

# 会议规范文档建设与功能实现\*

方安, 胡佳慧, 王军辉, 任慧玲, 葛红梅, 梁芳, 胡铁军  
(中国医学科学院医学信息研究所, 北京 100020)

**摘要:** 会议文献资源的编目数据具有多样性特征。基于国家科技图书文献中心联合目录系统, 从数据业务流程角度分析会议规范文档建设的功能需求; 针对书目数据的规范化处理、规范与书目数据的挂接、规范数据间关联关系的揭示三方面建设内容, 建立数据规范化与查重机制以及数据关联与规范挂接规则; 通过在国家科技图书文献中心联合目录系统实现会议规范文档功能, 促成对系统会议信息的规范控制和统一管理。

**关键词:** NSTL; 联合目录系统; 会议规范文档

**中图分类号:** G203

**DOI:** 10.3772/j.issn.1673-2286.2017.07.006

## 1 引言

在国家科技图书文献中心 (National Science and Technology Library, NSTL) 共建共享原则<sup>[1]</sup>的指导下, NSTL联合目录系统对北京市9家成员单位上载的书目数据进行统一管理, 并为NSTL联合数据加工系统与存储系统提供数据收割服务, 是NSTL各业务系统间数据关联与交互的重要纽带<sup>[2-3]</sup>。经过对系统的初步建设、升级及新时期的改造<sup>[3]</sup>, 目前NSTL联合目录系统已良好地实现在中心馆与9家成员单位间的正常业务运转, 并为NSTL其他业务系统提供数据共享。

规范文档建设是文献编目的重要环节<sup>[4-5]</sup>, 有利于减少因数据不规范导致的重复或错误著录, 提高编目成效; 基于规范文档实现对关联数据间关系的揭示, 可丰富关联检索功能, 为按需统计分析提供便利; 此外, 基于规范数据的评价结果具有更高的可信度。规范文档建设具有重要意义, 受到国内外图书文献相关研究机构的重视: 国际图书馆协会联合会就对数据规范控制提供指导性方针<sup>[6]</sup>; 在虚拟国际规范文档项目推动下, 德国、日本、英国等的国家图书馆分别开展基于规范文档的关联数据实践活动<sup>[7]</sup>; 中国高等教育文献保障系统、国家图书馆、香港中文名称规范数据库等除制定各自名称规范著录标准外, 还构建了适应本机构发展目标

的中文名称规范库<sup>[8]</sup>。然而, 由于会议名称和举办机构的不稳定性对会议名称数据的规范控制带来困难, 集中式会议名称规范尚待构建<sup>[9]</sup>。

NSTL联合目录系统收录的文献资源包括外文会议、科技丛书、期刊 (包括香港和台湾的期刊)、科技报告、文集汇编、图书、工具书, 截至2017年4月30日, 会议文献资源近12万条, 数据量位列第一, 约占系统总数据量的50%。然而, 由于会议文献资源出版不规律、举办机构频繁更新及著录方式存在差异等内外部因素, 使会议编目数据呈现多样性特征。会议规范文档建设迫在眉睫: 一方面, 会议文献资源具有时效性和新颖性<sup>[10]</sup>, 通过传统编目方式难以根据会议文献的出版方式、出版时间、主题内容及举办机构等信息实现对关联会议数据的关系揭示; 另一方面, 由于编目人员对著录规则的理解不同及著录习惯等方面差异, 特别是在多机构联合编目的环境下, 针对各机构上载同一书目数据的差异性, 需要通过规范文档的建设对其进行规范化处理, 实现数据关联与规范挂接。基于会议规范文档建设的复杂性及其在NSTL联合目录系统中功能实现的迫切需求, NSTL于2014年7月启动“NSTL联合目录系统运行功能完善与拓展”项目, 重点开展NSTL联合目录系统的规范文档建设。

目前围绕规范文档建设的研究主要是基于人名、

\* 本研究得到国家科技图书文献中心委托课题“NSTL联合目录系统运行功能完善与拓展” (编号: 2014XM043) 资助。

机构名等作者信息的规范文档建设, 研究目标集中于面向知识服务与分析评价。例如, 孙海霞等根据知识服务需要提出基于作者机构的规范文档构建策略<sup>[10]</sup>; 王星等通过对国内外机构规范文档建设的实践与理论研究, 探索基于大数据环境下机构规范文档的构建方式<sup>[5]</sup>; 赵捷等从资源揭示和利用层面构建基于规范文档的会议导航验证系统<sup>[11]</sup>; 杨代庆等从理论角度探讨会议规范文档的构建思路, 提出外文会议规范文档的建设策略<sup>[9]</sup>, 其所采用的技术方法为本文功能实现提供有力参考。

本文基于NSTL联合目录系统, 梳理数据业务流程并分析会议规范文档的建设需求, 针对会议文献资源在编目过程中表现的多样性特征, 提出规范文档建设方案, 并在系统实现对会议规范文档各目标功能的建设。

## 2 需求分析

### 2.1 建设内容

会议规范文档的作用在于实现对会议信息的规范控制和统一管理。根据规范文档的定义<sup>[4]</sup>, NSTL联合目录系统从会议名称规范和举办机构规范两方面进行会议规范库建设。从系统结构角度看, 各成员单位通过本地预处理工具实现与NSTL联合目录系统中心管理

系统的数据交互, 因此会议规范文档的建设需要考虑本地预处理工具和管理系统的功能实现。

### 2.2 业务流程

会议规范处理流程包括基于会议规范文档的数据查重、数据审核、数据维护、交互等重要环节(见图1)。NSTL联合目录系统成员单位的本地编目人员通过预处理工具上载书目数据, 并进行系统自动查重。首次查重是新数据入系统的正常业务处理流程, 检查系统是否存在重复数据。对于无重复的新数据, 通过二次查重检查系统是否已有与之匹配的规范, 若有匹配规范的数据, 将该条数据与查重得到的规范进行挂接, 更新相应的挂接信息; 若无匹配规范的数据, 进行人工挂接, 并生成规范数据, 从而实现数据与规范的挂接。由此可见, 从业务流程的运转角度看, NSTL联合目录系统会议规范文档建设主要包括书目数据规范化处理、规范与书目数据挂接、规范数据间关联关系揭示。

### 2.3 功能需求

#### 2.3.1 数据查重

针对本地上载入库的会议书目数据, 由系统自动化处理完成基于会议规范文档的数据查重过程; 针对查

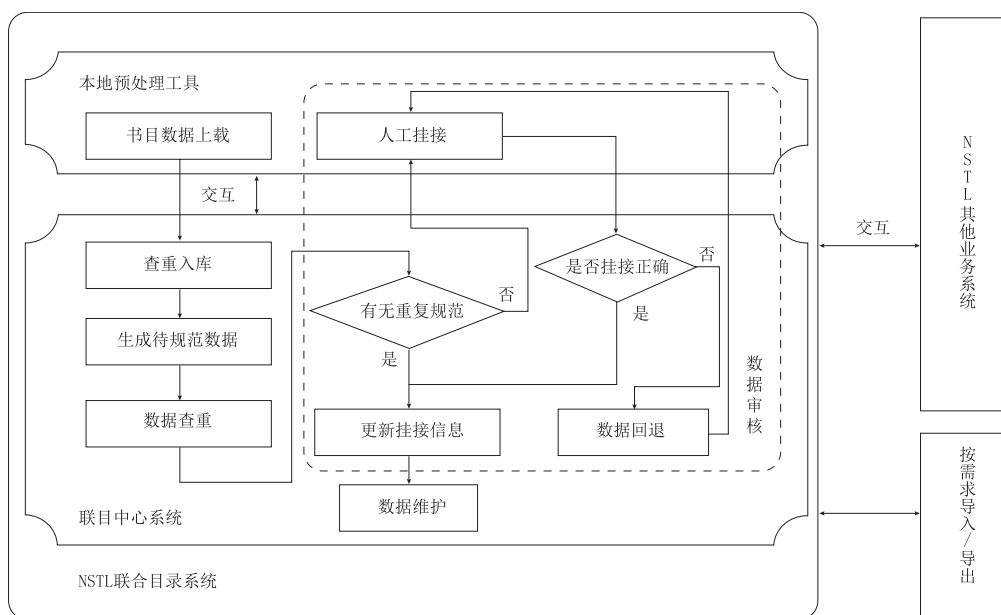


图1 会议规范处理流程

重后不确定的会议规范数据,由人工辅助查重。其中,自动化处理所依据的查重条件,根据相应元数据规范中会议名称规范元素集和举办机构规范元素集个性化定制。

### 2.3.2 数据审核

未通过系统自动查重匹配会议规范的数据,通过收割机制反馈至本地预处理工具进行人工查重,由NSTL联合目录中心馆管理系统对人工查重处理的数据进行审核。系统需要为数据审核提供两方面功能。

(1) 检索浏览。针对会议规范审核,支持根据会议名称全称、会议名称简称及成员馆名称等条件的检索,并支持对会议规范详细信息的浏览。(2) 数据回退。对有问题的会议规范记录数据,需将问题数据回退至本地预处理工具并提供回退理由,由本地编目人员处理后再提交审核。

### 2.3.3 数据维护

会议规范数据通过审核存入NSTL联合目录中心馆管理系统后,系统自动更新会议规范数据。未通过自动维护处理的数据,由人工进行维护。系统需要为数据维护提供三方面功能。(1) 检索浏览。支持对会议规范数据的检索及详细信息的浏览,其中详细信息包括书目数据与会议规范数据的挂接关系、会议父子关系。(2) 数据删除。若无相关会议挂接信息,可对会议规范数据其进行数据删除。(3) 数据归并。对有关联关系(如父会与子会)的会议规范数据,应将会议规范下的挂接关系进行归并处理,建立相应关联关系。

### 2.3.4 数据交互

NSTL联合目录系统是NSTL整个服务体系的关键环节,主要通过建立收割反馈机制调用OAI(Open Archives Initiative)协议实现数据交互。会议规范文档的数据交互包括三方面。(1) 与成员单位的数据交互。NSTL联合目录中心馆管理系统根据收割参数自动从预处理工具中收割需更新的会议规范数据,预处理工具从管理系统收割更新后的数据及其反馈处理意见。

(2) 与NSTL其他系统的数据交互。为NSTL其他业务系统收割会议规范数据提供接口,并支持NSTL联合目

录系统对相应反馈信息的收割。(3) 与第三方的数据交互。基于可扩展性的系统设计原则,支持对会议规范数据的按需导出。

### 2.3.5 附加功能

在NSTL联合目录系统实现会议规范文档建设过程中,需相应支持对会议规范数据的统计、会议规范文档的日志管理及在Web OPAC上增加对会议规范数据的检索浏览等功能。

## 3 建设方案

### 3.1 数据规范化机制

依据《NSTL联合目录系统元数据规范(2016年版)》和《国家科技图书文献中心会议录编目手册(2016年版)》,对会议名称和举办机构进行规范化处理。

由于著录习惯等因素,不同成员单位上载的同一条会议书目数据不相同。会议名称不规范直接影响会议文献的检全率和检准率,对会议数据的规范化操作主要包括过滤、替换和转换3种。(1) 过滤。针对会议编目数据,过滤会议名称和举办机构名称中的冠词(如“the”)、标点符号(如“^”)、多余空格等内容。(2) 替换。针对会议编目数据,替换因未按照规范书写而造成的序数词、基数词、连接符、机构缩写等内容。

(3) 转换。针对会议编目数据的小写英文字母和全角字符,分别将其转换为大写和半角格式。

### 3.2 查重机制与推荐规则

NSTL联合目录系统中心馆管理系统依据查重机制实现会议规范数据的自动化关联挂接,按照规范名称、挂接名称、索引依次对会议规范数据进行查重匹配,若返回匹配规范的查重结果,则读取会议规范信息并建立挂接关系;若无匹配规范的数据返回,则将该数据返回本地预处理工具进行人工核对。

本地预处理工具根据会议规范推荐规则实现人工挂接,其将待规范的数据依次与会议规范文档的挂接数据、简写名称、会议名称的匹配单词数进行对比,将疑似相同的会议规范作为相应会议数据的推荐规范。

### 3.3 实体关系模型与数据关联挂接

#### 3.3.1 实体关系模型

会议书目数据与会议规范文档通过书目ID、会议名称规范ID及举办机构规范ID实现关联挂接,会议规范文档实体关系模型如图2所示。其中,书目表包含书目ID、题名、标准号、会议名称、举办机构名称等信息,基于同一书目ID的书目表可能包含多个会议信息;会议名称规范表包含会议名称规范ID、会议规范全称和规范简称;举办机构规范表包含举办机构规范ID、举办机构规范全称和举办机构规范简称;在会议关联书目表提取会议书目表中的书目ID、会议全称、会议简称、会议举办年和会议举办届等重要信息,并通过会议名称从会议名称规范表读取相应会议名称规范ID(若无,则新增该规范ID);在举办机构关联书目表提取会议书目表中的书目ID、举办机构全称和简称等重要信息,并通过举办机构从举办机构规范表读取相应举办机构规范ID(若无,则新增该规范ID);会议规范关系表展示了会议1与会议N间的关联关系。

#### 3.3.2 关联挂接示例

根据图2所示的实体关系模型,会议书目数据与会议规范文档通过书目ID、会议名称规范ID以及举办机构规范ID实现关联挂接。其中,会议名称规范ID生成规则采用“HY/ZB+年份(4位)+流水(6位)+校验位(1位)”,校验位的计算方法参考ISBN校验规则。通过会议名称规范ID为“HY20140041993”的数据给出如下关联挂接关系示例。

(1) 会议书目数据与会议名称规范间的关联挂接关系。会议名称规范ID为“HY20140041993”的数据挂接两条会议书目数据,会议书目ID分别为“N2008E-MST0013948”和“N2012EMST0006733”,通过会议书目ID实现会议书目数据与会议名称规范间的关联。

(2) 会议父子关系。会议名称规范ID为“HY20140-041993”的数据包含两条子会议信息,子会议的会议名称规范ID分别为“HY20140194480”和“HY201-40271486”,通过会议名称规范ID实现父会与子会间的关联。

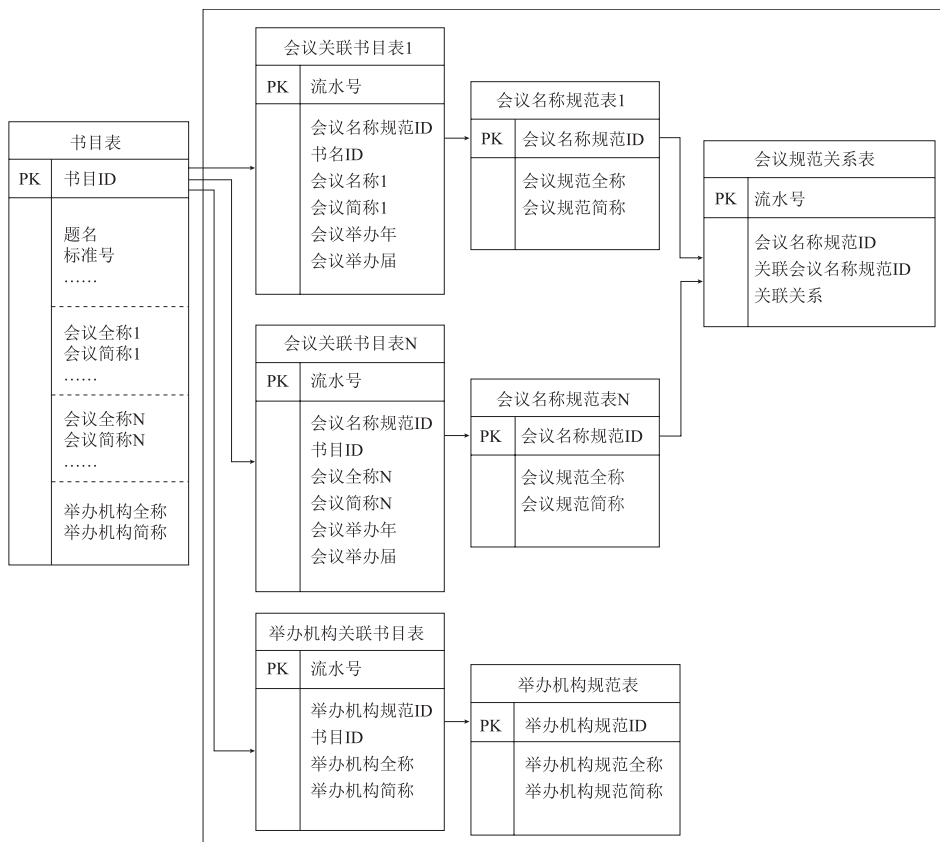


图2 会议规范文档实体关系模型

## 4 功能实现

根据上述建设方案, NSTL联合目录系统建设实现了会议规范文档各目标功能, 通过人工辅助系统处理实现书目数据的规范化、规范与书目数据的挂接及对规范数据间关联关系的揭示, 实现对会议信息进行规范控制和统一管理的建设目标。

针对会议名称规范和举办机构名称规范, 系统支持对名称规范详细信息的查看、数据归并、数据导出、数据删除和查看挂接等功能。

## 5 结语

基于NSTL联合目录系统, 本文从数据业务流程角度分析会议规范文档建设的功能需求, 并通过建立数据规范化与查重机制、数据关联和规范挂接规则, 对会议书目数据中的会议名称和举办机构进行规范化处理, 并在NSTL联合目录系统中实现规范与书目数据的挂接, 完成对规范数据间关联关系的揭示, 有助于对会议信息的规范控制和统一管理。本文在会议规范文档建设实践的经验, 将为NSTL联合目录系统下一步期刊规范文档的功能实现提供借鉴。

## 参考文献

- [1] 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.
- [2] 吴思竹, 胡铁军, 梁芳, 等. NSTL联合目录系统元数据的数据逻辑结构设计[J]. 图书馆杂志, 2014, 33(1):31-35.
- [3] 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.
- [4] 曾建勋. 推进规范文档建设[J]. 数字图书馆论坛, 2015(7):1.
- [5] 王星, 曾建勋, 苏静, 等. 机构规范文档构建方式研究[J]. 数字图书馆论坛, 2015(7):2-8.
- [6] IFLA[EB/OL]. [2017-06-23]. <https://www.ifla.org/>.
- [7] VIAF: the virtual international authority file[EB/OL]. [2017-06-23]. <http://www.viaf.org/>.
- [8] 石燕青. 中文个人名称规范文档共享研究及语义化探索[D]. 太原: 山西大学, 2016.
- [9] 杨代庆, 赵捷, 贾明静, 等. 外文会议规范文档构建研究[J]. 数字图书馆论坛, 2015(7):14-19.
- [10] 孙海霞, 李军莲. 学术论文作者机构规范文档构建[J]. 医学信息学杂志, 2015, 36(11):42-47.
- [11] 赵捷, 杨代庆, 王星. 基于规范文档的学术会议导航系统构建研究[J]. 数字图书馆论坛, 2015(7):20-24.

## 作者简介

方安, 男, 1976年生, 副研究馆员, 研究方向: 医学知识组织与数字图书馆。

胡佳慧, 女, 1987年生, 助理研究员, 研究方向: 医学数字资源保存与服务, E-mail: hu.jiahui@imicams.ac.cn。

王军辉, 男, 1982年生, 助理研究员, 研究方向: 医学信息资源管理与利用。

任慧玲, 女, 1971年生, 研究馆员, 研究方向: 信息资源建设。

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

梁芳, 女, 1953年生, 研究馆员, 研究方向: 信息资源建设。

胡铁军, 男, 1949年生, 研究员, 研究方向: 医学知识组织、数字图书馆、网络信息系统。

## Conference Authority File Construction and Function Realization

FANG An, HU JiaHui, WANG JunHui, REN HuiLing, GE HongMei, LIANG Fang, HU TieJun  
(Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China)

Abstract: The cataloging data of the conference literature resources present diversity characteristic. Based on NSTL union catalog system, the functional requirement of the conference authority file construction is analyzed from the data service process point. To construction three contents, i.e., the normalization of bibliographic data, the association between authority file and bibliographic data, the relationship reveal between authority data, mechanisms of data standardization and duplicate-checking are constructed, as well as the rule of data association and standard hanging; the standard control and unified management of conference information are achieved in the NSTL union catalog system through the conference authority file construction.

Keywords: NSTL; Union Catalog System; Conference Authority File

(收稿日期: 2017-04-21)