

ETH E-Collection: an Innovative Publication Platform for Scientists

Realisation of a Document Server for the ETH Zurich by the ETH Library

Ursula Jutzi-Müller

ETH Library
Swiss Federal Institute of Technology Zurich, Switzerland

Abstract. In January 2002 the ETH Library launched the new electronic publication platform ETH E-Collection. This publication channel gives scientists the ideal opportunity to make their publications available easily on the Internet. This collection was designed to bring together documents published outside traditional publishing houses, i.e. grey literature such as research reports, lecture notes, educational materials or newsletters. With more then 2000 documents within the collection (end of July 2002), the pilot phase is over and the new platform is already well known and very much appreciated within the scientific and learning community at the ETH. For the realisation of this document server the departments collection development and IT joined forces to collect documents, create metadata and find the suitable technical implementation. The whole realisation, including technical implementation, was done entirely by the ETH Library. Outside assistance was only required for the web design.

1 Introduction

The ETH E-Collection, a project of the Swiss Federal Institute of Technology (ETH) Zurich, grew out of a digitisation initiative for doctoral dissertations that was launched by the ETH Library in 1999. Within this very successful initiative doctoral dissertations of the ETH Zurich were made available online for world-wide use. The positive feedback from authors and the high usage of this collection encouraged ETH Library to set up an alternative publication platform for other

document types too. The expected readership includes the public interested in science and technology. The ETH E-Collection was chosen to be one of the main elements in the major strategic initiative ETH World (2.1), which represents the university's virtual communication and cooperation platform.

The "ETH Library" is the main library of the Swiss Federal Institute of Technology in Zurich. It is primarily dedicated to provide relevant information on all areas of science and technology as covered in teaching and research at ETH. Furthermore, it has a key role as the Swiss national centre for scientific and technological information.

With collections comprising over 5.6 million items, 6,500 current print journals and over 4,000 e-journals, the ETH Library is one of the largest sci-tech libraries in Europe. From this perspective it is obvious that the national scientific community is very interested in the library's activities and the performance of its services. To support both major roles, the management team of the library started the ETH E-Collection within the framework of the ETH World project to provide an innovative service for the members of the ETH Zurich (researchers, academics, students) and our customers throughout the world.

Apart from being a publication platform for the ETH community, the ETH E-Collection also supports learning and tuition at the ETH Zurich.

2 Milestones

The ETH E-Collection has been accessible for the public in the current design since January 2002.¹ Already in October 2000 the doctoral theses of the ETH Zurich were made available online for the public worldwide by the ETH Library. This pilot project was realised before the official start of the ETH World and became the foundation for the ETH E-Collection. It gave our employees the possibility to get important experience with different aspects regarding electronic publishing, such as formats, digitisation, archiving of electronic data and handling of new electronic materials in general.

In September 2001 the ETH E-Collection started with the first hundred documents in addition to the theses and activities for acquisitions were intensified. As a result of this acquisition effort we were able to make an official start to the collection for public use in January 2002. For this definite start, also a lot of effort was put into the creation of a new web design.

During the year 2002 many things have been realised. A workflow for these new document types within the library was settled, the websites were optimised, more and more documents were integrated and first trials for the integration of multi-media supplements were carried out successfully. As some of the learning

¹ ETH E-Collection: http://e-collection.ethbib.ethz.ch/index_e.html

materials are only for use within the ETH Zurich campus, an intranet solution for this material was also developed. Furthermore, to support our potential authors, the possibility of uploading documents directly to our server was implemented. This was necessary as some of the documents, e. g. material which includes maps or multi-media files, could quickly become too big to handle for the mail-servers and time-consuming roundabouts had to be carried out for the delivery of those documents.

As all the material is also accessible via the library's online catalogue, not only a few team members but all departments throughout the library were involved. Obtaining the support from all departments of the library was very important for the success of such a complex project.

All documents are available either via the library catalogue or via the website of the ETH E-Collection.

By the end of June the collection provided the following documents (Fig. 1):

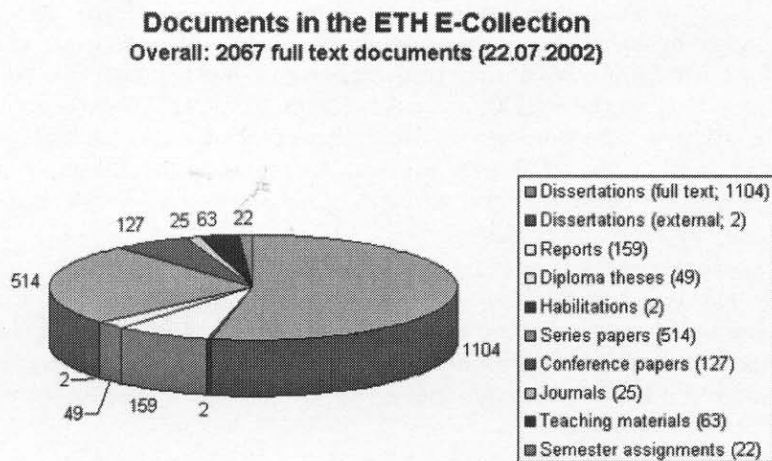


Fig. 1. Number of documents in the ETH E-Collection

3 Realisation

3.1 ETH World

The project ETH World can be described as follows: "ETH World is a strategic initiative to prepare ETH Zurich for the information age. Its objective is to create a

universal virtual communication and cooperation platform, supporting the activities of everyone working or studying at ETH. ETH World will augment the present physical locations, "Zentrum" and "Hnggerberg", by a virtual space, creating a third, virtual campus for ETH Zurich."²

Fortunately we were able to position the ETH E-Collection within the ETH World project. This aspect was very important for us not only in financial perspectives. ETH World gave us the opportunity to participate e.g. in their PR activities. This spring we were able to present the project at a lunch time session to the whole ETH community, which brought us into the news in the online newsletter ETH Life.³ With this great publicity success within the ETH community people are becoming acquainted with the E-Collection.

3.2 Acquisition

One of the main challenges for the project was and still is the acquisition of the documents from within the ETH community. As this has been one of the main reasons for the failure of similar projects we put much effort into this part of the project. Besides the well-known marketing activities like mailings (online and written), flyers etc. we were convinced that a different approach was required in the library world. In our case we decided to start an intensive telephone marketing campaign during which we tried to contact all of the members connected with teaching at the ETH Zurich personally.

Unfortunately there is no rule requiring scientists to deposit electronic documents at the library. Doctoral students are obliged to present the library with several printed copies, but there are no rules concerning the electronic version. Up to this date the university does not accept a doctoral thesis in electronic format only. However, PHD students are invited to give the library an electronic version of their thesis, if available.

If the PHD student does not have a suitable electronic document available, the library offers to digitise the printed copy.

As mentioned above, the lunch time session and the article in the ETH daily online paper, were a great help to make people aware of the new services. Whereas it is no problem acquiring dissertations, as generally all of our students are well aware of this service, which all of them rate excellent and therefore worth participating in, all efforts are necessary to persuade scientists to contribute to the collection.

² ETH World: <http://www.ethworld.ethz.ch/nw/ethworld/>

³ Article in ETH-Life:

<http://www.ethlife.ethz.ch/tages/show/0,1046,0-8-1734,00.html>

3.3 Technical Implementation

Web Realisation. Needless to say, that all new online services must have an attractive design which ensured easy access by the public. Navigation must be possible without further assistance through library staff.

For the design of the ETH E-Collection we chose an external partner to create a logo and the overall appearance of the collections site. The design should demonstrate both the close connection to the ETH Library, and the independence of the collection as a service on its own. We therefore tried to create a clear, plain platform, which could be recognised as having scientific content (Fig. 2).

As in the tradition of the ETH Zurich the main languages are German and English every web page exists in both languages and our customer has the possibility to switch the language on every level of the web site. The official languages of Switzerland, German, French and Italian are in this case not all supported but this bilingual realisation reflects the state of the art within the ETH Zurich.

The maintenance of the site is, of course, in the responsibility of the library, including all necessary updates and alterations.

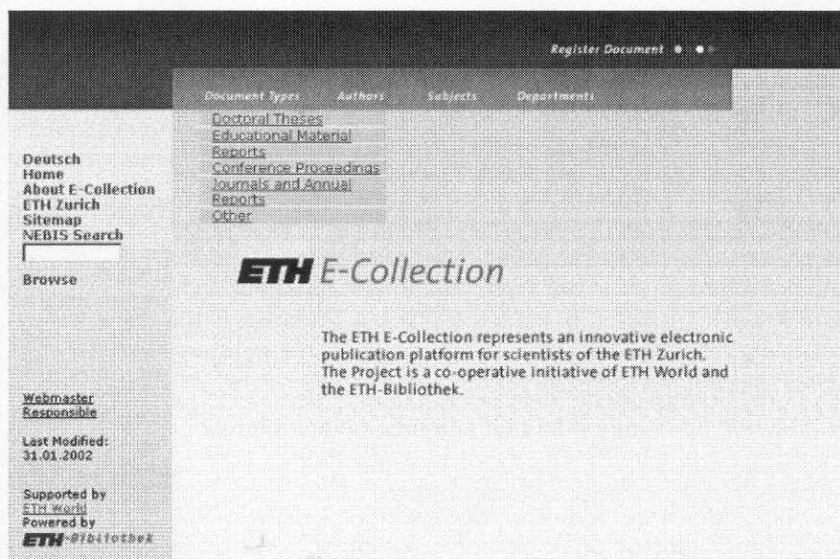


Fig. 2. Homepage of the ETH E-Collection with open layers for the document types

Workflow technical implementation. The collection development department and IT department developed a strategy to facilitate easy identification and presentation of all documents included in the collection (Fig. 3). Without great input from the IT department – especially from the programmers' side – the project could not have been realised within such a tight time-frame.

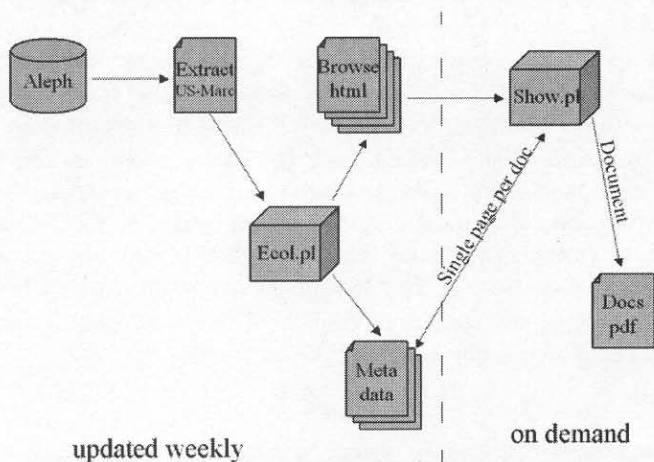


Fig. 3. Workflow technical implementation

In July 2002, E-Collection runs with Perl programs under O'Reilly WebSite Pro on a Windows NT fileserver and uses about 18 GB. Migration to Apache 2 on a new machine under Windows 2000 is planned for summer 2002.

As the library catalogue⁴ is seen as the main search tool for all library documents, regardless of their physical appearance, we tried to use the data in the catalogue to build up and maintain the web site. This procedure had already been chosen for our electronic resources such as databases or electronic journals.⁵ We attempt to avoid any redundant cataloguing and to keep the workflow as simple as possible.

Once a week the data are extracted from the library catalogue (USMARC). By means of a Perl program (ecol.pl) this data is used to extract metadata of these documents for further use and to generate web lists for browsing (Fig. 4). Browsing is possible via document type, author, subject (those relevant to the ETH) and department. The latter is, of course, not very useful for customers outside the ETH, but for our in-house customers this is a big advantage and also an incentive

⁴ NEBIS Library Catalogue: http://www.nebis.ch/index_e.html

⁵ ETH Library: Library Online: http://www.ethbib.ethz.ch/bibliothek_e.html

for them to submit their materials to the collection. Such data enables us to provide departmental lists of publications, if required.

The screenshot shows the ETH E-Collection website interface. At the top, there is a navigation bar with the logo 'ETH E-Collection' on the left and a link 'Register Document' on the right. Below the navigation bar, a breadcrumb trail reads 'Browse > Documents Types > Educational Materials'. The main content area is divided into two columns. The left column contains a sidebar with links: 'Deutsch', 'Home', 'About E-Collection', 'ETH Zurich', 'Sitemap', 'NEBIS Search' (with a search box), 'Browse', 'Webmaster Responsible', 'Last update: 16.07.2002', and 'Powered by ETH-Infotech'. The right column displays a list of documents under the heading 'Architecture, Built Environment and Environmental Design'. The list includes entries for Bonzi, Tamara; Faber, Michael H. (two entries); Girmscheid, Gerhard; Graf, Frank; Kerez, Christian; and Künzle, Otto. Each entry provides the title of the work, a link to 'Mutationen', and the institution and year of publication.

ETH E-Collection Register Document

Browse > Documents Types > Educational Materials

Deutsch
Home
About E-Collection
ETH Zurich
Sitemap

NEBIS Search

Browse

Webmaster Responsible

Last update:
16.07.2002

Powered by
ETH-Infotech

Architecture, Built Environment and Environmental Design

Bonzi, Tamara
Mutationen
[ETH, Eidgenössische Technische Hochschule Zurich, Departement Architektur], 2002

Faber, Michael H
Risk and safety in civil engineering [2002]
ETH, [2002]

Faber, Michael H
Risk and safety in civil engineering [2001]
ETH, 2001

Girmscheid, Gerhard
Sicherungsmaßnahmen mit Spritzbeton
Institut für Bauplanung und Baubetrieb, ETH Zurich, [2001]

Graf, Frank
Organismen im Lebensverbau: Voraussetzungen und Funktionen
WSL, [Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft], 2002

Kerez, Christian
Mutationen
[ETH, Eidgenössische Technische Hochschule Zurich, Departement Architektur], 2002

Künzle, Otto
Tragkonstruktionen I-IV, Aufgaben und Lösungen
ETHZ, Departement Architektur, Professur für Tragkonstruktionen, 1999-

Fig. 4. Browse list for educational material in architecture

As soon as a user selects an individual item from this browsing list, a second Perl program (show.pl) is called into action and, with the use of the metadata information and the html-templates, generates the web page for this document on the fly (Fig. 5).

A single page per document was crucial as we wanted a citeable URL. In addition we needed a way to present documents which consist of different parts in an efficient way to our customers. (e. g. dissertations will also provide a link to the abstracts; journals will of course have a link to every issue of the journal). Furthermore, additional links to the corresponding record in the library catalogue (NEBIS), to author's institute and to tools such as Acrobat Reader, provide users with additional help and information.

The 909 field in the USMARC record acts as the main identifier for the documents with regard to the E-Collection. An identification code for the institute is also entered in the 909 field. The Perl program (show.pl) matches these entries with a txt-file for the institutes which is updated regularly by our staff.

As the ETH E-Collection was also designed to support the teaching environment at the ETH Zurich, an intranet solution for all materials which are only for use in

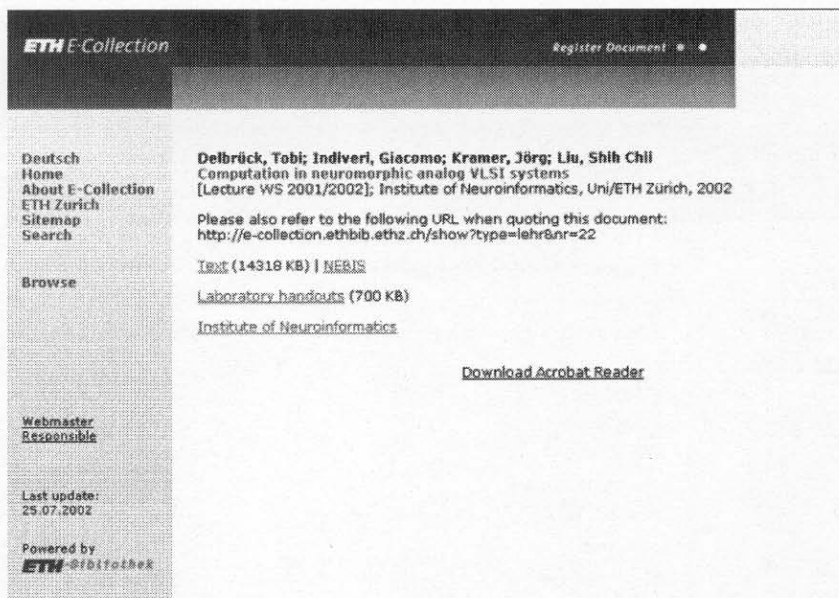


Fig. 5. Page created on the fly for a lecture note

the ETH Zurich, was developed. These documents are freely searchable, but via IP address verification (also controlled via the second show.pl program) the full text will only be displayed for users, who are within the IP-address-range of the ETH Zurich.

To make searching easier we provide a second search option besides the browse lists. As an interim solution we implemented a search field on the E-Collection site which submits the enquiry directly to the library catalogue and the search is automatically performed only on the ETH E-Collection documents. The first implementation using one of the well-known search engines had to be replaced by this NEBIS search because of absolutely intolerable indexing times.

4 Future

4.1 CMS

As a search within the full text of the documents would be of great use to our customers we have plans to either implement a search engine which is able to handle PDF documents properly or to integrate a type of CMS system. Besides the advantages of a full text search a CMS would make the exchange of metadata within the academic society easier. Moreover, this solution would make a connection to the Open Archives Initiative possible.

4.2 Co-operation

The first co-operative effort for the exchange of metadata was settled with Math Diss, a project of the Universities of Gttingen and Duisburg in Germany. For this project only the metadata for the mathematics dissertations are of interest.

A second partner is the project GetInfo of the Technical University Hannover (TIB).

5 Usage Statistics

To get reliable usage data our IT department put a lot of effort into analysing the log-files. The downloads for the full texts and the connection to the specific web pages were counted. Accesses of search engines were eliminated by program.

Usage statistics concerning online dissertations have been available since October 2000 (Fig. 6). This data is of great interest to the library as it can be compared with the check-out data of the print dissertations. For the counted downloads of full texts per month only 70 loans of print copies are counted for the same time frame.

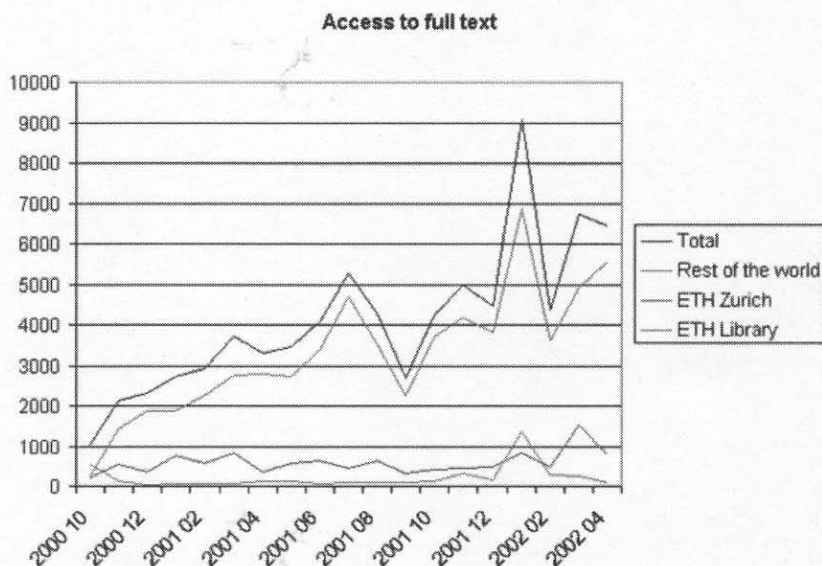


Fig. 6. Download of full text documents

6 Conclusion

The ETH E-Collection was chosen to be one of the main elements in the major strategic initiative ETH World, which represents the university's virtual communication and cooperation platform. Although the initial steps were taken outside ETH World, the integration of the ETH E-Collection in this innovative virtual space gives the library's project a good framework and a very interesting dimension.

The ETH Library collects electronic publications in a central system and ensures that they are catalogued according to international standards. This process of acquisition and registration of documents is currently being fully integrated into the library's organizational setup and processes.

The technical realisation of the project is now completed and we could look towards integrating CMS system to support our customers better and to improve the workflow.