

## **Stand und Folgen der Retrodigitalisierung des Althochdeutschen Wörterbuchs**

Das seit 1948 an der Sächsischen Akademie der Wissenschaften zu Leipzig beheimatete "Althochdeutsche Wörterbuch: Thesaurus" erarbeitet ein zehnbändiges Wörterbuch, in dem das gesamte in Texten, Glossierungen und Glossaren überlieferte deutsche Wortgut des 8. bis ca. 11. Jh. (bei Glossen alter Überlieferungstradition auch noch späterer Jahrhunderte) nach seinen belegten Formen unter Heranziehung der deutschen wie lateinischen Kontexte interpretiert und dokumentiert wird. Die Arbeitsergebnisse werden seit 1952 fortlaufend in gedruckten Lieferungen von je 40 Seiten à 80 Spalten publiziert. Die ersten fünf Bände mit den Buchstaben A bis L sind abgeschlossen, die Arbeiten am sechsten Band (M und N) im Gange.

Das Vorhaben nutzt und verwaltet zwar große Datenbestände – sei es das ca. 750.000 Zettel umfassende Belegarchiv, das von der Haager Konvention als schützenswertes nationales Kulturgut eingestuft wurde, sei es die umfangreiche Spezialbibliothek –, doch geschieht dies nach wie vor auf primär mechanische Weise. Auch die Arbeitsergebnisse werden zur Zeit noch ausschließlich in gedruckter, nicht in digitaler Form an die Benutzer weitergereicht. Ein digitales "Repositorium" im neueren Sinne des Wortes ist das Vorhaben angesichts dieser Umstände und angesichts der – verglichen mit Forschungseinrichtungen anderer Zielstellungen wie Bibliotheken – wiederum geringen Datenmengen also nicht. Überhaupt werden die Ressourcen, die große Wörterbuchunternehmen zu einer speziellen Forschungsaufgabe über lange Zeit zusammengetragen haben, normalerweise nicht gespeichert oder zugänglich gehalten: sie gelten nach Abschluss des Vorhabens als ausgewertet und werden, nicht zuletzt aus Platzgründen, aufgelöst.

Eine Digitalisierung des Zettelarchivs des Althochdeutschen Wörterbuchs wird aus verschiedenen Gründen vorerst nicht in Betracht gezogen. Vorrangig angestrebt wird eine digitale Nutzbarkeit der publizierten Bände des Wörterbuchs. Während 1999 ein erster Versuch über einen DFG-Antrag in diese Richtung scheiterte, konnte in diesem Jahr aus Mitteln der SAW eine Retrodigitalisierung als Volltexterkennung per Double-Keying in Angriff genommen werden. Die mittels TUSTEP verarbeiteten Daten sind vorerst nur zur arbeitsstelleninternen Nutzung bestimmt. Sie sind mit einer flachen XML-Auszeichnung versehen und stehen dem Vorhaben im rtf-Format zur Verfügung, so dass bereits mit der vorhandenen Computerausstattung ohne zusätzlichen informationstechnischen Aufwand auf die Daten zugegriffen und gleichzeitig ein langfristiger, hard- und softwareunabhängiger Zugriff gewährleistet werden kann.

Die Retrodigitalisierung bleibt nicht ohne Folgen für die künftige Arbeitsroutine. Zum einen muss die Codierung der vorhandenen projekteigenen Schriftsätze mit Sonderzeichen an UNICODE angepasst werden. Mit dem Übergang auf unicodebasierte Sonderzeichen muss auch ein unicodefähiges Textverarbeitungsprogramm (MS-Office 2010, Open-Office o. a.) eingeführt werden. Überhaupt muss die Auszeichnung der künftig zu erarbeitenden Daten mit der der Retrodigitalisierung übereinstimmen. Mit dem Satzunternehmen sind entsprechende Abstimmungen vorzunehmen. Wünschenswert wäre die Einführung eines Artikelredaktionssystems, doch sind die grundsätzlichen Voraussetzungen (Konzeptionierung und technische Umsetzung des Systems, digitale Aufbereitung einer kritischen Masse an Daten und Texten, kontinuierliche informationstechnische Betreuung) erst noch zu schaffen. Bis das Althochdeutsche Wörterbuch als Datenbank im Internet recherchierbar sein wird, ist noch ein gutes Stück Weg zurückzulegen.