Open Access to Scientific Literature and Research Data: A Window of Opportunity for Latin America

María del Rosario Molfino[1, 2], Claudía M. González[1,3]

[1]Universidad Nacional de La Plata, School of Humanities And Education Sciences. Library Science Department
[2]Universidad Nacional de La Plata. School of Information Technology Library
[3]Institute of Research in Humanities and Social Sciences (IdIHCS - UNLP/CONICET La Plata)
48 e/6 y 7, La Plata, Argentina
{rmolfino, cgonzalez}@fahce.unlp.edu.ar

Abstract: The advance that the international open access movement has had in the last decade may seem to suggest that we are witnessing an important change in the model of scientific communication. This paper introduces the fundamental concepts of this movement, and in turn tries to measure the impact it has had in Latin America based on the development of different strategies.

Keywords: Open Access, Open Access Movement, Academic and Scientific Production, Primary Data, Repositories.

Introduction

The issue of *Open Access* to scientific information (OA) becomes strong in the 90s, although it can be traced back to 1966 [1]. It emerges in a scenario marked by the economic difficulty of the state research system to maintain its information services based on the subscription to paid resources. The movement takes shape when institutions understand that they are trapped in a system where, in addition to subsidizing the research processes, they have to pay for access to sources of information held by publishing houses that, paradoxically, become holders of the intellectual property rights of a large part of the results of their activity. Consequently, the researchers start to suffer the non-access to what is produced by their own –local or disciplinary– community.

The solution seems to be on a new model that, based on the technology of Internet, calls on the scientific and academic community to consent the free distribution of their intellectual property. Most searches are for authors of non-published works –as the case of most post-graduate theses– and authors of journal articles who usually receive no profit in exchange. This situation clearly leaves them out of the debate on

open access to books, music and movies. The most important argument is full compatibility with the traditional publishing system with respect to peer review. The model proposes that institutions invest in structuring, on the one hand, low-cost digital publishing systems to produce free access journals with referees who donate their work (which is indeed the usual fashion), and on the other hand, it proposes the development of digital repositories to store their production, i.e. the creation of platforms that facilitate their authors the self-archiving of materials once they have been through the –traditional or not– publishing circuit, guaranteeing some kind of peer review [2].

This initiative is an integral part of a group of movements which may be gathered under the large umbrella of open access: *open access to knowledge* [3], *open access to information* [4], *open access to scientific data* [5], *free software* [6], *open source* [7], *open-source hardware* [8], locally known as *common property* [9]. Even though the process we are analyzing here may be classified under this "free culture", it is worth mentioning that, at present, it has already involved the political decision of several governments, the commitment of resources of several scientific organizations and a major pragmatic effort of the world leading publishing houses in order to advance to new global management and business models as survival strategy.

This paper is intended to introduce the main aspects of the *OA* movement, starting with the first conceptualization which was formally presented in the Budapest Initiative in 2001 and which was later consolidated in the Bethesda Statement and the Berlin Declaration in 2003. We also intend to specify the media proposed to achieve these goals and present current data which will show the global situation as well as the movement's particular situation in Latin America. A special interest is taken in the most recent derivation of the movement addressed to *open science data*, and we conclude by drafting of a window of opportunity for our environment.

Main Concepts

Cost Free

From the financial point of view, the cost-free availability of scientific content in Internet means a radical change in the scientific communication model. Since 1665 – the year when the first scientific journal was established–, this kind of publication gained a position at the core of the research system by becoming the quintessential means for knowledge sharing. The possibility of examining in academic articles the results obtained by others has allowed, for the past 350 years, a greater flow of ideas, information and data, which may be subjected to observation and improvement. According to Ulrich's Periodicals Directory –a global directory with information about journals– there are currently over 93,000 active titles in the world [10], producing about 1,7 million academic articles [11].

However, and notwithstanding the value that the scientific journals represent for the development of science, it is a well-known fact that for the past 20 years, the costs established by the publishing market have undermined the sustainability of this model. The Harvard University Library –one of the best financed libraries in the

world– issued a memorandum in April 2012 stating its concern regarding the cost faced for the subscription to services from the main journal publishers. With $3.75 million dollars being only a 10% of what the library acquires on a yearly-basis, they denounce a 145% increase in the past six years imposed by the main publishing groups (Elsevier, Springer y Willey) [12]. In Argentina, the Science and Technology Electronic Library (Biblioteca Electrónica de Ciencia y Tecnología), financed by the nation, reports a 58% increase in its 2010/2011 budget keeping a consistent subscription package, comprising barely fewer than 11,000 titles [13].

Considering that knowledge is a type of possession that may be shared without exhaustion and that the ubiquity offered by Internet is a fact for scientists around the world, the opportunity to boost the scientific communication model to its highest expression may be found within reach. If the cost of generating new knowledge and preparing a scientific article is fully afforded by the authors and their institutions, the cost of *peer review* –the other fundamental element of the model– and on-line publishing, seems to fail to justify the price imposed by publishers. For promotion agencies, that invest in generating science and then pay to have access to their results, adhering to and promoting an economic change in the model is a way of demanding a fairer balance between their input and output. For libraries, promoting free access is a way of breaking down barriers in order to guarantee the access to information as a universal right. Finally, for authors, this implies the opportunity to read and be read without restrictions.

Although it is worth mentioning that any kind of article openly published on the Internet cannot be read without a considerable investment in hardware, software and networks –an infrastructure that is directly afforded by users themselves or indirectly by the taxes they pay to sustain the public system–, the underlying intention is to break down the barrier of payment for content, so that more people can gain greater access to research literature.

Freedom of Use and Re-Use

Another basic element of OA, in addition to the cost free guarantee, is the abolition of certain restrictions in the rights of exploitation of works. As it is well-known, copyrights imply moral rights and exploitation rights. Moral rights are basically two-fold: acknowledgement of authorship and respect for the integrity of the work, and they do not expire. However, exploitation rights do expire. For a period of about 70 years after the death of the author, depending on the legislation of the country, the successors in title preserve their reproduction, distribution, public communication and work transformation rights. Based on the OA movement, the way in which this legislation has been transmitted to society has been one of the factors why authors have been hostages to publisher's commercial interests for a long time.

Scholars are not interested in selling, their effort is rather put on gaining acceptance of their ideas and progress by the community. For most scientific authors, a personal financial retribution as consideration for their contribution to knowledge is not as important as a compensation through peer acknowledgement and prestige. It may seem that while exploitation rights may lack importance, the other component of

copyrights is vital to obtain the acknowledgement of “Who's who” in science. By recognizing this duality, OA initiatives have worked on creating standardizing instruments that allow a real change in the way knowledge is spread, and they call on authors to exercise their right of freedom to choose what they want to do or not with their works.

The main contribution is provided by open licenses, including Copyleft in the field of Information technology and Creative Commons in the field of publications –which are the most popular ones. Both instruments generally seek to extend their application based on regulations themselves, i.e. imposing or suggesting for derived works the application of the same type of license. This would allegedly entail a cascade propagation effect that would end up changing the information ecosystem around us. The Creative Commons license –which is much less rigid than Copyleft– recognizes authorship, but it is more flexible because it offers alternatives regarding the creation of derived works.

As mentioned in the previous section, regarding scientific progress and its close relationship with the possibility of improving the ideas and results obtained by others, offering immediate (i.e. cost-free) access and admitting the re-use of information are two vital aspects. Being free not only to read what has been produced, but also to manipulate such content in order to produce new enriched works, is one of the substantial changes of the new model.

Gold Road and Green Road

Considering that OA to scientific literature entails that users may read, download, copy, distribute, print, search and/or link the complete text of scientific articles and use them for any other legitimate purpose, without any financial, legal, or technical limitations other than gaining access to Internet itself, there are basically two roads to reach open access: the authors publish their works in open access journals (gold road), and/or the authors self-archive their works in their personal websites or in an open access institutional or thematic repository (green road).

The first open access scientific journals appear in the early 1990, and according to Ulrich’s Periodicals Directory, they currently comprise over 11,000 titles; 6,798 of which have external review. Whether considering the total number of scientific journals or only the journals with peer review system, the percentage of OA titles is between the 12-13% of the total number of titles in the world [10]. The Directory of Open Access Journal (DOAJ), a multidisciplinary directory of open access journals supported by Lund University, currently includes over 7800 OA journals. This category includes: open and cost free for both readers and authors journals, and those which are paid only by authors, also journals that combine the traditional subscription model permitting authors to choose if they want to pay for open access publishing, and finally, journals that offer contents for free after an embargo period [14]. All of them pursue the same quality standards as the traditional subscription model journals, and are therefore starting to be indexed by the main specialized directories.

Institutional repositories, in turn, are collections of diverse academic materials digital objects, grouped following an institutional production or similar subject

criteria, and they always follow a well-defined policy. Some of them, among other features, admit self-archiving by the authors; guarantee interoperability and adjust to the Open Archives Initiative (OAI) protocol that enables their resources to be accessed by other systems; and are committed to long-term preservation. Their purpose is larger than the one pursued by gold road journals as they seek, in addition to spreading knowledge, to provide visibility to their institutional production, organize it and preserve it. Their origin dates back to 1991, when the thematic repository arXiv is founded, created by physicians to openly share their pre-prints. At present, according to the *OpenDOAR* directory of OA repositories, there are over 2100; distributed as follows: 47% in Europe, 21% in USA, 18% in Asia, 7% in South America and 7% in the rest of the world.

The fashion in which these two roads have evolved is subject to permanent study, mainly because leading science managers feel the need to define stimulus and funding policies in order to involve their institutions in OA projects. Recent examples are the report of the Danish Agency for Science [15] and the report prepared for the UK Ministry of Science, known as Finch Report [16].

Research Data

There is a branch of the *OA* movement that stands out mainly because of the advantage it presents related to the feature of re-use of information. This is a trend that works to make the data generated during research processes openly available. This is an major proposal because, in addition to generally maintaining OA elements, it includes a change in the scientific communication system by suggesting another form of distribution of new knowledge different from the hegemonic form that has followed the scientific journal article in the past three centuries.

Having access to data and models that scientists gather and create during their research involves depositing said data in an *ad hoc* repository. These data may have an associated journal article or not, since one of the aspects proposed by the model is to reduce the time span between the production of information and the moment when it reaches the community of interest to be reviewed. It is well-known that in many disciplines, such as medicine, the time required to complete the circuit until data reaches publication may be detrimental for the scientific community to solve an urgent problem. However, this possibility of obtaining faster feedback from colleagues, in a more collaborative proposal, involves developing effective communication mechanisms that turn data intelligible for others. Although all repositories require digital objects to be stored together with their corresponding meta-data records, in primary data repositories, this kind of record becomes paramount as it will provide information regarding where the data come from, how the data were acquired and treated, on which date, how the data were analyzed for a scientific purpose and how the data must be used. This matter becomes a real challenge when the purpose is also to reach multidisciplinary audiences. On the other hand, this same OA trend proposes that data provided in scientific journal works should be accessible to be re-used and should be presented in a way that allows for a direct correspondence with the parts of the article in question. Both the repository record and the journal article should instruct under which conditions data can be used.

Actors in Latin America

In Latin America we are witnessing not only the growth in the number of open access repositories and journals, but also the development of other related expressions and initiatives that have emerged in the domain of science and technology institutions. In order to have a more complete view of our regional reality, some statistical data extracted from reference websites will be shown and the main progress achieved will be described.

Growth Aspects

With regard to the green road, i.e. the development of institutional or thematic repositories, such as the one shown in Table 1, Latin America (excluding Mexico), is a bit behind as compared to other continents and North America. If we take into account the 20 repositories kept by Mexico (which in the Table are included in North America), Latin America holds 203 repositories, a 9.4% of the total number of repositories registered in *Open*DOAR [17]. At the top is Brazil with 62 (30.5%), followed by Colombia, Mexico and Argentina with 20 (9.9%) each, then there are Ecuador with 17 (8.4%), Peru and Venezuela with 13 (6.4%) and Chile with 10 (4.9%). Institutional repositories clearly prevail over disciplinary ones. Considering only the aforementioned countries, out of 175 repositories, 142 (81.1%) are institutional, 16 (9.1%) are disciplinary and the remaining 17 (10%) are aggregating and governmental repositories.

Europe	1018	47.1%
North America	463	21.4%
Asia	382	17.7%
South America	159	**7.4%**
Australasia	60	2.8%
Africa	54	2.5%
Caribbean	13	**0.6%**
Central America	11	**0.5%**
Other	2	0.1%
TOTAL	2162	100.0%

Table 1. Number of Repositories per Continent. Source OpenDOAR (28 June, 2012)

In terms of the gold road, we can observe that all regions are showing growth. Although Europe stands out, we can also note that Latin America has kept a growth rate similar to that of North America

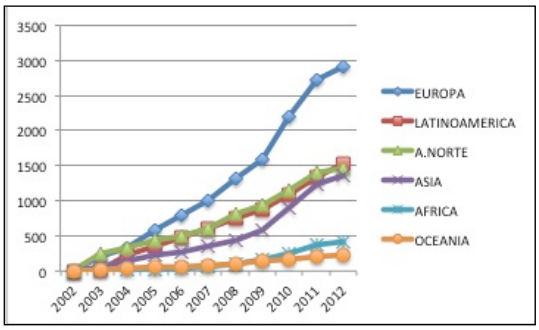

Fig. 1 Number of OA Journals per Continent. Source DOAJ (28 June, 2012)

If we estimate the annual growth rate between 2003 and 2011, Latin America is the region that accounts for the most significant growth (0.73), followed by Africa (0.62), Asia (0.46), Europe (0.38), Oceania (0.36) and North America (0.25). The high rate shown by Latin America is explained by the growth that exceeded a 1200% between 2003 y 2004.

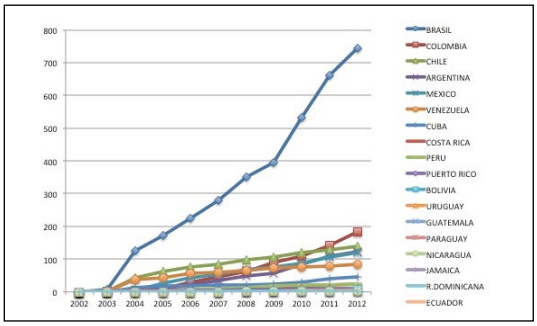

Fig. 2 Number of OA Journals per Latin American Countries. Source DOAJ (28 June, 2012)

When analyzing the increase per countries of *OA* journals in Latin America, the sharp difference shown by Brazil as compared to the rest is easily noticed. However, the annual growth rate between 2004 and 2011, exhibits better figures for Argentina (0.59), followed by Colombia (0.49) and Mexico (0.23).

Regulatory Aspects

The Berlin Declaration of 2003 is a milestone that initiates the political/governmental recognition at world level of the *OA principles.* The major German scientific organizations committed thereby to call on their scientists to publish in compliance with these principles, i.e. to follow any (or both) road(s). In turn, said institutions would guarantee publications of this kind to be recognized during the scientific career evaluation processes and they took a commitment to manage and maintain digital repositories with long term preservation of said materials.

This major agreement later drove other scientific and technical organizations from

several countries to adopt measures that generally range from signing adherence, preparing statements, establishing mandates, creating lines of credit for *OA* projects, to enacting laws. Similarly, in Latin America, Brazil and Argentina have bills of law which are being discussed at the Congress. Among the essential points included in said bills of law is the obligation to deposit works in open repositories for research financed with public funds. Peru has also started its way down this road with the Bill of law No. 01188 of May, 2012.

The Argentine bill of law defines as *"scientific-technical production the set of documents resulting from the scientific-technical activities that go through a quality review process, whether they are published or not. Primary research data may include, but are not limited to, text, numbers, equations, algorithms, images, audio and video, animations and software tools"*, Brazil's bill of law fails to consider primary data. Likewise, in Argentina, the Ministry of Science, Technology and Productive Innovation (Ministerio de Ciencia, Tecnología e Innovación Productiva, MINCyT) recently concluded the formalization and creation of the National Science and Technology Digital Repositories System (Res.469/2011)

On the other hand, academic institutions that keep institutional repositories have found in the concept of *mandated* deposit in their own repositories a way of putting into practice *OA* statements. And they do this not only because it provides greater accessibility to information, but also because the results of the research they produce and/or fund may achieve greater visibility and impact. ROARMAP (*Registry of Open Access Repository Material Archiving Policies*), a registry of *OA* repositories archiving policies at international level that provides systematic information regarding their most salient aspects, in June 2012, contained 419 registered mandates, 16 out of which belonged to Latin America. In Argentina, for example, two university mandates of a different nature are the UNLP mandated theses deposit (resolution 78/2011) and the obligation to use *Creative Commons* licenses for UNC academic publications (resolution 116/2010) –the existence of which is known to us, despite the nonresistance of their registration in ROARMAP.

The progress of OA to scientific information would receive major support and boost were it established by national legislation.

Both Brazil and Argentina have bills of law which are under the process of discussion at the Congress, *"Dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil e dá outras providências"* and *"Creación de repositorios digitales abiertos de ciencia y tecnología. Ciencia Abierta Argentina 2010"*, respectively.

In Argentina, beginning in 2009 policies of OA to scientific knowledge start to spread and became established in the academic field, helped by the boost provided by the Ministry of Science, Technology and Productive Innovation (MINCyT), which recently concluded with the formalization and creation of the National Science and Technology Digital Repositories System (Res.469/2011), and the drafting of the bill of law that has recently been passed by the House of Representatives.

Funding Aspects

The region of Latin America has already some important projects related to *OA* which

involve significant funding. The Network of Scientific Journals from Latin America and the Caribbean, Spain and Portugal (Red de Revistas Científicas de América Latina y el Caribe, España y Portugal) [18], promoted by Universidad Autónoma del Estado de México, and the SciELO (Scientific Electronic Library Online) [19] platform, developed by FAPESP-BIREME, which has been implemented by eight Latin American countries, gather their most relevant national production.

The Latin American Cooperation of Advanced Networks (Cooperación Latinoamericana de Redes Avanzadas, RedCLARA), with the support of the Inter-American Development Bank, is promoting a project to create the Latin American Federated Network of Institutional Repositories (Red Federada Latinoamericana de Repositorios Institucionales), which intends to share and provide visibility to scientific production from higher education and scientific research institutions of Latin America. The initiative has been supported by eight countries of the region: Colombia, Argentina, Brazil, Chile, Ecuador, Mexico, Peru and Venezuela.

Conclusion

Having an environment rich in Open Access intellectual products at Latin American level allows us to think of a scenario of collective knowledge development with greater capacity to set out the strategic problems of the region. The fact that researchers have the chance to access not only articles, but also the data sets on which research is based is a key point to increase scientific capacity. This involves taking care of the systematic organization of the information produced in the region, assuming a commitment regarding long term preservation and developing information retrieval systems capable of showing said information in the context where it was produced providing a real opportunity for re-use. Similarly, these results will not only be beneficial for the society, but in the medium term, caring for the technological and organizational infrastructure requirements that this kind of activities involve will also turn to be beneficial; and this also tend to dissolve the barriers between more and less developed countries.

References

1. Open Access Directory (OAD): Timeline of Open Access. (2011). http://oad.simmons.edu/oadwiki/Timeline [Accessed on: 25 March, 2011]
2. Suber, P.: Open access overview. (2010). http://www.earlham.edu/~peters/fos/overview.htm [Accessed on: 25 March, 2011]
3. Open Knowledge Foundation (OKF): Open knowledge definition. (2011). http://www.opendefinition.org/ [Accessed on: 25 March, 2011]
4. Assange, J.: What is WikiLeaks? (2011) http://213.251.145.96/About.html

[Accessed on: 25 March, 2011]
5. Murray-Rust, P.: Open data in science. (2008) http://precedings.nature.com/documents/1526/version/1/files/npre20081526-1.pdf [Accessed on: 25 March, 2011]
6. Stallman, R.: The GNU manifesto. (1985). http://www.gnu.org/gnu/manifesto.html [Accessed on: 25 March, 2011]
7. Open Source Initiative (OSI): The Open Source definition. (2011). http://www.opensource.org [Accessed on: 25 March, 2011]
8. Bewens, M. P2P Foundation (2011) http://blog.p2pfoundation.net [Accessed on: 25 March, 2011]
9. Verselli, A., Thomas, H.: Repensando los bienes comunes: análisis socio-técnico sobre la construcción y regulación de los bienes comunes. (2008). http://www.bienescomunes.org/lectura/ [Accessed on: 25 March, 2011]
10. Abadal, E.: Retos de las revistas en acceso abierto: cantidad, calidad y sostenibilidad económica. Hipertext.net, 10 (2012). http://www.upf.edu/hipertextnet/numero-10/retos-revistas-en-acceso-abierto.html [Accessed on: 25 March, 2011]
11. Willetts, D.: Public access to publicly-funded research. Publishers Association annual general meeting, London (2012). http://www.bis.gov.uk/news/speeches/david-willetts-public-access-to-research [Accessed on: 23 June 2012]
12. Harvard Faculty Advisory Council. Faculty Advisory Council memorandum on journal pricing, http://isites.harvard.edu/icb/icb.do?keyword=k77982&tabgroupid=icb.tabgroup143448 [Accessed on: 23 June 2012]
13. BECyT (Biblioteca Electrónica de Ciencia y Tecnología. MINCyT. Argentina), http://www.biblioteca.mincyt.gov.ar [Accessed on: 23 June 2012]
14. Abadal, E., Melero, R.: Las universidades y el apoyo institucional al Open Access. ThinkEPI, (2008) http://www.thinkepi.net/las-universidades-y-el-apoyo-institucional-al-open-access. [Accessed on: 28 June 2012]
15. FI & DEFF: Access to Research and Technical Information in Denmark http://www.fi.dk/publikationer/2011/adgang-til-forskningsresultater-og-teknisk-information-i-danmark (2011) [Accessed on: 26 June 2012]
16. Finch, J.: Accessibility, sustainability, excellence: how to expand access to research publications. Report of the Working on Expanding Access to Research Findings. http://www.researchinfonet.org/publish/finch/ (2012) [Accessed on: 25 June 2012]
17. OpenDOAR - The Directory of Open Access Repositories, http://www.opendoar.org/ [Accessed on: 25 June 2012]
18. Red de Revistas Científicas de América Latina y el Caribe, España y Portugal, http://redalyc.uaemex.mx/[Accessed on: 23 June 2012]
19. SciELO, Scientific Electronic Library Online, http://www.scielo.org /[Accessed on: 23 June 2012]