Integrating new data sources and data innovations for official statistics and partnerships with stakeholders

Workshop on Preparing Evidence-based Voluntary National Reviews for 2021 HLPF
(Session 3)

January 2021

SDGs implementation
Legal framework in Colombia

High Level SDG´s Commision
President: DNP

1. National Planning Department
2. Ministry of Finance
3. Ministry of Foreign Affairs
4. Social Prosperity Department
5. Ministry of Environment and Sustainable Development
6. Presidency Department
7. National Statistics Department

Technical Secretariat
DNP-DSEPP

Technical committee

- ✓ Members: ministries
- ✓ DANE
- ✓ Technical Secretariat: DNP-DSEPP
- ✓ Invited guests

Partnership

- ✓ Civil society organizations
- ✓ Private sector
- ✓ Academy
- ✓ **International Organizations**
- ✓ **Media**

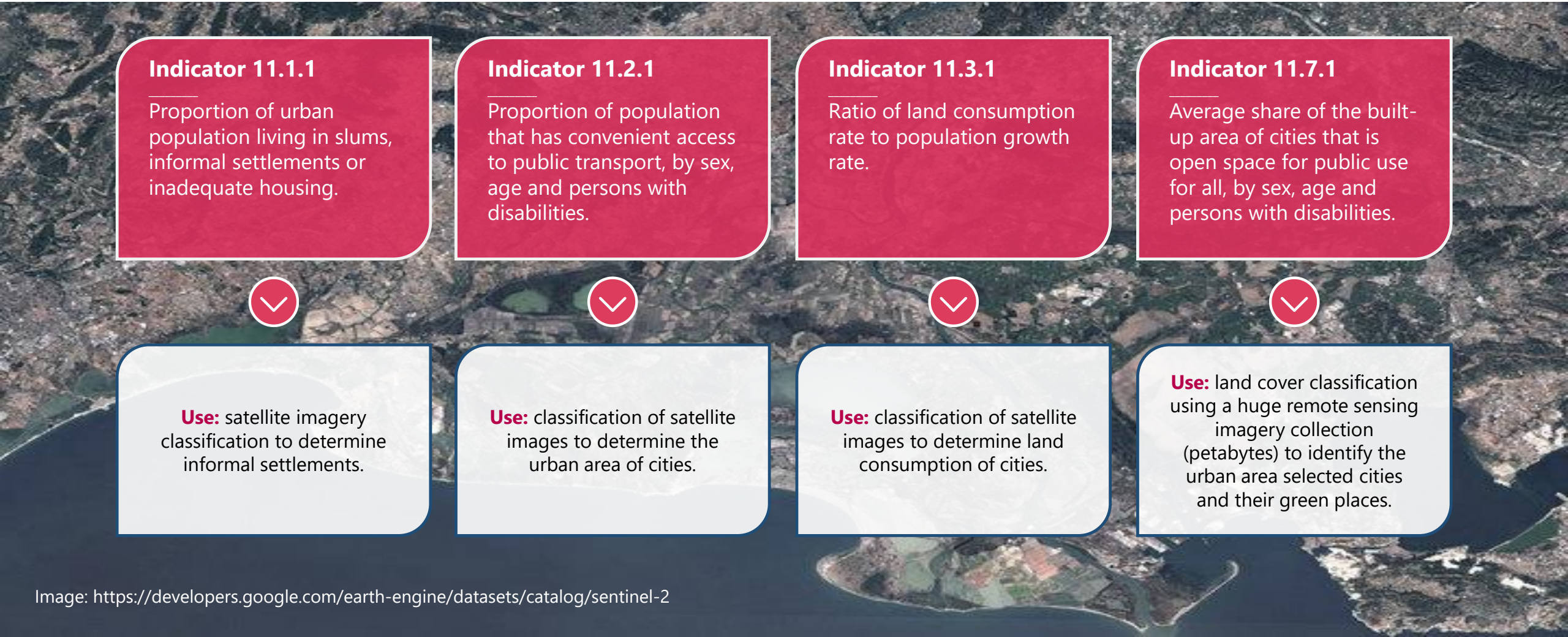
POLICY DOCUMENTS

- *Long Term*: CONPES 3918
  - National SDG Indicators Framework
  - A set of global indicators prioritized
- *Mid/short Term:*
  - National Development Plan 2015-2018
  - National Development Plan 2019-2022

INFORMACIÓN PARA TODOS

Use of satellite images for computing the SDG indicators from Goal 11

Using satellite images, it is possible to obtain historical and updated data on land cover to analyze the expansion of urban agglomerations.



Indicator 11.1.1

Proportion of urban population living in slums, informal settlements or inadequate housing.

Use: satellite imagery classification to determine informal settlements.

Indicator 11.2.1

Proportion of population that has convenient access to public transport, by sex, age and persons with disabilities.

Use: classification of satellite images to determine the urban area of cities.

Indicator 11.3.1

Ratio of land consumption rate to population growth rate.

Use: classification of satellite images to determine land consumption of cities.

Indicator 11.7.1

Average share of the built-up area of cities that is open space for public use for all, by sex, age and persons with disabilities.

Use: land cover classification using a huge remote sensing imagery collection (petabytes) to identify the urban area selected cities and their green places.

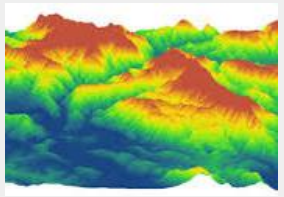
Image: https://developers.google.com/earth-engine/datasets/catalog/sentinel-2

SDG 9.1.1

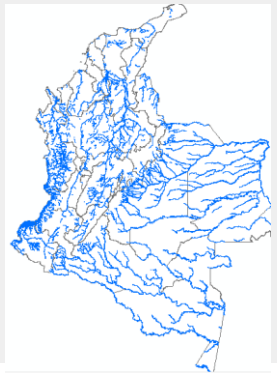
Sources of information

External Sources

Road Coverage
Source: Agustín Codazzi Geographical Institute – IGAC (2018)
Resolution/Scale: 1:100.000
Type of roads: Clasificación IGAC (1,2,3 y 4)



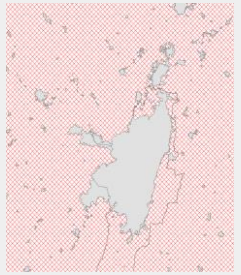
Digital Elevation Mode
Source : Agustín Codazzi Geographical Institute – IGAC (2018)
Resolución/Escala: 12,5 metros



Hydrographic Coverage
Source : Agustín Codazzi Geographical Institute – IGAC (2018)
Resolution/Scale : 1:100.000
Tipo de Hidrografía: Drenajes dobles, Drenajes sencillos, ciénagas y humedales)

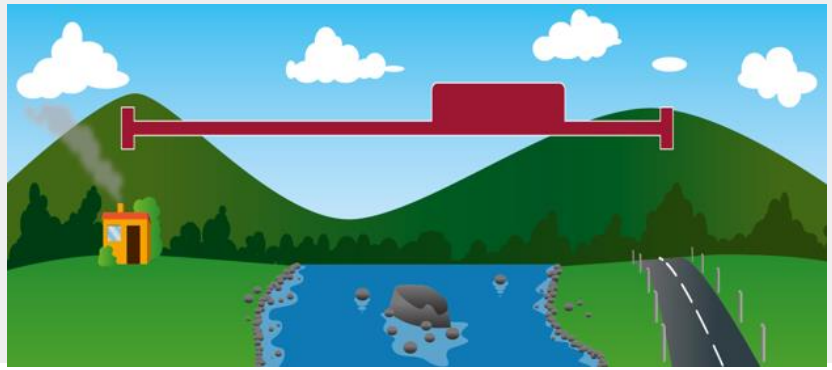
Internal Sources

National Population and Housing Census
Source : DANE (2018)
Resolution/Scale : National
Type: Housing and rural population



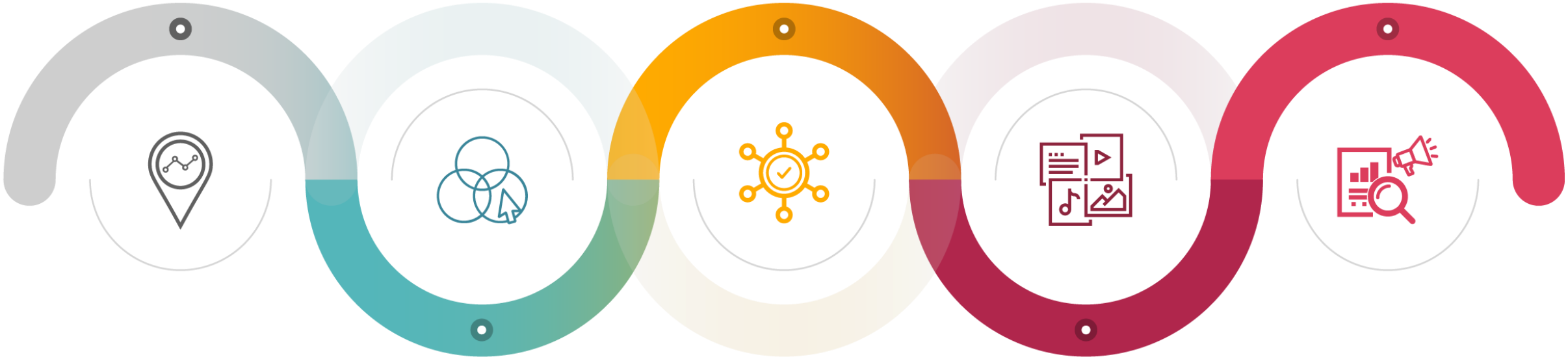
Limits Urban/Rural areas
Source: DANE National Geostatistical Framework (2018)
Resolution/Scale : Nacional
Type: Delimitation of urban/rural areas

Satellite Images



INFORMACIÓN PARA TODOS

Integration of statistical and geospatial information



Inputs

Geoespatial

- Fundamental data.
- Supplementary data.
- New data sources.

Statistical

- Censuses. Surveys.
- Administrative registers.
- Big data and other sources.

Principles

- Accessible & usable.
- Statistical and geospatial interoperability.
- Common geographies for dissemination of statistics.
- Geocoded unit record data in a data management environment.
- Used of fundamental geospatial infrastructure and geocoding.

Key elements

- Standards and good practices.
- National laws and policy.
- Technical infrastructure.
- Institutional collaboration.

Outputs

- Integration.
- Harmonized and standardise information.
- Interoperability and comparability.

These serve as inputs for:

- Analysis.
- Diffusion.
- Decision making.

Source: Adapted from "THE GLOBAL STATISTICAL GEOSPATIAL FRAMEWORK. WORKING PAPER - FOR EG-ISGI CONSULTATION. UNITED NATIONS EXPERT GROUP ON THE INTEGRATION OF STATISTICAL AND GEOSPATIAL INFORMATION"

ArcGIS Hub – Esri *Geovisor (work in progress)*

Home page of DANE's ArcGIS Hub portal for the spatial dissemination of SDG indicators

Dissemination of geo-referenced statistical information:

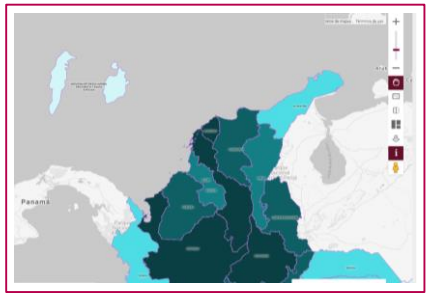
1. Vulnerability Geovisor

- Comorbidities.
- Adult population.
- Overcrowding.
- Population Density.
- Intergenerational risk.

2. Multidimensional Poverty Index Geovisor

- Educational conditions.
- Conditions of childhood and youth.
- Labor conditions.
- Health.
- Housing conditions and public services.

3. Population Census Geovisor



- Population.
- Dwellings.
- Socio-demographic indicators.
- Housing conditions and public services.

Geovisor

GLOBAL

IAEG-SDGs

- WG of Geospatial Information
- WG of SDMX
- WG on the new Measurement of Development Support

REGIONAL

Statistical coordination group for 2030 Agenda in Latin America and the Caribbean. Sustainable Development Goals – SDGs.

- Discussions about the Regional Framework
- Report on the activities of the IAEG-SDGs to all country members
- Fostering the availability of a synthetic measure of the 2030 Agenda

NATIONAL

Working Group with Agengies of the UN System in Colombia.

- Work on specific SDG indicators to be produced for Colombia.
- Design of a follow up scheme to track advances in statistical production (BAROMETER)

DANE has membership in these instances of the UN System

DATA FOR THE GOALS

DATA FOR NOW

Possibilities to strength the NSS

Go beyond methodologies.

Documentation process and develop of virtual material to allow other countries to reply the exercises

Expand possibilities on information sources and alternative methods

Barometer

Designed between UNFPA Colombia and DANE

D

1. Contact was made with the Custodian Agency
2. Possible sources have been identified
3. Origin of the sources

C

4. Sources of information are available
5. There was inter-institutional coordination
6. There is a Work Plan for data extraction
7. There is an agreement on the conceptual and methodological definition

B

8. The data for calculating the indicator is complete
9. The required data have the necessary statistical quality
10. The Custody Agency validated the data
11. DANE validated the data

A

12. There is clarity about statistical operation
13. Data collection has an established periodicity
14. It is recorded from and until when the information is available
15. The periodicity required for the dissemination of the indicator is guaranteed
16. The indicator is reported globally
17. The indicator is reported globally and is validated by the country

A: 76% - 100%	The indicator is produced
B: 51% - 75%	The indicator is not produced but can be produced with existing information sources
C: 26% - 50%	Some information is available but needs to be improved or supplemented to produce the indicator
D: 0% - 25%	No information is available to produce the indicator

*** In February, a first exercise was applied to establish the baseline. The second version was applied in June and one last time in November.*

IT Architecture

Objective: conceptualize and deepen the governance structure that allows strengthening the use of administrative records to produce statistical information

Progress: it started on December 1st 2020, with introduction of DANE team members and UNSD/D4N IT consultant and agreement to hold **regular meetings twice a week**.

Implementation phases*:

Phase 0 - Discovery and Inquiry (December 2020)
Phase 1 - Proposal formulation (January and February 2021)
Phase 2 - Implementation and prototypes (January and February 2021)
Phase 3 - Analysis of results (March 2021)
Phase 4 – Recommendations (March 2021)

INFORMACIÓN PARA TODOS

DATA FOR THE GOALS

DATA FOR NOW

Poverty mapping

Integration of alternative sources of information in the statistical process

First approach we are currently working on:

1 NO POVERTY

Currently DANE measures:

- MPI statistics at the department-level using household surveys (annually).
- MPI statistics at the municipality-level using census data (every 10 years).

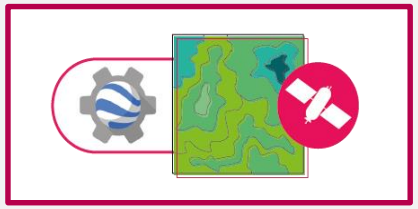
Goal:

- Measure MPI statistics at the municipality-level every year.

Sources:

- Household surveys.
- Spatially detailed Census data.
- Geospatial covariate datasets.

Methodology:

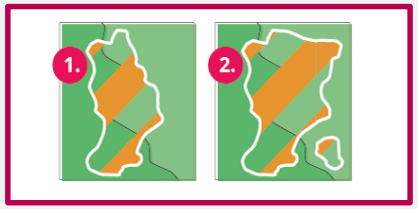


Compile

- Geospatial covariate datasets (eg. nighttime light consumption, vegetation index, accessibility via road to towns and cities).

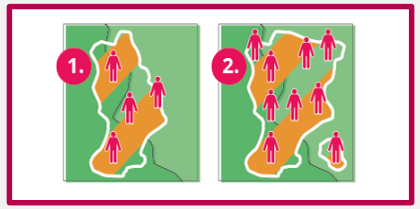
Input

- Survey clusters displaying the cluster-level MPI headcount ratio.



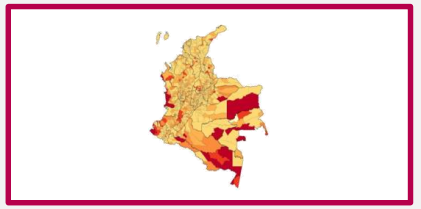
Modelling

- Generalized linear mixed model (model-based geostatistics).
- Bayesian geostatistical model.



Estimate

- The population living in poverty at the cluster level.



Results and validation

- Mapping MPI headcount ratio at the micro- scale (cluster-level) and macro – scale (municipality-level).
- Asses models' predictive performance.

Night-time lights as a proxy of economic activity during COVID-19 outbreak

Integration and comparison between night-time data and socioeconomic measures from firms

We are currently working on:

Currently DANE firm measures (per month):

- Production.
- Total production personnel.
- Power consumption.

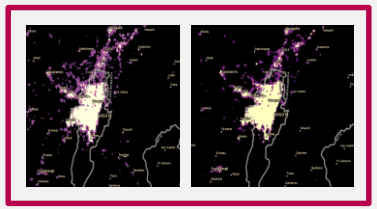
Goal:

- Measure changes in nighttime TOA radiance on a date before the COVID-19 outbreak and on a date during the lockdown. Correlate these changes with shifts in the economic variables by means of an econometric model.

Sources:

- Monthly Manufacturing Survey.
- Spatially detailed Census and firm data.
- Night-time light datasets (VNP46A1 - Radiance).

Methodology:



Compile night-time light data

- Night-time light datasets from NASA's VNP46A1 sensor in dates before (07/02/2020) and during (27/04/2020) the outbreak lockdown.

Process and compare images

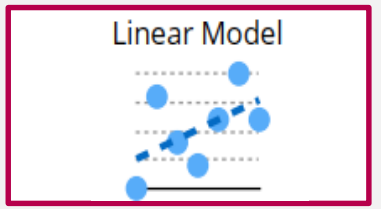
- Process the images (select suitable dates, apply cloud and vegetation masks, obtain the radiance difference between the two images, focusing on pixels with positive and negative values.

Georeference firms

- Locate the surveyed firms and for each one, calculate power consumption ratio.

Correlate

- Compare firm variables month by month and correlate them with changes in nighttime lights by building an econometric model.



Results and validation

- Assess model results.

INFORMACIÓN PARA TODOS

Early estimation of manufacturing production

Integration of administrative registers and survey´s

First approach we are currently working on:

Currently DANE measures:

- Estimate anual industry production growth with a lag of 45 days.

Goal:

- Early anual estimation of manufacturing production level with a lag of 15 days.

Sources:

- Manufacturing industry survey.
- Google trends.
- Energy consumption administrative register.

Methodology:

Compile

- Covariate datasets (i.e, energy consumption index, Google trends).

Input

- Manufacturing enterprises survey.
- Energy consumption administrative register.

Modelling

$$\Delta\%Y_t = \beta\Delta\%X_t$$

$$\varepsilon_t = \alpha\varepsilon_{t+1} + \epsilon$$

- Forecast evaluation.
- Integrated moving average model.

Estimate

- Annual production index growth.

Results and validation

- Asses models' predictive performance.

Microdata classification algorithm

Integration of social security register and household survey

First approach we are currently working on:

Currently DANE measures:

- In march and april, DANE measured the unemployment rate, but due to the changes in data collection, the infomality rate is not available for these periods due to a lack of information.

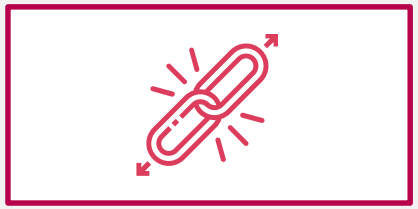
Goal:

- Estimate a dummy informality variable to impute the microdata of the household survey to recover the informality series.

Sources:

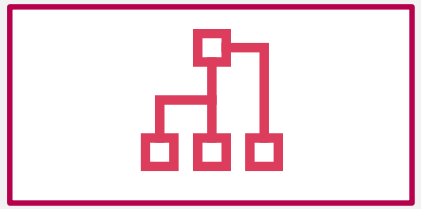
- Social security register – PILA.
- Household survey – GEIH.

Methodology:



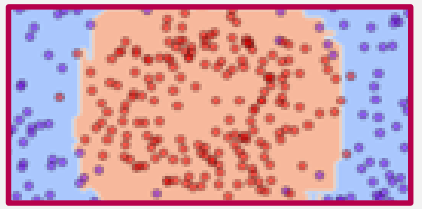
Compile

- Link PILA-GEIH.
- PILA status and GEIH covariates.



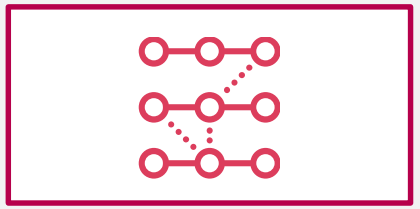
Modelling

- Machine learning.
- Classification model – Random Forest.



Estimate

- Random Forest model.
- Impute GEIH microdata



Results and validation

- Symmetric confusión matrix.
- Precision, Recall y F1 score greater than 0,8.
- The dummy indicator of the match is in the top 10 of predictors.

Experimental Statistics

Current Approach in DANE

Definition

Are the ones derived from projects in development that have at least one of the following innovative aspects:

- new sources of information
- the statistical methodology used
- a new topic not previously measured.

They are considered experimental because they still show room for improvement (harmonization, coverage and methodology) and they have not yet reached sufficient maturity in terms of the reliability, stability or quality assessment of the data. In Colombia the experimental statistics are official statistics by decree 2404 of 2019.

Quality Attributes

1. **Relevance**
2. **Opportunity**
3. **Accesibility**
4. **Interpretability**
5. **Coherence**
6. **Transparency**
7. Accuracy
8. Comparability
9. Continuity
10. Credibility

Attributes required for experimental statistics

DANE
INFORMACIÓN PARA TODOS

Thank you!

klchavezq@dane.gov.co
direccion@dane.gov.co

www.dane.gov.co

El futuro es de todos
Gobierno de Colombia