

FROM CENTRAL ADMINISTRATION OF HARDCOPY BACHELOR- AND MASTER THESES TOWARDS A DECENTRALIZED ETD-SYSTEM WITH VALUE ADDED SERVICES

Rudi Baccarne

Librarian, Catholic School for Higher Education, Geel, Belgium

Keywords: Thesis administration, ETD-workflow, ETD-training, Repository Software, Metadata

ABSTRACT

Thesis administration has been the responsibility of the Central Library of the *Katholieke Hogeschool Kempen* since 1980. A production process has been developed to publish hardcopy theses. This primarily manual workflow has been computerized over the years and since 1997 students have been obliged to submit an electronic thesis in addition to the hard copy. This automation process has been constantly evolving. At first, all metadata and links to full text documents were imported in the library catalogue. The next step was the development of an ETD repository for all Flemish institutes of higher education. A newly built software called 'DoKS' automates the whole process of thesis administration. (upload titles, metadata and full text, export functions to the library catalogue, external publisher, etc.) and is built in such a way that theses from different institutes are decentrally stored yet at the same time available through one interface. Standards and protocols were taken into account to guarantee interoperability (Open Archives Initiative, Dublin Core, XML, etc.). <http://doks.khk.be/eindwerk> is an example of a local DoKS repository. A central DoKS-server can be consulted at <http://www.doks.be>. At the moment the DoKS system is being optimized to generate value added services in response to the needs of private industry and academic staff. An alert and portfolio component has been integrated in order to make the ETD-system usable as a recruitment database. The ETD repositories are being used by the Flemish Chamber of Engineers as an integral part of a project aimed at stimulating entrepreneurship. A DoKS built-in scripting engine provides different automation opportunities (quality control, title and abstract listings, publishing, etc.). The paper will focus on the advantages of a central administration of theses and on value added services generated by an ETD-system.

1. INTRODUCTION

The task of thesis administration has been the responsibility of the library at the *Katholieke Hogeschool Kempen* since 1980. This means that in order to graduate every student has to report to the library's main desk with a boxed copy of his or her thesis. The box containing the thesis is then collected by a printing company, which reproduces the title page and cover, and binds the thesis. Approximately 800 theses a year are handled in this manner. Since 1997 students have also been requested to submit an electronic version in addition to the hardcopy version. Although this is not an obligation, more than 95% of the students now submit ETDs. We believe that this success is related to the fact that over the years the library has come to be regarded as an authority when it comes to thesis administration. The library has established the rules and procedures and taken the initiative for the automatization of thesis administration. The development of an ETD-system is perceived as a logical result by professors, students and the board of directors. In the long term the system will probably

make it easier to replace the workflow for hardcopy items with a workflow for ETDs. To date the number of hardcopy items for the library has been reduced from two to one and the ETD is beginning to replace the traditional file copy. The belief that libraries have to be involved to produce successful repositories is also stressed in the context of broader Open Access initiatives (Pepe *et al.* 2005) and the UNESCO ETD guide (McMillan 2001).

2. CENTRAL ADMINISTRATION OF THESES AT THE LIBRARY

2.1. Workflow Schema

The core of the workflow comprises a timing schedule. For each department specific dates are agreed about thesis instruction and collection of metadata and full text. These dates apply to both electronic and hardcopy versions. Figure 1 below shows the whole workflow from A to Z.

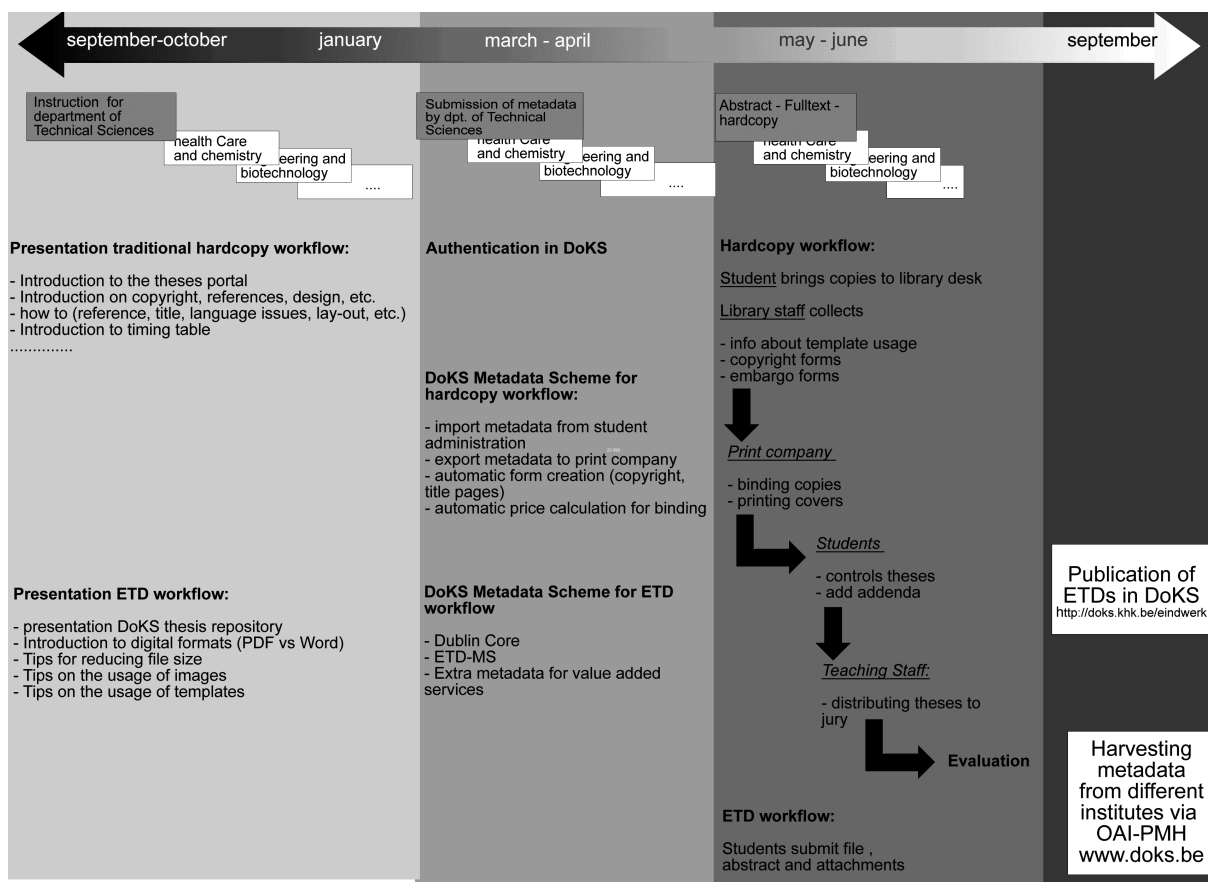


Figure 1: Thesis administration at the *Katholieke Hogeschool Kempen*

2.2. Some Key Factors

Compared to similar institutes in Belgium, few problems have been experienced in convincing students to submit ETDs. In addition to the library's authoritative reputation as regards thesis submission, the following factors play a crucial role.

a. Instructions for students

It is obligatory for every graduating student to follow a short course at the beginning of the academic year on certain aspects of the general workflow for thesis administration. Part of this course focuses on the submission of the ETD. Students are taught how copyright laws function and what they may or may not insert in their work, how to cite references correctly, the purpose of the different copyright forms, which templates to use and so on. The library

supplies a voluntary Word Template to author the thesis. The main aim is to achieve a common structure for theses and make students aware of the advantages of structured authoring. Although voluntary the template is used by a majority of the students. The supply of the template also originated from the idea of a workflow to convert Word documents to XML. Experiments have been set up, but as other ETD programs indicate these need extra employees, instructions and workshops to achieve reasonable results (Wendland 2003) (Dobratz 2001). The bulk of theses handled each year, the heterogeneity of the student population and staff limitations, make the organisation of such a workflow impossible. Students are free to choose a file format although the use of PDF is strongly recommended. Conversion facilities are available at the Institute.

b. Personal contact at the library desk

Given that every student has contact with the library desk when bringing in copies of their thesis it is very easy to distribute and collect copyright and embargo forms. The library also supplies support by means of scan facilities and assistance for conversion and compression of files. Furthermore, the feedback received at the library is an important instrument to optimize the workflow.

c. Seamless integration of hardcopy and ETD workflow.

All metadata about theses is collected electronically and serves both the hardcopy and the ETD workflow. A part of the metadata (author, department, degree title, address, etc.) is automatically imported from files received by the library from the Institute's general administration department whilst another part (title, abstract, language, volumes, contact details, number of desired copies, instructor, trainee post, trainee supervisor) is submitted by the students via the DoKS repository software. All gathered metadata is saved in XML and serves different purposes.

d. The thesis portal

All information about the thesis administration is available through the thesis portal. The portal also contains a step-by-step workflow to submit metadata and full text and automatically generates the necessary copyright forms. Students are obliged to use this site to submit metadata. These data are exported to a printing company, which reproduces title pages and binds the copies.

3. AUTOMATION OF WORKFLOW WITH DOKS

3.1. Concept

The DoKS software has been developed within the framework of a project funded by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT-Vlaanderen, www.iwt.be). The main goal of the project was to make theses from engineering departments of different institutes for higher education in Flanders available via the internet. To achieve this goal various meetings were organised with delegates from these institutes. As a result agreement has been reached about the use of specific metadata sets (Dublin Core, ETD-MS) and content classification. The latter is based on the classification scheme of the Flemish Research Database IWETO ("Ministry" n.d.). At present seven institutes use DoKS and more than 2000 ETDs are available. Other institutes for higher education are planning to install and use DoKS in the near future. The scope of the project is no longer restricted to engineering departments and theses from social, economic, health and chemistry departments, to name but a few, are also now available.

DoKS as a decentralized ETD-system

One of the main reasons for developing a new software instead of using an existing one was the need for a system that could be highly customized by users to answer local needs and needs specific for Flemish institutes of higher education. Furthermore every institute had to be capable of collecting extra metadata in addition to the commonly agreed use of Dublin Core and ETD-MS. At the time the DoKS project started (2002) there was little indication that existing software was being developed to fulfil this need. Examples of extra metadata are the desired number of copies, volumes, subtitles, trainee post, student's address and so on. As discussed in greater detail in the chapter on the architecture of DoKS, the choice by participating institutes of which database system and operating system to use has also become an issue. More general factors that influence the choice or development of repository software are the materials to be stored, long-term preservation and costs amongst others (Prudlo 2005).

Central OAI interface

By means of the Open Archives Protocol for Metadata Harvesting, different local DoKS repositories are made available through one interface. Until now our OAI harvester uses only the Dublin Core metadata, although participating institutes also collect metadata conform ETD-MS. Initial agreements on the field thesis.degree (level of education) will possibly have to be changed in accordance with amendments to the Bologna declaration. (Linde, P., Björklund, C., Svensson, A. 2005). We use the PKP OAI Harvester ("PKP Open Archives Harvester" n.d.) software to collect the metadata and are planning to adapt it to harvest local and ETD-MS metadata. All metadata is captured in XML and as a result it can be harvested in accordance with the OAI-PMH. (Van De Sompel, H., Young, Jeffrey A., Hickey, Thomas B. 2003)

3.2. Architecture

To meet the requirement that Doks should be installed and used at low cost by participating institutes, the software is exclusively built with open source Java Frameworks. The minimum costs involved for working with DoKS are the purchase of a server and in addition some staff expenses to maintain the system. In most cases however staff expenses can be compensated because DoKS automates different aspects of traditional theses administration. Time consuming activities such as cataloguing and generating title and abstract lists can be extensively automated.

The choice for Java makes DoKS accessible in different environments (UNIX, windows). Furthermore it is database independent and different authentication systems (LDAP, dedicated password) can be used. These are important issues, which have been raised at meetings organised for participating institutes. A detailed description of the software and the different frameworks involved (Struts, Hibernate, AspectWerkz, OAICAT, Apache Lucene) are available at the project website (<http://doks.khk.be>). At the moment preparations are underway to make the software available as open source project.

4. VALUE ADDED SERVICES

Private industry partners have been involved from the beginning and have supported the ETD project. Some are interested in the scholarly knowledge that will be made available, while others see it as a potential recruitment tool. Thanks to the use of some state-of-the-art Java frameworks we have been able to attract the attention of software companies as well. A Business Plan was developed to investigate whether DoKS can be self-maintaining in the long term by means of value added services for the industry.

4.1. ETDs for Recruitment Purposes

A thesis is a good indicator of the capabilities of a particular student. The layout of the thesis, the style of writing, the quality of the research, the consistency of the work are all elements that reveal a student's competencies. At present we are adding (or planning to add) extra data typical of a curriculum vitae or portfolio. Students will be held responsible for the maintenance of this data and after graduating students will have the right to access, edit or remove this data. A database combining these data with ETD records is unique and marketable as a recruitment tool. This proved a potential motivation for private partners to support the development of our ETD program. It is proposed that access to the combined data, will be based on a paid subscription. After subscription the user can authenticate with userID and password. A market survey, carried out as part of the Business Plan, indicated that companies are prepared to pay for such a subscription.

As a result of the bulk of thesis records added each year (appr. 800 for the *Katholieke Hogeschool Kempen* alone), companies can submit a profile and via this profile be alerted to new items. Given that privacy information about individuals is stored in the database, the scope of it goes beyond scholarship. Some juridical consequences must be taken into consideration. In Belgium such databases must be registered with the Commission for Protection of Privacy of the Federal Government (www.privacy.fgov.be).

4.2. Stimulating Entrepreneurship

The Flemish Chamber of Engineers is developing an award program based on the DoKS repositories to stimulate entrepreneurship. The aim of this project is to filter theses with a high commercial or innovative character, especially those that have the potential to develop into enterprises. The idea was born from the fact that the number of new businesses of an innovative nature started each year in Belgium is very low compared with other countries (De Clercq *et al* 2003). The project was launched at the *Katholieke Hogeschool Kempen* during the academic year 2004-2005 and is in the test stage. If successful it might prove an important factor in convincing new institutions to start an ETD program.

4.3. Services via the Built-in Scripting Engine

DoKS supplies a java and JavaScript-like scripting engine (Beanshell) for task automation, complex work flows, specialized import/export, etc. Administrators who are familiar with similar scripting languages will find it relatively easy to make new scripts. Standard the DoKS software ships with some import and export scripts. The use of this scripting engine has led to the automation of formerly manual processes such as MARC-export, collection of abstracts and titles, publishing and so on.

MARC export

Thesis records, including abstracts and links to the full text, have been imported in the library catalogue since 2001. Before DoKS, this process was rather complicated and needed conversions in various steps between different file formats. As a consequence a requirement of DoKS was that it contained an automated and simple workflow for the export of the XML-ETD records to MARC. The exported file had to be suitable for batch import into the library system. This conversion process is now handled by a script which generates a text file that can be converted directly to MARC with a standard MARC-editor. Alternatively, there are some open source java tools for the conversion from XML to MARC that might be worth considering ("Marc4J Project Home" n.d.).

Collection of abstracts and titles

Another service the library provides is the creation of abstract and title listings. Staff from different departments require listing of abstracts and titles for different purposes. Abstract listings are printed and bound in different formats and sent to thesis reviewers, trainee posts

and so on. The listings are generated automatically by executing a script based on variables such as department codes and year. These scripts are executed in a few seconds and the resulting lists can be delivered in text format or HTML amongst others.

Publishing

A specific export script is written for the publishing house responsible for the binding and the covers of the theses. For this purpose, students have to submit specific metadata including the number of copies required, volumes, subtitles and so on. The exported file also includes the price the students will have to pay for the published work.

5. CONCLUSION

Since 2002 an ETD portal has been in existence for theses from Flemish institutes for higher education. Due to the extensive experience in electronic theses and dissertations held by the central library at the *Katholieke Hogeschool Kempen* the latter has occupied the privileged position of designing a broader collaborative ETD portal. To achieve this, the repository software DoKS was designed with broader worldwide initiatives in mind and as a means of countering specific problems for Flemish institutes of higher education.

The success of an ETD-system depends certainly on the technology involved but perhaps more importantly on clear procedures regarding copyright, embargos and so on. In addition an environment - not just technological -, has to be created in a way that provokes as little resistance as possible towards the objective. Measures to breakdown barriers must be taken where different partners are involved. In the context of our project an important requirement was that institutes could use and install the software free of charge and that there were as few technological restrictions as possible. Another means of promoting co-operation was the creation of value added services towards students, teaching staff and other partners involved. The advantages experienced by using the system must motivate these users to submit ETDs and use DoKS. Finally experience and knowledge must be shared with other institutes in order to motivate them to participate in the project. At the moment investigations into whether services for private industry can make the system self maintaining are underway.

6. REFERENCES

Dobratz, S. (2001) Preparing for Conversion to SGML/XML, In: Moxley, J.M., Fox, E., Masiello, D., eds. *The Guide for Electronic Theses and Dissertations*, UNESCO. Retrieved June 14, 2005 from <http://etdguide.org>

De Clercq, D., Manigart, S., Clarysse, B., Crijns, H., De Sutter, M., Verzele, F. (2003). Global Entrepreneurship Monitor: Executive report for Belgium and Wallonia, Vlerick Leuven Gent Management School, 2003, 95 p. Retrieved 16 June, 2005 from <http://www.gemconsortium.org/document.asp?id=265>

Linde, P., Björklund, C., Svensson, A., (2005). Putting a National Portal for Undergraduate Theses into Production. In: Dobrev, M (ed.), Engelen, J. (ed.), *From Author to Reader: challenges for the Digital Content Chain, proceedings of the 2005 ELPUB Conference*. (pp. 91-96) *Katholieke Universiteit Leuven 8-9-10 June 2005*, Leuven: Peeters Publishing.

Marc4j Project Home. Retrieved June 16, 2005 from <http://marc4j.tigris.org/>

McMillan, G.(2001). Role of the library and Archives, In: Moxley, J.M., Fox, E., Masiello, D., eds. *The Guide for Electronic Theses and Dissertations*, UNESCO. Retrieved June 14, 2005 from <http://etdguide.org>

Ministry of Flanders, science and innovation, *Flemisch Research Database Iweto*, Retrieved June 14, 2005 from http://www2.vlaanderen.be/IWETO/DOCS/thema_pag.html

Pepe, A., Baron, T., Gracco, M., Le Meur, J.-Y., Robinson, N., Simko, T., et al., (2005) CERN Document Server Software: the integrated digital library. In: Dobrev, M (ed.), Engelen, J. (ed.), *From Author to Reader: challenges for the Digital Content Chain, proceedings of the 2005 ELPUB Conference*. (pp. 297-302) *Katholieke Universiteit Leuven 8-9-10 June 2005*, Leuven: Peeters Publishing.

PKP Open Archives Harvester. (n.d.) Retrieved June 16, 2005, from <http://www.pkp.ubc.ca/pkp-harvester/>

Prudlo, Marion, (April 2005) E-Archiving: An Overview of Some Repository Management Software Tools. *Ariadne*, issue 43. Retrieved June, 14, 2005 from <http://www.ariadne.ac.uk/issue43/prudlo/>

Van De Sompel, H., Young, Jeffrey A., Hickey, Thomas B. (2003). Using the OAI-PMH...Differently. *Dlib magazine*, July/August 2003, volume 9 nr. 7/8. Retrieved June 14, 2005 from <http://www.dlib.org/dlib/july03/young/07young.html>

Wendland, B., (2003) Workflow for a Standardized Treatment of Templates: history and present state. Retrieved June 15, 2005, from Electronic Publishing Group of CMS and University Library of Humboldt University Berlin Web Site: http://www.dissonline.de/tagungen/etd_2003/etd_workflow_hu.pdf