

Suchen, sammeln, schichten, vernetzen

Ein Metakatalog mit Mehrwert

Schulz, Daniela

[schulz\[at\]hab.de](mailto:schulz[at]hab.de)

Herzog August Bibliothek Wolfenbüttel und Universität zu Köln,
Deutschland

ORCID: 0000-0003-3167-5089

Gradl, Tobias

[tobias.gradl\[at\]uni-bamberg.de](mailto:tobias.gradl[at]uni-bamberg.de)

Otto-Friedrich-Universität Bamberg, Deutschland

ORCID: 0000-0002-1392-2464

Furth, Leon

[leon.fruth\[at\]uni-bamberg.de](mailto:leon.fruth[at]uni-bamberg.de)

Otto-Friedrich-Universität Bamberg, Deutschland

0009-0001-2128-3025

Zusammenfassung. Wissenschaftliche Ressourcen auffindbar zu machen und deren Forschungsdaten projekt- bzw. institutionenübergreifend nutzen zu können, stellt aktuell noch eine Herausforderung dar. Um zumindest den ersten Schritt zu erleichtern und einen zentralen Zugang zu bieten, wurde mit der Text+ Registry ein Meta-Katalog für Ressourcen unterschiedlicher Datendomänen entwickelt, der eine Brückenfunktion erfüllt. Der Beitrag skizziert die dahinterstehenden methodischen Überlegungen und das architektonische Design. Der Fokus liegt dabei auf dem Einspielen, Verknüpfen und Anreichern von (Meta)Daten aus der Domäne der Editionen mit dem Ziel, durch die Schichtung von Informationen aus verschiedenen Provenienzen sowie der Möglichkeit händischer Kuration möglichst gehaltvolle und FAIRe Ressourcenbeschreibungen zu erstellen.

Wissenschaftliche Ressourcen entwickeln sich in ganz unterschiedlichen Kontexten und aus verschiedenen Fragestellungen heraus. Die heterogenen Parameter, unter denen sie entstehen – z.B. disziplinärer Kontext, finanzielle und personelle Ausstattung, technische Expertise – führen dazu, dass selbst Ressourcen aus einer Domäne,

die sich grundlegend den gleichen Standards verschrieben haben, kaum ohne weiteres miteinander projekt- oder institutionsübergreifend nutzbar gemacht werden können. Zahlreiche Faktoren verhindern deren Interoperabilität. Oftmals beginnt das Problem jedoch schon bei der Auffindbarkeit dieser Ressourcen bzw. der aus ihnen resultierenden Forschungsdaten, da vorhandene Katalog- und Forschungsinformationssysteme ebenfalls unterschiedliche Stoßrichtungen und Schwerpunktsetzungen aufweisen und nicht miteinander kommunizieren.

Im Kontext des NFDI-Konsortiums Text+ entsteht mit der sog. Registry eine zentrale Plattform zur Katalogisierung, Beschreibung und Verknüpfung wissenschaftlicher Textressourcen. Das übergreifende Verzeichnissystem bietet nicht nur einen schnellen Zugang zu Ressourcen und Forschungsdaten, sondern bezieht auch Dienste, Repositorien und andere Entitäten wie Personen und Institutionen mit ein und vernetzt diese miteinander. Mit ihrem charakteristischen Ansatz des *information layering* und *model driven design* übernimmt die Text+ Registry damit eine wichtige Brückenfunktion und kann Lücken zwischen bestehenden Angeboten schließen, um so die Auffindbarkeit und Zugänglichkeit zu verbessern. Damit fungiert sie auch als Katalysator für die Bereitstellung von FAIRen Daten in der text- und sprachbasierten Forschung.

Die Text+ Registry zielt darauf ab, strukturierte, umfassende und kontextbezogene Beschreibungen von text- und sprachbasierten Ressourcen aus den drei Datendomänen Collections, lexikalische Ressourcen und Editionen bereitzustellen.¹ Innerhalb dieses breiteren Rahmens stellt der Bereich der Editionen ein besonders heterogenes und technisch anspruchsvolles Feld dar. Editionen unterscheiden sich stark in Bezug auf ihre medialen Formate, ihren editorischen Skopus, ihr historisches Quellenmaterial und die jeweiligen disziplinären Konventionen. Gedruckte Editionen werden in den gängigen Bibliothekskatalogen nicht erfasst, andere Formate i.d.R. nicht. Um diesen Herausforderungen zu begegnen, wurde in enger Zusammenarbeit mit Expert*innen und Interessenvertreter*innen aus den Bereichen Forschung, Infrastruktur und Bibliotheken ein spezielles Editionsdatenmodell² entwickelt, das die konzeptionelle Grundlage für

¹ Ausführlich dazu Gradl et al. 2024

² vgl. Measure 1 der Task Area Editions 2024 und M1 der Task Area Editions in Text+. 2024

eine konsistente und zweckmäßige Katalogisierung unterschiedlicher Editionstypen bildet. Dabei mussten verschiedene Arten von Daten, Beziehungen und Granularitäten berücksichtigt werden. Parallel dazu wurden praktische Verfahren für die Datenintegration entwickelt, die sowohl die automatische Aufnahme und Schichtung als auch die manuelle Anreicherung unterstützen.

Der Beitrag möchte den Mehrwert der Text+ Registry anhand konkreter Beispiele des Dateningests und der Anreicherung von (Meta)Daten aus dieser Domäne veranschaulichen (vgl. Abb. 1). Die gewählten Beispiele verdeutlichen dabei die unterschiedlichen Herausforderungen, mit denen die Registry bei der Aufnahme von (Meta)Daten aus diversen Quellen konfrontiert ist und die jeweiligen Lösungsansätze.

Seit den 1990er Jahren pflegt Patrick Sahle einen Katalog digitaler wissenschaftlicher Editionen. Da sich diese der traditionellen bibliographischen Beschreibung größtenteils entziehen, wurden (sofern vorhanden) neben Projektbeschreibungen relevante Informationen zu Titel, Herausgebenden, Fachgebiet, Material, Sprache, Epoche, beteiligten Institutionen, Erscheinungsjahren, ISBN Nummern und anderen Identifikatoren gesammelt, um die Einzeleditionen identifizier- und referenzierbar zu machen. Die Daten sind TEI kodiert und werden online im Framework eines angepassten TEIPublisher präsentiert. Der aktuelle Stand ist unter <https://git.uni-wuppertal.de/dhsfu/sde-catalog> abrufbar.

Das Forschungsinformationssystem AGATE³ enthält Informationen zu allen Forschungsprojekten, die seit 1979 durch die Deutsche Akademienunion gefördert wurden. Darunter finden sich auch viele Langzeit-Editionsprojekte. Das Current Research Information System (CRIS) bietet seine Metadaten als Linked Open Data mit verschiedenen Mappings zu externen Ressourcen wie der Research Organization Registry (ROR) und Wikidata über eine RESTful Metadata API und einen SPARQL Endpoint an. Von dort werden sie abgefragt und in die Text+ Registry integriert.

Die am Salomon Ludwig Steinheim Institut für deutsch-jüdische Geschichte an der Universität Duisburg-Essen angelagerte epigraphische Datenbank *epidat* ist eine Plattform zur Erforschung der epigraphischen Überlieferung jüdischer Friedhöfe. Sie enthält gegenwärtig 256 digitale Editionen mit 46806 Grabmalen. Die Daten

³ vgl. Wuttke et al. 2017

werden im EpiDoc (TEI XML for Epigraphic Documents) Format bereitgestellt.

Die Monumenta Germaniae Historica (MGH) bereiten seit 200 Jahren mittelalterliche Textquellen in Form kritischer Ausgaben für die Forschung auf. Die Quelleneditionen erscheinen traditionell als Printausgaben, die in der dMGH digitalisiert und Open Access zur Verfügung gestellt wurden. Neue Bände werden nach einer Schutzfrist von zwei Jahren in die dMGH aufgenommen. Im Rahmen der openMGH werden nun nach und nach die Editionstexte aller Bände in TEI-XML aufbereitet und als Downloads zur Verfügung gestellt.

Neben textuellen Ressourcen aus verschiedenen Katalogen, Datenbanken oder Forschungsinformationssystemen ermöglicht die Registry auch die Referenzierung von Software, die in Editionsprojekten eingesetzt wird. Durch die Anbindung an eine Instanz des Research Software Directory (RSD) ist sie für eine Interaktion mit dem Basisdienst [nfdi.software](https://nfdi.org/software) (derzeit in Entwicklung) vorbereitet, und darüber hinaus durch die Integration von Metadaten zu Software aus dem SSH Open Marketplace mit dem sozial- und geisteswissenschaftlichen Bereich der European Open Science Cloud (EOSC) verbunden.

Im Beitrag werden die grundlegenden technologischen und methodischen Überlegungen, die hinter der Text+ Registry stehen, vorgestellt und darüber hinaus Herausforderungen, Grenzen und Vorteile sowie zukünftige Entwicklungen diskutiert. Der Schichtenansatz zur Erstellung möglichst gehaltvoller Ressourcenbeschreibungen und das architektonische Design werden am Beispiel des Bereichs der Editionen skizziert.

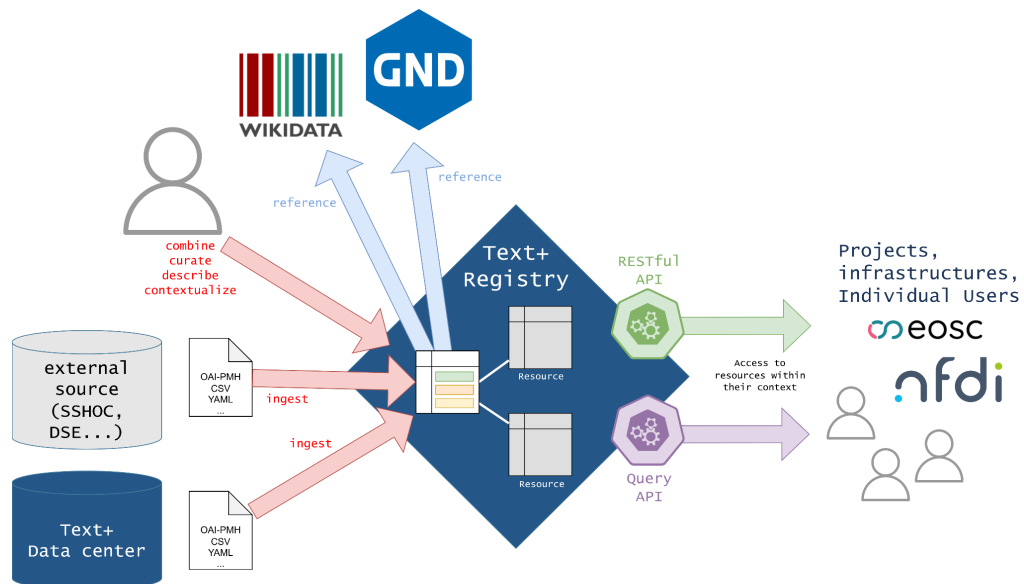


Abb. 1. Die Text+ Registry als zentraler Hub für Ingest, Anreicherung und Output (Tobias Gradl)

Bibliografie

- Gradl, Tobias, Christoph Kudella, Harald Lordick, Daniela Schulz (2024). "Towards a Registry for Digital Resources – The Text+ Registry for Editions", in: Datenbank Spektrum 24, pp. 151–160. <https://doi.org/10.1007/s13222-024-00479-0>
- Measure 1 der Task Area Editions. (2024). Datenmodell Editionenregistry (Text+) (1.1). Zenodo. <https://doi.org/10.5281/zenodo.12799883>.
- M1 der Task Area Editions in Text+. (2024). Die Editionsregistry (ERY) und das zugrundeliegende Datenmodell – Dokumentation und Spezifikation (1.1). Zenodo. <https://doi.org/10.5281/zenodo.13379995>
- Wuttke, Ulrike, Ott, Carolin, Adrian, Dominik, & Worthington, Simon (2017): AGATE. Concept for a European Academies Internet Gateway for the Humanities and Social Sciences. Zenodo. <https://doi.org/10.5281/zenodo.815916>