

Cite as:

Rodríguez-Armentia, N; Amat, CB: Is it worth to establishing institutional repositories? The strategies for Open Access to Spanish peer-reviewed articles. Forthcoming *Learned Publishing*, 2010.

Is it worth to establishing institutional repositories? The strategies for Open Access to Spanish peer-reviewed articles.

Nerea Rodríguez-Armentia

Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC)
PO Box 73
46100 Burjassot, Valencia Spain

Carlos B. Amat*

INGENIO [CSIC-UPV], Institute of Innovation and Knowledge Management
Universidad Politécnica de Valencia - Camino de Vera s/n
Ciudad Politécnica de la Innovación - Edificio 8E 5ª Planta
46022 - Valencia

[Outline

Abstract

Introduction

Sources and method

Selection of documents

Identification of open access articles

Variables

Results

OA to Spanish peer reviewed documents

Redundancy or repeated deposits

Subject area differences

Influence of commercial journals

Timeline (chronology) of deposits

Discussion

Conclusions]

Abstract

We examine open access to the Spanish scientific literature through investigation of a sample of peer reviewed articles in seven subject categories. Of the 28,259 papers published in 2000, 26.89 % were freely accessible with the share varying among disciplines. Articles in the Social and Behavioral Sciences were the most available for free. This disciplinary divide applies also to the strategies used to offer open access to documents. In Clinical Medicine, Life Sciences, Arts and Humanities and Social Sciences open access was mainly based on the publishers side while subject based repositories were dominant in Physical, Chemical and Earth Sciences and deposit in home pages was the preferred strategy in Engineering, Computing and Technology. Institutional and general repositories seem to play a minor role in providing free access to the Spanish peer reviewed literature. Papers published in commercial journals are less accessible than those that appear in non-commercial journals and we found overlaps in almost 20% of papers deposited. The fastest way to gain open access is to deposit in subject based repositories and the longest delays are related to deposits in home

* To whom correspondence should be addressed: cbenito@ingenio.upv.es

pages and specially to institutional repositories. Open access to Spanish peer reviewed articles is dominated by the passive mechanism of the gold road and the editorial strategy with self archiving practices in the minority and directed mainly towards subject based repositories and home page posting of the papers.

The resulting figures of this study could serve as a reference starting point in further study on the evolution of open access to the peer reviewed literature of Spain.

Introduction

Open access (OA) to the scientific literature is defined as the free, immediate and permanent online availability of the full texts of research and academic documents in digital form (Budapest Open Access Initiative, <http://www.soros.org/openaccess/read.shtml>). The OA principle has been in place since the early 1990s and based on the confluence of several motivations and initiatives.

The first contributing factor to OA was the previous existence of subject based repositories, constituted on the basis of a culture of document exchange among specialist communities. The Mathematical Physics Preprint Archive (mp_arc) and Arxiv, launched in 1991 were followed by Research Papers in Economics (RePEc) and CERN's E-print repository (CERN Document Server) both implemented in 1993. In the early 2000s two other repositories were established devoted to the Library and Information Science Literature: Digital Library of Information Science and Technology (DLIST) in 2002 and E-prints in Library and Information Science (E-LIS) in 2003.

A second factor, alongside increasing connectivity facilities and decreasing digital storage costs, was the development of free archiving software. The first such program was E-Prints, developed in the School of Electronics and Computer Science of the University of Southampton, UK, in 2000 which was followed by DSpace (MIT and Hewlett-Packard, 2002) and Flexible Extensible Digital Object Repository Architecture (FEDORA) publicly released in 2003. These are only the main software platforms used to organize and distribute digital objects. These technical developments were accompanied by the Open Archives Initiative and the Protocol for Metadata Harvesting (OAI-PMH) released on January 2001 to ensure system interoperability¹. The Internet and the web provide basic platforms for the individual distribution of research documents (e.g. through personal home pages or departmental servers). In contrast, archiving software allows for the organized compilation and subsequent diffusion of scholarly and research material in combination with crawling of their contents by Web search engines.

Accompanying these developments was the idea that public funding of research activities should imply a generalized availability to the resulting output, an idea that has gained support to become a common place in the OA culture. In fact, it can be traced in the series of declarations or initiatives issued in support of the OA movement since the seminal Budapest Open Access initiative. This idea is usually combined with a fourth argument in support of generalized availability of research papers, the so called journal affordability problem²: the progressive detachment between libraries' journal acquisition budgets and the costs of journal subscription.

Finally, the claim that OA maximise the visibility, usage and impact of the research papers freely available is a strong argument for it. Although several studies seem to confirm this generalized conviction, some recent investigations have begun to question it^{3,4}.

In line with the Budapest Initiative, strategies to increase access to scholarship and research documents have been summarized into two "roads" regularly labelled as the "green" and the "gold" roads to OA. The green road is based on authors' initiatives to self-archive their scientific production. The gold road relies mainly on publishers allowing access to their journal contents in various ways. Thus, initiatives from the editorial sector can be collectively considered as a passive mechanism that leads, with some degree of delay, to the availability of documents. There is, of course, some initiative when the authors follow the so called "open access option" in some journals. This option offer authors the possibility to have their articles immediately available for a fee. But a simple estimate reveals that this option is in minority: out of 16692 Spanish papers published in 2008 and recorded in PubMed, only 548 (3.28%) were published by BioMed Central, a publisher which has pioneered the "open access option" as its only publishing model.

But there are many means of gaining access to research literature: John Willinsky⁵ distinguishes up to ten models or formulae devoted to it, revealing a complex environment that has emerged as a result of the interrelations of these actions and initiatives.

Institutional repositories

An Institutional Repository (IR) comprises a digital collection of the academic and research output of a given community and the services that enable the storing and diffusion of that output. Its benefits are to ensure the long-term preservation of an institution's academic output; to increase its visibility and prestige, and act as an advertisement to attract funding sources, potential new faculty and students. For the individual, IRs provide a central archive of a researcher's work, increase its dissemination and thus, potentially, its impact on the research community, and can act as a full CV⁶. Institutional repositories can be considered along other issues as active strategies for providing free access to documents. They are active because they depend of the author's initiative to self archive and disseminate their works.

IRs are mostly dependent upon public organizations in terms of financial as well as human resources. This fact, the ambitious character of their goals and their difficulties in reaching a critical mass of deposits have led to the establishment of several certification criteria, auditing processes and evaluative frameworks. The recent review of Kim and Kim⁷ provides an overview of some of these initiatives although some others are ignored. Their main objective is "to develop diagnostic evaluation indicators to... pinpoint procedural weaknesses and... determine necessary customary solutions" for a consortium of university IRs. However, its framework contains only five (out of 29) indicators on content and thus their approach has been criticized for its mainly procedural or managerial perspective⁸. In one of the works not included in Kim and Kim's review, Dobratz and colleagues point out that evaluation comes after the conception, planning and realization stages of digital repositories implementation. However, their study also ignores some indicators of content, although these authors adopt a deliberate and explicit level of abstraction in the catalogue of criteria they propose⁹. Also not considered by Kim and Kim is the analysis of the Canadian repositories by Mary Westell, who claims that "the combination of input activity, usage and citation analysis will give the full picture of repository effectiveness"¹⁰. This is the model that governs most evaluation projects: see for example the works devoted to digital libraries and repositories in India¹¹ and Mexico¹² or the results of a survey of UK repository managers and users¹³.

All the studies referred to above seem to share an isolated view of IRs, both in respect to other complementary or alternative strategies and in relation with a more general vision of academic and research literature dynamics. In fact institutional repositories are the most recent OA provider and their role is far from clear.

The use of basic concepts from Ecology has been recently introduced in the field of digital repositories. The starting point is a recognition that "...there is a growing need to consider the interactions between repositories and between repositories and other systems" and that the main objective is to "inform the task of understanding and articulating the interactions between users, repositories, and services and the information environments in which they take place". This approach is in strong contrast to the isolated view in evaluative attempts. Fortunately, there are examples of a more systematic approach to the interactions taking place in the OA environment including a fruitful discussion of the relationship between extreme strategies towards the achievement of OA and some recent quantitative studies on the share of OA documents over total published documents in several research fields.

In studying the relationship between journals and repositories, Stephen Pinfield recognizes the existence of a potential for repositories and journals to interact with each other and proposes three main models for this interaction¹⁴. The key point made in his paper is the systemic approach and the acknowledgement that "Repositories and journals may interact in a future OA scholarly *communication and dissemination environment*" (emphasis added).

As for the quantitative studies, the first one recognizes that "There is thus an urgent need for reliable figures concerning the yearly volumes of journal publishing, and the share of the yearly volume which is available as open access via different channels" and argues "We also think that the ratio of open access articles to the overall number of articles published is a much more important indicator of the growing importance of OA than the number of OA titles compared to the number of titles in general"¹⁵. The second work states as its main purpose "To more completely and comprehensively capture the complex structure of OA in biomedicine"¹⁶. While Björk and collaborators make an estimate of the scientific literature as a whole and Matsubayashi et al.¹⁶ limit their study to the biomedical field, Altelman selects four subject categories. Her main purpose is to determine whether scholarly articles from disciplines with varying rates of open access adoption have a greater research impact if the articles are freely

available online than if they are not¹⁷ but she also offers a detailed account of the strategies followed for achieving free diffusion of published papers.

Spanish repositories

The first Spanish repository was to accommodate and preserve theses and dissertations from 20 Spanish universities. The idea was proposed in 1999 and the repository set up in 2001¹⁸. Results of some recent works find that the number of repositories has since grown to 56 OAI-PMH repositories in July 2008¹⁸ or 47 in December same year¹⁹ depending on how a repository is defined and revealing some confusion between repositories and digital libraries projects in some environments. These figures increase as additional criteria are handled, for example the participation of Spanish groups in international subject based or general repositories²⁰. The works referenced are good examples of the three current approaches to OA in the Spanish context. The first is promotional and somewhat didactic in nature. It is aimed at promoting the advantages of OA among scholars and researchers and promulgating the available strategies and tools within the library community. The second approach follows the surveys of the Digital Repository Infrastructure Vision for European Research (DRIVER) at a national level and makes repeated inventories of existing initiatives. The activities of the national "Open Access to Science" working group (<http://www.accesoabierto.net/>) are demonstrative of these approaches. The work of Peset and Ferrer²⁰ is so far the only example of the third approach to OA in Spain. It relates figures on OA activity with sociodemographic and economic data. Their main assumption seems to be that implementation of repositories is a natural outcome of achievement of a certain level of economic development. However, to put OA in a more immediate context, a more reasonable approach would be to relate the development of repositories with academic figures or other higher education statistics instead of using macroeconomic figures as a term of comparison.

Objectives

To obtain an accurate picture of Open Access in relation to Spanish research papers it is not enough to answer to the question of how many of them are freely available. A more systematic insight into the Spanish environment requires an investigation of the strategies adopted towards depositing materials and how they interrelate. Also, the existence (if any) of some sort of "disciplinary divide" among the various research cultures needs to be determined. It is necessary also to investigate whether co-authorship, especially with foreign colleagues, has some influence on where the papers are deposited. Another issue that deserves some study is the relationship between the commercial nature of the journals and the OA to their content. Finally, it is important to determine when an article becomes accessible and whether different strategies led to its faster diffusion. Besides these research questions, a secondary objective of this paper is to obtain background data that will form the foundation and quantitative reference for future studies of the same environment.

In the succeeding sections, we describe our sampling of a set of Spanish peer reviewed papers. Next, we identify which are available and compare the features of the available and not available portions of the Spanish research literature. We discuss our findings and provide some conclusions.

Sources and method

Selection of documents

We used Thomson ISI Current Contents Connect (CCC) to retrieve documents with at least one contributing Spanish author for the year 2000 published in journals indexed by ISI. This multidisciplinary current awareness resource provides access to complete bibliographic information for more than 8,000 of the world's leading scholarly journals and more than 2,000 books. Its contents are organized in seven subject editions: 1) Agriculture, Biology and Environmental Sciences; 2) Arts and Humanities; 3) Clinical Medicine; 4) Engineering, Computing and Technology; 5) Life Science; 6) Physical, Chemical and Earth Sciences, and 7) Social and Behavioural Sciences.

In order to extract these articles categorised, we ran seven searches, one for each CCC subject edition. Each of the searches was based on the following parameters: AD

(address/institution) = Spain and Publication year= 2000. No restrictions were imposed in relation to language or document type. CCC was chosen for its extensive coverage of Spanish journals and the eventual inclusion of other document types (such as book chapters). The seven subject area coverage allows us to investigate the thematic differences in the accessibility to the peer reviewed documents. The year 2000 is critical: the appearance of the first academic repositories in 2001 marked a kind of starting point to the culture of OA in Spain. Also data resulting from the present study provides a point of departure and a basis for comparison for future research. In addition to these facts, it is convenient to keep in mind that this study focus in what has been happening to papers published in 2000 in the eight years following their appearance.

We retrieved overall 28,259 documents. For each ISI category, a random sample consisting of 10% of the total number of articles was selected, thus the number of documents analyzed was 2,856.

Identification of open access articles

We conducted searches in Google and Google Scholar based on article title to find the availability of the free full-text versions of the target articles. In some cases, titles were edited or shortened to avoid problems related to punctuation in titles or transliteration of certain characters. Either PDF, HTML or other versions of full texts were considered successful hits. Cases where documents were located in repositories with access restrictions (e.g. corporate) and cases where the metadata were accessible but not the full text, were considered as not publicly available.

To avoid the possibility of journals included in packages of corporate subscriptions causing false hits, searches were performed from outside the corporate intranet.

Google was chosen because of its large coverage of the web and its popularity. It was considered that most potential readers will use Google as their primary resource for locating on-line versions of scientific papers. We discarded other resources and aggregators, such as Scientific Commons (www.scientificcommons.org) based on their lack of inclusion of personal and institutional home pages.

Variables

In addition to bibliographic data, we studied six variables for every sampled document: number of authors, number of institutions, number of Spanish institutions, share (percentage) of foreign contributing institutions, nature of the publisher and date of publication.

In some cases we were able to determine the exact date of publication. When this information was not available, we estimated it as: 1) the 15th day of the month for monthly journals, and 2) the first day of the second month for bimonthly journals.

Where full content of papers were available, the URL and date of 'deposit' (date when the document became publicly available regardless of the OA strategy used) were recorded. If the full text was found in multiple locations, all URLs were considered in order to study overlaps. Based on examination of URLs, the OA option was identified and recorded to enable analysis of the number of documents that are available via each OA strategy. The variants identified were: home page posting, institutional repository, subject based repository, editorial posting and general repository. General repositories, also termed "journal platform or portal site"¹⁶ or "other repository"¹⁷, have a regional or national basis and are associated with the dissemination of the technical and scientific literature published in developing countries, thereby increasing the visibility of literature that otherwise would be accessible only within their national borders²¹. In the cases of home page posting and institutional repositories, we identified country of deposit.

Comparative analysis of publication and 'deposit' date allowed us to study the interval between the publication date and the date when the document was made publicly available. Finally, the use of ISI areas allowed us to identify potential differences in the impact of OA in different areas of knowledge.

Results

OA to Spanish peer reviewed documents

Overall, results indicate that almost 27 % of the peer reviewed Spanish articles published in 2000 are openly accessible in the eight years period following their publication date. Table 1 shows that the highest proportion of articles is in to the Social and Behavioural Sciences subject category (56.31 %) while the minimum percentage (10.73 %) is found among Engineering, Computing and Technology articles. 87.71 % of the sampled documents are articles, 5.71 % are letters and 2.31 % review documents.

[Table 1 around here, please]

For the 768 accessible documents we have identified 918 deposits. Table 2 provides a cross tabulation of deposits in every subject area and for every access strategy, where values represent percentages for the number of deposits in each subject area.

Six different strategies to OA have been identified: Editorial posting, Home page, Subject Based Repository, Institutional Repository, General Repository and Other.

The last row in Table 2, which is the proportion of deposits via each strategy, presents some sort of hierarchy among the options followed: in general terms, the preferred one is the editorial strategy. Home pages and subject based repositories fall behind at a considerable distance and institutional and general repositories are minor options in OA pursuit.

[Table 2 around here, please]

The main subject based repositories, identified through the examination of the URLs of the accessible documents are Arxiv (104 out of 175 cases or 59.43 %) and PubMed Central (24 %). Institutional repositories were located mainly in Spain (50 % of deposits that follow this strategy) and USA (12 %) followed by Germany, Switzerland, UK and Russia.

Editorial posting is the preferred option both in Life Sciences and Clinical Medicine categories while posting papers in home pages is a rarely used strategy in these two categories although it is very important in Humanities and Engineering and also relevant in other fields. General repositories play a great role in Social and Behavioural Sciences. Subject based repositories are the main strategy in Physics and Engineering and institutional repositories play a minor role in all fields except Medicine and Humanities. Ten cases (1.06 %) were discarded because they used some other strategy (e.g. Google books) to provide accessibility. The differences among strategies in subject categories are significant (Chi square = 686.699 with 30 df and $p < 0.001$).

Commercial versus non-commercial journals

Articles were published in 1246 different journals from 33 different countries. USA, UK and Netherlands were the countries of publication of more than three quarters (76.58%) of the papers. Journals were categorized into two groups: those published by commercial companies and those produced by academic institutions (e.g. the University of Miami publishes the *Journal of Interamerican Studies and World Affairs*) or learned and scientific societies (*Texas Heart Institute Journal* which is published by the homonym institution). The source for this distinction was the field PU (publisher) in the data schema of the source database. Table 3 presents a cross tabulation for the influence of the type of journal (commercial or not) on the accessibility of the articles published in it. For this analysis, all sampled records (N=2856) were used. And the Table provides absolute values and percentages on total cases.

[Table 3 around here, please]

Chi square value (181.68 with 1 df and $p < 0.01$) suggests that there is some relationship between the journals publisher type and the accessibility of the articles. Risk estimate is 1.44 (confidence interval = 1.35-1.53) for not accessible articles published in commercial journals while the ratio for accessible articles is 0.45 (0.4-0.51). Thus, the relative risk difference shows that articles published in commercial journals are accessible only in 68 % of the cases compared to those published in non-commercial journals which would be 100 % freely available.

Redundancy (multiple posting)

One given article may be available at several places: a publisher's site and a personal home place, for instance. From the set of 768 articles, 118 were accessible at more than one place. Thus, redundancy or proportion of multiple posting is 16.4 % overall. The maximum value, applying to two cases, is six. There are strong differences among the subject categories: in Clinical Medicine there is not redundancy, while in Social and Behavioural Sciences the proportion of multiple posting is 58.62 %. ANOVA shows significant differences among the disciplinary groups ($F = 18.94$ with 6 df and $p < 0.01$) and post hoc analysis reveals three groups with the category of Social and Behavioural Sciences apart from the rest and responsible for the differences (Tukey's HSD = 1,72 for $\alpha = 0.05$).

The redundancy is not associated with either the number of contributing authors per paper or the number of participating institutions per paper. We find a slightly positive significant correlation (Pearson CC = 0.112 with $p < 0.01$) between redundancy and participation of foreign institutions in the publications, expressed as a percentage of non-Spanish institutions over the total collaborating institutions. In fact, a two-tailed T test shows significant differences between accessible and not accessible articles only for the variable foreign participating institutions ($t = 6.131$ with 1262 df and $p < 0.01$) while number of contributing authors and total number of institutions show no differences between these two groups.

The multiple posting follows several strategies as depicted in Figure 1. The vertices in the network represent the strategies followed and their diameters are proportional to the number of deposits made via that strategy. The strength of the edges denotes the number of cases where the linked strategies were applied to the same documents.

[Figure 1 around here, please]

The role of the editorial strategy, both in terms of both frequency and centrality, as the main path to OA to articles is obvious from the network. Institutional repositories are associated only with home pages posting, with weak ties to subject based repositories. In contrast, there is a strong relationship between home pages posting and subject based repositories.

Chronology

Time lapse between publication and deposit of a given paper could be determined for 695 articles. On average, an article is freely accessible 1.35 ± 1.93 years after its formal publication. However, time lapses varied depending on the access strategy used. The box plot in Figure 2 illustrates these differences and shows that the quickest way to achieve OA is via subject based repositories (0.007 ± 1.15 years, less than a month) while the longest delays are associated with home pages posting (4.49 ± 2.4) and specially to institutional repositories (5.5 ± 2 years).

[Figure 2 around here, please]

104 articles (14.96 %) were accessible at the time of publication and a very similar proportion ($N=102$) was deposited prior to publication.

Figure 2 also illustrates that half of the articles deposited in general repositories were immediately accessible. Agreements between these repositories and journal editors could be responsible for this effect.

In the case of deposit prior to publication, it becomes evident that what is deposited is an earlier version of the version of the record: the author's original, the submitted manuscript or the accepted manuscript, following the definitions by NISO/ALPSP²².

Discussion

Over a quarter of our sample of peer reviewed papers published by Spanish researchers in 2000 are accessible eight years after publication. It should be noted that the length of this temporal window makes this main quantitative finding difficult to compare with the results from similar studies. Björk and collaborators used 2006 as their target year and made searches in winter 2007¹⁵. So, their estimate of a 19.4 % of articles being freely accessible corresponds to less than 12 months. A similar time frame is used in the work by Matsubayashi et al. who studied the accessibility of PubMed articles published in the first nine months of 2005 and conducted the searches for full texts in January 2006 resulting in a share of OA papers of 26.6 %¹⁶.

If the figures of the present study are interpreted as meaning that it takes eight years to achieve a similar proportion (26.89 %) of freely accessible papers, then the logical conclusion is that Spanish papers are less accessible than the average, taking into account that the 19.4 % of the general study by Björk was achieved only a few months after publication. However, if the share of free Clinical Medicine papers in the Spanish sample (38.15 %) is compared with the general figure and with the proportion of OA papers from Spanish authors (24.7%) in Matsubayashi's study then it could be concluded that Spanish clinical articles are far more accessible than the average. Again, differences in time window used make it difficult to compare figures. Any interpretation of these raw quantitative data therefore, would be controversial and no clear conclusion can be drawn.

In any case, as it has been stated in the objectives of this study, our aim is not (or not only) to find the share of OA in a set of peer reviewed papers published in 2000, but to study which strategies have authors and publishers follow to disseminate their content openly. It is arguable that the emergence of Spanish repositories has led to some changes in current self archiving practices of authors. In support of this assumption we can contribute an additional estimate: The number of deposits in Spanish repositories (available in <http://roar.eprints.org/>) has grown at an annual rate of 40.08% in the period 2000-2009. For the same period, the corresponding figure for Spanish papers available through PubMed Central is 24% . Thus, it is obvious that any future estimation will lead to a distinct landscape, both from a quantitative as well as qualitative point of view.

Regarding the strategies for OA, the results of our study are consistent with those in previous work. In fact, the study on open access in the biomedical field reports that access through journal web sites (the strategy termed as editorial posting in our study) is predominant. PubMed Central is ranked second by Matsubayashi, as is the strategy of subject based repositories in the present study; in both studies, home page posting and institutional repositories are least used. The ranking was the same for the biomedical papers originated in Spain. In the general study by Björk et al. the only difference was the reduced paper of subject based repositories, which account in their estimate only for a 3.3 %. On the other hand, these authors carry out a separate estimation to determine that the number of articles directly accessible through the editorial strategy is 4.6% of all articles published in 2006. They add another 3.5% for papers accessible after some embargo period (delayed OA) or through author charges (hybrid OA) to reach a final share of 8.1% . These results contrast strongly with our findings and the fact that the editorial way (direct or indirect) is the prevailing strategy for Spanish peer reviewed papers to achieve OA.

The influence of the journal's publishing model (commercial or not) is investigated only by Matsubayashi et al¹⁶ who find that the total category of society publishers accounted for 58.9% of the OA articles, however they provide no further statistical analysis of their data. However their share coincides with the results of the present study and indicates a relationship between the non-commercial nature of the journal and the probability of an article to be freely accessible.

Repeated deposits are treated by Björk et al.¹⁵ who refer to them as "parallel publishing". Surprisingly, they find no overlaps where an article appears on a publisher's web site and is deposited in a repository. This is obviously an erroneous finding if we think, for example, of the articles hosted on journals web sites and later deposited, as part of an editor's policy, in PubMed Central.

It is striking that deposit in subject based repositories provide prompt and faster access than posting on the author's home pages. This might be due to the weight of the cultural component associated to the common use of subject based repositories. The delay of availability through institutional repositories may be affected by the lack of popularity of Spanish repositories due to their more recent emergence (since 2001). However, the majority of deposits that follow this strategy are in Spanish institutional repositories.

Finally, to consider deposits in institutional repositories to be the result of self archiving initiatives by the authors and hence, to view institutional repositories as an active mechanism is dubious. There would seem to be a generalized practice of batch uploading based on the time plots of records additions. See, for example, the timeline of publication activity of the most populated Spanish institutional repository with loading peaks at certain weekends (<http://en.scientificcommons.org/repository/digitalsic>).

Conclusions

If the green and gold routes to OA are interpreted (as suggested in the introductory paragraphs) as being active as opposed to passive diffusion mechanisms, thus it must be concluded that overall, Spanish peer reviewed literature is driven to OA by a major passive mechanism through the publisher website or editorial strategy. However, there is a disciplinary divide in terms of OA with a greater presence of active self archiving strategies in the Social Sciences, the Humanities and in the field of Agricultural, Biological and Environmental Sciences. This disciplinary divide also leads to a preferred strategy in each subject category or specialty culture.

Gold road and subject based repositories by far outweigh the other strategies for providing free and rapid dissemination of research results and despite the commercial nature of many of these journals, immediate or delayed access to their contents is assured via publishers' websites.

Behind the research questions posed at the introductory section of this paper relies another more general one: Is it worth establishing institutional repositories?. The answer will depend on the functions attributed to them. Long term preservation and especially the institutional visibility are major arguments for their implementation, however the role of institutional repositories as a unique mechanism to gain access to the research literature is less clear.

Further research is needed to observe the effects of the emergence of Spanish repositories on author's attitudes. But success or failure of institutional repositories should not be based solely on a growing evolution in the number of deposits. Attention should also be paid to the shift among active strategies, mainly to a decrease in home page posting in favor of institutional deposits. Thus, further research should focus on a systematic investigation of all OA options and not only on institutional repositories' evolution.

Acknowledgements

This work was partially supported by the research grant "Acceso abierto a la producción científica en España" project number CSO2008-05525-C02-01/SOC (Spanish Ministry of Science and Innovation). The authors would like to thank Elena Mas (School of Mathematics, University of Valencia) for her assistance in statistical testing and Dr. María Francisca Abad (Department History of Science and Documentation, University of Valencia) for her critical review of the draft manuscript.

References

1. Lagoze, C. and Sompel, H.D. (2003) The making of the Open Archives Initiative Protocol for Metadata Harvesting. *Library Hi Tech* **21**, 118-128.
2. Harnad, S. ; Brody, T. ; Vallieres, F. et al. (2008) The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review* **34**, 36-40.
<http://dx.doi.org/10.1016/j.serrev.2007.12.005>
3. Craig, I.D., Plume, A.M., McVeigh, M.E., Pringle, J. and Amin, M. (2007) Do open access articles have greater citation impact?: A critical review of the literature. *Journal of Informetrics* **1**, 239-248. <http://dx.doi.org/10.1016/j.joi.2007.04.001> Also in http://www.publishingresearch.net/Citations-SummaryPaper3_000.pdf.pdf (accessed 10 November 2009).
4. Davis, P.M., Lewenstein, B.V., Simon, D.H., Booth, J.G. and Connolly, M.J.L. (2008) Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ* **337**,

- a568. http://www.bmj.com/cgi/content/full/337/jul31_1/a568 (accessed 10 November 2009).
5. Willinsky, J. *The Access Principle: The Case for Open Access to Research and Scholarship*. (MIT Press: Cambridge, MA, 2005).
<http://mitpress.mit.edu/catalog/item/ebook.asp?tttype=2&tid=10611> (accessed 10 May 2009)
 6. Prosser, D. (2003) Institutional repositories and Open Access: The future of scholarly communication. *Information Services and Use* **23**, 167-170.
 7. Kim, Y. and Kim, H. (2008) Development and validation of evaluation indicators for a consortium of institutional repositories: A case study of dCollection. *Journal of the American Society for Information Science and Technology* **59**, 1282-1294.
 8. Jantz, R. and Wilson, M. (2008) Institutional Repositories: Faculty Deposits, Marketing, and the Reform of Scholarly Communication. *Journal of Academic Librarianship* **34**, 186-195.
 9. Dobratz, S. and Matthaei, B. (2003) Open Archives Activities and Experiences in Europe: An Overview by the Open Archives Forum. *D-Lib Magazine* **9**.
<http://www.dlib.org/dlib/january03/dobratz/o1dobratz.html> (accessed 11 May 2009)
 10. Westell, M. (2006) Institutional repositories: proposed indicators of success. *Library Hi Tech* **24**, 211-226.
 11. Mittal, R. and Mahesh, G. (2008) Digital libraries and repositories in India: An evaluative study. *Program* **42**, 286-302.
 12. Galina, I. and Giménez, J. Overview of the development of open access journals and repositories in Mexico. *ElPub 2008 - Conference on Electronic Publishing* 280-287 (25-27 June 2008). http://elpub.scix.net/cgi-bin/works/Show?280_elpub2008 (accessed 5 October 2009)
 13. Zuccala, A., Oppenheim, C. and Dhiensa, R. (2008) Managing and evaluating digital repositories. *Information Research* **13**. <http://informationr.net/ir/13-1/paper333.html> (accessed 9 September 2009)
 14. Pinfield, S. (2009) Journals and repositories: an evolving relationship? *Learned Publishing* **22**, 165-175.
 15. Björk, B., Roos, A. and Lauri, M. Global Annual Volume of Peer Reviewed Scholarly Articles and the Share Available Via Different Open Access Options. *ELPUB 2008* 287 (25-27 June 2008). http://elpub.scix.net/cgi-bin/works/Show?id=178_elpub2008 (accessed 11 May 2009)
 16. Matsubayashi, M. et al. Status of open access in the biomedical field in 2005. *J Med Libr Assoc.* **97**, 4-11 (2009).
 17. Antelman, K. (2004) Do Open-Access Articles Have a Greater Research Impact? *College and Research Libraries* **65**, 372-382.
 18. Alonso, J., Subirats, I. and Martínez Conde, M. *Informe APEI sobre acceso abierto*. (Asociación Profesional de Especialistas en Información: Gijón, 2008).en
<http://eprints.rclis.org/archive/00015107/01/informeapeiaccesoabierto.pdf>
 19. Melero, R., Abadal, E., Abad, M. and Rodríguez Gairín, J. *Situación de los repositorios institucionales en España: informe 2009*. (Grupo de investigación Acceso Abierto a la Ciencia: 2009). <http://www.accesoabierto.net/node/28> (accessed 1 May 2009)
 20. Peset, F. and Ferrer, A. (2008) Implantación de la Open Archives Initiative en España. *Information Research* **13**, paper 385. <http://informationr.net/ir/13-4/paper385.html> (accessed 12 May 2009)
 21. Marcondes, C. and Sayao, L. (2003) The SciELO Brazilian Scientific Journal Gateway and Open Archives. A Report on the Development of the SciELO-Open Archives Data Provider Server. *D-Lib Magazine* **9**.
<http://www.dlib.org/dlib/march03/marcondes/o3marcondes.html> (accessed 27 August 2009).
 22. Morgan, C. (2008) Journal article version nomenclature: the NISO/ALPSP recommendations. *Learned Publishing*, **21**, 273-277.
DOI: <http://dx.doi.org/10.1087/095315108X356699>
 23. Robertson, R.J. ; Mahey, M. and Allison, J. An ecological approach to repository and service interactions Version 1.2. JIS. CCETIS, UKOLN and University of York, March 2008.
<http://www.ukoln.ac.uk/repositories/digirep/images/a/a5/Introductoryecology.pdf> (accessed May 2009).

Table 1
Open access articles per subject category

Subject area	Published docs	Sampled docs	OA docs	OA %
ABES	3719	382	118	30.89
AH	599	52	9	17.31
CM	4633	464	177	38.15
ECT	2677	289	31	10.73
LS	7550	780	196	25.13
PCES	8094	794	179	22.54
SBS	987	103	58	56.31
Total	28259	2856	768	26.89

ABES: Agriculture, Biology and Environmental Sciences. AH: Arts and Humanities. CM: Clinical Medicine. ECT: Engineering, Computing and Technology. LS: Life Sciences. PCES: Physical, Chemical and Earth Sciences. SBS: Social and Behavioural Sciences.

Table 2
Strategy to OA followed per category*

Subject area	Editor	Home	GenRep	InsRep	SBRep
ABES	1.14	28.95	3.95	11.18	2.63
AH	44.44	44.44	0.00	0.00	11.11
CM	92.47	0.00	0.54	0.00	6.99
ECT	2.29	46.67	2.22	6.67	37.78
LS	80.34	3.42	1.28	2.99	11.97
PCES	15.38	23.50	0.85	8.97	51.28
SBS	37.62	26.73	27.72	2.97	1.98
General	53.49	16.55	4.27	5.31	19.25

* All figures in percentage.

Editor: access via publisher. Home: posting in home pages. GenRep: General repository. InsRep: Institutional repository. SBRep: Subject-based repository.

Table 3
Contingency analysis for accessibility of papers and commercial or institutional nature of journals

		Access		Total
		Accessible	Not accessible	
Publisher	Commercial	Count	391	1951
		% of Total	13.69	68.31
	No commercial	Count	401	905
		% of Total	14.04	31.69
Total		Count	792	2856
		% of Total	27.73	100

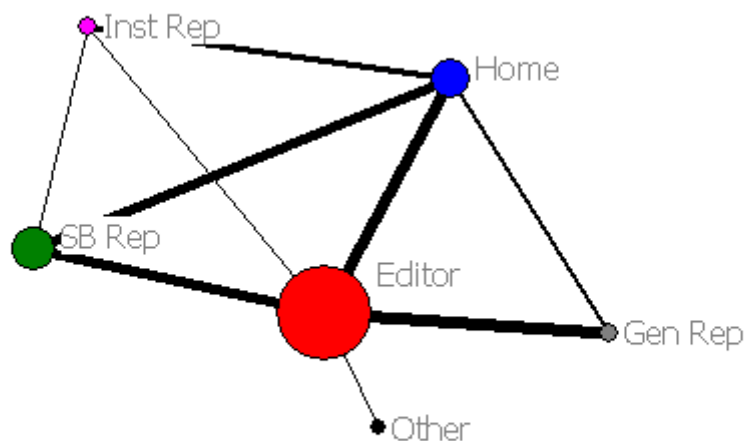


Figure 1. Relationships among the strategies for OA based on repeated deposit of the same articles.

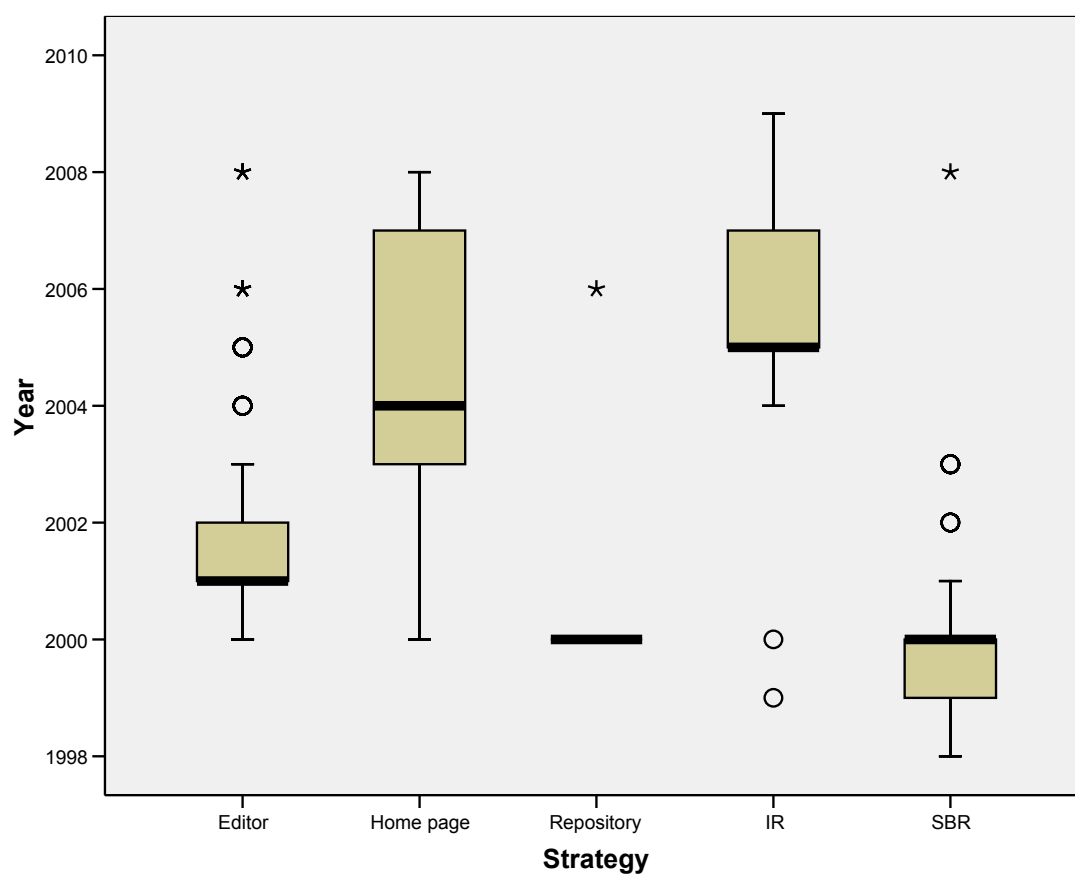


Figure 2. Time lapse between publication and deposit of the articles via the strategies identified.