A Bachelor and Master Theses Portal: Specific Needs and Business Opportunities for the DoKS Repository Tool

Rudi Baccarne

Central Library, Katholieke Hogeschool Kempen, Geel, Belgium
e-mail: rudi.baccarne@khk.be

Abstract

A few years ago a portal for bachelor and master theses from Flemish university colleges was established by means of the open source repository software DoKS. At present approximately 3500 theses from Flemish university colleges are available online. The growing use of the portal has led to a new communication stream that requires supervision and maintenance. Social software components amongst others are or will be integrated in the portal to give users a platform to perform tasks such as communicate, annotate and advertise. Although different local DoKS repositories and the concept of the DoKS application are similar to repositories and tools within the scientific community, the scope and the aim of a theses repository for university colleges are different. The main part of the database consists of applied research and the majority theses comprise trainee reports. Thus, in addition to students and instructors, the portal is attractive to key players in industry, non-profit institutions and private users with a particular interest in a theses subject. This paper examines the different opportunities and specific needs of a bachelor and master theses portal, illustrated by real life examples. Social software components can breath new life into former static text documents. Users can add comments, create blogs, add tables, illustrations and suchlike. Content sensitive advertisements enhance the content and usage of our theses records and create revenues that can be used to make new improvements. In addition we will discuss the need for new and strict procedures with regard to content control, copyright issues and embargos when a bulk collection of industry related theses are published online.

Keywords: print on demand; content sensitive advertising; Electronic Theses and Dissertations (ETD); social software

1 Introduction

In 2003 the library of the Katholieke Hogeschool Kempen (KHK) launched a portal [1] for electronic theses and curricula vitae of graduating students at Flemish university colleges. In order to create this portal a new software DoKS (Document and Knowledge Sharing) was built. The project is funded by the Institute for the Promotion of Innovation by Science and Technology in Flanders [2], private industry partners and non-profit organizations. One of the main reasons for developing a new software was the need for a system that could be highly customized by users and tailored to needs specific to Flemish university colleges. The need for a flexible way to add local metadata (awards, trainee posts, credit points, etc.) in addition to commonly agreed sets such as Dublin Core or ETD-MS was in particular a high priority for the different colleges that were interviewed during preparatory meetings.

The DoKS portals from different colleges are decentrally stored yet at the same time available through one interface via the OAI-PMH protocol. Apart from searching harvested metadata via the central OAI-harvester, the user can search the full text of Electronic theses and dissertations (ETDs) from different institutes by means of the Google custom search engine [3], the latter allowing specification of the websites to be searched.

The KHK portal [4] receives 1500 daily visitors and offers an almost daily feedback and thus provides an indication of its usefulness for the labour market, graduates and the industry amongst others. Although the concept is similar to ETD-repositories and tools from the scientific community, a theses repository of a university college has other target users in mind. From the start private industry partners and other organizations supported the development of the DoKS theses portal, their reasons for supporting the DoKS project varying in accordance with their core business or particular interest. (IT, recruitment, valorisation of knowledge, screening of publications on business related content, electronic publishing etc.). The business plan carried out in the framework of the DoKS project and feedback from users raised new ideas and commercial opportunities for consideration. In addition to the publication of ETDs, several new and sometimes secured services (curricula vitae, ratings, etc.) were added to the free service of rendering theses from Flemish university colleges available

worldwide. The DoKS software is available by means of an open source license at Sourgeforge (http://sourceforge.net/projects/doksproject). Reports and manuals are available via a wiki (http://doks.khk.be/wiki/) and the project website (http://doks.khk.be).

2 Specific Needs and Services for University Colleges

At international level, the focus of ETD-projects is on research theses from academic institutions. The main target audience for electronic theses projects is the research community. By making research theses more broadly available by means of open online repositories, researchers and research become more visible and as a consequence more widely cited. Advantages and value added services are generated for the researcher, his work and his place of work. [5, 6] Although university colleges are less focused on scientific research they show a growing interest in publishing electronic theses. From the beginning the DoKS repository tool was developed in accordance with the guidelines of the ETD community As a result the portal is interoperable and combinable to a larger extent. On the other hand, the increasing use of the site and the feedback gained has shown that a theses-portal with mainly bachelor and master theses has different needs and offers other opportunities to exploit.

2.1 Curricula vitae

In the framework of the DoKS project a business plan was carried out. Part of the business plan was a quantitative and a qualitative research of different target users of an ETD portal. One important conclusion drawn from the business plan and based on the findings of the user surveys was that there seemed to be considerable interest from the private sector in the use of the portal as a recruitment tool. As a consequence curricula vitae can be filed in a standardized way together with the ETD records. This renders the portal a simple and cheap alternative as a starting point in the search for new employees. The force of this system lies in the accuracy of the data it contains and specific search facilities to retrieve this data (see Figure 1). In addition it is possible to raise an alert when new CVs matching specific criteria are added. Entries to the system are made by students and/or institutes and the system is therefore more complete and unique than any other built *ad hoc* and by a third party. All companies interviewed were prepared to pay for such a service by means of a subscription or registration service.

Furthermore the business plan created for the DoKS project outlined the opportunities that could render the portal self maintaining. Once optimized and well positioned on the market the CV module should create enough revenues for the employment of a 50% employee for maintenance and administration of the system. At the moment different steps are being taken to convert this potential into reality. Different approaches to commercialize the CV module include partnerships, pay on demand, subscriptions and sponsorship to name but a few. It is clear that the graduating students also benefit from the CV module. The DoKS system automatically creates a CV for all graduating students based on data from student administration files. The student can complete this by adding extra data. The result is a CV that can be converted to a Europass CV by one mouse click. Statistics on the number of updated CVs confirm the students' interest. At the moment 1578 CV records are available belonging to KHK students who have graduated over the last two academic years. About 25 % of CV records were completed by students one month after the automatic creation of their CV and more than 70 % are updated before the end of an academic year. We believe that this enthusiasm is related tot the fact that the CVs are already formatted, half filled out and exportable to print and other formats (see Figure 1). In other words, the student does not have to make much effort to produce an attractive and easy-to-maintain CV. The idea of automatically generating many of the relevant metadata behind the scenes to avoid filling out lengthy electronic forms is relevant for different web environments. In the field of educational multimedia, for example, this is strongly expressed by Erik Duval in the slogan 'electronic forms must die' [7]. While going through the self-archiving wizard for their electronic thesis students must choose to what extent their CV will be available to employers. A majority choose the option 'CV fully available' thus allowing employers to contact the student directly. 'CV available without personal data (contact through DoKS website)' and 'CV not available' are two other options. Graduated students can access and update their CVs via a dedicated account for a limited period of two years.

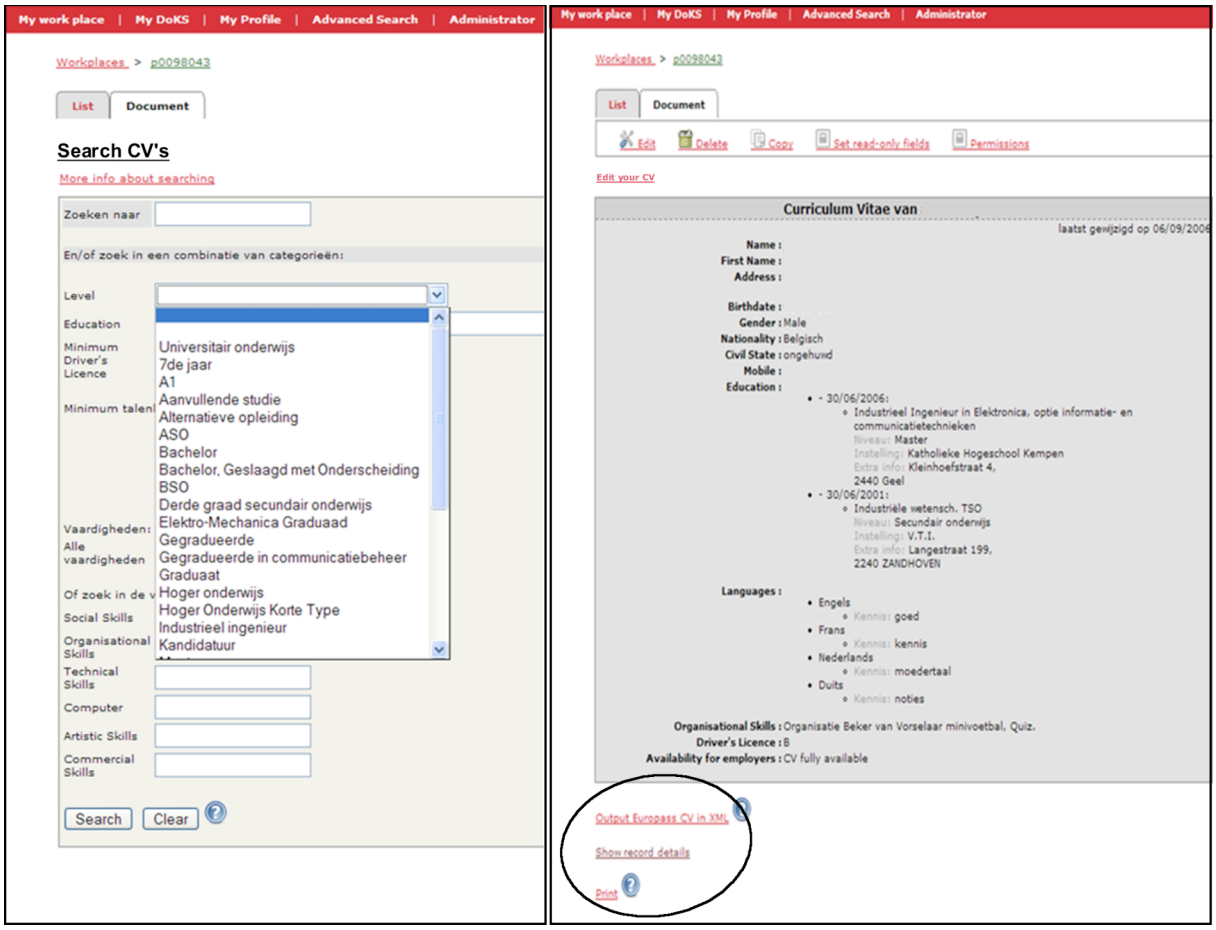

Figure 1: Search interface for CVs and basic example CV

In addition to the commercialization of the CV module, a good working DoKS portal can be exploited in several other ways. At the KHK the theses portal is linked to the Google AdSense program and creates revenues that are sufficient for replacement of hardware, new improvements, etc. Several other opportunities and partnerships with industry partners or non-profit organizations emerge once the portal is known to the different stakeholders. At the KHK this has led to partnerships with innovation agents, non profit partners and private industry partners that have a variety of reasons for supporting the maintenance of the portal [8, 9].

2.2 ETD-MS

ETD-MS is an interoperability metadata standard for electronic theses and dissertations [10]. The standard adds one element to the Dublin Core metadata elements, namely thesis.degree. This element has 4 qualifiers: thesis.degree.name, thesis.degree.level, thesis.degree.discipline, and thesis.degree.grantor. This standard is a result of the work established by the Networked Digital Library of Theses and Dissertations (NDLTD) which tries to coordinate the different worldwide initiatives. The aim of the standard is interoperability and is described as such on the NDLTD website [11] 'i.e., to make it possible to share information about ETDs. This will allow us to improve existing federated searches, create union databases, and provide greater consistency for researchers searching for theses and dissertations at different institutions.

To integrate bachelor and master theses in ETD union catalogues and repositories for scholarly communication we believe the level of education must be transparent and clearly distinguishable. This will help the end user to place the work in his context so he can judge it appropriately. The PKP-harvester software [12] we use for harvesting metadata from different local DoKS repositories did not however support ETD-MS. Therefore we recently created an ETD-MS plug-in [13] for this harvester. By using the plug-in users can perform searches on degrees and level of education to find graduates, their profiles and their learning outcomes.

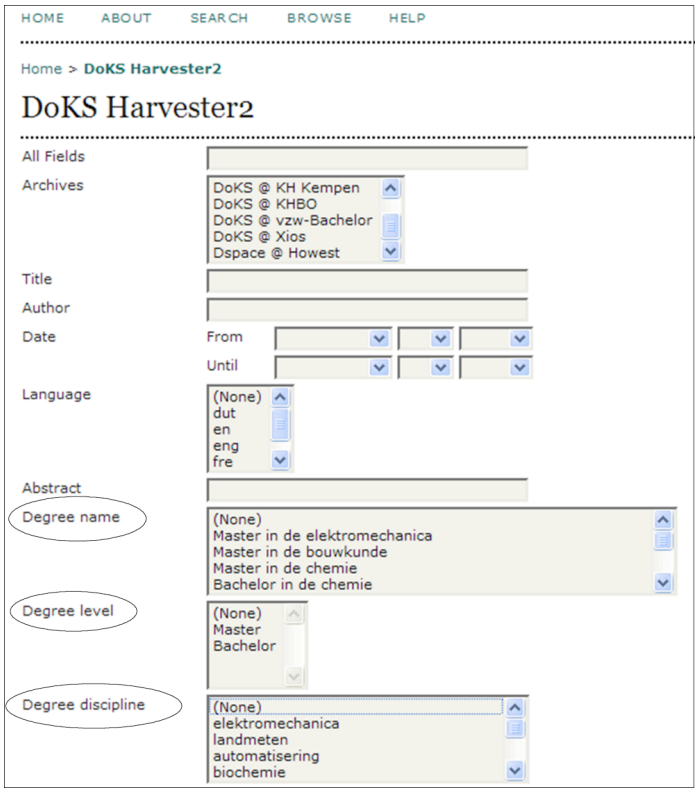

Figure 2: ETD-MS plug-in for PKPHarvester2

When other ETD programs consistently use the same standards, users can search records in a similar manner across records from different countries. The end user has an immediate knowledge of the type of the work, the level of the work, the educational program in which it was produced, the related degree and so on without necessarily knowing the language in which it was written.

2.3 Content (applied research, less academic, …)

At the KHK and by extension at most similar university colleges in Flanders nearly all theses comprise reports on work a student has done at a trainee post. As a consequence a thesis might contain confidential information. The industry partner where the trainee is placed has the option of requesting an embargo by means of strict procedures and dedicated forms. Although we expected to see a drastically increasing *a priori* demand for embargos once we started to publish ETDs, this seems not to be the case. Nevertheless the student must inform the trainee post about the online publication a clear demand for new embargos is seen once a thesis is online for a period of time. In some cases an embargo is requested because confidential information is published, but there seems to be several other reasons an industry partner does not want to see a trainee report published. This is often related to the high search engine ranking of our theses records - DoKS theses records are ranked higher than the web pages of the trainee posts - , old or false information on products is still available on the web, etc. In the future there are plans to give users a communication platform on which annotations can be made. In this way it is hoped that embargos can be avoided where the need for them goes beyond the publication of confidential information. On the other hand we see in literature [14] and also in practice a shift towards more transparency in domains (pharmaceutical industry, innovative IT companies) that were at first more resistant to the online publication of research data and material.

The power of the portal to serve the needs of innovation agents and intermediary organizations has resulted in the following collaborations:

Flemish Chamber of Engineers (VIK)
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 stemmed from the fact that the annual number of new start-up businesses of an innovative nature in Belgium is very low compared with other countries [15]. Furthermore over the years the Chamber has kept files of new businesses that emerged from the basis of an innovative idea in a thesis;

'Innovatiecentrum West-Vlaanderen' (West Flanders regional innovation centre)
A study [16] carried out by the regional innovation centre of West Flanders pointed out that of all theses established in the context of a trainee post only a low percentage resulted in an economic surplus value for the firms involved. An analysis of the study however showed that the economic valorisation could be increased by taking measures such as recruitment of the student after graduation or extra guidance by the college. To achieve this the regional innovation centre allocates awards for students and valorisation budgets for the firms. The innovation centre urges University colleges from the region to set up a DoKS portal in order to improve and accelerate selection procedures for theses that would be considered for a valorisation trajectory;

Indiegroup
Indiegroup is an organization that develops software for the innovation market. Integrating innovative content from the theses of university colleges can create a surplus for their software 'Cognistreamer'. Cognistreamer is a platform for open innovation concepts. By means of RSS and XML crosswalks innovative theses projects from DoKS could be integrated and selectively disseminated via Cognistreamer to organizations that are working on related subjects.

2.4 Statistics

2.4.1 Daily Visitors

The statistics in Figure 3 are based on figures from Google Analytics and cover the last full 12 months the KHK DoKS portal was online. The extremely sudden peak on the 11th of January has a logical explanation. At that time it was noted that a majority of visitors downloaded the full text of a DoKS thesis directly via a Google result list bypassing the DoKS website. For several reasons we have now decided to use a URL rewrite mechanism so that users are always transferred to a DoKS thesis record from where they can download the full text of the document. First of all the figures for the use of our portal were seriously underestimated. First and foremost, users were downloading documents from the site without knowing they were reading a thesis document from a bachelor or master student at a Flemish university college.

In addition it is clear that there is stable use of the website which at the moment has an average of 1500 daily visitors. The trend is downwards during the weekend and holidays and use increases use during the periods the students are working on a thesis and need the portal to submit data and full text. The statistics from Figure 6 indicate that the use of the site is strongly related to the revenues created by Google Ads with the same steep increase from January 2007 onwards.

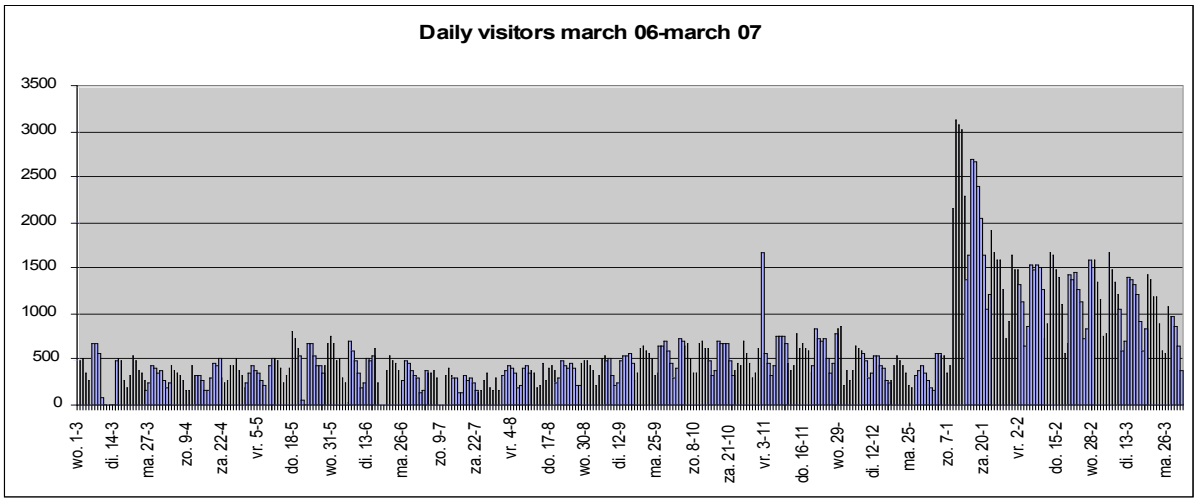

Figure 3: Daily visitors DoKS@KHK

2.4.2 Downloads

As shown in the graph below showing the number of theses downloads from the KHK-portal a 'long tail' curve emerges. This illustrates the wide and varied interest in theses content. The usage shown by the curve seems to be typical for e-business websites and indicates new economic mechanisms that are related to the internet. Theses that perhaps never came to light when they were stored physically at the library are, once online, consulted more than the most borrowed hardcopy theses from that same library. These new models and mechanisms are described by Chris Anderson with regard to e-business sites from the amusement industry (Amazon, Itunes, etc.) [17].

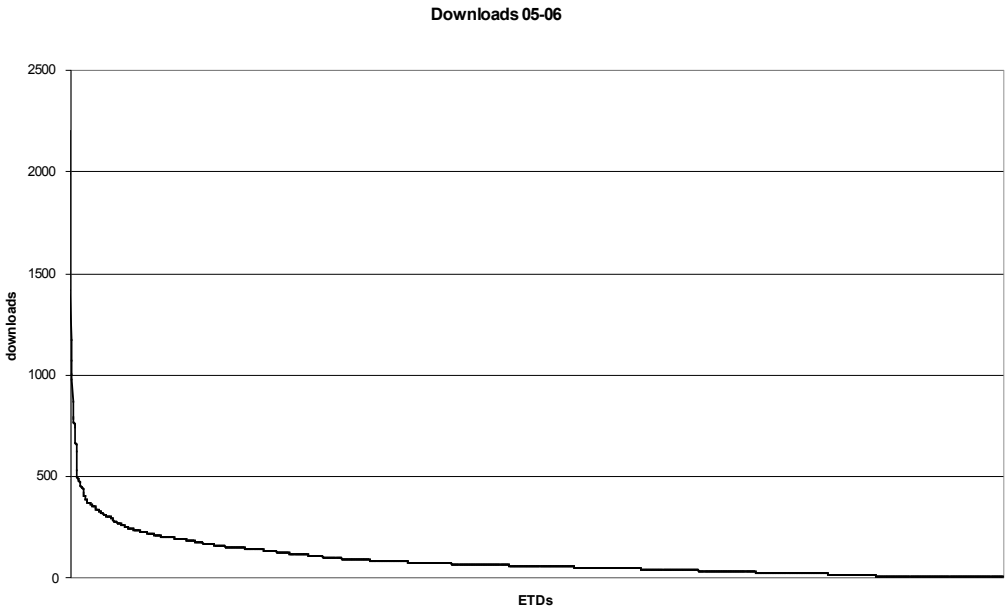

Figure 4: KHK-ETD downloads 05-06

2.5 User Feedback (categories)

When filled with a significant amount of content the DoKS repositories receive a high search engine ranking. As a result the number of daily visitors is significant. Among a variety of users, the feedback received in Flanders introduced several new business opportunities and interests. The feedback received can be categorized as follows:

- Job offers and offers for trainee posts
- Knowledge sharing
    - Collaboration proposals
    - Questions on thesis subject
    - Demand for annotations. Users want to comment on the content of a thesis and students want to add new views, opinions, corrections, etc.
- Editorial boards of journals
- Embargo requests from industry partners of the KHK
- Hardcopy requests (see section 3.3.5)
- Reporting on violations of the law with regards to:
    - professional confidentiality
    - copyright
    - privacy

2.6 Social Networking and Business Opportunities

The use and feedback on the portal clearly indicated the need to add social networking tools. Plans are being made for the future to integrate the features of the KNOSOS [18] platform in DoKS. Users who want to add, blog, annotate or tag to name but a few will be allocated to the collaborative working space provided by

KNOSOS. In preparation of a structured approach to overcome different needs, the first experiments have been set up. The following paragraphs describe the way in which we have already addressed some user demands.

2.6.1 A New Splash Page

Students are nowadays familiar with new technologies (Internet, multimedia, publishing, web 2.0, etc.) but are not supported in the use of them in a traditional hardcopy print environment. By following new internet trends, DoKS is able to keep track of the way young undergraduate students use the internet. We believe this is a necessary condition to conserve the enthusiasm of our most important supplier of information, the students themselves. As a result the record splash page (Figure 5) has been drastically changed in favour of a more user-friendly interface.

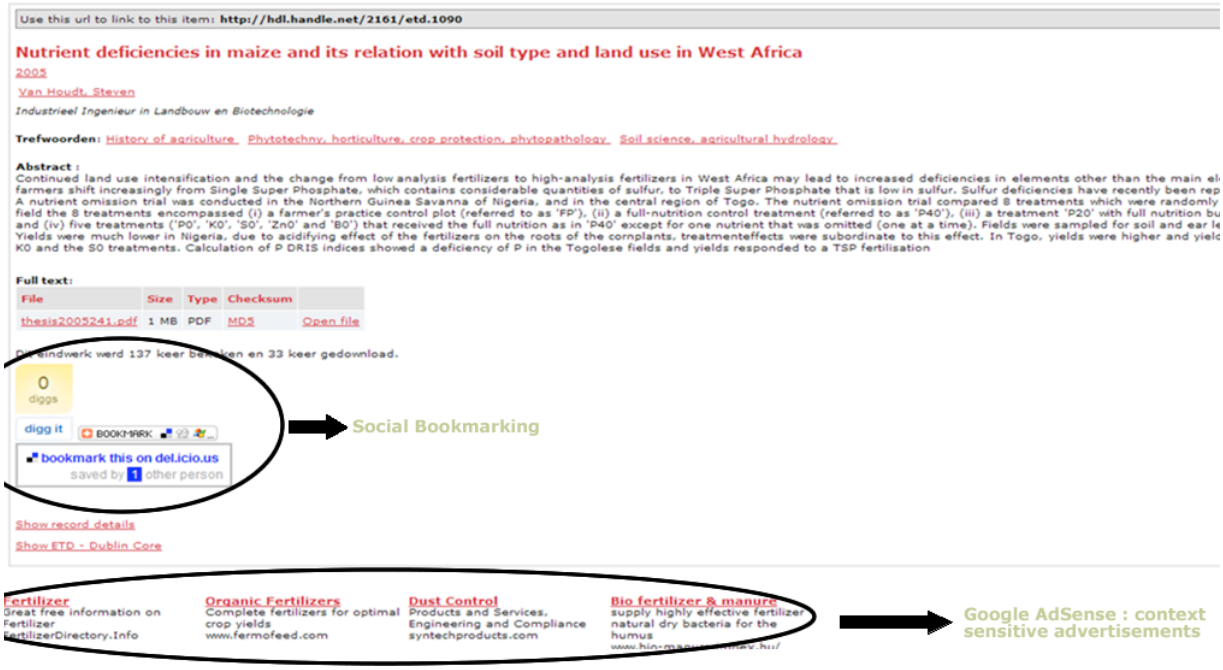

Figure 5: A thesis record splash page

In the following paragraphs the benefits of the major adjustments, namely the integration of social bookmarking tools and context sensitive advertisements will be discussed.

2.6.2 Social Bookmarking

In the light of our current subject classification which is deemed insufficient [19], a new opportunity is presented by the use of folksonomies or a tagging system. Furthermore, an interactive way of supplying keywords or tags perfectly matches the broader aims of the DoKS project, namely, knowledge sharing and community building. At the moment social bookmarking tools are provided on the theses records. By means of the 'Delicious Tagometer' (see Figure 5) it is easy to find out which other people have tagged a particular thesis record. This will lead you to the bookmark pages of people with common interests. It also allows you to see how what resources other users have tagged on the same subject. A desired feature that until now has not been available is a way of aggregating tags from different users of our portal in a tag cloud. Once such a feature is available we can provide this aggregation of tags to our users.

2.6.3 Instant Messaging

Although 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, another part (title, abstract, language, volumes, contact details, number of desired copies, instructor, trainee post, trainee supervisor) must be submitted by the students via the DoKS repository software. In addition the full text must be submitted by a self-archiving approach. Given that on an annual basis as many as 800 students submit their thesis data, they have to follow strict procedures and guidelines. By giving students the opportunity to ask questions whilst submitting data, the administrators can assist students directly should they encounter problems. In DoKS this communication is provided by a Meebo [20] widget and is similar to Instant Messaging systems the use of which

is very familiar to the students. Another attractive feature that comes with the integration of the Meebo widget is the ability to keep track of the number of concurrent users of your site.

2.6.4 Google AdSense

Once installed and filled with content a DoKS repository creates revenues via Google Adsense that can cover maintenance costs (upgrading hardware backup tapes) and/or new improvements. This is already the case for the repository of the Katholieke Hogeschool Kempen. The idea to start experimenting with Google Ads was based on a very practical mail question received from a user. The user had downloaded a thesis about laying out a private swimming pond. In the thesis prices for products were given which appeared to be much lower than the ones experienced by the user. Instead of contacting the student behind the particular theses we thought it might be easier to provide a direct path to suppliers of goods related to a thesis subject. Via Google AdSense relevant context sensitive ads are displayed on the pages containing theses records. The ads are related to the visitors' search and thus create a way to both monetize and enhance the theses records. We have currently been experimenting for almost a year with the system and evaluated that both benefits seems to be fulfilled. The ads are in most cases relevant and enhance our content.

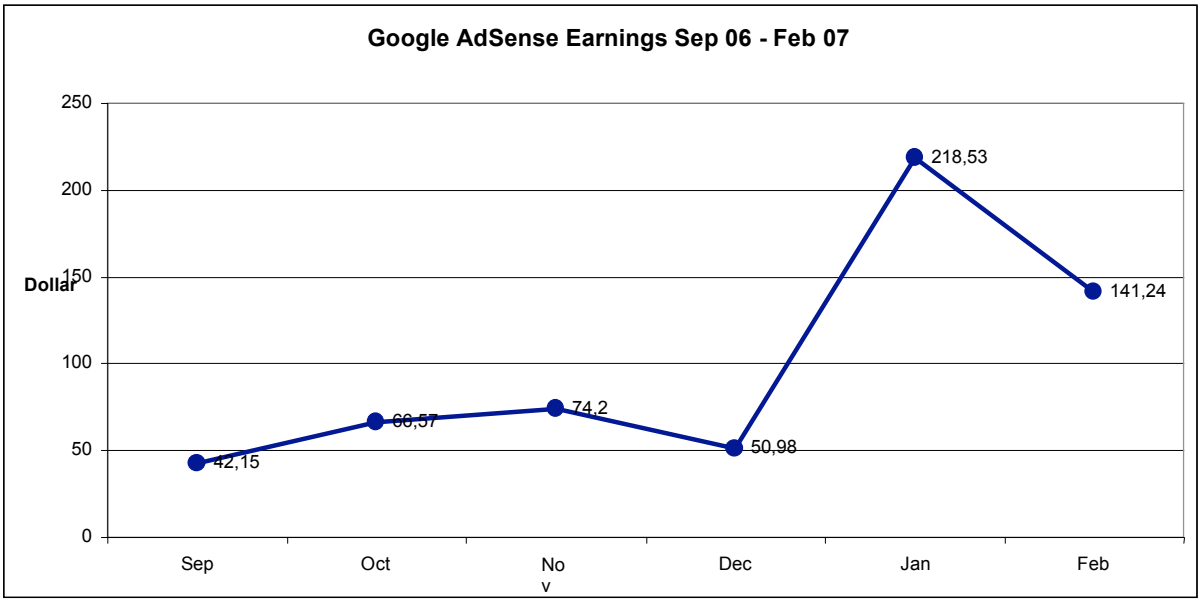

Figure 6: Google AdSense earnings Sep 06 – Feb 07

2.6.5 Print On Demand

With great surprise we noted a significant demand from our users to obtain a hardcopy version of the theses we published electronically. At first the intention was to deny these requests because it was thought that they would occur very occasionally and there were not the resources to give an appropriate answer. However more requests for hardcopies arose and by coincidence the DoKS portal caught the attention of a new player on the market of print on demand and self publishing, i.e. WWAOW (world wide association of writers) [21] This resulted very recently in a new collaboration and the first theses from different university colleges in Flanders are available via WWAOW (see Figure 7). Students are informed about this opportunity during the self-archiving procedure where they can choose whether they want to make their thesis available by means of the WWAOW website.

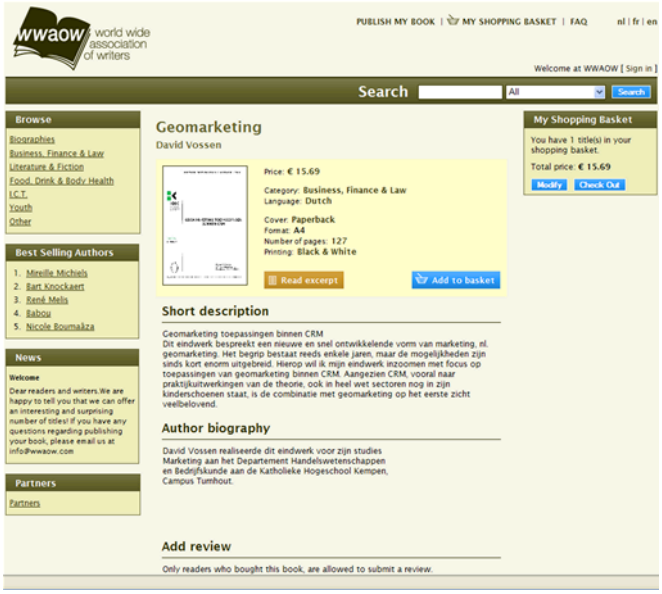

Figure 7: Print on demand via WWAOW

2.6.6 Alerting via Persistent Query Mechanisms and RSS

All authenticated users have a personal profile page in which they can store keywords (My Topics). By means of a persistent query mechanism search queries can be saved and can be executed again. This technique is used to provide an alert system. Once logged in a personal homepage is displayed (MyDoKS). On this homepage there appears a list showing documents which are new since the last time the user logged in.

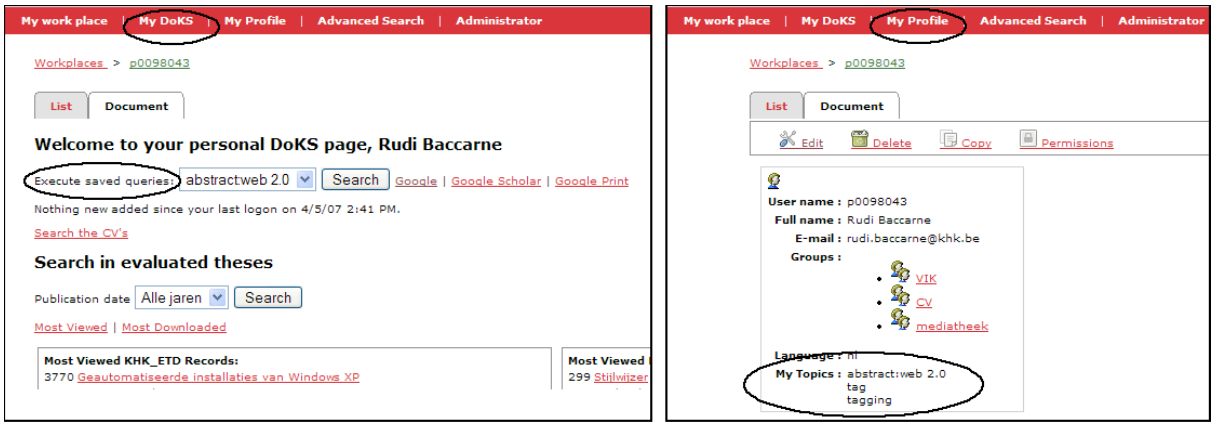

Figure 8: DoKS personal homepage and profile page

The built in RSS functionality can be used as an alert system as well. It is possible for example to subscribe to search queries via RSS, with the result that whenever a new item is published that reflects your query you will be alerted by you RSS reader. This RSS functionality is also a means to publish automatically updated lists. For example you can subscribe to a list of theses that are available by the print on demand system (see Figure 9), and new CVs to name but a few.

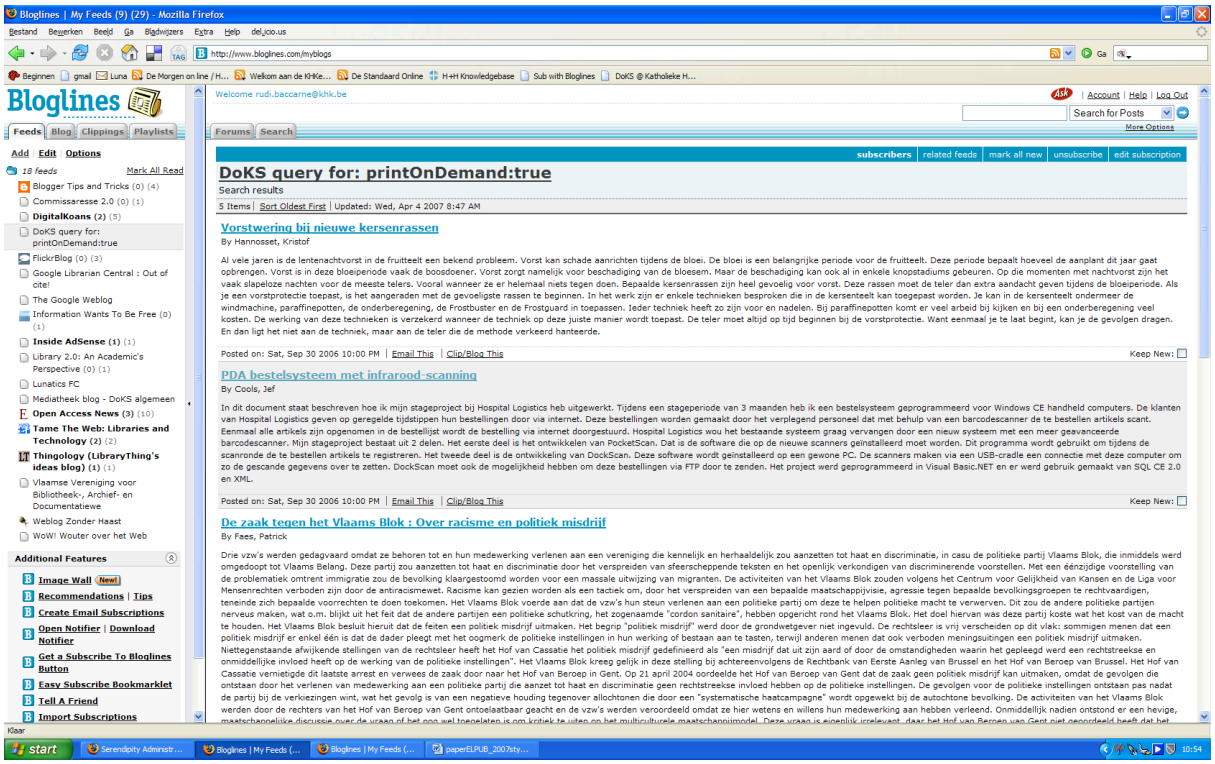

Figure 9: Example of a DoKS RSS search query subscription via Bloglines

4 Conclusion

Apart from being more visible to the scholarly community as well as to the labour market, students and university colleges will profit in the long term by contributing to the portal in several ways. The submitted work will have to meet certain conditions before being accepted for publication. Students learn about digital publishing and structured authoring. They have to deal with choices between different file and image formats, reducing file size and structured authoring, to name but a few. The wider availability of the work creates a mentality change amongst students and lecturers towards different phases in the electronic publication chain (citing, references, copyright, technical implications of electronic publishing, etc.). In the long term this will lead to better quality in electronic documents whilst at the same time students with a rather resistant attitude to computers, internet and the like will be introduced to the internet and electronic publishing.

When filled with a significant amount of content the DoKS repositories receive a high search engine ranking. As a result the number of daily visitors is significant. Among a variety of users, the feedback received at the KHK, has introduced several new business opportunities and interests. In this sense a successful repository can be seen as a powerful public relations tool. Because DoKS supplies a java and JavaScript-like scripting engine (Beanshell) for task automation, complex work flows, specialized import/export, etc. formerly manual processes such as MARC-export for the library, collection of abstracts and titles, publishing and so on are automated. Furthermore at the KHK DoKS is used to support services such as the employment agency and the research department. As a result the tool is highly appreciated by the users of the institution. Other benefits of integrating student scholarship in institutional repositories are discussed in many blogs and publications. A collection of similar and other arguments from which the quotation below is extracted is listed in the 'Law Librarian Blog' from Carol A. Parker [22]. 'The students' scholarship would attain visibility on a scale never before seen, and the students would enjoy the benefit of informing the subsequent work of others. Plagiarism should not be an issue because most institutional repositories are indexed at the full-text level, meaning that a simple Google search would quickly identify an existing paper that was later used without proper attribution. …

Digital collections of student work can also be used for publicity and outreach, especially with alumni. Many schools already inform alumni of recent faculty publications; alumni could also be informed of student scholarship published in repositories. Making student scholarship available in digital collections provides students with a connection to their schools after graduation.

Law schools would also be sending the message that they take student scholarship seriously. Knowing that their work will also be subject to scrutiny beyond the four walls of their professors' offices would give law students added incentive to produce better scholarship.'

References

[1] ETD portal for Flemish University colleges. See http://www.doks.be

[2] http://www.iwt.be

[3] See http://www.google.com/coop/

[4] http://doks2.khk.be/eindwerk

[5] The Open Citation Project - Reference Linking and Citation Analysis for Open Archives. *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies*. Retrieved 5 April 2007 http://opcit.eprints.org/oacitation-biblio.html

[6] PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B., (2007). *Sharing Detailed Research Data Is Associated with Increased Citation Rate*, PLoS ONE, March 21, 2007 Retrieved 5 April 2007 http://www.plosone.org/article/fetchArticle.action?articleURI=info:doi/10.1371/journal.pone.0000308

[7] DUVAL, E. (2004). We're on the road to…, in ED-MEDIA 2004. Lugano, Switzerland. Retrieved 6 april 2007 http://www.cs.kuleuven.ac.be/~hmdb/publications/files/pdfversion/41316.pdf

[8] For an overview of DoKS partners see: http://www.doks.be/partners.htm

[9] BACCARNE, R (2005). *From central administration of hardcopy Bachelor- and Master theses towards a decentralized ETD-system with value added services* / Conference Paper, ETD2005, the 8th International Electronic Theses and Dissertations Symposium. The Scientia, University of New South Wales, Sydney, Australia, 28-30 september 2005. Retrieved 5 April 2007 http://adt.caul.edu.au/etd2005/papers/045Baccarne.pdf

[10] For a full description of ETD-MS see: http://www.ndltd.org/standards/metadata/current.html

[11] http://www.ndltd.org

[12] http://pkp.sfu.ca/harvester_download

[13] The plug-in can be downloaded at: http://doks.khk.be/do/record/Get?dispatch=view&recordId=SDoc413ebf1711bbc5e40111bc0f15560001

[14] TAPSCOTT, D; WILLIAMS, A. D. , *The new Science of Sharing* In: Business Week, March 2, 2007 Retrieved 29 march 2007 http://www.businessweek.com/innovate/content/mar2007/id20070302_219704.htm

[15] DE CLERCQ, D.; MANIGART, S.; CLARYSSE, B.; CRIJNS, H.; DE SUTTER, M.; VERZELE, F. *Global Entrepreneurship Monitor: Executive report for Belgium and Wallonia, Vlerick Leuven Gent Management School*, 2003, 95 p. Retrieved 5 april, 2007 http://www.gemconsortium.org/document.asp?id=265

[16] VANNESTE, P.; BLOMME, E.; DESAEGER, A.; GRYMONPREZ, P. (red.), *Kennisvalorisatie als opstap naar innovatiebij KMO's en kleine non-profit organisaties*, VZW Kortrijks Ondernemerscentrum, Kortrijk, 2006, 53p.

[17] ANDERSON, C. *The Long Tail: Why the Future of Business is Selling Less of More*. New York : Hyperion, 2006.

[18] http://www.knosos.be

[19] For an essay on the insufficiency of traditional categorization methods for the electronic world, See: Clay Shirky. *Ontology is overrated: Categories, links, and tags*, 2005. Retrieved 10 April 2007 http://www.shirky.com/writings/ontology overrated.html.

[20] http://www.meebome.com/

[21] http://www.wwaow.com

[22] PARKER, C. A.[blog] http://lawprofessors.typepad.com/law_librarian_blog/2007/02/institutional_r.html