

UNIMARC и связанные данные.

UNIMARC and linked data

Gordon Dunsire

Freelance consultant

Edinburgh, UK

E-mail: gordon@gordondunsire.com

Mirna Willer

University of Zadar

Department of Library and Information
Science

Zadar, Croatia

E-mail: mwiller@unizd.hr

**Meeting: 187 — Advancing UNIMARC:
alignment and innovation — IFLA**

UNIMARC Programme (UNIMARC)

Перевод В. Скворцова,

Национальная Служба развития системы
форматов RUSMARC.

Резюме

Основная цель этой статьи – представить аргументы за и рекомендации для представления форматов UNIMARC для библиографических и авторитетных данных в RDF (Resource Description Framework), стандарте W3C для структурирования данных в Semantic Web и Linked Data Environment. Это делается в продолжение работ, уже начатых соответствующими группами IFLA в отношении представления ISBD и концептуальных моделей FRBR, FRAD и FRSAD. Авторы настоятельно рекомендуют, чтобы PUC выступил с предложением о финансировании IFLA работ по представлению форматов UNIMARC в RDF как исследовательского и эволюционного проекта.

Введение и основные предпосылки

«Выражение *Linked Data* (связанные данные) относится к установлению наилучших способов публикации и связывания данных в Web-пространстве». ^[1] При таком подходе данные представляются как простые выражения, использующие Resource Description Framework (RDF) и связанные посредством машиночитаемых идентификаторов, соответствующих синтаксису Uniform Resource Identifier (URI). Выражения RDF принимают форму трёхчастной структуры *субъект-предикат-объект* (subject-predicate-object), где *субъект* определяет предмет описания, *предикат* – специфический аспект описываемого предмета, и *объект* определяет или представляет значение этого аспекта. Поэтому выражение RDF обычно называют тройкой (triple). Основой тройки является *предикат*, который выражает RDF-свойство, а *субъект* и *объект* являются составляющими RDF-классов. Классы описывают предметы, в то время как свойства (properties) описывают взаимоотношения между этими предметами; классы и свойства – основные типы элементов RDF. Предмет (thing), описанный как класс, может представлять любой тип ресурса или сущности, в отношении которой мы собираемся построить выражение. Контролируемые термины, используемые как элементы *объекта* в тройках, могут представлять собой некий «словарь значений», созданный на основе Системы организации простых знаний (Simple Knowledge Organization System – SKOS ^[2]) – [семейства формальных языков для представления тезаурусов, классификационных схем, таксономий, систем предметных заголовков и других типов структурированных

словарей контролируемой лексики (Wikipedia). В оригинале: специального набора элементов RDF для простых тезаурусов и таксономий]. Объект также может быть представлен буквенной строкой данных, таких как имя лица, область издания, и т. д., не поддерживаемой словарём или контролируемой терминологией. Тройка – это по существу метаданные, то есть данные о данных, в этом случае – данные о *субъекте* (subject) тройки. *Linked data* должны, следовательно, представлять особый интерес для библиотечного сообщества, развивающего изолированные, ориентированные на пользователя, подходы к библиографическим метаданным в форме каталогов, регулируемых согласованными на международном уровне стандартами. Характерная черта *linked data* – их ориентация на Web, Семантический Web (Semantic Web), позволяющая производить разделение данных на глобальном уровне между множественными разнородными источниками. Опять же, это замечательная возможность для библиотек, которые обмениваются MARC-записями с 1960-х годов.

Библиотечные *linked data*, произведённые из существующих записей, созданных на основе международных стандартов, будут высокого качества и в большом количестве, представляя множество информационных ресурсов, предположительно интересующих пользователей Семантического Web'a. Один только WorldCat OCLC содержит свыше 230 млн. библиографических записей^[3]. Анализ MARCовского содержания^[4] даёт свыше 13 миллионов подполей на примерно 420 тыс. записей. Полагая, что каждое подполе генерирует одну тройку (triple), получим в среднем 31 тройку на запись. Эта цифра не уменьшается за счёт дублетности внутри WorldCat, поскольку легко компенсируется за счёт записей, не входящих в WorldCat, что даст по крайней мере миллиард троек из всего наследия записей. Не менее важны данные, создаваемые библиотеками для авторитетного контроля, включая имена лиц, наименования организаций, места, предметные входы и т. п., по-видимому, представляющие интерес для более широкой аудитории, нежели традиционные пользователи библиотек.

Возможность использования существующих библиотечных стандартов для создания новых троек (triples) и извлечения троек из унаследованных записей, требует их представления в RDF – или путём создания соответствующих RDF-элементов, или путём конвертирования в существующие элементы. Это не просто позволит Семантическому Web'у извлечь выгоду из библиотечных метаданных, но должно улучшить взаимодействие между библиографическими сущностями, атрибутами и отношениями, описанными различными, но связанными стандартами. Атрибуты RDF должны выбираться из различных стандартов и объединяться внутри одного приложения с целью соответствия их функциональным требованиям, с использованием Dublin Core Application Profile^[5] или онтологии, выраженной в RDF/OWL^[6].

IFLA, как организация, определяющая стандарты, должна быть особенно заинтересована во внедрении *linked data* и Семантического Web'a, в силу своей обязанности развивать и поддерживать библиографические модели и стандарты, и обеспечивать библиотечному сообществу, таким образом, лучшее обслуживание пользователей в изменяющихся технологических условиях. Кроме того, поддерживая развитие, ведущее к представлению её международно-согласованных стандартов в RDF, IFLA обеспечивает подлинность и достоверность метаданных, созданных библиотеками, что исключительно важно в окружении, позволяющем «каждому сказать что угодно о любом ресурсе», в то же время продвигая свой бренд за границы библиотечного сообщества. Используя чётко определённые взаимоотношения, «можно путём компьютерной обработки создать Web доверия [Godlbeck and Parsia]. Организация системы доверия в Семантическом Web'e позволит компьютерам более просто определять, какая информация исходит из авторитетных источников, а какая нет»^[7].

Первую инициативу пересмотра стандартов IFLA в контексте Web-технологий и сервисов можно отнести к 2006-му году, когда группа по пересмотру ISBD Секции каталогизации IFLA приняла к сведению рекомендацию своей группы по изучению Обозначений материала развивать XML схему для ISBD. С этой целью в 2008 году была организована ISBD/XML Study Group^[8], однако, поскольку работа группы по пересмотру FRBR (FRBR Review Group^[9]) над FRBR^[10], связанная с RDF, началась годом раньше, Study Group приняла решение вместо общей XML разметки рассмотреть представление ISBD как такого в RDF. Этот трёхлетний проект ныне находится в финальной фазе и должен быть завершён к декабрю 2011^{[11],[12]}. FRBR Review Group продолжала свою работу по представлению моделей IFLA, расширив её до моделей для авторитетных данных (FRAD^[13]) и предметных авторитетных данных (FRSAD^[14]). Все три модели, так же, как и ISBD, созданы с использованием Open Metadata Registry (OMR)^[15].

Необходимо, однако, упомянуть, что все эти инициативы разрабатывались в тесной взаимосвязи друг с другом, подпитывая общее развитие представления стандартов IFLA в RDF. Необходимо также упомянуть, что проводилось исследование возможности применения RDA (Resource Description and Access), как содержательного стандарта для формата UNIMARC, в добавление к и наряду с ISBD в контексте Семантического Web'a^[16].

Последние, третьи, редакции форматов UNIMARC для библиографических и авторитетных данных были опубликованы соответственно в 2007^[17] и 2009^[18] годах, с последующими апдейтами, подготовленными Постоянным комитетом IFLA по формату UNIMARC (Permanent UNIMARC Committee – PUC). В третьей редакции формата UNIMARC для авторитетных данных уже внедрены специфические свойства модели FRAD^[19], с целью сближения с этой моделью, в то время как его сближение с FRSAD по-прежнему ждёт своей очереди. Решения по сближению формата UNIMARC с FRBR, а также с новой, консолидированной редакцией ISBD^[20] – в процессе одобрения. Не говоря уже о том, форматы UNIMARC строго следуют другим стандартам IFLA, в случае библиографических данных – это, в частности, ISBD. Таким образом, работа по представлению формата UNIMARC в RDF в особенности для библиографических данных – есть расширение работы ISBD/XML Study Group, которая формирует основу этой статьи.

Пространство имён формата UNIMARC

RDF требует, чтобы классы и свойства имели машиночитаемые идентификаторы, соответствующие синтаксису Uniform Resource Identifier (URI)^[21]. Набор URI с базовой информацией о соответствующих классах и свойствах, опубликованный и используемый в единственном контексте, называется пространством имён. Все URI в пространстве имён состоят из общей строки символов, называемой базовым доменом, к которому добавляется строка отличия (distinguish string), называемая локальной частью. В частности, выгода такого подхода заключается в том, что базовый домен может приводиться в форме короткой аббревиатуры для пользователя и автоматически расширяться для машинной обработки. UNIMARC RDF элементы будут изначально создаваться и поддерживаться с использованием OMR, следуя тому же подходу, что и ISBD и семейство функциональных требований к моделям метаданных (FRBR, FRAD, FRSAD). В частности, OMR поддерживает многоязычные метки и другие аннотации. Это важное требование для IFLA стандартов, предназначенных для приложения в международном окружении и созданных для многоязычного взаимодействия. UNIMARC переведён с английского на китайский, хорватский, французский, итальянский, литовский,

португальский, русский и т.д. языки. RDF по существу нейтрален по отношению к языкам, поскольку предназначен для машинной обработки, но допускает, чтобы к одному и тому же элементу приписывались метки, определения, примечания об области использования и другие аннотации. Долговременная инфраструктура, необходимая для управления пространством имён IFLA будет исследоваться и развиваться IFLA Namespaces Technical Group, в соответствии с рекомендациями, данными в отчёте этой группы^[22].

Идентификация элементов формата UNIMARC через тег/подполе/позицию символа

В настоящей статье для идентификации элементов формата UNIMARC посредством тегов (поля), кодами подполей, индикаторами и позициями символов в блоке 1-- Блок кодированной информации, используются аббревиатурные коды по следующей схеме:

Тег + 1-ый индикатор + 2-ой индикатор + код подполя, при этом символ *b* используется для обозначения отсутствующего индикатора, а также вместо значка #, обозначающего пробел. Разделитель подполя \$ не требуется, поскольку каждый аббревиатурный код относится к единственному под полю, и шесть позиций дают код подполя, например:

010bba = ISBN

2001ba = Основное заглавие (значимое заглавие)

2000ba = Основное заглавие (незначимое заглавие)

Для Блока кодированной информации аббревиатурный код формируется с добавлением позиций символов, например:

100bba8 = Тип даты публикации

100bba17-19 = Код целевого назначения

100bba34-35 = Графика заглавия

Домены пространства имён

Одно, или больше пространств имён с соответствующими доменами потребуются для представления элементов формата UNIMARC и словарей в RDF.

Заемствование пространств имён

Наилучшей практикой является формирование RDF-элементов и словарей из уже существующих пространств имён, где это возможно: это сохраняет время и силы на развитие элементов и словарей и на их поддержку, проще развивать приложения и сервисы, работающие с метаданными, это стимулирует *mix-and-much* подход к приложениям, это развивает *web* связанных элементов и *linked data*. Наиболее важно, однако, убедиться, что используемые заново элементы, жёстко связаны с использующими их стандартами, с тем, чтобы какие-либо изменения в их прямом или косвенном значении (семантическое соседство) непосредственно отражались в родственном пространстве имён, чтобы предотвратить семантический дрейф между двумя пространствами имён.

Библиографический UNIMARC следует за ISBD, который уже имеет опубликованное пространство имён для наборов своих элементов и словарей (Область 0 – область формы содержания и типа носителя). Таким образом, у формата UNIMARC есть выбор:

1. Использовать классы и свойства ISBD, где это возможно, вместо создания своих собственных в пространстве имён библиографического формата UNIMARC. Это целесообразно только в том случае, если предлагаемые изменения в одном стандарте влекут за собой рассмотрение их влияния на другой, т.е. оба стандарта управляются и поддерживаются «как один».
2. Представить все элементы библиографического формата UNIMARC в своём собственном пространстве имён формата UNIMARC и связать с эквивалентными классами и свойствами в пространстве имён ISBD. Эта возможность должна выбираться, если UNIMARC и ISBD продолжают развиваться отдельно, даже если между ними существует тесная взаимосвязь.

Этот выбор должен быть сделан до того, как начнётся самостоятельное развитие пространства имён для библиографического формата UNIMARC.

В таблице 1 приводится пример соответствия потенциальных свойств (properties) библиографического формата UNIMARC существующим свойствам (properties) ISBD. Обратите внимание, что домен пространств имён как для properties формата UNIMARC, так и для ISBD не приводится ради краткости изложения.

UNIMARC	English label	ISBD property	English label
P205bba	has edition statement	P1008	has edition statement
P205bbb	has issue statement*	P1011	has additional edition statement
P205bbd	has parallel edition statement	P1009	has parallel edition statement
P205bbf	has statement of responsibility relating to edition	P1010	has statement of responsibility relating to edition
P205bbg	has subsequent statement of responsibility**	P1010	has statement of responsibility relating to edition

Таблица 1. Пример соответствия свойств библиографического формата UNIMARC существующим свойствам RDF ISBD. Библиографический формат UNIMARC, поле 205 Область издания.

* Различия в областях применения, например P205bbb, могут быть сглажены с использованием SKOS properties для альтернативных областей, как это делается для некоторых FRBR properties, имеющих различные области применения во FRAD.

** Пример, когда элемент формата UNIMARC более детален, чем элемент ISBD. Для сохранения элемента формата UNIMARC и более точной детализации, необходимо определение UNIMARC property в пространстве имён библиографического формата UNIMARC.

UNIMARC Authorities, как уже упоминалось «принимает во внимание атрибуты сущностей и их взаимоотношения, как определено в Функциональных требованиях к авторитетным данным (*Functional Requirements for Authority Data: Conceptual Model - FRAD*)» в следующих аспектах: «изменена терминология, определения полей и контрольного подполя \$5, Управление связями [...]. Блоки переименованы на 2-- Блок точек доступа, 4-- Блок вариантных точек доступа, 5-- Блок связанных точек доступа и 7-- Блок принятых точек доступа на других языках и/или в альтернативной графике, в то

время как теги обозначают имена сущностей, которые представляют контролируемые точки доступа, такие как Имя лица, Наименование организации, Заглавие». Однако FRAD – это модель, в то время как UNIMARC/A – носитель содержания на уровне приложения, поэтому эквивалентность определений FRAD и UNIMARC/A необходимо проверить. К тому же, в отличие от ISBD, FRAD – это расширение FRBR и в свою очередь использует подходящие элементы пространства имён последнего. Необходимо заметить, что между форматами UNIMARC для библиографических и авторитетных данных существует внутренняя взаимосвязь, означающая, что UNIMARC/A развивается, следуя изменениям и добавлениям в библиографическом формате. Согласование между UNIMARC/A и UNIMARC/B определяется подполем \$3 в полях точек доступа формата UNIMARC/B и специальным образом в разделе Guidelines for Use UNIMARC/A [В российской версии, в разделе «Соответствие полей российских форматов для библиографических записей и для авторитетных / нормативных записей». - прим. перев.]. Поэтому необходимо исследовать, как позиционировать UNIMARC/A в отношении Блока ответственности 7-- формата UNIMARC/B с одной стороны и FRAD с другой, чтобы определить взаимоотношение элементов для специфических случаев связанных данных. Например, если специфический элемент – точка доступа/входа в формате UNIMARC/B имеет соответствующую запись в UNIMARC/A, которая также должна быть опубликована в RDF, тогда между ними должна быть установлена цепочка связанных данных, иначе, точка доступа формата UNIMARC/B, например 700 – Имя лица – первичная ответственность, используемая без ссылки на UNIMARC/A через подполе \$3, может быть представлена как буквенная строка.

Специфические пространства имён формата UNIMARC

ISBD не охватывает точки доступа или заголовки и соответствующие им авторитетные записи, поэтому UNIMARC Authorities не имеет соответствующих ISBD классов и свойств. Поэтому требуется создать пространство имён для UNIMARC Authorities.

Пример в таблице 1 показывает, что элементы ISBD не доходят до уровня детализации элементов формата UNIMARC для библиографических данных, поэтому создание пространства имён, специфичного для UNIMARC Bibliographic, для элементов, не охватываемых ISBD также необходимо, независимо от использования элементов ISBD. UNIMARC Bibliographic и UNIMARC Authorities должны иметь отдельные пространства имён как отражение отдельной публикации текстов и различия в подобных кодах полей и подполей.

Предполагается, что пространства имён форматов UNIMARC должно следовать шаблонам, установленным рабочей группой IFLA Namespaces Task Group для ISBD и FRBR, FRAD и FRSAD.

Для элементов формата UNIMARC Authorities базовый домен пространства имён имеет вид:

<http://iflstandards.info/ns/unimarc/unimarca/elements/>

Для представления его пользователям может использоваться аббревиатура "unimarca". Обратите внимание, что это не URL, это «cool» URI.

Для элементов формата UNIMARC Bibliographic базовый домен пространства имён имеет вид:

<http://iflstandards.info/ns/unimarc/unimarcb/elements/>

Его аббревиатура "unimarcb".

Для словарей форматов UNIMARC используются отдельные базовые домены пространств имён для каждого словаря, следуя практике ISBD. Базовый домен состоит из строки специфического идентификатора словаря, добавленного к общему базовому домену для UNIMARCовских словарей:

<http://iflstandards.info/ns/unimarc/terms/> + строка идентификатора словаря

Специфическая строка идентификатора словаря не может базироваться на tag/indicators/subfield/character position code, поскольку некоторые словари, например Графика, приписываются более чем к одному тегу и в UNIMARC/B, и в UNIMARC/A.

UNIMARC/A: 100bba21-22

UNIMARC/B: 100bba34-35

Вместо этого может использоваться аббревиатура названия словаря, например:

<http://iflstandards.info/ns/unimarc/terms/graphicsmd> как пространство имён для словаря для специфического обозначения материала в 116bba1.

В этом случае для представления пользователю может использоваться аббревиатура "unimarcgsmd".

Заметим, что нет необходимости использовать базовый домен, чтобы показать, взят ли данный словарь из библиографического формата, или авторитетного. Словари могут использоваться как *объект* (object), обладающий соответствующими свойствами (properties), использующимися в тройке (triple), и использование словарей может быть представлено в application profile как схема кодировки словаря *Vocabulary encoding scheme*, связанная с соответствующими свойствами RDF. Такой подход изымает словарь из его специфического использования внутри формата UNIMARC и упрощает его использование другими сообществами в не-UNIMARCовских приложениях.

Application profiles

Один, или более DC Application profile потребуются для формата UNIMARC для представления заимствованных элементов, если что-либо: ISBD или FRAD классы или свойства, использование совокупных выражений, составленных из двух или более свойств (как в ISBD), использование специфических словарей в качестве Vocabulary Encoding Schemes, или другое, вынуждает использовать элементы в правильно сформированной записи библиографического или авторитетного формата UNIMARC, такие как статус – обязательный или повторяемый элемент.

Как UNIMARC/B, так и UNIMARC/A определяют обязательные, повторяемые и неповторяемые элементы в записи формата UNIMARC. К обязательным элементам в обоих форматах относятся: 001 Идентификатор записи, 100 Данные общей обработки (только некоторые элементы совпадают в обоих форматах) и 801 Источник записи, в то время как специфичными для каждого формата являются: 200\$a Основное заглавие в UNIMARC/B (помимо отдельных полей, специфичных для ресурсов различных типов) и 2-- Блок точек доступа в UNIMARC/A. Оба формата определяют повторяемые и неповторяемые элементы на уровне идентификаторов полей и подполей, например 010

ISBN – повторяемое, в то время как 010\$a Номер – неповторяемое в формате UNIMARC/B. В формате UNIMARC/A поле 220 Точка доступа – Родовое имя – повторяемое, но только для альтернативной графики, в то время как 220\$a Начальный элемент ввода – не повторяемое. Порядок следования идентификаторов подполей в записи формата UNIMARC не определён, поскольку определяется данными.

Мета-метаданные

Данные о специфике UNIMARCовской записи содержатся в маркере записи и в 1-- Блоке кодированной информации. Это мета-метаданные, то есть данные о метаданных. В RDF существует множество техник, которые могут использоваться для представления таких данных, например, квалификатор языка, который может добавляться к буквенной строке, например, заглавия, используемого как объект в тройке, например «@en» означает, что строка – на английском языке, «@fr» - на французском и т.д. Эти техники не требуют специфических UNIMARCовских элементов и исключаются из дальнейшего обсуждения в этой статье.

Существуют также мета-метаданные, специфичные для записей UNIMARC, такие как, например, данные ISO 2709, структурированного формата обмена записями, такие как длина записи, коды применения, длина индикатора и т.д. Такие элементы не имеют значения, когда метаданные представляются в RDF как тройки.

Словари Блока кодированной информации

Коды и соответствующие величины и определения, используемые в Блоке кодированной информации для описания ресурса (а не запись UNIMARC), наилучшим образом представляются как SKOS-словари, аналогично словарям области 0 ISBD.

Код формата UNIMARC для словарного выражения (term-code) может использоваться как локальная часть его URI. Например:

```
http://iflastandards.info/ns/unimarc/terms/graphicsmd#a = "collage"
```

Когда term-код выражен числом, ему должна предшествовать буква, например T (term), чтобы избежать проблем в XML с локальной частью, начинающейся с числового символа, что соответствует практике ISBD.

Использование term-кода таким образом сохраняет языковую независимость URI, позволяет избежать перегрузки URI семантикой, а также позволяет избежать путаницы, если выражение (английское) изменяется в будущем (скажем, с "collage" на "mixed-media two-dimensional sculpture").

Сам по себе UNIMARCовский код вполне может быть представлен через свойство skos:notation. В следующих примерных тройках, использующих это свойство, URI используется как *субъект* (subject), а term code – как величина *объекта* (object):

```
<http://iflastandards.info/ns/unimarc/terms/graphicsmd#a> skos:notation "a".  
(или, используя аббревиатуру пространства имён: unimarcgsmd#a skos:notation "a".)  
<http://iflastandards.info/ns/unimarc/terms/publicationdatatype#f> skos:notation "f".  
<http://iflastandards.info/ns/unimarc/terms/titlescript#ca> skos:notation "ca".
```

Подобным образом выражение (term) само по себе может быть представлено с использованием свойства skos:prefLabel с добавлением языкового квалификатора. Например:

<<http://iflastandards.info/ns/unimarc/terms/graphicscmd#a>> skos:prefLabel "collage"@en.

<<http://iflastandards.info/ns/unimarc/terms/publicationdatetype#f>> skos:prefLabel "monograph, date of publication uncertain"@en.

<<http://iflastandards.info/ns/unimarc/terms/titlescript#ca>> skos:prefLabel "Cyrillic"@en.

Таблица 2 даёт полный пример словаря Блока кодированной информации с метками на английском, итальянском и португальском языках, взятыми из официальных переводов формата UNIMARC.

N	PrefLabel@en	PrefLabel@it	PrefLabel@pt	Definition@en	ScopeNote@en
a	collage	collage	colagem	An original work created by affixing various materials (paper, wood, newspaper, cloth, etc.) to a surface.	
b	drawing	disegno	desenho	An original visual representation (other than a print or painting) made with pencil, pen, chalk, or other writing instrument on paper or similar non-rigid support.	
c	painting	pittura	pintura	An original visual representation produced by applying paint to a surface.	
d	photomechanical reproduction	riproduzione fotomeccanica	reprodução fotomecânica	Any picture produced in imitation of another picture through the use of a photographic process to transfer the image to a printing surface.	A snapshot made to document a painting or a Xerox copy of a print are considered photomechanical reproductions. Art reproductions, postcards, posters, and study prints are included here.
e	photonegative	fotonegativo	negativo fotográfico	A piece of film, a glass plate, or paper on which appears a "negative" image, i.e. directly opposite to a	Does not include negative photoprints, photoprints that are a combination of negative and positive images,

				"positive" image (photoprint), slide, or transparency. Used to produce a positive print.	photographs or solarized prints, all of which are considered to be techniques used when making photoprints.
f	photoprint	riproduzione fotografica	positivo fotográfico	A positive image made either directly or indirectly on a sensitized surface by the action of light or other radiant energy.	The term "photoprint" is used here as a more precise term than "photograph", which technically can cover both the print and the negative. Radiographs and opaque stereographs are included here.
h	picture	immagine	imagem	A twodimensional visual representation accessible to the naked eye and generally on an opaque backing.	This term is used when a more specific designation is unknown or not desired.
i	print	stampa	gravura	A design or picture transferred from an engraved plate, wood block, lithographic stone, or other medium.	Generally, there are four types: planographic print, relief print, intaglio print, and stencil print.
k	technical drawing	disegno tecnico	desenho técnico	A cross section, detail, diagram, elevation, perspective, plan, working plan, etc., made for use in an engineering or other technical context.	
m	master	master	matriz	Any plate, mould, matrix, die etc. which allows the reproduction of the same impression.	
z	other nonprojected graphic type	altro tipo di documento grafico non proiettabile	outro material gráfico não projectável	[Types other than collage, drawing, painting, photomechanical reproduction, photonegative,	Includes mixed media productions made by a combination of freehand and printing techniques

				photoprint, picture, print, technical drawing, master.]	when one or the other does not predominate. In some cases, where mixed media are applied, one must decide whether the creator intends the item to be a photoprint (even though it is painted over the photographic image). Hand colouring is considered a technique applied to a printing process; this aspect is covered by a character position 3. Computerproduced graphics and the various duplication masters (including spirit masters and transparency masters) are included here.
--	--	--	--	---	---

Таблица 2. Полный пример словаря Блока кодированной информации: словарь 116bba0 = Кодированные данные для изобразительных материалов: Специфическое обозначение изобразительного материала (Колонка N = код, также является частью URI).

Термины, определения и примечания о содержании взяты из следующих источников:

@en: <http://archive.ifla.org/VI/8/unimarc-concise-bibliographic-format-2008.pdf>

@it: <http://unimarc-it.wikidot.com/116>

@pt: http://www.ifla.org/files/uca/Unimarc_bib_3%C2%AAed_abrev.pdf

В заголовки Определения и Примечания о содержании включена разметка [Definition@en, ScoreNote@en] чтобы показать их происхождение из английского текста.

Использование внешних SKOS словарей

Некоторые из наборов кодированных величин формата UNIMARC в точности базируются на внешних словарях или терминологии. Например, язык инципита в 036bbaz или язык документа в различных местах поля 101 используют трёхсимвольный код, взятый из Приложения А, которая есть ни что иное, как список формата MARC для языков. Этот список доступен как SKOS-словарь^[23], который может использоваться непосредственно в

любой тройке формата UNIMARC. Коды географических регионов в 660bba также доступны в SKOS^[24].

Для обозначения страны публикации в 102bba используется 2-х символьный код из Приложения В, который является стандартом ISO 3166-1. Он не совпадает со списком стран формата MARC, который Библиотека Конгресса опубликовала как SKOS-словарь^[25]. Однако RDF-представление ISO 3166-1 – доступно^[26], хотя требуется дальнейшее исследование его применимости для формата UNIMARC.

Доступность SKOS-представлений для других внешних словарей, используемых в формате UNIMARC необходимо проверить. Если никаких SKOS-представлений найти нельзя, Постоянный Комитет по формату UNIMARC должен бы вступить в контакт с собственником словаря, чтобы обсудить его развитие и целесообразное представление в RDF.

Некоторые внутренние словари формата UNIMARC могут уже иметь подходящие SKOS-представления, даже если они не в точности базируются на внешних словарях и терминологиях. Возможный пример – набор символьных кодов, используемых в 100bba26-29. То есть какое-нибудь другое сообщество может уже иметь SKOS-представление подобного словаря, который содержит все выражения и коды, используемые форматом UNIMARC, и если так, UNIMARC может заимствовать необходимые SKOS URI. Это также требует дальнейшего исследования и проверки, например, с использованием онтологических поисковых машин, таких как Swoogle.

Кодировка дат

Кодированные элементы, которые представляют даты в определённом формате, могут быть представлены с использованием свойства rdfs:range. Например, даты публикации 1 и 2 в поле данных общей обработки библиографического формата UNIMARC могут использовать тройки (следуя соглашению URI, данному ниже):

```
unimarcb:100bba09-12 rdfs:range xsd:gYear.
```

```
unimarcb:100bba13-16 rdfs:range xsd:gYear.
```

Классы формата UNIMARC

Как и в случае ISBD, есть только один класс, который необходимо рассмотреть для библиографического формата UNIMARC, за исключением классов для синтаксических схем кодировки, необходимых для внедрения внешних элементов, как сказано ниже. Это ISBD-класс Resource, который можно использовать как домен для всех RDF-свойств библиографического формата UNIMARC. Нет необходимости создавать UNIMARC-класс для Resource.

Классы для формата UNIMARC/Authorities требуют дальнейшего исследования, особенно в отношении FRAD/FRSAD. Пространство имён FRAD имеет классы для всех сущностей FRAD, таких как Библиографическая сущность, Имя / Наименование, Идентификатор, Контролируемая точка доступа, Правила и организация, а также подклассы для Имени лица, Наименования организации, Родового имени и Названия произведения (Name of a Work); в нём также имеются подкласс для Наименования организации, поскольку его определение отличается от определения FRBR, и класс для Рода, который не определён как сущность в опубликованном документе FRBR. Пространство имён FRAD, однако, не включает классов для других сущностей, таких как Лицо, Произведение, Выражение,

Воплощение и т.д., поскольку они уже опубликованы в пространстве имён FRBR. В ходе анализа выясняется, что классы FRBR не удовлетворяют всем типам возможных классов формата UNIMARC/Authorities: Лицо, Наименование организации, Производство могли бы рассматриваться как идущие в одном ряду, но есть примеры, где некоторые типы сущностей формата UNIMARC/Authorities, или кандидаты в классы, требуют проведения специального анализа. Таково, например, Место, которое в FRBR определено просто как «A location», что может соответствовать Территориальному или географическому наименованию формата UNIMARC/Authorities, но не вполне **Месту, как точке доступа**, которое было вначале определено для записи места (страны или города), но впоследствии расширено, чтобы включить **Место и дату публикации, исполнения и т. д.**, к тому же, поскольку UNIMARC/Authorities – общий формат для авторитетных записей Имен/Наименований и Предметных авторитетных записей, класс «Место» должен рассматриваться и в контексте FRAD. Другой пример – это Производство/Выражение: UNIMARC/Authorities определяет среди типов его сущностей следующие: Заглавие, Типовое заглавие, Имя/Заглавие, Имя/Типовое заглавие, но в 3-ей редакции он не делает различия, относится ли данный тип сущности к Произведению или к Выражению. FRAD имеет подкласс для Имен, относящемуся к Произведению, но если UNIMARC определяет новый тип сущности Выражение, то необходимо добавить к его пространству имён подкласс Имя, относящееся к Выражению. Возможные классы вне областей, охватываемых FRBR и FRAD, это такие сущности UNIMARC/A, как Торговая марка, Форма, Жанр и Физические характеристики, которые FRAD сознательно исключает из своего рассмотрения.

Свойства в формате UNIMARC

В общем случае элементы формата UNIMARC тег/индикаторы/подполе будут представлены как RDF-свойства, в соответствии с практикой ISBD.

Не все элементы UNIMARC подходят или целесообразны для представления в качестве RDF-свойств. Это относится к элементам мета-метаданных, рассмотренным выше. Однако другие данные в маркере записи, такие, как тип записи (в обоих форматах) в действительности являются метаданными. Несмотря на то, что часть информации, содержащейся в маркере, должна быть представлена в теле записи, это не всегда так, поэтому некоторые элементы маркера записи потребуют RDF-представления. Это такие элементы, например, как Библиографический и Иерархический уровни в UNIMARC/B, или Тип сущности в UNIMARC/A, и это требует дополнительного исследования.

Аббревиатурные коды тег/индикаторы, используемые в настоящей статье, могут формировать локальную часть URI. Разделение элементов URI через слэш «/», предпочтительнее, чем через хэш «#», в тех случаях, когда имеется большое количество свойств. Локальная часть должна предваряться буквой, чтобы избежать проблем с XML. ISBD использует P (property) для свойств, и UNIMARC может следовать этому соглашению. Например, URI для Области издания может быть:

<http://iflstandards.info/ns/unimarc/unimarcb/elements/P205bba>

(или, используя аббревиатуру пространства имён: unimarcb:P205bba)

Такой подход не зависит от языка, избавляет от семантической нагрузки URI и позволяет избежать путаницы, если "заголовок" (caption) связан с изменениями тега, индикатора или подполя.

Сам по себе `caption` может быть представлен с использованием свойства `rdfs:label` и языкового квалификатора `RDF`. Заголовок (`caption`) может потребовать синтеза из отдельных заголовков. Следуя практике `ISBD`, метки `RDF`-свойств могут быть сделаны вербальными, путём предварения заголовка словом *has*. Например,

`unimarcb:P205bba reg:name "hasEditionStatement"`.

Таблица 3 даёт полный пример `RDF`-свойств взятых из одного поля без индикаторов.

URI	Label@en	Name
P205bba	has edition statement	hasEditionStatement
P205bbb	has issue statement	hasIssueStatement
P205bbd	has parallel edition statement	hasParallelEditionStatement
P205bbf	has statement of responsibility relating to edition	hasStatementOfResponsibilityRelatingToEdition
P205bbg	has subsequent statement of responsibility	hasSubsequentStatementOfResponsibility

Таблица 3. Полный пример `RDF`-свойств, представляющих поле формата `UNIMARC` без индикаторов: Область издания.

Как уже обсуждалось, каждая уникальная комбинация индикаторов и подполя внутри поля составляет отдельное `RDF`-свойство с отдельной соответствующей меткой. Метод достижения этого – снабжать заголовок подполя индикатором «`caption`», как показано в таблице 4:

URI	Label@en	Name
P206bba	has mathematical data statement (unstructured)	hasMathematicalDataStatementUnstructured
P206bbb	has statement of scale (unstructured)	hasStatementOfScaleUnstructured
P206bbc	has statement of projection (unstructured)	hasStatementOfProjectionUnstructured
P206bbd	has statement of coordinates (unstructured)	hasStatementOfCoordinatesUnstructured
P206bbe	has statement of zone (unstructured)	hasStatementOfZoneUnstructured
P206bbf	has statement of equinox (unstructured)	hasStatementOfEquinoxUnstructured
P2060ba	has mathematical data statement (structured)	hasMathematicalDataStatementStructured
P2060bb	has statement of scale (structured)	hasStatementOfScaleStructured
P2060bc	has statement of projection (structured)	hasStatementOfProjectionStructured
P2060bd	has statement of coordinates (structured)	hasStatementOfCoordinatesStructured
P2060be	has statement of zone (structured)	hasStatementOfZoneStructured
P2060bf	has statement of equinox (structured)	hasStatementOfEquinoxStructured

Таблица 4. Полный пример RDF-свойств, представляющих поле формата UNIMARC с одним двоичным индикатором: Область специфических сведений: картографические материалы - математические данные.

Когда в поле используются оба индикатора, со множеством значений для каждого индикатора, число потенциальных RDF-свойств существенно увеличивается за счёт комбинаторного взрыва, как показано в таблице 5:

URI	Label@en
P210bba	has place of publication, distribution, etc. (sequence of publication data not applicable or earliest available publisher; produced in multiple copies, usually published or publically distributed)
P210b1a	has place of publication, distribution, etc. (sequence of publication data not applicable or earliest available publisher; not published or publically distributed)
P2100ba	has place of publication, distribution, etc. (intervening publisher; produced in multiple copies, usually published or publically distributed)
P21001a	has place of publication, distribution, etc. (intervening publisher; not published or publically distributed)
P2101ba	has place of publication, distribution, etc. (current or latest publisher; produced in multiple copies, usually published or publically distributed)
P21011a	has place of publication, distribution, etc. (current or latest publisher; not published or publically distributed)

Таблица 5. Частный пример RDF-свойств, представляющих поле формата UNIMARC с двумя многозначными индикаторами: Публикация, распространение и др.

Общее число потенциальных свойств для этого примера вычисляется как 3 (число значений первого индикатора) умножить на 2 (число значений 2-го индикатора) и умножить на 8 подполей = 48.

В формате UNIMARC есть поля с гораздо большим числом комбинаций:

327 Примечания о содержании: $4 \times 2 \times 12 = 96$

620 Место и дата публикации, исполнения и т. д.: $7 \times 3 \times 15 = 315$

621 Место и дата, связанные с историей экземпляра: $7 \times 3 \times 16 = 336$

852 Местонахождение и шифр хранения: $7 \times 4 \times 16 = 448$

Выяснение, не являются ли отдельные комбинации нереальными и не требующими отдельных свойств, требует дополнительного исследования.

Совокупные выражения

Каждое повторяющееся поле с более чем одним подполем формирует совокупное выражение (aggregated statement). Необходимо сохранять подполя вместе для каждого вхождения поля, чтобы они не смешивались.

Например:

010bba: Международный стандартный номер книги – Номер + 010bbb: Международный стандартный номер книги – Уточнения.

Совокупные выражения представляются в RDF через Syntax encoding schemes (SES) с необходимостью следования практике ISBD. Заимствование ISBD элементов, представляющих собой сами по себе совокупные выражения, избавляет от необходимости развития эквивалентов формата UNIMARC.

Заключение и рекомендации

Вовлечение IFLA в деятельность по публикации её международно-согласованных стандартов в RDF, как первый шаг к признанию библиотечных метаданных в качестве авторитетных и достоверных для Семантического Web'a, уже произошло. Однако, хотя эти первые шаги охватывают все три концептуальные модели, FRBR, FRAD и FRASD и библиографический стандарт ISBD, необходима дальнейшая работа. Эта статья, представляя некоторые решения и поднимая вопросы для будущего анализа, доказывает необходимость представления форматов IFLA UNIMARC, как для библиографических, так и авторитетных данных, таким же образом. Авторы также доказывают, что координация работы по представлению стандартов IFLA должна проводиться более тесно, поскольку на практике они рассматриваются и используются согласованно, а также потому, что так их будущее развитие станет более эффективным и экономичным. Другой аспект работы по представлению стандартов в RDF, тот, что это позволяет взглянуть с новой точки зрения на сами стандарты, их структуру, точность в словесных выражениях и определениях, последовательность, взаимодействие с другими родственными библиотеками и различными сообществами, имеющими дело со стандартами метаданных, что вызвано новой технологической парадигмой Семантического Web'a.

Рекомендации Постоянному Комитету по формату UNIMARC для дальнейшего обсуждения и утверждения:

- Утвердить метод идентификации элементов и словарей формата UNIMARC.
- Принять решение по поводу изначального создания и поддержки элементов и словарей формата UNIMARC в OMR.
- Поддерживать и продвигать перевод классов и свойств формата UNIMARC на национальные языки.
- Принять решение – заимствовать ли существующее пространство имён ISBD для формата UNIMARC/B, или представить [в RDF] элементы UNIMARC/B и правильно связать с существующими классами и свойствами ISBD.
- Провести более глубокое исследование и определить – производить ли заимствование существующих пространств имён FRAD/FRBR и FRASD, или же представить [в RDF] все элементы UNIMARC/A и корректно связать с существующими FRAD/FRBR/FRASD классами/подклассами и свойствами.
- Утвердить шаблон для пространств имён элементов и словарей UNIMARC/B и UNIMARC/A.
- Обсудить и рассмотреть требования к Application Profiles для формата UNIMARC.
- Исследовать возможность SKOS представлений внешних словарей для формата UNIMARC.
- Изучить внутренние словари формата UNIMARC с точки зрения возможности их SKOS-представлений. Рассмотреть предложения к владельцам внешних словарей о развитии их SKOS-представлений.
- Провести более глубокое исследование целесообразности классов формата UNIMARC/A в отношении UNIMARC/B, FRAD/FRBR и FRASD.
- Провести более глубокое исследование «комбинаторного взрыва» свойств формата UNIMARC; определить, не являются ли некоторые ситуации невозможными и не требующими отдельных свойств.
- Рассмотреть и одобрить заимствование совокупных элементов ISBD, которые представляются в RDF с использованием Syntax encoding schemes (SES), которые избавляют от необходимости развития самостоятельных UNIMARCовских эквивалентов.

- Провести обзор родственных инициатив в развитии MARC21, в особенности Bibliographic Framework Transition Initiative, недавно анонсированной Библиотекой Конгресса^[27].

Авторы статьи настоятельно рекомендуют, чтобы PUC обратился в IFLA с просьбой о выделении финансирования развития представления формата UNIMARC в RDF, в качестве исследовательского и эволюционного проекта.

Библиография

- [1] Bizer, Christian; Tom Heath, Tim Berners-Lee. Linked data – the story so far. // International Journal on Semantic Web and Information Systems (IJSWIS), vol. 5, issue 3. (2009). Pre-print available at: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
- [2] W3C. SKOS Simple Knowledge Organization System - home page. 2010. Available at: <http://www.w3.org/2004/02/skos>.
- [3] OCLC. WorldCat facts and statistics. 2011. Available at: <http://www.oclc.org/worldcat/statistics/default.htm>.
- [4] Moen, William E.; Penelope Bernardino. Assessing metadata utilization: an analysis of MARC content designation use. // International Conference on Dublin Core and Metadata Applications, DC-2003--Seattle Proceedings. Available at: <http://dcpapers.dublincore.org/ojs/pubs/article/view/745/741>.
- [5] Coyle, Karen; Thomas Baker. Guidelines for Dublin Core application profiles. 2009. Available at: <http://dublincore.org/documents/profile-guidelines/index.shtml>.
- [6] W3C OWL Working Group. OWL 2 Web Ontology Language: document overview. 2009. Available at: <http://www.w3.org/TR/owl2-overview>.
- [7] Graves, Mike; Adam Constabaris, Dan Brickley. FOAF: connecting people on the Semantic Web. // Knitting the Semantic Web / Jane Greenberg, Eva Méndez, editors. Binghamton, NY: The Howarth Information Press, 2007. P. 196.
- [8] IFLA. Cataloguing Section. ISBD Review Group, ISBD/XML Study Group. 2008. Available at: www.ifla.org/en/node/1795.
- [9] IFLA. Cataloguing Section. FRBR Review Group. Meeting Report Durban, August 21, 2007. Available at: www.ifla.org/files/cataloguing/frbrg/meeting_2007.pdf.
- [10] Functional requirements for bibliographic records: final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München: Saur, 1998. Amended 2009, available at: www.ifla.org/en/publications/functional-requirements-for-bibliographic-records.
- [11] Dunsire, Gordon; Mirna Willer. Standard library metadata models and structures for the Semantic Web. // Library hi tech news, vol. 28, no. 3. (2011), pp. 1-12. Available at: <http://dx.doi.org/10.1108/07419051111145118>.
- [12] Willer, Mirna; Gordon Dunsire, and Boris Bosančić. ISBD and the Semantic Web. // J LIS.it Journal of Library and Information Science. Italy, vol. 1, no. 2, (2010), pp. 213-236. Available at: <http://dx.doi.org/10.4403/jlis.it-4536>.
- [13] Functional requirements for authority data : a conceptual model / edited by Glenn E. Patton ; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Final report, December 2008 / approved by the Standing Committees of the IFLA Cataloguing Section and IFLA Classification and Indexing Section, March 2009. München: K. G. Saur, 2009. Also available at: www.ifla.org/publications/functional-requirements-for-authority-data.
- [14] Functional requirements for subject authority data (FRSAD) : a conceptual model / edited by Marcia Lei Zeng, Maja Žumer and Athena Salaba ; IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR). Berlin/München: De Gruyter Saur,

2011. Final report available at: www.ifla.org/files/classification-and-indexing/functional-requirements-for-subjectauthority-data/frsad-final-report.pdf.
- [15] Open Metadata Registry. No date. Available at: <http://metadataregistry.org>.
- [16] Dunsire, Gordon. UNIMARC, RDA and the Semantic Web. In: International Cataloguing and Bibliographic Control (ICBC), vol. 39, no. 2 (April/June 2010). Based on a paper presented to the World Library and Information Congress: 75th IFLA General Conference and Assembly, 23-27 August 2009, Milan, Italy; available at: www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf.
- [17] UNIMARC Manual: Bibliographic Format / edited by Alan Hopkinson. 3rd edition. München: Saur, 2007.
- [18] UNIMARC Manual: Authorities Format / edited by Mirna Willer. 3rd edition. München: Saur, 2009.
- [19] Willer, Mirna. Foreword to the third edition. // UNIMARC Manual: Authorities Format / edited by Mirna Willer. 3rd edition. München: Saur, 2009. Pp. 7-9.
- [20] International standard bibliographic description (ISBD). Consolidated edition. Draft as of 2010-05-10. Available at: www.ifla.org/files/cataloguing/isbd/isbd_wwr_20100510_clean.pdf. Standard edition is expected to be published by IFLA 2011.
- [21] Semantic Web Education and Outreach (SWEO) Interest Group. Cool URIs for the Semantic Web. 2008. Available at: <http://www.w3.org/TR/cooluris>.
- [22] IFLA Namespaces Task Group. IFLA namespaces - requirements and options. 2010. Available at: http://www.ifla.org/files/classification-and-indexing/ifla-namespaces-requirements-optionsreport_corrected.pdf.
- [23] Library of Congress. No date. MARC list for languages. Available at: <http://id.loc.gov/vocabulary/languages.html>.
- [24] Library of Congress. No date. MARC list for geographic areas. Available at: <http://id.loc.gov/vocabulary/geographicAreas.html>.
- [25] Library of Congress. No date. MARC list for countries. Available at: <http://id.loc.gov/vocabulary/countries.html>.
- [26] Martin, Earle. ISO 3166 RDF representation. 2005. Available at: <http://downlode.org/Code/RDF/ISO-3166>.
- [27] Library of Congress. Bibliographic Framework Transition Initiative. 2011. Available at: <http://www.loc.gov/marc/transition>.