

# 国家开放信息资源建设策略研究<sup>\*</sup>

杨代庆, 丁道劲, 周锐

(中国科学技术信息研究所, 北京 100038)

**摘要:** 期刊价格上涨、用户学术交流模式改变等因素促使开放信息资源规模不断增加, 逐渐成为一项重要的学术信息资源来源。从国际、国家/地区、商业出版社三个层面揭示了国内外大规模开放信息资源建设现状, 指出从国家层面推进开放信息资源建设必须首先定位好政府参与的切入点、制订统一的元数据标准、处理好开放信息资源与商业订购资源的关系以及合理规避版权风险, 并基于此, 从基础资源构建、信息组织与服务等方面详细论述了开放信息资源建设的实施策略。

**关键词:** 开放获取; 信息资源建设; 战略保障

**中图分类号:** G250

**DOI:** 10.3772/j.issn.1673-2286.2015.08.003

学术期刊价格上涨一直是图书馆与出版商争论的焦点所在, 科技期刊涨价幅度大大超过图书馆订购经费以及消费指数的同期增幅<sup>[1]</sup>, 高昂的订阅价格和涨幅甚至一度引起诸如图书馆界、学术界抵制Elsevier等事件, 加之电子期刊严格的使用范围限制使科技知识的广泛传播和公共保存受到严重威胁。为了打破这些限制, 国际科技界提出了科技信息OA的理念, 于2002年2月发布《布达佩斯开放获取计划》, 推动科技文献的OA, 即用户通过互联网免费阅读、下载、复制和传播作品<sup>[2]</sup>。经过十余年的发展, 开放概念从期刊、图书等传统出版物逐步向其它资源形态延伸, 互联网中出现了大量用户可直接免费浏览和下载的学位论文、技术报告、教学课件等内容, 极大丰富了科技信息来源渠道。不可避免的是, OA期刊、机构知识库和开放出版的蓬勃发展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<sup>[3]</sup>, 加之政治、技术等不稳定因素存在, 仅向用

户提供网站级别存储位置的服务手段已不能完全满足用户需求, 亟待从国家层面探索和深化面向开放信息环境的资源建设与服务新模式, 实现高质量开放信息资源的大规模集成与发现, 以提供稳定、持续的开放学术资源服务, 在文献资源以外提供更多基于数据的服务, 进一步保障我国学术信息资源的战略安全。

## 1 国内外大规模开放信息资源建设特征

随着OA运动的逐步深化, 越来越多的学术信息资源的开放程度逐步从绿色OA扩展到金色OA, 出版商纷纷向作者和机构推出OA出版服务, 在所有同行评议的期刊中开放获取文章已达到7.7%<sup>[4]</sup>, 开放信息资源规模增长明显。相应地, 开放信息资源的存储位置也不再局限于独立的机构知识库, 资源建设逐步受到更为广泛的重视, 国家/地区性乃至全球性的开放信息资源集成平台越来越多。大型出版平台伴随OA出版业务的扩张也积累了一大批开放信息资源。

<sup>\*</sup> 本研究得到国家科技图书文献中心“NSTL开放资源建设——采集功能需求与方法策略和示范性采集加工”(编号: 2015XM23)资助。

## 1.1 全球性开放信息资源集成与发现平台发展多元化

伴随开放信息资源的数量增长以及来源的多样化,集成开放信息资源的元数据乃至全文是保存资源、提高资源利用率的必然选择。期刊、图书等开放信息资源均已建有大规模的资源集成平台。截止目前,开放期刊集成平台DOAJ可检索的开放期刊已超过一万种<sup>[5]</sup>;开放学位论文门户OATD中共收录了2,540,937条开放学位论文记录;开放图书平台DOAB中收录的图书数量相对较少,涵盖107家出版社的3,123种经过同行评议的学术图书。除非盈利性的开放信息资源集成平台外,Summon、EDS等商业性发现系统也集成了大量的开放信息资源。在我国,开放信息资源建设以开放期刊为主,建设规模较大的开放期刊集成平台包括中国教育图书进出口公司的Socolar开放信息资源一站式检索平台以及中国科学院文献情报中心(以下简称“中科院”)新近推出的OA期刊采集服务体系GoOA。此外,国家科技图书文献中心(以下简称“NSTL”)、中科院还建有开放会议录集成平台。从资源角度,建设模式分为元数据集成和全文集成两种模式<sup>[6]</sup>,建设粒度分为篇级和母体级两个层次,但是受版权限制、发展目标、技术开发等因素影响,较多开放信息资源集成平台并不实际存储资源全文,而是以元数据整合方式提供检索服务,将用户直接指向实际存储该资源的机构知识库、出版社平台等,资源长期保存功能不明显。同时,开放信息资源元数据来源较多,如OATD的元数据分别来自全球约1000家高等院校和研究机构的机构知识库、地区/国家级学位论文数据库以及OCLC Worldcat的电子学位论文编目数据,少部分数据是通过OAI-PMH协议收割获取<sup>[7]</sup>,这就导致平台在服务过程中可能存在元数据项缺失、资源链接失效等问题。

## 1.2 国家/地区学术信息资源存储库趋向开放利用

从国家角度而言,多数学术成果的形成离不开国家项目的支持,单纯集成元数据的方式也难以完全保证资源的长期保存需求,因此许多发达国家十分重视全文的长期保存,这一点在学位论文建设方面体现得尤为明显,例如加拿大国家图书馆的Theses Canada门户<sup>[8]</sup>、大英图书馆的电子学位论文在线服务门户ETThOS等。

这些国家级的学位论文库中的相当一部分内容已可供网站用户全文浏览/下载,在ETThOS中迄今可检索超过40万篇英国高等院校的博士论文,其中16万篇学位论文可以直接或通过机构知识库获取全文<sup>[9,10]</sup>。除了以资源类型单独建设外,国家/地区层面学术信息资源的另一个开放体系是以国家资助项目为范畴展开的,国家要求公共资金资助项目成果的开放呈缴。例如在医学研究领域,PubMed Central作为美国国家健康研究院(NIH)资助成果的存储库,收藏了美国在生物医学领域的重要成果,并要求在论文正式出版后的12个月向公众开放<sup>[11]</sup>。在我国,中国科学院与国家自然科学基金委员会于2014年5月15日发布了OA的政策声明,要求相应资助项目全部或部分资助的科研项目产出的论文,在学术期刊上发表研究论文的作者应在论文发表时,将同行评议后录用的最终审定稿,存储到国家自然科学基金委员会/中科院所属机构的知识库,不晚于发表后12个月OA。相比于全球性的开放信息资源集成与发现平台,国家级的开放信息资源存储库无论是在元数据规范还是全文存储方面均具有一定的优越性,但是它一般以国家/地区或项目资助来设定资源采集范围,可建设范围相对有限。

## 1.3 商业出版平台开放信息资源规模随出版社OA出版业务扩张增加

日趋开放的信息环境、政府政策支持使越来越多的出版商参与到开放信息资源传播中,参与方式从允许出版论文存储在机构知识库中扩展到OA出版,使出版平台中存储的开放信息资源数量逐步增加,特别是学术期刊的OA出版,据汤森路透Web of Science的数据显示,2013年约有11%的科研论文发布在OA期刊上,而在2002年该比例仅有2%<sup>[12]</sup>。目前,开放出版规模较大的出版社主要有BMC、HighWire、PloS one,其中BMC在2010年被德国大型学术出版商Springer收购,可见商业出版机构对OA出版业务的高度重视。美国ProQuest公司作为一家专业出版博硕士学位论文出版商也以OA出版方式在PQDT Open网站中提供2万余篇学位论文,网络用户可免费检索和获取全文。学位论文是否以OA方式出版由作者所在高校决定,每篇学位论文的OA出版费用为95美元<sup>[13]</sup>。不言而喻,商业出版商的出版平台仅仅存储自有出版物,同时出于商业利益考虑,较多以OA期刊论文与商业订购资源同时存储在同

一平台中,但是平台未向用户提供单独检索开放信息资源的功能,用户只能在检索结果中逐一浏览单篇论文是否标有开放标识来查找OA论文,这在较大程度上限制了开放信息资源的利用。

总体而言,OA期刊在开放信息资源领域发展相对较为成熟,因此它也是我国参与建设较多的资源类型,在其他类型的开放信息资源领域,我国现有的资源建设实践仍十分有限。从全球来看,开放信息资源建设多见于期刊、图书、学位论文领域,技术报告等其他类型的开放信息资源仍多以机构为单位进行组织揭示,由于互联网中的显示度有限,对用户知识背景和检索能力要求较高。因此,面向多类型的开放信息资源建设仍具有一定的探索价值,在考虑各类型资源特征的基础上,亟待处理好元数据收割与全文存储、开放信息资源与商业订购资源、开放信息资源与国家呈缴资源等之间的关系。

## 2 开放信息资源建设的关键问题分析

在传统的学术信息传播模式中,无论是出版商的盈利模式还是作品最终的版权归属均已形成行业内通行的规则,但是随着项目资助者(政府)对科研成果的保存与公共利用的重视,作者对自身作品希望保留的权利越来越多,开放信息环境催生出新的学术信息传播格局,除了在微观层面构建开放信息资源的发现、获取以及遴选机制外,从宏观层面确立开放信息资源存储政策、处理开放信息资源与其他资源的关系以及做好版权风险控制是开放信息资源建设得以有序开展的必要条件。

### 2.1 推动政府有效参与OA以及开放信息资源建设

政府资助科技项目是一种由政府出资购买的科技公共产品或公共服务<sup>[14]</sup>,通过OA方式可扩大科研成果的公共利用范围,提高科研资金的社会效益,因此政府理应参与到OA当中去。总体而言,政府的参与方式包括两个方面,一是通过出台强制性政策要求科研成果通过OA方式出版或在出版一段时间后实现OA,并建立国家开放资源登记系统进行统一管理;二是采取类似参与资助SCOAP3的做法从学术资源传播之初就支持OA,对通过OA方式出版的科研成果给予一定补助,从

而增加开放信息资源建设的话语权,保证开放信息资源收集的完整性与规范性。

### 2.2 制订开放信息资源元数据标准并统一推行

开放信息资源来源多样,根据本文对7000种开放期刊的调研结果来看,开放期刊可获取元数据字段差异较大,多数期刊仅提供题名、作者信息,较少提供摘要信息,且部分期刊信息仅到达刊级,未对单篇论文进行揭示。由此可见,为确保资源共享和兼容完整,提高开放信息资源的整体利用效率,需要针对期刊、图书等各类开放信息资源特征建立专门的元数据标准,并倡导广泛使用,以方便资源集成平台、出版商平台以及机构知识库等不同系统之间的兼容与互操作。

### 2.3 实现商业订购资源与开放信息资源的有机整合

开放信息资源建设不能将传统商业订购资源建设完全独立开来,一方面是由于商业资源与开放信息资源的状态随时存在相互转换的可能性,延迟开放信息资源介于商业订购资源与开放信息资源之间,另一方面在于开放信息资源建设的目的就是扩大用户的学术信息来源,因此不应该将开放信息资源与商业订购资源完全割裂开来,而是通过相应的元数据字段匹配或引用关系实现各类型开放信息资源内部以及开放信息资源与商业订购资源的互联,尽可能地通过用户使用情境推荐合适的信息来源。

### 2.4 寻求开放信息资源建设与服务的版权解决方案

一般而言,明确以OA方式传播的信息资源会增加CC版权标识,即作者在保留部分权利的前提下向社会公众开放,相对而言,CC协议在期刊论文发表中运用得较为常见,但是根据相关研究对7000种开放期刊的调研结果来看,仅有40%的期刊明确遵循某种使用协议,其余并未做出说明<sup>[15]</sup>,而在技术报告、学位论文中使用CC协议的开放信息资源数量则更为有限,这对信息服务机构开展开放信息资源建设造成了一定的版权障碍,需要在规划开放信息资源建设之初就根据当时的

版权环境确定相应的版权处理原则与方法。

### 3 国家开放学术资源建设实施策略

当前, 开放信息资源仍集中在外文资源领域, 我国无论是在开放信息资源建设基础设施还是开放政策方面均相对薄弱, 因此国家开放学术资源建设应采取“分步走”策略, 从外文资源建设扩展到中文资源, 在政策支持下形成从开放信息资源延伸到商业订购外文资源乃至呈缴资源, 最终形成真正的“大”数据仓储, 提供基于事实数据的数据分析和数据支持服务, 见图1。

#### 3.1 建设外文开放学术资源集成平台, 形成统一的外文资源发现系统

目前已有相当一部分外文学术成果正在以开放方式进行传播, 因此首先需要针对这部分外文开放学术资源制订元数据框架, 进行资源遴选、采集、查重、组织揭示等工作, 构建开放信息资源集成平台。文献型开放

学术资源主要涉及开放期刊、开放图书、开放会议录、开放报告、开放课件以及开放学位论文等类型, 元数据框架的形成应同时考虑不同资源类型特征以及后期与商业订购资源关联的需要, 在国际通用元数据标准或已有标准基础上结合开放信息资源的特征生成新的元数据标准。与此同时, 从质量、机读性等方面遴选出质量规范的开放信息资源进行采集。由于部分使用协议不明的开放学术资源的全文采集仍存在版权风险, 在元数据层面为了避免与来源网站形成流量竞争, 需要通过集成网站将用户指向来源网站; 在全文层面, 对使用协议中明确可采集全文进行采集, 并可通过发送邮件、电话等方式联系来源网站, 说明全文采集意向, 由于开放信息资源存在较强的易逝性, 因此对于在一段时间内无回应的开放信息资源可采取“先采集, 后补救”的处理方式, 在集成平台首页显眼处说明与资源来源方的合作意向。一般而言, 全文仅用于长期保存, 仅在来源网站无法正常提供服务或发生战争等突发事件时提供服务。由于各来源网站的元数据字段项目和内容存在较大差异, 后期需要根据元数据框架对采集元数据

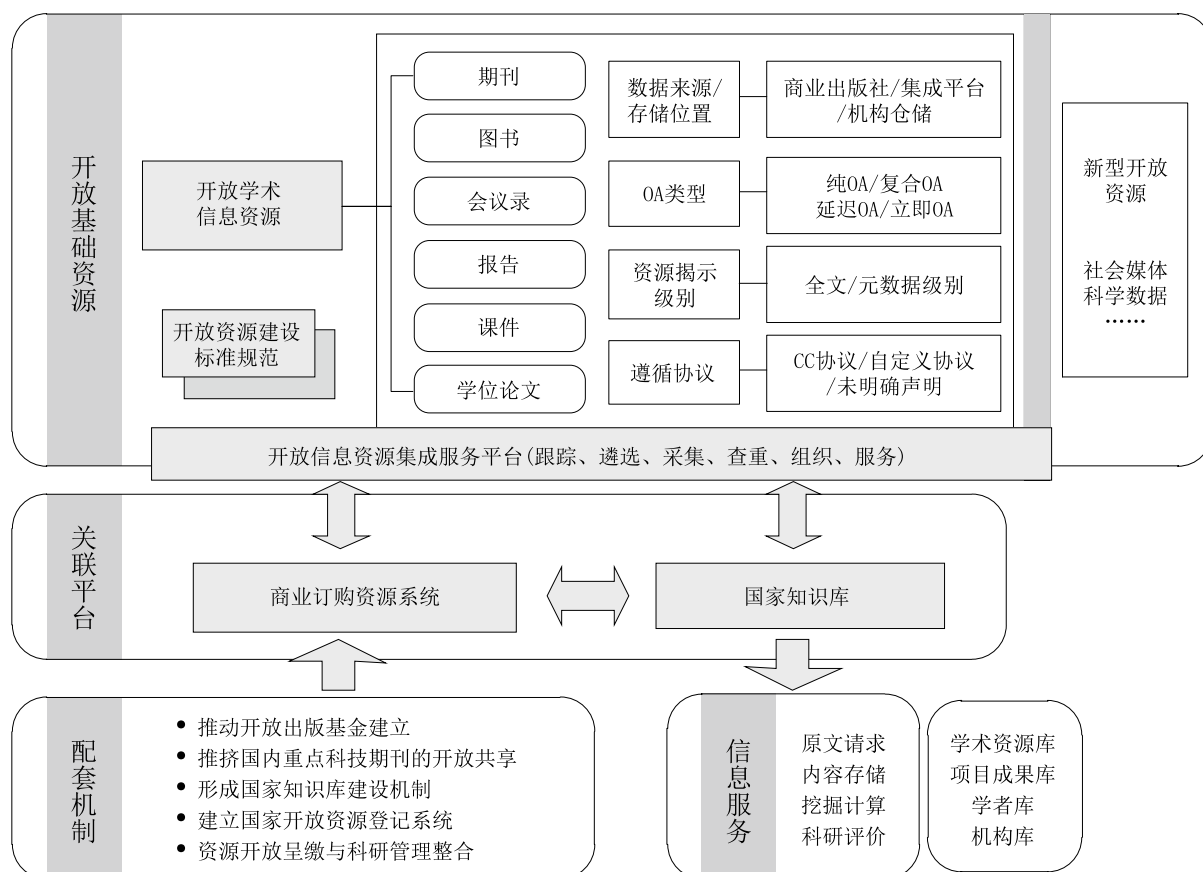


图1 国家开放学术资源建设分步实施策略

进行补充规范,便于提高检索效率以及数据统计分析。在集成文献型开放学术资源的基础上,积极拓展开放数据等其他类型开放数据的集成与发现服务,并利用社交网络数据对开放学术资源进行评价分析等。开放学术资源集成平台并不是一个孤立的信息检索平台,开放学术资源需要通过项目、作者、机构、学科等与商业订购资源实现融合构建,最终形成大规模的外文信息资源发现平台。

### 3.2 构建开放资源登记系统,与机构知识库互联互通,形成国家知识库

除专门面向外文资源进行专门资源集成外,开放信息资源建设还应着眼于国内公共科研项目成果的登记管理。在机构层面,我国越来越多的高校、科研机构正在着手建立机构知识库,例如厦门大学于2013年建设学者论文目录库,为校内课题的结题、考核、评奖、评职等提供依据;清华大学图书馆通过从数据库中抓取全校论文数据并进行院系划分,院系在此基础上再补充数据,并形成了清华大学机构知识库,用于统计学校发表论文情况,它们都在试图通过与本机构科研管理与评价工作结合为机构知识库谋求更多发展<sup>[16]</sup>。从国家层面来说,科研成果录入机构知识库仅仅是机构层面的存缴,对提供公共利用程度的作用有限,亟待建立国家开放资源登记制度与系统,对国内项目生成的中外文科研成果进行统一管理检索。在日本,名为JAIRO CLOUD的共同机构知识库项目在面向全国的机构知识库提供外部技术支持的基础上,日本国立情报学研究所已实现对部分元数据进行集成管理,系统提供的不是资源本身,而是相应资源的存储位置,从而形成国家知识库,它以整合元数据的方式进一步扩大了相对封闭的机构知识库资源的传播范围。结合前期构建的外文开放学术资源集成平台和外文资源发现系统,系统对已集成的开放信息资源可直接向用户提供全文浏览/下载服务,或直接指向相应的机构知识库或商业出版平台,根据用户使用情境进行推荐。

### 3.3 深度组织聚合基础数据,支持第三方数据服务,构建多层次信息服务体系

建设开放信息资源的目不仅仅在于资源长期保存本身,其最终目标是保证用户的稳定获取与利用。在

由开放信息资源、商业订购资源以及国家开放呈缴资源共同形成的大元数据体系下,除提供一般检索和全文下载或推送服务外,通过知识化组织元数据,强化关联数据的处理和形成,进行标题/图标/概念/活动等的标注、规范与数据关联,推进元数据的开放服务。基于数据支撑提供第三方的解决方案服务、研究人员开放的Application服务,逐步开拓和支持相应机构的文献计量分析、数据挖掘与监测、事实型情报分析等,实现大数据环境下的学术知识挖掘服务。同时,从海量信息中抽象分析出热点主题、有影响的学者、有实力的机构,并建立它们之间的学术关系,在不同研究阶段向研究者提供全面的网络分析与挖掘功能,针对机构/研究者提供科研绩效评价服务。在专注服务内容和形式的同时,应重视用户体验和服务效果。通过实地走访调查用户需求、服务数据统计和挖掘,评价和改进服务方式和内容。

## 4 结语

开放信息资源学术质量的不断提升使它逐步得到了学术界、图书馆界、政府的多方认同,但是其自身仍存在易逝性、质量不一等问题,建设规模之大,复杂程度之高,需要以NSTL等国家级信息服务机构为牵头单位,从国家层面出发分步进行开放学术资源建设,从开放信息资源集成平台逐步实现与商业订购资源、国家开放呈缴资源融合构建,形成“大元数据”体系,提供各层级信息服务。同时,开放信息资源建设牵涉多方利益,涉及到学术信息传播格局的重构,因此需要政府、学术界、图书馆界乃至出版界等利益主体的共同协作,进一步促进开放学术信息资源的生产与传播。

## 参考文献

- [1] 2013 Study of Subscription Prices for Scholarly Society Journals [EB/OL]. [2015-06-15]. [www.allenpress.com/system/files/pdfs/library/2013\\_AP\\_JPS.pdf](http://www.allenpress.com/system/files/pdfs/library/2013_AP_JPS.pdf).
- [2] 中国科技期刊开放获取平台关于OA[EB/OL]. [2015-06-15]. <http://www.oaj.cas.cn/aboutoa/index.jhtml>.
- [3] 孙坦.开放信息环境:学术图书馆信息资源建设的重定义与再造[J].中国图书馆学报, 2013,39(3): 9-17.
- [4] Laakso M, Bukvova H, Nyman L, et al. The Development of Open Access Journal Publishing from 1993 to 2009[J/OL]. [2015-06-15].

- <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0020961>.
- [5] DOAJ介绍[EB/OL]. [2015-06-15]. <https://doaj.org/>.
- [6] 陈雪飞, 黄金霞, 张建勇, 等. 国外开放科技资源项目实证研究及启示[J]. 图书馆建设, 2013(6):18-21, 25.
- [7] OATD[EB/OL]. [2015-06-15]. <http://oatd.org/faq.html>.
- [8] Theses Canada [EB/OL]. [2015-06-15]. <http://www.bac-lac.gc.ca/eng/services/theses/Pages/theses-canada.asp>.
- [9] 姚蓉. 学位论文资源建设与服务研究[M]. 北京: 国家图书馆出版社, 2012.
- [10] About EThOS[EB/OL]. [2015-06-15]. <http://ethos.bl.uk/About.do>.
- [11] Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research [EB/OL]. [2015-06-15]. <http://grants.nih.gov/grants/guide/notice-files/not-od-08-033.html>.
- [12] AAAS announces open-access journal [EB/OL]. [2014-10-17]. <http://blogs.nature.com/news/2014/02/aaas-announces-openaccessjournal.html>.
- [13] PQDT Open介绍[EB/OL]. [2015-06-15]. <http://www.proquest.com/about/terms-and-conditions.html>.
- [14] 贺德方. 科技报告的内涵作用与管理机制[J]. 情报学报, 2014, 33(8): 788-792.
- [15] 王星, 苏静, 杨代庆. 国外开放科技信息资源建设情况分析 & 启示[J]. 数字图书馆论坛, 2014(11): 8-13.
- [16] 曾婷, 黄丽, 邹荣, 等. 清华大学机构知识库的扩展服务研究与实践[EB/OL]. [2015-06-15]. [http://ir.las.ac.cn/handle/12502/6324?mode=full&submit\\_simple=Show+full+item+record](http://ir.las.ac.cn/handle/12502/6324?mode=full&submit_simple=Show+full+item+record).

## 作者简介

杨代庆, 男, 1975年生, 高级工程师, 研究方向: 数字资源管理。  
丁道劲, 女, 1988年生, 助理馆员, 研究方向: 数字资源建设。  
周锐, 男, 1987年生, 助理馆员, 研究方向: 资源分析。

## Strategy on National Development of Open Information Resources

YANG DaiQing, DING QiuJin, ZHOU Rui  
(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: Factors such as price increases and the change of users' mode of the academic exchange have contributed to the increasing scale of the open information resources which have been an important source of academic information resources. This paper firstly analyses the present situation of open information resources development on large scale. Then some key issues including government's role, unified metadata standard and copyright solution are put forward. In the end, some strategies are discussed.

Keywords: Open Access; Information Resources Development; Strategic Security

(收稿日期: 2015-07-03; 编辑: 雷雪)