

NSTL联合目录书目数据质量控制体系建设*

徐晶晶 葛红梅

(中国医学科学院医学信息研究所, 北京 100005)

摘要: 本文从规则、系统、人员三方面介绍国家科技图书文献中心(NSTL)联合目录书目数据质量控制体系建设经验,通过制定详细、严格的编目规则、控制词表、校验规则和查重规则,在本地自动化系统、NSTL联合目录预处理工具和NSTL联合目录中心书目数据库三级系统的辅助下,本地编目员、中心质检员和系统管理员三级人员对数据从编目源头到入库后的各环节进行监控,从而保证数据质量。

关键词: 联合目录;元数据;质量控制

中图分类号: G254 **DOI:** 10.3772/j.issn.1673-2286.2021.09.002

引文格式: 徐晶晶,葛红梅.NSTL联合目录书目数据质量控制体系建设[J].数字图书馆论坛,2021(9):11-18.

联合目录能够一站式揭示、报道多个文献收藏机构所藏文献情况,通常由若干文献收藏机构共同遵循统一著录原则和标准合作编制而成^[1]。目前国内具有代表性的联合目录有全国图书馆联合编目中心(Online Library Cataloging Center, OLCC)的全国图书馆联合馆藏目录和中国高等教育文献保障系统(China Academic Library & Information System, CALIS)的CALIS联合目录公共检索系统,其合作模式为联机编目,使用标准机读目录格式,按照相关文献著录标准和标引词表进行,但是由于参与单位众多,参与人员水平参差不齐,使用标准不统一、数据监控管理措施不够严格等原因,导致存在重复数据、简单记录数据多、著录格式不统一、缺少必备字段等问题^[2-3]。

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(National Science and Technology Library, NSTL)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^[4]。NSTL联合目录收录各成员馆订购或采集的各类科技文献资源。具体而言,文献类型包括期刊、会议录、科技丛书、文集汇编、科技报告、工具书、科技专著7种类型;载体类型分为印刷

版、电子版;媒介类型包括光盘、网络及其他如Email、USB等形式;采集类型有NSTL经费支持的印本资源、以印本方式订购的单品种电子资源、开放获取资源、全国开通回溯数据库资源、全国开通现刊数据库资源、成员馆捆绑订购的数据库资源、以书评方式赠阅的文献资源、本地自有经费采集的资源。截至2021年6月30日,已累积392 123条书目数据及2 523 722条期刊单册数据(签到数据)。

NSTL联合目录系统是NSTL多个数字业务系统的集成连接枢纽和唯一输出源头;NSTL联合目录建设是NSTL文献信息资源揭示服务的重要基础性工作;NSTL联合目录是NSTL经费支持的订购和采集资源的实际清单,发挥着至关重要的作用,下游各种文献信息服务和更细颗粒的加工揭示工作都以此为依据。因此,NSTL联合目录数据质量控制尤为重要,有着牵一发而动全身的影响。

1 NSTL联合目录书目数据质量控制体系

NSTL联合目录采用自主研发的系统,根据NSTL图书文献资源管理的需求,进行定制开发,自2012年上

* 本研究得到国家科技图书文献中心委托课题“NSTL捆绑全文数据库西文期刊集中著录建设”(编号:2020XM33)资助。

线以来在实际应用中不断升级改进,完善系统功能,经过多年探索和实践,逐渐形成一套利用规则、系统和人员相互作用、互相制约的相对完善可靠具有可行性的书目数据质量控制方案。

书目质量控制是“建立、组织、管理和维护特定书目记录文档的所有活动”^[5]。具体到NSTL联合目录,书目数据质量控制的重点主要有两个层面。

一是单条书目数据的质量,包括字段的丰富和完备性以及著录内容的准确性,错误的(特别是题名、标准号和责任者著录错误的)、不规范的书目数据会影响读者查找文献的查全率和查准率,从而降低该资源的利用率,同时错误或不完善数据的修改维护会牵涉到下游多个系统的数据同步更新。

二是整个书目数据库的质量,主要指数据是否重复,重复数据除了在服务上给读者甄别、选择带来困惑和不便外,同时也会给下游以篇为单位的文摘引文加工系统、全文存储系统和开放资源采集系统带来大量重

复劳动。

为了提升数据质量,NSTL联合目录运行维护项目组反复审视书目数据的生命周期,提取各流程环节涉及的关键因素,形成规则、系统、人员“三位一体”的质量控制体系(见图1)。通过严格周密的规则制定,明确需要的数据;通过系统对明显不符合规则的数据进行拦截;通过人员对规则体系、系统功能进行维护,尽可能利用系统和规则控制数据质量。具体而言,通过制定和选用相关编目规则、元数据规范、控制词表、校验规则、查重规则及其他补充规则,在成员馆本地自动化系统、NSTL联合目录预处理工具和NSTL联合目录中心书目数据库(下文简称“联目中心库”)三级系统的辅助下,本地编目员、中心质检员和系统管理员三级人员对数据从编目源头到入库后的各环节进行监控,最终通过“规则制定→编目人员培训&实践→系统自动校验→系统自动查重→人工质检→数据定期清洗”6个关键环节,对书目数据质量进行有效控制。

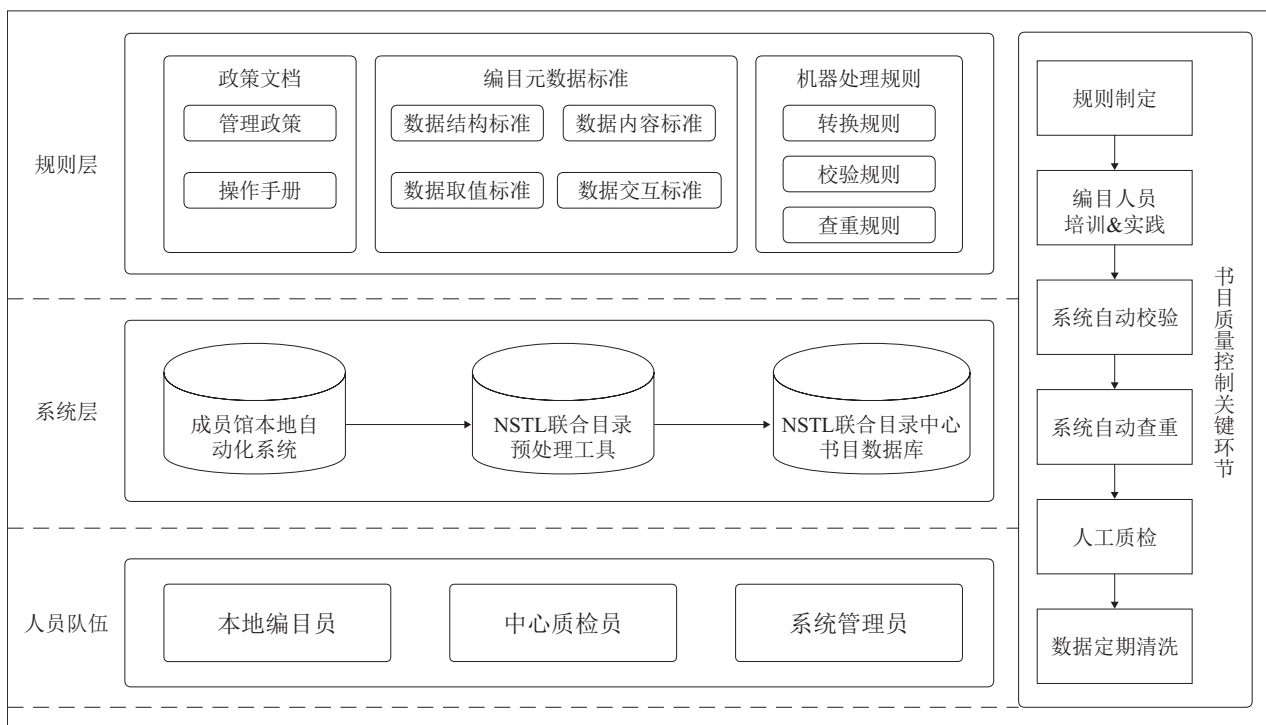


图1 NSTL联合目录书目数据控制体系

2 书目数据质量控制规则

NSTL联合目录书目数据质量控制规则体系包括管理政策和操作手册、编目元数据标准以及机器处理规则三部分(见表1)。管理政策和操作手册包括对中心编

目工作的规范要求如《国家科技图书文献中心联合目录系统管理实施办法》《NSTL文献资源界定和分类划定规则》,相关系统的操作手册如《国家科技图书文献中心联合目录系统使用手册》《国家科技图书文献中心预处理工具用户使用手册》等,保证NSTL编目工作操作

规范性和对相关概念理解的一致性。

编目元数据标准是书目数据质量控制规则体系的核心,可分为4类,即数据结构标准、数据内容标准、数据取值标准和数据交换标准。MARC属于结构标准,编目规则属于内容标准,分类法、主题词表、名称规范档等属于取值标准,2709格式或XML等属于交互标准^[6]。NSTL联合目录的9家成员单位在本地著录时使用CNMARC或MARC21格式,按照相关编目规则、取值标准(控制词表)建立本地书目数据,然后上传、转换和审核后,生成XML格式 NSTL联合目录元数据。

从本地MARC数据到NSTL联合目录元数据的过

程中,需要利用机器处理规则包括转换规则、校验规则和查重规则,对书目质量进行控制。转换规则即MARC数据与NSTL联合目录元数据的映射关系,本地书目通过预处理工具上传到NSTL联合目录主系统,在上传过程中,根据转换规则将本地MARC数据转换为NSTL联合目录元数据,并利用校验规则对转换后的联目元数据进行校验,对不符合校验规则的数据进行回退并返回意见。校验规则对元素的格式、取值范围、必备性和可重复性等进行约束。查重规则是数据进入联合目录主程序后,与联合目录已有书目数据进行查重时遵循的规则,将在下文系统控制部分详细阐述查重逻辑。

表1 NSTL联合目录书目数据质量控制规则体系

管理政策和操作手册	管理政策	《国家科技图书文献中心联合目录系统管理实施办法》 《NSTL文献资源界定和分类划定规则》		
	操作手册	《国家科技图书文献中心联合目录系统使用手册》 《国家科技图书文献中心预处理工具用户使用手册》		
	编目元数据标准	数据结构标准	《MARC21书目数据格式使用手册》 《中国机读目录格式使用手册》	
数据内容标准		《NSTL图书文献编目手册》(2020年版) 《NSTL连续出版物编目手册》(2020年版) 《NSTL西文会议文献编目工作规则》 《NSTL期刊编目和签到工作规则》		
		数据取值标准	《中国图书馆分类法》(第五版) 《世界各国和地区名称代码》(GB/T 2659—2000) 《语种名称代码 第2部分:3字母代码》(GB/T 4880.2—2000) 文献类型代码表、媒介类型代码表、采集类型代码表、成员馆代码表等NSTL联合目录自拟词表 推荐词表:OA权益相关描述词表、出版频率词表、责任方式词表	
			数据交互标准	《NSTL联合目录元数据规范》
			机器处理规则	转换规则
校验规则				对字段必备性、重复性、取值合法性等进行查验
查重规则		保证数据记录的唯一性		

2.1 NSTL编目手册

2004年NSTL联机联合编目指导委员会成立了标准规范研制课题组,对国际国内各类文献描述规则进行研究、分析并结合国内实际情况进行本地化改造,在参考《西文文献著录条例(修订扩大版)》^[7]、《新版中国机读目录格式使用手册》^[8]的基础上,制定NSTL联机联合编目系统数据制作的标准编目手册。该编目手册自2004年发布实施以来,根据NSTL文献采集政策和精细化管理的要求,历经2007年、2013年、2016年、2020年4次主要修订,目前为2020年发布的第五版。

2020年版编目手册分为《NSTL连续出版物编目手册》和《NSTL图书文献编目手册》两个分册,在充分揭示各类资源在文献类型、媒介类型、获取方式、权益等特征的基础上,按照统一体例说明各类型期刊和图书文献的编目规则,包括著录用语言和文字,著录信息源、字段和子字段的必备性和可重复性(利用字段简表对不同资源类型的著录项目的必备性和可重复性分别规定)、著录内容、著录格式等规则,并辅以实例说明,可以用以描述NSTL收录的各类型中西文资源。目前各成员单位根据2020年版NSTL编目手册,使用MARC格式和图书馆自动化系统进行本地资源编目,并且通过定

期培训、线上指导等方式加强编目员对编目规则的理解的一致性。

除了基本的编目手册之外,辅以《NSTL期刊编目和签到工作规则》对签到数据的规范性进行控制,《NSTL西文会议文献编目工作规则》对期刊中的会议文献进行分析描述。

2.2 控制词表

除了利用编目规则对编目工作进行规范外,采用多个控制词表对数据质量进行控制。成员馆在著录部分关键字段时,根据编目规则指导,利用控制词表限定范围内的术语进行描述,系统根据控制词表对著录的结果进行检查。控制词表有两类,一类为强制词表,包括:《世界各国和地区名称代码》(GB/T 2659—2000)、《语种名称代码 第2部分:3字母代码》(GB/T 4880.2—2000)、文献类型代码表、媒介类型代码表、采集类型代码表、成员馆代码表等,对系统校验和数据交互具有重大意义的字段使用强制词表进行著录并对著录结果进行检验;另一类为推荐词表,包括OA权益相关描述词表、出版频率词表、责任方式词表等,推荐词表可以更好地保障不同编目人员著录结果的一致性。在资源分类方面,除了统一使用《中国图书馆分类法》(第五版)^[9]之外,根据NSTL资源的学科特点,制定了《NSTL统一资源分类表》^[10],该分类法能与中图法进行自动映射,并且着重调整NSTL重点领域的类目层级并补充部分新兴和交叉领域类目,更适应科技文献分类标引和检索的需要。

2.3 NSTL联合目录元数据规范

2010年起,NSTL联合目录系统为适应多业务系统的协同运行和交互需求,摒弃了传统MARC描述书目数据格式,根据NSTL资源与服务特点,制定了基于DC的《NSTL联合目录系统基本元数据规范》,用以描述NSTL联合目录系统中的科技资源。2016年7月11日,《NSTL统一文献元数据标准》^[11]发布实施,NSTL联合目录系统作为NSTL主要业务系统之一,对原有NSTL联合目录元数据进行转换升级,形成新版《NSTL联合目录元数据规范》作为NSTL联合目录系统使用的数据格式与上下游系统进行数据交互。这套元数据方案删繁就简,摒弃MARC格式中复杂的字段间呼应关系和标点符号,仅保留对于资源描述和识别具有重要意义的描述字段,减轻数据转换和校验压力。

3 书目数据质量控制系统

NSTL联合目录系统分为“成员馆本地自动化系统→预处理工具→NSTL联合目录中心书目数据库”三级结构,成员馆通过本地自动化系统进行本地资源的描述,然后通过预处理工具转换上传到联目中心库。联目中心库对上传的数据进行查重处理并将处理结果通过预处理工具反馈给本地。书目数据质量的系统自动控制通过本地自动化系统字段设置、预处理工具校验和联目中心库查重三个环节实现(见图2)。

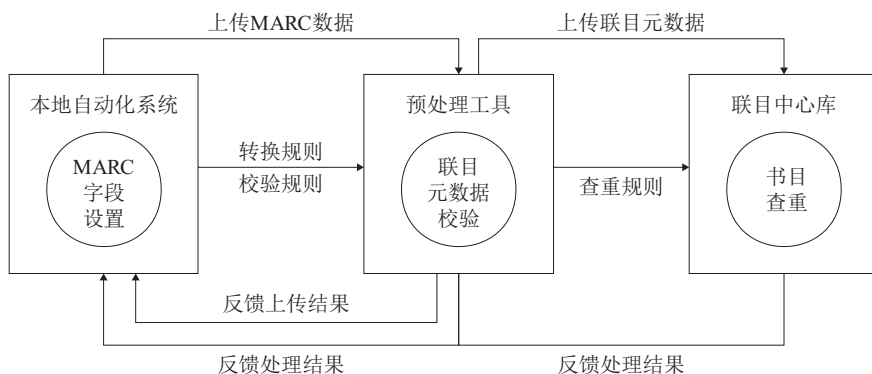


图2 NSTL联合目录书目质量系统控制机制

3.1 本地自动化系统

成员馆编目员根据编目规则在本地自动化系统进

行编目,将NSTL编目规则涉及的字段和子字段添加到自动化系统的MARC字段规则设置模块中,并设置相应的必备性和可重复性,在编目完成审校保存时,本地

自动化系统根据设置完成相应校验。由于不同文献类型、载体类型、采集方式的资源对字段的必备性、可重复性要求并不完全一致, 编目员可以根据自己负责的资源类型对字段进行个性化设置, 往往一位编目员需要同时处理多种类型的资源, 那么在成员馆自动化系统中的字段约束则是较为粗略的。

3.2 预处理工具

预处理工具作为本地自动化系统和联目中心库之间的桥梁, 主要功能有两个: ①利用转换规则将本地MARC数据转换成NSTL联合目录元数据; ②利用校验规则对转换后的元数据进行校验, 并拦截不符合要求的数据, 拦截数据主要有两类。

(1) 不符合字段必备性和可重复性要求的数据。

为了保证数据字段著录的完备性, NSTL联合目录对著录信息的详细级别要求比较高。除了规定所有资源必须具备题名、出版年、语种、出版国、分类号等基本信息外, 对于不同文献类型的资源分别规定字段的必备性和可重复性, 如期刊的出版频率、出版卷期范围, 会议的名称、时间、地点、举办机构, 丛书的丛书信息等都要求必备。尤其OA资源, NSTL联合目录要求OA资源必须具有OA标志、OA系统号并提供资源的平台信息、权益信息和获取链接等, 涉及多个字段和子字段, 且各字段间有关联关系, 则预处理工具处理数据时会对具有OA标志的资源去查找各子字段和字段著录是否符合要求, 这类判断比较复杂, 但是对于提高数据质量意义重大。OA资源特征字段描述见图3: 310字段著录资源的权益信息; 317字段著录资源来源平台信息; 856字段著录资源的获取链接、该链接所属平台、该链接提供的卷期范围; 905字段著录本馆馆藏信息, \$v子字段著录实际采集的馆藏范围, \$z子字段标识该资源为OA资源; 930\$b子字段著录OA系统的ID。

在校验过程中, 通过905\$z的OA标志, 判断该数据应该必备“310”“317”“856”“930”及相关子字段, 且310\$y、317\$a、856\$y的内容应该一致。

(2) 不符合取值格式要求的数据。取值格式的要求包括数值类型、数值长度、取值范围等不同类型。如对ISSN进行格式的规定, 必须为XXXX-XXXX, ISBN的长度仅能为10位或13位; 对语种、国家通过语种代码表、国家代码表等控制词表进行控制, 仅能使用词表中的词汇进行描述; 对资源的出版年、书目记录的创建时

间和修改时间等进行取值类型和长度的限制等。

310		\$cCC BY \$eFull-OA \$gImmediate-OA \$ySpringer Nature
317		\$aSpringer Nature \$hCompleteContent \$iPublisher \$uhttps://www.springer.com/cn \$5YK \$zOA
856	4	\$uhttps://www.springer.com/journal/10856 \$ySpringer Nature \$3v.1(1990)- \$zOA
905		\$aCN111024 \$cEP \$vv.22- 2010- \$yE \$nOnline \$zOA
930		\$bT20210310162713457

图3 OA资源特征字段著录示例

3.3 联目中心库

前述提到NSTL联合目录收录多种类型、来源的资源, 同一种资源可能通过不同的途径编目后进入NSTL联合目录系统, 如一种期刊, 可能会有2家以上的成员馆订购, 由于订购的资源载体类型不同, 可能对应不同的ISSN; 同时, 此刊可能是混合OA期刊, 根据NSTL资源采集政策, 可以增加OA采集, 并且可能被收录在成员馆订购的捆绑数据库中, 则该刊可能从多个途径进行编目后进入NSTL联合目录系统。NSTL联合目录系统需要对这些书目进行归并处理, 否则将会造成大量的冗余和错乱。具体书目界定逻辑可以利用RDA的实体-关系模型来厘清。从RDA的核心元素定义^[12]来看, 作品是指独特的知识或艺术形式, 内容表达是指作品的知识或艺术创作得以实现的方式, 作品和内容表达决定了资源本身在内容上的独特性。载体表现指作品的一个内容表达的物理体现, 不同载体表达在内容上

高度雷同。根据NSTL资源政策,不同文献类型、载体类型、采集类型的资源对应不同的采集、加工、服务政策,需要区别对待,但是又需要揭示同一种资源的不同形式之间的关系。NSTL联合目录建立新书目的标准基于载体表现,印本、CD、Online的资源在内容上具有较高的相似性,但是在物理体现上截然不同,在使用和服务中具有可替代性,因此为不同的载体表现分别建立分散书目,并且将具有不同载体表现的资源作为同一作品进行关联聚合,生成融合书目。单件指载体表现的一个样例或实例,内容和形式上完全相同,在联合目录中

对应不同的馆藏。

根据上述书目界定标准设计NSTL联合目录的查重策略,从作品级、载体表现级、单件级逐级检查,并且结合系统管理的需求配置相应查重和处理操作,着重对区分资源品种的关键字段进行查重,关键字段包括题名、ISSN/ISBN、载体类型、媒介类型、采集类型、本地书目ID;此外,对于多卷书增加分辑号/分辑名查重;对于会议录,增加会议名称和会议时间作为查重字段,具体查重逻辑见图4。

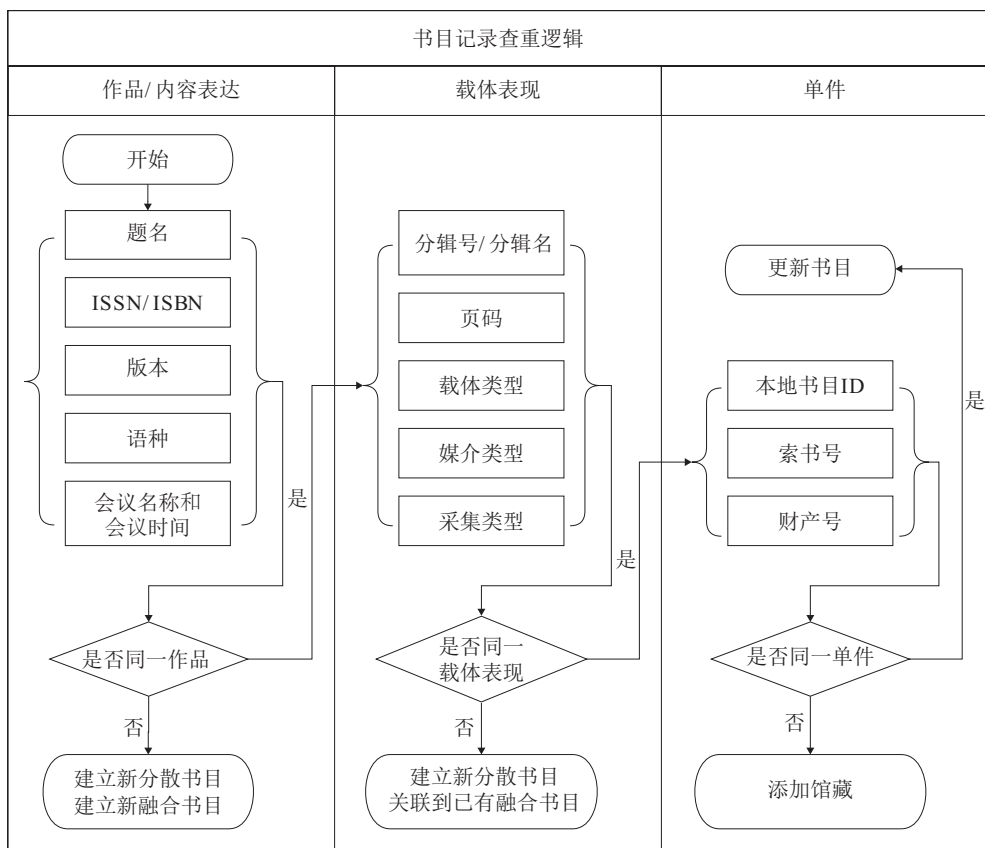


图4 查重逻辑图

查重步骤为:首先对每日入库数据的自查重,通过本地书目ID、题名、ISSN对数据进行查重,对于重复数据保留最后一次修改的记录,减轻重复入库给联目中心库查重的压力;所有当日上传的有效数据进入临时库根据对应的文献类型查重规则与联目中心库进行查重,可以系统自动判断处理的机器根据预先设置的入库规则进行自动处理操作,包括生成新书目、更新书目、添加馆藏等,如查重后对载体类型、媒介类型、采集类型完全一致的同种资源进行馆藏合并,不完全一

致的同种资源则生成新记录同时生成融合记录,融合记录下挂接同一种资源的不同形态的书目记录。其他疑似重复,系统无法自动判断的数据,留待人工质检。

由于上下游系统对于NSTL联合目录数据的依赖程度较高,系统自动查重采取比较严格的策略,对所有疑似重复的数据进行拦截。如NSTL联合目录查重的一个重点是ISSN与EISSN的交叉匹配,NSTL编目规则对于ISSN和EISSN的著录要求是有则必备,要求著录所有能够获取到的ISSN,以便提供更多的查重匹配信息,

在具体查重过程中会对所有ISSN和EISSN进行跨字段匹配(包括废止或注销的ISSN), 这样任何疑似重复的ISSN和EISSN在联目中心库中被检索到, 待审记录都会被判疑似重复, 等待人工审核。

4 书目数据质量控制人员

4.1 本地书目编目员

NSTL成员馆包括理、工、农、医、冶金、计量、机械、化工、标准等各个科技领域, 要求成员馆编目员除了具备图书馆专业知识还需要相关领域学科背景, 一般编目员可能只侧重其中一项技能。目前NSTL联合目录项目组通过定期培训、线上指导、线下走访等培养了一批熟悉NSTL联合目录编目规则的编目员, 数据质量稳步提升, 数据回退率明显下降。具体人员培养方案包括: 保证编目队伍的可持续发展, 通过成员馆老带新保持各成员馆编目队伍的稳定性, 通过定期培训保证不同成员馆之间对于编目规则理解的一致性, 对于新到馆的图情背景工作人员着重培养传统图书馆技能到NSTL特定编目规则的转换迁移, 有学科背景的馆员加强图书馆员相关资源描述和标引技能培训。

4.2 中心书目质检员

从各成员单位选用科班出身、编目经验丰富、对NSTL资源政策了解的资深编目员组建合格的质检员队伍。质检组负责的工作包括4个方面。①数据质检。对每天新上传的通过系统自动查重过滤后有疑似的数据进行人工甄别检查。②NSTL联合目录编目规范指导与答疑。通过在线工作群对成员馆编目员遇到的各种编目中的问题进行解答。③规则制定。追踪国内外资源描述进展, 分析NSTL联合目录资源描述需求变化, 定期更新维护编目规则。④组织培训。每年两次固定培训, 一次为编目规则系统培训, 尤其针对新修订的规则进行培训; 一次年底总结, 对各成员馆上传的书目质量进行总结评估, 针对易错案例重点培训。

4.3 系统管理员

在对新入库资源进行日常系统查重和人工质检外, 系统管理员需要定期对历史数据进行清洗。一是随着

对资源理解的不断加深, 查重策略不断优化, 校验规则不断细化, 需要对历史数据进行一致性处理; 二是对人工质检结果进行复核, 进一步减少人工失误导致的错误。除此之外, 系统管理员要及时纠正中心质检员、上下游系统反馈的问题数据。

5 结语

就体量而言, NSTL作为全国科技图书文献保障机构, NSTL联合目录相对于高校图书馆和公共图书馆系统联盟目录, 参与单位和人员相对较少, 收录文献类型相对集中, 中心与成员馆联系紧密, 有条件进行精细化管理, 对数据质量提出更高的要求。目前利用这套质量控制体系, 经过日常维护和2015年、2020年两次大规模数据清洗, NSTL联合目录的书目数据质量得到有效控制, 重复数据基本全部清除, 新增数据错误率控制在0.2%以内, 早期信息不完善数据应补尽补, 下游反馈问题数据量显著减少。准确完善的书目数据能够为多单位协调采集和订购资源提供有效参考, 为资源统一发布和服务提供数据支撑, 为篇级文献加工和全文采集任务管理提供有力协助, 有效减少了NSTL资源重复建设问题, 提高了资源利用率。

目前NSTL联合目录系统质量控制主要集中在字段完备性和数据重复性检查方面, 未来希望通过进一步优化校验规则、查重策略数据处理逻辑, 减轻人工质检的工作量, 质检人员的工作更多地集中在主动发现问题、提前预测问题而非事后补救和数据清洗。或可尝试利用语义分析和机器学习, 对字段内容进行进一步的控制, 进一步提高书目数据的准确性, 如对出版地和出版国的对应关系、对题名语种和作品语种著录的对应关系、对分类号和主题词的对应关系、对作者名称的著录格式等进行自动审核和规范^[13]。

参考文献

- [1] 曾建勋. 从联合目录到资源调度知识库[J]. 数字图书馆论坛, 2017(5): 1.
- [2] 肖乃菲. CALIS联合目录数据库的质量控制——以贵州省地方文献书目数据为例[J]. 大学图书情报学刊, 2016, 34(3): 78-83.
- [3] 袁乐乐. OLCC联合编目数据质量控制研究[J]. 图书情报导刊, 2019, 4(8): 24-27.
- [4] 葛红梅, 丁迺劲, 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.
- [5] 陈娟. 总分馆制下联合编目中文图书书目质量控制策略研究——以深圳市“图书馆之城”为例[J]. 图书馆研究, 2019, 49(4): 65-71.
- [6] 胡小菁. 文献编目: 从数字化到数据化[J]. 中国图书馆学报, 2019, 45(3): 49-61.
- [7] 中国图书馆学会《西文文献著录条例》修订组. 西文文献著录条例(修订扩大版)[M]. 北京: 科学技术文献出版社, 2003.
- [8] 国家图书馆. 新版中国机读目录格式使用手册[M]. 北京: 北京图书馆出版社, 2004.
- [9] 国家图书馆《中国图书馆分类法》编辑委员会. 中国图书馆分类法(第五版)[M]. 北京: 国家图书馆出版社, 2010.
- [10] NSTL统一资源分类表[EB/OL]. [2021-09-02]. <http://cmts.imicams.ac.cn/nstl/>.
- [11] NSTL统一文献元数据标准3.1[EB/OL]. [2021-09-02]. <http://spec.nstl.gov.cn/embed/metastandard.htm?metastandardid=370&base=base>.
- [12] RDA发展联合指导委员会. 资源描述与检索(RDA)[M]. RDA翻译工作组, 译. 北京: 国家图书馆出版社, 2014.
- [13] 周小敏. 全国图书馆联合编目中心书目数据质量控制[J]. 图书馆建设, 2011(2): 56-59.

作者简介

徐晶晶, 女, 1989年生, 硕士, 馆员, 研究方向: 资源描述、数字资源建设。

葛红梅, 女, 1979年生, 硕士, 副研究馆员, 通信作者, 研究方向: 图书馆信息组织、图书馆元数据建设, E-mail: ge.hongmei@imicams.ac.cn。

Construction of Bibliographic Data Quality Control System in NSTL Joint Directory

XU JingJing GE HongMei

(Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100005, China)

Abstract: This paper introduces the experience of bibliographic data quality control in NSTL joint directory from three aspects: rules, system and personnel. By making detailed and strict catalog rules, controlled vocabularies, verification rules and duplicate checking rules, with the help of local library information systems, pre-processing tool and the central database of NSTL Joint Directory System, the catalogers, the quality inspectors and the system administrators are able to monitor the data from before cataloging to after storage, so as to ensure the data quality.

Keywords: Joint Directory; Metadata; Quality Control

(收稿日期: 2021-09-02)