Institutional Repositories

Chapter in Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics

Maria-Soledad Ramírez-Montoya, Héctor G. Ceballos
Tecnológico de Monterrey

Introduction

In the context of the Open Access movement, repositories provide a window for scientific and academic communities to preserve and share knowledge. Glassman & Kang (2016) stress the importance of educational institutions developing technologies that support an open educational system. Accordingly, Ramírez (2015) mentions that open access is an opportunity to improve the transfer and dissemination of knowledge and discuss the relevance of repositories as a space to accommodate and recover scientific and educational production. In this sense, institutional repositories constitute a support for the open dissemination of knowledge, but also bring with them challenges that must be worked out in organizational cultures.

A repository is a technological platform of open access to knowledge, which is directed to the storage, preservation and diffusion of the production generated in the institutions. MacIntyre & Jones (2016) emphasize the support of institutional repositories for research and the visibility they can bestow on institutions. Fontes Ferreira & Souza de Silva (2015), emphasize the idea that the more academic tools people have, with the support of repositories, the better academic level they will reach, as well as a modern way of studying. Empirical studies have provided evidence of the positive role of repositories; for example, Koler-Povh, Mikoš & Turk (2014) presented a study on the institutional repository of the University of Ljubljana in early 2011, where they found that 89% of its visitors came from other institutions. This shows how important are institutional repositories to publish information and how communication between institutions and universities contribute to open education.

In contrast to the advantages provided by institutional repositories, it is also necessary to state the challenges existing in the environment of the open education movement. Davis, Carr, Hey, Howard, Millard, Morris & White (2010) present the problem of the poor response to open educational models, despite the great investment in infrastructure that is taking place. Cragin, Palmer, Carlson & Witt (2010) analyze the cognitive processes that lead authors to share or not their work, based on the study of their cognitive processes and the "open sharing" culture. Another challenge lies in the registration of resources so that knowledge can be "discovered" and used; for example, González (2016) found in his study that meta-information in articles of institutional repositories may be incomplete and that could generate difficulties to access this information. Hence the importance that authors and library staff should consider methods to improve the indexing of articles in order to be discovered.

The motivations for institutions to have a repository can be very varied. Being a set of centralized Web Services, the objective of an institutional repository is to organize, manage, preserve and disseminate digital materials, mainly scientific and academic production created by an institution and its members. In this way, institutional repositories provide institutions with the possibility of improving their position in rankings, ensuring the preservation of their organizational memory, gaining visibility and presence on the Web, increasing the impact (citation) of authors, encouraging scientific collaboration (internationalization), supporting innovation (research projects) and, recently, being able to respond to national policies of open sharing of the production generated in publicly financed projects.

The Open Access movement in the previous decade and institutional repositories developed by universities and academic libraries as part of this movement, have challenged the traditional model of the school communication system. Researchers such as Cullen & Chawner (2011) examined the growth of institutional repositories alongside the open access movement and reported the findings of a national survey of academics highlighting the conflict between the principles and rewards of the traditional school communication system and the benefits of the open access movement.

In different latitudes, the motivation to have institutional repositories is evident. Figure 1 shows the distribution by country of the 2,824 institutional repositories listed by February 7th, 2017 in the Directory of Open Access Repositories (OpenDOAR)[1]. As can be seen, United States is the country with more repositories listed (373), nevertheless, Europe together has more than three times this number of repositories.
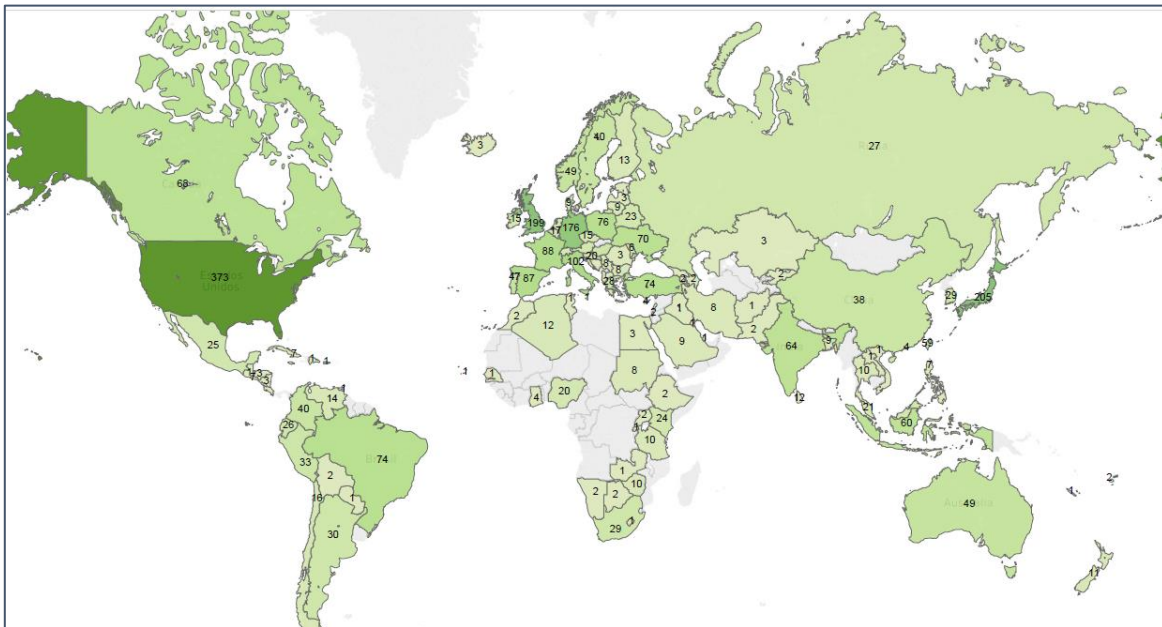

Figure 1. Number of repositories per Country (OpenDOAR, Feb. 7, 2017).

In repositories is stored scientific and academic production that is of interest for institutions. Scientific productions such as papers, articles, chapters, books, journals, datasets, are highly valued documents to share in the research community. Also academic productions such as cases, learning objects, videos, notes, presentations, are useful for the teaching and student community. This production must be cataloged in such a way that it can be found to be used. Shukair, Loutas, Peristeras and Sklar (2013), mention that in order to facilitate the discovery, access and reuse of resources, semantic interoperability must be considered, i.e. a thesaurus of concepts that describe the information and a new model of Metadata to identify the types of reusable resources, which facilitate their discovery and ensure a minimum consistency that is directly related to the definition of a conceptual architecture for a federated resource deposit.

Open Access

The Open Access movement aims at the online availability of scientific products free of cost and restrictions on the use of the information contained within. In the dissemination of scientific

---

[1] The Directory of Open Access Repositories – OpenDOAR. http://www.opendoar.org/

production, two main routes for Open Access (OA) can be observed: to publish in an open access journal (Gold OA) or to archive copies of the article in a repository or web site (Green OA). Björk, Laakso, Welling & Paetau (2013), present information about what is Green OA and its differences with Gold OA and analyze the legal considerations of the publications and the ease of its administration. They conclude that the Green OA technical infrastructure is becoming more robust due to the increase in the number of institutional repositories. Also Johnson, Fosci & Pinfield (2015) mention the differences between the two routes, which have different purposes, and also agree that the interest in Green OA is greater, since it is focused on institutional repositories of Universities. In addition, they mention aspects of the costs and of how each of them relates in the society.

Digital preservation and copyright control in institutional repositories must be strengthened to increase participation. Kim (2011), through a study demonstrated the motivation given to faculty staff to adopt Green Open Access, where the faculties were motivated by both preservation and copyright. Teachers contributed to institutional repositories to make their material extensively accessible and not lose the benefits of open access. Another very positive motivation for authors is the possibility of citation that gives them the open publication. Gargouri, Hajjem, Lariviere, Gingras, Carr, Brody & Harnad (2010) demonstrated that articles written through Green Open Access are cited more frequently than articles requiring subscription. The above is due to the quality advantage of these items as users have more freedom to choose which articles to cite.

As we see benefits there are also challenges faced by Green Open Access. Antonczak, Keegan & Cochrane (2016) claim that there are obstacles to creating repositories. The people who produce and create these repositories are creative, but they could create such repositories for purposes other than to benefit society. On the other hand, Wallace (2011) presents the development of PEER (Publishing and the Ecology of European Research), in which they have elaborated an observatory with more than forty-four thousand manuscripts in Green Open Access. One of the challenges of this development was the difference in manuscript types and metadata formats. The author focuses on the Green Open Access based on a study in the UK, which also compares the attitudes and behaviors of Open Access authors and users and the challenges of developing PEER's Green Open Access. Another challenge lies in the number of possibilities for open publishing.

Studies related to publication in repositories have also been carried out. Dawson & Yang (2016) did a study of the Registry of Open Access Repositories (ROAR) database to find trends in how major institutional repositories handle copyright and Open Access. The authors found that there is no clear way for scientific publications to achieve OA status and suggest that if you want to start an IR program, libraries need to be involved with the faculty to help them obtain copyright permits. It is necessary to educate faculty staff about copyright, as this is one of the main barriers to the growth of repositories. It is recommended that librarians become experts on the subject of copyright to assist in the process.

Rankings

The evaluation of repositories have generated the possibility of grouping them according to their characteristics. One of the most recognized rankings is Webometrics, which arose in the Superior Center of Scientific Investigation (CSIC). Aguillo, Ortega, Fernández, and Utrilla (2010) present Webometrics and evaluate the repositories according to the following characteristics:

a) Listed in OpenDOAR

b) Being institutional or thematic

c) Contains scientific articles

d) Included in the domain of the Institution

e) Each file is measured according to size in number of pages, should be PDF file

f) Must be found in google scholar and must have visibility in external links.

Another ranking focuses on success characteristics of repositories oriented to scientific work. Marcial & Hemminger (2010) studied a sample of 100 Scientific Data Repositories (SDRs) and generated different segments or classifications of them. The characteristics of the SDRs were explored to identify their role in determining groups and their relation to the success of the group: if it received funding support, if the support came from different sponsors, the size of the SDR and if there was a preservation policy.

Platforms and protocols

Open platforms are multiple and this has increased the possibilities for repositories. Pinfield, Salter, Bath, Hubbard, Millington, Anders & Hussain (2014) provide an analysis of different platforms from 2005 to 2012 in different parts of the world where they reviewed the growth of open access repositories, using data from the OpenDOAR project, as well as the growth of type proposals in each country. It shows the map of institutional repositories, their development and operation. The research focused mainly on North America, Western Europe and Australasia. Since 2010, there has been growth in repositories in East Asia, South America and Eastern Europe. Globally, repositories are predominantly institutional, multidisciplinary and in English. Typically they use open access software, but they have some problems with licensing. Tzoc (2016), on the other hand, carried out a study to see the platforms of institutional repositories most used in universities of the United States. The results of the 67 institutions taken into account for the study were: DSpace, CONTENTdm, Islandora and NC DOCKS.

Figure 2 shows the main platforms used by the 2,824 institutional repositories listed in OpenDOAR by February 7th, 2017. As can be seen, DSpace is the platform mainly used (48%), followed by EPrints (14%) and Digital Commons (5%), meanwhile 27% of the remaining repositories use diverse platforms.
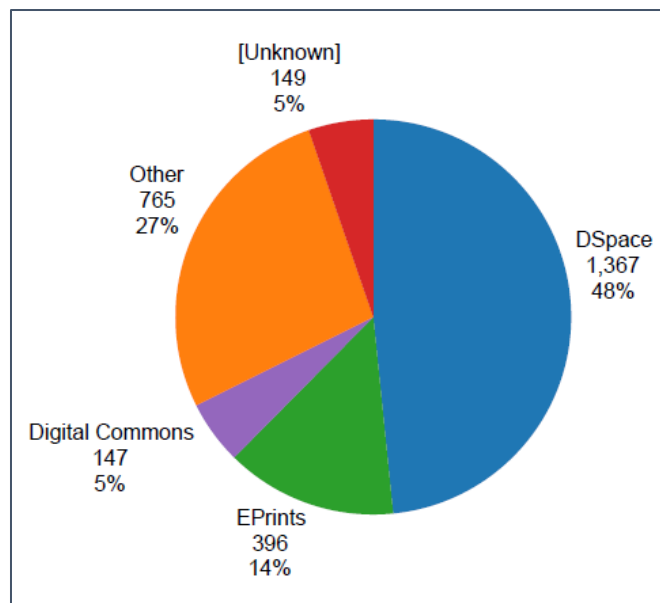

Figure 2. Institutional Repositories by platform (OpenDOAR, Feb. 7th, 2017).

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an automatic mechanism for transferring information between repositories[2]. Through this protocol, repositories become *data providers* by exposing structured metadata describing resources arranged in collections. Other repositories, called *service providers*, use OAI-PMH service requests to harvest that metadata.

Figure 3 shows the percentage of repositories supporting the protocol OAI-PMH (blue bars) and the percentage of resources hosted in those repositories, as listed in OpenDOAR by February 7th, 2017. As can be seen, despite 70% of institutional repositories support this protocol, less than 45% of the resources can be accessed through it.
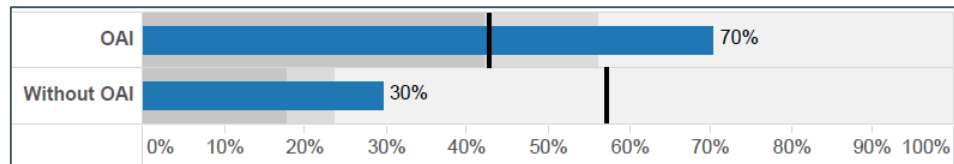

Figure 3. Repositories supporting the protocol OAI-PMH (OpenDOAR, Feb. 7th, 2017).

Metadata vocabularies used for describing resources is a key aspect for linking and visibility of repositories. Arlitsch & O'Brien (2012) seek to prove a theory that by transforming metadata schemas of institutional repositories will lead to an increase in Google Scholar indexing. The report indicates that repositories using Google Scholar's recommended data schemas and those expressed in HTML "meta tags" have a much better indexing rate. The ease with which search engine crawlers can navigate a repository affects the indexing rate as well. The importance of this research is that the lack of visibility in Google Scholar, will limit the visibility of the resources stored in these repositories, preventing a more important role in the increase on the number of citations.

Challenges for Institutional Repositories

Limitations of platforms and protocols for repositories are present in different stages. Way (2010) shows a problem of archiving articles in repositories, he conducted a study using Google Scholar to examine the open access availability of the *Library and Information Science* (LIS) and examined whether Google Scholar was able to find the links to the full texts in case there were open access versions of articles. It was found that article archiving is not commonly practiced but Google Scholar is an effective tool for finding LIS articles. Laakso & Björk (2016) describe how different methods and applications have appeared over the years so that repositories can be used in different ways and purposes. The issue of costs is also a limiting factor. Burns, Lana & Budd (2013) warn that little is known about the costs incurred by an academic library when implementing and managing institutional repositories and the value that these repositories offer to their communities.

Beyond technological issues, human factors limit the deposit and sharing of materials. Wu (2015) estimates that only 15% -30% of academics deposit their work in institutional repositories. These studies suggest that the most common reasons that contribute to mistrust are plagiarism and copyright issues, the impact of their citations, credibility, and the time required to enter objects into the repository. The problems with most impact in small institutions are the limited budget, lack of planning and lack of technological expertise. To promote the development of its repository, the university that publishes this paper proposes to use more of its Institutional Repository (IR) materials in the classroom, to start producing audiovisual content and to integrate IRs with classroom management systems. In addition, it is proposed to use the repositories as a space to

---

apply what was learned in class, publishing journals made completely based on the work of the students; this can integrate many disciplines such as graphic design and research.

Several analyzes have been conducted on institutional repositories, both technical and on usability. For example, in the perspective of users of Institutional Repositories, Jean, Rieh, Yakel, Markey & Samet (2009) present an investigation about the use that users give to Institutional Repositories and their perception about them. In their research questions one can appreciate a concern about how frequently users access the repository, for what purpose they do it, whether or not they use the content they find, and if they recommend its use to their colleagues. Berquist (2015) explains how repositories have grown and the implications they have for journals, researchers, universities, and so on. It outlines the problems that arise from the Institutional Repositories and calls for solutions for authors to work better with institutions. In 2016 Berquist focuses on how to improve the ways of working and minimizing the time and effort required for the deposit of articles in repositories.

Other studies have also been able to add value to the growth of institutional repositories. Schöpfel (2013) proposes five ways to add value to institutional repositories:

- **Quality**: this can be achieved by constraining the acceptance of any type of work, giving unrestricted access to the full text and having validation of the institution.
- **Metadata**: using standardized schemas you can enrich the description of content that is not explicitly written in the text. This increases the likelihood of finding resources in repositories.
- **Format**: institutional repositories must contain the complete text, not only metadata, independently of having files with formats such as MS Word and PDF. For scanned files, it is important to use OCR before indexing them.
- **Interoperability**: It is important that repositories connect to each other. For this it is necessary to have a standard format for describing resources, policies for the exchange of information, etc.
- **Services**: you must have an advanced search, display, download and navigation system to improve the user experience.

In addition, it is suggested to use another type of non-textual format and enrich the resources with 'Deep Access', linking them with other information, alternative representations and activities.

Open technologies and the possibilities they provide will have a predominant role in the next generation of repositories. Sotiriou, Riviou, Cherouvis, et al. (2016) present research on education using Open Spaces of Discovery, an initiative that pre-modernizes schools with a large implementation of open-scale methodologies using technological innovation. The researchers applied the model to different schools and observed how student connections and overall satisfaction grew, as technological maturity increased considerably, and as those who participated gradually shared their knowledge. Rodés-Paragarino, Gewerc-Barujel & Llamas-Nistal (2016) mention several studies on the use and adoption of repositories and other digital educational resources, making clear that there is a challenge to strengthen those that can still give repositories in terms of connection.

In the open educational movement, repositories still have a long way to go. Xia, Gilchrist, Smith, Kingery, Radecki, Wilhelm and Mahn (2012), mention that, although many policies have proven to be effective and positive, some others have no impact on the development of repositories. So that the policy of the open educational movement, by itself, will not change the existing practices of academic self-archiving. A clear aspect is evidenced by Ezema (2011), who presents the necessary elements for the development of the open educational movement and the use of repositories in Nigeria. The article proposes the development of a new technological culture, awareness, financing

of universities, development of ITC infrastructure and the presentation of theses and electronic dissertations.

Conclusions

Despite the growth of institutional repositories reported by OpenDOAR is linear, the number of records they hold increases exponentially. By February 7th 2017, the 2,824 institutional repositories listed in OpenDOAR accounted for 105.75 million records, with an average of 38,300 records per repository.

As described above, this growth is motivated both by the support given by universities to the open access movement and by the relevance of Web presence in university rankings. And while the metadata format and harvest protocols provide facilities for inspecting the contents of repositories, finding the right resources for research or teaching materials remains a difficult problem to solve. Resources are cataloged under diverse classification systems and despite efforts such as ORCID[3] there are no unique identifiers of authors associated with all available materials. Recognizing equivalent resources, authors, and affiliations across repositories is another hard task.

As can be seen, Data Analytics play an important role in this scenario. The question remains open whether the impact (citation) of scientific publications published in open access through institutional repositories is greater than that of restricted access publications, and whether this is independent of the quality of the journal or the paper. Identifying related resources and determining the characteristics of the most downloaded resources are other areas of application.

Finally, although institutional repositories represent a support tool for institutions, there is still much work to be done in organizational cultures in order to obtain the benefits that these can provide for academic and scientific work. The invitation to further contribute to studies that lead to the technological and academic growth of the potential of institutional repositories remains open.

References

Aguillo, I. F., Ortega, J. L., Fernández, M. & Utrilla, A. M. (2010). *Indicators for a Webometric Ranking of Open Access Repositories*. Digital csic Sitio web: http://digital.csic.es/bitstream/10261/32190/1/Ranking%20of%20Repositories.pdf

Antonczak, L., Keegan, H., & Cochrane, T. (2016). mLearning and Creative Practices: a Public Challenge?. *International Journal of Mobile and Blended Learning (IJMBL),* 8(4), 34-43.

Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories. *Library Hi Tech,* 30(1), 60–81. doi:10.1108/07378831211213210

Berquist, T. H. (2015). Open-Access Institutional Repositories: An Evolving Process? *American Journal of Roentgenology, 205*(3), 467-468. doi:10.2214/ajr.15.15234

Berquist, T. H. (2016). Authors and Institutional Repositories: Working Together to Reduce the Time to Final Decision. *American Journal of Roentgenology, 207*(1), 1-1. doi:10.2214/ajr.16.16611

---

[3] Open Research and Contributor ID (ORCID). https://orcid.org/

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2013). Anatomy of green open access. *Journal of the Association for Information Science and Technology, 65*(2), 237–250. doi:10.1002/asi.22963

Burns, C. S., Lana, A., & Budd, J. M. (2013). Institutional repositories: Exploration of costs and value. D-*Lib Magazine, 19*(1/2). doi:10.1045/january2013-burns

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368*(1926), 4023–4038. doi:10.1098/rsta.2010.0165

Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: A study of conflicting paradigms. *The Journal of Academic Librarianship, 37*(6), 460–470. doi:10.1016/j.acalib.2011.07.002

Davis, H. C., Carr, L., Hey, J. M. N., Howard, Y., Millard, D., Morris, D., & White, S. (2010). Bootstrapping a culture of sharing to facilitate open educational resources. *IEEE Transactions on Learning Technologies, 3*(2), 96–109. doi:10.1109/tlt.2009.34

Dawson, P. H. & Yang, S. Q. (2016) Institutional Repositories, Open Access and Copyright: What Are the Practices and Implications?. *Science & Technology Libraries, 35*(4). doi: 10.1080/0194262X.2016.1224994

Ezema, I. J. (2011). Building open access institutional repositories for global visibility of nigerian scholarly publication. *Library Review, 60*(6), 473-485. doi:10.1108/00242531111147198

Fontes Ferreira, A., Souza de Silva, A. L. (2015). Institucional Repositories: Accesibility to academic research- public property. *IEEE Explore*,172-175

Johnson, R., Fosci, M., & Pinfield, S. (2015). usiness process costs of implementing "gold" and "green" open access in institutional and national contexts. *Journal of the Association for Information Science and Technology, 67*(9), 2283-2295. doi: 10.1002/asi.23545

Gargouri, Y., Hajjem, C., Lariviére, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE, 5*(10) doi:10.1371/journal.pone.0013636

Glassman, M., & Kang, M. J. (2016). Teaching and learning through open source educative processes. *Teaching and Teacher Education, 60*, 281-290.

Gonzalez, L. (2016). Representing Serials Metadata in Institutional Repositories. *The Serials Librarian, 70*(1-4), 247-259.

Jean, B. S., Rieh, S. Y., Yakel, E., Markey, K., & Samet, R. (2009). Institutional repositories: What's the use? *Proceedings of the ASIST Annual Meeting*, 46. doi:10.1002/meet.2009.1450460315

Kim, J. (2011). Motivations of faculty Self-archiving in institutional repositories. *The Journal of Academic Librarianship, 37*(3), 246–254. doi:10.1016/j.acalib.2011.02.017

Laakso, M., & Björk, B. C. (2012). Anatomy of open access publishing: a study of longitudinal development and internal structure. *BMC medicine*, *10*(1), 124.

MacIntyre, R. & Jones, H. (2016). IRUS-UK: Improving Understanding of the Value and Impact of Institutional Repositories. *The Serials Librarian, 70* (1-4), 100-10. doi: 10.1080/0361526X.2016.1148423

Marcial, L. H., & Hemminger, B. M. (2010). Scientific data repositories on the web: An initial survey. *Journal of the American Society for Information Science and Technology, 61*(10), 2029–2048. doi:10.1002/asi.21339

Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H. S., & Hussain, A. (2014). Open-access repositories worldwide, 2005-2012: Past growth, current characteristics, and future possibilities. *Journal of the Association for Information Science and Technology, 65*(12), 2404–2421. doi:10.1002/asi.23131

Ramírez, M. S. (2015). Acceso abierto y su repercusión en la Sociedad del Conocimiento: Reflexiones de casos prácticos en Latinoamérica. *Education In The Knowledge Society (EKS), 16*(1), 103-118. doi:10.14201/eks2015161103118. Disponible en: http://catedra.ruv.itesm.mx/handle/987654321/873

Rodés-Paragarino, V., Gewerc-Barujel, A. & Llamas-Nistal, M. (Mayo de 2016) Use of Repositories of Digital Educational Resources: State-of-the-Art Review, IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 1(2), 73-78. doi: 10.1109/RITA.2016.2554000

Schöpfel, J. (2013). Adding Value to Electronic Theses and Dissertations in Institutional Repositories. *D-Lib Magazine* D-Lib Magazine, 19(3-4).doi: 10.1045/march2013-schopfel

Sotiriou, S., Riviou, K., Cherouvis, S., Chelioti, E., & Bogner, F. X. (2016). Introducing Large-Scale Innovation in Schools. *Journal of Science Education and Technology*, *25*(4), 541-549.

Shukair, G., Loutas, N., Peristeras, V., y Sklarß, S. (2013). Towards semantically interoperable metadata repositories: The asset description metadata schema. *Computers in Industry, 64*(1), 10-18.

Koler-Povh, T., Mikoš, M., & Turk, G. (2014). Institutional repository as an important part of scholarly communication. *Library Hi Tech, 32*(3), 423 – 434. doi: http://dx.doi.org/10.1108/LHT-10-2013-0146

Tzoc, E. (2016). Institutional repository software platforms at undergraduate libraries in the United States. *College & Undergraduate Libraries,23.* doi: 10.1080/10691316.2014.959230

Wallace, J. M. (2011). PEER: Green open access—Insight and evidence. *Learned Publishing*, *24*(4), 267-277.

Way, D. (2010). The open access availability of library and information science literature. *College and Research Libraries, 71*(4), 302–309.

Wu, M. (2015). The future of institutional repositories at small academic institutions: Analysis and insights. *D-Lib Magazine, 21.* doi: 10.1045/september2015-wu

Xia, J., Gilchrist, S. B., Smith, N. X., Kingery, J. A., Radecki, J. R., Wilhelm, M. L., ... & Mahn, A. J. (2012). A review of open access self-archiving mandate policies. *portal: Libraries and the Academy*, *12*(1), 85-102.

Bio

Maria Soledad Ramírez
Tecnológico de Monterrey
María Soledad Ramírez Montoya is researcher and director of postgraduate and continuing education at the Tecnológico de Monterrey's School of Education, Humanities and Social Sciences. She is the coordinator of the Research and Innovation Group of the Tecnológico de Monterrey, director of the International Council for Open Education (ICDE): "Latin America's Open Education Movement" and director of the UNESCO Chair: "Open Educational Movement for Latin America". She is also a research professor invited by the University of Salamanca where she advises thesis and teaches courses in the Doctoral Program Training on Knowledge Society. Her lines of research are the open educational movement, innovative teaching strategies, technological resources for education and training of educational researchers.

Héctor G. Ceballos
Tecnologico de Monterrey
Hector G. Ceballos is head of the Scientometrics office of Tecnologico de Monterrey, where he developed its Current Research Information System (CRISTec), and currently advices the development of the Institutional Repository meanwhile he conducts analysis on scientific collaboration and research group's performance. He is also collaborator researcher of the Intelligent Systems group at the School of Engineering and Sciences of Tecnologico de Monterrey, member of the Masters in Intelligent Systems' faculty staff at Campus Monterrey and adherent member of the Mexican Society on Computing (AMEXCOMP). His research lines are Semantic Web technologies, Bayesian Networks and Causality applied to Research Assets Management, Cultural Heritage Visualization and Smart Cities.