

## **Analoge Forschungsdaten neu aufbereitet**

### **Die Digitalisierung und Erschließung einer lexikografischen Belegsammlung im Projekt DEMel**

#### **Müller, Caroline**

caroline.mueller4[at]uni-rostock.de

Lehrstuhl für Romanische Sprachwissenschaft und Juniorprofessur für Digital Humanities Rostock, Deutschland

ORCID: 0000-0002-8591-7859

#### **Stephan, Robert**

robert.stephan[at]uni-rostock.de

Universitätsbibliothek Rostock, Deutschland

ORCID: 0000-0001-7605-7415

#### **Labahn, Karsten**

karsten.labahn[at]uni-rostock.de

Universitätsbibliothek Rostock, Deutschland

ORCID: 0000-0002-8482-807X

**Zusammenfassung.** Um analoge Forschungsdaten einer breiteren Öffentlichkeit zugänglich zu machen, ist ihre Digitalisierung und Erschließung unter Berücksichtigung der FAIR-Prinzipien notwendig. Ein Beispiel hierfür ist das Archiv des Diccionario del Español Medieval (DEM). Nach der Veröffentlichung der ersten drei Bände des Wörterbuchs musste das Projekt 2007 aus finanziellen Gründen eingestellt werden, wodurch die etwa 865.000 Zettel umfassende Belegsammlung ungenutzt blieb. Das seit 2016 von der DFG geförderte Projekt Diccionario del Español Medieval electrónico (DEMel) hat sich zum Ziel gesetzt, diese Forschungsdaten zugänglich zu machen. Die gesamte Belegsammlung wurde dafür digitalisiert und die wesentlichen Informationen manuell erfasst. Das Ergebnis ist eine Datenbank mit knapp 700.000 Belegen zu über 32.000 Stichwörtern, die über ein Webportal und eine REST-API frei zugänglich ist. Bis zum Projektende 2026 werden die Daten zudem in standardisierten Formaten bereitgestellt und in öffentlichen Repositorien publiziert, um ihre Interoperabilität und Nachnutzbarkeit zu verbessern. Das Poster zeigt, wie die analogen Forschungsdaten digitalisiert und aufbereitet worden sind.

Publikationen zum Forschungsdatenmanagement beschäftigen sich meist ausschließlich mit digitalen Daten. In den traditionellen Geisteswissenschaften, im Gegensatz zu den Digital Humanities, liegen jedoch auch heutzutage noch viele Forschungsdaten in analoger Form vor. Um eine breite Nachnutzung dieser Daten zu ermöglichen und sie gemäß den FAIR-Prinzipien (Wilkinson et al. 2016) aufzubereiten, ist daher in der Regel zunächst ihre Digitalisierung und Erschließung notwendig. Dies galt auch für das umfangreiche Archiv des Diccionario del Español Medieval (DEM).

Das DEM ist ein Wörterbuch zum mittelalterlichen Spanisch, an dem zwischen 1971 und 2007 unter der Leitung von Prof. Bodo Müller in Heidelberg gearbeitet wurde. Als Basis für das Wörterbuch wurde eine Sammlung von rund 865.000 Belegzetteln erstellt. Hierfür nutzte das Heidelberger Projektteam mehr als 600 literarische und nicht literarische Werke sowie Text- und Dokumentensammlungen aus dem 10. bis 15. Jahrhundert als Quellen. Ab 1987 erschienen die ersten Faszikeln des Wörterbuchs, deren philologische Qualität und fachliche Bedeutsamkeit von der Fachcommunity anerkannt wurde. Pharies (1991) schrieb zum Beispiel: "The appearance of the first fascicles of this excellent dictionary of medieval Spanish is a major event for Hispanists, who [...] have not had at their disposal a scientifically conceived general dictionary devoted to the older stages of the language". Nach der Veröffentlichung der ersten drei Bände mussten die Arbeiten 2007 aus finanziellen Gründen eingestellt werden. Daher wurden nur Artikel von „a“ bis „almohatac“ publiziert, wohingegen das Archiv des DEM Belegmaterial zu Wörtern von A bis Z enthält. Diese in Zettelkästen archivierten Forschungsdaten wurden nach Projektende nicht weiter genutzt und waren für die Öffentlichkeit unzugänglich. Das Projekt Diccionario del Español Medieval electrónico (DEMe) hat sich zum Ziel gesetzt, dies zu ändern, was umso relevanter ist, da es weiterhin kein vollendetes, umfangreiches Wörterbuch zu älteren Sprachstufen des Spanischen gibt.

DEMe ist ein Gemeinschaftsprojekt der Universitätsbibliothek Rostock, des Instituts für Romanistik und der Juniorprofessur für Digital Humanities der Universität Rostock sowie bis 2020 des Instituts für Romanistik der Universität Paderborn. Im Rahmen des seit 2016 von der DFG geförderten Projekts wurde die gesamte Belegsammlung digitalisiert und die wesentlichen Informationen von den Zetteln größtenteils manuell erfasst. Das Ergebnis ist eine Datenbank mit knapp 700.000 Belegen zu mehr als 32.000 Stichwörtern (Lemmata), die

bereits über ein Webportal<sup>1</sup> sowie eine REST-API frei zugänglich ist. Um die Auffindbarkeit der Daten zu gewährleisten, hat jedes Objekt (Lemma, Beleg, Quelle, Scan des Zettels etc.) einen eindeutigen Identifikator in Form einer PURL erhalten. In der aktuellen Projektphase mappen wir die Daten für eine bessere Interoperabilität und Nachnutzbarkeit in die standardisierten Formate TEI Lex-0<sup>2</sup> und OntoLex Lemon<sup>3</sup>. Zum Projektende 2026 werden sie in öffentlichen Repositorien wie RosDok<sup>4</sup> publiziert. Damit können sie zukünftig auch unabhängig von der DEMel-Webanwendung genutzt werden.

Das Poster veranschaulicht, wie wir die analogen lexikografischen Forschungsdaten digitalisiert haben, entsprechend den FAIR-Prinzipien aufbereiten und veröffentlichen. Dabei wird die verwendete technische Infrastruktur näher beleuchtet.

## Bibliografie

- Pharies, David 1991. "Diccionario del español medieval. By Bodo Müller. Heidelberg: Winter, 1987. Fascicles 1-4, 274 pages." *Hispanic Review* 59: 79–80.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 160018.  
<https://doi.org/10.1038/sdata.2016.18>.

---

<sup>1</sup> <https://demel.uni-rostock.de/>

<sup>2</sup> <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

<sup>3</sup> <https://www.w3.org/2019/09/lexicog/>

<sup>4</sup> <https://rosdok.uni-rostock.de/>