

面向语义网的书目框架(BIBFRAME): 功能需求及实现^{*}

□夏翠娟

摘要 针对 BIBFRAME 如何应用语义网技术以适应新的编目环境问题,结合最新发布的 BIBFRAME 本体词表和规范控制框架,从功能需求设计的技术细节着手,详细深入解读 BIBFRAME 面向语义网的技术架构和本体模型,并通过典型用例的分析,探索 BIBFRAME 在技术思路对其功能需求的实现方法,及其从数据层面而非依赖具体的应用系统来改善图书馆用户服务和改进采编业务流程的技术思路。BIBFRAME 利用关联数据技术、知识本体建模方法、基于网络资源的规范控制框架,为图书馆数据成为语义网的一部分提供了技术上的可行性。

关键词 MARC 书目框架 语义网 关联数据

1 背景与现状

图书馆作为人类文明的保存和传承机构,拥有海量的高质量数据,在网络时代,这些数据的价值远远超过图书馆领域和传统图书馆的服务范围。互联网已成为人们工作和生活不可或缺的一部分,以“谷歌”为代表的搜索引擎技术成为信息海洋的导航塔,但它们以关键词匹配和网页评级(PageRank)为主要检索技术,在检索结果的相关性上还有欠缺。图书馆长期以来在知识组织领域积累了大量的结构化数据,其中包含众多领域专家的知识,这些数据如果能成为万维网的组成部分,将在很大程度上丰富万维网的内容^[1]。图书馆在规范控制方面也有长期积累的方法和经验,如果应用于网络资源,将在很大程度上改善信息检索结果的相关度^[2]。过去的半个世纪中,图书馆的数据主要是以 MARC 格式编码, MARC 格式诞生于卡片目录时代,主要用于在图书馆领域内交换数据,但 MARC 格式的封闭性阻碍了图书馆数据在更广阔的网络世界的交换和利用,急需一种新的格式来取代^[3]。

自蒂姆·伯纳斯·李(Tim Berners Lee)于本世纪初倡导“语义网(Semantic Web)”运动、并于

2006 年提出“关联数据(Linked Data)”的概念和技术框架^[4]以来,图书馆界多年来致力于利用关联数据技术改造图书馆数据,使之适应网络环境,成为网络的一部分,瑞典、美国、匈牙利、英国、德国、西班牙、日本等国的国家图书馆,以及 OCLC,已陆续将自己的书目数据或规范数据发布为关联数据^[5]。关联数据作为语义网的一种轻量级的实现方式,基于目前成熟的网络技术如 HTTP、URI 等,使数据可以方便地在网络上展示、共享,同时采用 W3C 的语义网标准——RDF 数据模型对数据建模,使数据与数据之间的关系可被机器识别和处理,图书馆的数据要成为万维网甚至是语义网的一部分,关联数据是一种简便可行的方式^[6]。

2008 年,美国国会图书馆的“书目控制未来工作组(Working Group on the Future of Bibliographic Control)”发布了最终报告,呼吁图书馆界设计开发新的书目格式,以适应新的编目环境^[7]。2011 年 5 月,美国国会图书馆宣布发起“书目框架先导计划”,目的是开发出一套适用于网络的书目框架格式,最终可以取代 MARC,同时又兼容后 MARC 时代的内容标准如 RDA 和元数据标准如 DC 等^[8]。

^{*} 本文为国家社科基金青年项目“W3C 的 RDB2RDF 标准规范在关联数据服务构建中的应用”(编号:13CTQ008)和国家社科基金一般项目“关联数据的理论与应用研究”(编号:11BTQ041)的成果之一。

除国会图书馆外,还有德国国家图书馆、大英图书馆等6个大型图书馆参与早期试验,2012年底,国会图书馆发布书目框架的关联数据模型草案,正式推出书目框架格式(简称BIBFRAME)^[9],同时发布功能需求与用例,展示了BIBFRAME为网络而生的初衷和改造图书馆数据使之适应新环境的巨大潜力。2014年2月,美国国会图书馆在BIBFRAME网站上发布了基于知识本体的BIBFRAME词表,更新了基于网络资源的规范控制框架。BIBFRAME的发布在国内图书馆界也引起了关注,但由于对它的研究需要元数据、知识本体、RDF、SPARQL、OWL等语义网相关技术的综合知识和技术能力,因而相关的研究较少,目前为止只有2013年11月在成都举办的“全国第五次图书馆2.0研讨会”上有一个分会场以BIBFRAME为主题,只有一篇学术论文见诸国内图书馆学刊物,本文是其后续研究,针对BIBFRAME是如何应用语义网技术以适应新的编目环境问题,结合最新发布的BIBFRAME本体词表和规范控制框架,从功能需求的技术细节着手,详细深入解读BIBFRAME面向语义网的技术架构,通过对典型用例的分析,探索BIBFRAME的实现方法,以及将对图书馆用户服务和业务流程所带来的影响。

2 面向语义网的技术架构

2.1 采用关联数据的原则来组织、展示和分享数据

关联数据的技术框架可以概括为4个原则,它基于HTTP协议、URI这些基本的网络技术,用HTTP URI来标识一切事物,通过访问URI,还能获取更多关于该事物的结构化信息。HTTP协议是网络的基本架构,用HTTP URI来标识的资源在全球范围内可唯一识别和定位。

在MARC记录中,100字段的\$a子字段用来表示著者,字段的值只是一个字串:Cohn John,对于计算机来说,它没有除字面之外更多的意义,但在BIBFRAME中,会用一个URI来表示:<http://viaf.org/viaf/94196132/>。用一定的请求访问这个URI,可以获得请求所指定的返回结果信息,比如关于这个作者的生卒年、作品年表等,这些用一定格式编码(例如RDF/XML)的结构化信息可被计算机读取和处理。

用RDF三元组模型对数据建模,使得数据与数

据之间的关系可以用机器能理解的方式表达出来。RDF三元组由主体-谓词-客体组成,一个三元组中通常包含两个节点(主体和客体),它们之间用一个谓词来联接,表示二者之间的关系,三元组中的节点和谓词都可以用一个URI来表示,可以位于网络上的任何地方,例如:`<http://www.example.com/work/1>` `<http://bibframe.org/vocab/creator>` `<http://viaf.org/viaf/94196132/>`,意为作品`<http://www.example.com/work/1>`的著者(creator)是`<http://viaf.org/viaf/94196132/>`,而著者(creator)这个词来自于`<http://bibframe.org/vocab/>`这个受控词表,有着特定的可被机器读取和处理的意义。RDF三元组中的主体和客体作为节点,谓词是有向弧,一个由节点和弧构成的三元组被称为一个基本的RDF图(RDF Graph),一个RDF图中的节点又可以与其他不同的RDF图中的节点产生关联关系,它们之间由谓词关联,这样多个相互关联的RDF图构成RDF数据集,从而形成RDF数据所特有的立体网状图式结构。

书目数据以关联数据为基本数据模型意味着,书目数据不再像存储于关系数据库中的MARC数据那样,以记录为单位,而是以更小粒度的数据为单位,每个数据单元都是独立的存在,同时又可与其他数据单元建立可被机器理解的关联关系,数据是相互关联且富含语义的;而以HTTP URI标识的数据单元可以方便地被链接到其他数据集中,与不同来源的数据混搭整合,构建新的服务,这就使得书目数据在数据层面可以做到超越数据库系统,直接附生于网络的基础架构之上。

2.2 用SPARQL语言作为RDF数据查询语言。

SPARQL(SPARQL Protocol and RDF Query Language)的全称是SPARQL协议与RDF查询语言,由W3C的RDF数据访问工作小组设计维护,于2008年1月15日成为W3C的推荐标准。SPARQL语言在语法上也像关系数据库的查询语言SQL那样由SELECT、FROM、WHERE这三个主要部分组成,不同的是SPARQL的WHERE子句利用RDF的三元组模式来构造查询条件,构成基于图模式(Graph Pattern)的匹配。查询条件可以是一个三元组也可以是多个三元组,可以将查询变量(一般以?开头)置于主体、谓词、客体的位置上,三元组中主体、谓词、客体可以参引不同的领域。

下面是一个简单的 SPARQL 查询,包含前缀声明、SELECT-FROM-WHERE 子句、结果集过滤排序限制这几个基本的部分,意为查询国会图书馆某数据集中所有谓词为 bf: title 的实体中,得到其 bf: title 的值,并按值的升序排列,结果集限制为前 10 条:

```
PREFIX bf:<http://bibframe.org/vocab/>
SELECT ?title
FROM<http://loc.gov/catalog/dataset.rdf>
WHERE {
    ?x bf:title ?title .
}
ORDER BY ?title
Limit 10
```

与关系数据库面向数据结构的查询相比,SPARQL“主-谓-宾”结构的三元组模式是直接面向语义和知识的查询。SPARQL 的 FROM 子句指定的查询范围不再像 SQL 那样是结构多样的关系数据库表或视图,而是一个 RDF 文件,可以位于本地,也可以位于网络上的任何位置,只要有一个 URL,并且可以同时查询位于网络上不同位置的多个 RDF 文件,在某种程度上做到突破网域和领域的限制。

SPARQL 协议允许基于 HTTP 协议提交简单的 SPARQL 查询请求,关联数据应用系统通常会通过 SPARQL 端点来提供数据查询和消费服务,SPARQL 端点可看做是 SPARQL 协议的简单实现。例如 DBPedia 的 SPARQL 端点为: <http://dbpedia.org/sparql>,可以通过向 DBPedia 发送 HTTP GET 请求,返回的数据可用 XML 或 JSON 解析器解析。

```
http://dbpedia.org/sparql?query=SELECT+DISTINCT+?concept+WHERE+{+?s+a+?concept+}+LIMIT+50
```

该命令会向 DBPedia 的 SPARQL 端点发送如下 SPARQL 查询语句:

```
SELECT DISTINCT ?concept
WHERE {
    ?s a ?concept .
} LIMIT 50
```

2.3 良好定义的 RDF 数据序列化格式支持数据在网络范围内的传输和交互

RDF 的三元组模型是一种抽象的数据模型,须

经过序列化(Serialization)后才能成为可被机器读取和处理的数据。最新发布的 RDF1.1 支持多种序列化格式,如图 1 所示,除 RDFa、RDF、XML、N-Triples(N3)、Turtle 之外,还有支持多个 RDF 图的 JSON-LD、TriG、N-Quads。每种序列化格式为不同的目的设计,例如 JSON-LD 是可被 PHP、JAVA、C 等多种编程语言默认支持的 JSON 格式的一个子集,专为关联数据应用而设计,这种格式简捷清晰,一般用于机器与机器之间的数据交互,常用于关联数据消费接口的返回数据格式。而 RDF/XML 是 RDF 模型最为经典的序列化方式,与 XML Schema 配合使用,基于 XML 编码的 RDF 甚至能实现不同应用领域之间的互操作,但 RDF/XML 有着不利于人读和写的缺点。RDFa 是将 RDF 三元组嵌入 HTML 文档的一种序列化格式,适用于能方便修改 HTML 文档模板但难以介入系统体系架构的应用,如 Drupal 等内容管理系统^[10]。N-quads、N3 和 Turtle 和是一种纯文本的 RDF 序列化格式,适用于人读和写,N-quads 和 N3 是 Turtle 的一个子集,因其规定每一个三元组中的主体、谓词、客体都必须用完整的 URI 来表示,所以与 Turtle 和 RDF/XML 相比,文件会比较大,但这也是它的优点,因为它的每一行都可以单独解析,同时它也很适合压缩以减少在传输交换过程中的网络流量,所以 N3 适用于传输大型关联数据包,N-quads 和 N3 类似,只是在句法上有细微的差别,会在三元组的行末附加用于表示三元组所在上下文的信息,N3 的语法是: <subject> <predicate> <object>, N-quads 的语法是: <subject> <predicate> <object> <context> . .

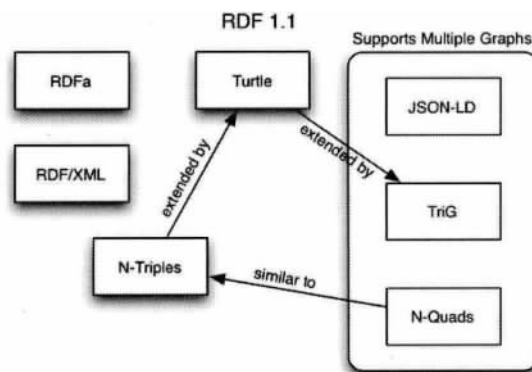


图 1 W3C 的 RDF1.1 推荐标准中支持的序列化格式及其关系^[11]

如: <<http://example.org/alice/foaf.rdf#me>> <

http://xmlns.com/foaf/0.1/name> " Alice" <http://example.org/alice/foaf.rdf> .。TriG 是专用于 RDF 数据集的序列化格式,以压缩文本的形式将一个完整的数据集打包在一个文本文件中。

2.4 大量使用基于事件驱动的网络触发器

网络触发器(Web Triggers)常被用于在网络站点之间进行主动的信息交互,是一种基于 REST 技术架构的点对点的交互技术,通过 HTTP Request 发送请求。主要理念在于主动“推”而非被动“拉”,由请求发起方定义请求的具体内容,接收方做出相应的回应。当一个站点有新的信息与另一个站点分享时,会主动发起连接和动作,而不是被动地等待被另一个站点发现,这种发起通常是由事件自动驱动的,如一条新信息发布、更改或删除时,触发一个程序,去另一个站点作相应的更新。

网络触发器的主要技术有 PingBack、TrackBack、Web Hooks, TrackBack 常用于博客内容更新的推送。GitHub 是一个著名的在线应用程序库,它有一种名为网络钩子的工具(Web Hooks),提供一种触发机制,当程序有更新时,会将更新主动同步到部署了这些应用程序的 Web 服务器上。Pubsubhubbub 是 Google 的一个推送订阅协议,作为 Atom 和 RSS 的扩展,提供强大的、简单的、开放的服务器端信息推送和订阅功能。

图书馆的许多业务工作需要跨网域的分布式交流和协作,例如联合编目,网络触发器可为这种协作提供便捷性,并基于更细微的粒度,一些必须人工完成的工作可以由触发器自动完成,减少交互成本。在基于书目框架格式进行编目时,大量使用网络触发器,有望为图书馆的编目工作带来流程的改善和效率的提高。

3 采用知识本体对书目数据建模

2012 年 12 月发布的 BIBFRAME 定义了核心模型和注释模型,其核心模型包含四个大类:创造性作品(下文简称作品)、实例、规范、注释^[12]。2014 年 2 月 20 日,书目框架模型作为一种本体词表发布^[13],命名空间为: http://bibframe.org/vocab/,前缀为 bf,同时提供 rdf 文件以供下载。在该文件中,语句<owl:Ontology rdf:about="http://bibframe.org/vocab/">定义该词表为一个 OWL 本体。除此之外,与核心模型四大类相关的其他资源也都

被抽象为与这四类同级的资源类,都作为 bf: Resource 类的子类,例如事件(Event)、关系(Related)、题名项(Title)、标识符(Identifier)、语种项(Language)等,这些在 MARC 中以文本出现的属性值,在 BIBFRAME 中都立体化了,而载体项和出版项这些带着明显印本书特征的属性,被归到更广泛的类目之下,如出版项是“提供者(Provider)”的一部分,这是因为 BIBFRAME 还要描述除印本资源外的数字资源¹⁴, Provider 被定义为所有与资源出版、印刷、发行、发布、生产相关的代理(Agent),包括人、机构等,图 2 是将 BIBFRAME 本体导入 Protege4.3 客户端后可可视化的结果,可以看出 BIBFRAME 包含的所有类:

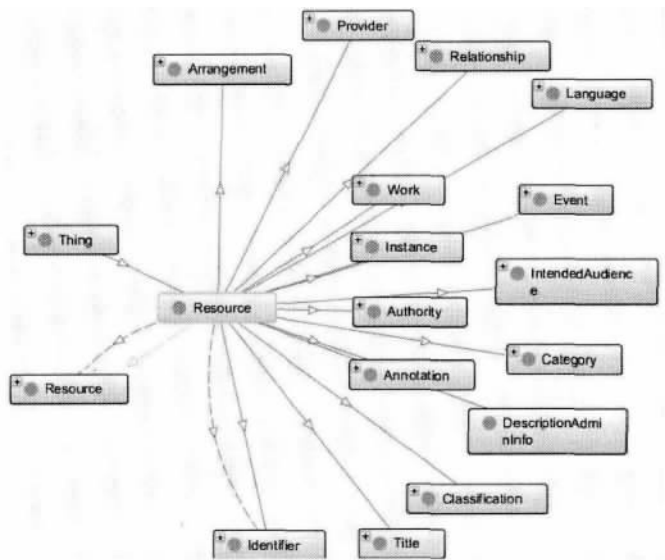


图 2 BIBFRAME 的本体模型

知识本体通常用属性来表达类与类之间的关系,每个类有自己的独有的属性和从父类继承的属性,属性是连接类与类之间的桥梁,以“作品”(Work)为例,它与大部分类之间都存在着关系,如图 3,方形代表类,不同颜色的有向箭头代表不同的关系,如紫色实线箭头表明一个类是另一个类的子类,虚线箭头代表一个类与另一个类之间有一个基于属性建立的关联关系,一般的惯例是类的开头字母大写,属性开头字母小写。这种类与类之间的关系可以用三元组表达出来,例如<http://bibframe/work/frbr-report>是一个作品,有一个属性是“著者项(creator)”,著者项的取值范围是关系(Related)类,该类有一个属性 relatedTo,指向一个机构(Organization),其 URI 为:<? http://bibframe/

auth/org/ifla>?,还有一个属性是 relationship,其值是一个文本,指向责任方式: Author,另一个可以用于指向责任方式的属性 relationshipUri,可以指向一个 URI,如 RDA 责任方式词表中的一个规范词条<http://rdvocab.info/roles/authorWork>。机构(Organization)是代理(Agent)的子类(subClassOf),而代理(Agent)又是核心类“规范(Authority)”的子类。

利用 RDF 三元组模型来表达这种关系就是:

```
<http://bibframe/work/frbr-report>
rdf:type bf:Work
<http://bibframe/work/frbr-report> bf:
creator <http://local/relationship001>
<http://localhost/relationship001>
rdf:type bf:Related
bf:relatedTo <http://bibframe/auth/org/
ifla>
bf:relationship "Author"
bf:relationshipUri <http://rdvocab.info/
roles/authorWork>
```

```
<http://bibframe/auth/org/ifla> rdf:
type bf:Organization
bf:Organization rdfs:subClassOf bf:Agent
bf:Agent rdfs:subClassOf bf:Authority
```

知识本体是领域知识的抽象,经过上述的形式化后,数据成为某一个类的实例,进而成为经过严格定义的、可识别可定位的语义单元,数据和数据之间的关系用来自规范词表的谓词(属性)来表达,可被机器理解和处理,因而数据不再是对机器而言没有意义的字符串,而具备了语义,BIBFRAME 的基于知识本体的模型和词表,为书目数据成为语义网的一部分奠定了基础。

BIBFRAME 的注释(Annotation)模型体现了 BIBFRAME 的创新之处,也是 BIBFRAME 开放性和灵活性的主要体现。注释模型建立在语义网以往的“开放注释”研究和探索的成果之上,它直接继承了 W3C 的开放注释本体(Open Annotation, OA),bf:Annotation 是 oa:Annotation 的一个子类,W3C 的 OA 本体定义了 Annotation 的模型和框架,如有 Body 和 Target 这两个主要部分,定义了书签、评论、描述、提问、回答、标注等注释行为,bf:Annotation

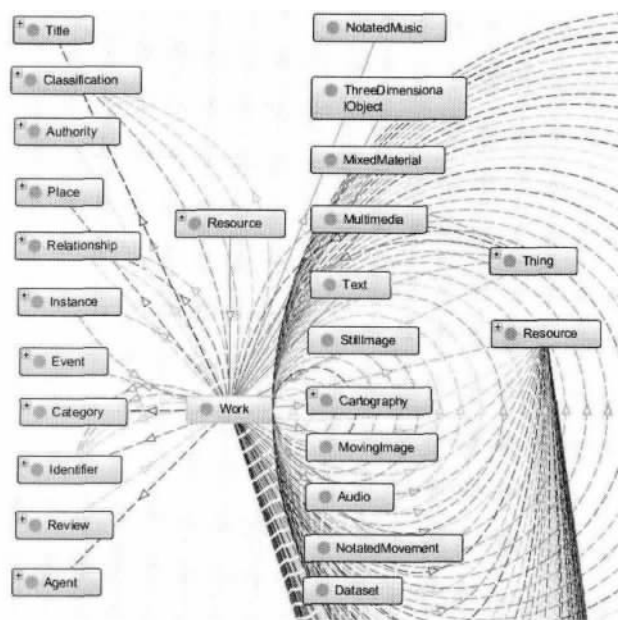


图3 BIBFRAME 作品的本体模型

继承了 oa:Annotation 的框架,并对其进行了扩展,定义了适合于书目数据的子类和属性,以及如何与作品、实体和规范建立关联的方法。BIBFRAME 的注释模型使得核心模型更抽象、更清晰,更规范。作品、实例、规范这三类并不能完全涵盖到全部的书目数据,有许多与这三类相关的信息,但又超出这三类本身,如关于作品的评论、关于出版者的介绍、关于作者的传记信息、某本书的封面、某复本的馆藏信息等,将这些复杂又分散的信息单独建模,使三类核心类的边界更为清晰,关系更易于描述,有助于 MARC 数据向 BIBFRAME 模型转换。注释模型中的实体相互独立,又与核心模型中的实体有千丝万缕的联系,关于作品的评论、关于出版者的介绍、关于作者的传记信息、某本书的封面,这些信息通常在书目数据系统之外,如关于作品的评论,可能位于 SNS 网站上,这样就便于为书目数据建立更广泛的外部联系,书目数据的触角可以伸向网络上的任意一个地方,使得书目数据成为网络的一部分。另一方面,也能提升用户服务,例如不同来源的书评可以方便地集成到书目服务平台中,出处和时间等信息一目了然,读者就可以无需打开每一个网站。

4 基于网络资源的规范控制

基于网络资源的规范控制是 BIBFRAME 的一个重要需求,也是进入网络时代后,对书目控制最基本最迫切的需求,在过去的十年中,图书馆界在这方

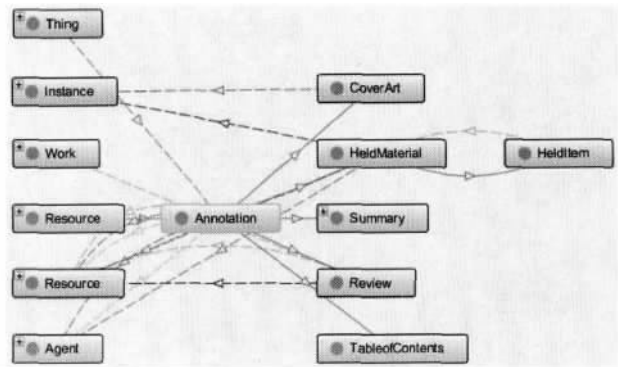


图4 BIBFRAME 注释的本体模型

面做出不少的努力^[15]。近年来,大量的规范词表被发布为关联数据,每一个规范词条都有一个 HTTP URI,可以在全网络范围内被参引。美国国会图书馆自 2009 年开始尝试将国会主题词表(LCSH)发布为关联数据,到目前为止,在 id. loc. gov 网站上,已经有包括 LC Name Authority File、LC Classification、Thesaurus for Graphic Materials 在内的十余个词表被发布成关联数据,OCLC 负责运营的国际虚拟规范档(VIAF),有十多个国家图书馆参与。

规范数据是 BIBFRAME 核心模型的重要组成部分,BIBFRAME 的规范控制与以往的网络规范控制方法并非竞争甚至替代的关系,而是提供一个轻量级的抽象层、一个框架,定义规范数据如何与作品和实例关联,以及如何指向已有的规范数据^[16]。

在最新发布 BIBFRAME 词表中,规范数据有 4 个子类:代理 (Agent)、主题 (Topic)、地点 (Place)、时间 (Temporal),其中代理还分人 (Person)、机构 (Organization)、会议 (Meeting) 等 5 个子类。2014 年 3 月 7 日发布的书目框架规范数据草案中,这些类都继承了来自父类 Authority 的一个属性:hasAuthority,该属性指向一个已有的规范数据 URI,如下 RDF 数据意为:Cutright, Paul Russell 是一个人 (bf: Person),它在国会图书馆的规范数据中有一个 URI,如果一个规范实体有多个可参引的 URI,可以用属性 bf: referenceAuthority 来表示,该属性可重复,理论上可指向无限个网络上的规范数据资源对象,如下 RDF/XML 格式的例子是一个关于人的规范数据:

```
<!-- BIBFRAME Authority -->
<bf:Person>
< bf: authorizedAccessPoint > Cutright,
Paul Russell, 1897 - </bf: authorizedAccess-
```

Point>

```
<bf:hasAuthority rdf:resource = http://
id. loc. gov/authorities/names/n80022903"/>
<bf:referenceAuthority rdf:resource = "
http://viaf.org/viaf/91312848"/>
</bf:Person>
```

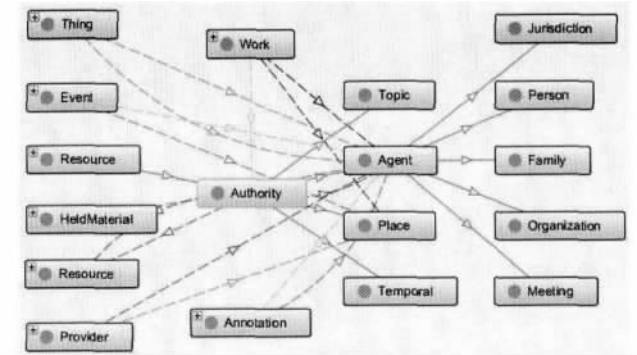


图5 BIBFRAME 规范数据的本体模型

5 用例及实现

国会图书馆给出了 15 个用例用于说明 BIBFRAME 的功能^[17],本文选取了部分用例,从 BIBFRAME 为用户服务改善和图书馆的业务流程改进的角度,分析了这些用例在网络环境下的适应性和实现方式。

5.1 对用户服务的改善

(1)方便地利用地理空间信息提供 SoLoMo 服务。

当某用户需要查找距离某地 30 英里范围内的图书馆中某指定的馆藏,可用如下 SPARQL 语句实现,用户输入作品题名 (bf: title" PhantomTollbooth"),限定查找条件为该作品的实例,并查找这些实例被哪些馆藏地收藏,从用户手持设备上获取所在地的经纬度坐标 40. 1583 83. 0742 30,用 SPARQL 语句实现如下:

```
SELECT ?work ?inst ?lib
WHERE {
?work ? bf: title" PhantomTollbooth"
? inst bf:instanceOf ?work
? holdings ? bf:holds ? inst
? holdings ? bf:heldBy ? lib
? lib ? gs:nearby(40. 1583 83. 0742 30).
}
```

(2)直接支持移动阅读。

查找某图书馆所藏的、关于某作品的、适合于某设备(指定媒体格式)的馆藏的 SPARQL 语句如下:

```
SELECT ?work ?inst
WHERE {
  ?work ?bf:title "Fahrenheit 451"
  ?work ?bf:hasInstance ?inst
  ?inst rdf:type bf:EBook
  ?inst bf:format <http://idpf.org/epub>
}
```

5.2 对编目工作的改善

(1)维护馆藏信息。

对大多数编目员来说,馆藏信息维护是主要的日常工作,例如,为某作品的电子书添加本地馆藏信息,首先,在联合编目系统中找到某作品(如题名为 *The Shadow Club*)的所有电子书,电子书是“实例”的一个子类,因而,需要 3 个查询条件,首先要限定题名为 *The Shadow Club* 的作品(Work),其次查找该作品的实例,再次标明实例的类型(type)为电子书(Ebook),用 SPARQL 语句表达如下:

```
SELECT ?work ?inst
WHERE {
  ?work ?bf:title ?TheShadowClub"
  ?work ?bf:hasInstance ??inst
  ?inst ?rdf:type ?bf:EBook
}
```

● 查到符合条件的实例后,拷贝实例到本地系统,赋予一个本地 URI:<http://local.library.org/examples/shusterman/test001/i1>

● 添加它与本地作品的关系:

<http://local.library.org/examples/shusterman/test001/w1>

bf:hasInstance ? <http://local.library.org/examples/shusterman/test001/i1>

● 为该实例添加馆藏信息:

<http://olentangydistrict.library.org/examples/shusterman/test001/i1>

bf:assertedBy ? <http://library.local.org/people/mary>

(2)补充完善“作品”信息。

一个“作品”出版后,或者再版?,或者被改编成电影戏剧等其他形式,或者被翻译成其他语种,会经

常地需要更新相应的信息,在书目框架格式中,作品以一个 HTTP URI 来唯一标识和定位,可在全球范围内共享,被全世界的图书馆员参引关于该作品的详细信息,一般是由多个 RDF 三元组组成的集合,当有新的信息产生时,只需经过简单的触发机制,就能及时更新到全球共享的三元组集合中。例如,某英文作品被翻译成印度语,如何将这个信息更新到作品信息中,过程如下:

● 在网络上的规范控制服务中查到该作者 (URI 为

<http://bibframe.org/auth/people/kurt_vonnegut>))的该作品(bf:title ?Bluebeard"):

```
SELECT ?inst ??work
WHERE {
  ?work ?bf:title ?Bluebeard"
  ?work ?bf:author ? <http://bibframe.org/auth/people/kurt_vonnegut>
}
```

● 拷贝该作品 (<http://local.library.org/bluebeard/test001/w1>)到本地,赋予新的 URI (<http://national.library.org/barbablu/test001/w1>),为作品添加额外的语种信息,即如下三元组:

<http://national.library.org/barbablu/test001/w1> bf:language
<http://id.loc.gov/vocabulary/languages/ita>

● 触发 Web Trigger,给原作品维护机构发更新请求,告知该作品有了一个新的语种:

POST http://bibframe.org/bluebeard/test001/w1

Content-Type: application/x-www-form-urlencoded; charset=utf-8

title=BIBFRAME;hasTranslation+Update+Request&.url=http://library.local.org/barbablu/test001/w1

● 于是在原作品的三元组集合中就增加了一条三元组陈述:

<http://national.library.org/barbablu/test001/w1>

bf:hasTranslation ? <http://national.library.org/barbablu/test001/w1>

(3)便于采购和编目的协作。

采购馆员在采购阶段就能够向自己的目录中添加“作品”、“实例”和“馆藏”信息,虽然此阶段编目馆员尚未为此资源添加全面的信息,但读者可以在资源采购阶段就能了解某资源已在“采购中”,随着采编流程的进行,信息逐步得到补充。当资源采购到位后,编目馆员开始编目,当她打开本馆的编目系统查找与该资源采购订单相关的信息时,会发现联合编目系统中已有该资源的书目信息更新,她接受了这些更新。

● 采购馆员在联编系统中查找与订单匹配的作品和实例:

```
< http://library. local. org/examples/king/
test001/w1>
```

```
bf: hasInstance < http://bibframe. org/exam-
ples/king/test001/il> ;
```

● 拷贝作品和实例并添加馆藏信息包括状态信息:

```
< http://library. local. org/examples/annota-
tion/king001/a1>
```

```
a bf:HeldItem;
```

```
a bf:HeldMaterial
```

```
a bf:Annotation;
```

```
bf: holds < http://bibframe. org/examples/
king/test001/il>;
```

```
bf: heldBy < http://bibframe. org/library/
delaware_public>;
```

```
bf: status " on order";
```

```
bf: assertedBy < http://bibframe. org/li-
brary/columbus_metro>.
```

● 采购馆员在联编系统中设置了触发器,在原始“作品”和本地“作品”之间添加了一条“注释”信息:

```
< http://library. local. org/annotation/
king001/w2>
```

```
a bf:Annotation;
```

```
a bf:Watcher;
```

```
bf: pingBack < http://library. local. org/
hub/? id= http://library. local. org/exam-
ples/king/test001/w1>;
```

● 当原始“作品”信息有更新时,会利用网络触发器发送通知给本地编目系统:

```
POST http://library. local. org/hub/? id=ht-
```

```
tp://library. local. org/examples/king/test001/w1
```

● 本地编目系统收到更新通知后,接受更新,系统会自动在本地“作品”中添加以下更新信息:

```
bf: subject " http://bibframe. org/auth/subject/
City and town life-North Carolina-Fiction. " ;
```

● 在本地实例中添加以下更新信息:

```
bf: publisher < http://bibframe. org/auth/
org/hardcase_crime> ;
```

在这个过程中,由于 BIBFRAME 书目数据模型面向网络的技术架构以及网络触发器的作用,采购和编目之间的协作更紧密深入和高效,联编系统的长处得到更充分地发挥。而实现这些功能并不依赖于复杂而多样的系统逻辑,而是基于数据本身,也就是说,在数据层面就能解决某些以前必须由系统才能解决的复杂计算,这无疑是一个革命性的变化。

5.3 规范档的维护

在图书馆的检索系统中,经常遇到一个作者有多个笔名的情况,在 BIBFRAME 中,可以很方便地利用本体语言来解决这个问题,用 owl:sameAs 在不同的规范实体之间建立“等同(Equivalence)”关系。

```
<http://local. library. org/person/14790462>
```

```
a bf:Person;
```

```
bf: label " Richard Bachman";
```

```
<http://local. library. org/person/97113511>
```

```
a bf:Person;
```

```
bf: label " Stephen King";
```

```
<http://local. library. org/person/14790462>
```

```
owl: sameAs <http://local. library. org/person/
97113511>.
```

这样,当读者检索 Richard Bachman 的作品时,同时也能检索到 Stephen King 的作品。

6 结语

BIBFRAME 在功能需求上着重支持语义网技术的利用,以适应新的网络环境,同时支持国际化和本地扩展,这相对于 MARC 来说,是一个革命性的进步,将会对图书馆的用户服务和编目工作带来颠覆性的变革。BIBFRAME 采用的知识本体建模法和基于网络资源的规范控制技术都是语义网的基石,为封闭的书目数据仓储向开放互联的语义网转

变夯实了基础,基于网络资源的规范控制框架为图书馆融入万维网,成为语义网的组成部分,发挥图书馆已有的高质量数据的潜在价值,提供了方法和路径。但 BIBFRAME 并非是一个激进的数据格式,它的首要功能仍然是要完全容纳已有的 MARC 数据,兼容 MARC 时代的系统和技术,以支持从 MARC 到 BIBFRAME 的平滑过渡和转换,因而它仍然要求兼容遗留的 MARC 数据和基于 MARC 数据的技术如 SRU、Z39.50、OAI 等,虽然 BIBFRAME 的关联数据模型相比于 FRBR 模型更为简化和清晰,数据格式比 MARC 更为开放和适应网络环境,但在词表的设计上并不能做到完全抛弃过去的包袱,面临着词表无限扩充,难以确立边界的问题,在技术实现上,也没有成熟的实现方案和成功的案例可供借鉴,功能需求设计的完美和实际实施的矛盾的博弈也将持续。本文主要在文献调研的基础上分析了 BIBFRAME 在语义网环境下的功能需求和实现方式,试图展示 BIBFRAME 对图书馆用户服务和业务流程的改善和改进的巨大可能,下一步的研究工作将继续从实践的层面利用 BIBFRAME 的关联数据模型和技术体系改造图书馆的传统 MARC 数据,对 BIBFRAME 使图书馆数据融入互联网成为语义网的一部分从而发挥更大价值的能力作进一步的验证。

参考文献

1 Library of Congress. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. 2012-11-21. [2013-09-12]. <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>

- 2 刘炜,张春景. 网络资源的规范控制. 现代图书情报技术, 2008(12):27-31
- 3 刘炜,夏翠娟. 书目数据新格式 BIBFRAME 及其应用. 大学图书馆学报, 2014(1):5-13
- 4 Tim, Berners. Lee. Design Issues. [2014-03-06]. <http://www.w3.org/DesignIssues/>
- 5 林海青等. 图书馆关联数据:机会与挑战. 中国图书馆学报, 2012(1):55-65
- 6 刘炜. 关联数据:概念、技术及应用展望. 大学图书馆学报, 2011(2):5-12
- 7 Working Group on The Future of Bibliographic Control. On The Record. 2008-01-09. [2014-02-08]. <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>
- 8 Library of Congress. Transforming our Bibliographic Framework. 2011-05-13. [2014-03-06]. <http://www.loc.gov/bibframe/news/framework-051311.html>
- 9 同 1
- 10 夏翠娟等. 关联数据的发布技术及实现——以 Drupal 为例. 中国图书馆学报, 2012(1):49-57
- 11 W3C RDF Working Group. RDF 1.1 Concepts and Abstract Syntax. 2014-02-25. [2014-03-03]. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
- 12 同 3
- 13 Library of Congress. BIBFRAME Vocabulary. [2014-04-08]. <http://bibframe.org/vocab/>
- 14 Library of Congress. BIBFRAME Resource Types Discussion Paper. 2013-06-25. [2014-04-08]. <http://bibframe.org/documentation/resource-types/>
- 15 同 2
- 16 Library of Congress. BIBFRAME Authorities. 2014-03-07. [2014-04-08]. <http://www.loc.gov/bibframe/docs/bibframe-authorities.html>
- 17 Library of Congress. BIBFRAME Use Cases and Requirements. 2013-08-21. [2013-11-10]. <http://bibframe.org/documentation/bibframe-usecases/>

作者单位:上海图书馆,上海,200031

收稿日期:2014年4月10日

BIBFRAME as a Semantic Web Oriented Bibliographic Framework

Xia Cuijuan

Abstract: This paper studies on how BIBFRAME adapt to the new cataloging environment with Semantic Web technologies by looking through the technical details of the functional requirements and implementation. It mainly focuses on the technology architecture based on Linked Data, the latest release of BIBFRAME vocabularies based on Ontology modeling, and the authority control framework in the network environment, which are all core ideas and technologies of Semantic Web. It also gives out several typical cases to draw a conclusion that BIBFRAME could improve library services and library cataloging workflow on the data level.

Keywords: MARC; Bibliographic Framework; Semantic Web; Linked Data