

1. OBJETIVO

Establecer las actividades necesarias para la validación previa de tablas, documentación, validación, aprobación, publicación en el catálogo y repositorio y administración de las Listas de Referencia (LR), con el fin de cumplir con los lineamientos definidos por el Modelo de Gobernanza de Datos.

2. DOCUMENTOS RELACIONADOS

Tipo de documento	Código	Título	Modo de uso	Clasificación documento
Procedimiento	PR-IIT-0472	Gestión de solicitudes y proyectos de analítica de datos	Digital	Interno
Resolución	Resolución 00002 de 2024	Por la cual se adicionan los numerales 8 y 9 al artículo 17 de la Resolución Número 000021 del 28 de enero de 2022 y se establece el Modelo de Gobernanza de Procesos y el Modelo de Gobernanza de Datos en la Unidad Administrativa Especial Dirección de Impuestos y Aduanas Nacionales – UAE DIAN	Digital	Interno
Circular	Circular 000004 de 2022	Consideraciones generales sobre la gobernanza de datos en la Entidad	Digital	Interno
Instructivo	IN-IIT-0315	Gestión de Dominios de Datos	Digital	Interno
Formato	FT-IIT-2880	Inventario de las listas de referencia	Digital	Interno
Formato	FT-IIT-2881	Caracterización de la lista de referencia	Digital	Interno

3. DEFINICIONES Y SIGLAS

- **Catálogo de Listas de Referencia:** repositorio institucional dispuesto para el almacenamiento, administración, consulta y control de las Listas de Referencia autorizadas por la entidad. Fuente: Lineamientos para la Gestión de Listas de Referencia – DIAN V.1.
- **Custodio de Datos:** funcionario responsable de la administración operativa de los dominios de datos, garantizando su disponibilidad, integridad y seguridad conforme a los lineamientos institucionales. Fuente: Resolución 000002 del 24 enero 2024 - Modelo de Gobernanza de Datos DIAN.
- **Dominio de Datos:** conjunto de datos lógicamente agrupados cuyo significado está asociado a un proceso y a un concepto de negocio. Fuente: Resolución 000002 del 24 enero 2024 - Modelo de Gobernanza de Datos DIAN.

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

- **Dueño del Dato:** responsable funcional del dominio de datos, encargado de definir y aprobar criterios, contenido y necesidades asociadas a su administración. Fuente: Resolución 000002 del 24 enero 2024 - Modelo de Gobernanza de Datos DIAN.
- **Gestor de Gobernanza de Datos:** funcionario encargado de apoyar la identificación, administración, actualización y control de los dominios de datos conforme al Modelo de Gobernanza de Datos. Fuente: Resolución 000002 del 24 enero 2024 - Modelo de Gobernanza de Datos DIAN.
- **Lista de Referencia (LR):** dominio de datos constituido por un conjunto finito de valores definidos y clasificados, utilizado para facilitar la clasificación o codificación de la información y administrado conforme al Modelo de Gobernanza de Datos. Fuente: Resolución 000002 del 24 enero 2024 - Modelo de Gobernanza de Datos DIAN.
- **Modelo de gobernanza de datos:** Conjunto de reglas que gobiernan el ciclo de vida y el flujo de datos, de acuerdo con su tipología con el fin de garantizar su calidad, uso, titularidad, compartición, seguridad y supresión, y que están orientadas a la generación de valor, minimizando los riesgos asociados a ello. Fuente: Resolución 000002 del 24 enero 2024 - Modelo de Gobernanza de Datos 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: DIAN (2023) Programa de calidad de datos del proyecto “Diseño e implementación del Modelo de Gobierno de Datos para la Dirección de Impuestos y Aduanas Nacionales – DIAN”.
- **PAMD:** Programa de Apoyo a Modernización de la Dian.
- **TABLA:** Lista o catálogo de cosas puestas por orden sucesivo o relacionadas entre sí. Fuente RAE. Para el caso interno del gobierno de datos corresponde a listas de valores, sin un proceso de estandarización, utilizadas por los sistemas de información.
- **Líder de dominio:** es el funcionario público delegado por el dueño del dato, experto en el conjunto de datos que conforman un dominio y que tiene responsabilidad de su administración y control. Asumen este rol, los subdirectores y/o los funcionarios públicos de nivel profesional, designados por éstos, de acuerdo con su ficha descriptiva del empleo. Fuente: Resolución DIAN No. 002 de 2024.
- **Vigencia:** periodo durante el cual un valor contenido en una Lista de Referencia se considera válido para su uso institucional. Fuente: Lineamientos para la Gestión de Listas de Referencia – DIAN V.1.
- **Metadatos:** son datos sobre los datos y articulan un contexto para determinados objetos de interés (recursos), en forma de descripción de recursos. (Adaptado MINTIC – G.INF.06 – Guía Técnica de Información – Gobierno del dato).

- **SIA:** Subdirección de Información y Analítica.

4. DESARROLLO DEL TEMA

El presente capítulo establece el conjunto de actividades, roles y lineamientos necesarios para la gestión integral de las LR en el marco del Modelo de Gobernanza de Datos. Estas listas constituyen un elemento fundamental para asegurar la estandarización, calidad y reutilización de la información, permitiendo la correcta clasificación y codificación de los datos utilizados en los diferentes sistemas de la entidad.

En este sentido, el proceso inicia con una fase de validación previa orientada a la identificación de posibles LR, a partir del análisis de tablas y atributos provenientes de distintos escenarios como las mesas de trabajo dentro del marco del PAMD, ejercicios de calidad de datos y caracterización de dominios de datos y actividades propias de la operación y monitoreo del modelo de gobernanza. Posteriormente, se define un flujo estructurado que abarca la validación de fuentes de información y la determinación de estándares aplicables, entendiendo el estándar como una regla, especificación o definición formalmente aprobada por una autoridad competente, interna o externa, que establece la forma uniforme en que un dato debe ser representado, clasificado, codificado o utilizado dentro de la organización. Los estándares constituyen la base para la creación y administración de las listas de referencia y demás mecanismos de control de datos garantizando la coherencia y alineación con las necesidades del negocio y la normativa vigente.

Finalmente, el capítulo desarrolla las etapas de documentación, validación, aprobación, publicación e Inventario de las LR, incluyendo su mantenimiento y control de cambios a lo largo de su ciclo de vida. De esta manera, se asegura no solo la disponibilidad y trazabilidad de las listas, sino también su adecuada gestión como activos de datos estratégicos para la entidad

4.1 Validación previa de las tablas

La identificación de (LR) corresponde al proceso mediante el cual se identifican conjuntos de valores controlados utilizados para clasificar o codificar información dentro de los sistemas de la entidad. Esta actividad es fundamental para asegurar la estandarización y reutilización de datos. Para tal actividad realizamos lo siguientes pasos:

4.1.1 Identificación de listas:

El Gestor de Gobernanza de Datos realiza la identificación de las listas, la cual, se puede presentar en tres escenarios, los cuales no son excluyentes: mesas de trabajo dentro del marco del PAMD, ejercicios de calidad de datos y caracterización de dominios de datos. A continuación, se explican dichos escenarios:

a. Mesas de trabajo dentro del marco del PAMD - en el marco de los ejercicios de PAMD, se realizan mesas de trabajo con las áreas de negocio y técnicas donde se identifican tablas o atributos que contienen valores repetitivos, categóricos o codificados. Durante este apartado se validan necesidades de estandarización, el uso de las tablas y la fuente u origen de la información dispuesta en la tabla.

b. Ejercicios de calidad de datos - durante la aplicación de las reglas de calidad o el monitoreo de los Key Data Element (KDE), se identifican campos con valores no homologados, inconsistentes o abiertos que pueden ser normalizados mediante una LR.

c. Caracterización de dominios de datos - durante el ejercicio se identifica con los usuarios, las posibles LR utilizadas o que hacen parte del dominio. Posterior el Gestor de Gobernanza de Datos establece si el conjunto de datos, de la posible LR, cumple con lo siguiente:

- Son un conjunto finito de valores.
- Están definidos y clasificados.
- Cuentan con una estructura definida
- Están avalados por una entidad que reglamenta, controla o supervisa la temática.
- Tienen vigencia en un periodo de tiempo dependiendo de su naturaleza.
- No cambian con las operaciones o transacciones en los que son utilizados.

En caso de cumplir con la totalidad de condiciones continuamos con el apartado 4.1.2

4.1.2 Validación de la fuente de información y determinación del estándar aplicable:

La validación de la fuente de información y determinación del estándar se realiza atendiendo el escenario del cual se identificó la tabla:

- a.** Mesas de trabajo dentro del marco del PAMD – el Gestor de Gobernanza de Datos realiza el análisis en conjunto con el delegado designado del módulo que se encuentra en verificación dentro del PAMD, con el propósito de identificar la fuente de origen de los valores contenidos en la tabla evaluada. En caso de que la fuente corresponda a un estándar previamente definido, se procederá con la documentación de la LR conforme a lo establecido en el numeral 4.2. Si la fuente no corresponde a un estándar identificado, el Gestor de Gobernanza de Datos, en primera instancia, realizará la búsqueda de un posible estándar aplicable, el cual será validado con el delegado del módulo. En caso de no encontrar un estándar, o que este no sea aplicable o no sea validado por el delegado, se escalará a una validación de segundo nivel con el apoyo del equipo de Gobernanza de Datos, con el fin de identificar alternativas de estandarización y someterlas nuevamente a validación. Finalmente, si no se logra identificar un estándar aplicable o no se obtiene validación por parte del delegado, se debe realizar una mesa de trabajo conjunta entre el equipo de

Gobernanza de Datos y el equipo del módulo en evaluación, con el propósito de definir y acordar el estándar que debe aplicarse a esta LR. Posterior se procederá con la documentación de la LR conforme a lo establecido en el numeral 4.2

- b. Ejercicios de calidad de datos - el Gestor de Gobernanza de Datos realiza el análisis en conjunto con el Custodio de Datos delegado por el Líder de Dominio con el propósito de identificar la fuente de origen de los valores en la tabla evaluada en el marco del dominio a verificar. En caso de que la fuente corresponda a un estándar previamente definido, se procederá con la documentación de la LR conforme a lo establecido en el numeral 4.2. Si la fuente no corresponde a un estándar identificado, el Gestor de Gobernanza de Datos, en primera instancia, realizará la búsqueda de un posible estándar aplicable, el cual será validado con el Custodio de Datos. En caso de no encontrar un estándar, o que este no sea aplicable o no sea validado por Custodio de Datos, se escalará a una validación de segundo nivel con el apoyo del equipo de Gobernanza de Datos, con el fin de identificar alternativas de estandarización y someterlas nuevamente a validación. Finalmente, si no se logra identificar un estándar aplicable o no se obtiene validación por parte del Custodio de Datos, se recomienda realizar una mesa de trabajo conjunta entre el equipo de Gobernanza de Datos y el Líder de Dominio junto con el Custodio de Datos, con el propósito de definir y acordar el estándar que debe ser adoptado. Posterior se procederá con la documentación de la LR conforme a lo establecido en el numeral 4.2
- c. Caracterización de dominios de datos - el Gestor de Gobernanza de Datos y el Custodio de Datos identifican cuales conjuntos de datos son susceptibles a constituirse como una LR. En caso de que las tablas correspondan una LR, se procederá con la documentación de la LR conforme a lo establecido en el numeral 4.2.

4.2 Documentación de la LR

Para la documentación de las LR, deben realizarse lo siguientes pasos:

4.2.1 Registro en el formato de Inventario de las listas de referencia

Una vez el Gestor de Gobernanza de Datos realiza identifica que la tabla corresponde a una la LR, debe ser registrada formalmente en el **Inventario de las listas de referencia FT-IIT-2880**; asignando un código a la lista, este formato se diligencia atendiendo lo establecido en la hoja de instrucción de diligenciamiento.

4.2.2 Registro en el formato de Caracterización de la LR

Una vez el Gestor de Gobernanza de Datos ha asignado un código a la LR, procede a realizar la documentación individual en el formato **Caracterización de la lista de referencia FT-IIT-2881**, el cual, está compuesto por las siguientes hojas:

- Estructura_LR.
- LR_Cargue.
- Anexo_1_Detalle_Campo_Estándar.

Dichas hojas se diligencian conforme a lo establecido en sus hojas de instrucción de diligenciamiento.

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

4.3 Validación y aprobación de la LR

Una vez documentada, la LR el Gestor de Gobernanza valida técnica y funcionalmente la LR según el escenario en el que se identificó:

- a. Mesas de trabajo dentro del marco del PAMD - se realiza en conjunto con el delegado designado del módulo que se encuentra en verificación dentro del PAMD.
- b. Ejercicios de calidad de datos - se ejecuta con el Custodio de Datos delegado por el Líder de Dominio.
- c. Caracterización de dominios de datos - la validación es realizada exclusivamente por el Gestor de Gobernanza.

Una vez se surten las validaciones necesarias y dependiendo de la criticidad para el proceso afectado por la LR, la aprobación se realizará a través del Líder de Dominio o del Dueño de los Datos, dependiendo de las circunstancias presentadas; esta LR se será administrada dentro del modelo de Gobernanza de Datos por el equipo de Gobierno de Datos de la Subdirección de Información y Analítica.

4.4 Publicación en el catálogo y repositorio

Las LR aprobadas deberán ser publicadas en dos instancias obligatorias. En primera instancia en Databricks, como catálogo institucional donde se dispone la LR para su consumo y uso en los sistemas informáticos de la entidad. En segunda Instancia la documentación de la LR deberá conservarse en el repositorio de SharePoint designado por la SIA, el cual contendrá los documentos soporte y anexos asociados al proceso de validación, aprobación y documentación de la LR en todas sus versiones, incluyendo la vigente y las anteriores.

Se considerará como LR válida aquella que se encuentre debidamente publicada tanto en Databricks como en el repositorio de SharePoint, siendo requisito indispensable el cumplimiento de ambas condiciones para su uso institucional. En todos los casos, la administración de estos repositorios será responsabilidad exclusiva de la SIA o quien haga sus veces.

Adicionalmente, una vez la LR sea cargada en el catálogo de Databricks, se deberá garantizar la correcta gestión de sus metadatos, incluyendo su adecuada definición, clasificación y etiquetado, con el fin de facilitar su identificación, entendimiento y uso por parte de los consumidores de información.

Finalmente, los controles de acceso al catálogo de Databricks serán gestionados por la Subdirección de Procesamiento de Datos o quien haga sus veces, en concordancia con lo establecido en marco normativo vigente.

4.5 Administración de la LR.

4.5.1 Verificar y actualizar la caracterización de la LR

El Gestor de Gobernanza de Datos deberá realizar la verificación periódica de la información individual de la LR registrada en el formato de **Caracterización de la lista de referencia FT-IIT-2881**. Esta verificación se llevará a cabo de acuerdo con la periodicidad definida en el campo “Frecuencia de consulta del estándar” del mencionado formato o en atención a solicitudes específicas realizadas por usuarios de los datos.

En caso de que, como resultado de la verificación, no se evidencien cambios en los valores de la LR, se mantendrá la versión vigente de la misma y únicamente se procederá a actualizar el campo “Fecha de revisión Lista de Referencia:” en el formato de Caracterización.

Por el contrario, si se identifican cambios en los valores de la LR, se deberá:

- Actualizar la información correspondiente en el formato de Caracterización de la lista de Referencia FT-IIT-2881.
- Generar una nueva versión de la LR, incorporando los cambios identificados, conforme a las instrucciones de diligenciamiento del formato.
- Gestionar el control de versiones y registro de cambios asociados.

Antes de la actualización de la LR en Databricks se socializa con la Subdirección de Procesamiento de Datos (SDP) y Subdirección de Soluciones y Desarrollo (SSD) para evaluar los posibles impactos de los cambios y actualizaciones. Conforme a esto se acuerda la fecha de actualización de los cambios en Databricks. De estas acciones de actualización se deberá mantener informados a los Custodios y Líderes de Dominio para que implementen las acciones requeridas de acuerdo con los cambios presentados.

4.5.2 Garantizar la disponibilidad de las LR frente a los sistemas de información (SIES)

Cuando se notifiquen cambios en la arquitectura o ruta de los sistemas donde se encuentran publicadas las diferentes LR por parte del área encargada, se deben realizar los acompañamientos y alistamientos necesarios que permitan garantizar la disponibilidad e integridad de la información contenida en las LR.

5. CONTROL DE CAMBIOS

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

Elaboró:	Luis Miguel Almarza Pino Elaboración técnica	Gestor	Subdirección de Información y Analítica
Revisó:	Jhonattan Alberto Bustos Ramírez Revisión metodológica	Analista	Subdirección de Procesos
	Pedro Alberto Camargo Casas Revisión técnica	Inspector	Subdirección de Información y Analítica
Aprobó:	Jesús Daniel Cañizares Osorio	Subdirector	Subdirección de Información y Analítica