

关联书目数据模型比较研究*

吴贝贝 (中国科学院文献情报中心)
夏翠娟 (上海图书馆)

摘要 关联数据模型是图书馆关联书目数据项目实施的关键。图书馆领域的一些典型的关联书目数据项目在具体实施中采用了不同的关联数据模型。本文选取5个典型的图书馆关联数据项目,从内容描述规则、本体模型及词表、数据模型及格式和数据消费方式4个方面进行了分析和比较研究。

关键词 书目数据 关联数据 本体模型

DOI: 10.13663/j.cnki.lj.2015.05.012

Comparative Study of Linked Bibliographic Data Models in Libraries

Wu Beibei (National Science Library, Chinese Academy of Sciences)
Xia Cuijuan (Shanghai Library)

Abstract Linked data model is very important in the implementation of linked bibliographic data projects. There are different linked data models adopted in the library circle. This article makes comparative analysis and study on 5 typical linked data projects from 4 perspectives: content description rules, ontology model and vocabularies, data model and format, and data consumption.

Key words Bibliographic data, Linked data, Ontology model

1 关联数据与图书馆

图书馆作为文献信息资源的聚集地,在管理文献资源的过程中产生了大量的原生数据,最具有代表性的是书目数据和规范数据,但在网络环境下并未得到广泛深入的利用。书目数据的交换已经存在并发展了很多年,仍然需要通过 Z39.50 等协议获取,各个书目记录之间彼此独立没有关联,基于网络资源的规范控制更是没有得到普遍实施。^[1]自数字图书馆概念出现以来,如何在分布式异构信息环境中使用户能够全面准确地获取信息和知识,如何实现不同的图书馆系统间实现数据集的无缝链接,如何揭示不同知识单元间的各种相关关系逐渐成为图书馆从业人员关注的焦点。^[2]伴随着互联网的发展,基于传统信息组织和服务方式的图书馆,逐渐被隔离在网络世界之外,图书馆

中大量的人类文化遗产和知识没有能在网络世界上发挥应有的功能,其价值也被淹没在海量而快速更新的互联网资源之中。图书馆界需要一种新的技术手段解决上述问题。Tim Berners Lee 于 2006 年提出关联数据的概念,网络从文档之间的相互关联,发展到数据之间的相互关联,海量相互关联的数据构成 Tim Berners Lee 所设想的数据网络。^[3]随着 W3C 等标准化组织的推波助澜,图书馆界成为关联数据应用的积极支持者和身先士卒者,大量的关联书目数据

* 本文系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”(项目编号:13CTQ008)和上海市哲学社会科学项目“基于书目控制的网络信息资源的规范控制方法”(批准号:2009JG502-BTQ040)的研究成果之一。

项目的实践表明,关联数据不仅为解决数字图书馆互操作遗留问题提供了技术可能,同时也为图书馆资源融入互联网世界,并发挥应用的功能、产生新的价值提供了新的契机。

2 关联书目数据发展现状

图书馆将书目数据和规范数据以关联数据的形式发布,与网络中的其他资源发生关联并相互利用,是当前图书馆应用关联数据的主要模式。^[4]2008年起,瑞典国家图书馆将瑞典联合目录(Swedish Union Catalogue,即LIBRIS)整体发布为关联数据,是图书馆界第一次在关联数据领域的尝试。瑞典联合目录在发布为关联数据的过程中,使用的词汇不仅仅局限于图书馆领域,同时借鉴了SKOS、FOAF和BIBO本体等词汇表,为图书馆数据与广阔的语义网的融合打开了新的局面。^[5]2009年,美国国会图书馆开始尝试将国会图书馆主题词表(LCSH)发布为关联数据,至今美国国会图书馆已经发布了包括“名称规范档”、“分类法”在内的40余个规范档。此后,英国、法国、德国等国家相继将本国图书馆的书目数据发布为关联数据。作为全球最大的书目数据生产、维护和服务机构,OCLC也于2012年6月开始将被图书馆广泛收藏的书目数据发布为关联数据。根据datahub^[6]中的记录并结合OCLC关联数据研究^[7]。表1对各个国家图书馆关联数据项目进行了梳理,总结出目前14个国家级图书馆发布的关联数据集,内容涉及书目数据和规范数据两部分,同时指出与其相关联的主要数据集。除了这些国家级图书馆的关联数据项目,还有一些跨机构跨领域的复杂关联数据项目,如欧洲数字图书馆Europeana等,这些项目构建了更具包容性和开放性的本体模型,来整合图书馆、档案馆、博物馆等机构的资源,以实现资源之间的关联、重用与共享。综上所述,西方国家图书馆界在关联数据领域进行了很多实践,致力于将图书馆数据从孤岛中抽离出来推向网络环境。但是在图书馆关联数据资源中,中文书目数据几乎空白。国内仅有上海图书馆、中国科学院文献情报中心以及中国科学院技术信息研究所对关联数据的理念进行追踪

和研究,并没有公开以关联数据的形式发布书目数据和规范数据。

3 典型的关联书目数据模型

一个关联的开放数据系统涉及很多方面的内容,其基础和核心包括领域本体模型、数据模型等。本文选取5个典型的图书馆关联数据项目:英国国家书目、OCLC的WorldCat、美国国会图书馆的BIBFRAME项目、欧洲数字图书馆Europeana项目、美国数字公共图书馆DPLA项目,从内容描述规则、本体模型及词表、数据模型及格式和数据消费方式四个方面进行比较分析^[8],研究它们的特点,从而为中文书目数据发布为关联数据提供借鉴和参考。其中,内容描述规则指的是对资源的种类和特征进行描述的规则,如AACR2或RDA;本体模型及词表指的是领域的概念模型及描述概念所用的术语词表,如FRBR;数据模型及格式指的是数据的抽象模型和编码格式,如RDF及其序列化格式RDF/XML等;数据消费方式指的是以何种方式向外界提供数据服务,如SPARQL端点或API等。

3.1 大英图书馆:英国国家书目数据模型(BL Data Model)

2011年7月,大英图书馆提出要将英国国家书目(British National Bibliography,以下简称BNB)数据从MARC21格式转换为关联数据的RDF/XML格式,并发布为关联数据。目前已经发布为关联数据的资源包括已出版的图书和连续出版物,并将扩展到期刊、报纸以及电子资源等。^[9]

BNB只是将现有的MARC书目记录直接转化为RDF格式,因而在本质上其内容描述规则仍然是传统的AACR2。BNB也没有完全从零开始设计独立的本体模型和词表,而是尽可能利用现有的本体词汇,但没有采用FRBR模型,基于两个原因,一是时间的原因,从MARC记录中标识出FRBR实体是一件非常耗时的工作,二是当时的开发者认为FRBR是一个过于复杂的模型。BNB也没有完全采用图书馆界的本体模型和词表,比如ISBD的很多元素对取值范围约束较少,也没有足够的类来封装

表 1 各国国家级图书馆关联数据项目统计

机 构	发布时间	书目数据	规范数据	关联的主要数据集
瑞典国家图书馆	2008 年	LIBRIS	LIBRIS	Wikipedia、DBPedia、VIAF、LC Authorities
美国国会图书馆			Library of Congress Name Authority File (NAF)、Library of Congress Subject Headings	DBPedia、Geonames、Stitch、Marc Codes List
英国国家图书馆	2010 年	British National Bibliography (BNB)		VIAF、LCSH、Lexvo、GeoNames、MARC country、MARC language、DDC、RDF Book Mashup
匈牙利国家图书馆	2010 年	Hungarian National Library (NSZL) catalog		DBPedia、VIAF
意大利众议院图书馆	2011 年	Bibliography of the Italian Parliament and electoral studies (BPR)		VIAF
法国国家图书馆	2011 年	data.bnf.fr — Bibliothèque nationale de France	RAMEAU subject headings	Agrovoc、Geonames、Dbpedia、Wikipedia、VIAF、DDC、LCSH
德国国家图书馆	2012 年	Deutsche Nationalbibliografie (DNB)、20th Century Press Archives	Gemeinsame Normdatei (GND)	DOI、DBPedia、VIAF、LCSH、Geonames、Chronicling America
西班牙国家图书馆	2012 年	datos.bne.es		DBPedia、LIBRIS、Lexvo、Sudoc bibliographic data、VIAF
OCLC	2012 年	WorldCat	FAST、VIAF、DDC	FAST、VIAF、DDC
新加坡国家图书馆管理局	尚未完成			
美国农业部国家农业图书馆			National Agricultural Library Thesaurus	
日本国会图书馆			National Diet Library of Japan subject headings、Web NDL Authorities—National Diet Library of Japan	LCSH
捷克国家技术图书馆			Polythematic Structured Subject Heading System	DBPedia、LCSH
芬兰国家图书馆			Yleinen suomalainen aineistointi—YSA	

MARC 记录中的自由文本, 于是选择了大量非图书馆界的本体词表, 如 W3C 的 Organization 本体和 Event 本体、FOAF、GeoNames 地理空间本体等, 同时也定义了自己的类和属性即大英图书馆术语 (British Library Terms, BLT), 其中有些类继承了或扩展了现有本体中的词汇。利用这些本体和词汇, BNB 将书目数据中的人、机构、地、时、事、主题等实体抽取出来作为对象来标识和描述, 明确地区分了现实世界中的真实对象和文献记录本身, 不仅实现了书目数据中不同实体间的关联, 还实现了与 Geonames、DDC、LCSH、VIAF 等外部数据集的关联。^{[10]2-3}

BNB 图书和连续出版物采用了不同的数据模型, 包含了不同的类和属性, 在功能上也有一定的区别。如图 1 图书数据模型所示, BNB 使用了以下词表中的类和属性来描述图书资源: BIBO、DC、FOAF、Event、SKOS、OWL、RDF Schema、Organization、ISBD、书目信息词表 Bio、GeoNames 等。

BNB 提供批量下载 RDF/XML 的格式的书

目数据包获得, 还分别提供了针对人和机器使用的 SPARQL 检索服务。

3.2 OCLC: Schema.org 书目扩展 (Schema-BibEX)

OCLC 早在 2009 年开始将自家产品杜威十进制分类法 (DDC) 发布为关联数据, 作为“术语服务”, 使 DDC 成为了关联数据云中的一员, 将 DDC 的应用扩展到其他领域。^[11]2012 年, OCLC 开始将 WorldCat 中的一批核心书目数据发布为关联数据, 标志着 OCLC 正式将其丰富的书目数据推向网络环境, 到 2014 年 4 月, OCLC 已发布了书目数据 1.97 亿条。

OCLC 将 WorldCat 书目记录发布为关联数据的初衷是使其书目数据能够在网络环境中被搜索引擎发现, 以得到更多的利用。为了实现这个目标, 采用 Schema.org 作为本体模型和词汇来源。^[12]Schema.org 是 2011 年由 Google、Yahoo!、Yandex 等三个搜索引擎巨头发起并设计的, 为网络资源的语义描述和标记提供了一个核心词表, 以帮助搜索引擎更好地发现并利用隐含的数据。Schema.org 为书目资源的描述

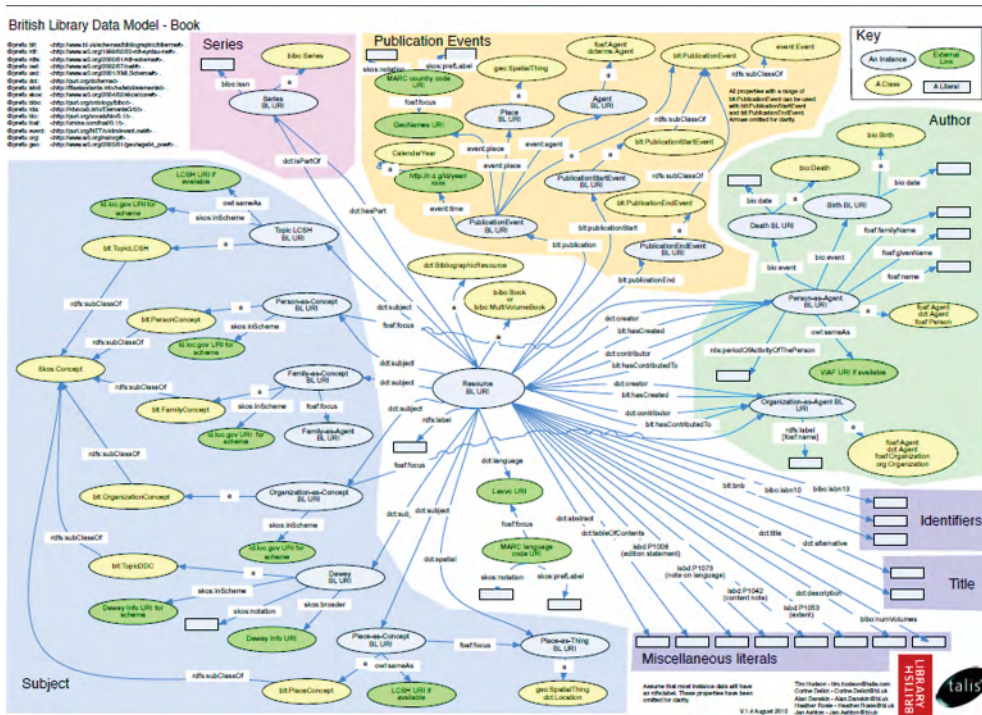


图 1 BNB 图书关联数据模型^{[10]4}

提供了基本的类和属性，目前主要的资源类型有论文、图书、创作作品、事件、人、地图、音乐记录、组织、地点、评论、列举、事物等 12 个，每一类下都有各自的属性。Schema.org 不是特定针对图书馆资源而设计，因此在描述资源的时候无法全面深入地反映图书馆的书目信息。为了弥补 Schema.org 的这一缺点，OCLC 基于 Schema.org 开发了一个图书馆扩展词汇集，即 SchemaBibEx，增加了 16 个资源类型和其他反映馆藏等信息的属性，用来配合 Schema.org 使之能更好地满足图书馆的元数据实践需求。

在实现方式上，采用在 HTML 网页中嵌入 RDF 三元组的方式来实现关联书目数据的发布，目前 WorldCat 上的每一个条目都被嵌入了 RDFa (RDF attribute) 和 Microdata 这两种数据格式。

与大英图书馆的关联书目数据相比，OCLC 在本体模型和词汇来源、数据格式以及实施方案及目标均有所不同。BNB 选择尽可能多的本体词汇实现与更多图书馆相关数据集的链接，而 OCLC 在将图书馆数据资源推向互联网世界的方向上更进一步，目标也更为明确，主要基于 Google 等产业界巨头所支持的 Schema.org 来自行扩展适用于图书馆书目描述需求的词汇，其目标是让 Google 等搜索引擎发现更多的图书馆资源，让图书馆资源在更广泛的范围内产生更大的价值。

表 2 BNB 和 WorldCat 关联数据模型比较

比较内容	英国国家书目 BNB	OCLC WorldCat
本体词汇来源	BIBO 本体、FOAF、ISBD、Org、OWL、SKOS、RDA、Event 本体、DC、BLT、Bio	Schema.org、SchemaBibEx
关联的数据集	VIAF、LCSH、Lexvo、Geonames、MARC 国家代码、MARC 语言代码、DDC、RDF Book Mashup	VIAF、FAST、DDC
数据编码格式	RDF/XML	RDFa、Microdata

3.3 美国国会图书馆：书目框架 (BIBFRAME)

2008 年，美国国会图书馆 (Library of Congress，以下简称 LC) 发布了关于未来书目控制的报告，探讨新的书目数据格式标准，以替代不能适应网络环境的 MARC，BIBFRAME 初次被提到。^[13]2012 年，LC 联合 Zepheria 公司共同设计了 BIBFRAME 模型，这一模型以网络为基础构架，以关联数据为设计原则，试图构建一个通用的描述框架并力图实现两个目标：整合现有的书目资源，满足图书馆具体的编目需求；创建一个新的书目环境，使图书馆融入到更广泛的网络环境。^[14]BIBFRAME 模型设计了 4 个核心类来建构主要的概念模型框架：创造性作品 (CreativeWork)、实例 (Instance)、规范 (Authority)、注释 (Annotation)，如图 2 所示，四个核心类下有很多子类，每个类根据资源描述的需要分别有多个属性。

在内容规则方面，虽然 BIBFRAME 试图构建一个可以与多种编目规则相适应的模型，但是模型的主要促成因素是新编目规则 RDA 的实施。^[15]BIBFRAME 模型对 RDA 所采用的 FRBR 模型进行了简化。如图 3 所示，FRBR 中的“作品”和“内容表达”在 BIBFRAME 中简化成为“创作的作品”；“载体表现”的内容在 Bibframe 中的“实例”中得到表达；RDA 与馆藏信息有关的“单件”则归并到 BIBFRAME 中的“注释”。^[16]通过对比可以发现，BIBFRAME 模型弱化了内容表达这一概念，原因是在 FRBR 的研究和实践中，现存书目记录的信息不足以将作品清晰地划分为不同的内容表达形式，而在网络环境中，对内容表达形式的区分也变得不那么重要。^[17]除了明确定义核心模型的四大类外，一些与四大类相关的其他资源也被抽象为与这四类同级的资源类，比如事件 (Event)、关系 (Related)、题名项 (Title)、标识符 (Identifier)、语种项 (Language) 等，都作为 bf:Resource 类的子类。

规范模型是 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。^[18] 规范类包含代理 (人、机构、家庭等) 地、时、主题等子类。

注释模型体现了 BIBFRAME 的创新之处,也是其开放性和灵活性的主要体现。它涵盖了诸如关于作品的评论、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,这些信息可来自于书目数据系统之外,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。^[19]



图2 BIBFRAME 关联数据模型^[20]

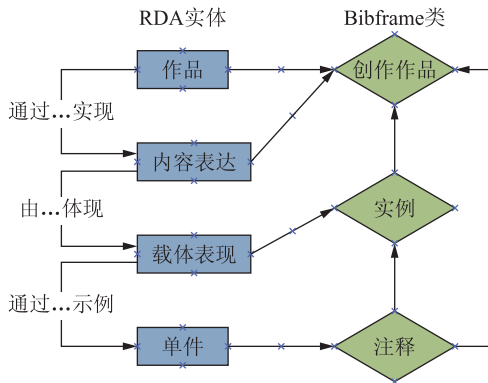


图3 RDA 实体和 BIBFRAME 类的对应关系^[16]

在本体词表的设计方面, BIBFRAME 没有复用已有的词表,而是创建了自己的词表和命名空间: <http://bibframe.org.vocab/>, 前缀为 bf, 独立管理术语的定义和更新, 目前共有 338 个术语, 其中类 63 个, 属性 275 个。

BIBFRAME 作为一种关联数据模型, 使用 RDF 来作为抽象的数据模型, 利用 SPARQL 语言作为 RDF 数据查询语言, 并大量使用基于事

件驱动的网络触发器来提供主动的数据交互。

3.4 欧洲数字图书馆 (Europeana): 欧洲数据模型 (EDM)

欧洲数字图书馆 (以下简称 Europeana) 项目的目标是在一个通用的元数据方案之下, 使用一个通用的平台将全欧洲的图书馆馆藏聚合起来。Europeana 包含来自 1500 个机构的超过 2 亿条记录和 1 千万个数字对象。^[21] 欧洲数据交换协议鼓励所有参与 Europeana 共享的元数据都发布为 Creative Commons CC0 1.0, 以便在欧洲范围内能够无限制的复用这些元数据。

Europeana 以关联数据的形式提供数据集, 项目仓储由 2 千万条记录组成, 使用欧洲数据模型 (Europeana Data Model, 以下简称为 EDM) 组织数据, 支持跨元数据提供商的复杂记录的关联, 同时这个元数据模型通过关注“对象”及其生命周期中的“事件”来追踪资源。

Europeana 采用来自 OAI-ORE 规范的概念和词表为资源描述提供了一个基础模型。OAI-ORE 词表和数据结构能够描述元数据陈述中供应商、对象和版本之间的复杂关系。EDM 为模型中绝大多数的类和属性提供了编目纲要, 《欧洲数据模型映射纲要》中详细说明了 EDM 和其他元数据方案的映射情况。EDM 数据集可以链接到很多的外部词表, 包括到 GeoNames 的地理名称链接、到 DBpedia 的人名链接等, 由于这些关联的存在, EDM 中管理元数据的内容规则更多的关注与这些链接相关的 RDF 句法。

如图 4 所示, EDM 的本体模型将资源划分为 3 个核心类: edm: ProvidedCHO、edm: WebResource、ore: Aggregation。其中 edm: ProvidedCHO 指的是被描述的对象本身, 可以与 BIBFRAME 中的 Work 层和 Instance 层相对应; edm: WebResource 是指 edm: ProvidedCHO 的数字表现形式; 而 ore: Aggregation 则涉及模型的聚合功能, 它是对提供商所提供资源的一个集合。EDM 的前两个核心类型将书目世界划分成了两个层次: 资源本身及它们的数字表达。而 ore: Aggregation 类则遵循了 ORE 的方法, 将提供商提供的所有对象及它们的数字表达聚合在一起成为一个整体, 定义了 edm: ProvidedCHO 和 edm: WebResource 之间的关

系。这种资源划分类型既可以追踪资源对象本身又可以记录资源的描述性元数据。图4展示了一个ORE聚合和EDM编目资源之间的关系。

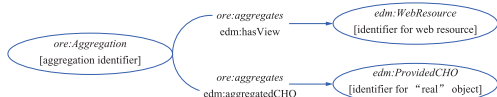


图4 European 关联数据模型^[22]

EDM 试图能够精确地获取资源提供商和元数据的版本信息。这样的目标使得 EDM 需要一个相对复杂的元数据方案，能够将关于资源对象的元数据和这个资源的对象联系起来。为此，EDM 增加了 ore:Proxy 类，这样尽管一个对象拥有不同的数字表现形式，但是依然能够将它们共有的描述信息集中在一起。

Europeana 项目提供 JSON 格式的数据下载，有些数据集也可以通过 RDF/XML 和 RDF/N-Triples 的形式获取。除了这两种形式，可以通过 Europeana SPARQL 终端查询数据，以 RDF/XML、JSON、N3/Turtle 和 N-Triples 格式返回数据。

3.5 美国数字公共图书馆 (DPLA)

美国数字公共图书馆 DPLA 项目于 2010 年发起，旨在将美国图书馆、档案馆和博物馆的资源进行整合。2013 年 4 月，DPLA 官方发起了一个发现平台，提供由 18 个成员机构贡献的原始数据集，包括 3200 个条目，超过 200 万条记录。^[23]

DPLA 的数据模型基于 RDF 建模，但是也采用了 RDF 驱动的序列化形式 JSON-LD，通过 API 进行输出，数据收割方法采用了 OAI-ORE 标准。在内容标准方面，DPLA 在本质上是一个数据收集和共享服务，因此没有针对具体的编目工作设计。数据模型中有一些针对特定属性的内容及格式规则，但主要是吸收了其他词表中的规则，例如 dpla:country 采用了 IOS 3166-1 国家代码。DPLA 元数据模型基于欧洲数据模型（简称为 EDM）建立，目前仍在完善中。这个元数据模型以 RDF 为基础，使用 DC、DCTERMS 和 EDM 属性收集描述性和管理性的元数据。如图 5 所示，DPLA 元数据模型共有 6 个类，分别是 dpla:SourceResource、ore:Aggregation、edm:WebResource、

dcmitype:Collection、edm:TimeSpan、dpla:Place。其中，最核心的类是 dpla:SourceResource，指各种文化遗产对象，包括书目世界中的各种“创作作品”，它类似于一个 ore:AggregatedResource 类型，因此 DPLA 使用了 EDM 模型中的 edm:aggregateCHO 属性与 ore:Aggregation 类发生关联。DPLA 的数据使用 JSON-LD 进行编码，以 RDF/XML 形式存储。

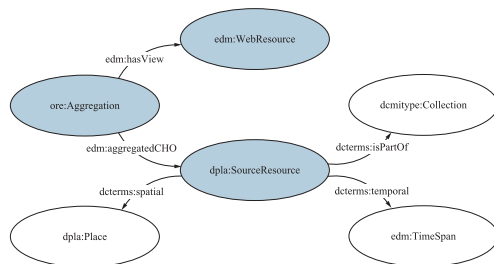


图5 DPLA 关联数据模型^[24]

4 比较研究

表 3 将本文选取的 5 个关联数据项目从内容描述规则、本体模型及词表、数据模型及格式和数据消费方式 4 个方面进行集中总结和对比，存在着相似性和差异性，归纳为以下几个特点：

(1) 图书馆利用关联数据技术的本质目的是将本地资源与外界资源联系起来，并将图书馆的资源呈现到网络上，为更多的人利用。BNB 和 OCLC 发布的主要是图书馆现存的大量的 MARC 格式的书目数据，通过选择不同的本体模型和转换路径将 MARC 数据转换成 RDF/XML 或 Microdata 等格式，以关联数据的形式发布，满足了图书馆资源外联和呈现于网络的基本需求。BNB 使用了多个来自图书馆领域或其他领域的本体词汇，实现了与多个外部数据集的关联，大大扩展了 BNB 的资源内容；OCLC 则通过使用搜索引擎认可的 Schema.org 词汇让书目记录成为网络资源的一部分。而 BIBFRAME、Europeana 和 DPLA 则是构建了更加宏观的模型，不拘泥于特定的内容规则，试图容纳图书馆、档案馆和博物馆的资源，实现数字资源的聚合。虽然 BIBFRAME 目前仍处在书目数据的转换阶段，但模型构建的初衷是在未来相当长的一段时间内，为图书馆、档案

馆、博物馆等文化机构创建一个以“网络”为中心的书目环境,使内在关联成为普遍存在。

(2) 在对书目世界的揭示方面, BNB、OCLC 和 BIBFRAME 关联数据模型从本质上是围绕图书馆书目资源展开的, 将与文献相关的一系列描述项目抽象成了上层概念, 每一个概念下都有深刻揭示与此概念相关的其他要素, 也就是详细的元数据描述。而 Europeana 和 DPLA 则更关注的是资源的聚合, 使用 ore:Aggregation 核心类, 将代表文化遗产的“作品”与其数字表达集合起来。Europeana 和 DPLA 并没有专门针对图书馆文献资源给出具体的元数据方案, 而只有对资源的通用描述, “作品”的数字表达大多来源于不同的提供者, 使用的元数据方案可能各不相同。

(3) 在互操作层面上, DPLA 采用了简化的 EDM 模型, 在数据模型中使用了 EDM 的子类, 对比它们的元数据方案, 可以发现它们共用了很多相同的词表如 OAI-ORE、DC 等, 这些很好地支持了两个关联数据项目之间的互操作。BIBFRAME 和 SchemaBibEx 则是都采用了简化的 FRBR 模型, 两者在一些实体如作品、实例、组织和个人等的定义方面具有类似性。二者均定义了“作品”类, 均将 FRBR “内容表达”作为类的属性, 实际上是弱化了“内容表达”这一概念。这些都为两个模型的映射

提供了基础, 同时这种冗余反映这两个有着不同动机和用户的项目之间的趋同性, 实际上 OCLC 也提供了 SchemaBibEx 与 BIBFRAME 的映射方案; 而 BIBFRAME 作为专为图书馆及其他文化继承机构设计的模型, 在描述深度上非常契合图书馆的需求, SchemaBibEx 由于背靠适用于网络世界的 Schema.org, 其对图书馆资源的描述深度不如 BIBFRAME, 描述广度则非 BIBFRAME 所及, 正体现了二者的互补性。^[25]

5 结语

关联数据不仅是一种技术手段, 在具体服务与图书馆相关的“人”与“机器”时, 要根据不同的应用目的, 设计相应的总体架构和数据模型。目前我国图书馆界对关联数据的研究和利用还处在探索阶段, 尚未将图书馆的书目数据等资源广泛发布为关联数据, 需要根据实际情况借鉴国外先进的关联数据项目经验。如果想在改变内容描述规则的前提下将书目数据发布为关联数据, 可以参照 BNB 的数据模型, 复用已有的本体词表, 将自己的书目数据和规范数据发布为关联数据; 如果想更多地融入互联网、使流行的搜索引擎更易于发现自己的资源, 可借鉴 OCLC, 背靠 Google 等业界巨头, 采用 Schema.org 数据模型; 而如果整合

表 3 典型关联书目数据模型比较研究

比较内容	BNB	OCLC WorldCat	BIBFRAME	Europeana	DPLA
内容描述规则	本质上是 AACR2	本质上是 AACR2	目前主要是 RDA, 但 BIBFRAME 设计意图为不局限于任何内容规则	定义了资源之间的各种关系, 没有特定的内容规则	除了多面元数据外, 其他大多没有定义内容规则
本体模型及词表	尽可能利用外部本体词汇, 如 BIBO、DC 等	Schema.org 和 Library 词汇扩展	主要是 BIBFRAME 词表	扩展使用外部终端资源	是简化的 EDM 模型, 有 DPLA 特定的元素
数据模型及格式	RDF/XML	RDFa 和 Microdata	RDF/XML、JSON	RDF/XML	RDF/XML
数据消费方式	SPARQL	基于 API	存在数据转换工具将 MARCXML 格式的数据转换成 BIBFRAME 数据、支持 SPARQL	SPARQL、APIs	基于 API

图书馆、档案馆、博物馆的文化资源，构建一个统一的信息基础设施平台时，可更多地关注 Europeana 和 DPLA 等资源聚合项目；作为 MARC 的替代者，以美国国会图书馆牵头、世界各大图书馆参与设计和试验的 BIBFRAME 模型，也值得关注和借鉴。在发布关联书目数据的同时，既要深入研究资源中知识内容关联的揭示以及信息组织深度序化等问题，又要考虑

到数据的消费与利用方式，如何更方便地为人所用，如何更深入广泛地融入整个网络环境中，成为网络的一部分，这都依赖于有一个科学的、可持续性发展的、能够不断生长的数据模型。本文对几个具有典型意义和行业影响的关联书目数据项目进行了深入分析和比较研究，以期对国内关联书目数据的实施起到一定的参考作用。

参考文献

[1] 刘炜. 关联数据: 概念、技术及应用展望[J]. 大学图书馆学报, 2011(2):5-12.

[2] 欧石燕. 面向关联数据的语义数字图书馆资源描述与组织框架设计与实现[J]. 中国图书馆学报, 2012, 38(202):58-71.

[3] Berners-Lee, T. On the Next Web[EB/OL]. [2015-02-25]. http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html.

[4] 林海青, 楼向英, 夏翠娟. 图书馆关联数据: 机会与挑战[J]. 中国图书馆学报, 2012(1):58-67.

[5] LIBRIS[EB/OL]. [2015-02-28]. <http://libris.kb.se/>.

[6] datahub[EB/OL]. [2015-02-28]. <http://datahub.io/>.

[7] Le Van R. OCLC Linked Data Research[EB/OL]. [2015-02-28]. <http://www.oclc.org/research/themes/datascience/linkeddada.html?urlm=168906>.

[8] Elings W M, Waibel G. Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums[J]. First Monday, 2007(3): 7-14.

[9] Taylor C. Libraries innovating with Linked Data[EB/OL]. [2015-02-28]. <http://collabor8now.com/2011/07/libraries-innovating-with-linked-data/>.

[10] Corine Deliot. Publishing the British National Bibliography as Linked Open Data[EB/OL]. [2015-02-28]. http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf.

[11] OCLC. Dewey summaries as linked data[EB/OL]. [2015-02-28]. <http://www.oclc.org/dewey/webservices.en.html>.

[12] Schema.org[EB/OL]. [2015-02-28]. <http://schema.org/>.

[13] Marcum D. A Bibliographic Framework for the Digital Age[EB/OL]. [2015-02-28]. www.loc.gov/bibframe/pdf/bibframework10312011.pdf.

[14] Coyle K. Bibliographic framework initiative[EB/OL]. [2015-03-01]. <http://www.loc.gov/bibframe/>.

[15] Coyle K. MARC vs RDA[EB/OL]. [2015-03-01]. <http://kcoyle.blogspot.com/2011/09/marc-vs-rda.html>.

[16] 吴贝贝, 宋文. 从 MARC 走向 BIBFRAME—后 MARC 时代的书目记录[J]. 图书情报工作, 2014(9):88.

[17] Shapiro D. Loving & Leaving FRBR[EB/OL]. [2015-03-02]. <http://www.slideshare.net/DebShap/frb-talk>.

[18] Library of Congress. BIBFRAME Authorities[EB/OL]. 2014-03-07[2015-02-28]. <http://www.loc.gov/bibframe/docs/bibframe-authorities.html>.

[19] 夏翠娟. 面向语义网的书目框架(BIBFRAME): 功能需求及实现[J]. 大学图书馆学报, 2014(6):61-69.

[20] Bibliographic framework as a web of data: linked data model and supporting services[R/OL]. [2015-03-02]. <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>.

[21] Haslhofer B, Momeni E. Europeana Connect, Results and Resources[EB/OL]. [2015-03-02]. http://www.europeanaconnect.eu/documents/europeana_ts_report.pdf.

[22] EDM primer[EB/OL]. [2015-03-02]. <http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation>.

[23] Welcome to the Digital Public Library of America[EB/OL]. [2015-03-02]. <http://dp.la/info/2013/04/18/message-from-the-executive-director>.

[24] DPLA Metadata Application Profile[EB/OL]. [2015-03-02]. <http://dp.la/info/developers/map/>.

[25] 刘炜, 夏翠娟. 书目数据新格式 BIBFRAME 及其应用[J]. 大学图书馆学报, 2014(1):5-13.

吴贝贝 女, 中国科学院文献情报中心, 硕士研究生。研究方向: 信息资源组织与建设。E-mail: wubb@mail.las.ac.cn 北京 100190
 夏翠娟 女, 上海图书馆(上海科学技术情报研究所), 高级工程师。研究方向: 数字图书馆开发。E-mail: cjxia@libnet.sh.cn 上海 200031

(收稿日期: 2015-02-03)