arbido

2014/2 Le catalogage des médias audiovisuels

Kwasnitza Stefan, Abteilungsleiter Informationszugang / stv. Direktor Bundesarchiv

Tags: audiovisuel électronique,

Der normierte Sucheinstieg – GND und RDF

Die Vereinigung unterschiedlicher digitaler Formate im gemeinsamen Suchraum erfordert mehr als die bisherigen Lösungen in der Erschliessungspraxis zu leisten vermögen. Um interoperable Daten zu verlinken werden sie mit RDFVokabularen beschrieben. Die Verwendung der GND ermöglicht normierte Sucheinstiege.

«My Dear Eugenia [,] I am home again safe […]»1 Mit diesen Worten beginnt der auf den 4. Februar 1863 datierte Brief eines gewissen John A. Bushnell aus Missouri. Er berichtet im amerikanischen Bürgerkrieg einer Freundin von den Kämpfen um Vicksburg, Mississippi. Das Erschliessungsprojekt Civil War On The Western Border der Kansas City Public Library vereint Quellenmaterial aus rund 25 Archiven der Bundesstaaten Missouri und Kansas aus den Jahren 1854–1865. Die digitalisierten Originaldokumente finden sich in der eigens entwickelten Oberfläche mit interaktiven Zeitachsen, Bildern, Karten, Nachschlagewerken, Vorlesungsvideos und wissenschaftlicher Literatur verknüpft2.

Der digitalisierte Brief von Bushnell lässt sich zusammen mit einer Transkription und Hintergrundmaterial wie dem C-Span Video einer Hochschulvorlesung zur Schlacht von Vicksburg aufrufen3. Eigentliches Herzstück der Anwendung ist ein zukunftsweisendes Visualisierungs-Werkzeug, das dem Benutzer ermöglicht, frei durch jenes Beziehungsnetz aus Personen, Gruppen, Orten und Ereignissen zu wandern, welches sich aus dem umfangreichen Archivmaterial knüpfen lässt. Ein Beziehungsgeflecht, das durch die Techniken des semantischen Webs entsteht. Denn wie im Projekt aus Kansas gründet die semantische Vernetzung auf der Zusammenführung von Metadaten verschiedener Bestände, die auf einer schmalen aber wirkungsvollen Schicht aus interoperablen Metadaten basieren.

Inhalte zusammenführen

Erschliessung zählt zu den zentralen Aufgaben von Gedächtnisinstitutionen wie Archiven, Museen und Bibliotheken. Dies gilt nicht nur für handschriftliches Material oder gedruckte Werke, sondern auch für Ton- und Bilddokumente. Während bis in die achtziger Jahren die Informationsexperten in Gedächtnisinstitution den Fokus auf die Schaffung elektronischer Metadaten legten, fanden die audiovisuellen Dokumente und ihre Eigenschaften in 90er Jahren Beachtung.

Dabei wirkt bis heute nach, dass in der Welt des Papiers die Praktiken in der Anwendung von Normen, Standards, Konzepten und Arbeitsabläufen bereits klarer definiert waren, während in der AV-Erschliessung nach Orientierung noch gesucht wurde. Im Vordergrund standen dabei für alle Bestände lediglich beschreibende Praktiken: Dokument und Kontext wurden erfasst, die inneren Strukturen der eigentlichen Objekte blieben wenig sichtbar.

Durch zunehmend skalierbare Infrastrukturen und geeignete Hard- und Softwarearchitekturen wurde die Bereitstellung digitaler Medien und digitalisierter Inhalte einfacher. In unterschiedlichsten Verfahren werden heute diverse Medientypen automatischen Analysen mittels Cloud Computing unterzogen: Digitale Audio-Video-Daten lassen sich in zeitsynchrone Textinhalte umwandeln, digitale Bilddaten werden mittels Mustererkennung gruppiert.

Die technischen Möglichkeiten überholen hierbei oftmals die bestehenden Strukturen der Erschliessung. In den zurückliegenden Jahrzehnten entstand eine Vielzahl an Datenformaten, Regelwerken und weiteren Standards, um die Datenmengen nachnutzbar zu halten4. Mit den neuen Verfahren der digitalen Welt liegen die Schwerpunkte nun stärker auf der Möglichkeit Inhalte zusammenzuführen. Damit erfahren die traditionellen Funktionen der Erschliessung, bzw. Beschreibung von Objekten mit Metadaten in der digitalen Welt eine Erweiterung und Neuausrichtung5.

Interoperabilität als Herausforderung

Nicht das massive Sammeln von Daten oder deren Verarbeitung, sondern das Teilen von Daten, ihre Interoperabilität, sind die grosse Herausforderung der Gegenwart. Die Frage nach der angemessenen Wahl von Datenformaten stellt sich für die Integration der jeweils eigenen Metadaten in institutionsübergreifende Suchmöglichkeiten. Metakataloge erbringen dabei aufgrund neuer Suchmaschinentechnologie und eigenem, zentralen Metadatenmanagement erheblichen Mehrwert in Ergänzung zum einzelnen Katalog der jeweiligen Institution6.

Die Erschliessung von AV-Beständen hat sich auch in diesen normierten Sucheinstiegen zu bewähren: Metadaten müssen alle Informationen enthalten, welche den Zugriff auf institutionsinterne und -externe Informationsquellen ermöglichen. Eine Herausforderung liegt dabei in der richtigen Normdatenpflege: Wie kann die schmale Schicht der Metadaten jene relevanten Daten liefern?

Domänenübergreifende Normdaten

Die Zukunft der Rechercheinstrumente für Personennamen liegt in riesigen Datenbanken, die es ermöglichen, mittels Metadaten Beziehungen zwischen den digitalen Objekten zu erzeugen. In der deutschsprachigen Welt bildet die Gemeinsame Normdatei (GND) mit 10 Mio. Beschreibungen für Personen, Schlagwörter oder Körperschaften ein weites semantisches Netz. Die GND selbst ist in VIAF (Virtual International Authority File) eingebunden, das die nationalen Personennamen-Normdateien virtuell zusammenführt7. Über ein maschinelles Matching-Verfahren werden in VIAF rund 14,5 Mio. Cluster gebildet. Diese Cluster lassen sich in eigene applikatorische Landschaften aufnehmen und bieten so die domänenübergreifende Integration von Normdaten. Ein praktisches Beispiel mit dem Ziel der institutionenübergreifenden Normdatenpflege ist das von der Deutschen Forschungsgemeinschaft (DFG) geförderte Kooperationsprojekt «IN2N» zwischen der Deutschen Nationalbibliothek und dem Deutschen Filminstitut8.

In Verbindung mit einer verstärkt auf personalisierten Algorithmen basierenden Suche strukturiert sich das Web zunehmend nach den Bedürfnissen einer «mobile-only» Generation. Auf das mobile Internet wird mit Haptik über kleine Bildschirme zugegriffen9. Limitierenden Faktoren stellen die Displaygrösse und die Bedienung dar. Kleinere Displays lassen keine mehrspaltigen Layouts zu, Bedienungselemente können nicht beliebig verkleinert werden, das Tippen wird aufwändiger10. Dies führt dazu, dass die Eingaben sich in der Regel auf ein Minimum beschränken. Diese Besonderheiten mobiler Endgeräte verändern das Verhalten und die Vorstellung der Internetnutzung. Komplexe Recherchen und schwer zu parametrisierende Suchfunktionalitäten lokaler Kataloge oder benutzerunfreundliche Oberflächen werden bereits heute konsequent gemieden11. In den ständig wachsenden Suchräumen ermöglicht die Nutzung der GND normierte einfache Sucheinstiege für die mobile Zwei-Wort-Suche, kooperativen Erschliessungsmöglichkeiten und die Vernetzung über die eigene Community hinaus.

Resource Description Framework

Um die Sichtbarkeit vorhandener Erschliessungleistungen in diesen globalen Datenstrukturen zu erhalten, sollten Metadaten eine dynamischen Rekombination und Wiederverwendung erlauben: Lokale Datenstrukturen müssen in globale Datenstrukturen übertragbar sein. Dies gelingt mit Hilfe des spartenunabhängigen Austauschformats RDF (Resource Description Framework). Das RDF-Modell und seine Serialisierung basieren auf der XML-Syntax. In Tripeln (Subjekt, Prädikat, Objekt) werden Aussagen über Ressourcen beschrieben.

Da die bestehenden RDF-Vokabulare häufig nicht alle spezifischen Regeln einer Institution abdecken, lassen sich weitere Vokabulare selbst definieren. Die Strukturen von eigenen, ergänzenden Vokabularen gilt es dann als Ontologien zur Nachnutzung offenzulegen.

Im Projekt Civil War On The Western Border der Kansas City Public Library wird RDF genutzt, um eine einheitliche Metadatenschicht aus unterschiedlichen Datastreams zu speisen12. In den RDF-Daten werden automatisiert Beziehungen geknüpft und diese wieder an die digitalen Objekte gekoppelt.

Stefan Kwasnitza

Stefan Kwasnitza (1977) arbeitet seit 2014 als Abteilungsleiter Informationszugang für das Schweizerische Bundesarchiv (BAR). Er war zuvor während fünf Jahren bei der Schweizerischen Nationalbibliothek tätig. Seit April 2019 ist er stellvertretender Direktor des BAR. Stefan Kwasnitza ist Historiker und verfügt über eine Weiterbildung als Webprogrammierer und ein Diplom in IT Business Management der Universität St. Gallen.

Résumé

Français
L’association de différents formats numériques dans un espace de recherche commun requiert davantage que les solutions proposées à ce jour dans la pratique du catalogage. Pour relier des données interopérables, ils sont décrits avec des vocabulaires RDF. L’utilisation de la GND («Gemeinsame Normdatei», angl. «Integrated Authority File») permet des entrées de recherche normalisées.