

ACTA DE REUNIÓN N°1 AYUDA DE MEMORIA

Ciudad: Bogotá

Lugar: Sesión virtual, plataforma Teams

Tema: Seguridad de datos gestionados en la nube- reserva estadística y protección de datos confidenciales

Hora: 8:00 a.m. a 10:00 a.m.

Fecha: 13/05/2022

Dependencia responsable: Secretaría Técnica CASEN (DIRPEN)

Participantes

Miembros de la Sala Especializada de Modernización Tecnológica

León Darío Parra Bernal
Valérie Gauthier Umaña

Departamento Administrativo Nacional de Estadística (DANE)

Asesora DIRPEN y Dirección General
Elizabeth Moreno Barbosa

Oficina de Sistemas
Luis Martín Barrera Pino

Oficina de sistemas
Gilberto Villalba Gamboa

Asesor de la Dirección General
Juan Sebastián Ordóñez

Dirección de Regulación, Planeación,
Estandarización y Normalización DIRPEN
Julieth Alejandra Solano

Coordinadora - Planificación y Articulación
Estadística
Mónica Patricia Pinzón

Profesional GIT Planificación y Articulación
Estadística (DIRPEN)
Ruth Constanza Triana

Profesional GIT Planificación y Articulación
Estadística (DIRPEN)
Andrés Felipe Villarraga

Profesional GIT Planificación y Articulación
Estadística (DIRPEN)
Sofía Sánchez Granados

DIRPEN – Gobierno de Datos
Mateo Cardona

DIRPEN- Dirección técnica
Anderson Leal Vélez

Coordinador- Prospectiva y Análisis de Datos
Alexander González

Prospectiva y Análisis de Datos, Equipo ODS
Grace Andrea Torres

Geoestadística
Mónica Lorena Ortiz

Orden del día

1. Solicitar autorización para grabar la reunión
2. Verificación del quórum
 - León Darío Parra
 - Valérie Gauthier Umaña
 - Juan Sebastián Ordóñez
 - Mateo Cardona
 - Julieth Alejandra Solano Villa
 - Elizabeth Moreno Barbosa
 - Mónica Patricia Pinzón
 - Luis Martín Barrera
 - Marly Esther de Moya
 - Alexander González
3. Propuesta de valor 2022- Memorias sobre la experiencia para el fortalecimiento del CASEN (Sofía Sánchez- 5 minutos).
4. Presentación: Contextualización de la situación del DANE en materia de seguridad de datos gestionados en la nube (Gilberto Villalba y Anderson Leal- 15 minutos).
5. Realimentación de los expertos sobre las preguntas enviadas y la presentación anterior "Contextualización de la situación del DANE en materia de seguridad de datos gestionados en la nube" (Expertos de la sala: Valérie Gauthier y León Parra- 10 minutos cada uno).
6. Presentación avances proyectos prospectiva (Alexander González y Grace Torres- 20 minutos).
7. Realimentación de los expertos sobre los proyectos presentados en el punto anterior (Expertos de la sala: Valérie Gauthier y León Parra- 10 minutos cada uno).
8. Conclusiones y cierre (Sofía Sánchez - 5 minutos).

Desarrollo

Objetivo

Presentar el contexto del DANE en materia de seguridad de datos gestionados en la nube y protección de datos confidenciales, a fin de recibir realimentación por parte de los expertos y fortalecer estos procesos.

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. Propuesta de valor 2022- Memorias sobre la experiencia para el fortalecimiento del CASEN (Sofía Sánchez- 5 minutos)

Sofía Sánchez inició la sesión abordando el contexto de la propuesta de valor. Se centró en el documento de recomendaciones 2021, donde se abordaron las dos líneas de investigación desarrolladas: 1) BigData y 2) Seguridad de la información.

En la primera, se abordaron dos proyectos de investigación: desarrollo y distribución de aplicaciones web para la recolección de datos y Comité de Administración de Datos (CAD). Con respecto a la segunda línea, se habló sobre el desarrollo de presentaciones sobre: la política de gobierno de registros administrativos y fuentes alternativas; el modelo de gestión de datos – Requerimientos de seguridad; la política de seguridad de la información DANE; la anonimización del censo de edificaciones; y, sobre criptografía, sus implicaciones, perspectivas y conceptos básicos.

En lo concerniente al trabajo bilateral adelantado desde el mes de febrero de 2022, se habló sobre la puesta en marcha del convenio con la Universidad EAN y la Universidad de los Andes, los avances en el proyecto de índice de noticias, la posibilidad de participar en eventos de investigación y la definición de líneas de investigación.

Por otro lado, señaló que a pesar de que se surtieron todas las fases del proceso para cubrir la vacante de experto en la sala, la selección quedó pendiente para el nuevo periodo del CASEN. Mónica Patricia Pinzón puntualizó que una vez el DANE realizó la invitación pública, dos aspirantes manifestaron su interés mediante un ensayo y presentaron la prueba de competencias, sin embargo, uno de ellos desistió de continuar en ese punto, mientras que el otro no obtuvo una calificación superior al 70 % en la prueba de integridad, requisito estipulado en la resolución 0446 de 2020 para hacer parte del CASEN.

Para finalizar, recordó a los expertos la importancia de desarrollar las memorias sobre la experiencia en el CASEN a modo de realimentación sobre los avances y buenas prácticas adelantadas en la sala en estos dos años para ver en qué se puede mejorar.

4. Presentación: Contextualización de la situación del DANE en materia de seguridad de datos gestionados en la nube (Gilberto Villalba y Anderson Leal– 15 minutos).

Gilberto Villalba explicó que la seguridad depende del proveedor y para el caso del DANE, tiene una defensa cibernética en capas, la cual requiere: conformidad a políticas para dar acceso, que haya seguridad en los datos, gestión de identidades y acceso, visibilidad y control en la cual el proveedor utiliza técnicas de machine learning para identificar las alertas más importantes, y seguridad y resiliencia de las aplicaciones. Aclaró que esto se da en un entorno híbrido.

Para hablar de la estrategia de seguridad del proveedor actual de nube del DANE, el expositor señaló los tres enfoques con los que cuenta: la estrategia de seguridad del proveedor de nube, que consiste en asegurar la plataforma de nube (zero trust); el de asegurar las identidades, aplicaciones y datos en la nube; y la protección y monitoreo entre nubes y on-premises. Además, dijo que la nube que se tiene es de segunda generación, es decir, hace la virtualización de redes aisladas, lo cual realiza contención de amenazas y reducción de riesgos.

El tercer punto desarrollado por Gilberto Villalba, fue el relacionado con la arquitectura segura-zero trust, que es un enfoque de seguridad que tiene el proveedor, que el DANE también quiere tener en cuenta en la implementación. Teniendo en cuenta el enfoque de seguridad expuesto, abordó la arquitectura conceptual que se está implementando con algunos retos en la entidad y que se están trabajando de la mano con el proveedor. Aclaró que a pesar de que es una nube, no se puede acceder por internet, ya que es una extensión de la on-premise por medio de una VPN. Resaltó el papel del firewall y los permisos, ya que, a pesar de ingresar en la red del DANE, cada usuario tiene acceso a la información de acuerdo a las políticas.

Anderson Leal abordó el tema de la seguridad de los datos en la nube. Se tiene la directriz de no subir algún microdato con riesgo de seguridad directamente en el entorno, debido a que esta información se almacenaría en el recurso de almacenamiento del proveedor que no se encuentra en Colombia (los servidores más cercanos se encuentran en EE. UU. y Brasil). Por lo tanto, se ha tomado la decisión desde el área directiva de usar la información, pero siempre con ámbito de anonimización y seguridad de los datos, el cual pueda garantizar la seguridad de los mismos.

Después, explicó la estructura del triángulo de acción de los proyectos a desarrollar en nube: usuario- grupo de gobierno de datos- grupo de gestión de datos. La cual funciona así:

- I. el usuario realiza una solicitud de automatización de algún procedimiento utilizando las facultades de la nube en términos de cómputo, algunos algoritmos especiales y también el tratamiento de algún tipo de información.
- II. para ello, el usuario se comunica con el grupo de gestión de datos, quienes atienden la solicitud junto con el grupo de gobierno de datos (seguridad de datos en DIRPEN) para establecer las políticas de seguridad: reglas de anonimización y/o encriptamiento de datos que se deban cumplir según corresponda a cada una de las entidades o de los objetos en cuestión.
- III. el grupo de gobierno de datos proporciona al grupo de gestión de datos las reglas de anonimización, quienes aplican en lo local.
- IV. así se obtienen como resultado las bases anonimizadas que cumplen con el aval de seguridad proporcionado.

- V. con ello se procede a realizar el cargue y tratamiento de la información directamente en nube de nuevo contactando al usuario para poder realizar o llevar a cabo las reglas de procedimiento.

5. Realimentación de los expertos sobre las preguntas enviadas y la presentación anterior “Contextualización de la situación del DANE en materia de seguridad de datos gestionados en la nube” (Expertos de la sala: Valérie Gauthier y León Parra- 10 minutos cada uno)

Valérie Gauthier inició su intervención señalando que no maneja el tema de normatividad. Sin embargo, manifestó su intención de contactarse con expertos en el tema tanto del aspecto legal como técnico. Además, planteó dos interrogantes:

- 1) ¿Los aspectos de seguridad dependen del tipo de datos? Es decir, ¿hay alguna información que tiene más protección que otra?

Anderson Leal respondió que en el caso de infraestructura todos se están yendo con la máxima seguridad, mientras que, en el caso de los datos, señaló que los datos que se han trabajado se van anonimizados. Precisó que como está en implementación, han surgido varias dudas frente a qué pasa si se sube un dato sin anonimizar, aunque la infraestructura tenga seguridad, ya que el proveedor tiene el poder, más aún si no se cuenta con un datacenter en el país, que es el que garantiza una seguridad física.

- 2) ¿Han avanzado en temas de seguridad diferencial?

Anderson señaló que la mayoría de las técnicas de anonimización para los datos utilizando la nube obedecen a aquellos que tengan la menor perturbación sobre los mismos, ya que los usuarios en la gran mayoría de los proyectos necesitan que los datos no se encuentren perturbados. Por lo cual, para ese caso se usan técnicas de enmascaramiento, cambios de rótulos, aleatoriedad en la posición de los registros, algún que otro tratamiento con los identificadores y los cuasi identificadores, según lo expresen en el grupo de gobierno y seguridad de la información.

Por último, aclaró que, para poder generar un ruido a la información, es necesario que al proporcionárselo se tenga la llave para quitárselo una vez se encuentren los datos cargados en la nube para poder realizar el procesamiento y esta información sea eliminada ya sea en memoria (en término de procesamiento) cuando se utiliza la herramienta en el proveedor en nube.

Por otro lado, el experto León Parra expresó no ser experto en el área de protección de datos, pero manifestó conocer diferentes colegas que se desempeñan en el tema de manejo de grandes volúmenes de información e interoperabilidad, por lo cual se podrían generar espacios de interlocución enfocados en normatividad y protocolos mediante la modalidad de seminario web, por ejemplo, con la invitación a Andrea Martínez Devia.

En cuanto a la normativa, destacó la ley 2157 de 2021 que modificó la ley 1266 de 2008 en materia de habeas data. Resaltó que Colombia se encuentra muy incipiente en el desarrollo de normatividad específica frente a protección de datos en la nube.

Compartió con la sala, un manual del BID y otro del banco asiático de desarrollo, los cuales se toman como referentes de buenas prácticas y de las oportunidades que hay en toda el área de la computación en la nube. Además, se comprometió a hacer llegar por correo a los asistentes de la sala algunos casos exitosos en Asia, en donde han hecho un mayor avance en términos de protección de datos y la gobernanza de los datos.

Luego, enfatizó en la importancia de la gobernanza de los datos. Esto, mediante la explicación de un diagrama extraído de un paper que habla sobre la taxonomía de la gobernanza de los datos. A partir de esto, señaló que hay tres puentes: 1) gobernanza en términos de TI, la cual se relaciona con toda la arquitectura que se tiene para proteger los datos y sobre todo para generar políticas en cuanto a la protección de datos, 2) la gobernanza en cuanto al manejo de la información y establece que es allí donde surge la pregunta frente a si todo debe estar anonimizado en la nube, 3) las políticas que hay detrás, es decir, el manual de buenas prácticas, definir cuál forma y la arquitectura desde el punto de vista más corporativo que van a manejar los protocolos.

Señaló que la unión entre políticas y procesos es esencial y que hay cuatro niveles que desde el punto de vista del DANE se deberían definir: 1) los estándares bajo los cuales se va a trabajar a nivel técnico, 2) los procesos para liberar y manejar estas metodologías, 3) las políticas y principios a desarrollar para el uso tradicional de la gobernanza de los datos, 4) arquitectura diferencial, es decir, involucrar a los diferentes actores para saber el nivel de accesibilidad que tiene en cada dato.

Gilberto Villalba, acoge la propuesta del profesor León de realizar un seminario web. Después, solicitó a los expertos abordar otra de las preguntas planteadas previamente sobre el conocimiento frente a grupos de investigación en el tema.

El profesor León se refirió al grupo ONTARE de la Universidad Ean que trabaja la línea de habeas data y manejo de datos, el cual que se podría articular con el DANE en el marco del convenio bilateral. Mientras que Valérie habló del Grupo de Estudios en Internet, Comercio Electrónico, Telecomunicaciones e Informática - GECTIC de la Universidad de los Andes que aborda la línea de protección de datos, legal tech, propiedad intelectual y ciberseguridad desde el punto de vista legal y la Asociación Colombiana de Datos y Privacidad- ADAPRI, que aborda el tema desde la mirada jurídica.

Luego, Gilberto preguntó a los expertos si tienen conocimiento sobre ejemplos o casos sobre riesgos que el DANE deba tener en cuenta. León habló sobre el Global Entrepreneurship Monitor- GEM, ya que hacia el año 2016, desarrollando un proyecto de BigData, en la medida que empezó a incrementarse el número de actores que tenían acceso a la información, el riesgo aumentó. El experto señaló que en tres países se filtró información y por algunos proveedores, esa información fue a dar a situaciones comerciales. Los comerciales usaron los datos para enviar información a las personas, siendo que la garantía que se le debe dar a los usuarios es que su información nunca

puede ser compartida. A partir de lo que pasó, se implementó todo lo desarrollado frente a la gobernanza de los datos y la anonimización de los datos.

Por último, Gilberto invitó a los expertos a compartir alguna experiencia que conocieran respecto a casos sobre cómo se han manejado ataques cibernéticos o secuestro de información en nube y on-premise. La experta Valérie comentó el conocido caso de WannaCry, pero manifestó que, al no trabajar en esa área, no conoce de primera mano. Sin embargo, planteó la posibilidad de establecer contacto con algún colega que haya estado trabajando con la inteligencia del gobierno (Wilson Prieto) para ver cómo lo manejan y qué buenas prácticas tienen.

3) Presentación avances proyectos prospectiva (Alexander González y Grace Torres- 20 minutos)

Alexander González realizó una introducción a su equipo de trabajo. Luego, presentó el objetivo y los avances de los seis proyectos priorizados para el año 2022.

Inició con los proyectos liderados por Grace Andrea Torres:

1. **Proyecto Índice de Noticias.** Cuyo objetivo consiste en probar un conjunto de modelos que permitan mejorar el cálculo de un estimador temprano para conocer el comportamiento aproximado del sector agrícola según las Divisiones CIU Rev. 4 A.C.-61 agrup. Como principales avances se encuentran: el ajuste a la metodología para responder al rediseño de la GEIH y la emisión de los datos del estimador temprano del primer cuatrimestre de 2022 para los sectores agrícola, turismo y construcción.

2. **Proyecto Medición proxy del indicador del ODS 10.3.1/16.b.1 utilizando fuentes no convencionales (redes sociales).** Objetivo: obtener mediciones complementarias de los indicadores del ODS 16, asociados a la percepción de la discriminación, a través de Facebook. Dentro de los avances están: Consolidación de la metodología para obtener mediciones complementarias de los indicadores del ODS 16. Dashboard de visualización de resultados. Publicación en la página web del DANE en la sección de estadísticas experimentales el próximo 17 de mayo.

3. **Proyecto Análisis y priorización de necesidades para la producción estadística.** Tiene por objetivo mejorar la priorización de las necesidades de información para promover una mejor gestión de recursos, una mejor rendición de cuentas y el involucramiento más incluyente de todos los grupos de interés. Avances: Dashboard de visualización y priorización de las necesidades de información estadística para 2020 y 2021, la creación de un procedimiento para la detección y análisis de necesidades de información estadística (versión para aprobación) y el instrumento de captura de las necesidades de información actualizado.

Continuó con los proyectos liderados por Raúl Gómez

4. **Proyecto Desarrollo y distribución de aplicaciones Web, en redes sociales, cómo método de recolección de información para estadísticas oficiales.** Objetivo: Probar la efectividad del método de recolección de información estadística oficial por medio de aplicaciones web y desarrollar una aplicación web y distribuirla por medio de redes sociales para recolectar información estadística para el cálculo del indicador ODS 16.b.1. Principales avances: App: <https://encuestapropuestaods.bubbleapps.io/version-test/facebook> y Cooperación GPSDD: Financiación Consultores, evaluación de hojas de vida y entrevistas.
5. **Proyecto Visualización de Indicador de Calidad de Bases de Datos.** Su objetivo se centra en presentar a los responsables de las OO.EE y a la sociedad en general, el nivel de calidad de las bases de datos, de las operaciones estadísticas evaluadas por el DANE con base en la norma técnica NTC PE 1000:2020 e indicadores asociados que permita a los responsables identificar posibilidades de mejora y a los usuarios identificar el alcance y las limitaciones en el uso de los datos. Avances: dashboard desarrollado y desplegado en entorno de prueba: https://dirpen.shinyapps.io/indicador_calidad_bd/ y el diseño y desarrollo de modelo de datos (MER) de los resultados de las evaluaciones de calidad del proceso estadístico.
6. **Detección de anomalías temáticas y operativas en el Censo Económico 2022.** Su objetivo es entrenar modelos de aprendizaje automático que permitan detectar casos que indiquen la posible ocurrencia de fallas en el proceso de entrevista y en los datos recolectados del Censo Económico, con el fin de facilitar el proceso de control y seguimiento del operativo. Dentro de los principales avances se encuentran: las pruebas en Censo Experimental 2021, el avance en diseño e implementación del flujo de datos y los documentos metodológicos.

4) Realimentación de los expertos sobre los proyectos presentados en el punto anterior (Expertos de la sala: Valérie Gauthier y León Parra- 10 minutos cada uno)

León manifestó su disposición para continuar trabajando articuladamente en los proyectos. Planteó dos retos a los que se enfrentan los proyectos: cómo ir articulando a otros actores que pueden estar relacionados con las temáticas y la interconexión con otras salas del CASEN para que de esta forma se logre alimentar mejor los insumos.

Valérie extendió la invitación para participar mediante la presentación de estos proyectos en la conferencia de Matemáticas Aplicadas Industriales a desarrollar en la ciudad de Medellín, con el objetivo de establecer lazos para trabajar de la mano con diferentes actores de la academia interesados en el tema. además, presentó la posibilidad de crear semilleros de investigación en la Universidad de los Andes para trabajar de la mano con el DANE en el marco del convenio que se quiere establecer más adelante.

Conclusiones y cierre (Sofía Sánchez- 5 minutos)

Elizabeth Moreno extendió un agradecimiento a los expertos y al equipo del DANE por reactivar el espacio de la sala. Puntualizó que, dentro de las apuestas de valor para este año, se tiene

programada la publicación del paper con la Universidad Ean, visibilizar estos proyectos y promover mediante espacios como el seminario web los desarrollos de la sala.

Por último, Sofía Sánchez cerró la reunión con los compromisos generados a lo largo de la sesión.

Compromisos

- 1. Tarea:** enviar el acta de la sesión
Responsable: DANE
Fecha entrega: a más tardar 19 de mayo
- 2. Tarea:** envío Doodle para agendar segunda sesión
Responsable: DANE
Fecha entrega: 30 de mayo de 2022
- 3. Tarea:** envío de documentación y papers
Responsable: Experto- León Parra
Fecha entrega: 13 de mayo de 2022
- 4. Tarea:** programar seminario web sobre protección de datos. Invitada especial: Andrea Martínez
Responsable: León Parra- experto de la sala
Fecha de entrega: a más tardar 27 de mayo de 2022
- 5. Tarea:** establecer contacto con Yesid y Sandra (expertos en el aspecto técnico- ciberseguridad), con Ángela, Andrés Umaña y Wilson Prieto (expertos en el aspecto de normatividad) y verificar la disponibilidad de agenda para realizar dos sesiones.
Responsable: Valérie Gauthier- experta de la sala
Fecha de entrega: a más tardar 27 de mayo de 2022

Expertos CASEN

DANE

Próxima reunión:

Responsable de convocar: Secretaría Técnica CASEN

Fecha: por definir

Anexos:

- Lista de asistencia

Nombres y Apellidos	Entidad	Dependencia	Número de contacto	Correo electrónico
León Darío Parra Bernal	Universidad Ean	Facultad de Administración, Finanzas Y Ciencias Económicas		
Valérie Gauthier Umaña	Universidad de los Andes	Departamento de Ingeniería de Sistemas y Computación		
Elizabeth Moreno Barbosa	DANE	DIRPEN- Dirección general		
Luis Martín Barrera Pino	DANE	Oficina de Sistemas		
Gilberto Villalba Gamboa	DANE	Oficina de Sistemas		
Juan Sebastián Ordóñez	DANE	Dirección general		
Mónica Patricia Pinzón	DANE	DIRPEN		
Ruth Constanza Triana	DANE	DIRPEN		
Andrés Felipe Villarraga	DANE	DIRPEN		
Sofía Sánchez Granados	DANE	DIRPEN		
Mateo Cardona	DANE	DIRPEN – Gobierno de Datos		
Anderson Leal Vélez	DANE	DIRPEN- Dirección técnica		
Alexander González	DANE	Prospectiva y Análisis de Datos		
Grace Andrea Torres	DANE	Prospectiva y Análisis de Datos		
Mónica Lorena Ortiz	DANE	Geoestadística		