
UNIMARC 到 RDF 项目：命名空间和关联数据

Mirna Willer

信息科学学院，扎达尔大学，扎达尔，克罗地亚
E-mail 地址: m.willer@unizd.hr

Gordon Dunsire

独立咨询师，爱丁堡，苏格兰
E-mail 地址: gordon@gordondunsire.com

Predrag Perožić

信息科学学院，扎达尔大学，扎达尔，克罗地亚
E-mail 地址: pperozic@gmail.com



版权属于 Mirna Willer, Gordon Dunsire, Predrag Perožić。本著作在创作公用署名 3.0 许可的条款下可以使用：<http://creativecommons.org/licenses/by/3.0/>。

中文翻译：刘华梅（中国国家图书馆中文采编部）
Chinese traslation by LIU Huamei (National Library of China)

文摘：

本文描述了用资源描述框架（RDF）表示 UNIMARC 的项目工作，这是语义网和关联数据的基础。当前的重点是基于 UNIMARC 书目格式，为字段标识符、子字段以及值词汇的编码信息块开发一套元素集。本文讨论了项目确定的改进 UNIMARC 标准的问题，特别是使其符合 ISBD 以及像 RDA（资源描述和检索）这样的书目标标准。本文还给出了来源于 UNIMARC 记录的关联开放数据的实例。

关键词：UNIMARC，资源描述框架（RDF），ISBD，映射，互操作

1 简介和背景

在波多黎各圣胡安举行的世界图书馆和信息大会：第 77 届国际图联大会上提交的，随后在 IFLA 杂志上发表了更新版本的一篇文章（Dunsire & Willer, 2011），讨论了用资源描述框架（RDF）

表示 UNIMARC 书目 (UNIMARC/B) 和 UNIMARC 规范 (UNIMARC/A) 编码格式的一个基本框架, 这是语义网和关联数据的语法基础。文章确定了具体问题, 并就解决这些问题和开发命名空间来容纳 UNIMARC 元素集和值词汇提出了一些建议, 这些值词汇遵循为其他 IFLA 书目标准建立的模式, 包括功能需求 (FR) 系列模型和国际标准书目著录 (ISBD)。UNIMARC 永久委员会 (PUC), 负责 UNIMARC 格式的维护, 在 2012 年已经同意继续开发这样的命名空间。虽然向 IFLA 专业委员会提出的一个申请没有被通过, 但 PUC 能够找到并分配足够的资金支持 2013 年发起的一个项目 (PUC, 2012)。

该项目的首要点是开发 UNIMARC 书目格式的命名空间。然而, 很明显, 将命名空间和其他相关标准分离, 只孤立地分析其格式, 将会产生一个片面的结果, 因为项目并不提供格式功能的信息前景。不过, 关系到 IFLA 和其他相关标准的格式配置, 会额外地将格式本身的差异和发展潜力反馈给开发者。本文考虑了 UNIMARC 书目格式到 ISBD 的关系, 考虑了 ISBD 统一版中记载的对应性的影响, UNIMARC 书目格式和 UNIMARC 规范格式的关系, 以及更广泛的层面, 他们和资源描述和检索 (RDA) 的匹配。

在这篇文章中, 语词 “元素集” 和 “值词汇” 遵循 W3C 图书馆关联数据孵化小组推荐的用法 (Isaac and others, 2011)。

2 命名空间创建的基本方法

UNIMARC 格式的文档只适用于机器可读形式, 如 Microsoft Word 或 Adobe RDF 文件。这些缺乏自动解析元素集和值词汇所需数据的必要结构, 像标签、定义和范围注释。然而, 为了人类识别这样的数据, 在布局上必须要有足够的结构。从文件中抽取数据的基本方法是人力介入的复制、粘贴, 及随后的编辑。

能获取的最精细的粒度, 在元素集情况下是 UNIMARC 子字段代码层, 在值词汇情况下是符号代码和相应语词。子字段是元素中最小的代码单元。子字段通常是和字段或字段标识符聚合在一起的, 遵循 ISO 2709 标准的三位数编码 (ISO, 2008)。不过, 一个字段标识符使用一个或两个指示符能更改子字段和字段标识符本身的语义。一个 UNIMARC 子字段的完整语义需要包括它的字段标识符和两个指示符。本项目使用的方法假设这是默认情况, 因此需为每个允许组合的字段标识符、指示符和子字段创建一个元素。例如, UNIMARC 书目格式 200 字段标识符¹ (题名与责任说明项) 有一个子字段, 编码为 “\$a”, 用于正题名。虽然第二个指示符没有使用, 第一个指示符可以取两个值中的一个, 以区分题名有检索意义或无检索意义的情况。题名的检索意义, 也就是正题名, 可以通过创建两个元素获取, 一个题名是有检索意义的, 另一个是无检索意义的。

如果一个字段标识符中使用了两个指示符, 则为每个子字段创建的元素数是每个指示符取值数的乘积; 如果第一个指示符可以取 3 个值, 第二个指示符可以取 2 个值, 则可创建 6 个元素。这样就确保每种不同的含义都能容纳。为了减少手工从 UNIMARC 文档中抽取所需数据的时间, 可使用一个扩展表来存储字段标识符中每个子字段的数据。字段标识符的行集用于重复记录第一指示符的不同允许值, 整个块重复的行集用于重复记录第二指示符的每个允许值。当每个重复的子字段集合中的指示符的值发生了变化了, 可快速使用接口实现在扩展表中复制单元格。每一行包含字段标识符, 指示符值, 和子字段代码, 连同对应的字段标识符、指示符和子字段名

¹ UNIMARC 书目, 200 题名与责任说明项

称。通过扩展表公式利用这些数据可以创建一个有点人为的但人类可读的表示子字段的 RDF 属性标签。例如，正题名的两个元素可以有这样两个标签“题名与责任说明项中的正题名（题名是有检索意义的）”，“题名与责任说明项中的正题名（题名是无检索意义的）”。

同样地，每个元素的统一资源标识符（URI）也是使用扩展表公式导出，连接字段标识符，指示符值和子字段代码字母形成一个部分的 URI，其在 UNIMARC 编码方案中是唯一的。URI 其余部分的最终形式，即命名空间的基础，在整个数据款目完成前不会被最终确定。例如，正题名 RDF 属性的 URI 可能分别是 http://iflastandards.info/ns/unimarc/unimarcb/elements/2XX/U2001_a 和 http://iflastandards.info/ns/unimarc/unimarcb/elements/2XX/U2000_a。下划线（_）是用来标记第二指示符的位置，在这个字段中是不使用的。不过这个例子并不是严格必需的，它允许从 UNIMARC 书目记录使用的实际编码自动推导 URI。这种方法可以用一个简单的计算机程序把任何 UNIMARC 记录转换成三元组数据。

正题名的两个不同属性可以通过增加第三个属性而“结合”在一起，如用 URI http://iflastandards.info/ns/unimarc/unimarcb/elements/2XX/U200__a，和标签“题名与责任说明项中的正题名”；也就是说，忽略了第一指示符的取值。第三个属性也可以申明另外两个属性的上级属性。这就允许自动生成三元组数据，将前两个属性“简化”成上级属性，而丢失了由指示符取值引起的检索意义的区分。需要区分检索意义的应用程序可以使用原始的三元组数据；不需要区分检索意义的程序使用简化的三元组。

这一“子属性阶梯”技术还可以扩展用于其他用 RDF 命名的书目格式。图 1 展示了来自 UNIMARC, ISBD 和 RDA 的正题名元素势图，只用了 RDF 语法（RDFS）的“subPropertyOf”属性。

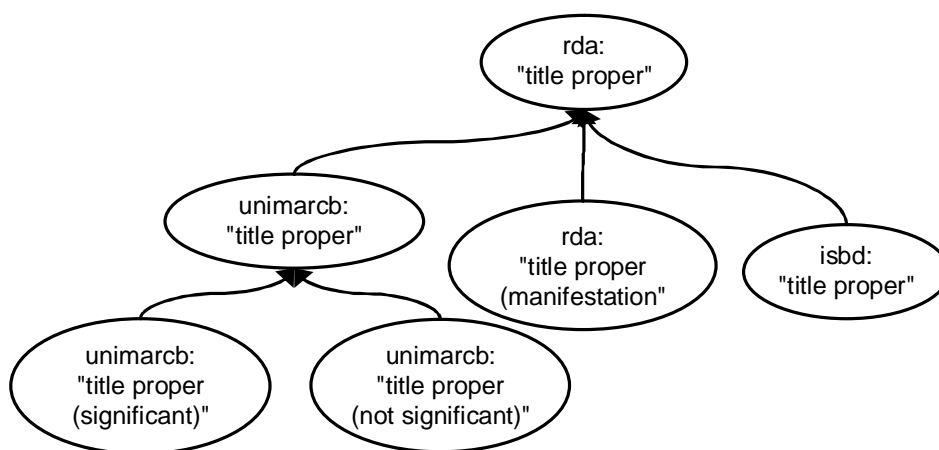


图1: 来自 UNIMARC, ISBD 和 RDA 元素集“正题名”的 RDF 图, 所有谓词是 RDFS 的 sub-property of” 属性, 为清楚起见, URIs 都是用缩写取代的。

为了清楚和简洁，图 1 中的 URIs 缩写成“基础”部分，例如“unimarcb”代表“http://iflastandards.info/ns/unimarc/unimarcb/elements/2XX/”，以及“局部”部分，例如，ISBD 的“P1004”，人类可读的标签如“正题名”用引号或方括号括起来。本文其余部分将使用类似的约定。

图 1 中的曲线图是基于 ISBD 和 RDA 元素集建议映射使用的无界的或“无约束”的属性(Dunsire, 2011; Dunsire & ISBD 评估小组, 2012)。这些映射也可以使用集成声明的元素, 所以本项目在扩展表中包括字段级聚合的数据, 以及在子字段代码普通模式存在的更精细的聚合。

从扩展表公式导出的 URIs 和标签完成后将被上传到 OMR (开放元数据注册)。上传还包括属性的定义和范围注释。要确定这些并不是直截了当的, 还需要大量的编辑干预。

2.1 转录问题

命名空间的其余内容, 像定义和范围注释能被复制和粘贴到扩展表或任何 RDF 词表管理系统, 但是这样做只能选取文档内容外的文本。当提交与词表其余部分结合时, 人类很容易从嵌入的范围注释、文本格式或用法说明中识别出定义, 但是分离时这种嵌入文本大多会变成噪声、模糊或过度复杂的元素的定义。定义和范围注释需要易读, 易懂, 明确和独立。

本项目已经确定了影响 UNIMARC 文档到命名空间转录的几种情况类型:

- 定义通常包含一些说明性短语如“包括”, “不包括”, “例如”, “如”等。一个实例: *unimarcb:U2000_a* 的定义“文献的主要题名, 包括交替题名, 但不包括其他题名信息(如副题名)和并列题名”。这些短语被视为其他 IFLA 标准转录过程中元素的范围注释部分。
- 定义直接引用其他元素。一个实例: *unimarcb:U2000_d* 的定义“出现在\$a或\$c子字段的正题名的另一种语言和/或文字的题名”。引用是有语法的, 使用字段、子字段代码和引用结构, 需要从定义的语义中分离。它们是应用程序配置文件中的正确表示。
- 定义包含用法短语。一个实例: *unimarcb:U2000_v* 的定义“用于标识一篇文献是另一篇文献的分卷号”。脱离了人工环境这些语词使定义更难阅读和理解。
- 定义包含格式信息。一个实例: *unimarcb:U2000_z* 的定义“出现在\$d子字段的并列正题名的语种代码标识”。怎样表示内容并不是元素的语义部分。这是一个和嵌入语法信息类似的问题。这种信息在领域和范围约束和应用程序配置文件中是最好的表示。

这些都需要人类详细审查通过编辑创建的 RDF 属性定义和注释的格式手册文本。这是费时的, 但最终会告知手册本身的未来发展。

3 匹配 UNIMARC 书目和 ISBD

2XX 区或块²包括著录字段, 字段标识符范围从 200 到 225, UNIMARC 书目格式是根据 ISBD 规定构建的。分析 UNIMARC 书目和 ISBD 之间的对应性, 和遵循一个或其他 IFLA 标准把数据发布为关联数据的方法和做法是相关的, 可以投入到互操作关系。不过, 应当指出, 这里有一个问题, 因为 UNIMARC 书目格式不提供有关记录内容的规则: 它只提供了一个指示符说明著录数据元素是否符合 ISBD 规定, 因此允许创建容纳格式上不遵循 ISBD 做法的数据。例如, UNIMARC 书目录头标字符位置 18 (著录格式)³取值有完全的, 部分的或非 ISBD 的。另外, 虽然它提供 ISBD 数据元素字段/子字段的定义, 但那些定义指的是 ISBD 文档⁴。

² UNIMARC 书目, 2—著录信息块

³ UNIMARC 书目, 记录头标, 18 著录格式

⁴ UNIMARC 书目, 1.3 定义

这在解释表1中给出的对应集合也是特别重要地，表1中的对应关系是基于各自定义的语义一致性，在含义上用相等（“=”），更宽泛（“>”）或更狭窄（“<”）分类。这里还有两个其他问题应该考虑：一是格式本身的特点和它指示内容的方法；二是考虑著录数据在资源中出现顺序的特点，因此应该在记录转录时约定好。虽然后一个问题被认为是开发和实施应用程序配置文件的重点（Willer,Dunsire & Bosančić,2010），下面还将展示它与命名空间元素的定义和对应性的分析也是相关的。

UNIMARC			ISBD	
属性	标签	A	属性	标签
200 2000_ 2001_	题名与责任说明项	=	P1159	题名与责任说明项
			P1170	题名说明
			P1012	题名
200__a	正题名	= <>	P1004	正题名
			P1117	同一作者的单个作品题名
			P1137	正题名的共同题名
200__b	一般资料标识			
200__c	其他责任者的正题名	=	P1118	不同作者的单个作品题名
200__d	并列正题名	=	P1005	并列题名
			P1182	并列题名的共同题名
			P1183	并列题名的从属题名
			P1184	并列题名的从属题名标识
200__e	其他题名信息	=	P1006	其他题名信息
			P1140	并列其他题名信息
200__f	第一责任说明	>	P1007	题名相关的责任说明
200__g	其他责任说明	<	P1007	题名相关的责任说明
			P1141	题名相关的并列责任说明
200__h	分辑（册）号	=	P1139	正题名的从属题名标识
200__i	分辑（册）名	=	P1138	正题名的从属题名
200__j	起讫日期			
200__k	批量日期			
200__r	正题名的题名页信息 (用于旧的专著出版物)			
200__v	卷标识			
200__z	并列正题名语种			
200__5	使用本字段的机构			

表1: UNIMARC书目格式与ISBD的对应，字段标识符200（草图）

区块中的第一个子字段，*unimarc*b:U200__a(正题名)是提到的三个问题的一个好例子。这是字段中的第一个数据元素，它的定义和ISBD中同样元素的定义是相对应的；因此它们在含义上是

“相等的”。但事实上，它同时也可以被认为在含义上比ISBD元素宽泛和狭窄。也就是说，UNIMARC书目格式最经济的指定目录的方法是指定第一个来自著录资源上下文的为第一个数据元素：不仅仅是“正题名”和“正题名中的共同题名”，也可以是“同一作者的单个作品题名”。“共同题名”可以和ISBD等同对应，也可以用一种算法计算以检测到同一字段标识符中存在的\$h子字段（分册号）或\$i子字段（分册名），即使方法不是100%准确。不过，在关联数据环境中每一个来自记录的数据三元组都潜在地和其他每一个三元组相关联的，这一计算方法在三元组发布之后是不合理的。同样的事实对第二种情况，元素“同一作者的单个作品题名”重复在\$a子字段编码。从这个例子我们可以推断在双方对应过程中信息价值已经被丢失：在UNIMARC书目到ISBD的映射中，UNIMARC书目属性用URI *unimarcb:U200__a*把*isbd:P1004*，*isbd:P1137*和*isbd:P1117*属性的含义都包含在内，反过来ISBD向UNIMARC映射时，分离的ISBD属性在一个UNIMARC属性中失去其特殊性。在应用程序配置文件实施的情况下，要尽可能的对应重复的\$a子字段到*isbd:P1117*，但这并不适用于*isbd:P1137*“共同题名”的情况。

在并列题名数据元素情况下这种对应缺失也是很明显的。对这种内容类型只有一种UNIMARC属性：*unimarcb:U200__d*（并列正题名）。它的定义是“出现在\$a或\$c子字段的正题名的另一种语言和/或文字的题名”，通过这一定义，意味着它的语义是和*unimarcb:U200__a*和所有重复子字段，及*unimarcb:U200__c*（其他责任者的正题名）相关的。只有数据元素的序列或顺序，是从资源转录的和通过应用程序配置文件处理的，可以说“*unimarcb:U200__d*是和这两个或更多题名属性对应的。虽然*isbd:P1005*（并列题名）的定义并没有说明和哪种正题名相关，但实例显示了含义是等同的。UNIMARC书目格式中的所有其他并列数据也在相关子字段予以揭示：UNIMARC书目规定“如果遵循ISBD规则在任何其他子字段的开始加‘=’，就必须明确记入”⁵。换句话说，UNIMARC书目元素通过其内容语法进行语义细化（“简化”的反义词），因为MARC代码本身不足以指定格式化数据的语义。实例如下说明⁶：

200 1#\$aBibliographica belgica\$fCommission belge de bibliographie\$f= Belgische
Commissie voor bibliografie

这里，子字段\$f的数据“Commission belge de bibliographie”和*isbd:P1007*（题名相关的责任说明）有着同样的语义，而子字段\$f的数据“Belgische Commissie voor bibliografie”和*isbd:P1141*（题名相关的并列责任说明）有着同样的语义。这一规定意味着UNIMARC书目命名空间应该包含“200__f=”的属性，字面是“并列属性”，除了*unimarcb:U200__a*和*unimarcb:U200__c*之外的其他所有情况。为了支持和ISBD并列数据属性的对应这些都应该创建。如果不这样做，ISBD到UNIMARC书目的对应将引起数据并列信息值的丢失。必须说明在处理所有其他2XX字段给出的并列数据时都要使用同样的规定。

我们还应该考虑UNIMARC书目中定义的处理：正如已经提到的，手册并没有定义ISBD数据元素，但提到了ISBD定义的特性。UNIMARC书目格式的第一版，和当前版本的简洁版采用同样的样式，只命名了字段和子字段数据元素而没有提供定义。接下来的版本命名为UNIMARC Handbook：书目格式，随后又命名为UNIMARC Manual：书目格式，基于同时代的ISBD版本都包含了定义。UNIMARC书目字段/子字段定义的状态应该考虑，从统一版取代ISBD(总则)和七个专门的ISBDs而引起的ISBD数据元素定义变化相关的格式维护方面，以及从指引用户到另一文档这一通用目的方面都应该考虑。表1中显示的UNIMARC/ISBD的对列，事实上，只有

⁵ UNIMARC 书目，200 题名与责任说明项，并列数据

⁶ UNIMARC 书目，200 题名与责任说明项，例 6

unimarcb:U200__a, *unimarcb:U200__f*和*unimarcb:U200__g*没有被认为在语义上和相应的ISBD属性是对等的。*unimarcb:U200__a*的情况在本文前面已经讨论过,然而另一种情况,很有必要区分第一和随后的责任说明,因为这两种数据元素所需的不同ISBD标点符号。不过,含义的不同分类,使得状态更模糊,第一责任说明不仅和正题名相关,还和UNIMARC书目所列举的其他元素相关,如:“题名的第一责任说明出现在\$a,\$c或\$d子字段,著作的分册号或分册名出现在\$h或\$i子字段”。ISBD的实例证明了同样的处理方法;或者,更好的说法,UNIMARC遵循ISBD规定的“用尽可能最好的方式”。此外,ISBD认为“第一责任说明和随后的责任说明的不同只是顺序的事情”,只把它作为单一的重复元素处理,因此没有单独的RDF属性来区分首次出现。如果要当作单独的元素表示,它就应该和*unimarcb:U200__g*对应,数据的信息价值在两个方向的对应是等同的。

在开发UNIMARC命名空间开始就应该决定不再重复使用ISBD命名空间的元素。和从FR系列模型识别上下文是不同的,这里同样元素的RDF属性可用于三种模型,例如类可以是“作品”或“个人”。基本的论点是每套元素集本身应该是一致的和完整的,因为要使它们能独立维护和进一步发展。这一决定是合理的,因为和其他UNIMARC书目2--字段对应时会显示在200字段标识符中提到的同样的问题。希望UNIMARC书目格式遵循ISBD统一版的变化而更新已存在的字段和子字段定义是不现实的,但是在那些新的方面是要遵循的,如区域0内容形式和媒介类型,因为格式是数据容纳器的基本。在任何情况下,例如,元素*unimarcb:U200__a*,当记录包含的数据遵循非ISBD著录做法时可以被用作任何其他目的。

这里应该提到字段或子字段名称变化引起的格式上的某些变化,因为它们直接影响UNIMARC书目RDF属性域的定义。特别是ISBD(ER)术语的变化,从“出版物/文献”到“资源”影响了207字段标识符⁷和206字段标识符⁸的重新命名。字段名称从“资料”变为“资源”。207字段和子字段的定义用“资源”取代了“文献”,206子字段的定义保留使用“文献”这一术语。应当指出,除了207字段,UNIMARC只在电子资源、连续性资源和整合资源的情况下遵循了语词的变化,因此术语“资源”被认为是著录特定类型资料的最终学术语。

2XX块还显示了与发展和维护各自文档相关的另一个问题。在ISBD统一版中“一般资料标识”(GMD)元素从ISBD元素1.2中移除了,并用新的ISBD区域0内容形式和媒介类型取代了。新版本对元素进行了重新编号,元素1.2从GMD变成了“并列题名”元素,等等。UNIMARC2012更新没有反映这一重新编号,而是保留了GMD的200字段\$b子字段。格式不能删除一个定义元素,预期它会在传统记录中使用;它只能使子字段废弃,不过,PUC并没有这样做。同时,还应该说明,ISBD评估小组也没有考虑选择不宜用的元素,也没有重新使用位置号码,因为这种情况同样也出现在UNIMARC书目格式;要考虑遗留下来的数据,以及这些数据将产生的连续性做法,为了一个原因或另一个原因就不能遵循新的定义区域。UNIMARC书目230字段标识符⁹也有同样的情况,它被定义为和ISBD(ER)类型和扩展资源(区域3)是等同的,但在ISBD统一版中却从同一区域删除了。而PUC并没有废弃这一字段。

⁷ UNIMARC 书目 207 资料特定细节项: 连续性资源的卷期编号

⁸ UNIMARC 书目 206 资料特定细节项: 测绘制图资料的数学数据

⁹ UNIMARC 书目 230 资料特定细节项: 电子资源特征

4 两种格式，两种命名空间？

UNIMARC规范格式和UNIMARC书目格式之间的对应性是基于结构兼容性，主要原因是两种类型记录的目的是在集成图书馆系统一起使用。这就意味着同一检索点的数据元素出现在两种格式的相同子字段，而字段标识符和字段名称因为书目记录和规范记录的不同功能而不同。

这种对应性在UNIMARC规范使用指南中有记载；从第三版截取的部分对应表，如图2所示。

		Guidelines for Use
(10) Correspondence Between UNIMARC/Authorities and UNIMARC/Bibliographic		
UNIMARC/Authorities Access Point Fields		Access point Usage in UNIMARC Bibliographic Fields
200	Personal name	700, 701, 702 4-- with embedded 700, 701, 702 600 604 with embedded 700, 701, 702
210	Corporate or meeting name	710, 711, 712 4-- with embedded 710, 711, 712 601 604 with embedded 710, 711, 712
215	Territorial or geographic name	710, 711, 712 4-- with embedded 710, 711, 712 601, 607 604 with embedded 710, 711, 712
216	Trademark	716
217	Printer/Publisher device	717 [to be defined]

图2：UNIMARC规范和UNIMARC书目对应表（部分）

在字段这一层面，两种格式的不同在于怎样为团体检索点的关系归类；也就是说，UNIMARC书目中三个分离的团体名称标签¹⁰，对应UNIMARC规范中的210字段标识符¹¹和215字段标识符¹²。71X字段标识符和210字段标识符中的子字段名称和定义在含义上是等同的，而215字段标识符被定义为“包含一个地理区域或地名检索点[.....]只有地理区域名称或只作为主题复分的附加成分时被认为是地理区域名称(215字段)；地理区域名称后接团体复分时被认为是团体名称(210字段)”。因此，UNIMARC书目的71X字段标识符和UNIMARC规范的215字段标识符的关系归类是有争议的。

子字段名称和定义之间的对应遵循前面提到的格式的结构和功能情况下被归为等同对应。有一个不同点是UNIMARC书目500字段标识符“首选检索点”和UNIMARC规范230字段标识符“规范检索点”的字段名称不同。一般来说，UNIMARC规范的术语遵循FRAD，而UNIMARC书目更

¹⁰ UNIMARC 书目 710 团体名称—主要责任，UNIMARC 书目 711 团体名称—等同责任，UNIMARC 书目 712 团体名称—次要责任

¹¹ UNIMARC 规范 210 规范检索点—团体名称

¹² UNIMARC 规范 215 规范检索点—地名或地理区域名称

接近RDA。两种UNIMARC格式的开发和维护已经实现了，到2012年每种格式都更新到了第三版。

两种格式的结构兼容性反映在它们的代码上，这驱动了它们之间的语法对应。子字段和字段标识符层面的语义对应只是局部的，决定为每种格式创建命名空间是合理的。

不过，语法对应的元素也是语义对应，UNIMARC书目扩展表的相关部分可以被UNIMARC规范循环利用，这样就节省了开发命名空间的大量时间。

5 UNIMARC数据作为RDF关联数据的实例

UNIMARC书目1XX块用编码作为数据值，编码本身含有标题或标签，有时是定义。每一套编码作为值词汇能被表示成RDF，用一个相关的编码为每个概念分配一个URI。这样的词汇实例是连续性资源的出版周期。每种周期值有其自己的编码，存储在UNIMARC数据记录110字段标识符的字符位置¹³。编码数据字段用固定的字符位置，而不用子字段界定元素，因此URI模式有些微地不同。因此“出版周期”元素有URI *unimarcb:U110__al*。作为项目的部分值词汇在OMR中已经加以表示（UNIMARC 出版周期，2013）。

这就允许关联三元组数据通过URI编码值发布。例如，一条日报（daily newspaper）的UNIMARC记录将包含一个110字段标识符，字符位置的编码值如表2所示。

字符位置	值	说明
0	c	newspaper
1	a	daily
2	a	regular
3	#	n/a
4-6	###	n/a
7	0	Not conference proceedings
8	x	n/a
9	x	n/a
10	0	No cumulative index, etc.

表2：日报110字段标识符的编码值

图3展示了前三个字符的关联数据图。周期“daily”的URI关联到编码或符号“a”，首选标签有英语、意大利语和葡萄牙语。

¹³ UNIMARC 书目 110 编码数据字段：连续性资源 \$a/1 出版周期

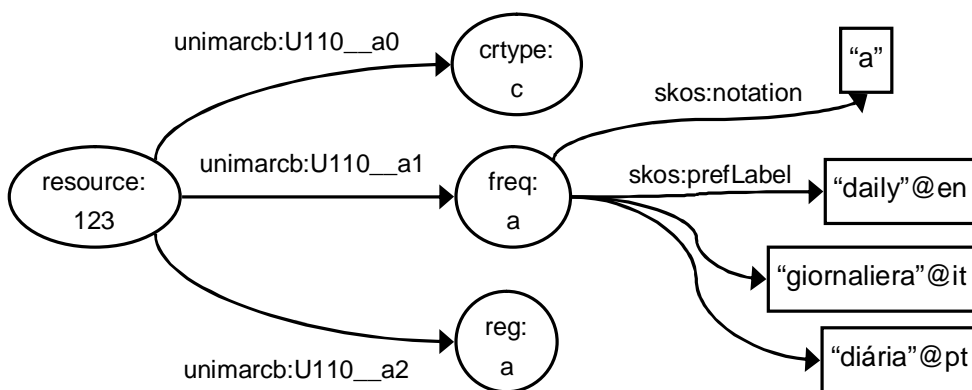


图3：日报 UNIMARC 记录编码信息块部分的 RDF 图

图 3 中的曲线图可以增加连续性资源类型代码“c”和出版规律代码“a”的符号和首选标签的相似关联，当它们的值词汇在项目稍后阶段发布时。UNIMARC 使用 RDF 语法的代码信息比使用自由文本标签更兼容。这在多语言环境有直接的好处，如图 3 所示，使用 RDF 内置语言识别系统。

6 映射值

对于像期刊和文集这样的连续性资源的出版周期，除了 UNIMARC 值词汇，OMR（开放元数据注册）还包括其他三种词汇，DC 集合级描述，MARC21 和 RDA（资源描述和检索）。从所有四种词汇中检索“daily”的结果，如图 4 所示。

The screenshot shows the Open Metadata Registry logo and a search results table for the term 'daily'. The table lists four results from different vocabularies, each with a concept, label, SKOS property, and language.

Search results for 'daily'				
Vocabulary	Concept	Label	SKOS property	Language
Dublin Core Collection Description Frequency Vocabulary	Daily	Daily	preferred label	en
MARC21-008: Frequency of continuing resource	daily	daily	preferred label	en
RDA Frequency	daily	daily	preferred label	en
UNIMARC: Continuing resources: Frequency of issue	daily	daily	preferred label	en
4 results				

图 4：开放元数据注册系统注册的值词汇检索“daily”的结果截屏
http://metadataregistry.org/conceptprop/search?concept_term=daily.

如果相似概念能被关联起来那么基于这四项标准的关联数据互操作性就能有所改善。为了决定

每种词汇中“daily”概念之间的关系，很有必要检查它们的定义和范围注释：

- UNIMARC 语词没有定义或范围注释。
- DC 语词有定义“一天发生一次的事件”。语词来自 MARC21 馆藏，853-855—标识说明及样式通用信息，子字段\$w。注意这是第五种词汇，虽然没有用 RDF 表示，因为子字段和 MARC21 书目有不同的语境。
- MARC21 语词有定义“一天一次”和范围注释“包括星期六和星期天”；这一信息来自 MARC21 书目手册（MARC21 2010）。
- RDA 语词有定义“资源发布或更新的频次每天一次，通常不包括非工作日”。也就是说，RDA 概念不包括星期六和星期天。

在没有定义的情况下，语词“daily”的本地用法建议 UNIMARC 概念覆盖每周 7 天，但这不应该并且也是不安全的就认为可以和 MARC21 概念精确匹配。DC 和 MARC21 概念比 RDA 概念宽泛，RDA 只包括每周 5 天。简单知识组织系统（SKOS）命名空间（SKOS 2009）的映射关系属性能用于创建相关三个概念的 RDF 图。用简洁的三元组语言（Beckett & Berners-Lee, 2011），序列如下：

```
@prefix cld: <http://purl.org/dc/cld/freq/> .
@prefix marc21: <http://marc21rdf.info/terms/continuingfreq#> .
@prefix rda: <http://rdvocab.info/termList/frequency/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix unimarc: <http://iflastandards.info/ns/unimarc/terms/continuingfreq#> .
# daily
unimarc:a skos:closeMatch cld:daily .
unimarc:a skos:closeMatch marc21:d .
cld:daily skos:closeMatch marc21:d .
rda:1001 skos:broadMatch cld:daily .
rda:1001 skos:broadMatch marc21:d .
rda:1001 skos:broadMatch unimarc:a .
```

相反的关系能自动推导如下：

```
cld:daily skos:closeMatch unimarc:a .
marc21:d skos:closeMatch unimarc:a .
marc21:d skos:closeMatch cld:daily .
cld:daily skos:narrowMatch rda:1001 .
marc21:d skos:narrowMatch rda:1001 .
unimarc:a skos:narrowMatch rda:1001 .
```

完整的 RDF 图如图 5 所示。

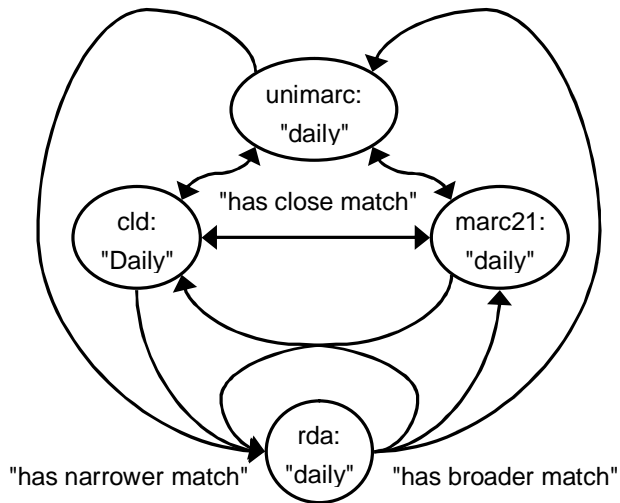


图 5: 来自 MARC21, RDA 和 UNIMARC 值词汇的出版周期“daily”的 RDF 图。为了清楚用 URIs 取代标签。

图 5 是“daily”来自三个值词汇的概念图。它可以扩展到其他词汇，像 DC（DC 集合描述工作组，2007）开发的集合自然增长周期性。这个图可以和图 3 结合，为书目关联数据全球互操作提议新的可能性。

7 未来工作

本项目正在进行，最初计划的重要工作安排在余下的 2013 年和 2014 年。这包括为编码信息块的所有编码集开发值词汇，和 UNIMARC 命名空间的内部映射，从 UNIMARC 到外部元素集和值词汇的映射，基于本文讨论的实例。本项目已确定的另外重大区域需要在项目和 UNIMARC 环境中的进一步工作。

在 UNIMARC 中，处理嵌入在记录数据中的语法和语义元数据，需要用 RDF 建模。嵌入式语法元数据的一个实例是非归档字符指示符，而嵌入式语义元数据的一个实例是使用“=”表示并列数据。在项目未来计划阶段，随着子字段聚合的表示这将不得不考虑。

考虑 UNIMARC 书目和 ISBD，起草 UNIMARC 书目到 ISBD 对应表，应该扩展到 ISBD 7 附注项和 ISBD 8 资源标识号和获得方式项，ISBD 到 UNIMARC 书目的对应表也应该开发。这一过程将不得不考虑开发不受限制的 UNIMARC 属性的必要性，也就是说，没有 RDF 领域或范围。对应实践揭露了 UNIMARC 书目和 ISBD 各种问题，应考虑维护机构之间的密切合作。希望 ISBD 到 UNIMARC 书目对应工作中将揭露更多其他问题。同样，字段标识符 2XX 子字段和 ISBD 之间的 UNIMARC 书目对应表应该被更新，以对应 ISBD 统一版的元素数。

两种 UNIMARC 格式结构的复杂性，满足它们被兼容的目的，同时考虑它们各自的功能，为 UNIMARC 书目 500 字段标识符、7XX 子字段和 UNIMARC 规范本身发布 RDF 属性的过程将进一步展示。建议工作并行进行。在 UNIMARC 书目 6XX 字段标识符主题数据元素的其他问题也将被揭露，将在 UNIMARC 规范子字段的层面处理，因为 UNIMARC 规范是一个综合性的名称/标题和主题规范格式。

8 参考文献

Beckett, David, and Tim Berners-Lee (2011). Turtle - Terse RDF triple language. Available at: <http://www.w3.org/TeamSubmission/turtle/>

Dublin Core Collection Description Task Group (2007). Dublin Core collection description frequency vocabulary. Available at: <http://dublincore.org/groups/collections/frequency/2007-03-09/>

Dunsire, Gordon (2011). Mapping ISBD and RDA element sets: briefing/discussion paper. Available at: <http://www.rda-jsc.org/docs/6JSC-Chair-4.pdf>

Dunsire, Gordon, and Mirna Willer (2011). UNIMARC and Linked Data. IFLA Journal, 37, 4 (December 2011) (pp314-326). Available at: http://www.ifla.org/files/hq/publications/ifla-journal/ifla-journal-37-4_2011.pdf

Dunsire, Gordon, and ISBD Review Group (2012). Alignment of the ISBD: International Standard Bibliographic Description element set with RDA: Resource Description & Access element set. Version 1.1. Available at: <http://www.rda-jsc.org/docs/6JSC-ISBD-Discussion-1-Alignment.pdf>

Hopkinson, Alan (ed.) (2008). UNIMARC Manual: Bibliographic Format. 3rd. ed. München: K. G. Saur.

IFLA, UNIMARC Core Activity, Permanent UNIMARC Committee Available at: www.ifla.org/unimarc/puc

IFLA, PUC (2012). UNIMARC Core Activity, Permanent UNIMARC Committee, Minutes of the Informal Meetings of the Permanent UNIMARC Committee, 13, 14 and 16 August 2012, IFLA Congress, Helsinki, Finland, Draft: 2012 October 31.

Isaac, Antoine, William Waites, Jeff Young, and Marcia Zeng (2011). Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets. W3C Incubator Group Report 25 October 2011. Available at: <http://www.w3.org/2005/Incubator/1ld/XGR-1ld-vocabdataset-20111025/>

ISBD (2011). ISBD : International standard bibliographic description. Consolidated ed. Berlin ; München : De Gruyter Saur.

ISO (2008). ISO 2709:2008: Information and documentation -- Format for information exchange.

MARC 21 (2010). 008 - Continuing Resources (NR). Available at: <http://www.loc.gov/marc/bibliographic/bd008s.html>

SKOS (2009). SKOS Simple Knowledge Organization System: Reference. W3C Recommendation 18 August 2009. Available at: <http://www.w3.org/TR/skos-reference/>

UNIMARC Frequency of issue (2013). UNIMARC: Continuing resources: Frequency of issue.
Available at: <http://metadataregistry.org/vocabulary/show/id/324.html>

Willer, Mirna (ed.) (2009). UNIMARC Manual: Authorities Format. 3rd. ed. München: K. G. Saur.

Willer, Mirna, Gordon Dunsire, and Boris Bosančić (2010). 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://leo.cilea.it/index.php/jlis/article/view/4536>