

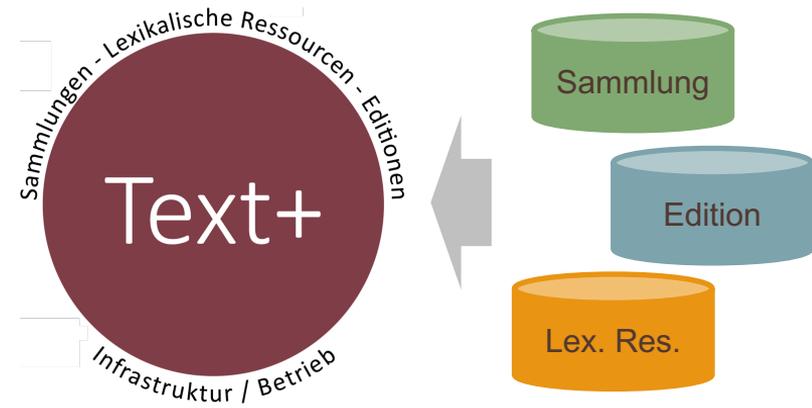
## Data Depositing Services und der Text+ Datenraum

Andreas Witt, Leibniz-Institut für Deutsche Sprache  
Andreas Henrich, Otto-Friedrich-Universität Bamberg  
Jonathan Blumtritt, Cologne Center for eHumanities (CCeH)  
Christoph Draxler, Bayerische Archiv für Sprachsignale  
Axel Herold, Berlin-Brandenburgische Akademie der Wissenschaften  
Marius Hug, Berlin-Brandenburgische Akademie der Wissenschaften  
Christoph Kudella, SUB Göttingen  
Peter Leinen, Deutsche Nationalbibliothek  
Philipp Wieder, Gesellschaft für wiss. Datenverarbeitung mbH Göttingen

[text-plus.org](http://text-plus.org)  
[office@text-plus.org](mailto:office@text-plus.org)



# Data Depositing in Text+: Was habe ich davon?



- » Nachhaltige Datenhaltung (FAIR/CARE)
- » Auffindbarkeit über Kataloge (z. B. VLO, Editions katalog)
- » Nachnutzbarkeit (z. B. LRS, Geo-Browser)

# FAIR Prinzipien

## » **Auffindbarkeit:**

- » Daten und Metadaten sollten sowohl von Menschen als auch von Maschinen leicht zu finden sein.



## » **Zugänglichkeit:**

- » Daten und Metadaten sollten verfügbar gemacht und langzeitarchiviert werden, sodass sie leicht von Menschen und Maschinen heruntergeladen und genutzt werden können.



## » **Interoperabilität:**

- » Die Daten sollten derart vorliegen, dass sie mit anderen Datensätzen von Menschen und Maschinen verknüpft werden können.



## » **Wiederverwendbarkeit:**

- » Zur Wiederverwendbarkeit trägt eine Beschreibung der Datensätze über Metadaten bei, sodass sie für weitere Forschungen nachnutzbar und mit anderen Datensätzen vergleichbar sind.



# Data Depositing bei und mit Datenzentren



siehe auch: <https://www.text-plus.org/forschungsdaten/daten-und-kompetenzzentren/>

# Daten- und Kompetenzzentren

# Was sind Daten- und Kompetenzzentren?

## » Datenzentren :=

- » Partneereinrichtungen mit Spezialisierung auf bestimmte Arten von Daten, die sie vorhalten
- » Betreiben eine Infrastruktur, die eine langfristige Bereitstellung und Archivierung von Daten ermöglicht
- » Nutzen zur Datenhaltung zertifizierte Repositorien
- » Stellen die Metadaten zu den Daten über Schnittstellen bereit
- » Stellen Schnittstellen zu weiteren Diensten von Text+ zur Verfügung (z. B. zur verteilten Suche)
- » Nehmen Daten gemäß der Spezialisierung auch von Dritten entgegen (mehr dazu später)

## » Kompetenzzentren :=

- » Partneereinrichtungen mit speziellen Kenntnissen und Fähigkeiten für bestimmte Arten von Daten
- » Beraten zur Erstellung und Archivierung von bestimmten Daten
- » Können auch an der Entwicklung von Diensten in Text+ beteiligt sein, die auf Schnittstellen aufbauen
- » Unterstützen bei der Archivierung an anderen Orten, benötigen keine eigene Archivinfrastruktur

# Welche Zentren gibt es?

- » für jede Datendomäne mit unterschiedlichen Schwerpunkten
  - » Von „allgemein“ bis „sehr speziell“
  - » Born Digital bis retrodigitalisiert
- » Unterschiedliche Sprachen, Epochen, Datenformate



# Daten- und Kompetenzzentren

- » Zertifizierung
- » Metadaten/ Metadaten-Harvesting zum Aufbau von Katalogen
- » APIs/Interfaces

# Zertifizierung der Datenzentren

- » Core Trust Seal (CTS)
  - » Nachhaltige Infrastruktur und Prozesse
  - » International etabliert
  - » Bewertung aufgrund von einer Kriterienliste
  - » Begutachtung durch ‚Qualified volunteer Reviewers‘ und Zertifizierung durch das CoreTrustSeal Board
  
- » Nestor e.V.
  - » Infrastruktur und Prozesse zur Langzeitarchivierung
  - » Basiert auf DIN 31644 „Kriterien für vertrauenswürdige digitale Langzeitarchive“
  - » Bewertung aufgrund von auf der DIN 31644 basierenden Kriterienliste
  - » Begutachtung durch Mitglieder der nestor-Arbeitsgruppe ‚nestor AG Zertifizierung‘

# Metadaten

- » Verschiedenste Konventionen und Normen
  - » Dublin Core/Dublin Core 15
  - » Lightweight Information Describing Objects (LIDO)
  - » Marc21
  - » ISO 24622-1 und ISO 24622-2 (CMDI)
  - » TEI-Header
  - » ...
- » Unterschiedliche Serialisierungen
  - » XML
  - » JSON/JSON-LD
  - » RDF (N-Tuples, Turtle, XML-RDF, JSON-LD)
- » „Öffentliche“ Information über Forschungsdaten
  - » Für Kataloge und Nachweissysteme
  - » Bereitgestellt über Schnittstellen
  - » Durchsuchbarkeit über Suchmaschinen
  - » Ermöglicht Zitation/Persistente Identifikation

# APIs/Interfaces

- » Auslieferung von Metadaten:
  - » OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting)
    - » REST-Interface-Spezifizierung
    - » Erlaubt neben Dublin Core auch andere Metadatenschemata
  - » SPARQL (**S**PARQL **P**rotocol **A**nd **R**DF **Q**uery **L**anguage)
    - » Insbesondere für Linked Data Anwendungen
    - » Erfordert Linked Data Darstellung der Metadaten im Backend
- » Zugriff auf Forschungsdaten in einer ortsverteilten Infrastruktur
  - » Föderierte Inhaltssuche (FCS)
  - » Basiert auf Webstandards
- » Services aus Repositorien: Zugriff auf Erschließungswerkzeuge

# Spezialisierungen der Datenzentren

## » Sammlungen

- » IDS
- » BBAW
- » DNB
- » SUB
- » Hamburg (Uni und Akademie)
- » LMU München
- » Uni des Saarlandes
- » Uni Duisburg-Essen
- » Uni Freiburg (K)
- » Uni Köln
- » Uni Tübingen
- » Uni Würzburg (K)

## » Lexikalische Ressourcen

- » IDS
- » BBAW
- » Sächsische Akademie
- » Uni Köln
- » Uni Trier
- » Uni Tübingen

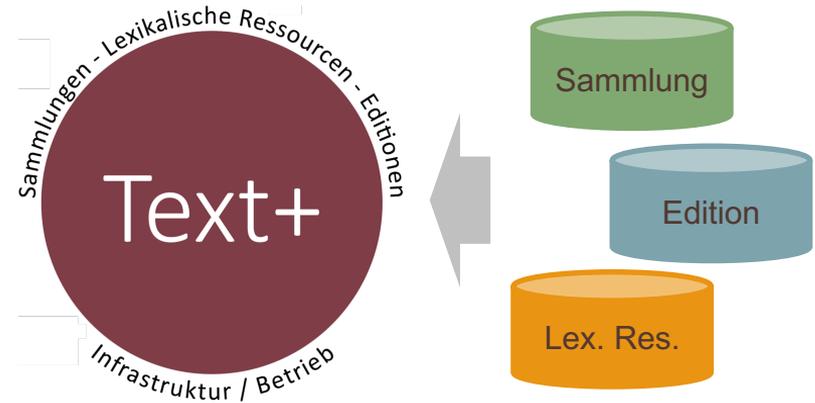
## » Editionen

- » BBAW
- » NRW Akademie
- » SUB
- » Akademie Mainz
- » Darmstadt (TU, HS, UB)
- » HAB
- » Leopoldina
- » Steinheim Institut

# Wege ins Text+ Universum

# Nutzung eines spezifischen Zentrums

- » Daten an ein Datenzentrum geben
- » Formate, Anforderungen absprechen
- » Ggf. Unterstützung bei der Anpassung, Aufbereitung erhalten
- » jährliche Ausschreibung zur Förderung von Kooperations-Projekten



# Selbst Text+ Zentrum werden

- » Die Daten nicht an Text+ „übergeben“
- » Daten selbst hosten und verwalten
- » Selbst ein Datenzentrum in Text+ betreiben
- » Unterstützung und Begleitung durch Text+

# Nutzung der „generischen“ Optionen

- » Text+ bietet das DARIAH Repository als "catch-all" Repository für Dateneigentümer, die die clusterspezifischen Kriterien nicht erfüllen
- » Text+ bietet in Zukunft einen Text+ Archiv-Service basierend auf dem CTS- und nestor-zertifizierten Langzeitarchiv der DNB, welches von der GWDG entwickelt und gepflegt wird

# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » Lexikalische Ressourcen: Axel Herold, BBAW
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » Lexikalische Ressourcen: Axel Herold, BBAW
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Collections: unstructured text

- » Digitale Daten ohne Strukturinformationen (Tiefenerschließung)
  - » OCR-te Texte
  - » Born Digital (pdf, e-pub)
- » Erschließung für die Forschung erfordert Vorarbeiten
  - » OCR
  - » Nachbearbeitung
    - » Boilerplate-Verarbeitung
    - » Qualitätssicherung
    - » Normalisierung
  - » Automatisierte Annotation zur weiteren Erschließung
    - » Seiten und Abschnittseinteilungen
    - » Tokenisierung
    - » Named Entity Recognition
    - » Verschlagwortung/Metadatengenerierung
    - » ...

# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » Lexikalische Ressourcen: Axel Herold, BBAW
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Collections: Historische Korpora

- » DTA – Archiv für historische, v. a. deutschsprachige Texte und Korpora am Zentrum Sprache der BBAW
- » (Maschinenlesbare) Korpusbeschreibungen, u. a. zur Integration der Ressourcen des Datenzentrums in die Text+-Infrastruktur
- » Kuration verschiedener neuer Korpora und Integration in die DTA-/DWDS-/ZDL-Infrastruktur

# Collections: Historische Korpora

## » Projektkontext:

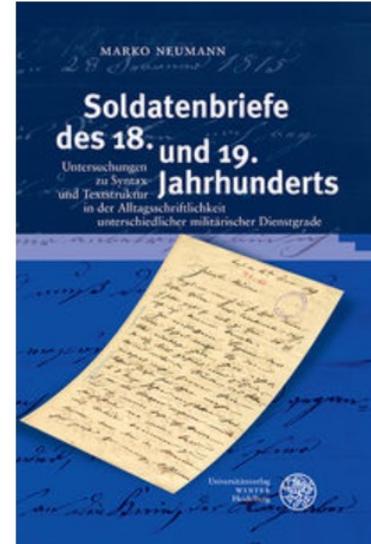
- Marko Neumann:  
Soldatenbriefe des 18. und 19. Jahrhunderts
- als Dissertation publiziert
- Anhang: Forschungsdaten (PDF-Download)

## Text+ User Story:

Soldiers' letters of the 18th and 19th centuries:

From the PDF edition to reusable, interoperable research data.

<https://www.text-plus.org/en/research-data/user-story-508/>



# Collections: Historische Korpora

## » Text+-Korpus: Soldatenbriefe (1745–1872)

- Erstellung einer Korpusbeschreibung (nach Schema)
- Anreicherung der Metadaten
- Transformation nach XML/TEI P5
- Veröffentlichung von (Meta-)Daten und Dokumentation auf github unter CC BY-SA 4.0

Soldatenbriefe des 18. und 19. Jahrhunderts

1745–1872

[Link](#) Landing-Page [Link](#) Suche

Sprache deu Format DTABF Format TCF

Faksimiles nein Transkription manually

Dokumente 170 Sätze 3319

Tokens 96023

Genre Gebrauchsliteratur::Brief

Lizenz CC BY-SA 4.0

Das Korpus „Soldatenbriefe“ umfasst 170 Briefe von Offizieren, Unteroffizieren und einfachen Soldaten, adressiert an die Familien in der Heimat. Die Briefe wurden im Zeitraum von 1745 bis 1872 auf Deutsch verfasst; ein deutlicher Schwerpunkt liegt auf Briefen aus den Koalitions- und Befreiungskriegen (1792–1815), dem Deutschen Krieg (1866) und dem Deutsch-Französischen Krieg (1870/71).

[Datensatz](#)

# Collections: Historische Korpora

## » Text+-Korpus: Soldatenbriefe (1745–1872)

- Linguistische Aufbereitung als Voraussetzung der Korpus-Suche [hier: Getränk | germanet]
- Korpusanalyse mittels DiaCollo [hier: Mutter]
- Korpusübergreifende Suche durch Integration in DWDS-Metakorpus „Historische Korpora“

n aus gehungerten Seelen mit einer Flase	<b>Cognac</b>	beschenkte.
daß sich sogar die Trinkfesten förmlich in	<b>Rothwein</b>	gebadet haben sollen; während wir.
während wir Ärzte seit ca 8 Tagen keinen	<b>Wein</b>	gesehn, sondern zu unserer guten al
er einfachen Krankenkost nur schlechten	<b>Schnaps</b>	getrunken haben und unsere Krank
nd unsere Kranken zwei Tage anstatt des	<b>Rothweins</b>	nur leichten Mosel erhielten .
In und ganz Schwämme nicht gesalzen das	<b>Wasser</b>	ganz sonders Bier gibt es Einzellnes
e nicht gesalzen das Wasser ganz sonders	<b>Bier</b>	gibt es Einzellnes da kostet die Maß
s Einzellnes da kostet die Maß 24 Kreuzer	<b>Wein</b>	gibt es genug den Schoppen zu 6 Kr
use war und hätte wenn ich ein par Eimer	<b>Wasser</b>	gehabt hätte das Feuer auslöschen l
Strümpfe Unterjacken Pfeifen Conjac und	<b>Schnaps</b>	, ich habe schon 2 paar Strümpfe un
sehr incomodirt, außerdem habe ich mir	<b>Chocolate</b>	zu wider gegessen + habe noch Vorr
Marceau) wie bei uns der Bankplatz, ein	<b>Caffe</b>	neben den andern + das eine feiner



# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » Lexikalische Ressourcen: Axel Herold, BBAW
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Gesprochene Sprache

- » Bayerisches Archiv für Sprachsignale
  - » Sprachtechnologie und Phonetik, Webdienste
- » Hamburger Akademie der Wissenschaften & Uni Hamburg
  - » Ethnologische und mehrsprachige Korpora, Gebärdensprache
- » Leibniz Institut für Deutsche Sprache
  - » Gesprächs- und Varietäten-Korpora
- » Universität des Saarlandes
  - » Stimmpathologie, slawische Sprache in Deutschland

# Repositories gesprochener Sprache

- » Metadaten in CMDI Format
  - » Zentrenübergreifende Meta- und Inhaltsdatensuche
  - » OAI-PMH Schnittstelle
  - » Core Trust Seal bzw. CLARIN zertifiziert
- » Download von Korpora nach einmaliger Authentifizierung
- » Unterstützung des Imports neuer Korpora
  - » Fachkundige Beratung während oder nach der Korpuserstellung
  - » Teilautomatisierung der Metadatengenerierung, z.B. via COALA Webdienst



# Bsp: BAS Repository

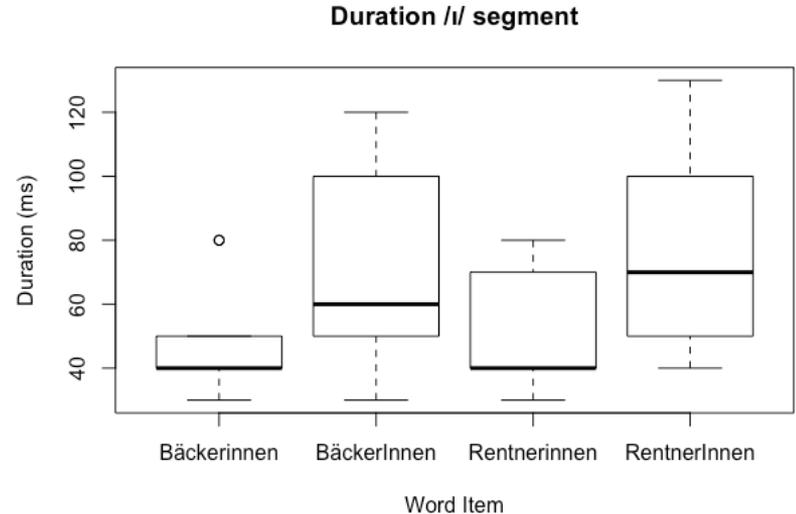
The screenshot shows the BAS CLARIN Repository homepage. At the top, there are logos for the Bavarian Archive for Speech Signals, CLARIN B CENTRE, CORE TRUST SEAL, and a 'CLICK ME FOR HELP' button. The main heading is 'BAS CLARIN Repository'. Below it, a paragraph explains that the repository contains all BAS resources accessible for academic users. A pink box contains a message: 'You are not yet authorized to have access to the BAS repository. Click here to login either via your academic institution or via your CLARIN IDP account. If you are not an academic, or if your academic institution is not part of the DFN-AAI, you can register here to get a CLARIN IDP account. Please read our privacy policies for AAI authentication.' A 'Log in' button is visible on the right. A left sidebar menu includes 'Repository', 'FAQ', 'Search', 'BASStat', 'About/Privacy', and 'Links'. The main content area shows 'Metadata' with fields for PID (11022/1009-0000-0001-231F-6) and CMDI, 'Collections' with 'OHD', and 'Corpora' with 'AbsolventInnen'.

This screenshot shows the same BAS CLARIN Repository homepage, but with a green box indicating full access: 'You are authorized to have full access to the BAS repository. Restrictions may still apply in case the resource is not freely accessible for academic users. For secure logout please follow this link and close your browser afterwards. Logout'. The 'Log in' button is now greyed out. The 'Corpora' section is expanded, showing a list of corpora with details for 'ASD' (Institut für deutsche Sprache) and 'ASD' (Audiolias Siebenbürgisch-Sächsischer Dialekt).

- 50+ Korpora
- technisch validiert
- Korrekturen, Ergänzungen und Versionskontrolle
- verschiedene Lizenzen

# Wie spricht man „BäckerInnen“?

- » Studentisches Projekt: ca. 600 Aufnahmen gelesener Sätze, automatisch segmentiert
- » In 29% der Fälle vermeiden die SprecherInnen die genderneutrale Form, in den restlichen sprechen sie das finale // deutlich länger bzw. mit Pause
- » „Sprachwandel ausgehend von Schriftform“
- » Als BAS Ressource frei verfügbar seit 2017 11022/1009-0000-0003-FF39-F



# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » **Lexikalische Ressourcen: Axel Herold, BBAW**
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Lexical Resources

- » „Wissen über Wörter“ aus sehr unterschiedlichen (linguistischen, lexikografischen, ...) Perspektiven  
→ heterogene Daten- und Ressourcentypen
- » Veröffentlichte vs. „graue“ Drucke/Manuskripte vs. (genuin) **digitale Ressourcen**  
→ heterogene technische Repräsentation
- » Mittelfristiges **Ziel**: konzeptionelle und technische **Harmonisierung** (Kuration und/oder Mapping)

# Lexical Resources: Inhaltliche Randbedingungen

» Sechs Institutionen, ähnliche Anforderungen:

- 1) „kompatible“ Lizenzierung
- 2) Umfangreiche, aussagekräftige Metadaten
- 3) Sammelgebiete (zentrenspezifisch, inhaltlich weit gefasst):
  - German Dictionaries in a European Context
  - Born-Digital Lexical Resources
  - Non-Latin Scripts

# Lexical Resources: Technische Randbedingungen

» Sechs Institutionen, momentan ähnliche Strategie:

- 1) TEI (for dictionaries), TEI Lex-0
- 2) Andere (eventuell proprietäre) XML-basierte Formate
- 3) Andere standardisierte Serialisierungsformate
- 4) Keine Bitstream-Preservation (z. B. keine Datenbank-Dumps, keine proprietären Binärformate)

→ Daten nicht nur mit proprietären Programmen verarbeitbar

# Lexical Resources: Ausblick

- » Für infrastrukturelle Einbindung (z. B. FCS, LOD): Anforderungen mittelfristig schärfen
- » Für reine Archivierung: möglichst agnostisch bleiben
- » Kurzfristig: erste Erfahrungen durch Flex-Funds-Projekte

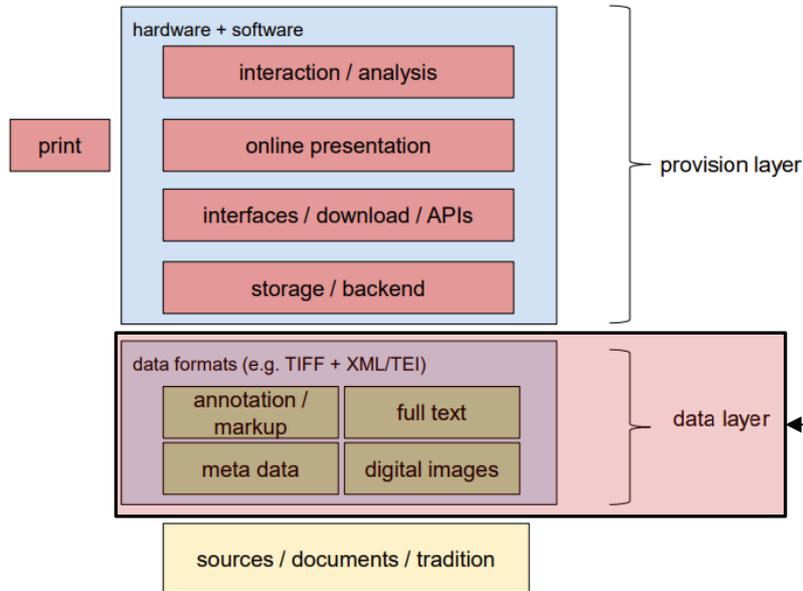
# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » Lexikalische Ressourcen: Axel Herold, BBAW
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Editions – Allgemeine Herausforderungen

- » **Unterschiedliche Editionstypen** mit jeweils eigenen Konventionen und Anforderungen
- » Zusätzlich **projektspezifische Anforderungen** (i.d.R. abgeleitet aus Forschungsfragen)
- » **Dynamisches Feld** („moving target“): Ständige Veränderung der technischen Möglichkeiten, bei gleichzeitigem Wandel der editorischen Methodiken/Anforderungen

# Editions – Schichtenmodell



## Datenschicht

Das Markup ist mit den **TEI** Guidelines durch eine internationale Community weitgehend standardisiert, **ABER**

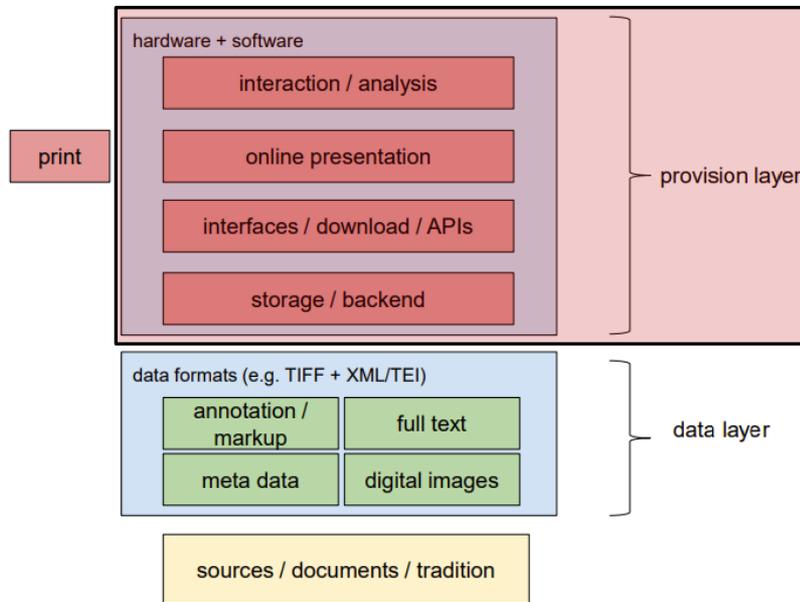
- *TEI ALL lässt sehr **viel Spielraum** zu, bietet z. T. mehrere Auszeichnungsmöglichkeiten für das selbe Phänomen oder semantisch ähnliche Phänomene*
- *für jedes Editionsprojekt muss aufgrund der projektspezifischen Anforderungen i. d. R. ein **eigenes TEI-Anwendungsprofil** erstellt werden*

**Heterogene und zunächst nicht-interoperable Forschungsdaten, Aber: Normdaten als**

- *Interoperabilitätsebene/Vernetzung*
- *Grundlage für übergreifende Dienste*

Originalabbildung: [https://www.i-d-e.de/wp-content/uploads/2019/11/08\\_Architekturen.pdf](https://www.i-d-e.de/wp-content/uploads/2019/11/08_Architekturen.pdf)

# Editions – Schichtenmodell



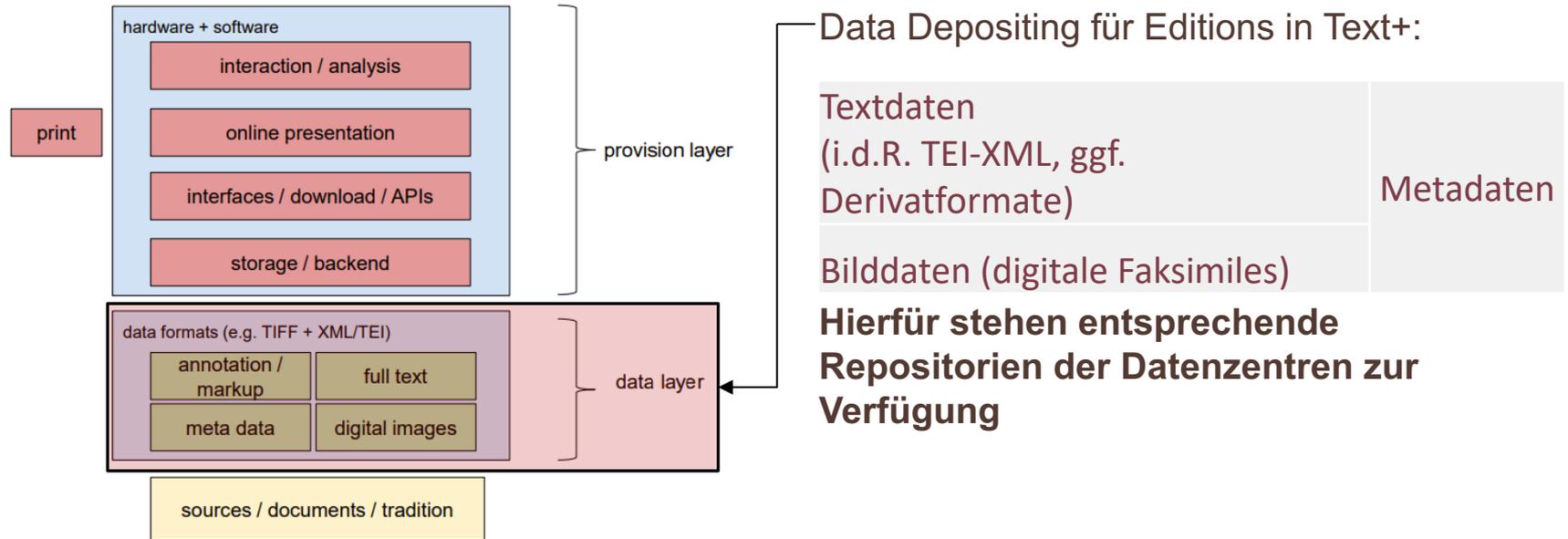
## Bereitstellungsschicht

= Forschungssoftware mit sehr starken Abhängigkeiten von

- **Technologien:** für den dauerhaften Betrieb muss dem technologischen Wandel kontinuierlich gefolgt werden („software rot“)
- **Personen:** Wissen um die technischen Funktionalitäten und Systemarchitektur liegt bei einzelnen Personen
- Zusätzlich: Häufig **dynamische Ressourcen**, die mit gängigen Webarchivierungsmechanismen nicht archivierbar sind

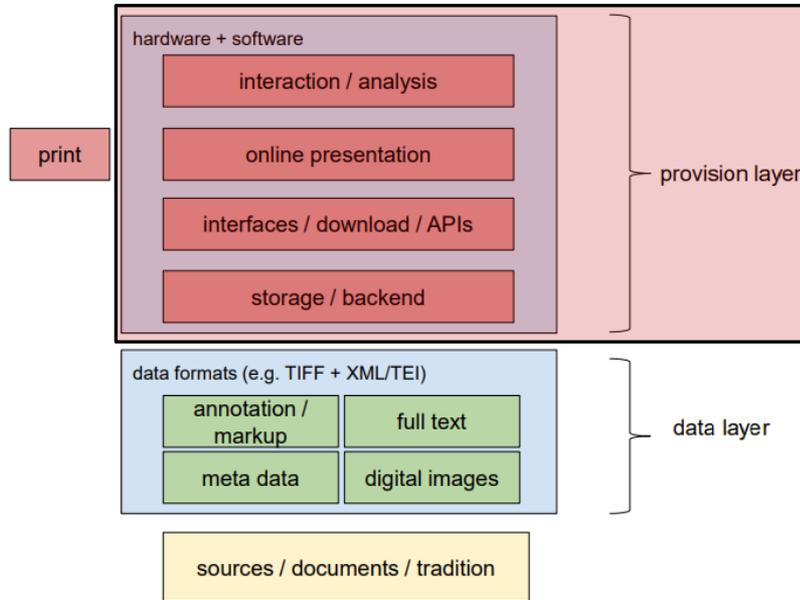
Originalabbildung: [https://www.i-d-e.de/wp-content/uploads/2019/11/08\\_Architekturen.pdf](https://www.i-d-e.de/wp-content/uploads/2019/11/08_Architekturen.pdf)

# Editions – Data Depositing



Originalabbildung: [https://www.i-d-e.de/wp-content/uploads/2019/11/08\\_Architekturen.pdf](https://www.i-d-e.de/wp-content/uploads/2019/11/08_Architekturen.pdf)

# Editions – Bereitstellungsschicht



## Bereitstellungsschicht

Sehr heterogene Software-Stacks, dabei zusätzlich Unterschiede in den eingesetzten:

- Datenbanken
- Indizes
- Image Servern
- Webservern
- Frontend-Webapplikationen
- Betriebsmethoden (z.B. VM vs. Container)

Derzeit keine Möglichkeit der Betriebsübernahme durch Text+, ABER:

- **Beratungsangebote**
- **Handreichungen zu Best Practices**
- **Monitoring und Zentrales Nachweissystem** (siehe „Registry“)
- **„Software Services“**
- **Standardisierungsaktivitäten**

Originalabbildung: [https://www.i-d-e.de/wp-content/uploads/2019/11/08\\_Architekturen.pdf](https://www.i-d-e.de/wp-content/uploads/2019/11/08_Architekturen.pdf)

# Beispiele & Impulse

- » Collections: unstructured text, Peter Leinen, DNB
- » Collections: historische Korpora, Marius Hug, BBAW
- » Collections: gesprochene Sprache, Christoph Draxler, BAS
- » Lexikalische Ressourcen: Axel Herold, BBAW
- » Editions: Ch. Kudella, SUB Göttingen & J. Blumtritt, CCeH
- » Bitstream Preservation / LTA: Philipp Wieder, GWDG

# Langzeitarchivierung in Text+

- » Heimat ist TA „Infrastructures/Operations“, Measure 2 “Accessibility”
- » Aufgabe: Bereitstellung eines Text+ Archive-Services
- » Deliverables
  - » IO2.2 „Text+ LZA Policy” (M12)
  - » IO2.7 „Text+ Archiv in Betrieb“ (M24)
  - » IO2.15 „Statistik zu archivierten Ressourcen im Text+ Archiv“

# Technische Basis

- » Langzeitarchiv der DNB\*
  - » OAIS-konform
  - » Zertifiziert: nestor Siegel, DSA
- » Basissoftware nennt sich koala+
  - » Entwickelt von GWDG
  - » Bitstream Preservation auf Tape/S3-konformen Speicher
  - » Simple Interface und Dashboard
  - » Eingesetzt für weitere großvolumige Projekt wie z.B. "Geschichte der Max Planck Gesellschaft"

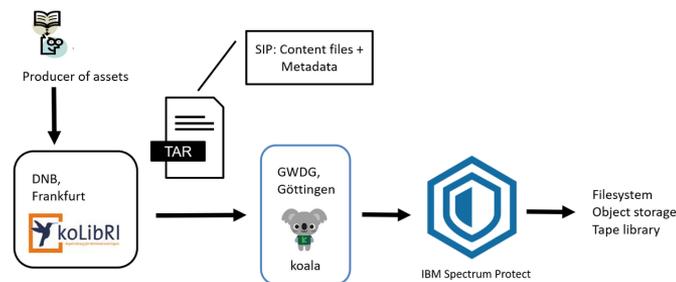


(\*)[https://www.dnb.de/DE/Professionell/Erhalten/LZA-System/lza-system\\_node.html](https://www.dnb.de/DE/Professionell/Erhalten/LZA-System/lza-system_node.html)

(+) <http://koala-docs.gwdg.de/>

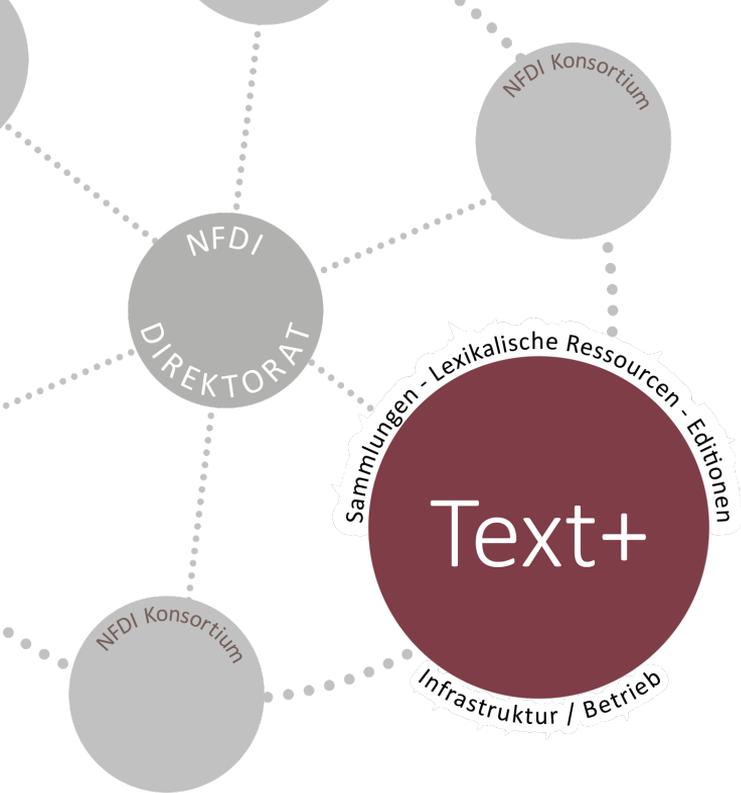
# BSP vs. LZA

- » Bitstream Preservation nur Teil des Archivierungsworkflows
  - » Teile des Prozesses liegen auf Seiten des Datenhalters
    - » Metadatenerzeugung oder –transformation
    - » Kuratierung
  - » Unterstützung der Integration in nutzerseitigen Workflow durch
    - » M5 „Software & Services“
- » Zertifizierungen (wie CoreTrustSeal) decken i.d.R. Verpflichtungen von Datenhalter & Betreiber des Archivs ab



# s. auch ...

- » AG Langzeitarchivierung der Sektion „Common Infrastructures“
  - » Konzept: <https://zenodo.org/record/6451456#.Yx7vofHP0-Q>
  - » Starke Beteiligung durch Text+
- » Falls Base4NFDI gefördert wird: Enge Abstimmung mit der AG & Einbringen der Text+ Anforderungen



## Vielen Dank für Ihre Aufmerksamkeit!

Andreas Witt, Leibniz-Institut für Deutsche Sprache  
Andreas Henrich, Otto-Friedrich-Universität Bamberg  
Jonathan Blumtritt, Cologne Center for eHumanities (CCeH)  
Christoph Draxler, Bayerische Archiv für Sprachsignale  
Axel Herold, Berlin-Brandenburgische Akademie der Wissenschaften  
Marius Hug, Berlin-Brandenburgische Akademie der Wissenschaften  
Christoph Kudella, SUB Göttingen  
Peter Leinen, Deutsche Nationalbibliothek  
Philipp Wieder, Gesellschaft für wiss. Datenverarbeitung mbH Göttingen

[text-plus.org](http://text-plus.org)   [office@text-plus.org](mailto:office@text-plus.org)

Text+ ist ein Konsortium der bundesweiten Initiative zum Aufbau einer nationalen Forschungsdateninfrastruktur (NFDI) und wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - Projektnummer 460033370

