

ACTA DE REUNIÓN N°4 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

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

Departamento Administrativo Nacional de Estadística (DANE)

Asesora DIRPEN y Dirección General
Elizabeth Moreno Barbosa

Coordinadora - Planificación y Articulación
Estadística
Mónica Patricia Pinzón

Oficina de Sistemas
Luis Martín Barrera Pino

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

Oficina de Sistemas
Mónica Patricia Pinzón Suárez

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

Oficina de sistemas
Gilberto Villalba Gamboa

DIRPEN- Dirección técnica
Anderson Leal Vélez

Asesor de la Dirección General
Juan Sebastián Ordóñez

Geoestadística
Mónica Lorena Ortiz

Orden del día

1. Solicitar autorización para grabar la reunión.
2. Verificación del quórum.
3. Síntesis de las reuniones anteriores (Sofía Sánchez– Tiempo: 5 minutos).
4. Presentación de las inquietudes del DANE sobre las oportunidades de uso de Blockchain en operaciones estadísticas (Luis Martín Barrera Oficina de Sistemas- Tiempo: 25 minutos).
5. Consideraciones de los expertos acerca de las inquietudes del DANE (León Darío Parra, Valérie Gauthier- Tiempo: 10 minutos por intervención).
6. Presentación Marco ético de los datos (Elizabeth Moreno B DIRPEN- Tiempo: 30 minutos).
7. Consideraciones de los expertos acerca del Marco ético de los datos (León Darío Parra, Valérie Gauthier- Tiempo: 5 minutos por intervención).
8. Socialización del documento final de la Sala (versión preliminar) (Sofía Sánchez Granados– GIT Planificación y Articulación Estadística, Tiempo: 10 minutos).
9. Consideraciones de los expertos acerca del documento final de la Sala (versión preliminar) (León Darío Parra, Valérie Gauthier- Tiempo: 5 minutos por intervención).
10. Conclusiones, compromisos y cierre de la reunión. (Mónica Pinzón Torres – Tiempo: 10 minutos).

Desarrollo

Objetivo

Presentar las inquietudes del DANE sobre las oportunidades de uso de Blockchain en operaciones estadísticas, así como el marco ético de los datos, con el fin de recibir realimentación por parte de los expertos y acoger sus recomendaciones y, realizar el cierre de la Sala con la presentación del documento de recomendaciones para el 2022.

1. Solicitar 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 de las reuniones anteriores (Sofía Sánchez – Tiempo: 5 minutos).

Sofía Sánchez inició su intervención realizando una síntesis de las reuniones realizadas durante el año y los principales aportes o recomendaciones brindadas por los expertos.

- 25 de marzo 2022- reunión bilateral con la experta Valérie Gauthier: como resultado de esta reunión se concretó la participación del DANE en el evento "Mapi2: segunda conferencia colombiana de Matemáticas aplicadas industriales".
- 30 de marzo 2022- reunión bilateral con el experto León Parra: como resultado de esta reunión, se estructuró un plan de trabajo para continuar con en el apoyo metodológico a proyectos DANE: Paper índice de noticias.
- 13 de mayo 2022- contexto DANE en materia de seguridad de datos gestionados en la nube y protección de datos confidenciales: se generó como compromiso, realizar reuniones internas con expertos en el aspecto técnico y jurídico de la seguridad de los datos en nube para realimentar los ejercicios adelantados por el DANE en la materia.
- 6 y 22 de junio- socialización y realimentación del Proyecto ODS sobre indicadores preliminares de discriminación y representatividad.

Se generaron las siguientes recomendaciones de los expertos de la sala en relación con la metodología:

1. Ampliar la técnica de Web Scraping a otras redes sociales.
2. Ampliar el entrenamiento de los modelos de aprendizaje supervisado y no supervisado con la combinación y comparación de técnicas.
3. Incluir en posteriores etapas del proyecto el análisis de imagen y video.
4. Incluir el análisis de sentimientos y el análisis de discurso de odio.

Se realizaron los siguientes comentarios en relación con los resultados:

1. Parametrizar en intervalos de tiempo, dado que los resultados pueden estar afectados por la coyuntura política.
2. El análisis horizontal con modelo no supervisado Zero Shot, se podría realizar excluyendo la variable política y dando lugar a enfatizar otras categorías.

3. Se sugiere estudiar la posibilidad de analizar el cruce entre categorías con el fin establecer posibles asociaciones entre grupos o tendencias.
 4. Revisar las categorías y etiquetas para las dimensiones de inclusividad, receptividad y representación política.
- 17 de junio 2022 (10 a 11 am)- reunión sobre el aspecto técnico de la seguridad de los datos con el grupo de ciberseguridad de la Universidad de los Andes.

Se generaron las siguientes recomendaciones:

1. Separar conceptualmente la "seguridad" y la "privacidad".
 2. Previo a migrar datos a la nube se deben verificar las necesidades, los requerimientos y prever las consecuencias.
 3. Revisar la tendencia actual de "privacidad diferencial" para determinar el modelo adecuado para el procesamiento de datos en nube de acuerdo con las características de la entidad.
- 17 de junio 2022 (11 a 12 am)- reunión sobre el aspecto normativo de la seguridad de los datos con expertos externos en el tema y 14 de julio 2022- reunión interna Oficina de Sistemas del DANE-experta Andrea Martínez (invitada a la sala) sobre las restricciones de negocio para la seguridad de los datos al adoptar nube.

Se destacan las siguientes recomendaciones:

1. Los principios de reserva estadística aplican aún para las excepciones.
2. Al migrar datos a la nube, se debe aplicar el máximo nivel de seguridad.
3. Las entidades públicas pueden usar la nube sin ninguna restricción legal.
4. Blindar los contratos con los proveedores de nube (debida diligencia) con anexos robustos que garanticen que el proveedor cumpla con la normatividad local.
5. Se sugiere revisar casos de buenas prácticas: DIAN, Ministerio de Salud, Colombia Compra Eficiente y Secretaría de Educación de Bogotá.

4. Presentación de las inquietudes del DANE sobre las oportunidades de uso de Blockchain en operaciones estadísticas (Luis Martín Barrera Oficina de Sistemas- Tiempo: 25 minutos).

Luis Martín Barrera, jefe de la Oficina de Sistemas no pudo asistir a la reunión, por lo cual Mónica Pinzón Suárez realizó la presentación. Primero, realizó una contextualización sobre la Producción de información estadística.

Señaló que el soporte a Sistemas de Información al interior del DANE se maneja de la siguiente forma: se resuelven las incidencias de fallo reportadas por los usuarios de los sistemas de información y se apoya con el procesamiento de datos en las fases de la producción estadística que lo requieran.

En relación con el desarrollo y mantenimiento de Sistemas de Información dijo que se realizan diferentes etapas del ciclo de vida del desarrollo de software para los sistemas de información requeridos por las diferentes áreas de la Entidad, cumpliendo con los lineamientos de la Oficina de Sistemas.

Después, Mónica habla respecto a la atención de solicitudes. Señala que en un primer momento dependen del tipo de requerimiento. Así, es como se tiene un tiempo de solución establecido de 65 días hábiles, bien sea para desarrollos de sistemas de información nuevos, para mantenimientos a sistemas de información activos o nuevos requerimientos (adquisición de un nuevo sistema de información, retiro de sistema de información en producción).

Por otro lado, según tipo de incidencia, se tiene un tiempo de solución: 24 horas hábiles cuando se trata de soportes, cargue de información o novedades.

Para hablar de la definición de la arquitectura utilizada por el DANE, Mónica se refirió a Mysurvey y al modelo de integración del DANE con las entidades del SEN para generar un ecosistema seguro de intercambio.

- **Definición de la arquitectura (My Survey)**

La arquitectura se diseñó con el objetivo de orientar los desarrollos de software a un estilo orientado a microservicios, permitiendo beneficios tales como: flexibilidad, reutilización de componentes, interoperabilidad, escalabilidad, reducción de tiempo y costos de implementación.

Estas permiten construir aplicaciones con diseño web adaptable basadas en componentes lo cual optimiza la reutilización de código fuente, comunicación asíncrona entre servicios mediante protocolos y estándares. Para la reutilización de los microservicios e interoperabilidad de los mismo es necesario contar con una documentación centralizada y estandarizada, teniendo en cuenta los ambientes de pruebas y producción de la entidad.

Por último, Mónica planteó a los expertos una serie de inquietudes frente al uso de Blockchain en operaciones estadísticas.

5. Consideraciones de los expertos acerca de las inquietudes del DANE (León Darío Parra, Valérie Gauthier- Tiempo: 10 minutos por intervención).

Valérie Gauthier inició su intervención con una presentación sobre Blockchain. Señaló que uno de los aspectos más importantes del Blockchain es que es una cadena de bloques. Se puede ver como

un libro contable, con páginas enumeradas, es decir, que uno no puede arrancar una página o cambiar un contenido porque se van a dar cuenta y, que puede estar cifrada o no.

La función de Hash es algo que achica un contenido a un tamaño de bits que uno quiera de tal manera que no pueda encontrar colisiones fáciles, es decir, dos bloques que lleguen a la misma función de Hash (puesto que así se podría modificar un bloque que tenga una preimagen) y que tampoco le pueda tener una preimagen o una segunda imagen.

Si se quiere un proyecto del que se quiera tener trazabilidad, de tal forma que no se pueda borrar alguna información en la mitad, es bueno hacerlo con Blockchain.

La experta señaló que la segunda idea importante de Blockchain se refiere a las redes de peer to peer. Es algo que funcionaba en sistemas desde antes. Uno podía tener un sistema centralizado donde si la información de un computador se borra, todos se borran o una información donde todos se conectan con todos. Aparece entonces el espíritu de la descentralización, puesto que mediante el Blockchain se eliminan los intermediarios y se realizan transacciones a mayor velocidad.

Frente a las preguntas realizadas por la Oficina de Sistemas, la experta dijo lo siguiente:

- En relación con los requisitos que se deben cumplir para migrar a la tecnología de Blockchain, se debe mirar qué tipo de Blockchain quiere el DANE, el impacto que tiene puesto que no sirve para todo. Se pueden hacer contratos inteligentes, o dejar trazabilidad.
- Los requisitos tecnológicos dependen de las plataformas.
- La seguridad depende de cómo se estructure el Blockchain. Hay muchos aspectos por explorar, por ejemplo, la firma que se utilice. Es decir, se podría generar una cadena de bloques descentralizada entre diferentes entidades o a nivel interno, que todos los usuarios puedan actualizar, pero que no permita eliminar la información. Además, debe estar cifrada con los mismos estándares frente a la reserva.

Por último, Valérie planteó la posibilidad de establecer contacto mediante una reunión con Binance- empresa de exchange con criptos que busca volverse el centro de la web 3.0. que podría poner la plataforma para desarrollar de la mano con el DANE proyectos para explorar las posibilidades del Blockchain, por ejemplo. Así mismo, planteó la opción de coordinar con un profesor de la Universidad de los Andes para tener un primer ejemplo. Mónica señaló que le parece muy interesante con el ánimo de solucionar dudas y poder validar esta tecnología que aún es muy novedosa.

El experto León Parra señaló que con la presentación de Valérie se resolvieron varias de las inquietudes. Después continuó su intervención haciendo un paralelo entre usar una estructura de Blockchain (ventajas y desventajas) y seguir con el lago de datos. Más aún para el caso del DANE que se está utilizando la plataforma de X-ROAD de interoperabilidad donde hay que ser cuidadoso entre cómo se maneja la información entre actores.

Frente a las preguntas realizadas por la Oficina de Sistemas, el experto dijo lo siguiente:

- En lo referente al mandato de reserva estadístico que debe cumplir la entidad, desde el DANE pueden desarrollar un token para manejo interno de la entidad, pero ya como tal para el manejo de los datos de los usuarios dado que ya se está con X-ROAD, tendrían que ser cuidadosos y ver si la cadena de bloques si cumple con la trazabilidad para garantizar la reserva de la información entre los diferentes usuarios y cuál sería la más adecuada.
- En cuanto el impacto de los procesos de desarrollo en la implementación de Blockchain, dijo que se tienen dos caminos: 1) elegir un operador que ya esté posicionado en el mercado como Binance o Ethereum, quienes cobran por las transacciones. 2) ejemplo: caso historias clínicas y trazabilidad (2019). El ejercicio pretendía mirar cómo las diferentes empresas que a estaban trabajando en el área de Blockchain podían desarrollar una prueba de token (utilizando alguna de las plataformas) para que la gente desarrollara código y probara. Se evidenció que la tecnología permite, por ejemplo, anonimizar la información, sin embargo, se presentaron barreras normativas para el manejo de la información con esta tecnología.
- Plantear la pregunta y el propósito único para utilizar el Blockchain, así como las limitaciones y alcance de unas iniciativas como etas en el DANE. Teniendo en cuenta que por el X-ROAD ya se tendría una limitante en donde se tiene que resguardar la trazabilidad de la información hacia otras entidades.

6. Presentación Marco ético de los datos (Elizabeth Moreno B DIRPEN- Tiempo: 30 minutos).

Elizabeth Moreno inició la presentación dando un contexto a los asistentes sobre el Comité de Administración de Datos- CAD. El CAD es unidad de articulación y coordinación que lidera la gestión y administración de datos, bajo la implementación de un esquema de gobernanza que garantiza la articulación entre la producción de información estadística con las políticas públicas en ambientes seguros.

Luego, señaló la normatividad asociada al marco ético de los datos:

- Decreto 2404 del 27 de diciembre de 2019
- Resolución número 0451 del 26 de marzo de 2020

Después, para referirse a la gestión ética de los datos, dijo que de acuerdo con Payam Moula y Per Sandin, ésta facilita a las organizaciones la toma de decisiones ante situaciones éticas. Al ser una herramienta que ayuda a una entidad o comunidad mejorar la calidad de sus deliberaciones con el fin de alcanzar decisiones y juicios éticamente informados (2017).

Además, facilita el diálogo entre los distintos grupos involucrados en el manejo de los datos y la concreción de los valores en acciones particulares.

Para entrar en materia frente al marco ético de los datos, dijo que se tomaron como referentes dos modelos. Frente al modelo de Davis, este parte de los valores y su concreción en acciones. Integrado por cuatro componentes: investigación, análisis, articulación y acción. De conformidad con Davis,

las organizaciones deben tratar de “alinear sus acciones con sus valores. El modelo del BID parte de los riesgos envueltos en el manejo de los datos y establece criterios para la gestión ética de los datos. Contempla cinco etapas: recolección, almacenamiento, análisis, compartición y archivo.

Luego, Elizabeth explicó la síntesis entre los dos modelos, señaló las etapas (recolección; almacenamiento; análisis; compartición; archivo y eliminación) y los componentes (análisis/ evaluación; articulación/ plan de acción y acción/ puesta en marcha del plan) del ciclo de vida de los datos. Esto, para concluir que la aplicación de los valores: privacidad y seguridad, derechos humanos, no maleficencia, confiabilidad: exactitud y precisión, relevancia, eficiencia y usabilidad; y transparencia se da de manera transversal.

En esta medida, explicó uno a uno en qué consisten los principios éticos de la gestión ética de los datos: privacidad y seguridad; derechos humanos; no maleficencia; confiabilidad: precisión y exactitud; relevancia; eficiencia y usabilidad; y, transparencia.

Por último, explicó el diseño institucional de la evaluación de la gestión ética de los datos del SEN. El cual consisten en dos sistemas: el Sistema de evaluación de las operaciones estadísticas y el Sistema de evaluación de la gestión ética de los datos. La estructura del sistema inicia con una fase de autodiagnóstico de las operaciones estadísticas y está liderada por las entidades que conforman el SEN, luego, a través del Sistema de Ética Estadística- SETE se lleva a cabo una fase de evaluación y acompañamiento y, por último, es el Comité de Administración de Datos- CAD quien da el aval.

7. Consideraciones de los expertos acerca del Marco ético de los datos (León Darío Parra, Valérie Gauthier- Tiempo: 5 minutos por intervención).

Los expertos de la sala felicitaron a Elizabeth y en general al equipo del DANE por el trabajo adelantado frente a la sistematización del documento preliminar del marco ético de los datos, señalaron que es un marco de referencia para la región y representa una herramienta muy valiosa para la sala, por lo cual agradecieron la socialización y el espacio para manifestar sus impresiones. Solicitaron el envío del documento para realizar comentarios al respecto con mayor profundidad.

8. Socialización del documento final de la Sala (versión preliminar) (Sofía Sánchez Granados– GIT Planificación y Articulación Estadística, Tiempo: 10 minutos).

Sofía Sánchez señaló que el documento tiene como objetivo presentar al SEN y a la sala general del CASEN, las recomendaciones realizadas para el año 2022 por los miembros de la Sala Especializada de Modernización Tecnológica de la Producción Estadística y el DANE.

Esto, teniendo en cuenta que, durante el año 2022, se abordaron las siguientes líneas de investigación: 1) Big Data e interoperabilidad y 2) Seguridad de la Información.

A lo largo del texto se resalta la metodología utilizada para el desarrollo de la sala, en donde se alternó la realización de:

1. Reuniones plenarios: en donde los pares del DANE (Oficina de Sistemas), presentaron diferentes temas de interés en el contexto de la entidad y las consecuentes preguntas, con el objetivo de recibir realimentación por parte de los expertos. Se destacan las reuniones realizadas sobre Bigdata y Blockchain.
2. Reuniones internas: las cuales contaron con el acompañamiento de invitados, que nutrieron el debate y aportaron desde sus campos del conocimiento a ampliar la mirada del DANE, en temas como la seguridad de los datos en la nube desde el aspecto jurídico y de ciberseguridad.
3. Trabajo bilateral con los expertos: reuniones realizadas con el objetivo de realizar seguimiento al cronograma de actividades de la sala y a los compromisos adquiridos por cada experto, de estas reuniones se destaca la articulación para llevar a cabo el acompañamiento a diferentes proyectos del DANE: el documento final del paper índice de noticias- 2022 “Colombian agricultural sector’s early estimator of GDP post-pandemic COVID19 using Google News and Google Trends”, las recomendaciones metodológicas y sobre los resultados al proyecto de cálculo de Indicadores del ODS 16 usando redes sociales y la participación en la conferencia de Matemáticas aplicadas e industriales – MAPI- desarrollada en la ciudad de Medellín del 8 al 10 de junio de 2022.

En esta medida, vale la pena señalar que la estructura del documento de recomendaciones inicia con el planteamiento de los objetivos, continúa con los antecedentes, el marco conceptual de la sala y las recomendaciones de los expertos para las líneas de interés.

9. Consideraciones de los expertos acerca del documento final de la Sala (versión preliminar) (León Darío Parra, Valérie Gauthier- Tiempo: 5 minutos por intervención).

Los expertos de la sala, León Parra y Valérie Gauthier dijeron estar de acuerdo con la estructura y contenido del documento de recomendaciones. Una vez se envíe el documento final, realizarán los comentarios a que haya lugar.

10. Conclusiones, compromisos y cierre de la reunión. (Mónica Pinzón Torres – Tiempo: 10 minutos).

Mónica Pinzón Torres realizó una recopilación del número de sesiones trabajadas durante la vigencia 2022. Para la línea de investigación de Bigdata e interoperabilidad, se realizaron tres reuniones. Mientras que, para la línea de Seguridad de la información, se llevaron a cabo cinco reuniones.

De conformidad con el número de sesiones (2) desarrolladas durante el año, cada uno de los expertos recibirá \$10.000.000 de pesos como retribución por las labores desempeñadas en el marco de la Sala y el Consejo Asesor técnico del Sistema Estadístico Nacional- CASEN. Luego, Mónica preguntó a los expertos si tenían alguna duda al respecto lo que ellos respondieron de manera

negativa. Por último, señaló que en los próximos días se estarían enviando un correo solicitando los respectivos documentos para realizar el trámite del pago.

Compromisos

1. **Tarea:** envío del acta de la sesión
Responsable: DANE
Fecha entrega: 29 de noviembre de 2022
2. **Tarea:** envío ensayos de experiencias CASEN 2020-2022
Responsable: DANE
Fecha entrega: 25 de noviembre de 2022
3. **Tarea:** envío del documento final de recomendaciones 2022
Responsable: DANE
Fecha entrega: diciembre de 2022
4. **Tarea:** participación en la sesión de cierre CASEN- presencial
Responsable: integrantes de la Sala de modernización tecnológica de la producción estadística
Fecha de entrega: 29 de diciembre de 2022 (10:00 a.m. a 12:00 m.)
5. **Tarea:** envío de la documentación para el pago (dos sesiones)
Responsable: expertos de la sala
Fecha de entrega: 9 de diciembre de 2022
6. **Tarea:** realizar contacto con la empresa Binance para plantear proyectos conjuntos
Responsable: Valérie Gauthier
Fecha de entrega: por definir

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		
Julieth Solano Villa	DANE	DIRPEN		
Elizabeth Moreno Barbosa	DANE	DIRPEN- Dirección general		
Luis Martín Barrera Pino	DANE	Oficina de Sistemas		
Mónica Patricia Pinzón Suárez	DANE	Oficina de Sistemas		
Gilberto Villalba Gamboa	DANE	Oficina de Sistemas		
Juan Sebastián Ordóñez	DANE	Dirección general		
Mónica Patricia Pinzón Torres	DANE	DIRPEN		
Ruth Constanza Triana	DANE	DIRPEN		
Sofía Sánchez Granados	DANE	DIRPEN		
Anderson Leal Vélez	DANE	DIRPEN- Dirección técnica		
Mónica Lorena Ortiz	DANE	Geoestadística		