

# 书目数据新格式 BIBFRAME 及其应用\*

□刘炜 夏翠娟

**摘要** MARC 格式被认为是目前制约图书馆数据开放利用的最大障碍,已经不适合当前的编目环境。基于书目控制的历史和当前背景及需求,通过剖析新的书目框架格式 BIBFRAME 的关联数据模型、需求与用例、工具与服务,分析其开放性、语义化、向后兼容性等特点,提出其在应用上将遇到技术门槛过高、与 FRBR/RDA 等已有标准的融合与冲突、与 Shema.org 的竞争与合作关系等问题。BIBFRAME 不仅是 MARC 的替代品,还将为图书馆数据融入更广阔的互联网环境带来巨大的潜力和可能性。

**关键词** MARC 书目框架 语义网 关联数据

书目数据是图书馆进行书目控制的重要基础,是揭示馆藏、开展服务的主要工具,也是传统图书馆对于知识进行组织和整序的最有价值的贡献。数字时代的书目数据虽然只是图书馆类型丰富的元数据中的一种,但它肩负着将纸媒带入数字时代的重任,同时又是出版物得以传播利用的重要桥梁,因此其重要性并没有降低。当前最大的问题是如何将书目数据开放成真正的万维网上的数据?这其实并不是系统实现问题,而是书目数据格式问题。由于语义技术的发展,人们才真正明白,MARC 格式是当前制约图书馆数据开放利用的最大障碍,网络时代呼唤新的书目数据格式<sup>[1]</sup>。

## 1 背景和历史

MARC 格式设计于上世纪 60 年代末。当时计算资源极其匮乏,计算机速度慢,存储介质昂贵,以磁带这种顺序存取介质为主,因此 MARC 记录格式(即 2709 格式)被设计为不定长字段的紧凑型格式,以记录为单元,每条记录代表一个编目单位。MARC 最初的设计目的是为了打印目录卡片,后来有条件的图书馆逐渐开放了检索终端,发展为后来的标准 OPAC 服务。由于磁带格式只能顺序检索而无法随机存取,因此在格式中设计了很多机器识

别码、记录及字段子字段标识,以便能通过倒排索引而迅速定位检索。后来虽然随机存储介质(磁盘)和关系型数据库技术得到普及,但 2709 格式依然作为 MARC 记录的交换格式,图书馆的 MARC 记录虽然都可以通过万维网查询,但除了按照客户机/服务器时代研发的 Z39.50 标准开放 API 接口之外,基本没有其它互操作方式,MARC 数据一直被牢牢地圈养在每个图书馆的 OPAC 范围内,缺乏方便地生成一个国家或地区的总书目或联合目录的技术和能力,更不用说开放给整个社会各相关行业使用。

当今的信息服务业有这样一个趋势:凡是互联网上不存在的,基本上是可以被忽略的。图书馆作为一种向大众传播知识的社会机构,如果遭到大众的忽略,就无法充分实现其职能,因而也无法实现其价值。自 MARC 产生到现在已经过去了半个世纪,随着技术的发展、信息大环境的改变,读者利用图书馆资源的对象和方式也发生了巨大的改变。MARC 格式的种种局限在网络时代越来越成为一个绊脚石,越来越阻碍图书馆行业成为一个与其地位相称的网络存在:首先,互联网让计算机不再是简单的数据处理工具,而成为全球信息空间的一个个节点,书目控制不再局限于某一台机器、某一个系统、某一家机构,而需要在图书馆以外的网络空间进行传输和

\* 本文为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)和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)的成果。

交换, MARC 因其领域上的封闭性已无法满足这种需求。其次, 图书馆要保存和处理的资源不仅限于传统的印本图书, 还包括诸如电子文献、网页、图像、多媒体等复杂的资源类型, 而 MARC 是为印本图书设计的, 虽然曾在网络资源的描述上做过一定的努力(856 字段), 但效果并不理想。

都柏林核心元数据组织 (DCMI) 很早就看到了 MARC 对网络应用的不适应, 从上世纪 90 年代中期开始发起了一轮新型元数据研发风暴, 试图为 MARC 等传统的文献描述方式找到替代者。都柏林元数据核心集、术语词表、抽象模型及应用纲要等都是 DCMI 为网络时代书目控制做出的重要贡献。对于图书馆元数据, DCMI 很早就提出一套名为“图书馆应用纲要”的元数据方案, 试图实现类似 MARC 的描述功能, 但这个方案只是一个属性元素集合, 相对于 MARC 这种既包含丰富得多的属性元素集、又含有著录和语法编码规则的“复合型”标准来说, 远不能满足实际需要, 因而一直缺乏大型图书馆的支持而始终停留于草案阶段, 未能成为 DCMI 的推荐规范。在数据编码方式上, MARC 虽然号称“机读”, 但是其各种内部格式只是被计算机读取, 而无法被机器“理解”。DCMI 尝试引入 RDF/XML 作为语义描述规范, 但由于图书馆通行的编目规则 (AACR2) 是和 MARC 配套的, 无法直接采用 RDF 编码, 而且语义技术的实用化也有一个过程, 当时还不足以支持大规模的应用, 因此 MARC 的地位长期以来难以撼动。

国际图联 1998 年出版了《书目记录的功能需求》(FRBR) 报告<sup>[2]</sup>, 提出与书目描述相关的三类实体、属性及其相互之间的关系, 指出传统的 MARC 无法满足报告中提出的这种层次化的、全网域的“查找、标识、搜索和获取 (FISO)”四方面的需求。随着以兼容过去并实现 FRBR 为目标的 RDA 编目规则的出台, 美国国会图书馆对 MARC21 进行了大量扩展, 人们对 MARC 格式的不满达到了一个顶点。

呼吁以新的书目格式取代 MARC 的呼声早已有之, 进入新千年之后更是不绝于耳。加州大学数字图书馆项目前经理、现任 OCLC 高级项目主管 Roy Tennant 在 2002 年公开喊出 MARC Must Die (MARC 必须死)<sup>[3]</sup>。但它为什么还死不了呢? 最主要的原因是一直没有一种能够满足网络时代书目数据功能需求的书目格式来取代它。

美国国会图书馆 2006 年 12 月成立了未来书目控制项目组, 2008 年 1 月提交最终报告 *On the Record*, 表明将书目控制向万维网推进的积极态度, 引发巨大争论。OCLC 领导了 W3C 的 Schema 书目扩展社区的研究工作, 通过对一种由 Google、Yahoo 和微软等提出的 Schema.org 编码推荐规范进行扩展, 在其 WorldCat 中尝试基于 Schema.org 的书目格式, 实现基于 FRBR 的 RDA 编目数据的网络化。所有这些探索都指向一个目标: 直接用万维网原生的技术——即以 URI、HTTP、HTML/XML 为代表的技术来处理 and 发布信息。

随着 W3C 等标准组织的推波助澜, 以及语义网、关联数据技术在图书馆内得到越来越多的应用, 美国国会图书馆已经在将 VIAF、LCSH 等规范词表发布成关联数据, 德国、英国、瑞典等国家图书馆纷纷进行了书目数据的关联数据化<sup>[4]</sup>, 推出新一代书目数据格式的时机已经成熟。在这种背景下, 基于关联数据技术的“书目框架 (BIBFRAME)”应运而生。

2011 年 5 月, 美国国会图书馆正式宣布与以语义技术起家的 Zepheira 公司合作开发 BIBFRAME, 同年 10 月, 国会图书馆发布项目计划, 2012 年 5 月, Zepheira 公司开始评估相关的项目并开始数据建模, 11 月书目框架发布第一份报告<sup>[5]</sup>, 公布了所用的关联数据模型草案, 同时在书目框架官网上发布用例与需求、不断更新的词表以及适当的工具和服务, 不列颠图书馆、德国国家图书馆、OCLC 等参与了早期测试。2013 年的 ALA 大会上, 书目框架成为热门话题<sup>[6]</sup>, 根据 Eric Miller 等人的报告, 目前国会图书馆仍然在根据早期测试的结果修正各种细则, OCLC 在研究如何将 WorldCat 中所有的数据转换成书目框架格式<sup>[7]</sup>。接下来的工作将更加具体化, 如继续更新术语词表, 用 XML 来对书目框架的 RDF 数据模型进行序列化的试验, 开发支持 MARC21 到书目框架的转换工具、及支持书目框架数据浏览的关联数据浏览器。平台提供商中, 目前有图书馆系统服务商 VTLIS 宣称其 Open Skies 平台支持书目框架<sup>[8]</sup>, 科罗拉多大学启动了“BIBFRAME 数据存储库”项目<sup>[9]</sup>, 在 Redis 图书馆服务平台上, 增加了一个模块, 通过 MODS-to-MARC 映射实现 MARC21 记录到 BIBFRAME 实体的转换。

尽管书目框架希望创造一种新的表达书目数据的方法用以取代 MARC,但它的目标比此更加宏伟。作为一项计划,BIBFRAME 要调研书目描述、数据创建、数据交换等各个方面。除了取代 MARC,它还能适应不同内容模型和边际规则,探索新的数据记录的组成方法,以及评估目前的交换协议。书目框架出现之后,废弃 MARC 的时间已经临近,美国国会图书馆和 OCLC 都宣布,将于 2016 年停止对 MARC 格式的支持,这就意味着 MARC 即将退出书目控制的历史舞台,书目框架将扮演重要的角色。

## 2 什么是 BIBFRAME ?

BIBFRAME 全称为“书目框架计划(The Bibliographic Framework Initiative)”,也曾称为“书目框架迁移计划(The Bibliographic Framework Transition Initiative)”,它的目的是开发一种“适应未来需求”的书目数据格式,因此也用它来命名这种新的书目数据格式。所谓格式,即适用于某种特定应用的编码和解码的打包方式,书目框架就是对于书目信息按照一定的结构和方式进行编码,并能够满足信息交换等功能需求而进行解码的一种“数据包”。书目框架应用了关联数据技术,能够支持图书馆及类似机构对各类馆藏资源进行描述和编码,其研发的第一个需求是支持数以亿计的 MARC 数据转换,将来所有新产生的数据将可以不再使用 MARC,从而逐步取代 MARC21 等世界各国沿用至今的 MARC,使得书目数据在整个万维网上得到方便的交换、发布和共享。

书目框架项目的最终目的,是要形成一套适应万维网的书目信息应用规范,主要包括三个方面:模型与实体、术语词表、编码及打包规则,模型与实体提供了书目框架的应用对象和描述深度,术语词表提供了规范控制的基础,编码和打包规则提供了机器理解的手段。另外书目框架还会开发一些用于数据校验、互操作测试等特定功能的辅助工具,以及发布一些应用平台信息和最佳实践等。目前书目框架的官网除了发布进行中的规范之外,作为一个项目管理网站(部分内容须密码登入),还兼有征求意见、讨论问题和普及宣传相关知识的功能。

### 2.1 框架模型

与 FRBR 类似,书目框架采用了实体—关系分

析法,对所涉及的实体、实体属性、实体关系、属性关系等进行了分析。但书目框架的分析更一般化,只将描述对象分为抽象的创造性作品(Creative Work)和实例(Instance)两层,另外根据规范控制和扩展描述的需要,单独定义了规范数据和注释数据,因此一共有四类实体(如图 1、2 所示):

- Creative Work(创造性作品)
- Instance(实例)
- Authority(规范数据)
- Annotation(注释数据)

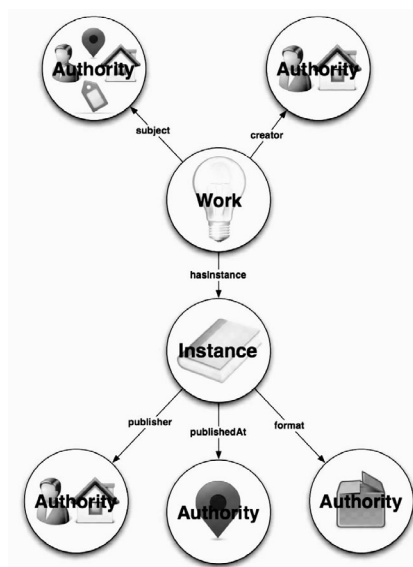


图 1 书目框架的核心模型<sup>[10]</sup>

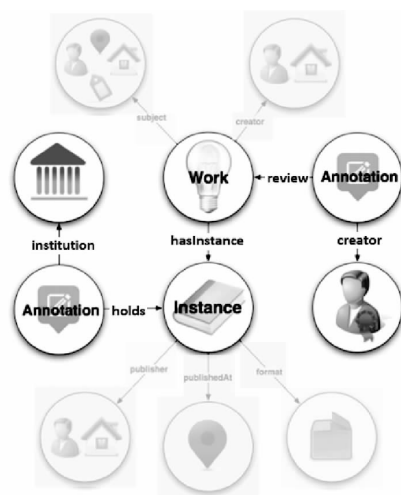


图 2 书目框架的注释模型<sup>[11]</sup>

前两者可以通过对 MARC 中各类字段属性进行梳理而“继承”,而规范数据与传统规范控制的做法有很大不同,它既要实现传统规范控制对作品、实

例及其相关的作者(人物及角色)、机构、主题、事件等要素的规范功能,又有许多网络资源规范控制的特点(如外链 VIAF 或 DBPedia 中的 RDF 数据)。书目框架定义的规范数据格式并非要取代其它的规范控制方法,而是作为一个轻型的抽象层,提供一种容器,使规范控制在万维网环境下更加有效地发挥作用。

对与作品或实例相关的实体的描述都作为“注释”,是书目框架的一种“创造”。它所涉及的各种注释包括目前已经纳入考虑范围的封面设计、书评、描述和馆藏描述四类,其中“描述”又分三个子类:摘要、文摘和目录,其所需描述的属性、关系等完全不同,但它们都是 `bf: Annotation` 的子类,这种完全异构的数据只有在书目框架这种采用了 RDF 描述的模型中才能方便地进行编码,因为我们可以把“注释”类当作一个容器,而不像关系数据库那样,一定要强求结构的统一。目前书目框架只定义了一些急需描述的基本类型,如上述四类和三个子类,主要是考虑到书目框架首先要兼容 MARC21,如果在转换 MARC21 数据时没有对应的映射字段,就无法满足对“框架”的需求。将来可能会根据需要扩展更多的类型,这样的模型也能够直接支持网络中由其它机构组织发布和维护的数据<sup>[12]</sup>。

## 2.2 术语词表

规范术语的发布是语义万维网技术的基础,通过为各类实体、属性、关系、取值等相关术语赋予 URI,形成词表,为语义编码(一般是采用 RDF 的各种序列化形式)提供了基本的结构单元。

从目前书目框架的实践来看,它已不再像 DC-MI 那样严格控制元素和术语的数量,而是根据实际需要原则,经过一定的内部讨论流程,随时发布和修订。目前术语的主要来源是 MARC、FRBR、RDA、DC、VAR 等相关业界的元数据规范。至 2013 年 9 月,已发布了近 300 个术语(10 月 1—10 日的美国政府关门也波及到美国国会图书馆,书目框架的术语页面至今没有恢复, MARC21 转换书目框架的试验项目只更新到 6 月 5 日)。

在众多的规范术语中,资源类型是较为特殊的一种:它既是书目框架描述对象的一种属性,又是书目框架所适用的资源类型。书目框架的“创造性作品”和“实例”两类都可以有资源类型属性。书目框架吸收了 MARC/MODS、RDA、DC、Schema.org 等

元数据规范中对资源类型的定义,将他们组合归并,初步形成一个由 14 种类型组成的词表(文字资料、地图、数据集、乐谱、动作谱、音频资料、图画、动画、三维物体、软件/多媒体、混合资料、资源集合、手稿和触觉资料),既具备科学性,又很实用,应用时可以任意组合,基本能够从概念上覆盖所有的图书馆及相关机构常见的资源对象,比 RDA 中利用内容类型、载体类型和媒体类型三方面组合的定义方式更加简便易行<sup>[13]</sup>。

## 2.3 用例与需求

软件工程中的“用例与需求”通常用于深度描述系统的使用场景和功能,让设计者和用户充分了解系统如何与外界互动,不同的用户角色可以利用系统做些什么,以及系统典型的开发目的和业务目标。多个用例可以较为全面地反映系统的应用需求。用例是普通读者了解系统复杂功能的一把钥匙,通常不会以专深的技术术语编写,而采用普通用户能够看懂并理解的日常用语。

目前书目框架的官方网站列出了 15 个用例,并归纳了 12 个方面的设计目标<sup>[14]</sup>。15 个用例中有 4 个是能够给用户带来的好处,11 个是给图书馆员(图书馆工作)带来的好处。从用例上看,书目框架给图书馆带来的变革将是巨大的。

通过这些用例和需求描述,我们能够很方便地看出什么是真正的“网络时代的书目格式”,它与传统的 MARC 格式在使用场景和满足需求方面,有多么巨大的不同。

例如它能够直接提供书目信息的 SoLoMo<sup>①</sup> 服务、选择“自适应”内容(根据类型、媒体类型、语种等)、进行规范控制(提供规范数据、人名规范名对应、规范档更新等)以及进行联合编目(合并、去重、加馆藏信息、复制数据等)。

## 2.4 工具、服务与演示

书目框架是关联数据的一种应用,是一种书目数据的 RDF Schema 和打包规则。之所以称为“框架”,是因为其提供了各种 RDF 序列化数据的打包方式。书目框架的四类数据应该有四种打包方式,

<sup>①</sup>注:SoLoMo,指 Social Local Mobile,即支持移动和本地化的社会化应用,在提供移动服务时能够感知地理位置,从而提供此时此地的个性化服务,同时这种个性化服务又是基于社会性网络的,可以附加很多诸如好友推荐、朋友动态之类的功能。SoLoMo 被认为是移动服务的一个趋势。

```

@prefix bf: <http://bibframe.org/vocab/>.
@prefix madsrdf: <http://www.loc.gov/mads/rdf/v1#>.

<http://bibframe.org/resources/works/003680557> bf:creator <http://bibframe.org/auths/names/n79038434>;
  bf:title "The robber bridegroom";
  a bf:Text,
    bf:Work;
  bf:language <http://bibframe.org/vocabulary/languages/eng>;
  bf:subject <http://bibframe.org/auths/subjects/sh2007102158>,
    <http://bibframe.org/auths/subjects/sh2008107229>,
    <http://bibframe.org/auths/subjects/sh2008107699>;
  bf:instance <http://bibframe.org/resources/instances/0036805570001>,
    <http://bibframe.org/resources/instances/0036805670001>,
    <http://bibframe.org/resources/instances/0036806010001>,
    <http://bibframe.org/resources/instances/0036806280001>;
  bf:derivedFrom <http://bibframe.org/resources/bibs/3681690>;
  bf:consolidates <http://bibframe.org/resources/bibs/1005582>,
    <http://bibframe.org/resources/bibs/8791933>,
    <http://bibframe.org/resources/bibs/8820540>;

<http://bibframe.org/auths/names/n79038434> bf:label "Welty, Eudora, 1909-2001.";
  madsrdf:authoritativeLabel "Welty, Eudora, 1909-2001"@en;
  a bf:Person,
    madsrdf:Authority;
  rdfs:label "Welty, Eudora, 1909-2001.".

```

图3 国会图书馆的 MARC 数据转换为书目框架格式的编码示例<sup>[17]</sup>

这些书目的编码除了必须支持 RDF/XML 之外,还可根据需要,支持 Ntriples, Turtle, RDF/JSON, RDFa 等形式,在应用中需采用内容协商和适当的媒体类型方式对 HTTP 请求做出反应。

为了让大家有一个直观的了解,书目框架项目网站上提供了两种工具<sup>[15]</sup>,可用来了解 MARC 数据转换成书目框架的效果。一种是比较服务,通过输入一个美国国会图书馆 MARC 书目记录的标识(MARC BIB 001 字段),就可以看到转换成书目框架格式前后的书目记录。另一种是转换服务,以 MARC/XML 格式粘贴或上载提交一个不大于 2MB 的书目数据文件之后,系统会利用 Exhibit 开发的工具将数据转换并显示成书目框架格式。这两种工具目前还是试验阶段,每隔几周会有调整和更新。将来书目框架网站上会提供更多的工具和服务<sup>[16]</sup>。

除了美国国会图书馆的大力倡导和投入外,书目框架项目也吸引了包括不列颠图书馆、德国国家图书馆、乔治华盛顿州立大学图书馆在内的许多著名图书馆参与,OCLC 也积极响应。目前在书目框架网站上提供了这些机构将自己的书目数据(MARC 格式)转换成书目框架格式的丰富例子,包括 MARC 书目记录转换成书目框架中的“作品”、“人”、“机构”、“实例”、“注释”等各类相互关联的实体的情况,并有 RDF/XML 格式的数据可供下载。这些演示初步展示了书目框架格式对不同 MARC

格式的包容性和适用性,验证了关联数据技术给书目数据带来的语义功能和数据间的互操作能力。

### 2.5 书目框架示例

图3 是国会图书馆的 MARC 数据转换为书目框架格式的一个简单的例子。该实例以 RDF/Turtle 格式编码,比 RDF/XML 编码方式更为简洁,易于人识读。该记录是关于一个作品(bf: Work)的数据,定义了该作品的 URI、题名(bf: title)和责任者(bf: creator),有 3 个主题(bf: subject)和 4 个实例(bf: instance),并描述了其来自于哪条书目记录(bf: derivedFrom)。底部代码描述了该作品的责任者,是一条人名规范数据。

### 3 BIBFRAME 的特点

知名图书馆技术专家 Karen Coyle 于 2004 年发文<sup>[18]</sup>认为,新的书目系统只满足 FRBR 的 FISO 四大需求,对于网络时代是不充分的,至少还要有“描述”、“发现”、“定位”、“购买”、“保存”和“推广”等功能。OCLC 的资深专家 Roy Tennant 也在同一期刊物上呼应了这个观点<sup>[19]</sup>,认为未来的书目数据要具有“通用性”、“可扩展性”、“开放性”、“透明性”、“低门槛”、“包容性”,支持“协同管理”、“模块化”、“层次结构”、“颗粒性”和“容错性”等,对未来的书目系统提出了很高的要求。

书目框架作为网络时代的书目数据格式,简而言之,主要在以下三个方面超越了人们对传统书目

数据的需求:(1)普适性。不仅体现在书目框架可描述的资源类型不限于描述传统图书,还可描述网络资源,还体现在其采用 RDF、URI 等标准化的数据模型和编码方式,可与网络上的数据很好地兼容和互操作。(2)语义化。因为其是采用 RDF 描述的,因而数据的最小单位不是记录,而是 RDF 的陈述,这就打破了 MARC 书目“记录”的格式,将记录分拆为一个或多个三元组,成为一个又一个相互关联的语义单元,为书目数据赋予了机器可处理的语义。(3)开放性。用 HTTP URI 来标识所有的“东西”是关联数据技术框架的基石,而 HTTP URI 是万维网的基本信息组织方式,可全球唯一定位,当书目数据中的一切实体都赋予了 URI,就可在全网范围内被定位和访问,突破了 MARC 数据的局限,打破了封闭的图书馆系统与开放网络之间的藩篱,使书目数据能够方便地与其他网络数据混搭,融入“数据的网络”,构建新的数据服务。

对书目框架种种特性的总结,基本可以归纳为以下四个特点:

#### 特点一:与 MARC 的兼容性

书目框架继承了 MARC 的很多特点,比如丰富的语义和严格的一致性,它能兼容作为内容模型的 FRBR,实现 MARC 数据的各种角色:规范数据、馆藏数据、分类数据等。它设计的首要任务是能够“容纳”当前所有的、各式各样的 MARC 记录,从而将图书馆行业数十年人工积累的编目记录悉数收入其中,平滑过渡到新的格式,由新的万维网时代的系统软件进行管理。但书目框架并不是 MARC 的严格等价物和简单的翻版,它尽可能地保证 MARC 的大多数语义描述能够保留,但也只是 MARC 可以转换成书目框架,而不保证书目框架完全可以转换成 MARC。

#### 特点二:超越书目记录格式

BIBFRAME 的开发目标是成为一种书目记录格式,但它不仅是书目记录的“格式”,也不仅适用于书目。取代目前图书馆主流书目记录格式 MARC,是书目框架计划最重要的目标之一,但它设计的抽象模型和技术框架显然不止于为书目记录提供一种格式,而是试图容纳多种领域本体、元数据方案、数据序列化格式的框架。书目框架试图描述的资源类型远不止书目,还包括博客文章、静态动态图像、3D 影像、计算机程序、软件、数据集、实物等等,并为

MARC/MODS、RDA、Schema.org、DCMI 等资源描述模式提供映射方案。因此书目框架的适用性并不局限于图书馆领域,也适合于博物馆、档案馆等机构,以及更广泛的类似机构。

特点三:严格区分抽象内容和它的物理/数字表现

在互联网环境中,必须要有一种方法将图书馆数据的抽象内容(题名和作者)与物理形态(页数、是否有插图等)区别开来。标识图书馆资源的创建者(如作者、出版者)与资源的内容属性(如主题)具有同等重要的地位。在书目框架的关联数据模型中,作品和规范数据属于抽象内容,实例数据属于载体表现,在描述时分开描述,一个作品与它的规范数据和实例之间是一对多的关系,都分别赋予 URI 后,可以更灵活准确地解释它们之间的关系,用 RDF 编码并序列化后,可被机器理解。

#### 特点四:面向语义网的书目数据格式

用 HTTP URI 来标识一切事物,是关联数据四原则中的第一条,也是在万维网上清晰标识信息实体、实现网络资源的规范控制的必须。书目框架继续把这一原则发扬光大,无论对作品,对人、地、机构等规范数据,还是每一个实例,甚至包括与“作品”相关的封面、书评、描述等都用 HTTP URI 来唯一标识。HTTP URI 的最大特点不仅仅在于唯一标识功能,还在于它的全球定位功能,以及从一个命名空间到另一个命名空间的跨网域链接。

“数据网”中实体之间要建立更多可被机器处理的关系,这种关系要跨越传统图书馆领域,覆盖到整个网络信息空间,以提高搜索引擎检索结果的相关性,帮助用户找到他们真正需要的信息,提升网络服务的效能。在网络范围内建立可被机器理解的语义关系,正是关联数据技术的长处之所在,书目框架的关联数据模型建立了表达并利用这些关系的基础。这样就提供了一种互联网环境下跨系统、跨领域的网络资源规范控制方法,这种设计使得书目框架既能够满足图书馆的特殊需求,又能与更广泛的相关应用进行整合。

## 4 BIBFRAME 的应用和问题

书目框架计划一经提出便受到业界瞩目,并非因为采用了特别先进的技术,而主要是因为美国国会图书馆的影响力。国会图书馆目前把书目框架作

为最重要的、具有最高优先级的项目进行开发。目前书目框架的开发尚未完成,主要还有三个方面的工作:第一、书目框架的规范控制超出了传统书目信息规范数据的边界,成为网络信息规范控制的一部分,但具体应该新增哪些内容的规范控制,尚未得到确定。第二、注释模型还存在争议,特别是其适用的类型以及将来如何扩展方面。例如对于将馆藏纳入到注释类型进行描述是否合适,尚有很多细节还未形成规范或达成共识。第三、目前的应用平台开发,图书馆管理系统的支持,以及演示、编目、转换、校验等工具软件还需要进一步丰富和完善。另外 MARC21 向书目框架进行映射转换过程中也有很多细节问题需要讨论,目前的测试主要集中于此,能否转换和容纳所有参与图书馆现有的 MARC21 数据,是书目框架能否成功的试金石,因此转换数据实现目前书目系统的基本功能,是当下书目框架应用的首要任务。

书目框架系出名门,人们对其寄予厚望,然而它毕竟自提出至今只有两年多时间,能否成功、在多大范围内成功还言之尚早,因此现阶段也是人们质疑最多的时候。归纳起来主要有以下几个问题。

#### 4.1 全新开发,技术门槛较高

从技术上来看,书目框架完全采用语义万维网技术,涉及到十余年来互联网和图书馆领域研究和应用的多项前沿技术,如 Web 发布、实体—关系(E-R)分析、RDF 数据模型、RDF 的序列化(编码)、XML 数据处理、关联数据的标准规范和应用实施、MARC 数据的 XML 化处理、书目数据的 FRBR 化、领域知识本体和元数据编码、规范控制等等,对于图书馆界的研究和应用人员来说,有较高的技术门槛,而对于系统开发商和平台提供商来说,提供可投入实际应用的解决方案也非易事。

另外书目框架的开发策略也与以前图书馆界的元数据方案研发有很大的不同,以往对于属性元素的选择非常慎重,对于各类规范词表的确定总是以尽可能优先复用为原则,而书目框架基本不考虑与其它元数据模式的复用和兼容,甚至不尊崇 FRBR 模型,完全摒弃了 MARC 用字段和子字段名来揭示书目的属性,也没有复用 DCMI 或 RDA 的术语词表,而完全在新的命名空间(bf: <http://www.bibframe.org/vocab/>)定义了一套新的术语。这种做法固然有助于新的数据格式从一开始就站上一个新

的起点,但也会给应用推广带来难度和阻力,且不符合语义万维网尽可能复用已有本体、词表的理念。

#### 4.2 与 FRBR 和 RDA 不完全一致

书目框架并非直接采用 FRBR,而是对 FRBR 进行了简化。作为一种新的书目数据格式,书目框架数据本来应该是 RDA 编目的直接结果,但目前由于 RDA 是 FRBR 模型的实践者,而书目框架简化了这种模型,从而造成书目框架与 RDA 编目数据不完全一致的现象。

好在这种简化是一种一般化,可以将书目框架看成是一种更大的“容器”,其通过采用一种“领域纲要”的机制来适应特定领域的特殊需求,因此实现 FRBR 的描述需求完全没有问题,从而也能经过扩展满足 FRBR 的功能需求。

“领域纲要”看起来是借鉴了 DCMI 的应用纲要,为书目框架应用于某个领域提供纲领性的指南。领域纲要规定了实体如何继承和分解,复用哪些词表、如何与 MARC 数据映射等具体问题。如果说书目框架是一栋房子,那么“领域纲要”则是房子里面的支架,用以呼应领域应用的具体需求。如需将 FRBR 化的书目数据转换成书目框架格式,需要定义“领域纲要”,明确规定 FRBR 中的 WEMI 四类实体如何转换成书目框架。如图 4 所示,“BIBFRAME WEMI Profile”是一个“领域纲要”,将 FRBR 中的 W(Work,作品)、E(Expression,内容表达)定义为书目框架中的“创造性作品”的修饰和限定,M(Manifestation,载体表现)、I(Item,单件)则作为“实例”的修饰和限定,这就是为什么书目框架虽然摒弃了 E 和 M,但仍然能够兼容 FRBR 的原因。

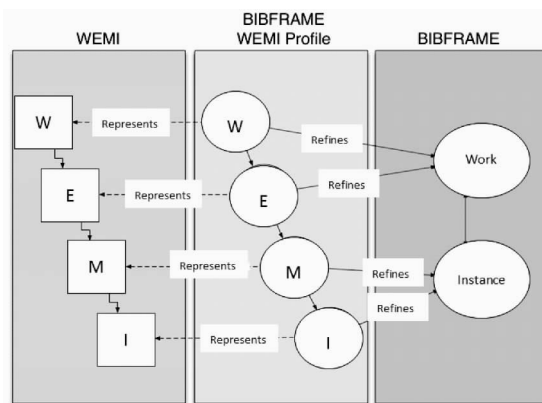


图 4 BIBFRAME 与 FRBR 的映射<sup>[20]</sup>

#### 4.3 与 Schema.org 书目扩展的竞争和合作

Schema.org 是互联网产业巨头 Google、Yahoo、Bing、Yandex 共同制定的一套网络资源描述规范,通过在网页中嵌入网络资源的语义描述(RDFa 形式),可使语义数据方便地以网页形式发布,并被搜索引擎检索以及进行后续(智能化)处理。

作为关联数据的一种实现方式,Schema.org 的崛起迅速引起了 OCLC 的注意。但 Schema.org 并不是为图书馆资源而设计,它缺少适用于书目数据的 FRBR 第一组概念(即 WEMI)、无法明确区分内容和载体、缺少对资源集合和连续出版物的定义等,因而不适合直接拿来作为描述图书馆资源的工具,必须进行扩展。于是 OCLC 的关联数据专家和图书馆技术专家为 Schema.org 设计了一套用于图书馆的书目扩展,即 SchemaBibEX。该扩展包括模型和词表两部分内容,如增加了“OCLCnumber”、“Holding”等馆藏特性元素用于描述物理载体。美国国会图书馆的书目框架计划是在 OCLC 的 SchemaBibEX 的模型发布数月后才发布的,二者描述的是同一类资源,存在竞争关系。鉴于国会图书馆的影响力,OCLC 立即开始了两者之间兼容和互操作的研究,谋求二者的合作,使其形成相互补充的关系。二者对书目数据描述能力的比较如图 5 所示。

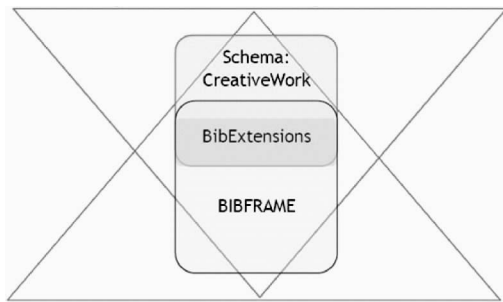


图 5 BIBFRAME 和 Schema.org

对书目数据描述能力的比较<sup>[22]</sup>

(注:倒立三角形代表 Schema.org 书目扩展覆盖的广度较大而深度较小,正立三角形代表 BIBFRAME 覆盖的深度较大而广度较小。)

作为 Schema.org 的一种领域应用,SchemaBibEX 可用于图书馆的语义数据在网上进行发布,因此书目框架也可以采用 SchemaBibEX 方式来发布,前提是 OCLC 主导的扩展规范能够兼容书目框架的元素和术语,特别是要解决具有相同或类似语义的元素和术语的处理问题。例如 OCLC 的扩展中所使用的 Work 与书目框架的 Work 容易引起歧

义,这种词表之间的相互关系,可以通过建立映射等方式来解决。但由于这两者都还处于应用的早期,许多问题尚未暴露,OCLC 与美国国会图书馆之间对于这两种书目格式的合作问题仍然存在许多不确定的因素<sup>[21]</sup>。

书目框架虽然还处在襁褓之中,但已经展示了其实现图书馆行业有史以来书目控制梦想的潜力。在业界,已有 VTLS、ExLibris 等公司加入了书目框架应用开发或测试的行列,一些图书馆开始尝试用大数据解决方案管理 RDF 数据<sup>[23]</sup>。每半年一次的项目通报会让业界不断看到新的进展,就目前而言,书目框架已经揭开了良好的序幕,展示了美好的前景。但是任何新生事物都是不完美的,面对未知的困难,它能够做到什么程度,目前下结论还为时过早。本文的介绍希望能让大家看到书目框架的巨大潜力和可能性,但要到达彼岸取得成功,还需要更多业界同行的参与、尝试、讨论和不断完善。

#### 参考文献

- 1 Library of Congress. A Bibliographic Framework for the Digital Age. 2011-10-31. [2013-09-16]. <http://www.loc.gov/bibframe/news/framework-103111.html>
- 2 IFLA. Functional Requirements for Bibliographic Records. [2013-09-16]. <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
- 3 Tennant, Roy. MARC Must Die. Library Journal (October 15, 2002): 26-27. [2013-09-16]. <http://libraryjournal.reviews-news.com/index.asp?layout=article&articleid=CA250046>
- 4 林海青,楼向英,夏翠娟. 图书馆关联数据:机会与挑战. 中国图书馆学报,2012(1):55-65
- 5 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/marcl-report-11-21-2012.pdf>
- 6 李恺. RDA、FRBR 和 BIBFRAME 的最新进展:2013 年 ALA 年会参会笔记. 数字图书馆论坛,2013(9):32-37
- 7 Ted Fons. LC Bibliographic Framework Initiative Update Forum (January 27, 2013). [2013-10-22]. [http://www.loc.gov/bibframe/pdf/ALAmw2013\\_\\_bibframe\\_\\_OCLC\\_\\_publicUpdate\\_\\_Fons.pdf](http://www.loc.gov/bibframe/pdf/ALAmw2013__bibframe__OCLC__publicUpdate__Fons.pdf)
- 8 VTLS. VTLS Open Skies: A smarter framework for library services. [2013-10-26]. <http://www.vtls.com/openskies>
- 9 Jeremy Nelson. Redis Library Services Platform (RLSP). [2013-09-26]. <http://tuttdemo.coloradocollege.edu/code4lib/redis-library-services-platform.html>
- 10 同 5



- 11 同 5
- 12 同 5
- 13 Library of Congress. BIBFRAME Resource Types Discussion Paper. 2013-06-25. [2013-09-16]. <http://bibframe.org/documentation/resource-types/>
- 14 Library of Congress. BIBFRAME Use Cases and Requirements. 2013-08-21. [2013-09-18]. <http://bibframe.org/documentation/bibframe-usecases/>
- 15 Library of Congress. BIBFRAME Tools. [2013-10-08]. <http://bibframe.org/tools/>
- 16 Library of Congress. BIBFRAME Frequently Asked Questions. [2013-09-20]. <http://www.loc.gov/bibframe/faqs/>
- 17 Kevin Ford. LC'S Bibliographic Framework Initiative: An Update. 2012-11-28. [2013-09-28]. <http://www.loc.gov/bibframe/pdf/kefo-swib12.pdf>
- 18 Karen Coyle. Future Considerations: The Functional Library Systems. *Library Hi Tech*, 2004, 22(2): 170-171
- 19 Roy Tennant. A Bibliographic Metadata Infrastructure for the Twenty-First Century. *Library Hi Tech*, 2004, 22(2): 176-177
- 20 Eric Miller. BIBFRAME Community Profile. ALA 2013. 2013-06-30. [2013-10-20]. <http://de.slideshare.net/zepheiraorg/alabibframe-lc20130630>
- 21 Carol Jean Godby. The Relationship between BIBFRAME and OCLC's Linked-Data Model of Bibliographic Description: A Working Paper. [2013-09-28]. <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>
- 22 同 20
- 23 Jeremy Nelson. Building a Library App Portfolio with Redis and Django. 2013-01-15. [2013-12-08]. <http://journal.code4lib.org/articles/7349>

作者单位: 上海图书馆, 上海, 200031

收稿日期: 2013 年 12 月 8 日

## Introduction to BIBFRAME as a Successor to MARC

Liu Wei Xia Cuijuan

**Abstract:** MARC, which is based on forty-year-old techniques for data management, can't meet the new requirements of today's bibliographic environment. The Bibliographic Framework Initiative (BIBFRAME) is developed not only as the replacement of MARC, but also as a successor to MARC because it could bring library community into the wider Internet world and become a part of the web. By analyzing the linked data model, the requirements and use cases, tools and services, the paper points out the problems of BIBFRAME as follows: how to work together with the existing standards such as FRBR and RDA, how to deal with the popular semantic standards schema.org and so on.

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

## 封面图片介绍: 广东药学院图书馆

广东药学院图书馆坐落于广州大学城广东药学院校区中部的荔枝山上。以“绿色环保”与“文化传承”为建筑理念的图书馆大楼倚山临水, 四季花开, 远处蜿蜒的行道旁芒果树郁郁葱葱, 与绿地上读书的学生形成一道美丽的风景。

图书馆馆舍面积为 3.1 万平方米, 有阅览座位 2900 个。主楼高十二层, 其中 1-6 层为大空间大构架的开放服务空间, 楼层高、空间大、采光透风好、功能区域布局合理, 为日常业务开展和读者学习提供了舒适的环境; 7-12 层为低层高的密集型储存书库, 经济实用, 易于管理。另有辅楼五层, 学术会议中心厅一层, 层高错落有致、别具一格。

截止到 2013 年 9 月底, 馆藏文献资源总量为 214.4 万册, 其中纸质资源 176.7 万册, 电子资源 37.7 万册, 引进中外文数据库 15 个。为服务于学校的医药学特色化发展, 图书馆以医药尤其药类文献作为重点馆藏, 医药类图书占总藏书的比例为 32%, 其中药类文献品种为华南地区图书馆中最多最全。