



## **Digitale Dienste in einer neuen digitalen Bibliothek – neue Methoden der Präsentation des Bibliothekskatalogs**

### **Asgeir Rekkavik**

Oslo Public Library, Deichmanske bibliotek  
Oslo, Norway  
E-mail: asgeir.rekkavik[at]kul.oslo.kommune.no

### **Kim Tallerås**

Oslo and Akershus University College of Applied Sciences  
Oslo, Norway  
E-mail: kim.talleras[at]hioa.no

### **Anne-Lena Westrum**

Oslo Public Library, Deichmanske bibliotek  
Oslo, Norway  
E-mail: anne-lena.westrum[at]kul.oslo.kommune.no

**Deutsche Übersetzung:** Andrea Köpcke  
Deutsche Nationalbibliothek

**Meeting:**

**80 — Inspired moments in cataloguing — Cataloguing**

### **Abstract:**

*Wenn ein Bibliotheksbenutzer den Online-Katalog nach Werken eines bestimmten Verfassers durchsucht, erhält er normalerweise eine lange Liste, die verschiedene Übersetzungen und Ausgaben aller Bücher dieses Autors umfasst, nach dem Titel oder dem Erscheinungsdatum sortiert. Als Versuch, etwas Ordnung in dieses Chaos zu bringen, wendete das Projekt PODE ein Verfahren automatisierter FRBRisierung an, das auf den Informationen basiert, die in MARC-Datensätzen enthalten sind.*

*Um zu demonstrieren, wie die gesamte Produktion eines Verfassers als kurze und klare Liste eindeutiger Werke präsentiert werden kann, experimentierte das Projekt auch mit einer RDF-Repräsentation von MARC-Datensätzen.*

*Das Ergebnis war die Anregung, eine vollständig konvertierte RDF-Repräsentation des Bibliothekskatalogs weiterzuverfolgen.*

## **Einleitung**

Die Pläne für ein neues Bibliotheksgebäude in Oslo, Norwegen, und die Diskussion über Bibliotheksdienstleistungen, die von diesen Plänen beeinflusst wurden, haben der Oslo Public Library) eine indirekte Gelegenheit gegeben, zu untersuchen, wie ihre Metadaten in neuen Kontexten und auf eine Weise, die zu besserem Service beiträgt, genutzt werden können.

Das Projekt Poda hat während der letzten Jahre im Hinblick auf Mash-Ups, Referenz-Modelle wie FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998), neue Generationen von OPACs und Linked Data mit deskriptiven Metadaten experimentiert. Diese Arbeit hat zu mindestens einer zentralen Einsicht geführt: man kann auf der Basis bereits existierender Metadaten keine besseren Dienste erschaffen, als die Qualität der Metadaten unterstützen wird.

Das Projekt Poda endete vor einem Jahr, aber die Ergebnisse und Erkenntnisse sind nun in ein Projekt zur Schaffung öffentlicher Dienste für die Bibliotheksbenutzer überführt worden. Das Projekt basiert auf der RDF-Repräsentation der MARC-Datensätze und hat zum Ziel, eine interaktive Schnittstelle zu schaffen, die den Bibliothekskatalog auf eine neue Weise präsentiert. Das Projekt wird den Bibliothekskatalog mit externen Datenquellen verknüpfen und ihn durch eine Touch-Screen-basierte Schnittstelle auf eine visuelle und intuitive Art präsentieren. Wir arbeiten außerdem an einer semantischen Sammlung von Bücherempfehlungen, die mit den RDF-Katalogdatensätzen verknüpft werden.

## **Das Projekt Poda**

Das Projekt Poda arbeitete mit neuen Verfahren der Nutzung von Metadaten, die in Bibliothekskatalogen zu finden sind, insbesondere mit Verfahren, die Kataloginhalte mit anderen Ressourcen zu mischen. Außerdem prüfte das Projekt die Möglichkeiten und Herausforderungen, die die Netz-Technologien hinsichtlich heutiger Systeme und Gepflogenheiten bieten.

Das Projekt verwendete herkömmliche, von den Bibliothekssystemanbietern unterstützte Protokolle und es konvertierte und bearbeitete Daten, um sie zu FRBRisieren und damit Linked Library Data „mit einer Bedeutung“ liefern zu können.

Das Projekt experimentierte auch mit einer RDF-Repräsentation von MARC-Datensätzen, um zu demonstrieren, wie die gesamte Produktion eines Verfassers als kurze und klare Liste eindeutiger Werke präsentiert werden kann, die einfach nach deren verschiedenen Expressionen und Manifestationen durchsucht werden kann. Außerdem wird die Darstellung durch die Verlinkung der Instanzen aus dem Datenbestand mit passenden oder entsprechenden Instanzen in externen Beständen mit zusätzlichen Informationen über Verfasser und Werke angereichert.

Unsere zweite zentrale Einsicht war: eine RDF-Repräsentation unseres Bibliothekskatalogs liefert uns einen besser zu verarbeitenden Datenbestand mit wesentlich mehr Möglichkeiten, Dienste für unsere Benutzer zu schaffen als die herkömmlichen MARC-Datenbestände.

## **FRBR und FRBRisierung – Den Weg durch Trefferlisten der Bibliothek finden**

Knut Hamsun (1859-1952) ist Norwegens berühmtester Romanautor und einer von drei norwegischen Nobelpreisträgern der Literatur. Seine literarische Produktion umfasst rund 30

Romane, ein paar Theaterstücke und Sammlungen von Kurzgeschichten, eine Gedichtsammlung und einige Sachbücher sowie biografische Schriften. Insgesamt beläuft sich Hamsuns Produktion auf rund 40 Werke; das ist eine Bibliografie, die ein Bibliotheksbenutzer einfach durchsuchen können sollte. Dennoch bietet sich dem Bibliotheksbenutzer ein ganz anderes Bild. Im Online-Katalog der Oslo Public Library ergibt die qualifizierte Suche nach "Hamsun, Knut" als Verfasser eine Liste von 568 Treffern (Stand: 7. Mai 2012). Das sind natürlich viel zu viele gelieferte Treffer für einen Autor, der 40 Bücher geschrieben hat. Beachten Sie, dass dies das Ergebnis für eine erweiterte qualifizierte Suche ist. Eine eher typische einfache Suche, welche die meisten Bibliotheksbenutzer versuchen würden, liefert eine noch längere Liste.

Das Problem ist, dass der Online-Katalog nicht zwischen den verschiedenen Werken eines Verfassers und den verschiedenen Versionen eines einzelnen Werkes unterscheidet. In unserer Liste von 568 Treffern, beziehen sich immerhin 63 auf verschiedene Repräsentationen eines Romans: *Hunger*. Dabei handelt es sich wohl um verschiedene Ausgaben, verschiedene Formate und Übersetzungen in unterschiedliche Sprachen.

Die Benutzer müssen natürlich auswählen können, ob sie das Buch, das Hörbuch oder den Film möchten, und sie müssen auswählen können, in welcher Sprache sie das Buch lesen möchten. Für die meisten Benutzer ist es aber eher störend als hilfreich wenn der OPAC sie zwischen mehr als 20 verschiedenen Ausgaben von *Hunger* auf Norwegisch auswählen lässt.

### **Bibliotheksstandards**

Der Bibliothekskatalog konzentriert sich traditionell auf das Beschreiben physischer Objekte. Jede Manifestation eines Buches wird durch einen einzelnen Datensatz repräsentiert, und es gibt keine funktionalen Verknüpfungen zwischen den Datensätzen, die Manifestationen desselben Werkes beschreiben. Ein Bibliotheksbenutzer, der den Online-Katalog nach einem bestimmten Titel durchsucht, mag sich daher in eine Warteliste eintragen, um eine bestimmte Ausgabe eines klassischen Romans auszuleihen, ohne zu bemerken, dass zahlreiche andere Ausgaben desselben Buches schon verfügbar sind. Ein anderer Benutzer mag letztendlich das Buch überhaupt nicht erhalten, wenn er zufällig eine Ausgabe ausgewählt hat, von der keine Exemplare mehr verfügbar sind. Das Projekt PODE hat seine Experimente auf die Hypothese aufgebaut, dass Bibliotheksbenutzer normalerweise daran interessiert sind, einen bestimmten Titel zu finden, nicht eine bestimmte Ausgabe dieses Titels, und dass dieses Interesse besonders typisch ist für belletristische Werke, bei denen verschiedene Ausgaben normalerweise identische Inhalte haben. Aus der Perspektive des Bibliothekssystems sollte der Benutzer im Idealfall die Möglichkeit haben, einen Titel zu bestellen, ohne zwischen verschiedenen Ausgaben wählen zu müssen. Natürlich sollten diejenigen, die sich für die Ausgaben interessieren, weiterhin die Möglichkeit haben, diese zu spezifizieren, aber wieso sollte jeder andere gezwungen sein, zu wählen?

Wie viele betont haben, wurden die derzeitigen Bibliotheksstandards früher als das Netz und die derzeitige Infrastruktur für Produktion, Vertrieb und Nutzung der Metadaten entwickelt. Außerdem wurden die Standards, die Bibliotheksdaten in die elektronische Sphäre überführten, Jahre vor der Erfindung des Entitäten-Relations-Modells und der relationalen Datenbanken entwickelt.

Das bedeutet einige Herausforderungen für die Implementierung von Referenzmodellen wie den FRBR (die Ausgaben von Werken separieren), welche auf der Analyse von Entitäten und Relationen basieren, wobei diese nicht in oder zwischen MARC-Datensätzen implementiert sind.

Das MARC-Format enthält technische Eintragungen und die Logik aus dem Kartenkatalog, für dessen Automatisierung es entwickelt wurde. Die Metadaten des Kartenkatalogs wurden von Menschen gelesen und interpretiert, ein Merkmal, das im MARC-Format fortgesetzt wird, welches die Erstellung von (separaten) Datensätzen vorschreibt, die größtenteils aus Textelementen bestehen, die für Menschen lesbar sind. Dies ist natürlich eine vereinfachte Beschreibung der Gepflogenheiten von Bibliotheks-Metadaten. Der Prozess des Erstellens eines MARC-Datensatzes ist geprägt von einer komplexen Beeinflussung von Katalogisierungsregeln wie AACR2 und ISBD, und die meisten der Hintergründe hinter den Textelementen sind in diesen Regeln zu finden. In einem Web-Kontext möchten wir, dass Maschinen die Daten verarbeiten und für uns interpretieren. Die Textelemente müssen absolut konsistent sein, damit Maschinen die Daten richtig interpretieren und eine nützliche Präsentation der Daten erstellen.

In relationalen Datenbanken und Linked-Data-Umgebungen ist die Best-Practice-Doktrin, Disambiguierung durch eindeutige Identifikatoren zu vermeiden – bzw. Primär-/Fremdschlüssel und URIs zu verwenden. MARC-Datensätze haben keine solchen eindeutigen Identifikatoren, die unseren Indexierungs-Tools und Suchmaschinen geholfen hätten, zwei Verfasser mit demselben Namen zu unterscheiden, oder vielleicht diese Verfasser zusammenzuführen, wenn sie aufgrund von übereinstimmenden Relationen zu denselben eindeutig identifizierten Werken wahrscheinlich eine Person repräsentieren. Diese Art von Herausforderungen setzt sich natürlich fort, wenn wir möchten, dass unsere Maschinen Relationen zwischen Dokumenten, Verfassern und verschiedenen FRBR-Entitäten identifizieren. Welche Bücher repräsentieren Manifestationen einer bestimmten Expression oder eines Werks? Wie identifizieren wir Werke wie Kurzgeschichten, wenn sie nur textlich in einem MARC-Fußnoten-Feld beschrieben sind? Für die meisten Bücher, die nach 1970 veröffentlicht wurden, haben wir ISBN-Nummern, die Manifestationen identifizieren, und wir haben Namen und Titel, über die wir Computer urteilen lassen können, aber ohne konsistente Kataloge und maschinenlesbare Identifikatoren ist dies alles eine schwierige Aufgabe. Es sind zumindest sorgfältige Aufräumarbeiten erforderlich.

### **FRBRisierung von MARC-Datensätzen**

Das Projekt PODE verwendete ein Tool, das von der Technisch-Naturwissenschaftlichen Universität Norwegens für die automatisierte FRBRisierung von MARC-Datensätzen entwickelt wurde (Aalberg, 2006). Das Tool nutzt XSL-Transformationen für Katalog-Exporte in das Format MarcXchange, um die Daten innerhalb von bibliografischen Datensätzen zu sortieren, abhängig davon zu welchen FRBR-Entitäten die Daten gehören.

Datensätze in Bibliothekskatalogen sind hauptsächlich Beschreibungen von Manifestationen; der individuelle Datensatz beschreibt eine bestimmte Ausgabe eines veröffentlichten Werks mit manifestations-spezifischen Informationen wie z. B. Zeit und Ort der Veröffentlichung, physische Beschreibung, ISBN usw. Aber der Datensatz wird auch Informationen enthalten, die sich auf die Expression und das Werk beziehen, zu denen diese Manifestation gehört. Die Daten beispielsweise, die den Namen des Verfassers und den Originaltitel eines Buchs enthalten, sind Informationen, die die Werk-Entität beschreiben, während Informationen über die Sprache des Dokuments und das Format etwas über die Expression aussagen. Das FRBRisierungs-Tool identifiziert, welche Felder sich auf welche FRBR-Entitäten beziehen, und nutzt dies, um jeden Datensatz in einen Werk-Teil, einen Expressions-Teil und einen Manifestations-Teil einzuteilen, mit FRBR-Relationen dazwischen. Wenn das Tool identische Werk-Beschreibungen für zwei verschiedene MARC-Datensätze auswirft, wird angenommen,

dass die Datensätze zwei verschiedene Verkörperungen desselben Werkes beschreiben. Das Tool wird auch Entitäten der Gruppe 2 und 3 erzeugen, wie z. B. Verantwortliche und Themen.

Zunächst ließ das Projekt die ausgewählte Sammlung von Datensätzen (908 Datensätze, welche die Manifestationen von Hamsun und einem anderen norwegischen Autoren, Per Petterson, beschreiben) unbereinigt durch das automatisierte FRBRisierungs-System laufen. Der erste Versuch lieferte Ergebnisse, die weit entfernt von einwandfrei waren. Dies lag hauptsächlich an fehlenden Informationen in den MARC-Datensätzen und an inkonsistenten Katalogisierungs-Gepflogenheiten. Um mit dem Experiment voranzukommen, musste das Projekt eine beträchtliche Anzahl von Datensätzen bereinigen, um Daten zu erhalten, die ausreichend aussagekräftig und konsistent waren.

### **Das Bereinigungsverfahren**

Kurz gefasst bestand das Bereinigungsverfahren hauptsächlich aus dem Identifizieren und Hinzufügen von Original- und Einheitstiteln in Datensätzen, in denen diese fehlten, dem Hinzufügen von Informationen über individuelle Werke und Kurzgeschichten, die in einem Band gesammelt sind, und dem Setzen von Indikatoren, um zwischen spezifischen Werktiteln und unspezifischen Titeln zu unterscheiden. Außerdem wurde einige Zeit darauf verwendet, Tippfehler und [andere] Fehler zu korrigieren. Insgesamt wurden annähernd 60 Stunden aufgewendet, um die Produktionen von Knut Hamsun und Per Petterson zu bereinigen, einschließlich der Zeit, die dafür gebraucht wurde, die Regeln für die Korrekturen festzulegen.

Eine der Hauptaufgaben in dem Bereinigungsverfahren war es, Datensätze für übersetzte Werke zu identifizieren und zu verbessern, die entweder keinen norwegischen Originaltitel hatten oder den Originaltitel nur in für Menschen lesbaren Fußnoten bereitstellten. Eine andere Arbeit befasste sich mit dem Hinzufügen von Einheitstiteln in Fällen, in denen sich die Titel nicht-übersetzt Titel vom Originaltitel unterschieden. Aufgrund der norwegischen Rechtschreibreformen wurden einige von Hamsuns Werken zu unterschiedlichen Zeiten unter verschiedenen Titeln veröffentlicht. Zum Beispiel wurde die Kurzgeschichte *Paa tourney* (Originaltitel) auch in den Schreibweisen *På tourné* und *På turné* veröffentlicht.

Eine weitere Aufgabe war es, Indikatoren zu setzen, um spezifische echte Werktitel von unspezifischen Titeln zu trennen. Unspezifische Titel kommen normalerweise in Werksammlungen eines Verfassers vor, in denen der Inhalt nicht vom Verfasser zusammengestellt und die Veröffentlichung ebenfalls von jemand anderem betitelt wurde. Veröffentlichungen wie diese sollten in einer FRBRisierten Bibliografie nicht als Werke aufgeführt werden, obwohl die einzelnen, in der Sammlung enthaltenen Romane oder Kurzgeschichten es sein sollten.<sup>1</sup>

### **Beispiele:**

a) 245 10 †aGrowth of the soil †cKnut Hamsun ; translated from the Norwegian by W. Worster

---

<sup>1</sup> Obwohl Relationen zwischen MARC-Feldern implizit durch einen Datensatz ausgedrückt werden, ist es nicht einfach, Beziehungen zwischen Feldern in den Fällen eindeutig und explizit auszudrücken, in denen ein Dokument aus verschiedenen Werken besteht.

(Englische Übersetzung eines Originalwerks von Hamsun. Erster Indikator auf eins gesetzt bedeutet, dass es sich um einen spezifischen Werktitel handelt.)

b) 245 00 †aTales of love and loss †cKnut Hamsun ; translated by Robert Ferguson

(Englische Sammlung von Kurzgeschichten, die ursprünglich nicht zusammen veröffentlicht worden waren. Erster Indikator auf null gesetzt bedeutet, dass es sich um einen unspezifischen Titel handelt.)

Weitere Informationen über das Bereinigungs-Verfahren siehe Anhang.

## Ergebnisse

Während der erste Versuch der FRBRisierung 149 Werke von Hamsun identifizierte, wurde die Liste nach dem Bereinigen der MARC-Datensätze weiter reduziert auf eine Anzahl von 84. Am Beispiel von Per Petterson beobachteten wir einen Anstieg der Anzahl von Werken von 14 auf 41, bedingt durch die Ergänzung von Titeln einzelner Kurzgeschichten und Essays zu einigen Katalogdatensätzen. Die resultierende Liste von Werken entsprach beinahe genau den aktuellen Bibliografien der zwei Autoren. Einzige Ausnahme war ein Werk von Hamsun, das in der Hamsun-Bibliografie der norwegischen Nationalbibliothek aufgeführt war, in der Sammlung der Stadtbibliothek von Oslo (Oslo public library) jedoch fehlte. Das Bereinigungs-Verfahren lieferte uns damit unbeabsichtigt eine Methode, in Bibliothekssammlungen fehlende Werke zu identifizieren. Dies ist sonst eine langwierige Prozedur, wenn man es mit langen Listen von Hunderten von Manifestationen zu tun hat.

## RDFisierung

Die FRBRisierten Datenbestände wurden mittels XSLT<sup>2</sup> und einer Konkordanz zwischen MARC-Feldern und RDF-Prädikaten, der im Projekt entwickelt wurde, in RDF konvertiert. Die Konkordanz verwendete hauptsächlich bekannte Vokabulare und Ontologien wie die Dublin-Core-Metadaten-Terme, die Bibliographic ontology, Core FRBR, FOAF und SKOS. Aber das Projekt bildete auch einige spezifischere Sub-Property, um unsere Daten genauer auszudrücken als es diese Vokabulare zulassen. Später stellte das Projekt fest, dass die RDA-Vokabulare<sup>3</sup> Prädikate enthalten, die viele der auszudrückenden, eher bibliotheksspezifischen Informationen bereits abdecken. In späteren Revisionen der Konkordanz haben einige von ihnen unsere eigenen Prädikate ersetzt.

Die vollständige Konkordanz finden Sie hier: <http://www.bibpode.no/blogg/?p=1573>.

Sobald die Daten in RDF konvertiert waren, wurden sie mit Verknüpfungen zu anderen Datenbeständen angereichert. Werke wurden mit Instanzen von DBpedia und Project Gutenberg verknüpft, während Personen mit DBpedia und VIAF verknüpft wurden. Um eine Liste von Werken chronologisch sortieren zu können, ergänzte das Projekt Informationen über die erste Veröffentlichung zu den Werk-Instanzen. Diese Informationen sind nicht einfach aus einem MARC-Datensatz zu extrahieren, wenn sie überhaupt enthalten sind. Mit diesem neuen angereicherten Datenbestand war das Projekt in der Lage, eine Web-Applikation zu entwickeln, die dem Endnutzer erlaubte, die vollständige Sammlung der Bücher dieser beiden Autoren in der Bibliothek zu durchsuchen, indem sie aus einer kurzen

---

<sup>2</sup> In späteren Arbeiten dieser Art haben wir anstatt XSLT Ruby-Skripts mit einer YAML-Mapping-Datei benutzt. Details s. <https://github.com/bensinober>

<sup>3</sup> <http://rdvocab.info/>

Liste von Werken auswählen konnten anstatt eine flache Liste mit Hunderten von Manifestationen zu durchsuchen. Außerdem konnte die Applikation dem Endnutzer relevante Informationen über Autoren aus DBpedia sowie Verknüpfungen zu digitalen Volltext-Versionen im Project Gutenberg liefern.

Es wurde eine einfache Web-Applikation entwickelt, die dem Endnutzer erlaubte, diesen Teil der Bibliothekssammlung nach FRBR-Entitäten gebündelt zu durchsuchen, mit zusätzlichen Informationen, die von externen Quellen zur Verfügung gestellt und durch die Verknüpfung von Daten zugänglich gemacht wurden.<sup>4</sup>

Die RDF-Repräsentation der beiden Autoren lieferte uns sogleich einen besser zu verarbeitenden Datenbestand. Die Erfahrung war so positiv, dass wir beschlossen, den gesamten Bibliothekskatalog zu konvertieren, um zu sehen, ob dies die Antwort auf unseren grundlegenden Wunsch war: bessere Dienste für unsere Benutzer zu schaffen.

### **Linked Data und RDF als Basis für neue Dienste**

In Verbindung mit den Projekten "Book recommendations" und "Active shelves" an der Oslo Public Library wurde viel Arbeit in die Konvertierung des Bibliothekskatalogs in das RDF-Format investiert. Der Aufwand, den wir jetzt in diesem Bereich betreiben haben, ist wesentlich detaillierter und gründlicher und bietet uns wesentlich mehr Möglichkeiten als die Experimente, die wir seinerzeit im Projekt PODE gemacht haben. Der Katalog der Oslo Public Library besteht aus ungefähr 420.000 Datensätzen, zurückdatierend bis auf die späten 1970er Jahre.

RDF ist ein Format, das Bibliothekskatalogdaten wirklich maschinenlesbar macht. Während Maschinenlesbarkeit im Zusammenhang mit dem MARC-Format bedeutet, dass Computer die *Zeichen* eines Katalogdatensatzes lesen, speichern und verarbeiten können, bedeutet Maschinenlesbarkeit in RDF, dass ein Computer die tatsächliche *Bedeutung* oder Semantik unserer Daten lesen kann. Kurz gefasst geht es bei der Konvertierung von Bibliotheksdaten in RDF um die „Übersetzung“ von Katalogdatensätzen in Sets von einfachen Aussagen, in denen wir sowohl für die Dinge, über die wir etwas aussagen wollen, als auch für die Informationen, die wir über diese Dinge liefern wollen, eindeutige Identifikatoren verwenden.

Sind wir in der Lage, neue Gedanken zu denken, ohne unsere alten Fähigkeiten zu vergessen? Wir sind nicht die erste Bibliothek auf der Welt, die ihre Bibliotheksdaten in RDF konvertiert. Aber während sich viel von der auf diesem Feld geleisteten Arbeit auf das Extrahieren und Konvertieren eines kleinen Untersets essentieller Informationen aus den Katalogdatensätzen konzentriert, haben wir versucht, „alles“ mitzunehmen. Die Leistung von Bibliothekskatalogisierern ist sorgfältig und unschätzbar und jeder, der in einer Bibliothek gearbeitet hat, weiß, dass es entscheidend ist, dass wir mehr Informationen über unsere Dokumente abrufen können als nur den Titel, den Namen des Verfassers, das Thema und die ISBN. Wir haben versucht, so viele Informationen wie möglich aus den MARC-Datensätzen zu erhalten, und sie so präzise und korrekt wie möglich auszudrücken. Wir haben uns, Feld für Feld, systematisch unseren Weg durch den norwegischen Standard NORMARC erschlossen und haben unsere Entscheidungen darüber getroffen, wie die Daten zu interpretieren sind und wie die Inhalte in RDF ausgedrückt werden könnten.

---

<sup>4</sup> <http://bibpode.no/linkedauthors/>

Unser vollständiges Konversions-Skript, das binäres MARC über ein YAML-Mapping nach RDF konvertiert, finden Sie hier: <https://github.com/digibib/marc2rdf>.

Gleichzeitig ist es wichtig, dass eine RDF-Version des Katalogs nicht einfach ein weiterer Datenbestand wird, der genau dasselbe nur in einer anderen Sprache aussagt. Die Katalogisierungsregeln und das MARC-Format wurden einige Jahrzehnte zuvor entwickelt, um die bestmöglichen Ergebnisse innerhalb bestimmter Einschränkungen zu erzielen, und einige dieser Einschränkungen sind heute nicht mehr so relevant wie sie es waren, als diese Standards entwickelt wurden. Außerdem zeichnen sich viele der Regeln durch die Doppelfunktion aus, Dokumente sowohl zu beschreiben als auch zu lokalisieren. Wenn die alten Regeln und die alten Ideen in das neue Format gebracht werden, werden die alten Einschränkungen mit ihnen einhergehen. Das RDF Format bietet uns Möglichkeiten, die MARC fehlen, und es ist wichtig, dass wir uns dies zunutze machen.

Ein spezifisches Beispiel: was bedeutet es, wenn ein Katalogdatensatz aussagt, dass Karl Marx der Autor vom "Manifest der Kommunistischen Partei" ist, wohingegen Friedrich Engels der Co-Autor ist? Diese Unterscheidung zwischen den Funktionen hat nicht wirklich etwas mit den tatsächlichen Funktionen der beiden Personen zu tun. Eigentlich haben wir überhaupt kein richtiges Verfasserfeld in MARC; es gibt das Feld „Haupteintrag“. Die Funktion dieses Felds ist, etwas darüber auszusagen, wo sich ein Dokument in einer Sammlung befinden sollte. Ein physisches Dokument kann nur an einem Ort sein, deshalb kann es nur einen Haupteintrag haben. Marx' Name wird auf der Haupttitelseite vor Engels' Namen aufgeführt; deshalb erhält er allein die Ehre, als Autor des Buches genannt zu werden, während Engels' Anteil auf das des Co-Autors reduziert wird.

Wenn wir diese Information in RDF ausdrücken, brauchen wir uns nicht um solcherlei künstliche und nicht-funktionale Unterscheidungen zu kümmern. Karl Marx ist ein Autor des Buches und Friedrich Engels ist es ebenso. Es gibt keinen Konflikt zwischen diesen Aussagen.

Ein weiteres Beispiel ist es, wenn Informationen in Katalogdatensätzen mehrdeutig und kontextabhängig sind. Nehmen Sie diese Nebeneintragung als ein Beispiel:

```
*700 0 $a King, Stephen
      $d 1947-
      $e forf.
      $j am.
      $t Rita Hayworth and Shawshank Redemption
```

Was ist in diesem Fall die Verbindung zwischen der Novelle "Rita Hayworth and Shawshank Redemption" und dem katalogisierten Dokument? Das hängt ganz vom Kontext ab. Wenn das katalogisierte Dokument ein Buch ist, bedeutet die Aussage sehr wahrscheinlich, dass die Novelle ein Teil des Dokuments ist. Wenn das katalogisierte Dokument eine DVD-Disk ist, bedeutet es wahrscheinlich, dass der Film auf der Geschichte basiert, die in der Novelle erzählt wird. Eine korrekte Interpretation der Daten erfordert, dass man etwas über Texte und Filme weiß und darüber, wie sie zusammenhängen können. Mit anderen Worten: sie erfordert einen menschlichen Leser.

Beide Beispiele sagen uns etwas über die Einschränkungen dessen, was wir im MARC-Format ausdrücken können. Das erste Beispiel sagt etwas darüber aus, wie Konzepte wie Haupt- und Nebeneintragungen nicht in erster Linie etwas sind, das dazu entwickelt wurde,

Bücher, Filme oder Musik zu beschreiben. Sie sind dazu konzipiert, physische Sammlungen zu organisieren und die Wege auszuwählen, die ein Bibliotheksbenutzer für die Suche in einem Kartenkatalog haben sollte. Das zweite Beispiel hat mit dem Grad der Maschinenlesbarkeit zu tun. Wenn die korrekte Interpretation der Daten von menschlichem Allgemeinwissen abhängt, können die Daten kaum als wirklich maschinenlesbar bezeichnet werden.

SPARQL ist die Abfragesprache, die wir benutzen, um RDF-Daten zugänglich zu machen. Wir können jede gewöhnliche Katalogsuche mit der Sprache SPARQL ausführen, außerdem erhalten wir einiges zusätzlich.

Die CCL-Bibliothekssuche bietet uns viele Wege, Suchanfragen zu kombinieren, um erweiterte und spezifische Suchen durchzuführen. Aber es gibt eine Beschränkung, um die Sie nicht herumkommen: wie auch immer Sie Ihre Suche konstruieren, was Sie als Ergebnis erhalten, sind Katalogdatensätze. Sie können jede Frage stellen, die Sie möchten, aber nur solange sie mit den Worten „Welche Bücher...“ beginnen.

Sie können nicht nach einer Liste mit Themen fragen, die in Dokumenten in einer bestimmten Sprache enthalten sind, oder nach einer Liste von Autoren, die über ein bestimmtes Thema geschrieben haben. Um Fragen wie diese zu beantworten, müssten Sie die *Bücher* finden, welche die Bibliothek in dieser Sprache oder zu diesem Thema hat, und dann den mühsamen Vorgang des Durchsuchens dieser Liste durchführen, um die verschiedenen Themen oder Autoren zu identifizieren.

Eine weitere Einschränkung bei traditionellen Bibliothekssuchanfragen ist, dass sie keine unbekanntentitäten enthalten können. Sie können beispielsweise eine CCL-Suche nicht so konstruieren, dass sie alle Bücher eines Verfassers eines bestimmten Titels liefert. Sie müssen den bekannten Titel nachschlagen, den Namen des Verfassers herausfinden und dann eine neue Suchanfrage nach Büchern dieses Verfassers starten. Das mag nicht sehr schwierig sein, aber was wenn die Informationen, die Sie benötigen, von mehr als einer unbekanntentität abhängen? Es ist hoffnungslos, eine Frage wie z. B. „Welche Romane kann ich in der Bibliothek finden, die ich auch als Adaption auf DVD finden kann?“ mit den herkömmlichen Bibliotheksrecherche-Tools beantworten zu können, es sei denn, Sie haben sehr viel Zeit. Mit der Unterstützung von SPARQL können Sie abfragen:

Suche alle Paare von Dokumenten, doc-1 und doc-2, so dass:

doc-1 ist ein Film auf DVD  
doc-2 ist ein Buch, das einen Roman enthält  
doc-1 ist basiert auf dem Werk X  
doc-2 ist eine Manifestation des Werkes X

## Schlussbemerkung

Wir werden unsere Arbeit an der RDF-Repräsentation unseres Katalogs fortsetzen. Der nächste Schritt wird sein, eine Suchfunktion zu implementieren, um zu sehen, ob die Ergebnisse besser sein werden als die, die wir zurzeit aus unserem OPAC erhalten. Sie können unsere Fortschritte unter: <http://digital.deichman.no/> verfolgen.

## Anhang – Bereinigung und Korrekturen

Die Hauptaufgabe war es, sicherzustellen, dass alle übersetzten Datensätze (sowohl bedingt durch Fremdsprachen als auch durch norwegische Sprachreformen) den Originaltitel des Werks im Feld 240 enthielten. Wenn die 240 fehlte, wurde sie automatisch ergänzt, basierend auf den Informationen im Fußnotenfeld 574, einem NORMARC-spezifischen Feld, das Informationen zum Originaltitel enthält. Wenn dieses Fußnotenfeld fehlte oder die automatische Konversion fehlschlug, wurde der Titel manuell hinzugefügt.

Ein weiterer zeitaufwändiger Teil der Arbeit war es, tatsächliche (Original-)Titel spezifischer Werke in den Feldern 245 und 740/700 $\ddagger$ t zu identifizieren und festzustellen sowie die entsprechenden Indikatoren gemäß den Korrekturregeln zu setzen.

Bei der Behandlung des zweiten Indikators im Feld 740 entschied das Projekt, den Wert 2 nur zu verwenden, wenn das Feld einen spezifischen Werktitel enthält. Für alle anderen Titel wird der Indikator auf 0 gesetzt, auch wenn sie analytische Einträge sind. Auch wenn diese Praxis nicht der Konvention der Verwendung von 700 $\ddagger$ a +  $\ddagger$ t für analytische Titeleinträge entspricht, wurde hierüber als der effizienteste Ansatz entschieden, um analytische Werke, die in eindeutiger Weise veröffentlicht wurden, zu entdecken.

Basierend auf den Ergebnissen des ersten Versuchs der FRBRisierung umfassten die Korrekturen in den 908 Datensätzen für Hamsun und Petterson:

- Korrigieren des Sprachencodes in 008 in 5 Datensätzen
- Ergänzen des Einheitstitels (oder “Originaltitels“ in präziser Übereinstimmung mit der NORMARC-Terminologie) in den 240-Feldern in 85 Datensätzen und Korrigieren von Tippfehlern in bestehenden 240-Feldern in 24 Datensätzen
- Korrigieren von Tippfehlern in 245 $\ddagger$ a (oder fehlerhafte ISBD-Syntax) in 6 Datensätzen
- Korrigieren des ersten Indikators in 245: vor der Korrektur hatten 137 Datensätze den Indikator 1=0 oder Leerzeichen, während der Indikator 1=1 in 774 Datensätzen verwendet wurde. Nach der Korrektur war die Verteilung 263 – 651.
- Korrigieren der 700-Felder. In den originären Datensätzen fanden wir 948 700 $\ddagger$ a und 545 700 $\ddagger$ t Felder. In den korrigierten Datensätzen sind die Zahlen reduziert auf entsprechend 917 mal 700 $\ddagger$ a und 481 mal 700 $\ddagger$ t. Die Veränderung beruht auf einer systematischeren Verwendung von 740-Feldern in allen Datensätzen, die denselben Verfasser haben (welcher in der 100 aufgeführt ist).
- Ändern des zweiten Indikators im Feld 700, um zu verdeutlichen, ob ein Eintrag ein eindeutiges Werk ist oder nicht.

## Referenzen

Aalberg, T. (2006). A Tool for Converting from MARC to FRBR. *ERCIM News*, (66). In: [http://www.ercim.eu/publication/Ercim\\_News/enw66/aalberg.html](http://www.ercim.eu/publication/Ercim_News/enw66/aalberg.html)