

The Miguel de Cervantes DL's Digital Theses Program

Julia Puig

Alejandro Bia

Miguel de Cervantes Digital Library

julia.puig@cervantesvirtual.com

alex.bia@cervantesvirtual.com

University of Alicante, apdo. 99, 03080, Alicante, Spain

<http://cervantesvirtual.com/>

Abstract

Ph.D. theses are a fundamental pillar for the application and development of knowledge. Generally, Ph.D. theses have a very limited reach. Their printed publication is rarely viable in a massive way. Often, only a partial publication is possible. Unfortunately, many original research projects, to which authors have devoted long working hours, lie unnoticed on the bookshelves of universities worldwide. Moreover, they are not easily accessed by potential scholars and interested researchers.

Electronic publishing makes it possible to reach every corner of the world and opens up new research and communication paths. The Miguel de Cervantes Digital Library welcomes both theses written in Spanish and theses dealing with the Spanish culture, coming from any part of the world. About 300 theses have already been published in our research section, from 33 national and international universities.

Theses in our research section are accessed through a specific catalogue where selection is possible by author, title or subject. The latter, also selected either by means of the CDU or the UNESCO classification. Alternatively, a quick search window can be used.

For every thesis, an author-related data sheet is added (including bibliography, specialization fields, knowledge areas), which also gives access to an optional discussion forum, allowing the constant exchange of ideas among researchers with similar interests. Other useful information such as the number of visitors is also provided.

In this article we describe the spirit and objectives of the MCDL's Theses Program, including copyright and legal issues, and digital preservation.

Introduction

The Miguel de Cervantes Digital Library (MCDL)¹ welcomes both theses written in Spanish and theses dealing with the Spanish culture, coming from any part of the world. But before going any further, it is essential to explain what the MCDL is. It is four different things:

1.- **A Digital Library:** The MCDL is a SYSTEMatically organized collection of all kinds of digital materials which are freely accessible through our Web site.

2 - **A Digital Publisher:** The MCDL produces its own digital resources using the latest technologies.

3.- **A research centre:** The MCDL is also focused on research and creation of advanced tools and methods to be applied to humanities research and to the development of digital libraries².

4.- **A communication channel:** The MCDL promotes the diffusion of the Hispanic culture and academic work.

The MCDL synthesizes these four concepts: It is a collection of digital materials; a digital publisher; a place where research takes place; and a communication instrument. Its main purpose, according to its foundational objectives, consists of promoting and spreading the Hispanic cultural values by the application of the most modern technologies to relevant literature, science and Hispano-American works.

Spanish scientific publishers joined the digital edition scene around the year 1995. Since then, the number of Web sites has increased a great deal, specially those of scientific magazines. Nevertheless, this progression had not quite the same pace as far as the digital edition of complete Ph.D. theses is concerned. The first step was taken by the University of Alicante through the MCDL in 1999. Afterwards, other universities followed such as the Universidad Complutense de Madrid, the Universidad Politécnica de Madrid and the TDC (project developed within the scope of the Catalan Digital Library), although in most cases the access to theses is limited. On the contrary, the Miguel de Cervantes Digital Library offers a free access service.

Access to the MCDL funds is possible by means of a general catalogue³ or alternatively by selecting specific specialized sections. One of these sections belongs to the Humanities Research Area⁴ and includes the Ph.D. Theses Section⁵. The Humanities Research Area is a project addressed to the research community and to specialists on different fields of science and the humanities.

1 <http://cervantesvirtual.com/>

2 According to the British Library Research and Innovation Centre, digital libraries are those that "use technology to acquire, store, preserve and facilitate access to data and publications originally edited on a digital version or digitalised from previously existent documents on a printed or audiovisual version or other".

3 The MCDL currently offers over ten thousand publications within its general catalogue: <http://cervantesvirtual.com/catalogo.shtml>

4 <http://cervantesvirtual.com/investigacion/index-en.shtml>

5 <http://cervantesvirtual.com/tesis/index-en.shtml>

Objectives

The main objective of the Ph.D. Theses Section consists of making possible the edition and Web publication of Ph.D. theses from universities that are linked to the rich plural Hispanic cultural heritage, using the most modern methods of digital edition.

It is interesting to think on the revolution that the Internet has meant for the publishing industry, especially with the academic world urging for the latest ideas and innovations. This was a giant achievement because information is now readily available in a very short time (Ph.D. theses are available on the net as soon as dissertations take place at universities) promoting as a result a rich interaction among scholars.

Although research works are generally published by researchers who are familiar with the net, they mostly use this channel to obtain information, and not for the exchange of ideas or as a source of bibliographic resources. This is a new comprehensive concept of communication and diffusion. Indeed, the purpose of our Ph.D. theses program is not only the mere publication of theses but to create a research section that is dynamic and alive.

Features of our edition

Apart from the subjectivity that any Internet publication implies (page layout, structure,...) our aim is to combine accessibility elements, scientific accuracy and standardization. With an author related data sheet⁶ we managed to join these requirements together with the publication of the full text, providing added value to authors and visitors.

Accessibility and technical accuracy

Theses are published in PDF format. In this way, we keep a faithful copy of the original contents and design, including charts and fonts. As explained earlier, the MCDL funds can be accessed by means of a general catalogue⁷. Similarly, the Humanities Research Area Web-page has a specific catalogue. A quick and an advanced search are also available.

The structure of this specific catalogue is simple:

- First, a directory that allows searches by author.

- Second, a directory to search by title.
- Third, a directory to search by university.
- Finally, we can search theses by the CDU and the UNESCO subject classifications.

These four access paths are essential in order to combine accessibility and scientific accuracy. All the publications in our Library are perfectly catalogued. The CDU and UNESCO subject classification features optimize the resource searches. In this way, users are able to group by subject any specific topic they are looking for. The simplicity of the SYSTEM makes it easier for the researcher to find information on any specific subject.

Structure of the data sheet

The author related data sheet⁸ covers the most relevant information on the author's lines of research:

- First of all, the author's name and links to all his/her publications within the Library.
- Then the thesis title, which is linked directly to the full thesis document.
- Finally, we can find the author's most relevant publications (bibliographic reviews), specialization fields, news, reviews, mentions, articles published⁹, degrees and qualifications.

By offering the visitor the possibility of looking up the latest reviews of the author's publications along with his/her lines of research, we offer a much wider vision of the document. Besides, authors may also add their own articles that have been previously edited by specialized magazines (with prior authorization of the publishers), which they expect to have a broader reach. This new project needs further developing, but the initial response to it has been very good, not only from article writers but also because of the reference made to the journal. This is therefore a two-way diffusion tool that benefits both the writer and the journal.

Discussion forums and online chats

We would also like to make a special reference to our discussion forums¹⁰, which have resulted in an increasing number of visits to our site. Forums provide researchers with a continuous information exchange flow. The Ph.D. Theses Section forums gather opinions, comments, questions and any other contribution from readers regarding authors and their work. Forum comments¹¹ sometimes

6 <http://cervantesvirtual.com/FichaObra.html?Ref=3628>

7 http://cervantesvirtual.com/tesis/tesis_catalogo.shtml#acatalogo

8 A data sheet, an enhanced electronic version of the traditional bibliographic cards of conventional libraries, accompanies every digital work of our library. In the case of theses, this card introduces the author and his/her work. See for instance: <http://cervantesvirtual.com/FichaObra.html?Ref=3628> or <http://cervantesvirtual.com/FichaObra.html?Ref=3971>. The number after "Ref" is the catalogue number used for every publication in our library.

9 The articles are also published in PDF format to keep their original layout and design, with maximum respect for the previous edition. <http://cervantesvirtual.com/FichaObra.html?portal=0&Ref=2624&otrasPublic=1>

10 Electronic discussion forums are provided for all author's data sheets, accessible by a link just after the main headings. You can also look them up together in one of our pages: http://cervantesvirtual.com/foros/foros_tesis.shtml

give way to replies from the authors, opening then a debate with fruitful results for all.

Our forums initiative have also resulted in online chats¹² about specific topics as requested by researchers. An online chat is a virtual meeting point for researchers and users where discussions are held about topics like literature, history, education, philosophy, literary creation, etc. To keep the scientific accuracy, once we identify a popular topic, we contact academic specialists who will develop their views online during the chat, with the help of a moderator. The text lines of these online chats are permanently published later on.

Access statistics

Another application that is quite useful for the author is the number of visitors to his/her thesis. The most read thesis has currently had over 9000 visits. This figure is growing day by day¹³. The MCDL monitors regularly the number of visitors to its more than 10.000 publications.

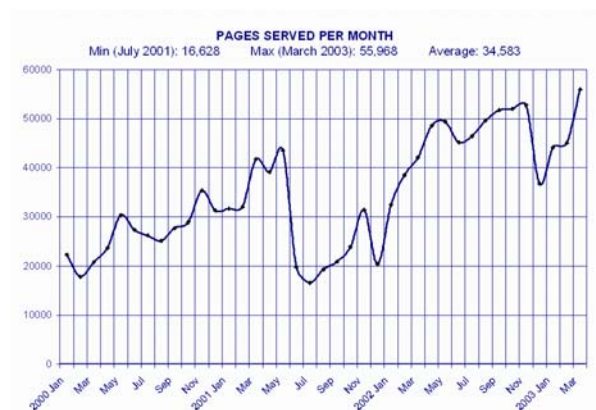


Figure 1: Web accesses to theses of the Miguel de Cervantes Digital Library (since Jan 2000).

The Humanities Research Area has performed an assessment survey on these visits and the Ph.D. Theses Section has produced a monthly visit chart out of these data. The scope of the survey starts the year after the MCDL opened to public, covering the period 2000-2003. We can draw some conclusions by studying the numbers of pages served (see figure 1):

1.- Visits increased as the Library consolidated into a stable and worldwide recognized foundation.

2.- We noticed how during the first months of the year 2000 the number of pages served increased slightly, in tune with the number of published theses. We had less

than a hundred theses published then, although the number was slightly increasing.

3.- The chart shows a significant fact: it is during the academic term months that the increase is greater. Obviously, the Library is visited by different types of users and it is interesting to see how this diversity has an effect on Ph.D. theses. The greater number of visits come from the academic field and these decrease as the holiday season begins.

4.- After the shy progress performed during the year 2000, the year 2001 followed with a similar pace, still suffering the same initial problems.

5.- Finally we entered a take-off phase: the number of visits increased significantly during the whole year 2003, even during those peculiar summer months. Once again, the Ph.D. Theses Section progress is similar to the progress made by the Library in general terms. Today the number of successfully served pages has reached the figure of **60.156.312**, with a daily average of 83.507. There is no doubt that all of our publications are benefiting from the growing popularity of digital editing, not only in our country but also at an international level. Our digital publication of Ph.D. theses reaches the whole international university community and the thirty-three universities presented in our catalogue give good evidence of this.

The number of visits to each thesis gives us the opportunity to find out which topics are the most popular among researchers. First, we have noticed that the subject of "sales and market research" is one of the topics of higher demand, followed by "didactics and education". Next, we find various subjects, going from *History* and *Literature*, to *Geography*, *Anthropology*, *Philosophy*, etc....Indeed, technology subjects are not the most visited¹⁴. Consequently, we venture to draw the conclusion that the topics of humanities knowledge are always alive in our educational field and somehow revive and face up to the fierce technologic research which encourages society to continuously demand the technologic dogma.

Copyrights and legal protection

The protection of the author's rights is an important issue to take into account when dealing with the digital publication of Ph.D. theses. Copyrights protection on the Web has introduced significant changes and challenges as far as the enforcement of national laws is concerned.

The globalization of telecommunication nets and accessibility from any part of the world to Internet information, have made it necessary to resort to international agreements and treaties to regulate copyrights, in order to provide rules for the new situation. E-commerce¹⁵ has

11 <http://cervantesvirtual.com/servlet/MuestraForo?obra=4530&comentario=4407>

12 <http://cervantesvirtual.com/tertulia/>

13 <http://cervantesvirtual.com/FichaObra.html?Ref=2244>

14 This page shows data about accesses to works which give us a general idea of the most demanded topics: <http://cervantesvirtual.com/servlet/SirveObras?estadistica=1&tipoObra=tesis&sumaParcial=1>

15 Hipola, P. "The Electronic Information Market". Jornadas Andaluzas de Documentación (meetings on documentation of Andalucía), November 1997, Sevilla, pp.21-25.

acquired an important economic and political position and therefore copyrights are critical for a stable environment.

We can remember at this point the OMPI (Organisation Mondiale de la Propriété Intellectuelle)¹⁶ International Conference that examined the impact of electronic commerce on intellectual property, and on the other hand the growing harmonization impulse given by the European Union in the Green Books¹⁷ on the matter.

¿How are these considerations put into practice within the MCDL? There is an essential point here: the edition of PhD thesis at the MCDL is free of charge, so the economical element disappears giving no way to electronic commerce. Legal protection is established by the current norms followed for all the works of the Library, but in addition we needed to cover a continuous demand for curricular recognition of such electronic non-commercial editions. The solution was simply to propose a "curricular"¹⁸ ISBN to defend the interests of authors that were publishing their research work for the first time. The Spanish Agency for ISBN proposed to cover, by means of a specific registry number for non-commercial publications, the so called "curricular" ISBN. There are many Spanish universities that have opened their doors to this initiative.

To conclude we can say that the MCDL has given an alternative answer to the academic world joining standardization and edition.

Long-term preservation of theses

According to Kenney et al [1], "the goal of digital preservation is to maintain the ability to display, retrieve, and use digital collections in the face of rapidly changing technological and organizational infrastructures and elements", in other words, to ensure the documents will be readable in the future.

The Miguel de Cervantes Digital Library has developed a digital preservation plan that includes the Theses section. The objective is to protect the documents against physical deterioration of the storage media, mishandling, obsolescence of file formats and to protect the integrity of the files in general.

The strategies employed are the usual: enduring care, periodical scheduled refreshing of the media and migration to fight obsolescence. With these purposes, preservation copies are kept in a fireproof safe located in a room with adequate environmental conditions to optimize their life expectancy. Periodic and SYSTEMatic integrity checks are performed to detect failures. Files on

optical media are regularly refreshed every two years by copying their content to new storage media. The older copies are kept as second backup copies in case failures are detected in the newest ones. As a rule, we try to stay aware of technological developments and particularly of format changes, to detect the need to migrate the files from older to newer formats before they become completely obsolete.

This may seem easy or obvious, but establishing a successful preservation plan is always a difficult task. We first had to raise the sufficient institutional commitment and top-level support to design and put into practice a long-term preservation plan. During the way, we had to fight other problems like the lack of financial resources and staff turnover. But the worst part was to convince the staff of the need and value of preservation. We had to teach them to think and work in terms of preservation. Before this, they were not only unaware of the problem, they just didn't care. So creating a conscience of the high values of preserving digital resources for future generations is a key aspect of a successful plan. Another key aspect, the impossibility of preserving everything, leads naturally to the awareness of the need to be selective with the targets of our preservation scheme. Here comes another challenge: to create a conscience of the need of selectiveness and to define the selection criteria.

Finally, a good preservation plan has to be periodically reviewed for many reasons: to encompass technology changes, to correct errors, to fine tune procedures, etc. So, once a year, we organize a meeting with the heads of the different areas of the digital library to discuss preservation issues and review the plan. This is very effective for keeping the preservation conscience alive and also to fight the consequences of staff turnover (i.e. unawareness of the problem and of the plan to tackle it).

Conclusions

The Ph.D. Theses Section of the MCDL struggles to fulfill its objective of spreading research knowledge to the global academic community through the Web. Our aim is to be able to offer a better service by optimizing searches of digital resources, by reducing waiting times for digital publication, by promoting dynamic scientific research communication and by developing efficient preservation strategies.

Our thesis section is alive and always growing. Daily experience results in the continuous integration of new ideas. Our goal is not only the mere publication of research work, but to build a rich and open communication chan-

16 OMPI treaty on commerce and intellectual property. 14th to 16th of September of 1999.

17 Green Book on copyrights and related rights in the Information Society. Proposal of the European Parliament and Council relative to the harmonization of certain aspects of copyrights.

18 http://cervantesvirtual.com/tesis/aviso_legal.shtml

nel for the global scientific community, with the aim of making true the Enlightenment dream of the "Republic of Letters".

Bibliography

- 1 ANNE R. KENNEY, OYA Y. RIEGER AND RICHARD ENTLICH : Moving Theory into Practice, Digital Imaging Tutorial (chapter 8: Digital Preservation). Copyright Cornell University Library/Department of Preservation and Conservation, 2001 <http://www.library.cornell.edu/preservation/tutorial/>.
- 2 HÍPOLA, P. El mercado de la información electrónica I Jornadas Andaluzas de la Documentación (proceedings) Sevilla Nov. 1997
- 3 HÍPOLA, P. AND MOYA, F. Proyectos EDI y normalización documental Revista Española de Documentación Científica 144 1991 408-419
- 4 PAJARES-TOSCA, SUSANA Espacio digital y publicaciones electrónicas Jornadas sobre publicación electrónica-Madrid Universidad Carlos III July 1999