

Semantik, Linked Data und Web-Präsentation – Grundlagen der Nachlasserschließung im Editionsportal www.pueckler-digital.de

Dr. Jana Kittelmann, Forschungsstelle Pückler-Archiv, Robinienweg 5, 03042 Cottbus. Tel.: 0355/7515138, jana.kittelmann@pueckler-museum.de

Dr. Christoph Wernhard, TU Dresden, Fakultät für Informatik, Knowledge Representation and Reasoning Group, Nöthnitzer Str. 46, 01187 Dresden, christoph.wernhard@tu-dresden.de

Mit dem Editionsportal www.pueckler-digital.de wollen wir beispielhaft zeigen, wie man mit wenig Ressourcen und einfachen Mitteln wertvolles, bislang nur am Standort recherchierbares Material systematisch erschließen und digital zugänglich machen kann. Anlass ist die Erschließung des Nachlasses Pückler-Carolath-Haugwitz. Das vornehmlich aus Briefen, aber auch aus Lebensdokumenten und Manuskripten bestehende Konvolut befindet sich seit dem Jahr 2009 in der Forschungsstelle Pückler-Archiv Branitz und ist in den Jahren 2012/2013 systematisch erschlossen worden. Die Bestandsbildner stammen aus dem familiären Umfeld des Gartenkünstlers und Schriftstellers Hermann von Pückler-Muskau (1785–1871). Den Hauptteil des Bestandes stellt mit 1491 Briefen die Korrespondenz zwischen Lucie von Pückler-Muskau (1775–1854) und ihrer Tochter Adelheid von Carolath-Beuthen (1797–1849) dar.

Unser Gegenstandsbereich umfasst Briefe, Personen, Orte, Zeitpunkte und -räume, historische und persönliche Ereignisse sowie Archivstandorte. Wir verwenden für diesen Gegenstandsbereich eine eigene konzeptuelle Modellierung, die sich an den im deutschsprachigen Raum maßgeblichen *Regeln zur Erschließung von Nachlässen und Autographen (RNA 2010)* orientiert, jedoch an unsere konkreten Anforderungen angepasst ist und sich leicht in ähnliche Modelle abbilden lässt. Eine Spezifikation unserer Modellierung wird als natursprachlicher Text veröffentlicht, begleitet von maschinenverarbeitbaren Dokumenten für unterschiedliche Anwendungen: einem in SQL spezifizierten relationalen Schema, einem RDF-Schema sowie einer OWL-Ontologie.

Bei Metadaten zu Archivmaterial wie Korrespondenzen ergibt sich ein bestimmtes Zusammenspiel von normierten Fakten und Zusatzinformation, die entsprechend den *RNA* in *Bemerkungen* als freier Text ausgedrückt werden. Beispielsweise lässt sich das Entstehungsdatum eines Briefs normiert darstellen. Das Datum kann aber genau oder geschätzt sein. Es kann in unterschiedlicher Weise vom Verfasser angegeben worden sein, wie vollständig, unvollständig, fehlend oder auch falsch. In unserer Modellierung lassen sich solche häufig vorkommenden Varianten mit Hilfe von vorgegebenen Werten für begleitende Bemerkungen so ausdrücken, dass sie auch inhaltlich maschinell erfasst werden können.

Der Erschließungsprozess zielt auf unterschiedliche Nutzungsformen, die kurz- und langfristig sind:

Präsentation im Web. Die in den Metadaten enthaltene Information wird als Hypertext im Web veröffentlicht, für ausgewählte Briefe ergänzt durch Grundtranskriptionen und Digitalisate in verschiedenen Auflösungen. Es werden verschiedene Navigationsmöglichkeiten über Hyperlinks zur Verfügung gestellt, die jedoch entsprechend der Grundkonzeption von HTML beim Entwurf der Webseiten fest bestimmt werden: Beispielsweise lassen sich Übersichten über Teilbestände der Briefe entsprechend verschiedener, aber fest vorgegebener Ordnungskriterien anzeigen. Die Web-Präsentation ist ohne weitere Hilfsmittel mit einem Web-Browser über das Internet zugänglich. Typischerweise finden Nutzer die Webseiten als Ergebnisse einer stichwortbasierten Suche mit einer

Suchmaschine. Auf diese Weise werden die Branitzer Bestände auch Nutzern bekannt gemacht, die nur nach relevanten Schlüsselworten gesucht, jedoch keine Kenntnis von den Beständen und keine explizite Archivsuche vorgenommen haben. Die Web-Präsentation kann die Arbeit im Archiv begleiten und vorbereiten.

Export in den zentralen Verbundkatalog für Autographen (Kalliope). Mit dem Eintrag unserer Metadaten in Kalliope machen wir das Material für Recherchen und Forschung in diesem Kontext verfügbar. Wir nutzen die Offline-Importschnittstellen von Kalliope, da sie uns Flexibilität in der Modellierung und Pflege der Metadaten erlaubt, wie die Generierung der verschiedenen hier beschriebenen Anwendungen sowie komplexere Konsistenzprüfungen. Unser System stellt die Metadaten in der von Kalliope benötigten Variante des Formats EAD (Encoded Archival Description) als XML-Dateien zur Verfügung.

Schnittstellen zu Technologien des Semantischen Web. Wir möchten zukünftige Entwicklungen und Web-Anwendungen unterstützen, indem wir unsere Metadaten im RDF Format als Web-Dokumente zur Verfügung stellen. Die Metadaten können dann mechanisch eingelesen und in zukünftigen Anwendungen des Semantischen Web mit anderen Wissensquellen verknüpft werden, etwa mit der Gemeinsamen Normdatei (GND), die ihre Datensätze auch im RDF-Format zur Verfügung stellt. Die konsequente Verwendung von dauerhaften und globalen Objektidentitäten in unserem Modell folgt der Grundstufe in der Architektur des Semantischen Web, wo diese als Uniform Resource Identifier (URI) bezeichnet werden.

Pflege der Daten als relationale Datenbank. Das auf dem mathematischen Konzept der Relation beruhende relationale Datenmodell erlaubt es, die an die *Regeln zur Erschließung von Nachlässen und Autographen* (RNA 2010) angelehnte konzeptuelle Modellierung unserer Metadaten in ein relationales Datenbankschema zu übertragen. Wir verwenden eine entsprechende Repräsentation unserer Metadaten als relationale Datenbank zur hausinternen Pflege. Über Schnittstellen entsprechend dem SQL-Standard wird diese Datenbank mit den Programmen zur Generierung der Anwendungssichten verbunden.

Die wesentlichen Aspekte stellen sich wie folgt dar: Das Ziel ist, Nachlassmaterialien zugänglich zu machen. Der Bestand, den wir hier vorstellen, ist heterogen und umfasst über 2000 Briefe, Lebensdokumente sowie Manuskripte. Die zur Verfügung stehenden finanziellen Mittel sind begrenzt. Eine historisch-kritische Edition des gesamten Bestands ist nicht verhältnismäßig. Es sollen unterschiedliche Anwendungen unterstützt werden. Anstelle eines geschlossenen Software-Systems ermöglichen offene Standards Anwendungen mit aktueller Standard-Software und bilden den Ansatzpunkt für künftige Entwicklungen, insbesondere im Bereich des Semantic-Web. Wir haben gesehen, dass vorhandene Standard-Modellierungen unseres Gegenstandsbereichs unsere Anforderungen nicht vollständig abdecken und mussten diese ergänzen. Darüber hinaus sind in unserer Repräsentation Metadaten verknüpft mit Informationen, die elektronisch verfügbar, aber semantisch nicht vollständig maschinell verarbeitbar sind, wie Grundtranskriptionen und Digitalisate. Der Erschließungsprozess umfasst mehrere Stufen, die softwareseitig unterstützt werden: Erfassung der Metadaten, Erstellung von Grundtranskriptionen und Kommentaren, Zusammenstellung von Personen- und Ortsdaten, Verknüpfung mit Normdaten (GND, Wikipedia, WW-Person), sowie die Erstellung und Präsentation von Digitalisaten.

Eckdaten zum Projekt insgesamt:

- Titel und Website: www.pueckler-digital.de

- Laufzeit: seit 2012

- Verwendete Tools: Datenvalidierung, Dokumentgenerierung: SWI-Prolog (Programmiersprache, ISO-Standard mit Erweiterungen); Texteditor: GNU Emacs; eigenes Format zur textbasierten Erstellung und Pflege von Metadaten und Grundtranskriptionen, Schema basierend auf den Richtlinien RNA; Kalliope Encoded Archival Description XML Format für Bestände/Objekte (basierend auf dem EAD Standard der Society of American Archivists/Library of Congress und XML, W3C Recommendation); SQL (ISO Standard); RDF (W3C Recommendation), geplant; Relationale Datenbanksysteme: OpenOffice Base und MySQL, basierend auf SQL

- Kontaktdaten: Editionsportal pueckler-digital.de, Forschungsstelle Pückler-Archiv, Robinienweg 5, 03042 Cottbus