

1. OBJETIVO

Establecer acciones que permitan gestionar, controlar y mejorar la calidad de los datos de la entidad, a partir de la definición e implementación de las reglas de calidad, con el fin de asegurar que los datos sean confiables, precisos, completos y consistentes para la toma de decisiones.

2. DOCUMENTOS RELACIONADOS

Tipo de documento	Código	Título	Modo de uso	Clasificación documento
Código	CG-PEC-0001	Código de buen gobierno – numeral 4.	Digital	Interno
Manual	MN-IIT-0072	Manual de Políticas y Lineamientos de Seguridad de la Información	Digital	Interno
Manual	MN-IIT-0062	Manual de Protección de Datos Personales	Digital	Interno
Instructivo	IN-IIT-0315	Gestión de dominios de datos.	Digital	Interno
Formato	FT-IIT-2879	Libro de dominios	Digital	Interno

3. DEFINICIONES Y SIGLAS

- **Calidad de datos:** se define como el grado en el que un dato cumple con los requisitos definidos por la entidad mediante las reglas de negocio y de calidad. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Causa raíz:** se refiere a la causa origen o fundamental de un problema o hallazgo relacionado con los datos de la entidad. Es la causa principal que ha llevado a la existencia de un problema o defecto en la calidad, integridad o confiabilidad de los datos. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Ciclo de vida del dato:** conjunto de etapas que cubren la creación u obtención de los datos, su transformación, uso, almacenamiento y disposición final, de acuerdo con su clasificación. Fuente: Circular número 000004 – 31 de marzo de 2022.
- **DGIT:** sigla de Dirección de Gestión de Innovación y Tecnología.
- **Dimensión de calidad:** es una característica medible de los datos, proporcionan una visión holística y completa de la calidad de los datos. Las dimensiones para evaluar en la DIAN son: *completitud, unicidad, validez, consistencia, precisión, integridad, oportunidad, razonabilidad e integridad*. Para la definición de las dimensiones de calidad que puede gestionar la DIAN se analizó la ISO/IEC 25012 - Data Quality Model. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Dominio de datos:** “conjunto de datos lógicamente agrupados que responden a las necesidades estratégicas, de negocio o de soporte de la entidad, y permiten ser administrados. Fuente: documento

con los dominios de información del proyecto “Diseño e implementación del Modelo de Gobierno de Datos para la Dirección de Impuestos y Aduanas Nacionales – DIAN”.

- **KDE (Key data element):** son los elementos (atributos) claves en el dominio de datos, que hace que el objetivo de la entidad para la cual fue creado se pueda cumplir, por lo tanto, requiere una calidad apropiada, un almacenamiento único (evitar duplicar los datos) y gestión estandarizada, con el fin de convertirse en una fuente de aprovisionamiento. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Plan de remediación:** es un conjunto de acciones y estrategias diseñadas para abordar y corregir los hallazgos identificados en el perfilamiento de datos. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Perfilamiento de datos:** es un proceso mediante el cual se realiza un análisis de los datos para identificar problemas de calidad y evaluar el nivel de cumplimiento de las reglas de calidad respecto a criterios como validez, consistencia y unicidad, entre otros. Teniendo en cuenta las necesidades de la entidad se pueden utilizar diferentes variables. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Programa de calidad de datos:** es un proceso gradual y sistemático, diseñado para establecer y mantener un orden en todos los dominios de datos de la entidad. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN – EY.
- **Reglas de calidad:** representan las expectativas de la calidad de datos, es decir, describen como deben ser los datos para que estos sean operables y utilizables, además deben contar con la definición de los umbrales y la dimensión que aplica para validar la regla de calidad. Fuente: adaptado de DAMA – DMBOK 2 Edition 2017.
- **Reglas de negocio:** es una declaración lógica o restricción que define o limita las acciones o comportamientos de los datos en las actividades de la entidad. Fuente: Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.
- **Repositorio o contenedor de datos:** son estructuras de almacenamiento las cuales se encuentran diseñadas para almacenar, gestionar y organizar conjuntos de datos de manera centralizada. Fuente: adaptado de Dama Internacional (2017). *DAMA DMBOK – Data Management body of Knowledge* [Versión 2.]. Editorial New Jersey. Roles y responsabilidades frente a los datos: establecidos mediante Resolución 00002 de 2024.
- **Umbral de calidad de datos:** establece el nivel aceptable de calidad que se debe alcanzar para considerar los datos como confiables y aptos para su uso. Fuente: Programa de calidad de datos – Consultoría Diseño e Implementación del Modelo de Gobierno de Datos para la DIAN - EY.

4. CALIDAD DE DATOS

4.1. Creación del proyecto de calidad de datos

Las tareas de calidad de datos se desarrollan a partir de los siguientes eventos:

- En desarrollo de las acciones de administración de dominios de datos a cargo de la Subdirección de Información y Analítica. De acuerdo con el instructivo IN-IIT-0315 Gestión de dominios de datos.
- Por solicitud de un líder de dominio.
- Durante la ejecución de proyectos de migración o integración de datos.
- Cuando se presenten cambios en las directrices normativas que impliquen la actualización o creación de reglas de calidad de los datos del respectivo dominio.
- Por solicitud de un área que explote y/o utilice los datos y encuentre condiciones inesperadas o inaceptables.

La viabilidad del proyecto de calidad de datos se evalúa por parte del líder de gestión de la gobernanza de datos a partir de los siguientes criterios:

- Propiedad de los datos: se verifica que el dominio de datos a gestionar su calidad tenga el visto bueno o aprobación de la dependencia dueña de los datos.
- Disponibilidad de la información: los datos del dominio deben estar disponibles en el repositorio único de datos.
- Disponibilidad de recursos: se valida la disponibilidad de recursos humanos y técnicos para atender la solicitud.

El proyecto de calidad de datos será ejecutado por el equipo de responsables del respectivo libro de dominio, de acuerdo con los roles registrados en el inventario de dominio de datos de la entidad.

El equipo elabora un plan de trabajo en el cual se establecen estrategias, actividades, condiciones y estándares que los datos deben cumplir para ser considerados de alta calidad. También se incluye el cronograma de ejecución y se designan las responsabilidades correspondientes.

El plan de trabajo debe ser aprobado por el líder de dominio.

4.2. Ejecución de las tareas de calidad de datos

4.2.1. Descubrimiento

- Identificación de la fuente de datos a intervenir con la participación del custodio de datos y el delegado de la Subdirección de Procesamiento de Datos.
- Validación del inventario de atributos del respectivo dominio a intervenir.
- Selección de los KDE a intervenir su calidad de datos.

"Se considera copia controlada los documentos publicados en el LMD"

- Definir, en conjunto con el custodio de datos, las reglas de calidad asociadas a los atributos a intervenir, así como los umbrales definidos para cada dimensión.

Las reglas de calidad se definen teniendo en cuenta los siguientes criterios:

- **Relevancia:** ¿Qué tanto contribuye la regla a los objetivos del negocio?
- **Impacto:** ¿Cuán significativamente la regla puede mejorar la precisión, consistencia e integridad de los datos?
- **Factibilidad:** Qué tan práctica y realista es la implementación de la regla correspondiente.

La definición y su evaluación se encuentra en la hoja del libro de dominios denominada “Calidad de datos”.

4.2.2. Diagnóstico

- Seleccionar, con el custodio, el conjunto de datos del dominio a los que se les van a aplicar las reglas de calidad para la ejecución del perfilamiento de los datos. Las pruebas funcionales se realizan sobre una muestra de los datos. Excepcionalmente, se realizará la prueba sobre el 100% de los datos dependiendo del volumen de información.
- Configurar las reglas de calidad en las herramientas de gobernanza de datos (DataQuality-Databricks y/o MS Purview). Esto implica establecer los parámetros de cada regla y ajustar la configuración de la herramienta para garantizar una evaluación eficiente y efectiva.
- Con base en la anterior configuración, el gestor de gobernanza de datos realiza pruebas funcionales y verifica que las reglas cumplan con los requisitos establecidos y funcionen de manera adecuada.
- Ejecutar el perfilamiento de los datos una vez configuradas las reglas de calidad en la herramienta seleccionada, para esto es necesario examinar y analizar los datos de manera sistemática para obtener información detallada sobre su estructura, contenido, calidad y patrones. NOTA: Cuando el tamaño de la muestra no pueda ser evaluado por el gestor de gobierno de datos se solicitará el perfilamiento a la Subdirección de Procesamiento de Datos.
- Una vez finalizado el perfilamiento, generar el informe para el líder de dominio y custodio, en el que se detallen los siguientes elementos:
 - ✓ Resumen de los hallazgos más importantes del perfilamiento de datos.
 - ✓ Descripción detallada de los datos que se han perfilado, incluyendo su origen, formato, volumen, y cualquier otra característica relevante.
 - ✓ Detalles de los resultados obtenidos del perfilamiento de datos, incluyendo estadísticas descriptivas, anomalías detectadas, valores atípicos, inconsistencias, duplicados, y cualquier otro hallazgo relevante.

4.2.3. Remediación

- Con base en el informe de perfilamiento se validan los hallazgos en conjunto con el líder de dominio. De ser necesario se realizan los ajustes correspondientes del perfilamiento para la aceptación del líder de dominio.
- De acuerdo con los resultados del perfilamiento se prioriza la gestión de los hallazgos para determinar la respectiva causa raíz de acuerdo con las definiciones dictadas por el líder de dominio.
- En conjunto con el custodio de datos y el líder técnico de calidad de datos se realiza el análisis de causa raíz para identificar los motivos por los cuales se generaron los hallazgos priorizados y de esta manera definir el plan de remediación.
- Elaborar el plan de remediación con el fin de mejorar la calidad de los datos y abordar los problemas identificados durante el perfilamiento o la evaluación. Estos planes de remediación podrían considerar acciones de enriquecimiento dependiendo de cada caso, considerando los siguientes aspectos:
 - ✓ Eventual ajuste del sistema transaccional fuente del dato con incumplimiento de regla de calidad. Lo anterior requiere solicitud de ajuste elevada por el área funcional a la Subdirección y Soluciones y Desarrollo.
 - ✓ Determinación de la fecha a partir de la cual los registros en el repositorio único de datos van a ser ajustados en la capa plata de acuerdo con la regla de calidad definida por el negocio. Nota: Los registros a los que se les aplique la regla de remediación deberán quedar identificados (marcados) para efectos de trazabilidad y control
 - ✓ Para los registros anteriores a la precitada fecha, el negocio deberá establecer la fecha histórica a partir de la cual se aplicaría la regla de remediación o si los registros se conservan con la inconsistencia detectada. En ambos casos, la Subdirección de Procesamiento de Datos debe garantizar que los registros a los que se aplica la regla de remediación queden identificados (marcados) para efectos de trazabilidad y control.
- Antes de su implementación, el plan de remediación debe ser aprobado por el dueño de los datos, previo visto bueno del líder de dominio.
- El dueño de los datos solicita a la DGIT la implementación del plan de remediación de acuerdo con el procedimiento de habilitación de cambios de TI (PR-IIT-0457).
- Dependiendo de la causa raíz, la ejecución del plan de remediación puede requerir la intervención de varias subdirecciones de la DGIT, aspecto que coordina el líder técnico de gobernanza de datos. En la medida en que avance el proceso de implementación, el líder técnico de gobernanza de datos generará informes de seguimiento para el líder de dominio y custodio, y mantendrá informado a los gestores de gobierno de datos para su monitoreo y control.

4.2.4. Monitoreo

- El gestor de gobierno de datos en conjunto con el custodio o líder de dominio, realizan el monitoreo de la ejecución de los planes de remediación, de manera que se pueda determinar su efectividad. Para este fin se debe:
 - ✓ Establecer la línea base de comparación, tomando el perfilamiento inicial.
 - ✓ Evaluar periódicamente los resultados del perfilamiento ejecutado para determinar el avance de las acciones.
 - ✓ Identificar y corregir problemas o desafíos que se presenten durante la implementación de las acciones del plan de remediación.
- El gestor de gobierno de datos ejecuta periódicamente las reglas de calidad en la herramienta seleccionada por la entidad para evaluar la efectividad del plan de remediación y los niveles de cumplimiento de calidad de los datos.
- Presentar al líder de dominio y al dueño de dato un informe con los resultados del proyecto de calidad de datos, indicando el nivel de calidad inicial y el resultante después de la implementación de los planes de remediación. Las principales acciones y los resultados de la aplicación del plan de remediación se registran en la herramienta de gobierno de datos y en la hoja “Matriz de Remediación” del libro de dominio correspondiente.

5. CONTROL DE CAMBIOS

Versión	Vigencia		Descripción de los cambios	Tipo de información
	Desde	Hasta		
1	30/06/2026		Versión inicial.	Información pública

Elaboró:	Edicson Alexander Rojas Gamba Ramsés Alfonso Cruz Bohórquez Técnica	Inspector Gestor	Subdirección de Información y Analítica
Revisó:	Ginna Zabransky Rodríguez Metodológica	Inspector	Subdirección de Procesos
	Pedro Alberto Camargo Casas Técnica	Inspector	Subdirección de Información y Analítica
Aprobó:	Jesús Daniel Cañizales	Subdirector	Subdirección de Información y Analítica