

Integrating Electronic Dissertations in a Regular Library Workflow

Alice Keller

ETH Library Zurich

alice.keller@library.ethz.ch

ETH-Bibliothek, Raemistrasse 101, CH-8092 Zurich, Switzerland

www.ethbib.ethz.ch/

Keywords: Electronic Dissertations, Workflow, Library Catalogue

Abstract

The Swiss Federal Institute of Technology (ETH) Zurich introduced electronic dissertations in the year 2000. During the first phase the ETH Library scanned print documents retrospectively and made them available over the internet. Later authors were encouraged to provide the library with their own digital versions. With this very pragmatic approach a large collection could be built up within a short time.

During the initial phase, a special project team was set up to develop the new service. However, the aim of the management was to integrate all electronic dissertations in the regular workflow of the library. Today all electronic documents are processed by the regular departments (exchange section, cataloguing and classification depts). Furthermore, electronic dissertations form part of the library's major electronic publishing initiative ETH E-Collection.

All electronic dissertations are catalogued in the library SYSTEM Aleph 500 according to international standards (MARC, AACR2). However, the library also wanted to offer direct access over the ETH E-Collection web pages. For this reason selected metadata is extracted daily from Aleph 500. Perl scripts are used to create browsing lists (title, author name and subject) and individual title pages for each document.

Besides offering access to internal dissertations, the ETH Library tries to enrich the library catalogue with references to as many external electronic dissertations as possible. This work is done by copy-cataloguing, mainly from the German national bibliography. The task is fully integrated in the regular cataloguing and classification departments.

This presentation gives an overview of the chances and problems a library experiences if it tries to integrate a large number of internal and external electronic dissertations into the regular workflow.

All doctoral dissertations are collected by the library. The ETH Library, or ETH-Bibliothek, is the main library of the Swiss Federal Institute of Technology. It is primarily dedicated to provide relevant information to all areas of science and technology as covered in teaching and research at the ETH. Furthermore, it has a key role as Swiss national centre for sci-tech information.

With collections comprising 5.7 million items, 6'000 current print journals and 5'300 e-journals, the ETH Library belongs to the largest special libraries of Europe.

As with all university libraries, collections and services are tailored to support teaching, learning and research. As these activities change throughout the years, the ETH Library constantly re-assesses, re-evaluates and re-thinks its vision and strategy. The library has defined a flexible e-library strategy that covers both digital text and image collections and defines developments in three directions¹. These directions can be described as follows:

1. offering new services in physical space (e.g. larger departmental libraries in place of smaller special libraries, PC working space for users),
2. offering traditional services in a virtual space (e.g. include abstracts in OPAC; extended collection of e-journals, e-texts and databases; electronic dissertations) and
3. developing new services in a virtual space (e.g. ETH E-Collection as alternative publication platform; innovative picture information SYSTEM EPICS; project MyLibrary).

As shown here, building a collection of electronic dissertations falls within the second category.

The collection of electronic dissertations was launched in autumn 2000. In order to offer a substantial collection right from the beginning, authors of the years 1999/2000 were approached SYSTEMatically. They were asked whether they would agree to have their doctoral dissertation made available online. Thanks to this initiative many titles dating back to 1999 are accessible. During this first phase the ETH Library scanned print documents and made them available over the internet.

The reactions of scientists to the new electronic publishing initiative were so positive that the ETH Library soon decided to widen the scope of the collection and integrate other document types, too. Thus the collection

History of the Electronic Dissertations at ETH

The Swiss Federal Institute of Technology (ETH) Zurich is the largest technical university in Switzerland and covers all areas of science and technology. There are currently 2'200 PhD students at the university, nearly half of whom are from abroad. The annual number of PhD titles granted lies between 500 and 600 which means that doctoral students spend about 4 years at the ETH Zurich.

¹ A. Keller, R. Mumenthaler and W. Neubauer: Electronic library services as part of the university's learning environment, ETH-Bibliothek, 2002 (<http://e-collection.ethbib.ethz.ch/show?type=inconf&nr=39>).

of electronic dissertations was extended and renamed ETH E-Collection in 2001².

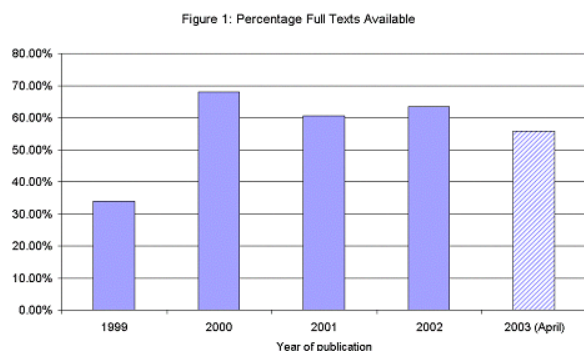
This new initiative ETH E-Collection received substantial support from the university and represents one of the key projects within the major strategic initiative ETH World³.

As table 1 shows, the electronic dissertations still form the major substance of the ETH E-Collection. However, the number of other document types is increasing rapidly. Whereas the output of doctoral theses remains constant (500-600 per annum), the number of other documents can be raised well above 600 per annum if scientists co-operate.

Table 1 : Contents of ETH E-Collection shown according to document type (March 03).

Document type	Number of documents
ETH dissertations (with full text)	1'300
ETH dissertations (only abstract)	892
Dissertations from other universities	4
Diploma theses	89
Reports	1'137
Conference papers / presentations	168
Other document types	132
Journals	34
Total	3'756

Getting Full Texts from the Authors



As in all electronic publishing projects persuading the author to co-operate is crucial for success. The author, as copyright holder, needs to agree to make her or his work available online. As already mentioned above, initially print dissertations were scanned. Full texts were put in the internet as soon as the written declaration of consent from the author was obtained. For this purpose all authors were approached personally with a letter outlining the aim of the project and asked to sign and return a form. As from 2001 authors were encouraged to provide the library with their own digital versions.

Currently doctoral students are given this information including the form of consent when they hand in their dissertation at the doctoral administration. This procedure ensures that all PhD students are (in principle) aware of the service. Furthermore, all documentation is available online on the website of the library⁴. If no written consent is granted, the ETH Library digitises the abstract and makes only this summary available online.

Naturally the ETH Library would like to make as many full texts available online as possible. However, despite very positive, even enthusiastic feedback from students and scientists, the library has not managed to raise this figure significantly above 60% during the last four years (figure 1). This means that although all doctoral dissertations of the ETH are recorded in the ETH E-Collection, only 60% of the records lead to a full text. (The other 40% lead to an electronic abstract only.)

Unfortunately, the reasons why authors withhold their consent are not fully known. Staff experience from the exchange division shows that only very few authors are actually opposed to the service. Some are apparently not fully aware of the service, others have lost the form, or don't care to make a contribution themselves. In German we summarize these reasons under: "vergessen, verlegt oder verreist".

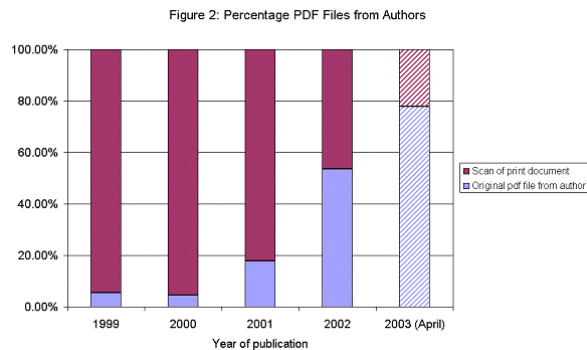
The fact that the number of full texts from 2003 is slightly lower than in 2002 is not worrying - we expect that more declarations of consent are still to come in!

The ETH Library now plans to launch an awareness campaign and approach individual authors more directly. Ideally an intensified co-operation between the doctoral administration of the university and the exchange division at the library would lead to a heightened awareness amongst PhD students.

2 ETH E-Collection: <http://e-collection.ethbib.ethz.ch/>.

3 ETH World: <http://www.ethworld.ethz.ch/>.

4 Advice to doctoral students: http://e-collection.ethbib.ethz.ch/about/dissertationen_e.html.



Whilst the library is struggling to raise the percentage of full texts beyond 60%, the number of original author files is rising steadily. As already mentioned above, dissertations were digitised retrospectively in the initial phase. Soon, however, the library started to encourage authors to send in (or rather upload) their own pdf full texts. Currently 77% of all new full texts are original pdf files provided by the authors (figure 2). This development, of course, marks a considerable improvement, both in respect to the size of the file and the quality of the images. In cases where the author cannot provide a file, the library continues to digitise the print copy. The costs for digitising are borne by the library.

Workflow of Electronic Dissertations at ETH Library

As in all projects a project team was given the task of designing and implementing the new service. Their main tasks were to set up the web site, develop the metadata exchange between library catalogue and website, and promote the new collection in the ETH community. Once the website was launched and all technical specifications clarified, the full responsibility for processing of electronic dissertations was passed on to the regular staff. Today all electronic documents are checked in and catalogued by the regular departments (exchange section, cataloguing and classification). The responsibility for the whole ETH E-Collection, the electronic publishing framework within which all doctoral dissertations are made available, lies with a new project team.

Although the electronic dissertations are heavily used and very popular, the print copy remains the official version of a doctoral thesis at the ETH Zurich. For this reason the library still receives three print copies of each dissertation irrespective of the online availability⁵. These copies are catalogued traditionally in the library catalogue NEBIS (Aleph) using MARC format and AACR2.

The internet address of the electronic version can be assigned using the specific prefix of the ETH E-Collection and the number of the dissertation.

Example:

`http://e-collection.ethbib.ethz.ch/show?type=diss&nr=14543`

At the same time the whole document is digitised by a commercial contractor. This procedure ensures that all abstracts can be made available online within less than a fortnight. The full text is only provided when the authors' declaration of consent arrives. This can happen very soon or only years later! If the author provides a pdf file, the exchange division checks it and switches it online. Alternatively, the scanned version is taken. If no consent arrives within a couple of months, the exchange division produces microfiches and sends them to the exchange partners upon request.

As more and more full texts are provided by authors, it will no longer be necessary to scan all incoming dissertations by default. This change is currently being evaluated.

Technical Implementation

All dissertations are integrated in the larger electronic publishing platform ETH E-Collection. Documents are primarily recorded in the library catalogue NEBIS and also shown on the website of the E-Collection. Metadata is extracted from the library catalogue on a daily basis and reformatted for the browsing lists and title pages in the ETH E-Collection website.

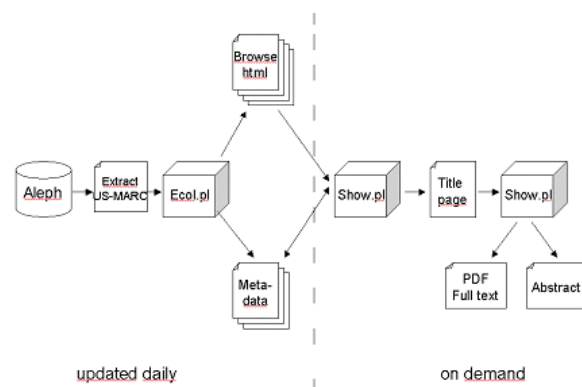


Figure 3: Technical workflow.

The technical workflow in figure 3 shows how metadata is exported from the library SYSTEM NEBIS (Aleph). A special MARC field (909EE \$a) was defined to identify documents for the ETH E-Collection. Using the exported metadata, a Perl programme (Ecol.pl) generates browsing lists for the web and metadata for the title page.

5 In addition the library receives an unbound copy which serves as master copy for microfilming or digitisation.

es. These browsing lists offer easy access via document type, author name or subject.

As soon as a document is selected from a browsing list, a second Perl programme (Show.pl) is launched that puts together the title page for an individual document using available metadata and the web template. These title pages are generated when requested, or rather on-the-fly. They offer a link to the library catalogue NEBIS, to the abstract and - where available - to the full text (fig. 4).



Figure 4: Title page of electronic dissertation in ETH E-Collection.

Currently the ETH Library is working at the implementation of the OAI protocol to make the metadata available in other collections or projects such as the NDLTD, GetInfo (TIB Hannover and FIZ Karlsruhe) or E-Helvetica (Swiss National Library)⁶.

Access to External Electronic Dissertations

It is a good idea to offer users online access to the dissertations of the home university. However, libraries are also challenged to offer users easy access to electronic dissertations of other universities. Naturally, these can be accessed either via the homepages of the respective universities or via harvesting initiatives such as NDLTD or TheO⁷. But for most library users the library catalogue of the home university remains the initial starting point for literature searches. For this reason the ETH Library chose to integrate selected metadata from foreign electronic dissertations into its OPAC.

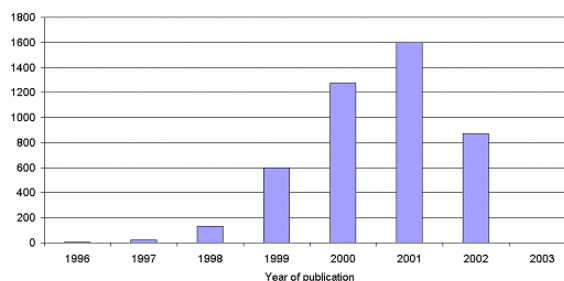
According to the collection policy of the ETH Library only dissertations in the field of science and technology are collected. Furthermore, dissertations need to have persistent URLs and offer full texts online (not only abstracts). To start with, the ETH Library decided to harvest metadata from German dissertations. These are archived by the universities and the Deutsche Bibliothek and recorded in the German National Bibliography. The

Deutsche Bibliothek guarantees persistent internet addresses and long term digital preservation.

Title lists of new electronic dissertations are extracted from the online version of the German National Bibliography on a monthly basis. Subject specialists work through the lists and mark which titles are of interest to the library. Cataloguing is done by copy-cataloguing only, however, full UDC classification is added by subject specialists. Thus over 4'500 references to German electronic dissertations have been added to the library catalogue NEBIS during the last two years.

Figure 5 shows the number of German electronic dissertations that could be included in the NEBIS library catalogue using the method of copy-cataloguing from the German National Bibliography. The low figure for 2002 suggests that not all electronic dissertations with publication year 2002 are yet available. By April 2003 only one dissertation from 2003 is available for copy-cataloguing as yet.

Figure 5: Records of German Electronic Dissertations included in Library Catalogue NEBIS



In order to catalogue electronic dissertations of other Swiss universities, cataloguing staff has to check the websites of all ten Swiss universities. This metadata is regularly copied into NEBIS. There are plans that in future the metadata of all Swiss electronic dissertations should be collected centrally in a national project. In comparison to the German situation there are only very few electronic dissertations available in Switzerland (apart from the ETH Zurich).

Ideally the ETH Library would also cover electronic dissertations of other countries, too. But it is clearly too much work to check all foreign universities for suitable documents. Alternatively the ETH Library could act as a Open Archive Initiative harvester. Even then the challenge remains how to integrate the data into the library catalogue. Furthermore, several checks with data from foreign universities show that URLs of electronic dissertations are very often anything but persistent!

A lot still remains to be done before we can offer users access to the vast collection of electronic dissertations through the library catalogue of the home university.

6 NDLTD: <http://www.ndltd.org/>, GetInfo: <http://www.getinfo-doc.de/>, E-Helvetica: <http://www.e-helvetica.admin.ch/>.

7 TheO: <http://www.iwi-iuk.org/dienste/TheO/>.