

科技部科技基础性工作专项资金重大项目 研究成果

项目名称：我国数字图书馆标准规范建设

子项目名称：专门数字对象描述元数据规范

项目编号：2002DEA20018

研究成果类型：研究报告

成果名称：学位论文资源分析报告

成果编号：CDLS-S05-008

成果版本：总项目组推荐稿

成果提交日期：2003年7月

撰写人：清华大学图书馆学位论文描述元数据规范课题组

项目版权声明

本报告研究工作属于科技部科技基础性工作专项资金重大项目《我国数字图书馆标准规范建设》的一部分，得到科技部科技基础性工作专项资金资助，项目编号为 2002DEA20018。按照有关规定，国家和《我国数字图书馆标准规范建设》课题组拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

本报告可以复制、转载、或在电子信息系统上做镜像，但在复制、转载或镜像时须注明真实作者和完整出处，并在明显地方标明“科技部科技基础性工作专项资金重大项目《我国数字图书馆标准规范建设》资助”的字样。

报告版权人不承担用户在使用本作品内容时可能造成的任何实际或预计的损失。

作者声明

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性和可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。科技部有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

《我国数字图书馆标准规范建设》课题组网站 (<http://cdls.nstl.gov.cn>) 作为本报告的第一发表单位，并可向其他媒体推荐此作品。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

学位论文资源分析报告

目 录

1. 学位论文的定义和特点	1
1.1 定义	1
1.2 特点	1
1.2.1 从载体形式上看	1
1.2.2 从责任者方面看	1
1.2.3 学位论文版权	2
1.2.4 学位论文的学位信息	3
2. 学位论文的著录对象分析	3
2.1 著录对象的四个层次	3
2.2 学位论文著录对象分析	4
3. 学位论文的著录单位、著录内容	4
3.1 著录单位	4
3.2 著录内容	5
4. 学位论文的使用和检索需求	5
5. 国内外可参考的学位论文元数据规范	5
5.1 国外可参考的学位论文元数据规范	5
5.2 国内可参考的学位论文元数据规范	7
参考文献	9

1. 学位论文的定义和特点

1.1 定义

具美国标准学会解释，学位论文是为了获取不同级别学位的候选资格、专业资格或其他授奖而提出的研究成果或研究结论的书面报告^[1]。简而言之，学位论文就是学生为了获取学位向学校或其他学术单位提交的学术研究论文。学位论文是高校、科研院所图书馆文献资源建设中非常宝贵的信息资源，其中蕴涵着大量的富有创造性的思维和成果，它不仅反映学校、科研机构的教学、科研水平，同时也直接反映各学科领域的前沿动态和最新发展，对后继的教学和科研活动有较高的参考价值。

1.2 特点

1.2.1 从载体形式上看

学位论文从其载体形式上来分，有两种类型：一种是印刷本的；一种是数字化的学位论文。数字化的学位论文从其产生方式上有两种：

- 由学生通过远程提交系统提交的论文

随着计算机的日益普及、信息技术的不断发展，多数学生采用通用的软件来编辑、撰写自己的论文，多数学生在提交印刷型论文时，也同时拥有本人论文的电子版。从 2000 年开始，国内相继有些学校要求学生在毕业之前，提交一份与印刷本论文内容和格式完全一样的电子版论文。在学生创作学位论文时，还有一些与论文相关的实验数据、实验报告、多媒体资源、软件系统、彩色图象、化学结构式的虚拟现实描述、声音文件等。这些作为学位论文的附加信息，也是很有价值的。在学生毕业之前，除了要求学生提交一份与印刷本相同的电子论文外，也可要求学生提交与论文相关的附加信息。通过远程提交方式收集电子论文，为收集与论文相关的“附加信息”提供了可能。

- 由印刷本论文扫描而形成的数字化论文

没有通过提交系统收集的论文，其数字化过程需要通过扫描而完成。一般是指没有建立提交系统的学校或在建立提交系统前，只收集了印刷本论文的学校，需要通过扫描方式来进行论文的数字化。

1.2.2 从责任者方面看

学位论文的选题一般都是本学科需要解决的比较重要的具有前沿性的理论或应用方面的课题，代表了专业的发展方向，本身也是导师研究内容的一个子课题或子项目。学位论文一般是在导师的指导下，以学生为主，由学生创作、编写而

成。

学位论文的著者除了学位申请者之外，没有其他任何的合作者，这一点与其他类型的文献，特别是期刊论文有所不同。因此，学位申请者应是论文的第一责任者。

学生的导师作为论文著者的指导者，在论文的创作和形成过程中起到了重要的作用，所以，导师应视为仅次于论文著者的其他责任者。

学位申请者要申请学位，需要通过论文答辩。论文答辩委员会成员采取不记名投票方式，就是否通过论文答辩和建议授予学位进行表决。论文答辩通过，由学位授予机构授予学位证书。论文答辩未通过，但经论文答辩委员会全体成员半数以上同意，并做出决议，可在两年内修改论文，重新申请答辩一次。如果论文答辩委员会未做出同意修改论文的决议，任何个人无权同意修改论文并重新组织答辩。为了提高论文质量，使得正式答辩时更有把握，有些单位在安排正式答辩之前，还为申请者安排了一次论文的预答辩，以期收到更多更好的关于论文的修改意见。因此，答辩委员会成员也可视为该论文的其他责任者。但其作用远次于导师，故列为可选项，是否予以著录，由资源创建者自定。

1.2.3 学位论文版权

学位论文不在正式出版物上发表，属于非正式出版物。印刷本学位论文设有公开、内部、秘密和保密四个级别。各个图书馆采用的服务办法基本相同：公开的论文不提供外借但提供室内阅览服务，内部、秘密和保密的论文在解密后，提供室内阅览服务。读者阅览任何一篇学位论文前，必须登记所查阅的论文篇目，若读者提出部分复印要求，应登记所复印的页码，图书馆负责保存登记档案。

因网络和计算机传递资源的便捷性，电子版论文的版权控制和服务问题远比印刷本论文复杂。与印刷本论文对应的电子版论文同样设有公开、内部、秘密和保密四个级别，但其涉及到网络信息传播权、电子论文复制权等复杂问题。学位论文是研究生利用本校的各种教学、科研资源，在导师的指导下，通过作者的创造性劳动取得的研究成果，它是学校、导师和学生共同努力的成果，其版权控制又与电子期刊、电子图书等有所不同。

在“高等院校知识产权保护管理规定”中第十三条规定：“在高等学校学习、进修或者开展合作项目研究的学生、研究人员，在校期间参与导师承担的本校研究课题或者承担学校安排的任务所完成的发明创造及其他技术成果，除另有协议外，应当归高等学校享有或持有。进入博士后流动站的人员，在进站前应就知识产权问题与流动站签定专门协议。”，第九条规定：“为完成高等学校的工作任务所创作的作品是职务作品，除第十条规定情况外，著作权由完成者享有。高等学校在其业务范围内对职务作品享有优先使用权。作品完成二年内，未经高等学校同意，作者不得许可第三人以与高等学校相同的方式使用该作品。”这应是高校使用学位

论文的基本依据和原则。

通过以上分析，学位论文的使用授权应该由学生授予学校，其授权模式可由学校拟定规范文本，由学生签字。学位授予机构（高等学校、科研院所）在尊重作者知识产权的前提下，有权按规定保存论文的印刷版和电子版，并提供目录检索与阅览服务。

1.2.4 学位论文的学位信息

学位信息是学位论文所特有的特征。国标 GB6864-86 中明确规定了学位名称和代码。

2. 学位论文的著录对象分析

2.1 著录对象的四个层次

国际图联颁布的研究报告《书目记录的功能需求》（简称FRBR）^[2]用“实体-关系模型”(E-R模型)来描述编目对象，把作为编目对象的实体划分为著作（work）、表现方式（expression）、表达形式（manifestation）、文献单元（item）4个层次。著作是知识或艺术创作内容，是一个抽象的实体，必须通过一定的表现方式，如文字、图像、乐曲等，才能被人感知。表现方式体现知识或艺术创作内容的物理形态，如手稿、图书、期刊、录音制品等；同一种表现方式可采用不同的表达形式，如某一版本的文字著作可以是印刷型的图书，也可以是缩微品或电子图书。文献单元是表达形式的某一物理实体，如图书馆收藏的某一本书或某一套多卷书。著作、表现方式、表达形式、文献单元这4个层次都可以成文编目对象，确定哪一层次为编目对象与不同类型文献的目录学特征有密切的关系。从下图1可看出4个层次的关系。

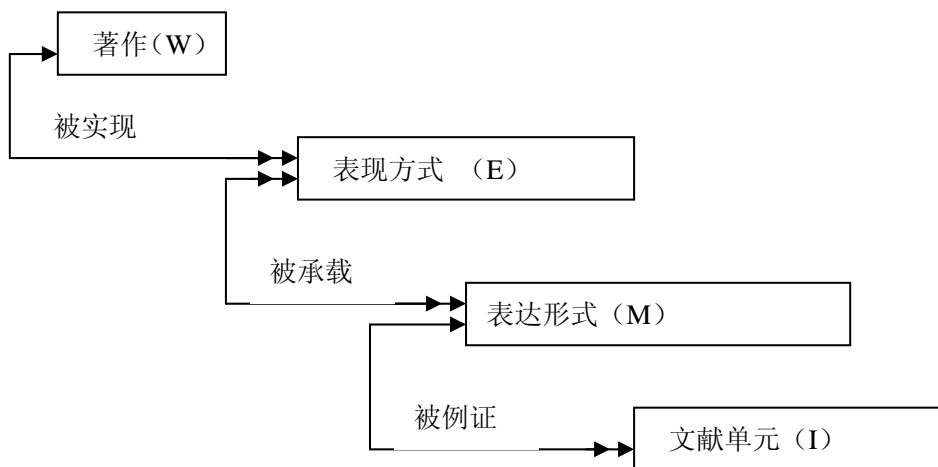


图1 著录对象4个层次之间的关系
(注：单箭头表示一，双箭头表示多的含义)

2.2 学位论文著录对象分析

根据前面著录对象的4个层次的分析,可以将学位论文著录对象之间关系描述如下:

◆ 著作(W)

例如:清华大学研究生蒋萌的硕士毕业论文

◆ 表现方式(E)

例如:论文题目是《EIM-601 程控交换机光纤互连接口板的设计》的学位论文

◆ 表达形式

- 学位论文存在多种表达形式;
- 学位论文的印刷版、电子版同时并存;
- 通过远程提交方式收集电子版学位论文,还可能附有与学位论文相关的音频、视频、动画、软件系统等“附加资源”

例如:该论文有印刷版和电子版两个版本

◆ 文献单元

- 文献单元是表达形式的某一物理实体
- 学位论文存在不同的文献单元
- 印刷版学位论文具有多个复本
- 电子版学位论文具有多个复本或格式

例如:电子版学位论文具有多种格式,每种格式又有多个复本

3. 学位论文的著录单位、著录内容

3.1 著录单位

确定著录单位也即确定以哪一个层次为基准著录资源对象,需要考虑可行性、是否易于著录、是否易于检索、是否节约编目人力等方面的问题。

从上面的分析可以看出,一篇学位论文有电子版和印刷版两个版本,每个版本可能又有多个复本,学位论文以单篇论文作为一个著录单位,不同版本的学位论文做等同处理。

这个著录单位,是一个集成的概念,包括电子版和印刷版两种不同版本,还包括与本论文相关的实验数据、实验报告、多媒体、音频、视频、动画、软件系统等“附加资源”。电子版和印刷版共同具有的特征,可以其电子版为主著录。涉及到印刷本的特性,需要从印刷本中提取著录项。例如:由于版权和其他应用环境等问题的限制,用户在网络上不能获得论文电子版全文的情况下,可到指定的馆藏地阅览其相应的印刷本,这样有关印刷本的馆藏地点(即馆藏号)需要著录。

3.2 著录内容

学位论文的著录要点是论文的题名、主题、文摘；著者、导师、学位的授予单位；资源形式以及网址、馆藏地、权限管理等。因为它们是提供用户发现、评价、选择和获取该论文的重要因素、也是我们创建该资源的目的所在。

通过远程提交方式收集的学位论文由学生（论文著者）和图书馆员共同著录，大部分的著录项目由学生完成，图书馆员检查学生的著录内容，同时完成印刷本论文的馆藏索取号、分类号等学生不能做的项目的著录。

通过扫描等数字化过程制作的论文或印刷本论文需要由图书馆员著录。

著录项目应有严格的规范控制和著录规则控制。

4. 学位论文的使用和检索需求

创建学位论文数据库的最终目的是为了使用。用户希望能够很容易、很方便地查找所需要的论文，这就要求学位论文元数据标准应具有广泛的通用性和兼容性；用户希望能够从不同角度快速、准确地发现所需要的资源，这就需要为其提供题名、著者、主题等不同的检索途径；对于所查到的论文，用户又希望能够对其实用性与权威性进行比较并迅速做出判断与选择；因此有关该论文的相关资料以及导师与学位授予机构的提供也很重要；还有，关于如何正确利用与获取论文的信息也不可或缺。

通过以上的分析，暂将著录元素集合确定为：题名、主要责任者、主题、资源描述、其他责任者、日期、资源类型、资源格式、资源唯一性标识、出版者、语言、权限管理、学位、相关资源 14 个元素。

同时应提供按学科、分类浏览的功能，最大限度满足用户需求。

5. 国内外可参考的学位论文元数据规范

5.1 国外可参考的学位论文元数据规范

电子版博、硕士学位论文元数据标准 ETDMS (an interoperability metadata standard for electronic theses and dissertations), 应用于 NDLTD(networked digital library of theses and dissertations)项目, 该项目由 virginia polytechnic institute and state university 图书馆和计算机系联合发起的世界范围内的学位论文共享项目, 该项目得到了美国教育部资助。到目前为止, 成员单位有 152 个, 其中包括世界范围内的 132 所大学, 20 所公共机构。目前, 可免费检索该系统内学位论文数据。

表 1 ETDMS 元素的定义和分析

元素	描述	注释
dc.title	论文题名	必选/可重复
dc.title.alternative	选择题名	任选/可重复
dc.creator	论文著者	必选/可重复
dc.subject	主题词或关键词, 应使用 scheme qualifier	必选/可重复
dc.description	文摘, 可与 dc.description.abstract 交互	任选/可重复
dc.description.abstract	文摘	任选/可重复
dc.description.note	论文的附加信息, 例如: 院系的评注	任选/可重复
dc.description.release	著作的版本的描述, 例如: 勘误表描述	任选/可重复
dc.publisher	对资源数字化或存档负责的团体	任选/可重复
dc.contributor	其它责任者, 论文指导教师或答辩委员会成员	任选/可重复
dc.contributor.role	责任方式, 例如: 导师、委员会成员、主席	任选
dc.date	出现在论文题名页的日期, 以 ISO8601 标准定义	必选
dc.type	资源类型和资源内容类型, 资源类型推荐著录成 electronic thesis or dissertation, 资源内容类型可参照 标准词表 http://dublincore.org/documents/dcmi-type-vocabulary/	任选/可重复
dc.format	论文存储或传递中使用的格式, 使用标准的MIME 类型, ftp://ftp.isi.edu/in-notes/iana/assignments/media-types/media-types mime 中不包括的, 著录为 unknown, 如果电子形式的论文不能用, 则可省略	任选/可重复
dc.identifier	论文唯一标识, 推荐使用 URI, URN, purl, handles 等	必选/可重复
dc.language	文献语种, 使用 ISO 639-2 或 RFC 标准, 默认的语种是 en	任选/可重复
dc.coverage	论文内容的时空范围	任选/可重复

续表 1

元素	描述	注释
dc.right	权限声明的信息	任选/可重复
thesis.degree.name	学位名称, 例如 <i>Masters in Operations Research</i>	任选/可重复
thesis.degree.level	学位级别, 例如 bachelor's, master's 等	任选/可重复
thesis.degree.discipline:	学科名称	任选/可重复
thesis.degree.grantor	授予学位的机构名称	任选/可重复

以上表 1 列出的所有元素都是可检索字段。

5.2 国内可参考的学位论文元数据规范

“calis 高校学位论文数据库”数据规范, 由 CALIS 全国工程文献中心项目管理组依照国家标准 GB/T 2901-92 《书目信息交换用磁带格式》, 及该标准推荐执行的“中国公共交换格式 (CCFC)”制定的。根据学位论文本身的特点, 本数据规范灵活地对 CCFC 中某些字段做了一些简单的调整。根据学位论文文献的特点, 采用 CCFC 分析级工作单 (1A) 对目标文献 (学位论文) 进行著录。依据该标准建立的“calis 高校学位论文数据库”已有数据约 10 万余条, 参建单位有百余所。

表 2: “calis 高校学位论文数据库”数据规范元素和子元素定义和说明

记录来源	02 0	00	@A	@B			
记录完整程度	02 1	00	@A	记录生成日期	022	00	@A
记录字符集	03 0	00	@B @C	记录语种	031*	00	@A
文献语种	04 0	00	@A*	摘要语种	041	00	@A*
文献载体	05 0	00	@A*	文献类型	060	00	@A
国际标准书号 (ISBN)	10 0	00	@A				

续表2

馆藏索取号	11 1	00	@A	文献号(论文编号)	120	00	@ A
保密级别	19 0	00	@A	学位级别	191	00	@ A
论文注解日期	19 2	00	@A @C	@D			@B
论文题名	20 0	01	@A @a @x	@L			
并列题名	21 0	01	@A @L				
个人责任者	30 0	10 20	@A @E				@a
个人责任单位	33 0	00	@A @D	@E			@B
字段连接关系	08 6	00	@A @C				@B
团体责任者	31 0	10 20	@A @D	@a @E			@B
出版单位	40 0	00	@A @C	@D			@B
出版日期	44 0	10	@A				
载体外形描述	46 0	00	@A @C	@D			@B
注释项	50 0	00	@A				
分类标识	61 0	00	@A @B				
主题标识	62 0	00 10	@A @B				

续表2

文 摘	60 0	00	@A
相关全文指 针	08 9	00	@A @B

calis 高校学位论文数据库”提供的检索点有：题名、论文作者、导师、作者专业、作者单位、摘要、分类号、主题。

参考文献

- [1] 邹志仁.情报学基础.南京:南京大学出版社,1987
- [2] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records: final report. Munchen: Sauer, 1998
- [3] ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations, version 1.00,June 2001. <http://www.ndltd.org/standards/metadata/current.html>
- [4]CALIS 工程文献中心.CALIS 全国工程中心建设“高校学位论文”、“高校会议论文”数据库子项目实施方案(讨论稿). 1998 年 12 月