



Date : 31/05/2006

## ETD building the nutrition for researchers

**Yu Zhenglu and Pan Yuntao**

Information methods research center, Institute of scientific & technical information of China, Beijing, P.R.China

<b>Meeting:</b>	<b>110 Social Science Libraries with Government Libraries</b>
<b>Simultaneous Interpretation:</b>	No
<p>WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL 20-24 August 2006, Seoul, Korea <a href="http://www.ifla.org/IV/ifla72/index.htm">http://www.ifla.org/IV/ifla72/index.htm</a></p>	

### **Abstract:**

*Theses and dissertations are valuable resources for researchers and the ETD format provides a good way to make them more useful. Government and libraries play an important role in the building of Chinese ETD database. This article introduces the present status of ETD database construction in China and the diversification of researchers' needs, as well as the trend of ETD database building in other countries and the analysis of some problem now in China . Some recommendations are given for ETD building to supply better service to researchers.*

## **1. Theses and Dissertations**

Monographs, journal articles, conference and research reports, theses and dissertations are very primary and important sources for research. Theses and dissertations are the important production of universities and institutions, which can reflect the research level and specialty fields of the organizations. Now there are more than 700 organizations in China who can grant doctor's or master's degrees, so there are an increasing number of theses and dissertations every year. What makes theses and dissertations different includes:

- 1) They are a special kind of writing. Writing them is to accept examination and get the degree.
- 2) Most of the research involved in them are focused on the fresh theory or problem which is very urgent and important or very popular. So there are the significant indicators of the science trends.
- 3) They contain many new ideas and results.
- 4) There is no strict length restrict, so the author can analyze and debate the problem totally.
- 5) There are a lot of original charts, data analysis and conferences in the theses and dissertations
- 6) Most of the theses and dissertations are not published and relatively inaccessible.

China restored and established its degree system following the reform and openness in 1978. In order to make better collection, the Degree Office of the State Council has issued some rules since 1983. Including that all the STM theses and dissertations should be sent to the Institute of Scientific & Technical Information of China and the ones in social sciences to the Chinese Academy of Social Sciences, and all the dissertations are sent to the National Library.

## **2. ETDs (electronic theses and dissertations) are the trends**

With the fast development of information industry, ETD is the trend of the theses and dissertations. The ETD building project is appearing everywhere. ETDs are conducive to better collection and preservation. They can promote the sharing and spread of knowledge and technology. They can also save the cost from the long term point of view. ETDs will accelerate the scientific research level of the units and improve the communication of among researchers.

In the mean time, researchers will save more time and money in finding related to theses and dissertations. For one retrieval, they can get more information about different authors. The information can be comprehensive and continuing. Researchers can get more spirits and avoid duplicated work.

ETDs have strong index function, through which we can find more valuable information in addition to theses and dissertations themselves. For example, to certain extent, we can find who are doing what in this field, how good is his/her research and what are the other projects he/she is doing. We can also know how many theses and dissertations are tutored by the same person. Different combination retrieval can make different results, yielding different value.

In fact, there are all kinds of people who need ETDs not only scholars but also people from government, commerce, education, military, etc.

From the economic point, ETDs can be divided into profitable ETDs and non-profitable ETDs. Based on the operators of ETDs, ETDs can be divided into: universities ETDs, Organizations combine ETDs and commercial ETDs.

### **3. ETDs in China**

ETD building is developing very quickly in China. A lot of universities have begun this project and established a series of measures. But the size and level of the ETD building is slower than the Internet instruction, and there are not so many big and powerful ETD databases in China.

There are presently 4 very popular ETD databases in China.

#### **1) CDDB (Chinese Dissertations Database )**

Institute of Scientific & Technical Information of China(ISTIC) is the legal deposit unit for theses and dissertations. ISTIC began to collect STM theses and dissertations in China since 1980.<Chinese Dissertations Bulletin> was born in 1984 which is a kind of index journal made by ISTIC. Wangfang data Ltd transformed it into Compact Disc <Chinese Dissertations Database>(CDDB) in 1995.

Based on the data of ISTIC, Wangfang data Ltd has made a good and rapid progress in ETD. Up to May 2006, the CDDB has almost reached 710,000 records in the CDDB. Now the full text database CDDBFT is being built which archives the theses and dissertations after 2000. Now there are more than 300,000 records in the CDDBFT and it increases 100,000 records every year.

There are three retrieval terms which can be combined in basic retrieval of CDDBFT. The detailed search result fields includes: author, tutor, granting unit, granting time, total pages, language, Chinese library classification number, preserve number, key words and abstract. Below the three index terms, there is classification of subject. The CDDBFT provides very strong dictionary index function.



In addition to theses and dissertation database, there are also full-text journals, full-text conference proceedings, patents, companies, celebrities and other database in Wanfang data Ltd, so Wanfang can provide some unique service to the clients. In the result of theses and dissertation retrieval, there are some hotlinks on the terms of author, professor, subject classification and keywords. In the hotlink of an author, We can find all his/her theses and dissertations, journal articles, conference proceedings, reports and some other information about commerce. The full text browser software is Adobe Acrobat.



2) Chinese Selected Doctoral Dissertations and Master's Theses Full-Text Databases(CDMD)

CDMD is developed by China National Knowledge Infrastructure electronic journal society and TTOD Ltd. CDMD archives theses and dissertations from 1999 to now and the records of CDMD has reached 270,000 up to Dec-31-2005.

There are two sets of retrieval navigations: knowledge classification navigation and subject navigation. CDMD provides three retrieval ways: basic, advance and special retrieval. The full text browser software is Cajviewer. With everyday update of Database Change Service Center, every mirror stand can also update everyday by internet or satellite transmission. Subject compact disc of CDMD updates every year.

CDMD has its own recompense rules:

Every Doctoral Dissertations RMB \$380+ CDMD retrieval card (value RMB \$300)  
Every Master's Theses RMB \$230+ CDMD retrieval card (value RMB \$200)

3) China Academic Library & Information System(CALIS) theses and dissertations database

This ETD database is supported by The China Academic Library & Information System(CALIS). The record of CALLS theses and dissertations database has reached 70,000 up to now. The joining members of this project are 83 universities now.

It collects metadata through OAI open archives initiative protocol for metadata harvesting and makes contact between metadata and local full text by the technology of openURL. So its model is centralized retrieval and distributed full text getting. All the joining universities should collect their own theses and dissertations database and submit it to the Tsinghua university in order to make unitive database.

The first 16 pages of theses and dissertations are free to the joining members. Readers can get full text through Tsinghua university commutative borrowing system or getting contact with the original university library. The copyright is controlled by the local system of the joining members.

In order to encourage the utilization, the organizers lay some brief information about recently theses and dissertations periodically on the university internet.

#### 4) PQDD ( ProQuest Digital Dissertations)

PQDD is published by American company ProQuest Information and Learning.(old name :UMI). By 2000, its Dissertation Abstracts database archived over 1.6 million dissertations and master's theses. The dissertations come from more than 1,000 universities all over the world. The earliest archiving year is 1861.

PQDD is the largest and powerful Dissertations database in the world. It has detailed operation procedure. Most of its systems are healthiness. If the requests of one dissertation are more than 7, the author can get the 10% of the sales as recompense. Part of the dissertations after 1997 can also provide the first 24 pages for free.

PQDD is a commercial organization. After many years development PQDD can provide very good service to students, universities and readers.

Some universities and institutions had combined into a group to order the dissertations form PQDD since 2002. China National Sci Tech Information I/E Corp is the sole agency. The China Academic Library & Information System(CALIS) is responsible for the organization. All the joining members can share the dissertations ordered by the group.

#### **4. What is the problems in China now**

ETD building is very popular in China now, but there are still a lot of aspects need to be improved and some problems need to be solved.

- 1) There is a system of theses and dissertations legal deposit; however, universities and institutions are not under the subjection of the archiving organizations, so the collecting work is not very smoothly and some theses and dissertations are not integrity.
- 2) The criterion of the abstract. Some abstracts of theses and dissertations are lengthy and include some charts and formulas. This does not accord with the international criterion.
- 3) The choosing of keywords. Some students do not get the meaning of keywords, so some keywords are far away from the topic of the theses and dissertations. It will recede the effect of retrieval.
- 4) The data's retrospect is very difficult. Because paper is difficult to preserve for a long time, the procedure of retrospect is complicated and costly.
- 5) The building of metadata criterion. The China Academic Library & Information System(CALIS) has put forward a plan to adopt standardization metadata criterion.
- 6) Copyright. The whole transfer process should protect the right of authors, libraries and users. The theses and dissertations transfer on internet can be regarded as publication, so authors' rights should be protected. Libraries make the database and release it, so their right should be protected. And users' rights to use the database

legally should be protected too.

According to copyright law in China, only original creation database can be protected. Original creation emphasizes that there should be some original creation in resource choosing and database building.

- 7) Some students don't agree to release their theses and dissertations. There are three common reasons: (1) the research results are going to be published on the professional journal, so they put off the release. (2) the theses and dissertations are related to secrets. (3) the research results include some patents.

## 5. ETD building in other countries.

### 1) Virginia Tech-ETD(USA)

USA has been paying more attention to the ETD building. Before the World War , some libraries and universities began a series of activities including: copying and microfilming the theses and dissertations etc. Virginia Tech-ETD (VT-ETD) is one of the pioneer in the field of ETD building. Virginia Tech university had begun to research the project of ETD with UMI company and some other organizations since 1987.

By now, submitting the ETD has become a basic request to graduates. The reason for building VT-ETD was that so few masters' theses were sent to UMI company which was the largest theses and dissertations collecting organization. VT-ETD is established by Virginia Tech University Library and Graduate School. Up to now, there are 7941 theses and dissertations in VT-ETD.

VT-ETD pays more attention to the facts analysis. From the table below, we can see that the accesses by commerce add up to 405,302 in 2002/03, occupying 49.4%. The accesses come from education, commerce, government, military, network and Non-Profit Organizations.

Accesses by US Domains

domain extension	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	Domain
.edu	15,314	112,876	372,348	525,855	941,736	207,911	Educational
.com	5,309	48,540	341,198	562,292	1,220,881	405,302	Commercial
.gov	282	1,362	7,564	14,870	31,403	20,183	Government
.mil	188	1,872	6,153	10,441	22,442	5,174	Military
.net	2,522	14,026	109,053	282,521	756,267	172,661	Networks
.org	375	3,132	9,291	9,851	26,810	9,460	Non-Profit Organizations

(<http://scholar.lib.vt.edu/theses/data/somefacts.html>)

The next table is about the percent availability of VT-ETD. From 1998 to 2004, there are more and more students agreeing to put their theses and dissertations available online. On the contrary, the students restricting from all access is decreasing.

Percent availability of VT ETDs

	1998	1999	7/2000	9/2001	2/2002	2/2004
<b>Approved ETDs available online</b>	<b>85.5%</b>	<b>82.6%</b>	<b>80.7%</b>	<b>75.6%</b>	<b>81.7%</b>	<b>88.1%</b>
Available without restrictions	63.1%	54.2%	53.5%	52.5%	52.5%	70.7%
Available to VT only	21.7%	26.1%	24.3%	26.2%	26.3%	26.5%
Available with mixed access	0.6%	2.3%	2.9%	3.1%	3.0%	2.8%
<b>Approved ETDs restricted from all access</b>	<b>14.5%</b>	<b>17.4%</b>	<b>19.3%</b>	<b>18.2%</b>	<b>18.3%</b>	<b>11.9%</b>

(<http://scholar.lib.vt.edu/theses/data/somefacts.html>)

## 2) Dissonline(GERMANY)

German National Library has begun to collect ETD on internet and built index with related metadata for future usage in 1998. The library has the charge of the transfer of metadata and long-term accesses to ETD. Later German National Library founded co-ordination agency in order to promote the ETD building. This project was supported by German research center committee DFG .

The Dissonline had archived 97 ETDs by 1998, 20,957 by 2003 and 24,158 by 2004.

The co-ordination agency's work includes:

- Contact point and information exchange
- Incitation of co-operation of the participating institutions and know-how transfer
- Co-ordination of technical advancements and adjustments
- Co-operation with national and international organization

## 3) United Kingdom

The British Library is not the theses and dissertations legal deposit organization, but archives a lot of theses and dissertations. The British Library found British theses and dissertations service center which has archived 150,000 theses and dissertations. With the development of information technology, the U. K established theses and dissertations online group in 1994. It can retrieve 460,000 ETD abstracts from 1716.



#### 4) Canada

The National Library of Canada has started the ETD plan and also founded a ETD consulting committee which researches ETD operation models and solves the problems occurring in the implement.

### **6.Recommedations to the ETD building in China**

First of all, set up enough criterions in the ETD building, such as writing format, table and chart format, abstract, keywords and subject classification.

Secondly, take more efforts in the field of related IT technology. There are a lot of researches on Metadata projects all over the world now. An item team from USA has put forward the ETD-ms : an Interoperability Metadata Standard for Electronic Theses and Dissertations. More extensive applying and powerful metadata in the library and information industry is the DC(Dublin Core)Metadata Element Set.

Thirdly, cultivate more professional human resources and build good service system and coordinate organizations. These measures can assure the collections of ETDs and make it more effective.

#### **References:**

- 1) Lu Cheng, A Comparison on the Dissertations Databases between PQDD and ProQuest [J], Library Tribun, 2005(25)01:85
- 2) Cui Huimin, Talk about the Intellectual Property Protection of Academic Dissertation [J], Library Construction, 2005(00)02:21
- 3) He Defang, Research on the National Dissertation Service System [J],Journal of the China Society for Scientific and Technical Information,2004(23)06:697
- 4) Wang Ping, The Copyright Problems of the Academic Thesis Database Building [J],Library Construction, 2006(00)01:17
- 5) Nikola Korb, formation of ETD's and related issues[C], 6th ETD Conference May 20 –24 2003, Berlin
- 6) [http://www.edu.cn/20010906/3000497\\_2.shtml](http://www.edu.cn/20010906/3000497_2.shtml)
- 7) <http://ckrd.cnki.net/grid20/Navigator.aspx?ID=2>
- 8) <http://scholar.lib.vt.edu/theses/data/somefacts.html>
- 9) [http://www.dissonline.de/index\\_e.htm](http://www.dissonline.de/index_e.htm)