

# 我国学位论文开放资源建设方式研究\*

丁道劲, 沈丹

(中国科学技术信息研究所, 北京 100038)

**摘要:** 在国外, 较多国家/地区已建有学位论文平台供网络用户免费检索和浏览论文全文, 学位论文的开放传播概念逐步得到强化。在分析国外类似平台的资源建设方式、版权管理特征的基础上, 结合我国学位论文资源建设现状, 分析推进我国学位论文开放获取在管理体制、版权管理等方面需要解决的问题, 指出建立各方利益平衡机制是实现学位论文开放获取的重要条件, 通过政策约束、专项资金支持等在元数据层面建设国家统一的学位论文发现系统, 在全文层面构建集中与分布相结合的全文存储体系, 为教育培养、管理决策等提供多方位信息服务。

**关键词:** 学位论文; 开放获取; 发现系统; 资源建设

**中图分类号:** G250

**DOI:** 10.3772/j.issn.1673-2286.2015.08.006

学位论文是现代学位制度的产物, 根据ISO 7144-1986对学位论文的定义, 学位论文是著者提出其研究和发现, 并自荐报请授予学位或专门职位资格的文献。国家标准GB 7713-1987《科学技术报告、学位论文和学术论文的编写格式》指出, 学位论文是表明作者从事科学研究取得创造性成果或有了新的见解, 并以此为内容撰写而成、作为提出申请授予相应学位的学术论文。由此可见, 学位论文作为学位候选人的知识创新成果, 本身具有较高的学术理论价值和现实应用价值。随着我国高等教育规模不断扩大, 学位论文资源体量日渐增加, 开放信息环境下以内部收藏、付费使用为主的传统中文学位论文管理方式难免受到冲击。与此同时, 国外越来越多高质量学位论文趋向免费利用, 学位论文正在逐渐成为开放资源的一个重要组成部分。

结合陈枝清等对开放获取知识库名录(The Directory of Open Access Repositories, OpenDOAR)的研究<sup>[1]</sup>, 以及学术期刊的开放获取实践, 本文将学位论文开放资源分为两个层次: 一是在一定使用许可下可提供用户不受机构限制的全文下载或全文浏览服务的数字化博、硕士论文; 二是提供元数据公共查询服务, 面向机构内用户提供下载服务的学位论文存储系统(例

如: 机构知识库)中的学位论文资源。以学位论文在元数据以及全文层次的开放方式, 本文针对学位论文开放资源建设方式展开专门研究, 以期拓宽学术信息资源建设视野, 促进学术信息资源的完整保存, 推动学位论文在更大范围内的共享与交流。

## 1 学位论文开放资源建设进展

从国内外学位论文开放资源的建设实践来看, 根据涵盖学位论文的范围可将学位论文开放平台分为三个层级, 一是加利福尼亚大学等高校/科研机构独立构建的机构知识库以及机构知识库导航系统, 例如香港地区博硕学位论文查询系统HKLIS DTC; 二是以同一国家/地区学位论文为资源建设对象形成的存储库或导航系统, 如英国的电子学位论文在线门户EthOS、加拿大的Theses Canada门户; 三是类似于NDLTD、OATD一类的国际性数据库, 它们大规模集成了全球不同来源的学位论文。上述类型一般均为非商业性数据库, 伴随开放获取出版商业模式的成熟, 由美国ProQuest公司运作的开放学位论文平台ProQuest Open中的开放学位论文数量也初见规模。

\* 本研究得到国家社会科学基金青年项目“面向文献相关性度量的共词和引文综合分析方法研究”(编号: 11CTQ019)资助。

## 1.1 高校(科研机构)机构知识库中学位论文开放资源建设现状

高校建设机构知识库的目的在于对本校的知识资产进行管理,学位论文作为学校教学成果重要的一部分,是机构知识库重要的文献类型之一。根据OpenDOAR的统计,截止2015年7月,全球已有1607个机构知识库存有博硕士学位论文<sup>[2]</sup>。以加利福尼亚大学为例,该校所有学位论文交由ProQuest公司出版,学生可以选择传统商业出版方式或开放获取出版,所有以开放获取方式出版的学位论文全文均将通过学校的机构知识库eScholarship向公众开放<sup>[3]</sup>。截止2015年7月,eScholarship已收藏该校Berkeley、UCLA、Merced等多个校区的12,051篇学位论文<sup>[4]</sup>,用户可免费在线浏览和下载PDF格式全文,同时可查询单篇学位论文全年/每月的点击量、下载量等使用数据。除了类似eScholarship这种由单一高校构建的机构知识库外,机构知识库集成检索系统也是促进机构知识库中学位论文开放获取的重要力量,例如香港地区博硕士学位论文查询系统HKLIS DTC提供了港澳地区8所高校学位论文检索服务,从而将用户指向各高校的机构知识库。无论是对机构知识库还是导航系统而言,充分尊重学位论文产出机构和作者个人意愿,处理好版权问题系统是得以正常运行的根本,例如加利福尼亚大学各校区对论文提交、延迟出版等做出了具体规定<sup>[5]</sup>,HKLIS DTC涉及的8所学校中有3所可以提供全文浏览服务,它们分别通过在PDF文件中增加水印、确认版权使用声明等方式规范用户使用。

## 1.2 国家电子学位论文存储库资源开放利用现状

与其他文献信息资源建设类似,学位论文的收藏与利用也经历了纸质、缩微到电子化的发展历程,加拿大、美国、英国、德国等发达国家的图书馆及相关信息机构竞相建立了学位论文全文数据库,在国家层面推出了学位论文在线服务门户<sup>[6]</sup>,部分学位论文全文数据库已免费向公众开放,有效地扩大了学位论文的传播与利用。例如,加拿大图书馆在2002年起启动电子学位论文计划并着手建立加拿大学位论文门户Theses Canada,公众通过这个网站可以检索到学位论文全文数据库收录的1998年以来的电子学位论文全文,用户无须注册即可免费下载全文;大英图书馆也在2009年建立电子学位论文

在线服务门户EthOS,目前网站中可检索元数据量达到400,000条,其中160,000篇学位论文可直接通过EthOS或相应机构知识库下载全文。一般而言,国家电子学位论文存储库的资源主要是通过国家学位论文项目吸引国内高校联合参与形成的,数据库主要通过OAI-PMH等开放协议自动收割参与机构库或访问机构FTP地址的方式实现数据的采集。在版权与使用授权方面,各国法律均充分尊重和肯定了作者的各项权益,图书馆多通过要求参与高校必须在研究生申请相应学位时签订学位论文非排他性许可授权协议,以获得学位论文作者的授权。同时,日本、俄罗斯等国的学位论文呈缴制度也在一定程度上对本国大规模的电子学位论文资源建设予以了政策保障。

## 1.3 全球性学位论文集成平台开放资源建设现状

与开放期刊、开放图书等开放资源类似,学位论文也建有专门的开放资源集成平台——开放获取学位论文门户(Open Access Theses and Dissertations, OATD)。OATD中收录了2,540,937条学位论文记录,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协议收割而来,全文仍存储在机构本地。博士与硕士学位论文数字图书馆网络(Networked Digital Library of Theses and Dissertations, NDLTD)的构建机制与OATD类似,也是通过元数据收割形成数据库,二者的最大区别在于OATD学位论文数据库中所有的资源均可自主获取全文,而在NDLTD学位论文数据库中并非所有资源都可以直接获取全文。由于集成性质,这一类开放学位论文资源平台能否提供全文以及全文开放时滞均由元数据来源机构以及作者决定<sup>[7]</sup>,例如NDLTD根据作者的要求将连接到的部分全文数据分为无限制下载、有限制下载、不能下载几种方式,在未获得版权所有人正式许可的情况下,学位论文的全部内容均不得以任何形式、任何媒介进行商业化发行。除了公益性学位论文平台外,美国ProQuest公司为满足各方对学位论文的开放获取需求也推出了开放获取出版服务,每篇收取95美元的出版费用。截止2015年7月,ProQuest Open已存储27,832篇学位论文,用户可免费检索和获取全文。与集成平台相比,ProQuest Open作为商业性出版平台,学位论文的组织加工质量更高,资源存储稳定性较强,但是数量相对有限。

## 2 我国推进学位论文开放获取面临的挑战

无论与国外发达国家还是我国的港澳台地区相比,受多方面因素影响,大陆地区学位论文资源建设相对封闭与分散,显然很难与当前趋向开放的学术信息环境相适应,并直接影响到学位论文的共享与交流。学位论文作为一项学术成果,研究意义不仅限于使作者获得学位,其价值也不局限于一时的评审,其研究价值和应用价值的判断更有赖于后期的共享利用,封闭的交流环境反而更容易滋生抄袭、造假等学术不端行为。因此,学位论文信息理应向社会公开,在广泛使用中检验我国研究生教育成果的质量,从而提高研究生教育管理透明度。但是从我国当前的发展环境来看,学位论文开放仍存在诸多挑战。

### 2.1 尚未建立统一的学位论文资源建设体系

中国学位管理工作与中国的学位与研究生教育息息相关,研究生招生人数的增长导致学位论文的数量也迅猛增加,受行政系统、学科等因素影响,我国的学位论文的多头管理问题较为明显。目前,根据相关规定,中国科学技术信息研究所、国家图书馆以及中国社会科学院图书馆是三家法定收藏机构,同时高校系统和中科院系统也根据自身优势对学位论文进行了广泛、系统、集中收藏。由于分属文化、教育、科研等多系统领导,缺乏统一的管理,受行政约束,各收藏单位之间缺乏协调,各系统之间相对封闭,尚未从国家层面集成元数据或统一存储全文,对大规模推进学位论文开放获取造成困难。

### 2.2 尚未形成统一的学位论文版权登记制度

总体而言,国内各高校对学位论文的版权管理模式归纳为两种:一是出台相关的学位论文著作权管理规定,二是学校与学位论文的作者签署学位论文使用授权许可协议。但是据调研结果显示,很多高校的使用授权许可协议并无版权归属的条款,学位论文的版权归属问题相对模糊<sup>[8]</sup>,统一的版权登记制度则有利于从根本上明确这一问题,对公开学位论文的作者予以相应的版权保护。但是我国当前尚未针对学位论文形成版权登记制度,且现行版权登记责任主体呈现“双重、多元”的混乱局面<sup>[9]</sup>,难以从法律层面确保公开学位论文作者权利。

## 2.3 缺乏可持续的科技信息资源开放获取机制

网络技术的发展使学术信息资源传播与存储的成本大大降低,但仍需要一定的经费和人力资源维持系统的正常运行,商业出版社的加入更使开放获取并非单纯具有公益属性。从国外科技信息资源的开放获取实践来看,特别是粒子(高能)物理领域SCOAP3项目,它通过招标采购开放获取出版服务寻求的正是公共利益与商业利益之间的平衡,而我国目前由于相关机构的科学研究教育经费有限,开放获取还只停留在公益服务阶段,所依赖的经费和存储条件难以得到稳定保障,万方数据、清华同方等商业性信息资源建设力量在科技信息资源开放获取工作中的参与度有待提高。

## 3 面向开放获取的国家学位论文资源建设体系建设

学位论文作为我国高等教育的教学与研究成果,在开题、中期、答辩等过程中经受层层把关,具有较高的学术质量,但是创新性知识和成果得不到广泛应用也必然会慢慢落伍成为数据库中的鸡肋,因此需要将开放获取理念融入我国现有学位论文管理工作中来,面向开放获取在国家层面构建学位论文资源体系。

### 3.1 寻求平衡的学位论文开放利益机制

学术期刊由于价格上涨明显引起学术界、图书馆界不满,但这也直接推动期刊论文的传播较早参与到开放获取运动中,从最初的机构仓储向开放获取出版扩展。从期刊的开放获取现状来看,许多国家政府已出台科研成果强制存缴、开放获取出版基金等支持性政策。Elsevier、Springer等大型学术出版商不断推出开放获取出版服务,足以说明当前的开放获取不再是单纯的公共行为,公共与商业利益共同主导的开放获取运动格局已然形成,学位论文的开放获取进程亦应如此。因此,我国推进学位论文开放获取的首要任务就是在各方利益特别是在作者与公众之间寻求平衡,通过建立统一的版权登记制度,明确作者因撰写学位论文获得正当权益,从而协调学位论文作者、作者所在高校、学位论文收藏单位、数据加工单位、广大公众因学位论文的传播和使用而产生的法律关系。除国家政策

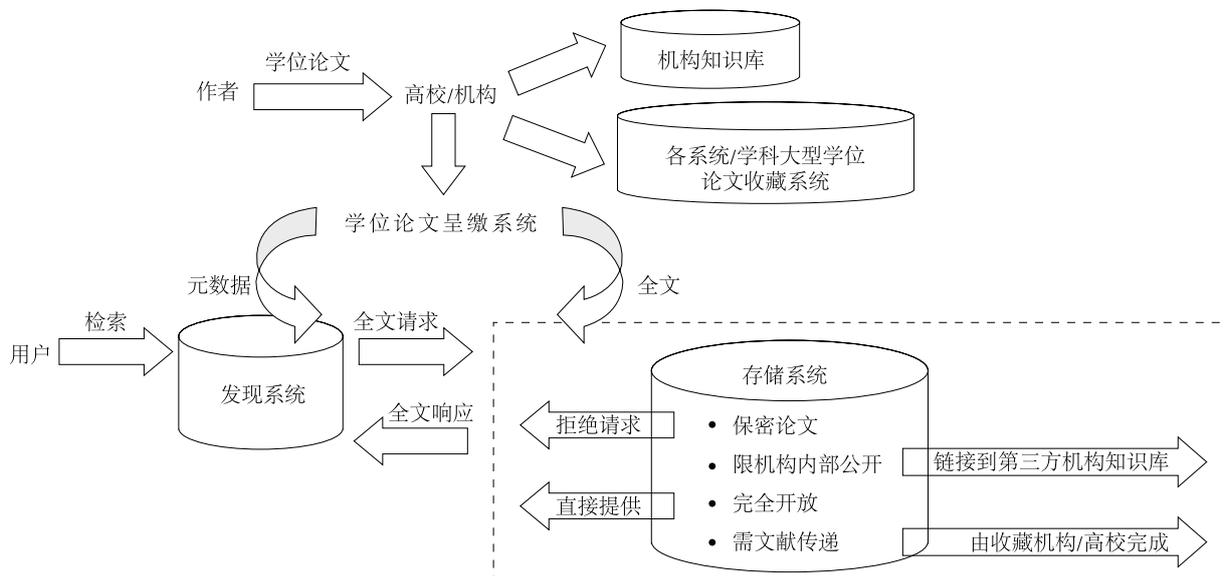


图1 国家学位论文资源建设体系示意图

法规外，在实际操作过程中还应特别要加强作者与高校权利让渡管理，在授权协议书中明确权益归属及是否允许全文公开等权利要点。同时，商业数据库商在数据加工与组织方面的人力、技术力量优势明显，可借鉴ProQuest公司的开放获取出版模式，以“前端付费”方式推进学位论文全文免费公开。

### 3.2 构建统一的国家学位论文资源发现系统

从加拿大、德国等国学位论文资源建设实践来看，以收割、呈缴元数据等方式在国家层面建设统一的学位论文资源系统已较为普遍，它是国家对高等教育成果的一种有效管理手段。我国当前学位论文分散在各个系统/高校，难以把握学位论文全局，需要在元数据层面对学位论文进行统一管理。国家教育部门应适时出台相关规定，强制要求高校将学位论文在发现系统中进行登记，作为学生毕业以及高校考核的重要依据。系统数据分为回溯和更新两部分，回溯数据可优先从国家图书馆、中国科学技术信息研究所等大型学位论文收藏系统获取，为保证数据的完整性、时效性，仍需要自动收割高校机构知识库、自建学位论文数据库元数据，并通过系统查重和元数据质量比较对回溯元数据集进行补充和优化。同时，由于各高校对学位论文的撰写要求不一，元数据字段项目和内容存在较大差异，后期需要根据先期制订的元数据框架对采集元数据进行补充规范，便于提高系统检索以及数据统计分析效率。系统

后期数据更新主要来源于高校每年登记的学位论文信息。发现系统本身不存储全文，但是可提供全文获取线索，根据用户使用情境链接到最优的存储位置，通过即时下载或文献传递获取全文。

### 3.3 建立集中与分布相结合的学位论文全文存储体系

促进学位论文的利用与交流是构建国家学位论文资源发现系统的主要建设意义之一，但是我国学位论文收藏较为分散，现阶段完全实现统一存储在管理体制上存在较大困难，在学位论文法定收藏的基础上吸引高校参与学位论文开放获取，通过机构合作形成集中与分布相结合的学位论文全文存储体系在我国当前具有一定可行性。在组织管理方式上，可由国家图书馆等国家级信息服务机构为主导、其他高校/收藏机构共同参与组成全国学位论文资源建设管理机构，形成协同的全国性学位论文数字长期保存机制。其中，高校可通过全国统一的“学位论文呈缴系统及技术平台”上传PDF格式的学位论文全文，作为国家信息资源战略储备，一般用于大规模数据统计与挖掘使用，仅在突发情况下启动全文服务；系统日常运行中，学位论文是否对外提供全文服务仍由高校决定，并由高校/论文收藏机构实际提供服务。为激励和鼓励高校/机构推进学位论文开放获取，国家应对开放获取学位论文保存给予稳定的维护费用资助，保证学位论文保存的连续性与完整性。

### 3.4 形成多元化学位论文信息服务体系

学位论文信息服务贯穿于学生选题到传播利用的整个过程,立足于本文提出建设的国家学位论文发现系统以及集中与分布相结合的全文存储体系,可形成不同层级的信息服务。除了学位论文检索、全文下载/传递、提供学生学术能力凭证等一般性信息服务外,大体量、完整、权威的学位论文元数据和全文数据也为开展学位论文其他服务提供可能,通过数据整合和数据挖掘衍生而来的学位论文引文、主题分析等结果既可用于评估学生选题的创新性,也可以用于高等教育管理部门按照地区、系统、学科、行业统计教育成果产出等,形成学位论文质量监控系统。但是,随着科研成果产出成果的多样化,单一的学位论文资源体系难以保证评估结果的完整性与准确性,学位论文有必要与期刊论文、会议论文等资源类型关联,特别是学位论文引文数据库目前存在较大空缺,建立跨信息资源类型的引文网络,揭示学位论文与中文期刊、专利、标准乃至报纸之间的引证关系,有助于多方位展示科研实体关系。

## 4 结语

在我国,学位论文作为非公开出版物,其大规模资源建设特别是全文开放在管理体制、版权管理等方面存在诸多挑战,利益关系复杂。但是从全球学位论文资源建设情况来看,大规模集成元数据已成为国家学位论文资源建设的主要方式,并且在其他资源类型开放获取运动的推动下,越来越多的学位论文全文可以直接获取,逐渐成为扩大学位论文共享交流范围的一种重

要渠道,它同时也为公众监督公共教育资金使用提供了窗口。因此,在同时尊重公共利益和作者/高校权益的基础上,有必要在元数据层面建设国家统一的学位论文发现系统,在全文层面构建集中与分布相结合的全文存储体系,通过政策约束、资金支持等方式鼓励越来越多的非保密学位论文被公众利用。

### 参考文献

- [1] 陈枝清,李欣.我国高校学位论文开放获取调查研究[J].图书与情报,2013(5):84-88.
- [2] Search or Browse for Repositories [EB/OL].[2015-06-19].<http://www.openoar.org/find.php>.
- [3] Electronic Dissertation/Thesis (ETD) Submission Information [EB/OL].[2015-06-19].<http://graduate.ucr.edu/ETD.html>.
- [4] Advanced Search[EB/OL].[2015-06-19].<http://escholarship.org/uc/search?keyword=&title=&creator=&publisher=&year=&yearmax=&type=dissertation&facet-affiliation=&facet-discipline=&rmode=save-adv-sel>.
- [5] Open Access Dissertations & Theses[EB/OL].[2015-06-19].<http://osc.universityofcalifornia.edu/scholarly-publishing/open-access/dissertations-theses/>.
- [6] 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,2012:56-80.
- [7] NDLTD网站介绍[EB/OL].[2015-06-19].<http://www.ndltd.org/find>.
- [8] 贺德方,赵嘉朱,姜爱荣,等.国家学位论文资源管理与共享系统研究[M].北京:科学技术文献出版社,2014:177-201.
- [9] 范继红.试论我国版权登记机关的统一[J].电子知识产权,2011(7):50-53.

### 作者简介

丁道劲,女,1988年生,助理馆员,研究方向:数字资源建设。  
沈丹,女,1986年生,馆员,研究方向:数字资源建设。

### Research on the Architecture of Open Access Theses and Dissertations Resources in China

DING QiuJing, SHEN Dan

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

Abstract: In foreign countries, many countries / regions have been built to provide a dissertation platform for Internet users to search and browse for free, the concept of open dissemination of dissertations gradually strengthened. On the basis of analyzing the way of resource construction, the characteristics of copyright management on foreign similar platform, the paper points out that the establishment of a mechanism for the balance among the interests of all parties is an important condition for the realization of the system. Through policy constraints, special funds supporting to construct the national unity dissertation discovery system at the metadata level, at the full-text level to build the full-text storage system of the combination of centralization and distribution, in order to provide comprehensive information services such as education training, management decisions.

Keywords: Dissertation; Open access; Discovery System; Resource Construction

(收稿日期: 2015-08-03; 编辑: 雷雪)