

ACTA DE REUNIÓN - AYUDA DE MEMORIA

Ciudad: Bogotá
Lugar: Sesión virtual, plataforma
Teams

Tema: Revisión de insumos y de la ficha
de líneas de investigación – Sala de
modernización tecnológica

Hora: 8:00 a.m. a 9:56 a.m.

Fecha: 14/08/2020

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

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

Directora Técnica de Regulación, Planeación,
Estandarización y Normalización (DIRPEN)
Julieth Alejandra Solano Villa

Jefe de la Oficina de Sistemas
Andrés Holguín Coral

Coordinadora Grupo Interno de Trabajo (GIT)
Planificación y Articulación Estadística (DIRPEN)
Mónica Patricia Pinzón Torres

Asesor Dirección General
Juan Sebastián Ordoñez

Asesora Dirección General y DIRPEN
Elizabeth Moreno Barbosa

Profesional GIT Planificación y Articulación
Estadística (DIRPEN)
María del Pilar Pulido

Orden del día

1. Solicitud de autorización para grabar la reunión
 2. Llamado a lista y verificación del quórum:
 - I. León Darío Parra Bernal representante sala CASEN
 - II. Valérie Gauthier Umaña
 - III. Jorge Andrés Gallego Durán
 - IV. Andrés Holguín Coral
 - V. Juan Sebastián Ordoñez
 3. Confirmación del delegado a la Sala General del CASEN
 4. Confirmación de acceso al grupo de Teams
 5. Presentación de insumos:
 - I. Ficha resumen producción de información tema TIC
 - II. Boletín sectorial
 - III. Inventario operaciones estadísticas
 - IV. Inventario registros administrativos
 - V. Inventario demandas insatisfechas
 - VI. Manuales y referentes internacionales
 6. Presentación y discusión línea 1
 7. Presentación y discusión línea 2
 8. Presentación y discusión línea 3
 9. Cierre y conclusiones
-

Desarrollo

Objetivo

Presentar a los expertos de la sala los insumos disponibles con los que cuenta el DANE en materia estadística sobre el tema objeto de la sala y presentar la ficha de las líneas de investigación de interoperabilidad, para precisar el alcance.

1. Solicitud de autorización para grabar la reunión

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

2. Llamado a lista y verificación del quórum

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

3. Confirmación del delegado a la Sala General del CASEN

Se ratificó a León Darío Parra Bernal como representante de la sala General CASEN

4. Confirmación de acceso al grupo de Teams

Julieth Solano, Directora del área DIRPEN, fue la encargada de la presentación. Inicialmente, se informó a los expertos sobre la creación de una carpeta compartida en la plataforma Teams, en la cual se anexaron los insumos que aportarían información a la discusión de la Sala de modernización tecnológica. El enlace para ingresar a la carpeta fue enviado a través del chat de la reunión: https://teams.microsoft.com/#/files/General?threadId=19%3A7726f169518a447b8155cb46a8fbc5ac%40thead.tacv2&ctx=channel&context=General&rootfolder=%252Fsites%252FDANE_SalaEspecializada_0365%252FShared%2520Documents%252FGeneral

5. Presentación de insumos

Se revisó la carpeta en la cual están disponibles los siguientes documentos de interés:

- I. Ficha resumen producción de información tema TIC
- II. Boletín sectorial del TIC's - Julio 2020
- III. Inventario de operaciones estadísticas
- IV. Inventario de registros administrativos
- V. Inventario de demandas insatisfechas
- VI. Manuales y referentes internacionales:

United nations global platform: Data for the world brochure

Se realiza anotación de participación en el grupo social media por parte del DANE – especialmente Twitter-, así como en el grupo de datos móviles (2017) e imágenes satelitales (Dirección Geoestadística) y Buró en administración del grupo.

6. Presentación y discusión línea 1: Big Data e Interoperabilidad

Julieth Solano, Directora del área DIRPEN, realizó la presentación de la línea indicando los comentarios realizados por los expertos, en la que León Darío Parra señaló estar de acuerdo en la línea, pero comentó que se debía hacer más precisión en la descripción, esto es, si será solo Big Data o Big Data y Data

Analytics. Asimismo, respecto al tema de interoperabilidad, acotó si sería solo de datos o sobre sistemas de información como un todo.

Por su parte, Valérie Gauthier Umaña manifestó que el término 'Big Data' hacía referencia a mucho y nada, y que no se entiende si es para analizarla o qué rol cumple. Por lo anterior, propuso una cuarta línea dirigida a la 'Analítica de Datos', sugiriendo que el nombre de esta podría ser 'Calidad de datos e interoperabilidad'.

Siguiendo con los criterios de planteamiento de la línea de investigación, la propuesta que se realizó es la de asesorar y evaluar el desarrollo del Sistema Estadístico Nacional (SEN) en cuanto a las plataformas y herramientas tecnológicas, así como las de analítica, necesarias para la generación de información estadística, a partir de la integración de estadísticas tradicionales como censos, registros administrativos y fuentes no tradicionales de tipo Big Data, integración con la información del DANE (piscina de datos), así como ayuda conceptual para los modelos de datos que se deben crear. Frente a esto, se señaló como soporte, el Decreto 2404 artículos 2.2.3.1.1 y 2.2.3.5.4, así como el Plan Nacional de Desarrollo 2018-2022 y el documento Conpes 3992.

Por otro lado, Jorge Andrés Gallego mencionó que tenía dudas sobre si las entidades que cuenten con bases de datos y quieran cruzarlas con información del DANE lo puedan hacer, y si el DANE hace cruces de base de datos. A lo anterior, Julieth Solano manifestó que esta actividad se realizaba de manera continua, y que es un desafío de las oficinas estadísticas por el tema de la reserva. Muchas veces este desafío se resuelve con infraestructuras tecnológicas de datos seguros suministrando información con datos anonimizados y agregados, para compilar, por ejemplo, registro base de empresas, de ministerios, etc. Por otra parte, manifestó que el DANE hace cruces y dispone de los datos tratados, y señala que existen mecanismos como las consultas especializadas en las que hay controles a la información y se aprovecha con un mayor grado de desagregación. A lo cual Jorge replicó que lo preguntaba por temas de privacidad para que el tema no sea tan asimétrico y se establezcan alternativas al respecto.

Andrés Holguín, Director de la Oficina de Sistemas, mencionó que el reto tecnológico es poder construir herramientas que permitan hacer los cruces sin tener acceso al microdato anonimizado, señaló que en otros países como Corea y Francia se pueden cruzar las variables sin acceso a estos datos y obtener los resultados buscados. También, que era importante tener en cuenta que no se trata únicamente de ofrecer seguridad restrictiva sino seguridad positiva que permita, con un costo razonable, acceder a herramientas para realizar cruces.

Luego de las observaciones, Julieth Solano continuó con la descripción de la línea anunciando que esta se enfoca en el aprovechamiento de la información proveniente de fuentes tradicionales y fuentes alternativas. León Darío Parra Bernal mencionó que, en relación con lo anterior, podría decirse que la línea estaría enfocada a mapear y analizar fuentes alternativas de información y su interoperabilidad con las fuentes o sistemas tradicionales.

A las anteriores observaciones, Valérie Gauthier Umaña indicó que la descripción se ajusta más al tema de analítica para la cuarta línea propuesta, y la temática a abordar sería más enfocada hacia buscar y

asegurar que los datos recopilados cumplan con estándares de calidad de datos e interoperabilidad, para facilitar su acceso y uso por parte de terceros.

Julieth Solano indicó que cuando se habla del aprovechamiento de la información se piensa en una fuente renuente como por ejemplo, los registros de telefonía móvil. Estas fuentes tienen buena cobertura y frecuencia de generación de información, pero hay desafíos de acceso y rigor de los indicadores a producirse. Se refiere a cómo se logra el aprovechamiento superando los desafíos, no sólo tecnológico sino también acceso, seguridad y privacidad. Es importante tener un concepto del CASEN junto con el Comité de Administración de Datos (CAD), con el fin de plantear un esquema de trabajo o proyecto para tener el acceso a la información. Movistar hace aprovechamiento de datos, pero envía al DANE los datos customizados, de caja negra, entonces es un desafío poder acceder a los datos.

Para este punto, Andrés Holguín indicó que en muchos países se está manejando estadísticas oficiales a partir de fuentes alternativas y que categoría de 'oficial' tiene un tono gris porque incorporan redes sociales, información abierta o compran información de acuerdo con la reglamentación estipulada. Se menciona también el tema de la oportunidad, en el que existen países que entregan información con rangos de confiabilidad a partir de fuentes alternativas, vinculando tendencias para dar mayor oportunidad a la información.

Continuando con la agenda de la sesión, Julieth Solano indicó que era necesario identificar mecanismos de interoperabilidad y recomendaciones de cómo superarlo, la entidad no ha tenido el apoyo esperado por parte de otras entidades en el proceso, ya que es difícil convencer a los actores para poder acceder a los datos, por esto, se requiere un concepto del CASEN para empezar a articular a las entidades y actores correspondientes.

A modo de complemento, León Darío Parra Bernal mencionó que antes del cómo implementarla es necesario revisar su factibilidad y el tema de calidad de la información, ya que va en sintonía con las posibles metodologías que existen para evaluar su calidad. Finalmente, en las fuentes alternativas, indicó que consideraba que el propósito es mapear y analizar la pertinencia de las diferentes fuentes alternativas más que incorporarlas, lo cual sería el siguiente paso.

Frente a este comentario, Julieth Solano mencionó que dependiendo de los temas hay avances, como lo es en el tema de pobreza con información de imágenes satelitales y nocturnas con información no tradicional.

Frente a las otras salas se mencionó que la sala de modernización tecnológica es transversal a todas las temáticas y se revisarán aspectos tecnológicos pensando en economías de escala.

En relación con los invitados a las diferentes sesiones de la sala, se mencionaron entidades como MinTic, MinCiencia, Consejería para Asuntos Económicos y de Transformación Digital. La propuesta de Valérie Gauthier fue invitar al DNP y a entidades como Catastro, Salud Data, las cuales cuentan con investigadores expertos en el área de analítica como es el caso de CAOBA y Juan Fernando Pérez Bernal, de la Universidad del Rosario.

Respecto al tema de interoperabilidad, Andrés Holguín mencionó que este se pensaba en el marco de 'interoperabilidad del Estado' y, desde el punto de vista regulatorio, es una camisa de fuerza que se tiene para trabajar la idea de implementar la lógica del ciudadano digital con herramientas tecnológicas de MinTic. Hay un movimiento de información dentro de las entidades del Estado y la desventaja que surge, es que mover registros administrativos (RR.AA) es complejo en esa lógica.

Por su parte, Julieth Solano mencionó que en EE.UU. un grupo de expertos se reunió para pensar en el futuro de las estadísticas en el sistema estadístico tomando en cuenta que este es descentralizado y realizaron una propuesta para producción estadística; sería importante analizarlo con el fin de adoptar buenas prácticas.

Frente al plan de trabajo, se indicó que es necesario establecer fuentes, plataformas y analítica sobre el aprovechamiento de información, tomando como referencia los comentarios recibidos por León Darío y Valérie Gauthier.

Igualmente, León Darío hizo referencia a los pasos, y posibles limitantes, en el tema de operabilidad. Indicó que es importante revisar el estado del mapeo y la línea de base; y luego de esto, pensar en el planteamiento de metodologías para las entidades si no se cuenta con la fotografía. Desde su perspectiva, como grupo no se puede planear sin tener en cuenta la realidad por parte de las entidades ya que no se ajustaría de manera acorte. Es conveniente revisar la línea base para determinar si se cuenta con datos e información que permita realizar una propuesta y establecer qué es lo que le conviene al país.

Valerié Gauthier mencionó que era importante separar las dos salas de Big Data e Interoperabilidad y que no era fácil establecer mecanismos, por esto, la conveniencia de incluir entidades para analizar los sistemas que usan y determinar necesidades. Lo anterior ayudaría a articular y trabajar con los diferentes actores. No se debe olvidar el tema de analítica y aprovechamiento, teniendo el tema de interoperabilidad y dato, seguiría el análisis de metodologías para dar recomendaciones de acuerdo a lo que hacen las entidades y luego seguir con ciberseguridad.

Juan Sebastián Ordoñez indicó que, de acuerdo con lo que Valérie Gauthier mencionó, se debe abrir la línea de analítica, pero no tuvo claridad del por qué se tiene que vincular el tema de seguridad, por lo que le surgió la pregunta sobre a qué hace referencia puntualmente con la propuesta realizada. Valérie Gauthier, mencionó, por ejemplo, que las entidades tienen información, pero no la aprovechan y esto se asocia con la analítica. 'Aprovechamiento' sería una línea y otra 'Cómo aprovechar la información'. Ciberseguridad es transversal a todas las líneas, así como la innovación.

Frente a la propuesta de trabajo, Julieth Solano sugirió que en la primera reunión se presentará el problema desde la óptica del DANE en relación con la interoperabilidad, así como el estado del arte de acuerdo con los referentes en torno al aprovechamiento e interoperabilidad y de esta manera, tener conclusiones y recomendaciones. Se planteó que la primera reunión fuera cerrada, la segunda con invitados y una última de cierre del tema.

En relación con este punto, Valérié Gauthier mencionó que se debía considerar la presencia de expertos en la reunión sobre el estado del arte, en la segunda expertos en el tema, y en la tercera, seguir con dos expertos.

Julieth Solano mencionó las observaciones generales de los expertos como el caso de León Darío Parra, quien manifestó que en general la línea de 'Big Data e interoperabilidad' a lo que podría enfocarse en una primera instancia es a mapear y analizar las buenas prácticas que se han manejado en otros países o ecosistemas de información y datos, y de estos recomendar qué fuentes alternativas de información y protocolos de interoperabilidad se podrían implementar para el Sistema de Estadística Nacional, con alcances y limitaciones; también hizo referencia a la opinión de Valérie Gauthier Umaña, quien consideró que se debería separar la parte de analítica y visibilidad de los datos, en una cuarta línea. Para esta cuarta línea se invitaría a las demás mesas para conocer los tipos de datos etc., pues menciona que, la ve muy técnica y es mejor que no sea así.

7. Presentación y discusión línea 2: Seguridad de la información

Julieth Solano, presentó la segunda línea a trabajar e hizo referencia al comentario de León Darío Parra Bernal, quien sugirió agregar el tema de protección de datos para que el nombre quedara: Seguridad de la información y protección de datos.

Como propósito se estableció asesorar y evaluar el desarrollo del Sistema Estadístico Nacional (SEN) en un esquema para el intercambio de datos, entre los miembros del SEN, que garantice la seguridad de la información, así como su uso de manera ética y adecuada.

Frente a la descripción de la línea de investigación, la propuesta es establecer recomendaciones y lineamientos de seguridad de la información.

La propuesta tiene como finalidad implementar los marcos definidos por MinTic entre los miembros del SEN, el intercambio de información; así como recomendar modelos de seguridad y control de la información del DANE y explorar y recomendar modelos de desidentificación/anonimización de la información. De acuerdo con lo anterior, León Darío Parra Bernal recomendó mapear y analizar las nuevas tecnologías y herramientas que se manejan en ciberseguridad para la trazabilidad de la información.

Los resultados esperados se enfocan en guías metodológicas y recomendaciones tecnológicas; también se señala la Ley de la reserva, lo cual se convierte en desafío frente a los requerimientos de las entidades para integrar datos y luego obtener información de estos. Esta línea es transversal a todas las salas.

Frente al tema de los invitados, se sugirió un delegado de MinTic y un delegado de Minciencias.

Asimismo, se propuso un esquema de trabajo con tres sesiones para cada línea de investigación: la primera para presentar el problema, la segunda para presentar el estado del arte y, una última, para conclusiones y recomendaciones.

Andrés Holguín, jefe del área de Sistemas, mencionó que existen muchos retos y es necesario revisar cuáles se van a afrontar. Entre estos destaca:

- La lógica de la seguridad del Estado es un modelo restrictivo y citó como ejemplos: las restricciones de navegación, el ingreso de portátiles al DANE, la implementación de un sistema de correo en la nube resultaba un reto al cruzar la frontera de la reserva estadística. De este modo, así como pasa en el DANE pasa en otras entidades del Estado.
- El tema cultural es otro factor que afecta, porque el hecho de prohibir todo no es garantía que exista seguridad.
- Entender cómo aprovechando las nuevas tecnologías puede encontrarse nuevas formas de hacer las cosas.
- Al igual que como pasa con el manejo de los datos de personas con la reserva estadística, la Ley 1581 excluye bases de datos del DANE, las cuales no requieren solicitar permiso o autorización a los propietarios de los datos. Frente a lo anterior, se presenta el reto de entender la ley dentro del sistema estadístico, teniendo en cuenta cómo se armoniza la metodología y herramientas a utilizar. Por ejemplo, para el diseño de políticas de datos del DANE.

8. Presentación y discusión línea 3: Innovación en métodos de recolección de información

Julieth Solano, realizó la presentación e hizo referencia al comentario realizado por León Darío Parra Bernal, quien mencionó que podría ser más genérica la línea de investigación denominándola 'Innovación en Sistemas de Captura y Recolección de la Información'.

Los criterios para definir la línea de investigación se enfocan en asesorar y evaluar el desarrollo del SEN en los métodos de medición para las estadísticas oficiales, así como los instrumentos, medios y fuentes de recolección de información que contribuyan a la incorporación de innovaciones.

También se mencionó que, en Canadá, para obtener la información transaccional, la oficina estadística tiene acceso directo a esta; por ejemplo, para el índice de precios. En Colombia es necesario realizar ajustes el proceso estadístico actual, pues, es costoso tanto su diseño como su puesta en marcha.

Los resultados esperados con la línea de investigación propuesta se enfocan en hojas de ruta para implementar nuevas soluciones a los métodos de recolección de información.

Como invitados a las próximas sesiones, León Darío Parra planteó que fueran delegados de MINTIC y Minciencias y, ocasionalmente, actores del sector privado que ya hayan implementado dichas herramientas. Valérie Gauthier Umaña propuso a la Consejería para Asuntos Económicos y de Transformación Digital, MinTIC, DNP, Superintendencia de Industria y Comercio, CEA de IoT, Jeimy Cano y Juan Fernando Pérez.

Para el plan de trabajo se adoptaría la misma propuesta de la línea anterior.

Finalmente, se compilaron los siguientes comentarios generales:

León Darío Parra: En relación con el componente de innovación, en la primera etapa de mapeo, la línea buscaría analizar algunos casos de estudio o buenas prácticas de actores que ya hayan implementado herramientas o métodos alternativos de captura.

Valérie Gauthier Umaña no tenía certeza sobre la pertinencia de invitar a empresas como Carvajal, que está trabajando en máquinas de voto electrónico y que puedan tener opiniones y estén interesados en alianzas.

Jorge Andrés Gallego mencionó que era importante reconocer el desafío que se presenta en los sesgos estadísticos cuando se utilizan nuevas formas de recolectar información y el costo asociado a estos; estos desafíos se evidenciaron por ejemplo, en las encuestas web que se empezaron a utilizar en el año 2000 y en el uso de los métodos tradicionales, como es el caso de los mecanismos presenciales. También se debe prestar atención a soluciones estadísticas a los sesgos cuando se usan nuevas fuentes y cómo se corrige.

Andrés Holguín mencionó que el problema identificado en cada aplicación para capturar datos es que se diseña para cada operación estadística (OO.EE) en particular. Al día de hoy, existen 125 sistemas de información distintos en el DANE, formularios de captura, monitoreo y control, monitoreo crítica, análisis de datos, producción de cuadros de salida; esto hace que la adopción tecnológica del DANE sea muy lenta. Todo este proceso inició en 1993 con el Censo y terminó en migración de captura en 2018.

Los datos de las OO.EE están en distintas "canecas" lo que hace que la correlación de datos sea difícil y catalogue esta forma de operación como un tema complejo. Los métodos de captura se deben construir y proponerse una caja de herramientas que se pueda usar en las OO.EE. Por ejemplo, captura web para no diseñar un nuevo métodos de captura, innovación para la inclusión de nuevas tecnologías, de internet de las cosas, de RR.AA, el traslado a muchas tiendas de grandes superficies (Éxito, Jumbo, Falabella) para conseguir información, siendo las tiendas de barrio, un espectro más complicado.

Juan Sebastián Ordoñez mencionó que se debe trabajar en complementar los mecanismos porque desde que se haga un mapeo se establecen las encuestas, se disminuyen costos, y es en esto, donde se debe tener en cuenta los hallazgos de calidad y qué tan vinculables son los datos. Se debe propender por generar estandarización de ID únicos o tablas de referencia, con el fin de tener seguridad y bases anonimizadas, con estándares de calidad, que permitan generar tablas para vincular fuentes de información, sistemas seguros, robustos, y que garanticen que no se viole ninguna ley de seguridad.

León Darío Parra mencionó que se debe dimensionar y priorizar el plan de trabajo porque todo es global, hay que mirar el alcance que se quiere lograr para el semestre, porque al abarcar tantas temáticas, puede que no se logra nada.

Luego de estos comentarios, Julieth Solano propuso realizar un cierre con comentarios, observaciones y solicitudes, a través de lo cual se estableció:

- Priorización en cómo se percibe la primera línea.
- Big data: fuentes y aprovechamiento analítica y procedimiento
- Interoperabilidad: estándares y plataformas MINTIC, cómo se implementan en un sistema estadístico
- Recolección de información un tema de seguridad con implicación transversal

León Darío Parra indicó que primero se debían abordar los insumos y mapear, para luego, entrar al tema de seguridad.

Valérie Gauthier indicó que era importante la utilidad de los datos y su transformación. Al comienzo se definen procesos de interoperabilidad, luego innovación y finalmente ciberseguridad al ser transversal.

9. Cierre y conclusiones

Julieth Solano concluyó que, de acuerdo con lo conversado, en la próxima sesión de 'Big Data e interoperabilidad', se analizarán los requerimientos o solicitudes para el planteamiento del problema. Teniendo en cuenta que para la primera línea se plantearon 4 o 5 reuniones con una periodicidad cercana, se indagó por la posibilidad de hacer la siguiente reunión en 15 días y manejar un tiempo de tres meses, esta periodicidad se iría evaluando de acuerdo con la exploración de las otras líneas el próximo año.

La manera de abordar la primera línea debe seguir los siguientes pasos:

- Planteamiento del problema: desafíos
- Referentes y estado del arte
- Propuestas y soluciones

León Darío Parra propuso realizar las reuniones los viernes en la tarde o jueves. Y pregunto si estas serían mensuales.

Valérie Ghautier indicó que no podía los viernes, por lo cual propuso que se realizarán los lunes en la tarde y que se brindarán fechas alternativas.

León Darío Parra manifestó que se debía proponer un calendario conjunto.

Julieth Solano indagó sobre observaciones generales y agradeció por la labor realizada por parte de los expertos.

Compromisos

- 1. Tarea:** Enviar acta
Responsable: DANE
Fecha entrega: 18 al 21 de agosto
- 2. Tarea:** Proponer un calendario de próximas reuniones de acuerdo a disponibilidad de expertos
Responsable: DANE
Fecha entrega: 18 al 21 de agosto

Próxima reunión:

Responsable de convocar: Secretaría Técnica
CASEN

Fecha: Finales del mes de agosto (por definir)

Anexos:

- Lista de asistencia

* Si el acta de reunión no requiere aprobación omite el siguiente campo.

Expertos CASEN

DANE