

# 国家科技图书文献中心业务流程再造和 系统建设方案设计

沈仲祺<sup>1</sup> 张建勇<sup>2</sup> 曾建勋<sup>3</sup>

(1. 国家科技图书文献中心, 北京 100038; 2. 中国科学院文献情报中心, 北京 100190;

3. 中国科学技术信息研究所, 北京 100038)

**摘要:** 本文根据国家科技图书文献中心知识化服务的发展战略, 分析中心在新的数字信息环境下业务流程和系统存在的问题, 从用户需求出发梳理中心的总体业务布局, 分析了资源建设、数据管理、系统服务等业务的建设重点和关联关系, 研究中心数字业务流程再造, 并设计相应的系统建设和改造策略。本方案的实施将提升中心知识化服务能力。

**关键词:** 业务布局; 流程再造; 业务规划; 系统建设

**中图分类号:** G250 **DOI:** 10.3772/j.issn.1673-2286.2020.07.001

**引用格式:** 沈仲祺, 张建勇, 曾建勋. 国家科技图书文献中心业务流程再造和系统建设方案设计 [J]. 数字图书馆论坛, 2020 (7) : 3-11.

国家科技图书文献中心(下文简称“中心”)自2000年成立以来,以资源共建共享为核心,以网络化、集成化的文献信息服务为特色,按照“统一采购、规范加工、联合上网、资源共享”原则,构建国家科技文献战略保障服务系统。经过20年发展,中心已经建立了相对稳定的业务流程,形成了资源建设、数据加工、网络系统、文献服务的业务工作框架,建成了面向全国在国内最大的科技文献信息保障服务系统。面对不断发展的信息环境,特别是资源数字化、开放化、数据化和知识化的发展潮流,中心提出基于大数据开放融合的知识服务发展战略,将推进文献资源发现和获取服务、知识评价和学科态势分析服务、语义关联和智能服务的知识化服务体系<sup>[1-2]</sup>。

美国、欧盟、英国等发达国家/组织面对新的数字信息环境,纷纷研究和制订发展战略和发展重点。美国国会图书馆发布“2019—2023战略规划和数字战略”<sup>[3]</sup>,提出以用户为中心、数据驱动、数字化作为基础的发展战略,重点实现资源优化和增强访问服务,使用户随时随地使用独特馆藏集合和服务;欧洲研究图书馆协会发布“欧洲研究图书馆协会2018—2022年发展规划”<sup>[4]</sup>,

提出研究型图书馆的3个战略方向,即创新的学术交流平台、数字技能与服务、可互操作且可扩展的研究基础设施。大英图书馆推出馆藏全球化战略<sup>[5]</sup>,提出从本地馆藏到全球服务的战略目标,积极推进馆藏全球化、类型多样化、服务开放化的进程;大英图书馆还提出元数据战略路线图,以保存和提供各种物理格式的数字文献内容,努力支持整个馆藏的可计算和智能化。

数字信息环境的核心特征是开放科学,目前开放获取已成为一种全球共识,从开放科技论文到开放数据,再到开放科学,以知识共享为特征的开放运动不断纵深发展。欧盟地平线2020研发计划<sup>[6]</sup>要求所有项目科研成果在2020年实现全部开放获取,欧洲11个国家的科研资助机构和欧洲研究理事会发布了开放获取S计划<sup>[7]</sup>;德国马普学会发起包括国家科技图书文献中心在内的全球机构签署参与的OA2020倡议,呼吁到2020年将所有订购期刊转换为开放出版。STM发布的2018年报告中指出开放获取市场约占市场出版总量的20%~22%,且这种上升趋势在未来几年还将继续<sup>[8]</sup>。开放科学开启了开放式科学研究与交流模式<sup>[9]</sup>,同时也对现有科技文献建设提出了挑战。

当前,科技资源的信息化、关联化、数据化和可计算成为知识创新的重大需求<sup>[10]</sup>,语义网、认知计算、深度学习等成为信息资源开发利用的核心驱动力,国际主流的科技文献服务平台加速向专业化、知识化转型,以知识发现与知识计算为核心的新一代知识服务体系呈现蓬勃发展之势,基于科技文献大数据的内容挖掘、语义搜索、智能问答、态势跟踪、情报与决策分析成为知识服务的发展重点。如美国国家医学图书馆开发的世界上最权威、最具影响力的生物医学信息服务系统PubMed<sup>[11]</sup>,谷歌公司推出的Google Scholar,美国化学文摘的SciFinder<sup>[12]</sup>,德国比勒费尔德大学图书馆开发的BASE-Search开放资源学术搜索引擎<sup>[13]</sup>,通过整合图书馆目录和开放获取资源,提供全球异构学术资源的免费检索与获取服务。施普林格·自然(Springer·Nature)推出的Scigraph(科研图谱)<sup>[14]</sup>,德国Transinsight公司与德累斯顿技术大学合作研发的Gopubmed<sup>[15]</sup>,基于语义词表和本体,实现语义网络、生物医学信息检索和生物医学热点可视化分析。Clarivate的德温特创新平台(Derwent Innovation)提供全球权威可靠的专利数据和科技文献检索与分析功能<sup>[16]</sup>。这些系统的发展极大地促进了知识服务能力。

在数字信息环境快速发展的背景下,中心研究确定了下一步需要重点建设的业务方向,优化和重组原有基于印本文献的采集加工和服务业务流程,建立适应数字业务环境变化的知识化服务业务布局,从而提高中心业务运行的效率,提升中心知识化服务的能力,为用户提供专业化的文献信息服务。

## 1 中心现有业务流程和系统分析

经过20年的发展,中心形成了以印本文献管理和服务为基础的数字业务流程,主要包括印本文献采购、文献编目、分布式数据加工、数据集成仓储发布和原文传递等相关环节(见图1)。但近年来随着数字文献的大量出现和开放获取资源的快速发展,基于印本建设的业务系统难以将数字资源和开放资源纳入已有的业务管理流程,中心订购的网络全文库及回溯库、成员馆订购的全文库、中心建设的开放资源库等均在资源管理的主流程之外,只能单独运行或提供简单链接服务。中心订购的专利数据、标准数据和中文期刊论文数据只是提供数据发布服务,没有统一进行关联揭示和管理。数据分散之后缺乏关联和融合,难以开展有价值的分析服务。

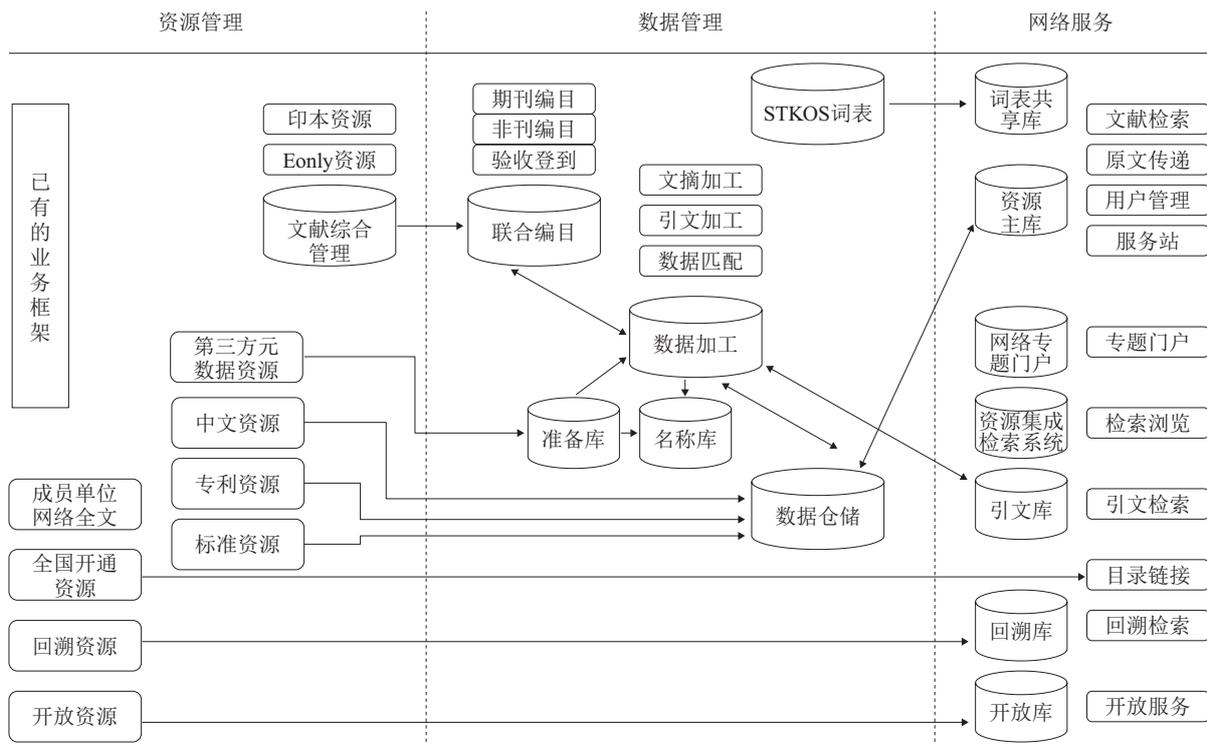


图1 现有业务系统布局 and 流程

中心系统已有的系统和服务分布在不同系统中,如检索和原文传递服务(订购资源)、回溯库服务系统(回溯资源)、引文服务(引文库)、开放资源服务系统(开放资源)、热点门户(网络专题资源)等都是不同历史阶段开发的独立系统,系统之间缺乏关联,彼此间没有衔接和配合关系。特别是面向用户的服务系统虽然经过了多次升级,但相对国外先进的同类系统服务功能比较单一,仅提供订购资源的文献检索和原文传递功能,且存在检索效率不高、检索不准确,以及缺乏准确理解用户检索意图、智能精准检索、检索结果分析功能等,用户体验不够友好,支付和订单管理复杂烦琐。现有系统的数据标准不统一,难以形成统一的大数据集合,业务系统和业务流程缺乏整体布局,欠缺有效的工作协同,导致系统服务效率低下,严重制约中心科技文献服务的深入发展。

随着开放科学和数字信息环境的不断发展,大数据和智能服务技术在信息服务领域的深入应用,国外发达国家加快了信息管理和服务的基础设施以及新技术的研究与应用,新的智能服务和知识服务系统不断出现。而中心的系统和服务水平与我国科研人员及社会大众期望值差距较大。根据用户需求和满意度调查统计,用户期望的服务是数据和系统高度融合的服务,通过访问系统发现各种类型(文献、专利、标准、多媒体、使用、评论)数据,不仅能提供文献检索和传递,也包括数据分析评价、热点监测、引证分析、学科态势等数据增值和专题服务<sup>[17]</sup>。

中心的发展定位是为全国用户提供普惠的公益文献信息服务,同时建设成为国家科技文献保障基地。中心要满足国家战略保障需求和全国用户的文献需求。国家战略保障需求要求从战略角度,满足国家对文献资源的可靠保存和保障使用的需求,特别是数字时代,资源的网络化和数字化特点带来的资源不可控问题,保证国内外文献资源的安全保存并长期可靠使用非常重要。用户的文献信息需求是中心发展的根本动力,中心不仅要满足用户的文献资源有效获取的需求,同时应满足用户专门而深入的信息需求。战略保障需求和用户需求都要求中心的业务系统、资源数据和业务紧密协同,相互有机衔接,形成可靠的业务流程和运行机制。

## 2 总体业务布局

按照中心的业务发展规划和方向,根据用户需求和现状分析,中心总体业务布局要满足知识化服务的要求,需要加强顶层设计,强化协同关联,规划具体的业务布局和业务系统,形成合理的数字业务流程。改造和调整原有分散的业务和系统,纳入新的业务流程,形成有机整体。从总体上来看,中心业务可以分为面向用户的业务和内部业务。其中,面向用户的业务包括从文献检索到知识服务的各类服务平台,提供一般文献传递服务和咨询服务,以及信息分析服务和专题服务等;内部业务包括资源组织和数据库建设,主要支撑用户服务的开展。网络和系统支撑整个中心业务的发展(见图2)。

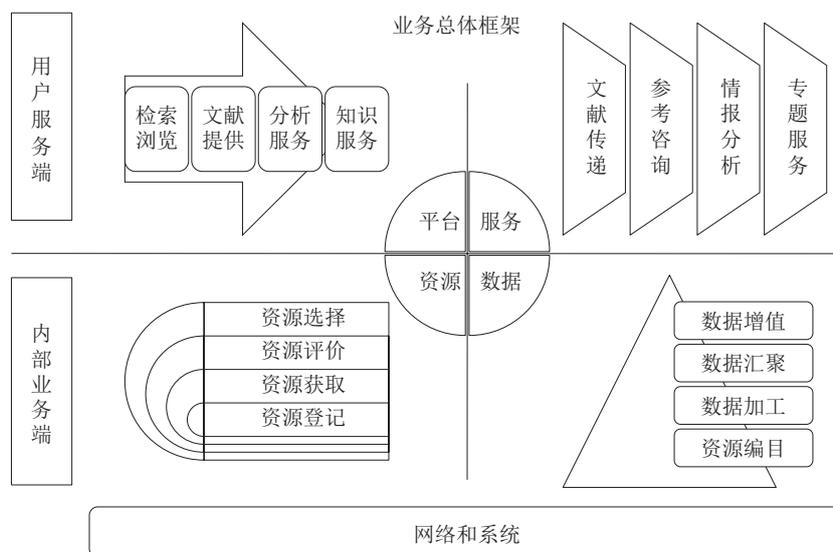


图2 中心业务总体框架

在该框架基础上,经过分析提出详细的业务布局(见图3)。在详细业务布局中,归纳和整理细化中心的各类业务,其中“①资源综合管理”扩展了要资源管理的范围和类型,除了继续管理印本资源外,要更多地管理数字资源、开放资源、回溯资源、成员单位全文资源、合作单位全文资源、第三方元数据资源。资源管理的核心仍然是资源的选择、评估、订购和获取,加强资源获取渠道的管理和资源版权的管理。资源管理的对象是资源集合。

数据管理包括全文存储、元数据集成加工管理、数据增值计算、语义标注和数据存储,形成完整的数据管理生态线。其中“②数字全文保存管理”要把中心获得的回溯全文、e-only全文和开放资源全文统一管理,

可靠保存,可供其他模块调用。“③元数据集成加工管理”包括编目管理、资源调度管理、文摘和引文元数据的加工汇聚管理,主要目标是统一处理来自资源管理的各类元数据,包括重复数据处理、数据融合、资源编目、数据加工和资源调度关系管理等。管理的数据对象是实体元数据。“④元数据增值和数据关系计算”包括数据属性增强、名称规范和引文关系计算、数据关系计算和实体标注,目的是提高元数据的完整性和质量,扩展数据关系。主题标引和分类标引是利用STKOS词表标注文献,提高数据的知识组织能力。“⑤数据仓储管理”重点建设各种元数据的存储和索引,为用户系统服务功能的实现建立数据基础。

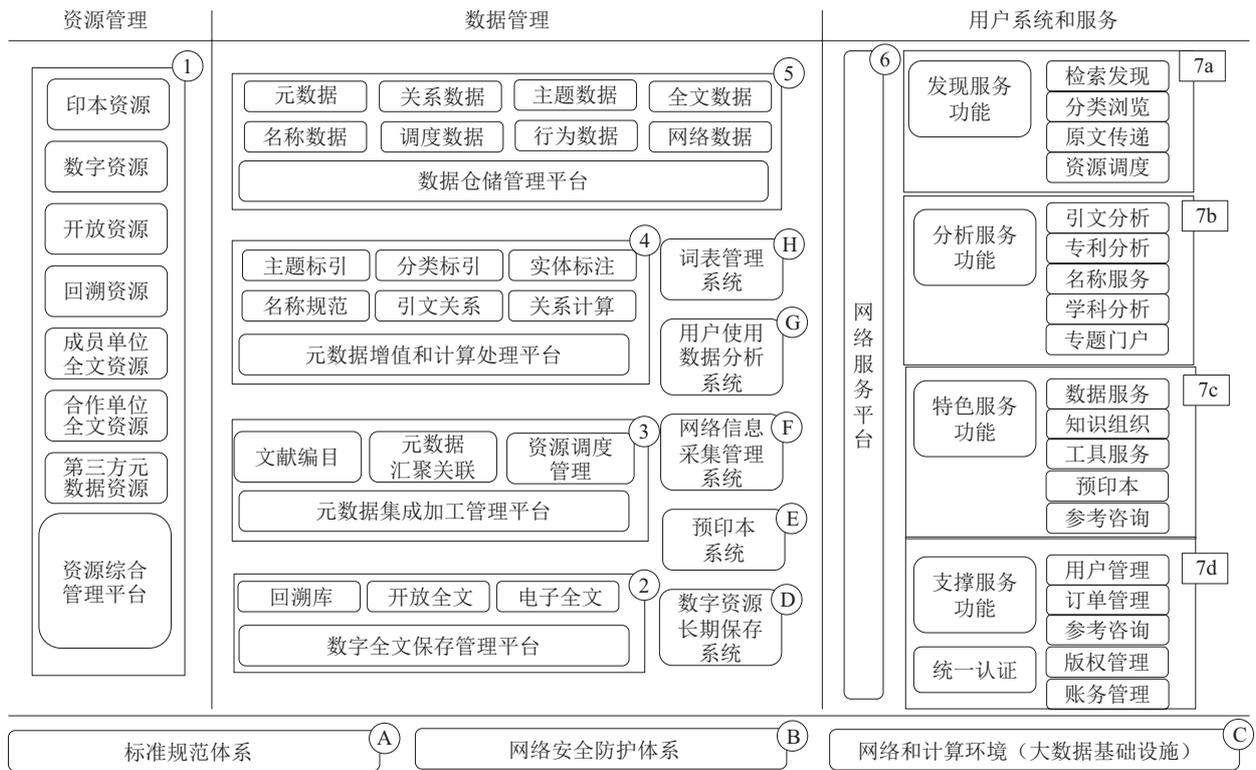


图3 中心新的业务布局

数字资源长期保存系统(D)作为国家数字资源长期战略保障基础,预印本系统(E)、网络信息采集管理系统(F)提供元数据输出接口和全文调用接口主流程。用户使用数据分析系统(G)主要处理来自网络服务平台的日志数据,形成可调用的用户画像数据和用户使用相关数据。词表管理系统(H)支持文献主题标引和分类工作。

用户系统和服务提供用户检索、分析评价和专题服务等,其中“⑥网络服务”根据用户需求建立分层次

的服务结构,具体包括发现服务(7a)面向所有人群提供服务;分析评价服务(7b)面向特定用户和注册用户提供深入的文献分析服务和学科态势分析服务,也包括网络专题信息门户和面向行业、企业和国家重大战略的专题门户;特色服务(7c)面向特殊人群提供定制数据、工具服务和知识组织体系和标准规范支持服务;支撑服务(7d)主要是用户管理、参考咨询、订单管理和支付管理等事务过程管理。

以上系统和业务要遵循统一的标准规范体系(A)

和网络安全防护体系(B)框架,应用系统要在统一的计算机基础设施(C)上建设。

### 3 具体业务流程和系统设计

#### 3.1 资源建设业务流程

现有资源建设偏重于印本资源的订购管理,相关系统包括文献综合管理系统和联合编目系统,主要是对印本资源引进订购和编目,资源建设业务流程再造的重点在于改变传统观念,推动资源开放的融合,改变单纯购买资源的业务模式,从资源保障和发现角度构建新的业务模式。资源描述将突破原有资源描述性编目范畴,还包括描述资源的版权状态、获取方式等。

在开放信息环境下资源建设业务重点在于资源的发现、评估、采集、合作共享。资源建设从印本扩展到数字资源、开放资源、合作共享资源、数据资源等。资源综合管理系统建设包括文献订购管理、采集渠道管理、合作渠道管理等,以及对中心采集资源的选择、获取、管理、评估等功能,系统管理各种过程文档(包括各类订购文档、合同、合作备忘录、获取权限等)。

#### 3.2 数字全文管理

全文管理在印本时代,体现为馆藏管理。数字信息

环境下开放资源、e-only资源和回溯资源的大量出现,需要集中管理和存贮原分散在各单位的电子全文。中心要建立统一的全文管理系统,建立全文文件的统一命名机制,并提供标准化接口方便其他系统调用全文文件。在元数据层面能输出符合统一文献元数据标准的元数据,供元数据集成加工管理系统调用。开放资源系统和回溯系统均应增加全文和元数据调用接口。

#### 3.3 元数据管理

文献元数据管理包括书目数据管理、文摘元数据集成加工管理、元数据增值计算,对应业务布局图3中的③和④部分。

##### 3.3.1 书目元数据集成管理

书目元数据管理任务是对各来源资源进行登记、编目、规范。书目元数据管理的范围包括中心订购的印本文献、开放资源和数字资源,以及第三方图书馆OPAC数据和第三方元数据,收集数据库描述信息和数据库收录资源信息,通过合作或其他方式获取其他机构所购买的数据库资源、机构IP地址范围等信息,建设形成支持中心发现系统的资源调度知识库,为用户提供情景敏感服务。书目元数据处理流程见图4。

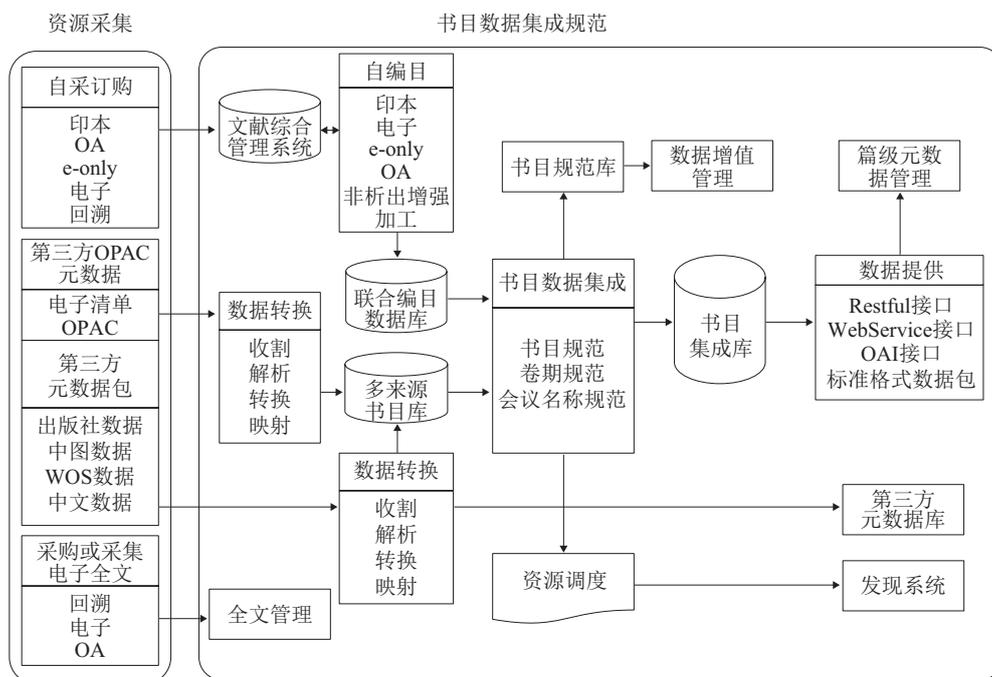


图4 书目元数据处理流程

### 3.3.2 篇级文摘数据集成加工管理

文摘元数据集成加工管理包括文摘元数据和引文元数据，并汇聚融合第三方的文摘元数据，建立与书目元数据系统处理结果的关联关系，形成统一的元数据库集成库。重点建立健全多来源元数据加工、匹配融合

机制，基于新的统一文献元数据标准规范，建设多源数据驱动的新型文献元数据分布式协同加工管理平台与人工弱干预数据融合工作系统，构建完整的数据生命周期管理体系，文摘元数据处理流程见图5。形成的元数据集可先期提供给发现系统并同时进入元数据增值计算环节进一步处理。

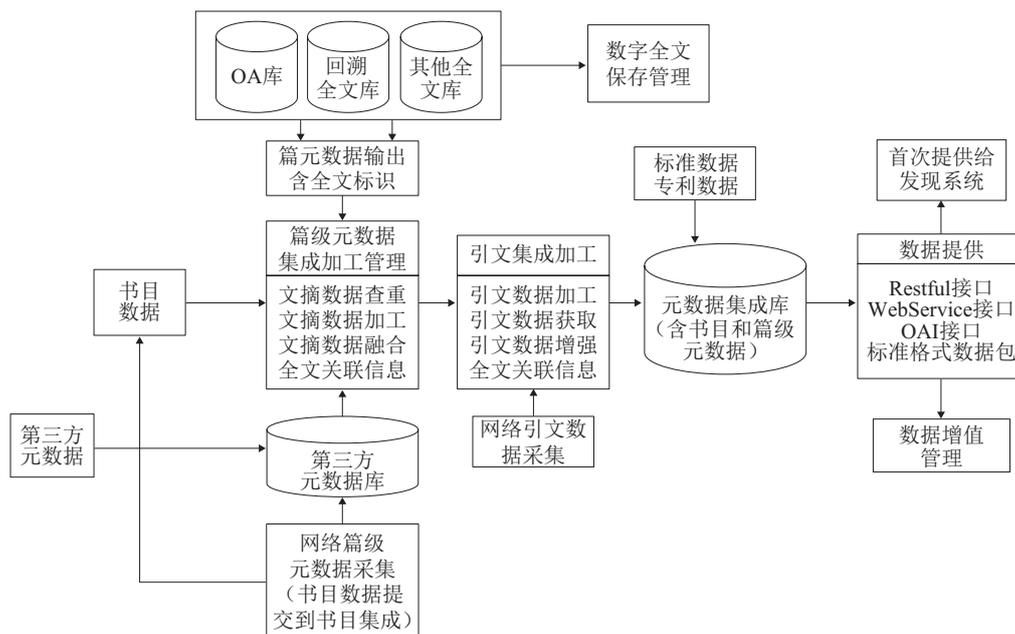


图5 文摘元数据处理流程

### 3.3.3 元数据增值计算

元数据增值计算包括名称规范、引文关系计算和主题标引，形成增值和富含语义的元数据库集成库。同时分析和计算用户访问行为数据，形成用户画像数据集。元数据增值计算和主题标引见图6。其中名称规范要实现数据库中各类名称（包括期刊名称、人名、机构名称、会议名称和基金名称）的规范处理，形成规范化的名称数据库。引文关系计算包括共现关系、被引次数、同被引关系计算等。开展基于STKOS词表的文献主题标引，不断提高主题标引的准确度。研究部署对文献内容中实体数据的抽取和管理，具体包括对文献内容中的概念、术语等命名实体进行识别和标引，对物种名称、基因、蛋白质、实验设备、科学数据DOI等实体信息进行识别、语句特征抽取和类别标注，为下一步关联科学数据建立基础，支持中心智能检索、关联分析和分析评价功能的实现。

### 3.4 服务系统建设

用户服务系统的建设以知识发现为目标，建立知识与知识、数据与数据、用户与用户、知识（数据）与用户之间的关联、计算及聚合关系，实现多载体、多类型、多来源资源统一集成揭示和多渠道获取，通过增值数据与关联计算结果，帮助用户发现相关的用户和知识信息，提供深层次的知识化服务。用户服务系统基本框架结构见图7。

发现系统为用户提供全面的文献发现、定位、获取服务，根据用户身份和权限获取相应的电子全文或原文传递服务。构建全文链接模块、原文传递模块和用户管理模块。分析评价系统包括引文分析、专利分析、学科分析、专题门户等，与发现系统在文献元数据层面保持一致性并能相互调用和关联。特色服务中的名称服务、数据服务和参考咨询服务与发现系统保持紧密联系，参考咨询服务嵌入发现系统中，用户可随时向服务人员发出咨询请求。

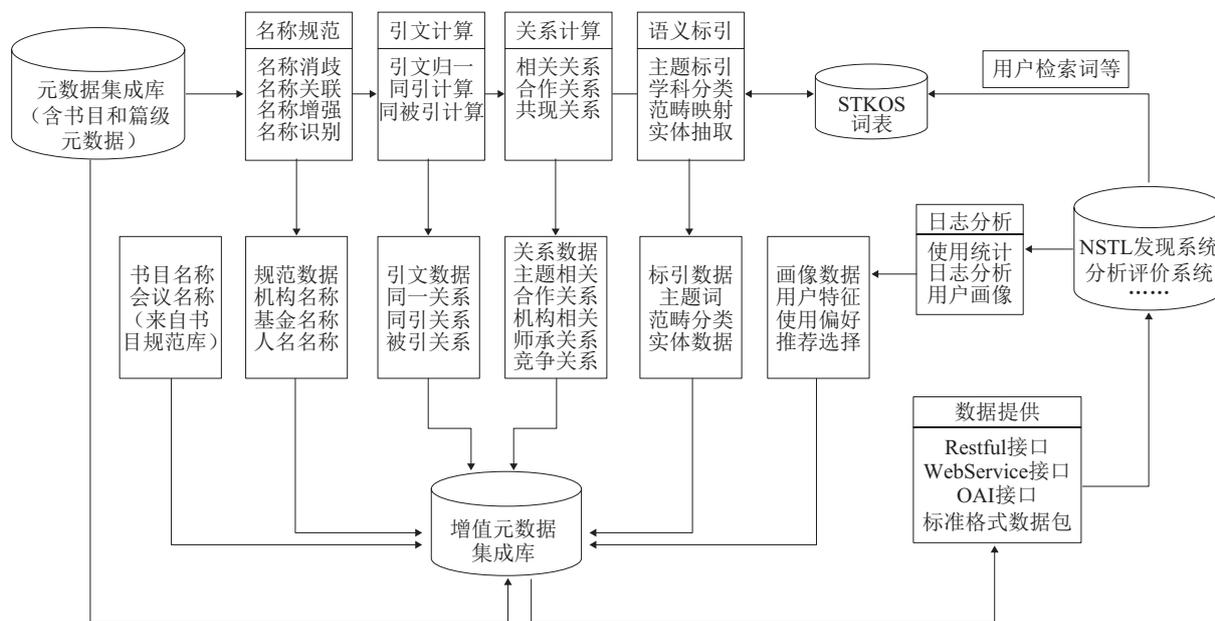


图6 数据增值计算和主题标引

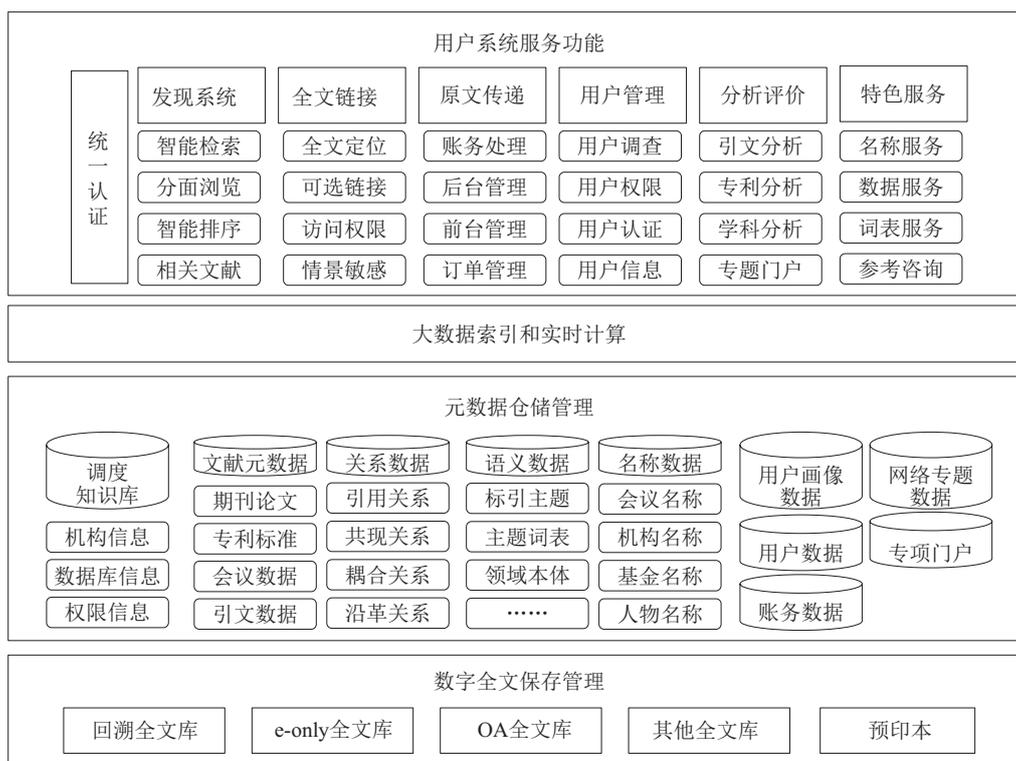


图7 用户服务系统基本框架

用户服务系统应能支持多种终端访问，包括PC端、Pad端和手机端的访问。加强用户注册和不同层次的用户身份认证，为非注册用户和注册用户提供差异化服务，通过统一认证系统，注册用户可以根据权限访问不

同的系统和服务。加强与成员单位、合作单位的身份统一认证功能，方便用户使用中心系统和服务。建立注册用户的信誉体系，为用户提供个性化深层次服务。

### 3.5 相关支撑工作

中心业务系统和流程的建设运行需要安全可靠的网络和统一的计算环境。各个业务系统建设要遵循中心发布的元数据制订指南和中心统一文献元数据标准,所有标准均要求在中心元数据登记系统中注册,并按照国家对网络和系统的有关安全规定,设计完善的网络安全和系统安全体系,开展常规的系统检测和修补工作。

## 4 总结

中心业务流程再造和系统建设方案经过多次专家讨论和修改,确定了最终方案并在2019年开始逐步实施,各成员单位根据方案申报相关任务,经评审后确定任务并由各项目团队开发建设。在推进建设过程中规划了系统建设路线图,确定各个系统的建设顺序和路径,并从组织机制上保证中心业务流程重组的效果,成立总体组和专家咨询组,负责审定重大建设问题和业务流程。目前方案中的资源综合管理系统、文摘元数据集成加工管理系统、名称规范、引文计算、主题标引系统已初步开始测试应用,书目元数据集成管理系统、资源调度系统正在开发,发现系统已开始提供检索服务。支持大数据管理和计算的基础技术环境已开始部署。2020年将启动分析评价系统、专题服务系统和用户统一认证以及实名注册模块的建设,已形成新业务流程雏形。中心业务流程再造和系统建设完成后将极大提高中心业务管理水平,增强为用户提供知识服务的能力,为中心创新发展打下坚实的基础。

### 参考文献

- [1] 国家科技图书文献中心十三五发展规划 [R]. 内部文件, 2015年.
- [2] 国家科技图书文献中心十四五发展规划 [R]. 内部文件, 2020年.
- [3] Library of Congress. The FY2019-2023 Strategic Plan of the Library of Congress [EB/OL]. [2019-10-15]. <https://www.loc.gov/strategic-plan>.
- [4] LIBER. Research Libraries Powering Sustainable Knowledge in the Digital Age: LIBER Europe Strategy 2018-2022 [EB/OL]. [2019-10-15]. <http://libereurope.eu/wp-content/uploads/2017/11/LIBER-Strategy-2018-2022.pdf>.
- [5] British Library Online Services. Living Knowledge: The British Library 2015-2023 [EB/OL]. [2019-10-15]. <https://www.bl.uk/britishlibrary/~media/bl/global/projects/living-knowledge/documents/living-knowledge-the-british-library-2015-2023.pdf>.
- [6] What is Horizon 2020? [EB/OL]. [2019-10-15]. <https://ec.europa.eu/programmes/horizon2020/what-horizon-2020>.
- [7] Why Plan S [EB/OL]. [2019-10-15]. <https://www.coalition-s.org/why-plan-s/>.
- [8] The STM Report [EB/OL]. [2019-10-15]. [https://www.stm-assoc.org/2018\\_10\\_04\\_STM\\_Report\\_2018.pdf](https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf).
- [9] 吴建中. 推进开放数据 助力开放科学 [J]. 图书馆杂志, 2018, 37(2): 4-10.
- [10] 张晓林. 颠覆性变革与后图书馆时代——推动知识服务的供给侧结构性改革 [J]. 中国图书馆学报, 2018, 44(1): 4-16.
- [11] NCBI. PubMed [EB/OL]. [2019-10-15]. <https://www.ncbi.nlm.nih.gov/pubmed>.
- [12] CSA SciFinder [EB/OL]. [2019-10-15]. <https://scifinder.cas.org/>.
- [13] BASE (Bielefeld Academic Search Engine) [EB/OL]. [2019-10-15]. <https://base-search.net/>.
- [14] Open Knowledge Maps: A visual interface to the world's scientific knowledge [EB/OL]. [2019-10-15]. <https://openknowledgemaps.org/>.
- [15] DOMS A, SCHROEDER M. GoPubMed: exploring PubMed with the Gene Ontology [J]. Nucleic Acids Research, 2005, 33: W783-W786.
- [16] Clarivate. Derwent Innovation [EB/OL]. [2019-10-15]. <https://clarivate.com/derwent/solutions/derwent-innovation/>.
- [17] NSTL用户需求和用户行为分析报告 [R]. 内部文件, 2019年.

### 作者简介

沈仲祺, 男, 1944年生, 高级工程师, 国家科技图书文献中心副主任, 研究方向: 数字图书馆。  
张建勇, 男, 1965年生, 硕士, 研究馆员, 通信作者, 研究方向: 数字图书馆, E-mail: zhangjy@mail.las.ac.cn。  
曾建勋, 男, 1965年生, 博士, 研究馆员, 研究方向: 数字资源管理。

Design of Business Process Reengineering and System Construction of National Science and Technology Library

SHEN ZhongQi<sup>1</sup> ZHANG JianYong<sup>2</sup> ZENG JianXun<sup>3</sup>

(1. National Science and Technology Library, Beijing 100038, China; 2. National Science Library, Chinese Academy of Sciences, Beijing 100190, China; 3. Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: In this paper, authors analyzed the problems existing in the system and business process of National Science and Technology Library in the new digital information environment, to reconstructed the NSTL's overall business model and key activities of the resource construction, data management, system services, special business based on user demand. To discuss the planning strategy and projects supported knowledge-based service capacity of NSTL.

Keywords: Business Model; Workflow; Operational Planning; Service System

(收稿日期: 2020-04-10)

## ■ 书 讯 ■

# 《汉语主题词表》

《汉语主题词表》自1980年问世以后,经1991年进行自然科学版修订,在我国图书情报界发挥了应有作用,曾经获得国家科学技术进步二等奖。为适应网络环境下知识组织与数据处理的需要,由中国科学技术信息研究所主持,并联合全国图书情报界相关机构,自2009年开始进行重新编制工作,拟分为工程技术卷、自然科学卷、生命科学卷、社会科学卷四大部分逐步完成。目前工程技术卷和自然科学卷已出版。

《汉语主题词表(工程技术卷)》共收录优选词19.6万条,非优选词16.4万条,等同率0.84,在体系结构、词汇术语、词间关系等方面进行了改进创新。《汉语主题词表(自然科学卷)》共收录专业术语12.4万条,包含数学、物理学、化学、天文学、测绘学、地球物理学、大气科学、地质学、海洋学、自然地理学等学科领域,收词系统、完整,语义关系丰富、严谨,每条词汇都有相应的学科分类号表现其专业属性,并与同义英文术语对应。同时,建立《汉语主题词表》网络服务系统,提供术语查询、文本主题分析、知识树辅助构建等服务。《汉语主题词表》可用于汉语文本分词、主题标引、语义关联、学科分类、知识导航和数据挖掘,是文本信息处理及检索系统开发人员不可或缺的工具。

《汉语主题词表(工程技术卷)》已于2014年由科学技术文献出版社出版,分为13个分册,总定价3 880元。

《汉语主题词表(自然科学卷)》已于2018年5月由科学技术文献出版社出版,分为5个分册,总定价1 247元。两卷均可分册购买。