

Accelerated
Data
Program

2006 - 2015

Latin America
Regional Assessment

Contents

Executive Summary	1
ADP Assessment	4
Contact Information of ADP Focal Points in Latinamerica	23
Partnership Information	23

The main objective of this report is to assess the quality and sustainability of ADP activities in the 5 countries selected in the region.

Those countries are:

- Bolivia
- Colombia
- Costa Rica
- Ecuador
- Perú

Executive Summary

Statistical development varies greatly in the Latin America region. On one part of the spectrum is Mexico, with a well developed and integrated statistical system and on the other are the small-islands (SIDS) in the Caribbean that have challenges unique to SIDS. Countries such as Colombia and Costa Rica have a strong drive to align reporting and statistical systems to correspond to OECD rules (Mexico is already a member). Regional integrating bodies such as the Andean Community (CAN by its spanish acronym) and the Central American Integration System (SICA by its spanish acronym) are developing greater roles in affecting and influencing the modernization of statistical systems. Demand for data in region varies but research institutons and think tanks such as CEPEI are at the forefront of using data in creative ways and becoming an influence for innovation in the region.

Out of 20 countries in continental Latin America, the ADP has been active in 17 of which 14 are considered currently active (defined as maintaining a National Data Archive). The Accelerated Data Program (ADP) began its activities with Mexico and most recently has provided some assistance to Brasil and Chile in developing their national data archives. The ADP has worked through regional organizations in the Caribbean with the exception being the Dominican Republic, where the ADP has a program of support and a regional workshop in Trinidad and Tobago. The ADP conducted a total of 48 activities in the region and over 700 persons trained in 176 institutions of which 44% were women. The pie chart illustrates how those activities were divided by country.

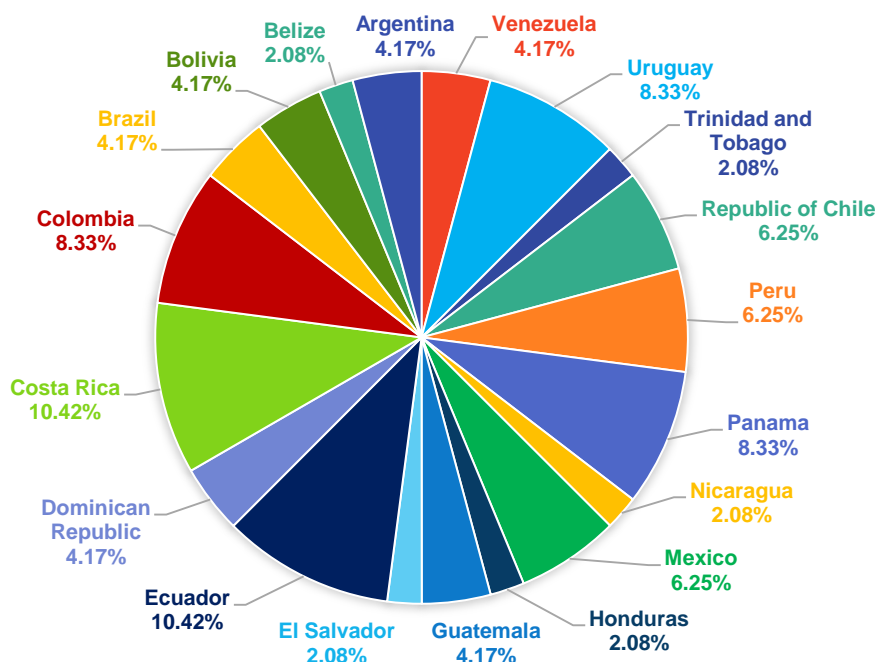


Figure 1: Distribution of ADP Activities on the Region

The activities of the ADP are summarized in the input and output visualizations below.

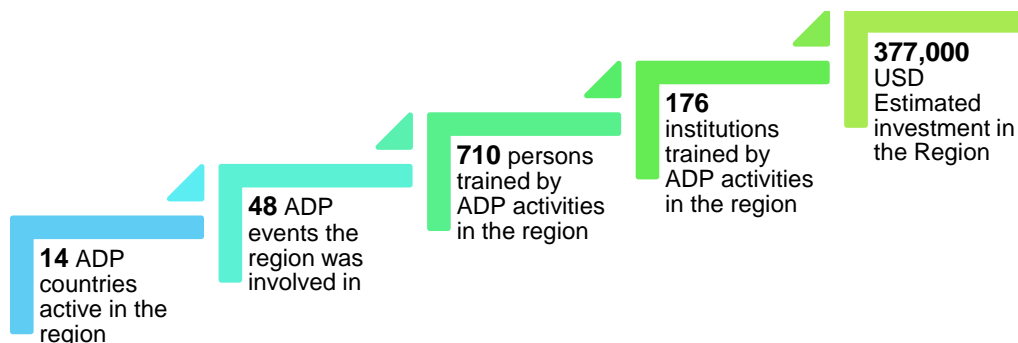


Figure 2: Inputs for implementing the ADP Initiative in the Region

The hexagon cluster below provides the summary of the outputs of the ADP in the region. These are the primary outputs as defined by the goals and objectives of the ADP.



Figure 3: Outputs of the Implementation of the ADP Initiative in the Region

ADP Assessment

The ADP assessment in Latinamerica reviewed the following 5 countries: Peru, Bolivia, Colombia, Costa Rica and Ecuador. The consultation workshop took place in Quito, Ecuador from 26-28 October, 2015.

This report is organized in the following manner:

- Metadata Assessment
- National Data Archives
- Progress in data dissemination and legal constraints
- Institutional commitment and user management

These four areas form the basis for defining key performance indicators which provide a relative measure of how countries performed in the region. The primary output monitored are the number of available studies on the National Data Archives. The figure below provides a view of the outputs per country evaluated.



Figure 4: Number of statistical operations documented and available on line by country (September 2015)

A final section will review the the data priorities in the changing environment provided by the need to respond to the calls for measuring sustainable development as defined by the Sustainable Development Goals (SDGs) and a look into the innovative approaches in the

data process reviewing the statistical processes as defined by the Generic Statistical Business Process Model (GSBPM).

The report also provides a series of recommendations and an estimate of the costs required to sustain the process of data development; catalyse change and assure higher quality data is made available at the right time.

Following the regional review, a more detailed country section is provided.

Section 1: Metadata Assessment

Metadata are descriptive elements that help researchers and policy makers assess the quality of the data and undertake more effective research. Different standards have been developed by different user communities and in the case of the ADP, the standard which was introduced was the Data Documentation Initiative (DDI). The DDI is an international effort to develop a standard for describing social science data. It is a standard that is designed to be interoperable and exchangeable for both human and machine exchange. The concept of DDI and documentation are used interchangeably. The metadata assessment evaluated the quality of the “documentation” from the standpoint of the utility of the information to a researcher. Details of the processes undertaken during the review are provided in the annex.

Table 1 below provides the results based on the standard Metadata Quality Assessment tool which was used to perform the evaluation. The scores below are based on a scale of 0-100 and evaluate the quality of the information under the primary headers provided.

Table 1: Summary of the DDI Quality Assessment Checklist on the 5 countries

	Colombia	Costa Rica	Ecuador	Perú	Bolivia
Average Score on Metadata Quality	81.4	84.8	67.4	78.0	75.2
Elements Evaluated for each study (Survey)					
• Identification of the study	94.1	97.0	97.5	99.8	90.6
• Version maintenance of the published document	83.3	82.1	76.6	83.1	83.3
• Overview of the study	99.1	98.2	98.3	100	98.7
• Geographic Coverage	98.1	91.0	100	67.0	79.6
• Producers and Sponsors	97.4	100	100	99.6	97.8
• Sampling	90.6	95.5	86.1	82.7	86.8
• Data collection	55.0	77.5	60.9	57.3	71.1
• Data processing	93.4	96.0	82.5	80.5	80.4
• Data appraisal	76.8	95.5	96.9	98.2	66.7
• Data access	28.4	42.0	76.7	43.3	48.1
• Data files	85.7	94.5	43.6	92.7	96.3
• Variable description	99.6	94.4	42.0	96.5	88.8
• External documents (survey questionnaires and reports)	46.3	65.7	95.3	48.9	28.2
• External documents that provide further explanation	94.5	89.3	91.0	58.4	55.0
• Metadata citation and use	75.1	66.9	50.2	47.3	51.3

The region overall ranks high in the quality of the metadata produced with an overall quality score of 77.36%. Costa Rica ranked first in the regional ranking of the assessment achieving an overall score of 84.8%. They scored high across all categories.

In the figure below we can see a relation between the number of projects documented by each country and their general range of scores for those outputs. Colombia, Costa Rica and Peru in general are close to 80% score in average, whereas Bolivia is near the 75% and Ecuador closer to 55%. This was due to greater inconsistencies of their metadata quality resulting in greater variation. Colombia is a good example of systematic quality metadata having a small range of variation with high average scores.

For the rest of countries assessed, Costa Rica, Perú and Bolivia, share a similar dispersion in their metadata quality but the lowest quality score in Costa Rica is still higher than the average of Peru and of Bolivia which is consistent with previous findings. Outliers are given by trailing points.

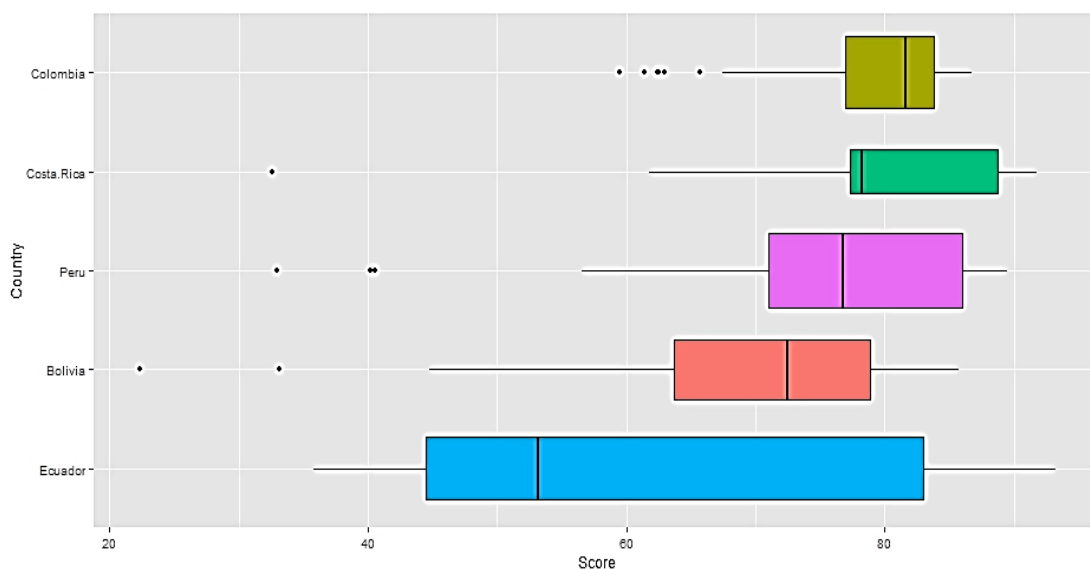


Figure 5: Range of scores of evaluated studied by country & score

Section 2: National Data Archive (NADA) Assessment

Through the International Household Survey Network (IHSN), the ADP promoted the use of an on-line archiving software known as the National Data Archive (NADA). The assessment of NADA catalog site was conducted through a comprehensive review evaluating items such as:

- Visibility of the catalog: Is the site well integrated with the primary website and does it have clear links and descriptions.
- Registration: The NADA requires users of data sets to register. This process requires an email exchange which is done at a machine level. The functionality of the registration is critical to maintaining users.
- Search and filters are key in browsing the archive and help present information to the user in more efficient ways.
- Data Dissemination Policy: This should be available and integrated into the NADA so researchers can quickly understand the processes of accessing data.
- Citations: The NADA has a functionality to track publications that are using national survey and administrative data. This functionality is considered important for monitoring use.
- Innovation measures innovative and creative presentation of the NADA to the user through on-line tools (such as a youtube clip)

The following table shows the average of each of the seven categories. The scores below are based on a range from 0-100.

Table 2: Summary Table of NADA Catalogues of countries in the Region

NADA Assessment category	Colombia	Costa Rica	Ecuador	Perú	Bolivia	LAC Score weighted over 100
TOTAL NADA Assessment Score	60.0%	52.9%	53.3%	44.4%	31.8%	48.5%
NADA visibility	50.0%	7.5%	45.0%	60.0%	11.8%	34.9%
NADA Registration	100.0%	100.0%	75.0%	75.0%	75.0%	85.0%
Search and Filters	100.0%	71.4%	75.0%	71.4%	87.5%	81.1%
Data dissemination policy	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Data access	100.0%	100.0%	100.0%	50.0%	50.0%	80.0%
Citations of surveys in publications	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Innovation	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Days to respond a data request	100.0%	100.0%	100.0%	100.0%	6.7%	81.3%

KEY FACTS:

- 100% of the institutions considered that the NADA Catalog portal is beneficial or helps researchers, but receive only 40% of requests for access to microdata because at least half of the microdata documented is not available online.
- 80% of institutions said they have a Microdata Dissemination Policy but only 60% have it available online in the main NSO website and no country has it linked to their NADA website.
- NADA catalogs of all the countries assessed are housed in the NSO Server.

Main Findings

- In general, the overall quality in the National Data Archives wasn't the best mainly to its recent implementation in the region. The average score for LatinAmerica is 48.5%.
- The element which accounted for the greatest loss of score was the absence of data dissemination policies. None of the countries monitored survey use by maintaining a catalogue of citations; innovative on-line promotion of the NADA was also not readily available.
- The categories that excelled were the registration process, search & filters and data access, although at least 50% or more microdata is still not available in the region through the NADA application online but it is accessible through alternative channels within the same institution.

The radar chart below provides a quick view of the regional averages of the NADA evaluation scores. The shaded box inside represents the relative size of the weighted scores (the outside "footprint"). For the evaluation a great deal of weight was placed on access to data.

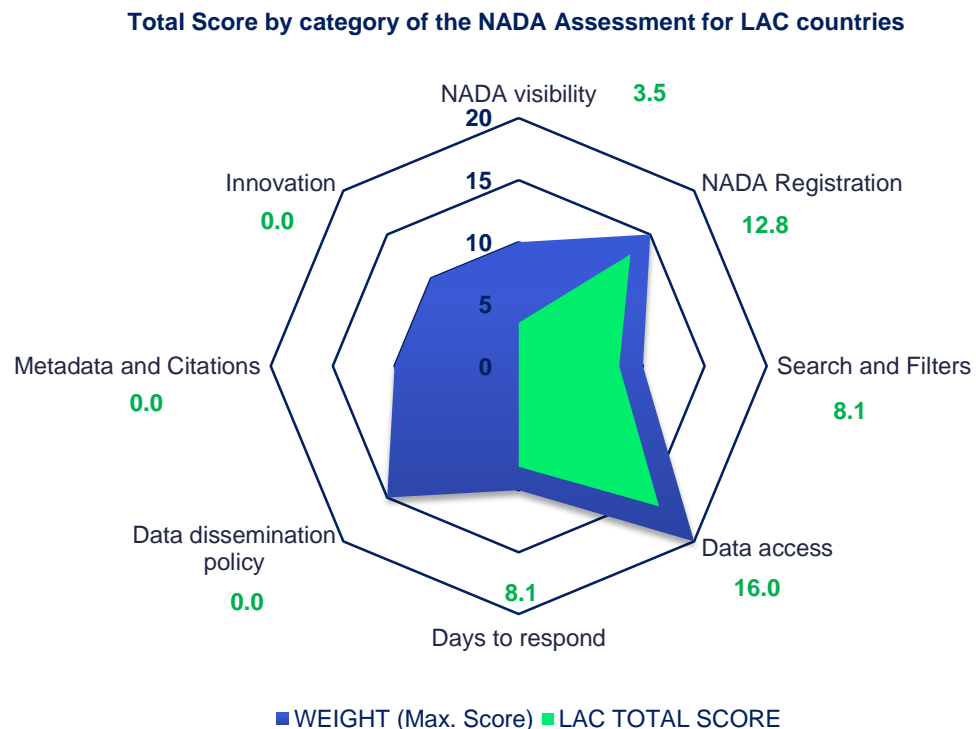


Figure 6: Total Score by category of the NADA Assessment for LAC countries



In many countries there is not an official dissemination policy in place. Nonetheless, most work with a "Best Practices" codebook instead. The processes of the dissemination policy are not in place except for Costa Rica

NADA Checklist findings:

- There are plenty of studies or projects posted on the NADA that do not correspond to microdata based surveys or administrative data. They are aggregated data or indexes that shouldn't be documented with the DDI & Dublin Core Standard. These would be better placed in a publication catalogue or a file management system like CKAN.
- NADA Catalogues of the assessed countries don't share the standard configuration found in catalogs from other regions (mission and objectives, quotes, policies and procedures, new products, innovations and contacts). More content could be provided so a user does not have to toggle out of the catalogue.
- In many cases customization of the NADA look and feel was not consistent with the look and feel of the NSO web site.
- A prominent link from the NSO home page to the NADA catalog was hard to find. A researchers should be able to quickly find the easy links to the NADA site.

Section 3: Status on Dissemination

Data Access

The National Data Archive – NADA Catalogue has four types of data access:

- Public use files: Microdata is downloadable through an on-line process
- Licensed files: Some data require licensing procedures and more in-depth reviews
- Data Enclave: Data is available on-site in controlled environments
- Data not available: No data is downloadable

It is important to note that in many cases, although the microdata is listed as "not available" there are other channels of delivering microdata access not specifically defined in the NADA system. These include: data for sale, or walk-in requests for data. Since it wasn't possible to reflect those other channels in the evaluation scores, it's important to acknowledge that the following scores correspond only to the online channels provided by the NADA.

The pie chart below provides an illustration of the kinds of data access available through the NADAs. About one-third of the documented datasets are readily available through the NADA catalogue.

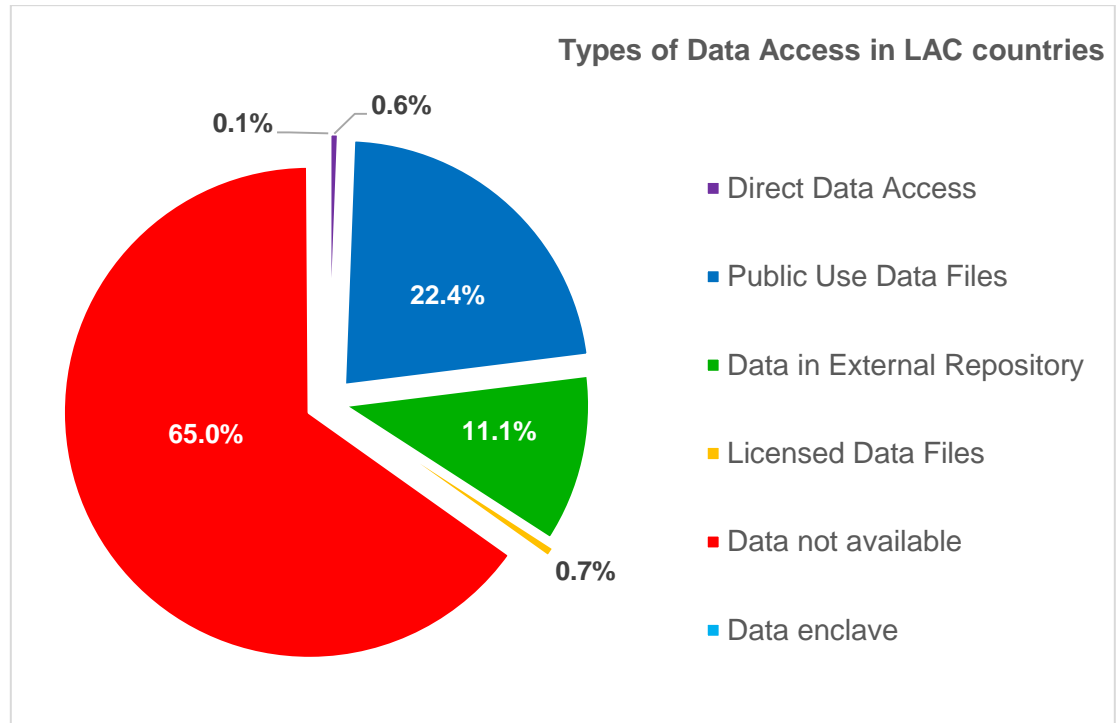


Figure 7: Types of Data Access in LAC countries

Status on the Dissemination Policy & Statistics Act

With the exception of Costa Rica, who recently published their dissemination policy & implementation in June 2014, the dissemination policy in many countries is not official. There are internal guidelines, codebook of best practices and a publication calendar that provide some dissemination information. The clear implementation is still lacking and data access is mainly done on a case by case basis.

The table below provides a review of the status of the data dissemination policy evolution in each country.

Table 3: Status on the Dissemination Policy & Statistics Act

LAC COUNTRIES	COLOMBIA	COSTA RICA	PERU	ECUADOR	BOLIVIA
Statistics Act allows Microdata Access	YES	YES	YES	NO Law under review	NO Preliminary bill
Is the Statistics act available on line	YES Decree Law March/2015	YES	YES	NO	NO
Dissemination Policy allows Microdata Access (or similar guidelines)	YES	YES June/2014	YES	NO	NO
It's for the NSO or NSS?	NSS	NSO	NSS	NSS	NSS
Is dissemination policy available on line?	NO	YES	NO	NO	NO
Regulation of Dissemination Policy includes Calendar of Publications, Type of Users, Type of Data Access per survey, NADA & DDI Standard	NO	YES	Codebook of Best Practices on the stage of implementation in the NSS	YES, partial & under review	Only Calendar of Publications
Other channels to Microdata Access available	YES	YES	YES	YES	YES
% Data Accessible through NADA Catalogue	45.1%	17.4%	56.2%	51.5%	2.4%
Progress of Dissemination Policy (if not approved or updated)	80% (Regulation in process)	100% in NSO pending NSS	60% (Implementation process)	60%	60% (Drafting process)

Section 4: Institutional Progress & Relation with Users

As part of the assessment a survey was sent to National Statistics Offices to elicit responses and identify areas where there could be relative strengths and weaknesses in the institutional processes. The responses to that survey provide the content to this section. The NSO survey covered areas such as:

- Institutionalization of data documentation activities

- Undertaking data curation activities¹
- Regional and sub-regional support for data support activities
- Relationship with users

ADP INSTITUTIONALIZATION

- 3 out of 5 of the NSOs (Colombia, Costa Rica, Perú) said the NSO trained staff have enough experience and skills to facilitate a workshop documentation data alone (without consultants)
- 4 countries reported their NSO staff had acquired the skills and expertise to manage, maintain and update the NADA Catalog independently (Colombia, Costa Rica, Bolivia and Ecuador).
- Only 1 country (Colombia) reported they allocate funds for microdata documentation and dissemination and are self-reliant.
- 4 countries (Colombia, Costa Rica, Bolivia and Ecuador) have included the activities of documentation & dissemination in their staff job descriptions. These are considered mandatory activities hence, are also considered as budgeted. Peru notes that exceptionally temporary documentation consultants can be hired.

PROCESS OF DOCUMENTATION AND DISSEMINATION IN THE REGION

- The only country that documents regularly is Costa Rica.
- Documentation tends to be done centrally. Decentralization and review of the quality can be done institutionally by contributions from special subject matter people and a central quality review.
- There is high staff turnover in some institutions which require a constant process of training.
- There tends to be a time lag between documenting the survey and posting it on the archive.
- In general, the NSOs in the region do not have the authority to share the databases of other members of the National Statistical System (NSS) and vice versa.

¹ Data Curation is defined as an institutional approach to data documentation that involves a more in-depth management of data and metadata within the archiving and dissemination processes of the NSO.

- Only Colombia has a distributed NADA. This is an architecture where various data producers in the NSS can post to the central NADA and distribute data as per their own policy.

REGIONAL AND SUBREGIONAL SUPPORT

- 3 countries were aware that there are consultant resources someone available to assist countries in microdata documentation and NADA management
- All countries responded that there is no regional organization working on data curation that could provide technical support (see more in Section 6).

RELATIONSHIP WITH USERS

- Only Colombia organizes user exchange workshops such as: dissemination and awareness events and microdata outreach to users. These are regularly budgeted each year.
- The only method under which the NSOs in the region generally receive feedback from its users is through e-mail.
- Only Peru and Colombia have specific procedures to take into account the feedback from users designed to improve data quality and data dissemination service.
- Only Colombia and Costa Rica reported a technical relationship with users.

Section 5: Key Performance Indicators

The Key Performance Indicators (KPIs) are designed to provide an overall score based on 4 different criteria each equally weighted. The figure below provides a summary of the components of the KPIs used to determine a final score.

The KPIs for LAC countries show:

- Colombia has the best overall performance in all areas with 92%.
- User Management is an area that all NSO struggled with, specially Bolivia since it just published its NADA Catalogue.
- The documentation activity is included in NSO budget in the sense that is part of the staff job description, but not separately.

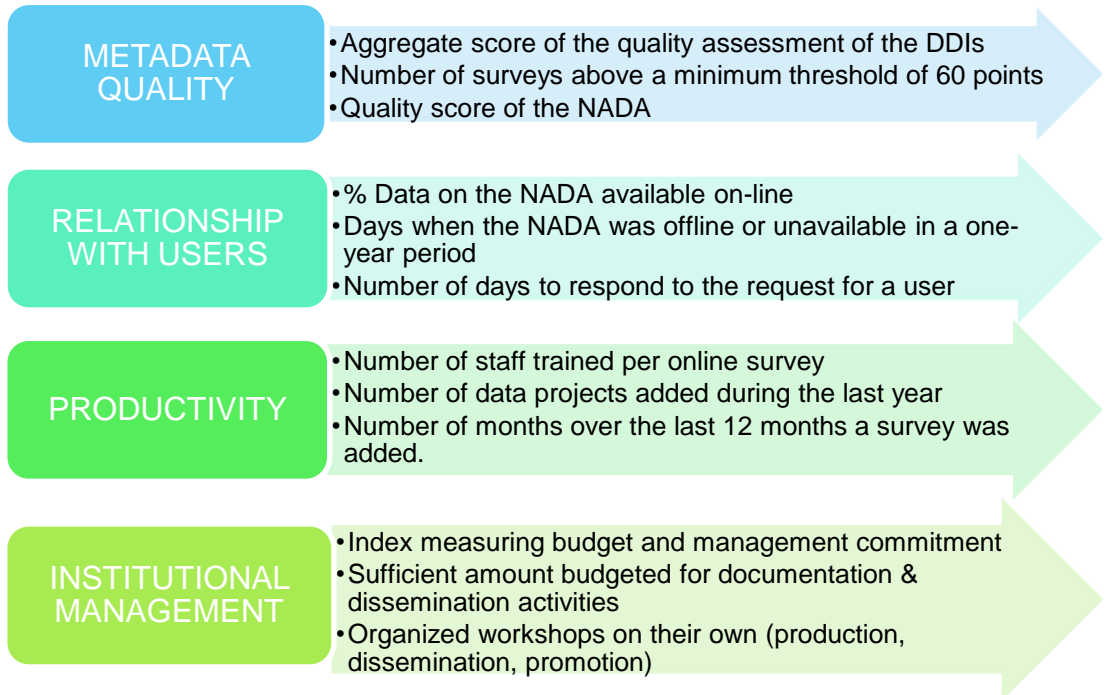


Figure 8: Concepts that conform the Key Performance Indicators

The figure below provides a review of the KPIs broken down by their main component.

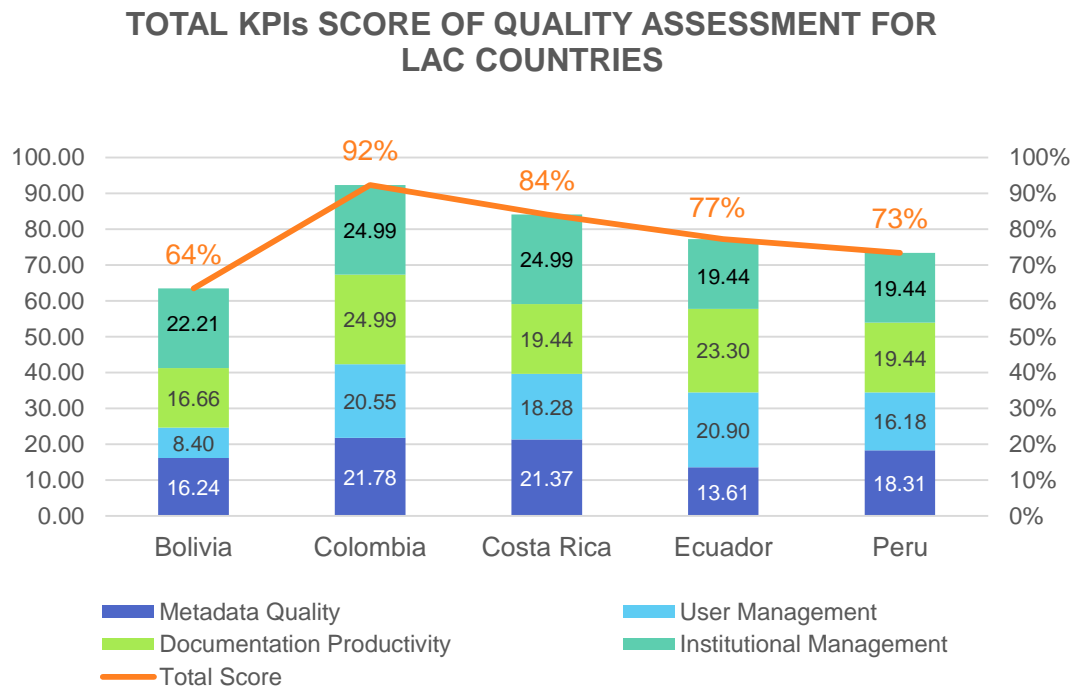


Figure 9: Key Performance Indicators for LAC countries

As can be seen from the figure above, Colombia can be considered the highest performer. However the differences between the top countries: Colombia, Costa Rica and Ecuador are not that great. These three countries can be considered to have adopted the tools and standards with greatest commitment. Bolivia can be seen as the country with the greatest room to improve, particularly in terms of user management, metadata quality and productivity.

Section 6: Innovation & Data Revolution

Tied very closely to the processes of data documentation and dissemination is the ability for an NSO to innovate and adapt to the changing data environment. Indeed, it could be said that the KPI discussed in Section 5 are proxies of a country's likelihood and ability to be innovative and open regarding data dissemination. As part of this workshop, two days were taken to assess the NSOs attitudes and activities under the Informing the Data Revolution Project (IDR) undertaken by PARIS21. As part of this two day assessment, a special

questionnaire was sent to the participating countries in order to assess the “bottle necks” or areas of resistance in improving the data process. This was not specific to microdata but rather evaluated all processes in an NSO using the Generic Statistical Business Process Model (GSBPM). The following are the conclusions:

1. General

The 5 participating LAC countries were a broadly homogenous group in terms of complexity of statistical systems throughout each NSS, statistical system platforms, IT infrastructures and statistical software and DBMS used and overall capacity. The group ranged from fairly advanced (Bolivia, Costa Rica, Peru) to very well advanced (Colombia, Ecuador) in terms of overall infrastructure and systems and could be compared to some European countries. The LAC group were clearly more advanced than other regions where the same assessment was undertaken.

2. NSS Integration and Coordination

Integration within the NSS tended to be reasonably well coordinated by the NSO although few standards were adopted throughout.

3. Regional support

It was agreed there is a need for regional training programmes and expertise, workgroup, knowledge transfer and also to keep abreast of the latest developments in standards, modernisation and innovation taking place in other countries and regions. As in the other workshops it was felt that regional coordination could benefit from using the UNECE-run HLG template. There are a number of regional and sub-regional agencies within LAC but it was felt that UNECLAC would be the best candidate for the regional coordination role. In the meantime there is space also for activities organised by sub-regional organisations like "Comunidad Andina" (CAN), SICA (Central America) and CariCom.

4. Training

There was a good level of awareness of DBMS, GSBPM and SDMX among the group. There was much interest among members of the group for learning more about big data initiatives and guidelines for use and in a number of cases (Colombia, Ecuador) the countries were about to embark on pilot projects so such knowledge would be timely. The group was unaware of current software sharing initiatives among the international statistics community and training should inform about the resources available (such as Sharing Advisory Board

[software inventory](#) and the Innovations Inventory). The group were shown the proposed regional training programme drafted following the South Asian Association for Regional Cooperation (SAARC) workshop and will be asked to indicate which elements would be of interest in their countries.

5. Collection

Data collection procedures were of a good level with all countries in the group showing fairly advanced use of hand-held devices and web-based tools.

6. Dissemination

Data dissemination was of a generally good level with no multiplicity of platforms as seen in the other regions. In general there was one platform dedicated to aggregated data and one for microdata.

7. GSBPM main “pain points”

The main processing bottlenecks identified in common in the group by GSBPM sub-process were:

- 2.5 – Design phase: processing and analysis
- 3.1 – Build phase: build data collection instruments
- 4.1 – Collect Phase: collect frame and collect samples
- 5.4 – Process phase: edit and impute data
- 6.4 – Analysis phase: apply disclosure control
- 7.5 – Dissemination phase: manage user support

8. Areas where specific technical needs were identified:

- Big Data sources, techniques and guidelines: the possibility to organise a regional Sandbox supported by UNECE Sandbox was discussed, to grant to regional NSOs the availability of tools and methods to produce Official Statistics from Big Data sources.

Section 7: Recommendations & Way Forward

Short Term ADP Type activities² for follow up

- Follow up with countries and assure detail of the quality assessments is properly transmitted
- IHSN should include a category “Other Dissemination Types” in order to correctly qualify data access in the region
- Undertaking the ADP Assessment Workshops for other countries of the region to have a base line for a full regional diagnosis on issues related to national microdata curation
- Develop an online Microdata Documentation Training for the region. This can be a self-paced course and managed by CANDANE, the Human Resources Training Center of DANE – Colombia as they already have experience in managing these learning tools. There is a need to update the Spanish language template and this could similarly be undertaken and maintained by CANDANE.
- Undertaking ADP activities in CARICOM countries. Jamaica is one of the strongest technical countries. It may be advisable to work with them as a model country.
- Selectively provide funding for documentation workshops designed to document past inventory.
- Encourage the extension of the NADA to manage other collections from data producers in the National Statistical System.
- Extension of the documentation practices to manage administrative data collections.

Mid and Long Term orientation (Additional Data and Metadata Topics)

From the survey sent to a few of the Latin-American Countries (Bolivia · Colombia · Peru · Ecuador · Costa Rica · El Salvador · Belize · Dominican Republic · Honduras · Guatemala · Nicaragua) the following prioritization of the 4 main topics considered for development in PARIS21 agenda are detailed. There is a strong interest in implementing SDMX principles (17%) followed by workshops harmonizing survey data and administrative data (14%), harmonizing data between Censuses and Civil Registration (12%) and Big Data Workshops - using Mobile data (12%). By offering technical assistance in this area the 4 prioritized topics below would respond to 55% of the demand.

² The ADP activities include data documentation, dissemination process and promoting data use as well as a continuous relation with users.

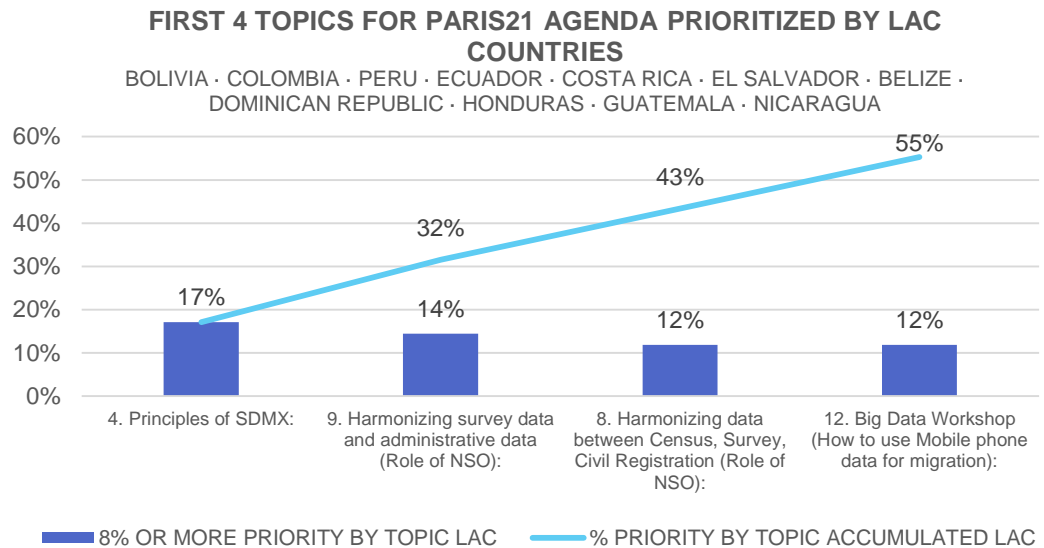


Figure 10: Primary technical topics prioritized by countries

Long Term Support

Fully sustainable and operational data management systems will always require support, if only for keeping relevant and sharing experiences among the various data producing and using entities in the region. Our point-of-view is that the best actions for assuring sustainable results will require national and regional contextualization; context that are focused on increasing data use and applications by innovative local and regional users and researchers. Harnessing this engine of innovation will ultimately provide the bottom-up force for developing responsive national statistical systems.

Regional Context

Regional support should be promoted particularly within organizations that have been given the legal mandate to undertake regional integration. This includes groups like SICA and CAN. In addition, there is a strong motive to align systems to OECD standards for inclusion into the OECD. This is a strong drive in Latin America. This is particularly the case with Costa Rica, Ecuador, Colombia and lately Perú. Regional organizations such as SICA and CAN should allocate funds to personnel and programs designed to support data development across the spectrum of data types and encourage harmonized data dissemination policies. This includes having high quality experts (data scientists) in a broad

context of data applications included but not limited to microdata documentation or data curation. These agencies and experts could undertake quality assessments and leverage economies of scale when required. Regional agencies should also develop expertise in sampling and harmonization of administrative system (in close collaboration with the OECD in order to cross pollinate lessons to the Latin America context)

Working with regional think tanks like CEPEI is important as these will leverage innovative solutions in research in relevant policy areas and we feel drive the process of developing the regional structure needed for Big Data applications (such as the Sand Box project expressed as an outcome of the IDR project). The group also recommended working with the CEPAL-CES and the working group on “Institutionalization” chaired by Colombia.

National Context

ADP activities in data documentation, dissemination and promoting data use should be included in national budgets and statistical development plans. Certain countries should be held as examples in respective fields and their lessons and best practices shared among countries on a peer to peer basis (exclusive of a regional hub). Countries like Suriname and Venezuela should be encouraged to mainline statistical process with the continent and not loose opportunities. New opportunities such as working with the new Argentinian government and national data producers could be effective over the next few years thereby adding to the pool and dynamism of the region.

Contact Information of ADP Focal Points in Latinamerica



Carmen Ugarteche Soriano
ADP Activities Coordinator for
Latinamerica & Caribbean
Tel. +591 72064833
Skype carmenugarteche
carmenugarteche@gmail.com



Federico Segui Stagno
National Data Archive
Coordinador for LAC
Skype federico.segui
federico.segui@outlook.com



Marco Antonio Gutiérrez
Consultor ADP
Skype
marcogr76@gmail.com



Franklin Hernández Cerrato
Consultor ADP
Skype hfrank2012
franklin_hernandez@hotmail.
com

Partnership Information

OECD/PARIS21

2 rue André Pascal. F-75775 Paris Cedex 16

Tel. +33.1.45.24.15.04

Fax +33.1.44.30.61.43

<http://www.paris21.org/Data-products>




PARIS21

www.PARIS21.org



THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

www.adp.ihsn.org