Chapter 7
Theses and Dissertations

Christiane Stock and Pierrette Paillassard
INIST-CNRS, France

7.1 Thesis and/or dissertation – terms and scope

According to Wikipedia\(^1\): "A dissertation (also called thesis or disquisition) is a document that presents the author's research and findings and is submitted in support of candidature for a degree or professional qualification. In some countries/universities, the word thesis is used as part of a Bachelors or Masters course, while dissertation is normally applied to a Doctorate." However, this usage is non consistent throughout the countries.

A doctoral thesis is the result of 3-4 years of research and the first valuable document in the career of a researcher. At the same time it is an administrative document necessary to obtain the doctoral degree. In some disciplines theses are considered as the result of teamwork and appear in the list of publications of the laboratory where the research was done (Mermet et al. 1998)

Limitations of the "study": the main focus of this chapter will be set on doctoral theses or doctoral dissertations. In many countries Master theses are not considered worth an effort of dissemination in the academic context. Considerations about the quality of the content may influence the decision of non-dissemination. We also exclude initiatives outside the academic or research context, e.g. initiatives taken by students to disseminate their works through associative websites. These private initiatives may not provide stability over time nor make any control of the input.

7.2 A short history of dissemination

Where can doctoral theses or dissertations be found?

Since they are produced in the universities, they were deposited in the university library and included in the library catalogue. Copies were made available to other universities through inter-library-loan.

---

Besides library catalogues doctoral theses were included early on in monthly/annual bulletins or indexes, then in databases on a national or international level, such as the catalogue of the British Library and its monthly bulletin of "British Reports, Theses and Translations".

In France a four-level national network for theses was established following a decree published by the French Ministry of Education in 1985. It included among others the systematic reproduction and dissemination of the documents on microfiche the dissemination of a copy to all French universities as well as the systematic referencing through a national database: Téléthèses. The latter has been integrated into the national union catalogue for university libraries, SUDOC*.

UMI*: the most important and long-lasting initiative worldwide is found in the United States: University Microfilm International (UMI*) in Ann Arbor collects (since 1938) abstracts and full texts of doctoral dissertations from North American and European universities which are indexed in Dissertation Abstracts International (paper and microform), and its internet database Dissertation Abstracts Online, now ProQuest Dissertations & Theses (PQDT) (UMI 2009). More than 90% of the doctoral dissertations accepted by North American universities are covered by UMI*. ProQuest UMI Dissertation Publishing has published over 2 million graduate works in 70 years.

SIGLE*: in Europe the SIGLE* database (System for Information on Grey Literature in Europe) played a major part in providing access to doctoral theses. Produced by EAGLE (European Association for Grey Literature Exploitation), it was entirely dedicated to the collection and dissemination of grey documents to which theses and dissertations belong. Its 15 members provided records facilitating the identification of the documents and held a copy for document delivery on demand. More than 275,000 records referenced doctoral theses, covering the period from 1980 to 2004, representing about one third of the database. The majority of the database is now in open access at http://opensigle.inist.fr.

NDLTD*: is another milestone in the doctoral theses universe. The Networked Digital Library of Theses and Dissertations (NDLTD*) was created in 1996 by Virginia Tech (Fox 1996), funded by the U.S. Department of Education and the Southeastern Universities Research Association in order to improve graduate education. The outcome of the project and further activities have greatly influenced the transition to electronic dissertations throughout the world. Work on electronic theses and dissertations started as early as 1987 when Virginia Tech developed the first SGML Document Type Definition (DTD) for theses and dissertations (Fox 1996).

Today’s proclaimed aim of NDLTD* is to promote the "adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations." It federates more than 100 members and offers access to almost 800,000 online documents through the NDLTD* Union Catalog3 based on OAI-PMH.

2 NDLTD, http://www.ndltd.org
3 NDLTD Union Catalogue, http://alcme.oclc.org/ndltd/
The NDLTD* website provides a host of tools and guides for electronic theses and dissertations developed by the federation.4

A comprehensive guide (ETD Guide 2009) covers on a technical level all issues for universities and students.

The ETD metadata scheme "ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations" has been adopted worldwide. It adds an element "theses.degree" to the basic Dublin Core metadata scheme with the following qualifiers5:

thesis.degree.name
thesis.degree.level
thesis.degree.discipline
thesis.degree.grantor


7.3 From the digital library to digital workflow -
   a new organisation for electronic theses

Electronic theses and dissertations can be produced in different ways. The easiest way is to digitize a paper copy. Converting a text format into a PDF file comes as a second. However, changing to the electronic version is not simply a change of the support of dissemination, but needs a whole new organization.

In France, first steps in the electronic era were taken in the 1980’s-90’s when universities digitized their theses, often in PDF format, and put them on the internet.

Thus INRIA*'s doctoral theses were accessed from the 90’s through a simple HTML webpage on its institutional website. Since 2005, students are encouraged to submit their document to the self archiving TEL-HAL*6.

This archive is today the most comprehensive French repository. It was created by the CCSD* and MathDoc*, one of the oldest French archives.

MathDoc* and Grisemine (now IRIS*) were among the first to digitize through paper copies. Since 1997 another of the top French engineering universities, INSA Lyon* offered its students different tools to produce electronic theses: converting a text format into a PDF file.

At the same time, a common project between Canadian and French universities (Montreal, Lyon) proposes a complete editorial chain in open source software called "Cyberdocs*". It covers aspects from document model to the conversion into a fully structured XML document using the TEI Lite DTD.

---

4  NDLTD-documentation, http://www.ndltd.org/resources
Following NDLTD*, the German project Dissonline* was one of the first to tackle the problem with a holistic approach: separate workgroups dealt with metadata issues, publication tools, development of the server platform OPUS, creation of a new workflow, the provision of training for students and administration staff, and last but not least help in resolving legal aspects. Universities had to revise their graduation regulation to accept electronic versions as valid documents. Authors rights and German copyright laws were other issues that need to be dealt with.

The successful outcome of the programme is to be seen in the catalogue of the German national library which includes more than 78,000 records of electronic dissertations.

France proceeded in several steps, defining first a national metadata scheme (TEF*), before working on the workflow and server issues. Valuable information can be found on the website ORI-OAI*.

### 7.4 Dissemination of ETDs today

Next to journal articles and eprints, electronic theses and dissertations (ETDs) are for various reasons the most frequent document type found in open archives.

- ETDs are a well defined and well referenced document type. Rules for deposit and citation are generally established on a national level, and international standards exist for specific information and theses metadata, contrary to other grey documents.
- ETDs are administrative documents, and students can be "obliged" to deposit their work in an archive or repository in order to obtain their diploma.

<table>
<thead>
<tr>
<th>Continent</th>
<th>Repositories registered</th>
<th>Repositories with theses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>686</td>
<td>372</td>
<td>54%</td>
</tr>
<tr>
<td>North America</td>
<td>404</td>
<td>160</td>
<td>39%</td>
</tr>
<tr>
<td>Australasia</td>
<td>77</td>
<td>54</td>
<td>70%</td>
</tr>
<tr>
<td>Asia</td>
<td>167</td>
<td>81</td>
<td>48%</td>
</tr>
<tr>
<td>South America</td>
<td>68</td>
<td>32</td>
<td>47%</td>
</tr>
<tr>
<td>Africa</td>
<td>24</td>
<td>17</td>
<td>70%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,426</strong></td>
<td><strong>716</strong></td>
<td><strong>50%</strong></td>
</tr>
</tbody>
</table>

According to the OpenDOAR* - the Directory of Open access repositories registry - 716 out of 1,426 repositories (50%) contain theses. Only for North America the

---

percentage is significantly lower. Here, we cannot distinguish between master’s theses and doctoral theses.

It should be noted that not all repositories with ETDs are registered in Open-DOAR*, while some sites included in the table may still be at an experimental stage and contain either very few records or only test records. The data above don’t provide any information on the real number of theses available through those repositories.

7.4.1 From auto-archiving to portals

The first archives for scientific publications including ETDs were based on the principle of self-archiving or deposit by the author, intended for fellow researchers. The documents submitted were reviewed for formal aspects or their contents; and the metadata were entirely supplied by the authors. An early example in France is TEL* (thèses en ligne), which has since been included into HAL*.

A second phase saw the creation of institutional repositories (IR) including the scientific production of a given university or research institute. IRs fill the role of a showcase to the world, but they are also increasingly used as a tool for managing research projects. We see new (old) actors such as librarians or administrative staff entering the scene. Metadata are controlled and even enhanced, which implies a gain in consistency, and the validity of the document is checked or controlled through the workflow.

Along with these institutional repositories we observe an increasing number of portals or websites offering a federated research, either for all types of scientific output, or dedicated to ETDs.

- On a worldwide level OAIster* provides access to electronic theses deposited in archives compliant with the OAI protocol for metadata harvesting the NDLTD Union Catalogue does likewise.
- The NARCIS* portal in the Netherlands receives input from all Dutch universities as well as from national research organizations. Its subset “Promise of science” is dedicated to doctoral theses.
- The Scandinavian portal DIVA* references doctoral and even more Master theses from 24 participating colleges and universities.
- EThOS* (Electronic Theses Online System) as a project funded by JISC (Joint Information Systems Committee) and RLUK (Research Libraries UK), opened in 2009 in its Beta version. It’s a "one stop shop" for secure access to research theses, simultaneously increasing the visibility of UK Higher education postgraduate research, and providing additional services such as print on demand or digitization of paper documents.
- In France, PASTEL*: theses from ParisTech (Paris Institute of Technology) includes 12 engineering schools. Besides providing free access to
ETDs, ParisTech developed since 2003 the project "ParisTech Graduate School" with access to open courseware.

- DART Europe is a partnership of research libraries in 14 European countries supported by LIBER (Association of European Research Libraries). DART-Europe is the European Working Group of NDLTD and allows for search of 110,000 full text theses.
- Since 2004, about 40 Canadian universities deposit metadata and electronic theses in the Theses Canada Portal. Moreover, theses and dissertations digitized between 1998 and 2002 are freely accessed. In 2008: it contained 300,000 ETDs on microform and 50,000 are also available electronically.
- The Australasian Digital Theses Program was funded by the Australian Research Council (ARC). The document format standard is PDF and the portal is OAI-PMH compliant. In 2009, 25,000 digital theses out of 150,700 from 41 universities are available in electronic format.

7.4.2 Quality issues and preservation

A viable website should at the same time ensure quality issues and offer long-term digital preservation.

Archiving and preservation aspects are treated with various approaches in terms of technologies used.

In Europe, DRIVER "a pan-European infrastructure for digital repositories" (Robinson, M. et al. 2009) emphasizes the need for targeted, refined and standardized harvesting. Thus, DRIVER recommends the following qualifications for “type” element:

"info:eu-repo/semantics/bachelorThesis",
"info:eu-repo/semantics/masterThesis",
"info:eu-repo/semantics/doctoralThesis (Bologna Convention)"

instead of various qualifications used:

"Cranfield:  <dc:type>Thesis or dissertation</dc:type>
<dc:type>Doctoral</dc:type>
<dc:type>PhD</dc:type>

DIVA:  <dc:type>text.thesis.doctoral</dc:type>

Humboldt:  <dc:type>Text</dc:type>
<dc:type>dissertation</dc:type>"

(Robinson, M. et al. 2009)

France developed a standard called TEF (AFNOR recommendation since 2006) to doctoral theses. TEF include a FRBR (Functional Requirements for Bibliographic Records) data representation and “defines a set of preservation metadata

---

10 Australasian Digital Theses Program, http://adt.caul.edu.au
that will permit long-term thesis preservation” (Boudia, D. et al. 2005) The preservation is maintained by PAC* (Archive Platform at CINES*) based on ISO standard 14721.

At the international level, NDLTD* with the "MetaArchive cooperative project*: digital preservation" uses LOCKSS11 (ETD 2009)

With regard to the quality of information, the risk of plagiarism has become an important obstacle to the deposit of ETDs in open archives (Davis et al. 2007), a phenomenon which increased with copy and paste facilities and the use of machine translation software.

However, plagiarism is not a recent problem and can be « solved » before deposit. A report by the Commission Éthique-Plagiat funded by the University of Geneva "La relation éthique-plagiat dans la réalisation des travaux personnels par les étudiants"(2008)12 underlines the importance of training students on document retrieval, and the role of evaluation and control by the authorities (teaching staff). The study recommends the use of software to detect plagiarism.

The attribution of a date stamp during the deposit proves the priority of the work.

7.4.3 Access to full text and confidentiality

Many academic and research institutions have defined policies, declaring the deposit of scientific works in their institutional (or national) repository a mandatory step for either the evaluation of a researcher or the defence of a doctoral dissertation. Academic regulations exist since 2005 for masters and PhD students at the University of Edinburgh13 and since 2006 at the University of Liège (Belgium)14 and at Leiden University (Netherlands)15

The need for restricted access to parts of the documents may create obstacles to these policies, but different kinds of solutions have been developed. Indeed a mandatory deposit does not necessarily include the authorization for worldwide dissemination of the full text. Aspects like confidentiality or quality criteria may lead to a restricted access to the document itself, e.g. through an intranet or on the campus, while the metadata are available to everyone. Users of such repositories find the information at different levels or stages of their visit.

Partial access to the full text: Authors may be charged with the submission process only once. What procedure to adopt if the dissertation is declared confi-
On the technical level software platforms allow the author to deposit the document once, but separated into different files, and to declare some parts under embargo. A user from outside the campus may come across a doctoral thesis for which the table of contents, introduction and conclusion as well as part 1 and 2 are freely accessible, but the link given for the other parts won’t work for him. The University of Leiden has adopted this practise for its repository, including even the date of the end of embargo.

Finland offers a different solution. Many Finnish Doctoral theses are based on articles. The full text documents accessible through the repository list the articles in the table of contents and include a summary in the corresponding chapter. The publications seem to be joined as appendices to the thesis, but are not included in the open access version.

A third way to deal with restricted parts of ETDs is shown by the University of Oslo repository DUO*. The bibliographic record includes the list of papers with a hyperlink to the commercial publisher. So the end-user may access these parts if he has a subscription to the journal or if he's willing to pay for it (example: http://wo.uio.no/as/WebObjects/theses.woa/wo/0.3.9)

**Records without theses:** An increasing number of repositories contain bibliographic records without full text. To support authors with their deposits and to increase their willingness to participate, many repository administrators have started to upload citations from external sources. In some cases the metadata for ETDs originate from library catalogues, to be completed by the author. The absence of the full text is then a phenomenon limited in time. Duplicate entries in repositories may be generated when metadata are added in the deposit process and are uploaded from a bibliographic file. For other universities the institutional repository should reflect the scientific production. Comprehensiveness of bibliographic records prevails over access to the documents.

How are the users informed about the presence/absence of full text documents? Again different solutions can be observed (for details see Stock 2008):

- A check box requesting “full text results only” in the search menu
- Use of icons in the list of results
- Information in the record display
- On specific pages dedicated to a community, a department, etc.

**Other solutions:** The Imperial College London Repository has chosen to give public access only to validated full-text dissertations, while an in-house platform is used for the workflow and administrative purposes (Jones 2007)

In France all doctoral theses must be referenced in the national bibliography, whereas access to the full text (paper or electronic form) may be "confidential". The present metadata scheme TEF* (TEF 2007) allows to identify and exclude passages not to be disseminated, and anticipates the co-existence of a complete version and a public version of the document.

What happens if the repositories are harvested by service providers?
Several checks with OAIster* showed that in some cases the service succeeds in eliminating OAI records without full text (e.g. Lund), while not in other cases. Hasselt University exposes to metadata harvesting only those documents for which the full text is available (Goovaerts 2007).

7.4.4 Masters theses and other student’s works

PhD theses are often subject to academic regulations concerning their dissemination. Ideally, the degree is only obtained when the document is published in print or electronic form. Master’s theses are far less controlled by legal dispositions. Conservation of the paper copy in the local library is not always guaranteed and consultation by other students may be subject to authorization.

Disseminating master’s theses through repositories is a common occurrence in northern countries. The Scandinavian portal DiVA* provides access to theses at different levels, from master thesis to first term paper. The question whether or not to include master's theses in open archives gave way to ardent exchanges on German discussion lists in the past and is not clear cut in France either. Repositories are seen as a showcase for scientific output, and student's works on a master or even bachelor level don't count as scientific publications, therefore don't belong in this category.

Dissemination of master theses may follow different objectives: making one’s work known to fellow students and subsequent as well as to alert future employers. Thus THESA*'s references of cutting edge theses subjects are geared to the economic world, whereas "DUMAS - Dépôt Universitaire de Mémoires Après Soutenance"16 aims to increase the visibility of master theses as well as the teaching activity of the universities. Others like (mémSIC17) instead make a selection of the best works.

7.5 Outlook

The landscape of online access to theses has changed in many ways over the past years. Digitizing paper copies was re-placed by workflows covering every stage of the production and dissemination of a thesis, including metadata and quality issues. Access through a list of titles on a webpage gave way to portals regrouping multiple repositories. One of the major actors in this process is NDLTD.

Technical developments have greatly facilitated the dissemination of ETDs, increasing their visibility to a worldwide level. However, legal issues especially copyrighted parts of third authors used in the thesis became an obstacle. Different solutions have been found and adopted so far. Indeed the growing number of the-

16 DUMAS - Dépôt Universitaire de Mémoires Après Soutenance, http://dumas.ccsd.cnrs.fr/
ses "available" on the internet comes with an increasing diversification as to the kind of access available to the full text and to its contents. Repositories mix full text entries with records without documents or theses with only partial access. Institutions alert users about these differences, but not in a consistent way. And, this diversification extends to other students works.

In an earlier study (Stock 2008) we observed important differences between countries with regards to the number of theses available online, as well as to the percentage written in English. Scandinavian countries as well as Belgium (and the Netherlands) are highly tolerant with regards to the language choice for a thesis: 50 to 90 percent of ETDs appear in English. France on the other hand has a low percentage of theses that are not written in French. "Pioneer deposits" of theses in repositories are mostly written in English and seem to indicate the willingness to give the widest access possible to one's work both through the choice of language and through the internet.

The growing complexity of the ETD landscape calls for explicit policies in the future to inform the user of a given repository on its deposit, validation, access and reuse of a thesis. Useful tools like OAIster* or the sites of service providers who harvest their records should be examined with a critical eye, especially when the primary need of the end-user is access to full text.

Glossary


**AFNOR**: Association Française de Normalisation (French standardisation organisation) http://www.afnor.org/

**Australasian Digital Theses Program**: http://adt.caul.edu.au/

**CCSD**: Centre pour la Communication Scientifique Directe (CNRS unit) http://www.ccsd.cnrs.fr/?lang=en

**CINES**: Centre informatique national de l’enseignement supérieur http://www.cines.fr/

**CNRS**: Centre National de la Recherche Scientifique (French National Research Organisation): http://www.cnrs.fr/index.php

**Cyberthèses**: thesis electronic archive and diffusion program http://www.cybertheses.org/?q=en/node/32

**Cyberdocs**: collaborative development site of Cyberdocs platform http://www.cyberdocs.org/en

**DART-Europe**: E-theses Portal http://www.dart-europe.eu/basic-search.php

**DissOnline**: German project http://www.dissonline.de/eng/links/index.htm

**DRIVER**: Digital Repository Infrastructure Vision for European Research http://www.driver-repository.eu/


**DUO**: Digitale utgivelser ved UiO http://www.duo.uio.no/englishindex.html
EThOS: Electronic Theses Online System http://www.ethos.ac.uk/
INIST: Institut de l’Information Scientifique et Technique (CNRS institute for scientific and technical information): http://international.inist.fr/
INRIA: French national institute for research in computer science and control: http://www.inria.fr/index.en.html
IRIS: digital library of University of Lille 1 https://iris.univ-lille1.fr/dspace/
LIBER: Association of European Research Libraries http://www.libereurope.eu/
MetaArchive cooperative: digital preservation http://www.metaarchive.org/
NARCIS: portal of Dutch universities http://www.narcis.info
NDLTD: Networked Digital Library of Theses and Dissertations http://www.ndltd.org/
OAiStor: union catalogue of digital resources http://www.oaister.org/
OpenDOAR-the Directory of Open access repositories: http://www.opendoar.org/
OpenSIGLE: System for Information on Grey Literature in Europe: http://opensigle.inist.fr/?locale=en
ORI-OAI: Outil de Référencement et d’Indexation en réseaux de portails compatibles OAI-PMH http://www.ori-oai.org/display/ORIOAI/ORIOAIORG
ROAR: Registry of Open Access Repositories http://roar.eprints.org/
SIGLE: see OpenSIGLE
STAR: Signalement des thèses, archivage et recherche (referencing, archiving and retrieval of ETDs): http://www.abes.fr/abes/page,428,star.html
THESA: THESA provides information on doctoral theses currently under way in the French accredited higher education establishments (Grandes Ecoles) http://thesa.inist.fr/eng/Accueil.htm
Theses Canada Portal: http://www.collectionscanada.gc.ca/thesescanada/index-e.html
SUDOC: Système Universitaire de Documentation (academic union catalogue of serials and monographs): http://www.sudoc.abes.fr/DB=2.1/LNG=EN/START_WELCOME

All websites visited in June 2010
Bibliography


All websites visited in June 2010