

# 开放关联趋势下NSTL规范控制 工作走向\*

张建勇 周毅 刘峥

(中国科学院文献情报中心, 北京 100190)

**摘要:** 本文旨在探索开放关联趋势下国家科技图书文献中心 (National Science and Technology Library, NSTL) 规范控制的发展方向, 促进NSTL规范数据资源在网络时代充分发挥应用价值。本文首先分析新的时代背景下规范控制的发展趋势, 包括规范内容和数据模型的转变。一方面, 规范内容逐渐由规范名称到唯一标识符到真实世界实体, 各图书情报机构纷纷将规范数据发布为关联数据并建立与其他数据集的连接; 另一方面, 规范数据模型由MARC转为一系列以实体关系为基础的模型。接着, 在回顾NSTL的规范控制工作的基础上, 分析其进展与局限。并且基于发展趋势与NSTL现状, 分析NSTL规范控制未来的发展方向。最后, 提出三方面的发展建议: 一是推进规范数据从封闭单一的“记录”数据走向开放关联的实体关系数据; 二是推动规范数据向更广泛范围内的第三方开放, 开发基于关联规范数据的数据服务; 三是促进跨领域规范数据融合, 实现联合档案馆、图书馆、博物馆等的关联应用和服务。

**关键词:** 规范控制; 关联数据; 唯一标识符; 国家科技图书文献中心

中图分类号: G254 DOI: 10.3772/j.issn.1673-2286.2020.07.006

引用格式: 张建勇, 周毅, 刘峥. 开放关联趋势下NSTL规范控制工作走向 [J]. 数字图书馆论坛, 2020 (7): 44-51.

现实世界中, 同一事物或个人经常会因不同的描述角度或经历拥有不同的名称, 如个人因婚姻、收养、写作等可能有多个名称, 作品因版本不同可能有多个名称。规范控制的目的是汇集同一实体的不同名称形式, 并区分具有相同名称的同一实体。图书馆界对规范控制的典型定义是: 为确保标目在检索款目及书目系统中的唯一性和稳定性而建立、维护、使用规范款目和规范文档的过程<sup>[1]</sup>。然而这一定义并没有抽象出规范控制的实质, 局限于描述卡片目录时代和机读目录时代的规范控制工作过程。规范控制的本质是实现基于概念的描述和匹配<sup>[2]</sup>。

在网络时代, 网络资源迅速增长, 体量巨大, 事物的名称种类也越来越多样化, 需要规范的概念不再只

是图书馆书目记录中的检索点。传统规范控制的效率问题、规范名称的一致性问题、规范数据的应用领域扩展问题等都对图书馆规范控制的发展提出挑战。在需求的驱动下, 国际图书馆界的规范控制由规范名称转向唯一标识符, 再走向开放关联。随着万维网联盟 (World Wide Web Consortium, W3C) 发起的关联开放数据运动的推动, 国外图书馆纷纷将书目数据或规范数据发布为关联数据, 提供关联数据服务, 如德国国家图书馆、美国国会图书馆等, 图书馆的规范数据正走向开放关联。

NSTL作为一个基于网络环境的科技文献信息资源服务机构, 其目标是建设成数字时代的国家科技文献信息资源的保障基地、国家科技文献信息服务的集成

\*本研究得到国家科技图书文献中心 (NSTL) 项目“名称规范数据库建设” (编号: 科1817)、 “STKOS超级科技词表内容建设机制与发展研究 (理学部分)” (编号: 科1802) 资助。

枢纽、国家科技文献信息服务发展的支持中心。NSTL拥有来自于多种数据源且数量庞大的信息资源,其包含的大量科研实体信息需要规范控制。在开放关联趋势下,NSTL的规范控制应如何进一步发展是需要深入研究的课题。

本文首先分析规范控制的发展趋势,包括规范内容的变化和规范数据模型的演变,在此基础上回顾并评估NSTL已经开展的规范控制工作并探讨开放关联环境下NSTL规范控制可能的发展方向。

## 1 开放关联环境下规范控制发展的趋势

### 1.1 规范内容由名称转向实体

在卡片目录时代和机读目录时代通常采用的做法,是在同一实体的多个名称形式中选定一个规范名称<sup>[3]</sup>。但一旦有规范名称修改,就会引起规范记录的修改,接着引起连接规范记录的书目记录修改,工作量难以估计。因此,图书馆改为在书目记录中记录规范记录的控制号来解决这一问题<sup>[4]</sup>。控制号是MARC规范记录的唯一性标识号,由创建、使用或发行规范记录的机构设置<sup>[5]</sup>。然而控制号仅在本地数据库有意义,对于更大范围内或国际规范数据交换时,需要在世界范围内有意义的唯一标识符。图书馆界开始意识到标识符的作用,并尝试建立多种标识符系统,包括ISADN、ORCID、ISNI等。标识符系统中不仅包含实体的名称信息,还包含实体的其他信息,对规范控制的内容开始由名称转向真实世界实体。

20世纪70年代,国际图书馆协会联合会(International Federation of Library Associations and Institutions, IFLA)曾建议使用国际标准权威机构数据编号(ISADN)用于连接相关的规范记录,但因其昂贵的开销而搁置<sup>[6]</sup>。2003年,联机计算机图书馆中心(Online Computer Library Center, OCLC)和德国国家图书馆、美国国会图书馆决定开发个人名称的虚拟规范文档VIAF<sup>[7]</sup>。VIAF现已拥有30多个国家和地区的40多个组织的规范数据<sup>[8]</sup>,其规范对象也扩展到人员名称、团体名称、地理名称、作品、书目名称等<sup>[9]</sup>。同时图书馆界内外出现大量的身份管理系统,用于管理人员、机构等实体的信息,除了名称以外,还包括身份、联系地址等其他信息。这些系统同样可以起到规范控制的汇集和消歧的作用。其中,管理人员或者机构的唯一标识

符有ORCID<sup>[10]</sup>、iAuthor<sup>[11]</sup>、ResearcherID<sup>[12]</sup>、Scopus Author ID<sup>[13]</sup>、ISNI<sup>[14]</sup>等。作品方面,数字对象唯一标识符(DOI)<sup>[15]</sup>、国际标准书号(ISBN)、国际标准连续出版物号(ISSN)、国际标准音乐作品码(ISWC)、国际标准音像号(ISAN)等也可以唯一识别作品。

### 1.2 规范数据集从封闭孤立走向开放关联

尽管唯一标识符系统在全球范围内提供服务,但其能覆盖的范围有限,一个系统想要识别所有研究者或组织几乎是不现实的。各个系统唯一识别的人员或机构存在交叉,一个作者很可能同时拥有ORCID、ResearcherID和ISNI。因此,为实现更大范围的唯一识别,有必要建立系统之间的连接。

2006年Bernets-Lee提出了关联数据的概念,他认为语义网不仅仅是将数据发布到网上,而是要建立它们之间的连接,人或者机器才能探索数据网络<sup>[16]</sup>。他同时提出关联数据的四原则:一是使用URI命名事物;二是使用HTTP URI以便于用户查找事物名称;三是当用户查找URI时,通过RDF、SPARQL等标准提供有用的信息;四是包含指向其他URI的链接,以便于用户发现更多的内容<sup>[17]</sup>。关联数据的提出为图书馆规范数据进一步统一提供了新途径。关联数据四原则中要求使用URI命名事物,即以URI为唯一标识符识别事物。包含指向其他URI链接的原则体现在不同数据集的关联。关联数据的优点在于支持数据重用,便于扩大数据内外部联接,形成数据网络,从而有利于提高数据集的质量,提升数据价值。关联数据为全球范围内规范数据的关联提供了新路径。

2007年W3C启动开放关联数据运动(Linking Open Data, LOD),号召人们将现有数据发布为关联数据。截至2020年5月,在LOD上发布的数据集由2007年的12个增长到1 255个<sup>[18]</sup>。多个图书馆或规范数据项目将其规范数据发布为关联数据。VIAF将其数据发布为关联数据<sup>[19]</sup>。美国国会图书馆自2009年开始将由其发布的所有词表发布为关联数据,并提供关联数据服务<sup>[20]</sup>。2010年德国国家图书馆将其规范数据GND发布为关联数据<sup>[21]</sup>。日本国会图书馆<sup>[22]</sup>、匈牙利国家图书馆<sup>[23]</sup>,以及国内的上海图书馆<sup>[24]</sup>等也将自身的规范数据发布成关联数据。此外,博物馆界发起的“Open Authority”项目,试图利用社会性网络,采用众包的方式,共同开发图书馆、博物馆、美术馆的规范控制服务,实现跨机

构类型的规范数据共建共享<sup>[25]</sup>。

在将规范数据发布为关联数据后,各关联数据集逐步建立了与其他关联数据集的关联,在提高自身数据质量的同时,实现更大范围的数据关联与共享。各数据集之间的关联对象包括与百科类数据集的关联、与其他规范数据集的关联以及与其他相关数据集的关联。其中与百科类数据集的关联既可以丰富规范数据集的内容,也可以借助百科网站的信息实现规范对象的语义消歧,通常关联的百科类数据集有Wikidata、DBpedia、百度百科等。与其他同类型的规范数据集或身份识别系统关联,则是通过实体对齐与实体融合,实现规范数据范围的扩大与内容的丰富,同时通过身份识别系统的唯一标识符实现规范对象的唯一识别,如与VIAF、ORCID等的关联。与相关数据集的关联包括与GeoNames等的关联,如VIAF建立了与DBpedia、GND等数据集的关联<sup>[26]</sup>,德国国家图书馆建立了与DBpedia、VIAF、ORCID、GeoNames等数据集关联<sup>[21]</sup>。

### 1.3 规范数据模型向实体关系演变

当规范控制不再局限于名称形式的字符串层面,转而以真实世界的实体本身为规范对象后,MARC格式用数字编码、元素粒度不一等的局限逐渐凸显,新的数据模型被提出。IFLA发布的“功能需求”系列模型,将文献信息的描述由外部特征转向真实实体。在开放关联环境下,各国图书馆采取复用已有数据模型或建立自定义本体作为数据模型的方式将自身的规范数据发布为关联数据。

1998年IFLA发布FRBR(书目记录的功能需求)。1999年4月,IFLA成立规范记录的功能需求与编号(FRANAR)工作组,定义规范记录的功能需求,延续FRBR的工作。FRANAR工作组在2005年公布了FRAD(规范数据的功能需求)。FRAD将一条规范记录定义为一个实体的信息集合,为规范记录的功能需求分析及规范数据的国际共享提供一个分析框架。FRAD定义了16个实体,包括个人、家族团体作品、内容表达、载体表现、单件(FRBR中的四层书目)等。在FRAD中,名称、标识符分别被独立出来作为实体,其内容分别是各种形式的名称和标识符。名称实体通过“称谓”关系与规范对象实体相连。标识符实体通过“被标识”关系与规范对象实体相连。

2010年6月,IFLA推出FRSAD(主题规范数据的功能需求)。

FRSAD中定义THEMA为表示用作作品主题的实体。NOMEN作为表示名称、标识符和检索点的实体。THEMA与NOMEN之间存在“有称谓……”“是……的称谓”的关系<sup>[27]</sup>。

FRBR、FRAD、FRSAD虽然是同系列的模型,共同构成书目世界完整的模型体系,但3个模型存在不一致,造成应用上的不便,如“个人的名称”在FRBR和FRAD中隶属于不同实体的属性<sup>[28]</sup>。2017年IFLA将FRBR、FRAD、FRSAD整合成为IFLA-LRM模型<sup>[29]</sup>。IFLA-LRM被设计用于关联数据环境,并支持和促进关联数据环境中书目数据的使用。在IFLA-LRM中最终定义了11个实体,并将实体分为三个层级。其中RES是新模型中的顶层,是其他10个实体的超类,表示其他任何实体。NOMEN作为第二级实体之一,由FRSAD中的“NOMEN”和FRAD中的“名称”合并而成。NOMEN与RES之间存在“是……的称谓”的关系。

在IFLA“功能需求”系列模型的影响下,2011年,美国国会图书馆开发了旨在适应未来万维网和广义网络世界的书目数据描述模型BIBFRAME。BIBFRAME简化了FRBR的思想,在BIBFRAME 1.0中提出4个核心类,即作品(work)、实例(instance)、规范(Authority)、注释(Annotation)。规范作为核心类是一种资源,代表与BIBFRAME作品、实例或注释相关的个人、家庭、团体、管辖权、会议、地点、主题或时间表达。在规范表达上,BIBFRAME使用属性“bf:authorizedAccessPoint”表示规范检索点,使用属性“bf:has Authority”指向外部规范,取值为URI<sup>[30]</sup>。2016年BIBFRAME更新到2.0版本,BIBFRAME 2.0表示为OWL本体。BIBFRAME 2.0取消了规范类以及创作者(bf:creator)、规范检索点(bf:authorizedAccessPoint)等属性,将需要规范的对象定义为真实世界的实体,而不是通过名称来识别,包括person、organization、place等<sup>[31]</sup>。

各图书馆机构或项目在将规范数据发布为关联数据时,根据自身的数据特点,通过两种方式构建数据模型。一是自定义的本体作为数据模型发布关联数据,在自定义本体中复用多种广泛使用的标准词表,或者与已有标准词表对齐,如德国国家图书馆建立了GND本体<sup>[32]</sup>、上海图书馆建立了人名规范库本体(shlnames)<sup>[33]</sup>等。二是尽量复用已有的数据模型,以一种现有标准词表为核心词表,以其他标准词表为补充配合使用,以达到对数据的准确描述。2014年,OCLC的国际规范文档VIAF参

考Wikidata的做法,以Schema.org为核心词表,辅之其他标准词表作为补充<sup>[19]</sup>。

## 2 NSTL的规范控制工作

开放关联环境下,国际上规范数据控制工作的变化为国内文献信息资源建设提供了发展借鉴,同时也为规范控制工作带来了新的挑战。在国际上实体管理标准与技术越来越成熟的情况下,国内的相关实践屈指可数。NSTL作为国家科技文献保障体系的核心组成部分,近年在对科研实体的规范控制中取得重大进展。

NSTL的文献数据资源来自多种渠道,包括自加工的目录数据、篇章数据和引文数据,也包括来自第三方的数据,如Web of Science、CUP、Wiley、OUP等来源的数据。这些多来源的异构数据总量已达到千万量级。数据中包含大量的科研实体信息(人员、机构、基金、期刊等),这些科研实体信息以不同的元数据标准描述,分散在各自的数据库中,不能唯一识别,也不能建立相互之间的关系。

对此,NSTL在其“十三五”规划中提出要拓展国家科技文献信息元数据资源多渠道采集方式,全面采集、整合集成和转换规范多类型的资源内容,构建统一完整的元数据框架,构建中国科技信息资源的“大”元数据体系。同时提出开展多粒度的科技文献信息深度组织与知识揭示<sup>[34]</sup>。在“十三五”规划的指导下,NSTL设计了《NSTL统一文献元数据标准》(以下简称《统一标准》)<sup>[35]</sup>,以此为元数据格式,将不同来源的数据解析、清洗、转换、集成,形成统一的元数据存储和管理系统,为NSTL的规范控制工作提供了数据基础。

### 2.1 名称规范元数据标准设计

为满足名称规范数据描述和数据交互需要,NSTL遵循《统一标准》的规定,将该标准中名称规范描述部分和名称规范关系部分独立成单个标准,包括贡献者、机构、基金等,形成《名称规范元数据标准》。《统一标准》的设计遵循模块化设计原则。其具体体现是根据实体关联方法分析抽象出资源对象的实体关系模型,以实体为单位形成描述元素集,对资源的描述就是对模型中不同实体描述的集合,描述各类实体的元素集可以复用。如机构实体的描述可以用来描述研究者所在机构,也可以用来描述出版机构。《名称规范元数据

标准》复用了《统一标准》中的贡献者、机构、基金等元素集,沿袭了模块化原则,这些元素集以实体为单位汇集了所有描述人员、机构、基金规范信息的元数据。

《统一标准》遵循最小粒度原则,将数据描述粒度尽可能细致到最小层面,如机构名称可以细分为一级机构名称、二级机构名称,地址信息可分为国家、州或省、城市以及综合的地址信息描述。在《统一标准》中采用元素与属性相结合的方式进行数据描述。其中元素是元数据的基本单元,属性用于描述和限定元素相关信息。如日期由元素“date”描述,要限定日期的类型,需要同时采用属性“日期类型(date-type)”。“出生日期”表示为“<date date-type=‘birth’>……</date>”。《名称规范元数据标准》复用《统一标准》的元素集,同样沿袭了最小粒度原则,使得规范对象的各类信息都可以在元数据标准中得到描述。元素与属性相结合的方式避免元素数量过大,又能深入细致地描述信息资源,具有表达灵活性。

《名称规范元数据标准》中还描述了规范对象的名称规范关系。规范关系的根节点下包含规范对象的唯一标识符和多个其他形式的标识符,即一条规范数据由该规范对象的一组数据组成,每个规范对象有一个唯一标识的规范ID。因《名称规范元数据标准》中以实体为单位建立描述元素集,且每一个规范对象有唯一标识符标识,因此可以揭示规范对象之间的关系,如作者与机构的关系、基金与机构的关系等。

### 2.2 名称规范系统建设

NSTL以集成融合的数据管理系统为数据来源,以《名称规范元数据标准》为元数据格式,从中抽取科研实体信息,构建NSTL名称规范系统。NSTL名称规范系统的建设首先是数据抽取,根据《名称规范元数据标准》,从文献信息中析出人员、机构、基金、期刊信息。其中,人员有姓名、联系方式、出生与死亡日期、所属机构等信息;机构有机构名称、联系方式、地址等信息;基金有基金名称、日期、主题、介绍等信息。

其次是规范控制,规范控制又包含两方面的工作:一是科研实体合并消歧,针对不同类型的规范对象,采用不同的消歧算法,在消歧过程中充分利用唯一标识符的唯一识别作用。如人员消歧算法中,首先采用ORCID、ResearcherID等唯一标识符进行识别,接着采用姓名、邮箱、研究领域、其他名称、所在机构、合作

者信息等进行区分。二是规范关系的揭示,规范关系描述同一规范对象的不同表达形式的信息,通过揭示规范关系,形成规范数据。一条规范数据是关于单个实体的一组记录,这组记录可能包含来自多个不同的数据源的多条记录。NSTL名称规范系统为每一条规范数据赋予一个规范ID唯一识别实体,通过实体关联识别与规范关系揭示,可以依托原文献信息挖掘实体与实体之间的关系,如人员与机构间的隶属关系、机构与基金之间的发起关系。

### 2.3 名称规范控制研究

NSTL在“十三五”规划中提出推进NSTL元数据的开放利用,按照通用格式规范发布开放元数据,争取以关联数据方式发布开放元数据。逐步建立跨界数据关联利用机制,逐步建立NSTL文献资源与国内外各类数据源的关联<sup>[36]</sup>。在“十三五”规划的指导下,NSTL在规范控制方面开展了一系列的研究工作,包括各类型规范数据建设研究,在NSTL名称规范库基础上的各类规范数据规范化处理与增强,以及面向未来的规范数据语义化研究。

各类规范数据建设方面,王军辉等<sup>[37]</sup>研究了期刊规范文档建设方案与应用场景并进行了实践,借鉴FRBR/FRAD模型的理念将NSTL联合目录中的每条期刊书目视为一个物理品种实体。随后在实践中以乌利希期刊资源完善了期刊的沿革关系,以PubMed期刊资源为例实现期刊别名整合,并开发了原型系统<sup>[38]</sup>。方安等<sup>[39]</sup>研究了会议规范文档建设与功能。通过NSTL联合目录系统建设实现了会议规范文档的目标功能,达到对会议信息进行规范控制和统一管理的建设目标。

NSTL名称规范库基础上的各类规范数据规范化处理与增强方面,包括从多来源母体数据的人工规范化处理、机构和人名信息的增强及处理、基金信息的增强和处理等方面,研究如何充分利用已有外部数据库提供的相关数据,丰富本地名称规范数据。

面向未来的语义化研究方面,开展名称规范数据的语义表示路径研究。通过调研现有信息服务机构和出版商在语义表示方面的研究项目和实践,分析关联数据发布项目中的数据模型,吸取经验,构建NSTL名称规范数据的数据模型,探索NSTL关联数据发布方案,推动NSTL规范数据走向语义化。

## 3 NSTL规范控制未来发展展望

NSTL的《名称规范元数据标准》以模块化设计为原则,每一种规范对象为一个模块且模块可复用,避免了MARC21在一条规范记录中描述多种实体信息的混合描述情况,体现了实体-关系思想。最小粒度原则使得规范对象的信息能够被深入细致地描述。每一规范对象赋予唯一标识符,实现规范对象的唯一识别,方便揭示规范对象之间的关系。NSTL名称规范系统利用计算机技术,依托自身丰富的资源,实现了大数量级的科研实体规范控制。科研实体的规范控制为NSTL文献检索发现系统提供了支撑,为提高其检索效果以及开发科研动态揭示、科研合作网络发现等知识服务提供了数据基础。NSTL规范控制的研究为提高NSTL规范系统的建立、规范数据的提高以及规范数据的未来发展进行了预研,积累了理论与方法基础。

在当前规范控制内容从名称选择向实体管理转变,规范数据模型由平面化的MARC或元素集转向以实体关系为基础的数据模型的趋势下,NSTL规范控制仍然有进一步拓展的空间。如《名称规范元数据标准》以单条记录为描述单元,没有完全形成以真实世界实体为描述对象的描述框架或本体,不能支持数据集的重用和开放关联。另外,NSTL名称规范数据主要用于支持NSTL文献检索发现系统,其规范ID只能在本地系统中唯一识别科研实体,不能为NSTL其他系统或第三方机构重用。NSTL名称规范数据从文献数据中析出,描述规范对象的数据有限,需要借助融合其他数据集获得质量提升。因此,研究提出下一步NSTL规范控制的发展走向。

### 3.1 推进规范数据从封闭单一的记录数据走向开放关联的实体关系数据

国内外先行者开展的多项关联数据发布实践提供了丰富的经验。NSTL可依托已经开展的语义化研究成果,在研究现有国内外信息服务机构或出版商关联数据实践的基础上,探索符合NSTL名称规范系统应用需要的关联数据模型,将NSTL名称规范数据发布为关联数据。并逐步建立关联规范数据集与外部数据集的关联,包括与VIAF等规范关联数据集的关联,与DBpedia、维基百科、百度百科等百科类知识的关联,以及与ORCID、ISNI等唯一标识符系统的关联。通过开放关

联提高数据集的质量,提高规范数据的应用价值。

### 3.2 推动规范数据向第三方开放并开发基于关联规范数据的服务

德国国家图书馆发布的《2017—2020优先战略》<sup>[40]</sup>提出要推进规范文档向所有类型的文化机构开放,提高公众对规范文档的认识和使用,实现规范文档GND基础设施现代化。德国国家图书馆开发了基于GND数据服务Entity Facts<sup>[41]</sup>,为规范文档中的实体提供机器可读的说明,使得规范数据更加容易集成到其他应用程序中。NSTL规范数据在寻求与外部资源关联融合以及提高规范数据质量的同时,也可以采取措施推动规范数据在更广范围内的应用,开发基于名称规范数据的数据服务,拓展服务范围,创新服务方式。服务范围拓展方面,研究规范数据提供服务的机制,使规范数据不仅为NSTL文献检索发现系统提供规范控制,同时也能对其他图书馆的系统或其他文化遗产机构的资源系统相融合;数据服务开发方面,支持语义搜索、科研评价、数据挖掘等服务。在关联数据支持下,当用户在数据库中查询规范数据时,可以获取实体的属性信息,还可以通过关联数据集中的关系查询到与之相关联的实体。通过对某一机构或某一作者的出版成果进行统计分析可以用于评价作者或机构等实体的科研水平。此外,关联数据化的规范数据可以支持一系列的计量学分析,包括作者合作分析、研究趋势分析、研究演化分析、区域分布分析等。

### 3.3 促进跨领域规范数据的融合及跨领域关联数据应用的开发

多国国家级机构建立了文化遗产资源的统一平台,寻求图书馆、档案馆、博物馆等文化遗产机构的联合。博物馆界发起的“Open Authority”试图联合图书馆、美术馆、博物馆共同进行规范控制。欧盟数字图书馆(Europeana)与数千个欧洲档案馆、图书馆和博物馆合作,共享文化遗产。其在2011—2015年的战略规划中将聚合并建立开放可信的欧洲文化遗产资源作为未来5年四大发展战略之一<sup>[42]</sup>。2012年10月,Europeana将2 000万条文字、图像、视频和声音的元数据转换为关联数据<sup>[43]</sup>。美国数字公共图书馆(Digital Public Library of America, DPLA)旨在将美国的图书馆、档案馆和博

物馆的资源聚合起来,将18个成员机构,超过200万条记录发布为关联数据。在2019—2022年的战略规划中,DPLA表示要不断扩大文化遗产资源汇聚网络<sup>[44]</sup>。这些案例均体现跨领域资源融合的趋势。NSTL作为国家级科技文献信息的资源保障基地、服务集成枢纽和服务发展支持中心,其下拥有多个成员单位。寻求与其他文化遗产机构联合,实现联合档案馆、图书馆、博物馆等的关联应用和服务,将形成合力,将规范控制工作推上新台阶。

## 4 结论

传统的规范控制方式已不适用于当前的需求环境,规范控制的内容与数据模型均发生了变革。开放关联为规范控制提供了新路径,在W3C开放关联运动的推动下,规范数据走向开放关联。规范数据的数据模型也由MARC走向以实体-关系为基础的描述框架或本体。国外图书馆纷纷将规范数据发布为关联数据并建立与其他数据集的关联。在这样的趋势下,NSTL制定了模块化、细粒度化的《名称规范元数据标准》。以此为数据模型建立的NSTL名称规范库系统实现了大量级的科研实体规范控制,NSTL规范控制取得重要成果。未来NSTL可以将已有研究成果应用于实践,推进规范数据从封闭单一的“记录”数据走向开放关联的实体关系数据,并推动规范数据向更广范围内的第三方开放,开发基于关联规范数据的数据服务。同时,促进跨领域规范数据融合,实现联合档案馆、图书馆、博物馆等的关联应用和服务。通过服务范围的拓展和服务方式的创新,以期充分实现图书馆资源在网络时代应有的价值。

## 参考文献

- [1] 黄俊贵. 规范控制概说[J]. 高校图书馆工作, 1999(3): 4-11.
- [2] 刘炜, 张春景, 夏翠娟. 万维网时代的规范控制[J]. 中国图书馆学报, 2015(3): 24-35.
- [3] 胡小菁. 规范控制: 从名称选择到实体管理[J]. 数字图书馆论坛, 2018(1): 2-7.
- [4] NIU J. Evolving landscape in name authority control[J]. Cataloging & classification Quarterly, 2013, 51(4): 404-419.
- [5] 国家图书馆MARC21使用手册课题组. MARC21书目数据格式使用手册[M]. 北京: 北京图书馆出版社, 2005.

- [6] Mandatory Data Elements for Internationally Shared Resource Authority Records [EB/OL]. [2019-11-15]. <https://archive.ifa.org/ubcim/p1996-2/mlar.htm>.
- [7] BENNETT R, HENGEL-DITTRICH C, O'NEILL E T, et al. Vial (virtual international authority file) : Linking die deutsche bibliothek and library of congress name authority files [C] // World Library and Information Congress: 72nd IFLA General Conference and Council. 2006.
- [8] OCLC访问VIAF [EB/OL]. [2019-11-15]. <https://www.oclc.org/zh-Hans/viaf.html>.
- [9] VIAF [EB/OL]. [2019-11-15]. <https://viaf.org/>.
- [10] ORCID [EB/OL]. [2019-11-15]. <https://orcid.org/about/what-is-orcid/mission>.
- [11] 中国科学家在线 (iAuthor) [EB/OL]. [2019-11-15]. <http://iauthor.las.ac.cn/welcome/index>.
- [12] ResearchID. Welcome to the new Web of Science ResearcherID [EB/OL]. [2019-11-15]. <http://www.researcherid.com/#rid-for-institutions>.
- [13] What is the Scopus Author Identifier? [EB/OL]. [2019-11-15]. [https://service.elsevier.com/app/answers/detail/a\\_id/11212/supporthub/scopus/](https://service.elsevier.com/app/answers/detail/a_id/11212/supporthub/scopus/).
- [14] International Standard Name Identifier (ISO 27729) [EB/OL]. [2019-11-15]. <http://www.isni.org/>.
- [15] The DOI System [EB/OL]. [2019-11-15]. <https://www.doi.org/>.
- [16] 大卫·伍德, 玛莎·扎伊德曼, 卢克·鲁思, 等. 关联数据: 万维网上的结构化数据 [M]. 蒋楠, 译. 北京: 人民邮电出版社, 2018: 3.
- [17] BERNERS-LEE T. Linked data [EB/OL]. [2019-11-15]. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [18] The Linked Open Data Cloud [EB/OL]. [2020-05-27]. <https://lod-cloud.net/>.
- [19] VIAF RDF Changes [EB/OL]. [2019-11-15]. <https://outgoing.typepad.com/outgoing/2015/04/viaf-rdf-changes.html>.
- [20] ID. LOC. GOV-Linked Data Service [EB/OL]. [2019-11-15]. <http://id.loc.gov/>.
- [21] GND [EB/OL]. [2019-11-15]. <https://lod-cloud.net/dataset/dnb-gemeinsame-normdatei>.
- [22] Web NDL Authorities (について). About the RDF Model [EB/OL]. [2019-11-15]. [https://id.ndl.go.jp/information/model\\_en/](https://id.ndl.go.jp/information/model_en/).
- [23] National Széchenyi Library (national library of Hungary) on the semantic web [EB/OL]. [2019-11-15]. [http://nektar.oszk.hu/wiki/Semantic\\_web](http://nektar.oszk.hu/wiki/Semantic_web).
- [24] 上海图书馆人名规范库 [EB/OL]. [2019-11-15]. <http://names.library.sh.cn/mrgf/home/index>.
- [25] Defining open authority in museums [EB/OL]. [2019-11-21]. <https://museumplanner.org/wp-content/uploads/2012/10/defining-open-authority-in-museums>.
- [26] VIAF: The Virtual International Authority File [EB/OL]. [2019-11-18]. <https://lod-cloud.net/dataset/viaf>.
- [27] Functional Requirements for Subject Authority Data [EB/OL]. [2019-11-18]. <https://www.ifa.org/node/5849>.
- [28] 袁硕. FR家族概念模型统一版——《FRBR图书馆参考模型》初探 [J]. 数字图书馆论坛, 2017 (1): 31-36.
- [29] IFLA Library Reference Model (LRM) [EB/OL]. [2019-11-18]. <https://www.ifa.org/publications/node/11412>.
- [30] BIBFRAME Authorities [EB/OL]. [2019-11-18]. <http://www.loc.gov/bibframe/docs/bibframe-authorities.html#subjects>.
- [31] What's New in BIBFRAME 2.0 [EB/OL]. [2019-11-18]. <https://www.loc.gov/bibframe/docs/bibframe2-whatsnew.html>.
- [32] VERSIONSHISTORIE DES LINKED-DATA-SERVICE [EB/OL]. [2019-11-18]. [https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/LDS/lds\\_versionshistorie.html?nn=250612](https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/LDS/lds_versionshistorie.html?nn=250612).
- [33] 人名规范库本体 (shInames) [EB/OL]. [2019-11-18]. <http://data.library.sh.cn/ont/ontology/tree?g=http://ont.library.sh.cn/graph/shInames>.
- [34] 彭以祺, 吴波尔, 沈仲祺. 国家科技图书文献中心“十三五”发展规划 [J]. 数字图书馆论坛, 2016 (11): 12-20.
- [35] 沈仲祺, 张建勇. 文献元数据设计指南和实践 [M]. 北京: 科学技术文献出版社, 2017.
- [36] 张建勇, 于倩倩, 黄永文, 等. NSTL统一文献元数据标准的设计与思考 [J]. 数字图书馆论坛, 2016 (2): 33-38.
- [37] 王军辉, 方安, 任慧玲, 等. 期刊规范文档建设方案与应用场景研究 [J]. 数字图书馆论坛, 2015 (7): 9-13.
- [38] 王军辉, 虞昌亮, 方安, 等. 期刊规范文档建设实践研究 [J]. 图书馆杂志, 2017, 36 (5): 41-46, 120.
- [39] 方安, 胡佳慧, 王军辉, 等. 会议规范文档建设与功能实现 [J]. 数字图书馆论坛, 2017 (7): 30-34.
- [40] 2017-2020 STRATEGIC PRIORITIES [EB/OL]. [2019-11-21]. <https://d-nb.info/1126595101/34>.
- [41] Entity Facts [EB/OL]. [2019-11-21]. [https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projektarchiv/2014/entityFacts/entityFacts\\_node.html](https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projektarchiv/2014/entityFacts/entityFacts_node.html).
- [42] Europeana. Strategic Plan 2011-2015 [EB/OL]. [2019-11-21]. <https://pro.europeana.eu/post/strategic-plan-2011-2015>.

- [43] Europeana pro, linked open data [EB/OL]. [2020-06-06]. <https://pro.europeana.eu/page/linked-open-data>.  
[44] DPLA. Strategic Roadmap, 2019-2022 [EB/OL]. [2019-11-21]. <https://pro.dp.la/about-dpla-pro/strategic-plan>.

## 作者简介

张建勇, 男, 1965年生, 硕士, 研究馆员, 研究方向: 数据库建设和数据管理。

周毅, 女, 1989年生, 硕士, 馆员, 通信作者, 研究方向: 知识组织与知识管理, E-mail: zhoyi@mail.las.ac.cn。

刘峥, 女, 1979年生, 博士, 副研究馆员, 研究方向: 知识组织与知识管理。

Research on the Development Direction of Authority Control of NSTL in the Trend of Open and Linking

ZHANG JianYong ZHOU Yi LIU Zheng

( National Science Library, Chinese Academy of Sciences, Beijing 100190, China )

Abstract: Considering of the trend of open and linked data, this paper tries to explore the development direction of NSTL authority data, and promote the full application of NSTL authority data in the network era. The paper analyzes the development trend of authority control first, including the change of authority control content and data model. On the one hand, the content of authority control gradually changed from the authority name to the unique identifier of entities in the real world, and many library and information institutions have published authority data as linked data. On the other hand, authority data model has been transformed from MARC to a series of models based on entity and relationships. Then this paper analyzes the progress and limitations of work that NSTL has done of authority control. Finally, the future development direction of NSTL authority control is analyzed, based on the trend of opening and linking and the current status of NSTL. Three suggestions have been put forward: 1) transform the authority data from closed and single "record" data to open and linked data; 2) make the authority data to be opened to more third parties, and develop more data service based on linked authority data; 3) promote the integration of cross-domain authority data and realize the associated applications and services of different institutions such as archives, libraries, and museums.

Keywords: Authority Control; Linked Data; Unique Identifier; NSTL

(收稿日期: 2020-04-10)