Accelerated Data Program

2006 - 2015

Colombia
Country Assessment

Content

Executive Summary

The ADP has been implemented in Colombia through the national statistics office or the *Departamento Administrativo Nacional de Estadística* (DANE). The process of engaging the country began in 2008 when the DANE joined the ADP initiative.
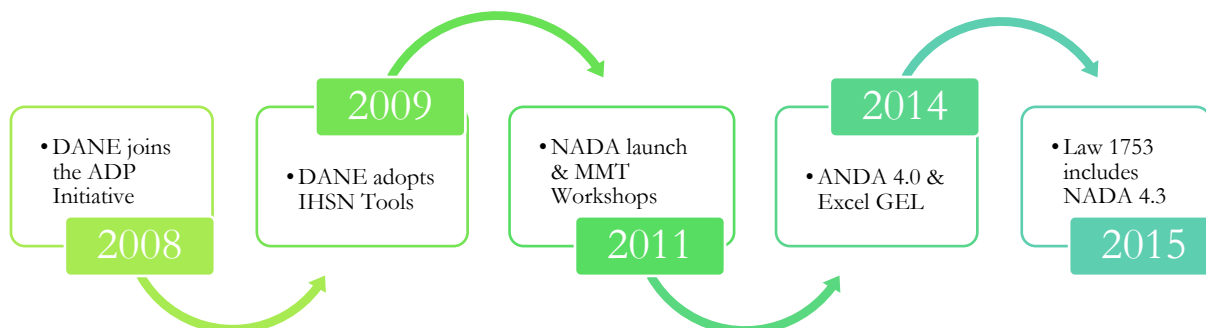

Figure 1: Timeline of the Implementation of ADP in the Country

Since its inception, the ADP funded 9 specific country interventions in direct country level support. (See Annex 2). The specific activities break down as follows:

- Documentation of statistical operations at the NSO
- Extending the process into the National Statistical System
- Setting up the National Data Archive
- Promoting the Use of microdata

The DANE has demonstrated that it is an active and innovative partner in the ADP process. It remains at the forefront of applying innovative approaches to data use and dissemination. Among the countries evaluated Colombia ranked first because of the completeness & quality of their projects' documentation, their efforts in engaging users and producers through the NADA Catalogue and retrieving feedback from them to improve on data quality and data access as well as expanding DDI & Dublin Core Standards of documentation in the National Statistics System, since they were successful in internalizing the activities the ADP promotes.

DANE has fully institutionalized the documentation processes and dedicated USD$39.000 per year of its own budget to maintain the processes. It can be considered that DANE is a mature country in the documentation and dissemination process and is able to adapt to the new data environment pro-actively. The implementation of these activities resulted in more than 177 Surveys documented with 80 projects with microdata available to download directly online (available on the National Data Archive - NADA)

Also, it's important to acknowledge the initiative DANE has taken regarding training activities with a specialized online program offered by their staff training center (CANDANE) on the IHSN Microdata Management Toolkit, which offered the benefit of opening the course to members of other NSO agencies in the region besides DANE and the Colombian National Statistics System.

Additionally is the second country in the region, besides Uruguay, who has implemented a Microdata Outreach Workshop for over a 100 participants, both users and producers, to gain feedback on availability, access and use of data as well as ways to collaborate in the future, promoting the NADA Catalogue and engaging the audience through didactic activities and debates.

Section 1: DDI Assessment

The DDI is a metadata standard adapted by the World Bank. It was originally intended to form a common standard for researchers to exchange information on research projects. For this reason the custodian of the standards remains in large part within the ICPSR. The tools developed for the assessment evaluate the main metadata elements and are designed to check completeness and coherence of the metadata such that a researcher would be able to capitalize on the available information.

The main areas evaluated in the DDI Checklist are:

1. *Description of the document:* Key descriptive elements to define the document.
2. *Description of study:* Overview, Coverage, Producers & Sponsors, Sampling, Data Collection, Data Processing, Data Appraisal, Data Access.
3. Description of the data file & Variables: Content, producer, version, literal questions text, universe, variable labels and categories, methods of derivation and imputation and confidentiality, etc.
4. *External Reference Materials:* Using the Dublin Core Standard is a summarized description of Questionnaires, coding information, technical and analytical reports, interviewer manuals, data processing and analysis software, photos, maps, etc.
5. *Resources related to the study:* Metadata, Citation and Use, IHSN Catalogue.

The table below provides a summary of the scores of the National Data Archive of DANE:

Table 1: DDI Assessment Summary of NADA Catalogue of DANE – COLOMBIA

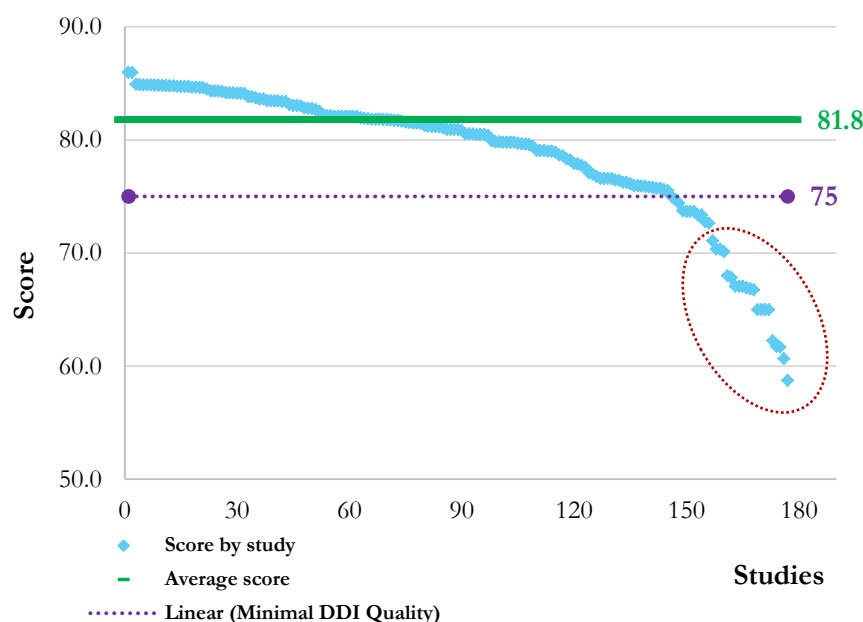| RESULTS OF COLOMBIA	SCORE
Regional Ranking	1RST
Number of statistical projects in NADA	177
Average Score	81.8%
Highest Score	86.0%
Lowest Score	58.7%
Critical Categories	DATA COLLECTION, DATA ACCESS, EXTERNAL RESOURCES 1
Difficulty to Access Data	53% of projects don't have microdata available.
Main Findings	• Lack of documentation in organizing and preparing • Inconsistencies in documenting Access Policies, Quote of the source, Copyright • Missing the incorporation of results/outputs reports

For more information on the scores in a disaggregated form please refer to the Annex 1.

DANE ranked first in the regional ranking of the assessment achieving an overall score of 81.8% of quality in their published metadata. Nonetheless, there were 3 specific areas that presented issues: Data Collections, Data Access and External Resources. There is a gap of information in the documentation regarding pilot testing, conformation and supervision of teams in the field and conceptual basis for data collection formats, as well as data access policies, provide a format for citations and copyright statements. Additionally, questionnaires weren't integrated in all projects.

Table 2: DDI Assessment Summary of NADA Catalogue of DANE – COLOMBIA

ADP Implementation in COLOMBIA by Sept. 2015	NSO – DANE	NSS – SEN	TOTAL COUNTRY
Number of Surveys Documented (Available on NADA)	145	32	177
Number of Microdata available in NADA	80	0	80

The graph below plots the individual scores ranked from highest to lowest. The pattern can give some indication as to the reasons for decreasing quality. The detail can be seen below with DANE's review scores breakdown (See circled grouping).

82% of the DDIs show they have comply with the minimal requirement to be considered high quality documentation whereas only 18% don't meet the minimal standard and the metadata quality needs to improve.

This means that with little time & effort, the Colombian agencies can improve the observations and achieve an overall high quality in all documentation published.

Figure 2: Assessed studies by score
(Under/Above of average score and the Minimal DDI Quality)

Section 2: NADA Assessment

The assessment of NADA catalog site was conducted by applying a Comprehensive Review of NADA, which consists of 41 items grouped into 7 categories, the evaluation was conducted on specific aspects about: Configuration and Visibility, Registration, Search and Filters, Data Dissemination Policy, Data Access, User and Citations, and Innovations (see Annex 2). The following table shows the average of each of the seven categories that grouped the 41 questions that assess the content and operation of the NADA catalog site. Because the maximum score for each category is different, scores are expressed as a percentage of the total possible score for each category.

Table 3. Score of NADA assessment by category

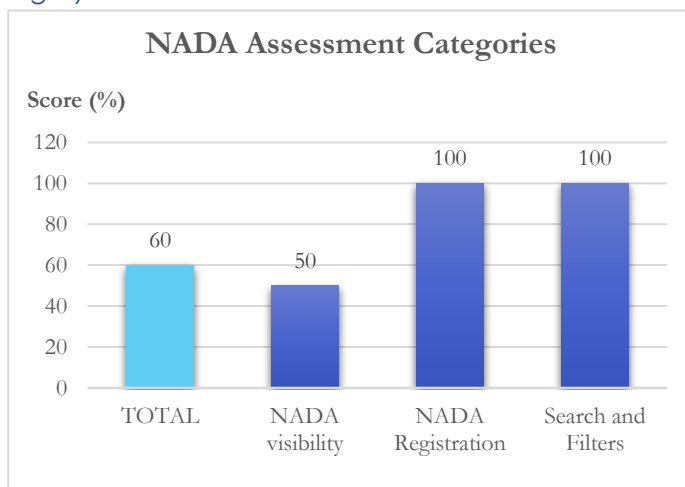| Category	Score (%)	
TOTAL	**60.0**	⬌
NADA visibility	**50.0**	⬇
NADA Registration	**100.0**	⬆
Search and Filters	**100.0**	⬆
Data dissemination policy	**0.0**	⬇
Data access	**100.0**	⬆
Metadata and Citations	**0.0**	⬇
Innovation	**0.0**	⬇

Figure 3: Score of NADA Assessment by Category

The overall score achieved by DANE's NADA Catalogue was 60%. The areas that presented more problems were Visibility of the Website for absence a specific tab for policies and procedures as well as Contact information. Also, the absence of a Data dissemination policy, Citations and Innovation was evident. Nonetheless, a Decree Law just

approved reinforces the use of NADA Catalogue as the preferred dissemination tool and formally adopts the use of DDI & Dublin Core Standards in the metadata production of the National Statistics System in Colombia and promotes anonymized microdata access online. The categories that excelled were the registration process, search & filters and data access.

Section 3: Status on Dissemination

Data Access and User Management

At **September/2015,** date in which was conducted the assessment of statistical projects contained in the **NADA catalog from DANE** of Colombia, there **were 177 published studies**, which were the subject of this assessment.

Table 4: Users Statistics in Colombia's Central NADA Catalogue

Number of registered users (active and non-active) in NADA	**7420**
Number of Active registered users	6745
Number of disactivated users (blocked)	675
Number of Licensed Requests (total)	10
Number of pending licensed requests	0
Number of citations on-line	0
Number of surveys with data unavailable in-country but available elsewhere (DHS, MICS)	0

The NADA catalog of DANE from Colombia has four types of data access: public use files, licensed data access, data enclave and no data available. The public use files can be downloaded easily into an appropriate format, once the user has completed the registration. Licensed data files requires users to be registered and to fill out a data request form, accepting the terms and conditions of the Agreement to data access. If the request is granted, the system immediately shows a confirmation message and a link is provided to follow up it. For data enclave or available via an access request at the headquarters of DANE, NADA catalog shows information describing the process and the conditions under which such consultation should be carried out and provides an email account to make the user's specific request.
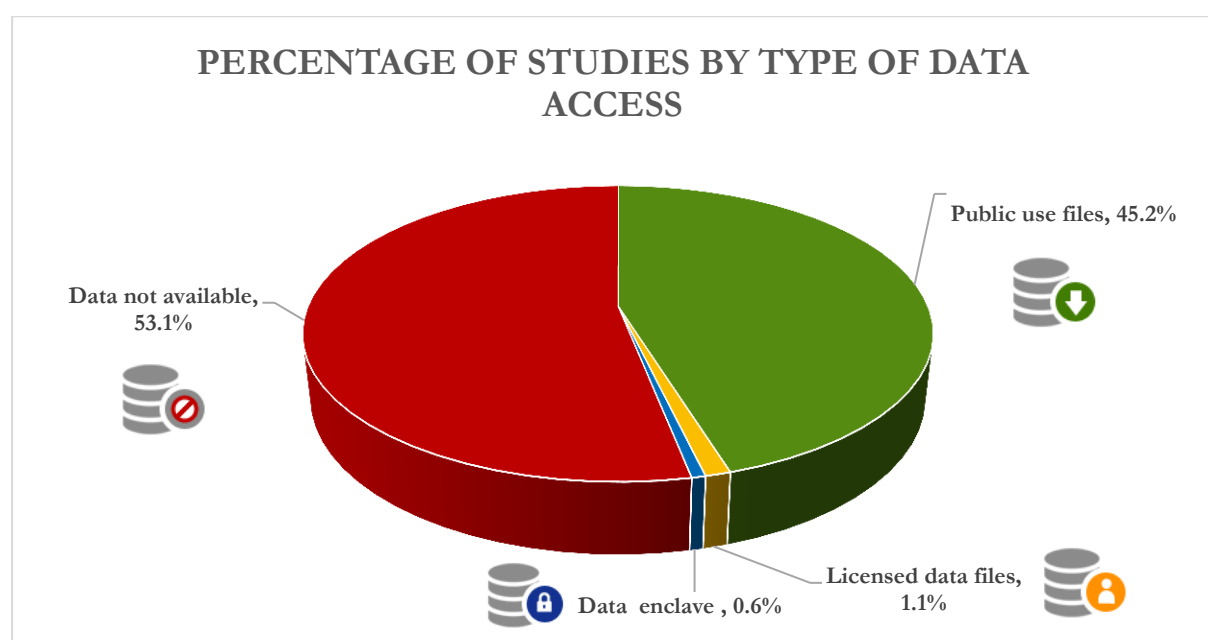

Figure 4: Studies by type of data access

An overall look to the statistical projects, for 45% the data is available and where 53% of the microdata is not available. This mainly because in order to disseminate datasets of other NSS agencies, DANE needs to sign specific agreements

with each institutions, even though a Central NADA Catalogue is in place and each agency has the capacity to manage data access of its statistical projects individually. Also, there are other ways to access the data by requesting a consultation or process a query, but for confidential datasets, especially in economics, the dataset are not available to download.

Status on the Dissemination Policy & Statistics Act

There is a Dissemination Policy stipulated within a Decree Law 3851 and its implementation started from on March, 2015 initially for DANE and in a second stage for the agencies of the National Statistics System. DANE is in charge of centralizing and providing a quality certification as well as establishing normativity on standardization & harmonization. This process has been recently approved and its implementation is being coordinated at the time of this assessment.

Table 5: STATUS OF DISSEMINATION POLICY & STATISTICS ACT

COLOMBIA	NSO	NSS	COMENTS & PROGRESS
Statistics Act allows Microdata Access	Decree-law 3851approved on March 2015		It's a mandate that allows microdata sharing at a level of National Statistics System, allowing DANE to centralize it and certify its quality.
Is the Statistics act available on line (provide link)	YES		http://www.dane.gov.co/index.php/esp/acerca-del-dane/normatividad/114-top-menu/acerca-del-dane/4002-decretos
Dissemination Policy allows Microdata Access	YES	YES	The Decree Law allows microdata sharing at a level of National Statistics System, allowing DANE to centralize it and certify its quality.
Is dissemination policy available on line? Provide link	YES		Is included in the Decree Law approved in March 2015
Regulation of Dissemination Policy includes Calendar of Publications, Type of Users, Type of Data Access per survey, NADA & DDI Standard	NO	NO	Even if there is no mandatory regulation, there are internal guidelines within DANE's management such as Administrative Resolutions that provide similar results to a dissemination policy regulation.
Other channels to Microdata Access available			
% Data Accessible through NADA Catalogue	45.1%	0%	Since the Decree Law was recently approved, the NSS is still implementing changes to make their data available.
Comments and Progress of Dissemination Policy (if not approved/updated)	80%	0%	DANE has 80% of progress in the Dissemination Policy, the Regulation still needs to be drafted. The NSS process will start after DANE's guidelines are concluded.

Section 4: Institutional Progress & Relation with Users

In DANE, the survey budgeting process at the NSO includes documentation of microdata, there is a separate process manual for the data documentation and a specific unit monitors the data dissemination called DIRPEN, who is also assigned to microdata documentation and NADA Catalogue management. Within this specialized unit there are staff members who have expertise to facilitate the data documentation workshop by themselves, and they also have expertise to maintain and update the NADA Catalogue independently. DANE is proficient in most IT applications related to data management and Statistical Packages with the exception of Data warehouses and Big Data, but they don't budget for training on any of these.

The documentation of microdata is irregular and not scheduled along the execution of the data collection, nonetheless, the documentation & dissemination processes is fully institutionalized for each statistical project, since is part of the job description of each technician involved in the process who work permanently in the agency. Hence, a program of permanent training was need it and since ADP inception, DANE has organized several data documentation workshops by its own budget and publish survey metadata in less than a month after documentation. They also created

an online course as an innovation to make the training process more accessible and suited to the staff needs, based on the Microdata Management Toolkit Workshop imparted by ADP Consultants at a local level.

Regarding data access, DANE shares datasets from other agencies within the National Statistical System on its NADA portal only with a signed agreement, but the management of licensed data and datasets access is strictly the NSS agency competence.

Estimated Budget of Documentation & Dissemination of a Survey in DANE

In the following table there is an estimation of the budget allocated for documentation and dissemination activities:

Table 6: Estimated Budget of Documentation & Dissemination of a Survey in DANE (in USD)

DANE's Curation Estimated Budget	Yes/No	Amount USD	Comments
Is there a dedicated line in the budget to documentation in the NSO budget?	Yes	USD 39000/YEAR	Overall Estimated budget for: documentation & Dissemination Activities
Is there a dedicated amount in each survey for documenting?	No		
Are consultants hired to document the surveys using internal sources?	No		
Are there dedicated personnel that document as part of the Job Descriptions? **If Yes, How many and the amount of time dedicated.**	Yes	2016 USD	Full time staff to document a survey
	Yes	9984 USD	Professional full time
	Yes	5760 USD	Technician full time
	Yes	4736 USD	Leader Part time
	Yes	4000 USD	IT Department – Part time
	Yes	6400 USD	Training – Support activities
	Yes	4000 USD	Systems – Support activities
Are there staff who review the DDIs? How many and time dedicated?	Yes		The ADP Coordination Unit who has full time dedication to DDIs validation

Regarding user management, DANE keeps records of the registered users and communicates with them. The channel to get feedback from users regarding its data dissemination is done through email, were there are specific procedures to consider the user's feedback for future survey design, or improving the data dissemination system. Also, DANE has taken the lead in the region on organizing user-producer workshops, that are budgeted each year for the promotion of data (aggregated, or microdata) dissemination were NADA Catalogue is showcased too.

Since institutionalization is one of the most important aspects to evaluate the effectiveness of the activities a SWOT analysis was undertaken. On the balance DANE has fully institutionalize the activities ADP promotes, showing initiative and proactivity in tackling the gaps and being capable of ensuring ADP activities sustainability in time. The hardest part will be expanding this standard to all NSS agencies since their statistical strength is not so developed as in DANE and the time necessary to internalize the tools within their regular activities will be a challenge.
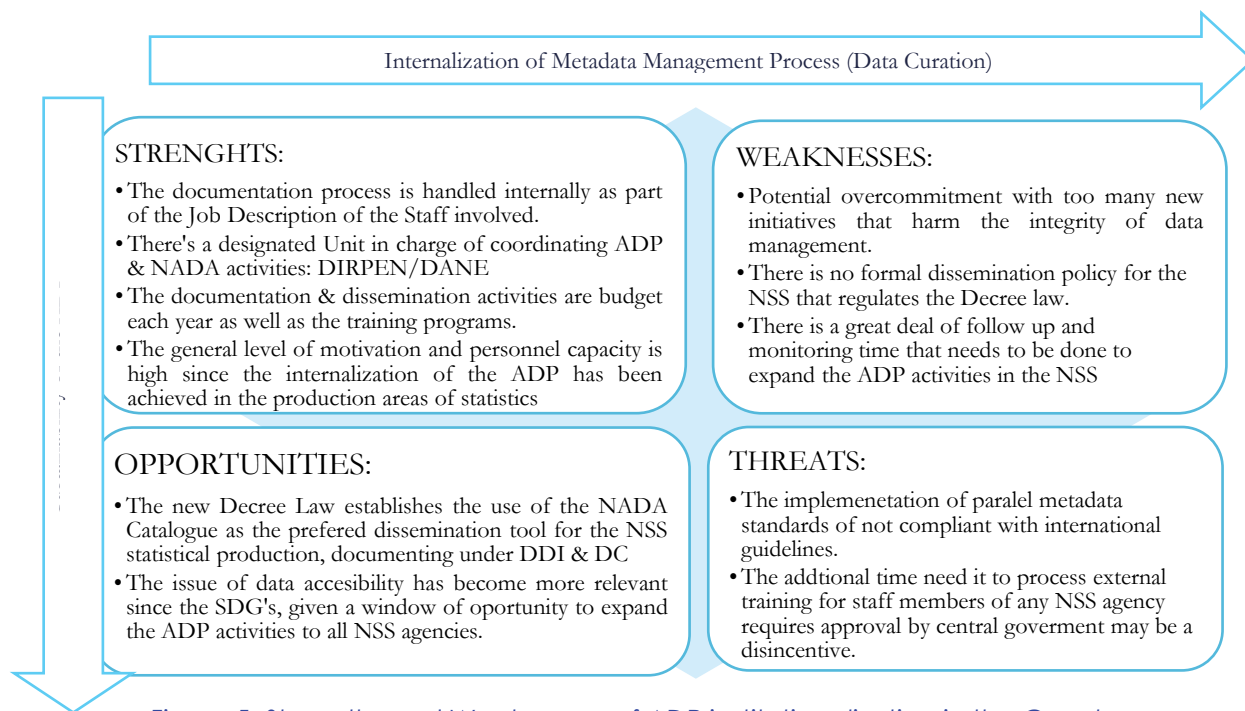
STRENGHTS:
- The documentation process is handled internally as part of the Job Description of the Staff involved.
- There's a designated Unit in charge of coordinating ADP & NADA activities: DIRPEN/DANE
- The documentation & dissemination activities are budget each year as well as the training programs.
- The general level of motivation and personnel capacity is high since the internalization of the ADP has been achieved in the production areas of statistics

WEAKNESSES:
- Potential overcommitment with too many new initiatives that harm the integrity of data management.
- There is no formal dissemination policy for the NSS that regulates the Decree law.
- There is a great deal of follow up and monitoring time that needs to be done to expand the ADP activities in the NSS

OPPORTUNITIES:
- The new Decree Law establishes the use of the NADA Catalogue as the prefered dissemination tool for the NSS statistical production, documenting under DDI & DC
- The issue of data accesibility has become more relevant since the SDG's, given a window of oportunity to expand the ADP activities to all NSS agencies.

THREATS:
- The implemenetation of paralel metadata standards of not compliant with international guidelines.
- The addtional time need it to process external training for staff members of any NSS agency requires approval by central goverment may be a disincentive.

Figure 5: Strengths and Weaknesses of ADP institutionalization in the Country

Section 5: Key Performance Indicators

Ultimately the objective of the assessment is to evaluate the performance of the country based on key performance indicators (KPIs) constructed taking into account 4 dimensions: Metadata Quality, Relationship with Users, Productivity and Institutional Management, where dimension is composed of 3 KPIs, were each category is measured over 8.33 maximus score. After running the assessment on all four dimensions, DANE ranked first in the regional ranking of the quality assessment achieving an overall score of 92.31% of quality in their published metadata in their National Data Archive.

Is worth noting that metadata has a Web presence, hence, a poor quality of documentation and poor performance of the National Data Archive website results in misuses of information, generating confusion and misinterpretations by the user. For a detailed breakdown of the performance indicators scores, please refer to the following graph to see the scores of each KPI.
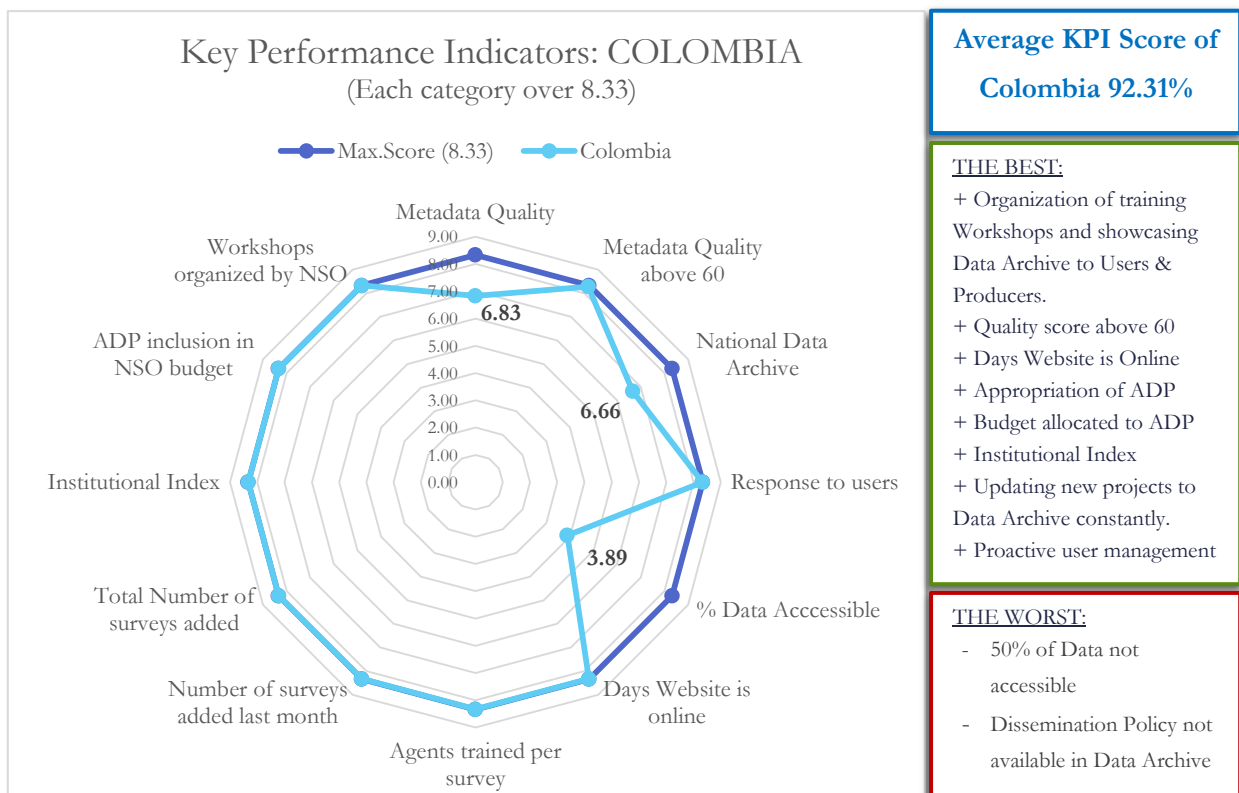
Figure 6: Key Performance Indicators: COLOMBIA

Section 6: Innovation & Data Revolution

Colombia has implemented an innovation context and based on the survey on a diagnosis under the Generic Statistical Business Process Model (GSBPM), in which the areas related to the ADP activities are the ones related to the processes that intervene in the Data Curation: Process, Analyse, and Disseminate. The comments of DANE on each process that presents problems are detailed below:

- Design/ Design variable description: Improve standardization of variables.
- Design/Design collection: Balancing the questionnaire design with the outputs. Improve collecting schemes for administrative data, there is no systematization in this process
- Design/Design frame and sample: Improve common sample frameworks.
- Design/Design processing and analysis: Improve in the statistical treatment of administrative records in integrating data from different sources and anonymization techniques that will be applied.
- Design/Design production systems and workflows: Improve microdata & metadata processing and backup.
- Build/Build collection instrument: Instrument design for collecting administrative records
- Build/Build or enhance process components: Improve on integration of different sources of data & backup.
- Collect/Create frame and select sample: Technical assistance on coordinating samples between cases of the same statistical management process.
- Process/Integrate data: Improve integration of data specially regarding to combining multiple data.
- Process/Classify and code: Identify opportunities of automatized coding
- Process/Edit and impute: Technical assistance for imputation mechanisms, normalization & generalization of those methods to the institution good practices.
- Analyze Apply disclosure control: Improve anonymization techniques specially for economic statistics
- Disseminate/Update output systems: Technical assistance for establishing the metadata production in previous stages to address only quality validation in this particular process.

- Disseminate/Promote dissemination products: Assistance for improving strategies for disseminating statistics to broaden the scope of users. (data visualization among others)
- Disseminate/Manage user support: Technical assistance on systematizing client consults through the quality management system, to include new needs or the modification of the existing.

DANE has express they are able to provide technical assistance to other countries and/or agencies can in the advances that have for metadata and microdata dissemination, implementation and accompaniment to technical staff and other users who have joined the National Data Archive NADA.
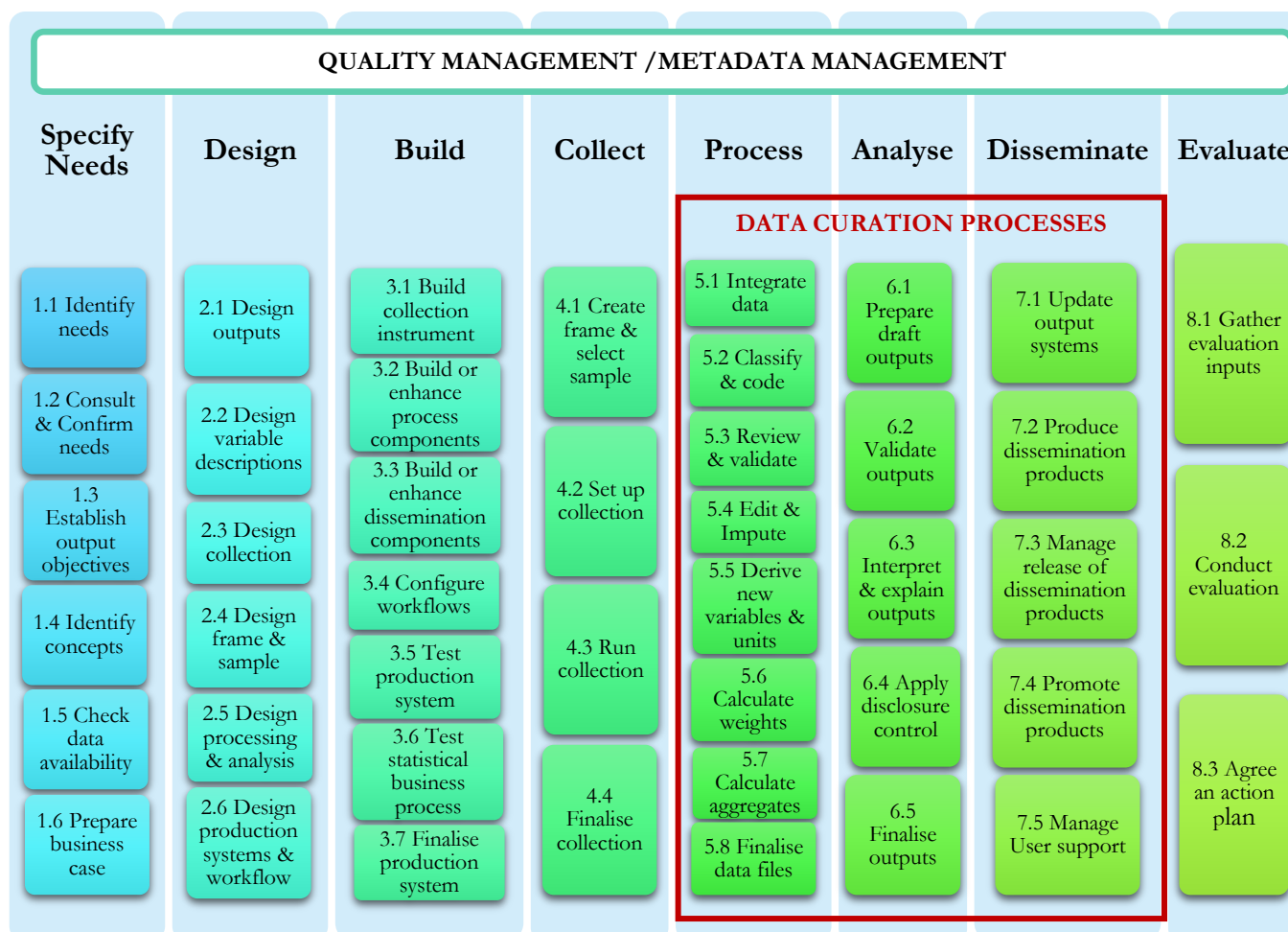
QUALITY MANAGEMENT /METADATA MANAGEMENT							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
				DATA CURATION PROCESSES			
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & Confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components		5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.2 Set up collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.2 Conduct evaluation
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows		5.4 Edit & Impute			
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system	4.3 Run collection	5.5 Derive new variables & units	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process	4.4 Finalise collection	5.6 Calculate weights			8.3 Agree an action plan
		3.7 Finalise production system		5.7 Calculate aggregates	6.5 Finalise outputs	7.5 Manage User support	
				5.8 Finalise data files			

Figure 7: The Generic Statistical Business Process Model (GSBPM)

Section 7: Conclusions & Recommendations

Objective 1. Quality Assessment of metadata published in DANE's NADA Catalogue

- During the metadata quality assessment, only a few gaps in documentation were identified in the following sections: sampling, data collection and data processing, mainly due to instructions on the field description not entirely accurate, or because the documentation of these elements in the metadata template isn't mandatory.
- The documentation of the data access policy is not consistent with the criteria of the assessment template in the following elements: official policies and procedures to disseminate data, structure to citation of databases and the copyright statement, resulting in low scores in the assessment of this category.
- Documenting different time periods of the same project in diferentiated studies is not exactly an error but causes NADA Catalog shows duplicated documentation even when the conceptual design and structure of datasets has not changed over time.
- The default settings of DANE's NADA catalog don't show elements that were documented in the metadata editor, such as: Questionnaires description, geographic unit and ending date in time period.
- A recommendation is not to use the same documentation for labels and literal questions.
- DANE is working on developing a documentation template for administrative registers.

Objective 2. Status on the Statistics Act and Dissemination Policy of Microdata

- While there are major advances in the field, is necessary to develop the regulation of data Disclosure Policy (microdata) approved in large as part of the Decree Law in March 2015 for the entire national statistical system, initially DANE in the short term and then with the agencies of the National Statistical System.
- The dissemination policy (Decree Law) and its Regulations should be available on the website of the National Data Archive. (Decree 3851 of 2006, Article 1)

Objective 3. Assessment of the funtionality of NADA Catalogue and quality of the service to microdata users.

- During the DANE's assessment of the National Data Archive - NADA it was evident the Website layout is different in relation to other NADAs from other regions, since it lacked content on some sections that should be present such as: mission and purposes, citations, policies and procedures, acknowledgements, news, innovations and contacts, causing low scores on the NADA assessment.
- The functioning of the NADA Catalog will render greater benefits when it is formalized as the preferred method of dissemination in the institutional policy
- Opportunities to measure the quality of service of microdata users.
- The inclusion of a new category of "other access to data" in the classification of Data Access is necesary in the next version of the Catalogue ANDA.

Objective 4. ADP sustainability and User engagement

- Sustainability of the activities of ADP is assured as it depends on three fundamental aspects that DANE has successfully implemented: a) The inclusion of the tasks of documentation and dissemination within the job description of the staff and the budget of the institution. b) A plan of internal training in the use of tools Microdata IHSN managed by the DANE's training area (CANDANE), as the initiative taken to consolidate the online course, based on the presentations and tools provided by the ADP. c) Awareness workshops and permanet feedback & relationship with users.

Section 8: Looking Forward

In an effort to consider NSO needs in setting priorities for the PARIS21 Agenda post 2015, each country was asked to fill out a form stating the priority of 15 products that PARIS21 is considering to develop.

The results for DANE are presented in the following graph, where SDMX and Innovation & Data Visualization take the first place, followed by Dissemination Policy Design and User Management, Harmonization of Data of different sources, Big Data, NSDS & SDGs and Data Curation Manuals. By offering assistance in the first 3 topics 64% of Colombia's demand would be covered.
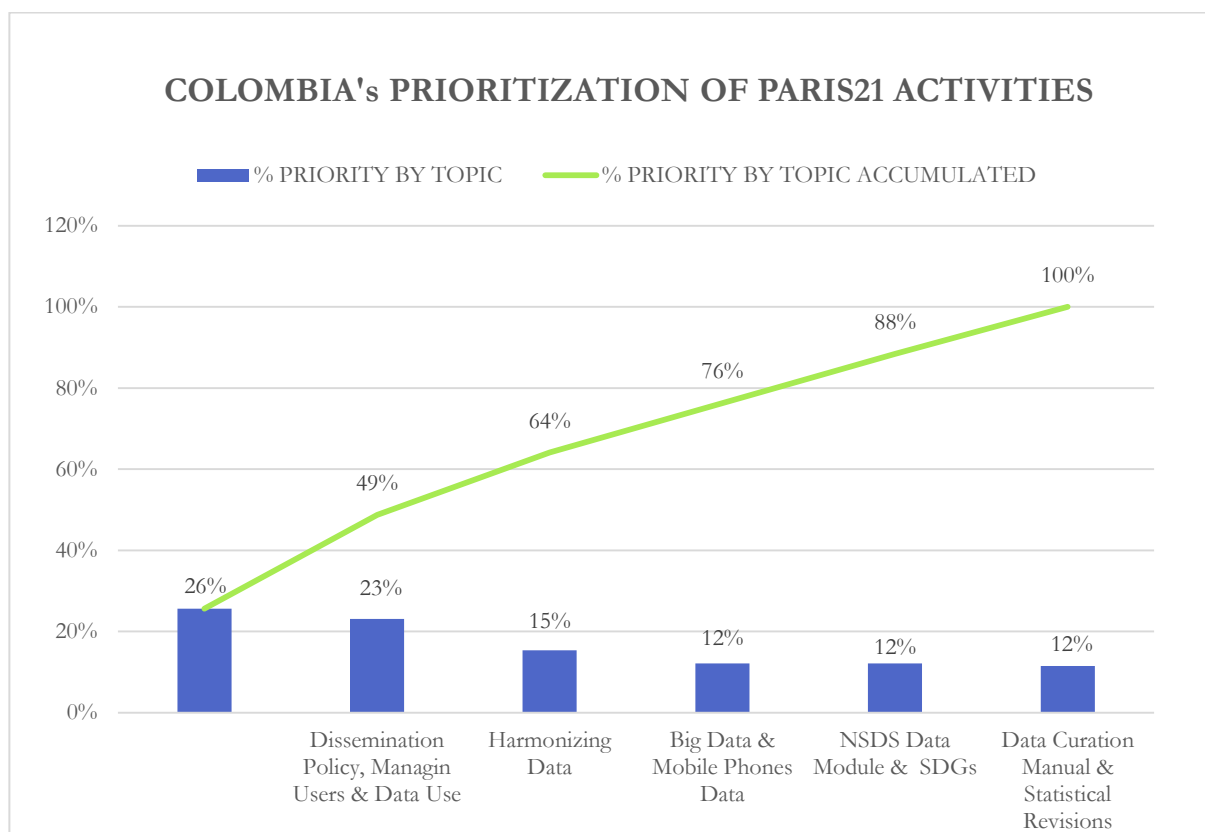

Figure 8: COLOMBIA's PRIORITIZATION OF PARIS21 ACTIVITIES

Accounting for the country expression of needs and in the survey drafted by the Informing a Data Revolution Initiative, the main areas where specific technical needs were identified are Big Data sources, techniques and guidelines through a regional Sandbox supported by UNECE Sandbox, to grant DANE the availability of tools and methods to produce Official Statistics from Big Data sources.

ANNEXES

Annex 1: Score by metadata category for DANE's NADA Catalogue DDIs

Category	Score (%)		Main comment for improvement
IDENTIFICATION	94.1	⬆	The reference year should not appear in the subtitle.
VERSION	83.3	⬆	It is recommended the status of the version field of data files, is changed to mandatory.
OVERVIEW	99.1	⬆	Avoid documentation that does not contribute to the understanding or use of data.
COVERAGE	98.1	⬆	It is necessary to modify the filling instructions to clearly establish differences between coverage and geographical unit.
PRODUCERS AND SPONSORS	97.4	⬆	-
SAMPLING	90.6	⬆	It is necessary to modify the filling instructions to clearly establish criteria for targeted sampling and sampling frame for administrative records data.
DATA COLLECTION	55.0	⬇	It is necessary to modify the filling instructions to document information about pilot testing, conformation of enumeration team and conceptual basis for data collection formats.
DATA PROCESSING	93.4	⬆	It is necessary to modify the filling instructions to document information about kind of data entry and software used to edit data.
DATA APPRAISAL	76.8	➡	-
DATA ACCESS	28.4	⬇	It is necessary to modify the filling instructions to document better information about data access policies, format to citations and copyright statements
DATA FILES	85.7	⬆	Do not incorporate documentation about structure of the data files in the description of data base content..
VARIABLE	99.6	⬆	The labels should be clear and precise, avoiding duplicate documentation of literal questions and labels.
EXTERNAL RESOURCES 1	46.3	⬇	Questionnaires must be integrated in all projects
EXTERNAL RESOURCES 2	94.5	⬆	Verify that the links of external repositories for downloading files to work properly. Favoring downloading on PDF format.
METADATA, CITATION AND USE	75.1	➡	Ensure that the version of NADA allows download documentation in PDF and XML. Incorporate Citations tool on NADA
IHSN CATALOGUE	60.4	⬇	Make a reconciliation exercise between NADA and IHSN catalog so that both sites contain the same projects

Average score of Colombia NADA Catalogue: 81.8%

Annex 2: Summary of ADP implementation in COLOMBIA

ADP implementation in COLOMBIA up to September/2015

ADP implementation in COLOMBIA up to Sept/2015	NSO – DANE	NSS – SEN	TOTAL COUNTRY
Year ADP Started	2008	2012	2008 - 2015
Number ADP interventions in the Country	**4**	**5**	**9**
Metadata Production Workshops	2	3	5
National Data Archive (NADA)	2	1	3
User Outreach Workshops		1	1
Number of Institutions Trained in Data Documentation	1	12	13
Number of Institutions attending User Outreach Workshops			
Number of persons attending ADP training events	33		
Number of women trained in documentation	17		
Number of Surveys Documented (Available on NADA)	145	32	177
Number of Microdata available in NADA	80	0	80

DANE

www.dane.gov.co

PARIS21

www.PARIS21.org

THE WORLD BANK
IBRD · IDA | WORLD BANK GROUP

www.adp.ihsn.org