

**DANE** 70 AÑOS  
INFORMACIÓN PARA TODOS

# Desde DANE como apoyamos al ciclo de política pública con datos

Subdirección



GOBIERNO DE COLOMBIA

21 abril de 2023

DANE (desde October 1953) es el Instituto Nacional de Estadística de Colombia responsable de la planeación, compilación, análisis y difusión de las estadísticas oficiales del país. DANE es responsable de la realización del Censo Nacional de Población y Vivienda cada 10 años (últimos 2005 y 2018), entre otras 129 operaciones estadísticas entre censos, encuestas y registros administrativos.



## Ecosistema de datos de las estadísticas oficiales - SEN

● Sistema Estadístico Nacional – SEN

● Subsistema Departamental

● Subsistema Municipal

○ Usuarios de información estadística

\* Subsistema de otras entidades territoriales

- Provincias
- Áreas metropolitanas
- Territorios indígenas
- Territorio colectivo de comunidades negras
- Asociaciones de municipios
- Territorios PDET



<sup>1</sup> **Artículo 155 de la Ley 1955 de 2019.** Desde la oferta, además de las entidades del orden nacional y subnacional que producen información estadística; están las personas jurídicas, públicas o privadas, que prestan servicios públicos y las personas jurídicas que posean, produzcan o administren registros administrativos en el desarrollo de su objeto social, que sean insumos necesarios para la producción de estadísticas oficiales. Dentro de las organizaciones de la sociedad civil (OSC) se incluyen: las organizaciones no gubernamentales (ONG), entidades sin ánimo de lucro (ESAL) las cuales a su vez, incluyen las cámaras de comercio; gremios; organizaciones sociales, comunitarias, sindicales, de profesionales, étnicas, académicas; entre otras.

**129 Operaciones Estadísticas**  
(diferentes métodos de recolección, regularidad y cobertura)

Agricultural  
5

Economics  
11

Commerce  
and Industry  
19

Indexes  
74

National  
Accounts  
46

Social  
21

Technology  
and  
Innovation  
4

Government  
and  
Institutions  
6

Specify Needs

Design

Build

Collect

Process

Analyse

Disseminate

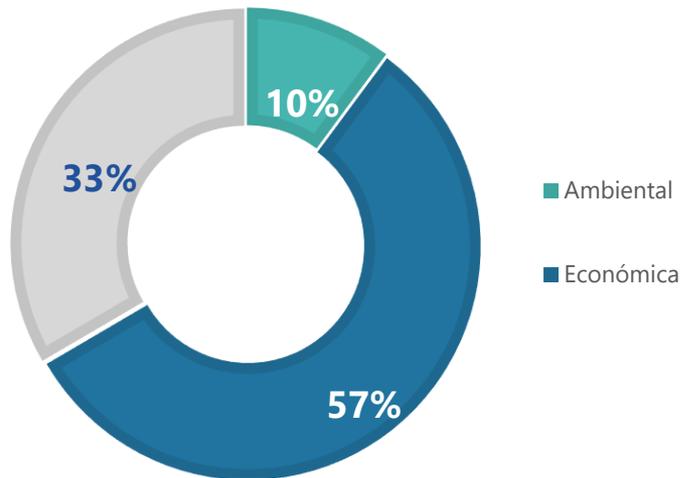
Evaluate

**Generic Statistical Business Process Model (GSBPM)**

## DANE principal productor de estadísticas oficiales

### Operaciones Estadísticas\*

Participación de las OO.EE por temática



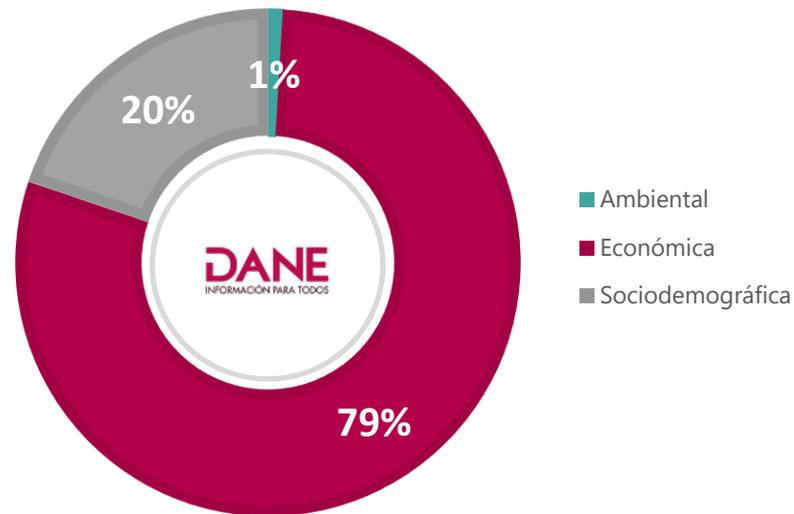
El país tiene **401 operaciones estadísticas** producidas por **103 entidades**.

\* OO.EE activas

Los datos e información estadística del SEN son activos disponibles para la investigación aplicada y la toma de decisiones.

### Operaciones Estadísticas DANE \*

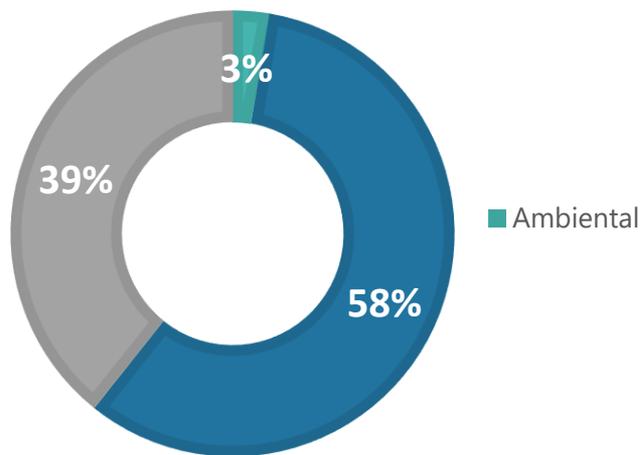
Participación de los RR.AA. por temática



Del total de **operaciones estadísticas**, el DANE produce actualmente **106**

## Fuentes secundarias disponibles y su uso estadístico

Participación de los RR.AA. por temática



El país tiene **527 Registros administrativos** producidos por **95 entidades**.

Los datos e información estadística del SEN son activos disponibles para la investigación aplicada y la toma de decisiones.

Participación de los principales usos de RR.AA. por el DANE



El DANE aprovecha estadísticamente **233 Registros administrativos**.



## Interés estadístico del nuevo Plan Nacional de Desarrollo - PND 2022-2026

Las principales necesidades de información y análisis de datos son:

Censo Económico – 2024.

Conteo Intercensal de Población – 2025.

Mayor granularidad de las estimaciones. Estadísticas oficiales para grupos poblacionales más específicos como víctimas, género, etnicidad, pero también estimaciones a niveles más desagregados de nivel geográfica como municipios, localidades, barrios para todas las OOE no sólo del DANE sino del SEN para el seguimiento del cumplimiento de los indicadores.

De acuerdo con las bases del Plan Nacional de Desarrollo 2022-2026 “**Colombia potencia mundial de la vida**”, se cuenta con los siguientes actores diferenciales para el cambio:



**Genero y OS**  
Mujer  
LGBTIQ+



**Ciclo de vida**  
Niños y niñas  
Jóvenes  
Adulto Mayor



**Discapacidad**



**Víctimas  
Desplazados**



**Grupos  
étnicos**



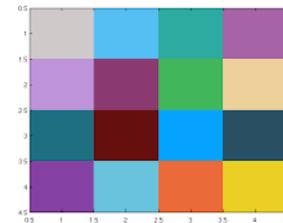
**Territorio  
Municipal y Sub  
Ruralidad  
PDET**



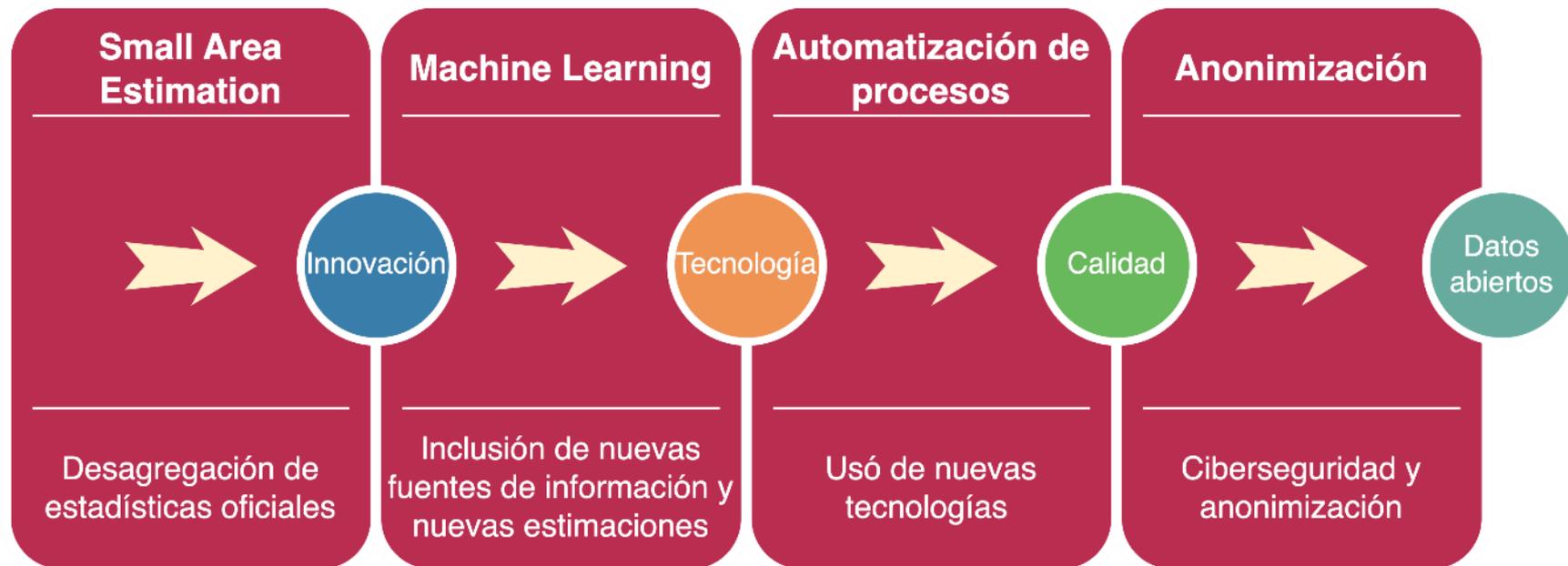
**Campesinos**



**Economía  
Popular**



**Interseccionalidad**



## Toma de decisiones basadas en información

Dada la alta disponibilidad de datos se genera de una mayor demanda de análisis con el fin de **tomar decisiones basados en datos**. En la política pública se está dando el mismo enfoque lo que aumenta la demanda de información por parte de otras entidades, la academia, organismos gubernamentales y la ciudadanía en **publicar información con mayor desagregación y detalle**.

## Uso de los datos en el ciclo de política

Medición de variables de interés (Pobreza, Mercado laboral, crecimiento económico, entre otros) mensual, trimestral y anualmente.

Identificación del problema



Análisis de los usuarios de los Microdatos en ANDA:

1. Identificación de población objetivo
2. Simulación de escenarios

Diseño y formulación



Rendición de cuentas



Publicación de estadísticas oficiales

Implementación

Focalización geográfica usando información del censo, información geoespacial

Seguimiento y evaluación

Uso de microdatos para la medición de resultados y efectos de la política pública.

## Objetivos de desarrollo sostenible

Los indicadores de los Objetivos de Desarrollo Sostenible deberían desglosarse, siempre que fuera pertinente, por:

- Ingreso
- Sexo
- Edad
- Raza
- Etnicidad
- Estado migratorio
- Discapacidad
- Ubicación geográfica

u otras características, de conformidad con los Principios Fundamentales de las Estadísticas Oficiales (resolución 68/261 de la Asamblea General).



# Requerimientos y necesidades

## Requerimientos



## Necesidades

### Requerimientos

Se le ha solicitado al DANE en diferentes ocasiones generar la desagregación de variables actualizadas a través de encuestas o censos.

Ej: Pobreza a nivel municipal, Pobreza para grupos étnicos

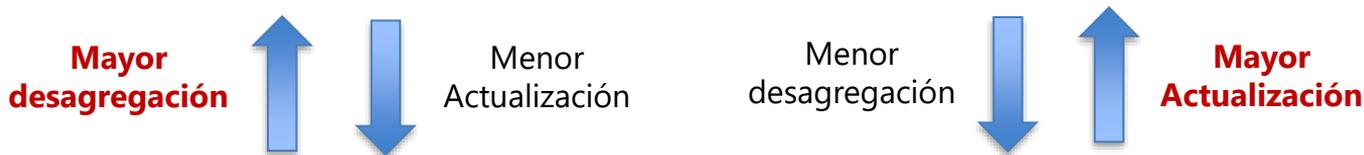
### Necesidades

Esto implicaría el aumento del tamaño de muestra e inclusión de nuevos estratos en los diseños muestrales de nuestras encuestas

Aumentar la periodicidad del Censo Nacional de Población y Vivienda.



## Pobreza – Información Disponible



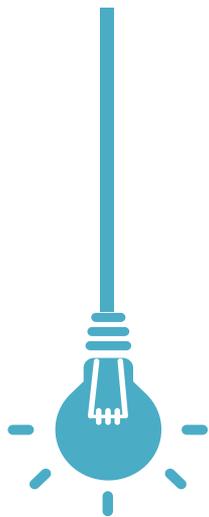
CENSO NACIONAL  
DE POBLACIÓN Y VIVIENDA 2018 • COLOMBIA

GEH MERCADO  
LABORAL

Encuesta Nacional de  
Calidad de Vida

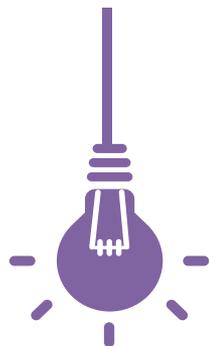
Variables	<ol style="list-style-type: none"> <li>Índice de pobreza multidimensional</li> <li>15 privaciones del IPM</li> </ol>	<ol style="list-style-type: none"> <li>Pobreza Monetaria</li> <li>Pobreza extrema</li> <li>Coficiente Gini</li> </ol>	<ol style="list-style-type: none"> <li>Índice de pobreza multidimensional</li> <li>15 privaciones del IPM</li> </ol>
Desagregación geográfica	<ol style="list-style-type: none"> <li>Sección censal</li> <li>Municipio</li> <li>Departamento</li> <li>Urbano/rural</li> <li>Nacional</li> </ol>	<ol style="list-style-type: none"> <li>24 departamentos</li> <li>Urbano/rural</li> <li>Nacional</li> </ol>	<ol style="list-style-type: none"> <li>33 departamentos</li> <li>Urbano/rural</li> <li>Nacional</li> </ol>
Periodicidad	10 años	Anual	Anual

## Fuentes de información disponibles para la integración de datos



Censos

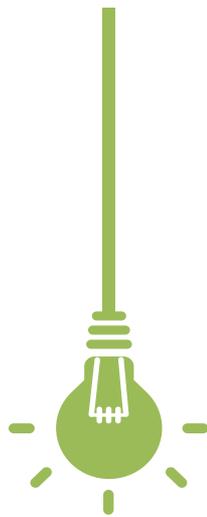
CENSO NACIONAL  
DE POBLACIÓN Y VIVIENDA 2018 • COLOMBIA



Registros  
administrativos

RELAB

SEN<sup>v2.0</sup>  
Sistema Estadístico  
Nacional • Colombia



Encuestas

GEIA MERCADO  
LABORAL

ECV

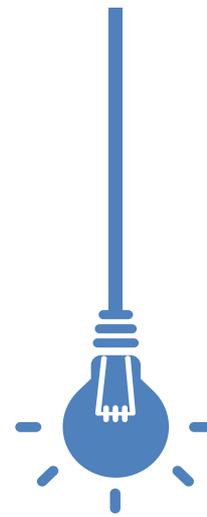
ENA



Información  
geoespacial



Imágenes  
satelitales



Fuentes  
alternas

Web scraping  
APPS  
IA  
Big Data

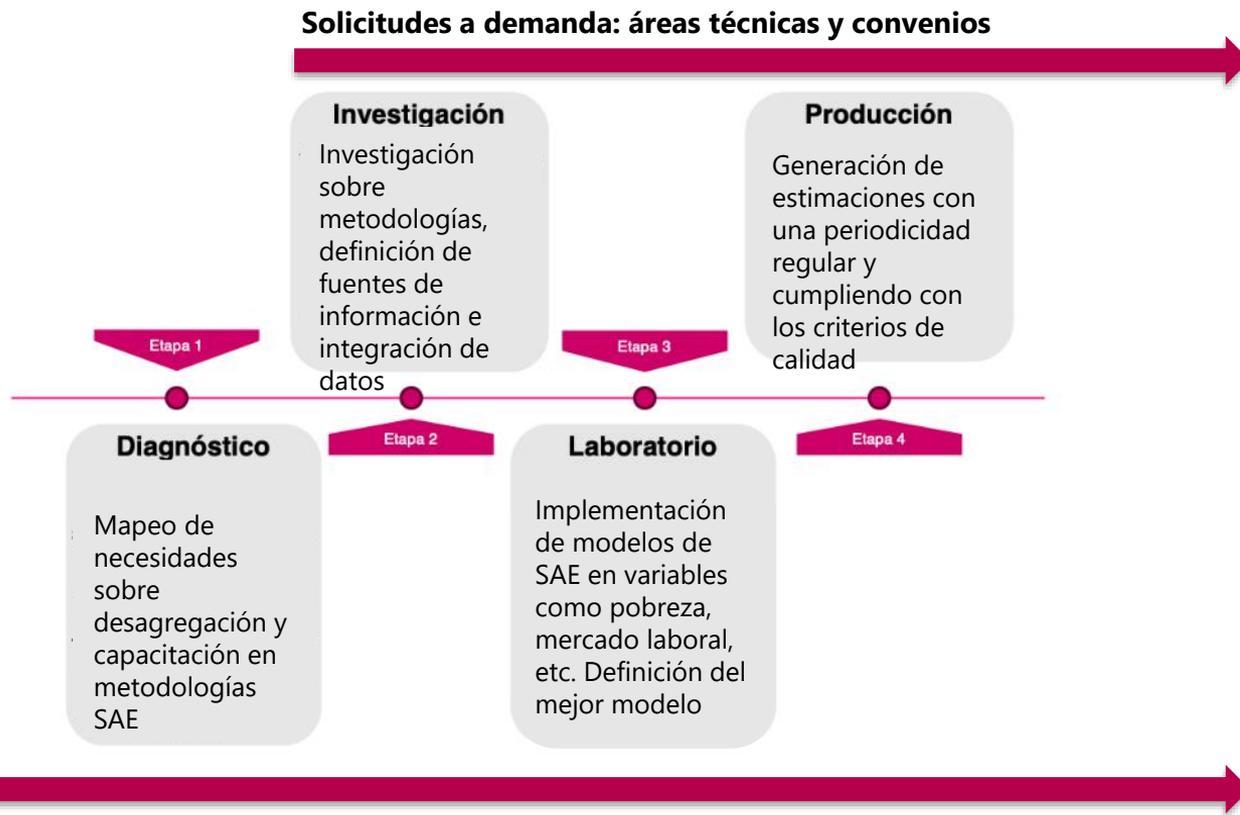
**DANE** 70 AÑOS  
INFORMACIÓN PARA TODOS

**Cómo hemos avanzado en DANE?**



**GOBIERNO DE COLOMBIA**

## Plan de trabajo SAE



**Investigación, prueba y aplicación de nuevas metodologías, como: Machine Learning y Big Data para la inclusión de variables provenientes de: fuente alternas, temporales, espaciales, registros administrativos, etc.**

En el marco de la iniciativa Data4Now el DANE propuso estimar la pobreza monetaria a nivel municipal en 2018 usando información convencional (CNPV-2018 y GEIH-2018) y no convencional (RR.AA.).

- En línea con el propósito transversal de La Agenda 2030 para el Desarrollo Sostenible *para que nadie se quede atrás*, la iniciativa Data4Now busca incentivar el uso de métodos y herramientas que mejoren la oportunidad, la cobertura y la calidad de los datos a través de la colaboración, el apoyo técnico, y el intercambio de información.
- En reconocimiento de Colombia con los ODS, especialmente con el ODS **1 Poner fin a la pobreza en todas sus formas y en todo el mundo**, y con el objetivo de contar con datos que representen a una mayor cantidad de personas y que permitan una mejor toma de decisiones, el **DANE** ha comenzado a predecir la Pobreza Monetaria a nivel municipal, anualmente, a través de técnicas de estimación de áreas pequeñas.



## Use of satellite images for computing the SDG 11 indicators

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.

# Poverty mapping

## Integration of alternative sources of information in the statistical process

### First approach we are currently working on:



#### 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).

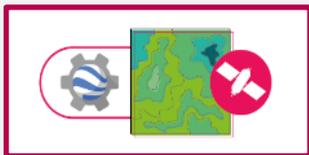
#### Goal:

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

#### Sources:

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

### Methodology:



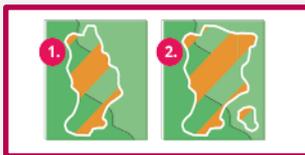
#### Compile

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



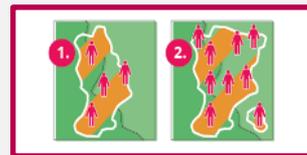
#### Input

- 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.



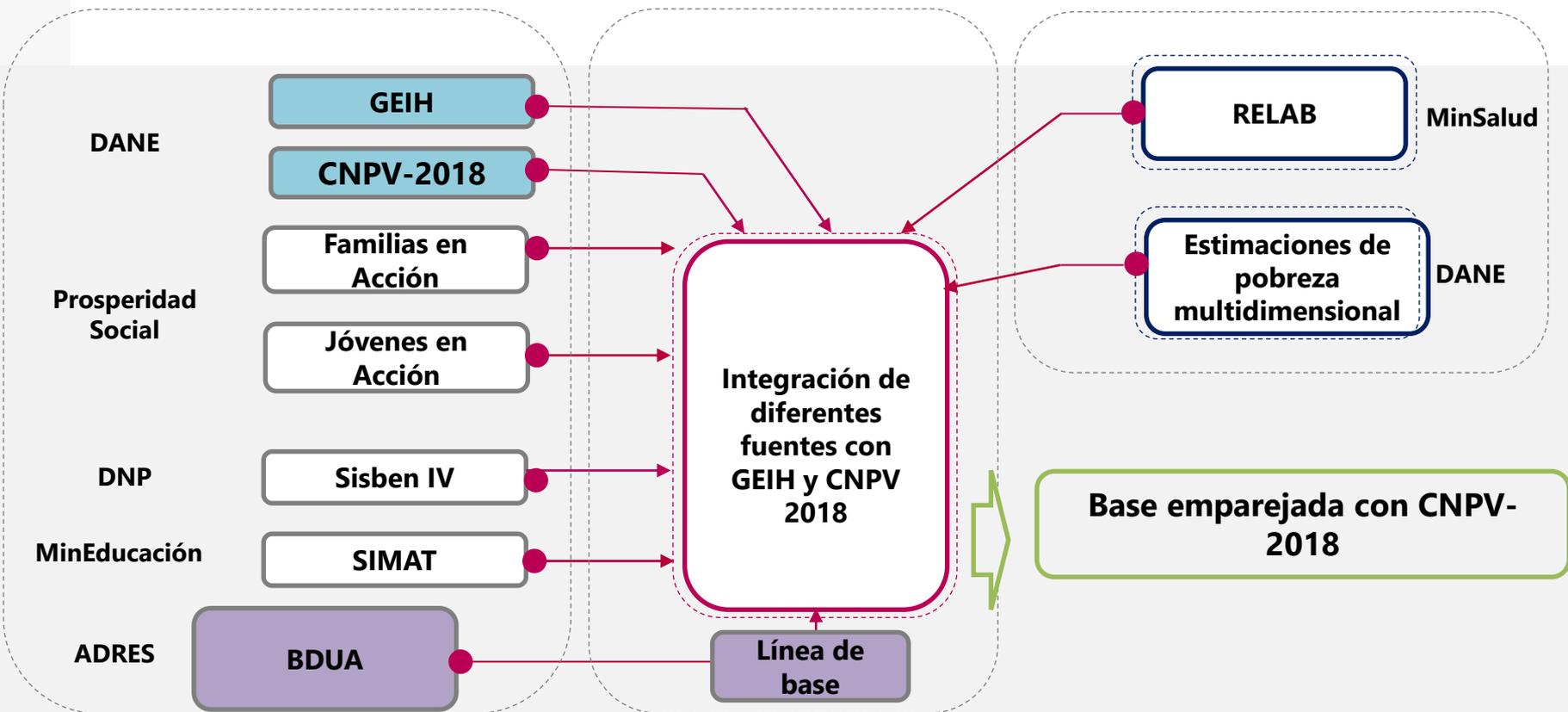
## Mapas de pobreza

Implementar modelos de Estimación de Áreas Pequeñas aprovechando la integración de información disponible en DANE y las entidades del SEN, nos dan una solución a las solicitudes de desagregación sin aumentar los costos de los operativos de recolección.

Los mapas de pobreza son el producto más representativo de este tipo de modelos, que al desarrollarlos permitirá a los tomadores de decisiones poder realizar análisis, como:



## Fuentes de información



## Dissemination of geo-referenced statistical information:

### 1. Vulnerability Geovisor



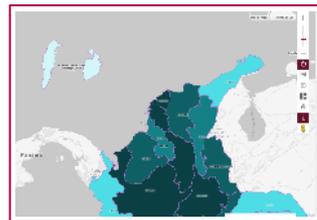
- Comorbidities.
- Adult population.
- Overcrowding.
- Population Density.
- Intergenerational risk.

### 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





# 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 informality 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:

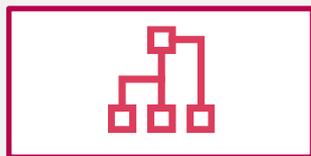
- Social security register – PILA.
- Household survey – GEIH.

### Methodology:



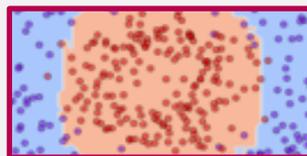
#### Compile

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



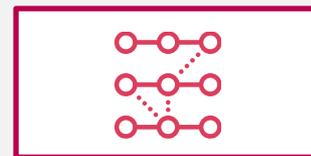
#### Modelling

- Machine learning.
- Classification model – Random Forest.



#### Estimate

- Random Forest model.
- Impute GEIH microdata



#### Results and validation

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

# 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.

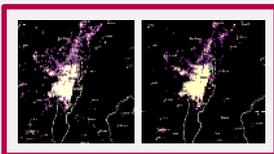
#### Goal:

- 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).



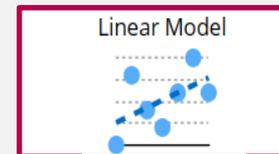
#### Georeference firms

- 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.

Migración con modelos de área (anterior presentación de Mariana Ospina).

Estimaciones para planificación familiar a través de modelos de áreas pequeñas (anterior presentación de Carlos Ramírez y Lina Sánchez).

MUCHO TRABAJO AUN MAS POR HACER!!



**DANE** 70 AÑOS  
INFORMACIÓN PARA TODOS

**Desde DANE como  
apoyamos al ciclo  
de política pública  
con datos**



**GOBIERNO DE COLOMBIA**