

文献资源组织中元数据集成建设 问题研究*

杨永清 黄毕惠

(四川大学图书馆, 成都 610065)

摘要: 海量数字资源的出版和信息环境的巨变给图书馆传统的文献资源组织方式带来了极大挑战, 传统编目生成的粗粒度书目数据正在丧失文献组织与检索的主导地位, 细粒度、语义化、规模化、关联化的文献元数据集成管理对图书馆的资源发现和智慧服务具有重要意义。文献服务机构需要重视文献元数据体系建设, 对文献元数据进行结构化、系统化、关联化处理, 注重元数据统一集成管理中的元数据收集、元数据标准制定、元数据整合、元数据协同共建与共享, 构建国家层面的“大数据”级元数据集成体系, 才能释放和提升文献资源价值, 满足用户资源发现、知识关联和智慧服务的新需求。

关键词: 文献元数据; 元数据集成管理; 资源组织; 资源发现; 知识服务

中图分类号: G250 **DOI:** 10.3772/j.issn.1673-2286.2020.09.004

引文格式: 杨永清, 黄毕惠. 文献资源组织中元数据集成建设问题研究[J]. 数字图书馆论坛, 2020 (9): 24-31.

随着信息技术的迅猛发展和信息环境的巨大改变, 图书馆的文献资源类型不断扩展, 纸本资源、多载体资源、数据库资源、网络资源、开放资源等层出不穷, 呈现出数量庞大、来源多样、结构各异、种类繁多等特点, 给用户快速、准确地查找和利用资源造成了困扰, 也使习惯于实体资源组织的图书馆对数字环境下涌现出来的巨量数字资源整合深感无力。大量隐性资源、闲置资源不仅造成资源浪费, 还难以满足用户获取、关联、挖掘、重组文献资源中蕴含的知识资源的迫切需求。资源发现和知识服务都依赖于海量资源的合理类聚和深度揭示, 而此种资源组织方式必须有元数据做最基础的支撑, 可以说, 元数据是图书馆提供资源获取和知识服务的基础, 是图书馆管理和服务乃至生存与发展的重要资本。基于此, 文献元数据体系建设实践蓬勃兴起, 欧洲数字图书馆于2011年设计了《Europeana数据模型》^[1]; 2012年美国数字公共图书馆提出了《元数据应用纲要》, 2015年更新了强化关联数据的MAP4.0版^[2]; 大英图书馆于2015年发布了《释放价值: 大英图书馆2015—2018年馆藏元数据战略》^[3], 图书馆

界正纷纷加快设计和实施元数据建设规划。

1 文献元数据

元数据是关于数据的数据, 是描述信息资源或数据的一种结构化的数据。图书馆元数据种类丰富, 包含描述元数据、语义元数据、技术元数据、业务元数据、管理元数据等^[4]。文献元数据是图书馆元数据的重要组成部分, 它是以文献资源为对象, 用于实现文献资源的描述、解释、定位、存取、发现、评估、选择、关联、挖掘等功能, 使文献资源更易管理、检索和利用的结构化信息。简言之, 文献元数据能够描述文献资源本身的特征和属性, 支持文献资源的组织和整合, 实现文献资源的检索利用、数据挖掘、知识发现和关联。图书馆文献元数据建设的目标, 是将海量、多源、异构的资源元数据通过规范整理、分类聚合、组织保存等手段整合在一起, 统一实施元数据生命全周期的集中管理和利用, 以改进元数据整合集成效果, 释放隐藏元数据, 提高资源发现效率, 推动资源组织向知识组织转型。

*本研究得到国家社会科学基金项目“开放共享背景下图书馆书目元数据组织模式的演变与发展研究”(编号: 18BTQ026)资助。

2 文献元数据集成建设研究概况

2.1 国外研究概况

进入21世纪, 国外的文献元数据集成研究与实践日益蓬勃。在元数据战略规划方面, 美国数字公共图书馆(DPLA)在其元数据政策声明中提出通过开发和创建元数据集成库和集成工具, 开展元数据收集、描述、索引、数据增强等工作, 保障元数据的共享与重用^[5]; 大英图书馆在其馆藏元数据战略中提出通过建设丰富、一致、权威性、持续性的元数据资源体系, 将馆藏资源与社会资源关联起来, 提高资源使用效率^[3]。在元数据体系框架方面, 奥地利学者Ibrahim等^[6]设计了元数据集成体系结构和系统组件; Bansal等^[7]提出了用于大数据集成的语义提取、转换和加载框架; Lemos等^[8]比较分析了多个网络多媒体知识组织系统及系统框架。在元数据标准方面, 美国国会图书馆的合作编目计划(PCC)致力于元数据描述标准的制定、名称规范档和主题规范档(主题词、分类法、叙词表等)的创建与修订^[9]; ProQuest公司的Summon系统制定了统一的元数据描述框架, 将所有的资源类型映射到Summon统一架构下, 实现各种类型资源的同步发现与揭示^[10]。在元数据集成建设实践方面, 有资源发现系统商和免费共享平台两大主流, 以Summon、Primo、EDS为代表的资源发现系统商, 每个系统内集成的各层级文献元数据规模均是数以亿计。免费共享平台以国际区域联盟和各国国家图书馆的文献资源集成共享平台为代表, 欧盟委员会的Europeana、美国数字公共图书馆的DPLA、澳大利亚国家图书馆的Trove、新西兰国家图书馆的DigitalNZ、日本国立国会图书馆的NDL等, 这些平台制定了元数据的模型与标准, 如Europeana的数据模型(EDM)、NDL的国立国会图书馆都柏林核心元数据描述(DC-NDL)^[11]等, 并整合了各自国家图书馆、博物馆、档案馆等文献机构的数字资源, 为公众提供文献资源的广泛使用和共享。

2.2 国内研究概况

随着数字资源数量激增, 我国也在积极探索文献资源整合及文献元数据集成建设, 研究内容涉及元数据整合平台、系统架构、集成实践等多个层面。在发现系统资源整合研究方面, 窦天芳等^[12]在分析资源发现

系统逻辑结构和功能的基础上, 指出元数据质量及与图书馆集成系统整合方式是资源发现的关键问题, 提出图书馆发展前景是基于海量元数据仓储开展数据驱动的知识服务; 彭佳等^[13]基于发现系统探讨构建文献资源深度聚合框架体系, 实现语义层面的文献组织与整合。在元数据集成管理系统研究方面, 赵捷等^[14]在系统需求分析的基础上, 提出了系统架构、系统功能和集成管理流程的设计方案; 丁道劲等^[15]基于国外元数据集成实践经验, 提供了一个多来源元数据集成组织管理框架, 从基础层、操作层和应用层3个层次探讨了多源元数据的集成管理方法; 丁道劲等^[16]设计了文献元数据集成管理框架和管理流程, 提出了元数据集成管理的推进策略; 姜恩波等^[17]针对书目元数据介绍了一个集元数据检索、转化和整合功能为一体的开源元数据管理发布平台。在元数据规范化处理研究方面, 赵捷等^[14]提出了构建规则库和规范档来解决多源异构元数据的统一规范方案。在元数据建设实践方面, 梁惠玮等^[18]介绍了国家图书馆元数据仓储建设的目标、流程和元数据整合方法。综上所述, 我国虽然在文献资源元数据的集成管理研究和实践方面取得了一定的成果, 但在文献资源元数据的多渠道获取、元数据资产管理、元数据标准化规范化建设、元数据大规模集成建设和共享应用等方面的研究还具有一定的局限性, 需要进一步研究和探索。

3 文献元数据建设的若干问题

元数据是资源发现和知识服务的基础, 发现的机制是通过文献资源元数据的集成整合, 建立元数据仓储, 编制统一的元数据预索引, 来实现简捷高效的资源检索与服务。从实践效果来看, 当前发现系统检索出的资源数量庞大、相关性不强、检索精准度不高、识别选择困难, 这些现象都与元数据的集成管理问题密切相关, 如元数据仓储架构、元数据质量、元数据测评、元数据服务等^[19]。除此之外, 笔者认为元数据集成管理还存在诸多方面的问题, 如元数据统一集成、元数据标准规范、元数据权益确立、元数据协同建设等, 这些都是必须加以关注和解决的问题。

3.1 多源异构元数据统一集成问题

图书馆文献资源可以大致分为实体资源和数字资

源。长期以来,对实体文献资源的组织揭示主要是基于编目工作建立MARC书目记录,通过图书馆的OPAC系统来进行检索和利用。经过五十余年的建设与发展,全球已经积累了数量丰富的书目记录,以OCLC为例,截至2019年9月,OCLC的WorldCat中书目记录数量已经超过4.64亿条^[20]。MARC书目记录的编制经过数十年的发展虽然已经具备了规模效应和固定模式,对实体资源的计算机检索和利用作出了积极而重大的贡献,但由于其揭示粒度粗、无法实现关联开放等缺陷也深受诟病。针对数字资源而言,目前大型数据库收录的数字资源数量庞大,其涵盖的元数据数量均数以亿计,对数字资源的组织揭示也主要是通过商用系统来实现。不同系统中的多种资源给用户一站式查找和获取文献造成了障碍;而且,文献资源的元数据格式类型众多,加之来源不同、粒度不一、形态各异,需要通过元数据的融合、聚类 and 重组,对图书馆分散的、独立存储于不同系统中的文献资源元数据进行大规模的统一集成整合,构建一个多源异构的元数据有机体,在语义层面实现文献资源的类聚与重组,达成对多源异构资源一站式检索的目标,将“资源发现”提升为“知识发现”,满足用户获取知识服务(如数据挖掘、引文分析、科学数据管理、科研前沿预测等)的迫切需求。

3.2 元数据标准规范的制定问题

元数据的标准规范与元数据建设质量密切相关。对于多源异构的元数据集成不可避免地要面临元数据层次级别不同、粒度粗细不均、著录形式不一、格式和语义冲突等问题。以元数据层级为例,传统的实体资源元数据绝大多数是书目级元数据,而大多数的数据库资源元数据是篇章级元数据,层级不同导致元数据粒度不均。以元数据著录为例,同一作品会出现多个题名(如红楼梦、石头记),同一责任者也可能有不同的名称(如鲁迅、周树人),同样的数字还会有不同的表达形式(如阿拉伯数字、罗马数字)等。以元数据格式为例,有marc、dc、xml等,不同的元数据格式带来了格式转换和字段映射问题。以相同资源元数据的唯一性为例,不同来源的元数据中,同一资源由于被多个数据库收录,会出现多条元数据,且同一资源的多条元数据之间还存在差异,带来了元数据的判重和归并问题。没有统一的元数据标准规范,将无法实现元数据的统一描述、规范标引、集成整合和元数据互操作。

3.3 元数据权益和资产注册问题

元数据权益有两方面的含义,一是指图书馆收集获取元数据的权利,二是指元数据的资产归属认定。收集获取元数据的权利主要是针对商用数据库资源而言。由于全球数字出版模式的兴起,文献资源大量以数字化的形态出品,数据库资源正成为图书馆的主力资源。据全国高校图工委统计,许多本科及以上层次高校图书馆的电子资源购置费比例已经接近或达到图书馆资源建设总经费的70%^[21]。但图书馆花巨资购买的数据库资源大多数都只有使用权,而没有电子全文的长期保存权和拥有权;图书馆引进的商用发现系统也是以系统接口的方式来提供检索服务,文献资源元数据并未存储于图书馆本地系统,图书馆对文献资源元数据缺乏自主权。图书馆收集数据库资源元数据困难,究其原因元数据具有信息资源增值功能,很多数据库内容提供商为了商业竞争而垄断元数据资源,不愿意将元数据提交给图书馆,这种状况导致图书馆无法利用海量的基础元数据来开展数据挖掘、计算分析等工作,更谈不上知识发现与知识服务。

元数据资产归属认定是针对大规模元数据统一集成体系中元数据的注册机制而言。该元数据体系集合了出版发行部门、内容提供商、图书情报机构等跨领域的众多主体机构的元数据。缺乏清晰的元数据资产权属认定,会对元数据的收割、管理、保存、更新、协调共享、馆藏资产清点等工作带来一系列的负面影响,也不利于形成联盟层面甚至是国家层面元数据大规模的集成与应用。

3.4 元数据多主体协同建设问题

语义检索、数据挖掘、知识关联和发现的实现必须基于海量元数据资源的集成管理,即需要构建“大数据”化的文献资源元数据体系。目前我国国家图书馆和大型的图书馆联盟都在积极致力于资源元数据体系建设,如国家图书馆通过文津搜索系统来收割公共图书馆的文献元数据,并与自建元数据和外购商用资源元数据进行整合集成,构建大型元数据仓储^[22];中国高等教育文献保障系统(CALIS)通过e读学术搜索收集了高校成员馆提交的免费资源、购买的商用数据资源及合作机构提供的资源,涵盖期刊、学位论文、普通图书、工具书、年鉴、报纸等资源类型,数据量已超过3亿条^[23];

国家科技图书文献中心(NSTL)除了资源的自主加工以外,还与出版社、内容提供商、数据库代理商等机构开展合作集成整合资源,其类型包含图书、期刊、会议论文、科技报告、专利、标准、计量规程等。虽然上述大型机构和联盟的元数据集成整合实践卓有成效,但距“大数据”级还相差甚远,同时在各跨界领域还没有形成统一的资源描述和元数据标准,还不能实现元数据的共享、复用和互操作,亟须相关的上下游行业和文献服务机构开展元数据的协同建设。

4 多主体协同共建的文献元数据集成管理

数据关联和知识发现必须以“大数据”级的文献资源元数据体系为基础,仅靠单一图书馆或单个图书馆联盟无法形成海量的元数据资源。达成资源发现和知识服务的目标,实施多主体协同共建文献资源元数据体系的策略是行之有效的路径。在元数据采集、元数据更新、元数据管理、新技术应用等方面,图书馆需要与信息产业链上的出版商、内容提供商、平台构建商、其他信息服务机构开展协同合作,才能满足终端用户更深层次的信息需求;而在元数据服务、终端用户需求收集、馆藏特色资源数字出版、科研成果信息获取等方面,信息产业链的上下游机构需要与作为信息传播中介的图书馆协同合作,才能促使自己的可持续发展。因此,文献元数据集成体系建设与上述产业链上的多个主体相关,多主体、多来源、多类型的异构资源元数据整合成“大数据”级的文献资源元数据体系,采用多主体协同共建模式既可实现多方共赢,又能满足用户所需。多主体协同共建的重点内容有三方面,一是元数据的统一集成,二是元数据的标准制定,三是元数据的资产注册。应该以此为突破口,加强元数据的统一集成、细粒度加工和语义关联,强化文献元数据的集成管理,破解元数据建设中遭遇的各种难题。

4.1 文献资源元数据统一集成体系建设

元数据统一集成的实质就是将分散在相对独立的各个系统和平台中的各类型元数据进行融合、聚类 and 重组,形成一个统一、高效、开放的元数据有机体系。文献资源元数据集成体系的建设涉及元数据收集、处理、集成组织、服务应用等多个层面(见图1)。建立元

数据采集、处理、整合流程,针对多源异构元数据制定统一的元数据标准,实现元数据的统一规范和互操作,支持元数据的共享和复用;在对描述型元数据进行整合的同时,也需要对语义元数据进行规范整合,揭示元数据所表达的各类实体之间、实体与属性之间的关系,并利用元数据属性进行聚类重组,统一整合分散于各个系统之中的元数据,构建细粒度、语义丰富的知识网络体系,便于计算机对文本内容的理解和对语义内容的计算,实现文献检索服务向真正意义上的知识服务的转型。

4.1.1 元数据收集

图书馆文献资源由馆藏实体资源、馆藏数字资源(外购和自建)、共享联盟资源、其他领域文献机构资源和互联网资源等构成。因此,文献资源元数据包含书目元数据、商用数字资源元数据、自建数字资源元数据、联盟资源元数据、其他机构资源元数据和网络资源元数据等,涵盖母级元数据、卷期元数据、篇章级元数据等多层级元数据。

对上述元数据进行收集,应明确各类型文献资源元数据的来源:①书目元数据,存储于图书馆本地的自动化管理系统中,是图书馆最传统最基础的元数据,是图书馆数十年编目工作积累的成果,包含馆藏图书、期刊、音像制品等物理资源的编目数据;②商用数字资源元数据,来源于图书馆外购的数据库资源,需要图书馆与数据库商、数据库代理商或资源内容提供商进行协商,签订元数据收割协议,按协议规定的方式(如FTP、OAI、API、HTTP等)开展元数据的采集;③自建数字资源元数据,是图书馆根据学校教学科研发展和图书馆服务需要自行建设的特色资源数据库,图书馆拥有该类元数据的自主权;④联盟元数据,是图书馆参与合作共建共享的资源元数据,需要遵循联盟共同达成的一致性协议,采用OAI-PMH等标准访问接口方式来采集元数据;⑤网络资源元数据,包括各类免费资源和开放获取资源元数据,可以采用搜索引擎等方式来对相关资源元数据进行收集。

4.1.2 元数据处理

对元数据进行处理是构建多元化异构元数据统一集成体系必不可少的环节。元数据处理包含如下内容。

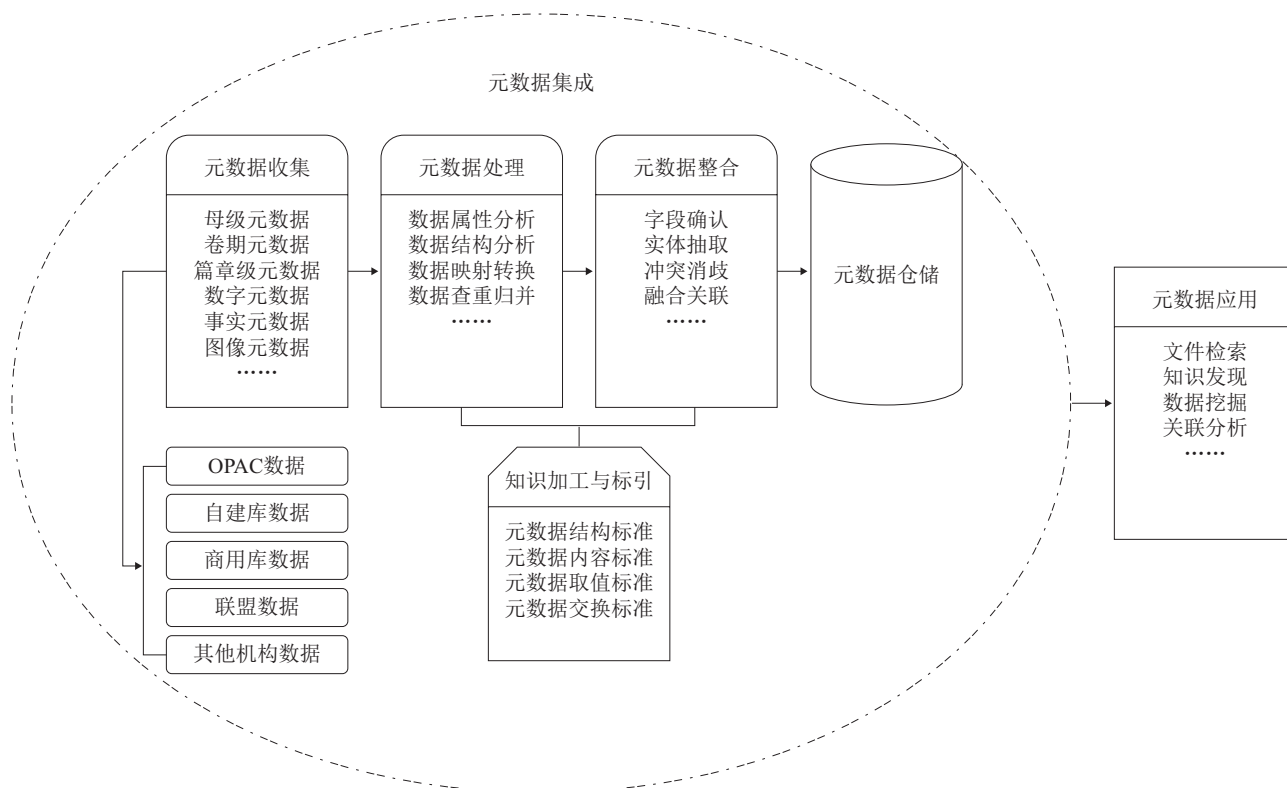


图1 文献元数据集成架构

①元数据分析。对各类型元数据进行结构分析和属性分析，了解元数据的内容结构、句法结构、语义结构及其属性关系。②异构数据同构化。首先对来自不同系统、不同类型的数字资源定义各自一致的元数据结构；其次设计异构数据的相互映射机制，通过数据转换消除数据结构差异；再次梳理不同类型元数据之间的关联关系，如书目母体元数据与其目次元数据的关系、实体资源元数据与同种数字资源元数据的关系、描述型元数据与规范元数据的关系等。③元数据查重归并。不同系统中针对同一文献资源可能存在多个重复的元数据，由于元数据制作机构标准不统一，重复的元数据之间还有可能存在差异，首先要发现和确认重复数据，制定不同类型元数据的查重策略，设计查重匹配规则，再对不同层级的各类型元数据进行机器自动比对，对重复冗余的元数据进行修改、剔除和合并，以实现来自不同系统的相同元数据整合揭示的唯一性，提高文献资源发现的效率。

4.1.3 元数据整合

文献资源元数据来自不同机构的不同数据库，这

些数据库的元数据字段定义各不相同，采用的资源描述标准也存在差异。这就要求对元数据字段进行准确的分析、确认和匹配。制定元数据字段匹配策略时，要注重字段的完整性、规范性、唯一性和语义明确性。

梳理元数据表达的各种实体及实体之间的相互关系，是元数据整合的重要一环。文献资源中包含多种实体，既可以是具体的人、事、物，也可以是抽象的概念或关系，如人名、地名、机构、事件、概念、公理、生成方式、相互关系等。分析元数据实体及其关系，制定元数据实体抽取策略，建立实体名称规范化规则，并通过实体消歧解决数据冲突和语义冲突，实现元数据实体融合。

元数据整合时应应对元数据的多种形式分别进行整合，如对同种资源的不同载体形式进行整合，对同种资源的不同版本形式进行整合，对同种资源的不同语种形式进行整合等。还应根据元数据的类型和层次，对元数据进行分层级整合。在对描述型元数据进行整合的同时，还要注重对知识元数据的整合，注意采用分类法、词表、领域词典、本体、语义网络等知识组织工具与规范文档进行知识标引和知识加工，建立知识关系体系，实现知识实体的识别、深层次聚合和知识关系网

络的构建。

4.2 文献资源元数据标准规范体系构建

元数据标准体系的建设是图书馆数据基础设施建设的重要组成部分。由于文献资源元数据情况复杂,既有实体资源的书目级元数据,又有数字资源的篇章级元数据;既有图书馆自建元数据,又有其他来源元数据,因此在开展元数据标准体系建设时,要充分考虑各种复杂情况,按照统一的标准来开展实体资源和数字资源的统一集成管理。元数据标准体系的建设涉及标准、规范和规则的建立,大致可分为结构标准、内容标准、取值标准和交换标准^[24]。①结构标准。元数据结构标准定义元数据格式,如MARC、DC、BIBFRAME、CDF等。②内容标准。元数据内容标准描述元数据元素选取、著录和构成规划,如ISBD、RDA等。③取值标准。元数据取值标准确定元数据词表和标引词的选取,如分类法、主题词表、名称规范档等。④交换标准。元数据交换标准(编码标准)记录元数据的存储和交换,如MARC(2709格式)、XML等。

不同类型的资源元数据有不同的元数据标准,同种类型的资源元数据来源于不同的机构也可能有不同的元数据标准,因此广泛收集的底层元数据不可避免地存在元数据冲突和差异。为了屏蔽元数据冲突,需建立统一的元数据标准体系,结合目标任务,兼顾不同元数据的多项特征,通过选用国际上成熟的各类型元数据标准或自制元数据标准,设计能兼容各类型、各领域元数据特点的元数据统一标准规范体系;并以统一标准体系为蓝本,开展原生元数据的生产(自建元数据),或建立外来元数据标准与统一标准体系的映射关系,通过映射转换,使文献元数据体系建立于统一的标准之上,实现各类型元数据的统一管理,提高资源发现效率,延伸和深化知识服务。

4.3 文献资源元数据采集权益和资产注册机制的确立

4.3.1 外购数字资源元数据采集

没有文献资源元数据做支撑,无法实现基于海量元数据体系的数据挖掘、数据分析、知识服务等功能,这凸显了元数据采集工作的重要性。图书馆重金外购

的商用数据库资源,其文献元数据掌握在数据库商手中。面对元数据采集困境,图书馆应大胆主张元数据权益,与数据库商开展谈判沟通,从互利互惠的角度平衡双方诉求,签订元数据收集与利用协议。协议涉及元数据收割、保存、更新、使用、开发等各个方面,如元数据提供条件(免费提供还是少量付费)、元数据提交方式(如OAI、API、FTP、HTTP等)、元数据更新周期、元数据质量要求、元数据使用范围、元数据开发许可等。该协议对元数据的知识产权、涉及费用、双方权利与义务等都应有明确的约定,以避免后续纠纷给各自带来不利的影响。同时,应注重对多种类型文献资源的元数据收集,不仅注重对常规的图书、期刊等资源元数据的收集,还要重视对多介质资源和特藏资源元数据的收集,如音频、视频、缩微资料、科学报告、科研数据、课件资料、图像资料、动漫资料等,以支持特色资源的整合,满足各种个性化的需求。

4.3.2 元数据资产注册

在基于“大数据”级别的大规模元数据统一集成体系中,文献资源元数据来源于不同领域的不同机构,数据分布广泛,体量巨大。通过元数据资产注册,可以了解元数据的资产归属和分布情况,记录元数据的版权属性,登记元数据所属机构,揭示元数据的使用范围和使用对象。元数据资产注册的内容包含元数据名称、存储类型、数据对象、数据业务规则、数据管理机构、数据馆藏属性(如数据来源、数据URL、数据存放位置)等(见图2)。元数据资产注册可以推动元数据的共享利用,分清元数据的权属关系,有利于元数据的管理、更新和协调使用,保障不同行业、不同机构协同建设的文献资源元数据集成体系平稳、持续地运行。

4.4 多主体元数据协同体系建设的驱动策略

元数据协同建设是元数据共享复用的基础。多主体元数据协同建设首先要做好顶层设计,制定跨界协同发展战略,建立长效合作机制,打造元数据协同建设与服务平台;要制定管理制度,争取政策支持,针对涉及的工作环节建立多个研究工作小组,保障平稳运行,实现多方共赢,确保元数据建设协同模式不断创新与发展。其次,要认真研讨元数据协同建设内容,注重商用资源、网络资源、开放资源、图书馆机构知识库等多源

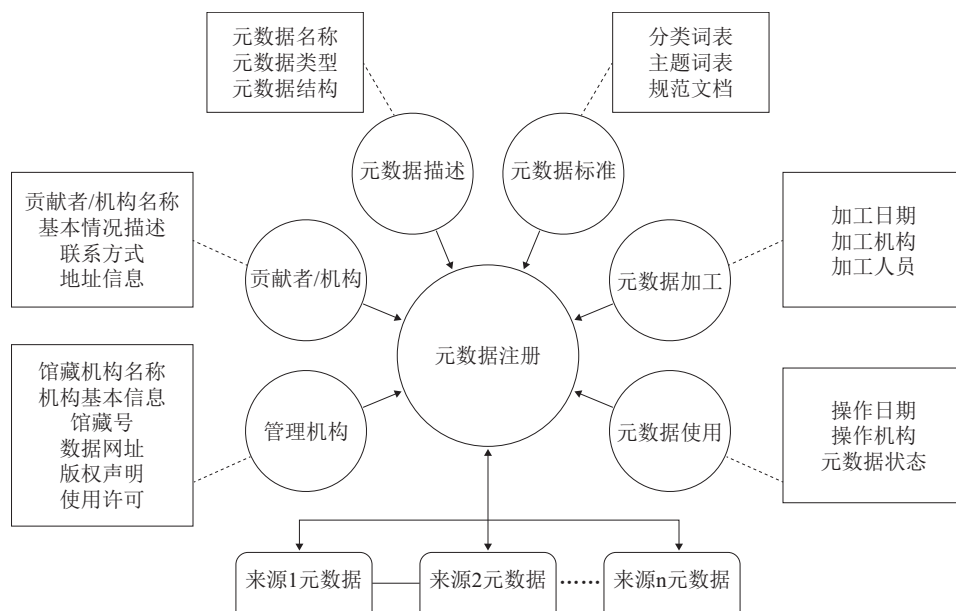


图2 文献元数据注册示意图

元数据的收集渠道,重视元数据注册与管理,研究元数据查重、匹配、映射、转换机制。再次,要积极开展元数据统一标准的制定,充分考虑各方原有标准之间的差异,建立多方原有标准与统一标准的映射关系,设计能兼容多方标准特点的元数据统一标准体系。然后,要积极开展文献资源长期保存合作,与出版商、数据库商签署多方协议,在全国范围内通过协同合作建立数字资源长期保存机制,保障数字资源的长期使用需求。最后,要强化元数据协同建设与服务平台功能,使其不仅是文献提供、资源评估、知识服务等图书馆精细化服务的平台,还是出版商、数据库商开展商业活动需要的资源展示推广平台以及用户需求信息和个性化出版信息征集平台。总之,兼顾多方需求,开展协同共建,实现携手共赢,达成元数据协同建设目标是未来大势所趋和时代发展所向。

5 结论

在数字资源占据主导地位的今天,沿用旧的文献资源整合模式已无法适应资源发现和知识服务的需求,亟须图书馆融入开放互联环境,参与大数据发展战略,重构资源组织模式,建立文献元数据体系,细粒度地深入揭示文献资源的外部形态、内容特征以及相互之间的关联关系。通过不同层级文献元数据集成建设,运用各类词表、分类法、规范文档等知识组织工

具,对文献单元以及知识单元进行结构化、系统化、关联化处理,实现海量、多源、异构文献元数据的统一集成;主张元数据权益,推动元数据注册;制定统一的元数据标准规范,解决多源元数据的格式冲突和语义冲突;打造元数据识别与关联体系,主动融入社会资源的大数据环境,全面释放馆藏资源价值;建立基于信息产业链的跨界元数据合作建设与服务机制,共同构建国家层面的大规模元数据集成体系,更好地适应基于开放互联环境的细粒度知识组织,以达成资源发现、数据挖掘和知识服务的目标。

参考文献

- [1] Europeana. Europeana Data Model [EB/OL]. [2019-03-21]. <https://pro.europeana.eu/page/edm-documentation>.
- [2] DPLA. Metadata Application profile [EB/OL]. [2020-06-02]. <https://pro.dp.la/hubs/metadata-application-profile>.
- [3] British Library. Unlocking The Value: The British Library's Collection Metadata Strategy 2015-2018 [EB/OL]. [2019-05-03]. <http://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2015-2018.pdf>.
- [4] 曾建勋. 资源建设需要推进元数据战略 [J]. 数字图书馆论坛, 2017 (11): 1.
- [5] DPLA. The Digital Public Library of America Policy Statement on Metadata [EB/OL]. [2019-02-16]. <https://pro.dp.la/hubs/>

- metadata-application-profile.
- [6] IBRAHIM I K, SCHWINGER W. Data Integration in Digital Libraries: Approaches and Challenges [EB/OL]. [2019-06-18]. <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=4A1120E952ABF2D152EA95BC46DF6F5C?doi=10.1.1.216.3957&rep=rep1&type=pdf>.
- [7] BANSAL S K, KAGEMANN S. Semantic Extract-Transform-Load Framework for Big Data Integration [EB/OL]. [2019-07-29]. https://repository.asu.edu/attachments/160603/content/CO_COMSI-2014-09-0144.R1_Bansal.pdf.
- [8] LEMOS D S, LUCAS D, SOUZA R, et al. Knowledge Organization Systems for the representation of multimedia resources on the web: a comparative analysis [J]. Knowledge Organization, 2020, 47 (4): 300-319.
- [9] PCC. PCC (Program for Cooperative Cataloging) Strategic Directions [EB/OL]. [2019-11-06]. <https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf>.
- [10] Summon. Summon网络级发现服务简介及功能说明 [EB/OL]. [2019-05-22]. <https://wenku.baidu.com/view/04d67d07561252d380eb6ea8.html>.
- [11] 国立国会図書館. 电子图书馆(数字图书馆)项目的概要 [EB/OL]. [2019-01-11]. <https://www.ndl.go.jp/zh/aboutus/dlib/index.html>.
- [12] 窦天芳, 姜爱蓉. 资源发现系统功能分析及应用前景 [J]. 图书情报工作, 2012, 56 (7): 38-43.
- [13] 彭佳, 郑巧英. 信息资源聚合与组织研究——以发现系统为例 [J]. 图书馆杂志, 2016 (3): 80-85.
- [14] 赵捷, 董微. 面向发现服务的图书馆元数据集成管理系统构建研究 [J]. 数字图书馆论坛, 2018 (7): 11-21.
- [15] 丁道劲, 马袁燕, 李勃慧. 多来源元数据集成中的组织管理框架研究 [J]. 数字图书馆论坛, 2017 (12): 58-62.
- [16] 丁道劲, 曾建勋. 文献元数据集成管理研究 [J]. 情报学报, 2019, 38 (6): 568-577.
- [17] 姜恩波, 金晶. 图书馆元数据管理与发布平台 [J]. 图书馆杂志, 2013 (6): 64-69, 75.
- [18] 梁惠玮, 萨蕾. 数字图书馆推广工程面向数字资源整合的元数据仓储构建 [J]. 国家图书馆学报, 2012 (5): 27-32.
- [19] 程颖. 资源发现系统元数据的问题与思考 [J]. 图书情报工作, 2015, 59 (9): 104-110, 126.
- [20] OCLC. worldcat内部: 图书馆资源全球目录 [EB/OL]. [2019-12-28]. <https://www.oclc.org/zh-Hans/worldcat/inside-worldcat.html>.
- [21] 教育部高等学校图书情报工作指导委员会. 2018年960所高校图书馆电子资源购置费统计表 [EB/OL]. [2019-12-30]. <http://www.scal.edu.cn/sites/default/files/attachment/tjpg/3-2018年960所高校图书馆电子资源购置费统计表.pdf>.
- [22] 杨东波, 邢军. 国家图书馆“天津搜索”的设计与实现 [J]. 国家图书馆学报, 2014 (3): 93-98.
- [23] e读学术搜索 [EB/OL]. [2019-12-28]. <http://www.calis.edu.cn/pages/list.html?id=12f95c90-3ca2-4f49-bdae-195693a5e118>.
- [24] 胡小菁. 文献编目: 从数字化到数据化 [J]. 中国图书馆学报, 2019 (3): 49-61.

作者简介

杨永清, 男, 1966年生, 副研究馆员, 研究方向: 文献资源组织、数字图书馆研究, E-mail: yyq@scu.edu.cn。
黄毕惠, 女, 1966年生, 研究馆员, 研究方向: 文献编目、知识组织、数字图书馆研究。

Research on the Construction of Metadata Integration in Literature Resources Organization

YANG YongQing HUANG BiHui
(Sichuan University Library, Chengdu 610065, China)

Abstract: The publication of massive digital resources and the great changes in information environment have brought great challenges to the traditional organization of library literature resources. The coarse-grained bibliographic data generated by traditional cataloging is losing its leading position in literature organization and retrieval. The integrated management of fine-grained, semantic, large-scale and related literature metadata is of great significance to library resource discovery and intelligent service. Literature service institutions should pay attention to the construction of literature metadata system, deal with literature metadata in a structured, systematic and related way, pay attention to metadata collection, metadata standard setting, metadata integration, metadata collaborative construction and sharing, and build a national "big data"-level metadata integration system, so as to release and enhance the value of literature resources and meet the new requirements of user resource discovery, knowledge association and intelligent service.

Keywords: Literature Metadata; Metadata Integration Management; Resource Organization; Resource Discovery; Knowledge Services

(收稿日期: 2020-07-15)