

Berlin-Brandenburgische Akademie der Wissenschaften
TELOTA / DFG-Projekt „Personendaten-Repositoryum“
Jägerstraße 22/23
10117 Berlin

tel +49 / 30 / 20370–264
fax +49 / 30 / 20370–600
web <http://pdr.bbaw.de>
mail roeder@bbaw.de



berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

DAS PERSONENDATEN-REPOSITORYUM (PDR)

EINFÜHRUNG

Die *Berlin-Brandenburgische Akademie der Wissenschaften* (BBAW) ist die größte außeruniversitäre Forschungseinrichtung der Region. Im Jahr 2002 wurde hier TELOTA (*The Electronic Life Of The Academy*), eine Initiative für akademisch angewandte Informationstechnologie, ins Leben gerufen. Diese unterstützt die Akademievorhaben mit der Entwicklung informationstechnischer Lösungen für Forschungsarbeit und digitale Publikation.

Mit dem DFG-Projekt „Aufbau eines Repositoriums für biografische Daten historischer Personen des 19. Jahrhunderts“ – kurz: *Personendaten-Repositoryum* – werden bisherige Ansätze der Datenvernetzung und elektronischen Biografik weiterentwickelt. Es erforscht anhand von Personeninformationen des „langen 19. Jahrhunderts“ (1789–1914), wie sich heterogene Datenbestände miteinander verbinden und präsentieren lassen. Ziel des Projektes ist die Bereitstellung eines dezentralen Softwaresystems, welches Lehr- und Forschungseinrichtungen, Archiven und Bibliotheken ermöglicht, biographische Informationen aus verschiedenen Beständen über einen gemeinsamen Zugang zu nutzen.

Das Projekt untergliedert sich in drei Teile: 1) Der Entwurf eines geeigneten Datenmodells, welches unterschiedlichen Perspektiven und Forschungsmethoden gerecht wird, 2) der Datenaustausch mit Kooperationspartnern im In- und Ausland, und 3) die Entwicklung einer Software-Lösung auf der Basis eines zu evaluierenden Framework. Das Projekt wurde von TELOTA bei der DFG beantragt und ist für die Laufzeit von zwei Jahren bewilligt worden. Im Juli 2009 wurde die Arbeit mit drei wissenschaftlichen Mitarbeitern und drei studentischen Hilfskräften aufgenommen.

1. DATENMODELLIERUNG

Zur Strukturierung heterogener biographischer Daten verfolgt das Projekt einen neuartigen Ansatz, der bereits in einem Vortrag auf dem Workshop „Personendateien – Elektronisches Publizieren“ im September 2009 in Leipzig vorgestellt wurde (siehe Link unten). Eine Person wird darin nicht als einzelner Datensatz definiert, sondern vielmehr als die Menge aller Aussagen, die zu ihr getroffen werden. Damit ist es möglich, sowohl sich ergänzende als auch sich widersprechende Aussagen nebeneinander abzubilden, was grundlegenden Problemen biografischen Arbeitens Rechnung trägt.

In dem erwähnten Vortrag von Niels-Oliver Walkowski hieß es: „Biografisches Arbeiten verstanden als Erzeugung von identifizierenden Narrationen vollzieht semantische Konstruktionsleistungen, zu denen ein Mensch den Anlass gab, der aber nicht mit ihm zusammenfällt. Eine Folge sind konkurrierende

Narrationen, Polysemie und Kontingenz, die nicht Ausdruck mangelnder Kenntnis, sondern den Voraussetzungen biografischen Arbeitens an sich geschuldet sind.“

Gerade also, um verschiedenen Forschungsansätzen und Perspektiven gerecht zu werden, ist die kleinste Dateneinheit des Personendaten-Repositorys nicht eine Person, sondern eine einzelne Aussage zu einer Person, die in dem Datenmodell „Aspekt“ genannt wird. Ein Aspekt bündelt Bezüge zu Personen, Orten, Daten und einer Quelle. Dadurch wird es möglich sein, durch eine entsprechende Abfrage weitere Narrationen zu erzeugen, bei denen nicht unbedingt eine einzelne Person, sondern auch ein Zeitraum oder ein Ort die erste Dimension bilden könnte. Daraus gewonnene Erkenntnisse erweitern wiederum das Personenwissen.

Zudem werden die Aspekte einerseits mit den jeweiligen Quellen, andererseits mit geläufigen Identifikationssystemen, etwa mit PND und LCCN, verknüpft. Dadurch bleibt die wissenschaftliche Transparenz und die Kompatibilität mit bestehenden und zukünftigen Systemen gewährleistet.

2. KOOPERATIONEN

Da das Personendaten-Repository die Daten über seine Partner bezieht und sich selbst auf die Organisation der Daten konzentriert, anstatt eigene Datenbestände zu erarbeiten, ist Zusammenarbeit ein essenzieller Bestandteil des Projektes. Von richtungweisender Bedeutung sind dabei die Ziele der *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, die auch von der Berlin-Brandenburgischen Akademie der Wissenschaften unterzeichnet wurde und welche den wissenschaftlichen Austausch durch digitale Technik unter Gewährleistung von Quellen- und Autorentransparenz befördert.

Zunächst wurden einige Vorhaben der Berlin-Brandenburgischen Akademie der Wissenschaften dazu eingeladen, ihre Daten auf der zukünftigen Repositoryn-Plattform verfügbar und verknüpfbar zu machen. Darunter fallen die *Marx-Engels-Gesamtausgabe*, die *Alexander-von-Humboldt-Forschungsstelle*, die *Protokolle des Preußischen Staatsministeriums* sowie das Altmitgliederverzeichnis aus dem Akademiearchiv. Damit liegen bereits Daten zu mehreren zehntausend Personen vor; weitere Kooperationen mit Akademievorhaben sind in Vorbereitung, um ca. 200.000 Einträge zu erreichen.

Parallel dazu werden externe Institutionen als Partner herangezogen. Die Kooperationsmöglichkeiten reichen von der Übernahme unstrukturierter Datenbestände bis hin zum Austausch mit anderen Repositoryn. So sind bereits Kontakte mit mehreren Editionsprojekten, Bilddatenbanken und Verbund-Datenbanken geknüpft worden, die im weiteren Verlauf des Projektes zu vertiefen sind. Die Grundlage für Austausch und Veröffentlichung der Personendaten schafft im besten Falle eine Vereinbarung im Sinne des *Open Access*, wobei auch individuell davon abweichende Verabredungen getroffen werden können.

Da beabsichtigt ist, nicht nur die Daten, sondern auch die Infrastruktur des Personendaten-Repositorys zur Verfügung zu stellen, ist das Projekt auch für solche Vorhaben interessant, deren historischer Rahmen das 19. Jahrhundert nicht berührt. In diesem Fall kann eine Abmachung über die Einrichtung einer gleichartigen Infrastruktur mit selbst bestimmbareren Inhalten, also eine Art „Schwester-Repository“, geschlossen werden.

3. ENTWICKLUNG

Im Rahmen der Vorbereitungen für die praktische Umsetzung des Personendaten-Repositorys wurde eine Liste etablierter Software-Pakete erstellt. Eine Evaluation zeigt, welches die Kernkomponente der Datenhaltung des PDR bildet. Da die Software auch anderen Projekten zur Verfügung steht, gehören Dokumentation, Konfigurierbarkeit, Skalierbarkeit, Erweiterbarkeit und Schnittstellen zu den wesentlichen Anforderungen der Evaluation.

Als bereits längerfristig genutzte Software zur Erfassung von Personendaten existiert bereits innerhalb der BBAW der von der TELOTA-Initiative entwickelte Archiv-Editor (Link s. u.). Dieser wird im Rahmen des PDR eine zentrale Rolle als Werkzeug für die Eingabe von Personendaten spielen und wird entsprechend der veränderten Anforderungen weiterentwickelt.

Neben der Arbeit an der Repositorien- und Archivsoftware fällt auch die Konvertierung von Datenbeständen in den Entwicklungsbereich. Dabei werden sowohl manuelle als auch automatische Verfahren eingesetzt. Die Zerlegung der Biogramme in Einzelaspekte erfolgt zunächst anhand einer einfachen Syntaxanalyse, wird dann über Abkürzungs-, Orts- und Personenverzeichnisse tiefstrukturiert und zum Abschluss manuell korrigiert. Begonnen wurde mit dem Personenregister der *Protokolle des Preußischen Staatsministeriums*, welche sich als abgeschlossenes Projekt und mit über 22.000 Kurzbiogrammen als hervorragende Ausgangsbasis anbot. Als zweite Quelle für Personendaten des 19. Jahrhunderts wurden die umfangreichen Register der *Marx-Engels-Gesamtausgabe* ausgewählt. Weitere Akademienvorhaben, etwa die *Alexander-von-Humboldt-Forschungsstelle*, haben ihre Daten ebenfalls bereits zur Verfügung gestellt.

AUSBLICK

Da unser Projekt in großem Maße auf Kooperationen setzt, ist es unser Wunsch, mit interessierten Projekten aus allen denkbaren Fachgebieten ins Gespräch zu kommen und Partnerschaften zu schließen. Ebenso freuen wir uns auf Fragen, Hinweise und Anregungen sowie auf den Austausch von theoretischen und technischen Lösungsansätzen. Ein Workshop, zu dem sowohl ähnliche Projekte, Kooperationspartner und interessiertes Fachpublikum eingeladen werden, ist für Herbst 2010 geplant.

WEITERE INFORMATIONEN

Webseite des Personendaten-Repositorys	pdr.bbaw.de
TELOTA-Initiative der BBAW	www.telota.de
Archiv-Editor (Einführung und Download)	www.telota.de/telota-lab/personendb
Workshop "Personendateien" in Leipzig	www.saw-leipzig.de/aktuelles/personendateien

TERMINE

PDR-Workshop

27. – 29. September 2010

Berlin-Brandenburgische Akademie der Wissenschaften