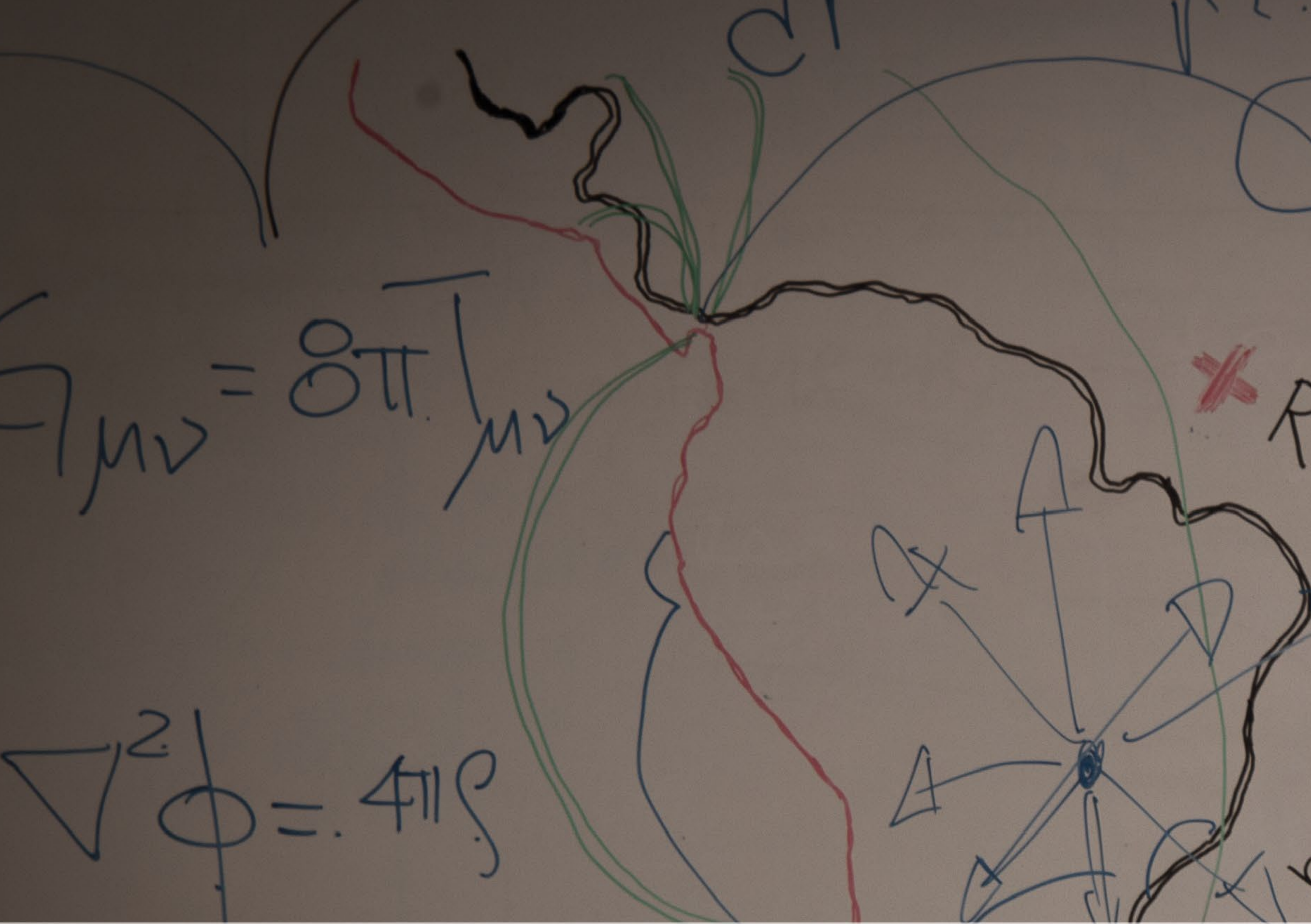


THE DIGITAL OPEN SCIENCE ECOSYSTEM IN LATIN AMERICA AND THE CARIBBEAN

NOVEMBER 2021

The Digital Open Science Ecosystem in Latin America and the Caribbean

August 2022

Authors:
Bianca Amaro, LA Referencia
Carlos Barrios, SCALAC
Luis Eliécer Cadenas, RedCLARA
Federico Cetrangolo, LA Referencia
Leandro Ciuffo, RNP
Rafael Mayo-García, CIEMAT
Lautaro Matas, LA Referencia
Phillipe Navaux, SCALAC
Luis Núñez, UIS
Washington L. Ribeiro de Carvalho Segundo, IBICT

Edition and design: María José López Pourailly, RedCLARA.

Santiago, Chile
26-08-2022

Contents

1. Introduction

Open Science is the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of society, be they researchers, amateurs, or professionals. [1]

As a working model for scientific activity, Open Science has the potential to promote greater economic and social development in Latin America, favouring the efficient and safe utilization of the resources dedicated to funding science and facilitating the processes for knowledge transfer and dissemination in a region rife with challenges and inequities. The incorporation of transparent, collaborative, and open scientific practices also improves the quality of results, their impact, and their reproducibility.

This transparency and accessibility strengthen the ability to translate scientific advances into innovations and allow the interaction of the scientific world with open innovation systems that need to be strengthened to facilitate the creation of wealth and strengthen productivity and the accumulation of human capital in our society.

The draft text of the UNESCO Recommendation on Open Science highlights the five key pillars on which it is supported: open scientific knowledge, open science infrastructure, scientific communication, open engagement of societal actors, and open dialogue with other knowledge systems. [2]

The current dimensions and characteristics of scientific activity are such that its productivity is contingent on certain key factors such as access to global collaboration networks, the effective management of data and scientific publications (e.g., the possibility of searching, accessing, interacting with, and reusing these data and the related scientific publications), computing power, connection speed, and the availability of shared research infrastructures.

This collection of necessary infrastructures serves as the base for the digital ecosystem needed to support and promote Open Science. Latin America has a long tradition in this area that can provide the foundations to develop important leadership in the deployment and strengthening of these ecosystems, increasing our ability to incorporate ourselves into existing collaborations and to promote the development of new scientific collaborations.

This document proposes a vision for this digital Open Science ecosystem that we must finish building in an orderly manner, analysing its components and interactions, while suggesting possible courses of action for its achievement.

2. The Digital Ecosystem of Science

Like most other human activities, scientific activity has been digitized. The work of researchers requires a set of basic services that allow them to:

- Collect, curate, store, publish, maintain, and secure the data produced in the course of their research.
- Access computer resources of all kinds, including traditional and non-traditional high-performance computing capabilities, heterogeneous computing architectures, large-scale or quantum cyber infrastructures.
- Publish and share their articles or work and link them permanently with the data and the algorithms used to process them and which are essential for their reproducibility (this also requires infrastructures that combine massive data analysis capabilities, simulation, and artificial intelligence).
- Have flexible collaboration tools that allow them to integrate their work with that of other researchers and leave a proper record of the process used to draw their conclusions.
- Use collaborative development environments and platforms that allow the traceable, federated, and safe use, assessment, and contribution of open resources that will allow the construction and dissemination of knowledge by the different communities.

The need for infrastructures and organizations that support these capabilities is evident globally, and the countries that are leaders in terms of productivity have an abundance of both. In the case of Europe, for example, this ecosystem is comprised of a wide multiplicity of organizations that coordinate, develop, manage, and make these infrastructures accessible to the European scientific community. High-capacity networks and connectivity; advanced computer centres (including supercomputer or quantum resources); massive data centres; centres of excellence; digital innovation hubs; large, shared research infrastructures, and a European Open Science Cloud make up the ecosystem that supports its scientific activity. With minor variations, the same applies to Canada, the United States, Japan, China, and others. In these cases, there is a clear alignment between the institutions that fund science, the scientific community, and the group of institutions that support the promotion and development of these infrastructures.

In Latin America, organizations have been created over several decades that have contributed to the construction of this digital ecosystem which is still incomplete. Many of these initiatives resulted from collaborative scientific initiatives in areas such as high energy physics, astrophysics, or environmental sciences. These organizations are currently in a position to offer the infrastructure pillar mentioned above and to integrate it in each country at the national level, facilitating the regional interconnection of national science, technology, and innovation systems, and their relationship with other global ecosystems. Examples of these organizations include:

- National research and education networks.
- CLARA, the Latin American Cooperation of Advanced Networks.

- LA Referencia, the Federated Network of Institutional Repositories of Scientific Publications.
- SCALAC, the Advanced Computing System for Latin America and the Caribbean.
- The research infrastructures of major international experiments that generate data, such as the European Southern Observatory (ESO, https://www.eso.org/), the Pierre Auger Observatory (PAO, https://www.auger.org/) or the Latin American Giant Observatory (LAGO, http://lagoproject.net).

This ecosystem is lacking in regional efforts and capabilities to have an integrated open data infrastructure that can be used to complete the tools required by researchers. This document proposes a reference architecture for this ecosystem comprised of infrastructures and organizations, as well as some alternatives to build the capabilities that are lacking.

2.1. The Latin American Cooperation of Advanced Networks

CLARA was born in 2003 in Valle de Bravo, Mexico. There, 13 Latin American national research and education networks came together to create an organization that would help build the interconnection infrastructure needed to connect with each other and with the rest of the world. Back then, the interconnection took place by default at the Network Access Points (NAPs) located in Florida. By that time, national research and education networks had already been in existence for about a decade, at least in the countries that initiated these efforts.

These national networks were created to provide interconnection services and internet access to universities and research centres in each country. After more than five years of cooperation within the framework of the Latin American Network Forum, the creation of CLARA was a natural step towards the formalization of an international cooperation which, in practice, had already been operating for many years.

Since its creation and thanks to the ongoing support of the European Commission, CLARA managed to deploy RedCLARA, an infrastructure unique in its nature and characteristics and with major connectivity capabilities, an essential element for the integration of an Open Science ecosystem. Investments have made it possible to build a regional public good with assets and investments in excess of one hundred million dollars.

Co-funded by four Latin American national networks and the European Commission, over the past four years the BELLA (Building the Europe Link to Latin America) project has deployed a direct high-speed connection between Latin America and Europe, through a new submarine cable that connects Sines, in Portugal, with Fortaleza, in Brazil. The first phase of this project deployed several high-capacity terrestrial links, connecting Brazil, Argentina, Chile, Panama, and Ecuador. In 2020, a direct and high-capacity connection with Africa was established, creating opportunities for cooperation with African networks. Figure 1 shows RedCLARA's current topology.

Fig. 1: RedCLARA infrastructure.

As shown in Fig. 1, several Latin American countries are either not connected or have low-capacity connections. RedCLARA aspires to finish connecting these countries so that they will have equal opportunities to be part of the regional and global science, technology, and innovation ecosystem. [3] These connectivity plans are complemented and strengthened by other Latin American projects such as the Humboldt cable, which seeks to connect Chile with Oceania and Asia, and the cable to Antarctica, which would connect the continent via Patagonia (Fig. 3).

Fig. 2: BELLA 2030 topology

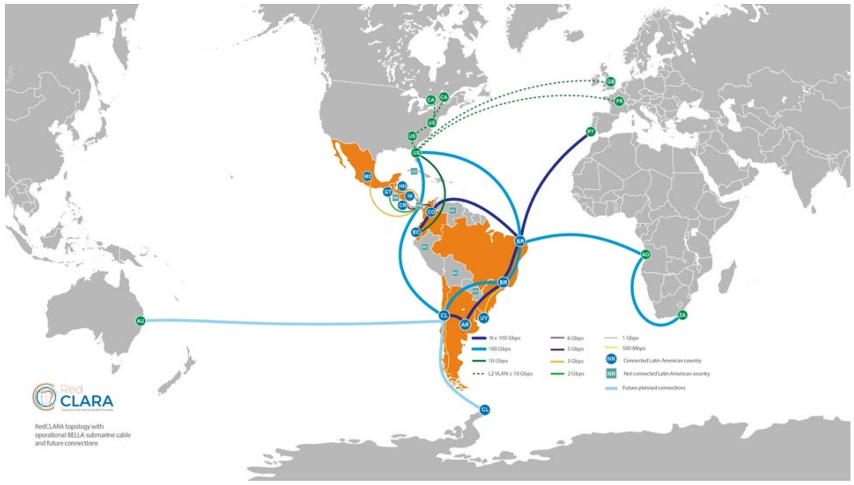


Fig. 3: Humboldt cable and cable to Antarctica.

2.2 Latin American Network of Open Access Repositories (LA Referencia)

The Federated Network of Institutional Repositories of Scientific Publications, or simply LA Referencia, is the Latin American network of Open Science repositories created and supported by the governments of the region. By means of agreements, it supports national Open Science strategies in Latin America and, since its inception, has implemented a platform of technology services with interoperability standards that have allowed sharing and bringing visibility to the scientific production of higher education institutions and scientific research. In this context, scientific production is understood to be not only the literature, but also the data and metadata of the inputs, products, and processes of scientific research.

LA Referencia was born from a cooperation agreement signed in Buenos Aires in 2012 that reflects the political decision to provide open access to the scientific production of Latin America as a regional public good, with an emphasis on the results financed with public funds.

The initiative operates under the principle of a regional federated network where each participating country is a national node that can decide which metadata of its scientific production it will contribute. LA Referencia's central node acts as a regional aggregator, creating value-added services on top of the collection, and interoperating with other regional and global networks, directing access to the sources that originate the content, i.e., the repositories.

The metadata of scientific articles, doctoral and master's theses, and other academic and scientific documents contributed by universities and research institutions of the countries that are currently part of LA Referencia —Argentina, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Spain, Mexico, Panama, Peru, and Uruguay— are thus integrated.

The network is governed by a board of directors comprised of representatives of government science and technology agencies of the member countries, along with RedCLARA. It is an international discussion space with high political legitimacy. Members of the LA Referencia board of directors are elected by the governments of the region, generally from among the country's executives and decisionmakers. Within this space, agreements are discussed and defined in the field of open access and Open Science, promoting common standards and developing technologies that are transferable and interoperable among its partners. As a result, the public ministries and national science and technology organizations of the member countries work in coordination, together with RedCLARA, which actively participates by contributing technology infrastructure, resource management, and experience in the management of regional public goods.

After more than eight years of work, LA Referencia has developed and consolidated its own software platform, LRHarvester, a comprehensive solution that is transferred to each node of the network. Not only does the platform facilitate the aggregation, validation, and enrichment of the metadata, but it has also advanced and now offers value-added services in response to interoperability challenges with other global initiatives such as OpenAIRE. Its new version seeks to satisfy the growing need for integration with current research information systems (CRIS) and the possibility of producing statistics that can be used at decision-making levels.

Born as a regional project, during its almost 10 years of existence, LA Referencia (see Fig. 4) has been consolidated at the international level as the natural environment and reference for the discussion of national and regional policies on Open Science issues and is also frequently cited as a success story of a collaborative network sustained over time despite various differences and political changes.

Fig. 4: Members of LA Referencia.

2.3 Ecuadorian Research and Innovation Network (REDI)

REDI, the Ecuadorian Research and Innovation Network, was created in 2015 as a repository for scientific publications and the CVs of Ecuadorian researchers. In 2021, the need to broaden the network's scope to support the development of the national research and innovation ecosystem became evident. As a result, CEDIA develops the platform using an architecture based on CRIS (Figure 1) which provides global visibility to information on scientific publications, researcher profiles, patents, collaboration groups, trends in research topics, scientific events, and technology services offered by Ecuadorian institutions.

Fig. 5: Ecuadorian Research and Innovation Network.

REDI has records for 170,349 scientific publications by 38,918 authors in 50 higher education institutions, harvested from DSpace repositories, OJS journals and bibliographic databases such as SciELO, Springer, DBLP, and others

The platform complies with the FAIR principles (Findability, Accessibility, Interoperability, and Reuse) using semantic web technologies that facilitate standardized access to, visualization, and reuse of the metadata registered in the platform.

The next step for the REDI project will be to help higher education institutions so they can adopt CRIS systems, which support institutional research and, in turn, provide information to REDI, thus generating statistics and research indicators to support decision making at various institutional levels, and streamline the articulation of the different actors who are part of the research ecosystem.

2.4 Scientific Journals

Given their ability to combine speed in the publication of articles with the evaluation of the articles by specialists in the field, scientific journals are the most important medium for scientific communication. They allow linking authors and content and guarantee the scientific nature of what is published in the shortest possible amount of time by considering peer reviewed papers. In Latin America, these publications typically adhere to the Open Access to scientific information movement and the practice of Open Access, which is considered a highly influential model.

Starting in the late 1980s, the Open Access to scientific information movement grew stronger as a reaction to the exorbitant prices charged by the oligopolies of commercial scientific publishing houses. The prices of scientific journals became an impediment even for the libraries of major institutions, restricting access to scientific publications even for the producers of science themselves. This scenario led to the Gold Open Access model, which promotes the creation of scientific journals that publish articles licensed for sharing and reuse via Creative Commons licenses or similar.

Under the influence of this global movement and faced with the impossibility of paying for commercial scientific journals, many ideas were brought up to solve this problem. In the countries of the global north, seeking to abolish access and reuse restrictions, Open Access journals that charge researchers a fee for processing or submitting articles have become popular. This practice removes restrictions on access and use of articles but restricts the publication of articles in the case of researchers with no funding for their research. It also poses a problem in that it opens up the possibility of the creation of predatory journals that take advantage of the Open Access perspective to obtain easy financial gains by charging for the publication of articles that have not gone through the formal scientific evaluation process.

In Latin America, the practice of Open Access was typically built with institutional funding (mainly public) for the production and maintenance of these publications. This model eliminates the restrictions that limit access and use of the articles, as well as the publication cost that falls on the researcher, as the publications are maintained by the higher education and research institutions themselves. Despite its advantages, the model has some setbacks related to the lack of incentives for national science funding agents, who prioritize traditional, restrictive models to the detriment of Open Science. As a result of this lack of funding, it is difficult for journals to adapt, so journals have limited staff and lack the personnel that their sustainability and prominence within the scientific community requires. Generally speaking, despite usually complying with formal scientific evaluation processes, Latin American journals suffer from low editorial quality, which prevents the indexing of a large number of journals by international indexers and platforms and limits the impact of Latin American science worldwide, undermining its international qualifications.

To solve these problems, initiatives such as SciELO and Redalyc have emerged that seek a better adaptation and increased sustainability of the Latin American Open Access model and the expansion of its capabilities to promote internationalization and the resulting increase on the impact of Latin American science.

2.4.1 SciELO Platform

The SciELO model was developed in 1997 as a collaboration between the São Paulo State Research Foundation (FAPESP) and the Latin American and Caribbean Centre on Health Sciences Information (BIREME/PAHO/WHO). The project was developed within the framework of a special program by FAPESP, with the support of CAPES and CNPq. Originally established in Brazil in 1998, that same year it was adopted by CONICYT Chile, thus starting the SciELO Network of national journal collections which, from its very inception, included an important selection of materials on public health. In 2021, twenty-three years after its creation, SciELO is present in 17 countries (South Africa, Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Ecuador, Spain, Antilles, Mexico, Paraguay, Peru, Portugal, Uruguay, and Venezuela). Journal collections are managed, financed, and developed at the national level, typically with the support of a science and technology organization (science council, ministry of science, or other similar entity). Collections are coordinated by a recognized and highly prestigious institution and developed with the support of a scientific committee representative of each country's community of periodical publications.

The services offered by SciELO include the indexing of journals and other communication objects with an evaluation of bibliometric performance, digital storage and preservation, open access publication, dissemination, and interoperability. The permanent objective is to promote the quality of the articles and other documents contained in the journal collections accessible through the https://www.scielo.org/ portal. SciELO manages preprints through the SciELO Preprints server (https://preprints.scielo.org/) and in the future through SciELO PostPrint, data files, and other materials through the SciELO data repository (https://data.scielo.org/), and SciELO Books (https://books.scielo.org/).

At the same time, the journals published on SciELO are adjusting their editorial policies in favour of Open Science communication practices. The services offered by the SciELO platform follow national and international standards, which are implemented by means of various policies, procedures, and technologies. Adopted standards include the use of the Creative Commons open access licensing system, SciELO Publishing Schema for marking XML documents that follows the NISO Journal Article Tag Suite (JATS), the ISO language and country codes, common Digital Object Identifiers (DOI) for research communication objects, ORCID for researchers, Contributor Roles Taxonomy (CRediT) to identify the contribution of authors of collaborative articles, the LOCKSS digital preservation system through the Carianian Network, and the COUNTER Code of Practice. The application of the SciELO Editorial Model is documented by a series of manuals and guides widely available from its website in Spanish, English, and Portuguese.

The SciELO Network currently manages around 1,300 active journals, which publish more than 50,000 documents annually. The SciELO Network includes a total of more than 950,000 documents, with a daily average of more than 1.5 million downloads as measured by the COUNTER protocol. All of these documents are indexed by Google Scholar, LA Referencia, the SciELO Citation Index on the Web of Science platform, and catalogued in Latindex. In 2020, 50,470 research papers were published, 70% of which are indexed in the Scopus database and 50% in WoS, the bibliometric indexes most commonly used by national evaluation systems. Likewise, in 2020, 30% of the journals and articles in the SciELO Network were ranked in the first and second quartiles of the SCImago Journal Rank (SJR) indicator.

2.4.2 Redalyc Platform

Redalyc is a project based on the model of science as a common good with a scientific communication approach that focuses on the Latin American region. Its purpose is to provide visibility, interoperability, and services for digital production and XML-JATS markup of scientific journals.

Redalyc currently indexes and provides services not only to Latin American journals but to publications worldwide, provided that they are non-commercial open access peer-reviewed journals (diamond open access journals without APC) and pass the scientific and editorial quality evaluation process. Redalyc currently indexes more than 1,400 scientific journals from 31 countries and more than 600 institutions, and a collection that has more than 800,000 full-text articles.

Redalyc envisions an ecosystem where knowledge managed by the academic community circulates freely and outside the commercial market, by means of the intrinsic cooperation between the Green and the Diamond Open Access models. Redalyc has striven to help return the knowledge produced by researchers to their institutions.

Thus, Redalyc has developed processes to standardize the affiliation data of the authors of each article, which allows grouping and recovering the scientific production published in the journals by the authors' institution or country. In this sense, Redalyc has developed a service to contribute to the strengthening of institutional repositories and national repository networks, through the creation of an OAI-PMH access point for each of the institutions with scientific production published in journals indexed by Redalyc.

As a result, Redalyc offerd information from more than 10,000 institutions, available to be harvested by their repositories. This allows institutions to know and recover the scientific production of their researchers published in Open Access journals and use this information to populate their institutional or national repositories. OAI-PMH points offer, for example, more than 92,000 scientific articles from 2,743 Colombian institutions, 274,000 articles from 7,254 Brazilian institutions, 122,000 articles from 2,788 Mexican institutions, and 45,000 articles from 1,516 Argentinian institutions, to name a few.

2.4.3 Latindex

Latindex is the product of the cooperation of a network of institutions working in coordination to gather and disseminate information on serial scientific publications produced in Ibero-America. The idea of creating Latindex dates back to 1995 at the National Autonomous University of Mexico (UNAM) and became a regional cooperation network in 1997.

Latindex has two information products: the Directory, which provides bibliographic and contact data of all registered journals; and the Catalog 2.0, which contains a selection of journals that meet the highest quality standards according to the Latindex methodology. It includes scientific research, technical-professional, and scientific and cultural journals published in Latin America, the Caribbean, Spain, and Portugal. It also offers information on journals with Ibero-American content published anywhere in the world. Journals can be printed or online and relate to any scientific discipline.

2.4.4 Preprint Repositories Dedicated to Scientific Journals

2.4.4.1 EmeRI: Emerging Research Information

The Emerging Research Information (EmeRI) preprint repository was implemented in 2020 through a cooperation between the Brazilian Association of Scientific Editors (ABEC) and the Brazilian Institute of Information in Science and Technology (IBICT), with the aim of providing services to journals and editors in order to accelerate the dissemination of the results of emerging scientific research. EmeRI bases these services on the availability of preprint files.

EmeRI's proposal came about in response to the demands of some Brazilian scientific editors who saw the need to speed up the availability of articles submitted to their journals, especially in the face of the Coronavirus pandemic.

EmeRI promotes barrier-free access to scientific information and supports the openness and speed of the scientific process.

2.4.4.2 SciELO Preprints

The SciELO Preprints Collection is an integral part of SciELO. It is a collection of non-peer reviewed manuscripts within the SciELO Network of national and thematic journal collections. As such, preprints benefit from the same services as journal articles, including indexing and preservation. It runs on top of Open Preprint Systems (OPS) software which is a free, non-profit online preprint file and distribution server developed and maintained by the Public Knowledge Project (PKP). To send preprints, the SciELO Preprints server directly interfaces with SciELO journals, as well as with any delivery system running the SWORD protocol.

2.5 Supercomputing Infrastructure in Latin America and the Caribbean (SCALAC)

The Advanced Computing System for Latin America and the Caribbean (SCALAC, https://scalac.redclara.net) is an initiative that came directly out of the RedCLARA user communities, who, having participated for more than a decade in various infrastructure projects for grid-type distributed computing, brought together their experience to formalize this collaboration. Projects such as EELA, EELA-2, and GISELA allowed not only providing infrastructure but also training human resources at various levels and for various roles: from science users to computer architecture managers, application developers, and researchers working in computer science and engineering.

Progressively, national systems such as Brazil's National High-Performance Processing System (SINAPAD) and other national and university centres have joined to become an open high-performance distributed computing community. Computing projects also joined the initiative, including the Ibero-American Supercomputing Network (RISC), the Ibero-American High-Performance Computing Network (RICAP), and the France-Brazil collaboration, which allowed the integration of Brazil into the French advanced computing ecosystem Grid5000 as well as to PlanetLab.

SCALAC's main objective is to organize and integrate advanced computing capabilities and regional knowledge competencies across the region using the connectivity provided by RedCLARA and the national research and education networks, to provide not-for-profit services that serve common interests but require high-performance computing. It should be noted that much emphasis is placed on training, the dissemination of results, and collaboration among the countries of the region as well as with the rest of the world. SCALAC was officially launched in Colombia in 2013, during the completion of RedCLARA's Grid and Scientific Computing community program. Although it is strategically closely linked to RedCLARA (not only in terms of its interconnectivity but also in terms of its activity), it is an independent non-profit organization legally incorporated in Costa Rica.

The national centres and systems that are part of SCALAC currently provide computing capabilities for researchers in the region whose institutions are part of the national research and education networks (see Table 1).

Other national centres and initiatives have joined the system thanks to the support of national research and education networks (e.g., the Colombian Network for Advanced Computing, with the support of RENATA, Colombia's national research and education network). SCALAC also integrates international collaborations such as RISC2 (a network that supports the coordination of high-performance computing research between Europe and Latin America), led by the National Supercomputing Centre of Spain (BSC-CNS), the Americas HPC Collaboration (with the participation of actors from the United States Department of Energy), and ComputeCanada, the Canadian advanced computing system, among others.

Table 1: SCALAC supercomputing centres interconnected through RedCLARA

Computing Centre	Total computing capacity	Country	Website
Computational Simulation Centre for Technological Applications	48 TFlops	Argentina	https://csc.conicet.gov.ar/
National Scientific Computer Laboratory (includes the Santos Dumont Advanced Computer Nucleus)	5.4 PFlops	Brazil	https://www.gov.br/mcti/pt-br/rede-mcti/lncc and https://sdumont.lncc.br
National High-Performance Computing System (includes the platforms of the federal university system and some other institutions)	1.3 PFlops	Brazil	https://www.lncc.br/sinapad/
National Laboratory for High Performance Computing (NLHPC)	266	Chile	https://www.nlhpc.cl/
Supercomputing and Scientific Calculation, Industrial University of Santander	105	Colombia	https://www.sc3.uis.edu.co/
Los Andes University	48	Colombia	https://www.uniandes.edu.co
National Advanced Computing Collaboratory (CNCA)	85	Costa Rica	https://kabre.cenat.ac.cr/
Abacus: Laboratory for Applied Mathematics and High-Performance Computing	429	Mexico	https://www.abacus.cinvestav.mx

Mexican Supercomputing Network (RedMex via CUDI)	200	Mexico	http://www.redmexsu.mx/
University of the Republic	98	Uruguay	https://www.udr.uy
Ecuadorian Supercomputing Service (via CEDIA)	230	Ecuador	https://cedia.edu.ec/

Source: SCALAC

2.6 Researcher CV Platforms

2.6.1 The Lattes Platform

This initiative originated in Brazil and made Latin America a pioneer in this field, with the creation in 1999 of the Lattes platform (https://lattes.cnpq.br/), an information system that currently (2021) contains more than 7.2 million researcher CVs, 377,000 of which correspond to holders of PhD degrees (with 7,000 researchers having declared higher doctorate degrees) and 653,000 exclusively at the master's level. This information system is developed and maintained by CNPq, the National Council for Scientific and Technological Development.

The content described by the Lattes platform is very rich and became the object of studies with the support of various analysis tools, such as ScriptLattes and LattesDataExplorer. These tools perform "data recovery" and process the CVs hosted on the platform. The platform contains more than 4.5 million journal articles, 10.3 million conference papers, 1.4 million theses and dissertations, 1.2 million book chapters, 520,000 books, 750,000 newspaper and magazine items, 75,000 patents, and 920 technical works.

2.6.2 CVAR Platform

Created in 2011 by the Argentine Ministry of Science, Technology, and Productive Innovation (MINCYT) as an exhaustive survey of the country's science and technology system, the CVar platform currently holds more than 60,000 CVs.

CVar maintains a unified and standardized online record of the curricula vitae of scientists and technologists of all the institutions in the country and is constantly updated. The www.sicytar.mincyt.gob.ar website allows searching by research topic, name of the scientist, area, province, institution, and other variables.

Among other functions, the platform provides statistical information to support planning and decision-making, serves as a tool for institutional requests, publishes curricular information, allows drafting standard CVs that will be accepted by any institution, and promotes the exchange of information.

2.7 Current Research Information Systems (CRIS)

Current Research Information Systems (CRIS) store and disseminate all the information related to the research conducted by an institution and/or country, including the names of the researchers, publications, patents, and data sets they have generated, research projects, and their sources of funding (De Castro, 2019). Their creation began in Europe during the 1990s and euroCRIS9 —the European Organization for International Research Information— was founded in 2002. Having a CRIS-type infrastructure in Latin America would help build proper indicators to assess scientific production beyond traditional publications.

The UNESCO Recommendation on Open Science raises the need to advance in long-term investments linked to digital infrastructure. Priorities include the development of CRIS research management systems. For its implementation, UNESCO recommends that the scientific community should be responsible for the governance of these platforms, while governments should be responsible for their funding.

2.7.1 The BrCris Project

In 2014, IBICT began the process of studying the creation of a Brazilian National Research Information System (CRIS). This study culminated in 2020 with the launch of a research project for the development of a Brazilian national CRIS system, called BrCris (see Fig. 5).

Likewise, in Brazil there is a network of more than 110 institutional scientific publication repositories, 125 theses and dissertation libraries, and 1,000 open access scientific journals. This network of open access sources comes together at the Brazilian node of LA Referencia, the Oasisbr Portal < https://oasisbr.ibict.br/ >, which currently aggregates more than 2 million items and 670 full-text documents. This portal was created and coordinated by the Brazilian Institute of Information in Science and Technology (IBICT).

The Brazilian scientific research ecosystem also comprises other important participants. Examples include the Sucupira Platform < https://sucupira.capes.gov.br/ >, which was developed and is maintained by the Coordination for the Improvement of Higher Education Personnel (CAPES), with data collected for the evaluation of the technical and scientific production generated within the framework of Brazil's academic master's and doctoral programs.

The architecture selected for BrCris aggregates various sources, which are centrally organized in the form of a knowledge network. Like Oasisbr, the software selected for storing and organizing the content compiled for BrCris is the LA Referencia Platform. The platform allows data to be exported to different service and visualization platforms.

Fig. 6: BrCRIS Project.

2.7.2 The PeruCRIS Project

PeruCRIS is the name of the project by Concytec Peru that seeks to establish, develop, and operate the National Information Network on Science, Technology and Technological Innovation (CTI) for the purpose of consolidating and managing scientific and academic information from across Peru, generating statistics to support decision-making at the institutional, regional, sectoral, and national levels, as well as bringing global visibility to the activities, capabilities, and scientific production of Peruvian researchers.

The #PeruCRIS project extends beyond the construction of a computer platform, as it comprises a series of measures and actions for the operation of the National Information Network on Science, Technology and Technological Innovation.

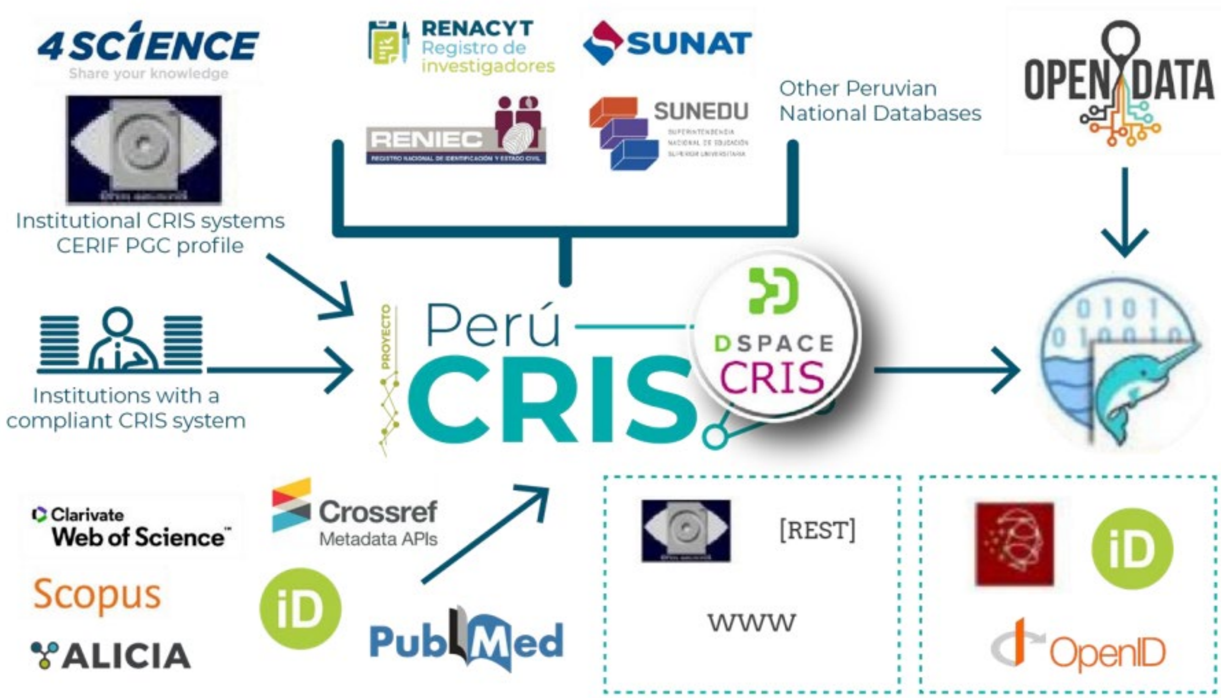


Fig. 7: The PeruCRIS Project

2.8 Services and Platforms for Analysis and Experimentation

The services that an Open Science digital ecosystem must provide to scientists and researchers include open access to platforms, infrastructures, and instruments for experimentation. These infrastructures or instruments range from large astronomical observatories and particle accelerators, through resources that one or several research groups have at their disposal at a local lab, to sensors and devices deployed to obtain data from the environment.

In the case of large infrastructures, the cost of building them means that open and shared access is a necessity. These infrastructures are too expensive not to take advantage of them collectively. In the case of other types of instruments, sharing can lead to a more efficient use of the investment. In both cases, proper digital support is essential for their use.

Moreover, digital transformation processes are creating other data sources that serve as valuable inputs to support research in various scientific disciplines. With the introduction of new technologies such as the Internet of Things, artificial intelligence, blockchain and others, a profound transformation is taking place in terms of the opportunities and ways of developing scientific knowledge.

The IoT, for example, can be used to generate very detailed and granular data from a wide variety of events. In the future, the number of connected devices will grow exponentially, and adequate infrastructure will be required to be able to deploy and leverage this information. The sheer volume of data and the model required to process it has led to the development of new computational paradigms such as edge computing, which offer the possibility of reaching the computational scales needed to receive, process, and take action based on the collected data within the required response times.

In turn, this ability to collect data through IoT devices is directly linked to the deployment of 5G and Wi-Fi 6 services, two technologies that modify the connectivity architecture, taking processing capabilities to the edges and bringing computing capabilities closer to the devices that need to be managed or from which data must be collected. Thus, if we think of a very large environment both in terms of surface and number of sensors, we will need a computational infrastructure that allows conducting different analyses in different layers, for example, in an IoT environment:

- Analysis in the sensor and edge computing (ARM, FPGA...), for example, for a first data calibration.
- Fog computing, see computing clusters for a first analysis of previously filtered data.
- Larger supercomputers to perform analyses based on more precise and computational capacity and/or memory demanding calculations.
- This layer should already have the additional storage infrastructure that may not be needed in the previous cases, as the data is deleted after their processing and being sent to downstream servers.
- Implementation of containers adapted to each type of analysis conducted on virtualized environments.

Complementing this, one might consider large-scale experiments that continuously provide data, both experimental and synthetic, based on simulations. In this case, an infrastructure that integrates different capabilities would also be needed:

- Data processing and cleaning tools.
- Connection with parallel and distributed infrastructures for storage and simulation.
- These types of platforms could also include *ad hoc* virtualized environments.
- Categorization of data at different levels for their subsequent leveraging and temporary embargo by those participating in the collaboration.

As can be easily concluded, both IoT infrastructure and the infrastructure for large-scale experiments are not disjoint sets and both may require capabilities initially designed for the other for reasons of each one's specific data generation environment.

Thus, artificial intelligence, big data, and computing capabilities are having a significant impact on the productivity of researchers (and even of regular citizens, who have increasingly greater access to data from any scientific, technological, or social field). Ideally, a researcher should have high-bandwidth connectivity to infrastructures and devices, regardless of whether they are large and scarce or small and plentiful. These connections combine various communication technologies, ranging from fibre optics for high-capacity connections, to wireless technologies such as 5G or Wi-Fi 6. This connection must allow the researcher to access and manipulate the instruments associated with the experiment, as well as to collect, clean, and store the data produced in repositories intended for this purpose. Computing tools and access to algorithms and processing techniques, incorporating new technologies such as artificial intelligence that facilitate the processing of increasingly large volumes of data. Examples of this type of environment currently in operation include the test bed services offered by the Brazilian National Research Network (RNP) (see https://www.rnp.br/servicos/testbeds).

3. Some Preliminary Conclusions

3.1 Impact of the Digital Transformation

As the Digital Economy Report published by UNCTAD in 2021 indicates, emerging digital technologies are built on digital data. Highlights of these technologies include data analytics, artificial intelligence, block chain technology, the Internet of Things, cloud computing, and all internet-based services.

The relevance that these technologies as a whole are gaining in terms of trade balances, innovation, and economic development makes it essential for the different countries to discuss their regulatory implications. They also have implications in terms of equity and national security.

In addition, the COVID-19 pandemic greatly accelerated the digital transformation processes: in 2020, global internet bandwidth increased by 35%, and this trend is expected to continue in the coming years. This acceleration accentuates a series of imbalances in terms of equity at the global level. In this sense, the Digital Economy Report (UNCTAD, 2020) notes that only 20% of people in least developed countries use the Internet; when they do, they face major difficulties in terms of poor quality of service at relatively high costs compared to those paid by users in developed countries. Likewise, rural populations and women exhibit higher levels of exclusion.

The United States and China are among the countries with the highest digital technology utilization indicators. Together, the two countries account for 50% of the world's data centres, the highest rates of 5G adoption, 94% of all funding of Artificial Intelligence start-ups, 70% of the world's Artificial Intelligence researchers, and 90% of the market capitalization of digital platforms (UNCTAD, 2020).

As for Latin America and the Caribbean, although the use of digital data has increased considerably just as in the rest of the world, today the region is facing important challenges to reduce the digital divide. According to World Bank data, approximately 200 million people in the region lack basic digital infrastructure, while many others have some basic infrastructure, but with insufficient quality of service and higher costs compared to other regions. Just as it did worldwide, the pandemic accelerated the digital transformation process; however, this process moved forward while access gaps continued to grow.

There are several reasons why these gaps continue to grow: the high cost of connectivity in relation to household income, directly related to the lack of competition among communication service providers; complex regulatory frameworks; a lack of telecommunications infrastructure, and inefficiency in the provision of services.

The inequity in reaping the benefits of the digital transformation processes means that it is essential to develop regional policies that will lead to an increase of digital infrastructure, data centres, etc. Multilateral funding organizations can play a key role, becoming strategic partners for the private sector to be able to assume the risk of the investments needed to improve the services associated with digital transformation, especially in rural or geographically difficult-to-access areas.

3.2 International Cooperation in Open Science

Throughout the document, various collaborations articulated through different projects have been mentioned (see BELLA in Section 2.1, OpenAIRE in Section 2.2, EELA, CHAIN or RICAP in Section 2.4, to name but a few) where cooperation with European, North American, African, or Asian organizations has represented a significant step in the development of Open Science in Latin America.

Thus, it is obvious that an essential element for developing an Open Science system in the region is cooperation with funding agencies and key institutions in other regions of the world. In the field of advanced networks (Internet2, Geánt, Ubuntunet Alliance, RedCLARA) there has always been fruitful and lasting collaborations that must now be extended to initiatives such as the Research Data Alliance, the European Open Science Cloud, EuroHPC, etc., to allow the exchange of knowledge and, ultimately, the creation of a global Open Science ecosystem.

For historical reasons and given the success that has been achieved, the European Commission, through its Framework Programs (6th Framework Program and Horizon 2020 in the past, Horizon Europe from 2021 onwards), has been a driving force in the development of advanced infrastructures in Latin America and its role for the future would appear to be key.

Beyond potential calls for research and innovation actions (RIA) or coordination and support actions (CSA) that can fund the various advances identified in this document, coordination with the Commission itself through its General Directorates can be the first relevant step to achieve the agreed objectives. Only by defining an ambitious yet realistic roadmap will the different actions here promoted be successful.

3.3 Data Infrastructure, a Major Absentee

In Latin America, there is no regional infrastructure to store scientific data. Each country and, in some cases, each university or research centre deploys its own solutions. This leads to a fragmented approach that hinders the development and implementation of the FAIR principles listed earlier in this document (Findable, Accessible, Interoperable, Reusable).

What is the advantage of having an infrastructure that promotes these principles for the data? By having these properties, the data produced by the different scientific communities become a common resource that can be used by researchers, eliminating the duplication of efforts and the inefficient use of the scarce resources that are generally assigned to science in Latin America.

With a common infrastructure such as the one we have described, it would be easier for investments in scientific data collection and processing to contribute at the regional level, indirectly expanding national investments by giving them access to a greater amount of evidence and resources for the same cost.

The proposal, therefore, is to implement an architecture such as the one proposed in various scientific and technological fields, identifying those that are initially better prepared so that they can serve as an example for other communities to adopt a similar system for their specific architecture.

4. A Possible Roadmap and Some Recommendations

Taking into consideration all of the above, a series of recommendations are succinctly presented below that we hope will contribute to the effective implementation of a digital Open Science ecosystem in Latin America and the Caribbean.

4.1 Expand the connectivity of the organizations that are part of the Latin American Open Science ecosystem

Universities, research centres, science and technology organizations, and other actors that are part of the scientific ecosystem must be well connected with each other and with their peers. Data connections that can properly support science require high bandwidth, low latency, high levels of cybersecurity, and interconnection with similar characteristics to other research and education networks worldwide. There are significant digital divides in Latin America that deepen the knowledge gap with respect to the leading countries in scientific research. Appropriate connectivity is only available to part of the ecosystem, which leaves out a significant set of organizations that are unable to connect with the required capacity. Joining the efforts of governments, international cooperation agencies, development banks and other entities to emphasize the creation of infrastructures dedicated to science is a very important step. The regional ecosystem of organizations has very relevant assets that must be coherently integrated into a continental strategy. RedCLARA has proposed a vision of connectivity for the year 2030. With the support of the European Commission and the national research and education networks of Latin America, the aim is to connect countries that are not yet connected to the BELLA infrastructure. It is important to complement these efforts and generate synergies to achieve maximum effectiveness and efficiency in the use of the resources available to these projects.

4.2 Advance by integrating the lessons learned from successful projects that have managed to build important capabilities in the region

We must integrate the lessons learned from projects that have been successful in developing capabilities. The most recent example are the achievements of the BELLA project, which has deployed a new submarine cable connecting Europe and Latin America in which a very significant amount of dedicated capacity has been reserved for science and education and which represents a means for integration with European digital science and education ecosystems. Other projects such as Amlight have built capabilities by integrating advanced connectivity with the United States and Canada. In the future, these projects may serve as a reference for the plans to deploy the Humboldt cable between Chile and Australia, integrating the continent with the science and education ecosystems of Asia, or the plans to deploy direct connectivity to Antarctica, an achievement that would also represent a major contribution to the global scientific community.

4.3 Take advantage of the services developed by the large-scale scientific experiments of the region

Analysis of the capabilities that the large-scale experiments conducted in the region (ESO, PAO, LAGO, etc.) can offer beyond the mere use of the connectivity provided by the cable between Latin America and Europe, as they have probably already integrated FAIR and simulation capabilities in virtualized environments.

4.4 Expansion and strengthening of the services offered by LA Referencia

Analysis of the capabilities that the large-scale experiments conducted in the region (ESO, PAO, LAGO, etc.) can offer beyond the mere use of the connectivity provided by the cable between Latin America and Europe, as they have probably already integrated FAIR and simulation capabilities in virtual environments.

4.5. Selection and implementation of use cases

Selection of use cases implemented at the continental level that may serve as the first users of the proposed digital ecosystem and that complement those of item 3 above. Candidates might include:

- Areas of science related to Bioinformatics, such as the activities developed by the CABANA project (https://www.cabana.online/es).
- Areas related to teaching, with courses covering topics of regional interest, such as those addressed by the project for training in High Energy Physics, LaCONGA (https://laconga.redclara.net/), or Energy with initiatives such as the United Nations Global Network of Regional Sustainable Energy Centres GN-SEC, UNIDO (https://training.gn-sec.net/login/index.php).

4.6 Human resources

Analysis of the human resource capabilities needed from the main actors to properly estimate the next steps: RedCLARA, LA Reference, SCALAC, users of the ecosystem.

4.7 Support from multilateral development banks

The support of multilateral banks for regional projects and organizations has been very limited due to the characteristics of these organizations. It is important to structure international funding schemes in the fields of science so that the necessary synergies can be generated, not merely strengthening national capabilities, which in isolation or poorly integrated, are insufficient.

4.8 Furthering cooperation with Europe and its digital ecosystems

There is a rich history of European Commission support which has enabled the construction of a significant part of the required regional infrastructure. It is important to further these ties and advance in the search for all possible supports, whether through agreements with large European infrastructures and research centres, their computer networks, their centres of excellence in digital transformation technologies, and funding programs and calls for projects within the context of the Horizon Europe program. In this sense, it is important to identify projects by Horizon Europe, Digital Europe, and others that are likely to contribute solutions, ideas, and capabilities to the Latin American Open Science digital ecosystem.

References

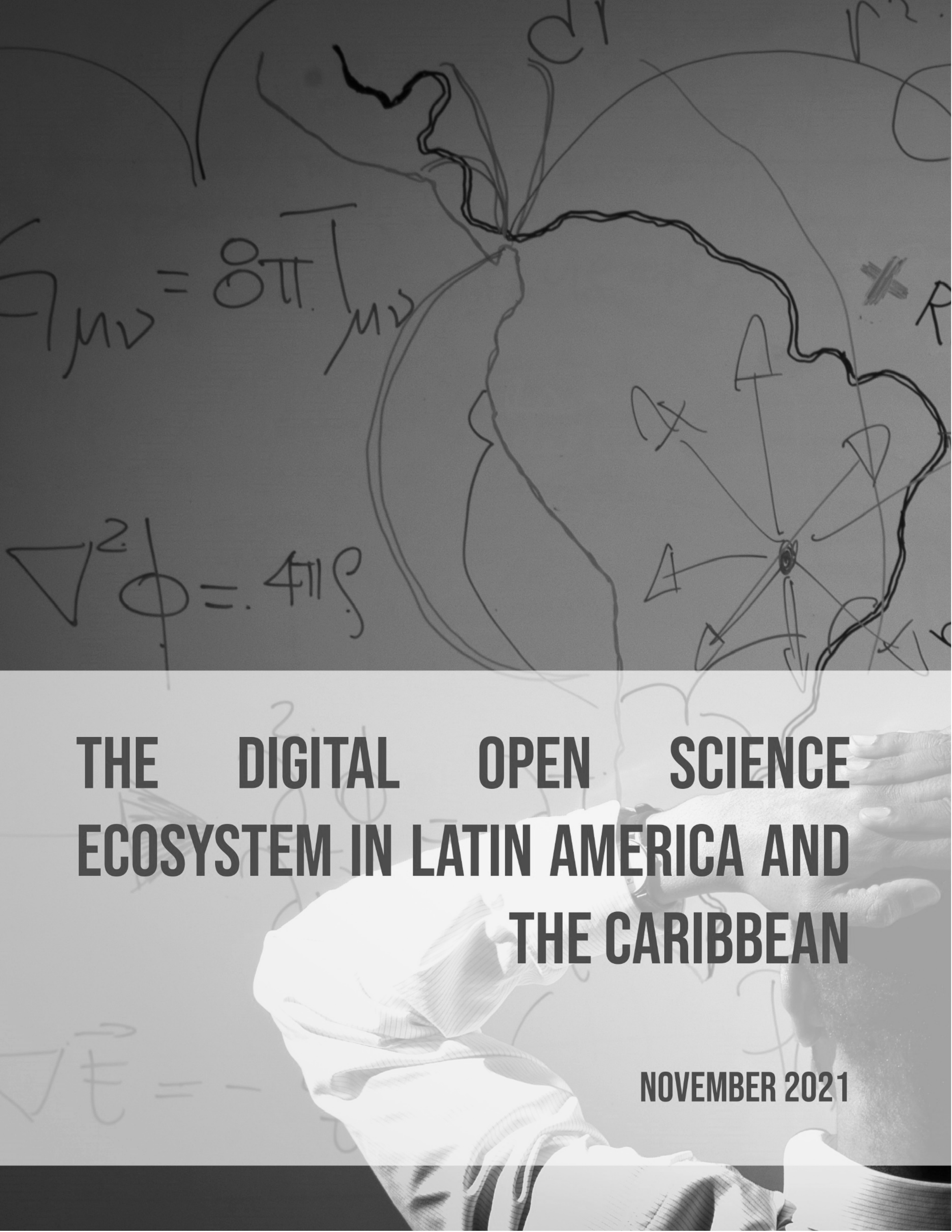
[1] M. Woelfle, P. Olliaro, and M. H. Todd, "Open science is a research accelerator," *Nat. Chem.*, vol. 3, No. 10, pp. 745-748, Oct. 2011, doi: 10.1038/nchem.1149.

[2] Unesco, "Proyecto de recomendación sobre la ciencia abierta." UNESCO, Sept. 08, 2021. Accessed: Nov. 10, 2021. [Online]. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000378841_spa

[3] Carlos Seaton and Luis Cadenas, "Bella 2030: Alianza digital entre América Latina, el Caribe y Europa." Jun. 01, 2021. Accessed: Sept. 06, 2021. [Online]. Available at: https://www.redclara.net/images/docs/BELLA2030-Alianza-Digital-ES.pdf

Acknowledgments

We would like to thank REDALYC, SciELO, and Latindex for their important textual contributions during the preparation of this document.



THE DIGITAL OPEN SCIENCE ECOSYSTEM IN LATIN AMERICA AND THE CARIBBEAN

NOVEMBER 2021