

Diseño Conceptual del Modelo para la Gestión de Datos Maestros en Colombia – Versión 4

Fase 2

Proyecto Diseño de un Modelo de Gestión de Datos Maestros para Colombia

Marzo 13 de 2022



MINISTERIO DE TECNOLOGÍAS
DE LA INFORMACIÓN Y LAS
COMUNICACIONES



DEPARTAMENTO
NACIONAL DE PLANEACIÓN



BANCO DE DESARROLLO
DE AMÉRICA LATINA



Pontificia Universidad
JAVERIANA
Colombia



Alianza **CAOBA**
Centro de Excelencia y Apropiación
en Big Data y Data Analytics

Equipo de trabajo

DNP

Javier Alfonso Lesmes Patiño - Coordinador Grupo Transformación y Economía Digital
Diana Paola Ramírez Roa – Líder administrativo/Supervisor Técnico
Nicolas Enrique Chibuque Pérez - Apoyo administrativo/Supervisor Técnico
Andrés Ocampo Sáenz - Apoyo administrativo/Supervisor Técnico

MinTIC

Luis Clímaco Córdoba Gómez - Subdirector de Estándares y Arquitectura de TI
Jairo Alberto Riascos Muñoz - Líder técnico/Supervisor técnico

Banco de desarrollo para América Latina – CAF

María Isabel Mejía - Ejecutiva Senior de la Dirección de Transformación Digital en CAF

Alianza CAOBA

Hugo Sin Triana - Director
Pedro F. Pérez Arteaga – Líder de analítica
Juan Sebastián Sandino – Líder de innovación
Astrid Herrera – Líder de Comunicaciones
Adriana Suárez – Diseñadora

Control de cambios

VERSIÓN	FECHA	MODIFICACIÓN	RESPONSABLE
1.0	29/11/2022	Definición de los conceptos base para el desarrollo del Diseño Conceptual para la gestión de datos maestros	Marta Silvia Tabares Sandra Liliana Santos
2.0	06/12/2022	Ajustes en el título del documento y el Equipo de Trabajo. Adición de los objetivos específicos de acuerdo con el contrato de la consultoría.	Marta Silvia Tabares Sandra Liliana Santos
2.1	13/01/2023	Cambios en la estructura de contenidos Avance en las secciones de: Arquitectura Gestión Gobierno	Marta Silvia Tabares Sandra Liliana Santos
3.0	15/02/2023	Ajustes de acuerdo a observaciones del equipo DNP-MinTIC Actualización de la matriz de roles y responsabilidades de la gobernanza para la GDM Actualización del Capítulo de Calidad de datos Actualización del documento solo con el alcance del entregable Fase 2	Marta Silvia Tabares Sandra Liliana Santos
4.0	13/03/2023	Ajustes de acuerdo a observaciones del equipo DNP-MinTIC Refinamiento de los capítulos: - Catálogo de Datos - Calidad de la Gestión de Datos Maestros - Seguridad y protección de Datos Maestros	Marta Silvia Tabares Sandra Liliana Santos
5.0	10/04/2023	Ajustes observaciones finales	Marta Silvia Tabares Pedro F. Pérez Arteaga

Contenido

1. Objetivo General y Específicos de la Elaboración del Diseño Conceptual necesario para la Gestión de Datos Maestros	9
2. Metodología de Trabajo	10
3. Marco Teórico	11
3.1. Concentrador de Datos Maestros	11
3.2. Catálogo de Datos	12
3.3. Metadatos	13
3.4. Registro de Oro	13
4. Estrategia para la Gestión de Datos Maestros	15
4.1. Paso 1. Requisitos y Modelos de Referencia	17
4.1.1. <i>Requisitos Estratégicos para la Gestión de Datos Maestros del Gobierno Nacional</i>	17
4.1.2. <i>Modelos de Referencia y Componentes para la Gestión de Datos Maestros</i>	17
4.1.3. <i>Principios que fundamentan la Gestión y el Gobierno de los Datos Maestros</i>	19
<i>Principio 1: Participación de Entidades y el Concentrador GDM</i>	20
<i>Principio 2: Maximizar los beneficios del Concentrador GDM</i>	21
<i>Principio 3: La Gestión de los Datos Maestros es asunto de todas las entidades participantes</i>	22
<i>Principio 4: Gestión de Datos Maestros como Servicio</i>	22
<i>Principio 5: Armonización del Marco Normativo</i>	23
<i>Principio 6: Responsabilidad de la Entidad Administradora del Concentrador GDM</i>	24
<i>Principio 7: Calidad de los datos</i>	25
<i>Principio 8: Fácil búsqueda, accesibilidad, interoperabilidad y reutilización</i>	26
<i>Principio 9: Seguridad y Protección de los datos</i>	27
<i>Principio 10: Confianza pública y gestión ética de los datos</i>	28
<i>Principio 11: Estandarización</i>	29
4.1.4. <i>Políticas definidas para la gestión de datos maestros</i>	29
5. Diseño de la Arquitectura para la Gestión de Datos Maestros del Gobierno Nacional	33
5.1. Estilos de la Arquitectura	33
5.1.1. <i>Estilo de Registro o Federado</i>	34
5.1.2. <i>Estilo Centralizado</i>	35
5.1.3. <i>Estilo de Consolidación</i>	37
5.1.4. <i>Estilo de Coexistencia</i>	38
5.1.5. <i>Híbrida</i>	38
5.2. Arquitectura de Referencia	39
5.3. Modelos de Datos	41

5.3.1. Dominio de Personas	42
5.3.2. Dominio de Empresas	47
5.3.3. Dominio de Lugares	53
5.4. Modelos de Datos de Referencia	59
5.4.1. Tabla de referencia – hubref_direccion	59
5.4.2. Tabla de referencia – hubref_usuario	62
5.4.3. Tabla de referencia – hub_referencia	63
5.5. Ciclo de Vida de los Datos Maestros	66
5.6. Ciclo de vida de los datos de referencia	71
6. Gobierno para la Gestión de Datos Maestros	72
6.1. Gobierno de Datos multidominio	73
6.2. Marco para la Gestión y Gobierno de Datos Maestros	74
6.3. Roles y Responsabilidades del Gobierno de Datos Maestros	75
6.4. Procesos de Estrategia y Gobierno para la Gestión de Datos Maestros	79
7. Catálogo de Datos Maestros	84
7.1. Introducción	84
7.2. Proceso de Gestión del Catálogo de Datos Maestros y de Referencia	84
7.3. Instrumentos de trabajo para gestionar el Catálogo de Datos Maestros y de Referencia	87
7.3.1. Glosario de las Entidades	87
7.3.2. Definición de Metadatos - Diccionarios de Datos	88
7.3.3. Perfil de los activos de Datos	88
7.3.4. Metadatos a nivel de Entidad/Tabla	89
7.3.5. Metadatos a nivel de Atributo/Columna	90
7.3.6. Metadatos a nivel de Dominio de Valor	91
8. Aseguramiento de la Calidad de la Gestión de Datos Maestros	92
8.1. Perfilamiento de Datos Maestros	92
8.2. Dimensión y reglas de Calidad del dato	93
8.2.1. Dimensión de Exactitud	94
8.2.2. Dimensión de Completitud	95
8.2.3. Dimensión de Consistencia	96
8.2.4. Dimensión de Unicidad	97
8.2.5. Dimensión de Validez	97
8.2.6. Dimensión de Oportunidad	97
8.2.7. Dimensión de Vigencia	98
8.2.8. Dimensión de Conformidad	98
8.2.9. Dimensión de Integridad	99
8.2.10. Dimensión de Precisión	99
8.3. Reglas de emparejamiento (matching)	99

8.4. Reglas de Supervivencia	101
8.5. Medición de la calidad de los datos.....	102
9. Seguridad de Datos Maestros	105
10. Privacidad de los de Datos Maestros	108
Glosario	111
Temas para Refinamiento en Entregable 3	112
Referencias.....	115

Lista de Figuras

FIGURA 1. GESTOR DE DATOS MAESTROS (ALLEN & CERVO, 2015).	12
FIGURA 2. HOJA DE RUTA PARA LA REALIZACIÓN DE LA ESTRATEGIA DE LA GDM DEL GOBIERNO NACIONAL.	16
FIGURA 3. MODELO DE DOMINO CRUZADO – MULTI-DOMINIO Y MULTI-ENTIDAD PARA LA GESTIÓN DE DATOS MAESTROS DEL GOBIERNO DE COLOMBIA.	19
FIGURA 4. ARQUITECTURA HÍBRIDA MULTI-DOMINIO. UNA ADAPTACIÓN DE (ALLEN & CERVO, 2015).	40
FIGURA 5. ARQUITECTURA FEDERADA. UNA ADAPTACIÓN DE (ALLEN & CERVO, 2015) CON FUENTES ASOCIADAS AL PROYECTO.	41
FIGURA 6. MODELO LÓGICO DEL REGISTRO DE ORIGEN RNEC PERSONAS - FUENTE: REGISTRADURÍA NACIONAL DE ESTADO CIVIL.	43
FIGURA 7. MODELO DE DATOS DE LA TABLA MAESTRA DE PERSONAS HUBDPER_MAESTRAPERSONA – DOMINIO DE PERSONAS.	45
FIGURA 8. MODELO LÓGICO DEL REGISTRO ORIGEN CCRUES. FUENTE: DEFINIDA DESDE EL DICCIONARIO DE DATOS PROPORCIONADO POR CONFECÁMARAS.	48
FIGURA 9. MODELO DE DATOS DE LA ENTIDAD MAESTRA DE RUES EN EL ROR – DOMINIO DE EMPRESAS.	51
FIGURA 10. MODELO FÍSICO DEL REGISTRO ORIGEN DE CATASTRO – IGACRIC. FUENTE: IGAC (IGAC-REPOSITORIO, 2022).	54
FIGURA 11. MODELO DE DATOS DE LA MAESTRA DE CATASTRO HUBDLUG_MAESTRACASTRO – DOMINIO LUGARES.	56
FIGURA 12. MODELO DE REFERENCIA DE DIRECCIÓN – HUBREF_DIRECCION.	60
FIGURA 13. MODELO DE REFERENCIA DE USUARIO – TABLA HUBREF_USUARIO.	62
FIGURA 14. MODELO DE REFERENCIA – TABLA HUB_REFERENCIA.	64
FIGURA 15. PROCESO DEL CICLO DE VIDA DE LA GESTIÓN DE DATOS MAESTROS MULTI-DOMINIO HÍBRIDO.	67
FIGURA 16. FLUJO DE TRABAJO PARA GESTIONAR LA EXTRACCIÓN O CARGA DE LOS DATOS MAESTROS O DE REFERENCIA DE ORIGEN.	68
FIGURA 17. SUBPROCESO DE INTEGRACIÓN DE DATOS.	70
FIGURA 18. CICLO DE VIDA DE LOS DATOS DE REFERENCIA.	71
FIGURA 19. RELACIÓN DE LAS ENTIDADES PARTICIPANTES CON LOS DOMINIOS DE DATOS.	74
FIGURA 20. MARCO PARA LA GESTIÓN Y GOBIERNO DE DATOS MAESTROS EN COLOMBIA.	75
FIGURA 21. MATRIZ DE ROLES Y RESPONSABILIDADES PARA EL CONCENTRADOR GDM DEL GOBIERNO NACIONAL.	76
FIGURA 22. PROCESO ESTRATEGIA PARA LA GESTIÓN DE DATOS MAESTROS DEL GOBIERNO NACIONAL – NIVEL ESTRATÉGICO.	80
FIGURA 23. PROCESO GESTIÓN DE DATOS MAESTROS DEL GOBIERNO NACIONAL – NIVEL TÁCTICO.	81
FIGURA 24. PROCESO DATAOps PARA LA GESTIÓN DE DATOS MAESTROS – NIVEL OPERATIVO.	83
FIGURA 25. PROCESO PARA LA GESTIÓN DEL CATÁLOGO DE DATOS MAESTROS Y DE REFERENCIA.	85
FIGURA 26. PROCESO DE GESTIÓN DE LA CALIDAD DE DATOS MAESTROS.	92
FIGURA 27. JERARQUÍA PARA EL ASEGURAMIENTO DE LA CALIDAD DE LOS DATOS MAESTROS Y DE REFERENCIA.	94

Lista de Tablas

TABLA 1. ACTIVIDADES DEFINIDAS PARA EL DESARROLLO DE LA FASE 2. (FUENTE: ELABORACIÓN PROPIA)	10
TABLA 2. INSTRUMENTO DE CALIDAD 1 - DEFINICIÓN DE PRINCIPIOS ESTRATÉGICOS PARA LA GESTIÓN Y GOBIERNO DE LA GDM.	19
TABLA 3. VENTAJAS Y DESVENTAJAS - ESTILO DE ARQUITECTURA REGISTRO O FEDERADA.	35
TABLA 4. VENTAJAS Y DESVENTAJAS – ESTILO ARQUITECTURA CENTRALIZADO	36
TABLA 5. VENTAJAS Y DESVENTAJAS – ESTILO ARQUITECTURA CONSOLIDADO	37
TABLA 6. VENTAJAS Y DESVENTAJAS – ESTILO ARQUITECTURA DE COEXISTENCIA	38
TABLA 7. VENTAJAS Y DESVENTAJAS – ESTILO ARQUITECTURA HÍBRIDA.	39
TABLA 8. DICCIONARIO DE DATOS DE LA MAESTRA DE PERSONAS RNECPERSONAS .	44
TABLA 9. DICCIONARIO DE DATOS MAESTRA DE PERSONA – HUBDPER_MAESTRAPERSONA.	45
TABLA 10. DICCIONARIO DE DATOS MAESTRA DE PERSONA – HUBDPER_UBICACIONPER.	47
TABLA 11. DICCIONARIO DE DATOS DE LA MAESTRA DE EMPRESAS CCRUES.	49
TABLA 12. DICCIONARIO DE DATOS DE LA TABLA MAESTRA DEL RUES HUBDEPM_MAESTRARUES – DOMINIO EMPRESAS.	51
TABLA 13. DICCIONARIO DE DATOS DE LA MAESTRA DE LUGARES IGACRIC.	54
TABLA 14. DICCIONARIO DE DATOS DE LA TABLA MAESTRA DE CATASTRO HUBDLUG_MAESTRACATASTRO – DOMINIO LUGARES.	57
TABLA 15. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA DE DIRECCIÓN HUBREF_DIRECCION.	60
TABLA 16. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_DIVIPOLA.	61
TABLA 17. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_DEPARTAMENTO.	62
TABLA 18. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_COMUNA.	62
TABLA 19. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_USUARIO.	63
TABLA 20. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUB_REFERENCIA.	64
TABLA 21. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUB_TIPOREFERENCIA.	65
TABLA 22. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUB_TIPOREFERENCIA.	66
TABLA 23. CARACTERÍSTICAS DEL REGISTRO DE ORO DE ORIGEN (ROO).	68
TABLA 19. INSTRUMENTO DE CATÁLOGO DE DATOS MAESTROS 1 - DEFINICIÓN DE TÉRMINOS DEL NEGOCIO BASE PARA LA GDM.	88
TABLA 20. PERFIL DE LOS ACTIVOS DE DATOS.	89
TABLA 21. METADATOS A NIVEL DE ENTIDAD/TABLA	89
TABLA 22. METADATOS A NIVEL ATRIBUTO/COLUMNA.	90
TABLA 23. METADATOS AL NIVEL DE DOMINIO DE VALOR	91
TABLA 24. MEDICIONES REPRESENTATIVAS PARA LA GESTIÓN DE DATOS MAESTROS.	103
TABLA 25. DIMENSIONES Y MEDICIÓN DE CALIDAD DE DATOS.	103
TABLA 26. CATEGORÍAS DE RIESGOS DE SEGURIDAD DE LOS DATOS.	105
TABLA 27. TIPOS DE DATOS DEFINIDOS EN EL CONPES 3920.	109

1. Objetivo General y Específicos de la Elaboración del Diseño Conceptual necesario para la Gestión de Datos Maestros

1.1. Objetivo General

Diseñar el modelo de gestión de datos maestros para Colombia que incluya el detalle de los procesos de gestión de datos maestros, la definición de roles y responsabilidades de las entidades involucradas.

1.2. Objetivos Específicos

1. Revisar los modelos de gestión de datos maestros disponibles a nivel internacional y que sean adaptables para Colombia. ([sección 3.2](#))
2. Elaborar un documento con la definición del catálogo de datos maestros en el que se relacionen las fuentes y sistemas de información y diccionarios de datos estandarizados utilizando el lenguaje común de intercambio de información. ([capítulo 6](#))
3. Definir los flujos para el intercambio, estandarización, actualización, distribución y consulta de datos maestros, integrando el rol de las entidades productoras y consumidoras de dichos datos.
4. Definir los procesos de gestión de datos maestros relacionados con la uniformidad, precisión, administración, consistencia, calidad, semántica y responsabilidad de los datos maestros. ([sección 5.3](#))
5. Definir lineamientos de buenas prácticas de gestión de la calidad para datos maestros que puedan ser adaptables para datos de referencia. ([capítulo 7](#))
6. Elaborar el documento con el diseño del modelo de gestión de datos maestros, en el que se tenga en cuenta la adopción de estándares técnicos y semánticos definidos por MinTIC, y que incluya una propuesta de integración para el consumo de datos de referencia. ([capítulos 4 al 6](#))
7. Elaborar una propuesta de indicadores para hacer seguimiento a la implementación del modelo de gestión de datos maestros. ([sección 6.6](#))

2. Metodología de Trabajo

La metodología para lograr el objetivo específico de la consultoría, “OE2: Diseñar el modelo de gestión de datos maestros para Colombia que incluya el detalle de los procesos de gestión de datos maestros, la definición de roles y responsabilidades de las entidades involucradas,” está definida a partir de tres actividades como se muestra en la Tabla 1.

Tabla 1. Actividades definidas para el desarrollo de la Fase 2. (Fuente: Elaboración Propia)

Actividad	Tareas	Responsables	Entregable
1. Definir los requisitos técnicos y procesos mínimos de gestión.	<ul style="list-style-type: none"> Definir los requisitos que guiarán el diseño conceptual Identificar los modelos de referencia, elementos de modelo de arquitectura y procesos de gestión que requiera la gestión de datos maestros en un contexto multi-Dominio y multi-Entidad Clasificar los elementos o componentes arquitectónicos mínimos requeridos para definir la arquitectura multi-Dominio y multi-Entidad para el gobierno colombiano Definir la arquitectura base de alto nivel que guiará el diseño del modelo de gestión de datos maestros 	Consultoría CAOBA Apoyo: Equipo DNP-MinTIC	Documento entregable Fase 2 – Informe parcial 1
2. Diseñar modelo de gestión de datos maestros.	<ul style="list-style-type: none"> Definir el conjunto integral de elementos y guías prácticas que orientarán el diseño arquitectónico desde los puntos de vista de: Gestión de la calidad, Gestión de metadatos, Gestión de Datos de Referencia, e Integración de Datos. Diseñar a alto nivel los modelos de datos, arquitectura de datos, y el ciclo de vida de los datos que definirán la gestión de datos maestros a nivel nacional. Esto en conjunto con los procesos requeridos para la gestión de los datos maestros 	Consultoría CAOBA	Documento Entregable Fase 2 – Avance
3. Definir el gobierno de datos - Interacción y restricciones	<ul style="list-style-type: none"> Definir el conjunto integral de elementos y guías prácticas que orientarán las disciplinas clave de la gestión de datos maestros: Gobierno de datos y Administración de datos maestros Refinar el documento final de la fase 	Consultoría CAOBA Apoyo: Equipo DNP-MinTIC	Documento Entregable Fase 2 – Entrega Final

Inicialmente, se identifican los requisitos que guiarán el desarrollo de cada uno de los elementos que serán parte del diseño conceptual para la gestión de datos maestros, así como los modelos de referencia y los componentes mínimos requeridos para definir la arquitectura deseada. Una vez se comprenden cada uno de los componentes o procesos que deben ser parte de la arquitectura, se procede a hacer un análisis de la forma cómo las entidades y sus datos maestros o registros únicos van a participar en la gestión de datos maestros del nivel nacional. Esto

determinará la forma ideal de abordar la gestión de datos maestros desde diferentes vistas arquitectónicas, y así definir el estilo arquitectónico más propicio para gestionar los datos maestros del nivel nacional.

3. Marco Teórico

3.1. Concentrador de Datos Maestros

Allen y Cervo en (Allen & Cervo, 2015) definen “un concentrador en un dominio con arquitectura como un gran gestor de calidad de los datos que entran y salen a este. Un concentrador de datos normalmente tendrá varios sistemas internos conectados como sistemas radiales y puede tener una o más fuentes de datos externos (como datos de proveedores) utilizados en el concentrador de datos. Estos flujos de datos entrantes y salientes son áreas clave donde se deben asignar administradores de datos. Estos administradores de datos deben estar muy familiarizados con los datos y los procesos transaccionales o de proveedores que se conectan al centro. La Figura 1 proporciona un ejemplo de este concepto.

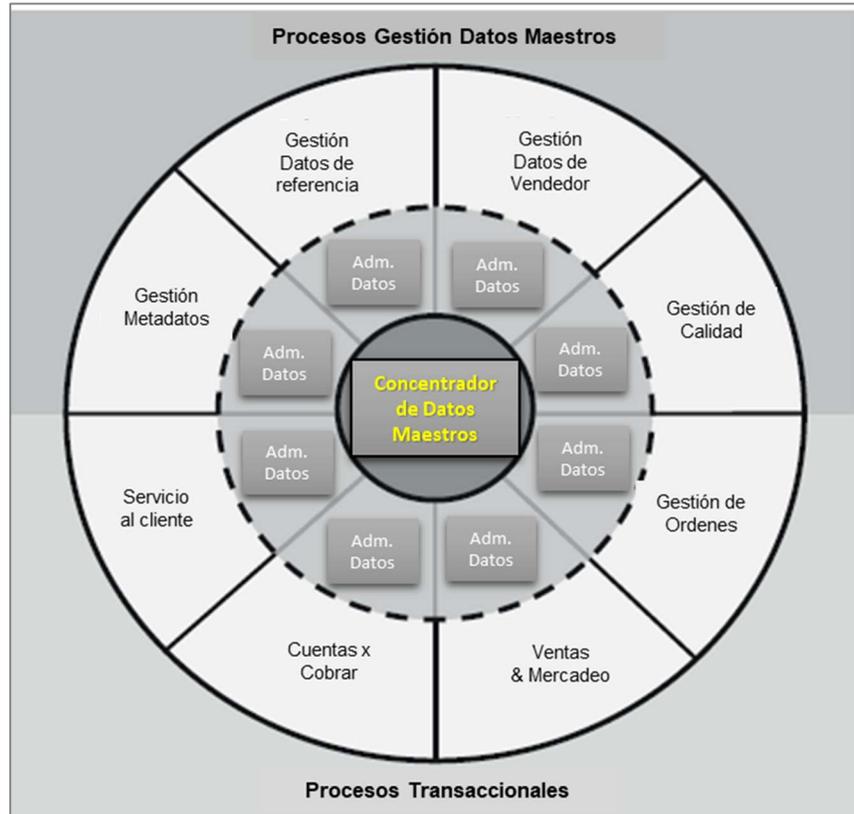


Figura 1. Gestor de Datos Maestros (Allen & Cerro, 2015).

El concentrador para la Gestión de Datos Maestros (GDM) propuesto en este proyecto y denominado “Concentrador GDM” se refiere al repositorio donde se almacenan, procesan y consultan los datos maestros del Gobierno Nacional. Este se define desde un modelo centralizado, el cual requiere menos mapeo y normalización de datos, un solo proceso de extracción, transformación y carga (ETL) por cada dominio y entidad maestra, pocos o ningún proceso de resolución de entidades, controles y balances y otros procesos de gestión de calidad de datos. Esto se traducirá en un administrador de datos que debe participar la toma de decisiones a los administradores de datos en los sistemas de origen. El Concentrador GDM contendrá tanto las aplicaciones para la GDM como la plataforma de bases de datos o sistemas de bases de datos para la Gestión de datos maestros. En esta se define la arquitectura de datos más conveniente para gestionar y gobernar los datos maestros y de referencia del gobierno de Colombia.

3.2. Catálogo de Datos

Ziaidi y Duncan en (Zaidi, Simoni, Edjlali, & Duncan, 2017) definen “Un catálogo de datos mantiene un inventario de activos de datos a través del descubrimiento, descripción y organización de conjuntos de datos. El catálogo

proporciona contexto para permitir que los analistas de datos, los científicos de datos, los administradores de datos y otros consumidores de datos encuentren y comprendan un conjunto de datos relevante con el fin de extraer valor comercial.” (Zaidi, Simoni, Edjlali, & Duncan, 2017). El Catálogo de Datos es un instrumento de calidad que facilita el gobierno de los datos desde la gestión de los metadatos de las entidades maestras y de referencia que se definen en el Concentrador GDM. Talent en (Talent, 2023) hace el siguiente ejemplo, para orientar la finalidad de un catálogo de datos “Para ilustrar, piense en un catálogo en línea para encontrar libros en una biblioteca. El catálogo es una fuente de verdad administrada centralmente donde los lectores pueden buscar cualquier libro. El catálogo conserva los metadatos de cada libro, incluidos el título, el autor, el resumen, la fecha de publicación y la ubicación en la estantería. También puede incluir calificaciones y recomendaciones de otros lectores, y una forma de dejar sus propias reseñas. El catálogo en línea permite a los usuarios de la biblioteca encontrar y acceder a lo que necesitan más rápido que si recorrieran la biblioteca estante por estante. Por supuesto, todos los metadatos deben mantenerse bien organizados y actualizados para que los usuarios confíen en ellos y los utilicen.” (Talent, 2023).

3.3. Metadatos

Los metadatos son datos sobre datos. La norma ISO 15489 define los metadatos de los registros como: “información estructurada o semiestructurada, que permite la creación, gestión y uso de registros a lo largo del tiempo y dentro y entre dominios”. Los metadatos se refieren a “nombres, fechas de creación y cualquier otra información contextual que describa los datos en su lago de datos o almacén de datos. Todos estos metadatos agregan información significativa a sus conjuntos de datos. Esto mejora la usabilidad de los datos y los convierte en un activo real para su organización. Un catálogo de todos los metadatos hace posible la búsqueda y recuperación de cualquier dato.” (Talent, 2023). Allen y Cervo en (Allen & Cervo, 2015) indican que “la gestión de metadatos es quizás la capacidad de gestión de datos menos explorada dentro de un MDM multidominio. Debido a que las empresas a menudo comienzan sus planes de MDM con un solo dominio, la cantidad de elementos maestros puede ser relativamente pequeña y esto puede no justificar tener una gestión de metadatos dedicada. Pero a medida que aumenta la cantidad de dominios, hay más elementos de datos maestros que necesitan control y se realizan más actividades de integración de datos, por lo que también aumenta la necesidad de un enfoque de gestión de metadatos más formalizado y centralizado.” (Allen & Cervo, 2015).

3.4. Registro de Oro

Allen y Cervo en (Allen & Cervo, 2015) plantean que el Registro de Oro es “un conjunto único de elementos de datos maestros relacionados con una sola instancia de una entidad. Una sola instancia de una entidad significa una persona, o una empresa, o un paciente, o una cuenta, o un producto, y así sucesivamente, independientemente de cuántos registros de datos existan para cada uno de ellos. El registro de oro es, en esencia, la única versión de la verdad para la ocurrencia de una entidad determinada.” (Allen & Cervo, 2015).

4. Estrategia para la Gestión de Datos Maestros

La Gestión de Datos Maestros del Gobierno Nacional es una apuesta estratégica del Plan Nacional de Infraestructura de Datos (DNP-PNID, 2021). La estrategia que se describe a lo largo de este documento da directriz, a través de una hoja de ruta, al qué, cómo, quién, cuándo, y donde se deben gestionar y gobernar los datos maestros del Gobierno Nacional. Para lograr esto, se usan buenas prácticas y modelos de referencia estandarizados a nivel mundial y nacional que guían la forma de beneficiar el recurso humano y tecnológico con los que actualmente cuenta el gobierno de Colombia y sus entidades públicas para la administración de sus datos maestros. En el desarrollo de esta estrategia también se aprovechan las oportunidades que brinda el Gobierno Digital, las cuales darán seguridad y permitirán unir esfuerzos paralelos para apoyar el desarrollo económico del país desde el manejo de datos únicos confiables para los servicios transaccionales y la toma de decisiones a nivel nacional. Esto lleva finalmente a darle sentido estratégico a la recopilación, el aprovechamiento, la limpieza, la escalabilidad, y la integración de los datos maestros generados desde y para diferentes entidades del Estado. Para lograr esto se definen los siguientes propósitos que guiarán el desarrollo de la estrategia planteada para el desarrollo del Modelo de Gestión de Datos Maestros:

Propósito General

Definir un Modelo de Gestión de Datos Maestros del Gobierno Nacional que garantice el uso de buenas prácticas gestión y gobierno de los datos únicos cuyas fuentes de datos sean las entidades del Estado, así como sus usuarias. Esto con el fin de mejorar la toma de decisiones y nivel del diseño e implementación de políticas públicas asociada a la administración de los datos únicos del nivel nacional.

Propósitos Específicos

- Garantizar un diseño de arquitectura repetible y medible a partir de la definición de guías, modelos, arquitectura, y buenas prácticas para la gestión de datos maestros y de referencia
- Proporcionar alta calidad de los datos maestros gestionados por el Gobierno Nacional, en armonía con los datos maestros proporcionados por las entidades participantes
- Generar los mecanismos y tecnologías requeridas para que los datos maestros y de referencia del Gobierno Nacional sean gestionados desde un Concentrador de Datos que provea el acceso a datos únicos que apoyarán la estrategia y la operación de las entidades públicas o privadas que lo demanden.

Los propósitos anteriormente definidos se lograrán a partir de una serie de pasos/actividades definidas en la hoja de ruta ilustrada en la Figura 2. Esta divide el desarrollo de la Gestión de Datos Maestros del Gobierno Nacional – pilar del Plan Nacional de Infraestructura de Datos (PNID), en cinco grandes pasos que guiarán la implementación de un concentrador de datos maestros y de referencia basado en una arquitectura flexible,

escalable, y de alta calidad que garantice su sostenibles a largo plazo. El paso (1) uno se desarrolla a lo largo de este documento, como parte del alcance definido en este proyecto.

- **Paso 1.** Estrategia, Arquitectura, y Gobierno: Definición de la estrategia y el gobierno que se deben establecer para el diseño e implementación del Concentrador GDM. Además, de la definición de la arquitectura de referencia de datos maestros a alto nivel y los procesos que se deberán tener en cuenta para la gestión de los datos maestros.
- **Paso 2.** Implementación del Concentrador GDM: Construcción de los componentes que le dan vida a la arquitectura lógica y física y a los procesos que automatizarán la GDM en una plataforma tecnológica específica.
- **Paso 3.** Entrega del Concentrador de Datos Maestros. Desarrollo de las pruebas que se aplicarán a cada uno de los componentes de datos definidos en el concentrador de datos maestros.
- **Paso 4.** Puesta en marcha del Concentrador GDM. Una vez sean aceptadas las pruebas, se dará vía libre para poner en marcha el Gestor de Datos Maestros del Gobierno Nacional.

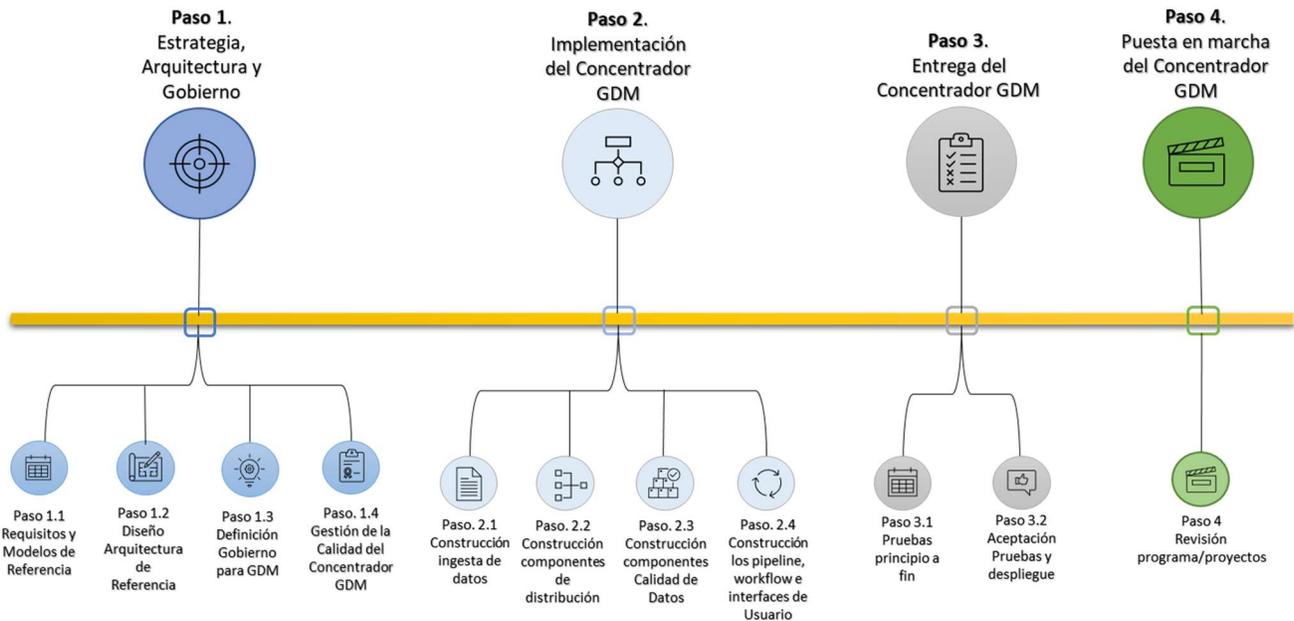


Figura 2. Hoja de Ruta para la realización de la Estrategia de la GDM del Gobierno Nacional.

4.1. Paso 1. Requisitos y Modelos de Referencia

4.1.1. Requisitos Estratégicos para la Gestión de Datos Maestros del Gobierno Nacional

El diseño conceptual para la gestión de datos maestros del nivel nacional deberá cumplir con los siguientes requisitos.

R1: Definir un enfoque lógico para la planificación, la implementación y la gestión continua de los datos maestros en el contexto Multi-Dominio y Multi-Entidad, tanto desde la perspectiva de un administrador de programas como la de un administrador de datos.

R2: Definir el conjunto integral de elementos y guías prácticas que orientarán las disciplinas clave de la gestión de datos maestros: Gobierno de datos, Administración de datos, Gestión de la calidad, Gestión de Metadatos, Gestión de Datos de Referencia, e Integración de datos.

R3: Diseñar los modelos de datos, arquitectura de datos, y el ciclo de vida de los datos que definirán la gestión de datos maestros del nivel nacional. Esto en conjunto con los procesos requeridos para la gestión de los datos maestros.

R4: Definir las reglas de negocio que determinan la interacción con los datos maestros y las funciones que respaldan las capacidades de gestión de las entidades, seguridad, auditoría y detección de eventos. Esto incluye las funciones de cargas de datos, limpieza, vinculación y duplicación, etc.

R5: Definir el contexto y los contenidos que orientarán la planificación del programa de la gestión de datos maestros a nivel nacional, las técnicas de participación de las diferentes entidades, control de cambios, y control de calidad, que serán compartidos con los participantes de la gestión de datos maestros.

Estos requisitos son la guía para definir los diferentes componentes que deberá tener una arquitectura de alto nivel para la gestión de datos maestros del Gobierno de Colombia.

4.1.2. Modelos de Referencia y Componentes para la Gestión de Datos Maestros

Para agilizar el desarrollo de los diferentes módulos de la hoja de ruta se seleccionaron dos modelos de referencia:

- a) El Modelo de Datos Maestros Multi-Dominio (MDM-MD) de (Allen & Cervo, 2015). Este provee componentes de estrategia, gobierno, perfilado de datos, arquitectura, etc., los cuales se adaptarán para definir la gestión de datos maestros del Gobierno Nacional en los módulos del Paso 1.

- b) Se adoptó el modelo de procesos de referencia para la gestión de datos maestros propuesto por (Reichert, Otto, & Österle, 2013), con el objetivo de mostrar la funcionalidad del modelo propuesto desde tres áreas de procesos que agrupan diferentes actividades generales:
- Estrategia: hoja de ruta, propósitos, y principios de gestión y gobierno.
 - Gobierno: marco de trabajo, roles y responsabilidades, procesos de gestión y gobierno.
 - Operaciones: arquitectura de datos, calidad y catálogo de datos.

La Figura 3 muestra el Modelo de Domino Cruzado y Multi-Entidad adaptado para la Gestión de Datos Maestros del Gobierno de Colombia. Este diagrama involucra los dominios básicos, alcance de este proyecto: Personas, Empresas, y Lugares; además, se propone un grupo de Otros Dominios o Dominios Futuros: transporte, vehículos, propiedades y direcciones. De igual forma, las Entidades se clasifican en dos grupos:

- Entidades Actuales (definidas en el alcance de este proyecto – óvalo continuo y fondo azul): Registraduría Nacional de Estado Civil, el Departamento Administrativo Nacional de Estadísticas – DANE, la Confederación de Cámaras de Comercio – Confecámaras, y el Instituto Agustín Codazzi – IGAC, en relación con los dominios de datos de Personas, Empresas y Lugares.
- Entidades Futuras (identificadas en el estado del arte - óvalo punteado): Migración Colombia, Ministerio de Transporte/Secretarías de Tránsito, y la Superintendencia de Notariado y Registro – SNR, en relación con el dominio de Personas y los dominios futuros de Transporte, Vehículos, y Propiedades respectivamente.

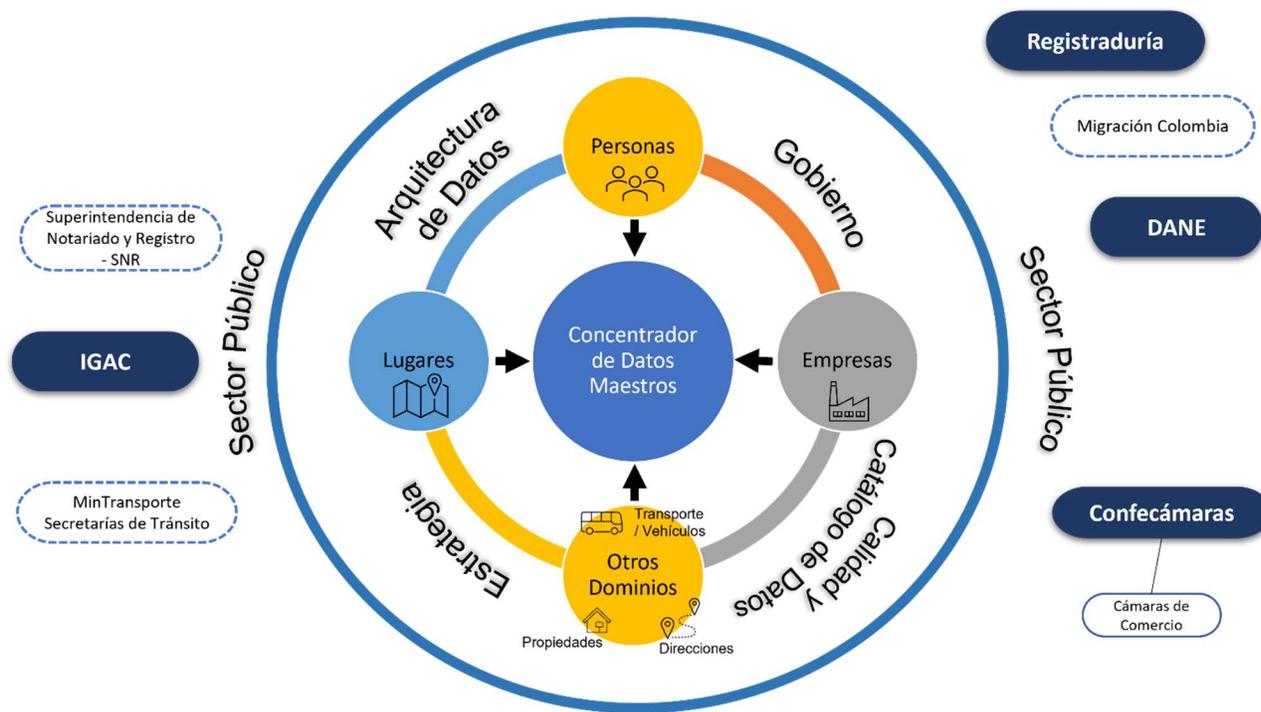


Figura 3. Modelo de domino cruzado – Multi-Dominio y Multi-Entidad para la Gestión de Datos Maestros del Gobierno de Colombia.

4.1.3. Principios que fundamentan la Gestión y el Gobierno de los Datos Maestros

Los principios, que se definen a continuación, son reglas y directrices generales que definen y guían al PNID desde diferentes puntos de vista para dar cumplimiento a la definición, implementación, y puesta en marcha de la Gestión y Gobierno de los datos maestros del Gobierno Nacional. Estos principios deben ser la base de las políticas y reglas (o normas) que darán vida a la arquitectura de datos base del Concentrador GDM (Concentrador para la Gestión de Datos Maestros). Por esto se requiere que dichas políticas y reglas sean duraderas y rara vez modificadas, de tal forma que se pueda medir la gestión y el gobierno implementado. En la Tabla 2 se presenta un instrumento de trabajo para definir los principios que apoyan la estrategia.

Tabla 2. Instrumento de Calidad 1 - Definición de Principios estratégicos para la gestión y gobierno de la GDM.

Nombre	<ul style="list-style-type: none"> • Debe representar la esencia de la meta sobre la cual se toman decisiones de calidad • Debe ser fácil de recordar. • En el nombre o declaración del principio no deben mencionarse plataformas tecnológicas específicas. • Evitar palabras ambiguas
Declaración	<ul style="list-style-type: none"> • El principio se debe comunicar de forma sucinta y sin ambigüedades • Es vital que la declaración de cada principio sea inequívoca

Razón Fundamental	<ul style="list-style-type: none"> • Resaltar los beneficios gubernamentales y para las entidades del Estado y particulares de adherirse al principio, usando terminología común a todos los participantes. • Señalar la similitud de los principios de gestión de la calidad de la GDM y la tecnología asociada con los principios que rigen las operaciones del Estado. • Describir la relación con otros principios y las intenciones con respecto a una interpretación equilibrada. • Describir situaciones en las que se dé prioridad a un principio o tenga más peso que otro para tomar una decisión.
Trascendencia	<ul style="list-style-type: none"> • Resaltar los requisitos, tanto para la entidad participante como para el Concentrador GDM, para llevar a cabo el principio, en términos de recursos, costos y actividades/tareas. • Describir el impacto en la adopción del principio en los diferentes sectores o entidades del gobierno o privadas participantes del Concentrador GDM • Definir el nivel de importancia del impacto: alto, medio o bajo, dependiendo de la importancia del principio para el Concentrador GDM y las entidades participantes.

A continuación, se declaran los principios que deberán regir la estrategia y el gobierno de la GDM del Gobierno Nacional.

Principio 1: Participación de Entidades y el Concentrador GDM

Declaración

Los principios definidos para la gestión y el gobierno de los datos maestros del Gobierno Nacional aplican a aquellas entidades que hacen parte del Concentrador GDM teniendo en cuenta su distinción de origen sean estas públicas o privadas.

Razón Fundamental

Logra un nivel consistente y medible de los datos maestros para los tomadores de decisiones del Gobierno Nacional y otros sectores públicos o privado, es si y sólo si todas las entidades participantes cumplan la políticas y reglas establecidas para la gestión de datos maestros a través de la colaboración efectiva entre estas y la entidad que administre los datos del Concentrador GDM.

Trascendencia

- Sin este principio, las exclusiones, la preferencia y la incoherencia socavarían rápidamente la gestión de los datos maestros definida para el Gobierno Nacional.
- La gestión de los datos maestros no comenzará hasta que se examine el cumplimiento de los principios definidos.

- Un conflicto con un principio se resolverá entre las entidades participantes y el Gobierno Nacional. Se analizarán las causas y sus implicaciones. Si fuese necesario se harán los cambios requeridos o se eliminará del marco de referencia establecido para la GDM.

Principio 2: Maximizar los beneficios del Concentrador GDM

Declaración

Las decisiones de gestión de los datos almacenados en el Concentrador GDM son tomadas para proporcionar el máximo beneficio del tratamiento de los datos maestros desde el Concentrador GDM tanto para el Gobierno Nacional y las entidades participantes.

Razón Fundamental

- Las decisiones que se tomen alrededor de la GDM deben impactar positivamente al Gobierno Nacional y a las entidades participantes para lograr mayor valor de los datos maestros a largo plazo tanto en la transaccionalidad como en la toma de decisiones.
- La gestión de los datos maestros debe estar alineada con el Plan Nacional de Infraestructura de Datos y el Modelo de Gobernanza de la infraestructura de datos.
- Ninguna entidad participante del Concentrador GDM, en particular, restará valor a los beneficios que aporte éste al conglomerado. Además, cada entidad implementará los diferentes lineamientos requeridos por el Concentrador GDM.

Trascendencia

- Lograr el máximo beneficio para todas las entidades participantes y el Gobierno Nacional requerirá un análisis para determinar si requiere cambios en la forma como se planifican y administran los datos maestros; la tecnología por sí sola no provocará los cambios requeridos. La gestión del cambio institucional orientará este tipo de cambios.
- Algunas entidades deben ajustar las políticas o normas actuales que han definido para la gestión de sus datos maestros para alcanzar el mayor beneficio de la gestión de datos maestros a nivel nacional.
- Las prioridades de desarrollo de sistemas para la gestión de datos maestros se deben establecer en cada una de las entidades participantes para armonizarse de manera más asertiva con la GDM del Gobierno Nacional.
- Los componentes del Concentrador GDM del Gobierno Nacional pueden ser accedidos por las entidades participantes de acuerdo con las políticas vigentes al momento en que este se encuentre en operación.

La gestión de datos maestros en cada una de las entidades participantes debe llevarse a cabo de acuerdo los lineamientos definidos en el PNID y en la documentación técnica vigente para la Administración de los datos maestros además de otros documentos que se vayan declarando que apliquen a la gestión de datos maestros del Gobierno de Colombia.

Principio 3: La Gestión de los Datos Maestros es asunto de todas las entidades participantes

Declaración

Todas las entidades públicas que hagan parte del Concentrador GDM deben participar en las decisiones de gestión de los datos maestros de los dominios priorizados en la resolución 460 de 2022 (MinTIC, 2022), así como otras entidades o dominios que se definan en el tiempo, de tal forma que se logran los objetivos estratégicos del Gobierno Nacional en el contexto de la Gestión de datos Maestros o de Referencia.

Razón Fundamental

- Los usuarios del Concentrador GDM son entidades públicas registradas como participantes de la Gestión de Datos Maestros del Gobierno Nacional, ya sea en rol de proveedora o consumidora de datos.
- Para garantizar que la GDM del Gobierno Nacional esté alineada con el PNID – Datos Maestros, todas las entidades públicas o privadas participantes del Concentrador GDM deberán aceptar la responsabilidad de desarrollar los elementos administrativos, reglamentario, y tecnológico requerido por el equipo de gobierno del Concentrador GDM para la gestión y el gobierno de sus datos maestros. Esto implica la definición y cumplimiento diferentes acuerdos establecidos por el Concentrador GDM para la gestión de los datos maestros y de referencia. Esto aplica tanto para la entidad fuente o proveedoras como para las entidades consumidoras de los datos maestros del Concentrador GDM.
- La gestión de datos maestros, y la gestión del concentrador de GDM requiere de la disposición de perfiles técnicos y estratégicos de las entidades públicas y privadas a apoyar de manera conjunta las metas y objetivos definidos para la Gestión de Datos Maestros.
- El Concentrador GDM requiere de la disposición de perfiles técnicos y estratégicos de las entidades participantes públicas y privadas, para apoyar de manera conjunta las metas y objetivos definidos para su adecuada gestión.

Trascendencia

- Se requiere de la articulación de las entidades y empresas involucradas en la Gestión de los Datos Maestros del Gobierno Nacional para atender la demanda de la gobernanza para la gestión e implementación de la estrategia de datos maestros.
- Para la adopción de este principio, es necesario la planeación y disposición de recursos físicos, humanos y financieros por parte de las entidades públicas y privadas vinculadas al concentrador de datos.

Principio 4: Gestión de Datos Maestros como Servicio

Declaración

La gestión de los datos maestro definida como servicio se garantiza a partir de una arquitectura híbrida para el Concentrador GDM. Esta habilita aspectos como automatización de los flujos de datos, agilidad para emparejar y combinar datos desde diferentes fuentes, aseguramiento de datos únicos robustos del gobierno de Colombia para consumo de diferentes entidades públicas y privadas, y flexibilidad y escalabilidad del Concentrado GDM a través del tiempo.

Razón Fundamental

- El Concentrador GDM brinda agilidad al Gobierno Nacional para operar y tomar decisiones basadas en datos maestros con un alto nivel de verdad y flujos de información sin límites.
- Los servicios de datos son definidos entre el Gobierno Nacional y las entidades públicas y privadas participantes del Concentrador GDM. Así se proporciona un contexto apoyado por procesos, objetivos, reglas, políticas, interfaces de servicios y componentes de servicios. De igual forma, implementa la orquestación de servicios de datos maestros.

Trascendencia

- La orientación al servicio impone requisitos únicos a la infraestructura, y las implementaciones deben usar estándares abiertos para lograr la interoperabilidad y la transparencia de la ubicación de los datos de la fuente.
- Las implementaciones realizadas en el Concentrador GDM son especificadas e implementadas en el marco de trabajo acordado entre el Gobierno Nacional y las entidades participantes.
- La gobernanza definida se refleja en representación e implementación del servicio.

Principio 5: Armonización del Marco Normativo

Declaración

Los procesos de gestión de datos maestros del Gobierno Nacional están alineados con las leyes, políticas y reglamentaciones de Colombia pertinentes y vigentes en relación con la operación del concentrador de datos maestros.

Razón Fundamental

- Las políticas de la GDM del Gobierno Nacional deben cumplir con las leyes, políticas y reglamentos vigentes definidas para el tratamiento de datos maestros y de referencia. Esto no impedirá las mejoras en los procesos asociados a la gestión de los datos maestros que conduzcan a cambios en las políticas y reglamentaciones del nivel nacional.

Trascendencia

- El Gobierno Nacional debe tener en cuenta el cumplimiento de las leyes, los reglamentos y las políticas internas y externas con respecto a la recopilación, retención y gestión de datos maestros.
- Se debe crear un instrumento normativo que defina la gobernanza y estrategia de GDM del estado colombiano.
- El equipo que sea parte del marco de gobierno de datos maestros, propuesto para el Gobierno de Colombia, debe capacitarse en función del marco normativo, las reglas de funcionales (de negocio) y de tecnológica definidas para la administración de datos maestros y de referencia que se gestionen en el Concentrador GDM.
- La eficiencia, la necesidad y el sentido común no son los únicos impulsores. Los cambios en la ley y los cambios en las regulaciones pueden generar cambios en los procesos o aplicaciones del gestor de datos maestros del nivel nacional.

Principio 6: Responsabilidad de la Entidad Administradora del Concentrador GDM

Declaración

La Entidad Administradora del Concentrador GDM asignado para la Gestión de Datos Maestros del Gobierno Nacional es responsable de definir e implementar procesos, soluciones e infraestructura de TI requerida para que el Concentrador GDM cumplan con los requisitos definidos en cuanto a funcionalidad, niveles de servicio, costo y tiempo de operación.

Razón Fundamental

- Alinear de manera efectiva las expectativas definidas para el Concentrador GDM con las capacidades y los costos de cada una de las entidades que hacen parte de éste, De tal forma que todos los proyectos definidos para la viabilidad de la hoja de ruta propuesta sean rentables para el Gobierno Nacional.

Trascendencia

- Se debe crear un proceso para priorizar los proyectos para la administración del Concentrador GDM del Gobierno Nacional.
- La función del Administrador de datos del Concentrador GDM debe definir procesos, crear lineamientos y guías para que las entidades vinculadas al concentrador alineen sus capacidades con los requerimientos técnicos para integrarse al concentrador.
- Se deben abordar y resolver todas las necesidades tecnológicas como, construcción de modelos de bases de datos, desarrollo de aplicaciones y servicios web, infraestructura tecnológica, entre otras, requeridas para la puesta en marcha e implementación de la estrategia para la GDM del Gobierno Nacional.
- El instrumento normativo se debe definir entre la entidad que opere la estrategia de datos básicos y el concentrador de GDM.

Los principios que se declaran a continuación fueron definidos por el PNID (DNP-PNID, 2021) los cuales están asociados a: calidad de los datos; fácil búsqueda, accesibilidad, interoperabilidad y reutilización; seguridad y protección de los datos; confianza pública y gestión ética de los datos; y estandarización. Estos se adaptan directamente a la Gestión de Datos Maestros del Gobierno Nacional.

Principio 7: Calidad de los datos

Declaración

Los datos maestros y de referencia que son parte de la infraestructura de datos del Gobierno Nacional deben cumplir con las características de calidad desde las siguientes dimensiones: exactitud, completitud, integridad, actualización, coherencia, relevancia, accesibilidad y confiabilidad.

Razón Fundamental

- Uno de los beneficios de un entorno de la GDM es la capacidad de compartir datos entre el Gobierno Nacional y las entidades participantes.
- A medida que crece el nivel de gestión de los datos maestros del Gobierno Nacional, las entidades tienen mayor dependencia de estos. Así, es necesario que los administradores de datos de las entidades proveedoras y consumidoras en conjunto con el administrador del Concentrador GDM tomen acciones preventivas relacionadas a la calidad de los datos maestros y de referencia.
- Dado que los datos pueden perder su integridad por diferentes razones; el Administrador de los datos de la entidad fuente garantizará la integridad de los datos con procesos previamente establecidos bajo una gobernanza vigente.

Trascendencia

- Este principio compromete el mantenimiento y gestión de los conjuntos de datos maestros dispuestos en la infraestructura de datos.
- La gobernanza asociada a la seguridad y protección de los datos maestros defina el Gobierno Nacional debe tratar factores relacionados a la "propiedad" de los datos producidos por las entidades proveedoras, y la disponibilidad de esos datos para satisfacer las necesidades de todas de las entidades consumidoras del Concentrador GDM a través del registro de oro definido en cada dominio. Esto implica que puede ser necesario un cambio cultural de la "propiedad" de los datos del orden nacional. De igual forma, las entidades proveedoras deben asegurar el cumplimiento de las normativas definidas por el gobierno nacional alrededor de la protección de datos.
- El Administrador de datos, en conjunto con los equipos tácticos definidos en el marco de gobierno de datos, serán responsables de cumplir con los requisitos de calidad definidos sobre los datos maestros y de referencia del Gobierno Nacional. Para lograr, ellos deberán utilizará las herramientas del Concentrador GDM, como el

Catálogo de Datos Maestros, el cual apoya la gestión de metadatos. Además, la solución tecnológica sobre la cual se implemente la gestión de los datos maestros deberá facilitar la creación de políticas y reglas de calidad definidas para el Concentrador GDM.

- Es esencial que el Catálogo de Datos Maestros del Concentrador GDM tenga la capacidad de brindar confianza a las entidades participantes (productoras o consumidoras) sobre los datos en función de atributos como "fuente de datos"
- Es esencial que la arquitectura del Concentrador GDM y el Catálogo de Datos Maestros faciliten la identificación de las fuentes de los datos verdaderos, de tal forma que el Administrador de Datos pueda definir reglas de presentación relacionadas a los convenios establecidos con las entidades consumidoras.
- La información requerida por el Catálogo de Datos Maestros debe capturarse desde el proceso de extracción (ingesta) de los datos fuentes. Así, es posible garantizar calidad, integridad, y linaje durante su ciclo de vida. Esto implica la definición de procesos y guías que consoliden la gestión del catálogo.
- Como resultado de compartir datos maestros entre el Gobierno Nacional y las entidades participantes, el Administrador del Concentrador GDM en conjunto con el Administrador de datos de la Entidades son los responsables de la exactitud y actualidad de datos gestionados en el Concentrador GDM.

Principio 8: Fácil búsqueda, accesibilidad, interoperabilidad y reutilización

Declaración

- Los datos que integran la infraestructura de datos del Estado se deben caracterizar por los principios FAIR (por sus siglas en inglés), tomados de las propiedades de los datos científicos. Estos deben ser de fácil búsqueda, de fácil acceso, interoperables y dispuestos para su reutilización.
- Dado que la infraestructura de datos está integrada por distintos tipos de datos, según su naturaleza, este principio está condicionado al principio de "tan abierto como sea posible, tan cerrado como sea necesario".

Razón Fundamental

- El acceso a los datos maestros, correctamente definido, conduce a la eficiencia y eficacia en la toma de decisiones y permite una respuesta oportuna a las solicitudes de información y la prestación de servicios de búsqueda, accesibilidad, interoperabilidad, y reutilización de datos maestros.
- El uso de los datos maestros almacenados en el Concentrador GDM del Gobierno Nacional debe considerarse desde una perspectiva que permitir el acceso a una amplia variedad de entidades públicas y privadas de acuerdo con convenios establecidos entre las partes. Así, se ahorran tiempos administrativos y se mejora la consistencia de los datos.

Trascendencia

- En particular con relación al principio de accesibilidad, de acuerdo con la Ley 1712 de 2014 este hace parte del núcleo del derecho de transparencia y acceso a la información pública, que debe armonizarse con los principios de finalidad y circulación restringida que rigen el derecho al habeas data (Ley 1266 de 2008) y con los atributos de la información .de carácter de tipo personal, propios de la protección y privacidad de la información, conforme con la normativa de protección de datos personales (Ley 1581 de 2012).
- Debe promulgarse una cultura alrededor del acceso a los datos maestros del Gobierno Nacional. Esta, para garantizar que todas las entidades públicas y privadas participantes del Concentrador GDM del Gobierno Nacional entiendan la relación entre el valor de los datos, su administración, el intercambio y la accesibilidad a los datos.
- La forma en que se accede y se entregan a los datos maestros debe ser lo suficientemente adaptable para cumplir con las necesidades de una amplia gama de entidades públicas y privadas, y sus correspondientes métodos de acceso.
- El acceso a los datos maestros debe estar definido al interior del en el Concentrador GDM a través de políticas, normas o reglas. Esto evitará errores de interpretación por parte de los usuarios definidos en Concentrador GDM.
- El acceso a los datos maestros no otorga necesariamente al usuario derechos de acceso para modificar o divulgar los datos maestros. Esto lleva a crear un conjunto de restricciones y reglas asociadas al tipo de usuario y su relación con el uso de los datos maestros. Es decir, se debe diferenciar si son usuarios autorizados para solo consulta, o usuarios para acceder a datos que van a ser utilizados por las entidades fuente o consumidoras en sus propios modelos de datos maestros, o si son usuarios activos en el consumo de servicios asociados a la interoperabilidad o si es un usuario que puede reutilizar los datos de la fuente, como lo es el mismo Concentrador GDM.

Principio 9: Seguridad y Protección de los datos

Declaración

- En cualquiera de las fases del ciclo de vida de los datos maestros se debe garantizar la seguridad y la protección de los datos personales o que se consideren sensibles (de acuerdo con lo definido por la Ley 1581 de 2012).
- Así mismo, en los espacios de intercambios de datos entre distintos actores, se debe propender por la aplicación de prácticas que minimicen el riesgo de identificación y re-identificación de datos personales, aplicando el principio de responsabilidad demostrada, y privacidad desde el diseño y por defecto.
- Lo anterior con el fin de preservar la confidencialidad, integridad, y disponibilidad, y brindar confianza a las partes constitutivas de toda la infraestructura de datos.

Razón Fundamental

- La Seguridad y Protección definida para los datos maestros almacenados en el Concentrador GDM, debe equilibrarse con la legislación vigente del gobierno de Colombia que explícitamente defina la necesidad de restringir la disponibilidad de información clasificada, confidencial y sensible.
- Las leyes y reglamentos del gobierno de Colombia vigentes que salvaguardan la seguridad y la privacidad de los datos del nivel nacional deben ser la base para la definición de políticas y reglas que permitan el acceso a los datos almacenados en el Concentrador GDM. Así, Así, se evitarán interpretaciones erróneas y uso inapropiado por parte de las entidades participantes.

Trascendencia

- La arquitectura y procesos en los cuales se defina la GDM del Gobierno Nacional debe proveer los mecanismos para que los datos maestros estén protegidos contra el uso y la divulgación no autorizados bajo la normatividad establecida por las leyes nacionales y particulares a cada entidad.
- La seguridad y privacidad de los datos maestros almacenados en el Concentrador GDM deben conservar las características de protección de datos sensible definida por las entidades proveedoras.
- La entidad proveedora de datos maestros que actúa como Administradora de los datos del registro de oro o los datos de referencia que serán consumidos por el Concentrador GDM periódicamente. La entidad proveedora no perderán potestad sobre sus datos, ni la competencia de aplicar principios de protección y seguridad de la información de acuerdo con la ley de protección de datos personales.

Principio 10: Confianza pública y gestión ética de los datos

Declaración

- La calidad de los datos que integran la infraestructura de datos el Estado, y su modelo de Gobernanza deben propender por la creación de un escenario de confianza para la generación, acceso, intercambio y reutilización de datos entre los distintos actores.
- A su vez, esta confianza se fortalece con la integración de la gestión ética de los datos durante su ciclo de vida, y la reducción de sesgos de discriminación, injusticia y parcialidad con el fin de proteger los derechos humanos.
- Este principio debe propender por generar confianza a todos los actores que hacen parte de la infraestructura de datos del Estado colombiano y a la sociedad en general, sobre el contenido y veracidad de los datos recabados y puestos a disposición de los diferentes usuarios, siendo esto un propósito para generar valor público en un entorno de confianza y seguridad digital.

Razón Fundamental

- El Concentrador GDM y los procesos implementados en éste, garantizarán la confianza en el tratamiento de los datos al definir los diferentes principios aquí descritos. Los cuales a su vez asegurarán la gestión ética de los datos bajo la definición de una gobernanza que así lo procure.

Trascendencia

- La arquitectura del Concentrador GDM y los procesos, estándares, normatividad, participantes, y acceso, entre otros aspectos asociados a la gestión de los datos maestros deben ser implementados bajo los principios definidos en esta sección.
- La implementación del Concentrador GDM del Gobierno Nacional debe garantizar la integridad de los datos maestros y establecer parámetros de confiabilidad y seguridad requerida por los participantes y la legislación establecida en el PNID.

Principio 11: Estandarización

Declaración

- La infraestructura de datos del Estado debe operar bajo criterios de estandarización a través de todo su ciclo de gestión. Se debe considerar sobre los datos en sí mismos y en los metadatos.
- La aplicación de este principio debe caracterizarse por una baja dependencia de esquemas complejos, facilitando la interoperabilidad y el lenguaje común de intercambio entre sistemas de información, y la posibilidad de integrarlos e intercambiar datos a través de servicios web (web services).

Razón Fundamental

- Los buenos procesos de gestión de datos maestros requieren inicialmente la definición de estándares de datos.
- Específicamente para la GDM Multi-Dominio y bajo una arquitectura híbrida, se debe garantizar la aplicación de criterios de estandarización para los diferentes pasos del ciclo de vida de los datos maestros. Esto permite tener datos maestros bajo un contexto de datos comunes para las diferentes entidades participantes.

Trascendencia

- La arquitectura Multi-Dominio e híbrida y procesos en los cuales se defina la GDM del Gobierno Nacional debe proveer los mecanismos para que los datos maestros sean estandarizados en diferentes fases del ciclo de vida. Así se garantiza el almacenamiento de datos correctos y el buen uso de ellos por diferentes usuarios.

4.1.4. Políticas definidas para la gestión de datos maestros

Una política es una orientación o directriz que se debe divulgar para dar mayor claridad y entendimiento de los propósitos que se formulan dentro del contexto de la gestión de datos maestros. Las políticas se implementan a través de: reglas para los sistemas que crean o modifican datos maestros o de referencia. Las políticas se convierten en guías o lineamientos para orientar diferentes acciones sobre los datos del Concentrador GDM, y

por lo tanto deben ser acatadas por los administradores de datos del Gobierno Nacional y de las Entidades proveedoras o consumidoras de dichos datos.

Cada política que se define a continuación permite observar un problema que se podría repetir una y otra vez dentro de la gestión de los datos maestros o de referencia del Concentrador GDM. En este sentido, las políticas son criterios generales de ejecución que complementan el logro de los objetivos y facilitan la implementación de los principios que se describen desde la estrategia definida para el Concentrador GDM. Por esta razón las políticas, reglas, y estándares permiten medir el logro de la calidad de los datos.

Las siguientes políticas están asociadas directamente para soportar los principios definidos en la anterior.

Las siguientes políticas son algunas de las que podrían soportar los principios del 1 al 6.

- **Política 1.** La administración y gestión de los datos maestros y de referencia almacenados en el concentrador de datos maestros es potestad del Gobierno Nacional.
- **Política 2.** Las entidades públicas del Gobierno Nacional, participantes del concentrador como proveedoras o consumidoras de datos maestros, deberán establecer un convenio formalizado de acuerdo con la normatividad vigente establecida por el Gobierno Nacional para la gestión de los datos maestros del nivel nacional.
- **Política 3.** Cada entidad participante del gestor de datos maestros del Gobierno Nacional, cuyo rol sea identificado como Registro de Origen, será la propietaria y la administradora de los datos que se proveen para forma parte del registro de oro de al menos una entidad maestra en cada dominio definido en el GDM del Gobierno Nacional.
- **Política 4.** Los datos del registro de oro coexistirán en el Registro de Origen y el Registro de Referencia bajo un convenio establecido entre el Gobierno Nacional y la entidad fuente o dueña de los datos del registro de oro.
- **Política 5.** La Administración de los datos maestros y de referencia gestionados en el Concentrador GDM del Gobierno Nacional es responsabilidad de los grupos tácticos del gobierno de datos Concentrador GDM. Así mismo, la operación de estos datos será responsabilidad de los grupos del nivel operativo.

Las siguientes políticas son algunas de las que podrían soportar el principio 7.

- **Política 6.** Los modelos físicos de datos maestros y de referencia definidos en el Concentrador GDM, así como las restricciones definidas para estos, se deben gestionar en relación con la arquitectura de datos maestros establecida para la gestión y gobierno de estos.
- **Política 7.** El administrador de la calidad de los datos almacenados en el Concentrador GDM debe supervisar y medir la gestión de los datos maestros o de referencia en relación con los sistemas de origen que crean o modifican los datos maestros, y los sistemas que consumen datos maestros.

- **Política 8.** Los metadatos de todas y cada una de las entidades maestras y de referencia definida en cada dominio deberán ser definidos, descritos y gestionados en el catálogo de datos maestros con base en dos aspectos relevantes: el registro de oro definido para cada entidad maestra y
- **Política 9.** Los datos maestros y de referencia almacenados y gestionados en el Concentrador GDM deberán regirse por todas y cada una de las reglas de calidad definidas en el Concentrador GDM.
- **Política 10.** La vigencia de los datos maestros en el Concentrador GDM debe ser validada constantemente con base en una política pública definida para este fin.
- **Política 11.** La actualización de la estructura del registro de oro de una entidad maestra debe concertarse entre el administrador de datos maestros del Concentrador GDM y el administrador de la fuente del registro de origen. Se debe documentar y aprobar el impacto que dicho cambio traería a las entidades consumidoras de los datos maestros.
- **Política 12.** La actualización de la estructura de los datos de referencia debe concertarse entre el administrador de datos maestros del Concentrador GDM y el administrador de la fuente del registro de origen. Se debe documentar y aprobar el impacto que dicho cambio traería a las entidades consumidoras de los datos maestros.
- **Política 13.** Las operaciones de escritura: Creación, Actualización, o Borrado realizadas en los datos del registro de oro de las diferentes entidades maestras definidas en cada dominio del Concentrador GDM, se deben gestionar mediante un canalizador (pipeline) que garantice la trazabilidad, exactitud, completitud, consistencia, y seguridad de los datos maestros.
- **Política 14.** Las operaciones de escritura: Creación, Actualización, o Borrado realizadas en los datos de referencia gestionados en Concentrador GDM deben provenir de requerimientos de fuentes externas que han sido certificadas por el Administrador de Datos o fuentes que provienen de entidades participantes del Concentrador GDM. En cada uno de los casos, estas operaciones se deben gestionar mediante un canalizador (pipeline) que garantice la trazabilidad, exactitud, completitud, consistencia, y seguridad de los datos maestros.
- **Política 15.** En el catálogo de datos maestros se debe mantener el histórico de cambios de todas y cada una de las estructuras de las entidades maestras y de referencia definidas en el Concentrador GDM. La gestión de históricos se debe gestionar mediante un canalizador (pipeline) que garantice la trazabilidad, exactitud, completitud, consistencia, y seguridad de los datos maestros.

Las siguientes políticas son algunas de las que podrían soportar el principio 8.

- **Política 16.** Las operaciones de consulta realizadas sobre cada una de las entidades maestras definidas en el Concentrador GDM deben obedecer a un convenio establecido entre las entidades consumidoras y el Administrador de los datos maestros del Gobierno Nacional donde se definan el tipo de consultas y los roles/usuarios de cada entidad que pueden consultar los datos maestros. Las consultas se deben gestionar mediante un canalizador (pipeline) que garantice la consulta a través de un mecanismo que provea los datos requeridos y seguridad de estos.

- **Política 17.** El acceso a los datos de las entidades maestras se implementa formal y explícitamente de acuerdo con los parámetros mínimos establecidos en los acuerdos entre el Administrador del Concentrador GDM y el Administrador de datos de cada entidad consumidora, teniendo en cuenta diferentes niveles de acceso y activos de datos a acceder: tabla/atributo, usuario autorizado, otros activos autorizados para acceso. El protocolo de acceso se activa y gestiona desde un proceso predeterminado definido desde un canalizador (pipeline) que garantiza el acceso y la trazabilidad de los datos maestros.
- **Política 18.** La interoperabilidad requerida en el ambiente de la arquitectura híbrida se debe garantizar a través de X-ROAD¹. Las reglas de intercambio serán implementadas allí para garantizar el acceso a los datos de diferentes dominios y entidades maestras.
- **Política 19.** En el contexto de una arquitectura híbrida, el Registros de Origen se identifica como una fuente de información válida, confiable y autorizada, que puede y debe ser reutilizada digitalmente por el Concentrador GDM en el Registro de Referencia como el registro de oro de una entidad maestra del que mínimamente una entidad pública es responsable de la recopilación, el uso, la actualización y la preservación de la información registrada en este a través del tiempo.

Las siguiente(s) políticas son algunas de las que podrían soportar el principio 9.

- **Política 20.** Cualquiera usuario que se cree e inicie sesión en el Concentrador GDM deberá leer y aceptar la política de uso aceptable. Esta define las restricciones de uso de cada recurso definido en el Concentrador GDM, así como el comportamiento adecuado e inadecuado cuando los usuarios acceden a esos recursos.
- **Política 21.** Las contraseñas de los usuarios que acceden al Concentrador GDM debe tener una longitud y complejidad establecida, así como la frecuencia con la que caducan, y el procedimiento para restablecer las contraseñas olvidadas.
- **Política 22.** El ambiente de seguridad definido para el Concentrador GDM debe facilitar la auditoría relacionada a los intentos de acceso y los cambios en la configuración del sistema.
- **Política 23.** El ambiente de seguridad de proveer los informes de incidentes de seguridad para ser analizados y tomar las acciones preventivas o correctivas necesarias.

¹ X-Road® es un software de código abierto y una solución de ecosistema que proporciona un intercambio de datos unificado y seguro entre organizaciones (p.ej., del sector público y privado) (Wikipedia: <https://en.wikipedia.org/wiki/X-Road>).

5. Diseño de la Arquitectura para la Gestión de Datos Maestros del Gobierno Nacional

La arquitectura que le da vida a la gestión de datos maestros en el Concentrador GDM del Gobierno Nacional, se diseña a partir de los requisitos estratégicos R3 y R4 definidos en la [sección 3.1.1](#), los cuales se especifican a continuación para dar mayor claridad de las acciones a realizar.

R3: Diseñar los modelos de datos, arquitectura de datos, y el ciclo de vida de los datos que definirán la gestión de datos maestros del nivel nacional. Esto en conjunto con los procesos requeridos para la gestión de los datos maestros.

- R3.1: el estilo de arquitectura para la gestión de datos maestros debe conservar, en lo posible, los datos maestros en su origen.
- R3.2: el estilo de arquitectura para la gestión de datos maestros debe gestionar los registros de oro de cada dominio en el concentrador central, conservando su confiabilidad: integridad y seguridad, de tal forma que se garantice a los consumidores datos exactos y correctos.
- R3.3: definir el ciclo de vida de los datos que serán almacenados en el Concentrador GDM de acuerdo con el estilo de arquitectura seleccionada para la gestión de datos maestros, de tal forma que cada dominio determine las restricciones de acceso y uso del registro de oro.
- R3.4: definir los modelos de datos que almacenarán los registros de oro de cada dominio en el concentrador, teniendo en cuenta tanto datos maestros como datos de referencia.
- R3.5: definir las reglas de integración y sincronización que se deben tener en cuenta para gestionar los datos maestros y de referencia entre los datos provenientes de las entidades fuente y el Concentrador GDM.

R4: Definir las reglas de negocio que determinan la interacción con los datos maestros y las funciones que respaldan las capacidades de gestión de las entidades, seguridad, auditoría y detección de eventos. Esto incluye las funciones de cargas de datos, limpieza, vinculación y duplicación, etc.

5.1. Estilos de la Arquitectura

La arquitectura para la gestión de datos maestros normalmente se define desde cuatro estilos comunes de implementación de la gestión de datos maestros. A continuación, se identifican las características técnicas, ventajas y desventajas de cada estilo arquitectónico. Esto se toma como un insumo para realizar un análisis con respecto a los requerimientos de gestión definidos en el requisito R3.

5.1.1. Estilo de Registro o Federado

Este estilo es totalmente opuesto al enfoque centralizado ya que cada sistema de origen mantiene el control de sus propios datos y sigue siendo el que provee las consultas requeridas sobre los datos maestros. Este estilo de arquitectura mantiene una conexión virtual entre la entidad fuente y el Concentrador GDM. Es decir, los datos maestros de origen no se replican en el Concentrador GDM, sino que se accede a ellos a través de identificadores que se van almacenando en el Concentrador GDM periódicamente o a demanda, y presenta el registro de oro a las entidades consumidoras como una serie de "vistas" que mapean el identificador y su correspondiente registro de oro de origen.

Otras características:

- Se utiliza principalmente para detectar datos duplicados mediante la ejecución de algoritmos de limpieza y coincidencia en los datos de sus diversos sistemas de origen.
- Asigna identificadores globales únicos a los registros coincidentes para ayudar a identificar una única versión de la verdad.
- Este estilo no devuelve datos a los sistemas de origen, por lo que los cambios en los datos maestros se siguen realizando a través de los sistemas de origen existentes. En su lugar, limpia y compara la información de referencia cruzada de identificación y asume que el sistema de origen puede administrar la calidad de sus propios datos.
- Se almacena la información necesaria para hacer coincidir y proporcionar el vínculo entre los registros correspondientes, y se puede acceder a una vista de estos datos según sea necesario.
- Cuando se necesita una vista única y completa del dato maestro, utiliza cada sistema de referencia para crear una vista de 360 grados en tiempo real. Sin embargo, se requiere un gobierno central de los datos para garantizar que el registro de oro sea confiable. En otras palabras, este estilo de arquitectura es viable si se implementa en un modelo de gestión de datos maestros federado. La Tabla 3 muestra las ventajas y desventajas del este estilo de arquitectura.

Tabla 3. Ventajas y Desventajas - Estilo de Arquitectura Registro o Federada.

Ventajas	Desventajas
<ul style="list-style-type: none"> • Debido a que cada aplicación mantiene sus propios datos, los cambios en el código de la aplicación para implementar este modelo suelen ser mínimos y, por lo general, los usuarios actuales de la aplicación no necesitan conocer el sistema de datos maestros. • Este estilo proporciona una vista de datos de solo lectura sin modificar los datos maestros y es una forma útil de eliminar duplicaciones y obtener un acceso constante a sus datos maestros. • Ofrece una integración de datos rápida y de bajo costo con el beneficio de una intrusión mínima en sus sistemas de aplicaciones. • Se puede utilizar un enfoque de estilo de registro para analizar los datos y evitar el riesgo de sobrescribir información en los sistemas de origen. Esto lo ayudará a evitar posibles fallas en el cumplimiento u otras repercusiones regulatorias (que pueden variar de una entidad a otra) que podrían ocurrir si se modifican los datos de origen. • Evita las batallas políticas en torno a la propiedad de los datos, ya que sitúa los datos maestros en su origen permitiendo acceder a ellos y los agrega dinámicamente. 	<ul style="list-style-type: none"> • Cada consulta sobre datos maestros es una consulta distribuida en todas las entradas de los datos deseados en todas las bases de datos de la aplicación • Agregar una aplicación al Concentrador GDM significa agregar columnas a la tabla de coincidencia de claves, lo cual no es un gran problema, pero también puede significar cambiar las consultas para incluir la nueva fuente de información. • Aunque este estilo de arquitectura ayuda a encontrar datos duplicados, no ayuda a limpiarlos. Por ejemplo: si una persona tiene muchos registros con diferentes números de teléfono, no hay forma de determinar cuál usar. En otras palabras, si se tiene una gran cantidad de sistemas de origen repartidos en diferentes partes, puede ser difícil establecer una fuente autorizada. • No aborda el tema de datos maestros inexactos, redundantes o inconsistentes; simplemente. La calidad de los datos depende de las fuentes de datos federadas. • La gestión de la información, como el control de versiones, la reversión, el control de auditoría o la validación, es manejada por los sistemas individuales y puede ser incoherente o inadecuada. No es posible crear y validar relaciones de datos para datos almacenados en varios sistemas. • Es un diseño orientado solo para el acceso. Si es necesario actualizar muchos sistemas para completar una transacción, la solución federada tendrá que crear una lógica compleja de protocolo de enlace y reversión para que sea viable.

5.1.2. Estilo Centralizado

Este estilo también se denomina de Transacción o Repositorio o Empresarial ya que almacena y mantiene atributos de datos maestros utilizando algoritmos de vinculación, limpieza, coincidencia y enriquecimiento para mejorar los datos. Los datos mejorados se pueden volver a publicar en su sistema de origen respectivo. La

colección completa de datos maestros para una empresa se almacena en una sola base de datos, incluidos todos los atributos requeridos por todas las aplicaciones que usan los datos maestros. Todas las aplicaciones que consumen crean o mantienen datos maestros se modifican para usar los datos maestros en el concentrador, en lugar de los datos maestros mantenidos anteriormente en la base de datos de la aplicación, lo que convierte al concentrador de datos maestros en el sistema de entrada (SOE por sus siglas en inglés) así como el sistema de registro (SOR por sus siglas en inglés). En otras palabras, el concentrador admite la fusión de registros maestros y los sistemas de origen pueden suscribirse a las actualizaciones publicadas por el sistema central para brindar una consistencia completa. Sin embargo, este estilo requiere una intrusión en los sistemas de origen para las interacciones. La Tabla 4 muestra las ventajas y desventajas de esta arquitectura.

Tabla 4. Ventajas y Desventajas – Estilo arquitectura Centralizado

Ventajas	Desventajas
<ul style="list-style-type: none"> • No hay problemas para mantener múltiples versiones del mismo registro de cliente en múltiples aplicaciones sincronizadas, porque todas las aplicaciones usan el mismo registro • El beneficio real de este estilo es que sus datos maestros son precisos y completos en todo momento • Hay menos posibilidades de registros duplicados porque solo hay un conjunto de datos, por lo que los duplicados son relativamente fáciles de detectar. 	<ul style="list-style-type: none"> • No siempre es fácil o incluso posible cambiar sus aplicaciones para usar nuevas fuentes o nuevos datos maestros • Crear un modelo de datos que incluya todos los datos necesarios, sin que sea tan grande que sea imposible de usar (es decir, tiene múltiples aplicaciones que requieren diferentes formatos de dirección) • Las políticas de seguridad y visibilidad a nivel de atributos de datos se definen totalmente desde el concentrador • ¿Qué hacer con los elementos de datos que no son utilizados por todas las aplicaciones? • Puede ser extremadamente costoso y llevar mucho tiempo implementarlo, ya que requiere cambios en las aplicaciones que mantienen y consumen los datos maestros. • Deberá transformar y cargar todas las bases de datos actuales en el concentrador, eliminando los duplicados en el proceso. • Deberá descubrir cómo manejar el historial, ya que está cambiando sus bases de datos para usar una nueva clave para todos sus datos maestros, por lo que debe lidiar con muchos años de historia que se creó usando diferentes claves para los datos maestros.

5.1.3. Estilo de Consolidación

En este estilo, los datos maestros de origen generalmente se consolidan a partir de múltiples fuentes en un concentrador para crear una única versión de la verdad, también conocida como el registro de oro. Ideal para reportería o analítica que está localizada en la bodega de datos. Los datos de múltiples fuentes se pueden limpiar, combinar e integrar para ofrecer un único registro completo para uno o más dominios de datos maestros. Un registro de oro se almacena en el concentrador central y se utiliza para informes y referencias. Sin embargo, cualquier actualización realizada en los datos maestros se aplica luego a las fuentes originales. La Tabla 5 muestra las ventajas y desventajas de este estilo de arquitectura.

Tabla 5. Ventajas y desventajas – Estilo arquitectura Consolidado

Ventajas	Desventajas
<ul style="list-style-type: none"> • No intrusivo al día a día del registro de datos de las diferentes áreas funcionales de una entidad • Las bodegas de datos son la plataforma donde se hace la consolidación de los datos maestros • Los beneficios dependerán del éxito de la estrategia de la inteligencia que se haga sobre los datos • No hay problemas para mantener múltiples versiones del mismo registro de cliente en múltiples aplicaciones sincronizadas, porque todas las aplicaciones usan el mismo registro • El beneficio real de este estilo es que sus datos maestros son precisos y completos en todo momento • Hay menos posibilidades de registros duplicados porque solo hay un conjunto de datos, por lo que los duplicados son relativamente fáciles de detectar. 	<ul style="list-style-type: none"> • No hace limpieza de fuentes • No siempre es fácil o incluso posible cambiar sus aplicaciones para usar nuevas fuentes o nuevos datos maestros • Crear un modelo de datos que incluya todos los datos necesarios, sin que sea tan grande que sea imposible de usar (es decir, tiene múltiples aplicaciones que requieren diferentes formatos de dirección) • Las políticas de seguridad y visibilidad a nivel de atributos de datos se definen totalmente desde el concentrador • No es claro qué hacer con los elementos de datos que no son utilizados por todas las aplicaciones • Puede ser extremadamente costoso y llevar mucho tiempo implementarlo, ya que requiere cambios en las aplicaciones que mantienen y consumen los datos maestros. • Deberá transformar y cargar todas las bases de datos actuales en el concentrador, eliminando los duplicados en el proceso. • Deberá descubrir cómo manejar el historial, ya que está cambiando sus bases de datos para usar una nueva clave para todos sus datos maestros, por lo que debe lidiar con muchos años de historia que se creó usando diferentes claves para los datos maestros.

5.1.4. Estilo de Coexistencia

El estilo de Coexistencia permite construir un registro de oro de la misma manera que el estilo de Consolidación, pero los datos se dominan en los sistemas de origen y luego se sincronizan con el concentrador, por lo que los datos pueden coexistir en armonía y seguir ofreciendo una única versión de la verdad. Todos los atributos del modelo de datos maestros deben ser coherentes y limpios antes de cargarlos en el sistema de gestión de datos maestros. Sin embargo, este estilo debe considerarse cuando se requiere vincular los datos controlados de forma centralizada con los sistemas de origen. La tabla 6 muestra las ventajas y desventajas de usar esta arquitectura.

Tabla 6. Ventajas y desventajas – Estilo arquitectura de Coexistencia

Ventajas	Desventajas
<ul style="list-style-type: none"> • Modelo distribuido a gran escala • Enfocado en servicios distribuidos 	<ul style="list-style-type: none"> • Mayor necesidad de duplicar datos • Requiere gobernanza local y global • Mayor riesgo en el control y la seguridad de los datos

5.1.5. Híbrida

Este estilo de arquitectura incluye características de los estilos de Registro y Centralizado. Es decir, el concentrador mantiene una conexión virtual con las bases de datos de origen a través de los identificadores únicos proporcionados por las fuentes, las cuales a su vez son responsables de la calidad de sus datos. El Concentrador GDM hace una réplica del registro de oro de cada entidad maestra que proviene de las fuentes, es la autoridad de la calidad de los datos, realiza la gestión multidominio, define y mantiene el gobierno de los datos maestros, define el modelo de datos centralizado, gestiona las relaciones o acuerdos entre las entidades proveedoras/consumidoras, consolida entidades de referencia, gestiona el ciclo de vida de los datos maestros, y publica los datos maestros a las entidades consumidoras. La tabla 7 muestra las ventajas y desventajas de usar esta arquitectura.

Tabla 7. Ventajas y desventajas – Estilo arquitectura Híbrida.

Ventajas	Desventajas
<ul style="list-style-type: none"> • Manejo de un registro de oro lo suficientemente robusto que la mayoría de las consultas de diferentes entidades se satisfagan desde el concentrador • Dado que los datos del concentrador GDM son una copia, los sistemas de origen pueden seguir utilizando su propia versión de los datos. • La creación y el mantenimiento de datos continúan en las fuentes originales y los atributos replicados en el concentrador se mantienen actualizados. • El concentrador GDM es un sistema de referencia que integra muchas fuentes, aplica la resolución de entidades y otras funciones de calidad de datos para dominar los datos y pone a disposición una versión única de la verdad para los sistemas de consumo. • El concentrador GDM es un sistema de referencia que integra muchas fuentes, aplica la resolución de entidades y otras funciones de calidad de datos para dominar los datos y pone a disposición una versión única de la verdad para los sistemas de consumo. 	<ul style="list-style-type: none"> • Debe atender la sincronización, los conflictos de actualización y los problemas de latencia de replicación. • Reconoce las deficiencias de un modelo de transacción y la necesidad de modificar las aplicaciones para consultar un concentrador GDM para todos los datos maestros. • Existe la necesidad de sincronización continua y sus muchas facetas, como la sincronización por lotes o en tiempo real, qué interfaces son bidireccionales, actualizaciones en conflicto de múltiples fuentes, qué atributos se ven afectados para cada fuente, etc.

5.2. Arquitectura de Referencia

El estilo de arquitectura híbrida Multi-Dominio se propone como el estilo de arquitectura que debe ser adoptado para la gestión de datos maestros del Gobierno Nacional, conservando las siguientes características:

- Mantener los registros de origen de los datos maestros de cada dominio en las bases de datos de las aplicaciones origen y mantener las claves en el Concentrador GDM bajo las reglas de coexistencia y calidad requeridas y definidas en el gobierno de datos.
- Replicar los atributos de los registros de origen de cada entidad maestra de cada dominio en el Concentrador GDM.
- Mantener los históricos de cambios/actualizaciones a un mismo registro – auditoría.

- Asegurar, en el Concentrador GDM, la calidad y la seguridad de los datos que provienen de las fuentes, y a la vez identificar la gobernanza correspondiente que mantengan los requerimientos de los propietarios de los datos en el catálogo del Concentrador GDM.
- Implementar, en el Concentrador GDM, la gobernanza requerida para que los administradores de los datos maestros realicen de forma correcta la gestión de los registros de oro de cada dominio.

Las Figuras 4 y 5 muestran la arquitectura híbrida multidominio propuesta para la gestión de datos maestros para el Estado colombiano. Particularmente, para el dominio de Personas se puede llegar a constituir un registro de oro compuesto por información proporcionada por la Registraduría como: la identificación, donde el proveedor debe ser la Registraduría y posiblemente con información de Personas complementaria proveniente de otras entidades participantes que aporten a su vez a otras tantas entidades públicas y privadas.

La Figura 4 presenta los componentes y su interacción. Se resalta el entendimiento de esta arquitectura ya que va a ser parte de la solución propuesta.

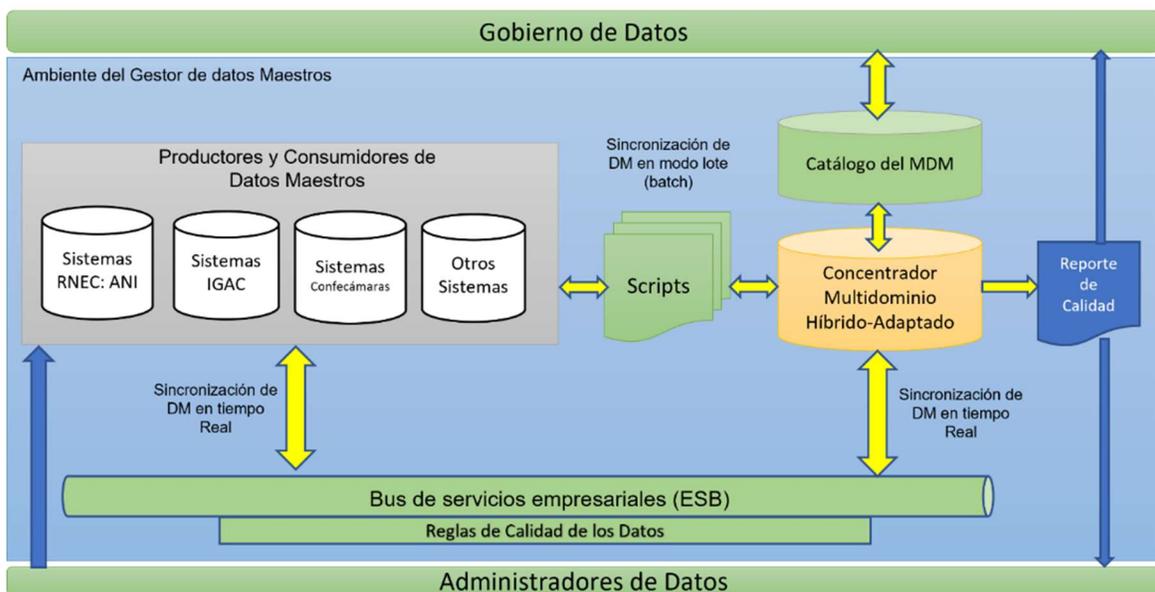


Figura 4. Arquitectura híbrida Multi-Dominio. Una adaptación de (Allen & Cervo, 2015).

La figura 5 muestra la representación del funcionamiento de la arquitectura federa que es parte de la arquitectura.

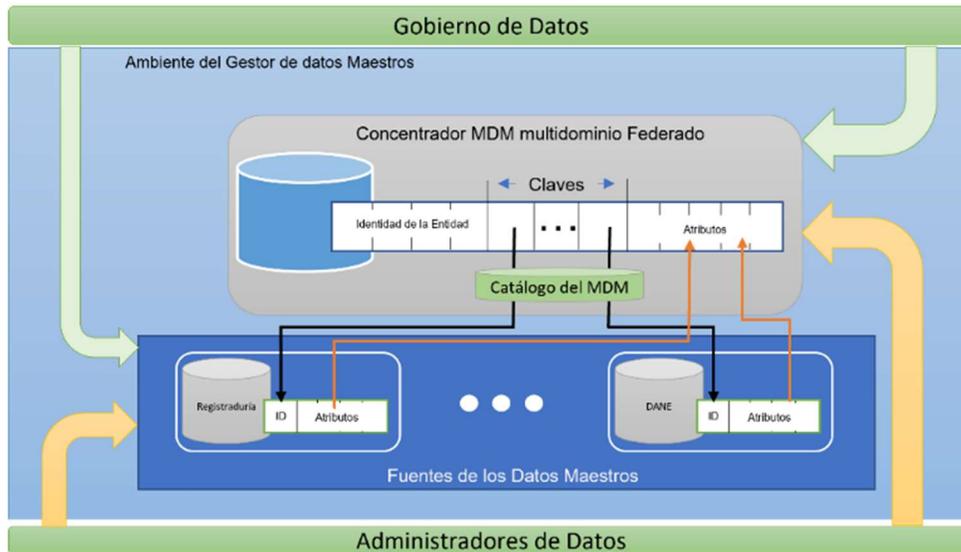


Figura 5. Arquitectura Federada. Una adaptación de (Allen & Cerro, 2015) con fuentes asociadas al proyecto.

5.3. Modelos de Datos

Los modelos de datos maestros que se presentan en esta sección corresponden a tres tipos de modelos que definen a cada dominio de datos: Personas, Empresas, y Lugares:

- Registros de Origen. Modelos de datos cuyas tablas definen y describen el registro de oro de cada entidad fuente.
- Registro de Referencia. Modelos de datos cuyas tablas definen y describen el registro de oro que se replicará desde cada entidad maestra al modelo de datos centralizado del Concentrador GDM del Gobierno Nacional.
- Modelos de Datos de Referencia. Modelos de datos cuyas tablas contiene datos de referencia internos o externos al Concentrador GDC, y cuyo propósito es complementar la descripción de las entidades maestras logrando así integridad referencial.

Para los **Registros de Referencia** de cada dominio, el estándar de nombramiento de tablas maestras es **hubd<abreviación dominio>_<complemento>** y se debe interpretar de la siguiente forma:

- Prefijo: **hubd<abreviación dominio>_**:
 - hub: concentrador
 - “d”: indica que es una tabla de un dominio específico

- <abreviación dominio>: abreviatura de tres caracteres del nombre del dominio al que pertenece la entidad maestra, p.ej: **per**: abreviatura del dominio persona.
- Complemento: <**complemento**> nombre de la tabla maestra o sustantivo que representa la información que se almacena en la tabla.

5.3.1. Dominio de Personas

- **Registro de Origen**

La Figura 6 muestra el modelo físico del Registro de Origen del dominio de Personas. La Tabla **rnecpersonas** proveerá los datos y metadatos que se almacenarán en el concentrador GDM del Gobierno Nacional. El estándar de nombramiento es el siguiente:

- **rnec**: sigla de Registraduría Nacional del Estado Civil – Entidad fuente
- **personas**: entidad maestra de la fuente

Información complementaria para el diseño de los modelos de este dominio:

En el desarrollo de este proyecto, la entidad fuente no proporcionó un diccionario de datos que permitiera conocer las tablas, atributos y tipos de datos del modelo de origen. Sin embargo, para lograr el objetivo del proyecto se buscó en internet la información más cercana posible al registro civil de nacimiento (<https://consultasrc.registraduria.gov.co:28080/ProyectoSCCRC/>, 2023)².

² El diseño del del modelo se hace por experiencia de los autores y la información de consulta disponible desde el Registro Civil.

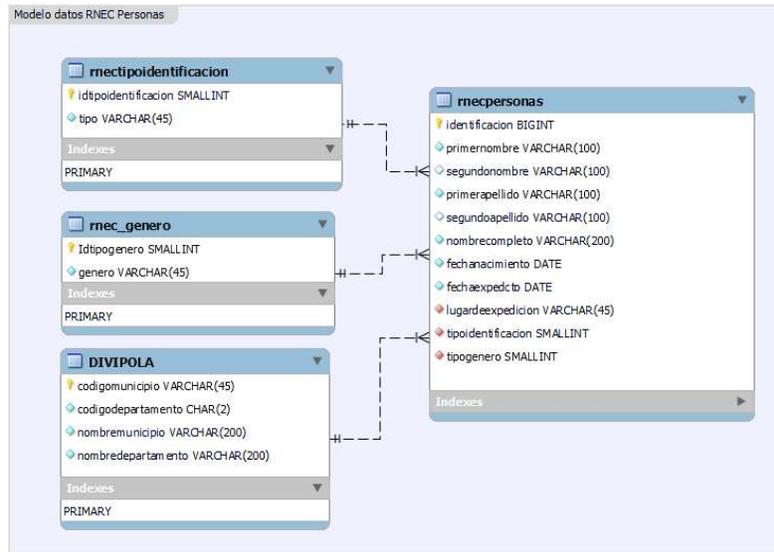


Figura 6. Modelo lógico del registro de origen RNEC Personas - Fuente: Registraduría Nacional de Estado Civil.

La Tabla 8 describe el diccionario de datos de la tabla maestra de Personas **rncpersonas** que provee la fuente de la RNEC.

Tabla 8. Diccionario de datos de la maestra de personas *rnecpersonas*.

Nombre Columna	Descripción Columna	Tipo de Dato de la Columna
Identificación	Identificación única de cada persona natural que está registrado civilmente en el estado colombiano	BIGINT
Primernombre	Primer Nombre de la persona como está en el registro civil de nacimiento	VARCHAR(100)
segundonombre	Segundo Nombre de la persona como está en el registro civil de nacimiento	VARCHAR(100)
primerapellido	Primer apellido de la persona como está en el registro civil de nacimiento	VARCHAR(100)
segundoapellido	Segundo apellido de la persona como está en el registro civil de nacimiento	VARCHAR(100)
nombrecompleto	Nombre completo conformada por primer nombre, segundo nombre, primer apellido, segundo apellido	VARCHAR(500)
fechanacimiento	Fecha de nacimiento registrada en el registro civil de la persona asociada a "identificación"	DATE
fechaexpedcto	Fecha asignada al momento del registro de la identificación	DATE
lugardeexpedicion	Lugar de expedición del documento de identificación	INT
tipoidentificacion	Tipos de documento definidos por RNEC	SMALLINT
tipogenero	Género de la persona identificada en el registro civil de nacimiento	SMALLINT

- **Registro de Referencia**

La Figura 7 muestra el modelo de datos de la entidad maestra de Personas que representa el dominio Personas. Está compuesto por tablas maestras (color azul) y tablas de referencia (color rosa). La tabla maestra **hubdper_maestrapersona** almacenará el registro de oro con información correspondiente a la identidad de cada ciudadano colombiano. La Registraduría Nacional del Estado Civil es la entidad fuente que certificará los datos de origen que se proveerían desde los sistemas ANI (Archivo Nacional de Identificación) y RCN (Registro Civil de Nacimiento).

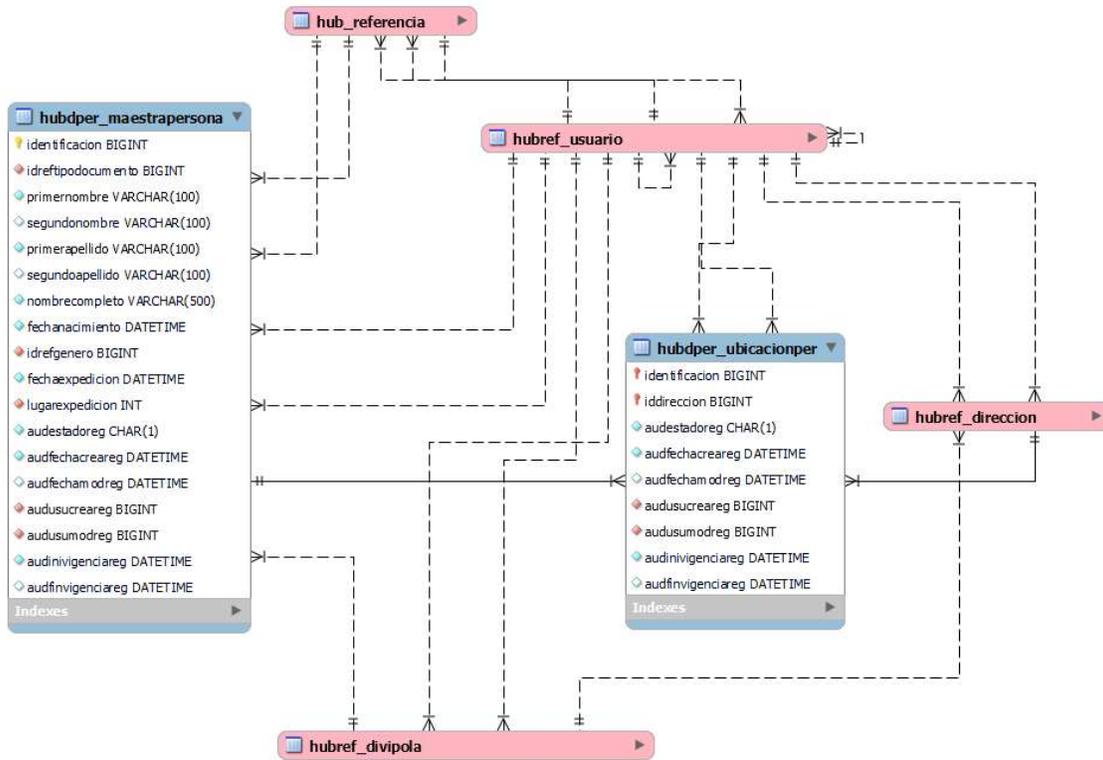


Figura 7. Modelo de datos de la tabla maestra de personas hubper_maestrapersona – dominio de Personas.

La tabla **hubper_ubicacionper** es una tabla maestra que se propone como un complemento a la descripción de identificación que entrega el registro de oro de la tabla **hubper_maestrapersona**. El dueño de esta tabla es el Concentrador GDM y almacena la relación de cada individuo con la tabla de referencia **hubref_direccion**.

La Tabla 9 presenta el diccionario de datos de la tabla maestra **hubper_maestrapersona**, En el diccionario no se muestran los datos de auditoría de las tablas ya que se definieron en una sección inicial de este capítulo.

Tabla 9. Diccionario de Datos Maestra de Persona – hubper_maestrapersona.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foranea	Sensibilidad
Identificación	Identificación única de cada persona natural que está registrado civilmente en el estado colombiano	BIGINT	No	Si	No	PII

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foranea	Sensibilidad
idreftipodocumento	Tipos de documento definidos por RNEC	SMALLINT	No	No	hub_referencia (idreferencia)	No
Primernombre	Primer Nombre de la persona como está en el registro civil de nacimiento	VARCHAR(100)	No	No	No	PII
Segundonombre	Segundo Nombre de la persona como está en el registro civil de nacimiento	VARCHAR(100)	Si	No	No	PII
Primerapellido	Primer apellido de la persona como está en el registro civil de nacimiento	VARCHAR(100)	No	No	No	PII
Segundoapellido	Segundo apellido de la persona como está en el registro civil de nacimiento	VARCHAR(100)	Si	No	No	PII
Nombrecompleto	Nombre completo conformada por primer nombre, segundo nombre, primer apellido, segundo apellido	VARCHAR(500)	No	No	No	PII
Fechanacimiento	Fecha de nacimiento registrada en el registro civil de la persona asociada a "identificación"	DATE	No	No	No	Sin información
Fechaexpedicion	Fecha asignada al momento del registro de la identificación	DATE	No	No	No	Sin información
Lugarexpediton	Lugar de expedición del documento de identificación	INT	No	No	hubref_divipola (código_municipio)	Sin información
Idrefgenero	Género de la persona identificada en el registro civil de nacimiento	SMALLINT	No	No	hub_referencia (idreferencia)	Sin información

La Tabla 10 presenta el diccionario de datos de la tabla maestra **hubdper_ubicacionper**, Las tuplas de esta tabla se crean tomando como referencia la relación entre persona y dirección que pueden proveer tablas maestras como **hubdemp_maestrarues**, y otras maestras futuras que contienen esta relación. En el diccionario no se muestran los datos de auditoría de las tablas ya que se definieron en una sección inicial de este capítulo.

Tabla 10. Diccionario de Datos Maestra de Persona – hubdper_ubicacionper.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foranea	Sensibilidad
Identificación	Identificación única de cada persona natural que está registrado civilmente en el estado colombiano	BIGINT	No	Si	hubdper_maestrapersona (identificación)	PII
Iddireccion	Tipos de documento definidos por RNEC	SMALLINT	No	No	Hubref_direccion (iddireccion)	No

5.3.2. Dominio de Empresas

- **Registro de Origen**

La Figura 8 muestra el modelo físico del Registro de Origen del dominio de Empresas. La tabla **ccruces** proveerá los datos y metadatos que se almacenaran en el concentrador de datos maestros del Gobierno Nacional. El estándar de nombramiento es el siguiente:

- **cc:** Abreviación de Cámara de Comercio – Entidad fuente
- **rues:** entidad maestra de la fuente

Información complementaria para el diseño de los modelos de este dominio:

- INSTRUCCIONES PARA DILIGENCIAR EL FORMULARIO DEL REGISTRO ÚNICO EMPRESARIAL Y SOCIAL RUES (Cámara de Comercio Bogotá, 2022).
- Ejemplo de consulta de datos RUES (eInforma, 2022)

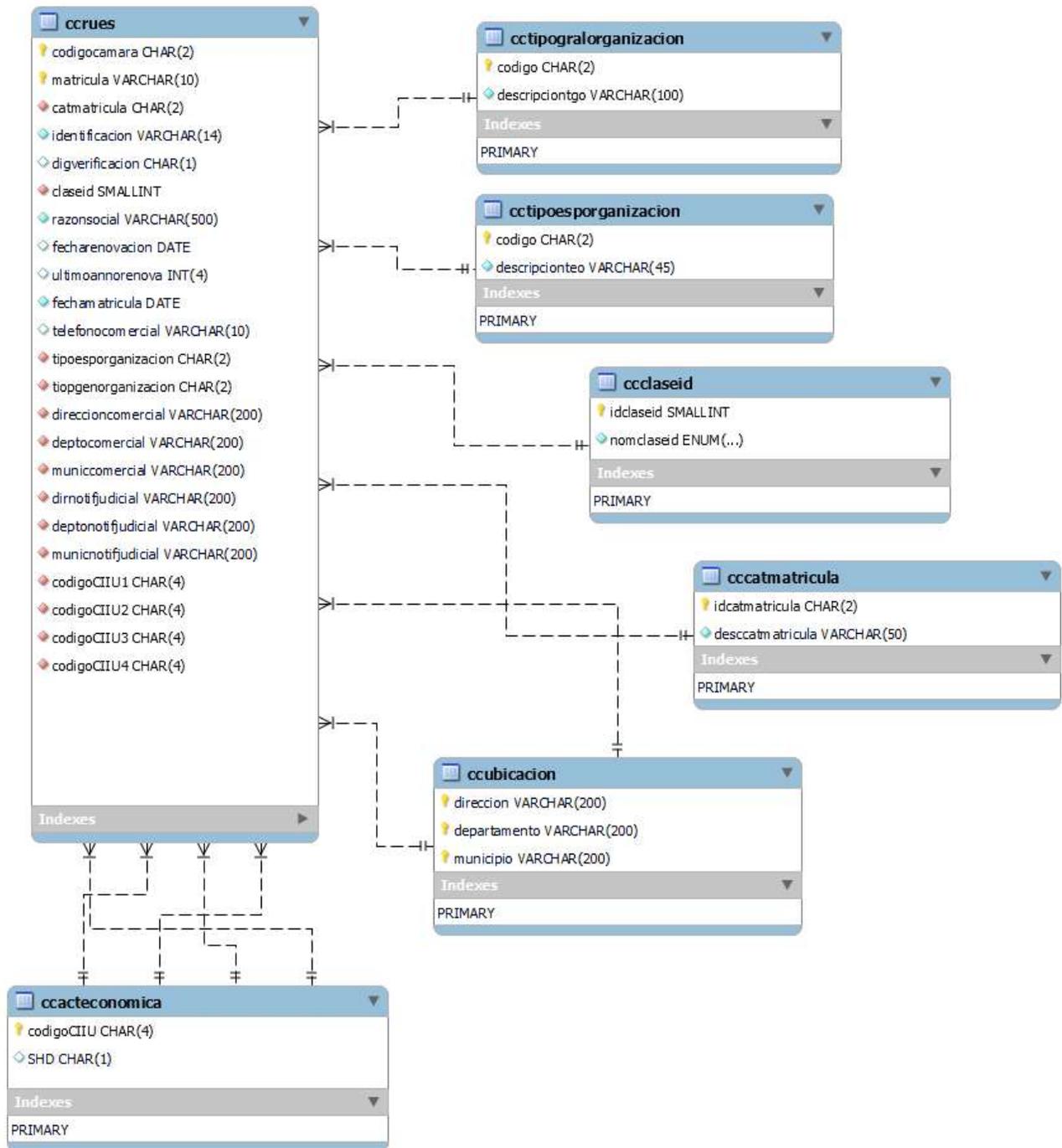


Figura 8. Modelo Lógico del Registro Origen **ccruces**. Fuente: Definida desde el diccionario de datos proporcionado por Confecámaras.

La Tabla 11 describe el diccionario de datos de la tabla maestra de Empresas **ccruces**.

Tabla 11. Diccionario de datos de la maestra de empresas ccruces.

Nombre Columna	Descripción Columna	Tipo de Dato de la Columna
codigocamara	Código de la Cámara de Comercio	CHAR(2)
matricula	Matricula del funcionamiento del negocio	VARCHAR(10)
catmatricula	La categoría de la matricula	CHAR(2)
Identificación	Número de Identificación o NIT de la persona natural o jurídica que está registrada en el RUES	VARCHAR(14)
Razonsocial	Nombre de la razón social de la empresa o persona jurídica registrada en el RUES	VARCHAR(500)
digverificacion	Digito de verificación	CHAR(1)
direccioncomercial	Dirección comercial de la persona natural o jurídica que está registrada en el RUES	VARCHAR(200)
claseid	Clase de identificación definida en el formulario RUES	SMALLINT
fecharenovacion	Fecha de renovación de la matricula	DATE
ultimoannorenova	Último año de renovación de la matricula	INT(4)
fechamatricula	Fecha de la matricula	DATE
telefonocomercial	Teléfono comercial	VARCHAR(10)
tipesporganizacion	Tipo específico de organización definido en el formulario RUES	CHAR(2)
tipogenorganizacion	Tipo general de la organización definido en el formulario RUES	CHAR(2)
deptocomercial	Departamento donde se encuentra ubicada la persona natural o jurídica que está registrada en el RUES	VARCHAR(200)
municcomercial	Municipio donde se encuentra ubicada la persona natural o jurídica que está registrada en el RUES	VARCHAR(200)
dirnotifjuridica	Dirección comercial de notificación jurídica que está registrada en el RUES	VARCHAR(200)
deptonotifjuridica	Departamento de notificación jurídica que está registrada en el RUES	VARCHAR(200)

Nombre Columna	Descripción Columna	Tipo de Dato de la Columna
municotifjuridica	Municipio de notificación jurídica que está registrada en el RUES	VARCHAR(200)
codigoCIU1	Código CIU de la Actividad Económica Principal de la empresa registrada en el RUES	CHAR(4)
codigoCIU2	Código CIU de la Actividad Económica Principal de la empresa registrada en el RUES	CHAR(4)
codigoCIU3	Código CIU de la Actividad Económica Principal de la empresa registrada en el RUES	CHAR(4)
codigoCIU3	Código CIU de la Actividad Económica Principal de la empresa registrada en el RUES	CHAR(4)

- **Registro de Referencia**

La figura 9 muestra el modelo de datos de la entidad maestra del RUES que representa el dominio Empresas. Está compuesto por la tabla maestra `hubdemp_maestrarues` y tablas de referencia `hub_referencia`, `hubref_direccion`, `hubref_acteconomica`, y `hubref_usuario`. La tabla maestra `hubdemp_maestrarues` almacenará el registro de oro con información correspondiente a la identidad de cada empresa registrada en Colombia. Confecámaras es la entidad fuente que certificará los datos de origen.

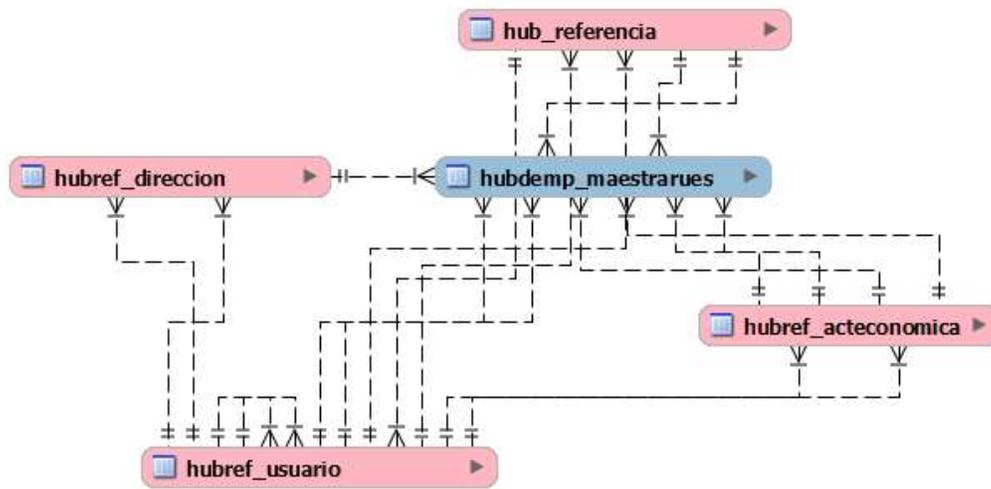


Figura 9. Modelo de datos de la entidad maestra de RUES en el ROR – dominio de Empresas.

La tabla 12 presenta el diccionario de datos de la tabla maestra **hubdemp_maestrarues**.

Tabla 12. Diccionario de Datos de la tabla maestra del RUES **hubdemp_maestrarues** – dominio Empresas.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave foránea	Sensibilidad
Identificación	Número de Identificación o NIT de la persona natural o jurídica que está registrada en el RUES	VARCHAR(14)	No	Si	No	PII
razonsocialnombre	Nombre de la razón social de la empresa o persona natural registrada en el RUES	VARCHAR(500)	No	No	No	PII
direccioncomercio	Dirección comercial de la empresa o persona jurídica	INT	No	No	hubref_direccion (iddireccion)	

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave foránea	Sensibilidad
	registrada en el RUES					
idrefclaseidentificacion	Clase se identificación definida en el formulario RUES	SMALLINT	No	No	hub_referencia (idreferencia)	
idreftipoesporganizacion	Tipo específico de organización definido en el formulario RUES	CHAR(2)	No	No	hub_referencia (idreferencia)	
CIIU1	Código CIIU de la Actividad Económica Principal de la empresa registrada en el RUES	CHAR(4)	No	No	hubref_acteconomica (claseciiu)	
CIIU2	Código CIIU de la Actividad Económica secundaria de la empresa registrada en el RUES	CHAR(4)	Si	No	hubref_acteconomica (claseciiu)	
CIIU3	Código CIIU de la Actividad Económica terciaria de la empresa registrada en el RUES	CHAR(4)	Si	No	hubref_acteconomica (claseciiu)	
CIIU4	Código CIIU de la Actividad Económica cuaternaria de la empresa	CHAR(4)	Si	No	hubref_acteconomica (claseciiu)	

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave foránea	Sensibilidad
	registrada en el RUES					

5.3.3. Dominio de Lugares

- **Registro de Origen**

La Figura 10 muestra el modelo físico del Registro de Origen del dominio de Lugares. La tabla **igacric** proveerá los datos y metadatos que se almacenaran en el concentrador de datos maestros del Gobierno Nacional. El estándar de nombramiento es el siguiente:

- **igac:** Abreviación de Instituto Agustín Codazzi – Entidad fuente
- **ric:** abreviación de Registro de Identificación Catastral

Información complementaria para el diseño de los modelos de este dominio: Diccionario de Datos: (IGAC-LADM_COL, 2022)

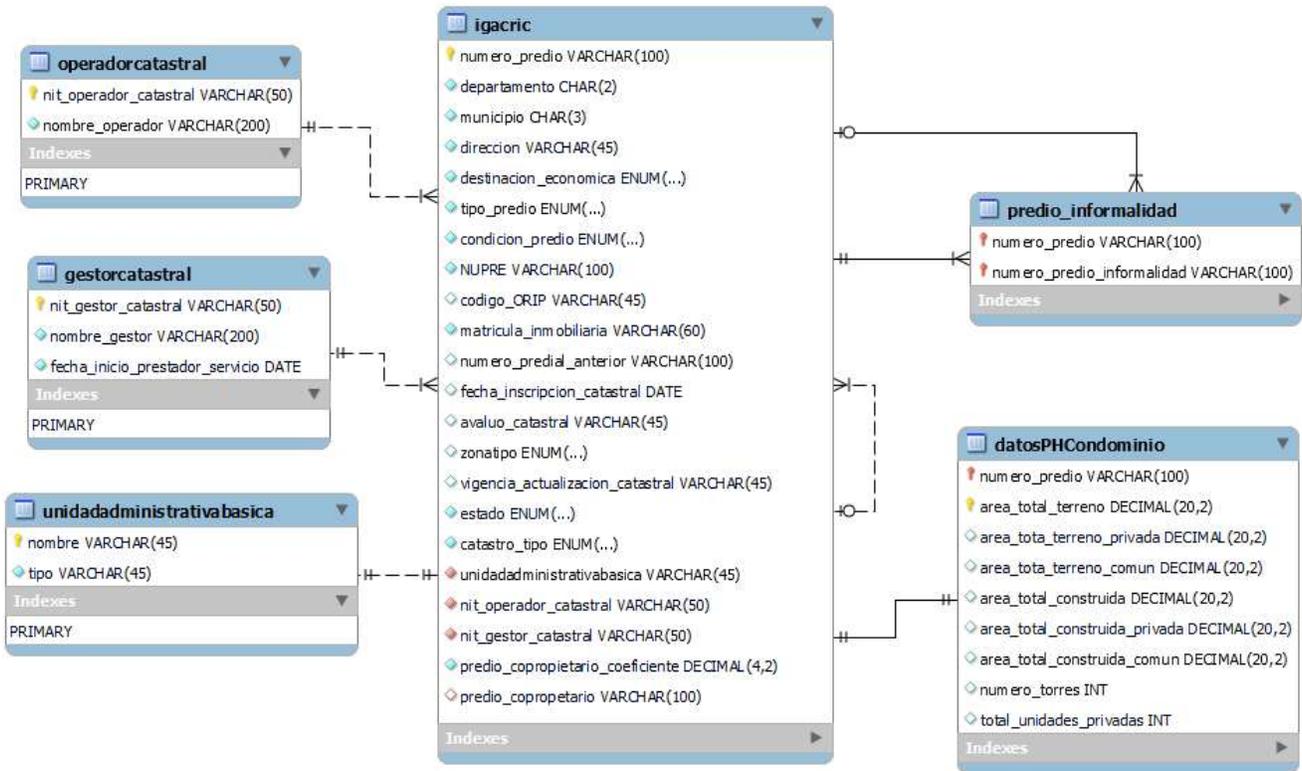


Figura 10. Modelo Físico del Registro Origen de Catastro – igacric. Fuente: IGAC (IGAC-Repositorio, 2022).

La Tabla 13 describe el diccionario de datos de la tabla maestra de Lugares **igacric**.

Tabla 13. Diccionario de datos de la maestra de lugares igacric.

Nombre Columna	Descripción Columna	Tipo de Dato de la Columna
numero_predio	Nuevo código numérico que se le asigna a cada predio y busca localizarlo inequívocamente en los documentos catastrales, según el modelo determinado por el IGAC	VARCHAR(100)
departamento	Departamento del predio	CHAR(2)

municipio	Municipio del predio	CHAR(2)
direccion	Dirección del predio	BIGINT
destinacion_economica	Destinación económica del predio	ENUM
tipo_predio	Corresponde al dato de tipo de predio incorporado en las bases de datos registrales.	ENUM
condicion_predio	Condición del predio definido por el IGAC	ENUM
NUPRE	Es un código homologado para identificar los inmuebles tanto en los sistemas de información catastral como registral.	VARCHAR(100)
codigo_ORIP	Circulo registral	VARCHAR(45)
matricula_Inmobiliaria	Matricula inmobiliaria	VARCHAR(60)
numero_predial_anterior	Anterior código numérico que se le asigna a cada predio y busca localizarlo inequívocamente en los documentos catastrales, según el modelo determinado por el IGAC	VARCHAR(100)
fecha_inscripcion_catastral	Es la fecha en la cual se abrió el Folio de matrícula inmobiliaria.	DATE
avaluo_catastral	Valor catastral del predio, obtenido mediante investigación y análisis estadístico del mercado inmobiliario y la metodología de aplicación correspondiente.	VARCHAR(45)
zona_tipo	Es la zona (rural o urbana) en la cual se encuentra localizado el predio	ENUM
vigencia_actualizacion_catastral	Fecha de vigencia de la actualización catastral	VARCHAR(45)
Estado	Se refiere a la situación actual en la que se encuentra una matrícula inmobiliaria.	ENUM
catastro_tipo	Tipo Catastro	ENUM
unidadAdministrativa	Nombre que recibe la unidad administrativa básica, en muchos casos toponímico, especialmente en predios.	VARCHAR(45)
nit_operador_catastral	Código NIT del operador catastral	VARCHAR(50)

nit_gestor_catastral	Código NIT del gestor catastral	VARCHAR(50)
predio_copropietario_coeficiente	Numero predio copropietario	DECIMAL (4,2)
predio_copropietario	Numero predio copropietario	VARCHAR(100)

- **Registro de Referencia**

La Figura 11 muestra el modelo de datos de la entidad maestra de Catastro que representa el dominio Lugares. Está compuesto por la tabla maestra `hubdlug_maestracatastro` y tablas de referencia `hub_referencia`, `hubref_direccion`, y `hubref_usuario`. La tabla maestra `hubdlug_maestracatastro` almacenará el registro de oro con información correspondiente a la identidad de cada empresa registrada en Colombia. Confecámaras es la entidad fuente que certificará los datos de origen.

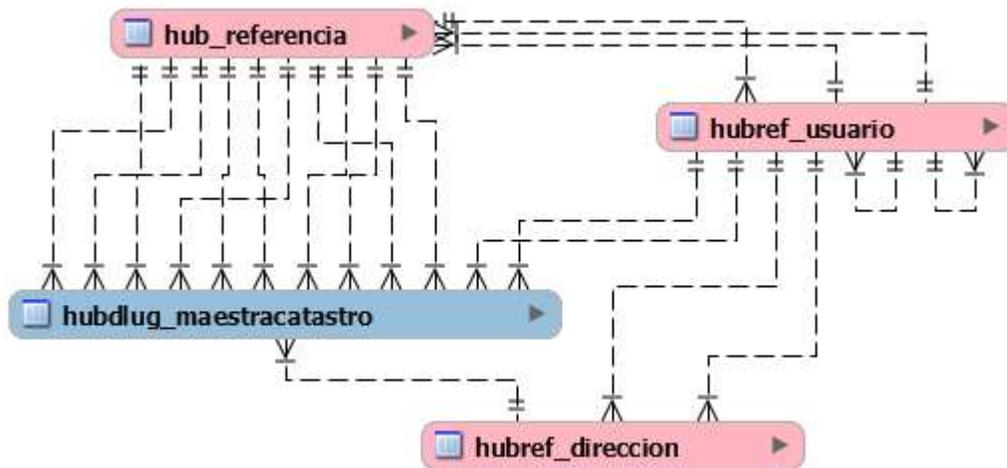


Figura 11. Modelo de Datos de la maestra de Catastro `hubdlug_maestracatastro` – dominio Lugares.

La tabla 14 presenta el diccionario de datos de la tabla maestra `hubdlug_maestracatastro`.

Tabla 14. Diccionario de Datos de la tabla maestra de Catastro hubdlug_maestracatastro – dominio Lugares.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave foránea	Sensibilidad
Numeropredial	Nuevo código numérico que se asigna a cada predio y busca localizarlo inequívocamente en los documentos catastrales, según el modelo determinado por el IGAC	VARCHAR(100)	No	Si	No	No
Codigodireccion	Ubicación registrada del predio	BIGINT	No	No	hubref_direccion (iddireccion)	No
Idrefdestinacioneconomica	Destinación económica del predio	ENUM	No	No	hub_referencia (idreferencia)	No
Idreftipopredio	Corresponde al dato de tipo de predio incorporado en las bases de datos registrales	ENUM	No	No	hub_referencia (idreferencia)	No
Idrefcondicionpredio	Condición del predio definido por el IGAC	ENUM	No	No	hub_referencia (idreferencia)	No
Nupre	Código homologado para identificar los inmuebles tanto en los sistemas de información catastral como registral	VARCHAR(100)	No	No	No	No
Codigoorip	Circulo registral	VARCHAR(45)	No	No	No	No
Matriculainmobiliaria	Matricula inmobiliaria	VARCHAR(60)	No	No	No	No
Numeropredialanterior	Número de predial anterior solo en casos en que al mismo predio se le haya asignado un nuevo número	VARCHAR(100)	Si	No	No	No
Fechainscripcioncatastral	Es la fecha en la cual se abrió el	DATE	No	No	No	No

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave foránea	Sensibilidad
	Folio de matrícula inmobiliaria.					
Avaluocatastral	Valor catastral del predio, obtenido mediante investigación y análisis estadístico del mercado inmobiliario y la metodología de aplicación correspondiente.	VARCHAR(45)	No	No	No	No
Idreftipozona	Es la zona (rural o urbana) en la cual se encuentra localizado el predio	ENUM	No	No	hub_referencia (idreferencia)	No
Vigenciaactualcatastral	Fecha de vigencia de la actualización catastral	VARCHAR(45)	Si	No	No	No
Idrefestado	Se refiere a la situación actual en la que se encuentra una matrícula inmobiliaria.	ENUM	No	No	hub_referencia (idreferencia)	No
Idrefipocatastro	Tipo Catastro	ENUM	No	No	hub_referencia (idreferencia)	No
Idrefunidadadministrativa	Nombre que recibe la unidad administrativa básica, en muchos casos toponímico, especialmente en predios.	VARCHAR(45)	No	No	hub_referencia (idreferencia)	No
Idrefoperadorcatastral	Codigo NIT del operador catastral	VARCHAR(50)	No	No	hub_referencia (idreferencia)	No
Idrefgestorcatastral	Codigo NIT del gestor catastral	VARCHAR(50)	No	No	hub_referencia (idreferencia)	No

5.4. Modelos de Datos de Referencia

Los modelos de referencia relacionan tablas que almacenan datos de referencia y cuya función es complementar la funcionalidad de las entidades maestras definidas en cada dominio³. Son llamados datos de referencia porque no pertenecen a los datos propios de la función maestra.

El estándar de nombramiento de tablas de referencia presentadas en la figura 12 se presentan en dos formatos: uno, **hubref_<nombretabla>** el cual se debe interpretar de la siguiente forma:

- Prefijo: **hubref_**:
 - o hub: concentrador
 - o ref: indica que es una tabla de referencia
- Complemento: **<nombretabla>** nombre de la tabla o sustantivo que representa la información que se almacena en la tabla.

El otro es el formato **hub_<nombretabla>**, el cual se definió solo para tres tablas creadas específicamente para manejar de forma dinámica los atributos de referencia con valores que provienen normalmente de listas de valores pequeñas y poco variables en el tiempo.

5.4.1. Tabla de referencia – hubref_direccion

La figura 12 muestra el modelo de datos de la entidad de referencia Dirección. Está compuesto por la tabla de referencia **hubref_direccion** acompañada de las tablas de referencia **hubref_divipola**, **hubref_departamento**, **hubref_comuna**, y **hubref_usuario**. Los datos son extraídos al interior del Concentrador GDM, específicamente de las tablas **hubdemp_maestrarues** y **hupdlug_maestracatastro**, las cuales se capturan en el registro de referencia de dirección. Estos son datos propios del Concentrador GDM, para los cuales se aplican reglas de supervivencia de tal forma que se logre un registro de referencia único y exacto para ser almacenado en la tabla **hubref_direccion**.

³ La funcionalidad de los datos, en general, se refiere a la función que estos datos tienen en conjunto para apoyar la calidad de los datos en la transaccionalidad y en la toma de decisiones

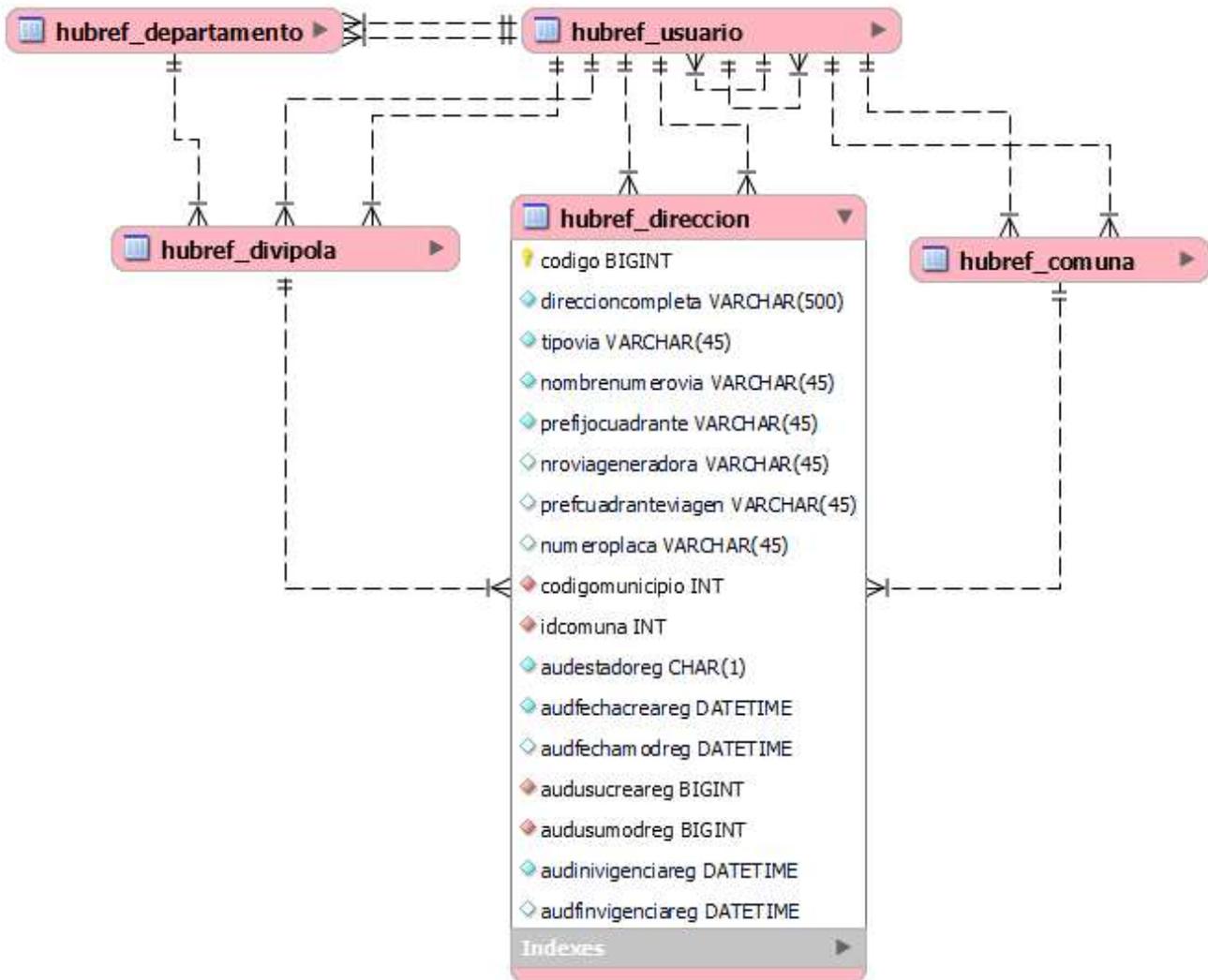


Figura 12. Modelo de Referencia de Dirección – hubref_direccion.

La tabla 15 presenta el diccionario de datos de la tabla maestra **hubref_direccion**.

Tabla 15. Diccionario de Datos de la tabla de referencia de Dirección hubref_direccion.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
Código	Código único de registro en la tabla direcciones	BIGINT	No	Si	No	No
Direccioncompleta	Dirección completa	VARCHAR(500)	No	No	No	No

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
Tipovia	Tipo de vía sobre la cual está ubicada la dirección	VARCHAR(45)	No	No	No	No
Nombrenumerovia	Nombre o numero de la vía sobre la cual está ubicada la dirección	VARCHAR(45)	No	No	No	No
Prefijocuadrante	Prefijo de cuadrante de la dirección	VARCHAR(45)	No	No	No	No
Nroviageneradora	Número de vía generadora	VARCHAR(45)	Si	No	No	No
cuadranteviageneradora	Prefijo de cuadrante de la vía generadora sobre la cual está ubicada la dirección	VARCHAR(45)	Si	No	No	No
Numeroplaca	Número de la placa de la dirección	VARCHAR(45)	Si	No	No	No
Idcomuna	Código de la comuna a la cual pertenece la dirección	INT	No	No	hubref_comuna (idref_comuna)	No
Codigomunicipio	Nombre del Municipio - Departamento de la dirección	INT	No	No	hubref_divipola (código_municipio)	No

Las tablas 16, 17, y 18 presentan el diccionario de datos de la tablas de referencia **hubref_divipola** y **hubref_departamento**, y **hubref_comuna**.

Tabla 16. Diccionario de Datos de la tabla de referencia hubref_divipola.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
codigo_municipio	Identificador único del municipio - creado por el DANE	INT	No	Si	No	No
nombre_municipio	Nombre del municipio - creado por el DANE	VARCHAR(100)	No	No	No	No
Tipo	Categoría de municipio - creado por el DANE	VARCHAR(45) NOT NULL	No	No	No	No
codigodepartamento	Código departamento - creado por el DANE - clave foránea	INT	No	No	Si	No

Tabla 17. Diccionario de Datos de la tabla de referencia hubref_departamento.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
codigodepartamento	Identificador único de departamento asignado por el DANE	INT NOT NULL	No	Si	No	No
Nombre	Nombre del departamento asignado por el DANE	VARCHAR(100)	No	No	No	No

Tabla 18. Diccionario de Datos de la tabla de referencia hubref_comuna.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
idref_comuna	Código único de la comuna - definida por cada municipio	INT	No	Si	No	No
descripcion	Nombre único de la comuna - definida por cada municipio	VARCHAR(100)	No	No	No	No

5.4.2. Tabla de referencia – hubref_usuario

La figura 13 muestra el modelo de datos de la entidad de referencia Usuario. Está compuesto por la tabla de referencia **hubref_usuario**. Los datos son capturados al configurar la seguridad del sistema en el Concentrador GDM, por lo tanto, son datos propios del Concentrador GDM.

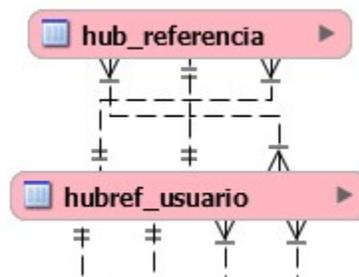


Figura 13. Modelo de Referencia de Usuario – Tabla hubref_usuario.

La tabla 19 presenta el diccionario de datos de la tabla de referencia **hubref_usuario**.

Tabla 19. Diccionario de Datos de la tabla de referencia hubref_usuario.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
Idusuario	Identificación única del usuario autorizado del concentrador GDM	BIGINT UNSIGNED	No	Si	No	Si
nombrecompleto	Nombre completo del usuario autorizado del concentrador GDM	VARCHAR(100)	No	No	No	No
email	Correo electrónico del usuario autorizado del concentrador GDM	VARCHAR(45)	No	No	No	Si
contrasenna	Contraseña del usuario autorizado del concentrador GDM	VARCHAR(2000)	No	No	No	Si
idrefestadousuario	Estado del usuario (activo, inactivo)	BIGINT	No	No	hub_referencia (idreferencia)	No
idrefrolgobierno	Rol asignado dentro del gobierno del concentrador GDM	BIGINT	No	No	hub_referencia (idreferencia)	No

5.4.3. Tabla de referencia – hub_referencia

La figura 14 muestra el modelo de datos de la entidad de referencia ‘Referencia’, la cual se diseñó como una tabla de metadatos donde se definen listas de valores de referencia. Está compuesto por la tabla de referencia **hub_tiporeferencia** y **hub_valorreferencia**. Los datos son capturados al configurar diferentes valores de referencia que garantizan la integridad referencial de las tablas maestras y de referencia. Por lo tanto, son datos propios del Concentrador GDM.

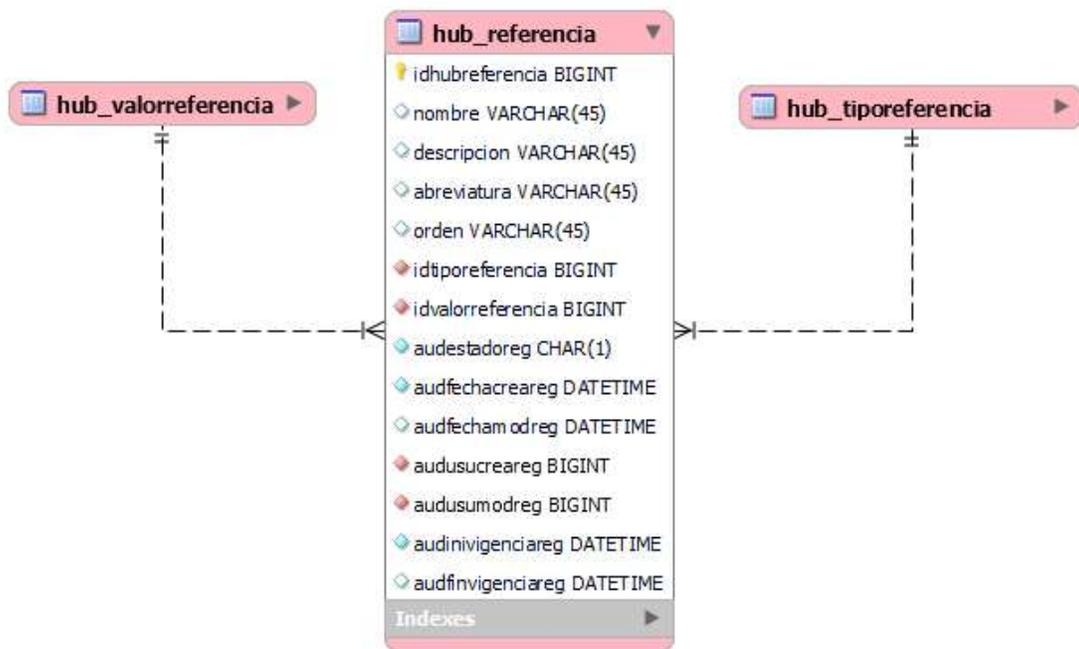


Figura 14. Modelo de Referencia – Tabla hub_referencia.

La tabla 20 presenta el diccionario de datos de la tabla de referencia **hub_referencia**.

Tabla 20. Diccionario de Datos de la tabla de referencia hub_referencia.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
idhubreferencia	Un identificador único del recurso de datos	BIGINT	No	Si	No	No
nombre	Nombre de la referencia que será usada por una tabla maestra o de referencia	VARCHAR(45)	Si	No	No	No
descripcion	Descripción de la referencia	VARCHAR(45)	Si	No	No	No
abreviatura	Abreviatura de la referencia	VARCHAR(45)	Si	No	No	No
orden	El número que indica la Posición en que se ubica en la tabla que la referencia	VARCHAR(45)	Si	No	No	No
idtiporeferencia	Código que hace referencia al tipo de	BIGINT	No	No	Si	No

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
	referencia de la tabla hub_tiporeferencia					
idvalorreferencia	Código que hace referencia al valor de referencia de la tabla hub_valorreferencia	BIGINT	No	No	Si	No

La tabla 21 presenta el diccionario de datos de la tabla de referencia **hub_tiporeferencia**.

Tabla 21. Diccionario de Datos de la tabla de referencia hub_tiporeferencia.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
nombrereferencia	Nombre del tipo de referencia	VARCHAR(100)	No	No	No	No
descripcionref	Descripción extendida del tipo de referencia: a que aplica, quien la debe usar, etc.	VARCHAR(500)	No	No	No	No
nombrecampo	Nombre del campo o etiqueta que representará el uso del campo cuando se requiera en diferentes tablas	VARCHAR(200)	No	No	No	No
tipodato	Tipo del dato que definirá el tipo de valor que será almacenado	VARCHAR(100)	No	No	No	No
longitudvalor	La longitud máxima del valor que debe almacenar el tipo de dato cuando el campo sea usado	VARCHAR(100)	No	No	No	No
presentacioncampo	Se refiere a la forma como el campo debe ser presentado para ser usado	VARCHAR(200)	No	No	No	No

La tabla 22 presenta el diccionario de datos de la tabla de referencia **hub_valorreferencia**.

Tabla 22. Diccionario de Datos de la tabla de referencia hub_tiporeferencia.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
idvalorreferencia	Un identificador único del recurso de datos	BIGINT	No	Si	No	No
nombre	Nombre del valor de referencia	VARCHAR(45)	No	No	No	No
valor	Valor de referencia	VARCHAR(45)	No	No	No	No

5.5. Ciclo de Vida de los Datos Maestros

El ciclo se activa cada vez que una fuente de datos maestros, de un dominio específico, crea o modifica la estructura de datos o alguno de los valores que será almacenado en el Concentrador GDM. Por ejemplo, para el dominio de personas, el evento es lanzado por la Registraduría del Estado Civil – Dominio de Personas; para el dominio de empresas, el evento es lanzado por Confecámaras, y para el dominio de Lugares, el evento es lanzado por el IGAC. Las responsabilidades de los diferentes roles definidos en el marco del gobierno de datos se definen a continuación.

- **Nivel táctico.** Los responsables del diseño, gestión y mejoramiento de este proceso y sus subprocesos. Además, cuando el Concentrador GDM entre en la fase de implementación, cada uno de los subprocesos del ciclo de vida más otros requeridos por la solución tecnológica deberán ser implementados como canalizadores (pipelines).
- **Nivel Operativo.** En el ambiente de producción, los equipos del nivel operativo son los responsables de ejecutar los trabajos definidos para que de forma automática ejecutarán los flujos de los diferentes subprocesos. La Figura 15 muestra las actividades que forma parte del flujo del ciclo de vida de los datos maestros.

La figura 15 muestra las actividades que forma parte del flujo del ciclo de vida de los datos maestros.

Actividades del proceso

- **Decisión: ¿repositorio virtual con réplica del Registro de Oro (RO)?**

El ciclo de vida inicia validando el tipo de evento que recibe para acceder a una fuente de origen o replicar fuente de origen en registro de oro del Concentrador GDM. Esto es:

Si no hay réplica, se activa la actividad de **Habilitar consulta del registro de oro de la fuente de origen**, en la cual se define el acceso a registro de oro localizado en un repositorio virtual que utiliza un índice maestro para apuntar a los registros originales en sus conjuntos de datos de origen que representan el registro de oro. En este caso, el

Concentrador GDM se comporta como una especie de “pasarela” que controla el acceso de diferentes entidades consumidoras usando diferentes reglas de presentación.

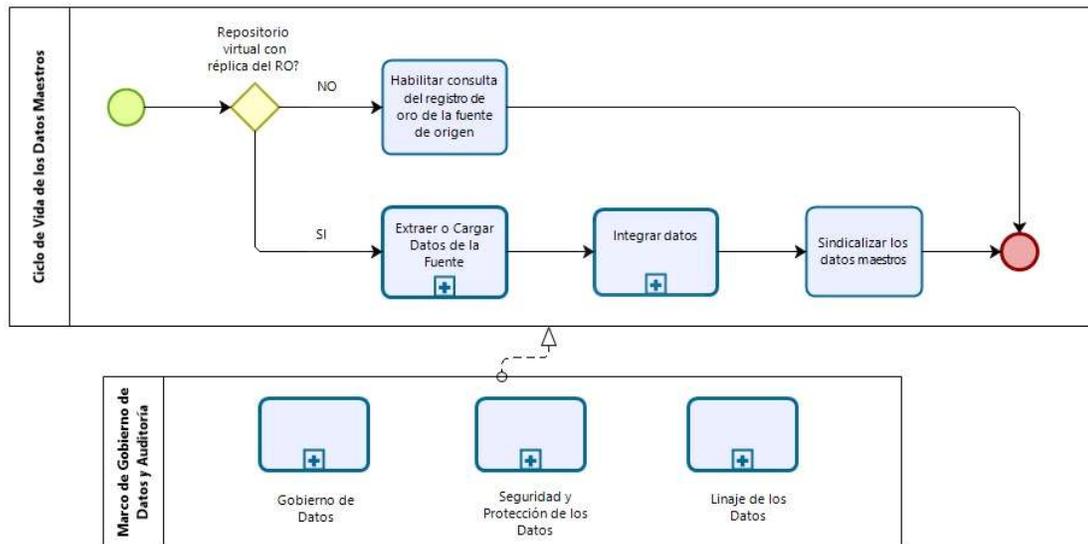


Figura 15. Proceso del ciclo de vida de la gestión de datos maestros Multi-Dominio híbrido. (Fuente: Elaboración Propia)

Si hay réplica se realizan las siguientes actividades:

- **Extraer o cargar datos de la fuente**

Este subproceso se activa cuando hay un acuerdo entre la entidad fuente y el Concentrador GDM a través del X-ROAD para hacer réplica del registro de oro de las entidades maestras o entidades de referencia definidas en cada dominio. La figura 16 muestra el flujo de trabajo de este subproceso el cual está soportado por los subprocesos de Gobierno de Datos, Seguridad y Protección de Datos, y el linaje de datos.

El proceso inicia una vez la entidad fuente ha certificado sus entidades maestras y de referencia, se genera un evento para que el Concentrador GDM a través del X-ROAD proceda a la extracción de entidades maestras y de entidades de referencia, de acuerdo con el período de actualización de datos acordados entre las partes. La carga es un proceso excepcional que se podría convenir en algún momento y por eso se considera como una posible acción. Una vez se haga la extracción/carga se procede a copia de los datos extraídos al área de llegada para luego aplicarles reglas de detección y sincronización según sea requerido por cada uno de los dominios.

Específicamente la sincronización se requiere para alistar los datos para la resolución, vinculación y consolidación de entidades de forma periódica (por ejemplo, todas las noches). La sincronización se configura con períodos de tiempo exactos para que los gestores de datos maestros o de referencia de las fuentes no acumulen datos de diferentes períodos; así estos gestores podrán reconocer qué datos ya están almacenados en el Concentrador GDM con sus correspondientes metadatos requeridos para la sincronización. Los entornos de sincronización actuales presentan dos alternativas: La primera, a través de una representación física común del registro de oro, y la segunda mediante el uso de un enfoque virtual o híbrido el cual siempre proporciona los registros de origen más recientes. Así, los datos se publican o sincronizan según el estilo de arquitectura. Finalmente se crean las tablas de extracción que permitirán que el Concentrador GDM inicie la gestión de los datos maestros y de referencia teniendo en cuenta la actualización de su catálogo de datos maestros. Las acciones que se ejecuten en el flujo dependerán de los acuerdos definidos con las entidades por ley o decreto u otro instrumento que lo formalice legalmente.

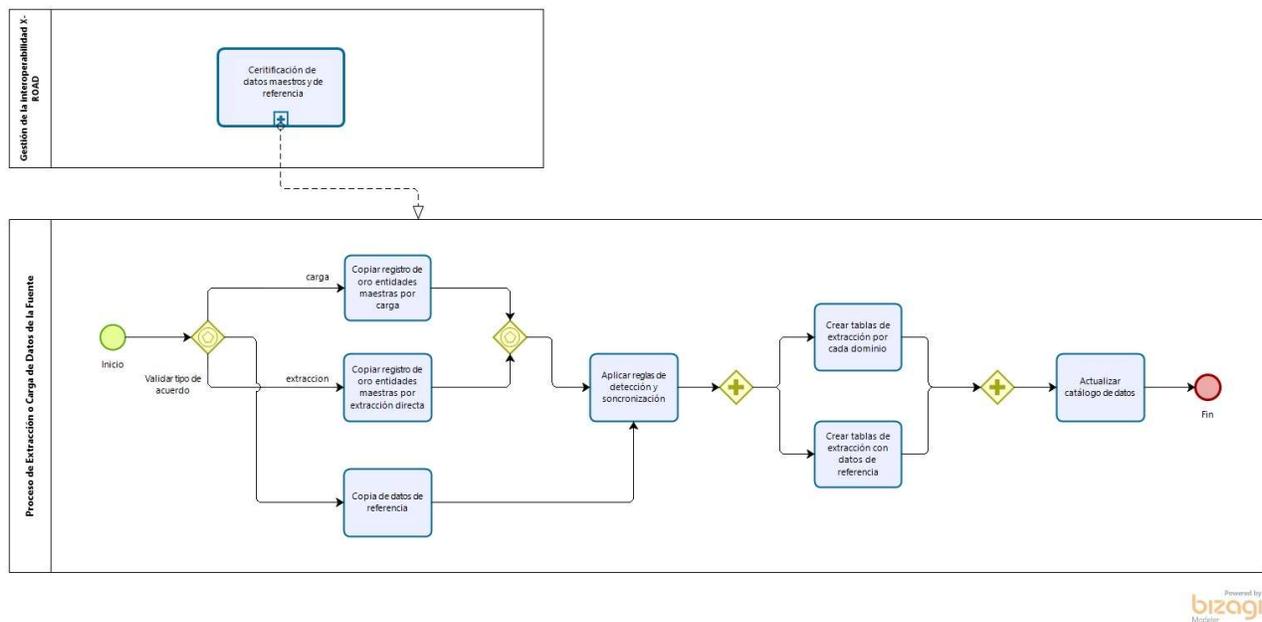


Figura 16. Flujo de trabajo para gestionar la extracción o carga de los datos maestros o de referencia de origen. (Fuente: Elaboración Propia)

La tabla 23 presenta las características que deben acompañar el registro de oro del registro de origen (ROO).

Tabla 23. Características del registro de oro de origen (ROO).

Fuente de verdad	Entidad: <entidad dueña o creadora de los datos fuente> Tabla: <nombre tabla certificada que contiene el registro de oro de origen>.
------------------	---

Periodicidad	Diaria, Semanal, Mensual o en tiempo real
Metadatos tabla/archivo de origen	Formato de archivo Fecha creación Fecha última modificación Nombre usuario certificador de los datos de origen
Metadatos del registro de oro de origen	Atributos del registro de origen
Metadatos auditoria por cada tupla del registro de oro de origen	Fecha creación Fecha última modificación Evento sobre el registro (N:nuevo, M:modificado, Anulado) Nombre usuario certificador del evento
Tupla del registro de oro	Conjunto de datos del registro de oro de la tabla maestra de origen

- **Integrar los datos**

La integración de datos es un proceso que se realiza siempre y cuando varias fuentes certificadas aporten a la resolución de entidad del registro de oro en cada dominio. Este proceso se realiza con las diferentes entidades maestras y de referencia que ha sido replicadas en la actividad anterior. Para lograr esto se requiere aplicar diferentes reglas de calidad y así garantizar que cada registro de oro es un objeto confiable en el Concentrador GDM. La figura 17 muestra el subproceso de Integrar Datos. Este subproceso se divide en dos grandes bloques para el aseguramiento de la calidad. En el primero se aplican reglas básicas de exactitud, completitud, consistencia, oportunidad, unicidad, integridad, vigencia, precisión, y conformidad, las cuales hacen que el registro de oro de la fuente sea correctamente almacenado en el Concentrador GDM. La descripción de cada una de estas tareas de aplicación de reglas se define en capítulo [8. Aseguramiento de la Calidad de la Gestión de Datos Maestros](#). Estas reglas de calidad deben aplicarse para lograr tres acciones:

- **Prevención.** Evitar que datos incorrectos o erróneos sean almacenados en la base de datos del Concentrador GDM
- **Detección.** Tomar un enfoque proactivo para identificar datos incorrectos que ya están en las bases de datos del Concentrador GDM.
- **Reparación.** Arreglar o borrar los datos incorrectos.

La ejecución de estas tareas genera un reporte de auditoría y activará una serie de registros en el histórico de cambios. El segundo bloque parte de una preparación inicial de las tablas maestras por cada dominio. Luego se procede a aplicar reglas de emparejamiento, mapeo, enriquecimiento y supervivencia según sea el caso

requerido por cada dominio. Estas tareas también actualizarán el reporte de auditoría y los datos históricos de cambios. Una vez culmine este conjunto de tareas, se actualizan las entidades maestras y se genera un modelo de verdad o certificado del registro de oro como un conjunto único de elementos de datos maestros relacionados con una sola instancia de una entidad.

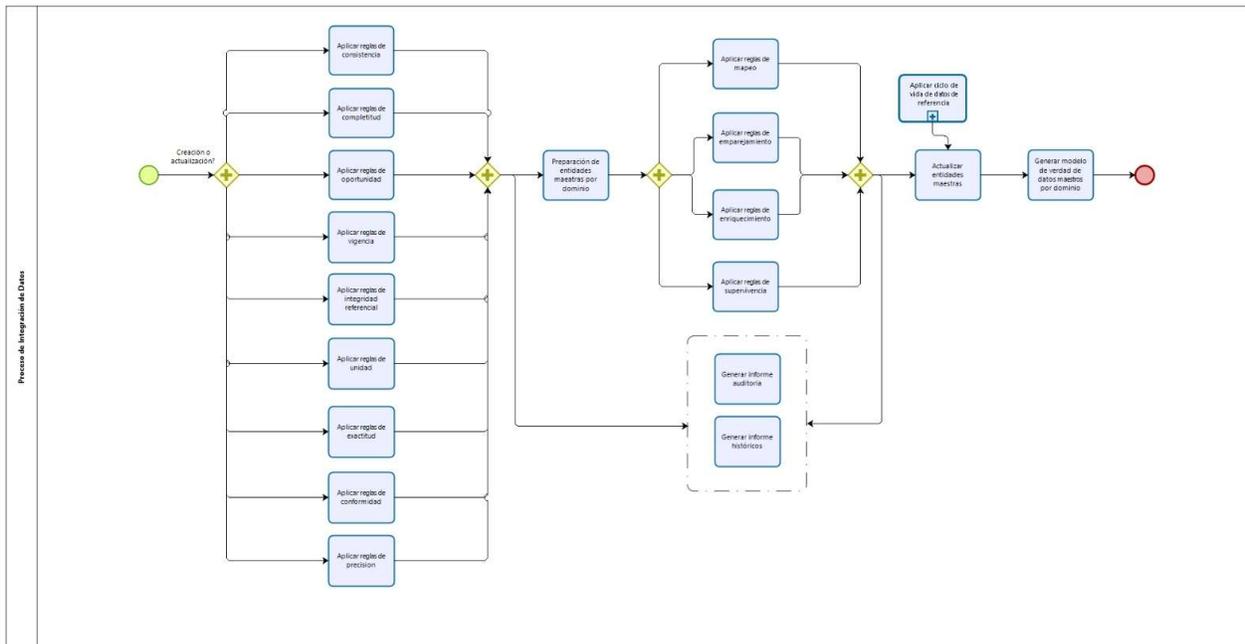


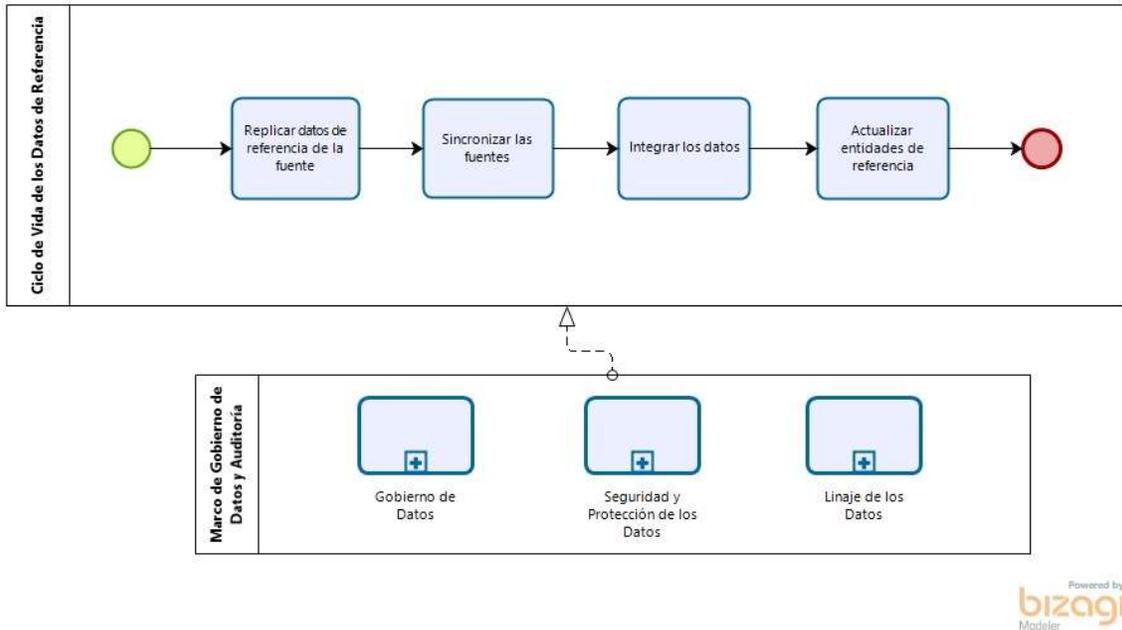
Figura 17. Subproceso de Integración de Datos. (Fuente: Elaboración Propia)

- **Sindicalizar los datos maestros**

Una vez se crean los datos maestros y de referencia y se certifican desde cada uno de los dominios del Concentrador GDM, se procede a definir y a administrar vistas virtuales de los datos maestros que serán consultados por las entidades consumidoras. Esto es posible proporcionando servicios de sindicación de datos para definir para cada entidad consumidora qué requiere cada entidad consumidora, quién puede ver dicha información, y cómo se presenta esa información. La sindicación elimina la complejidad de proveer la misma información a múltiples fuentes y proporciona un único punto de acceso desde diferentes vistas previamente definidas.

5.6. Ciclo de vida de los datos de referencia

La Figura 18 muestra el ciclo de vida de los datos de referencia. Los datos de referencia pueden venir de fuentes externas o son creados desde el Concentrador GDM.



Powered by
bizagi
Modeler

Figura 18. Ciclo de vida de los datos de referencia. (Fuente: Elaboración Propia.)

Actividades del proceso:

Una vez la plataforma de interoperabilidad X-ROAD ha validado las mínimas características de los acuerdos realizados entre las entidades proveedoras y el Concentrador GDM, el ciclo de vida de los datos de referencia inicia duplicando los datos de referencia de la fuente teniendo en cuenta los datos de certificación definidos en la tabla 3. Este proceso está soportado por los subprocesos de Gobierno de Datos, Seguridad y Protección de los Datos, y el linaje de datos.

Cuando la entidad de referencia se apoya en una única fuente externa (p.ej., Actividad Económica – Entidad DANE) el proceso continúa a la última tarea de actualizar datos de referencia en el Concentrador GDM. Por el contrario, cuando una entidad de referencia proviene de varias fuentes externas para alimentar una sola entidad de referencia (p.ej., Dirección – Entidades: IGAC, Confecámaras, Supersalud, etc.) se debe proceder a la sincronización y posteriormente a realizar la integración realizando un subproceso como se muestra en la figura

18, donde se aplican reglas de calidad básicas y avanzadas para poder lograr el registro único de la entidad de referencia. En este caso se evalúa si es posible la consolidación de un solo "registro único" que seleccione valores de atributos "sobrevivientes" entre registros de diferentes fuentes de origen vinculadas al modelo de datos de las entidades maestras definidas en el Concentrador GDM. Cualquier fuente externa que alimenta las entidades de referencia debe ser validada y aprobada por el Administrador de datos del Concentrador GDM. Una vez se han transformado los datos de la(s) fuente (s) se generan los datos que van a ser cargados en el modelo de referencia destino del Concentrador GDM.

6. Gobierno para la Gestión de Datos Maestros

En la gestión de datos maestros Multi-Dominio, el gobierno es entendido como el componente transversal que cruza los dominios de datos. En otras palabras, es un componente estratégico común a diferentes dominios, así como la calidad, la seguridad de los datos, a la gestión de metadatos en un estilo de arquitectura híbrida que facilite la sincronización del concentrador de datos maestros con sus fuentes.

El marco de trabajo que se define para la GDM del Gobierno Nacional se basa en el Plan Nacional de Infraestructura de Datos - Documento Técnico y Hoja de Ruta (DNP-PNID, 2021), subsección 6.3. Aprovechamiento de Datos – 6.3.2. Gobierno de Datos, la cual define:

- “Para la gestión del ciclo de vida de los datos, y en consecuencia alcanzar su aprovechamiento es necesario contar con un gobierno de datos que permita garantizar la calidad del dato, la seguridad y privacidad de los datos, hacer un apropiado almacenamiento y consolidar un proceso de documentación sobre cada fase del ciclo de vida del dato, aunado a una gestión adecuada de los metadatos.
- La óptima gestión de datos requiere de un marco que acoja un gobierno de datos, entendido como el ejercicio de diseñar, controlar y monitorizar todo lo relativo a los datos desde un enfoque holístico, en el que participen los implicados, desde el gobierno y el departamento de TI hasta un consejo de gestión de datos que represente a las partes interesadas.
- El propósito del gobierno de datos es asegurar que la gestión de los datos se cumpla con los propósitos misionales de las entidades públicas o empresas, buscando obtener una reducción de costos en lo que respecta a su gestión y a su protección (en lo que respecta al cumplimiento de la normativa y a la preservación de la privacidad).”

Además, La gobernanza definida para el Concentrador GDM del Gobierno Nacional toma como base en el decreto 1389 de 2022 (Función Pública, 2022) en el ARTÍCULO 2.2.24.2.1, Modelo de Gobernanza de la Infraestructura de Datos, el cual define el como “el conjunto de elementos políticos, técnicos, legales y organizacionales que permiten la articulación de actores, instancias, normas, políticas, planes, programas, estrategias, metodologías, compromisos, procesos y procedimientos para implementar, fortalecer, gestionar y manejar la infraestructura

de datos, con la finalidad de generar valor público, social y económico, a través de los datos.” Específicamente, en esta sección se desarrolla el primer objetivo definido en el ARTÍCULO 2.2.24.2.2. Objetivos del Modelo de Gobernanza de la Infraestructura de Datos: “Definir los actores, instancias, normas, políticas, planes y proyectos, para implementar, fortalecer, gestionar y dar sostenibilidad a la infraestructura de datos.”

6.1. Gobierno de Datos multidominio

El gobierno de datos multidominio se define para la arquitectura híbrida definida en el capítulo cuatro. En este tipo de arquitectura, se rescata la gobernanza de datos federados ya que “establece una estructura que admite decisiones, políticas, estándares y el intercambio de información entre múltiples entidades semiautónomas, como dominios de datos, comités, funciones de la entidad, áreas de tecnologías de información, y proyectos.” (Allen & Cervo, 2015). Además, se tiene en cuenta la importante de identificar qué funcionalidades definidas por las diferentes entidades proveedoras/consumidoras que requieren de la colaboración que proporciona el contexto multidominio a través de las funciones cruzadas de gobierno, seguridad, gestión de metadatos, e integración de datos en el Concentrador GDM, a través de cada uno de sus dominios. La Figura 19 muestra las funcionalidades de las diferentes entidades que requieren gestión de datos en los diferentes dominios. Por

ejemplo, los datos de información personal son proporcionados por la entidad Registraduría en el dominio de personas; estos mismos son consumidos por Confecámaras, IGAC-SNR, y DANE.

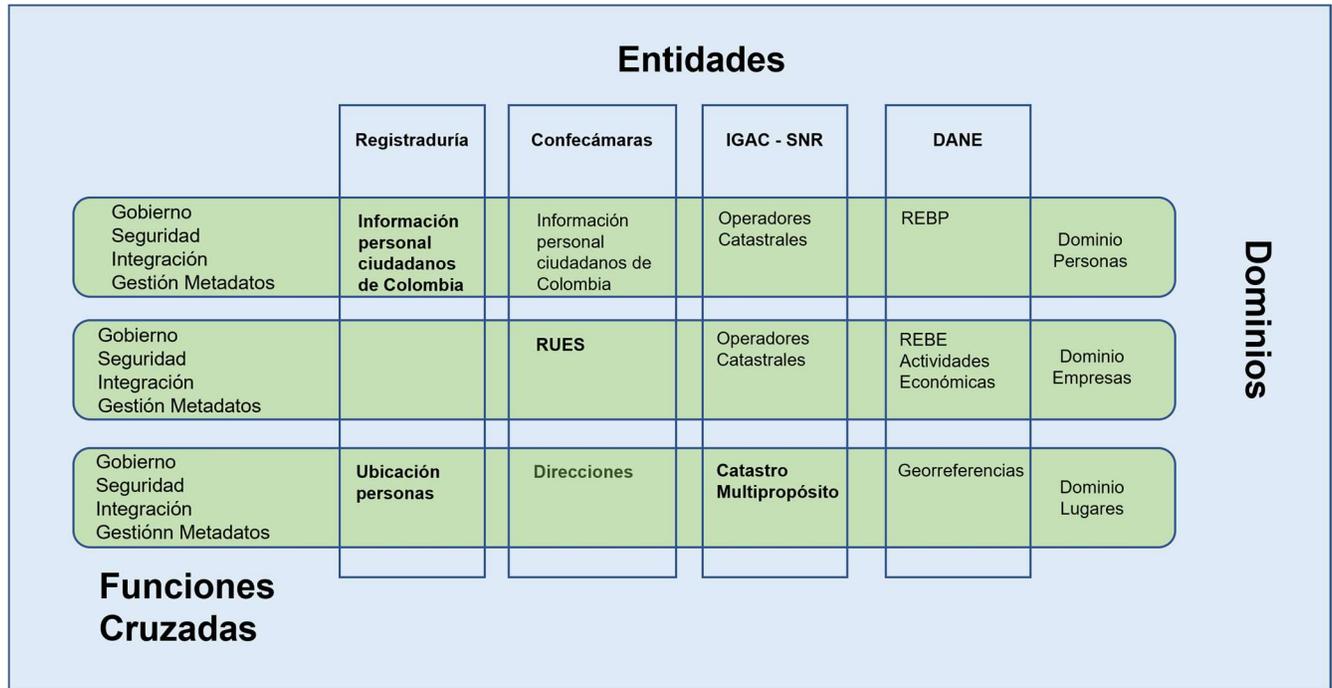


Figura 19. Relación de las Entidades Participantes con los dominios de datos.

Bajo este esquema, el proceso de gobierno de datos maestros del Gobierno Nacional debe operar para un enfoque híbrido, es decir, el gobierno debe ser consensuado entre el administrador de datos del Concentrador GDM y el de las entidades, para tomar decisiones acerca de políticas, reglas, y estándares de gobierno de datos que impacten en todas las funciones que proveen o requieren datos maestros. Además, para hacer gestión de las tecnologías de información, la planificación e implementación de la estrategia en cada dominio. Así, cada entidad define un gobierno de datos más específico con orientación del administrador de datos del Concentrador GDM, y con un enfoque de gestión de calidad para cada dominio con la supervisión del equipo estratégico de gobierno. Los roles que apoyarán la gobernanza de las funciones cruzadas, se describen con mayor detalle en la [sección 5.3. Roles y Responsabilidades del Gobierno de Datos Maestros](#).

6.2. Marco para la Gestión y Gobierno de Datos Maestros

La Figura 20 muestra los componentes del gobierno definido para la gestión de datos maestros del Gobierno Nacional. Los roles involucrados en los tres niveles de gobierno: estratégico, táctico y operativo cuentan con

responsabilidades asignadas para gestionar la gobernanza definida a partir del catálogo de datos maestros, el esquema de aseguramiento de la calidad, la seguridad de datos, y el cumplimiento y privacidad de datos. Estos componentes internos que le dan vida al gobierno de datos maestros interactúan con los dominios de datos maestros a través de procesos de admisión, acceso a herramientas y servicios, políticas y estándares, métricas y reportes, entrenamiento, y comunicación. Estos a su vez se apoyan en el gobierno de datos de los dominios los cuales son administrados por las entidades proveedoras.

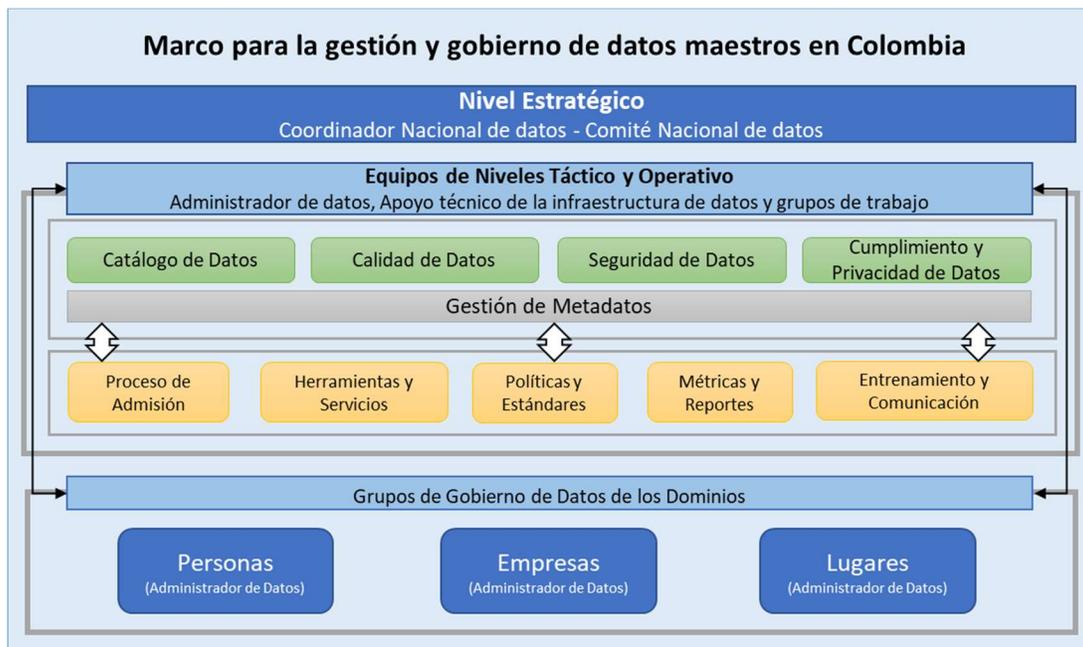


Figura 20. Marco para la gestión y gobierno de datos maestros en Colombia.

6.3. Roles y Responsabilidades del Gobierno de Datos Maestros

Los roles y responsabilidades que deben participar en la ejecución de la hoja de ruta mostrada en el capítulo [3. Estrategia para la Gestión de Datos Maestros](#), se describen usando la matriz de responsabilidades propuesta por (Kendrick, 2013). Estas responsabilidades se describen a continuación:

- **R:** Responsable. Es el responsable de ejecutar el trabajo para lograr el entregable o tarea. Hay al menos un rol con un tipo de participación de responsable, aunque otros pueden ser delegados para ayudar en el trabajo requerido.
- **A:** Aprobador. Es el responsable último de la correcta y minuciosa competencia del entregable o tarea, y el que delega el trabajo a los responsables. En otras palabras, es quien firma (aprueba) el trabajo que proporciona el responsable. Debe haber un solo responsable especificado para cada tarea o entregable.

- **A:** Asesor Técnico. Es aquella persona experta en la materia, a quien se le solicita su opinión, y con quien hay comunicación bidireccional.
- **I:** Informado (Provisto de información). Es la persona a quien se mantiene actualizado sobre el progreso; a menudo solo cuando se completa la tarea o el entregable, y con quien solo existe una comunicación unidireccional.
- **V:** Validador de Política: Corresponde al rol encargado desde el nivel estratégico, de validar lo que se defina en el nivel táctico y Operativo.

La Figura 21 muestra la asignación de las responsabilidades a cada uno de los roles identificados para la Gestión de Datos Maestros en el Concentrador GDM.

ROLES	Coordinador Nacional de Datos	Comité Nacional de Datos	Grupos técnicos de trabajo	Entidad Administradora del concentrador de GDM	Administrador de datos - O responsable de datos (entidades públicas de origen de datos del GDM)	Apoyo técnico de la infraestructura (MINTIC)	Equipos Dataops de la Entidad Administradora del concentrador de GDM	Operador Cloud de la Entidad Administradora del concentrador de GDM	Dataops - (entidades públicas de origen de datos del GDM)
	Nivel Estratégico			Nivel Táctico			Nivel Operativo		
Fase I Estrategia, Arquitectura y Gobierno									
Proceso estrategia para la gestión de datos maestros del Estado - Nivel Estratégico	R	R	A	NA	NA	NA	NA	NA	NA
Proceso de gestión de datos maestros del Estado - Nivel Táctico	V	V	R	E/R	R	E/R	R	I	I
Proceso de dataops para la gestión de datos maestros - Nivel Operativo	I	I	I	I	R	I	R	R	R
Proceso de diseño de la arquitectura para la gestión de datos maestros	I	I	I	R	A/I	A	I	I	I
Fase II Construcción									
Proceso de construcción de componentes y activos de datos	I	I	I	R	R/A	A	R	R	R
Fase III Entrega									
Proceso de pruebas y despliegue	I	I	I	R	R/A	A	R	R	R
Fase IV Operación									
Proceso de operación	I	I	I	R	R/A	A	R	R	R

Figura 21. Matriz de Roles y Responsabilidades para el Concentrador GDM del Gobierno Nacional.

A continuación, se describen los roles y responsabilidades propuestos para la GDM del Gobierno Nacional, los cuales parten de la definición descrita en el ARTÍCULO 2.2.24.2.4. Niveles del Modelo de Gobernanza de la Infraestructura de Datos, y el ARTÍCULO 2.2.24.3.1. Instancias del Modelo de Gobernanza de la Infraestructura de Datos.

1. **Nivel estratégico.** “Es el nivel en el que se definen las políticas, estrategias y prioridades para el desarrollo de la infraestructura de datos. Determina los objetivos a largo plazo y el modo en que las partes interesadas han de interactuar entre sí.” En este nivel se orientará la toma de decisiones y la participación de diferentes entidades del estado y privadas que accederán al concentrador de datos maestros y harán parte de la gestión de los datos maestros y datos de referencia que ellas mismas proveen o consuman. Los roles que apoyan la gestión de los datos maestros en este nivel son:

- **El Coordinador Nacional de datos.** Entidad que actuará dentro del gobierno para la GDM de acuerdo con lo estipulado en el ARTÍCULO 2.2.24.3.2 Coordinador Nacional de datos, y el ARTÍCULO 2.2.24.3.3. Funciones del Coordinador Nacional de Datos.
 - **Comité Nacional de datos.** Grupo estratégico que actuará dentro del gobierno para la GDM de acuerdo con lo estipulado en el ARTÍCULO 2.2.24.3.4. Comité Nacional de Datos, el ARTÍCULO 2.2.24.3.5. Conformación del Comité Nacional de datos, y el Artículo 2.2.24.3.6. Funciones del Comité Nacional de Datos.
2. **Nivel táctico.** “Es el nivel en el que se elaboran los planes, programas, iniciativas, proyectos, procesos y procedimientos para alcanzar los objetivos definidos por el nivel estratégico. Efectúa el control de la gestión realizada por el nivel operacional y soporta las decisiones que se toman y que afectan a las múltiples partes interesadas.” Los roles que apoyan la gestión de los datos maestros en este nivel son:
- **Grupos técnicos de trabajo de datos.** Equipo que actuará dentro del gobierno para la GDM de acuerdo con lo estipulado en el ARTÍCULO 2.2.24.3.10. Grupos técnicos de trabajo de datos. Estos grupos de trabajo deben estar conformados por las entidades que son productoras de los dominios, y algunas entidades usuarias de esos dominios.
 - **Entidad Administradora del Concentrador de GDM.** Entidad pública del nivel nacional, encargada o responsable de la administración del concentrador de gestión de datos maestros. Este rol comprende los siguientes roles:
 - **Administrador de datos.** Dependencia o funcionario del nivel directivo de los sujetos obligados que según el decreto 1389 está encargado de Liderar la planeación, coordinación e implementación de estrategias en la entidad para la debida gestión de los datos durante todo su ciclo de vida y liderar la definición e implementación del modelo de gobierno de datos de la entidad, y coordinar las acciones en el nivel táctico y operativo necesarias para su adecuada implementación, entre otras funciones dadas.
 - **Arquitecto del GDM.** El Gobierno Nacional designará un funcionario o un equipo de funcionarios que cumplan con las capacidades requeridas para tomar el rol de arquitectos del GDM. Ellos deben tener las competencias para diseñar y construir todos y cada uno de los componentes requeridos para el correcto funcionamiento de la plataforma de gestión de los datos maestros (o Concentrador GDM) y la evolución de su arquitectura. En caso de que la implementación de las actividades o productos a ser realizados por este rol sean contratados con un tercero, este equipo será el interlocutor que defina los requisitos, guíe y apruebe lo definido en cada contratación para la mejora continua de la arquitectura del concentrador de datos maestros.

- **Gestor de Catálogo GDM.** El rol de gestor del catálogo de datos maestros es un funcionario o equipo de funcionarios responsables de la precisión y vigencia de datos maestros definidos para cada uno de los dominios y en cada modelo de datos que sean definidos; siendo este un recurso para las entidades del Estado y los ciudadanos en general. El gestor del catálogo sirve como el principal punto de contacto para todas las entidades participantes del Concentrador GDM. El rol también incorpora una estrategia con respecto a la réplica de los datos maestros proporcionados por diferentes entidades del nivel nacional y los procesos que los integran al Catálogo de Datos Maestros.
- **Gestores de métricas y reportes.** Responsable de recopilar, analizar y presentar datos relacionados con el rendimiento y la actividad del Concentrador GDM a través de reportes y/o informes que muestran los resultados del análisis en formatos visuales entre estos gráficos y tablas, lo cual permitirá identificar tendencias, patrones, oportunidades de mejora, comprender y actuar sobre la gestión de los datos.
- **Responsables del acuerdo a nivel servicio o reglas de negocio.** Responsable de definir las condiciones que debe cumplir el servicio, entre estas: documentar los objetivos de nivel de servicio, especificación del servicio, gestión del servicio, y especifica las responsabilidades, para el Concentrador GDM y la entidad pública origen de los datos.
- **Administrador de datos o responsable de datos (entidad pública de origen de los datos del GDM).** Dependencia o funcionario que actuará dentro del gobierno para la GDM de acuerdo con lo estipulado en el ARTÍCULO 2.2.24.3.9. Administrador de datos, desde las entidades públicas de origen de datos del GDM.
- **Apoyo técnico de la infraestructura de datos (MinTIC).** Equipo que actuará dentro del gobierno para la GDM de acuerdo con lo estipulado en el ARTÍCULO 2.2.24.3.11. Apoyo técnico para el desarrollo del Modelo de Gobernanza de la Infraestructura de Datos. Debido a que este rol lo ejerce MinTIC, desde la dirección de Gobierno Digital, se contará con un grupo que acompañe la adopción de lineamientos sobre el gobierno de datos maestros en las entidades públicas.

El gobierno nacional designará un equipo de funcionarios que tengan las capacidades de tomar decisiones para diseñar, construir, y gestionar las políticas, normas, estándares, y lineamientos de gobierno que reflejen los decretos u otros instrumentos definidos por el nivel estratégico, en armonía con la arquitectura GDM. Estas actividades y productos deben de ser implementados por el equipo en las tecnologías que soportarán la gestión de datos maestros en el concentrador del gobierno nacional. En caso de que la implementación de las actividades o productos a ser realizado por este rol sean contratadas con un tercero, este equipo será el interlocutor que defina los requisitos, guíe y apruebe lo definido en cada contratación para la mejora continua de los procesos que facilitan el gobierno del concentrador de datos maestros.

3. **Nivel operativo.** “El nivel operativo está conformado por los funcionarios que estén directamente relacionados con la creación, procesamiento, almacenamiento, intercambio, uso y análisis, archivo y preservación de los datos en todas las áreas y niveles del gobierno. (ARTÍCULO 2.2.24.3.12)” Los roles que apoyan la gestión de los datos maestros en este nivel son:
- **Equipos DataOps de entidad administradora del concentrador GDM.** Equipo conformado por funcionarios que asumen los siguientes roles: el ingeniero de datos, quien es el responsable de la curaduría, limpieza, y disponibilidad de los datos maestros; los arquitectos del Concentrador GDM, quienes son los responsables de conceptualizar la plataforma o marco de trabajo del Concentrador GDM, además trabajan directamente con las entidades dueñas de los datos fuente y consumidores para hacer la gestión de requisitos; el ingeniero dataOps es quien proporciona a los ingenieros de datos orientación y soporte de diseño sobre pipelines de información, revisiones de código, todos los nuevos procesos y flujos de trabajo entorno a la utilización de datos almacenados en el Concentrador GDM. Además, los ingenieros de dataOps supervisan los flujos de datos, la infraestructura de datos, y la gobernanza de toda la operación de ingeniería de datos.
 - **Operador Cloud de la entidad administradora de GDM.** Normalmente, este rol es contratado por una empresa que actuará como tercero y se encargará de proveer el personal y la infraestructura para hacer la ingeniería de datos requerida por la arquitectura sobre la cual esté diseñado el concentrador para la GDM. Este grupo debe ser coordinado por el Ingeniero DataOps.
 - **Dataops – (entidades públicas de origen de datos del GDM).** Cada entidad proveedora deberá garantizar que tiene un equipo de Datops que soporten las necesidades requeridas u ofrecidas para la gestión de sus datamos mestros o de referencia.

6.4. Procesos de Estrategia y Gobierno para la Gestión de Datos Maestros

La Figura 22 describe el flujo de la implementación de la Estrategia para la Gestión de los Datos Maestros (GDM). El nivel Estratégico es el responsable de definir los lineamientos y la estrategia; el nivel táctico el encargado de definir cómo ejecutar y qué acciones llevar a cabo en la estrategia; y el operativo es el encargado de ejecutar las tareas definidas por el nivel táctico.

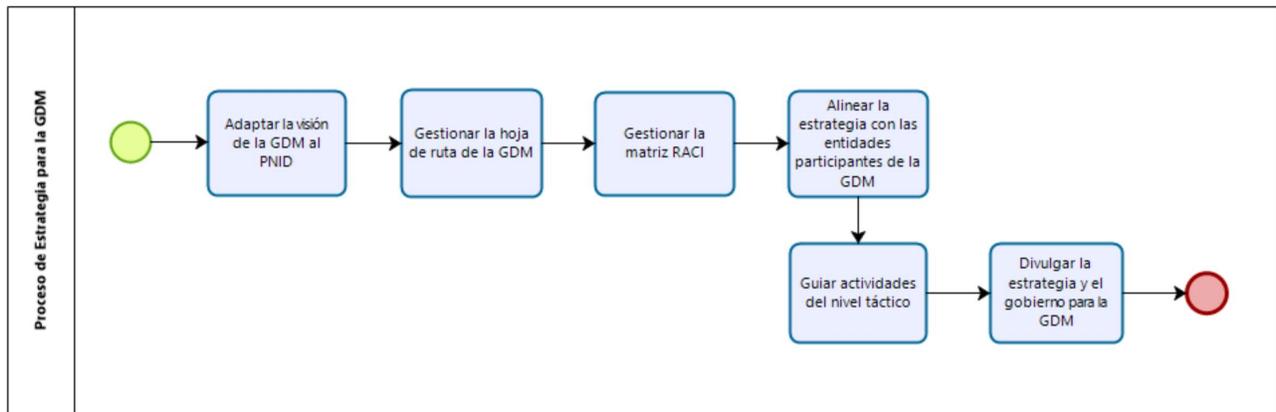


Figura 22. Proceso Estrategia para la Gestión de Datos Maestros del Gobierno Nacional – Nivel Estratégico.

Actividades del proceso:

- **Adaptar la visión de la GDM al PNID.** En esta actividad, el equipo del nivel estratégico analiza y actualiza la estrategia o el gobierno del Concentrador GDM del gobierno nacional, de acuerdo con los cambios requeridos por ley o por necesidades funcionales de los datos maestros. Así podrá mantener vigentes los siguientes componentes estratégicos:
 - Objetivos a mediano y largo plazo que determinan el valor del Concentrador GDM del Gobierno Nacional.
 - Principios que rigen la estrategia y gobierno del Concentrador GDM.
- **Gestionar la hoja de ruta de la GDM.** En esta actividad, el equipo del nivel estratégico define y gestiona el cumplimiento de tareas que habilitan los componentes, recursos y tiempos definidos para la ejecución de la hoja de ruta de la GDM. Esta actividad debe ser coordinada con el nivel estratégico.
- **Gestionar la matriz RACI.** En esta actividad, el equipo del nivel estratégico gestiona los roles y responsabilidad de los diferentes actores que le dan vida a la GDM en los diferentes pasos de la hoja de ruta.
- **Alinear la estrategia con las entidades participantes de la GDM.** En esta actividad, el equipo del nivel estratégico en conjunto con el representante del nivel táctico analiza el impacto de los cambios aprobados en la actividad de Gestionar la matriz RACI, y crean o actualizan directrices que mantengan alineadas las entidades participantes a la GDM del Gobierno Nacional.
- **Guiar actividades del nivel táctico.** En esta actividad, el equipo del nivel estratégico proporciona mecanismos, tecnologías o guías para que el equipo del nivel táctico lleve a cabo las tareas necesarias para lograr la implementación de las decisiones tomadas en la actividad de alinear la estrategia con las entidades participantes de la GDM.
- **Divulgar la estrategia y el gobierno para la GDM.** En esta actividad, el equipo del nivel estratégico divulga las novedades que dan directriz a la GDM del Gobierno Nacional.

La Figura 23 muestra el proceso de Gestión de Datos Maestros ejecutado por el nivel táctico. En el Nivel Táctico: el Administrador de Datos es Responsable, el Apoyo técnico de la Infraestructura de Datos actúa como Asesor Técnico, y los Grupos técnicos de trabajo de Datos son responsables por la ejecución de diferentes subprocesos asociados a la gestión de activos de gobierno y activos de datos. En el nivel Operativo el Equipo DataOps acompaña a los grupos técnicos ya que es el responsable por la ejecución del ciclo de vida de los datos. El nivel estratégico es Informado o actúa como Asesor Técnico.

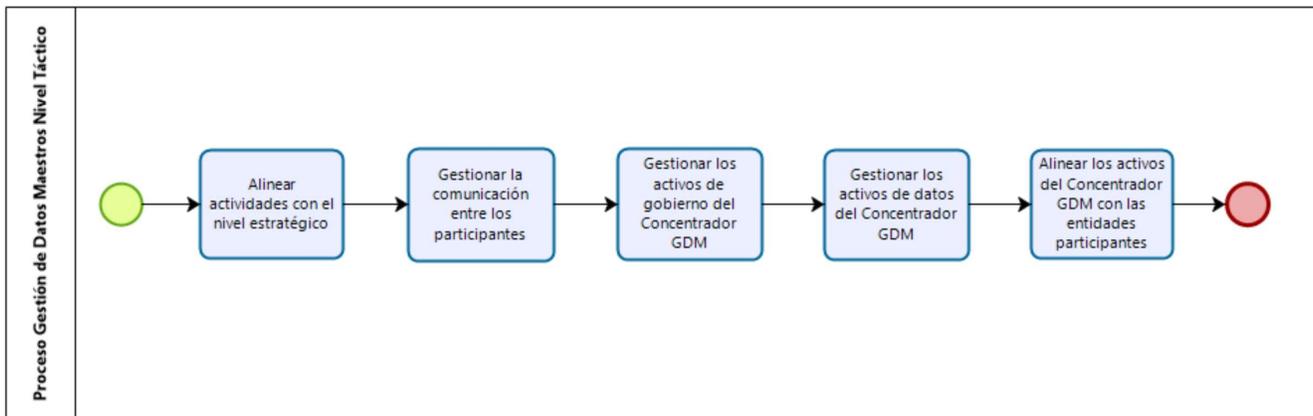


Figura 23. Proceso Gestión de Datos Maestros del Gobierno Nacional – Nivel Táctico.

Actividades del proceso:

- **Alinear actividades con el nivel estratégico.** En esta actividad, el equipo del nivel táctico toma las directrices definidas por el nivel estratégico y las aplica o ajusta en diferentes activos de gobierno o de datos. Para su ejecución crea y mantiene planes, programas y diferentes tareas que estén bajo la responsabilidad de este nivel.
- **Gestionar la comunicación entre los participantes.** En esta actividad, el equipo del nivel táctico inicialmente creará un plan de comunicaciones que le dará vida a la relación efectiva entre los diferentes niveles del gobierno de la GDM y las entidades participantes. Dicho plan deberá ser revisado con el equipo estratégico para mantener una excelente comunicación entre los participantes.
- **Gestionar los activos de gobierno del Concentrador GDM.** En esta actividad, el equipo del nivel táctico define y mantiene los activos de gobierno que deben ser gestionados para que el gobierno de datos maestros tenga vigencia en el tiempo. Los activos de gobierno le dan vida al cumplimiento de decretos o acuerdos u otros instrumentos definidos por el nivel estratégico. Algunos de ellos son: políticas, procesos, normas, estandarización, calidad de los datos, seguridad, etc.

- **Gestionar los activos de datos del Concentrador GDM.** En esta actividad, el equipo del nivel táctico define y mantiene los activos de datos que deben ser gestionados para que el concentrador de datos maestros tenga vida en el tiempo⁴. Algunos de los activos de datos son: catálogo de datos, administración de datos, ciclo de vida de los datos, modelos, arquitectura, flujos de trabajo, sincronización e integración de datos, datos de referencia, operaciones CRUD (Create, Retrieve, Update, Delete), etc.
- **Alinear los activos del Concentrador GDM con las entidades participantes.** En esta actividad, el equipo del nivel táctico da a conocer y transfiere el conocimiento requerido para que las entidades participen en los diferentes procesos y procedimientos asociados a la gestión de los activos de gobierno y de datos definidos para el Concentrador GDM que deberán ser ejecutadas por el nivel operativo.

La Figura 24 muestra el proceso DataOps o Ciclo de vida de los datos y soporte a los datos, los cuales permiten gestionar los datos maestros de la organización una vez se ponga en marcha el Concentrador GDM. Gartner define: “DataOps es una práctica de gestión de datos colaborativa centrada en mejorar la comunicación, la integración y la automatización de los flujos de datos entre los administradores de datos y los consumidores de datos en una organización. El objetivo de DataOps es entregar valor más rápido mediante la creación de una entrega predecible y una gestión de cambios de datos, modelos de datos y artefactos relacionados. DataOps utiliza tecnología para automatizar el diseño, la implementación y la gestión de la entrega de datos con los niveles adecuados de gobernanza, y utiliza metadatos para mejorar la usabilidad y el valor de los datos en un entorno dinámico.”⁵

En este proceso el Nivel Operativo es responsable de la Ejecución de cada una de sus actividades. El rol de Apoyo técnico de la Infraestructura de Datos es el Responsable de los diferentes planes de ejecución de la operación. El nivel táctico actúa como Asesor Técnico o es Informado. El nivel Estratégico es Informado cuando se requiera.

⁴ Esta actividad es una de las responsabilidades del administrador de datos, de acuerdo con el decreto 1389 de 2022 (Función Pública, 2022).

⁵ <https://www.gartner.com/en/information-technology/glossary/data-ops>

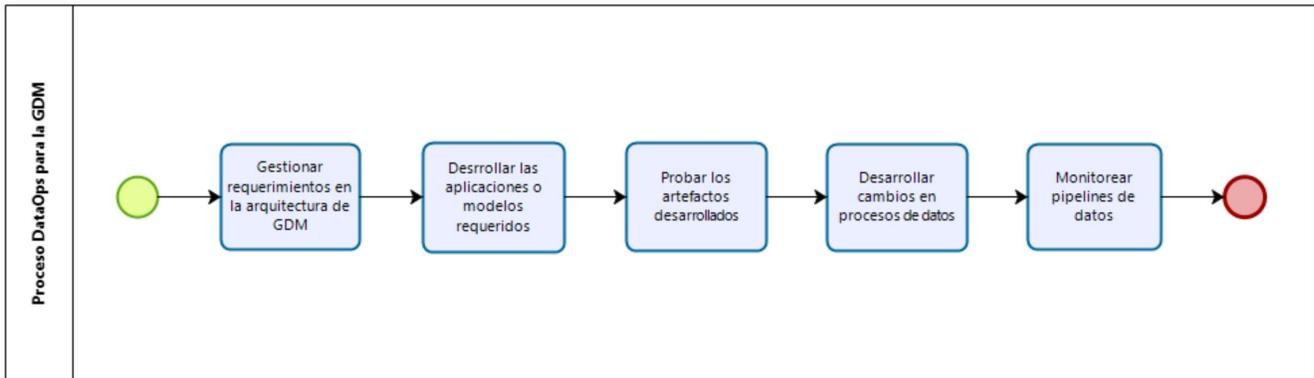


Figura 24. Proceso DataOps para la Gestión de Datos Maestros – Nivel Operativo.

Actividades del proceso:

- **Gestionar requerimientos en la Arquitectura GDM.** En esta actividad, el equipo del nivel operativo idéntica y analiza requisitos asociados a cambios en la arquitectura para la GDM, los modelos, el catálogo de datos maestros y los requerimientos de calidad, y privacidad de datos. Para su ejecución cuenta con acceso a la plataforma de GDM en la cual está implementada la arquitectura.
- **Desarrollar las aplicaciones o modelos requeridos.** En esta actividad, el equipo del nivel operativo usa métodos de la ingeniería de software y los aplica al desarrollo o mantenimiento de pipelines de datos que definen el flujo de datos desde el origen hasta el consumo.
- **Probar los artefactos desarrollados.** En esta actividad, el equipo del nivel operativo incorpora como parte del desarrollo y postproducción pruebas automatizadas de pipelines de datos.
- **Desarrollar cambios en procesos de datos.** En esta actividad, el equipo del nivel operativo incorpora como parte del desarrollo y postproducción la necesidad para orquestación de herramientas. Promueve/habilita con procesos y tecnología Entrega/implementación continua de datos como código desde la prueba hasta la producción. Esta actividad puede ser realizada por el rol Operador Cloud el cual debe ser contratado a un tercero.
- **Monitorear pipelines de datos.** En esta actividad, el equipo del nivel operativo incorpora como parte del desarrollo y postproducción el monitoreo de pipelines de datos y alertas por incumplimiento/falla. Esta actividad puede ser realizada por el rol Operador Cloud el cual debe ser contratado a un tercero.

7. Catálogo de Datos Maestros

7.1. Introducción

(Wells, 2020) define un catálogo de datos como *“una colección de metadatos, combinados con herramientas de gestión y búsqueda de datos, que ayuda a los analistas y otros usuarios de datos a encontrar los datos que necesitan; sirve como un inventario de los datos disponibles y proporciona información para evaluar la idoneidad de los datos para usos previstos.”* Los catálogos son introducidos por las arquitecturas de grandes volúmenes que requieren gestionar y gobernar el ciclo de vida de los datos con mayor detalle y así facilitar y controlar el uso de los datos con mayor precisión tanto para la operación de las entidades como para la analítica.

El elemento base de un catálogo son los conjuntos de datos almacenados ya sea en un lago de datos, una bodega de datos, un repositorio de datos maestros, o cualquier otro recurso de datos compartido. Los conjuntos de datos, además de tener sus propios metadatos, son enriquecidos con otros metadatos de personas, búsqueda, procesamiento, y proveedores.

- **Metadatos de personas.** Describen quiénes trabajan con los datos, ya sean administradores, gestores de metadatos, etc., que estén gestionados del concentrador de datos.
- **Metadatos de la búsqueda.** Etiquetas o palabras clave que facilitan el hallazgo de los datos que son requeridos por los usuarios.
- **Metadatos de procesamiento.** Describe las transformaciones y derivaciones que se aplican a medida que los datos se gestionan a lo largo de su ciclo de vida
- **Metadatos de proveedores.** Datos adquiridos de las fuentes externas, informando sobre fuentes y restricciones de suscripción o licencia.

7.2. Proceso de Gestión del Catálogo de Datos Maestros y de Referencia

En la arquitectura de referencia propuesta se incluye el catálogo de datos maestros como un elemento activo de la arquitectura, el cual tiene la función de proporcionar un ambiente para gestionar los metadatos de los activos de datos que son parte del Concentrador GDM. La Figura 25 muestra el flujo de trabajo definido para gestionar el catálogo de datos maestros y de referencia.

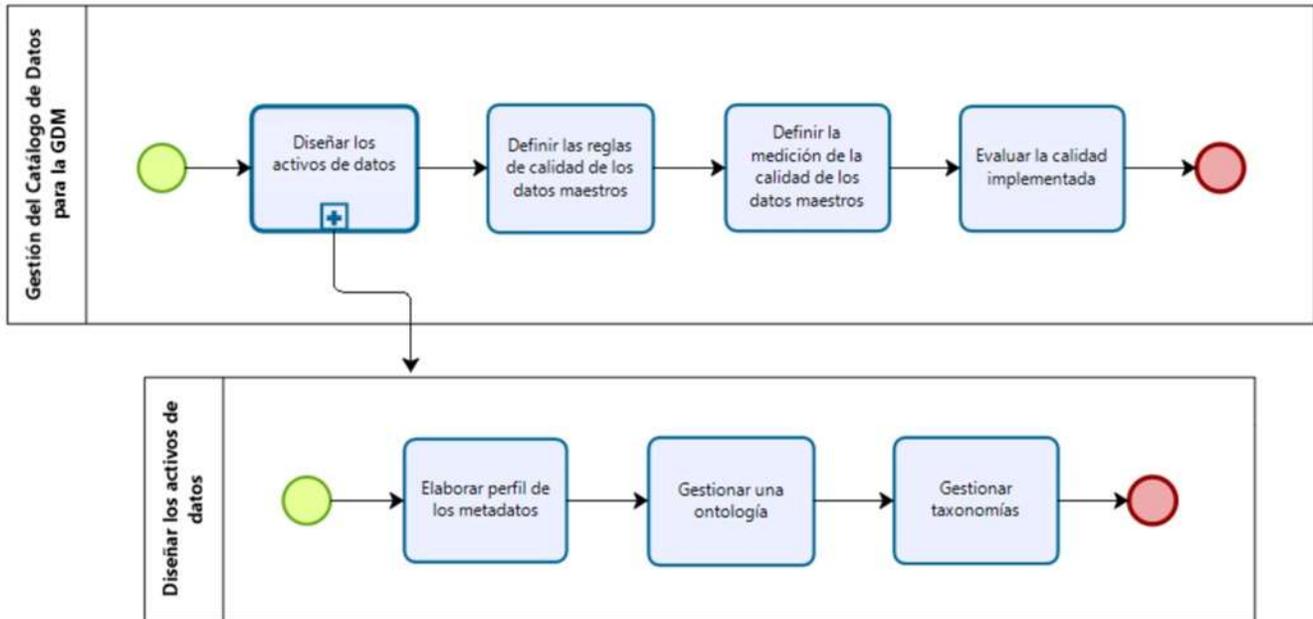


Figura 25. Proceso para la Gestión del Catálogo de Datos Maestros y de Referencia.

Actividades del proceso y subproceso de la gestión del Catálogo para la GDM:

- **Diseñar los activos de Datos.** Este es un subproceso que solo se activa cuando se crea o actualizan los metadatos de los activos de datos que hacen parte del Concentrador GDM. Los activos de datos autorizados por el nivel estratégico del marco de referencia del gobierno, que se registran y actualizan en el Catálogo, corresponden a las tablas y atributos maestros o de referencia que harán posible la gestión de datos en el Concentrador GDM. De igual forma, en el momento de la creación o actualización se debe definir la gobernanza asociada a cada activo de datos. Esta se refiere al control de acceso para roles específicos que se pueden mapear en el catálogo de datos maestros definido, que a su vez están dentro del marco de gobierno propuesto. De esta forma, el catálogo de datos maestros se puede utilizar para administrar el gobierno desde los activos de datos y sus los metadatos relacionados a la sensibilidad de un dato que necesita protección (ver metadatos a nivel columna). Para escenarios en los cuales los metadatos se caracterizan como confidenciales, estos deben contar con permisos especiales a fin de conocer quién puede acceder a ellos o quien puede consultarlos. Este subproceso está definido por un flujo de trabajo conformado por tres actividades, como se muestra en la Figura 24, las cuales son:
 - **Elaborar perfil de los metadatos.** El perfil de datos que se almacena en el catálogo de datos maestros define información adicional sobre tablas, atributos claves, valores válidos y rangos, etc., así como la asociación entre atributos en tablas o documentos, prioridad entre los mismos campos entre múltiples tablas o documentos y restricciones específicas de dominio, etc. En esta actividad también

se recopilan metadatos estructurales de las fuentes de datos y se analizan tablas de bases de datos o activos de datos en el catálogo de datos maestros. Adicionalmente, como parte de las actividades a adelantar al elaborar los perfiles respectivos se deben categorizar los activos de datos en dominios de datos u otros grupos. La plataforma o herramientas que automatice la gestión del catálogo del Concentrador GDM debe proporcionar la forma de rastrear bases de datos asociadas al Concentrador GDM y recuperar metadatos automáticamente, así como hacer recuentos de filas, valores superiores en una columna, recuentos nulos, recuentos de valores distintos y mucho más.

- **Gestionar una ontología.** A partir de la definición de los perfiles de los metadatos, es importante diseñar y gestionar una ontología que permita establecer la relación entre datos estructurados internos y datos de fuentes externas. Aunque el diseño lógico establece una relación por integridad referencial (sistema de reglas utilizado para asegurar que los registros de tablas relacionadas son válidos y no se borran o cambian datos relacionados de forma accidental para evitar producir errores de integridad), algunas veces las relaciones funcionales a nivel de atributos no se logran ver fácilmente. Estas relaciones se pueden documentar en el catálogo de datos maestros como enlaces con nombre entre diferentes tipos de entidades.
- **Gestionar Taxonomías.** En la definición de los datos maestros o de referencia se pueden dar estructuras con jerarquías de datos. Los registros de origen muchas veces provienen de sistemas de información o sistemas funcionales de las entidades o de bodegas de datos donde son naturales las jerarquías de datos definidos por los diseñadores de datos que los crearon. En una arquitectura híbrida para la gestión de datos maestros, este tipo de diseño no se da con frecuencia al almacenar el registro de oro de las entidades maestras en estructuras relacionales en tercera forma normal. Sin embargo, a medida que se van involucrando tipos de datos no estructurados que pueden provenir de diseños multidimensionales, es importante identificar este tipo de diseños y analizar el impacto en el registro de oro de cada dominio. En caso de que el diseño lo requiera, dichas jerarquías se pueden almacenar en el catálogo de datos maestros y visualizarlas a través de una interfaz de usuario con la funcionalidad de desglose en una solución tecnológica para la gestión de datos maestros. El diseñador debe validar entre la necesidad de uso de este tipo de estructuras con respecto al desempeño requerido al momento de consultar los datos maestros.
- **Crear los activos de datos para la GDM.** En el proceso de extracción o consulta de los datos en cada dominio, los productos del catálogo de datos maestros se deberán identificar o escanear (manual o automáticamente) los metadatos en las fuentes estructuradas para el descubrimiento y la catalogación. En caso de que los datos presenten algún inconveniente en esta actividad de descubrimiento, estos deben curarse manualmente bajo las reglas definidas para dicha curación. En esta actividad se recopilan metadatos estructurales de las fuentes de datos ya sea cargando un diccionario de datos existente o creando uno de cada fuente de datos. Algunas herramientas de catálogo de datos proporcionadas por proveedores de plataformas para la gestión de datos maestros (p.ej., IBM, Oracle, Informatica, etc.) pueden rastrear bases de datos y recuperar metadatos automáticamente. Además, durante esta actividad, el catálogo proporcionará elementos para normalizar

datos de varias fuentes con el objetivo de garantizar una única fuente de referencia para las entidades consumidoras de datos en el Concentrador GDM.

- **Reutilizar casos de uso.** Para lograr una buena gestión de los datos maestros es importante tener claro y documentado el uso de las tablas o atributos y los datos relacionados que están almacenados en el Concentrador GDM. Una buena práctica es describirlo a través de casos de uso. Esto dará pie a tener casos de uso para reutilizarlos en otros, lo que ahorra tiempo para volver a trabajar en el etiquetado y la extracción. Por eso es importante durante la definición, marcar relaciones entre datos, haciendo conexiones entre tablas de bases de datos, activos de datos y dominios, etc. relacionados con una misma entidad maestra o de referencia. Esto se puede hacer manualmente o mediante los algoritmos avanzados de una herramienta proporcionada por la plataforma donde se alojarán los datos maestros y de referencia (p.ej., IBM, Oracle, Informatica, etc.).
- **Visualizar Linaje.** Esta actividad parte de crear un linaje de datos identificando los activos de datos que se integran a partir de dos o más conjuntos de datos y analizando los procesos ETL⁶ que se realizan en cada activo de datos. El linaje de los datos es esencial para el administrador. Este le permite abordar de forma ágil problemas en la transformación de los datos y visualizar la historia de los datos desde su carga hasta su utilización final.

7.3. Instrumentos de trabajo para gestionar el Catálogo de Datos Maestros y de Referencia

Las actividades del proceso para gestionar el catálogo de datos maestros están respaldadas por un conjunto de instrumentos que le dan vida a la gestión de los metadatos. Estos instrumentos para la gestión del catálogo de datos maestros que a continuación se presentan, servirán de guía a los roles tácticos y operativos para que puedan proporcionar una buena gestión de los datos maestros. El archivo anexo denominado “**Catalogo de Datos del HUB-GDM.xls**” es una versión del catálogo de datos maestros definido para los activos de datos planteados en el alcance de este proyecto, y además con una versión ampliada y diligenciada de los diferentes instrumentos de trabajo.

7.3.1. Glosario de las Entidades

“Un glosario de las entidades rige los conceptos organizaciones junto con las definiciones asociadas y las relaciones entre esos términos. Los glosarios deben brindar definiciones claras, accesibles en toda la entidad para las partes interesadas técnicas y no técnicas por igual, para mantener la coherencia de los términos y ayudar a todas las partes interesadas a permanecer en la misma página.” (CMS, 2022). En la Tabla 18 se define el instrumento sugerido para recolectar el glosario de la entidad. Este ayudará a descubrir los términos usados en

⁶ "Extract, transform and load (extracción, transformación y limpieza), en administración de sistemas de información y almacenes de datos." Fuente: ETL - <https://es.wikipedia.org>

el día a día de la función integral de la entidad; luego se convertirán en activos de datos y metadatos. La fila del encabezado muestra las etiquetas que agrupan la información a ser identificada y recolectada en cada columna de la tabla. La fila 2 define la guía (preguntas o sentencias) que regirá el diligenciamiento de la información requerida en cada columna. Y en la fila 3 se presenta un ejemplo de un elemento de la entidad asociado al dominio de Personas. Esta tabla reflejará la base conceptual de las entidades y atributos que se definen desde el nivel de la función misional y que luego convertirán en activos de datos.

Tabla 24. Instrumento de Catálogo de Datos Maestros 1 - Definición de términos del negocio base para la GDM.

Término de la entidad	Definición del Término	Categoría	Términos relacionados	Fuente del término	Dominio del dato	Enlace a los datos	Término del Administrador
¿Cuál es el nombre del dato o concepto que se describe?	Describa los datos o conceptos subyacentes al término de la entidad	¿A qué parte de la entidad pertenece esta información?	¿Qué palabras o expresiones están asociados con este término?	¿Dónde se encuentran los datos asociados con este término? ¿De dónde vino?	¿En qué dominio se incluyen estos datos?	Vínculo con los datos ¿Qué datos específicos (p. ej., conjuntos de datos o bases de datos) tenemos sobre este término?	¿Quién es responsable de este término o rige los datos o conceptos subyacentes?
Datos personales del ciudadano	Identificación única de cada persona ciudadano/ciudadana de Colombia	Registro Civil de Nacimiento y el Archivo Nacional de Identificación	Datos de identificación tarjeta de identidad, cedula de ciudadanía	Bases de datos de la RNEC	Personas	ANI, RCN	Administrador de las BD ANI, RCN

7.3.2. Definición de Metadatos - Diccionarios de Datos

En un catálogo de datos maestros, así como en otros sistemas de datos, se “buscan establecer las pautas para crear definiciones claras, concisas y sin ambigüedades de los componentes de datos.” (CMS-DD, 2022). “Un diccionario de datos es una colección de descripciones de objetos o elementos de datos en un activo de datos. Los diccionarios de datos proporcionan información estructural y semántica sobre los datos. Es una colección de términos, definiciones y atributos sobre elementos de datos que se utilizan o capturan en un sistema de información o base de datos específicos.” (CMS, 2022).

7.3.3. Perfil de los activos de Datos

Los activos de datos que se van a definir y almacenar en el Concentrador GDM son diversos. Los de mayor relevancia son las entidades maestras y las entidades de referencia. Adicional a estas entidades, también son activos de datos elementos que componen a las entidades como procedimientos almacenados, funciones, canalizaciones (pipelines), y flujos de trabajo que definen la gestión de los datos, etc. Cada uno de estos se debe registrar en el catálogo de datos maestros. En la Tabla 19 se define el instrumento sugerido para definir el perfil

de los activos de datos. Este ayudará a identificar y describir los activos de datos de los cuales se gestionarán sus metadatos. La fila del encabezado muestra las etiquetas que agrupan la información que se debe identificar y recolectar en cada columna de la tabla. En la fila 2 se describe la guía para diligenciar la información requerida en cada columna, y en la fila 3 se presenta un ejemplo de un activo de datos, en este caso del dominio de Personas.

Tabla 25. Perfil de los activos de datos.

Nombre del activo de datos	Descripción del activo de datos	Tipo de activo de Datos	Versión del activo de datos	Fecha Creación	Fecha actualización
Nombre dado al recurso	Narrativa breve (p. ej., 2 a 4 oraciones) que describe las características esenciales del conjunto de datos para ayudar al consumidor a comprender su contenido y relevancia.	Un activo de datos puede ser un archivo de salida de un sistema o una aplicación, una base de datos, un documento o una página web. Un activo de datos también incluye un servicio que se puede proporcionar para acceder a los datos desde una aplicación.	Versión del activo de dato. Debe ajustarse al sistema de versionado definido para la GDM	Fecha en que se crea el activo de datos	Fecha de la última actualización del activo de datos
HUB_Maestra_Personas	Base de Datos que almacena el modelo de la entidad maestra de Personas cuyo origen de los datos está en las bases de datos federadas de RNEC: ANI y RCN	Sistema o plataforma de bases de datos usada para el almacenamiento, procesamiento, y consulta de estructuras de datos maestros.	1.0	15/06/2023	02/88/2023

7.3.4. Metadatos a nivel de Entidad/Tabla

Los metadatos a nivel de Entidad/Tabla son etiquetas y descripciones que definen este tipo de activo como un objeto o componente único en el concentrador GMD. En la Tabla 20 se precisa el instrumento sugerido para definir y describir los metadatos de este nivel. La fila del encabezado muestra las etiquetas que agrupan la información que se debe identificar y recolectar de cada tabla maestra o de referencia definida en el concentrador. En la fila 2 se describe la guía para diligenciar la información requerida en cada columna, y en la fila 3 se presenta un ejemplo de un metadato de Entidad/Tabla, en el dominio de Personas.

Tabla 26. Metadatos a nivel de Entidad/Tabla

Nombre de la Entidad	Definición de la Entidad	Nombre de la Tabla	Descripción de la Tabla	Identificación del requisito de la entidad	Categoría de Seguridad de la Entidad	Sistema de Almacenamiento
----------------------	--------------------------	--------------------	-------------------------	--	--------------------------------------	---------------------------

Etiqueta o título que se le da la entidad lógica que representa una entidad del negocio	Descripción corta del objetivo de manejo de datos a partir de esta entidad	Nombre técnico basado en los estándares establecidos por los grupos del nivel táctico en el gobierno de datos	Descripción corta del objetivo de manejo de los datos que serán almacenados en la tabla	Número o código del requisito que define la necesidad de almacenamiento	Categoría de seguridad definida por la entidad que se garantizan en el Concentrador GDM bajo la implementación de la norma ISO 27001.	Activo de dato: repositorio o instancia de base de datos donde se almacenará la tabla
Personas	Entidad para almacenar la información personal de los ciudadanos de Colombia	HUB_DPER_Maestra_Personas_ROR	Tabla para almacenar la información personal de los ciudadanos de Colombia	IDDOC-00001	CONFIDENCIALIDAD=ALTA; INTEGRIDAD=ALTA; DISPONIBILIDAD=MODERADA	HUB_Maestra_Personas

7.3.5. Metadatos a nivel de Atributo/Columna

Los metadatos a nivel de Atributo/Columna son etiquetas y descripciones que definen las restricciones de calidad de cada columna que son parte de una Entidad/Tabla. Este tipo de activo se define como un objeto único en el concentrador GMD. En la Tabla 21 se define el instrumento sugerido para definir y describir los metadatos de este nivel. La fila del encabezado muestra las etiquetas que agrupan la información que se debe identificar y recolectar en cada columna de cada tabla maestra o de referencia. En la fila 2 se describe la guía para diligenciar la información requerida en cada columna, y en la fila 3 se presenta un ejemplo de un metadato de Atributo/Columna en el del dominio de Personas.

Tabla 27. Metadatos a nivel atributo/Columna.

Nombre Tabla	Nombre Columna	Descripción Columna	Tipo de Dato de la Columna	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad	Fuente Técnica
Nombre de la Tabla Física	Nombre del campo físico	Descripción del campo	Tipo de Dato	Es nulo u opcional	El campo es Clave Primaria (Si/No)	El campo es Clave Foránea (Si/No)	Clasificación de la Sensibilidad del campo (definición en el documento)	Fuente Técnica del Elemento del Dato
HUB_DPER_Maestra_Personas_ROR	MPER_Identificación	Identificación única de cada persona natural que está registrado civilmente en el estado colombiano	BIGINT	No	Si	No	Alta	No conocido

7.3.6. Metadatos a nivel de Dominio de Valor

Los metadatos a nivel de Dominio de Valor son etiquetas y descripciones que permiten identificar atributos/columnas que restringen el valor a almacenar desde un conjunto de valores restringidos. Estos valores son definidos por la entidad en cada dominio y normalmente se restringen en el sector gobierno por una norma o decreto al interior de la entidad que administra el registro de origen. En la Tabla 22 se define el instrumento sugerido para definir y describir los metadatos de este nivel. La fila del encabezado muestra las etiquetas que agrupan la información que se debe identificar y recolectar en cada columna que tenga como tipo de dato un Enumerador (Enumeration type) en cada tabla maestra o de referencia. En la fila 2 se describe la guía para diligencia la información requerida en cada columna, y en las filas de la 3 a la 6 se presenta un ejemplo de un metadato de Dominio de valor para la columna “HUB_DPER_Tipodocumento_ROR_Id_tipodocumento” en el del dominio de Personas.

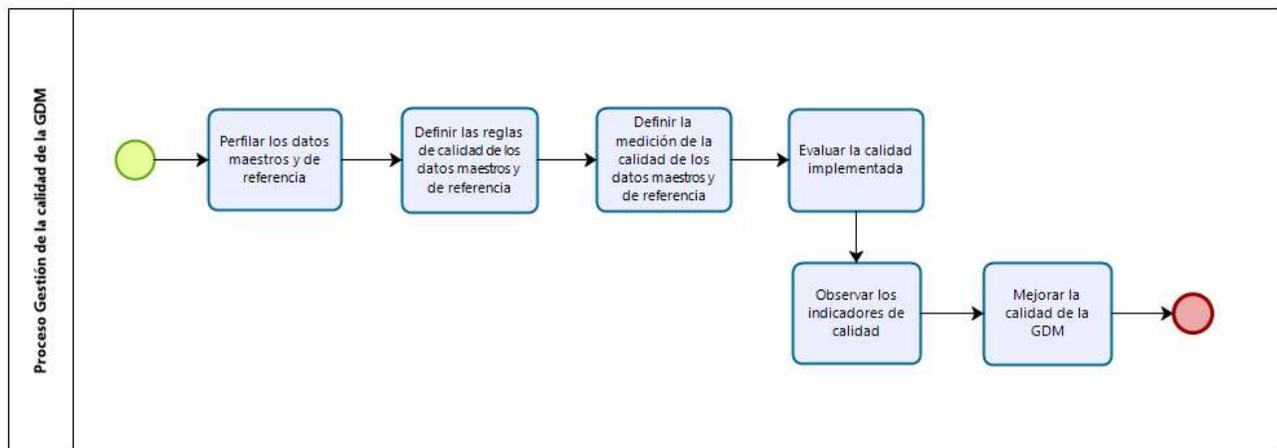
Tabla 28. Metadatos al nivel de Dominio de Valor

Nombre de la tabla física	Descripción del campo físico	Enumerador dentro del campo	Código/Número del valor de un campo	Descripción del Valor
Etiqueta o nombre dato a la tabla en la cual se define el atributo/columna	Nombre dado al campo físico definido	número secuencia de declaración del valor	Valor definido por política o norma de la entidad que luego se convierte en restricción de valor	Descripción del código/valor
HUB_DPER_Maestra_Personas_ROR	HUB_DPER_Tipodocumento_ROR_Id_tipodocumento	1	Nuip	Número Único de Identificación Personal https://www.registraduria.gov.co/Glosario-de-identificacion.html
HUB_DPER_Maestra_Personas_ROR	HUB_DPER_Tipodocumento_ROR_Id_tipodocumento	2	Nip	Número de Identificación personal https://www.registraduria.gov.co/Glosario-de-identificacion.html
HUB_DPER_Maestra_Personas_ROR	HUB_DPER_Tipodocumento_ROR_Id_tipodocumento	3	TI	Tarjeta de Identidad https://www.registraduria.gov.co/Glosario-de-identificacion.html
HUB_DPER_Maestra_Personas_ROR	HUB_DPER_Tipodocumento_ROR_Id_tipodocumento	4	CC	Cedula Ciudadania https://www.registraduria.gov.co/Glosario-de-identificacion.html

8. Aseguramiento de la Calidad de la Gestión de Datos Maestros

En el **Principio 7. Calidad de los Datos** se describe la base sobre la cual la gestión de la calidad de los datos maestros del Gobierno Nacional debe definir las políticas o reglas de calidad que garanticen diferentes dimensiones de la calidad de los datos como: completitud, validez, unicidad o singularidad, consistencia, puntualidad, y precisión de los datos maestros del Gobierno Nacional. Esto incluye elementos tales como catálogo de datos maestros el cual maneja datos de referencia de la información de los datos maestros, metadatos, calidad de los datos que incluye elementos como con perfil de los datos, reglas de reconocimiento, coincidencia y corrección del nombre de la entidad, emparejamiento entre datos de diferentes fuentes, protocolos, estándares a nivel de códigos o nombres de lugares y direcciones, etc.

La Figura 26 muestra el proceso de Gestión de la Calidad de Datos Maestros. Este proceso es un enfoque iterativo de la calidad de los datos. Inicia con la actividad de creación de perfiles de datos para luego definir reglas, diseñar, implementar, y monitorear la gestión de la calidad. El proceso es circular y mide y mejora los datos maestros de forma continua.



Powered by
bizagi
Modeler

Figura 26. Proceso de Gestión de la Calidad de Datos Maestros.

8.1. Perfilamiento de Datos Maestros

La elaboración de perfiles de datos es una técnica para explorar datos de fuentes de información existentes y recopilar información sobre esos datos. Los perfiles permiten descubrir la calidad de los datos mediante la

determinación de la precisión, integridad y consistencia de los datos. Como parte del proceso de resolución de entidades, la elaboración de perfiles de datos es esencial para definir qué elementos de datos usar para identificar de manera única a una entidad, qué operaciones se requieren para transformar los datos en un estándar de calidad mínimo requerido, cómo agrupar registros similares y, finalmente, como armar un registro de oro. Para definir el perfil de los datos que serán almacenados en el concentrador de datos maestros se definen dos tipos de registros:

- **Registro de origen (*Record of Origin – ROO*)**. Se refiere al lugar físico donde los datos se crean con la mayor precisión posible. Una ROO también puede denominarse fuente autorizada. Para el alcance de este proyecto se definen un registro de origen por cada dominio ubicados en cada una de las entidades participantes: Registraduría, Confecámaras e IGAC. Se usará una identificación específica denominada ROOB – ROO Base, cuando el ROO se comporte solo como entidad de referencia u origen principal de los datos de origen cuando hubiese otras fuentes de origen para la misma entidad maestra,
- **Registro de referencia (*Record of Reference – ROR*)**. Se refiere al concentrador o centro de datos que sirve como una única fuente de verdad para la gestión de datos maestros. La creación y el uso de un ROR confiable, el que extrae o conecta sus datos con los ROO definidos, es clave para establecer la consistencia de los datos maestros que requiere el GDM del Gobierno Nacional.

La creación de perfiles de datos son la base para medir la calidad general de los datos maestros y consecuentemente provee información para el monitoreo y la generación de reportes continuos del comportamiento de los datos en momentos de tareas de integración y sincronización. En el programa estratégico para la GDM Multi-Dominio, se requiere la creación de perfiles de datos en varias etapas del proceso (Allen & Cervo, 2015). Los perfiles se definen a partir de dimensiones de calidad establecidas para la gestión de datos. Las entidades fuentes proveedoras de los registros de oro, deben garantizar que los perfiles definidos por cada dimensión se cumplan desde el registro de origen. De esta forma, las reglas que se definan en el Concentrador GDM facilitarán la clasificación de posibles fallas al tener certeza de que la fuente comparte sus datos con las mínimas reglas de calidad.

8.2. Dimensión y reglas de Calidad del dato

Las reglas son los elementos de validación que permiten establecer condiciones o restricciones sobre los datos transaccionales, maestros o de referencia. La Figura 27 muestra a través de un ejemplo la jerarquía de elementos estratégicos que son parte de la definición de cada regla que asegura la calidad de los datos maestros y de referencia del Concentrador GDM. Una regla de calidad existe porque inicialmente se declara desde la estrategia. El Principio 7. Calidad de Datos tiene asociadas las políticas de la 6 a la 15, las cuales habilitan la definición de las reglas de calidad y de vigencia de los datos.

El estándar de numeración de cada regla es **R#Principio.#Política.#Consecutivo**.

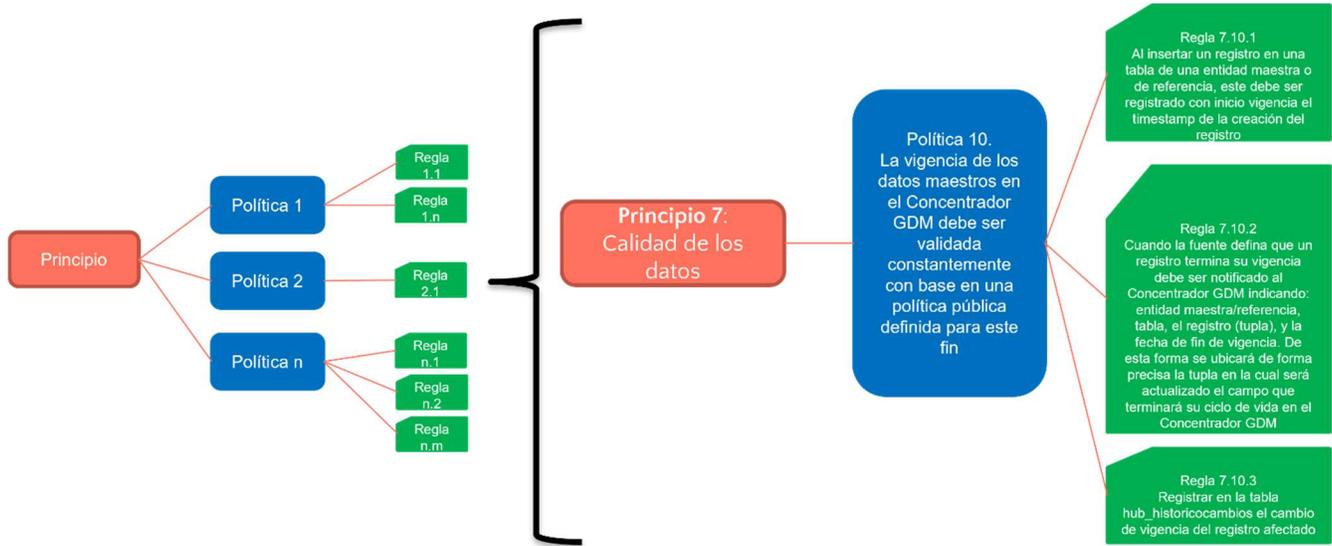


Figura 27. Jerarquía para el aseguramiento de la calidad de los datos maestros y de referencia.

8.2.1. Dimensión de Exactitud

Es el grado en el que los datos corresponden a valores correctos conocidos en el mundo real, en este caso, las entidades maestras de los ROO, por lo tanto, son proporcionados por una "fuente de verdad" reconocida o establecida. El Perfil que garantiza la medición de la exactitud se logra de la siguiente forma:

- Comparando los valores del registro de oro con los valores estándar o valores de origen proporcionados por las entidades

Reglas asociadas a la dimensión exactitud.

- **R7.13.1. Mapeo metadatos.**
Mapear uno a uno de los metadatos del registro de origen con cada uno de los atributos del registro de oro de referencia del Concentrador GDM.

ROO::Atributo (nombre, tipo, restricciones de calidad) = **ROR::Atributo** (nombre, tipo, restricciones de calidad)

Si no hay coincidencia total se reporta error por inconsistencia tipo estructural

- **R7.13.2, Copia Registro de Oro.** correspondiente. Cada uno de los registros de oro proporcionados por la fuente se debe duplicar de forma exacta en la tabla **hubdper_maestrapersona**.

ROO::Atributo::campo → ROR::Atributo::campo.

- ✓ Esta regla garantiza que la autenticación digital se conserva en el Concentrador GDM.

Si no hay coincidencia total en los metadatos que provienen de la fuente se genera un error y se reporta en el informe de calidad “inconsistencia tipo estructural en el Reporte de Calidad y Auditoría”.

8.2.2. Dimensión de Completitud

Es el grado en que los datos se completan en función de la definición del registro de oro definido en la ROO y finalmente se duplicará en el ROR. Las reglas de negocio definidas por cada entidad participante en la generación de un registro de oro definen la completitud del registro de oro del ROR. Los perfiles que garantiza la completitud pueden identificar: falta el registro en sí, falta un valor en un atributo (atributo nulo), falta un valor de referencia, truncamiento de datos.

Reglas asociadas a la dimensión de completitud.

- **R7.13.3. Campos nulos.** Si un ROO de la entidad maestra llega con campos nulos al Concentrador GDM, la estructura de la entidad maestra del ROR no admite el registro de oro fuente.
- **R7.13.4, totalregistros.** La entidad fuente debe entregar totales de control de las entidades maestras disponibles desde el ROO. Una vez los datos fuentes se almacenen en la base de datos de extracción se hace la correspondiente validación.

(validar si X-ROAD se puede encargar de esta operación. En caso de que si, el Concentrador GDM, solo recibiría datos completos desde el número de registros).

- **R7.13.5, valorreferencia.** Los atributos de referencia que complementen el registro de oro de la entidad fuente debe conservarse en la entidad maestra definida en el Concentrador GDM.

ROO::<Atributodereferencias> = ROR <Atributodereferencia>

- **R7.13.6, valorcompleto.** Los valores del registro de oro del ROO que sean vulnerables a registrarse de forma incompleta, debe ser completados bajo reglas de negocio definidas por la entidad del ROO antes de ser certificados para el Concentrador GDM.

Este tipo de certificación es válida si y solo si en el acuerdo entre el ROO y el ROR se especifica cuales atributos de las entidades maestras del ROO tienen valores de este tipo y las reglas que llevan que se aplican para proveer valores completos, no truncados, y certificados para ser llevados al ROR.

8.2.3. Dimensión de Consistencia

Los perfiles que garantiza consistencia de los datos son: registros diferentes entre el origen y el destino, atributos diferentes entre el origen y el destino, datos funcionalmente idénticos manejados de forma diferente por dos entidades, datos a lo largo del tiempo diferentes, representación de datos entre los sistemas de origen y destino diferentes.

Reglas asociadas a la dimensión de consistencia.

- **R7.13.7. registrofuentedestino.** El ROO garantiza que las entidades maestras almacenan un registro maestro único con detalles precisos sobre cada persona, empresa o lugar relacionado con la misión de cada la entidad fuente. Así, cada ROR será un reflejo de la misma consistencia en el Concentrador GDM.
- **R7.13.8. datos a través del tiempo.** El ROO debe garantizar la historia de cambios de los datos maestros incluyendo la afectación en otras entidades maestras por cambios que se realizan en el tiempo por errores en los datos o por nuevas reglas de negocio.

La fuente debe declarar en el acuerdo realizado con el Concentrador GDM que acepta realizar el método de cambios en datos maestros o de referencia o en las estructuras. Además, la forma cómo se garantiza la historia para el uso correcto de los datos a través del tiempo. Este tipo de consistencia está asociada a la vigencia de los datos.

- **R7.13.9. representación datos maestros.** Los administradores de las entidades consumidoras deben acordar con el administrador del Concentrador GDM la forma de consumir los datos maestros. Se pueden presentar dos casos:
 - Las entidades consumidoras consultan las entidades maestras del Concentrador a través de reglas de presentación definidas desde la necesidad de la entidad consumidora.
 - El Concentrador GDM define un protocolo de actualización de las tablas maestras de los gestores de datos maestros de las entidades consumidoras.
- **R7.13.10. datos creados en el Concentrador GDM.** El Administrador de Datos del Concentrador GDM puede proponer la creación de datos maestros que enriquezcan el consumo de los datos maestros proporcionados por las entidades fuente. Se deben crear nuevas estructuras y procedimientos almacenados o funciones que garanticen la consistencia y otros atributos de calidad base del

Concentrador GDM. La propuesta debe ser aprobada por el nivel estratégico del Concentrador GDM. Una vez se apruebe el equipo táctico debe definir el ciclo de vida de esos nuevos datos.

8.2.4. Dimensión de Unicidad

El grado en que se permite que los datos tengan valores duplicados. Un evento o entidad solo debe registrarse una sola vez. Los perfiles que garantiza la unicidad al identificar: la misma entidad se representa con diferentes identidades, la misma entidad se representa varias veces con la misma identidad

Reglas asociadas a la dimensión de unicidad.

- **R7.13.11. Unicidad datos maestros.** La fuente tiene que garantizar que cada registro de la entidad maestra es único en el grupo de datos que certifica para ser duplicados en el Concentrado GDM.
- **R7.13.12. Unicidad datos de referencia.** Las fuentes que proveen los datos de referencia tienen que garantizar que cada registro de la entidad de referencia es único en el grupo de datos que certificados para ser duplicados en el Concentrado GDM.

8.2.5. Dimensión de Validez

El grado en que los datos se ajustan a las reglas de funcionalidad definidas por las entidades origen de los datos para que el contenido sea aceptable. Esto puede incluir: formato, patrón, tipo de datos, lista de valores válidos, dominio, alcance. Los perfiles que garantiza la validez de los datos: reglas de negocio o cálculos, datos para el rango de valores, secuencia inválida.

Regla asociada a la dimensión de validez.

- **R7.13.13. validez de los datos fuente.** Las fuentes que proveen los datos maestros o de referencia deben certificar que esos datos son resultado de aplicar correctamente reglas de negocio y definirlos bajo un formato que incluye tipo de dato, lista de valores, rango de valores, etc.

8.2.6. Dimensión de Oportunidad

El grado en que los cambios en los datos están disponibles dentro del marco de tiempo requerido por la entidad. Los perfiles que garantiza la identificación de falta oportunidad de los datos: los datos llegan tarde para ser

usados en un contexto de entidades consumidoras que requieran de la oportunidad de actualización del registro de oro en el concentrador, y retraso de tiempo en sistemas de tiempo real de origen.

Regla asociada a la dimensión de validez.

- **R7.13.13. Acuerdo de servicio.** El Administrador de Datos formalizará un acuerdo de servicio con el administrador de cada dominio para definir los tiempos de carga o extracción y los tiempos de consumo de los datos maestros.

8.2.7. Dimensión de Vigencia

El grado de vigencia de los datos se trata del reflejo de la condición de actualidad frente al estado capturado en el conjunto de datos. Los perfiles que pueden garantizar la vigencia de los datos: la dirección ha cambiado, número de teléfono desactualizado, lugar no existente, persona difunta.

Regla asociada a la dimensión de vigencia.

- **Regla 7.10.1.** Al insertar un registro en una tabla de una entidad maestra o de referencia, este debe ser registrado con inicio vigencia con el *timestamp* de la creación del registro
- **Regla 7.10.2.** Cuando la fuente defina que un registro termina su vigencia debe ser notificado al Concentrador GDM indicando: entidad maestra/referencia, tabla, el registro (tupla), y la fecha de fin de vigencia. De esta forma se ubicará de forma precisa la tupla en la cual será actualizado el campo que terminará su ciclo de vida en el Concentrador GDM.
- **Regla 7.10.3.** Registrar en la tabla **hub_historicocambios** el cambio de vigencia del registro afectado.

8.2.8. Dimensión de Conformidad

Es el grado de datos almacenados en un formato no estándar. Los valores de datos de los mismos atributos deben representarse en un formato y tipos de datos uniformes, tanto en la fuente como en el Concentrador GDM. Los perfiles que pueden garantizar la conformidad de los datos: conformidad de formato, conformidad del tipo de datos,

Regla asociada a la dimensión de conformidad.

- **R7.13.13. EstructuraFuenteDestino.** El ROR debe garantizar estructuras idénticas a las estructuras definidas por el ROO.

ROO:: Atributo (nombre, tipo, restricciones de calidad) = **ROR::Atributo** (nombre, tipo, restricciones de calidad)

8.2.9. Dimensión de Integridad

Es el grado en que los elementos de datos contienen contenido coherente en varias bases de datos. En otras palabras, es el grado en que se implementa una restricción relacional definida entre dos conjuntos de datos. Los perfiles que identifican la falta de calidad: integridad referencial y cardinalidad.

Regla asociada a la dimensión de vigencia.

- **R7.13.13. EstructuraFuenteDestino.** El ROO debe garantizar que la integridad referencial esté definida e incluida dentro de los datos y estructuras que provea para el Concentrador GDM.

8.2.10. Dimensión de Precisión

El grado en que los datos han sido redondeados o agregados. La precisión en la medición de la calidad de los datos es más un concepto derivado diseñado para verificar errores de redondeo y agregados en los datos. Los perfiles que garantizan precisión en los datos: precisiones numéricas, precisión de tiempo, precisión de granularidad.

Regla asociada a la dimensión de vigencia.

- **R7.13.14. CampofuenteDestino.** El ROO debe garantizar que la precisión de los datos numéricos cumpla con las reglas de negocio de la entidad fuente. La estructura de la entidad maestra o de referencial del ROR debe garantizar la misma precisión que provea el ROO.

8.3. Reglas de emparejamiento (matching)

Reglas de emparejamiento de columnas. Este tipo de regla se define para hacer coincidir registros de objetos base en función de los valores de las columnas que ha definido como columnas de coincidencia, como apellido, nombre, dirección, etc. Este es el método más utilizado para identificar coincidencias.

Esta regla solo se aplicará en el dominio de personas, específicamente en la maestra de personas, para la cual se haría el emparejamiento entre el registro de oro proporcionado por la entidad Registraduría y otras fuentes que capturen los mismos datos: identificación, primer nombre, segundo nombre, primer apellido, segundo apellido, fecha de nacimiento, tipo de documento, y fecha de expedición del documento. En el alcance de este proyecto solo se aplicaría la regla con datos de las personas naturales capturadas desde el RUES en el dominio de Empresas.

Reglas de emparejamiento de clave primaria. Se usa para hacer coincidir registros de dos sistemas⁷ que usan las mismas claves principales para los registros. No es común que dos sistemas de origen diferentes utilicen claves primarias idénticas. Sin embargo, cuando esto ocurre, las coincidencias de clave principal son rápidas y muy precisas.

Esta regla solo se aplicará en el dominio de personas, específicamente en la maestra de personas, para la cual se haría el emparejamiento entre el registro de oro proporcionado por la entidad Registraduría y otras fuentes que capturen los mismos datos: identificación, primer nombre, segundo nombre, primer apellido, segundo apellido, fecha de nacimiento, tipo de documento, y fecha de expedición del documento. En el alcance de este proyecto solo se aplicaría la regla con datos de las personas naturales capturadas desde el RUES en el dominio de Empresas.

Tipos de Emparejamiento:

- Emparejamiento Exacto. Los valores de emparejamiento deben ser exactos tanto para valores nulos como los no nulos.
- Emparejamiento Aproximado. Los valores que no son un emparejamiento exacto pero similar al valor con el que se compara. Las coincidencias están determinadas por los tokens de emparejamiento que comparten algunos valores y esto depende de la población registrada en la fuente que es quien debe definir dicha regla. Este tipo de regla se aplica normalmente en nombres o descriptores tipo cadena de caracteres que, aunque se escriben diferente significan o se entienden cómo lo mismo, y por eso se le asigna un token. Este tipo emparejamiento debe ser implementado en el Concentrador GDM para el caso del registro de referencia de Direcciones con el mayor grado de verdad.
- Emparejamiento filtrado. Este tipo de reglas usan el nivel de búsqueda “Estrecho”, porque la coincidencia aproximada debe ser lo suficientemente limitada para que los resultados sean los mismos que los de un emparejamiento exacto.

⁷ La comparación es binaria, así haya más de un sistema o esquema a ser comparado.

8.4. Reglas de Supervivencia

Las reglas de supervivencia definen un proceso de verificación de datos campo a campo correspondiente a cada registro definido para identificar la "única fuente de verdad", o el registro que captura toda la información pertinente y necesaria ya sea de personas, lugares, o empresas. Una vez que se verifican todos los datos reales en cada campo de cada esquema involucrado, se pueden pasar a una versión única, precisa y completa de cada registro de oro de la entidad maestra definida para cada dominio. En otras palabras, la supervivencia define qué datos sobreviven en los registros de oro.

Durante el proceso de comparación de registros o esquemas de las fuentes, se utilizan varias técnicas para seleccionar u obtener al mejor candidato posible. En este proceso se deben tener en cuenta la estructura y la fuente de los datos, cómo y qué se captura, y qué tipo de datos se almacenan, para realizar la supervivencia. Para seleccionar el registro sobreviviente se usan comúnmente tres técnicas:

- **Más reciente:** se pueden ordenar los registros con sello de fecha en orden del más al menos reciente.
- **Más frecuente:** coincide con registros repetidos que contienen la misma información
- **Más completo:** considera que el campo está completo como una indicación de la precisión del registro.

Específicamente en el alcance de este proyecto, las reglas o proceso de supervivencia no son necesarias ya que, en cada dominio, la entidad maestra almacena el registro de oro consultado o replicado del registro de verdad proporcionado por la entidad pública considerada como la única fuente de verdad proveedora del registro de oro. Las entidades que provean datos complementarios al registro de oro, como por ejemplo la dirección de ubicación de la persona, aplicarán reglas de emparejamiento sobre los campos del registro de oro a fin de asociar de forma correcta dicha dirección a la maestra de personas.

- Una regla de supervivencia define, para los atributos de las entidades con coincidencia aproximada y con coincidencia de clave, cómo se calculan los valores de los registros de oro. Está compuesto por:
 - Una regla de consolidación, que define cómo consolidar valores de registros duplicados (detectados por el comparador) en un solo registro de oro.
 - Una regla de anulación que define cómo los valores creados por los usuarios anulan el valor consolidado en el registro dorado.

Una regla de consolidación la cual define, utilizando una "Estrategia de Consolidación", cómo elegir los mejores valores en el proceso de consolidación.

8.5. Medición de la calidad de los datos

La medición de la calidad de los datos es la medición más representativa en un proceso de gestión de datos maestros y de referencia, así como medidas asociadas a la carga correcta de los datos maestros hacia el Concentrador GDM. La Tabla 24 presenta dos grupos de medidas representativas en la gestión de datos maestros y de referencia.

Tabla 29. Mediciones representativas para la gestión de datos maestros.

Actividad	Descripción
Calidad de los Datos	<p>Medición Calidad de Datos con enfoque “Proporción de éxito a fracaso”. <i>Esta es una relación entre el total de registros disponibles y los registros defectuosos encontrados por una de las dimensiones de calidad de datos.</i> En la Tabla 25 se muestra el cálculo de puntuación de los indicadores clave de rendimiento (<i>Key Performance Indicators - KPI</i>)</p> <p>Incluye mediciones basadas en las dimensiones de calidad (ver tabla 25)</p>
Ingesta y consumo de datos y servicios	<p>La confianza en la relación entre la cantidad de registros que ingresan sin errores desde la fuente y el total de registros entrantes por cada variable filtrada:</p> $\text{Confianza} = \frac{\text{Total registros válidos}}{\text{Total registros entrantes}}$ <p>El soporte es la relación entre la cantidad de registros que ingresan desde la fuente sin errores y el total teórico de medición por unidad de tiempo (día, semana, mes, etc.).</p> $\text{Soporte} = \frac{\text{Total registros válidos}}{\text{Valor teórico de la medición}}$

En la Tabla 25 se describen las dimensiones de calidad de datos que permitirán subdividir las reglas en grupos, cada uno con su forma particular de medirse (Plotkin, 2014).

Tabla 30. Dimensiones y Medición de Calidad de Datos.

Id_Dim	Dimensión de Calidad del dato	KPI
EXA	Exactitud	# de registros con datos inexactos / # total de registros
CTT	Compleitud	# de registros con datos incompletos / # total de registros
CON	Consistencia	# de registros con datos de inconsistencia / # total de registros
UCD	Unicidad	# de registros no únicos / # total de registros
VAL	Validez	# de registros con datos inválidos / # total de registros
CFD	Conformidad	# de registros con datos no confirmados / # total de registros
INT	Integridad	# de registros con datos con problemas de integridad / # total de registros
PRE	Precisión	# de registros con datos de imprecisión / # total de registros

9. Seguridad de Datos Maestros

(ISO/IEC 27001:2022, 2022) declara los requerimientos de seguridad asociados al manejo de la información. Las categorías de posibles riesgos de seguridad son los siguientes (Samonas & Coss, 2014).

1. **Confidencialidad.** Se refiere a la liberación de información no autorizada. Es el grado de autoridad que tiene una persona para leer y aprovechar la información almacenada en un repositorio de datos o en un sistema de archivos o de documentos. Esta categoría de preocupación a veces se extiende al "análisis de tráfico", en el que el intruso solo observa los patrones de uso de la información. A partir de esos patrones, el intruso puede inferir algún contenido de información. Esta categoría también incluye el uso no autorizado de un programa propietario.

2. **Integridad.** Se refiere a la modificación de información no autorizada: Es el grado de autoridad que tiene una persona para realizar cambios en la información almacenada.

3. **Disponibilidad.** Denegación de uso no autorizado. Es el grado de autoridad que tiene una persona para acceder o usar información.

En la Tabla 26 se presenta el impacto de cada uno de cada objetivo de seguridad de la información.

Tabla 31. Categorías de riesgos de seguridad de los datos.

Objetivo de seguridad	Impacto en la seguridad		
	Bajo	Moderado	Alto
Confidencialidad	Se puede esperar que la divulgación no autorizada de información tenga un efecto adverso limitado en las operaciones de la organización, los activos de la organización o las personas.	Se puede esperar que la divulgación no autorizada de información tenga un efecto adverso grave en las operaciones de la organización, los activos de la organización o las personas.	Se puede esperar que la divulgación no autorizada de información tenga un efecto adverso severo o catastrófico en las operaciones de la organización, los activos de la organización o las personas.
Integridad	Se puede esperar que la modificación o destrucción no autorizada de la información tenga un efecto adverso limitado en las operaciones de la organización, los activos de la organización o las personas.	Se puede esperar que la modificación o destrucción no autorizada de la información tenga un efecto adverso grave en las operaciones de la organización, los activos de la organización o las personas.	Se puede esperar que la modificación o destrucción no autorizada de la información tenga un efecto adverso severo o catastrófico en las operaciones de la organización, los activos de la organización o las personas.

Objetivo de seguridad	Impacto en la seguridad		
	Bajo	Moderado	Alto
Disponibilidad	Se podría esperar que la interrupción del acceso o uso de la información o un sistema de información tenga un efecto adverso limitado en las operaciones de la organización, los activos de la organización o las personas.	Se podría esperar que la interrupción del acceso o uso de la información o un sistema de información tenga un efecto adverso grave en las operaciones de la organización, los activos de la organización o las personas.	Se podría esperar que la interrupción del acceso o uso de la información o un sistema de información tenga un efecto adverso severo o catastrófico en las operaciones de la organización, los activos de la organización o las personas.

Para dar cumplimiento a la ley en lo que corresponde a la seguridad de los datos maestros del Gobierno de Colombia, la Ley 1712 de 2014 del gobierno de Colombia se refiere a la Información pública clasificada como “aquella información que estando en poder o custodia de un sujeto obligado en su calidad de tal, pertenece al ámbito propio, particular y privado o semiprivado de una persona natural o jurídica por lo que su acceso podrá ser negado o exceptuado, siempre que se trate de las circunstancias legítimas y necesarias y los derechos particulares o privados consagrados en el artículo 18 de esta ley.

Además, declara en el artículo 4°. Principios para el Tratamiento de datos personales, el cual incluye los siguientes principios:

- Principio de legalidad en materia de Tratamiento de datos
- Principio de finalidad
- Principio de libertad
- Principio de veracidad o calidad
- Principio de transparencia
- Principio de acceso y circulación restringida
- Principio de seguridad
- Principio de confidencialidad

En el contexto de la definición del Concentrador de Datos maestros, la arquitectura de referencia cuenta con un componente de arquitectura denominado seguridad y protección de datos. Este dispone de un conjunto de componentes que permiten autorizar a diferentes roles a realizar operaciones sobre los datos. Específicamente, facilita la definición de permisos y privilegios de los usuarios sobre los datos y las estructuras de datos. Además, el proceso del ciclo de vida de los datos maestros y de referencia define la réplica de datos maestros y de referencia desde las diferentes entidades proveedoras, como el mecanismo que asegurar que dichos datos no serán intervenidos en ninguno de los procesos definidos realizados al interior del Concentrador GDM. De esta forma, las entidades proveedoras deberán asegurar que cumplen toda la ley que cobije normatividad asociada a la seguridad y protección de los datos de los ciudadanos de Colombia.

10.Privacidad de los de Datos Maestros

El tratamiento de datos sensibles cada día toma mayor fuerza debido al incremento de datos personales expuestos constantemente al capturarse desde diferentes tecnologías, sistemas de información, formularios web, redes sociales, etc., de baja confiabilidad, o sin niveles de restricción de acceso. Esto facilita extracción no autorizada de datos confidenciales o sensibles de datos personales, financieros, etc.

La gestión de datos maestros del Gobierno Nacional definida para el Concentrador GDM estará sujeta al cumplimiento de las normas establecidas para protección de datos personales definidas en la Ley 1581 de 2012 (Departamento Administrativo de la Función Pública, 2012) y sus actualizaciones, reforzada con el decreto 1377 de 2013 (Ministerio de Industria y Comercio, 2013). Específicamente, esta ley define conceptos fundamentales que guiarán la identificación y manejo de los activos de datos con tratamiento especial. Estos activos deberán perfilarse con el registro de origen de cada una de las entidades participantes como proveedoras de datos. La ley define los siguientes conceptos:

- **“Datos sensibles:** Se entiende por datos sensibles aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquello que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición, así como los datos relativos a la salud, a la vida sexual, y los datos biométricos.
- **Transferencia:** La transferencia de datos tiene lugar cuando el Responsable y/o Encargado del Tratamiento de datos personales, ubicado en Colombia, envía la información o los datos personales a un receptor, que a su vez es Responsable del Tratamiento y se encuentra dentro o fuera del país.

El marco de gobierno propuesto para el Concentrador GDM debe estar alineado con los decretos establecidos para el tratamiento de datos sensibles que proveerán las entidades fuentes. Es decir, el gobierno de datos de las fuentes debe tener implementada la normatividad vigente expedida por el Gobierno Nacional para el tratamiento de datos. De esta forma, la “transferencia” de datos que se establezca entre las entidades fuentes y el Concentrador GDM mantiene el tratamiento de datos personales en el registro de oro del Dominio de Personas,

En el CONPES 3920 se compilan diferentes tipos de datos de acuerdo con las normatividades Ley 1266 de 2008, la Ley 1581 de 2012, la Ley 1712 de 2014 y el Decreto 1074 de 2015 específicas para datos con tratamiento especial durante la gestión de datos maestros. En la Tabla 27 se presenta la clasificación propuesta en el CONPES.

Tabla 32. Tipos de Datos definidos en el CONPES 3920.

Tipología	Descripción
Sensible	Afecta la intimidad y puede generar discriminación
Privado	Naturaleza íntima, solo es relevante para su titular
Semiprivado	Su conocimiento puede interesar a un grupo de persona o a la sociedad en general
Público	No pertenece a las tipologías anteriores o es definido por la ley
Abierto	Primario o sin procesar, en formatos estándar e interoperables, libre, y si restricciones. Puede ser usado y reutilizado
Clasificado	Puede generar daños a la intimidad, vida, salud, o seguridad de una persona; o a los secretos comerciales, industriales y profesionales.
Reservado	Acceso prohibido por la Constitución o la ley

Específicamente, Ley 1581 de 2012, los Datos sensibles se entiende por datos sensibles aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición así como los datos relativos a la salud, a la vida sexual y los datos biométricos.

Niveles de sensibilidad de los datos Maestro y de Referencia

- **Datos de alta sensibilidad.** Son datos que se comprometen o destruyen en una transacción autorizada, lo cual ocasionaría un impacto catastrófico en la organización o en las personas.
- **Datos de sensibilidad media,** Son datos destinados solo para uso interno, pero si se comprometen o destruyen, no tendrían un impacto catastrófico en la organización o las personas.
- **Datos de baja sensibilidad.** Son datos destinados al uso abierto (open data).

Las entidades proveedoras de datos maestros o de referencia para el Concentrador GDM, deben cumplir con el tratamiento de sus datos maestros a partir de la ley que regula los datos sensibles. Es potestad de cada entidad definir el grado de sensibilidad de sus datos maestros teniendo en cuenta los niveles antes presentados.

El Catálogo de datos, en los metadatos al nivel de columnas, provee un metadato etiquetado como “sensibilidad” que deberá ser proporcionado por las fuentes. El acuerdo entre la fuente y el Concentrador GDM deberá describir de forma clara y explícita el grado de sensibilidad por cada metadato proporcionado.

Ley 1266 de 2008 define las siguientes tipologías de datos:

- **Dato personal.** Es cualquier pieza de información vinculada a una o varias personas determinadas o determinables o que puedan asociarse con una persona natural o jurídica. Los datos impersonales no se sujetan al régimen de protección de datos de la presente ley. Cuando en la presente ley se haga referencia a un dato, se presume que se trata de uso personal. Los datos personales pueden ser públicos, semiprivados o privados;
- **Dato público.** Es el dato calificado como tal según los mandatos de la ley o de la Constitución Política y todos aquellos que no sean semiprivados o privados, de conformidad con la presente ley. Son públicos, entre otros, los datos contenidos en documentos públicos, sentencias judiciales debidamente ejecutoriadas que no estén sometidos a reserva y los relativos al estado civil de las personas;
- **Dato semiprivado.** Es semiprivado el dato que no tiene naturaleza íntima, reservada, ni pública y cuyo conocimiento o divulgación puede interesar no sólo a su titular sino a cierto sector o grupo de personas o a la sociedad en general, como el dato financiero y crediticio de actividad comercial o de servicios a que se refiere el Título IV de la presente ley.
- **Dato privado.** Es el dato que por su naturaleza íntima o reservada sólo es relevante para el titular.

En el catálogo de datos a nivel de columna se crean define la columna “Tipología del Dato Ley 1266 de 2008”.

Teniendo en cuenta la relevancia de la Seguridad y privacidad de la información en el Marco de la infraestructura de datos, se requiere que este tema sea desarrollado y articulado desde El coordinador de Datos con la Coordinación Nacional de Seguridad digital.

Glosario

Canalizador de datos (data pipeline), “Las canalizaciones de datos recopilan, transforman y almacenan datos para mostrarlos a las partes interesadas en una variedad de proyectos de datos.”

Temas para Refinamiento en Entregable 3

De acuerdo con las observaciones realizadas por el equipo DNP y MinTIC, los siguientes temas serán trabajados y detallados en el documento del diseño de la arquitectura de referencia para la gestión de datos maestros de la Fase 3.

Capítulos	Observación	Comentarios/Sugerencias	Entregable
4.3. Modelos de Datos	Refinamiento del capítulo	“Se puede incluir aca una mención a que los requerimientos de cada dominio, y las consideraciones a tener en cuenta para la definición de los modelos que están siendo planteados son abordados en el gobierno de datos del GDM o en su defecto que se extienda la definición que se realiza sobre la importancia que tiene el modelo en cada uno de los dominios y lo que representa.”	Pasa a Entregable Fase 3
	Documentación de las tablas de los modelos de datos	<p>“Frente al diseño de base de datos, con el fin de no hacer observaciones en cada una de las tablas haré un listado de ellas, si es posible hacerlo después de cada imagen de cada modelo:</p> <ul style="list-style-type: none"> -Colorcar un apartado de las siglas usadas en el modelo HUB, ROOR,ROO, etc. -Después de cada imagen, colocar cada tabla, su descripción y para se usa -Una descripción de la figura, explicación de la relación de las tablas y el porque se definió de esa manera -Seria necesario que se incluyera como es el manejo de la auditoria y manejo de históricos de cada modulo de base de datos. -En las tablas de diccionario de datos, incluir la columna "Obligatoriedad" y asi mismo diligenciar el valor para cada campo. -Sugiero hacer unas sesiones tecnicas de Caoba, Mintic y DNP para que se socialice el modelo, estará bien entenderlo y poder escuchar en los casos que se aplique propuestas de mejora u optimización (en este caso veo redundancia con la tabla de personas, se podría simplificar las tipologias o tipos, en un solo listado de valores de referencia, etc).” 	Pasa a Entregable Fase 3

Capítulos	Observación	Comentarios/Sugerencias	Entregable
4.5. Ciclo de Vida de los Datos Maestros	Redacción - Entendimiento	“Sugiero revisar redacción "El ciclo de vida de la gestión de los datos maestros se constituye como un elemento esencial para el funcionamiento adecuado de la arquitectura definida para la GDM, dado que permite analizar el ciclo de vida de los datos maestros a un nivel estratégico, táctico y operativo, lo que ayuda a identificar áreas de mejora en las entidades y organizaciones donde la GDM fortalece los procesos de gestión de datos" o la redacción que mejor se ajuste al lenguaje utilizado en el documento.”	Pasa a Entregable Fase 3
	Explicación de las actividades del proceso	“Creo que para cada una de las actividades del proceso es relevante colocar la información como se presenta para Integrar y Sincronizar datos maestros por cada dominio - es decir que se mencione la relevancia de la actividad y luego se explique cual es el proceso que se realiza en cada una de ellas.”	Pasa a Entregable Fase 3
		“Se debe especificar quien o quienes son los responsables de cada actividad del proceso planteado.”	Pasa a Entregable Fase 3
	Aclaración en la actividad: Limpiar datos consolidados	"Una vez se tengan las soluciones adecuadas para la recopilación y la consolidación en el sistema híbrido de GDM, se pasa a la siguiente fase." “A que tipo de soluciones se refiere? de datos? Detallar.”	Pasa a Entregable Fase 3
	Complementar descripción de la actividad: Limpiar datos consolidados -Prevención. Evitar que datos incorrectos o erróneos sean almacenados en la bases de datos de su organización -Detección. Tomar un enfoque proactivo	“Se puede incluir aquí una mención a las principales herramientas que pueden ser empleadas? A parte del comentario anterior, también se sugiere que la prevención se describa en función del riesgo, podría ser: "Evitar que datos incorrectos o erróneos sean almacenados en la bases de datos de su organización".”	Pasa a Entregable Fase 3

Capítulos	Observación	Comentarios/Sugerencias	Entregable
	<p>para identificar datos incorrectos que ya están en las bases de datos.</p> <p>-Reparación. Arreglar o borrar los datos incorrectos.</p>		
6. Catálogo de Datos Maestros	Refinamiento del capítulo		Queda en Entregable Fase 2
7. Calidad de la Gestión de Datos Maestros	Refinamiento del capítulo	<p>“Agradecemos Incluir un párrafo que describa la sección 7.2. Reglas y Estándares de Calidad</p> <p>CAOBA: Se harán los ajustes de este numeral en el documento entregable de la Fase 3 debido a que se dependía de la aprobación de los modelos de datos propuestos para describir los estándares y procedimientos de calidad.”</p>	Queda en Entregable Fase 2
8. Seguridad de Datos Maestros	Refinamiento del capítulo	<p>“Revisar, en este punto se seguridad, como se pueden mitigar los impactos mencionados.</p> <p>También es importante mencionar como se presenta o se refleja</p> <p>CAOBA: Se debe analizar con el equipo si la solicitud está dentro del alcance de la Fase 2 a la AUDITORIA de los datos para estas situaciones de impacto y en general.”</p>	Queda en Entregable Fase 2

Referencias

- DANE. (15 de Septiembre de 2022). *Directorio Estadístico*. Obtenido de <https://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-informacion/registro-estadistico-de-empresas>
- Eurostat. (2010). *Manual de recomendaciones de Registros de Empresas*.
- ISO. (2012). *ISO standar 19152*. Obtenido de Geographic information — Land Administration Domain Model (LADM): <https://www.iso.org/standard/51206.html>
- IGAC. (2022). *IGAC - LADM_COL*. Obtenido de <https://igac.gov.co/es/catastro-multiproposito/ladm-col>
- DNP. (2012). Bogotá.
- Horst, N. L., Bjerre, S., Lind, M., & Hvingel, L. (2014). The Basic Data Programme – A Danish Infrastructure Model for Public Data. *PERSPEKTIV NO. 24*.
- Denmark, E. C. (2021). Digital Public Administration factsheet 2021 - Denmark.
- Norway, E. C. (2020). *Digital Government Factsheet 2019 Norway*.
- Entur. (30 de 8 de 2022). *Stops- and Timetable data*. Obtenido de <https://developer.entur.org/stops-and-timetable-data>
- Helleland, L. H. (2021). *Norwegian data centres - sustainable, digital powerhouses*. Norwegian Ministry of Local Government and Modernisation.
- NordREG. (2021). *Implementation of data hubs in the Nordic countries*. Oslo: Nordic Electric Regulators.
- Norway, European Commision. (2022). *Digital Public Administration 2021*. Ministry of Local Government and Modernisation. (2019). *One Digital Public Sector - Digital strategy for the public sector 2019–2025*. Oslo.
- OECD. (2015). *Open Government Data Review of Poland: Unlocking the Value of Government Data*.
- OSCE. (2017). *compendium of good practices in identity management in the osce region*. Miodowa, Poland: OSCE Office for Democratic Institutions and Human Rights (ODIHR).
- OSCE-ODIHR. (2009). *Guidelines on Population Registration*. Warsaw, Poland: Homework.
- Danish Government; Local Government Denmark. (2012). *Good Basic Data for Everyone – A Driver for Growth and Efficiency*. Denmark.
- Ng, S. T., Xu, F. J., Yang, Y., & Lu, M. (2017). A Master Data Management Solution to Unlock the Value of Big Infrastructure Data for Smart, Sustainable and Resilient City Planning. *Procedia Engineering, 196*, 939-947. doi:<https://doi.org/10.1016/j.proeng.2017.08.034>
- Haneem, Kama, F. &, Azmi, N. &, Azizan, A. &, sam, A. &, Yusop, S. &, . . . Hafiza. (2017). Master Data Definition and the Privacy Classification in Government Agencies: Case Studies of Local Government. *Advanced Science Letters, 23*. doi:10.1166/asl.2017.7317
- Haneem, F., Kama, N., Taskin, N., Pauleen, D., & Bakar, N. A. (2019). Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management, 45*, 25-43. doi:<https://doi.org/10.1016/j.ijinfomgt.2018.10.007>
- Qodarsih, Yudhoatmojo, N. &, Hidayanto, S. &, & Achmad. (2018). Master Data Management Maturity Assessment: A Case Study in the Supreme Court of the Republic of Indonesia. 1-7. doi:10.1109/CITSM.2018.8674373

- Naufal, Kusumasari, & M., Alam, T. &, & Ekky. (2020). Analysis and Design of Master Data Monitoring Application using Open Source Tools: A Case Study at Government Agency. 196-200. doi:10.1109/ICOIACT50329.2020.9332155
- Pratama, Astana, F. &, Yudhoatmojo, S. &, Hidayanto, S. &, & Achmad. (2018). Master Data Management Maturity Assessment: A Case Study of Organization in Ministry of Education and Culture. 1-6. doi:10.1109/IC3INA.2018.8629524
- Kiggundu, S. A., & Okike, E. U. (November de 2015). An architectural design framework for Population Registration and National Identification System in Uganda. *Data modeling International Journal of Computer Science and Information Security (IJCSIS)*, 13. Obtenido de <https://sites.google.com/site/ijcsis/>
- Karan, E. P., Irizarry, J., & Haymaker, J. (2015). *BIM and GIS integration and interoperability based on semantic web technology*.
- DNP; MinTIC; SIC. (2019). *CONPES 3975 Política Nacional para la Transformación Digital e Inteligencia Artificial*. Bogotá. Obtenido de <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3975.pdf>
- DNP; MinTIC; SIC. (2018). *CONPES 3920 Política Nacional de Explotación de Datos (Big data)*. Bogotá. Obtenido de <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>
- DNP, MinInterior, MinHacienda, MinAgricultura, MinSalud, MinTrabajo, . . . MinCiencias. (2021). *CONPES 4023 Política para la Reactivación, la Repotenciación y el Crecimiento Sostenible*. Obtenido de <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4023.pdf>
- DAMA International. (2017). *DMBOK: Data Management Body of Knowledge*. Basking Ridge, New Jersey: Technics Publications.
- MinTIC. (2022). *Plan Nacional de Infraestructura de Datos - Resolución 460 de 2022*. Obtenido de https://www.mintic.gov.co/portal/715/articulos-198952_resolucion_00460_2022.pdf
- DNP - CONPES 3958. (2019). *CONPES 3958 - ESTRATEGIA PARA LA IMPLEMENTACIÓN DE LA POLÍTICA PÚBLICA DE CATASTRO MULTIPROPÓSITO*. Bogotá.
- DANE - Guía Diseño RA AEI. (2021). *GUÍA PARA EL DISEÑO DE REGISTROS ADMINISTRATIVOS DE ACTIVIDADES ECONÓMICAS INFORMALES*.
- DANE - Resolución 114. (2020). *RESOLUCIÓN 114 Por la cual la Unidad Administrativa Especial Dirección de Impuestos y Aduanas Nacionales -DIAN adopta la Clasificación de Actividades Económicas CIIU Rev. 4 A.C. (2020)*.
- MinTIC Resolución 460 PNID. (2022). *Plan Nacional de Infraestructura de Datos - Resolución 460 de 2022*. Obtenido de https://www.mintic.gov.co/portal/715/articulos-198952_resolucion_00460_2022.pdf
- DANE GIT DEST. (2022). *Sistema de Registros Estadísticos*.
- DANE - METODOLOGÍA REG ADMIN. (2021). *METODOLOGÍA GENERAL PARA LA CONSTRUCCIÓN Y ACTUALIZACIÓN DEL REGISTRO ESTADÍSTICO BASE DE POBLACIÓN – REBP 2020*.
- CCMA - Glosario. (s.f.). *GLOSARIO - TÉRMINOS QUE USA LA CCMA O QUE TIENEN RELACIÓN CON SU ACTIVIDAD*. Medellín. Obtenido de <https://www.camaramedellin.com.co/Portals/0/Transparencia/Documentos/Glosario.pdf>
- MinTIC - Marco de interoperabilidad. (2019). *Marco de interoperabilidad para Gobierno Digital*.
- MinTIC - G.INF.02. (2019). *G.INF.02 Guía técnica de Información- Administración del dato maestro*. Colombia: MinTIC. Obtenido de https://www.mintic.gov.co/arquitecturati/630/articulos-9254_recurso_pdf.pdf

- DANE - Guía para carga a las fuentes. (2020). *GUÍA PARA LA REDUCCIÓN DE LA CARGA A LAS FUENTES*.
- MinTIC - Resolución 460 PNID. (2022). *Resolución 460 de 2022 Plan Nacional de Infraestructura de Datos Colombia*. Obtenido de https://www.mintic.gov.co/portal/715/articulos-198952_resolucion_00460_2022.pdf
- DNP - Comité Política Catastro Multipropósito. (2022). *Comité de Implementación de la Política de Catastro Multipropósito (ppt)*. Bogotá.
- IGAC - Portal. (2022). *IGAC - Que estamos haciendo?* Obtenido de <https://www.igac.gov.co/es/contenido/areas-estrategicas/catastro/que-estamos-haciendo>
- IGAC - Avances Catastro Multipropósito. (2021). *Avances en la implementación de la política de Catastro Multipropósito*.
- Wallgren, Wallgren, A., & Britt, y. (2021). *Hacia un sistema estadístico integrado y basado en registros*. New York: BID.
- DANE - Convenio DANE y Registraduría. (2022). *Dane Información para todos*. Obtenido de En marcha convenio entre el DANE y la Registraduría: <https://www.dane.gov.co/index.php/actualidad-dane/4403-en-marcha-convenio-entre-el-dane-y-la-registraduria>
- DANE - REBP. (2018). *Registro Estadístico Base Población REBP*. Colombia. Obtenido de <https://www.dane.gov.co/files/investigaciones/poblacion/informes-estadisticas-sociodemograficas/2021-10-01-Registro-Estadistico-Base-Poblacion-REBP-2018.pdf>
- DANE - Informe Estadística Sociodemográfica. (2022). *Informes de Estadística Sociodemográfica Número 14 - Principales ciudades y sus conurbaciones en Colombia*. Bogotá: DANE.
- IGAC - Subdirección de Geografía y Cartografía. (2022). *IGAC - Subdirección de Geografía y Cartografía*. Obtenido de <https://www.igac.gov.co/es/contenido/areas-estrategicas/subdireccion-de-geografia-y-cartografia>
- IGAC - Reporte de Información Catastral. (2022). *REPORTE DE INFORMACIÓN CATASTRAL*. Obtenido de <https://igac.gov.co/catastro-multiproposito/reporte-de-informacion-catastral>
- MinSalud. (2022). *Planilla Integrada de Liquidación de Aportes*. Obtenido de <https://www.minsalud.gov.co/proteccionsocial/Paginas/pila.aspx>
- Agency for Data Supply and Infrastructure. (2022). *The Danish Basic Data Program, the Data-Distributor (Datafordeleren) and the Covid-19 Pandemic*. Genova.
- Jonker, R., Kooistra, F., Cepariu, D., Etten, J. v., & Swartjes, S. (2011). *Effective master data management*. KPMG.
- Otto, B., Hüner, K., & Österle, H. (2012). Toward a functional reference model for master data quality management. *Information Systems and e-Business Management*, 10, pages395–425. doi:<https://doi.org/10.1007/s10257-011-0178-0>
- Reichert, A., Otto, B., & Österle, H. (2013). A Reference Process Model for Master Data Management. *11th International Conference on Wirtschaftsinformatik*, (págs. 817-830). Leipzig, Germany.
- Gamero, A., García, J., & Raymundo, C. (2019). Reference Model with a Lean Approach of Master Data Management in the Peruvian Microfinance Sector. *2019 8th International Conference on Industrial Technology and Management*, (págs. 56-60).
- Allen, M., & Cervo, D. (2015). *Multi-Domain Master Data Management*. Morgan Kaufmann.
- Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, 15(B), 1068-1076. doi:<https://doi.org/10.1016/j.chb.2014.09.030>

- Chrissis, M. B., Konrad, M., & Shrum, S. (2011). *CMMI for Development: Guidelines for Process Integration and Product Improvement (SEI Series in Software Engineering) 3rd Edición*. Addison-Wesley Professional.
- NIFO - National Interoperability Framework Observatory. (25 de Oct de 2022). Obtenido de <https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/digital-public-administration-factsheets>
- Kroon, J., & Paas, A. (2014). *Using registers as the basis to statistical production – linking issues and potential bottlenecks*. IFC Bulletin No 37.
- Vassil, K. (2015). *Estonian e-Government Ecosystem - Foundation, applications, outcomes*. Institute of Government and Politics - University of Tartu.
- European Commission - Estonia. (2022). *Digital Public Administration factsheet 2022 - Estonia*.
- European Commission - Poland. (2022). *Digital Public Administration factsheet 2022 Poland*.
- Ko, C., Adywiratama, A. D., Hidayanto, & Nizar, A. (2021). Master Data Management Maturity Model (MD3M) Assessment: A Case Study in Secretariat of Presidential Advisory Council. *2021 9th International Conference on Information and Communication Technology (ICoICT)*, (págs. 359-362).
- Qodarsih, N., Yudhoatmojo, S. B., & Hidayanto, A. N. (2018). Master Data Management Maturity Assessment A Case Study in the Supreme Court of the Republic of Indonesia. *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*.
- A., A. R., P., G. D., R., M. F., A., N. F., B., P. A., & Ruldeviyani, Y. (2019). Master Data Management Maturity Assessment : A Case Study of A Pasar Rebo Public Hospital. *ICACISIS 2019*, (págs. 47-503).
- Krismawati, D., Ruldeviyani, Y., & Rusli, R. (2019). Master Data Management Maturity Model: A Case Study at Statistics Business Register in Statistics Indonesia. *2019 International Conference on Information and Communications Technology (ICOIACT)*, (págs. 931-936).
- Rishartati, P., Rahayuningtyas, N. D., Maulina, J., Adetia, A., & Ruldeviyani, Y. (2019). Maturity Assessment and Strategy to Improve Master Data Management of Geospatial Data Case Study: Statistics Indonesia. *2019 5th International Conference on Science and Technology (ICST)*. Yogyakarta, Indonesia.
- DNP. (2017). *Análisis de duplicidad en la información recolectada por medio de Registros Únicos*.
- DANE-RegAdmitivos. (12 de 10 de 2022). <https://www.dane.gov.co/index.php/sistema-estadistico-nacional-sen/registros-administrativos/programa-de-fortalecimiento>.
- DANE. (2018). *Guía de Estándares para Registros Administrativos*.
- DANE Conceptos estandarizados. (10 de 10 de 2022). *SISTEMA DE CONSULTA DE CONCEPTOS ESTANDARIZADOS*. Obtenido de <https://conceptos.dane.gov.co/conceptos/transversal>
- ICDE Infraestructura . (10 de 10 de 2022). *ICDE Infraestructura colombiana de datos espaciales*. Obtenido de <https://www.icde.gov.co/diccionario-de-terminos/sistema-nacional-de-informacion-catastral-sinic>
- Departamento Administrativo de la Función Pública. (1970). *DECRETO 1260 DE 1970*. Obtenido de https://www.funcionpublica.gov.co/eva/gestornormativo/norma_pdf.php?i=8256
- ICDE. (10 de 10 de 2022). *ICDE - Datos Fundamentales*. Obtenido de <https://www.icde.gov.co/datos-y-recursos/datos-fundamentales>
- MinTIC - Marco de Referencia de Arquitectura Empresarial. (10 de 10 de 2022). *Marco de Referencia de Arquitectura Empresarial*. Obtenido de <https://www.mintic.gov.co/arquitecturati/630/w3-propertyvalue-8161.html>

- DANE - SEN. (10 de 10 de 2022). *Sistema Estadístico Nacional (SEN)*. Obtenido de <https://www.dane.gov.co/index.php/sistema-estadistico-nacional-sen/normas-y-estandares/nomenclaturas-y-clasificaciones/nomenclaturas/codificacion-de-la-division-politica-administrativa-de-colombia-divipola>
- DAMA Internacional. (10 de 10 de 2022). *DAMA Internacional*. Obtenido de <https://www.dama.org/cpages/home>
- Registraduría . (2022). *ACCESO A LA BASE DE DATOS ANI*. Obtenido de <https://www.registraduria.gov.co/-Consultas-ANI-.html>
- MinSalud. (2022). *ADRES*. Obtenido de <https://www.adres.gov.co/entidades-territoriales/bdua>
- Función Pública. (2022). Decreto 1389 de 2022 - "lineamientos generales para fortalecer la gobernanza de infraestructura de datos y crear el Modelo de Gobernanza de la infraestructura de Datos".
- Kendrick, T. (2013). *The Project Management Tool Kit, 3rd Edition*. AMACOM.
- Plotkin, D. (2014). *Data Stewardship - An Actionable Guide to Effective Data Management and Data Governance*. Elsevier Inc.
- Cámara de Comercio Bogotá. (27 de 12 de 2022). *Formulario del Registro Único Empresarial Social (RUES)*. Obtenido de Trámites y Consultas: <https://www.ccb.org.co/Tramites-y-Consultas/Mas-informacion/Formulario-del-Registro-Unico-Empresarial-y-Social-RUES>
- eInforma. (27 de 12 de 2022). *Directorio de Empresas de Colombia*. Obtenido de <https://directorio-empresas.einforma.co/informacion-empresa/importadora-bombonera-he-sas>
- IGAC-LADM_COL. (4 de 10 de 2022). *Diccionario de datos Modelo de Intercambio Catastro Registro LADM_COL V. 1.0*. Obtenido de https://gitlab-ladm-col.igac.gov.co/root/LADM_COL/-/blob/master/Catastro_Multiproposito/2_Aplicacion/docs/Diccionario_de_Datos_Modelo_de_Aplicacion_Intercambio_Catastro_Registro_v1.0.pdf
- IGAC-Repositorio. (7 de 10 de 2022). *LADM-COL*. Obtenido de <https://igac.gov.co/es/catastro-multiproposito/ladm-col>
- DNP-PNID. (2021). *Plan Nacional de Infraestructura de Datos - Documento Técnico y Hoja de Ruta*. Bogotá: MinTIC, DNP, Presidencia de la República.
- DNP-PNID. (2021). *Plan Nacional de Infraestructura de Datos - Documento Técnico y Hoja de Ruta*. Bogotá: MinTIC, DNP, Presidencia de la República.
- Wells, D. (1 de 2020). *Introduction to Data Catalogs*. Obtenido de <https://www.alation.com/wp-content/uploads/dave-wells-intro-to-data-catalogs-alation.pdf>
- CMS. (15 de 1 de 2022). *Business Glossary Quick Reference Guide*. Obtenido de <https://www.cms.gov/files/document/business-glossary-quick-reference-guide.pdf>
- CMS-DD. (15 de 1 de 2022). *Data Definition Quick Reference Guide*. Obtenido de <https://www.cms.gov/files/document/data-definition-quick-reference-guide.pdf>
- ISO/IEC 27001:2022. (13 de 12 de 2022). *ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements*. Obtenido de <https://www.iso.org/standard/82875.html>
- Samonas, S., & Coss, D. (2014). THE CIA STRIKES BACK: REDEFINING THE CIA STRIKES BACK: REDEFINING IN SECURITY. *Journal of Information System Security*, 10(3), 21-45.
- Departamento Administrativo de la Función Pública. (2012). *Ley 1581 de 2012 - Protección de Datos Personales*.
- Ministerio de Industria y Comercio. (2013). *Decreto Régimen General de Protección de Datos*.
- Zaidi, E., Simoni, G. D., Edjlali, R., & Duncan, A. D. (2017). *Data Catalogs Are the New Black in Data Management and Analytics*. Gartner.

Talent. (27 de 2 de 2023). *What is a data catalog, and do you need one?* Obtenido de <https://www.talend.com/resources/what-is-data-catalog/#:~:text=Gartner%20defines%20the%20data%20catalog,description%2C%20and%20organization%20of%20datasets>.