

Consejo Asesor Técnico del Sistema Estadístico Nacional (CASEN) 2023-2025

Sala Especializada para la Modernización Tecnológica de la Producción Estadística ACTA No. 3

Ciudad: Bogotá D.C.

Enlace: <https://acortar.link/MOzZ11>

Tema: Recibir realimentación y recomendaciones por parte de los miembros de la sala.

Hora: 2:00 p. m. a 4:00 p. m.

Fecha: 26/04/2024

Dependencia responsable: Secretaría Técnica del CASEN.

Participantes

Miembros de la Sala Especializada para la Modernización Tecnológica de la Producción Estadística del CASEN

Mario Linares Vásquez.

Nicolás Cardozo Álvarez.

León Darío Parra.

Departamento Administrativo Nacional de Estadística (DANE)

Julieth Alejandra Solano Villa, directora técnica de la Dirección de Regulación, Planeación, Estandarización y Normalización (DIRPEN) y secretaria técnica CASEN.

Luis Martin Barrera Pino, jefe Oficina de Sistemas (OSIS).

Natalia Ximena Arteaga Gutiérrez, asesora subdirección.

Andrea Roncancio, coordinadora Grupo Interno de Trabajo (GIT) Prospectiva y Analítica de Datos (PAD), DIRPEN.

Andrea Catherine Neira Bustamante, designada OSIS.

Alejandro Sandoval Pineda, designado OSIS.

Daniel Mauricio Montenegro Reyes, designado OSIS.

Luis Fernando Barajas Duarte, designado OSIS.

Victoria Adriana González Ramírez, designada DSCN.

Juan José Rubio Mesa, designado DIMPE.

Secretaría Técnica del CASEN - DIRPEN

Julieth Alejandra Solano Villa, directora técnica DIRPEN y secretaria técnica CASEN.

Sandra Yaneth Cortés Gamba, responsable general del CASEN, GIT PAE.

Elizabeth Moreno Barbosa, asesora.

Derly Vivian Lizarazo García, responsable de la Sala Especializada para la Modernización Tecnológica de la Producción Estadística.

Yennifer Castillo Murcia, GIT Planificación y Articulación Estadística (PAE).

Agenda

Tiempo	Actividad	Responsable
2:00 p. m. a 2:05 p. m.	Instalación, verificación del quorum y registro fotográfico.	Derly Lizarazo, responsable de la sala.
2:05 p. m. a 2:10 p. m.	Apertura del evento.	Julieth Solano, directora técnica DIRPEN y secretaria técnica del CASEN.
2:10 p. m. a 2:15 p. m.	Síntesis de la reunión anterior y compromisos.	Derly Lizarazo, responsable de la sala.
2:15 p. m. a 2:30 p. m.	Espacio de realimentación y recomendaciones por parte de los miembros de la sala, sobre el documento <i>Resumen de los aspectos teóricos y conceptuales de la línea ciclo de vida</i> .	Miembros de la sala (5 minutos): Mario Linares Nicolás Cardozo León Darío Parra Modera: Alejandro Sandoval, designado OSIS.
2:30 p. m. a 3:05 p. m.	Presentación de la línea 1. <i>Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencia de datos, 1.1.2. Exploración de herramientas y plataformas para cada etapa del ciclo de vida de los modelos de ciencia de datos.</i>	Alejandro Sandoval, designado OSIS.
3:05 p. m. a 3:50 p. m.	Presentación de la línea 1.2. <i>Proyectos de ciencia de datos, App Diversa y espacio de realimentación y</i>	Modera: Andrea Roncancio, coordinadora GIT PAD.

Tiempo	Actividad	Responsable
	recomendaciones por parte de los miembros de la sala.	
3:50 p. m a 3:55 p. m.	Compromisos de la reunión.	Derly Lizarazo, responsable de la sala.
3:55 p. m a 4:00 p. m.	Conclusiones y cierre.	Julieth Solano, directora técnica DIRPEN y secretaria técnica del CASEN.

Objetivo: recibir realimentación y recomendaciones por parte de los miembros de la sala, sobre el documento *Resumen de los aspectos teóricos y conceptuales de la línea ciclo de vida* y presentar la línea 1. *Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencia de datos, 1.1.2. Exploración de herramientas y plataformas para cada etapa del ciclo de vida de los modelos de ciencia de datos*, la línea 1.2. *Proyectos de ciencia de datos*, la App diversa y dar un espacio de realimentación y recomendaciones por parte de los miembros de la sala.

Desarrollo

1. Instalación, verificación del quórum y apertura de la reunión

Se realizó la verificación del quórum, contando con la participación de los tres miembros de la sala, los designados de las Oficina de Sistemas, del GIT Prospectiva y Analítica de Datos de la DIRPEN, de la Dirección de Síntesis y Cuentas Nacionales (DSCN) y de la Dirección de Metodología y Producción Estadística (DIMPE), así como de la secretaría técnica del CASEN y su equipo de trabajo. Posterior a ello, se realizó el registro fotográfico y se le dio la palabra a Julieth Solano, directora técnica de la Dirección de Regulación, Planeación, Estandarización y Normalización (DIRPEN) y secretaría técnica del CASEN, quien realizó la apertura a la reunión.

2. Síntesis de la reunión anterior y compromisos

Derly Lizarazo, responsable de la sala, realizó una breve síntesis de la reunión anterior, señalando la aprobación del plan de trabajo 2024-2025 e inicio de la línea de investigación *No. 1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencia de datos*. actividad: definición del marco teórico asociado con su respectiva realimentación por parte de los miembros de la sala.

Posteriormente recordó los compromisos establecidos como remitir el material de trabajo y el asunto de consulta (15 de abril) para la próxima reunión (la línea de investigación *No. 1: Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencia de datos* y el documento *Resumen de los aspectos teóricos y conceptuales relacionados con el ciclo de vida de modelos en proyectos de ciencia de datos*) para la

correspondiente realimentación por parte de los miembros de la sala y para la Línea de investigación 1. *Ciencia de datos, 1.2. Proyectos de ciencia de datos* y la App Diversa.

3. Espacio de realimentación y recomendaciones por parte de los miembros de la sala, sobre el documento: resumen de los aspectos teóricos y conceptuales de la línea ciclo de vida

En este espacio, los tres miembros de la sala precisaron que luego de revisar el documento evidenciaron que se habían incluido los términos recomendados y que faltaría incluir los conceptos de: inteligencia artificial general, inteligencia artificial débil y fuerte.

El profesor Mario Linares, preguntó si el alcance y la divulgación del documento sería a nivel interno del DANE o se podría enviar a las entidades para generar cultura alrededor de la inteligencia artificial e incluirlos en los cursos y sugirió que podría ser uno de los documentos que debe generar la sala. A lo que Julieth Solano, respondió que se había pensado inicialmente que sería a nivel interno del DANE, pero que se tendría en cuenta la sugerencia realizada por el profesor Linares.

4. Presentación de la línea 1. *Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencia de datos, 1.1.2. Exploración de herramientas y plataformas para cada etapa del ciclo de vida de los modelos de ciencia de datos*

Alejandro Sandoval designado de la OSIS realizó la presentación e indicó que el asunto de consulta era establecer el listado de infraestructura y herramientas con su descripción, aspectos clave, beneficios y limitaciones posterior a la discusión y las observaciones por parte de los expertos. Así mismo, se mencionaron tres tipos de infraestructura (local, nube o modelo híbrido) sus beneficios y limitaciones.

Frente a la gestión de paquetes y entornos virtuales se identificaron Conda, PIP, Pyenv, Poetry, Renv y Pactrak y se presentaron los beneficios y las limitaciones. Las sugerencias de los miembros de la sala fueron:

- Parte de la evaluación de estos manejadores debería tener en cuenta los procesos de verificación (Safety).
- Se compartió la herramienta que revisa la seguridad de los paquetes.
- Se recomendó hacerlos parte de Jupiter o el Colap.

Sobre el control de versiones de código, se mencionaron Bitbucket, Gitlab, Github, Apache Subversión, así como los beneficios y las limitaciones. Las sugerencias de los miembros de la sala fueron:

- Diferenciar lo que es el sistema de control de versiones y que son las plataformas para control de versiones.
- Falta incluir el de Microsoft (Asus de Box).

- Cambiar lo que se hacía con SNB por GIT.
- Se recomendó el uso del Github, ya que tiene sus propios protocolos para la fuga de información.

Sobre el control de versiones de datos se mencionaron Collibria, Alation, Data Versión Control, sus beneficios y sus limitaciones y frente a esto se recomendó revisar Kit LFSY, Dolto, DLT o GIT FNS.

Respecto al control de gestión de ciclo de vida de modelos de datos y la distribución y la ejecución de aplicaciones, no se presentaron recomendaciones por parte de los miembros de la sala.

Se concluyó, explicando sobre el desarrollo, el monitoreo y la gestión de proyectos, donde los miembros de la sala sugirieron incluir Datadog, que es una versión paga.

5. Presentación de la línea 1.2. *Proyectos de ciencia de datos, App Diversa y espacio de realimentación y recomendaciones por parte de los miembros de la sala*

Andrea Roncancio, coordinadora del GIT Prospectiva a Analítica de Datos, presentó el contexto sobre la línea de proyectos de ciencia de datos App Diversa. Luego continuó Omar Beltrán, que describió la necesidad basada en: el creciente uso de dispositivos móviles, la generación de datos por parte de los usuarios, la posibilidad de ofrecer valor y servicios a los usuarios, la experiencia atractiva de uso, la difusión a través de redes sociales y la disponibilidad de plataformas de desarrollo de bajo costo.

Se presentaron los hitos de la App Diversa, el desarrollo del prototipo de formulario de recolección de datos en Drupal y las generalidades de la herramienta. Asimismo, la investigación de buenas prácticas en recolección de datos voluntarios y diferentes gráficas relacionadas con patrones de uso internet en Colombia.

Luego de la exposición, se leyeron las preguntas orientadoras para realimentación por parte de los miembros de la sala. Estas fueron:

- ¿Qué opinan del desarrollo?
- ¿Cómo se puede mejorar el proceso de desarrollo?
- ¿Qué opinan de las buenas prácticas de recolección de datos voluntarios?

Los comentarios recibidos por parte de los tres miembros fueron:

- Se debe testear al usuario para que no abandone la aplicación y así obtener un buen resultado, una buena experiencia de usuario; esto hace parte de las fases que tiene un proyecto de este tipo.
- Dado que se está buscando estandarización, es importante contar con una guía de accesibilidad, interfaz gráfica y experiencia de usuario.

- Más allá de estandarizar la herramienta, es importante estandarizar como se maneja la interfaz gráfica y el proceso de desarrollo.
- Se puede categorizar o construir un manual de diseño, donde se tengan definidas herramientas de acuerdo con la encuesta, el estilo o los atributos que se van a tener en cuenta.
- Con Drupal se debe tener cuidado con la seguridad porque utiliza parches por lo que es necesario tener una línea de investigación continua en el área de trabajo para estar pendiente de probar la herramienta, la seguridad, de cuáles son los últimos parches, dado que esto garantiza la calidad de la herramienta.
- El mecanismo de enganche es importante porque es lo que lleva a que una persona responda una encuesta de manera voluntaria (valor monetario, valor emocional).

Compromisos

Tarea	Envío de acta de la presente reunión para revisión y aprobación de los miembros y participantes de la sala.
Responsable	Responsable de la sala: Derly Lizarazo.
Fecha entrega	09/05/2024.
Tarea	Remitir material de trabajo y asunto de consulta para los siguientes temas: 1. Reunión 10 de mayo: línea de investigación <i>1. Ciencia de datos, 1.2. Proyectos de ciencia de datos / 1.2.1. Índice de noticias</i> . Se remitirá lo correspondiente el 3 de mayo. 2. Reunión 24 de mayo: línea de investigación <i>1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencias de datos, No. 1.1. Revisión de prácticas de Ingeniera de Software para Machine Learning</i> . Se remitirá lo correspondiente el 10 de mayo.
Responsable	Responsable de la sala: Derly Lizarazo.
Fecha entrega	3 y 10/05/2024.
Tarea	Avances en la propuesta del taller de cultura de la arquitectura de datos
Responsable	Derly Lizarazo, Sandra Cortés, Mauricio
Fecha entrega	

Próxima reunión

Responsable de convocar: Secretaría Técnica del CASEN.

Fecha: 10 de mayo de 2024 de 2:00 p. m. a 4:00 p. m. (presencial en el DANE).