

Sprachwandel digital erfahrbar machen

Forschungsdatenmanagement in der Variationslinguistik

Ganswindt, Brigitte

ganswindt[at]uni-marburg.de

Forschungszentrum Deutscher Sprachatlas, Universität Marburg,
Deutschland

Engsterhold, Robert

engsterhold[at]uni-marburg.de

Forschungszentrum Deutscher Sprachatlas, Universität Marburg,
Deutschland

Zusammenfassung: Am Beispiel des historischen und weltweit einmaligen „Sprachatlas des Deutschen Reichs“ und seiner Datengrundlage (über 50.000 handgeschriebene Fragebögen in Kurrentschrift) werden die Anforderungen an ein jeweils zeitgemäßes Forschungsdatenmanagement in den Bereichen der Sprachgeographie und Variationslinguistik vorgestellt und diskutiert. Der Fokus des Beitrags liegt dabei zum einen auf den Besonderheiten der historischen Materialsammlung und ihrer Bedeutung für die Regionalsprachenforschung sowie für die Analyse von Sprachwandelprozessen. Zum anderen wird die Materialerschließung und die weltweit zugängliche, kostenfreie Bereitstellung der Daten im Internet vorgestellt. Um stets aus dem aktuellen Stand der Technik zu bleiben sind kontinuierliche Modernisierungen im Forschungsdatenmanagement erforderlich, die im Vortrag präsentiert werden sollen.

1 Der „Sprachatlas des Deutschen Reichs“

1.1 Entstehungsgeschichte

In den Jahren 1876 bis 1887 verschickte der Marburger Bibliothekar und Germanist Georg Wenker an alle Volksschulen im ehemaligen Deutschen Reich Fragebogen (so genannte „Wenkerbogen“), die zur Erhebung der Lokaldialekte des Deutschen dienten. Die lokalen Gewährspersonen waren aufgefordert, 40 vorformulierte hochsprachliche Sätze schriftlich in ihren jeweiligen Dialekt zu übertragen. Diese Sätze waren so konstruiert, dass typische lautliche und einige grammatische Eigenschaften der betreffenden Dialekte

durch die Übersetzung hervortreten mussten. Nach Abschluss der in mehreren Etappen mit unterschiedlichen Fragebogen durchgeführten Erhebung hatte Wenker circa 45.000 Fragebogen aus fast 41.000 Schulorten vorliegen. Einschließlich der späteren Nacherhebungen in weiteren deutschsprachigen Regionen umfasst das Korpus historische Forschungsdatensätze zu über 50.000 Orten.

Die Wenkerbogen stellen die Datengrundlage zum „Sprachatlas des Deutschen Reichs“ dar, einem handgezeichneten Sprachatlas, der auf 1.668 Kartenblättern (eine Gesamtkarte besteht jeweils aus drei Teilkarten) phonetisch-phonologische sowie morphologische Besonderheiten der deutschen Dialekte präsentiert. Dieser Atlas und die dazugehörigen Wenkerbogen stellen seit Anbeginn ihrer Entstehung bis heute ein einmaliges Forschungsinstrument der Sprachwissenschaft dar. Heute liegt ihr besonderer Wert darin, dass sie nicht nur den analytischen Vergleich jüngerer und älterer Sprachstufen ermöglichen, sondern auch die oftmals einzige Erschließung heute nicht mehr existenter Sprachvarietäten sichern.

Aufgrund seines übergroßen Formates und der Verwendung von bis zu 22 Farben pro Karte konnte der „Sprachatlas des Deutschen Reichs“ nie gedruckt werden. Viele Forschergenerationen mussten daher stets nach Marburg reisen, um vor Ort mit dem einzigartigen Material – einer Kompletterschließung einer Nationalsprache – zu arbeiten.

1.2 Digitalisierung der historischen Forschungsdaten

Dies änderte sich erst ab dem Jahr 2000, in dem mit der Digitalisierung dieser umfangreichen Forschungsdaten im DFG-geförderten Projekt DiWA („Digitaler Wenker-Atlas“) begonnen wurde. Erst durch die Digitalisierung des Materials konnte der Atlas und seine Datengrundlage überhaupt publiziert und einer breiten fachwissenschaftlichen Community sowie interessierten Laien im Internet frei zur Verfügung gestellt werden.

Für die Digitalisierung der über 50.000 Wenkerbogen und der 1.668 Karten des „Sprachatlas des Deutschen Reichs“ wurde eine externe Firma beauftragt, die die Daten mit einem Gesamtvolumen von über 5TB auf DVDs zur Verfügung stellte. Zu Beginn des Projektes Anfang der 2000er Jahre war ausreichend Festplattenplatz zwar schon theoretisch vorhanden, aber finanziell oft nicht erschwinglich, weswegen für die Bereitstellung der Karten und Bogen nach Möglichkeiten gesucht wurde, die Dateigröße zu reduzieren. In der ersten Version des „Digitalen Wenker-Atlas“ wurde sich für ein proprietäres System von ERDAS entschieden, deren Dateiformat ECW, ein Wavelet-Format, die

Datenmenge stark reduzieren konnte. Dieses Kernmodul des DiWA-Projektes basierte auf einem JAVA-Plugin, das allerdings nach einigen Jahren nicht mehr von modernen Browsern ausgeführt werden konnte. Ab dem Jahr 2008 konnte in dem Akademiengeförderten Langzeitprojekt „Regionalsprache.de“¹ (REDE) ein Sprachgeographisches Informationssystem (SprachGIS) entwickelt werden, in das alle in DiWA verfügbaren Karten und Daten im Zuge einer Servermigration übertragen werden konnten. Zusätzlich werden im SprachGIS alle historischen Daten mit modernen Sprachatlanten (ca. 12.500 Rasterkarten sowie 14.200 in eine Datenbank überführte Vektorkarten), circa 12.300 Tonaufnahmen deutscher Dialekte sowie einer umfangreichen Online-Bibliographie zusammengeführt, die explorative und vergleichende Analysen der Dialektologie und Variationslinguistik mit Datenquellen verschiedenster Art ermöglichen. Technisch baut das SprachGIS rein auf open-source-Lösungen auf, da sich hier Standards und frei verfügbare Systeme für die Darstellung von Karten oder großen Bilddateien entwickelt haben. Dementsprechend wurde das SprachGIS als seinerzeit sehr moderne JavaScript Web-App angelegt. Das Herz dieser Anwendung ist das OpenLayers-Framework, ursprünglich in der Version 2² verwendet.

Die Bereitstellung der Rasterkarten erfolgt nach dem WMS-Standard und die Anwendung mapserv³ dient als Kartenserver. Mit der IIPImage-Software⁴ werden die Wenkerbogen bereitgestellt und mittels einer leicht modifizierten Version des zugehörigen IIPMooViewer in der Web-App dargestellt. Als interaktive JavaScript-Anwendung ermöglicht das SprachGIS nicht nur ein passives Anschauen von Karten, sondern bietet auch die Möglichkeit, selbst Karten zu erstellen, Analysen mit eigenen linguistischen Daten auszuführen und mit historischen Karten in Kontext zu setzen. Zudem lassen sich die selbsterstellten Karten für Publikationen in angemessenen Bildformaten und Auflösungen exportieren.

2 Modernisierung des Forschungsdatenmanagements

Zwar wird die SprachGIS Web-App kontinuierlich weiterentwickelt und betreut und befindet sich bereits in der dritten Hauptversion,

¹ Kehrein 2020, Ganswindt et al. 2015, <https://www.regionalsprache.de>.

² <https://openlayers.org/two/>.

³ <https://mapserver.org>.

⁴ <https://iipimage.sourceforge.io>.

nichtsdestoweniger ist eine grundlegende Modernisierung mittlerweile erforderlich.

Aktuell wird daran gearbeitet, die bereits verwendeten Standards zu modernisieren sowie weitere Standards zu verwenden, um so eine bessere Integrierung in eine moderne, vernetzte Wissenschaftslandschaft zu ermöglichen. Daher ist eine Umstellung auf OpenLayers 6 in Vorbereitung, die nicht nur eine bessere und schnellere Darstellung von Raster- und Vektorkarten ermöglicht, sondern auch Bilddateien nach dem IIIF-Standard⁵ darstellen kann. Außerdem wurden die Rasterbildformate von ECW auf JP2 umgestellt. Seit 2020 existiert ein Prototyp⁶ für einen neuen Wenkerbogen-Viewer, der auf dem IIIF 2.1-Standard basiert und neben Darstellung und Verortung der Bögen auch die Möglichkeit bietet, Transliterationen der Bogen zu betrachten oder selbst zu erstellen. Zunächst als internes Projekt begonnen, wurden circa 1.200 Transliterationen (Überführung der handschriftlichen Kurrentschrift der Bogen in eine Datenbank) normalisiert und online gestellt. Eine Erweiterung zu einem „Citizen Science“-Projekt ist derzeit in Planung.

Die Überarbeitung des SprachGIS bietet neben einer modernen Benutzeroberfläche auch eine REST-API, um Daten zur direkten Einbindung in andere Projekte oder Anwendungen zur Verfügung zu stellen. Zudem sollen Dereferenzierbarkeit und semantische Annotationen eine Einbindung in das Semantische Netz ermöglichen. Damit einhergehend wird im Rahmen der Hessischen Forschungsdateninfrastruktur (HeFDI) unser Repository derzeit überarbeitet und an ein aktuelles Forschungsdatenmanagement angepasst.

Im Vortrag sollen anhand der Wenker-Materialien und dem jeweils erforderlichen Umgang mit diesen Daten die Anforderungen und Herausforderungen an ein zeitgemäßes Forschungsdatenmanagement aufgezeigt werden. Erst die zeitgemäße Materialerschließung hat zu zahlreichen, teils wegweisenden Ergebnissen der dialektologischen und variationslinguistischen Fachgeschichte geführt, die ohne eine Interaktion zwischen Geisteswissenschaften und Informatik nicht möglich gewesen wären. Ganz entscheidend ist dafür natürlich die kostenfreie Verfügbarkeit der Forschungsdaten im Internet, die WissenschaftlerInnen weltweit bequemen Zugriff auf die Daten für ihre eigene Forschung ermöglicht.

⁵ <https://iiif.io>.

⁶ <https://wenker.online.uni-marburg.de/wenker>.

Bibliografie

Ganswindt, Brigitte, Roland Kehrein, Alfred Lameli,
„Regionalsprache.de (REDE)“, in *Regionale Variation des
Deutschen – Projekte und Perspektiven*, edited by Roland
Kehrein, Alfred Lameli Stefan Rabanus (Berlin/Boston: De
Gruyter, 2015), 421–453.

Kehrein, Roland, „Das Wenker-Material im REDE SprachGIS“, in
*Minderheitensprachen und Sprachminderheiten: Deutsche
und seine Kontaktsprachen in der Dokumentation der
Wenker-Materialien* (Hildesheim/Zürich/New York: Olms,
2020), 21–56.