

Building a Global Conceptual Framework on Citizen Data

Instructions: Click on the link to access each author's presentation.

Organiser: Haoyi Chen

Chair: Faryal Ahmed

Discussant: [Parthasarathi Lahiri](#)

Participants:

[Graciela Marquez:](#) Citizen Data: National Survey on Sexual and Gender Diversity (ENDISEG) – 2021

[Andrea Ramírez Pisco:](#) Building a Citizen-Generated Data roadmap for the National Statistical System of Colombia

Sarah Omache:*

* Work presentation not available or non-existent



Citizen Data

NATIONAL SURVEY ON SEXUAL AND GENDER DIVERSITY (ENDISEG) - 2021

Graciela Márquez
President
INEGI - MÉXICO



Content

- **Background**
 - Previous surveys
- **The road to ENDISEG 2021**
 - Partnerships through the design of ENDISEG
 - Training interviewers
 - Citizen engagement
- **ENDISEG : Technical characteristics**
- **Lessons learnt**





Background Previous Surveys

Previous Surveys

Year	Name	Characteristics	Public entities	Civil Society Organizations	Level of Citizen Participation
2014	Homophobia and the World of Work/Labour Environment (survey)	2,284 completed questionnaires Nationwide call to participate	National Commission of Human Rights (CNDH)	Espolea, A.C.	Partnership
2015	Knowing our Diversity (survey)	793 completed questionnaires. Participants on the 37 th Pride Parade. 3 087 completed online questionnaires.	National Pedagogical University	Asociación Civil Actúa, DF..	Partnership
2018	Discrimination based on Sexual Orientation or Gender Identity (survey)	9 959 completed online questionnaires. Voluntary participation of LGBTI+ population.	National Council to Prevent Discrimination (CONAPRED) and CNDH	-	Information



Background

The road towards
ENDISEG

ENDISEG: Technical
characteristics

Lessons learnt



The road to ENDISEG



The road to ENDISEG: public consultations

2016

National Survey on Discrimination

CONAPRED requested to include a module on gender identity and sexual orientation.

INEGI proposed to further examine the inclusion of such questions in other statistical programmes.

2018



2020 Population and Housing Census

INEGI proposed a specialized survey given the complexity of the topic.

CNDH and Civil Society Organizations requested to include questions on LGBTI+ rights



Background

The road towards
ENDISEG

ENDISEG: Technical
characteristics

Lessons learnt

ENDISEG: Design through partnerships

2018

February

INEGI, CNDH and CSOs agreed to establish a working group to collect data on the LGBTI+ population.

INEGI conducted exploratory in-depth interviews to self-declared LGBTI+ people.

April June

INEGI and CONAPRED review of the collection instrument for the pilot exercise.

2019

July

Budget and further adjustments of the collection instrument by INEGI and CONAPRED.

Participants in working meetings:

- Government: INEGI, CNDH, CONAPRED
- Civil Society Organizations: Federación Mexicana de Educación Sexual y Sexología, Colectivo Reinas Chulas Cabaret, Almas Cautivas, Letra S, Fundación Arcoiris, Balance A.C.
- Review of the instrument design for the survey tests to check wording and flow of the questions.
- Test of the preliminary questionnaire in 22 households.
- Test of the preliminary questionnaire to self-declared LGBTI+ people.



ENDISEG: Design through partnerships

Budget and further adjustments of the collection instrument by INEGI and CONAPRED.

2019



Participants in working meetings:

- Government: INEGI, CNDH, CONAPRED
- Civil Society Organizations: Almas Cuativas, Fundación Arcoiris, Armario Abierto, Intersecta.
- Academia and researchers
- Logistics for the pilot exercise in 32 states and final adjustments to the collection instrument.

December

Survey test.

32 states.
5 000 households.

2021

March

INEGI team conducted in-depth interviews to 37 self-declared LGBTI+ people.

May

Participants in working meeting:

INEGI and CONAPRED.
Colectivo Brújula Intersexual.
Review of the inclusion of the intersexual population question in the survey.



Background

The road towards
ENDISEG

ENDISEG: Technical
characteristics

Lessons learnt

Training interviewers

Active role of **Civil Society Organizations** and **CONAPRED** throughout interviewers' training period.



Clarification
of concepts



Conferences and mock
interviews with self-
declared volunteers



Citizen engagement in ENDISEG



Previous citizen-driven initiatives



Building objectives jointly



Active participation in working groups & reviews of preliminary collection instruments.



Constant sharing of technical skills, knowledge & experiences





ENDISEG: Technical characteristics

ENDISEG: TECHNICAL CHARACTERISTICS

Collection period



From August 23rd, 2021 to January 16th, 2022

Observation unit



- Housing units
- Households
- People aged 15 years or over

Collection method



Electronic questionnaire; face-to-face interviews, and **self-reported** by **audio** for sensitive topics.

Sampling scheme

$$n^2 \sum_{z^2} p^x = ()$$

- Probabilistic
- Three-stage
- Stratified
- By clusters

Geographical coverage



National and by state (only for the LGBTI+ population estimates).

Sample size



- **43 737 housing** units representing **36.4 million** housing units.
- **44 189** people aged 15 years or over on the effective sample, representing **97.2 million** people aged 15 years or over.

ENDISEG: questionnaire structure

1 Household informant



Characteristics of housing units



Household identification



Sociodemographic characteristics

2 Selected person



Personal characteristics



Childhood



Adolescence



Opinion and social openness


Audio interview



Emotional health



Sexuality



Sexual orientation



Social rejection



Gender identity





LESSONS LEARNED



LESSONS LEARNED

ENDISEG showed how INEGI can go beyond conventional public consultations.



Innovations derived from citizen data can be applied to other statistical programmes.



Information generated through citizen engagement enjoy greater trust amongst users



Models of citizens participation enrich design & collection methods by closing critical gaps from traditional approaches.



Citizens & NSO mutually benefit from better data and improved learning of information generation processes.



Empowered citizens become agents of change.





Thank you!





Building a Citizen-Generated Data roadmap for the National Statistical System of Colombia

Andrea Ramírez Pisco
Deputy Director – DANE



Content

1. National Statistical System of Colombia

- Law 2335 of 2023
- Data Ecosystem of the SEN

2. Citizen-Generated Data in Colombia

- Motivation
- Roadmap 2024

3. Examples citizen-Generated Data in Colombia

- App Diversa
- Vital Statistics
- Instituto Humbolt

Goals and challenges

National Statistical System of Colombia -NSS-



Law No. 2335 of 2023

“By which provisions are issued on official statistics in the country”

The Congress of
Colombia,

DECREES

ARTICLE 1. OBJECTIVE AND APPLICATION SCOPE.

This Law establishes the general legal framework for the planning, production, dissemination, and administration of official statistics in the country.

The provisions of this Law shall apply to statistical operations, administrative records and data collected or obtained for statistical purposes by the producers of official statistics within the framework of the National Statistical System (NSS).

National Statistical System (NSS) Sistema Estadístico Nacional (SEN)

Definition and structure

¿What is the NSS?

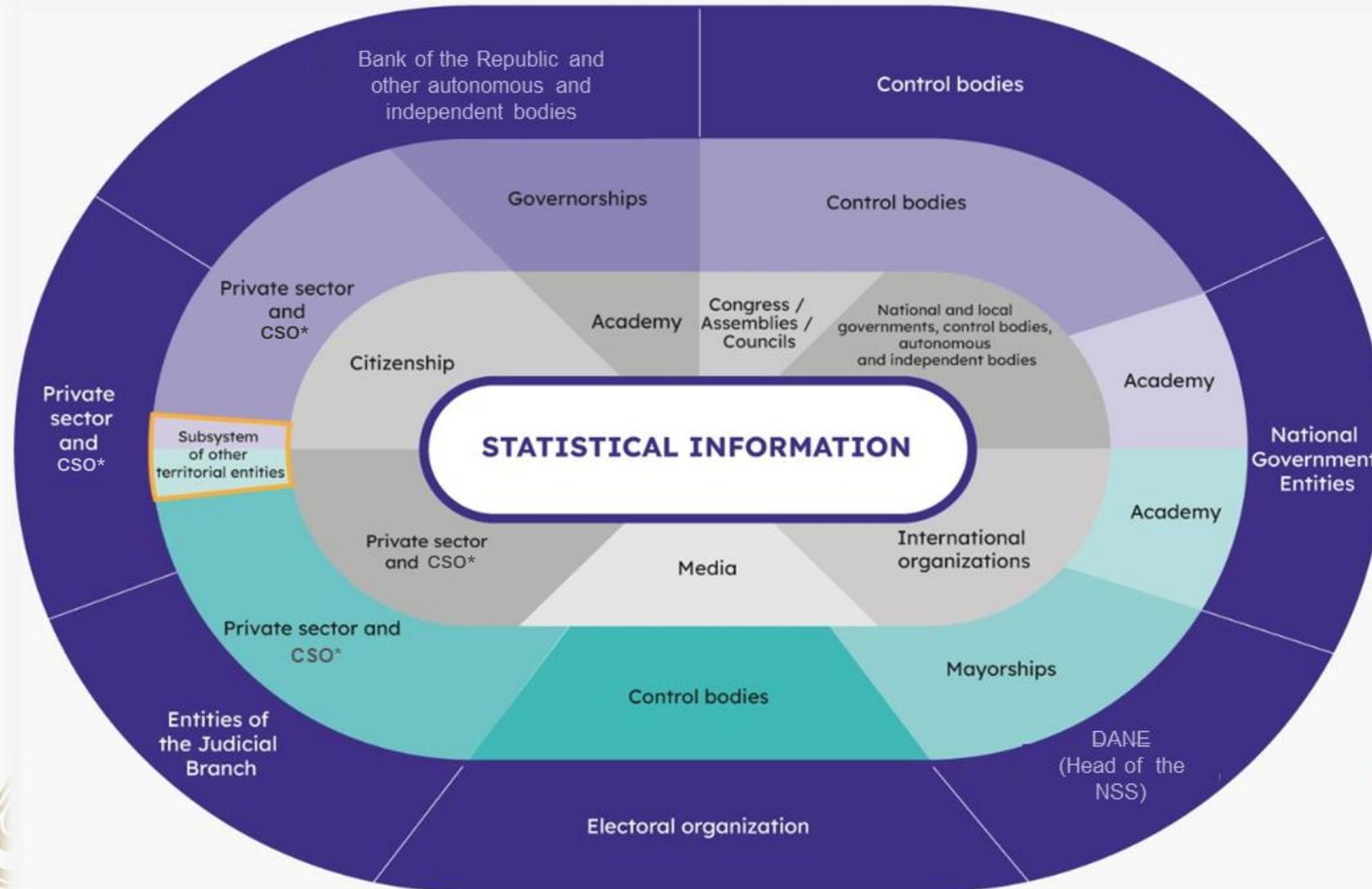
Articulated set of components that ensure the production and dissemination of official statistics at the national and local levels that the country requires, in an organized and systematic manner.

¿What are the components of the NSS?

- Entities and organizations that produce statistical information
- People and organizations responsible for administrative records
- Users
- Technical processes and instruments for coordination.
- Technological infrastructure
- Policies, principles and information sources
- Human talent

National Statistical System (NSS) – Sistema Estadístico Nacional (SEN)

Data Ecosystem of the SEN



Citizen- Generated Data in Colombia



Citizen – generated Data Copenhagen Framework of Citizen Science

'Citizen Data' are defined as data originated from initiatives where citizens either **initiate** or are **sufficiently engaged** in **the design and/or collection stages of the data value chain**, regardless of whether these data is integrated into official statistics.

Operational definition of data generated by citizens – Copenhagen Framework

Characteristics of the Citizen generated data:



Types of initiative
for data collection



Stages of the data
value chain in which
citizens participate

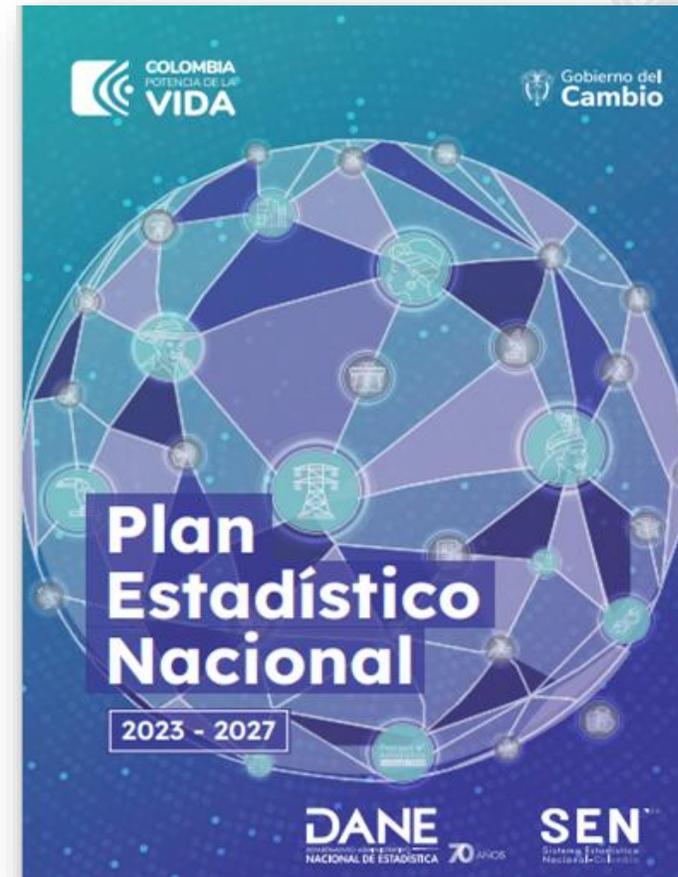


Levels of citizen
participation

Colombian National Statistical Plan

Document of **recommendations, good practices, guidelines, mechanisms and collaborations for citizen participation** in the generation, validation and use of statistical data.

Strategy 5, Action 15, Goal 35 of the Colombian National Statistical Plan (PEN) 2023 – 2027



Citizen-generated Data in Colombia
Mechanisms to achieve the objective

Next quality framework actions

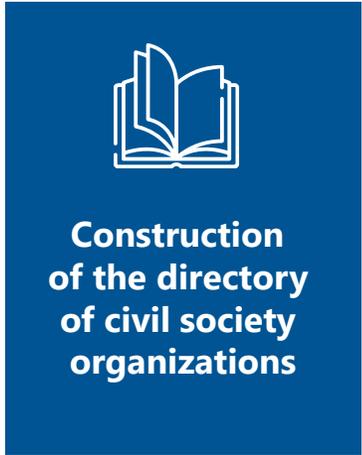
Workplan Citizen Generated Data –CGD- in DANE

Literature review and international references



February 2024

Construction of the directory of civil society organizations

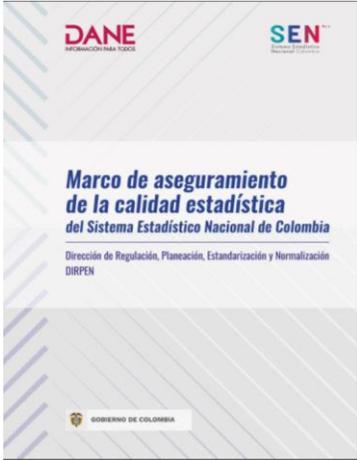


February – June 2024

Event with civil society organizations
June



Statistical quality assurance framework



December 2024

Identification of the ecosystem actors



Document of recommendations, good practices and experiences

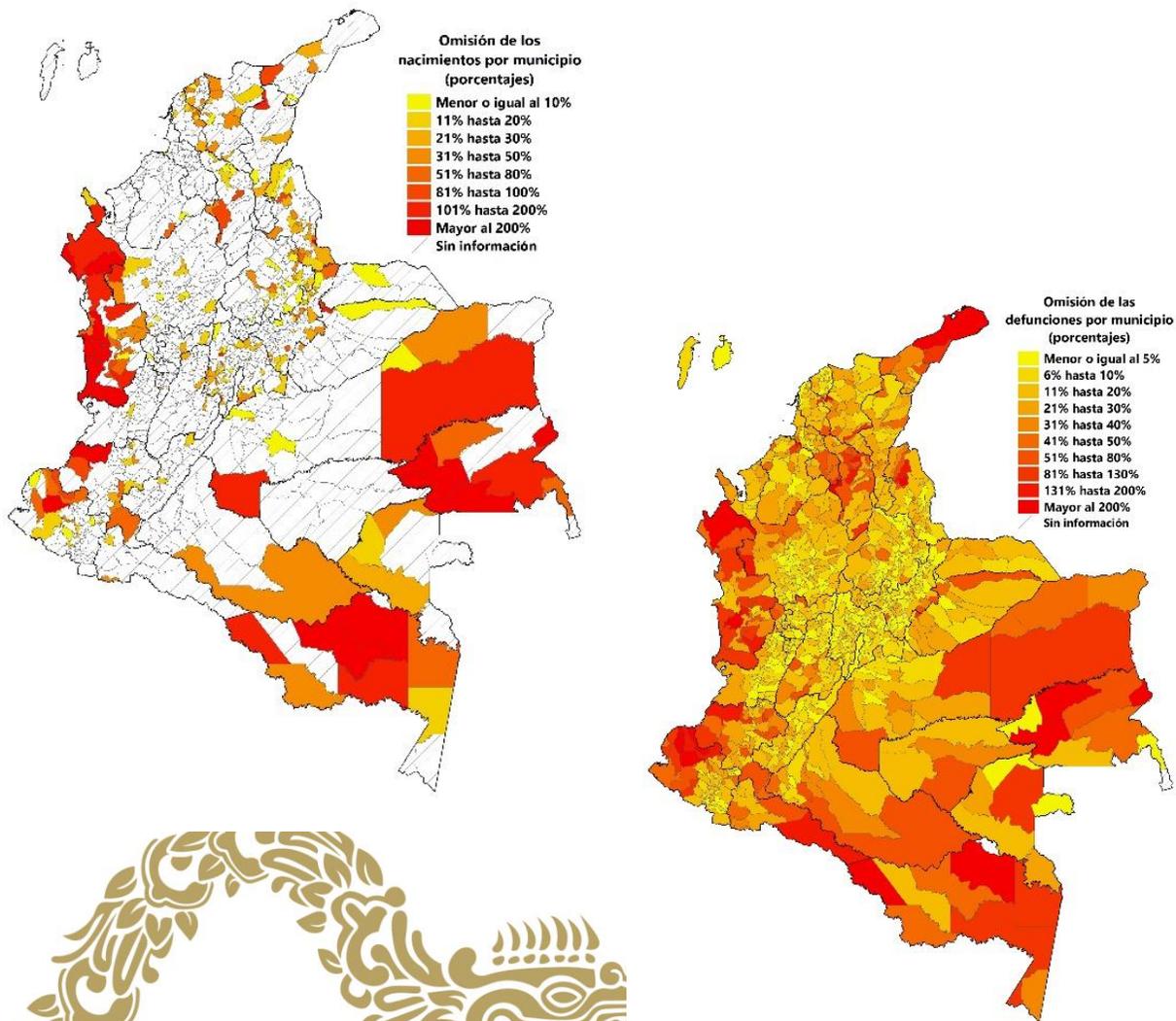
Pilot projects with CSOs

Examples citizen- Generated Data in Colombia



Citizen-generated Data in Colombia

Notification strategy for reducing the under-registration of births in dispersed places

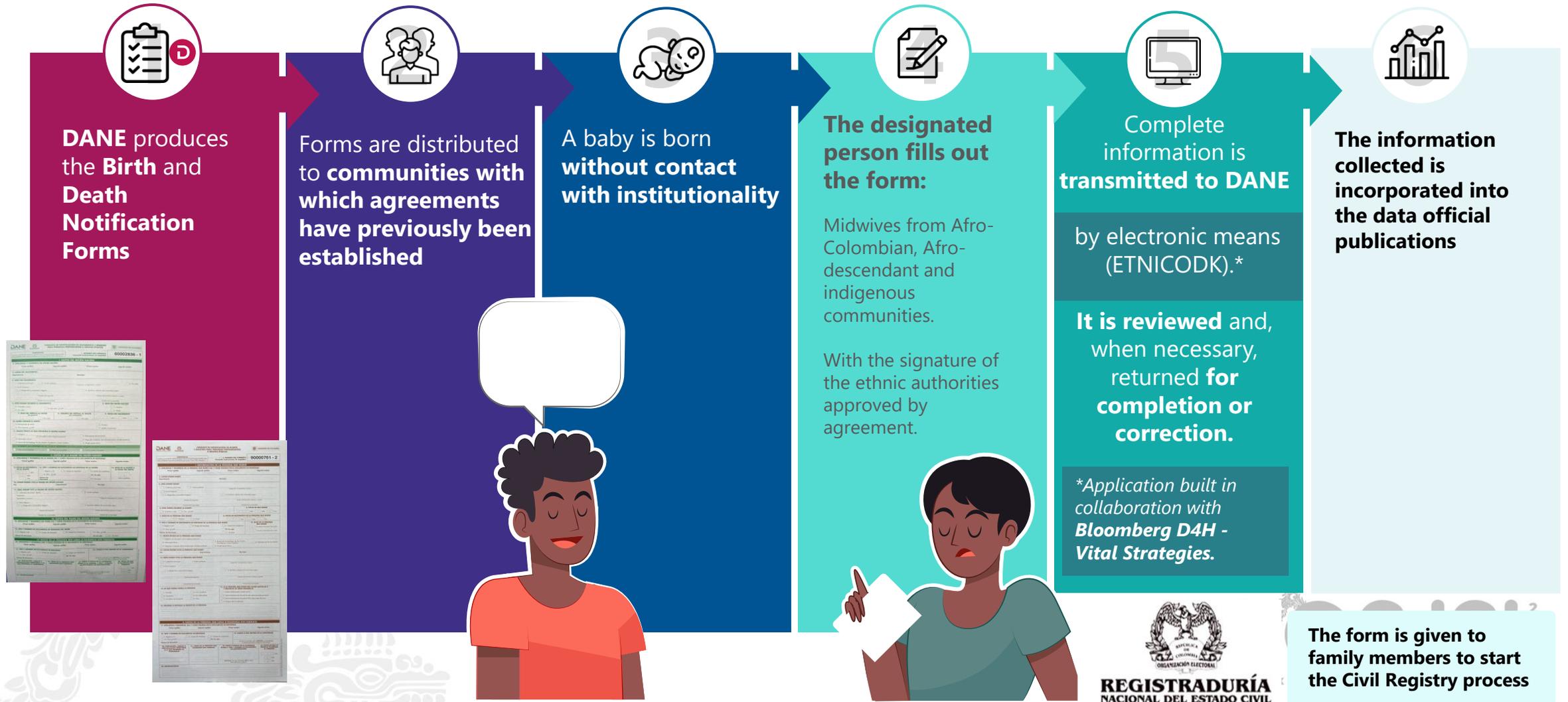


In the cross-referencing of information with the CNPV 2018, it was identified that the peripheral areas of the country are those with the greatest underreporting of Vital Statistics.

For this reason, the process of implementing a **notification strategy started and currently focuses on reducing the under-registration of births in dispersed places**, so that they are included both in national statistics and in the civil registry.

This is a guarantee of access to identification rights principally in the ethnic-racial groups of the country.

Birth Notification Strategy for Ethnic-Racial Groups



Due to logistical constraints, only the under-registration of births in ethnic communities is currently being addressed

Legal Guarantee of Identity through Reporting Formats for Ethnic-Racial Groups



DANE focuses on the ethnic group, and through fieldwork carries out awareness and training processes with the community on the importance of the civil registration of births and deaths and of being part of vital statistics.

After raising awareness, a formal agreement is reached with the ethnic authorities to begin reporting births and deaths that occur in their community.

Together



The National Registry of Civil Status and DANE, together, generate an administrative act authorizing a certain ethnic group to notify both births and deaths that occur within its territory.



With this authorization, all local Registries and notaries' offices with registry functions, are obliged to issue the Civil Registries of Birth and Death with the antecedent already generated.



Citizen-generated Data in Colombia

Project carried out-App Diversa



Pilot strategy that seeks to **innovate in data collection methods** to be implemented in statistical production processes.



Execution period:

Two months (March 1st to April 30th, 2023)



Users:

People over 18 years old



Main objective:

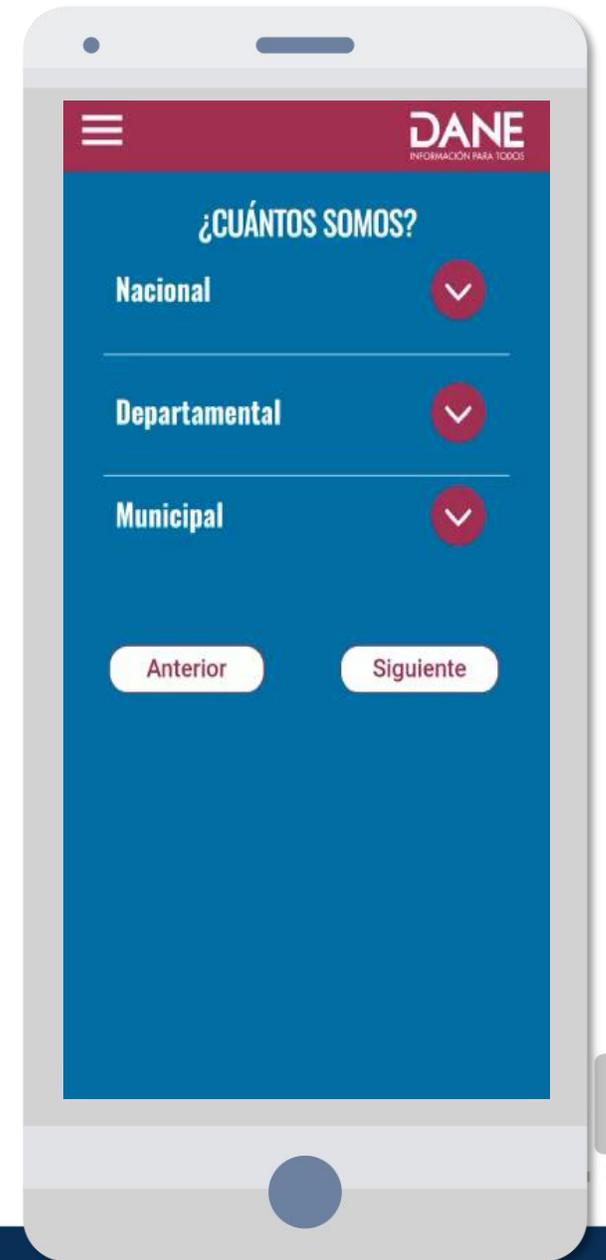
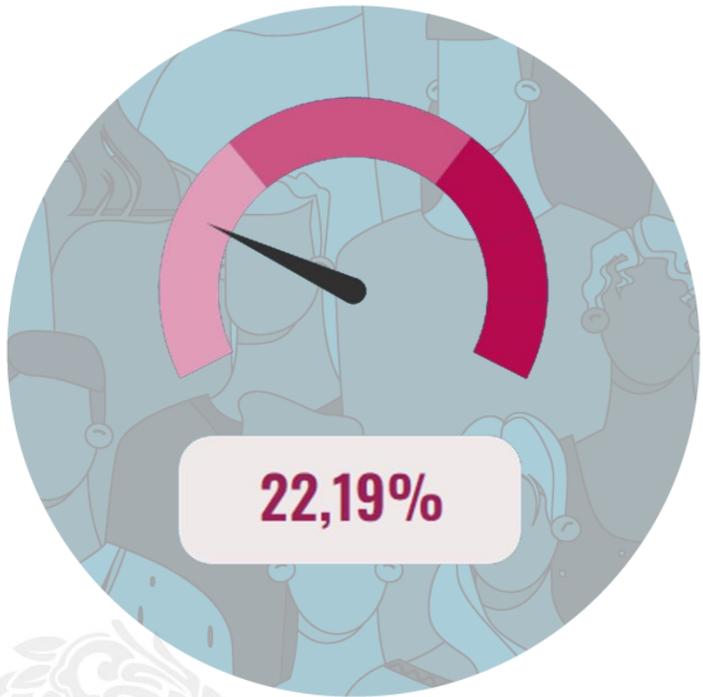
Measure the perception of people experiencing discrimination-related events.



It promotes **two-way interactions**, obtaining feedback from citizens.



Discrimination Risk Indicator



General results



Five-star ratings
(69%)



Age of users:
Adults between
18 and 47 years old.



Sex of users:
Women were the users
who contributed the most.



**Socio-economic level
(Strata):**
Strata 2, 3 and 4 showed
greater representation.



Place of residence:
37.7% of users reside in **Bogotá**, 11.2% in **Antioquia**, 5.3% in **Córdoba**, being the departments with the highest representation.



Ethnic group:
92.7% of users do not belong to any ethnic group, 5.2% belong to the black, mulatto, afrodescendants, afro-Colombian groups, and 1.4% belong to indigenous groups.



Disability:
92.5% of users do not have any special health condition or disability.

Citizen-generated Data in Colombia

eBird

The Von Humboldt Institute has been a pioneer in Colombia using citizen data for collecting, processing, and analyzing data for producing environmental statistics.

The Institute has undertaken numerous global biodiversity projects like **eBird**, a global initiative for birding, or the Colombian National Biodiversity System in which **citizens could gather data and collaboratively participate in different phases of the scientific method.**

In the case of e-Bird, citizens could register information on a web repository on birding or they can use the eBird Mobile app. The **people's sightings of birds could contribute to conservation decisions, peer-reviewed research, and fostering awareness of environmental protection.**

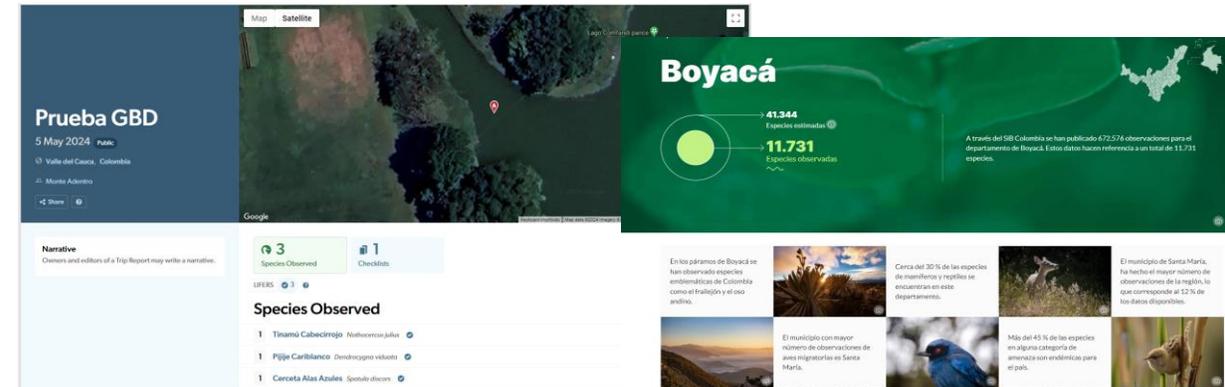
Colombia Partners

eBird Colombia

El portal regional de eBird Colombia es un proyecto colaborativo que se lleva a cabo gracias a la participación de las siguientes organizaciones:



COLABORADORES



Goals and challenges



GOALS

- **Community Empowerment:** Engaging citizens in data collection empowers them to take an active role in addressing issues affecting their communities.
- **Advocacy and Awareness:** Data collected from citizens can be used to advocate for policy changes, raise awareness about specific issues, and mobilize resources.
- **Participatory Decision-Making:** Involving citizens in data collection processes promotes democratic values and can lead to more inclusive decision-making processes.

CHALLENGES

- **Data Quality:** Ensuring the accuracy and reliability of citizen-generated data can be challenging, as it may be subject to biases, errors, or manipulation.
- **Digital Divide:** Access to technology and digital literacy can limit the participation of certain groups in citizen-generated data initiatives, leading to potential biases in the data.
- **Sustainability:** Maintaining citizen engagement in data collection efforts over the long term requires ongoing resources and support, which can be challenging to sustain.
- **Data Integration:** Integrating citizen-generated data with existing datasets and information systems can be complex, particularly if different data sources use different formats or standards.

Thank you

ANDREA RAMIREZ PISCO

Deputy director – DANE
aramirezp@dane.gov.co

Discussion

Partha Lahiri
Joint Program in Survey Methodology (JPSM)

ISI-IAOS Conference, Mexico City, May 15-17, 2024

Three presentations

- Graciela Marquez, INEGI, Mexico
- Andrea Ramirez Pisco, DANE, Colombia
- Sarah Omache, KNBS, Kenya

Presentation by Graciela Marquez

- Good collaboration between Government and Civil society organizations in constructing survey instrument and data collection.
- Self reporting by audio for sensitive questions helps minimizing measurement errors
- Probability sample survey could provide robust estimates for general population at the national level
- Estimates for LGBTQ+ population could be challenging due to small sample size.

Presentation by Andrea Ramirez Pisco

- Workplan (literature review, construction of list of CSOs, events with CSOs, statistical quality assurance framework) for CGD in DANE is commendable.
- Nice Examples of CGD:
 - Reduction of under registration of births in dispersed places
 - App Diversa use to estimate discriminant risk indicator
 - eBird for environmental data generation
- Challenges
 - Data quality
 - Data integration

Presentation by Sarah Omache

- Recognition of CDG as rich sources of data to fill in data gaps
- Development of a framework for validating CDG using PARIS21 guidelines and borrowing best practices used by different countries
- Support to CSOs
- Challenges in determining gateway entry and scoring matrix

The Copenhagen framework of Citizen Science

- Inclusivity
- Engagement
- Quality
- Ethics
- Openness
- Impact

Some comments

- Data sources used by NSS: Census Data, Administrative records, probability samples
- Alternative data sources: nonprobability samples, citizen generated data, various big data
- Can quality of CGD be assessed using probability pilot samples?
- Can alternative data sources be used in dealing with emerging problems (e.g., producing statistics more frequently, more granularity,...)?
- Capacity building