



UNIMARC und FRBR – können wir beides haben?

Trond Aalberg

Norwegian University of Science and Technology
Trondheim, Norway

Jan Pisanski and

Maja Žumer

University of Ljubljana
Ljubljana, Slovenia

Übersetzung:

*Susanne Oehlschläger und Andrea Köpcke
Deutsche Nationalbibliothek*

Meeting:

**187 — Advancing UNIMARC: alignment and innovation — IFLA
UNIMARC Programme (UNIMARC)**

Abstract:

Mit dem Modell der Funktionalen Anforderungen an bibliografische Datensätze (FRBR) wurde der Bibliothekscommunity ein neuer konzeptioneller Bezugsrahmen für bibliografische Daten vorgestellt. Auf der einen Seite baut das Modell auf aktuellen bibliografischen Standards und auf der Praxis auf, aber auf der anderen Seite repräsentiert es einen dramatischen Wandel: weil es ein Modell für bibliografische Informationen ist, die als Entitäten und Beziehungen ausgedrückt werden. Die Angleichung des UNIMARC-Standards an die FRBR ist wichtig, um die Implementierung des Modells in Bibliothekssystemen, den Austausch von bibliografischen Daten, die auf dem FRBR-Modell basieren, und die Weiterverwendung von UNIMARC-Daten als maschinenlesbare Informationen für ein breites Spektrum von bedeutungssensitiven Anwendungen zu erleichtern. IN diesem Vortrag analysieren wir die Fähigkeit von UNIMARC, einfache und komplexe FRBR-Strukturen zu enkodieren und diskutieren verschiedene Alternativen, diese zu enkodieren und versuchen, die am besten geeignete Lösung zu identifizieren.

Hintergrund

Projekte und Prototypen, die sich auf das Herauslösen von FRBR-Entitäten und Beziehungen aus MARC-basierten Datensätzen konzentrieren, haben viele Möglichkeiten und Probleme der Verwendung von MARC-Formaten in diesem neuen Kontext gezeigt. Das Werk, die Expression, durch die es realisiert ist, und damit in

Beziehung stehende Personen und Körperschaften können mit einer gewissen Erfolgsquote durch Aufbereitung bestehender Datensätze abgeleitet werden, aber die meisten Interpretationen werden Fehler produzieren, weil die Daten nicht dafür konzipiert sind, so bearbeitet zu werden (Aalberg, 2006, Manguinhas, Freire, Borbinha, 2010). Das Problem der richtigen Identifizierung von FRBR-Entitäten ist allgemein bekannt (z. B. Hegna, Murtooma, 2002, Hickey, O'Neill, 2005, Pisanski, Žumer, Aalberg, 2009) und entsteht typischerweise durch die Katalogisierungspraxis und Datenanforderungen, die nicht im Einklang mit dem Modell sind.

Ein anderes Problem, dem wenig Aufmerksamkeit geschenkt wurde, das aber für die Anpassung der FRBR an MARC-Standards noch grundlegender ist, ist das Problem der sauber definierten Darstellung von komplexeren Entitäten-Beziehungs-Strukturen in MARC. Datensätze, die Ressourcen mit vielen verschiedenen Teilen beschreiben, möglicherweise kombiniert mit vielen Personen/Körperschaften, Schlagworteinträgen und anderen in Beziehung stehenden Werken sind leider im Allgemeinen derart repräsentiert, dass sie informativ und für den Menschen lesbar sind, wenn sie angezeigt werden, aber meist unmöglich korrekt und konsistent durch Software interpretiert werden können. Einerseits beschränkt das den Wert und die Tauglichkeit dieser Datensätze für Bibliothekssysteme, die auf den FRBR basieren, andererseits handelt es sich um ein Problem, das den Bestand als Ganzes betrifft. Der Katalog wird unvollständig bleiben, wenn wir solche Datensätze außen vor lassen und wenn wir versuchen, solche Datensätze zu interpretieren, werden wir wahrscheinlich eine Reihe von Fehlern einbringen. Einen Formalsachtitel als einzelnes, eigenes Werk zu interpretieren, produziert ein bedeutungsloses Werk und ignoriert die tatsächlichen Werke, die für die Endnutzer wichtig sein können. Alternativtitel oder Titel von in Beziehung stehenden Werken als eigenen Inhalt zu interpretieren, stellt eine weitere Fehlerquelle dar. Selbst wenn die Zahl der falsch interpretierten Datensätze klein ist, werden diese Fehler typischerweise eskalieren, je weiter wir in der Abstraktionsebene gehen. Zur Liste von Werken, die unterschiedliche Lösungen für produktive Verfassernamen wie Agatha Christie identifizieren können, gehört normalerweise eine große Zahl von falsch identifizierten Werken. Einer der Gründe dafür ist, dass die Datensätze nicht die korrekte FRBR-Struktur für den Inhalt haben.

MARC und FRBR-Strukturen

Während der frühen Jahre von MARC entwickelten sich verschiedene Formate, die den nationalen Anforderungen und der jeweiligen Katalogisierungspraxis geschuldet waren. Heute dominieren UNIMARC und MARC 21 die Welt der bibliografischen Informationen, weil Bibliotheken die Vorteile der Anwendung von einem dieser international anerkannten Formate im Gegensatz zu nationalen oder Anbieter-spezifischen Formaten anzuerkennen scheinen. Das Format ISO 2709, das die Syntax und die zugrundeliegende Struktur aller MARC-Formate definiert, ist konzipiert, um die vielen Aspekte einer Ressource zu beschreiben, unter Verwendung einer flachen Struktur von Kontroll- und Datenfeldern. Jedes Datenfeld hat darüber hinaus Indikatoren und Unterfelder und Datensätze, die auf dem Standard basieren, können ziemlich komplexe Informationen enthalten, aber es gibt keinen Standard-Formalismus zur Darstellung einer Informationseinheit, die ein Set von miteinander in Beziehung stehenden Entitäten beschreibt. Allerdings wurden verschiedene Methoden und Verfahren innerhalb des Formalismus des Standards entwickelt, um die Kodierung von komplexeren bibliografischen Informationen sowohl in UNIMARC als auch in MARC 21 zu erleichtern.

Einige dieser Methoden sind relevant für die Kodierung von FRBR, weil sie zu Datensätzen führen, die konsistent als ein Set von miteinander in Beziehung stehenden FRBR-Entitäten interpretiert werden. Trotz einer beabsichtigten Interoperabilität durch Konversion zwischen diesen Formaten, gibt es Unterschiede, die im Hinblick darauf relevant sind, wie Formate verwendet werden können, um FRBR-Strukturen zu kodieren.

UNIMARC basiert auf Feldern, die anders nummeriert und benannt sind als die äquivalenten Felder in MARC 21, und hat eine Gruppierung in Blöcke, die eine andere Organisation der Informationen repräsentiert. In UNIMARC finden wir getrennte Blöcke für beschreibende Informationen, in Beziehung stehende Titel und geistige Verantwortlichkeit. Die korrespondierenden Informationen werden in MARC 21 in den Haupteintragungsfeldern, den Titel- und titelbezogenen Feldern und in den Nebeneintragungsfeldern gefunden. Einige Datenfelder in UNIMARC sind weniger mehrdeutig als die korrespondierenden Felder in MARC 21. In UNIMARC werden immer unterschiedliche Felder für Namen und in Beziehung stehende Titel verwendet, während z. B. das Feld 700 in MARC 21 Nebeneintragung Personennamen nur einen Personennamen enthalten kann oder einen Titel, in dem der Personennamen Teil der Titelidentifikation ist. Der engere Geltungsbereich dieser Felder in UNIMARC hat auch Einfluss darauf, wie bestimmte FRBR-Strukturen im Datensatz beschrieben werden. Der 5xx-Block von UNIMARC für den in Beziehung stehenden Titel sieht nur Alternativtitel vor, die mit einer zu katalogisierenden Manifestation in Verbindung stehen, und wird immer ein Titel für die Manifestation sein, die die einzelne enthaltene Expression oder das Werk realisiert. Die Titel von anderen in Beziehung stehenden Werken oder Expressionen müssen unter Verwendung von Verknüpfungsfeldern katalogisiert werden. Ein Titel in einem Nebeneintragungsfeld würde in der Praxis entweder eine analytische Titeleintragung sein oder ein Titel, der entweder ein Alternativtitel für das enthaltene Werk/die enthaltene Expression oder für ein in Beziehung stehendes Werk/eine in Beziehung stehende Expression ist. Der zweite Indikator kann verwendet werden, um anzugeben, dass das Feld ein analytischer Eintrag ist, aber der Zweck des Feldes im Kontext der FRBR ist unspezifisch, wenn es keinen Wert für diesen Indikator gibt. Weiterhin erlaubt UNIMARC wiederholbare Hauptsachtitel und Verfasserangaben, während MARC 21 keine wiederholbaren Unterfelder für die Titelangabe hat und sich für die weitere Strukturierung innerhalb der Felder auf die ISBD-Zeichensetzung verlässt. Beide Formate haben Felder für unformatierte Anmerkungen zum Inhalt, aber MARC 21 hat ein zusätzliches Feld für formatierte Anmerkungen mit wiederholbaren Unterfeldern.

Beide Formate haben Verknüpfungseintragungsfelder zur Beschreibung von in Beziehung stehenden Ressourcen (entweder enthalten oder extern zu den Manifestationen), aber es gibt wesensbedingte Unterschiede bei der Verwendung dieser Felder. Im Hinblick auf die FRBR scheint UNIMARC konzipiert zu sein, um eine einzelne enthaltene Expression zu beschreiben, und bevorzugt die Verwendung von Verknüpfungseintragungsfeldern für strukturierte Beschreibungen von mehreren enthaltenen Expressionen oder in Beziehung stehenden Werken/Expressionen. In MARC 21 kann dies durch die Verwendung von Nebeneintragungen erreicht werden, und Verknüpfungseintragungsfelder sind für diesen Zweck weniger relevant. Ein weiterer Hauptunterschied zwischen den Verknüpfungsfeldern ist die Methode der eingebetteten Verlinkung, die in UNIMARC zu finden ist. Dies ermöglicht eine flexiblere und präzisere Beschreibung der Zielentitäten unter Verwendung einer geeigneten

Kombination von einfachen Datenfeldern als die Verwendung des definierten Sets von verlinkten Unterfeldern.

Beide Formate haben Unterfelder für kontrollierte Funktionsbezeichnungs-codes zur Beschreibung der Funktion der Entitäten der Gruppe 2. In UNIMARC scheint dies systematischer verwendet zu werden. In MARC 21 gibt es ein zusätzliches Feld für die Funktionsbezeichnung. MARC 21 wurde kürzlich um Unterfelder für Beziehungsinformationen in den Nebeneintragungsfeldern erweitert. Das Unterfeld \$8 zur Feldverknüpfung ist einzigartig für MARC 21 und kann verwendet werden, um ein Datenfeld mit anderen spezifischen Datenfeldern zu verknüpfen. Es gibt zusätzlich weitere kleinere Unterschiede wie beispielsweise, ob bestimmte Unterfelder wiederholbar sind oder nicht. Obwohl beide Formate äquivalente Informationen über dieselbe Manifestation aufnehmen können, haben die Unterschiede einen gewissen Einfluss auf die Beschreibung.

Eine gewisse Unterstützung für solche Strukturen kann sowohl in UNIMARC als auch in MARC 21 gefunden werden, beispielsweise die Verknüpfungsfelder, die Verwendung von Indikatoren, Felder, die separate und individuelle Entitäten repräsentieren können (oder nicht), Funktionsbezeichnungs-codes usw. Allerdings erfüllen diese Eigenschaften nur zum Teil die Anforderungen und eine systematische Herangehensweise ist notwendig, um die gegenwärtigen Möglichkeiten und künftigen Anforderungen zur Repräsentation des FRBR-Modells unter Verwendung von MARC zu identifizieren. Dennoch – auch wenn die Formate eine explizitere Kodierung ermöglichen würden, die theoretisch dem FRBR-Modell entspricht, bleibt immer noch eine Vielzahl von Katalogisierungsanwendungen, die an verschiedene Anforderungen und Materialarten angepasst wurde (z. B. Mcgrath, Bisko, 2008).

Komplexe FRBR-Strukturen in UNIMARC

Im Folgenden werden wir eine pragmatische Herangehensweise verfolgen, um die Fähigkeit von UNIMARC zur Kodierung komplexer FRBR-Strukturen zu analysieren. Zunächst präsentieren wir eine einfache Struktur und dann geben wir ein paar Beispiele für komplexere Strukturen, diskutieren verschiedene Alternativen, um diese zu kodieren, und versuchen, die geeignetste Lösung zu identifizieren. Der zugrunde liegende Formalismus in unserer Analyse ist, dass die FRBR das bibliografische Universum in einem Katalog als Graph¹ unterschiedlich typisierter Knoten (FRBR-Entitäten) mit typisierten Links (Beziehungen) dazwischen modellieren. Jeder bibliografische Datensatz wird immer einen bestimmten Teil dieses Universums repräsentieren. Die einzige Anforderung in unserer Analyse ist, dass es möglich sein soll, Algorithmen zu schreiben, die automatisch eine Sammlung von Datensätzen interpretieren und einen Graph von FRBR-Entitäten bilden, der mit dem beabsichtigten FRBR-Graph, der durch die Datensätze beschrieben wird, korrespondiert.

Im FRBR-Modell sind alle Entitäten individuelle, aber miteinander in Beziehung stehende Untersuchungseinheiten, die wir verarbeiten können wollen. Die Anforderungen an die Verarbeitung jedes MARC-Datensatzes als ein Set von miteinander in Beziehung stehenden Entitäten kann in den folgenden Fragen zusammengefasst werden:

¹ Eigentlich ein Set von Graphen, weil es Unterabteilungen des bibliografischen Universums geben kann ohne bekannte Beziehungen zum Rest.

- Was ist das Set an Entitäten, das durch jeden Datensatz beschrieben wird?
- Welche sind die Merkmale, die jede Entität beschreiben?
- Was ist die Identität jeder Entität?
- Wie sind diese Entitäten miteinander verbunden?

Entitäten

Eine beachtliche Zahl von Datensätzen in einem Bibliothekskatalog wird eine ziemlich einfache FRBR-Struktur haben und Entitäten, die implizit oder explizit beschrieben sind, können entweder aus den Tags und Codes abgeleitet werden oder durch die Interpretation des Modells selbst. Die meisten Datensätze beschreiben typischerweise eine einzelne Veröffentlichung, die im FRBR-Modell als Manifestation angesehen wird. Nach dem Modell wissen wir, dass diese Manifestation mindestens eine Expression verkörpert, die die Realisierung nur eines Werks darstellt. Personen und andere Entitäten der Gruppe 2 können glücklicherweise aus spezifischen Feldern identifiziert werden und sind deshalb relativ einfach zu erkennen. Außerdem können wir andere Tags und Codes verwenden, um weitere individuelle Werke und Expressionen zu identifizieren, wie beispielsweise die, die durch Schlagworteinträge beschrieben sind.

Leider entspricht dieser einfache, standardmäßige Blick auf einen Datensatz nicht immer der tatsächlichen FRBR-Struktur, welche der Datensatz repräsentiert. Manifestationen können mehrere Expressionen verkörpern und es kann andere Entitäten in der Beschreibung geben, die nur in Verbindung stehen, aber nicht enthalten sind. Die Anwendung der standardmäßigen und einfachen Interpretation, dass jede Manifestation eine einzige Expression verkörpert, ist nicht nur eine grobe Vereinfachung, sondern sie bedeutet auch, dass wir das FRBR-Modell nicht so verwenden, wie es gedacht ist. Allerdings stellt die Identifizierung des richtigen Sets der FRBR-Entitäten in einem Datensatz eine Herausforderung dar. Enthaltene Expressionen können in Anmerkungen zum Inhalt beschrieben sein und es gibt Felder, die in einem Kontext ein individuelles Werk identifizieren können, in einem anderen jedoch nicht.

Merkmale

Die Informationen im Datensatz werden, mit Ausnahme der Felder, die den Datensatz selbst beschreiben, verschiedene Eigenschaften der Entitäten beschreiben. Für die einfache Struktur einer Manifestation, die eine einzelne Expression verkörpert, ist das ein Problem, das auf ein Mapping reduziert werden kann, bei dem jedes Feld oder Unterfeld einem bestimmten Entitätentyp zugeordnet werden kann. Die Bedeutung und Verwendung spezifischer Felder kann im Hinblick darauf, welchen Entitätentyp sie beschreiben, mehrdeutig sein, aber das ist eher ein semantisches als ein strukturelles Problem. Andererseits kann es ein strukturelles Problem sein, wenn der Datensatz eine Manifestation beschreibt, die zwei Expressionen verkörpert. Wenn wir einen Wert für die Sprache haben, können wir vermuten, dass er sich auf beide Expressionen bezieht. Wenn wir mehrere Werte für die Sprache haben, gelten dann beide Sprachen für beide Expressionen oder sind die Expressionen in verschiedenen Sprachen? Welche Sprache wird im letztgenannten Fall in welcher Expression verwendet?

Identität von Entitäten

Bibliografische Informationen basieren auf der Verwendung von beschreibender Identifizierung, wobei die Untersuchungseinheiten durch ein Set von Merkmalen identifiziert werden, das nach Katalogisierungsregeln erstellt wird. Das ist im

Wesentlichen nicht viel anders als die Verwendung von Schlüsselmerkmalen in relationalen Datenbanken, aber bibliografische Informationssysteme sind „offene Weltsysteme“, in denen es möglich ist, sich auf jede bekannte bibliografische Entität zu beziehen. Selbst wenn es uns möglich ist, das richtige Set an Entitäten zu identifizieren und wir alle Werte für die Merkmale dem richtigen Entitätentyp zugeordnet haben, haben wir leider oft keine ausreichenden Informationen, um die Entitäten konsistent automatisiert zu identifizieren. Die beschreibende Identifizierung, die bei der Katalogisierung verwendet wird, ist für Menschen gedacht, die in der Lage sind, die Daten im Zusammenhang zu interpretieren und aus jeglicher Information, die sie finden, den größten Nutzen zu ziehen. Wenn wir aus einer übersetzten Expression das Werk identifizieren durch die Verwendung des Verfassers, des Einheitstitels und der Form des Werkes, sind dies die Schlüsselwerte, die mit dem Werk zusammenpassen müssen, wie es auch in anderen Zusammenhängen auftaucht. Leider ist das nicht immer der Fall. Als enthaltenes Werk kann es durch den Originaltitel (als Einheitstitel) identifiziert werden, aber als Inhaltseintrag oder als Eintrag für ein in Beziehung stehendes Werk kann es unter dem Titel katalogisiert sein, unter dem es im Allgemeinen in der Sprache des Katalogs bekannt ist.

Verbindungen

Die letzte strukturelle Herausforderung sind die Beziehungen. Das ist sowohl eine Frage der Beziehungsart als auch der Identifizierung der Entitäten-Endpunkte, die durch die Beziehung miteinander verbunden werden. Die Information über die Art einer Beziehung kann implizit sein oder kann aus einem Feldtag, einem Indikator oder aus spezifischen Unterfeldern, wie der Funktionsbezeichnung im Fall der Entitäten der Gruppe 2, abgeleitet werden. Bei einer Manifestation, die eine einzelne Expression verkörpert, sind die Art und die Endpunkte der Beziehungen implizit, aber noch einmal, das ist nicht der Fall, wenn wir komplexere Strukturen haben. Eine Manifestation, die zwei Expressionen verkörpert und einen einzigen Schlagworteintrag mit dem Personennamen enthält, hat eine mehrdeutige Struktur. Die Art der Beziehung ist bekannt, aber wir kennen die Endpunkte nicht. Bezieht sich der Schlagworteintrag auf beide Werke oder nur auf eines der beiden? Basierend auf der Art die richtige Verbindung zwischen Entitäten herzustellen, ist nur möglich, wenn es auf einer Seite der Beziehung eine einzige Entität gibt. Das Problem des Verbindens von Entitäten ist noch schwieriger, wenn die Art der Beziehung unbekannt ist, was der Fall ist, wenn zum Beispiel keine Funktionsbezeichnungs-codes verwendet werden.

Manifestationen mit einer einzigen Expression

Manifestationen mit einer einzigen Expression sind der einfachste Fall eines bibliografischen Datensatzes und Bücher, die einen einzigen Roman enthalten, sind gute Beispiele. Dies ist außerdem die am häufigsten vorkommende Struktur in den Datensätzen eines Bibliothekskatalogs. Theoretisch sind viele dieser Veröffentlichungen in Wirklichkeit komplexer, doch auch in einem FRBR-Kontext würden wir im Allgemeinen Cover Art, kurze Einführungen und andere Expressionen von geringerer Bedeutung vernachlässigen, weil wir eine ausgewogene Auswahl dessen treffen müssen, was als eindeutige Entität zu behandeln ist. Entitäten nur um des Modells willen zu ergänzen, ist nicht gerechtfertigt, da es nur Unruhe in das Bibliothekssystem bringen würde.

Weil Manifestationen Expressionen verkörpern und jede Expression immer die Realisierung eines (und nur eines) Werkes ist, können wir ableiten, dass diese Entitäten

existieren. Wenn es nur eine Expression gibt und folgerichtig nur ein Werk, ist es die vorrangige Herausforderung herauszufinden, welche Felder und Unterfelder welche Entität beschreiben, ebenso wie die entsprechenden Schlüsselmerkmale für diese Entitäten auszuwählen.

Die Entitäten der Gruppe 2 zu diesem Bild zu ergänzen, bereitet nicht notwendigerweise Probleme. Eindeutige Entitäten der Gruppe 2 werden in UNIMARC gut unterstützt durch die Nutzung des 700-Blocks für die Verfasserschaft, die Nutzung von Normdaten und die Nutzung von Funktionsbezeichnungs-codes. Der Funktionsbezeichnungscode ist wesentlich, da er die Hauptinformationsquelle dafür ist, wie eine Entität der Gruppe 2 zu einer Entität der Gruppe 1 in Beziehung steht. Wenn der Funktionsbezeichnungscode in einem Feld 700 Personennamen 070 ist, wissen wir, dass diese Person der Verfasser ist. In der FRBR-Terminologie wird dies eine „hat geschaffen“-Beziehung zu einem Werk sein, und weil es nur ein Werk gibt, ist die Verknüpfung recht offenkundig. Wenn der Funktionsbezeichnungscode in einer Eintragung Personennamen 730 ist, wissen wir, dass die Person der Übersetzer ist und folglich die Verknüpfung zur Expression führen sollte.

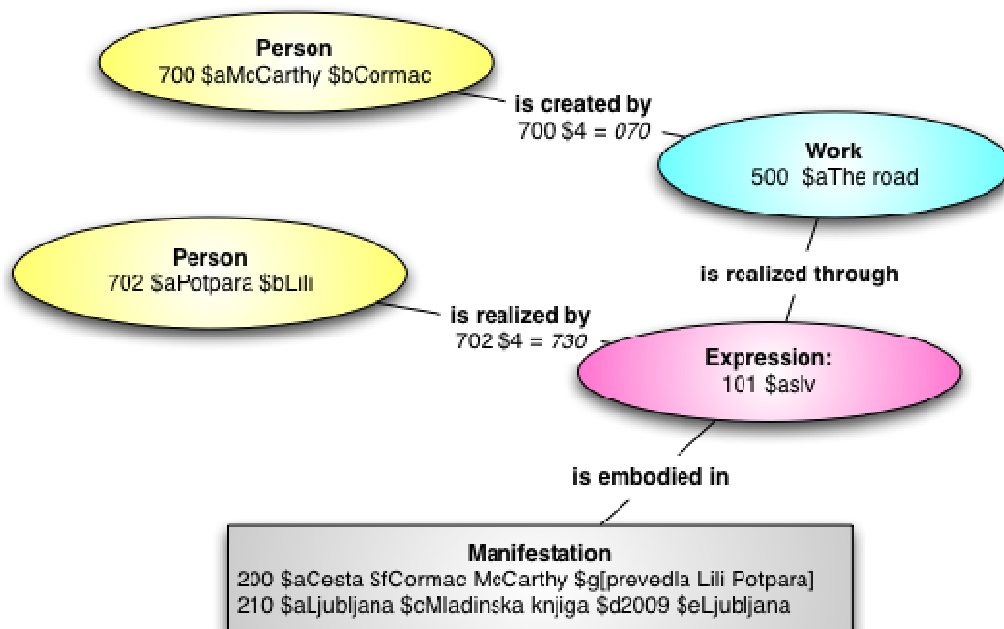


Abb. 1: Eine Manifestation, die eine Expression verkörpert, mit Beispielen wesentlicher Informationen aus dem Datensatz, der genutzt wird, um Entitäten zu identifizieren und Beziehungen herzustellen (Slowenische Nationalbibliographie). Die Leerzeichen zwischen den Unterfeldern wurden für die Lesbarkeit eingefügt.

Das Beispiel in Abbildung 1 veranschaulicht diesen Fall anhand einer slowenischen Übersetzung von „*The Road*“ des amerikanischen Schriftstellers *Cormac McCarthy*. Auch durch die einfache Abfolge der Felder im MARC-Format kann die Struktur direkt aus dem Datensatz abgeleitet werden. Probleme in Bezug auf solch einfache Strukturen sind hauptsächlich auf unzureichend beschriebene und identifizierte Entitäten sowie auf fehlende Funktionsbezeichnungs-codes zurückzuführen.

Manifestationen mit mehreren Expressionen

Manifestationen können auch mehr als eine Expression verkörpern: Kapitel oder Artikel, die im Ganzen enthalten sind, einzelne Bände oder physisch eigenständige Bestandteile.

Dies ist eine Form der Aggregation, die orthogonal zu dem Gedanken einer Manifestation ist, die (Unter-)Manifestationen als Teile hat, oder eines Werkes, das andere Werke als Teile hat. Solche Aggregationen sind oft im Datensatz beschrieben, aber die Struktur der miteinander in Beziehung stehenden Entitäten kann aufgrund der Beschränkungen von MARC mehrdeutig sein. Das Beispiel in Abbildung 2 zeigt eine Veröffentlichung, die zwei von Agatha Christies Romanen auf Portugiesisch von zwei verschiedenen Übersetzern enthält. In diesem Datensatz deuten sowohl die wiederholten Unterfelder in 200 Titelangabe als auch die wiederholte Fußnote in 304 an, dass es zwei Werke gibt. Weil es nur einen Verfasser in dem Datensatz gibt, können wir annehmen, dass dieser Verfasser beide Werke geschaffen hat. Informationen wie die Sprache können den Expressionen zugeordnet werden, basierend auf der gleichen Annahme, dass sich die eine Sprachenangabe auf beide bezieht. Die Übersetzer sind gut identifiziert, aber es gibt keine eindeutige Verknüpfung, welche 702 Personeneintragung zu welcher Expression gehört. Dieser Mangel an eindeutigen Verknüpfungen zwischen Entitäten mag auf den ersten Blick nicht wichtig erscheinen, aber er zeigt beispielhaft ein größeres Problem. Wenn die zwei Romane von verschiedenen Verfassern wären, würden wir die gleiche Art von Problem für die „*ist geschaffen von*“-Beziehung haben, was zu unzureichender oder fehlerhafter Identifizierung der Werke führen würde.

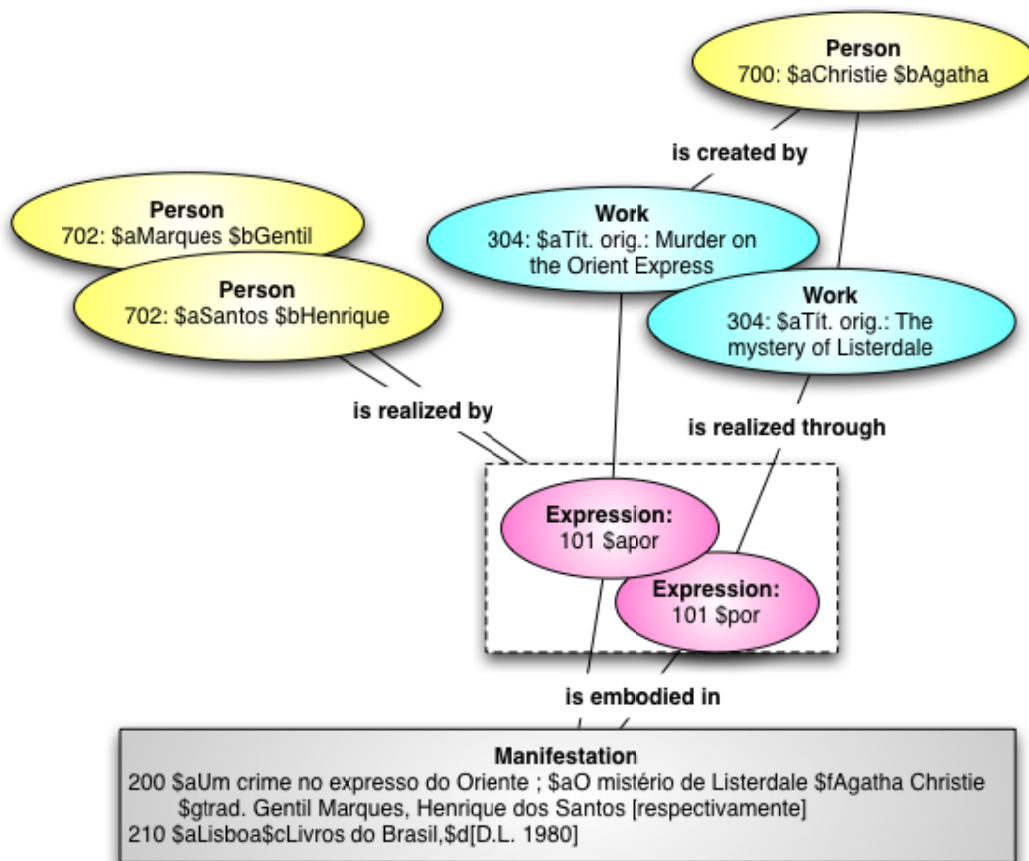


Abb. 2: Beispiel für eine Manifestation, die zwei Expressionen verkörpert (The Portuguese Union Catalogue).

Die Beispiele in Tabelle 1 zeigen einige Datensätze mit dieser Art der Struktur. Das erste Beispiel beschreibt eine Sammlung von Kriminalromanen von verschiedenen Verfassern. Der Inhalt kann teilweise durch die wiederholbaren Unterfelder in der Titelangabe identifiziert werden, aber die Auflistung der Originaltitel in einem Fußnotenfeld 304 ist eine bessere Quelle, da sie alle enthaltenen Werke umfasst. Theoretisch könnte es möglich sein, die Sequenz Titel und Personennameneintragen zu nutzen, um den Verfasser mit dem richtigen Werk zu verknüpfen, aber wie dargestellt, folgt die Reihenfolge von 70X-Feldern nicht der Titelabfolge in der Fußnote.

Das zweite Beispiel beschreibt eine Sammlung von Kriminalromanen von Agatha Christie. Dieser Datensatz verwendet nach der Methode der eingebetteten Felder ein Verknüpfungsfeld, das unter Verwendung der Felder 200 beziehungsweise 700 alle Titel und den Verfasser enthält. Zusätzlich enthält er eine Fußnote zur Inhaltsangabe in 327 mit demselben Set von Titeln. In diesem Fall gibt es mehrere Titel im Verknüpfungseintragungsfeld, die eindeutig ein und demselben Verfasser zuzuordnen sind.

Das dritte Beispiel beschreibt eine schwedische Übersetzung von Lord of the Rings von J. R. R. Tolkien. In diesem Fall gibt es mehrere Verknüpfungseintragungen, eine für jede Expression/jedes Werk. Die 500-Felder enthalten den Originaltitel, geeignet für die Werkidentifizierung, die 200-Felder würden in diesem Fall gut als Titel der Expressionen dienen, und die Einbindung des Verfassers in allen Eintragungen macht

die Struktur eindeutig. Der Übersetzer ist allerdings nur in der Verfasserangabe genannt und wird schwierig als Teil der Identifizierung der Expression zu nutzen sein.

<p>101: \$apor 200: \$aOs crimes da rua morgue \$fEdgar A. Poe \$gTrad. João Costa\$aO cachimbo de Maigret \$fGeorges Simennon \$gTrad. J. Lima da Costa 210: \$aLisboa\$cLivros do Brasil,\$d1997 304 \$aTit. orig.: The murders in the Rue Morgue ; Murder on the links ; La pipe de Maigret ; The bronze door and other stories 700 \$aPoe \$bEdgar Allan \$f1809-1849 701 \$aChandler \$bRaymond \$f1888-1959\$4070 701 \$aChristie \$bAgatha \$f1890-1976\$4070 701 \$aSimenon \$bGeorges \$f1903-1989\$4070 702 \$aCosta \$bJ. Lima da \$4730 702 \$aCosta \$bJoão Alves da \$f1948-\$4730 702 \$aRodrigues \$bFernanda Pinto\$4730</p>
<p>101 \$aeng 200 \$a1950s omnibus \$fAgatha Christie 210 \$aLondon \$cHarperCollins \$d2006 327 \$aThey came to Baghdad \$aDestination unknown \$aOrdeal by innocence \$aThe pale horse 423 \$1200 \$aThey came to Baghdad \$aDestination unknown \$aOrdeal by innocence \$aThe pale Horse \$1700 \$aChristie \$bAgatha \$f1890-1976 \$4070 700 \$aChristie \$bAgatha \$f1890-1976 \$4070</p>
<p>101 \$aswe 200 \$aHärskarringen \$fJ. R. R. Tolkien \$göversättning av Åke Ohlmarks \$gillustrationer av Alan Lee 210 \$aStockholm \$cNorstedts \$d2002 \$e[Ljubljana] \$gMladinska knjiga 300 \$aPrevod dela: The lord of the rings 327 \$aSagan om ringen = The fellowship of the ring \$aSagan om de två tornen = The two towers \$aSagan om konungens återkomst = The return of the king 423 \$1200 \$aSagan om ringen \$1500 \$aThe fellowship of the ring \$1700 \$aTolkien \$bJohn Ronald Reuel \$4070 423 \$1200 \$aSagan om de två tornen \$1500 \$aThe two towers \$1700 \$aTolkien \$bJohn Ronald Reuel \$4070 423 \$1200 \$aSagan om konungens återkomst \$1500 \$aThe return of the king \$1700 1 \$aTolkien \$bJohn Ronald Reuel \$4070 500 \$aThe lord of the rings 700 \$aTolkien \$bJohn Ronald Reuel \$4070</p>

Tabelle 1: Verschiedene Manifestationen, die zwei oder mehr Expressionen verkörpern (die ersten beiden Beispiele sind aus dem Portuguese union catalogue, das letzte Beispiel ist aus der Slowenischen Nationalbibliographie).

Das dritte Beispiel veranschaulicht außerdem einen weiteren strukturellen Aspekt. Die Person in der 700-Eintragung kann der Trilogie „Lord of the Rings“ (als Ganzes) zugeordnet werden, weil es sinnvoll ist, anzugeben, dass J. R. R. Tolkien der geistige Schöpfer von „Lord of the Rings“ war, welches das Werk ist, auf das man sich normalerweise bezieht.

In den beiden anderen Beispielen würde eine solche Angabe völlig falsch sein. Ein Werk zu erstellen für den Titel *1950s omnibus* und Agatha Christie damit als Schöpferin dieses Werkes zu verbinden, ist eine falsche Angabe – sie ist lediglich die Verfasserin der Romane, die es enthält. Beim ersten Beispiel wäre es ebenso falsch, alle Verfasser als geistige Schöpfer der Sammlung zu assoziieren. Das Vorhandensein eines Felds 500 Einheitstitel im letzten Beispiel kann darauf hinweisen, dass wir das Ganze als ein Werk

ansehen sollten, aber dies ist eine Annahme, die nur deshalb für diese Beispiele funktioniert, weil das letzte eine Übersetzung ist.

Die Expressionen/Werke aufzulisten, indem man eine der Techniken verwendet, drückt die gegenwärtige „Inhalts-“Struktur aus, aber die Verwendung von Verknüpfungseintragungsfeldern ist bei weitem die präziseste Methode, da sie die Einbindung aller Informationen ermöglicht, die für die Identifizierung jeder Expression und jedes Werkes benötigt werden. Die Nutzung mehrerer Felder ist am besten geeignet, da sie die Differenzierung zwischen den Eintragungen ermöglicht und auf einer Struktur basiert, die der natürlichen Gruppierung von Informationen entspricht. Ein Fallstrick bei der Nutzung jeder dieser Techniken ist, dass es Informationen in anderen Feldern geben kann, die sich auf alle enthaltenen Expressionen beziehen, wie z. B. die Sprache. Wenn es nur eine Sprachenangabe gibt, ist es normal anzunehmen, dass sie für alle Expressionen gültig ist. Zwei oder mehrere Sprachenangaben werden nur sinnvoll sein, wenn sie für alle Expressionen gültig sind. Die Sprache zu den enthaltenen Verknüpfungen zu ergänzen, könnte eine Lösung für den eher selten vorkommenden Fall von Expressionen in verschiedenen Sprachen sein, aber das Problem besteht ebenso für andere Informationen, wie z. B. für Schlagworteintragungen. In MARC 21 wird die Verbindung von Feldern durch die Nutzung der Feldverknüpfung mit \$8 unterstützt, was auch in UNIMARC eine mögliche Lösung für das Problem ist.

Beziehungen zu anderen Werken

Das FRBR-Modell definiert eine Vielzahl von Beziehungen zwischen Entitäten der Gruppe 1, zusätzlich zu den grundlegenden Beziehungen „*ist verkörpert*“ und „*ist realisiert*“.

Die Gründe für die Ergänzung solcher Beziehungen zu einem Datensatz sind eine bessere Beschreibung der katalogisierten Ressource zu geben und Benutzern zu ermöglichen, in Beziehung stehende Werke zu finden. Im Kontext von FRBR sind solche Beziehungen noch wichtiger, weil sie einen Mechanismus zur Unterstützung von neuen Formen von Benutzerinteraktionen bieten, wie z. B. recherche- und retrieval-basiertes Browsen und Suchen. Die Verknüpfungseintragungsfelder sind der Hauptmechanismus für die Erstellung solcher Beziehungen in UNIMARC, aber nur wenige der in den FRBR beschriebenen Arten von Beziehungen werden speziell unterstützt. Alle anderen Beziehungen werden durch die Verwendung der generischen 488-Verknüpfungseintragung katalogisiert werden müssen, wahrscheinlich unter Nutzung einer 311 Fußnote. Das letztgenannte ist leider kein besonders guter Mechanismus. Zum einen ist es eine nicht-standardisierte textliche Beschreibung, die nur für die Anzeige gedacht ist. Zum anderen wird es scheitern, wenn es verschiedene Verknüpfungseintragungsfelder gibt und die Fußnote nur für ein spezifisches Feld gültig ist. Die Methode der eingebetteten Verlinkung ist der gängigen Unterfeld-Technik grundsätzlich überlegen, weil sie die detaillierte Identifizierung einer Entität ermöglicht, indem sie genau dasselbe Set von Datenfeldern nutzt, das benutzt würde, wenn die Entität durch die Nutzung der üblichen Datenfelder in einem Datensatz beschrieben würde. Diese Technik ermöglicht eine ziemlich präzise Identifizierung von Werken, Expressionen und Manifestationen, aber es bedarf eines besseren Mechanismus, um die Beziehungen zu typisieren.

Auch wenn die Verknüpfungsfelder und die Nutzung der eingebetteten Verlinkung gut geeignet schienen, um Beziehungen auszudrücken, gibt es bei dieser Technik eine große Herausforderung, wenn es um Manifestationen geht, die mehrere Expressionen

verkörpern. Wenn alle enthaltenen Expressionen durch Datenfelder mit Verknüpfungseintragungen beschrieben werden, und es ein oder mehrere Verknüpfungseintragungsfelder gibt, die Werk-zu-Werk- oder Expression-zu-Expression-Beziehungen beschreiben, gibt es keine Möglichkeit zu unterscheiden, zu welcher der enthaltenen Expressionen die Beziehung gehört. Wir können annehmen, dass sie zu allen gehört, aber man kann davon ausgehen, dass diese Vereinfachung falsche Interpretationen verursachen wird.

Eine weitere wichtige Beziehungsart ist die von Werken, die Entitäten der Gruppe 1 oder 2 zum Thema haben. In UNIMARC gibt es spezielle Datenfelder für Namen von Personen, Körperschaften oder Familien als Thema sowie für Titel als Thema. Das Feld 604 ist für Namen und Titel vorgesehen und basiert auf der Methode der eingebetteten Felder. Erneut gibt es ein Problem, wenn die enthaltenen Expressionen (oder eher die entsprechenden Werke) verschiedene Themen haben, wie z. B. im Falle von Essay-Sammlungen usw. Es gibt keine spezifische Möglichkeit anzuzeigen, welches Thema zu welchem Werk gehört.

Abbildung 3 stellt die Gruppierung dar, die durch die Nutzung von Verknüpfungseintragungsfeldern und der Methode der eingebetteten Felder erlangt werden kann. Die verschiedenen „üblichen“ Felder beschreiben die Manifestation und alle Informationen, die für alle enthaltenen Expressionen/Werke gültig sind. Zwei Verknüpfungseintragungsfelder werden benutzt, um zwei verschiedene verkörperte Expressionen und die Werke, deren Realisierungen sie sind, zu beschreiben, ebenso wie die Personen, die zu den spezifischen Expressionen und Werken gehören. Durch das Tag des Verknüpfungsfelds wissen wir, dass es sich um enthaltene Expressionen handelt, die zu der im Datensatz beschriebenen Manifestation gehören (z. B. durch Nutzung des Feldes 423, um den früheren Beispielen zu folgen). Zusätzliche Verknüpfungseintragungen gibt es für ein in Beziehung stehendes Werk und ein Thema. Die Verbindung zwischen diesen Eintragungen sowie die Eintragungen für die enthaltenen Expressionen werden dennoch nicht beschrieben und können nicht präzise interpretiert werden.

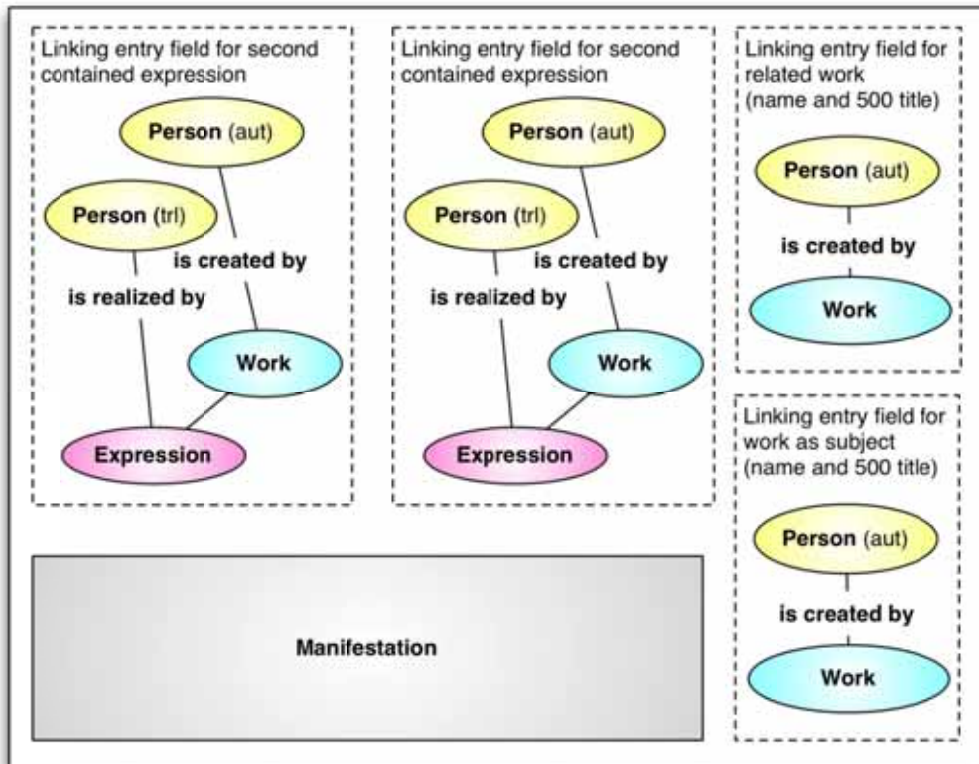


Abb. 3: Beispiel für die Nutzung von Verknüpfungseintragungsfeldern und der Methode der eingebetteten Felder.

Schlussfolgerungen

In diesem Vortrag haben wir einige der Probleme diskutiert und untersucht, die bestehen, wenn die FRBR in UNIMARC kodiert werden. Das Problem der Identitäten- und Merkmalszuordnung stammt hauptsächlich aus der Katalogisierungspraxis (und aus Katalogisierungsregeln) und wird durch unzureichende Informationen (wie beispielsweise dem Fehlen von Titeln, die die Werke genau identifizieren) oder Datenfelder verursacht, die im Kontext der FRBR mehrdeutig sind. Einige dieser Probleme werden in RDA angesprochen und dem werden hoffentlich Änderungen an den Datenfeldern in den MARC-Formaten folgen.

Das Problem, ein Set von Entitäten mit korrekt verlinkten Beziehungen in UNIMARC zu definieren und zu beschreiben, ist eine andere Sache. Das gegenwärtige Format ist ausreichend zur Beschreibung von Manifestationen mit einer einzigen Expression, aber die Beispiele, die wir aufgeführt haben, zeigen, dass für komplexere Strukturen heutzutage eine Vielzahl von Lösungen verwendet wird, die entweder auf den Anmerkungen zum Inhalt basieren, auf wiederholbaren Unterfeldern in der Titelangabe oder auf Verknüpfungsfeldern. Die beiden erstgenannten verwenden beschreibende Informationen, die nicht ausreichend sind, weil sie schwierig zu verarbeiten sind und in unzureichender Identifizierung der Entitäten münden. Glücklicherweise scheint die Verwendung von Verknüpfungsfeldern eine geeignete Lösung zu sein, um viele der Probleme mit den komplexeren Strukturen zu lösen. Dies kann als Zwischenlösung angesehen werden, die der grundlegenden MARC-Struktur eine neue Tiefe hinzufügt, die zum FRBR-Modell zu passen scheint. Diese Technik imitiert die Verwendung von mehrfach verlinkten Datensätzen, aber mit dem Vorteil, dass alle Informationen

innerhalb eines Datensatzes gehalten werden. Mit der Methode der eingebetteten Felder können wir alle Informationen einschließen, die wir brauchen, um alle Entitäten zu beschreiben und zu identifizieren, die mit dem „Zweig“ der enthaltenen Expression in Beziehung stehen.

Ein besonderes Problem, mit dem wir uns beschäftigt haben, ist die Mehrdeutigkeit, die dann vorkommt, wenn die Informationen in einem bestimmten Feld sich auf mehrere enthaltene Expressionen oder Werke beziehen können oder es in Beziehung stehende Werke/Expressionen gibt, die sich nur auf bestimmte enthaltene Werke/Expressionen beziehen. Bei inhaltsbezogenen Datenfeldern kann das durch Regeln dafür gelöst werden, was in die eingebetteten Felder der Verknüpfungseintragung einbezogen werden soll, aber für Beziehungen zwischen Werken in separaten Verknüpfungseintragungen ist diese Lösung nicht möglich. Ein Feldverknüpfungsmechanismus, vergleichbar mit dem Unterfeld \$8 in MARC 21 könnte das Problem lösen, entweder durch Hinzufügen eines Unterfelds für diesen Zweck (das nicht mit anderen eingebetteten Unterfeldern kollidieren sollte) oder durch Verwendung des nicht verwendeten ersten Indikators dafür (nur zutreffend zum Verbinden von Verknüpfungsfeldern).

Wenn wir möchten, dass die Verknüpfungsfelder vollständig die FRBR-Beziehungen unterstützen, benötigen wir entweder speziell konzipierte Eintragungen in Verknüpfungsfeldern oder einen Mechanismus, mit dem wir die Art der Beziehung durch Hinzufügen eines Identifikators ausdrücken können. Das erstgenannte würde eine große Anzahl von Verknüpfungsfeldern mit sich bringen, das letzte erfordert ein eigens dafür konzipiertes Unterfeld, das schwierig einzuführen ist, weil es mit den eingebetteten Unterfeldern kollidieren könnte.

Wir haben nur Fälle beschrieben, die sich auf mehrere Expressionen beziehen, die in einer Manifestation verkörpert sind, aber es ist uns bewusst, dass noch mehr Arten von Strukturen zu untersuchen sind, wie zum Beispiel Manifestationen, Expressionen und Werke, die mehrere Teile oder Nachträge oder ähnliches haben. Einige Probleme können durch Verknüpfungsfelder gelöst werden, aber es kann Fälle geben, in denen andere Lösungen notwendig sind.

Schließlich möchten wir einige allgemeinere Empfehlungen für die künftige Entwicklung von UNIMARC herausheben.

Perspektivenwechsel von der Darstellung zur Verarbeitung

Die FRBR in einem Bibliothekssystem zu implementieren, bedeutet mehr als nur das Format oder das zugrundeliegende Datenmodell zu ändern. Die FRBR repräsentieren ein logisches Modell der fraglichen Domäne und Implementierungen sind abhängig von Austauschformaten und Speichermodellen, die das Modell widerspiegeln. Die eigentlichen Anforderungen an bibliografische Datensätze hängen in Zukunft maßgeblich davon ab, was wir von den Daten erwarten. Jüngste Entwicklungen bei Benutzerschnittstellen haben die Verwendung von FRBR-ähnlichen Gruppierungen der Ergebnismenge untersucht, und wir sehen die Vorzüge des Modells immer deutlicher. Eine wesentliche Herausforderung der Zukunft ist es, die tatsächlichen Anforderungen an Daten zu definieren, im Hinblick auf die Verarbeitung und auf Abfragen, welche von den Daten unterstützt werden sollen, und nicht im Hinblick darauf, wie sie angezeigt werden.

Überarbeitung des Standards

Das FRBR-Modell hat die RDA beeinflusst und wird dementsprechend künftig die Standards zur Kodierung der bibliografischen Datensätze beeinflussen. Einige Veränderungen wurden im Format MARC 21 vorgenommen und es ist wichtig, dass UNIMARC das Format entsprechend modernisieren kann. Eine bessere Unterstützung der FRBR führt nicht notwendigerweise zu einer rückwärtigen Inkompatibilität, aber die FRBR nicht zu unterstützen wird zwangsweise zu einer künftigen Inkompatibilität führen.

Definition von Best Practices zur Kodierung der FRBR

Die FRBR in MARC zu kodieren, ist nicht notwendigerweise eine Frage davon, dem Datensatz mehr Informationen hinzuzufügen und damit zusätzliche Erschließungskosten zu produzieren. Das Hauptproblem besteht in der Verwendung einer Vielzahl von Lösungen in verschiedenen Katalogen ebenso wie innerhalb desselben Katalogs. Einige Erschließungspraktiken halten wir für weniger sinnvoll für künftige FRBR-Implementierungen als andere. Im aktuellen Standard gibt es keine Beispiele, die sich auf die FRBR beziehen, und es fehlen generell Richtlinien oder Best-Practice-Beispiele zur Kodierung von Informationen in einer FRBR-kompatiblen Art und Weise. Selbst ohne Änderung des Formats und ohne Änderung von Katalogisierungsregeln und -praxis ist es möglich, Best-Practice-Lösungen für eine Reihe von häufig auftretenden Fällen vorzuschlagen.

Korrekte Versionierung

Die meisten Bibliothekskataloge enthalten Datensätze, die über viele Jahre oder gar Jahrzehnte erstellt und zusammengetragen wurden. Katalogisierungspraxis und -formate können sich im Laufe der Zeit verändert haben, was zu kleineren und größeren Unterschieden führt. Migrationen von nationalen auf international anerkannte Formate können syntaktisch mit den Zielstandards konform sein, aber es bleiben immer Reste der früheren Praxis zurück, die nicht migriert werden können. Ein Wechsel zu einer anderen Verwendung des Formats, die an das FRBR-Modell angepasst ist, ist ein Wechsel, der unter Verwendung einer Art von Versionierung vollzogen werden sollte, so dass wir in der Lage sind, diese Datensätze zu identifizieren. Der aktuelle Identifier der Version 005 von UNIMARC ist einfach das Datum der letzten Aktualisierung und gibt nichts anderes an. Wenn ein Bibliothekskatalog einige Datensätze enthalten soll, die auf bestimmte Arten, die FRBR auszudrücken und zu verarbeiten, zugeschnitten sind, während andere es nicht sind, dann ist eine Minimalanforderung, dass wir in der Lage sein sollen, diese beiden durch Verwendung eines Versionenidentifiers zu unterscheiden.

Literaturangaben

Aalberg, T. (2006). A process and tool for the conversion of MARC records to a normalized FRBR implementation. *ICADL 2006, LNCS*, vol. 4312, 283-292.

Hegna, K. and Murtomaa, E. (2002). Data mining MARC to find: FRBR? *68th IFLA Council and General Conference, 18-27 August 2002, Glasgow, Scotland*.

<http://archive.ifla.org/IV/ifla68/papers/053-133e.pdf>

Hickey, T. and O'Neill, E. (2005). FRBRizing OCLC's WorldCat. *Cataloging & Classification Quarterly*, 39 (3/4), 239-251.

IFLA International Federation of Library Associations and Institutions. Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional Requirements for Bibliographic Records: final report*. Munich, Germany: KG Saur

Manguinhas, H., Freire N. and Borbinha, J. (2010). FRBRization of MARC records of multiple catalogs. *Proceedings of ACM/IEEE Joint Conference on Digital Libraries*. 225-234.

Mcgrath, K. and Bisko, L. (2008). Identifying FRBR work-level data in MARC bibliographic records for moving images. *Code4Lib Journal*. 1 (5), December 2008.
<http://journal.code4lib.org/articles/775>

Pisanski, J., Žumer, M. and Aalberg, T. (2009). Frbrisation: towards a bright new future for national bibliographies. *World Library and Information Congress: 75th IFLA General Conference and Council, 23-27 August 2009, Milan, Italy*. <http://www.ifla.org/files/hq/papers/ifla75/77-pisanski-en.pdf>