The Participation and Web Visibility of University Digital Repositories in the European Context

Participación y visibilidad web de los repositorios digitales universitarios en el contexto europeo

ABSTRACT
This study focuses on academic institutional repositories as tools that allow us open access to scientific and academic production. Specifically, we analyze the Top 50 European University repositories differentiating, firstly, those repositories linked to Spanish universities compared to those belonging to universities throughout Europe and, secondly, repositories that only include research content as opposed to those that also include teaching content. Specifically, this work complements previous studies on the consolidation of the repositories, focusing on the analysis of the competitive environment by considering their participation and relative visibility shares. The analysis, using competitive maps and comparative advantage method, allows us to identify European university repositories that lead their segments, in terms of their levels of participation and web visibility in the market. In general, without distinguishing by segments, results show that the leadership at European level in terms of participation is held by the University College of London (UK) and the repository of the University of Umea (Sweden) is the leader in visibility.

RESUMEN
El presente estudio se centra en los repositorios institucionales universitarios como instrumentos que posibilitan el acceso en abierto a la producción científica y académica. Se analizan los Top50 repositorios universitarios europeos diferenciando, en primer lugar, aquellos repositorios vinculados a universidades españolas frente a los pertenecientes a universidades del resto de Europa y, en segundo lugar, los repositorios que incluyen en sus contenidos exclusivamente resultados de investigación frente a aquellos que también albergan recursos docentes. En concreto, este trabajo complementa estudios previos sobre la consolidación de los repositorios, profundizando en el análisis del entorno competitivo a partir de sus cuotas relativas de participación y de visibilidad web. El análisis efectuado, a través del diseño de mapas competitivos y la aplicación del método de la ventaja relativa, permite identificar los repositorios universitarios europeos líderes en sus segmentos respecto a sus niveles de participación y visibilidad web en el mercado. A nivel general, sin establecer diferencias por segmentos, los resultados muestran que el liderazgo a nivel europeo, en términos de participación, lo ostenta el University College of London (Reino Unido), mientras que el repositorio de la Universidad de Umea (Suecia) es líder en visibilidad.

KEYWORDS / PALABRAS CLAVE
Repository, scientific communication, open access, universities, scientific production, learning, objects, participation, penetration.
Repositorio, comunicación científica, acceso abierto, universidades, producción científica, recursos, aprendizaje, participación, penetración.
1. Introduction

The study of digital repositories is currently very important because since The Budapest Declaration (BOAI, 2002), which established the first formal definition of the open access movement (ratified and expanded in The Bethesda and Berlin Declarations of 2003), the implantation and development of repositories of electronic documents have increased substantially. According to the Ranking Web of World Repositories, there are more than 1,500 digital repositories in 2012. The importance of repositories in the communication of scientific knowledge and their role in strengthening the cooperative spirit in scientific research have led to the need to analyze them.

Coinciding with the rise of the World Wide Web in the 90s, projects linked with the open access movement began to appear. This entailed free Internet access, with no economic or copyright restrictions, to the scientific literature (Suber, 2005). The arXiv repository of pre-publications, founded in 1991 in the field of Physics, is considered to be the pioneer in the development of digital repositories.

If we focus on the strategies that characterize the implantation and development of the open access movement, it is self-archiving or the green route that began and nurtured the digital repositories (Harnad & al., 2004; Sánchez & Melero, 2006). Apart from publication in journals, this strategy means the placing of a copy of a study in a stable repository that allows free on-line access. In this context, the term «repository» entails an expansion of the preservation and conservation characteristics of an archive since, apart from storing information, a repository has other functions such as the supply, management, recovery, visualization and reutilization of digital documents (Pinfield, 2009). In this sense, open access to a repository adds easy availability of content that may come from various sources to the advantages of no cost or unlimited access to information. Independently of their role of provider of data and/or services (Hernández, Rodríguez & Bueno, 2007), repositories can be implemented by institutions, thematic communities, research centers or other groups. This study focuses on the study of institutional repositories which, according to The Budapest Declaration (BOAI, 2002), arose in response to the need for academic institutions to conserve and preserve their intellectual property and make it available to the education and research community.

There is much debate around the content of repositories; some authors (Crow, 2002; Johnson, 2002) defend education-learning as one of the key functions of university, believing that teaching materials should be included along with research results. Taking this point further, repositories specializing in teaching design could be a tool for educational staff to learn different teaching strategies such as a detailed explanation of the steps to be taken in their implementation (Marcelo, Yot & Mayor, 2011). Other authors oppose this position, supporting the premise that the purpose of an institutional repository is the diffusion of research results and hold that the key factor is free access to these results (Harnad, 2005; Sánchez & Melero, 2006).

Notwithstanding this open debate, Lynch (2003) defines the institutional repository in the area of universities as a collection of services that a university offers to the members of its community for the management and diffusion of digital materials created by the institution and its members. Hence, it is an organizational obligation to manage digital material that includes its long-term preservation, its organization and its access or distribution (Lynch & Lippincott, 2005). In line with Crow (2002), institutional repositories comply with 2 of the strategic factors of universities. First, these repositories constitute a critical component of the academic communication system by expanding access to research, increasing competition and reducing the monopolistic power of the journals. Second, they can be quantitative indicators of the quality of a university and they can demonstrate the scientific, social and economic importance of academic activity; thus increasing the visibility, status and public value of the institution. In a broad sense, university repositories collect part of the intellectual production of universities, in that they are where the organization, preservation and diffusion of digital documents derived from academic work take place.

The study of repositories is a current hot topic (Barreto & García, 2009; Ezema, 2011; Galina, 2011). And within this field there are various lines of research, such as those focused on the analysis of the technical factors around the implementation of repositories (Koopman & Kipnis, 2009; Subirats & al., 2008), on attitudes to self-archiving (Carr & Brody, 2007; Chuk & McDonald, 2007; Xia & Sun, 2007), on free access and the impact of citations (Davis, 2010; Gaulé & Maystre, 2011; Giglia, 2010) and on the evolution of repositories (Keefe, 2007; Krishnamurthy & Kempraj, 2011; Peset & Ferrer, 2008; Wray, Mathieu & Teets, 2009). This study belongs in the last of these lines and aims to analyze the competitive environment of university repositories through the volume of digital content, participation of a repository in the supply of digital content and the web visibility of a repository.
The study also uses a double segmentation to consider the geographical context of the universities that host the repositories and the type of digital content stored in them.

After these initial considerations, the following section describes the methodology and identifies the sources of information and variables used. Next we present the results differentiated by geographical area and by the content type of the repositories and we finish with the conclusions derived from the study.

2. Material and methods

2.1. Methodology

After describing the current situation of university repositories in Europe, we use information visualization (Chen, 2003) to analyze their competitive environment through a comparative map. More precisely, we use a variant of the dispersion diagram that positions Spanish university repositories against those of the rest of Europe in terms of 2 dimensions: their participation and visibility shares compared to the other competitors in their segment; each repository is represented by a circle which is indicative of the volume of digital documents derived from the academic production of the host universities. The analysis by geographical area considers an additional segmentation around the content of the repositories, differentiating those with content derived exclusively from research from those that also include teaching resources (mixed repositories).

The final position occupied by a repository in the diagram described above allows us to identify the leaders in the analyzed dimensions (repositories with relative shares above 1). If there is no single repository that leads in both dimensions, the leader is identified through the relative advantage method. This method implies initially obtaining the advantages, in terms of participation and of visibility, for the 2 repositories that are leaders in each dimension. Next we compare the above advantages, with the dimension that has the greater relative advantage being the identification criteria for the leader repository.

2.2. Data and variables

The university repositories to be analyzed are identified using the Ranking Web of World Repositories (RWWR) of the Spanish National Research Council (Aguillo & al., 2010). Using the latest available edition (April 2012), we select the 50 main repositories linked to European universities, discarding those with incomplete information on the number of entries in the analysis period (see Table 1). This ranking also provides the degree of visibility of the selected repositories. We use the Registry of Open Access Repositories (ROAR) to find the size of the repositories through the accumulated number of entries from the foundation date until 31st December 2011. The evolution of entries during 2011 gives us the participation share for this period for each repository. Finally, the Academic Ranking of World Universities (ARWU) allows us to identify universities and their geographical distribution.

Using the above information, we construct the following variables that allow us to analyze the competitive environment of the Top50 European university repositories.

1) Relative participation share (CPRijk) of university repository i (i=1...l) in geographical area j (j=1 (Spain), 2(Rest de Europe)) and of type k (k=1 (mixed), 2(research), so that:

\[ CPR_{ijk} = \frac{CP_{ijk}}{CP_{C(j)k}} (i \neq C1) \]

with

\[ CPR_{ijk} = \frac{CP_{ijk}}{CP_{C(j)k}} (i \neq C1) \]

\[ CPR_{ijk} = \frac{RT_{ijk}}{RT_{C(j)k}}, \text{where:} \]

- CPRijk: Participation share of repository i in geographical area j and of type k.

This study belongs in the last of these lines and aims to analyze the competitive environment of university repositories through the volume of digital content, participation of a repository in the supply of digital content and the web visibility of a repository. The study also uses a double segmentation to consider the geographical context of the universities that host the repositories and the type of digital content stored in them.
• CPC1jk: Highest participation share of the repositories in geographical area j of type k.
• CPRC1jk: Relative participation share of the repository with the highest participation share in geographical area j of type k.
• CPC2jk: Participation share of the 2nd best competitor in geographical area j of type k.
• RTijk: Total entries in repository i in geographical area j of type k in the year 2011.
RTjk: Total entries of the repositories in geographical area j of type k in the year 2011.

2) Relative visibility share (CVRijk) of repository i (i=1,…, Ij) in geographical area j (j=1(Spain), 2(Rest of Europe)) of type k (k=1(mixed), 2(research)), so that:

\[
CVR_{ijk} = \frac{CV_{ijk}}{CV_{C1jk}}, \quad (2)
\]

with \(CVR_{C1jk} = \frac{CV_{C1jk}}{CVC1jk}\) and \(CV_{ijk} = \frac{V_{ijk}}{V_{i}}\), where:

• CVijk: Visibility share of repository i in geographical area j of type k.
• CVC1jk: Highest visibility share of repositories in geographical area j of type k.
• CVRC1jk: Relative visibility share of the repository with the highest visibility share in geographical area j of type k.
• CVC2jk: Visibility share of the 2nd highest competitor in geographical area j of type k.
• Vijk: Visibility of repository i in geographical area j of type k.
• Vi: Visibility of the repositories in geographical area j of type k.

Bearing in mind that the degree of visibility (V) of repository i (i=1,…, 50) is:

\[
V_{i} = \frac{1}{E_{linki}}, \quad (3)
\]

where Elinki represents the position in visibility terms provided by the RWWR, obtained by the number of external links received by repository i (Aguillo et al., 2010).

3) Size (TR) of repository i (i=1… 50) until day T (31 December 2011):

\[
TR_{i} = \sum_{t=1}^{T} DD_{i}, \quad (4)
\]

where:

• DDt: Number of digital documents of repository i on day t.
• Fi: Foundation date of repository i.

Hence, using Equation (1) we quantify the relative participation shares of the repositories –as a measure of the degree of participation of each repository in the supply of digital content stored in all the repositories considered-, differentiating Spanish repositories from those of the rest of Europe and repositories with only research content from mixed repositories. For the generic case of a given repository in a concrete segment, the relative participation share is the quotient between its participation share and the highest share...
of its segment; for the repository with the highest participation share, we divide its share by the 2nd highest share. For the calculation of participation shares we consider the number of entries received by the repository in 2011 compared with the number of entries of all the repositories in the segment in the same period. With a similar method, equation (2) finds the relative visibility shares of repositories by segments—as a measure of the level of market penetration—considering visibility as the inverse of the position in terms of this variable given by the RWWM. Finally, in equation (4) referring to the size of the repository—as a measure of digital academic production—we consider the number of digital documents accumulated in the repository from the foundation date until the 31st December 2011.

3. Analysis and results

The Top50 European repositories analyzed are distributed so that 12% belong to Spanish universities and the remaining 88% to universities from the rest of Europe. In terms of content, 56% only store research results and 44% are mixed repositories. The repositories considered have an average of 33,630 digital documents, ranging from the 234,760 entries of the University College of London (United Kingdom) and the 1,502 of the University of Oulu (Finland).

Looking at the analysis of the competitive environment of European repositories without differentiating by segments, the repository of the University of Umeå (Sweden) is leader in visibility and the University College of London (United Kingdom) is leader in participation. In terms of the segmentations by geographical area (Spain versus the rest of Europe) and by content type (research versus mixed), Figure 1 shows only the leading repositories in the 3 dimensions analyzed. Each repository is represented in terms of its relative participation and visibility shares, and its size.

The comparative analysis using the double segmentation, and initially focusing on the Spanish repository market, shows that the repositories of the Autónoma University of Barcelona and the Polytechnic of Madrid have relative participation shares above 1. Therefore, the repositories of these universities are leaders in the supply of digital content, with the Polytechnic of Madrid being leader in the research only segment and the Autónoma University of Barcelona leader in the mixed segment. Turning to visibility, the leading Spanish repositories are the Polytechnic of Cataluña and the Autónoma University of Barcelona for research only and mixed repositories respectively. Given that visibility is related to the number of links received by each repository, these 2 universities are leaders in terms of market penetration.

Moving on to the rest of Europe, we find that the University of Liège (Belgium) and the University College of London (United Kingdom) are leaders in participation in the research only and mixed segments respectively. The leaders in terms of penetration are the University of Umeå (Sweden) for research repositories and the University of Utrecht (Netherlands) in the mixed segment.

The size of the bubbles in Figure 1, which shows the supply of digital content, gives us the highest volume repositories for the segments considered. The University Carlos III of Madrid has the largest research repository and the Autónoma University of Barcelona has the largest mixed repository. In the rest of Europe, the repositories of the University of Amsterdam (Netherlands) and the University College of London (United Kingdom) are the largest in the research and mixed segments, respectively.

Apart from the Autónoma University of Barcelona, which is the leader in participation and penetration in Spanish mixed repositories, there are no repositories that lead in both dimensions; some lead in participa-
tion and others in penetration. Hence, in these cases and for the other segments, we find the leading repository in the 2 segments by applying the relative advantage method described in the previous section. The application of this method shows that the leader in the Spanish research repositories segment is the Polytechnic of Cataluña; the University of Umea (Sweden) is the leader in research repositories in the rest of Europe; and, finally, the leader of the rest of Europe mixed repositories segment is the University of Utrecht (Netherlands).

To go further into the characterization of repositories that do not lead in any of the dimensions considered, figure 2 identifies repositories with content supply and relative participation and visibility shares that are above average for the non-leaders group. We obtain these average values through the maximum and minimum values in each dimension.

Looking at figure 2 and focusing on the Spanish repositories, we find three repositories that stand out for their above average values for relative participation and visibility shares. While the research only repository of the University Carlos III of Madrid and the mixed repository of the University of Alicante stand out in terms of participation, the repositories of the universities of Complutense of Madrid, Alicante and Carlos III of Madrid stand out in terms of market penetration. With regard to repositories from the rest of Europe, the research repositories of the universities of Milan (Italy), Amsterdam (Netherlands) and Glasgow (United Kingdom), and the mixed repositories of the Federal Polytechnic School of Lausanne (Switzerland) and the University of Southampton (United Kingdom) stand out in participation. In terms of penetration, notable repositories are the research repository of the University of Humboldt (Germany) and the mixed repositories of the universities of Oulu (Finland), Stuttgart (Germany), Saint Gallen (Switzerland) and Southampton (United Kingdom).

To synthesize the information in Figures 1 and 2, Table 2 shows the leading repositories and those that are above average in the dimensions of participation, visibility and size for the segments considered. This table shows the absolute leader repositories in their segments after applying the relative advantage method; in other words, those that lead in both participation and penetration.

4. Discussion and conclusions

In response to the cementing of the position of free access as a model of scientific communication in the scientific-academic world and the growing number of institutional repositories, we propose the need to evaluate this type of application. This study analyzes the market of the Top50 European university repositories, differentiating within the same competitive environment repositories linked to Spanish universities from those pertaining to universities from the rest of Europe and further differentiating repositories that only store research results from those that also include teaching resources. Concretely, this study complements previous studies on the consolidation of repositories that focus on the volume of digital content derived from the production of universities. As a new contribution, we go deeper into the analysis of the competitive environment of the repositories through their relative participation and web visibility shares, which identify the leading repositories in a double segmentation by geography and content type.

Looking at the Spanish repositories, there are currently 6 Spanish university repositories in the Top 50 European institutional repositories: the universities of Autónoma of Barcelona, Polytechnic of Cataluña, Alicante, Complutense of Madrid, Polytechnic of Madrid and Carlos III of Madrid. Looking further into...
the national context, the first positions in the dimensions analyzed are held by the research repository of Carlos III University of Madrid and the mixed repository of the Autónoma University of Barcelona. However, the Polytechnic of Madrid holds first place in participation in research repositories and the Autónoma University of Barcelona leads the mixed repositories segment. In terms of market penetration, the Polytechnic of Cataluña and the Autónoma University of Barcelona have the leading research and mixed repositories, respectively.

Turning to the rest of Europe, we find that the University of Amsterdam (Netherlands) and the University College of London (United Kingdom) have the largest repositories, the former in the research segment and the latter in the mixed segment. The universities of Liège (Belgium) and the University College of London (United Kingdom) are leaders in participation in the research and mixed segments, respectively. The leaders in terms of penetration are the research repository of the University of Umeå (Sweden) and the mixed repository of the University of Utrecht (Netherlands).

Accordingly, the leading universities in relative participation share give more importance to the basic functions of storage and preservation that characterize institutional repositories. These universities develop their repositories as a complement to the traditionally used options for presenting academic production. In this sense, they are using their repositories to make themselves better known by offering open access to a wide variety of the teaching and/or research output of their academic staff. In the terms of penetration, leading positions in web visibility of academic output strengthen the function of diffusion of own knowledge of the repository as a means of communication. Therefore, leading positions in both participation and penetration allow a university to not only make itself better known than others, with regard to its academic output, but to also increase possible access to this academic output. In this sense, the leading repositories in the dimensions considered gain importance as means of communication of teaching and research knowledge, with emphasis on the functions of storage, preservation and diffusion of knowledge.

Although this study characterizes the main university repositories in terms of volume of digital content, participation in the supply of this content and web visibility, there is scope to continue this line of research with a causal analysis to identify the determining factors of the leading positions in the different dimensions. Among other aspects, factors such as the language of the repository, the diversity of the content, the size of the institution or its funding could be analyzed to see whether they influence the leading positions. Similarly, and taking the premise that a large presence in the market through high content volume is not the only important factor, researchers could also investigate the quality of the content stored in repositories as an additional key factor in the evolution of these instruments that give open access to scientific output; this could be another future research line.

References


<table>
<thead>
<tr>
<th>Repositorios institucionales universitarios</th>
<th>España</th>
<th>Resto de Europa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posición</td>
<td>Participación</td>
<td>Visibilidad web</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| | | | |
| | | | |

Table 2. Principal institutional university repositories.


