

图书馆馆藏元数据发展策略分析

张茹

(太原市图书馆, 太原 030024)

摘要: 图书馆作为元数据生产与提供的重要机构, 在元数据制订、管理及服务中积累了丰富的经验。本文首先探讨图书馆馆藏元数据可能产生的价值, 基于此对当前图书馆元数据发展存在的问题进行分析, 认为其固有的管理观念、方式及手段一定程度上阻碍了元数据作用的有效发挥; 最后针对性地提出包含元数据资产管理、质量、服务等在内的发展策略, 旨在保障元数据在未来应用中能够被发现、获取、互操作及重用。

关键词: 馆藏元数据; 元数据策略; 图书馆

中图分类号: G254 DOI: 10.3772/j.issn.1673-2286.2021.04.010

引文格式: 张茹. 图书馆馆藏元数据发展策略分析[J]. 数字图书馆论坛, 2021 (4) : 68-72.

元数据是描述资源特征的结构化数据, 发挥着定位、查找、发现资源的基本功能。图书馆作为馆藏元数据的生产与提供的重要机构, 在元数据制订、管理及服务中积累了丰富的经验。馆藏元数据作为图书馆关键资产, 推动着机构内外部数据获取、交换及重用, 通过数据深层次开发与再利用, 提升了图书馆信息服务能力。

我国图书馆发展经历了从重视资源建设到提升服务水平两个主要阶段^[1]。资源建设强调的是资源的规模化及数字化, 旨在通过丰富馆藏资源数量及其便利的信息技术手段, 提升资源的可获取性。当资源达到一定规模后, 如何提升图书馆服务能力及水平成为图书馆的关注焦点, 其中学科、参考咨询、个性化、知识服务等成为此阶段的研究热点。但服务的开展始终围绕着资源且受限于资源的开发利用程度。馆藏元数据记录作为资源的替代品, 其质量水平也成为制约图书馆发展的重要因素。网络开放性及用户需求广泛化推动了图书馆与外部数据交换频繁, 元数据作为主要交换载体, 推动着图书馆馆藏数据共享、重组、集成。国外图书馆界已意识到元数据的重要性, 纷纷制订了元数据发展战略。英国国家图书馆制订了《2019—2023年大英图书馆馆藏元数据发展战略》, 通过为馆藏元数据实施统一的系统及标准基础设施, 旨在达到开发利用数据资

产、资源发现及其广泛社区利用目的^[2]。OCLC于2020年发布了《下一代元数据转变》报告, 提出元数据应适应关联数据环境, 从定义标识符、馆藏类型多样化、元数据服务角度提出转变思路^[3]。正值我国图书馆启动“十四五”战略规划编制之际, 只有将元数据纳入到图书馆长期发展规划中, 在图书馆内部达成共识, 才能推动元数据在未来发展中发挥其关键作用。

1 图书馆馆藏元数据的发展价值

图书馆馆藏元数据是馆藏的数字表现形式, 反映着馆藏整体水平, 包括书目数据库、商业数据库及机构库、学位论文库、特藏库等自建库。馆藏元数据是图书馆元数据的核心, 从元数据创建到使用, 形成图书馆关键业务流程, 支撑着图书馆各项服务功能。

1.1 元数据已上升为重要数据基础设施

图书馆作为知识中心, 收藏着人类各种宝贵的文化知识, 担负着知识加工与利用的重任。随着数字化技术与方法的广泛应用, 机器存储及处理数据能力增强, 以元数据形式为核心的馆藏数据, 将融入以数据驱动为核心的数字化经济中, 全面服务于用户学习、生产、

研发、管理、决策等各个环节,实现社会全面创新及科学化发展。元数据作为重要数据基础设施,支撑着图书馆内外部的数据交换,图书馆元数据建设成为一项长期持续发展的建设项目。图书馆应着眼于建设标准化、开放性、持续性的数据基础设施,旨在推动数据开放互联和知识融通^[4]。通过标准性表示及开放性存取,将元数据作为可开发利用的知识宝藏,嵌入社会各领域各个层面的应用中,以发挥图书馆的社会功能。

1.2 元数据是智慧图书馆发展的基石

从实体图书馆、数字图书馆到智慧图书馆的发展,旨在不断提升图书馆知识服务能力,将分布异构的海量信息资源进行全面高效的知识发现与获取、组织与整合、开发与利用^[5]。智慧图书馆有机地将人与资源、资源与资源、人与人之间建立联系^[6],通过自动感知、情境相关、语义计算实现高度融合共享的图书馆服务。智慧图书馆的实现依赖于针对多源数据的元数据采集、生成、聚类、关联、集成、重组等过程,通过对资源及用户的数据进行细颗粒度表示及其计算,才会无处不在地按需提供适合的服务。实际上,用户特征提取、情景分析、资源推荐、与各类应用的数据交换都是借助于元数据生成及互操作方法,将馆藏资源组织成知识的形态,主动参与到用户知识交流过程,实现知识的探索、发现、传播、交流与创造^[7]。

1.3 元数据是图书馆服务质量提升的保障

图书馆服务开展需要在充分了解用户需求的基础上,提供针对性的信息内容。用户需求的分析及其馆藏资源的深度开发与利用,都需对元数据进行一系列处理,通过对用户及资源不同程度的描述,形成描述元数据、管理元数据、结构元数据及使用元数据。对用户数据及馆藏资源元数据进行匹配的基础上,实现馆藏元数据的聚类、整合、连接,支持浏览检索、咨询、推荐、专题服务。从信息服务到知识服务,从集成服务到个性化服务,很大程度上取决于元数据建设水平及其馆员的元数据素养。

2 图书馆馆藏元数据管理所存在的问题

从元数据标准选取、元数据记录创建、维护到集

成应用,馆藏元数据经过一系列处理流程以实现馆藏资源的管理与应用。元数据管理贯穿于图书馆各环节中,高质量的馆藏资源取决于元数据管理水平。从我国当前图书馆元数据发展水平看,固有的管理观念、方法及手段一定程度上阻碍了元数据作用的有效发挥。

2.1 元数据管理意识较为薄弱

编目工作外包导致编目业务逐渐弱化,图书馆编目人员数量减少。但随着不同来源数字资源出现,资源共享及集成需求不断增加,围绕元数据生成及合并活动频繁,外部信息环境对元数据管理提出了更高要求,涉及新技术使用、元数据质量评估、元数据再利用等。图书馆管理者不仅要意识到元数据的价值,还应设置专门的管理部门。当前国外许多图书馆设置了如元数据馆员等新兴岗位^[8],并开展馆员元数据素养教育,而当前国内的馆员只熟悉传统的编目环境,不了解图书馆内部的元数据类型,没有认识到元数据的重要作用,缺乏元数据管理意识,对元数据的创建、合并、维护到更新全流程管理所涉及的技术及工具熟悉度不高。据相关统计,仅有20%的馆员完全了解图书馆数据,25%的馆员掌握数据处理技能^[9],整体表现出元数据素养能力低下,馆员元数据素养提升已成为图书馆目前发展的紧迫任务。

2.2 缺乏有效的元数据管理手段及方法

图书馆积累了丰富的针对实体馆藏资源的管理手段及方法。随着资源数字化处理技术的成熟,纸质资源已不再是馆藏核心,数字资源成为支持服务开展的重要资源形态。数据资源的管理、开发及利用都在元数据层面开展,需要图书馆管理人员充分了解数字资源的特征,掌握新型的元数据管理方法。但当前图书馆依然沿用纸质资源的管理手段,注重资源的采购,忽视资源的二次加工,重视资源本身的外部形态管理,忽视资源内容层面的描述与揭示。资源描述基于传统的编目管理思想,以单件为核心所形成的扁平化描述方式,难以表述资源之间、资源内容之间的有机联系,阻碍了其与外部数据的集成应用。随着所采集的数字资源数量及种类增加,图书馆管理人员并未意识到对元数据进行统一管理及其深层加工,从而导致图书馆内部的元

数据处于分散的、独立存储于不同系统中的现实状况,数据孤岛严重,缺乏对馆藏资源元数据大规模的统一集成整合,影响数据重用及共享。据统计,各个高校图书馆平均拥有20个以上数据库,但近90%的图书馆缺乏有效地资源整合途径,用户不能充分实现跨库、跨类型的一站式检索^[10]。

2.3 元数据质量管理水平不高

图书馆书目元数据主要由书商提供,多采用简要级别的数据描述,元数据整体质量表现良莠不齐,数据的一致性、规范性、完整性相对较低^[11],尤其以MARC元数据标准所形成的书目记录,格式复杂,大部分图书馆MARC数据尚未转换成XML等通用形式,与其他数据进行交换具有一定的难度。当前图书馆缺乏明确的元数据质量评估指标及评估方法,忽视元数据的重用性,导致我国图书馆元数据构建成本高,利用率低,闲置于图书馆内部,多次开发利用少,并未发挥其应有的数据价值。

3 图书馆馆藏元数据的发展策略

当前语义网及数据分析处理技术的发展,推动了大范围的数据集成应用。馆藏元数据作为图书馆的关键资产,意味着需与图书馆外部资源进行深度整合,将馆藏元数据有机融入更丰富的情境中,以发挥图书馆的社会服务功能。图书馆馆藏元数据发展策略是推动图书馆可持续发展的基础。建立统一、标准化的元数据基础设施,提升元数据质量水平,实现元数据访问与协作、开放与重用,通过图书馆元数据价值的发挥,以进一步提升图书馆业务效率,深化服务内容。

3.1 馆藏元数据资产管理策略

资产强调新价值的产生与创造。图书馆投入一定成本,依赖馆员的智慧,通过对资源对象加工创建形成元数据,以确保其在元数据再利用中产生新的价值。基于元数据重用性,Greenberg首次提出元数据资产概念^[12],认为高质量的元数据未来不断地被再利用,会产生更大的价值。元数据资产管理策略旨在使图书馆将元数据上升为重要资产,采用科学有效的管理手段与方法,实现元数据的持续增值。确定元数据资产管理的内容

及范围,形成元数据资产认定和登记机制,制订相关制度及规范;构建中心式的元数据资产管理体系,对元数据统一管理;实现对元数据标准的管理,包含标准的需求分析、制订、实施、维护及评估;规划元数据全生命周期,形成从创建、集成到应用的全链条管理体系;明确元数据共享开放策略,涉及数据共享方式、数据开放协议与端口、数据安全及交换,推动元数据资产价值流通;评估元数据资产价值,制订价值评估体系,从应用对象、范围、使用频度、效果等方面入手实施评估,促进资产价值最大化。

3.2 馆藏元数据质量管理策略

OCLC焦点小组认为,图书馆的战略目标如促进发现和使用、丰富用户体验、支持整个学术生命周期都是基于高质量元数据而实现的^[3]。元数据质量战略旨在通过提升元数据质量水平,使元数据具有良好的互操作及可重用性,带动深层次开发与利用,实现元数据资产效益最大化。元数据质量持续提升依赖于科学的数据质量管理方法及其闭环式质量改进流程,从质量定义、质量评价、质量问题分析到质量优化实现全面质量管理。质量定义应明确高质量的馆藏元数据的评估指标,应适用于当前数据网络环境发展,包含资源描述采用高流行度的元数据标准,尤其选用内容标准,确保每个元素的含义清晰明确;元数据记录颗粒度细,数据之间建立丰富的语义关系;元数据具有可识别性,拥有永久标识符;具有一致性的数据模型,并采用XML、RDF语言进行形式化描述;数据开放度高,提供数据开放端口,确保用户多途径获取到各种格式的元数据^[15]。图书馆应定期对元数据质量进行评估,并进行质量问题分析,发布元数据质量评估报告,确定质量提升目标,以不断提升图书馆元数据质量水平。

3.3 元数据服务策略

元数据服务策略旨在发展以元数据为核心的服务,包含元数据创建及维护、元数据注册、数据仓储^[13]、元数据素养培训及教育等服务,通过对多种来源的元数据集中管理及重用推动元数据的跨社区应用。元数据创建及维护服务旨在构建成熟稳定的基础服务设施,以实现元数据的高质量生成及管理。明确元数据管理规章制度,规范元数据创建及维护处理流程,开发元数

据管理系统及技术,实现对元数据高效率地管理。元数据注册服务利用注册管理平台实现对元数据的集中存储和控制,促进元数据的发现、重用、管理。构建图书馆领域的术语服务系统,以支持对领域内术语的注册、更新、发布及其使用。通过用户管理、元数据上传、下载、发布和维护等功能实现数据的统一更新与管理,方便图书馆内外部用户获取到元数据并进行再利用。数据仓储服务旨在将不同来源的元数据记录集中在数据仓储中心,统一向外提供元数据记录服务。包括采用一定的数据收割机制将来自不同来源的数据采集,不同格式数据的去重、整合及语义关联,提供对外数据获取接口,以实现元数据提供服务,通过一次生产,多次利用,以实现馆藏元数据再利用。元数据素养培训及教育服务旨在培养馆员及其科研工作者的元数据素养能力,提升信息环境下创建、使用、维护元数据的能力。要求馆员熟悉各种类型的元数据标准,掌握元数据标准设计方法,熟悉各种类型的元数据处理工具,能够高质量地创建元数据记录,实施元数据互操作。

4 结语

FAIR原则的发布推动了开放科学的进程,旨在提升机器自动发现和使用数据的能力^[14]。图书馆馆藏元数据策略与FAIR原则保持一致,强调元数据的可发现、可获取、互操作及重用性,支撑着数字图书馆向智慧图书馆的转变,推动着图书馆可持续发展。图书馆需要充分认识到元数据的发展价值,从资产管理、质量、服务三个层面入手,全面提升图书馆元数据管理水平及效率。图书馆馆藏元数据发展受限于许多因素,如图书馆长短期目标、管理方法、技术力量、人员素质等,不同类型的图书馆也会因定位差异,馆藏元数据发展策略侧重点有所不同,如单个图书馆侧重于元数据的资产管理、质量管理策略,而国家图书馆、联盟图书馆更侧重于元数据服务策略。而就元数据服务而言,单个图书馆的元数据创建及维护策略是重点,数据仓储服务及元数据注册服务更青睐于技术力量较强的图书馆。因此后期研究中,将深入分析各个不同服务策略实施的目标、条件及其适用类型,以推动不同图书馆联合,共同提升元数据服务能力。

参考文献

- [1] 吴瑾,刘德偲,王磊,等.从资源驱动走向服务主导——2017年全国图书馆新型服务能力建设学术研讨会综述[J].图书情报工作,2017,61(13):133-139.
- [2] The British Library. Foundations for the Future: The British Library's New Collection Metadata Strategy 2019-2023 [EB/OL]. (2019-04-12) [2021-02-02]. <https://blogs.bl.uk/digital-scholarship/2019/04/the-british-librarys-new-collection-metadata-strategy.html>.
- [3] SMITH-YOSHIMURA K. Transitioning to the Next Generation of Metadata Transitioning of Metadata [R/OL]. (2020-09-01) [2021-02-01]. <https://www.oclc.org/content/dam/research/publications/2020/oclcresearch-transitioning-next-generation-metadata.pdf>.
- [4] 张静,姜永常.知识构建:数字图书馆知识服务能力的根本保障[J].情报理论与实践,2010,33(9):28-31.
- [5] 夏翠娟.面向人文研究的“数据基础设施”建设——试论图书馆学对数字人文的方法论贡献[J].中国图书馆学报,2020,46(3):24-37.
- [6] 王世伟.再论智慧图书馆[J].图书馆杂志,2012,31(11):2-7.
- [7] 谢蓉,刘炜. SoLoMo与智慧图书馆[J].大学图书馆学报,2012(3):5-10,79.
- [8] 贾延霞,赵秀君.编目员的未来:编目员的职责拓展研究[J].图书馆建设,2011(7):31-33,37.
- [9] 廖利.大数据背景下高校图书馆员数据素养调查与分析[J].宁波教育学院学报,2017,19(4):105-108.
- [10] 毕忠东,周敬治.高校图书馆数字资源建设与整合现状调查与分析[J].情报资料工作,2009(5):54-56.
- [11] 卜书庆.与图书馆共存亡的元数据编制工作[J].国家图书馆学刊,2015,24(6):3-6.
- [12] Greenberg, J. Metadata capital via a linked data HIVE [C/OL]. [2021-02-10]. <https://journals.lib.washington.edu/index.php/acro/article/view/14680/12320>.
- [13] ZENG M L, QIN J. Metadata [M]. New York: Neal-Schuman Publishers, 2014.
- [14] WILKINSON M D, DUMONTIER M, MONS B. The FAIR Guiding Principles for scientific data management and stewardship [J/OL]. [2021-02-02]. <https://www.nature.com/articles/sdata201618/>.
- [15] 贾君枝.词表生态系统:构成要素及关联关系[J].中国图书馆学报,2020,46(4):60-68.

作者简介

张茹, 女, 1969年生, 本科, 馆员, 研究方向: 数字资源, E-mail: 13934524988@163.com。

The Analysis of the Development of Collection Metadata Strategy in Library

ZHANG Ru

(Taiyuan City Public Library, Taiyuan 030024, China)

Abstract: As an important organization of metadata production and supply, library has accumulated rich experience in metadata creation, management and service. This paper first discusses the possible value of library metadata and analyzes the issues faced by the current development of library metadata, and considers that traditional management ideas, ways and means hinder the effective play of metadata. Finally, the metadata strategy including metadata asset management, quality and service is proposed to enhance the value of metadata and ensure that metadata can be findable, accessible, interoperable, and reusable in future applications.

Keywords: Collection Metadata; Metadata Strategy; Library

(收稿日期: 2021-03-05)

2021年第十二届全国知识组织与知识链接 学术交流会征文通知

党的十九届五中全会通过的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》中明确提出“构建国家科研论文和科技信息高端交流平台”的重要任务,对科技信息工作提出明确要求。在此新阶段下,亟需加强知识组织与知识链接的新理念、新思路、新方法的研究,以推动我国科技情报工作适应科技自立自强总体要求,促进科技信息资源开放融合,推进国家数字资源保障体系重构。为此,中国科学技术信息研究所、国家图书馆、国家科技图书文献中心和中国科学技术情报学会拟于2021年9—10月召开第十二届“全国知识组织与知识链接学术交流会”。现面向全国图书情报界及相关领域的专家、学者征文。具体会议时间、地点另行通知。

一、征文主题

1. 知识组织理论与方法研究进展
2. 知识组织系统构建与应用
3. 智能环境下叙词表、本体编制
4. 元数据管理及其标准
5. 大数据清洗分析与资源组织
6. 智慧图书馆资源发现系统
7. 科学评价与学科分析
8. 计算型情报分析与突破性技术预测
9. 知识图谱及可视化技术在图书馆的应用
10. 其他

欢迎广大图书馆学、情报学、出版传播相关教研人员,以及相关信息组织技术人员踊跃投稿。会议将评选优秀论文,并由会议主办方颁发优秀论文荣誉证书。优秀论文将推荐到相关期刊发表。

二、征文要求

1. 文章要求: 观点明确、主题突出; 必须为未经发表的论文; 以电子邮件方式提供Word格式文档; 正文字数应在5 000字以上。
2. 来稿请提供: 中英文题目、中英文作者及单位、中英文摘要和关键词、正文、参考文献; 文后请附作者简介, 包括联系电话、电子邮箱、通信地址及邮政编码等。
3. 截稿日期: 2021年8月31日。

三、联系方式

联系人: 宋官钰、喻菲 联系电话: 010-58882324; 010-88545322 征文邮箱: KOLink@istic.ac.cn