

书目资源的开放关联化实现研究

范炜 胡春晖

(四川大学公共管理学院信息管理技术系, 成都 610064)

摘要: 虽然关联数据在国内图书情报领域的认知已普及,但在书目资源上的应用推进仍较缓慢。本文引介联合国粮食及农业组织发布的《书目数据开放关联化指南》(LODE-BD),通过借鉴国外书目资源的关联数据实践经验,以期促进国内信息资源序化理论与实践的融合。本文重点辨析了基于RDF的三对基本概念,解读了M2B概念模型,分析了元数据模式的通用属性组设计,介绍了LODE-BD的决策树选择流程创新做法。最后在分析LODE-BD的基础上,提出书目资源开放关联化的5圈层结构,为国内书目资源的开放关联化提供系统化实施参考。

关键词: 书目资源; LODE-BD; 关联数据

中图分类号: G254.0 **DOI:** 10.3772/j.issn.1673-2286.2021.01.002

引文格式: 范炜,胡春晖.书目资源的开放关联化实现研究[J].数字图书馆论坛,2021(1):10-17.

2001年, Berners-Lee等^[1]在《科学美国人》上发表《语义网:一种对计算机有意义的网络内容新形式将引发一场新革命》,系统化阐述语义网理念与发展愿景,至今已有20年。语义网与关联数据为各行业数据资源的开放、互访、关联、集成提供了技术路线与规范。开放只是第一步,开放的数据需要进一步通过特定的“联系”,关联在一起,进而为检索、分析、发现提供现实基础。在中文语境中, LOD (Linked Open Data) 的理解与表述应为“关联化的开放数据”,不宜将两者随意并列和颠倒词序。在LOD的实现动作上,“关联化的开放数据”,需要先开放数据,再对开放数据进行关联化处理,这个过程就是“数据的开放关联化”。开放化与关联化是逐步递进和加成的,使得数据更具智能化分析潜质,即智慧数据概念所倡导的大数据深化方式^[2]。

以关联数据发布的各领域数据集逐年增多,规模效应逐渐体现。McCrae等开发并维护的关联数据云图(LOD Cloud)为关联数据发布情况提供了一个可视化概览^[3]。关联数据云图收录的数据集需要符合关联数据发布原则且有一定数量和格式要求。截至2020年12月,关联数据云图包含1 255个数据集和16 174个数据集链接^[4]。书目资源数据集在关联数据云图中占有一定比例,涉及国外各个图书馆书目及相关词表,是关联数

据技术赋能书目资源的现实应用,为书目资源打开了更加开放与潜在关联的新局面。不过,目前尚未出现中文书目资源的关联数据集连入关联数据云图。

在数据智能广泛流行的当下,数据是生产要素,数据是基础性资源^[5],各种应用服务都依赖于广泛而坚实的数据基础。长期以来,图书情报领域在书目资源建设方面积累了大量高度结构化的优质数据,关于书目资源描述的国家标准也已出台,作为推荐性标准使用^[6]。书目资源不仅支撑文献检索与知识发现,还为新兴的数字人文提供研究数据。国内对数据基础设施的这一提法^[7],应景且及时。书目资源要发挥更大的数据价值,需要在原有结构化基础上,借助关联数据的语义化特征,加强结构化的表征形式,增强开放化的关联集成。

随着数字化到数据化的进程深化,文献编目、网络元数据、语义网技术等实践应用越来越丰富。在国内图书情报领域,关联数据理念已得到广泛普及,相关技术与方法讨论众多,但国内书目资源的开放关联化仍未见到有影响力和规模化的实质推进,目前发布成关联数据的书目资源数量和规模相对有限。究其原因,很重要的一个现实鸿沟是:懂关联数据技术的工程师不懂书目工作理论,书目工作者对如何驾驭关联数据有技术障碍。“如何做”的问题依然是业务与技术结合的难点所

在。虽然信息资源描述工作标准化与语义网技术工具应用各自相对成熟,但是将二者结合起来,应用层面的具体步骤指引资料相对缺乏。相比之下,国外书目资源的关联数据项目则非常活跃,总结和积累了大量实践经验。国内要解决这一问题,在理解关联数据技术本质的基础上,需要跟踪和汲取国外实践经验,并结合本土化实际,设计适合中国国情的书目资源开放关联化方案。

LODE-BD (Linked Open Data Enabled Bibliographical Data) 是联合国粮食及农业组织 (FAO) 在多年关联数据项目实践基础上发布的一份面向书目数据拥有者或提供者的实践指南。它给出了书目资源发布成关联数据的指导原则与推荐做法,强调将书目数据发布成“有意义的 (meaningful) 且有用的 (useful)” 开放关联化书目数据^[8]。其中, Enabled 有“赋予…作用/功能…实现”的含义,一定程度上具有“赋能”意味。为便于中文理解,将 LODE-BD 意译为“书目数据的开放关联化”,赋能动作与实现结果蕴含在开放化为前提的关联化动作序列之中。

LODE-BD 从书目资源概念模型、元数据描述方案、取值编码等逐步递进,其本质上是一套经项目实践检验的书目资源开放关联化实现方案。LODE-BD 首个版本发布于 2011 年。2.0 版本发布于 2015 年,主要变化是增加了 LODE-BD 与 Schema.org 的元数据映射。2020 年,LODE-BD 发布了 3.0 版,基于联合国粮食及农业组织一站式信息系统 AGRIS 的示范项目,集成了美国农业部数据集元数据,增加了科学数据资源元数据,参考了万维网联盟 W3C 于 2020 年 2 月更新的数据目录词汇 (Data Catalog Vocabulary, DCAT) 第 2 版,同时根据都柏林核心元数据倡议 (Dublin Core Metadata Initiative, DCMI) 2020 年 DC 元数据的修订变化做出相应更新调整。其中, M2B 概念模型、元数据模式与决策树选择流程具有鲜明特点,值得深入研究与推广借鉴。本文通过对 LODE-BD 3.0 进行全貌式解读分析与观点性讨论,以期为国内书目资源开放关联化从理念认识走向实际应用提供参考。

1 LODE-BD 发展概述

自 2010 年起,联合国粮食及农业组织开展了虚拟开放存取农业与水产仓储 (Virtual Open Access Agriculture & Aquaculture Repository, VOA3R) 项目,项目成员包括 13 个国家/地区的 17 个机构,共建共享

书目数据,讨论元数据项和标准词汇集用于农业、食品及环境相关主题的科学数据交换与共享,目前已建成 8 个开放仓储库。

VOA3R 在元数据交换与采集上使用了两种元数据方案,一种是 DC 元素集,另一种是元数据对象描述模式 (Metadata Object Description Schema)。VOA3R 提供了《元数据编码指南》,为数据提供者如何编码书目数据给出实践参考。《元数据编码指南》包括来自标准化命名空间的属性,通过规范文档与受控词表作为关联中介,发布 RDF 三元组,提交 RDF 数据集到 VOA3R。

VOA3R 实际扮演了服务提供者角色,在传播渠道和可获取方面进行扩展,同时也是书目数据发布为 RDF 的提倡者,促进了农业与水产领域的关联数据应用。VOA3R 项目的长期实践孕育了 LODE-BD, LODE-BD 可看作 VOA3R 项目的成果分享与推广。虽然 LODE-BD 产生自 VOA3R 项目的农业与水产相关主题书目资源,但其本身不涉及农业和水产专指概念,而是适用于各种主题领域的书目资源描述与数据生成,可为各类文献 (期刊论文、专著、学位论文、会议论文、研究报告、学习对象等) 提供书目数据结构化的推荐做法。

在书目资源开放关联化目标下,LODE-BD 着力于如何复用广为接受的元数据元素集和取值词汇表,生成开放关联化的书目数据,从而促进数据共享与再利用,提升资源发现。LODE-BD 以具体问题为导向,首先提出 5 个问题: ① 什么样的元数据标准提供将书目资源发布成关联数据的描述基础; ② 要实现有意义的的数据分享,需要的最小属性集是什么; ③ 是否存在可直接利用的元数据模式和应用纲要来生成数据; ④ 已发布成关联数据的受控词表有哪些; ⑤ 本地数据库如何转换成关联数据集。然后,在这 5 个问题引导下,通过决策树方式提供书目资源描述所需常用属性的选择策略。每棵决策树给出各类行动点及其匹配的编码建议,渐进式回答以上 5 个问题。LODE-BD 允许数据提供者根据开发阶段、内部数据结构和现实情况,选择具体可行的做法。

下文将由内及外,从概念建模、元数据模式的通用属性组设计、决策树选择流程等对 LODE-BD 展开分析。

2 LODE-BD 的概念建模

2.1 三对基本概念辨析

目前,书目资源的开放关联化讨论一般是在关联数

据语境下开展的。无论是数据概念建模还是后续的元数据描述,都需要有明确一致的基本概念认识。LODE-BD在RDF图模型的三元组结构上明确了资源与实体、属性与元数据元素、字符串与URI的关系。以下结合语义网与关联数据,对三对基本概念进行辨析。

(1) 资源与实体(主)。在信息资源描述中,资源也就是事物(thing)的含义,是最抽象的概念化存在,能够代表任何形式的对象,始终是描述的主体,处在RDF三元组的主语位置上。用资源的抽象性引出对概念模型的根本性认识,这是使用“资源”一词的初衷。在概念模型中,抽象的资源被定义为实体。虽然实体也是抽象概念,但与资源相比,其是资源的具象化。在概念建模中,两者都是概念化的存在,资源比实体更抽象。基于此解释,书目资源内涵最广,代表着应用场景与领域。书目资源中包含不同的实体,这些实体自身有其特有的属性刻画,实体之间有着各种关系存在。LODE-BD将资源作为起始的分析对象,随着决策树流程而逐步具象化。

(2) 属性与元数据元素(谓)。语义网对元数据产生了非常重要的影响。DCMI在推进DC元数据发展进程中,XML与RDF作为数据交换的编码序列化格式对元数据产生了重要的影响,也是抽象到具体的有形表征。RDF是当前元数据讨论与使用的主要技术语境,元数据本身定位在RDF三元组的谓语位置上。资源的属性(property)对应具象化为元数据元素(element),属性与元数据元素是一对多的关系。也就是说,一个属性根据需要可以选择不同的元数据元素,通过元数据的命名空间(namespace)前缀进行区分。例如,版权属性可以使用dc:rights或dcterms:rights,以及其他元数据(如schema.org)的版权有关属性,选择灵活性较高。

在中文翻译中,英文property与attribute都可以被翻译为属性一词。但是,作为元数据专门术语,两者之间有差异,需要明确区分。根据Johnston对DC元数据元素限定的解释^[9],property是资源的描述,不仅包含attribute(特性),还包括aspect(方面)、characteristic(特征)以及relation(关系)等。元数据中的property定义突出两点:一是property内涵大于attribute;二是谈及property,包含对关系的认识,这也符合RDF宾语位置的定位,属性与关系的位置一致性。在中文元数据讨论中具体区分时,可将attribute译为特性,与property(属性)区分开。

(3) 字符串与URI(宾)。字符串(string)与URI

是取值的两种重要类型,对应的是RDF三元组的宾语位置。字符串是文本的一种表示,对应RDF的文字(literal)类型。用字符串表示取值,意味着链接的终点,不具有可关联性。URI是资源的唯一识别符,对应RDF的非文字(non-literal)类型,是实现HTTP可持续一致性访问的保障,也是形成链接的基础。书目资源开放关联化的操作核心就是实现尽可能多的URI表示,单纯字符串文本无法进一步实现自动链接。在关联语境中,取值弱化字符串的文本化表象,强化作为事物的实体化内涵。结合书目工作来说,编目中的主题标引一般取自受控词表(如叙词表、规范文档等),若受控词表本身已进化发布为关联数据,则直接赋予主题对应概念的URI即可。该概念URI在标引中被多次赋值,通过该概念URI自动汇集关联起与该概念相关的书目资源。举例来说,LODE-BD给出大米概念的URI赋值是http://aims.fao.org/aos/agrovoc/c_6599.html,就用到了已经开放关联化的联合国粮食及农业组织的AGROVOC叙词表。

以上三对概念的明确一致化是书目资源开放关联化基础的一般认识,也是概念建模的基础要素。

2.2 LODE-BD概念模型M2B解析

长期以来,编目规则、元数据标准等信息资源描述方案是数据结构化编码的直接参考依据。全面详尽的著录项、元数据元素、字段制定背后是有一定概念模型作为基础的。换句话说,信息资源描述方案的形成是通过自上而下方式,从概念模型的抽象设计开始,再逐步具象化展开形成的。信息资源描述需要解决这些问题:需要描述的对象是什么;需要描述对象的哪些方面(属性);需要揭示对象的哪些关系。要有效使用编目规则、元数据标准开展资源描述和结构化数据编码,应避免一开始就陷入众多属性与关系的细节纠缠中,因此对概念模型的认识尤为重要。

LODE-BD提出的概念模型被命名为“有意义的书目元数据”(Meaningful Bibliographic Metadata, M2B)。该概念模型基于IFLA的FRBR家族模型,并与最新的整合巩固模型LRM保持一致^[10]。M2B是对书目资源的高度抽象化,具备通用性与简约性特点。M2B仅定义了3个基本实体,即资源(resource)、行为者(agent)与希玛(thema)。在3个基本实体中,希玛是thema的中文直译,仅是方便其有一个中文称谓,没必要刻意翻译。

希玛有主题 (subject) 的含义, 但又不仅仅是主题。FRSAD标准制定者特意选取希腊文thema是将主题上升到与资源等同的抽象层面, 开拓出概念实体与其命名词语 (命名法) 的关系空间, 涵盖主题词表、本体、知识组织体系等^[11]。

M2B不对资源实体进行例化式的子实体划分, 而是将其作为不同类型的资源实体, 如期刊论文、会议报告等。关系定义主要分为两类。一类是不同实体间的关系。例如, 特定资源实体与行为者、希玛实体两两之间的关系。根据上下文, 结合中文理解, 这里的特定资源实体即上面提及的不同类型的资源实体。另一类则是同一实体内部之间的关系。例如, 资源与资源、行为者与行为者、希玛与希玛的关系。

M2B概念模型对规范控制的强化是其突出特点, 体现在取值部分。例如, 行为者的名称规范约束, 希玛的主题规范等。需要注意的是, M2B对规范控制的实施前提是受控词表的开放关联化。开放关联化词表 (Linked Open Vocabularies, LOV) 平台汇集了已发布成关联数据的受控词表、规范文档、本体以及元数据等, 是M2B规范控制实施的取值参考来源。截至2020年12月, LOV平台已收录738个受控词表相关词汇集^[12]。

在LODE-BD版本修订中, M2B概念模型的基本实体定义没有发生大的变化, 而在关系定义的细化与扩展上有所变化。概念模型新版本在关系定义上进一步抽象, 增加了同一实体内的关系定义。这里对LODE-BD3.0中M2B概念模型进行中文对照展示 (见图1)。

中间竖排的3个实体是概念模型的主干, 以资源实体为中心, 资源与希玛、资源与行为者两两实体之间存在互反的双向关系。资源与希玛是主题标引关系, 资源与行为者是负责与责任者关系。每个实体右上角都有自连接指向箭头, 表示同一实体的相关关系。值得注意的是, 每个实体左侧都给出关联云图样式, 意味着各个实体都有其对应的关联数据云图存在。资源实体左侧未标名称, 其表示书目资源关联数据云图, 即LODE-BD最终实施的结果与成效。希玛左侧标注的是取值词汇, 其表示受控词表关联数据云图, 也就是之前提到的LOV。行为者左侧标注的是名称规范, 其表示与个人、团体等行为者相关的名称规范关联数据云图。

3 LODE-BD元数据模式的通用属性组设计

经过充分研究已有开放仓储库实践, LODE-BD精

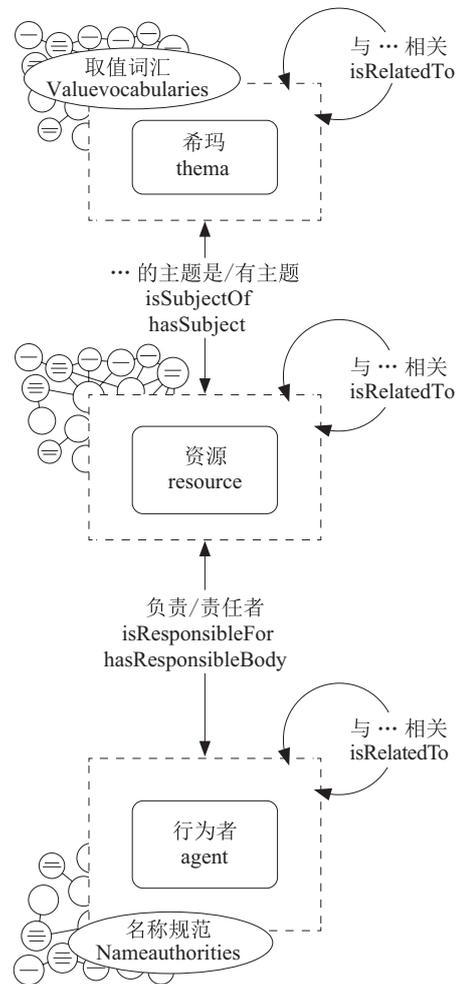


图1 LODE-BD M2B概念模型图示

心设计了9个属性组及其对应的若干通用属性。关系单独作为第9组, 处理书目资源实体之间和行为者之间的关系。

在属性设计成型后, 对属性所对应的元数据复用情况进行明确, 主要复用了7个元数据标准。具体元数据标准包括DC元数据元素集 (dc)、DCMI元数据术语 (dcterms)、书目本体 (bibo)、澳大利亚政府定位服务元数据 (agls)、预印本术语 (eprint)、溯源本体 (prov) 及数据目录词汇 (dcat)。以上括号里是各自的命名空间前缀。

LODE-BD切实贯彻了元数据复用思想, 为开放关联化的自动链接打下基础。表1对LODE-BD 9组属性、M2B属性推荐和元数据复用进行了翻译与概要介绍。

表1中, 属性一栏“++”号表示必备属性, “+”号表示强烈推荐属性。需求一栏分为两种情况, 一种是分析性描述、另一种是非分析性描述。对应描述的粒度需求程度分为4种: M表示必备, HR表示强烈推荐, R表示

推荐, O表示可选。取值控制一栏包含以下几种情况: 无是指不受限制、自由取值; 名称规范是指包括人名、团体与会议等; 语法编码规则是指标准化或特定的数据格式, 如日期格式的ISO 8601; 受控列表是指有限的取

值选项; 受控词表是指从各类知识组织体系中取值, 分类法是受控词表的一种类型; 受控资源ID专门用于资源之间的关系。通过表1可以了解LODE-BD从属性组到具体属性, 再到具体元数据的复用选择。

表1 LODE-BD属性概览

属性组	M2B属性推荐				元数据复用		
	属性	需求		取值控制	通用		专用
		分析性	非分析性				
1.题名信息	题名++	M	M	无	复用dc	复用dcterms	复用dcterms、bibo、dcat、agls、eprint等
	替代题名	O	O	无			
2.责任者	创建者+	HR	HR	无或名称规范			
	贡献者	O	O	无或名称规范			
	发布者+	R	HR	无或名称规范			
	[其他责任者]	O	O	无或名称规范			
3.外部特征	日期++	M	M	语法编码规则			
	识别符+	HR	HR	语法编码规则			
	语种++	M	M	受控列表			
	格式/媒介+	HR	HR	受控列表			
	版本	R	R	无			
	来源+	R	HR	无			
	[其他外部特征]	O	O	无			
4.馆藏/位置信息	位置++	M	M	无或规则			
	其他位置	O	O	无			
5.主题信息	主题词+	HR	HR	受控词表			
	分类	O	O	受控词表/分类法			
	自由关键词	R	R	无			
	地理名称	O	O	受控词表			
6.内容描述	描述/摘要	HR	HR	无			
	类型/形式	R	R	受控词表			
7.知识产权	版权+ 使用条款	R	R	无			
8.用法	适用对象	O	O	受控列表			
	指导方法	O	O	无或受控列表			
	文学指征	O	O	受控列表			
	教育水平	O	O	受控列表			
9.关系	[资源之间的关系]	HR	O	受控资源ID			
	[行为者之间的关系]	O	O	无或名称规范			

4 LODE-BD决策树选择流程

LODE-BD创新性地采用决策树方式, 以作用点(acting point)作为选择性分支引导, 从抽象到具象逐步细化, 最终得到适合的属性与取值推荐。决策树从资

源实体出发, 根据不同选择分支, 得到具体元数据元素及其取值类型的步骤, 符合认知与选择的渐进流程。这一方式是可实施的编码策略风格的体现, 在当下众多标准规范、技术推荐文档之中, 开辟出一条高效可行的书目资源开放关联化的实现路径。

LODE-BD对9组属性分别进行了详细介绍,每个属性按照决策树流程图与决策过程表相结合的方式展开,决策树流程图侧重直观过程,决策过程表更具形式化,两者应结合使用。

决策树流程图一般以资源(圆角矩形)为起点,通过一步步条件判断(菱形),最终给出可供选择的元数据元素及其对应的取值推荐。此处选取创建者的属性选择进行示例说明,以便直观了解决策树做法(见图2)。对创建者信息的描述中,分为是否使用规范文档两种情况。在不使用的规范文档(具体指人名规范)情况下,建议使用dc:creator,取值为字符串(1a)。在使用规范文档情况下,若规范文档没有无可用的关联数据,建议使用dc:creator,取值为字符串(2a);若规范文档有可用的关联数据,建议使用dcterms:creator,取值推荐使用URI(实心圆圈)或字符串(虚线圆圈)(2b)。决策过程表是决策树流程图的表格化再现,由步骤(序号数字)、条件、判断、行动、取值类型、元数据元素、取值示例7列组成,是对决策树流程图的步骤说明,并提供元数据元素不同取值类型(规范化、非规范化、字符串、URI)对应的取值示例。

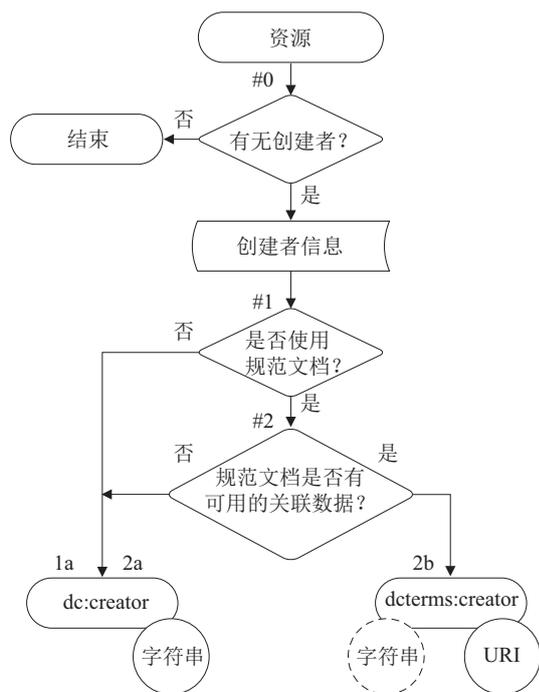


图2 创建者的决策树流程图

决策树是LODE-BD的创新所在, M2B概念模型、通用属性组和元数据复用三者都是实现决策树的要件。有关决策树的内容占LODE-BD一半篇幅以上,极

尽详细地给出所有属性落实到元数据元素和对应取值的做法,适合作为案头常备查阅之用。

5 总结与启示

关联数据已有多年发展,以LODE-BD为代表的国外书目资源关联数据实践已经走出了一条发展道路,提供了切实可行的参考指南。LODE-BD以LOD打底,将概念模型M2B、RDF图模型和元数据模式复用等,通过决策树选择流程有机集成在一起。决策树做法具有创新性,值得肯定与大力推广。

由于篇幅所限,本文对LODE-BD仅做了概要性、原理性分析解读,属性选择与取值的决策树详细内容参见指南全文,有关元数据标准、受控词表关联数据化、序列化编码、关联数据利用以及与schema.org的映射转换内容并未专门展开讨论。

通过LODE-BD的分析可以看出,目前关联数据聚集效应的形成主要是遵照开放世界假设,借助RDF三元组结构上的开放互访性,“主-谓-宾”三者尽可能URI化,实现自动链接识别与联系建立的HTTP访问技术基础,其技术路线成熟,实现难度不高。书目资源开放关联化的具体做法是,在RDF中尽量将谓语位置的属性(property)和宾语位置的文字(literal)两处,充分利用资源描述规范和各类词表资源,通过已有命名空间的元数据标准/应用纲要的URI及受控词汇的URI表示,在资源描述与数据结构化编码阶段建立起链接的自动识别与关联基础。进一步通过映射变换(crosswalk)、对齐(alignment)及互操作(interoperability)等语义化处理手段,增强关联的黏度与广度。

书目资源拥有良好的数据化实施基础,但也因诸多因素未能见到成熟稳定的中文书目关联数据项目和数据集发布。除开放的政策与利益有关问题外,仅就数据而言,主要还是缺乏对书目资源开放关联化的系统思考和渐进式落地。LODE-BD所带来的启示是,书目工作中的编目规则、概念模型、元数据、数据格式、数据库等都是走向开放关联化的重要组成部分,要有系统化的集成认识,将这些方面有机联系在一起,共同实现书目资源的开放关联化。

在LODE-BD的启发下,进一步系统化思考书目资源的开放关联化,可以构建出由内及外的5层结构,分别由概念模型、逻辑模型、元数据模式、序列化编码、开放关联组成,如图3所示。

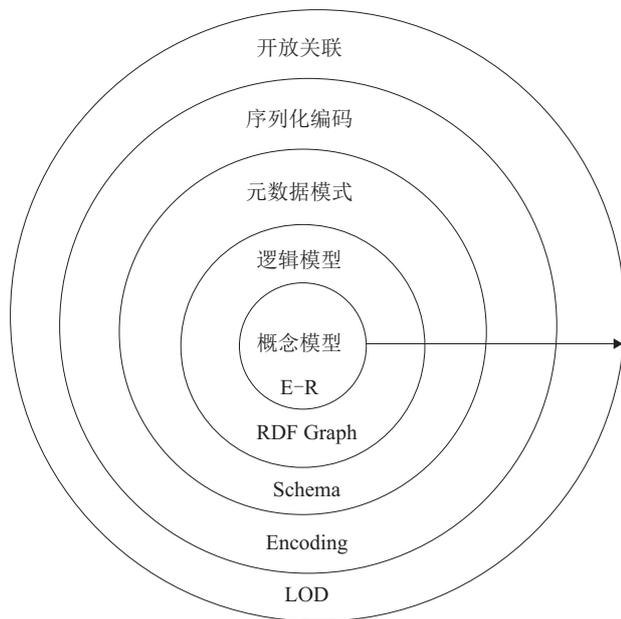


图3 书目资源开放关联化的5圈层渐进图

(1) 概念模型。书目资源的概念模型既是资源描述的认识起点，也是自上而下地形成概览式框架。概念模型位于书目资源开放关联化圈层的核心位置，体现出基础重要性。一个稳固且开放的概念模型从根本上决定了后续实现的走向。

概念模型在书目工作中的重要性显现，是从编目规则细节制定到数据建模的关注，特别是对实体-关系（Entity-Relationship, E-R）方法论在书目资源描述上的运用。书目资源的概念模型是运用E-R方法论，围绕书目资源，定义相关的实体、属性与关系。书目资源的概念模型不止一种，除FRBR之外，美国国会图书馆提出了BIBFRAME，还有来自开放网络社区的各种书目模型，如BIBO本体、schema.org的书目描述内容等。

概念模型作为顶层设计，独立于计算机实现环境，是一种认知分析成果，可作为标准、推荐进行推广。但是，要发挥概念模型的作用，需要在应用场景中考虑计算机处理环境，往往与逻辑模型直接绑定进行分析。逻辑模型作为概念模型进入计算机处理域的直接中介，在现实应用中常常被当作概念模型的替代或两者一体化论述。

(2) 逻辑模型。逻辑模型是在数据处理技术语境中，对概念模型的具象化翻译、映射或转换。参照数据库设计的三级模式，关系型数据库的逻辑模式是表定义。以关联数据集为目标，书目资源数据化的逻辑模型对应的则是RDF图（Graph）^[13]。RDF图较之属性图、

UML类图、OWL等，其三元组结构简单且开放关联性易于建立，是关联数据实现的主流方式。值得注意的是，目前逻辑模型和RDF的有关认识与表述存在模糊和不一致，需要进一步明确认识。

(3) 元数据模式。在概念模型与其对应的逻辑模型基础上，元数据模式（Schema）是对如何描述书目资源进行具体设计与约定。通用元数据标准、各种行业元数据标准以及网络社区元数据方案（事实性标准）为有效的规范化描述提供丰富的参考资源。充分复用已有元数据标准来定义属性与取值是发挥元数据作用与价值的主要手段。元数据模式的形成是书目资源数据化的实现核心，起到承上启下的作用。

(4) 序列化编码。书目资源数据化的成果要以数据记录、数据集和数据库等形式存在。数据编码是数据化实现阶段的关键一步。序列化（serialization）编码是对数据存储、交换与共享有相应要求，以元数据模式为编码具体依据，具体格式对应逻辑模型RDF图的各种数据格式，如RDF/XML、RDFa、Turtle、N-Triples以及最新推荐标准JSON-LD等^[14]。序列化编码得到既定设计要求的结构化与语义化数据，为下一步开放关联做好准备。

(5) 开放关联。开放关联是数据化最外围的一层，其主要作用是将序列化编码后的数据通过关联数据发布机制开放到网络上，进入开放关联数据环境中，与其他开放的数据集建立起关联，促进关联数据云图的形成与壮大。开放关联主要涉及三个方面：一是技术，W3C已推出关联数据平台推荐标准^[15]，因此技术实现上目前没有障碍；二是已有命名空间元数据元素与受控词汇的复用，它形成关联的基础，是由第三圈层元数据模式决定；三是数据开放政策与许可方式，由数据拥有者和提供者决定，这是开放关联最难以控制的人为方面。这三方面的共同作用，才能够达成开放关联目的。

书目资源开放关联化5圈层是一种图示化的理论抽象，有助于理解书目资源开放关联化的内在逻辑层次。5圈层是方法论意义的顶层认识，LODE-BD则在行动层面提供了实践指导，希望本文能为国内中文书目资源发布成关联数据提供行动参考，期待未来有更多中文书目资源的示范性关联数据集发布。

参考文献

[1] BERNERS-LEE T, HENDLER J, LASSILA O. The semantic

- web [J]. Scientific American, 2001, 284 (5) : 34-43.
- [2] 曾蕾, 王晓光, 范炜. 图档博领域的智慧数据及其在数字人文研究中的角色 [J]. 中国图书馆学报, 2018, 44 (1) : 17-34.
- [3] The Linked Open Data Cloud Maintainers [EB/OL]. [2020-12-23]. <https://lod-cloud.net/#about>.
- [4] The Linked Open Data Cloud Diagram [EB/OL]. [2020-12-23]. <https://lod-cloud.net/#diagram>.
- [5] 中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见 [EB/OL]. [2020-12-23]. http://www.gov.cn/zhengce/2020-04/09/content_5500622.htm.
- [6] 宋文, 朱学军. 《资源描述》国家标准及对我国信息资源描述标准体系的思考 [J]. 数字图书馆论坛, 2016 (12) : 21-27.
- [7] 夏翠娟. 面向人文研究的“数据基础设施”建设——试论图书馆学对数字人文的方法论贡献 [J]. 中国图书馆学报, 2020, 46 (3) : 24-37.
- [8] SUBIRATS I, ZENG M L. Linked Open Data Enabled Bibliographical Data (LODE-BD) 3.0-A practical guide on how to select appropriate encoding strategies for producing Linked Open Data Enabled Bibliographical Data. Rome, FAO [EB/OL]. [2020-12-23]. <https://doi.org/10.4060/cb2209en>.
- [9] JOHNSTON P. Element Refinement in Dublin Core™ Metadata [EB/OL]. [2020-12-23]. <http://www.dublincore.org/specifications/dublin-core/dc-elem-refine>.
- [10] IFLA图书馆参考模型 [EB/OL]. [2020-12-19]. https://www.ifla.org/files/assets/cataloguing/frbr-irm/ifla-irm-august-2017_rev201712-ch.pdf.
- [11] Functional Requirements for Subject Authority Data, A Conceptual Model (FRSAD) [EB/OL]. [2020-12-19]. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.
- [12] Linked Open Vocabularies [EB/OL]. [2020-12-19]. <https://lov.linkeddata.es/dataset/lov>.
- [13] RDF 1.1: On Semantics of RDF Datasets [EB/OL]. [2020-12-19]. <https://www.w3.org/TR/2014/NOTE-rdf11-datasets-20140225/>.
- [14] RDF Syntax [EB/OL]. [2020-12-19]. <https://www.w3.org/wiki/RdfSyntax>.
- [15] Linked Data Platform 1.0 [EB/OL]. [2020-12-19]. <https://www.w3.org/TR/ldp/>.

作者简介

范炜, 男, 1981年生, 博士, 硕士生导师, 副教授, 通信作者, 研究方向: 信息存储与检索, E-mail: fanw@scu.edu.cn。
胡春晖, 男, 1997年生, 硕士研究生, 研究方向: 信息检索与情报服务。

Publishing the Bibliographic Resources as Linked Open Data

FAN Wei HU ChunHui

(Department of Information Management Technology, School of Public Administration, Sichuan University, Chengdu 610064, China)

Abstract: While the notion of linked data is popular in China, the application development of linked data for Chinese bibliographic resources is lacking. This article introduces the Linked Open Data Enabled Bibliographical Data (LODE-BD) which proposed by FAO. It's helpful to introduce the practical experience of linked data for foreign bibliographic resource which promote the integration of information resource ordering theory and practice. It discusses the three pairs of basic concepts based on RDF, interprets the M2B conceptual model, analyzes the design of general attribute group of metadata schema, and introduces the innovative approach of decision tree selection process of LODE-BD. Finally, it proposes a 5-circle framework with a systematic thinking, in order to provide the linked open data implementation strategy for the Chinese bibliographic resources with LODE-BD.

Keywords: Bibliographic Resources; LODE-BD; Linked Data

(收稿日期: 2020-12-24)