

Die Öffnung bibliothekarischer Daten - Eine Spielwiese

Seit der ersten Freigabe von Katalogdaten durch das hbz im Jahr 2010 haben mittlerweile fast alle großen Bibliotheksverbünde sowie die Deutsche Nationalbibliothek nachgezogen.

Die meisten dieser Daten werden als (Linked) Open Data unter der Creative Commons Zero Lizenz zur Verfügung gestellt, d.h. die Daten sind gemeinfrei, sie gehören allen und dürfen zu beliebigen Zwecken und ohne Auflagen genutzt werden.

Es scheint mir aber noch wenige Services, Portale oder Tools zu geben, die sich mit diesen Daten und ihrer Nutzung beschäftigen, und ich habe den Eindruck, dass die Phase des Kennenlernens und Spielens mit den Möglichkeiten erst begonnen hat.

Von den veröffentlichenden Institutionen wird gerne auf die Möglichkeiten hingewiesen, die im RDF Format liegen, in welchem die meisten Daten zum Download angeboten werden, nämlich die Adressierung und Einbindung im Rahmen des World Wide Web. Das wird auch bereits gemacht, z.B. von der Wikipedia.

Der Vortrag will, an einem Beispiel, nur ein paar Gedanken dazu beitragen.

Schon im Studium hat mich fasziniert, dass neben der Geschichte der großen deutschen Literaturwerke und ihrer Autoren diejenige der Übersetzungen und Übersetzer ein recht kärgliches Schattendasein gefristet hat. Abgesehen von den Fällen, wo ein literaturgeschichtlich relevanter deutscher Autor zugleich als Übersetzer tätig war, sind mir keine kritischen Werkeditionen bekannt, die Übersetzungen zum Gegenstand haben. So wird es wohl irgendwann bestimmt eine kritische Edition der Tieckschen Don Quixote-Übertragung geben, jedoch so rasch keine der Braunfelsschen, obwohl beide eine ähnliche Wirkung entfaltet haben. Die bibliografische Lage war zu meiner Studienzeit ebenfalls mager. Mittlerweile gibt es aber schon eine ganze Reihe Bibliografien zu verschiedenen Nationalliteraturen in deutscher Übersetzung. So z.B. die in einem DFG-Projekt entstandene *Bibliographie niederländischer Literatur (von den Anfängen bis 1830) in deutscher Übersetzung* unter der Leitung von Prof. Dr. Konst an der Freien Universität Berlin, um nur eine zu nennen.

Ohne auch in der Kunst des Bibliografierens ein Experte zu sein, kann ich mir vorstellen, dass es ein mühevolleres Unterfangen ist, überhaupt erst einmal eine Basis für ein solches Unterfangen zu schaffen.

Die jetzt freigegebenen Titeldaten der deutschen Bibliotheksverbünde könnten für das Projekt der *Weltliteratur in deutschen Übersetzungen* jedoch äußerst hilfreich sein.

Folgende Bestände können bereits bezogen und verarbeitet werden:

DNB - Deutsche Nationalbibliothek (ca. 11 Mio. Titeldatensätze)

BVB - Bibliotheksverbund Bayern und KOBV - Kooperativer Bibliotheksverbund

Berlin-Brandenburg (ca. 23 Mio. Titeldatensätze)

GBV - Bibliotheksverbund der Länder Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen und der Stiftung Preußischer Kulturbesitz (ca. 37 Mio. Titeldatensätze; nur ausgewählte Datenbestände)

HBZ - Verbundkatalog NRW (ca. 18 Mio. Titeldatensätze)

HeBIS - Hessisches Bibliotheksinformationssystem (ca. 17,5 Mio. Titeldatensätze)

SWB - Verbundkatalog des Südwestdeutschen Bibliotheksverbund (ca. 17,5 Mio. Titeldatensätze)

Die Suche nach den relevanten Übersetzungen war bislang nicht so einfach. Über die OPAC-Services der Bibliotheken und Verbünde konnte - auch in den erweiterten Suchen - nur auf ausgewählte Datenfelder der MAB- oder MARC-Datensätze zugegriffen werden. Zugriffe per API waren erlaubnispflichtig und auf spezielle Nutzergruppen beschränkt.

Autorennamen konnten dann freilich in vielen Schreibweisen vorkommen und die Information, dass es sich um eine Übertragung ins Deutsche handelte, konnte auf unterschiedlichste Weise in den vielen Datenfeldern verborgen sein.

Erst seit den 90er Jahren wurde beispielsweise der dreistellige Sprachencode im Feld 37b eines MAB-Titeldatensatzes erfasst - und auch dann keineswegs immer.

Die häufig verwendete Zeichenfolge "[Übers.]" konnte in den MAB-Feldern 100-104 einen Hinweis liefern.

Zeichenfolgen wie "Übers. aus dem", "Übers.:", "Deutsch von", "In der Übersetzung von", "durch ... verdolmetscht" usw. konnten in verschiedensten Feldern eines Titeldatensatzes auftauchen.

Ebenso das Kürzel "<dt.>" im Einheitssachtitel oder einer Nebeneintragung.

Die freien Daten ermöglichen nun aber das Schreiben eigener Filter.

Zudem ist auch die Gemeinsame Normdatei, welche die verschiedenen Ansetzungsformen von Personennamen liefert, mittlerweile frei verfügbar.

Will man nun die in allen deutschen Bibliotheken zu einem Autor vorhandenen Titel in deutscher Übersetzung finden, so scannt man die Daten unter allen Ansetzungsformen des Namens sowie den o.g. Kriterien. Einfach - aber bislang so nicht möglich.

Gleiches gilt natürlich für die Suche nach einem Titel, wenn man z.B. die Don Quixote Übersetzungen des 17. und 18. Jahrhunderts zusammenstellen möchte (was freilich heute kein Desiderat mehr ist).

Der kleine Vortrag möchte an dieser Stelle (in seinem Kern) Beispiele der freien Titeldaten, den Aufbau eines Suchscriptes, die dadurch erzielbaren Ergebnisse und mögliche Weiterverarbeitungen zeigen.

Hat man einmal die Möglichkeit geschaffen, die relevanten Datensätze herauszufiltern, sind natürlich weitere Anwendungsmöglichkeiten denkbar. Die RDF-Triple der offenen Daten lassen insbesondere zu, dass jedes Objekt, z.B. eine Titelansetzung oder ein Autorname, selbst zum Subjekt anderer Triple wird, die auf weitere Ressourcen im Web verweisen.

So könnte ein gedachtes Portal nicht nur die bibliografischen Informationen, sondern z.B. auch gleich den Zugriff auf die Digitalisate aus den wachsenden Sammlungen an Retrodigitalisierungen der Bibliotheken bieten. Informationen zu den Autoren, Lebensdaten, Historisches etc.

Unter Auswertung der Druckorte ist ferner eine Landschaft der literarischen Übersetzungstätigkeit in Kartenform denkbar.