

REPOSITÓRIUM - IMPLEMENTING DSPACE IN PORTUGUESE: LESSONS FOR THE FUTURE AND RESEARCH PATHWAYS

ELOY RODRIGUES¹; ANA ALICE BAPTISTA²; ISABEL RAMOS²;
MARIA FERNANDA SARMENTO E SOUZA³

¹ Serviços de Documentação da Universidade do Minho, Portugal.
eloy@sdum.uminho.pt

² Departamento de Sistemas de Informação, Universidade do Minho, Portugal.
analice@dsi.uminho.pt; iramos@dsi.uminho.pt

³ Centro de Estudos de Venenos e Animais Peçonhentos, Universidade Estadual Paulista, Brasil.
fsarmento@cevap.org.br

In this paper we describe the implementation of the Minho University (UMinho) Institutional Repository (RepositóriUM), Portugal. UMinho translated and has been using the first Portuguese version of DSpace, an Open Source software system developed in a partnership of the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP). The five phases established for the implementation plan are pointed out in this work (1. installation, translation, and configuration; 2. upload of thesis and dissertations; 3. establishment of pilot communities; 4. making the repository public, and 5. opening to the overall UMinho community), as well as, the reasons that led to the choice of the six pilot communities of different areas of knowledge (Information Systems; Polymers Engineering; Biological Engineering; Management, and one from the Humanities and another from the Social Sciences areas) to participate in this implementation period. In this paper we wish to share the lessons learnt, and the pathways for research, in particular for cross-cultural research, according to some questions and observations obtained during the first 10 months of experience. With this information as a background, some research pathways have been identified and are starting to be further developed in the context of MSc or PhD research at UMinho.

Keywords: open access; scholarly communication; institutional repositories; DSpace

INTRODUCTION

The creation of RepositóriUM was decided in the framework of Minho University (UMinho) project e-UM – Virtual Campus, submitted to e-U (electronic university) initiative, established by the Portuguese government. The project was committed to Documentation Services of Minho University (SDUM), the organizational unit of the University responsible for UMinho libraries and the authors of the project's proposal.

The first stage was a study of the state of the art on institutional repositories, and evaluation of the open sources platforms for their creation.

In April 2003 it was decided to use DSpace (<http://www.dspace.org>) - an Open Source software system developed in a partnership of the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP) - for the creation of Minho University Institutional Repository. On the 4th of November 2002 the launching of DSpace was announced. DSpace is defined by its creators as “a groundbreaking digital library system to capture, store, index, preserve, and redistribute the intellectual output of a university's research faculty in digital formats.” [1]. Among other features, DSpace is an open source software system (1) with a simple but effective software architecture, (2) using state-of-the-art technologies, (3) intended to be implemented in institutional repositories and (4) very much targeted to open access on scholar material. Additionally, a DSpace Federation

was created, composed of some of the most important North American universities, and this fact by itself is considered to be a proof of its importance and of some degree of reliability that the system will be used in the forthcoming years.

The UMinho was the first university in the Portuguese speaking worldwide community (of about 200 million speakers) and one of the first ones in Europe to translate and implement DSpace.

The RepositóriUM (<http://repositorium.sdum.uminho.pt>) - our instance of DSpace - is already running since July 2003. In this paper we wish to share the lessons learnt, and the pathways for research, in particular for cross-cultural research, that emerged in these 10 months of experience.

In the next section we will describe the implementation of DSpace in Minho University and the resulting institutional repository – RepositóriUM. Then we will make an analysis of the selected communities adherence to the system. In section four we describe some of the problems found and in section five we draw some research pathways. After the conclusions we talk about the future work, some of which is already running at Minho University.

IMPLEMENTING DSPACE

In the scope of the e-UM Virtual Campus activities, the UMinho Documentation Services (SDUM) initiated the DSpace activities in May, with its translation. The implementation plan consisted of 5 main phases:

1. Installation, translation, and configuration – This phase comprises all steps been done since the software was downloaded to our computers until it was ready for wide community use.
2. Upload of Thesis and Dissertations – feeding the system with thesis and dissertations developed in UMinho. Former UMinho PhD and MSc students were asked to send the digital (or paper) copy of their thesis and dissertations (as well as the authorization for publication) to the SDUM. These were then uploaded to the system and their metadata were edited and validated accordingly to the planned rules. This phase started in July 2003.
3. Establishment of Pilot communities – six (6) pilot communities were invited by SDUM to join the RepositóriUM with their scientific resources, like articles submitted to conferences or journals, grey literature, etc. Four (4) out of these six communities accepted to participate and sent resources to the SDUM to be uploaded to the system. Two communities did not send documents to the RepositóriUM.
4. Making the Repository Public – On November 20, 2003, and only with a slight shift on the calendar previously planned, the RepositóriUM was made public. A ceremony and press conference with the UMinho Dean was organized.
5. Opening to the Overall UMinho Community – Formally, RepositóriUM was open to all the UMinho communities since the public release of RepositóriUM. But, due to the holyday season, only in January 2004 a general call for participation was issued. The establishment of Pilot Communities (phase 2) was a keystone to this phase, once that some lessons learnt on that phase are helpful on this one.

Phase 1 was mostly technical and it run smoothly and on schedule. The main difficulties had to do with the minor adaptations derived from the UMinho organizational context that is different from the one at MIT. Some translations also resulted in inline code modifications. All changes done were carefully documented. Moreover, it was decided from the first day that no major changes to the DSpace code would be performed, in order that it could be easily upgraded.

In a technological point of view everything run as smoothly as expected. On the other hand, the social questions raised were not so easy to solve. The other phases in the plan resulted in the identification of a variety of problems, some of which could already be solved and some of which will require more time to study and, probably, more strategical and in-depth solutions.

ANALYSIS OF THE COMMUNITIES ADHERENCE TO THE SYSTEM

There were calls for deposit of thesis and dissertations in RepositóriUM. The first ones were issued in the mailing list of the University (UM-Net) on July 1st, 2003. In August, there were not calls and only 2 documents were received. The next call occurred on September 18 during the meeting at SDUM. From September 29, there was also a highlight in the homepage of SDUM about the deposit of thesis and dissertations.

We observed that people only used to send their documents on days right after the SDUM calls. Some days later, document traffic declined greatly, increasing again after another call two or three weeks later, and so on, until September. Figures 1, and 2 show the relation between the SDUM calls and document reception at SDUM's desks.

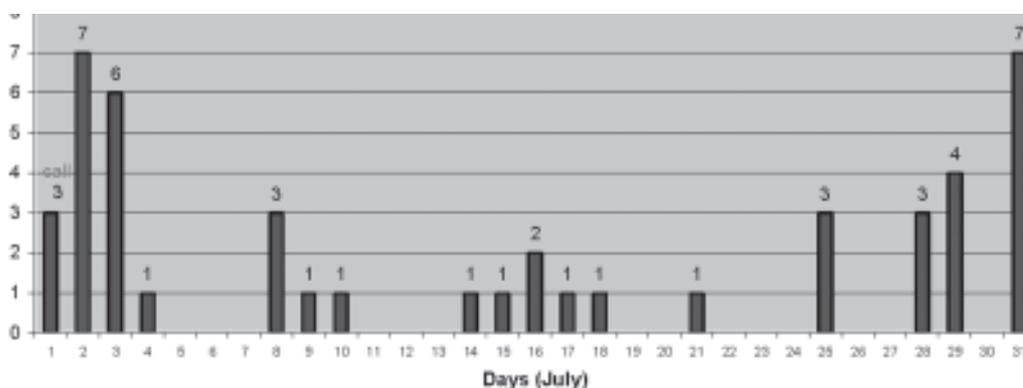


FIGURE 1 - NUMBER OF DOCUMENTS SUBMITTED TO THE REPOSITÓRIUM AFTER JULY 1ST 2003, WHEN THERE WAS THE FIRST CALL TO UMINHO COMMUNITY.

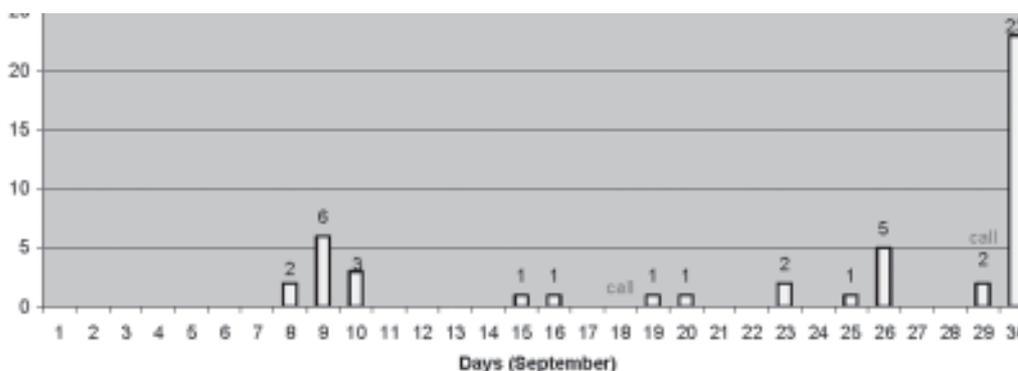


FIGURE 2 - NUMBER OF DOCUMENTS SUBMITTED TO THE REPOSITÓRIUM AFTER THE MEETING WITH THE DIRECTORS OF THE UMINHO PILOT DEPARTMENTS ON SEPTEMBER 18 AND THE HIGHLIGHT ON THE HOMEPAGE ON SEPTEMBER 29.

TABLE 1: TOTAL NUMBER OF DOCUMENTS SUBMITTED TO REPOSITÓRIUM UP TO MAKING IT PUBLIC.

Month	Number of documents	%
July	46	16.43
August	2	0.71
September	48	17.14
October	69	24.64
November	115	41.08
Total	280	100

Of the total number of papers included in the RepositóriUM, a higher percentage was observed between October and November, with a substantial increase in November.

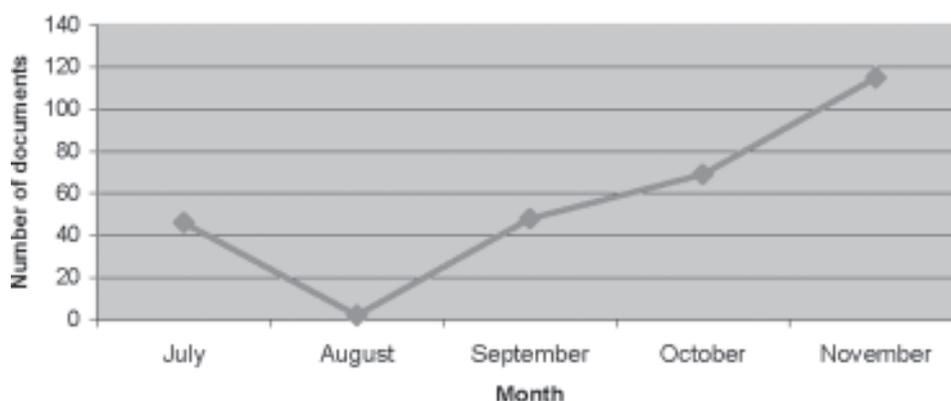


FIGURE 3 - CURVE OF GROWTH TENDENCY RELATED TO THE INCLUSION OF DOCUMENTS IN THE REPOSITÓRIUM.

In September 2003, simultaneously with the uploading of thesis and dissertations the following six units of UMinho were invited to use RepositóriUM as pilot-communities: (1) Information Systems; (2) Polymers Engineering; (3) Biological Engineering; (4) Management and two other more, one from the Humanities and another from the Social Sciences areas. These communities were chosen mainly for one or more of the following reasons: (a) strong research record (mainly by the number of publications and excellent external evaluation); (b) desire to have communities from different disciplines, geographical localization and institutional affiliation (different Schools/Institutes); (c) personal contacts in these communities that would (in principle) make them more active in adopting the technology.

As the workflow of deposit of thesis and dissertations was different from the collections and there were several from authors not affiliated with the pilot-communities, a Library of UMinho community was created, to store all thesis and dissertations.

The first four communities responded in a fairly way to the challenge, while the other two didn't at all. Although we didn't have the opportunity to make a detailed study, we may point out some reasons that we believe may have contributed to this failure:

- The repository was being made available as a project of the UMinho, it did not result of an expressed or unexpressed need of the researchers. In this case, there is a need for the project leader to spend a considerable effort championing the idea.
- Researchers were not aware of the big potential of the technology made available to them. It was not easily understood the benefits in storing in the repository the papers that were already published in well-established journals.
- Researchers in those departments are usually subject to substantial workload, trying to cope with their research, teaching, and management responsibilities. They saw the task of tracking past publications and upload them as a supplementary burden.
- Some of the communities are not comfortable and used to share their most recent research work. The effort seemed more of sharing knowledge with colleagues than make it visible internationally.
- Researchers less comfortable in using the broad range of information technologies now available felt considerable difficulty in using the interface of DSpace. Some researchers felt that it took considerable time to upload only one paper and gave up the task.

The delivery of resources to the RepositóriUM had a great variability accordingly to the scientific communities. From Table 2, it can be seen that the communities that were more responsive are those of Polymers and Information systems.

In this case we do not take into account the number of documents related to the SDUM community, once that it gathers thesis and dissertations stored by all UMinho communities, including the ones involved in the study.

TABLE 2: DISTRIBUTION OF DOCUMENTS BY COMMUNITIES.

Communities	Number of documents up to November/2003	Number of documents up to April/2004
Biological Engineering	22	22
Polymers Engineering	115	115
Information Systems	37	63
Management	15	15
Community from Humanities	0	0
Community from Social Sciences	0	0
TOTAL	189	215

Table 2 shows the number of documents uploaded to the RepositóriUM until two periods: Novembre/2003 and April/2004. These values indicate that five out of the six communities were not modified while one of them has improved its score. Two out of the six communities are already above the average on the second period measured and two out of the six still keeps a null score what indicates a total lack of interest.

Data allowed observing the differentiated reaction between the selected communities. There are two extremes of acceptance, where two communities are above the mean and the others below.

It was expected that all communities were close to the mean. In this case could be put the hypothesis that all of them had taken into account the solicitations of their management authorities. The fact that there are two communities with a null score, in the first and second measured periods, highlights the need to proceed with a qualitative study for this community in order to understand what is behind this behaviour of apparent disregard.

MAIN PROBLEMS FOUND

As in most of other institutional repositories the main challenge was populating the system. The response to the calls for thesis and dissertations was limited: 111 documents, most of them from the last 3 years (2000 to 2003). As for the contents from the pilot-communities most of them were collected by one or several pivots at each community, and then batch uploaded.

The number of self-archived documents, prior and after the formal opening of RepositóriUM is very small (less than 10 documents). There are several explanations for this fact, but most of them are related to the fact that faculty generally is not used to self-archive their own content on institutional repositories or other external archives.

Self-archiving is a new task, sometimes perceived as an additional burden in overcrowded schedules, not yet part of the habits and routines of researchers and academic staff. Additionally, most of researchers are not aware of the copyright licenses they sign when they publish in journals and that is a significant problem for self-archiving. To self-archive, authors must know their rights: what can they archive (preprints, postprints, etc.) and with witch access rights (open worldwide, restricted to their organization, etc.).

Besides the cultural, social and legal issues, there were also some questions regarding the quality control of the metadata associated with the documents deposited in RepositóriUM, namely the authority control of author names.

RESEARCH PATHWAYS

The creation of UMinho Institutional Repository and the project progress described in the previous sections of this paper have been supporting the identification of some research pathways and the beginning of further developments in the context of MSc or PhD research at UMinho. In particular, four researchers are actively working in trying to contribute to clarify part of the following issues:

6. How do researchers (in both roles of producers and consumers of information) react to this technology? What is involved in this reaction? What motivates it?

7. How does the reaction change (and how this change is expressed) across disciplines? And across cultures?

8. What are the important factors that help this technology to be better accepted? What can motivate researchers to use it? And how can they be motivated? What management initiatives should be implemented to foster a better communication and sharing of research results through the intensive use of the UMinho Institutional Repository?

9. What is the applicability of DSpace to other kind of materials? What should be maintained and what should be changed? And with other users?

10. What other functionalities could be added to DSpace in order to foster relationships between researchers?

While the use of DSpace is still in its beginning, there are already some insights that can already be drawn. The project started as an initiative of the Documentation Services of Minho University. This initiative resulted from the clear understanding of the importance of the RepositóriUM to the dissemination of the University research, both internally and externally. However, while most of the involved departments understand this importance, many researchers expressed less enthusiasm for the reasons listed in section 3. The few initiatives to champion the repository led to an increase in the upload of research papers in the Information Systems (Table 2). This still restricted experience is motivating us to plan similar initiatives in other departments.

Also, there is some evidence that by disseminating the system's statistics showing that stored papers are being downloaded by nationally and internationally can increase the researchers' interest in the repository. However, this may not be true to the whole scientific communities of the University. In those scientific fields where there is a culture of sharing research in its early stage, researchers will probably feel motivated to make the effort to upload their most recent papers. In the scientific fields where such a culture does not exist, the easy access provided by the repository may be a key factor sustaining the resistance to use the repository.

Table 2 shows that the Biological Engineering and Management departments show little adherence to the repository, and Humanities and Social Sciences communities did not manifest. One of the reasons may be the fact that researchers in these departments have less experience in using information systems regularly. This fact may result in an increased tension performing the required tasks to upload a paper.

CONCLUSIONS

The University of Minho was the first institution in the Portuguese speaking worldwide community (of about 200 million speakers) and one of the first ones in Europe to translate and implement DSpace.

The RepositóriUM (our instance of DSpace) first phase implementation plan was divided in five main steps, starting in May 2003 with the system translation. The last step comprised the opening to the overall UMinho community and was initiated in November 2003.

During these months of activity (from May 2003 until April 2004) we faced many problems and acquired a lot of experience we are able to share with other scientists.

The experience of the project clearly supports all empirical evidence obtained worldwide about communication and collaborative systems. Collaboration, communication and sharing of knowledge are not implemented merely by the implementation of information systems that support them. There is a need for management to create and nurture a culture of collaboration, communication and sharing through management initiatives. Cultural changes are not easy to carry out and may

have unexpected results since they are about changing perceptions, feelings and understandings deeply ingrained in a community.

Deeper and very well targeted studies will be needed to try to find solutions and propose pathways for the implementers, especially in what regards to social and political constraints and behaviours not always taken into account by computer scientists.

Some of these studies are already being developed in the Information Systems Department at UMinho.

FUTURE WORK

There are two kinds of future work planned: development and research. The development efforts will be performed mainly by the SDUM and the research efforts will be performed mainly by UMinho scientific departments, in particular the Information Systems Department.

In the second semester of 2004 the development of RepositóriUM will be led in several directions. From the previous experience, the most important strategic vector is to promote author initiated deposit, by self-archiving or mediated upload, of contents. In this direction, SDUM will:

- Enhance the usability of the deposit interface and the support information (guides for the deposit of different types of documents, contextual helps and examples) of the deposit process;
- Develop a copyright help and clearinghouse service, integrated in the deposit workflow;
- Integrate a end user license step (using the Creative Commons standard licenses) in the deposit workflow;
- Integrate the deposit of thesis and dissertations submitted and approved by Minho University into RepositóriUM with the administrative/academic processes and information systems;

Another important strategic issue is the demonstration of value and advantages of RepositóriUM to the individual researchers and the organizational units where they work. The communities that are already using the system are being surveyed regarding added value services and functionalities that they consider important to be included in RepositóriUM. The final list of developments is still being defined, but it will certainly include:

- Statistics and reports of access and download of papers, at individual and community level;
- Lists of publications by author and organic units (generation of WebPages of references, generation of the publications section of periodical evaluation reports that researchers must produce, etc.);
- Citation analysis services of papers included in the RepositóriUM;

Regarding the research activities, some of the planned work is already being done while some other is planned to begin in the near future.

The use of the RepositóriUM project is being studied by a PhD student that is surveying attitudes and behaviours of researchers from the Minho University different knowledge fields and cultures towards the use of the repository. Attitudes and behaviours will be surveyed using a measurement tool being developed. In parallel, a qualitative study will be performed to identify the reasons underlying the behaviours patterns identified by the survey.

The Information Systems department is also participating in the design of an R&D project proposed for EU funding. In this project, the described survey will be expanded to include several

other European countries. In this way, we will be able to compare the behaviour patterns of the same scientific communities in different institutional and national cultures. We expect that the broadening of the study help us in identifying a list of success factors and barriers in implementing institutional repositories and a list of strategies and changes to be adopted by academic institutions to successfully implement an institutional repository.

An MSc student is developing his work on finding out which statistics and system usage feedback information should be retrieved to motivate scholars to include their materials. He will then develop and integrate the necessary code to generate these statistics and present them to the final user in a friendly way.

Another MSc student has been studying metadata element sets for e-learning environments. A DSpace instance (papadocs – <http://papadocs.dsi.uminho.pt>) for students' assignments has been installed and is running. Some adaptations (more than the ones made for the RepositóriUM) were needed and have already been made. Other adaptations and integration with e-learning environments are to be studied in the next future.

Another researcher is working on informal communication mechanisms to be added both to papadocs and to RepositóriUM. Some developments have already been made and are now being tested. Information to the DSpace community will be released as soon as a stable version is achieved.

REFERENCES

(1) DSpace Federation - DSpace at MIT. (2003). Retrieved 28/01/2004, 2004, from <http://DSpace.org/introduction/index.html>