面向发现服务的图书馆元数据集成管理系统构建研究*

赵捷 董微
（中国科学技术信息研究所，北京 100038）

摘要：元数据集成管理系统是图书馆发现系统的重要组成之一。为构建该系统，本文对面向发现的元数据集成管理研究现状进行调研与总结。针对发现服务面临的主要数据问题，提出在元数据集成管理过程中，采用基于异构数据同构化、元数据统一规范、查重归并与语义冲突处理方法的解决方案。在此基础上，对系统构建需求进行分析并给出系统信息架构、集成管理流程、系统功能的设计。基于相关设计，提出一种基于领域驱动设计的系统实现方法。

关键词：图书馆系统；发现系统；发现服务；元数据集成管理；信息系统构建
中图分类号：TP317.1
DOI：10.3772/j.issn.1673-2286.2018.07.002

信息技术的快速发展极大地改变了现有信息环境，来源广泛的纸本资源、电子资源等各类文献资源对图书馆的资源建设产生深远影响，形成新的挑战。随着文献原生数字资源、回溯数字资源、开放数字资源等的大量引进，图书馆文献数字资源来源更加广泛，呈现出数量庞大、种类繁多、载体多样、分布广泛、交叉重叠、结构多样等特征与形态，给用户使用图书馆信息资源带来困扰，也给图书馆的相关工作带来不便，利用图书馆系统实现各类文献资源的整合，进而对文献资源进行深度揭示和有效利用是图书馆研究中不断探索的热点问题之一。

传统的图书馆系统在向读者提供图书馆文献服务过程中，多采用联邦数据库系统，基于中间件的数据集成系统等资源整合技术对分散的、异构的、高度自治的数据源进行集成[1]，采用此类方法构建的本地图书馆系统所需编写的数据访问接口程序数量取决于所集成的数据源数量，系统较庞大、操作较复杂、易用性较差，且受网络环境、远程数据管理系统性能等因素影响，系统性能不够稳定，用户体验不佳。随着谷歌、百度等搜索网站的服务模式日益深入人心，其新颖、亲民、方便、快捷的服务特性深受广大用户喜爱。反观图书馆系统，受传统理念影响，系统亲和度、易用性等难以有大的提升，急需创新。面对这些挑战，为适应新环境，图书馆系统设计的理念也在变化，越来越多图书馆系统开始探索以用户为中心，借鉴搜索网站的服务模式，引入发现服务机制，基于资源发现系统（简称发现系统）向读者提供更加方便快捷的文献服务。目前，发现系统已在图书馆界开始盛行[2-3]。

1 面向发现服务的元数据集成管理研究现状

发现服务主要通过发现系统实现，其工作原理是按照统一的标准规范将多来源渠道、类型众多、结构各异的元数据汇聚在一起，通过映射转换、整理规范、分析查重、聚合归并、重新组织、集中保存等手段，将多源异构元数据聚集为一个有机的整体，并在此基础上提供类似谷歌、百度的统一搜索服务。实现资源发现服务，对

*本研究得到国家社会科学基金项目“机构规范文档结构及构建方式研究”（编号：15BTQ015）资助。

多源异构元数据进行集成管理是不可或缺的部分，面向发现服务的元数据集成管理是对多源异构元数据进行整合集成，进而实施集中管理的过程。元数据整合集成[4]是将分布在不同结构数据源中的不同格式、不同性质特征的数据，逻辑地或物理地集成到统一数据集合中的过程，使用户能够以透明或直接的方式访问这些数据源。元数据集成管理主要借助计算机软硬件技术、数据库技术等，以集中方式对数据全生命周期进行管理的过程，管理内容主要包括数据收集、处理、组织、定位、存储、使用与维护等，管理对象主要包括机构、人员、流程、文件、代码、规则、规范、脚本、模型、指标、日志等实体与数据对象，以及描述其属性与关系的元数据。

元数据整合集成可分为逻辑集成、物理集成和综合集成3种方法[5]。早期的元数据整合集成主要采用逻辑集成方法，又称模式集成、虚拟数据库整合、系统平台整合、数据聚合或数据互操作方法，较典型的是基于联邦数据库的整合集成系统[6]和基于中间件的整合集成系统[7]，如整合馆藏OPAC书目数据大多基于联邦数据库方法，而目前图书馆较通用的逻辑集成方法常基于中间件方法。逻辑集成方法的基本思想是在构建整合集成系统时采用模式识别方法将分散的、异构的不同数据源的数据视图集成为全局模式，使用户能够按照全局模式透明地访问这些数据源。全局模式描述了数据源共享数据的结构、语义及操作等，用户直接在全局模式的基础上提交请求，由整合集成系统处理这些请求，转换成各数据源在本地数据视图基础上能够执行的请求。

面向发现服务的元数据集成主要采用物理集成方法或综合集成方法。物理集成方法又称数据复制、统一物理集成方法，是一种基于数据仓储的整合集成方法，将分散的、异构的非本地数据源的数据复制到与其相关的本地数据源上，并对本地数据源的数据规范性与一致性进行维护，从而提高其共享利用效率，数据复制可以是整体复制，也可以是仅对变化数据的复制。物理集成方法可以使用户直接访问本地数据源上的数据，减少甚至避免对非本地数据源上数据的访问，从而提高数据访问效率。综合集成方法是将逻辑集成和物理集成方法混合在一起使用的方法。逻辑集成和物理集成各有一定的局限性。逻辑集成适用于被集成的数据源规模较大、数据更新较频繁、数据实效性要求较高、数据不允许复制、用户查询模式多变且难以预测等场合；物理集成适用于数据分布较广、数据更新频次要求相对不高、查询获取响应时间要求相对较高、用户查询模式较稳定等场合，采用综合集成的目的是突破这两种方法的局限。

图书馆收录的文献资源类型多种多样，包括图书、专著、期刊、会议、报告、专利、学位论文、影音图片等。相关文献元数据源自图书馆、大学、出版机构、科研机构、数据库商等，呈现出数量庞大、类型众多、来源广泛、格式多样、结构各异、交叉重叠、质量良莠不一的特征，对其进行整合集成有时甚至会遇到数据格式不能转换或数据转换格式后信息丢失等问题，严重阻碍资源共享。针对此类问题，面向发现服务构建图书馆元数据集成管理系统，对多源异构元数据进行整合集成与集中管理，已成为增强图书馆服务品质的必然选择。针对多源异构元数据的整合集成与集中管理，许多学者进行了深入的研究与探索。针对档案元数据，金更达等[8]给出基于统一的元数据标准构建元数据注册系统、元数据仓储系统，通过相关系统集中管理档案元数据；针对多来源开放期刊元数据，李颖等[9]提出基于XML元数据交换、数字权益管理等技术构建资源整合平台，通过该平台整合集成开放期刊元数据；针对多来源文献书目元数据，傅红梅等[10]提出基于Web2.0信息资源组织、开放互联等构建文献资源知识库系统，通过该系统管理书目元数据；针对类型众多的图书馆数字资源元数据，任慧玲等[11]提出对多类型数字资源元数据进行集成管理的设想；为应对快速增长的海量数字资源给图书馆带来的压力与挑战，姜爱蓉[12]提出基于元数据集成、组织与揭示的集成管理思路。相关研究多未涉及面向发现服务的元数据集成管理，而此方面研究正处在起步阶段，相关文献报道尚不多见。结合工作实际，本文给出一种面向发现服务的图书馆元数据集成管理系统构建方案。

2 面向发现服务的元数据关键问题研究

元数据是发现服务的基础与保障，发现服务对元数据问题极为敏感，在元数据整合集成过程中，相关问题如果得不到较好的解决，将对发现服务产生不利的影响，因而受到业内持续关注[13-14]。通过对比分析与归纳汇总，现有的发现服务主要存在三方面的元数据问题：第一是元数据标准不统一，相关元数据在结构上存在差异，因而难以集成，难以实现一站式搜索服务，给用户的查询浏览带来不便；第二是元数据著录加工规则、质量

控制方法、规范控制方法等不一致，相关元数据在一致性、规范性、正确性、有效性等方面存在较大差异，因而难以集成和管理，搜索服务的准确性难以保证，给用户准确查找文献造成不便；第三是元数据查重与聚合归并过分依赖计算机，机器是按预先订制的规则进行自动匹配与自动处理，但面对多变的数据，相关规则不能全面适用，会出现遗漏，造成聚合归并不彻底，因而无法实现唯一揭示，搜索结果经常出现重复记录，给用户准确定位、快速获取资源造成不便。解决这些相关问题，需要在元数据集成管理过程中采取对应的措施。

2.1 异构数据同构化

对于多源异构元数据，来源不同则描述方案不同，类型不同描述方案亦有所不同，主要表现在基本框架、元素数量、描述、定义等存在一定差异，因而无法直接对其进行集成合并。解决相关问题需要制定统一的元数据描述方案，并据此对异构异类元数据进行同构化处理。在图1所示的统一描述概念模型及对应描述方案中，将实体进行抽象与归并，明确实体间相互关系，统一元素的描述与定义，参考美国国家信息标准组织发布的提高发现服务透明度的推荐做法——开放发现倡议（Open Discovery Initiative: Promoting Transparency In Discovery，ODI）[15]。方案元素集包含或可以衍生ODI推荐的两类构建发现服务集中索引所需元素集，分别是核心元素集（Core Metadata Elements）和扩展元素集（Enriched Content Elements）。

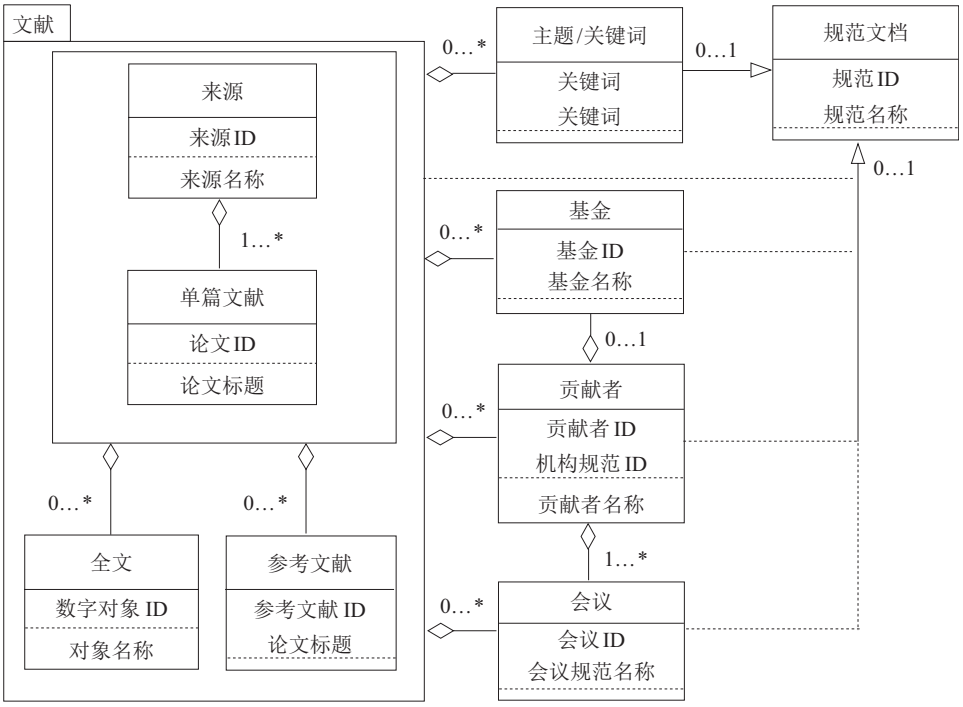

图1 元数据统一描述方案概念模型

对多源异构元数据进行同构化处理，需要针对每一来源的元数据，建立相关元数据描述方案与统一描述方案的映射转换关系并形成对应规则，在此基础上，通过映射转换等方法实现相关数据的同构化。随着来源不断增多、类型不断丰富，映射转换规则也相应增多。这种情况下，需要构建规则库，对多源异构元数据进行同构化处理。规则库（Rule Base）是将规则组织在一起并在计算机中进行保存的知识库，规则的主要作用是对标目进行约束与限制，以保证标目表达的唯一性。为便于随时调用，相关规则需完整保存，构建规则库，保存所有映射转换规则，基于规则库对多源异构元数据进行同构化处理可以较好满足这一需求。规则库建设初期可以将已经形成的规则写入数据库，后期以动态维护方式不断补充与优化完善。

2.2 元数据统一规范

图书馆元数据既有自加工数据也有其他来源的数据，由于各自的著录加工规则不一致，内容表达各异，数据不一致问题随处可见。一旦发现重复记录，需要对其进行聚合归并，由于重复记录的一些内容表达不一致，导致在格式、语义等层面出现冲突，因而难以进行处理。解决此类问题，需要按照统一的规范控制方法对多来源元数据进行规范化处理。对多源异构数据进行统一规范较为复杂烦琐，需要在实践中不断认识与总结，找到规律和对应的处理方法[16]。构建规则库和规范档（Authority File）是解决这一问题的有效途径[17-18]，多源异构元数据的统一规范可采用基于规则库和规范库的解决方案。

规则库主要用于统一国别/地区、语种、时间、卷期等表达方式，包括著录规则、检测规则及与之对应的转换规则等，著录规则与元数据著录规范相对应且可被计算机识别，如代码范围、正则表达式、数据字典、时间日期表达式等，检测规则和转换规则是对不符合著录规则的数据进行检测进而特殊处理的规则，是在统一规范过程中逐步积累形成的规则。

规范库主要用于对期刊、会议、责任者等名称进行规范控制，由规范档组成。规范档是由规范记录组成的计算机文档，其作用是对标目进行规范化控制，以保证标目的前后一致性和唯一性。图书馆多源异构元数据主要涉及品种（如期刊、会议）、机构（如文献责任机构、作者所在机构、出版发行机构）等实体。规范库主要包括母体、会议、机构等规范档，主要用于描述相关实体的属性与关系：实体属性包括唯一标识、规范名称、名称变体等元素；实体关系主要包括属分关系、等同关系、同义关系、相关关系、沿革关系等。规范档不但要将相关实体的属性信息保存在规范记录中，还要能够记录相关实体的关系信息，在规范档中为每一个实体分配唯一标识，唯一标识不但用来唯一识别单个实体，还能建立与其他相关实体的联系。以机构规范档为例，机构规范表描述了规范机构名称、起止年代等规范属性，机构规范表与其他表通过规范机构ID实现关联。机构表与机构规范表的关联可以表达相关机构间的同义关系，机构规范表与隶属机构表、从属机构表的关联可以表达机构间的属分关系，机构规范表与同级机构表的关联可以表达机构间的等同关系，机构规范表与相关机构表间的关联可以表达机构间的相关关系。机构规范档实体关系概念模型如图2所示。规范库建设按照循序渐进的原则逐步丰富与完善。

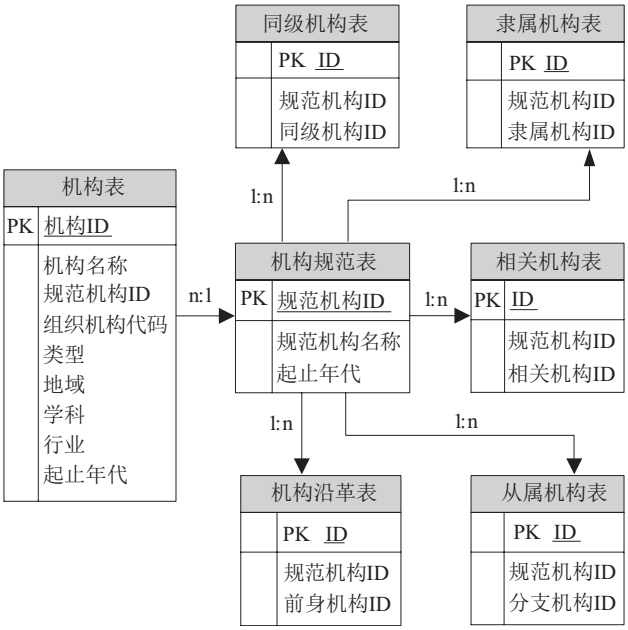

图2 机构规范档E-R图

2.3 查重归并与语义冲突处理

发现服务的重要功能之一是对多来源文献资源进行唯一揭示。为此，应保证关于同一文献的多条重复元数据记录能够聚合归并为一条记录，使来源出处归并在一起。文献元数据涉及多个层面，既有关于母体层面的元数据，也有关于卷期、篇级等层面的元数据，因此需要分别针对不同层面的元数据进行查重。查重的实质是检索，因此构建有效的检索策略必不可少。针对不同类型文献、不同层面的元数据，查重策略不尽相同，查重策略数量多且在不断调整优化，系统运用相关规则进行查重势必导致软件系统的频繁调整。针对这一问题，采用基于查重规则库的解决方案，将检索策略以规则形式存入规则库，基于规则库以模糊检索方式进行自动查重。查重结果按相似度排序，当相似度均小于预先设定的阈值时，以人工方式对查重规则补充完善或优化调整，目的是不断提高查重匹配精度，进而最大限度降低人工干预。查重规则库包括母体、卷期、篇级三个层面的规则（见表1），每一层面的规则根据文献类型由一组规则构成。

针对重复记录，通常采用优选方法选出质量好、内容较丰富的数据[19]，此种方式处理重复数据，不能达

表1 查重规则库中主要规则

文献类型	层 级	对应规则
期刊	母体	由ISSN、刊名、馆藏号等构成
	卷期	由上一层级规则与母体（卷期）中的年、卷、期等共同构成
	篇级	由上一层级规则与篇级唯一标识、题名、作者、单位等共同构成
会议	母体	由会议名称、主办单位、届次等构成
	卷期	由上一层级规则与母体（文集）题名、馆藏号等共同构成
	篇级	由上一层级规则与篇级唯一标识、题名、作者、单位等共同构成
文集汇编	母体	由母体（文集）题名、馆藏号等构成
	篇级	由上一层级规则与篇级标识、题名、作者、单位等共同构成
图书专著	母体	由题名、作者、出版者、版次等构成
研究报告	篇级	由题名、作者、单位等构成
学位论文	篇级	由题名、作者、学位、授予单位等构成

到逐渐增加数据厚度、提高数据品质的目的，因此，本研究采用图3所示的多来源重复记录处理机制对相关重复数据进行聚合归并。

对于重复的记录，其内容、格式、著录方式等存在较大差异，会出现数据冲突，特别是语义冲突。利用计算机遍历重复记录各字段内容可以发现数据冲突，聚合重复记录，采用拆分归并、叠加归并、替换归并等方法可以解决数据冲突。元数据聚合归并需忠实于原文，处理时需打开重复记录所指向的原文文件，通过内容比较进行合理操作。元数据聚合归并具有重复性且呈现一定规律性，将这种规律转换为归并规则并保存到规则库中，由计算机按照规则库中的相应规则自动完成相关操作可以大幅提高相关重复数据处理的自动化水平。聚合归并规则库的构建是一个积累重复数据处理知识的过程，需要不断认识、积累和完善。

3 系统设计

3.1 需求分析

发现系统主要基于集中索引机制（Central Index）或集中索引检索引擎（Centrally Indexed Search Engine），集中索引机制以集成仓储为基础，采用格式、索引和排名一致的算法，将相关数据与服务索引整合并进行封装保存。发现系统信息体系结构如图4所示，元

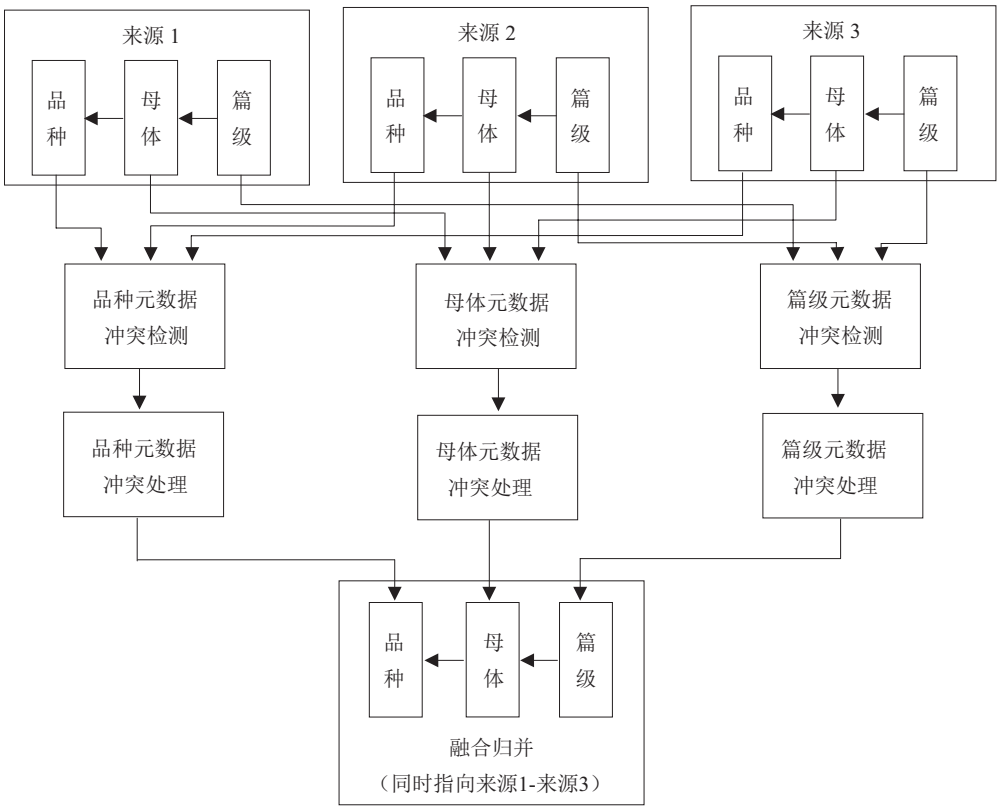

图3 多来源重复记录处理机制

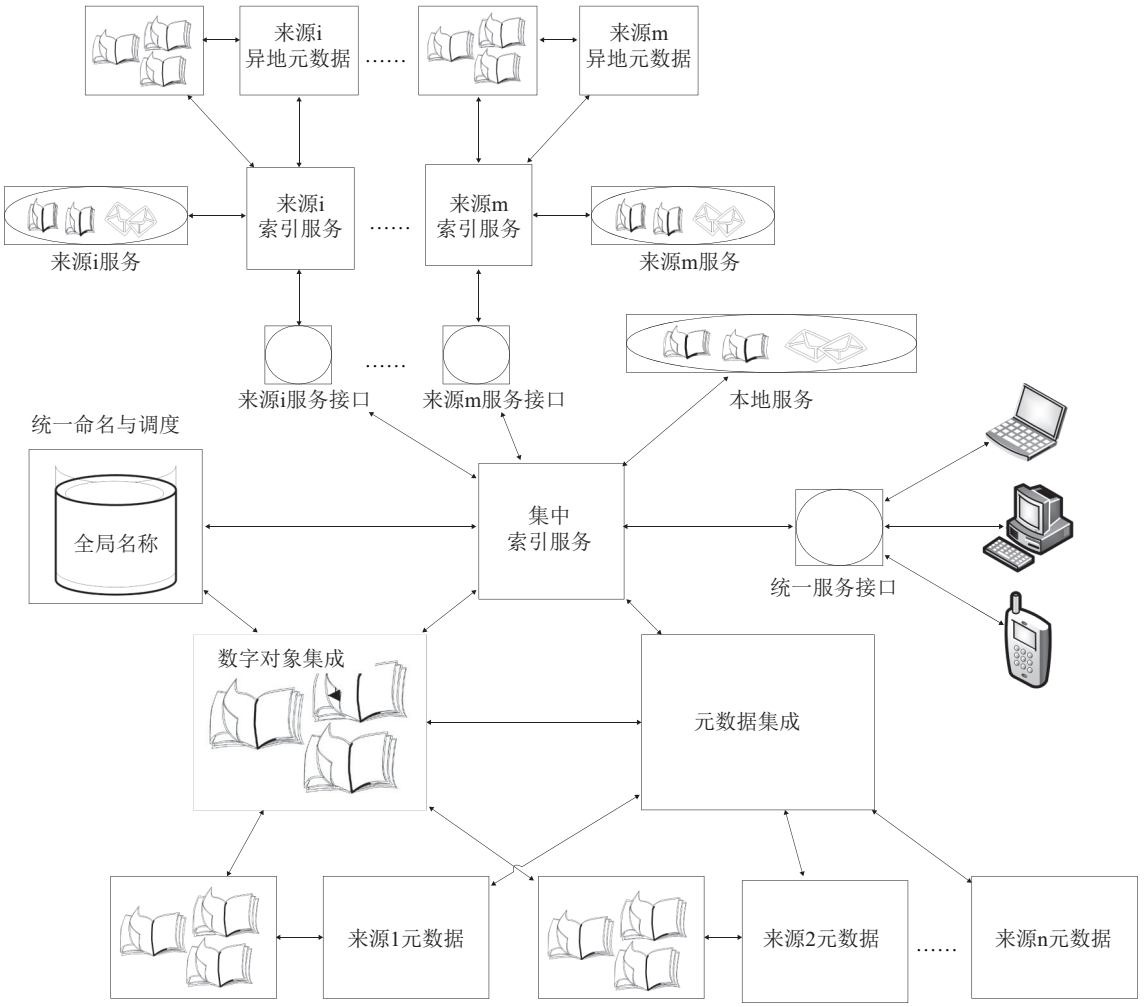

图4 发现系统信息体系结构

数据集成管理系统在对海量多源异构元数据汇聚的基础上，通过高度集成与统一管理，形成结构统一、内容规范、内外关联的元数据集成仓储，从而为发现服务提供必要的数据支撑与保障。

面向发现服务对海量多源异构元数据进行集成管理，工作量大、过程烦琐复杂，出于成本等因素的考虑，在大多环节上采用计算机技术进行自动处理，缺乏人机交互。由于计算机是按照人的指令处理相关数据，而相关指令主要基于处理相关数据问题的知识，没有人工参与，一些问题无法及时发现，因而导致系统性的数据问题时常出现。针对此类问题，现阶段的元数据集成管理还需人工介入，目的是及时发现新的数据问题，及时总结提炼，形成相应的处理知识并及时加以运用，进而在保证数据质量的前提下不断提高集成管理自动化水平。人工介入是全流程的参与，在组织内部需要形成专门的工作团队，以流程化、知识化、自动化的工作方式协同开展图书馆多源异构元数据的集成管理。

面向发现服务的元数据集成管理是一项技术性、专业性很强的工作，离不开平台的支撑，搭建元数据集成管理系统就是为满足这一需求。具体来说，首先，系统的构建围绕发现系统的构建展开，形成完善的信息体系架构，提供充分的数据访问接口，满足发现系统调取全文数据及相关配套数据的需要；其次，围绕元数据集成仓储建设，提供采集备份、转换校验、清理规范、整合集成、集中保存、检索查询、统计分析、监测预警等管理功能，以满足发现系统集中索引服务的需要；最后，围绕相关工作实际，固化业务流程、建立完善的知识积累与运用机制，提供强大的自动化处理能力和海量数据分析处理能力，确保系统易用好用、快捷高效、安全可靠、开放扩展，以满足相关工作团队开展元数据

集成管理等业务工作的需要。

3.2 系统信息架构设计

元数据集成管理系统在发现系统信息体系中起枢纽作用。一方面将不同来源的元数据汇聚在一起，另一方面为其他相关系统提供数据支撑与保障。元数据集成管理系统的来源数据主要有馆际联合目录（OPAC）数据、自加工数据及其他方数据，集成后产生的、含有多来源指向的母体、卷期、篇级等元数据，可以为索引服务系统构建集中索引提供基础元数据，为统一命名与调度系统管理数字对象提供相关来源链接数据，为统一检索与服务接口系统进行信息组织与构建、资源导航体系构造等提供相关规范数据。具体设计的元数据

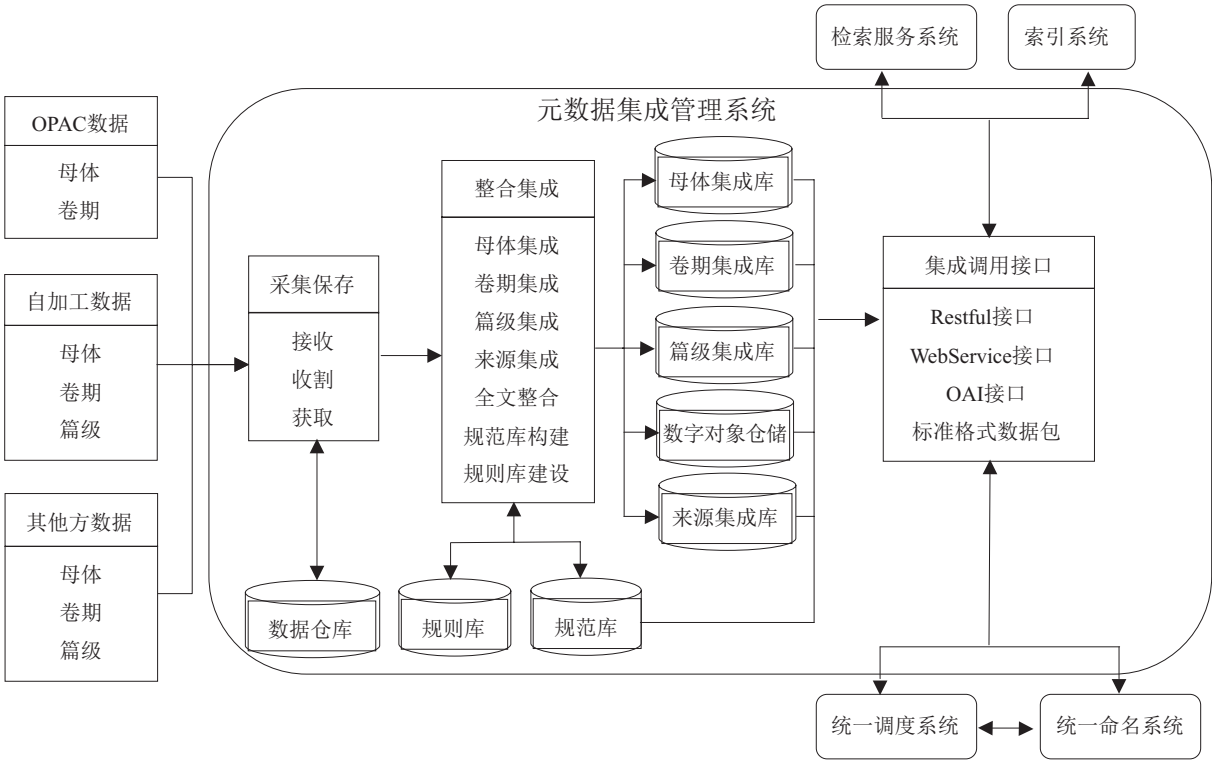

图5 元数据集成管理系统信息架构

集成管理系统信息架构如图5所示。

3.3 集成管理流程设计

流程再造是以满足用户需求为宗旨，借助先进信息技术和管理手段，对传统业务模式、流程、组织形态等进行全面梳理与根本性改造，以期使相关业务在效率、效益、质量等方面发生显著性转变。为进一步提升服务水平，许多图书馆基于流程再造理论重塑业务流程[20-22]，并据此对相关业务进行管理。元数据集成管理工作是图书馆重要工作内容之一，面对大量涌入的描述定义各异、文献类型多样、质量良莠不一、相互交叉重叠的多源异构元数据，梳理、整合集成相关数据以满足发现系统服务要求，对进度、质量、时效等要求越来越高，如果不借助先进的技术手段和流程化管理手段，相关需求难以满足。为实现多源异构元数据的流程化管理，基于流程再造理论，结合工作需要，采用图6所示的基于协同的工作流模型，建模方法采用基于事件驱动过程链（Event-Driven Process Chains，EPC）的方法[23]，该方法使各环节通过事件状态相互衔接，一个环节的事件状态发生改变将导致相邻环节的事件状态随之发生变化，这种变化将传至全流程各个环节。

图6所示的流程涉及采集管理、任务管理、加工管理和质量管理4个基本环节，各环节职责划分明确，避免因职责界限不清、接口不明造成相互推诿、不负责任等问题，其主要职责分别为：①采集管理，按批次接收

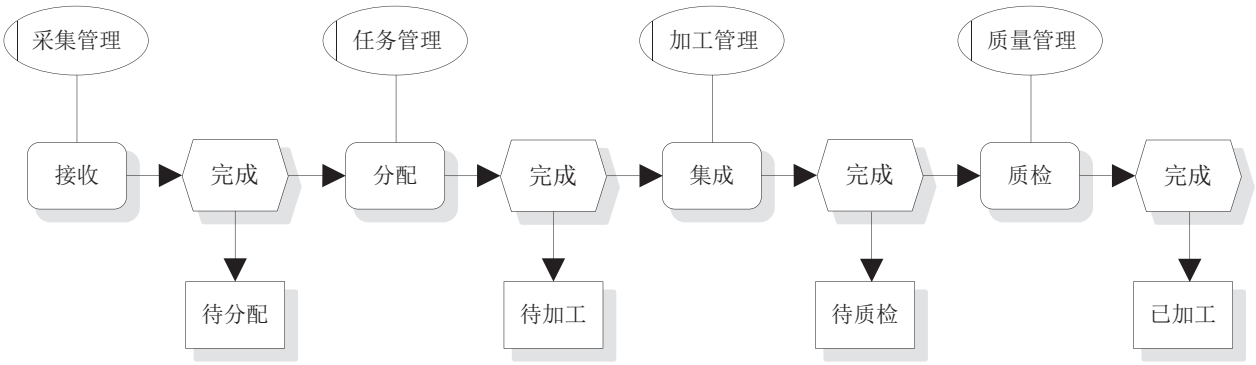

图6 元数据集成管理工作流模型

数据并对其进行转换解析与校验，接收完成后系统将所接收的数据包置于“待分配”状态；②任务管理，根据数据加工人员的工作状态，将“待分配”数据包分配给数据加工人员，分配完成后系统将“待分配”数据包置于“待加工”状态；③加工管理，对“待加工”数据包进行整合集成，整合集成后系统将“待加工”数据包置于“待质检”状态；④质量管理，针对“待质检”数据包进行质量检验与集中保存，质检完成后系统将“待质检”数据包置于“已加工”状态。

3.4 系统功能设计

系统的使用人员包括组织内部承担不同工作职责、具备不同角色的工作人员。为满足组织内部各类人员的使用要求，系统功能构成及资源配置需完整全面，但根据实际业务需要，并非所有人员都有权使用完整的功能与资源。为便于使用，更为保证系统安全和数据安全，需依据使用人员的职能与角色赋予相应的系统功能操控权限和相关数据调用权限，为满足这种安全访问控制需要，采用基于角色的访问控制机制（Role-Based Access Control，RBAC）[24]。RBAC属于强制访问控制技术，基本原理是以角色为中间媒介建立用户与权限的联系，克服直接向用户授权的不足。借助角色这一实体控制用户对系统资源的访问，有效提升系统安全性，改善易用性，降低管理复杂度，以及减少授权工作复杂性。元数据集成管理系统对多源异构元数据进行集成管理，作为相关业务的工作平台，为保证集成管理工作正常开展，元数据集成管理系统需提供访问控制、采集管理、任务管理、加工管理、质量管理、数据管理和系统管理方面的功能，整体功能架构如图7所示。

4 系统实现方法

4.1 系统开发思想

支持相关业务人员以流程化、知识化、自动化方式集中管理多源异构元数据，业务环节较多，相关环节的人员角色与权限有所不同，所涉及的管理方法差异较大，数据对象多样繁杂且愈加复杂，这些因素导致整个业务逻辑需要不断增多，功能需要不断扩展与强化，性能需要不断提升。为适应这种变化的需求，采用基于领域驱动设计（Domain-Driven Design，DDD）[25]的方法构建系统，目的是让业务人员参与到系统构建中，通过与软件开发人员的协同配合，降低开发复杂度，缩短开发周期，确保所开发系统的性能与功能能够更好地满足业务需求，从而保证系统的可用性、可维护性和可扩展性。DDD是一种用来指导面向复杂领域的软件开发项目的思维方式，该方法采用分层方式构建系统，核心是让参与开发的业务人员构建领域模型（Domain Model）。领域是系统的核心业务，领域模型是领域概念的抽象，是通用语言，由实体、值对象、领域服务、聚合或聚合根、仓储、工厂等模型元素构成。按照DDD设计思想，系统架构可以分为接口层（Interface）、应用层（Application）、领域层（Domain）和基础设施层（Infrastructure）。

接口层包含服务于客户端的所有表现逻辑及与其他系统进行交互的接口与通信设施，依赖于应用层与领域层。

应用层包含各类应用服务，依赖于领域层和基础设施层。应用服务只负责协调并委派业务逻辑，不负责业务逻辑的实现，业务逻辑大多由领域层对象承载和处理。

图7 元数据集成管理系统功能架构

领域层是整个系统的核心层，维护使用面向对象技术实现的领域模型，几乎全部的业务逻辑会在该层实现。在该层中采用模型元素对业务逻辑进行封装，这种与其他层的松耦合，能够以最小代价实现其他层的调整或替换。

基础设施层为接口层、应用层和领域层提供支撑。所有与具体平台、框架相关的实现会在基础设施层中提供，避免其他层掺杂进这些实现，从而“污染”领域模型。该层最常见的设施是对象持久化的具体实现。

4.2 技术方法

对DDD思想的核心支持与技术实现是DDDLib库，是一整套支持DDD思想的类库，可以使开发者轻松创建符合DDD思想的项目，更适合业务复杂、规则灵活多变、分析抽象要求高的项目。多源异构元数据集成管理系统的开发基于DDDLib开发框架，在现有的相关开源软件中，选用基于Java EE企业级应用的开源开发平台Koala[26]。借助Koala平台提供的DDDLib、Hibernate、JPA等基础类库及开发工具，以流水线的生产方式加快系统开发进程。采用DDDLib框架开发的系统在体系结构上采用B/S结构和图8所示的多层体系架构，能够较好地支撑工作团队对多源异构元数据进行集成管理。系统开发完成并投入使用，运行实践表明，系统构建方案切实可行，系统性能与功能的不断扩展与强化，满足多源异构元数据的集成管理需要。

图8 基于DDDLib的开发框架

5 结语

系统的构建采用迭代式的开发路线。建设初期，侧重系统基础框架与基本功能的实现、业务流程的打通和规则规范的摸索，在证实方案可行性、系统基本成型的基础上，经过持续的功能扩展、性能优化和加工知识的积累，系统管理能力和自动化水平不断提升。在系统的有效支撑下，组织内部数据加工管理工作团队能够基于本地网络以流程化、知识化、自动化的工作方式对馆藏期刊、会议、科技报告等文献的多源异构元数据进行集成管理，组织内部其他人员还可以利用系统中的海量多来源数据，借助系统提供的比较分析、报表编制等功能开展相关业务工作，工作效率得到显著提升。

参考文献

[1] 莫秀娟. 资源整合技术研究 [J]. 图书馆学研究，2011（1）：69-73.

[2] 陈定权，卢玉红，杨敏. 图书馆资源发现系统的现状与趋势 [J]. 图书情报工作，2012，56（7）：44-48.

[3] SADEH T. From search to discovery [J]. Bibliothek-Forschung und Praxis，2015，39（2）：212-224.

[4] 马文峰，杜小勇. 数字资源整合方式研究 [J]. 图书情报工作，2005，49（5）：67-71.

[5] 陈跃国，王京春. 数据集成综述 [J]. 计算机科学，2004，31（5）：48-51.

[6] 张兵，张荣肖，潘玉平. 联邦数据库系统 [J]. 计算机系统应用，1995（1）：50-54.

[7] 李爱华，陶宏才. 一种基于XML的数据集成中间件系统方案 [J]. 计算机与现代化，2007（6）：36-39.

[8] 金更达，何嘉荪. 档案信息资源集成管理中的元数据问题及对策研究 [J]. 中国图书馆学报，2006，32（4）：56-59.

[9] 李颖，梁冰，乔晓东. 多国OA期刊资源整合平台的构建及其权益管理功能的引入——基于下一代NSTL体系的国际合作 [J]. 数字图书馆论坛，2011（8）：32-36.

[10] 傅红梅，张建勇，路纳新，等. 构建用户需求驱动的下一代联机目录 [J]. 图书馆建设，2012（2）：35-38.

[11] 任慧玲，曹海霞. STM数字出版对图书馆资源建设的影响 [J]. 数字图书馆论坛，2014（5）：2-6.

[12] 姜爱蓉. 图书馆系统的过去、现在与未来 [J]. 数字图书馆论坛，2015（8）：2-7.

[13] 秦鸿. 关于发现系统的问题与思考 [J]. 数字图书馆论坛，2012（7）：17-20.

[14] 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.

[15] Open Discovery Initiative Working Group. Open Discovery Initiative:Promoting Transparency in Discovery [EB/OL]. [2017-11-21]. https://groups.niso.org/apps/group_public/download.php/11606/rp-19-201x_ODI_draft_for_comments_final.pdf.

[16] 杨青云，赵培英，杨冬青，等. 数据质量评估方法研究 [J]. 计算机工程与应用，2004（9）：3-4，15.

[17] 刘芳，李敏，任洪敏，等. 基于规则库的数据质量评估方法 [J]. 计算机系统应用，2017，26（11）：165-169.

[18] 曾建勋. 推进规范文档建设 [J]. 数字图书馆论坛，2015（7）：1.

[19] 满靖，闫健卓，王普. 信息集成中的数据冲突解决策略 [J]. 计算机与信息技术，2006（1）：66-67.

[20] 陈能华，周淑云. 图书馆业务流程重组的动因与意义 [J]. 图书馆学报，2004，30（5）：39-42.

[21] 徐军华. 1993-2009年图书馆业务流程重组（BPR）研究综述 [J]. 图书馆论坛，2010，30（4）：7-9.

[22] 吕淑丽. 国内图书馆流程再造研究综述 [J]. 图书馆界，2013（3）：50-53.

[23] KINDLER E. On the semantics of EPCs：a framework for resolving the vicious circle [C] //International Conference on Business Process Management. Potsdam（DE）：Computer Science Department，University of Paderborn，Germany，2004.

[24] JAEGER T，PRAKASH A. Requirements of role-based access control for collaborative systems [C] //Proceedings of the First ACM Workshop on Role-based Access Control. Gaithersburg，MD（US）：Software Systems Research Laboratory，Department of Electrical Engineering and Computer Science，University of Michigan，1996.

[25] ERIC EVANS. Domain-driven design：tackling complexity in the heart of software [M]. Boston：Addison-Wesley，2004.

[26] Koala开发平台介绍 [EB/OL]. [2017-12-21]. http://openkoala.org/introduction.html.

作者简介

赵捷，男，1959年生，研究馆员，中国科学技术信息研究所信息资源中心副主任，研究方向：数字图书馆研究、信息系统研究、知识组织与知识链接研究，E-mail：zhaojie@istic.ac.cn。

董微，女，1987年生，博士，研究方向：数据挖掘、网络爬取，E-mail：dongw@istic.ac.cn。

Research on the Construction of Library Metadata Integration Management System of Discovering-Oriented Service

ZHAO Jie DONG Wei

(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: Metadata integration management system is an important component of library discovery system. To construct the system, this article has investigated and summarized the research status about metadata integration management of discovering-oriented service. To solve the main data problem during discovery service, this paper has proposed specific solutions about constructing isomorphic data, unifying isomerous data, duplicate checking and semantic conflict processing. And on this basis, system construction requirements are analyzed. This article has put forward system information framework, integration management workflow, system function, and so on. Based on relevant design, this paper has proposed a system implementation method based on domain-driven design.

Keywords: Library System; Discovery System; Discovery Service; metadata integration system; Information System Construction

（收稿日期：2018-06-04）