



DANE

SEN^{2.0}
Sistema Estadístico
Nacional-Colombia

CONSEJO TÉCNICO ASESOR DEL SISTEMA ESTADÍSTICO NACIONAL - CASEN

SALA ESPECIALIZADA PARA LA MODERNIZACIÓN TECNOLÓGICA DE LA PRODUCCIÓN ESTADÍSTICA

DOCUMENTO DE RECOMENDACIONES Y PRODUCTOS GENERADOS POR LA SALA

Diciembre de 2025

**DEPARTAMENTO ADMINISTRATIVO
NACIONAL DE ESTADÍSTICA
-DANE**

B. PIEDAD URDINOLA CONTRERAS
Directora

ANDREA RAMÍREZ PISCO
Subdirectora

ÁLVARO FERNANDO GUZMÁN LUCERO
Secretaria General

DIRECTORES TÉCNICOS

CÉSAR MAURICIO LÓPEZ ALFONSO
Dirección de Metodología y
Producción Estadística

JAVIER PLAZAS ECHEVERRI
Dirección de Censos y Demografía

CARLOS ALBERTO HERNÁNDEZ LOZANO
Dirección de Regulación, Planeación,
Estandarización y Normalización

JUAN PABLO CARDOSO TORRES
Dirección de Síntesis y Cuentas
Nacionales

ELKIN ERNESTO RAMÍREZ NIÑO
Dirección de Geoestadística

DIANA MARÍA BOHÓRQUEZ LOSADA
Dirección de Difusión, Mercadeo y
Cultura Estadística

DIANA MARÍA ROJAS ORDUS
Dirección de Recolección y Acopio

© DANE, 2025

Prohibida la reproducción total o parcial sin
permiso o autorización del Departamento
Administrativo Nacional de Estadística
Colombia

MARIO LINARES VÁSQUEZ
NICOLÁS CARDOZO ÁLVAREZ
LEÓN DARIO PARRA BERNAL

Expertos

Consejo Técnico Asesor del Sistema
Estadístico Nacional -CASEN

GERMAN LEONIDAS ORJUELA BORDA
DERLY VIVIAN LIZARAZO GARCÍA

Coordinación de Planificación y Articulación
Estadística

Contenido

INTRODUCCIÓN	4
OBJETIVO GENERAL	5
ANTECEDENTES	5
3. RECOMENDACIONES	6
3.1 Proyectos de Ciencia de datos	6
3.1.1 Modelo tecnológico para la producción de registros estadísticos base	7
de unidades económicas y población	7
3.1.2 APP Diversa.....	9
3.1.3 Modelo de evaluación de la calidad de bases de datos de operaciones estadísticas.....	10
3.2 Desarrollo de Sistemas de Información.....	11
3.2.1 Guía para el proceso de interoperabilidad	12
3.2.2 Procedimiento para ingreso seguro mediante Single Sign-On (SSO)	12
3.2.3 Guía para la construcción de un sistema de información estadística	13
3.2.4 Módulo de sistemas de información del Sistema de Identificación y Caracterización de Oferta y Demanda Estadística del SEN – SICODE.....	16
4. PRODUCTOS DE LA SALA	17

INTRODUCCIÓN

El Departamento Administrativo Nacional de Estadística -DANE, a partir de la Ley 2335 de 2023 en su artículo 14, establece la creación del Consejo Asesor Técnico del Sistema Estadístico Nacional (CASEN), que tiene como objeto asesorar al DANE y al Gobierno Nacional sobre cuestiones de importancia estratégica para las estadísticas oficiales del país. Está compuesto por la sala general y cinco salas especializadas en diferentes temáticas: i) Sala especializada para la modernización tecnológica de la producción estadística, ii) Sala especializada de salud, bienestar social y demografía, iii) Sala especializada de gobierno, seguridad y justicia, iv) Sala especializada de geografía, medio ambiente y ordenamiento territorial y v) Sala especializada de economía. Cada sala especializada estará integrada por tres (3) expertos.

La sala especializada para la Modernización tecnológica de la producción estadística está conformada por los expertos Mario Linares Vásquez, Nicolás Cardozo Álvarez y León Darío Parra Bernal, Carlos Alberto Hernández Lozano como secretaria técnica de la sala y los delegados de las diferentes direcciones del DANE: Dirección de Regulación, Planeación, Estandarización y Normalización (DIRPEN) y Oficina de Sistemas (OSIS). La metodología que se utilizó para el desarrollo de la sala consistió en reuniones virtuales, en donde se presentaron los avances en temas de interés en el contexto de la entidad y se propusieron preguntas orientadoras, con el objetivo de recibir realimentación por parte de los expertos.

Este documento tiene como objetivo presentar al Sistema Estadístico Nacional (SEN) y a la sala general del CASEN, las recomendaciones realizadas por los miembros de la Sala especializada para la modernización tecnológica de la producción estadística y el DANE, frente a las líneas de investigación propuestas: i) Proyectos de ciencia de datos; y ii) Desarrollo de sistemas de información.

A continuación, se presentan el objetivo general del documento, los antecedentes y concluye con las recomendaciones de los expertos para las líneas de investigación trabajadas en la sala en el 2025.

OBJETIVO GENERAL

Presentar las recomendaciones realizadas por los miembros de la Sala especializada para la modernización tecnológica de la producción estadística durante el 2025, en las líneas de investigación: i) Proyectos de datos; y ii) Desarrollo de sistemas de información.

ANTECEDENTES

Para la formulación del plan de trabajo 2025 se consideraron las iniciativas adelantadas por el DANE en años anteriores, relacionadas con proyectos de ciencia de datos y el desarrollo de sistemas de información. Estas acciones se orientaron a fortalecer la capacidad técnica del Sistema Estadístico Nacional (SEN), garantizar la calidad de los procesos y promover la interoperabilidad entre las diferentes entidades productoras de información estadística.

El propósito fundamental fue consolidar productos y herramientas que contribuyeran a la modernización de la infraestructura tecnológica y metodológica del SEN, asegurando que las soluciones propuestas respondieran a estándares internacionales y a las necesidades específicas del contexto nacional.

Línea de proyectos de ciencia de datos, en esta línea se incluyeron los siguientes temas estratégicos:

1. Modelo tecnológico para la producción de registros estadísticos base de unidades económicas y población, cuyo objetivo fue establecer un marco conceptual y recomendaciones técnicas que permitieran definir un modelo robusto para la generación de registros estadísticos. Este modelo busca optimizar la integración de fuentes administrativas y mejorar la cobertura y calidad de la información.
2. APP Diversa, iniciativa desarrollada durante 2024, que realizó pruebas piloto en el primer semestre de 2025. Este proyecto se orientó a evaluar la viabilidad técnica y metodológica de la aplicación, con el fin de identificar oportunidades de mejora, garantizar la pertinencia de los procesos y asegurar que la herramienta cumpla con los requerimientos del SEN.

3. Modelo de evaluación de la calidad de bases de datos de operaciones estadísticas, diseñado para establecer criterios y métricas que permitan valorar la consistencia, integridad y confiabilidad de las bases de datos. Se buscó incorporar observaciones y aportes de expertos para fortalecer los componentes del modelo y asegurar su aplicabilidad en diferentes operaciones estadísticas.

Línea de desarrollo de sistemas de información, en esta línea se contemplaron los siguientes temas:

1. Guía para el proceso de interoperabilidad, documento elaborado en 2024 que requería comentarios finales por parte de expertos antes de su difusión. Esta guía constituye un insumo clave para garantizar la integración eficiente de sistemas y la circulación segura de datos entre entidades.
2. Procedimiento para ingreso seguro mediante Single Sign-On (SSO), orientado a validar con los expertos del CASEN el mecanismo de autenticación y autorización para el desarrollo y mantenimiento de sistemas de información. El objetivo fue recibir recomendaciones que permitieran actualizar el procedimiento y fortalecer la seguridad de los accesos.
3. Guía para la construcción de un sistema de información estadística, cuyo propósito fue socializar el documento con expertos para recoger observaciones y recomendaciones que contribuyeran a su mejora. Esta guía busca estandarizar los procesos de diseño y desarrollo de sistemas, asegurando su alineación con las necesidades del SEN.
4. Módulo de sistemas de información del Sistema de Identificación y Caracterización de Oferta y Demanda Estadística del SEN – SICODE, orientado a presentar el módulo a expertos para recopilar observaciones y sugerencias que permitan optimizar su funcionamiento. Este componente es fundamental para consolidar la plataforma SICODE como herramienta estratégica para la gestión de la oferta y demanda estadística en el país.

3. RECOMENDACIONES

3.1 Proyectos de Ciencia de datos

El objetivo principal de esta línea de investigación consiste en aplicar técnicas y métodos avanzados de análisis para optimizar la recopilación, el procesamiento y la presentación de los datos estadísticos generados por el DANE. Esta orientación busca fortalecer la calidad, la

oportunidad y la interoperabilidad de la información, garantizando que los productos estadísticos respondan a las necesidades del Sistema Estadístico Nacional (SEN) y a los estándares internacionales.

En este contexto, y como parte del trabajo liderado por la Oficina de Sistemas del DANE, con el acompañamiento del (CASEN), se elaboró la Guía para el Proceso de Interoperabilidad. Este documento constituye un instrumento estratégico para facilitar la integración de sistemas y el intercambio seguro de datos entre las diferentes entidades productoras de información estadística, contribuyendo así a la modernización tecnológica y al fortalecimiento de la infraestructura del SEN.

A continuación, se presentan las recomendaciones realizadas por los expertos en los diferentes temas:

3.1.1 Modelo tecnológico para la producción de registros estadísticos base de unidades económicas y población

Las principales recomendaciones fueron las siguientes:

- Incluir una capa de observabilidad y monitoreo: implementar mecanismos de observabilidad y monitoreo que permitan detectar errores en tiempo real durante el procesamiento de datos. Esta práctica facilita la identificación temprana de fallos, reduce tiempos de respuesta y contribuye a la estabilidad del sistema.
- Optimizar el proceso de vinculación de datos mediante catálogos indexados: utilizar catálogos indexados y estructuras de búsqueda eficientes, como *hash tables* y algoritmos de búsqueda binaria, para mejorar el rendimiento en procesos de *data linkage*. Esto permitirá reducir la complejidad computacional y acelerar la integración de grandes volúmenes de información.
- Explorar el uso de GPU para procesamiento paralelo: evaluar la incorporación de unidades de procesamiento gráfico (GPU) para ejecutar operaciones en paralelo sobre grandes conjuntos de datos. Esta estrategia puede incrementar significativamente la velocidad de cálculo en tareas intensivas, como análisis estadístico y modelado.

- Distribuir catálogos en múltiples bases de datos (sharding): aplicar técnicas de *sharding* para distribuir catálogos en diferentes bases de datos, evitando cuellos de botella y mejorando la escalabilidad del sistema. Esta práctica es clave para garantizar un desempeño óptimo en entornos con alta demanda de consultas.
- Micro optimización del código en operaciones críticas: se resalta la importancia de realizar micro optimización en el código, especialmente en operaciones de comparación de cadenas, con el fin de reducir tiempos de ejecución y mejorar la eficiencia general del sistema.
- Fortalecer la gobernanza de datos con controles y auditoría: incorporar mecanismos de validación, control y auditoría dentro de la estructura de gobernanza de datos. Esto permitirá garantizar la integridad, trazabilidad y cumplimiento normativo en el manejo de información.
- Realizar pruebas funcionales con usuarios externos: llevar a cabo pruebas funcionales con usuarios externos para validar la experiencia de uso y la utilidad del sistema. Este enfoque asegura que las soluciones desarrolladas respondan a las necesidades reales de los actores del SEN.
- Explorar algoritmos de matching probabilístico: evaluar el uso de algoritmos de emparejamiento probabilístico, como el índice de Jaccard y *Propensity Score Matching*, para mejorar la precisión en procesos de vinculación de registros y reducir errores en la integración de datos.
- Considerar la virtualización de datos como estrategia de eficiencia: analizar la implementación de técnicas de virtualización de datos, orientadas a optimizar el acceso y reducir costos asociados al almacenamiento físico, sin comprometer la disponibilidad ni la seguridad de la información.
- Implementar estrategias avanzadas de búsqueda segmentada: explorar métodos como *Bicon Guide Search*, que segmentan los datos en subconjuntos para facilitar búsquedas más rápidas y eficientes, especialmente en entornos con grandes volúmenes de información.

3.1.2 APP Diversa

Las recomendaciones principales son:

Mejora de la experiencia de usuario en la encuesta

- Visibilidad de instrucciones: se recomienda que las indicaciones para responder las preguntas sean más claras y visibles. Por ejemplo, incluir asteriscos o mensajes explícitos que indiquen que una pregunta es obligatoria, evitando confusión y errores en el diligenciamiento.
- Tamaño y estilo del texto: aumentar el tamaño de los textos aclaratorios (como “seleccione una opción por fila”) y emplear negrita o colores contrastantes para garantizar que los usuarios no los pasen por alto.
- Validación anticipada: implementar validaciones durante el flujo de la encuesta, en lugar de realizarlas únicamente al final, para reducir frustraciones y prevenir el abandono del cuestionario.

Incentivos para aumentar la participación:

- Campañas con referidos en redes sociales: promover el uso de hashtags y menciones para que los usuarios inviten a otros a participar, generando mayor alcance orgánico.
- Rifas o sorteos simbólicos: aunque no se cuente con presupuesto significativo, se podrían ofrecer kits institucionales, libros o visitas guiadas al DANE como incentivos no monetarios.
- Eventos con figuras públicas: vincular personalidades reconocidas, como Brigitte Baptiste, en eventos de lanzamiento para aumentar el interés y la confianza en la encuesta.
- Alianzas con organizaciones: establecer acuerdos con programas como “Hola Colombia”, que han tenido experiencias exitosas ofreciendo incentivos a poblaciones específicas, como bonos de mercado para migrantes.
- Explorar sistemas de puntos: analizar la viabilidad de ofrecer puntos redimibles (por ejemplo, Puntos Colombia) como recompensa por la participación.

Articulación institucional: se destaca la importancia de establecer vínculos estratégicos con entidades públicas, académicas y organizaciones sociales para:

- Difusión en espacios confiables (Ministerio de Igualdad, ICBF, Defensoría del Pueblo, universidades, ONGs).
- Ofrecer rutas de atención a quienes reporten situaciones sensibles.
- Validar el contenido de las encuestas con expertos temáticos, garantizando pertinencia y enfoque inclusivo.

Inclusión y multilingüismo:

- Pruebas con comunidades indígenas y rurales: realizar pruebas de usabilidad con estas poblaciones, considerando sus particularidades en el uso de tecnología.
- Traducción a lenguas nativas: incorporar versiones en idiomas indígenas para evitar exclusión y garantizar representatividad.

Fortalecimiento de la narrativa comunicacional:

- Explicar el propósito de los datos: comunicar claramente por qué se solicita la información y cómo será utilizada, fortaleciendo la confianza.
- Construcción de país: desarrollar una narrativa que conecte la participación con el impacto en políticas públicas, inclusión y mejora de servicios.
- Evitar sesgos de autoselección: diversificar los canales de difusión para llegar a públicos que no usan medios digitales con frecuencia.
- Garantizar representatividad: mostrar que los datos no se emplean únicamente para estadísticas, sino para acciones concretas que beneficien a los participantes.

3.1.3 Modelo de evaluación de la calidad de bases de datos de operaciones estadísticas

Las principales recomendaciones son:

- Automatización del proceso: se observó que el flujo actual es altamente manual, lo que genera ineficiencias y riesgos de error. Se propuso automatizar la validación y la generación de reportes directamente sobre la base de datos, eliminando la dependencia de archivos intermedios en Excel y reduciendo pasos redundantes.
- Adopción de metodologías modernas para análisis de calidad de datos: se sugirió explorar herramientas y enfoques especializados, como: CDQ (Comprehensive Data Quality) como marco metodológico para evaluar y mejorar la calidad de datos o lenguajes declarativos como DataLog, que permiten definir reglas lógicas y validarlas sobre conjuntos de datos de manera eficiente.
- Procesamiento paralelo para grandes volúmenes de datos: se recomendó investigar soluciones que soporten procesamiento distribuido y paralelo, mencionando tecnologías como: Apache Hadoop y su ecosistema para almacenamiento y procesamiento masivo, Apache Spark, incluyendo librerías específicas para calidad de datos o Databricks y frameworks en Python como DQX, orientados al análisis avanzado de calidad. Estas alternativas permitirían mejorar la eficiencia y escalabilidad, reduciendo tiempos de ejecución y optimizando recursos.

- Migración hacia plataformas modernas: se coincidió en que Pentaho está rezagado frente a herramientas actuales para Big Data, lo que limita el rendimiento y la capacidad de respuesta. Se recomendó migrar hacia soluciones más robustas, como: Apache Spark para procesamiento distribuido o servicios en la nube (AWS) para escalabilidad y gestión eficiente de grandes volúmenes de datos.
- Medidas transitorias ante limitaciones de migración: se señaló que, si la migración no es inmediata, se debe aplicar partición de datos (sharding) para evitar represamientos y reducir riesgos en procesos prolongados (que pueden tardar más de 12 horas).
- Fortalecimiento de la gobernanza de datos: se sugirió incorporar buenas prácticas, incluyendo: indicadores de seguimiento por nivel (estratégico, táctico y operativo), matrices de riesgo y semáforos para monitoreo en tiempo real y protocolos de seguridad alineados con estándares internacionales.
- Comunicación orientada a la toma de decisiones: se indicó que los resultados deben presentarse de forma más visual y narrativa, mostrando el impacto en costos, eficiencia y riesgos, para facilitar la comprensión por parte de los tomadores de decisiones.
- Control de riesgos por procesos manuales: se recalcó que mantener procesos manuales incrementa riesgos de sesgos y errores, por lo que se deben aplicar controles y mecanismos de mitigación.

3.2 Desarrollo de Sistemas de Información

Esta línea se centra en fortalecer la arquitectura y seguridad de los sistemas de información del Sistema Estadístico Nacional (SEN) mediante el diseño e implementación de lineamientos técnicos y metodológicos que garanticen la interoperabilidad, el desarrollo seguro y la eficiencia en la construcción de sistemas estadísticos. Esta línea busca consolidar herramientas como la Guía para el Proceso de Interoperabilidad, la Guía para la Construcción de un Sistema de Información Estadística y la presentación del módulo de sistemas de información del SICODE, asegurando que las soluciones tecnológicas cumplan con estándares de calidad, protección de datos y escalabilidad, en beneficio de la modernización y sostenibilidad del ecosistema estadístico nacional en emplear técnicas y métodos de análisis avanzados para mejorar la recopilación, el procesamiento y la presentación de datos estadísticos producidos desde el DANE.

Frente a los diferentes temas las recomendaciones fueron las siguientes:

3.2.1 Guía para el proceso de interoperabilidad

Este tema fue revisado en 2024 y, posteriormente, en 2025 se llevó a cabo una reunión para evaluar el documento actualizado conforme a las recomendaciones de los expertos. Como comentarios finales, los especialistas sugirieron incorporar aspectos relacionados con ciberseguridad, gestión de identidades y protocolos de protección frente a vulnerabilidades.

3.2.2 Procedimiento para ingreso seguro mediante Single Sign-On (SSO)

Las principales recomendaciones son:

- Integración con autenticación multifactor (MFA): se discutió la necesidad de que el proceso de login contemple métodos adicionales de autenticación, como MFA, además de usuario y contraseña.
- Uso de protocolos estándar para autenticación: se mencionó la implementación de protocolos como OAuth y OpenID Connect para permitir integración con diferentes frontends, incluso móviles.
- Principio de mínimo privilegio en la asignación de roles y permisos: se destacó la importancia de evitar excesos en accesos mediante la aplicación del principio de mínimo privilegio.
- Definición de “reinos” o dominios para control de roles y permisos: se sugirió segmentar roles y permisos en reinos particulares para mayor seguridad.
- Administración de usuarios y roles basada en matrices definidas por el negocio: se indicó que la gestión debe depender de matrices configuradas en el SSO según los requerimientos del negocio.
- Aclarar el alcance del procedimiento: Especificar que aplica principalmente a nuevos sistemas desarrollados por los grupos SIPA y AIOC, pero con posibilidad de incluir sistemas heredados progresivamente e indicar que no se limita solo a encuestas, sino que está abierto a cualquier sistema que requiera acceso seguro.
- Modificar términos para evitar confusión: Cambiar el término “administrador de la encuesta” por “administrador del sistema” o uno más general, para reflejar que aplica a sistemas más allá de encuestas.
- Ajustar el objetivo y el nombre del procedimiento: el objetivo debe aclarar que el procedimiento trata sobre la implementación de autenticación basada en SSO para sistemas nuevos, se sugiere que el nombre sea más específico: “*Procedimiento para habilitar/implementar ingreso seguro mediante Single Sign-On*” e indicar que contribuye a controles preventivos, pero no los abarca todos.

- Evitar ambigüedades en auditorías: ajustar semánticamente el documento para mayor claridad y evitar interpretaciones erróneas y realizar un análisis para mejorar la estructura y la redacción del procedimiento.
- Considerar otros controles de seguridad en la documentación: aunque el procedimiento se centra en autenticación y software, se debe reconocer que el ingreso seguro también involucra controles como: hardening de hardware y red, sistema operativo, certificados SSL/TLS, autenticación multifactor y aclarar que estos controles son gestionados por otras áreas, pero mencionarlos como parte de la estrategia de defensa en profundidad.
- Incluir en el procedimiento la gestión de actualizaciones y vulnerabilidades del SSO (Keycloak): incorporar una sección que indique qué hacer ante: nuevas versiones de Keycloak, vulnerabilidades reportadas (CVE) y ajustes necesarios para mantener la seguridad.
- Hacer obligatoria la activación de controles de auditoría automáticos: ajustar la redacción para que no se interprete como opcional, especificar que los registros automáticos (inicio/cierre de sesión, intentos fallidos, acceso por IP, etc.) deben activarse por defecto, salvo excepciones justificadas y modificar la nota para indicar que se activan automáticamente según lo habilite el negocio o arquitectura, no solo a demanda.
- Aclarar la evidencia en la actualización del caso en la mesa de servicios: definir claramente qué se debe registrar como evidencia: el resultado de la validación del servicio probado y la documentación en el formato de pruebas del programador.

3.2.3 Guía para la construcción de un sistema de información estadística

Las principales recomendaciones son:

- Asignación de roles en el componente organizacional: se observó que el documento se enfoca únicamente en roles relacionados con el manejo de datos, dejando por fuera otros roles clave en el gobierno del sistema como tal (mantenimiento, operación, desarrollo), se sugirió ampliar la definición para incluir roles técnicos, administrativos y estratégicos que garanticen el funcionamiento integral del sistema.

- Alineación con objetivos institucionales: se recomendó incluir explícitamente cómo el sistema de información se alinea con la misión, visión y estrategia de la organización, esto evitaría que el sistema se perciba como una iniciativa aislada sin conexión con las metas institucionales.
- Gestión del cambio: se propuso incorporar un apartado sobre gestión del cambio, dado que los sistemas evolucionan con el tiempo, esto incluye adaptación a nuevas políticas, tecnologías, necesidades de usuarios y roles institucionales.
- Atributos de calidad tecnológica: se sugirió incluir atributos como: alta disponibilidad, escalabilidad, desempeño, accesibilidad y experiencia de usuario. Estos son esenciales para sistemas estadísticos que manejan grandes volúmenes de datos y deben ser amigables para usuarios no técnicos.
- Arquitectura empresarial: se recomendó alinear el documento con el marco de arquitectura empresarial promovido por MinTIC. Esto incluye dominios como arquitectura de datos, seguridad, infraestructura y soluciones.
- Claridad sobre el desarrollo del sistema: se observó que el texto da la impresión de que el sistema se basa solo en software existente, se sugiere aclarar que también puede incluir desarrollos a medida o híbridos.
- Normativa tecnológica: se propuso incluir referencias a guías, estándares y resoluciones nacionales e internacionales que regulan el desarrollo de sistemas tecnológicos.
- Guía de ciencia de datos del DANE: Se recomendó articular el documento con esta guía, ya que ofrece un marco metodológico útil y reconocido.
- Falta de enfoque técnico: Se señaló que el documento tiene un enfoque muy general y carece de elementos propios de la ingeniería de software (ciclos de desarrollo, pruebas, documentación técnica).
- Bill of Materials (BoM): se sugirió incluir esta práctica como parte de la trazabilidad tecnológica, dado que Permite identificar componentes reutilizados, licencias, y origen del software para facilitar mantenimiento y seguridad.

- **Gobernanza de datos:** se indicó que el documento no desglosa suficientemente la gobernanza en niveles estratégico, táctico y operativo, se recomendó definir funciones, roles, responsabilidades y alcances en cada nivel.
- **Interoperabilidad:** se sugirió incluir estándares como ISO, EI, SDMX para garantizar compatibilidad y eficiencia.
- **Seguridad de la información:** se Recomendó incorporar normativas como ISO 27001 y 27032, estas regulan el diseño, manejo y protección de datos en todas las fases del sistema.
- **Monitoreo y evaluación:** se propuso establecer protocolos y manuales para seguimiento y evaluación continua del sistema; esto incluye indicadores de uso, funcionalidad y seguridad.
- **Actualización tecnológica:** se sugirió definir periodicidad y responsables para la actualización de componentes tecnológicos y de seguridad (contraseñas, accesos, hash, etc.).
- **Gestión automática de metadatos:** se mencionaron herramientas basadas en inteligencia artificial como Parsley para la actualización automática de metadatos; se señaló que este tipo de soluciones podrían explorarse en la literatura sobre bases de datos y data warehouses.
- **Fortalecimiento desde la arquitectura empresarial:** se propuso orientar la guía hacia la identificación de requerimientos funcionales y no funcionales, esto facilitaría decisiones sobre desarrollo interno, contratación externa o adopción de plataformas.
- **Articulación con arquitecturas de solución:** se sugirió que cada entidad pueda adaptar una arquitectura de referencia a su contexto tecnológico, esto permitiría mayor flexibilidad sin perder coherencia metodológica.
- **Integración con iniciativas existentes:** se recomendó vincular el documento con otras guías del DANE y del SEN, como la de ciencia de datos y la APP Diversa.

3.2.4 Módulo de sistemas de información del Sistema de Identificación y Caracterización de Oferta y Demanda Estadística del SEN – SICODE

Las principales recomendaciones son:

- Se observó que el proceso actual es altamente manual, lo que genera ineficiencias y riesgos de error, se propuso automatizar la validación y generación de reportes directamente sobre la base de datos, evitando la dependencia de archivos intermedios en Excel y reduciendo pasos redundantes.
- Se sugirió explorar herramientas y metodologías modernas para análisis de calidad de datos como: CDQ (Comprehensive Data Quality) como marco metodológico y lenguajes especializados como DataLog, que permiten definir reglas lógicas y validarlas sobre conjuntos de datos de manera eficiente.
- Se recomendó investigar soluciones que soporten procesamiento paralelo para grandes volúmenes de datos, mencionando tecnologías como: Apache Hadoop y su ecosistema, Apache Spark (incluyendo librerías para calidad de datos) o Herramientas basadas en Databricks y frameworks en Python como DQX para análisis de calidad. Se enfatizó que estas alternativas podrían mejorar la eficiencia y escalabilidad, reduciendo tiempos de ejecución y optimizando recursos.
- El uso de Pentaho está rezagado frente a herramientas modernas para Big Data, y que su uso prolongado limita el rendimiento y la capacidad de respuesta: se recomendó migrar hacia plataformas más robustas como: Apache Spark para procesamiento distribuido, servicios en la nube como AWS (Amazon Web Services) para escalabilidad y gestión eficiente de grandes volúmenes. Se indicó que, si no es posible migrar de inmediato, se debe aplicar partición de datos para evitar represamientos y reducir riesgos en procesos largos (que pueden tardar más de 12 horas).
- Se sugirió incorporar buenas prácticas de gobernanza de datos, incluyendo: indicadores de seguimiento por nivel (estratégico, táctico y operativo), matrices de riesgo y semáforos para monitorear problemas en tiempo real y protocolos de seguridad de la información alineados con estándares internacionales.
- Se indicó que la comunicación de resultados debe ser más visual y narrativa, orientada a tomadores de decisiones, mostrando impacto en costos, eficiencia y riesgos.
- Se recalcó que mantener procesos manuales incrementa riesgos de sesgos y errores, por lo que se deben aplicar controles para mitigarlos.

4. PRODUCTOS DE LA SALA

Las recomendaciones dadas por los expertos de la sala se vieron reflejadas en la Guía para el proceso de interoperabilidad, en el presente documento, así como las actas resultantes de cada reunión.