

●张学福 孟连生

论国家博硕士学位论文数字资源保障体系建设

摘 要 建设国家博硕士学位论文数字资源保障体系是国家信息资源战略储备的需要,对促进成果交流和共享,提高科研水平,具有重要意义。目前,国内外学位论文数字资源建设中的论文收集、加工、服务方面,有集中式和分布式两种模式;版权处理上有商业和非商业模式。我国国家学位论文数字资源保障体系建设的论文收集、加工、服务及版权问题处理,应分别情况,采用不同模式。参考文献1。

关键词 学位论文 数字资源 保障体系 国家中心

分类号 G253

ABSTRACT The authors discuss the necessity and importance of a national system of digital resources of doctoral dissertations and master's theses, summarize the patterns for the development of such resources, and propose some suggestions to promote the development. 2 refs.

KEY WORDS Dissertations. Digital resource. Resource system. National center.

CLASS NUMBER G253

1 建立国家博硕士学位论文数字资源保障体系的意义和可行性

1.1 建设国家博硕士学位论文数字资源保障体系的意义

1.1.1 我国信息资源战略储备的需要

博硕士学位论文作为非公开出版物,其印刷版对非收藏单位和作者个人以外的人来说不易获得。但它们具有较高的学术研究价值和使用价值,经过20多年的积累已成为国家的一种重要资源,是我国科学研究、经济建设和社会发展不可缺少的。为了有效地开发、利用这些资源,需要从国家战略的角度考虑这类资源的整体建设。

1.1.2 避免和减少科研工作的低水平重复

学位论文是研究人员选择研究方向和确定论文题目的首选参阅文献。但由于学位论文的非公开出版特征,印本学位论文的流通范围极其狭小。在没有获得相关论文信息的条件下,人们可能没有意识到其他的研究生已经通过学位论文完成了该研究。从可获得的全球性数字学位论文(Electronic Theses and Dissertations, ETDs)资源中,人们从世界上任何地方都能快速检索到与其相关的学位论文。这样就可以避免选题重复,防止科研中走弯路,减少不必要的人力资源与投资浪费。

1.1.3 有助于增加大学和研究生间的资源共享,实现大学的国际合作

过去学位论文除了答辩委员外,其他人很少能阅读到。存储在学位论文网络数字图书馆中的论文可以被千千万万的用户阅读。随着越来越多的ETDs被提供使用,研究生了解其他研究生的研究工作更加容易。由于许多ETDs包括指导教师的E-mail地址和其他联系信息,如果找不到研究

生本人,可以联系指导教师。通过联系,相互间可以就相关的概念、研究方法的应用、研究遗留未决的问题等进一步讨论。通过交流会增加相互间的友谊、合作等。

1.1.4 有助于提高对大学研究成果的存取能力

一篇学位论文可能以数字图书馆系统或相似机制的媒介形式出现。这些论文很可能经常被数以万计的对相关研究感兴趣的研究人员查阅。如果阅读ETDs的数量是阅读印本论文的10或100倍,那么很明显,实行ETD计划后存取大学研究成果的数量将会显著增加。

1.1.5 有利于推动学位论文的传播交流

通过ETDs,研究生们可以有效地交流。在这种研究环境中可以使用彩色照片(使用数码相机或扫描仪)、数字视频序列、音频文档、医学图像和其他数字表示方法。因此通过使用ETDs,将有助于广泛的、更有效的研究成果的交流。

1.1.6 可以提供学位论文的产出状况,为决策提供依据

通过ETDs,可以对大学使用和产生电子学位论文的现状进行调查,了解我国学位论文的产出状况,为相应部门的决策提供依据。

1.2 国家博硕士学位论文数字资源保障体系建设的可行性

1.2.1 国家对电子学位论文资源建设事业的支持

国家制定了一系列政策支持文献信息资源建设^[1]。例如,2004年国家以科技基础条件平台工作重点项目资助学位论文服务体系的研究;CALIS高校学位论文库是国务院批准的我国高等教育“211工程”总体规划中CALIS的子项目,获得国家经费资助,等等。

1.2.2 相关技术支撑

博硕士学位论文数字资源保障体系建设所需的数字化信息存储与管理技术、信息访问与检索技术、信息传播技术、

安全防护技术和原文传递技术等相关技术比较成熟。

1.2.3 资源基础支撑

高校系统、中国国家图书馆、中科院文献情报中心等积累了大量学位论文的文摘、全文资源。CALIS 学位论文库收录文摘数据 7 万条;中国国家图书馆学位论文库现提供 3 万余篇博士论文全文;中国科学院学位论文数据库收录数据 28000 多条等。

1.2.4 实践基础和国外经验的支撑

高校系统、中国国家图书馆、中科院文献情报中心等在建构型、全文型学位论文建设过程中积累了大量实践经验,建立了学位论文信息检索和服务系统,积累了一大批相关方面的人才。UTOG、NDLTD 和 ADT 等的实践可以给我们提供非常丰富的经验。这些都为我国博硕士学位论文数字资源保障体系建设打下坚实的基础。

2 国内外博硕士学位论文数字信息资源建设剖析

2.1 学位论文的收集

学位论文的收集有集中式、分布式两种模式。集中式收集以美国、英国和我国的台湾地区为代表,分布式收集的典型代表为 NDLTD 和以 CALIS 为代表的中国高校系统。

1999 年美国国会图书馆与 UMI 公司签署协议,委托 UMI 收集所有博士论文,国会图书馆购买由其加工后的全部电子版博士论文。UMI 收集过程为研究生将学位论文电子版、印刷版本、论文目录统一交给学校图书馆,学校图书馆汇总后交给 UMI。美国这种集中模式的优点是:UMI 采用商业运作,收集论文比较齐全;国会图书馆完整购买一份,可以有一个完整备份。缺点是学位论文的使用必须从 UMI 购买,要求有大量资金作为保障。

1988 年,英国版权、设计与专利法案规定学位论文为非公开出版物,是法定送缴物。学位论文提交要遵循一定的标准,学位论文通过送缴方式收集。学位论文提交时学生可以从下面三种情况中选择其一:无限制使用、校园内使用和仅通过原文传递获得。收集过程是研究生将学位论文电子版、印刷版本交给学校图书馆,学校图书馆汇总后呈缴到大英图书馆。主要收集国内博士论文。英国这种集中式模式的优点是:国家以法律的形式规定学位论文必须集中呈缴到国家图书馆,收集论文比较齐全;学校有本校学生学位论文电子版。缺点是学位论文的使用采用原文传递形式,不能免费获得。

台湾新修订的学位授予法规定学位论文是法定呈缴物,必须统一呈缴到台湾中央图书馆。收集过程是研究生通过网络提交到学校,经学校各系所行政人员审核通过后统一呈缴到中央图书馆。台湾这种集中式模式的优点是:以法定上缴形式收集学位论文,收集论文比较齐全;学校有本校学生学位论文电子版,免费使用。缺点是未获得授权的论文不能提供服务。

NDLTD 是由美国国家科学基金会和联合国教科文组织经费支持,200 余家来自全球的单位参加的开放联盟,收集博士和硕士论文。它采用由各个学校分散收集学位论文的电子版、NDLTD 集中收集学位论文元数据的模式。收集过程为研究生向所在学校提交学位论文的电子版全文;会员单位为 NDLTD 提供论文的元数据,并由其汇总后加入到元数据库中。NDLTD 分布式收集模式的优点是:由学位授予单位直接收集学位论文电子版,并由其提供全文服务,学生容易接受;免费收集;可以充分调动学校的积极性;提供所有成员单位学位论文的目录、摘要;免费提供技术支持。缺点是部分学生不愿意免费授权提供学位论文,学位论文的完整性受到限制。

以 CALIS 为代表的中国高校系统的收集模式与 NDLTD 相似,由学位授予单位直接收集学位论文电子版,CALIS 中心集中所有论文的元数据。收集过程为研究生上缴电子版学位论文到学校,学校把元数据提交给 CALIS 中心,由其汇总,免费提供给成员单位。CALIS 学位论文分布式收集模式的优点是:由学位授予单位直接收集学位论文电子版,可以充分调动学校的积极性,同时学生容易接受;免费收集;提供各个成员单位学位论文的目录、摘要;为各个单位免费提供文摘系统技术支持。缺点是各个授予单位提供的全文检索系统标准不统一,不便于集成检索;各个单位技术、人员等条件不同,学位论文数字化技术差别较大;学位论文的使用限制在校园内。

2.2 学位论文的加工

学位论文的加工也分为集中式和分布式两种模式。集中式加工模式的代表是美国的 UMI、英国的大英图书馆、中国国家图书馆和台湾中央图书馆,分布式加工模式的代表是 NDLTD 和 CALIS。

UMI 公司集中收集学位论文后,由 UMI 公司集中加工,生成格式、标准统一的学位论文数据库资源,包括文摘库和全文库,美国版权局为 UMI 提供一切软件和技术支持。其优点是数据库采用统一的格式;美国版权局提供统一的软件和硬件支持;经费由 UMI 公司提供,资金有保障;有完整的目录、文摘信息。缺点是所有学位论文集中到一起,统一加工所用时间较长。

大英图书馆集中收集学位论文后由大英图书馆集中加工,把学位论文元数据编入大英图书馆的 OPAC 和英国 BLPC (The British Library Public Catalogue);制作原文缩微胶片;进行电子文本的格式转换等。其优点是数据库采用统一的格式;有完整的目录、文摘信息。缺点也是加工周期长。

中国国家图书馆集中收集印刷版的学位论文,然后由收藏中心负责元数据加工、全文影像库建设。其优点是数据库采用统一的格式;有完整的目录信息。缺点是加工周期长;没有文摘信息;电子学位论文是对印刷版扫描后获得的,存储空间大,同时也很难进行全文检索。

台湾中央图书馆集中收集学位论文的电子版,然后由该馆负责元数据加工、全文格式转换等工作。其优点是数据库采用统一的格式;有完整的目录、文摘信息。缺点是加工时间长。

NDLTD采用OAI协议收集元数据,把采集到的元数据集中加入到学位论文元数据库中,学位论文电子版全文由各学校在本地区分布式建设。其优点是数据采用统一的标准格式;有完整的目录、文摘信息,同时在其中提供全文的链接(已授权的论文);免费为会员单位提供系统平台;分布式加工,加工周期短。缺点是提供的全文链接受全文所在服务器的工作状态限制。

CALIS把各个成员单位提交的学位论文元数据集中加工,并把其加入到学位论文元数据库中,电子版论文全文由各个学校自己加工建设,构建自己的全文数据库。其优点是学位论文元数据库格式统一;有完整的会员单位学位论文的目录、文摘信息;免费为会员单位提供文摘系统平台;分布式加工,加工周期短。缺点是全文数据库格式不统一;全文系统的软、硬件资源由授予单位自己负责;目录、文摘中没有全文的链接。

2.3 学位论文的服务

(1)集中服务模式:分为商业模式和非商业模式。

商业模式的代表为PQDD和CDMD。PQDD由UMI公司以商业形式运作,为用户免费提供学位论文的文摘和论文全文的前24页。PQDD用户采用购买全部或部分数据库(集团用户)和使用原文传递的方式获取学位论文全文。CDMD由清华同方公司以商业形式运作,用户采用购买全部或部分数据库(集团用户)和购买用户卡(个人用户)的方式获取学位论文全文。商业模式的优点是用户购买数据库或用户卡后就可以在一定时间内使用论文全文数据库;数据库使用不受时间(每天的时间)和地理位置的限制;可以多次下载论文全文。缺点是需要投入大量的资金;论文使用只能在定购期限内,不能保证长期使用。

非商业模式的代表为大英图书馆文献提供中心和中国国家图书馆学位论文收藏中心。

大英图书馆文献提供中心:用户通过Index to Theses等进行检索,选择适合的学位论文,然后由其通过原文传递提供全文服务。

中国国家图书馆学位论文收藏中心:用户可免费浏览3万余种博士论文全文的前24页。学位论文采用“到馆服务和原文传递服务”方式。

这种非商业的集中式服务模式的优点是:用户不需要投入大量的资金。缺点是只能通过原文传递方式和到馆服务方式获得全文;下载的原文电子版只能在几天(如14天)之内使用。

(2)分布式服务模式的代表为NDLTD和我国“CALIS学位论文库+各大学学位论文库”。

NDLTD为用户免费提供检索和浏览服务,用户可以免费查询NDLTD联盟中所有成员机构的电子版硕士论文,并可免费获得论文的题录和详细摘要及部分可获取的学位论文全文。用户首先检索元数据库,然后通过OPENURL技术访问分布在各个学校的论文全文。全文的服务由论文提供单位提供,是否提供全文,是否收费,均由论文提供单位决定。其优点是用户可以免费获得学位论文全文;缺点是授权的学位论文全文数量有限。

CALIS学位论文库提供学位论文的文摘信息,全文信息的获取需要与论文作者所在高校的图书馆联系,通过原文传递等途径获取。其优点是成员单位可以免费使用文摘信息。缺点是不能免费获取全文;各个高校全文数据库建设标准不统一,不便于提供集成服务。

2.4 学位论文的版权问题

版权问题解决方式分为商业模式和非商业模式。

商业模式的代表为UMI,非商业模式的代表为NDLTD、VT-ETD和英国大英图书馆。

UMI公司1999年与美国版权局和美国国会图书馆签署协议,有权与美国国会图书馆一起收集论文,与美国版权局一起存放论文。协议对整个流程都做了详细规定,如UMI有权对数据库内容增加或删减而不用通知用户。一旦出现了问题,一律按照UMI与学生、图书馆或是研究机构签署的协议,责任分担。学生在与UMI签订协议时,必须把自己的要求注明,否则出现问题一律由学生本人承担。学生有权对自己论文缩微版本复制或传播,但无权对他人电子版论文、论文摘要的某一部分或全部内容复制或传播。学校等机构在取得学生授权后,可以整体和UMI签订协议,学校也可建立自己的论文库,集成资源。商业模式的优点是公司给予版权人一定报酬,完全解决版权问题;缺点是用户利用时必须通过购买才能获得。

NDLTD采取的方法为其拥有保存和使用电子学位论文元数据的权利,包括使用这些元数据进行非商业目的的检索;按照版权法,学位论文版权归作者所有;与学生签订非专属授权协议,该协议分为三种:论文在全球范围内免费开放;只允许联盟成员访问;只允许联盟成员访问指定内容。

VT-ETD采取的方法为其拥有对论文的元数据加工、编制索引、对外提供检索服务的权利,并且通过与学生的协商,可以与PQDD、NDLTD等进行合作;学生是著作权的拥有者,学校建议学生自己保留版权,也可以考虑保留部分的版权(诸如特殊的数字、表格、多媒体等);学生提交论文时,可以对自己的论文以电子形式访问的公开程度进行选择,例如自由使用(推荐)、仅局限于校内使用、一段时间内的全面禁止(一般是涉及到专利等问题)。

NDLTD和VT-ETD方法的优点是它们拥有保存和以非商业目的使用电子学位论文元数据的权利;与学生签订非专属授权协议解决版权问题;可以免费使用部分学位论文全

文;联盟内或校内免费使用学位论文(保密论文除外)。缺点是不能保证所有撰写非保密论文的学生自愿签订非专属授权协议。

英国的博士论文作为非正式出版物受版权、设计和专利法案保护,该法案禁止在没有版权人同意的情况下发表论文中的任何重要部分的内容。学位论文的版权属于作者,但当研究过程开始前,在研究者和授予机构之间这种所有权可以通过书面协议进行明确的或是承诺部分的转让。如果授予机构实际雇用研究人员从事研究工作,版权属于授予机构。大英图书馆(BL)要求作者与其签订协议,内容主要有在需要时允许其拷贝博士论文,并把它们出借或出售给个人或图书馆,BL提供的拷贝与原作品一样受法案的同等保护;BL向作者提供销售利润的10%作为版税,每年支付一次。

英国采用方法的优点是以法案的形式规定学位论文的版权属于作者;以作为作者支付版税的方式解决版权问题。缺点是用户不能免费使用学位论文;版权授权只是局限在大英图书馆内,远程用户只能使用原文传递方式。

3 对我国国家博硕士学位论文数字信息资源保障体系建设的建议

3.1 学位论文的收集模式

考虑到我国的实际情况,完全的分布式收集和完全的集中式收集都不太适合。因此,可以考虑采取分布式和集中式相结合的模式收集学位论文,即各个学校或研究机构收集本单位的电子学位论文,同时呈缴到国家法定收藏机构一份,进行集中保存。该模式包括两个层次:国家层面的集中式收集和学校或研究机构层面的分布式收集。

对国家层面来说,我国对电子形式的学位论文没有明确的规定,而且电子版学位论文也没有必要存放在三个单位中。借鉴英国等存放在国家图书馆的惯例,可考虑修改国家学位管理条例,明确授权国家图书馆为电子版学位论文的法定收藏机构,或是成立国家学位论文虚拟收藏中心,以法律形式授权其收藏电子版学位论文,解决目前三家印刷版学位论文法定收藏单位对电子版学位论文建设的冲突问题,解决电子版学位论文的集中存储问题。

对学校或研究机构层面来说,借鉴NDLTD模式和CALIS模式,学生把电子版学位论文提交到相应学校或研究机构的学位论文提交系统中,各授予单位通过各自的学位论文提交系统收集本单位的电子版学位论文,构建自己的全文数据库。同时把汇总的电子版学位论文呈缴到中国国家图书馆或国家学位论文虚拟收藏中心一份。要求研究生向学校或研究机构提交学位论文采用统一的标准格式(如Word等)。

3.2 学位论文的加工模式

国家层面的集中式收集和学校或研究机构层面的分布式收集模式,决定了我国的学位论文加工模式有两种:国家层面集中式加工和学校或研究机构层面的分布式的加工。

比较前面案例剖析中集中式和分布式加工的优缺点,采用学校或研究机构层面的分布式加工模式比较适合。为保证学位论文提交、数据加工的一致性,应该由国家资助开发统一学位论文的提交平台、加工和全文服务平台。论文的提交、加工和服务平台应考虑与目前国内相关部门全文系统的兼容性,借鉴这些系统的优点,有效利用现有的全文学位论文资源。

在统一的提交、加工和服务平台上,各个学校或研究机构就可以分布式建设学位论文数据库;国家图书馆或国家学位论文虚拟收藏中心只是简单地把学校或研究机构层面提供的数据汇总就可以把相应数据整合到全文数据库中。另外由其负责国家电子版学位论文的目录、摘要的加工整合工作,在其中增加电子学位论文全文的链接(国家中心和具体授予单位两个链接)。

3.3 学位论文的服务模式

考虑到我国的实际情况,在集中式收集和学校或研究机构层面的分布式收集模式的基础上,采用商业模式和NDLTD那种完全免费的模式,都不太适合。NDLTD模式有可能损害某些部门的利益,不利于调动学校和研究机构的积极性。

可以考虑采用非商业模式的集中式和分布式相结合的模式,即国家层面和学校或研究机构层面共同提供学位论文全文服务,国家层面同时集中提供学位论文目录、文摘的服务。

国家层面和学校或研究机构层面收集的电子版学位论文在学校或研究机构层面的成员单位间免费提供服务。这样参与建设单位可以以少量建设本单位学位论文全文库的资金,享受成员单位提供的所有学位论文,可以充分调动学校或研究机构的积极性,体现学位论文资源的共建共享。用户可以根据网络情况选择使用国家中心或是成员单位(授予单位)的服务器,方便用户使用。对非成员单位的用户采用原文传递的方式提供服务(学生授权在全国范围免费服务的除外),用户可以向国家中心和学校或研究机构提出原文传递申请。关于原文传递规则可以参考大英图书馆的规则来制定。

参考文献

- 1 张学福.论国家文献信息资源保障体系建设.中国图书馆学报,2004(4)

张学福 中科院文献情报中心、中科院研究生院博士研究生。黑龙江大学信息资源管理研究基地教授,黑龙江大学信息管理学院副院长。通信地址:北京中关村北四环西路33号。邮编100080。

孟连生 中国科学院文献情报中心博士生导师,国家科学数字图书馆项目管理中心办公室主任。通信地址同上。

(来稿时间:2004-09-17)