

Einsatz und Bewertung komponentenbasierter Metadaten in einer föderierten Infrastruktur für Sprachressourcen am Beispiel der CMDI

Zusammenfassung

Die Arbeit setzt sich mit dem Einsatz der Component Metadata Infrastructure CMDI im Rahmen der föderierten Infrastruktur CLARIN auseinander. Diese serviceorientierte Forschungsumgebung für text- bzw. sprachorientierte Fachbereiche mit seinen spezifischen Rahmenbedingungen stellt besondere Anforderungen an die Erstellung und Verwendung strukturierter Metadaten. Zu diesen Rahmenbedingungen gehören eine eher kleinteilige Organisationsstruktur, ein sehr breites Spektrum unterstützter Ressourcentypen sowie die Bereitstellung föderierter und stark integrierter Anwendungen für den Nutzer.

Es werden diverse Problemfälle, die im Laufe des Projektes offensichtlich wurden, aufgezeigt und systematisch analysiert. Auf dieser Basis werden entsprechende Lösungsansätze entwickelt, die auf die Spezifika der CMDI abgestimmt sind. Für die Erarbeitung konkreter Lösungsstrategien werden unterschiedliche Verfahren adaptiert und für die Qualitätsanalyse von Metadaten sowie zur Optimierung ihres Einsatzes in einer föderierten Umgebung vorgeschlagen. Konkret betrifft dies vor allem die Übernahme von Modellierungsstrategien der „Linked Data“-Community, die Übernahme von Prinzipien und Qualitätsmetriken der objektorientierten und testgetriebenen Softwareentwicklung für Metadatenkomponenten sowie den Einsatz von Zentralitätsmaßen der Graph- bzw. Netzwerkanalyse für die Bewertung des Zusammenhalts des gesamten Metadatenverbundes. Entsprechende Ansätze tragen der aktuellen Entwicklung von Metadatenschemata und -instanzen Rechnung die sich im zunehmenden Maße von statischen zu dynamischen Entwicklungsprozessen entwickeln. Das sehr flexible CMD-Modell unterstützt bereits wichtige Aspekte dieses neuen Ansatzes durch Modularisierung und verbesserte Wiederverwendbarkeit.

Dabei wird im Rahmen der Arbeit vor allem die Analyse verwendeter Schema- bzw. Schemabestandteile in den Vordergrund gestellt und auf die Betrachtung verwendeter Individuenvokabulare der konkreten Metadateninstanzen weitgehend verzichtet. Es wird allerdings gezeigt, dass die derzeit in CLARIN praktizierte nachträgliche Bearbeitung im Rahmen eines instanzbasierten Kurationsprozesses zur verbesserten Nutzung vorhandener Metadatenbestände führen kann.