

开放信息环境:学术图书馆信息资源建设的重定义与再造

孙 坦

摘 要 本文从数据密集型科学研究与 e-Science、学术交流模式变革与开放获取、语义网与知识表示技术三个方面分析开放信息环境的形成与内涵,指出学术图书馆信息资源建设面临开放信息环境的要求和数字图书馆服务边缘化的挑战,围绕学术图书馆如何构建开放信息资源体系和构建开放知识资源体系两个关键问题,提出学术图书馆资源建设模式转变的目标是构建网络开放信息环境中的信息(知识)资源设施 Information Commons。为此,需重构业务布局,重新定义业务交互模式,重新定位资源建设人员的角色。图3。参考文献19。

关键词 开放信息环境 信息资源建设 学术图书馆 重定义

分类号 G250

Open Information Environment: Redefinition and Recreation in Information Resource Construction of Academic Library

Sun Tan

ABSTRACT The article analyses the formation and definition of open information environment from data-intensive scientific research and e-Science, the changes of scholarly communication and open access, semantic web and knowledge representation technology, discusses the needs and challenges of open information environment and marginalization in services of digital library for the information resource construction of academic library, provides the goal of shifting the resource construction model in academic library——constructing information commons in open information network from two key points about how to construct and reengineer open information system. For this purpose, it is necessary to reconstruct work arrangement, redefine patterns of operation exchange and reposition the role of people in resource construction. 3 figs. 19 refs.

KEY WORDS Open information system. Information resource construction. Academic library. Redefinition.

随着信息技术的深入发展,人类科学研究又一次面临范式的转变,e-Science 环境已现端倪,数据密集型科学研究第四范式呼之欲出。同时,商业出版模式的垄断趋势对学术交流效率的束缚催生了开放获取运动的蓬勃发展,语义技术的不断成熟则推动信息知识化以及知识关联、重组和计算成为可能。因此,科学研究范式演进、开放获取运动和

语义技术的发展共同推动信息资源形态与传播利用方式发生根本性变化。学术图书馆的传统信息资源建设模式无法应对这种变化,必须重新定义信息资源建设的内涵,并通过对信息资源建设中资源定义、发现,评估选择,采集与存储,结构化描述与知识化组织等多个方面的改变与新建,以完成信息资源建设模式的再造。

1 开放信息环境的内涵与特征

1.1 数据密集型科学研究与 e-Science

2009 年微软研究院出版了《第四范式:数据密集型科学发现》(The Fourth Paradigm: Data-intensive Scientific Discovery),提出了一种新的科学研究范式:基于数据密集型计算的科学研究第四范式,系统介绍了地球与环境科学、生命与健康科学、数字信息基础设施和数字化学术信息交流等方面基于海量数据的科研活动、过程、方法和基础设施^[1]。在宏观浅层次上,我们可以这样理解数据密集型科研:在实验科学、理论分析和模拟三个科学范式基础上,在计算科学和数字信息环境发展的推动下,大量的科学数据和相应的技术允许科学家不再仅仅依靠实验、理论分析和模拟仿真开展科学研究,而是在综合运用上述方法的基础上,通过对科研过程中产生的海量科学数据(文献也是数据)进行计算开展科学研究。

e-Science 是一种科研模式和科研组织模式,也是一个基于网络、跨越时空限制的数字科研环境。数据密集型科学是 e-Science 的一个组成部分,它试图从研究方法上将理论、实验和计算仿真统一起来,而从科研资源上将设备、数据和人员集成起来,以便开展基于协同、共享和计算的大科学研究。2007 年,吉姆·格雷(Jim Gray)在美国国家研究理事会计算机科学与电信委员会发表著名演讲《e-Science: 科学方法的一次革命》^[1],首次系统阐述了科学研究第四范式,并指出目前每个学科正演变为两个分支:X-Info 和 Comp-X(其中 X 代表一个学科),即任何一门学科均向学科信息学和计算科学演化,如生物学正在演化为生物信息学和计算生物学。e-Science 为这种趋势提供了及时的基础设施环境。2000 年,英国研究理事会总干事约翰·泰勒(John Taylor)率先提出了“e-Science”概念。2001 年,英国 e-Science 核心计划正式实施,成立了国家 e-Science 中心、七大研究理事会 e-Science 中心和十大地区中心,启动了 300 个以上的 e-Science 项目。同时,欧盟也通过第五研究框架计划(The

Fifth Frame Work Programme,FP5)启动了 e-Science 建设,并一直持续到第六和第七框架。美国也通过国家科学基金会(NSF)和美国能源部(DOE)资助启动 e-Science 计划(e-Infrastructure)^[2]。

数字网络技术和计算技术的发展推动了数据密集型科学研究和 e-Science 的诞生与强力崛起,为人类科学研究带来了革命性改变。之所以称之为“革命性改变”,不仅因为这是科学研究范式的转移,也因为这种改变不仅仅体现在科学研究方法和模式上,而是覆盖全局,包括学术交流模式和学术信息生态环境也随之发生了根本性改变,广泛开放、零时差传播和语义互操作的知识资源生态环境正在形成。

1.2 学术交流模式变革与开放获取运动

如图 1 所示,吉姆·格雷(Jim Gray)在那篇关于科学研究第四范式的著名演讲中也讨论了即将到来的学术交流革命,认为互联网能把所有科学数据与文献联系在一起,创建一个数据和文献能够交互操作的世界,以提高科学的“信息速率”,促进研究人员的科学生产力^[1]。

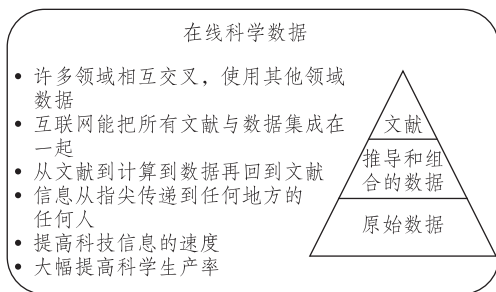


图 1 科学数据与文献的交互操作

美国学者 J. M. Hurd 研究了科学交流的变化,在 Garvey/Griffith 模型的基础上提出了 2020 年的学术信息交流模型^[3]。该模型体现了科学交流系统的演变,包括交流的基本方式的变化和现有的基于纸本载体交流系统中所不存在的新的功能的出现,主要讨论了出版机构和信息服务商的角色转变。知识界及专业学术出版社协会(ALPSP)提出的未来的学术交流模型^[4],则指出未来的学术交流系统,应该考虑以数据集、模拟、软件和动态知识的

表述等为信息单元,能够对复杂文档进行管理,将不同形式、不同区域的学术交流产物有机地整合起来。

美国国家科学基金会(NSF)在其2007年的报告《学术交流的未来:为学术互联空间建立基础设施》中又提出了一个名为“学术互联空间(Cyber-scholarship)”的发展路线图^[5],期望整合原始数据以及各种来源的新模型,利用最新的工具去分析、可视化,以及模拟复杂的相互关系。建立更普遍、更广泛、交互性和功能性更适合研究机构的数字化环境,包括人、数据、信息、工具、仪器以及高水平的计算操作、存储和数据传输能力。建立有效的执行组织,克服跨学科合作中的阻碍,使新型的科研组织和支撑环境服务于个人、团体和组织,改革他们的研究内容、方式以及参与人员。

众多关于新的学术交流模式的讨论均体现出一个共同特征:在数字开放信息环境中如何提高学术交流中的“信息速率”。以商业出版和同行评议为主要特征的传统学术交流模式严重背离了数字信息环境与新型学术交流模式的要求,开放获取运动的出现与蓬勃发展成为必然。2001年12月,“布达佩斯开放获取倡议”(BOAI)正式发布。截止到2012年底,全球已有8500多种开放出版学术期刊和3340个开放获取知识库,开放获取论文已经达到4000万篇^[6]。2011年,布达佩斯宣言相关组织再次召开会议,明确提出10年后在世界上的任何国家或地区任何学科领域的同行评议学术论文均实现开放存储与开放获取^[7]。10年来开放获取的典型变化是包括商业出版机构在内的相关各方从抵制走向开放,并在适应数字信息环境新型学术交流模式与尊重各方权益的平衡中探索出很多有效的举措。目前,在基于现有同行评议的开放获取出版方面主要有三种模式:①作者付费模式,如Springer出版社的BioMed Central(BMC)对250种同行评议期刊实行开放获取出版,作者付费,读者免费访问、下载和传播^①。②科研资助机构通过法

案强制存缴模式,如美国国家生物技术信息中心(NCBI)创建的PubMed Central(PMC),完整收录期刊972种,其中美国国立健康研究院(NIH)资助论文的期刊301种^②。③开放出版资助联盟模式,如国际粒子物理开放出版资助联盟SCOAP3(Sponsoring Consortium for Open Access Publishing in Particle Physics)通过全球化合作资助粒子物理领域12种学术期刊开放出版,实现该领域内90%以上的学术论文开放获取^[8]。其中,作者付费模式和开放出版资助联盟模式属于金色OA(GOLD OA),而机构知识存缴模式则属于绿色OA(GREEN OA)。

1.3 语义网络与知识表示技术

新型学术交流模式与开放信息环境为数据密集型科学研究提供了坚实的基础,但要满足知识计算的要求,则必须有语义网络和知识表示与计算技术的支持。

2008年,Mills Davis提出网络发展趋向于两个基本态势,即一方面从资源内容层面上知识关联和推理功能逐步增强,网络资源表示方式呈现从无语义或惰语义表示到强语义表示和推理能力逐步增强的趋势。另一方面从人类参与的层面上网络与人的互动逐渐增强,最终趋向于类似于人类理解和学习在网络。语义网时代,则以富含语义的个体集为典型支撑,在人工智能、语义搜索、语义网站、语义平台、智能代理等方面得以发挥,使机器能够实现逻辑推理的操作^[9]。

面对开放信息环境中数据资源体量、复杂性和异构性的增加,语义方法成为解决异构互操作的必然选择。Mills Davis将知识表示方式分为目录和分类法、词表、概念模型、逻辑理论和量子物理、价值论等五个阶段。目录和分类法阶段实现了语法层次上的互操作,词表阶段实现了结构化的互操作,概念模型、逻辑理论和量子物理、价值论阶段目的是实现语义上的互操作。目前处于概念模型阶段,以RDF/S、UML、OWL为研究内容。这个观点

① <http://www.biomedcentral.com>

② <http://www.ncbi.nlm.nih.gov/pmc>

充分表明了语义网和知识表示在开放信息环境中支持异构互操作的必要性。Peter Fox 和 James Hendler 指出,e-Science 基础设施开发者越来越需要基于语义的方法、工具和中间件,从而能够推动科学知识的建模,进行基于逻辑的假设检测,实现基于语义的数据集成,构建各类应用组合,并能够对来源于不同科学领域和系统的数据实现集成化的知识发现和分析^[10]。

2 图书馆信息资源建设面临新要求与挑战

开放信息环境中,新的科学研究范式和学术交流模式对信息利用的要求是:在 e-Science 基础设施(网络、工具和中间件等)支撑下,对开放综合、异构互联的各类信息对象进行基于知识单元和语义互操作的集成分析与计算。Patrick Wendel 在英国 e-Science 先导项目 Discovery Net 中提出了“信息网格:动态数据集成”(InfoGrid: Dynamic Data Integration)的信息利用模式(见图2)^[11],支持用户按需访问各类数据源和结构化信息,并创建自己的数据集进行集成分析。

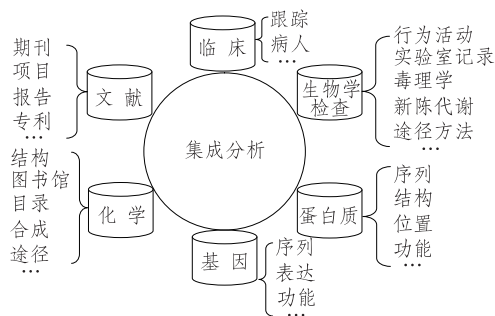


图2 动态数据集成

尽管图书馆界始终积极拥抱数字化网络化和知识组织系统,但是数字化商业信息资源仍是学术图书馆的根基,信息资源保障牢牢占据着学术图书馆服务的核心。开放获取运动正在将学术图书馆从传统服务中驱离,而第三方在关联检索和语义知识发现方面的飞速进展更迅速填补着新型科学研

究与学术交流的需求空白,并日益侵蚀着学术图书馆服务创新的新空间。因此,第二代基于集成服务的数字图书馆提供的馆藏数字信息资源集成检索与获取服务面临着被迅速边缘化的严峻挑战。

2.1 开放信息环境颠覆图书馆“馆藏”根基

随着开放获取运动的蓬勃发展,OA 期刊、机构知识库和开放出版等将导致图书馆在学术期刊保障方面的垄断地位迅速弱化。学术专著作为学术图书馆保障服务的另一主要部分也承受着数字出版与营销模式变革的双重挑战。截至 2008 年,全球 STM 图书出版六强机构(Springer, Taylor & Francis, Wiley, Elsevier, CUP, OUP)已经占据了全球 STM 图书出版的约 70%^[12],其数字化比例达到约 80%,占每年出版 STM 图书总数的 50% 左右^[13]。在电子图书的营销模式方面,出版机构不再依托集成商生产和发布。目前,六大出版机构已经全部退出集成商平台,在服务上针对移动设备不断增强电子图书使用功能,在营销模式上尽管图书馆仍是主要市场,但正迅速倾向于终端销售模式。因此,未来学术图书馆很可能面临无资源可采购的窘境。

另一方面,越来越多的综合科技信息资源开放呈现在互联网上,包括第三方开放资源集合、种类繁多的综合科技资源与服务登记系统、开放会议、开放课件、咨询报告、技术报告、行业(产业)报告、专利与标准信息以及经济社会和法律信息等。此外,学术社交网络和社会媒体提供了大量用户创建的信息资源,知名的学术社交网络 Academia. edu 目前已有 200 万科研人员注册,约占全球学者的 10%,每天增加用户约 4000 人,月访问量达到 450 万次^①。同时,随着 e-Science 和大科学工程的发展,越来越多的科学数据实行开放服务,科学数据保存服务中心应运而生。上述开放的信息资源与科学数据对传统图书馆的资源建设困境无疑是雪上加霜,但对重构图书馆信息资源体系而言却是难得的机遇。

① <http://www.academia.edu>

2.2 数字图书馆服务的边缘化挑战

目前,主流数字图书馆仍处于基于文献资源的集成服务模式:信息资源体系仍局限于期刊、图书等传统信息资源种类;在信息描述与组织方面,基于描述性元数据的文献集成与系统互操作仍是主流,信息资源主要呈现为 PDF 文档和自由文本形式的文献对象,缺乏结构化、数据化、语义化描述与知识组织;在信息服务方面,主要提供基于元数据的关键词检索和部分概念检索,缺乏语义搜索与知识发现服务。

越来越多的第三方知识服务产品日益侵蚀着学术图书馆的服务拓展空间。从 Web of Science 到 Scopus,从 TDA 到 Aureka,从 Pubmed 到 NPG,全方位地对学术图书馆的文献服务、信息检索、计量分析、关联发现等各层次的服务带来巨大冲击和替代。

(1) 增强出版、语义出版和融汇型期刊。面对开放获取和新型学术交流模式的影响,商业出版机构从早期的抵制抗拒逐渐转变为积极面对和把握新的发展机遇,启动了出版的创新转型。

自然出版集团(NPG)在学术交流未来方向中明确提出了语义出版的模式^[14],Nature 杂志将改变现有 PDF 格式的出版模式,将所有文章转换为一个论文元数据库和一组结构化数据集,表现为基于 RDF 描述的结构化的、可交互和查询的图、表和文本。传统的“检索”已消失,信息的发现是通过结构化的、交互化的、可查询的图表与文本实现,服务已更多地转为科学信息内容的分析、可视化与分发功能。英国皇家化学学会(RSC)期刊 Molecular Biosystems 则通过增强 HTML 标记机制,对文本中被化学术语数据库收录的词汇和来自基因、序列和细胞知识本体的词汇进行高亮显示,并提供外部链接^[15]。Elsevier 启动了“Article of the Future Project”,力图实现科学论文格式的彻底变革,并已经推出了 SciVerse 和科学工作流解决方案 Reaxys

Xcelerate。增强出版和语义出版通过结构化和语义化实现了基于知识单元的语义交互,完成了从传统期刊向融汇型期刊的转型。

同时,以爱思唯尔(Elsevier)和汤森路透(Thomson Reuters)为代表的商业出版机构和数据库商更加致力于对自身海量信息的开发利用,通过关联、计量和挖掘直接面向用户提供科研评价、分析和战略咨询服务,对学术图书馆的传统文献服务带来巨大挑战。

(2) 关联检索、语义丰富与知识发现

随着新型科学研究模式在生命科学领域的率先应用,在以 W3C 为代表的各类结构化、数据化、语义化标准的支持和语义网技术的推动下,以知识关联、重组和语义互操作为特征的新型知识产品与知识服务也集中呈现在该领域,如 PubMed Entrez、NxOpinion、NCIt、NCBO、VIVO 等。

VIVO 是美国康奈尔大学开发的一个语义网应用项目,用于支持生命科学领域研究人员的相关知识发现^①。该项目通过一个广泛适用于科研领域的科研本体和生命科学领域本体支持对海量科研人员、仪器设备、科研活动和项目、科学数据、科研成果以及文献资源等进行结构化、语义化的知识表示与可视化展示,支持科研人员的关联发现。NCBO(The National Center for Biomedical Ontology)利用 NCBO Annotator 工具基于 NCI Thesaurus(参考术语本体)和连接到 UMLS 中的 200 多个本体,实现了对自由文本的大规模语义标注,使非结构化的自由文本数据变得结构化、标准化^②。为生物医学研究人员知识密集型工作提供支撑,提供在线工具及基于网络的资源(如 BioPortal),使研究人员可以在生物医学研究及临床实践的各个方面实现对不同本体的访问、审查与集成,搜索与特定的本体术语相关的生物医学资源^③。

新型知识产品与知识服务呈现出显著的开放化与个性化特征,普遍在开放信息环境或用户环境

① <http://www.vivoweb.org>

② <http://www.bioontology.org/annotator-service>

③ <http://bioportal.bioontology.org/ontologies>

中,支持科研人员在语义层面集成化构建、管理和分析各类知识资源。

3 学术图书馆资源建设模式再造

通过上述分析可以发现,在开放信息环境和数据密集型科学研究中,学术图书馆文献保障服务功能迅速被边缘化,必须重新定义学术图书馆在新型学术交流中的角色定位和贡献模式。

吉姆·格雷在演讲中批评了美国国家科学基金会(NSF)的数字图书馆项目全部聚焦于传统图书馆的元数据,而非真正的数字图书馆,并建议建设支持科学数据与文献及其他各类信息开放互连和语义互操作的新型数字图书馆,以美国国家医学图书馆支撑生物科学的相同方式,建立支撑其他科学的数字图书馆^[1]。张晓林教授提出开放获取、开放知识、开放创新推动开放知识服务模式,研究图书馆必须推进范式再转变,打造开放知识服务支持平台,构建由 Information Commons、Knowledge Tools Commons、Research & Creation Commons 组成的公共的 Knowledge Service Support Commons^[6]。他明确提出了将开放信息转换为可计算的开放知识的步骤:①支持开放信息的开放再使用;②支持开放信息的结构化;③支持结构化信息的语义化;④支持语义化信息的开放关联化。

学术图书馆信息资源建设模式重构的目标是:构建网络开放信息环境中的信息(知识)资源设施 Information Commons,即面向开放信息环境构建覆盖各类信息对象的开放信息资源体系,通过知识表示技术转换为可计算的开放知识资源体系,支撑面向用户驱动的开放知识服务的数字知识资源生态环境(见图3)。

3.1 构建开放信息资源体系

面向开放信息环境,学术图书馆必须打破传统馆藏建设模式,创新资源建设理念,在整个数字空间(而不仅局限于图书馆)发现、评估、登记、采集、描述和组织各类信息资源,植根于用户环境构建开放综合、支持用户定制集成的信息资源体系。

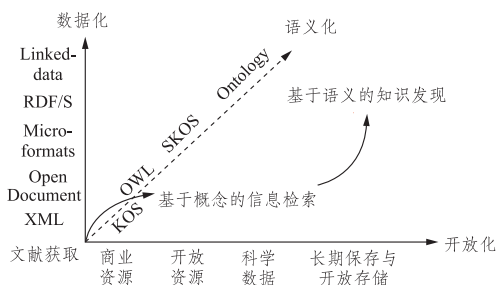


图3 学术图书馆信息资源建设模式再造方向

(1) 开放化、综合化。在继续夯实商业信息资源体系的基础上,向开放信息资源扩展,包括开放期刊、机构知识库、开放会议、开放课件、开放报告、行业信息、产业信息、科学数据、科研工具、仪器设备以及广泛的经济社会和法律信息资源等。

(2) 评估与质量控制。针对缺乏同行评议的学术资源以及广泛开放的网络信息构建评估体系,建立评价指标体系、遴选标准与质量控制机制,包括建立不同类型开放资源的基于内容质量、技术约束、访问管理、利用许可、用户反馈等方面的评价指标以及开放协议、访问接口、知识产权和法律限定、存储约束等。

(3) 开放资源登记、发现与集成。规范化地集成登记各类开放资源是一种有效解决开放资源发现的模式,如开放获取期刊目录 DOAJ,开放获取图书目录 DOAB,澳大利亚研究数据发现服务(Research Data Australia),美国研究数据目录 Databib,英国经济与社会数据服务(Economic and Social Data Service),开放获取仓储登记系统 OpenDOAR 和 ROAR,综合开放学术资源登记系统 Informain 和 IESR 等。登记系统要统筹考虑资源类型的选择、资源的覆盖范围、资源的质量控制体系、遴选方案、规范的元数据描述和知识化组织、系统的服务机制、可持续发展机制等多种因素。

(4) 使用政策与协议。尽管以创作共用协议(Creative Commons)和数据共用协议(Open Data Commons)为代表的开放获取政策在不断试验、完善和发展,但各类开放资源在开放程度、方式和用途约束以及对竞争、隐私和公益保护的政策也不尽相同。使用政策与协议贯穿于开放资源的评估、遴

选、登记描述、集成组织和服务全过程,包括:①不同程度、方式和用途下的开放信息资源的利用许可政策(包括元数据、全文、引文、附加数据等的再利用约束);②开放资源获取途径、技术使用条件和限制条件与技术规则(元数据接口和内容获取接口、机器可访问的内容级别等);③开放信息资源获取和利用过程中多种层次上的标准和协议,如 Web service (SOAP、Restful)、SRU/SRW、OAI-ORE、OAI-PMH、SWORD、Open Data Protocol、SparQL 查询标准等。

(5)开放仓储与长期保存。在使用许可框架内推动开放资源本地化开放仓储与长期保存是保障资源可持续使用和计算挖掘等再利用的有效举措。该领域需要重点关注开放获取资源长期保存的资源体系构建、权益机制管理、技术保障机制等。

3.2 打造知识表示能力

打造知识表示能力,将开放信息资源体系转换为可计算的开放知识资源是重构学术图书馆信息资源建设的核心。目前,针对自由文本的知识表示与语义化路径主要包括:利用开放关联数据(LOD)实现语义关联、利用本体实现开放信息语义化、利用实体识别实现开放信息语义化和利用文本挖掘实现开放信息语义化等。其中,开放关联数据和本体标注的知识表示方式是学术图书馆资源建设重构的最佳选择,也是未来核心能力的主要体现。

(1)知识组织体系。利用知识组织体系实现信息资源的语义化表示主要包括两种模式:①基于传统知识组织体系扩展的语义网模式。例如,UMLS 通过定义语义类型并将集成的词表概念映射到语义类型构建语义网(Semantic Network)。STERNA (Semantic Web-based Thematic European Reference Network Application)则是利用 RDF 对信息元数据进行结构化描述,并映射到 SKOS 描述的知识组织体系构建参考网络支持异构资源的语义检索。②本体模式。主要是利用传统知识组织体系(叙词表、分类表、术语表等)构建本体,并对信息资源进行语义标注或注释,如联合国粮农组织的渔

业本体(FOS)、美国国家生物信息中心的生物医学本体(NCIt)和欧盟的本网网络项目(NeOn)等。

(2)概念识别与语义标注。开放信息资源转换为可计算的开放知识的关键是实现语义标注。目前,基于词表、本体等知识组织体系进行概念识别与语义标注及扩展的主要方法包括:①根据概念或属性关系进行扩展标注。②基于本体语义相似度计算获取相关的概念并生成新的标注^[16-17]。③根据已有的本体之间的映射,找到其他本体中的相关概念词进行扩展标注^[18]。基于本体的语义相似度计算方法包括:基于距离的语义相似度计算、基于内容的语义相似度计算、基于属性的语义相似度计算和混合式语义相似度计算^[19]。

(3)用户知识环境构建。支持用户知识环境构建是学术图书馆开放知识服务的重要内容,也是资源建设打造知识表示能力的重要组成部分。主要包括两个关键内容:①开放资源与服务配置引擎,支持用户对各类信息资源和知识资源的个性化定制集成,构建个人知识库。②开放知识组织引擎,支持用户定制、完善个性化的知识组织体系并实现对个人知识库的知识表示与关联,构建个人知识网络。

3.3 学术图书馆资源建设再造

面向开放信息环境知识服务的需求,学术图书馆资源建设必须重新构建业务布局,重新定义业务交互模式,重新定位资源建设馆员的角色。

(1)重构业务布局。传统资源建设业务社会化是学术图书馆资源建设谋求业务布局重构空间的有效途径。在传统资源建设业务转型基础上,重点围绕开放信息资源体系构建和打造知识表示能力构建新的业务布局。包括:信息资源开放利用政策研究,开放信息资源评估、登记与集成,知识组织体系建设与应用,知识表示技术开发与应用,用户知识环境建设,开放存储与长期保存。其中,能够融合资源质量、使用许可政策、技术规则、资源内容、使用环境与用途等全面描述开放资源的元数据标准规范、知识组织体系与知识表示将成为新的业务核心。

(2)重新定义业务交互模式。传统学术图书馆资源建设业务交互主要局限于“采、分、编、检、流”框架内的图书馆员与信息资源之间的交互,封闭于馆藏内部。新型的业务交互模式呈现出显著的开放化特征,扩展为三个部分:①开放信息资源体系构建过程中,资源建设人员在整个数字空间内与各类信息对象及权益主体的交互;②在知识组织体系建设与应用过程中与用户和领域专家的交互;③在用户支持环境构建中与用户和领域专家的交互。因此,开放、协同成为新型业务交互模式的主体。

(3)重新定位资源建设人员的角色。数据密集型科学研究需要培育数据科学家群体,而数据索引专家、图书馆学家都是数据科学家群体的组成部分。新型学术图书馆模式中,资源建设人员将扮演一个全新的角色:数据图书馆员。他们承担的主要职责包括:①以数据的视角发现、评估、登记、采集、描述与组织、知识表示与语义标注的审核与质量控制等;②与用户和领域专家协同创建、维护和应用词表、本体、参考网络等各类知识组织体系;③与学

科馆员、用户和领域专家共同创建和维护用户知识环境。

4 结语

无论怎样变化,学术图书馆保存和传播人类科学知识和学术成果的使命和职能无疑将会长期存在。但是,面向新型科学研究范式和开放语义环境的形成,学术图书馆信息资源建设必须更有作为,通过开放化、结构化和语义化知识关联体系构建,抢占新型学术交流模式的核心位置,打造学术图书馆开放知识服务的基础,支持数据型数字图书馆发展知识关联、计算、挖掘等知识服务的核心竞争力。

致谢:本文是在中国科学院国家科学图书馆资源建设部承担的战略扫描、开放资源、知识组织等多个项目基础上完成的,对项目组成员表示感谢。

参考文献

- [1] Hey T, Tansley S, Tolle K. 第四范式:数据密集型科学发现[M]. 潘教峰,张晓林,等,译. 北京:科学出版社, 2012. (Hey T, Tansley S, Tolle K. The fourth paradigm: Data-intensive scientific discovery[M]. Pan Jiaofeng, Zhang Xiaolin, et al, Trans. Beijing: Science Press, 2012.)
- [2] 孙坦. 数字化科研——e-Science 研究[M]. 北京:电子工业出版社, 2009. (Sun Tan. Digitalization study: e-Science[M]. Beijing: Electronics Industry Press, 2009.)
- [3] Hurd J M. The transformation of scientific communication: A model for 2020[J]. Journal of the American Society for Information Science, 2000, 51(14):1279-1283.
- [4] Warner S. The transformation of scholarly communication[J]. Learned Publishing, 2005, 18(3): 177-187.
- [5] Arms W Y, Larsen R L. The future of scholarly communication: Building the infrastructure for cyberscholarship[R]. Workshop Report (September 2007), National Science Foundation and Joint Information Systems Committee, 2007.
- [6] 张晓林. 开放获取、开放知识、开放创新推动开放知识服务模式——3O 汇聚与研究图书馆范式再转变[J]. 现代图书情报技术, 2013(2): 2-10. (Zhang Xiaolin. Open access, open knowledge, and open innovation pushes for open knowledge services——3O convergence and a new paradigmatic shift for research libraries[J]. New Technology of Library and Information Service, 2013(2): 2-10.)
- [7] Ten years on from the Budapest open access initiative: Setting the default to open[R/OL]. [2012-03-07]. <http://www.opensocietyfoundations.org/openaccess/boai-10-recommendations>.
- [8] 曾燕, 郑建程, 赵艳, 等. SCOAP3: 开放出版新模式及其影响[J]. 图书情报工作, 2013(1): 37-42. (Zeng Yan,

- Zheng Jiancheng, Zhao Yan, et al. SCOAP3: A new model for open access publishing[J]. Library and Information Service, 2013(1): 37-42.)
- [9] Davis M. Semantic wave 2008 report: Industry roadmap to Web 3.0 and multibillion dollar market opportunities[R/OL]. [2013-03-07]. http://www.isoco.com/pdf/Semantic_Wave_2008-Executive_summary.pdf.
- [10] Fox P, Hendler J. 语义 e-Science:在下一代数字化推动的科学研究中实现语义编码[M]// Tony Hey, Stewart Tansley, Kristin Tolle. 第四范式:数据密集型科学发现. 北京:科学出版社, 2012: 149-154. (Fox P, Hendler J. Semantic e-Science: Implementing semantic coding in the scientific research motivated by next generation of digitization [M]//Tony Hey, Stewart Tansley, Kristin Tolle. The fourth paradigm: Data-intensive scientific discovery. Beijing: Science Press, 2012:149-154.)
- [11] Wendel P. Discovery net: A UK e-Science pilot project for grid-based knowledge discovery services[C/OL]. Data Mining and Exploration Middleware for Distributed and Grid Computing, September 18-19, 2003.
- [12] What's new at Springer[EB/OL]. [2013-03-07]. http://www.metecbooks.ru/presentation/Springer_ebook_June_202006.ppt.
- [13] 黄飞燕. 专业电子图书市场发展趋势研究[D]. 中国科学院研究生院, 2008. (Huang Feiyan. Research on the growing trend of professional and electronic book market[D]. Graduate School of the Chinese Academy of Sciences, 2008.)
- [14] Timo Hannay. Current experiments & future directions in scholarly communication[EB/OL]. Nature Publishing Group. [2013-03-07]. http://old.cni.org/tfms/2007b.fall/CNI_Hannay.pdf.
- [15] Ginsparg P. 以数据为中心的世界中的文本[M]// Tony Hey, Stewart Tansley, Kristin Tolle. 第四范式:数据密集型科学发现. 北京:科学出版社, 2012:190. (Ginsparg P. Texts in the data-centric world[M]//Tony Hey, Stewart Tansley, Kristin Tolle. The fourth paradigm: Data-intensive scientific discovery. Beijing: Science Press, 2012:190.)
- [16] Lee W J, Raschid L, Srinivasan P, et al. Using annotations from controlled vocabularies to find meaningful associations[J]. Data Integration in the Life Sciences Lecture Notes in Computer Science, Volume 4544, 2007: 247-263.
- [17] Caviedes J E, Cimino J J. Towards the development of a conceptual distance metric for the UMLS[J]. Journal of Biomedical Informatics, 2004, 37(2): 77-85.
- [18] Euzenat J, Shvaiko P. Ontology matching[M]. Springer, 2007.
- [19] 孙海霞, 钱庆, 成颖. 基于本体的语义相似度计算方法研究综述[J]. 现代图书情报技术, 2010(1): 51-56. (Sun Haixia, Qian Qing, Cheng Ying. Review of ontology-based semantic similarity measuring[J]. New Technology of Library and Information Service, 2010(1): 51-56.)

孙 坦 中国科学院国家科学图书馆研究馆员, 博士生导师。

通讯地址: 北京中关村北四环西路 33 号。邮编: 100190。

(收稿日期: 2013-03-07)