

基于关联数据的 非物质文化遗产资源聚合研究

翟姗姗 著



科学出版社

国家科技支撑计划（2012BAH83F00）、文化遗产传承与数字化
保护协同创新中心招标课题（2015XT007）及教育部人文社会
科学研究青年基金项目（BYJC870029）联合资助

基于关联数据的非物质文化遗产 资源聚合研究

翟姗姗 著

科学出版社

北 京

内 容 简 介

本书引入关联数据的理念与技术方法,在对非物质文化遗产、资源聚合相关理论进行系统梳理的基础上,基于元数据与本体对非物质文化遗产资源进行了语义化标注,并开展了非物质文化遗产关联数据构建与发布的关键技术研究,建立了基于关联数据的非物质文化遗产资源聚合与服务平台,同时选取具有地方代表性的非物质文化遗产项目“楚剧”进行实证研究。由此证明引入关联数据的理论与技术方法,是实现非物质文化遗产资源的规范化描述、语义化揭示、多维度关联的有效途径,对提高非物质文化遗产资源聚合与保护、优化知识服务质量有着重要的作用。

本书内容丰富,应用性强,可供数字图书馆、信息组织、信息资源管理等领域从事相关研究的专家学者及院校相关专业师生参考使用。

图书在版编目(CIP)数据

基于关联数据的非物质文化遗产资源聚合研究/翟姗姗著. —北京:科学出版社,2015

ISBN 978-7-03-045479-9

I. ①基… II. ①翟… III. ①数字技术—应用—文化遗产—资源管理—研究 IV. ①G04-39

中国版本图书馆 CIP 数据核字(2015)第 200203 号

责任编辑:余 丁 阙 瑞 / 责任校对:桂伟利

责任印制:徐晓晨 / 封面设计:迷底书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2015 年 6 月第 一 版 开本:720×1 000 1/16

2015 年 6 月第一次印刷 印张:13 3/4 插页:2

字数:266 000

定价:69.00 元

(如有印装质量问题,我社负责调换)

前言

五千多年的历史、五十六个民族多元性文化生态，造就了我国多姿多彩的非物质文化遗产（简称非遗）。这些非物质文化遗产蕴含着特有的精神价值、思维方式和民族文化，体现着中华民族的生命力和创造力，是各民族智慧的结晶，也是全人类文明的象征。当前，非物质文化遗产资源的保护既面临着机遇也必须迎接新环境的挑战。以非物质文化遗产资源为基础，研究其数字化关键技术，建立非物质文化遗产数字化技术规范，开发资源聚合与服务平台，实现对非物质文化遗产资源的描述、关联、聚合与服务，对发展我国文化事业，促进社会精神文明建设，提升我国的民族凝聚力和自豪感都具有重要意义。然而，在非物质文化遗产资源聚合的实际过程中存在着资源语义化描述程度不高、关联强度不够、知识服务水平较低等问题，如何整合多学科的方法实现非物质文化遗产资源精确化描述与语义揭示，并建立起资源间的知识关联，从而有效实现非物质文化遗产资源聚合，更好地满足用户知识需求并实现非物质文化遗产合理保护，就成为了当前亟待解决的问题。

本书引入关联数据的理念与技术方法，在对非物质文化遗产、资源聚合相关理论进行系统梳理的基础上，基于元数据与本体对非物质文化遗产资源进行了语义化标注，并开展了非物质文化遗产关联数据构建与发布的关键技术研究，建立了基于关联数据的非物质文化遗产资源聚合与服务平台，同时选取具有地方代表性的非物质文化遗产项目“楚剧”进行实证研究，以期验证本书所提出的基于关联数据的非物质文化遗产资源聚合方法的科学性与有效性。

本书共7章。第1章主要探讨基于关联数据的非物质文化遗产资源聚合这一科学问题的研究背景、国内外研究现状、研究思路与技术路线、创新之处以及本书的组织结构。第2章是全文的理论基础。包括三个部分，分别是语义网与语义标注相关理论、关联数据理论基础、非物质文化遗产资源聚合有关理论。第3~5章是本书的核心章节，按照“资源标注——多维关联——语义聚合”的思路展开研究。第3章，非物质文化遗产资源描述及其语义标注。这是实现非物质文化遗产资源聚合的基础性工作，旨在为下一阶段进行非物质文化遗产资源多维关联提供结构化的资源描述；第4章，非物质文化遗产关联数据的创建与发布。其目的是建立起资源间的语义关联关系，这是实现非物质文化遗产资源聚合的技术手段；第5章，基于关联数据的非物质文化遗产资源聚合方案设计。本章在实现非物质文化遗产资源语义标注与多维关联的基础上，通过搭建资源聚合与服务平台，实现了非物质文化遗产资源在语义层面的聚合。第6章以国家第一批非物质文化遗产保护项目——楚剧为例开展了

实证研究，本章在分析了楚剧资源聚合的必要性与可行性之后，实现了楚剧资源的标注、关联及聚合，验证了基于关联数据的非物质文化遗产资源聚合方法的可行性与有效性。第7章，总结与展望。一方面，对全书所作出的研究工作与研究内容进行归纳和总结；另一方面，对研究不足及今后的研究方向进行了探讨。

本书的研究表明，引入关联数据的理论与技术方法，是实现非物质文化遗产资源的规范化描述、语义化揭示、多维度关联的有效途径，对提高非物质文化遗产资源聚合与保护、优化知识服务质量有着重要的作用。

本书的研究得到“十二五”国家科技支撑计划“荆楚文化遗产数字化公共服务关键技术研究及示范”（2012BAH83F00）项目、文化遗产传承与数字化保护协同创新中心招标课题“荆楚非物质文化遗产数字资源知识聚合与应用研究”（2015XT007）及教育部人文社会科学研究青年基金项目“基于用户兴趣挖掘的电子政务门户知识整合研究”（BYJC870029）的支持。在项目的研究和本书的编写过程中，黄永林教授、夏立新教授、詹一虹教授、谈国新教授、王学东教授、娄策群教授、夏南强教授、王伟军教授等众多专家悉心指导，提出了许多宝贵的意见和建议，作者所在研究团队也给予了大力的支持，在此一并表示诚挚的谢意。

在项目研究与本书撰写过程中，尽管作者投入了大量精力，付出了艰辛努力，但仍难免存在错误和疏漏，恳请各位专家、同行批评指教。

翟姗姗

2015年5月

目 录

前言	
第 1 章 绪论	1
1.1 研究背景及研究意义	1
1.2 国内外研究现状	3
1.2.1 关联数据研究现状	4
1.2.2 非物质文化遗产资源聚合研究现状	7
1.2.3 研究述评	13
1.3 研究内容与技术路线	14
1.4 研究的创新之处	17
1.5 组织结构	18
第 2 章 相关理论基础	20
2.1 语义网及语义标注相关理论	20
2.1.1 语义网的发展	20
2.1.2 语义标注概述	21
2.1.3 语义网环境下元数据、本体及语义标注间的关系	23
2.2 关联数据有关理论	24
2.2.1 概念解析	25
2.2.2 类型划分	26
2.2.3 支撑技术	26
2.2.4 关键实现技术	30
2.3 非物质文化遗产资源聚合概述	37
2.3.1 有关概念的内涵与外延	37
2.3.2 相关理论	38
2.3.3 维度分析	40
2.3.4 聚合方式的演进	41
第 3 章 非物质文化遗产资源描述及其语义标注	43
3.1 非物质文化遗产资源分类体系	43
3.1.1 分类现状	44

3.1.2	现有分类方法的局限	46
3.1.3	分类体系的构建	47
3.2	非物质文化遗产资源元数据标准规范	53
3.2.1	常用元数据标准	53
3.2.2	二维元数据框架构建	54
3.2.3	元数据结构的实现	57
3.3	非物质文化遗产资源本体构建	73
3.3.1	本体构建原则	73
3.3.2	本体构建方法与流程	74
3.3.3	本体构建的工具	75
3.3.4	本体构建的实现——以中国传统戏剧为例	76
3.4	基于本体的非物质文化遗产语义标注模型	83
3.4.1	语义描述的层次	83
3.4.2	语义描述模板	84
3.4.3	语义标注模型的构建	85
第4章	非物质文化遗产关联数据的创建与发布	88
4.1	非物质文化遗产关联数据创建与发布的基本原则	88
4.2	非物质文化遗产关联数据创建与发布的流程	89
4.3	基于语义的非物质文化遗产关联模型构建	92
4.3.1	核心资源及其属性	93
4.3.2	项目类及其核心属性	98
4.3.3	机构类及其核心属性	99
4.3.4	相关人员类及其核心属性	101
4.3.5	数据库集类及其核心属性	102
4.3.6	相关事件类及其核心属性	106
4.4	非物质文化遗产关联数据创建与发布平台的设计	109
4.4.1	DHC 在非物质文化遗产关联数据创建与发布中的应用	109
4.4.2	非物质文化遗产关联数据创建与发布架构	112
4.4.3	非物质文化遗产关联数据创建与发布平台的设计与实现	114
第5章	基于关联数据的非物质文化遗产资源聚合方案设计	126
5.1	非物质文化遗产资源聚合的原则	126
5.2	关联数据在非物质文化遗产资源聚合中的应用优势分析	129
5.3	基于关联数据的非物质文化遗产资源聚合框架	131
5.3.1	非物质文化遗产资源聚合宏观模式	131

5.3.2	非物质文化遗产资源聚合微观模式	140
5.3.3	基于关联数据的非物质文化遗产资源语义聚合框架构建	142
5.4	基于关联数据的非物质文化遗产资源聚合与服务平台的设计与实现	145
5.4.1	基于关联数据的非物质文化遗产资源聚合与服务平台设计目标	145
5.4.2	基于关联数据的非物质文化遗产资源聚合与服务平台整体设计思路	147
5.4.3	基于关联数据的非物质文化遗产资源聚合与服务平台关键功能实现	149
第 6 章	基于关联数据的非物质文化遗产资源聚合实证研究——以楚剧为例	161
6.1	楚剧资源聚合的必要性与可行性分析	161
6.2	楚剧资源采集及语义描述	163
6.2.1	楚剧资源收集及元数据标准建立	164
6.2.2	楚剧领域本体开发	167
6.2.3	楚剧资源语义标注	172
6.3	基于楚剧资源的关联数据创建与发布	177
6.4	非物质文化遗产资源聚合与服务平台应用实例——楚剧资源展示	183
6.5	与传统信息系统的对比分析	196
第 7 章	总结与展望	198
参考文献		202
彩图		

第1章 绪 论

本章在对当前非物质文化遗产资源聚合中所涉及的理论与方法进行深入分析的基础上,指出当前非物质文化遗产资源聚合中存在的一系列问题,提出基于关联数据的非物质文化遗产资源聚合方法,并对国内外研究现状、研究内容与技术路线以及创新之处进行了阐述,最后介绍了本书的组织结构。

1.1 研究背景及研究意义

本书的选题系“十二五”国家科技支撑计划“荆楚文化遗产数字化公共服务关键技术研究及示范”(2012BAH83F00)、文化遗产传承与数字化保护协同创新中心招标课题“荆楚非物质文化遗产数字资源知识聚合与应用研究”(2015XT007)及教育部人文社会科学研究青年基金项目“基于用户兴趣挖掘的电子政务门户知识整合研究”(BYJC870029)研究成果之一。

一方面,非物质文化遗产资源建设依托于文化部文化共享工程,由国家牵头、地方参与,对地方志、特色文献、民风民俗等一系列独具地方特色的文化资源进行加工整理。作为一类特殊的馆藏资源,对其进行聚合具有典型的示范意义;另一方面,非物质文化遗产是人类创造力、想象力、智慧和劳动的结晶,对其进行有效的资源聚合和合理保护是当前该领域亟待解决的实际问题。所以,研究基于关联数据的非物质文化遗产资源聚合具有重要的理论意义和实际应用价值。

1. 研究背景

非物质文化遗产与物质文化遗产一样,都是人类伟大文明的结晶。两者作为现存的文化记忆,物质文化遗产的可见性,决定了其在保护和传承上的优势,而非物质文化遗产的无形性和活态性,使其所包含的文化记忆更容易随着时代的变迁和时间的推移而被人们忽略和遗忘。非物质文化遗产的保护,不仅仅关乎某项具体文化领域的生存问题,更是对人类精神文明内涵与意义的深刻理解。当前,非物质文化遗产保护面临着一系列问题,如消亡现象严重、工业化与城市化的影响较大、文化生态空间的异化等,在这种情况下,对非物质文化遗产资源实现有效聚合就显得十分重要。随着多媒体技术和网络技术的发展,非物质文化遗产资源聚合与保护有了新的方法途径——数字化遗产保护技术。

我国文化遗产的数字化保护工作正在如火如荼地开展,并已经初见成效,尤其随着互联网的兴起,非物质文化遗产的数字化保护得到了进一步的发展。概括起来,非物质文化遗产数字化保护的关键技术涉及数字化建模、遥感技术、虚拟现实技术、数字化辅助设计系统、破碎文化修复及复原模拟技术、数字化舞蹈编排与声音驱动技术等,这些数字化技术的发展为非物质文化遗产资源聚合与保护提供了坚实的技术基础。然而,我们并不满足其数字化的保护和保存,更要通过新的方法与手段对其进行重新阐释,赋予它们新的内涵,并对其资源进行合理组织与整合,更加有效的实现非物质文化遗产资源的知识传承与利用。

非物质文化遗产资源的异构性、多源性使其在聚合时存在较大的困难,主要表现为资源描述语义程度低、集成度不高、关联深度与广度不够、服务质量和水平有待提升等。一方面,不同的非物质文化遗产相关信息机构往往采用不同的硬件平台、数据库和数据描述标准,且对内部资源实行独立管理,因而造成了大量的分布式异构资源的存在。传统的资源组织与整合方式只能解决这些资源在物理及逻辑结构上的异构,而对资源间的语义异构现象无法提供有效的解决途径;另一方面,当前的非物质文化遗产资源聚合是基于数字化技术的,也就是资源聚合的对象往往是数字资源,如音频、视频、图片等,但与非物质文化遗产有关的其他资源却没有囊括其中,如某一领域的知识资源、非遗数据库资源、人员/机构数据等,更无法实现这些资源与开放环境下的网络资源间的关联。此外,非物质文化遗产资源聚合的目的就是保护与传承,并最大限度地为用户提供知识服务,使用户在获取非物质文化遗产资源的同时,也能够汲取资源所包含的文化知识及民族文化精髓,而当前非物质文化遗产资源聚合效果及其服务水平很难满足用户对知识获取的迫切需求,人们虽处于信息海洋中,却经常面临信息匮乏、知识匮乏的困境^[1]。所以,目前非物质文化遗产资源所能实现的聚合广度、深度以及所提供的服务质量并不尽如人意。

近年来,国内外研究人员利用元数据、本体及领域相关主题词表等知识组织工具,面向网络资源开展了有效的组织和利用工作,开启了资源聚合的新篇章,而语义网的发展及关联数据的提出,更是极大促进了这项工作的开展,目的就是实现资源的语义描述、关联、聚合和利用,采用 URI 和 RDF 格式发布、链接各类数据。随着关联数据逐渐成为语义网的研究热点,也鉴于其自身所具有的技术特点,使其能够被应用于资源聚合中。目前关联数据用于资源融合领域主要集中在企业信息融合、金融数据融合、图书馆信息融合等,而通过关联数据组织、聚合和利用非物质文化遗产资源的研究还很少见。因此,有必要引入关联数据的理念,应用关联数据相关技术方法,加强网络环境下非物质文化遗产资源的深度组织、聚合和有效利用,提高其知识服务能力,并不断创新服务内容和面向用户的服务功能,缩小与其他应用领域的差距。同时,也使得非物质文化遗产资源中的知识内容和文化精髓能够通过这种方式进行传承和延续。

2. 研究意义

鉴于非物质文化遗产领域中的资源分布式异构现象,通过现有的资源聚合方式不能解决其语义聚合的难题,且针对资源描述语义化程度低下、聚合程度不高、关联深度与广度不够、服务质量和水平较低等实际情况,本书旨在提出一种基于关联数据的非物质文化遗产资源语义聚合方法,选题具有重要的理论与实际应用价值。

(1) 理论研究意义:①本书系统深入的探索了语义网环境下,基于关联数据的非物质文化遗产资源聚合相关理论,其研究成果对于建立和完善网络环境下非物质文化遗产资源聚合的理论体系具有重要价值,也是非物质文化遗产保护与传承的必然要求;②本书深入分析了元数据技术、本体技术及关联数据技术在非物质文化遗产知识揭示与描述、知识关联与发布、知识聚合与集成、知识可视化展示与语义检索等方面的应用,对建立和完善网络环境下非物质文化遗产资源聚合的方法体系具有较大的借鉴意义;③本书采用了语义网、语义标注、关联数据、资源聚合、知识服务等相关理论与方法,构建了基于关联数据的非物质文化遗产资源语义聚合框架,对促进语义网、关联数据、资源聚合、知识服务等相关理论和方法的发展有一定的推动作用。

(2) 实际应用意义:①本书针对非物质文化遗产数字化保护及资源聚合的现状与存在问题,提出了基于元数据、本体及关联数据的解决思路和方法,该方法可以较为充分的发挥人、资源聚合技术及语义网技术在知识标注、组织、管理和服务上的优势,具有一定的科学性与合理性,能够实现非物质文化遗产资源的描述精确化(基于元数据与本体的语义标注)、深度结构化(明确知识对象、属性及其关系)和语义关联化(多维度揭示资源间的语义关联关系),促进非物质文化遗产资源信息组织向知识组织的转变;②本书针对非物质文化遗产的资源类型及其特征,提出以关联数据的形式对其进行发布、集成与服务,这对提升非物质文化遗产资源聚合效率,改进知识服务模式和质量,拉近与网络用户之间的距离,传承民族文化精髓来说,都起着非常重要的推动作用,将更加凸显非物质文化遗产及其资源保护的重要性;③深度聚合与开发利用非物质文化遗产资源,探索面向用户的非物质文化遗产知识服务与语义检索,以楚剧为应用示范,将该领域内多种类型的信息资源进行整合并建立丰富的语义关联关系,并基于关联数据实现楚剧资源的应用与服务,为实现非物质文化遗产其他领域资源聚合积累了宝贵经验。

1.2 国内外研究现状

国内外研究现状的阐述主要从关联数据及非物质文化遗产资源聚合两个方面展开。首先对关联数据研究现状进行全面概括,其次从数字化技术、理论、方法、应用四个方面对非物质文化遗产资源聚合的研究现状进行了梳理,然后聚焦于基于关联数据的非物质文化遗产资源聚合研究中存在的不足,并提出了本书所研究的科学问题。

1.2.1 关联数据研究现状

国外学者对关联数据的研究主要集中于学术会议、项目研究、语义关联模型、技术工具、实际应用方面,国内则更加关注理论层面的研究。

(1) 国外研究现状

2006年万维网创始人Tim在《关联数据笔记》中第一次提出了关联数据的概念,即将以前没有进行关联的数据链接起来,并构建一个能够被机器所理解的富含语义关系的数据网络(web of data),在此基础上实现更加智能的应用^[2]。同时,Tim又提出了关联数据的四个基本原则,得到了学术界和各应用领域的广泛认同,并针对关联数据创建、发布及应用开展了大量的研究与实践。

学术会议方面。截至2014年9月已经召开了七届LDOW(Linked Data on the Web)研讨会,会议主要围绕关联数据创建与发布、关联关系自动构建、关联数据应用、关联数据融合等方面展开^[3]。此外,2010年召开的DC元数据年会^[4]、2011年的语义网技术大会、DC与元数据应用研讨会等都涉及关联数据研究的多个领域。

项目研究方面。Dbpedia项目从Wikipedia词条里提取出结构化数据,可以将其其他数据集与Wikipedia在数据节点上相链接,开发多种创新应用^[5];OREChem项目^[6]将现有化学类数据源发布到LOD云中,应用于化学领域的知识本体构建;Linking open drug data项目^[7]将不同数据源的医药数据进行关联,并提供相关的医药服务;在农业领域方面,联合国粮农组织(FAO)将AGROVOC叙词表发布为关联数据,并建立了多个词表之间的关联关系,以此作为农业领域关联数据创建与发布的依据^[8];2010年欧盟又启动了LOD2项目,其目标是研究面向企业级的关联数据创建、发布、浏览工具^[9]。

语义关联模型方面。目前已经有一些较为成熟和通用的本体模型(领域本体、规范的词汇集词表等)可以复用,如LOD社区发布的LOD数据云中,收录了多种知名数据集,诸如DBpedia、FOAF、GeoNames、MusicBrainz等,这些数据集涉及地理、医学、媒体、社会网络等众多领域。相关领域本体或者通用词表在构建关联数据语义模型、语义聚合和互操作等方面发挥了重要作用。

技术工具方面。关联数据已经逐渐从理论走向应用,也已经出现了大量关联数据创建与发布的工具,大致可以分为三类:①关系型数据库RDF转化工具。最具有代表性的是D2R,其作用是将关系型数据库中的数据转化为虚拟RDF数据,由于这一关联数据创建与发布工具较容易更新,且数据空间复杂度较低,所以成为一种使用率极高的关联数据创建与发布工具。此类型的关联数据工具还有Triplify,其作用是将关系型数据库发布为真实的RDF数据^[10]。②直接生成RDF数据的关联数据工具。Virtuoso universal server是一种商业级关联数据工具,可以通过一个SPARQL端点将数据转化为RDF数据^[11];SparqPlug则是从HTTP文本中直接抽取关联数据

并以 RDF 格式序列化输出^[12]。③其他发布 RDF 数据的工具。这类工具应用较多的是 Pubby 和 Talis platform。Pubby 能将 URI 请求转化为潜在 RDF 数据查询语言 SPARQ^[13], Talis 则是一个能够提供 RDF 或关联数据存储的软件服务平台^[14]。

应用研究方面。关联数据自 2006 年产生以来得到了广泛的认同和应用,越来越多的组织与个人加强了对关联数据的应用研究,涉及大众传媒、商业企业、政府部门、图书馆等众多领域。Michael 将关联数据的应用分成内容再利用、语义标签、综合提问应答系统以及事件数据管理系统四个方面^[15]。在图书馆领域,作为信息收集、组织与服务的专门性机构,图书馆的书目数据、主题标目等都可以发布为关联数据供用户使用。2010 年 W3C 成立了图书馆关联数据孵化小组,其职责就是为了帮助图书馆创建和发布关联数据,增强图书馆数字资源的互操作。美国国会图书馆则是通过 SKOS 将传统的主题标目转化成 Web 可以理解、处理和使用的形式^[16]。2008 年瑞典国家图书馆将其联合目录 (LIBRIS) 发布为关联数据,并创建了与 lcsb.info 和 Wikipedia/DBpedia 的链接^[17]。在企业应用领域,随着互联网的发展,多媒体资源在网络中占据了较大的比例,如何对多媒体资源进行内容检索成为了一个难点。面对这种困难,Google、BBC 等商业公司都进行了关联数据创建与发布的尝试,如 2011 年 Google 支持并开放了微格式和 RDFa 两种标记结构化数据的标准^[18],BBC 的音乐可被逆向引用至 HTML、RDF、JSON 等,用以获取节目、创建者、艺术家等信息的 URI^[19]。在电子政务领域,美国、澳大利亚、西班牙等国家及其地方政府,都开始尝试以关联数据方式发布政府公开数据,最具影响力的是 2010 年由 Tim 和 Nigel 等创建的英国政府开放数据网站 data.gov.uk。

总体来说,国外对关联数据的研究较为全面、系统,强调关联数据的应用与实践,并产生了一系列的应用项目,成功的实践案例也有很多。

(2) 国内研究现状

国内针对关联数据的研究也取得了较大的进展,其研究成果主要来自图书情报领域,相关学界也积极开展了各领域关联数据的构建与应用。上海图书馆的刘炜等较早关注了关联数据,介绍了关联数据的基本概念以及国外的研究与应用现状^[20]。通过对现有文献的分析,国内关联数据的研究重点主要集中在关联数据创建与发布方法、关联数据应用两个层面。

关联数据如何进行创建与发布,目前并没有指南性的文件,但在发布过程中大多都遵守了关联数据的四项基本原则。学者们对关联数据发布方式的研究,一般都是通过构建理论模型并借助相关的工具予以实现。沈志宏、刘筱敏等以科技文献、科学数据为例,以关联数据的发布技术框架为研究对象,规范了关联数据发布流程^[21];邓兰兰、李春旺提出适用于关联数据发布的相似度计算综合描述方法,在一定程度上解决了当前关联创建中相关资源集无法实现自动化的问题^[22];张春景等探讨了关联数据在发布、消费和再创造过程中涉及的相关协议问题,如知识共享协议、开放数据许可、

开放政府许可协议等,并对这些协议的应用范围提出了建议^[23];陶俊等基于 R2R 语言的特点,从词汇与数据实例两个层面提升了 RDF 链接创建能力^[24];中国科学技术信息研究所的相关成员,对关联数据构建原则、关联关系自动构建、语义关联模型、语义模式映射问题进行了深入研究,初步实现了书目关联数据的创建与发布^[25,26];钱国富采用关联数据标准来发布政府数据,以此构建一个政府数据本体^[27];王忠义等则从数字图书馆资源粒度出发,将关联数据划分为浅层关联数据、中层关联数据及深层关联数据,并探讨了这三种不同粒度的关联数据如何进行创建与发布^[28]。

有关关联数据应用方面的研究较多,取得的成果也较为丰富。许多学者从理论出发,分析了关联数据在具体领域的应用现状、存在的问题以及今后的发展方向,如沈志宏通过描述关联数据的原则及技术体系,介绍了关联数据的多种典型应用,提出在具体构建关联数据时所需要面对的问题^[29]。他同样以关联数据互联技术为研究对象,深入分析了关联数据的互联技术下一阶段的发展方向^[30];刘炜对目前采用大数据技术发布关联数据的方法和途径作出了梳理,也对关联数据在大数据领域的应用前景进行了规划^[31];黄永文等通过介绍关联数据开展的一些活动,提出关联数据可以应用于用户信息检索服务中的多个方面,并从多个层面分析了关联数据应用于信息检索服务的相关技术^[32];李亚婷等认为 Web 环境下的关联数据应用主要集中在两个层面,即语义标签和 Web 资源集成^[33];王思丽等则以中国科学院机构知识库平台 CASIR 为例,研究了利用关联数据实现机构知识库语义扩展的相关技术,为机构知识库从基础服务到语义集成服务的转变奠定了基础^[34];武汉大学的马费成教授还基于关联数据设计实现了依托武汉大学免费网络学术资源的集成实验系统^[35],这是关联数据在高校应用中的一次成功实践。无论是关联数据在信息检索、知识服务、知识库构建还是在语义网、大数据环境下的各类应用,以及所研究出的创建模式、发布工具,都为关联数据的普及提供了很大的帮助。

关联数据应用最为广泛的,应该还是在图书馆领域。图书馆是信息获取、整理、分析、组织、服务的专门性机构,关联数据可以在图书馆资源建设的多个环节中得到一定程度的应用:①将图书馆资源发布为关联数据。目前在图书馆领域,已经实现了将 SKOS 系统(分类表、词表)、书目数据等发布为关联数据。郑征总结了当前图书馆领域常用的关联词表^[36];刘炜、胡小菁、钱国富等研究了以 RDA 编目的书目数据以关联形式进行发布的优势及特点,并对其发布过程进行了详细描述^[37];黄永文则系统提及了数字图书馆各领域的关联数据创建与发布状况^[38]。②基于关联数据实现图书馆资源组织与语义检索。南京大学欧石燕教授提出了基于本体与关联数据的数字图书馆资源语义聚合与测评方法,通过构建面向关联数据的语义数字图书馆资源描述与组织框架,实现了一个数字图书馆资源整合演示系统^[39,40];白海燕等提出基于本体构建语义关联关系,并通过关联数据的语义描述和统一存储对书目数据进行语义化组织^[41];赵悦则是在比较元数据与关联数据基础上,以书目数据、

知识组织系统数据及规范词表为基础,梳理关联数据在图书馆书目组织中的应用情况^[42]。③利用关联数据扩展资源知识发现及知识服务。关联数据允许图书馆与网络中更为广泛的信息资源进行关联,以达到知识发现的作用,从而为用户提供更为高效的知識服务。李楠等通过分析知识发现与关联数据间的关系,构建了基于关联数据的知识发现模型,并对其应用体系做出了探讨^[43,44];钟远薪等介绍了OPAC混搭关联数据的方法及应用情况,指出这种方法旨在建立用户所提供的信息间的关联,以此提高图书馆资源发现的概率^[45]。

与国外研究相比,国内对关联数据的研究还处于不断发展的阶段,关联数据的应用并不十分广泛,在普通用户中也并未普及,且已开展的研究中侧重于对关联数据概念、原则、创建与发布流程以及应用现状的介绍与综述,对其进行实际构建的尝试并不多,而在非物质文化遗产领域的应用几乎没有提及。

综上所述,国内外对关联数据的研究目前虽然已经取得了一些成果,在应用实践方面也积累了一定的经验,但也存在着一些不足,主要包括以下几点。

第一,在理论研究方面。虽然大多数研究都涉及关联数据创建与发布的过程,但一般在集中讨论其发布方式和发布工具,而对关联数据如何创建并未做出深入解释。即使在构建关联数据语义关系模型时,大多数研究只考虑了资源的外部特征,如题名、创建者、时间、地点等,缺乏对资源内容本身特征以及核心属性的描述,没有从语义层面揭示资源中概念与概念间的关系,关联层次不够。

第二,在应用实践方面。通过对已发布的关联数据集的梳理,可以看出,大部分数据资源的类型较为分散,资源间无法实现有效兼容和语义互操作。尤其针对特定的研究领域,其资源的多样性、异构性无法得到体现,更无法实现专业领域资源的多维度聚合和数据关联。

第三,大多数关联数据的创建与发布是以实现资源有效整合为目标,但这并不是其最终目标。资源聚合的最终目标是实现资源的共建共享,为用户提供更合理、高效的服务。所以,关联数据的创建与发布也必须是以用户需求为导向,但在非物质文化遗产资源组织的实际过程中,关联数据的相关理念、技术与方法并没有得到应用,也缺乏基于关联数据的资源聚合与服务平台的构建。

1.2.2 非物质文化遗产资源聚合研究现状

非物质文化遗产资源聚合的目的就是要对非物质文化遗产资源进行保护、传承和利用,由于非物质文化遗产所具有的无形性、传承性、社会性及多元性,使其资源在聚合与保护的 actual 过程中存在很大的困难^[46]。由于非物质文化遗产资源及其聚合的概念并没有被明确定义,相关研究也较少,所以针对非物质文化遗产资源聚合的研究,是在资源聚合研究的基础上,结合非物质文化遗产领域及其资源的特点加以归纳。

通过对现有资料的阅读和代表性文献的调研,非物质文化遗产资源聚合的研究主要围绕四个方面展开。

(1) 非物质文化遗产资源数字化技术研究

随着计算机的广泛应用和网络技术的迅速发展,信息数字化深刻影响着社会发展的各个方面,数字资源已经逐渐成为社会核心资源之一,数字化技术、网络技术的数字化,也已经成为实现非物质文化遗产资源聚合的重要途径。

早在20世纪70年代,各文化管理部门和机构单位都加强了计算机技术进行文化信息保护的研究和探索。在国外,1992年所开展的世界记忆项目、2004年的数字化保护丝绸之路文化遗产项目、2004年的世界非物质文化遗产保护研究项目、2005年中国丽江的世界记忆项目都表现出数字化技术的重要程度。在国内,对非物质文化遗产的数字化保护研究主要有浙江大学国家重点实验室的民间表演艺术的数字化抢救保护与开发的关键技术研究、浙江大学计算机学院现代工业设计研究所的楚文化编钟乐舞数字化技术研究和云南斑铜工艺品数字化辅助设计系统,以及南京大学的三峡文化遗产数字化等^[47]。此外,各大数字博物馆也为我国利用计算机及多媒体技术实现对濒危非物质文化遗产保护、传承与再创造提供了大量帮助,如中国国际友谊博物馆工程、南京博物馆数字化、山东大学考古数字博物馆项目等^[48]。这些博物馆对非物质文化遗产资源数字化项目的探索,为实现非物质文化遗产资源聚合提供积累了有益的方法及技术经验。

现阶段,我国非物质文化遗产领域已经积累了多种资源数字化保护方法,如数字化采集和存储技术、数字化复原和再现技术、数字化展示与传播技术、虚拟现实技术等。通过这些最新的数字图像和互联网技术的综合运用,将非物质文化遗产资源进行整理、归类,并通过数字化技术进行记录、编辑和管理,使人们能够对无形的非物质文化遗产有清晰的认识和了解,能够不受时间和空间的限制,感受到非物质文化遗产传承的魅力,也使得非物质文化遗产资源聚合工作能够有效的开展^[49]。

(2) 非物质文化遗产资源聚合理论研究

国内对非物质文化遗产资源聚合理论的研究主要集中在图书情报、计算机科学以及文化文艺三个领域,主要涉及非物质文化遗产资源聚合相关概念、理论模型、建议对策等方面。

① 相关概念研究

非物质文化遗产资源聚合从实质上来说,就是对非物质文化遗产资源的聚集与整合,其最终目的就是为了资源的保护、传承和利用。而对非物质文化遗产、非物质文化遗产资源、非物质文化遗产资源聚合的内涵方面的理解,还存在着一定的分歧。杨怡指出研究非物质文化遗产领域尚存的几个问题,如定义、物质载体、明确的标准规范等^[50];贺军学对非物质文化遗产资源聚合及其保护的本质、原则、主体

价值进行了理论探讨;龙红则是通过对非物质文化遗产概念用语的认识过程,提出非物质文化遗产三个层面的认知^[51]。

② 理论模型研究

在资源聚合理论模型方面,图书馆学界有着较为丰富的研究成果可供非物质文化遗产领域借鉴,其研究重点集中在宏观和微观两层面,宏观层面是对资源聚合模式、模型及相关策略的归纳,微观层面则是从知识粒度出发,对资源聚合所涉及的知识结构进行全面研究。

有学者从资源聚合的主体出发研究网络资源聚合相关策略,如李枫林、胡昌平提出以用户为导向的网络信息资源整合,是合理利用信息资源的最佳途径^[52];邓胜利将网络信息资源整合的目标确定为利用各信息机构的互补性实现跨机构的资源集成,并制定了以网络信息资源、信息技术、信息机构和信息人才为要素的信息整合策略^[53]。有的学者则关注于资源本身的融合,如邱均平在分析图书馆资源语义化方法存在问题的基础上,构建了基于计量分析方法的馆藏资源语义化模型,以实现馆藏资源的深度聚合^[54];黄传慧、李娟在分析了基于OPAC、基于跨库检索、基于资源导航、基于超级链接、基于知识本体这五种数字资源整合模式后,提出基于本体的数字资源整合是未来资源整合模式的发展趋势^[55]。随着语义网的普及和信息技术的发展,基于语义的资源聚合逐渐成为研究重点,贺德方对当前资源整合工具及国内外研究现状进行了概述,提出了三种基于语义的馆藏资源整合方式,同时构建了实现资源聚合服务所需要的语义描述框架^[56];牟冬梅的研究明确指出数字图书馆知识组织语义互联成为其资源聚合必须解决的关键性问题,并构建了语义互联的应用模型,具有一定的应用价值^[57]。

随着多源异构数据聚合研究的深入,人们已经不满足对异构数据在逻辑及物理上的简单集成和封装,常规的数据集成已经不能满足用户在语义层面的知识需求,人们更关心的资源聚合之后数据所能展现出的语义含义,那么就必须扩展资源聚合的深度,从知识节点、知识单元等微观层面对资源聚合开展更为深入的研究。赵蓉英对知识网络的概念、类型、结构等基本问题进行了系统地研究,并探讨了知识网络的结构组成,重点对知识节点及其类型进行了深入分析^[58,59];文庭孝提出可以用网络广度、深度及关联强度等指标来测量知识网络中网络节点间的关联关系^[60,61];邱均平针对图书馆文献资源,从文献特征关联、利用过程关联、知识关联及用户需求关联四个维度探讨了馆藏资源多维度聚合模型,串联了资源与用户间的多维关联路径,从多个层面揭示了资源中概念与概念之间的关联关系^[62];王忠义等借助多阶段知识发现理论,提出一种数据集内的关联关系知识发现方法^[63];曾建勋则认为知识链接是资源聚合的基础,下一步的知识链接将朝着开放式跨界链接、基于用户行为的个性化推荐链接、知识要素链接、知识组织链接、知识要素链接以及语义关联链等方面发展^[64]。