****

# POLICIES FOR OPEN SCIENCE AND SCIENTIFIC DATA

# IN LATIN AMERICA

## LA Referencia

## February 2018

**Technical document by LA Referencia Council to define regionals priorities in open science and scientific open data. Action decided at the General Council Meeting of September 2017, in Buenos Aires, Argentina.**

[Introduction 2](#_Toc517101254)

[Model 4](#_Toc517101255)

[Concept 4](#_Toc517101256)

[Focus 5](#_Toc517101257)

[Disciplines and Metadata 5](#_Toc517101258)

[Scientific Journals 6](#_Toc517101259)

[Licenses 6](#_Toc517101260)

[Technology 7](#_Toc517101261)

[Incentives 7](#_Toc517101262)

[Costs 7](#_Toc517101263)

[Globalization 8](#_Toc517101264)

[FAIR Principles 8](#_Toc517101265)

[Examples of some components 9](#_Toc517101266)

[Conclusions 9](#_Toc517101267)

### Introduction

LA Referencia is a platform of governmental organisms of science and technology of Latin America that fosters the Open Access (OA) to scientific production, with emphasis on the results financed entirely or partly with public funds. It is a result of the national Open Access strategies, through the Institutional Repositories (IR), known as the "green route", forming a Regional Network of Open Access Repositories to Science.

This network works to develop OA policies, technologies and guidelines, with the aim of strengthening the repositories that offer the results of research.   Its strategies and lines of work have as a basic principle the conviction that knowledge is a public good and that a great part of these results have been financed with public contributions from institutions or countries.[[1]](#footnote-1)

There are several statements of support for Open Access around the world, published by institutions such as the Global Research Council, the G8 Science Ministers and the European Commission. In turn, public or semi-public funding agencies have explicitly stated, as a financing condition, that their results should be published in Open Access. An example of this is the European Community's Horizon2020 program, which requires that the results and the scientific data from its funding must be available to all citizens.

Currently, the tendency to promote Open Science (OS) within the Science and Technology (S&T) organisms, represents new challenges and goals for LA Referencia[[2]](#footnote-2)

*Open science generally refers to efforts to make the results of publicly funded research more accessible in digital format for the scientific community, business sector or society at large.* *Open science is the meeting between the old tradition of openness in science and the tools of information and communication technologies (ICTs) that have reformed the scientific enterprise and require* *a critical look from policy makers seeking to promote* *long-term* *research* *and also innovation.*

This general definition of OS comes from the OECD and recognizes the tradition of the Public Goods theory in this field.

This concept must be understood as a mean of improving the quality and impact of the research and not as an end, as the document states. Similarly, the OS is a general framework of assumptions about how the dissemination of knowledge should be and covers areas such as citizen science, scientific data, peer review and open procedures, new metrics, among others. In addition, other authors identify schools or paradigms to analyze the phenomenon, such as infrastructure (which deals with technological architecture), public policies (which deals with access to knowledge creation), measurement (impact and alternative measures), democratic (which refers to access to knowledge) and pragmatic (collaborative research). [[3]](#footnote-3)

It is clear that the multiple aspects that make up the OS must be built from different areas and actors. In this context, given the identity and mission of LA Referencia, the objective is a contribution, in addition to its efforts in Open Access to scientific publications, to scientific data, through infrastructure, services, guidelines, etc. LA Referencia will specifically promote actions in a specific area of OS: the open access to scientific data from the perspective of public goods (metadata guidelines, interoperability, technology, policies).

Open and free access to research data, especially those supported with public funds, enable to:

* Reproduce and validate the results in a more appropriate way. Generate new research and knowledge from existing open scientific data;
* Provide more transparency to research funds;
* Preserving the integrity of research;
* Avoid duplication of efforts

In addition, an important fact is that Latin America already has national laws of OA (Peru and Argentina in 2013, and Mexico in 2014), which include scientific data as part of the policy. Chile, through CONICYT, is promoting good practices of scientific data and Brazil, through Ibict, has launched an Open Data Manifesto. However, directories that enumerate data repositories show a very incipient reality in the region in terms of numbers.[[4]](#footnote-4)

On the other hand, in the last two meetings of the Directing Council of LA Referencia it was decided that actions should be carried out in the OS field, for which a set of decisions that are included in this document have already been defined (San Luís de Potosí, 2016; Buenos Aires, 2017).

The following pages detail the elements to be considered and the priorities for action. It is also clear that this is not a technological issue only, but a cultural change in which scientific data must be a significant result of the research process.

### Model

The region has particularities in relation to the rest of the world with regard to the development of a scientific data strategy:

* The culture of scientific production in Latin America is not commercial. In general, its volume in terms of Repositories and scientific journals is the "double” in OA mode in comparison with its participation in articles published in known databases (Web of Science, Scopus).
* Unlike the rest of the world, R&D funding continues to come mainly from the state: in a direct or indirect fashion.
* An important part of LA Referencia members controls various aspects of the value chain. From R&D funding, production evaluation, support to scientific journals of excellence, to the joint purchase of the produced results. For this reason, articulated public policy projects at a national level are possible.

### Concept

Given the current debate, it is essential to differentiate "open primary scientific data" from the trend called "open government data". The latter aims to place government data in a way that is accessible to all citizens. The first, in addition to its openness, has special complexities due to the need for metadata standards, privacy, interoperability, persistent identifiers, reproducibility and preservation, among others.

### Focus

It is essential to move forward in operational definitions that give focus to the regional action. To this end, LA Referencia decided that, in a first phase, the focus is on the scientific data that support and validate publications; on the scientific data collected or generated with public funding, and those that are based on Data Management Plans (DMP) and that will be part of the research proposals to be funded with public resources.

At the same time, in search for solutions, it will address issues related to software and tools needed to understand and reuse the data mentioned in the previous paragraph along with interoperable metadata standards[[5]](#footnote-5). Another area of exchange of experience is the issue related to harmonization or definition of a common set of elements that are requested to researchers from less developed countries in order to avoid duplication of resources at the regional level.

Finally, it is fundamental to cooperate, from a point of view as the one expressed in this chapter, with initiatives that aim to encourage and propose data management plans in the institutions that produce them.

### Disciplines and Metadata

Although scientific data at regional levels are produced in all areas of knowledge and one component of all scientific policy is the decision not to favor one discipline over another, there are few open scientific data initiatives in Latin America, with some exceptions (data portals, institutional repositories, mandates and management plans, legislation).

At the same time, we note that there are international networks which are consolidated or have tradition in this field (eg.: astronomy, biodiversity, oceanography); networks that integrate many researchers in the region, who, in turn, are used to depositing, on their platforms the research they produce.

We also detected a resistance, from the community, about the shift from the "international deposit" to the "local deposit" paradigm (either in the institutional repository or in national thematic portals). Even stronger is the resistance to the opening of data by those who do not have this habit, in many cases because of the fear of "who will appropriate the data and how it will be used" with a sense of ownership over information, and in others because the practice is seen as one more burden on the researcher's activity.

In this sense, LA Referencia will support pilots in multidisciplinary data projects. The decision is to work in a flexible scenario and work together for greater articulation and interoperability that facilitates the institutions and countries of the network to obtain their own scientific output.

Finally, it is considered that there are areas where greater articulation would be important, as it has a strong impact on regional development and quality of life, such as Biodiversity and the Environment.

Moreover, even if it is obvious, it should be remembered that the priority of LA Referencia is the articulation with services and scientific data repositories that are open, interoperable and non-commercial, providing open access to research data, even with a possible temporary embargo.

Finally, in relation to metadata schemas, there is a set subject to the specificity of each discipline (eg.: astronomy, genomics, biodiversity, social sciences, etc.)[[6]](#footnote-6). LA Referencia decided that, as a general recommendation for the description of all types of scientific data to be deposited in the repositories, it will adopt Datacite v. 4.1.

This decision is based on the fact that this scheme is the most widely used and allows description for multidisciplinary repositories. Also encourages, thematic repositories to add to this basic and common scheme, if feasible, the elements that correspond to more specific schemes which are standard in its subject area. The main objective of this decision is to ensure interoperability and avoid duplication of efforts.

### Scientific Journals

The action in data must be complementary to the initiatives of scientific journals. Latin America is different from other regions and is distinguished by an OA model that, for the most part, has no APC costs. Therefore, the main focus is the design of alternatives that allow, within technology and metadata standards, interoperability between scientific data repositories and publications. Indeed, it is necessary to work with the large journals portals in the region to facilitate the linkages between papers and data and vice versa, exploring the design of alternatives that enable interoperability between scientific data repositories and publications. In this context, we see as very valuable to expand the use of DOI for the data that can be linked in the publications.

In addition, science production must be embedded in research processes to avoid duplication of efforts. We must move forward and collaborate on proposals where reporting is straightforward and simple for our research producers.

### Licenses

It was agreed to advance in some guidelines or recommendations on the type of licenses that are most convenient for scientific data, aiming to facilitate re-use with legal certainty.

In this regard, it is recommended to use Creative Commons licenses, which is the standard in Latin America, and also the most widespread in the world. Likewise, it is suggested that the most convenient alternative, in order to allow reuse and, at the same time, to guarantee long-term OA is Attribution 4.0 International (CC BY 4.0) or the Attribution-NonCommercial-ShareAlike 4.0 International (CC BY- NC-SA 4.0).

### Technology

The design of this network is not simple in terms of the location of the nodes and repositories. In some cases, the option is services in the cloud or hosting outside the country. In others, the option is for systems located the country due to the existing legislation, issues of information privacy, among others.

In this context, it is convenient to maintain a flexible model that facilitates interoperable and federated infrastructures. In fact, an approach based on national data nodes is advisable. For this, the technology for metadata harvesting should be open source and transferable to the national nodes. Therefore, La Referencia will evaluate, in 2018-2019, the development of a scientific data harvester, that can collect, validate and transform metadata from several systems in development with standard protocols (OAI-PMH).

Another aspect is the technology of data repositories. Still there is no "dominant" platform for the implementation of these repositories as was the DSpace for the publications.  This will depend in part on the disciplines and its specific networks. But to promote the infrastructures in institutions, it is convenient technologies that are open source, free and transferable. The experience in the region with publication repositories has shown that reach greater use and innovation. Finally, it prevents the premature “closure” of technologies and standards that may lead to "locking" situations. Therefore, LA Referencia will continue to privilege interoperability in the development of solutions for scientific data, working with the existing platforms.

### Incentives

Since OS is a cultural change, it is considered necessary that institutions and agencies implement actions, recommendations and mandates to promote it. One action that will have impact is that activities related to the management and dissemination of scientific data are considered in the evaluation of academic careers.

On the other hand, it is critical that research institutions adopt explicit recommendations or mandates on scientific data management.  In fact, the evaluation models of researchers, must start from the institutions and become a common aspiration; otherwise, public action is not the result of a constructive dialogue, but will be interpreted from the point of view of imposition. The initial role of modifying evaluation models must come from the research community.

### Costs

The question of who should finance infrastructures and recurrent costs is extremely relevant. As long as the use of PMDs is extended, specific items will be included. Also contributions from some governments at national level. Finally, is vital that institutions provide resources in infrastructure areas as part of the institutional mission of managing its production.

However, there is a relevant number of S&T institutions in the region that simply do not have adequate resources. This is a situation that requires strong intra-regional and international collaboration and cooperation, as well as viable technologies to implement and transfer, and recommendations based on best practices that avoid duplication of efforts. This condition defines the difference between the developed countries and Latin America. In fact, it is convenient to analyze the ICSU's "Open Data in a Big Data World" statement, which introduces elements and considerations in this field for developing countries[[7]](#footnote-7).

In this sense, LA Referencia is an instrument to advance faster, together and in search of solutions with economies of scale.

### Globalization

In regard to the international dimension, LA Referencia, together with COAR (Confederation of Open Access Repositories), participates, since January 2015, in the OpenAIRE project, which is the OA platform for the Horizon2020 programme. One of its tasks was to expand the adaptation/adoption of guidelines and services in Latin America to demonstrate the feasibility of interoperability.

Today, OpenAIRE is also leading a number of initiatives in open science and scientific data in Europe and will continue this common effort at the international level, which strengthens the insertion of our research and offers tangible benefits and synergies. In parallel, the idea is to explore relationships with other global referents, such as ICSU-Codata, RDA and Datacite, among others.

The integration of LA Referencia with global networks and projects from other regions results in the design of strategies and pilots that have global impact.

### FAIR Principles

The previous elements are aligned with the policy of "findable, accessible, interoperable and reusable", or simply ‘FAIR’, which is being implemented in Horizon 2020 and European countries, among others. [[8]](#footnote-8)

"FAIR data management in general terms, your research data should be 'FAIR', that **is findable, accessible, interoperable and re-usable**. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.” [[9]](#footnote-9)

A recent European Community Declaration on the *European* *Science* *Cloud* says: "The implementation of the FAIR principles should be pragmatic and neutral from the point of view of technology, covering the four dimensions: location, accessibility, interoperability and reuse. FAIR principles are not practical standards. Disciplinary areas should develop its specific notions of FAIR data in a coordinated manner, determining the desired level. The FAIR principles must be applied not only to research data, but also to algorithms, tools, workflows, protocols, services and other types of digital data-related research." [[10]](#footnote-10)

### Examples of some components

* Findable: use of persistent identifiers, use of enriched metadata, search engine registration.
* Accessible: recovery by identifiers using standard, open, free and deployable protocols.
* Interoperable: formal, accessible and shared language of representation. Vocabularies that follow FAIR principles.
* Reusable: clear and visible use licenses; information on the origin of the data; standards used in its collection; authors; etc.

### Conclusions

Based on the above, it is reaffirmed that the approximation of the LA Referencia is characterized by the following elements:

* A model based on the federation of national nodes that aims to collect scientific data metadata under common policy, technology and guidelines agreements.
* Use of open and transferable standards and technologies with a non-commercial approach.
* Generation of public goods that facilitate the construction of alternatives without creating early barriers.
* International connection with similar initiatives to advance interoperability and to avoid duplication of efforts.

1. The public good applies to knowledge and justifies that large part of R & D is public. Traditionally they are produced by the state. From a more economic perspective, it is a good that is available to all, and its use by one person does not diminish the use for others. In essence, non-excludable and non-rivalrous products and services, with generally important positive externalities. [↑](#footnote-ref-1)
2. Making Open Science a Reality. OECD, 15 Oct 2015 DOI: 10.1787/5jrs2f963zs1-en

   <https://www.innovationpolicyplatform.org/content/open-science> [↑](#footnote-ref-2)
3. Fecher, B. and Friesike, S. (2013). Open Science: One Term, Five Schools of Thought. In: Bartling, S. and Friesike (Eds.), Opening Science New York, NY: Springer, pp. 17-47. Available at: http://book.openingscience.org/basics\_background/open\_science\_one\_term\_five\_schools\_of\_thought.html [↑](#footnote-ref-3)
4. R3DATA <http://service.re3data.org/browse/by-country/>

   ICSU World Data System. <https://www.icsu-wds.org/services/data-portal> [↑](#footnote-ref-4)
5. For example, the pilot of Horizon2020 of the European Community says: “The Open Research Data Pilot applies to both the data and any metadata, documentation, software and tools needed to understand and reuse the data. At a base level, the Open Research Data Pilot applies to all data needed to validate results in scientific publications, but researchers may also choose to share other curated and/or raw data that they feel has value to others and can be shared.”

   <https://www.openaire.eu/public-documents?id=808&task=document.viewdoc> [↑](#footnote-ref-5)
6. <http://www.dcc.ac.uk/resources/metadata-standards> [↑](#footnote-ref-6)
7. ICSU (2015). "Open Data in a Big Data World": The International Council of Science (ICSU); the InterAcademy Partnership (IAP); the International Social Science Council (ISSC); and The World Academy of Sciences for the advancement of science in developing countries (TWAS). <http://www.icsu.org/news-centre/news/top-news/leading-science-groups-urge-global-accord-on-open-data-in-a-big-data-world> [↑](#footnote-ref-7)
8. . Nature (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, Article number: 160018 (2016) ​doi:10.1038/sdata.2016.18 <http://www.nature.com/articles/sdata201618> [↑](#footnote-ref-8)
9. H2020 Program. Guidelines on FAIR data management. http://ec.europa.eu/research/participants/data/ref/h2020/grants\_manual/hi/oa\_pilot/h2020-hi-oa-data-mgt\_en.pdf [↑](#footnote-ref-9)
10. EUROPEAN COMMISSION. DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION. The Director-General.. EOSC Declaration. Brussels, 26 October 2017. <https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf> [↑](#footnote-ref-10)