

国外关联书目数据模型比较研究*

郑华敏

(东北师范大学信息科学与技术学院, 长春 130024)

摘要: 文章针对目前中文书目数据关联现状, 对国外典型关联书目数据模型进行调研分析, 主要从核心框架、功能实现、映射情况三方面对四个关联书目数据模型进行比较研究, 以期为我国图书馆书目数据关联化提供可行性建议。

关键词: 书目数据; 关联模型; 本体; 映射

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2018.07.008

图书馆是高校的文献资料情报中心, 是为教学和科学研究提供服务的研究机构, 保存大量珍贵书目资源^[1]。2006年, 在Tim Berners Lee^[2]首次提出关联数据的概念后, 国外许多图书馆都积极探索关联数据, 在构建书目数据关联框架的基础上, 将MARC数据映射到RDF进行发布, 这种方法不仅可以将书目数据发布为关联数据集, 还为图书馆书目资源整体知识发现和规范内容描述奠定基础。本文主要对美国国会图书馆书目模型(BIBFRAME)、英国国家书目数据模型(BNB)、欧洲数据模型(EDM)和瑞典联合目录(LIBRIS)进行整体描述, 从核心框架、功能实现、映射情况三方面对其进行比较研究, 以期为我国图书馆书目数据关联化提供可行性建议。

1 典型关联书目数据模型概述

本文主要从内容描述、词表、数据发布三方面对四个关联书目数据模型及其主要特征进行阐述。

1.1 BIBFRAME 2.0关联书目数据模型

BIBFRAME是美国国会图书馆(Library of Congress, LC)为取代MARC而开展的书目数据项目,

新框架旨在开拓关联数据的可能性, 提高各类资源的可发现性, 从而使图书馆成为真正意义上的数据关联共享中心。

(1) 内容描述方面。BIBFRAME 2.0间接借鉴FRBR实体-关系模型, 重点关注资源之间的关系。BIBFRAME 2.0中的创作作品对应FRBR中的实体作品和内容表达, 实例对应载体表现。在FRBR中, 单件是书目实体WEMI中的最低层, 指“一种载体表现的单一样本”, 图书馆书目记录下的馆藏记录也被称为单件记录, 两者比较接近。BIBFRAME 2.0取消了规范核心类, 因为在关联数据中身份控制不再由名称形式上的规范, 而是通过URI来体现其可操作性, 规范检索点与其他名称一样, 只是一个显示标签^[3]。

(2) 词表方面。BIBFRAME在公布时为了保证模型空间稳定性没有复用任何词表, 也没有表明与现有词表的关系, 但这种保守且不适应关联数据实践的做法已经开始改变。RDF、RDFS、WEB注释模型, 以及其他的成熟关联数据模型已逐渐在BIBFRAME 2.0中使用。

(3) 数据发布方面。BIBFRAME 2.0使用SPARQL语言作为RDF数据查询语言。RDF三元组是一种抽象的数据模型, 通常使用RDF/XML对其进行序列化处理, 成为被机器可读的数据。除此之外, BIBFRAME 2.0还大量使用基于事件驱动的网络触发器等。

*本研究得到国家社会科学基金项目“图书馆书目数据社会化应用研究”(编号: 15BTQ018)资助。

1.2 BNB关联书目数据模型

2011年,大英图书馆将英国国家书目(British National Bibliography, BNB)数据发布为关联数据并将书目资源分为图书和连续出版物^[4],采用不同的数据模型对其进行内容描述,在功能上也有所区别。本文主要以图书关联数据模型为例进行阐述。

(1) 内容描述方面。BNB的数据模型被划分为出版事件数据模型、主题数据模型、责任者数据模型和其他数据模型四大部分。同时BNB将书目资源实体分为书目资源、连续出版物、主题和概念四大类,各个实体都有其相对应的URI^[5]。

(2) 词表方面。BNB没有重新设计词表和本体模型,而是尽可能采用已有的本体词汇,根据实体类型构建相应数据模型。BNB采用14个词表来描述连续出版物和图书的命名空间,每个词表都有其独特的作用。词表分为通用词表和专用词表,通用词表主要描述实体的一般属性(包括DC、RDF、RDFS、OWL等);专用词表有其专用的描述实体,如British Library Terms (BLT)是大英图书馆创建的用来描述书目资源的属性和类,以及其间关系的词表。

(3) 数据发布方面。BNB将MARC21记录进行一系列字符集转换、数据标准化和匹配处理,然后将数据置于更广泛的环境中,从关联数据集中选择高质量资源进行关联以生成关联数据。文件被转换为RDF/XML和N-Triples格式并上传到网站,用户可通过SPARQL端点访问数据。

1.3 EDM欧洲关联书目数据模型

欧洲数字图书馆(Europeana)将图书馆与博物馆、档案馆、画廊等连接起来,基于EDM组织数据,将1 500个机构的超过2亿条记录和1 000万个数字对象以关联数据的形式提供数据集,为欧洲的文化遗产创造新的共同接入点。

(1) 内容描述方面。Europeana的本体模型有三大核心类,即edm:ProvidedCHO、edm:WebResource和ore:Aggregation。其中,edm:ProvidedCHO代表被描述的对象本身,包括绘画、电影、乐谱、图书等;edm:WebResource表示被描述对象的数字表现形式;ore:Aggregation遵循ORE规则,将数字文化资源对象和其

数字表达聚合在一起构成数据集^[6]。

(2) 词表方面。EDM是Europeana语义元素(ESE)的一项重大改进,利用本体描述语言OWL对欧洲丰富文化遗产进行收集、管理、发布和共享。EDM使用OAI-ORE词汇表和数据结构表示提供者、对象和元数据之间的复杂关系,从而协调不同版本的资源并追踪与这些资源相关的描述性陈述^[6]。

(3) 数据发布方面。Europeana可通过文件下载和API将数据输出为JSON文件。此外,Europeana还可通过SPARQL端点访问数据,此端点可以返回JSON、RDF/XML、N3/Turtle和N-Triples格式的序列化数据。

1.4 LIBRIS瑞典国家图书馆关联书目数据模型

2008年,瑞典国家图书馆将LIBRIS发布为关联数据,是世界上首次被整体发布为关联数据的联合目录,这标志着书目数据开始正式融入互联网环境。

(1) 内容描述方面。LIBRIS的实体类型根据FRBR模型进行实体抽取,包括作品、人、概念和机构四大类。LIBRIS的成员使用同一个图书馆集成管理系统(ILS)进行编目,通过RDF服务器封装件实现HTTP访问ILS,然后根据请求递送描述书目记录、规范记录及书目资源链接的RDF,并创建唯一标识符URI。当有正确的请求时,URI被反向解析并通过一定的方式,利用HTTP的内容协商机制传输RDF。

(2) 词表方面。LIBRIS在发布关联数据的过程中,使用的词汇并没有局限于图书馆领域,还利用了一些如DC元数据、FOAF、SKOS等被广为使用和理解的标准。除此之外,LIBRIS数据库还添加了DBpedia和Wikipedia等外部资源,为用户提供更多相关信息。

(3) 数据发布方面。LIBRIS^[7]的SPARQL server8当前使用Sesame 2 Native Store作为三重存储,具有良好的查询性。同时基于内容协商协议,可以通过相同URL传送多种格式的数据。

2 关联书目数据模型核心框架比较分析

本文对上述四种主要的关联书目数据模型的核心框架进行比较分析,如表1所示。

表1 关联书目数据模型核心框架比较

图书馆名称	实体类型	URI格式	发布格式	词表来源
LC	Work、Instance、Item	<http://bibframe.example.org/item/item5>	HTML RDF/XML	BIBFRAME、PREMIS EBUCore/PBCore
大英图书馆	Resource、Series、Agent、Concept	http://bnb.data.bl.uk/id/resource/<bnb-number>	HTML RDF/XML Turtle JSON	DC、BIBO、Bio、FOAF、SKOS、RDA、British Library Terms、EventOntology、ISBD、Org、OWL、RDF SchemaWGS84GeoPositionig
Europeana	edm:ProvidedCHO edm:WebResource ore:Aggregation	http://data.europeana.eu/item/<number>/BibliographicResource<number>	HTML N-Triple RDFJSON	DC、FOAF、SKOS、OWL、RDF、OAI-ORE、EDM、CIDOC-CRM、WGS84 Geo Positioning、CC、RDAU
瑞典国家图书馆	Work、Person Concept、Organization	http://libris.kb.se/resource/bib/<number>	HTML RDF/XMLN3	DC、BIBO、FOAF、SKOS、LIBRIS

2.1 实体类型

实体的本质是对词表中“相对关系”界定模糊的定义进行详细明确的形式化说明,丰富传统词表内容。本文介绍的书目数据模型中,BIBFRAME和LIBRIS均以FRBR模型为基础模型,修订后的BIBFRAME2.0中bf: Creative Work和bf: Instance保留原本的语义,新增类bf: Item在语义上接近FRBR Item实体,以此替代现有bf: Annotation的子类即bf: HeldItem和bf: HeldMaterial。BNB针对图书和连续出版物抽取不同的实体类型,并根据实体类型构建模型。在构建过程中,BNB避免复制MARC的复杂结构,为使数据集在图书馆领域以外同样发挥作用,图书和连续出版物的实体设计具有普适性,同时将出版物事件模型进行扩展以覆盖数据的生命周期。EDM则使用本地定义的属性来跟踪贡献的文化遗产对象,为连接核心类,EDM还依赖ore: Aggregation定义资源之间关系的聚合类; EDM侧重跟踪文化遗产客体本身和其周围的描述性元数据,区分对象和描述该对象的元数据,遵循一对一原则。

本文介绍的四种关联书目数据模型可分为两大类:第一类是根据已有模型进行实体抽取,不仅因为FRBR模型经过实践证明具备可实施性,还因为FRBR模型与BIBFRAME、LIBRIS的书目数据特点相契合,符合其发展需求;第二类是机构自行抽取,BNB和EDM都是根据自身书目数据特点进行抽取,并构建相应的关联数据模型。

2.2 URI格式

关联书目数据的重要步骤是使用唯一标识符(URI)来标识资源。URI的通用结构是<基地址>/<实体类型名称>/<标识符>。采用URI标识数字资源,实际上是对实体进行URI命名,以便有效地识别实体,建立实体间关联,确保实体的唯一性和可获取性。BIBFRAME2.0将实体分为work、instance和item。以work为例,其主要是对作品的概念进行阐述,不涉及具体内容,包括worktitle、language、creator等。

```
<bf:Work rdf:about=" http://example/001" >
  <bf:workTitle rdf:resource=" http://example/001/
title01" />
  <bf:creator rdf:resource=" http://example/001/
person01" />
```

BNB将书目资源分为图书和连续出版物,实体分为resource、agent、concept和series。其中,agent包括人名、机构名、家族名和地名;concept包括主题词、分类号及各类型名称实体。人名、机构名、家族名和地名既可作为agent对待(即资源关联的责任者名),又可作为concept对待(即资源涉及的人名、机构名、家族名和地名等),因此对应两个URI。

以人名为例,在agent中的URI为: http://bnb.data.bl.uk/id/person/{person-name};在concept中的URI为: http://bnb.data.bl.uk/id/concept/person/lcsh/{person-name}。

2.3 发布格式

关联数据的发布格式取决于不同的关联序列化方法,当前的关联数据发布格式可以分为四类:HTML类型(HTML、RDFa、Microdata)是为用户理解和和使用书目数据而设计的,可以提供数据服务;XML类型(RDF/XML、RDF/XML-ABBREV)是W3C的标准推荐格式,但其存在复杂度高、可读性差的缺点;N3类型(N3、Turtle、N-Triple、N-Quads、TriG、TriX)简化了XML的复杂度,提升了可读性和互动性;JSON类型(RDF/JSON、JSON-LD)是目前普遍使用的数据交换格式,适用于互联网,但难以书写和阅读^[8]。书目数据的发布格式一般包括多种类型,这样才能同时满足机器和用户的需求。

四个图书馆均采用HTML类型和XML类型的发布格式,RDF/XML是W3C的推荐标准,也是语义网技术的基础,可以被大部分机器语言识别和处理,所以也是很多图书馆数据发布格式的选择。此外,LIBIRIS选择可读性和互操作性较好的N3类型,BNB和EDM选择JSON类型。JSON类型目前被很多数字网络服务采用,是可同时兼顾可读性和标准性的数据发布格式。

2.4 词表

书目数据的词表包括复用已发布的通用词表和

独创词表(见表2)。BIBFRAME在创建之初便没有复用任何通用词表,随着关联数据的不断实践,BIBFRAME2.0直接采用RDF和RDFS两个关联数据基础命名空间。如rdf:Resource代替bf:Resource,rdfs:label代替bf:label等。BIBFRAME2.0将有可能继续引入更多已有词表^[9]。而大英图书馆、Europeana及瑞典国家图书馆不仅复用了通用词表,还根据自身书目数据特点形成独创词表。独创词表包括大英图书馆的BLT、Europeana的EDM,以及瑞典图书馆的LIBRIS等。

除LC外,大部分图书馆都选择复用通用词表增强数据的互操作性,而LC也在改变原有做法,逐渐创建新的词表。大英图书馆是复用通用词表最多的图书馆,这也与其使用不同领域词表、实现与多个外部数据集相关联的指导思想相契合。在图书馆独创词表中,各图书馆均有与其数据模型相对应的词表。此外,BNB通过研究并结合自身情况独创了BLT,用以描述书目资源的属性和类及资源间的关系。该词表是BNB根据自身实际情况增加其他本体中没有的属性和类。Europeana的OAI-ORE是EDM的核心数据结构,基于开放归档对象重用和交换模型,该模型是Web资源聚合描述和交换的参考模型^[10]。ORE聚合用于表示数据提供者对欧洲的贡献。

综上所述,BIBFRAME期望将数据标准设计成一种开放的交换格式,尽量不规定特定的编目原则,同时侧重规范定义、数据建模和元数据质量。BNB将重

表2 关联书目数据模型词表分类表

词表	词表名称	词表用途
通用词表	DC	都柏林核心元数据
	RDF	一种用于描述网络资源的通用语言
	OWL	一种语义标记语言,用于在万维网发布和共享本体
	BIBO	用于描述书目信息的词表
	FOAF	描述人与人之间关系的词表
	SKOS	描述受控词表概念信息的词表
	WGS84 Geo Positioning	描述地理领域概念与知识的词表
独创词表	BIBFRAME	LC创建的用于描述书目信息的词表
	British Library Terms	大英图书馆创建的用于描述书目资源的属性和类及其关系的词表
	OAI-ORE Web	Europeana创建的资源聚合描述和交换的参考模型。ORE聚合用于表示数据提供者对欧洲的贡献,EDM将ESE与Web架构和关联数据原理相结合,识别和展示Web资源
	LIBRIS	致力于建设瑞典175个成员馆的学术和研究图书馆的联合目录

点放在数据的主要实体上,而并不试图复制MARC21的复杂结构和内容,同时为提高互操作性,尽量使用现有的RDF词汇表和本体来描述实体和关系。EDM同样基于RDF进行数据模型构建,鉴于其成熟性,EDM的记录范围比BIBFRAME更大,且具有强大的数据收集和转换服务。LIBRIS在UNIMARC中保留源格式,然后在镜像数据库中转换为XML以提供Web

服务和API,同时在RDF中使用并行自制生产数据转换工具。

3 关联书目数据模型功能比较分析

本文对上述4种关联书目数据模型的关联情况、功能实现及应用范围进行比较,如表3所示。

表3 关联书目数据模型功能比较

图书馆名称	关联情况		功能实现	应用范围
LC	图书馆资源	图书馆中BIBFRAME格式的全部在线目录	适应图书馆、文化机构和用户创建的各种形式的书目资源元数据,最终在互联网中创建并共享这些元数据	包括LC在内的15个机构的书目数据
	外部资源	BNF、ISNI、LC NAF、VIAF、Wikidata		
大英图书馆	图书馆资源	LCSH、VIAF、MARC Country、MARC Language Codes、Dewey.info	使用不同领域词表,实现与多个外部数据集关联,丰富书目资源语义	整个英国和爱尔兰的书目资源,包括连续出版物和图书
	外部资源	DBpedia、Geonames、RDF Book Mashup、Lexvo、ISNI		
Europeana	图书馆资源	VIAF、GEMET、AAT、DNB	不拘泥于特定规则,将图书馆、博物馆、档案馆资源进行关联并共享	Europeana中愿意公开的资源,包括文本、图像、视频、音频等
	外部资源	DBpedia、Geonames、Iconclass		
瑞典国家图书馆	图书馆资源	LCSH、VIAF	关联数据的发布是图书馆界第一次在关联数据领域的尝试	LIBRIS 175个成员馆的书目数据
	外部资源	Wikipedia、DBpedia		

3.1 关联情况

图书馆采用关联书目数据模型的本质是为了使图书馆书目信息资源不仅可以内部互相关联,还可以与外界资源相关联,从而使图书馆资源在互联网上得到最优化利用。所以,各关联书目数据模型的数据关联化步骤基本相同,都是将各类资源在技术支持下纳入关联化序列,建立各个实体属性间的关联。

在书目数据的关联过程中,首先是各图书馆间的资源互联(大英图书馆和瑞典国家图书馆、法国图书馆、德国图书馆等书目数据集都互相关联);其次在关联外部资源时,通常选择公共领域中一些重要的开放数据集(如Wikipedia、DBpedia等);最后,各图书馆均借鉴由权威机构发布的一些通用词表和规范文档,包括由OCLC发表并集合各图书馆有关人名和机

构的规范文档VIAF、LC发布的主题词表LCSH及用于描述地理信息概念的词表Geonames。

3.2 功能实现

BIBFRAME不局限于某一机构特定的内容规则,构建了开放包容的本体模型,旨在将书目数据链入关联数据模型,同时还纳入新兴的数据标准和模型(包括FRBR和RDA)。BIBFRAME强调将书目资源解构为明确的内容描述,同时使用URI取代文本或文字值^[11]。但由于BIBFRAME仍在不停地修订,关联数据模型未完全确定,尚处于测试阶段。BNB根据不同类型的书目资源构建不同的本体模型,以不同的语义操作技术为支撑,将MARC格式的书目数据映射为RDF格式,并将大量图书馆内部资源和外部其他领域的词表进行关联扩

展, 最终实现书目数据实体间的关联, 以及与外部资源的共享, 满足图书馆与外部数据集相关联的需求。

Europeana开展的跨领域、跨机构的复杂关联数据项目, 已经成功搭建了数字文化资源聚合门户。作为互联网门户网站, Europeana可为全欧洲数百万书籍、绘画、电影、博物馆对象和数字化档案记录提供接口, 也是为文化代理人之间的合作提供知识交流平台。LIBRIS在发布关联数据的过程中使用的词表不局限于图书馆领域, 同时借鉴FOAF、BIBO等词表, 为图书馆书目数据资源与外界资源的融合打开新局面。

3.3 应用范围

目前已有7个机构和8个馆藏正在使用BIBFRAME进行书目数据的关联测试, LC根据测试结果对BIBFRAME进行不断补充和修正。大英图书馆收录了英国和爱尔兰自1950年以来出版发行的所有新书和连续出版物, 包括后来的CD-ROM、DVD-ROM、电子期刊、电子图书、数据集、网页等。Europeana的关联数据集来自欧盟成员国200多个不同文化机构的数据。Europeana将欧洲的画廊、博物馆、图书馆和档案馆资源进行数字化描述, 让研究人员获得更丰富的材料和数据。LIBRIS的175个成员馆均使用同一个ILS进行编目, 目前整个系统拥有600万条书目数据, 瑞典国家图书馆已经开发出一套组件使系统能够处理RDF格式数据。

综上所述, BIBFRAME利用关联数据使得图书馆元数据变得更易于访问和实用, 以满足用户前期未被满足的需求和新兴需求。BNB负责描述馆藏数据并记录英国的出版产品, 通过采取开放的元数据战略, 增加图书馆数据的社会价值, 保持图书馆服务的相关性。EDM试图超越欧洲文化资源部门的信息视角, 采用基于语义网的开放跨域框架, 不仅支持数据提供商提供的多种元数据, 还可以通过关联外部资源丰富数据库。LIBRIS希望为研究人员、教师和学生提供最大可能的数据发现和可利用性, 旨在创建简单、直观和易于使用的数据关联模型。

4 关联书目数据模型映射比较

关联书目数据模型通常采用RDF格式对书目数据进行描述, 因此在进行书目数据的关联过程中, 从MARC到RDF的映射就成为其中的关键技术。2016

年, BIBFRAME官网上发布了BIBFRAME2.0术语词表, 包括分类视图、清单视图和RDF视图三种视图。其中, RDF视图取代模型视图, 使BIBFRAME能够更具体、更迅速地适应关联数据大环境。如在一般属性中的bf:identifiedBy与MARC中的“0-标识块”对应, 在BIBFRAME词表中对bf:identifiedBy的描述强调了唯一性, 因此在映射时只与001字段、014字段、040字段进行了对应。

BNB的映射表根据模型实体也将书目数据划分为责任者信息、出版事件信息、主题信息及其他信息(包括书目资源类别信息、题名信息、丛书系列信息、语言信息、标识信息和附注类信息)对各个属性进行映射。如在责任者信息中常用的MARC字段有100字段、110字段、700字段等, 需要映射的主要款目标目是Dct:creator, 附加款目标目是Dct:contributor。

Europeana的映射规则也是根据模型实体展开的, 包括EDM核心类(edm: ProvidedCHO、edm: WebResource、ORE:aggregation)和contextual类(edm:agent、edm:place、edm:timespan、skos:concept、cc:license)^[12]。Europeana鼓励供应商从现有数据中提供尽可能多的属性以创建完整描述, 虽然没有必要使用所有可用的EDM属性, 但必须保证其规范性, 还要尽可能使用最精确的属性。如使用子属性dcterms: spatial或dcterms: temporal代替更宽泛的dc: coverage。

目前我国图书馆采用CNMARC格式描述书目数据, 但与MARC格式仍存在细微差异(见表4)。以CNMARC中的600字段为例, 600字段表示个人主题名称, 在BIBFRAME中根据子字段的的不同, 600字段和不同的BIBFRAME属性相对应, 如600的子字段为\$a(款目要素)时, 对应类别为“主题词和分类信息”中的属性Bf:subject。在BNB中对应主题信息中的Dct:subject(LCSH), 其中命名空间Dct是都柏林核心集的升级版, 取值来自LC标题表(LCSH)。在EDM中对应核心类的edm: ProvidedCHO中的Dc:subject, 其命名空间是DC元数据。

综上所述, BIBFRAME只采用了自身的命名空间, 因为BIBFRAME在发布词表时没有复用任何通用词表, 而是选择重新定义每个属性和类。BNB在进行书目资源描述时采用14个词表, 包括成熟的规范词表和大英图书馆的独创词表。词表作为命名空间被使用, 每个命名空间都有独特的作用。BNB灵活地将各领域的通

表4 常用CNMARC字段与各个关联书目数据模型映射比较

CNMARC	BIBFRAME	BNB	EDM
600 (个人主题名称)	Bf:subject	Dct:subject (LCSH)	Dc:subject
5__ (相关题名块)	Bf:title	Dct:title	Dc:title
010 (国际标准书号)	Bf:isbn	Bibo:isbn10、isbn13	Dc:isbn
205 (版本说明)	Bf:edition	Isbd:P1008 (edition statement)	Dcterms:has Version
210 (出版发行项)	Bf:date Bf:place Bf:agent	Event:time Event:place Event:agent	Dc:date Edm:place Edm:agent
300 (一般性附注)	Bf:note	Dct:description	dcterms:provenance
101 (作品语种)	Bf:language	Dct:language	Dc:language

用词表运用到自身书目资源的描述中,使其在各个实体中发挥最大作用。EDM采用的命名空间主要是DC元数据和自身命名空间(如edm),在contextual类中也复用了许多词表(如SKOS、FOAF、WGS84等)。

5 启示

目前,我国图书馆的关联数据工作正处于探索和研究阶段,尚未将图书馆的书目数据大量发布为关联数据并共享。本文从关联书目数据模型的核心框架、功能比较、映射比较三个维度对LC、大英图书馆、Europeana及瑞典国家图书馆进行详细分析。国外这些成熟的书目数据项目对我国将来的关联数据研究和实践提供了丰富经验。

(1) 复用通用词表。通过分析发现,国外在书目数据关联过程中采用的词表包括已有的由权威机构发布的通用词表和图书馆独创词表。在关联数据领域经常通过复用通用词表的类和属性来标识同一事物,如果完全采用自身命名空间会使词表过于庞大而难以维护。因此,我国在创建关联数据模型时可借鉴大英图书馆的做法,尽可能复用通用词表。这样不仅可以减轻图书馆的工作量和对本地元数据管理的负担,提升数据互操作性,还可以使关联开放的书目数据在数据模型上一致,便于关联数据的管理、共享和利用。

(2) 参考已有模型进行建模。数据模型的建立基本与词表的选用同时进行,拥有明确的实体模型非常重要。如BIBFRAME和LIBIRS均以FRBR为基础框架进行模型构建。图书馆在创建关联书目数据模型本体时应尽量建立在已有模型基础上。首先,已有模型均由

权威机构发布,经过实践的考验并具有一定的稳定性。其次,明确的实体模型有利于书目资源中复杂关系的描述和知识发现,并能提升书目数据关联的灵活性和扩展性。最后,在选择已有模型作为基础框架时,应结合中文书目数据特点,以完整准确地描述我国图书馆资源内容。

(3) 与外界资源建立广泛关联。国外关联书目数据在资源关联方面主要以图书馆内部资源为主,如主题词表、规范文档等,而对外界资源的关联具有较大局限性,主要关联了一些重要的开放数据集(如Wikipedia、DBpedia等),关联内容和领域较单一。在密集型科研中,图书馆应积极探索新的服务方式,为用户提供更精准的服务^[13]。因此,我国书目资源在关联过程中应尽量扩大关联范围,与各领域资源建立关联,同时将关联书目数据渗透到其他领域,丰富书目数据的关联性,使图书馆书目数据资源在互联网中发挥重要作用。

参考文献

- [1] 魏来,高希然.大数据背景下高校数据馆员的角色定位[J].情报资料工作,2015,34(5):90-94.
- [2] BERNERS-LEE T. On the Next Web [EB/OL]. [2018-05-25]. http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html.
- [3] TANIGUCHI S. Examining BIBFRAME 2.0 from the viewpoint of RDA metadata schema [J]. Cataloging & Classification Quarterly, 2017, 55 (6): 387-412.
- [4] DELIOT C. Publishing the British National Bibliography as Linked Open Data [EB/OL]. [2018-05-25]. <http://www.bl.uk/>

- bibliographic/pdfs/publishing_bnb_as_lod.pdf.
- [5] 张鹏图. 大英图书馆书目数据的关联化分析 [J]. 国家图书馆学报, 2015, 24 (4): 103-113.
- [6] EDM primer [EB/OL]. [2018-05-25]. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.
- [7] LIBRIS [EB/OL]. [2018-05-25]. <http://libris.kb.se/>.
- [8] 邹美辰, 胡瀛. 欧美国家图书馆书目数据关联化案例研究 [J]. 图书馆理论与实践, 2016 (11): 61-66, 70.
- [9] 胡小菁. BIBFRAME核心类演变分析 [J]. 中国图书馆学报, 2016, 42 (3): 20-26.
- [10] HASLHOFER B, ISAAC A. Data.europeana.eu: the europeana linked open data pilot [C] //International Conference on Dublin Core and Metadata Applications. Dublin Core Metadata Initiative, 2011: 94-104.
- [11] CASALINI M. BIBFRAME and Linked Data practices for the stewardship of research knowledge [C] //Digital Humanities. Connecting Libraries and Research. Berlin:IFLA-Satellite-Meeting, 2017.
- [12] Europeana Data Model – Mapping Guidelines v2.3 [EB/OL]. (2016-11-18) [2018-05-25]. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM%20Mapping%20Guidelines%20v2.3_112016.pdf.
- [13] 黄金霞, 马雨萌. 大数据时代开放信息资源的数据服务能力思考 [J]. 数字图书馆论坛, 2016 (8): 54-59.

作者简介

郑华敏, 女, 1994年生, 硕士研究生, 研究方向: 信息服务, E-mail: 1604609940@qq.com。

Comparative Research on Linked Bibliographic Data Models Abroad

ZHENG HuaMin

(Northeast Normal University School of Information Science and Technology, Changchun 130024, China)

Abstract: According to the status quo of data association in current Chinese books, the data models of typical bibliography abroad are investigated and analyzed. Mainly from the main framework, function implementation, mapping of the four bibliographic data models comparison study, the paper finally provides a feasible proposal for the bibliographic data association for our library.

Keywords: Bibliographic Data; Association Model; Ontology; Mapping

(收稿日期: 2018-06-05)

书讯

《汉语主题词表(自然科学卷)》

《汉语主题词表》共分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分。继2014年《汉语主题词表(工程技术卷)》出版后,由中国科学技术信息研究所主持,并联合全国图书情报界相关机构,又研制推出《汉语主题词表(自然科学卷)》。

全书共收录专业术语12.4万条,包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域,收词系统、完整,语义关系丰富、严谨,每条词汇都有相应的学科分类号表现其专业属性,并与同义英文术语对应。同时,建立《汉语主题词表》网络服务系统,提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表(自然科学卷)》按学科相关原则分为5个分册,可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘,是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表(自然科学卷)》已于2018年5月由科学技术文献出版社出版,全书396万字,总定价1247元,可分册购买。