



# 68th IFLA Council and General Conference

## August 18-24, 2002

---

**Code Number:** 053-133-R  
**Division Number:** IV  
**Professional Group:** Cataloguing  
**Joint Meeting with:** -  
**Meeting Number:** 133  
**Simultaneous Interpretation:** Yes

### Поиск данных в MARC: FRBR?

#### **Кнут ХЕГНА**

Библиотека Университета Осло  
Oslo, Norway  
E-mail: [Knut.Hegna@ub.uio.no](mailto:Knut.Hegna@ub.uio.no)

#### **Эва МУРТОМАА**

Библиотека Университета Хельсинки  
Helsinki, Finland  
E-mails: [Eeva.Murtomaa@helsinki.fi](mailto:Eeva.Murtomaa@helsinki.fi)

---

### **Резюме:**

*В настоящем докладе подводятся итоги одного из проектов, в ходе которого данные, представленные в MARC-форматах, из двух национальных библиографий были проанализированы в свете модели данных, представленной ИФЛА в исследовании FRBR -- «Функциональные требования к библиографическим записям». В ходе проекта мы обнаружили, что даже хотя информация в записях MARC несет в себе атрибуты, релевантные для идентификации таких явлений, как произведение, выражение и проявление, все же точность и формальный синтаксис слишком просты для того, чтобы программы могли работать с ними должным образом. Некоторые из этих результатов могут быть использованы для улучшения вывода результатов поиска в ОПАСах. Для пользовательского интерфейса ОПАСов проект дал два предложения, основанных на идеях из исследования «Функциональных требований...» и на результатах проекта.*

Полный отчет см.: <http://folk.uio.no/knuthe/dok/frbr/datamining.pdf>

---

## **1. Функциональные требования к библиографическим записям и MARC**

Первый вопрос, который мы себе задали, был таков: «Можем ли мы найти структуру FRBR в существующих записях MARC – формата?» Посмотрев на

отдельную запись, мы видим, что в ней имеется информация о произведении, выражении и проявлении.

```
*008880325          no          esp
*02000 $a 82-991075-3-0 $b h. $c Nkr 60. 00
*04110 $a espnor
*08200 $a 839.822[S]
*10010 $a Ibsen, Henrik $d 1828 – 1906
*24510 $a Puppejmo (1879) $c Henrik Ibsen ; tradukis :
        Odd Tangerud ; lingve kontrolita de
        Esperantista Verkista Asocio (EVA)
*26000 $a Hokksund $b Eldenjo Odd Tangerud $c 1987
*26900 $a [Drammen] : Tngen –Trykk
*30000 $a [1], 57 s. $c 24 cm
*50000 $a Originaltittel : Et dukkehjem. –
        Originalutgave : København : Gyldendal, 1879
*99100 \ $a Tangerud, Odd
```

Эта запись из Норвежской национальной библиографии описывает документ, в котором содержится вариант знаменитой пьесы Генрика Ибсена «Кукольный дом» на языке эсперанто. Заглавие произведения мы находим в 500 поле примечаний, связи с создателем произведения – в поле 100. Эта информация может идентифицировать произведение, описанное в записи.

Более того, через коды языков в полях 008 и 041 мы находим информацию о выражении, а связи с лицом, ответственным за перевод, - частично в сведениях об ответственности и в поле 991.

Информация о проявлении содержится в полях, описывающих обрабатываемый документ: 245, 260, 269, 300 и др.

Итак, ответ на наш вопрос будет: «**ДА**». Почему? Потому что библиографическая запись может:

- описывать и произведение, и проявление;
- содержать в себе ссылки к выражению;
- содержать некоторые взаимосвязи с добавочными описаниями, примечаниями и предметными описаниями.

В какой-то степени элементы модели FRBR присутствуют в MARC записи!

Но, к сожалению, ответ также: «**Нет**». Почему?

- потому что правила каталогизации хорошо приспособлены для карточного каталога и печатной библиографии, а не для модели FRBR.
- центральная информация часто записывается так, чтобы она была более удобна глазам и разуму человека, а не компьютеру.

Рассматривать отдельную запись бесполезно. Мы решили проанализировать множества записей, полученные в результате поиска в финской и норвежской национальных библиографиях. Мы ограничились поиском по автору.

В конце названия нашего проекта поставлен вопросительный знак. Мы не знали, каких результатов добьёмся. Мы предполагали, что это исследование не будет окончательно завершено. Мы полагали, что, по крайней мере, нам удастся определить ряд проблем, связанных с правилами каталогизации и форматом MARC.

Для начала мы проанализировали таблицы из исследования по FRBR, показывающих, какие атрибуты явлений важны для идентификации произведения, выражения и проявления. Затем мы попытались установить соответствие записей MARC и этих атрибутов.

Чтобы получить названия произведений мы смотрели унифицированные или оригинальные заглавия. Прежде всего, мы проверили, имеется ли поле 241 (оригинальное заглавие). В национальной библиографии Норвегии это поле не используется, а в финской – используется. Если поля 241 не было, мы проверяли поля 500 и 505 на наличие слов *Оригинальное заглавие* на двух языках. В норвежских записях: *Originaltittel*: сокращенно – *Orig. tit.* и во множ. числе *Originaltittler*. В финских записях: *Alkuteos*: и во множ. числе *Alkuteokset*. Заглавие оригинала всегда следует после этих слов. Если при такой проверке в 50X полях оригинальное заглавие найдено не было, мы переходили к другим полям заглавия, рассматривая их в следующем порядке: 248 \$h (основное заглавие части – в финских записях) и 240 \$a. Последним из возможных решений было решение выбрать оригинальное заглавие из поля 245.

Обычно заглавия, включенные в поле 505, повторяются в полях добавочных описаний на заглавие (745 \$a в Финляндии и 740 \$t в Норвегии), но способа решить, являются ли приведенные в этих полях данные заглавием, нет. Иногда в этих полях содержатся оригинальные заглавия, а иногда – другая информация о заглавии. Ни 745, ни 740 поле не содержат информации, определяющей это заглавие. Во многих случаях 74X поля – это добавочные описания, относящиеся к другому произведению, а не к рассматриваемому, но каким-то образом связанные с ним.

Мы полагали, что вместе со связью с создателем произведения в каждой записи можно определить произведения или произведения. Основываясь на этом, располагая этими данными, мы могли бы выявить идентичные произведения, находящиеся в нескольких записях, и дифференцировать их с другими произведениями. Сделав это, мы использовали другие данные, чтобы дифференцировать/выявить разные выражения, идентифицируемые по коду языка и переводчику. Для идентификации проявлений использовались другие данные.

Таблица 1: Атрибуты (определения), идентифицирующие произведение в формате MARC

Атрибут FRBR	Оценка по FRBR	NORMARC	FinMARC	Отобрано
Заглавие произведения	Высокая	241 \$a	241 \$a	Да
		500 \$a	500 \$a	Да
		505 \$a	505 \$a	Да

		240 \$a 245 \$a	248 \$a 240 \$a 245 \$a	Да Да Да
Связь с лицом, ответствен ным за...	Высокая	100 \$a 70010 \$a	100 \$a \$h 70010 \$a \$h	Да
Предпола- гаемая продолжи тельность	Высокая	?	?	Нет
Форма произведе ния	Умеренная	Интерпрета- ция Дьюи	008 п.24-27, 29- 30, 33-34	

И вот некоторые результаты:

Мы видим, что имеется 744 записи, в которых Генрик Ибсен выступает как автор или как лицо, на которое дано добавочное описание. Мы смогли выделить из этих записей 914 идентификаторов произведений, и по этим 914 идентификаторам мы определили 220 отдельных произведений. Выглядит прекрасно, но Ибсен написал 26 пьес и несколько сборников стихов. Откуда же тогда 220 произведений?

*Таблица 2: Результаты использования процедуры сокращения числа произведений в записях из норвежской (Н) и финской (Ф) национальной библиографии*

<b>АВТОР</b>	<i>Число записей</i>	<i>Число строк произведений</i>	<i>Число строк отдельных произведений</i>
Ибсен, Генрик (Н)	<b>744</b>	<b>914</b>	<b>220</b>
Вассмо, Хербйорг(Н)	<b>149</b>	<b>159</b>	<b>19</b>
Гаардер, Йостейн (Н)	<b>237</b>	<b>237</b>	<b>14</b>
Солстад, Даг (Н)	<b>92</b>	<b>93</b>	<b>35</b>
Куннас, Маури (Ф)	<b>130</b>	<b>133</b>	<b>41</b>
Янссон, Туве (Ф)	<b>576</b>	<b>595</b>	<b>92</b>
Линна, Вайно (Ф)	<b>168</b>	<b>168</b>	<b>20</b>

Имелись записи, в которых недоставало информации или информация была неверной: сборники и избранные пьесы, оригинальные или унифицированные заглавия которых не могли быть идентифицированы, хотя они и присутствовали (12); 33 записи, в которых подзаголовочные данные были включены в основное заглавие; 14 записей, в которых современное написание заглавия отличается от оригинального заглавия; 10 записей вообще без оригинальных заглавий, т.е. заглавием произведения было только заглавие на иностранном языке.

## 2. Интерфейс пользователя

Итак, мы определили некую структуру. Второй вопрос, который мы задали себе после этого, был таким: «Какую пользу мы можем получить от этой структуры, выводя для пользователя результаты поиска?»

Сначала мы посмотрели в таблицы FRBR, определив, какие атрибуты важны для выбора одного элемента из ряда сходных элементов и установили, каким меткам MARC соответствуют эти атрибуты.

Чтобы упростить выбор, важно также решить, какого рода сортировка является наиболее подходящей. Принцип сортировки должен быть понятен для пользователя и должен варьировать в зависимости от того, какой уровень элемента представлен.

Мы полагаем, что результат поиска должен выводиться в соответствии с проведённым поиском и с самими результатами. Если вы ищете определенное лицо, в список должны войти его (её) произведения, расположенные в некотором порядке – алфавитном или, возможно, хронологическом, или в соответствии со списком функций, которые он (она) выполнял по отношению к этим документам: автор, иллюстратор, переводчик.

В первом интерфейсе выдача представлена в виде наложенных друг на друга карточек. Верхние части карточек – это горизонтальная ось, направленная «вглубь» экрана, который образует систему из двух осей, а поверхность карточки – это вертикальные оси, дифференцирующие информацию под верхним заголовком каждой карточки.

Такими двумя осями могут быть

*авторы – произведения,*

*произведения – выражения,*

*выражения – проявления,*

*предметы – произведения,*

в зависимости от того, какой именно поиск проведён. Если пользователь ищет конкретного автора и получено несколько записей, на горизонтальной оси будет выведен список имен и фамилий лиц, а на вертикальной – произведения каждого из авторов. С другой стороны, если при поиске найдена только одна запись, на горизонтальной оси будут указаны произведения этого писателя, а на вертикальной – различные выражения каждого их произведений.

Это можно продемонстрировать на примере произведений Йостейна Гаардера, расположенных вдоль горизонтальной оси, и их выражениях вдоль вертикальной оси под каждым произведением. Количество выражений, идентифицированных

для каждого произведения, указано в скобках после заглавия произведения. Количество проявлений для каждого выражения указывается после информации о выражении.

Пользователь может выбрать адекватное выражение, активизируя, таким образом, отдельное окно, в котором под этим выражением указаны все его проявления. Проявления выбранного выражения сортируются хронологически, для каждой записи указывается год издания. В окне имеются кнопки для различных функций («сохранить», «печать» и т.д.) и пока окно не будет явно закрыто, оно не исчезнет. Это значит, что пользователь может работать с тем числом выражений, какое ему необходимо, причем с каждым в отдельном окне.

Второй интерфейс пользователя представляет произведения в виде ветвей списка результатов, изображенного в виде дерева. При выводе на экран произведения сортируются в алфавитном порядке (по заглавию оригинала), а число, указанное после заглавия, обозначает, сколько выражений этого произведения идентифицировано.

Узлы произведений можно раскрыть, при на следующем уровне («листья») представлены выражения. Выражения сортируются так же, как и для уже описанного выше интерфейса: язык оригинала наверху, и под ним языки в алфавитном порядке.

Эти листья активны, щелчком мыши можно открыть окно выражения, как описано выше.

### **3. РЕЗЮМЕ**

Одна из главных причин шума, выявленного в ходе нашего эксперимента, заключается в неквалифицированном использовании поля 700, и в том, что системы индексируют эту информацию вместе с полем 100, независимо от того, какую реальную функцию выполняло это лицо.

Мы полагаем, что указание этой функции в поле 700 должно быть обязательным, и что список функций должен быть расширен. Системы должны использовать указание функции для представления конечному пользователю более структурированного списка результатов поиска.

Это даст системам возможность показывать все функции этого лица, представленные в базе данных, давая конечному пользователю возможность в ходе библиографического поиска выбрать все или некоторые функции.

Кроме того, наше исследование показало, что для анализа было бы полезно, если бы заглавия оригиналов можно было вводить более последовательно, в отдельные, повторяющиеся поля.

Коды языков – это один из наиболее важных атрибутов для идентификации различных выражений одного и того же произведения. Они также очень важны для того, чтобы выяснить, является ли выражение переводом, или нет.

Нормализация данных сама по себе является средством связи. Использование авторитетных файлов в библиографической среде помогает установить структуру навигации, как собирая явления, так и дифференцируя их. Чаще всего используются авторитетные файлы имен и названий организаций. Наша работа показывает, что, для того, чтобы идентифицировать одно и то же произведение под одним и тем же заголовком, существует необходимость и в авторитетных файлах произведений.