

## ACTA DE REUNIÓN N°4 AYUDA DE MEMORIA

Ciudad: Bogotá

Lugar: Sesión virtual, plataforma Teams

Tema: Presentación proyectos DANE y aportes expertos

Hora: 9:00 a.m. a 11:00 a.m.

Fecha: 03/08/2021

Dependencia responsable: Secretaría Técnica CASEN (DIRPEN)

### Participantes

#### ***Miembros de la Sala Especializada de Modernización Tecnológica***

León Darío Parra Bernal  
Jorge Andrés Gallego Durán  
Valérie Gauthier Umaña

Coordinadora - Planificación y Articulación Estadística

Mónica Patricia Pinzón

#### ***Departamento Administrativo Nacional de Estadística (DANE)***

Directora Técnica DIRPEN  
Julieth Solano

Profesional GIT Planificación y Articulación Estadística (DIRPEN)

Ruth Constanza Triana

Asesora DIRPEN y Dirección General  
Elizabeth Moreno Barbosa

DIRPEN – Gobierno de Datos  
Mateo Cardona

Oficina de Sistemas  
Marly De Moya Amaris

DIRPEN  
Francisco Javier Lara

Oficina de Sistemas  
Laura Milena Muñoz

DIRPEN  
Alexander González Coca

Oficina de Sistemas  
Camilo Andrés Cáceres

Profesional GIT Planificación y Articulación Estadística (DIRPEN)

Juan José Galeano

Asesor de la Dirección General  
Juan Sebastián Ordoñez

Profesional GIT Planificación y Articulación Estadística (DIRPEN)

María José Prada

Asesora de la Dirección General  
Karen Lizeth Chávez

Dirección de Censos y Demografía  
Lelio Alejandro Arias

## Orden del día

1. Solicitud autorización para grabar la reunión
2. Verificación del quórum
  - I. León Darío Parra Bernal
  - II. Valérie Gauthier Umaña
  - III. Jorge Andrés Gallego Duran
  - IV. Mónica Patricia Pinzón
  - V. Elizabeth Moreno Barbosa
  - VI. Juan Sebastián Ordoñez
3. Síntesis reunión del pasado 13 de julio (María José Prada – 5 minutos)
4. Discusión de preguntas orientadoras pendientes (Expertos de la sala – 20 minutos)
5. Presentación Comité de Administración de Datos - CAD (Mateo Cardona – 20 minutos)
6. Presentación proyecto Pobreza Multidimensional (Lelio Alejandro Arias – 30 minutos)
7. Presentación Migración del Censo de Edificaciones (Laura Milena Muñoz, Francisco Lara – 20 minutos)
8. Retroalimentación de los expertos respecto a los proyectos presentados (Expertos de la sala – 20 minutos)
9. Conclusiones y cierre (María José Prada - 5 minutos)

## Desarrollo

### Objetivo

Presentar los proyectos desarrollados por el DANE, a fin de recibir retroalimentación por parte de los expertos y fortalecer estos procesos, teniendo en cuenta las preguntas orientadoras que surgen de la presentación de cada proyecto.

### 1. Solicitud autorización para grabar la reunión

Se inició la grabación de la sesión con previa autorización de los participantes.

## **2. Verificación del quórum**

Se verificó el quórum tanto de los expertos de la Sala, como del Departamento Administrativo Nacional de Estadística (DANE).

## **3. Síntesis reunión del pasado 13 de julio (María José Prada – 5 minutos)**

María José Prada inició la reunión haciendo una recopilación de los temas discutidos en las reuniones previas. Mencionó que, en la reunión del 13 de julio, Juan Sebastián Ordoñez presentó los proyectos de Automatización de Procesos de Calidad y Cruces Difusos y que de estos quedó pendiente discutir las preguntas orientadoras con los expertos. Asimismo, se mencionó que durante la presente sesión se haría la exposición del proyecto del Comité de Administración de Datos que quedó pendiente en la sesión del 13 de julio por temas de tiempo.

Por otro lado, se mencionaron los comentarios de los expertos Valerie y Jorge realizados en la sesión anterior y se explicó que se debe definir el plan de trabajo para recibir el apoyo del profesor León Darío Parra en el proyecto de Automatización de Procesos

Del mismo modo, Julieth Solano, se encargó de hacer el saludo inicial de la reunión y mencionó que, aunque no pudo estar en las reuniones anteriores, ya se encuentra al día con los temas que se han discutido y que está atenta a todas las sugerencias y apoyos que puedan aportar los expertos.

## **4. Discusión de preguntas orientadoras pendientes (Expertos de la sala – 20 minutos)**

Esta sección de la reunión inició con la lectura de las preguntas orientadoras que quedaron pendientes acerca del proyecto de Automatización de Procesos de Calidad. Posteriormente, el profesor León Darío respondió a estas preguntas mencionando que es pertinente dividir el instrumento de automatización para poder evaluarlo por partes y evitar errores. Asimismo, mencionó frente a otra de las preguntas que vale la pena reforzar el glosario para fortalecer los términos que pueden generar dificultades de comprensión.

Luego de esta intervención, el profesor Jorge Gallego preguntó sobre los incentivos que existen para diligenciar los formularios. Además, mencionó que definitivamente se debe hacer una división para evaluar el proceso. Continúo mencionando que pueden existir dificultades en cuanto a las capacitaciones de los funcionarios para llevar a cabo el proyecto, dado que en el sector público hay una constante rotación de contratistas, que pueden generar costos muy altos para capacitar continuamente. Por último, mencionó que este mismo proceso de capacitación puede ralentizar la aplicación del proyecto por lo que hay un *trade off* entre calidad y efectividad.

Después, Valerie Gauthier mencionó que más allá de fortalecer los glosarios se debe entender lo que hace falta en el proceso, mencionó que se puede hacer uso de preguntas frecuentes o propuestas similares para apoyar mejor el diligenciamiento de los formularios.

Para responder a los comentarios de los expertos, Karen Chávez, respondió a Jorge Gallego, frente al tema del *trade off* entre calidez y efectividad, mencionando que la idea precisamente es automatizar ciertos procesos, incluso los de capacidad instalada para evitar ciertos costos de capacitación y garantizar los accesos a información a través de la nube.

Mateo Cardona mencionó respecto a la pregunta del profesor Jorge Gallego, que es muy válido segmentar los procesos de automatización. Frente a la sugerencia de Valerie de usar preguntas frecuentes, mencionó que se pueden usar también otros medios virtuales como videos y presentaciones para facilitar los diligenciamientos. Por último, respecto al *trade off* mencionado por Jorge, Juan Sebastián Ordoñez respondió que para él es más importante tener calidad en los procesos en un principio para generar cultura estadística dentro de las otras entidades que ahora están en el proceso de convertirse en productores de estadísticas, para que estas sean coherentes. A esto, Jorge respondió que está de acuerdo, sin embargo, hace hincapié en la magnitud e importancia del tema ya que es urgente definir si ciertas políticas pueden continuar.

Para concluir el punto, Julieth mencionó que la revisión de pares pretende resolver todas estas dudas y *trade offs* para garantizar que la información esté disponible de manera oportuna y que a la vez el Comité de Administración de Datos tiene como objetivo realizar la gobernanza y la gestión de todos los registros y datos. A lo anterior, Jorge respondió que va a compartir el trabajo que se hace dentro del Behavioral Insights Team del Reino Unido en donde se trabajan estos procesos engorrosos con el fin de facilitar el trabajo, haciendo uso de incentivos monetarios y de comportamiento.

Para cerrar la sección, se hizo la lectura de las preguntas orientadoras del tema de Cruces Difusos, a lo que Jorge Gallego respondió que el tema es demasiado técnico y no tiene conocimiento. Valerie respondió a una de las preguntas haciendo sugerencias sobre el tipo de funciones de seguridad que se pueden usar (Funciones Hash) enfatizando en que, si se usan, los errores no podrán corregirse pero que igualmente son válidas y que incluso son usadas en los bancos. Del mismo modo, León Darío Parra, mencionó que no tiene comentarios al respecto porque no conoce sobre el tema.

## **5. Presentación Comité de Administración de Datos - CAD (Mateo Cardona – 20 minutos)**

Mateo Cardona inició la presentación del Comité de Administración de Datos mencionando que este tiene como objetivo principal ser un ente colaborativo entre todas las entidades que hacen parte de él. También se mencionó que, en el último año, teniendo en cuenta la pandemia, ha funcionado como medio para proponer políticas públicas relacionadas. Sin embargo, la idea es generar un piloto para fusionar las cuatro entidades que hacen parte del CAD, para que a partir de aprovechar los datos disponibles se les pueda generar más valor.

Se mencionó que el DANE dentro del CAD además de tener a cargo la secretaría técnica también puede aportar desde la perspectiva de las metodologías que se puedan llevar a cabo dentro de todas las entidades que hacen parte. Además, dentro del proyecto piloto planteado se busca que los formatos de necesidades sean lo suficientemente efectivos para que los miembros del CAD puedan responder oportunamente a los requerimientos. Esto es también parte de tener un Gobierno de Datos eficaz.

Mateo, durante su exposición también mencionó los riesgos que existen dentro del manejo de los datos y las propuestas de mitigación de riesgos que se han planteado para solucionarlos. Asimismo, comentó sobre los retos que tiene el CAD que incluyen definir el proceso de articulación de entidades y terminó su intervención explicando el cronograma que se ha manejado dentro del proyecto, en donde actualmente se están evaluando todos los procesos, teniendo en cuenta el *trade off* mencionado anteriormente, sobre calidad y agilidad de procesos.

## **6. Presentación proyecto Pobreza Multidimensional (Lelio Alejandro Arias – 30 minutos)**

Alejandro Arias inició la presentación del proyecto mencionando que con este busca generar un Indicador de Pobreza Multidimensional usando aprendizaje de máquinas aplicado a imágenes satelitales a partir del censo hecho en el 2018. Empezó mencionando y mostrando algunos mapas en donde se ve que, aunque previamente se ha intentado revisar este mismo tema, los resultados no han logrado ser completos dado que como los censos son tan complejos no hay información censal de territorios colombianos alejados que son precisamente los que mayores niveles de pobreza pueden tener. Mencionó también que Data 4 Now, apoyó al proyecto y que se aplicó un modelo de Redes Neuronales y vecinos más cercanos para identificar la información principal de las imágenes que se descargaron de Sentinel-2.

En el proceso de revisión de las imágenes también se hace una revisión de características lumínicas para determinar si los territorios analizados son rurales o urbanos. En este proceso se usan los métodos de Random Forest y Gradient Boosting Regression, en los cuales como resultado se obtiene que hay un patrón de pobreza multidimensional en donde hay mayor concentración en la Región Amazónica y Pacífica y menor en la Región Andina, como ya se había encontrado en estudios anteriores.

A manera de ejemplo, se revisó el visor de información que se encuentra en la página del DANE, en donde se ubica el municipio de Uribia en la Guajira, en el cual se observa que, en las zonas rurales, la pobreza alcanza niveles de 95% y en las urbanas el 50%. Del mismo modo, Alejandro mencionó que, toda la información se encuentra disponible en la página y que, aunque el proyecto en su primera fase ya terminó, se encuentran en desarrollo nuevas fases que buscan mejorar la precisión de los resultados, sobre todo en la resolución de las imágenes. Lo anterior, busca mejorar la implementación de políticas públicas dado que usando esta información se puede hacer una adecuada focalización.

Alejandro también mencionó que a partir del apoyo de los expertos le gustaría tener sugerencias para tener una nueva capa en el método de redes neuronales y en los procesos de regresión para evitar el proceso iterativo y que no continúen los patrones en las regiones donde hay menos información, puesto que no se puede garantizar esta información.

Respecto a este proyecto, el profesor León Darío hizo un par de recomendaciones: en primer lugar, sugirió usar el trabajo de Volvinic - Aracely Damian para complementar la revisión de literatura frente al tema de Pobreza Multidimensional, sobre todo respecto a la heterogeneidad de las regiones por lo que se debe hacer un análisis cualitativo. También mencionó que en las proyecciones del modelo se debe revisar la robustez y se debe hacer un análisis cualitativo para caracterizar mejor los datos e identificar más variables.

### **7. Presentación Migración del Censo de Edificaciones (Laura Milena Muñoz, Francisco Lara – 20 minutos)**

En esta presentación Laura Muñoz inició explicando el objetivo del proyecto de Migración del Censo de Edificaciones. Posteriormente, Ana María Mondragón explicó que este Censo pretende conocer la producción y composición de la actividad de construcción en Colombia y que para el DANE existe una variable que permite conocer la evolución de las obras y los tipos de construcciones.

En este punto, Francisco Lara comentó que el proyecto consiste en cambiar parte de la metodología del Censo, revisando las variables y llegando a la Anonimización de las variables, identificando la ubicación de las observaciones y los periodos, verificando datos atípicos para así consolidar la base de datos del Censo. A partir de la revisión de los datos atípicos para lograr la Anonimización se pasó a una segunda fase, ya que se descubrió que para las áreas y obras grandes se generan sesgos, por lo que se deben clasificar las obras por área para poder lograr datos anónimos coherentes y de calidad.

Laura mencionó que el proceso de limpieza se realizó en SAS y que el tratamiento de los datos se realizó en conjunto con el GIT PAD del DANE. Seguidamente, se explicó el cronograma y las fases del proyecto que aún no se han llevado a cabo, que incluyen mesas de trabajo y migración de los datos fuera de SQL.

Posteriormente, se explicaron los resultados preliminares del proyecto, demostrando que la migración a la plataforma Zeppelin en el Lago de Datos, reduce los tiempos de tratamiento de datos de manera significativa, se pasa de horas a segundos por lo cual la migración es válida. Además, se comprueba que con ambas metodologías se llega a los mismos resultados. Por último, Laura hizo la lectura de las preguntas orientadoras a los expertos y finalizó la exposición presentando al equipo DANE que participó en el proyecto.

## **8. Retroalimentación de los expertos respecto a los proyectos presentados (Expertos de la sala – 20 minutos)**

La retroalimentación inició con comentarios para el proyecto CAD, en donde Jorge mencionó que le parece muy interesante y comenta que le enorgullece que el proyecto haya avanzado y que esto pueda tener repercusiones positivas en temas de política pública. Comentó que los datos de los registros administrativos se pueden aprovechar en muchos contextos y que hay una gran necesidad de aplicar proyectos como el CAD en otras entidades.

Posteriormente, Mateo comentó sobre los retos del CAD, afirmando que le gustaría recibir las perspectivas de los expertos sobre cómo garantizar la seguridad de la información si las evaluaciones de procesos las hacen consultoras externas y terceros. Por último, Juan Sebastián mencionó que en el caso de que las entidades se muestren renuentes para compartir sus datos es necesario identificar métodos para generar más confianza para que haya más difusión de información, sobre todo cuando son externos los que hacen la revisión. A esto, Jorge comentó que una solución podría ser crear una Sala Especializada DANE a donde lleguen los datos, para disminuir la aversión de las entidades a compartir sus datos.

Además, Valerie comentó que los tres proyectos le parecieron muy interesantes y que continuar con la aplicación del CAD es muy importante.

Por otro lado, sobre el proyecto de Pobreza Multidimensional, León Darío reiteró sus sugerencias de hacer un análisis tanto cuantitativo como cualitativo y profundizar en la robustez de los modelos de Machine Learning. En este punto, Jorge compartió un link (<https://www.pnas.org/content/114/50/13108>), en donde se encuentra información relacionada con imágenes satelitales de Google Street View que se pueden aplicar al proyecto sobre todo en el caso de las ciudades.

Sobre el proyecto de la Migración del Censo de Edificaciones, León Darío mencionó que va a compartir un par de artículos que tienen un marco de referencia sobre aprendizaje de máquinas y que no están publicados, que incluyen diferentes modelos innovadores que pueden ser útiles para resolver dudas del proyecto.

## **9. Conclusiones y cierre (María José Prada - 5 minutos)**

La reunión finalizó con la intervención de María José Prada, enumerando los compromisos pendientes para la próxima sesión que incluyen la definición del plan de trabajo del apoyo del profesor León Parra en dos proyectos y los pasos a seguir para oficializar el Capstone entre la Universidad del Rosario y el DANE.

## Compromisos

- 1. Tarea:** enviar el acta de la sesión  
**Responsable:** DANE  
**Fecha entrega:** a más tardar el 6 de agosto
- 2. Tarea:** envío Doodle para agendar quinta reunión  
**Responsable:** DANE  
**Fecha entrega:** 4 de agosto
- 3. Tarea:** envío documentación proyectos a trabajar en la sala  
**Responsable:** DANE  
**Fecha entrega:** antes de la realización de la quinta reunión
- 4. Tarea:** revisar plan de trabajo de aportes pendientes de expertos  
**Responsable:** Expertos de la sala  
**Fecha de entrega:** Próxima reunión

---

Expertos CASEN

DANE

## Próxima reunión:

**Responsable de convocar:** secretaría técnica CASEN

**Fecha:** Por confirmar