

# Referenzierung des digitalen kulturellen (Text-)Erbes

## Digitale Quellenkritik und Modellierung von Metadaten

### Althage, Melanie

melanie.althage@hu-berlin.de  
Humboldt-Universität zu Berlin, Germany

### Dreyer, Malte

malte.dreyer@cms.hu-berlin.de  
Humboldt-Universität zu Berlin, Germany

### Guescini, Rolf

rolf.guescini@hu-berlin.de  
Humboldt-Universität zu Berlin, Germany

### Hiltmann, Torsten

torsten.hiltmann@hu-berlin.de  
Humboldt-Universität zu Berlin, Germany

### Lüdeling, Anke

anke.luedeling@hu-berlin.de  
Humboldt-Universität zu Berlin, Germany

### Odebrecht, Carolin

carolin.odebrecht@hu-berlin.de  
Humboldt-Universität zu Berlin, Germany

## Forschungsgegenstand

Historische Textdaten wie etwa Urkunden, Briefe, Tagebücher aber auch literarische Texte sind integraler Bestandteil unseres kulturellen Erbes und insofern für viele geisteswissenschaftliche Fachbereiche wie die Sprach-, Literatur- und Geschichtswissenschaften die empirische Forschungsgrundlage. Für deren Referenzierung und Erschließung wurden insbesondere in den historischen Disziplinen und Methodiken zur Einordnung und Kritik solcher Quellen geprägt (grundlegend etwa: Droysen 1868: 16-19; Bernheim 1907: 113-134), die wiederum auch für andere wie die oben genannten Fachbereiche intellektuelle Zugänge zu Forschungsressourcen ermöglichen können. Die Anpassung der Quellenkritik an die spezifischen Eigenschaften digitalisierter und genuin-digitaler Quellen ist dagegen ein interdisziplinär zu verortendes Desiderat in dem emergierenden komplementären digitalen Forschungsparadigma „Digital Humanities“ (für die Geschichtswissenschaften etwa: Föhr 2019; Margulies 2009).<sup>1</sup> Mit domänen-spezifischen Zugängen, die sich Beschreibungsansätze aus den genannten Fachbereichen zu eigen machen, können (historische) Textdaten trans- und interdisziplinär kritisch eingeordnet, gefunden, erschlossen und wiederverwendet werden (vgl. FAIR Data Wilkinson et al. 2016).

In unserem Beitrag zeigen wir, wie eine domänenspezifische Datendokumentation, -kritik und -referenzierung mit standardisierten Methoden (TEI ODD Burnard 2013, UML-Modell für Metadaten Odebrecht 2019) für das LAUDATIO-Repositorium (Guescini, Schulz und Odebrecht 2021)<sup>2</sup> und damit für die historisch arbeitenden Textwissenschaften umgesetzt werden kann. Die Dokumentation und Kritik nehmen sich dabei die historische Quellenkritik als Vorbild. Entsprechend werden in LAUDATIO sowohl äußerlich beschreibende als auch aus den Inhalten extrahierte Metadaten verwendet, um eine Einordnung und Erschließung der textuellen Quellen zu ermöglichen und zugleich neue Wege zu deren korpusimmanenter wie -übergreifender Recherche und Analyse zu bieten. Ein Schwerpunkt wird hierbei zunächst auf Personen- und Ortsbezüge gesetzt (wie zum Beispiel mit GND). Dabei leiten uns drei Fragenkomplexe an:

I) Wie können wir diese multiplen Perspektiven (Kritik, Erschließung und Dokumentation) auf eine Quelle und über mehrere Quellen hinweg in einem Modell abbilden? Wie können wir Referenzen zu Personen und Orten in verschiedensten Quellen abrufbar machen? Wie gelingt die Einordnung der Quellen in etablierte Taxonomien beziehungsweise Ontologien?

II) Wir erweitern diesen Forschungsgegenstand um die folgende Perspektive: Wie ordnen sich Transkripte und Annotationen in dieses Informationsgefüge ein? Wie lassen sich deren digitale Repräsentationen und deren (gewollte) Manipulationen so beschreiben, dass auch dieser Teil der digitalen Quellenkritik sind?

III) Wie kann aus dieser Informationsdichte eine Exploration erfolgen, die den Forschenden die so versammelten Quellen als Forschungsdaten auffindbar, zugänglich, interoperabel und wiederverwendbar macht?

Um sich der Beantwortung dieser Fragen anzunähern und um im Rahmen einer nachhaltigen Infrastruktur als Weiterentwicklung des bestehenden Forschungsdatenrepositoriums LAUDATIO entsprechende Lösungsansätze umzusetzen, haben wir in einem ersten Schritt in einem interdisziplinären Team aus zentraler Infrastruktureinheit, Korpuslinguistik, Forschungsdatenmanagement und Geschichtswissenschaften 56 User Stories entwickelt, die wir in der Posterpräsentation vorstellen und mit der Community diskutieren.

## Anwendungsperspektive

Diese User Stories und deren Akzeptanzkriterien stellen für uns die wesentliche Implementierungsgrundlage der Weiterentwicklung von LAUDATIO über die bisherige sprachwissenschaftliche Zielgruppe hinaus dar. LAUDATIO will explizit ein Angebot an Geisteswissenschaftler:innen verschiedenster Disziplinen zur nachhaltigen Publikation und Nachnutzung textbasierter Forschungsdaten machen. Eine breite Diskussion und ggf. Anpassung/Erweiterung ist daher essentiell. Nur so können wir LAUDATIO als Publikationsplattform in der Nutzerdomäne erweitern und einen wichtigen Beitrag für Open Science leisten sowie das Forschungsdatenmanagement unseres kulturellen Texterbes unterstützen.

## User Stories für die Erschließung von Textdaten

Die User Stories repräsentieren die Bedarfe von historisch Forschenden, die das Repositorium und dessen Filter-/Suchinterfaces

zur Forschungsdatenpublikation und -recherche verwenden. Zwei entscheidende Befunde, die sich aus der Entwicklung der User Stories ergeben haben, betreffen die Umsetzung der FAIR-Prinzipien sowie die Berücksichtigung der Forschungs- und Arbeitsprozesse insbesondere im Bereich des Datenmanagements der Zielgruppe(n). Nachfolgend zeigen wir ein Beispiel für eine User Story mit Akzeptanzkriterien, die sich aus den Forschungsprozessen und Workflows der nutzenden Community ableiten:

User Story: *Als Geschichtswissenschaftler:in möchte ich eine Liste von Orten, die in den Dokumenten genannt sind, um nach für mich relevanten Dokumenten zu recherchieren.*

Akzeptanzkriterien:

- Dynamische Generierung von Ortslisten aus dem Metadatenindex
- Metadatenindex hält verschiedene Typen von Orten vor (Publikationsort, Ort genannt in Dokument)
- Orte sind, wo möglich, mit Geodaten referenziert
- Auswahl und Anzeige der Ortsliste über die Expertensuche
- Auswahl und Anzeige der Ortsliste über die Metadatenanzeige pro Korpus und pro Dokument

Der thematische Schwerpunkt liegt bei uns in der Modellierung flexibler, unterschiedlich granularer und miteinander verschränkter Referenzierungen, die auf Korpus-, Text- und Annotations-ebene ansetzen und miteinander interagieren. Personen, Orte und Werke sind Entitäten, die für eine digitale Quellenkritik entscheidende Informationen tragen. Ein umfassendes Desiderat für die Weiterentwicklung von LAUDATIO sind die Akzeptanzkriterien, welche das messbare Ergebnis und damit den Erfolg der User Story für die nutzende Community beschreib-, validierbar und im Rahmen der Entwicklung implementierbar machen. Wir diskutieren auf der Postersession nicht nur die User Stories sondern auch die Akzeptanzkriterien.

**Odebrecht, Carolin** (2019): „A Model-to-model Approach for Developing Corpus Metadata. An “Odd” TEI Customization for Encoding Metadata“, in: Digital Humanities 2019 Conference Proceedings. Utrecht University, 9-12-7.2019, Utrecht.

**Wilkinson, M. D. et al.** (2016): „The FAIR Guiding Principles for scientific data management and stewardship“, in: *Scientific Data* 3, DOI: 10.1038/sdata.2016.18.

## Fußnoten

1. Vgl. auch weitere Ansätze zur Diskussion einer digitalen Quellenkritik *Virtuelles BarCamp zu den theoretischen Aspekten einer digitalen Quellenkritik* : <https://vdhd2021.hypotheses.org/288> [letzter Zugriff 15.Juli 2021].
2. <https://www.laudatio-repository.org/> besucht am 12.07.2021.

## Bibliographie

**Bernheim, Ernst** (1907): *Einleitung in die Geschichtswissenschaft*. Leipzig: Göschen'sche Verlagsbuchhandlung.

**Burnard, Lou** (2013): „Resolving the Durand Conundrum“, in: *Journal of the Text Encoding Initiative* 6 <http://journals.openedition.org/jtei/842> [letzter Zugriff 13.07.2021].

**Droysen, Johann Gustav** (1868): *Grundriss der Historik*. Leipzig: Veit.

**Föhr, Pascal** (2019): *Historische Quellenkritik im Digitalen Zeitalter*. Glückstadt: Verlag Werner Hülsbusch.

**Guescini, Rolf / Schulz, Konstantin / Odebrecht, Carolin.** (2021). *Laudatio Repository - Long-term Access and Usage of Deeply Annotated Information Docker Images (Version 0.1)*. Zenodo. <http://doi.org/10.5281/zenodo.5101524>

**Margulies, Simon B.** (2009): *Digitale Daten als Quelle der Geschichtswissenschaft. Eine Einführung* (Kölner Beiträge zu einer geisteswissenschaftlichen Fachinformatik, Bd. 2). Hamburg: Verlag Dr. Kovač.