

NSTL联合目录资源描述发展与思考*

葛红梅¹ 丁道劲² 徐晶晶¹ 李军莲¹ 梁芳¹ 任慧玲¹

(1. 中国医学科学院医学信息研究所, 北京 100005; 2. 中国科学技术信息研究所, 北京 100038)

摘要: 在系统梳理国家科技图书文献中心 (National Science and Technology Library, NSTL) 联合目录资源描述的发展历程及现状、分析科技文献资源变化以及中心新业务流程需求变化的基础上, 提出新时期NSTL联合目录元数据体系重构方案, 思考NSTL联合目录资源描述未来发展方向。

关键词: 联合目录; 资源描述; 元数据

中图分类号: G254 DOI: 10.3772/j.issn.1673-2286.2020.08.004

引用格式: 葛红梅, 丁道劲, 徐晶晶, 等. NSTL联合目录资源描述发展与思考[J]. 数字图书馆论坛, 2020(8): 22-28.

联合目录通常指图书馆及其他文献收藏机构, 为实现一定范围内文献资源共建共享, 遵循一定的著录规则, 建设而成的文献书目清单或数据库。最有影响力的是OCLC公司推出的WorldCat, 汇集世界多个知名图书馆的馆藏, 是全球收录范围最广、文献量最大、使用用户最多的联机联合目录数据库。国内比较有影响的联合目录集中在科技期刊书目, 如中国科学院文献情报中心主办的覆盖全国专业图书馆的《全国期刊联合目录数据库》、中国高等教育文献保障系统 (China Academic Library & Information System, CALIS) 主办的汇集高校图书馆的《CALIS 西文期刊联合目录》、上海图书馆主办的服务华东地区各图书馆的《华东地区外国和港台科技期刊预订联合目录》。联合目录的建设推进各图书馆之间合作分工, 在有限经费下馆藏各有专长, 通过馆际互借, 为用户提供丰富的资源, 同时促进文献资源的利用^[1]。

NSTL联合目录是在计算机和网络环境下, 国家科技图书文献中心 (以下简称“中心”) 9家成员馆遵循一套资源描述规则、书目交互协议, 按学科分工合作, 建立的具有统一标准的科技文献资源联合目录数据库。NSTL联合目录系统是支撑NSTL联合目录的业务平台, 也是中心多个数字业务系统的集成连接枢纽和唯

一输出源头, 发挥着至关重要的作用。作为中心文献信息资源揭示服务的重要基础性工作, 随着中心几大关键数字业务平台的建立及优化, 业务平台中流转的文献资源范围和业务工作内容持续增加, 渐已形成多业务系统工作流程数字化协同的运行机制, 对中心资源的揭示、管理以及与相关系统数据交换提出了全新的需求, 因而NSTL联合目录资源描述也在不断发展中。

NSTL联合目录建设包括两个层次, 即9家成员单位本地范围内书目描述和中心联合目录数据库建设。成员馆按照中心统一制定的编目标准规范进行分散编目、上传; 中心联合目录数据库对各成员馆书目数据进行格式转换、数据合并等处理, 最终形成规范的文献资源书目数据库。成员馆基于图书馆自动化系统, 使用CNMARC或USMARC进行书目著录; 中心联合目录数据库著录格式则经历CNMARC、DC元数据不同发展阶段, 资源描述规则也不断发展完善。

1 NSTL联合目录资源描述建设沿革

1.1 印本文献MARC联机联合编目

在国家“九五”攻关项目“全国西文文献信息计算

*本研究得到国家科技图书文献中心委托课题“NSTL联合目录系统运行管理” (编号: 2019XM11) 资助。

机联合编目系统”支持下,中心开始建设联合目录系统及书目数据。联合目录系统建设成果是国家科技文献信息计算机联机编目系统Bridge2000及后续升级系统Bridge2K;书目数据包括中心9家成员馆的印本期刊和西文会议录;与中心数据加工业务、中心网络服务系统建立数据输出机制。

联合编目系统创建之初,就明确中心与成员馆的业务分工,既重视中心联机联合编目著录规则的建设,也重视本地编目工作的规范化。经过分析中心资源特征,调研各种机读目录格式,联合编目系统最终确定采用中国机读目录格式(CNMARC)为中心书目的著录格式。依据《西文文献著录条例(修订扩大版)》《中国机读目录格式使用手册》著录规则,针对中心资源特点编制中心联合编目工作指南,2004—2008年先后制定《NSTL联机联合编目工作管理试行办法》《NSTL期刊编目和签到工作规则》《NSTL西文会议文献编目工作规则》《NSTL电子资源编目手册》《NSTL电子资源管理暂行办法》《NSTL联编书目 签到质量控制操作细则》等规范^[2]。

成员馆单位编目人员水平参差不齐,既有文献编目专家,也有无任何编目基础的人员。针对当时现状,为确保中心联机联合编目工作的正常运转和编目数据的规范,中心联合编目项目组从细化著录规则和人员培训两方面推进业务开展。2004年成立编目规范研究课题组,按照国际标准,划定文献类型,按照统一格式,以分册形式编写连续出版物、科技报告、学位论文、西文图书等多种类型文献编目手册,详细说明每一种文献类型著录内容。统一著录详简级别和著录用语和文字,规定每一种文献著录项目的必备性和重复性,明确每一字段著录数据的主要信息源或规定信息源,规范统一的著录导语,启用一致的代码表(世界各国和地区代码表、语种代码表、国际地理区域代码表)、分类表,保障不同编目人员著录资源的一致性。在人员培训方面,有计划地开展编目业务培训,既有系统的编目著录规则基础培训,也有结合书目数据的典型样例分享、错误数据反馈、个人业务指导等。经过不断努力,中心成员馆编目人员都很快胜任了该工作,中心编目数据规范质量得到保障。

中心联合目录既承担传统书目功能,对文献进行组织揭示,为用户提供查阅中心资源检索途径,也要为中心下游系统提供书目数据和馆藏数据基础。因此在书目著录规则上也具有显著特点。如原有加工系统根据科

技期刊ISSN判断期刊品种,因此在连续出版物编目规则中,设置912字段著录科技期刊ISSN和题名变化的断点;对于单篇文献(论文、科技报告等)的著录按丛书或通用名称进行集中著录,统一著录一条书目记录;多卷会议录,以物理本为描述对象,分散著录揭示;同一文献,不同载体或媒介,以印本为基础只著录一条书目记录;同一文献,多家订购,则在同一条书目下,合并馆藏信息。

1.2 NSTL联合目录系统元数据规范

2010年,中心提出要实现文献遴选、编制订单、发单订购、验收登到、编目揭示、文摘加工、元数据收割、网络发布、全文服务在内的全流程业务系统化、精细化管理,提升各项业务工作的联动性、规范化。中心于2010年前后建设以“文献综合管理系统”“联合目录系统”“联合数据加工系统”为核心的数字业务工作平台,以保证业务工作的高效运作。新的数字业务系统建设均采用基于XML的元数据,对书目数据库的格式和内容规范提出了新的要求。原有的联机联合目录采用的CNMARC格式是参照卡片目录方式制定的,著录规则和数据格式均存有手工时代的烙印,书目数据具有自我封闭性,与其他系统的数据交互需要实时开展数据解析与格式转换,一定程度上限制了数据在著录规则和内容表达上的一致性^[3]。

多系统业务工作流程的运行机制和建设模式对联合编目系统的功能、数据的规范、质量的控制提出了更高的要求,为适应中心多业务系统的协同运行和交互需求,NSTL联合目录系统决定在充分调研书目数据著录规范、著录格式的基础上,结合中心建设的资源特点,制定新的NSTL联合目录系统资源描述规则。这一时期中心资源特点主要表现为:资源载体表达形式比较单一,出现多载体形式的资源仅占2.1%;多语种文献较少,多内容表达形式所占比例仅为0.5%;连续出版物出版的多变性和衍生性,如期刊发生继承、吸收等改刊情况比例较高,约为20%。同时,基于FRBR概念模型的RDA处在研究和试用阶段,AACR2仍是目前图书馆自动化管理系统通用的著录规范与格式,中心书目数据来自于各成员馆格式多样的基于MARC的书目数据。在此背景下,NSTL联合目录系统决定对DC元数据规则进行扩展,制定联合目录系统元数据著录规范。

2012年新的NSTL联合目录系统正式上线运行,

新的联合目录系统基于中心特殊的需求建立了一套《NSTL联合目录系统基本元数据标准规范》(以下简称《元数据规范》)进行资源描述与检索,属国内外先行之列。《元数据规范》全面揭示中心经费支持各成员单位订购的文献资源的外部特征、内容特征和馆藏特征。

《元数据规范》按照中心文献资源类型的界定对基本元数据进行归类,形成科技期刊、会议文献、科技丛书、文集汇编、工具书、科技报告、科技专著、学位论文8种类型的基本元数据。《元数据规范》共有9个元素集,150个元素。9个元素集分别是描述信息元素集、责任者/导师信息元素集、丛编信息元素集、期刊历史变革元素集、ISSN历史变革元素集、卷期信息元素集、期刊连接元素集、馆藏信息元素集、管理信息元素集。《元数据规范》采用《信息和文献工作 都柏林核心元数据元素集》(ISO15836—2003)的元数据定义方法定义基本元数据属性,包括标签、名称、定义、必备性、可重复性、修饰词、著录规则、注释和示例。与原有MARC规则相比,元素定义更加明确,增加电子资源描述相关元素,同时减少数据冗余。如加区分会议主办者、承办者的类型元素,增加便于签到浏览显示的期刊签到规范年、规范卷、规范期,电子资源需要著录载体变化附注、URL、馆藏载体类型和媒介类型,减少原MARC多字段著录出版年等重复字段。

1.3 NSTL联合目录系统元数据体系

2013年以后,中心在传统印本资源基础上,加大数字资源建设力度,全国开通的回溯数据库/现刊数据库达到一定规模,并组织开展OA资源建设^[4]。与之相应的,联合目录项目组在原有NSTL联合目录元数据规范基础上进一步补充、完善数字资源权益与获取信息的相关元素集及元素。根据数字资源特征,设置获取信息元素集,揭示品种级的期刊获取信息,考虑到一种期刊可能存在多个获取来源,需要分别对获取来源的地址和可获取的卷期区间范围进行描述;设置数据库描述信息、权益信息、管理信息等相关元素,构成数据库元数据,对数据库统一管理。

在资源描述的基础上,为更好地管理文献品种及多角度服务用户,中心开展了期刊规范文档、会议规范文档的建设,并将其纳入NSTL联合目录系统日常工作中。规范文档是指由规范记录组成的数据集合,包含检

索点或名称规范标目、关系属性等信息。规范文档作为一种工具,通过受控词表或规范用语控制来保证规范文档的准确性和一致性,建立规范标目与相关的不同标目形式之间的参照关系,引导用户根据多种标目形式找到相应的规范标目,进而完成针对相应目录的操作和管理。NSTL联合目录系统期刊规范^[5]重点建设期刊沿革和刊名别名,将不同时期的期刊书目通过沿革沿革建立关联,实现对期刊书目信息的统一管理和规范控制。会议规范是指描述会议规范的一组数据。NSTL联合目录系统会议规范库^[6]包括会议名称规范和会议举办机构规范。会议名称规范汇集不同表现形式的同一会议,并建立会议规范名称与原始著录名称、书目数据的关联。会议举办机构规范汇集不同表现形式的同一机构,并建立会议举办机构规范名称与原始著录名称、书目数据的关联。通过NSTL联合目录系统会议规范库的建设,为会议文献多层次、多维度、多途径组织揭示奠定基础。

在规范制定以及OAI交互方面,摒弃按文献类型分别制定的模式,而是通过整合不同文献类型资源描述需求,汇集各种文献类型的书目描述规范和规范文档描述规范,系统间统一Schema验证,制定了一套通用的NSTL联合目录元数据体系。不同文献类型的资源,以元数据体系为基础,根据元素集及元素的必备性、属性值,建设不同的元数据实例。元数据体系中,不同载体、不同媒介类型的资源,也同样使用同一套元数据体系进行描述。NSTL联合目录元数据体系总体结构见图1,分为规范元数据、书目元数据、数据库元数据和管理信息元数据。

2 中心“十三五”发展对NSTL联合目录资源描述的新要求

2.1 配合业务流程改造,建设NSTL联合目录系统2.0

中心“十三五”发展规划提出,要优化国家科技文献资源保障体系,巩固现有印本文献资源,加强开放数字资源建设,扩展数字资源^[7]。在中心整体规划下,2018年起中心业务流程改造,NSTL联合目录系统迎来新的发展阶段,建设NSTL联合目录系统2.0。NSTL联合目录系统2.0旨在对中心订购及采集的不同文献类型、载体类型、媒介类型,以及采集来源的各类印本资

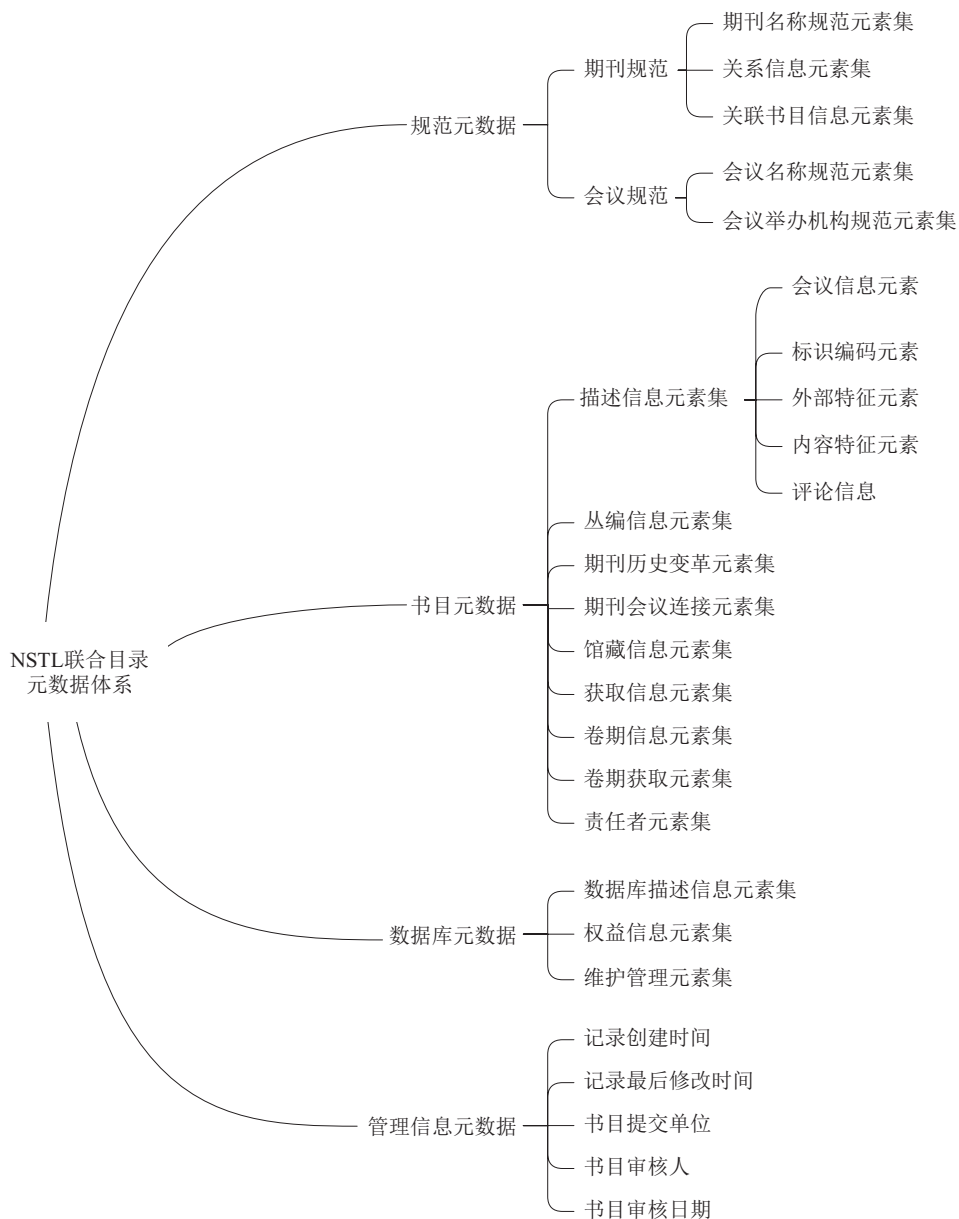


图1 NSTL联合目录元数据体系总体结构

源、数字资源母体实现区别描述、统一管理，实现书目/签到数据的转换、增值、汇聚，形成统一的单册母体信息库，并对期刊历史沿革信息、会议名称规范信息进行建设。

2.2 文献资源描述对象特征复杂

从文献类型上看，NSTL联合目录系统2.0收录资源主要有期刊、会议录、科技丛书、科技报告、文集汇编、工具书和科技专著7种，需要对不同文献类型的资

源的特征信息进行描述，如期刊的期刊沿革信息、会议录的会议信息、丛书的丛编信息等。

从采集方式上看，中心之前资源工作的重点为印本保障，资源载体形式主要为印本，近些年中心加大了电子资源的采集力度，现有电子资源有以印本方式订购的E-only资源（单品种订购电子资源）、OA资源、全国开通回溯期刊资源、全国开通现刊资源、成员馆捆绑订购的数据库资源等，需要对不同采集来源的资源的获取方式、使用权益等信息进行描述，从而保证资源获取的可靠性和使用的合法性。

从媒介类型上看,电子资源又可以区分出Online、CD-rom、USB、Email 4种,不同媒介类型的资源在描述信息源的选取和服务方式上各有不同。

2.3 遵循NSTL统一文献元数据标准

2016年7月11日,NSTL统一文献元数据标准发布实施。NSTL联合目录系统作为中心业务系统,将启用NSTL统一文献元数据标准,并基于此进行个性化补充完善。根据NSTL统一文献元数据标准,资源母体描述的对象是单一物理本或数字资源中一个论文集合^[8-9]。NSTL联合目录系统改造重点是实现从物理品种到物理本管理的转变,即建立基于单册的母体管理,对每一物理本的外部特征、馆藏信息、获取信息等进行详细描述,因此需要对书目的层级结构进行调整。

2.4 优化规范库,促进书目关联融合

期刊规范是根据期刊作为一种连续出版物,具有复杂的沿革关系特征建立的规范文档,用以对期刊沿革信息进行纵向规范控制,对内可以指导期刊采编工作,对外可以为用户检索发现提供更好的补充。

总之,要依据中心资源特征描述的需求,建立一个从多个维度体现资源对象所涉及的特征属性以及它们之间的关联、层级关系的中心资源描述元数据模型,以促进各成员馆书目的关联融合与资源发现。

3 重新构建NSTL联合目录元数据体系

3.1 建立一套标准体系,分散著录文献

重构后的NSTL联合目录元数据体系旨在使用同一套元数据方案对不同书目层级、不同文献类型、不同载体类型、不同采集方式的文献进行统一描述。通过核心元素集对资源的基本书目信息、馆藏信息、关联资源信息、馆藏信息和管理信息进行描述,在此基础上,根据资源的书目层级、文献类型、载体类型、采集方式等,建立不同的书目数据,从不同维度选用对应的特征元素集对资源特征进行揭示。

3.2 资源物理本(单册)、物理本集合分层级组织揭示并关联

NSTL联合目录系统元数据体系的特点在于适用范围扩大,扩充电子资源订购、描述与获取信息;与会议规范文档、期刊规范文档成为一个体系。NSTL联合目录系统元数据体系调整后将从三个层级对各类资源进行描述,包括单册书目、期刊品种书目和期刊规范记录。

单册书目,即一个物理本对应的书目。对非刊文献而言,即一般意义上对每一本非刊文献著录的结果。对期刊而言,指期刊的某一卷期对应的书目,是由一般意义上的“期刊书目+签到数据”组合而成的书目。不同载体类型、不同媒介类型、不同采集来源的单册均分别进行描述,形成单册分散书目,同时建立不同单册之间的关联,汇聚成单册融合书目。对非刊文献来说,在联合目录中存在单册分散书目和单册融合书目两种记录(如图2所示)。

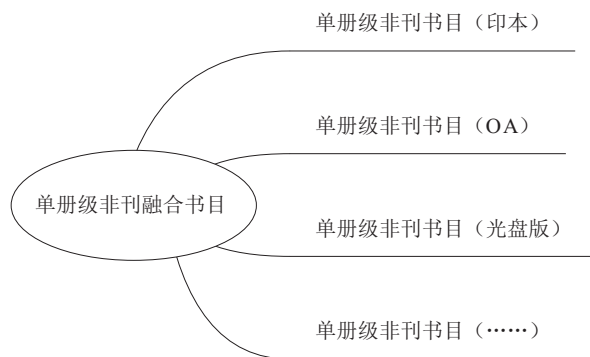


图2 NSTL联合目录系统体系非刊书目描述结构

期刊品种书目,即一般意义上的期刊书目,按品种对期刊的出版频率、沿革历史等信息进行描述。期刊品种书目也分为分散书目和融合书目,不同载体、媒介、采集方式的期刊品种对应不同的期刊品种分散书目,同一种期刊的分散书目之间进行关联,汇聚成期刊品种融合书目。

期刊规范记录,即在期刊品种融合书目的基础上对期刊历史沿革关系进行规范管理,给有沿革关系的资源赋予相同的逻辑ID,形成期刊规范记录。通过逻辑ID,将一个有历史沿革的期刊,其所有的书目都关联。

对期刊来说,则存在单册分散书目、单册融合书目、期刊品种分散书目、期刊品种融合书目、期刊规范记录等各种记录(如图3所示)。

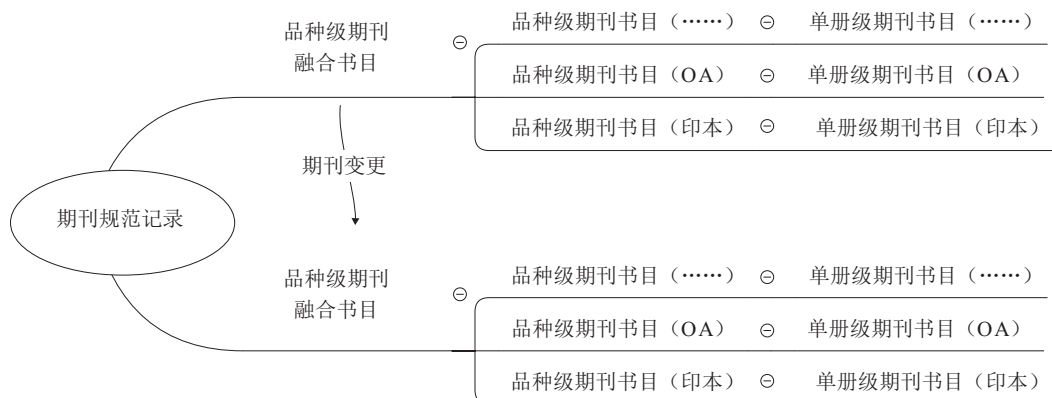


图3 NSTL联合目录系统体系期刊描述结构

NSTL联合目录元数据体系通过融合书目、规范记录实现数据关联,增加多个资源实体^[10]之间的关系,能被多渠道获取,如期刊品种书目数据和期刊物理本的单册书目数据进行关联;成员馆本地书目与中心库数据通过书目ID、多家馆藏合并等建立关联;期刊的当前书目数据、历史书目数据通过期刊规范记录建立关联,展现期刊历史沿革;印本资源和电子资源的分散书目,通过融合书目建立不同载体实体之间的关联。

4 NSTL联合目录资源描述意义和未来发展方向

通过梳理NSTL联合目录资源描述变革,可知NSTL联合目录资源描述规则伴随中心资源扩展、业务规范等不断优化,保障了NSTL联合目录的质量。在NSTL联合目录描述范围方面,2013年起开展以印本方式订购的E-only期刊描述实践,2016年开展对OA文献资源(期刊、会议录、科技报告、图书)描述实践,2017完成中心全国开通现刊文献描述实践,2018年完成中心全国开通回溯数据库期刊描述实践,保障了NSTL联合目录资源收录的完整性和准确性,为下游资源深度加工和服务提供可靠的数据源。

NSTL联合目录在印本文献资源、数字文献资源共享中发挥了文献揭示、规范导航、组织管理的作用,但是面对当前的资源快速增长及NSTL发现系统建设的需求,NSTL联合目录资源描述,在未来应加大资源获

取源头的管理、重视版权信息,为多来源馆藏资源集成与协同服务提供保障。

参考文献

- [1] 丁道劲. 从联合目录到资源调度: 馆藏资源共享服务模式新探 [J]. 图书馆学研究, 2020 (1): 55-59.
- [2] 梁芳, 白海燕, 胡铁军. NSTL联机联合编目工作的发展与实践 [J]. 数字图书馆论坛, 2010 (10): 30-35.
- [3] 夏翠娟, 吴建中. 从门户到平台——图书馆目录的转型 [J]. 图书馆论坛, 2015, 35 (7): 1-7.
- [4] 任慧玲, 葛红梅, 刘蕾, 等. NSTL开放获取资源组织揭示方式研究 [J]. 数字图书馆论坛, 2015 (8): 28-31.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.

作者简介

葛红梅,女,1979年生,副研究馆员,研究方向:图书馆信息组织、图书馆元数据建设。

丁道劲,女,1988年生,馆员,研究方向:信息资源构建与分析。

徐晶晶,女,1989年生,馆员,研究方向:资源描述、数字资源建设。

李军莲,女,1972年生,研究馆员,研究方向:资源建设、知识组织。

梁芳,女,1953年生,研究馆员,研究方向:文献资源建设与管理、文献编目标注与规范等。

任慧玲,女,1971年生,研究馆员,通信作者,研究方向:信息资源建设, E-mail: ren.huiling@imicams.ac.cn。

On the Development and Thinking of the Resource Description of NSTL Joint Directory

GE HongMei¹ DING QiuJin² XU JingJing¹ LI JunLian¹ LIANG Fang¹ REN HuiLing¹

(1. Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100005, China; 2. Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: By summarizing the development course and current situation of resource description of NSTL joint directory, analyzing the changes of scientific and technological literature resources publishing, and the changes of requirements due to NSTL's new business flow, puts forward the reconstruction scheme of NSTL joint directory metadata system in the new period, and thinks about the creative ideas of resource description development of NSTL joint directory.

Keywords: NSTL Joint Directory; Resource Description; Metadata

(收稿日期: 2020-04-10)

■ 书 讯 ■

《〈汉语主题词表〉构建研究》

《汉语主题词表》自1980年诞生以来,在我国情报检索语言发展历程中发挥了应有作用。作为国际上最为大型的综合性主题词表,一方面体现了我国图书情报领域众多专家的集体智慧,带动了我国众多专业性汉语主题词表的编制与应用;另一方面提升了我国文献资源信息组织能力和水平,强化了中文文献的文档索引和查全查准率。2009年,中国科学技术信息研究所开始了《汉语主题词表》在网络环境下的改造与重修工作。《〈汉语主题词表〉构建研究》记录十年来新型《汉语主题词表》的探索经历与实践过程,主要表述汉语类主题词表的发展演变路径,记录《汉语主题词表》的历史作用与现实意义,展示新形势下《汉语主题词表》重构的框架、思路、流程和模式,归纳《汉语主题词表》构建中探索的标准规范、基础词库、协同编制平台和范畴体系,阐述概念建设、关系构建和范畴分类等核心构建业务,详述《汉语主题词表》服务系统。曾建勋等著的《〈汉语主题词表〉构建研究》,2020年由科学技术文献出版社出版。