressourcen bei TELOTA

- **DH-Abteilung der BBAW**
 - **Softwareentwicklung für den gesamten Forschungszyklus** in geisteswissenschaftlichen Forschungsvorhaben
 - Digitalisierung, Datenerfassung, Annotation, Publikation, Visualisierung, Abfrage, Analyse, Interoperabilität, Speicherung
 - Digitale Editionen
 - **Datenzentrum & Forschungsinfrastruktur** für die (digitalen) Geisteswissenschaften und das Forschungsdatenmanagement
 - **KI-Lab**
- **ca. 40 Projekte**
- **Open Access, Open Data, Open Source**



Schichtenmodell einer digitalen Edition



Technische Komplexität

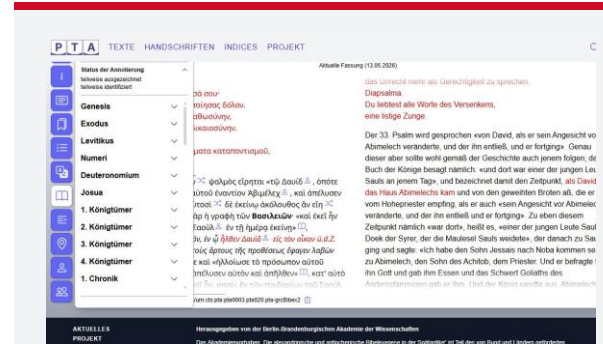


Berlin-Brandenburgische
AKADEMIE DER WISSENSCHAFTEN



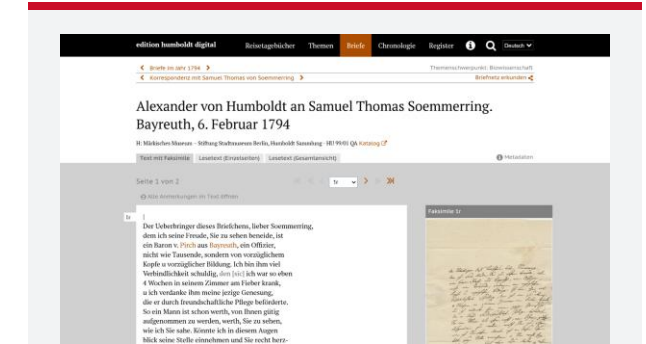
Statische Edition
**Jahresberichte für
deutsche Geschichte der
Zwischenkriegszeit**

pom.bbaw.de/JDG



Datenbankgestützte Edition
Patristic Text Archive

pta.bbaw.de



Hochvernetzte Edition
edition humboldt digital

edition-humboldt.de

technischer Pflege- und Erhaltungsaufwand

Suche im Browser

APIs und Schnittstellen

Interaktive
Visualisierungen

Dynamisch generierte
Inhalte und Ansichten

Statische Websites

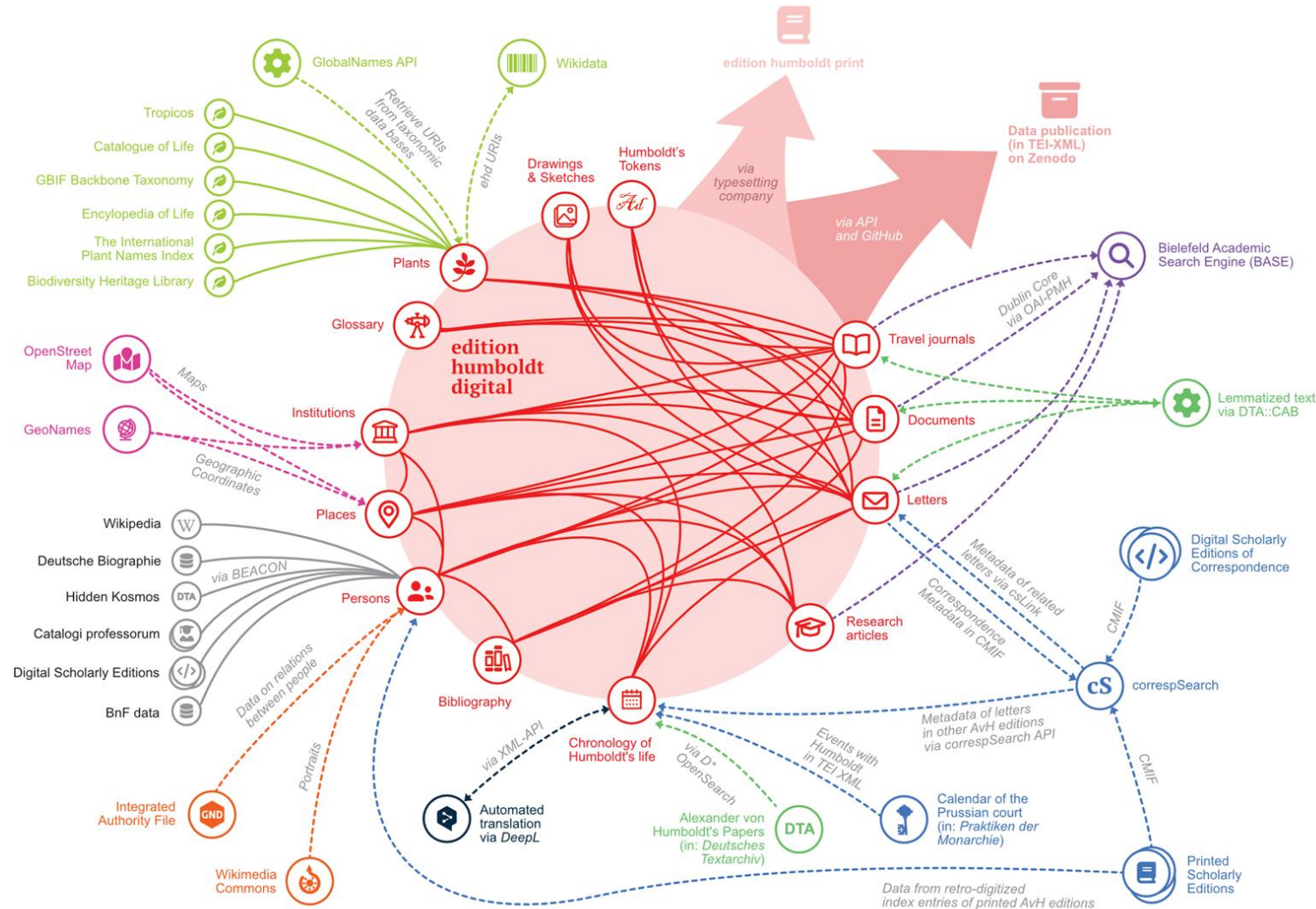
Datenbank im Hintergrund

Serverseitige Suche,
Register

Einbindung externer
Dienste

Weitere digitale Editionen der BBAW: bbaw.de/bbaw-digital/digital-humanities/digitale-editionen

Die hochvernetzte Edition



Mehr Vernetzung erweitert die Nutzungsmöglichkeiten, aber erhöht den dauerhaften Kurationsbedarf.

Erhaltungsumfang einer digitalen Ressource

**Vernetztes und
dynamisches System
erhalten**

externe APIs, Daten und
Dienste, dynamische oder KI-
generierte Ansichten

→ **kontinuierlich
überwachen und kuratieren**

Nutzbare Edition erhalten

Oberfläche, Suche, Register,
Visualisierungen, APIs

→ **betreiben, testen und
migrieren**

**Zitierfähige Fassung
erhalten**

Version, statische Seiten,
Lizenz, PID, WARC

→ **einfrieren, referenzieren
und paketieren**

**Forschungsdaten
erhalten**

Digitalisate, Texte,
Annotationen, Metadaten,
Dokumentation

→ **archivieren und validieren**

FAIR-Prinzipien: auffindbar, zugänglich, interoperabel, nachnutzbar

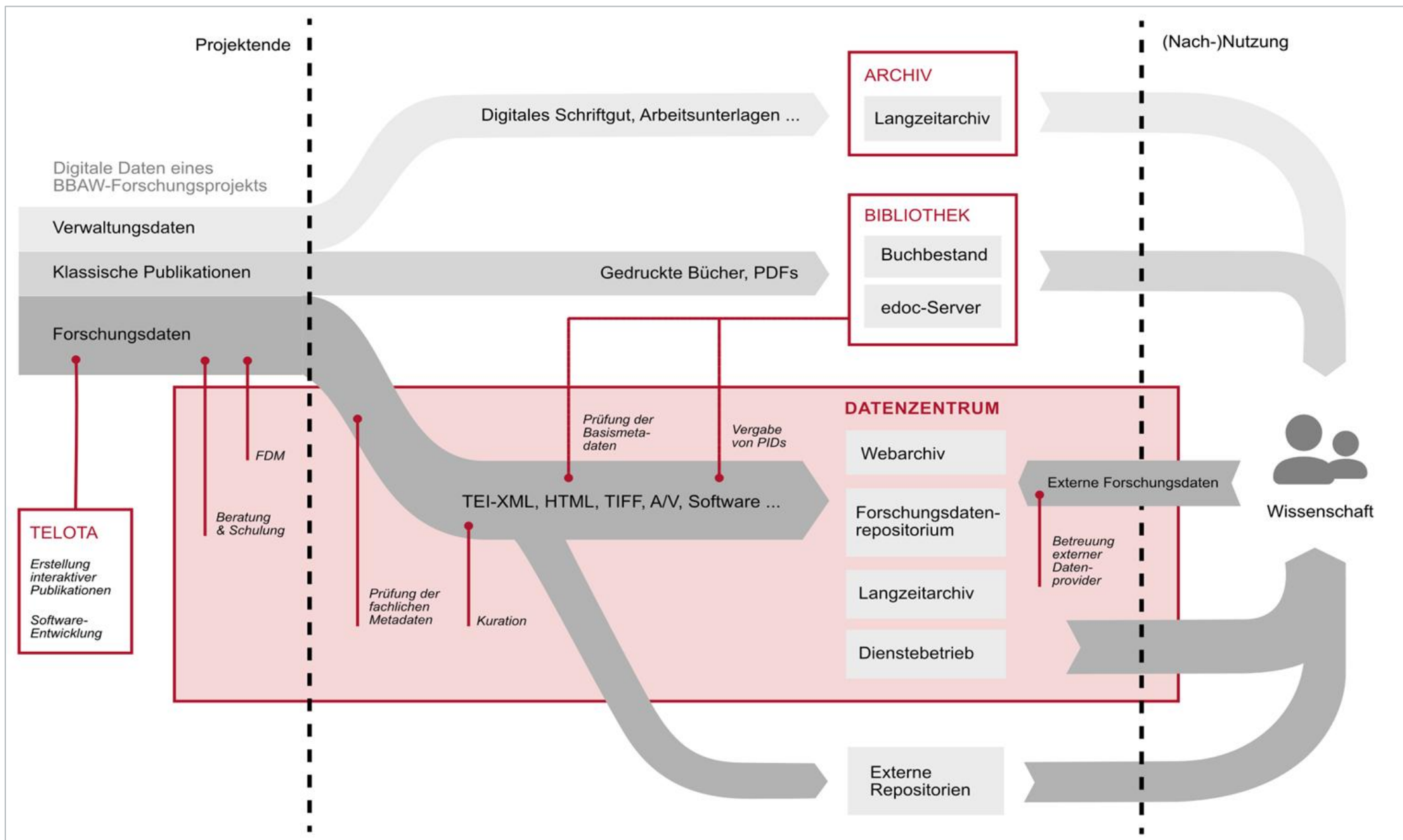
mehr Funktionen und Abhängigkeiten
mehr Pflegeaufwand, Kosten und institutionelle
Verantwortung

Vorschlag für Archivierungs- und Servicelevel

Servicelevel / Archivierungsstufe	AS 1	AS 2	AS 3	AS 4	Beschreibung SL
SL 1	passive LZA / 0.01	passive Webarchivierung / 0.01	passive Repositorienablage / 0.05	passiver Dienstbetrieb / 0.05	Bereitstellung von Infrastruktur, alle Ingest- und Kuratierungsarbeiten verbleiben beim Datengebenden
SL 2	funktionale LZA / 0.02	funktionale Webarchivierung / 0.1	funktionale Repositorienablage / 0.2	funktionaler Dienstbetrieb / 0.2	Ingestunterstützung, Funktionstests
SL 3	serviceorientierte LZA / 0.1	serviceorientierte Webarchivierung / 0.4	serviceorientierte Repositorienablage / 0.4	serviceorientierter Dienstbetrieb / 0.6	Monitoring und Reports, Harvesting, Beratung, umfassender Funktionserhalt
SL 4	aktive LZA / 0.2	aktive Webarchivierung / 0.6	aktive Repositorienablage / 1.2	aktiver Dienstbetrieb / 1.2	Entwicklungsarbeiten, Portale, eigene Metadatenformate, APIs
Beschreibung AS	Langzeitarchivierung nach OAIS	Webarchivierung mit Ablage in LZA	Forschungsdaten-Repositorium	Serverlösungen (correspSearch), Digitale Sammlungen	Beispiele: Clarin Center SL3AS3 / Bibliographische Annalen SL2AS4 / JDG SL3AS4 / Webarchiv SL1AS2 / correspSearch SL4AS4

Berlin-Brandenburgische Akademie der Wissenschaften, 2025, Initiative Datenzentrum

Archivierung, Betrieb und Kuration erzeugen unterschiedliche dauerhafte Kosten.



Arbeitsbereich des Datenzentrums zwischen Projekten, TELOTA, Archiv und Bibliothek (Initiative Datenzentrum, 2026)

Daueraufgaben

1

Sicher betreiben

Updates, Monitoring, Backup und Wiederherstellung

2

Technik erneuern

Formate, Code und Abhängigkeiten migrieren

3

Verfügbarkeit sichern

Versionen, PIDs und dauerhafte Referenzen

4

Daten und Dokumentation kuratieren

Metadaten, Normdaten, Integrität und Dokumentation pflegen



Technisch: Software alert

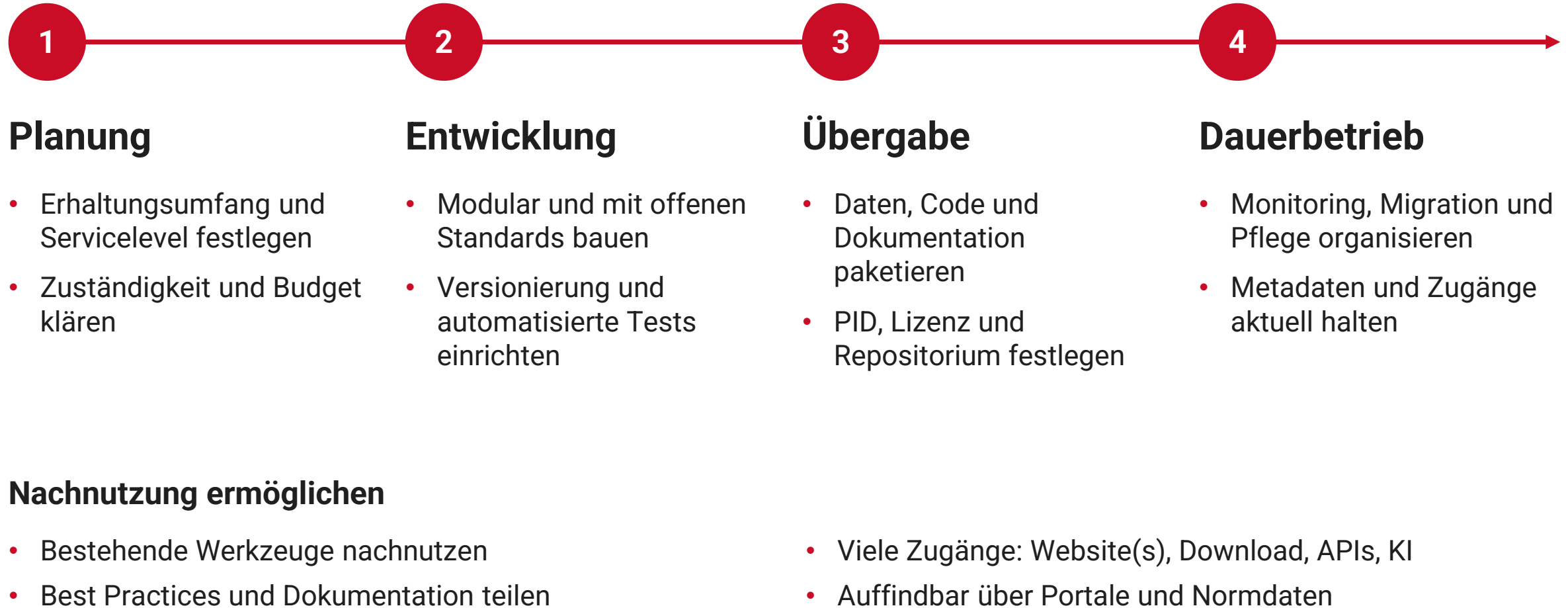
- Frameworks, Bibliotheken und Betriebssysteme veralten
- Sicherheitsupdates erzwingen Migrationen
- Externe Dienste und Normdaten ändern ihre Schnittstellen
- Suche, Register und Visualisierungen lassen sich nicht einfach archivieren
- Individuelle Lösungen ohne Dokumentation und Tests



Institutionell: Projekte enden

- Projekte enden, Ressourcen sollen dauerhaft verfügbar bleiben
- Personal wechselt, implizites Wissen geht verloren
- Zuständigkeit nach Projektende ist oft ungeklärt
- Betrieb und Pflege sind nicht (dauerhaft) finanziert
- Rechte und Lizenzen sind nicht immer geklärt

Nachhaltigkeit im Lebenszyklus einer Ressource



Seit 2025 bei TELOTA: Infrastruktur, Beratung und Methoden für die Forschungsvorhaben der BBAW

Läuft

Beratung und KI-Werkstätten

Schulungen für die Vorhaben der BBAW

KI-Richtlinie der BBAW

in Kraft seit Juni 2025

Texterkennung (OCR/HTR)

eigene eScriptorium-Instanz in
Kooperation mit dem ZIB

Prototyp Leibniz-Datierung

KI-gestützte Datierung von Manuskripten,
mit dem HPI

In Vorbereitung

Editionsunterstützung

KI-Werkzeuge für das digitale Edieren

KI-gestützte Kuration

Monitoring, Migration, Qualitätskontrolle

Neue Zugänge

Editionen für KI-Werkzeuge erschließen,
z. B. über Chatbot und MCP-Server

Dokumentation für KI

so dokumentieren, dass Menschen und
KI-Agenten damit arbeiten können

Nächster Schritt

Kooperationen ausweiten

KI-Dienste der GWDG nutzen

Eigene Modelle trainieren

Texterkennung und Sprachmodelle

Qualitätskriterien entwickeln

Prüfmaßstäbe für KI-Ergebnisse

Wissen & Methoden teilen

Workflows offen publizieren

KI verändert wie digitale Ressourcen entstehen und erhalten werden.



Migration und Modernisierung

KI analysiert gewachsenen Code, überführt Daten in Standardformate und erzeugt neue Oberflächen.



Automatisierte Qualitätskontrolle

KI kann Daten, Links und Oberflächen laufend prüfen und Auffälligkeiten melden.



Unterstützung bei Fehlerdiagnose und Fehlerbehebung

Agenten finden Ursachen, aktualisieren Abhängigkeiten, migrieren Code und schreiben Tests.



Unterstützung bei kontinuierlichem Monitoring

Monitoring erkennt Probleme früh, um unbemerktes Veralten zu verhindern.

Voraussetzung: ausführliche Dokumentation (Spezifikation und Screenshots), Tests und versionierte Daten.

Die Dokumentation wird zum Prompt für KI-Agenten für die Kuration. → Context Engineering

Erprobung im KI-Lab von TELOTA und im Rahmen der Text+ Task Area Editions.

Dokumentation als Bauplan für KI

Dokumentation

so genau spezifiziert, dass sie als Prompt dient

§ 1 Datenmodell und TEI-Schema

§ 2 Designentscheidungen

§ 3 Architektur und Schnittstellen

§ 4 Ansichten und Layout

§ 5 Suche, Register, Funktionen



Prompt /
Kontext



Generative KI

interpretiert die Spezifikation



Rekonstruierte Edition



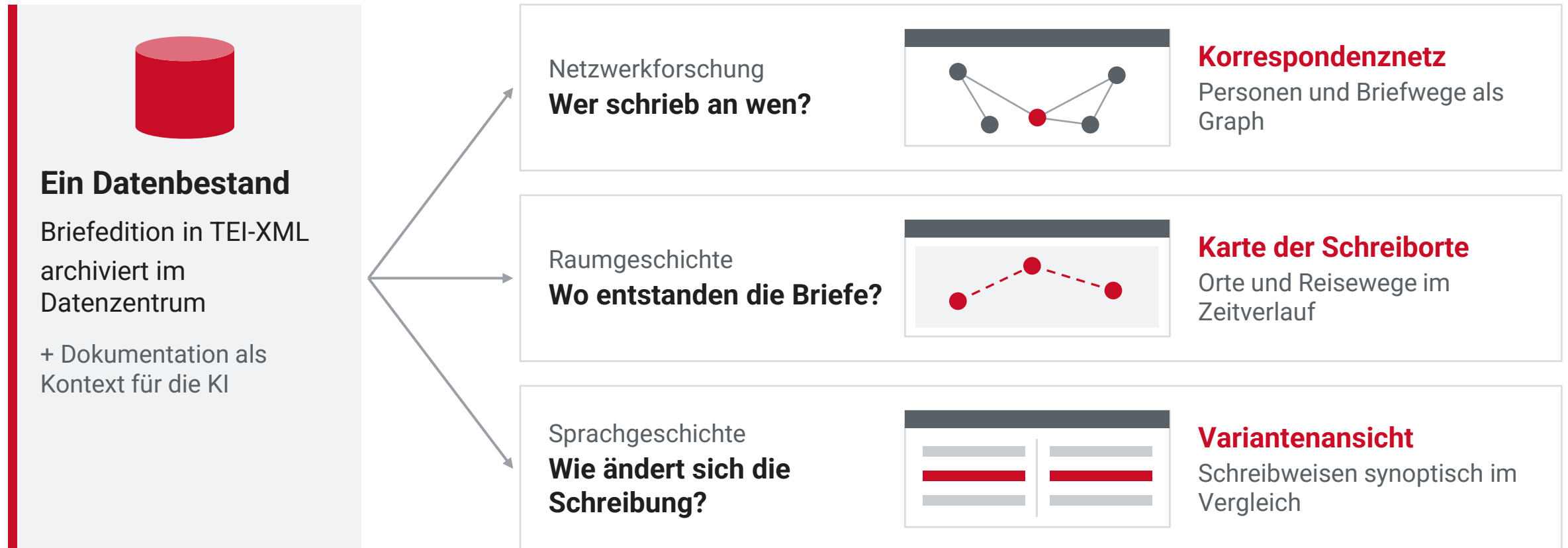
Ergebnis

Zustand, Ansicht und Funktion annähernd
wiederhergestellt

Keine echte Reproduzierbarkeit: Jeder Lauf ist eine neue Interpretation, Halluzinationen bleiben möglich.

Dieselben Daten, viele Ansichten

Frontend on demand: KI erzeugt die Ansicht beim Aufruf, zugeschnitten auf die jeweilige Frage



Ökologisch fragwürdig: Jeder Aufruf erzeugt die Ansicht neu und verbraucht erneut Rechenzeit und Energie.

Schwer zitierbar: Die Ansicht existiert nur im Moment. Ohne archivierte Fassung gibt es keinen stabilen Beleg.



Neue Aufwände

- Aufwand verlagert sich zu Monitoring, Evaluation und Orchestrierung.
- KI kann Gesamtkosten erhöhen, auch wenn einzelne Aufgaben billiger werden.
- Lizenzen, Rechenzeit und API-Nutzung verursachen laufende Kosten.
- Jeder Modellwechsel erfordert neue Tests und Anpassungen.



Menschen bleiben verantwortlich

- Nachlassbewusstsein
- Fachliche Qualitätskontrolle
- Prompts, Modellversionen und KI-Anteile werden dokumentiert
- juristische Verantwortung

Qualität und Reproduzierbarkeit

- Welche Modell- und Werkzeugversion wurde verwendet?
 - Welcher Prompt bzw. Kontext wurde übergeben?
 - Welcher Output wurde übernommen oder verworfen?
 - Ist der Output deterministisch oder probabilistisch?
-

Verantwortung

- Wer verantwortet das wissenschaftliche Ergebnis?
- Wie kennzeichnen wir KI-generierte Inhalte in der Ressource?

Zitierbarkeit

- Wie wird eine dynamisch generierte Ansicht dauerhaft zitiert?
 - Sind KI-Ergebnisse nachvollziehbar und wissenschaftlich nutzbar?
-

Rechte und Datenschutz

- Dürfen Daten und Quellen an externe KI-Dienste gehen?
 - Wem gehören KI-generierte Inhalte?
-

Abhängigkeit von Anbietern

- Was passiert, wenn ein Modell geändert oder abgeschaltet wird?
- Wie können wir offene Modelle selbst betreiben?

1

Gut modellierte und verlässliche Daten

Daten, Metadaten, Dokumentation und dauerhaft zitierfähige Versionen

2

Ein Bauplan in Form von dokumentierten Entscheidungen

Datenmodelle, Schnittstellen, Funktionsbeschreibungen, Spezifikationen und Tests

3

Eine Institution und klare Verantwortung

Infrastruktur, Zuständigkeit, fachliche Kontrolle, dauerhafte Finanzierung

Demnächst mit KI?

Ja, aber nur auf nachhaltigen Fundamenten.
KI kann Kuration und Rekonstruktion unterstützen.
Nachhaltigkeit bleibt eine institutionelle Aufgabe.

Vielen Dank!

Alexander Czmieł

czmiel@bbaw.de

TELOTA

www.telota.de

Datenzentrum

www.bbaw.de/forschung/datenzentrum

KI-Lab

www.bbaw.de/bbaw-digital/ki