

国家图书馆关联数据注册与服务系统设计及实现

蔡颖

(国家图书馆, 北京 100081)

摘要: 关联数据技术的发展与应用是图书馆遇到的新的机遇与挑战, 国家图书馆借此契机建设了关联数据注册与服务系统。该系统具备对RDF数据的通用处理能力, 根据数据注册、审核、发布流程配备完整的状态管理和转换机制, 可以实现涵盖关联数据整个生命周期的管理。系统设计遵守关联数据四原则, 开发了包括数据注册、数据发布、数据服务、数据展示和数据管理等功能模块。为发布数据提供通用展示和个性化展示界面, 目前系统已注册并发布3个数据集, 共计258 469个资源, 约5 049 162条三元组, 并向用户提供资源检索、数据下载、API接口等数据开放共享服务。

关键词: 关联数据; 数据服务; 中国分类主题词表; RDF

中图分类号: G250.76

DOI: 10.3772/j.issn.1673-2286.2019.11.005

2006年万维网的发明者Berners-Lee^[1]首次提出关联数据(Linked Data)的概念。关联数据可以视作一种轻量级的语义网实现方式。它采用资源描述框架(Resource Description Framework, RDF)三元组数据模型, 使用统一资源标识符(Uniform Resource Identifier, URI)标识资源实体, 发布类数据及实例数据, 并且通过HTTP协议来揭示、获取这些数据, 关联数据主要强调数据之间的相互关联关系, 以及语义信息可以被机器理解^[2]。关联数据的提出可以有效解决自主、异类、异构的海量数据的知识发现问题^[3]。关联数据的提出受到广泛关注和重视, 各界纷纷加入关联数据的研究与实践中, 截至2019年3月, 关联开放数据(Linking Open Data, LOD)项目已经发布包括政府数据、生命科学、地理数据等各个领域共1 239个数据集^[4]。

1 图书馆界关联数据应用情况

关联数据可以将图书馆的海量资源与外界资源进行有效地连接, 从而解决图书馆资源“数据孤岛”的现状, 进一步优化和开拓图书馆的资源发现系统, 提升

图书馆资源发现的服务水平。万维网联盟(World Wide Web Consortium, W3C)报告认为, 在关联数据运动中图书馆可以担任领导者的角色。2010年, W3C正式成立图书馆关联数据孵化小组, 为图书馆建立、应用关联数据提供帮助^[5]。对此, 图书馆界也给予充分的重视, 主要将关联数据应用于图书馆的各种规范词表、图书馆记录集、图书馆的元数据元素集等多个方面。

1.1 国外图书馆的探索与实践

国外图书馆关联数据建设大多先从规范词表数据开始, 然后推广至书目数据。规范词表关联数据的典型应用包括美国国会图书馆选择简单知识组织系统(Simple Knowledge Organization System, SKOS)将美国国会图书馆主题词表(Library of Congress Subject Headings, LCSH)发布成为关联数据; 联机计算机图书馆中心(Online Computer Library Center, OCLC)发布由美国、法国、德国3国图书馆的联合项目虚拟国际规范文档(Virtual International Authority File, VIAF), VIAF数据是目前LOD项目中最大的规

范名称数据集^[6]；德国国家图书馆早在2010年就发布了330万条名称规范记录和近19万条主题规范记录^[7]。书目数据的典型应用是瑞典国家图书馆发布的瑞典联合目录（Library Information System for Sweden, LIBRIS），瑞典是第一个将书目数据发布为关联数据的国家^[8]。在构建关联数据本体时，各国图书馆尽可能复用已出现的本体词汇，规范数据选择的本体词汇一般为SKOS、FOAF、OWL等，书目数据选择的本体词汇更多倾向于DC、DCT、BIBO、RDFS。如英国国家图书馆尽可能地利用现有的本体词汇；德国国家图书馆则逐步按照数据模型开发的需要增添新的本体词汇来源^[9]。而编目格式从传统的MARC21格式转向基于关联数据模型的新格式BIBFRAME，美国国会图书馆在2011年发表书目数据转换声明，2013年正式发布创作作品、实例、规范和注释4个核心类组成的BIBFRAME模型，2015年10月美国国会图书馆发布BIBFRAME2.0并对核心类作出了调整^[10]。

1.2 国内图书馆的探索与实践

国内图书馆开展相关的实践较晚，近年来国内图书馆界逐渐意识到在关联数据方面与国外图书馆界的差距，以关联数据的理论研究为开端，配合以实验性质的项目，越来越多地将关联数据融入图书馆实际的工作与应用。中国科学技术信息研究所对发布工具D2R进行了研究并开展实践，实现书目数据的关联数据发布。宁波市数字图书馆利用PHP开源语音编程开发相关的关联数据发布组件，实现企业名录和专业人才关联数据的发布。卜书庆^[11]进行了基于SKOS的《中国分类主题词表》（以下简称《中分表》）语义化描述和可视化研究。上海图书馆对社会大众开放了数字人文项目的数据平台，以关联数据的方式在互联网公开发布项目所用的基础知识库、文献知识库和本体词表，包括人名规范库、华人姓氏列表、地理名词表、机构名录、中国历史纪年表等，并以多种方式提供数据消费接口，供开发人员调用，以促进数据的开放获取、共享和重用^[12]。关联数据在国外各大图书馆中有很多典型的应用，虽然国内图书馆界也陆续开展了一些实验课题和应用项目，但图书馆中最具价值的资源如《中分表》、书目数据还没有正式发布成为关联数据，与国外图书馆界关联数据的应用存在一定的差距。

2 国家图书馆关联数据平台设计方案

2.1 设计思路

国家图书馆关联数据注册与服务系统的建设总目标是建立涵盖关联数据整个生命周期管理能力和对RDF数据具备通用处理能力的系统。图书馆拥有海量词表规范、书目数据等有价值的数据，在发布关联数据时通常选择这些已有的数据集并通过D2R等开源工具转换成关联数据结构，系统支持其他工具产生的关联数据的批量导入。在设计时考虑到关联数据生命周期的完整性及未来关联数据应用的业务场景，也开发了相应的注册模块来满足关联数据直接在系统中生产的要求。

在建设系统之前，国家图书馆在关联数据方面具备一定的理论研究基础和实践成果，其中包括《中分表》、国图公开课的关联数据描述规范与其MARC转换方案，故将《中分表》和国图公开课作为两个关联数据集在系统发布。除此以外，选取2 500条馆藏文献，将其书目数据转换成关联数据格式，作为本平台对书目数据的探索实践。系统能够进行关联数据的注册、管理、发布与可视化展示，已经实现叙词表、视频、书目3种不同类型数据集的发布与服务。

2.1.1 设计原则

系统设计的一个重要原则是通用性，即建成关联数据统一的、通用的注册与服务平台。这要求系统具有良好的通用性和可扩展性，对RDF/XML、JSON、N-Triples和Turtle等不同格式的文件，以及不同本体结构的RDF数据均具备存储和处理能力。系统的建设还可有效避免每发布一套数据集就建设一个平台的现象，防止资源浪费和重复建设。

系统设计遵循开放性、成熟性、标准性、可靠性和安全性原则。具有标准接口、API或者通信协议，体系结构、资源建设与服务设计坚持开放性原则。系统建设过程中力求采用先进可靠的软件技术和长期经受市场考验的成熟工具或产品，确保系统的高性能和稳定性。系统设计还要充分考虑各种安全风险来保证平台的安全运行以及数据的安全性。

除以上建设常规系统时应考虑的设计原则，作为关联数据注册与服务系统，在系统设计时还必须遵守

关联数据的四原则^[1]：使用URI作为一切事物的标识名称；使用HTTP URI让一切事物的名称可以查看；使用唯一的RDF资源模型在网络上发布结构化数据；使用超链接以连接任何类型的事物。

2.1.2 总体架构

系统总体架构如图1所示，由数据层、业务层和展示层组成。数据层以Virtuoso、MySQL、Redis数据库

为支撑，开源数据库Virtuoso支持对RDF三元组形式的数据存储，MySQL用来存储系统中的用户数据、日志文件等，Redis数据库的最大特点是将数据存放在内存中，读写速度优于MySQL，因此将系统中如数据统计业务场景数据的热数据存放在Redis中。业务层提供数据注册、数据服务、数据管理、用户管理和统计报表等功能。展示层通过模块引擎渲染和Ajax交互实现了对数据进行可视化展示、接口服务等。

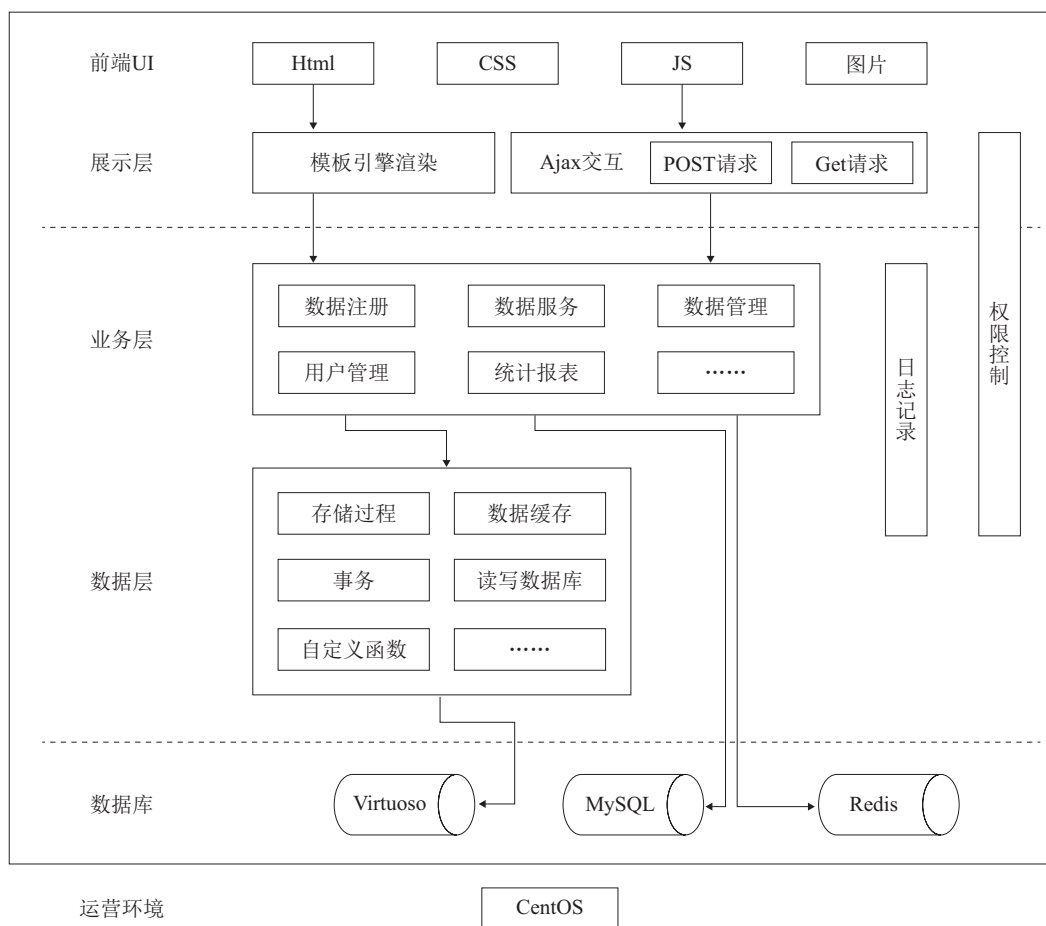


图1 国家图书馆关联数据注册与服务系统架构

2.1.3 URI设计

URI是网络资源的身份标识，URI具有唯一性和必须性，有了URI的资源就像是网络认证的“合法公民”，能够对其进行管理、存储、跟踪、调用等生命周期的管理^[13]。

系统中URI设计遵循国际上通用的COOL URI设计原则以及统一模式，国家图书馆关联数据的URI

按照规范词表、本体、取值词表、信息资源、非信息资源和数据集分为6个类别。国家图书馆的URI的统一模式为“http://{domain}/{type}/{concept}/{IRI}”。通常{domain}由服务器地址和政府类型组成，国家图书馆的{domain}结合中国国家图书馆现有域名，采用“data.nlc.cn”来表示；{type}表示被标识对象的类型；{concept}表示具体对象集合；{IRI}是系统内部唯一标识符。国家图书馆URI具体模式见表1。

表1 国家图书馆关联数据URI模式

标识对象类型词	说 明	URI
thesaurus	规范词表, 如名称规范档、中图分类法等	http://data.nlc.cn/thesaurus/
ontology	本体	http://data.nlc.cn/ontology/
vocab	取值词表, 如语言、国家、性别等	http://data.nlc.cn/vocab/
entity	非信息资源, 如人物、地点、机构等	http://data.nlc.cn/entity/
resource	信息资源, 如文档、图像、视频、音频等数字化的资源对象	http://data.nlc.cn/resource/
dataset	数据集资源	http://data.nlc.cn/dataset

2.2 功能需求与实现

2.2.1 数据注册

系统将SKOS、DC等W3C组织体系定义的词汇规范作为平台的词汇规范基础, 包含2018年W3C发布的所有的词汇规范和词汇属性, 既支持在此基础上词汇属性的扩展, 也支持注册新的或自定义的词汇规范。

形成完整、不断更新扩展的词汇规范之后, 在系统中可以通过选择不同词汇规范中的属性词汇, 来构建不同的知识本体。知识本体是领域知识被抽象后所形成的可以共享重用的概念模型^[14], 也是对概念体系可共享的规范说明, 每一个本体对应系统中一类关联数据或一个关联数据集。在创建和选定对应的本体后, 就可以进行数据注册。系统支持单条注册和批量灌装两种数据注册方式, 批量灌装默认支持RDF/XML、JSON、N-Triples和Turtle 4种序列化格式的文件导入

和存储。单条注册在系统后台根据选择的本体模型创建关联数据的三元组, 通过页面来填写三元组的主语、谓语和宾语, 从而直接在系统中产生关联数据。由于本次注册的《中分表》、国图公开课和馆藏文献3个数据集数据量较大, 都是采用在系统外部通过开源工具转换成关联数据格式后, 再通过批量灌装的方式在系统中进行注册的。

RDF三元组由主语、谓语、宾语共同构成, 当宾语为普通文本时, 该三元组被称为文本三元组; 当宾语为URI标识时, 则该三元组被称作RDF链接^[15]。当RDF链接中的宾语通过URI标识不断与其他数据集建立跨域链接, 不同数据集中丰富的实体关系被有效地联通, 最终将分散的数据聚合起来, 形成相互关联的数据网络。本系统在数据注册时, 宾语可以填写普通文本, 也可以在填写文本后对文本在系统内搜索匹配的资源实体或概念, 选择相应的实体或概念后, 可以添加其URI标识作为宾语, 从而形成RDF链接, 如图2所示。

图2 国家图书馆关联数据注册与服务系统中建立RDF链接

2.2.2 数据发布

在Web上发布关联数据即通过支持各种协商机制

的Web服务器在网上发布成关联数据应用, 通过API、REST、SOAP或其他Web服务方式, 支持HTTP查询和关联数据消费^[16]。系统实现发布功能时根据要发布关

联数据的数据量大小、数据更新频率、数据的存储方式和访问方式来选择不同的发布方式。在系统中对于数据量较小的情况,采用静态发布的方式,即发布静态的RDF文件;对于数据量大、更新频率高的情况,将数据批量存储在Virtuoso数据库中,使用Virtuoso自带的SPARQL Endpoint(端点)进行发布,SPARQL Endpoint也是关联数据的一种常见发布方式^[17]。

目前系统已经发布《中分表》、国图公开课、馆藏文献3个数据集,为方便这3个数据集的共享与复用,系统将这3个数据集的本体也在Web上进行了发布。

2.2.3 数据服务

关联数据发布后会提供开放数据服务、RDF DUMP服务和SPARQL查询服务。

国家图书馆关联数据注册与服务系统也提供以上3种服务方式,首先开放数据服务,提供关联数据的检索和浏览功能。在系统中提供通用检索、精确检索和组合检索3种检索方式。浏览服务主要提供数据集、词表的元数据及内容的浏览功能。针对整个数据集或者词表的浏览,用户可以选择某一个条件作为检索点,检索后可以定位检索词在数据集或词表中的位置,并可以上下浏览整个有序列表。

RDF DUMP服务是一种把数据以多种RDF文件格式发布的方式。在系统中,数据以RDF文件格式发布后,基于用户权限,可提供数据集或者词表的整体下载、元数据下载、内容下载及单个概念属性的下载。下载方式默认为RDF/XML序列化格式,并支持其他序列

化格式的下载功能（如NTriple、N3、Turtle）。

系统还提供标准的对外服务API接口及接口说明，依据关联数据开放和共享的原则，鼓励其他机构、平台遵循该接口规范，获取本系统的关联数据或直接复用本系统的资源实体或概念URI。接口服务既支持通过HTTP协议访问的方式获取数据集或词表，又支持使用RDF的标准化检索语言SPARQL访问，支持机器对数据集和词表内容的检索。在web界面上，用户可以通过SPARQL查询终端自行构建SPARQL查询语句以查询数据集或词表中的特定内容，并且可以保存构建的查询语句。

2.2.4 数据展示

系统具备通用展示和个性化展示2种关联数据展示方式。通用展示方式有2种,分别是基于列表的三元组属性展示以及知识图谱的展示。基于列表的展示方式,以“属性-属性-值”的形式展示每一条关联数据资源的三元组,这种可视化方式充分地展现了关联数据RDF链接的特性,使用户可以沿着有向边浏览整个数据集^[18]。此外,系统基于关联数据中的语义关系用Vue.js编程的方式实现关联数据知识图谱的可视化展示,如图3表示《中分表》中的2个资源,图中不仅可以看到每个资源的属性信息,还可以直观地看到资源之间的关系,从展示的关联关系中发现数据间隐含的知识。系统可以通过这两种通用的可视化展示方式对系统中注册的新数据集进行数据展示。

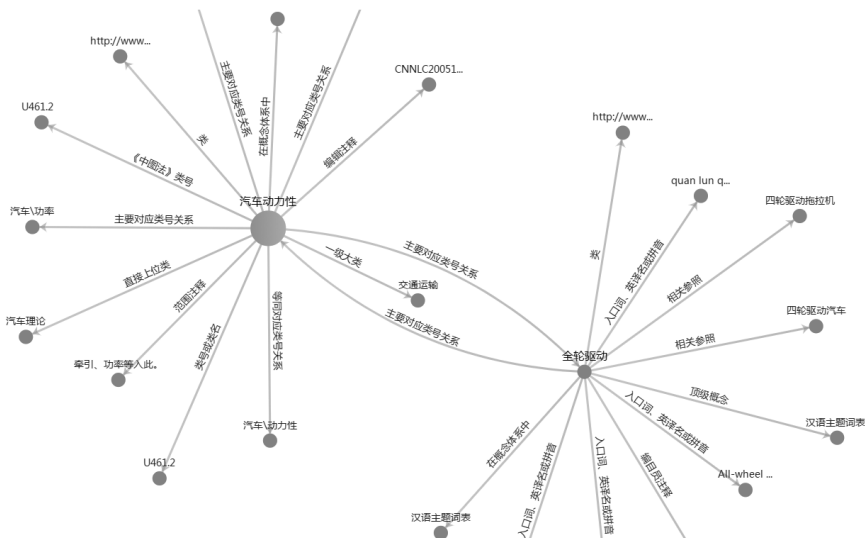


图3 国家图书馆关联数据注册与服务系统资源可视化示例

此外,对《中分表》、国图公开课、馆藏文献3个数据集,根据其各自的特点开发了个性化展示页面。对于中分表的数据,考虑其叙词表的特点,通过分类法的树状结构进行展示,基于关联数据间的语义关系还实现了主题分类和主题词之间的联动。国图公开课数据集个性化展示页面,呈现了课程的图片、主讲人的头像图片,并提供视频资源的链接。馆藏文献数据集根据其特点,展示书封并提供阅读地址的链接。个性化展示页面中的任何资源均可一键跳转到关联数据的可视化展示页面。

2.2.5 数据管理

系统针对不同数据的特点使用不同的数据库存储。对于用户数据、系统日志等数据在系统中使用关系型数据库MySQL进行存储。对于RDF数据使用Open Link Virtuoso数据库存储,Virtuoso数据库是一种开源的TripleStore数据库,它与传统的关系型数据库以及NoSQL数据库相比,不仅可以实现RDF在关系型数据库中的存储,还支持SPARQL语法查询^[19]。因此,选用Virtuoso数据库对RDF数据进行存储。

系统提供数据查询界面,可查询系统中的数据集、词表、元数据和数据内容。查询方式分为一般查询、高级查询和SPARQL查询3种。查询结果可以根据数据状态和数据提交者进行筛选。提供按用户角色进行查询的功能,使数据提交者、数据审核者以及系统管理者等用户能够快速定位到自己需要处理的数据。

系统中的数据在其生命周期的不同阶段可分为未完成、候选、合格、失效、被替代5种状态,根据其发布状态又可分为未发布、已发布和屏蔽状态。系统具备对数据的各种状态进行管理的功能,提供操作界面,使具有相应权限的用户可以对数据的状态进行修改,也可以对数据从一种状态转变为另一种状态的进程进行跟踪。

对于系统中的词表类数据,提供版本控制的功能,用来追踪和记录注册词表的修改、调整情况,以便数据管理者能够完整地追溯词表的变更历史,利于词表的更新维护。

2.3 系统建设成果

系统为多样化的数据开发了通用的关联数据注

册、发布及服务功能,覆盖关联数据生命周期全流程的管理。系统内根据数据注册、审核、发布流程配备了完整的状态管理和转换机制。从数据可视化的角度,除通用展示以外还为不同数据集设计不同的个性化展示界面。

目前系统已注册并发布3个数据集,每个数据集根据其数据情况和本体结构不同又包含多个子集。中分表数据集包括类目、主题词、概念体系和概念集合4个子集。国图公开课分为课程、分集、主讲人、组织机构、资料信息、知识链接和知识库7个子集。馆藏文献分为馆藏信息、实体文献、电子资源3个子集。目前系统共发布了14个数据子集,258 469个资源,以及5 049 162条三元组,详情见表2。

表2 国家图书馆关联数据系统数据发布统计

	子集/个	资源/个	三元组/条
中分表	4	237 758	4 702 428
国图公开课	7	16 182	218 754
馆藏文献	3	4 529	127 980
总计	14	258 469	5 049 162

系统面向人机用户可以提供多种数据服务。以《中分表》为例,面向系统的使用者,《中分表》可以提供浏览词表中的类目、主题词、概念体系或概念集合,通过搜索获取与查询字段相匹配的类目、主题词、概念体系或概念集合,通过URI标识查询并获取某个术语及其内容,支持SQARQL语言检索《中分表》的内容等术语服务。面向机器,通过API接口可以获取系统的《中分表》数据集中的术语列表、单个术语资源,以及类目、主题词、概念体系和概念集合4个数据集。国家图书馆关联数据注册与服务系统发布的以《中分表》为主的资源实体和概念URI,希望这部分数据在国内图书馆界关联数据的发布中起到知识组织的作用,可以作为其他关联数据平台RDF链接的对象,为我国关联数据聚合起到桥梁的作用。

3 未来展望

数字图书馆的建设、发展和信息技术的发展密切相关,信息技术每一次小的发展都会对数字图书馆的发展方向产生巨大的影响,为数字图书馆的建设进程起到指导和推动作用^[20]。在“互联网+公共文化”的背

景与理念下,国家图书馆更要积极拥抱语义网和关联数据技术带来的机遇与挑战。虽然目前已经开发完成关联数据注册与服务系统的基础建设,但还需要不断优化调整,使平台功能更加合理和完善。在RDF链接构建方面,目前系统只支持单条注册时人工挂接系统内的实体或概念URI,批量注册的关联数据是在系统外部通过各种方式建立RDF链接后再导入系统的,虽然目前发布的关联数据中拥有大量的RDF链接,但这部分功能大部分不是在本系统中实现的。所以在系统中对批量灌装的数据如何进行自动或半自动的RDF链接构建,如何构建和聚合系统外部的RDF链接是下一步亟须探索和解决的问题。在术语服务方面,面向机器的术语服务还略显薄弱,基于《中分表》良好的数据基础,下一步系统将通过API接口提供获取术语上下位概念、获取术语顶级概念、获取词表元数据内容、通过URI获取对应实体或概念等功能,以实现术语服务在检索发现、查询扩展、语义推理等方面的应用。

国家图书馆拥有大量的特色资源,在统一通用的关联数据注册与服务平台建设完成后,可以逐渐扩充系统中的资源。将国家图书馆拥有的规范记录和古籍善本、民国文献等特色资源逐步发布成关联数据集,将书目数据分批转换并发布为关联数据。系统还应具备“众包”思想,为相关企事业单位和公众开放相应权限的注册界面,充分借助社会力量扩展系统资源,并面向互联网用户建立数据修改与反馈机制,提供意见反馈和数据修改的界面,借助公众力量提升数据质量。此外,应该看到关联数据在图书馆界良好的发展趋势,参照美国国会图书馆基于关联数据模型的编目模式,思考用关联数据模型代替中文MARC书目数据的可能性,建立富含语义的中文编目格式,尝试在系统中建立对应的本体模型,开发相应的功能使图书馆馆员在系统中进行编目并直接发布成为关联数据。

在系统中的资源越来越丰富之后,可以以建立知识库等形式,作为数字人文服务的数据基础。再利用关联数据的技术从数据和其复杂的关系网络中挖掘事实和知识,使用数据分析相关技术帮助用户探索海量资源,再利用可视化工具形成知识地图,从而辅助人文研究。对于用户来说,系统在帮助其汇集跨领域、跨学科知识的同时,分析用户的知识背景、使用习惯、兴趣和社会关系等,通过逻辑推理和语义扩展明确其学习需求,推荐用户所需和感兴趣的知识,为用户提供更加智能化、便捷化、个性化的知识服务。对于图书馆工作者

来说,关联数据中丰富、互联的语义关系为图书馆馆员在进行科技查新等参考咨询相关工作时提供了更广泛的资源支持,可以通过关联数据的建设,进一步加强图书馆科技查新和科技咨询等科研服务的水平。

参考文献

- [1] BERNERS-LEE T. Linked Data [EB/OL]. (2009-06-18). [2019-02-15]. <http://www.w3.org/De-signIssues/LinkedData.html>.
- [2] Structured Dynamics LLC. Linked Data FAQ [EB/OL]. (2010-03-06) [2019-02-15]. http://structureddynamics.com/linked_data.html.
- [3] 沈志宏, 张晓林. 关联数据及其应用现状综述 [J]. 现代图书情报技术, 2010 (11): 1-9.
- [4] The Linked Open Data Cloud [EB/OL]. [2019-11-05]. <https://lod-cloud.net/#about>.
- [5] 潘有能, 张悦. 关联数据研究与应用进展 [J]. 情报科学, 2011 (1): 124-130.
- [6] 崔春, 毕强. 虚拟国际规范文档 (VIAF) 项目进展 [J]. 图书馆情报工作, 2014, 58 (6): 129-134.
- [7] 张海玲. 图书馆书目数据的关联数据化研究——以德国国家图书馆为例 [J]. 图书馆论坛, 2013, 33 (1): 120-125.
- [8] 徐华. 关联数据在国外图书馆中的应用及其借鉴意义 [J]. 图书馆学研究, 2011 (16): 87-89.
- [9] 姜恩波. 欧美国家级图书情报机构资源关联数据化比较研究 [J]. 图书馆建设, 2015 (7): 19-23.
- [10] 胡小菁. BIBFRAME核心类演变分析 [J]. 中国图书馆学报, 2016 (3): 20-26.
- [11] 卜书庆. 知识组织系统构建与知识服务研究 [M]. 北京: 国家图书馆出版社, 2014: 96-103.
- [12] 上海图书馆开放数据平台 [EB/OL]. [2019-11-05]. <http://data.library.sh.cn/index>.
- [13] 许磊, 夏翠娟, 刘炜, 等. 关联数据URI设计规范探讨 [J]. 国家图书馆学报, 2016 (5): 22-32.
- [14] 刘炜, 李大玲, 夏翠娟. 元数据与知识本体 [J]. 图书馆杂志, 2004, 23 (6): 50-54.
- [15] 司徒俊峰, 曹树金, 谢莉. 论基于关联数据的知识链接构建与应用 [J]. 图书情报工作, 2013, 57 (16): 123-129.
- [16] 刘炜, 夏翠娟, 张春景. 大数据与关联数据: 正在到来的数据技术革命 [J]. 现代图书情报技术, 2013 (4): 2-9.
- [17] 夏翠娟, 刘炜, 赵亮, 等. 关联数据发布技术及其实现——以

- Drupal为例[J]. 中国图书馆学报, 2012, 38(1): 49-57.
- [18] 曲佳彬, 欧石燕. 关联数据可视化研究进展分析[J]. 图书与情报, 2018, 182(4): 57-67.
- [19] 陈涛, 夏翠娟, 刘炜, 等. 关联数据的可视化技术研究与实现[J]. 图书情报工作, 2015, 59(17): 113-119.
- [20] 魏大威, 张伟. 国家数字图书馆建设与发展思考[J]. 图书馆研究与工作, 2017(1): 12-15.

作者简介

蔡颖, 女, 1986年生, 硕士, 工程师, 研究方向: 数字图书馆建设与服务, E-mail: caiy@nlc.cn。

Design and Implementation of Linked Data Registration and Service System of National Library of China

CAI Ying
(National Library of China, Beijing 100081, China)

Abstract: The development and application of linked data technology is a new opportunity and challenge for the library. The National Library of China took this opportunity to build the linked data registration and service system. The system has general processing ability of RDF data, it provides a complete state management and transformation mechanism according to the data registration, audit and publishing process, which can also realize the management of the whole life of the linked data. The design of the system follows the four principles of linked data, including data registration, data publishing, data service, data display and data management. At present, the system has registered and published three data sets, a total of 258 469 resources, about 5 049 162 triples. And provide users with data open sharing services such as resource retrieval, data download, API interface.

Keywords: Linked Data; Data Service; CCT; RDF

(收稿日期: 2019-10-27)