

CNMARC与BIBFRAME的映射基础 及其实现

许磊

(上海图书馆, 上海 200031)

摘要: 关联数据已经成为图书馆资源开放的重要技术手段。书目数据作为图书馆最核心的数据资源, 在网络中开放共享是必然趋势。关联书目数据发布的关键一步是MARC字段与本体词表之间的映射。本文在梳理元数据映射、元数据元素与知识本体之间的关系基础上, 对CNMARC字段与BIBFRAME本体词表的语义映射关系进行总结, 并以典型的字段映射进行说明。

关键词: CNMARC; BIBFRAME; 元数据映射

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2018.11.004

图书馆拥有丰富的结构化数据, 在互联网时代应当成为重要的信息分发机构。但现实是用户的第一信息获取途径已经不再是图书馆^[1-2]。在开放数据运动的推动下, 数据在互联网中的可发现才能发挥其更大的价值^[3]。而现有的图书馆数据是以不兼容于网络通行标准的MARC和ISO2709格式进行存储和交换, 这极大限制了图书馆数据的开放和共享, 也进一步边缘化了图书馆信息中心的角色地位。

自2006年关联数据概念被首次提出, 其开放、互联的属性天然地成为开放数据运动的技术助力。关联数据利用成熟的网络技术(如RDF、URI、HTTP等), 在互联网上发布和关联结构化的数据, 并提供HTML、API或SPARQL便于人机的访问。图书馆利用关联数据技术将数据发布到网络上有助于资源的可发现性和可利用性^[4]。在国外的文化遗产领域, 关联数据应用日渐成熟, 并形成一套从数据的建模、清洗、转换、发布、消费、评估全流程的实现路径与软件工具^[5]。相比国外的蓬勃发展, 国内研究虽已从理论的探讨或案例、技术、工具的介绍, 进入实践阶段, 但无论是参与机构数量还是数据集规模都仍显不足^[6]。上海图书馆(以下简称“上图”)是国内应用关联数据技术发布、开放、消费数据较为成熟的机构, 不仅设计了适用于不同文献类型的领域本

体, 也提供了基于关联数据技术的开放数据服务^[7], 并逐步建成面向知识服务的数字人文平台^[8]。上图也是国内唯一大规模发布关联书目数据的机构^[9]。2017年, 上图开始尝试将上海市文献联合编目中心的中文书目数据发布为关联数据, 探索关联书目数据的应用可能性, 研究BIBFRAME词表在中文编目环境中的适用性, 为编目技术、流程的变革积累经验与技术。

关联书目数据可以通过映射将结构化的MARC数据转换成高质量的元数据发布到网络上, 为全网域的潜在用户提供服务。中文书目数据的关联化, 先需要解决的是CNMARC与本体词表的映射与转换。白林林等^[10]在参考MARC21与RDF转换的基础上, 构建了CNMARC到RDF的映射实现。李勇文^[11]从理论层面探讨了BIBFRAME词表在中文书目数据中的研究范式。吴贝贝^[12]采用元数据映射方法, 形成了CNMARC字段与BIBFRAME词表的映射。周小萍^[13]、许磊^[14]在实践层面实现了CNMARC字段与BIBFRAME词表之间更具体的映射关系。但现有的研究没有对CNMARC与BIBFRAME之间的映射基础进行深入研究。

CNMARC与BIBFRAME之间的映射本质是两种描述标准的互操作。本文尝试在资源描述的语义互操作基础上, 对本体词表与元数据元素间的联系与区别进

行讨论,实现对CNMARC与BIBFRAME的映射基础更深入地分析,在此基础上实现两者之间的映射。

1 CNMARC与BIBFRAME的映射基础

1.1 资源描述的语义互操作

随着数字资源的快速增长,通用的或仅适用于特定领域的元数据标准被不断提出。元数据的多元化发展满足了领域内资源描述和管理的需求。但不兼容的元数据标准阻碍了跨系统的数据整合与服务。因此,促进跨系统数据交换与使用的元数据互操作性成为其重要原则之一^[15]。元数据互操作的主要实现方法有应用纲要、映射、复用、开放搜寻等^[16]。其中元数据映射是常用的方法之一。映射是一组系列的转换,目的是将存储在原始元数据标准中的元素内容进行适当的修正,以存储到目标元数据元素中。具体来讲,一个完整的元数据标准包括3个方面,即语义结构、内容结构和语法结构。语义是指元数据元素的定义与解释,结构是元素之间的关系,语法是元数据的形式化编码方案^[17]。对应于这3个部分,元数据互操作也需解决3个层面的问题。元数据映射不仅包含对元数据语义结构中元素定义、语义关系和编码规则的协调,也暗含对内容结构中各个元素关系的匹配。而对于语法结构的互操作,更多的是利用一种开放的可互操作的描述方法来对元数据进行置标。RDF就是一个标准的框架,通过设计支持语义、语法和结构方面的通用协议机制,满足人们对元数据的编码、交换和再利用^[18]。

现有的书目数据多以MARC格式著录,而关联数据是要用本体词表进行描述。关联书目数据的发布,就需要确定MARC元数据元素与本体词表属性之间的对应关系,即元数据元素与本体词表之间的映射。

1.2 本体词表与元数据schema的联系与区别

本体即领域本体,是对领域知识的抽象化、概念化,是共享的概念模型明确的形式化规范说明^[19]。它提供了描述领域知识的结构化术语词表。术语分为类(Class)和属性(Property)两种。类是抽象层面的泛指,通常是对一类实体对象共性的总结;属性是对类的各种特征的抽象,用于表示类之间的关系。属性在RDF三元组中是作为谓语,它连接主语和宾语。本体词表可

以通过RDF schema结构的属性公理定义属性的定义域(Domain)和值域(Range)特征,规定其连接的主语和宾语的类型。根据定义域和值域的不同,属性可以分为对象属性(Object Properties)和数据类型属性(Datatype Properties)。数据属性连接的是实体到数据值。对象属性连接实体到实体,其值域类可以作为另一个属性的定义域类,而这也正是RDF框架的优势所在,即灵活、开放、可扩展。

元数据是对某个潜在信息性对象做出的陈述。一个元数据陈述就是含有主谓宾的三元组陈述。在元数据模式中,谓语常被称为元素,它通过元素-值配对构成关于某个资源的唯一陈述^[20]。被描述的资源是主语,而宾语是用于描述主语的值。值的类型可以分成字符串数据和对象数据。这对应了本体词表中的2种类型属性。MARC是书目信息的元数据模式。一条MARC数据是对整个文献资源中各种信息对象的陈述集合。MARC模式中,字段子字段就是谓语元素。分解到每一个字段或子字段及其取值,就是对多个或单个信息对象做出的唯一陈述。MARC字段与本体词表的映射,就是对字段子字段的元素-值配对陈述与本体词表中定义的属性进行对比。

1.3 书目框架BIBFRAME与MARC的关系

MARC标准经过50多年的发展,其线性的结构、匮乏的语义、封闭的生态已不适用于开放互联的数据时代。图书馆领域一直在探索制定后MARC时代的书目标准。BIBFRAME即是这种探索的重要成果之一。BIBFRAME是美国国会图书馆牵头,以取代MARC为基本目标而开发的一种面向“未来网络世界中书目描述”的书目本体^[21]。BIBFRAME采用实体-关系模型对书目数据中实体及其关系进行分析和标识,并使机器能够理解这个模型。在2016年发布更新版中,BIBFRAME词表发布了包含3个核心类在内的共186个类,以及190个属性。3个核心类分别是作品(Work)、实例(Instance)和单件(Item)。最新的三层模型与国际图联统一模型(IFLA-LRM)的四层模型更加接近,增加了与IFLA-LRM和RDA的兼容性,更便于编目员进行编目操作^[22]。

作品层是抽象的内容层次,不指向特定的物理对象,是对不同题名不同表达形式的同一作品的聚合。BIBFRAME按照表达类型,将作品分为文本、地图、

数据集、静态图像、动态图像、音频、乐谱、舞谱、物体、多媒体、混合资料11种子类，帮助用户查找识别选择不同类型的作品。不同的作品拥有共同的主题、代理、事件等属性特征。实例是载体层次，是抽象作品的具体载体表现。BIBFRAME细分了印刷型、手稿、档案、触摸、电子型5种实例类型。实例层共享不同作品实例的版本、载体形态、出版发行等属性特征。单件是BIBFRAME2.0中新增的核心类，是实例的单一样本。BIBFRAME单件对应了IFLA-LRM和RDA的单件层，满足图书馆灵活、轻便、准确地记录馆藏数据^[23]。

BIBFRAME190个属性中有127个对象属性，63个数据类型属性。在BIBFRAME词表网页端分别使用“Used with”和“Expected Value”表示属性的定义域和值域。RDF三元组的主语必须是实体，即属性的定义域取值必须是类。当属性定义了值域，则该属性是值域类的专属属性，即属性三元组的主语只能是定义域类。而当属性没有定义值域，即“Used with”是“Unspecified”，说明属性三元组的主语没有限定，可以是任意的实体类。对于值域的不同，区分了两种不同类型的属性。数据属性的期待值是字符串(Literal)，对象属性的期待值是实体类。如对象属性bf:heldBy，表示文献单件的馆藏者。它的定义域是实体类bf:Item，值域是类bf:Agent。类bf:Agent可以作为其他属性的定义域，这样就可以进一步描述bf:Agent的属性和关系。

BIBFRAME和MARC是不同技术环境下对同一类对象进行描述的标准。前者定义本体词表，利用RDF三元组结构对书目数据中的实体对象进行属性和关系的描述；后者定义字段与子字段，利用元素-值的结构对书目数据进行整体性的描述。虽然两者在结构框架、交换机制等方面存在差别，但BIBFRAME继承性地取代MARC的初衷，就已暗含了两者的对应关系。即MARC格式可以通过映射与BIBFRAME词表建立对应关系，实现数据的重新编码与发布。

MARC标准以文献载体为著录对象，文献载体不同的特征利用不同的字段描述。一个字段即是对其某一方面特征的描述，其潜在语义结构是“…”“有/是/关于”“…”的主谓宾结构。而在关联数据语境下，文献载体不再是不可分割的整体，文献所含的对象与关系成为我们关注的焦点。因此，MARC格式与BIBFRAME词表之间的映射，需要对MARC字段进行拆解与分层，以辨别每个字段描述了哪一层对象的属性和关系。MARC格式通过字段的定义和范围说明其含义，而字

段含义隐含了对实体对象的指定。这个字段的实体对象的属性与关系对应到子字段的描述。子字段记录了具体的特征值，也就是对象的属性和关系。

2 CNMARC与BIBFRAME的映射实现

2.1 映射策略

语义映射是两个元数据标准的元素对应关系。在遵循语义相近原则、精确匹配原则和最广泛兼容原则下^[24]，元素映射应尽可能完整准确地保留源元数据的语义信息，最小化信息的损耗。

元素映射需要对原始的MARC数据进行分析，统计字段使用情况，明晰字段的映射范围。MARC字段可分为必备字段、特定资源必备字段、有则必备字段和选择使用字段。首先，编目机构在制定本馆著录细则时，会根据需要规定不同的字段必备性。其次，某些沿用自磁带技术的字段及其取值代码，在映射到本体词表后不再有意义，需要排除在映射表外或规定其取值规则。最后，针对语义冗余的MARC字段，需要根据实际进行映射的取舍或优先级判断。

BIBFRAME将书目数据分成作品-实例-单件(Work-Instance-Item)的核心实体层，它们各有自己的属性和关系。CNMARC记录是按照标志块、编码信息块等10个功能块构成的平面层次结构。每个功能块混杂了作品、实例或单件的属性和关系。在明确字段映射范围后，映射表在编制时就要确定MARC字段对应BIBFRAME的哪个核心层。

一个完整的或规范性的映射不仅有语义映射，也需要有元数据转换说明^[25]。元数据转换说明是对内容适当修正做出的规定，即对遵从原始元数据格式的元素值进行修正，以满足目标元数据的内容结构要求，实践中表现的是“映射取值规则”等。MARC格式通过子字段标识来存储不同含义的内容值，且不同子字段之间有着前后顺序，即一个字段中的多个子字段顺序组配在一起表达一个完整的含义。同时，在没有子字段标识区分不同内容值含义时，就需要有另一种标识符号系统满足对不同内容值的含义及起止进行说明的需求。ISBD就提供了标准的标识符解决方案，“由书目机构提供，置于每一书目著录项目或著录单元(第一著录项目的第一单元除外)信息之前或书目著录项目之中的标识符号”。因此，映射取值规则中明确规定了

取值内容要保留原始的子字段顺序,并使用ISBD标识符进行内容取值的字符串拼接。除了对原始格式内容格式的规定,目标格式内容也要在取值规则中说明。如BIBFRAME定义的属性bf:data取值是字符串,但在最佳实践中时间的格式建议采用ISO国际标准格式,以便于数据的共享和处理。

2.2 映射关系

元数据标准是为不同领域或不同目的而设计的,因此不同元数据元素之间存在不同的语义对应关系。除了在非MARC环境中没有意义的字段外,CNMARC字段与BIBFRAME词表之间的映射关系还可以分为以下4种。

(1) 一对一关系。一对一关系是两个元数据之间映射最常见的类型,表示一个源术语的语义信息与唯一一个目标术语的语义有极高的相似性或完全等同。如010\$aISBN字段映射到BIBFRAME就是bf:Instance--bf:identifiedBy--bf:Isbn这个三元组。

(2) 一对多关系。一对多关系表示源术语的语义外延较宽泛,有多个目标术语的语义在其语义范围内,这时可以建立起一对多的语义映射关系。如200题名字段,同一个字段需要映射到Work和Instance两层不同的三元组。

另外还存在一种特殊的一对多关系。一是字段在指示符不同的时候,含义不同,同一个字段就会映射到不同的三元组。典型的如团体责任者字段,当第一指示符是0时,表示团体,映射后三元组的宾语就是Organization;当是1时,则为Meeting。二是定长字段同一栏位根据取值需要映射到不同或多个三元组。如头标区第6位,当取值为a时,表示文献记录类型是印刷文字,它表达了两层含义:第一层是作品的内容表现形式为文字,映射到BIBFRAME三元组就是bf:Work--rdf:type--bf:Text;第二层表示文献的物质载体是印刷型,映射的三元组则为bf:Instance--rdf:type--bf:Print。

(3) 多对一关系。多对一关系与一对多关系正好相反,表示源术语的语义外延较宽窄,这些术语的语义可以被语义外延较宽的一个目标术语所包含,这时就可以建立多对一的映射关系。在这类映射中,需要进行映射优先级判断或对关系类型进行说明。第一类如出版发行时间字段的210\$d\$h与100\$a第9~16位,两个字段都表示文献的出版发行时间,前者是转录字段,后者是规范的

公元纪年。出于时间规范化处理,出版发行时间的子字段映射到bf:ProvisionActivity--bf:data--Literal三元组时,100\$a第9~16位的优先级高于210\$d\$h。而在映射到bf:Instance--bf:provisionActivityStatement--Literal三元组时,210\$d\$h的优先级有高于100\$a第9~16位。第二类的典型如3XX附注类字段。不同的3XX字段表示不同含义的附注信息,但在BIBFRAME词表中只有一个bf:note属性构成的三元组用于附注说明。但在映射中为了区分不同的附注类型,可以用bf:Note--bf:noteType--Literal三元组进行说明。

(4) 无映射关系。无映射关系表示源元数据术语在目标元数据中无法找到对应的元素,两者术语的语义不存在任何重合的情况。此类字段的映射,需要复用其他词表的术语或自定义术语。而在关联数据实践中,复用已有的成熟词表,以增强数据之间的共享便利性是推荐的最佳实践。因此,CNMARC字段无法映射到BIBFRAME词表时,可以在关联开放词表(linked open vocabularies, LOV)中查询适用的词表,如主题类字段的映射^[14]。

2.3 映射实现

CNMARC与BIBFRAME之间的映射,是以字段子字段为节点的MARC元数据模式与本体词表的三元组框架之间的映射。当两者的主谓宾三元组陈述相同时,即可进行映射。

具体的字段映射如210字段。210字段著录了与文献有关的出版发行信息。此处的文献是载体层面的概念,对应到BIBFRAME模型就是实例层(Instance)。BIBFRAME词表用对象属性bf:provisionActivity表示与资源出版发行制作等有关的时间、地点、代理。它的定义域是bf:Instance,值域是bf:ProvisionActivity。属性bf:provisionActivity的三元组映射到CNMARC就是210字段。210字段再细分到子字段,\$a表示出版发行地,对应类bf:Place,它的可用属性是未规定定义域的bf:place,可以是任何的实体类。因此类bf:Place通过对象属性bf:place与类bf:ProvisionActivity构成一个三元组陈述。\$b表示出版发行者,对应类bf:Agent,使用未规定定义域的对象属性bf:agent与类bf:ProvisionActivity构成一个三元组陈述。\$c表示出版发行时间,直接用未规定定义域的数据类型属性bf:data构成一个三元组陈述。完整的210字段映射如图1所示。

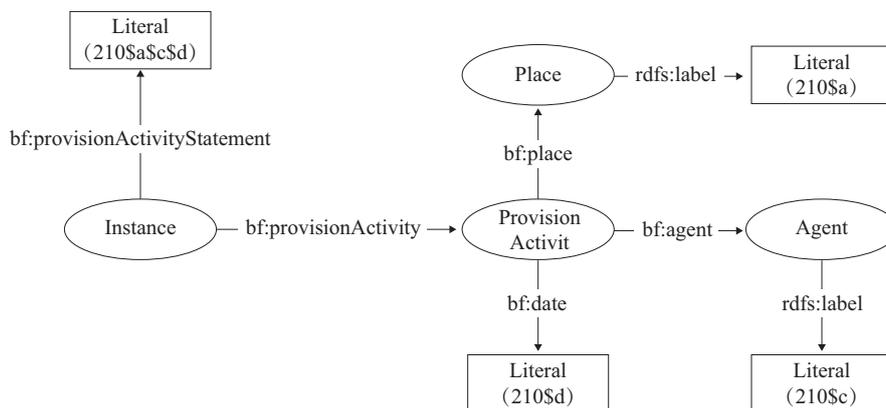


图1 出版发行项映射

又如定长类字段，其1位或多位取值构成一组语义信息，此类字段需要按位进行映射。定长字段105记录了与图书有关的图片类型、内容特征、索引等代码。0~3位图片代码和10位索引代码都是文献的附加内容，随不同的版本而不同，不属于作品的固有特征，因此映射到

BIBFRAME的实例层(Instance)。4~9, 11~12位是内容体裁代码，其属性值不随着版本的不同而发生变化，是作品的固有特征，因此映射到BIBFRAME的作品层(Work)。完整的105字段映射见表1。

表1 字段105与BIBFRAME映射对照表

CNMARC	BIBFRAME词汇三元组				
	subject	predicate	object (subject)	predicate	object
105\$a0~3	Instance	illustrativeContent	Illustration		
105\$a4~9, 11~12	Work	genreForm	GenreForm		
105\$a10	Instance	supplementaryContent	SupplementaryContent	rdfs:label	有索引

以关联数据实体识别为视角，BIBFRAME将书目数据分为作品、实例和单件3个核心层。实例和单件的概念与MARC标准相近，映射较为简单；作品是书目数据中最重要的实体，作品层映射关系的确定是整个映射表编制的关键与难点。

作品的一个重要识别元素就是正题名。CNMARC数据中题名著录涉及200、225、5XX、461和462等多个字段，每个字段又有\$a\$h\$i子字段著录题名的各数据单元。在上海图书馆关联书目数据项目(以下简称“上图项目”)中，将作品分为普通图书、汇编文献、年鉴和集刊、丛书4种。

对于普通图书，作品的首选题名著录于500字段。因此，500字段的映射优先级高于200字段。200字段的正题名又可以分为单一正题名和交替题名。单一正题名时，作品题名将直接映射到200字段；对于交替题名，有检索意义的题名分别著录在517字段中。因此，作品题名优先从第一个517字段获取。其他的517字段作为作品

的变异题名。

年鉴和集刊类文献是指定期出版的期刊类图书。这两类图书每期的责任者或题名的分卷信息有可能不同。但本质上它们都是一个系列的文献，是一个更大作品的组成部分。对于这类文献，使用“超级作品”的概念。超级作品是由人工整理的年鉴集刊的作品信息，是一个没有载体实例的虚拟实体，起到聚类的作用。具体到年鉴，在上海联编的CNMARC数据中年鉴的首选题名著录于540字段，其取值优先级高于200字段；集刊类书目数据的首选题名信息首先从500字段取值，如果没有再从200提取。这两类作品与超级作品之间的关系用“bf:partOf”表示。

丛书是一组相互关联而又各自独立的文献，每种文献除具有各自的题名外，还有一个整组文献的总题名。丛书本质上也是一个作品，它与系列内各作品之间是“bf:hasSeries”的关系。在上图项目中，丛书只有在225\$f时才作为作品发布，如果没有则作为实例的属性。

而对于无总题名的汇编文献,机器无法判断题名字段与责任者字段之间的配对关系,也就无法识别作品,这类数据在现阶段暂不考虑。

另外,上图项目在利用“创作者+首选题名”作品集信息键对作品进行去重聚类时,如果多条MARC记录属于同一件作品,就通过自定义属性“shlib:source”保留CNMARC记录的唯一标识号以说明属性值的来源。

3 结语

发布关联书目数据,映射表的设计是一个基础性工作。CNMARC是书目领域的元数据标准,BIBFRAME则是以RDF为基础开发的本体词表。一方面,元数据标准本身可以看成是知识本体的一种形式;另一方面,本体在某种程度上也可以看成是“元”元数据^[19]。两者都是对资源做出陈述性描述的标准。在此理解基础上,利用元数据映射的基本原则和步骤,根据数据分析和两者的术语语义关系,建立从源标准到目标标准的术语映射表。

现阶段映射表已完成CNMARC普通图书类字段的映射。但依旧存在诸多问题。①BIBFRAME词表的很多属性词只用“Used with”说明其建议的定义域,而没有严格定义。其实很多属性可以同时用于作品、实例或单件层。这为映射表的编制带来了一定的困扰。②关联书目数据的发布先要解决作品的识别问题。作品作为抽象概念,导致其属性与关系认定的模糊性。特别是在利用“创作者+首选题名”作品键进行作品对象去重时,即使不同MARC数据的作品键相同,但作品层的其他属性与关系也有可能是不一样的,如摘要、主题、读者对象等。因此,在利用BIBFRAME词表描述的数据实现FRBR作品层聚合时,需要明确BIBFRAME作品与FRBR作品是不同的,也不是简单地对应了FRBR的作品和内容表达。一方面,BIBFRAME词表中定义的属性词“Used with”为“bf:Work”和“bf:Instance”时,映射表需根据项目需求决定映射到哪一层对象;另一方面,也可以在BIBFRAME作品-实例-单件的三层模型结构之上,引入更抽象的超级作品层,即超级作品-BIBFRAME作品-实例-单件的四层结构,在超级作品层上实现聚类。③对于无总题名的汇编类数据,可以考虑在CNMARC格式中启用新字段著录作品的责任者与题名信息;或者考虑众包方式,将这类作品的识别由用户人工完成。

接下来,在第一阶段关联数据发布后,结合转换效

果的分析,继续对现有的映射表进行完善,特别是作品层映射关系的明确,同时也将扩展到非书类字段的映射,最终实现所有CNMARC字段与BIBFRAME词表的映射。

参考文献

- [1] 周艳玫,刘东苏,王衍喜,等.大学生信息行为调查分析与信息服务对策[J].图书情报工作,2015(6):61-67.
- [2] BORREGO A, ANGLADA L. Faculty information behaviour in the electronic environment [J]. New Library World, 2016, 117 (3/4): 173-185.
- [3] 廖建军. 美国政府“开放获取”政策及其对美国的影响[J]. 图书馆, 2018(4): 58-62, 105.
- [4] TIMOTHY W C, MYUNG-JA H, WILLIAM F W. Library Marc records into linked open data: challenges and opportunities [J]. Journal of Library Metadata, 2013, 13 (2/3): 163-196.
- [5] 2018 International Linked Data Survey for Implementers [EB/OL]. [2018-10-02]. <https://www.oclc.org/research/themes/data-science/linkedata/linked-data-survey.html>.
- [6] 潘煦,阳广元.近年来国内基于关联数据的数字图书馆研究综述[J].图书馆理论与实践,2016(7):34-38.
- [7] 张磊,夏翠娟.面向数字人文的图书馆开放数据服务研究——以上海图书馆开放数据应用开发竞赛为例[J].图书馆杂志,2018(3):33-38,48.
- [8] 夏翠娟,张磊,贺晨芝.面向知识服务的图书馆数字人文项目建设:方法、流程与技术[J].图书馆论坛,2018(1):1-9.
- [9] 夏翠娟,许磊.中文关联书目数据发布方案研究[J].数字图书馆论坛,2018(1):8-16.
- [10] 白林林,贾君枝.关联数据中CNMARC到RDF的映射实现[J].国家图书馆学报,2015,24(4):94-102.
- [11] 李勇文.书目框架(BIBFRAME)在中文书目数据中的应用范式探讨[J].图书情报工作,2016,60(2):101-105,145.
- [12] 吴贝贝. BIBFRAME在中文编目环境中的应用研究[J].农业图书情报学报,2017,29(9):91-95.
- [13] 周小萍. CNMARC与BIBFRAME的映射及转换研究[J].图书馆杂志,2018,37(8):21-29.
- [14] 许磊. CNMARC与BIBFRAME映射及其实现——以上海联编中文普通图书数据为例[C]//回顾与展望:新媒体时代下信息组织方法的创新与发展——第五届全国文献编目工作研讨会论文集.2017.
- [15] CHAN L M, ZENG M L. Metadata interoperability and

- standardization—a study of methodology part I [J]. D-Lib Magazine, 2006, 12 (6) : 3.
- [16] 杨蕾, 李金芮. 国外公共数字文化资源整合元数据互操作方式研究 [J]. 图书与情报, 2015 (1) : 15-21.
- [17] 赵亮, 楼向英, 张春景, 等. 元数据应用: 语义、结构与句法 [J]. 图书馆杂志, 2004 (7) : 49-55.
- [18] 刘嘉. 元数据导论 [M]. 北京: 华艺出版社, 2002: 85-87.
- [19] 刘炜, 李大玲, 夏翠娟. 元数据与知识本体 [J]. 图书馆杂志, 2004 (6) : 50-54, 49.
- [20] 杰弗里·波梅兰茨. 元数据 [M]. 李梁, 译. 北京: 中信出版社, 2017: 27-30.
- [21] Library of Congress. BIBFRAME Frequently Asked Questions 1: What is the Bibliographic Framework Initiative? [EB/OL]. [2018-10-03]. <http://www.loc.gov/bibframe/faqs/#q01>.
- [22] 王景侠. 书目框架 (BIBFRAME) 模型演进分析及启示 [J]. 数字图书馆论坛, 2016 (10) : 67-72.
- [23] 胡小菁. BIBFRAME核心类演变分析 [J]. 中国图书馆学报, 2016 (3) : 20-26.
- [24] 肖珑, 赵亮. 中文元数据概论与实例 [M]. 北京: 北京图书馆出版社, 2007: 311-312.
- [25] 徐维. 透视元数据映射概念 [J]. 情报理论与实践, 2004 (6) : 649-650, 631.

作者简介

许磊, 男, 1989年生, 助理馆员, 研究方向: 中文编目、关联书目数据, E-mail: leixu@libnet.sh.cn.

Foundations of Mapping from CNMARC to BIBFRAME

XU Lei
(Shanghai Library, Shanghai 200031, China)

Abstract: Linked data has become an important technical means for the opening of library resources. As the core data resource of the library, bibliographic data is an inevitable trend in the opening and sharing of the network. A key step in the publication of the linked bibliographic data is the mapping between the MARC fields and the ontology vocabularies. Based on metadata mapping, relationship between metadata elements and knowledge ontology, this paper summarizes the semantic mapping relationship between CNMARC fields and BIBFRAME ontology vocabularies, and describes it with typical field mapping.

Keywords: CNMARC; BIBFRAME; Metadata Mapping

(收稿日期: 2018-10-07)